Minimaxity in Estimation of Restricted and Non-restricted Scale Parameter Matrices

Hisayuki Tsukuma
Toho University

Tatsuya Kubokawa
University of Tokyo

September 2012

Minimaxity in Estimation of Restricted and Non-restricted Scale Parameter Matrices

Hisayuki Tsukuma[*] and Tatsuya Kubokawa[†]

September 4, 2012

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. In this paper, we address the classical problem and succeed in construction of such a sequence, which establishes minimaxity of the best equivariant estimator. We also derive unified conditions for a sequence of prior distributions to be least favorable in the general estimation problem with an invariance structure. These unified conditions are applied to both restricted and non-restricted cases of parameters, and we give a couple of examples which show minimaxity of the best equivariant estimators under restrictions of the covariance matrix.

AMS 2010 subject classifications: Primary 62C20; secondary 62F10, 62C10.

Key words and phrases: 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)).

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).

[*]Faculty of Medicine, Toho University, 5-21-16 Omori-nishi, Ota-ku, Tokyo 143-8540, Japan, E-Mail: tsukuma@med.toho-u.ac.jp

[†]Faculty of Economics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan, E-Mail: tatsuya@e.u-tokyo.ac.jp

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 minimaxity of the best equivariant and unrestricted estimators even in the case of restricted location and scale parameters.

Although the invariance approach is quite general and hard to understand, the least favorable approach gives an understandable proof for minimaxity of the best equivariant estimator. 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 one of the most interesting issues 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. In Section 2, we describe exactly the general estimation problem with an invariance structure which is similar to Hora and Buehler (1966). Using the method given in Girshick and Savage (1951) and Kubokawa (2004) based on a sequence of prior distributions, we derive sufficient and unified conditions for minimaxity of the best equivariant estimator, which can be applied to restricted and non-restricted cases of parameters.

Section 3 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 Section 3, and this sequence is shown to satisfy the general conditions for mimimiaxity given in Section 2. These results are extended to a class of elliptical distribution models including the matrix-variate F distribution.

Section 4 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 general result given in Section 2. It is shown that the best equivariant estimator is further improved on by the isotonic regression method.

The general conditions given in Section 2 and the arguments given in Sections 3 and 4 have the potential to apply to various restrictions of covariance and precision matrices. In Section 5, we provide an example of restriction based on determinant of covariance matrix, and state concluding remarks. Some proofs are given in the appendix.

2 General Conditions for Minimaxity by Girshick-Savage's Method

In this section, we formulate the general estimation problem with an invariance structure and derive sufficient and unified conditions for minimaxity of the best equivariant estimator.

Let X be an observable random variable. Let $(\mathcal{X}, \mathcal{B}_X)$ be a measurable space of X and $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ be a family of identifiable probability measures with parameter space Θ. We assume the following conditions.

(A1) There exist a group $\mathcal{G} = \{g\}$ and a measurable space $(\mathcal{G}, \mathcal{B}_G)$ on which there exists a left invariant Haar measure γ satisfying

$$\gamma(gG) = \gamma(G) \quad \text{for all } g \in \mathcal{G} \text{ and all } G \in \mathcal{B}_G.$$

Each $g \in \mathcal{G}$ induces a one-to-one transformation $\overline{g}$ from Θ onto itself defined by $P_{\overline{g}\theta}(gA) = P_\theta(A)$ for any $A \in \mathcal{B}_X$ and any $\theta \in \Theta$. The induced space $\overline{\mathcal{G}} = \{\overline{g} : \overline{g} \in \mathcal{G}\}$ is measurable.

(A2) There exists a one-to-one correspondence $X \leftrightarrow (T_x, U_x)$ between $\mathcal{X}$ and $\mathcal{G} \times \mathcal{U}_X$ such that gX corresponds to (gT_x, U_x) and $\mathcal{U}_X$ is a measurable space. The statistic U_x is maximal invariant under the transformation $\mathcal{G}$.

(A3) There exists a one-to-one correspondence $\theta \leftrightarrow \overline{g}_\theta$ between Θ and $\overline{\mathcal{G}}$ such that $\overline{g}\theta$ corresponds to $\overline{g}\,\overline{g}_\theta$ for all $\overline{g} \in \overline{\mathcal{G}}$. The correspondence of $\overline{g}_\theta$ in $\mathcal{G}$ is denoted by g_θ.

(A4) There exist a conditional probability density function $p(g_\theta^{-1}t_x|u_x)$ given $U_x = u_x$ such that for all $A \in \mathcal{B}_X$,

$$P_\theta[A] = \int_A p(g_\theta^{-1}t_x|u_x)p_x(u_x)\gamma(\mathrm{d}t_x)\gamma_x(\mathrm{d}u_x),$$

where $p_x(\cdot)$ is a marginal density function of U_x with respect to a measure $\gamma_x(\cdot)$ on $\mathcal{U}_X$.

Under these assumptions, we consider estimation of g_θ under the loss function $\ell(g_\theta, d) = L(g_\theta^{-1}d)$, which is invariant under the transformation $g_\theta \to gg_\theta$ and $d \to gd$. Any equivariant estimator $\delta(T_x|U_x)$ satisfies that $\delta(gT_x|U_x) = g\delta(T_x|U_x)$ for any $g \in \mathcal{G}$. The best equivariant estimator of g_θ is

$$\delta^{BE}(T_x|U_x) = \arg\min_\delta \int_\mathcal{G} L(g_1^{-1}\delta)p(g_1^{-1}T_x|U_x)\nu(\mathrm{d}g_1), \tag{2.1}$$

where $\nu(\cdot)$ is the measure defined by

$$\nu(\mathrm{d}g) = \gamma(\mathrm{d}g^{-1}).$$

This is a right invariant Haar measure. Since $\gamma(\cdot)$ is left invariant, it is noted that $\gamma(h\mathrm{d}g) = \gamma(\mathrm{d}g)$ and $\gamma((\mathrm{d}g)h) = \Delta(h)\gamma(\mathrm{d}g)$ for $h, g \in \mathcal{G}$, where $\Delta(\cdot)$ is a modular function.

The best equivariant estimator is the generalized Bayes estimator against the right invariant measure $\nu(\mathrm{d}g)$. Kiefer (1957) showed that the best equivariant estimator is *minimax* if the group $\mathcal{G}$ is *amenable*, namely, if there is a sequence of probability measures $\gamma_j(\cdot)$

3

on $\mathcal{G}$ that is asymptotically invariant in the sense that $\lim_{j\to\infty} \int \{\psi(ag) - \psi(a)\} \gamma_j(\mathrm{d}a) = 0$ for every $g \in \mathcal{G}$ and every bounded measurable function ψ on $\mathcal{G}$. However, the best equivariant estimator is not necessarily minimax when the parameter space is restricted. Kubokawa, *et al.* (2012) used this general framework for predicting the predictive density function, and showed minimaxity of the generalized Bayes predictive density function under restriction of parameters based on the arguments of Girshick and Savage (1951).

We here show minimaxity of the best equivariant estimator based on the method of Girshick and Savage (1951) in the context of point estimation in both cases of restricted and non-restricted spaces. To this end, we provide the following unified conditions for minimaxity.

(A5) In the case that Θ is restricted, assume that the restriction is equivalently expressed as $g_\theta \in P$. Also, it is assumed that $P \subset \mathcal{G} \subset \mathbb{R}^r$; namely, $\mathcal{G}$ is a subset of r dimensional Euclidean space $\mathbb{R}^r$ and P is a restricted space of $\mathcal{G}$. In the case that Θ is not restricted, P is identical to $\mathcal{G}$.

(A6) There exist a sequence of subsets $P_k (\subset P)$ and one-to-one functions $\varphi_k(\cdot)$ between $P_k \leftrightarrow \Xi \subset \mathbb{R}^r$ with $\xi = \varphi_k(g_\theta)$ for $g_\theta \in P_k$ where P_k, $\varphi_k(\cdot)$ and Ξ satisfy the following conditions:

(A6-1) $\cup_{k=k_0}^{\infty} P_k = P$ for some $k_0 \geq 1$.

(A6-2) Let $V(P_k) = \int_{P_k} \nu(\mathrm{d}g_\theta)$. Let $\gamma_k(\cdot)$ be an induced measure defined by $\gamma_k(A) = \nu(\varphi_k^{-1}(A))$ for $A \in \Xi$. Then, $\varphi_k(P_k) = \Xi = \prod_{i=1}^{r}[-1 + a_{i,k}, 1 + b_{i,k}]$ and

$$\int_{\varphi_k(P_k)} f(\xi)\gamma_k(\mathrm{d}\xi)/V(P_k) \geq \frac{1}{2^r + c_k} \int I\Big(\xi \in \prod_{i=1}^{r}[-1 + a_{i,k}, 1 + b_{i,k}]\Big)f(\xi)\mathrm{d}\xi, \qquad (2.2)$$

where $f(\cdot) \geq 0$, $I(\cdot)$ is the indicator function, and $\lim_{k\to\infty} a_{i,k} = \lim_{k\to\infty} b_{i,k} = \lim_{k\to\infty} c_k = 0$ for $i = 1, \ldots, r$.

(A6-3) Assume that $g_\theta g \in P_k$ is equivalent to $[\varphi_k^{-1}(\xi)]g \in P_k$ or $g \in \tilde{P}_k(\xi)$. For any small enough $\varepsilon > 0$ and any $\xi \in \prod_{i=1}^{r}[-1 + a_{i,k} + \varepsilon, 1 + b_{i,k} - \varepsilon]$, there exists a sequence of subsets P_k^* such that P_k^* does not depend on ξ, $P_k^* \subset \tilde{P}_k(\xi)$ and $\cup_{k=k_1}^{\infty} P_k^* = \mathcal{G}$ for some $k_1 \geq 1$.

Theorem 2.1 *Assume conditions from* **(A1)** *to* **(A6)**. *Then, the best equivariant estimator δ^{BE} is minimax.*

Proof. We can show this theorem along the same lines as in Kubokawa (2004) and Kubokawa *et al.* (2012) who modified the method of Girshick and Savage (1951). Consider the sequence of prior distributions given by

$$\pi_k(g_\theta)\nu(\mathrm{d}g_\theta) = \begin{cases} \{V(P_k)\}^{-1}\nu(\mathrm{d}g_\theta) & \text{if } g_\theta \in P_k \\ 0 & \text{otherwise.} \end{cases}$$

This yields the Bayes estimators

$$\delta_k^\pi(t_x|u_x) = \arg\min_\delta \int_{g_1 \in P_k} L(g_1^{-1}\delta)p(g_1^{-1}t_x|u_x)\nu(\mathrm{d}g_1),$$

4

with conditional Bayes risks

$$r_k(\pi_k, \delta_k^\pi | u_x) = \frac{1}{V(P_k)} \int_{g_\theta \in P_k} \int L(g_\theta^{-1} \delta_k^\pi(t_x | u_x)) p(g_\theta^{-1} t_x | u_x) \gamma(\mathrm{d}t_x) \nu(\mathrm{d}g_\theta).$$

It is easy to check that $r_k(\pi_k, \delta_k^\pi | u_x) \leq r_k(\pi_k, \delta^{BE} | u_x) = R_0(u_x)$, so it is sufficient to show that $\liminf_{k \to \infty} r_k(\pi_k, \delta_k^\pi | u_x) \geq R_0(u_x)$, where $R_0(u_x) = \int L(\delta^{BE}(t_x | u_x)) p(t_x | u_x) \gamma(\mathrm{d}t_x)$. Making the transformation $s_x = g_\theta^{-1} t_x$ yields

$$r_k(\pi_k, \delta_k^\pi | u_x) = \frac{1}{V(P_k)} \int_{g_\theta \in P_k} \int L(g_\theta^{-1} \delta_k^\pi(g_\theta s_x | u_x)) p(s_x | u_x) \gamma(\mathrm{d}s_x) \nu(\mathrm{d}g_\theta), \qquad (2.3)$$

where $\delta_k^\pi(g_\theta s_x | u_x)$ is expressed as

$$\delta_k^\pi(g_\theta s_x | u_x) = \arg\min_\delta \int_{g_1 \in P_k} L(g_1^{-1} \delta) p(g_1^{-1} g_\theta s_x | u_x) \nu(\mathrm{d}g_1).$$

Now, make the transformation $g_2 = g_\theta^{-1} g_1$ with $\nu(\mathrm{d}g_1) = \Delta(g_\theta) \nu(\mathrm{d}g_2)$. Then,

$$\delta_k^\pi(g_\theta s_x | u_x) = \arg\min_\delta \int_{g_\theta g_2 \in P_k} L(g_2^{-1} g_\theta^{-1} \delta) p(g_2^{-1} s_x | u_x) \Delta(g_\theta) \nu(\mathrm{d}g_2),$$

namely, we have $g_\theta^{-1} \delta_k^\pi(g_\theta s_x | u_x) = \delta_k^*(s_x | u_x, g_\theta)$ where

$$\delta_k^*(s_x | u_x, g_\theta) = \arg\min_\delta \int_{g_\theta g_2 \in P_k} L(g_2^{-1} \delta) p(g_2^{-1} s_x | u_x) \nu(\mathrm{d}g_2).$$

In view of the assumptions, there exists a transformation $\xi = \varphi_k(g_\theta)$ satisfying the condition **(A6)**. Note that $g_\theta g_2 \in P_k$ is equivalent to $\varphi_k^{-1}(\xi) g_2 \in P_k$, or

$$g_2 \in \{[\varphi_k^{-1}(\xi)]^{-1} g; g \in P_k\} \equiv \tilde{P}_k(\xi).$$

Then, the Bayes estimator $\delta_k^*(s_x | u_x, g_\theta)$ is rewritten as

$$\delta_k^*(s_x | u_x, \varphi_k^{-1}(\xi)) = \arg\min_\delta \int_{g_2 \in \tilde{P}_k(\xi)} L(g_2^{-1} \delta) p(g_2^{-1} s_x | u_x) \nu(\mathrm{d}g_2),$$

and the conditional Bayes risk (2.3) is rewritten as

$$r_k(\pi_k, \delta_k^\pi | u_x) = \frac{1}{V(P_k)} \int_{\xi \in \varphi_k(P_k)} \int L(\delta_k^*(s_x | u_x, \varphi_k^{-1}(\xi))) p(s_x | u_x) \gamma(\mathrm{d}s_x) \gamma_k(\mathrm{d}\xi).$$

It is noted that from **(A6-2)**, for any small $\varepsilon > 0$,

$$\varphi_k(P_k) = \prod_{i=1}^r [-1 + a_{i,k}, 1 + b_{i,k}] \supset \prod_{i=1}^r [-1 + a_{i,k} + \varepsilon, 1 + b_{i,k} - \varepsilon] \equiv I_{k,\varepsilon}.$$

Then from (2.2), the conditional Bayes risk is evaluated as

$$r_k(\pi_k, \delta_k^\pi | u_x) \geq \frac{1}{2^r} \int I(\xi \in I_{k,\varepsilon}) \int L(\delta_k^*(s_x | u_x, \varphi_k^{-1}(\xi))) p(s_x | u_x) \gamma(\mathrm{d}s_x) \mathrm{d}\xi.$$

5

For $\xi \in I_{k,\varepsilon}$, from **(A6-3)**, it can be seen that $\delta_k^*(s_x|u_x, \varphi_k^{-1}(\xi)) \to \delta^{BE}(s_x|u_x)$ as $k \to \infty$. Hence, Fatou's lemma is used to bound the Bayes risks as

$$\liminf_{k\to\infty} r_k(\pi_k, \delta_k^\pi|u_x)$$

$$\geq \frac{1}{2^r} \int \liminf_{k\to\infty} I(\xi \in I_{k,\varepsilon}) \int p(s_x|u_x) \cdot \liminf_{k\to\infty} L(\delta_k^*(s_x|u_x, \varphi_k^{-1}(\xi)))\gamma(\mathrm{d}s_x)\mathrm{d}\xi$$

$$= \frac{1}{2^r} \int_{|-1+\varepsilon,1-\varepsilon|^r} \mathrm{d}\xi \int p(s_x|u_x)L(\delta^{BE}(s_x|u_x))\gamma(\mathrm{d}s_x)$$

$$= (1-\varepsilon)^r R(\theta, \delta^{BE}(t_x|u_x)) = (1-\varepsilon)^r R_0(u_x)$$

From the arbitrariness of $\varepsilon > 0$, it follows that $\liminf_{k\to\infty} r_k(\pi_k, \delta_k^\pi|u_x) \geq R_0(u_x)$, completing the proof of Theorem 2.1. ∎

In some examples, handling g_θ^{-1} as a parameter is more convenient than treating g_θ. Namely, in assumption (A4), consider the conditional probability density function $p(h_\theta t_x|u_x)$ and the probability

$$P_\theta[A] = \int_A p(h_\theta t_x|u_x)p_x(\mathrm{d}t_x)\gamma(\mathrm{d}t_x)\gamma_x(\mathrm{d}u_x),$$

where $h_\theta = g_\theta^{-1}$. In this case, the best equivariant estimator given in (2.1) is

$$\delta^{BE}(T_x|U_x) = \arg\min_\delta \int_{\mathcal{G}} L(h_1\delta)p(h_1 T_x|U_x)\gamma(\mathrm{d}h_1),$$

for the left invariant measure $\gamma(\cdot)$. The minimaxity of the best equivariant estimator under restriction can be similarly shown by modifying assumptions (A5) and (A6) as follows:

(A5′) In the case that Θ is restricted, assume that the restriction is equivalently expressed as $h_\theta \in P$. Also, it is assumed that $P \subset \mathcal{G} \subset \mathbb{R}^r$; namely, $\mathcal{G}$ is a subset of r dimensional Euclidean space $\mathbb{R}^r$ and P is a restricted space of $\mathcal{G}$. In the case that Θ is not restricted, P is identical to $\mathcal{G}$.

(A6′) There exist a sequence of subsets P_k and one-to-one functions $\varphi_k(\cdot)$ between $P_k \leftrightarrow \Xi \subset \mathbb{R}^r$ with $\xi = \varphi_k(h_\theta)$ for $h_\theta \in \mathcal{G}$ such that P_k and $\varphi_k(\cdot)$ satisfy the following conditions:

(A6-1′) $\cup_{k=k_0}^\infty P_k = P$ for some $k_0 \geq 1$.

(A6-2′) Let $V(P_k) = \int_{P_k} \gamma(\mathrm{d}h_\theta)$. Let $\gamma_k(\cdot)$ be an induced measure defined by $\gamma_k(A) = \gamma(\varphi_k^{-1}(A))$ for $A \in \Xi$. Then, $\varphi_k(P_k) = \Xi = \prod_{i=1}^r[-1 + a_{i,k}, 1 + b_{i,k}]$ and

$$\int_{\varphi_k(P_k)} f(\xi)\gamma_k(\mathrm{d}\xi)/V(P_k) \geq \frac{1}{2^r + c_k} \int I\Big(\xi \in \prod_{i=1}^r[-1 + a_{i,k}, 1 + b_{i,k}]\Big)f(\xi)\mathrm{d}\xi,$$

where $f(\cdot) \geq 0$, $I(\cdot)$ is the indicator function, and $\lim_{k\to\infty} a_{i,k} = \lim_{k\to\infty} b_{i,k} = \lim_{k\to\infty} c_k = 0$ for $i = 1, \ldots, r$.

(A6-3′) Assume that $hh_\theta \in P_k$ is equivalent to $h \in \tilde{P}_k(\xi)$. For any small enough $\varepsilon > 0$ and any $\xi \in \prod_{i=1}^r [-1 + a_{i,k} + \varepsilon, 1 + b_{i,k} - \varepsilon]$, there exists a sequence of subsets P_k^* such that P_k^* does not depend on ξ, $P_k^* \subset \tilde{P}_k(\xi)$ and $\cup_{k=k_1}^\infty P_k^* = \mathcal{G}$ for some $k_1 \geq 1$.

Under assumptions **(A1)**-**(A4)**, **(A5′)** and **(A6′)**, we can check each step in the proof of Theorem 2.1 by replacing g_θ^{-1}, g_1^{-1}, g_2^{-1}, $\nu(\cdot)$, $g_2 = g_\theta^{-1} g_1$ and $g_\theta g_2 \in P_k$ with h_θ, h_1, h_2, $\gamma(\cdot)$, $h_2 = h_1 h_\theta^{-1}$ and $h_2 h_\theta \in P_k$, respectively. Then, it can be verified that the best equivariant estimator is minimax.

Corollary 2.1 *Assume conditions* **(A1)**-**(A4)**, **(A5′)** *and* **(A6′)**. *Then, the best equivariant estimator δ^{BE} is minimax.*

3   A Least Favorable Sequence of Prior Distributions in Estimation of Covariance Matrix

We now construct a least favorable sequence of prior distributions in estimation of the normal covariance and precision matrices. The problem of finding such a sequence for the normal covariance matrix has been an open question for a long time since Stein (1956) and James and Stein (1961). Thus, the derived sequence, which satisfies the general conditions given in Theorem 2.1, brings a solution to one of the most interesting issues in statistical decision theory.

3.1   Estimation of normal covariance and precision matrices

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

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

where C is a normalizing constant and $\gamma(\mathrm{d}\mathbf{T}) = (\prod_{i=1}^p t_{ii}^{-i})\mathrm{d}\mathbf{T}$, which is left-invariant measure on $\mathcal{T}^+$. Denote a loss function by $L(\mathbf{\Theta}\boldsymbol{\delta}\mathbf{\Theta}^t)$, where $\boldsymbol{\delta}$ is an estimator of $\mathbf{\Sigma}$.

For all $\mathbf{A} \in \mathcal{T}^+$, the group transformation with respect to $\mathcal{T}^+$ on a random matrix $\mathbf{T}$ and a parameter matrix $\mathbf{\Theta}$ is defined by $(\mathbf{T}, \mathbf{\Theta}) \to (\mathbf{A}\mathbf{T}, \mathbf{\Theta}\mathbf{A}^{-1})$. Then the equivariant estimator with respect to the group $\mathcal{T}^+$ has the form $\mathbf{T}\mathbf{D}\mathbf{T}^t$, where $\mathbf{D}$ is a diagonal matrix independent of $\mathbf{T}$. The best diagonal matrix $\mathbf{D}$ which yields the best equivariant estimator $\boldsymbol{\delta}^{BE}$, say, depends on the loss function L.

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

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

Then, we consider the sequence of prior distributions given by

$$\pi_k(\mathbf{\Theta})\mathrm{d}\mathbf{\Theta} = \frac{\gamma(\mathrm{d}\mathbf{\Theta})}{V(P_k)}I(\mathbf{\Theta} \in \boldsymbol{P}_k), \quad k = 1, 2, \ldots, \tag{3.1}$$

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 following theorem shows that the sequence (3.1) is least favorable.

Theorem 3.1 *The best equivariant estimator $\boldsymbol{\delta}^{BE}$ is minimax.*

Proof. The minimaxity of the best equivariant estimator of $\boldsymbol{\Sigma}$ can be shown by checking the conditions (A6-1′)-(A6-3′) of Corollary 2.1 for $P = \mathcal{G} = \mathcal{T}^+$. It holds that $\cup_{k=1}^{\infty} P_k = \mathcal{T}^+$, which satisfies (A6-1′). For (A6-2′), 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 $\boldsymbol{\xi} = \varphi_k(\boldsymbol{\Theta})$. Then, we obtain $\varphi_k(P_k) = [-1,1]^{p(p+1)/2}$ and

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

for $\boldsymbol{\xi} = (\xi_{11}, \xi_{21}, \xi_{22}, \ldots, \xi_{p1}, \ldots, \xi_{pp})^t$ and $\gamma(\mathrm{d}\boldsymbol{\Theta}) = (\prod_{i=1}^{p} \theta_{ii}^{-i})\mathrm{d}\boldsymbol{\Theta}$. Also, it follows that

$$\int_{\varphi_k(P_k)} f(\boldsymbol{\xi})\gamma_k(\mathrm{d}\boldsymbol{\xi})/V(P_k) = 2^{-p(p+1)/2}\int f(\boldsymbol{\xi})I(\boldsymbol{\xi}\in[-1,1]^{p(p+1)/2})\mathrm{d}\boldsymbol{\xi}.$$

Finally, we shall show condition (A6-3′). Note that "$\boldsymbol{Y}\boldsymbol{\Theta}\in P_k$" is equivalent to "$\boldsymbol{Y}\in P_k'(\boldsymbol{\Theta})$", where

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

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

$$\begin{aligned}
\tilde{P}_k(\boldsymbol{\xi}) &= \Big\{\boldsymbol{Y}\in\mathcal{T}^+ : k^{-1} < y_{ii}k^{\xi_{ii}} < k \ \ (i = 1,\ldots,p) \text{ and} \\
&\qquad - y_{ii}k^{c_{ij}+\xi_{ii}} < y_{ij}k^{\xi_{jj}} + \sum_{m=j+1}^{i} y_{im}\xi_{mj}k^{c_{mj}+\xi_{mm}} < y_{ii}k^{c_{ij}+\xi_{ii}} \ \ (i > j)\Big\} \\
&= \{\boldsymbol{Y}\in\mathcal{T}^+ : k^{-(1+\xi_{ii})} < y_{ii} < k^{1-\xi_{ii}} \ \ (i = 1,\ldots,p) \text{ and} \qquad\qquad (3.2) \\
&\qquad L_{ij}(\boldsymbol{Y},\boldsymbol{\xi}) < y_{ij} < U_{ij}(\boldsymbol{Y},\boldsymbol{\xi}) \ (i > j)\},
\end{aligned}$$

with

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

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

8

Recall that $c_{ij} = 3(i - j) - 1$. Also, assume that $|\xi_{ij}| < 1 - \varepsilon$ for any $\varepsilon > 0$ and $i \geq j$. Then, ξ_{ij} satisfies that $1 - \xi_{ij} > \varepsilon$ and $1 + \xi_{ij} > \varepsilon$. For each y_{ii}, it follows from (3.2) that the set $\{y_{ii} \in \mathbb{R} : k^{-(1+\xi_{ii})} < y_{ii} < k^{1-\xi_{ii}}\}$ includes

$$\{y_{ii} \in \mathbb{R} : k^{-\varepsilon} < y_{ii} < k^\varepsilon\}. \tag{3.3}$$

As proved in Lemma 3.1 given below, we express the upper and lower bounds of y_{ij} as

$$U_{ij}(\boldsymbol{Y}, \boldsymbol{\xi}) > y_{ii}\varepsilon(1 - k^{-1})^{i-j-1}k^{c_{ij}+\xi_{ii}-\xi_{jj}}, \tag{3.4}$$

$$L_{ij}(\boldsymbol{Y}, \boldsymbol{\xi}) < -y_{ii}\varepsilon(1 - k^{-1})^{i-j-1}k^{c_{ij}+\xi_{ii}-\xi_{jj}} \tag{3.5}$$

for $i > j$. Hence from (3.3) and (3.4), it is observed that

$$\begin{aligned} U_{ij}(\boldsymbol{Y}, \boldsymbol{\xi}) &> k^{-\varepsilon}\varepsilon(1 - k^{-1})^{i-j-1}k^{3(i-j)-1-2+(1+\xi_{ii})+(1-\xi_{jj})} \\ &> k^{-\varepsilon}\varepsilon(1 - k^{-1})^{i-j-1}k^{3(i-j-1)+2\varepsilon} \\ &> \varepsilon(1 - k^{-1})^{i-j-1}k^\varepsilon \equiv U_{ij}^*, \end{aligned}$$

since $i - j - 1 \geq 0$. Similarly, using (3.3) and (3.5) gives that $L_{ij}(\boldsymbol{Y}, \boldsymbol{\xi}) < -\varepsilon(1 - k^{-1})^{i-j-1}k^\varepsilon \equiv L_{ij}^*$. Let

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

Then, $P_k^* \subset \tilde{P}_k(\boldsymbol{\xi})$ and $\cup_{k=1}^\infty P_k^* = \mathcal{T}^+ = \mathcal{G}$, which satisfies (A6-3′). Hence from Corollary 2.1, the best equivariant estimator of covariance matrix $\boldsymbol{\Sigma}$ is minimax. ∎

To complete the proof of Theorem 3.1, we need to show the following lemma.

Lemma 3.1 *For $i > j$,*

$$U_{ij}(\boldsymbol{Y}, \boldsymbol{\xi}) > y_{ii}\varepsilon(1 - k^{-1})^{i-j-1}k^{c_{ij}+\xi_{ii}-\xi_{jj}}, \tag{3.6}$$

$$L_{ij}(\boldsymbol{Y}, \boldsymbol{\xi}) < -y_{ii}\varepsilon(1 - k^{-1})^{i-j-1}k^{c_{ij}+\xi_{ii}-\xi_{jj}}. \tag{3.7}$$

Proof. We show the inequality (3.6) by induction. For $j = i - 1$, $U_{i,i-1}(\boldsymbol{Y}, \boldsymbol{\xi}) = y_{ii}(1 - \xi_{i,i-1})k^{c_{i,i-1}+\xi_{ii}-\xi_{i-1,i-1}} > y_{ii}\varepsilon k^{c_{i,i-1}+\xi_{ii}-\xi_{i-1,i-1}}$, so that the inequality (3.6) holds for $j = i - 1$. For $\ell = i - 1, \ldots, j + 1$, suppose that the inequality

$$U_{i\ell}(\boldsymbol{Y}, \boldsymbol{\xi}) > y_{ii}\varepsilon(1 - k^{-1})^{i-\ell-1}k^{c_{i\ell}+\xi_{ii}-\xi_{\ell\ell}}, \tag{3.8}$$

namely,

$$y_{i\ell} < y_{ii}\varepsilon(1 - k^{-1})^{i-\ell-1}k^{c_{i\ell}+\xi_{ii}-\xi_{\ell\ell}}. \tag{3.9}$$

Then, we verify the inequality when $\ell = j$. Using the assumption (3.8) and (3.9) of induction, we see that

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

9

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

$$U_{ij}(\boldsymbol{Y}, \boldsymbol{\xi}) > y_{ii} k^{c_{ij} + \xi_{ii} - \xi_{jj}} \left\{ \varepsilon - \sum_{m=j+1}^{i-1} \varepsilon (1 - k^{-1})^{i-m-1} k^{-1} \right\}$$

$$= y_{ii} k^{c_{ij} + \xi_{ii} - \xi_{jj}} \varepsilon \left\{ 1 - \frac{1}{k} - \left(1 - \frac{1}{k} \right) \frac{1}{k} - \left(1 - \frac{1}{k} \right)^2 \frac{1}{k} - \cdots - \left(1 - \frac{1}{k} \right)^{i-j-2} \frac{1}{k} \right\}$$

$$= y_{ii} k^{c_{ij} + \xi_{ii} - \xi_{jj}} \varepsilon \left(1 - \frac{1}{k} \right)^{i-j-1},$$

which proves the inequality (3.6). Since the inequality (3.7) can be shown similarly, we get Lemma 3.1. ∎

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

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

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

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

for $d_i = (n + p - 2i + 1)^{-1}$. From Theorem 3.1, $\boldsymbol{\delta}^{BE}$ is minimax relative to $L_S(\boldsymbol{\delta}, \boldsymbol{\Sigma})$, and the least favorable sequence of prior distributions is given in (3.1). Similarly, we use the following invariant loss functions

$$L_Q(\boldsymbol{\delta}, \boldsymbol{\Sigma}) = \operatorname{tr}(\boldsymbol{\Sigma}^{-1} \boldsymbol{\delta} - \boldsymbol{I}_p)^2 = \operatorname{tr}(\boldsymbol{\Theta} \boldsymbol{\delta} \boldsymbol{\Theta}^t - \boldsymbol{I}_p)^2,$$

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

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.

3.2   Extension to elliptical distributions

The results given in Subsection 3.1 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

$$f_\phi(\boldsymbol{T}|\boldsymbol{\Theta})\gamma(\mathrm{d}\boldsymbol{T}) = |\boldsymbol{\Theta}\boldsymbol{T}|^n \phi(\boldsymbol{\Theta}\boldsymbol{T}\{\boldsymbol{\Theta}\boldsymbol{T}\}^t)\gamma(\mathrm{d}\boldsymbol{T}), \tag{3.10}$$

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

$$\boldsymbol{\delta}_\phi^{BE} = \arg\min_{\boldsymbol{\delta}} \int L(\boldsymbol{\Theta}\boldsymbol{\delta}\boldsymbol{\Theta}^t) f_\phi(\boldsymbol{T}|\boldsymbol{\Theta})\gamma(\mathrm{d}\boldsymbol{\Theta}) = \boldsymbol{T}\boldsymbol{D}_\phi\boldsymbol{T}^t,$$

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 3.2 *Assume that the random matrix is distributed as an elliptical distribution with the p.d.f. (3.10). Then 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

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

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 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 $\boldsymbol{\Theta T}^{-1}$ gives the expression $\boldsymbol{\delta}_F^{BE} = \boldsymbol{T}\boldsymbol{D}_F^{BE}\boldsymbol{T}^t$ where

$$\boldsymbol{D}_F^{BE} = \left[\int \boldsymbol{\Theta}^t\boldsymbol{\Theta} f_F(\boldsymbol{I}_p|\boldsymbol{\Theta})\gamma(\mathrm{d}\boldsymbol{\Theta})\right]^{-1}. \tag{3.11}$$

The value of $\boldsymbol{D}_F^{BE}$ is hard to derive, but we can obtain it in the following proposition which will be proved in the appendix.

Proposition 3.1 *Suppose that $v > 2$. The exact value of $\boldsymbol{D}_F^{BE}$ defined in (3.11) 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 3.2 suggests 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).

4 Estimation under Restriction of Lower Triangular Matrix

4.1 Minimaxity under order-restricted diagonal elements

Consider the unique reparametrization for $\boldsymbol{\Sigma}$ of the form $\boldsymbol{\Gamma\Sigma\Gamma}^t = \boldsymbol{\Lambda}$, 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

$$\theta_{11}^{-2} \geq \theta_{22}^{-2} \geq \cdots \geq \theta_{pp}^{-2}, \quad \text{or, equivalently, } \theta_{11} \leq \theta_{22} \leq \cdots \leq \theta_{pp}. \tag{4.1}$$

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

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

Then, we consider the sequence of prior distributions given by

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

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 4.1 *For an elliptical distribution* (3.10) *with the restriction* (4.1), *the best equivariant and unrestricted estimator* $\boldsymbol{\delta}_\phi^{BE}$ *is minimax.*

Proof. The minimaxity is verified by checking the conditions (A6-1′)-(A6-3′) of Corollary 2.1. Let $\mathcal{G} = \mathcal{T}^+$, $P = \{\boldsymbol{\Theta} \in \mathcal{T}^+ : \theta_{11} \leq \theta_{22} \leq \cdots \leq \theta_{pp}\}$ and $P_k = P_k^L$ for $\mathcal{G}$, P and P_k in the assumptions (A5′) and (A6′). It is seen that $\cup_{k=1}^{\infty} P_k = P \subset \mathcal{T}^+$. The function $\boldsymbol{\xi} = \varphi_k(\boldsymbol{\Theta})$ is defined by

$$\xi_{11} = \frac{2\log\theta_{11}}{\log k}, \qquad \xi_{ii} = \frac{1}{\log k}\log\frac{\theta_{ii}}{\theta_{i-1,i-1}} - 1 \quad (i = 2, \ldots, p),$$

$$\xi_{ij} = \frac{\theta_{ij}}{\theta_{ii}k^{c_{ij}}} \quad (1 \leq j < i \leq p).$$

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

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

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

For (A6-3′), "$\boldsymbol{Y}\boldsymbol{\Theta} \in P_k$" is rewritten as "$\boldsymbol{Y} \in P_k'(\boldsymbol{\Theta})$", where

$$P_k'(\boldsymbol{\Theta}) = \Big\{\boldsymbol{Y} \in \mathcal{T}^+ : k^{-1/2} \leq \{\boldsymbol{Y}\boldsymbol{\Theta}\}_{11} \leq k^{1/2}, \ 1 \leq \frac{\{\boldsymbol{Y}\boldsymbol{\Theta}\}_{ii}}{\{\boldsymbol{Y}\boldsymbol{\Theta}\}_{i-1,i-1}} \leq k^2 \ (i = 2, \ldots, p)$$

$$\text{and} - k^{c_{ij}}\{\boldsymbol{Y}\boldsymbol{\Theta}\}_{ii} \leq \{\boldsymbol{Y}\boldsymbol{\Theta}\}_{ij} \leq k^{c_{ij}}\{\boldsymbol{Y}\boldsymbol{\Theta}\}_{ii} \ (1 \leq j < i \leq p)\Big\}$$

$$= \Big\{\boldsymbol{Y} \in \mathcal{T}^+ : k^{-1/2} \leq y_{11}\theta_{11} \leq k^{1/2}, \ 0 \leq \frac{y_{ii}\theta_{ii}}{y_{i-1,i-1}\theta_{i-1,i-1}} \leq k^2 \ (i = 2, \ldots, p)$$

$$\text{and} - k^{c_{ij}}y_{ii}\theta_{ii} \leq \sum_{m=j}^{i} y_{im}\theta_{mj} \leq k^{c_{ij}}y_{ii}\theta_{ii} \ (1 \leq j < i \leq p)\Big\}.$$

12

It is noted that $\theta_{ii} = k^{\xi_{11}/2 + \sum_{\ell=2}^{i}(1+\xi_{\ell\ell})}$ for $i \geq 2$. 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

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

with

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

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

Let $|\xi_{ij}| < 1 - \varepsilon$ for $i \geq j$ and any small $\varepsilon > 0$. We here use the same arguments as in the preceding section and can easily show that, for $i > j$,

$$U_{ij}(\boldsymbol{Y}, \boldsymbol{\xi}) > y_{ii} \varepsilon (1 - k^{-1})^{i-j-1} k^{c_{ij} + \sum_{\ell=j+1}^{i}(1+\xi_{\ell\ell})},$$
$$L_{ij}(\boldsymbol{Y}, \boldsymbol{\xi}) < -y_{ii} \varepsilon (1 - k^{-1})^{i-j-1} k^{c_{ij} + \sum_{\ell=j+1}^{i}(1+\xi_{\ell\ell})},$$

which yields

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

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

Let

$$P^*_k = \{\boldsymbol{Y} \in \mathcal{T}^+ : L^*_{ij} < y_{ij} < U^*_{ij} \quad (1 \leq j \leq i \leq p)\}.$$

It then holds that $P^*_k \subset \tilde{P}_k(\boldsymbol{\xi})$ and $\cup_{k=1}^{\infty} P^*_k = \mathcal{T}^+ = \mathcal{G}$, which satisfies (A6-3′). Hence the proof is complete. ∎

4.2 Improvement by the isotonic regression method

In the previous subsection, the best equivariant and unrestricted estimator is shown to remain minimax under the restriction (4.1). In this subsection, it is shown that the best equivariant estimator can be further improved on by the isotonic regression method under the restriction (4.1).

Let $\boldsymbol{V} \sim \mathcal{W}_p(n, \boldsymbol{\Sigma})$, and the Cholesky decomposition is written as $\boldsymbol{V} = \boldsymbol{T}\boldsymbol{T}^t$ with $\boldsymbol{T} = (t_{ij}) \in \mathcal{T}^+$. 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}. Then the best equivariant estimator with respect to the Stein loss is rewritten as $\boldsymbol{\delta}^{BE} = \boldsymbol{T}_1 \boldsymbol{\Phi}^{BE}(\boldsymbol{t}) \boldsymbol{T}_1^t$ for $\boldsymbol{t} = (t_{11}^2, \ldots, t_{pp}^2)$, where $\boldsymbol{\Phi}^{BE}(\boldsymbol{t}) = \text{diag}(d_1 t_{11}^2, \ldots, d_p t_{pp}^2)$ with $d_i = (n + p - 2i + 1)^{-1}$. 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 $\boldsymbol{\Phi}^{BE}(\boldsymbol{t})$ 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 $\boldsymbol{\Phi}^{BE}(\boldsymbol{t})$. Let $\boldsymbol{\Phi}^{BE}(\boldsymbol{t}) = \operatorname{diag}(\phi_1^{BE}, \ldots, \phi_p^{BE})$ with $\phi_i^{BE} = d_i t_{ii}^2$, and let $\boldsymbol{\Phi}^{IR}(\boldsymbol{t}) = \operatorname{diag}(\phi_1^{IR}, \ldots, \phi_p^{IR})$, where $\{\phi_1^{IR}, \ldots, \phi_p^{IR}\}$ is a solution of minimizing $\sum_{i=1}^p (\lambda_i - \phi_i^{BE})^2$ subject to $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$.

Theorem 4.2 *Assume that $\boldsymbol{V} \sim \mathcal{W}_p(n, \boldsymbol{\Sigma})$. Under the restriction $\theta_{11}^{-2} \geq \theta_{22}^{-2} \geq \cdots \geq \theta_{pp}^{-2}$, $\boldsymbol{\delta}^{IR} = \boldsymbol{T}_1 \boldsymbol{\Phi}^{IR}(\boldsymbol{t}) \boldsymbol{T}_1^t$ is minimax estimator dominating $\boldsymbol{\delta}^{BE}$ relative to the Stein loss.*

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

Lemma 4.1 (Fenchel 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

$$\sup_{\boldsymbol{x} \in \mathcal{K}} f(\boldsymbol{x}) = - \sup_{\boldsymbol{y} \in \mathcal{K}^*} f^*(\boldsymbol{y}) \tag{4.2}$$

if either $\operatorname{ri}(\operatorname{dom} f) \cap \operatorname{ri}(\mathcal{K}) \neq \emptyset$ or $\operatorname{ri}(\operatorname{dom} f^*) \cap \operatorname{ri}(\mathcal{K}^*) \neq \emptyset$, where ri means relative interior and $\operatorname{dom} f = \{\boldsymbol{x} \in \mathbb{R}^p : f(\boldsymbol{x}) > -\infty\}$. Denote by $\boldsymbol{x}^*$ and $\boldsymbol{y}^*$, respectively, solutions of the left- and right-hand sides of (4.2). 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 4.2. Let $\boldsymbol{\Phi}(\boldsymbol{t}) = \operatorname{diag}(\phi_1, \ldots, \phi_p)$ 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 $\boldsymbol{\delta} = \boldsymbol{T}_1 \boldsymbol{\Phi}(\boldsymbol{t}) \boldsymbol{T}_1^t$ is written as

$$R(\boldsymbol{\delta}, \boldsymbol{\Sigma}) = E[\operatorname{tr} \boldsymbol{\Sigma}^{-1} \boldsymbol{T}_1 \boldsymbol{\Phi}(\boldsymbol{t}) \boldsymbol{T}_1^t - \log|\boldsymbol{\Sigma}^{-1} \boldsymbol{T}_1 \boldsymbol{\Phi}(\boldsymbol{t}) \boldsymbol{T}_1^t| - p]$$
$$= E[\operatorname{tr} \boldsymbol{\Lambda}^{-1} \boldsymbol{U} \boldsymbol{\Phi}(\boldsymbol{t}) \boldsymbol{U}^t - \log|\boldsymbol{\Lambda}^{-1} \boldsymbol{\Phi}(\boldsymbol{t})| - \log|\boldsymbol{U} \boldsymbol{U}^t| - p],$$

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

$$E[\operatorname{tr} \boldsymbol{\Lambda}^{-1} \boldsymbol{U} \boldsymbol{\Phi}(\boldsymbol{t}) \boldsymbol{U}^t] = E\left[\sum_{i=1}^p \phi_i \{\boldsymbol{U}^t \boldsymbol{\Lambda}^{-1} \boldsymbol{U}\}_{ii} \right] = E\left[\sum_{i=1}^p \phi_i \sum_{j=i}^p u_{ji}^2 / \lambda_j \right].$$

It is noted that $u_{ii} = 1$ and that, from the Bartlett decomposition, $t_{ii}^2 / \lambda_i \sim \chi_{n-i+1}^2$ for $i = 1, \ldots, p$ and $u_{ji} | t_{ii} \sim \mathcal{N}(0, \lambda_j / t_{ii}^2)$ for $j > i$. Thus, we obtain

$$E[\operatorname{tr} \boldsymbol{\Lambda}^{-1} \boldsymbol{U} \boldsymbol{\Phi}(\boldsymbol{t}) \boldsymbol{U}^t] = E\left[\sum_{i=1}^p \phi_i E\left[\frac{1}{\lambda_i} + \frac{u_{i+1,i}^2}{\lambda_{i+1}} + \cdots + \frac{u_{pi}^2}{\lambda_p} \,\Big|\, \boldsymbol{t} \right] \right]$$
$$= E\left[\sum_{i=1}^p \phi_i (\lambda_i^{-1} + (p-i)/t_{ii}^2) \right],$$

14

which yields

$$R(\boldsymbol{\delta}, \boldsymbol{\Sigma}) = E\left[\sum_{i=1}^{p}\{\phi_i(\lambda_i^{-1} + (p-i)/t_{ii}^2) - \log(\phi_i/\lambda_i)\}\right] - p. \tag{4.3}$$

Let $\boldsymbol{\phi}^{BE} = (\phi_1^{BE}, \ldots, \phi_p^{BE})^t$ and $\boldsymbol{\phi}^{IR} = (\phi_1^{IR}, \ldots, \phi_p^{IR})^t$. For $i = 1, \ldots, p$, let $\xi_i = \lambda_i^{-1}$ and $a_i = (p-i)/t_{ii}^2$. 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

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

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

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

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

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

The subgradient of $-\ell(\boldsymbol{\xi}|\boldsymbol{\phi}^{BE})$ is equal to $(\phi_1^{BE} - \xi_1^{-1}, \ldots, \phi_p^{BE} - \xi_p^{-1})^t$, so Lemma 4.1 (d) implies that the supremum of $\ell^*(\boldsymbol{\eta}|\boldsymbol{\phi}^{BE})$ attains at

$$\widehat{\boldsymbol{\eta}} = (\hat{\xi}_1^{-1} - \phi_1^{BE}, \ldots, \hat{\xi}_p^{-1} - \phi_p^{BE})^t.$$

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

$$-\sup_{\substack{\boldsymbol{\eta} \in \mathcal{K}^* \\ \boldsymbol{\eta}+\boldsymbol{\phi}^{BE}\succ\boldsymbol{0}_p}} \ell^*(\boldsymbol{\eta}|\boldsymbol{\phi}^{BE}) = -\ell^*(\widehat{\boldsymbol{\eta}}|\boldsymbol{\phi}^{BE}) = -\sum_{i=1}^{p}\log \phi_i^{IR} - p - \sum_{i=1}^{p}\phi_i^{BE}a_i. \tag{4.4}$$

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

$$- \sup_{\substack{\boldsymbol{\eta} \in \mathcal{K}^* \\ \boldsymbol{\eta} + \boldsymbol{\phi}^{BE} \succ \mathbf{0}_p}} \ell^*(\boldsymbol{\eta} | \boldsymbol{\phi}^{BE}) \leq - \sum_{i=1}^{p} \log(\phi_i^{BE} e^{a_i(\phi_i^{IR} - \phi_i^{BE})}) - p - \sum_{i=1}^{p} \phi_i^{BE} a_i$$

$$= - \sum_{i=1}^{p} \log \phi_i^{BE} - p - \sum_{i=1}^{p} \phi_i^{IR} a_i. \tag{4.5}$$

Combining (4.4) and (4.5) gives that $\sum_{i=1}^{p}(\phi_i^{IR} a_i - \log \phi_i^{IR}) \leq \sum_{i=1}^{p}(\phi_i^{BE} a_i - \log \phi_i^{BE})$, or, equivalently,

$$\sum_{i=1}^{p} \{\phi_i^{IR} a_i - \log(\phi_i^{IR}/\lambda_i)\} \leq \sum_{i=1}^{p} \{\phi_i^{BE} a_i - \log(\phi_i^{BE}/\lambda_i)\}. \tag{4.6}$$

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

$$\sum_{i=1}^{p} (\phi_i^{IR} - \phi_i^{BE}) \xi_i = \sum_{i=1}^{p} (\phi_i^{IR} - \phi_i^{BE}) \lambda_i^{-1} \leq 0. \tag{4.7}$$

Combining (4.6) and (4.7), we can see that

$$\sum_{i=1}^{p} \{\phi_i^{IR}(\lambda_i^{-1} + (p-i)/t_{ii}^2) - \log(\phi_i^{IR}/\lambda_i)\} \leq \sum_{i=1}^{p} \{\phi_i^{BE}(\lambda_i^{-1} + (p-i)/t_{ii}^2) - \log(\phi_i^{BE}/\lambda_i)\}$$

with probability one. Thus, it follows from (4.3) that $R(\boldsymbol{\delta}^{IR}, \boldsymbol{\Sigma}) \leq R(\boldsymbol{\delta}^{BE}, \boldsymbol{\Sigma})$, which implies that $\boldsymbol{\delta}^{IR}$ is minimax estimator improving $\boldsymbol{\delta}^{BE}$. ∎

5 Concluding Remarks

In this paper, we have considered the general estimation problem with the invariance structure, and have derived sufficient and unified conditions under which a sequence of prior distributions is least favorable, namely, the best equivariant estimator is minimax. The unified conditions 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). Interestingly, our general conditions are satisfied by the suggested sequence of prior distributions of the covariance matrix. We have also applied the general results given in Theorem 2.1 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 general conditions in Theorem 2.1 and the arguments used in the proofs of Theorem 3.1 have the potential to apply to various restrictions of covariance and precision matrices. For example, consider the restriction $|\boldsymbol{\Sigma}| \leq c$ for positive c, namely, $\prod_{i=1}^{p} \theta_{ii}^2 \geq 1/c$. Then, we can get the following theorem which will be shown in the appendix.

Theorem 5.1 *For an elliptical distribution* (3.10) *with the restriction* $|\mathbf{\Sigma}| \le c$, *the best equivariant estimator of* $\mathbf{\Sigma}$ *is minimax.*

The results given in this paper can be extended 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 $\boldsymbol{V} \sim \mathcal{W}_p(n, \mathbf{\Sigma})$ and $\boldsymbol{X} \sim \mathcal{N}_p(\boldsymbol{\mu}, \mathbf{\Sigma})$. Our results can be easily extended to this model, and minimaxity for the best equivariant estimator of $\mathbf{\Sigma}$ is established.

Acknowledgments

References

[1] Berger, J.O. (1985). *Statistical decision theory and Bayesian analysis (2nd ed.)*, Springer-Verlag, New York.

[2] Bondar, J.V. and Milnes, P. (1981). Amenability: A survey for statistical applications of Hunt-Stein and related conditions on groups, *Z. Wahrsch. verw. Gebiete*, **57**, 103–128.

[3] Calvin, J.A. and 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, *Ann. Statist.*, **19**, 850–869.

[4] Eaton, M.L. and Olkin, I. (1987). Best equivariant estimators of a Cholesky decomposition, *Ann. Statist.*, **15**, 1639–1650.

[5] Girshick, M.A. and Savage, L.J. (1951). Bayes and minimax estimates for quadratic loss functions, in *Proc. Second Berkeley Symp. Math. Statist. Probab., Vol. 1*, University of California Press, Berkeley, pp.53–74.

[6] Hora, R.B. and Buehler, R.J. (1966). Fiducial Theory and Invariant Estimation, *Ann. Math. Statist.*, **37**, 643–656.

[7] James, W. and Stein, C. (1961). Estimation with quadratic loss, in *Proc. Fourth Berkeley Symp. Math. Statist. Probab., Vol. 1*, University of California Press, Berkeley, pp.361–379.

[8] Kiefer, J. (1957). Invariance, minimax sequential estimation, and continuous time processes, *Ann. Math. Statist.*, **28**, 573–601.

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

[10] Kubokawa, T. (2004). Minimaxity in estimation of restricted parameters, *J. Japan Statist. Soc.*, **34**, 229–253.

[11] Kubokawa, T., Marchand, E., Strawderman, W.E. and Turcotte, J.-P. (2012). Minimaxity in predictive density estimation with parametric constraints, Discussion Paper Series, CIRJE-F-843, Faculty of Economics, The University of Tokyo.

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

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

[14] Robertson, T., Wright, F.T. and Dykstra, R.L. (1988). *Order restricted statistical inference*, John Wiley & Sons, New York.

[15] Rockafellar, R.T. (1970). *Convex analysis*, Prinston Univ. Press.

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

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

[18] Strawderman, W.E. (2000). Minimaxity, *J. Amer. Statist. Assoc.*, **95**, 1364–1368.

A Appendix

A.1 Proof of Proposition 3.1

It is noted that the integral in D_F^{BE} given in (3.11) is invariant under transformation $\boldsymbol{\Theta} \to \boldsymbol{B\Theta B}$ 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:

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

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

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

which yields

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

$$= C|\boldsymbol{I}_{p-1} + \boldsymbol{\Theta}_{11}\boldsymbol{\Theta}_{11}^t|^{-a+1/2}(1 + \theta_{pp}^2)^{-a+(p-1)/2}\left(\prod_{i=1}^{p}\theta_{ii}^{n-i}\right)$$

$$\times |\boldsymbol{G}|^{-1/2}(1 + \boldsymbol{\theta}_{21}^t\boldsymbol{G}^{-1}\boldsymbol{\theta}_{21})^{-a}\mathrm{d}\boldsymbol{\Theta}_{11}\mathrm{d}\boldsymbol{\theta}_{21}\mathrm{d}\theta_{pp}$$

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

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

$$= \begin{pmatrix} (1 + \beta_{p-1})\boldsymbol{H}_{p-1} + \beta_{p-1}\boldsymbol{I}_{p-1} & \boldsymbol{0}_{p-1} \\ \boldsymbol{0}_{p-1}^t & \alpha_p \end{pmatrix},$$

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

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

$$h_p = E_*[\{\boldsymbol{\Theta}^t\boldsymbol{\Theta}\}_{pp}] = \alpha_p,$$

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

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

and

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

$$= (v+p-2) \sum_{j=i}^{p-1} \left(\frac{1}{v+j-2} - \frac{1}{v+j-1} \right)$$

$$= \frac{p-i}{v+i-2},$$

which gives h_i in Proposition 3.1. This completes the proof of Proposition 3.1.

A.2 Proof of Theorem 5.1

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

$$P_k = \left\{ \boldsymbol{\Theta} \in \mathcal{T}^+ : 1 \leq \prod_{i=1}^{p} \theta_{ii} \leq k^{2p}, \quad 1/k \leq \theta_{ii} \leq k \quad (i = 2, \dots, p), \quad \text{and} \right.$$

$$\left. - k^{c_{ij}} \theta_{ii} \leq \theta_{ij} \leq k^{c_{ij}} \theta_{ii} \quad (1 \leq j < i \leq p) \right\}.$$

Note that $\cup_{k=1}^{\infty} P_k = P$, which fulfills (A6-1') of Corollary 2.1. To check condition (A6-2') of Corollary 2.1, we observe that $V(P_k) = \int_{P_k} \gamma(\mathrm{d}\boldsymbol{\Theta}) = p2^{p(p+1)/2} (\log k)^p \prod_{i=1}^{p} \prod_{j=1}^{i-1} k^{c_{ij}}$. A set of functions

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

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

$$\int_{\varphi_k(P_k)} f(\boldsymbol{\xi}) \gamma_k(\mathrm{d}\boldsymbol{\xi}) / V(P_k) = 2^{-p(p+1)/2} \int f(\boldsymbol{\xi}) I(\boldsymbol{\xi} \in [-1, 1]^{p(p+1)/2}) \mathrm{d}\boldsymbol{\xi}.$$

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

$$P_k'(\boldsymbol{\Theta}) = \left\{ \boldsymbol{Y} \in \mathcal{T}^+ : 1 \leq \prod_{i=1}^{p} \{\boldsymbol{Y}\boldsymbol{\Theta}\}_{ii} \leq k^{2p}, \ 1/k \leq \{\boldsymbol{Y}\boldsymbol{\Theta}\}_{ii} \leq k \text{ for } i = 2, \dots, p, \text{ and} \right.$$

$$\left. - k^{c_{ij}} \{\boldsymbol{Y}\boldsymbol{\Theta}\}_{ii} \leq \{\boldsymbol{Y}\boldsymbol{\Theta}\}_{ij} \leq k^{c_{ij}} \{\boldsymbol{Y}\boldsymbol{\Theta}\}_{ii} \text{ for } i > j \right\}.$$

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

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

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

The same arguments as in Section 3.1 yields that $\tilde{P}_k(\boldsymbol{\xi}) \supset P_k^* = \{\boldsymbol{Y} \in \mathcal{T}^+ : L_{ij}^* < y_{ij} < U_{ij}^* \quad (1 \le j \le i \le p)\}$, where, for a small enough $\varepsilon > 0$,

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

Therefore, we observe that $\cup_{k=1}^{\infty} P_k^* = \mathcal{T}^+ = \mathcal{G}$. Since this satisfies (A6-3') of Corollary 2.1, the proof is complete.