Available online at www.sciencedirect.com

ScienceDirect

Journal of Multivariate Analysis 99 (2008) 751 – 775

Journal of Multivariate Analysis

www.elsevier.com/locate/jmva

Asymptotic distribution of Wishart matrix for block-wise dispersion of population eigenvalues

Yo Sheena[a,*], Akimichi Takemura[b]

[a]*Department of Economics, Shinshu University, 3-1-1 Asahi, Matsumoto, Nagano 3908621, Japan*
[b]*Graduate School of Information Science and Technology, University of Tokyo, Japan*

Received 10 November 2006
Available online 7 April 2007

Abstract

This paper deals with the asymptotic distribution of Wishart matrix and its application to the estimation of the population matrix parameter when the population eigenvalues are block-wise infinitely dispersed. We show that the appropriately normalized eigenvectors and eigenvalues asymptotically generate two Wishart matrices and one normally distributed random matrix, which are mutually independent. For a family of orthogonally equivariant estimators, we calculate the asymptotic risks with respect to the entropy or the quadratic loss function and derive the asymptotically best estimator among the family. We numerically show (1) the convergence in both the distributions and the risks are quick enough for a practical use, (2) the asymptotically best estimator is robust against the deviation of the population eigenvalues from the block-wise infinite dispersion.

AMS 2000 subject classification: 62H10; 62C20

Keywords: Covariance matrix; Wishart distribution; Quadratic loss; Stein's loss; Asymptotic risk

1. Introduction

Suppose that a p-dimensional random vector $\boldsymbol{y}$ has the covariance matrix $\boldsymbol{\Sigma}$. The inference for $\boldsymbol{\Sigma}$ has been studied in enormous amount of literature and is still an important topic from both theoretical and practical points of view. Often we assume some structure of $\boldsymbol{\Sigma}$, i.e., restriction on its parameter space $\{\boldsymbol{\Sigma}|\boldsymbol{\Sigma} > 0\}$. A structure, in some cases, arises from a theoretical reason behind

* Corresponding author. Fax: +81 263 37 2340.
E-mail address: sheena@shinshu-u.ac.jp (Y. Sheena).

0047-259X/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2007.04.001

the data. In other cases, it appears as a result of exploratory analysis such as principle component analysis or exploratory factor analysis.

For example suppose that $\boldsymbol{y}$ is generated in the following multivariate linear model;

$$\boldsymbol{y} = \boldsymbol{B}\boldsymbol{x} + \boldsymbol{e}, \tag{1}$$

where $\boldsymbol{B}$ is a $p \times m$ coefficient (factor loading) matrix with rank $\boldsymbol{B} = m$, $\boldsymbol{x}$ is a latent $m \times 1$ random vector (common factor) and $p \times 1$ vector $\boldsymbol{e}$ is an error term (unique factor) which is independently distributed from $\boldsymbol{x}$. If we further assume that $\boldsymbol{e}$ has $\sigma^2 \boldsymbol{I}_p$ ($\boldsymbol{I}_p$: p-dimensional identity matrix) as its covariance matrix, $\boldsymbol{\Sigma}$ is written as

$$\boldsymbol{\Sigma} = \boldsymbol{B}\boldsymbol{\Sigma}_x \boldsymbol{B}' + \sigma^2 \boldsymbol{I}_p,$$

where $\boldsymbol{\Sigma}_x$ is the nonsingular covariance matrix of $\boldsymbol{x}$. In this case $\boldsymbol{\Sigma}$ has the eigenvalues $\lambda_1 \geqslant \cdots \geqslant \lambda_p$ given by

$$\lambda_i = \begin{cases} \tau_i + \sigma^2 & \text{if } i = 1, \ldots, m, \\ \sigma^2 & \text{if } i = m+1, \ldots, p, \end{cases} \tag{2}$$

where $\tau_i > 0$, $i = 1, \ldots, m$, are the eigenvalues of $\boldsymbol{B}\boldsymbol{\Sigma}_x \boldsymbol{B}'$. It is often observed that σ^2 is quite small compared to τ_i's, which means that the first group of eigenvalues $(\lambda_1, \ldots, \lambda_m)$ is very large compared to the second group $(\lambda_{m+1}, \ldots, \lambda_p)$. In this paper we call this state as "(two-)block-wise dispersion" of the population eigenvalues.

What would happen to the sample covariance matrix, when the eigenvalues of population covariance matrix are "infinitely" dispersed? This is an interesting question from a theoretical standpoint. Takemura and Sheena [14] and Sheena and Takemura [10] deal with this problem under "total dispersion" of population eigenvalues, namely

$$(\lambda_2/\lambda_1, \lambda_3/\lambda_2, \ldots, \lambda_p/\lambda_{p-1}) \to \boldsymbol{0}.$$

This paper is a generalization of Takemura and Sheena [14] from a theoretical point of view, while the practical motivation is as follows; as we saw above, we often come across a practical situation where the population eigenvalues are block-wise dispersed. It is helpful for the inference on $\boldsymbol{\Sigma}$ in practical situations to understand the behavior of the sample covariance matrix, when the population eigenvalues are block-wise "infinitely" dispersed. The state of the population eigenvalues being infinitely dispersed is a theoretical approximation, but understanding the limiting behavior leads to a better insight on its neighborhood where the eigenvalues are "largely" dispersed.

Now we formally state the framework of this paper. Let $\boldsymbol{S} = (s_{ij})$ be distributed according to Wishart distribution $\boldsymbol{W}_p(n, \boldsymbol{\Sigma})$, where p is the dimension, n is the degrees of freedom, and $\boldsymbol{\Sigma}$ is the covariance matrix. The spectral decompositions of $\boldsymbol{\Sigma}$ and $\boldsymbol{S}$ are given by

$$\boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{\Gamma}', \quad \boldsymbol{S} = \boldsymbol{G}\boldsymbol{L}\boldsymbol{G}',$$

where $\boldsymbol{G}, \boldsymbol{\Gamma} \in \mathcal{O}(p)$, the group of $p \times p$ orthogonal matrices, and $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_p)$, $\boldsymbol{L} = \mathrm{diag}(l_1, \ldots, l_p)$, are diagonal matrices with the eigenvalues $\lambda_1 \geqslant \cdots \geqslant \lambda_p > 0$, $l_1 \geqslant \cdots \geqslant l_p > 0$ of $\boldsymbol{\Sigma}$ and $\boldsymbol{S}$, respectively. We use the notations $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_p)$ and $\boldsymbol{l} = (l_1, \ldots, l_p)$ hereafter. By the requirement that

$$\widetilde{\boldsymbol{G}} = (\widetilde{g}_{ij}) = \boldsymbol{\Gamma}'\boldsymbol{G}$$

has positive diagonal elements, the spectral decomposition $\boldsymbol{S} = \boldsymbol{GLG}'$ is almost surely uniquely determined. Then almost surely there exists a one-to-one correspondence between the set $\{\boldsymbol{S}|\boldsymbol{S} > 0\}$ and $\mathcal{L} \times \mathcal{O}^+(p)$, where

$$\mathcal{L} = \{\boldsymbol{l}|l_1 > \cdots > l_p > 0\}, \quad \mathcal{O}^+(p) = \{\widetilde{\boldsymbol{G}} \in \mathcal{O}(p)|\widetilde{g}_{ii} > 0,\ 1 \leqslant i \leqslant p\}.$$

Let m (m_i in Section 2.3) denote the dividing point of the first block and the second block of the eigenvalues. Now we parameterize $\boldsymbol{\lambda}$,$\boldsymbol{l}$ as follows:

$$\lambda_i = \begin{cases} \xi_i \alpha & \text{if } i = 1, \ldots, m, \\ \xi_i \beta & \text{if } i = m+1, \ldots, p, \end{cases} \tag{3}$$

$$l_i = \begin{cases} d_i \alpha & \text{if } i = 1, \ldots, m, \\ d_i \beta & \text{if } i = m+1, \ldots, p. \end{cases} \tag{4}$$

In this paper we always consider ξ's are given and fixed. We also use the notations,

$$\boldsymbol{\Xi} = \mathrm{diag}(\xi_1, \ldots, \xi_p), \quad \boldsymbol{\xi} = (\xi_1, \ldots, \xi_p),$$
$$\boldsymbol{D} = \mathrm{diag}(d_1, \ldots, d_p), \quad \boldsymbol{d} = (d_1, \ldots, d_p).$$

We will investigate the asymptotic distribution of $\boldsymbol{S}$ as β/α goes to 0 while $\boldsymbol{\Xi}$ is fixed and its application to the estimation of $\boldsymbol{\Sigma}$. The state $\beta/\alpha \approx 0$ means that the eigenvalues of $\boldsymbol{\Sigma}$ are two-block-wise "largely" dispersed. In the following, the notation $\beta/\alpha \to 0$ means a limiting operation $n \to \infty$ with arbitrary sequences $\alpha_n, \beta_n, n = 1, 2, \ldots$, such that $\beta_n/\alpha_n \to 0$.

We briefly describe the content of the following sections. In Section 2.1 we prepare a local coordinate system of $\mathcal{O}^+(p)$ around $\boldsymbol{I}_p$. In Section 2.2 we present our main results on asymptotic distributions and we further discuss the case of multi-block-wise infinite dispersion in Section 2.3. Section 3 deals with the estimation of $\boldsymbol{\Sigma}$ from decision-theoretic framework. In Section 3.1 we introduce orthogonally equivariant estimators and two loss functions and in Section 3.2 we calculate the asymptotic risks. We concentrate on the special case of block-wise identity covariance matrices in Section 3.3, which is practically important, and we propose the best estimator for the case with respect to each loss function. In Section 3.4 the convergence speed of both distributions and risks are numerically evaluated. Together with the application to discriminant analysis, the numerical comparisons show the superiority of the new estimators.

Before concluding this subsection, we introduce some notational conventions in this paper. In the sections other than Section 2.3, we always consider a same two-block partition of matrices. For $\boldsymbol{A} = (a_{ij})$, a $p \times p$ matrix, $\boldsymbol{A}_{ij}$ $(1 \leqslant i, j \leqslant 2)$ denotes the (i, j)-block in the partition

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{pmatrix}, \quad \boldsymbol{A}_{11} : m \times m, \quad \boldsymbol{A}_{22} : (p-m) \times (p-m).$$

If $\boldsymbol{A}$ is block diagonal, i.e., $\boldsymbol{A}_{12} = \boldsymbol{A}_{21} = \boldsymbol{0}$, we write

$$\boldsymbol{A} = \mathrm{diag}(\boldsymbol{A}_{11}, \boldsymbol{A}_{22}) = \begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A}_{22} \end{pmatrix}.$$

For the particular case of diagonal matrix $\boldsymbol{A} = \mathrm{diag}(a_1, \ldots, a_p)$, we simply write $\boldsymbol{A}_1, \boldsymbol{A}_2$ instead of $\boldsymbol{A}_{11}, \boldsymbol{A}_{22}$, i.e., $\boldsymbol{A}_1 = \mathrm{diag}(a_1, \ldots, a_m)$, $\boldsymbol{A}_2 = \mathrm{diag}(a_{m+1}, \ldots, a_p)$. Let $\boldsymbol{a} = (a_{ij})_{1 \leqslant j < i \leqslant p}$ denote the vector of the elements in the lower triangular part of $\boldsymbol{A}$, which is correspondingly partitioned as $\boldsymbol{a} = (\boldsymbol{a}_{11}, \boldsymbol{a}_{22}, \boldsymbol{a}_{21})$, where

$$\boldsymbol{a}_{11} = (a_{ij})_{1 \leqslant j < i \leqslant m}, \quad \boldsymbol{a}_{22} = (a_{ij})_{m+1 \leqslant j < i \leqslant p}, \quad \boldsymbol{a}_{21} = (a_{ij})_{1 \leqslant j \leqslant m < i \leqslant p}.$$

If $\boldsymbol{a}$ is a p-dimensional row vector, i.e., $\boldsymbol{a} = (a_1, \dots, a_p)$, then we make a partition of $\boldsymbol{a}$ as

$$\boldsymbol{a} = (\boldsymbol{a}_1, \boldsymbol{a}_2), \quad \boldsymbol{a}_1 = (a_1, \dots, a_m), \quad \boldsymbol{a}_2 = (a_{m+1}, \dots, a_p).$$

We write etr $X = \exp(\mathrm{tr}\, X)$ for a square matrix X.

2. Asymptotic distribution

2.1. Local coordinates

We consider a local coordinate of $\mathcal{O}^+(p)$, $\boldsymbol{u} = (u_{ij})_{1 \leqslant j < i \leqslant p}$, around the identity matrix $\boldsymbol{I}_p$. For the proof of the existence of such coordinate, see Takemura and Sheena [14, Appendix B]. We have the following open sets U, V and functions ϕ_{ij}, $1 \leqslant i \leqslant j \leqslant p$;

$$\mathbf{0} \in U \subset R^{p(p-1)/2},$$
$$\boldsymbol{I}_p \in V \subset \mathcal{O}^+(p),$$

and $\phi_{ij}(\boldsymbol{u})$ is a C^∞ function on U such that $\boldsymbol{G}(\boldsymbol{u}) = (g_{ij}(\boldsymbol{u}))$ defined by

$$\begin{cases} g_{ij}(\boldsymbol{u}) = \phi_{ij}(\boldsymbol{u}), & 1 \leqslant i \leqslant j \leqslant p, \\ g_{ij}(\boldsymbol{u}) = u_{ij}, & 1 \leqslant j < i \leqslant p \end{cases} \tag{5}$$

is a one-to-one function from U onto V. Using V we can construct a finite open covering of $\mathcal{O}^+(p)$ as follows. For $\boldsymbol{H}_1 \in \mathcal{O}^+(m)$, $\boldsymbol{H}_2 \in \mathcal{O}^+(p-m)$, let

$$V(\boldsymbol{H}_1, \boldsymbol{H}_2) = \mathrm{diag}(\boldsymbol{H}_1, \boldsymbol{H}_2) V \cap \mathcal{O}^+(p) = \{\boldsymbol{G} | \boldsymbol{G} = \mathrm{diag}(\boldsymbol{H}_1, \boldsymbol{H}_2)\boldsymbol{G}^*,\ \exists \boldsymbol{G}^* \in V\} \cap \mathcal{O}^+(p).$$

denote the open neighborhood of $\mathrm{diag}(\boldsymbol{H}_1, \boldsymbol{H}_2)$. Let

$$\mathcal{O}(m, p-m) = \{\mathrm{diag}(\boldsymbol{H}_1, \boldsymbol{H}_2) | \boldsymbol{H}_1 \in \mathcal{O}^+(m), \boldsymbol{H}_2 \in \mathcal{O}^+(p-m)\},$$

then

$$\mathcal{O}(m, p-m) \subset \bigcup_{\boldsymbol{H}_1 \in \mathcal{O}^+(m), \boldsymbol{H}_2 \in \mathcal{O}^+(p-m)} V(\boldsymbol{H}_1, \boldsymbol{H}_2).$$

Since $\mathcal{O}(m, p-m)$ is compact, we can choose a finite number of sets $O^{(\tau)} = V(\boldsymbol{H}_1^{(\tau)}, \boldsymbol{H}_2^{(\tau)})$, $\tau = 1, \dots, T$, such that $\bigcup_{\tau=1}^T O^{(\tau)} \supset \mathcal{O}(m, p-m)$.

For $O^{(\tau)}$, $1 \leqslant \tau \leqslant T$, we can use $\boldsymbol{u}$ as a local coordinate since $\boldsymbol{G}$ in $O^{(\tau)}$ can be uniquely expressed as $\boldsymbol{H}^{(\tau)}\boldsymbol{G}(\boldsymbol{u})$ with some $\boldsymbol{u}$ in U, where

$$\boldsymbol{H}^{(\tau)} = \mathrm{diag}(\boldsymbol{H}_1^{(\tau)}, \boldsymbol{H}_2^{(\tau)}), \quad \tau = 1, \dots, T. \tag{6}$$

Now we have $\boldsymbol{u}$ as a local coordinate on each $O^{(\tau)}$, $\tau = 1, \dots, T$. We need another local coordinate to investigate the asymptotic behavior of $\boldsymbol{S}$. Let $\boldsymbol{q} = (q_{ij})_{1 \leqslant j < i \leqslant p}$ be defined as follows as another coordinate on $O^{(\tau)}$ for a fixed τ, $\tau = 1, \dots, T$; if $1 \leqslant j \leqslant m < i \leqslant p$,

$$\begin{aligned} q_{ij} &= l_j^{1/2} \lambda_i^{-1/2} \sum_{t=m+1}^{p} (\boldsymbol{H}_2^{(\tau)})_{i-m, t-m} u_{tj} \\ &= \alpha^{1/2} \beta^{-1/2} d_j^{1/2} \xi_i^{-1/2} \sum_{t=m+1}^{p} (\boldsymbol{H}_2^{(\tau)})_{i-m, t-m} u_{tj} \end{aligned} \tag{7}$$

and $q_{ij} = u_{ij}$ otherwise. If we use matrices $\boldsymbol{Q} = (q_{ij}),\ \boldsymbol{U} = (u_{ij})$ and their partitions, (7) is the same as

$$\boldsymbol{Q}_{21} = \alpha^{1/2}\beta^{-1/2}\boldsymbol{\Xi}_2^{-1/2}\boldsymbol{H}_2^{(\tau)}\boldsymbol{U}_{21}\boldsymbol{D}_1^{1/2}, \quad \boldsymbol{Q}_{11} = \boldsymbol{U}_{11}, \quad \boldsymbol{Q}_{22} = \boldsymbol{U}_{22}. \tag{8}$$

Conversely

$$\boldsymbol{U}_{21} = \alpha^{-1/2}\beta^{1/2}\boldsymbol{H}_2^{(\tau)\prime}\boldsymbol{\Xi}_2^{1/2}\boldsymbol{Q}_{21}\boldsymbol{D}_1^{-1/2}, \quad \boldsymbol{U}_{11} = \boldsymbol{Q}_{11}, \quad \boldsymbol{U}_{22} = \boldsymbol{Q}_{22}, \tag{9}$$

or

$$u_{ij} = \begin{cases} \alpha^{-1/2}\beta^{1/2} \displaystyle\sum_{t=m+1}^{p} (\boldsymbol{H}_2^{(\tau)})_{t-m,i-m} q_{tj} \xi_t^{1/2} d_j^{-1/2} & \text{if } 1 \leqslant j \leqslant m < i \leqslant p, \\ q_{ij} & \text{otherwise.} \end{cases} \tag{10}$$

2.2. Main results

The following theorem says that $\widetilde{\boldsymbol{G}}$ asymptotically separates into two orthogonal matrices $\widetilde{\boldsymbol{G}}_{11}, \widetilde{\boldsymbol{G}}_{22}$ on the diagonal blocks.

Theorem 1. (1) *As* $\beta/\alpha \to 0$, $\widetilde{\boldsymbol{G}}_{21} \xrightarrow{p} \boldsymbol{0}$.
(2) $\lim_{\beta/\alpha \to 0} P(\widetilde{\boldsymbol{G}} \in O) = 1$ *for any open set* $O \subset \mathcal{O}^+(p)$ *including* $\mathcal{O}(m, p-m)$.

Proof. Since 2 is easily proved from 1, we only prove 1 here. Let

$$\bar{\boldsymbol{S}} = (\bar{s}_{ij}) = \boldsymbol{\Lambda}^{-1/2}\boldsymbol{\Gamma}'\boldsymbol{S}\boldsymbol{\Gamma}\boldsymbol{\Lambda}^{-1/2} = \boldsymbol{\Lambda}^{-1/2}\widetilde{\boldsymbol{G}}\boldsymbol{L}\widetilde{\boldsymbol{G}}'\boldsymbol{\Lambda}^{-1/2} \sim \boldsymbol{W}_p(n, \boldsymbol{I}_p),$$

Suppose $1 \leqslant j \leqslant m < i \leqslant p$. Note that

$$\bar{s}_{ii} = (\tilde{g}_{i1}^2 l_1 + \cdots + \tilde{g}_{ip}^2 l_p)\lambda_i^{-1}.$$

Therefore,

$$\tilde{g}_{ij}^2 \leqslant \bar{s}_{ii}\frac{\lambda_i}{l_j} = \bar{s}_{ii}\frac{\lambda_j}{l_j}\frac{\lambda_i}{\lambda_j} \leqslant \bar{s}_{ii}\frac{\lambda_j}{l_j}\frac{\xi_i}{\xi_j}\frac{\beta}{\alpha}. \tag{11}$$

Since $\bar{s}_{ii}$ is distributed independently of $\boldsymbol{\Sigma}$, for any $\epsilon > 0$, there exists M such that

$$P(\bar{s}_{ii} < M) > 1 - \epsilon, \quad \forall \boldsymbol{\Sigma}. \tag{12}$$

Besides, from the result of Lemma 1 of Takemura and Sheena [14], for any $\epsilon > 0$, there exists C such that

$$P\left(\frac{\lambda_j}{l_j} < C\right) > 1 - \epsilon, \quad \forall \boldsymbol{\Sigma}. \tag{13}$$

From (12) and (13) we have

$$\bar{s}_{ii}\frac{\lambda_j}{l_j}\frac{\beta}{\alpha} \xrightarrow{p} 0 \quad \text{as } \frac{\beta}{\alpha} \to 0.$$

From this fact and (11) we have

$$\tilde{g}_{ij}^2 \xrightarrow{p} 0 \quad \text{as } \frac{\beta}{\alpha} \to 0, \quad 1 \leqslant \forall j \leqslant m < \forall i \leqslant p. \qquad \square$$

Next we state a rather technical lemma, which will be used in the proofs of some theorems. Consider a random variable $x(\boldsymbol{G}, \boldsymbol{l}, \boldsymbol{\lambda}, \alpha, \beta)$. We are often interested in the asymptotic expectation of $x(\boldsymbol{G}, \boldsymbol{l}, \boldsymbol{\lambda}, \alpha, \beta)$ as $\beta/\alpha \to 0$ while $\boldsymbol{\Gamma}$ is fixed. For fixed $\boldsymbol{\Gamma}$ and $\boldsymbol{H}^{(\tau)} (\tau = 1, \ldots, T)$, somewhat abusing the notation, let

$$x(\boldsymbol{d}, \boldsymbol{q}, \xi, \alpha, \beta; \boldsymbol{\Gamma}, \boldsymbol{H}^{(\tau)}) = x(\boldsymbol{\Gamma}\boldsymbol{H}^{(\tau)}\boldsymbol{G}(\boldsymbol{u}(\boldsymbol{d}, \boldsymbol{q}, \xi, \alpha, \beta)), \boldsymbol{l}(\boldsymbol{d}, \alpha, \beta), \boldsymbol{\lambda}(\xi, \alpha, \beta), \alpha, \beta) \tag{14}$$

for emphasizing the right-hand side as the function of $(\boldsymbol{d}, \boldsymbol{q}, \xi, \alpha, \beta)$, where $\boldsymbol{G}(\boldsymbol{u})$, $\boldsymbol{u}(\boldsymbol{d}, \boldsymbol{q}, \xi, \alpha, \beta)$, $\boldsymbol{l}(\boldsymbol{d}, \alpha, \beta)$, $\boldsymbol{\lambda}(\xi, \alpha, \beta)$ are, respectively, defined by (5), (10), (4) and (3). For $\boldsymbol{u} = (\boldsymbol{u}_{11}, \boldsymbol{u}_{22}, \boldsymbol{u}_{21})$, we have

$$\lim_{\beta/\alpha \to 0} \boldsymbol{u}(\boldsymbol{d}, \boldsymbol{q}, \xi, \alpha, \beta) = \lim_{\beta/\alpha \to 0} (\boldsymbol{u}_{11}(\boldsymbol{q}_{11}), \boldsymbol{u}_{22}(\boldsymbol{q}_{22}), \boldsymbol{u}_{21}(\boldsymbol{d}, \boldsymbol{q}, \xi, \alpha, \beta)) = (\boldsymbol{q}_{11}, \boldsymbol{q}_{22}, \boldsymbol{0}), \tag{15}$$

hence

$$\lim_{\beta/\alpha \to 0} \boldsymbol{G}(\boldsymbol{u}(\boldsymbol{d}, \boldsymbol{q}, \xi, \alpha, \beta)) = \boldsymbol{G}(\boldsymbol{q}_{11}, \boldsymbol{q}_{22}, \boldsymbol{0}). \tag{16}$$

Lemma 1. *Suppose that there exist some* $a < \frac{1}{2}$ *and* $b > 0$ *such that*

$$|x(\boldsymbol{\Gamma}\boldsymbol{G}, \boldsymbol{l}, \boldsymbol{\lambda}, \alpha, \beta)| \leqslant b \operatorname{etr}(a\boldsymbol{G}\boldsymbol{L}\boldsymbol{G}'\boldsymbol{\Lambda}^{-1}) \text{ a.e. in } (\boldsymbol{G}, \boldsymbol{l}) \tag{17}$$

and suppose that for each τ, $\tau = 1, \ldots, T$, $\lim_{\beta/\alpha \to 0} x(\boldsymbol{d}, \boldsymbol{q}, \xi, \alpha, \beta; \boldsymbol{\Gamma}, \boldsymbol{H}^{(\tau)})$ *exists and equals to a function*

$$\bar{x}_{\boldsymbol{\Gamma}}(\boldsymbol{H}^{(\tau)}\boldsymbol{G}(\boldsymbol{q}_{11}, \boldsymbol{q}_{22}, \boldsymbol{0}), \boldsymbol{d}, \boldsymbol{Q}_{21}, \xi). \tag{18}$$

Then

$$\lim_{\beta/\alpha \to 0} E[x(\boldsymbol{G}, \boldsymbol{l}, \boldsymbol{\lambda}, \alpha, \beta)]$$

$$= E[\bar{x}_{\boldsymbol{\Gamma}}(\operatorname{diag}(\boldsymbol{G}_{11}(\boldsymbol{W}_{11}), \boldsymbol{G}_{22}(\boldsymbol{W}_{22})), (\boldsymbol{d}_1(\boldsymbol{W}_{11}), \boldsymbol{d}_2(\boldsymbol{W}_{22})), \boldsymbol{Z}_{21}, \xi)], \tag{19}$$

where the expectation on the right side of (19) *is taken with respect to the following mutually independent distributions*

$$\boldsymbol{W}_{11} \sim \boldsymbol{W}_m(n, \boldsymbol{\Xi}_1),$$

$$\boldsymbol{W}_{22} \sim \boldsymbol{W}_{p-m}(n-m, \boldsymbol{\Xi}_2),$$

$$\boldsymbol{Z}_{21} \sim \boldsymbol{N}_{(p-m)\times m}(\boldsymbol{0}, \boldsymbol{I}_{p-m} \otimes \boldsymbol{I}_m), \tag{20}$$

and $\boldsymbol{G}_{ss}(\boldsymbol{W}_{ss}), \boldsymbol{d}_s(\boldsymbol{W}_{ss})$, $s = 1, 2$, *are the components in the unique spectral decomposition of* $\boldsymbol{W}_{ss}$ *for* $s = 1, 2$;

$$\boldsymbol{W}_{11} = \boldsymbol{G}_{11}\boldsymbol{D}_1\boldsymbol{G}_{11}', \quad \boldsymbol{D}_1 = \operatorname{diag}(d_1, \ldots, d_m), \quad \boldsymbol{d}_1 = (d_1, \ldots, d_m),$$

$$\boldsymbol{W}_{22} = \boldsymbol{G}_{22}\boldsymbol{D}_2\boldsymbol{G}_{22}', \quad \boldsymbol{D}_2 = \operatorname{diag}(d_{m+1}, \ldots, d_p), \quad \boldsymbol{d}_2 = (d_{m+1}, \ldots, d_p). \tag{21}$$

The proof is omitted. See Sheena and Takemura [11] for the proof.

The following theorem on the asymptotic distributions is actually a corollary of Lemma 1. Let

$$\widetilde{\boldsymbol{W}}_{11} = \widetilde{\boldsymbol{G}}_{11}\boldsymbol{D}_1\widetilde{\boldsymbol{G}}'_{11},$$

$$\widetilde{\boldsymbol{W}}_{22} = \widetilde{\boldsymbol{G}}_{22}\boldsymbol{D}_2\widetilde{\boldsymbol{G}}'_{22},$$

$$\widetilde{\boldsymbol{Z}}_{21} = \alpha^{1/2}\beta^{-1/2}\boldsymbol{\Xi}_2^{-1/2}\widetilde{\boldsymbol{G}}_{21}\boldsymbol{D}_1^{1/2},$$

where all the elements on the right-hand side are defined in Section 1.

Theorem 2. *As* $\beta/\alpha \to 0$,

$$\widetilde{\boldsymbol{W}}_{11} \xrightarrow{d} \boldsymbol{W}_m(n, \boldsymbol{\Xi}_1),$$

$$\widetilde{\boldsymbol{W}}_{22} \xrightarrow{d} \boldsymbol{W}_{p-m}(n-m, \boldsymbol{\Xi}_2),$$

$$\widetilde{\boldsymbol{Z}}_{21} \xrightarrow{d} \boldsymbol{N}_{(p-m)\times m}(\boldsymbol{0}, \boldsymbol{I}_{p-m} \otimes \boldsymbol{I}_m)$$

and $\widetilde{\boldsymbol{W}}_{11}, \widetilde{\boldsymbol{W}}_{22}, \widetilde{\boldsymbol{Z}}_{21}$ *are asymptotically mutually independently distributed.*

Proof. Let $\boldsymbol{\Theta}_{11} : m \times m$ symmetric matrix, $\boldsymbol{\Theta}_{22} : (p-m) \times (p-m)$ symmetric matrix and $\boldsymbol{\Theta}_{21} : m \times (p-m)$ matrix. Consider the moment generating function

$$\begin{aligned} x(\boldsymbol{G}, \boldsymbol{l}, \boldsymbol{\lambda}, \alpha, \beta) &= \exp(\operatorname{tr} \widetilde{\boldsymbol{W}}_{11}\boldsymbol{\Theta}_{11} + \operatorname{tr} \widetilde{\boldsymbol{W}}_{22}\boldsymbol{\Theta}_{22} + \operatorname{tr} \widetilde{\boldsymbol{Z}}_{21}\boldsymbol{\Theta}_{21}) \\ &= \exp\left(\sum_{s=1}^{2} \operatorname{tr} \widetilde{\boldsymbol{W}}_{ss}\boldsymbol{\Theta}_{ss} + \operatorname{tr} \widetilde{\boldsymbol{Z}}_{21}\boldsymbol{\Theta}_{21}\right). \end{aligned}$$

For $\boldsymbol{H}^{(\tau)} = \operatorname{diag}(\boldsymbol{H}_1^{(\tau)}, \boldsymbol{H}_2^{(\tau)}),\ \boldsymbol{H}_1^{(\tau)} \in \mathcal{O}^+(m),\ \boldsymbol{H}_2^{(\tau)} \in \mathcal{O}^+(p-m)$, we have

$$\begin{aligned} x(\boldsymbol{\Gamma}\boldsymbol{H}^{(\tau)}\boldsymbol{G}(\boldsymbol{u}), \boldsymbol{l}, \boldsymbol{\lambda}, \alpha, \beta) = \exp\Bigg\{ &\sum_{s=1}^{2} \operatorname{tr} (\boldsymbol{H}^{(\tau)}\boldsymbol{G}(\boldsymbol{u}))_{ss}\boldsymbol{D}_s(\boldsymbol{H}^{(\tau)}\boldsymbol{G}(\boldsymbol{u}))'_{ss}\boldsymbol{\Theta}_{ss} \\ &+ \operatorname{tr} \alpha^{1/2}\beta^{-1/2}\boldsymbol{\Xi}_2^{-1/2}(\boldsymbol{H}^{(\tau)}\boldsymbol{G}(\boldsymbol{u}))_{21}\boldsymbol{D}_1^{1/2}\boldsymbol{\Theta}_{21} \Bigg\}. \end{aligned}$$

From (5)

$$(\boldsymbol{H}^{(\tau)}\boldsymbol{G}(\boldsymbol{u}))_{21} = \boldsymbol{H}_2^{(\tau)}\boldsymbol{U}_{21},$$

hence from (8)

$$\alpha^{1/2}\beta^{-1/2}\boldsymbol{\Xi}_2^{-1/2}(\boldsymbol{H}^{(\tau)}\boldsymbol{G}(\boldsymbol{u}))_{21}\boldsymbol{D}_1^{1/2} = \boldsymbol{Q}_{21}.$$

This leads to

$$x(\boldsymbol{d}, \boldsymbol{q}, \boldsymbol{\xi}, \alpha, \beta; \boldsymbol{\Gamma}, \boldsymbol{H}^{(\tau)}) = \exp\left\{\sum_{s=1}^{2} \operatorname{tr} (\boldsymbol{H}^{(\tau)}\boldsymbol{G}(\boldsymbol{u}))_{ss}\boldsymbol{D}_s(\boldsymbol{H}^{(\tau)}\boldsymbol{G}(\boldsymbol{u}))'_{ss}\boldsymbol{\Theta}_{ss} + \operatorname{tr} \boldsymbol{Q}_{21}\boldsymbol{\Theta}_{21}\right\},$$

with $\boldsymbol{u} = \boldsymbol{u}(\boldsymbol{d}, \boldsymbol{q}, \boldsymbol{\xi}, \alpha, \beta)$. Therefore from (16)

$$\lim_{\beta/\alpha\to 0} x(\boldsymbol{d}, \boldsymbol{q}, \boldsymbol{\xi}, \alpha, \beta; \boldsymbol{\Gamma}, \boldsymbol{H}^{(\tau)})$$
$$= \exp\left\{\sum_{s=1}^{2} \mathrm{tr}\,(\boldsymbol{H}^{(\tau)}\boldsymbol{G}(\boldsymbol{q}_{11}, \boldsymbol{q}_{22}, \boldsymbol{0}))_{ss}\boldsymbol{D}_s(\boldsymbol{H}^{(\tau)}\boldsymbol{G}(\boldsymbol{q}_{11}, \boldsymbol{q}_{22}, \boldsymbol{0}))'_{ss}\boldsymbol{\Theta}_{ss} + \mathrm{tr}\,\boldsymbol{Q}_{21}\boldsymbol{\Theta}_{21}\right\}.$$

From Lemma 1,

$$\lim_{\beta/\alpha\to 0} E[\exp(\mathrm{tr}\,\widetilde{\boldsymbol{W}}_{11}\boldsymbol{\Theta}_{11} + \mathrm{tr}\,\widetilde{\boldsymbol{W}}_{22}\boldsymbol{\Theta}_{22} + \mathrm{tr}\,\widetilde{\boldsymbol{Z}}_{21}\boldsymbol{\Theta}_{21})]$$
$$= E\left[\exp\left\{\sum_{s=1}^{2} \mathrm{tr}\,\boldsymbol{G}_{ss}(\boldsymbol{W}_{ss})\boldsymbol{D}_s(\boldsymbol{W}_{ss})\boldsymbol{G}_{ss}(\boldsymbol{W}_{ss})'\boldsymbol{\Theta}_{ss} + \mathrm{tr}\,\boldsymbol{Z}_{21}\boldsymbol{\Theta}_{21}\right\}\right]$$
$$= E[\mathrm{etr}\,\boldsymbol{W}_{11}\boldsymbol{\Theta}_{11}]E[\mathrm{etr}\,\boldsymbol{W}_{22}\boldsymbol{\Theta}_{22}]E[\mathrm{etr}\,\boldsymbol{Z}_{21}\boldsymbol{\Theta}_{21}],$$

where in the second and third equations the expectations are taken with respect to the distributions (20) in Lemma 1. □

2.3. Multi-block partition

In this section, we extend Theorem 2 into multi-block cases. We partition $(1, \ldots, p)$ into k blocks;

$$\begin{array}{ll} \text{1st block} & (m_0+1, \cdots, m_1), \\ \text{2nd block} & (m_1+1, \ldots, m_2), \\ & \vdots \\ k\text{th block} & (m_{k-1}+1, \cdots, m_k), \end{array}$$

where

$$m_0 = 0 < m_1 < m_2 < \cdots < m_k = p.$$

Let $[i]$, $i = 1, \ldots, p$, denote the block containing i, i.e.,

$$[i] = s \quad \text{if } m_{s-1} + 1 \leqslant i \leqslant m_s.$$

We also use the notations $\bar{m}_s = m_s - m_{s-1},\ s = 1, \ldots, k$, for the block sizes.

Correspondingly to the above partition, we make the following partition of a $p \times p$ matrix $\boldsymbol{A} = (a_{ij})$:

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{A}_{11} & \cdots & \boldsymbol{A}_{1k} \\ \vdots & \ddots & \vdots \\ \boldsymbol{A}_{k1} & \cdots & \boldsymbol{A}_{kk} \end{pmatrix}, \quad \boldsymbol{A}_{st} : \bar{m}_s \times \bar{m}_t \text{ matrix}, \quad 1 \leqslant s, t \leqslant k.$$

For a diagonal matrix $\boldsymbol{A} = \mathrm{diag}(a_1, \ldots, a_p)$, we use the notation

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{A}_1 & & \boldsymbol{0} \\ & \ddots & \\ \boldsymbol{0} & & \boldsymbol{A}_k \end{pmatrix}, \quad \boldsymbol{A}_s = \mathrm{diag}(a_{m_{s-1}+1}, \ldots, a_{m_s}), \quad s = 1, \ldots, k.$$

Consider the following parametrization of $\boldsymbol{l}$, $\boldsymbol{\lambda}$

$$\lambda_i = \xi_i \alpha_{[i]}, \quad 1 \leqslant i \leqslant p,$$

$$l_i = d_i \alpha_{[i]}, \quad 1 \leqslant i \leqslant p.$$

In this subsection we again consider that ξ_i's are fixed. Now we define $\widetilde{\boldsymbol{W}}_{ss}$, $\widetilde{\boldsymbol{Z}}_{st}$, $1 \leqslant t < s \leqslant k$;

$$\widetilde{\boldsymbol{W}}_{ss} = \widetilde{\boldsymbol{G}}_{ss} \boldsymbol{D}_s \widetilde{\boldsymbol{G}}'_{ss},$$

$$\widetilde{\boldsymbol{Z}}_{st} = \alpha_t^{1/2} \alpha_s^{-1/2} \boldsymbol{\Xi}_s^{-1/2} \widetilde{\boldsymbol{G}}_{st} \boldsymbol{D}_t^{1/2}.$$

where notations of the right-hand side are defined in Section 1. The following theorem is the extension of Theorem 2.

Theorem 3. *As* $(\alpha_2/\alpha_1, \alpha_3/\alpha_2, \ldots, \alpha_k/\alpha_{k-1}) \to \boldsymbol{0}$,

$$\widetilde{\boldsymbol{W}}_{ss} \xrightarrow{d} \boldsymbol{W}_{\alpha_s}(n - m_{s-1}, \boldsymbol{\Xi}_s), \quad 1 \leqslant s \leqslant k,$$

$$\widetilde{\boldsymbol{Z}}_{st} \xrightarrow{d} \boldsymbol{N}_{\bar{m}_s \times \bar{m}_t}(\boldsymbol{0}, \boldsymbol{I}_{\bar{m}_s} \otimes \boldsymbol{I}_{\bar{m}_t}), \quad 1 \leqslant t < s \leqslant k,$$

and $\widetilde{\boldsymbol{W}}_{ss} (1 \leqslant s \leqslant k)$, $\widetilde{\boldsymbol{Z}}_{st} (1 \leqslant t < s \leqslant k)$ *are asymptotically mutually independently distributed.*

Proof. Though we can prove the theorem in the same manner as the proof of Theorem 2, it is notationally too cumbersome. Instead we will prove the theorem by using Theorem 2 recursively. Let $r_1 = \alpha_1$ and $r_t = \alpha_t/\alpha_{t-1},\ t = 2, \ldots k$, then $\prod_{t=1}^{s} r_t = \alpha_s,\ s = 1, \ldots, k$. Note for $1 \leqslant i \leqslant p$,

$$l_i = d_i \alpha_{[i]} = d_i \prod_{t=1}^{[i]} r_t, \quad \lambda_i = \xi_i \alpha_{[i]} = \xi_i \prod_{t=1}^{[i]} r_t.$$

We consider the moment generating function

$$E\left[\exp\left(\operatorname{tr} \sum_{s=1}^{k} \widetilde{\boldsymbol{W}}_{ss} \boldsymbol{\Theta}_{ss} + \operatorname{tr} \sum_{1 \leqslant t < s \leqslant k} \widetilde{\boldsymbol{Z}}_{st} \boldsymbol{\Theta}_{st}\right)\right],$$

where $\boldsymbol{\Theta}_{ss} (1 \leqslant s \leqslant k)$ and $\boldsymbol{\Theta}_{st} (1 \leqslant t < s \leqslant k)$ are, respectively, an $\bar{m}_s \times \bar{m}_s$ symmetric matrix and an $\bar{m}_t \times \bar{m}_s$ matrix. We have

$$\begin{aligned}
&\lim_{(a_2/a_1, \ldots, a_k/a_{k-1}) \to 0} E\left[\exp\left(\operatorname{tr} \sum_{s=1}^{k} \widetilde{\boldsymbol{W}}_{ss} \boldsymbol{\Theta}_{ss} + \operatorname{tr} \sum_{1 \leqslant t < s \leqslant k} \widetilde{\boldsymbol{Z}}_{st} \boldsymbol{\Theta}_{st}\right)\right] \\
&= \lim_{(r_2, \ldots, r_k) \to 0} E\left[\exp\left(\operatorname{tr} \sum_{s=1}^{k} \widetilde{\boldsymbol{W}}_{ss} \boldsymbol{\Theta}_{ss} + \operatorname{tr} \sum_{1 \leqslant t < s \leqslant k} \widetilde{\boldsymbol{Z}}_{st} \boldsymbol{\Theta}_{st}\right)\right] \\
&= \lim_{r_2 \to 0} \cdots \lim_{r_k \to 0} E\left[\exp\left(\operatorname{tr} \sum_{s=1}^{k} \widetilde{\boldsymbol{W}}_{ss} \boldsymbol{\Theta}_{ss} + \operatorname{tr} \sum_{1 \leqslant t < s \leqslant k} \widetilde{\boldsymbol{Z}}_{st} \boldsymbol{\Theta}_{st}\right)\right].
\end{aligned}$$

We omit technical arguments on uniform convergences, which guarantee the decomposition of $\lim_{(r_2, \ldots, r_k) \to 0}$ in the second line into step-by-step limiting operations $\lim_{r_2 \to 0} \cdots \lim_{r_k \to 0}$ in the third line.

Consider the partitions;

$$\widetilde{\boldsymbol{G}} = \begin{pmatrix} & & & \widetilde{\boldsymbol{G}}_{1k} \\ & \widetilde{\boldsymbol{G}}^{(k-1)} & & \vdots \\ & & & \widetilde{\boldsymbol{G}}_{k-1k} \\ \widetilde{\boldsymbol{G}}_{k1} & \cdots & \widetilde{\boldsymbol{G}}_{kk-1} & \widetilde{\boldsymbol{G}}_{kk} \end{pmatrix},$$

where

$$\widetilde{\boldsymbol{G}}^{(k-1)} = \begin{pmatrix} \widetilde{\boldsymbol{G}}_{11} & \cdots & \widetilde{\boldsymbol{G}}_{1k-1} \\ \vdots & \ddots & \vdots \\ \widetilde{\boldsymbol{G}}_{k-11} & \cdots & \widetilde{\boldsymbol{G}}_{k-1k-1} \end{pmatrix}.$$

Define $\boldsymbol{D}^*$, $\boldsymbol{\Xi}^*$ as partitioned matrices;

$$\boldsymbol{D}^* = \begin{pmatrix} \boldsymbol{L}^{(k-1)} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{D}_k \end{pmatrix}, \quad \boldsymbol{\Xi}^* = \begin{pmatrix} \boldsymbol{\Lambda}^{(k-1)} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Xi}_k \end{pmatrix},$$

where

$$\boldsymbol{L}^{(k-1)} = \mathrm{diag}(l_1, \ldots, l_{m_{k-1}}), \quad \boldsymbol{\Lambda}^{(k-1)} = \mathrm{diag}(\lambda_1, \ldots, \lambda_{m_{k-1}}).$$

Let $\alpha = 1,\ \beta = \alpha_k = \prod_{t=1}^{k} r_t$. Then

$$\boldsymbol{L} = \begin{pmatrix} \boldsymbol{L}^{(k-1)}\alpha & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{D}_k\beta \end{pmatrix}, \quad \boldsymbol{\Lambda} = \begin{pmatrix} \boldsymbol{\Lambda}^{(k-1)}\alpha & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Xi}_k\beta \end{pmatrix}.$$

Since as $r_k \to 0$, $\beta/\alpha \to 0$, from Theorem 2, we have

$$\begin{aligned} &\boldsymbol{S}^{(k-1)} = \widetilde{\boldsymbol{G}}^{(k-1)}\boldsymbol{L}^{(k-1)}\widetilde{\boldsymbol{G}}^{(k-1)\prime} \xrightarrow{d} \boldsymbol{W}_{m_{k-1}}(n, \boldsymbol{\Lambda}^{(k-1)}), \\ &\widetilde{\boldsymbol{W}}_{kk} \xrightarrow{d} \boldsymbol{W}_{\bar{m}_k}(n - m_{k-1}, \boldsymbol{\Xi}_k), \\ &\widetilde{\boldsymbol{Z}}_{kt} \xrightarrow{d} \boldsymbol{N}_{\bar{m}_k \times \bar{m}_t}(\boldsymbol{0}, \boldsymbol{I}_{\bar{m}_k} \otimes \boldsymbol{I}_{\bar{m}_t}), \quad 1 \leqslant t \leqslant k-1, \end{aligned}$$

and the asymptotic distributions are mutually independent. Therefore

$$\begin{aligned} &\lim_{r_k \to 0} E\left[\exp\left(\mathrm{tr}\sum_{s=1}^{k}\widetilde{\boldsymbol{W}}_{ss}\boldsymbol{\Theta}_{ss} + \mathrm{tr}\sum_{1\leqslant t<s\leqslant k}\widetilde{\boldsymbol{Z}}_{st}\boldsymbol{\Theta}_{st}\right)\right] \\ &\quad = E\left[\exp\left(\mathrm{tr}\sum_{s=1}^{k-1}\widetilde{\boldsymbol{W}}_{ss}(\boldsymbol{S}^{(k-1)})\boldsymbol{\Theta}_{ss} + \mathrm{tr}\sum_{1\leqslant t<s\leqslant k-1}\widetilde{\boldsymbol{Z}}_{st}(\boldsymbol{S}^{(k-1)})\boldsymbol{\Theta}_{st}\right)\right] \\ &\qquad \times E[\mathrm{etr}\,\widetilde{\boldsymbol{W}}_{kk}\boldsymbol{\Theta}_{kk}] \times \prod_{t=1}^{k-1} E[\mathrm{etr}\,\widetilde{\boldsymbol{Z}}_{kt}\boldsymbol{\Theta}_{kt}], \end{aligned}$$

where the expectations on the right-hand side is taken with respect to the above asymptotic distributions. If we apply Theorem 2 again to $\boldsymbol{S}^{(k-1)}$ and recursively to the upper-left block Wishart distribution which asymptotically arises, we gain the result. □

Note that Theorem 3 reduces to Theorem 2 of Takemura and Sheena [14] for the extreme case of 1-element blocks $\bar{m}_s = 1, s = 1, \ldots, p$. Therefore Theorem 3 is a generalization of Theorem 2 of Takemura and Sheena [14].

3. Application to estimation of $\boldsymbol{\Sigma}$

3.1. Loss functions and orthogonally equivariant estimators

In this section, we apply the asymptotic result on the distribution of $\boldsymbol{S}$ to the estimation of $\boldsymbol{\Sigma}$ when β/α vanishes. We take a decision-theoretic approach to evaluate the performance of the estimators. We deal with the two loss functions; one is Stein's loss (entropy loss) function

$$L_1(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}) = \operatorname{tr}(\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}) - \log|\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}| - p, \tag{22}$$

and the other is a scale-invariant quadratic loss function

$$L_2(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}) = \operatorname{tr}(\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1} - \boldsymbol{I}_p)^2. \tag{23}$$

The associated risk functions are denoted as

$$R_d(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}) = E[L_d(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma})], \quad d = 1, 2.$$

The classical estimator of $\boldsymbol{\Sigma}$ is the unbiased estimator

$$\widehat{\boldsymbol{\Sigma}}^{\mathrm{U}} = n^{-1}\boldsymbol{S},$$

which has been widely used for many statistical analysis, especially with statistical software packages. However, as James and Stein [2] showed, this estimator is neither minimax nor admissible with Stein's loss function (22). The same drawback with respect to the quadratic loss function (23) was reported by Olkin and Selliah [6]. Following these initiative papers, much literature has been written seeking for a superior estimator to $\widehat{\boldsymbol{\Sigma}}^{\mathrm{U}}$. See Pal [7] for the review on the estimation of $\boldsymbol{\Sigma}$. In this paper we only refer to orthogonally equivariant estimators proposed by Stein [12], Dey and Srinivasan [1] and Krishnamoorthy and Gupta [3]. An estimator of the form

$$\widehat{\boldsymbol{\Sigma}} = \boldsymbol{G}\boldsymbol{\Psi}(\boldsymbol{L})\boldsymbol{G}', \quad \boldsymbol{\Psi}(\boldsymbol{L}) = \operatorname{diag}(\psi_1(\boldsymbol{l}), \ldots, \psi_p(\boldsymbol{l}))$$

is called *orthogonally equivariant*; i.e., $\widehat{\boldsymbol{\Sigma}}(\boldsymbol{GSG}') = \boldsymbol{G}\widehat{\boldsymbol{\Sigma}}(\boldsymbol{S})\boldsymbol{G}',\ \forall \boldsymbol{G} \in \mathcal{O}(p)$.

Stein [12] and Dey and Srinivasan [1] proposed the orthogonally equivariant estimator, $\widehat{\boldsymbol{\Sigma}}^{\mathrm{SDS}}$, defined by

$$\psi_i(\boldsymbol{l}) = l_i \Delta_i^{\mathrm{JS}}, \quad 1 \leqslant i \leqslant p,$$

where $\Delta_i^{\mathrm{JS}} = (n + p + 1 - 2i)^{-1}$. $\widehat{\boldsymbol{\Sigma}}^{\mathrm{SDS}}$ is of simple form but dominates $\widehat{\boldsymbol{\Sigma}}^{\mathrm{U}}$ with substantially better risk w.r.t. the loss function (22). It is also a minimax estimator. See Dey and Srinivasan [1] and Sugiura and Ishibayashi [13] for more details. Order preservation among $\psi_i(\boldsymbol{l}),\ i = 1, \ldots, p$, is discussed in Sheena and Takemura [9].

The orthogonally equivariant estimator $\widehat{\boldsymbol{\Sigma}}^{\mathrm{KG}}$ is defined by

$$\psi_i(\boldsymbol{l}) = l_i \Delta_i^{\mathrm{OS}}, \quad 1 \leqslant i \leqslant p,$$

where Δ_i^{OS} is given by

$$(\Delta_1^{\mathrm{OS}}, \ldots, \Delta_p^{\mathrm{OS}})' = \boldsymbol{A}^{-1}\boldsymbol{b}$$

with a $p \times p$ matrix $\boldsymbol{A} = (a_{ij})$ and a $p \times 1$ vector $\boldsymbol{b} = (b_i)$ defined by

$$a_{ij} = \begin{cases} (n+p-2i+1)(n+p-2i+3) & \text{if } i = j, \\ (n+p-2i+1) & \text{if } i > j, \\ (n+p-2j+1) & \text{if } j > i, \end{cases}$$

$$b_i = n+p+1-2i, \quad i = 1, \ldots, p.$$

$\widehat{\boldsymbol{\Sigma}}^{\mathrm{KG}}$ is conjectured to be a minimax estimator which dominates $\widehat{\boldsymbol{\Sigma}}^{\mathrm{U}}$ w.r.t. the loss function (23). This was proved by Sheena [8] for the case $p = 2$.

In this section we only consider orthogonally equivariant estimators given by

$$\psi_i(\boldsymbol{l}) = c_i l_i, \quad 1 \leqslant i \leqslant p \tag{24}$$

with some constant c_i $(1 \leqslant i \leqslant p)$, or in the matrix expression,

$$\boldsymbol{\Psi}(\boldsymbol{L}) = \boldsymbol{L}^{1/2}\boldsymbol{C}\boldsymbol{L}^{1/2}, \quad \boldsymbol{C} = \mathrm{diag}(c_1, \ldots, c_p).$$

It is interesting that $\widehat{\boldsymbol{\Sigma}}^{\mathrm{SDS}}$ and $\widehat{\boldsymbol{\Sigma}}^{\mathrm{KG}}$ are also the minimum risk estimators among the estimators of the form (24), respectively, for $L_1(\cdot,\cdot)$ and $L_2(\cdot,\cdot)$ when all the population eigenvalues are dispersed. See Takemura and Sheena [14] for more details.

3.2. Asymptotic risk

This subsection is devoted to the calculation of the asymptotic risks $\widetilde{R}_d(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma})$

$$\widetilde{R}_d(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}) = \lim_{\beta/\alpha \to 0} R_d(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}), \quad d = 1, 2,$$

for an orthogonally equivariant estimator defined by (24). Note that

$$\begin{aligned} R_1(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}) &= E[\mathrm{tr}\, \boldsymbol{G}\boldsymbol{L}^{1/2}\boldsymbol{C}\boldsymbol{L}^{1/2}\boldsymbol{G}'\boldsymbol{\Gamma}\boldsymbol{\Lambda}^{-1}\boldsymbol{\Gamma}'] - \log|\boldsymbol{C}| - E[\log|\boldsymbol{\Sigma}^{-1/2}\boldsymbol{S}\boldsymbol{\Sigma}^{-1/2}|] - p \\ &= E[\mathrm{tr}\, \boldsymbol{G}\boldsymbol{L}^{1/2}\boldsymbol{C}\boldsymbol{L}^{1/2}\boldsymbol{G}'\boldsymbol{\Gamma}\boldsymbol{\Lambda}^{-1}\boldsymbol{\Gamma}'] - \sum_{i=1}^{p} \log c_i - \sum_{i=1}^{p} E[\log \chi^2_{n-i+1}] - p. \end{aligned} \tag{25}$$

$$\begin{aligned} R_2(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}) &= E[\mathrm{tr}\, (\boldsymbol{G}\boldsymbol{L}^{1/2}\boldsymbol{C}\boldsymbol{L}^{1/2}\boldsymbol{G}'\boldsymbol{\Gamma}\boldsymbol{\Lambda}^{-1}\boldsymbol{\Gamma}' - \boldsymbol{I}_p)^2] \\ &= E[\mathrm{tr}\, (\boldsymbol{G}\boldsymbol{L}^{1/2}\boldsymbol{C}\boldsymbol{L}^{1/2}\boldsymbol{G}'\boldsymbol{\Gamma}\boldsymbol{\Lambda}^{-1}\boldsymbol{\Gamma}')^2] - 2E[\mathrm{tr}\, \boldsymbol{G}\boldsymbol{L}^{1/2}\boldsymbol{C}\boldsymbol{L}^{1/2}\boldsymbol{G}'\boldsymbol{\Gamma}\boldsymbol{\Lambda}^{-1}\boldsymbol{\Gamma}'] + p. \end{aligned} \tag{26}$$

For the evaluation $E[\log|\boldsymbol{\Sigma}^{-1/2}\boldsymbol{S}\boldsymbol{\Sigma}^{-1/2}|]$, see e.g., Muirhead [5, (10), p. 132].

We start with the following lemma, the proof of which is omitted. (For the proof see Sheena and Takemura [11].)

Lemma 2.

$$\begin{aligned} &\lim_{\beta/\alpha \to 0} E[\mathrm{tr}\, \boldsymbol{G}\boldsymbol{L}^{1/2}\boldsymbol{C}\boldsymbol{L}^{1/2}\boldsymbol{G}'\boldsymbol{\Gamma}\boldsymbol{\Lambda}^{-1}\boldsymbol{\Gamma}'] \\ &\quad = E[\mathrm{tr}\, \boldsymbol{G}_{11}\boldsymbol{D}_1^{1/2}\boldsymbol{C}_1\boldsymbol{D}_1^{1/2}\boldsymbol{G}_{11}'\boldsymbol{\Xi}_1^{-1}] + E[\mathrm{tr}\, \boldsymbol{G}_{22}\boldsymbol{D}_2^{1/2}\boldsymbol{C}_2\boldsymbol{D}_2^{1/2}\boldsymbol{G}_{22}'\boldsymbol{\Xi}_2^{-1}] \\ &\qquad + (p-m)\mathrm{tr}\, \boldsymbol{C}_1, \end{aligned} \tag{27}$$

$$\lim_{\beta/\alpha\to 0} E[\mathrm{tr}\,(\boldsymbol{G}\boldsymbol{L}^{1/2}\boldsymbol{C}\boldsymbol{L}^{1/2}\boldsymbol{G}'\boldsymbol{\Gamma}\boldsymbol{\Lambda}^{-1}\boldsymbol{\Gamma}')^2]$$
$$= E[\mathrm{tr}\,(\boldsymbol{G}_{11}\boldsymbol{D}_1^{1/2}\boldsymbol{C}_1\boldsymbol{D}_1^{1/2}\boldsymbol{G}'_{11}\boldsymbol{\Xi}_1^{-1})^2] + E[\mathrm{tr}\,(\boldsymbol{G}_{22}\boldsymbol{D}_2^{1/2}\boldsymbol{C}_2\boldsymbol{D}_2^{1/2}\boldsymbol{G}'_{22}\boldsymbol{\Xi}_2^{-1})^2]$$
$$+2(p-m)E[\mathrm{tr}\,\boldsymbol{C}_1^2\boldsymbol{D}_1^{1/2}\boldsymbol{G}'_{11}\boldsymbol{\Xi}_1^{-1}\boldsymbol{G}_{11}\boldsymbol{D}_1^{1/2}] + 2\mathrm{tr}\,\boldsymbol{C}_1 E[\mathrm{tr}\,\boldsymbol{\Xi}_2^{-1}\boldsymbol{G}_{22}\boldsymbol{D}_2^{1/2}\boldsymbol{C}_2\boldsymbol{D}_2^{1/2}\boldsymbol{G}'_{22}]$$
$$+(p-m)(p-m+2)\sum_{i=1}^m c_i^2 + 2(p-m)\sum_{1\leqslant i<s\leqslant m} c_i c_s, \tag{28}$$

where the expectations on the right-hand side in (27) *and* (28) *are taken with respect to the distributions in* (20) *and the decompositions in* (21).

Now suppose that under the distribution of $\boldsymbol{W}_{ss}$, $s = 1, 2$, in (20) and their spectral decomposition in (21), we estimate $\boldsymbol{\Xi}_s$, $s = 1, 2$, by the following orthogonally equivariant estimators

$$\widehat{\boldsymbol{\Xi}}_1 = \boldsymbol{G}_{11}\boldsymbol{D}_1^{1/2}\boldsymbol{C}_1\boldsymbol{D}_1^{1/2}\boldsymbol{G}'_{11}, \qquad \boldsymbol{C}_1 = \mathrm{diag}(c_1, \ldots, c_m),$$
$$\widehat{\boldsymbol{\Xi}}_2 = \boldsymbol{G}_{22}\boldsymbol{D}_2^{1/2}\boldsymbol{C}_2\boldsymbol{D}_2^{1/2}\boldsymbol{G}'_{22}, \qquad \boldsymbol{C}_2 = \mathrm{diag}(c_{m+1}, \ldots, c_p),$$

then the risks w.r.t. each loss function (22), (23) are given by

$$R_{11}(\widehat{\boldsymbol{\Xi}}_1, \boldsymbol{\Xi}_1) = E[\mathrm{tr}\,(\widehat{\boldsymbol{\Xi}}_1\boldsymbol{\Xi}_1^{-1}) - \log|\widehat{\boldsymbol{\Xi}}_1\boldsymbol{\Xi}_1^{-1}| - m],$$
$$R_{21}(\widehat{\boldsymbol{\Xi}}_2, \boldsymbol{\Xi}_2) = E[\mathrm{tr}\,(\widehat{\boldsymbol{\Xi}}_2\boldsymbol{\Xi}_2^{-1}) - \log|\widehat{\boldsymbol{\Xi}}_2\boldsymbol{\Xi}_2^{-1}| - p + m],$$
$$R_{12}(\widehat{\boldsymbol{\Xi}}_1, \boldsymbol{\Xi}_1) = E[\mathrm{tr}\,(\widehat{\boldsymbol{\Xi}}_1\boldsymbol{\Xi}_1^{-1} - \boldsymbol{I}_m)^2],$$
$$R_{22}(\widehat{\boldsymbol{\Xi}}_2, \boldsymbol{\Xi}_2) = E[\mathrm{tr}\,(\widehat{\boldsymbol{\Xi}}_2\boldsymbol{\Xi}_2^{-1} - \boldsymbol{I}_{p-m})^2].$$

The following theorem gives the decomposition of the asymptotic risk, $\widetilde{R}_d(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma})$, into the risks R_{1d}, R_{2d} and the residuals R_{3d} for $d = 1, 2$.

Theorem 4. *For* $d = 1, 2$,

$$\widetilde{R}_d(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}) = R_{1d}(\widehat{\boldsymbol{\Xi}}_1, \boldsymbol{\Xi}_1) + R_{2d}(\widehat{\boldsymbol{\Xi}}_2, \boldsymbol{\Xi}_2) + R_{3d},$$

where

$$R_{31} = (p-m)\sum_{i=1}^m c_i,$$

and

$$R_{32} = 2(p-m)E[\mathrm{tr}\,\boldsymbol{C}_1^2\boldsymbol{D}_1^{1/2}\boldsymbol{G}'_{11}\boldsymbol{\Xi}_1^{-1}\boldsymbol{G}_{11}\boldsymbol{D}_1^{1/2}] + 2\mathrm{tr}\,\boldsymbol{C}_1 E[\mathrm{tr}\,\boldsymbol{\Xi}_2^{-1}\boldsymbol{G}_{22}\boldsymbol{D}_2^{1/2}\boldsymbol{C}_2\boldsymbol{D}_2^{1/2}\boldsymbol{G}'_{22}]$$
$$+ (p-m)(p-m+2)\sum_{i=1}^m c_i^2 + 2(p-m)\sum_{1\leqslant i<s\leqslant m} c_i c_s - 2(p-m)\sum_{i=1}^m c_i.$$

All the expectations are taken with respect to the distributions (20) *and the decompositions* (21).

Proof. From (25),

$$R_{11}(\widehat{\boldsymbol{\Sigma}}_1, \boldsymbol{\Sigma}_1) = E[\operatorname{tr} \boldsymbol{G}_{11}\boldsymbol{D}_1^{1/2}\boldsymbol{C}_1\boldsymbol{D}_1^{1/2}\boldsymbol{G}_{11}'\boldsymbol{\Xi}_1^{-1}] - \sum_{i=1}^{m}\log c_i - \sum_{i=1}^{m} E[\log\chi^2_{n-i+1}] - m,$$

$$R_{21}(\widehat{\boldsymbol{\Sigma}}_2, \boldsymbol{\Sigma}_2) = E[\operatorname{tr} \boldsymbol{G}_{22}\boldsymbol{D}_2^{1/2}\boldsymbol{C}_2\boldsymbol{D}_2^{1/2}\boldsymbol{G}_{22}'\boldsymbol{\Xi}_2^{-1}] - \sum_{i=m+1}^{p}\log c_i - \sum_{i=m+1}^{p} E[\log\chi^2_{n-i+1}] - p + m.$$

Using (27) together with the above result, we have the result for $\widetilde{R}_1(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma})$. From (26),

$$R_{12}(\widehat{\boldsymbol{\Sigma}}_1, \boldsymbol{\Sigma}_1) = E[\operatorname{tr}(\boldsymbol{G}_{11}\boldsymbol{D}_1^{1/2}\boldsymbol{C}_1\boldsymbol{D}_1^{1/2}\boldsymbol{G}_{11}'\boldsymbol{\Xi}_1^{-1})^2] - 2E[\operatorname{tr} \boldsymbol{G}_{11}\boldsymbol{D}_1^{1/2}\boldsymbol{C}_1\boldsymbol{D}_1^{1/2}\boldsymbol{G}_{11}'\boldsymbol{\Xi}_1^{-1}] + m,$$

$$R_{22}(\widehat{\boldsymbol{\Sigma}}_2, \boldsymbol{\Sigma}_2) = E[\operatorname{tr}(\boldsymbol{G}_{22}\boldsymbol{D}_2^{1/2}\boldsymbol{C}_2\boldsymbol{D}_2^{1/2}\boldsymbol{G}_{22}'\boldsymbol{\Xi}_2^{-1})^2] - 2E[\operatorname{tr} \boldsymbol{G}_{22}\boldsymbol{D}_2^{1/2}\boldsymbol{C}_2\boldsymbol{D}_2^{1/2}\boldsymbol{G}_{22}'\boldsymbol{\Xi}_2^{-1}] + p - m.$$

Using (28) and (27) together with the above result, we have the result for $\widetilde{R}_2(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma})$. □

3.3. Minimum asymptotic risk estimator

Consider the model (1) and suppose $\tau_1 = \cdots = \tau_m (= \tau)$ in (2). Then $\alpha = \tau + \sigma^2$ and $\beta = \sigma^2$ and

$$\boldsymbol{\Xi}_1 = \boldsymbol{I}_m, \quad \boldsymbol{\Xi}_2 = \boldsymbol{I}_{p-m}. \tag{29}$$

This assumption may not be very realistic. However, note that it is trivially satisfied in the one-factor model $m = 1$, which is frequently used in practice. In this subsection we focus on the estimation of $\boldsymbol{\Sigma}$ under the condition (29). In this case, since we have no unknown parameters anymore, the asymptotic risk is uniquely determined, hence we can derive the "best" i.e., minimum asymptotic risk estimator among the orthogonally equivariant estimators of the form (24). The following theorem gives the asymptotic risk for the case (29).

Theorem 5. *If* $\boldsymbol{\Xi}_1 = \boldsymbol{I}_m$, $\boldsymbol{\Xi}_2 = \boldsymbol{I}_{p-m}$, *then the asymptotic risk* $\widetilde{R}_d(\hat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma})$, $d = 1, 2$, *is given by the following function of* $\boldsymbol{c} = (c_1, \ldots, c_p)'$:

$$\widetilde{R}_1(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}) = \sum_{i=1}^{p}(b_i c_i - \log c_i) - \sum_{i=1}^{p} E[\log\chi^2_{n-i+1}] - p, \tag{30}$$

$$\widetilde{R}_2(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}) = \boldsymbol{c}'\boldsymbol{A}\boldsymbol{c} - 2\boldsymbol{b}'\boldsymbol{c} + p, \tag{31}$$

where $\boldsymbol{b} = (b_1, \ldots, b_p)'$ *is given by*

$$b_i = \begin{cases} E[d_i] + p - m & \text{if } 1 \leqslant i \leqslant m, \\ E[d_i] & \text{if } m+1 \leqslant i \leqslant p, \end{cases}$$

and $p \times p$ *symmetric matrix* $\boldsymbol{A} = (a_{ij})$ *is given by*

$$a_{ij} = \begin{cases} E[d_i^2 + 2(p-m)d_i] + (p-m)(p-m+2) & \text{if } 1 \leqslant i = j \leqslant m, \\ E[d_i^2] & \text{if } m+1 \leqslant i = j \leqslant p, \\ p-m & \text{if } 1 \leqslant i \neq j \leqslant m, \\ E[d_j] & \text{if } 1 \leqslant i \leqslant m < j \leqslant p, \\ E[d_i] & \text{if } 1 \leqslant j \leqslant m < i \leqslant p, \\ 0 & \text{otherwise}. \end{cases}$$

All the expectations are taken with respect to the distribution (20) *and the decompositions* (21) *with* $\boldsymbol{\Xi}_1 = \boldsymbol{I}_m,\ \boldsymbol{\Xi}_2 = \boldsymbol{I}_{p-m}$.

Proof. Evaluating $R_{jd}(\widehat{\boldsymbol{\Sigma}}_j, \boldsymbol{\Sigma}_j),\ 1 \leqslant j, d \leqslant 2$ in Theorem 4 when $\boldsymbol{\Xi}_1 = \boldsymbol{I}_m,\ \boldsymbol{\Xi}_2 = \boldsymbol{I}_{p-m}$, we have the following results:

$$\begin{aligned} R_{11}(\widehat{\boldsymbol{\Xi}}_1, \boldsymbol{\Xi}_1) &= E[L_1(\widehat{\boldsymbol{\Xi}}_1, \boldsymbol{I}_m)] = E[\operatorname{tr} \widehat{\boldsymbol{\Xi}}_1 - \log|\widehat{\boldsymbol{\Xi}}_1| - m] \\ &= E\left[\sum_{i=1}^m d_i c_i - \log|\boldsymbol{W}_{11}|\right] - \sum_{i=1}^m \log c_i - m \\ &= \sum_{i=1}^m E[d_i]c_i - E[\log|\boldsymbol{W}_{11}|] - \sum_{i=1}^m \log c_i - m, \end{aligned}$$

$$R_{21}(\widehat{\boldsymbol{\Xi}}_2, \boldsymbol{\Xi}_2) = \sum_{i=m+1}^p E[d_i]c_i - E[\log|\boldsymbol{W}_{22}|] - \sum_{i=m+1}^p \log c_i - p + m,$$

$$\begin{aligned} R_{12}(\widehat{\boldsymbol{\Xi}}_1, \boldsymbol{\Xi}_1) &= E[L_2(\widehat{\boldsymbol{\Xi}}_1, \boldsymbol{I}_m)] = E[\operatorname{tr}(\widehat{\boldsymbol{\Xi}}_1 - \boldsymbol{I}_m)^2] \\ &= E[\operatorname{tr} \widehat{\boldsymbol{\Xi}}_1^2 - 2\operatorname{tr} \widehat{\boldsymbol{\Xi}}_1] + m = E\left[\sum_{i=1}^m d_i^2 c_i^2 - 2\sum_{i=1}^m d_i c_i\right] + m \\ &= \sum_{i=1}^m E[d_i^2]c_i^2 - 2\sum_{i=1}^m E[d_i]c_i + m, \end{aligned}$$

$$R_{22}(\widehat{\boldsymbol{\Xi}}_2, \boldsymbol{\Xi}_2) = \sum_{i=m+1}^p E[d_i^2]c_i^2 - 2\sum_{i=m+1}^p E[d_i]c_i + p - m.$$

Next we calculate R_{32} in Theorem 4 when $\boldsymbol{\Xi}_1 = \boldsymbol{I}_m$, $\boldsymbol{\Xi}_2 = \boldsymbol{I}_{p-m}$. Note that

$$\begin{aligned} 2(p-m)E[\operatorname{tr} \boldsymbol{C}_1^2 \boldsymbol{D}_1^{1/2} \boldsymbol{G}_{11}' \boldsymbol{\Xi}_1^{-1} \boldsymbol{G}_{11} \boldsymbol{D}_1^{1/2}] &= 2(p-m)E[\operatorname{tr} \boldsymbol{C}_1^2 \boldsymbol{D}_1] \\ &= 2(p-m)\sum_{i=1}^m E[d_i]c_i^2, \end{aligned}$$

$$2\operatorname{tr} \boldsymbol{C}_1 E[\operatorname{tr} \boldsymbol{\Xi}_2^{-1} \boldsymbol{G}_{22} \boldsymbol{D}_2^{1/2} \boldsymbol{C}_2 \boldsymbol{D}_2^{1/2} \boldsymbol{G}_{22}'] = 2\left(\sum_{i=1}^m c_i\right)\left(\sum_{i=m+1}^p E[d_i]c_i\right).$$

Therefore,

$$R_{32} = \sum_{i=1}^{m} c_i^2 \{(p-m)(p-m+2) + 2(p-m)E[d_i]\}$$
$$+ 2(p-m) \sum_{1 \leqslant i < s \leqslant m} c_i c_s + 2 \sum_{1 \leqslant i \leqslant m < s \leqslant p} c_i c_s E[d_s] - 2(p-m) \sum_{i=1}^{m} c_i .$$

Combining above results, we see that (30) and (31) hold. □

Corollary 1. *The minimum asymptotic risk with respect to the loss function $L_1(\cdot,\cdot)$ is given by*

$$\sum_{i=1}^{p} \log b_i - \sum_{i=1}^{p} E[\log \chi^2_{n-i-1}].$$

It is attained by $\widehat{\boldsymbol{\Sigma}}^{\mathrm{MA}_1}$ given by $c_i = b_i^{-1}$, $i = 1, \ldots, p$. The minimum asymptotic risk with respect to the loss function $L_2(\cdot,\cdot)$ is given by

$$p - \boldsymbol{b}'\boldsymbol{A}^{-1}\boldsymbol{b}.$$

It is attained by $\widehat{\boldsymbol{\Sigma}}^{\mathrm{MA}_2}$ given by $\boldsymbol{c} = \boldsymbol{A}^{-1}\boldsymbol{b}$.

Proof. The results are easily obtained by the minimization $\sum_{i=1}^{p}(b_i c_i - \log c_i)$ or $\boldsymbol{c}'\boldsymbol{A}\boldsymbol{c} - 2\boldsymbol{b}'\boldsymbol{c}$. □

The calculation of the asymptotic risks in Theorem 5 and the c_i's of $\widehat{\boldsymbol{\Sigma}}^{\mathrm{MA}_1}$ and $\widehat{\boldsymbol{\Sigma}}^{\mathrm{MA}_2}$ requires the evaluation of $E[d_i]$, $E[d_i^2]$, $i = 1, \ldots, p$, that is, the first and the second moment of the eigenvalues of the Wishart distribution with the identity covariance matrix. Generally, we need to make use of Monte Carlo simulation or numerical integration for the evaluation of the moments of the eigenvalues. However, when p is small and n is appropriately even or odd depending on p, the analytic evaluation is feasible. See Appendix in Sheena and Takemura [11] for this evaluation.

Tables 1–3 give c_i's for $\widehat{\boldsymbol{\Sigma}}^{\mathrm{U}}, \widehat{\boldsymbol{\Sigma}}^{\mathrm{SDS}}, \widehat{\boldsymbol{\Sigma}}^{\mathrm{KG}}, \widehat{\boldsymbol{\Sigma}}^{\mathrm{MA}_1}, \widehat{\boldsymbol{\Sigma}}^{\mathrm{MA}_2}$ when $p = 4$ with several values of n. (Tables for $p = 3$ are given in Sheena and Takemura [11].) The value of c_i's for the minimum asymptotic risk estimators $\widehat{\boldsymbol{\Sigma}}^{\mathrm{MA}_1}, \widehat{\boldsymbol{\Sigma}}^{\mathrm{MA}_2}$ is calculated by the aforementioned analytic method. As it is well known, $n^{-1}l_i$ $(i = 1, \ldots, p)$ tends to overestimate the corresponding eigenvalue of $\boldsymbol{\Sigma}$ when i is small, while it tends to underestimate the corresponding eigenvalue of $\boldsymbol{\Sigma}$ when i is large. The estimators $\widehat{\boldsymbol{\Sigma}}^{\mathrm{SDS}}, \widehat{\boldsymbol{\Sigma}}^{\mathrm{KG}}$ modify this tendency by increasing weight $c_1 < \cdots < c_p$. It is seen from the tables that $\widehat{\boldsymbol{\Sigma}}^{\mathrm{MA}_1}, \widehat{\boldsymbol{\Sigma}}^{\mathrm{MA}_2}$ enlarge the weight difference within each block in most cases; for example when $p = 4, m = 2$, the relation between c_i's of $\widehat{\boldsymbol{\Sigma}}^{\mathrm{SDS}}$ $(\widehat{\boldsymbol{\Sigma}}^{\mathrm{KG}})$ (say $c_i^{\mathrm{SDS}}(c_i^{\mathrm{KG}})$, $i = 1, \ldots, 4$) and those of $\widehat{\boldsymbol{\Sigma}}^{\mathrm{MA}_1}$ $(\widehat{\boldsymbol{\Sigma}}^{\mathrm{MA}_2})$ (say $c_i^{\mathrm{MA}_1}(c_i^{\mathrm{MA}_2})$, $i = 1, \ldots, 4$) is found as

$$c_1^{\mathrm{MA}_1} < c_1^{\mathrm{SDS}} < c_2^{\mathrm{SDS}} < c_2^{\mathrm{MA}_1}, \quad c_3^{\mathrm{MA}_1} < c_3^{\mathrm{SDS}} < c_4^{\mathrm{SDS}} < c_4^{\mathrm{MA}_1},$$

and

$$c_1^{\mathrm{MA}_2} < c_1^{\mathrm{KG}} < c_2^{\mathrm{KG}} < c_2^{\mathrm{MA}_2}, \quad c_3^{\mathrm{MA}_2} < c_3^{\mathrm{KG}} < c_4^{\mathrm{KG}} < c_4^{\mathrm{MA}_2}.$$

Table 1
$p = 4, m = 1$

	$\widehat{\Sigma}^{U}$	$\widehat{\Sigma}^{SDS}$	$\widehat{\Sigma}^{KG}$	$\widehat{\Sigma}^{MA_1}$	$\widehat{\Sigma}^{MA_2}$
$n = 5$					
c_1	0.2000	0.1250	0.0822	0.1250	0.0759
c_2	0.2000	0.1667	0.0973	0.1200	0.0927
c_3	0.2000	0.2500	0.1222	0.3333	0.2310
c_4	0.2000	0.5000	0.1746	1.5000	0.6931
Asy.Risk1	3.0752	2.0603		1.5303	
R.R.R.		33.00		50.24	
Asy.Risk2	4.0000		1.8435		1.4655
R.R.R.			53.91		63.36
$n = 7$					
c_1	0.1429	0.1000	0.0690	0.1000	0.0647
c_2	0.1429	0.1250	0.0796	0.0883	0.0726
c_3	0.1429	0.1667	0.0959	0.2000	0.1559
c_4	0.1429	0.2500	0.1259	0.5956	0.3816
Asy.Risk1	1.8508	1.2955		0.9241	
R.R.R.		30.01		50.07	
Asy.Risk2	2.8571		1.4923		1.1116
R.R.R.			47.77		61.10
$n = 9$					
c_1	0.1111	0.0833	0.0600	0.0833	0.0571
c_2	0.1111	0.1000	0.0681	0.0707	0.0602
c_3	0.1111	0.1250	0.0798	0.1429	0.1179
c_4	0.1111	0.1667	0.0990	0.3497	0.2553
Asy.Risk1	1.3436	0.9790		0.6852	
R.R.R.		27.13		49.00	
Asy.Risk2	2.2222		1.2591		0.9083
R.R.R.			43.34		59.13
$n = 11$					
c_1	0.0909	0.0714	0.0533	0.0714	0.0513
c_2	0.0909	0.0833	0.0596	0.0593	0.0517
c_3	0.0909	0.1000	0.0685	0.1111	0.0949
c_4	0.0909	0.1250	0.0819	0.2413	0.1890
Asy.Risk1	1.0585	0.7956		0.5496	
R.R.R.		24.84		48.08	
Asy.Risk2	1.8182		1.0927		0.7730
R.R.R.			39.90		57.49
$n = 21$					
c_1	0.0476	0.0417	0.0346	0.0417	0.0341
c_2	0.0476	0.0455	0.0372	0.0338	0.0311
c_3	0.0476	0.0500	0.0404	0.0526	0.0483
c_4	0.0476	0.0556	0.0444	0.0879	0.0782
Asy.Risk1	0.5127	0.4183		0.2769	
R.R.R.		18.41		45.99	
Asy.Risk2	0.9524		0.6708		0.4526
R.R.R.			29.57		52.47

Table 1 *contd.*

	$\widehat{\Sigma}^{U}$	$\widehat{\Sigma}^{SDS}$	$\widehat{\Sigma}^{KG}$	$\widehat{\Sigma}^{MA_1}$	$\widehat{\Sigma}^{MA_2}$
$n = 51$					
c_1	0.0196	0.0185	0.0170	0.0185	0.0169
c_2	0.0196	0.0192	0.0176	0.0154	0.0148
c_3	0.0196	0.0200	0.0182	0.0204	0.0197
c_4	0.0196	0.0208	0.0189	0.0278	0.0266
Asy.Risk1	0.2016	0.1777		0.1122	
R.R.R.		11.88		44.36	
Asy.Risk2	0.3922		0.3207		0.2055
R.R.R.			18.22		47.59

Table 2
$p = 4, m = 2$

	$\widehat{\Sigma}^{U}$	$\widehat{\Sigma}^{SDS}$	$\widehat{\Sigma}^{KG}$	$\widehat{\Sigma}^{MA_1}$	$\widehat{\Sigma}^{MA_2}$
$n = 5$					
c_1	0.2000	0.1250	0.0822	0.1034	0.0762
c_2	0.2000	0.1667	0.0973	0.2308	0.1261
c_3	0.2000	0.2500	0.1222	0.2000	0.1173
c_4	0.2000	0.5000	0.1746	1.0000	0.3988
Asy.Risk1	3.0752	2.2687		1.9819	
R.R.R.		26.23		35.55	
Asy.Risk2	4.0000		1.8668		1.7317
R.R.R.			53.33		56.71
$n = 7$					
c_1	0.1429	0.1000	0.0690	0.0820	0.0632
c_2	0.1429	0.1250	0.0796	0.1724	0.1055
c_3	0.1429	0.1667	0.0959	0.1304	0.0885
c_4	0.1429	0.2500	0.1259	0.4286	0.2425
Asy.Risk1	1.8508	1.4334		1.2107	
R.R.R.		22.55		34.59	
Asy.Risk2	2.8571		1.5273		1.3728
R.R.R.			46.54		51.95
$n = 9$					
c_1	0.1111	0.0833	0.0600	0.0682	0.0546
c_2	0.1111	0.1000	0.0681	0.1362	0.0910
c_3	0.1111	0.1250	0.0798	0.0980	0.0719
c_4	0.1111	0.1667	0.0990	0.2632	0.1727
Asy.Risk1	1.3436	1.0774		0.8908	
R.R.R.		19.81		33.70	
Asy.Risk2	2.2222		1.2992		1.1422
R.R.R.			41.54		48.60
$n = 11$					
c_1	0.0909	0.0714	0.0533	0.0586	0.0482
c_2	0.0909	0.0833	0.0596	0.1119	0.0798
c_3	0.0909	0.1000	0.0685	0.0790	0.0609
c_4	0.0909	0.1250	0.0819	0.1872	0.1337
Asy.Risk1	1.0585	0.8700		0.7080	
R.R.R.		17.81		33.11	

Table 2 *contd.*

	$\widehat{\Sigma}^{U}$	$\widehat{\Sigma}^{SDS}$	$\widehat{\Sigma}^{KG}$	$\widehat{\Sigma}^{MA_1}$	$\widehat{\Sigma}^{MA_2}$
Asy.Risk2	1.8182		1.1337		0.9792
R.R.R.			37.65		46.14
$n = 21$					
c_1	0.0476	0.0417	0.0346	0.0349	0.0330
c_2	0.0476	0.0455	0.0372	0.0577	0.0531
c_3	0.0476	0.0500	0.0404	0.0410	0.0352
c_4	0.0476	0.0556	0.0444	0.0735	0.0615
Asy.Risk1	0.5127	0.4477		0.3477	
R.R.R.		12.68		32.18	
Asy.Risk2	0.9524		0.7013		0.5722
R.R.R.			26.36		39.92
$n = 51$					
c_1	0.0196	0.0185	0.0170	0.0162	0.0153
c_2	0.0196	0.0192	0.0176	0.0227	0.0211
c_3	0.0196	0.0200	0.0182	0.0173	0.0163
c_4	0.0196	0.0208	0.0189	0.0248	0.0232
Asy.Risk1	0.2016	0.1857		0.1377	
R.R.R.		7.90		31.73	
Asy.Risk2	0.3922		0.3331		0.2544
R.R.R.			15.06		35.13

Table 3
$p = 4, m = 3$

	$\widehat{\Sigma}^{U}$	$\widehat{\Sigma}^{SDS}$	$\widehat{\Sigma}^{KG}$	$\widehat{\Sigma}^{MA_1}$	$\widehat{\Sigma}^{MA_2}$
$n = 4$					
c_1	0.2500	0.1429	0.0919	0.1071	0.0852
c_2	0.2500	0.2000	0.1111	0.2500	0.1670
c_3	0.2500	0.3333	0.1449	0.6000	0.2383
c_4	0.2500	1.0000	0.2174	1.0000	0.1698
Asy.Risk1	4.8592	3.6569		3.4447	
R.R.R.		24.74		29.11	
Asy.Risk2	5.0000		2.0872		1.9697
R.R.R.			58.26		60.61
$n = 6$					
c_1	0.1667	0.1111	0.0749	0.0812	0.0678
c_2	0.1667	0.1429	0.0873	0.1667	0.1248
c_3	0.1667	0.2000	0.1072	0.3733	0.2028
c_4	0.1667	0.3333	0.1461	0.3333	0.1209
Asy.Risk1	2.2985	1.7446		1.5186	
R.R.R.		24.10		33.93	
Asy.Risk2	3.3333		1.6702		1.5097
R.R.R.			49.89		54.71
$n = 8$					
c_1	0.1250	0.0909	0.0642	0.0660	0.0569
c_2	0.1250	0.1111	0.0733	0.1250	0.0999

Table 3 *contd.*

	$\widehat{\Sigma}^{U}$	$\widehat{\Sigma}^{SDS}$	$\widehat{\Sigma}^{KG}$	$\widehat{\Sigma}^{MA_1}$	$\widehat{\Sigma}^{MA_2}$
c_3	0.1250	0.1429	0.0870	0.2591	0.1670
c_4	0.1250	0.2000	0.1108	0.2000	0.0966
Asy.Risk1	1.5538	1.2032		0.9929	
R.R.R.		22.57		36.10	
Asy.Risk2	2.5000		1.3948		1.2111
R.R.R.			44.21		51.56
$n = 10$					
c_1	0.1000	0.0769	0.0565	0.0560	0.0493
c_2	0.1000	0.0909	0.0636	0.1000	0.0833
c_3	0.1000	0.1111	0.0737	0.1944	0.1385
c_4	0.1000	0.1429	0.0896	0.1429	0.0810
Asy.Risk1	1.1828	0.9327		0.7412	
R.R.R.		21.15		37.34	
Asy.Risk2	2.0000		1.1991		1.0067
R.R.R.			40.05		49.66
$n = 20$					
c_1	0.0500	0.0435	0.0358	0.0326	0.0303
c_2	0.0500	0.0476	0.0386	0.0500	0.0455
c_3	0.0500	0.0526	0.0421	0.0808	0.0694
c_4	0.0500	0.0588	0.0465	0.0588	0.0450
Asy.Risk1	0.5385	0.4484		0.3218	
R.R.R.		16.73		40.24	
Asy.Risk2	1.0000		0.7122		0.5395
R.R.R.			28.78		46.05
$n = 50$					
c_1	0.0200	0.0189	0.0172	0.0151	0.0146
c_2	0.0200	0.0196	0.0179	0.0200	0.0192
c_3	0.0200	0.0204	0.0186	0.0271	0.0256
c_4	0.0200	0.0213	0.0193	0.0213	0.0192
Asy.Risk1	0.2069	0.1836		0.1207	
R.R.R.		11.28		41.65	
Asy.Risk2	0.4000		0.3293		0.2236
R.R.R.			17.68		44.11

The tables also give asymptotic risk comparison w.r.t. L_1 among the estimators $\widehat{\Sigma}^{U}, \widehat{\Sigma}^{SDS}, \widehat{\Sigma}^{MA_1}$ (see "Asy.Risk1") and that w.r.t. L_2 among the estimators $\widehat{\Sigma}^{U}, \widehat{\Sigma}^{KG}, \widehat{\Sigma}^{MA_2}$ (see "Asy.Risk2"). The risks are analytically calculated except for evaluating $\sum_{i=1}^{p} E[\log \chi^2_{n-i+1}]$ by Monte Carlo simulation method using 10^5 random numbers. "R.R.R." under "Asy.Risk1" or "Asy.Risk2" shows the risk reduction rate defined by

$$\text{R.R.R. of } \widehat{\Sigma} = \frac{\text{The risk of } \widehat{\Sigma}^{U} - \text{The risk of } \widehat{\Sigma}}{\text{The risk of } \widehat{\Sigma}^{U}} \times 100.$$

It has been observed that $\widehat{\Sigma}^{SDS}$ and $\widehat{\Sigma}^{KG}$ drastically reduce the risk of Σ^{U} when the population eigenvalues are close to each other. Lin and Perlman [4] report that when $\Sigma = I_p$, R.R.R. of

$\widehat{\Sigma}^{\mathrm{SDS}}$ often reaches 70%. See also Sugiura and Ishibayashi [13] for a risk comparison by elabarate simulation. In the situation of the block-wise dispersion, the risk reduction rate of these estimators rarely approaches 50%. Especially when n is as large as 50, the rate is always under 20%. On the other hand, the risk reduction rates of $\widehat{\Sigma}^{\mathrm{MA}_1}$ and Σ^{MA_2} are constantly over 30% and often reach 50% irrespective of the values of n. It is interesting that $\widehat{\Sigma}^{\mathrm{MA}_2}$ always outperforms $\widehat{\Sigma}^{\mathrm{MA}_1}$ in view of R.R.R.

3.4. Simulation studies

In this subsection, we evaluate the performance of $\widehat{\Sigma}^{\mathrm{MA}_1}$, $\widehat{\Sigma}^{\mathrm{MA}_2}$ by Monte Carlo simulation under the situation (29). As we saw in the previous subsection, in view of the asymptotic risks, $\widehat{\Sigma}^{\mathrm{MA}_1}$, $\widehat{\Sigma}^{\mathrm{MA}_2}$ provide better risk reduction compared to $\widehat{\Sigma}^{\mathrm{SDS}}$, $\widehat{\Sigma}^{\mathrm{KG}}$. In practical point of view, however, it is important to see how largely the population eigenvalues must be dispersed so that the use of $\widehat{\Sigma}^{\mathrm{MA}_d}$, $d = 1, 2$, is recommended. The convergence speed of the distributions given in Theorem 2, which is an interesting topic by itself, is closely related to this problem.

To see the convergence speed in both distributions and risks, we carried out Monte Carlo Simulation for the case $p = 4$, $m = 1$. We took 11 values 1.0, 0.8, 0.6, 0.4, 0.2, $10^{-i}(i = 1, \dots, 6)$ in the convergence parameter β, while α is fixed at 1. We took three different values of n and generated 10^6 random Wishart matrices under given n, β. The result is given in Table 4. (The result for $p = 3, m = 1$ is given in Sheena and Takemura [11].) The upper part of each table shows the speed of the distributional convergence in Theorem 2. Note that when $\Xi_1 = \boldsymbol{I}_m$, $\Xi_2 = \boldsymbol{I}_{p-m}$, the asymptotic distribution of a diagonal element of $\widetilde{\boldsymbol{W}}_{ss}$, $s = 1, 2$, is a χ^2 distribution. The labels in the tables are given as follows with $\chi^2_n(\alpha)$, $z(\alpha)$ denoting the lower α percentage points of χ^2 distribution with n degrees of freedom and the standard normal distribution, respectively ;

$$\text{Prob 1a} = P(\widetilde{\boldsymbol{W}}_{11} \leqslant \chi^2_n(0.05)), \quad \text{Prob 1b} = P(\widetilde{\boldsymbol{W}}_{11} \leqslant \chi^2_n(0.95)),$$

$$\text{Prob 2a} = P((\widetilde{\boldsymbol{W}}_{22})_{11} \leqslant \chi^2_{n-1}(0.05)), \quad \text{Prob 2b} = P((\widetilde{\boldsymbol{W}}_{22})_{11} \leqslant \chi^2_{n-1}(0.95)),$$

$$\text{Prob 3a} = P((\widetilde{\boldsymbol{W}}_{22})_{33} \leqslant \chi^2_{n-1}(0.05)), \quad \text{Prob 3b} = P((\widetilde{\boldsymbol{W}}_{22})_{33} \leqslant \chi^2_{n-1}(0.95)),$$

$$\text{Prob 4a} = P((\widetilde{\boldsymbol{Z}}_{21})_{11} \leqslant z(0.05)), \quad \text{Prob 4b} = P((\widetilde{\boldsymbol{Z}}_{21})_{11} \leqslant z(0.95)),$$

$$\text{Prob 5a} = P((\widetilde{\boldsymbol{Z}}_{21})_{31} \leqslant z(0.05)), \quad \text{Prob 5b} = P((\widetilde{\boldsymbol{Z}}_{21})_{31} \leqslant z(0.95)).$$

In the lower part of each table, “Risk 1_*” and “Risk 2_*” show the risks of the corresponding estimator $\widehat{\Sigma}^{*}$, respectively, for L_1 and L_2. The tables show that

1. The convergence of the diagonal elements of $\widetilde{\boldsymbol{W}}_{ss}$, $s = 1, 2$, is so rapid that when $\beta = 0.1$, the asymptotic distribution already gives a good approximation for the exact distribution. When $\beta = 0.1$, every probability of the diagonal elements is within 0.01 deviation from the exact asymptotic probability.
2. The convergence speed of $\widetilde{\boldsymbol{Z}}$ is quite slow compared to that of the diagonal elements of $\widetilde{\boldsymbol{W}}_{ss}$, $s = 1, 2$. For a good approximation as above, β must be as small as 10^{-5} or 10^{-6}.
3. The risks also rapidly converge to the asymptotic risks so that $\beta = 0.1$ is small enough to give a good approximation. Actually all the risks in the tables when $\beta = 0.1$ are within the $\pm 5\%$ interval centered at the exact asymptotic risk.
4. The risk of $\widehat{\Sigma}^{\mathrm{MA}_d}$, $d = 1, 2$, is always lower than that of the competing estimators. Most notably their superiority in risk is kept even when the population eigenvalues are all equal.

Table 4
Convergence speed for $p = 4, m = 1$

	1	0.8	0.6	0.4	0.2	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	Asymp.
$n = 11$												
Prob 1a	0.5852	0.4760	0.3396	0.1881	0.0751	0.0554	0.0517	0.0497	0.0495	0.0498	0.0489	0.0500
Prob 2a	0.3156	0.2620	0.1937	0.1166	0.0660	0.0557	0.0505	0.0494	0.0510	0.0501	0.0492	0.0500
Prob 3a	0.3146	0.2620	0.1920	0.1172	0.0651	0.0560	0.0515	0.0503	0.0496	0.0494	0.0503	0.0500
Prob 4a	0.1955	0.1794	0.1533	0.1124	0.0650	0.0431	0.0283	0.0284	0.0315	0.0449	0.0507	0.0500
Prob 5a	0.1312	0.1206	0.1005	0.0685	0.0342	0.0205	0.0131	0.0131	0.0172	0.0391	0.0497	0.0500
Prob 1b	0.9761	0.9676	0.9610	0.9547	0.9521	0.9510	0.9500	0.9510	0.9495	0.9508	0.9501	0.9500
Prob 2b	0.9986	0.9977	0.9948	0.9863	0.9690	0.9581	0.9505	0.9510	0.9507	0.9509	0.9493	0.9500
Prob 3b	0.9985	0.9979	0.9947	0.9861	0.9681	0.9570	0.9509	0.9515	0.9499	0.9508	0.9508	0.9500
Prob 4b	0.6525	0.6774	0.7152	0.7789	0.8576	0.8979	0.9251	0.9276	0.9310	0.9454	0.9501	0.9500
Prob 5b	0.5899	0.6184	0.6607	0.7327	0.8304	0.8751	0.9091	0.9115	0.9185	0.9399	0.9508	0.9500
Risk 1_U	1.0566	1.0583	1.0552	1.0592	1.0577	1.0583	1.0603	1.0544	1.0574	1.0573	1.0559	1.0585
Risk 1_SDS	0.6514	0.6572	0.6714	0.7092	0.7558	0.7781	0.7954	0.7920	0.7943	0.7942	0.7927	0.7956
Risk 1_MA_1	0.4064	0.4154	0.4367	0.4738	0.5104	0.5295	0.5485	0.5471	0.5484	0.5478	0.5468	0.5496
Risk 2_U	1.8199	1.8213	1.8147	1.8170	1.8175	1.8199	1.8210	1.8147	1.8206	1.8180	1.8176	1.8182
Risk 2_KG	1.0173	1.0214	1.0291	1.0493	1.0749	1.0876	1.0939	1.0921	1.0929	1.0926	1.0915	1.0927
Risk 2_MA_2	0.5967	0.6075	0.6326	0.6767	0.7268	0.7516	0.7728	0.7724	0.7737	0.7733	0.7719	0.7730
$n = 21$												
Prob 1a	0.7030	0.5419	0.3317	0.1428	0.0601	0.0532	0.0503	0.0505	0.0498	0.0495	0.0509	0.0500
Prob 2a	0.4043	0.3183	0.2017	0.0985	0.0579	0.0547	0.0508	0.0495	0.0492	0.0505	0.0505	0.0500
Prob 3a	0.3995	0.3222	0.2019	0.0975	0.0581	0.0527	0.0507	0.0504	0.0496	0.0497	0.0493	0.0500
Prob 4a	0.2413	0.2156	0.1748	0.1172	0.0601	0.0421	0.0297	0.0292	0.0344	0.0480	0.0503	0.0500
Prob 5a	0.1720	0.1533	0.1185	0.0711	0.0331	0.0201	0.0141	0.0137	0.0211	0.0484	0.0505	0.0500
Prob 1b	0.9830	0.9737	0.9627	0.9557	0.9502	0.9506	0.9503	0.9504	0.9497	0.9505	0.9505	0.9500
Prob 2b	0.9989	0.9977	0.9929	0.9809	0.9617	0.9548	0.9507	0.9501	0.9488	0.9500	0.9514	0.9500
Prob 3b	0.9988	0.9975	0.9935	0.9805	0.9628	0.9549	0.9509	0.9502	0.9501	0.9496	0.9487	0.9500
Prob 4b	0.5985	0.6278	0.6881	0.7757	0.8657	0.8998	0.9262	0.9272	0.9339	0.9476	0.9494	0.9500
Prob 5b	0.5303	0.5632	0.6291	0.7282	0.8368	0.8794	0.9118	0.9130	0.9219	0.9479	0.9504	0.9500
Risk 1_U	0.5121	0.5136	0.5135	0.5116	0.5115	0.5128	0.5110	0.5127	0.5115	0.5109	0.5119	0.5127
Risk 1_SDS	0.3503	0.3552	0.3677	0.3871	0.4056	0.4134	0.4167	0.4183	0.4172	0.4169	0.4177	0.4183
Risk 1_MA_1	0.2241	0.2315	0.2461	0.2568	0.2650	0.2715	0.2759	0.2772	0.2759	0.2764	0.2765	0.2769
Risk 2_U	0.9512	0.9514	0.9537	0.9503	0.9516	0.9521	0.9477	0.9535	0.9520	0.9501	0.9505	0.9524
Risk 2_KG	0.6059	0.6109	0.6233	0.6429	0.6607	0.6669	0.6692	0.6708	0.6700	0.6695	0.6707	0.6708
Risk 2_MA_2	0.3510	0.3622	0.3861	0.4114	0.4326	0.4433	0.4516	0.4532	0.4521	0.4524	0.4525	0.4526
$n = 51$												
Prob 1a	0.8209	0.5805	0.2691	0.0916	0.0533	0.0492	0.0498	0.0504	0.0501	0.0504	0.0500	0.0500
Prob 2a	0.5101	0.3626	0.1647	0.0721	0.0560	0.0522	0.0501	0.0502	0.0501	0.0504	0.0500	0.0500
Prob 3a	0.5098	0.3610	0.1669	0.0722	0.0555	0.0533	0.0500	0.0507	0.0479	0.0498	0.0506	0.0500
Prob 4a	0.2912	0.2595	0.1878	0.1118	0.0604	0.0415	0.0303	0.0291	0.0403	0.0507	0.0491	0.0500
Prob 5a	0.2191	0.1863	0.1307	0.0689	0.0308	0.0196	0.0133	0.0148	0.0313	0.0501	0.0499	0.0500
Prob 1b	0.9891	0.9762	0.9649	0.9548	0.9507	0.9501	0.9504	0.9501	0.9504	0.9505	0.9497	0.9500
Prob 2b	0.9992	0.9970	0.9889	0.9700	0.9573	0.9526	0.9502	0.9498	0.9503	0.9507	0.9504	0.9500
Prob 3b	0.9990	0.9973	0.9891	0.9712	0.9565	0.9521	0.9499	0.9513	0.9503	0.9506	0.9492	0.9500
Prob 4b	0.5383	0.5836	0.6703	0.7803	0.8683	0.9022	0.9272	0.9286	0.9411	0.9494	0.9503	0.9500
Prob 5b	0.4666	0.5129	0.6101	0.7334	0.8386	0.8789	0.9081	0.9153	0.9312	0.9496	0.9503	0.9500
Risk 1_U	0.2018	0.2022	0.2019	0.2017	0.2019	0.2017	0.2020	0.2017	0.2020	0.2023	0.2018	0.2016
Risk 1_SDS	0.1566	0.1592	0.1658	0.1721	0.1758	0.1768	0.1780	0.1777	0.1780	0.1783	0.1779	0.1777
Risk 1_MA_1	0.1037	0.1083	0.1109	0.1088	0.1104	0.1113	0.1125	0.1124	0.1123	0.1124	0.1125	0.1122

Table 4 *contd.*

	1	0.8	0.6	0.4	0.2	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	Asymp.
Risk 2_U	0.3923	0.3939	0.3920	0.3916	0.3920	0.3924	0.3927	0.3919	0.3931	0.3929	0.3920	0.3922
Risk 2_KG	0.2896	0.2938	0.3038	0.3127	0.3179	0.3194	0.3208	0.3203	0.3211	0.3215	0.3208	0.3207
Risk 2_MA_2	0.1785	0.1867	0.1943	0.1959	0.2010	0.2033	0.2057	0.2055	0.2054	0.2056	0.2059	0.2055

It seems that $\widehat{\mathbf{\Sigma}}^{\mathrm{MA}_d}$, $d = 1, 2$, has robustness to the deviation from the dispersion of the population eigenvalues.

Because of the robustness, $\widehat{\mathbf{\Sigma}}^{\mathrm{MA}_d}$, $d = 1, 2$, seem to be useful for various applications. Now as the last topic in this section, apart from a decision-theoretic approach, we evaluate these new estimators' performance in discriminant analysis. We use a well-known example of Fisher's iris data. The data consist of 50 samples from each of the three groups(species) with 4-dimensional variable (x_1:sepal length(cm), x_2:sepal width(cm), x_3:petal length(cm), x_4:petal width(cm)). We downloaded the data from the website http://www-unix.oit.umass.edu/~statdata. We let $\boldsymbol{x}_j^{(i)}$, $i = 1, 2, 3$, $j = 1, \ldots, 50$ denote the jth sample in the ith group. The estimators to be tested are the traditional estimators $\widehat{\mathbf{\Sigma}}^{\mathrm{U}}$, $\widehat{\mathbf{\Sigma}}^{\mathrm{SDS}}$, $\widehat{\mathbf{\Sigma}}^{\mathrm{KG}}$ and the new estimators $\widehat{\mathbf{\Sigma}}^{\mathrm{MA}_1}$, $\widehat{\mathbf{\Sigma}}^{\mathrm{MA}_2}$ which are formulated under the condition $p = 4, m = 1$.

We carry out cross validations. Suppose a learning data set $\boldsymbol{y}_j^{(i)}$, $j = 1, \ldots, N$, is chosen from the ith group, $i = 1, 2, 3$. Estimates for the population covariance matrix of the ith group are calculated from $\widehat{\mathbf{\Sigma}}^{\mathrm{U}}$, $\widehat{\mathbf{\Sigma}}^{\mathrm{SDS}}$, $\widehat{\mathbf{\Sigma}}^{\mathrm{KG}}$, $\widehat{\mathbf{\Sigma}}^{\mathrm{MA}_1}$, $\widehat{\mathbf{\Sigma}}^{\mathrm{MA}_2}$ based on

$$\boldsymbol{A}^{(i)} = \sum_{j=1}^{N} (\boldsymbol{y}_j^{(i)} - \bar{\boldsymbol{y}}^{(i)})(\boldsymbol{y}_j^{(i)} - \bar{\boldsymbol{y}}^{(i)})',$$

where $\bar{\boldsymbol{y}}^{(i)} = N^{-1} \sum_{j=1}^{N} \boldsymbol{y}_j^{(i)}$. As a discriminant function, we use a Mahalanobis distance based on each estimates $\widehat{\mathbf{\Sigma}}^{\mathrm{U}}(\boldsymbol{A}^{(i)})$, $\widehat{\mathbf{\Sigma}}^{\mathrm{SDS}}(\boldsymbol{A}^{(i)})$, $\widehat{\mathbf{\Sigma}}^{\mathrm{KG}}(\boldsymbol{A}^{(i)})$, $\widehat{\mathbf{\Sigma}}^{\mathrm{MA}_1}(\boldsymbol{A}^{(i)})$, $\widehat{\mathbf{\Sigma}}^{\mathrm{MA}_2}(\boldsymbol{A}^{(i)})$, that is, for a test data $\boldsymbol{x}$

$$MD_i^* = (\boldsymbol{x} - \bar{\boldsymbol{y}}^{(i)})' \widehat{\mathbf{\Sigma}}^*(\boldsymbol{A}^{(i)})^{-1} (\boldsymbol{x} - \bar{\boldsymbol{y}}^{(i)}), \quad i = 1, 2, 3.$$

The eigenvalues of the covariance matrix within each group is as follows:

$$\begin{aligned} &\text{Group 1:} \quad (0.234, 0.039, 0.027, 0.009), \\ &\text{Group 2:} \quad (0.482, 0.075, 0.056, 0.011), \\ &\text{Group 3:} \quad (0.688, 0.107, 0.057, 0.036). \end{aligned} \tag{32}$$

We observe that (1) in each group, the largest eigenvalue is about 6 times as large as the second largest eigenvalue, (2) the second largest eigenvalue is about 3–7 times as large as the smallest eigenvalue. We are interested in the performance of $\mathbf{\Sigma}^{\mathrm{MA}_d}$, $d = 1, 2$, with the population eigenvalues in (32) which are considered as a deviation from (∞, c, c, c), the ideal eigenvalues for $\mathbf{\Sigma}^{\mathrm{MA}_i}$, $i = 1, 2$.

Table 5
10-sample-set

Learning data set	$\widehat{\Sigma}^{U}$	$\widehat{\Sigma}^{SDS}$	$\widehat{\Sigma}^{KG}$	$\widehat{\Sigma}^{MA_1}$	$\widehat{\Sigma}^{MA_2}$
1	82.50	83.33	83.33	81.67	82.50
2	85.83	85.00	85.00	85.00	85.00
3	82.50	82.50	82.50	82.50	82.50
4	81.67	83.33	82.50	85.83	84.17
5	76.67	77.50	77.50	79.17	79.17
Average	81.83	82.33	82.17	82.83	82.67

Table 6
5-sample-set

Learning data set	$\widehat{\Sigma}^{U}$	$\widehat{\Sigma}^{SDS}$	$\widehat{\Sigma}^{KG}$	$\widehat{\Sigma}^{MA_1}$	$\widehat{\Sigma}^{MA_2}$
1	66.67	71.85	68.89	75.56	75.56
2	78.52	80.00	78.52	85.19	82.96
3	41.48	41.48	41.48	44.44	42.96
4	43.70	46.67	45.93	53.33	50.37
5	88.89	88.15	88.89	92.59	90.37
6	73.33	78.52	77.78	89.63	88.15
7	64.44	68.89	67.41	73.33	71.85
8	73.33	75.56	72.59	82.96	79.26
9	73.33	75.56	72.59	82.96	79.26
10	69.63	72.59	71.85	82.22	77.78
Average	67.33	69.93	68.59	76.22	73.85

We made three types of cross validations.

1. Leave-one-out: For a chosen $(i, j),\ i = 1, 2, 3,\ j = 1, \ldots, 50$, leave $\boldsymbol{x}_j^{(i)}$ out from the whole data to be a test data, and use the rest as a learning data set. We repeat this trial for every possible (i, j). Consequently 150 trials were carried out.
2. 10-sample-set: First choose $\boldsymbol{x}_1^{(i)}, \ldots, \boldsymbol{x}_{10}^{(i)},\ i = 1, 2, 3$, as a learning data set and use all the rest as a test data. Next use $\boldsymbol{x}_{11}^{(i)}, \ldots, \boldsymbol{x}_{20}^{(i)},\ i = 1, 2, 3$, as a learning data set and the others as a test data. Repeatedly change a learning data set until every data is used once as a learning data. Totally we carried out $600(= 120 \times 5)$ trials.
3. 5-sample-set: First choose $\boldsymbol{x}_1^{(i)}, \ldots, \boldsymbol{x}_5^{(i)},\ i = 1, 2, 3$, as a learning data set and use all the rest as a test data. Next use $\boldsymbol{x}_6^{(i)}, \ldots, \boldsymbol{x}_{10}^{(i)},\ i = 1, 2, 3$, as a learning data set and the others as a test data. Repeatedly change a learning data set until all data are used once as a learning data. Totally we carried out $1350(= 135 \times 10)$ trials.

We summarize the result on the correct classification percentage ("C.C.P." for abbreviation) of each discriminant function.

1. Leave-one-out: All the discriminant functions returned the same classification for every test data and scored 96.67% of C.C.P. The misclassification occurred at the sample $\boldsymbol{x}_{19}^{(2)}, \boldsymbol{x}_{21}^{(2)}, \boldsymbol{x}_{23}^{(2)}$,

$\boldsymbol{x}_{34}^{(2)}, \boldsymbol{x}_{32}^{(3)}$. With as much as 49 learning data, all the discriminant functions work quite correctly and make no differences among the functions.

2. 10-sample-set: See Table 5 for the C.C.P. in each learning data set and the average. Depending on the learning data set, different discriminant functions records the best C.C.P, but the margins are small and negligible. It seems that even 10-sample-learning set is too large to differentiate the functions.
3. 5-sample-set: See Table 6 for the C.C.P. in each learning data set and the average. In every learning data set, the functions based on $\widehat{\boldsymbol{\Sigma}}^{\mathrm{MA}_d}$, $d = 1, 2$, outperform the other functions. Especially $\widehat{\boldsymbol{\Sigma}}^{\mathrm{MA}_1}$ always keeps the highest C.C.P. In total, $\widehat{\boldsymbol{\Sigma}}^{\mathrm{MA}_1}$ and $\widehat{\boldsymbol{\Sigma}}^{\mathrm{MA}_2}$ record better C.C.P. than $\widehat{\boldsymbol{\Sigma}}^{\mathrm{U}}$ by 8.89% and 6.52%, respectively, while the margins of $\widehat{\boldsymbol{\Sigma}}^{\mathrm{SDS}}$ and $\widehat{\boldsymbol{\Sigma}}^{\mathrm{KG}}$ over $\widehat{\boldsymbol{\Sigma}}^{\mathrm{U}}$ are, respectively, 2.60% and 1.26%.

References

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

[2] W. James, C. Stein, Estimation with quadratic loss, in: Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, vol. 1, University of California Press, Berkeley, CA, 1961, pp. 361–380.

[3] K. Krishnamoorthy, A.K. Gupta, Improved minimax estimation of a normal precision matrix, Canad. J. Statist. 17 (1989) 91–102.

[4] S.P. Lin, M.D. Perlman, A Monte Carlo comparison of four estimators for a covariance matrix, in: P.R. Krishnaiah (Ed.), Multivariate Analysis, vol. 6, North-Holland, Amsterdam, 1985, pp. 411–429.

[5] R.J. Muirhead, Aspects of Multivariate Statistical Theory, Wiley, New York, 1982.

[6] I. Olkin, J.B. Selliah, Estimating covariances in a multivariate normal distribution, in: S.S. Gupta, D. Moore (Eds.), Statistical Decision Theory and Related Topics, vol. 2, Academic Press, New York, 1977, pp. 313–326.

[7] N. Pal, Estimating the normal dispersion matrix and the precision matrix from a decision-theoretic point of view: a review, Statist. Papers 34 (1993) 1–26.

[8] Y. Sheena, On minimaxity of some orthogonally invariant estimators of bivariate normal dispersion matrix, J. Japan Statist. Soc. 32 (2002) 193–207.

[9] Y. Sheena, A. Takemura, Inadmissibility of non-order-preserving orthogonally invariant estimators of the covariance matrix in the case of Stein's loss, J. Multivariate Anal. 41 (1992) 117–131.

[10] Y. Sheena, A. Takemura, An asymptotic expansion of Wishart distribution when the population eigenvalues are infinitely dispersed, Statist. Method. 4 (2007) 158–184.

[11] Y. Sheena, A. Takemura, Asymptotic distribution of Wishart matrix for block-wise dispersion of population eigenvalues, `arXiv:math.ST/0609275`, 2006.

[12] C. Stein, Estimating the covariance matrix. Unpublished manuscript, 1977.

[13] N. Sugiura, H. Ishibayashi, Reference prior Bayes estimator for bivariate normal covariance matrix with risk comparison, Comm. Statist. Theory Methods 26 (1997) 2203–2221.

[14] A. Takemura, Y. Sheena, Distribution of eigenvalues and eigenvectors of Wishart matrix when the population eigenvalues are infinitely dispersed and its application to minimax estimation of covariance matrix, J. Multivariate Anal. 94 (2005) 271–299.