Journal of Multivariate Analysis **70**, 66–85 (1999)
Article ID jmva.1998.1816, available online at http://www.idealibrary.com on IDEAL®

Bayesian Statistical Inference on Elliptical Matrix Distributions*

Kai-Tai Fang

Hong Kong Baptist University and Institute of Applied Mathematics, The Chinese Academy of Sciences, Beijing, People's Republic of China

and

Runze Li

University of North Carolina at Chapel Hill

Received July 21, 1997

In this paper we are concerned with Bayesian statistical inference for a class of elliptical distributions with parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. Under a noninformative prior distribution, we obtain the posterior distribution, posterior mean, and generalized maximim likelihood estimators of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. Under the entropy loss and quadratic loss, the best Bayesian estimators of $\boldsymbol{\Sigma}$ are derived as well. Some applications are given. © 1999 Academic Press

AMS 1991 subject classifications: 62C10, 62F10, 62F15.

Key words and phrases: elliptical matrix distributions; entropy loss; posterior mean; quadratic loss.

1. INTRODUCTION

Statistical decision theory and Bayesian statistical inference are related at a number of levels. Many statistical decision problems can be successfully solved by Bayesian methods (see Berger, 1985). In classical multivariate analysis, Bayesian methods are often adapted to prove the admissibility, minimaxicity and invariance of estimators. Press (1982) systematically applied the Bayesian methods to classical multivariate analysis. However, it is valuable to implement Bayesian methods to *generalized multivariate analysis*, multivariate analysis based on elliptical matrix distributions (Fang and Zhang, 1990). Bayesian statistical inference in generalized multivariate analysis will be developed in this paper.

* This work was supported by a Hong Kong RGC grant, the Statistical Ressearch and Consultancy Centre, Hong Kong Baptist University.

0047-259X/99 $30.00

There are many references on Bayesian inference for linear and nonlinear regression models with non-normal error terms. Zellner (1976) considered the Bayesian analysis of regression models with multivariate Student-t error terms, and Singh, Misra and Pandey (1995) studied a generalized class of estimators in linear regression models with multivariate Student-t distributed error term. Jammalamadaka, Tiwari and Chib (1987) and Chib, Tiwari, and Jammalamadaka (1988) studied this model with errors being distributed according to a scale mixtures of normal distribution. Osiewalski (1991) and Chib, Osiewalski, and Steel (1991) investigated Bayesian inference for nonlinear regression models with scale mixtures of normally distributed errors. Osiewalski and Steel (1993a) gave a generalization to linear regression model whose error terms have a multivariate elliptical distribution. Later Osiewalski and Steel (1993b) studied robust Bayesian statistical inference for l_q-spherical models. When $q=2$, the l_q-spherical distribution reduces to a spherical distribution. Fernández, Osiewalski and Steel (1994) investigated the robustness of Bayesian estimators of location parameters in continuous multivariate location-scale models. Their study contains the case of the multivariate spherical distributions with unknown location parameters.

Let $\mathbf{X}$ be an $n \times p$ random matrix, which can be expressed in terms of its elements, columns and rows as

$$\mathbf{X} = (x_{ij}) = (\mathbf{x}_1, ..., \mathbf{x}_p) = (\mathbf{x}_{(1)}, ..., \mathbf{x}_{(n)})'. \tag{1.1}$$

Here $\mathbf{x}_{(1)}, ..., \mathbf{x}_{(n)}$ can be regarded as a sample of size n from a p-dimensional population. When $\mathbf{x}_{(1)}, ..., \mathbf{x}_{(n)}$ are independently and identically distributed according to $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, the matrix $\mathbf{X}$ is called to have a matrix variate normal distribution and is denoted by $\mathbf{X} \sim N_{n \times p}(\mathbf{1}_n \boldsymbol{\mu}', \mathbf{I}_n \otimes \boldsymbol{\Sigma})$. In this paper we always assume that the first two moments of $\mathbf{X}$ exist. Denote

$$\bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_{(i)}, \tag{1.2}$$

and

$$\mathbf{S} = \sum_{i=1}^{n} (\mathbf{x}_{(i)} - \bar{\mathbf{x}})(\mathbf{x}_{(i)} - \bar{\mathbf{x}})'. \tag{1.3}$$

In the context of classical multivariate analysis, the samples are independently and identically distributed according to a normal distribution. However the covariance of samples may be different in heteroscedastic models. In this situation, we may assume that given σ, the samples are independently and identically distributed according to $N_p(\boldsymbol{\mu}, \sigma^2\boldsymbol{\Sigma})$ and regard σ as a nuisance parameter. In Bayesian context, we may further assume that σ has a prior distribution $H(\sigma)$, and consequently $\mathbf{x}_{(1)}, ..., \mathbf{x}_{(n)}$ have the joint density

$$\int (2\pi\sigma^2)^{-(np/2)} \, |\boldsymbol{\Sigma}|^{-n/2} \exp\left\{-\frac{1}{2\sigma^2} \operatorname{tr} \boldsymbol{\Sigma}^{-1}(\mathbf{X}-\mathbf{1}\boldsymbol{\mu}')' \, (\mathbf{X}-\mathbf{1}\boldsymbol{\mu}')\right\} dH(\sigma)$$

which is a scale mixtures of normal distributions. The class of matrix Student-t distributions is a subclass of scale mixtures of normal distributions, in which $1/\sigma^2$ is distributed according to a Gamma distribution. Although $\mathbf{x}_{(1)}, ..., \mathbf{x}_{(n)}$ are independent samples for given σ, statistical inference in fact is based on the (marginal) distribution of the random matrix $\mathbf{X}$, whose rows are not independent. From this point of view, we may assume that the samples, the rows of the random matrix $\mathbf{X}$, are dependent in the context of generalized multivariate analysis. There are studies on the topic of statistical inference based on independently and identially distributed samples from a multivariate elliptical distribution. The derived results are usually large sample properties. In practice, when a large samples cannot be obtained, statistical inference based on elliptical matrix distributions would be preferred (cf. Dawid, 1977, and Anderson, 1993).

Based on multivariate normal samples, one can easily construct a random matrix having a spherical matrix distribution. Indeed, if the rows $\mathbf{x}_{(i)}$ of $\mathbf{X}$ are independent samples from a normal distribution $N_p(\mathbf{0}, \boldsymbol{\Sigma})$, and $\mathbf{D}(\mathbf{X}'\mathbf{X})$ is a $p \times q$ projection direction matrix uniquely determined by $\mathbf{X}'\mathbf{X}$, then $\mathbf{XD}(\mathbf{X}'\mathbf{X})$ has a left-spherical distribution, as $\mathbf{X}$ is left-spherical (see below for definition). On the other hand, if $\mathbf{X} \sim N_{n\times p}(\mathbf{1}_n\boldsymbol{\mu}', \mathbf{I}_n \otimes \boldsymbol{\Sigma})$, we can choose a row-orthogonal $(n-1)\times n$ constant matrix $\mathbf{A}$ such that $\mathbf{A}\mathbf{1}_n = \mathbf{0}$, and $\mathbf{Y} = \mathbf{AX}$ is distributed according to $N_{(n-1)\times p}(\mathbf{0}, \mathbf{I}_{n-1} \otimes \boldsymbol{\Sigma})$. As a result $\mathbf{YD}(\mathbf{Y}'\mathbf{Y})$ has a left-spherical distribution. Läuter (1996) and Läuter, Glimm and Kropf (1996) pioneered the use of these facts and constructed some powerful exact t- and F-tests for normal means. Later Fang, Li and Liang (1998) and Liang (1998) proposed some tests for multinormality based on these facts. From the references mentioned above, it can be seen that the theory of spherical and elliptical matrix distributions is useful in both distribution theory and other statistical branches.

There are several ways to define elliptical matrix distributions. Four classes of elliptical matrix distributions are defined and discussed by Dawid (1977) and Anderson and Fang (1990a, 1990b). Let us start with left-spherical matrix distributions $LS_{n\times p}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$, the largest class of elliptical matrix distributions among the four classes of elliptical matrix distributions studied in Fang and Zhang (1990). The notation $\mathbf{X} \sim LS_{n\times p}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$

means that $\mathbf{X}$ is distributed according to an elliptical matrix distribution with the probability density function (pdf)

$$d_{n,p}\,|\boldsymbol{\Sigma}|^{-n/2}\,g(\boldsymbol{\Sigma}^{-1/2}(\mathbf{X}-\mathbf{1}\boldsymbol{\mu}')'\,(\mathbf{X}-\mathbf{1}\boldsymbol{\mu}')\,\boldsymbol{\Sigma}^{-1/2}),$$

where g is a real function. Throughout this exposition $d_{n,p}$ will denote the normalizing constant, and $\mathbf{1}'=(1,...,1)_{1\times n}$. When implementing Bayesian methods for $LS_{n\times p}(\boldsymbol{\mu},\boldsymbol{\Sigma},g)$, one may be interested in situations where $\mathbf{X}$ has a pdf of the form

$$d_{n,p}\,|\boldsymbol{\Sigma}|^{-n/2}\,g(\boldsymbol{\Sigma}^{-1}(\mathbf{X}-\mathbf{1}\boldsymbol{\mu}')'\,(\mathbf{X}-\mathbf{1}\boldsymbol{\mu}')).$$

This is equivalent to impose the following condition on $g: g(\mathbf{AB})=g(\mathbf{BA})$ for any $p\times p$ positive definite symmetric matrices $\mathbf{A}$ and $\mathbf{B}$. Through straightforward linear algebra calculations, it can be proved that this condition is equivalent to that $g(\mathbf{A})$ depends on $\mathbf{A}$ only through its eigenvalues, and therefore Herz (1955) called this class of functions as *symmetric* functions. In this case the function $g(\mathbf{A})$ can be expressed as $g(\lambda(\mathbf{A}))$, where $\lambda(\mathbf{A})=\mathrm{diag}(\lambda_1,...,\lambda_p)$ are the eigenvalues of $\mathbf{A}$, and the density of $\mathbf{X}$ becomes

$$d_{n,p}\,|\boldsymbol{\Sigma}|^{-n/2}\,g(\lambda(\boldsymbol{\Sigma}^{-1}(X-\mathbf{1}\boldsymbol{\mu}')'\,(\mathbf{X}-\mathbf{1}\boldsymbol{\mu}'))). \tag{1.4}$$

For convenience, we write $\mathbf{X}\sim SS_{n\times p}(\boldsymbol{\mu},\boldsymbol{\Sigma},g)$, where g is a real function and does not depend on n in the rest of the paper. For simplicity, we shall use the notation $g_\lambda(\cdot)$ to replace $g(\lambda(\cdot))$. If $\mathbf{X}\sim SS_{n\times p}(\boldsymbol{\mu},\boldsymbol{\Sigma},g)$, then the characteristic function of $\mathbf{X}$ has the form $\exp(i\mathbf{T}'\mathbf{1}\boldsymbol{\mu}')\,\phi(\lambda(\boldsymbol{\Sigma}^{-1}\mathbf{T}'\mathbf{T}))$, where $\phi(\cdot)$ is called the *characteristic generator of* $\mathbf{X}$.

Let

$$\begin{aligned}\mathscr{F}=\{|\boldsymbol{\Sigma}|^{-n/2}\,g_\lambda(\boldsymbol{\Sigma}^{-1}\,(\mathbf{X}-\mathbf{1}\boldsymbol{\mu}')'\,(\mathbf{X}-\mathbf{1}\boldsymbol{\mu}'))\mid g_\lambda(\mathbf{X}'\mathbf{X})\\ \text{is a density and } \boldsymbol{\mu}\in R^p,\ \boldsymbol{\Sigma}>0\}\end{aligned}$$

be a class of densities. It is known that $(\bar{\mathbf{x}},\mathbf{S})$ is a complete sufficient statistic for the family $\mathscr{F}$, and $\mathbf{S}$ is positive definite with probability one if and only if $n>p$ (see Fang and Zhang, 1990, Sections 4.1 and 4.3). Therefore, throughout this paper we assume that $n>p$.

In this paper we take a noninformative prior distribution for $(\boldsymbol{\mu},\boldsymbol{\Sigma}):\pi(\boldsymbol{\mu},\boldsymbol{\Sigma})\ \propto\ |\boldsymbol{\Sigma}|^{-(p+1)/2}$. In Section 2, the posterior distribution and marginal distributions of $(\boldsymbol{\mu},\boldsymbol{\Sigma})$, the posterior mean and generalized maximum likelihood estimators (GMLE) of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, the value which maximizes the posterior distribution $\pi(\boldsymbol{\mu},\boldsymbol{\Sigma}\mid\mathbf{X})$ with respect to $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, are derived. The confidence regions for $\boldsymbol{\mu}$ are obtained. In Section 3, the best Bayesian estimators of $\boldsymbol{\Sigma}$ under both entropy and quadratic losses are studied. Section 4 gives applications of the theory developed in previous sections.

In the following sections, $|\mathbf{A}|$, $\mathrm{tr}(\mathbf{A})$ and $\lambda(\mathbf{A}) = \mathrm{diag}(\lambda_1, ..., \lambda_p)$, where $\lambda_1 \geqslant \cdots \geqslant \lambda_p$ will denote the determinant, trace and eigenvalues of $\mathbf{A}$, respectively. The t-distribution with m degrees of freedom is denoted by t_m, and the F-distribution with n and m degrees of freedom is denoted by $F_{n,m}$. A multivariate gamma function will be denoted by $\Gamma_p(\cdot)$, where $\Gamma_p(n/2) = \pi^{(p-1)p/2} \prod_{i=1}^n \Gamma((n-i+1)/2)$.

2. THE BAYESIAN STATISTICAL INFERENCE FOR $\boldsymbol{\mu}$

In this section we will derive the posterior distribution and marginal distributions of $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. The posterior means of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ will be obtained as well.

Suppose that $\mathbf{X} \sim SS_{n\times p}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$. It follows from Theorem 4.2.1 of Fang and Zhang (1990) that the conditional pdf of $(\bar{\mathbf{x}}, \mathbf{S})$ given $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ is

$$f(\bar{\mathbf{x}}, \mathbf{S} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) = c_{n,p} \, |\mathbf{S}|^{(n-p)/2-1} \, |\boldsymbol{\Sigma}|^{-n/2} \, g_\lambda(\boldsymbol{\Sigma}^{-1}(\mathbf{S} + n(\bar{\mathbf{x}} - \boldsymbol{\mu})(\bar{\mathbf{x}} - \boldsymbol{\mu})')), \tag{2.1}$$

where $c_{n,p} = n^{p/2} \pi^{(n-1)p/2} [\Gamma_p((n-1)/2)]^{-1} d_{n,p}$. The choice for the prior distribution of $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ will be the noninformative prior distribution, i.e.,

$$\pi(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-(p+1)/2} (d\boldsymbol{\mu})(d\boldsymbol{\Sigma}), \tag{2.2}$$

where the notation $\propto$ means "to be proportional to."

THEOREM 2.1. *Suppose that* $\mathbf{X} \sim SS_{n\times p}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$ *and the prior distribution of* $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *is* (2.2). *Then*

(a) *The posterior pdf of* $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *is*

$$f(\boldsymbol{\mu}, \boldsymbol{\Sigma} \mid \bar{\mathbf{x}}, \mathbf{S}) = \lambda_{n,p} \left(\frac{n}{\pi}\right)^{p/2} \Gamma(n/2)[\Gamma((n-p)/2)]^{-1} \, |\mathbf{S}|^{(n-1)/2} \, |\boldsymbol{\Sigma}|^{-(n+p+1)/2}$$
$$\times g_\lambda(\boldsymbol{\Sigma}^{-1}(\mathbf{S} + n(\bar{\mathbf{x}} - \boldsymbol{\mu})(\bar{\mathbf{x}} - \boldsymbol{\mu})')), \tag{2.3}$$

where $\lambda_{n,p} = \pi^{np/2} [\Gamma_p(n/2)]^{-1} d_{n,p}$.

(b) The posterior density of $\boldsymbol{\mu}$ is

$$f(\boldsymbol{\mu} \mid \bar{\mathbf{x}}, \mathbf{S}) = \frac{\Gamma(n/2)}{\pi^{p/2} \Gamma((n-p)/2)} \left|\frac{1}{n}\mathbf{S}\right|^{-1/2} \left(1 + (\bar{\mathbf{x}} - \boldsymbol{\mu})' \left(\frac{1}{n}\mathbf{S}\right)^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu})\right)^{-n/2} \tag{2.4}$$

which is a multivariate Pearson type VII distribution $MPVII_p(\bar{\mathbf{x}}, (1/n)\mathbf{S}, g_{N,m})$ with $N = n/2$ and $m = 1$ (see Fang, Kotz, and Ng, 1990, p. 83).

(c) The posterior means of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}^{-1}$ are

$$E(\boldsymbol{\mu} \mid \bar{\mathbf{x}}, \mathbf{S}) = \bar{\mathbf{x}},$$

$$E(\boldsymbol{\Sigma}^{-1} \mid \bar{\mathbf{x}}, \mathbf{S}) = \frac{n-p-1}{n-p}\, n(-2\phi'(0))\, \mathbf{S}^{-1}, \tag{2.5}$$

respectively, where $\phi(\cdot)$ is the characteristic generator of $\mathbf{X}$.

Remark 2.1. As mentioned in the previous section, $\mathbf{S}$ is positive definite with probability one if and only if $n > p$, which is also a necessary and sufficient condition for the existence of the posterior density in (2.4). The class of multivariate Pearson VII distributions has been studied in detail in Fang, Kotz, and Ng (1990). Comparing (2.4) to the multivariate Students-t distribution, one can easily deduce that (2.4) only has moments up to order $n-p$ exclusively.

Before proving this theorem we need two lemmas which are also useful for other results considered in this paper.

LEMMA 2.1. *Let*

$$h(\boldsymbol{\Sigma}) = \int_{\mathbf{G}>0} |\mathbf{G}|^{(n-p-1)/2}\, g_\lambda(\boldsymbol{\Sigma}^{-1}\mathbf{G}))\, d\mathbf{G}, \tag{2.6}$$

where g_λ is given by (1.4) *and* $\mathbf{G} > 0$ *means that* $\mathbf{G}$ *is positive definite. Then*

$$h(\boldsymbol{\Sigma}) = \lambda_{n,p}^{-1}\, |\boldsymbol{\Sigma}|^{n/2}, \tag{2.7}$$

where $\lambda_{n,p}$ is given in Theorem 2.1.

Proof. Suppose that $\mathbf{X}$ has a density of the form (1.4). Let $\mathbf{G} \triangleq (\mathbf{X} - \mathbf{1}\boldsymbol{\mu}')'\,(\mathbf{X} - \mathbf{1}\boldsymbol{\mu}')$. Then the pdf of $\mathbf{G}$ is

$$\lambda_{n,p}\, |\boldsymbol{\Sigma}|^{-n/2}\, |\mathbf{G}|^{(n-p-1)/2}\, g_\lambda(\boldsymbol{\Sigma}^{-1}\mathbf{G})$$

and (2.7) follows. ∎

LEMMA 2.2. *In the previous notation we have that*

$$\int (1+n(\bar{\mathbf{x}}-\boldsymbol{\mu})' \mathbf{S}^{-1}(\bar{\mathbf{x}}-\boldsymbol{\mu}))^{-n/2}\, d\boldsymbol{\mu} = |\mathbf{S}|^{1/2} \frac{\pi^{p/2}\Gamma((n-p)/2)}{n^{p/2}\Gamma(n/2)}, \tag{2.8}$$

and

$$\int |\mathbf{S}+n(\bar{\mathbf{x}}-\boldsymbol{\mu})(\bar{\mathbf{x}}-\boldsymbol{\mu})'|^{-n/2}\, d\boldsymbol{\mu} = |\mathbf{S}|^{-(n-1)/2} \frac{\pi^{p/2}\Gamma((n-p)/2)}{n^{p/2}\Gamma(n/2)}. \tag{2.9}$$

Proof. Let $\mathbf{y}=\boldsymbol{\mu}-\bar{\mathbf{x}}$. The left hand side of (2.8) becomes

$$\begin{aligned}\int (1+n\mathbf{y}'\mathbf{S}^{-1}\mathbf{y})^{-n/2}\, d\mathbf{y} &= |\mathbf{S}|^{1/2} \int (1+n\mathbf{z}'\mathbf{z})^{-n/2}\, d\mathbf{z}\\ &= \frac{|\mathbf{S}|^{1/2}}{n^{p/2}} \int (1+\mathbf{u}'\mathbf{u})^{-n/2}\, d\mathbf{u}.\end{aligned}$$

By formula (1.32) of Fang, Kotz, and Ng (1990), the latter becomes

$$\frac{|\mathbf{S}|^{1/2}\,\pi^{p/2}}{n^{p/2}\Gamma(p/2)} \int_0^\infty y^{p/2-1}(1+y)^{-n/2}\, dy = \frac{|\mathbf{S}|^{1/2}\,\pi^{p/2}}{n^{p/2}\Gamma(p/2)}\, B\left(\frac{p}{2}, \frac{n-p}{2}\right)$$

which implies (2.8). The formula (2.9) is a consequence of (2.8). ∎

Proof of Theorem 2.1. By the Bayes formula we have that

$$f(\boldsymbol{\mu}, \boldsymbol{\Sigma} \mid \bar{\mathbf{x}}, \mathbf{S}) = \frac{f(\bar{\mathbf{x}}, \mathbf{S} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})\, \pi(\boldsymbol{\mu}, \boldsymbol{\Sigma})}{\int f(\bar{\mathbf{x}}, \mathbf{S} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})\, \pi(\boldsymbol{\mu}, \boldsymbol{\Sigma})\, d\boldsymbol{\mu}\, d\boldsymbol{\Sigma}} \tag{2.10}$$

whose denominator can be simplified as

$$\begin{aligned}&\int f(\bar{\mathbf{x}}, \mathbf{S} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})\, \pi(\boldsymbol{\mu}, \boldsymbol{\Sigma})\, d\boldsymbol{\mu}\, d\boldsymbol{\Sigma}\\ &= c_{n,p}\, |\mathbf{S}|^{(n-p)/2-1} \int d\boldsymbol{\mu} \int |\boldsymbol{\Sigma}|^{-(n+p+1)/2}\, g_\lambda(\boldsymbol{\Sigma}^{-1}(\mathbf{S}+n(\bar{\mathbf{x}}-\boldsymbol{\mu})(\bar{\mathbf{x}}-\boldsymbol{\mu})'))\, d\boldsymbol{\Sigma}\\ &\qquad (\text{let } \mathbf{Y}=\boldsymbol{\Sigma}^{-1})\\ &= c_{n,p}\, |\mathbf{S}|^{(n-p)/2-1} \int d\boldsymbol{\mu} \int |\mathbf{Y}|^{(n-p-1)/2}\, g_\lambda(\mathbf{Y}(\mathbf{S}+n(\bar{\mathbf{x}}-\boldsymbol{\mu})(\bar{\mathbf{x}}-\boldsymbol{\mu})'))\, d\mathbf{Y}.\end{aligned}$$

By Lemmas 2.1 and 2.2, the latter becomes

$$c_{n,p}\,|\mathbf{S}|^{(n-p)/2-1}\int \lambda_{n,p}^{-1}\,|\mathbf{S}+n(\bar{\mathbf{x}}-\boldsymbol{\mu})(\bar{\mathbf{x}}-\boldsymbol{\mu})'|^{-n/2}\,d\boldsymbol{\mu}$$

$$=c_{n,p}\,|\mathbf{S}|^{(n-p)/2-1}\,\lambda_{n,p}^{-1}\,|\mathbf{S}|^{-(n-1)}\,\frac{\pi^{p/2}\Gamma((n-p)/2)}{n^{p/2}\Gamma(n/2)}$$

$$=\frac{c_{n,p}\pi^{p/2}\Gamma((n-p)2)}{\lambda_{n,p}n^{p/2}\Gamma(n/2)}\,|\mathbf{S}|^{-(p+1)/2}.$$

Using (2.3), (2.9), and (2.10), assertion (a) follows from the above result. From (a), assertion (b) follows by

$$\int_{\boldsymbol{\Sigma}>0}|\boldsymbol{\Sigma}|^{-(n+p+1)/2}\,g_\lambda(\boldsymbol{\Sigma}^{-1}(\mathbf{S}+n(\bar{\mathbf{x}}-\boldsymbol{\mu})(\bar{\mathbf{x}}-\boldsymbol{\mu})'))\,d\boldsymbol{\Sigma}\qquad(\text{let }\mathbf{Y}=\boldsymbol{\Sigma}^{-1})$$

$$=\int_{\mathbf{Y}>0}|\mathbf{Y}|^{(n-p-1)/2}\,g_\lambda(\mathbf{Y}(\mathbf{S}+n(\bar{\mathbf{x}}-\boldsymbol{\mu})(\bar{\mathbf{x}}-\boldsymbol{\mu})'))\,d\mathbf{Y}$$

$$=\frac{\Gamma_p(n/2)}{\pi^{np/2}}\,|\mathbf{S}+n(\bar{\mathbf{x}}-\boldsymbol{\mu})(\bar{\mathbf{x}}-\boldsymbol{\mu})'|^{-n/2}$$

$$=\frac{\Gamma_p(n/2)}{\pi^{np/2}}\,|\mathbf{S}|^{-n/2}\,(1+n(\bar{\mathbf{x}}-\boldsymbol{\mu})'\,\mathbf{S}^{-1}(\bar{\mathbf{x}}-\boldsymbol{\mu}))^{-n/2}.$$

For proving (c), write $l_{n,p}=\lambda_{n,p}(n/\pi)^{p/2}\,\Gamma(n/2)[\Gamma((n-p)/2)]^{-1}\,|\mathbf{S}|^{(n-1)/2}$. It follows from (a) that

$$E(\boldsymbol{\mu}\mid\bar{\mathbf{x}},\mathbf{S})=l_{n,p}\int\boldsymbol{\mu}\,|\boldsymbol{\Sigma}|^{-(n+p+1)/2}\,g_\lambda(\boldsymbol{\Sigma}^{-1}(\mathbf{S}+n(\bar{\mathbf{x}}-\boldsymbol{\mu})(\bar{\mathbf{x}}-\boldsymbol{\mu})'))\,d\boldsymbol{\mu}\,d\boldsymbol{\Sigma}$$

$$(\text{let }\mathbf{Y}=\boldsymbol{\Sigma}^{-1})$$

$$=l_{n,p}\int(\bar{\mathbf{x}}+(\boldsymbol{\mu}-\bar{\mathbf{x}}))\,|\mathbf{Y}|^{(n-p-1)/2}$$

$$\times g_\lambda(\mathbf{Y}(\mathbf{S}+n(\bar{\mathbf{x}}-\boldsymbol{\mu})(\bar{\mathbf{x}}-\boldsymbol{\mu})'))\,d\boldsymbol{\mu}\,d\mathbf{Y}$$

$$(\text{let }\mathbf{a}=\boldsymbol{\mu}-\bar{\mathbf{x}})$$

$$=l_{n,p}\bar{\mathbf{x}}\int|\mathbf{Y}|^{n-p-1)/2}\,g_\lambda(\mathbf{Y}(\mathbf{S}+n(\bar{\mathbf{x}}-\boldsymbol{\mu})(\bar{\mathbf{x}}-\boldsymbol{\mu})'))\,d\boldsymbol{\mu}\,d\mathbf{Y}$$

$$+l_{n,p}\int\mathbf{a}\,|\mathbf{Y}|^{n-p-1}\,g_\lambda(\mathbf{Y}(\mathbf{S}+n\mathbf{a}\mathbf{a}'))\,d\mathbf{a}\,d\mathbf{Y}$$

$$\triangleq\mathbf{I}_1+\mathbf{I}_2\,.$$

By Lemma 2.1, it follows that

$$\mathbf{I}_1 = l_{n,p}\lambda_{n,p}^{-1}\bar{\mathbf{x}} \int |\mathbf{S} + n(\bar{\mathbf{x}}-\boldsymbol{\mu})(\bar{\mathbf{x}}-\boldsymbol{\mu})'|^{-n/2}\, d\boldsymbol{\mu} = \bar{\mathbf{x}}.$$

Let $\mathbf{b} = \mathbf{S}^{-1/2}\mathbf{a}$, then Lemma 2.1 implies that

$$\mathbf{I}_2 = l_{n,p}\lambda_{n,p}^{-1}\, |\mathbf{S}|^{(n+1)/2}\, \mathbf{S}^{1/2} \int \mathbf{b}(1 + n\mathbf{b}'\mathbf{b})^{-n/2}\, d\mathbf{b} = \mathbf{0}.$$

So we have $E(\boldsymbol{\mu} \mid \bar{\mathbf{x}}, \mathbf{S}) = \bar{\mathbf{x}}$. Finally, we derive the posterior mean of $\boldsymbol{\Sigma}^{-1}$:

$$\begin{aligned}
E(\boldsymbol{\Sigma}^{-1} \mid \bar{\mathbf{x}}, \mathbf{S}) &= l_{n,p} \int \boldsymbol{\Sigma}^{-1}\, |\boldsymbol{\Sigma}|^{-(n+p+1)/2} \\
&\qquad \times g(\lambda(\boldsymbol{\Sigma}^{-1}(\mathbf{S} + n(\bar{\mathbf{x}}-\boldsymbol{\mu})(\bar{\mathbf{x}}-\boldsymbol{\mu})')))\, d\boldsymbol{\mu}\, d\boldsymbol{\Sigma} \\
&= l_{n,p} \int \mathbf{Y}\, |\mathbf{Y}|^{(n-p-1)/2}\, g(\lambda(\mathbf{Y}(\mathbf{S} + n(\bar{\mathbf{x}}-\boldsymbol{\mu})(\bar{\mathbf{x}}-\boldsymbol{\mu})'))\, d\boldsymbol{\mu}\, d\mathbf{Y} \\
&= l_{n,p}\lambda_{n,p}^{-1} \int |\mathbf{S} + n(\bar{\mathbf{x}}-\boldsymbol{\mu})(\bar{\mathbf{x}}-\boldsymbol{\mu})'|^{-n/2}\, d\boldsymbol{\mu} \\
&\qquad \times \int \lambda_{n,p}\mathbf{Y}\, |\mathbf{S} + n(\bar{\mathbf{x}}-\boldsymbol{\mu})(\bar{\mathbf{x}}-\boldsymbol{\mu})'|^{n/2} \\
&\qquad \times |\mathbf{Y}|^{(n-p-1)/2}\, g_\lambda(\mathbf{Y}(\mathbf{S} + n(\bar{\mathbf{x}}-\boldsymbol{\mu})(\bar{\mathbf{x}}-\boldsymbol{\mu})'))\, d\mathbf{Y} \\
&= l_{n,p}\lambda_{n,p}^{-1} \int n(-2\phi'(0))(\mathbf{S} + n(\bar{\mathbf{x}}-\boldsymbol{\mu})(\bar{\mathbf{x}}-\boldsymbol{\mu})')^{-1} \\
&\qquad \times |\mathbf{S} + n(\bar{\mathbf{x}}-\boldsymbol{\mu})(\bar{\mathbf{x}}-\boldsymbol{\mu})'|^{-n/2}\, d\boldsymbol{\mu} \\
&= \left(\frac{n}{\pi}\right)^{n/2} \Gamma(n/2)[\Gamma((n-p)/2)]^{-1} \cdot n(-2\phi'(0))\, \mathbf{S}^{-1} \\
&\qquad \times \int (\mathbf{I} + n\mathbf{S}^{-1/2}(\bar{\mathbf{x}}-\boldsymbol{\mu})(\bar{\mathbf{x}}-\boldsymbol{\mu})'\, \mathbf{S}^{-1/2})^{-1} \\
&\qquad \times |\mathbf{I} + n\mathbf{S}^{-1/2}(\bar{\mathbf{x}}-\boldsymbol{\mu})(\bar{\mathbf{x}}-\boldsymbol{\mu})'\, \mathbf{S}^{-1/2}|^{-n/2}\, |\mathbf{S}|^{-1/2}\, d\boldsymbol{\mu} \\
&= (\pi)^{-p/2}\, \Gamma(n/2)[\Gamma((n-p)/2)]^{-1}\, n(-2\phi'(0)\, \mathbf{S}^{-1} \\
&\qquad \times \int (\mathbf{I} + \mathbf{a}\mathbf{a}')(1 + \mathbf{a}'\mathbf{a})^{-n/2}\, d\mathbf{a}
\end{aligned}$$

$$((\mathbf{I}+\mathbf{aa}')^{-1}=\mathbf{I}-(1+\mathbf{a}'\mathbf{a})^{-1}\,\mathbf{aa}')$$

$$=\pi^{-p/2}\Gamma(n/2)[\Gamma((n-p)/2)]^{-1}\cdot n(-2\phi'(0))\,\mathbf{S}^{-1}$$

$$\times\left(\int \mathbf{I}(1+\mathbf{a}'\mathbf{a})^{-n/2}\,d\mathbf{a}-\int \mathbf{aa}'(1+\mathbf{a}'\mathbf{a})^{(n+2)/2}\,d\mathbf{a}\right.$$

$$=(-2\phi'(0))\,\mathbf{S}^{-1}-n(n-p)^{-1}\,(-2\phi'(0))\,\mathbf{S}^{-1}$$

$$=n(n-p-1)(n-p)^{-1}\,(-2\phi'(0))\,\mathbf{S}^{-1}. \quad \blacksquare$$

THEOREM 2.2. *Suppose that* $\mathbf{X}\sim SS_{n\times p}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$, *the prior distribution on* $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *is* (2.2), *and* g *satisfies the following conditions*:

(i) *nonincreasing, i.e.,* $\boldsymbol{\Sigma}_1 \geqslant \boldsymbol{\Sigma}_2 \geqslant 0$ *implies* $g_\lambda(\boldsymbol{\Sigma}_1) \leqslant g_\lambda(\boldsymbol{\Sigma}_2)$;

(ii) *continuous, i.e.,* $g_\lambda(\boldsymbol{\Sigma}) \to g_\lambda(\boldsymbol{\Sigma}_0)$ *as* $\boldsymbol{\Sigma}\to\boldsymbol{\Sigma}_0$ *along nonnegative definite matrices* $\boldsymbol{\Sigma}_0 \geqslant 0$. *Then the GMLE of* $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *is* $(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}) = (\bar{\mathbf{x}}, \mathbf{S}^{1/2}(\boldsymbol{\Sigma}^*)^{-1}\,\mathbf{S}^{1/2})$, *where the value* $\boldsymbol{\Sigma}^*$ *maximizes*

$$|\mathbf{W}|^{(n+p+1)/2}\, g_\lambda(\mathbf{W}), \tag{2.11}$$

with respect to $\mathbf{W}(>0)$.

Proof. Since g is nonincreasing in the sense of condition (i), for any given $\boldsymbol{\Sigma}>0$, the generalized likelihood function $L(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, which is $f(\boldsymbol{\mu}, \boldsymbol{\Sigma} \mid \bar{\mathbf{x}}, \mathbf{S})$ in (2.3), as a function of $\boldsymbol{\mu}$, arrives at its maximum at $\boldsymbol{\mu}=\bar{\mathbf{x}}$ and the concentrated likelihood is

$$L(\bar{\mathbf{x}}, \boldsymbol{\Sigma}) \propto |\mathbf{S}|^{(n-1)/2}\,|\boldsymbol{\Sigma}|^{-(n+p+1)/2}\, g_\lambda(\boldsymbol{\Sigma}^{-1/2}\mathbf{S}\boldsymbol{\Sigma}^{-1/2})$$

$$=|\mathbf{S}|^{(n-1)/2}\,|\boldsymbol{\Sigma}|^{-(n+p+1)/2}\, g_\lambda(\mathbf{S}^{1/2}\boldsymbol{\Sigma}^{-1}S^{1/2}).$$

Write $\mathbf{W}=\mathbf{S}^{1/2}\boldsymbol{\Sigma}^{-1}\mathbf{S}^{1/2}$. Then

$$L(\bar{\mathbf{x}}, \boldsymbol{\Sigma}) \propto |\mathbf{W}|^{(n+p+1)/2}\, g_\lambda(\mathbf{W}).$$

By Lemma 4.1.4 of Fang and Zhang (1990), $|\mathbf{W}|^{(n+p+1)/2}\, g_\lambda(\mathbf{W})$ attains its maximum at some positive definite matrix, say, $\boldsymbol{\Sigma}^*$. $\blacksquare$

For the purpose of constructing the confidence regions for $\boldsymbol{\mu}$, we present the following results:

THEOREM 2.3. *Suppose that* $\mathbf{X} \sim SS_{n \times p}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$ *and the prior distribution on* $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *is* (2.2). *Then*

$$\mathscr{L}\left\{\frac{(\mu_j - \bar{x}_j)}{\sqrt{s_{jj}/n(n-p)}} \,\middle|\, \mathbf{x}, \mathbf{S}\right\} = t_{n-p}, \qquad j = 1, ..., p,$$

where $\mathscr{L}\{y \mid z\}$ *denotes the conditional distribution of* y *for given* z, $\mathbf{S} = (s_{ij})$, $\boldsymbol{\mu} = (\mu_1, ..., \mu_p)'$, *and* $\mathbf{x}' = (\bar{x}_1, ..., \bar{x}_p)'$. *Furthermore*

$$\mathscr{L}\left\{\frac{n-p}{p}(\boldsymbol{\mu} - \bar{\mathbf{x}})' \left(\frac{1}{n}\mathbf{S}\right)^{-1} (\boldsymbol{\mu} - \bar{\mathbf{x}}) \mid \bar{\mathbf{x}}, \mathbf{S}\right\} = F_{p, n-p}.$$

It is noted by Theorems 2.1 and 2.2 that the Bayesian estimators of $\boldsymbol{\Sigma}$ are dependent on g_λ or equivalently on the characteristic generator of the random matrix $\mathbf{X}$, we will discuss this matter at the end of next section. On the other hand, it can be found from Theorems 2.1, 2.2 and 2.3 that the posterior distribution of $\boldsymbol{\mu}$, and therefore Bayesian inference for $\boldsymbol{\mu}$, is not dependent on the distribution of $\mathbf{X}$ as long as $\mathbf{X} \sim SS_{n \times p}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$. We may regard this property as the robustness of estimators of $\boldsymbol{\mu}$ and it can be extended to more general classes of distributions. This robustness property has been found in Osiewalski and his co-authors' work. Under a prior which is the counterpart of (2.2) for $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_p$, Osiewalski and Steel (1993a) studied multivariate elliptical regression models and obtained the posterior density of slope parameters, which is similar to that in (2.4). Osiewalski and Steel (1993b) further studied robust Bayesian inference for l_q-spherical models. The related arguments for multivariate location-scale models were developed in Fernández, Osiewalski and Steel (1994). The extension from multivariate elliptical distributions to the class of vector-spherical matrix distributions (cf. Fang and Zhang, 1990, Sect. 3.1.4) is trivial, but the extension to $SS_{n \times p}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$ is technical and of interest.

3. THE BAYESIAN STATISTICAL INFERENCE OF $\boldsymbol{\Sigma}$

Two loss functions which have been suggested and considered in the literature by James and Stein (1961), Olkin and Selliah (1977), Haff (1980), and Dey and Srinivasan (1985) are

$$L_1(\boldsymbol{\Sigma}, \mathbf{D}) = \operatorname{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{D}) - \log |\boldsymbol{\Sigma}^{-1}\mathbf{D}| - p, \tag{3.1}$$

$$L_2(\boldsymbol{\Sigma}, \mathbf{D}) = \operatorname{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{D} - \mathbf{I})^2 \tag{3.2}$$

which are entropy loss and quadratic loss, respectively. The estimation of the scatter matrix $\boldsymbol{\Sigma}$ based on multivariate normal distribution has been well studied (see, e.g., Section 4.3 of Muirhead 1982), and the corresponding results based on elliptical matrix distributions are considered in Section 4.4.2 of Fang and Zhang (1990). In this section, we consider Bayesian estimation for $\mathbf{\Sigma}$ under the entropy loss and the quadratic loss.

Consider the estimator of the form $\alpha\mathbf{S}$. Since

$$E(\mathbf{S} \mid \boldsymbol{\mu}, \mathbf{\Sigma}) = (n-1)(-2\phi'(0))\mathbf{\Sigma},$$

where $\phi(\cdot)$ is the characteristic generator of $\mathbf{X}$, the statistic $(1/(n-1)(-2\phi'(0)))\mathbf{S}$ is an unbiased estimator of $\mathbf{\Sigma}$. However, it is inadmissible under the loss function L_1 (cf. Corollary 2 on p. 150 of Fang and Zhang, 1990), and it is not the best Bayesian estimator of $\mathbf{\Sigma}$, having the form $\alpha\mathbf{S}$, from the following theorem.

THEOREM 3.1. *Suppose that* $\mathbf{X} \sim SS_{n\times p}(\boldsymbol{\mu}, \mathbf{\Sigma}, g)$ *with the finite second moment, and the prior distribution of* $(\boldsymbol{\mu}, \mathbf{\Sigma})$ *is given by* (2.2). *Then under the entropy loss* (3.1), *the best Bayesian linear estimator is*

$$\hat{\mathbf{\Sigma}}_0 = (n-p)[(n-p-1)\, n(-2\phi'(0))]^{-1}\, \mathbf{S}, \tag{3.3}$$

where $\phi(\cdot)$ *is the characteristic generator of* $\mathbf{X}$.

Proof. Assume $\mathbf{D} = \alpha\mathbf{S}$, $\alpha > 0$. Then under entropy loss L_1, the posterior risk of $\mathbf{D}$ by Theorem 2.1 is

$$\begin{aligned} R_1(\mathbf{\Sigma}, \mathbf{D}) &= E[L_1(\mathbf{\Sigma}, \mathbf{D}) \mid \bar{\mathbf{x}}, \mathbf{S}] \\ &= E[\operatorname{tr}(\mathbf{\Sigma}^{-1}\alpha\mathbf{S}) \mid \bar{\mathbf{x}}, \mathbf{S}] - E[\log(|\mathbf{\Sigma}^{-1}\alpha\mathbf{S}|) \mid \bar{\mathbf{x}}, \mathbf{S}] - p \\ &= \operatorname{tr}(\alpha E[\mathbf{\Sigma}^{-1} \mid \bar{\mathbf{x}}, \mathbf{S}]\mathbf{S}) - E[\log(|\mathbf{\Sigma}^{-1}\alpha\mathbf{S}|) \mid \bar{\mathbf{x}}, \mathbf{S}] - p \\ &= \alpha \operatorname{tr}\{[(n-p-1)\, n(-2\phi'(0))](n-p)^{-1}\, \mathbf{S}^{-1}\mathbf{S} \\ &\quad - p \log \alpha - E[\log(|\mathbf{\Sigma}^{-1}\mathbf{S}|) \mid \bar{\mathbf{x}}, \mathbf{S}] - p \\ &= \alpha p[(n-p-1)\, n(-2\phi'(0))](n-p)^{-1} - p \log \alpha \\ &\quad - E[\log(|\mathbf{\Sigma}^{-1}\mathbf{S}) \mid \bar{\mathbf{x}}, \mathbf{S}) - p. \end{aligned} \tag{3.4}$$

The proof is completed by noting that the value of α which minimizes the right-hand of (3.4) is $\alpha_0 = (n-p)[(n-p-1)\, n(-2\phi'(0))]^{-1}$. ∎

It is easy to obtain the best Bayesian linear estimator of $\mathbf{\Sigma}$ under the entropy loss by Theorem 3.1. Now we consider a nonlinear estimator of $\mathbf{\Sigma}$. For instance, consider estimator $\mathbf{h}_1(\mathbf{S})$ satisfying

$$\mathbf{L}\mathbf{h}_1(\mathbf{B})\mathbf{L}' = \mathbf{h}_1(\mathbf{L}\mathbf{B}\mathbf{L}')$$

for each $\mathbf{L} \in LT^+(p) = \{\mathbf{A} \mid \mathbf{A}$ is a $p \times p$ lower triangle matrix with positive diagonal elements$\}$.

Let $\mathbf{S} = \mathbf{TT}'$, where $\mathbf{T} \in LT^+(p)$. By the discussion of Section 4.3 of Muirhead (1982), it suffices to consider only

$$\mathbf{h}_1(\mathbf{S}) = \mathbf{T\Delta T}', \tag{3.5}$$

where $\mathbf{\Delta} = \mathrm{diag}(\delta_1, ..., \delta_p) > 0$. Let

$$\mathbf{M} = \mathbf{T}'\mathbf{\Sigma}^{-1}\mathbf{T} = \mathbf{L}'\mathbf{L}, \qquad \text{where} \quad \mathbf{L} = (\boldsymbol{l}_{(1)}, ..., \boldsymbol{l}_{(p)})' \in LT^+(p). \tag{3.6}$$

THEOREM 3.2. *Suppose that* $\mathbf{X} \sim SS_{n \times p}(\boldsymbol{\mu}, \mathbf{\Sigma}, g)$ *with the finite second moment, and the prior distribution of* $(\boldsymbol{\mu}, \mathbf{\Sigma})$ *is given by* (2.2). *Then under the entropy loss, the best Bayesian estimator having the form* (3.5) *is*

$$\mathbf{h}_1^*(\mathbf{S}) = \mathbf{T\Delta}^*\mathbf{T}', \tag{3.7}$$

where $\mathbf{S} = \mathbf{TT}'$, $\mathbf{T} \in LT^+(p)$, $\mathbf{\Delta}^* = diag(\delta_1, ..., \delta_p)$, $\delta_i^{-1} = E\boldsymbol{l}'_{(i)}\boldsymbol{l}_{(i)}$, *and* $\mathbf{L} = (\boldsymbol{l}_{(1)}, ..., \boldsymbol{l}(p))'$ *is defined by* (3.6) *and* $\mathbf{M}$ *has a pdf*

$$k_{n,p}\, |\mathbf{M}|^{(n-p)/2-1} \int_{R^p} g_\lambda(\mathbf{M} + n\mathbf{vv}'))\, d\mathbf{v}, \tag{3.8}$$

where $k_{n,p}$ *is the normalizing constant.*

Before proving the theorem we need the following lemma.

LEMMA 3.1. *Let* $\mathbf{X}$ *and* $\mathbf{Y}$ *be two random matrices. Let* $p(\mathbf{X}, \mathbf{Y})$, $h(\mathbf{Y})$, *and* $p(\mathbf{X} \mid \mathbf{Y})$ *be the joint density of* $\mathbf{X}$ *and* $\mathbf{Y}$, *the marginal density of* $\mathbf{Y}$ *and the conditional density of* $\mathbf{X}$ *given* $\mathbf{Y}$, *respectively. Let* $\mathbf{Z} = \mathbf{t}(\mathbf{Y})$ *be a transformation where* $\mathbf{Z}$ *has the same number of independent variables as* $\mathbf{Y}$. *Then the conditional distribution of* $\mathbf{X}$ *given* $\mathbf{Z}$ *is*

$$p(\mathbf{X} \mid \mathbf{Z}) = \frac{p(\mathbf{X}, \mathbf{t}(\mathbf{Y}))}{h(\mathbf{t}(\mathbf{Y}))}. \tag{3.9}$$

Proof. It is known that the conditional distribution of $\mathbf{X}$ given $\mathbf{Z}$ can be expressed as

$$p(\mathbf{X} \mid \mathbf{Z}) = \frac{p(\mathbf{X}, \mathbf{Z})}{h(\mathbf{Z})}. \tag{3.10}$$

Let J be the Jacobian determinant of the transformation $\mathbf{t}$. Then the joint density of $\mathbf{X}$ and $\mathbf{Z}$ is $p(\mathbf{X}, \mathbf{t}(\mathbf{Y}))\, |J|$ and the density of $\mathbf{Z}$ is $h(\mathbf{t}(\mathbf{Y}))\, |J|$. Put them into (3.10) and the assertion (3.9) follows. ∎

Proof of Theorem 3.2. We derive the pdf of $\mathbf{M}$ defined in (3.6) first. From (2.3) we can find the conditional pdf of $\mathbf{\Sigma}$ given $\bar{\mathbf{x}}$ and $\mathbf{S}$ as

$$f(\mathbf{\Sigma} \mid \bar{\mathbf{x}}, \mathbf{S}) = c\, |\mathbf{S}|^{(n-1)/2}\, |\mathbf{\Sigma}|^{-(n+p+1)/2} \int g_\lambda(\mathbf{\Sigma}^{-1}(\mathbf{S} + n(\bar{\mathbf{x}} - \boldsymbol{\mu})(\bar{\mathbf{x}} - \boldsymbol{\mu})'))\, d\boldsymbol{\mu}$$

$$= c\, |\mathbf{S}|^{(n-1)/2}\, |\mathbf{\Sigma}|^{-(n+p+1)/2} \int g_\lambda(\mathbf{\Sigma}^{-1}(\mathbf{S} + n\mathbf{u}\mathbf{u}'))\, d\mathbf{u}, \tag{3.11}$$

where c is the normalizing constant and is not always the same in the proof. The pdf $f(\mathbf{\Sigma} \mid \bar{\mathbf{x}}, \mathbf{S})$ is independent of $\bar{\mathbf{x}}$ and (3.11) gives $f(\mathbf{\Sigma} \mid \mathbf{S})$. By Lemma 3.1 we have

$$f(\mathbf{\Sigma} \mid \mathbf{T}) = c\, |\mathbf{T}\mathbf{T}'|^{(n-1)/2}\, |\mathbf{\Sigma}|^{-(n+p+1)/2} \int g_\lambda(\mathbf{\Sigma}^{-1}(\mathbf{T}\mathbf{T}' + n\mathbf{u}\mathbf{u}'))\, d\mathbf{u}.$$

Let $\mathbf{A} = \mathbf{\Sigma}^{-1}$ and $\mathbf{v} = \mathbf{T}^{-1}\mathbf{u}$. The pdf of $\mathbf{A}$ given $\mathbf{T}$ is

$$f(\mathbf{A} \mid \mathbf{T}) = c\, |\mathbf{T}'\mathbf{T}|^{n/2}\, |\mathbf{A}|^{(n-p-1)/2} \int g_\lambda(\mathbf{T}'\mathbf{A}\mathbf{T}(\mathbf{I} + n\mathbf{v}\mathbf{v}'))\, d\mathbf{v}.$$

Let $\mathbf{M} = \mathbf{T}'\mathbf{A}\mathbf{T} = \mathbf{T}'\mathbf{\Sigma}^{-1}\mathbf{T}$. The pdf of $\mathbf{M}$ given $\mathbf{T}$ is

$$\begin{aligned} f(\mathbf{M} \mid \mathbf{T}) &= c\, |\mathbf{M}|^{(n-p-1)/2} \int g_\lambda(\mathbf{M}(\mathbf{I} + n\mathbf{v}\mathbf{v}'))\, d\mathbf{v} \\ &= c\, |\mathbf{M}|^{(n-p)/2-1} \int g_\lambda(\mathbf{M} + n\mathbf{v}\mathbf{v}')\, d\mathbf{v} \end{aligned}$$

which is independent of $\mathbf{T}$ and (3.8) follows. From the above proof we find that $f(\mathbf{M} \mid \bar{\mathbf{x}}, \mathbf{\Sigma}) = f(\mathbf{M})$. Therefore,

$$\begin{aligned} R_1(\mathbf{\Sigma}, \mathbf{h}_1(\mathbf{S})) &= E\{[\operatorname{tr}(\mathbf{\Sigma}^{-1}\mathbf{h}_1(\mathbf{S})) - \log(|\mathbf{\Sigma}^{-1}\mathbf{h}_1(\mathbf{S})|) - p] \mid \bar{\mathbf{x}}, \mathbf{S}\} \\ &= E\{[\operatorname{tr}(\mathbf{T}'\mathbf{\Sigma}^{-1}\mathbf{T}\mathbf{\Delta}) - \log(|\mathbf{T}'\mathbf{\Sigma}^{-1}T\mathbf{\Delta}|) - p] \mid \bar{\mathbf{x}}, \mathbf{S}\} \\ &= E[\operatorname{tr}(\mathbf{M}\Delta) - \log(|\mathbf{M}\mathbf{\Delta}|) - p] \\ &= \sum_{i=1}^{p} [\delta_i E(\boldsymbol{l}'_{(i)}\boldsymbol{l}_{(i)}) - \log \delta_i] - E(\log |\mathbf{M}|) - p. \end{aligned}$$

Note that $R_1(\mathbf{\Sigma}, \mathbf{h}_1(\mathbf{S}))$ attains its minimum at $\delta_i = (E(\boldsymbol{l}'_{(i)}\boldsymbol{l}_{(i)}))^{-1}$, $i = 1, ..., p$. The proof is completed. ▌

Now let us consider the best Bayesian estimator of the form (3.5) under the quadratic loss.

THEOREM 3.3. *Suppose that* $\mathbf{X} \sim SS_{n\times p}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$ *with the finite second moment, and the prior distribution on* $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *is given by* (2.2). *Then, under the quadratic loss* (3.2), *the best Bayesian estimator of* $\boldsymbol{\Sigma}$ *if the form* (3.5) *is*

$$\mathbf{h}_2^*(\mathbf{S}) = \mathbf{T}\boldsymbol{\Delta}^*\mathbf{T}', \tag{3.12}$$

where $\mathbf{T}$ *is given in Theorem* 3.2 *and* $\boldsymbol{\Delta}^* = \mathrm{diag}(\delta_1, ..., \delta_p)$, *and* $(\delta_1, ..., \delta_p)'$ *is the solution of the equation*

$$\mathbf{B}\boldsymbol{\delta} = g, \tag{3.13}$$

where $\mathbf{g} = (E(\boldsymbol{l}'_{(1)}\boldsymbol{l}_{(1)}), ..., E(\boldsymbol{l}'_{(p)}\boldsymbol{l}_{(p)}))'$, $\mathbf{B} = (E(\boldsymbol{l}'_{(i)}\boldsymbol{l}_{(j)})^2)_{p\times p}$, *and* $\mathbf{L} = (\boldsymbol{l}_{(1)}, ..., \boldsymbol{l}_{(p)})'$ *is defined in* (3.6).

Proof. We have

$$\begin{aligned}
R_2(\boldsymbol{\Sigma}, \mathbf{h}_1(\mathbf{S})) &= E[\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{h}_1(\mathbf{S}) - \mathbf{I})^2 \mid \bar{\mathbf{x}}, \mathbf{S}]\\
&= E[\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{T}\boldsymbol{\Delta}\mathbf{T}' - \mathbf{I})^2 \mid \bar{\mathbf{x}}, \mathbf{S}]\\
&= E[\mathrm{tr}(\mathbf{T}'\boldsymbol{\Sigma}^{-1}\mathbf{T}\boldsymbol{\Delta} - \mathbf{I})^2 \mid \bar{\mathbf{x}}, \mathbf{S}]\\
&= E[\mathrm{tr}(\mathbf{M}\boldsymbol{\Delta} - \mathbf{I})^2]\\
&= E\,\mathrm{tr}(\mathbf{L}\boldsymbol{\Delta}\mathbf{L}' - \mathrm{I})^2\\
&= \boldsymbol{\delta}\mathbf{B}\boldsymbol{\delta} - 2\boldsymbol{\delta}'\mathbf{g} + p,
\end{aligned}$$

As $\partial R_2(\boldsymbol{\Sigma}, \mathbf{h}_1(\mathbf{S}))/\partial\boldsymbol{\delta} = 2\mathbf{B}\boldsymbol{\delta} - 2\mathbf{g} = 0$, $\boldsymbol{\Delta}^*$ should satisfy (3.13). Since $R_2(\boldsymbol{\Sigma}, h_1(\mathbf{S}))$ is a quadratic form of $\boldsymbol{\delta}$ and $R_2(\boldsymbol{\Sigma}, \mathbf{h}_1(\mathbf{S})) \geqslant 0$, i.e., it has a lower bound, Eq. (3.13) is consistent. Moreover $R_2(\boldsymbol{\Sigma}, \mathbf{h}_1(\mathbf{S}))$ can attain its minimum value. ▮

From the derived Bayesian estimators of $\boldsymbol{\Sigma}$, it can be found that these estimators depend on g_λ or equivalently on the characteristic generator of the random matrix $\mathbf{X}$. Hence unlike the Bayesian estimators of $\boldsymbol{\mu}$, the Bayesian estimators of $\boldsymbol{\Sigma}$ are not invariant among the elliptical matrix distributions. Now let us compare these Bayesian estimators of $\boldsymbol{\Sigma}$ with the classical ones obtained in Section 4.4.2 of Fang and Zhang (1990).

The best classical estimator of $\boldsymbol{\Sigma}$, having the form $\alpha\mathbf{S}$, is $[(n-1)\{-2\phi'(0)\}]^{-1}\,\mathbf{S}$, under the loss L_1, while the corresponding Bayesian estimator is $(n-p)[(n-p-1)\ n\{-2\phi'(0)\}]^{-1}\,\mathbf{S}$. For normal distribution $N_{n\times p}(\mathbf{1}\boldsymbol{\mu}', \mathbf{I}_n \otimes \boldsymbol{\Sigma})$, the classical one is $(n-1)^{-1}\,\mathbf{S}$, while the Bayesian one is $(n-p)[(n-p-1)n]^{-1}\,\mathbf{S}$. Note that $p \geqslant 1$ and $n > p$, we have $n^{-1} < (n-p)^{-1}$, and $1/(n-1) < (n-p)/(n-p-1)n$. The inequality implies that the Bayesian estimator of $\boldsymbol{\Sigma}$ is greater than the classical one in that sense that $\mathbf{A} > \mathbf{B}$ if $\mathbf{A} - \mathbf{B}$ is positive definite. From the form of the estimator, we find that the magnitude of the deviation from multinormality on the estimator of $\mathbf{\Sigma}$, having the form $\alpha\mathbf{S}$, is only a multiple of $\phi'(0)$. Severe deviation from multinormality may lead to an estimator of $\mathbf{\Sigma}$ which is quite different from the sample covariance matrix.

Comparing Theorem 4.4.5 in Fang and Zhang (1990) with Theorem 3.2 of this paper, it can be found that the estimators of $\mathbf{\Sigma}$ have a close form, but they are different. For the classical estimators, $\Delta^* = \operatorname{diag}(\{E(\boldsymbol{l}_1'\boldsymbol{l}_1)\}^{-1}, ..., \{E(\boldsymbol{l}_p'\boldsymbol{l}_p)\}^{-1})$, where $\boldsymbol{l}_1, ..., \mathrm{l}_p$ are the columns of $\mathbf{L}$, defined in (3.6) (cf. Theorem 4.2.1 of Fang and Zhang (1990)), while for the Bayesian estimator, $\Delta^* = \operatorname{diag}(\{E(\boldsymbol{l}_{(1)}'\boldsymbol{l}_{(1)})\}^{-1}, ..., \{E(\boldsymbol{l}_{(p)}'\boldsymbol{l}_{(p)})\}^{-1})$, where $\boldsymbol{l}_{(1)}, ..., \boldsymbol{l}_{(p)}$ are the rows of $\mathbf{L}$. For the normal distribution $N_{n\times p}(\mathbf{1}\boldsymbol{\mu}', \mathbf{I}_n \otimes \mathbf{\Sigma})$, we can obtain the distributions of elements of $\mathbf{L}$ (see, e.g., Theorem 3.2.4 of Muirhead (1982)). Further we can derive that $\Delta^* = \operatorname{diag}(\delta_1, ..., \delta_p)$, with $\delta_i = (n+p-2i)^{-1}$ for the classical estimator, while $\delta_i = (n-p+2i)^{-1}$ for the Bayesian counterpart.

The difference between the Bayesian estimator and the classical one under the loss L_2 is similar to that under the loss L_1. The estimators depends on the density of $\mathbf{X}$ through the distribution of $\mathbf{M}$, which is more complicated than that in Theorem 3.1. Except normal distribution case, it may involve numerical integration methods to compute δ_i in Theorem 3.2 and $\mathbf{B}$ and $\mathbf{g}$ in Theorem 3.3.

4. SOME EXAMPLES

In this section we apply the theory developed in the previous sections to several subclasses of elliptical matrix distributions.

EXAMPLE 4.1 (Kotz's Type Elliptical Matrix Distributions). If the pdf of $\mathbf{X}$ is of the form

$$f(\mathbf{X} \mid \boldsymbol{\mu}, \mathbf{\Sigma}) = d_{n,p}\, |\mathbf{\Sigma}|^{-n/2}\, |\mathbf{\Sigma}^{-1}(\mathbf{X}-\mathbf{1}\boldsymbol{\mu}')'\,(\mathbf{X}-\mathbf{1}\boldsymbol{\mu}')|^{N-(p+1)/2} \times \exp\{-r[\operatorname{tr}(\mathbf{\Sigma}^{-1}(\mathbf{X}-\mathbf{1}\boldsymbol{\mu}')'\,(\mathbf{X}-\mathbf{1}\boldsymbol{\mu}'))]^s\}, \tag{4.1}$$

where N, r, s are known parameters satisfying $2N+n > p+1$, and s, $r > 0$, $\mathbf{X}$ is said to have a Kotz's type elliptical matrix distribution.

By a direct calculation, it follows from Theorem 2.1 that

$$f(\boldsymbol{\mu}, \mathbf{\Sigma} \mid \bar{\mathbf{x}}, \mathbf{S}) = \lambda_{n,p} \left(\frac{n}{\pi}\right)^{p/2} \frac{\Gamma(n/2)}{\Gamma((n-p)/2)}\, |\mathbf{S}|^{(n-1)/2}\, |\mathbf{\Sigma}|^{-(N+n/2+p+1)} \times |\mathbf{S}+n(\bar{\mathbf{x}}-\boldsymbol{\mu})(\bar{\mathbf{x}}-\boldsymbol{\mu})'|^{N-(p+1)/2} \times \exp\{-r\operatorname{tr}[\mathbf{\Sigma}^{-1}(\mathbf{S}+n(\bar{\mathbf{x}}-\boldsymbol{\mu})(\bar{\mathbf{x}}-\boldsymbol{\mu})')]^s\},$$

and $E(\boldsymbol{\mu} \mid \bar{\mathbf{x}}, \mathbf{S}) = \bar{\mathbf{x}}$. To calculate $E(\boldsymbol{\Sigma}^{-1} \mid \mathbf{x}, \mathbf{S})$, it is necessary to compute $\phi'(0)$. By Theorem 4 of Li (1993), when $s > 1/2$,

$$\phi(x) = \sum_{k=1}^{\infty} \frac{\Gamma\{[(2N+n-p-1)\,p/2+k]/s\}}{((2N+n-p-1)\,p/2)^{[k]}\,\Gamma((2N+n-p-1)\,p/2s)} \times \frac{1}{k!} \cdot \frac{((2N+n-p-1)/2)^{[k]}}{(n/2)^{[k]}} \left(-\frac{1}{4r^s}\right)^k \cdot x^k,$$

where $a^{[k]} = a(a+1)\cdots(a+k-1)$. Therefore

$$\phi'(0) = \frac{\Gamma\{[(2N+n-p-1)\,p/2+1]/s\}}{[(2N+n-p-1)\,p/2]\,\Gamma((2N+n-p-1)\,p/2s)} \cdot \frac{(2N+n-p-1)/2}{n/2} \cdot \left(-\frac{1}{4r^s}\right)$$

and

$$E(\boldsymbol{\Sigma}^{-1} \mid \bar{\mathbf{x}}, \mathbf{S}) = \frac{(n-p-1)}{(n-p)} \cdot \frac{\Gamma\{[(2N+n-p-1)\,p/2+1]/s\}}{\Gamma[(2N+n-p)\,p/2s]\,pr^s}\,\mathbf{S}^{-1}.$$

In particular, when $s = 1$

$$E(\boldsymbol{\Sigma}^{-1} \mid \bar{\mathbf{x}}, \mathbf{S}) = \frac{(n-p-1)}{(n-p)} \cdot \frac{(2N+n-p-1)}{2r}\,\mathbf{S}^{-1}.$$

When $\mathbf{X}$ has a matrix normal distribution $(N = (p+1)/2,\ r = 1/2$ and $s = 1)$, we have

$$E(\boldsymbol{\Sigma}^{-1} \mid \bar{\mathbf{x}}, \mathbf{S}) = \frac{n-p-1}{n-p}\,n\mathbf{S}^{-1}.$$

Next we will find the generalized maximum likelihood estimate of $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Since

$$\begin{aligned} |\mathbf{W}|^{(n+p+1)/2}\,g_\lambda(\mathbf{W})) &= |\mathbf{W}|^{(2N+n+p+1)/2}\exp(-r(\operatorname{tr}(\mathbf{W}))^s) \\ &= (w_1 \cdots w_p)^{(2N+n+p+1)/2}\exp(-r(w_1 + \cdots + w_p)^s), \end{aligned} \tag{4.2}$$

where $(w_1, ..., w_p)$ are the eigenvalues of $\mathbf{W}$. To maximize (4.2) with respect to $w_i's$, $(w_1, ..., w_p)$ should satisfy the equations

$$w_i^{-1} = \frac{2rs}{2N+n+p+1}\,(w_1 + \cdots + w_p)^{s-1}, \qquad i = 1, ..., p. \tag{4.3}$$

It can be shown that

$$w_1 = w_2 = \cdots = w_p = \left(\frac{2rs}{p^{s-1}(2N+n+p+1)} \right)^{-1/s}$$

is a solution of (4.3). So the GMLE of $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is

$$(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}) = \left(\bar{\mathbf{x}}, \left[\frac{2rs}{p^{s-1}(2N+n+p+1)} \right]^{1/s} \mathbf{S} \right).$$

In particular, when $s=1$, $N=(p+1)/2$ and $r=1/2$ (normal case),

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n+2(p+1)} \mathbf{S}.$$

By Theorem 3.1, under the entropy loss the best Bayesian linear estimator of $\boldsymbol{\Sigma}$, having of the form $\alpha\mathbf{S}$, is

$$\hat{\boldsymbol{\Sigma}}_0 = \frac{n-p}{n-p-1} \cdot \frac{\Gamma[(2N+n-p)\,p/2s]\,pr^s}{\Gamma\{[(2N+n-p-1)\,p/2+1]/s\}} \mathbf{S}, \qquad \text{for} \quad s>1/2.$$

EXAMPLE 4.2 (Pearson Type II Elliptical Matrix Distributions). If the pdf of $\mathbf{X}$ is of the form

$$f(\mathbf{X} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) = d_{n,p}\, |\boldsymbol{\Sigma}|^{-n/2}\, |\mathbf{I}_p - \boldsymbol{\Sigma}^{-1}(\mathbf{X}-\mathbf{1}\boldsymbol{\mu}')'\,(\mathbf{X}-\mathbf{1}\boldsymbol{\mu}')|^{m-(p+1)/2},$$

where $0 \leqslant \boldsymbol{\Sigma}^{-1/2}(\mathbf{X}-\mathbf{1}\boldsymbol{\mu}')'\,(\mathbf{X}-\mathbf{1}\boldsymbol{\mu}')\boldsymbol{\Sigma}^{-1/2} \leqslant \mathbf{I}_p$ in the nonnegative definite sense, and $m>(p+1)/2$. It is follows from Theorem 2.1 that

$$f(\boldsymbol{\mu}, \boldsymbol{\Sigma} \mid \bar{\mathbf{x}}, \mathbf{S}) = \lambda_{n,p} \left(\frac{n}{\pi} \right)^{p/2} \frac{\Gamma(n/2)}{\Gamma[(n-p)/2]}\, |\mathbf{S}|^{(n-1)/2}$$
$$\times |\boldsymbol{\Sigma}|^{-(n+p+1)/2}\, |\mathbf{I} - \boldsymbol{\Sigma}^{-1}(\mathbf{S} + n(\bar{\mathbf{x}}-\boldsymbol{\mu})(\bar{\mathbf{x}}-\boldsymbol{\mu})')|^{m-(p+1)/2}.$$

Then by Theorem 3.3 of Li (1992),

$$E(\boldsymbol{\mu} \mid \bar{\mathbf{x}}, \mathbf{S}) = \bar{\mathbf{x}},$$

$$E(\boldsymbol{\Sigma}^{-1} \mid \bar{\mathbf{x}}, \mathbf{S}) = \frac{n-p-1}{n-p} \cdot \frac{n}{n+2m} \mathbf{S}^{-1}$$

and the GMLE of $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is

$$(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}) = \left(\bar{\mathbf{x}}, \frac{n+2m}{n+p+1} \mathbf{S} \right).$$

Under the entropy loss, the best Bayesian linear estimate of $\mathbf{\Sigma}$, having of the form $\alpha\mathbf{S}$, is

$$\hat{\mathbf{\Sigma}}_0 = \frac{n-p}{n-p-1} \cdot \frac{n+2m}{n} \mathbf{S}.$$

ACKNOWLEDGMENT

The authors thank a referee for his valuable comments, pointing out references on Bayesian statistical inference for regression models with errors being distributed according to an elliptical distribution and helping improve the presentation of the paper. The authors also thank Dr. Hongbin Fang for his useful comments. The authors are grateful to Mr. K. Owzar for helping to improve the presentation of the paper.

REFERENCES

1. T. W. Anderson, Nonnormal multivariate distributions: inference based on elliptical contoured distributions, *in* "Multivariate Analysis: Future Directions" (C. R. Rao, Ed.), pp. 1–24, Elsevier, Amsterdam, 1993.
2. T. W. Anderson and K. T. Fang, On the theory of multivariate elliptically contoured distributions and their applications, *in* "Statistical Inference in Elliptically Contoured and Related Distributions" (K. T. Fang and T. W. Anderson, Eds.), pp. 1–24, Allerton Press, New York, 1990a.
3. T. W. Anderson and K. T. Fang, Inference in multivariate elliptically contoured distributions based on maximum likelihood, *in* "Statistical Inference in Elliptically Contoured and Related Distributions" (K. T. Fang and T. W. Anderson, Eds.), pp. 201–216, Allerton Press, New York, 1990b.
4. J. O. Berger, "Statistical Decision Theory and Bayesian Analysis," Springer-Verlag, New York, 1985.
5. S. Chib, J. Osiewalski, and M. F. J. Steel, Posterior inference on the degrees of freedom parameter in multivariate-t regression models, *Econom. Lett.* **37** (1991), 391–397.
6. S. Chib, R. C. Tiwari, and S. R. Jammalamadaka, Bayes prediction in regressions with elliptical errors, *J. Econometrics* **38** (1988), 349–360.
7. A. P. Dawid, Spherical matrix distributions and a multivariate model, *J. Roy. Statist. Soc. Ser. B* **39** (1977), 254–261.
8. D. K. Dey and C. Srinivasan, Estimation of covariance matrix under Stein's loss, *Ann. Statist.* **13** (1985), 1581–1591.
9. K. T. Fang, S. Kotz, and K. W. Ng, "Symmetric Multivariate and Related Distributions," Chapman & Hall, London, 1989.
10. K. T. Fang, R. Z. Li, and J. J. Liang, A multivariate version of Ghosh's T_3-plot to detect non-multinormality, *Comput. Statist. Data Anal.* **28** (1998), 371–386.
11. K. T. Fang and Y. T. Zhang, "Generalized Multivariate Analysis," Springer-Verlag, Berlin, and Science Press, Beijing, 1990.
12. C. Fernández, J. Osiewalski, and M. F. J. Steel, The continuous multivariate location-scale model revisited: A tale of robustness, *Biometrika* **81** (1994), 588–594.
13. L. R. Haff, Empirical Bayes estimation of the multivariate normal covariance matrix, *Ann. Statist.* **8** (1980), 586–597.
14. C. S. Herz, Bessel functions of matrix argument, *Ann. Math.* **61** (1955), 474–523.

15. W. James and C. Stein, Estimation with quadratic loss, *in* "Proc. Fourth Berkeley Symp. Math. Statist. Prob.," Vol. 1, pp. 361–379, University of California Press, Berkeley, 1961.
16. S. R. Jammalamadaka, R. C. Tiwari, and S. Chib, Bayes prediction in the linear model with spherically symmetric errors, *Econom. Lett.* **24** (1987), 39–44.
17. J. Läuter, Exact t and F tests for analyzing studies with multiple endpoints, *Biometrics* **52** (1996), 964–970.
18. J. Läuter, F. Glimm, and S. Kropf, New multivariate tests for data with an inherent structure, *Biometrical J.* **38** (1996), 5–23.
19. R. Z. Li, Some subclasses of spherical matrix, *J. Grad. School Acad. Sinica* **9** (1992), 125–140.
20. R. Z. Li, The characteristic functions of spherical matrix distributions, *Statist. Probab. Lett.* **17** (1993), 273–279.
21. J. J. Liang, "Testing Multinormality, Spherical and Elliptical Symmetry," Ph.D. dissertation, Department of Mathematics, Hong Kong Baptist University, 1998.
22. R. J. Muirhead, "Aspects of Multivariate Statistical Theory," Wiley, New York, 1982.
23. I. Olkin and J. B. Selliah, Estimating covariance in a multivariate normal distribution, *in* "Statistical Decision Theory and Related Topics, II" (S. S. Gupta and D. Moore, Eds.), pp. 313–326, Academic Press, New York, 1977.
24. J. Osiewalski A note on Bayesian inference in a regression model with regression errors, *J. Econometrics* **48** (1991), 183–193.
25. J. Osiewalski and M. F. J. Steel, Robust Bayesian inference in elliptical regression models, *J. Econometrics* **57** (1993a), 345–363.
26. J. Osiewalski and M. F. J. Steel, Robust Bayesian inference in l_q-regression models, *Biometrika* (1993b), 456–460.
27. S. J. Press, "Applied Multivariate Analysis," 2nd ed., Holt, Rinehart & Winston, New York, 1982.
28. P. K. Sing, S. Misra, and S. K. Pandey, A generalized class of estimators in linear regression models with multivariate-t distributed error, *Statist. Probab. Lett.* **23** (1995), 171–178.
29. A. Zellner, Bayesian and non-Bayesian analysis of the regression model with multivariate Student-t error terms, *J. Amer. Statist. Assoc* **71** (1976), 400–405.