

Research Article

Posterior Propriety of an Objective Prior in a 4-Level Normal Hierarchical Model

Chengyuan Song ¹, Dongchu Sun,^{1,2} Kun Fan,¹ and Rongji Mu¹

¹KLATASDS-MOE, School of Statistics, East China Normal University, Shanghai 200062, China

²Department of Statistics, University of Nebraska-Lincoln, Lincoln, NE 68588, USA

Correspondence should be addressed to Chengyuan Song; songch1990@gmail.com

Received 19 October 2019; Revised 3 January 2020; Accepted 6 January 2020; Published 14 February 2020

Academic Editor: Luigi Rodino

Copyright © 2020 Chengyuan Song et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The use of hierarchical Bayesian models in statistical practice is extensive, yet it is dangerous to implement the Gibbs sampler without checking that the posterior is proper. Formal approaches to objective Bayesian analysis, such as the Jeffreys-rule approach or reference prior approach, are only implementable in simple hierarchical settings. In this paper, we consider a 4-level multivariate normal hierarchical model. We demonstrate the posterior using our recommended prior which is proper in the 4-level normal hierarchical models. A primary advantage of the recommended prior over other proposed objective priors is that it can be used at any level of a hierarchical model.

1. Introduction

Bayesian hierarchical models have a wide range of modern applications including engineering [1], astrophysics [2], economics [3], environmental sciences [4], climatology [5], survival analysis [6], and genetics [7]. It is wonderful here to stay, but hyperparameter priors are often chosen in a casual fashion. Statisticians often use improper priors to express ignorance or to provide good frequency properties. However, it is dangerous to implement the Gibbs sampler without checking that the posterior is proper. As Hobert and Casella [8] pointed, without proper precaution, simple noninformative priors can be misused, sometimes unknowingly, and lead to other difficulties, such as the non-convergence of the Gibbs sampler. Therefore, it is hazardous to skip the demonstration at the risk of making the inference from an improper posterior distribution. There are many examples of this in the statistical and other literatures. See especially, Wang et al. [9]; Hobert and Casella [8, 10]; Berger and Strawderman [11]; Berger et al. [12]; Speckman et al. [13]; Roy and Dey [14]; Michalak and Morris [15]; Ramos

et al. [16]; Tomazella et al. [6]; etc. The importance of the posterior propriety motivates us to explore it.

The normal hierarchical distribution has received enormous attention and is also of substantial importance in contemporary statistical theory and application. Ning [17] proposed a 2-level multivariate normal hierarchical model for the degradation data of multiple units with change point. Wang and Coit [18] handled the reliability prediction problem using a multivariate normal distribution model, considering multiple degradation measures. Heuristically, one could also build the multivariate normal hierarchical model with adding a normal distribution for unknown mean vectors to deal with this reliability prediction problem. The unknown parameters because of either convenience or a lack of prior information are often modeled with improper objective priors. Some references of objective priors can be found in Berger and Strawderman [11]; Berger [19]; Pollo et al. [20]; and Ferreira et al. [21]. However, formal approaches to objective Bayesian analysis, such as the Jeffreys-rule approach or reference prior approach, are only implementable in simple hierarchical settings. It is thus common to

use less formal approaches, such as utilizing formal priors from nonhierarchical models in hierarchical settings. However, this can be fraught with danger. For instance, nonhierarchical Jeffreys-rule priors for variances or covariance matrices result in improper posterior distributions if they are used at higher levels of a hierarchical model (see [12]).

Berger et al. [12] addressed the question of choice of hyperpriors in the following 2-level normal hierarchical model:

$$\mathbf{y} \sim N_p(\boldsymbol{\theta}, \mathbf{I}_p), \boldsymbol{\theta} \sim N_p(\mathbf{b}, \mathbf{Q}), \quad (1)$$

where

$$\begin{aligned} \mathbf{y}_{p \times 1} &= \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_m \end{pmatrix}, \\ \boldsymbol{\theta}_{p \times 1} &= \begin{pmatrix} \boldsymbol{\theta}_1 \\ \boldsymbol{\theta}_2 \\ \vdots \\ \boldsymbol{\theta}_m \end{pmatrix}, \\ \mathbf{b}_{p \times 1} &= \begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\beta} \\ \vdots \\ \boldsymbol{\beta} \end{pmatrix}, \\ \mathbf{Q}_{p \times p} &= \mathbf{I}_m \otimes \mathbf{V}, \end{aligned} \quad (2)$$

for $i = 1, 2, \dots, m$, in which the \mathbf{y}_i are the $k \times 1$ observation vectors and the $\boldsymbol{\theta}_i$ are the $k \times 1$ unknown mean vectors (thus $p = mk$), $\boldsymbol{\beta}$ is a $k \times 1$ unknown “hypermean” vector, and \mathbf{V} is an unknown $k \times k$ “hypercovariance matrix.” Some commonly priors for the hyperparameter $\boldsymbol{\beta}$ can be considered. For example, the constant prior and conjugate normal prior. For the prior of the covariance matrix \mathbf{V} , a large literature can be found. For example, the articles by Dey and Srinivasan [22]; Lin and Perlman [23]; Haff [24]; Yang and Berger [25]; Daniels and Kass [26, 27]; Ledoit and Wolf [28]; Sun and Berger [29]; Rajaratnam et al. [30]; Hoff [31]; And Berger et al. [12] explored posterior propriety and admissibility of various priors. But no overall conclusion was reached as to a specific prior in their paper. Following the seminal work of Berger et al. [12]; Berger et al. [32] rec-

ommended a particular objective prior for choice of $\pi(\boldsymbol{\beta})$ and $\pi(\mathbf{V})$:

$$\begin{aligned} \pi(\boldsymbol{\beta}) &\propto \frac{1}{[1 + \|\boldsymbol{\beta}\|^2]^{(k-1)/2}}, \\ \pi(\mathbf{V}) &= \frac{1}{|\mathbf{V}|^{1-1/(2k)} \prod_{i < j} (v_i - v_j)}, \end{aligned} \quad (3)$$

where $v_1 > v_2 > \dots > v_k$ are the ordered eigenvalues of \mathbf{V} . Their recommendation is based on considerations of the posterior propriety, admissibility, ease of implementation (including computational considerations), and performance. A primary advantage of the recommended prior over other proposed objective priors is that it can be used at any level of a hierarchical model. But the property of the resulting posterior is not clear, and it represents a very difficult question to answer. Michalak and Morris [15] pointed out that except in the simplest models when improper priors are used, it can be daunting and time-consuming to verify that the resulting posterior distribution is proper. While Berger et al. [32] suspected that the propriety holds in any level normal hierarchical models, they were not strictly able to prove it. They only showed that the posterior is proper in a 3-level hierarchical model. In this paper, we follow the story and consider the posterior propriety of the recommended prior in a 4-level normal hierarchical model. We demonstrate that the posterior using the recommended prior is still proper in the 4-level normal hierarchical model.

The structure of this paper is as follows. Section 2 presents a 4-level normal hierarchical model. We propose the recommended hyperpriors for “hypermean” vectors and “hypercovariance matrix.” In Section 3, we demonstrate that the posterior using the recommended prior is proper in the 4-level normal hierarchical model. In Section 4, we consider MCMC computation from the posterior. Section 5 gives the performance of this prior, in comparison with other objective priors that were studied in Berger et al. [12], presenting strong numerical evidence of the superiority of (3). Some concluding remarks are provided in Section 6.

2. A 4-Level Normal Hierarchical Model

Consider the following 4-level normal hierarchical model:

$$\begin{cases} \text{Level 1: } \mathbf{y}_i = \boldsymbol{\theta}_i + N_k(0, \mathbf{I}_k), & i = 1, 2, \dots, m, \\ \text{Level 2: } \boldsymbol{\theta}_i = \mathbf{Z}_i \boldsymbol{\beta} + N_k(0, \mathbf{V}), & \boldsymbol{\beta}^t = (\boldsymbol{\beta}_1^t, \dots, \boldsymbol{\beta}_s^t), \\ \text{Level 3: } \boldsymbol{\beta}_j = \mathbf{T}_j \boldsymbol{\eta} + N_p(0, \mathbf{W}), & j = 1, 2, \dots, s, \boldsymbol{\eta}^t = (\boldsymbol{\eta}_1^t, \dots, \boldsymbol{\eta}_n^t), \\ \text{Level 4: } \boldsymbol{\eta}_l = \boldsymbol{\xi} + N_q(0, \boldsymbol{\Sigma}), & l = 1, 2, \dots, n, \end{cases} \quad (4)$$

where the \mathbf{y}_i are the $k \times 1$ observation vectors; the $\boldsymbol{\theta}_i$ are the $k \times 1$ unknown mean vectors; the $\boldsymbol{\beta}_j$ are the $p \times 1$ unknown

vectors; the $\boldsymbol{\eta}_l$ are the $q \times 1$ unknown vectors; $\boldsymbol{\xi}$ is an unknown q -dimensional “hypermean” vectors; the \mathbf{Z}_i are the

$k \times sp$ known matrices; the \mathbf{T}_j are the $p \times nq$ known matrices; and \mathbf{V} , \mathbf{W} , and $\mathbf{\Sigma}$ are the unknown $k \times k$, $p \times p$, and $q \times q$ ‘‘hypercovariance matrix,’’ respectively. Following the assumption of Berger et al. [32], let $k = ps$ and $p = nq$. We also assume that $m \geq 2$, $s \geq 2$, $n \geq 2$, and $q \geq 1$.

There are four unknown parameters: $(\boldsymbol{\xi}, \mathbf{V}, \mathbf{W}$, and $\mathbf{\Sigma})$. From Berger et al. [32], the recommended (independent) priors for $(\boldsymbol{\xi}, \mathbf{V}, \mathbf{W}$, and $\mathbf{\Sigma})$ become

$$\begin{aligned} \pi(\boldsymbol{\xi}) &\propto \frac{1}{(1 + \|\boldsymbol{\xi}\|^2)^{(q-1)/2}}, \quad \boldsymbol{\xi} \in \mathbb{R}^q, \\ \pi(\mathbf{V}) &\propto \frac{1}{|\mathbf{V}|^{1-1/(2k)} \prod_{1 \leq i < j \leq k} (v_i - v_j)}, \quad \mathbf{V} > 0, \\ \pi(\mathbf{W}) &\propto \frac{1}{|\mathbf{W}|^{1-1/(2p)} \prod_{1 \leq i < j \leq p} (w_i - w_j)}, \quad \mathbf{W} > 0, \\ \pi(\mathbf{\Sigma}) &\propto \frac{1}{|\mathbf{\Sigma}|^{1-1/(2q)} \prod_{1 \leq i < j \leq q} (\sigma_i - \sigma_j)}, \quad \mathbf{\Sigma} > 0, \end{aligned} \quad (5)$$

where $v_1 > \dots > v_k > 0$ are the ordered eigenvalues of \mathbf{V} , $w_1 > \dots > w_p > 0$ are the ordered eigenvalues of \mathbf{W} , and $\sigma_1 > \dots > \sigma_q > 0$ are the ordered eigenvalues of $\mathbf{\Sigma}$.

As discussed by Berger et al. [12], the prior $\pi(\boldsymbol{\xi})$ can be represented by the following hierarchical structure:

$$\begin{aligned} \boldsymbol{\xi} | \lambda &\sim N_q(0, \lambda \mathbf{I}_q), \\ [\lambda] &\sim \lambda^{-(1/2)} e^{-1/2\lambda}. \end{aligned} \quad (6)$$

For both intuitive and technical reasons, it is convenient to write $\mathbf{V} = \mathbf{O}_1 \mathbf{V}^* \mathbf{O}_1'$, $\mathbf{W} = \mathbf{O}_2 \mathbf{W}^* \mathbf{O}_2'$, and $\mathbf{\Sigma} = \mathbf{O}_3 \mathbf{\Sigma}^* \mathbf{O}_3'$, where $\mathbf{V}^* = \text{diag}(v_1, \dots, v_k)$, $\mathbf{W}^* = \text{diag}(w_1, \dots, w_p)$, $\mathbf{\Sigma}^* = \text{diag}(\sigma_1, \dots, \sigma_q)$, and \mathbf{O}_i being the matrix of eigenvectors corresponding to \mathbf{V}^* , \mathbf{W}^* , and $\mathbf{\Sigma}^*$, $i = 1, 2, 3$, respectively. Consider the one-to-one transformation from \mathbf{V} to $(\mathbf{V}^*, \mathbf{O}_1)$ and rewrite the prior as

$$\pi(\mathbf{V}) d\mathbf{V} = \pi(\mathbf{V}^*, \mathbf{O}_1) d\mathbf{V}^* d\mathbf{O}_1, \quad (7)$$

where $d\mathbf{V} = \prod_{i \leq j} dv_{ij}$, $d\mathbf{V}^* = \prod_{i=1}^k dv_i$, and $d\mathbf{O}_1$ denotes the invariant Haar measure over the space of orthonormal matrices $\mathcal{O} = \{\mathbf{O}_1: \mathbf{O}_1' \mathbf{O}_1 = \mathbf{I}_k\}$ (see Anderson [33] for definition). From Farrell [34], the functional relationship between $\pi(\mathbf{V})$ and $\pi(\mathbf{V}^*, \mathbf{O}_1)$ is

$$\pi(\mathbf{V}^*, \mathbf{O}_1) = \pi(\mathbf{O}_1 \mathbf{V}^* \mathbf{O}_1') \prod_{i < j} (v_i - v_j). \quad (8)$$

Therefore, the prior for \mathbf{V} becomes the prior of $(\mathbf{V}^*, \mathbf{O}_1)$:

$$\pi(\mathbf{V}^*, \mathbf{O}_1) \propto \frac{1}{|\mathbf{V}^*|^{1-(1/2k)}}. \quad (9)$$

Use of the invariant prior on \mathbf{O}_1 (essentially a uniform prior over rotations) is natural and noncontroversial. This transformation reveals a significant difficulty of any prior that can be written as a function of $|\mathbf{V}|$; in the $(\mathbf{V}^*, \mathbf{O}_1)$ space, such priors contain the term $\prod_{i < j} (v_i - v_j)$, which gives low mass to close eigenvalues and hence effectively force the eigenvalues apart. It is a criticism of inverse Wishart and

Jeffreys priors. This is contrary to the common intuition, in that one often chooses a prior that pushes the eigenvalues closer together.

Similarly, one can obtain the prior of $(\mathbf{W}^*, \mathbf{O}_2)$ and $(\mathbf{\Sigma}^*, \mathbf{O}_3)$, given by

$$\begin{aligned} \pi(\mathbf{W}^*, \mathbf{O}_2) &\propto \frac{1}{|\mathbf{W}^*|^{1-(1/2p)}}, \\ \pi(\mathbf{\Sigma}^*, \mathbf{O}_3) &\propto \frac{1}{|\mathbf{\Sigma}^*|^{1-(1/2q)}}. \end{aligned} \quad (10)$$

3. Posterior Propriety

In this section, we consider the posterior propriety of the recommended prior in the 4-level normal hierarchical model (4).

3.1. The Case $q \geq 2$

Theorem 1. Consider the hierarchical Bayes model (4) with $q \geq 2$. Assume that \mathbf{Z} has rank ps and $\mathbf{Z}\mathbf{T}$ has rank nq . Then, the posterior distribution is always proper.

Proof. For the technical reason, we define

$$\begin{aligned} \mathbf{y} &= \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_m \end{pmatrix}_{mk \times 1}, \\ \boldsymbol{\theta} &= \begin{pmatrix} \boldsymbol{\theta}_1 \\ \boldsymbol{\theta}_2 \\ \vdots \\ \boldsymbol{\theta}_m \end{pmatrix}_{mk \times 1}, \\ \mathbf{1}_n &= \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}_{n \times 1}, \\ \mathbf{Z} &= \begin{pmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \\ \vdots \\ \mathbf{Z}_m \end{pmatrix}_{mk \times sp}, \\ \mathbf{T} &= \begin{pmatrix} \mathbf{T}_1 \\ \mathbf{T}_2 \\ \vdots \\ \mathbf{T}_s \end{pmatrix}_{sp \times nq}. \end{aligned} \quad (11)$$

Then, (4) is equivalent to

$$\begin{cases} \text{Level 1: } (\mathbf{y} | \boldsymbol{\theta}) \sim N_{mk}(\boldsymbol{\theta}, \mathbf{I}_{mk}), \\ \text{Level 2: } (\boldsymbol{\theta} | \boldsymbol{\beta}, \mathbf{V}) \sim N_{mk}(\mathbf{Z}\boldsymbol{\beta}, \mathbf{I}_m \otimes \mathbf{V}), \\ \text{Level 3: } (\boldsymbol{\beta} | \boldsymbol{\eta}, \mathbf{W}) \sim N_{sp}(\mathbf{T}\boldsymbol{\eta}, \mathbf{I}_s \otimes \mathbf{W}), \\ \text{Level 4: } (\boldsymbol{\eta} | \boldsymbol{\xi}, \boldsymbol{\Sigma}) \sim N_{nq}(\mathbf{I}_n \otimes \boldsymbol{\xi}, \mathbf{I}_n \otimes \boldsymbol{\Sigma}). \end{cases} \quad (12)$$

We marginalize out over $\boldsymbol{\theta}$, $\boldsymbol{\beta}$, and $\boldsymbol{\eta}$, one by one. It yields the marginal distribution of \mathbf{y} given $(\boldsymbol{\xi}, \mathbf{V}, \mathbf{W}, \text{ and } \boldsymbol{\Sigma})$ is

$$(\mathbf{y} | \boldsymbol{\xi}, \mathbf{V}, \mathbf{W}, \boldsymbol{\Sigma}) \sim N_{mk}(\mathbf{G}\boldsymbol{\xi}, \Delta), \quad (13)$$

where $\mathbf{G}_{mk \times q} = \mathbf{ZT}(\mathbf{I}_n \otimes \mathbf{I}_q)$ and

$$\Delta = \mathbf{I}_{mk} + \mathbf{I}_m \otimes \mathbf{V} + \mathbf{Z}(\mathbf{I}_s \otimes \mathbf{W})\mathbf{Z}' + \mathbf{ZT}(\mathbf{I}_n \otimes \boldsymbol{\Sigma})\mathbf{T}'\mathbf{Z}'. \quad (14)$$

Note that $\boldsymbol{\xi} | \lambda \sim N_p(0, \lambda \mathbf{I}_q)$, and marginalizing out $\boldsymbol{\xi}$ yields the marginal distribution of \mathbf{y} given $(\mathbf{V}, \mathbf{W}, \boldsymbol{\Sigma}, \text{ and } \lambda)$:

$$(\mathbf{y} | \mathbf{V}, \mathbf{W}, \boldsymbol{\Sigma}, \lambda) \sim N_{mk}(0, \Delta + \lambda \mathbf{G}\mathbf{G}'). \quad (15)$$

Therefore, the likelihood of \mathbf{y} given $(\mathbf{V}, \mathbf{W}, \boldsymbol{\Sigma}, \text{ and } \lambda)$ is

$$\begin{aligned} L(\mathbf{V}, \mathbf{W}, \boldsymbol{\Sigma}, \lambda) &\propto \frac{1}{|\Delta + \lambda \mathbf{G}\mathbf{G}'|^{1/2}} \exp\left(-\frac{1}{2}\mathbf{y}'(\Delta + \lambda \mathbf{G}\mathbf{G}')^{-1}\mathbf{y}\right) \\ &\leq \frac{1}{|\mathbf{I}_{mk} + \mathbf{I}_m \otimes \mathbf{V}|^{h_1} |\mathbf{I}_{mk} + \mathbf{Z}(\mathbf{I}_s \otimes \mathbf{W})\mathbf{Z}'|^{h_2} |\mathbf{I}_{mk} + \mathbf{ZT}(\mathbf{I}_n \otimes \boldsymbol{\Sigma})\mathbf{T}'\mathbf{Z}'|^{h_3} |\mathbf{I}_{mk} + \lambda \mathbf{G}\mathbf{G}'|^{h_4}}, \end{aligned} \quad (16)$$

where $h_1 = (1/2mk) + \epsilon$, $h_2 = (1/2sp) + \epsilon$, $h_3 = (1/2) - (1/2mk) - (1/2sp) - (1/2q) - 3\epsilon$, $h_4 = 1 - h_1 - h_2 - h_3 = (1/2q) + \epsilon$, and ϵ is a positive constant, assuring $h_3 \geq 0$. Note

that $h_1 + h_2 + h_3 + h_4 = 1/2$. Let $m(\mathbf{y})$ be the marginal density of \mathbf{y} . Clearly,

$$\begin{aligned} m(\mathbf{y}) &= \int \int \int \int L(\mathbf{V}, \mathbf{W}, \boldsymbol{\Sigma}, \lambda) \pi(\mathbf{V}) \pi(\mathbf{W}) \pi(\boldsymbol{\Sigma}) \pi(\lambda) d\mathbf{V} d\mathbf{W} d\boldsymbol{\Sigma} d\lambda \\ &\leq \int \frac{\pi(\mathbf{V})}{|\mathbf{I}_{mk} + \mathbf{I}_m \otimes \mathbf{V}|^{h_1}} d\mathbf{V} \times \int \frac{\pi(\mathbf{W})}{|\mathbf{I}_{mk} + \mathbf{Z}(\mathbf{I}_s \otimes \mathbf{W})\mathbf{Z}'|^{h_2}} d\mathbf{W} \times \int \frac{\pi(\lambda)}{|\mathbf{I}_{mk} + \lambda \mathbf{G}\mathbf{G}'|^{h_4}} d\lambda \times \int \frac{\pi(\boldsymbol{\Sigma})}{|\mathbf{I}_{mk} + \mathbf{ZT}(\mathbf{I}_n \otimes \boldsymbol{\Sigma})\mathbf{T}'\mathbf{Z}'|^{h_3}} d\boldsymbol{\Sigma}. \end{aligned} \quad (17)$$

Note that

$$\int \frac{\pi(\mathbf{V})}{|\mathbf{I}_{mk} + \mathbf{I}_m \otimes \mathbf{V}|^{h_1}} d\mathbf{V} = \int \frac{\mathbf{1}_{\{v_1 > v_2 > \dots > v_k\}}}{|\mathbf{I}_k + \mathbf{V}|^{(1/2k) + m\epsilon} |\mathbf{V}|^{1 - (1/2k)} \prod_{1 \leq i < j \leq k} (v_i - v_j)} d\mathbf{V}. \quad (18)$$

From relationship (8) between \mathbf{V} and $(\mathbf{V}^*, \mathbf{O}_1)$,

$$\int \frac{\pi(\mathbf{V})}{|\mathbf{I}_{mk} + \mathbf{I}_m \otimes \mathbf{V}|^{h_4}} d\mathbf{V} = \iint \frac{\mathbf{1}_{\{v_1 > v_2 > \dots > v_k\}}}{|\mathbf{I}_k + \mathbf{V}^*|^{(1/2k) + m\epsilon} |\mathbf{V}^*|^{1 - (1/2k)}} d\mathbf{V}^* d\mathbf{O}_1 \leq \prod_{i=1}^k \int_0^\infty \frac{1}{(1 + v_i)^{(1/2k) + m\epsilon} v_i^{1 - (1/2k)}} dv_i < \infty. \quad (19)$$

Next, consider the integration over \mathbf{W} . From the matrix identity $|\mathbf{A} + \mathbf{X}\mathbf{B}\mathbf{X}'| = |\mathbf{A}||\mathbf{B}||\mathbf{B}^{-1} + \mathbf{X}'\mathbf{A}^{-1}\mathbf{X}|$, we obtain

$$|\mathbf{I}_{mk} + \mathbf{Z}(\mathbf{I}_s \otimes \mathbf{W})\mathbf{Z}'| = |\mathbf{W}|^s \left| (\mathbf{I}_s \otimes \mathbf{W}^{-1}) + \mathbf{Z}'\mathbf{Z} \right|. \quad (20)$$

Let $z_0 > 0$ be the smallest eigenvalue of $\mathbf{Z}'\mathbf{Z}$, then

$$\left| (\mathbf{I}_s \otimes \mathbf{W}^{-1}) + \mathbf{Z}'\mathbf{Z} \right| \geq \left| \mathbf{W}^{-1} + z_0 \mathbf{I}_{sp} \right|^s. \quad (21)$$

Thus,

$$\left| \mathbf{I}_{mk} + \mathbf{Z}(\mathbf{I}_s \otimes \mathbf{W})\mathbf{Z}' \right| \geq \left| \mathbf{I}_p + z_0 \mathbf{W} \right|^s. \quad (22)$$

Using the above inequality, it yields

$$\begin{aligned} &\int \frac{\pi(\mathbf{W})}{|\mathbf{I}_{mk} + \mathbf{Z}(\mathbf{I}_s \otimes \mathbf{W})\mathbf{Z}'|^{h_2}} d\mathbf{W} \\ &\leq \int \frac{1}{|\mathbf{I}_k + z_0 \mathbf{W}^*|^{(1/2p) + s\epsilon} |\mathbf{W}^*|^{1 - (1/2p)}} d\mathbf{W}^* \end{aligned} \quad (23)$$

$$\leq \prod_{j=1}^p \int_0^\infty \frac{1}{(1 + z_0 w_j)^{(1/2p) + s\epsilon} w_j^{1 - (1/2p)}} dw_j < \infty.$$

Let $u_1 \geq \dots \geq u_q > 0$ be the eigenvalues of $\mathbf{G}'\mathbf{G}$. Using the matrix identity $|\mathbf{A} + \mathbf{X}\mathbf{B}\mathbf{X}'| = |\mathbf{A}||\mathbf{B}||\mathbf{B}^{-1} + \mathbf{X}'\mathbf{A}^{-1}\mathbf{X}|$ again, it follows that

$$|\mathbf{I}_{mk} + \lambda \mathbf{G}\mathbf{G}'| = \left| \lambda \mathbf{I}_q \right| \left| \frac{1}{\lambda} \mathbf{I}_q + \mathbf{G}'\mathbf{G} \right| \geq (1 + u_p \lambda)^q. \quad (24)$$

Therefore,

$$\int \frac{1}{|\mathbf{I}_{mk} + \lambda \mathbf{G}\mathbf{G}'|^h} \pi(\lambda) d\lambda \leq \int_0^\infty \frac{1}{\lambda^{1/2} (1 + u_p \lambda)^{(1/2)+q\epsilon}} e^{-2/\lambda} d\lambda < \infty. \quad (25)$$

Last, to consider the integration over Σ . Let t_0 be the smallest eigenvalue of $\mathbf{T}'\mathbf{Z}'\mathbf{Z}\mathbf{T}$. Then,

$$|\mathbf{I}_{mk} + \mathbf{Z}\mathbf{T}(\mathbf{I}_n \otimes \Sigma)\mathbf{T}'\mathbf{Z}'| \geq |\mathbf{I}_q + t_0 \Sigma^*|^n. \quad (26)$$

Thus,

$$\begin{aligned} \int \frac{\pi(\Sigma)}{|\mathbf{I}_{mk} + \mathbf{Z}\mathbf{T}(\mathbf{I}_n \otimes \Sigma)\mathbf{T}'\mathbf{Z}'|^h} d\Sigma &\leq \int \frac{1}{|\mathbf{I}_k + t_0 \Sigma^*|^{n((1/2)-(1/2mk)-(1/2sp)-(1/2q)-3\epsilon)} |\Sigma^*|^{1-(1/2q)}} d\Sigma^* \\ &\leq \prod_{l=1}^q \int_0^\infty \frac{1}{(1 + t_0 \sigma_l)^{n((1/2)-(1/2mk)-(1/2sp)-(1/2q)-3\epsilon)} \sigma_l^{1-(1/2q)}} d\sigma_l, \end{aligned} \quad (27)$$

which is finite if

$$n \left(\frac{1}{2} - \frac{1}{2mk} - \frac{1}{2sp} - \frac{1}{2q} - 3\epsilon \right) + 1 - \frac{1}{2q} > 1. \quad (28)$$

Since ϵ can be chosen arbitrarily small, the integration over Σ is finite if

$$n(q-1) > 1 + \frac{nq}{mk} + \frac{nq}{sp} = 1 + \frac{1}{ms} + \frac{1}{s}. \quad (29)$$

Since q, n, m , and $s \geq 2$, (29) is true. The theorem is proved. \square

3.2. The Case $q = 1$

Theorem 2. Consider the hierarchical Bayes model (4) with $q = 1$. Assume that \mathbf{Z} has rank ps and $\mathbf{Z}\mathbf{T}$ has rank n . Then, the posterior distribution is proper if

$$\frac{1}{m} + \frac{1}{s} + \frac{2}{n} + \frac{1}{sn} + \frac{1}{msn} < 1. \quad (30)$$

Proof. In the case $q = 1$, note that Σ is just a variance, so that

$$\Delta = \mathbf{I}_{mk} + \mathbf{I}_m \otimes \mathbf{V} + \mathbf{Z}(\mathbf{I}_s \otimes \mathbf{W})\mathbf{Z}' + \Sigma \mathbf{Z}\mathbf{T}\mathbf{T}'\mathbf{Z}'. \quad (31)$$

The prior of ξ becomes

$$\pi(\xi) \propto 1. \quad (32)$$

Integrating out ξ in (22) with its constant prior and dropping all exponential terms of the likelihood (as they are less than one), the marginal likelihood for \mathbf{V} and W satisfies

$$L(\mathbf{V}, W) < \frac{1}{|\Delta|^{1/2} (\mathbf{G}'\Delta^{-1}\mathbf{G})^{1/2}}. \quad (33)$$

Thus, for the posterior propriety, we need to verify

$$\int \int \frac{1}{|\Delta|^{1/2} (\mathbf{G}'\Delta^{-1}\mathbf{G})^{1/2}} \pi(\mathbf{V})\pi(W)\pi(\Sigma) d\Sigma d\mathbf{V} dW < \infty. \quad (34)$$

Denote t_1 be the largest eigenvalue of $\mathbf{T}'\mathbf{Z}'\mathbf{Z}\mathbf{T}$. Let $\lambda_0(\mathbf{A})$ and $\lambda_1(\mathbf{A})$ denote the minimum and maximum eigenvalue of \mathbf{A} , respectively. Then,

$$\begin{aligned} \lambda_1(\Delta) &\leq \lambda_1 \left((1 + v_1)\mathbf{I}_{mk} + w_1 \mathbf{Z}\mathbf{Z}' + \Sigma \mathbf{Z}\mathbf{T}\mathbf{T}'\mathbf{Z}' \right) \\ &\leq 1 + v_1 + z_1 w_1 + t_1 \Sigma \\ &\leq (1 + v_1)(1 + z_1 w_1)(1 + t_1 \Sigma). \end{aligned} \quad (35)$$

Noting that \mathbf{G} is an $mk \times 1$ vector, it follows that

$$\begin{aligned} \mathbf{G}'\Delta^{-1}\mathbf{G} &> \lambda_0(\Delta^{-1})\mathbf{G}'\mathbf{G} = \lambda_1^{-1}(\Delta)\mathbf{G}'\mathbf{G} \\ &> \frac{1}{(1 + v_1)(1 + z_1 w_1)(1 + t_1 \Sigma)} \mathbf{G}'\mathbf{G}. \end{aligned} \quad (36)$$

Let $b_1 = (1/2mk) + (1/2m) + \epsilon$ and $b_2 = (1/2sp) + (1/2s) + \epsilon$, where ϵ is a small positive constant, satisfying $(1/2) - b_1 - b_2 \geq 0$. Then,

$$\begin{aligned} |\Delta|^{1/2} &\geq |\mathbf{I}_{mk} + \mathbf{I}_m \otimes \mathbf{V}|^{b_1} |\mathbf{I}_{mk} + \mathbf{Z}(\mathbf{I}_s \otimes \mathbf{W})\mathbf{Z}'|^{b_2} |\mathbf{I}_{mk} + \Sigma \mathbf{Z}\mathbf{T}\mathbf{T}'\mathbf{Z}'|^{(1/2)-b_1-b_2} \\ &\geq |\mathbf{I}_k + \mathbf{V}|^{mb_1} |\mathbf{I}_p + z_0 \mathbf{W}|^{sb_2} (1 + t_0 \Sigma)^{n((1/2)-b_1-b_2)}. \end{aligned} \quad (37)$$

Clearly,

$$m(\mathbf{y}) \leq C \iiint \frac{[(1 + v_1)(1 + z_1 w_1)(1 + t_1 \Sigma)]^{1/2} \pi(\mathbf{V})\pi(W)\pi(\Sigma)}{|\mathbf{I}_k + \mathbf{V}|^{mb_1} |\mathbf{I}_p + z_0 \mathbf{W}|^{sb_2} (1 + t_0 \Sigma)^{n(1/2-b_1-b_2)}} d\mathbf{V} dW d\Sigma, \quad (38)$$

where $C = (\mathbf{G}'\mathbf{G})^{-(1/2)}$. Note that

$$\begin{aligned} \int \frac{(1+v_1)^{1/2}}{|\mathbf{I}_k + \mathbf{V}|^{mb_1}} \pi(\mathbf{V}) d\mathbf{V} &\leq \int \frac{(1+v_1)^{1/2}}{|\mathbf{I}_k + \mathbf{V}|^{mb_1} |\mathbf{V}|^{1-(1/2k)}} d\mathbf{V}^* \\ &\leq \int_0^\infty \frac{1}{(1+v_1)^{mb_1-(1/2)} v_1^{1-(1/2k)}} dv_1 \times \prod_{i=2}^k \int_0^\infty \frac{1}{(1+v_i)^{mb_1} v_i^{1-1/2k}} dv_i, \end{aligned} \quad (39)$$

which is finite since $mb_1 - (1/2) + 1 - (1/2k) = 1 + m\epsilon > 1$. Similarly,

$$\begin{aligned} \int \frac{(1+z_1 w_1)^{1/2}}{|\mathbf{I}_k + \mathbf{W}|^{sb_2}} \pi(\mathbf{W}) d\mathbf{W} &\leq \int_0^\infty \frac{(1+z_1 w_1)^{1/2}}{(1+w_1)^{sb_2} w_1^{1-(1/2p)}} dw_1 \\ &\times \prod_{j=2}^p \int_0^\infty \frac{1}{(1+w_j)^{sb_2} w_j^{1-(1/2p)}} dw_j, \end{aligned} \quad (40)$$

which is finite since $sb_2 + 1 - (1/2p) - (1/2) = 1 + s\epsilon > 1$. Next, consider the integration over Σ . It follows that

$$\int_0^\infty \frac{(1+t_1 \Sigma)^{1/2}}{(1+t_0 \Sigma)^{n((1/2)-b_1-b_2)} \Sigma^{1/2}} d\Sigma < \infty, \quad (41)$$

if $n((1/2) - b_1 - b_2) = n((1/2) - (1/2mk) - (1/2m) - (1/2sp) - (1/2s) - 2\epsilon) > 1$. Since ϵ can be chosen arbitrarily small, $p = n$ and $k = sn$; this is true if

$$\frac{1}{m} + \frac{1}{s} + \frac{2}{n} + \frac{1}{sn} + \frac{1}{msn} < 1. \quad (42)$$

This completes the proof. \square

4. Computation

In this section, we consider the MCMC computation from the posterior arising from the model (4). The normal hierarchical models are typically handled today by Gibbs sampling. One difficulty of computation is to sample the covariance matrix efficiently. The main new development discussed in Berger et al. [35] is a new and efficient computational algorithm for dealing with priors on covariance matrices as in (3). It overcomes the computational bottleneck mentioned in the introduction.

Fact 1. Here are the full conditional distributions:

- (a) For $i = 1, \dots, m$, the conditional posterior of θ_i given β and \mathbf{V} is the usual conjugate posterior density:

$$N_k \left((\mathbf{I}_k + \mathbf{V}^{-1})^{-1} (\mathbf{y}_i + \mathbf{V}^{-1} \mathbf{Z}_i \beta), (\mathbf{I}_k + \mathbf{V}^{-1})^{-1} \right). \quad (43)$$

- (b) Define

$$\begin{aligned} \mathbf{M} &= \mathbf{I}_s \otimes \mathbf{W}^{-1} + [(\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}' (\mathbf{I}_m \otimes \mathbf{V}) \mathbf{Z} (\mathbf{Z}'\mathbf{Z})^{-1}]^{-1}, \\ \mathbf{u} &= (\mathbf{I}_s \otimes \mathbf{W}^{-1}) \mathbf{T} \eta + [(\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}' (\mathbf{I}_m \otimes \mathbf{V}) \mathbf{Z} (\mathbf{Z}'\mathbf{Z})^{-1}]^{-1} (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}' \theta. \end{aligned} \quad (44)$$

The conditional posterior of β given $(\theta, \eta, \mathbf{V}, \mathbf{W})$ is $N_{sp}(\mathbf{M}^{-1} \mathbf{u}, \mathbf{M}^{-1})$.

- (c) Defining $r_1 = (m/2) + 1 - (1/3k)$, the conditional posterior density of \mathbf{V} for given (θ, β) can be written as

$$\pi(\mathbf{V} | \theta, \beta) \propto \frac{1}{|\mathbf{V}|^{r_1} \prod_{i < j} (v_i - v_j)} \text{etr} \left\{ -\frac{1}{2} \mathbf{V}^{-1} \mathbf{H}_1 \right\}. \quad (45)$$

Here and in the following $\text{etr}(\mathbf{A})$ represents $\exp(\text{trace}(\mathbf{A}))$ for a squared matrix \mathbf{A} , and

$$\mathbf{H}_1 = \mathbf{H}_1(\theta, \beta) = \sum_{i=1}^m (\theta_i - \mathbf{Z}'_i \beta) (\theta_i - \mathbf{Z}'_i \beta)'. \quad (46)$$

- (d) Define

$$\begin{aligned} \mathbf{K} &= \mathbf{I}_n \otimes \Sigma^{-1} + [(\mathbf{T}'\mathbf{T})^{-1} \mathbf{T}' (\mathbf{I}_s \otimes \mathbf{W}) \mathbf{T} (\mathbf{T}'\mathbf{T})^{-1}]^{-1}, \\ \mathbf{v} &= \mathbf{1}_n \otimes \Sigma^{-1} \xi + [(\mathbf{T}'\mathbf{T})^{-1} \mathbf{T}' (\mathbf{I}_s \otimes \mathbf{W}) \mathbf{T} (\mathbf{T}'\mathbf{T})^{-1}]^{-1} (\mathbf{T}'\mathbf{T})^{-1} \mathbf{T}' \beta. \end{aligned} \quad (47)$$

The conditional posterior of η given $(\beta, \xi, \mathbf{W}, \Sigma)$ is $N_{nq}(\mathbf{K}^{-1} \mathbf{v}, \mathbf{K}^{-1})$.

- (e) Defining $r_2 = (s/2) + 1 - (1/2p)$, the conditional posterior density of \mathbf{W} for given (β, η) can be written as

$$\pi(\mathbf{W} | \beta, \eta) \propto \frac{1}{|\mathbf{W}|^{r_2} \prod_{i < j} (w_i - w_j)} \text{etr} \left\{ -\frac{1}{2} \mathbf{W}^{-1} \mathbf{H}_2 \right\}, \quad (48)$$

where $\mathbf{H}_2 = \sum_{j=1}^s (\beta_j - \mathbf{T}'_j \eta) (\beta_j - \mathbf{T}'_j \eta)'$.

- (f) To compute with the recommended prior for ξ , use the equivalent representation as follows:

$$\begin{aligned} \xi | \lambda &\sim N_q(0, \lambda \mathbf{I}_q), \\ [\lambda] &\propto \lambda^{-1/2} e^{-1/2\lambda}. \end{aligned} \quad (49)$$

$$[\lambda | \xi] \propto \frac{1}{\lambda^{(q+1)/2}} \exp \left\{ -\frac{1 + \|\xi\|^2}{2\lambda} \right\}. \quad (50)$$

$$N_q \left(\left(\frac{1}{\lambda} \mathbf{I}_q + n \Sigma^{-1} \right)^{-1} \Sigma^{-1} \sum_{i=1}^n \eta_i, \left(\frac{1}{\lambda} \mathbf{I}_q + n \Sigma^{-1} \right)^{-1} \right). \quad (51)$$

(g) Defining $r_3 = (n/2) + 1 - (1/2q)$, the conditional posterior density of Σ for given (β, η) can be written as

$$\pi(\Sigma | \beta, \eta) \propto \frac{1}{|\Sigma|^{r_3} \prod_{i < j} (\sigma_i - \sigma_j)} \text{etr} \left\{ -\frac{1}{2} \Sigma^{-1} \mathbf{H}_3 \right\}, \quad (52)$$

where $\mathbf{H}_3 = \sum_{l=1}^n (\eta_l - \xi)(\eta_l - \xi)'$.

(i) For sample λ from its full conditional, the Inverse Gamma $((q-1)/2, (1 + \|\xi\|^2)/2)$ density

If $q = 1$, this step is not needed as the hyperprior, for ξ is constant.

(ii) Given $(\lambda, \Sigma, \eta_i)$, Gibbs sampling of ξ can be done from its full conditional, which is (when $q = 1$, set $\lambda = \infty$)

Sampling of θ, β, η , and ξ can simply be carried out with a Gibbs step, as its full conditional will be a normal distribution. To sample the covariance matrix \mathbf{V}, \mathbf{W} , and Σ from Fact 1, the Metropolis–Hastings Algorithm [36] and Hit-and-Run Method Chen and Schmeiser [37] could be used, based on proposal distributions that generate full-candidate matrices. But the two well-known methods are both inefficient, especially for high-dimensional data. Fortunately, Berger et al. [35] proposed a new and efficient computational algorithm for sampling the covariance matrix \mathbf{V}, \mathbf{W} , and Σ from Fact 1, respectively.

Take sampling \mathbf{V} from (45) as an example. For the conditional density of \mathbf{V} given in (45), we use the eigenvalue-vector decomposition $\mathbf{V} = \mathbf{O}_1 \mathbf{V}^* \mathbf{O}_1'$, where \mathbf{O}_1 is the orthogonal and \mathbf{V}^* is the diagonal matrix of ordered eigenvalues. For \mathbf{H}_1 defined in (46), note that $\text{tr}(\mathbf{V}^{-1} \mathbf{H}_1) = \text{tr}(\mathbf{O}_1 (\mathbf{V}^*)^{-1} \mathbf{O}_1' \mathbf{H}_1) = \text{tr}((\mathbf{V}^*)^{-1} \mathbf{O}_1' \mathbf{H}_1 \mathbf{O}_1)$, so the conditional density of \mathbf{V} given in (45) can be transformed to

$$\pi(\mathbf{V}^*, \mathbf{O}_1 | \theta, \beta) \propto \frac{1}{|\mathbf{V}^*|^{r_1}} \text{etr} \left\{ -\frac{1}{2} (\mathbf{V}^*)^{-1} \mathbf{O}_1' \mathbf{H}_1 \mathbf{O}_1 \right\} 1_{\{v_1 > v_2 > \dots > v_k\}}. \quad (53)$$

Gibbs sampling of \mathbf{V}^* : following Lemma 1 in Berger et al. [12], i.e.,

$$\iint g(\mathbf{O}_1 \mathbf{V}^* \mathbf{O}_1') 1_{\{v_1 > \dots > v_k\}} d\mathbf{V}^* d\mathbf{O}_1 = \frac{1}{k!} \iint g(\mathbf{O}_1 \mathbf{V}^* \mathbf{O}_1') d\mathbf{V}^* d\mathbf{O}_1, \quad (54)$$

we can first sample \mathbf{V}^* given $(\mathbf{O}_1, \mathbf{H}_1)$, from

$$\begin{aligned} \pi(\mathbf{V}^* | \mathbf{O}_1, \theta, \beta) &\propto \frac{1}{\prod_{i=1}^k v_i^{r_1}} \text{etr} \left(-\frac{1}{2} (\mathbf{V}^*)^{-1} \mathbf{O}_1' \mathbf{H}_1 \mathbf{O}_1 \right) \\ &= \prod_{i=1}^k \frac{1}{v_i^{r_1}} \exp \left(-\frac{c_i}{v_i} \right), \end{aligned} \quad (55)$$

where c_i is the (i, i) element of $\mathbf{O}_1' \mathbf{H}_1 \mathbf{O}_1 / 2$. Thus, we can directly sample v_i independently from inverse gamma $(r_1 - 1, c_i)$ distributions, and then simply rearrange the v_i , so that $v_1 \geq \dots \geq v_k$.

Gibbs sampling of \mathbf{O}_1 : from (19), the conditional density of \mathbf{O}_1 given \mathbf{V}^* and \mathbf{H}_1 is

$$\pi(\mathbf{O}_1 | \mathbf{V}^*; \theta, \beta) \propto \text{etr} \left(-\frac{1}{2} \mathbf{H}_1 \mathbf{O}_1 (\mathbf{V}^*)^{-1} \mathbf{O}_1' \right). \quad (56)$$

Write $\mathbf{H}_1 = \mathbf{L} \mathbf{L}'$, where \mathbf{L} is the matrix of normalized eigenvectors ($\mathbf{L} \mathbf{L}' = \mathbf{I}_k$) and $\mathbf{U} = \text{diag}(u_1, \dots, u_k)$ is the diagonal matrix of corresponding eigenvalues with $u_1 \geq \dots \geq u_k$. Denote $\Gamma = \mathbf{L}' \mathbf{O}_1$. Then, the conditional density of Γ given \mathbf{V}^* and \mathbf{H}_1 is

$$\pi(\Gamma | \mathbf{V}^*; \theta, \beta) \propto \text{etr} \left(-\frac{1}{2} \mathbf{U} \Gamma (\mathbf{V}^*)^{-1} \Gamma' \right). \quad (57)$$

Hoff [31] introduced an MCMC algorithm to sample from the posterior of \mathbf{O}_1 by updating two randomly selected columns of \mathbf{O}_1 . Alternatively, Berger et al. [35] suggested updating two randomly selected rows of Γ (essentially equivalent to Hoff's method in full rank cases, but considerably faster in situations of less than full rank). For example, to update the first two rows of Γ , we write $\Gamma = \text{diag}(\mathbf{\Omega}, \mathbf{I}_{k-2}) (\Gamma'_{12}, \Gamma'_{-12})'$, where Γ'_{12} is the first 2 rows of Γ , Γ'_{-12} is the rest $k-2$ rows of Γ , and

$$\mathbf{\Omega} = \mathbf{D}_\epsilon \mathbf{\Omega}_0 = \begin{pmatrix} \epsilon_1 & 0 \\ 0 & \epsilon_2 \end{pmatrix} \begin{pmatrix} \cos \omega & -\sin \omega \\ \sin \omega & \cos \omega \end{pmatrix}. \quad (58)$$

Here, $\omega \in (-\pi/2, \pi/2]$ and $\epsilon_i = \pm 1$ for $i = 1, 2$. Write $\mathbf{U}_1 = \text{diag}(u_1, u_2)$ and $\mathbf{U}_2 = \text{diag}(u_3, \dots, u_k)$. Then, the conditional posterior of ω , $\pi(\omega | \Gamma'_{12}, \Gamma'_{-12}, (\mathbf{V}^*)^{-1}; \mathbf{H}_1)$ is

$$\begin{aligned} &\propto \text{etr} \left\{ -\frac{1}{2} \begin{pmatrix} \mathbf{U}_1 & 0 \\ 0 & \mathbf{U}_2 \end{pmatrix} \begin{pmatrix} \mathbf{\Omega} & 0 \\ 0 & \mathbf{I}_{k-2} \end{pmatrix} \begin{pmatrix} \Gamma'_{12} \\ \Gamma'_{-12} \end{pmatrix} (\mathbf{V}^*)^{-1} \right. \\ &\quad \left. \cdot (\Gamma'_{12}, \Gamma'_{-12}) \begin{pmatrix} \mathbf{\Omega}' & 0 \\ 0 & \mathbf{I}_{k-2} \end{pmatrix} \right\} \end{aligned} \quad (59)$$

$$\propto \text{etr} \left\{ -\frac{1}{2} \mathbf{U}_1 \mathbf{\Omega}_0 \Gamma'_{12} (\mathbf{V}^*)^{-1} \Gamma'_{12} \mathbf{\Omega}_0' \right\}.$$

Write

$$\Gamma'_{12} (\mathbf{V}^*)^{-1} \Gamma'_{12} = \begin{pmatrix} \cos \phi & -\sin \phi \\ \sin \phi & \cos \phi \end{pmatrix} \begin{pmatrix} s_1 & 0 \\ 0 & s_2 \end{pmatrix} \begin{pmatrix} \cos \phi & \sin \phi \\ -\sin \phi & \cos \phi \end{pmatrix}, \quad (60)$$

where $\phi \in (-\pi/2, \pi/2]$ and $s_1 > s_2$. From (59), the conditional posterior of θ is

$$\begin{aligned} \pi(\theta | \Gamma'_{12}, \mathbf{V}^*; \theta, \beta) &\propto \exp \{ c_0 \cos^2(\omega + \phi) \}, \\ \theta &\in (-\pi/2, \pi/2], \end{aligned} \quad (61)$$

where $c_0 = -(1/2)(s_1 - s_2)(u_1 - u_2) \leq 0$. Let $\alpha = \cos^2(\omega + \phi)$. Then, the full conditional density of α is

$$\begin{aligned} \pi(\alpha | \Gamma'_{12}, \mathbf{V}^*; \theta, \beta) &\propto e^{c_0 \alpha} \alpha^{-(1/2)} (1 - \alpha)^{-(1/2)}, \\ \alpha &\in [0, 1]. \end{aligned} \quad (62)$$

Sampling $\alpha \in [0, 1]$ can be done with a rejection sampler with proposal Beta(1/2, 1/2) distribution.

From the studies in Berger et al. [35], the new method substantially outperforms the Metropolis and Hit-and-Run

TABLE 1: The Bayes risk.

Prior for $(\mathbf{V}, \mathbf{W}, \text{ and } \Sigma)$	Prior for ξ	(ξ_1, \mathbf{V}_1)	(ξ_1, \mathbf{V}_2)	(ξ_2, \mathbf{V}_1)	(ξ_2, \mathbf{V}_2)
Constant prior	$\pi_1(\xi)$	234.280	301.450	244.154	312.565
	$\pi_2(\xi)$	219.234	284.716	253.791	309.367
	$\pi_3(\xi)$	214.406	267.673	223.430	275.486
Reference prior	$\pi_1(\xi)$	208.342	280.193	229.257	286.841
	$\pi_2(\xi)$	201.971	264.365	243.239	294.634
	$\pi_3(\xi)$	197.123	255.153	208.372	265.474

algorithms in moderate dimensions and succeeds for k up to 100, whereas the other methods break down in much lower dimensions.

5. Simulation

From the mean square error (MSE) perspective, we compare the performance of 6 objective hyperpriors (considered in Berger et al. [12]), created from the product of three objective hyperpriors for ξ :

- (1) Constant prior: $\pi_1(\xi) = 1$
- (2) Conjugate prior: $\pi_2(\xi) \sim N_k(0, \mathbf{I}_q)$
- (3) Recommended prior: $\pi_3(\xi) \propto [1 + \|\xi\|^2]^{-((q-1)/2)}$ and two objective hyperpriors for \mathbf{V}
 - (1) Constant prior: $\pi(\mathbf{V}) = 1, \pi(\mathbf{W}) = 1$ and $\pi(\Sigma) = 1$
 - (2) Recommended reference prior:

$$\begin{aligned} \pi(\mathbf{V}) &\propto \frac{1}{|\mathbf{V}|^{1-1/(2k)} \prod_{1 \leq i < j \leq k} (v_i - v_j)}, \quad \mathbf{V} > 0, \\ \pi(\mathbf{W}) &\propto \frac{1}{|\mathbf{W}|^{1-1/(2p)} \prod_{1 \leq i < j \leq p} (w_i - w_j)}, \quad \mathbf{W} > 0, \\ \pi(\Sigma) &\propto \frac{1}{|\Sigma|^{1-1/(2q)} \prod_{1 \leq i < j \leq q} (\sigma_i - \sigma_j)}, \quad \Sigma > 0. \end{aligned} \quad (63)$$

Except for the constant prior and recommended reference prior, Berger et al. [12] also studied the nonhierarchical independence Jeffreys prior ($\pi(\mathbf{V}) = |\mathbf{V}|^{-(k+1)/2}$) and the hierarchical independence Jeffreys prior ($\pi(\mathbf{V}) = |\mathbf{I} + \mathbf{V}|^{-(k+1)/2}$) in the model (1). From Berger et al. [12], the nonhierarchical independence Jeffreys prior for the covariance matrix cannot be used for the hierarchical models since it results in improper posterior. For the hierarchical independence Jeffreys prior, the common sampling methods are very inefficient to sample \mathbf{V} from the posterior distribution. Therefore, we do not consider either of the above-mentioned two Jeffreys priors.

Set $q = n = 2, p = s = 4, m = k = 16$. We generate $\mathbf{Z} = (\mathbf{z}_1, \dots, \mathbf{z}_{mk})'$, where $\mathbf{z}_i \stackrel{iid}{\sim} N_{sp}(0, \mathbf{I}_{sp})$ for $i = 1, \dots, mk$. Similarly, $\mathbf{T} = (\mathbf{t}_1, \dots, \mathbf{t}_{sp})'$, where $\mathbf{t}_j \stackrel{iid}{\sim} N_{nq}(0, \mathbf{I}_{nq})$ for $j = 1, \dots, sp$. We simulate the Bayes risks of the posterior means resulting from the 6 priors in every combination of the following cases:

- (i) $\xi_1 = \mathbf{1}_q$ or $\xi_2 = 30\mathbf{1}_q$;

- (ii) $(\mathbf{V}_1 = \mathbf{I}_k, \mathbf{W}_1 = \mathbf{I}_p, \Sigma_1 = \mathbf{I}_q)$ or
 $(\mathbf{V}_2 = \text{diag}\{4k - 3, \dots, 5, 1\},$
 $\mathbf{W}_2 = \text{diag}\{4p - 3, \dots, 5, 1\},$
 $\Sigma_2 = \text{diag}\{4q - 3, \dots, 5, 1\}).$

Each choice of $(\xi, \mathbf{V}, \mathbf{W}, \text{ and } \Sigma)$ specifies a ‘‘true’’ hierarchical model for the simulation and we wish to compute the Bayes risk corresponding to the posterior means for θ that arise from each of the 6 objective priors. To simulate these risks, we generate 2000 random $\theta_l \sim N_{mk}(\mathbf{1}_m \otimes \beta, \mathbf{I}_m \otimes \mathbf{V})$ and generate observations $\mathbf{y}_l | \theta_l \sim N_{mk}(\theta_l, \mathbf{I}_{mk}), l = 1, \dots, 2000$. To obtain the posterior means, $\delta^\pi(\mathbf{y}_l)$, of θ_l , under the 6 priors, we run 100,000 MCMC cycles after 3,000 burn-in cycles, using the algorithms described above. Finally, we approximate the Bayes risk as the average observed mean squared error (MSE) as follows:

$$\text{MSE} = \frac{1}{2000} \sum_{l=1}^{2000} (\theta_l - \delta^\pi(\mathbf{y}_l))' (\theta_l - \delta^\pi(\mathbf{y}_l)). \quad (64)$$

The results are given in Table 1. The recommended prior (in the last line) dominates all the others in terms of risk. Within each three-row segment, comparison of the first row with the third row is a comparison between using the constant prior for ξ and the recommended prior. The gains with the recommended prior are large. Comparison of each of the first three rows with each of the last three rows is a comparison between using the constant prior for \mathbf{V} versus the recommended prior. The gains with the recommended prior are also large. The risk results present the strong numerical evidence of the superiority of the recommended priors (5).

6. Conclusion

In this paper, we follow Berger et al.’s [12] work and study a 4-level normal hierarchical model. We demonstrate that the posterior using the recommended prior is still proper in the 4-level normal hierarchical model. Herein, we do not demonstrate Berger et al.’s [32] conjecture, i.e., the posterior using the recommended prior is always proper in any level normal hierarchical models. But our method provides one useful guideline for the completion of the story.

Besides, the normal hierarchical models are typically handled by Gibbs sampling. One difficulty of computation is to sample the covariance matrix efficiently. The common sampling methods for covariance matrices, for example, Metropolis–Hastings algorithm [36] and Hit-and-Run method [25], are inefficient for higher dimensions. To

overcome the computational bottleneck of sampling covariance matrices, we can use a powerful and efficient method proposed by Berger et al. [35] for sampling the conditional density of covariance matrices. Therefore, there is no difficulty in computation for the 4-level normal hierarchical model with the recommended priors (5). In addition, the simulation result presents the strong numerical evidence of the superiority of the recommended priors (5).

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The research was supported by the Chinese 111 Project (Grant no. B14019), the National Natural Science Foundation of China (Grant nos. 11671146 and 11901519), and the China Postdoctoral Science Foundation (Grant no. 2019M661416).

References

- [1] D. Shermom, *Systems Cost Engineering Program Affordability Management and Cost Control*, Gower Publishing, Farnham, UK, 2017.
- [2] P. Léna, D. Rouan, F. Lebrun, F. Mignard, and D. Pelat, *Observational Astrophysics*, Springer Science & Business Media, Berlin, Germany, 2012.
- [3] K. Shimotsu, "Exact local whittle estimation of fractional integration with unknown mean and time trend," *Econometric Theory*, vol. 26, no. 2, pp. 501–540, 2010.
- [4] N. Shoari, J.-S. Dubé, and S. E. Chenouri, "Estimating the mean and standard deviation of environmental data with below detection limit observations: considering highly skewed data and model misspecification," *Chemosphere*, vol. 138, pp. 599–608, 2015.
- [5] C. Chenouri and R. Hale, *Time Series Analysis in Meteorology and Climatology: an Introduction*, Vol. 7, John Wiley & Sons, Hoboken, NJ, USA, 2012.
- [6] V. L. D. Tomazella, S. R. de Jesus, F. Louzada, S. Nadarajah, and P. L. Ramos, "Reference bayesian analysis for the generalized lognormal distribution with application to survival data," *Statistics and Its Interface*, vol. 13, no. 1, pp. 139–149, 2020.
- [7] J. Schafer and K. Strimmer, "A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics," *Statistical Applications in Genetics and Molecular Biology*, vol. 4, p. 32, 2005.
- [8] J. P. Hobert and G. Casella, "The effect of improper priors on Gibbs sampling in hierarchical linear mixed models," *Journal of the American Statistical Association*, vol. 91, no. 436, pp. 1461–1473, 1996.
- [9] C. Wang, J. Rutledge, and D. Gianola, "Bayesian analysis of mixed linear models via gibbs sampling with an application to litter size of iberian pigs," *Genetics Selection Evolution*, vol. 26, pp. 1–25, 1994.
- [10] J. P. Hobert and G. Casella, "Functional compatibility, Markov chains, and Gibbs sampling with improper posteriors," *Journal of Computational and Graphical Statistics*, vol. 7, no. 1, pp. 42–60, 1998.
- [11] J. O. Berger and W. E. Strawderman, "Choice of hierarchical priors: admissibility in estimation of normal means," *The Annals of Statistics*, vol. 24, no. 3, pp. 931–951, 1996.
- [12] J. O. Berger, W. Strawderman, and D. Tang, "Posterior propriety and admissibility of hyperpriors in normal hierarchical models," *The Annals of Statistics*, vol. 33, no. 2, pp. 606–646, 2005.
- [13] P. L. Speckman, J. Lee, and D. Sun, "Existence of the MLE and propriety of posteriors for a general multinomial choice model," *Statistica Sinica*, vol. 19, pp. 731–748, 2009.
- [14] V. Roy and D. K. Dey, "Propriety of posterior distributions arising in categorical and survival models under generalized extreme value distribution," *Statistica Sinica*, vol. 24, pp. 699–722, 2014.
- [15] S. E. Michalak and C. N. Morris, "Posterior propriety for hierarchical models with log-likelihoods that have norm bounds," *Bayesian Analysis*, vol. 11, no. 2, pp. 545–571, 2016.
- [16] P. L. Ramos, F. Louzada, and E. Ramos, "Posterior proprieties of the nakagami-m distribution using noninformative priors and applications in reliability," *IEEE Transactions on Reliability*, vol. 67, no. 1, pp. 105–117, 2018.
- [17] S. Ning, *Bayesian degradation analysis considering competing risks and residual-life prediction for two-phase degradation*, PhD thesis, Ohio University, Athens, OH, USA, 2012.
- [18] P. Wang and D. W. Coit, "Reliability and degradation modeling with random or uncertain failure threshold," in *Proceedings of the 2007 Annual Reliability and Maintainability Symposium*, pp. 392–397, IEEE, Orlando, FL, USA, January 2007.
- [19] J. Berger, "The case for objective bayesian analysis," *Bayesian Analysis*, vol. 1, no. 3, pp. 385–402, 2006.
- [20] M. Pollo, V. Tomazella, G. Gilardoni, P. L. Ramos, M. J. Nicola, and F. Louzada, "Objective bayesian inference for repairable system subject to competing risks," 2018, <https://arxiv.org/abs/1804.06466>.
- [21] P. H. Ferreira, E. Ramos, P. L. Ramos et al., "Objective bayesian analysis for the lomax distribution," *Statistics & Probability Letters*, vol. 159, Article ID 108677, 2019.
- [22] D. K. Dey and C. Srinivasan, "Estimation of a Covariance Matrix under Stein's Loss," *The Annals of Statistics*, vol. 13, no. 4, pp. 1581–1591, 1985.
- [23] S. P. Lin and M. D. Perlman, "A monte carlo comparison of four estimators of a covariance matrix," in *Proceedings of the Sixth International Symposium on Multivariate Analysis VI*, Pittsburgh, PA, USA, 1985.
- [24] L. R. Haff, "The variational form of certain Bayes estimators," *The Annals of Statistics*, vol. 19, no. 3, pp. 1163–1190, 1991.
- [25] R. Yang and J. O. Berger, "Estimation of a covariance matrix using the reference prior," *The Annals of Statistics*, vol. 22, no. 3, pp. 1195–1211, 1994.
- [26] M. J. Daniels and R. E. Kass, "Nonconjugate Bayesian estimation of covariance matrices and its use in hierarchical models," *Journal of the American Statistical Association*, vol. 94, no. 448, pp. 1254–1263, 1999.
- [27] M. J. Daniels and R. E. Kass, "Shrinkage estimators for covariance matrices," *Biometrics*, vol. 57, no. 4, pp. 1173–1184, 2001.
- [28] O. Ledoit and M. Wolf, "A well-conditioned estimator for large-dimensional covariance matrices," *Journal of Multivariate Analysis*, vol. 88, no. 2, pp. 365–411, 2004.
- [29] D. Sun and J. O. Berger, "Objective priors for the multivariate normal model (with discussion)," in *Proceedings of the*

- Valencia/ISBA 8th World Meeting on Bayesian Statistics*, pp. 525–564, Alicante, Spain, June 2007.
- [30] B. Rajaratnam, H. Massam, and C. M. Carvalho, “Flexible covariance estimation in graphical Gaussian models,” *The Annals of Statistics*, vol. 36, no. 6, pp. 2818–2849, 2008.
 - [31] P. D. Hoff, “A hierarchical eigenmodel for pooled covariance estimation,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 71, no. 5, pp. 971–992, 2009.
 - [32] J. O. Berger, D. Sun, and C. Song, “An objective prior for hyperparameters in normal hierarchical models,” *Journal of Multivariate Analysis*, 2019.
 - [33] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, Wiley, Hoboken, NJ, USA, 1984.
 - [34] R. H. Farrell, *Multivariate Calculation. Use of the Continuous Groups*, Springer, New York, NY, USA, 1985.
 - [35] J. O. Berger, D. Sun, and C. Song, “Bayesian analysis of the covariance matrix of multivariate normal distribution with a new class of priors,” *The Annals of Statistics*, 2019.
 - [36] W. K. Hastings, “Monte Carlo sampling methods using Markov chains and their applications,” *Biometrika*, vol. 57, no. 1, pp. 97–109, 1970.
 - [37] M.-H. Chen and B. Schmeiser, “Performance of the gibbs, hit-and-run, and Metropolis samplers,” *Journal of Computational and Graphical Statistics*, vol. 2, no. 3, pp. 251–272, 1993.