Ann Inst Stat Math (2015) 67:261–285
DOI 10.1007/s10463-014-0449-x

Minimaxity in estimation of restricted and non-restricted scale parameter matrices

Hisayuki Tsukuma · Tatsuya Kubokawa

Received: 14 August 2013 / Revised: 10 December 2013 / Published online: 6 March 2014

Abstract In estimation of the normal covariance matrix, finding a least favorable sequence of prior distributions has been an open question for a long time. This paper addresses the classical problem and accomplishes the specification of such a sequence, which establishes minimaxity of the best equivariant estimator. This result is extended to the estimation of scale parameter matrix in an elliptically contoured distribution model. The methodology based on a least favorable sequence of prior distributions is applied to both restricted and non-restricted cases of parameters, and we give some examples which show minimaxity of the best equivariant estimators under restrictions of scale parameter matrix.

Keywords Bayesian inference · Equivariance · Least favorable prior · Minimax estimation · Restricted parameter space · Statistical decision theory

1 Introduction

In statistical decision theory of point estimation, minimaxity is a crucial principle, and it is used as an intelligible criterion for measuring quality of estimators. There are two well-known approaches to finding a minimax estimator or establishing minimaxity of a specific estimator: one is the invariance approach and the other is the least favorable prior approach (Strawderman 2000).

H. Tsukuma (✉)
Faculty of Medicine, Toho University, 5-21-16 Omori-nishi, Ota-ku, Tokyo 143-8540, Japan
e-mail: tsukuma@med.toho-u.ac.jp

T. Kubokawa
Faculty of Economics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
e-mail: tatsuya@e.u-tokyo.ac.jp

The invariance approach is based on invariance under a group transformation. A relationship between invariance and minimaxity is often referred to as the Hunt–Stein theorem. It is an elegant theorem and requires invariance of the estimation problem and amenability of the group. For more details and generalizations of the Hunt–Stein theorem, see (Kiefer 1957). Equivalent conditions for amenability of groups were extensively reviewed by Bondar and Milnes (1981).

On the other hand, the least favorable prior approach is a Bayesian method with a least favorable prior distribution or a least favorable sequence of prior distributions (see Berger 1985). A valuable merit of the least favorable prior approach is that it has a wide range of applications. For instance, this approach is of use in the case of a restricted parameter space. Actually, Kubokawa (2004) applied this approach to show that minimaxity of the best equivariant and unrestricted estimators holds true even in the case of restricted location and scale parameters. See also Marchand and Strawderman (2005a, b), who gave the same results as in Kubokawa (2004).

Although the invariance approach is quite general, the least favorable approach gives an alternative and possibly more direct approach. In estimation of covariance matrix of a multivariate normal distribution, the best equivariant estimator under the group transformation of lower triangular matrices with positive diagonal elements, which is also called the James–Stein estimator, is known to be minimax by the Hunt–Stein theorem from the invariance approach. It is noted that other minimax estimation problems related to the covariance matrix have been studied in Selliah (1964), Eaton and Olkin (1987), Krishnamoorthy and Gupta (1989) and others. However, no least favorable sequence of prior distributions has been found since Stein (1956) and James and Stein (1961). This is a most interesting issue in statistical decision theory. Moreover, in the case that the parameter space is restricted, it is not clear whether the best equivariant estimator maintains the minimax property. In this paper, we address these problems and succeed in constructing least favorable sequences of prior distributions in restricted and non-restricted cases of the covariance matrix.

The outline of this paper is as follows. Section 2 addresses the important issue on minimaxity in estimation of the covariance matrix of a multivariate normal distribution model from the least favorable prior approach. An explicit formula of a least favorable sequence of prior distributions is presented in Sect. 2. The result under the normal model is extended to a class of elliptical distribution models including the matrix-variate F distribution.

Section 3 deals with the case that the covariance matrix is limited to a restricted parameter space, which is motivated by Pourahmadi (1999). Then, we construct a least favorable sequence of prior distributions, and establish minimaxity of the best equivariant estimator by applying the same arguments given in Sect. 2. It is shown that the best equivariant estimator is further improved on by the isotonic regression method.

The methods given in Sects. 2 and 3 have the potential to be applied to various restrictions of scale parameter matrix. In Sect. 4, we provide an example of restriction based on determinant of scale parameter matrix, and state concluding remarks. Some proofs are given in the appendix.

2 A least favorable sequence of prior distributions in estimation of covariance matrix

We here construct a least favorable sequence of prior distributions in estimation of the normal covariance and precision matrices. Also, the interesting result is extended to an elliptically contoured distribution model.

2.1 Estimation of normal covariance and precision matrices

Consider the estimation of $\boldsymbol{\Sigma}$ based on a $p \times p$ random matrix $\boldsymbol{V}$ having the Wishart distribution $\mathcal{W}_p(n, \boldsymbol{\Sigma})$. Let $\mathcal{T}^+$ be the set of $p \times p$ lower triangular matrices with positive diagonal entries. By the Cholesky decomposition, $\boldsymbol{\Sigma}^{-1}$ and $\boldsymbol{V}$ can be written as $\boldsymbol{\Sigma}^{-1} = \boldsymbol{\Theta}^t\boldsymbol{\Theta}$ and $\boldsymbol{V} = \boldsymbol{T}\boldsymbol{T}^t$ for $\boldsymbol{\Theta} = (\theta_{ij}) \in \mathcal{T}^+$ and $\boldsymbol{T} = (t_{ij}) \in \mathcal{T}^+$. The probability density function of $\boldsymbol{T}$ is

$$f_W(\boldsymbol{T}|\boldsymbol{\Theta})\gamma(\mathrm{d}\boldsymbol{T}) = C|\boldsymbol{\Theta}\boldsymbol{T}|^n \exp\left[-\frac{1}{2}\mathrm{tr}\,[(\boldsymbol{\Theta}\boldsymbol{T})(\boldsymbol{\Theta}\boldsymbol{T})^t]\right]\gamma(\mathrm{d}\boldsymbol{T}),$$

where C is a normalizing constant and $\gamma(\mathrm{d}\boldsymbol{T}) = (\prod_{i=1}^p t_{ii}^{-i})\mathrm{d}\boldsymbol{T}$, which is left-invariant measure on $\mathcal{T}^+$.

Let $L(\boldsymbol{A})$ be a continuous scalar-valued function of a $p \times p$ matrix $\boldsymbol{A}$. Assume that $L(\boldsymbol{A}) \geq 0$ for any $p \times p$ matrix $\boldsymbol{A}$ and that $L(\boldsymbol{A}) = 0$ if and only if $\boldsymbol{A} = \boldsymbol{I}_p$ for the $p \times p$ identity matrix $\boldsymbol{I}_p$. Define a loss function as $L(\boldsymbol{\Theta}\boldsymbol{\delta}\boldsymbol{\Theta}^t)$, where $\boldsymbol{\delta}$ is an estimator of $\boldsymbol{\Sigma}$. It is noted that $L(\boldsymbol{\Theta}\boldsymbol{\delta}\boldsymbol{\Theta}^t) = 0$ if $\boldsymbol{\delta} = \boldsymbol{\Sigma} = (\boldsymbol{\Theta}^t\boldsymbol{\Theta})^{-1}$. Some examples of loss functions will be given later in this subsection. Denote the risk function by $R(\boldsymbol{\delta}, \boldsymbol{\Sigma}) = \int_{\mathcal{T}^+} L(\boldsymbol{\Theta}\boldsymbol{\delta}\boldsymbol{\Theta}^t) f_W(\boldsymbol{T}|\boldsymbol{\Theta})\gamma(\mathrm{d}\boldsymbol{T})$.

For all $\boldsymbol{A} \in \mathcal{T}^+$, the group transformation with respect to $\mathcal{T}^+$ on a random matrix $\boldsymbol{T}$ and a parameter matrix $\boldsymbol{\Theta}$ is defined by $(\boldsymbol{T}, \boldsymbol{\Theta}) \to (\boldsymbol{A}\boldsymbol{T}, \boldsymbol{\Theta}\boldsymbol{A}^{-1})$. Then the best equivariant estimator with respect to the group $\mathcal{T}^+$ can be written by

$$\boldsymbol{\delta}^{BE} = \boldsymbol{\delta}^{BE}(\boldsymbol{T}) = \arg\min_{\boldsymbol{\delta}} \int_{\boldsymbol{\Theta}\in\mathcal{T}^+} L(\boldsymbol{\Theta}\boldsymbol{\delta}\boldsymbol{\Theta}^t) f_W(\boldsymbol{T}|\boldsymbol{\Theta})\gamma(\mathrm{d}\boldsymbol{\Theta}).$$

This estimator has the form $\boldsymbol{T}\boldsymbol{D}\boldsymbol{T}^t$, where $\boldsymbol{D}$ is a diagonal matrix independent of $\boldsymbol{T}$. The best diagonal matrix $\boldsymbol{D}$, which yields the best equivariant estimator depends on the loss function L. The best equivariant estimator $\boldsymbol{\delta}^{BE}$ has a constant risk R_0, say, and thus the supremum risk over the parameter space is the same as R_0, namely, $\sup_{\boldsymbol{\Theta}\in\mathcal{T}^+} R(\boldsymbol{\delta}^{BE}, \boldsymbol{\Sigma}) = R_0$.

Let $c_{ij} = 3(i - j) - 1$. Define a set P_k of $\boldsymbol{\Theta}$ by

$$P_k = \{\boldsymbol{\Theta} \in \mathcal{T}^+ : 1/k < \theta_{ii} < k \ (i = 1, \dots, p) \text{ and } -k^{c_{ij}}\theta_{ii} < \theta_{ij} < k^{c_{ij}}\theta_{ii} \ (1 \leq j < i \leq p)\}.$$

Then, we consider the sequence of prior distributions given by

$$\pi_k(\boldsymbol{\Theta})\mathrm{d}\boldsymbol{\Theta} = \frac{\gamma(\mathrm{d}\boldsymbol{\Theta})}{V(P_k)} I(\boldsymbol{\Theta} \in P_k), \quad k = 1, 2, \ldots, \tag{1}$$

where $\gamma(\mathrm{d}\boldsymbol{\Theta}) = (\prod_{i=1}^p \theta_{ii}^{-i})\mathrm{d}\boldsymbol{\Theta}$, $V(P_k) = \int_{P_k} \gamma(\mathrm{d}\boldsymbol{\Theta}) = 2^{p(p+1)/2}(\log k)^p \prod_{i=1}^p \prod_{j=1}^{i-1} k^{c_{ij}}$, and $I(\cdot)$ denotes the indicator function.

The prior distributions (1) yield the Bayes estimators

$$\boldsymbol{\delta}_k^\pi = \boldsymbol{\delta}_k^\pi(\boldsymbol{T}) = \arg\min_{\boldsymbol{\delta}} \int_{\boldsymbol{Z}\in P_k} L(\boldsymbol{Z}\boldsymbol{\delta}\boldsymbol{Z}^t) f_W(\boldsymbol{T}|\boldsymbol{Z})\pi_k(\boldsymbol{Z})\mathrm{d}\boldsymbol{Z},$$

with the Bayes risks

$$r_k(\pi_k, \boldsymbol{\delta}_k^\pi) = \frac{1}{V(P_k)} \int_{\boldsymbol{\Theta}\in P_k} \int_{\mathcal{T}^+} L(\boldsymbol{\Theta}\boldsymbol{\delta}_k^\pi(\boldsymbol{T})\boldsymbol{\Theta}^t) f_W(\boldsymbol{T}|\boldsymbol{\Theta})\gamma(\mathrm{d}\boldsymbol{T})\pi_k(\boldsymbol{\Theta})\mathrm{d}\boldsymbol{\Theta}.$$

If the above Bayes risks converge to the supremum risk of $\boldsymbol{\delta}^{BE}$ as $k \to \infty$, namely, $\lim_{k\to\infty} r_k(\pi_k, \boldsymbol{\delta}_k^\pi) = R_0$, then $\boldsymbol{\delta}^{BE}$ is minimax and the sequence $\pi_k(\boldsymbol{\Theta})\mathrm{d}\boldsymbol{\Theta}$ is least favorable (see Proposition 2 of Strawderman 2000). This is proved in the following theorem.

Theorem 1 *The sequence (1) is least favorable, and the best equivariant estimator* $\boldsymbol{\delta}^{BE}$ *is minimax.*

Proof We can show this theorem along the same lines as in Kubokawa (2004) and Kubokawa et al. (2013) who modified the method of Girshick and Savage (1951).

It is easy to check that $r_k(\pi_k, \boldsymbol{\delta}_k^\pi) \le r_k(\pi_k, \boldsymbol{\delta}^{BE}) = R_0$, therefore it is sufficient to show that $\liminf_{k\to\infty} r_k(\pi_k, \boldsymbol{\delta}_k^\pi) \ge R_0$. Making the transformation $\boldsymbol{L} = \boldsymbol{\Theta}\boldsymbol{T}$ yields

$$r_k(\pi_k, \boldsymbol{\delta}_k^\pi) = \frac{1}{V(P_k)} \int_{\boldsymbol{\Theta}\in P_k} \int_{\mathcal{T}^+} L(\boldsymbol{\Theta}\boldsymbol{\delta}_k^\pi(\boldsymbol{\Theta}^{-1}\boldsymbol{L})\boldsymbol{\Theta}^t) f_W(\boldsymbol{L}|\boldsymbol{I}_p)\gamma(\mathrm{d}\boldsymbol{L})\pi_k(\boldsymbol{\Theta})\mathrm{d}\boldsymbol{\Theta}, \tag{2}$$

where $\boldsymbol{\delta}_k^\pi(\boldsymbol{\Theta}^{-1}\boldsymbol{L})$ is expressed as

$$\boldsymbol{\delta}_k^\pi(\boldsymbol{\Theta}^{-1}\boldsymbol{L}) = \arg\min_{\boldsymbol{\delta}} \int_{\boldsymbol{Z}\in P_k} L(\boldsymbol{Z}\boldsymbol{\delta}\boldsymbol{Z}^t) f_W(\boldsymbol{L}|\boldsymbol{Z}\boldsymbol{\Theta}^{-1})\pi_k(\boldsymbol{Z})\mathrm{d}\boldsymbol{Z}.$$

Now, make the transformation $\boldsymbol{Y} = \boldsymbol{Z}\boldsymbol{\Theta}^{-1}$ with $\mathrm{d}\boldsymbol{Z} = (\prod_{i=1}^p \theta_{ii}^{p-i+1})\mathrm{d}\boldsymbol{Y}$. We then have

$$\boldsymbol{\delta}_k^\pi(\boldsymbol{\Theta}^{-1}\boldsymbol{L}) = \arg\min_{\boldsymbol{\delta}} \int_{\boldsymbol{Y}\boldsymbol{\Theta}\in P_k} L(\boldsymbol{Y}\boldsymbol{\Theta}\boldsymbol{\delta}\boldsymbol{\Theta}^t\boldsymbol{Y}^t) f_W(\boldsymbol{L}|\boldsymbol{Y})\pi_k(\boldsymbol{Y})\mathrm{d}\boldsymbol{Y},$$

namely, $\boldsymbol{\Theta}\boldsymbol{\delta}_k^\pi(\boldsymbol{\Theta}^{-1}\boldsymbol{L})\boldsymbol{\Theta}^t = \boldsymbol{\delta}_k^*(\boldsymbol{L}|\boldsymbol{\Theta})$ where

$$\boldsymbol{\delta}_k^*(\boldsymbol{L}|\boldsymbol{\Theta}) = \arg\min_{\boldsymbol{\delta}} \int_{\boldsymbol{Y}\boldsymbol{\Theta}\in P_k} L(\boldsymbol{Y}\boldsymbol{\delta}\boldsymbol{Y}^t) f_W(\boldsymbol{L}|\boldsymbol{Y})\pi_k(\boldsymbol{Y})\mathrm{d}\boldsymbol{Y}.$$

Hence, the Bayes risk (2) can be rewritten as

$$r_k(\pi_k, \boldsymbol{\delta}_k^{\pi}) = \frac{1}{V(P_k)} \int_{\boldsymbol{\Theta}\in P_k} \int_{\mathcal{T}^+} L(\boldsymbol{\delta}_k^*(\boldsymbol{L}|\boldsymbol{\Theta})) f_W(\boldsymbol{L}|\boldsymbol{I}_p)\gamma(\mathrm{d}\boldsymbol{L})\pi_k(\boldsymbol{\Theta})\mathrm{d}\boldsymbol{\Theta}. \tag{3}$$

Let $\xi_{ii} = \log\theta_{ii}/\log k$ for $i = 1, \ldots, p$ and let $\xi_{ij} = \theta_{ij}/(k^{c_{ij}}\theta_{ii})$ for $i > j$. This correspondence is denoted by the function $\boldsymbol{\xi} = \varphi_k(\boldsymbol{\Theta})$ for $\boldsymbol{\xi} = (\xi_{11}, \xi_{21}, \xi_{22}, \ldots, \xi_{p1}, \ldots, \xi_{pp})^t$. Then, we obtain $\varphi_k(P_k) = (-1, 1)^{p(p+1)/2}$ and

$$\gamma_k(\mathrm{d}\boldsymbol{\xi}) = (\log k)^p \left(\prod_{i=1}^{p}\prod_{j=1}^{i-1} k^{c_{ij}}\right) \mathrm{d}\boldsymbol{\xi} = \gamma(\mathrm{d}\boldsymbol{\Theta})$$

for $\gamma(\mathrm{d}\boldsymbol{\Theta}) = (\prod_{i=1}^{p} \theta_{ii}^{-i})\mathrm{d}\boldsymbol{\Theta}$. Note that "$\boldsymbol{Y\Theta} \in P_k$" is equivalent to "$\boldsymbol{Y} \in P_k'(\boldsymbol{\Theta})$", where

$$\begin{aligned} P_k'(\boldsymbol{\Theta}) &= \{\boldsymbol{Y} \in \mathcal{T}^+ : k^{-1} < \{\boldsymbol{Y\Theta}\}_{ii} < k \ \ (i = 1, \ldots, p) \text{ and} \\ &\qquad - k^{c_{ij}}\{\boldsymbol{Y\Theta}\}_{ii} < \{\boldsymbol{Y\Theta}\}_{ij} < k^{c_{ij}}\{\boldsymbol{Y\Theta}\}_{ii} \ \ (i > j)\} \\ &= \Big\{\boldsymbol{Y} \in \mathcal{T}^+ : k^{-1} < y_{ii}\theta_{ii} < k \ \ (i = 1, \ldots, p) \text{ and} \\ &\qquad - k^{c_{ij}} y_{ii}\theta_{ii} < \sum_{m=j}^{i} y_{im}\theta_{mj} < k^{c_{ij}} y_{ii}\theta_{ii} \ \ (i > j)\Big\}. \end{aligned}$$

By the function $\boldsymbol{\xi} = \varphi_k(\boldsymbol{\Theta})$, "$\boldsymbol{Y} \in P_k'(\boldsymbol{\Theta})$" is expressed as "$\boldsymbol{Y} \in \tilde{P}_k(\boldsymbol{\xi})$", where

$$\begin{aligned} \tilde{P}_k(\boldsymbol{\xi}) &= \Big\{\boldsymbol{Y} \in \mathcal{T}^+ : k^{-1} < y_{ii}k^{\xi_{ii}} < k \ \ (i = 1, \ldots, p) \text{ and} \\ &\qquad -y_{ii}k^{c_{ij}+\xi_{ii}} < y_{ij}k^{\xi_{jj}} + \sum_{m=j+1}^{i} y_{im}\xi_{mj}k^{c_{mj}+\xi_{mm}} < y_{ii}k^{c_{ij}+\xi_{ii}} \ (i > j)\Big\} \\ &= \Big\{\boldsymbol{Y} \in \mathcal{T}^+ : k^{-(1+\xi_{ii})} < y_{ii} < k^{1-\xi_{ii}} \ \ (i = 1, \ldots, p) \text{ and} \\ &\qquad L_{ij}(\boldsymbol{Y}, \boldsymbol{\xi}) < y_{ij} < U_{ij}(\boldsymbol{Y}, \boldsymbol{\xi}) \ \ (i > j)\Big\} \end{aligned} \tag{4}$$

with

$$U_{ij}(\boldsymbol{Y}, \boldsymbol{\xi}) = y_{ii}k^{c_{ij}+\xi_{ii}-\xi_{jj}} - \sum_{m=j+1}^{i} y_{im}\xi_{mj}k^{c_{mj}+\xi_{mm}-\xi_{jj}},$$

$$L_{ij}(\boldsymbol{Y}, \boldsymbol{\xi}) = -y_{ii}k^{c_{ij}+\xi_{ii}-\xi_{jj}} - \sum_{m=j+1}^{i} y_{im}\xi_{mj}k^{c_{mj}+\xi_{mm}-\xi_{jj}}.$$

We express $\boldsymbol{\delta}_k^*(\boldsymbol{L}|\boldsymbol{\Theta})$ as

$$\boldsymbol{\delta}_k^*(\boldsymbol{L}|\boldsymbol{\xi}) = \arg\min_{\boldsymbol{\delta}} \int_{\boldsymbol{Y}\in\tilde{P}_k(\boldsymbol{\xi})} L(\boldsymbol{Y}\boldsymbol{\delta}\boldsymbol{Y}^t) f_W(\boldsymbol{L}|\boldsymbol{Y})\pi_k(\boldsymbol{Y})\mathrm{d}\boldsymbol{Y}$$

and the Bayes risk (3) as

$$r_k(\pi_k, \boldsymbol{\delta}_k^\pi) = \frac{1}{2^q}\int_{\boldsymbol{\xi}\in\varphi_k(P_k)}\int_{\mathcal{T}^+} L(\boldsymbol{\delta}_k^*(\boldsymbol{L}|\boldsymbol{\xi})) f_W(\boldsymbol{L}|\boldsymbol{I}_p)\gamma(\mathrm{d}\boldsymbol{L})\mathrm{d}\boldsymbol{\xi} \tag{5}$$

for $q = p(p+1)/2$.

It is noted that, for any small $\varepsilon > 0$,

$$\varphi_k(P_k) = \prod_{i=1}^{q}(-1, 1) \supset \prod_{i=1}^{q}(-1+\varepsilon, 1-\varepsilon) \equiv I_\varepsilon.$$

Then, the Bayes risk (5) is evaluated as

$$r_k(\pi_k, \boldsymbol{\delta}_k^\pi) \ge \frac{1}{2^q}\int_{\boldsymbol{\xi}\in I_\varepsilon}\int_{\mathcal{T}^+} L(\boldsymbol{\delta}_k^*(\boldsymbol{L}|\boldsymbol{\xi})) f_W(\boldsymbol{L}|\boldsymbol{I}_p)\gamma(\mathrm{d}\boldsymbol{L})\mathrm{d}\boldsymbol{\xi}.$$

As proved in Lemma 1 given below, we see that, for $\boldsymbol{\xi} \in I_\varepsilon$, $\boldsymbol{\delta}_k^*(\boldsymbol{L}|\boldsymbol{\xi}) \to \boldsymbol{\delta}^{BE}(\boldsymbol{L})$ as $k \to \infty$. Hence, Fatou's lemma is used to bound the Bayes risk from below as

$$\begin{aligned}
\liminf_{k\to\infty} r_k(\pi_k, \boldsymbol{\delta}_k^\pi) &\ge \frac{1}{2^q}\int_{\boldsymbol{\xi}\in I_\varepsilon}\int_{\mathcal{T}^+} f_W(\boldsymbol{L}|\boldsymbol{I}_p)\cdot\liminf_{k\to\infty} L(\boldsymbol{\delta}_k^*(\boldsymbol{L}|\boldsymbol{\xi}))\gamma(\mathrm{d}\boldsymbol{L})\mathrm{d}\boldsymbol{\xi}\\
&= \frac{1}{2^q}\int_{\boldsymbol{\xi}\in I_\varepsilon}\mathrm{d}\boldsymbol{\xi}\int_{\mathcal{T}^+} f_W(\boldsymbol{L}|\boldsymbol{I}_p) L(\boldsymbol{\delta}^{BE}(\boldsymbol{L}))\gamma(\mathrm{d}\boldsymbol{L})\\
&= (1-\varepsilon)^q R(\boldsymbol{\delta}^{BE}(\boldsymbol{L}), \boldsymbol{I}_p) = (1-\varepsilon)^q R_0.
\end{aligned}$$

From the arbitrariness of $\varepsilon > 0$, it follows that $\liminf_{k\to\infty} r_k(\pi_k, \boldsymbol{\delta}_k^\pi) \ge R_0$, completing the proof of Theorem 1. □

To complete the proof of Theorem 1, we need the following lemma, whose proof will be given in the Appendix.

Lemma 1 *For $\boldsymbol{\xi} \in I_\varepsilon$, it holds that $\boldsymbol{\delta}_k^*(\boldsymbol{L}|\boldsymbol{\xi}) \to \boldsymbol{\delta}^{BE}(\boldsymbol{L})$ as $k \to \infty$.*

In estimation of the covariance matrix $\boldsymbol{\Sigma}$, Stein (1956) employed the so-called Stein loss function given by

$$L_S(\boldsymbol{\delta}, \boldsymbol{\Sigma}) = \operatorname{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{\delta} - \log|\boldsymbol{\Sigma}^{-1}\boldsymbol{\delta}| - p = \operatorname{tr}\boldsymbol{\Theta}\boldsymbol{\delta}\boldsymbol{\Theta}^t - \log|\boldsymbol{\Theta}\boldsymbol{\delta}\boldsymbol{\Theta}^t| - p, \tag{6}$$

and the best equivariant estimator is given by $\boldsymbol{\delta}^{BE} = \boldsymbol{T}\boldsymbol{D}^{BE}\boldsymbol{T}^t$, where

$$\boldsymbol{D}^{BE} = \left[\int_{\mathcal{T}^+}\boldsymbol{\Theta}^t\boldsymbol{\Theta} f_W(\boldsymbol{I}_p|\boldsymbol{\Theta})\gamma(\mathrm{d}\boldsymbol{\Theta})\right]^{-1} = \operatorname{diag}(d_1, \ldots, d_p)$$

for $d_i = (n+p-2i+1)^{-1}$. From Theorem 1, $\boldsymbol{\delta}^{BE}$ is minimax relative to the Stein loss (6), and the least favorable sequence of prior distributions is given in (1). Similarly, we use the following invariant loss functions

$$L_Q(\boldsymbol{\delta}, \boldsymbol{\Sigma}) = \mathrm{tr}\,(\boldsymbol{\Sigma}^{-1}\boldsymbol{\delta} - \boldsymbol{I}_p)^2 = \mathrm{tr}\,(\boldsymbol{\Theta}\boldsymbol{\delta}\boldsymbol{\Theta}^t - \boldsymbol{I}_p)^2,$$
$$L_P(\boldsymbol{\delta}, \boldsymbol{\Sigma}) = \mathrm{tr}\,\boldsymbol{\Sigma}\boldsymbol{\delta}^{-1} - \log|\boldsymbol{\Sigma}\boldsymbol{\delta}^{-1}| - p = \mathrm{tr}\,(\boldsymbol{\Theta}\boldsymbol{\delta}\boldsymbol{\Theta}^t)^{-1} - \log|(\boldsymbol{\Theta}\boldsymbol{\delta}\boldsymbol{\Theta}^t)^{-1}| - p,$$

so that the resulting best equivariant estimators are minimax. It is noted that the loss L_P is used in estimation of the precision matrix $\boldsymbol{\Sigma}^{-1}$ rather than of the covariance matrix $\boldsymbol{\Sigma}$. For more details of estimation with respect to L_Q and L_P, see Selliah (1964) and Krishnamoorthy and Gupta (1989), respectively.

The same arguments can be used for estimation of $\boldsymbol{\Theta}$ based on the Cholesky decomposition. The best equivariant estimator of $\boldsymbol{\Theta}$ is given in Eaton and Olkin (1987), and its minimaxity can be shown by the arguments based on the sequence of prior distributions.

2.2 Extension to elliptical distributions

The results given in Sect. 2.1 can be extended to a class of elliptical distributions whose probability density function (p.d.f.) with respect to $\gamma(\mathrm{d}\boldsymbol{T})$ is given by

$$f_\phi(\boldsymbol{T}|\boldsymbol{\Theta})\gamma(\mathrm{d}\boldsymbol{T}) = |\boldsymbol{\Theta}\boldsymbol{T}|^n\phi(\boldsymbol{\Theta}\boldsymbol{T}\{\boldsymbol{\Theta}\boldsymbol{T}\}^t)\gamma(\mathrm{d}\boldsymbol{T}) \tag{7}$$

for an integrable function $\phi(\cdot)$. It is assumed that ϕ satisfies $\phi(\boldsymbol{A}) = \phi(\boldsymbol{BAB})$ for a squared matrix $\boldsymbol{A}$ and a diagonal matrix $\boldsymbol{B}$ with diagonal elements ± 1. Consider here the problem of estimating the scale parameter matrix $\boldsymbol{\Sigma} = (\boldsymbol{\Theta}^t\boldsymbol{\Theta})^{-1}$ relative to an invariant loss $L(\boldsymbol{\Theta}\boldsymbol{\delta}\boldsymbol{\Theta}^t)$. Then, the best equivariant estimator is given by

$$\boldsymbol{\delta}_\phi^{BE} = \boldsymbol{\delta}_\phi^{BE}(\boldsymbol{T}) = \arg\min_{\boldsymbol{\delta}} \int_{\boldsymbol{\Theta}\in\mathcal{T}^+} L(\boldsymbol{\Theta}\boldsymbol{\delta}\boldsymbol{\Theta}^t) f_\phi(\boldsymbol{T}|\boldsymbol{\Theta})\gamma(\mathrm{d}\boldsymbol{\Theta}) = \boldsymbol{T}\boldsymbol{D}_\phi\boldsymbol{T}^t, \tag{8}$$

where $\boldsymbol{D}_\phi$ is a diagonal matrix whose diagonal elements are constants depending on the functions ϕ and L. Then, we obtain the following theorem based on the same arguments as in the previous subsection.

Theorem 2 *Assume that the random matrix is distributed as an elliptical distribution with the p.d.f. (7). Then the sequence (1) is least favorable, and the best equivariant estimator* $\boldsymbol{\delta}_\phi^{BE}$ *is minimax.*

An example of $f_\phi(\boldsymbol{T}|\boldsymbol{\Theta})\gamma(\mathrm{d}\boldsymbol{T})$ is the p.d.f. of matrix-variate F distribution or matrix-variate beta distribution, which is expressed as

$$f_F(\boldsymbol{T}|\boldsymbol{\Theta})\gamma(\mathrm{d}\boldsymbol{T}) = C|\boldsymbol{\Theta}\boldsymbol{T}|^n|\boldsymbol{I}_p + \boldsymbol{\Theta}\boldsymbol{T}\{\boldsymbol{\Theta}\boldsymbol{T}\}^t|^{-(v+n+p-1)/2}\gamma(\mathrm{d}\boldsymbol{T}),$$

where C is a normalizing constant, and v is a positive constant. The best equivariant estimator of $\boldsymbol{\Sigma} = (\boldsymbol{\Theta}^t\boldsymbol{\Theta})^{-1}$ relative to the Stein loss (6) is $\boldsymbol{\delta}_F^{BE} =$

$[\int \boldsymbol{\Theta}^t \boldsymbol{\Theta} f_F(\boldsymbol{T}|\boldsymbol{\Theta})\gamma(\mathrm{d}\boldsymbol{\Theta})]^{-1}$. Making transformation from $\boldsymbol{\Theta}$ to $\boldsymbol{\Theta}\boldsymbol{T}^{-1}$ gives the expression $\boldsymbol{\delta}_F^{BE} = \boldsymbol{T}\boldsymbol{D}_F^{BE}\boldsymbol{T}^t$ where

$$\boldsymbol{D}_F^{BE} = \left[\int_{\mathcal{T}^+} \boldsymbol{\Theta}^t \boldsymbol{\Theta} f_F(\boldsymbol{I}_p|\boldsymbol{\Theta})\gamma(\mathrm{d}\boldsymbol{\Theta}) \right]^{-1}. \tag{9}$$

A direct calculation from (9) yields the following lemma, which will be proved in the appendix.

Lemma 2 *Suppose that $v > 2$. The exact value of $\boldsymbol{D}_F^{BE}$ defined in (9) is given by $\boldsymbol{D}_F^{BE} = \mathrm{diag}\,(h_1^{-1}, \ldots, h_p^{-1})$ for*

$$h_i = \frac{n-i+1}{v+i-3} \cdot \frac{v+p-2}{v+i-2} + \frac{p-i}{v+i-2} \qquad (i = 1, \ldots, p).$$

Theorem 2 establishes that the best equivariant estimator $\boldsymbol{\delta}_F^{BE}$ is minimax. It is noted that $\boldsymbol{\delta}_F^{BE}$ is the same as a minimax estimator obtained by Muirhead and Verathaworn (1985) from the invariance approach. Our notation (n, p, v) corresponds to $(n_1, m, n_2 - m + 1)$ in their notation. See Section 3 of Muirhead and Verathaworn (1985).

3 Estimation under restriction of lower triangular matrix

3.1 Minimaxity under order-restricted diagonal elements

Consider the unique reparametrization for $\boldsymbol{\Sigma}$ of the form $\boldsymbol{\Gamma}\boldsymbol{\Sigma}\boldsymbol{\Gamma}^t = \boldsymbol{\Lambda}$, where $\boldsymbol{\Gamma} = (\gamma_{ij})$ is a lower triangular matrix with unit diagonal elements, $\gamma_{ii} = 1$, and $\boldsymbol{\Lambda} = \mathrm{diag}\,(\lambda_1, \ldots, \lambda_p)$ with positive diagonal elements λ_i. Pourahmadi (1999) has pointed out a statistical meaning of the γ_{ij} and λ_i in analysis of longitudinal data, and showed that they are interpreted as the autoregressive coefficients and the innovation (residual) variances, respectively.

In the previous section, we used the Cholesky decomposition $\boldsymbol{\Sigma}^{-1} = \boldsymbol{\Theta}^t\boldsymbol{\Theta}$ where $\boldsymbol{\Theta} = (\theta_{ij}) \in \mathcal{T}^+$. It then follows that $\lambda_i = \theta_{ii}^{-2}$ and $\gamma_{ij} = \theta_{ij}/\theta_{ii}$ because $\boldsymbol{\Sigma}^{-1} = \boldsymbol{\Gamma}^t\boldsymbol{\Lambda}^{-1}\boldsymbol{\Gamma}$ and the Cholesky decomposition is unique. In this section, we consider the restriction $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$, namely,

$$\theta_{11}^{-2} \geq \theta_{22}^{-2} \geq \cdots \geq \theta_{pp}^{-2}, \quad \text{or, equivalently, } \theta_{11} \leq \theta_{22} \leq \cdots \leq \theta_{pp}. \tag{10}$$

This signifies that the innovation variances decrease as time goes. For simple explanation of this constraint, see Pourahmadi (1999, Section 2.6).

For $c_{ij} = 3(i-j) - 1$, define a set P_k^L of $\boldsymbol{\Theta}$ by

$$P_k^L = \Big\{\boldsymbol{\Theta} \in \mathcal{T}^+ : k^{-1/2} \leq \theta_{11} \leq k^{1/2},\ 1 \leq \theta_{ii}/\theta_{i-1,i-1} \leq k^2\ (i = 2, \ldots, p) \text{ and}$$
$$-k^{c_{ij}}\theta_{ii} \leq \theta_{ij} \leq k^{c_{ij}}\theta_{ii}\ (1 \leq j < i \leq p)\Big\}.$$

Then, we consider the sequence of prior distributions given by

$$\pi_k(\boldsymbol{\Theta})\mathrm{d}\boldsymbol{\Theta} = \frac{\gamma(\mathrm{d}\boldsymbol{\Theta})}{V(P_k^L)} I(\boldsymbol{\Theta} \in P_k^L), \quad k = 1, 2, \ldots, \tag{11}$$

where $\gamma(\mathrm{d}\boldsymbol{\Theta}) = (\prod_{i=1}^p \theta_{ii}^{-i})\mathrm{d}\boldsymbol{\Theta}$ and $V(P_k^L) = 2^{p(p+1)/2-1}(\log k)^p \prod_{i=1}^p \prod_{j=1}^{i-1} k^{c_{ij}}$.

Theorem 3 *For an elliptical distribution (7) with the restriction (10), the best equivariant and unrestricted estimator $\boldsymbol{\delta}_\phi^{BE}$, given in (8), is minimax and the sequence (11) is least favorable.*

Proof An outline of the proof is given because the proof is similar to that of Theorem 1.

Using the same lines from beginning to (3) in the proof of Theorem 1, we express the Bayes risk with respect to P_k^L as

$$r_k(\pi_k, \boldsymbol{\delta}_k^\pi) = \frac{1}{V(P_k^L)} \int_{\boldsymbol{\Theta}\in P_k^L} \int_{\boldsymbol{L}\in\mathcal{T}^+} L(\boldsymbol{\delta}_k^*(\boldsymbol{L}|\boldsymbol{\Theta})) f_\phi(\boldsymbol{L}|\boldsymbol{I}_p)\gamma(\mathrm{d}\boldsymbol{L})\pi_k(\boldsymbol{\Theta})\mathrm{d}\boldsymbol{\Theta}, \tag{12}$$

where

$$\boldsymbol{\delta}_k^*(\boldsymbol{L}|\boldsymbol{\Theta}) = \arg\min_{\boldsymbol{\delta}} \int_{\boldsymbol{Y\Theta}\in P_k^L} L(\boldsymbol{Y\delta Y}^t) f_\phi(\boldsymbol{L}|\boldsymbol{Y})\pi_k(\boldsymbol{Y})\mathrm{d}\boldsymbol{Y}.$$

Since we easily see that, for all k, $r_k(\pi_k, \boldsymbol{\delta}_k^\pi) \le R_\phi = \int_{\boldsymbol{T}\in\mathcal{T}^+} L(\boldsymbol{\delta}_\phi^{BE}(\boldsymbol{T})) f_\phi(\boldsymbol{T}|\boldsymbol{I}_p)$ $\gamma(\mathrm{d}\boldsymbol{T})$, it will be shown that $\liminf_{k\to\infty} r_k(\pi_k, \boldsymbol{\delta}_k^\pi) \ge R_\phi$.

Let $P = \{\boldsymbol{\Theta} \in \mathcal{T}^+ : \theta_{11} \le \theta_{22} \le \cdots \le \theta_{pp}\}$. It is seen that $\cup_{k=1}^\infty P_k^L = P \subset \mathcal{T}^+$. The function $\boldsymbol{\xi} = \varphi_k(\boldsymbol{\Theta})$ is defined by

$$\xi_{11} = \frac{2\log\theta_{11}}{\log k}, \quad \xi_{ii} = \frac{1}{\log k}\log\frac{\theta_{ii}}{\theta_{i-1,i-1}} - 1 \quad (i = 2, \ldots, p),$$
$$\xi_{ij} = \frac{\theta_{ij}}{\theta_{ii}k^{c_{ij}}} \quad (1 \le j < i \le p).$$

Then, it follows that $\varphi_k(P_k^L) = [-1, 1]^{p(p+1)/2}$ and

$$\gamma_k(\mathrm{d}\boldsymbol{\xi}) = 2^{-1}(\log k)^p \left(\prod_{i=1}^p \prod_{j=1}^{i-1} k^{c_{ij}}\right) \mathrm{d}\boldsymbol{\xi} = \gamma(\mathrm{d}\boldsymbol{\Theta})$$

for $\gamma(\mathrm{d}\boldsymbol{\Theta}) = (\prod_{i=1}^p \theta_{ii}^{-i})\mathrm{d}\boldsymbol{\Theta}$.

It is noted that "$\boldsymbol{Y\Theta} \in P_k^L$" is written as "$\boldsymbol{Y} \in P_k'(\boldsymbol{\Theta})$", where

$$\begin{aligned} P_k'(\boldsymbol{\Theta}) &= \Big\{ \boldsymbol{Y} \in \mathcal{T}^+ : k^{-1/2} \le \{\boldsymbol{Y\Theta}\}_{11} \le k^{1/2},\ 1 \le \tfrac{\{\boldsymbol{Y\Theta}\}_{ii}}{\{\boldsymbol{Y\Theta}\}_{i-1,i-1}} \le k^2 \\ &\qquad (i = 2, \dots, p) \quad \text{and} - k^{c_{ij}}\{\boldsymbol{Y\Theta}\}_{ii} \le \{\boldsymbol{Y\Theta}\}_{ij} \le k^{c_{ij}}\{\boldsymbol{Y\Theta}\}_{ii} \\ &\qquad\qquad (1 \le j < i \le p) \Big\} \\ &= \Big\{ \boldsymbol{Y} \in \mathcal{T}^+ : k^{-1/2} \le y_{11}\theta_{11} \le k^{1/2},\ 0 \le \tfrac{y_{ii}\theta_{ii}}{y_{i-1,i-1}\theta_{i-1,i-1}} \le k^2 \\ &\qquad (i = 2, \dots, p) \quad \text{and} - k^{c_{ij}} y_{ii}\theta_{ii} \le \textstyle\sum_{m=j}^{i} y_{im}\theta_{mj} \le k^{c_{ij}} y_{ii}\theta_{ii} \\ &\qquad\qquad (1 \le j < i \le p) \Big\}. \end{aligned}$$

It also follows that $\theta_{ii} = k^{\xi_{11}/2+\sum_{\ell=2}^{i}(1+\xi_{\ell\ell})}$ for $i \ge 2$. Thus, by the function $\boldsymbol{\xi} = \varphi_k(\boldsymbol{\Theta})$, "$\boldsymbol{Y} \in P_k'(\boldsymbol{\Theta})$" is expressed as "$\boldsymbol{Y} \in \tilde{P}_k(\boldsymbol{\xi})$", where

$$\tilde{P}_k(\boldsymbol{\xi}) = \{\boldsymbol{Y} \in \mathcal{T}^+ : L_{ij}(\boldsymbol{Y}, \boldsymbol{\xi}) \le y_{ij} \le U_{ij}(\boldsymbol{Y}, \boldsymbol{\xi}) \quad (1 \le j \le i \le p)\}$$

with

$$L_{ij}(\boldsymbol{Y}, \boldsymbol{\xi}) = \begin{cases} k^{-(1+\xi_{11})/2} & \text{if } i = j = 1, \\ y_{i-1,i-1}k^{-(1+\xi_{ii})} & \text{if } i = j \ge 2, \\ -y_{ii}k^{c_{ij}+\sum_{\ell=j+1}^{i}(1+\xi_{\ell\ell})} & \\ \quad - \sum_{m=j+1}^{i} y_{im}\xi_{mj}k^{c_{mj}+\sum_{\ell=j+1}^{m}(1+\xi_{\ell\ell})} & \text{if } i > j, \end{cases}$$

$$U_{ij}(\boldsymbol{Y}, \boldsymbol{\xi}) = \begin{cases} k^{(1-\xi_{11})/2} & \text{if } i = j = 1, \\ y_{i-1,i-1}k^{1-\xi_{ii}} & \text{if } i = j \ge 2, \\ y_{ii}k^{c_{ij}+\sum_{\ell=j+1}^{i}(1+\xi_{\ell\ell})} & \\ \quad - \sum_{m=j+1}^{i} y_{im}\xi_{mj}k^{c_{mj}+\sum_{\ell=j+1}^{m}(1+\xi_{\ell\ell})} & \text{if } i > j. \end{cases}$$

The function $\boldsymbol{\xi} = \varphi_k(\boldsymbol{\Theta})$ transforms $\boldsymbol{\delta}_k^*(\boldsymbol{L}|\boldsymbol{\Theta})$ to

$$\boldsymbol{\delta}_k^*(\boldsymbol{L}|\boldsymbol{\xi}) = \arg\min_{\boldsymbol{\delta}} \int_{\boldsymbol{Y} \in \tilde{P}_k(\boldsymbol{\xi})} L(\boldsymbol{Y\delta Y}^t) f_\phi(\boldsymbol{L}|\boldsymbol{Y})\pi_k(\boldsymbol{Y})\mathrm{d}\boldsymbol{Y}.$$

Let $|\xi_{ij}| < 1 - \varepsilon$ for $i \ge j$ and any small $\varepsilon > 0$. We here use the same way as in the proof of Lemma 1 and can easily show that $P_k^* = \{\boldsymbol{Y} \in \mathcal{T}^+ : L_{ij}^* < y_{ij} < U_{ij}^*\ (1 \le j \le i \le p)\} \subset \tilde{P}_k(\boldsymbol{\xi})$, where

$$L_{ij}^* = \begin{cases} k^{-\varepsilon/2} & \text{if } i = j, \\ -\varepsilon(1 - k^{-1})^{i-j-1}k^{\varepsilon} & \text{if } i > j, \end{cases}$$

$$U_{ij}^* = \begin{cases} k^{\varepsilon/2} & \text{if } i = j, \\ \varepsilon(1 - k^{-1})^{i-j-1}k^{\varepsilon} & \text{if } i > j. \end{cases}$$

It then holds that $\cup_{k=1}^{\infty} P_k^* = \mathcal{T}^+$, which implies that, for $\boldsymbol{\xi} \in I_\varepsilon$, $\boldsymbol{\delta}_k^*(\boldsymbol{L}|\boldsymbol{\xi}) \to \boldsymbol{\delta}_\phi^{BE}(\boldsymbol{L})$ as $k \to \infty$. We, hence obtain $\liminf_{k\to\infty} r_k(\pi_k, \boldsymbol{\delta}_k^\pi) \geq R_\phi$ in the same way as the last part after (5) in the proof of Theorem 1. □

3.2 Improvement by the isotonic regression method

In the previous subsection, the best equivariant and unrestricted estimator is shown to remain minimax under the restriction (10). In this subsection, it is shown that the best equivariant estimator can be further improved on by the isotonic regression method under the restriction (10).

Let $\boldsymbol{T}_1$ be the lower triangular matrix with unit diagonal elements such that, for $i > j$, the (i, j) off-diagonal elements are t_{ij}/t_{jj}, where the t_{ij} are elements of $\boldsymbol{T} = (t_{ij}) \in \mathcal{T}^+$. Then, the best equivariant estimator with respect to the Stein loss (6) is rewritten as

$$\boldsymbol{\delta}_\phi^{BE} = \boldsymbol{T}\boldsymbol{D}_\phi\boldsymbol{T}^t = \boldsymbol{T}_1\boldsymbol{\Psi}^{BE}(\boldsymbol{t})\boldsymbol{T}_1^t$$

for $\boldsymbol{t} = (t_{11}^2, \ldots, t_{pp}^2)$, where $\boldsymbol{\Psi}^{BE}(\boldsymbol{t}) = \mathrm{diag}\,(d_1 t_{11}^2, \ldots, d_p t_{pp}^2)$ and the d_i are positive constants. If $\theta_{11}^{-2} \geq \theta_{22}^{-2} \geq \cdots \geq \theta_{pp}^{-2}$ on $\boldsymbol{\Sigma} = (\boldsymbol{\Theta}^t\boldsymbol{\Theta})^{-1}$, we should modify the ordering property of diagonal elements of $\boldsymbol{\Psi}^{BE}(\boldsymbol{t})$ as long as $\Pr(d_1 t_{11}^2 \geq \cdots \geq d_p t_{pp}^2) < 1$.

To revise the unnatural ordering, we apply the isotonic regression to diagonal elements of $\boldsymbol{\Psi}^{BE}(\boldsymbol{t})$. Let $\boldsymbol{\Psi}^{BE}(\boldsymbol{t}) = \mathrm{diag}\,(\psi_1^{BE}, \ldots, \psi_p^{BE})$ with $\psi_i^{BE} = d_i t_{ii}^2$, and let $\boldsymbol{\Psi}^{IR}(\boldsymbol{t}) = \mathrm{diag}\,(\psi_1^{IR}, \ldots, \psi_p^{IR})$, where $\{\psi_1^{IR}, \ldots, \psi_p^{IR}\}$ is a solution of minimizing $\sum_{i=1}^p (\lambda_i - \psi_i^{BE})^2$ subject to $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$. For more details of the isotonic regression, refer to Robertson et al. (1988).

Theorem 4 *For an elliptical distribution (7) with the restriction (10), suppose that* $\Pr(d_1 t_{11}^2 \geq \cdots \geq d_p t_{pp}^2) < 1$. *Then* $\boldsymbol{\delta}_\phi^{IR} = \boldsymbol{T}_1\boldsymbol{\Psi}^{IR}(\boldsymbol{t})\boldsymbol{T}_1^t$ *is minimax estimator dominating* $\boldsymbol{\delta}_\phi^{BE}$ *relative to the Stein loss (6).*

We verify this theorem via the following lemma. For details of the lemma, see Rockafellar (1970) and Calvin and Dykstra (1991).

Lemma 3 (Fenchel's duality theorem) *Let $f(\boldsymbol{x})$ be a concave function defined in $\mathbb{R}^p$, and let $\mathcal{K}$ be a closed convex cone in $\mathbb{R}^p$. Define the concave conjugate of $f(\boldsymbol{x})$ and the dual cone of $\mathcal{K}$ as, respectively,*

$$f^*(\boldsymbol{y}) = \inf_{\boldsymbol{x}\in\mathbb{R}^p} \left\{ \sum_{i=1}^p x_i y_i - f(\boldsymbol{x}) \right\}, \quad \mathcal{K}^* = \left\{ \boldsymbol{y} \in \mathbb{R}^p : \sum_{i=1}^p x_i y_i \leq 0, \ \forall \boldsymbol{x} \in \mathcal{K} \right\}.$$

Then, we have

$$\sup_{\boldsymbol{x}\in\mathcal{K}} f(\boldsymbol{x}) = - \sup_{\boldsymbol{y}\in\mathcal{K}^*} f^*(\boldsymbol{y}) \tag{13}$$

if either ri(dom f) $\cap$ ri($\mathcal{K}$) $\neq \emptyset$ or ri(dom f^*) $\cap$ ri($\mathcal{K}^*$) $\neq \emptyset$, where ri *means relative interior and* dom $f = \{\boldsymbol{x} \in \mathbb{R}^p : f(\boldsymbol{x}) > -\infty\}$. *Denote by* $\boldsymbol{x}^*$ *and* $\boldsymbol{y}^*$*, respectively, solutions of the left- and right-hand sides of (13). It then holds that* (a) $\boldsymbol{x}^* \in \mathcal{K}$, (b) $\boldsymbol{y}^* \in \mathcal{K}^*$, (c) $(\boldsymbol{x}^*)^t \boldsymbol{y}^* = \sum_{i=1}^p x_i^* y_i^* = 0$ *and* (d) $-\boldsymbol{y}^*$ *is a subgradient of* $-f$ *at* $\boldsymbol{x}^*$.

Proof of Theorem 4 The proof can be done along the same lines as in Tsukuma and Kubokawa (2011). Let $\boldsymbol{\Psi}(\boldsymbol{t}) = \mathrm{diag}\,(\psi_1, \ldots, \psi_p)$ whose diagonal elements are functions of $\boldsymbol{t}$. Recall that $\boldsymbol{\Sigma}^{-1} = \boldsymbol{\Gamma}^t \boldsymbol{\Lambda}^{-1} \boldsymbol{\Gamma}$ with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$. Then, the risk of estimator $\boldsymbol{\delta}_\phi = \boldsymbol{T}_1 \boldsymbol{\Psi}(\boldsymbol{t}) \boldsymbol{T}_1^t$ is written as

$$\begin{aligned} R(\boldsymbol{\delta}_\phi, \boldsymbol{\Sigma}) &= E[\mathrm{tr}\,\boldsymbol{\Sigma}^{-1} \boldsymbol{T}_1 \boldsymbol{\Psi}(\boldsymbol{t}) \boldsymbol{T}_1^t - \log|\boldsymbol{\Sigma}^{-1} \boldsymbol{T}_1 \boldsymbol{\Psi}(\boldsymbol{t}) \boldsymbol{T}_1^t| - p] \\ &= E[\mathrm{tr}\,\boldsymbol{\Lambda}^{-1} \boldsymbol{U} \boldsymbol{\Psi}(\boldsymbol{t}) \boldsymbol{U}^t - \log|\boldsymbol{\Lambda}^{-1} \boldsymbol{\Psi}(\boldsymbol{t})| - \log|\boldsymbol{U}\boldsymbol{U}^t| - p], \end{aligned}$$

where the second equality follows from the transformation $\boldsymbol{U} = (u_{ij}) = \boldsymbol{\Gamma} \boldsymbol{T}_1$. The first term of the last right-hand side is expressed as

$$E[\mathrm{tr}\,\boldsymbol{\Lambda}^{-1} \boldsymbol{U} \boldsymbol{\Psi}(\boldsymbol{t}) \boldsymbol{U}^t] = E\left[\sum_{i=1}^p \psi_i \{\boldsymbol{U}^t \boldsymbol{\Lambda}^{-1} \boldsymbol{U}\}_{ii}\right] = E\left[\sum_{i=1}^p \psi_i \sum_{j=i}^p u_{ji}^2/\lambda_j\right]. \tag{14}$$

Let

$$a_i = a_i(\boldsymbol{t}) = E\left[\sum_{j=i+1}^p u_{ji}^2/\lambda_j \,\middle|\, \boldsymbol{t}\right]. \tag{15}$$

It follows that $u_{ii} = 1$ for $i = 1, \ldots, p$. Note that a_i depends on the probability density function (7) and that if (7) is the normal density then $a_i = (p-i)/t_{ii}^2$, since the Bartlett decomposition leads to the fact that $t_{ii}^2/\lambda_i \sim \chi^2_{n-i+1}$ for $i = 1, \ldots, p$ and $u_{ji}|t_{ii} \sim \mathcal{N}(0, \lambda_j/t_{ii}^2)$ for $j > i$. Combining (14) and (15), we obtain

$$E[\mathrm{tr}\,\boldsymbol{\Lambda}^{-1} \boldsymbol{U} \boldsymbol{\Psi}(\boldsymbol{t}) \boldsymbol{U}^t] = E\left[\sum_{i=1}^p \psi_i (\lambda_i^{-1} + a_i)\right],$$

which yields

$$R(\boldsymbol{\delta}_\phi, \boldsymbol{\Sigma}) = E\left[\sum_{i=1}^p \{\psi_i(\lambda_i^{-1} + a_i) - \log(\psi_i/\lambda_i)\}\right] - p. \tag{16}$$

Let $\boldsymbol{\psi}^{BE} = (\psi_1^{BE}, \ldots, \psi_p^{BE})^t$ and $\boldsymbol{\psi}^{IR} = (\psi_1^{IR}, \ldots, \psi_p^{IR})^t$. For $i = 1, \ldots, p$, let $\xi_i = \lambda_i^{-1}$. Denote $\mathbb{R}_+^p = \{\boldsymbol{x} \in \mathbb{R}^p : x_i > 0 \text{ for each } i\}$. Then, $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_p)^t$ belongs to

$$\mathcal{K} = \{\boldsymbol{\xi} \in \mathbb{R}_+^p : \xi_1 \leq \cdots \leq \xi_p\}.$$

Also, denote the dual cone of $\mathcal{K}$ by

$$\mathcal{K}^* = \{\boldsymbol{\eta} \in \mathbb{R}^p : \boldsymbol{\eta}^t \boldsymbol{x} \leq 0 \text{ for any } \boldsymbol{x} \in \mathcal{K}\}.$$

Let the objective function be

$$\ell(\boldsymbol{\xi}|\boldsymbol{\psi}^{BE}) = \sum_{i=1}^{p} \{\log \xi_i - \psi_i^{BE}(\xi_i + a_i)\},$$

which is the concave function of $\boldsymbol{\xi}$. It is noted from Robertson et al. (1988) that ψ_i^{IR}'s are the same as certain solutions $\hat{\xi}_i^{-1}$ of maximizing $\sum_{i=1}^{p}\{\log \xi_i - \psi_i^{BE}\xi_i\}$ subject to $\boldsymbol{\xi} \in \mathcal{K}$ and, moreover, the $\hat{\xi}_i$'s are equivalent to solutions of maximizing $\ell(\boldsymbol{\xi}|\boldsymbol{\psi}^{BE})$ subject to $\boldsymbol{\xi} \in \mathcal{K}$.

The concave conjugate function of $\ell(\boldsymbol{\xi}|\boldsymbol{\psi}^{BE})$ is given by

$$\begin{aligned}
\ell^*(\boldsymbol{\eta}|\boldsymbol{\psi}^{BE}) &= \inf_{\boldsymbol{\xi} \in \mathbb{R}_+^p} \left\{ \sum_{i=1}^{p} \xi_i \eta_i - \ell(\boldsymbol{\xi}|\boldsymbol{\psi}^{BE}) \right\} \\
&= \inf_{\boldsymbol{\xi} \in \mathbb{R}_+^p} \left[\sum_{i=1}^{p} \{\xi_i(\eta_i + \psi_i^{BE}) - \log \xi_i\} \right] + \sum_{i=1}^{p} \psi_i^{BE} a_i \\
&= \sum_{i=1}^{p} \log(\eta_i + \psi_i^{BE}) + p + \sum_{i=1}^{p} \psi_i^{BE} a_i
\end{aligned}$$

and the domain of $\ell^*(\boldsymbol{\eta}|\boldsymbol{\psi}^{BE})$ is $\{\boldsymbol{\eta} \in \mathbb{R}^p : \boldsymbol{\eta} + \boldsymbol{\psi}^{BE} \succ \mathbf{0}_p\}$, where "$\succ$" stands for "is componentwise greater than".

The subgradient of $-\ell(\boldsymbol{\xi}|\boldsymbol{\psi}^{BE})$ is equal to $(\psi_1^{BE} - \xi_1^{-1}, \ldots, \psi_p^{BE} - \xi_p^{-1})^t$, so Lemma 3 (d) implies that the supremum of $\ell^*(\boldsymbol{\eta}|\boldsymbol{\psi}^{BE})$ is attained at

$$\widehat{\boldsymbol{\eta}} = (\hat{\xi}_1^{-1} - \psi_1^{BE}, \ldots, \hat{\xi}_p^{-1} - \psi_p^{BE})^t.$$

Since $\hat{\xi}_i^{-1} = \psi_i^{IR}$, we can see that

$$-\sup_{\substack{\boldsymbol{\eta} \in \mathcal{K}^* \\ \boldsymbol{\eta} + \boldsymbol{\psi}^{BE} \succ 0_p}} \ell^*(\boldsymbol{\eta}|\boldsymbol{\psi}^{BE}) = -\ell^*(\widehat{\boldsymbol{\eta}}|\boldsymbol{\psi}^{BE}) = -\sum_{i=1}^{p} \log \psi_i^{IR} - p - \sum_{i=1}^{p} \psi_i^{BE} a_i. \tag{17}$$

It is noted that $\psi_i^{BE} e^{a_i(\psi_i^{IR}-\psi_i^{BE})} > 0$ for each i. Replacing ψ_i^{IR} by $\psi_i^{BE} e^{a_i(\psi_i^{IR}-\psi_i^{BE})}$ in the above expression yields

$$-\sup_{\substack{\boldsymbol{\eta}\in\mathcal{K}^* \\ \boldsymbol{\eta}+\boldsymbol{\psi}^{BE}\succ 0_p}} \ell^*(\boldsymbol{\eta}|\boldsymbol{\psi}^{BE}) \le -\sum_{i=1}^p \log(\psi_i^{BE} e^{a_i(\psi_i^{IR}-\psi_i^{BE})}) - p - \sum_{i=1}^p \psi_i^{BE} a_i$$
$$= -\sum_{i=1}^p \log \psi_i^{BE} - p - \sum_{i=1}^p \psi_i^{IR} a_i. \tag{18}$$

Combining (17) and (18) gives that $\sum_{i=1}^p (\psi_i^{IR} a_i - \log \psi_i^{IR}) \le \sum_{i=1}^p (\psi_i^{BE} a_i - \log \psi_i^{BE})$, or, equivalently,

$$\sum_{i=1}^p \{\psi_i^{IR} a_i - \log(\psi_i^{IR}/\lambda_i)\} \le \sum_{i=1}^p \{\psi_i^{BE} a_i - \log(\psi_i^{BE}/\lambda_i)\}. \tag{19}$$

From the fact that $\widehat{\boldsymbol{\eta}} \in \mathcal{K}^*$ and $\boldsymbol{\xi} \in \mathcal{K}$, it follows that $\widehat{\boldsymbol{\lambda}}^t \boldsymbol{\xi} \le 0$, namely,

$$\sum_{i=1}^p (\psi_i^{IR} - \psi_i^{BE})\xi_i = \sum_{i=1}^p (\psi_i^{IR} - \psi_i^{BE})\lambda_i^{-1} \le 0. \tag{20}$$

Combining (19) and (20), we can see that

$$\sum_{i=1}^p \{\psi_i^{IR}(\lambda_i^{-1} + a_i) - \log(\psi_i^{IR}/\lambda_i)\} \le \sum_{i=1}^p \{\psi_i^{BE}(\lambda_i^{-1} + a_i) - \log(\psi_i^{BE}/\lambda_i)\}$$

with probability one. Thus, it follows from (16) that $R(\boldsymbol{\delta}_\phi^{IR}, \boldsymbol{\Sigma}) \le R(\boldsymbol{\delta}_\phi^{BE}, \boldsymbol{\Sigma})$, which implies that $\boldsymbol{\delta}_\phi^{IR}$ is a minimax estimator improving on $\boldsymbol{\delta}_\phi^{BE}$. $\square$

4 Concluding remarks

In this paper, we have considered some problems related to estimating covariance matrix of multivariate normal distribution model, and have used sequences of prior distributions to show minimaxity of the resulting best equivariant estimator. The approaches to minimaxity can be used in both restricted and non-restricted cases of parameters. The most striking result of the paper is that we have succeeded in deriving an explicit formula for a least favorable sequence of prior distributions for the covariance matrix. This has been an open question for a long time, since Stein (1956) and James and Stein (1961). The least favorable prior approach to minimaxity is constructive and pedagogic. We have also applied the same arguments given in Theorem 1 to the restricted case of lower triangular matrix for establishing minimaxity of the best equivariant and unrestricted estimators, which is further improved on by the isotonic regression method.

The least favorable prior approach used in the proofs of Theorem 1 has the potential to be applied to various restrictions of covariance and precision matrices. For example, consider the restrictions $|\boldsymbol{\Sigma}| \leq c$ and $|\boldsymbol{\Sigma}| \geq c$ for positive c, namely, $\prod_{i=1}^{p} \theta_{ii}^2 \geq 1/c$ and $\prod_{i=1}^{p} \theta_{ii}^2 \leq 1/c$. The case of $|\boldsymbol{\Sigma}| \geq c$ has been treated by Marchand and Strawderman (2012), who established minimaxity of the best equivariant estimator based on a different method from our approach. See Marchand and Strawderman (2012) for more details. For the case of $|\boldsymbol{\Sigma}| \leq c$, we can get the following theorem which is shown in the appendix.

Theorem 5 *For an elliptical distribution (7) with the restriction $|\boldsymbol{\Sigma}| \leq c$, the best equivariant estimator of $\boldsymbol{\Sigma}$, given in (8), is minimax.*

The least favorable sequences of prior distributions addressed in this paper are extended to general estimation problems with an invariance structure and a unified methodology can be presented for both restricted and non-restricted cases. See Tsukuma and Kubokawa (2012).

The methods given in this paper can be applied to more general models with both location and scale parameters. For instance, we can handle the case that a sample mean vector is available, which can be described as $\boldsymbol{V} \sim \mathcal{W}_p(n, \boldsymbol{\Sigma})$ and $\boldsymbol{X} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Our results can be easily extended to this model, and minimaxity for the best equivariant estimator of $\boldsymbol{\Sigma}$ is established.

Appendix A

A.1 Proof of Lemma 1

For $i > j$, let $U_{ij}^* = \varepsilon(1 - k^{-1})^{i-j-1} k^{\varepsilon}$ and $L_{ij}^* = -\varepsilon(1 - k^{-1})^{i-j-1} k^{\varepsilon}$. Define

$$P_k^* = \{\boldsymbol{Y} \in \mathcal{T}^+ : k^{-\varepsilon} < y_{ii} < k^{\varepsilon} \quad (i = 1, \dots, p) \quad \text{and} \quad L_{ij}^* < y_{ij} < U_{ij}^* \quad (i > j)\}.$$

Recall that $\tilde{P}_k(\boldsymbol{\xi})$ is given in (4). If it is proved that $P_k^* \subset \tilde{P}_k(\boldsymbol{\xi})$ for all k and any $\boldsymbol{\xi} \in I_{\varepsilon}$, then $\boldsymbol{\delta}_k^*(\boldsymbol{L}|\boldsymbol{\xi}) \to \boldsymbol{\delta}^{BE}(\boldsymbol{L})$ as $k \to \infty$ because $\cup_{k=1}^{\infty} P_k^* = \mathcal{T}^+$. Then, the proof of this lemma will be complete.

Denote, for $i = 1, \dots, p$,

$$R_{ii} = \{y_{ii} \in \mathbb{R} : k^{-(1+\xi_{ii})} < y_{ii} < k^{1-\xi_{ii}}\},$$
$$R_{ii}^* = \{y_{ii} \in \mathbb{R} : k^{-\varepsilon} < y_{ii} < k^{\varepsilon}\},$$

and, for $i > j$,

$$R_{ij} = \{y_{ij} \in \mathbb{R} : L_{ij}(\boldsymbol{Y}, \boldsymbol{\xi}) < y_{ij} < U_{ij}(\boldsymbol{Y}, \boldsymbol{\xi})\},$$
$$\tilde{R}_{ij} = \{y_{ij} \in \mathbb{R} : -y_{ii} B_{ij}(\boldsymbol{\xi}) < y_{ij} < y_{ii} B_{ij}(\boldsymbol{\xi})\},$$
$$R^*_{ij} = \{y_{ij} \in \mathbb{R} : L^*_{ij} < y_{ij} < U^*_{ij}\},$$

where $B_{ij}(\boldsymbol{\xi}) = \varepsilon(1-k^{-1})^{i-j-1} k^{c_{ij}+\xi_{ii}-\xi_{jj}}$. It is here noted that $\tilde{P}_k(\boldsymbol{\xi}) = \cap_{i=1}^{p} \cap_{j=1}^{i} R_{ij}$ and $P^*_k = \cap_{i=1}^{p} \cap_{j=1}^{i} R^*_{ij}$. To prove that $P^*_k \subset \tilde{P}_k(\boldsymbol{\xi})$, we will establish the following inclusion relations

$$\tilde{P}_k(\boldsymbol{\xi}) \supset \bigcap_{i=1}^{p} \left\{ \left(\bigcap_{j=1}^{i-1} \tilde{R}_{ij} \right) \cap R_{ii} \right\} \supset P^*_k. \tag{21}$$

For proof of the first inclusion relation in (21), we inductively show that for each i

$$\bigcap_{j=1}^{i-1} R_{ij} \supset \left(\bigcap_{j=1}^{i-2} R_{ij} \right) \cap \tilde{R}_{i,i-1} \supset \left(\bigcap_{j=1}^{i-3} R_{ij} \right) \cap \left(\bigcap_{j=i-2}^{i-1} \tilde{R}_{ij} \right) \supset \cdots \supset \bigcap_{j=1}^{i-1} \tilde{R}_{ij}. \tag{22}$$

It is observed that $1 - \xi_{ij} > \varepsilon$ and $1 + \xi_{ij} > \varepsilon$ for $i \geq j$ since $\boldsymbol{\xi} \in I_\varepsilon = (-1+\varepsilon, 1-\varepsilon)^q$. For $j = i-1$, it follows that

$$\begin{aligned} U_{i,i-1}(\boldsymbol{Y}, \boldsymbol{\xi}) &= y_{ii}(1 - \xi_{i,i-1}) k^{c_{i,i-1}+\xi_{ii}-\xi_{i-1,i-1}} > y_{ii} \varepsilon k^{c_{i,i-1}+\xi_{ii}-\xi_{i-1,i-1}} \\ &= y_{ii} B_{i,i-1}(\boldsymbol{\xi}), \\ L_{i,i-1}(\boldsymbol{Y}, \boldsymbol{\xi}) &= -y_{ii}(1 + \xi_{i,i-1}) k^{c_{i,i-1}+\xi_{ii}-\xi_{i-1,i-1}} < -y_{ii} \varepsilon k^{c_{i,i-1}+\xi_{ii}-\xi_{i-1,i-1}} \\ &= -y_{ii} B_{i,i-1}(\boldsymbol{\xi}). \end{aligned}$$

We thus have $R_{i,i-1} \supset \tilde{R}_{i,i-1}$, which yields

$$\bigcap_{j=1}^{i-1} R_{ij} = \left(\bigcap_{j=1}^{i-2} R_{ij} \right) \cap R_{i,i-1} \supset \left(\bigcap_{j=1}^{i-2} R_{ij} \right) \cap \tilde{R}_{i,i-1}.$$

This implies that the first inclusion relation in (22) is true.

Next, suppose that the $(i-\ell-1)$-th inclusion relation in (22) holds true, namely,

$$\left(\bigcap_{j=1}^{\ell+1} R_{ij} \right) \cap \left(\bigcap_{j=\ell+2}^{i-1} \tilde{R}_{ij} \right) \supset \left(\bigcap_{j=1}^{\ell} R_{ij} \right) \cap \left(\bigcap_{j=\ell+1}^{i-1} \tilde{R}_{ij} \right).$$

We then verify the $(i-\ell)$-th inclusion relation of (22),

$$\left(\bigcap_{j=1}^{\ell} R_{ij}\right) \cap \left(\bigcap_{j=\ell+1}^{i-1} \tilde{R}_{ij}\right) \supset \left(\bigcap_{j=1}^{\ell-1} R_{ij}\right) \cap \left(\bigcap_{j=\ell}^{i-1} \tilde{R}_{ij}\right), \tag{23}$$

by means of proving that

$$R_{i\ell} \cap \left(\bigcap_{j=\ell+1}^{i-1} \tilde{R}_{ij}\right) \supset \bigcap_{j=\ell}^{i-1} \tilde{R}_{ij}. \tag{24}$$

For $j = \ell+1, \ldots, i-1$, $\tilde{R}_{ij}$ implies that $|y_{ij}/y_{ii}| < B_{ij}(\boldsymbol{\xi})$. Then, the upper bound of $R_{i\ell}$ is evaluated below as

$$\begin{aligned}
U_{i\ell}(\boldsymbol{Y}, \boldsymbol{\xi}) &= y_{ii} k^{c_{i\ell}+\xi_{ii}-\xi_{\ell\ell}} \left\{1 - \xi_{i\ell} - \sum_{m=\ell+1}^{i-1} \frac{y_{im}}{y_{ii}} \xi_{m\ell} k^{c_{m\ell}-c_{i\ell}+\xi_{mm}-\xi_{ii}}\right\} \\
&> y_{ii} k^{c_{i\ell}+\xi_{ii}-\xi_{\ell\ell}} \left\{1 - \xi_{i\ell} - \sum_{m=\ell+1}^{i-1} \left|\frac{y_{im}}{y_{ii}}\right| \cdot |\xi_{m\ell}| k^{c_{m\ell}-c_{i\ell}+\xi_{mm}-\xi_{ii}}\right\} \\
&> y_{ii} k^{c_{i\ell}+\xi_{ii}-\xi_{\ell\ell}} \left\{1 - \xi_{i\ell} - \sum_{m=\ell+1}^{i-1} B_{im}(\boldsymbol{\xi}) \cdot |\xi_{m\ell}| k^{c_{m\ell}-c_{i\ell}+\xi_{mm}-\xi_{ii}}\right\}.
\end{aligned}$$

Noting that $1 - \xi_{i\ell} > \varepsilon$ and $|\xi_{m\ell}| < 1$, we obtain

$$\begin{aligned}
U_{i\ell}(\boldsymbol{Y}, \boldsymbol{\xi}) &> y_{ii} k^{c_{i\ell}+\xi_{ii}-\xi_{\ell\ell}} \left\{\varepsilon - \sum_{m=\ell+1}^{i-1} B_{im}(\boldsymbol{\xi}) k^{c_{m\ell}-c_{i\ell}+\xi_{mm}-\xi_{ii}}\right\} \\
&= y_{ii} k^{c_{i\ell}+\xi_{ii}-\xi_{\ell\ell}} \left\{\varepsilon - \sum_{m=\ell+1}^{i-1} \varepsilon (1-k^{-1})^{i-m-1} k^{c_{im}+c_{m\ell}-c_{i\ell}}\right\}.
\end{aligned}$$

Recall that $c_{ij} = 3(i-j)-1$. Since $c_{im} + c_{m\ell} - c_{i\ell} = 3(i-m) - 1 + 3(m-\ell) - 1 - 3(i-\ell) + 1 = -1$, it is observed that

$$\begin{aligned}
U_{i\ell}(\boldsymbol{Y}, \boldsymbol{\xi}) &> y_{ii} k^{c_{i\ell}+\xi_{ii}-\xi_{\ell\ell}} \left\{\varepsilon - \sum_{m=\ell+1}^{i-1} \varepsilon (1-k^{-1})^{i-m-1} k^{-1}\right\} \\
&= y_{ii} k^{c_{i\ell}+\xi_{ii}-\xi_{\ell\ell}} \varepsilon \left\{1 - \frac{1}{k} - \left(1-\frac{1}{k}\right)\frac{1}{k} - \left(1-\frac{1}{k}\right)^2 \frac{1}{k}\right. \\
&\qquad \left. - \cdots - \left(1-\frac{1}{k}\right)^{i-\ell-2} \frac{1}{k}\right\}
\end{aligned}$$

$$= y_{ii} k^{c_{i\ell}+\xi_{ii}-\xi_{\ell\ell}} \varepsilon \left(1-\frac{1}{k}\right)^{i-\ell-1} = y_{ii} B_{i\ell}(\boldsymbol{\xi}).$$

Similarly, it is seen that, when $|y_{ij}/y_{ii}| < B_{ij}(\boldsymbol{\xi})$ for $j = \ell+1, \ldots, i-1$,

$$\begin{aligned} L_{i\ell}(\boldsymbol{Y}, \boldsymbol{\xi}) &= -y_{ii} k^{c_{i\ell}+\xi_{ii}-\xi_{\ell\ell}} \left\{1+\xi_{i\ell} + \sum_{m=\ell+1}^{i-1} \frac{y_{im}}{y_{ii}} \xi_{m\ell} k^{c_{m\ell}-c_{i\ell}+\xi_{mm}-\xi_{ii}}\right\} \\ &< -y_{ii} k^{c_{i\ell}+\xi_{ii}-\xi_{\ell\ell}} \left\{1+\xi_{i\ell} - \sum_{m=\ell+1}^{i-1} \left|\frac{y_{im}}{y_{ii}}\right| \cdot |\xi_{m\ell}| k^{c_{m\ell}-c_{i\ell}+\xi_{mm}-\xi_{ii}}\right\} \\ &< -y_{ii} B_{i\ell}(\boldsymbol{\xi}), \end{aligned}$$

so that

$$\begin{aligned} R_{i\ell} \cap \left(\bigcap_{j=\ell+1}^{i-1} \tilde{R}_{ij}\right) &= \{y_{i\ell} \in \mathbb{R} : L_{i\ell}(\boldsymbol{Y}, \boldsymbol{\xi}) < y_{i\ell} < U_{i\ell}(\boldsymbol{Y}, \boldsymbol{\xi})\} \cap \left(\bigcap_{j=\ell+1}^{i-1} \tilde{R}_{ij}\right) \\ &\supset \{y_{i\ell} \in \mathbb{R} : -y_{ii} B_{i\ell}(\boldsymbol{\xi}) < y_{i\ell} < y_{ii} B_{i\ell}(\boldsymbol{\xi})\} \cap \left(\bigcap_{j=\ell+1}^{i-1} \tilde{R}_{ij}\right) \\ &= \bigcap_{j=\ell}^{i-1} \tilde{R}_{ij}. \end{aligned}$$

Hence we get (24), which immediately gives the inclusion relation (23). Inductively repeating the relation (23), we obtain the inclusion relations (22) for each i, which establishes the first inclusion relation in (21).

Next, the second inclusion relation in (21) is shown. Noting that $1-\xi_{ij} > \varepsilon$ and $1+\xi_{ij} > \varepsilon$ for $i \geq j$, we obtain, for $i = 1, \ldots, p$,

$$R_{ii} = \{y_{ii} \in \mathbb{R} : k^{-(1+\xi_{ii})} < y_{ii} < k^{1-\xi_{ii}}\} \supset \{y_{ii} \in \mathbb{R} : k^{-\varepsilon} < y_{ii} < k^{\varepsilon}\} = R_{ii}^* \tag{25}$$

and also, for $i > j$,

$$\begin{aligned} B_{ij}(\boldsymbol{\xi}) &= \varepsilon(1-k^{-1})^{i-j-1} k^{3(i-j)-1+\xi_{ii}-\xi_{jj}} \\ &= \varepsilon(1-k^{-1})^{i-j-1} k^{3(i-j-1)+(1+\xi_{ii})+(1-\xi_{jj})} \\ &> \varepsilon(1-k^{-1})^{i-j-1} k^{3(i-j-1)+2\varepsilon} \\ &> \varepsilon(1-k^{-1})^{i-j-1} k^{2\varepsilon}, \end{aligned} \tag{26}$$

where the last inequality follows from the fact that $i-j-1\geq 0$. When $y_{ii}\in R^*_{ii}$, using the inequality (26) gives, for $i>j$,

$$y_{ii}B_{ij}(\boldsymbol{\xi})>k^{-\varepsilon}B_{ij}(\boldsymbol{\xi})>k^{-\varepsilon}\times\varepsilon(1-k^{-1})^{i-j-1}k^{2\varepsilon}=\varepsilon(1-k^{-1})^{i-j-1}k^{\varepsilon},$$

which yields that

$$\begin{aligned}R^*_{ii}\cap\tilde{R}_{ij}&=R^*_{ii}\cap\{y_{ij}\in\mathbb{R}:-y_{ii}B_{ij}(\boldsymbol{\xi})<y_{ij}<y_{ii}B_{ij}(\boldsymbol{\xi})\}\\&\supset R^*_{ii}\cap\{y_{ij}\in\mathbb{R}:-\varepsilon(1-k^{-1})^{i-j-1}k^{\varepsilon}<y_{ij}<\varepsilon(1-k^{-1})^{i-j-1}k^{\varepsilon}\}\\&=R^*_{ii}\cap R^*_{ij}.\end{aligned}\tag{27}$$

Using (25) and (27) gives

$$\begin{aligned}\bigcap_{i=1}^{p}\left\{\left(\bigcap_{j=1}^{i-1}\tilde{R}_{ij}\right)\cap R_{ii}\right\}&\supset\bigcap_{i=1}^{p}\left\{\left(\bigcap_{j=1}^{i-1}\tilde{R}_{ij}\right)\cap R^*_{ii}\right\}\\&\supset\bigcap_{i=1}^{p}\left\{\left(\bigcap_{j=1}^{i-1}R^*_{ij}\right)\cap R^*_{ii}\right\}=P^*_k,\end{aligned}$$

which establishes the second inclusion relation in (21). It hence holds that $P^*_k\subset\tilde{P}_k(\boldsymbol{\xi})$ for all k, which completes the proof of Lemma 1.

A.2 Proof of Lemma 2

It is noted that the integral in D^{BE}_F given in (9) is invariant under transformation $\boldsymbol{\Theta}\to\boldsymbol{B}\boldsymbol{\Theta}\boldsymbol{B}$ where $\boldsymbol{B}$ is a diagonal matrix such that all diagonal elements are respectively either one or minus one. Denote by E_* the expectation with respect to the probability density function $f_F(\boldsymbol{I}_p|\boldsymbol{\Theta})\gamma(\mathrm{d}\boldsymbol{\Theta})$ and let

$$\boldsymbol{H}_p=E_*[\boldsymbol{\Theta}^t\boldsymbol{\Theta}]=\int\boldsymbol{\Theta}^t\boldsymbol{\Theta}f_F(\boldsymbol{I}_p|\boldsymbol{\Theta})\gamma(\mathrm{d}\boldsymbol{\Theta}).$$

Partition $\boldsymbol{\Theta}$ into four blocks as follows:

$$\boldsymbol{\Theta}=\begin{pmatrix}\boldsymbol{\Theta}_{11}&\boldsymbol{0}_{p-1}\\\boldsymbol{\theta}^t_{21}&\theta_{pp}\end{pmatrix},$$

where the sizes of $\boldsymbol{\Theta}_{11}$ and $\boldsymbol{\theta}_{21}$ are, respectively, $(p-1)\times(p-1)$ and $(p-1)\times 1$. It is seen that

$$\begin{aligned}|\boldsymbol{I}_p+\boldsymbol{\Theta}\boldsymbol{\Theta}^t| &= \begin{vmatrix}\boldsymbol{I}_{p-1}+\boldsymbol{\Theta}_{11}\boldsymbol{\Theta}_{11}^t & \boldsymbol{\Theta}_{11}\boldsymbol{\theta}_{21}\\ \boldsymbol{\theta}_{21}^t\boldsymbol{\Theta}_{11}^t & 1+\theta_{pp}^2+\boldsymbol{\theta}_{21}^t\boldsymbol{\theta}_{21}\end{vmatrix}\\ &= |\boldsymbol{I}_{p-1}+\boldsymbol{\Theta}_{11}\boldsymbol{\Theta}_{11}^t|\{1+\theta_{pp}^2+\boldsymbol{\theta}_{21}^t\boldsymbol{\theta}_{21}\\ &\quad -\boldsymbol{\theta}_{21}^t\boldsymbol{\Theta}_{11}^t(\boldsymbol{I}_{p-1}+\boldsymbol{\Theta}_{11}\boldsymbol{\Theta}_{11}^t)^{-1}\boldsymbol{\Theta}_{11}\boldsymbol{\theta}_{21}\}\\ &= |\boldsymbol{I}_{p-1}+\boldsymbol{\Theta}_{11}\boldsymbol{\Theta}_{11}^t|\{1+\theta_{pp}^2+\boldsymbol{\theta}_{21}^t(\boldsymbol{I}_{p-1}+\boldsymbol{\Theta}_{11}^t\boldsymbol{\Theta}_{11})^{-1}\boldsymbol{\theta}_{21}\},\end{aligned}$$

which yields

$$\begin{aligned}f_F(\boldsymbol{I}_p|\boldsymbol{\Theta})\gamma(\mathrm{d}\boldsymbol{\Theta}) &= C|\boldsymbol{I}_{p-1}+\boldsymbol{\Theta}_{11}\boldsymbol{\Theta}_{11}^t|^{-a}(1+\theta_{pp}^2)^{-a}(1\\ &\quad+\boldsymbol{\theta}_{21}^t\boldsymbol{G}^{-1}\boldsymbol{\theta}_{21})^{-a}\left(\prod_{i=1}^p\theta_{ii}^{n-i}\right)\mathrm{d}\boldsymbol{\Theta}_{11}\mathrm{d}\boldsymbol{\theta}_{21}\mathrm{d}\theta_{pp}\\ &= C|\boldsymbol{I}_{p-1}+\boldsymbol{\Theta}_{11}\boldsymbol{\Theta}_{11}^t|^{-a+1/2}(1+\theta_{pp}^2)^{-a+(p-1)/2}\left(\prod_{i=1}^p\theta_{ii}^{n-i}\right)\\ &\quad\times|\boldsymbol{G}|^{-1/2}(1+\boldsymbol{\theta}_{21}^t\boldsymbol{G}^{-1}\boldsymbol{\theta}_{21})^{-a}\mathrm{d}\boldsymbol{\Theta}_{11}\mathrm{d}\boldsymbol{\theta}_{21}\mathrm{d}\theta_{pp}\end{aligned}$$

with $a=(v+n+p-1)/2$ and $\boldsymbol{G}=(1+\theta_{pp}^2)(\boldsymbol{I}_{p-1}+\boldsymbol{\Theta}_{11}^t\boldsymbol{\Theta}_{11})$. Hence, the marginal distribution of $\{(v+p-1)/(n-p+1)\}\theta_{pp}^2$ is the F distribution with $n-p+1$ and $v+p-1$ degrees of freedom, and the conditional distribution of $(v+n)^{1/2}\boldsymbol{\theta}_{21}$ given $\boldsymbol{\Theta}_{11}$ and θ_{pp} is the $(p-1)$-dimensional t distribution with $v+n$ degrees of freedom, mean zero and scale matrix $\boldsymbol{G}$. Letting $\boldsymbol{H}_{p-1}=E_*[\boldsymbol{\Theta}_{11}^t\boldsymbol{\Theta}_{11}]$, we obtain

$$E_*[\theta_{pp}^2]=\frac{n-p+1}{v+p-3}$$

and

$$E_*[\boldsymbol{\theta}_{21}\boldsymbol{\theta}_{21}^t]=E_*\left[\frac{1+\theta_{pp}^2}{v+n-2}(\boldsymbol{I}_{p-1}+\boldsymbol{\Theta}_{11}^t\boldsymbol{\Theta}_{11})\right]=\frac{1}{v+p-3}(\boldsymbol{I}_{p-1}+\boldsymbol{H}_{p-1}),$$

which implies that

$$\begin{aligned}\boldsymbol{H}_p &= \begin{pmatrix}E_*[\boldsymbol{\Theta}_{11}^t\boldsymbol{\Theta}_{11}+\boldsymbol{\theta}_{21}\boldsymbol{\theta}_{21}^t] & \boldsymbol{0}_{p-1}\\ \boldsymbol{0}_{p-1}^t & E_*[\theta_{pp}^2]\end{pmatrix}\\ &= \begin{pmatrix}(1+\beta_{p-1})\boldsymbol{H}_{p-1}+\beta_{p-1}\boldsymbol{I}_{p-1} & \boldsymbol{0}_{p-1}\\ \boldsymbol{0}_{p-1}^t & \alpha_p\end{pmatrix},\end{aligned}$$

where $\alpha_p=E_*[\theta_{pp}^2]=(n-p+1)/(v+p-3)$ and $\beta_{p-1}=1/(v+p-3)$.

Similarly, let $\boldsymbol{A}_i$ be the $i \times i$ left upper corner of $\boldsymbol{\Theta}$ and denote $\boldsymbol{H}_i = E_*[\boldsymbol{A}_i^t \boldsymbol{A}_i]$. Then, it holds that

$$\alpha_i = E_*[\theta_{ii}^2] = \frac{n-i+1}{v+i-3}$$

and

$$\boldsymbol{H}_i = \begin{pmatrix} (1+\beta_{i-1})\boldsymbol{H}_{i-1} + \beta_{i-1}\boldsymbol{I}_{i-1} & \boldsymbol{0}_{i-1} \\ \boldsymbol{0}_{i-1}^t & \alpha_i \end{pmatrix}$$

with $\beta_{i-1} = (1+\alpha_i)/(v+n-2) = 1/(v+i-3)$. Solving these inductively yields

$$h_i = E_*[\{\boldsymbol{\Theta}^t\boldsymbol{\Theta}\}_{ii}] = \alpha_i \prod_{j=i}^{p-1}(1+\beta_j) + \sum_{j=i}^{p-1} \beta_j \prod_{k=j+1}^{p} (1+\beta_k) \quad (i = 1, \ldots, p-1),$$
$$h_p = E_*[\{\boldsymbol{\Theta}^t\boldsymbol{\Theta}\}_{pp}] = \alpha_p,$$

where $\beta_p = 0$. It is observed that

$$\prod_{j=i}^{p-1}(1+\beta_j) = \frac{v+p-2}{v+i-2}$$

and

$$\begin{aligned}\sum_{j=i}^{p-1} \beta_j \prod_{k=j+1}^{p} (1+\beta_k) &= \sum_{j=i}^{p-1} \frac{v+p-2}{(v+j-2)(v+j-1)} \\ &= (v+p-2)\sum_{j=i}^{p-1}\left(\frac{1}{v+j-2} - \frac{1}{v+j-1}\right) \\ &= \frac{p-i}{v+i-2},\end{aligned}$$

which gives h_i in Lemma 2. This completes the proof of Lemma 2.

A.3 Proof of Theorem 5

Without loss of generality, we take $c = 1$. Let $P = \{\boldsymbol{\Theta} \in \mathcal{T}^+ : |\boldsymbol{\Theta}|^2 \geq 1\} = \{\boldsymbol{\Theta} \in \mathcal{T}^+ : \prod_{i=1}^{p} \theta_{ii}^2 \geq 1\}$. Define

$$P_k = \Big\{\boldsymbol{\Theta} \in \mathcal{T}^+ : 1 \leq \prod_{i=1}^{p} \theta_{ii} \leq k^{2p}, \quad 1/k \leq \theta_{ii} \leq k \quad (i = 2, \ldots, p), \quad \text{and}$$
$$-k^{c_{ij}}\theta_{ii} \leq \theta_{ij} \leq k^{c_{ij}}\theta_{ii} \quad (1 \leq j < i \leq p)\Big\}$$

with $c_{ij} = 3(i-j) - 1$. Note that $\cup_{k=1}^{\infty} P_k = P$. It is observed that

$$V(P_k) = \int_{P_k} \gamma(\mathrm{d}\boldsymbol{\Theta}) = p2^{p(p+1)/2}(\log k)^p \prod_{i=1}^{p}\prod_{j=1}^{i-1} k^{c_{ij}},$$

where $\gamma(\mathrm{d}\boldsymbol{\Theta}) = (\prod_{i=1}^{p} \theta_{ii}^{-i})\mathrm{d}\boldsymbol{\Theta}$.

Define the sequence of prior distributions as

$$\pi_k(\boldsymbol{\Theta})\mathrm{d}\boldsymbol{\Theta} = \frac{\gamma(\mathrm{d}\boldsymbol{\Theta})}{V(P_k)} I(\boldsymbol{\Theta} \in P_k), \quad k = 1, 2, \ldots.$$

The resulting Bayes estimators are denoted by

$$\boldsymbol{\delta}_k^{\pi} = \boldsymbol{\delta}_k^{\pi}(\boldsymbol{T}) = \arg\min_{\boldsymbol{\delta}} \int_{\mathbf{Z}\in P_k} L(\mathbf{Z}\boldsymbol{\delta}\mathbf{Z}^t) f_{\phi}(\boldsymbol{T}|\mathbf{Z})\pi_k(\mathbf{Z})\mathrm{d}\mathbf{Z}$$

with the Bayes risks

$$r_k(\pi_k, \boldsymbol{\delta}_k^{\pi}) = \frac{1}{V(P_k)} \int_{\boldsymbol{\Theta}\in P_k} \int_{\mathcal{T}^+} L(\boldsymbol{\Theta}\delta_k^{\pi}(\boldsymbol{T})\boldsymbol{\Theta}^t) f_{\phi}(\boldsymbol{T}|\boldsymbol{\Theta})\gamma(\mathrm{d}\boldsymbol{T})\pi_k(\boldsymbol{\Theta})\mathrm{d}\boldsymbol{\Theta}.$$

If it can be proved that $\lim_{k\to\infty} r_k(\pi_k, \boldsymbol{\delta}_k^{\pi}) = R_{\phi}$, where $R_{\phi} = R(\boldsymbol{\delta}_{\phi}^{BE}, \boldsymbol{\Sigma})$, then we obtain the result of Theorem 5. Since the proof of Theorem 5 is very similar to those of Theorems 1 and 3, we only show a convergence corresponding to $\boldsymbol{\delta}_k^{\pi}$, as given in Lemma 1.

A set of functions

$$\xi_{11} = \frac{1}{p}\sum_{i=1}^{p} \frac{\log\theta_{ii}}{\log k} - 1, \quad \xi_{ii} = \frac{\log\theta_{ii}}{\log k} \ \ (i = 2, \ldots, p),$$
$$\xi_{ij} = \frac{\theta_{ij}}{k^{c_{ij}}\theta_{ii}} \ \ (1 \le j < i \le p)$$

is denoted by $\boldsymbol{\xi} = \varphi_k(\boldsymbol{\Theta})$, which implies that $\varphi_k(P_k) = [-1, 1]^{p(p+1)/2}$. It then follows that

$$\gamma_k(\mathrm{d}\boldsymbol{\xi}) = p(\log k)^p \left(\prod_{i=1}^{p}\prod_{j=1}^{i-1} k^{c_{ij}}\right) \mathrm{d}\boldsymbol{\xi} = \gamma(\mathrm{d}\boldsymbol{\Theta})$$

for $\gamma(\mathrm{d}\boldsymbol{\Theta}) = (\prod_{i=1}^{p} \theta_{ii}^{-i})\mathrm{d}\boldsymbol{\Theta}$.

Replacing the θ_{ij} in P_k by the $\{\boldsymbol{Y}\boldsymbol{\Theta}\}_{ij}$, we obtain

$$P_k'(\boldsymbol{\Theta}) = \Big\{ \boldsymbol{Y} \in \mathcal{T}^+ : 1 \le \prod_{i=1}^{p} \{\boldsymbol{Y\Theta}\}_{ii} \le k^{2p},\ 1/k \le \{\boldsymbol{Y\Theta}\}_{ii} \le k \text{ for } i = 2, \ldots, p,$$
$$\text{and } - k^{c_{ij}} \{\boldsymbol{Y\Theta}\}_{ii} \le \{\boldsymbol{Y\Theta}\}_{ij} \le k^{c_{ij}} \{\boldsymbol{Y\Theta}\}_{ii} \text{ for } i > j \Big\}.$$

The function $\boldsymbol{\xi} = \varphi_k(\boldsymbol{\Theta})$ implies that $\theta_{11} = k^{p(1+\xi_{11}) - \sum_{\ell=2}^{p} \xi_{\ell\ell}}$, $\theta_{ii} = k^{\xi_{ii}}$ for $i = 2, \ldots, p$, and $\theta_{ij} = \xi_{ij} k^{c_{ij} + \xi_{ii}}$ for $i > j$. The intervals "$1 \le \prod_{i=1}^{p} \{\boldsymbol{Y\Theta}\}_{ii} \le k^{2p}$" and "$1/k \le \{\boldsymbol{Y\Theta}\}_{ii} \le k$" are equivalent to, respectively,

$$k^{-p(1+\xi_{11})} \prod_{\ell=2}^{p} y_{\ell\ell}^{-1} \le y_{11} \le k^{p(1-\xi_{11})} \prod_{\ell=2}^{p} y_{\ell\ell}^{-1},$$
$$k^{-(1+\xi_{ii})} \le y_{ii} \le k^{1-\xi_{ii}} \quad (i = 2, \ldots, p).$$

Hence, $P_k'(\boldsymbol{\Theta})$ becomes

$$\tilde{P}_k(\boldsymbol{\xi}) = \{\boldsymbol{Y} \in \mathcal{T}^+ : L_{ij}(\boldsymbol{Y}, \boldsymbol{\xi}) \le y_{ij} \le U_{ij}(\boldsymbol{Y}, \boldsymbol{\xi}) \text{ for } 1 \le j \le i \le p\},$$

where

$$L_{ij}(\boldsymbol{Y}, \boldsymbol{\xi}) = \begin{cases} k^{-p(1+\xi_{11})} \prod_{\ell=2}^{p} y_{\ell\ell}^{-1} & \text{if } i = j = 1, \\ k^{-(1+\xi_{ii})} & \text{if } i = j \ge 2, \\ -y_{ii} k^{c_{ij} + \xi_{ii} - \xi_{jj}} & \\ \quad - \sum_{m=j+1}^{i} y_{im} \xi_{mj} k^{c_{mj} + \xi_{mm} - \xi_{jj}} & \text{if } 1 \le j < i \le p, \end{cases}$$

$$U_{ij}(\boldsymbol{Y}, \boldsymbol{\xi}) = \begin{cases} k^{p(1-\xi_{11})} \prod_{\ell=2}^{p} y_{\ell\ell}^{-1} & \text{if } i = j = 1, \\ k^{1-\xi_{ii}} & \text{if } i = j \ge 2, \\ y_{ii} k^{c_{ij} + \xi_{ii} - \xi_{jj}} & \\ \quad - \sum_{m=j+1}^{i} y_{im} \xi_{mj} k^{c_{mj} + \xi_{mm} - \xi_{jj}} & \text{if } 1 \le j < i \le p. \end{cases}$$

The same arguments as in the proof of Lemma 1 yield that $\tilde{P}_k(\boldsymbol{\xi}) \supset P_k^* = \{\boldsymbol{Y} \in \mathcal{T}^+ : L_{ij}^* < y_{ij} < U_{ij}^*\ (1 \le j \le i \le p)\}$, where, for a small enough $\varepsilon > 0$,

$$L_{ij}^* = \begin{cases} k^{-\varepsilon} & \text{if } i = j, \\ -\varepsilon (1 - k^{-1})^{i-j-1} k^{\varepsilon} & \text{if } 1 \le j < i \le p, \end{cases}$$
$$U_{ij}^* = \begin{cases} k^{\varepsilon} & \text{if } i = j, \\ \varepsilon (1 - k^{-1})^{i-j-1} k^{\varepsilon} & \text{if } 1 \le j < i \le p. \end{cases}$$

Then, we observe that $\cup_{k=1}^{\infty} P_k^* = \mathcal{T}^+$. This implies that, for $\boldsymbol{\xi} \in I_\varepsilon$,

$$\boldsymbol{\delta}_k^*(\boldsymbol{L}|\boldsymbol{\xi}) = \arg\min_{\boldsymbol{\delta}} \int_{\boldsymbol{Y}\in\tilde{P}_k(\boldsymbol{\xi})} L(\boldsymbol{Y}\boldsymbol{\delta}\boldsymbol{Y}^t) f_\phi(\boldsymbol{L}|\boldsymbol{Y})\pi_k(\boldsymbol{Y})\mathrm{d}\boldsymbol{Y} \to \boldsymbol{\delta}_\phi^{BE}(\boldsymbol{L})$$

as $k \to \infty$. Therefore, combining the above convergence and the same arguments as in the proof of Theorem 1 leads to the proof of Theorem 5.

Acknowledgments The authors are deeply grateful to the two reviewers for their helpful comments and suggestions on this paper. The first author acknowledges support in part from Grant-in-Aid for Scientific Research (23500361), Japan. The second author acknowledges support from Grant-in-Aid for Scientific Research (21540114 and 23243039), Japan.

References

Berger, J. O. (1985). *Statistical decision theory and Bayesian analysis* (2nd ed.). New York: Springer.

Bondar, J. V., Milnes, P. (1981). Amenability: A survey for statistical applications of Hunt–Stein and related conditions on groups. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, *57*, 103–128.

Calvin, J. A., Dykstra, R. L. (1991). Maximum likelihood estimation of a set of covariance matrices under Löwner order restrictions with applications to balanced multivariate variance components models. *Annals of Statistics*, *19*, 850–869.

Eaton, M. L., Olkin, I. (1987). Best equivariant estimators of a Cholesky decomposition. *Annals of Statistics*, *15*, 1639–1650.

Girshick, M. A., Savage, L. J. (1951). Bayes and minimax estimates for quadratic loss functions. *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley, pp. 53–74.

James, W., Stein, C. (1961). Estimation with quadratic loss. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1*, University of California Press, Berkeley, pp. 361–379.

Kiefer, J. (1957). Invariance, minimax sequential estimation, and continuous time processes. *Annals of Mathematical Statistics*, *28*, 573–601.

Krishnamoorthy, K., Gupta, A. K. (1989). Improved minimax estimation of a normal precision matrix. *Canadian Journal of Statistics*, *17*, 91–102.

Kubokawa, T. (2004). Minimaxity in estimation of restricted parameters. *Journal of the Japan Statistical Society*, *34*, 229–253.

Kubokawa, T., Marchand, E., Strawderman, W. E., Turcotte, J.-P. (2013). Minimaxity in predictive density estimation with parametric constraints. *Journal of Multivariate Analysis*, *116*, 382–397.

Marchand, E., Strawderman, W. (2005a). Improving on the minimum risk equivariant estimator of a location parameter which is constrained to an interval or a half-interval. *Annals of the Institute of Statistical Mathematics*, *57*, 129–143.

Marchand, E., Strawderman, W. (2005b). On improving on the minimum risk equivariant estimator of a scale parameter under a lower-bound constraint. *Journal of Statistical Planning and Inference*, *134*, 90–101.

Marchand, E., Strawderman, W. (2012). A unified minimax result for restricted parameter spaces. *Bernoulli*, *18*, 635–643.

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

Pourahmadi, M. (1999). Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation. *Biometrika*, *86*, 677–690.

Robertson, T., Wright, F. T., Dykstra, R. L. (1988). *Order restricted statistical inference*. New York: Wiley.

Rockafellar, R. T. (1970). *Convex analysis*. Princeton: Princeton University Press.

Selliah, J. B. (1964). Estimation and testing problems in a Wishart distribution, Technical reports No.10, Department of Statistics, Stanford University, Stanford.

Stein, C. (1956). Some problems in multivariate analysis, Part I, Technical reports No.6, Department of Statistics, Stanford University, Stanford.

Strawderman, W. E. (2000). Minimaxity. *Journal of the American Statistical Association*, *95*, 1364–1368.

Tsukuma, H., Kubokawa, T. (2011). Modifying estimators of ordered positive parameters under the Stein loss. *Journal of Multivariate Analysis*, *102*, 164–181.

Tsukuma, H., Kubokawa, T. (2012). Minimaxity in estimation of restricted and non-restricted scale parameter matrices, Discussion Paper Series, CIRJE-F-858. Faculty of Economics, The University of Tokyo.