

## **Stein-type Estimators for Mean Vector in Two-Sample Problem of Multivariate Student-t Populations with Common Covariance Matrix**

**Shahjahan Khan**

*Department of Mathematics & Computing, University of Southern Queensland,  
Toowoomba, Qld. 4530, Australia, e-mail: khans@usq.edu.au*

**A.K. Md. E. Saleh**

*School of Mathematics & Statistics, Carleton University, Ottawa, Canada*

[Received March 19, 2002; Revised June 30, 2002; Accepted July 15, 2002]

### **Abstract**

The paper obtains two Stein-type estimators of the mean vector of a multivariate Student-t population with unknown shape parameter based on two random samples from populations having common but unknown covariance matrix by using the preliminary test approach to shrinkage estimation. The properties of the shrinkage and the positive rule shrinkage estimators are investigated in terms of bias and the quadratic risk criteria. The relative performances of the estimators, and dominance over the usual maximum likelihood estimator are discussed. Comparisons of the estimators for the multivariate normal model are also included under different conditions. The uniform dominance of the positive rule shrinkage estimator over the shrinkage and maximum likelihood estimator is established.

**Keywords and Phrases:** Multivariate normal, Student-t, inverted gamma, noncentral chi-squared and  $F$ -distributions; maximum likelihood, preliminary test, shrinkage and positive rule-shrinkage estimators; bias, quadratic risk and relative efficiency.

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

## **1 Introduction**

Conventional estimators, like the maximum likelihood and least squares estimators, are based on the sample responses alone. Such estimators disregard any non-sample



information, either in the form of a prior distribution of the parameter(s) or as a suspected value of it. Improved estimator, on the other hand, incorporates both the sample and non-sample information in the definition of the estimators. Among the popular improved estimators, the well known preliminary test estimator (Bancroft, 1944) and James-Stein type shrinkage estimator (Stein 1956, and James and Stein 1961) perform better than the conventional estimators under certain conditions. There has been many studies on the topic, mainly for the linear models and one-sample problems. Khan and Hoque (2002) considered the two-sample problem for the multivariate normal populations. This paper makes a wider assumption by considering the Student-t population and deals with the estimation of the location vector using both the sample and non-sample prior information regarding the value of the mean vectors. Sampling from Student-t population is not straight forward as in the case of the normal population. Khan (1997) adopted the mixture distribution of the normal and the inverted gamma distributions to define samples from Student-t population. In this paper, we pursue the same approach to obtain Student-t samples and define unrestricted estimator of the location vector.

Consider an independent random sample,  $\mathbf{X}_{11}, \mathbf{X}_{12}, \dots, \mathbf{X}_{1n_1}$  from a  $p$ -variate normal population with unknown mean vector  $\boldsymbol{\mu}_1$ . Then consider another independent random sample  $\mathbf{X}_{21}, \mathbf{X}_{22}, \dots, \mathbf{X}_{2n_2}$  from a second  $p$ -variate normal population with unknown mean vector  $\boldsymbol{\mu}_2$ . Let  $\tau^2 \Sigma$  be the unknown but common covariance matrix of the two populations. Then the joint density of the two independent samples can be given by

$$f(\mathbf{X}_1, \mathbf{X}_2; \boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \tau^2 \Sigma) = (2\pi)^{-\frac{p(n_1+n_2)}{2}} (\tau^2)^{-\frac{p(n_1+n_2)}{2}} |\Sigma|^{-\frac{n_1+n_2}{2}} e^{-\frac{1}{2\tau^2}(\mathbf{Q}_1 + \mathbf{Q}_2)} \quad (1)$$

where  $\mathbf{Q}_i = \sum_{j=1}^{n_i} (\mathbf{X}_{ij} - \boldsymbol{\mu}_i)' \Sigma^{-1} (\mathbf{X}_{ij} - \boldsymbol{\mu}_i)$  and  $\mathbf{X}_i$  is a  $p \times n_i$ -dimensional matrix of the  $i$ -th sample for  $i = 1, 2$ .

Now assume that  $\tau$  follows an inverted gamma distribution with the density function

$$f(\tau) = \frac{2}{\Gamma(\frac{\nu}{2})} \left(\frac{\nu}{2}\right)^{\nu/2} (\tau)^{-(\nu+1)} e^{-\frac{\nu}{2\tau^2}}, \quad \tau > 0 \quad (2)$$

where  $\nu$  is the shape parameter. It is well known that the mixture distribution of the two samples is  $p(n_1 + n_2)$ -dimensional Student-t distribution with the joint density function

$$f(\mathbf{X}_1, \mathbf{X}_2, \boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \Sigma, \nu) = K(\nu, n_1, n_2, p) |\Sigma|^{-\frac{n_1+n_2}{2}} \times \left[ 1 + \frac{1}{\nu} \sum_{i=1}^2 \sum_{j=1}^{n_i} (\mathbf{X}_{ij} - \boldsymbol{\mu}_i)' \Sigma^{-1} (\mathbf{X}_{ij} - \boldsymbol{\mu}_i) \right]^{-\frac{\nu+p(n_1+n_2)}{2}} \quad (3)$$



where

$$K(\nu, n_1, n_2, p) = \frac{\Gamma\left(\frac{\nu+p(n_1+n_2)}{2}\right)}{(\pi\nu)^{\frac{p(n_1+n_2)}{2}} \Gamma\left(\frac{\nu}{2}\right)}$$

is the normalizing constant. Note that the dispersion matrix of  $\mathbf{X}_{ij}$  is  $\frac{\nu}{\nu-2}\Sigma$  for  $j = 1, 2, \dots, n_i$  and  $i = 1, 2$ , and

$$\text{Cov}(\mathbf{X}_1, \mathbf{X}_2) = \frac{\nu}{\nu-2}\Sigma \otimes \begin{bmatrix} I_{n_1} & O_{n_2} \\ O_{n_1} & I_{n_2} \end{bmatrix} \quad (4)$$

where  $I_{n_i}$  is the identity matrix of order  $n_i$  and  $O_i$  is a  $n_i \times n_i$  order matrix of zeros for  $i = 1, 2$ , and  $\otimes$  is the Kronecker product between matrices. Thus the elements in each column of  $\mathbf{X}_1$  and  $\mathbf{X}_2$  are correlated, but the columns of  $\mathbf{X}_1$  and  $\mathbf{X}_2$  are dependent but uncorrelated. This is a special property of the multivariate Student-t distribution (cf Khan, 1992 and Anderson, 1993). Khan (1998) obtained the preliminary test estimator for the above two-sample problem in the presence of uncertain prior information and with diagonal covariance matrix, while Khan (1997) proposed the likelihood based inference for the mean vectors of a similar model. Most of the textbooks on multivariate analysis cover the multivariate Student-t distribution but the original work on this distribution is due to Cornish (1954). For convenience, we assume that  $\nu > 1$ , and hence the study of the Cauchy and sub-Cauchy distributions are beyond the scope of this paper.

There has been an increasing trend in the use of the Student-t model in recent years. Fisher (1956, p.133) warned against the consequences of inappropriate use of the traditional normal model. Fisher (1960, p.46) analyzed Darwin's data (cf. Box and Taio, 1992, p. 133) by using a non-normal model. Fraser and Fick (1975) analyzed the same data by the Student-t model. Zellner (1976) provided both Bayesian and frequentist analyses of the multiple regression model with Student-t errors. Fraser (1979) illustrated the robustness of the Student-t model. Prucha and Kelegian (1984) proposed an estimating equation for the simultaneous equation model with the Student-t errors. Ullah and Walsh (1984) investigated the optimality of different types of tests used in econometric studies for the multivariate Student-t model. The interested readers may refer to the more recent work of Singh (1988), Lange et al. (1989), Giles (1991), Khan (1992), Anderson (1993), Spanos (1994), Lucus (1997) and Khan (1998) for different applications of the Student-t models.

We wish to define and study Stein-type shrinkage estimators for the mean vector of the first population,  $\boldsymbol{\mu}_1$  by applying the preliminary test approach. Assume that uncertain prior information regarding the equality of the mean vectors,  $\boldsymbol{\mu}_1$  and  $\boldsymbol{\mu}_2$ , is available, and that can be expressed by the null hypothesis,  $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ , but we do not have enough evidence in support of  $H_0$ . Following Bancroft (1944) the preliminary test estimator of  $\boldsymbol{\mu}_1$  can be obtained that removes the uncertainty in the above null hypothesis. Such an estimator is an extreme choice between the usual maximum likelihood estimator of  $\boldsymbol{\mu}_1$  and the overall sample mean  $\bar{\mathbf{X}}$  (say), depending on the



rejection or acceptance of the  $H_0$  at a given level of significance of the test. Moreover, it depends on the choice of the level of significance. Khan (1998) investigated the properties of the preliminary test estimator for the two-sample problem with unknown but common diagonal covariance matrix. In this paper, we define the Stein-type estimators of  $\boldsymbol{\mu}_1$  by applying the same approach. The properties of the estimators are also investigated based on the unbiasedness, and risk under quadratic loss function criteria. The relative performance of the estimators under different conditions are also discussed.

## 2 The Stein-Type Estimators

It is well known that the maximum likelihood estimator (mle) of  $\boldsymbol{\mu}_i$  is  $\tilde{\boldsymbol{\mu}}_i = \bar{\mathbf{X}}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \mathbf{X}_{ij}$  for  $i = 1, 2$ , and that the restricted mle (restricted by  $H_0$ ) of  $\boldsymbol{\mu}_1$  is  $\hat{\boldsymbol{\mu}}_1 = \bar{\mathbf{X}} = \frac{1}{n_1 + n_2} \sum_{i=1}^2 \sum_{j=1}^{n_i} \mathbf{X}_{ij}$ , the overall sample mean. Interested readers may see Zellner (1976) and Khan (1998), for instance, for the detailed derivation of such results. Similarly, it can be easily seen that the unrestricted and restricted mle's of  $\Sigma$  become

$$\tilde{\Sigma} = \frac{1}{n_1 + n_2} \mathbf{S} \quad \text{and} \quad \hat{\Sigma} = \frac{1}{n_1 + n_2} \mathbf{R} \quad (5)$$

where

$$\mathbf{S} = \sum_{i=1}^2 \sum_{j=1}^{n_i} (\mathbf{X}_{ij} - \bar{\mathbf{X}}_i)(\mathbf{X}_{ij} - \bar{\mathbf{X}}_i)'$$

and

$$\mathbf{R} = \sum_{i=1}^2 \sum_{j=1}^{n_i} (\mathbf{X}_{ij} - \bar{\mathbf{X}})(\mathbf{X}_{ij} - \bar{\mathbf{X}})'$$

For details on the derivation of multivariate Student-t model parameters see Khan (1997).

Based on the above estimators, a test statistic for testing the null hypothesis,  $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$  is found to be

$$T^2 = \frac{n_1 n_2}{n_1 + n_2} (\bar{\mathbf{X}}_2 - \bar{\mathbf{X}}_1)' \mathbf{S}^{-1} (\bar{\mathbf{X}}_2 - \bar{\mathbf{X}}_1) \quad (6)$$

which follows a modified Hotelling's  $T^2$  distribution (cf. Anderson 1985, p. 109). Finally, note that conditional on a given value of  $\tau$ ,

$$T^2 \sim \frac{p}{m} F_{p,m}(\Delta_\tau) \quad (7)$$



where  $F_{p,m}(\Delta_\tau)$  follows a non-central  $F$ -distribution with  $p$  and  $m = n_1 + n_2 - p - 1$  degrees of freedom (d.f.), and non-centrality parameter

$$\Delta_\tau = (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)' \frac{\boldsymbol{\Sigma}^{-1}}{\tau^2} (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1) = \frac{\boldsymbol{\delta}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\delta}}{\tau^2}. \quad (8)$$

The power function and other properties of the test can be found in Khan (1997). In this paper, we use the above  $T^2$ -statistic for the definition of the Stein-type shrinkage estimators of  $\boldsymbol{\mu}_1$ .

The *shrinkage estimator* (SE) of  $\boldsymbol{\mu}_1$  can be defined as

$$\hat{\boldsymbol{\mu}}_1^S = \tilde{\boldsymbol{\mu}}_1 + MCT^{-2}(\tilde{\boldsymbol{\mu}}_2 - \tilde{\boldsymbol{\mu}}_1) \quad (9)$$

where  $M = \frac{n_2}{n_1 + n_2}$  and  $0 < C < \frac{2(p-2)}{(N-p+3)}$  is the shrinkage constant with  $N = n_1 + n_2 - 2$ . The above estimator is a Stein-type estimator (cf. Stein 1956, and James and Stein 1961) and it dominates the usual maximum likelihood estimator of  $\boldsymbol{\mu}_1$  when  $p \geq 3$ . However, as  $T^2 \rightarrow 0$ , the above shrinkage estimator becomes unstable. To avoid this difficulty we define the *positive rule shrinkage estimator* (PRSE) of  $\boldsymbol{\mu}_1$  for the Student-t model as follows:

$$\hat{\boldsymbol{\mu}}_1^{S+} = \hat{\boldsymbol{\mu}}_1 + (1 - CT^{-2})I(T^2 > C)(\tilde{\boldsymbol{\mu}}_1 - \hat{\boldsymbol{\mu}}_1) \quad (10)$$

where  $I(T^2 > C)$  is an indicator function with only two possible values, 0 or 1. The above PRSE can be expressed as

$$\hat{\boldsymbol{\mu}}_1^{S+} = \hat{\boldsymbol{\mu}}_1^S + M(\tilde{\boldsymbol{\mu}}_2 - \tilde{\boldsymbol{\mu}}_1)I(T^2 > C) - CMT^{-2}(\tilde{\boldsymbol{\mu}}_2 - \tilde{\boldsymbol{\mu}}_1)I(T^2 \leq C). \quad (11)$$

Often such a representation makes the derivation of bias and risk functions easier.

### 3 The Bias of the Estimators

In this section we compute the expressions for the bias of the estimators. The following representations of the SE and PRSE are useful for the evaluation of the terms under the expectation operator in the definition of bias:

$$\begin{aligned} (\hat{\boldsymbol{\mu}}_1^S - \boldsymbol{\mu}_1) &= (\tilde{\boldsymbol{\mu}}_1 - \boldsymbol{\mu}_1) + CMT^{-2}(\tilde{\boldsymbol{\mu}}_2 - \tilde{\boldsymbol{\mu}}_1) \\ (\hat{\boldsymbol{\mu}}_1^{S+} - \boldsymbol{\mu}_1) &= (\tilde{\boldsymbol{\mu}}_1 - \boldsymbol{\mu}_1) + CMT^{-2}(\tilde{\boldsymbol{\mu}}_2 - \tilde{\boldsymbol{\mu}}_1) + M(\tilde{\boldsymbol{\mu}}_2 - \tilde{\boldsymbol{\mu}}_1)I(T^2 \leq C) \end{aligned} \quad (12)$$

$$-CMT^{-2}(\tilde{\boldsymbol{\mu}}_2 - \tilde{\boldsymbol{\mu}}_1)I(T^2 \leq C). \quad (13)$$

The following theorems provide the bias functions of the SE and the PRSE.

**Theorem 3.1.** *For the Student-t model stated in Section 1, the bias of the SE of  $\boldsymbol{\mu}_1$  is given by*

$$B(\hat{\boldsymbol{\mu}}_1^S; \boldsymbol{\mu}_1) = CmM\boldsymbol{\delta}E^{(2)}[\chi_{p+2}^{-2}(\Delta^*)] \quad (14)$$



where

$$E^{(2)}[\chi_{p+2}^{-2}(\Delta^*)] = \sum_{r=0}^{\infty} \frac{1}{(p+2r)} \frac{\Gamma(\frac{\nu}{2} + r)}{r! \Gamma(\frac{\nu}{2})} \frac{\left(\frac{\Delta^*}{\nu-2}\right)^r}{\left(1 + \frac{\Delta^*}{\nu-2}\right)^{\frac{\nu}{2}+r}} \quad (15)$$

in which  $\Delta^* = \frac{\nu-2}{\nu} \Delta$  with  $\Delta = n_1 M \delta' \Sigma^{-1} \delta$ .

**Proof.** By definition, the bias of the SE of  $\mu_1$  is

$$B(\hat{\mu}_1^S; \mu_1) = E[\hat{\mu}_1 - \mu_1] = E[(\tilde{\mu}_1 - \mu_1) + CMT^{-2}(\tilde{\mu}_2 - \tilde{\mu}_1)] \quad (16)$$

by using (12). Since  $E(\tilde{\mu}_1 - \mu_1) = \mathbf{0}$ , we get

$$B(\hat{\mu}_1; \mu_1) = CME[T^{-2}(\tilde{\mu}_2 - \tilde{\mu}_1)]. \quad (17)$$

Now, for the evaluation of the term under expectation, consider the following transformation:

$$Y = \sqrt{n_1 M} \frac{\Sigma^{-1/2}}{\tau} (\tilde{\mu}_2 - \tilde{\mu}_1). \quad (18)$$

Then, for a given value of  $\tau$ , we have

$$Y \sim N_p \left( \sqrt{n_1 M} \frac{\Sigma^{-1/2}}{\tau} \delta, I_p \right), \quad (19a)$$

$$(\tilde{\mu}_2 - \tilde{\mu}_1) = \frac{\tau \Sigma^{1/2}}{\sqrt{n_1 M}} Y \sim N_p \left( \delta, \frac{\tau^2 \Sigma}{n_1 M} \right), \quad (19b)$$

$$Y'Y = n_1 M (\tilde{\mu}_2 - \mu_1)' \frac{\Sigma^{-1}}{\tau^2} (\tilde{\mu}_2 - \tilde{\mu}_1) \sim \chi_p^2(\Delta_\tau), \text{ and} \quad (19c)$$

$$T^2 = \frac{Y'Y}{\chi_m^2} = \frac{p}{m} F_{p,m}(\Delta_\tau) \quad (19d)$$

where  $\chi_p^2(\Delta_\tau)$  is a noncentral chi-squared variable with  $p$  degrees of freedom and noncentrality parameter  $\Delta_\tau$ ;  $\chi_m^2$  is a central chi-squared variable with  $m$  degrees of freedom; and  $F_{p,m}(\Delta_\tau)$  is a noncentral  $F$ -variable with  $p$  and  $m$  degrees of freedom and noncentrality parameter  $\Delta_\tau$ . Therefore, we can write

$$\begin{aligned} E[T^{-2}(\tilde{\mu}_2 - \mu_1)|\tau] &= E \left[ \frac{\chi_m^2}{Y'Y} \frac{\tau \Sigma^{1/2}}{\sqrt{n_1 M}} Y \right] \\ &= m \delta E \chi_{p+2}^{-2}(\Delta_\tau). \end{aligned} \quad (20)$$

by applying Theorem 1 from the Appendix B.2 of Judge and Bock (1979, pp. 321–324). Finally taking expectation on the last term of the right hand side of (20) with respect to the distribution of  $\tau$ , we get

$$E^{(2)}[\chi_{p+2}^2(\Delta^*)] = E_\tau \{ E[\chi_{p+2}^{-2}(\Delta_\tau) | \tau] \}$$



$$= \sum_{r=0}^{\infty} \frac{1}{(p+2r)} \frac{\Gamma\left(\frac{\nu}{2} + r\right)}{r! \Gamma\left(\frac{\nu}{2}\right)} \frac{\left(\frac{\Delta^*}{\nu-2}\right)^r}{\left(1 + \frac{\Delta^*}{\nu-2}\right)^{\frac{\nu}{2}+r}}. \quad (21)$$

Hence the theorem.

Clearly the value of  $\Delta^*$  is zero when  $\nu = 2$ , regardless of the value of  $\Delta$ . Therefore, for the remainder of the paper we assume that  $\nu > 2$ , so that  $\Delta^* \geq 0$  iff  $\Delta \geq 0$ . Note that for a small  $\nu$  the value of  $\Delta^*$  is significantly smaller than that of  $\Delta$ .

**Theorem 3.2.** *For the multivariate Student-t model defined in Section 1, the bias of the PRSE of  $\boldsymbol{\mu}_1$  is given by*

$$\begin{aligned} B(\hat{\boldsymbol{\mu}}_1^{S+}; \boldsymbol{\mu}_1) &= B(\hat{\boldsymbol{\mu}}_1^S; \boldsymbol{\mu}_1) + M\delta G_{p+2,m}^{(2)}(q; \Delta^*) \\ &\quad - \frac{(p-2)}{m+2} m M \delta E^{(2)}[\chi_{p+2}^{-2}(\Delta^*) I(F_{p+2,m}(\Delta^*) \leq q)] \end{aligned} \quad (22)$$

where  $B(\hat{\boldsymbol{\mu}}_1^S; \boldsymbol{\mu}_1)$  is the bias of the SE,  $\hat{\boldsymbol{\mu}}_1^S$ ;

$$\begin{aligned} q &= \frac{m(p-2)}{(m+2)(p+2)}; \\ G_{p+2,m}^{(2)}(q; \Delta^*) &= \sum_{r=0}^{\infty} \frac{\Gamma\left(\frac{p+2+m}{2} + r\right)}{\Gamma\left(\frac{p+2}{2} + r\right) \Gamma\left(\frac{m}{2}\right)} \frac{B_{u_q}\left(\frac{m}{2}, \frac{p+2}{2} + r\right)}{r!} \frac{\Gamma\left(\frac{\nu}{2} + r\right)}{\Gamma\left(\frac{\nu}{2}\right)} \\ &\quad \times \frac{\left(\frac{\Delta^*}{\nu-2}\right)^r}{\left(1 + \frac{\Delta^*}{\nu-2}\right)^{\frac{\nu}{2}+r}}; \end{aligned} \quad (23)$$

and

$$\begin{aligned} E^{(2)}[\chi_{p+2}^{-2}(\Delta^*) I(F_{p+2,m}(\Delta^*) \leq q)] &= \sum_{r=0}^{\infty} \frac{\Gamma\left(\frac{p+2+m}{2}\right)}{\Gamma\left(\frac{p+2r}{2}\right) \Gamma\left(\frac{m}{2}\right)} \\ &\quad \times \frac{\Gamma\left(\frac{\nu+2r}{2}\right) B_{u_q}\left(\frac{m}{2}, \frac{p+2+2r}{2}\right)}{r! \Gamma\left(\frac{\nu}{2}\right) \Gamma\left(\frac{p+2+2r}{2}\right)} \frac{\left(\frac{\Delta^*}{\nu-2}\right)^r}{\left(1 + \frac{\Delta^*}{\nu-2}\right)^{\frac{\nu}{2}+r}}, \end{aligned} \quad (24)$$

in which  $u_q = \frac{(m+2)(p+2)}{(m+2)(p+2)+m(p-2)}$  and  $B_{u_q}(a; b)$  is an incomplete beta function with arguments 'a' and 'b'.



**Proof.** By definition, the bias of the PRSE of  $\mu_1$  is

$$\begin{aligned}
 B(\mu_1^{S+}; \mu_1) &= E(\hat{\mu}_1^{S+} - \mu_1) \\
 &= E[(\hat{\mu}_1^S - \mu_1) + M(\tilde{\mu}_2 - \tilde{\mu}_1) I(T^2 \leq C) \\
 &\quad - CMT^{-2}(\tilde{\mu}_2 - \tilde{\mu}_1) I(T^2 \leq C)] \\
 &= B(\hat{\mu}_1^S; \mu_1) + ME[(\tilde{\mu}_2 - \tilde{\mu}_1) + I(T^2 \leq C)] \\
 &\quad - CME[T^{-2}(\tilde{\mu}_2 - \tilde{\mu}_1) I(T^2 \leq C)]. \tag{25}
 \end{aligned}$$

Now applying the transformation in (18), conditional on a given value of  $\tau$ , we have

$$\begin{aligned}
 E[(\tilde{\mu}_2 - \tilde{\mu}_1) I(T^2 \leq C) | \tau] &= E \left[ \frac{\tau \Sigma^{1/2}}{\sqrt{n_1 M}} Y I \left( \frac{Y'Y}{\chi_m^2} \leq C \right) \right] \\
 &= \delta E \left[ I \left( \frac{\chi_{p+2}^2(\Delta_\tau)}{\chi_m^2} \leq C \right) \right] \\
 &= \delta E \left[ I \left( F_{p+2, m}(\Delta_\tau) \leq \frac{m}{p+2} C \right) \right] \\
 &= \delta G_{p+2, m}(q; \Delta_\tau) \tag{26}
 \end{aligned}$$

where  $q = \frac{m}{p+2} \times \frac{p-2}{m+2}$  in which the optimal value of the shrinkage constant  $C$ ,  $C_{opt} = \frac{p-2}{m+2}$  has been used, and the result from the Appendix B.2 of Judge and Bock (1978) has been applied. Note that here  $C$  is optimal in the sense of minimizing the quadratic risk (cf. Ahmed and Saleh, 1989).

Similarly, conditional on a given value of  $\tau$ , we get

$$\begin{aligned}
 E[T^{-2}(\tilde{\mu}_2 - \tilde{\mu}_1) I(T^2 \leq C) | \tau] &= E \left[ \frac{\chi_m^2}{Y'Y} \frac{\tau \Sigma^{1/2}}{\sqrt{n_1 M}} Y I \left( \frac{Y'Y}{\chi_m^2} \leq C \right) \right] \\
 &= m \delta E[\chi_{p+2}^{-2}(\Delta_\tau) I(F_{p+2, m}(\Delta_\tau) \leq q)]. \tag{27}
 \end{aligned}$$

Computing the expectation on (26) with respect to the inverted gamma distribution we obtain

$$E_\tau \{ E[(\tilde{\mu}_2 - \tilde{\mu}_1) I(T^2 \leq C) | \tau] \} = \delta G_{p+2, m}^{(2)}(q; \Delta^*) \tag{28}$$

and that on (27) yields

$$E_\tau \{ T^{-2}(\tilde{\mu}_2 - \tilde{\mu}_1) I(T^2 \leq C) | \tau \} = m \delta E^{(2)} \left[ \chi_{p+2}^{-2}(\Delta^*) I(F_{p+2, m}(\Delta^*) \leq q) \right]. \tag{29}$$

Finally combining the results of (28) and (29), and substituting in (25) we get the expression in (22). Hence the theorem.



Under  $H_0$ , both SE and PRSE of  $\boldsymbol{\mu}_1$  are unbiased estimators. However, the bias of the estimators depends on the value of  $\boldsymbol{\delta}$ , the departure of  $\boldsymbol{\mu}_2$  from  $\boldsymbol{\mu}_1$ , under the alternative hypothesis. Since the difference of the bias of the PRSE and the SE is given by

$$M\boldsymbol{\delta}G_{p+2,m}^{(2)}(q; \Delta^*) - \frac{(p-2)}{(m+2)}mM\boldsymbol{\delta}E^{(2)}[\chi_{p+2}^{-2}(\Delta^*)I(F_{p+2,m}(\Delta^*) \leq q)], \quad (30)$$

the PRSE has a smaller bias than the SE if

$$G_{p+2,m}^{(2)}(q; \Delta^*) < \frac{(p-2)}{(m+2)}mE^{(2)}[\chi_{p+2}^{-2}(\Delta^*)I(F_{p+2,m}(\Delta^*) \leq q)], \quad (31)$$

or equivalently if

$$\sum_{r=0}^{\infty} \Gamma\left(\frac{p+2+m+2r}{2}\right) < \frac{(p-2)}{(m+2)} \sum_{r=0}^{\infty} \frac{\Gamma\left(\frac{p+2+m}{2}\right)}{\Gamma\left(\frac{p+2r}{2}\right)}. \quad (32)$$

The SE has a smaller bias than the PRSE if the opposite inequality holds. Since for given  $p$ ,  $n_1$  and  $n_2$ ,  $m$  is always fixed, the infinite sum on the left hand side of (32) will always be greater than that on the right hand side. Therefore, the SE always has a smaller bias than the PRSE under the alternative hypothesis.

## 4 The Quadratic Risk

Let  $\hat{\boldsymbol{\theta}}$  be an estimator of  $\boldsymbol{\theta}$  based on a sample of size  $n$  from a population with mean  $\boldsymbol{\theta}$  and covariance matrix  $\Omega$ . The quadratic risk of  $\hat{\boldsymbol{\theta}}$  in estimating  $\boldsymbol{\theta}$  is defined as the expected loss of the estimator as follows:

$$R(\hat{\boldsymbol{\theta}}; \Omega) = E[n(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})'\Omega^{-1}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})]. \quad (33)$$

Note that the quadratic loss function of  $\hat{\boldsymbol{\theta}}$  in estimating  $\boldsymbol{\theta}$  is given by

$$L(\hat{\boldsymbol{\theta}}; \Omega) = n(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})'\Omega^{-1}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}). \quad (34)$$

In this section, we derive the risks for the SE and PRSE of  $\boldsymbol{\mu}_1$  for the above loss function.

**Theorem 4.1.** *For the multivariate Student-t model defined in Section 1, the quadratic risk of the SE of  $\boldsymbol{\mu}_1$  is given by*

$$R(\hat{\boldsymbol{\mu}}_1^S; \boldsymbol{\mu}_1) = p - CmM[2(p-2) - C(m+2)]E^{(0)}[\chi_p^{-2}(\Delta^*)] \quad (35)$$



where

$$E^{(0)}[\chi_p^{-2}(\Delta^*)] = \sum_{r=0}^{\infty} \frac{1}{(p-2+2r)} \frac{\Gamma(\frac{\nu}{2}+r)}{r!\Gamma(\frac{\nu}{2})} \frac{\left(\frac{\Delta^*}{\nu-2}\right)^r}{\left(1+\frac{\Delta^*}{\nu-2}\right)^{\frac{\nu}{2}+r}}. \quad (36)$$

**Proof.** By definition, the quadratic risk of the SE of  $\boldsymbol{\mu}_1$  is

$$R(\hat{\boldsymbol{\mu}}_1^S; \boldsymbol{\mu}_1) = E_{\tau} \left\{ E \left[ n_1 (\hat{\boldsymbol{\mu}}_1^S - \boldsymbol{\mu}_1)' \frac{\Sigma^{-1}}{\tau^2} (\hat{\boldsymbol{\mu}}_1^S - \boldsymbol{\mu}_1) | \tau \right] \right\}. \quad (37)$$

Now applying the representation of  $(\hat{\boldsymbol{\mu}}_1^S - \boldsymbol{\mu}_1)$  as given in (12) in the above quadratic form, and completing the multiplication we get, for a given value of  $\tau$

$$\begin{aligned} R(\hat{\boldsymbol{\mu}}_1^S; \boldsymbol{\mu}_1 | \tau) &= n_1 E[(\tilde{\boldsymbol{\mu}}_1 - \boldsymbol{\mu}_1)' \frac{\Sigma^{-1}}{\tau^2} (\tilde{\boldsymbol{\mu}}_1 - \boldsymbol{\mu}_1) | \tau] \\ &\quad + n_1 C^2 M^2 E \left[ T^{-4} (\tilde{\boldsymbol{\mu}}_2 - \boldsymbol{\mu}_1)' \frac{\Sigma^{-1}}{\tau^2} (\tilde{\boldsymbol{\mu}}_2 - \tilde{\boldsymbol{\mu}}_1) | \tau \right] \\ &\quad + 2n_1 C M E \left[ T^{-2} (\tilde{\boldsymbol{\mu}}_1 - \boldsymbol{\mu}_1)' \frac{\Sigma^{-1}}{\tau^2} (\tilde{\boldsymbol{\mu}}_2 - \tilde{\boldsymbol{\mu}}_1) | \tau \right]. \end{aligned} \quad (38)$$

Applying the following result on the conditional expectation,

$$E[(\tilde{\boldsymbol{\mu}}_1 - \boldsymbol{\mu}_1) | (\tilde{\boldsymbol{\mu}}_2 - \tilde{\boldsymbol{\mu}}_1)] = -M\{(\tilde{\boldsymbol{\mu}}_2 - \tilde{\boldsymbol{\mu}}_1) - (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)\} \quad (39)$$

in the last term of the r.h.s. of (38), and simplifying, the quadratic risk of the SE becomes

$$\begin{aligned} R(\hat{\boldsymbol{\mu}}_1^S; \boldsymbol{\mu}_1 | \tau) &= p + C^2 M^2 n_1 E \left[ T^{-4} (\tilde{\boldsymbol{\mu}}_2 - \tilde{\boldsymbol{\mu}}_1)' \frac{\Sigma^{-1}}{\tau^2} (\tilde{\boldsymbol{\mu}}_2 - \tilde{\boldsymbol{\mu}}_1) | \tau \right] \\ &\quad - 2CM^2 n_1 \left\{ E \left[ T^{-2} (\tilde{\boldsymbol{\mu}}_2 - \tilde{\boldsymbol{\mu}}_1)' \frac{\Sigma^{-1}}{\tau^2} (\tilde{\boldsymbol{\mu}}_2 - \tilde{\boldsymbol{\mu}}_1) | \tau \right] \right. \\ &\quad \left. - (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)' E \left[ T^{-2} \frac{\Sigma^{-1}}{\tau^2} (\tilde{\boldsymbol{\mu}}_2 - \tilde{\boldsymbol{\mu}}_1) | \tau \right] \right\} \end{aligned} \quad (40)$$

where  $E[n_1 (\tilde{\boldsymbol{\mu}}_1 - \boldsymbol{\mu}_1)' \frac{\Sigma^{-1}}{\tau^2} (\tilde{\boldsymbol{\mu}}_1 - \boldsymbol{\mu}_1) | \tau] = p$  is the quadratic risk of the mle of  $\boldsymbol{\mu}_1$ . Then using the transformation in (18) and the results in the Appendix B.2 of Judge and Bock (1978) we have, conditional on  $\tau$ ,



$$\begin{aligned}
& E[T^{-4}(\tilde{\boldsymbol{\mu}}_2 - \tilde{\boldsymbol{\mu}}_1)' \frac{\Sigma^{-1}}{\tau^2} (\tilde{\boldsymbol{\mu}}_2 - \tilde{\boldsymbol{\mu}}_1) | \tau] \\
&= \frac{1}{n_1 M} E \left[ \frac{\chi_m^4}{Y'Y} \right] = \frac{m(m+2)}{n_1 M} E[\chi_p^{-2}(\Delta_\tau)] \quad (41a)
\end{aligned}$$

$$E[T^{-2}(\tilde{\boldsymbol{\mu}}_2 - \tilde{\boldsymbol{\mu}}_1)' \frac{\Sigma^{-1}}{\tau^2} (\tilde{\boldsymbol{\mu}}_2 - \tilde{\boldsymbol{\mu}}_1) | \tau] = \frac{m}{n_1 M} \quad (41b)$$

$$E[T^{-2} \frac{\Sigma^{-1}}{\tau^2} (\tilde{\boldsymbol{\mu}}_2 - \tilde{\boldsymbol{\mu}}_1) | \tau] = m \frac{\Sigma^{-1}}{\tau^2} \delta E[\chi_{p+2}^{-2}(\Delta_\tau)]. \quad (41c)$$

Substituting the above results in (40) we get

$$\begin{aligned}
R(\hat{\boldsymbol{\mu}}_1^S; \boldsymbol{\mu}_1 | \tau) &= p + C^2 M m (m+2) E[\chi_p^{-2}(\Delta_\tau)] - 2 C M m \\
&\quad + 2 C M m \Delta_\tau E[\chi_{p+2}^{-2}(\Delta_\tau)]. \quad (42)
\end{aligned}$$

Using the identity

$$\Delta_\tau E[\chi_{p+2}^{-2}(\Delta_\tau)] = 1 - (p-2) E[\chi_p^{-2}(\Delta_\tau)] \quad (43)$$

in the last term of (42), and simplifying, the expression in (42) reduces to

$$R(\hat{\boldsymbol{\mu}}_1^S; \boldsymbol{\mu}_1 | \tau) = p - C M m [2(p-2) - C(m+2)] E[\chi_p^{-2}(\Delta_\tau)]. \quad (44)$$

The expression for the quadratic risk of the SE of  $\boldsymbol{\mu}_1$ , as given in (35), is then obtained by taking expectation on  $R(\hat{\boldsymbol{\mu}}_1^S; \boldsymbol{\mu}_1 | \tau)$  with respect to the distribution of  $\tau$ . Hence the proof.

**Theorem 4.2.** For the multivariate Student-t model defined in Section 1, the quadratic risk of the PRSE of  $\boldsymbol{\mu}_1$  is given by

$$\begin{aligned}
R(\hat{\boldsymbol{\mu}}_1^{S+}; \boldsymbol{\mu}_1) &= R(\hat{\boldsymbol{\mu}}_1^S; \boldsymbol{\mu}_1) \\
&\quad - \Delta^* \left\{ M G_{p+4,m}^{(4)}(q_4; \Delta^*) + 2 G_{p+2,m}^{(2)}(q_2; \Delta^*) + 2 C m \right. \\
&\quad \times \left. E^{(2)}[\chi_{p+2}^{-2}(\Delta^*) I(F_{p+2,m}(\Delta^*) \leq q_2)] \right\} \\
&\quad - M \left\{ p G_{p+2,m}^{(2)}(q_2; \Delta^*) - 2 C m G_{p,m}^{(0)}(q_0; \Delta^*) \right\} \\
&\quad - C^2 M m (m+2) E^{(2)}[\chi_{p+2}^{-2}(\Delta^*) I(F_{p+2,m}(\Delta^*) \leq q_2)] \quad (45)
\end{aligned}$$

where  $R(\hat{\boldsymbol{\mu}}_1^S; \boldsymbol{\mu}_1)$  is the quadratic risk of the SE of  $\boldsymbol{\mu}_1$  as given in (35);

$$G_{p+h,m}^{(h)}(q_h; \Delta^*) = \sum_{r=0}^{\infty} \frac{\Gamma\left(\frac{p+h+m}{2} + r\right)}{\Gamma\left(\frac{p+h}{2} + r\right) \Gamma\left(\frac{m}{2}\right)} \frac{\Gamma\left(\frac{\nu}{2} + r\right) B_{u_h}\left(\frac{m}{2}, \frac{p+h}{2} + r\right)}{r! \Gamma\left(\frac{\nu}{2}\right)} \frac{\left(\frac{\Delta^*}{\nu-2}\right)^r}{\left(1 + \frac{\Delta^*}{\nu-2}\right)^{\frac{\nu}{2}+r}} \quad (46)$$



with  $u_h = \frac{(m+2)(p+h)}{(m+2)(p+h)+m(p-2)}$  for  $h = 0, 2, 4$ ; and

$$\begin{aligned} & E^{(2)}[\chi_{p+2}^{-2}(\Delta^*)I(F_{p+2,m}(\Delta^*) \leq q_2)] \\ &= \sum_{r=0}^{\infty} \frac{\Gamma\left(\frac{p+2+m}{2}\right)}{\Gamma\left(\frac{p}{2}+r\right)\Gamma\left(\frac{m}{2}\right)} \frac{\Gamma\left(\frac{\nu}{2}+r\right)B_{u_2}\left(\frac{m}{2}, \frac{p+2}{2}+r\right)}{r!\Gamma\left(\frac{\nu}{2}\right)\Gamma\left(\frac{p+2}{2}+r\right)} \frac{\left(\frac{\Delta^*}{\nu-2}\right)^r}{\left(1+\frac{\Delta^*}{\nu-2}\right)^{\frac{\nu}{2}+r}}, \quad (47) \end{aligned}$$

in which  $B_{u_2}(a, b)$  is the incomplete beta function with arguments 'a' and 'b', and evaluated at  $u_2$ .

**Proof.** By definition, the quadratic risk of the PRSE of  $\mu_1$  is given by

$$R(\hat{\mu}_1^{S+}; \mu_1) = E_{\tau} \left\{ E[n_1(\hat{\mu}_1^{S+} - \mu_1)' \frac{\Sigma^{-1}}{\tau^2} (\hat{\mu}_1^{S+} - \mu_1) | \tau] \right\}. \quad (48)$$

Using the presentation of  $(\hat{\mu}_1^{S+} - \mu_1)$  as given in (13), expanding the resulting quadratic forms and simplifying the terms, the above risk function, conditional on a given value of  $\tau$ , can be written as

$$\begin{aligned} R(\hat{\mu}_1^{S+}; \mu_1 | \tau) &= E[n_1(\tilde{\mu}_1 - \mu_1)' \frac{\Sigma^{-1}}{\tau^2} (\tilde{\mu}_1 - \mu_1) | \tau] \\ &+ C^2 M^2 n_1 E[T^{-4}(\tilde{\mu}_2 - \tilde{\mu}_1)' \frac{\Sigma^{-1}}{\tau^2} (\tilde{\mu}_2 - \tilde{\mu}_1) | \tau] \\ &+ M^2 n_1 E[(\tilde{\mu}_2 - \tilde{\mu}_1)' \frac{\Sigma^{-1}}{\tau^2} (\tilde{\mu}_2 - \tilde{\mu}_1) I(T^2 \leq C) | \tau] \\ &+ C^2 M^2 n_1 E[T^{-4}(\tilde{\mu}_2 - \tilde{\mu}_1)' \frac{\Sigma^{-1}}{\tau^2} (\tilde{\mu}_2 - \tilde{\mu}_1) I(T^2 \leq C) | \tau] \\ &+ 2CM n_1 E[(\tilde{\mu}_1 - \mu_1)' \frac{\Sigma^{-1}}{\tau^2} (\tilde{\mu}_2 - \tilde{\mu}_1) | \tau] \\ &+ 2M n_1 E[(\tilde{\mu}_1 - \mu_1)' \frac{\Sigma^{-1}}{\tau^2} (\tilde{\mu}_2 - \tilde{\mu}_1) I(T^2 \leq C) | \tau] \\ &- 2CM n_1 E[T^{-2}(\tilde{\mu}_1 - \mu_1)' \frac{\Sigma^{-1}}{\tau^2} (\tilde{\mu}_2 - \tilde{\mu}_1) I(T^2 \leq C) | \tau] \\ &+ 2CM^2 n_1 E[T^{-2}(\tilde{\mu}_2 - \tilde{\mu}_1)' \frac{\Sigma^{-1}}{\tau^2} (\tilde{\mu}_2 - \tilde{\mu}_1) I(T^2 \leq C) | \tau] \\ &- 2C^2 M^2 n_1 E[T^{-4}(\tilde{\mu}_2 - \tilde{\mu}_1)' \frac{\Sigma^{-1}}{\tau^2} (\tilde{\mu}_2 - \tilde{\mu}_1) I(T^2 \leq C) | \tau] \\ &- 2CM^2 n_1 E[T^{-2}(\tilde{\mu}_2 - \tilde{\mu}_1)' \frac{\Sigma^{-1}}{\tau^2} (\tilde{\mu}_2 - \tilde{\mu}_1) I(T^2 \leq C) | \tau], \quad (49) \end{aligned}$$



where  $I^2(A) = I(A)$  for the indicator function,  $I(A)$ , of the set  $A$ , has been used.

Clearly, the first term in (49) is  $p$ , the quadratic risk of the mle of  $\boldsymbol{\mu}_1$ . For the evaluation of the other 9 terms, we use the transformation in (18), and apply the results from Judge and Bock (1978), as before. Let  $t_i$  denote the  $i$ -th term in (49) for  $i = 1, 2, \dots, 10$ . Then we have that  $t_1 = p$ ; and

$$\begin{aligned} t_2 &= C^2 M^2 n_1 E \left[ \frac{\chi_4^2}{Y'Y} \frac{1}{n_1 M} \right] = C^2 M m(m+2) E[\chi_p^{-2}(\Delta_\tau)]; \\ t_3 &= M^2 n_1 E \left[ \frac{Y'Y}{n_1 M} I \left( \frac{Y'Y}{\chi_m^2} \leq C \right) \right] \\ &= M p G_{p+2}(q_2; \Delta_\tau) + M \Delta_\tau G_{p+4,m}(q_4; \Delta_\tau); \\ t_4 &= C^2 M^2 n_1 E \left[ \frac{\chi_m^4}{Y'Y} \frac{1}{n_1 M} I \left( \frac{Y'Y}{\chi_m^2} \leq C \right) \right] \\ &= C^2 M m(m+2) E \left[ \chi_{p+2}^{-2}(\Delta_\tau) I(F_{p+2,m}(\Delta_\tau) \leq q_2) \right]; \\ t_5 &= -2 C M^2 n_1 E \left[ \frac{\chi_m^2}{n_1 M} \right] + 2 C \Delta_\tau E \left[ \frac{\chi_m^2}{Y'Y} \right] \\ &= -2 C M m + 2 C m \Delta_\tau E[\chi_{p+2}^{-2}(\Delta_\tau)], \end{aligned}$$

where the result on the conditional expectation of  $(\tilde{\boldsymbol{\mu}}_1 - \boldsymbol{\mu}_1)$ , given  $(\tilde{\boldsymbol{\mu}}_2 - \tilde{\boldsymbol{\mu}}_1)$ , has been applied from (39);

$$\begin{aligned} t_6 &= -2 M^2 n_1 E \left[ \frac{Y'Y}{n_1 M} I \left( \frac{Y'Y}{\chi_m^2} \leq C \right) \right] + 2 M n_1 \boldsymbol{\delta}' \frac{\Sigma^{-1}}{\tau^2} E \left[ \frac{\tau \Sigma^{1/2}}{\sqrt{M n_1}} Y I \left( \frac{Y'Y}{\chi_m^2} \leq C \right) \right] \\ &= -2 M p G_{p+2,m}(q_2; \Delta_\tau) - 2 M \Delta_\tau G_{p+4,m}(q_4; \Delta_\tau) + 2 \Delta_\tau G_{p+2,m}(q_2; \Delta_\tau) \end{aligned}$$

in which the previous result on the conditional expectation has been used and the expression has been simplified;

$$\begin{aligned} t_7 &= 2 C M^2 n_1 E \left[ \frac{\chi_m^2}{M n_1} I \left( \frac{Y'Y}{\chi_m^2} \leq C \right) \right] + 2 C M n_1 E \left[ \frac{\chi_m^2}{Y'Y} \boldsymbol{\delta}' \frac{\Sigma^{-1}}{\tau \sqrt{M n_1}} Y I \left( \frac{Y'Y}{\chi_m^2} \leq C \right) \right] \\ &= 2 C M m G_{p,m}(q; \Delta_\tau) + 2 C m \Delta_\tau E[\chi_{p+2}^{-2}(\Delta_\tau) I(F_{p+2,m}(\Delta_\tau) \leq q_2)] \end{aligned}$$

when simplified after using the earlier conditional expectation and results from Judge and Bock (1978);

$$\begin{aligned} t_8 &= 2 C M^2 n_1 E \left[ \frac{\chi_m^2}{M n_1} I \left( \frac{Y'Y}{\chi_m^2} \leq C \right) \right] = 2 C M m G_{p,m}(q; \Delta_\tau); \\ t_9 &= -2 C^2 M^2 n_1 E \left[ \frac{\chi_m^4}{M n_1} I \left( \frac{Y'Y}{\chi_m^2} \leq C \right) \right] \\ &= -2 C^2 M m(m+2) E[\chi_{p+2}^{-2}(\Delta_\tau) I(F_{p+2,m}(\Delta_\tau) \leq q_2)]; \text{ and} \end{aligned}$$



$$t_{10} = -2CM^2n_1E\left[\frac{\chi_m^2}{Mn_1}I\left(\frac{Y'Y}{\chi_m^2} \leq C\right)\right] = -2CMmG_{p,m}(q; \Delta_\tau). \quad (50)$$

Now collecting  $t_1 - t_{10}$  in (49), conditional on a given value of  $\tau$ , the quadratic risk of the PRSE of  $\mu_1$  becomes, on regrouping of the terms and simplification,

$$\begin{aligned} R(\hat{\mu}_1^{S+}; \mu_1|\tau) &= \left\{p + C^2Mm(m+2)E[\chi_{p+2}^{-2}(\Delta_\tau)] - 2CMm + 2Cm\Delta_\tau E[\chi_{p+2}^{-2}(\Delta_\tau)]\right\} \\ &- \Delta_\tau \left\{MG_{p+4,m}(q_4; \Delta_\tau) + 2G_{p+2,m}(q_2; \Delta_\tau) + 2CmE[\chi_{p+2}^{-2}(\Delta_\tau)I(F_{p+2,m}(\Delta_\tau) \leq q_2)]\right\} \\ &- M\{pG_{p+2,m}(q_2; \Delta_\tau) - 2CmG_{p,m}(q; \Delta_\tau)\} \\ &- C^2Mm(m+2)E[\chi_{p+2}^{-2}(\Delta_\tau)I(F_{p+2,m}(\Delta_\tau) \leq q_2)] \end{aligned} \quad (51)$$

where the terms inside the first curly brackets is  $R(\hat{\mu}_1^S; \mu_1|\tau)$ , the quadratic risk of the SE of  $\mu_1$ , for a given value of  $\tau$ ;  $q_h = \frac{m}{p+h} \times \frac{p-2}{m+2}$  for  $h = 0, 2, 4$  and in which  $C_{opt} = \frac{p-2}{m+2}$ , as before.

The final expression in (45) is obtained by computing the expected value of the  $R(\hat{\mu}_1^{S+}; \mu_1|\tau)$  with respect to the distribution of  $\tau$ . Hence the proof.

## 5 Analysis of Risk

It is well known that the quadratic risk of the mle of  $\mu_1$  is a constant and is equal to the dimension of the vector  $\mathbf{X}_i$  for  $i = 1, 2$ . Hence we have, for the model under consideration,  $R(\tilde{\mu}_1; \mu_1) = p$ . Thus the risk function of  $\tilde{\mu}_1$  does not depend on  $\nu$  and  $\Delta$ .

The risk of the SE,  $\hat{\mu}_1^S$  is a monotone function of  $\Delta^*$ . The minimum of  $R(\hat{\mu}_1^S; \mu_1)$  is attained when  $\Delta^* = 0$  with a value of  $\left\{p - \frac{(p-2)^2}{(m+2)}mM\right\}$ , and the maximum value of  $R(\hat{\mu}_1^S; \mu_1)$  approaches to  $p$  as  $\Delta^* \rightarrow \infty$ . Therefore, the quadratic risk of the SE is always smaller than that of the mle of  $\mu_1$  for all values of  $\Delta$  and  $\nu > 2$ . It can be shown that for the multivariate normal model,  $R(\hat{\mu}_1^S; \mu) \rightarrow p$  as  $\Delta$  grows larger. However, since  $\Delta^* < \Delta$  for all  $\nu > 2$ , the risk function of the SE for the multivariate normal model approaches to that of the mle faster than (for a smaller value of  $\Delta$ ) that for the multivariate Student-t model. Furthermore, as  $\nu \rightarrow \infty$ , the difference in the values of  $\Delta$  and  $\Delta^*$  becomes negligible, and as a result both  $R(\tilde{\mu}_1; \mu_1)$  and  $R(\hat{\mu}_1^S; \mu_1)$  approach to  $p$  in about the same pace. Nevertheless, for the smaller values of  $\nu$ , the risk of the mle grows larger and approaches to  $p$  faster than that of the SE as  $\Delta$  increases. Therefore, for the multivariate Student-t model the Stein-type shrinkage estimator dominates the usual mle of the mean vector, and this domination is over a wider range of values of  $\Delta^*$  than that of  $\Delta$  for the multivariate normal model. Thus  $\tilde{\mu}_1$  for  $\mu_1$  is inadmissible for the two-sample Student-t problem. Khan and Hoque (2002) provided similar results for the the independent normal models.



The risk function of the PRSE of  $\mu_1$  is also a function of  $\Delta^*$ . From (45) it is clear that the risk of the PRSE is smaller than that of the SE for all values of  $\Delta^*$ . Similar to the risk function of the SE, the risk function of PRSE,  $R(\hat{\mu}_1^{S+}; \mu_1)$  increases as the value of  $\Delta^*$  grows larger. It also approaches to the risk of the mle as  $\Delta^* \rightarrow \infty$ . But for any particular value of  $\Delta^*$ , the amount of risk of PRSE is always at least as low as that of the SE. Hence the PRSE always dominates over the SE uniformly over all the values of  $\Delta^*$ . Therefore, the SE is inadmissible against the PRSE for all values of  $\Delta$  and  $\nu > 2$ . This is also true for the multivariate normal model, since the normal model is a special (limiting) case of the Student-t model.

Comparing the risk of the PRSE of the multivariate normal model with the multivariate Student-t model, it is observed that the later approaches to the risk of the mle ( $p$ ) slower than the former as  $\Delta^*$  is always smaller than  $\Delta$  for all values of  $\nu > 2$ . Hence a relatively larger departure of the value of  $\mu_1$  from its value under the null hypothesis will not increase the amount of the risk of the PRSE for the Student-t model as much as it would do for its normal counter-part. For the same amount of departure of  $\mu_1$  from its hypothesized value will lead to a higher level of risk for the PRSE of the normal model than the Student-t model.

Apart from removing the instability of the SE for very small values of  $T^2$  statistic, the PRSE overperforms the SE with respect to the quadratic risk criterion. Moreover, the PRSE is uniformly superior to the SE for all values of  $\Delta^*$ . Since the SE dominates the mle of  $\mu_1$ , the performance picture of the estimators can be summarized as follows

$$\hat{\mu}_1^{S+} \succ \hat{\mu}_1^S \succ \tilde{\mu}_1 \quad (52)$$

where the symbol “ $\succ$ ” stands for domination.

## 6 The Relative Efficiency

The relative efficiency of the SE with respect to the mle of  $\mu_1$  is given by

$$\begin{aligned} \eta(\hat{\mu}_1^S : \tilde{\mu}) &= \frac{R(\tilde{\mu}_1; \mu_1)}{R(\hat{\mu}_1^S; \mu_1)} \\ &= \left\{ 1 - \frac{mM(p-2)^2}{(m+2)p} E[\chi_p^{-2}(\Delta^*)] \right\}^{-1}, \end{aligned} \quad (53)$$

which is a monotone decreasing function of  $\Delta^*$ . The maximum relative efficiency of the SE relative to the mle is attained at  $\Delta^* = 0$  with the maximum value  $\left\{ 1 - \frac{M(p-2)m}{(m+2)p} \right\}^{-1}$ .

On the other hand  $\eta(\hat{\mu}_1^S : \tilde{\mu}_1) \rightarrow 1$  as  $\Delta^*$  grows larger. Thus for a smaller value of  $\Delta^*$  the SE performs better than the mle of  $\mu_1$ . However, for a larger value of  $\Delta^*$ , the relative efficiency of the SE with respect to the mle may not be significantly different



from 1. Nevertheless, the SE dominates over the mle of  $\mu_1$  for all values of  $\Delta^*$ . But this domination is more significant for all the values of  $\Delta^*$  in the neighborhood of zero.

The relative efficiency of the PRSE with respect to the SE of  $\mu_1$  is given by

$$\eta(\hat{\mu}_1^{S+}; \hat{\mu}_1^S) = \frac{R(\hat{\mu}_1^S; \mu_1)}{R(\hat{\mu}_1^{S+}; \mu_1)} = \{1 - \psi_1(\Delta^*)\}^{-1} \quad (54)$$

where  $\psi_1(\Delta^*) = \left\{ R(\hat{\mu}_1^S; \mu) \right\}^{-1} \{ \psi_2(\Delta^*) \}$  in which

$$\begin{aligned} \psi_2(\Delta^*) &= \Delta^* \left\{ MG_{p+4,m}^{(4)}(q_4; \Delta^*) + 2G_{p+2,m}^{(2)}(q_2; \Delta^*) \right. \\ &\quad \left. + 2CmE^{(2)}[\chi_p^{-2}(\Delta^*)I(F_{p+2,m}(\Delta^*) \leq q)] \right\} \\ &\quad + M \left\{ pG_{p+2,m}^{(2)}(q_2; \Delta^*) - 2CmG_{p,m}^{(0)}q_0; \Delta^* \right\} \\ &\quad + C^2 Mm(m+2)E^{(2)}[\chi_p^{-2}(\Delta^*)I(F_{p+2,m}(\Delta^*) \leq q_2)], \quad (55) \end{aligned}$$

which is a monotone decreasing function of  $\Delta^*$ . Therefore,  $\psi_1(\Delta^*)$  increases as  $\psi_2(\Delta^*)$  increases when  $\Delta^*$  grows smaller. Thus the relative efficiency of the PRSE with respect to the SE approaches its maximum at  $\Delta^* = 0$ . On the other extreme, the relative efficiency of the PRSE of  $\mu_1$  reduces to 1 as  $\Delta^* \rightarrow \infty$ . Thus, for any value of  $\Delta^*$  the relative efficiency of the PRSE with respect to the SE is at least 1. Hence the PRSE uniformly performs better than the SE for all values of  $\Delta^*$ , and this is more so when the value of  $\Delta^*$  is not too far from zero.

Finally, since the relative efficiency of the SE with respect to the mle is at least 1 for all values of  $\Delta^*$ , and the relative efficiency of the PRSE with respect to the SE is at least 1 for all values of  $\Delta^*$ , the PRSE performs even better than the mle with a higher level of relative efficiency.

## 7 Concluding Remarks

The foregoing study reveals that the mle of  $\mu_1$  for the two-sample multivariate Student-t problem is unbiased, as it uses the information from the first sample alone. On the other hand both the SE and PRSE of  $\mu_1$  are biased. These estimators used information from both the samples as well as the uncertain prior information about the equality of the location vectors. The amount of bias of the later two estimators depends on the value of  $\Delta$  which measures the extent of deviation of the actual value of  $\mu_1$  from its value specified under the null hypothesis. In general, the PRSE has a higher bias than the SE under the alternative hypothesis. However, under the null hypothesis  $\Delta = 0$  and all three estimators are unbiased. In real life, we are seldom sure about



the validity of the null hypothesis, and we are seldom sure of the correctness of the null hypothesis and hence on the value of  $\Delta$ . Thus on the basis of the criterion of unbiasedness, the mle of  $\mu_1$  is a better choice over the SE or PRSE, when the null hypothesis is in doubt.

It is evident from the analysis of the quadratic risk that the mle of  $\mu_1$  has the highest risk among the three estimators under investigation. Under the null hypothesis the dominance picture of the estimators is clear and has been stated in (52). Similar to the multivariate normal problem the mle is not admissible relative to the SE or PRSE for the two-sample multivariate Student-t problem when  $\nu > 2$  when the null hypothesis is true. Moreover, under the alternative, the SE dominates over the mle as the SE has a higher relative efficiency than the mle. Again, the PRSE of the  $\mu_1$  has a higher relative efficiency than the SE for all  $\Delta$  and  $\nu > 2$ . Therefore, in the face of uncertainty on the value of the null hypothesis, the PRSE overperforms the other two estimators of  $\mu_1$ . When the objective is to minimize the quadratic risk, the biased estimators, both SE and PRSE, perform better than the unbiased estimator, the mle. Finally, the PRSE uniformly dominates over the SE as well as the mle for all values of  $\Delta$  and any number of degrees of freedom,  $\nu$ .

## Acknowledgment

The first author thankfully acknowledges the excellent research facilities provided by the Sultan Quabbos University, Muscat, Oman. This research was partially funded by the University of Southern Queensland, Australia through the Academic Development Leave program.

## References

- Ahmed, S.E. and Saleh, A.K. Md. E. (1989). Pooling multivariate data. *Jou. Statist. Comput. Simul.*, 31, 149–167.
- 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.
- Bancroft, T.A. (1944). On biases in estimation due to the use of preliminary tests of significance. *Annals of Mathematical Statistics*, 15, 190–204.
- Box, G.E.P. and Taio, G.C. (1990). *Bayesian Inference in Statistical Analysis*. Wiley, New York.
- Cornish, E.A. (1954). Multivariate  $t$ -distribution associated with a set of normal sample deviates. *Australian Journal of Physics*, 7, 531–542.
- Fisher, R.A. (1956). *Statistical Methods in Scientific Inference*. Oliver and Boyd, London.
- Fisher, R.A. (1960). *The Design of Experiments*. 7th ed., Hafner, New York.



- Fraser, D.A.S. and Fick, G.H. (1975). Necessary analysis and its implementation. *In Proc. Symposium on Statistics and Related Topics*. ed. by A.K. Md. E. Saleh, Carleton University, Canada, 5.01–5.30.
- Fraser, D.A.S. (1979). *Inference and Linear Models*, McGraw–Hill, New York.
- Giles, J.A. (1991). Pre-testing for linear restrictions model with spherically symmetric disturbances. *Journal of Econometrics*, 50, 377–398.
- James, W. and Stein, C. (1961). Estimation with quadratic loss. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley, California, 361–379.
- Judge, G.C. and Bock, M.E. (1978). *Statistical Implications of Pre-test and Stein-rule Estimators in Econometrics*. North Holland, Amsterdam.
- Khan, S. (1992). *Predictive Inference for Multilinear Model with Multivariate Student-t Errors*. Unpublished Ph.D. Thesis, Department of Statistical and Actuarial Sciences, University of Western Ontario, Canada.
- Khan, S. (1997). Likelihood based inference on the mean vectors of two multivariate Student-t populations. *Far East Journal of Theoretical Statistics*, 1, 1–17.
- Khan, S. (1998). On the estimation of mean vector of Student-t population with uncertain prior information. *Pakistan Journal of Statistics*, 14, 161–175.
- Khan, S. and Z. Hoque (2002). James-Stein estimators for the mean vector of a multivariate normal population based on independent samples from two normal populations with common covariance structure. *Pakistan Journal of Statistics*. Special Issue Vol. 18, No.1. Forthcoming.
- 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.
- Khan, S. and Saleh, A.K.Md.E. (1998). Comparison of estimators of the mean based on  $p$ -samples from multivariate Student-t population. *Communications in Statistics: Theory & Methods*, 27(1), 193–210.
- Lange, K.L. Little, R.J.A., and Taylor, J.M.G. (1989). Robust statistical modeling using the  $t$ -distribution. *Journal of American Statistical Association*, 84, 881–896.
- Lucas, A. (1997). Robustness of the Student-t based  $M$ -estimator. *Communications in Statistics*, 26(5), 1165–1182.
- Prucha I.R. and Kelegian, H.H. (1984). The structure of simultaneous equation estimators: A generalization towards non-normal disturbances, *Econometrica*, 52, 721–736.
- 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.
- Spanos, A. (1994). On modeling heteroskedasticity: The Student's  $t$  and elliptical linear regression models. *Econometric Theory*, 10, 286–315.
- Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. *Proceedings of the Third Berkeley Symposium on Math.*



- Statist. and Probability*, University of California Press, Berkeley, 1, 197–206.
- Ullah, A. and Walsh, V.Z. (1984). On the robustness of LM, LR and W tests in regression models. *Econometrica*, 52, 1055–1066.
- Zellner, A. (1976). Bayesian and non-Bayesian analysis of the regression model with multivariate Student-t error term. *Journal of American Statistical Association*, 66, 601–616.