# Objective Bayesian meta-analysis based on generalized multivariate random effects model

**Olha Bodnar and Taras Bodnar**

# Objective Bayesian meta-analysis based on generalized multivariate random effects model

Olha Bodnar[1] and Taras Bodnar[2]

[1]Unit of Statistics, School of Business, Örebro University, SE-70182 Örebro, Sweden
[2]Department of Mathematics, Stockholm University, SE-10691 Stockholm, Sweden

May 7, 2021

### Abstract

Objective Bayesian inference procedures are derived for the parameters of the multivariate random effects model generalized to elliptically contoured distributions. The posterior for the overall mean vector and the between-study covariance matrix is deduced by assigning two noninformative priors to the model parameter, namely the Berger and Bernardo reference prior and the Jeffreys prior, whose analytical expressions are obtained under weak distributional assumptions. It is shown that the only condition needed for the posterior to be proper is that the sample size is larger than the dimension of the data-generating model, independently of the class of elliptically contoured distributions used in the definition of the generalized multivariate random effects model. The theoretical findings of the paper are applied to real data consisting of ten studies about the effectiveness of hypertension treatment for reducing blood pressure where the treatment effects on both the systolic blood pressure and diastolic blood pressure are investigated.

***Keywords***: Multivariate random-effects model; Jeffreys prior; reference prior; propriety; elliptically contoured distribution; multivariate meta-analysis

**JEL Classification:** C11; C13; C15; C16

# 1 Introduction

Random effects model is a well established quantitative tool when the results of several studies are combined in a single values as it is usually done in meta-analysis and interlaboratory comparison studies which are widely spread in medicine, physics, chemistry, and in many other fields of science (see, e.g., Brockwell and Gordon (2001); Ades et al. (2005); Viechtbauer (2005, 2007); Sutton and Higgins (2008); Riley et al. (2010); Strawderman and Rukhin (2010); Cornell et al. (2014); Novianti et al. (2014); Roever (2016); Bodnar et al. (2017); Rukhin (2017a,b); Wynants et al. (2018); Michael et al. (2019); Veroniki et al. (2019)). In most of applications considered in the literature, the aim is to infer the common mean of the measurement results on a single variable, while the inference procedures for the hetorogeneity parameter have recently been derived by Rukhin (2013); Langan et al. (2017); Ma et al. (2018); Bodnar (2019) among others. Both methods of the frequentist and Bayesian statistics have been established to deal with the problem and applied in practice (see, Paule and Mandel (1982); DerSimonian and Laird (1986); Lambert et al. (2005); Guolo (2012); Turner et al. (2015); Bodnar et al. (2017)).

Although statistical theory to analyse the univariate random effects model has been developed and successfully implemented in many applications, new challenges arise when several features are measured simultaneously and have to be combined into a single (multivariate) result. One possibility is based on the application of the univariate random effects to each feature separately. However, important information about the dependence structure present in the joint distribution of the features might be lost in this case. Another approach is to generalize the existent univariate methods to the multivariate case by deriving new statistical procedures which can capture the dependencies present between several features and efficiently combine the (multivariate) results of several studies. Moreover, the assumption of normality, which is commonly imposed in meta-analysis or in interlaboratory comparison studies, is not obviously fulfilled (see, Baker and Jackson (2008); Lee and Thompson (2008); Bodnar et al. (2016); Jackson and White (2018); Wang and Lee (2020)) and more sophisticated statistical models which take the heavy-tailed behaviour into account should be considered in many applications. This makes an additional difficult in the practical implementation of the random effects model, since only a few observations are present in most cases and the advanced asymptotic methods cannot be longer used. For instance, Davey et al. (2011) pointed out that 75 % of meta-analyses reported in the Cochrane Database of Systematic Reviews (CDSR) contained five or fewer studies.

Multivariate random effects model has increased its popularity in the literature recently (see, Gasparrini et al. (2012), Wei and Higgins (2013), Jackson and Riley (2014), Liu et al. (2015), Noma et al. (2019), Negeri and Beyene (2020), Jackson et al. (2020)). Statistical inferences for the model parameters, which are the common mean vector and the heterogeneity matrix, were initially derived from the viewpoint of the frequentist statistics. Jackson et al. (2010) extended the DerSimonian and Laird approach to the multivariate data, while Chen et al. (2012) presented the method based on the restricted maximum likelihood approach. These two procedures from frequentist statistics constitute the commonly used methods in multivariate meta-analysis (see, e.g., Jackson et al. (2013), Schwarzer et al. (2015), Jackson et al. (2020)). Paul et al. (2010) derived Bayesian inferences procedures for the parameters of the two-dimensional random effects model based on the Laplace approximation, while Nam et al. (2003) provided results in a multivariate case. Both the papers discussed Bayesian inference obtained when informative priors are employed.

Following Bernstein-von Mises theorem (see, Bernardo and Smith (2000)), a prior has a minor

impact on the posterior when the sample size is large. When a sample of a small size is available, which is a common situation in practice (cf., Davey et al. (2011)), the application of an incorrectly chosen informative prior can be very influential on the resulting Bayesian inference procedures for the model parameters. This challenge becomes even more pronounced in case of Bayesian inference for parameters of a multivariate model.

The contribution of the paper to the existent literature on multivariate random effects model and multivariate meta-analysis is done in several directions. First, we develop objective Bayesian inference procedures for the parameters of the multivariate random effects model. In particular, we derive the analytical expression of the Fisher information matrix and the two noninformative priors: Berger and Bernardo reference prior and Jeffreys prior. Employing these two priors, the expressions of the corresponding posterior distributions are obtained and the conditions for their propriety are established. Second, we weaken the assumption of multivariate normal distribution and replace it by a general class of multivariate distributions, the so-called elliptically contoured distributions (see, Gupta et al. (2013)).

The rest of the paper is structured as follows. In Section 2, the generalized multivariate random effects model is introduced and two noninformative priors, Berger and Bernardo reference prior and Jeffreys prior, are derived. The posterior distribution for model parameters are obtained in Section 3, while the conditions for posterior propriety are stated in Section 3.1. In Section 4 numerical procedures are developed to draw samples from the derived posterior distributions. Results for two special families of elliptically contoured distributions are provided in Section 5, while an empirical illustration is presented in Section 6. Final remarks are given in Section 7. The proofs of technical results are moved to the appendix (Section 8).

# 2   Model and noninformative priors

We consider an extension of the (normal) multivariate random effects model with density function given by

$$p(\mathbf{X}|\boldsymbol{\mu}, \boldsymbol{\Psi}) = \frac{1}{\sqrt{\det(\boldsymbol{\Psi} \otimes \mathbf{I} + \mathbf{U})}} f\left(\text{vec}(\mathbf{X} - \boldsymbol{\mu}\mathbf{1}^{\top})^{\top}(\boldsymbol{\Psi} \otimes \mathbf{I} + \mathbf{U})^{-1}\text{vec}(\mathbf{X} - \boldsymbol{\mu}\mathbf{1})\right), \quad (1)$$

where $\mathbf{X}$ is a $(p \times n)$ matrix, $\boldsymbol{\mu}$ is a $p$-dimensional vector, $\boldsymbol{\Psi}$ is a $(p \times p)$ matrix, $\mathbf{1}$ is a vector of ones, $\mathbf{I}$ is the identity matrix of an appropriate order, and $\mathbf{U}$ is a $(pn \times pn)$ deterministic matrix. The symbol $\otimes$ denotes the Kronecker product, while vec stands for the *vec* operator. The model (1) extends the univariate approach suggested in Bodnar et al. (2016) to the multivariate case and can also be used when several correlated features obtained from different studies should be combined together.

In a special case of $f(z) = \exp(-z/2)/(2\pi)^{pn/2}$ and $\mathbf{U} = \text{diag}(\mathbf{U}_1, ..., \mathbf{U}_n)$ with $\mathbf{U}_i$: $p \times p$ for $i = 1, ..., n$, the model (1) can be written as

$$\mathbf{x}_i = \boldsymbol{\mu} + \boldsymbol{\lambda}_i + \boldsymbol{\varepsilon}_i \quad \text{with} \quad \boldsymbol{\lambda} \sim \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Psi}) \quad \text{and} \quad \boldsymbol{\varepsilon} \sim \mathcal{N}_p(\mathbf{0}, \mathbf{U}_i), \quad (2)$$

where $\{\boldsymbol{\lambda}_i\}_{i=1,...,n}$ and $\{\boldsymbol{\varepsilon}_i\}_{i=1,...,n}$ are mutually independent. The presentation (2) defines the normal multivariate random effects model. Motivated by the normal multivariate random effects model, it is assumed that $\mathbf{U} = \text{diag}(\mathbf{U}_1, ..., \mathbf{U}_n)$ holds in (1).

In many applications in medicine, physics, and chemistry the aim is to infer $\boldsymbol{\mu}$ given observation matrix $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$. In the applications of these fields the information about the scale matrix $\mathbf{U}$

is usually provided by the participating organizations (see, Lambert et al. (2005), Turner et al. (2015), Bodnar and Elster (2014b), Jackson et al. (2020)). As a result, it is assumed to be a known symmetric positive definite matrix. On the other side, the matrix $\mathbf{\Psi}$ is treated as an unknown quantity with the aim to capture the additional variability in data when several observations taken at different places and times are pooled together. The matrix $\mathbf{\Psi}$ is usually treated as an additional nuisance parameter of the model.

By (1), the conditional distribution of $\mathbf{X}$ given $\boldsymbol{\mu}$ and $\mathbf{\Psi}$ belongs the class of the matrix-variate elliptical contoured distributions (see, e.g., Gupta et al. (2013) for the definition and properties of this matrix-variate family of distributions). This assertion will be denoted by $\mathbf{X}|\boldsymbol{\mu}, \mathbf{\Psi} \sim E_{p,n}(\boldsymbol{\mu}\mathbf{1}^\top, \mathbf{\Psi} \otimes \mathbf{I} + \mathbf{U}, f)$ ($p \times n$-dimensional matrix-variate elliptically contoured distribution with location matrix $\boldsymbol{\mu}\mathbf{1}^\top$, dispersion matrix $(\mathbf{\Psi} \otimes \mathbf{I} + \mathbf{U})$, and density generator $f(.)$. Following the definition of matrix-variate elliptically contoured distributions (see, Gupta et al. (2013, Theorem 2.7)), the function $f(.)$ should be a non-negative Lebesgue measurable function on $[0, \infty)$ such that

$$\int\limits_0^\infty t^{pn-1} f(t^2) dt < \infty.$$

## 2.1 Noninformative priors: Berger and Bernardo reference prior and Jeffreys prior

In many practical applications no information or only vague information is available about the model parameters. In such cases, especially when additionally the sample size is small or the model dimension is large in comparison to the sample size, the usage of an informative prior can be questionable. As a possible solution to this problem, noninformative priors were developed and employed in the derivation of Bayesian inference. Historically, the first noninformative prior was suggested by Laplace (1812) who proposed to assign a constant prior to the parameters of the model. This prior is also known in the literature as the constant prior or the uniform prior. Although the uniform prior works well when Bayesian inference is determined for location parameters of a statistical model, its application does not obviously lead to good results for other types of model parameters. One of the most crucial critiques of the uniform prior is that it is invariant under transformations of parameters.

As a solution, Jeffreys (1946) proposed to compute a non-informative prior as the square root of the determinant of the Fisher information matrix. Although this approach leads to a prior which is invariant under transformations of model parameters, some difficulties arise in the case of multi-parameter statistical models (see, Held and Bové (2014)). The approach of Jeffreys was further extended in Berger and Bernardo (1992) who suggested the so-called reference prior (see, also Berger et al. (2009) for the properties of the reference prior). The idea used in the derivation of the reference prior is based on the sequential maximization of the Shannon mutual information (see, Bodnar and Elster (2014a)) which determines the distance between the prior and posterior.

In Theorem 1 the analytical expression of the Fisher information matrix is provided, which is then used in the derivation of both the Berger and Bernardo reference prior and the Jeffreys prior for the parameters of the generalized multivariate random effects model (1).

**Theorem 1.** *The Fisher information matrix for model (1) with* $\mathbf{U} = diag(\mathbf{U}_1, ..., \mathbf{U}_n)$ *is given by*

$$\mathbf{F} = \begin{pmatrix} \mathbf{F}_{11} & \mathbf{O} \\ \mathbf{O} & \mathbf{F}_{22} \end{pmatrix} \tag{3}$$

*where*

$$\mathbf{F}_{11} = \frac{4J_1}{pn} \sum_{i=1}^{n} (\boldsymbol{\Psi} + \mathbf{U}_i)^{-1}, \tag{4}$$

$$\mathbf{F}_{22} = \mathbf{G}_p^{\top} \left[ \left( \frac{J_2}{2pn + p^2 n^2} - \frac{1}{4} \right) vec \left( \sum_{i=1}^{n} (\boldsymbol{\Psi} + \mathbf{U}_i)^{-1} \right) vec \left( \sum_{j=1}^{n} (\boldsymbol{\Psi} + \mathbf{U}_j)^{-1} \right)^{\top} \right.$$

$$\left. + \frac{2J_2}{2pn + p^2 n^2} \sum_{i=1}^{n} \left( (\boldsymbol{\Psi} + \mathbf{U}_i)^{-1} \otimes (\boldsymbol{\Psi} + \mathbf{U}_i)^{-1} \right) \right] \mathbf{G}_p \tag{5}$$

*with*

$$J_i = \mathbb{E} \left( (R^2)^i \left( \frac{f'(R^2)}{f(R^2)} \right)^2 \right), \tag{6}$$

*where $R^2 = vec(\mathbf{Z})^{\top} vec(\mathbf{Z})$ with $\mathbf{Z} \sim E_{p,n}(\mathbf{O}_{p,n}, \mathbf{I}_{p \times n}, f)$ standard matrix-variate elliptically contoured distribution with density generator $f(.)$ and $\mathbf{G}_p$ stands for the duplication matrix.*

The results of Theorem 1 show that the Fisher information matrix depends on the type of elliptical distribution only over the two univariate constants $J_1$ and $J_2$ which are fully determined by density generator $f(.)$. Moreover, the Fisher information matrix $\mathbf{F}$ is finite if $J_1 < \infty$ and $J_2 < \infty$. Thus, it is assumed throughout the paper that the density generator $f(.)$ is chosen such that these two conditions are fulfilled. Although the expectations in the definition of $J_1$ and $J_2$ cannot always be analytically computed, they can easily be approximated via simulations by drawing samples from the corresponding standard elliptically contoured distribution. Finally, $J_1$ is present in $\mathbf{F}_{11}$ as a multiplicative constant and, thus, both the Berger and Bernardo reference prior and the Jeffreys prior depend on $J_2$ only as shown below.

Since $\mathbf{F}$ is block-diagonal and it does not depend on $\boldsymbol{\mu}$, the Jeffreys prior for $\boldsymbol{\mu}$ and $\boldsymbol{\Psi}$ depends on $\boldsymbol{\Psi}$ only and it is given by

$$\pi_J(\boldsymbol{\mu}, \boldsymbol{\Psi}) = \pi_J(\boldsymbol{\Psi}) \propto \sqrt{\det(\mathbf{F})} = \sqrt{\det(\mathbf{F}_{11})} \sqrt{\det(\mathbf{F}_{22})}, \tag{7}$$

where $\mathbf{F}_{11}$ and $\mathbf{F}_{22}$ are given in (4) and (5), respectively.

Moreover, using the block-diagonal structure of $\mathbf{F}$ and the fact that $\mathbf{F}$ does not depend on $\boldsymbol{\mu}$, we immediately obtain the Berger and Bernardo reference prior $\pi_R(\boldsymbol{\mu}, \boldsymbol{\Psi})$ for the generalized multivariate random effects model (1) from the corollary to Proposition 5.29 in Bernardo and Smith (2000). This result is summarized in Theorem 2.

**Theorem 2.** *For model (1) with $\mathbf{U} = diag(\mathbf{U}_1, ..., \mathbf{U}_n)$ and grouping $\{\boldsymbol{\mu}, \boldsymbol{\Psi}\}$ (i.e. with $\boldsymbol{\Psi}$ as the nuisance parameter), the Berger and Bernardo reference prior is given by*

$$\pi_R(\boldsymbol{\mu}, \boldsymbol{\Psi}) = \pi_R(\boldsymbol{\Psi}) \propto \sqrt{det(\mathbf{F}_{22})}, \tag{8}$$

*where $\mathbf{F}_{22}$ is given in (5).*

Under additional restrictions imposed on matrix $\mathbf{U}$ and density generator $f(.)$, several simplifications of the expressions of both the Jeffreys prior and the reference prior are obtained and are presented in Corollary 1 and Corollary 2. For example, when the normal multivariate random effects model (2) is assumed, then we get

**Corollary 1.** *For model (2) and grouping $\{\boldsymbol{\mu}, \boldsymbol{\Psi}\}$ (i.e. with $\boldsymbol{\Psi}$ as the nuisance parameter), the following results hold:*

*(i) the Berger and Bernardo reference prior is given by*

$$\pi_R(\boldsymbol{\mu}, \boldsymbol{\Psi}) = \pi_R(\boldsymbol{\Psi}) \propto \sqrt{det\left(\mathbf{G}_p^\top \left[\sum_{i=1}^n ((\boldsymbol{\Psi} + \mathbf{U}_i)^{-1} \otimes (\boldsymbol{\Psi} + \mathbf{U}_i)^{-1})\right]\mathbf{G}_p\right)}, \tag{9}$$

*(ii) the Jeffreys prior is given by*

$$\pi_J(\boldsymbol{\mu}, \boldsymbol{\Psi}) = \pi_J(\boldsymbol{\Psi}) \propto \pi_R(\boldsymbol{\Psi})\sqrt{det\left(\sum_{i=1}^n (\boldsymbol{\Psi} + \mathbf{U}_i)^{-1}\right)}. \tag{10}$$

*Proof of Corollary 1:* (i) Using that $f(u) = \exp(-u/2)/(2\pi)^{pn/2}$, we get that $f'(u) = -\frac{1}{2}f(u)$ and, consequently,

$$J_2 = \frac{1}{4}\mathbb{E}\left((\text{vec}(\mathbf{Z})^\top \text{vec}(\mathbf{Z}))^2\right) = \frac{2pn + p^2n^2}{4}$$

Hence, under model (2) we obtain

$$\mathbf{F}_{22} = \mathbf{G}_p^\top \left[\frac{2J_2}{2pn + p^2n^2}\sum_{i=1}^n \left((\boldsymbol{\Psi} + \mathbf{U}_i)^{-1} \otimes (\boldsymbol{\Psi} + \mathbf{U}_i)^{-1}\right)\right]\mathbf{G}_p,$$

which leads to the expression presented in the statement of the corollary.

(ii) The result follows from the part (i) and the block-diagonality of $\mathbf{F}$.

$\square$

If the generalized multivariate random effects model is assumed to be homoscedastic, that is the equality $\mathbf{U}_1 = ... = \mathbf{U}_n = \mathbf{V}$ holds, then the Berger and Bernardo reference prior and the Jeffreys prior are given by

**Corollary 2.** *Under the assumption of Theorem 2, assume that $\mathbf{U}_1 = ... = \mathbf{U}_n = \mathbf{V}$. Then*

*(i) the Berger & Bernardo reference prior is given by*

$$\pi_R(\boldsymbol{\mu}, \boldsymbol{\Psi}) = \pi_R(\boldsymbol{\Psi}) \propto det(\boldsymbol{\Psi} + \mathbf{V})^{-(p+1)/2}, \tag{11}$$

*(ii) the Jeffreys prior is given by*

$$\pi_J(\boldsymbol{\mu}, \boldsymbol{\Psi}) = \pi_J(\boldsymbol{\Psi}) \propto det(\boldsymbol{\Psi} + \mathbf{V})^{-(p+2)/2}. \tag{12}$$

*Proof of Corollary 2:* (i) Under the condition $\mathbf{U}_1 = ... = \mathbf{U}_n = \mathbf{V}$, we get

$$\begin{aligned}
\mathbf{F}_{22} = \mathbf{G}_p^\top \Bigg[ &\left(\frac{J_2}{2pn + p^2n^2} - \frac{1}{4}\right) \text{vec}\left(n(\boldsymbol{\Psi} + \mathbf{V})^{-1}\right)\text{vec}\left(n(\boldsymbol{\Psi} + \mathbf{V})^{-1}\right)^\top \\
&+ \frac{2nJ_2}{2pn + p^2n^2}\left((\boldsymbol{\Psi} + \mathbf{V})^{-1} \otimes (\boldsymbol{\Psi} + \mathbf{V})^{-1}\right)\Bigg]\mathbf{G}_p.
\end{aligned}$$

The application of the properties of determinants involving the duplication matrix $\mathbf{G}_p$ (see, Lütkepohl (1996, Section 4.2.3)), we obtain

$$
\det\left\{(\mathbf{G}_p^\top \mathbf{G}_p)^{-1}\mathbf{G}_p^\top\left[\left(\frac{J_2}{2pn+p^2n^2}-\frac{1}{4}\right)\text{vec}\left(n(\mathbf{\Psi}+\mathbf{V})^{-1}\right)\text{vec}\left(n(\mathbf{\Psi}+\mathbf{V})^{-1}\right)^\top\right.\right.
$$
$$
\left.\left.+\ \frac{2nJ_2}{2pn+p^2n^2}\left((\mathbf{\Psi}+\mathbf{V})^{-1}\otimes(\mathbf{\Psi}+\mathbf{V})^{-1}\right)\right]\mathbf{G}_p\right\}
$$
$$
=\ \left(\frac{2nJ_2}{2pn+p^2n^2}\right)^{\frac{p(p+1)}{2}}\left(1+\left(\frac{1}{2}-\frac{2pn+p^2n^2}{8J_2}\right)pn\right)\det(\mathbf{\Psi}+\mathbf{V})^{-(p+1)},
$$

from which the expression of the Berger and Bernardo reference prior follows.

(ii) The result for the Jeffreys prior follows from part (i) and the equality

$$
\det(\mathbf{F}_{11})=\left(\frac{4J_1}{p}\right)^p\det(\mathbf{\Psi}+\mathbf{V})^{-1}.
$$

$\square$

It is remarkable that both the Berger and Bernardo reference prior and the Jeffreys prior under the assumption of homoscedasticity do not depend on the type of elliptically contoured distribution. In particular, the formulas from Corollary 2 can be used for the normal multivariate random effects model (2).

# 3 Posterior

In the derivation of the posterior we consider a prior for $\boldsymbol{\mu}$ and $\mathbf{\Psi}$ which is a function of $\mathbf{\Psi}$ only, that is $\pi(\mathbf{\Psi})$. Such a prior is an extension of both the Berger and Bernardo reference prior and the Jeffreys prior and, consequently, the derived posterior can be used to deduce the posteriors obtained when the Berger and Bernardo reference prior and the Jeffreys prior are employed as important special cases.

Under such a general prior the joint posterior for $\boldsymbol{\mu}$ and $\mathbf{\Psi}$ is obtained from (1) and it is given by

$$
\begin{aligned}
\pi(\boldsymbol{\mu},\mathbf{\Psi}|\mathbf{X}) \ &\propto\ \pi(\mathbf{\Psi})p(\mathbf{X}|\boldsymbol{\mu},\mathbf{\Psi})\\
&=\ \frac{\pi(\mathbf{\Psi})}{\sqrt{\det(\mathbf{\Psi}\otimes\mathbf{I}+\mathbf{U})}}f\left(\text{vec}(\mathbf{X}-\boldsymbol{\mu}\mathbf{1}^\top)^\top(\mathbf{\Psi}\otimes\mathbf{I}+\mathbf{U})^{-1}\text{vec}(\mathbf{X}-\boldsymbol{\mu}\mathbf{1})\right),
\end{aligned}\tag{13}
$$

with $\mathbf{U}=\text{diag}(\mathbf{U}_1,...,\mathbf{U}_n)$. In Theorem 3 it shown that the conditional reference posterior for $\boldsymbol{\mu}$ belongs to the family of elliptically contoured distributions.

**Theorem 3.** *Under the generalized multivariate random effects model (1) with $\mathbf{U}=diag(\mathbf{U}_1,...,\mathbf{U}_n)$, the conditional posterior $\pi(\boldsymbol{\mu}|\mathbf{\Psi},\mathbf{x})$ is given by*

$$
\pi(\boldsymbol{\mu}|\mathbf{\Psi},\mathbf{X})\ \propto\ f_{\mathbf{\Psi},\mathbf{X}}\left((\boldsymbol{\mu}-\tilde{\mathbf{x}}(\mathbf{\Psi}))^\top\left(\sum_{i=1}^n(\mathbf{\Psi}+\mathbf{U}_i)^{-1}\right)(\boldsymbol{\mu}-\tilde{\mathbf{x}}(\mathbf{\Psi}))\right),\tag{14}
$$

*where*

$$
f_{\mathbf{\Psi},\mathbf{X}}(u)=f\left(\sum_{i=1}^n(\mathbf{x}_i-\tilde{\mathbf{x}}(\mathbf{\Psi}))^\top(\mathbf{\Psi}+\mathbf{U}_i)^{-1}(\mathbf{x}_i-\tilde{\mathbf{x}}(\mathbf{\Psi}))+u\right)\qquad u\geq 0,\tag{15}
$$

*with*

$$\tilde{\mathbf{x}}(\boldsymbol{\Psi}) = \left(\sum_{i=1}^{n}(\boldsymbol{\Psi}+\mathbf{U}_i)^{-1}\right)^{-1}\sum_{i=1}^{n}(\boldsymbol{\Psi}+\mathbf{U}_i)^{-1}\mathbf{x}_i. \tag{16}$$

*Proof of Theorem 3:* It holds that

$$\text{vec}(\mathbf{X}-\boldsymbol{\mu}\mathbf{1}^\top)^\top(\boldsymbol{\Psi}\otimes\mathbf{I}+\mathbf{U})^{-1}\text{vec}(\mathbf{X}-\boldsymbol{\mu}\mathbf{1}) = \sum_{i=1}^{n}(\mathbf{x}_i-\boldsymbol{\mu})^\top(\boldsymbol{\Psi}+\mathbf{U}_i)^{-1}(\mathbf{x}_i-\boldsymbol{\mu})$$

$$= (\boldsymbol{\mu}-\tilde{\mathbf{x}}(\boldsymbol{\Psi}))^\top\left(\sum_{i=1}^{n}(\boldsymbol{\Psi}+\mathbf{U}_i)^{-1}\right)(\boldsymbol{\mu}-\tilde{\mathbf{x}}(\boldsymbol{\Psi}))+\sum_{i=1}^{n}(\mathbf{x}_i-\tilde{\mathbf{x}}(\boldsymbol{\Psi}))^\top(\boldsymbol{\Psi}+\mathbf{U}_i)^{-1}(\mathbf{x}_i-\tilde{\mathbf{x}}(\boldsymbol{\Psi}))$$

with $\tilde{\mathbf{x}}(\boldsymbol{\Psi})$ as in (16).

Hence,

$$\pi(\boldsymbol{\mu}|\boldsymbol{\Psi},\mathbf{X}) \propto \frac{\pi(\boldsymbol{\Psi})}{\sqrt{\det(\boldsymbol{\Psi}\otimes\mathbf{I}+\mathbf{U})}}$$

$$\times \ f\left((\boldsymbol{\mu}-\tilde{\mathbf{x}}(\boldsymbol{\Psi}))^\top\left(\sum_{i=1}^{n}(\boldsymbol{\Psi}+\mathbf{U}_i)^{-1}\right)(\boldsymbol{\mu}-\tilde{\mathbf{x}}(\boldsymbol{\Psi}))+\sum_{i=1}^{n}(\mathbf{x}_i-\tilde{\mathbf{x}}(\boldsymbol{\Psi}))^\top(\boldsymbol{\Psi}+\mathbf{U}_i)^{-1}(\mathbf{x}_i-\tilde{\mathbf{x}}(\boldsymbol{\Psi}))\right)$$

$$\propto \ f_{\boldsymbol{\Psi},\mathbf{X}}\left((\boldsymbol{\mu}-\tilde{\mathbf{x}}(\boldsymbol{\Psi}))^\top\left(\sum_{i=1}^{n}(\boldsymbol{\Psi}+\mathbf{U}_i)^{-1}\right)(\boldsymbol{\mu}-\tilde{\mathbf{x}}(\boldsymbol{\Psi}))\right),$$

where $f_{\boldsymbol{\Psi},\mathbf{X}}(.)$ is given in (15). $\qquad\square$

As a straightforward consequence of the result in Theorem 3 by substituting $\pi(\boldsymbol{\Psi})$ with $\pi_R(\boldsymbol{\Psi})$ and $\pi_J(\boldsymbol{\Psi})$ from (8) and (7) respectively, we get the conditional reference posterior $\pi(\boldsymbol{\mu}|\boldsymbol{\Psi},\mathbf{x})$ for the generalized multivariate random effects model (1) and the conditional posterior when the Jeffreys prior is used.

Moreover, from the proof of Theorem 3 we also get the marginal posterior for $\boldsymbol{\Psi}$ as given by

**Corollary 3.** *Under the generalized multivariate random effects model (1) with $\mathbf{U}=diag(\mathbf{U}_1,...,\mathbf{U}_n)$, the conditional posterior $\pi(\boldsymbol{\Psi}|\mathbf{X})$ is given by*

$$\pi(\boldsymbol{\Psi}|\mathbf{X}) \ \propto \ \frac{\pi(\boldsymbol{\Psi})}{\sqrt{det(\sum_{i=1}^{n}(\boldsymbol{\Psi}+\mathbf{U}_i)^{-1})}\prod_{i=1}^{n}\sqrt{det(\boldsymbol{\Psi}+\mathbf{U}_i)}}$$

$$\times \ \int_0^\infty u^{p-1}f\left(u^2+\sum_{i=1}^{n}(\mathbf{x}_i-\tilde{\mathbf{x}}(\boldsymbol{\Psi}))^\top(\boldsymbol{\Psi}+\mathbf{U}_i)^{-1}(\mathbf{x}_i-\tilde{\mathbf{x}}(\boldsymbol{\Psi}))\right)du, \tag{17}$$

*Proof of Corollary 3:* Using the transformation $\boldsymbol{\nu}=\left(\sum_{i=1}^{n}\boldsymbol{\Psi}+\mathbf{U}_i\right)^{-1}(\boldsymbol{\mu}-\tilde{\mathbf{x}}(\boldsymbol{\Psi}))$ with the Jacobian $1/\sqrt{\det(\sum_{i=1}^{n}(\boldsymbol{\Psi}+\mathbf{U}_i)^{-1})}$, the marginal posterior for $\boldsymbol{\Psi}$ is expressed as

$$\pi(\boldsymbol{\Psi}|\mathbf{X}) \ \propto \ \int_{\mathbb{R}^p}\frac{\pi(\boldsymbol{\Psi})}{\prod_{i=1}^{n}\sqrt{\det(\boldsymbol{\Psi}+\mathbf{U}_i)}}f\left(\text{vec}(\mathbf{X}-\boldsymbol{\mu}\mathbf{1}^\top)^\top(\boldsymbol{\Psi}\otimes\mathbf{I}+\mathbf{U})^{-1}\text{vec}(\mathbf{X}-\boldsymbol{\mu}\mathbf{1})\right)d\boldsymbol{\mu}$$

$$= \ \frac{\pi(\boldsymbol{\Psi})}{\sqrt{\det(\sum_{i=1}^{n}(\boldsymbol{\Psi}+\mathbf{U}_i)^{-1})}\prod_{i=1}^{n}\sqrt{\det(\boldsymbol{\Psi}+\mathbf{U}_i)}}$$

$$\times \ \int_{\mathbb{R}^p}f\left(\boldsymbol{\nu}^\top\boldsymbol{\nu}+\sum_{i=1}^{n}(\mathbf{x}_i-\tilde{\mathbf{x}}(\boldsymbol{\Psi}))^\top(\boldsymbol{\Psi}+\mathbf{U}_i)^{-1}(\mathbf{x}_i-\tilde{\mathbf{x}}(\boldsymbol{\Psi}))\right)d\boldsymbol{\nu}.$$

The rest of the proof follows by noting that (cf., Gupta et al. (2013, Lemma 2.1)):

$$\int_{\mathbb{R}^p} g(\mathbf{t}^\top \mathbf{t}) d\mathbf{t} = \frac{2\pi^{p/2}}{\Gamma(p/2)} \int_0^\infty u^{p-1} g(u^2) du,$$

where $\Gamma(.)$ stands for the gamma function. $\hfill\square$

The posterior mean vector and the posterior covariance matrix of $\boldsymbol{\mu}$ are derived from Theorem 3 by using the rule of iterated expectations. They are given by

$$
\begin{aligned}
\mathbb{E}\left(\boldsymbol{\mu}|\mathbf{X}\right) &= \mathbb{E}(\mathbb{E}(\boldsymbol{\mu}|\boldsymbol{\Psi},\mathbf{X})|\mathbf{X}) = \mathbb{E}(\tilde{\mathbf{x}}(\boldsymbol{\Psi})|\mathbf{X}) \\
&= \mathbb{E}\left(\left(\sum_{i=1}^n (\boldsymbol{\Psi}+\mathbf{U}_i)^{-1}\right)^{-1} \sum_{i=1}^n (\boldsymbol{\Psi}+\mathbf{U}_i)^{-1}\mathbf{x}_i \Bigg| \mathbf{X}\right)
\end{aligned}
\tag{18}
$$

and

$$
\begin{aligned}
\mathbb{V}ar\left(\boldsymbol{\mu}|\mathbf{X}\right) &= \mathbb{E}(\mathbb{V}ar(\boldsymbol{\mu}|\boldsymbol{\Psi},\mathbf{X})|\mathbf{X}) + \mathbb{V}ar(\mathbb{E}(\boldsymbol{\mu}|\boldsymbol{\Psi},\mathbf{X})|\mathbf{X}) \\
&= \mathbb{E}\left(C(\boldsymbol{\Psi})\left(\sum_{i=1}^n (\boldsymbol{\Psi}+\mathbf{U}_i)^{-1}\right)^{-1} \Bigg| \mathbf{X}\right) \\
&\quad + \mathbb{V}ar\left(\left(\sum_{i=1}^n (\boldsymbol{\Psi}+\mathbf{U}_i)^{-1}\right)^{-1} \sum_{i=1}^n (\boldsymbol{\Psi}+\mathbf{U}_i)^{-1}\mathbf{x}_i \Bigg| \mathbf{X}\right),
\end{aligned}
\tag{19}
$$

where

$$C(\boldsymbol{\Psi}) = \mathbb{E}\left(R_{\boldsymbol{\Psi},\mathbf{X}}^2\right) \quad \text{with} \quad R_{\boldsymbol{\Psi},\mathbf{X}}^2 = \mathbf{z}_{\boldsymbol{\Psi},\mathbf{X}}^\top \mathbf{z}_{\boldsymbol{\Psi},\mathbf{X}}, \tag{20}$$

where $\mathbf{z}_{\boldsymbol{\Psi},\mathbf{X}} \sim E_p(\mathbf{0}_p, \mathbf{I}_p, f_{\boldsymbol{\Psi},\mathbf{X}})$.

## 3.1 Propriety

For the derivation of the conditions required for the propriety of the posterior $\pi(\boldsymbol{\mu}, \boldsymbol{\Psi}|\mathbf{X})$, we use the following lemma:

**Lemma 1.** *Let $\mathbf{A} > 0$ be a symmetric and positive definite matrix, and let $\mathbf{B} \geq 0$ be a symmetric and positive semidefinite matrix. Then*

$$\mathbf{A}^{-1} \otimes \mathbf{A}^{-1} - (\mathbf{A}+\mathbf{B})^{-1} \otimes (\mathbf{A}+\mathbf{B})^{-1} \geq 0$$

*is positive semidefinite.*

The proof of Lemma 1 is given in the appendix. In Theorem 4 we formulate the conditions required for the propriety of the posterior.

**Theorem 4.** *Consider the generalized multivariate random effects model (1) with $\mathbf{U} = diag(\mathbf{U}_1, ..., \mathbf{U}_n)$. Let $f(u)$ be a non-increasing function in $u \geq 0$ and $\frac{J_2}{2pn+p^2n^2} - \frac{1}{4} \leq 0$ where $J_2$ is defined in (6).*

1. *If $n \geq p$, then the posterior $\pi(\boldsymbol{\mu}, \boldsymbol{\Psi}|\mathbf{X})$ derived under the Jeffreys prior $\pi_J(\boldsymbol{\Psi})$ is proper.*

2. *If $n \geq p + 1$, then the posterior $\pi(\boldsymbol{\mu}, \boldsymbol{\Psi}|\mathbf{X})$ derived under the Berger and Bernardo reference prior $\pi_R(\boldsymbol{\Psi})$ is proper.*

*Proof of Theorem 4:* First, we derive an upper bound for the determinant of two diagonal blocks of the Fisher information matrix derived in Theorem 1. Let $\tilde{\phi} = \min_{i=1,\dots,n} \min_{j=1,\dots,p} \phi_{i,j}$ denote the minimum of the eigenvalues $\{\phi_{i,j}\}_{j=1,\dots,p}$ computed for the matrices $\mathbf{U}_i$, $i = 1,\dots,n$. Since $\boldsymbol{\Psi} + \tilde{\phi}\mathbf{I}$ and $\mathbf{U}_i - \tilde{\phi}\mathbf{I}$ are symmetric and positive semi-definitewe obtain from Lemma 1.(ii) that

$$(\boldsymbol{\Psi} + \tilde{\phi}\mathbf{I})^{-1} \otimes (\boldsymbol{\Psi} + \tilde{\phi}\mathbf{I})^{-1} - (\boldsymbol{\Psi} + \mathbf{U}_i)^{-1} \otimes (\boldsymbol{\Psi} + \mathbf{U}_i)^{-1} \geq \mathbf{0},$$

for $i = 1,\dots,n$ and, hence,

$$n(\boldsymbol{\Psi} + \tilde{\phi}\mathbf{I})^{-1} \otimes (\boldsymbol{\Psi} + \tilde{\phi}\mathbf{I})^{-1} - \sum_{i=1}^{n}(\boldsymbol{\Psi} + \mathbf{U}_i)^{-1} \otimes (\boldsymbol{\Psi} + \mathbf{U}_i)^{-1} \geq \mathbf{0}.$$

The inequality $J_2/(2pn + p^2n^2) - 1/4 \leq 0$ yields

$$\det(\mathbf{F}_{22}) \quad \leq \quad \det\left(\frac{2J_2}{2pn + p^2n^2}\mathbf{G}_p^\top \sum_{i=1}^{n}\left((\boldsymbol{\Psi} + \mathbf{U}_i)^{-1} \otimes (\boldsymbol{\Psi} + \mathbf{U}_i)^{-1}\right)\mathbf{G}_p\right),$$

while the application of Theorem 18.1.6 of Harville (1997) implies

$$\det(\mathbf{F}_{22}) \quad \leq \quad \det\left(\frac{2nJ_2}{2pn + p^2n^2}\mathbf{G}_p^\top\left((\boldsymbol{\Psi} + \tilde{\phi}\mathbf{I})^{-1} \otimes (\boldsymbol{\Psi} + \tilde{\phi}\mathbf{I})^{-1}\right)\mathbf{G}_p\right)$$

$$= \quad \left(\frac{2J_2}{2p + p^2n}\right)^{p(p+1)/2}\det(\boldsymbol{\Psi} + \tilde{\phi}\mathbf{I})^{-(p+1)}. \tag{21}$$

Let

$$c_{11}(p,n) = \left(\frac{4J_1}{p}\right)^{p/2} \quad \text{and} \quad c_{22}(p,n) = \left(\frac{2J_2}{2p + p^2n}\right)^{p(p+1)/2}.$$

Using that

$$(\det(\boldsymbol{\Psi} \otimes \mathbf{I} + \mathbf{U}))^{-1} = \prod_{i=1}^{n}\det\left((\boldsymbol{\Psi} + \mathbf{U}_i)^{-1}\right) \leq \det\left((\boldsymbol{\Psi} + \tilde{\phi}\mathbf{I})^{-1}\right)^n = \det(\boldsymbol{\Psi} + \tilde{\phi}\mathbf{I})^{-n} \tag{22}$$

and that $f(.)$ is non-increasing, we get that the kernel of the posterior $\pi(\boldsymbol{\mu}, \boldsymbol{\Psi}|\mathbf{X})$ derived under the Jeffreys prior is bounded by

$$g_J(\boldsymbol{\mu}, \boldsymbol{\Psi}|\mathbf{X}) = \frac{\pi_J(\boldsymbol{\Psi})}{\sqrt{\det(\boldsymbol{\Psi} \otimes \mathbf{I} + \mathbf{U})}}f\left(\text{vec}(\mathbf{X} - \boldsymbol{\mu}\mathbf{1}^\top)^\top(\boldsymbol{\Psi} \otimes \mathbf{I} + \mathbf{U})^{-1}\text{vec}(\mathbf{X} - \boldsymbol{\mu}\mathbf{1})\right)$$

$$= \quad c_{11}(p,n)c_{22}(p,n)\frac{\sqrt{\det(\mathbf{F}_{22})}}{\sqrt{\det(\boldsymbol{\Psi} \otimes \mathbf{I} + \mathbf{U})}}\sqrt{\det\left(\sum_{i=1}^{n}(\boldsymbol{\Psi} + \mathbf{U}_i)^{-1}\right)}$$

$$\times \quad f\left((\boldsymbol{\mu} - \tilde{\mathbf{x}}(\boldsymbol{\Psi}))^\top\left(\sum_{i=1}^{n}(\boldsymbol{\Psi} + \mathbf{U}_i)^{-1}\right)(\boldsymbol{\mu} - \tilde{\mathbf{x}}(\boldsymbol{\Psi})) + \sum_{i=1}^{n}(\mathbf{x}_i - \tilde{\mathbf{x}}(\boldsymbol{\Psi}))^\top(\boldsymbol{\Psi} + \mathbf{U}_i)^{-1}(\mathbf{x}_i - \tilde{\mathbf{x}}(\boldsymbol{\Psi}))\right)$$

$$\leq \quad c_{11}(p,n)c_{22}(p,n)\det(\boldsymbol{\Psi} + \tilde{\phi}\mathbf{I})^{-(n+p+1)/2}$$

$$\times \quad \sqrt{\det\left(\sum_{i=1}^{n}(\boldsymbol{\Psi} + \mathbf{U}_i)^{-1}\right)}f\left((\boldsymbol{\mu} - \tilde{\mathbf{x}}(\boldsymbol{\Psi}))^\top\left(\sum_{i=1}^{n}(\boldsymbol{\Psi} + \mathbf{U}_i)^{-1}\right)(\boldsymbol{\mu} - \tilde{\mathbf{x}}(\boldsymbol{\Psi}))\right), \tag{23}$$

where (23) is proportional to the kernel of a generalized matrix-variate beta type II distribution with parameters $(p+1)/2$, $n/2$, and $\tilde{\phi}\mathbf{I}$ (see, e.g., Gupta and Nagar (2000)), which is a proper density as soon as $p \leq n$. The expression (23) is the kernel of the $p$-dimensional elliptically countered distribution with density generator $f(.)$. Since $f(.)$ is a density generator of a $p \times n$-dimensional matrix-variate

elliptically contoured distribution, the integral of (23) over $\boldsymbol{\mu} \in \mathbb{R}^p$ converges. Hence, the joint posterior $\pi(\boldsymbol{\mu}, \boldsymbol{\Psi}|\mathbf{X})$ derived under the Jeffreys prior is proper under the condition $p \leq n$.

In the case of the Berger and Bernardo reference prior we get that the posterior for $\boldsymbol{\mu}$ and $\boldsymbol{\Psi}$ is bounded by

$$
\begin{aligned}
g_R(\boldsymbol{\mu}, \boldsymbol{\Psi}|\mathbf{X}) &= \frac{\pi_R(\boldsymbol{\Psi})}{\sqrt{\det(\boldsymbol{\Psi} \otimes \mathbf{I} + \mathbf{U})}} f\left(\text{vec}(\mathbf{X} - \boldsymbol{\mu}\mathbf{1}^\top)^\top (\boldsymbol{\Psi} \otimes \mathbf{I} + \mathbf{U})^{-1} \text{vec}(\mathbf{X} - \boldsymbol{\mu}\mathbf{1})\right) \\
&= c_{22}(p, n) \frac{\sqrt{\det(\mathbf{F}_{22})}}{\sqrt{\det(\boldsymbol{\Psi} \otimes \mathbf{I} + \mathbf{U})}} \frac{1}{\sqrt{\det\left(\sum_{i=1}^n (\boldsymbol{\Psi} + \mathbf{U}_i)^{-1}\right)}} \\
&\quad \times \sqrt{\det\left(\sum_{i=1}^n (\boldsymbol{\Psi} + \mathbf{U}_i)^{-1}\right)} f\left((\boldsymbol{\mu} - \tilde{\mathbf{x}}(\boldsymbol{\Psi}))^\top \left(\sum_{i=1}^n (\boldsymbol{\Psi} + \mathbf{U}_i)^{-1}\right)(\boldsymbol{\mu} - \tilde{\mathbf{x}}(\boldsymbol{\Psi}))\right).
\end{aligned}
$$

It holds that (see, Lütkepohl (1996, p.55))

$$
n^p \det\left(\frac{1}{n} \sum_{i=1}^n (\boldsymbol{\Psi} + \mathbf{U}_i)^{-1}\right) \geq n^p \prod_{i=1}^n \det\left(((\boldsymbol{\Psi} + \mathbf{U}_i)^{-1})^{1/n}\right) = n^p \prod_{i=1}^n \det(\boldsymbol{\Psi} + \mathbf{U}_i)^{-1/n},
$$

which together with (22) implies that

$$
\begin{aligned}
g_R(\boldsymbol{\mu}, \boldsymbol{\Psi}|\mathbf{X}) &\leq n^{-p/2} c_{22}(p, n) \frac{\sqrt{\det(\mathbf{F}_{22})}}{\sqrt{\det(\boldsymbol{\Psi} \otimes \mathbf{I} + \mathbf{U})^{1-1/n}}} \\
&\quad \times \sqrt{\det\left(\sum_{i=1}^n (\boldsymbol{\Psi} + \mathbf{U}_i)^{-1}\right)} f\left((\boldsymbol{\mu} - \tilde{\mathbf{x}}(\boldsymbol{\Psi}))^\top \left(\sum_{i=1}^n (\boldsymbol{\Psi} + \mathbf{U}_i)^{-1}\right)(\boldsymbol{\mu} - \tilde{\mathbf{x}}(\boldsymbol{\Psi}))\right) \\
&\leq n^{-p/2} c_{22}(p, n) \det(\boldsymbol{\Psi} + \tilde{\phi}\mathbf{I})^{-(n+p)/2} \quad\quad\quad\quad\quad\quad\quad\quad\quad\quad (24) \\
&\quad \times \sqrt{\det\left(\sum_{i=1}^n (\boldsymbol{\Psi} + \mathbf{U}_i)^{-1}\right)} f\left((\boldsymbol{\mu} - \tilde{\mathbf{x}}(\boldsymbol{\Psi}))^\top \left(\sum_{i=1}^n (\boldsymbol{\Psi} + \mathbf{U}_i)^{-1}\right)(\boldsymbol{\mu} - \tilde{\mathbf{x}}(\boldsymbol{\Psi}))\right). \quad (25)
\end{aligned}
$$

The last line coincides with (23) and it is integrable in $\mathbb{R}^p$. Moreover, (24) is proportional to the kernel of a generalized matrix-variate beta type II distribution with parameters $(p+1)/2$, $(n-1)/2$, and $\tilde{\phi}\mathbf{I}$ (see, e.g., Gupta and Nagar (2000)), which is a proper density as soon as $p+1 \leq n$. Thus, the joint posterior $\pi(\boldsymbol{\mu}, \boldsymbol{\Psi}|\mathbf{X})$ derived under the Berger and Bernardo reference prior is proper under the condition $p + 1 \leq n$. $\qquad\square$

# 4 Drawing samples from the posterior distribution: Metropolis-Hastings algorithm

In this section we develop algorithms to draw samples $(\boldsymbol{\mu}^{(b)}, \boldsymbol{\Psi}^{(b)})$ from the posterior derived under the Berger and Bernardo reference prior and the Jeffreys prior. The idea is based on the application of the Markov chain Monte Carlo based on the Metropolis-Hastings algorithm, a popular approach is Bayesian statistics (see, e.g., Givens and Hoeting (2012)). Recently, Hill and Spall (2019) provided a comprehensive discussion of the stationarity and convergence of the algorithm, that depends on the chosen proposal from which the samples are generated. A good proposal distribution should have the support which covers the support of the target distribution, i.e., of the posterior for $\boldsymbol{\mu}$ and $\boldsymbol{\Psi}$. Also, it should ensure that that the constructed Markov chain has good mixing properties and it will not stack in a single point.

As a proposal, we suggest to use the special case of the posterior distribution derived under each of the considered prior in the case $\mathbf{U}_1 = ... = \mathbf{U}_n = \mathbf{O}$. The two proposals are then defined for all positive semi-definite matrices, thus having the same supports as the two posteriors derived under the Berger and Bernardo reference prior and the Jeffreys prior. More precisely, ignoring the normalizing constants the proposal under the Berger and Bernardo reference prior is given by

$$q_R(\boldsymbol{\mu}, \boldsymbol{\Psi}|\mathbf{X}) = \det(\boldsymbol{\Psi})^{-(n+p+1)/2} f\left(\operatorname{tr}\left(\boldsymbol{\Psi}^{-1} \sum_{i=1}^{n} (\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^\top\right)\right),$$

and it is expressed as

$$q_J(\boldsymbol{\mu}, \boldsymbol{\Psi}|\mathbf{X}) = \det(\boldsymbol{\Psi})^{-(n+p+2)/2} f\left(\operatorname{tr}\left(\boldsymbol{\Psi}^{-1} \sum_{i=1}^{n} (\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^\top\right)\right)$$

under the Jeffreys prior.

Let

$$\bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i \quad \text{and} \quad \mathbf{S} = \frac{1}{n-1} \sum_{i=1}^{n} (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^\top, \tag{26}$$

In using that

$$\sum_{i=1}^{n} (\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^\top = (n-1)\mathbf{S} + n(\boldsymbol{\mu} - \bar{\mathbf{x}})(\boldsymbol{\mu} - \bar{\mathbf{x}})^\top, \tag{27}$$

which implies

$$\det(\sum_{i=1}^{n} (\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^\top) = \det((n-1)\mathbf{S})\left(1 + \frac{n}{n-1}(\boldsymbol{\mu} - \bar{\mathbf{x}})^\top \mathbf{S}^{-1}(\boldsymbol{\mu} - \bar{\mathbf{x}})\right)$$

we get

$$q_R(\boldsymbol{\mu}, \boldsymbol{\Psi}|\mathbf{X}) \propto \left(1 + \frac{1}{n-p}\frac{n(n-p)}{n-1}(\boldsymbol{\mu} - \bar{\mathbf{x}})^\top \mathbf{S}^{-1}(\boldsymbol{\mu} - \bar{\mathbf{x}})\right)^{-n/2} \tag{28}$$

$$\times \quad \det(\boldsymbol{\Psi})^{-(n+p+1)/2} \det(\sum_{i=1}^{n} (\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^\top)^{n/2} f\left(\operatorname{tr}\left(\boldsymbol{\Psi}^{-1} \sum_{i=1}^{n} (\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^\top\right)\right)$$

and

$$q_J(\boldsymbol{\mu}, \boldsymbol{\Psi}|\mathbf{X}) \propto \left(1 + \frac{1}{n-p+1}\frac{n(n-p+1)}{n-1}(\boldsymbol{\mu} - \bar{\mathbf{x}})^\top \mathbf{S}^{-1}(\boldsymbol{\mu} - \bar{\mathbf{x}})\right)^{-(n+1)/2} \tag{29}$$

$$\times \quad \det(\boldsymbol{\Psi})^{-(n+p+2)/2} \det(\sum_{i=1}^{n} (\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^\top)^{(n+1)/2} f\left(\operatorname{tr}\left(\boldsymbol{\Psi}^{-1} \sum_{i=1}^{n} (\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^\top\right)\right).$$

The expression of the proposal $q_R(\boldsymbol{\mu}, \boldsymbol{\Psi}|\mathbf{X})$ derived under the Berger and Bernardo reference prior is proportional to the joint density function of $\boldsymbol{\mu}$ and $\boldsymbol{\Psi}$ with $\boldsymbol{\Psi}|\boldsymbol{\mu}, \mathbf{X} \sim GIW_p(n+p+1, \sum_{i=1}^{n}(\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^\top, f)$ (generalized $p$-dimensional inverse Wishart distribution with $n+p+1$ degrees of freedom, scale matrix $\sum_{i=1}^{n}(\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^\top$, and density generator $f$, see, e.g., Sutradhar and Ali (1989)) and $\boldsymbol{\mu}|\mathbf{X} \sim t_p\left(n-p, \bar{\mathbf{x}}, \frac{(n-1)\mathbf{S}}{n(n-p)}\right)$ ($p$-dimensional multivariate $t$-distribution with $n-p$ degrees of freedom, location vector $\bar{\mathbf{x}}$, and scale matrix $\frac{(n-1)\mathbf{S}}{n(n-p)}$). Similarly, we get that the proposal

under the Jeffreys prior is proportional to the joint density function of $\boldsymbol{\mu}$ and $\boldsymbol{\Psi}$ with $\boldsymbol{\Psi}|\boldsymbol{\mu}, \mathbf{X} \sim GIW_p(n + p + 2, \sum_{i=1}^{n}(\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^\top, f)$ and $\boldsymbol{\mu}|\mathbf{X} \sim t_p\left(n - p + 1, \bar{\mathbf{x}}, \frac{(n-1)\mathbf{S}}{n(n-p+1)}\right)$.

We finally note that both proposals (28) and (29) are proper under the conditions $n \geq p + 1$ and $n \geq p$, respectively, which coincides with the conditions needed for the propriety of the posteriors derived under the Berger and Bernardo reference prior and the Jeffreys prior in Theorem 4. As a result, the suggested proposal possesses the similar tail behaviour as the corresponding posteriors and, thus, they are good candidates for the construction of the Markov chains.

---

**Algorithm 1** Metropolis-Hastings algorithm for drawing realizations from $\pi(\boldsymbol{\mu}, \boldsymbol{\Psi}|\mathbf{X})$ as in (13) under the Berger and Bernardo reference prior (8)

---

(1) **Initialization:** Choose the initial values $\boldsymbol{\mu}^{(0)}$ and $\boldsymbol{\Psi}^{(0)}$ for $\boldsymbol{\mu}$ and $\boldsymbol{\Psi}$ and set $b = 0$.

(2) **Generating new values of $\boldsymbol{\mu}^{(w)}$ and $\boldsymbol{\Psi}^{(w)}$ from the proposal:**

   (i) For given data $\mathbf{X} = (\mathbf{x}_1, ..., \mathbf{x}_n)$, generate $\boldsymbol{\mu}^{(w)}$ from $t_p\left(n - p, \bar{\mathbf{x}}, \frac{(n-1)\mathbf{S}}{n(n-p)}\right)$ with $\bar{\mathbf{x}}$ and $\mathbf{S}$ as in (26);

   (ii) Using data $\mathbf{X}$ and the drawn in step (i) $\boldsymbol{\mu}^{(w)}$, generate $\boldsymbol{\Psi}^{(w)}$ from $\boldsymbol{\Psi}|\boldsymbol{\mu} = \boldsymbol{\mu}^{(w)}, \mathbf{X} \sim GIW_p(n + p + 1, \sum_{i=1}^{n}(\mathbf{x}_i - \boldsymbol{\mu}^{(w)})(\mathbf{x}_i - \boldsymbol{\mu}^{(w)})^\top, f)$.

(3) **Computation of the Metropolis-Hastings ratio:**

$$MH^{(b)} = \frac{\pi(\boldsymbol{\mu}^{(w)}, \boldsymbol{\Psi}^{(w)}|\mathbf{X})q_R(\boldsymbol{\mu}^{(b-1)}, \boldsymbol{\Psi}^{(b-1)}|\mathbf{X})}{\pi(\boldsymbol{\mu}^{(b-1)}, \boldsymbol{\Psi}^{(b-1)}|\mathbf{X})q_R(\boldsymbol{\mu}^{(w)}, \boldsymbol{\Psi}^{(w)}|\mathbf{X})}.$$

(4) **Moving to the next state of the Markov chain:**

   (i) Generate $U^{(b)}$ from the uniform distribution on $[0, 1]$;

   (ii) If $U^b < \min\left\{1, MH^{(b)}\right\}\pi(\boldsymbol{\mu}^{(b)}$, then set $\boldsymbol{\mu}^{(b)} = \boldsymbol{\mu}^{(w)}$ and $\boldsymbol{\Psi}^{(b)} = \boldsymbol{\Psi}^{(w)}$ (Markov chain moves to the new state). Otherwise, set $\boldsymbol{\mu}^{(b)} = \boldsymbol{\mu}^{(b-1)}$ and $\boldsymbol{\Psi}^{(b)} = \boldsymbol{\Psi}^{(b-1)}$ (Markov chain stays in the previous state).

(5) Return to step (2), increase $b$ by 1, and repeat until the sample of size $B$ is accumulated.

---

The Metropolis-Hastings algorithm for generating a draw from $\pi(\boldsymbol{\mu}, \boldsymbol{\Psi}|\mathbf{X})$ derived under the Berger and Bernardo reference prior is given in Algorithm 1. A similar algorithm with minor changes is constructed to draw a sample from the posterior $\pi(\boldsymbol{\mu}, \boldsymbol{\Psi}|\mathbf{X})$ derived under the Jeffreys prior. It is summarized in Algorithm 2.

# 5   Several families of elliptical distributions

In this section we apply the obtained theoretical results in case of two special families of elliptically contoured distribution: normal distribution and $t$-distribution.

**Algorithm 2** Metropolis-Hastings algorithm for drawing realizations from $\pi(\boldsymbol{\mu}, \boldsymbol{\Psi}|\mathbf{X})$ as in (13) under the Jeffreys prior (7)

(1) **Initialization:** Choose the initial values $\boldsymbol{\mu}^{(0)}$ and $\boldsymbol{\Psi}^{(0)}$ for $\boldsymbol{\mu}$ and $\boldsymbol{\Psi}$ and set $b = 0$.

(2) **Generating new values of $\boldsymbol{\mu}^{(w)}$ and $\boldsymbol{\Psi}^{(w)}$ from the proposal:**

   (i) For given data $\mathbf{X} = (\mathbf{x}_1, ..., \mathbf{x}_n)$, generate $\boldsymbol{\mu}^{(w)}$ from $t_p\left(n - p + 1, \bar{\mathbf{x}}, \dfrac{(n-1)\mathbf{S}}{n(n-p+1)}\right)$ with $\bar{\mathbf{x}}$ and $\mathbf{S}$ as in (26);

   (ii) Using data $\mathbf{X}$ and the drawn in step (i) $\boldsymbol{\mu}^{(w)}$, generate $\boldsymbol{\Psi}^{(w)}$ from $\boldsymbol{\Psi}|\boldsymbol{\mu} = \boldsymbol{\mu}^{(w)}, \mathbf{X} \sim GIW_p(n + p + 2, \sum_{i=1}^{n}(\mathbf{x}_i - \boldsymbol{\mu}^{(w)})(\mathbf{x}_i - \boldsymbol{\mu}^{(w)})^{\top}, f)$.

(3) **Computation of the Metropolis-Hastings ratio:**

$$MH^{(b)} = \frac{\pi(\boldsymbol{\mu}^{(w)}, \boldsymbol{\Psi}^{(w)}|\mathbf{X})q_J(\boldsymbol{\mu}^{(b-1)}, \boldsymbol{\Psi}^{(b-1)}|\mathbf{X})}{\pi(\boldsymbol{\mu}^{(b-1)}, \boldsymbol{\Psi}^{(b-1)}|\mathbf{X})q_J(\boldsymbol{\mu}^{(w)}, \boldsymbol{\Psi}^{(w)}|\mathbf{X})}.$$

(4) **Moving to the next state of the Markov chain:**

   (i) Generate $U^{(b)}$ from the uniform distribution on $[0, 1]$;

   (ii) If $U^b < \min\left\{1, MH^{(b)}\right\}\pi(\boldsymbol{\mu}^{(b)},$ then set $\boldsymbol{\mu}^{(b)} = \boldsymbol{\mu}^{(w)}$ and $\boldsymbol{\Psi}^{(b)} = \boldsymbol{\Psi}^{(w)}$ (Markov chain moves to the new state). Otherwise, set $\boldsymbol{\mu}^{(b)} = \boldsymbol{\mu}^{(b-1)}$ and $\boldsymbol{\Psi}^{(b)} = \boldsymbol{\Psi}^{(b-1)}$ (Markov chain stays in the previous state).

(5) Return to step (2), increase $b$ by 1, and repeat until the sample of size $B$ is accumulated.

## 5.1 Normal multivariate random effects model

In the case of the normal multivariate random effects model (2), we have

$$f(u) = K_{p,n} \exp(-u/2) \quad \text{with} \quad K_{p,n} = (2\pi)^{-pn/2}, \tag{30}$$

which directly yields

$$f_{\boldsymbol{\Psi},\mathbf{X}}(u) = \frac{1}{(2\pi)^{pn/2}} \exp\left(-\frac{u}{2}\right) \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{x}_i - \tilde{\mathbf{x}}(\boldsymbol{\Psi}))^\top (\boldsymbol{\Psi} + \mathbf{U}_i)^{-1}(\mathbf{x}_i - \tilde{\mathbf{x}}(\boldsymbol{\Psi}))\right).$$

The last equality leads to the conclusion that the conditional posterior for $\boldsymbol{\mu}$ given $\boldsymbol{\Psi}$ is a multivariate normal distribution expressed as

$$\boldsymbol{\mu}|\boldsymbol{\Psi}, \mathbf{X} \sim \mathcal{N}\left(\left(\sum_{i=1}^{n}(\boldsymbol{\Psi} + \mathbf{U}_i)^{-1}\right)^{-1}\sum_{i=1}^{n}(\boldsymbol{\Psi} + \mathbf{U}_i)^{-1}\mathbf{x}_i, \left(\sum_{i=1}^{n}(\boldsymbol{\Psi} + \mathbf{U}_i)^{-1}\right)^{-1}\right), \tag{31}$$

while the marginal posterior for $\boldsymbol{\Psi}$ is given by

$$
\begin{aligned}
\pi(\boldsymbol{\Psi}|\mathbf{X}) \quad &\propto \quad \frac{\pi(\boldsymbol{\Psi})}{\sqrt{\det(\sum_{i=1}^{n}(\boldsymbol{\Psi} + \mathbf{U}_i)^{-1})}\prod_{i=1}^{n}\sqrt{\det(\boldsymbol{\Psi} + \mathbf{U}_i)}} \\
&\times \quad \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{x}_i - \tilde{\mathbf{x}}(\boldsymbol{\Psi}))^\top (\boldsymbol{\Psi} + \mathbf{U}_i)^{-1}(\mathbf{x}_i - \tilde{\mathbf{x}}(\boldsymbol{\Psi}))\right).
\end{aligned} \tag{32}
$$

The posterior mean vector and the posterior covariance matrix of $\boldsymbol{\mu}$ are obtained as in (18) and (19) with $C(\boldsymbol{\Psi}) = 1$. Finally, we note that the posterior $\pi(\boldsymbol{\mu}, \boldsymbol{\Psi}|\mathbf{X})$ is proper for $n \geq p+1$ for the Berger and Bernardo reference prior and for $n \geq p$ for the Jeffreys prior following Theorem 4, since $\exp(-u/2)$ is a decreasing function in $u$ and $\frac{J_2}{2pn + p^2 n^2} - \frac{1}{4} = 0$.

All the derived expressions for the normal multivariate random effects model, like conditional posterior for $\boldsymbol{\mu}$, posterior mean vector, etc., depend on the marginal posterior for $\boldsymbol{\Psi}$ and thus cannot be computed analytically. In the univariate case, Bodnar et al. (2016) suggested a numerical procedure for the computation of such quantities based on the evaluation of one-dimensional integral. In the multivariate case $\boldsymbol{\Psi}$ is a matrix now and since it should be positive semidefinite it imposes further complications on the numerical integration. For that reason we opt for the simulation-based approach as described in Section 4.

For generating samples from the posterior $\pi(\boldsymbol{\mu}, \boldsymbol{\Psi}|\mathbf{X})$ we apply Algorithm 1 under the Berger and Bernardo reference prior and Algorithm 2 under the Jeffreys prior where the inverse generalized Wishart distribution becomes the inverse Wishart distribution with $n + p + 1$ and $n + p + 2$ degrees of freedom, respectively. Other parts of the algorithms remain the same without changes.

Alternatively, one can modify these two algorithms using the properties of the normal distribution. Under the Berger and Bernardo reference prior, another proposal distribution can be constructed by using (28) with (30). Namely, from (26) and (27) we get

$$
\begin{aligned}
&\exp\left(-\frac{1}{2}\text{tr}\left(\boldsymbol{\Psi}^{-1}\sum_{i=1}^{n}(\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^\top\right)\right) \\
&= \quad \exp\left(-\frac{n-1}{2}\text{tr}\left(\boldsymbol{\Psi}^{-1}\mathbf{S}\right)\right)\exp\left(-\frac{n}{2}(\boldsymbol{\mu} - \bar{\mathbf{x}})^\top \boldsymbol{\Psi}^{-1}(\boldsymbol{\mu} - \bar{\mathbf{x}})\right).
\end{aligned}
$$

This leads to the algorithm derived under the Berger and Bernardo reference prior which is summarized in Algorithm 3. A similar approach can also be used when the Jeffreys prior is employed with the

---

**Algorithm 3** Metropolis-Hastings algorithm for drawing realizations from $\pi(\boldsymbol{\mu}, \boldsymbol{\Psi}|\mathbf{X})$ as in (13) under the Berger and Bernardo reference prior (8) in the normal multivariate random effects model

---

(1) **Initialization:** Choose the initial values $\boldsymbol{\mu}^{(0)}$ and $\boldsymbol{\Psi}^{(0)}$ for $\boldsymbol{\mu}$ and $\boldsymbol{\Psi}$ and set $b = 0$.

(2) **Generating new values of $\boldsymbol{\mu}^{(w)}$ and $\boldsymbol{\Psi}^{(w)}$ from the proposal:**

   (i) For given data $\mathbf{X} = (\mathbf{x}_1, ..., \mathbf{x}_n)$, generate $\boldsymbol{\Psi}^{(w)}$ from $\boldsymbol{\Psi}^{(w)}|\mathbf{X} \sim IW_p(n + p, (n - 1)\mathbf{S})$;

   (ii) using data $\mathbf{X}$ and the drawn in step (i) $\boldsymbol{\Psi}^{(w)}$, generate $\boldsymbol{\mu}^{(w)}$ from $\boldsymbol{\mu}|\boldsymbol{\Psi} = \boldsymbol{\Psi}^{(w)}, \mathbf{X} \sim$
   $N_p\left(\bar{\mathbf{x}}, \dfrac{\boldsymbol{\Psi}^{(w)}}{n}\right)$ with $\bar{\mathbf{x}}$ and $\mathbf{S}$ as in (26).

(3) **Computation of the Metropolis-Hastings ratio:**

$$MH^{(b)} = \frac{\pi(\boldsymbol{\mu}^{(w)}, \boldsymbol{\Psi}^{(w)}|\mathbf{X})q_R(\boldsymbol{\mu}^{(b-1)}, \boldsymbol{\Psi}^{(b-1)}|\mathbf{X})}{\pi(\boldsymbol{\mu}^{(b-1)}, \boldsymbol{\Psi}^{(b-1)}|\mathbf{X})q_R(\boldsymbol{\mu}^{(w)}, \boldsymbol{\Psi}^{(w)}|\mathbf{X})}.$$

(4) **Moving to the next state of the Markov chain:**

   (i) Generate $U^{(b)}$ from the uniform distribution on $[0, 1]$;

   (ii) If $U^b < \min\left\{1, MH^{(b)}\right\}\pi(\boldsymbol{\mu}^{(b)}$, then set $\boldsymbol{\mu}^{(b)} = \boldsymbol{\mu}^{(w)}$ and $\boldsymbol{\Psi}^{(b)} = \boldsymbol{\Psi}^{(w)}$ (Markov chain moves to the new state). Otherwise, set $\boldsymbol{\mu}^{(b)} = \boldsymbol{\mu}^{(b-1)}$ and $\boldsymbol{\Psi}^{(b)} = \boldsymbol{\Psi}^{(b-1)}$ (Markov chain stays in the previous state).

(5) Return to step (2), increase $b$ by 1, and repeat until the sample of size $B$ is accumulated.

---

only change in step (2) of Algorithm 3, where $\mathbf{\Psi}^{(w)}|\mathbf{X} \sim IW_p(n+p,(n-1)\mathbf{S})$ should be replaced by $\mathbf{\Psi}^{(w)}|\mathbf{X} \sim IW_p(n+p+1,(n-1)\mathbf{S})$.

The performance of two algorithms for drawing samples from the posterior distribution is studied in Figures 1 for the normal multivariate random effects model when the Berger and Bernardo reference prior and the Jeffreys prior are employed. The notation 'Algorithm A' corresponds to the case where $\boldsymbol{\mu}$ is drawn from the marginal distribution and $\mathbf{\Psi}$ is generated from the conditional distribution as in Algorithms 1 and 2 with density generator $f(.)$ as in (30), while the notation 'Algorithm B' corresponds to the case when $\mathbf{\Psi}$ is generated from the marginal distribution and $\boldsymbol{\mu}$ is obtained from the conditional distribution as in Algorithm 3.
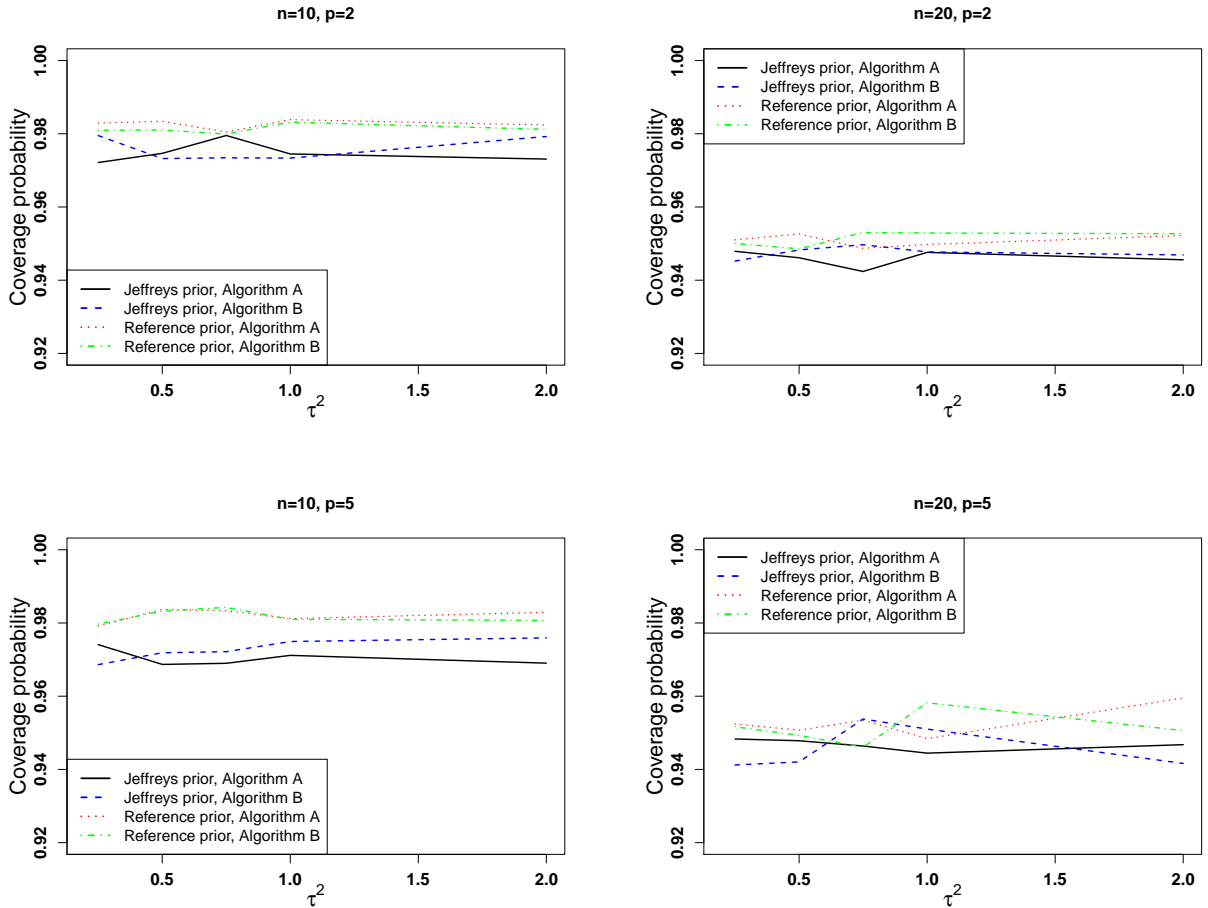


Figure 1: Coverage probabilities of the 95% credible intervals for the first component of $\boldsymbol{\mu}$ as a function of $\tau^2$ under the assumption of the normal multivariate random effects model when the Berger and Bernardo reference prior and the Jeffreys prior are employed. We set $p \in \{2,5\}$ and $n \in \{10,20\}$.

As a performance measure we use the empirical coverage probability of the credible interval constructed for $\mu_1$, which is computed based on 5000 independent repetitions. In each simulation run, the data matrix $\mathbf{X}$ is drawn from the normal multivariate random effects model (2) with the same $\boldsymbol{\mu}$, $\mathbf{\Psi} = \tau^2\mathbf{\Xi}$, and $\mathbf{U} = diag(\mathbf{U}_1,...,\mathbf{U}_p)$. The elements of $\boldsymbol{\mu}$ are generated from the uniform distribution on $[1,5]$. The eigenvalues of $\mathbf{\Xi}$, $\mathbf{U}_1$, ... , $\mathbf{U}_{p-1}$, and $\mathbf{U}_p$ are generated from the uniform distribution on $[1,4]$, while the eigenvectors are simulated from the Haar distribution. The results in Figure 1 are

17

obtained for $p \in \{2, 5\}$, $n \in \{10, 20\}$, and $\tau^2 \in \{0.25, 0.5, 0.75, 1, 2\}$.

In Figure 1 we observe that the credible intervals obtained by employing the Berger and Bernardo reference prior leads to wider credible intervals constructed for $\mu_1$, although the difference between the two non-informative priors is not large. The empirical coverage probabilities computed by using the two numerical procedures of drawing samples from the posterior distribution are above the chosen significance level of 95% in almost all of the considered cases. Furthermore, when the sample size increases to $n = 20$, then the constructed credible intervals possess almost perfect coverage probabilities of 95% for both considered dimensions $p \in \{2, 5\}$.

## 5.2 $t$ multivariate random effects model

In the case of the $t$ multivariate random effects model it holds that

$$f(u) = K_{p,n,d}(1 + u/d)^{-(pn+d)/2} \quad \text{with} \quad K_{p,n,d} = (\pi d)^{-pn/2}\frac{\Gamma\left((d+pn)/2\right)}{\Gamma\left(d/2\right)}. \tag{33}$$

Hence,

$$
\begin{aligned}
\pi(\boldsymbol{\mu}, \boldsymbol{\Psi}|\mathbf{X}) \quad &\propto \quad \frac{\pi(\boldsymbol{\Psi})}{\sqrt{\prod_{i=1}^{n} \det(\boldsymbol{\Psi} + \mathbf{U}_i)}} \\
&\times \quad \left(1 + \frac{1}{d}(\boldsymbol{\mu} - \tilde{\mathbf{x}}(\boldsymbol{\Psi}))^\top \left(\sum_{i=1}^{n}(\boldsymbol{\Psi} + \mathbf{U}_i)^{-1}\right)(\boldsymbol{\mu} - \tilde{\mathbf{x}}(\boldsymbol{\Psi})) \right. \\
&+ \quad \left. \frac{1}{d}\sum_{i=1}^{n}(\mathbf{x}_i - \tilde{\mathbf{x}}(\boldsymbol{\Psi}))^\top(\boldsymbol{\Psi} + \mathbf{U}_i)^{-1}(\mathbf{x}_i - \tilde{\mathbf{x}}(\boldsymbol{\Psi}))\right)^{-(pn+d)/2} \\
&= \quad \frac{\pi(\boldsymbol{\Psi})}{\sqrt{\prod_{i=1}^{n} \det(\boldsymbol{\Psi} + \mathbf{U}_i)}} \left(1 + \frac{1}{d}\sum_{i=1}^{n}(\mathbf{x}_i - \tilde{\mathbf{x}}(\boldsymbol{\Psi}))^\top(\boldsymbol{\Psi} + \mathbf{U}_i)^{-1}(\mathbf{x}_i - \tilde{\mathbf{x}}(\boldsymbol{\Psi}))\right)^{-(pn+d)/2} \\
&\times \quad \left(1 + \frac{1}{pn+d-p}\frac{pn+d-p}{d + \sum_{i=1}^{n}(\mathbf{x}_i - \tilde{\mathbf{x}}(\boldsymbol{\Psi}))^\top(\boldsymbol{\Psi} + \mathbf{U}_i)^{-1}(\mathbf{x}_i - \tilde{\mathbf{x}}(\boldsymbol{\Psi}))} \right. \\
&\times \quad \left. (\boldsymbol{\mu} - \tilde{\mathbf{x}}(\boldsymbol{\Psi}))^\top \left(\sum_{i=1}^{n}(\boldsymbol{\Psi} + \mathbf{U}_i)^{-1}\right)(\boldsymbol{\mu} - \tilde{\mathbf{x}}(\boldsymbol{\Psi}))\right)^{-(pn+d)/2},
\end{aligned}
$$

which shows that the conditional posterior of $\boldsymbol{\mu}$ given $\boldsymbol{\Psi}$ is

$$
\pi(\boldsymbol{\mu}|\boldsymbol{\Psi}, \mathbf{X}) \quad \propto \quad \left(1 + \frac{1}{pn+d-p}\frac{pn+d-p}{d + \sum_{i=1}^{n}(\mathbf{x}_i - \tilde{\mathbf{x}}(\boldsymbol{\Psi}))^\top(\boldsymbol{\Psi} + \mathbf{U}_i)^{-1}(\mathbf{x}_i - \tilde{\mathbf{x}}(\boldsymbol{\Psi}))}\right. \tag{34}
$$

$$
\times \quad (\boldsymbol{\mu} - \tilde{\mathbf{x}}(\boldsymbol{\Psi}))^\top \left.\left(\sum_{i=1}^{n}(\boldsymbol{\Psi} + \mathbf{U}_i)^{-1}\right)(\boldsymbol{\mu} - \tilde{\mathbf{x}}(\boldsymbol{\Psi}))\right)^{-(pn+d)/2}, \tag{35}
$$

i.e., $\boldsymbol{\mu}$ conditionally on $\boldsymbol{\Psi}$ and $\mathbf{X}$ has a $p$-dimensional $t$-distribution with $pn+d-p$ degrees of freedom, location parameter $\tilde{\mathbf{x}}(\boldsymbol{\Psi})$ and dispersion matrix

$$
\frac{d + \sum_{i=1}^{n}(\mathbf{x}_i - \tilde{\mathbf{x}}(\boldsymbol{\Psi}))^\top(\boldsymbol{\Psi} + \mathbf{U}_i)^{-1}(\mathbf{x}_i - \tilde{\mathbf{x}}(\boldsymbol{\Psi}))}{pn+d-p} \left(\sum_{i=1}^{n}(\boldsymbol{\Psi} + \mathbf{U}_i)^{-1}\right)^{-1}.
$$

Moreover, the marginal posterior for $\boldsymbol{\Psi}$ can also be deduced and it is expressed as

$$
\begin{aligned}
\pi(\boldsymbol{\Psi}|\mathbf{X}) \quad &\propto \quad \frac{\pi(\boldsymbol{\Psi})}{\sqrt{\det(\sum_{i=1}^{n}(\boldsymbol{\Psi} + \mathbf{U}_i)^{-1})}\prod_{i=1}^{n}\sqrt{\det(\boldsymbol{\Psi} + \mathbf{U}_i)}} \\
&\times \quad \left(1 + \frac{1}{d}\sum_{i=1}^{n}(\mathbf{x}_i - \tilde{\mathbf{x}}(\boldsymbol{\Psi}))^\top(\boldsymbol{\Psi} + \mathbf{U}_i)^{-1}(\mathbf{x}_i - \tilde{\mathbf{x}}(\boldsymbol{\Psi}))\right)^{-(pn+d)/2}.
\end{aligned} \tag{36}
$$

To this end, the posterior mean vector and the covariance matrix of $\boldsymbol{\mu}$ are obtained as in (18) and (19) with

$$C(\boldsymbol{\Psi}) = \frac{d + \sum_{i=1}^n (\mathbf{x}_i - \tilde{\mathbf{x}}(\boldsymbol{\Psi}))^\top (\boldsymbol{\Psi} + \mathbf{U}_i)^{-1} (\mathbf{x}_i - \tilde{\mathbf{x}}(\boldsymbol{\Psi}))}{pn + d - p - 2}.$$

Finally, we note that the constant $J_2$ can analytically be computed in the case of the $t$ multivariate random effects model and it is expressed as (see, Bodnar (2019, Section 3.2))

$$J_2 = \frac{pn(pn+2)(pn+d)}{4(pn+2+d)}.$$

The application of the last expression leads to the following formulas of the Berger and Bernardo reference prior

$$\pi_R(\boldsymbol{\Psi}) \quad \propto \quad \left( \det \left\{ \mathbf{G}_p^\top \left[ \frac{pn+d}{2(pn+2+d)} \sum_{i=1}^n \left( (\boldsymbol{\Psi} + \mathbf{U}_i)^{-1} \otimes (\boldsymbol{\Psi} + \mathbf{U}_i)^{-1} \right) \right. \right. \right. \tag{37}$$
$$\left. \left. \left. - \frac{1}{2(pn+2+d)} \mathrm{vec} \left( \sum_{i=1}^n (\boldsymbol{\Psi} + \mathbf{U}_i)^{-1} \right) \mathrm{vec} \left( \sum_{j=1}^n (\boldsymbol{\Psi} + \mathbf{U}_j)^{-1} \right)^\top \right] \mathbf{G}_p \right\} \right)^{1/2},$$

while the Jeffreys prior is given by (7) with $\pi_R(\boldsymbol{\Psi})$ as in (37). Furthermore, since $(1 + u/d)^{-(pn+d)/2}$ is a decreasing function in $u$ and

$$\frac{J_2}{2pn + p^2 n^2} - \frac{1}{4} = \frac{1}{4} \left( \frac{pn+d}{pn+d+2} - 1 \right) < 0,$$

the posterior $\pi(\boldsymbol{\mu}, \boldsymbol{\Psi} | \mathbf{X})$ is proper for $n \geq p + 1$ for the Berger and Bernardo reference prior and for $n \geq p$ for the Jeffreys prior due to Theorem 4.

Algorithm 1 and Algorithm 2 are used to draw samples from the posterior $\pi(\boldsymbol{\mu}, \boldsymbol{\Psi} | \mathbf{X})$ derived by employing the Berger and Bernardo reference prior and the Jeffreys prior, respectively. Under the special case of the $t$ multivariate random effects model, the step (ii) of the both algorithm is performed by computing $\boldsymbol{\Psi}^{(b)} = \frac{\xi^{(b)}}{d} \boldsymbol{\Omega}^{(b)}$ where $\xi^{(b)}$ and $\boldsymbol{\Omega}^{(b)}$ are simulated independently from $\chi_d^2$-distribution and the inverse Wishart distribution with parameter matrix $(n-1) \sum_{i=1}^n (\mathbf{x}_i - \boldsymbol{\mu}^{(b)})(\mathbf{x}_i - \boldsymbol{\mu}^{(b)})^\top$ and degrees of freedom equal to $n + p + 1$ under the Berger and Bernardo reference prior and $n + p + 2$ under the Jeffreys prior.

The modification of Algorithms 1 and 2 similar to the one derived for the normal multivariate random effects model can also be obtained under the assumption of the $t$-distribution. The application of the equality

$$\left( 1 + \frac{1}{d} \mathrm{tr} \left( \boldsymbol{\Psi}^{-1} \sum_{i=1}^n (\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^\top \right) \right)^{-(pn+d)/2} = \left( 1 + \frac{n-1}{d} \mathrm{tr} \left( \boldsymbol{\Psi}^{-1} \mathbf{S} \right) \right)^{-(pn+d)/2}$$
$$\times \quad \left( 1 + \frac{1}{np+d-p} \frac{n(np+d-p)}{d+(n-1)\mathrm{tr}\left(\boldsymbol{\Psi}^{-1}\mathbf{S}\right)} (\boldsymbol{\mu} - \bar{\mathbf{x}})^\top \boldsymbol{\Psi}^{-1} (\boldsymbol{\mu} - \bar{\mathbf{x}}) \right)^{-(pn+d)/2}$$

leads to another numerical procedure described in Algorithm 4 when the posterior $\pi(\boldsymbol{\mu}, \boldsymbol{\Psi} | \mathbf{X})$ is derived by applying the the Berger and Bernardo reference prior. Under the Jeffreys prior the step (2) of Algorithm 4 should be modified by generating $\boldsymbol{\Omega}^{(w)}$ from $IW_p(n + p + 1, (n-1)\mathbf{S})$.

In order to investigate the properties of the two proposed algorithms, we conduct a simulation study for the t multivariate random effects model designed similarly to the one presented in Section

---

**Algorithm 4** Metropolis-Hastings algorithm for drawing realizations from $\pi(\boldsymbol{\mu}, \boldsymbol{\Psi}|\mathbf{X})$ as in (13) under the Berger and Bernardo reference prior (8) in the $t$ multivariate random effects model

---

(1) **Initialization:** Choose the initial values $\boldsymbol{\mu}^{(0)}$ and $\boldsymbol{\Psi}^{(0)}$ for $\boldsymbol{\mu}$ and $\boldsymbol{\Psi}$ and set $b = 0$.

(2) **Generating new values of $\boldsymbol{\mu}^{(w)}$ and $\boldsymbol{\Psi}^{(w)}$ from the proposal:**

    (i) For given data $\mathbf{X} = (\mathbf{x}_1, ..., \mathbf{x}_n)$, generate $\boldsymbol{\Psi}^{(w)} = \frac{\xi^{(w)}}{d}\boldsymbol{\Omega}^{(w)}$ where $\xi^{(w)}$ and $\boldsymbol{\Omega}^{(w)}$ are simulated independently with from $\xi^w \sim \chi_d^2$ from $\boldsymbol{\Omega}^{(w)}|\mathbf{X} \sim IW_p(n+p, (n-1)\mathbf{S})$;

    (ii) Using data $\mathbf{X}$ and the drawn in step (i) $\boldsymbol{\Psi}^{(w)}$, generate $\boldsymbol{\mu}^{(w)}$ from $\boldsymbol{\mu}|\boldsymbol{\Psi} = \boldsymbol{\Psi}^{(w)}, \mathbf{X} \sim t_p\left(pn + d - p, \bar{\mathbf{x}}, \frac{d+(n-1)\mathrm{tr}\left((\boldsymbol{\Psi}^{(w)})^{-1}\mathbf{S}\right)}{n(np+d-p)}\boldsymbol{\Psi}^{(w)}\right)$ with $\bar{\mathbf{x}}$ and $\mathbf{S}$ as in (26) and the symbol $t_p(m, \mathbf{r}, \mathbf{R})$ stands for the multivariate $p$-dimensional $t$-distribution with $m$ degrees of freedom, location vector $\mathbf{r}$, and scale matrix $\mathbf{R}$.

(3) **Computation of the Metropolis-Hastings ratio:**

$$MH^{(b)} = \frac{\pi(\boldsymbol{\mu}^{(w)}, \boldsymbol{\Psi}^{(w)}|\mathbf{X})q_R(\boldsymbol{\mu}^{(b-1)}, \boldsymbol{\Psi}^{(b-1)}|\mathbf{X})}{\pi(\boldsymbol{\mu}^{(b-1)}, \boldsymbol{\Psi}^{(b-1)}|\mathbf{X})q_R(\boldsymbol{\mu}^{(w)}, \boldsymbol{\Psi}^{(w)}|\mathbf{X})}.$$

(4) **Moving to the next state of the Markov chain:**

    (i) Generate $U^{(b)}$ from the uniform distribution on $[0, 1]$;

    (ii) If $U^b < \min\left\{1, MH^{(b)}\right\}\pi(\boldsymbol{\mu}^{(b)}$, then set $\boldsymbol{\mu}^{(b)} = \boldsymbol{\mu}^{(w)}$ and $\boldsymbol{\Psi}^{(b)} = \boldsymbol{\Psi}^{(w)}$ (Markov chain moves to the new state). Otherwise, set $\boldsymbol{\mu}^{(b)} = \boldsymbol{\mu}^{(b-1)}$ and $\boldsymbol{\Psi}^{(b)} = \boldsymbol{\Psi}^{(b-1)}$ (Markov chain stays in the previous state).

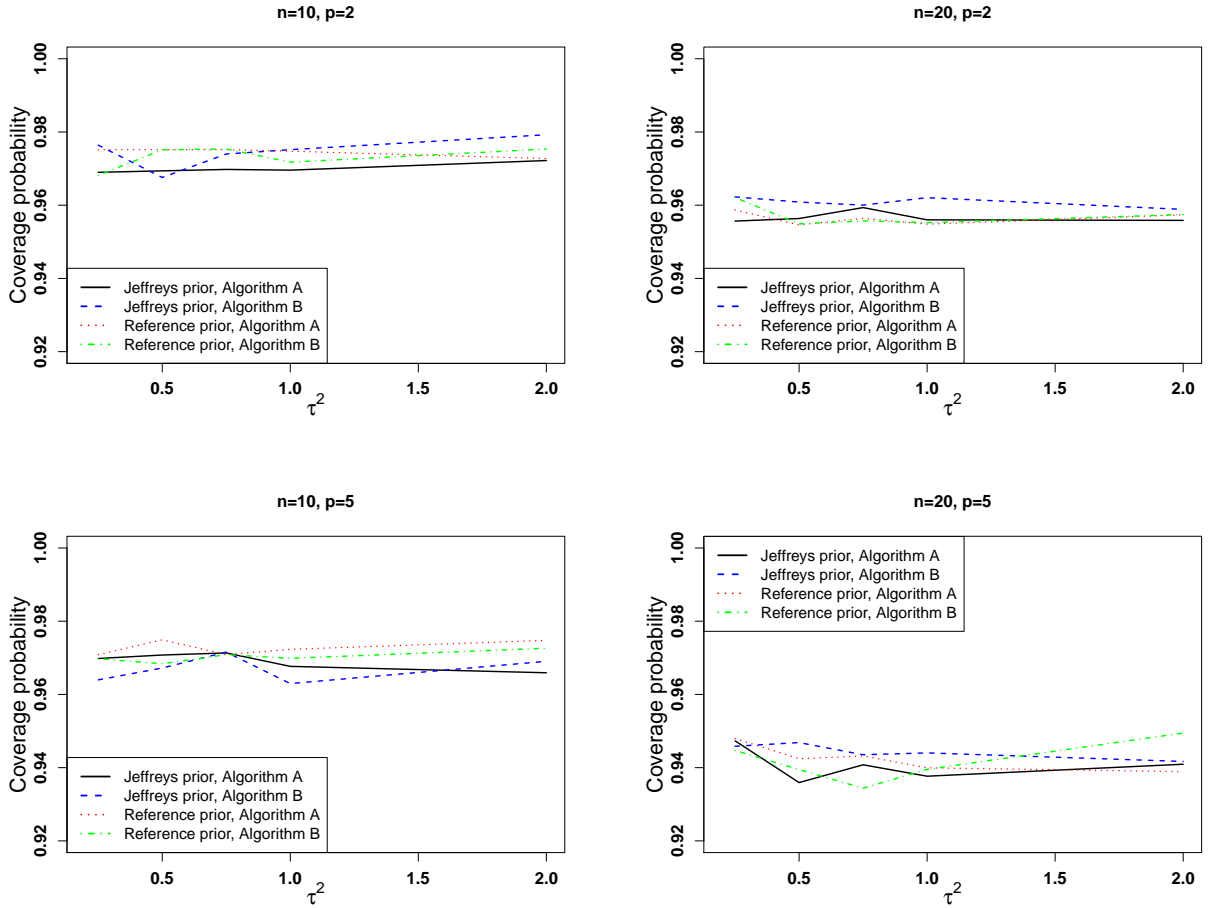(5) Return to step (2), increase $b$ by 1, and repeat until the sample of size $B$ is accumulated.

---

Figure 2: Coverage probabilities of the 95% credible intervals for the first element of $\boldsymbol{\mu}$ as a function of $\tau^2$ under the assumption of the $t$ multivariate random effects model when the Berger and Bernardo reference prior and the Jeffreys prior are employed. We set $p \in \{2, 5\}$ and $n \in \{10, 20\}$.

5.1 for the normal multivariate random effect models. Also, we use the same notations 'Algorithm A' and 'Algorithm B' to distinguish between the two procedures to draw the sample from the posterior distributions derived by employing the Berger and Bernardo reference prior and the Jeffreys prior.

We use the empirical coverage probability of the credible interval constructed for $\mu_1$ as a performance measure and compute it based on 5000 independent repetitions. In each simulation run, the data matrix $\mathbf{X}$ is simulated from the t multivariate random effects model, i.e., from the model (1) with $f(.)$ as in (33). The model parameters $\boldsymbol{\mu}$, $\boldsymbol{\Psi} = \tau^2 \boldsymbol{\Xi}$, and $\mathbf{U} = diag(\mathbf{U}_1, ..., \mathbf{U}_p)$ are chosen in the same way as the corresponding parameters of the normal multivariate random effects model in Section 5.1. Finally, we set $p \in \{2, 5\}$, $n \in \{10, 20\}$, and $\tau^2 \in \{0.25, 0.5, 0.75, 1, 2\}$.

The results of the simulation study are depicted in Figure 2. Similarly to the findings obtained for the normal multivariate random effects model, the empirical coverage probabilities are larger than the chosen significance level of 95% in almost all of the considered cases, independently whether the Markov chains are constructed following Algorithm A or Algorithm B. The application of the Berger and Bernardo reference prior leads to a slightly larger values of the empirical coverage probabilities. Finally, the coverage probabilities are close to 95% when the sample size is $n = 20$ independently of

the chosen values of $p \in \{2, 5\}$.

# 6    Empirical illustration

In this section we illustrate the derived theoretical findings on real data consisting of results obtained in ten studies that assess the effectiveness of hypertension treatment for reducing blood pressure. The treatment effects on the systolic blood pressure and diastolic blood pressure are investigated in the studies where the negative values document positive effect of the treatment. The data are provided in Table 1 and are taken from Jackson et al. (2013) where the treatment effects in each study are provided together with the covariance matrices $\mathbf{U}_i$ which are assumed to be known throughout this section.

| Study | $X_{i;1}$ (SBP) | $X_{i;2}$ (DBP) | $\sqrt{U_{i;11}}$ (SBP) | $\rho_{i;12} = \dfrac{U_{i;12}}{\sqrt{U_{i;11}U_{i;22}}}$ | $\sqrt{U_{i;22}}$ (DBP) |
|---|---|---|---|---|---|
| 1 | -6.66 | -2.99 | 0.72 | 0.78 | 0.27 |
| 2 | -14.17 | -7.87 | 4.73 | 0.45 | 1.44 |
| 3 | -12.88 | -6.01 | 10.31 | 0.59 | 1.77 |
| 4 | -8.71 | -5.11 | 0.30 | 0.77 | 0.10 |
| 5 | -8.70 | -4.64 | 0.14 | 0.66 | 0.05 |
| 6 | -10.60 | -5.56 | 0.58 | 0.49 | 0.18 |
| 7 | -11.36 | -3.98 | 0.30 | 0.50 | 0.27 |
| 8 | -17.93 | -6.54 | 5.82 | 0.61 | 1.31 |
| 9 | -6.55 | -2.08 | 0.41 | 0.45 | 0.11 |
| 10 | -10.26 | -3.49 | 0.20 | 0.51 | 0.04 |

Table 1: Data collected in 10 studies about the effectiveness of hypertension treatment with the aim to reduce blood pressure. The variables $X_{i;1}$ and $X_{i;2}$ denote the treatment effects on the systolic blood pressure (SBP) and the diastolic blood pressures (DBP) from the $i$th study, while $\mathbf{U}_i = (U_{i;lj})_{lj=1,2}$ is the corresponding covariance matrix.

Multivariate meta-analysis is performed by using data from Table 1 under the assumption of the normal multivariate random effects model (Section 5.1) and the $t$ multivariate random effects model (Section 5.2) when the Berger and Bernardo reference prior and the Jeffreys prior are employed. The samples from the joint posterior distribution $\pi(\boldsymbol{\mu}, \boldsymbol{\Psi}|\mathbf{X})$ are drawn by applying two versions of the Metropolis-Hastings algorithm which are described in Section 5.1 for the normal multivariate random effects model and denoted by Algorithm A and Algorithm B, respectively. For each type of the Metropolis-Hastings algorithm, the distributional class of the multivariate random effects model, and the chosen prior, $10^5$ realizations from the posterior distribution $\pi(\boldsymbol{\mu}, \boldsymbol{\Psi}|\mathbf{X})$ are drawn with 10% used as burn-in sample.

The first two columns of Table 2 present the results for the normal multivariate random effects model, while the results for the $t$ multivariate random effects model with $d = 3$ degrees of freedom are shown in the third and the fourth columns of the table. For each chosen prior, random effects model, and numerical algorithm to draw a sample from the posterior distribution, we compute the posterior mean and posterior median as two Bayesian point estimators for the overall mean vector

22

| | Normal random effects model | | t random effects model | |
|---|---|---|---|---|
| | $\mu_1$ (SBP) | $\mu_2$ (DBP) | $\mu_1$ (SBP) | $\mu_2$ (DBP) |
| Jeffreys prior, Algorithm A | | | | |
| post. mean | -9.79 | -4.05 | -10.15 | -4.67 |
| post. median | -9.60 | -4.27 | -10.10 | -4.66 |
| post. sd. | 0.88 | 0.93 | 1.08 | 0.60 |
| cred. inter. | [-11.73, -8.00] | [-5.61, -2.66] | [-12.44,-8.11] | [-5.87,-3.51] |
| Jeffreys prior, Algorithm B | | | | |
| post. mean | -9.78 | -4.37 | -10.03 | -4.66 |
| post. median | -9.84 | -4.37 | -9.97 | -4.65 |
| post. sd. | 0.74 | 0.50 | 1.13 | 0.61 |
| cred. inter. | [-11.46, -8.39] | [-5.38, -3.38] | [-12.46, -7.99] | [-5.90, -3.51] |
| Berger and Bernardo reference prior, Algorithm A | | | | |
| post. mean | -9.81 | -4.49 | -10.11 | -4.67 |
| post. median | -9.87 | -4.44 | -10.06 | -4.66 |
| post. sd. | 1.04 | 0.59 | 1.16 | 0.64 |
| cred. inter. | [-12.06, -8.00] | [-5.78, -3.42] | [-12.58, -7.97] | [-5.96, -3.43] |
| Berger and Bernardo reference prior, Algorithm B | | | | |
| post. mean | -9.70 | -4.51 | -10.08 | -4.68 |
| post. median | -9.72 | -4.53 | -10.03 | -4.65 |
| post. sd. | 1.01 | 0.58 | 1.13 | 0.64 |
| cred. inter. | [-11.88, -8.06] | [-5.67, -3.49] | [-12.50, -7.97] | [-5.94, -3.42] |
| REML | | | | |
| estimator | -9.50 | -4.43 | – | – |
| stand. error | 0.77 | 0.48 | – | – |
| Method of moments, Jackson et al. (2010) | | | | |
| estimator | -9.13 | -4.30 | – | – |
| stand. error | 0.54 | 0.36 | – | – |
| Method of moments, Jackson et al. (2013) | | | | |
| estimator | -9.17 | -4.31 | – | – |
| stand. error | 0.55 | 0.36 | – | – |

Table 2: Results of Bayesian inference (posterior mean, posterior median, posterior standard deviation, 95% credible interval) for the parameters of the normal multivariate random effects model obtained for the data from Table 1 by employing the Berger and Bernardo reference prior and the Jeffrey prior. The samples from the posterior distributions are drawn by Algorithm A and Algorithm B defined in Section 5.1. The last three panels of the table include the results of the restrictive maximum likelihood estimator of Chen et al. (2012), and two methods of moment estimators from Jackson et al. (2010) and Jackson et al. (2013).

together with the posterior standard deviation and 95% probability symmetric credible interval. In the case of the $t$ multivariate random effects model we multiply $\mathbf{U}_i$ by $\frac{d-2}{d}$ in order to ensure that the within-study covariance matrix calculated under the assumption of the $t$ multivariate random effects model coincides with the one given in Table 1. Finally, for comparison purposes, we also include the results obtained by three approaches of the frequentist statistics which are given in Table 3 of Jackson et al. (2013).
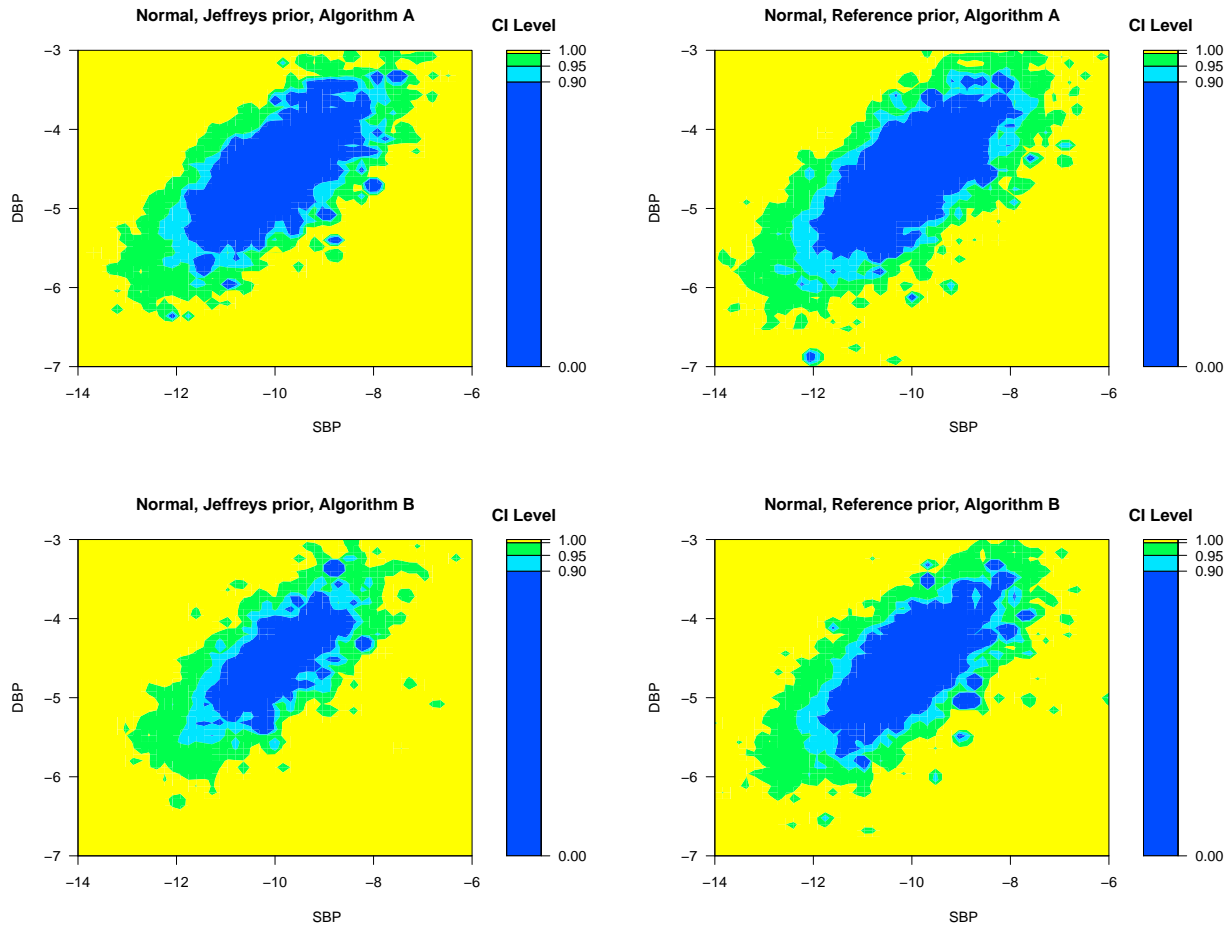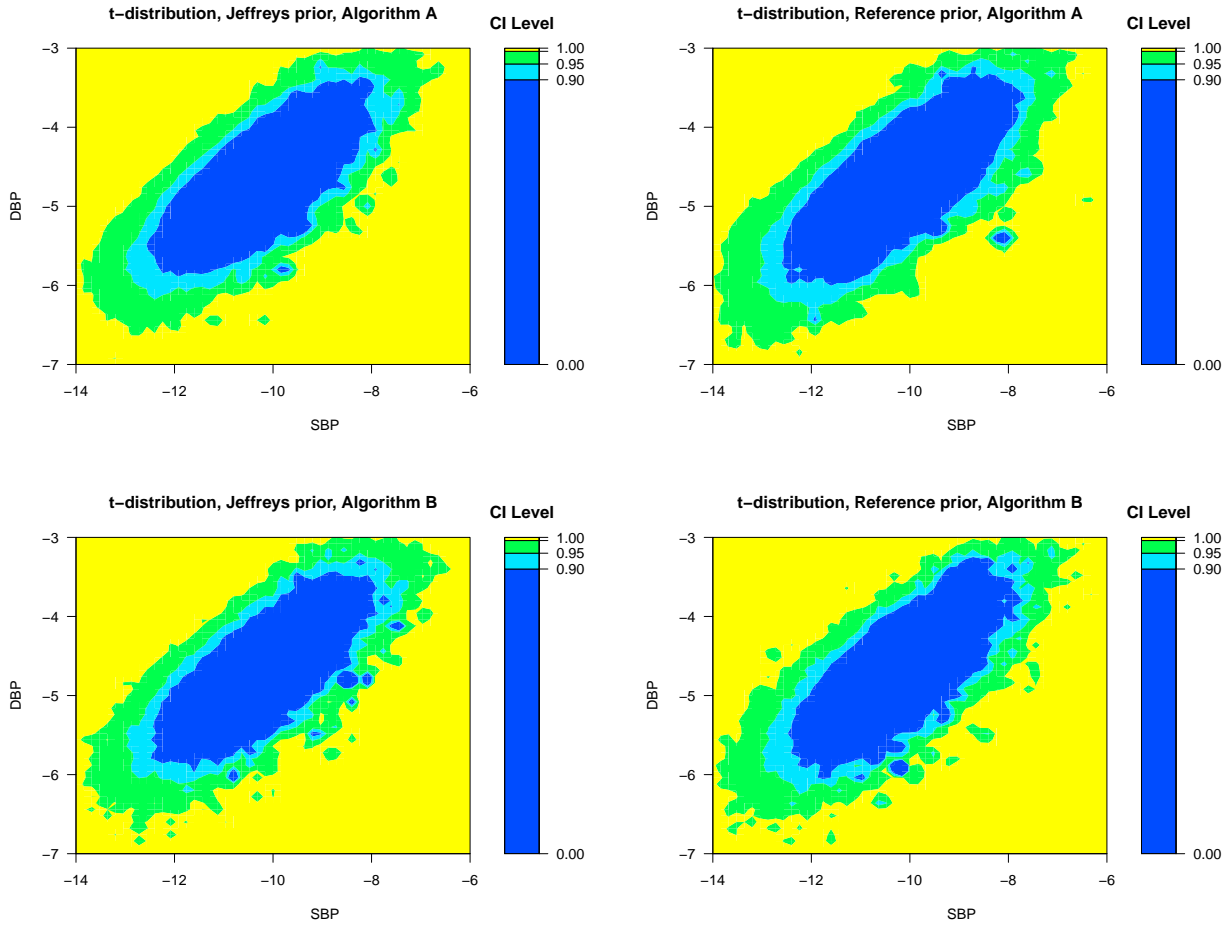


Figure 3: Credible sets for $\mu_1$ (SBP) and $\mu_2$ (DBP) obtained from the posterior distribution $\pi(\boldsymbol{\mu}|\mathbf{X})$ derived for the location parameters of the normal multivariate random effects model by employing the Berger and Bernardo reference prior and the Jeffrey prior and using data from Table 1. The samples from the posterior distributions are drawn by Algorithm A and Algorithm B defined in Section 5.1.

All Bayesian point estimators derived under the assumption of the normal multivariate random effects model are very similar and they are almost always slightly smaller than those obtained by the frequentist approaches. In contrast, the computed Bayesian standard errors are larger than those computed by the frequentist approaches, especially when the two methods of moments are used in their computation. These results are in line with statistical theory and reflect the fact that the Bayesian methods in contrast to the frequentist approaches take automatically the uncertainty about the between-study covariance matrix $\boldsymbol{\Psi}$ into account, while the frequentist methods usually ignore that $\boldsymbol{\Psi}$ is an unknown nuisance parameter of the model which has to be estimated before the inferences

for the overall mean vector are constructed. Finally, we have that the credible intervals obtained by employing the Berger and Bernardo reference prior are wider than those obtained by using the Jeffreys prior. Similar results are also obtained in the simulation study of Section 5.1 (see, Figure 1), where the larger values of the coverage probabilities are documented for the Berger and Bernardo reference prior. Such a result was also documented in the univariate case in Bodnar (2019). Finally, in the case of the $t$ multivariate random effects model, the estimated elements of the overall mean vector $\boldsymbol{\mu}$ become even smaller than those observed under the of the normal multivariate random effects model, while the corresponding Bayesian standard deviations increase reflecting the impact of heavy tails of the $t$-distribution.



Figure 4: Credible sets for $\mu_1$ (SBP) and $\mu_2$ (DBP) obtained from the posterior distribution $\pi(\boldsymbol{\mu}|\mathbf{X})$ derived for the location parameters of the normal multivariate random effects model by employing the Berger and Bernardo reference prior and the Jeffrey prior and using data from Table 1. The samples from the posterior distributions are drawn by Algorithm A and Algorithm B defined in Section 5.2.

The two-dimensional credible regions at significance levels 0.9 (dark blue), 0.95 (light blue), and 0.99 (green) for the elements of the mean vector $\boldsymbol{\mu}$ are depicted in Figure 3 for the normal multivariate random effects model and in Figure 4 for the $t$ multivariate random effects model with $d = 3$ degrees of freedom. The credible regions obtained under the assumption of the $t$-distribution are very similar, independently whether the Berger and Bernardo reference prior or the Jeffreys prior is employed, and

the chosen algorithm to draw samples from the posterior distribution. That is not longer the case in Figure 3, where the credible regions computed for the normal multivariate random effects model with Jeffreys prior and using Algorithm B appear to be slightly narrower. Finally, the credible intervals obtained under the assumption of the $t$-distribution are always wider reflecting the influence of heavy tails.

# 7 Summary

Multivariate random effects model is one of the mostly used statistical tool in multivariate meta-analysis where the aim is to combine multiple values obtained in several studies into a single value. The parameters of the multivariate random effects model are usually estimated from the viewpoint of frequentist statistics, while several subjective Bayesian approaches based on the informative priors exist in the literature. Although both methods provide a good fit of the model to real data when the sample size is relatively large due to the asymptotic theorems of the frequentist statistics and the Bernstein-von-Mises theorem in Bayesian statistics, the results might be different when a sample of small size is present which is the case in the majority of meta-analyses. When the sample size is not large enough the asymptotic approximation might deviate considerable from the exact sample distribution of the estimated parameters or/and the influence of the chosen informative prior might have a significant impact on the posterior. Methods of the objective Bayesian statistics propose a solution to the challenges related to the insufficient sample size by endowing the models parameters with noninformative prior. In particular, the Berger and Bernardo reference prior is derived by maximizing the Shannon mutual information, i.e. by choosing the prior with the smallest impact on the posterior.

Flexible objective Bayesian procedures for the parameters of the multivariate random effects model are developed by employing two noninformative priors, the Berger and Bernardo reference prior and the Jeffreys prior. The analytical expressions of both the priors are obtained and the corresponding posteriors are derived. The results are established for a general class of multivariate random effects models which include the normal multivariate random effects model as a special case. Moreover, the propriety of the posteriors is proved under a weak condition, which requires that the sample size is larger than the dimension of the data generating-model only, independently of the specific class of the multivariate random effects model. Finally, the Metropolis-Hastings algorithm has been developed in the paper to draw samples from the posterior derived for the parameters of the model. Via simulations, it is shown that the considered numerical procedures lead to similar results in the case of the normal multivariate random effects model and the $t$ multivariate random effects model. In an empirical illustration based on data consisting of ten studies about the effectiveness of hypertension treatment for reducing blood pressure, a positive effect of the treatments on both the systolic blood pressure and diastolic blood pressure are found.

# Acknowledgement

# 8 Appendix

In this section the proofs of theoretical results are given.

*Proof of Theorem 1:* Under model (1) the log-likelihood ignoring the constant term is given by

$$
\begin{aligned}
L(\boldsymbol{\mu}, \boldsymbol{\Psi}; \mathbf{X}) &= -\frac{1}{2}\log(\det(\boldsymbol{\Psi} \otimes \mathbf{I} + \mathbf{U})) + \log\left(f\left(\mathrm{vec}(\mathbf{X} - \boldsymbol{\mu}\mathbf{1}^\top)^\top(\boldsymbol{\Psi} \otimes \mathbf{I} + \mathbf{U})^{-1}\mathrm{vec}(\mathbf{X} - \boldsymbol{\mu}\mathbf{1})\right)\right) \\
&= -\frac{1}{2}\sum_{i=1}^{n}\log(\det(\boldsymbol{\Psi} + \mathbf{U}_i)) + \log\left(f\left(\sum_{i=1}^{n}(\mathbf{x}_i - \boldsymbol{\mu})^\top(\boldsymbol{\Psi} + \mathbf{U}_i)^{-1}(\mathbf{x}_i - \boldsymbol{\mu})\right)\right). \quad (38)
\end{aligned}
$$

Hence,

$$
\frac{\partial L(\boldsymbol{\mu}, \boldsymbol{\Psi}; \mathbf{X})}{\partial \boldsymbol{\mu}^\top} = -2\frac{f'\left(\sum_{i=1}^{n}(\mathbf{x}_i - \boldsymbol{\mu})^\top(\boldsymbol{\Psi} + \mathbf{U}_i)^{-1}(\mathbf{x}_i - \boldsymbol{\mu})\right)}{f\left(\sum_{i=1}^{n}(\mathbf{x}_i - \boldsymbol{\mu})^\top(\boldsymbol{\Psi} + \mathbf{U}_i)^{-1}(\mathbf{x}_i - \boldsymbol{\mu})\right)}\sum_{i=1}^{n}(\mathbf{x}_i - \boldsymbol{\mu})^\top(\boldsymbol{\Psi} + \mathbf{U}_i)^{-1}.
$$

Next, we compute the partial derivative of the log-likelihood function with respect to $\mathrm{vech}(\boldsymbol{\Psi})$, where vech denote for the *vech* operator with the following relation to vec (see, Harville (1997, p. 365))

$$
\mathbf{vec}(\boldsymbol{\Psi}) = \mathbf{G}_p\mathbf{vech}(\boldsymbol{\Psi}),
$$

where $\mathbf{G}_p$ is the duplication matrix (see, Magnus and Neudecker (2019, Section 3)).

From the properties of the differential of a determinant (see, Magnus and Neudecker (2019, Section 8)) we get

$$
\begin{aligned}
\mathbf{d}\det(\boldsymbol{\Psi} + \mathbf{U}_i) &= \det(\boldsymbol{\Psi} + \mathbf{U}_i)(\mathrm{vec}((\boldsymbol{\Psi} + \mathbf{U}_i)^{-1}))^\top\mathbf{d}\,\mathrm{vec}(\boldsymbol{\Psi} + \mathbf{U}_i) \\
&= \det(\boldsymbol{\Psi} + \mathbf{U}_i)(\mathrm{vec}((\boldsymbol{\Psi} + \mathbf{U}_i)^{-1}))^\top\mathbf{d}\,\mathrm{vec}(\boldsymbol{\Psi}) \\
&= \det(\boldsymbol{\Psi} + \mathbf{U}_i)(\mathrm{vech}((\boldsymbol{\Psi} + \mathbf{U}_i)^{-1}))^\top\mathbf{G}_p^\top\mathbf{G}_p\mathbf{d}\,\mathrm{vech}(\boldsymbol{\Psi}).
\end{aligned}
$$

Thus,

$$
\frac{\partial\det(\boldsymbol{\Psi} + \mathbf{U}_i)}{\partial\mathrm{vech}(\boldsymbol{\Psi})^\top} = \det(\boldsymbol{\Psi} + \mathbf{U}_i)\mathrm{vech}((\boldsymbol{\Psi} + \mathbf{U}_i)^{-1})^\top\mathbf{G}_p^\top\mathbf{G}_p
$$

and

$$
\frac{\partial\log\det(\boldsymbol{\Psi} + \mathbf{U}_i)}{\partial\mathrm{vech}(\boldsymbol{\Psi})^\top} = \mathrm{vech}((\boldsymbol{\Psi} + \mathbf{U}_i)^{-1})^\top\mathbf{G}_p^\top\mathbf{G}_p. \quad (39)
$$

Similar, using the properties of the differential of inverse matrix (see, Magnus and Neudecker (2019, Section 8)), we obtain

$$
\mathbf{d}\,(\boldsymbol{\Psi} + \mathbf{U}_i)^{-1} = -(\boldsymbol{\Psi} + \mathbf{U}_i)^{-1}\mathbf{d}\,(\boldsymbol{\Psi} + \mathbf{U}_i)(\boldsymbol{\Psi} + \mathbf{U}_i)^{-1} = -(\boldsymbol{\Psi} + \mathbf{U}_i)^{-1}\mathbf{d}\,\boldsymbol{\Psi}(\boldsymbol{\Psi} + \mathbf{U}_i)^{-1}
$$

and, consequently,

$$
\begin{aligned}
\text{vec}(\mathbf{d}\ (\boldsymbol{\Psi} + \mathbf{U}_i)^{-1}) &= \mathbf{d}\ \text{vec}((\boldsymbol{\Psi} + \mathbf{U}_i)^{-1}) \\
&= -\text{vec}((\boldsymbol{\Psi} + \mathbf{U}_i)^{-1}\mathbf{d}\ \boldsymbol{\Psi}(\boldsymbol{\Psi} + \mathbf{U}_i)^{-1}) \\
&= -\left((\boldsymbol{\Psi} + \mathbf{U}_i)^{-1} \otimes (\boldsymbol{\Psi} + \mathbf{U}_i)^{-1}\right)\mathbf{d}\text{vec}(\boldsymbol{\Psi}) \\
&= -\left((\boldsymbol{\Psi} + \mathbf{U}_i)^{-1} \otimes (\boldsymbol{\Psi} + \mathbf{U}_i)^{-1}\right)\mathbf{G}_p\mathbf{d}\ \text{vech}(\boldsymbol{\Psi}).
\end{aligned}
$$

Hence,

$$
\frac{\partial \text{vec}(\boldsymbol{\Psi} + \mathbf{U}_i)^{-1}}{\partial \text{vech}(\boldsymbol{\Psi})^{\top}} = -\left((\boldsymbol{\Psi} + \mathbf{U}_i)^{-1} \otimes (\boldsymbol{\Psi} + \mathbf{U}_i)^{-1}\right)\mathbf{G}_p. \tag{40}
$$

The application of (39) and (40) leads to

$$
\begin{aligned}
\frac{\partial L(\boldsymbol{\mu}, \boldsymbol{\Psi}; \mathbf{X})}{\partial \text{vech}(\boldsymbol{\Psi})^{\top}} &= -\frac{1}{2}\sum_{i=1}^{n}\frac{\partial \log(\det(\boldsymbol{\Psi} + \mathbf{U}_i))}{\partial \text{vech}(\boldsymbol{\Psi})^{\top}} \\
&+ \frac{\partial}{\partial \text{vech}(\boldsymbol{\Psi})^{\top}}\log\left(f\left(\sum_{i=1}^{n}(\mathbf{x}_i - \boldsymbol{\mu})^{\top}(\boldsymbol{\Psi} + \mathbf{U}_i)^{-1}(\mathbf{x}_i - \boldsymbol{\mu})\right)\right) \\
&= -\frac{1}{2}\sum_{i=1}^{n}\text{vech}((\boldsymbol{\Psi} + \mathbf{U}_i)^{-1})^{\top}\mathbf{G}_p^{\top}\mathbf{G}_p + \frac{f'\left(\sum_{i=1}^{n}(\mathbf{x}_i - \boldsymbol{\mu})^{\top}(\boldsymbol{\Psi} + \mathbf{U}_i)^{-1}(\mathbf{x}_i - \boldsymbol{\mu})\right)}{f\left(\sum_{i=1}^{n}(\mathbf{x}_i - \boldsymbol{\mu})^{\top}(\boldsymbol{\Psi} + \mathbf{U}_i)^{-1}(\mathbf{x}_i - \boldsymbol{\mu})\right)} \\
&\times \sum_{i=1}^{n}\left((\mathbf{x}_i - \boldsymbol{\mu})^{\top} \otimes (\mathbf{x}_i - \boldsymbol{\mu})^{\top}\right)\frac{\partial \text{vec}(\boldsymbol{\Psi} + \mathbf{U}_i)^{-1}}{\partial \text{vech}(\boldsymbol{\Psi})^{\top}} \\
&= -\frac{1}{2}\sum_{i=1}^{n}\text{vec}((\boldsymbol{\Psi} + \mathbf{U}_i)^{-1})^{\top}\mathbf{G}_p - \frac{f'\left(\sum_{i=1}^{n}(\mathbf{x}_i - \boldsymbol{\mu})^{\top}(\boldsymbol{\Psi} + \mathbf{U}_i)^{-1}(\mathbf{x}_i - \boldsymbol{\mu})\right)}{f\left(\sum_{i=1}^{n}(\mathbf{x}_i - \boldsymbol{\mu})^{\top}(\boldsymbol{\Psi} + \mathbf{U}_i)^{-1}(\mathbf{x}_i - \boldsymbol{\mu})\right)} \\
&\times \sum_{i=1}^{n}\left((\mathbf{x}_i - \boldsymbol{\mu})^{\top} \otimes (\mathbf{x}_i - \boldsymbol{\mu})^{\top}\right)\left((\boldsymbol{\Psi} + \mathbf{U}_i)^{-1} \otimes (\boldsymbol{\Psi} + \mathbf{U}_i)^{-1}\right)\mathbf{G}_p,
\end{aligned}
$$

The first block of the Fisher information matrix is given by

$$
\mathbf{F}_{11} = \mathbb{E}\left[\left(\frac{\partial L(\boldsymbol{\mu}, \boldsymbol{\Psi}; \mathbf{X})}{\partial \boldsymbol{\mu}^{\top}}\right)^{\top}\frac{\partial L(\boldsymbol{\mu}, \boldsymbol{\Psi}; \mathbf{X})}{\partial \boldsymbol{\mu}^{\top}}\right] = 4\sum_{i=1}^{n}\sum_{j=1}^{n}(\boldsymbol{\Psi} + \mathbf{U}_i)^{-1}\mathbf{H}_{ij}(\boldsymbol{\Psi} + \mathbf{U}_j)^{-1} \tag{41}
$$

with

$$
\begin{aligned}
\mathbf{H}_{ij} &= \left(\prod_{k=1}^{n}\sqrt{\det(\boldsymbol{\Psi} + \mathbf{U}_k)}^{-1/2}\right)\int_{\mathbb{R}^{np}}(\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_j - \boldsymbol{\mu})^{\top} \\
&\times \left(\frac{f'\left(\sum_{k=1}^{n}(\mathbf{x}_k - \boldsymbol{\mu})^{\top}(\boldsymbol{\Psi} + \mathbf{U}_k)^{-1}(\mathbf{x}_k - \boldsymbol{\mu})\right)}{f\left(\sum_{k=1}^{n}(\mathbf{x}_k - \boldsymbol{\mu})^{\top}(\boldsymbol{\Psi} + \mathbf{U}_k)^{-1}(\mathbf{x}_k - \boldsymbol{\mu})\right)}\right)^2 f\left(\sum_{k=1}^{n}(\mathbf{x}_k - \boldsymbol{\mu})^{\top}(\boldsymbol{\Psi} + \mathbf{U}_k)^{-1}(\mathbf{x}_k - \boldsymbol{\mu})\right)\mathbf{dX} \\
&= \int_{\mathbb{R}^{np}}(\boldsymbol{\Psi} + \mathbf{U}_i)^{1/2}\mathbf{z}_i\mathbf{z}_j^{\top}(\boldsymbol{\Psi} + \mathbf{U}_j)^{1/2}\left(\frac{f'\left(\sum_{k=1}^{n}\mathbf{z}_k^{\top}\mathbf{z}_k\right)}{f\left(\sum_{k=1}^{n}\mathbf{z}_k^{\top}\mathbf{z}_k\right)}\right)^2 f\left(\sum_{k=1}^{n}\mathbf{z}_k^{\top}\mathbf{z}_k\right)\mathbf{dZ},
\end{aligned}
$$

where the last equality follows from the transformation $\mathbf{x}_k = \boldsymbol{\mu} + (\boldsymbol{\Psi} + \mathbf{U}_k)^{1/2}\mathbf{z}_k$ for $k = 1, ..., n$ with the Jacobian equal to $\left(\prod_{k=1}^{n}\sqrt{\det(\boldsymbol{\Psi} + \mathbf{U}_k)}^{1/2}\right)$ and $\mathbf{Z} = (\mathbf{z}_1, \ldots, \mathbf{z}_n)$. Moreover, we get that

$$
\text{vec}(\mathbf{Z}) = (\boldsymbol{\Psi} \otimes \mathbf{I} + \mathbf{U})^{-1/2}\text{vec}(\mathbf{X} - \boldsymbol{\mu}\mathbf{1})
$$

and, consequently, $\mathbf{Z} \sim E_{p,n}(\mathbf{O}_{p,n}, \mathbf{I}_{p\times n}, f)$ with $\mathbf{O}_{p,n}$ $p \times n$ zero matrix (see, Theorem 2.13 in Gupta et al. (2013)).

Since

$$\mathbf{z}_j^\top (\boldsymbol{\Psi} + \mathbf{U}_j)^{1/2} \left( \frac{f'\left(\sum_{k=1}^n \mathbf{z}_k^\top \mathbf{z}_k\right)}{f\left(\sum_{k=1}^n \mathbf{z}_k^\top \mathbf{z}_k\right)} \right)^2 f\left(\sum_{k=1}^n \mathbf{z}_k^\top \mathbf{z}_k\right)$$

is an odd function on a symmetric region, we get that $\mathbf{H}_{ij} = \mathbf{O}_{p,p}$. Furthermore, using that $\mathbf{z}_i / \sqrt{\mathrm{vec}(\mathbf{Z})^\top \mathrm{vec}(\mathbf{Z})}$ and $\mathrm{vec}(\mathbf{Z})^\top \mathrm{vec}(\mathbf{Z})$ with $\mathrm{vec}(\mathbf{Z})^\top \mathrm{vec}(\mathbf{Z}) = \sum_{k=1}^n \mathbf{z}_k^\top \mathbf{z}_k$ are independent (see, Gupta et al. (2013, Theorem 2.15)), we obtain

$$\mathbf{H}_{ii} \;=\; J_1 (\boldsymbol{\Psi} + \mathbf{U}_i)^{1/2} \mathbb{E}\left( \frac{\mathbf{z}_i \mathbf{z}_i^\top}{\mathrm{vec}(\mathbf{Z})^\top \mathrm{vec}(\mathbf{Z})} \right) (\boldsymbol{\Psi} + \mathbf{U}_j)^{1/2}, \tag{42}$$

where

$$J_1 \;=\; \mathbb{E}\left( \mathrm{vec}(\mathbf{Z})^\top \mathrm{vec}(\mathbf{Z}) \left( \frac{f'\left( \mathrm{vec}(\mathbf{Z})^\top \mathrm{vec}(\mathbf{Z}) \right)}{f\left( \mathrm{vec}(\mathbf{Z})^\top \mathrm{vec}(\mathbf{Z}) \right)} \right)^2 \right). \tag{43}$$

To this end, we note that the distribution of $\mathbf{z}_i / \sqrt{\mathrm{vec}(\mathbf{Z})^\top \mathrm{vec}(\mathbf{Z})}$ does not depend on the type of elliptical distribution, i.e., on density generator $f(.)$, and it is the same as in case of the normal distribution. Let $\mathbf{Z}_N \sim \mathcal{N}_{p,n}(\mathbf{O}_{p,n}, \mathbf{I}_{p \times n})$ (standard matrix-variate normal distribution) and let $\mathbf{z}_{i,N}$ denote its $i$th column. Then, it holds that

$$\mathbb{E}\left( \frac{\mathbf{z}_i \mathbf{z}_i^\top}{\mathrm{vec}(\mathbf{Z})^\top \mathrm{vec}(\mathbf{Z})} \right) = \mathbb{E}\left( \frac{\mathbf{z}_{i,N} \mathbf{z}_{i,N}^\top}{\mathrm{vec}(\mathbf{Z}_N)^\top \mathrm{vec}(\mathbf{Z}_N)} \right) \frac{\mathbb{E}(\mathrm{vec}(\mathbf{Z}_N)^\top \mathrm{vec}(\mathbf{Z}_N))}{\mathbb{E}(\mathrm{vec}(\mathbf{Z}_N)^\top \mathrm{vec}(\mathbf{Z}_N))}$$

$$=\; \frac{\mathbb{E}\left( \mathbf{z}_{i,N} \mathbf{z}_{i,N}^\top \right)}{\mathbb{E}(\mathrm{vec}(\mathbf{Z}_N)^\top \mathrm{vec}(\mathbf{Z}_N))} = \frac{1}{pn} \mathbf{I}_p, \tag{44}$$

where we used that $\mathrm{vec}(\mathbf{Z}_N)^\top \mathrm{vec}(\mathbf{Z}_N) \sim \chi^2_{pn}$ ($\chi^2$-distribution with $pn$ degrees of freedom).

Summarizing (41), (42), and (43), we get

$$\mathbf{F}_{11} = \frac{4J_1}{pn} \sum_{i=1}^n (\boldsymbol{\Psi} + \mathbf{U}_i)^{-1}.$$

Next, we compute the nondiagonal block $\mathbf{F}_{21}$ of the Fisher information matrix. The transformation $\mathbf{x}_k = \boldsymbol{\mu} + (\boldsymbol{\Psi} + \mathbf{U}_k)^{1/2} \mathbf{z}_k$ for $k = 1, ..., n$ yields

$$\mathbf{F}_{21} \;=\; \mathbb{E}\left[ \left( \frac{\partial L(\boldsymbol{\mu}, \boldsymbol{\Psi}; \mathbf{X})}{\partial \boldsymbol{\mu}^\top} \right)^\top \frac{\partial L(\boldsymbol{\mu}, \boldsymbol{\Psi}; \mathbf{X})}{\partial \mathrm{vech}(\boldsymbol{\Psi})^\top} \right] = \sum_{i=1}^n \sum_{j=1}^n (\boldsymbol{\Psi} + \mathbf{U}_i)^{-1/2} \left( \int_{\mathbb{R}^{pn}} \mathbf{M}_{ij}(\mathbf{Z}) d\mathbf{Z} \right) \mathbf{G}_p,$$

where

$$\mathbf{M}_{ij}(\mathbf{Z}) \;=\; \frac{f'\left( \sum_{k=1}^n \mathbf{z}_k^\top \mathbf{z}_k \right)}{f\left( \sum_{k=1}^n \mathbf{z}_k^\top \mathbf{z}_k \right)} \mathbf{z}_i \left[ \mathrm{vec}((\boldsymbol{\Psi} + \mathbf{U}_j)^{-1})^\top \right.$$

$$+\; 2 \frac{f'\left( \sum_{k=1}^n \mathbf{z}_k^\top \mathbf{z}_k \right)}{f\left( \sum_{k=1}^n \mathbf{z}_k^\top \mathbf{z}_k \right)} \left( \mathbf{z}_j^\top \otimes \mathbf{z}_j^\top \right) \left( (\boldsymbol{\Psi} + \mathbf{U}_j)^{-1/2} \otimes (\boldsymbol{\Psi} + \mathbf{U}_j)^{-1/2} \right) \Bigg]$$

which is an odd function on a symmetric region. Hence,

$$\int_{\mathbb{R}^{pn}} \mathbf{M}_{ij}(\mathbf{Z}) d\mathbf{Z} = \mathbf{O} \quad \text{for all} \quad i, j \in \{1, ..., n\},$$

and, consequently, $\mathbf{F}_{21} = \mathbf{O}$.

Similarly, using the transformation $\mathbf{x}_k = \boldsymbol{\mu} + (\boldsymbol{\Psi} + \mathbf{U}_k)^{1/2}\mathbf{z}_k$ for $k = 1, ..., n$, we get

$$
\begin{aligned}
\mathbf{F}_{22} &= \mathbb{E}\left[\left(\frac{\partial L(\boldsymbol{\mu}, \boldsymbol{\Psi}; \mathbf{X})}{\partial \mathrm{vech}(\boldsymbol{\Psi})^\top}\right)^\top \frac{\partial L(\boldsymbol{\mu}, \boldsymbol{\Psi}; \mathbf{X})}{\partial \mathrm{vech}(\boldsymbol{\Psi})^\top}\right] \\
&= \mathbb{E}\Bigg\{\mathbf{G}_p^\top\Bigg[\sum_{i=1}^n \sum_{j=1}^n \frac{1}{4}\mathrm{vec}((\boldsymbol{\Psi} + \mathbf{U}_i)^{-1})\mathrm{vec}((\boldsymbol{\Psi} + \mathbf{U}_j)^{-1})^\top \\
&\quad + \frac{1}{2}\frac{f'\left(\sum_{k=1}^n \mathbf{z}_k^\top \mathbf{z}_k\right)}{f\left(\sum_{k=1}^n \mathbf{z}_k^\top \mathbf{z}_k\right)}\left((\boldsymbol{\Psi} + \mathbf{U}_i)^{-1/2} \otimes (\boldsymbol{\Psi} + \mathbf{U}_i)^{-1/2}\right)(\mathbf{z}_i \otimes \mathbf{z}_i)\mathrm{vec}((\boldsymbol{\Psi} + \mathbf{U}_j)^{-1})^\top \\
&\quad + \frac{1}{2}\frac{f'\left(\sum_{k=1}^n \mathbf{z}_k^\top \mathbf{z}_k\right)}{f\left(\sum_{k=1}^n \mathbf{z}_k^\top \mathbf{z}_k\right)}\mathrm{vec}((\boldsymbol{\Psi} + \mathbf{U}_i)^{-1})\left(\mathbf{z}_j^\top \otimes \mathbf{z}_j^\top\right)\left((\boldsymbol{\Psi} + \mathbf{U}_j)^{-1/2} \otimes (\boldsymbol{\Psi} + \mathbf{U}_j)^{-1/2}\right) \\
&\quad + \left(\frac{f'\left(\sum_{k=1}^n \mathbf{z}_k^\top \mathbf{z}_k\right)}{f\left(\sum_{k=1}^n \mathbf{z}_k^\top \mathbf{z}_k\right)}\right)^2\left((\boldsymbol{\Psi} + \mathbf{U}_i)^{-1/2} \otimes (\boldsymbol{\Psi} + \mathbf{U}_i)^{-1/2}\right) \\
&\quad \times (\mathbf{z}_i \otimes \mathbf{z}_i)\left(\mathbf{z}_j^\top \otimes \mathbf{z}_j^\top\right)\left((\boldsymbol{\Psi} + \mathbf{U}_j)^{-1/2} \otimes (\boldsymbol{\Psi} + \mathbf{U}_j)^{-1/2}\right)\mathbf{G}_p\Bigg]\Bigg\} \\
&= \mathbf{G}_p^\top\Bigg[\sum_{i=1}^n \sum_{j=1}^n \frac{1}{4}\mathrm{vec}((\boldsymbol{\Psi} + \mathbf{U}_i)^{-1})\mathrm{vec}((\boldsymbol{\Psi} + \mathbf{U}_j)^{-1})^\top \\
&\quad + \frac{1}{2}\left((\boldsymbol{\Psi} + \mathbf{U}_i)^{-1/2} \otimes (\boldsymbol{\Psi} + \mathbf{U}_i)^{-1/2}\right)\mathbf{D}_i\mathrm{vec}((\boldsymbol{\Psi} + \mathbf{U}_j)^{-1})^\top \\
&\quad + \frac{1}{2}\mathrm{vec}((\boldsymbol{\Psi} + \mathbf{U}_i)^{-1})\mathbf{D}_j^\top\left((\boldsymbol{\Psi} + \mathbf{U}_j)^{-1/2} \otimes (\boldsymbol{\Psi} + \mathbf{U}_j)^{-1/2}\right) \\
&\quad + \left((\boldsymbol{\Psi} + \mathbf{U}_i)^{-1/2} \otimes (\boldsymbol{\Psi} + \mathbf{U}_i)^{-1/2}\right)\mathbf{D}_{ij}\left((\boldsymbol{\Psi} + \mathbf{U}_j)^{-1/2} \otimes (\boldsymbol{\Psi} + \mathbf{U}_j)^{-1/2}\right)\mathbf{G}_p\Bigg]
\end{aligned}
$$

with

$$
\mathbf{D}_i = \mathbb{E}\left(\frac{f'\left(\mathrm{vec}(\mathbf{Z})^\top \mathrm{vec}(\mathbf{Z})\right)}{f\left(\mathrm{vec}(\mathbf{Z})^\top \mathrm{vec}(\mathbf{Z})\right)}(\mathbf{z}_i \otimes \mathbf{z}_i)\right) = J\mathbb{E}\left(\frac{(\mathbf{z}_i \otimes \mathbf{z}_i)}{\mathrm{vec}(\mathbf{Z})^\top \mathrm{vec}(\mathbf{Z})}\right) \tag{45}
$$

and

$$
\mathbf{D}_{ij} = \mathbb{E}\left(\left(\frac{f'\left(\mathrm{vec}(\mathbf{Z})^\top \mathrm{vec}(\mathbf{Z})\right)}{f\left(\mathrm{vec}(\mathbf{Z})^\top \mathrm{vec}(\mathbf{Z})\right)}\right)^2 (\mathbf{z}_i \otimes \mathbf{z}_i)\left(\mathbf{z}_j^\top \otimes \mathbf{z}_j^\top\right)\right) = J_2\mathbb{E}\left(\frac{(\mathbf{z}_i \otimes \mathbf{z}_i)\left(\mathbf{z}_j^\top \otimes \mathbf{z}_j^\top\right)}{(\mathrm{vec}(\mathbf{Z})^\top \mathrm{vec}(\mathbf{Z}))^2}\right) \tag{46}
$$

where we use that $\mathbf{Z}/\sqrt{\mathrm{vec}(\mathbf{Z})^\top \mathrm{vec}(\mathbf{Z})}$ and $\mathrm{vec}(\mathbf{Z})^\top \mathrm{vec}(\mathbf{Z})$ are independent (see, Gupta et al. (2013, Theorem 2.15)), $\sum_{k=1}^n \mathbf{z}_k^\top \mathbf{z}_k = \mathrm{vec}(\mathbf{Z})^\top \mathrm{vec}(\mathbf{Z})$, and define

$$
J_2 = \mathbb{E}\left((\mathrm{vec}(\mathbf{Z})^\top \mathrm{vec}(\mathbf{Z}))^2 \left(\frac{f'\left(\mathrm{vec}(\mathbf{Z})^\top \mathrm{vec}(\mathbf{Z})\right)}{f\left(\mathrm{vec}(\mathbf{Z})^\top \mathrm{vec}(\mathbf{Z})\right)}\right)^2\right). \tag{47}
$$

and

$$
J = \mathbb{E}\left(\mathrm{vec}(\mathbf{Z})^\top \mathrm{vec}(\mathbf{Z})\frac{f'\left(\mathrm{vec}(\mathbf{Z})^\top \mathrm{vec}(\mathbf{Z})\right)}{f\left(\mathrm{vec}(\mathbf{Z})^\top \mathrm{vec}(\mathbf{Z})\right)}\right).
$$

Let $R^2 = \mathrm{vec}(\mathbf{Z})^\top \mathrm{vec}(\mathbf{Z})$. Then the density of $R^2$ is given by $f_{R^2}(r) = r^{pn/2-1}f(r)$ (cf., Theorem

2.16 in Gupta et al. (2013)) and

$$
\begin{aligned}
J &= E\left(R^2 \frac{f'(R^2)}{f(R^2)}\right) = \int_0^\infty r \frac{f'(r)}{f(r)} r^{pn/2-1} f(r) \mathrm{d}r = \int_0^\infty r^{pn/2} f'(r) \mathrm{d}r \\
&= \left. r^{pn/2} f(r) \right|_0^\infty - \frac{np}{2} \int_0^\infty r^{pn/2-1} f(r) \mathrm{d}r = -\frac{pn}{2},
\end{aligned}
$$

where we use that $\int_0^\infty r^{n/2-1} f(r) \mathrm{d}r = 1$ because $r^{n/2-1} f(r)$ is the density of $R^2$ which also implies that $r^{n/2-1} f(r) = o(r^{-1})$ as $r \to \infty$, i.e. $r^{n/2} f(r) = o(1)$ as $r \to \infty$.

Moreover, since the distribution of $\mathbf{z}_i / \sqrt{\mathrm{vec}(\mathbf{Z})^\top \mathrm{vec}(\mathbf{Z})}$ does not depend on the type of elliptical distribution, we get with $\mathbf{Z}_N \sim \mathcal{N}_{p,n}(\mathbf{O}_{p,n}, \mathbf{I}_{p \times n})$ that

$$
\begin{aligned}
\mathbb{E}\left(\frac{(\mathbf{z}_i \otimes \mathbf{z}_i)}{\mathrm{vec}(\mathbf{Z})^\top \mathrm{vec}(\mathbf{Z})}\right) &= \mathbb{E}\left(\frac{(\mathbf{z}_{i,N} \otimes \mathbf{z}_{i,N})}{\mathrm{vec}(\mathbf{Z}_N)^\top \mathrm{vec}(\mathbf{Z}_N)}\right) \\
&= \mathbb{E}\left(\frac{(\mathbf{z}_{i,N} \otimes \mathbf{z}_{i,N})}{\mathrm{vec}(\mathbf{Z}_N)^\top \mathrm{vec}(\mathbf{Z}_N)}\right) \frac{\mathbb{E}\left(\mathrm{vec}(\mathbf{Z}_N)^\top \mathrm{vec}(\mathbf{Z}_N)\right)}{\mathbb{E}\left(\mathrm{vec}(\mathbf{Z}_N)^\top \mathrm{vec}(\mathbf{Z}_N)\right)} = \mathbb{E}\left(\mathbf{z}_{i,N} \otimes \mathbf{z}_{i,N}\right) \frac{1}{pn} = \frac{1}{pn} \mathrm{vec}(\mathbf{I}_p),
\end{aligned}
$$

where the last equality follows from the results in Ghazal and Neudecker (2000, p. 67)

Similarly for $i \neq j \in \{1, ..., n\}$ we get

$$
\begin{aligned}
&\mathbb{E}\left(\frac{(\mathbf{z}_i \otimes \mathbf{z}_i)\left(\mathbf{z}_j^\top \otimes \mathbf{z}_j^\top\right)}{(\mathrm{vec}(\mathbf{Z})^\top \mathrm{vec}(\mathbf{Z}))^2}\right) \\
&= \mathbb{E}\left((\mathbf{z}_{i,N} \otimes \mathbf{z}_{i,N})\left(\mathbf{z}_{j,N}^\top \otimes \mathbf{z}_{j,N}^\top\right)\right) \frac{1}{\mathbb{E}\left((\mathrm{vec}(\mathbf{Z}_N)^\top \mathrm{vec}(\mathbf{Z}_N))^2\right)} \\
&= \mathbb{E}\left(\mathbf{z}_{i,N} \otimes \mathbf{z}_{i,N}\right) \mathbb{E}\left(\mathbf{z}_{j,N}^\top \otimes \mathbf{z}_{j,N}^\top\right) \frac{1}{2pn + p^2 n^2} \\
&= \frac{1}{2pn + p^2 n^2} \mathrm{vec}(\mathbf{I}_p) \mathrm{vec}(\mathbf{I}_p)^\top.
\end{aligned}
$$

Finally, for $i = j \in \{1, ..., n\}$ we obtain

$$
\begin{aligned}
&\mathbb{E}\left(\frac{(\mathbf{z}_i \otimes \mathbf{z}_i)\left(\mathbf{z}_i^\top \otimes \mathbf{z}_k^\top\right)}{(\mathrm{vec}(\mathbf{Z})^\top \mathrm{vec}(\mathbf{Z}))^2}\right) \\
&= \mathbb{E}\left((\mathbf{z}_{i,N} \otimes \mathbf{z}_{i,N})\left(\mathbf{z}_{i,N}^\top \otimes \mathbf{z}_{i,N}^\top\right)\right) \frac{1}{\mathbb{E}\left((\mathrm{vec}(\mathbf{Z}_N)^\top \mathrm{vec}(\mathbf{Z}_N))^2\right)} \\
&= (\mathbf{I}_{p^2} + \mathbf{K}_p + \mathrm{vec}(\mathbf{I}_p) \mathrm{vec}(\mathbf{I}_p)^\top) \frac{1}{2pn + p^2 n^2},
\end{aligned}
$$

where the last equality follows from Theorem 4.1 in Magnus et al. (1979) and $\mathbf{K}_p$ is the commutation matrix.

The properties of the *vec* operator and the properties of the duplication and commutation matrices, namely (see, Magnus and Neudecker (2019, Theorem 3.12 and Theorem 3.13))

$$
\frac{1}{2}((\mathbf{I}_{p^2} + \mathbf{K}_p)(\mathbf{A} \otimes \mathbf{A})\mathbf{G}_p = (\mathbf{A} \otimes \mathbf{A})\mathbf{G}_p
$$

yield

$$
\begin{aligned}
&\left((\boldsymbol{\Psi} + \mathbf{U}_i)^{-1/2} \otimes (\boldsymbol{\Psi} + \mathbf{U}_i)^{-1/2}\right) \mathrm{vec}(\mathbf{I}_p) \\
&= \mathrm{vec}\left((\boldsymbol{\Psi} + \mathbf{U}_i)^{-1/2}(\boldsymbol{\Psi} + \mathbf{U}_i)^{-1/2}\right) = \mathrm{vec}\left((\boldsymbol{\Psi} + \mathbf{U}_i)^{-1}\right)
\end{aligned}
$$

and

$$\left((\boldsymbol{\Psi} + \mathbf{U}_i)^{-1/2} \otimes (\boldsymbol{\Psi} + \mathbf{U}_i)^{-1/2}\right) \left(\mathbf{I}_{p^2} + K_p\right) \left((\boldsymbol{\Psi} + \mathbf{U}_i)^{-1/2} \otimes (\boldsymbol{\Psi} + \mathbf{U}_i)^{-1/2}\right) \mathbf{G}_p$$
$$= 2 \left((\boldsymbol{\Psi} + \mathbf{U}_i)^{-1} \otimes (\boldsymbol{\Psi} + \mathbf{U}_i)^{-1}\right) \mathbf{G}_p.$$

Hence,

$$\mathbf{F}_{22} = \mathbf{G}_p^\top \Bigg[ \left(\frac{J_2}{2pn + p^2 n^2} - \frac{1}{4}\right) \mathrm{vec}\left(\sum_{i=1}^n (\boldsymbol{\Psi} + \mathbf{U}_i)^{-1}\right) \mathrm{vec}\left(\sum_{j=1}^n (\boldsymbol{\Psi} + \mathbf{U}_j)^{-1}\right)^\top$$
$$+ \frac{2J_2}{2pn + p^2 n^2} \sum_{i=1}^n \left((\boldsymbol{\Psi} + \mathbf{U}_i)^{-1} \otimes (\boldsymbol{\Psi} + \mathbf{U}_i)^{-1}\right) \Bigg] \mathbf{G}_p,$$

which complete the proof of the theorem. $\square$

*Proof of Lemma 1:* First, we show that $(\mathbf{A} + \mathbf{B}) \otimes (\mathbf{A} + \mathbf{B}) - \mathbf{A} \otimes \mathbf{A} \geq \mathbf{0}$. For any vector $\mathbf{c} = \mathrm{vec}(\mathbf{C})$ we get

$$\mathbf{c}^\top \left((\mathbf{A} + \mathbf{B}) \otimes (\mathbf{A} + \mathbf{B}) - \mathbf{A} \otimes \mathbf{A}\right) \mathbf{c}$$
$$= \mathrm{vec}(\mathbf{C})^\top \left((\mathbf{A} + \mathbf{B}) \otimes (\mathbf{A} + \mathbf{B}) - \mathbf{A} \otimes \mathbf{A}\right) \mathrm{vec}(\mathbf{C})$$
$$= \mathrm{tr}(\mathbf{C}^\top (\mathbf{A} + \mathbf{B}) \mathbf{C} (\mathbf{A} + \mathbf{B})) - tr(\mathbf{C}^\top \mathbf{A} \mathbf{C} \mathbf{A})$$
$$= \mathrm{tr}(\mathbf{C}^\top (\mathbf{A} + \mathbf{B}) \mathbf{C} (\mathbf{A} + \mathbf{B})) - tr(\mathbf{C}^\top (\mathbf{A} + \mathbf{B}) \mathbf{C} \mathbf{A}) + tr(\mathbf{C}^\top (\mathbf{A} + \mathbf{B}) \mathbf{C} \mathbf{A}) tr(\mathbf{C}^\top \mathbf{A} \mathbf{C} \mathbf{A})$$
$$= \mathrm{tr}(\mathbf{C}^\top (\mathbf{A} + \mathbf{B}) \mathbf{C} \mathbf{B}) + tr(\mathbf{C}^\top \mathbf{B} \mathbf{C} \mathbf{A})$$
$$= \mathrm{tr}((\mathbf{A} + \mathbf{B})^{1/2} \mathbf{C} \mathbf{B} \mathbf{C}^\top (\mathbf{A} + \mathbf{B})^{1/2}) + tr(\mathbf{A}^{1/2} \mathbf{C}^\top \mathbf{B} \mathbf{C} \mathbf{A}^{1/2}) \geq 0.$$

Since $\mathbf{A}$ and $\mathbf{B}$ are symmetric, we get that $(\mathbf{A} + \mathbf{B}) \otimes (\mathbf{A} + \mathbf{B})$ and $\mathbf{A} \otimes \mathbf{A}$ are symmetric. Hence, the application of Theorem 18.3.4 in Harville (1997) leads to the second statement of the lemma.

$\square$

# References

Ades, A. E., Lu, G., and Higgins, J. (2005). The interpretation of random-effects meta-analysis in decision models. *Medical Decision Making*, 25(6):646–654.

Baker, R. and Jackson, D. (2008). A new approach to outliers in meta-analysis. *Health care management science*, 11(2):121–131.

Berger, J. and Bernardo, J. M. (1992). On the development of reference priors. In Bernardo, J. M., Berger, J., Dawid, A. P., and Smith, A. F. M., editors, *Bayesian Statistics*, volume 4, pages 35–60. Oxford: University Press.

Berger, J., Bernardo, J. M., and Sun, D. (2009). The formal definition of reference priors. *The Annals of Statistics*, 37(2):905–938.

Bernardo, J. M. and Smith, A. F. M. (2000). *Bayesian theory*. Chichester: John Wiley.

Bodnar, O. (2019). Non-informative Bayesian inference for heterogeneity in a generalized marginal random effects meta-analysis. *Theory of Probability and Mathematical Statistics*, 100:7–23.

Bodnar, O. and Elster, C. (2014a). Analytical derivation of the reference prior by sequential maximization of Shannon's mutual information in the multi-group parameter case. *Journal of Statistical Planning and Inference*, 147:106–116.

Bodnar, O. and Elster, C. (2014b). On the adjustment of inconsistent data using the Birge ratio. *Metrologia*, 51(5):516.

Bodnar, O., Link, A., Arendacká, B., Possolo, A., and Elster, C. (2017). Bayesian estimation in random effects meta-analysis using a non-informative prior. *Statistics in Medicine*, 36(2):378–399.

Bodnar, O., Link, A., and Elster, C. (2016). Objective Bayesian inference for a generalized marginal random effects model. *Bayesian Analysis*, 11(1):25–45.

Brockwell, S. E. and Gordon, I. R. (2001). A comparison of statistical methods for meta-analysis. *Statistics in Medicine*, 20(6):825–840.

Chen, H., Manning, A. K., and Dupuis, J. (2012). A method of moments estimator for random effect multivariate meta-analysis. *Biometrics*, 68(4):1278–1284.

Cornell, J. E., Mulrow, C. D., Localio, R., Stack, C. B., Meibohm, A. R., Guallar, E., and Goodman, S. N. (2014). Random-effects meta-analysis of inconsistent effects: A time for change. *Annals of Internal Medicine*, 160(4):267–270.

Davey, J., Turner, R. M., Clarke, M. J., and Higgins, J. (2011). Characteristics of meta-analyses and their component studies in the Cochrane Database of Systematic Reviews: A cross-sectional, descriptive analysis. *BMC Medical Research Methodology*, 11(1):160.

DerSimonian, R. and Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials*, 7(3):177–188.

Gasparrini, A., Armstrong, B., and Kenward, M. (2012). Multivariate meta-analysis for non-linear and other multi-parameter associations. *Statistics in Medicine*, 31(29):3821–3839.

Ghazal, G. A. and Neudecker, H. (2000). On second-order and fourth-order moments of jointly distributed random matrices: A survey. *Linear Algebra and its Applications*, 321(1-3):61–93.

Givens, G. H. and Hoeting, J. A. (2012). *Computational Statistics*, volume 710. John Wiley & Sons.

Guolo, A. (2012). Higher-order likelihood inference in meta-analysis and meta-regression. *Statistics in Medicine*, 31(4):313–327.

Gupta, A. K. and Nagar, D. K. (2000). *Matrix Variate Distributions*. Chapman and Hall/CRC.

Gupta, A. K., Varga, T., and Bodnar, T. (2013). *Elliptically Contoured Models in Statistics and Portfolio Theory*. Springer, New York.

Harville, D. A. (1997). *Matrix Algebra from Statistician's Perspective*. Springer, New York.

Held, L. and Bové, D. S. (2014). *Applied Statistical Inference: Likelihood and Bayes*. Springer Science & Business Media.

Hill, S. D. and Spall, J. C. (2019). Stationarity and convergence of the metropolis-hastings algorithm: Insights into theoretical aspects. *IEEE Control Systems Magazine*, 39(1):56–67.

Jackson, D. and Riley, R. D. (2014). A refined method for multivariate meta-analysis and meta-regression. *Statistics in Medicine*, 33(4):541–554.

Jackson, D. and White, I. R. (2018). When should meta-analysis avoid making hidden normality assumptions? *Biometrical Journal*, 60(6):1040–1058.

Jackson, D., White, I. R., and Riley, R. D. (2013). A matrix-based method of moments for fitting the multivariate random effects model for meta-analysis and meta-regression. *Biometrical Journal*, 55(2):231–245.

Jackson, D., White, I. R., and Riley, R. D. (2020). Multivariate meta-analysis. In Schmid, C. H., Stijnen, T., and White, I. R., editors, *Handbook of Meta-Analysis*, pages 163–186. CRC Press.

Jackson, D., White, I. R., and Thompson, S. G. (2010). Extending DerSimonian and Laird's methodology to perform multivariate random effects meta-analyses. *Statistics in Medicine*, 29(12):1282–1297.

Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society A*, 186:453–461.

Lambert, P. C., Sutton, A. J., Burton, P. R., Abrams, K. R., and Jones, D. R. (2005). How vague is vague? A simulation study of the impact of the use of vague prior distributions in mcmc using winbugs. *Statistics in Medicine*, 24(15):2401–2428.

Langan, D., Higgins, J. P., and Simmonds, M. (2017). Comparative performance of heterogeneity variance estimators in meta-analysis: A review of simulation studies. *Research Synthesis Methods*, 8(2):181–198.

Laplace, P. S. (1812). *Théorie Analitique des Probabilités*. Paris: Courcier.

Lee, K. J. and Thompson, S. G. (2008). Flexible parametric models for random-effects distributions. *Statistics*

*in medicine*, 27(3):418–434.

Liu, D., Liu, R. Y., and Xie, M. (2015). Multivariate meta-analysis of heterogeneous studies using only summary statistics: efficiency and robustness. *Journal of the American Statistical Association*, 110(509):326–340.

Lütkepohl, H. (1996). *Handbook of Matrices*. Wiley Chichester.

Ma, X., Lin, L., Qu, Z., Zhu, M., and Chu, H. (2018). Performance of between-study heterogeneity measures in the Cochrane library. *Epidemiology*, 29(6):821–824.

Magnus, J. R. and Neudecker, H. (2019). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley & Sons.

Magnus, J. R., Neudecker, H., et al. (1979). The commutation matrix: some properties and applications. *The Annals of Statistics*, 7(2):381–394.

Michael, H., Thornton, S., Xie, M., and Tian, L. (2019). Exact inference on the random-effects model for meta-analyses with few studies. *Biometrics*, 75(2):485–493.

Nam, I.-S., Mengersen, K., and Garthwaite, P. (2003). Multivariate meta-analysis. *Statistics in Medicine*, 22(14):2309–2333.

Negeri, Z. F. and Beyene, J. (2020). Robust bivariate random-effects model for accommodating outlying and influential studies in meta-analysis of diagnostic test accuracy studies. *Statistical Methods in Medical Research*, 29(11):3308–3325.

Noma, H., Maruo, K., Gosho, M., Levine, S. Z., Goldberg, Y., Leucht, S., and Furukawa, T. A. (2019). Efficient two-step multivariate random effects meta-analysis of individual participant data for longitudinal clinical trials using mixed effects models. *BMC Medical Research Methodology*, 19(1):33.

Novianti, P. W., Roes, K. C. B., and van der Tweel, I. (2014). Estimation of between-trial variance in sequential meta-analyses: A simulation study. *Contemporary Clinical Trials*, 37(1):129–138.

Paul, M., Riebler, A., Bachmann, L., Rue, H., and Held, L. (2010). Bayesian bivariate meta-analysis of diagnostic test studies using integrated nested Laplace approximations. *Statistics in Medicine*, 29(12):1325–1339.

Paule, R. C. and Mandel, J. (1982). Consensus values and weighting factors. *Journal of Research of the National Bureau of Standards*, 87(5):377–385.

Riley, R. D., Lambert, P. C., and Abo-Zaid, G. (2010). Meta-analysis of individual participant data: Rationale, conduct, and reporting. *BMJ*, 340:c221.

Roever, C. (2016). *bayesmeta: Bayesian Random-Effects Meta-Analysis*. R package version 1.2.

Rukhin, A. L. (2013). Estimating heterogeneity variance in meta-analysis. *Journal of the Royal Statistical Society: Ser. B*, 75:451–469.

Rukhin, A. L. (2017a). Estimation of the common mean from heterogeneous normal observations with unknown variances. *Journal of the Royal Statistical Society: Ser. B*, 79(5):1601–1618.

Rukhin, A. L. (2017b). Research synthesis when some within-study uncertainties are absent. *Metrologia*, 54(6):874.

Schwarzer, G., Carpenter, J. R., and Rücker, G. (2015). *Meta-Analysis with R*. Springer.

Strawderman, W. E. and Rukhin, A. L. (2010). Simultaneous estimation and reduction of nonconformity in interlaboratory studies. *Journal of the Royal Statistical Society: Ser. B*, 72:219–234.

Sutradhar, B. C. and Ali, M. M. (1989). A generalization of the wishart distribution for the elliptical model and its moments for the multivariate t model. *Journal of Multivariate Analysis*, 29(1):155–162.

Sutton, A. J. and Higgins, J. (2008). Recent developments in meta-analysis. *Statistics in Medicine*, 27(5):625–650.

Turner, R. M., Jackson, D., Wei, Y., Thompson, S. G., and Higgins, J. (2015). Predictive distributions for between-study heterogeneity and simple methods for their application in Bayesian meta-analysis. *Statistics in Medicine*, 34(6):984–998.

Veroniki, A. A., Jackson, D., Bender, R., Kuss, O., Langan, D., Higgins, J. P., Knapp, G., and Salanti, G. (2019). Methods to calculate uncertainty in the estimated overall effect size from a random-effects meta-analysis. *Research Synthesis Methods*, to appear.

Viechtbauer, W. (2005). Bias and efficiency of meta-analytic variance estimators in the random-effects model.

*Journal of Educational and Behavioral Statistics*, 30(3):261–293.

Viechtbauer, W. (2007). Confidence intervals for the amount of heterogeneity in meta-analysis. *Statistics in Medicine*, 26(1):37–52.

Wang, C.-C. and Lee, W.-C. (2020). Evaluation of the normality assumption in meta-analyses. *American Journal of Epidemiology*, 189(3):235–242.

Wei, Y. and Higgins, J. P. (2013). Bayesian multivariate meta-analysis with multiple outcomes. *Statistics in Medicine*, 32(17):2911–2934.

Wynants, L., Riley, R., Timmerman, D., and Van Calster, B. (2018). Random-effects meta-analysis of the clinical utility of tests and prediction models. *Statistics in Medicine*, 37(12):2034–2052.