ESTIMATION UNDER THE MATRIX VARIATE ELLIPTICAL MODEL

Janet van Niekerk [1]
Department of Statistics, Faculty of Natural and Agricultural Sciences, University of Pretoria, Pretoria, South Africa
e-mail: *Janet.vanNiekerk@up.ac.za*

Andriette Bekker
Department of Statistics, Faculty of Natural and Agricultural Sciences, University of Pretoria, Pretoria, South Africa

Mohammad Arashi
Department of Statistics, School of Mathematical Sciences, University of Shahrood, Shahrood, Iran.
Department of Statistics, Faculty of Natural and Agricultural Sciences, University of Pretoria, Pretoria, South Africa

Daan J. de Waal
Department of Mathematical Statistics and Actuarial Science, Faculty of Natural and Agricultural Sciences, University of the Free State, Bloemfontein, South Africa.
Department of Statistics, Faculty of Natural and Agricultural Sciences, University of Pretoria, Pretoria, South Africa

Key words: Bayesian inference, Bessel function of matrix argument, characteristic matrix, matrix variate elliptical model, maximum posterior mode, normal-inverse Wishart, normal-Wishart, squared error loss.

Abstract: The problem of estimation within the matrix variate elliptical model is addressed. In this paper a subjective Bayesian approach is followed to derive new estimators for the parameters of the matrix variate elliptical model by assuming the previously intractable normal-Wishart prior. These new estimators are compared to the estimators derived under a normal-inverse Wishart prior as well as the objective Jeffreys' prior which results in the maximum likelihood estimators, using different measures. A valuable contribution is the development of algorithms for the simulation of the posterior distributions of the matrix variate parameters with emphasis on the new proposed estimators. A simulation study as well as Fisher's Iris data set are used to illustrate the novelty of these new estimators and to investigate the accuracy gained by assuming the normal-Wishart prior.

[1] Corresponding author.

AMS: 62F15, 62H12, 62H10, 60E05

1. Introduction

The objective of this paper is to derive estimators for the parameters of the matrix variate elliptical model from a subjective Bayesian viewpoint. Subjective analysis generally produces more admissible results compared to objective analysis, since added information is used. Although objective Bayesian analysis for the matrix variate elliptical model was considered by Fang and Li (1999), very few results and estimators for the Bayesian analysis of model (1) exist. This paper contributes to the literature by presenting a more general subjective Bayesian estimation framework for the *matrix* case in (1). Prior information will be reflected by using the normal-inverse Wishart and the normal-Wishart prior distributions respectively (see Van Niekerk, Bekker, Roux and Arashi, 2013, for the multivariate elliptical model). The squared error loss function as well as the loss function defined by Das and Dey (2010) will be used for the Bayesian inference. The matrix variate elliptical model with density function

$$f(\boldsymbol{X}) = c_{s,p}|\boldsymbol{\Sigma}|^{-\frac{s}{2}}|\boldsymbol{\Omega}|^{-\frac{p}{2}}h\left[tr(\boldsymbol{X}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})\boldsymbol{\Omega}^{-1}\right] \tag{1}$$

is the underlying data generating model. The aim is to estimate the location and scale matrices, $\boldsymbol{\mu}_{p\times s}$ and $\boldsymbol{\Sigma}_{p\times p}$, respectively where $\boldsymbol{\Omega}_{s\times s}$ is assumed to be a known hyperparameter (see Section 6 for discussion of this assumption). The considered prior distributions will thus be for $\boldsymbol{\mu}_{p\times s}$ and $\boldsymbol{\Sigma}_{p\times p}$ (see Fang and Zhang, 1990). The density function in (1) can also be written as (see Chu, 1973):

$$f(\boldsymbol{X}|\boldsymbol{\mu},\boldsymbol{\Sigma}) = \int_0^\infty w(z) f_{N_{\boldsymbol{\mu},z^{-1}\boldsymbol{\Sigma}\otimes\boldsymbol{\Omega}}}(\boldsymbol{X}|\boldsymbol{\mu},\boldsymbol{\Sigma})dz \tag{2}$$

for a scalar function $w(z)$ where $f_{N_{\boldsymbol{\mu},z^{-1}\boldsymbol{\Sigma}\otimes\boldsymbol{\Omega}}}(\cdot)$ is the matrix variate normal density function having the same expected value as $\boldsymbol{X}$ and where $z^{-1}\boldsymbol{\Sigma}\otimes\boldsymbol{\Omega}$ is a scale matrix and $\otimes$ denotes the Kronecker product. Note that $\int_0^\infty w(z)dz = 1$ and that (2) is different from the class of normal scale mixtures. For normal scale mixtures, the weighting function is actually a density function where $0 \leq w(z) \leq 1$ for all values of z. However, for the elliptical model given in (1), the weighting function does not have the same restriction but rather $-1 \leq w(z) \leq 1$, for all values of z. Note that $w(z)$ is identified from the inverse Laplace transform of the density function of the particular elliptical model. See Arashi, Iranmanesh, Norouzirad and Salarzadeh-Jenatabadi (2014) and Arashi, Saleh and Tabatabaey (2013) for more details.

In Section 2, an incomplete type II Bessel function of matrix argument is defined. This follows from representing the cumulative distribution of the largest characteristic root of $\boldsymbol{\Sigma}$ for the Wishart prior. Section 3 deals with derivations of the posterior distributions, Bayes estimators of the parameters of the matrix variate elliptical model for the normal-inverse Wishart prior. The normal-Wishart prior forms the base of Section 4. The newly developed results of Sections 3 and 4 will be applied to particular subfamilies of the matrix variate elliptical distribution in Section 5. In Section 6 algorithms for simulating these posterior distributions and calculating the proposed estimators are presented as an illustration of the theoretical results, followed by the application to the well-known Fisher's Iris dataset.

The benefit of this approach is that the Wishart distribution can now be utilized as a prior for the scale matrix opposed to the inverse-Wishart distribution as applied previously. This is demonstrated clearly in Section 6 by the comparative simulation study where new algorithms are introduced for the calculation of the proposed estimators from a simulated dataset, followed by the application to Fisher's Iris dataset.

2. An incomplete Bessel function

The type II Bessel function of Herz (1955), $B_\delta(\boldsymbol{N})$, of matrix argument is defined by

$$B_\delta(\boldsymbol{N}) = \int_{\boldsymbol{U}>\boldsymbol{0}} |\boldsymbol{U}|^{-\delta-\frac{p+1}{2}} etr(-\boldsymbol{N}\boldsymbol{U} - \boldsymbol{U}^{-1}) d\boldsymbol{U}.$$

The integral is absolutely convergent for $Re(\boldsymbol{N}) > \boldsymbol{0}$ if and only if $-Re(\delta) > \frac{p-1}{2}$. We now present an extension to this function, as follows:

Definition 1 The incomplete type II Bessel function of matrix argument, denoted by $B_\delta(\boldsymbol{N};\boldsymbol{Q})$, is defined by

$$B_\delta(\boldsymbol{N};\boldsymbol{Q}) = \int_{\boldsymbol{0}<\boldsymbol{U}<\boldsymbol{Q}} |\boldsymbol{U}|^{-\delta-\frac{p+1}{2}} etr(-\boldsymbol{N}\boldsymbol{U} - \boldsymbol{U}^{-1}) d\boldsymbol{U},$$

where $Re(\boldsymbol{N}) > \boldsymbol{0}, \boldsymbol{Q} > \boldsymbol{0}$ and $-Re(\delta) > \frac{p-1}{2}$.

Remark 1 Replacing $\boldsymbol{N}$ by $\boldsymbol{\Delta}^{\frac{1}{2}}\boldsymbol{N}\boldsymbol{\Delta}^{\frac{1}{2}}, \boldsymbol{\Delta} > \boldsymbol{0}$ and changing the matrix variate $\boldsymbol{M} = \boldsymbol{\Delta}^{\frac{1}{2}}\boldsymbol{U}\boldsymbol{\Delta}^{\frac{1}{2}}$ with Jacobian $|\boldsymbol{\Delta}|^{-\frac{p+1}{2}}$ gives the following more general form:

$$B_\delta(\boldsymbol{\Delta},\boldsymbol{N};\boldsymbol{R}) = |\boldsymbol{\Delta}|^{\delta} \int_{\boldsymbol{0}<\boldsymbol{M}<\boldsymbol{R}} |\boldsymbol{M}|^{-\delta-\frac{p+1}{2}} etr(-\boldsymbol{N}\boldsymbol{M} - \boldsymbol{\Delta}\boldsymbol{M}^{-1}) d\boldsymbol{M} \tag{3}$$

with $\boldsymbol{R} = \boldsymbol{\Delta}^{\frac{1}{2}}\boldsymbol{Q}\boldsymbol{\Delta}^{\frac{1}{2}}$.

3. Normal-inverse Wishart prior

Let z be a known positive scalar and $\boldsymbol{\Sigma}$ be a positive definite random matrix of dimension p. Let $\boldsymbol{\Psi} = z^{-1}\boldsymbol{\Sigma}$ follow an inverse Wishart distribution with parameter $\boldsymbol{\Phi}$ and m degrees of freedom. For any $z > 0$, a generated variate of $\boldsymbol{\Psi}$ will produce a generated variate of $\boldsymbol{\Sigma}$ since $\boldsymbol{\Sigma} = z\boldsymbol{\Psi}$. Now assume a normal-inverse Wishart prior for $(\boldsymbol{\mu},\boldsymbol{\Psi})$ with prior distributions for $\boldsymbol{\mu}$ and $\boldsymbol{\Psi}$ respectively, $\boldsymbol{\mu}|\boldsymbol{\Psi} \sim N_{p,s}(\boldsymbol{\theta}_{p\times s}, \frac{1}{n_0}\boldsymbol{\Psi}_{p\times p} \otimes \boldsymbol{\Omega}_{s\times s})$ and $\boldsymbol{\Psi} \sim W^{-1}(\boldsymbol{\Phi}, p, m)$. From eq. 2.2.1, p. 55 and Definition 4.2.1, p. 111 of Gupta and Nagar (2000) the prior densities are

$$\pi(\boldsymbol{\mu}|\boldsymbol{\Psi}) = (2\pi)^{-\frac{sp}{2}} |\frac{1}{n_0}\boldsymbol{\Psi}|^{-\frac{s}{2}} |\boldsymbol{\Omega}|^{-\frac{p}{2}} etr\left[-\frac{n_0}{2}\boldsymbol{\Psi}^{-1}(\boldsymbol{\mu}-\boldsymbol{\theta})\boldsymbol{\Omega}^{-1}(\boldsymbol{\mu}-\boldsymbol{\theta})'\right], \boldsymbol{\mu} \in \boldsymbol{R}^{p\times s}$$

and

$$\pi(\boldsymbol{\Psi}) = \left[\Gamma_p\left(\frac{m-p-1}{2}\right)\right]^{-1} |\frac{1}{2}\boldsymbol{\Phi}|^{\frac{1}{2}(m-p-1)} |\boldsymbol{\Psi}|^{-\frac{1}{2}m} etr\left[-\frac{1}{2}\boldsymbol{\Psi}^{-1}\boldsymbol{\Phi}\right], \boldsymbol{\Psi} > \boldsymbol{0}, \boldsymbol{\Phi} > \boldsymbol{0} \text{ and } m > 2p$$

with the joint prior density function as

$$\pi(\boldsymbol{\mu},\boldsymbol{\Psi}) \propto |\boldsymbol{\Psi}|^{-\frac{1}{2}(m+s)} etr\left[-\frac{1}{2}\boldsymbol{\Psi}^{-1}\left(n_0(\boldsymbol{\mu}-\boldsymbol{\theta})\boldsymbol{\Omega}^{-1}(\boldsymbol{\mu}-\boldsymbol{\theta})' + \boldsymbol{\Phi}\right)\right]$$

with known hyperparameters $\boldsymbol{\theta}, \boldsymbol{\Omega}$, n_0 and m where $etr(\cdot)$ denotes $exp[tr(\cdot)]$. Now

$$\begin{aligned}
& \pi(\boldsymbol{\mu},\boldsymbol{\Sigma}|z) \\
\propto\ & \pi(\boldsymbol{\mu},\boldsymbol{\Psi})|J(\boldsymbol{\Psi}\to\boldsymbol{\Sigma})| \\
\propto\ & z^{-\frac{p(p+1-m-s)}{2}}|\boldsymbol{\Sigma}|^{-\frac{1}{2}(m+s)} \\
\times\ & etr\left[-\frac{1}{2}z\boldsymbol{\Sigma}^{-1}\left(n_0(\boldsymbol{\mu}-\boldsymbol{\theta})\boldsymbol{\Omega}^{-1}(\boldsymbol{\mu}-\boldsymbol{\theta})'+\boldsymbol{\Phi}\right)\right], \qquad (4)
\end{aligned}$$

since the Jacobian is $J(\boldsymbol{\Psi}\to\boldsymbol{\Sigma}) = z^{\frac{-p(p+1)}{2}}$. Following Arashi et al. (2014), the conjugate prior for the matrix variate elliptical model can be obtained as

$$\pi(\boldsymbol{\mu},\boldsymbol{\Sigma}) \propto \int_0^\infty \pi(\boldsymbol{\mu},\boldsymbol{\Sigma}|z)w(z)dz.$$

This representation of a prior distribution coincides with the representation in (2). It should be noted that in (2), z is not a random variable.

3.1. Posterior distributions

The likelihood function is obtained from (2) as follows

$$\begin{aligned}
& L(\boldsymbol{\mu},\boldsymbol{\Sigma}|\overline{\boldsymbol{X}},\boldsymbol{V}) \\
=\ & \prod_{i=1}^{n}\int_0^\infty w(z)f_{N_{\boldsymbol{\mu},z^{-1}\boldsymbol{\Sigma}\otimes\boldsymbol{\Omega}}}(\boldsymbol{x}|\boldsymbol{\mu},\boldsymbol{\Sigma})dz \\
\propto\ & \int_0^\infty w(z)|z^{-1}\boldsymbol{\Sigma}|^{-\frac{ns}{2}}|\boldsymbol{\Omega}|^{-\frac{np}{2}}etr\left[-\frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{X}_i-\boldsymbol{\mu})'z\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_i-\boldsymbol{\mu})\boldsymbol{\Omega}^{-1}\right]dz \\
=\ & \int_0^\infty w(z)z^{\frac{snp}{2}}|\boldsymbol{\Sigma}|^{-\frac{sn}{2}}|\boldsymbol{\Omega}|^{-\frac{np}{2}} \\
\times\ & etr\left[-\frac{1}{2}z\boldsymbol{\Sigma}^{-1}\left[\boldsymbol{V}+n(\overline{\boldsymbol{X}}-\boldsymbol{\mu})\boldsymbol{\Omega}^{-1}(\overline{\boldsymbol{X}}-\boldsymbol{\mu})'\right]\right]dz, \qquad (5)
\end{aligned}$$

where

$$\boldsymbol{V} = \sum_{i=1}^{n}(\boldsymbol{X}_i-\overline{\boldsymbol{X}})\boldsymbol{\Omega}^{-1}(\boldsymbol{X}_i-\overline{\boldsymbol{X}})'. \qquad (6)$$

From (4) and (5) the joint posterior density function is

$$\begin{aligned}
q(\boldsymbol{\mu},\boldsymbol{\Sigma}|\overline{\boldsymbol{X}},\boldsymbol{V}) \propto\ & \pi(\boldsymbol{\mu},\boldsymbol{\Sigma})L(\boldsymbol{\mu},\boldsymbol{\Sigma}|\overline{\boldsymbol{X}},\boldsymbol{V}) \\
=\ & \int_0^\infty z^{-\frac{p(p+1-m-s-ns)}{2}}w(z)|\boldsymbol{\Sigma}|^{-\frac{1}{2}(m+sn+s)} \\
& \times etr\left[-\frac{1}{2}z\boldsymbol{\Sigma}^{-1}\left(n_0(\boldsymbol{\mu}-\boldsymbol{\theta})\boldsymbol{\Omega}^{-1}(\boldsymbol{\mu}-\boldsymbol{\theta})'+\boldsymbol{\Phi}\right)\right] \\
& \times etr\left[-\frac{1}{2}z\boldsymbol{\Sigma}^{-1}\left[\boldsymbol{V}+n(\overline{\boldsymbol{X}}-\boldsymbol{\mu})\boldsymbol{\Omega}^{-1}(\overline{\boldsymbol{X}}-\boldsymbol{\mu})'\right]\right]dz.
\end{aligned}$$

Theorem 1 The marginal posterior distribution of the location matrix, $\boldsymbol{\mu}$, under the model in (1) and prior (4) is a matrix variate t-distribution with parameters $\boldsymbol{b}$,$\boldsymbol{W}$ and $\frac{1}{n+n_0}\boldsymbol{\Omega}$ and degrees of freedom $(m+ns-2p)$ with density function

$$
\begin{aligned}
q(\boldsymbol{\mu}|\overline{\boldsymbol{X}},\boldsymbol{V}) &= \frac{\Gamma_p\left(\frac{1}{2}(m+ns+s-p-1)\right)}{\pi^{\frac{1}{2}sp}\Gamma_p\left(\frac{1}{2}(m+ns-p-1)\right)}|\frac{1}{n+n_0}\boldsymbol{\Omega}|^{-\frac{1}{2}p}|\boldsymbol{W}|^{-\frac{1}{2}m} \\
&\quad \times|\boldsymbol{I}_p+\boldsymbol{W}^{-1}(\boldsymbol{\mu}-\boldsymbol{b})\left(\frac{1}{n+n_0}\boldsymbol{\Omega}\right)^{-1}(\boldsymbol{\mu}-\boldsymbol{b})'|^{-\frac{m+ns+s-p-1}{2}} \qquad (7)
\end{aligned}
$$

with $\boldsymbol{W}=\frac{nn_0}{n+n_0}(\overline{\boldsymbol{X}}-\boldsymbol{\theta})\boldsymbol{\Omega}^{-1}(\overline{\boldsymbol{X}}-\boldsymbol{\theta})'+\boldsymbol{\Phi}+\boldsymbol{V}$ and $\boldsymbol{\mu}\in\boldsymbol{R}^{p\times s}$ and with $\boldsymbol{b}=\frac{1}{n+n_0}\left(n\overline{\boldsymbol{X}}-n_0\boldsymbol{\theta}\right)$.

Proof. See Appendix A. ■

Let

$$\varsigma(\alpha)=\int_0^\infty z^\alpha w(z)dz$$

provided this integral exists. The convergence of this integral is dependent on the specific weight function $w(z)$ chosen. Consider the matrix variate t-distribution with parameters $\boldsymbol{\mu}_{p\times s}$ and $\boldsymbol{\Sigma}_{p\times p}\otimes\boldsymbol{\Omega}_{s\times s}$ and ν_0 degrees of freedom, as an example. Then from Table 1, p. 648 of Chu (1973), the associated weight function is the inverse gamma distribution with density function: $w(z)=\frac{(\frac{\nu_0}{2})^{\frac{\nu_0}{2}}z^{\frac{\nu_0}{2}-1}\exp(\frac{-\nu_0 z}{2})}{\Gamma(\frac{\nu_0}{2})}$, hence $\varsigma(\alpha)$ will only exist if $\alpha>-\frac{\nu_0}{2}$.

Theorem 2 The marginal posterior density function of the characteristic matrix, $\boldsymbol{\Sigma}$, under the model in (1) and prior (4) is

$$
\begin{aligned}
q(\boldsymbol{\Sigma}|\overline{\boldsymbol{X}},\boldsymbol{V}) &= \frac{2^{-\frac{p(m+ns-p-1)}{2}}}{\Gamma_p\left(\frac{m+ns-p-1}{2}\right)}|\boldsymbol{W}|^{\frac{m+ns-p-1}{2}}|\boldsymbol{\Sigma}|^{-\frac{1}{2}(m+ns)} \\
&\quad \times\sum_{k=0}^{\infty}\frac{\left(-\frac{1}{2}tr\left[\boldsymbol{\Sigma}^{-1}\boldsymbol{W}\right]\right)^k}{k!}\varsigma(-\frac{p(p+1-m-ns)-2k}{2}) \qquad (8)
\end{aligned}
$$

with $\boldsymbol{W}$ as defined in Theorem 1 and $\boldsymbol{\Sigma}>\boldsymbol{0}$, provided $\varsigma(-\frac{p(p+1-m-ns)-2k}{2})$ exists.

Proof. See Appendix A. ■

3.2. Statistical properties

In this section some statistical properties of the posterior distribution of $\boldsymbol{\Sigma}$ is derived. The characteristic function for $\boldsymbol{\Sigma}$ is obtained followed by the joint density function of eigenvalues of $\boldsymbol{\Sigma}$, as well as the density function of the largest eigenvalue of $\boldsymbol{\Sigma}$ amongst others.

Theorem 3 The characteristic function of $\boldsymbol{\Sigma}$ under the model in (1) and prior (4) is

$$
\begin{aligned}
\varphi(\boldsymbol{T}) &= \frac{2^{-\frac{p(m+ns-p-1)}{2}}(-i)^{\frac{p}{2}(m+ns)-\frac{p(p+1)}{2}}}{\Gamma_p\left(\frac{m+ns-p-1}{2}\right)}|\boldsymbol{W}|^{\frac{m+ns-p-1}{2}}|\boldsymbol{T}|^{\frac{1}{2}(m+ns)-\frac{p+1}{2}} \\
&\quad \times\int_0^\infty z^{-\frac{p(p+1-m-ns)}{2}}w(z)B_{\frac{m+ns-p-1}{2}}\left(-\frac{i}{2}z\boldsymbol{T}\boldsymbol{W}\right)dz \qquad (9)
\end{aligned}
$$

with $\boldsymbol{W}$ as defined in Theorem 1, $\boldsymbol{T}_{p\times p}$ is a real positive definite arbitrary matrix and $B_\delta(\cdot)$ is the type II Bessel function of matrix argument (see (1)).

Proof. From Theorem 2 follows that the characteristic function of $\boldsymbol{\Sigma}$ is

$$\begin{aligned}
\varphi(\boldsymbol{T}) &= E\left[etr(i\boldsymbol{T}\boldsymbol{\Sigma})\,|\overline{\boldsymbol{X}},\boldsymbol{V}\right] \\
&= \frac{2^{-\frac{p(m+ns-p-1)}{2}}}{\Gamma_p\left(\frac{m+ns-p-1}{2}\right)}|\boldsymbol{W}|^{\frac{m+ns-p-1}{2}}\int_0^\infty z^{-\frac{p(p+1-m-ns)}{2}}w(z) \\
&\quad\times\int_{\boldsymbol{\Sigma}>\boldsymbol{0}}|\boldsymbol{\Sigma}|^{-\frac{1}{2}(m+ns)}etr\left[i\boldsymbol{T}\boldsymbol{\Sigma}\right]etr\left[-\frac{1}{2}z\boldsymbol{\Sigma}^{-1}\boldsymbol{W}\right]d\boldsymbol{\Sigma}dz.
\end{aligned}$$

Using eq. 1.6.18 on p. 39 of Gupta and Nagar (2000), (9) follows. ■

Theorem 4 The joint density function of the eigenvalues $\boldsymbol{\Lambda} = diag(\lambda_1,...\lambda_p)$, $\lambda_p > ... > \lambda_1 > 0$ of $\boldsymbol{\Sigma}$ under the model in (1) and prior (4) is

$$\begin{aligned}
g(\boldsymbol{\Lambda}) &= \frac{\pi^{\frac{1}{2}p^2}}{\Gamma_p\left(\frac{p}{2}\right)}\prod_{i<j}^{p}(\lambda_i-\lambda_j)\frac{1}{\Gamma_p\left(\frac{m+ns-p-1}{2}\right)}|\boldsymbol{W}|^{\frac{m+ns-p-1}{2}}\prod_{i=1}^{p}\lambda_i^{-\frac{1}{2}(m+ns)} \\
&\quad\times\sum_{k=0}^{\infty}\sum_{\kappa}\frac{(-1)^k 2^{-\frac{p(m+ns-p-1)}{2}-k}C_\kappa\left(\boldsymbol{\Lambda}^{-1}\right)C_\kappa(\boldsymbol{W})}{k!C_\kappa(\boldsymbol{I}_p)} \\
&\quad\times\varsigma\left(-\frac{p(p+1-m-ns)}{2}+k\right)
\end{aligned}$$

with $\boldsymbol{W}$ as defined in Theorem 1 and $C_\kappa(\cdot)$ is the zonal polynomial (see Gupta and Nagar, 2000) corresponding to κ, provided $\varsigma\left(-\frac{p(p+1-m-ns)}{2}+k\right)$ exists.

Proof. See Appendix A. ■

Theorem 5 The cumulative distribution function of $\boldsymbol{\Sigma}$ under the model in (1) and prior (4) is given by

$$P(\boldsymbol{\Sigma}<\boldsymbol{A}) = \sum_{l=0}^{\infty}\sum_{k=0}^{\infty}\sum_{\kappa}\frac{2^{-k-l}(-1)^l\left[tr\left(\boldsymbol{A}^{-1}\boldsymbol{W}\right)\right]^l}{k!l!}\varsigma(l+k)C_\kappa\left(\boldsymbol{A}^{-1}\boldsymbol{W}\right) \tag{10}$$

for any $\boldsymbol{A} > \boldsymbol{0}$, provided $\varsigma(l+k)$ exists with $\boldsymbol{W}$ as defined in Theorem 1.

Proof. See Appendix A. ■

Remark 2 Note that $\lambda_{(p)} < a$ is equivalent to $\boldsymbol{\Sigma} < a\boldsymbol{I}_p$ since $\boldsymbol{H\Lambda H}' = \boldsymbol{\Sigma}$. To obtain the cumulative distribution function of $\lambda_{(p)}$, the largest eigenvalue of $\boldsymbol{\Sigma}$ under the model in (1) and prior (4), the previous theorem can be used with $\boldsymbol{A} = a\boldsymbol{I}_p$. Also, $\lambda_{(1)} > b$ is equivalent to $\boldsymbol{\Sigma} > b\boldsymbol{I}_{p.}$

Theorem 6 The cumulative distribution function of $\lambda_{(p)}$, the largest eigenvalue of $\boldsymbol{\Sigma}$ under the model in (1) and prior (4) is

$$F_{\lambda_{(p)}}(a) = \sum_{l=0}^{\infty}\sum_{k=0}^{\infty}\sum_{\kappa}\frac{(2a)^{-k-l}(-1)^l\left[tr(\boldsymbol{W})\right]^l}{k!l!}\varsigma(l+k)C_\kappa(\boldsymbol{W}) \tag{11}$$

provided $\varsigma(l+k)$ exists and with $\boldsymbol{W}$ as defined in Theorem 1.

Proof. From (10) the cumulative distribution function of $\lambda_{(p)}$, the largest eigenvalue of $\boldsymbol{\Sigma}$ under the model in (1) and prior (4) is

$$\begin{aligned} F_{\lambda_{(p)}}(a) &= P(\boldsymbol{\Sigma} < a\boldsymbol{I}_p) \\ &= \sum_{l=0}^{\infty}\sum_{k=0}^{\infty}\sum_{\kappa}\frac{2^{-k-l}(-1)^l\left[tr\left(\frac{1}{a}\boldsymbol{W}\right)\right]^l}{k!l!}\varsigma(l+k)C_\kappa\left(\frac{1}{a}\boldsymbol{W}\right). \end{aligned}$$

■

3.3. Bayesian inference

In this section Bayes estimators are derived for the matrix parameters utilizing the two loss functions mentioned in Section 1.

Theorem 7 Under the squared error loss function, the Bayes estimator of the location matrix, $\boldsymbol{\mu}$, under the model in (1) and prior (4) is the posterior mean (PM estimator) of $\boldsymbol{\mu}$, therefore

$$\widehat{\boldsymbol{\mu}} = \boldsymbol{b} = \frac{1}{n+n_0}\left(n\overline{\boldsymbol{X}} - n_0\boldsymbol{\theta}\right). \tag{12}$$

Proof. From Theorem 4.3.1, p. 135 of Gupta and Nagar (2000), the marginal posterior distribution of the location matrix, $\boldsymbol{\mu}$ is a matrix variate t-distribution with parameters $\boldsymbol{b}$,$\boldsymbol{W}$, $\frac{1}{n+n_0}\boldsymbol{\Omega}$ and degrees of freedom $(m+ns-2p)$ from (7). Under the squared error loss function the Bayes estimator for $\boldsymbol{\mu}$ is

$$\widehat{\boldsymbol{\mu}} = E\left[\boldsymbol{\mu}|\overline{\boldsymbol{X}},\boldsymbol{V}\right] = \boldsymbol{b}.$$

■

Remark 3 Under the squared error loss function, the Bayes estimator of $\boldsymbol{\Sigma}$, under the model in (1) and prior (4) is the posterior mean (PM estimator) of $\boldsymbol{\Sigma}$, hence

$$\widehat{\boldsymbol{\Sigma}} = E\left[\boldsymbol{\Sigma}|\overline{\boldsymbol{X}},\boldsymbol{V}\right].$$

Remark 4 Under the loss function $L(\boldsymbol{\omega},\widehat{\boldsymbol{\omega}}) = \log\left[\frac{q\left(\boldsymbol{\omega}|\overline{\boldsymbol{X}},\boldsymbol{V}\right)}{q\left(\widehat{\boldsymbol{\omega}}|\overline{\boldsymbol{X}},\boldsymbol{V}\right)}\right]$ (see Theorem 1 of Das and Dey, 2010), the Bayes estimators of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, respectively, are the modes of the respective posterior distributions (MAP estimators).

Lemma 8 The h^{th} posterior moment of $|\boldsymbol{\Sigma}|$ under the model in (1) and prior (4) is

$$m_h = E\left[|\boldsymbol{\Sigma}|^h|\overline{\boldsymbol{X}},\boldsymbol{V};z\right] = \frac{2^{-ph}\varsigma(ph)\Gamma_p\left(\frac{m+ns-p-1-2h}{2}\right)}{\Gamma_p\left(\frac{m+ns-p-1}{2}\right)}|\boldsymbol{W}|^h \tag{13}$$

with $\boldsymbol{W}$ as defined in Theorem 1, provided $\varsigma(ph)$ exists.

Proof. From (8) and Definition 4.2.1, p. 111 of Gupta and Nagar (2000) the h^{th} posterior moment is given by

$$\begin{aligned} E\left[|\boldsymbol{\Sigma}|^h|\overline{\boldsymbol{X}},\boldsymbol{V};z\right] &= \frac{2^{-\frac{p(m+ns-p-1)}{2}}}{\Gamma_p\left(\frac{m+ns-p-1}{2}\right)}|\boldsymbol{W}|^{\frac{m+ns-p-1}{2}}\int_0^\infty z^{-\frac{p(p+1-m-ns)}{2}}w(z) \\ &\quad \times \int_{\boldsymbol{\Sigma}>\boldsymbol{0}}|\boldsymbol{\Sigma}|^{-\frac{1}{2}(m+ns-2h)}etr\left[-\frac{1}{2}z\boldsymbol{\Sigma}^{-1}\boldsymbol{W}\right]d\boldsymbol{\Sigma}dz \end{aligned}$$

and (13) follows. ■

Theorem 9 Under the squared error loss function, the Bayes estimator of $|\boldsymbol{\Sigma}|$ under the model in (1) and prior (4) is

$$\widehat{|\boldsymbol{\Sigma}|} = \frac{2^{-p}\varsigma(p)\Gamma_p\left(\frac{m+ns-p-3}{2}\right)}{\Gamma_p\left(\frac{m+ns-p-1}{2}\right)}|\boldsymbol{W}| \tag{14}$$

with $\boldsymbol{W}$ as defined in Theorem 1, provided $\varsigma(p)$ exists.

Proof. The result is immediate from (13) with $h=1$. ■

Remark 5 The results derived for the model in (1) and prior (4) simplifies for $s=1$ to the results obtained for the multivariate elliptical model and for $s=1, p=1$ to the univariate elliptical model (see Van Niekerk et al., 2013).

The question arises whether the normal-Wishart prior performs as well as or better than the normal-inverse Wishart prior. To this end, we need to study the Bayesian analysis of the elliptical model under the latter prior structure, which is the focus of the forthcoming section.

4. Normal-Wishart prior

In this section the normal-Wishart prior for the matrix variate elliptical model is considered. Similarly to Section 2, the prior distributions for $\boldsymbol{\mu}$ and $\boldsymbol{\Psi}=z^{-1}\boldsymbol{\Sigma}$ respectively are from eq. 2.2.1 on p. 55 and eq. 3.2.1 on p. 87 of Gupta and Nagar (2000), $\boldsymbol{\mu}|\boldsymbol{\Psi}\sim N_{p,s}(\boldsymbol{\theta}_{p\times s},\frac{1}{n_0}\boldsymbol{\Psi}_{p\times p}\otimes\boldsymbol{\Omega}_{s\times s})$ and $\boldsymbol{\Psi}\sim W(\boldsymbol{\Phi},p,m)$ with density functions

$$\pi(\boldsymbol{\mu}|\boldsymbol{\Psi}) = (2\pi)^{-\frac{sp}{2}}|\frac{1}{n_0}\boldsymbol{\Psi}|^{-\frac{s}{2}}|\boldsymbol{\Omega}|^{-\frac{p}{2}}etr\left[-\frac{n_0}{2}\boldsymbol{\Psi}^{-1}(\boldsymbol{\mu}-\boldsymbol{\theta})\boldsymbol{\Omega}^{-1}(\boldsymbol{\mu}-\boldsymbol{\theta})'\right],\ \boldsymbol{\mu}\in\boldsymbol{R}^{p\times s}$$

and

$$\pi(\boldsymbol{\Psi}) = 2^{-\frac{mp}{2}}\left[\Gamma_p(\frac{m}{2})\right]^{-1}|\boldsymbol{\Phi}|^{-\frac{1}{2}m}|\boldsymbol{\Psi}|^{\frac{1}{2}(m-p-1)}etr\left[-\frac{1}{2}\boldsymbol{\Psi}\boldsymbol{\Phi}^{-1}\right],\ \boldsymbol{\Psi}>\boldsymbol{0},\ \boldsymbol{\Phi}>\boldsymbol{0}\text{ and }m\geq p$$

with the joint prior density function as

$$\pi(\boldsymbol{\mu},\boldsymbol{\Psi}) \propto |\boldsymbol{\Psi}|^{\frac{1}{2}(m-p-1-s)}etr\left[-\frac{n_0}{2}\boldsymbol{\Psi}^{-1}(\boldsymbol{\mu}-\boldsymbol{\theta})\boldsymbol{\Omega}^{-1}(\boldsymbol{\mu}-\boldsymbol{\theta})'\right]etr\left[-\frac{1}{2}\boldsymbol{\Psi}\boldsymbol{\Phi}^{-1}\right].$$

It then follows that

$$\pi(\boldsymbol{\mu},\boldsymbol{\Sigma}|z) \propto z^{-\frac{p(m-s)}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}(m-p-1-s)}etr\left[-\frac{n_0z}{2}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}-\boldsymbol{\theta})\boldsymbol{\Omega}^{-1}(\boldsymbol{\mu}-\boldsymbol{\theta})'\right]etr\left[-\frac{1}{2z}\boldsymbol{\Sigma}\boldsymbol{\Phi}^{-1}\right]. \tag{15}$$

4.1. Posterior distributions

From (5) and (15) the joint posterior density function follows as,

$$\begin{aligned} q(\boldsymbol{\mu},\boldsymbol{\Sigma}|\overline{\boldsymbol{X}},\boldsymbol{V}) \quad &\propto \quad \int_0^\infty z^{-\frac{p(m-s-sn)}{2}} w(z)|\boldsymbol{\Sigma}|^{\frac{1}{2}(m-p-1-s-ns)} etr\left[-\frac{1}{2z}\boldsymbol{\Sigma}\boldsymbol{\Phi}^{-1}\right] \\ &\quad \times etr\left[-\frac{n_0 z}{2}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}-\boldsymbol{\theta})\boldsymbol{\Omega}^{-1}(\boldsymbol{\mu}-\boldsymbol{\theta})'\right] \\ &\quad \times etr\left[-\frac{1}{2}z\boldsymbol{\Sigma}^{-1}\left[\boldsymbol{V}+n(\overline{\boldsymbol{X}}-\boldsymbol{\mu})\boldsymbol{\Omega}^{-1}(\overline{\boldsymbol{X}}-\boldsymbol{\mu})'\right]\right] dz \end{aligned}$$

with $\boldsymbol{V}$ as defined in (6).

Theorem 10 The marginal posterior density function of the location matrix, $\boldsymbol{\mu}$, under the model in (1) and prior in (15) is

$$\begin{aligned} q(\boldsymbol{\mu}|\overline{\boldsymbol{X}},\boldsymbol{V}) \quad &= \quad \frac{|\boldsymbol{\Omega}|^{-\frac{p}{2}}(n+n_0)^{\frac{sp}{2}}}{(2\pi)^{\frac{sp}{2}} B_{\frac{-m+ns}{2}}\left(\frac{1}{4}\boldsymbol{\Phi}^{-1}\boldsymbol{Y}\right)} |2\boldsymbol{\Phi}|^{-\frac{s}{2}} \\ &\quad \times B_{\frac{-m+s+ns}{2}}\left(\frac{1}{4}\boldsymbol{\Phi}^{-1}\left(\boldsymbol{Y}+(n+n_0)(\boldsymbol{\mu}-\boldsymbol{b})\boldsymbol{\Omega}^{-1}(\boldsymbol{\mu}-\boldsymbol{b})'\right)\right) \end{aligned} \tag{16}$$

with $\boldsymbol{Y}=\frac{nn_0}{n+n_0}(\overline{\boldsymbol{X}}-\boldsymbol{\theta})\boldsymbol{\Omega}^{-1}(\overline{\boldsymbol{X}}-\boldsymbol{\theta})'+\boldsymbol{V}$ and $\boldsymbol{\mu}\in\boldsymbol{R}^{p\times s}$.

Proof. See Appendix B. ■

Remark 6 The marginal posterior distribution of the location matrix in matrix elliptical models is robust with respect to departures from normality, under the non-conjugate normal-Wishart prior.

Theorem 11 The marginal posterior density function of the characteristic matrix, $\boldsymbol{\Sigma}$, under model (1) and prior (15) is

$$\begin{aligned} q(\boldsymbol{\Sigma}|\bar{\boldsymbol{x}},\boldsymbol{V}) \quad &= \quad \frac{2^{\frac{p(-m+ns)}{2}}}{B_{\frac{-m+ns}{2}}\left(\frac{1}{4}\boldsymbol{\Phi}^{-1}\boldsymbol{Y}\right)} |\boldsymbol{\Phi}|^{\frac{1}{2}(-m+ns)}|\boldsymbol{\Sigma}|^{\frac{1}{2}(m-p-1-ns)} \\ &\quad \times \sum_{k=0}^{\infty}\sum_{l=0}^{\infty} \frac{(-2)^{-k-l}\left(tr\left[\boldsymbol{\Sigma}\boldsymbol{\Phi}^{-1}\right]\right)^k\left(tr\left[\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}\right]\right)^l}{k!l!} \\ &\quad \times \varsigma\left(-\frac{p(m-sn)}{2}-k+l\right) \end{aligned} \tag{17}$$

with $\boldsymbol{Y}$ as defined in Theorem 10 and $\boldsymbol{\Sigma}>\boldsymbol{0}$, provided $\varsigma\left(-\frac{p(m-sn)}{2}-k+l\right)$ exists.

Proof. See Appendix B. ■

4.2. Statistical properties

Theorem 12 The characteristic function of $\boldsymbol{\Sigma}$ under model (1) and prior (15) is

$$\begin{aligned} \varphi(\boldsymbol{T}) \quad &= \quad \frac{2^{\frac{p(-m+ns)}{2}}}{B_{\frac{-m+ns}{2}}\left(\frac{1}{4}\boldsymbol{\Phi}^{-1}\boldsymbol{Y}\right)} |\boldsymbol{\Phi}|^{\frac{1}{2}(-m+ns)} \int_0^\infty z^{-\frac{p(m-sn)}{2}} w(z) \\ &\quad \times \left|\frac{1}{2z}\boldsymbol{\Phi}^{-1}-i\boldsymbol{T}'\right|^{\frac{1}{2}(m+ns)} B_{\frac{-m+ns}{2}}\left[\frac{1}{4}\left(\boldsymbol{\Phi}^{-1}-2\,iz\boldsymbol{T}'\right)\boldsymbol{Y}\right] dz \end{aligned}$$

with $\boldsymbol{Y}$ as defined in Theorem 10, $\boldsymbol{T}_{p\times p}$ is a real positive definite arbitrary matrix.

Proof. From (17) the characteristic function of $\boldsymbol{\Sigma}|\overline{\boldsymbol{X}},\boldsymbol{V}$ is given by

$$\begin{aligned}\varphi(\boldsymbol{T}) &= \frac{2^{\frac{p(-m+ns)}{2}}}{B_{\frac{1}{2}(-m+ns)}\left(\frac{1}{4}\boldsymbol{\Phi}^{-1}\boldsymbol{Y}\right)}|\boldsymbol{\Phi}|^{\frac{1}{2}(-m+ns)}\int_0^\infty z^{-\frac{p(m-sn)}{2}}w(z)\\ &\times\int_{\boldsymbol{\Sigma}>\boldsymbol{0}}|\boldsymbol{\Sigma}|^{\frac{1}{2}(m-p-1-ns)}etr\left[-\frac{1}{2z}\boldsymbol{\Sigma}\boldsymbol{\Phi}^{-1}+i\boldsymbol{T}'\boldsymbol{\Sigma}\right]etr\left[-\frac{z}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}\right]d\boldsymbol{\Sigma}dz.\end{aligned}$$

Applying eq. 1.6.18, p. 39 of Gupta and Nagar (2000) the proof is complete. ∎

Theorem 13 The joint density function of the eigenvalues $\boldsymbol{\Lambda}=diag(\lambda_1,...\lambda_p)$, $\lambda_p>...>\lambda_1>0$ of $\boldsymbol{\Sigma}$ under model (1) and prior (15) has the form

$$\begin{aligned}g(\boldsymbol{\Lambda}) &= \frac{\pi^{\frac{1}{2}p^2}}{\Gamma_p\left(\frac{p}{2}\right)}\prod_{i<j}^{p}(\lambda_i-\lambda_j)\frac{2^{\frac{p}{2}(-m+ns)}}{B_{\frac{-m+ns}{2}}\left(\frac{1}{4}\boldsymbol{\Phi}^{-1}\boldsymbol{Y}\right)}|\boldsymbol{\Phi}|^{\frac{1}{2}(-m+ns)}\prod_{i=1}^{p}\lambda_i^{\frac{1}{2}(m-p-1-ns)}\\ &\times\sum_{k=0}^\infty\sum_\kappa\sum_{l=0}^\infty\sum_\lambda\frac{1}{k!l!}\left(-\frac{1}{2}\right)^{k+l}\varsigma\left(-\frac{p(m-sn)+2k-2l}{2}\right)\\ &\times\sum_{\phi\in\kappa,\lambda}\frac{C_\phi^{\kappa,\lambda}\left(\boldsymbol{\Lambda},\boldsymbol{\Lambda}^{-1}\right)C_\phi^{\kappa,\lambda}\left(\boldsymbol{\Phi}^{-1},\boldsymbol{Y}\right)}{C_\phi\left(\boldsymbol{I}_p\right)}\end{aligned}$$

with $\boldsymbol{Y}$ as defined in Theorem 10, provided $\varsigma\left(-\frac{p(m-sn)+2k-2l}{2}\right)$ exists.

Proof. See Appendix B. ∎

Theorem 14 The cumulative distribution function of $\boldsymbol{\Sigma}$ under model (1) and prior (15) is

$$P(\boldsymbol{\Sigma}<\boldsymbol{A})=\frac{2^{\frac{p(-m+ns)}{2}}\varsigma\left(-\frac{p(m-sn)}{2}\right)B_{\frac{-m+ns}{2}}\left(\frac{1}{4}\boldsymbol{\Phi}^{-1}\boldsymbol{Y};\boldsymbol{I}_p\right)}{B_{\frac{-m+ns}{2}}\left(\frac{1}{4}\boldsymbol{\Phi}^{-1}\boldsymbol{Y}\right)}|\boldsymbol{\Phi}|^{\frac{1}{2}(-m+ns)}|\boldsymbol{A}|^{\frac{1}{2}(m-ns)}$$

for any $\boldsymbol{A}>\boldsymbol{0}$ and $B_{\frac{m-ns}{2}}(\cdot;\cdot)$ the incomplete type II Bessel function of matrix argument (see (3)), with $\boldsymbol{Y}$ as defined in Theorem 10 and provided $\varsigma\left(-\frac{p(m-sn)}{2}\right)$ exists.

Proof. See Appendix B. ∎

Theorem 15 The cumulative distribution function of $\lambda_{(p)}$, the largest eigenvalue of $\boldsymbol{\Sigma}$ under model (1) and prior (15) has the form

$$F_{\lambda_{(p)}}(a)=\frac{2^{\frac{p}{2}(-m+ns)}a^{\frac{p}{2}(m-ns)}\varsigma\left(-\frac{p(m-sn)}{2}\right)B_{\frac{-m+ns}{2}}\left(\frac{1}{4}\boldsymbol{\Phi}^{-1}\boldsymbol{Y};\boldsymbol{I}_p\right)}{B_{\frac{-m+ns}{2}}\left(\frac{1}{4}\boldsymbol{\Phi}^{-1}\boldsymbol{Y}\right)}|\boldsymbol{\Phi}|^{\frac{1}{2}(-m+ns)} \tag{18}$$

with $\boldsymbol{Y}$ as defined in Theorem 10 and provided $\varsigma\left(-\frac{p(m-sn)}{2}\right)$ exists.

Proof. From (10) the cumulative distribution function of $\lambda_{(p)}$ follows easily. ∎

4.3. Bayesian inference

Theorem 16 Under the squared error loss function, the Bayes estimator (PM estimator) of the location matrix, $\boldsymbol{\mu}$, under model (1) and prior (15) is

$$\widehat{\boldsymbol{\mu}} = \boldsymbol{b} = \frac{1}{n+n_0}\left(n\overline{\boldsymbol{X}} - n_0\boldsymbol{\theta}\right). \tag{19}$$

Proof. Under the squared error loss function the Bayes estimator of $\boldsymbol{\mu}$ is the posterior mean, i.e. $\widehat{\boldsymbol{\mu}} = E\left[\boldsymbol{\mu}|\overline{\boldsymbol{X}},\boldsymbol{V}\right]$. Note that the expected value of $(\boldsymbol{\mu}-\boldsymbol{b})$ is

$$\begin{aligned} E\left[\boldsymbol{\mu}-\boldsymbol{b}|\overline{\boldsymbol{X}},\boldsymbol{V}\right] &= \frac{|\boldsymbol{\Omega}|^{-\frac{p}{2}}(n+n_0)^{\frac{sp}{2}}}{(2\pi)^{\frac{sp}{2}}B_{\frac{1}{2}(-m+ns)}\left(\frac{1}{4}\boldsymbol{\Phi}^{-1}\boldsymbol{Y}\right)}|2\boldsymbol{\Phi}|^{-\frac{s}{2}}\int_{\boldsymbol{\mu}}(\boldsymbol{\mu}-\boldsymbol{b}) \\ &\quad \times B_{\frac{1}{2}(-m+s+ns)}\left(\frac{1}{4}\boldsymbol{\Phi}^{-1}\left(\boldsymbol{Y}+(n+n_0)(\boldsymbol{\mu}-\boldsymbol{b})\boldsymbol{\Omega}^{-1}(\boldsymbol{\mu}-\boldsymbol{b})'\right)\right)d\boldsymbol{\mu}. \end{aligned}$$

This is an integral of an odd function and hence $E\left[\boldsymbol{\mu}-\boldsymbol{b}|\overline{\boldsymbol{X}},\boldsymbol{V}\right] = \mathbf{0}$. Therefore from Theorem 10 the Bayes estimator of the location matrix, $\boldsymbol{\mu}$, is $\widehat{\boldsymbol{\mu}} = \boldsymbol{b} = \frac{1}{n+n_0}\left(n\overline{\boldsymbol{X}} - n_0\boldsymbol{\theta}\right)$. ∎

Lemma 17 The h^{th} posterior moment of $|\boldsymbol{\Sigma}|$ under model (1) and prior (15) is

$$m_h = E\left[|\boldsymbol{\Sigma}|^h|\overline{\boldsymbol{X}},\boldsymbol{V}\right] = \frac{2^{ph}\varsigma(ph)B_{\frac{-m+ns-2h}{2}}\left(\frac{1}{4}\boldsymbol{\Phi}^{-1}\boldsymbol{Y}\right)}{B_{\frac{-m+ns}{2}}\left(\frac{1}{4}\boldsymbol{\Phi}^{-1}\boldsymbol{Y}\right)}|\boldsymbol{\Phi}|^h \tag{20}$$

with $\boldsymbol{Y}$ as defined in Theorem 10, provided $\varsigma(ph)$ exists.

Proof. From (17), the h^{th} posterior moment is given by

$$\begin{aligned} & E\left[|\boldsymbol{\Sigma}|^h|\overline{\boldsymbol{X}},\boldsymbol{V};z\right] \\ = \; & \frac{2^{\frac{p(-m+ns)}{2}}}{B_{\frac{-m+ns}{2}}\left(\frac{1}{4}\boldsymbol{\Phi}^{-1}\boldsymbol{Y}\right)}|\boldsymbol{\Phi}|^{\frac{1}{2}(-m+ns)}\int_0^\infty z^{-\frac{p(m-sn)}{2}}w(z) \\ & \times\int_{\boldsymbol{\Sigma}>\mathbf{0}}|\boldsymbol{\Sigma}|^{\frac{1}{2}(m-p-1-ns+2h)}etr\left[-\frac{1}{2z}\boldsymbol{\Sigma}\boldsymbol{\Phi}^{-1}\right]etr\left[-\frac{z}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}\right]d\boldsymbol{\Sigma}dz. \end{aligned}$$

Applying eq. 1.6.18, p. 39 of Gupta and Nagar (2000), the proof is complete. ∎

Theorem 18 Under the squared error loss function, the Bayes estimator of $|\boldsymbol{\Sigma}|$ under model (1) and prior (15) is

$$\widehat{|\boldsymbol{\Sigma}|} = \frac{2^p\varsigma(p)B_{\frac{-m+ns-2}{2}}\left(\frac{1}{4}\boldsymbol{\Phi}^{-1}\boldsymbol{Y}\right)}{B_{\frac{-m+ns}{2}}\left(\frac{1}{4}\boldsymbol{\Phi}^{-1}\boldsymbol{Y}\right)}|\boldsymbol{\Phi}| \tag{21}$$

with $\boldsymbol{Y}$ as defined in Theorem 10, provided $\varsigma(p)$ exists.

Proof. The result is immediate from (20) with $h=1$. ∎

Remark 7 The results derived for model (1) and prior (15) simplifies for $s=1$ to the results obtained for the multivariate elliptical model and for $s=1, p=1$ to the univariate elliptical model (see Van Niekerk et al., 2013).

5. Particular subfamilies

In this section the newly developed results will be applied to the matrix variate normal distribution and the matrix variate t-distribution as special cases of the matrix variate elliptical model.

Remark 8 (See Remark 4).The marginal posterior distribution of $\boldsymbol{\mu}$ for all matrix variate elliptical distributions and a normal-inverse Wishart prior is given in (7) and for the normal-Wishart prior in (16). The Bayes estimator of $\boldsymbol{\mu}$ for both prior structures considered is from (12) and (19),

$$\widehat{\boldsymbol{\mu}}_B = \boldsymbol{b} = \frac{1}{n+n_0}\left(n\overline{\boldsymbol{X}} - n_0\boldsymbol{\theta}\right).$$

5.1. Matrix variate normal distribution

Let $\boldsymbol{X}$ follow a matrix variate normal distribution with parameters $\boldsymbol{\mu}_{p\times s}$ and $\boldsymbol{\Sigma}_{p\times p}\otimes\boldsymbol{\Omega}_{s\times s}$. Then from Table 1, p. 648 of Chu (1973), the associated weight function is:

$$w(z) = \delta(z-1) = \lim_{d\to 0}\frac{1}{d\sqrt{\pi}}\exp\left(-\frac{(z-1)^2}{d^2}\right), \tag{22}$$

where $\delta(\cdot)$ is the Dirac delta function.

- First we consider the normal-inverse Wishart case. The marginal posterior density function of $\boldsymbol{\Sigma}$ for the normal-inverse Wishart prior is obtained by using (8) and (22),

$$\begin{aligned} q(\boldsymbol{\Sigma}|\bar{\boldsymbol{x}},\boldsymbol{V})_{IW} &= \frac{2^{-\frac{p(m+ns-p-1)}{2}}}{\Gamma_p\left(\frac{m+ns-p-1}{2}\right)}|\boldsymbol{W}|^{\frac{m+ns-p-1}{2}} \\ &\quad\times|\boldsymbol{\Sigma}|^{-\frac{1}{2}(m+ns)}\int_0^\infty z^{-\frac{p(p+1-m-ns)}{2}}\delta(z-1)etr\left[-\frac{1}{2}z\boldsymbol{\Sigma}^{-1}\boldsymbol{W}\right]dz \end{aligned}$$

with $\boldsymbol{W}$ as defined in Theorem 1. We note that $\int_0^\infty f(x)\delta(x)dx = f(0)$ with $x = z-1$, $f(x) = (x+1)^{-\frac{p(p+1-m-ns)}{2}}etr\left[-\frac{1}{2}(x+1)\boldsymbol{\Sigma}^{-1}\boldsymbol{W}\right]$. Therefore $f(0) = etr\left[-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{W}\right]$, and

$$\begin{aligned} q(\boldsymbol{\Sigma}|\bar{\boldsymbol{x}},\boldsymbol{V})_{IW} &= \frac{2^{-\frac{p(m+ns-p-1)}{2}}}{\Gamma_p\left(\frac{m+ns-p-1}{2}\right)}|\boldsymbol{W}|^{\frac{m+ns-p-1}{2}}|\boldsymbol{\Sigma}|^{-\frac{1}{2}(m+ns)}f(0) \\ &= \frac{2^{-\frac{p(m+ns-p-1)}{2}}}{\Gamma_p\left(\frac{m+ns-p-1}{2}\right)}|\boldsymbol{W}|^{\frac{m+ns-p-1}{2}}|\boldsymbol{\Sigma}|^{-\frac{1}{2}(m+ns)}etr\left[-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{W}\right]. \end{aligned}$$

Therefore $\boldsymbol{\Sigma}\,|\,\bar{\boldsymbol{x}},\boldsymbol{V} \sim W^{-1}(\boldsymbol{W},p,m+ns)$ (see Bekker and Roux, 1995). From (22) and (14) the Bayes estimator of $|\boldsymbol{\Sigma}|$ is

$$|\widehat{\boldsymbol{\Sigma}}|_{B,IW} = \frac{2^{-p}\Gamma_p\left(\frac{m+ns-p-3}{2}\right)}{\Gamma_p\left(\frac{m+ns-p-1}{2}\right)}|\boldsymbol{W}|\int_0^\infty z^p w(z)dz = \frac{2^{-p}\Gamma_p\left(\frac{m+ns-p-3}{2}\right)}{\Gamma_p\left(\frac{m+ns-p-1}{2}\right)}|\boldsymbol{W}|;$$

the cumulative distribution function of $\lambda_{(p)}$, the largest eigenvalue of $\boldsymbol{\Sigma}|\bar{\boldsymbol{x}},\boldsymbol{V}$ from (11) is

$$\begin{aligned} F_{\lambda_{(p)}}(a) &= \frac{2^{-\frac{p(m+ns-p-1)}{2}}}{\Gamma_p\left(\frac{m+ns-p-1}{2}\right)}|\boldsymbol{W}|^{\frac{m+ns-p-1}{2}} etr\left[-\frac{1}{2a}\boldsymbol{W}\right] \\ &\times\sum_k\sum_\kappa \frac{\left(-\frac{1}{2}(m+ns)+p+1\right)_\kappa(-a)^k\Gamma_p\left(\frac{1}{2}(m+ns)-\frac{p+1}{2},-\kappa\right)}{k!} \\ &\times|\frac{1}{2}\boldsymbol{W}|^{-\left[\frac{1}{2}(m+ns)-\frac{p+1}{2}\right]}C_\kappa\left(\frac{1}{2}\boldsymbol{W}\right) \end{aligned}$$

with $\boldsymbol{W}$ as defined in Theorem 1.

- Secondly, the normal-Wishart prior is considered. The marginal posterior distribution of $\boldsymbol{\Sigma}$ is obtained by using (17) and (22),

$$q(\boldsymbol{\Sigma}|\bar{\boldsymbol{x}},\boldsymbol{V})_W = \frac{2^{\frac{p(-m+ns)}{2}}|\boldsymbol{\Phi}|^{\frac{1}{2}(-m+ns)}}{B_{\frac{-m+ns}{2}}\left(\frac{1}{4}\boldsymbol{\Phi}^{-1}\boldsymbol{Y}\right)|}|\boldsymbol{\Sigma}|^{\frac{1}{2}(m-p-1-ns)}etr\left[-\frac{1}{2}\boldsymbol{\Sigma}\boldsymbol{\Phi}^{-1}\right]etr\left[-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}\right]$$

(see Bekker and Roux, 1995) with $\boldsymbol{Y}$ as defined in Theorem 10. From (22) and (21) the Bayes estimator of $|\boldsymbol{\Sigma}|$ is

$$|\widehat{\boldsymbol{\Sigma}}|_{B,W} = \frac{2^p B_{\frac{-m+ns-2}{2}}\left(\frac{1}{4}\boldsymbol{\Phi}^{-1}\boldsymbol{Y}\right)}{B_{\frac{-m+ns}{2}}\left(\frac{1}{4}\boldsymbol{\Phi}^{-1}\boldsymbol{Y}\right)}|\boldsymbol{\Phi}|;$$

from (22) and (18) the cumulative distribution function of $\lambda_{(p)}$ is

$$F_{\lambda_{(p)}}(a) = \frac{2^{\frac{p}{2}(-m+ns)}a^{\frac{p}{2}(m-ns)}B_{\frac{-m+ns}{2}}\left(\frac{1}{4}\boldsymbol{\Phi}^{-1}\boldsymbol{Y};\boldsymbol{I}_p\right)}{B_{\frac{-m+ns}{2}}\left(\frac{1}{4}\boldsymbol{\Phi}^{-1}\boldsymbol{Y}\right)}|\boldsymbol{\Phi}|^{\frac{1}{2}(-m+ns)}$$

with $\boldsymbol{Y}$ as defined in Theorem 10.

5.2. Matrix variate t-distribution

Let $\boldsymbol{X}$ follow a matrix variate t-distribution with parameters $\boldsymbol{\mu}_{p\times s}$ and $\boldsymbol{\Sigma}_{p\times p}\otimes\boldsymbol{\Omega}_{s\times s}$ and ν_0 degrees of freedom. Then from Table 1, p. 648 of Chu (1973), the associated weight function is the inverse gamma distribution with density function:

$$w(z) = \frac{\left(\frac{\nu_0}{2}\right)^{\frac{\nu_0}{2}}z^{\frac{\nu_0}{2}-1}\exp\left(\frac{-\nu_0 z}{2}\right)}{\Gamma\left(\frac{\nu_0}{2}\right)}. \tag{23}$$

- As before we consider the normal-inverse Wishart prior. The marginal posterior distribution of $\boldsymbol{\Sigma}$ using (23) and (1) is,

$$\begin{aligned} q(\boldsymbol{\Sigma}|\bar{\boldsymbol{x}},\boldsymbol{V})_{IW} &= \frac{2^{-\frac{p(m+ns-p-1)}{2}}\left(\frac{\nu_0}{2}\right)^{\frac{\nu_0}{2}}}{\Gamma_p\left(\frac{m+ns-p-1}{2}\right)\Gamma\left(\frac{\nu_0}{2}\right)}|\boldsymbol{W}|^{\frac{m+ns-p-1}{2}}|\boldsymbol{\Sigma}|^{-\frac{1}{2}(m+ns)} \\ &\times\int_0^\infty z^{-\frac{p(p+1-m-ns-\nu_0+2)}{2}}\exp\left[-\frac{z}{2}\left(tr\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{W}\right)+\nu_0\right)\right]dz \end{aligned}$$

with $\boldsymbol{W}$ as defined in Theorem 1. Note that from Gradshteyn and Ryzhik (2007)

$$\begin{aligned}
&\int_0^\infty z^{-\frac{p(p+1-m-ns-\nu_0+2)}{2}+1-1} \exp\left[-\frac{z}{2}\left(tr\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{W}\right)+\nu_0\right)\right] dz \\
= \quad & \frac{\Gamma\left(1+\frac{p(m+ns+\nu_0-p-1)}{2}\right)}{\left(\frac{tr\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{W}\right)+\nu_0}{2}\right)^{1+\frac{p(m+ns+\nu_0-p-1)}{2}}},
\end{aligned}$$

provided $-\frac{p(p+1-m-ns-\nu_0+2)}{2}+1>0$. Therefore

$$\begin{aligned}
q(\boldsymbol{\Sigma}|\bar{\boldsymbol{x}},\boldsymbol{V})_{IW} \quad = \quad & \frac{2^{-\frac{p(m+ns-p-1)}{2}}\left(\frac{\nu_0}{2}\right)^{\frac{\nu_0}{2}}\Gamma\left(1+\frac{p(m+ns+\nu_0-p-1)}{2}\right)}{\Gamma_p\left(\frac{m+ns-p-1}{2}\right)\Gamma\left(\frac{\nu_0}{2}\right)} \\
& \times\left(\frac{tr\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{W}\right)+\nu_0}{2}\right)^{1+\frac{p(m+ns+\nu_0-p-1)}{2}} |\boldsymbol{W}|^{\frac{m+ns-p-1}{2}}|\boldsymbol{\Sigma}|^{-\frac{1}{2}(m+ns)}.
\end{aligned}$$

From (23) and (14) the Bayes estimator of $|\boldsymbol{\Sigma}|$ is

$$|\widehat{\boldsymbol{\Sigma}}|_{B,IW}=\frac{2^{-p}\Gamma_p\left(\frac{m+ns-p-3}{2}\right)}{\Gamma_p\left(\frac{m+ns-p-1}{2}\right)}|\boldsymbol{W}|\int_0^\infty z^p\frac{\left(\frac{\nu_0}{2}\right)^{\frac{\nu_0}{2}}z^{\frac{\nu_0}{2}-1}\exp\left(-\frac{\nu_0 z}{2}\right)}{\Gamma\left(\frac{\nu_0}{2}\right)}dz.$$

Hence,

$$|\widehat{\boldsymbol{\Sigma}}|_{B,IW}=\frac{2^{-p-\frac{\nu_0}{2}}\nu_0^{-p}\Gamma_p\left(\frac{m+ns-p-3}{2}\right)\Gamma\left(\frac{\nu_0}{2}+p\right)}{\Gamma_p\left(\frac{m+ns-p-1}{2}\right)\Gamma\left(\frac{\nu_0}{2}\right)}|\boldsymbol{W}|.$$

The cumulative distribution function of $\lambda_{(p)}$, the largest eigenvalue of $\boldsymbol{\Sigma}|\bar{\boldsymbol{x}},\boldsymbol{V}$ from (11) is

$$\begin{aligned}
F_{\lambda_{(p)}}(a) \quad = \quad & \sum_{l=0}^\infty\sum_{k=0}^\infty\sum_\kappa\frac{(2a)^{-k-l}(-1)^{k+l}\left[tr\left(\boldsymbol{W}\right)\right]^l\left(-\frac{1}{2}(m+ns)+p+1\right)_\kappa}{k!l!\Gamma_p\left(\frac{m+ns-p-1}{2}\right)} \\
& \times\frac{\Gamma_p\left(\frac{1}{2}(m+ns)-\frac{p+1}{2},-\kappa\right)\Gamma\left(l+k+\frac{\nu_0}{2}\right)}{\left(\frac{\nu_0}{2}\right)^{l+k}\Gamma\left(\frac{\nu_0}{2}\right)}C_\kappa\left(\boldsymbol{W}\right).
\end{aligned}$$

- Secondly, the normal-Wishart prior is the focus. The marginal posterior distribution of $\boldsymbol{\Sigma}$ using (17) and (23) is,

$$\begin{aligned}
q(\boldsymbol{\Sigma}|\bar{\boldsymbol{x}},\boldsymbol{V})_W \quad = \quad & \frac{2^{\frac{p(-m+ns)}{2}}\left(\frac{\nu_0}{2}\right)^{\frac{\nu_0}{2}}}{B_{\frac{-m+ns}{2}}\left(\frac{1}{4}\boldsymbol{\Phi}^{-1}\boldsymbol{Y}\right)\Gamma\left(\frac{\nu_0}{2}\right)}|\boldsymbol{\Phi}|^{\frac{1}{2}(-m+ns)}|\boldsymbol{\Sigma}|^{\frac{1}{2}(m-p-1-ns)} \\
& \times\int_0^\infty z^{-\frac{p(m-sn-\nu_0)+2}{2}}etr\left[-\frac{1}{2z}\boldsymbol{\Sigma}\boldsymbol{\Phi}^{-1}\right]\exp\left[-\frac{z}{2}\left(tr\left[\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}\right]+\nu_0\right)\right]dz,
\end{aligned}$$

with $\boldsymbol{Y}$ as defined in Theorem 10. Note that

$$\begin{aligned}
&\int_0^\infty z^{-\frac{p(m-sn-v_0)+1}{2}} etr\left[-\frac{1}{2z}\boldsymbol{\Sigma}\boldsymbol{\Phi}^{-1}\right]\exp\left[-\frac{z}{2}\left(tr\left[\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}\right]+2v_0\right)\right]dz \\
= \;& 2\left[\frac{tr\left(\boldsymbol{\Sigma}\boldsymbol{\Phi}^{-1}\right)}{\left(tr\left[\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}\right]+2v_0\right)}\right]^{-\frac{p(m-sn-v_0)-1}{4}} K_{\frac{p(-m+sn+v_0)+1}{2}}\left(\sqrt{tr\left(\boldsymbol{\Sigma}\boldsymbol{\Phi}^{-1}\right)\left(tr\left[\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}\right]+2v_0\right)}\right)
\end{aligned}$$

from Erdelyi, Magnus, Oberhettinger and Tricomi (1953) where $K_\delta(\cdot)$ is the Bessel function of the third kind (see Erdelyi et al., 1953). Hence

$$\begin{aligned}
q(\boldsymbol{\Sigma}|\bar{\boldsymbol{x}},\boldsymbol{V})_W \;=\;& \frac{2^{\frac{p(-m+ns)}{2}+1}\left(\frac{v_0}{2}\right)^{\frac{v_0}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}(m-p-1-ns)}|\boldsymbol{\Phi}|^{\frac{1}{2}(-m+ns)}}{B_{\frac{-m+ns}{2}}\left(\frac{1}{4}\boldsymbol{\Phi}^{-1}\boldsymbol{Y}\right)\Gamma\left(\frac{v_0}{2}\right)}\left[\frac{tr\left(\boldsymbol{\Sigma}\boldsymbol{\Phi}^{-1}\right)}{\left(tr\left[\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}\right]+2v_0\right)}\right]^{-\frac{p(m-sn-v_0)-1}{4}} \\
&\times K_{-\frac{p(m-sn-v_0)-1}{2}}\left(\sqrt{tr\left(\boldsymbol{\Sigma}\boldsymbol{\Phi}^{-1}\right)\left(tr\left[\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}\right]+2v_0\right)}\right).
\end{aligned}$$

From (23) and (21) the Bayes estimator of $|\boldsymbol{\Sigma}|$ is

$$\begin{aligned}
\widehat{|\boldsymbol{\Sigma}|}_{B,W} &= \frac{2^p B_{\frac{-m+ns-2}{2}}\left(\frac{1}{4}\boldsymbol{\Phi}^{-1}\boldsymbol{Y}\right)}{B_{\frac{-m+ns}{2}}\left(\frac{1}{4}\boldsymbol{\Phi}^{-1}\boldsymbol{Y}\right)}|\boldsymbol{\Phi}|\int_0^\infty z^p\frac{\left(\frac{v_0}{2}\right)^{\frac{v_0}{2}}z^{\frac{v_0}{2}-1}\exp(-v_0 z)}{\Gamma\left(\frac{v_0}{2}\right)}dz \\
&= \frac{2^{p-\frac{v_0}{2}}\Gamma\left(p+\frac{v_0}{2}\right)B_{\frac{-m+ns-2}{2}}\left(\frac{1}{4}\boldsymbol{\Phi}^{-1}\boldsymbol{Y}\right)}{v_0^p\Gamma\left(\frac{v_0}{2}\right)B_{\frac{-m+ns}{2}}\left(\frac{1}{4}\boldsymbol{\Phi}^{-1}\boldsymbol{Y}\right)}|\boldsymbol{\Phi}|.
\end{aligned}$$

The cumulative distribution function of $\lambda_{(p)}$, the largest eigenvalue of $\boldsymbol{\Sigma}$ from (18) is

$$\begin{aligned}
F_{\lambda_{(p)}}(a) &= \frac{2^{\frac{p}{2}(-m+ns)}a^{\frac{p}{2}(m-ns)}B_{\frac{-m+ns}{2}}\left(\frac{1}{4}\boldsymbol{\Phi}^{-1}\boldsymbol{Y};\boldsymbol{I}_p\right)}{B_{\frac{-m+ns}{2}}\left(\frac{1}{4}\boldsymbol{\Phi}^{-1}\boldsymbol{Y}\right)}|\boldsymbol{\Phi}|^{\frac{1}{2}(-m+ns)}\int_0^\infty z^{-\frac{p(m-sn)}{2}}w(z)dz \\
&= \frac{2^{\frac{p}{2}(-m+ns-v_0)}a^{\frac{p}{2}(m-ns)}v_0^{\frac{p(m-sn)}{2}}\Gamma\left(-\frac{p(m-sn)}{2}+\frac{v_0}{2}\right)B_{\frac{-m+ns}{2}}\left(\frac{1}{4}\boldsymbol{\Phi}^{-1}\boldsymbol{Y};\boldsymbol{I}_p\right)}{\Gamma\left(\frac{v_0}{2}\right)B_{\frac{-m+ns}{2}}\left(\frac{1}{4}\boldsymbol{\Phi}^{-1}\boldsymbol{Y}\right)}|\boldsymbol{\Phi}|^{\frac{1}{2}(-m+ns)}.
\end{aligned}$$

6. Applications

In this section we utilize the proposed results to justify the use of the normal-Wishart prior instead of the normal-inverse Wishart prior in some cases.

6.1. Simulation study

In this section a simulation study is done to calculate the new proposed estimators and to compare these estimators with the maximum likelihood estimators and the conventional Bayes estimators derived under the inverse Wishart prior. The Frobenius norm is used as a comparative measure of the bias of the different estimators. Some other interesting results of the posterior distributions of $\boldsymbol{\Sigma}$ are displayed and discussed.

6.1.1. Algorithms

Gibbs sampling (see Chapter 11 of Gelman, Carlin, Stern and Rubin, 1995) is used to simulate the posterior samples since the posterior density functions in (8) and (17) cannot be solved analytically to use simulation methods such as the inverse probability transform method.
The posterior distributions of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ with the normal-inverse Wishart prior is then simulated as follows:

Algorithm 1.

1. Initialize $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$
2. Repeat the following steps for $t = 1, \ldots, 100000$ times:
   a. Generate $\boldsymbol{\mu}_t \sim N_p(\boldsymbol{\theta}, \frac{1}{n_0}\boldsymbol{\Sigma}_{t-1})$.
   b. Calculate $\boldsymbol{D}_t = n_0(\boldsymbol{\mu}_t - \boldsymbol{\theta})(\boldsymbol{\mu}_t - \boldsymbol{\theta})'$ and $\boldsymbol{A}_t^* = \boldsymbol{V} + n(\overline{\boldsymbol{X}} - \boldsymbol{\mu}_t)(\overline{\boldsymbol{X}} - \boldsymbol{\mu}_t)'$.
   c. Generate the random matrix $\boldsymbol{\Sigma}_t \sim W^{-1}(\boldsymbol{A}_t^* + \boldsymbol{D}_t, p, s + ns + m_1)$.
3. Discard the first couple of observations, i.e. the posterior observations are $\boldsymbol{\mu}_{1000}, \ldots, \boldsymbol{\mu}_{100000}$ and $\boldsymbol{\Sigma}_{1000}, \ldots, \boldsymbol{\Sigma}_{100000}$.

The posterior distributions of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ with the normal-Wishart prior is then simulated by using Gibbs sampling with a Metropolis-Hastings (see Hastings, 1970) algorithm as follows:

Algorithm 2.

1. Initialize $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$
2. Repeat the following steps for $t = 1, \ldots, 100000$ times:
   a. Generate $\boldsymbol{\mu}_t \sim N_p(\boldsymbol{\theta}, \frac{1}{n_0}\boldsymbol{\Sigma}_{t-1})$
   b. Calculate $\boldsymbol{D}_t = n_0(\boldsymbol{\mu}_t - \boldsymbol{\theta})(\boldsymbol{\mu}_t - \boldsymbol{\theta})'$ and $\boldsymbol{A}_t^* = \boldsymbol{V} + n(\overline{\boldsymbol{X}} - \boldsymbol{\mu}_t)(\overline{\boldsymbol{X}} - \boldsymbol{\mu}_t)'$
   c. Metropolis-Hastings algorithm:
      i. Generate the random matrices $\boldsymbol{\Sigma}_1 \sim W(\boldsymbol{\Phi}_1, p, m_2)$ and $\boldsymbol{\Sigma}_2 \sim W^{-1}(\boldsymbol{\Phi}_2, p, m_2^*)$ such that $E[\boldsymbol{\Sigma}_1] = cE[\boldsymbol{\Sigma}_2]$.
      ii. Calculate $\boldsymbol{\Sigma}^* = w\boldsymbol{\Sigma}_1 + (1-w)\boldsymbol{\Sigma}_2$ for some $0 < w < 1$.
      iii. If $min\left(\frac{f^*[\boldsymbol{\Sigma}^*|\boldsymbol{\mu}_t]}{f^*[\boldsymbol{\Sigma}_{t-1}|\boldsymbol{\mu}_t]}, 1\right) > u$ where u is a random uniform$(0,1)$ variate, then $\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}^*$ else $\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}_{t-1}$, with $f^*[\boldsymbol{\Sigma}|\boldsymbol{\mu}_t] \propto |\boldsymbol{\Sigma}|^{-0.5(m_2-p-1-s-ns)} etr\left[-\frac{1}{2}\boldsymbol{\Sigma}\boldsymbol{\Phi}^{-1}\right] etr\left[-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{D}_t + \boldsymbol{A}_t^*\right)\right]$.
3. Discard the first couple of observations, i.e. the posterior observations are $\boldsymbol{\mu}_{1000}, \ldots, \boldsymbol{\mu}_{100000}$ and $\boldsymbol{\Sigma}_{1000}, \ldots, \boldsymbol{\Sigma}_{100000}$.

Remark 9 The value of c in Algorithm 2 will determine the efficiency of the algorithm, if it is chosen to be close to 1 the algorithm will be more efficient.

6.1.2. Results

A sample of size n is simulated from a multivariate normal $(s=1)$ distribution with a dimensionality of p and a zero mean and identity covariance matrix, i.e. $\boldsymbol{X}_{n\times p} \sim N_p(\boldsymbol{0}_p, \boldsymbol{I}_p)$. The priors are $\boldsymbol{\mu}|\boldsymbol{\Sigma} \sim N_p(\boldsymbol{\theta}, \frac{1}{n_0}\boldsymbol{\Sigma})$ and $\boldsymbol{\Sigma} \sim W^{-1}(\boldsymbol{\Phi}, p, m_1)$ and $\boldsymbol{\Sigma} \sim W(\boldsymbol{\Phi}, p, m_2)$, respectively, with $\boldsymbol{\theta} = 0.5 \times \boldsymbol{1}_p, \boldsymbol{\Phi} = 4 \times \boldsymbol{I}_p, \boldsymbol{\Phi}_1 = \boldsymbol{\Phi}_2 = \boldsymbol{\Phi}$, $n = 5, n_0 = 1, p = 3, m_1 = 9.5$, $m_2 = m_2^* = 3, w = 0.5$.

Note that the hyperparameters are assumed to be known and the degrees of freedom of the priors are chosen such that the priors have the same first moment. An empirical Bayes approach can be used as an alternative estimation method of the hyperparameters. The Bayes estimators under the squared error loss (PM estimators) of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, respectively, for the two priors are:

$$\widehat{\boldsymbol{\mu}}_{SEL,IW} = \begin{bmatrix} -0.42 & -0.33 & -0.26 \end{bmatrix}, \widehat{\boldsymbol{\mu}}_{SEL,W} = \begin{bmatrix} -0.41 & -0.33 & -0.26 \end{bmatrix}$$

$$\widehat{\boldsymbol{\Sigma}}_{SEL,IW} = \begin{bmatrix} 0.99 & 0.01 & 0.01 \\ 0.01 & 0.99 & 0 \\ 0.01 & 0 & 1 \end{bmatrix} \text{ and } \widehat{\boldsymbol{\Sigma}}_{SEL,W} = \begin{bmatrix} 1.04 & 0.41 & 0.19 \\ 0.41 & 0.98 & 0.19 \\ 0.19 & 0.19 & 1 \end{bmatrix}.$$

The MAP estimators (see Theorem 1 of Das and Dey, 2010) of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$,respectively, for the two priors are:

$$\widehat{\boldsymbol{\mu}}_{MAP,IW} = \begin{bmatrix} 0.01 & 0.03 & 0.23 \end{bmatrix}, \widehat{\boldsymbol{\mu}}_{MAP,W} = \begin{bmatrix} 0.05 & 0.19 & 0.35 \end{bmatrix}$$

$$\widehat{\boldsymbol{\Sigma}}_{MAP,IW} = \begin{bmatrix} 0.27 & -0.07 & -0.03 \\ -0.07 & 0.36 & 0.04 \\ -0.03 & -0.04 & 0.56 \end{bmatrix} \text{ and } \widehat{\boldsymbol{\Sigma}}_{MAP,W} = \begin{bmatrix} 0.72 & 0.35 & 0.29 \\ 0.35 & 0.54 & 0.1 \\ 0.29 & 0.1 & 0.34 \end{bmatrix}.$$

The sample estimates are given by:

$$\widehat{\boldsymbol{\mu}} = \bar{\boldsymbol{x}} = \begin{bmatrix} -0.24 & -0.01 & 0.23 \end{bmatrix} \text{ and } \widehat{\boldsymbol{\Sigma}} = \boldsymbol{S} = \begin{bmatrix} 0.78 & 0.49 & -0.24 \\ 0.49 & 0.66 & -0.37 \\ -0.24 & -0.37 & 0.73 \end{bmatrix}.$$

The Frobenius norm is used as a measure of closeness of the various estimates to the true parameter value and is defined as follows

$$||\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}||_F = \sqrt{tr\left(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\right)'\left(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\right)}.$$

It is known that

$$\lim_{r\to\infty} ||\left(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\right)^r||_F^{\frac{1}{r}} = \rho\left(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\right)$$

where $\rho\left(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\right)$ is the spectral radius. In Figures 1 and 2, respectively, $||\left(\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}\right)^r||_F^{\frac{1}{r}}$ and $||\left(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\right)^r||_F^{\frac{1}{r}}$ is plotted against r.

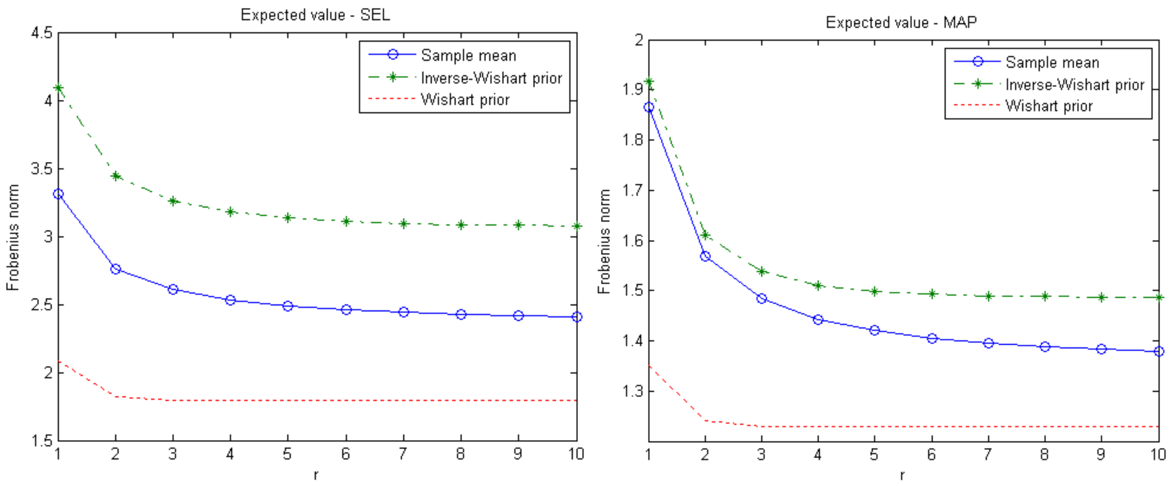

Figure 1: Frobenius norm for $\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}$ for the PM (left) and MAP (right) estimators.

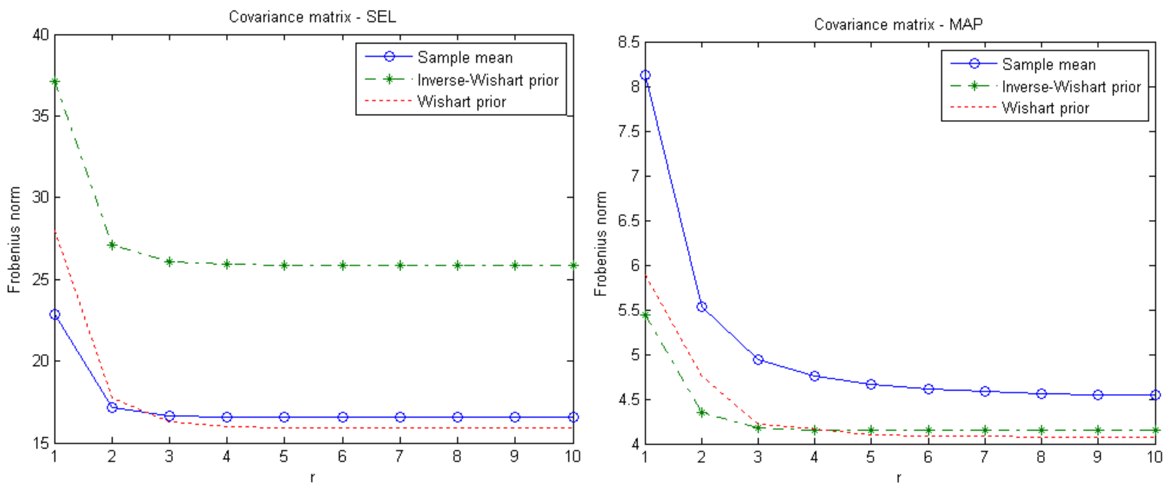

Figure 2: Frobenius norm for $\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}$ for the PM (left) and MAP (right) estimators.

It is clear that the lowest value of the Frobenius norm corresponds to the estimator with the smallest bias. In Figures 1 and 2 it is evident under the squared error loss function and the loss function in Remark 4, respectively, that the Bayes estimator from the Wishart prior is superior to the Bayes estimator from the inverse-Wishart prior and also the sample estimate.

Some confidence intervals for $|\mathbf{\Sigma}|$ were calculated for various parameter values using a bootstrap Jackknife method and are given in Table 1.

Table 1: 90% confidence intervals, $p = 3$, $n = 5$, $s = 1$ and $w = 0.2$.

		$m_1 = 7, m_2 = 7$	$m_1 = 10, m_2 = 5$	$m_1 = 50, m_2 = 4$
$n_0 = 1$	Wishart prior	$(1.18; 13.16)$	$(0.14; 2.5)$	$(0.04; 0.61)$
	Inverse-Wishart prior	$(0.17; 1.97)$	$(0.22; 1.93)$	$(0.52; 1.58)$
	Jeffreys prior	$(0.05; 0.47)$	$(0.05; 0.47)$	$(0.05; 0.47)$
$n_0 = 10$	Wishart prior	$(1.38; 13.48)$	$(0.23; 3.28)$	$(0.09; 1.05)$
	Inverse-Wishart prior	$(0.17; 1.92)$	$(0.22; 1.98)$	$(0.52; 1.57)$
	Jeffreys prior	$(0.05; 0.48)$	$(0.05; 0.48)$	$(0.05; 0.48)$

The Jeffreys prior (see Section 3.1. of Sun and Berger, 2006) is also included for comparison purposes. From Table 1 it is evident that the performance of the Jeffreys prior is unsatisfactory since the true value of $|\mathbf{\Sigma}| = 1$ is not contained in the interval. Concerning the two subjective priors, no one prior is superior to the other, and the choice of prior is left to the practitioner. However, for demonstrating our assertion regarding the superiority of the normal-Wishart as the prior distribution we consider the performance of the ratio of the two largest eigenvalues.

To be more specific, Figure 3 illustrates the cumulative distribution function of the ratio of the two largest eigenvalues of the posterior distribution of $\mathbf{\Sigma}$ for the inverse-Wishart and Wishart priors respectively,

$$\lambda_r = \frac{\lambda_W}{\lambda_{IW}}.$$

If this ratio is larger than one, we conclude that the Wishart prior leads to larger posterior eigenvalues which in turn will result in fewer principal components to explain a larger portion of the variation in the data.

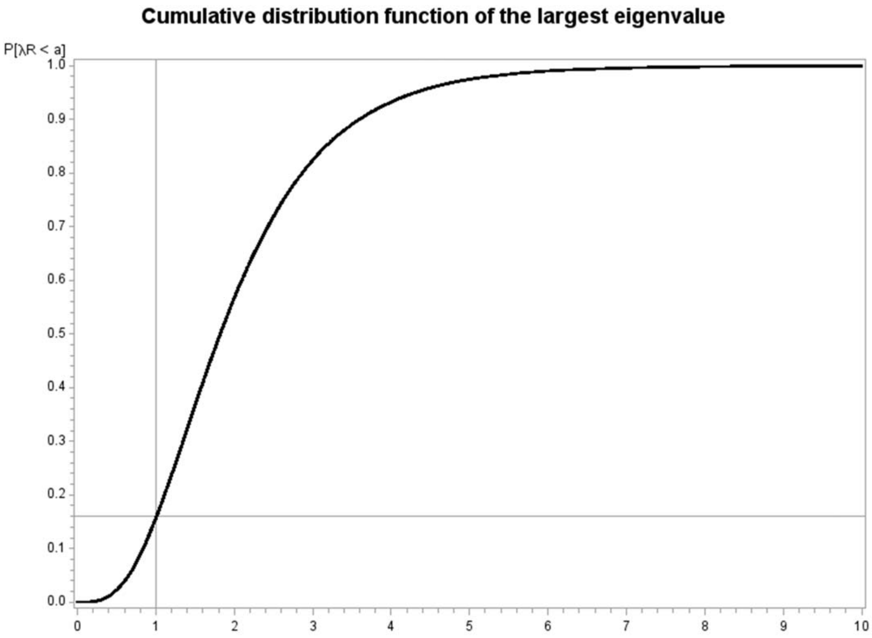

Figure 3: Cumulative distribution function of the ratio of the largest eigenvalue of the Wishart and inverse-Wishart priors, $p = 3$, $m_1 = 4$, $m_2 = 7$, $n = 5$, $n_0 = 1$, $s = 1$ and $w = 0.5$.

It is evident from Figure 3 that $P[\lambda_r < 1] = 0.16$ and hence $P[\lambda_r > 1] = 0.84$, so that when the Wishart prior is applied it leads to fewer principal components included in the model with probability 0.84. It is for this reason that the Wishart prior is preferred over the widely used and accepted inverse-Wishart prior.

6.2. Fisher's Iris dataset

The Iris dataset is a very well-known dataset of dimension four and a sample size of 150. The data set was introduced by Sir Ronald Fisher (1936) as an example of discriminant analysis. The data set consists of 50 samples from each of three species of Iris (Iris setosa, Iris virginica and Iris versicolor). Four features were measured from each sample: the length and the width of the sepals and petals, in centimeters. For the purpose of applying the results to this dataset only a subset of the data is used — a subset of the *Iris setosa* subspecies is used, and hence the sample size is 10. The reason for this is that each of the subspecies' measures have been shown to follow a four dimensional multivariate normal distribution.

Let $\boldsymbol{X}_{50\times 4}$ = sample of *Iris setosa* measurements $\sim N_4(\boldsymbol{\mu},\boldsymbol{\Sigma})$ with $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ unknown. Assume the prior distributions are $\boldsymbol{\mu}|\boldsymbol{\Sigma} \sim N_p(\boldsymbol{\theta},\boldsymbol{\Sigma})$ and $\boldsymbol{\Sigma} \sim W^{-1}(\boldsymbol{\Phi},4,5)$ and $\boldsymbol{\Sigma} \sim W(\boldsymbol{\Phi},4,9)$, respectively, and as illustration $\boldsymbol{\theta} = \overline{\boldsymbol{X}}$ and $\boldsymbol{\Phi} = \boldsymbol{S} = \frac{1}{50}\sum_{i=1}^{50}(\boldsymbol{X}_i - \overline{\boldsymbol{X}})(\boldsymbol{X}_i - \overline{\boldsymbol{X}})'$.

The posterior distributions are simulated by using algorithm 1 as in Subsection 6.1. The cumulative distribution function of the largest eigenvalue of the posterior distribution of $\mathbf{\Sigma}$ is given in Figure 4 for the normal-inverse Wishart and the normal Wishart prior.

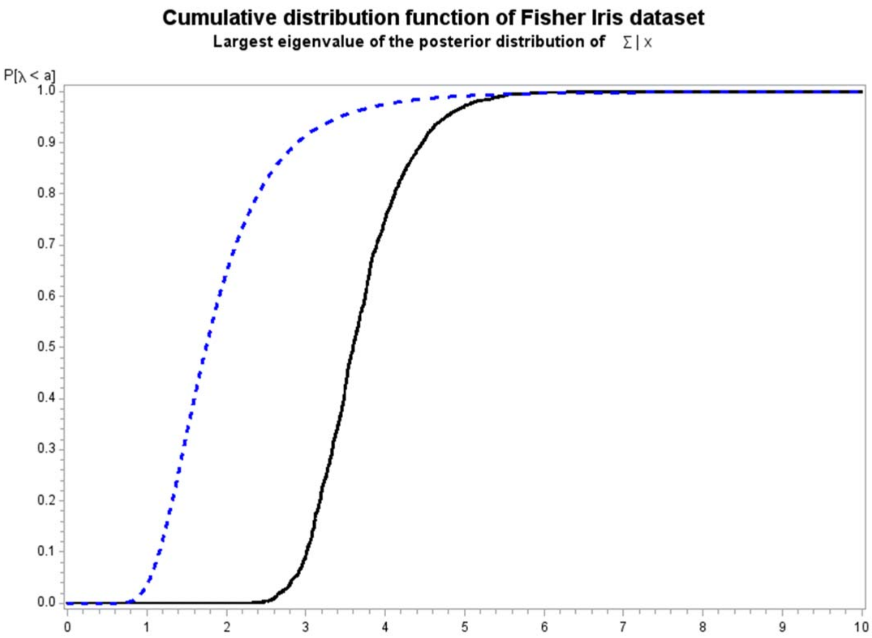

Figure 4: The cumulative distribution function of the largest eigenvalue of $\mathbf{\Sigma}$ under prior 4 (- - -) and 15 (——), $p = 4$, $m_1 = 5$, $m_2 = 9$, $n = 10$, $n_0 = 10$, $s = 1$.

7. Conclusion

Our main contributions in this paper are summarized as follows:

- New estimators for the matrix parameters of the matrix variate elliptical model was proposed from a subjective Bayesian viewpoint.
- The normal-inverse Wishart and normal-Wishart priors were considered for the location and scale matrices of the underlying model.
- The Bayes estimators of the parameters, under two loss functions, as well as the joint posterior density functions and marginal posterior density functions were derived, with the matrix variate normal and matrix variate t-distribution as particular subfamilies.
- The Bayes estimator under SEL of the location parameter is a robust estimator in the sense that it is independent of the prior distribution of the scale parameter.

- For both priors, the posterior distributions for the location and scale matrices of the matrix variate normal model were simulated, using the two new algorithms, the Bayes estimators under the two loss functions were calculated and compared using the Frobenius norm.

- The estimator derived under the normal-Wishart prior displayed superior performance in the simulation study and this justifies the use of a normal-Wishart prior in the Bayesian analysis of the normal model.

Acknowledgements

The authors would like to hereby acknowledge the support of the StatDisT group. This work is based upon research supported by the National Research foundation, Grant (Re:CPRR3090132066 No 91497).

References

ARASHI, M., IRANMANESH, A., NOROUZIRAD, M., AND SALARZADEH-JENATABADI, H. (2014). Bayesian analysis in multivariate regression models with conjugate priors. *Statistics*, **48** (6).

ARASHI, M., SALEH, A. K. M. E., AND TABATABAEY, S. M. M. (2013). Regression model with elliptically contoured errors. *Statistics*, **47** (6), 1266–1284.

BEKKER, A. AND ROUX, J. J. J. (1995). Bayesian multivariate normal analysis with a Wishart prior. *Communications in Statistics – Theory and Methods*, **24** (10), 2485 – 2497.

CHU, K. C. (1973). Estimation and decision for linear systems with elliptical random processes. *IEEE Transactions on Automatic Control*, **Ac18**, 499 – 505.

DAS, S. AND DEY, D. K. (2010). On Bayesian inference for multivariate generalized gamma distribution. *Statistics and Probability Letters*, **80** (19-20), 1492 – 1499.

ERDELYI, A., MAGNUS, W., OBERHETTINGER, F., AND TRICOMI, F. G. (1953). *Higher Transcendental Functions I*. McGraw-Hill Inc.: New York.

FANG, K. T. AND LI, R. Z. (1999). Bayesian statistical inference on elliptical matrix distributions. *Journal of Multivariate Analysis*, **70**, 66 – 85.

FANG, K. T. AND ZHANG, Y. (1990). *Generalized Multivariate Analysis*. Springer-Verlag and Science Press: Berlin and Beijing.

GELMAN, A., CARLIN, J. B., STERN, H. S., AND RUBIN, D. B. (1995). *Bayesian Data Analysis*. Chapman and Hall: London.

GRADSHTEYN, I. S. AND RYZHIK, I. M. (2007). *Table of Integrals, Series and Products*. Academic Press: San Diego.

GUPTA, A. K. AND NAGAR, D. K. (2000). *Matrix Variate Distributions*. Chapman and Hall: Boca Raton, FL.

HASTINGS, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, **57** (1), 97 – 109.

HERZ, C. Z. (1955). Bessel functions of matrix argument. *Annals of Mathematics*, **61**, 474 – 523.

SUN, D. AND BERGER, J. (2006). Objective bayesian analysis for the multivariate normal model. *In Proc. Valencia / ISBA 8th World Meeting on Bayesian Statistics*. Alicante, Spain.

VAN NIEKERK, J., BEKKER, A., ROUX, J. J. J., AND ARASHI, M. (2013). Subjective Bayesian analysis for the elliptical model. *Communications in Statistics – Theory and Methods*, **in Press**.

Appendices

Appendix A and B can be found at `http://www.up.ac.za/media/shared/115/appendix.zp52697.pdf`.

Manuscript received, 2015-06-22, revised, 2015-09-07, accepted, 2015-09-09.