AISM (2007) 59: 211–233
DOI 10.1007/s10463-006-0044-x

Xiaoqian Sun · Dongchu Sun

Estimation of a multivariate normal covariance matrix with staircase pattern data

Received: 20 January 2005 / Revised: 1 November 2005 / Published online: 11 May 2006

Abstract In this paper, we study the problem of estimating a multivariate normal covariance matrix with staircase pattern data. Two kinds of parameterizations in terms of the covariance matrix are used. One is Cholesky decomposition and another is Bartlett decomposition. Based on Cholesky decomposition of the covariance matrix, the closed form of the maximum likelihood estimator (MLE) of the covariance matrix is given. Using Bayesian method, we prove that the best equivariant estimator of the covariance matrix with respect to the special group related to Cholesky decomposition uniquely exists under the Stein loss. Consequently, the MLE of the covariance matrix is inadmissible under the Stein loss. Our method can also be applied to other invariant loss functions like the entropy loss and the symmetric loss. In addition, based on Bartlett decomposition of the covariance matrix, the Jeffreys prior and the reference prior of the covariance matrix with staircase pattern data are also obtained. Our reference prior is different from Berger and Yang's reference prior. Interestingly, the Jeffreys prior with staircase pattern data is the same as that with complete data. The posterior properties are also investigated. Some simulation results are given for illustration.

Keywords Maximum likelihood estimator · Best equivariant estimator · Covariance matrix · Staircase pattern data · Invariant Haar measure · Cholesky decomposition · Bartlett decomposition · Inadmissibility · Jeffreys prior · Reference prior

1 Introduction

Estimating the covariance matrix $\mathbf{\Sigma}$ in a multivariate normal distribution with incomplete data has been brought to statisticians' attention for several decades. It is

X. Sun (✉) · D. Sun
Department of Statistics, University of Missouri, Columbia, MO 65211, USA
E-mail: sunxia@missouri.edu

well-known that the maximum likelihood estimator (MLE) of $\mathbf{\Sigma}$ from incomplete data with a general missing-data pattern cannot be expressed in closed form. Anderson (1957) listed several general cases where the MLEs of the parameters can be obtained by using conditional distribution. Among these cases, the staircase pattern (also called monotone missing-data pattern) is much more attractive because the other listed missing-data patterns do not actually have enough information for estimating the unconstrained covariance matrix. For this pattern, Liu (1993) presents a decomposition of the posterior distribution of $\mathbf{\Sigma}$ under a family of prior distributions. Jinadasa and Tracy (1992) obtain a complicated form for the maximum likelihood estimators of the unknown mean and the covariance matrix in terms of some sufficient statistics, which extends the work of Anderson and Olkin (1985). Recently, Kibria, Sun, Zidek and Le (2002) discussed estimating $\mathbf{\Sigma}$ using a generalized inverted Wishart (GIW) prior and applied the result to mapping $PM_{2.5}$ exposure. Other related references may include Liu (1999), Little and Rubin (1987), and Brown, Le and Zidek (1994).

In this paper, we consider a general problem of estimating $\mathbf{\Sigma}$ with the staircase pattern data. Two convenient and interesting parameterizations, Cholesky decomposition and Bartlett decomposition, are used. Section 2 describes the model and setup. In Sect. 3, we consider the Cholesky decomposition of $\mathbf{\Sigma}$ and derive a closed form expression of the MLE of $\mathbf{\Sigma}$ based on a set of sufficient statistics different from those in Jinadasa and Tracy (1992). We also show that the best equivariant estimator of $\mathbf{\Sigma}$ with respect to the lower-triangular matrix group uniquely exists under the Stein invariant loss function, resulting in the inadmissibility of the MLE. We find a method to compute the best equivariant estimator of $\mathbf{\Sigma}$ analytically. By applying Bartlett decomposition of $\mathbf{\Sigma}$, Sect. 4 deals with the Jeffreys prior and a reference prior of $\mathbf{\Sigma}$. Surprisingly, the Jeffreys prior of $\mathbf{\Sigma}$ with the staircase pattern data is the same as the usual one with complete data. The reference prior, however, is different from that for complete data given in Yang and Berger (1994). The properties of the posterior distributions under both the Jeffreys and the reference priors are also investigated in this section.

In Sect. 5, an example is given for computing the MLE, the best equivariant estimator and the Bayesian estimator under the Jeffreys prior. Section 6 presents a Markov Chain-Monte Carlo (MCMC) algorithm for Bayesian computation of the posterior under the reference prior. Some numerical comparisons are briefly studied among the MLE, the best equivariant estimator of the covariance matrix with respect to the lower-triangular matrix group, the Bayesian estimators under the Jeffreys prior and the reference prior with respect to the Stein loss. Several proofs are given in Appendix.

2 The staircase pattern observations and the loss

Assume that the population follows the multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $\mathbf{\Sigma}$, namely, $(X_1, X_2, \ldots, X_p)' \sim N_p(\mathbf{0}, \mathbf{\Sigma})$. Suppose that the p variables are divided into k groups $\mathbf{Y}_i = (X_{q_{i-1}+1}, \ldots, X_{q_i})'$, for $i = 1, \ldots, k$, where $q_0 = 0$ and $q_i = \sum_{j=1}^{i} p_i$. Rather than obtaining a random sample of the complete vector $(X_1, X_2, \ldots, X_p) = (\mathbf{Y}_1', \ldots, \mathbf{Y}_k')$, we observe independently the simple random sample of $(\mathbf{Y}_1', \ldots, \mathbf{Y}_i')'$ of size n_i. We could rewrite these observations as the staircase pattern observations,

$$\begin{cases} \mathbf{Y}_{j1}, & j = 1, \ldots, m_1; \\ (\mathbf{Y}'_{j1}, \mathbf{Y}'_{j2})', & j = m_1 + 1, \ldots, m_2; \\ \cdots\cdots\cdots & \cdots\cdots\cdots \\ (\mathbf{Y}'_{j1}, \cdots, \mathbf{Y}'_{jk})', & j = m_{k-1} + 1, \cdots, m_k, \end{cases} \tag{1}$$

where $m_i = \sum_{j=1}^{i} n_j$, $i = 1, \ldots, k$. For convenience, let $m_0 = 0$ hereafter. Such staircase pattern observations are also called monotone samples in Jinadasa and Tracy (1992). Write

$$\mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} & \cdots & \mathbf{\Sigma}_{1k} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} & \cdots & \mathbf{\Sigma}_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{\Sigma}_{k1} & \mathbf{\Sigma}_{k2} & \cdots & \mathbf{\Sigma}_{kk} \end{pmatrix},$$

where $\mathbf{\Sigma}_{ij}$ is a $p_i \times p_j$ matrix. The covariance matrix of $(\mathbf{Y}'_1, \ldots, \mathbf{Y}'_i)'$ is given by

$$\mathbf{\Sigma}_i = \begin{pmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} & \cdots & \mathbf{\Sigma}_{1i} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} & \cdots & \mathbf{\Sigma}_{2i} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{\Sigma}_{i1} & \mathbf{\Sigma}_{i2} & \cdots & \mathbf{\Sigma}_{ii} \end{pmatrix}. \tag{2}$$

Clearly $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_{11}$ and the likelihood function of $\mathbf{\Sigma}$ is

$$\begin{aligned} L(\mathbf{\Sigma}) &= \prod_{i=1}^{k} \prod_{j=m_{i-1}+1}^{m_i} \frac{1}{|\mathbf{\Sigma}_i|^{\frac{1}{2}}} \\ &\quad \times \exp\left\{-\frac{1}{2}(\mathbf{Y}'_{j1}, \ldots, \mathbf{Y}'_{ji})\mathbf{\Sigma}_i^{-1}(\mathbf{Y}'_{j1}, \ldots, \mathbf{Y}'_{ji})'\right\} \\ &= \prod_{i=1}^{k} \frac{1}{|\mathbf{\Sigma}_i|^{\frac{n_i}{2}}} \operatorname{etr}\Big(-\frac{1}{2}\mathbf{\Sigma}_i^{-1}\mathbf{V}_i\Big), \end{aligned} \tag{3}$$

where

$$\mathbf{V}_i = \sum_{j=m_{i-1}+1}^{m_i} \begin{pmatrix} \mathbf{Y}_{j1} \\ \vdots \\ \mathbf{Y}_{ji} \end{pmatrix} (\mathbf{Y}'_{j1}, \ldots, \mathbf{Y}'_{ji}), \quad i = 1, \ldots, k. \tag{4}$$

Clearly $(\mathbf{V}_1, \ldots, \mathbf{V}_k)$ are sufficient statistics of $\mathbf{\Sigma}$ and are mutually independent.

To estimate $\mathbf{\Sigma}$, we consider the Stein loss

$$L(\hat{\mathbf{\Sigma}}, \mathbf{\Sigma}) = \operatorname{tr}(\hat{\mathbf{\Sigma}}\mathbf{\Sigma}^{-1}) - \log|\hat{\mathbf{\Sigma}}\mathbf{\Sigma}^{-1}| - p. \tag{5}$$

This loss function has been commonly used for estimating $\mathbf{\Sigma}$ for a completed sample, for example, Dey and Srinivasan (1985), Haff (1991), Yang and Berger (1994), and Konno (2001).

3 The MLE and the best equivariant estimator

3.1 The MLE of $\boldsymbol{\Sigma}$

For the staircase pattern data, Jinadasa and Tracy (1992) show the closed form of the MLE of $\boldsymbol{\Sigma}$ based on sufficient statistics $\mathbf{V}_1, \ldots, \mathbf{V}_k$. However, the form is quite complicated. We will show that the closed form of the MLE of $\boldsymbol{\Sigma}$ can be easily obtained by using a Cholesky decomposition, the lower-triangular squared root of $\boldsymbol{\Sigma}$ with positive diagonal elements, i.e.,

$$\boldsymbol{\Sigma} = \boldsymbol{\Phi}\boldsymbol{\Phi}', \tag{6}$$

where $\boldsymbol{\Phi}$ has the blockwise form,

$$\boldsymbol{\Phi} = \begin{pmatrix} \boldsymbol{\Phi}_{11} & \mathbf{0} & \cdots & \mathbf{0} \\ \boldsymbol{\Phi}_{21} & \boldsymbol{\Phi}_{22} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Phi}_{k1} & \boldsymbol{\Phi}_{k2} & \cdots & \boldsymbol{\Phi}_{kk} \end{pmatrix}.$$

Here $\boldsymbol{\Phi}_{ij}$ is $p_i \times p_j$, and $\boldsymbol{\Phi}_{ii}$ is lower triangular with positive diagonal elements. Also, we could define,

$$\boldsymbol{\Phi}_i = \begin{pmatrix} \boldsymbol{\Phi}_{11} & \mathbf{0} & \cdots & \mathbf{0} \\ \boldsymbol{\Phi}_{21} & \boldsymbol{\Phi}_{22} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Phi}_{i1} & \boldsymbol{\Phi}_{i2} & \cdots & \boldsymbol{\Phi}_{ii} \end{pmatrix}, \quad i = 1, \cdots, k. \tag{7}$$

Note that $\boldsymbol{\Phi}_1 = \boldsymbol{\Phi}_{11}$, $\boldsymbol{\Phi}_k = \boldsymbol{\Phi}$ and $\boldsymbol{\Phi}_i$ is the Cholesky decomposition of $\boldsymbol{\Sigma}_i$, $i = 1, \ldots, k$. The likelihood function of $\boldsymbol{\Phi}$ is

$$\begin{aligned} L(\boldsymbol{\Phi}) &= \prod_{i=1}^{k} |\boldsymbol{\Phi}_i\boldsymbol{\Phi}_i'|^{-\frac{n_i}{2}} \prod_{j=m_{i-1}+1}^{m_i} \\ &\quad \times \exp\left\{-\frac{1}{2}(\mathbf{Y}_{j1}', \ldots, \mathbf{Y}_{ji}')(\boldsymbol{\Phi}_i\boldsymbol{\Phi}_i')^{-1}(\mathbf{Y}_{j1}', \ldots, \mathbf{Y}_{ji}')'\right\}. \end{aligned} \tag{8}$$

It is necessary that an estimator of a covariance matrix is indeed positive definite. Fortunately, with a Cholesky decomposition, it will guarantee that the resulting estimator of $\boldsymbol{\Sigma}$ is positive definite if each diagonal element of the corresponding estimator of $\boldsymbol{\Phi}_{ii}$ is positive for all $i = 1, \ldots, k$.

Furthermore, we could define another set of sufficient statistics of $\boldsymbol{\Sigma}$. We first define

$$\begin{cases} \mathbf{W}_{111} = \sum_{j=1}^{m_k} \mathbf{Y}_{j1}\mathbf{Y}_{j1}', \\ \mathbf{W}_{i11} = \sum_{j=m_{i-1}+1}^{m_k} (\mathbf{Y}_{j1}', \ \ldots, \ \mathbf{Y}_{j,i-1}')'(\mathbf{Y}_{j1}', \ \ldots, \ \mathbf{Y}_{j,i-1}'), \\ \mathbf{W}_{i21} = \mathbf{W}_{i12}' = \sum_{j=m_{i-1}+1}^{m_k} \mathbf{Y}_{ji}(\mathbf{Y}_{j1}', \ \ldots, \ \mathbf{Y}_{j,i-1}'), \\ \mathbf{W}_{i22} = \sum_{j=m_{i-1}+1}^{m_k} \mathbf{Y}_{ji}\mathbf{Y}_{ji}', \end{cases} \tag{9}$$

for $i = 2, \ldots, k$.. Next, we define the inverse matrix of $\boldsymbol{\Phi}$. i.e., $\boldsymbol{\Delta} = \boldsymbol{\Phi}^{-1}$. Clearly

$\boldsymbol{\Delta}$ is also lower triangular. We write $\boldsymbol{\Delta}$ as a block matrix with similar structure as $\boldsymbol{\Phi}$ in Eq. (7). Define

$$\boldsymbol{\Upsilon}_i = (\boldsymbol{\Delta}_{i1}, \ldots, \boldsymbol{\Delta}_{i,i-1}), \quad i = 2, \ldots, k. \tag{10}$$

Then the likelihood function of $\boldsymbol{\Delta}$ is

$$\begin{aligned} L(\boldsymbol{\Delta}) &= \prod_{i=1}^{k} |\boldsymbol{\Delta}_i|^{n_i} \prod_{j=m_{i-1}+1}^{m_i} \exp\left\{-\frac{1}{2}(\mathbf{Y}'_{j1}, \ldots, \mathbf{Y}'_{ji})\boldsymbol{\Delta}'_i\boldsymbol{\Delta}_i(\mathbf{Y}'_{j1}, \ldots, \mathbf{Y}'_{ji})'\right\} \\ &= \prod_{i=1}^{k} |\boldsymbol{\Delta}_{ii}|^{m_k - m_{i-1}} \mathrm{etr}\Big(-\frac{1}{2}\boldsymbol{\Delta}_{ii}\mathbf{W}_{i22\cdot 1}\boldsymbol{\Delta}'_{ii}\Big) \\ &\quad \times \prod_{i=2}^{k} \mathrm{etr}\left\{-\frac{1}{2}(\boldsymbol{\Upsilon}_i + \boldsymbol{\Delta}_{ii}\mathbf{W}_{i21}\mathbf{W}_{i11}^{-1})\mathbf{W}_{i11}(\boldsymbol{\Upsilon}_i + \boldsymbol{\Delta}_{ii}\mathbf{W}_{i21}\mathbf{W}_{i11}^{-1})'\right\}, \end{aligned} \tag{11}$$

where

$$\mathbf{W}_{122\cdot 1} = \mathbf{W}_{111} \text{ and } \mathbf{W}_{i22\cdot 1} \ = \ \mathbf{W}_{i22} - \mathbf{W}_{i21}\mathbf{W}_{i11}^{-1}\mathbf{W}_{i12}, \ \ i = 2, \ldots, k. \tag{12}$$

It is easy to see that

$$\widetilde{\mathbf{W}} \equiv \{\mathbf{W}_{111}, (\mathbf{W}_{i11}, \mathbf{W}_{i21}, \mathbf{W}_{i22}) : i = 2, \ldots, k\} \tag{13}$$

is sufficient statistics of $\boldsymbol{\Delta}$ or $\boldsymbol{\Sigma}$. Recall that the sufficient statistics $\mathbf{V}_1, \ldots, \mathbf{V}_k$, defined in Eq. (4), are mutually independent. Here $\mathbf{W}_{111}$, $(\mathbf{W}_{211}, \mathbf{W}_{221}, \mathbf{W}_{222}), \ldots,$ $(\mathbf{W}_{k11}, \mathbf{W}_{k21}, \mathbf{W}_{k22})$ are no longer mutually independent.

We are ready to derive a closed form expression of the MLE of $\boldsymbol{\Sigma}$.

Theorem 3.1 *If*

$$n_k \ \geq \ p, \tag{14}$$

the MLE $\widehat{\boldsymbol{\Sigma}}_M$ of $\boldsymbol{\Sigma}$ exists, is unique, and is given by

$$\begin{cases} \widehat{\boldsymbol{\Sigma}}_{1M} \ \equiv \widehat{\boldsymbol{\Sigma}}_{11M} \ = \ \frac{\mathbf{W}_{111}}{m_k}, \\ \widehat{\boldsymbol{\Omega}}_{iM} \ \equiv (\widehat{\boldsymbol{\Sigma}}_{i1M}, \ldots, \widehat{\boldsymbol{\Sigma}}_{i,i-1,M}) \ = \ \mathbf{W}_{i21}\mathbf{W}_{i11}^{-1}\widehat{\boldsymbol{\Sigma}}_{i-1,M}, \\ \widehat{\boldsymbol{\Sigma}}_{iiM} = \frac{\mathbf{W}_{i22\cdot 1}}{m_k - m_{i-1}} + \mathbf{W}_{i21}\mathbf{W}_{i11}^{-1}\widehat{\boldsymbol{\Sigma}}_{i-1,M}^{-1}\mathbf{W}_{i11}^{-1}\mathbf{W}_{i12}, \quad i = 2, \ldots, k. \end{cases} \tag{15}$$

Proof It follows from Eq. (11) that the log-likelihood function of $\boldsymbol{\Delta}$ is

$$\begin{aligned} \log L(\boldsymbol{\Delta}) = -\frac{1}{2}\sum_{i=1}^{k}\Big\{\mathrm{tr}(\boldsymbol{\Delta}_{ii}\mathbf{W}_{i22\cdot 1}\boldsymbol{\Delta}'_{ii}) - (m_k - m_{i-1})\log|\boldsymbol{\Delta}'_{ii}\boldsymbol{\Delta}_{ii}|\Big\} \\ -\frac{1}{2}\sum_{i=2}^{k}\mathrm{tr}\left\{\big(\boldsymbol{\Upsilon}_i + \boldsymbol{\Delta}_{ii}\mathbf{W}_{i21}\mathbf{W}_{i11}^{-1}\big)\mathbf{W}_{i11}\big(\boldsymbol{\Upsilon}_i + \boldsymbol{\Delta}_{ii}\mathbf{W}_{i21}\mathbf{W}_{i11}^{-1}\big)'\right\}. \end{aligned} \tag{16}$$

It is clear that $\mathbf{W}_{111}, \mathbf{W}_{i11}$ and $\mathbf{W}_{i22\cdot 1}$, $i = 2, \ldots, k$ are positive definite with probability one if the condition (14) holds. Thus, from Eq. (16), the MLE of $\boldsymbol{\Delta}$ uniquely exists and so does the MLE of $\boldsymbol{\Sigma}$. Also, from Eq. (16), the MLE $\widehat{\boldsymbol{\Delta}}$ satisfies

$$\begin{cases} (\widehat{\boldsymbol{\Delta}}_{ii}' \widehat{\boldsymbol{\Delta}}_{ii})^{-1} = \frac{\mathbf{W}_{i22\cdot 1}}{m_k - m_{i-1}}, \quad i = 1, \ldots, k, \\ \qquad \widehat{\boldsymbol{\Upsilon}}_i = (\widehat{\boldsymbol{\Delta}}_{i1}, \ldots, \widehat{\boldsymbol{\Delta}}_{i,i-1}) = -\widehat{\boldsymbol{\Delta}}_{ii} \mathbf{W}_{i21} \mathbf{W}_{i11}^{-1}, \quad i = 2, \ldots, k. \end{cases} \tag{17}$$

Because $\boldsymbol{\Sigma} = (\boldsymbol{\Delta}'\boldsymbol{\Delta})^{-1}$, it follows

$$\begin{cases} \boldsymbol{\Sigma}_{11} = (\boldsymbol{\Delta}_{11}' \boldsymbol{\Delta}_{11})^{-1}, \\ \boldsymbol{\Sigma}_i \ = (\boldsymbol{\Delta}_i' \boldsymbol{\Delta}_i)^{-1}, \\ \boldsymbol{\Omega}_i \ = -\boldsymbol{\Delta}_{ii}^{-1} \boldsymbol{\Upsilon}_i (\boldsymbol{\Delta}_{i-1}' \boldsymbol{\Delta}_{i-1})^{-1}, \\ \boldsymbol{\Sigma}_{ii} = (\boldsymbol{\Delta}_{ii}' \boldsymbol{\Delta}_{ii})^{-1} + \boldsymbol{\Delta}_{ii}^{-1} \boldsymbol{\Upsilon}_i (\boldsymbol{\Delta}_{i-1}' \boldsymbol{\Delta}_{i-1})^{-1} \boldsymbol{\Upsilon}_i' (\boldsymbol{\Delta}_{ii}')^{-1}, \quad i = 2, \ldots, k. \end{cases} \tag{18}$$

Combining Eq. (17) with (18), the desired result follows.

From Theorem 3.1, the unique existence of the MLE of the covariance matrix will depend only on the sample size n_k for the whole variables $(X_1, X_2, \ldots, X_p)'$ in the model with staircase pattern data. Intuitively, this is understandable because only the observations from the whole variables $(X_1, X_2, \ldots, X_p)'$ can give the whole description of the covariance matrix. In fact, if $n_k < p$, the MLE of the covariance matrix will no longer exist uniquely no matter how many observations of partial variables $(\mathbf{Y}_1', \ldots, \mathbf{Y}_i')'$ there are $i = 1, \ldots, k-1$. In this paper, we assume that $n_k \geq p$ such that the MLE of $\boldsymbol{\Sigma}$ uniquely exists.

We also notice that it is difficult to evaluate the performance of the MLE of $\boldsymbol{\Sigma}$ because of its complicated structure and the dependence among $\widetilde{\mathbf{W}}$. Some simulation study will be explored in Sect. 6.

3.2 The best equivariant estimator

Now we try to improve over the MLE $\hat{\boldsymbol{\Sigma}}_M$ under the Stein loss Eq. (5). Let $\mathcal{G}$ denote the group of lower-triangular p by p matrices with positive diagonal elements. Note that the problem is invariant under the action of $\mathcal{G}$,

$$\boldsymbol{\Sigma} \to \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}', \ \ \mathbf{V}_i \to \mathbf{A}_i \mathbf{V}_i \mathbf{A}_i', \quad i = 1, \ldots, k, \tag{19}$$

where $\mathbf{A} \in \mathcal{G}$ and $\mathbf{A}_i$ is the upper left q_i by q_i sub-matrix of $\mathbf{A}$, $i = 1, \ldots, k$. It is not obvious to see how $\widetilde{\mathbf{W}} = \{\mathbf{W}_{111}, (\mathbf{W}_{i11}, \mathbf{W}_{i21}, \mathbf{W}_{i22}) : i = 2, \ldots, k\}$ changes after the transformation (19). In fact, after the transformation (19), $\widetilde{\mathbf{W}}$ will change as follows:

$$\begin{cases} \mathbf{W}_{111} \to \mathbf{A}_{11} \mathbf{W}_{111} \mathbf{A}_{11}'; \\ \mathbf{W}_{i11} \to \mathbf{A}_{i-1} \mathbf{W}_{i11} \mathbf{A}_{i-1}', \\ \mathbf{W}_{i21} \to \mathbf{B}_i \mathbf{W}_{i11} \mathbf{A}_{i-1}' + \mathbf{A}_{ii} \mathbf{W}_{i21} \mathbf{A}_{i-1}', \\ \mathbf{W}_{i22} \to \mathbf{B}_i \mathbf{W}_{i11} \mathbf{B}_i' + \mathbf{B}_i \mathbf{W}_{i12} \mathbf{A}_{ii}' + \mathbf{A}_{ii} \mathbf{W}_{i21} \mathbf{B}_i' + \mathbf{A}_{ii} \mathbf{W}_{i22} \mathbf{A}_{ii}', \quad i = 2, \ldots, k, \end{cases}$$

where $\mathbf{B}_i = (\mathbf{A}_{i1}, \ldots, \mathbf{A}_{i,i-1})$, $i = 2, \ldots, k$. Although it seems impossible to find a general form of equivariant estimator with respect to $\mathcal{G}$, an improved estimator of $\boldsymbol{\Sigma}$ over the MLE $\widehat{\boldsymbol{\Sigma}}_M$ may be derived under an invariant loss based on the similar idea of Eaton (1970). The Haar invariant measures will play an important role in finding a better estimator over the MLE $\widehat{\boldsymbol{\Sigma}}_M$. For $\boldsymbol{\Phi} = (\phi_{ij})_{p\times p} \in \mathcal{G}$, from Example 1.14 of Eaton (1989), the right Haar invariant measure on the group $\mathcal{G}$ is

$$\nu_{\mathcal{G}}^r(\boldsymbol{\Phi})\,\mathrm{d}\boldsymbol{\Phi} = \prod_{i=1}^{p} \phi_{ii}^{-p+i-1}\,\mathrm{d}\boldsymbol{\Phi}, \tag{20}$$

while the left Haar invariant measure on the group $\mathcal{G}$ is

$$\nu_{\mathcal{G}}^l(\boldsymbol{\Phi})\,\mathrm{d}\boldsymbol{\Phi} = \prod_{i=1}^{p} \phi_{ii}^{-i}\,\mathrm{d}\boldsymbol{\Phi}. \tag{21}$$

To get a better estimator over the MLE $\widehat{\boldsymbol{\Sigma}}_M$, we need the following lemma. Hereafter, we exploit the notation for matrix variate normal distribution given by Definition 2.2.1 of Gupta and Nagar (2000), that is, $\mathbf{X}_{p\times n} \sim N_{p,n}(\mathbf{M}_{p\times n}, \boldsymbol{\Sigma}_{p\times p} \otimes \boldsymbol{\Psi}_{n\times n})$ if and only if $\mathrm{vec}(\mathbf{X}') \sim N_{pn}(\mathrm{vec}(\mathbf{M}'), \boldsymbol{\Sigma} \otimes \boldsymbol{\Psi})$.

Lemma 3.1 *Let $\boldsymbol{\Phi}$ be the Cholesky decomposition of $\boldsymbol{\Sigma}$ defined by* (7) *and $\boldsymbol{\Delta} = \boldsymbol{\Phi}^{-1}$ with a similar block partition as* (7)*, and $\boldsymbol{\Upsilon}_i$ be given by* (10)*. For $\widetilde{\mathbf{W}}$ defined by* (9)*, if the condition* (14) *holds, then the posterior $p(\boldsymbol{\Delta} \mid \widetilde{\mathbf{W}})$ of $\boldsymbol{\Delta}$ under the prior $\nu_{\mathcal{G}}^r(\boldsymbol{\Phi})$ in* (20) *has the following properties:*

(a) *$p(\boldsymbol{\Delta} \mid \widetilde{\mathbf{W}})$ is proper;*
(b) *Conditional on $\widetilde{\mathbf{W}}$, $\boldsymbol{\Delta}_{11}, (\boldsymbol{\Upsilon}_2, \boldsymbol{\Delta}_{22}), \ldots, (\boldsymbol{\Upsilon}_k, \boldsymbol{\Delta}_{kk})$ are mutually independent;*
(c) $\boldsymbol{\Delta}_{ii}\mathbf{W}_{i22\cdot1}\boldsymbol{\Delta}_{ii}' \mid \widetilde{\mathbf{W}} \sim W_{p_i}(m_k - m_{i-1} - q_{i-1}, \mathbf{I}_{p_i})$*, $i = 1, \ldots, k$;*
(d) $\boldsymbol{\Upsilon}_i \mid \boldsymbol{\Delta}_{ii}, \widetilde{\mathbf{W}} \sim N_{p_i,q_{i-1}}(-\boldsymbol{\Delta}_{ii}\mathbf{W}_{i21}\mathbf{W}_{i11}^{-1}, \mathbf{I}_{p_i} \otimes \mathbf{W}_{i11}^{-1})$*, $i = 2, \ldots, k$;*
(e) *The posterior mean of $\boldsymbol{\Delta}'\boldsymbol{\Delta}$ is finite.*

Here $\mathbf{W}_{i22\cdot1}$ is defined by (12)*.*

Proof Combining (21) with the likelihood function (11), we will easily conclude that the posterior distribution of $\boldsymbol{\Delta}$,

$$\begin{aligned} p(\boldsymbol{\Delta} \mid \widetilde{\mathbf{W}}) \propto{}& \prod_{j=1}^{p_1} \delta_{jj}^{m_k-j} \exp\left\{-\frac{1}{2}\mathrm{tr}(\boldsymbol{\Delta}_{11}\mathbf{W}_{111}\boldsymbol{\Delta}_{11}')\right\} \\ &\times \prod_{i=2}^{k}\prod_{j=1}^{p_i} \delta_{q_{i-1}+j,q_{i-1}+j}^{m_k-m_{i-1}-q_{i-1}-j} \exp\left\{-\frac{1}{2}\mathrm{tr}(\boldsymbol{\Delta}_{ii}\mathbf{W}_{i22\cdot1}\boldsymbol{\Delta}_{ii}')\right\} \\ &\times \prod_{i=2}^{k} \exp\left[-\frac{1}{2}\mathrm{tr}\left\{\left(\boldsymbol{\Upsilon}_i + \boldsymbol{\Delta}_{ii}\mathbf{W}_{i21}\mathbf{W}_{i11}^{-1}\right)\mathbf{W}_{i11}\right.\right. \\ &\qquad\left.\left.\times\left(\boldsymbol{\Upsilon}_i + \boldsymbol{\Delta}_{ii}\mathbf{W}_{i21}\mathbf{W}_{i11}^{-1}\right)'\right\}\right]. \end{aligned} \tag{22}$$

The condition (14) will assure that $\mathbf{W}_{111}$, $\mathbf{W}_{i11}$ and $\mathbf{W}_{i22\cdot1}$ are positive definite with probability one, $i = 2, \ldots, k$. Thus (b), (c) and (d) hold. For (a), $p(\boldsymbol{\Delta} \mid \widetilde{\mathbf{W}})$ is proper if and only if each posterior $p(\boldsymbol{\Delta}_{ii} \mid \widetilde{\mathbf{W}})$ is proper, $i = 1, \ldots, k$, which also is guaranteed by the condition (14).

For (e), because

$$\boldsymbol{\Delta}'\boldsymbol{\Delta} = \begin{pmatrix} \sum_{i=1}^{k} \boldsymbol{\Delta}'_{i1}\boldsymbol{\Delta}_{i1} & \sum_{i=2}^{k} \boldsymbol{\Delta}'_{i1}\boldsymbol{\Delta}_{i2} & \cdots & \boldsymbol{\Delta}'_{k1}\boldsymbol{\Delta}_{kk} \\ \sum_{i=2}^{k} \boldsymbol{\Delta}'_{i2}\boldsymbol{\Delta}_{i1} & \sum_{i=2}^{k} \boldsymbol{\Delta}'_{i2}\boldsymbol{\Delta}_{i2} & \cdots & \boldsymbol{\Delta}'_{k2}\boldsymbol{\Delta}_{kk} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Delta}'_{kk}\boldsymbol{\Delta}_{k1} & \boldsymbol{\Delta}'_{kk}\boldsymbol{\Delta}_{k2} & \cdots & \boldsymbol{\Delta}'_{kk}\boldsymbol{\Delta}_{kk} \end{pmatrix}, \tag{23}$$

so the posterior mean of $\boldsymbol{\Delta}'\boldsymbol{\Delta}$ is finite if the following three conditions hold:

(i) $E(\boldsymbol{\Delta}'_{ii}\boldsymbol{\Delta}_{ii} \mid \widetilde{\mathbf{W}}) < \infty$, $1 \le i \le k$;
(ii) $E(\boldsymbol{\Delta}'_{ij}\boldsymbol{\Delta}_{ii} \mid \widetilde{\mathbf{W}}) < \infty$, $1 \le j < i \le k$;
(iii) $E(\boldsymbol{\Delta}'_{ij}\boldsymbol{\Delta}_{is} \mid \widetilde{\mathbf{W}}) < \infty$, $1 \le j, s < i \le k$.

To finish the proof of Lemma 3.1, we just need to show (i) because

$$E(\boldsymbol{\Delta}'_{ij}\boldsymbol{\Delta}_{ii} \mid \widetilde{\mathbf{W}}) = -\mathbf{B}_{ij}\mathbf{W}_{i11}^{-1}\mathbf{W}_{i12}E(\boldsymbol{\Delta}'_{ii}\boldsymbol{\Delta}_{ii} \mid \widetilde{\mathbf{W}}), \tag{24}$$

$$E(\boldsymbol{\Delta}'_{ij}\boldsymbol{\Delta}_{is} \mid \widetilde{\mathbf{W}}) = \mathbf{B}_{ij}\Big\{p_i\mathbf{W}_{i11} + \mathbf{W}_{i11}^{-1}\mathbf{W}_{i12}E(\boldsymbol{\Delta}'_{ii}\boldsymbol{\Delta}_{ii} \mid \widetilde{\mathbf{W}})\mathbf{W}_{i21}\mathbf{W}_{i11}^{-1}\Big\}\mathbf{B}'_{is}, \tag{25}$$

for $1 \le j,\ s < i \le k$, where

$$\mathbf{B}_{ij} = (\mathbf{0}_{p_j\times p_1}, \ldots, \mathbf{0}_{p_j\times p_{j-1}}, \mathbf{I}_{p_j}, \mathbf{0}_{p_j\times p_{j+1}}, \ldots, \mathbf{0}_{p_j\times p_{i-1}})_{p_j\times q_{i-1}}.$$

Here we obtain (24) and (25) by using the result of part (d). From (22), we get

$$p(\boldsymbol{\Delta}_{ii} \mid \widetilde{\mathbf{W}}) \propto \prod_{j=1}^{p_i} \delta_{q_{i-1}+j,q_{i-1}+j}^{m_k-m_{i-1}-q_{i-1}-j} \exp\left\{-\frac{1}{2}\mathrm{tr}\big(\boldsymbol{\Delta}_{ii}\mathbf{W}_{i22\cdot 1}\boldsymbol{\Delta}'_{ii}\big)\right\}.$$

Lemma 3 of Sun and Sun (2005) shows that $E(\boldsymbol{\Delta}'_{ii}\boldsymbol{\Delta}_{ii} \mid \widetilde{\mathbf{W}})$ is finite if $m_k - m_{i-1} - q_{i-1} - j \ge 0$, $j = 1, \ldots, p_i$, $i = 1, \ldots, k$, which is equivalent to $n_k \ge p$.

Because the MLE $\widehat{\boldsymbol{\Sigma}}_M$ belongs to a class of $\mathcal{G}$-equivariant estimators, the following theorem shows that the best $\mathcal{G}$-equivariant estimator $\widehat{\boldsymbol{\Sigma}}_B$, which is better than $\widehat{\boldsymbol{\Sigma}}_M$, uniquely exists.

Theorem 3.2 *For the staircase pattern data* (1) *and under the Stein loss* (5)*, if the condition* (14) *holds, the best $\mathcal{G}$-equivariant estimator $\widehat{\boldsymbol{\Sigma}}_B$ of $\boldsymbol{\Sigma}$ exists uniquely and is given by*

$$\widehat{\boldsymbol{\Sigma}}_B = \{E(\boldsymbol{\Delta}'\boldsymbol{\Delta} \mid \widetilde{\mathbf{W}})\}^{-1}. \tag{26}$$

Proof For any invariant loss, by Theorem 6.5 in Eaton (1989), the best equivariant estimator of $\boldsymbol{\Sigma}$ with respect to the group $\mathcal{G}$ will be the Bayesian estimator if we take the right invariant Haar measure $\nu_{\mathcal{G}}^r(\boldsymbol{\Phi})$ on the group $\mathcal{G}$ as a prior. The Stein loss (5) is an invariant loss, which can be written as a function of $\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}$. Thus, to find the best equivariant estimator of $\boldsymbol{\Sigma}$ under the group $\mathcal{G}$, we just need to minimize

$$H(\widehat{\boldsymbol{\Sigma}}) = \int \Big\{\mathrm{tr}(\hat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}) - \log|\hat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}| - p\Big\} L(\boldsymbol{\Sigma})\nu_{\mathcal{G}}^r(\boldsymbol{\Phi})\,\mathrm{d}\boldsymbol{\Phi}.$$

Making the transformation $\mathbf{\Phi} \to \mathbf{\Delta} = \mathbf{\Phi}^{-1}$ and noticing that $\nu_{\mathcal{G}}^{l}(\mathbf{\Delta})\,\mathrm{d}\mathbf{\Delta} = \nu_{\mathcal{G}}^{r}(\mathbf{\Phi})\,\mathrm{d}\mathbf{\Phi}$, we get

$$H(\widehat{\mathbf{\Sigma}}) \propto \int \left\{\mathrm{tr}(\widehat{\mathbf{\Sigma}}\mathbf{\Delta}'\mathbf{\Delta}) - \log|\widehat{\mathbf{\Sigma}}| - \log|\mathbf{\Delta}'\mathbf{\Delta}| - p\right\} p(\mathbf{\Delta} \mid \widetilde{\mathbf{W}})\,\mathrm{d}\mathbf{\Delta},$$

where $p(\mathbf{\Delta} \mid \widetilde{\mathbf{W}})$ is given by Lemma 3.1. So $H(\widehat{\mathbf{\Sigma}})$ attains a unique maximum at $\widehat{\mathbf{\Sigma}} = \{E(\mathbf{\Delta}'\mathbf{\Delta} \mid \widetilde{\mathbf{W}})\}^{-1}$ if $E(\mathbf{\Delta}'\mathbf{\Delta} \mid \widetilde{\mathbf{W}})$ exists. The result then follows by Lemma 3.1(e).

Remark 1 By Kiefer (1957), the best $\mathcal{G}$-equivariant estimator $\widehat{\mathbf{\Sigma}}_B$ of $\mathbf{\Sigma}$ is minimax under the Stein loss (5) because the group $\mathcal{G}$ is solvable.

Notice that the best $\mathcal{G}$-equivariant estimator $\widehat{\mathbf{\Sigma}}_B$ is uniformly better than the MLE $\widehat{\mathbf{\Sigma}}_M$ under the Stein loss (5) because the MLE $\widehat{\mathbf{\Sigma}}_M$ is equivariant under the lower-triangular group $\mathcal{G}$. However, it is still unclear how to get the explicit expressions of the risks for $\widehat{\mathbf{\Sigma}}_M$ and $\widehat{\mathbf{\Sigma}}_B$ under the staircase pattern model. Some simulation results will be reported in Sect. 6.

3.3 An algorithm to compute the best equivariant estimate

Eaton (1970) gave the explicit form of the best $\mathcal{G}$-equivariant estimator $\widehat{\mathbf{\Sigma}}_B$ for $k = 2$. In the following we will show how to compute $\widehat{\mathbf{\Sigma}}_B$ analytically for general k. The key point is to get the closed form of $E(\mathbf{\Delta}'\mathbf{\Delta} \mid \widetilde{\mathbf{W}})$. From (23), (24) and (25), we only need to compute each $E(\mathbf{\Delta}_{ii}'\mathbf{\Delta}_{ii} \mid \widetilde{\mathbf{W}})$, and this can easily be obtained by Eaton (1970) as follows:

$$E(\mathbf{\Delta}_{ii}'\mathbf{\Delta}_{ii} \mid \widetilde{\mathbf{W}}) = (\mathbf{K}_i')^{-1}\mathbf{D}_i\mathbf{K}_i^{-1} \tag{27}$$

where $\mathbf{K}_i$ is the Cholesky decomposition of $\mathbf{W}_{i22\cdot 1}$, $\mathbf{D}_i = \mathrm{diag}(d_{i1}, \ldots, d_{ip_i})$, and

$$d_{ij} = m_k - m_{i-1} - q_{i-1} + p_i - 2j + 1. \tag{28}$$

Combining (23), (24), (25) and (27), we could derive the closed form of $E(\mathbf{\Delta}'\mathbf{\Delta} \mid \widetilde{\mathbf{W}})$, and compute $\widehat{\mathbf{\Sigma}}_B$.

Algorithm for computing $\widehat{\mathbf{\Sigma}}_B$

Step 1: For the staircase pattern observations (1), compute $\widetilde{\mathbf{W}}$ based on (9).
Step 2: Compute $E(\mathbf{\Delta}_{ii}'\mathbf{\Delta}_{ii} \mid \widetilde{\mathbf{W}})$ based on (27) and (28), $i = 1, \ldots, k$.
Step 3: Compute $E(\mathbf{\Delta}_{ij}'\mathbf{\Delta}_{ii} \mid \widetilde{\mathbf{W}})$ and $E(\mathbf{\Delta}_{ij}'\mathbf{\Delta}_{is} \mid \widetilde{\mathbf{W}})$ based on (24) and (25) respectively, $1 \le j,\ s < i \le k$.
Step 4: Obtain $E(\mathbf{\Delta}'\mathbf{\Delta} \mid \widetilde{\mathbf{W}})$ based on the expression (23).
Step 5: $\widehat{\mathbf{\Sigma}}_B = \{E(\mathbf{\Delta}'\mathbf{\Delta} \mid \widetilde{\mathbf{W}})\}^{-1}$.

From Theorem 3.1, the MLE $\widehat{\mathbf{\Sigma}}_M$ has a closed form explicitly in terms of $\widetilde{\mathbf{W}}$. We know that the best equivariant estimator is $\widehat{\mathbf{\Sigma}}_B = \{E(\mathbf{\Delta}'\mathbf{\Delta} \mid \widetilde{\mathbf{W}})\}^{-1}$ although we are able to give the closed form expression of $E(\mathbf{\Delta}'\mathbf{\Delta} \mid \widetilde{\mathbf{W}})$ in terms of $\widetilde{\mathbf{W}}$. An interesting question is: could we derive a closed form expression for $\{E(\mathbf{\Delta}'\mathbf{\Delta} \mid \widetilde{\mathbf{W}})\}^{-1}$ directly? At the time we are unable to do so although this is not too restrictive in computing $\widehat{\mathbf{\Sigma}}_B$.

In Sect. 4.4, we could see that the Bayesian estimator of $\mathbf{\Sigma}$ under the Jeffreys prior $\pi_J(\mathbf{\Sigma})$ has a similar form as that of the best $\mathcal{G}$-equivariant estimator $\widehat{\mathbf{\Sigma}}_B$. Again, we are not sure if $\hat{\mathbf{\Sigma}}_J$ could be expressed explicitly in terms of $\widetilde{\mathbf{W}}$ directly.

4 The Jeffreys and reference priors and posteriors

4.1 Bartlet decomposition

To derive the Jeffreys and reference priors of $\mathbf{\Sigma}$, we may use the parameterizations of the Cholesky decomposition given by (6). It is more convenient to consider a one-to-one transformation by Bartlett (1933),

$$\begin{cases} \mathbf{\Gamma}_{11} = \mathbf{\Sigma}_{11}, \\ \mathbf{\Lambda}_i \widehat{=} (\mathbf{\Gamma}_{i1}, \dots, \mathbf{\Gamma}_{i,i-1}) = (\mathbf{\Sigma}_{i1}, \dots, \mathbf{\Sigma}_{i,i-1})\mathbf{\Sigma}_{i-1}^{-1}, \\ \mathbf{\Gamma}_{ii} = \mathbf{\Sigma}_{ii} - (\mathbf{\Sigma}_{i1}, \dots, \mathbf{\Sigma}_{i,i-1})\mathbf{\Sigma}_{i-1}^{-1}(\mathbf{\Sigma}_{i1}, \dots, \mathbf{\Sigma}_{i,i-1})', \quad i = 2, \dots, k, \end{cases} \tag{29}$$

where $\mathbf{\Sigma}_i$ is given by (2). It is easy to see that

$$\mathbf{\Sigma}_i^{-1} = \begin{pmatrix} \mathbf{I}_{p_1} & \mathbf{0} & \cdots & \mathbf{0} \\ -\mathbf{\Gamma}_{21} & \mathbf{I}_{p_2} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ -\mathbf{\Gamma}_{i1} & -\mathbf{\Gamma}_{i2} & \cdots & \mathbf{I}_{p_i} \end{pmatrix}' \begin{pmatrix} \mathbf{\Gamma}_{11}^{-1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{\Gamma}_{22}^{-1} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{\Gamma}_{ii}^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{I}_{p_1} & \mathbf{0} & \cdots & \mathbf{0} \\ -\mathbf{\Gamma}_{21} & \mathbf{I}_{p_2} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ -\mathbf{\Gamma}_{i1} & -\mathbf{\Gamma}_{i2} & \cdots & \mathbf{I}_{p_i} \end{pmatrix} \tag{30}$$

for $i = 2, \dots, k$. For convenience, let

$$\mathbf{\Gamma} = \begin{pmatrix} \mathbf{\Gamma}_{11} & \mathbf{\Gamma}_{21}' & \cdots & \mathbf{\Gamma}_{k1}' \\ \mathbf{\Gamma}_{21} & \mathbf{\Gamma}_{22} & \cdots & \mathbf{\Gamma}_{k2}' \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{\Gamma}_{k1} & \mathbf{\Gamma}_{k2} & \cdots & \mathbf{\Gamma}_{kk} \end{pmatrix}. \tag{31}$$

Notice that $\mathbf{\Gamma}$ is symmetric and the transformation from $\mathbf{\Sigma}$ to $\mathbf{\Gamma}$ is one-to-one. In fact, the generalized inverted Wishart prior in Kibria, Sun, Zidek and Le (2002) is to define each $\mathbf{\Gamma}_{ii}$ as an inverted Wishart distribution and each $\mathbf{\Lambda}_i$ as matrix-variate normal distribution. Moreover, the decomposition (30) may be viewed as a block counterpart of the decomposition considered by Pourahmadi (1999, 2000), and Daniels and Pourahmadi (2002), etc. Using $\mathbf{\Gamma}$ is more efficient than using $\mathbf{\Sigma}$ itself to deal with the statistical inference of the covariance matrix with staircase pattern data.

4.2 Fisher information and the Jeffreys prior

First of all, we give the Fisher information matrix of $\mathbf{\Gamma}$ as follows.

Theorem 4.1 *Let* $\boldsymbol{\theta} = (\text{vech}'(\mathbf{\Gamma}_{11}), \text{vec}'(\mathbf{\Lambda}_2), \text{vech}'(\mathbf{\Gamma}_{22}), \dots, \text{vec}'(\mathbf{\Lambda}_k), \text{vech}'(\mathbf{\Gamma}_{kk}))'$, *where* $\text{vec}(\cdot)$ *and* $\text{vech}(\cdot)$ *are two matrix operators described in detail in Henderson and Searle (1979), and for notational simplicity,* $\text{vec}'(\cdot) \equiv (\text{vec}(\cdot))'$, $\text{vech}'(\cdot) \equiv (\text{vech}(\cdot))'$. *Then the Fisher information matrix of* $\boldsymbol{\theta}$ *in multivariate normal distribution with staircase pattern data is given by*

$$\mathbf{I}(\boldsymbol{\theta}) = \text{diag}(\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3, \dots, \mathbf{A}_{2k-2}, \mathbf{A}_{2k-1}), \tag{32}$$

where

$$\mathbf{A}_{2i-2} = (m_k - m_{i-1})\boldsymbol{\Sigma}_{i-1} \otimes \boldsymbol{\Gamma}_{ii}^{-1}, \quad i = 2, \ldots, k;$$
$$\mathbf{A}_{2i-1} = \frac{m_k - m_{i-1}}{2}\mathbf{G}_i'(\boldsymbol{\Gamma}_{ii}^{-1} \otimes \boldsymbol{\Gamma}_{ii}^{-1})\mathbf{G}_i, \quad i = 1, \ldots, k$$

with $\mathbf{G}_i$ *uniquely satisfying* $\mathrm{vec}(\boldsymbol{\Gamma}_{ii}) = \mathbf{G}_i \cdot \mathrm{vech}(\boldsymbol{\Gamma}_{ii})$, $i = 1, \ldots, k$.

The proof of the theorem is given in Appendix. We should point out that the method in the proof also works for the usual complete data. We just need to replace $m_k - m_{i-1}$ by the sample size n of the complete data. Thus the Jeffreys prior for the staircase pattern data will be the same as that for the complete data. The details can be seen in the following theorem.

Theorem 4.2 *The Jeffreys prior of the covariance matrix* $\boldsymbol{\Sigma}$ *in multivariate normal distribution with staircase pattern data is given by*

$$\pi_J(\boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-(p+1)/2}, \tag{33}$$

which is the same as the usual Jeffreys prior for the complete data.

Proof It is well-known that for $\mathbf{A}_{p\times p}$ and $\mathbf{B}_{q\times q}$, $|\mathbf{A} \otimes \mathbf{B}| = |\mathbf{A}|^q|\mathbf{B}|^p$. Also, from Theorem 3.13 (d) and Theorem 3.14(b) of Magnus and Neudecker (1999), it follows

$$|\mathbf{G}_i'(\boldsymbol{\Gamma}_{ii}^{-1} \otimes \boldsymbol{\Gamma}_{ii}^{-1})\mathbf{G}_i| \;=\; |\mathbf{G}_i^{+}(\boldsymbol{\Gamma}_{ii} \otimes \boldsymbol{\Gamma}_{ii})\mathbf{G}_i^{+\prime}|^{-1} \;=\; 2^{p_i(p_i-1)/2}|\boldsymbol{\Gamma}_{ii}|^{-p_i-1},$$

where $\mathbf{G}_i^{+}$ stands for the Moore–Penrose inverse matrix of $\mathbf{G}_i$. Thus from (32), we can easily get the Jeffreys prior of $\boldsymbol{\Gamma}$,

$$\begin{aligned}\pi_J(\boldsymbol{\Gamma}) &\propto |\boldsymbol{\Gamma}_{11}|^{-(p_1+1)/2}|\boldsymbol{\Gamma}_{22}|^{-(p_1+p_2+1)/2}\ldots|\boldsymbol{\Gamma}_{kk}|^{-(p_1+\cdots+p_k+1)/2}\prod_{i=1}^{k-1}|\boldsymbol{\Sigma}_i|^{p_{i+1}/2}\\ &= \prod_{i=1}^{k}|\boldsymbol{\Gamma}_{ii}|^{(p-2q_i-1)/2}.\end{aligned} \tag{34}$$

Moreover, the Jacobian of the transformation from $\boldsymbol{\Gamma}$ to $\boldsymbol{\Sigma}$ is given by

$$J(\boldsymbol{\Gamma} \to \boldsymbol{\Sigma}) \;=\; \prod_{i=1}^{k-1}|\boldsymbol{\Sigma}_i^{-1} \otimes I_{p_{i+1}}| \;=\; \prod_{i=1}^{k-1}|\boldsymbol{\Sigma}_i|^{-p_{i+1}} \;=\; \prod_{i=1}^{k-1}|\boldsymbol{\Gamma}_{ii}|^{q_i-p}. \tag{35}$$

Consequently, the Jeffreys prior of $\boldsymbol{\Sigma}$ becomes

$$\pi_J(\boldsymbol{\Sigma}) \;=\; \pi_J(\boldsymbol{\Gamma}) \cdot J(\boldsymbol{\Gamma} \to \boldsymbol{\Sigma}),$$

which is given by (33) after simple calculation.

4.3 The reference prior

Suppose that each $p_i \geq 2$, $i = 1, \ldots, k$. Because $\mathbf{\Gamma}_{ii}$ is still positive definite, it can be decomposed as $\mathbf{\Gamma}_{ii} = \mathbf{O}_i'\mathbf{D}_i\mathbf{O}_i$, with $\mathbf{O}_i$ an orthogonal matrix with positive elements for the first row and $\mathbf{D}_i$ a diagonal matrix, $\mathbf{D}_i = \text{diag}(d_{i1}, \ldots, d_{ip_i})$, with $d_{i1} \geq \cdots \geq d_{ip_i}$, $i = 1, \ldots, k$. With the similar idea in Yang and Berger (1994), a reference prior may be derived in the following proposition.

Theorem 4.3 *Suppose that $p_i \geq 2$, $i = 1, \ldots, k$. For the multivariate normal distribution with staircase pattern data, the reference prior $\pi_R(\mathbf{\Sigma})$ of $\mathbf{\Sigma}$ is given as follows, providing the group ordering used lists $\mathbf{D}_1, \ldots, \mathbf{D}_k$ before $(\mathbf{O}_1, \ldots, \mathbf{O}_k, \mathbf{\Lambda}_2, \ldots, \mathbf{\Lambda}_k)$ and for each i, the $\{d_{ij}\}$ are ordered monotonically (either increasing or decreasing):*

$$\pi_R(\mathbf{\Sigma})\, d\mathbf{\Sigma} \propto \left\{ \prod_{i=1}^{k} |\mathbf{\Gamma}_{ii}|^{-1} \prod_{1 \leq s < t \leq p_i} (d_{is} - d_{it})^{-1} \right\} d\mathbf{\Gamma}. \tag{36}$$

Proof Because we have obtained the Fisher information matrix of $\mathbf{\Gamma}$ in Theorem 4.1, the proof is thus similar to that of Theorem 1 in Yang and Berger (1994) based on a general algorithm for computing ordered group reference priors in Berger and Bernardo (1992). □

Notice that although $\prod_{i=1}^{k} |\mathbf{\Gamma}_{ii}| = |\mathbf{\Sigma}|$, there is no ordering between the eigenvalues of $\mathbf{\Gamma}_{ii}$ and those of $\mathbf{\Gamma}_{jj}$, $1 \leq i \neq j \leq k$. Because the reference prior (36) depends on the group partition of the variables, it is totally different from the reference prior obtained by Yang and Berger (1994). Although the Jacobian of the transformation from $\mathbf{\Gamma}$ to $\mathbf{\Sigma}$ is given by (35), we actually just gave the explicit form of the reference prior of $\mathbf{\Gamma}$ rather than $\mathbf{\Sigma}$ itself because it is hard to express the eigenvalues of $\mathbf{\Gamma}_{ii}$ in terms of $\mathbf{\Sigma}$. However, this will not result in any difficulty in simulation work because we just need to simulate $\mathbf{\Gamma}$ first and then get the required sample by transformation for further Bayesian computation of $\mathbf{\Sigma}$. The details will be shown in Sect. 6

Remark 2 If $p_i = 1$ for some i, the corresponding reference prior under the condition of the above proposition will be obtained by replacing the part $|\mathbf{\Gamma}_{ii}|^{-1} \prod_{1 \leq s < t \leq p_i} (d_{is} - d_{it})^{-1}$ in (36) by $1/|\mathbf{\Gamma}_{ii}|$.

4.4 Properties of the posteriors under $\pi_J(\mathbf{\Sigma})$ and $\pi_R(\mathbf{\Sigma})$

The following two lemmas give the properties of the posterior distributions of the covariance matrix under the Jeffreys prior (33) and the reference prior (36). Both proofs can be seen in Appendix.

Lemma 4.1 *Let $\mathbf{\Gamma}$ be defined by* (31) *and* (29). *Then under the Jeffreys prior $\pi_J(\mathbf{\Sigma})$ given by* (33) *or the equivalent Jeffreys prior $\pi_J(\mathbf{\Gamma})$ given by* (34)*, the posterior $p_J(\mathbf{\Gamma} \mid \widetilde{\mathbf{W}})$ has the following properties if the condition* (14) *holds:*

(a) *$p_J(\mathbf{\Gamma} \mid \widetilde{\mathbf{W}})$ is proper*

(b) $\boldsymbol{\Gamma}_{11}, (\boldsymbol{\Lambda}_2, \boldsymbol{\Gamma}_{22}), \ldots, (\boldsymbol{\Lambda}_k, \boldsymbol{\Gamma}_{kk})$ *are mutually independent*
(c) *The marginal posterior of* $\boldsymbol{\Gamma}_{ii}$ *is* $IW_{p_i}(\mathbf{W}_{i22\cdot 1}, m_k - m_{i-1} + q_i - p)$*,* $i = 1, \ldots, k$
(d) $(\boldsymbol{\Lambda}_i \mid \boldsymbol{\Gamma}_{ii}, \widetilde{\mathbf{W}}) \sim N_{p_i, q_{i-1}}(\mathbf{W}_{i21}\mathbf{W}_{i11}^{-1}, \boldsymbol{\Gamma}_{ii} \otimes \mathbf{W}_{i11}^{-1})$*,* $i = 2, \ldots, k$
(e) *The posterior mean of* $\boldsymbol{\Sigma}^{-1}$ *is finite.*

Here the definition of an IW_p *(inverse Wishart) distribution with dimension p follows that on page* 268 *of* Anderson (1984).

Lemma 4.2 *Let* $\boldsymbol{\Gamma}$ *be defined by* (31) *and* (29). *Under the reference prior* $\pi_R(\boldsymbol{\Sigma})$ *given by* (36)*, the posterior* $p_R(\boldsymbol{\Gamma} \mid \widetilde{\mathbf{W}})$ *has the following properties if the condition (14) holds:*

(a) $p_R(\boldsymbol{\Gamma} \mid \widetilde{\mathbf{W}})$ *is proper*
(b) $\boldsymbol{\Gamma}_{11}, (\boldsymbol{\Lambda}_2, \boldsymbol{\Gamma}_{22}), \ldots, (\boldsymbol{\Lambda}_k, \boldsymbol{\Gamma}_{kk})$ *are mutually independent*
(c) *For* $i = 1, \ldots, k$*,*

$$p_R(\boldsymbol{\Gamma}_{ii} \mid \widetilde{\mathbf{W}}) \propto \frac{1}{|\boldsymbol{\Gamma}_{ii}|^{(m_k - m_{i-1} - q_{i-1})/2+1}} \times \prod_{1 \le s < t \le p_i} \frac{1}{d_{is} - d_{it}} \exp\left\{-\frac{1}{2}\operatorname{tr}(\boldsymbol{\Gamma}_{ii}^{-1}\mathbf{W}_{i22\cdot 1})\right\};$$

(d) *For* $i = 2, \ldots, k$*,* $(\boldsymbol{\Lambda}_i \mid \boldsymbol{\Gamma}_{ii}, \widetilde{\mathbf{W}}) \sim N_{p_i, q_{i-1}}(\mathbf{W}_{i21}\mathbf{W}_{i11}^{-1}, \boldsymbol{\Gamma}_{ii} \otimes \mathbf{W}_{i11}^{-1})$*;*
(e) *The posterior mean of* $\boldsymbol{\Sigma}^{-1}$ *is finite.*

From Yang and Berger (1994), the Bayesian estimator of $\boldsymbol{\Sigma}$ under the Stein loss function (5) is $\widehat{\boldsymbol{\Sigma}} = \left\{E(\boldsymbol{\Sigma}^{-1} \mid \widetilde{\mathbf{W}})\right\}^{-1}$. So we have the following theorem immediately.

Theorem 4.4 *If the condition* (14) *holds, the Bayesian estimators of* $\boldsymbol{\Sigma}$ *under* $\pi_J(\boldsymbol{\Sigma})$ *and* $\pi_R(\boldsymbol{\Sigma})$ *under the Stein loss function* (5) *uniquely exist.*

Based on Lemma 4.1, an algorithm for computing Bayesian estimator $\widehat{\boldsymbol{\Sigma}}_J$ under $\pi_J(\boldsymbol{\Sigma})$ can be proposed, which is similar to that for computing the best equivariant estimator $\widehat{\boldsymbol{\Sigma}}_B$ in Sect. 3.3. In Sect. 6, we will present an MCMC algorithm for computing Bayesian estimator $\widehat{\boldsymbol{\Sigma}}_R$ under $\pi_R(\boldsymbol{\Sigma})$.

5 The case when $k = 3$

As an example, we apply the algorithm of computing $\widehat{\boldsymbol{\Sigma}}_B$ in Sect. 3.3 for $k = 3$. In this case,

$$\widehat{\boldsymbol{\Sigma}}_B = \left\{E(\boldsymbol{\Delta}'\boldsymbol{\Delta} \mid \widetilde{\mathbf{W}})\right\}^{-1} \equiv \begin{pmatrix} \widehat{\boldsymbol{\Omega}}_{11B} & \widehat{\boldsymbol{\Omega}}_{12B} & \widehat{\boldsymbol{\Omega}}_{13B} \\ \widehat{\boldsymbol{\Omega}}_{21B} & \widehat{\boldsymbol{\Omega}}_{22B} & \widehat{\boldsymbol{\Omega}}_{23B} \\ \widehat{\boldsymbol{\Omega}}_{31B} & \widehat{\boldsymbol{\Omega}}_{32B} & \widehat{\boldsymbol{\Omega}}_{33B} \end{pmatrix}^{-1},$$

where

$$
\begin{aligned}
\widehat{\boldsymbol{\Omega}}_{11B} &= (\mathbf{K}_1')^{-1}\mathbf{D}_1\mathbf{K}_1^{-1} \\
&\quad + \sum_{i=2}^{3} \mathbf{B}_{i1}\left\{p_i\mathbf{W}_{i11}^{-1} + \mathbf{W}_{i11}^{-1}\mathbf{W}_{i12}(\mathbf{K}_i')^{-1}\mathbf{D}_i\mathbf{K}_i^{-1}\mathbf{W}_{i21}\mathbf{W}_{i11}^{-1}\right\}\mathbf{B}_{i1}', \\
\widehat{\boldsymbol{\Omega}}_{12B} &= \widehat{\boldsymbol{\Omega}}_{21B}' \;=\; -\mathbf{W}_{211}^{-1}\mathbf{W}_{212}(\mathbf{K}_2')^{-1}\mathbf{D}_2\mathbf{K}_2^{-1} \\
&\qquad\qquad + \mathbf{B}_{31}\left\{p_3\mathbf{W}_{311}^{-1} + \mathbf{W}_{311}^{-1}\mathbf{W}_{312}(\mathbf{K}_3')^{-1}\mathbf{D}_3\mathbf{K}_3^{-1}\mathbf{W}_{321}\mathbf{W}_{311}^{-1}\right\}\mathbf{B}_{32}', \\
\widehat{\boldsymbol{\Omega}}_{13B} &= \widehat{\boldsymbol{\Omega}}_{31B}' \;=\; -\mathbf{B}_{31}\mathbf{W}_{311}^{-1}\mathbf{W}_{312}(\mathbf{K}_3')^{-1}\mathbf{D}_3\mathbf{K}_3^{-1}, \\
\widehat{\boldsymbol{\Omega}}_{22B} &= (\mathbf{K}_2')^{-1}\mathbf{D}_2\mathbf{K}_2^{-1} \\
&\quad + \mathbf{B}_{32}\left\{p_3\mathbf{W}_{311}^{-1} + \mathbf{W}_{311}^{-1}\mathbf{W}_{312}(\mathbf{K}_3')^{-1}\mathbf{D}_3\mathbf{K}_3^{-1}\mathbf{W}_{321}\mathbf{W}_{311}^{-1}\right\}\mathbf{B}_{32}', \\
\widehat{\boldsymbol{\Omega}}_{23B} &= \widehat{\boldsymbol{\Omega}}_{32B}' \;=\; -\mathbf{B}_{32}\mathbf{W}_{311}^{-1}\mathbf{W}_{312}(\mathbf{K}_3')^{-1}\mathbf{D}_3\mathbf{K}_3^{-1}, \\
\widehat{\boldsymbol{\Omega}}_{33B} &= (\mathbf{K}_3')^{-1}\mathbf{D}_3\mathbf{K}_3^{-1}.
\end{aligned}
$$

Here,

$$
\mathbf{B}_{21} = \mathbf{I}_{p_1}, \quad \mathbf{B}_{31} = (\mathbf{I}_{p_1}, \mathbf{0}_{p_1\times p_2}), \quad \mathbf{B}_{32} = (\mathbf{0}_{p_2\times p_1}, \mathbf{I}_{p_2}). \tag{37}
$$

To see the difference between the best equivariant estimator $\widehat{\boldsymbol{\Sigma}}_B$ and the maximum likelihood estimator $\widehat{\boldsymbol{\Sigma}}_M$ when $k = 3$, we could rewrite $\widehat{\boldsymbol{\Sigma}}_M$ as follows:

$$
\widehat{\boldsymbol{\Sigma}}_M \equiv \begin{pmatrix} \widehat{\boldsymbol{\Omega}}_{11M} & \widehat{\boldsymbol{\Omega}}_{12M} & \widehat{\boldsymbol{\Omega}}_{13M} \\ \widehat{\boldsymbol{\Omega}}_{21M} & \widehat{\boldsymbol{\Omega}}_{22M} & \widehat{\boldsymbol{\Omega}}_{23M} \\ \widehat{\boldsymbol{\Omega}}_{31M} & \widehat{\boldsymbol{\Omega}}_{32M} & \widehat{\boldsymbol{\Omega}}_{33M} \end{pmatrix}^{-1},
$$

where

$$
\begin{aligned}
\widehat{\boldsymbol{\Omega}}_{11M} &= m_3(\mathbf{K}_1')^{-1}\mathbf{K}_1^{-1} \\
&\quad + \sum_{i=2}^{3}(m_3 - m_{i-1})\mathbf{B}_{i1}\mathbf{W}_{i11}^{-1}\mathbf{W}_{i12}(\mathbf{K}_i')^{-1}\mathbf{K}_i^{-1}\mathbf{W}_{i21}\mathbf{W}_{i11}^{-1}\mathbf{B}_{i1}', \\
\widehat{\boldsymbol{\Omega}}_{12M} &= \widehat{\boldsymbol{\Omega}}_{21M}' = -(m_3 - m_1)\mathbf{W}_{211}^{-1}\mathbf{W}_{212}(\mathbf{K}_2')^{-1}\mathbf{K}_2^{-1} \\
&\qquad\qquad + (m_3 - m_2)\mathbf{B}_{31}\mathbf{W}_{311}^{-1}\mathbf{W}_{312}(\mathbf{K}_3')^{-1}\mathbf{K}_3^{-1}\mathbf{W}_{321}\mathbf{W}_{311}^{-1}\mathbf{B}_{32}', \\
\widehat{\boldsymbol{\Omega}}_{13M} &= \widehat{\boldsymbol{\Omega}}_{31M}' = -(m_3 - m_2)\mathbf{B}_{31}\mathbf{W}_{311}^{-1}\mathbf{W}_{312}(\mathbf{K}_3')^{-1}\mathbf{K}_3^{-1}, \\
\widehat{\boldsymbol{\Omega}}_{22M} &= (m_3 - m_1)(\mathbf{K}_2')^{-1}\mathbf{K}_2^{-1} \\
&\quad + (m_3 - m_2)\mathbf{B}_{32}\mathbf{W}_{311}^{-1}\mathbf{W}_{312}(\mathbf{K}_3')^{-1}\mathbf{K}_3^{-1}\mathbf{W}_{321}\mathbf{W}_{311}^{-1}\mathbf{B}_{32}', \\
\widehat{\boldsymbol{\Omega}}_{23M} &= \widehat{\boldsymbol{\Omega}}_{32M}' = -(m_3 - m_2)\mathbf{B}_{32}\mathbf{W}_{311}^{-1}\mathbf{W}_{312}(\mathbf{K}_3')^{-1}\mathbf{K}_3^{-1}, \\
\widehat{\boldsymbol{\Omega}}_{33M} &= (m_3 - m_2)(\mathbf{K}_3')^{-1}\mathbf{K}_3^{-1}.
\end{aligned}
$$

Finally, the Bayesian estimator $\hat{\boldsymbol{\Sigma}}_J$ under the Jeffreys prior $\pi_J(\boldsymbol{\Sigma})$ can be calculated analytically,

$$
\widehat{\boldsymbol{\Sigma}}_J \equiv \begin{pmatrix} \widehat{\boldsymbol{\Psi}}_{11J} & \widehat{\boldsymbol{\Psi}}_{12J} & \widehat{\boldsymbol{\Psi}}_{13J} \\ \widehat{\boldsymbol{\Psi}}_{21J} & \widehat{\boldsymbol{\Psi}}_{22J} & \widehat{\boldsymbol{\Psi}}_{23J} \\ \widehat{\boldsymbol{\Psi}}_{31J} & \widehat{\boldsymbol{\Psi}}_{32J} & \widehat{\boldsymbol{\Psi}}_{33J} \end{pmatrix}^{-1},
$$

where

$$
\begin{aligned}
\widehat{\boldsymbol{\Omega}}_{11J} &= (m_3 + p_1 - p)\mathbf{W}_{111}^{-1} + p_2\mathbf{W}_{211}^{-1} \\
&\quad +(m_3 - m_1 + p_2 - p)\mathbf{W}_{211}^{-1}\mathbf{W}_{212}\mathbf{W}_{222\cdot 1}^{-1}\mathbf{W}_{221}\mathbf{W}_{211}^{-1} \\
&\quad +\mathbf{B}_{31}\left\{p_3\mathbf{W}_{311}^{-1} + (m_3 - m_2 + p_3 - p)\mathbf{W}_{311}^{-1}\mathbf{W}_{312}\mathbf{W}_{322\cdot 1}^{-1}\mathbf{W}_{321}\mathbf{W}_{311}^{-1}\right\}\mathbf{B}_{31}', \\
\widehat{\boldsymbol{\Omega}}_{21J} &= \widehat{\boldsymbol{\Omega}}_{12J}' \;=\; (m_3 - m_1 + p_2 - p)\mathbf{W}_{222\cdot 1}^{-1}\mathbf{W}_{221}\mathbf{W}_{211}^{-1} \\
&\quad +\mathbf{B}_{32}\left\{p_3\mathbf{W}_{311}^{-1} + (m_3 - m_2 + p_3 - p)\mathbf{W}_{311}^{-1}\mathbf{W}_{312}\mathbf{W}_{322\cdot 1}^{-1}\mathbf{W}_{321}\mathbf{W}_{311}^{-1}\right\}\mathbf{B}_{31}', \\
\widehat{\boldsymbol{\Omega}}_{31J} &= \widehat{\boldsymbol{\Omega}}_{13J}' \;=\; (m_3 - m_2 + p_3 - p)\mathbf{W}_{322\cdot 1}^{-1}\mathbf{W}_{321}\mathbf{W}_{311}^{-1}\mathbf{B}_{31}', \\
\widehat{\boldsymbol{\Omega}}_{32J} &= \widehat{\boldsymbol{\Omega}}_{23J}' \;=\; (m_3 - m_2 + p_3 - p)\mathbf{W}_{322\cdot 1}^{-1}\mathbf{W}_{321}\mathbf{W}_{311}^{-1}\mathbf{B}_{32}', \\
\widehat{\boldsymbol{\Omega}}_{22J} &= (m_3 - m_1 + p_2 - p)\mathbf{W}_{222\cdot 1}^{-1} \\
&\quad +\mathbf{B}_{32}\left\{p_3\mathbf{W}_{311}^{-1} + (m_3 - m_2 + p_3 - p)\mathbf{W}_{311}^{-1}\mathbf{W}_{312}\mathbf{W}_{322\cdot 1}^{-1}\mathbf{W}_{321}\mathbf{W}_{311}^{-1}\right\}\mathbf{B}_{32}', \\
\widehat{\boldsymbol{\Omega}}_{33J} &= (m_3 - m_2 + p_3 - p)\mathbf{W}_{322\cdot 1}^{-1}.
\end{aligned}
$$

Here $\mathbf{B}_{31}$ and $\mathbf{B}_{32}$ are given in (37).

Unfortunately, it is almost impossible to calculate the Bayesian estimator $\widehat{\boldsymbol{\Sigma}}_R$ analytically under the reference prior $\pi_R(\boldsymbol{\Sigma})$. Some simulation work will be given in the next section.

6 Simulation studies

In this section, we evaluate the performances of the MLE $\widehat{\boldsymbol{\Sigma}}_M$, the best $\mathcal{G}$-equivariant estimator $\widehat{\boldsymbol{\Sigma}}_B$, the Bayesian estimator $\widehat{\boldsymbol{\Sigma}}_J$ for the Jeffreys prior and the Bayesian estimator $\widehat{\boldsymbol{\Sigma}}_R$ for the reference prior. For a given $\widetilde{\mathbf{W}}$, we know how to compute the above estimators except $\widehat{\boldsymbol{\Sigma}}_R$ from Sects. 3 and 4.4. However, it is still impossible to compute the risks for the above four estimators analytically. So we will derive their approximate risks by simulation instead. For a given $\widetilde{\mathbf{W}}$, we first show how to simulate $\widehat{\boldsymbol{\Sigma}}_R$ because there is no closed form for this estimator.

Because the transformation from $\boldsymbol{\Sigma}^{-1}$ to $\boldsymbol{\Gamma}$ is one-to-one, we consider simulating from the posterior $p_R(\boldsymbol{\Gamma} \mid \widetilde{\mathbf{W}})$ of $\boldsymbol{\Gamma}$ and then obtaining the samples of $\boldsymbol{\Sigma}^{-1}$ from $p_R(\boldsymbol{\Sigma}^{-1} \mid \widetilde{\mathbf{W}})$ by using the transformation described in Sect. 4. From Lemma 4.2, we just need to simulate from the posteriors $p_R(\boldsymbol{\Gamma}_{11} \mid \widetilde{\mathbf{W}})$, $p_R(\boldsymbol{\Lambda}_2, \boldsymbol{\Gamma}_{22} \mid \widetilde{\mathbf{W}}), \ldots, p_R(\boldsymbol{\Lambda}_k, \boldsymbol{\Gamma}_{kk} \mid \widetilde{\mathbf{W}})$, respectively. By Lemma 4.2(b) and (c), the question turns out to simulate from the posterior $p_R(\boldsymbol{\Gamma}_{ii} \mid \widetilde{\mathbf{W}})$ of $\boldsymbol{\Gamma}_{ii}$ for each $i = 1, \ldots, k$. Yang and Berger (1994) provided a hit-and-run algorithm based on the exponential matrix transformation. Berger, Strawderman and Tang (2005) presented an easier Metropolis–Hastings algorithm and investigated some properties of the algorithm such as MCMC switching frequency and convergence. Here we adopt the Metropolis–Hastings algorithm to simulate from the marginal posterior distribution $p_R(\boldsymbol{\Gamma}_{ii} \mid \widetilde{\mathbf{W}})$ of $\boldsymbol{\Gamma}_{ii}$. For convenience, we will denote the marginal posterior distribution $p_R(\boldsymbol{\Gamma}_{ii} \mid \widetilde{\mathbf{W}})$ of $\boldsymbol{\Gamma}_{ii}$ as $f_i(\boldsymbol{\Gamma}_{ii})$, $i = 1, \ldots, k$ and $U(0, 1)$ as the uniform distribution between 0 and 1.

An algorithm for simulating from the marginal posterior distribution of $\boldsymbol{\Gamma}_{ii}$:

Step 0: Choose an initial value $\widehat{\boldsymbol{\Gamma}}_{ii}^{(0)}$ and set $l = 0$. Usually, we choose the MLE of $\boldsymbol{\Gamma}_{ii}$ as $\widehat{\boldsymbol{\Gamma}}_{ii}^{(0)}$, which is based on (29) and (15).
Step 1: Generate a random variable $\boldsymbol{\Gamma}_{ii} \sim g_i(\boldsymbol{\Gamma}_{ii})$ and another random variable $U \sim U(0, 1)$, where $g_i(\boldsymbol{\Gamma}_{ii})$ is a probing distribution, which will be given later.
Step 2: Compute

$$R_i = \frac{f_i(\boldsymbol{\Gamma}_{ii})}{f_i(\boldsymbol{\Gamma}_{ii}^{(l)})} \cdot \frac{g_i(\boldsymbol{\Gamma}_{ii}^{(l)})}{g_i(\boldsymbol{\Gamma}_{ii})}, \quad i = 1, \ldots, k.$$

Step 3: If $U > R_i$, set $\boldsymbol{\Gamma}_{ii}^{(l+1)} = \boldsymbol{\Gamma}_{ii}^{(l)}$, $l = l + 1$, and return to *Step 1*.
If $U \leq R_i$, set $\boldsymbol{\Gamma}_{ii}^{(l+1)} = \boldsymbol{\Gamma}_{ii}$, $l = l + 1$, and return to *Step 1*.

According to Berger, Strawderman and Tang (2005), an efficient probing distribution $g_i(\boldsymbol{\Gamma}_{ii})$ can be chosen as

$$g_i(\boldsymbol{\Gamma}_{ii}) \;\propto\; |\boldsymbol{\Gamma}_{ii}|^{-(m_k - m_{i-1} - q_{i-1} + p_i + 1)/2} \exp\left\{-\frac{1}{2} tr(\boldsymbol{\Gamma}_{ii}^{-1} \mathbf{W}_{i22\cdot 1})\right\},$$

which is $IW_{p_i}(\mathbf{W}_{i22\cdot 1}, m_k - m_{i-1} - q_{i-1})$, $i = 1, \ldots, k$. Therefore, the resulting R_i in Step 2 becomes

$$R_i = \frac{\prod_{1 \leq s < t \leq p_i} (d_{is}^{(l)} - d_{it}^{(l)})}{\prod_{1 \leq s < t \leq p_i} (d_{is} - d_{it})} \times \frac{|\boldsymbol{\Gamma}_{ii}|^{(p_i - 1)/2}}{|\boldsymbol{\Gamma}_{ii}^{(l)}|^{(p_i - 1)/2}}, \quad i = 1, \ldots, k.$$

Suppose that we now have the sample $\{\boldsymbol{\Gamma}_{ii}^{(0)}, \boldsymbol{\Gamma}_{ii}^{(1)}, \ldots, \boldsymbol{\Gamma}_{ii}^{(n)}\}$ by using the above algorithm for each $i = 1, \ldots, k$. Then, by Lemma 4.2(c), simulate $\boldsymbol{\Lambda}_i^{(l)} \sim N_{p_i, q_{i-1}}(\mathbf{W}_{i21}\mathbf{W}_{i11}^{-1}, \boldsymbol{\Gamma}_{ii}^{(l)} \otimes \mathbf{W}_{i11}^{-1})$, $l = 1, \ldots, n$, $i = 2, \ldots, k$. Consequently, by (31), we may have the sample $\{\boldsymbol{\Gamma}^{(1)}, \ldots, \boldsymbol{\Gamma}^{(n)}\}$ of size n being appropriately from the posterior $p_R(\boldsymbol{\Gamma} \mid \widetilde{\mathbf{W}})$.

Once we have $\{\boldsymbol{\Gamma}^{(1)}, \ldots, \boldsymbol{\Gamma}^{(n)}\}$, an MCMC sample $\{(\boldsymbol{\Sigma}^{-1})^{(1)}, \ldots, (\boldsymbol{\Sigma}^{-1})^{(n)}\}$ of size n for each $\widetilde{\mathbf{W}}$ can be obtained by using the transformation (30). So the Bayesian estimator $\widehat{\boldsymbol{\Sigma}}_R$ can be approximately calculated as

$$\widehat{\boldsymbol{\Sigma}}_R = \left\{\frac{1}{n} \sum_{j=1}^{n} (\boldsymbol{\Sigma}^{-1})^{(j)}\right\}^{-1}.$$

We report several examples for staircase pattern models with nine variables below. The results of each example are obtained based on 1,000 samples from the corresponding staircase pattern model. To get the Bayesian estimate of $\boldsymbol{\Sigma}$ under the reference prior for each of those 1,000 samples, we ran 10,000 cycles after 500 burn-in cycles by applying the algorithm described above. In fact, our simulation shows that taking 500 samples and running 5,500 cycling with 500 burn-in cycles is accurate enough to compute the risks of four estimators. Notice that $R(\widehat{\boldsymbol{\Sigma}}_R, \boldsymbol{\Sigma})$ depends on the true parameter $\boldsymbol{\Sigma}$ while $R(\widehat{\boldsymbol{\Sigma}}_M, \boldsymbol{\Sigma})$, $R(\widehat{\boldsymbol{\Sigma}}_B, \boldsymbol{\Sigma})$ and $R(\widehat{\boldsymbol{\Sigma}}_J, \boldsymbol{\Sigma})$ are independent of $\boldsymbol{\Sigma}$.

Example 1 For the case of $p_1 = 3, p_2 = 4, p_3 = 2, n_1 = 1, n_2 = 2, n_3 = 10$, we have $R(\widehat{\boldsymbol{\Sigma}}_M, \boldsymbol{\Sigma}) = 6.3325$, $R(\widehat{\boldsymbol{\Sigma}}_B, \boldsymbol{\Sigma}) = 4.4445$, $R(\widehat{\boldsymbol{\Sigma}}_J, \boldsymbol{\Sigma}) = 4.4472$, $R(\widehat{\boldsymbol{\Sigma}}_R, \boldsymbol{\Sigma}) = 4.1801$, when the true parameter $\boldsymbol{\Sigma} = \text{diag}(1, 2, \ldots, 9)$.

Example 2 For the case of $p_1 = 3, p_2 = 4, p_3 = 2, n_1 = 4, n_2 = 6, n_3 = 10$, we have $R(\widehat{\mathbf{\Sigma}}_M, \mathbf{\Sigma}) = 6.9642$, $R(\widehat{\mathbf{\Sigma}}_B, \mathbf{\Sigma}) = 4.9364$, $R(\widehat{\mathbf{\Sigma}}_J, \mathbf{\Sigma}) = 5.3014$, $R(\widehat{\mathbf{\Sigma}}_R, \mathbf{\Sigma}) = 4.7620$, when the true parameter $\mathbf{\Sigma} = \text{diag}(1, 2, \ldots, 9)$.

Example 3 For the case of $p_1 = 2, p_2 = 3, p_3 = 4, n_1 = 3, n_2 = 2, n_3 = 10$, we have $R(\widehat{\mathbf{\Sigma}}_M, \mathbf{\Sigma}) = 5.6764$, $R(\widehat{\mathbf{\Sigma}}_B, \mathbf{\Sigma}) = 4.2538$, $R(\widehat{\mathbf{\Sigma}}_J, \mathbf{\Sigma}) = 3.8374$, $R(\widehat{\mathbf{\Sigma}}_R, \mathbf{\Sigma}) = 3.5454$, when the true parameter $\mathbf{\Sigma} = \text{diag}(9, 8, \ldots, 1)$.

Theoretically, we have shown that the best equivariant estimator $\widehat{\mathbf{\Sigma}}_B$ beats the MLE $\widehat{\mathbf{\Sigma}}_M$ in any case. Our simulation results also show that the Bayesian estimator $\widehat{\mathbf{\Sigma}}_R$ is always the best one among the four estimators while the MLE $\widehat{\mathbf{\Sigma}}_M$ is the poorest. It must be admitted, however, that all these conclusions are quite tentative, being based on only a very limited simulation study.

7 Comments

In this paper, we apply two kinds of parameterizations to deal with the estimation of the multivariate normal covariance matrix with staircase pattern data. One is based on the Cholesky decomposition of the covariance matrix and is convenient to get the MLE and the best equivariant estimator with respect to the group of the lower triangular matrices under an invariant loss. The other is based on Bartlett decomposition and is convenient to deal with the Bayesian estimators with respect to the Jeffreys prior and the reference prior. Simulation study shows that the Bayesian estimator with respect to our reference prior is recommended under the Stein loss. Notice that the Bayesian estimator with respect to Yang-Berger's reference prior (Yang and Berger, 1994) performs well in estimation of the covariance matrix of the multivariate normal distribution with the complete data. However, it seems to be difficult to apply this prior to the staircase pattern model discussed in this paper. It is even hard to evaluate the performance of the resulting posterior distribution of the covariance matrix. Another advantage of our reference prior is that it can reduce the dimension of the variables in MCMC simulation. Because our reference prior depends on a group partition of variables, it is still unclear how to choose the best one from a class of reference priors. More study about this prior will be investigated in the future.

8 Appendix: Proofs

Proof of Theorem 4.1 For brevity, we just give the proof for $k = 2$. It is easy to extend it to the case for general k. Because the likelihood function of $\mathbf{\Gamma}$ is proportional to

$$\frac{1}{|\mathbf{\Sigma}_{11}|^{\frac{n_1}{2}}} \exp\left\{-\frac{1}{2}\sum_{j=1}^{m_1} \mathbf{Y}'_{j1}\mathbf{\Sigma}_{11}^{-1}\mathbf{Y}_{j1}\right\}$$
$$\times \frac{1}{|\mathbf{\Sigma}|^{\frac{n_2}{2}}} \exp\left\{-\frac{1}{2}\sum_{j=m_1+1}^{m_2} (\mathbf{Y}'_{j1}, \mathbf{Y}'_{j2})\mathbf{\Sigma}^{-1}\begin{pmatrix}\mathbf{Y}_{j1}\\ \mathbf{Y}_{j2}\end{pmatrix}\right\}$$

$$= \frac{1}{|\boldsymbol{\Gamma}_{11}|^{\frac{m_2}{2}}|\boldsymbol{\Gamma}_{22}|^{\frac{n_2}{2}}} \exp\left\{-\frac{1}{2}\sum_{j=1}^{m_1}\mathbf{Y}'_{j1}\boldsymbol{\Gamma}_{11}^{-1}\mathbf{Y}_{j1} - \frac{l^*}{2}\right\},$$

where

$$\begin{aligned} l^* &= \sum_{j=m_1+1}^{m_2} (\mathbf{Y}'_{j1}, \mathbf{Y}'_{j2}) \begin{pmatrix} \mathbf{I}_{p_1} & -\boldsymbol{\Lambda}'_2 \\ \mathbf{0} & \mathbf{I}_{p_2} \end{pmatrix} \begin{pmatrix} \boldsymbol{\Gamma}_{11}^{-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma}_{22}^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{I}_{p_1} & \mathbf{0} \\ -\boldsymbol{\Lambda}_2 & \mathbf{I}_{p_2} \end{pmatrix} \begin{pmatrix} \mathbf{Y}_{j1} \\ \mathbf{Y}_{j2} \end{pmatrix} \\ &= \sum_{j=m_1+1}^{m_2} (\mathbf{Y}'_{j1},\ \mathbf{Y}'_{j2} - \mathbf{Y}'_{j1}\boldsymbol{\Lambda}'_2) \begin{pmatrix} \boldsymbol{\Gamma}_{11}^{-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma}_{22}^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{Y}_{j1} \\ \mathbf{Y}_{j2} - \boldsymbol{\Lambda}_2\mathbf{Y}_{j1} \end{pmatrix} \\ &= \sum_{j=m_1+1}^{m_2} \mathbf{Y}'_{j1}\boldsymbol{\Gamma}_{11}^{-1}\mathbf{Y}_{j1} + \sum_{j=m_1+1}^{m_2} (\mathbf{Y}_{j2} - \boldsymbol{\Lambda}_2\mathbf{Y}_{j1})'\boldsymbol{\Gamma}_{22}^{-1}(\mathbf{Y}_{j2} - \boldsymbol{\Lambda}_2\mathbf{Y}_{j1}). \end{aligned}$$

The log-likelihood becomes

$$\begin{aligned} \log L = \text{const} - \frac{m_2}{2}\log|\boldsymbol{\Gamma}_{11}| - \frac{1}{2}\sum_{j=1}^{m_2}\mathbf{Y}'_{j1}\boldsymbol{\Gamma}_{11}^{-1}\mathbf{Y}_{j1} \\ -\frac{n_2}{2}\log|\boldsymbol{\Gamma}_{22}| - \frac{1}{2}\sum_{j=m_1+1}^{m_2}(\mathbf{Y}_{j2} - \boldsymbol{\Lambda}_2\mathbf{Y}_{j1})'\boldsymbol{\Gamma}_{22}^{-1}(\mathbf{Y}_{j2} - \boldsymbol{\Lambda}_2\mathbf{Y}_{j1}). \end{aligned}$$

Thus the Fisher information matrix $I(\boldsymbol{\theta})$ will have the following block structure,

$$\mathbf{I}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22} & \mathbf{A}_{23} \\ \mathbf{0} & \mathbf{A}'_{23} & \mathbf{A}_{33} \end{pmatrix}.$$

Similar to the calculation of $\mathbf{C}$ in Sect. 4 of McCulloch (1982), we can get

$$\mathbf{A}_{11} = \frac{m_2}{2}\mathbf{G}'_1(\boldsymbol{\Gamma}_{11}^{-1} \otimes \boldsymbol{\Gamma}_{11}^{-1})\mathbf{G}_1 \quad \text{and} \quad \mathbf{A}_{33} = \frac{n_2}{2}\mathbf{G}'_2(\boldsymbol{\Gamma}_{22}^{-1} \otimes \boldsymbol{\Gamma}_{22}^{-1})\mathbf{G}_2.$$

Now we calculate $\mathbf{A}_{22}$. Because

$$\frac{\partial \log L}{\partial \boldsymbol{\Lambda}_2} = \sum_{j=m_1+1}^{m_2} \boldsymbol{\Gamma}_{22}^{-1}(\mathbf{Y}_{j2} - \boldsymbol{\Lambda}_2\mathbf{Y}_{j1})\mathbf{Y}'_{j1},$$

we have

$$\frac{\partial \log L}{\partial \mathrm{vec}(\boldsymbol{\Lambda}_2)} = \sum_{j=m_1+1}^{m_2} (\mathbf{Y}_{j1} \otimes \boldsymbol{\Gamma}_{22}^{-1})(\mathbf{Y}_{j2} - \boldsymbol{\Lambda}_2\mathbf{Y}_{j1}).$$

Therefore,

$$\begin{aligned} A_{22} &= E\left[\frac{\partial \log L}{\partial \mathrm{vec}(\boldsymbol{\Lambda}_2)} \cdot \left(\frac{\partial \log L}{\partial \mathrm{vec}(\boldsymbol{\Lambda}_2)}\right)'\right] \\ &= n_2 E\big[(\mathbf{Y}_{m_2,1} \otimes \boldsymbol{\Gamma}_{22}^{-1})(\mathbf{Y}_{m_2,2} - \boldsymbol{\Lambda}_2\mathbf{Y}_{m_2,1})(\mathbf{Y}_{m_2,2} - \boldsymbol{\Lambda}_2\mathbf{Y}_{m_2,1})'(\mathbf{Y}'_{m_2,1} \otimes \boldsymbol{\Gamma}_{22}^{-1})\big] \\ &= n_2 E\big[(\mathbf{Y}_{m_2,1} \otimes \boldsymbol{\Gamma}_{22}^{-1})\boldsymbol{\Gamma}_{22}(\mathbf{Y}'_{m_2,1} \otimes \boldsymbol{\Gamma}_{22}^{-1})\big] \\ &= n_2 \boldsymbol{\Sigma}_1 \otimes \boldsymbol{\Gamma}_{22}^{-1}. \end{aligned}$$

In addition,

$$\frac{\partial \log L}{\partial \boldsymbol{\Gamma}_{22}} = -\frac{n_2}{2}\boldsymbol{\Gamma}_{22}^{-1} + \frac{1}{2}\boldsymbol{\Gamma}_{22}^{-1}\sum_{j=m_1+1}^{m_2}(\mathbf{Y}_{j2} - \boldsymbol{\Lambda}_2\mathbf{Y}_{j1})(\mathbf{Y}_{j2} - \boldsymbol{\Lambda}_2\mathbf{Y}_{j1})'\boldsymbol{\Gamma}_{22}^{-1}.$$

Similar to the calculation of **B** in Sect. 4 of McCulloch (1982), we have

$$\begin{aligned}
\mathbf{A}_{23}' &= E\left[\frac{\partial \log L}{\partial \mathrm{vech}(\boldsymbol{\Gamma}_{22})}\left(\frac{\partial \log L}{\partial \mathrm{vec}(\boldsymbol{\Lambda}_2)}\right)'\right] \\
&= \mathbf{G}_2' E\left[\frac{\partial \log L}{\partial \mathrm{vec}(\boldsymbol{\Gamma}_{22})}\left(\frac{\partial \log L}{\partial \mathrm{vec}(\boldsymbol{\Lambda}_2)}\right)'\right] \\
&= \mathbf{G}_2' E\Bigg[\mathrm{vec}\left\{-\frac{n_2}{2}\boldsymbol{\Gamma}_{22}^{-1} + \frac{1}{2}\boldsymbol{\Gamma}_{22}^{-1}\sum_{j=m_1+1}^{m_2}(\mathbf{Y}_{j2} - \boldsymbol{\Lambda}_2\mathbf{Y}_{j1})(\mathbf{Y}_{j2} - \boldsymbol{\Lambda}_2\mathbf{Y}_{j1})'\boldsymbol{\Gamma}_{22}^{-1}\right\} \\
&\qquad \times \sum_{j=m_1+1}^{m_2}(\mathbf{Y}_{j1} \otimes \boldsymbol{\Gamma}_{22}^{-1})(\mathbf{Y}_{j2} - \boldsymbol{\Lambda}_2\mathbf{Y}_{j1})\Bigg] \\
&= \mathbf{0}.
\end{aligned}$$

The last equality follows because $\mathbf{Y}_{j1}$ and $\mathbf{Y}_{j2} - \boldsymbol{\Lambda}_2\mathbf{Y}_{j1}$ are independent and $E(\mathbf{Y}_{j1}) = \mathbf{0}$. The proof of Theorem 4.1 is completed.

Proof of Lemma 4.1 The likelihood function of $\boldsymbol{\Gamma}$ is given by

$$\begin{aligned}
L(\boldsymbol{\Gamma}) &= |\boldsymbol{\Gamma}_{11}|^{-m_k/2}\exp\left\{-\frac{1}{2}\sum_{j=1}^{m_k}\mathbf{Y}_{j1}'\boldsymbol{\Gamma}_{11}^{-1}\mathbf{Y}_{j1}\right\} \\
&\quad \times |\boldsymbol{\Gamma}_{22}|^{-(m_k-m_1)/2}\exp\left\{-\frac{1}{2}\sum_{j=m_1+1}^{m_k}(\mathbf{Y}_{j2} - \boldsymbol{\Lambda}_2\mathbf{Y}_{j1})'\boldsymbol{\Gamma}_{22}^{-1}(\mathbf{Y}_{j2} - \boldsymbol{\Lambda}_2\mathbf{Y}_{j1})\right\} \\
&\quad \times |\boldsymbol{\Gamma}_{kk}|^{-(m_k-m_{k-1})/2}\exp\left\{-\frac{1}{2}\sum_{j=m_{k-1}+1}^{m_k}\left(\mathbf{Y}_{jk} - \boldsymbol{\Lambda}_k\begin{pmatrix}\mathbf{Y}_{j1}\\ \vdots \\ \mathbf{Y}_{j,k-1}\end{pmatrix}\right)'\right. \\
&\qquad \left. \times \boldsymbol{\Gamma}_{kk}^{-1}\left(\mathbf{Y}_{jk} - \boldsymbol{\Lambda}_k\begin{pmatrix}\mathbf{Y}_{j1}\\ \vdots \\ \mathbf{Y}_{j,k-1}\end{pmatrix}\right)\right\} \\
&= |\boldsymbol{\Gamma}_{11}|^{-m_k/2}\exp\left\{-\frac{1}{2}\mathrm{tr}(\boldsymbol{\Gamma}_{11}^{-1}\mathbf{W}_{111})\right\} \times \prod_{i=2}^{k}|\boldsymbol{\Gamma}_{ii}|^{-(m_k-m_{i-1})/2} \\
&\quad \times \exp\left[-\frac{1}{2}\mathrm{tr}\left\{\boldsymbol{\Gamma}_{ii}^{-1}(\mathbf{W}_{i22} - \boldsymbol{\Lambda}_i\mathbf{W}_{i12} - \mathbf{W}_{i21}\boldsymbol{\Lambda}_i' + \boldsymbol{\Lambda}_i\mathbf{W}_{i11}\boldsymbol{\Lambda}_i')\right\}\right]
\end{aligned}$$

and thus the posterior of $\boldsymbol{\Gamma}$

$$
\begin{aligned}
p_J(\boldsymbol{\Gamma} \mid \widetilde{\mathbf{W}}) &\propto |\boldsymbol{\Gamma}_{11}|^{-(m_k+2q_1-p+1)/2} \exp\left\{-\frac{1}{2}\mathrm{tr}(\boldsymbol{\Gamma}_{11}^{-1}\mathbf{W}_{111})\right\} \mathrm{d}\boldsymbol{\Gamma}_{11} \\
&\quad \times \prod_{i=2}^{k} |\boldsymbol{\Gamma}_{ii}|^{-(m_k-m_{i-1}+2q_i-p+1)/2} \\
&\quad \times \exp\left[-\frac{1}{2}\mathrm{tr}\left\{\boldsymbol{\Gamma}_{ii}^{-1}(\mathbf{W}_{i22} - \boldsymbol{\Lambda}_i\mathbf{W}_{i12} - \mathbf{W}_{i21}\boldsymbol{\Lambda}_i' + \boldsymbol{\Lambda}_i\mathbf{W}_{i11}\boldsymbol{\Lambda}_i')\right\}\right] \\
&= \prod_{i=1}^{k} |\boldsymbol{\Gamma}_{ii}|^{-(m_k-m_{i-1}+q_i+p_i-p+1)/2} \exp\left\{-\frac{1}{2}\mathrm{tr}(\boldsymbol{\Gamma}_{ii}^{-1}\mathbf{W}_{i22\cdot1})\right\} \\
&\quad \times \prod_{i=2}^{k} |\boldsymbol{\Gamma}_{ii}|^{-q_{i-1}/2} \\
&\quad \times \exp\left[-\frac{1}{2}\mathrm{tr}\left\{\boldsymbol{\Gamma}_{ii}^{-1}(\boldsymbol{\Lambda}_i - \mathbf{W}_{i21}\mathbf{W}_{i11}^{-1})\mathbf{W}_{i11}(\boldsymbol{\Lambda}_i - \mathbf{W}_{i21}\mathbf{W}_{i11}^{-1})'\right\}\right],
\end{aligned}
$$

which shows the parts (b)(c) and (d). Here the condition (14) is required to guarantee that $\mathbf{W}_{i11}$, and $\mathbf{W}_{i22\cdot1}$ are positive definite with probability one for each $i = 1, \ldots, k$.

For (a), we know that $p_J(\boldsymbol{\Gamma} \mid \widetilde{\mathbf{W}})$ is proper if and only if $p_J(\boldsymbol{\Gamma}_{ii} \mid \widetilde{\mathbf{W}})$ is proper for each $i = 1, \ldots, k$, and this is still guaranteed by the condition (14).

For (e), from (30), we get

$$
\begin{aligned}
\boldsymbol{\Sigma}^{-1} &= \begin{pmatrix} \mathbf{I}_{p_1} & \mathbf{0} & \cdots & \mathbf{0} \\ -\boldsymbol{\Gamma}_{21} & \mathbf{I}_{p_2} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ -\boldsymbol{\Gamma}_{k1} & -\boldsymbol{\Gamma}_{k2} & \cdots & \mathbf{I}_{p_k} \end{pmatrix}' \begin{pmatrix} \boldsymbol{\Gamma}_{11}^{-1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma}_{22}^{-1} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{\Gamma}_{kk}^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{I}_{p_1} & \mathbf{0} & \cdots & \mathbf{0} \\ -\boldsymbol{\Gamma}_{21} & \mathbf{I}_{p_2} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ -\boldsymbol{\Gamma}_{k1} & -\boldsymbol{\Gamma}_{k2} & \cdots & \mathbf{I}_{p_k} \end{pmatrix} \\
&\hat{=} \begin{pmatrix} \boldsymbol{\Psi}_{11} & \boldsymbol{\Psi}_{12} & \cdots & \boldsymbol{\Psi}_{1k} \\ \boldsymbol{\Psi}_{21} & \boldsymbol{\Psi}_{22} & \cdots & \boldsymbol{\Psi}_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Psi}_{k1} & \boldsymbol{\Psi}_{k2} & \cdots & \boldsymbol{\Psi}_{kk} \end{pmatrix},
\end{aligned}
$$

where

$$
\begin{aligned}
\boldsymbol{\Psi}_{ii} &= \boldsymbol{\Gamma}_{ii}^{-1} + \sum_{s=i+1}^{k} \boldsymbol{\Gamma}_{si}'\boldsymbol{\Gamma}_{ss}^{-1}\boldsymbol{\Gamma}_{si}, \quad i = 1, \ldots, k-1, \\
\boldsymbol{\Psi}_{kk} &= \boldsymbol{\Gamma}_{kk}^{-1}, \\
\boldsymbol{\Psi}_{ij} &= \boldsymbol{\Psi}_{ji}' \;=\; \boldsymbol{\Gamma}_{ii}^{-1}\boldsymbol{\Gamma}_{ij} + \sum_{s=i+1}^{k} \boldsymbol{\Gamma}_{si}'\boldsymbol{\Gamma}_{ss}^{-1}\boldsymbol{\Gamma}_{sj}, \quad 1 \le j < i \le k-1, \\
\boldsymbol{\Psi}_{ki} &= \boldsymbol{\Psi}_{ik}' \;=\; \boldsymbol{\Gamma}_{kk}^{-1}\boldsymbol{\Gamma}_{ki}, \quad i = 1, \ldots, k-1.
\end{aligned}
$$

So $E(\boldsymbol{\Sigma}^{-1} \mid \widetilde{\mathbf{W}})$ is finite if and only if the posterior means of the following items are finite: (i) $\boldsymbol{\Gamma}_{ii}^{-1}$, $i = 1, \ldots, k$; (ii) $\boldsymbol{\Gamma}_{ii}^{-1}\boldsymbol{\Lambda}_i$, $i = 2, \ldots, k$ and (iii) $\boldsymbol{\Lambda}_i'\boldsymbol{\Gamma}_{ii}^{-1}\boldsymbol{\Lambda}_i$, $i = 1, \ldots, k$. Based on Lemma 4.1(d), we get

$$E(\boldsymbol{\Gamma}_{ii}^{-1}\boldsymbol{\Lambda}_i \mid \boldsymbol{\Gamma}_{ii}, \widetilde{\mathbf{W}}) = \boldsymbol{\Gamma}_{ii}^{-1}\mathbf{W}_{i21}\mathbf{W}_{i11}^{-1},$$
$$E(\boldsymbol{\Lambda}_i'\boldsymbol{\Gamma}_{ii}^{-1}\boldsymbol{\Lambda}_i \mid \boldsymbol{\Gamma}_{ii}, \widetilde{\mathbf{W}}) = p_i\mathbf{W}_{i11}^{-1} + \mathbf{W}_{i11}^{-1}\mathbf{W}_{i12}\boldsymbol{\Gamma}_{ii}^{-1}\mathbf{W}_{i21}\mathbf{W}_{i11}^{-1}.$$

To complete the proof of part (e), we just need to show each posterior mean of $\boldsymbol{\Gamma}_{ii}^{-1}$ is finite, $i = 1, \ldots, k$. By Lemma 4.1(c), we will easily see that the marginal posterior of $\boldsymbol{\Gamma}_{ii}^{-1}$ is proper if the condition (14) holds and follows Wishart distribution with parameters $\mathbf{W}_{i22\cdot 1}$ and $m_k - m_{i-1} + q_i - p$. Consequently, it concludes

$$E(\boldsymbol{\Gamma}_{ii}^{-1} \mid \widetilde{\mathbf{W}}) = (m_k - m_{i-1} + q_i - p)\mathbf{W}_{i22\cdot 1}^{-1}, \quad i = 1, \ldots, k.$$

The results thus follow.

Proof of Lemma 4.2 Similar to the proof of Lemma 4.1, we have

$$\begin{aligned} p_R(\boldsymbol{\Gamma} \mid \widetilde{\mathbf{W}}) \propto \prod_{i=1}^{k} |\boldsymbol{\Gamma}_{ii}|^{-(m_k - m_{i-1} - q_{i-1} + 2)/2} \prod_{1 \le s < t \le p_i} \frac{1}{d_{is} - d_{it}} \\ \times \exp\left\{-\frac{1}{2}\mathrm{tr}(\boldsymbol{\Gamma}_{ii}^{-1}\mathbf{W}_{i22\cdot 1})\right\} \\ \times \prod_{i=2}^{k} |\boldsymbol{\Gamma}_{ii}|^{-q_{i-1}/2} \\ \times \exp\left[-\frac{1}{2}\mathrm{tr}\left\{\boldsymbol{\Gamma}_{ii}^{-1}(\boldsymbol{\Lambda}_i - \mathbf{W}_{i21}\mathbf{W}_{i11}^{-1})\mathbf{W}_{i11} \times (\boldsymbol{\Lambda}_i - \mathbf{W}_{i21}\mathbf{W}_{i11}^{-1})'\right\}\right], \end{aligned}$$

which concludes (b)(c) and (d). For (a), $p_R(\boldsymbol{\Gamma} \mid \widetilde{\mathbf{W}})$ is proper if and only if $p_R(\boldsymbol{\Gamma}_{ii} \mid \widetilde{\mathbf{W}})$ is proper for each $i = 1, \ldots, k$. As stated by Yang and Berger (1994), the marginal posterior of $\boldsymbol{\Gamma}_{ii}$

$$\begin{aligned} p_R(\boldsymbol{\Gamma}_{ii} \mid \widetilde{\mathbf{W}})\, d\boldsymbol{\Gamma}_{ii} \propto |\mathbf{D}_i|^{-(m_k - m_{i-1} - q_{i-1})/2 - 1} \\ \times \exp\left\{-\frac{1}{2}\mathrm{tr}(\mathbf{O}_i'\mathbf{D}_i^{-1}\mathbf{O}_i\mathbf{W}_{i22\cdot 1})\right\} \mathrm{d}\mathbf{D}_i\, \mathrm{d}\mathbf{H}_i \end{aligned}$$

and is proper because it is bounded by an inverse Gamma distribution if $n_k \ge p$, where d $\mathbf{H}_i$ denotes the *conditional invariant Haar measure* over the space of orthogonal matrices $\mathcal{O}_i = \{\mathbf{O}_i : \mathbf{O}_i'\mathbf{O}_i = \mathbf{I}_{p_i}\}$ (see Sect. 13.3 of Anderson (1984) for definition).

For (e), similar to the proof of part (e) of Lemma 4.1, we still just need to prove $E(\boldsymbol{\Gamma}_{ii}^{-1} \mid \widetilde{\mathbf{W}}) < \infty$ with respect to the posterior distribution of $\boldsymbol{\Gamma}_{ii}$ given by Lemma 4.2 (c), $i = 1, \ldots, k$. Note that $E(\boldsymbol{\Gamma}_{ii}^{-1} \mid \widetilde{\mathbf{W}}) < \infty$ if and only if $E\left[\{\mathrm{tr}\left(\boldsymbol{\Gamma}_{ii}^{-2}\right)\}^{1/2} \mid \widetilde{\mathbf{W}}\right] < \infty$. Considering that $\{\mathrm{tr}(\boldsymbol{\Gamma}_{ii}^{-2})\}^{1/2} = \left(\sum_{s=1}^{p_i} d_{is}^{-2}\right)^{1/2} \le p_i^{1/2} d_{ip_i}^{-1}$, where $d_{i1} \ge \cdots \ge d_{ip_i}$ are the eigenvalues of $\boldsymbol{\Gamma}_{ii}$, we have $E(\boldsymbol{\Gamma}_{ii}^{-1} \mid \widetilde{\mathbf{W}}) < \infty$ if $E(d_{ip_i}^{-1} \mid \widetilde{\mathbf{W}}) < \infty$. By a similar method used in the proof of Theorem 5 in Ni and Sun (2003), we easily get that $E(d_{ip_i}^{-1} \mid \widetilde{\mathbf{W}}) < \infty$ if the condition (14) holds. Hence part (e) follows. The proof is thus complete.

Acknowledgements The authors are grateful for the constructive comments from a referee. The authors would like to thank Dr. Karon F. Speckman for her carefully proof-reading this paper. The research was supported by National Science Foundation grant SES-0351523, and National Institute of Health grant R01-CA100760.

References

Anderson, T. W. (1957). Maximum likelihood estimates for a multivariate normal distribution when some observations are missing. *Journal of the American Statistical Association*, *52*, 200–203.
Anderson, T. W. (1984). *An introduction to multivariate statistical analysis*. New York: Wiley.
Anderson, T. W., Olkin, I. (1985). Maximum-likelihood estimation of the parameters of a multivariate normal distribution. *Linear Algebra and its Applications*, *70*, 147–171.
Bartlett, M. S. (1933). On the theory of statistical regression. *Proceedings of the Royal Society of Edinburgh*, *53*, 260–283.
Berger, J. O., Bernardo, J. M. (1992). On the development of reference priors. *Proceedings of the Fourth Valencia International Meeting, Bayesian Statistics*, *4*, 35–49.
Berger, J. O., Strawderman, W., Tang, D. (2005). Posterior propriety and admissibility of hyperpriors in normal hierarchical models. *The Annals of Statistics*, *33*, 606–646.
Brown, P. J., Le, N. D., Zidek, J. V. (1994). Inference for a covariance matrix. In: P. R. Freeman, A. F. M. Smith, (Eds.), Aspects of uncertainty. A tribute to D. V. Lindley. New York: Wiley
Daniels, M. J., and Pourahmadi, M. (2002). Bayesian analysis of covariance matrices and dynamic models for longitudinal data. *Biometrika 89*, 553–566.
Dey, D. K., Srinivasan, C. (1985). Estimation of a covariance matrix under Stein's loss. *The Annals of Statistics*, *13*, 1581–1591.
Eaton, M. L. (1970). Some problems in covariance estimatio (preliminary report). Tech. Rep. 49, Department of Statistics, Stanford University.
Eaton, M. L. (1989). *Group invariance applications in statistics*. Hayward: Institute of Mathematical Statistics.
Gupta, A. K., Nagar, D. K. (2000). *Matrix variate distributions*. New York: Chapman & Hall.
Haff, L. R. (1991). The variational form of certain Bayes estimators. *The Annals of Statistics*, *19*, 1163–1190.
Henderson, H. V., Searle, S. R. (1979). Vec and vech operators for matrices, with some uses in Jacobians and multivariate statistics. *The Canadian Journal of Statistics*, *7*, 65–81.
Jinadasa, K. G., Tracy, D. S. (1992). Maximum likelihood estimation for multivariate normal distribution with monotone sample. *Communications in Statistics, Part A – Theory and Methods*, *21*, 41–50.
Kibria, B. M. G., Sun, L., Zidek, J. V., Le, N. D. (2002). Bayesian spatial prediction of random space-time fields with application to mapping PM2.5 exposure. *Journal of the American Statistical Association*, *97*, 112–124.
Kiefer, J. (1957). Invariance, minimax sequential estimation, and continuous time process. *The Annals of Mathematical Statistics*, *28*, 573–601.
Konno, Y. (2001). Inadmissibility of the maximum likelihood estimator of normal covariance matrices with the lattice conditional independence. *Journal of Multivariate Analysis*, *79*, 33–51.
Little, R. J. A., Rubin, D. B. (1987). *Statistical analysis with missing data*. New York: Wiley.
Liu, C. (1993). Bartlett's decomposition of the posterior distribution of the covariance for normal monotone ignorable missing data. *Journal of Multivariate Analysis*, *46*, 198–206.
Liu, C. (1999). Efficient ML estimation of the multivariate normal distribution from incomplete data. *Journal of Multivariate Analysis*, *69*, 206–217.
Magnus, J. R., Neudecker, H. (1999). *Matrix differential calculus with applications in statistics and econometrics*. New York: Wiley.
McCulloch, C. E. (1982). Symmetric matrix derivatives with applications. *Journal of the American Statistical Association*, *77*, 679–682.
Ni, S., Sun, D. (2003). Noninformative priors and frequentist risks of bayesian estimators of vector-autoregressive models. *Journal of Econometrics*, *115*, 159–197.
Pourahmadi, M. (1999). Joint mean-covariance models with applications to longitudinal data: unconstrained parameterisation. *Biometrika*, *86*, 677–690.

Pourahmadi, M. (2000). Maximum likelihood estimation of generalised linear models for multivariate normal covariance matrix. *Biometrika 87*, 425–435.
Sun, D., Sun, X. (2005). Estimation of the multivariate normal precision and covariance matrices in a star-shape model. *Annals of the Institute of Statistical Mathematics 57*, 455–484.
Yang, R., Berger, J. O. (1994). Estimation of a covariance matrix using the reference prior. *The Annals of Statistics*, *22*, 1195–1211.