J. Jpn. Soc. Comp. Statist., **xx**(200x), 1–32

SIMULTANEOUS ESTIMATION OF SCALE MATRICES IN TWO-SAMPLE PROBLEM UNDER ELLIPTICALLY CONTOURED DISTRIBUTIONS

Hisayuki Tsukuma* **and Yoshihiko Konno†**

Number of pages: Text (18), References (2), Tables (10), Figures (0), Total (32)

Abbreviated title: Estimating scale matrices under elliptically contuored distributions

Corresponding author: Hisayuki Tsukuma,

E-mail: tsuku@kk.iij4u.or.jp

*Graduate School of Science and Technology, Chiba University, 1-33 Yayoi-cho, Chiba 263-8522, Japan
E-mail: 97um0204@g.math.s.chiba-u.ac.jp

†Graduate School of Science and Technology, Chiba University, 1-33 Yayoi-cho, Chiba 263-8522, Japan
E-mail: konno@math.s.chiba-u.ac.jp

Key words: covariance matrix, Stein's loss, Stein-Haff identity, two-sample problem

ABSTRACT

Two-sample problems of estimating $p \times p$ scale matrices are investigated under elliptically contoured distributions. Two loss functions are employed; one is sum of Stein's loss functions of one-sample problem of estimating a normal covariance matrix and the other is a quadratic loss function for $\Sigma_2\Sigma_1^{-1}$, where Σ_1 and Σ_2 are $p \times p$ scale matrices of elliptically contoured distribution models. It is shown that improvement of the estimators obtained under the normality assumption remains robust under elliptically contoured distribution models. A Monte Carlo study is also conducted to evaluate the risk performances of the improved estimators under three elliptically contoured distributions.

1. Introduction

Since the pioneer works of Stein (1956) and James and Stein (1961), there has been a great deal of effort to construct improved estimators for a covariance matrix of a multivariate normal distribution. The literature includes Haff (1980, 1982, 1991) and Dey and Srinivasan (1985). Two sample analogue of estimating covariance matrices has been also considered by several authors such as Muirhead and Verathaworn (1985) and Loh (1991a, 1991b). On the other hand Kubokawa and Srivastava (1999) showed that improvement of minimax estimators for a covariance matrix obtained under the normality assumption remains robust under elliptically contoured distribution models. In this paper, following the set-up considered by Loh (1991a, 1991b), we examine two-sample problems of estimating scale matrices of elliptically contoured distributions.

The precise set-up of the problems is as follows: Let $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$ be $N_1 \times p$ and $N_2 \times p$ random matrices which take multivariate linear models of the form

$$\boldsymbol{Y}_1 = \boldsymbol{C}_1\boldsymbol{\beta}_1 + \boldsymbol{\epsilon}_1 \qquad \text{and} \qquad \boldsymbol{Y}_2 = \boldsymbol{C}_2\boldsymbol{\beta}_2 + \boldsymbol{\epsilon}_2. \tag{1}$$

Here $\boldsymbol{\epsilon}_i\,(i = 1,\, 2)$ are $N_i \times p$ random matrices, $\boldsymbol{C}_i$ are known $N_i \times m$ matrices with full rank, and $\boldsymbol{\beta}_i$ are unknown $m \times p$ matrices. We also assume that the error matrices $\boldsymbol{\epsilon}_1$ and $\boldsymbol{\epsilon}_2$ are marginally distributed as elliptically contoured distributions. But we assume the two forms of the joint density functions of error matrices: First, two error matrices $\boldsymbol{\epsilon}_1$ and $\boldsymbol{\epsilon}_2$ are independently distributed and have the joint density function

$$\Pi_{i=1}^2 |\boldsymbol{\Sigma}_i|^{-N_i/2} g_i(\mathrm{tr}(\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\epsilon}_i'\boldsymbol{\epsilon}_i)), \tag{2}$$

where $\boldsymbol{\Sigma}_i\,(i = 1,\, 2)$ are $p \times p$ unknown positive definite matrices and g_i are nonnegative real-valued functions: Secondly two error matrices $\boldsymbol{\epsilon}_1$ and $\boldsymbol{\epsilon}_2$ are uncorrelatedly distributed and have the joint density function

$$|\boldsymbol{\Sigma}_1|^{-N_1/2}|\boldsymbol{\Sigma}_2|^{-N_2/2} g(\mathrm{tr}(\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\epsilon}_1'\boldsymbol{\epsilon}_1 + \boldsymbol{\Sigma}_2^{-1}\boldsymbol{\epsilon}_2'\boldsymbol{\epsilon}_2)), \tag{3}$$

where g is nonnegative real-valued function. Here $|\boldsymbol{P}|, \mathrm{tr}(\boldsymbol{P})$ and $\boldsymbol{P}'$ stand for the determinant, the trace and the transpose of a square matrix $\boldsymbol{P}$, respectively.

Following the approaches due to Loh (1991a, 1991b), we consider two sets of estimation problems as follows.

(i) Under the model (1) with the assumption (2), the problem of estimating $(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2)$ with unknown parameters $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$ is considered under a loss function

$$L_1(\widehat{\boldsymbol{\Sigma}}_1, \widehat{\boldsymbol{\Sigma}}_2, \boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2) = \sum_{i=1}^{2} \{\mathrm{tr}(\widehat{\boldsymbol{\Sigma}}_i \boldsymbol{\Sigma}_i^{-1}) - \log |\widehat{\boldsymbol{\Sigma}}_i \boldsymbol{\Sigma}_i^{-1}| - p\}, \tag{4}$$

where $\widehat{\boldsymbol{\Sigma}}_i$, $i = 1, 2$, are estimators of $\boldsymbol{\Sigma}_i$, respectively. This loss function is a natural extension of Stein's loss function in the one-sample case.

(ii) Under the model (1) with the assumption (3), the problem of estimating $\boldsymbol{\zeta} = \boldsymbol{\Sigma}_2 \boldsymbol{\Sigma}_1^{-1}$ with unknown parameters $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$ is considered under a loss function

$$L_2(\widehat{\boldsymbol{\zeta}}, \boldsymbol{\zeta}) = \mathrm{tr}\{\boldsymbol{\Sigma}_2^{-1}(\widehat{\boldsymbol{\zeta}} - \boldsymbol{\zeta})\boldsymbol{S}_1(\widehat{\boldsymbol{\zeta}} - \boldsymbol{\zeta})'\}/\mathrm{tr}\boldsymbol{\zeta}, \tag{5}$$

where $\widehat{\boldsymbol{\zeta}}$ is an estimator of $\boldsymbol{\zeta}$ and $\boldsymbol{S}_1 = \boldsymbol{Y}_1'(\boldsymbol{I}_{N_1} - \boldsymbol{C}_1(\boldsymbol{C}_1'\boldsymbol{C}_1)^{-1}\boldsymbol{C}_1')\boldsymbol{Y}_1$. This estimation problem is related to estimation of the common mean of two multivariate distributions. See a possible motivation for Loh (1991b). Furthermore, the eigenvalues of $\boldsymbol{\zeta}$ are important, for example, in the problem of testing the null hypotheses $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$ against the alternative hypotheses $\boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2$. For estimating these eigenvalues, see Muirhead and Verathaworn (1985), Muirhead (1987), and DasGupta (1989).

This paper is organized in the following way. In Section 2, we treat the problem (i). We adapt the extended Stein and Haff identity due to Kubokawa and Srivastava (1999) for two sample set-up (which is stated in Section 4) and obtain sufficient conditions under which an alternative estimator improves upon the James-Stein estimator $(\boldsymbol{T}_1\boldsymbol{D}_1\boldsymbol{T}_1', \boldsymbol{T}_2\boldsymbol{D}_2\boldsymbol{T}_2')$ with respect to the loss function (4). Here $\boldsymbol{T}_i$, $i = 1, 2$, is the lower triangular matrix with positive diagonal elements such that $\boldsymbol{S}_i = \boldsymbol{T}_i\boldsymbol{T}_i'$ and $\boldsymbol{D}_i$ is diagonal matrix with the j-th diagonal element $1/(N_i - m + p + 1 - 2j)$, $j = 1, 2, \ldots, p$, where $\boldsymbol{S}_i = \boldsymbol{Y}_i'(\boldsymbol{I}_{N_i} - \boldsymbol{C}_i(\boldsymbol{C}_i'\boldsymbol{C}_i)^{-1}\boldsymbol{C}_i')\boldsymbol{Y}_i$, $i = 1, 2$. Simulation study is conducted to evaluate risk performances of alternative estimators under the multivariate normal distribution, the matrix-variate t-distribution, and the matrix-variate Kotz-type distribution. Since these matrix-variate distributions except the normal distribution are not independent sampling, we also conduct simulation study based on independently and identically sampling model from the multivariate t-distribution and the Kotz-type distribution, respectively. Finding in this Section is that the estimators obtained under the error distribution (2) (i.e., which is different from independently and identically sampling model) perform well under independently and identically sampling from the elliptically contoured error models. In Section 3, we treat the problem (ii). In this problem, we treat the joint density function (3) only since we fail to obtain the suitable integration-by-parts formula under the joint density (2) to get improved estimators. We first obtain the best estimator among the constant multiple of $\boldsymbol{S}_2\boldsymbol{S}_1^{-1}$. Next we consider several types of improvement over the best constant multiple of $\boldsymbol{S}_2\boldsymbol{S}_1^{-1}$ and conduct simulation study in the much same way as that of Section 2. The proofs of the results obtained in Sections 2 and 3 put into Section 4.

2. Simultaneous estimation of $(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2)$

To consider the estimation problem of $(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2)$, we shall employ the loss function (4) and evaluate performance of estimators of $(\widehat{\boldsymbol{\Sigma}}_1, \widehat{\boldsymbol{\Sigma}}_2)$ by means of their risk function, i.e., $\boldsymbol{E}[L_1(\widehat{\boldsymbol{\Sigma}}_1,\, \widehat{\boldsymbol{\Sigma}}_2,\, \boldsymbol{\Sigma}_1,\, \boldsymbol{\Sigma}_2)]$, where the expectation is taken with respect to the joint distributions of two error distributions (2).

2.1. Class of estimators

First we introduce estimators obtained from one-sample problem of estimating a normal covariance matrix. We define the usual estimator

$$(\widehat{\boldsymbol{\Sigma}}_1^{US}, \widehat{\boldsymbol{\Sigma}}_2^{US}) = (\boldsymbol{S}_1/n_1, \boldsymbol{S}_2/n_2), \tag{6}$$

where $n_i = N_i - m$, $i = 1,\, 2$. Also we put the James-Stein estimator

$$(\widehat{\boldsymbol{\Sigma}}_1^{JS},\, \widehat{\boldsymbol{\Sigma}}_2^{JS}) = (\boldsymbol{T}_1\boldsymbol{D}_1\boldsymbol{T}_1',\, \boldsymbol{T}_2\boldsymbol{D}_2\boldsymbol{T}_2'), \tag{7}$$

where $\boldsymbol{T}_i$, $i = 1,\, 2$, is the lower triangular matrix with positive diagonal elements such that $\boldsymbol{S}_i = \boldsymbol{T}_i\boldsymbol{T}_i'$ and $\boldsymbol{D}_i$ is diagonal matrix with the j-th diagonal element $1/(n_i + p + 1 - 2j)$, $j = 1,\, 2, \ldots, p$. Note that the James-Stein estimator (7) is invariant under the group of transformations given by

$$\boldsymbol{\Sigma}_i \to \boldsymbol{P}_i\boldsymbol{\Sigma}_i\boldsymbol{P}_i', \qquad \boldsymbol{S}_i \to \boldsymbol{P}_i\boldsymbol{S}_i\boldsymbol{P}_i', \qquad i = 1,\, 2,$$

where $\boldsymbol{P}_i$ is any $p \times p$ lower triangular matrix with positive diagonal elements. ¿From the argument from James and Stein (1961), we can see that the estimator (7) has smaller risk than that of the estimator (6). Furthermore, applying Proposition 1 in Kubokawa and Srivastava (1999), we can see that improvement of the estimator (7) over the usual estimator (6) remains robust for all possible functions g_1 and g_2 in (2). To improve upon the estimator

(7) by using both $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ simultaneously, we adapt argument due to Loh (1991) and consider a class of invariant estimator under the group of the transformations

$$\boldsymbol{\Sigma}_i \to \boldsymbol{Q}\boldsymbol{\Sigma}_i\boldsymbol{Q}', \qquad \boldsymbol{S}_i \to \boldsymbol{Q}\boldsymbol{S}_i\boldsymbol{Q}', \qquad i = 1,\, 2, \tag{8}$$

where $\boldsymbol{Q}$ is any $p \times p$ nonsingular matrix. From Loh (1991a), we can see that an invariant estimator under the above group transformations has the form

$$(\widehat{\boldsymbol{\Sigma}}_1^{EQ},\, \widehat{\boldsymbol{\Sigma}}_2^{EQ}) = (\boldsymbol{B}^{-1}\boldsymbol{\Psi}(\boldsymbol{F})\boldsymbol{B}'^{-1},\, \boldsymbol{B}^{-1}\boldsymbol{\Phi}(\boldsymbol{F})\boldsymbol{B}'^{-1}). \tag{9}$$

Here we assume that $\boldsymbol{B}$ is a nonsingular matrix such that $\boldsymbol{B}(\boldsymbol{S}_1 + \boldsymbol{S}_2)\boldsymbol{B}' = \boldsymbol{I}_p$, $\boldsymbol{B}\boldsymbol{S}_2\boldsymbol{B}' = \boldsymbol{F}$, and $\boldsymbol{F} = \mathbf{diag}(f_1,\, f_2,\, \ldots,\, f_p)$ with $f_1 \geq f_2 \geq \cdots \geq f_p > 0$, and that $\boldsymbol{\Psi}(\boldsymbol{F}) = \mathbf{diag}(\psi_1(\boldsymbol{F}),\, \psi_2(\boldsymbol{F}),\, \ldots, \psi_p(\boldsymbol{F}))$ and $\boldsymbol{\Phi}(\boldsymbol{F}) = \mathbf{diag}(\phi_1(\boldsymbol{F}),\, \phi_2(\boldsymbol{F}),\, , \ldots, \phi_p(\boldsymbol{F}))$ are diagonal matrices whose elements are functions of $\boldsymbol{F}$. In the sequel of the paper, we abbreviate $\boldsymbol{\Psi}$, $\boldsymbol{\Phi}$, ψ_i, $\phi_i\,(i = 1,\, 2,\, \ldots,\, p)$ for $\boldsymbol{\Psi}(\boldsymbol{F})$, $\boldsymbol{\Phi}(\boldsymbol{F})$, $\psi_i(\boldsymbol{F})$, $\phi_i(\boldsymbol{F})$, respectively.

2.2. A sufficient conditions for improvement upon the James-Stein estimator

Now we state the main result in this section.

Theorem 1 *Suppose that we wish to estimate* $(\boldsymbol{\Sigma}_1,\, \boldsymbol{\Sigma}_2)$ *simultaneously under the loss function* (4). *An invariant estimator* $(\widehat{\boldsymbol{\Sigma}}_1^{EQ}, \widehat{\boldsymbol{\Sigma}}_2^{EQ})$ *is better than the James-Stein estimator* $(\widehat{\boldsymbol{\Sigma}}_1^{JS}, \widehat{\boldsymbol{\Sigma}}_2^{JS})$ *for arbitrary* g_1 *and* g_2 *in* (2) *if*

(i) $$p - (n_1 - p - 1)\sum_{j=1}^{p}\frac{\psi_j}{1 - f_j} - 2\sum_{j=1}^{p}\left[\psi_j + f_j\frac{\partial\psi_j}{\partial(1 - f_j)} + \psi_j\sum_{k \neq j}\frac{f_k}{f_k - f_j}\right] \geq 0,$$

(ii) $$p - (n_2 - p - 1)\sum_{j=1}^{p}\frac{\phi_j}{f_j} - 2\sum_{j=1}^{p}\left[\phi_j + (1 - f_j)\frac{\partial\phi_j}{\partial f_j} + \phi_j\sum_{k \neq j}\frac{1 - f_k}{f_j - f_k}\right] \geq 0,$$

(iii) $$\sum_{j=1}^{p}\left[-(\log d_{1j} + \log d_{2j}) + \log\frac{\psi_j}{1 - f_j} + \log\frac{\phi_j}{f_j}\right] \geq 0,$$

where $d_{1j} = 1/(n_1 + p + 1 - 2j)$ *and* $d_{2j} = 1/(n_2 + p + 1 - 2j)$, $j = 1,\, 2,\, \ldots,\, p$.

As a special case of $(\widehat{\boldsymbol{\Sigma}}_1^{EQ}, \widehat{\boldsymbol{\Sigma}}_2^{EQ})$, we introduce the estimator due to Loh (1991)

$$(\widehat{\boldsymbol{\Sigma}}_1^{LO},\, \widehat{\boldsymbol{\Sigma}}_2^{LO}) = (\boldsymbol{B}^{-1}\boldsymbol{\Psi}^{LO}\boldsymbol{B}'^{-1},\, \boldsymbol{B}^{-1}\boldsymbol{\Phi}^{LO}\boldsymbol{B}'^{-1}), \tag{10}$$

where $\boldsymbol{\Psi}^{LO} = \mathbf{diag}(\psi_1^{LO},\, \ldots,\, \psi_p^{LO})$ and $\boldsymbol{\Phi}^{LO} = \mathbf{diag}(\phi_1^{LO},\, \ldots,\, \phi_p^{LO})$ with the j-th diagonal elements $\psi_j^{LO} = (1 - f_j)/(n_1 - p - 1 + 2j)$ and $\phi_j^{LO} = f_j/(n_2 + p + 1 - 2j)$, $j = 1,\, 2,\, \ldots,\, p$, respectively.

Immediately we get the following corollary from Theorem 1.

Corollary 1 *The estimator* $(\widehat{\boldsymbol{\Sigma}}_1^{LO}, \widehat{\boldsymbol{\Sigma}}_2^{LO})$ *is better than* $(\widehat{\boldsymbol{\Sigma}}_1^{JS}, \widehat{\boldsymbol{\Sigma}}_2^{JS})$ *for arbitrary* g_1 *and* g_2 *in* (2).

Remark : When the error $(\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_2)$ have the joint density (3), we can see the dominance results similar to those in Theorem and Corollary above.

2.3. Numerical studies

Furthermore we introduce the Dey-Srinivasan estimator

$$(\widehat{\boldsymbol{\Sigma}}_1^{DS}, \widehat{\boldsymbol{\Sigma}}_2^{DS}) = (\boldsymbol{H}_1\boldsymbol{D}_1\boldsymbol{K}_1\boldsymbol{H}_1', \boldsymbol{H}_2\boldsymbol{D}_2\boldsymbol{K}_2\boldsymbol{H}_2'), \tag{11}$$

where $\boldsymbol{H}_i$, $i = 1, 2$, is a $p \times p$ orthogonal matrix such that $\boldsymbol{S}_i = \boldsymbol{H}_i\boldsymbol{K}_i\boldsymbol{H}_i'$ and $\boldsymbol{K}_i = \mathbf{diag}(k_{i1}, k_{i2}, \dots, k_{ip})$ with $k_{i1} \geq k_{i2} \geq \cdots \geq k_{ip} > 0$.

Note that the Dey-Srinivasan estimator (11) is invariant under the group of transformations given by

$$\boldsymbol{\Sigma}_i \to \boldsymbol{O}_i\boldsymbol{\Sigma}_i\boldsymbol{O}_i', \qquad \boldsymbol{S}_i \to \boldsymbol{O}_i\boldsymbol{S}_i\boldsymbol{O}_i', \qquad i = 1, 2,$$

where $\boldsymbol{O}_i$ is any $p \times p$ orthogonal matrix. Furthermore from Proposition 1 in Kubokawa and Srivastava (1999), we can see that the estimator (11) improves upon the estimator (7). However, it is difficult to compare $(\widehat{\boldsymbol{\Sigma}}_1^{LO}, \widehat{\boldsymbol{\Sigma}}_2^{LO})$ with $(\widehat{\boldsymbol{\Sigma}}_1^{DS}, \widehat{\boldsymbol{\Sigma}}_2^{DS})$ analytically. Therefore, to compare the risk performances of these estimators, we carry out Monte Carlo simulations.

Our simulations are based on 10,000 independent replications. We consider three-type of error distributions which are given in the following.

1. <u>The matrix-variate normal distribution</u>: The joint density function $(\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_2)$ is given by

$$\Pi_{i=1}^2 c_{i1}|\boldsymbol{\Sigma}_i|^{-N_i/2}\exp[-(1/2)\mathrm{tr}(\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\epsilon}_i'\boldsymbol{\epsilon}_i)],$$

   where $c_{i1} = (2\pi)^{-N_ip/2}$.

2. <u>The t-distribution</u>: The joint density function $(\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_2)$ is given by

$$\Pi_{i=1}^2 c_{i2}|\boldsymbol{\Sigma}_i|^{-N_i/2}\{1 + (1/v_i)\mathrm{tr}(\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\epsilon}_i'\boldsymbol{\epsilon}_i)\}^{-(v_i+N_ip)/2},$$

   where $c_{i2} = \Gamma[\{v_i + N_ip\}/2]/\{(\pi v_i)^{N_ip/2}\Gamma[v_i/2]\}$, $v_i > 0$. Here we denote by $\Gamma(\,\cdot\,)$ the Gamma function.

3. <u>The Kotz-type distribution</u>: The joint density function ($\boldsymbol{\epsilon}_1$, $\boldsymbol{\epsilon}_2$) is given by

$$\Pi_{i=1}^{2} c_{i3}|\boldsymbol{\Sigma}_i|^{-N_i/2}\{\mathrm{tr}(\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\epsilon}_i'\boldsymbol{\epsilon}_i)\}^{u_i-1}\exp[-r_i\{\mathrm{tr}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\epsilon}_i'\boldsymbol{\epsilon}_i\}^{s_i}],$$

where $r_i > 0$, $s_i > 0$, $2u_i + N_i p > 2$, and

$$c_{i3} = \frac{s_i\Gamma[N_i p/2] r_i^{\{u_i+N_i p/2-1\}/s_i}}{\pi^{N_i p/2}\Gamma[\{u_i + N_i p/2 - 1\}/s_i]}.$$

For generating a random number of the Kotz-type distribution above, see Fang, Kotz, and Ng (1990) for example.

For Monte Carlo simulations, we took $N_1 = N_2 = 15$, $m = 1$, and $p = 3$ and we also put $v_1 = v_2 = 3$ for the t-distribution and $(u_i, r_i, s_i) = (5, 0.1, 2)$, $i = 1, 2$, for the Kotz-type distribution. We also suppose that $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = (0, 0, 0)'$ and that the parameter $\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_1^{-1}$ is the diagonal matrix with typical elements. The estimated risks of these cases are given by Tables 1–3 and their estimated standard errors are in parentheses.

In Tables 1–3, 'US', 'JS', 'DS', and 'LO' stand for the usual estimator ($\widehat{\boldsymbol{\Sigma}}_1^{US}$, $\widehat{\boldsymbol{\Sigma}}_2^{US}$), the James-Stein estimator ($\widehat{\boldsymbol{\Sigma}}_1^{JS}$, $\widehat{\boldsymbol{\Sigma}}_2^{JS}$), the Dey-Srinivasan estimator ($\widehat{\boldsymbol{\Sigma}}_1^{DS}$, $\widehat{\boldsymbol{\Sigma}}_2^{DS}$), and the Loh estimator ($\widehat{\boldsymbol{\Sigma}}_1^{LO}$, $\widehat{\boldsymbol{\Sigma}}_2^{LO}$), respectively and 'AI' stands for average of improvement in risk over ($\widehat{\boldsymbol{\Sigma}}_1^{US}$, $\widehat{\boldsymbol{\Sigma}}_2^{US}$).

We also carried out simulations when the rows of $\boldsymbol{\epsilon}_i (i = 1, 2)$ have densities

$$|\boldsymbol{\Sigma}_i|^{-N_i/2} h(\boldsymbol{e}_{ij}'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{e}_{ij}), \qquad \text{for } j = 1, \ldots, N_i,$$

where $\boldsymbol{\epsilon}_i = (\boldsymbol{e}_{i1}', \boldsymbol{e}_{i2}', \ldots, \boldsymbol{e}_{iN_i}')'$. That is, the rows of each error matrix $\boldsymbol{\epsilon}_i$ are independently and identically distributed (i.i.d.) as an elliptically contoured distribution.

For Monte Carlo simulations, we suppose that the rows of $\boldsymbol{\epsilon}_i$ follow the vector-valued t-distributions, i.e., the density function of the random vectors $\boldsymbol{e}_{ij}$ ($i = 1, 2$, $j = 1, 2, \ldots, N_i$) are given by

$$c_{i4}\,|\boldsymbol{\Sigma}_i|^{-1/2}(1 + \boldsymbol{e}_{ij}'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{e}_{ij}/v_i)^{-(v_i+p)/2}, \tag{12}$$

where $v_i > 0$ and $c_{i4} = \Gamma[(v_i + p)/2]/\{(\pi v_i)^{p/2}\Gamma[v_i/2]\}$, and we also suppose that the rows of $\boldsymbol{\epsilon}_i$ follow the vector-valued Kotz-type distributions, i.e., the density functions of the random vectors $\boldsymbol{e}_{ij}$ ($i = 1, 2$, $j = 1, 2, \ldots, N_i$) are given by

$$c_{i5}|\boldsymbol{\Sigma}_i|^{-1/2}\{\boldsymbol{e}_{ij}'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{e}_{ij}\}^{l_i-1}\exp[-r_i\{\boldsymbol{e}_{ij}'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{e}_{ij}\}^{s_i}], \tag{13}$$

where $r_i > 0,\ s_i > 0,\ 2u_i + p > 2$, and

$$c_{i5} = \frac{s_i \Gamma[p/2] r_i^{\{u_i+p/2-1\}/s_i}}{\pi^{p/2}\Gamma[\{u_i + p/2 - 1\}/s_i]}.$$

For simulations, we take $N_1 = N_2 = 15,\ m = 1$, and $p = 3$ and we put $v_1 = v_2 = 3$ for the t-distributions and $(u_i, r_i, s_i) = (5, 0.1, 2),\ i = 1, 2$, for the Kotz-type distributions. The estimated risks of these cases are given by Tables 4 and 5.

The results of Monte Carlo simulations indicate that

1. When the eigenvalues of $\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_1^{-1}$ are close together, AI of LO is large;
2. DS is better than LO;
3. AIs of all alternative estimators are relatively small under non-normal error;
4. AIs of LO and DS are substantial under independently and identically sampling set-up from non-normal distribution, although we cannot prove improvement of alternative estimators over the usual estimator under this situation. Hence, these results suggest that the improvement under densities (2) remains robust even if the rows of errors are i.i.d.

3. Estimation of $\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_1^{-1}$

In this section we consider the problem (ii) given in Section 1 under elliptical error with density (3) and we treat the problem under the loss function

$$L_2(\hat{\boldsymbol{\zeta}}; \boldsymbol{\zeta}) = \mathrm{tr}\{\boldsymbol{\Sigma}_2^{-1}(\hat{\boldsymbol{\zeta}} - \boldsymbol{\zeta})\boldsymbol{S}_1(\hat{\boldsymbol{\zeta}} - \boldsymbol{\zeta})'\}/\mathrm{tr}\boldsymbol{\zeta}, \tag{14}$$

as considered by Loh (1988). Recall that $\boldsymbol{S}_i = \boldsymbol{Y}_i'(\boldsymbol{I}_{N_i} - \boldsymbol{C}_i(\boldsymbol{C}_i'\boldsymbol{C}_i)^{-1}\boldsymbol{C}_i')\boldsymbol{Y}_i$ for $i = 1, 2$. As pointed out in Loh (1991b), the problem (ii) is invariant under the group of transformations given by (8) and the estimators which is invariant under this transformation group has the form

$$\hat{\boldsymbol{\zeta}} = \boldsymbol{A}^{-1}\boldsymbol{\Xi}\boldsymbol{A},$$

where $\boldsymbol{A}$ is a nonsingular matrix such that $\boldsymbol{A}\boldsymbol{S}_1\boldsymbol{A}' = \boldsymbol{I}_p$ and $\boldsymbol{A}\boldsymbol{S}_2\boldsymbol{A}' = \boldsymbol{L}$ with $\boldsymbol{L} = \mathbf{diag}(l_1, \ldots, l_p)$ $(l_1 \ge l_2 \ge \cdots \ge l_p > 0)$ and further $\boldsymbol{\Xi} = \mathbf{diag}(\xi_1, \ldots, \xi_p)$ whose diagonal elements are functions of $\boldsymbol{L}$.

3.1. The best constant multiplier of $\boldsymbol{S}_2\boldsymbol{S}_1^{-1}$

Consider a class of estimators of the form $\hat{\boldsymbol{\zeta}}^{US} = \alpha\boldsymbol{S}_2\boldsymbol{S}_1^{-1}$, where α is a constant. Then this estimator can be rewritten as

$$\hat{\boldsymbol{\zeta}}^{US} = \boldsymbol{A}^{-1}\boldsymbol{\Xi}^{US}\boldsymbol{A},$$

where $\boldsymbol{\Xi}^{US}$ is diagonal matrix whose the j-th diagonal element is $\xi_j^{US} = \alpha l_j$, $j = 1, 2, \ldots, p$.

Theorem 2 *For any function g in* (3)*, the best usual estimator of $\boldsymbol{\zeta}$ under the loss function* (14) *is given by*

$$\hat{\boldsymbol{\zeta}}^{BU} = \boldsymbol{A}^{-1}\boldsymbol{\Xi}^{BU}\boldsymbol{A}, \tag{15}$$

where $\boldsymbol{\Xi} = \mathbf{diag}(\xi_1^{BU}, \ldots, \xi_p^{BU})$ with $\xi_j^{BU} = [(n_1 - p - 1)/(n_2 + p + 1)]l_j$.

3.2. Improved Estimators

We next discuss an improvement on the estimator (15). It is expected that the eigenvalues of $\boldsymbol{S}_2\boldsymbol{S}_1^{-1}$ are more spread out than those of $\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_1^{-1}$. To reduce the biases of the estimators for eigenvalues, we consider

$$\boldsymbol{\Xi}^{LO} = \mathbf{diag}(\xi_1^{LO}, \ldots, \xi_p^{LO}), \qquad \xi_j^{LO} = (n_1 - p - 1)l_j/(n_2 + p + 3 - 2j), \tag{16}$$

for $j = 1, 2, \ldots, p$. Then we have the following theorem:

Theorem 3 *Under the loss function* (14)*, $\hat{\boldsymbol{\zeta}}^{LO} = \boldsymbol{A}^{-1}\boldsymbol{\Xi}^{LO}\boldsymbol{A}$ is better than $\hat{\boldsymbol{\zeta}}^{BU}$ for any function g in* (3)*.*

Further we consider an improved estimator on $\hat{\zeta}^{LO}$ as in Loh (1988). Define the Berger-type estimator as

$$\hat{\boldsymbol{\zeta}}^{BE} = \boldsymbol{A}^{-1}\boldsymbol{\Xi}^{BE}\boldsymbol{A}, \tag{17}$$

where $\boldsymbol{\Xi}^{BE} = \mathbf{diag}(\xi_1^{BE}, \ldots, \xi_p^{BE})$ with

$$\xi_j^{BE} = \xi_j^{LO} + \frac{c}{b+u}, \quad u = \sum_{j=1}^{p}\left(\frac{n_2 + p + 3 - 2j}{(n_1 - p - 1)l_j}\right)^2.$$

Here $c : R^+ \mapsto R$ is a differentiable function of u and b is a suitable positive constant.

Then we have the following theorem:

Theorem 4 *Assume that*

(I) *$p \geq 3$, $n_1 \geq p$ and $n_2 \geq p$;*

(II) *$c(u) \geq 0$ and $c'(u) \geq 0$ for all $u \geq 0$;*

(III) $\sup_u c(u)/\sqrt{b} \leq 4(p^2+p-4)(n_2-p+3)/[\sqrt{p}(n_1-p-1)(n_2-p+7)]$.

Then, under the loss function (14), $\hat{\boldsymbol{\zeta}}^{BE}$ is better than $\hat{\boldsymbol{\zeta}}^{LO}$ for any function g in (3).

Since $l_1 \geq \cdots \geq l_p$, the diagonal elements of $\boldsymbol{\Xi}$ should have the ordering property, i.e., $\xi_1 \geq \cdots \geq \xi_p$. Hence, to improve on $\hat{\boldsymbol{\zeta}}^{LO}$, we can consider the estimator, for example, with Stein's isotonic regression on the ξ_j^{LO}'s (see Lin and Perlman, 1985).

Further we can also consider the Stein-type estimator (Stein, 1977)

$$\hat{\boldsymbol{\zeta}}^{ST} = \boldsymbol{A}^{-1}\boldsymbol{\Xi}^{ST}\boldsymbol{A}, \tag{18}$$

where $\boldsymbol{\Xi}^{ST} = \mathbf{diag}(\xi_1^{ST}, \ldots, \xi_p^{ST})$ with

$$\boldsymbol{\xi}_j^{ST} = (n_1-p-1)l_j \Bigg/ \left(n_2+p+1+2\sum_{k\neq j}\frac{l_k}{l_j-l_k}\right), \qquad j=1,\,2,\,\ldots,\,p.$$

The derivation of this estimator is given by Loh (1991b).

By applying methods of Berger and of Stein, the Stein-Berger estimator is given by

$$\hat{\boldsymbol{\zeta}}^{SB} = \boldsymbol{A}^{-1}\boldsymbol{\Xi}^{SB}\boldsymbol{A},$$

where $\boldsymbol{\Xi}^{SB} = \mathbf{diag}(\xi_1^{SB}, \ldots, \xi_p^{SB})$ with

$$\xi_j^{SB} = \bar{\xi}_j^{ST} + \frac{c}{b+u}\,(j=1,\,2,\,\ldots,\,p), \quad u = \sum_{j=1}^{p}\left(\frac{n_2+p+3-2j}{(n_1-p-1)l_j}\right)^2,$$

and the $\bar{\xi}_j^{ST}$'s are constructed by Stein's isotonic regression on the ξ_j^{ST}'s. However, we cannot analytically compare $\hat{\boldsymbol{\zeta}}^{BE}$ with $\hat{\boldsymbol{\zeta}}^{SB}$ and hence, in the next subsection, we examine these risk performances by using a numerical study.

3.3. Numerical studies

We have carried out Monte Carlo simulations (10,000 runs) to observe the risk performances of several estimators in the previous subsection.

1. The matrix-variate normal distribution: The joint density function of $(\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_2)$ is given by

$$\gamma_1|\boldsymbol{\Sigma}_1|^{-N_1/2}|\boldsymbol{\Sigma}_2|^{-N_2/2}\exp[-(1/2)\mathrm{tr}\{\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\epsilon}_1'\boldsymbol{\epsilon}_1 + \boldsymbol{\Sigma}_2^{-1}\boldsymbol{\epsilon}_2'\boldsymbol{\epsilon}_2\}],$$

where $\gamma_1 = (2\pi)^{-(N_1+N_2)p/2}$.

2. The t-distribution: The joint density function of $(\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_2)$ is given by

$$\gamma_2 |\boldsymbol{\Sigma}_1|^{-N_1/2} |\boldsymbol{\Sigma}_2|^{-N_2/2} [1 + (1/v) \mathrm{tr}\{\boldsymbol{\Sigma}_1^{-1} \boldsymbol{\epsilon}_1' \boldsymbol{\epsilon}_1 + \boldsymbol{\Sigma}_2^{-1} \boldsymbol{\epsilon}_2' \boldsymbol{\epsilon}_2\}]^{\{-(v+(N_1+N_2)p\}/2},$$

   where $\gamma_2 = \Gamma[\{v + (N_1 + N_2)p\}/2]/\{(\pi v)^{(N_1+N_2)p/2} \Gamma[v/2]\}$, $v > 0$.

3. The Kotz-type distribution: The joint density function of $(\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_2)$ is given by

$$\gamma_3 |\boldsymbol{\Sigma}_1|^{-N_1/2} |\boldsymbol{\Sigma}_2|^{-N_2/2} [\mathrm{tr}\{\boldsymbol{\Sigma}_1^{-1} \boldsymbol{\epsilon}_1' \boldsymbol{\epsilon}_1 + \boldsymbol{\Sigma}_2^{-1} \boldsymbol{\epsilon}_2' \boldsymbol{\epsilon}_2\}]^{u-1}$$
$$\times \exp[-r \mathrm{tr}\{\boldsymbol{\Sigma}_1^{-1} \boldsymbol{\epsilon}_1' \boldsymbol{\epsilon}_1 + \boldsymbol{\Sigma}_2^{-1} \boldsymbol{\epsilon}_2' \boldsymbol{\epsilon}_2\}^s],$$

   where $r > 0$, $s > 0$, $2u + (N_1 + N_2)p > 2$, and

$$\gamma_3 = \frac{s\Gamma[(N_1 + N_2)p/2] r^{\{u+(N_1+N_2)p/2-1\}/s}}{\pi^{(N_1+N_2)p/2} \Gamma[\{u + (N_1 + N_2)p/2 - 1\}/s]}.$$

The estimated risks with the error above are given in Tables 6–8. For Monte Carlo simulations, we took $N_1 = N_2 = 15$, $m = 1$, and $p = 3$ and we also put $v = 5$ for the t-distribution and $(u, r, s) = (5, 0.1, 2)$ for the Kotz-type distribution. We supposed that $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = (0, 0, 0)'$ and that the parameter $\boldsymbol{\Sigma}_2 \boldsymbol{\Sigma}_1^{-1}$ is the diagonal matrix with typical elements.

In tables, 'BU', 'LO', 'BE', and 'SB' stand for $\hat{\boldsymbol{\zeta}}^{BU}, \hat{\boldsymbol{\zeta}}^{LO}, \hat{\boldsymbol{\zeta}}^{BE}$ respectively and $\hat{\boldsymbol{\zeta}}^{SB}$, and 'AI' stands for the average of improving over risk of $\hat{\boldsymbol{\zeta}}^{BU}$. For $\hat{\boldsymbol{\zeta}}^{BE}$ and $\hat{\boldsymbol{\zeta}}^{SB}$, we set $b = 100$ and

$$c = \frac{2(p^2 + p - 4)(n_2 - p + 3)\sqrt{b}}{\sqrt{p}(n_1 - p - 1)(n_2 - p + 7)}.$$

We also studied simulations when the rows of the error are independently and identically distributed as in the densities (12) and (13). We assume that the distributions of rows of the errors are (12) and are (13) in the previous section. This results are given in Tables 9 and 10. We got the similar results as those in Tables 7 and 8.

We summarize our numerical results in Tables 6–10 as follows:

1. When the diagonal elements of $\boldsymbol{\Sigma}_2 \boldsymbol{\Sigma}_1^{-1}$ are close together, AIs of BE and SB are substantially large under normal error as well as under non-normal error;

2. Among the alternative estimators, AI of SB is the largest when the diagonal elements of $\boldsymbol{\Sigma}_2 \boldsymbol{\Sigma}_1^{-1}$ are equal or when the smallest diagonal element of $\boldsymbol{\Sigma}_2 \boldsymbol{\Sigma}_1^{-1}$ is far from the others and the others are large.

3. When the largest diagonal element of $\mathbf{\Sigma}_2\mathbf{\Sigma}_1^{-1}$ is far from the second, AIs are relatively small.

4. The results in Tables 9 and 10 suggest that the improvement by $\hat{\boldsymbol{\zeta}}^{LO}$ and $\hat{\boldsymbol{\zeta}}^{BE}$ under density (3) remains robust even if the rows of errors are i.i.d.

4. Proofs of Theorems

In this section we show theorems and corollaries in Sections 2 and 3. To give the proofs, we state a canonical form of our problems and list useful lemmas. Listed lemmas consist of two ingredients. First, we adapted the integration-by-parts formula from Kubokawa and Srivastava (1999) for our problems. We introduce two types of integration-by-parts formulas which concerns the joint density function (2) and the joint density function (3), respectively. Second, we quote lemmas on eigenstructure from Loh (1988).

4.1. Preliminaries

To derive a canonical form, write $(\mathbf{\Gamma}_i\boldsymbol{C}_i)' = ((\boldsymbol{C}_i'\boldsymbol{C}_i)^{1/2}, \mathbf{0})$, $i = 1, 2$, where $\mathbf{\Gamma}_i$ is an $N_i \times N_i$ orthogonal matrix. Also put $\boldsymbol{\theta}_i = (\boldsymbol{C}_i'\boldsymbol{C}_i)^{1/2}\boldsymbol{\beta}_i$ and $n_i = N_i - m$. Furthermore, write $(\boldsymbol{X}_i',\ \boldsymbol{Z}_i')' = \mathbf{\Gamma}_i\boldsymbol{Y}_i$, where $\boldsymbol{X}_i$ and $\boldsymbol{Z}_i$ are $m \times p$ and $n_i \times p$ matrices respectively. Then the densities (2) are rewritten as

$$|\mathbf{\Sigma}_i|^{-N_i/2} g_i[\mathrm{tr}\{\mathbf{\Sigma}_i^{-1}(\boldsymbol{X}_i - \boldsymbol{\theta}_i)'(\boldsymbol{X}_i - \boldsymbol{\theta}_i)\} + \mathrm{tr}(\mathbf{\Sigma}_i^{-1}\boldsymbol{Z}_i'\boldsymbol{Z}_i)] \tag{19}$$

for $i = 1, 2$.

Next we introduce notation for integration-by-parts formula with respect to the joint densities (2). Let

$$G_i(x) = \frac{1}{2}\int_x^{+\infty} g_i(t)dt$$

and let $\boldsymbol{\theta} = (\boldsymbol{\theta}_1,\ \boldsymbol{\theta}_2)$ and $\mathbf{\Sigma} = (\mathbf{\Sigma}_1, \mathbf{\Sigma}_2)$. For a function $U \equiv U(\boldsymbol{X}_1,\ \boldsymbol{X}_2,\ \boldsymbol{Z}_1,\ \boldsymbol{Z}_2)$, define

$$\boldsymbol{E}_{\theta,\,\Sigma}^{g_1 g_2}[U] = \int U \times \left\{\prod_{i=1}^{2} |\mathbf{\Sigma}_i|^{-N_i/2} g_i(w_i)\right\} d\boldsymbol{X}_1 d\boldsymbol{X}_2 d\boldsymbol{Z}_1 d\boldsymbol{Z}_2, \tag{20}$$

$$\boldsymbol{E}_{\theta,\,\Sigma}^{G_1 g_2}[U] = \int U \times |\mathbf{\Sigma}_1|^{-N_1/2}|\mathbf{\Sigma}_2|^{-N_2/2} G_1(w_1) g_2(w_2) d\boldsymbol{X}_1 d\boldsymbol{X}_2 d\boldsymbol{Z}_1 d\boldsymbol{Z}_2, \tag{21}$$

$$\boldsymbol{E}_{\theta,\,\Sigma}^{g_1 G_2}[U] = \int U \times |\mathbf{\Sigma}_1|^{-N_1/2}|\mathbf{\Sigma}_2|^{-N_2/2} g_1(w_1) G_2(w_2) d\boldsymbol{X}_1 d\boldsymbol{X}_2 d\boldsymbol{Z}_1 d\boldsymbol{Z}_2, \tag{22}$$

where $w_i = \mathrm{tr}\{\mathbf{\Sigma}_i^{-1}(\boldsymbol{X}_i - \boldsymbol{\theta}_i)'(\boldsymbol{X}_i - \boldsymbol{\theta}_i)\} + \mathrm{tr}(\mathbf{\Sigma}_i^{-1}\boldsymbol{Z}_i'\boldsymbol{Z}_i)$ $(i = 1,\ 2)$.

Put $\boldsymbol{S}_i = \boldsymbol{Z}_i'\boldsymbol{Z}_i$ and recall that $\boldsymbol{S}_i = \boldsymbol{Y}_i'\{\boldsymbol{I}_{N_i} - \boldsymbol{C}_i(\boldsymbol{C}_i'\boldsymbol{C})^{-1}\boldsymbol{C}_i'\}\boldsymbol{Y}_i$ for $i = 1, 2$ and let $\boldsymbol{H} \equiv \boldsymbol{H}(\boldsymbol{S}_1, \boldsymbol{S}_2) = (h_{ij})$ be a $p \times p$ matrix such that the (j, k)-element h_{jk} is a function of $\boldsymbol{S}_1 = (s_{1\cdot jk})$ and $\boldsymbol{S}_2 = (s_{2\cdot jk})$. For $i = 1, 2$, let

$$\{\boldsymbol{D}_{S_i}\boldsymbol{H}\}_{jk} = \sum_{a=1}^{p} d_{i\cdot ja} h_{ak}, \tag{23}$$

where

$$d_{i\cdot ja} = \frac{1}{2}(1 + \delta_{ja})\frac{\partial}{\partial s_{i\cdot ja}}$$

with $\delta_{ja} = 1$ for $j = a$ and $\delta_{ja} = 0$ for $j \neq a$. Also put $\boldsymbol{Z}_i = (\boldsymbol{z}_{i1}', \ldots, \boldsymbol{z}_{in_i}')'$ and $\boldsymbol{z}_{ij} = (z_{i\cdot j1}, \ldots, z_{i\cdot jp})$ for $i = 1, 2$, and $j = 1, 2, \ldots, n_i$. Hence we have $\boldsymbol{S}_i = \boldsymbol{Z}_i'\boldsymbol{Z}_i = \sum_{j=1}^{n_i} \boldsymbol{z}_{ij}'\boldsymbol{z}_{ij}$ for $i = 1, 2$.

Now we adapt the extended Stein-Haff identity due to Kubokawa and Srivastava (1999) for our problem. The difference between derivations of our identity and of that by Kubokawa and Srivastava (1999) is what expectation for the variables of integration are multiplied. Hence, we state the following formula without the proof.

Lemma 1 *Let*

$$\boldsymbol{H}_i \equiv \boldsymbol{H}_i\Bigg(\sum_{j_1=1}^{n_1} \boldsymbol{z}_{1j_1}'\boldsymbol{z}_{1j_1}, \sum_{j_2=1}^{n_2} \boldsymbol{z}_{2j_2}'\boldsymbol{z}_{2j_2}\Bigg), \qquad i = 1, 2,$$

be a $p \times p$ matrix whose element is differentiable with respect to $z_{i\cdot jk}$ ($j_i = 1, 2, \ldots, n_i$, $k = 1, 2, \ldots, p$). Furthermore, assume that

(a) $\boldsymbol{E}_{\theta,\Sigma}^{g_1 g_2}\Big[\big|\mathrm{tr}\boldsymbol{H}_i\boldsymbol{\Sigma}_i^{-1}\big|\Big]$ $(i = 1, 2)$ *is finite;*

(b) $\displaystyle\lim_{z_{i\cdot jk}\to\pm\infty} |z_{i\cdot jk}|\boldsymbol{H}_i\Bigg(\sum_{j_1=1}^{n_1} \boldsymbol{z}_{1j_1}'\boldsymbol{z}_{1j_1}, \sum_{j_2=1}^{n_2} \boldsymbol{z}_{2j_2}'\boldsymbol{z}_{2j_2}\Bigg)\Bigg(\sum_{j_i=1}^{n_i} \boldsymbol{z}_{1j_i}'\boldsymbol{z}_{1j_i}\Bigg)^{-1} G_i(z_{i\cdot jk} + a) = 0$ *for any real a.*

Then we have

$$\begin{aligned}\boldsymbol{E}_{\theta,\Sigma}^{g_1 g_2}[\mathrm{tr}(\boldsymbol{H}_1\boldsymbol{\Sigma}_1^{-1})] + \boldsymbol{E}_{\theta,\Sigma}^{g_1 g_2}[\mathrm{tr}(\boldsymbol{H}_2\boldsymbol{\Sigma}_2^{-1})] &= \boldsymbol{E}_{\theta,\Sigma}^{G_1 g_2}\big[(n_1 - p - 1)\mathrm{tr}(\boldsymbol{H}_1\boldsymbol{S}_1^{-1}) + 2\mathrm{tr}(\boldsymbol{D}_{S_1}\boldsymbol{H}_1)\big] \\ &\quad + \boldsymbol{E}_{\theta,\Sigma}^{g_1 G_2}\big[(n_2 - p - 1)\mathrm{tr}(\boldsymbol{H}_2\boldsymbol{S}_2^{-1}) + 2\mathrm{tr}(\boldsymbol{D}_{S_2}\boldsymbol{H}_2)\big]\end{aligned}$$

for $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ and $\boldsymbol{\Sigma} = (\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2)$.

To derive the integration-by-parts formula with respect to the density (3), we make an orthogonal transformation $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$ to rewrite the density (3) as

$$|\boldsymbol{\Sigma}_1|^{-N_1/2}|\boldsymbol{\Sigma}_2|^{-N_2/2}g\left\{\sum_{i=1}^{2}\left[\mathrm{tr}\{\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{X}_i-\boldsymbol{\theta}_i)'(\boldsymbol{X}_i-\boldsymbol{\theta}_i)\}+\mathrm{tr}(\boldsymbol{\Sigma}_i^{-1}\boldsymbol{Z}_i'\boldsymbol{Z}_i)\right]\right\}, \tag{24}$$

where $\boldsymbol{X}_i$, $\boldsymbol{Z}_i$ and $\boldsymbol{\theta}_i$ are defined in the same way to obtain (19). For a real-valued function U, denote

$$\boldsymbol{E}^g_{\theta,\,\Sigma}[U]=\int U\times\left(\prod_{i=1}^{2}|\boldsymbol{\Sigma}_i|^{-N_i/2}\right)g(w)d\boldsymbol{X}_1d\boldsymbol{X}_2d\boldsymbol{Z}_1d\boldsymbol{Z}_2,$$
$$\boldsymbol{E}^G_{\theta,\,\Sigma}[U]=\int U\times\left(\prod_{i=1}^{2}|\boldsymbol{\Sigma}_i|^{-N_i/2}\right)G(w)d\boldsymbol{X}_1d\boldsymbol{X}_2d\boldsymbol{Z}_1d\boldsymbol{Z}_2,$$

where $w=\sum_{i=1}^{2}[\mathrm{tr}\{\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{X}_i-\boldsymbol{\theta}_i)'(\boldsymbol{X}_i-\boldsymbol{\theta}_i)\}+\mathrm{tr}(\boldsymbol{\Sigma}_i^{-1}\boldsymbol{Z}_i'\boldsymbol{Z}_i)]$, $G(x)=(1/2)\int_x^{+\infty}g(t)dt$, $\boldsymbol{\theta}=(\boldsymbol{\theta}_1,\,\boldsymbol{\theta}_2)$, and $\boldsymbol{\Sigma}=(\boldsymbol{\Sigma}_1,\,\boldsymbol{\Sigma}_2)$. ¿From preliminaries as above, we get the integration-by-parts formula for the density (3):

Lemma 2 *Assume that*

$$\boldsymbol{H}\equiv\boldsymbol{H}(\boldsymbol{S}_1,\,\boldsymbol{S}_2) \tag{25}$$

is differentiable with respect to $z_{i\cdot jk}$ $(i=1,\,2,\;j=1,2,\;\ldots,\,n_i,\;k=1,\,2,\,\ldots,\,p)$ *and that*

(a) $\boldsymbol{E}^g_{\theta,\,\Sigma}\left[\left|\mathrm{tr}(\boldsymbol{H}\boldsymbol{\Sigma}_i^{-1})\right|\right]$ *is finite for* $i=1,2$*;*

(b) $\lim_{z_{i\cdot jk}\to\pm\infty}|z_{i\cdot jk}|\boldsymbol{H}\big(\boldsymbol{S}_1,\boldsymbol{S}_2\big)\big(\boldsymbol{S}_i\big)^{-1}G(z_{i\cdot jk}+a)=0$ *where* $\boldsymbol{S}_i=\sum_{j_i=1}^{n_i}\boldsymbol{z}_{ij_i}'\boldsymbol{z}_{ij_i}$*,* a *is any real number and* $i=1,2$*.*

Then, for $i=1,2$*,*

$$\boldsymbol{E}^g_{\theta,\,\Sigma}[\mathrm{tr}(\boldsymbol{H}\boldsymbol{\Sigma}_i^{-1})]=\boldsymbol{E}^G_{\theta,\,\Sigma}\big[(n_i-p-1)\mathrm{tr}(\boldsymbol{H}\boldsymbol{S}_i^{-1})+2\mathrm{tr}(\boldsymbol{D}_{S_i}\boldsymbol{H})\big]. \tag{26}$$

Furthermore, we need the following lemmas to show main theorems and their corollaries.

Lemma 3 [Loh, 1991a] *Let* $\boldsymbol{S}_1$ *and* $\boldsymbol{S}_2$ *be* $p\times p$ *symmetric and positive-definite matrices. Also let* $\boldsymbol{B}$ *be nonsingular matrix such that* $\boldsymbol{B}(\boldsymbol{S}_1+\boldsymbol{S}_2)\boldsymbol{B}'=\boldsymbol{I}_p$*,* $\boldsymbol{B}\boldsymbol{S}_2\boldsymbol{B}'=\boldsymbol{F}$ *where* $\boldsymbol{F}=\mathbf{diag}(f_1,\,f_2,\,\ldots,\,f_p)$ *with* $f_1\ge f_2\ge\cdots\ge f_p$*. Furthermore, let* $\boldsymbol{\Psi}=\mathbf{diag}(\psi_1,\,\psi_2,\,\cdots,\,\psi_p)$ *and* $\boldsymbol{\Phi}=\mathbf{diag}(\phi_1,\,\phi_2,\,\cdots,\,\phi_p)$*, where the* ψ_j *and the* ϕ_j $(j=1,\,2,\,\ldots,\,p)$ *are differentiable* *functions of* $\boldsymbol{F}$*. Then we have*

$$\mathrm{tr}(\boldsymbol{D}_{S_1}\boldsymbol{B}^{-1}\boldsymbol{\Psi}\boldsymbol{B}'^{-1}) = \sum_{j=1}^{p}\left[\psi_j + f_j\frac{\partial\psi_j}{\partial(1-f_j)} + \psi_j\sum_{k\neq j}\frac{f_k}{f_k - f_j}\right],$$
$$\mathrm{tr}(\boldsymbol{D}_{S_2}\boldsymbol{B}^{-1}\boldsymbol{\Phi}\boldsymbol{B}'^{-1}) = \sum_{j=1}^{p}\left[\phi_j + (1-f_j)\frac{\partial\phi_j}{\partial f_j} + \phi_j\sum_{k\neq j}\frac{1-f_k}{f_j - f_k}\right],$$

where $\boldsymbol{D}_{S_1}$ *and* $\boldsymbol{D}_{S_2}$ *are defined as* (23).

Lemma 4 [Loh, 1991b] *Let* $\boldsymbol{S}_1$ *and* $\boldsymbol{S}_2$ *be* $p\times p$ *symmetric and positive-definite matrices. Also let* $\boldsymbol{A}$ *be nonsingular matrix such that* $\boldsymbol{A}\boldsymbol{S}_1\boldsymbol{A}' = \boldsymbol{I}_p$, $\boldsymbol{A}\boldsymbol{S}_2\boldsymbol{A}' = \boldsymbol{L}$ *where* $\boldsymbol{L} = \mathbf{diag}(l_1,\, l_2\,\ldots,\, l_p)$ *with* $l_1 \geq l_2 \geq \cdots \geq l_p$*. Let* $\boldsymbol{\Xi} = \mathbf{diag}(\xi_1,\, \xi_2,\, \ldots,\, \xi_p)$*, where the* ξ_j $(j = 1,\, 2, \ldots,\, p)$ *are differentiable functions of* $\boldsymbol{L}$*. Then we have*

$$\mathrm{tr}(\boldsymbol{D}_{S_1}\boldsymbol{A}^{-1}\boldsymbol{\Xi}\boldsymbol{A}'^{-1}) = \sum_{j=1}^{p}\left[\xi_j - l_j\frac{\partial\xi_j}{\partial l_j} + \xi_j\sum_{k\neq j}\frac{l_k}{l_k - l_j}\right],$$
$$\mathrm{tr}(\boldsymbol{D}_{S_2}\boldsymbol{A}^{-1}\boldsymbol{\Xi}^2\boldsymbol{A}'^{-1}) = \sum_{j=1}^{p}\left[2\xi_j\frac{\partial\xi_j}{\partial l_j} + \xi_j^2\sum_{k\neq j}\frac{1}{l_j - l_k}\right],$$

where $\boldsymbol{D}_{S_1}$ *and* $\boldsymbol{D}_{S_2}$ *are defined as* (23).

4.2. Proofs of Theorem 1 and Corollary 1 in Section 2

Proof of THEOREM 1: ¿From Lemmas 1 and 3, the risk of the estimator $(\widehat{\boldsymbol{\Sigma}}_1,\, \widehat{\boldsymbol{\Sigma}}_2)$ can be expressed as

$$\begin{aligned} R_1(\widehat{\boldsymbol{\Sigma}}_1,\, \widehat{\boldsymbol{\Sigma}}_2) = \boldsymbol{E}_{\theta,\,\Sigma}^{G_1g_2}&\left[(n_1-p-1)\sum_{j=1}^{p}\frac{\psi_j}{1-f_j} + 2\sum_{j=1}^{p}\left\{\psi_j + f_j\frac{\partial\psi_j}{\partial(1-f_j)} + \psi_j\sum_{k\neq j}\frac{f_k}{f_k-f_j}\right\}\right] \\ +\boldsymbol{E}_{\theta,\,\Sigma}^{g_1G_2}&\left[(n_2-p-1)\sum_{j=1}^{p}\frac{\phi_j}{f_j} + 2\sum_{j=1}^{p}\left\{\phi_j + (1-f_j)\frac{\partial\phi_j}{\partial f_j} + \phi_j\sum_{k\neq j}\frac{1-f_k}{f_j-f_k}\right\}\right] \\ +\boldsymbol{E}_{\theta,\,\Sigma}^{g_1g_2}&\left[-\sum_{j=1}^{p}\left\{\log\frac{\psi_j}{1-f_j} + \log\frac{\phi_j}{f_j}\right\} + \sum_{i=1}^{2}\{-\log|\boldsymbol{S}_i| + \log|\boldsymbol{\Sigma}_i| - p\}\right]. \end{aligned} \tag{27}$$

Similarly the risk of the James-Stein estimator $(\widehat{\boldsymbol{\Sigma}}_1^{JS}, \widehat{\boldsymbol{\Sigma}}_2^{JS})$ is given by

$$\begin{aligned} R_1(\widehat{\boldsymbol{\Sigma}}_1^{JS},\, \widehat{\boldsymbol{\Sigma}}_2^{JS}) &= \boldsymbol{E}_{\theta,\,\Sigma}^{G_1g_2}[p] + \boldsymbol{E}_{\theta,\,\Sigma}^{g_1G_2}[p] \\ &+\boldsymbol{E}_{\theta,\,\Sigma}^{g_1g_2}\left[-\sum_{j=1}^{p}\left\{\log d_{1j} + \log d_{2j}\right\} + \sum_{i=1}^{2}\{-\log|\boldsymbol{S}_i| + \log|\boldsymbol{\Sigma}_i| - p\}\right], \end{aligned} \tag{28}$$

where $d_{ij} = 1/(n_i + p + 1 - 2j)$. Hence, comparing the integrands with respect to each expectation of (20), (21), and (22) in the rhs of the equations (27) and (28), we complete the proof. □

Proof of COROLLARY 1: ¿From Theorem 1, it suffices to show that

$$p - (n_1 - p + 1)\sum_{j=1}^{p} d_{1j}^* - 2\sum_{j=1}^{p}\left[(1-f_j)d_{1j}^*\sum_{k\neq j}\frac{f_k}{f_k - f_j}\right] \geq 0, \tag{29}$$

$$p - (n_2 - p + 1)\sum_{j=1}^{p} d_{2j}^* - 2\sum_{j=1}^{p}\left[f_j d_{2j}^*\sum_{k\neq j}\frac{1-f_k}{f_j - f_k}\right] \geq 0, \tag{30}$$

where $d_{1j}^* = 1/(n_1 - p - 1 + 2j)$ and $d_{2j}^* = 1/(n_2 + p + 1 - 2j)$. We here note that the last term of the lhs in (29) is evaluated as

$$\begin{aligned} -\sum_{j=1}^{p}\left\{(1-f_j)d_{1j}^*\sum_{k\neq j}\frac{f_k}{f_k - f_j}\right\} &= \sum_{j<k}\frac{f_k(1-f_j)d_{1j}^* - f_j(1-f_k)d_{1k}^*}{f_j - f_k} \\ &= \sum_{j<k}\frac{f_k(1-f_j)(d_{1j}^* - d_{1k}^*)}{f_j - f_k} - \sum_{j<k} d_{1k}^* \\ &\geq -\sum_{j<k} d_{1k}^*, \end{aligned}$$

where the last inequality in the above display is derived by the fact that $f_k < f_j < 1$ and $d_{1k}^* < d_{1j}^*$ for $j < k$. Since $\sum_{j<k} d_{1k}^* = \sum_{j=1}^{p}(j-1)d_{1j}^*$, we get

$$\begin{aligned} p - (n_1 - p + 1)\sum_{j=1}^{p} d_{1j}^* - 2\sum_{j=1}^{p}\left\{(1-f_j)d_{1j}^*\sum_{k\neq j}\frac{f_k}{f_j - f_k}\right\} \\ \geq p - \sum_{j=1}^{p}(n_1 - p - 1 + 2j)d_{1j}^* = 0. \end{aligned}$$

The proof of the inequality (30) can proceed similarly. Note that the last term of the lhs in (30) is evaluated as

$$-\sum_{j=1}^{p}\left[f_j d_{2j}^*\sum_{k\neq j}\frac{1-f_k}{f_j - f_k}\right] = \sum_{j>k}\frac{f_k(1-f_j)(d_{2k}^* - d_{2j}^*)}{f_j - f_k} - \sum_{j>k} d_{2k}^* \geq -\sum_{j=1}^{p}(p-j)d_{2j}^*.$$

Putting the above inequality into the lhs of (30), we get the desired result. □

4.3. Proofs of Theorems 2, 3, and 4 in Section 3

First, we give the following lemma:

Lemma 5 *For estimation of* $\boldsymbol{\zeta} = \boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_1^{-1}$ *in model* (3)*, we consider an estimator of the form*

$$\hat{\boldsymbol{\zeta}} = \boldsymbol{A}^{-1}\boldsymbol{\Xi}\boldsymbol{A}, \tag{31}$$

where $\boldsymbol{A}$ *is a nonsingular matrix such that* $\boldsymbol{A}\boldsymbol{S}_1\boldsymbol{A}' = \boldsymbol{I}_p$ *and* $\boldsymbol{A}\boldsymbol{S}_2\boldsymbol{A}' = \boldsymbol{L}$ *with* $\boldsymbol{L} = \mathbf{diag}(l_1, l_2, \ldots, l_p)$ *with* $l_1 \geq l_2 \geq \cdots \geq l_p > 0$*, and* $\boldsymbol{\Xi} = \mathbf{diag}(\xi_1, \xi_2, \ldots, \xi_p)$*. Then, under the loss function* (5)*, the risk of* $\hat{\boldsymbol{\zeta}}$ *is given by*

$$\begin{aligned} R_2(\hat{\boldsymbol{\zeta}};\, \boldsymbol{\zeta}) = \boldsymbol{E}^G_{\theta,\Sigma}\Big[n_1 + \sum_j \Big\{ &\frac{n_2 - p - 1}{l_j}\xi_j^2 + 2\xi_j^2 \sum_{k \neq j} \frac{1}{l_j - l_k} + 4\xi_j \frac{\partial \xi_j}{\partial l_j} \\ &-2(n_1 - p + 1)\xi_j + 4\xi_j \sum_{k \neq j} \frac{l_k}{l_j - l_k} + 4 l_j \frac{\partial \xi_j}{\partial l_j}\Big\} \Big/ \operatorname{tr}\boldsymbol{\zeta}\Big]. \end{aligned} \tag{32}$$

Proof. We can write the risk function as

$$R_2(\hat{\boldsymbol{\zeta}};\, \boldsymbol{\zeta}) = \boldsymbol{E}^g_{\theta,\Sigma}[\operatorname{tr}\{\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_1^{-1}\boldsymbol{S}_1 + \boldsymbol{\Sigma}_2^{-1}\boldsymbol{A}^{-1}\boldsymbol{\Phi}^2{\boldsymbol{A}'}^{-1} - 2\boldsymbol{\Sigma}_1^{-1}\boldsymbol{A}^{-1}\boldsymbol{\Phi}{\boldsymbol{A}'}^{-1}\}/\operatorname{tr}\boldsymbol{\zeta}].$$

Noting that $\boldsymbol{E}^g_{\theta,\Sigma}[\boldsymbol{S}_1] = \boldsymbol{E}^G_{\theta,\Sigma}[n_1\boldsymbol{\Sigma}_1]$ and using Lemmas 2 and 4, we get (32). □

Proof of THEOREM 2: Substituting αl_j for ξ_j in (32), we get

$$R_2(\hat{\boldsymbol{\zeta}}^{US};\, \boldsymbol{\zeta}) = \boldsymbol{E}^G_{\theta,\Sigma}[n_1 + \{(n_2 + p + 1)\alpha^2 - 2(n_1 - p - 1)\alpha\} \sum_{j=1}^p l_j / \operatorname{tr}\boldsymbol{\zeta}]. \tag{33}$$

Hence we can see that (33) is minimized at $\alpha = (n_1 - p - 1)/(n_2 + p + 1)$. □

Proof of THEOREMS 3: The proof proceed in similar way as in that of Theorem 3.5 in Loh (1988). We reproduce it for reader's convenience. Write $\tilde{d}_j = (n_1 - p - 1)/(n_2 + p + 3 - 2j)$, $j = 1, 2, \ldots, p$. ¿From (33), we have

$$(\operatorname{tr}\boldsymbol{\zeta})\{R_2(\hat{\boldsymbol{\zeta}}^{BU};\, \boldsymbol{\zeta}) - \boldsymbol{E}^G_{\theta,\Sigma}[n_1]\} = \boldsymbol{E}^G_{\theta,\Sigma}\left[-\frac{(n_1 - p - 1)^2}{n_2 + p + 1}\sum_{i=1}^p l_i\right].$$

¿From Lemma 5, we have

$$\begin{aligned} (\operatorname{tr}\boldsymbol{\zeta})\{R_2(\hat{\boldsymbol{\zeta}}^{LO};\, \boldsymbol{\zeta}) - \boldsymbol{E}^G_{\theta,\Sigma}[n_1]\} = \boldsymbol{E}^G_{\theta,\Sigma}\Bigg[\sum_{j=1}^p \Bigg\{ &(n_2 - p + 3)\tilde{d}_j^2 l_j + 2\sum_{k<j} \frac{\tilde{d}_k^2(l_j^2 - l_k^2)}{l_j - l_k} \\ &+2\sum_{k<j} \frac{l_j^2(\tilde{d}_j^2 - \tilde{d}_k^2)}{l_j - l_k} - 2(n_1 - p - 1)\tilde{d}_j l_j + 4\sum_{k<j} \frac{l_j l_k(\tilde{d}_j - \tilde{d}_k)}{l_j - l_k}\Bigg\}\Bigg]. \end{aligned} \tag{34}$$

Noting that

$$2\sum_{k<j}\frac{l_j^2(\tilde{d}_j^2-\tilde{d}_k^2)}{l_j-l_k}+4\sum_{k<j}\frac{l_jl_k(\tilde{d}_j-\tilde{d}_k)}{l_j-l_k}\le 0$$

and that

$$\sum_{j=1}^{p}\sum_{k<j}\frac{\tilde{d}_k^2(l_j^2-l_k^2)}{l_j-l_k}=\sum_{j=1}^{p}\Big\{(p-j)\tilde{d}_j^2+\sum_{k<j}\tilde{d}_k^2\Big\}l_j,$$

we can see that the rhs of (34) is less than

$$\begin{aligned}&\boldsymbol{E}^G_{\theta,\Sigma}\left[\sum_{j=1}^{p}\Big\{(n_2+p+3-2j)\tilde{d}_j^2-2(n_1-p-1)\tilde{d}_j+2\sum_{k<j}\tilde{d}_k^2\Big\}l_j\right]\\&\quad=\boldsymbol{E}^G_{\theta,\Sigma}\left[\sum_{j=1}^{p}\Big\{-\frac{(n_1-p-1)^2}{n_2+p+1}-2(n_1-p-1)^2\Big(\frac{j-1}{(n_2+p+1)(n_2+p+3-2j)}\right.\\&\qquad\qquad\left.-\sum_{k<j}\frac{1}{(n_2+p+3-2k)^2}\Big)\Big\}l_j\right].\end{aligned}$$

Furthermore, from mathematical induction on j, we can see that

$$\frac{j-1}{(n_2+p+1)(n_2+p+3-2j)}\ge\sum_{k<j}\frac{1}{(n_2+p+3-2k)^2}.$$

¿From this inequality, we finally get

$$(\mathrm{tr}\boldsymbol{\zeta})\{R_2(\hat{\boldsymbol{\zeta}}^{LO};\boldsymbol{\zeta})-\boldsymbol{E}^G_{\theta,\Sigma}[n_1]\}\le(\mathrm{tr}\boldsymbol{\zeta})\{R_2(\hat{\boldsymbol{\zeta}}^{BU};\boldsymbol{\zeta})-\boldsymbol{E}^G_{\theta,\Sigma}[n_1]\},$$

which completes the proof. □

Proof of THEOREM 4: The proof proceed in the same way as in that of Theorem 3.5 in Loh (1988). However, we reproduce it for reader's convenience when c is a positive constant which satisfies Assumption (III). Put $\alpha_j=c/\{\tilde{d}_jl_j(b+u)\}$, $j=1,2,\ldots,p$. Hence $\xi_j^{BE}=\tilde{d}_jl_j(1+\alpha_j)$. Also note that

$$\frac{\partial\xi_j^{BE}}{\partial l_j}=\tilde{d}_j(1+\alpha_j)+\tilde{d}_jl_j\frac{\partial\alpha_j}{\partial l_j}=\tilde{d}_j(1+\alpha_j)+\frac{c}{l_j(b+u)}\left[\frac{2}{\tilde{d}_j^2l_j^2(b+u)}-1\right].$$

¿From tedious calculation, we have

$$\begin{aligned}&(\mathrm{tr}\boldsymbol{\zeta})\{R_2(\hat{\boldsymbol{\zeta}}^{BE};\boldsymbol{\zeta})-R_2(\hat{\boldsymbol{\zeta}}^{LO};\boldsymbol{\zeta})\}\\&\quad=\boldsymbol{E}^G_{\theta,\Sigma}\sum_{j=1}^{p}\Big[(n_2-p-1)\tilde{d}_j^2l_j(2\alpha_j+\alpha_j^2)+4\sum_{k<j}\frac{\tilde{d}_j^2l_j^2\alpha_j-\tilde{d}_k^2l_k^2\alpha_k}{l_j-l_k}+4\tilde{d}_j^2l_j(2\alpha_j+\alpha_j^2)\end{aligned}$$

$$+4\tilde{d}_j^2 l_j^2(1+\alpha_j)\frac{\partial \alpha_j}{\partial l_j} - 2(n_1-p+1)\tilde{d}_j l_j \alpha_j + 4\sum_{k<j}\frac{\tilde{d}_j l_j \alpha_j l_k - \tilde{d}_k l_k \alpha_k l_j}{l_j - l_k}$$

$$+4 l_j \tilde{d}_j \alpha_j + 4\tilde{d}_j l_j^2 \frac{\partial \alpha_j}{\partial l_j}\Bigg]$$

$$= \boldsymbol{E}^G_{\theta,\Sigma}\sum_{j=1}^p\Bigg[(n_2-p+3)\tilde{d}_j^2 l_j \alpha_j^2 + 4(\tilde{d}_j^2+\tilde{d}_j)l_j^2\frac{\partial \alpha_j}{\partial l_j} + 4\tilde{d}_j^2 l_j^2 \alpha_j \frac{\partial \alpha_j}{\partial l_j} - 4 l_j \alpha_j \tilde{d}_j^2 (p-j)$$

$$+4\sum_{k<j}\frac{\tilde{d}_j^2 l_j^2 \alpha_j - \tilde{d}_k^2 l_k^2 \alpha_k}{l_j - l_k} + 4\sum_{k<j}\frac{\tilde{d}_j l_j \alpha_j l_k - \tilde{d}_k l_k \alpha_k l_j}{l_j - l_k}\Bigg]. \qquad (35)$$

Now we observe that

$$\sum_{j=1}^p\Bigg(\sum_{k<j}\bigg\{\frac{\tilde{d}_j^2 l_j^2 \alpha_j - \tilde{d}_k^2 l_k^2 \alpha_k}{l_j - l_k} + \frac{\tilde{d}_j l_j \alpha_j l_k - \tilde{d}_k l_k \alpha_k l_j}{l_j - l_k}\bigg\} - l_j \alpha_j \tilde{d}_j^2 (p-j)\Bigg) \le \frac{c(p-p^2)}{2(b+u)}. \quad (36)$$

Furthermore, noting that $\tilde{d}_1 < \tilde{d}_2 < \cdots < \tilde{d}_p$ and that $\sum_{j=1}^p 1/\{\tilde{d}_j^2 l_j^2 (b+u)\} = u/(b+u) \le 1$, we have

$$\sum_{j=1}^p (\tilde{d}_j^2+\tilde{d}_j) l_j^2 \frac{\partial \alpha_j}{\partial l_j} = \frac{c}{b+u}\sum_{j=1}^p\Bigg[-1 + \frac{2}{\tilde{d}_j^2 l_j^2 (b+u)} - \tilde{d}_j + \frac{2\tilde{d}_j}{\tilde{d}_j^2 l_j^2 (b+u)}\Bigg]$$

$$\le -\frac{c}{b+u}\big[p-2-\tilde{d}_p + \sum_{j=1}^{p-1}\tilde{d}_j\big] \le -\frac{c}{b+u}(p-2). \qquad (37)$$

The last inequality follows from $\sum_{j=1}^{p-1}\tilde{d}_j - \tilde{d}_p \ge 0$. Since $l_1 > l_2 > \cdots > l_p$, we have

$$4\sum_{j=1}^p \tilde{d}_j^2 l_j^2 \alpha_j \frac{\partial \alpha_j}{\partial l_j} = \frac{4c^2}{(b+u)^2}\sum_{j=1}^p\Bigg[\frac{2}{\tilde{d}_j^2 l_j^3 (b+u)} - \frac{1}{l_j}\Bigg] \le \frac{4c^2}{(b+u)^2}\Bigg[\frac{2}{l_p} - \sum_{j=1}^p \frac{1}{l_j}\Bigg],$$

which gives

$$\sum_{j=1}^p\big\{(n_2-p+3)\tilde{d}_j^2 l_j \alpha_j^2 + 4\tilde{d}_j^2 l_j^2 \alpha_j(\partial\alpha_j/\partial l_j)\big\} \le \frac{c^2}{(b+u)^2}\Bigg[\frac{8}{l_p} + (n_2-p-1)\sum_{j=1}^p\frac{1}{l_j}\Bigg]$$

$$\le (n_2-p+7)\frac{c^2\tilde{d}_p}{(b+u)^2}\sum_{j=1}^p\frac{1}{\tilde{d}_j l_j} \le \frac{(n_1-p-1)(n_2-p+7)c^2\sqrt{p}}{2(n_2-p+3)(b+u)\sqrt{b}}. \qquad (38)$$

The last inequality follows from the inequality

$$\max_{y_j>0}\frac{\sum_{j=1}^p y_j}{b+\sum_{j=1}^p y_j^2} \le \frac{\sqrt{p}}{2\sqrt{b}}.$$

Finally putting (36)–(38) into the rhs of (35), we have

$$\mathrm{tr}\boldsymbol{\zeta}\big\{R_2(\hat{\boldsymbol{\zeta}}^{BE};\boldsymbol{\zeta}) - R_2(\hat{\boldsymbol{\zeta}}^{LO};\boldsymbol{\zeta})\big\}$$

$$\le \boldsymbol{E}^G_{\theta,\Sigma}\Bigg[\frac{c}{b+u}\bigg\{\frac{(n_1-p-1)(n_2-p+7)c\sqrt{p}}{2(n_2-p+3)\sqrt{b}} - 2(p^2+p-4)\bigg\}\Bigg] \le 0.$$

The last inequality follows from Assumption (III) of Theorem 4. □

REFERENCES

Berger, J. (1980). Improving on inadmissible estimators in continuous exponential families with applications to simultaneous estimation of gamma scale parameters. *Ann. Statist.* **8** 545–571.

DasGupta, A. (1989). A general theorem on decision theory for nonnegative functionals: With applications. *Ann. Statist.* **17** 1360–1374.

Dey, D.K. and Srinivasan, C. (1985). Estimation of covariance matrix under Stein's loss. *Ann. Statist.* **13** 1581–1591.

Fang, K.T., Kotz, S., and Ng, K.W. (1990). *Symmetric multivariate and related distributions*. Chapman and Hall, New York.

Fang, K.T. and Zhang, Y.T. (1990). *Generalized Multivariate Analysis*. Springer-Verlag, Berlin and Science Press, Beijing.

Haff, L.R. (1980). Empirical Bayes estimation of the multivariate normal covariance matrix. *Ann. Statist.* **8**, 586–597.

Haff, L.R. (1982). Identities for the inverse Wishart distribution with computational results in linear and quadratic discrimination. *Sankhyā, ser.* **B 44**, 245–258.

Haff, L.R. (1991). The variational form of certain Bayes estimators. *Ann. Statist.* **19**, 1163–1190.

James, W. and Stein, C. (1961). Estimation with quadratic loss. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* **1**, 361–380. Univ. California Press, Berkeley.

Johnson, M. (1987). *Multivariate Statistical Simulation*. John Wiley and Sons, New York.

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.

Lin, S.P. and Perlman, M.D. (1985). A Monte Carlo comparison of four estimators for a covariance matrix. In *Multivariate Analysis VI* (P. R. Krishnaiah, ed.) 411–429. North-Holland, Amsterdam.

Loh, W.L. (1988). Estimating covariance matrices. Ph. D. dissertation, Dept. Statist., Stanford Univ.

Loh, W.L. (1991a). Estimating covariance matrices. *Ann. Statist.* **19** 283–296.

Loh, W.L. (1991b). Estimating covariance matrices II. *J. Multivariate Anal.* **36** 163–174.

Muirhead, R.J. (1987). Developments in eigenvalue estimation. In *Advances in Multivariate Statistical Analysis* (A.K. Gupta, Ed.), 277–288, Reidel, Boston.

Muirhead, R.J. and Verathaworn, T. (1985). On estimating the latent roots of $\Sigma_1\Sigma_2^{-1}$. In *Multivariate Anal.* (P. R. Krishnaiah, Ed.), Vol. 6, 431–447. North-Holland, Amsterdam.

Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proc. Third Berkeley Symp. Math. Statist. Prob.* **1**, 197–206. Univ. California Press, Berkeley.

Stein, C. (1977). Lectures on the theory of estimation of many parameters. In *Studies in the Statistical Theory of Estimation I* (I. A. Ibragimov and M. S. Nikulin, eds.).

Table 1. Estimated risks for estimation of $(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2)$ under normal distributions

(Estimated standard errors are in parentheses)

$\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_1^{-1}$	*UB*	*ST*	AI	*DS*	AI	*LO*	AI
$\mathbf{diag}(1, 1, 1)$	0.9356 (0.0038)	0.8938 (0.0037)	4.47%	0.6649 (0.0033)	28.9%	0.7517 (0.0035)	19.7%
$\mathbf{diag}(10, 1, 1)$	0.9356 (0.0038)	0.8938 (0.0037)	4.47%	0.7068 (0.0034)	24.5%	0.8324 (0.0035)	11.0%
$\mathbf{diag}(100, 1, 1)$	0.9356 (0.0038)	0.8938 (0.0037)	4.47%	0.7100 (0.0034)	24.1%	0.8468 (0.0036)	9.48%
$\mathbf{diag}(1000, 1, 1)$	0.9356 (0.0038)	0.8938 (0.0037)	4.47%	0.7099 (0.0034)	24.1%	0.8480 (0.0036)	9.36%
$\mathbf{diag}(10, 5, 1)$	0.9356 (0.0038)	0.8938 (0.0037)	4.47%	0.7397 (0.0036)	20.9%	0.8331 (0.0035)	11.0%
$\mathbf{diag}(10, 10, 1)$	0.9356 (0.0038)	0.8938 (0.0037)	4.47%	0.7748 (0.0042)	17.2%	0.8325 (0.0035)	11.0%
$\mathbf{diag}(100, 10, 1)$	0.9356 (0.0038)	0.8938 (0.0037)	4.47%	0.7523 (0.0035)	19.6%	0.8812 (0.0036)	5.81%
$\mathbf{diag}(100, 100, 1)$	0.9356 (0.0038)	0.8938 (0.0037)	4.47%	0.8046 (0.0076)	14.0%	0.8470 (0.0036)	9.47%
$\mathbf{diag}(1000, 10, 1)$	0.9356 (0.0038)	0.8938 (0.0037)	4.47%	0.7510 (0.0035)	19.7%	0.8879 (0.0036)	5.10%
$\mathbf{diag}(1000, 100, 1)$	0.9356 (0.0038)	0.8938 (0.0037)	4.47%	0.7694 (0.0035)	17.8%	0.8878 (0.0036)	5.11%
$\mathbf{diag}(1000, 1000, 1)$	0.9356 (0.0038)	0.8938 (0.0037)	4.47%	0.7941 (0.0129)	15.1%	0.8482 (0.0036)	9.34%

Table 2. Estimated risks for estimation of $(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2)$ under t-distributions

(Estimated standard errors are in parentheses)

$\mathbf{\Sigma}_2\mathbf{\Sigma}_1^{-1}$	UB	ST	AI	DS	AI	LO	AI
$\mathbf{diag}(1,1,1)$	10.116 (0.4976)	10.078 (0.5042)	0.38%	9.4165 (0.4812)	6.92%	9.6676 (0.4888)	4.44%
$\mathbf{diag}(10,1,1)$	10.116 (0.4976)	10.078 (0.5042)	0.38%	9.5315 (0.4838)	5.78%	9.8952 (0.4937)	2.19%
$\mathbf{diag}(100,1,1)$	10.116 (0.4976)	10.078 (0.5042)	0.38%	9.5425 (0.4843)	5.67%	9.9337 (0.4938)	1.81%
$\mathbf{diag}(1000,1,1)$	10.116 (0.4976)	10.078 (0.5042)	0.38%	9.5423 (0.4843)	5.67%	9.9370 (0.4937)	1.77%
$\mathbf{diag}(10,5,1)$	10.116 (0.4976)	10.078 (0.5042)	0.38%	9.6207 (0.4876)	4.90%	9.9112 (0.4973)	2.03%
$\mathbf{diag}(10,10,1)$	10.116 (0.4976)	10.078 (0.5042)	0.38%	9.7393 (0.4886)	3.73%	9.9116 (0.4961)	2.02%
$\mathbf{diag}(100,10,1)$	10.116 (0.4976)	10.078 (0.5042)	0.38%	9.6664 (0.4887)	4.45%	10.040 (0.4980)	0.76%
$\mathbf{diag}(100,100,1)$	10.116 (0.4976)	10.078 (0.5042)	0.38%	9.8365 (0.4921)	2.77%	9.9501 (0.4961)	1.64%
$\mathbf{diag}(1000,10,1)$	10.116 (0.4976)	10.078 (0.5042)	0.38%	9.6613 (0.4887)	4.50%	10.058 (0.4980)	0.58%
$\mathbf{diag}(1000,100,1)$	10.116 (0.4976)	10.078 (0.5042)	0.38%	9.7189 (0.4895)	3.93%	10.058 (0.4980)	0.58%
$\mathbf{diag}(1000,1000,1)$	10.116 (0.4976)	10.078 (0.5042)	0.38%	9.8939 (0.5104)	2.20%	9.9531 (0.4960)	1.61%

Table 3. Estimated risks for estimation of $(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2)$ under Kotz-type distributions

(Estimated standard errors are in parentheses)

$\mathbf{\Sigma}_2\mathbf{\Sigma}_1^{-1}$	*UB*	*ST*	AI	*DS*	AI	*LO*	AI
diag$(1,1,1)$	4.6140 (0.0060)	4.5725 (0.0059)	0.90%	4.5145 (0.0059)	2.16%	4.5363 (0.0059)	1.68%
diag$(10,1,1)$	4.6140 (0.0060)	4.5725 (0.0059)	0.90%	4.5250 (0.0059)	1.93%	4.5572 (0.0059)	1.23%
diag$(100,1,1)$	4.6140 (0.0060)	4.5725 (0.0059)	0.90%	4.5257 (0.0059)	1.91%	4.5609 (0.0059)	1.15%
diag$(1000,1,1)$	4.6140 (0.0060)	4.5725 (0.0059)	0.90%	4.5257 (0.0059)	1.91%	4.5612 (0.0059)	1.14%
diag$(10,5,1)$	4.6140 (0.0060)	4.5725 (0.0059)	0.90%	4.5330 (0.0059)	1.76%	4.5572 (0.0059)	1.23%
diag$(10,10,1)$	4.6140 (0.0060)	4.5725 (0.0059)	0.90%	4.5435 (0.0059)	1.53%	4.5570 (0.0059)	1.24%
diag$(100,10,1)$	4.6140 (0.0060)	4.5725 (0.0059)	0.90%	4.5366 (0.0059)	1.68%	4.5696 (0.0059)	0.96%
diag$(100,100,1)$	4.6140 (0.0060)	4.5725 (0.0059)	0.90%	4.5516 (0.0062)	1.35%	4.5607 (0.0059)	1.16%
diag$(1000,10,1)$	4.6140 (0.0060)	4.5725 (0.0059)	0.90%	4.5365 (0.0059)	1.68%	4.5713 (0.0059)	0.93%
diag$(1000,100,1)$	4.6140 (0.0060)	4.5725 (0.0059)	0.90%	4.5408 (0.0059)	1.59%	4.5713 (0.0059)	0.93%
diag$(1000,1000,1)$	4.6140 (0.0060)	4.5725 (0.0059)	0.90%	4.5503 (0.0068)	1.38%	4.5610 (0.0059)	1.15%

Table 4. Estimated risks for estimation of $(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2)$ under t-distributions (i.i.d.)

(Estimated standard errors are in parentheses)

$\mathbf{\Sigma}_2\mathbf{\Sigma}_1^{-1}$	UB	ST	AI	DS	AI	LO	AI
$\mathbf{diag}(1,1,1)$	8.3045 (0.2110)	7.7993 (0.1879)	6.08%	7.0405 (0.1845)	15.2%	7.2989 (0.1849)	12.1%
$\mathbf{diag}(10,1,1)$	8.3045 (0.2110)	7.7993 (0.1879)	6.08%	7.1560 (0.1847)	13.8%	7.5063 (0.1853)	9.61%
$\mathbf{diag}(100,1,1)$	8.3045 (0.2110)	7.7993 (0.1879)	6.08%	7.2055 (0.1853)	13.2%	7.6093 (0.1861)	8.37%
$\mathbf{diag}(1000,1,1)$	8.3045 (0.2110)	7.7993 (0.1879)	6.08%	7.2099 (0.1855)	13.2%	7.6232 (0.1865)	8.20%
$\mathbf{diag}(10,5,1)$	8.3045 (0.2110)	7.7993 (0.1879)	6.08%	7.1335 (0.1841)	14.1%	7.4849 (0.1853)	9,87%
$\mathbf{diag}(10,10,1)$	8.3045 (0.2110)	7.7993 (0.1879)	6.08%	7.2077 (0.1838)	13.2%	7.5056 (0.1853)	9.62%
$\mathbf{diag}(100,10,1)$	8.3045 (0.2110)	7.7993 (0.1879)	6.08%	7.2281 (0.1842)	13.0%	7.6732 (0.1859)	7.60%
$\mathbf{diag}(100,100,1)$	8.3045 (0.2110)	7.7993 (0.1879)	6.08%	7.3523 (0.1852)	11.5%	7.6110 (0.1862)	8.35%
$\mathbf{diag}(1000,10,1)$	8.3045 (0.2110)	7.7993 (0.1879)	6.08%	7.2351 (0.1846)	12.9%	7.7301 (0.1865)	6.92%
$\mathbf{diag}(1000,100,1)$	8.3045 (0.2110)	7.7993 (0.1879)	6.08%	7.3381 (0.1852)	11.6%	7.7286 (0.1865)	6.93%
$\mathbf{diag}(1000,1000,1)$	8.3045 (0.2110)	7.7993 (0.1879)	6.08%	7.3524 (0.1867)	11.5%	7.6243 (0.1864)	8.19%

Table 5. Estimated risks for estimation of $(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2)$ under Kotz-type distributions (i.i.d.)

(Estimated standard errors are in parentheses)

$\mathbf{\Sigma}_2\mathbf{\Sigma}_1^{-1}$	*UB*	*ST*	AI	*DS*	AI	*LO*	AI
$\mathbf{diag}(1,1,1)$	1.5235 (0.0037)	1.5228 (0.0037)	0.04%	1.1793 (0.0030)	22.6%	1.3040 (0.0033)	14.4%
$\mathbf{diag}(10,1,1)$	1.5235 (0.0037)	1.5228 (0.0037)	0.04%	1.2461 (0.0032)	18.2%	1.4343 (0.0035)	5.85%
$\mathbf{diag}(100,1,1)$	1.5235 (0.0037)	1.5228 (0.0037)	0.04%	1.2489 (0.0032)	18.0%	1.4503 (0.0036)	4.80%
$\mathbf{diag}(1000,1,1)$	1.5235 (0.0037)	1.5228 (0.0037)	0.04%	1.2487 (0.0032)	18.0%	1.4517 (0.0036)	4.71%
$\mathbf{diag}(10,5,1)$	1.5235 (0.0037)	1.5228 (0.0037)	0.04%	1.3037 (0.0038)	14.4%	1.4425 (0.0034)	5.32%
$\mathbf{diag}(10,10,1)$	1.5235 (0.0037)	1.5228 (0.0037)	0.04%	1.3743 (0.0057)	9.79%	1.4337 (0.0035)	5.89%
$\mathbf{diag}(100,10,1)$	1.5235 (0.0037)	1.5228 (0.0037)	0.04%	1.3174 (0.0034)	13.5%	1.5069 (0.0037)	1.09%
$\mathbf{diag}(100,100,1)$	1.5235 (0.0037)	1.5228 (0.0037)	0.04%	1.4237 (0.0136)	6.55%	1.4498 (0.0036)	4.84%
$\mathbf{diag}(1000,10,1)$	1.5235 (0.0037)	1.5228 (0.0037)	0.04%	1.3166 (0.0034)	13.6%	1.5141 (0.0037)	0.62%
$\mathbf{diag}(1000,100,1)$	1.5235 (0.0037)	1.5228 (0.0037)	0.04%	1.3374 (0.0034)	12.2%	1.5141 (0.0037)	0.62%
$\mathbf{diag}(1000,1000,1)$	1.5235 (0.0037)	1.5228 (0.0037)	0.04%	1.4000 (0.0233)	8.10%	1.4512 (0.0036)	4.74%

Table 6. Estimated risks for estimation of $\mathbf{\Sigma}_2\mathbf{\Sigma}_1^{-1}$ under normal distributions

(Estimated standard errors are in parentheses)

$\mathbf{\Sigma}_2\mathbf{\Sigma}_1^{-1}$	*BU*	*LO*	AI	*BE*	AI	*SB*	AI
$\mathbf{diag}(1,1,1)$	6.192 (0.033)	5.391 (0.033)	12.9%	4.822 (0.032)	22.1%	4.424 (0.032)	28.6%
$\mathbf{diag}(10,1,1)$	6.186 (0.042)	6.008 (0.042)	2.87%	5.853 (0.041)	5.38%	5.931 (0.042)	4.12%
$\mathbf{diag}(100,1,1)$	6.178 (0.047)	6.160 (0.047)	0.30%	6.142 (0.047)	0.60%	6.151 (0.047)	0.44%
$\mathbf{diag}(1000,1,1)$	6.176 (0.047)	6.174 (0.047)	0.03%	6.172 (0.047)	0.06%	6.173 (0.047)	0.05%
$\mathbf{diag}(10,5,1)$	6.198 (0.037)	5.833 (0.037)	5.89%	5.705 (0.036)	7.96%	5.671 (0.036)	8.67%
$\mathbf{diag}(10,10,1)$	6.205 (0.037)	5.790 (0.036)	6.69%	5.691 (0.036)	8.28%	5.517 (0.036)	11.1%
$\mathbf{diag}(100,10,1)$	6.184 (0.044)	6.124 (0.044)	0.97%	6.106 (0.044)	1.27%	6.167 (0.044)	0.27%
$\mathbf{diag}(100,100,1)$	6.205 (0.038)	5.845 (0.037)	5.80%	5.835 (0.037)	5.97%	5.564 (0.037)	10.3%
$\mathbf{diag}(1000,10,1)$	6.177 (0.047)	6.171 (0.047)	0.09%	6.169 (0.047)	0.12%	6.176 (0.047)	0.02%
$\mathbf{diag}(1000,100,1)$	6.183 (0.044)	6.134 (0.044)	0.79%	6.132 (0.044)	0.82%	6.178 (0.044)	0.07%
$\mathbf{diag}(1000,1000,1)$	6.204 (0.038)	5.849 (0.037)	5.73%	5.848 (0.037)	5.75%	5.566 (0.037)	10.3%

Table 7. Estimated risks for estimation of $\mathbf{\Sigma}_2\mathbf{\Sigma}_1^{-1}$ under t-distributions

(Estimated standard errors are in parentheses)

$\mathbf{\Sigma}_2\mathbf{\Sigma}_1^{-1}$	*BU*	*LO*	AI	*BE*	AI	*SB*	AI
diag$(1, 1, 1)$	10.26 (0.145)	8.954 (0.131)	12.8%	8.033 (0.122)	21.7%	7.359 (0.114)	28.3%
diag$(10, 1, 1)$	10.14 (0.146)	9.860 (0.144)	2.77%	9.608 (0.141)	5.25%	9.745 (0.145)	3.91%
diag$(100, 1, 1)$	10.10 (0.152)	10.07 (0.152)	0.29%	10.04 (0.152)	0.58%	10.05 (0.152)	0.44%
diag$(1000, 1, 1)$	10.09 (0.153)	10.09 (0.153)	0.03%	10.08 (0.153)	0.06%	10.08 (0.153)	0.04%
diag$(10, 5, 1)$	10.20 (0.143)	9.609 (0.137)	5.83%	9.401 (0.135)	7.86%	9.341 (0.134)	8.46%
diag$(10, 10, 1)$	10.24 (0.145)	9.568 (0.139)	6.56%	9.408 (0.137)	8.11%	9.123 (0.135)	10.9%
diag$(100, 10, 1)$	10.12 (0.148)	10.03 (0.147)	0.91%	9.999 (0.147)	1.21%	10.11 (0.148)	0.16%
diag$(100, 100, 1)$	10.22 (0.145)	9.633 (0.139)	5.71%	9.616 (0.139)	5.87%	9.169 (0.135)	10.3%
diag$(1000, 10, 1)$	10.09 (0.153)	10.08 (0.152)	0.09%	10.08 (0.152)	0.12%	10.09 (0.153)	0.02%
diag$(1000, 100, 1)$	10.11 (0.148)	10.04 (0.147)	0.74%	10.04 (0.147)	0.77%	10.12 (0.149)	-0.02%
diag$(1000, 1000, 1)$	10.21 (0.145)	9.631 (0.139)	5.65%	9.630 (0.139)	5.66%	9.165 (0.135)	10.2%

Table 8. Estimated risks for estimation of $\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_1^{-1}$ under Kotz-type distributions

(Estimated standard errors are in parentheses)

$\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_1^{-1}$	*BU*	*LO*	AI	*BE*	AI	*SB*	AI
diag$(1,1,1)$	1.072 (0.006)	0.933 (0.006)	12.9%	0.835 (0.006)	22.1%	0.766 (0.006)	28.5%
diag$(1,1,1)$	1.074 (0.007)	1.043 (0.007)	2.87%	1.017 (0.007)	5.37%	1.030 (0.007)	4.14%
diag$(100,1,1)$	1.074 (0.008)	1.071 (0.008)	0.30%	1.068 (0.008)	0.59%	1.070 (0.008)	0.44%
diag$(1000,1,1)$	1.074 (0.008)	1.074 (0.008)	0.03%	1.074 (0.008)	0.06%	1.074 (0.008)	0.05%
diag$(10,5,1)$	1.073 (0.006)	1.010 (0.006)	5.89%	0.988 (0.006)	7.95%	0.982 (0.006)	8.47%
diag$(10,10,1)$	1.072 (0.006)	1.000 (0.006)	6.70%	0.982 (0.006)	8.28%	0.953 (0.006)	11.1%
diag$(100,10,1)$	1.074 (0.008)	1.064 (0.007)	0.96%	1.061 (0.007)	1.25%	1.072 (0.008)	0.25%
diag$(100,100,1)$	1.072 (0.006)	1.009 (0.006)	5.81%	1.008 (0.006)	5.97%	0.961 (0.006)	10.3%
diag$(1000,10,1)$	1.074 (0.008)	1.073 (0.008)	0.09%	1.073 (0.008)	0.12%	1.074 (0.008)	0.01%
diag$(1000,100,1)$	1.074 (0.008)	1.066 (0.008)	0.78%	1.065 (0.008)	0.81%	1.074 (0.008)	0.06%
diag$(1000,1000,1)$	1.072 (0.006)	1.010 (0.006)	5.73%	1.010 (0.006)	5.75%	0.961 (0.006)	10.3%

Table 9. Estimated risks for estimation of $\mathbf{\Sigma}_2\mathbf{\Sigma}_1^{-1}$ under t-distributions (i.i.d.)

(Estimated standard errors are in parentheses)

$\mathbf{\Sigma}_2\mathbf{\Sigma}_1^{-1}$	BU	LO	AI	BE	AI	SB	AI
diag$(1,1,1)$	20.73 (0.883)	19.65 (0.883)	5.21%	18.92 (0.888)	8.72%	17.89 (0.876)	13.7%
diag$(10,1,1)$	20.48 (0.963)	20.28 (0.964)	0.97%	20.06 (0.965)	2.03%	20.01 (0.960)	2.26%
diag$(100,1,1)$	20.57 (1.140)	20.55 (1.140)	0.06%	20.53 (1.140)	0.18%	20.52 (1.140)	0.20%
diag$(1000,1,1)$	20.63 (1.193)	20.63 (1.193)	0.00%	20.63 (1.193)	0.02%	20.62 (1.193)	0.02%
diag$(10,5,1)$	20.14 (0.825)	19.67 (0.825)	2.35%	19.47 (0.826)	3.31%	19.08 (0.816)	5.25%
diag$(10,10,1)$	20.02 (0.792)	19.49 (0.793)	2.63%	19.34 (0.793)	3.38%	18.78 (0.783)	6.22%
diag$(100,10,1)$	20.33 (1.008)	20.29 (1.008)	0.20%	20.26 (1.008)	0.34%	20.27 (1.006)	0.27%
diag$(100,100,1)$	19.97 (0.808)	19.52 (0.809)	2.23%	19.50 (0.809)	2.31%	18.85 (0.799)	5.59%
diag$(1000,10,1)$	20.55 (1.149)	20.55 (1.149)	0.00%	20.55 (1.149)	0.01%	20.55 (1.149)	0.01%
diag$(1000,100,1)$	20.32 (1.014)	20.29 (1.014)	0.14%	20.29 (1.014)	0.15%	20.29 (1.012)	0.16%
diag$(1000,1000,1)$	19.98 (0.810)	19.53 (0.810)	2.21%	19.53 (0.810)	2.21%	18.87 (0.801)	5.56%

Table 10. Estimated risks for estimation of $\mathbf{\Sigma}_2\mathbf{\Sigma}_1^{-1}$ under Kotz-type distributions (i.i.d.)

(Estimated standard errors are in parentheses)

$\mathbf{\Sigma}_2\mathbf{\Sigma}_1^{-1}$	*BU*	*LO*	AI	*BE*	AI	*SB*	AI
diag$(1,1,1)$	7.815 (0.029)	6.333 (0.029)	19.0%	5.297 (0.028)	32.2%	4.703 (0.029)	39.8%
diag$(10,1,1)$	7.761 (0.042)	7.424 (0.041)	4.35%	7.150 (0.041)	7.88%	7.328 (0.043)	5.58%
diag$(100,1,1)$	7.752 (0.048)	7.714 (0.048)	0.49%	7.682 (0.048)	0.90%	7.703 (0.049)	0.63%
diag$(1000,1,1)$	7.753 (0.049)	7.749 (0.049)	0.05%	7.746 (0.049)	0.09%	7.748 (0.049)	0.06%
diag$(10,5,1)$	7.782 (0.035)	7.122 (0.034)	8.49%	6.903 (0.034)	11.3%	7.001 (0.033)	10.0%
diag$(10,10,1)$	7.799 (0.034)	7.022 (0.033)	9.96%	6.854 (0.033)	12.1%	6.629 (0.034)	15.0%
diag$(100,10,1)$	7.756 (0.045)	7.633 (0.045)	1.59%	7.602 (0.045)	2.00%	7.731 (0.046)	0.33%
diag$(100,100,1)$	7.796 (0.035)	7.115 (0.035)	8.73%	7.097 (0.035)	8.96%	6.684 (0.035)	14.3%
diag$(1000,10,1)$	7.752 (0.049)	7.740 (0.049)	0.16%	7.736 (0.049)	0.21%	7.750 (0.049)	0.03%
diag$(1000,100,1)$	7.756 (0.045)	7.654 (0.045)	1.32%	7.651 (0.045)	1.36%	7.751 (0.046)	0.07%
diag$(1000,1000,1)$	7.795 (0.035)	7.123 (0.035)	8.62%	7.122 (0.035)	8.64%	6.689 (0.035)	14.2%