

Estimation of the Trace of the Scale Matrix of the Scale Mixture of Multivariate Normal Distributions

Anwarul H. Joarder

M. Amir Hossain

Institute of Statistical Research and Training

University of Dhaka

Dhaka 1000

Bangladesh

ABSTRACT

A class of estimators for the trace of the scale matrix of the scale mixture of the multivariate normal distributions has been considered. The estimation strategy has been developed in the light of a quadratic loss function. The conditions under which the class of proposed estimators outperform the class of usual estimators have been derived.

1. INTRODUCTION

We consider the estimation of the trace of the scale matrix of the scale mixture of the multivariate normal distributions. The multivariate normal distribution and the multivariate t -distribution are the two important special cases of the scale mixture of the multivariate normal distributions. This class of distributions contains thin tailed as well as fat tailed distributions and hence is important in modelling many real world data.

It may be mentioned that many authors have observed that the empirical distributions of rates of return of common stocks have somewhat fatter tails than that of the normal distribution. The evidence provided by Mandelbrot (1963) and Fama (1965) suggests that one could explicitly account for the observed fat tails by using symmetric stable distributions. The multivariate t -distribution has also fatter tails and can, therefore, characterize rates of return on common stocks.

The estimation of the trace of the covariance matrix (scale matrix) of multivariate normal distribution has been considered by Olkin and Selliah (1977) under weighted squared error loss function. Dey (1988) considers the estimation of the trace of the covariance

Journal of Information & Optimization Sciences

Vol. 16 (1995), No. 3, pp. 565-572

© Analytic Publishing Co.

0252-2667/95 \$2.00+.25

matrix of the multivariate normal distribution under a squared error loss function.

Consider p -dimensional ($p \geq 2$) random vectors (not necessarily independent) $X_1, X_2, \dots, X_N$ having a joint probability density function (p.d.f.) given by

$$f(x_1, x_2, \dots, x_N) = \int_0^\infty \frac{|\tau^2 \Lambda|^{-Np/2}}{(2\pi)^{Np/2}} \exp \left(-\frac{1}{2} \sum_{j=1}^N (x_j - \theta)' (\tau^2 \Lambda)^{-1} (x_j - \theta) \right) h(\tau) d\tau \quad \dots (1.1)$$

where $x_j = (x_{1j}, x_{2j}, \dots, x_{pj})'$ is a p -dimensional column vector, θ an unknown vector of location parameters and Λ an unknown positive definite matrix of scale parameters while $h(\tau)$ is the p.d.f. of a suitable nondegenerate continuous random variable τ with positive support. The p.d.f. in (1.1) is the joint p.d.f. of N vectors each having a scale matrix of a multivariate normal distributions.

If τ has the p.d.f. given by

$$h(\tau) = \frac{2\tau^{-(v+1)}}{\Gamma(v/2)(2/v)^{v/2}} \exp \left(-\frac{v}{2\tau^2} \right)$$

then the p.d.f. in (1.1) defines a multivariate t -distribution given by

$$f(x_1, x_2, \dots, x_N) = \frac{K(v, Np)}{|\Lambda|^{N/2}} \left(v + \sum_{j=1}^N (x_j - \theta)' \Lambda^{-1} (x_j - \theta) \right)^{-(v+Np)/2} \quad (1.2)$$

where $x_j = (x_{1j}, x_{2j}, \dots, x_{pj})'$ is a p -dimensional column vector, θ an unknown vector of location parameters and Λ is an unknown positive definite matrix of scale parameters while the scalar v is assumed to be a known positive constant. The normalizing constant $K(v, Np)$ is given by

$$K(v, Np) = \frac{v^{v/2} \Gamma((v+Np)/2)}{\pi^{Np/2} \Gamma(v/2)}.$$

As mentioned earlier, the distribution in (1.2) represents the scale mixture of the multivariate normal distributions ; it has been considered, among others, by Zellner (1976) and Sutradhar and Ali (1986) in the context of a stock market problem, Dey (1990) and Singh (1991).

We now propose a class of estimators for the trace of the scale matrix of the scale mixture of the multivariate normal distributions.

Let $\delta = \text{tr}(\Lambda)$. In estimating δ by an estimator $u(A)$, where A is the sample sum of product matrix, we consider the following loss function

$$L(u(A), \delta) = (u(A) - \delta)^2 \quad (1.3)$$

where $u(A)$ is any estimator of $\text{tr}(\Lambda)$. The risk function is defined as usual by taking expectation over the loss function i.e.

$$R(u(A), \delta) = E[L(u(A), \delta)].$$

In particular, consider the following two classes of non-negative estimators of $\delta = \text{tr}(\Lambda)$

$$\text{class of usual estimators } \tilde{\delta} = c_1 \text{tr}(A), \quad (1.4)$$

$$\text{class of proposed estimators } \hat{\delta} = c_1 \text{tr}(A) - c_2 p |A|^{1/p}, \quad (1.5)$$

where c_1 is a known constant and c_2 is a number such that $\hat{\delta}$ is non-negative.

In this paper we prove a dominance theorem that the class of proposed estimators $\hat{\delta}$ dominates the class of usual estimators $\tilde{\delta}$ of $\delta = \text{tr}(\Lambda)$ in the sense of smaller risk i.e.

$$R(\hat{\delta}, \delta) = E(\hat{\delta} - \delta)^2 < R(\tilde{\delta}, \delta) = E(\tilde{\delta} - \delta)^2.$$

2. SOME PRELIMINARIES

We now prove two lemmas on the expectation of the Wishart matrix based on a sample (not necessarily random) from the scale matrix of the multivariate normal distributions which will be required in the sequel. It follows from (1.1) that

$$A | \tau \sim W_p(n, \tau^2 \Lambda),$$

where for given τ , $W_p(n, \tau^2 \Lambda)$ is the usual Wishart distribution with parameters $n = N - 1$ and $\tau^2 \Lambda$, and consequently, the p.d.f. of the Wishart matrix based on the scale mixture of the multivariate normal model is given by

$$m(A) = \int_0^\infty \frac{|\tau^2 \Lambda|^{-n/2} |A|^{(n-p-1)/2}}{2^{np/2} \Gamma_p(n/2)} \exp\left(-\frac{1}{2} \text{tr}((\tau^2 \Lambda)^{-1} A)\right) h(\tau) d\tau, A > 0, \quad n \geq p \quad (2.1)$$

where

$$\Gamma_p(\alpha) = \pi^{p(p-1)/4} \prod_{i=1}^p \Gamma\left(\alpha - \frac{1}{2}(i-1)\right), \alpha > \frac{p-1}{2}. \quad (2.2)$$

LEMMA 2.1. Let the sum of product matrix (Wishart matrix) A

Let $\delta = \text{tr}(\Lambda)$. In estimating δ by an estimator $u(A)$, where A is the sample sum of product matrix, we consider the following loss function

$$L(u(A), \delta) = (u(A) - \delta)^2 \quad (1.3)$$

where $u(A)$ is any estimator of $\text{tr}(\Lambda)$. The risk function is defined as usual by taking expectation over the loss function i.e.

$$R(u(A), \delta) = E[L(u(A), \delta)].$$

In particular, consider the following two classes of non-negative estimators of $\delta = \text{tr}(\Lambda)$

$$\text{class of usual estimators } \tilde{\delta} = c_1 \text{tr}(A), \quad (1.4)$$

$$\text{class of proposed estimators } \hat{\delta} = c_1 \text{tr}(A) - c_2 p |A|^{1/p}, \quad (1.5)$$

where c_1 is a known constant and c_2 is a number such that $\hat{\delta}$ is non-negative.

In this paper we prove a dominance theorem that the class of proposed estimators $\hat{\delta}$ dominates the class of usual estimators $\tilde{\delta}$ of $\delta = \text{tr}(\Lambda)$ in the sense of smaller risk i.e.

$$R(\hat{\delta}, \delta) = E(\hat{\delta} - \delta)^2 < R(\tilde{\delta}, \delta) = E(\tilde{\delta} - \delta)^2.$$

2. SOME PRELIMINARIES

We now prove two lemmas on the expectation of the Wishart matrix based on a sample (not necessarily random) from the scale matrix of the multivariate normal distributions which will be required in the sequel. It follows from (1.1) that

$$A | \tau \sim W_p(n, \tau^2 \Lambda),$$

where for given τ , $W_p(n, \tau^2 \Lambda)$ is the usual Wishart distribution with parameters $n = N - 1$ and $\tau^2 \Lambda$, and consequently, the p.d.f. of the Wishart matrix based on the scale mixture of the multivariate normal model is given by

$$m(A) = \int_0^\infty \frac{|\tau^2 \Lambda|^{-n/2} |A|^{(n-p-1)/2}}{2^{np/2} \Gamma_p(n/2)} \exp\left(-\frac{1}{2} \text{tr}((\tau^2 \Lambda)^{-1} A)\right) h(\tau) d\tau, A > 0, \quad n \geq p \quad (2.1)$$

where

$$\Gamma_p(\alpha) = \pi^{p(p-1)/4} \prod_{i=1}^p \Gamma\left(\alpha - \frac{1}{2}(i-1)\right), \alpha > \frac{p-1}{2}. \quad (2.2)$$

LEMMA 2.1. Let the sum of product matrix (Wishart matrix) A

have the p.d.f. given by (2.1). Then the r th moment of $|A|$ is given by

$$E(|A|^r) = 2^{pr} \frac{\Gamma_p(n/2 + r)}{\Gamma_p(n/2)} |\Lambda|^r E(\tau^{2pr}).$$

PROOF. If $A \sim W(n, \Lambda)$, it is well known (see for example Muirhead, 1986) that

$$E_w(|A|^r) = 2^{pr} \frac{\Gamma_p(n/2 + r)}{\Gamma_p(n/2)} |\Lambda|^r$$

where E_w is the expectation over usual Wishart matrix. Hence it follows that for any integer r

$$\begin{aligned} E(|A|^r) &= E[E_w(|A|^r | \tau)] \\ &= E \left[2^{pr} \frac{\Gamma_p(n/2 + r)}{\Gamma_p(n/2)} |\tau^2 \Lambda|^r \right] \\ &= 2^{pr} \frac{\Gamma_p(n/2 + r)}{\Gamma_p(n/2)} |\Lambda|^r E(\tau^{2pr}). \quad \square \end{aligned}$$

LEMMA 2.2. Consider the p.d.f. given by (2.1). Then for any real number k satisfying $n + 2k \geq 0$, the following result hold :

$$E(|A|^k A) = 2^{kp} (n + 2k) \frac{\Gamma_p(n/2 + k)}{\Gamma_p(n/2)} |\Lambda|^k \Lambda E(\tau^{2kp+2}).$$

PROOF. Since $A | \tau \sim W_p(n, \tau^2 \Lambda)$ where $\tau^{-2} \sim G(v/2, 2/v)$, it follows from Dey (1988, p. 140) that for any real number k

$$\begin{aligned} E(|A|^k A) &= E[E_w(|A|^k A | \tau)] \\ &= E \left[2^{kp} (n + 2k) \frac{\Gamma_p(n/2 + k)}{\Gamma_p(n/2)} |\tau^2 \Lambda|^k (\tau^2 \Lambda) \right] \\ &= 2^{kp} (n + 2k) \frac{\Gamma_p(n/2 + k)}{\Gamma_p(n/2)} |\Lambda|^k \Lambda E(\tau^{2kp+2}). \quad \square \end{aligned}$$

3. THE MAIN RESULTS

THEOREM 3.1. Consider the scale mixture of the multivariate normal distributions given by (1.1). Then with respect to the loss function given by (1.3) the class of proposed estimators $\hat{\delta}$ defined by (1.5) dominates over the class of usual estimators $\tilde{\delta}$ defined by (1.4) in the sense of smaller risk for any c_2 satisfying any of the following two conditions :

1. For a given c_1 satisfying

$$c_1 < \frac{E(\tau^2)}{E(\tau^4)} \frac{p}{np+2} \text{ (i.e. } d^* < 0),$$

$$c_2 \text{ satisfies } d^* < c_2 < 0 \quad (3.1)$$

where

$$d^* = \left(c_1 \frac{np+2}{p} - \frac{E(\tau^2)}{E(\tau^4)} \right) \frac{\Gamma_p(n/2+1/p)}{\Gamma_p(n/2+2/p)}. \quad (3.2)$$

2. For a given c_1 satisfying

$$c_1 > \frac{E(\tau^2)}{E(\tau^4)} \frac{p}{np+2} \text{ (i.e. } d^* > 0),$$

$$c_2 \text{ satisfies } 0 < c_2 < d^* \quad (3.3)$$

where d^* is given by (3.2).

PROOF. Let $D(\Lambda, c_2) = R(\hat{\delta}, \delta; c_2) - R(\tilde{\delta}, \delta)$. Then it is easy to show that

$$D(\Lambda, c_2) = E[(\hat{\delta} - \delta)^2 - (\tilde{\delta} - \delta)^2] = E[\hat{\delta}^2 - \tilde{\delta}^2 - 2(\hat{\delta} - \tilde{\delta})\delta].$$

Hence by virtue of

$$\hat{\delta}^2 = \{\tilde{\delta} - c_2 p \mid A \mid^{1/p}\}^2 = \tilde{\delta}^2 - 2c_2 p \tilde{\delta} \mid A \mid^{1/p} + c_2^2 p^2 \mid A \mid^{2/p},$$

we have

$$D(\Lambda, c_2) = E[-2c_2 p \tilde{\delta} \mid A \mid^{1/p} + c_2^2 p^2 \mid A \mid^{2/p} - 2(-c_2 p \mid A \mid^{1/p}) \text{tr}(\Lambda)]$$

$$= p E[-2c_1 c_2 \mid A \mid^{1/p} \text{tr}(\Lambda) + c_2^2 p \mid A \mid^{2/p} + 2c_2 \mid A \mid^{1/p} \text{tr}(\Lambda)].$$

It then follows from Lemma 2.1 and Lemma 2.2 that

$$D(\Lambda, c_2) = -2c_1 c_2 \text{tr} \left[(n+2/p) E(\tau^4) \frac{\Gamma_p(n/2+1/p)}{\Gamma_p(n/2)} \mid \Lambda \mid^{1/p} \Lambda \right]$$

$$+ 2c_2 \left[2E(\tau^2) \frac{\Gamma_p(n/2+2/p)}{\Gamma_p(n/2)} \mid \Lambda \mid^{1/p} \right] \text{tr}(\Lambda)$$

$$+ c_2^2 p \left[4E(\tau^4) \frac{\Gamma_p(n/2+2/p)}{\Gamma_p(n/2)} \mid \Lambda \mid^{2/p} \right]$$

$$= 4c_2 p E(\tau^4) \frac{\mid \Lambda \mid^{2/p}}{\Gamma_p(n/2)} \times$$

$$\times \left[-c_1 \frac{np+2}{p} + \frac{E(\tau^2)}{E(\tau^4)} \right] \Gamma_p(n/2+1/p) \frac{tr(\Lambda/p)}{|\Lambda|^{1/p}} + c_2 \Gamma_p(n/2+2/p) \Big].$$

Hence

$$D(\Lambda, c_2) = 4p \frac{\Gamma_p(n/2+2/p)}{\Gamma_p(n/2)} |\Lambda|^{2/p} E(\tau^4) c_2 \left(c_2 - \frac{tr(\Lambda/p)}{|\Lambda|^{1/p}} d^* \right) \quad (3.4)$$

where d^* is given by (3.2).

As stated earlier, in order that $\hat{\delta}$ dominates $\tilde{\delta}$ it is necessary and sufficient that $D(\Lambda, c_2) < 0$. However, $D(\Lambda, c_2) < 0$ if and only if

$$\frac{tr(\Lambda/p)}{|\Lambda|^{1/p}} d^* < c_2 < 0, \text{ or } 0 < c_2 < \frac{tr(\Lambda/p)}{|\Lambda|^{1/p}} d^*.$$

The above conditions involve $tr(\Lambda)$ and $|\Lambda|$ which are unknown quantities. Now let $\xi_1, \xi_2, \dots, \xi_p$ be the characteristic roots of Λ . Then $tr(\Lambda/p) = (\xi_1 + \xi_2 + \dots + \xi_p)/p$ and $|\Lambda|^{1/p} = (\xi_1 \xi_2 \dots \xi_p)^{1/p}$. It then follows from the well-known inequality between arithmetic mean and geometric mean that $tr(\Lambda/p) \geq |\Lambda|^{1/p}$. Hence $\hat{\delta}$ dominates $\tilde{\delta}$ if $d^* < c_2 < 0$, or $0 < c_2 < d^*$. This means that

$$d^* < 0 \text{ if and only if } c_1 < \frac{E(\tau^2)}{E(\tau^4)} \frac{p}{np+2},$$

$$\text{while } d^* > 0 \text{ if and only if } c_1 > \frac{E(\tau^2)}{E(\tau^4)} \frac{p}{np+2}.$$

Hence the class of proposed estimators $\hat{\delta}$ dominates the class of usual estimators $\tilde{\delta}$ if c_2 satisfies the conditions mentioned in the theorem. Thus the theorem is proved. $\square$

It may be remarked that for the case when

$$c_1 = \frac{E(\tau^2)}{E(\tau^4)} \frac{p}{np+2},$$

we have $d^* = 0$. In this case it is seen from (3.4) that $D(\Lambda, c_2) \geq 0$ so that there exists no proposed estimator $\hat{\delta}$ dominating the usual estimator $\tilde{\delta}$. However, $D(\Lambda, c_2) = 0$ only if $c_2 = 0$ in which case the two estimators coincide.

We now look for narrower range of c_2 (compared to the range of c_2 obtained in Theorem 3.1) in which the risk of the class of proposed estimators will be much lower than that of the class of usual estimators. It is seen from (3.4) that for given Λ , the Risk Difference

$D(\Lambda, c_2)$ is a polynomial of degree 2 in c_2 , and it is minimized at

$$c_2 = \frac{\text{tr}(\Lambda/p)}{|\Lambda|^{1/p}} \frac{d^*}{2}$$

Now by virtue of the inequality between arithmetic mean and geometric mean of the characteristic roots of Λ , it follows that the optimal value of c_2 satisfies the following conditions :

$$d^* < c_2 \leq d^*/2 \text{ if } d^* < 0 \quad (3.5)$$

and

$$d^*/2 \leq c_2 < d^* \text{ if } d^* > 0. \quad (3.6)$$

In the light of the above results we have the following version of Theorem 3.1 narrowing down the range of c_2 .

THEOREM 3.2. *Consider the scale mixture of the multivariate normal distributions given by (1.1). Then with respect to the loss function given by (1.3) the class of proposed estimators $\hat{\delta}$ defined by (1.5) dominates over the class of usual estimators $\tilde{\delta}$ defined by (1.4) in the sense of smaller risk for any c_2 satisfying any of the following two conditions :*

1. *For a given c_1 satisfying*

$$c_1 < \frac{E(\tau^2)}{E(\tau^4)} \frac{p}{np+2} \text{ (i.e. } d^* < 0),$$

c_2 satisfies $d^ < c_2 \leq d^*/2$ where d^* is given by (3.2).*

2. *For a given c_1 satisfying*

$$c_1 > \frac{E(\tau^2)}{E(\tau^4)} \frac{p}{np+2} \text{ (i.e. } d^* > 0),$$

c_2 satisfies $d^/2 \leq c_2 < d^*$ where d^* is defined by (3.2).*

REFERENCES

1. D. K. Dey (1988), Simultaneous estimator of eigenvalues, *Ann. Inst. Statist. Math.*, Vol. 40, pp. 137-147.
2. D. K. Dey (1990), Estimation of scale parameters in mixture distributions, *Canad. J. Statist.*, Vol. 18, pp. 171-178.
3. E. F. Fama (1965), The behaviour of stock market prices, *J. Business*, Vol. 38, pp. 34-105.
4. B. Mandelbrot (1963), The variation of certain speculative prices, *J. Business*, Vol. 36, pp. 394-419.
5. R. J. Muirhead (1986), A note on some Wishart expectations, *Metrika*, Vol. 33, pp. 247-251.

6. I. Olkin and J. B. Selliah (1977), Estimating covariance in a multivariate normal distribution, in *Statistical Decision Theory and Related Topics II*, (eds. S. S. Gupta and D. Moore), pp. 313-326, Academic Press, New York.
7. R. S. Singh (1991), James-Stein rule estimators in linear regression models with multivariate t -distribution error, *Austr. J. Statist.*, Vol. 33, pp. 145-158.
8. B. C. Sutradhar and M. M. Ali (1986), Estimation of the parameters of the regression model with a multivariate t error variable, *Commun. Statist. Theor. Math.*, Vol. 15, pp. 429-450.
9. A. Zellner (1976), Bayesian and non-Bayesian analysis of the regression model with multivariate Student- t error term, *J. Amer. Statist. Assoc.*, Vol. 71, pp. 400-405 (correction, Vol. 71, p. 1000).

Received October, 1994