# King Fahd University of Petroleum & Minerals

## DEPARTMENT OF MATHEMATICAL SCIENCES

Technical Report Series

TR 234

July 1998

# Estimation of the trace of the scaled convariance matrix of a multivariate t-model using a power transformation

A.H. Joarder, Sarjinder Singh

DHAHRAN 31261 ● SAUDI ARABIA ● www.kfupm.edu.sa/math/ ● E-mail: mathdept@kfupm.edu.sa

# Estimation of the trace of the scaled covariance matrix of a multivariate t-model using a power transformation

A.H. Joarder
Department of Mathematical Sciences
King Fahd University of Petroleum and Minerals
Dhahran, Saudi Arabia 31261

and

Sarjinder Singh
Australian Bureau of Statistics
PO Box 10, Belconen
ACT, Australia 2616

## Abstract

The trace of the scaled covariance matrix of the multivariate $t$-distribution is considered for estimation using a power transformation. The proposed estimator always dominates the usual maximum likelihood estimator in the sense of having smaller risk under a quadratic loss function. The dominance behaviour is proved analytically as well as computationally by using Monte-Carlo simulation.

*Key Words and Phrases: scale matrix; estimation of trace of scaled covariance matrix; multivariate t-distribution, Monte-Carlo simulation*

## 1. Introduction

In this paper we consider the estimation of the trace of the scaled covariance matrix of the multivariate $t$-distribution. The trace of the scaled covariance matrix representing the total variation in the component variables is important in many statistical analyses e.g. in principal component analysis.

We assume $N$ $p$-dimensional ($p \geq 2$) random vectors (not necessarily independent) $X_1, X_2, \ldots, X_N$ have a joint p.d.f. (probability density function) given by

$$
\begin{aligned}
&f(x_1, x_2, \ldots, x_N) \\
&= \frac{|\Sigma|^{-N/2}}{C(\nu,\ Np)(2\pi)^{Np/2}} \left( 1 + \sum_{j=1}^{N} (x_j - \mu)'(\nu\Sigma)^{-1}(x_j - \mu) \right)^{-(\nu+Np)/2} \qquad (1.1)
\end{aligned}
$$

where the normalizing constant $C(\nu,\ Np)$ is given by

$$
C(\nu,\ Np) = \frac{(\nu/2)^{Np/2}\,\Gamma(\nu/2)}{\Gamma((\nu + Np)/2)} \qquad (1.2)
$$

where $x_j = (x_{1j}, x_{2j}, \ldots, x_{pj})'$, $\mu$ is an unknown $p \times 1$ vector of location parameters, $\Sigma$ is a $p \times p$ unknown positive definite matrix of scale parameters and $\nu$ ($> 4$) is assumed to be a known positive constant. Each $p$-dimensional random vector $X_j$ ($j = 1, 2, \ldots, N$) has a multivariate $t$-distribution with mean vector $\mu$ and covariance matrix $\Sigma/(1 - 2/\nu)$.

The joint p.d.f. in (1.1) represents the multivariate $t$-model; it has been considered, among others, by Zellner (1976) in the context of stock market problems. Lange, Little and Taylor (1989) applied multivariate $t$-model to a variety of situations. For a discussion of the justification of the multivariate $t$-model the reader is referred to Kelejian and Prucha (1985) and Joarder and Singh (1997).

Olkin and Selliah (1977) considered the estimation of the trace of the covariance matrix of the multivaraite normal distribution under a weighted squared error loss function. Joarder (1995) considers estimation of the trace of the scale matrix of the multivariate t-distribution under a squared error loss function following Dey (1988). It may be remarked that we are estimating the trace of the scale matrix $\Sigma$ (which is hereinafter refered to scaled covariance matrix) instead of the trace of the covariance matrix since the scale matrix $\Sigma$ determines the covariance matrix upto a known constant $\nu/(\nu - 2)$.

The estimation of the trace of the covariance matrix of the multivariate normal distribution by the maximum likelihood method is meaningful because the large sample properties of maximum likelihood estimators follow from independently and identically distributed observations. This is not necessarily true for the model considered in (1.1) unless $\nu \to \infty$. Hence we use MSE (Mean Square Error Criteron)to estimate the trace of scaled covariance matrix.

Let $\delta = tr(\Sigma)$ be the trace of the scaled covariance matrix $\Sigma$ to be estimated. The usual estimator of $\delta$ is given by $\tilde{\delta} = c_0\ tr(A)$ where

$$A = \sum_{j=1}^{N}(X_j - \bar{X})(X_j - \bar{X})'$$

with $\bar{X} = (\bar{X}_1, \bar{X}_2, \ldots, \bar{X}_p)'$ and $\bar{X}_i = \sum_{j=1}^{N} X_{ij}/N, (i = 1, 2, \ldots, p)$, is the sample sum of product matrix (Wishart matrix) based on the multivariate $t$-model. In this paper, we propose an improved estimator of $\delta = tr(\Sigma)$ based on a power transformation of the Wishart matrix.

## 2. Proposed Estimation Strategy

The proposed estimator is given by

$$\hat{\delta} = c_0\ tr(A)\left(\frac{|A|^{1/p}}{tr(A)/p}\right)^c \tag{2.1}$$

where $c_0$ is a known positive constant, $c$ is a constant chosen so that the mean square error of the proposed estimator $\hat{\delta}$ is minimized. Since $|A|^{1/p}/tr(A/p)$ is usually less than 1, by using the binomial expansion up to the first order of approximaton, we have

$$\hat{\delta} = c_0\ tr(A) + c_0\ c\ (p|A|^{1/p} - tr(A)). \tag{2.2}$$

In order to appraise the efficiency of the proposed estimator relative to the usual estimator given by

$$\tilde{\delta} = c_0\ tr(A) \tag{2.3}$$

we use the mean square error (MSE) criterion. The estimators are based on the sample sum of product matrix $A$ having probability density function

$$g(A) = \frac{|\Sigma|^{-n/2}\ |A|^{(n-p-1)/2}}{C(\nu,\ np)\ 2^{np/2}\ \Gamma_p(n/2)}\ (1 + tr(\Sigma^{-1}A)/\nu)^{-(\nu+np)/2} \tag{2.4}$$

where $A > 0, n = N - 1 \geq p$, and $\Gamma_p(\alpha)$ is the generalized gamma function defined by

$$\Gamma_p(\alpha) = \pi^{p(p-1)/4} \prod_{i=1}^{p} \Gamma((2\alpha - i + 1)/2)$$

(see e.g. Joarder and Mahmood, 1998).

## 3. The Main Results

The main results are presented in this section in the form of two theorems.

**Theorem 3.1** The MSE of the usual estimator $\tilde{\delta}$ is given by

$$MSE(\tilde{\delta}) = \left[\frac{n\ c_0}{1-2/\nu}\left(\frac{nc_0}{1-4/\nu} - 2\right) + 1\right](tr\Sigma)^2 + \frac{2n\ c_0^2}{(1-2/\nu)(1-4/\nu)}\ tr(\Sigma^2).$$

**Proof.** We have

$$\begin{aligned} MSE(\tilde{\delta}) =& E(\tilde{\delta} - \delta)^2 \\ =& c_0^2 E(tr\ A)^2 + \delta^2 - 2\delta c_0 E(tr\ A). \end{aligned}$$

The expected value of the Wishart matrix is given by $E(A) = n\Sigma/(1 - 2/\nu)$. Then the proof is immediate by Joarder (1998).

**Theorem 3.2** The MSE of the proposed estimator $\hat{\delta}$ is given by

$$MSE(\hat{\delta}) = MSE(\tilde{\delta}) + c\beta_1 + c^2\beta_2 \tag{3.1}$$

where

$$\beta_1 = 2E[(c_0\ tr(A) - \delta)(p|A|^{1/p} - tr(A))] \tag{3.2}$$

$$\text{and } \beta_2 = c_0^2\ E[p|A|^{1/p} - tr(A)]^2. \tag{3.3}$$

**Proof:** The MSE of the proposed estimator is given by

$$MSE(\hat{\delta}) = E\left[c_0\ tr(A)\left(\frac{|A|^{1/p}}{tr(A/p)}\right)^c - \delta\right]^2$$

which, up to a first order approximation, is given by

$$\begin{aligned} MSE(\hat{\delta}) =& E[c_0\ tr(A) + cc_0(p|A|^{1/p} - tr(A)) - \delta]^2 \\ =& MSE(\tilde{\delta}) + c\beta_1 + c^2\beta_2 \end{aligned}$$

where $\beta_1$ and $\beta_2$ are defined in the theorem.

Using the results in Joarder (1998), it follows that

$$\begin{aligned} \beta_1 =& \frac{2}{1-2/\nu}\left[2\left(\frac{2c_0\ (n/2+1/p}{1-4/\nu} - p\right)\ \frac{\Gamma_p(n/2+1/p)}{\Gamma_p(n/2)}\ |\Sigma|^{1/p}\ tr(\Sigma) \right. \\ & \left. -n\left(\frac{n\ c_0}{1-4/\nu} - 1\right)\ (tr\Sigma)^2 - \frac{4n\ c_0}{1-4/\nu}\ tr(\Sigma^2)\right] \end{aligned} \tag{3.4}$$

and

$$\begin{aligned} \beta_2 =& \frac{c_0^2}{(1-2/\nu)(1-4/\nu)}\left[\frac{4p\ |\Sigma|^{1/p}}{\Gamma_p(n/2)}\ \left(p\ \Gamma_p(n/2+2/p)|\Sigma|^{1/p} \right.\right. \\ & \left.\left. -2\ \Gamma_p(n/2+1/p)\ tr(\Sigma)\right) + n(n(tr\ \Sigma)^2 + 2\ tr(\Sigma^2))\right]. \end{aligned} \tag{3.5}$$

The MSE($\hat{\delta}$) minimizes at $c = -\beta_1/(2\beta_2)$ and the minimum value is given by

$$Min.MSE(\hat{\delta}) = MSE(\tilde{\delta}) - \beta_1^2/(4\beta_2). \tag{3.6}$$

This proves that the proposed estimator is always better than the usual estimator in the sense of having smaller mean square error. However the proposed estimator is not operational as it involves the unknown parameter $c$. The operational value of c is given by

$$\hat{c} = \frac{-\hat{\beta}_1}{2\hat{\beta}_2} \tag{3.7}$$

which can be obtained by estimating $\beta_1$ and $\beta_2$ given by (3.4) and (3.5) respectively. The estimation of $\beta_1$ and $\beta_2$ can be done by replacing $\Sigma$ by the the usual estimator $c_0A$. If $c_0 = 1/(n+1)$, the estimators would be maximum likelihood estimators. Finally we suggest to use the following estimator of $\delta = tr(\Sigma)$

$$\hat{\delta}^* = c_0\ tr(A)\left(\frac{|A|^{1/p}}{tr(A)/p}\right)^{\hat{c}} \tag{3.8}$$

with $\hat{c}$ given by (3.7) where $\hat{\beta}_1$ and $\hat{\beta}_2$ are the estimators of $\beta_1$ and $\beta_2$ respectively.

## 3. Empirical Study

A Monte Carlo study has been carried out by generating 100 Wishart matrices from different Wishart distributions based on the multivariate t-model. We estimated the value of $c$ for each matrix and observed that $\hat{c}$ lies between 0 and 1. Thus the binomial expansion in (2.2) of the proposed estimator given by (2.1) holds good since it requires $c$ to be positive. The Percent Relative Efficiency of the proposed estimator obtained by putting $\hat{c}$ in (2.1) is presented in Table 1 for $n = 25$ and $p = 3$. The following three diagonal matrices

$$\Sigma = Diag(1,1,1), \ \Sigma = Diag(4,2,1), \ \Sigma = Diag(25,1,1)$$

considered in Table 1 are due to Dey (1988).

**Table 1.** Percent Relative Efficiency of the proposed estimator over the usual estimator.

| $\nu$ | $Diag(1,1,1)$ | $Diag(4,2,1)$ | $Diag(25,1,1)$ |
|---|---|---|---|
| 05 | 105.32 | 130.31 | 153.90 |
| 10 | 102.13 | 117.56 | 148.76 |
| 15 | 101.53 | 112.07 | 127.15 |

A general scale matrix of order $3\times3$ given by $\Sigma = ((\sigma_{ik}))$ where $\sigma_{11} = 25, \sigma_{12} = -2, \sigma_{13} = 4, \sigma_{22} = 4, \sigma_{23} = 1, \sigma_{33} = 9$ is also considered. In this case, the Percent Relative Efficiency of the proposed estimator over the usual one is given by 170.44, 165.57 and 158.11 for $\nu = 5$, 10 and 15 respectively.

Thus we conclude that the proposed estimator $\hat{\delta}^*$ in (3.8) remains always more efficient than the usual estimator $\tilde{\delta}$ given by (2.3) analytically as well as emperically.

### ACKNOWLEDGEMENTS

The first draft of the paper was written while both the authors had been working at Department of Econometrics, Monash University, Australia. This research has been supported in part by a grant to the first author from Faculty of Business and Economics Research Committee, Monash University, Australia.

## REFERENCES


Dey, D.K. (1988). Simultaneous estimation of eigenvalues. *Ann. Inst. Statist. Math.,* **40,** 137–147.

Joarder, A.H. (1995). Estimation of the trace of the scale matrix of a multivariate $t$-model. *Proceeding of the Econometrics Conference at Monash University* , Monash University, Australia, 467–474.

Joarder, A.H. (1998). Some useful Wishart expectations based on the multivariate $t$-model. **39**, 223–229.

Joarder, A.H. and Mahmood, M.(1998). On the statistical theory of a t-population. Technical Report No. 231, Department of Mathematical Sciences. King Fahd University of Petroleum and Minerals. Saudi Arabia.

Joarder, A.H. and Singh, Sarjinder (1996). Estimation of the trace of the scale matrix of a multivariate $t$-model using regression estimator. **28**, 161–168.

Kelejian, H.H. and Prucha, I.R. (1985). Independent or uncorrelated disturbances in linear regression: An illustration of difference. *Economic Letters,* **19**, 35-38.

Olkin, I. and Selliah, J.B. (1977). Estimating covariance in a multivariate normal distribution. *Statistical Decision Theory and Related Topics II,* (eds. S.S. Gupta and D. Moore), 313–326, Academic Press, New York.

Lange, K.L., Little, R.J.A. and Taylor, J.M.G (1989). Robust statistical modelling using the t distribution. *J. Amer. Statist. Assoc.*, **84**, 881–896.

Zellner, A. (1976). Bayesian and non-Bayesian analysis of the regression model with multivariate Student-t error term. *J. Amer. Statist. Assoc.,* **71**, 400–405.