King Fahd University of Petroleum & Minerals

DEPARTMENT OF MATHEMATICAL SCIENCES

Technical Report Series

TR 222

June 1997

Estimation of the Mean Vector of the Multivariate Student-t Populatin Under Uncertain Prior Information

Shahjahan Khan

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

ESTIMATION OF THE MEAN VECTOR OF THE MULTIVARIATE STUDENT-t POPULATION UNDER UNCERTAIN PRIOR INFORMATION

by

Shahjahan Khan
Department of Mathematical Sciences
King Fahd University of Petroleum and Minerals
Dhahran 31261, Saudi Arabia

ABSTRACT

This paper provides different estimators of the mean vector of a multivariate Student-t population in the presence of uncertain prior information. The usual maximum likelihood estimator, restricted estimator and preliminary test estimators are obtained. For each of the estimators the bias and the risk under the squared error loss function is derived. The relative performance of the estimators are studied based on the unbiasedness and minimum risk criteria. Some remarks and recommendations are provided as to the choice of the estimators under different conditions.

AMS 1990 Subject Classification: Primary 62F30, Secondary 62H12 and 62F10.

Keywords and Phrases: Maximum likelihood, restricted and preliminary test estimators; Student-t, normal, inverted gamma and distributions; noncentral distributions; bias and risk functions; optimal level of significance.

1 INTRODUCTION

Consider a random sample $X_{11}, X_{12}, \ldots, X_{1n_1}$, of size n_1 from a p-dimensional Student-t population with mean vector $\boldsymbol{\mu}_1$, covariance matrix $\sigma^2 I_p$ and shape parameter ν. We wish to estimate $\boldsymbol{\mu}_1$ while another sample $X_{21}, X_{22}, \ldots, X_{2n_2}$ of size n_2 is available from the same study or from a separate study of the same variables with mean vector $\boldsymbol{\mu}_2$ but same spread and shape parameters. Based on the experience in dealing with the variables or from the nature of the data generating process it is suspected that the population mean vectors of the populations from which the samples were selected are equal. Such an *uncertain prior information* about the value of the mean vectors can be expressed in the form of a null hypothesis, $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$, and could be used in the process of estimation of $\boldsymbol{\mu}_1$. The informative data from the second sample can also be incorporated in an appropriate fashion to define a 'better' estimator of $\boldsymbol{\mu}_1$.

In addition to the usual maximum likelihood estimator (mle) of $\boldsymbol{\mu}_1$ based on the first sample, we define the pooled estimator of $\boldsymbol{\mu}_1$ as the weighted mean of the combined samples. The preliminary test estimator of $\boldsymbol{\mu}_1$ is also defined as a function of the mle of $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ as well as the test statistic appropriate to testing the null hypothesis, $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$. The above three estimators of $\boldsymbol{\mu}_1$ are compared in terms of the criteria of unbiasedness and risk under quadratic loss function. The relative performances of the estimators are discussed under various conditions. The risk analysis provides a clear picture of the comparative behavior of the estimators that can be used to determine as to which estimator should be used in what situation.

Throughout the paper, the Student-t distribution is viewed as a mixture distribution of the normal and inverted Gamma distributions that simplifies the computations and derivations of the paper. However, the results of this paper are valid for the normal model as a special case when the shape parameter of the Student-t model grows very large. Moreover, since the results are derived for any arbitrary value of the shape parameter it also covers the Cauchy distribution on the other extreme in terms of the value of the shape parameter.

2 The Student-t Model

The normal distribution is not robust and can't be used as a sole model for the distribution of errors. Fisher (1956) discussed the danger of high sensitivity of the mathematical model with normal errors for slight departure from its assumptions. No wonder Fisher (1960, p. 37) analyzed Darwin's data (cf. Box and Tiao, 1990, p. 153) using non-normal distribution. The failure of normal distribution to model the data with 'outlier' and heavier or flat tails has generated a wide range of research in the area of non-normal distributions. Most of these studies are based on the Student-t model or elliptical models. A brief list of the literature on this topic is found in Chmielewiski (1981). Other recent work includes Fang and Zhang (1980), Khan and Haq (1990), Fang and Anderson (1990), Gupta and Vargava (1993), Anderson (1993), and Khan and Saleh (1997). Fraser (1979) demonstrated that the results for the linear models based on the Student-t distribution of the errors are applicable to the normal models, but not the vice-versa. Prucha and Kelejian (1984) critically discussed the reasons for the non-suitability of the normal distribution in many problems, and advocated the Student-t model as a 'better' alternative. As a matter of fact, the Student-t distribution is a more 'typical' member than the normal distribution of the elliptical family of distributions and hence is capable of modelling many symmetrical distributions, including the normal distribution as a limiting case. Therefore, the model considered in this paper is more general and covers a wider class of distributions for different values of the shape parameter.

Let $\boldsymbol{X}_i$ be a p-variate normal variable with mean vector $\boldsymbol{\mu}_i$ and covariance matrix

$\tau^2 I_p$ for $i = 1, 2$. Assume that τ follows an Inverted Gamma distribution with the shape parameter ν and density function

$$p(\tau;\nu) = \left(\frac{2}{\nu/2}\right)\left(\frac{\nu}{2}\right)^{\nu/2} \tau^{-(\nu+1)} e^{-\frac{\nu}{2\tau^2}}, \qquad \tau > 0. \tag{2.1}$$

Then we write, $[\boldsymbol{X}_i|\tau] \sim N_p\ (\boldsymbol{\mu}_i; \tau^2 I_p)$ and $\tau \sim IG(\nu, 1)$. It is well known that the mixture distribution of $\boldsymbol{X}_i$ is Student-t and is obtained by completing the integral

$$p(\boldsymbol{x}_i; \boldsymbol{\mu}, \Sigma, \nu) = \int_{\tau=0}^{\infty} N_p(\boldsymbol{\mu}_i, \tau^2 I_p) IG(\nu, 1) d\tau \tag{2.2}$$

where $\Sigma = \dfrac{\nu}{\nu-2} I_p = \sigma^2 I_p$. Thus the density function of $\boldsymbol{X}_i$ is obtained as

$$p(\boldsymbol{x}_i; \boldsymbol{\mu}, \Sigma, \nu) = k_i(\nu, p) \left|\Sigma\right|^{-1/2} \left[\nu + (\boldsymbol{x}_i - \boldsymbol{\mu}_i)' \Sigma^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu}_i)\right]^{-\frac{\nu+p}{2}} \tag{2.3}$$

where ν is the degree of freedom (df) of the Student-t distribution and

$$k_i(\nu, p) = \frac{\Gamma\left(\frac{\nu+p}{2}\right)\nu^{\nu/2}}{(\pi)^{p/2}\Gamma\left(\frac{\nu}{2}\right)}$$

is the normalizing constant. We write $\boldsymbol{X}_i \sim t_p(\boldsymbol{\mu}_i, \Sigma, \nu)$. For the above model, we derive the maximum likelihood estimator of $\boldsymbol{\mu}_i$ and Σ, while a method of moment estimator of ν is proposed by Singh (1988).

For the random sample $\boldsymbol{X}_{11}, \boldsymbol{X}_{12}, \ldots, \boldsymbol{X}_{1n_1}$ from the first population, the likelihood function is given by

$$p(\boldsymbol{x}_{11}, \ldots, \boldsymbol{x}_{1n_1}; \boldsymbol{\mu}_1, \Sigma, \nu) = k_{n_1}(\nu, p) \left[\nu + \sum_{j=1}^{n_1} (\boldsymbol{x}_{1j} - \boldsymbol{\mu}_1)' \Sigma^{-1} (\boldsymbol{x}_{1j} - \boldsymbol{\mu}_1)\right]^{-\frac{\nu+n_1 p}{2}} \tag{2.4}$$

where $k_{n_1}(\nu, p) = \dfrac{\Gamma\left(\frac{\nu+n_1 p}{2}\right)\nu^{\nu/2}}{(\pi)^{\frac{n_1 p}{2}}\Gamma\left(\frac{\nu}{2}\right)}$.

Note that $\boldsymbol{X}_{1j}$'s are dependent but uncorrelated, a typical property of the multivariate Student-t distribution (cf. Anderson 1993). Moreover, each $\boldsymbol{X}_{1j}$ marginally follows the multivariate Student-t distribution with mean $\boldsymbol{\mu}_1$ and covariance Σ for $j = 1, 2, \ldots, n_1$. The likelihood function of the second sample $\boldsymbol{X}_{21}, \boldsymbol{X}_{22}, \ldots, \boldsymbol{X}_{2n_2}$ can be obtained in a similar way. Also note that $\boldsymbol{X}_{2j} \sim t_p(\boldsymbol{\mu}_2, \Sigma, \nu)$ for $j = 1, 2, \ldots, n_2$, and $\boldsymbol{X}_{2j}$'s are dependent but uncorrelated.

Our primary objective is to estimate $\boldsymbol{\mu}_1$. First we wish to derive the maximum likelihood estimator for $\boldsymbol{\mu}_1$. Then we define another estimator of $\boldsymbol{\mu}_1$ as a weighted mean of the mle of $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$. Finally, we provide the preliminary test estimator of $\boldsymbol{\mu}_1$. This later estimator is a function of an appropriate test statistic. The bias and quadratic risk of the three estimators are obtained. The relative performances of the estimators are investigated and appropriate comparisons are provided under different conditions.

3 The Estimators and Test Statistic

The log-likelihood function of the two samples can be written as

$$\ln L(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \sigma^2 I_p) \propto -\frac{(n_1+n_2)p}{2} \ln \sigma^2 - \frac{\nu + (n_1+n_2)p}{2} \times$$
$$\ln\left[1 + \frac{1}{\nu\sigma^2}\left\{\sum_{j=1}^{n_1}(\boldsymbol{x}_{1j} - \boldsymbol{\mu}_1)'(\boldsymbol{x}_{1j} - \boldsymbol{\mu}_1) + \sum_{j=1}^{n_2}(\boldsymbol{x}_{2j} - \boldsymbol{\mu}_2)'(\boldsymbol{x}_{2j} - \boldsymbol{\mu}_2)\right\}\right] \quad (3.1)$$

Now $\dfrac{\partial \ln L(\cdot)}{\partial \boldsymbol{\mu}_1} = 0$, $\dfrac{\partial \ln L(\cdot)}{\partial \boldsymbol{\mu}_2} = 0$ and $\dfrac{\partial \ln L(\cdot)}{\partial \sigma^2} = 0$ yield the following normal equations:

$$\frac{\nu + (n_1+n_2)p}{\nu\sigma^2}\left[1 + \frac{1}{\nu\sigma^2}\left\{\sum_{j=1}^{n_1} Q_{1j} + \sum_{j=1}^{n_2} Q_{2j}\right\}\right]^{-1} \sum_{j=1}^{n_1}(\boldsymbol{x}_{1j} - \boldsymbol{\mu}_1) = 0 \quad (3.2)$$

$$\frac{\nu + (n_1+n_2)p}{\nu\sigma^2}\left[1 + \frac{1}{\nu\sigma^2}\left\{\sum_{j=1}^{n_1} Q_{1j} + \sum_{j=1}^{n_2} Q_{2j}\right\}\right]^{-1} \sum_{j=1}^{n_2}(\boldsymbol{x}_{2j} - \boldsymbol{\mu}_2) = 0 \quad (3.3)$$

$$\frac{(n_1+n_2)p}{\sigma^2} = \frac{\nu+(n_1+n_2)p}{\nu\sigma^4}\left[1+\frac{1}{\nu\sigma^2}\left\{\sum_{j=1}^{n_1} Q_{1j}+\sum_{j=1}^{n_2} Q_{2j}\right\}\right]^{-1} \times$$

$$\left[\sum_{j=1}^{n_1} Q_{1j}+\sum_{j=1}^{n_2} Q_{2j}\right] = 0 \tag{3.4}$$

where $Q_{ij} = \sum_{j=1}^{n_i}(\boldsymbol{x}_{ij}-\boldsymbol{\mu}_i)'(\boldsymbol{x}_{ij}-\boldsymbol{\mu}_i)$ for $i=1,2$.

Solving the normal equations we get the mle of $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$ and σ^2 as follows:

$$\tilde{\boldsymbol{\mu}}_1 = \overline{\boldsymbol{X}}_1,\ \tilde{\boldsymbol{\mu}}_2 = \overline{\boldsymbol{X}}_2 \tag{3.5}$$

$$\tilde{\sigma}^2 = \frac{1}{(n_1+n_2)p}\left[\sum_{i=1}^{2}\sum_{j=1}^{n_i}(\boldsymbol{X}_{ij}-\tilde{\boldsymbol{\mu}}_i)'(\boldsymbol{X}_{ij}-\tilde{\boldsymbol{\mu}}_i)\right]. \tag{3.6}$$

Under the null hypothesis, $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \boldsymbol{\mu}_0$, the restricted mle of σ^2 is obtained as

$$\hat{\sigma}^2 = \frac{1}{(n_1+n_2)p}\left[\sum_{i=1}^{2}\sum_{j=1}^{n_i}(\boldsymbol{x}_{ij}-\boldsymbol{\mu}_0)'(\boldsymbol{x}_{ij}-\boldsymbol{\mu}_0)\right]. \tag{3.7}$$

Now we can write the maximum of the unrestricted likelihood function as

$$L(\tilde{\boldsymbol{\mu}}_1,\tilde{\boldsymbol{\mu}}_2,\tilde{\sigma}^2) = k(\tilde{\sigma}^2)^{-\frac{(n_1+n_2)p}{2}}\left[1+\frac{(n_1+n_2)p}{\nu}\right]^{-\frac{\nu+(n_1+n_2)p}{2}} \tag{3.8}$$

and that of the restricted likelihood function as

$$L(\boldsymbol{\mu}_0,\hat{\sigma}^2) = k\left(\tilde{\sigma}^2\right)^{-\frac{(n_1+n_2)p}{2}}\left[1+\frac{(n_1+n_2)p}{\nu}\right]^{-\frac{(n_1+n_2)p}{2}} \tag{3.9}$$

where $k = \dfrac{\Gamma\left(\frac{\nu+(n_1+n_2)p}{2}\right)}{(\pi)^{\frac{(n_1+n_2)p}{2}}\Gamma\left(\frac{\nu}{2}\right)}$.

From (3.8) and (3.9) the likelihood ratio statistic becomes

$$\lambda = \left(\frac{\tilde{\sigma}^2}{\hat{\sigma}^2}\right)^{-\frac{(n_1+n_2)p}{2}} \tag{3.10}$$

which yields

$$(\lambda)^{2/(n_1+n_2)p} = \frac{\sum_{i=1}^{2}\sum_{j=1}^{n_i}(\boldsymbol{x}_{ij}-\tilde{\boldsymbol{\mu}}_i^2)'(\boldsymbol{x}_{ij}-\tilde{\boldsymbol{\mu}}_i)}{\sum_{i=1}^{2}\sum_{j=1}^{n_i}(\boldsymbol{x}_{ij}-\boldsymbol{\mu}_0)'(\boldsymbol{x}_{ij}-\boldsymbol{\mu}_0)}. \tag{3.11}$$

Now denoting the right hand side of (3.11) by T^2, it can be shown that under H_0 the statistic

$$F = \frac{(n_1 + n_2 - 2)p}{(n_1 + n_2 - p - 1)} T^2 \tag{3.12}$$

follows an F-distribution with p and m degrees of freedom with $m = (n_1 + n_2 - p - 1)$ (see Anderson, 4.2, p. 109).

Therefore, we can use the T^2 statistic to test $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$. As discussed by Anderson (1993) and many others, the above test is robust and is applicable to the entire family of elliptical models, not only for normal or Student-t model.

The foregoing derivations allow us to define three different estimators of $\boldsymbol{\mu}_1$. First, the *unrestricted estimator* (UE) of $\boldsymbol{\mu}_1$ is the mle, $\tilde{\boldsymbol{\mu}}_1 = \overline{X}_1$. Then the *restricted estimator* (RE) of $\boldsymbol{\mu}_1$ is defined as

$$\hat{\boldsymbol{\mu}}_1 = \frac{n_1 \tilde{\boldsymbol{\mu}}_1 + n_2 \tilde{\boldsymbol{\mu}}_2}{n_1 + n_2}, \tag{3.13}$$

the overall mean of the two samples.

Finally, the *preliminary test estimator* (PTE) of $\boldsymbol{\mu}_1$ is defined as

$$\hat{\boldsymbol{\mu}}_1{}^{pt} = \tilde{\mu}_1 I(T^2 > T_\alpha^2) + \hat{\boldsymbol{\mu}}_1 I(T^2 \leq T_\alpha^2) \tag{3.14}$$

where $I(A)$ is an indicator function of the set A, and T_α^2 is the value of T^2 statistic such that $P(T^2 > T_\alpha^2) = \alpha$. Clearly, $\hat{\boldsymbol{\mu}}_1^{pt}$ is a linear combination of UE and RE. In fact, it is a choice between $\tilde{\boldsymbol{\mu}}_1$ and $\hat{\boldsymbol{\mu}}_1$ depending on the rejection or acceptance of the null hypothesis for a given level of significance. Our study will include the investigation of the properties of the above three estimators under various conditions and using different criteria.

4 Derivation and Analysis of Bias

The criterion of unbiasedness is a well-known basis to judge the quality of an estimator. In this section, we investigate the three different estimators of $\boldsymbol{\mu}_1$ using this criterion. The bias of the UE is given by

$$B_1(\tilde{\boldsymbol{\mu}}_1; \boldsymbol{\mu}_1) = E(\tilde{\boldsymbol{\mu}}_1 - \boldsymbol{\mu}) = \mathbf{0}. \tag{4.1}$$

Thus the UE of $\boldsymbol{\mu}_1$ is an unbiased estimator.

Theorem 4.1. *The bias of the RE of $\boldsymbol{\mu}_1$ is given by*

$$B_2(\hat{\boldsymbol{\mu}}_1; \boldsymbol{\mu}_1) = E(\hat{\boldsymbol{\mu}}_1 - \boldsymbol{\mu}_1) = M\boldsymbol{\delta} \tag{4.2}$$

where $M = \dfrac{n_2}{n_1 + n_2}$ *and* $\boldsymbol{\delta} = (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)$.

Proof. The RE of $\boldsymbol{\mu}_1$ can be written as $\hat{\boldsymbol{\mu}}_1 = \dfrac{n_1}{n_1 + n_2}\tilde{\boldsymbol{\mu}}_1 + M\tilde{\boldsymbol{\mu}}_2$, and hence $\hat{\boldsymbol{\mu}}_1 - \boldsymbol{\mu}_1 = \dfrac{n_1}{n_1 + n_2}(\tilde{\boldsymbol{\mu}}_1 - \boldsymbol{\mu}_1) + M(\tilde{\boldsymbol{\mu}}_2 - \boldsymbol{\mu}_1)$. Therefore $E(\hat{\boldsymbol{\mu}}_1 - \boldsymbol{\mu}_1) = M\boldsymbol{\delta}$.

The RE of $\boldsymbol{\mu}_1$ is biased. The size of the bias grows unboundedly as $\boldsymbol{\delta}$ grows large, for given sample sizes. Thus the estimator based on both the samples does not perform better than the UE with respect to the unbiasedness criterion. The quadratic bias of $\hat{\boldsymbol{\mu}}_1$, $M^2\boldsymbol{\delta}'\boldsymbol{\delta} = QB_2$ can be investigated to study the mode of change in QB_2 with a change in the value of $\boldsymbol{\delta}$, the difference between the two unknown means. However, under H_0 the RE is unbiased, as $\boldsymbol{\delta}$ becomes $\mathbf{0}$. But, in real life, one may not be sure on the validity of H_0, and hence the RE is most likely to be unbiased.

Theorem 4.2. *The expression of bias for the PTE of $\boldsymbol{\mu}_1$ is given by*

$$B_3(\hat{\boldsymbol{\mu}}_1^{pt}; \boldsymbol{\mu}_1) = E(\hat{\boldsymbol{\mu}}_1^{pt} - \boldsymbol{\mu}_1) = M\boldsymbol{\delta}G^{(2)}_{p+2,m}(\ell_\alpha; \Delta^*) \tag{4.3}$$

where

$$G^{(2)}_{p+2,m}(\ell_\alpha;\Delta^*)=\sum_{r=0}^{\infty}\frac{\Gamma\left(\frac{p+2+m}{2}+r\right)}{\Gamma\left(\frac{p+2}{2}+r\right)\Gamma\left(\frac{m}{2}\right)}\xi_r(h_2)\xi_r(\nu)\xi_r(\nu,\Delta^*) \tag{4.4}$$

in which $\ell_\alpha=\dfrac{p}{p+2}F_{p,m}(\alpha)$ *with* $F_{p,m}(\alpha)$ *being such that*

$$P\{F_{p,m}>F_{p,m}(\alpha)\}=\alpha;$$

$$\xi_r(h_2)=I_{h_2}\left(\frac{m}{2};\frac{p+2+2r}{2}\right),\ \textit{incomplete beta function with } h_2=\frac{1}{1+\ell_\alpha};$$

$$\xi_r(\nu)=\frac{\Gamma\left(\frac{\nu}{2}+r\right)}{r!\Gamma\left(\frac{\nu}{2}\right)};\quad \textit{and}$$

$$\xi_r(\nu,\Delta^*)=\frac{[\Delta^*/(\nu-2)]^r}{[1+\Delta^*/(\nu-2)]^{\nu/2+r}} \tag{4.5}$$

with $\Delta^*=\dfrac{\nu-2}{\nu}\Delta$ *in which* $\Delta=n_1M\,\boldsymbol{\delta}'\boldsymbol{\delta}$ *and* $\boldsymbol{\delta}=\boldsymbol{\mu}_2-\boldsymbol{\mu}_1$.

Proof. From the definition of $\hat{\boldsymbol{\mu}}_1^{pt}$, we have

$$\hat{\boldsymbol{\mu}}_1^{pt}-\boldsymbol{\mu}_1=M(\tilde{\boldsymbol{\mu}}_2-\tilde{\boldsymbol{\mu}}_1)I(T^2\le T^2_\alpha)+(\tilde{\boldsymbol{\mu}}_1-\boldsymbol{\mu}_1).$$

Thus

$$\begin{aligned}B_3(\hat{\boldsymbol{\mu}}_1^{pt};\boldsymbol{\mu}_1)&=ME[(\tilde{\boldsymbol{\mu}}_2-\tilde{\boldsymbol{\mu}}_1)I(T^2\le T^2_\alpha)]\\&=ME_\tau\{E[(\tilde{\boldsymbol{\mu}}_2-\tilde{\boldsymbol{\mu}}_1)I(T^2\le T^2_\alpha)|\tau]\}.\end{aligned} \tag{4.6}$$

Now, conditional on τ, $(\tilde{\boldsymbol{\mu}}_2-\tilde{\boldsymbol{\mu}}_1)$ has a p-variate normal distribution with mean $\boldsymbol{\delta}$ and covariance $\dfrac{\tau^2}{Mn_1}I_p$. And hence

$$\boldsymbol{Y}=\frac{\sqrt{n_1M}}{\tau}\ (\tilde{\boldsymbol{\mu}}_2-\tilde{\boldsymbol{\mu}}_1) \tag{4.7}$$

follows a multivariate normal distribution with mean $\dfrac{\sqrt{n_1M}}{\tau}\boldsymbol{\delta}$ and covariance matrix I_p. Therefore, the quadratic form

$$\boldsymbol{Y}'\boldsymbol{Y}=\frac{n_1M}{\tau^2}(\tilde{\boldsymbol{\mu}}_2-\tilde{\boldsymbol{\mu}}_1)'(\tilde{\boldsymbol{\mu}}_2-\tilde{\boldsymbol{\mu}}_1) \tag{4.8}$$

has a noncentral chi-squared distribution with p degrees of freedom and noncentrality parameter $\Delta_\tau = \frac{\Delta}{\tau^2}$ where $\Delta = n_1 M \boldsymbol{\delta}'\boldsymbol{\delta}$.

Also, it can be shown that

$$T^2 = \frac{n_1 M}{S^2}(\tilde{\boldsymbol{\mu}}_2 - \tilde{\boldsymbol{\mu}}_1)'(\tilde{\boldsymbol{\mu}}_2 - \tilde{\boldsymbol{\mu}}_1) = \frac{Y'Y}{\chi^2_m} \tag{4.9}$$

where $S^2 = \sum_{i=1}^{2}\sum_{j=1}^{n_i}(\boldsymbol{X}_{ij} - \tilde{\boldsymbol{\mu}}_i)'(\boldsymbol{X}_{ij} - \tilde{\boldsymbol{\mu}}_i)$, used in (3.11). Applying the above results we get

$$E\left[(\tilde{\boldsymbol{\mu}}_2 - \tilde{\boldsymbol{\mu}}_1) I(T^2 \le T^2_\alpha)|\tau\right] = \boldsymbol{\delta} E\left[I\left(\frac{\chi^2_{p+2}(\Delta_\tau)}{\chi^2_m} \le \frac{p}{p+2}F(\alpha)_{p,m}\right)\right] = \boldsymbol{\delta} G_{p+2,m}(\ell_\alpha : \Delta_\tau) \tag{4.10}$$

where $\ell_\alpha = \frac{p}{p+2}F_{p,m}(\alpha)$ and $G_{p+2,m}(\ell_\alpha; \Delta_\tau)$ is the distribution function of a noncentral F-distribution with $p+2$ and m degrees of freedom and evaluated at ℓ_α.

The readers may check Appendix B2 of Judge and Bock (1978) for details on expectation on Borel measurable functions and noncentral distributions. Finally, we get the expression of bias of $\hat{\boldsymbol{\mu}}_1^{pt}$ by taking expectation on $G_{p+2,m}(\ell_\alpha; \Delta_\tau)$ with respect to the distribution of τ.

The bias of the PTE is $\mathbf{0}$ when $\boldsymbol{\delta} = \mathbf{0}$. That is, PTE is unbiased under H_0, but it is biased under the alternative hypothesis. The amount of bias of the PTE is less than that of the RE, since $G^{(2)}_{p+2,m}(\ell_\alpha; \Delta^*)$ is bounded by 0 and 1. Thus PTE reduces the bias of the RE (note RE is a special case of the PTE). Judging from the criterion of unbiasedness the PTE is preferable to the RE, and UE is preferable to the PTE.

5 Quadratic Risk

Let $\boldsymbol{\mu}^*$ be an estimator of $\boldsymbol{\mu}$ based on a sample of size n, and Ω be a positive definite matrix. Then the quadratic loss function of $\boldsymbol{\mu}^*$ in estimating $\boldsymbol{\mu}$ is given by

$$L(\boldsymbol{\mu}^*, \Omega) = n(\boldsymbol{\mu}^* - \boldsymbol{\mu})'\Omega^{-1}(\boldsymbol{\mu}^* - \boldsymbol{\mu}_1) \tag{5.1}$$

and the quadratic risk is defined as

$$R(\boldsymbol{\mu}^*, \Omega) = E\{L(\boldsymbol{\mu}^*, \Omega)\}. \tag{5.2}$$

The above loss function is also known as Mahalanobish distance. Following the above specification, we derive the risk of the three estimators of $\boldsymbol{\mu}_1$ defined in Section 3.

Theorem 5.1. *For the multivariate Student-t model, the quadratic risk of the UE is given by*

$$R_1(\tilde{\boldsymbol{\mu}}_1, \Sigma) = p. \tag{5.3}$$

Proof. By definition, we have

$$R_1(\tilde{\boldsymbol{\mu}}_1, \Sigma) = E_\tau\left[E\left\{\frac{n_1(\tilde{\boldsymbol{\mu}}_1 - \boldsymbol{\mu}_1)'(\tilde{\boldsymbol{\mu}}_1 - \boldsymbol{\mu}_1)}{\tau^2}\right\}|\tau\right]. \tag{5.4}$$

Since for a given τ, $(\tilde{\boldsymbol{\mu}}_1 - \boldsymbol{\mu}_1)\sqrt{n_1}/\tau$ has a standard normal distribution, the conditional expectation is the mean of the χ^2_p-variable. Hence the proof.

Clearly, the UE of $\boldsymbol{\mu}_1$ has a constant risk which equals the dimension of the population.

Theorem 5.2. *For the multivariate Student-t model, the quadratic risk of the RE is given by*

$$R_2(\hat{\boldsymbol{\mu}}_1, \Sigma) = \frac{n_1}{n_1 + n_2}p + M\Delta. \tag{5.5}$$

where $\Delta = \boldsymbol{\delta}'\boldsymbol{\delta}$.

Proof. From the definition of the quadratic risk, the risk of the RE is given by

$$R_2(\hat{\boldsymbol{\mu}}_1, \Sigma) = E_\tau \left[E \left\{ n_1 \frac{(\hat{\boldsymbol{\mu}}_1 - \boldsymbol{\mu}_1)'(\hat{\boldsymbol{\mu}}_1 - \boldsymbol{\mu}_1)}{\tau^2} \right\} | \tau \right]. \tag{5.6}$$

Now conditional on τ, $\hat{\boldsymbol{\mu}}_1$ is normally distributed with mean $\frac{n_1\boldsymbol{\mu}_1 + n_2\boldsymbol{\mu}_2}{n_1 + n_2}$ and covariance matrix $\frac{\tau^2}{n_1 + n_2} I_p$. Then $\frac{\sqrt{n_1 + n_2}}{\tau}(\hat{\boldsymbol{\mu}}_1 - \boldsymbol{\mu}_1)$ follows a normal distribution with mean $\frac{n_2}{\tau\sqrt{n_1 + n_2}}(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)$ and identity covariance matrix. Therefore, $\frac{(n_1 + n_2)}{\tau^2}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_1)'(\hat{\boldsymbol{\mu}}_1 - \boldsymbol{\mu}_1)$ has a non-central χ^2 distribution with p degrees of freedom and noncentrality parameter $\lambda = Mn_2\frac{\boldsymbol{\delta}'\boldsymbol{\delta}}{\tau^2} = \frac{n_2}{n_1}\Delta_\tau$. Hence, we get

$$E\{n_1(\hat{\boldsymbol{\mu}}_1 - \boldsymbol{\mu}_1)'(\hat{\boldsymbol{\mu}}_1 - \boldsymbol{\mu}_1)|\tau\} = \frac{n_1}{n_1 + n_2}(p + \lambda) = \frac{n_1}{n_1 + n_2}p + M\Delta_\tau \tag{5.7}$$

where $\Delta_\tau = n_1 M \frac{\boldsymbol{\delta}'\boldsymbol{\delta}}{\tau^2} = \frac{n_1 M}{\tau^2}\Delta$.

The final result is obtained by taking expectation of the right hand side of (5.7) with respect to $IG(\nu, 1)$.

The risk of the RE is smaller than that of the UE when $\Delta = 0$, that is, the null hypothesis holds good. Thus under H_0, the biased estimator, RE dominates the unbiased estimator UE. However, as Δ grows large the risk of the RE increases unboundedly and exceeds the risk of the UE for some value of Δ. In fact, $R_2(\hat{\boldsymbol{\mu}}_1; \Sigma)$ is a straight line function of Δ with slope M and intercept $\frac{n_1}{n_1 + n_2}p$. Moreover, the risk curves of UE and RE intersect at $\Delta = p$.

Theorem 5.3. *For the Student-t model, the quadratic risk of the PTE is given by*

$$R_3(\hat{\boldsymbol{\mu}}_1^{pt}; \Sigma) = p + M \left[\Delta^* \left\{ 2G^{(2)}_{p+2,m}(\ell_\alpha; \Delta^*) - G^{(4)}_{p+4,m}(\ell^*_\alpha; \Delta^*) \right\} - pG^{(2)}_{p+2,m}(\ell_\alpha; \Delta^*) \right] \tag{5.8}$$

where

$$G^{(4)}_{p+4,m}(\ell^*_m;\Delta^*) = \sum_{r=0}^{\infty}\frac{\Gamma\left(\frac{p+4+m}{2}+r\right)}{\Gamma\left(\frac{p+4}{2}+r\right)\Gamma\left(\frac{m}{2}\right)}\xi_r(h_4)\xi_r(\nu)\xi_r(\nu,\Delta^*) \tag{5.9}$$

in which $\xi_r(h_4) = I_{h_4}\left(\frac{m}{2},\frac{p+4}{2}+r\right)$, *incomplete beta function with* $h_4 = \frac{p+4}{p+4+pF_{p,m}(\alpha)}$ *and* $\ell^*_\alpha = \frac{p}{p+4}F_{p,m}(\alpha)$.

Proof. As before, first we will compute $E\left\{\frac{n_1(\hat{\mu}_1^{pt}-\mu_1)'(\hat{\mu}_1^{pt}-\mu_1)}{\tau^2}|\tau\right\}$, the conditional quadratic risk of $\hat{\mu}_1^{pt}$, for a given value of τ. From the definition of the PTE, we can write

$$(\hat{\mu}^{pt}-\mu_1) = (\tilde{\mu}_1-\mu_1) + M(\tilde{\mu}_2-\tilde{\mu}_1)I(T^2\le T^2_\alpha), \tag{5.10}$$

and hence

$$\begin{aligned} E\left\{\frac{n_1(\hat{\mu}_1^{pt}-\mu_1)'(\hat{\mu}_1^{pt}-\mu_1)}{\tau^2}|\tau\right\} &= E\left\{\frac{n_1(\tilde{\mu}_1-\mu_1)'(\tilde{\mu}_1-\mu_1)}{\tau^2}|\tau\right\} \\ &+M^2E\left\{n_1\frac{(\tilde{\mu}_2-\tilde{\mu}_1)'(\tilde{\mu}_2-\tilde{\mu}_1)}{\tau^2}I(T^2\le T^2_\alpha)|\tau\right\} \\ &+2ME\left\{n_1\frac{(\tilde{\mu}_1-\mu_1)'(\tilde{\mu}_2-\tilde{\mu}_1)}{\tau^2}I(T^2\le T^2_\alpha)|\tau\right\}. \end{aligned} \tag{5.11}$$

First note that the conditional expectation of $(\tilde{\mu}_1-\mu_1)$, given $(\tilde{\mu}_2-\tilde{\mu}_1)$, is $-M\{(\tilde{\mu}_2-\tilde{\mu}_1)-(\mu_2-\mu_1)\}$. Hence the third term on the right hand side of (5.11) becomes

$$\begin{aligned} &-2M^2E\left\{\frac{n_1(\tilde{\mu}_2-\tilde{\mu}_1)'(\tilde{\mu}_2-\tilde{\mu}_1)}{\tau^2}I(T^2\le T^2_\alpha)|\tau\right\} \\ &+2M^2(\mu_2-\mu_1)'E\left\{\frac{n_1(\tilde{\mu}_2-\tilde{\mu}_1)}{\tau^2}I(T^2\le T^2_\alpha)|\tau\right\}. \end{aligned} \tag{5.12}$$

Now the first term on the right hand side of (5.11) is p, the risk of the UE. For the evaluation of the remaining terms, recall the transformation $\boldsymbol{Y} = \frac{\sqrt{n_1 M}}{\tau}(\tilde{\mu}_2-\tilde{\mu}_1)$, as given in (4.7) and the results thereafter. Using those results, and applying theorems from Appendix B2 of Judge and Bock (1978), we get

$$E\left\{\frac{n_1(\tilde{\boldsymbol{\mu}}_2-\tilde{\boldsymbol{\mu}}_1)'(\tilde{\boldsymbol{\mu}}_2-\tilde{\boldsymbol{\mu}}_1)}{\tau^2}I(T^2\leq T_\alpha^2)|\tau\right\}$$

$$=\frac{p}{M}G_{p+2,m}(\ell_\alpha;\Delta_\tau)+\frac{\Delta_\tau}{M}G_{p+4,m}(\ell_\alpha^*;\Delta_\tau). \tag{5.13}$$

Similarly,

$$(\boldsymbol{\mu}_2-\boldsymbol{\mu}_1)'E\left\{\frac{n(\tilde{\boldsymbol{\mu}}_2-\tilde{\boldsymbol{\mu}}_1)}{\tau^2}I(T^2\leq T_\alpha^2)|\tau\right\}=\frac{\Delta_\tau}{M}G_{p+2,m}(\ell_\alpha;\Delta_\tau). \tag{5.14}$$

Using (5.12 – 5.14) in (5.11) and simplifying, conditional on τ, the risk of $\hat{\boldsymbol{\mu}}_1^{pt}$ becomes

$$R_3(\hat{\boldsymbol{\mu}}_1^{pt},\Sigma|\tau)=p+M\left[\Delta_\tau\left\{2G_{p+2,m}(\ell_\alpha^*;\Delta_\tau)-G_{p+4,m}(\ell_\alpha;\Delta_\tau)\right\}-pG_{p+2,m}(\ell_\alpha;\Delta_\tau)\right]. \tag{5.15}$$

Then taking expectation on (5.15) with respect to $IG(\nu,1)$ the final expression in (5.8) is obtained.

6 Analysis of Risks

In the previous section, the risks of the three estimators have been obtained. The risks are used as a criterion to compare the relative performances of the estimators. We provide the risk analysis both under the null and the alternative hypotheses. Based on the principle of dominance, we have the following theorem.

Theorem 6.1. *For the multivariate Student-t model, under H_0, the dominance of the three estimators is given by*

$$R_1(\tilde{\boldsymbol{\mu}}_1;\Sigma)\geq R_3(\hat{\boldsymbol{\mu}}_1^{pt};\Sigma)\geq R_2(\hat{\boldsymbol{\mu}}_1;\Sigma). \tag{6.1}$$

Proof. The risk of the RE under H_0 reduces to $\frac{n_1}{n_1+n_2}p$, which is less than p, the risk of UE. Thus

$$R_1(\tilde{\boldsymbol{\mu}}_1;\Sigma) > R_2(\hat{\boldsymbol{\mu}}_1;\Sigma). \tag{6.2}$$

Then the risk of the PTE under H_0 becomes $p - MpG_{p+2,m}(\ell_\alpha;0)$. Since $0 < M < 1$ and $0 \leq G_{p+2,m}(\ell_\alpha;0) \leq 1$, the risk of the PTE is less than UE. Hence

$$R_1(\tilde{\boldsymbol{\mu}}_1;\Sigma) > R_3(\hat{\boldsymbol{\mu}}_1^{pt};\Sigma). \tag{6.3}$$

Finally, the risk of the PTE can be expressed as

$$\frac{n_1}{n_1+n_2}p + Mp[1 - G_{p+2,m}(\ell_\alpha;0)]. \tag{6.4}$$

Noting that the term in the square bracket is positive but less than 1, the second term in (6.4) is positive. Therefore the expression in (6.4) is greater than the risk of the RE under H_0, that is,

$$R_3(\hat{\boldsymbol{\mu}}_1^{pt};\Sigma) > R_2(\hat{\boldsymbol{\mu}}_1;\Sigma). \tag{6.5}$$

Combining (6.2) – (6.5) the proof is completed. The above theorem holds good for all choices of α, the level of significance for the PTE and any choice of ν.

Under the alternative hypothesis, the following theorem provides the comparison of risks of the three estimators.

Theorem 6.2. *For the multivariate Student-t model, under the alternative hypothesis,*

(a) *the RE is better than the UE when* $\Delta < p$;

(b) *the PTE is better than the UE when*

$$\Delta^* < \frac{pG^{(2)}_{p+2,m}(\ell_\alpha; \Delta^*)}{G^{(4)}_{p+4,m}(\ell^*_\alpha; \Delta^*) - 2G^{(2)}_{p+2,m}(\ell_\alpha; \Delta^*)} \quad \textit{and} \tag{6.6}$$

(c) *the PTE is better than the RE whenever*

$$\Delta^* > \frac{p[1 - G^{(2)}_{p+2,m}(\ell_\alpha; \Delta^*)]}{2G^{(2)}_{p+2,m}(\ell_\alpha; \Delta^*) - G^{(4)}_{p+4,m}(\ell^*_\alpha; \Delta^*) - \frac{\nu-2}{\nu}}. \tag{6.7}$$

In each of the above cases, the opposite conclusion is true if the inequality is reversed.

Proof. The difference between the risks of the UE and RE is given by

$$D_{12} = p - \frac{n_1}{n_1 + n_2} p - M\Delta = Mp - M\Delta. \tag{6.8}$$

The RE has a smaller risk than the UE iff $D_{12} > 0$. That happens whenever $p > \Delta$. Therefore, the RE performs better than the UE if $0 < \Delta < p$.

The risk difference of the UE and PTE can be written as

$$D_{13} = pMG^{(2)}_{p+2,m}(\ell_\alpha; \Delta^*) + \Delta^* M \left\{2G^{(2)}_{p+2,m}(\ell_\alpha; \Delta^*) - G^{(4)}_{p+4,m}(\ell^*_\alpha; \Delta^*)\right\}. \tag{6.9}$$

So, the PTE has a smaller risk iff $D_{12} > 0$, that is, if

$$\Delta^* \left\{2G^{(2)}_{p+2,m}(\ell_\alpha; \Delta^*) - G^{(4)}_{p+4,m}(\ell^*_\alpha; \Delta^*)\right\} > -pG^{(2)}_{p+2,m}(\ell_\alpha; \Delta^*), \tag{6.10}$$

or equivalently, whenever

$$\Delta^* < \frac{pG^{(2)}_{p+2,m}(\ell_\alpha; \Delta^*)}{G^{(4)}_{p+4,m}(\ell^*_\alpha; \Delta^*) - 2G^{(2)}_{p+2,m}(\ell_\alpha; \Delta^*)}. \tag{6.11}$$

Now the risk difference of the RE and PTE becomes

$$D_{23} = \frac{n_1}{n_1+n_2}p + M\Delta - p + pMG^{(2)}_{p+2,m}(\ell_\alpha;\Delta^*) - \Delta^* M\left\{2G^{(2)}_{p+2,m}(\ell_\alpha;\Delta^*) - G^{(4)}_{p+4,m}(\ell^*_\alpha;\Delta^*)\right\} \quad (6.12)$$

which can be expressed as

$$D_{23} = -Mp\left[1 - G^{(2)}_{p+2,m}(\ell_\alpha;\Delta^*)\right] + M\Delta^*\left[\frac{\nu-2}{\nu} - \left\{2G^{(2)}_{p+2,m}(\ell_\alpha;\Delta^*) - G^{(4)}_{p+4}(\ell^*_\alpha\Delta^*)\right\}\right]. \quad (6.13)$$

Therefore, the PTE performs better than the RE iff $D_{23} > 0$, that is, whenever,

$$M\Delta^*\left[\frac{\nu-2}{\nu} - 2G^{(2)}_{p+2,m}(\ell_\alpha;\Delta^*) + G^{(4)}_{p+4,m}(\ell^*_\alpha;\Delta^*)\right] > Mp\left[1 - G^{(2)}_{p+2,m}(\ell_\alpha;\Delta^*)\right]$$

$$\Delta^* > \frac{p[1 - G^{(2)}_{p+2,m}(\ell_\alpha;\Delta^*)]}{\left[\frac{\nu-2}{\nu} - 2G^{(2)}_{p+2,m}(\ell_\alpha;\Delta^*) + G^{(4)}_{p+4,m}(\ell^*_\alpha;\Delta^*)\right]}. \quad (6.14)$$

Hence the proof.

From part (a) of the theorem, the RE overperforms the UE when $0 < \Delta < p$. Outside this interval, that is, whenever $\Delta > p$, the UE is better than the RE. At the point, $\Delta = p$, the risk curves of the two estimators intersect. The curve of the risk function of the RE starts under that of the UE having the lowest point at $\Delta = 0$, and passes through the risk curve of the UE at a point $\Delta = p$ and then grows upward unboundedly as Δ goes further away from 0. Therefore, under the alternative hypothesis neither RE nor UE uniformly dominates one another for all Δ.

Based on part (b) of the theorem, the PTE dominates the UE whenever

$$0 < \Delta < \frac{\nu-2}{\nu}pG^{(2)}_{p+2,m}(\ell_\alpha;\Delta^*)\left[G^{(4)}_{p+4,m}(\ell^*\alpha;\Delta^*) - 2G^{(2)}_{p+2,m}(\ell_\alpha;\Delta^*)\right]^{-1}; \quad (6.15)$$

otherwise the opposite is true. Thus there is no uniform domination of one over the other for all values of Δ. Similar to the risk of the RE, the risk curve of the PTE starts from below the risk line of the UE. But then it goes higher and higher after crossing the risk line of the UE from below. However, the risk curve of the PTE gradually approaches to that of the UE as Δ grows larger and larger. This phenomenon also depends on the choice of α. For a larger value of α the risk of the PTE approaches that of the UE more quickly than for a smaller α. The same comment is valid for the shape parameter, that is, for a larger number of degrees of freedom, $R_3(\hat{\boldsymbol{\mu}}_1^{pt};\Sigma)$ approaches to $R_1(\tilde{\boldsymbol{\mu}}_1;\Sigma)$ faster than for a smaller number of degrees of freedom.

To compare the risks of the RE and PTE, we note from part (c) of the theorem that the RE overperforms the PTE when

$$0<\Delta<\frac{\nu-2}{\nu}p\left[1-G^{(2)}_{p+2,m}(\ell_\alpha;\Delta;^*)\right]\left[2G^{(2)}_{p+2,m}(\ell_\alpha;\Delta^*)-G^{(4)}_{p+4,m}(\ell^*\alpha;\Delta^*)-\frac{\nu-2}{\nu}\right]^{-1}. \tag{6.16}$$

Thus for the value of Δ outside the limits in (6.16), the PTE dominates the RE. When the value of Δ is close to 0, the UE has a smaller risk than the PTE. Whereas the risk of the RE sharply increases as Δ becomes large and grows unboundedly after crossing the risk curve of the UE from below. On the other hand, although the risk of the PTE also crosses the risk curve of the UE from below it starts declining after reaching a maximum for some value of Δ. Thus, under the alternative the PTE provides a compromise between the UE and RE. The compromise is better in the sense of having smaller risk when the value of α is large.

7 Concluding Remarks

From the foregoing discussions it is evident that the PTE is a function of the choice of the level of significance, α for the F-test. Therefore, the question arises as to which value of α should be selected and why? The issue has been discussed by several authors. Akaike (1973) proposed an information criterion based method for the selection of an optimal value of α. Hirano (1977) illustrated the method for some preliminary test estimators. Khan and Saleh (1997) applied this method to determine optimal values of α for the shrinkage preliminary test estimator with varying sample sizes.

In the real life problems, the value of Δ is not known. Thus the choice of the best estimator may not be straightforward. The actual preference may depend on the knowledge or perception of the value of Δ. If the experimenter is confident that the departure of the value of $\boldsymbol{\mu}_1$ is not too far from its value under the null hypothesis, the RE may be preferred. Whereas, the PTE should be a better choice if such departure is likely to be very large. In the case when the departure is moderate one may be better off with the UE. However, in the face of uncertainty on the size of the departure, the PTE provides a better compromise for extreme values of the departure, leaving UE as a choice for moderate size of departure.

ACKNOWLEDGEMENT

The author acknowledges the excellent research facility provided by the King Fahd University of Petroleum and Minerals during the preparation and finalization of this paper.

REFERENCES

1. Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B.N. Petrov an F. Csaki (eds). Second International Symposium on Information Theory. Akademiai Budapest, 267–281.

2. Anderson, T.W. (1993). Nonnormal multivariate distributions: Inference based on elliptically contoured distributions. In Multivariate Analysis: Future Directions. ed by C.R. Rao. North-Holland, Amsterdam, 1–24.

3. Bancroft, T.A. (1944). On biases in estimation due to the use of the preliminary tests of significance. Ann. Math. Statist., Vol. 15, 190–204.

4. Box, G.E.P. and Tiao, G.C. (199). Bayesian Inference in Statistical Analysis, Wiley, New York.

5. Chmielewski, M.A. (1981). Elliptically symmetric distributions. International Statistical Review, Vol. 49, 67–74.

6. Fang, K.T. and Anderson, T.W. (1990). Statistical Inference in Elliptically Contoured and Related Distributions. Allerton Press Inc., New York.

7. Fisher, R.A. (1956). Statistical Methods in Scientific Inference. Oliver and Boyd, London.

8. Fraser, D.A.S. (1979). Inference and Linear Models, McGraw–Hill, New York.

9. Gupta, A.K. and Varga, T. (1993). Elliptically Contoured Models in Statistics. Kluwer Academic Publishers, London.

10. Han, C.P. an Bancroft, T.A. (1968). On pooling means when variance is unknown. Jou. Amer. Statist. Assoc., Vol. 63, 1333–1342.

11. Hirano, K. (1977). Estimation procedure based on preliminary test shrinkage technique and information criteria. Ann. Inst. Statist. Math. **29**, Part A, 21–34.

12. Judge, G.G. and Boek, M.E. (1978). The Statistical Implications of Pre-test and Stein-rule Estimators in Econometrics. North-Holland, New York.

13. Khan, S. and Haq, M.S. (1990). Prediction distribution for the linear model with multivariate Student-t error, Communications in Statistics: Theory & Methods. Vol. 19, 4705–4712.

14. Khan, S. and Saleh, A.K.Md.E. (1997). Shrinkage pre-test estimator of the intercept parameter for a regression model with multivariate Student-t errors. Biometrical Journal, **39**, 131–147.

15. Prucha I.R. and Kelegian, H.H. (1984). The structure of simultaneous equation estimators: A generalization towards non-normal disturbances, Econometrica, Vol. 52, 721–736.

16. Singh, R.S. (1988). Estimation of error variance in linear regression models with errors having multivariate Student-t distribution with unknown degrees of freedom. Economics Letters, **27**, 47–53.

17. Tabatabaey, S.M.M. (1995). Preliminary Test Approach Estimation: Regression Model With Spherically Symmetric Errors. Unpublished Ph.D. Thesis, Carleton University, Canada.