# On the comparison of the pre-test and shrinkage estimators for the univariate normal mean

**Shahjahan Khan**
Department of Mathematics and Computing
University of Southern Queensland
Toowoomba, Queensland, Australia
and
**A.K.Md.E. Saleh**
School of Mathematics and Statistics
Carleton University, Ottawa, Canada

**Abstract**

The estimation of the mean of an univariate normal population with unknown variance is considered when uncertain non-sample prior information is available. Alternative estimators are defined to incorporate both the sample as well as the non-sample information in the estimation process. Some of the important statistical properties of the restricted, preliminary test, and shrinkage estimators are investigated. The performances of the estimators are compared based on the criteria of unbiasedness and mean square error in order to search for a 'best' estimator. Both analytical and graphical methods are explored. There is no superior estimator that uniformly dominates the others. However, if the non-sample information regarding the value of the mean is close to its true value, the shrinkage estimator over performs the rest of the estimators.

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

**Keywords and Phrases**: Uncertain non-sample prior information; maximum likelihood, restricted, preliminary test and shrinkage estimators; bias, mean square error and relative efficiency; normal, Student-t, non-central chi-square and F distributions; and incomplete beta ratio.

## 1 Introduction

Traditionally the classical estimators of unknown parameters are based exclusively on the sample data. Such estimators disregard any other kind of non-sample prior information in its definition. The notion of inclusion of non-sample information to the estimation of parameters has been introduced to 'improve' the quality of the estimators. The natural expectation is that the inclusion of additional information would result in a better estimator. In some cases this may be true, but in many other cases the risk of worse consequences can not be ruled out. A number of estimators have been introduced in the literature that, under particular situation, over performs the traditional exclusive sample data based unbiased estimators when judged by criteria such as the mean square error and square error loss function.

There has been many studies in the area of the 'improved' estimation following the seminal work of Bancroft (1944) and later Han and Bancroft (1968). They developed the preliminary test estimator that uses uncertain non-sample prior information (not in the form of prior distributions), in addition to the sample information. Stein (1956) introduced the Stein-rule (shinkage) estimator for multivariate normal population that dominates the usual maximum likelihood estimators under the square error loss function. In a series of papers Saleh and Sen (1978, 1985) explored the preliminary test approach to Stein-rule estimation. Many authors have contributed to this area, notably Sclove et al. (1972), Judge and Bock (1978), Stein (1981), Maatta and Casella (1990), and Khan (1998), to mention a few. Ahmed and Saleh (1989) provided comparison of several improved estimators for two multivariate normal populations with a common covariance matrix. Later Khan and Saleh (1995, 1997) investigated the problem for a family of Student-t populations. However, the relative performance of the preliminary test and shrinkage estimators of the univariate normal mean has not been investigated.

Let $Y_1, Y_2, \cdots, Y_n$ be a random sample of size $n$ from a normal population with unknown mean $\theta$ and unknown but common variance $\sigma^2$. In the conventional notation we write, $Y \sim N(\theta,\ \sigma^2)$. Also assume that *uncertain non-sample prior information* on the value of $\theta$ is available, either from previous study or from practical experience of the researchers or experts. Let the non-sample prior information be expressed in the form of a null hypothesis, $H_0 :\ \theta = \theta_0$ which may be true, but not sure. We wish to incorporate both the sample data and the uncertain non-sample prior information in estimating the mean $\theta$. First we obtain the *unrestricted* maximum likelihood estimator (mle) of the unknown mean $\theta$ and the common variance $\sigma^2$

from the likelihood function of the sample. Based on the unrestricted and *restricted* (by the null hypothesis) mle of $\sigma^2$, we derive the likelihood ratio test for testing $H_0: \ \theta = \theta_0$ against $H_A: \ \theta \neq \theta_0$. Then use the test statistic, as well as the sample and non-sample information to define the preliminary test and shrinkage estimators of the unknown population mean.

It is well known that the mle of the population mean is unbiased. We wish to search for an alternative estimator of the mean that is biased but may well have some superior statistical property in terms of another more popular statistical criterion, namely the mean square error. In this process, we define three biased estimators: the restricted estimator (RE) with a *coefficient of distrust*, the preliminary test estimator (PTE) as a linear combination of the mle and the RE, and the shrinkage estimator (SE) by using the preliminary test approach. We investigate the bias and the mean square error functions, both analytically and graphically to compare the performance of the estimators. The relative efficiency of the estimators are also studied to search for a better choice. Extensive computations have been used to produce graphs and tables to critically check various affects on the properties of the estimators. Table 1 provides the minimum and maximum efficiency of the PTE for different values of the level of significance and varying sample sizes. Comparison of the relative efficiency of the PTE and SE as well as the maximum and minimum relative efficiency of the SE are given in Table 2. The analysis reveals the fact that although there is no uniformly superior estimator that bits the others, the SE dominates the other two biased estimators if the non-sample information regarding the value of $\theta$ is not too far from its true value. In practice, the non-sample information is usually available from past experience or expert knowledge, and hence it is expected that such an information will not be too far from the true value.

The next section deals with the specification of the model and definition of the unrestricted estimators of $\theta$, $\sigma^2$ as well as the derivation of the likelihood ratio test statistic. The three alternative 'improved' estimators are defined in section 3. The expressions of bias and mse functions of the estimators are obtained in section 4. Comparative study of the relative efficiency of the estimators are included in section 5. Some concluding remarks are given in section 6.

# 2 The Model and Some Preliminaries

Let us express the $n$ sample responses in the following convenient form

$$\boldsymbol{Y}_n = \theta \mathbf{1}_n + \mathbf{e} \tag{2.1}$$

where $\boldsymbol{Y}_n = (y_1, \ldots, y_n)'$ is an $n \times 1$ vector of observations, $\mathbf{1}_n = (1, \ldots, 1)'$ – a vector of $n$-tuple of one's, $\theta$ is a scalar unknown parameter (mean) and $\mathbf{e} = (e_1, \ldots, e_n)'$ is a vector of errors with independent components which is distributed as $N_n(\mathbf{0},\ \sigma^2 I_n)$ where

$$E(\mathbf{e}) = \mathbf{0} \quad \text{and} \quad E(\mathbf{e}\mathbf{e}') = \sigma^2 I_n. \tag{2.2}$$

Here, $\sigma^2$ stands for the variance of each of the error component in $\mathbf{e}$ and $I_n$ is the identity matrix of order $n$. From the exclusive sample information, the *unrestricted estimator* (UE) of $\theta$ is the usual maximum likelihood estomator (mle) given by

$$\widetilde{\theta}_n = (\mathbf{1}_n'\mathbf{1}_n)^{-1}\mathbf{1}_n'\boldsymbol{Y}_n = \bar{Y} \tag{2.3}$$

where $\bar{Y}$ is the sample mean. Note that $\mathbf{1}_n'\mathbf{1}_n = n$ and $\mathbf{1}_n'\boldsymbol{Y}_n = \sum_{i=1}^n Y_i$. It is well known that the sampling distribution of the mle of $\theta$ is normal with mean, $E(\widetilde{\theta}_n) = \theta$ and variance, $E(\widetilde{\theta}_n - \theta)^2 = \frac{\sigma^2}{n}$. Therefore, $\widetilde{\theta}_n$ is unbiased for $\theta$, and hence the mse is the same as its variance. Hence, the bias and the mse of $\widetilde{\theta}_n$ are given by

$$B_1(\widetilde{\theta}_n) = 0 \quad \text{and} \quad M_1(\widetilde{\theta}_n) = \frac{\sigma^2}{n} \quad \text{respectively.} \tag{2.4}$$

We compare the above bias and mse functions with those of the three biased estimators, and search for a 'best' option. It is well known that the mle of $\sigma^2$ is

$$\mathcal{S}_n^2 = \frac{1}{n}(\boldsymbol{Y}_n - \widetilde{\theta}_n\mathbf{1}_n)'(\boldsymbol{Y}_n - \widetilde{\theta}_n\mathbf{1}_n). \tag{2.5}$$

This estimator is biased. However, an unbiased estimator of $\sigma^2$ is given by

$$\mathcal{S}_n^2 = \frac{1}{n-1}(\boldsymbol{Y}_n - \widetilde{\theta}_n\mathbf{1}_n)'(\boldsymbol{Y}_n - \widetilde{\theta}_n\mathbf{1}_n). \tag{2.6}$$

The unbiased estimator of $\sigma^2$ has a scaled $\chi^2$ distribution with shape parameter $\nu = (n-1)$. Also, it is well-known that $\widetilde{\theta}_n$ and $\mathcal{S}_n^2$ are independently distributed.

To be able to use the uncertain non-sample prior information in the estimation of the mean, it is essential to remove the element of uncertainty concerning it's value. Fisher suggested to express the uncertain non-sample prior information in the form of a null hypothesis, $H_0 : \theta = \theta_0$ and treat it as a nuisance parameter. He proposed to conduct an appropriate statistical test on the null-hypothesis against the alternative $H_A : \theta \neq \theta_0$ to remove the uncertainty in the non-sample prior information. For the problem under study, an appropriate test is the likelihood ratio test (LRT). The LRT for testing the null-hypothesis is given by the test statistic

$$\mathcal{L}_\nu = \frac{\sqrt{n}(\widetilde{\theta}_n - \theta_0)}{\mathcal{S}_n}. \tag{2.7}$$

The above statistic $\mathcal{L}_\nu$, under $H_A$, follows a non-central Student-$t$ distribution with $\nu = (n-1)$ degrees of freedom (d.f.), with the non-centrality parameter $\frac{1}{2}\Delta^2$, where

$$\Delta^2 = \frac{n(\theta - \theta_0)^2}{\sigma^2}. \tag{2.8}$$

Equivalently, we may say that $\mathcal{L}_\nu^2$, under $H_A$, follows a non-central $F$-distribution with $(1,\ \nu)$ degrees of freedom having the same non-centrality parameter $\frac{1}{2}\Delta^2$. Under the null-hypothesis $\mathcal{L}_\nu$ and $\mathcal{L}_\nu^2$ follow a central Student-$t$ distribution and an $F$-distribution respectively with the same degrees of freedom. This test statistic was used by T.A. Bancroft (1944) to define the PTE, and we use the same statistic to define the shrinkage estimator by following the preliminary test approach to the shrinkage estimation.

# 3 Alternative Estimators

As part of incorporating the uncertain non-sample prior information into the estimation process, first we combine the exclusive sample based estimator, $\widetilde{\theta}_n$ with the non-sample prior information presented in the form of a null hypothesis, $H_0 : \theta = \theta_0$ in some reasonable way. First, consider a simple linear combination of $\theta_0$ and $\widetilde{\theta}_n$ as

$$\hat{\theta}_n(d) = d\widetilde{\theta}_n + (1-d)\theta_0, \quad 0 \le d \le 1. \tag{3.1}$$

This estimator of $\theta$ is called the *restricted estimator* (RE), where $d$ is the *degree of distrust* in the null hypothesis, $H_0 : \theta = \theta_0$. Now, $d = 0$, means there is *no distrust* in the $H_0$ and we get $\hat{\theta}_n(d=0) = \theta_0$, while $d = 1$ means there is *complete distrust* in the $H_0$ and we get $\hat{\theta}_n(d=1) = \widetilde{\theta}_n$. If $0 < d < 1$, the degree of distrust is an intermediate value which results in an interpolated value between $\theta_0$ and $\widetilde{\theta}_n$ given by (3.1). The restricted estimator, as defined above, is normally distributed with mean and variance given by

$$E(\hat{\theta}_n(d)) = d\theta + (1-d)\theta_0 \quad \text{and} \quad \mathrm{Var}(\hat{\theta}_n(d)) = \frac{d^2\sigma^2}{n} \quad \text{respectively.} \tag{3.2}$$

Following Bancroft (1944) we define a preliminary test estimator (PTE) of the mean as

$$\begin{aligned} \hat{\theta}_n^{\mathrm{PTE}}(d) &= \hat{\theta}_n I(|t_\nu| < t_{\alpha/2}) + \widetilde{\theta}_n I(|t_\nu| \ge t_{\alpha/2}) \\ &= \widetilde{\theta}_n - (1-d)(\widetilde{\theta}_n - \theta_0) I(|t_\nu| < t_{\alpha/2}) \end{aligned} \tag{3.3}$$

where $I(A)$ is the indicator function of the set $A$ and $t_{\alpha/2}$ is the critical value chosen for the two-sided $\alpha$-level test based on the Student-t distribution with $\nu = (n-1)$ degrees of freedom. A simple form of the above preliminary test estimator is

$$\hat{\theta}_n^{\mathrm{PTE}} = \theta_0 I(|t_\nu| < t_{\alpha/2}) + \widetilde{\theta}_n I(|t_\nu| \geq t_{\alpha/2}), \tag{3.4}$$

which is a special case of (3.3) when $d = 0$. Note that, the $\hat{\theta}_n^{\mathrm{PTE}}(d)$ is a convex combination of $\hat{\theta}_n(d)$ and $\widetilde{\theta}_n$, and $\hat{\theta}_n^{\mathrm{PTE}}(d=0)$ is a convex combination of $\theta_0$ and $\widetilde{\theta}_n$. We may rewrite (3.3) as

$$\hat{\theta}_n^{\mathrm{PTE}}(d) = \widetilde{\theta}_n - (1-d)(\widetilde{\theta}_n - \theta_0) I(F < F_\alpha) \tag{3.5}$$

where $F_\alpha$ is the $(1-\alpha)^{th}$ quantile of a central $F$-distribution with $(1,\ \nu)$ degrees of freedom. For $d = 0$, we get (3.5) as

$$\hat{\theta}_n^{\mathrm{PTE}}(d=0) = \widetilde{\theta}_n - (\widetilde{\theta}_n - \theta_0) I(F < F_\alpha). \tag{3.6}$$

The PTE is an extreme choice between $\hat{\theta}_n(d)$ and $\widetilde{\theta}_n$. Hence it does not allow any smooth transition between the two extreme values. Also, it depends on the preselected level of significance of the test. To overcome these problems, we consider the shrinkage estimator (SE) of $\theta$ defined as follows:

$$\hat{\theta}_n^s = \theta_0 + \left\{1 - \frac{cS_n}{|\sqrt{n}(\widetilde{\theta}_n - \theta_0)|}\right\} \big(\widetilde{\theta}_n - \theta_0\big). \tag{3.7}$$

Note that in this estimator $c$ is a constant function of $\nu$. Now, if $|t_\nu| = \frac{|\sqrt{n}|\widetilde{\theta}_n - \theta_0)|}{S_n}$ is large, $\hat{\theta}_n^s$ tends towards $\widetilde{\theta}_n$, while for small $|t_\nu|$ equaling $c$, $\hat{\theta}_n^s$ tends towards $\theta_0$ similar to the preliminary test estimator. The shrinkage estimator does not depend on the level of significance, unlike the preliminary test estimator.

# 4 Some Statistical Properties

In this section, we derive the bias and the mean square error (mse) functions of the SE. Also, we discuss some of the important features of these functions.

First the bias and the mse of the RE, $\hat{\theta}_n(d)$ are found to be

$$B_2(\hat{\theta}_n(d)) = -(1-d)\Delta, \quad \Delta = \frac{\sqrt{n}(\theta - \theta_0)}{\sigma} \tag{4.1}$$

$$M_2(\hat{\theta}_n(d)) = \frac{\sigma^2}{n}[d^2 + (1-d)^2\Delta^2] \tag{4.2}$$

where $\Delta^2$ is the *departure constant* from the null-hypothesis. The value of this constant is 0 when the null hypothesis is true; otherwise it is always positive. The statistical properties of the three estimators depend on the value of the above departure constant. The performance of the estimators change with the change in the value of $\Delta$. We investigate this feature in a greater detail in the forthcoming sections.

## 4.1 The Bias and the MSE of the PTE

From the definition, the expression of bias of the PTE is

$$\begin{aligned} E(\hat{\theta}_n^{\text{PTE}}(d) - \theta) &= E(\widetilde{\theta}_n - \theta) - (1-d)E\{(\widetilde{\theta}_n - \theta_0)I(F < F_\alpha)\} \\ &= -(1-d)\frac{\sigma}{\sqrt{n}}E\Big\{\frac{\sqrt{n}(\widetilde{\theta}_n - \theta_0)}{\sigma}I\Big(\frac{n(\widetilde{\theta}_n - \theta_0)^2}{\mathcal{S}_n^2} < F_\alpha\Big)\Big\}. \end{aligned} \tag{4.3}$$

Note $Z = \sqrt{n}(\widetilde{\theta}_n - \theta_0)/\sigma$ is distributed as $N(\Delta, 1)$, where $\Delta = \frac{\sqrt{n}}{\sigma}(\theta - \theta_0)$, and $(n-1)\mathcal{S}_n^2/\sigma^2$ is distributed (independently) as a central chi-square variable with $\nu = (n-1)$ degrees of freedom.

Evaluating the expression in (4.3) the bias function of $\hat{\theta}_n^{\text{PTE}}(d)$ is found to be

$$\begin{aligned} B_3\Big(\hat{\theta}_n^{\text{PTE}}(d)\Big) &= -(1-d)\frac{\sigma}{\sqrt{n}}\Delta G_{3,\nu}\Big(\frac{1}{3}F_\alpha; \Delta^2\Big) \\ &= -(1-d)(\theta - \theta_0)G_{3,\nu}\Big(\frac{1}{3}F_\alpha; \Delta^2\Big), \end{aligned} \tag{4.4}$$

where $G_{m,n}(\cdot; \Delta^2)$ is the c.d.f. of a non-central F-distribution with $(m, n)$ degrees of freedom and non-centrality parameter $\Delta^2$. The above c.d.f. involves incomplete beta function ratio with appropriate arguments. This bias function of the PTE depends on the *coefficient of distrust* and the *departure constant*, among other things. To evaluate the expression in (4.3) we used the following theorem.

**Theorem 4.1**. *If $\mathcal{Z} \sim \mathcal{N}(\Delta, 1)$ and $\phi(\mathcal{Z}^2)$ is a Borel measurable function, then*

$$E\{\mathcal{Z}\phi(\mathcal{Z}^2)\} = \Delta E\phi(\chi_3^2(\Delta^2)). \tag{4.5}$$

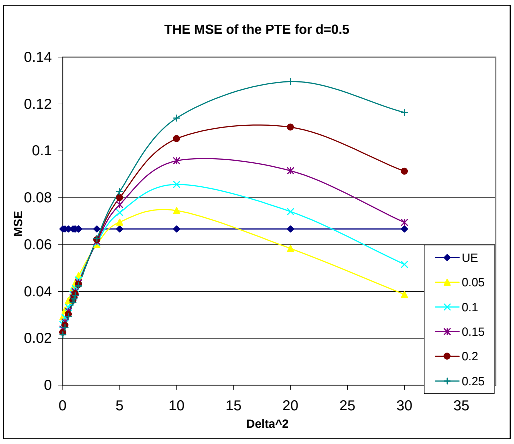
Furthermore, to obtain the mean square error of $\hat{\theta}_n^{\text{PTE}}(d)$ we need the following theorem.

**Theorem 4.2**. *If $Z \sim \mathcal{N}(\Delta, 1)$ and $\phi(Z^2)$ is a Borel measurable function, then*

$$E\{Z^2\phi(Z^2)\} = E\left[\phi(\chi_3^2(\Delta^2)\right] + \Delta^2 E\left[\phi\left(\chi_5^2(\Delta^2)\right)\right]. \tag{4.6}$$

The proof of the above two theorems are given in Appendix B2 of Judge and Bock (1978).

h



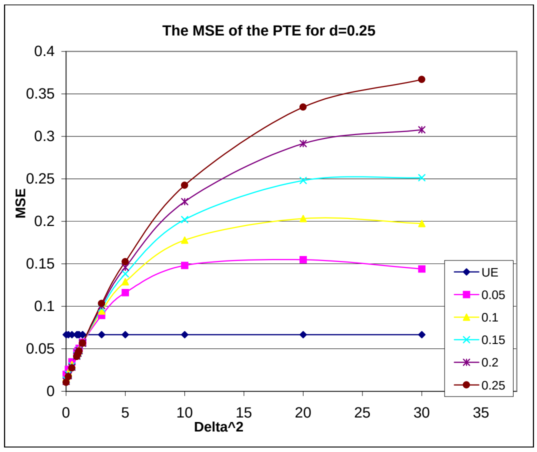




Figure 1: Graph of $M_3\left(\hat{\theta}_n^{\text{PTE}}(d)\right)$ as a function of $\Delta^2$

From the definition, the mse expression of the PTE is

$$\begin{aligned} M_3(\hat{\theta}_n^{\text{PTE}}(d)) &= E(\hat{\theta}_n^{\text{PTE}}(d)-\theta)^2 \\ &= E(\tilde{\theta}_n-\theta)^2+(1-d)^2E(\tilde{\theta}_n-\theta_0)^2I(F<F_\alpha) \\ &\qquad -2(1-d)E(\tilde{\theta}_n-\theta)(\tilde{\theta}_n-\theta_0)I(F<F_\alpha) \\ &= \frac{\sigma^2}{n}+(1-d)^2E\Big\{(\tilde{\theta}_n-\theta_0)^2I(F<F_\alpha)\Big\} \\ &\qquad -2(1-d)E\Big\{[(\tilde{\theta}-\theta_0)-(\theta-\theta_0)](\tilde{\theta}-\theta_0)I(F<F_\alpha)\Big\}. \end{aligned} \tag{4.7}$$

After completing the evaluation of all the terms on the R.H.S. of the above expression in (4.7), the mse function of the PTE becomes,

$$\begin{aligned} M_3(\hat{\theta}_n^{\text{PTE}}(d)) = \frac{\sigma^2}{n}\Big[1-(1-d^2)G_{3,\nu}\Big(\frac{1}{3}F_\alpha;\Delta^2\Big) \\ +(1-d)\Delta^2\Big\{2G_{3,\nu}\Big(\frac{1}{3}F_\alpha;\Delta^2\Big)-(1+d)G_{5,\nu}\Big(\frac{1}{5}F_\alpha;\Delta^2\Big)\Big\}\Big]. \end{aligned} \tag{4.8}$$

Figure 1, displays the behavior of the mse function of the PTE for different values of $\alpha$ with the change in the value of $\Delta^2$. The two graphs illustrate the different features for two values of the *coefficient of distrust*, $d = 0.25$ and $d = 0.50$.

**Some Properties of MSE of PTE**

(a) Under the null hypothesis $\Delta^2 = 0$, and hence the mse of $\hat{\theta}_n^{\text{PTE}}(d)$ equals

$$\frac{\sigma^2}{n}\left[1-(1-d^2)G_{3,\nu}\left(\frac{1}{3}F_\alpha;0\right)\right] < \frac{\sigma^2}{n}. \tag{4.9}$$

Thus, at $\Delta^2 = 0$ PTE of $\theta$ performs better than $\widetilde{\theta}_n$, the UE. As $\alpha \to 0$, $G_{3,\nu}\Big(\frac{1}{3}F_\alpha; 0\Big) \to 1$, then

$$\frac{\sigma^2}{n}\left[1 - (1-d^2)G_{3,\nu}\left(\frac{1}{3}F_\alpha; 0\right)\right] \to \frac{d^2\sigma^2}{n}, \tag{4.10}$$

which is the mse of $\hat{\theta}_n(d)$. On the other hand, if $F_\alpha \to 0$, $G_{3,\nu}\Big(\frac{1}{3}F_\alpha; 0\Big) \to 0$, then

$$\frac{\sigma^2}{n}\left[1 - (1-d^2)G_{3,\nu}\left(\frac{1}{3}F_\alpha; 0\right)\right] \to \frac{\sigma^2}{n}, \tag{4.11}$$

which is the mse of $\widetilde{\theta}_n$.

(b) As $\Delta^2 \to \infty$, $G_{m,\nu}\Big(\frac{1}{m}F_\alpha; \Delta^2\Big) \to 0$, this means the expression at (4.8) tends towards $\frac{\sigma^2}{n}$, the mse of the UE.

(c) Since $G_{3,\nu}\Big(\frac{1}{3}F_\alpha; \Delta^2\Big)$ is always greater than $G_{5,\nu}\Big(\frac{1}{5}F_\alpha; \Delta^2\Big)$ for any value of $\alpha$, replacing $G_{5,\nu}\Big(\frac{1}{5}F_\alpha; \Delta^2\Big)$ by $G_{3,\nu}\Big(\frac{1}{3}F_\alpha; \Delta^2\Big)$, (4.8) becomes

$$\begin{aligned} &\geq \frac{\sigma^2}{n}\Big[1 + (1-d^2)G_{3,\nu}\Big(\frac{1}{3}F_\alpha; \Delta^2\Big)\{(1-d)\Delta^2 - (1+d)\}\Big] \qquad (4.12) \\ &\geq \frac{\sigma^2}{n} \quad \text{whenever} \quad \Delta^2 > \frac{1+d}{1-d}. \end{aligned}$$

On the other hand, (4.8) may be rewritten as

$$\frac{\sigma^2}{n}\Big[1 + (1-d)G_{3,\nu}\Big(\frac{1}{3}F_\alpha; \Delta^2\Big)\{2\Delta^2 - (1+d)\} - (1-d^2)G_{5,\nu}\Big(\frac{1}{5}F_\alpha; \Delta^2\Big)\Big] \tag{4.13}$$

$$\leq \frac{\sigma^2}{n} \quad \text{whenever} \quad \Delta^2 < \frac{1+d}{2}. \tag{4.14}$$

This means that the mse of $\hat{\theta}_n^{\text{PTE}}(d)$ as a function of $\Delta^2$ crosses the constant line $M_1(\widetilde{\theta}_n) = \frac{\sigma^2}{n}$ in the interval

$$\left(\frac{1+d}{2}, \frac{1+d}{1-d}\right). \tag{4.15}$$

(d) A general picture of the mse graph may be described as follows: The mse-function begins with the smallest value $\frac{\sigma^2}{n}\Big[1 - (1-d^2)G_{3,\nu}\Big(\frac{1}{3}F_\alpha; 0\Big)\Big]$ at $\Delta^2 = 0$. As $\Delta^2$ grows, the function increases monotonically crossing the constant line $\frac{\sigma^2}{n}$ in the interval $\Big(\frac{1+d}{2}, \frac{1+d}{1-d}\Big)$ and reaches a maximum in the interval $\Big(\frac{1+d}{1-d}, \infty\Big)$ then monotonically decreases towards $\frac{\sigma^2}{n}$ as $\Delta^2 \to \infty$.

### 4.1.1 Determination of optimum $\alpha$ for the PTE

Clearly the (mse and hence the) relative efficiency of the preliminary test estimator compared with the unrestricted estimator depends on the level of significance $\alpha$ of the test of null-hypothesis and the departure parameter $\Delta^2$.

**Table 1: Minimum and Maximum Efficiency of PTE**

| $\alpha/n$ | | 10 | 15 | 20 | 25 | 30 | 35 | 40 |
|---|---|---|---|---|---|---|---|---|
| .05 | $E^*$ | 4.2577 | 4.0063 | 3.8912 | 3.8252 | 3.7825 | 3.7525 | 3.7304 |
| | Eo | 0.3350 | 0.3600 | 0.3720 | 0.3790 | 0.3836 | 0.3868 | 0.3893 |
| | $\Delta_0$ | 6.1009 | 5.5303 | 5.2857 | 5.1514 | 5.0656 | 5.0063 | 4.9629 |
| .10 | E* | 2.5564 | 2.4529 | 2.4052 | 2.3778 | 2.3600 | 2.3475 | 2.3383 |
| | Eo | 0.4500 | 0.4722 | 0.4828 | 0.4889 | 0.4929 | 0.4958 | 0.4979 |
| | $\Delta_0$ | 4.8042 | 4.4762 | 4.3337 | 4.2535 | 4.2022 | 4.1657 | 4.1397 |
| .15 | E* | 1.9523 | 1.8939 | 1.8669 | 1.8513 | 1.8412 | 1.8341 | 1.8288 |
| | Eo | 0.5405 | 0.5601 | 0.5693 | 0.5747 | 0.5782 | 0.5807 | 0.5825 |
| | $\Delta_0$ | 4.1626 | 3.9403 | 3.8429 | 3.7875 | 3.7520 | 3.7272 | 3.7090 |
| .20 | E* | 1.6406 | 1.6033 | 1.5860 | 1.5760 | 1.5695 | 1.5649 | 1.5616 |
| | Eo | 0.6174 | 0.6345 | 0.6425 | 0.6471 | 0.6502 | 0.6523 | 0.6539 |
| | $\Delta_0$ | 3.7616 | 3.6006 | 3.5278 | 3.4873 | 3.4612 | 3.4430 | 3.4295 |
| .25 | E* | 1.4514 | 1.4259 | 1.4141 | 1.4073 | 1.4029 | 1.3998 | 1.3975 |
| | Eo | 0.6844 | 0.6990 | 0.7059 | 0.7098 | 0.7124 | 0.7143 | 0.7156 |
| | $\Delta_0$ | 3.4811 | 3.3612 | 3.3066 | 3.2757 | 3.2557 | 3.2417 | 3.2314 |
| .30 | E* | 1.3256 | 1.3077 | 1.2994 | 1.2946 | 1.2914 | 1.2892 | 1.2876 |
| | Eo | 0.7430 | 0.7553 | 0.7611 | 0.7645 | 0.7666 | 0.7682 | 0.7693 |
| | $\Delta_0$ | 3.2756 | 3.1824 | 3.1402 | 3.1150 | 3.0997 | 3.0890 | 3.0811 |
| .35 | E* | 1.2374 | 1.2245 | 1.2185 | 1.2150 | 1.2128 | 1.2112 | 1.2100 |
| | Eo | 0.7941 | 0.8043 | 0.8090 | 0.8117 | 0.8135 | 0.8148 | 0.8157 |
| | $\Delta_0$ | 3.1163 | 3.0429 | 3.0110 | 2.9922 | 2.9801 | 2.9716 | 2.9653 |
| .40 | E* | 1.1732 | 1.1639 | 1.1596 | 1.1571 | 1.1555 | 1.1543 | 1.1535 |
| | Eo | 0.8381 | 0.8463 | 0.8501 | 0.8523 | 0.8538 | 0.8548 | 0.8555 |
| | $\Delta_0$ | 2.9903 | 2.9335 | 2.9077 | 2.8930 | 2.8835 | 2.8768 | 2.8719 |
| .45 | E* | 1.1256 | 1.1189 | 1.1158 | 1.1140 | 1.1128 | 1.1120 | 1.1114 |
| | Eo | 0.8756 | 0.8820 | 0.8850 | 0.8867 | 0.8878 | 0.8887 | 0.8892 |
| | $\Delta_0$ | 2.8885 | 2.8442 | 2.8240 | 2.8125 | 2.8050 | 2.7979 | 2.7941 |
| .50 | E* | 1.0899 | 1.0851 | 1.0829 | 1.0816 | 1.0808 | 1.0802 | 1.0797 |
| | Eo | 0.9068 | 0.9118 | 0.9140 | 0.9153 | 0.9162 | 0.9168 | 0.9172 |
| | $\Delta_0$ | 2.8054 | 2.7695 | 2.7541 | 2.7461 | 2.7403 | 2.7362 | 2.7332 |

Legends: $\alpha$ is the level of significance
$n$ is the sample size
E* is the maximum efficiency
Eo is the minimum efficiency
$\Delta_0$ is the value of Delta at which minimum efficiency occur.

Let the relative efficiency of the PTE with respect to the UE be denoted by $E(\alpha;\Delta^2)$ which is given by

$$E(\alpha;\Delta^2) = [1 + g(\Delta^2)]^{-1}, \tag{4.16}$$

where

$$g(\Delta^2) = 1 + (1-d)\Delta^2\Big\{2G_{3,\nu}\Big(\frac{1}{3}F_\alpha;\Delta^2\Big) - (1+d)G_{5,\nu}\big(\frac{1}{5}F_\alpha;\Delta^2\big)\Big\} \\ -(1-d^2)G_{3,\nu}\Big(\frac{1}{3}F_\alpha;\Delta^2\Big). \tag{4.17}$$

The efficiency function attains its maximum at $\Delta^2 = 0$ for all $\alpha$ given by

$$E(\alpha;0) = \Big[1-(1-d^2)G_{3,\nu}\Big(\frac{1}{3}F_\alpha;0\Big)\Big]^{-1} \geq 1. \tag{4.18}$$

As $\Delta^2$ departs from the origin, $E(\alpha;\Delta^2)$ decreases monotonically crossing the line $E(\alpha;\Delta^2) = 1$ in the interval $\left(\frac{1+d}{2}, \frac{1+d}{1-d}\right)$, to a minimum at $\Delta^2 = \Delta^2_{\text{min}}$, then from that point on increases monotonically towards 1 as $\Delta^2 \to \infty$ from below. Now, for $\Delta^2 = 0$ and level of significance varying, we have

$$\max_\alpha E(\alpha,0) = E(0,0) = d^{-2}. \tag{4.19}$$

As a function of $\alpha$, $E(\alpha;0)$ decreases as $\alpha$ increases. On the other hand, $E(\alpha;\Delta^2)$ as a function of $\Delta^2$ is decreasing, and the curves $E(0;\Delta^2)$ and $E(1/2;\Delta^2) = 1$ intersect at $\Delta^2 = \frac{1+d}{1-d}$. In general $E(\alpha;\Delta^2)$ and $E(\alpha;\Delta^2)$ always intersect in the interval $\left(0, \frac{1+d}{1-d}\right)$. The value of $\Delta^2$ at the intersection decreases as $\alpha$ increases. Therefore, for two different levels of significance say, $\alpha_1$ and $\alpha_2$, $E(\alpha_1;\Delta^2)$ and $E(\alpha_2;\Delta^2)$ intersects below 1. In order to choose an optimum level of significance with maximum relative efficiency we adopt the following rule: If it is known that $0 \leq \Delta \leq \frac{1+d}{1-d}$, $\hat{\theta}_n$ is always chosen since $E(0,\Delta^2)$ is maximum for all $\Delta^2$ in this interval. Generally, $\Delta^2$ is unknown. In this case there is no way of choosing a uniformly best estimator of $\theta$. Thus, we pre-assign a tolerable relative efficiency, say, $E_0$. Then, consider the set

$$A_\alpha = \Big\{\alpha | E(\alpha;\Delta^2) \geq E_0\Big\}. \tag{4.20}$$

An estimator $\hat{\theta}_n^{\text{PTE}}(d)$ is chosen which maximizes $E(\alpha;\Delta^2)$ over all $\alpha \in A_\alpha$ and $\Delta^2$. Thus, we solve the following equation for $\alpha$

$$\max_\alpha \min_{\Delta^2} E(\alpha;\Delta^2) = E_0. \tag{4.21}$$

The solution $\alpha^*$ provides a *maximin rule* for the optimum level of significance of the preliminary test. For practitioners, Table 1 provides the minimum and maximum relative efficiencies of the PTE and the values of $\Delta$ at which the minimum relative efficiency occur for selected sample sizes and varying values of $\alpha$ when $d = 0$. If the value of $\Delta$ is in the interval (0, 1) and known then the restricted estimtor (RE) is the best. However, the value of $\Delta$ is generally not known, in such a case we pre-assign a minimum tolerable relative efficiency (say, $E_0^{min}$) of the PTE and look for the appropriate level of significance (say $\alpha^*$) for the given sample size (say, $n_0$) from Table 1. As an example of selecting an optimal level of significance, if one wishes to have a guaranteed minimum relative efficiency of $E_0^{min} = 0.80$ of the PTE with a sample size of $n_0 = 20$ s/he has to select a level of significance, $\alpha_0 = 0.35$.

## 4.2 The Bias and MSE of the SE

Now, following Balforine and Zacks (1992) we compute the bias and the mse of the SE, $\hat{\theta}_n^s$. The bias of the SE is given by

$$\begin{aligned} B_4(\hat{\theta}_n^s) &= E(\hat{\theta}_n^s - \theta) = -cE\frac{S_n(\widetilde{\theta} - \theta_0)}{|\sqrt{n}(\widetilde{\theta}_n - \theta_0)|} \\ &= -\frac{c}{\sqrt{n}}E(S_n)E\left\{\frac{Z}{|Z|}\right\} \end{aligned} \tag{4.22}$$

where $Z = \frac{\sqrt{n}(\widetilde{\theta}_n - \theta_0)}{\sigma} \sim \mathcal{N}(\Delta, 1)$. Now, we use the following theorem to evaluate $E\left\{\frac{Z}{|Z|}\right\}$.

**Theorem 4.3**. *If $Z \sim \mathcal{N}(\Delta, 1)$ and $\phi(Z^2)$ is a Borel measurable function, then*

$$E\left\{\frac{Z}{|Z|}\right\} = 1 - 2\Phi(-\Delta). \tag{4.23}$$

*where $\Phi(\cdot)$ is the c.d.f. of the standard normal distribution.* The proof of the theorem is straightforward.

From the expression of the above bias function, the quadratic bias of the SE, $QB_4(\hat{\theta}_n^s)$ is obtained as

$$B_4(\hat{\theta}_n^s) = c^2K_n^2\{1 - 2\Phi(-\Delta)\}^2 = c^2K_n^2\{2\Phi(\Delta) - 1\}^2 \tag{4.24}$$

where $K_n = \sqrt{\frac{2}{n-1}}\frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})}$.

As $\Delta^2 \to 0$, $QB_4(\hat{\theta}_n^s) \to 0$ and as $\Delta^2 \to \infty$, $QB_4(\hat{\theta}_n^s) \to K_n^2c^2$. Therefore, $QB_4(\hat{\theta}_n^s)$ is a non-decreasing monotonic function of $\Delta^2$. Thus, unless $\Delta^2$ is near the origin, the quadratic bias of the SE is significantly large.

In order to compute the mse of $\hat{\theta}_n^s$ we consider

$$E(\hat{\theta}_n^s - \theta)^2 = E(\widetilde{\theta}_n - \theta)^2 + c^2E(S_n^2)E\left\{\frac{(\widetilde{\theta} - \theta_0)^2}{[\sqrt{n}(\widetilde{\theta}_n - \theta_0)]^2}\right\} \tag{4.25}$$

$$-2cE\Big\{\frac{(\widetilde{\theta}-\theta)(\widetilde{\theta}-\theta_0)}{|\sqrt{n}(\widetilde{\theta}_n-\theta_0)|}\Big\}E(S_n)$$
$$= \frac{\sigma^2}{n}+\frac{c^2\sigma^2}{n}-2c\frac{\sigma^2 K_n}{n}\Big\{E(|Z|)-\Delta E\Big(\frac{Z}{|Z|}\Big)\Big\}.$$

where $Z \sim \mathcal{N}(\Delta, 1)$. To find $E(|Z|)$, we have the following theorem.

**Theorem 4.4**. *If $Z \sim \mathcal{N}(\Delta, 1)$, then*

$$E(|Z|) = \sqrt{\frac{2}{\pi}}e^{-\Delta^2/2} + \Delta\{2\Phi(\Delta) - 1\} \tag{4.26}$$

*where $\Phi(\cdot)$ is the c.d.f. of the standard normal variable.*

**Proof**: The p.d.f. of $|Z|$ is

$$f_{|Z|}(z) = \phi(z-\Delta) + \phi(z+\Delta) \tag{4.27}$$

and hence we have,

$$\begin{aligned} E(|Z|) &= \int_0^\infty z\phi(z-\Delta)dz + \int_0^\infty z\phi(z+\Delta)dz \qquad (4.28)\\ &= \int_\Delta^\infty z\phi(z)dz + \int_{-\Delta}^\infty z\phi(z)dz + \Delta\Big\{\int_{-\Delta}^\infty \phi(z)dz - \int_\Delta^\infty \phi(z)dz\Big\}\\ &= \int_\Delta^\infty z\phi(z)dz + \int_{-\Delta}^\infty z\phi(z)dz + \Delta\{2\Phi(\Delta)-1\}\\ &= \sqrt{\frac{2}{\pi}}e^{-\Delta^2/2} + \Delta\{2\Phi(\Delta)-1\}. \end{aligned}$$

Therefore, the mse of $\hat{\theta}_n^s$ is given by

$$M_4(\hat{\theta}_n^s) = \frac{\sigma^2}{n}\Big\{1 + c^2 - 2cK_n\sqrt{\frac{2}{\pi}}e^{-\Delta^2/2}\Big\}. \tag{4.29}$$

The value of $c$ which minimizes (4.29) depends on $\Delta^2$ and is given by

$$c^* = \sqrt{\frac{2}{\pi}}K_n e^{-\Delta^2/2}. \tag{4.30}$$

To make $c^*$ independent of $\Delta^2$, we choose $c^0 = \sqrt{\frac{2}{\pi}}K_n$. Thus, optimum $M_4(\hat{\theta}_n^s)$ reduces to

$$M_4(\hat{\theta}_n^s) = \frac{\sigma^2}{n}\Big\{1 - \frac{2}{\pi}K_n^2\Big[2e^{-\Delta^2/2} - 1\Big]\Big\}. \tag{4.31}$$

We compare the above mse with that of the other estimators in the next section.

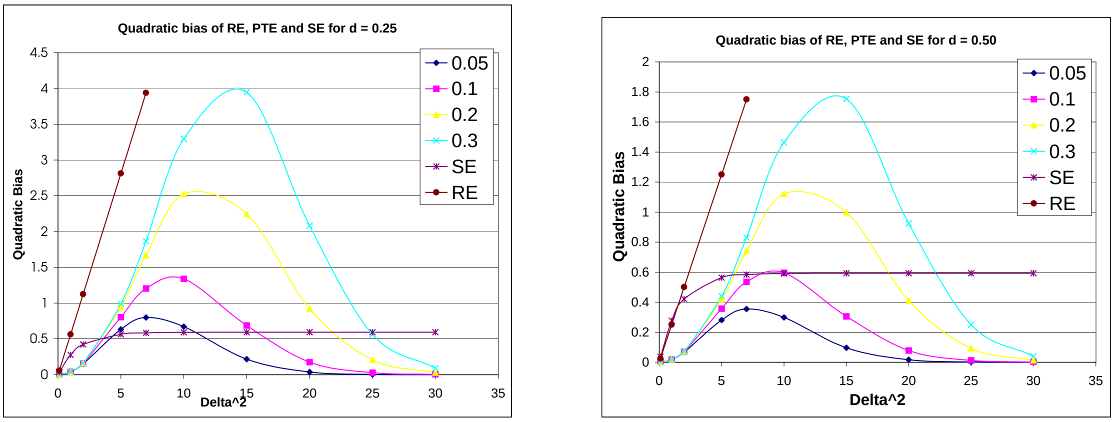
h 


Figure 2: Graph of the quadratic bias of the PTE and SE as a function of $\Delta^2$

# 5 Comparative Study

In this section we compare the bias of the three estimators. Also, we define the relative efficiency functions of the estimators, and analyze these functions to compare the relative performances of the estimators.

## 5.1 Comparing Quadratic Bias Functions

First, we note that the quadratic bias of the RE, PTE and SE are given by

$$\begin{aligned} QB_2(\hat{\theta}_n(d)) &= (1-d)^2\Delta^2 \qquad (5.1)\\ QB_3\Big(\hat{\theta}_n^{\rm PTE}(d)\Big) &= (1-d)^2\Delta^2\Big\{G_{3,\nu}\Big(\frac{1}{3}F_\alpha;\Delta^2\Big)\Big\}\\ QB_4(\hat{\theta}_n^s) &= c^2K_n^2\{2\Phi(\Delta)-1\}^2. \end{aligned}$$

The graph of $QB_2(\hat{\theta}_n)$, $QB_3\Big(\hat{\theta}_n^{\rm PTE}(d)\Big)$ and $QB_4(\hat{\theta}_n^s)$ are given in Figure 2. It is clear that

$$QB_1(\widetilde{\theta}_n) \le QB_3\Big(\hat{\theta}_n^{\rm PTE}(d)\Big) \le QB_2(\hat{\theta}_n(d)). \qquad (5.2)$$

Thus, $\hat{\theta}_n^{\rm PTE}(d)$ has smaller quadratic bias than $\hat{\theta}_n(d)$. Hence, under the null-hypothesis $B_1(\widetilde{\theta}_n) = B_2(\hat{\theta}_n(d)) = B_3\Big(\hat{\theta}_n^{\rm PTE}(d)\Big) = 0$. In selecting estimators with smallest quadratic bias we choose $\hat{\theta}_n^{\rm PTE}(d)$ over $\hat{\theta}_n(d)$ but $\widetilde{\theta}_n$ is the best.

The quadratic bias of the SE is higher than that of the PTE for all $\alpha$ when $\Delta$ is near 0. But, starting from a moderate value of $\Delta$ the quadratic bias of the

SE becomes constant and lower than that of the PTE. However, as $\Delta \to \infty$, the situation reverses again.

## 5.2 The Relative Efficiency

First we define the relative efficiency functions of the biased estimators as the ratio of the reciprocal of the mse functions. Then we compare the relative performance of the estimators by using the relative efficiency criterion.

**Comparing RE against UE**

The relative efficiency of $\hat{\theta}_n(d)$ compared to $\widetilde{\theta}_n$ is denoted by $\mathrm{RE}(\hat{\theta}_n(d) : \widetilde{\theta}_n)$ and is obtained as

$$\mathrm{RE}(\hat{\theta}_n(d) : \widetilde{\theta}_n) = [d^2 + (1-d)^2\Delta^2]^{-1}. \tag{5.3}$$

We observe the following based on (5.3).

(i) If the non-sampling information is correct, i.e., $\Delta^2 = 0$, the $\mathrm{RE}(\hat{\theta}_n(d) : \widetilde{\theta}_n) = d^{-2} > 1$ and $\hat{\theta}_n(d)$ is more efficient than $\widetilde{\theta}_n$. Thus, under the null hypothesis the biased estimator, RE performs better than the unbiased estimator, UE.

(ii) If the non-sampling information is incorrect, i.e., $\Delta^2 > 0$ we study the expression in (5.3) as a function of $\Delta^2$ for a fixed $d$-value. As a function of $\Delta^2$, (5.3) is a decreasing function with its maximum value $d^{-2}(> 1)$ at $\Delta^2 = 0$ and minimum value 0 at $\Delta^2 = +\infty$. It equals 1 at $\Delta^2 = \frac{1+d}{1-d}$. Thus, if $\Delta^2 \in [0, \frac{1+d}{1-d})$, $\hat{\theta}_n$ is more efficient than $\widetilde{\theta}_n$, and outside this interval $\widetilde{\theta}_n$ is efficient. For example, if $d = \frac{1}{2}$, the interval in which $\hat{\theta}_n(d)$ is more efficient than $\widetilde{\theta}_n$ is $[0, 3)$, while $\widetilde{\theta}_n$ is more efficient in $[3, \infty)$ than $\hat{\theta}_n(d)$. The maximum efficiency of $\hat{\theta}_n(d)$ over $\widetilde{\theta}_n$ is 4.

**Comparing PTE against UE**

Now, we consider the relative efficiency of the PTE compared to the UE. It is given by

$$\begin{aligned} RE\Big(\hat{\theta}_n^{\mathrm{PTE}}(d) : \widetilde{\theta}_n\Big) &= \Big[1 - (1-d^2)G_{3,\nu}\Big(\frac{1}{3}F_\alpha;\Delta^2\Big) + (1-d)\Delta^2 \\ &\quad \times\Big\{2G_{3,\nu}\Big(\frac{1}{3}F_\alpha;\Delta^2\Big) - (1+d)G_{5,\nu}\Big(\frac{1}{5}F_\alpha;\Delta^2\Big)\Big\}\Big]^{-1} \end{aligned} \tag{5.4}$$

for any fixed $d$ $(0 \le d \le 1)$ and at a fixed level of significance $\alpha$. As $F_\alpha \to \infty$, $RE\Big(\hat{\theta}_n^{\mathrm{PTE}}(d) : \widetilde{\theta}_n\Big) \to [1 - (1-d^2) + (1-d)^2\Delta^2]^{-1} = [d^2 + (1-d)^2\Delta^2]^{-1}$ which is the relative efficiency of $\hat{\theta}_n(d)$ compared to $\widetilde{\theta}_n$. On the other hand, as $F_\alpha \to 0$, $RE\Big(\hat{\theta}_n^{\mathrm{PTE}}(d) : \widetilde{\theta}_n\Big) \to 1$. This means the relative efficiency of the PTE is the same as the unrestricted estimator, $\widetilde{\theta}_n$. Note that under the null hypothesis, $\Delta^2 = 0$, and the relative efficiency espression (5.4) equals

$$\Big[1 - (1-d^2)G_{3,\nu}\Big(\frac{1}{3}F_\alpha; 0\Big)\Big]^{-1} \geq 1, \tag{5.5}$$

which is the maximum value of the relative efficiency. Thus the relative efficiency function monotonically decreases crossing the 1-line for $\Delta^2$-value between $\frac{1+d}{2}$ and $\frac{1+d}{1-d}$, to a minimum for some $\Delta^2 = \Delta^2_{\text{min}}$ and then monotonically increases, to approach the unit value from below. The relative efficiency of the preliminary test estimator equals unity whenever

$$\Delta^2_* = \frac{(1+d)}{\Big\{2 - (1+d)\frac{G_{5,\nu}(\frac{1}{5}F_\alpha;\Delta^2)}{G_{3,\nu}(\frac{1}{3}F_\alpha;\Delta^2)}\Big\}}, \tag{5.6}$$

where $\Delta^2_*$ lies in the interval $\Big(\frac{1+d}{2}, \frac{1+d}{1-d}\Big)$. This means that

$$RE\Big(\hat{\theta}^{\text{PTE}}_n(d) : \widetilde{\theta}_n\Big) \lesseqqgtr 1 \;\text{ according as }\; \Delta^2 \lesseqqgtr \Delta^2_*. \tag{5.7}$$

Finally, as $\Delta^2 \to \infty$, $\text{RE}(\hat{\theta}^{\text{PTE}}_n(d) : \widetilde{\theta}_n) \to 1$. Thus, the preliminary test estimator is more efficient than the unrestricted estimator whenever $\Delta^2 < \Delta^2_*$, otherwise $\widetilde{\theta}_n$ is more efficient. As for the relative efficiency of $\hat{\theta}^{\text{PTE}}_n(d)$ compared to $\hat{\theta}_n(d)$ we have

$$\text{RE}\Big(\hat{\theta}^{\text{PTE}}_n(d) : \hat{\theta}_n\Big) = [d^2 + (1-d)^2\Delta^2][1 + g(\Delta^2)]^{-1} \tag{5.8}$$

where

$$\begin{aligned} g(\Delta^2) \;&= (1-d)\Delta^2\Big\{2G_{3,\nu}\Big(\tfrac{1}{3}F_\alpha;\Delta^2\Big) - (1+d)G_{5,\nu}\Big(\tfrac{1}{5}F_\alpha;\Delta^2\Big)\Big\} \\ &\qquad -(1+d^2)G_{3,\nu}\Big(\tfrac{1}{3}F_\alpha;\Delta^2\Big). \end{aligned} \tag{5.9}$$

Under the null-hypothesis,

$$\text{RE}\Big(\hat{\theta}^{\text{PTE}}_n(d) : \hat{\theta}_n(d)\Big) = d^2\Big[1 - (1-d^2)G_{3,\nu}\Big(\frac{1}{3}F_\alpha; 0\Big)\Big]^{-1} \geq d^2. \tag{5.10}$$

At the same time we consider the result at (5.5). In combination, we obtain

$$d^2 \leq \text{RE}(\hat{\theta}^{\text{PTE}}_n(d) : \hat{\theta}_n(d)) \leq 1 \leq \text{RE}(\hat{\theta}^{\text{PTE}}_n(d) : \widetilde{\theta}_n). \tag{5.11}$$

For general $\Delta^2 > 0$, we have

$$\text{RE}\Big(\hat{\theta}^{\text{PTE}}_n(d) : \hat{\theta}_n(d)\Big) \lesseqqgtr 1 \;\text{ according as} \tag{5.12}$$

h

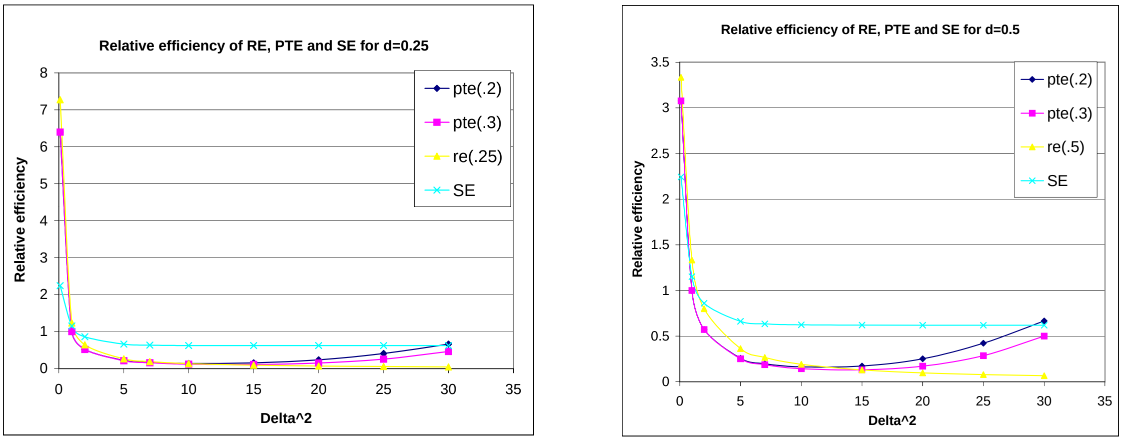



Figure 3: Graph of the relative efficiency of the RE, PTE and SE relative to UE

h

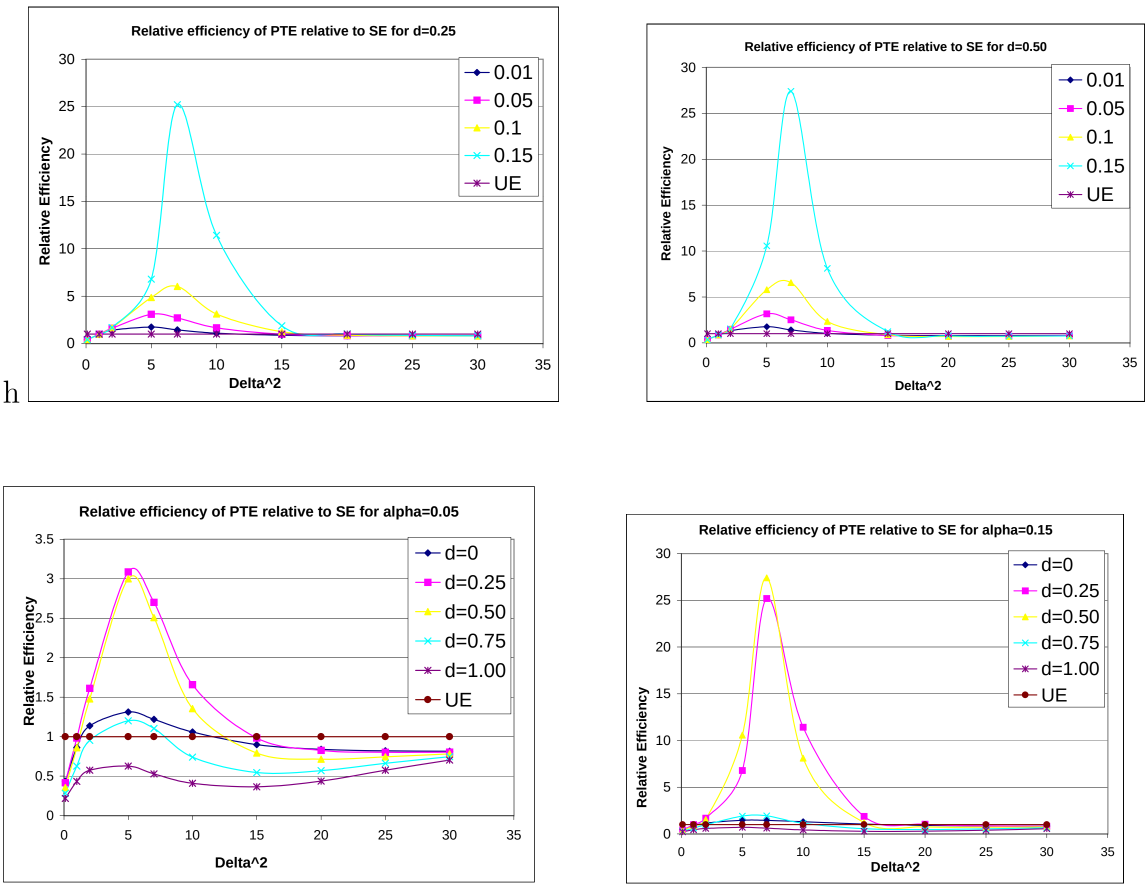



Figure 4: Graph of the relative efficiency of PTE relative to SE for selected values of $d$ and $\alpha$

$$\Delta^2 \lesseqgtr \frac{1+d}{1-d} \frac{\left\{1 - G_{3,\nu}\left(\frac{1}{3}F_\alpha; \Delta^2\right)\right\}}{\left\{1 - 2G_{3,\nu}\left(\frac{1}{3}F_\alpha; \Delta^2\right) - (1+d)G_{5,\nu}\left(\frac{1}{5}F_\alpha; \Delta^2\right)\right\}}. \quad (5.13)$$

Finally, as $\Delta^2 \to \infty$, $\text{RE}(\hat{\theta}_n^{\text{PTE}}(d); \hat{\theta}_n(d)) \to 0$. Thus, except for a small interval around 0, $\hat{\theta}_n^{\text{PTE}}(d)$ is more efficient than $\hat{\theta}_n(d)$.

**Comparing SE against UE**

The relative efficiency of $\hat{\theta}_n^s$ compared to $\widetilde{\theta}_n$ is given by

$$\text{RE}(\hat{\theta}_n^s : \widetilde{\theta}_n) = \left[1 - \frac{2}{\pi}K_n^2\left\{2e^{-\Delta^2/2} - 1\right\}\right]^{-1}. \quad (5.14)$$

Under the null-hypothesis $\Delta^2 = 0$, and hence

$$\text{RE}(\hat{\theta}_n^s : \widetilde{\theta}_n) = \left[1 - \frac{2}{\pi}K_n^2\right]^{-1} \geq 1. \quad (5.15)$$

In general, $\text{RE}(\hat{\theta}_n^s : \widetilde{\theta}_n)$ decreases from $[1 - \frac{2}{\pi}K_n^2]^{-1}$ at $\Delta^2 = 0$ and crosses the 1-line at $\Delta^2 = ln\ 4$ and then goes to the minimum value

$$\left[1 + \frac{2}{\pi}K_n^2\right]^{-1} \quad \text{as } \Delta^2 \to \infty. \quad (5.16)$$

Thus, the loss of efficiency of $\hat{\theta}_n^s$ relative to $\widetilde{\theta}_n$ is

$$1 - \left[1 + \frac{2}{\pi}K_n^2\right]^{-1} \quad (5.17)$$

while the gain in efficiency is

$$\left[1 - \frac{2}{\pi}K_n^2\right]^{-1} \quad (5.18)$$

which is achieved at $\Delta^2 = 0$. Thus, for $\Delta^2 < ln\ 4$, $\hat{\theta}_n^s$ performs better than $\widetilde{\theta}_n$, otherwise $\widetilde{\theta}_n$ performs better. The property of $\hat{\theta}_n^s$ is similar to the preliminary test estimator but does not depend on the level of significance.

**Comparing SE against PTE**

To compare the relative performances of the SE and the PTE, first note that the SE is superior to PTE when the null hypothesis is true and the level of significance, $\alpha$ is not too large. This is regardless of the value of the coefficient of distrust, $d$. However, as the value of $\Delta$ increases and or $\alpha$ grows larger the relative efficiency picture changes.

Table 2 provides a brief comparison of the performance of the PTE and SE relative to the UE when $d = 0$. The first two rows of the table gives the maximum and minimum relative efficiency of the SE for selected sample sizes. In general, the maximum relative efficiency of the SE increases as the sample size grows larger. Whereas the minimum relative efficiency has the opposite trend. The maximum relative efficiency of the SE is observed at $\Delta = 0$ regardless of the sample size. But the value of $\Delta$ at which the minimum relative efficiency is observed varies with the change in the sample size. Nevertheless, unlike that of the PTE, the relative efficiency of the SE remains constant with respect to the change in the value of $\Delta$ once it reaches its minimum.

**Table 2: Minimum and Maximum Efficiency of SE and Efficiency of PTE at $\Delta_0$ for Selected $\alpha$**

| $\alpha/n$ | | 10 | 15 | 20 | 25 | 30 | 35 | 40 |
|---|---|---|---|---|---|---|---|---|
| | $E^{max}$ | 2.6261 | 2.6692 | 2.6903 | 2.7029 | 2.7112 | 2.7171 | 2.7215 |
| | $E_{min}$ | 0.6176 | 0.6152 | 0.6141 | 0.6135 | 0.6131 | 0.6128 | 0.6125 |
| .05 | $E_{\Delta_0}$ | 0.6408 | 0.6466 | 0.6498 | 0.6518 | 0.6532 | 0.6542 | 0.6550 |
| | Eo | 0.3350 | 0.3600 | 0.3720 | 0.3790 | 0.3836 | 0.3868 | 0.3893 |
| | $\Delta_0$ | 6.1009 | 5.5303 | 5.2857 | 5.1514 | 5.0656 | 5.0063 | 4.9629 |
| .15 | $E_{\Delta_0}$ | 0.6827 | 0.6892 | 0.6924 | 0.6943 | 0.6955 | 0.6964 | 0.6971 |
| | Eo | 0.5405 | 0.5601 | 0.5693 | 0.5747 | 0.5782 | 0.5807 | 0.5825 |
| | $\Delta_0$ | 4.1626 | 3.9403 | 3.8429 | 3.7875 | 3.7520 | 3.7272 | 3.7090 |
| .25 | $E_{\Delta_0}$ | 0.7133 | 0.7182 | 0.7206 | 0.7220 | 0.7229 | 0.7236 | 0.7241 |
| | Eo | 0.6844 | 0.6990 | 0.7059 | 0.7098 | 0.7124 | 0.7143 | 0.7156 |
| | $\Delta_0$ | 3.4811 | 3.3612 | 3.3066 | 3.2757 | 3.2557 | 3.2417 | 3.2314 |
| .35 | $E_{\Delta_0}$ | 0.7361 | 0.7395 | 0.7410 | 0.7420 | 0.7426 | 0.7420 | 0.7433 |
| | Eo | 0.7941 | 0.8043 | 0.8090 | 0.8117 | 0.8135 | 0.8148 | 0.8157 |
| | $\Delta_0$ | 3.1163 | 3.0429 | 3.0110 | 2.9922 | 2.9801 | 2.9716 | 2.9653 |
| .45 | $E_{\Delta_0}$ | 0.7536 | 0.7555 | 0.7564 | 0.7569 | 0.7572 | 0.7576 | 0.7578 |
| | Eo | 0.8756 | 0.8820 | 0.8850 | 0.8867 | 0.8878 | 0.8887 | 0.8892 |
| | $\Delta_0$ | 2.8885 | 2.8442 | 2.8240 | 2.8125 | 2.8050 | 2.7979 | 2.7941 |

Legends: $\alpha$ is the level of significance
$n$ is the sample size
$E^{max}$ is the maximum efficiency of SE
$E_{mim}$ is the minimum efficiency of SE
$E_0$ is the minimum efficiency of PTE
$E_{\Delta_0}$ is the efficiency of SE at $\Delta_0$
$\Delta_0$ is the value of $\Delta$ at which the minimum efficiency of PTE occurs.

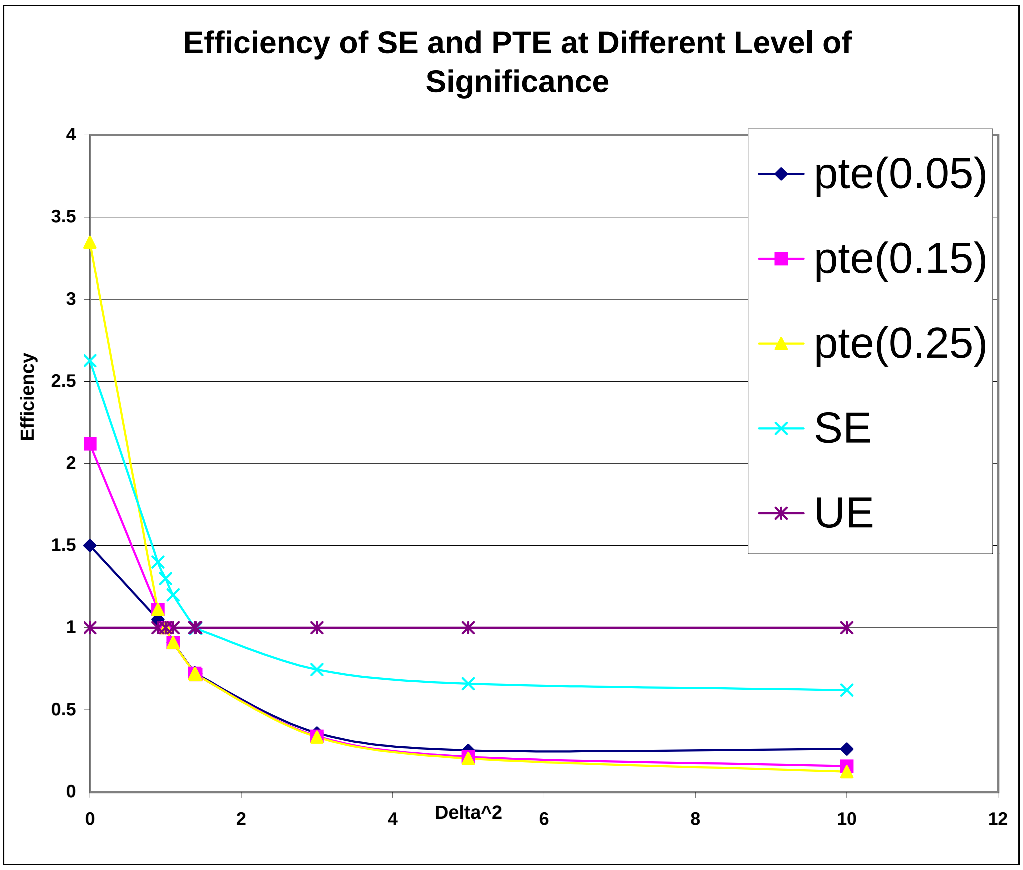
In the remaining rows of Table 2 the minimum relative efficiency of the PTE ($E_0$) for selected level of significance, as well as the corresponding relative efficiency of the SE ($E_{\Delta_0}$) at $\Delta_0$ have been recorded along with $\Delta_0$, the value of $\Delta$ at which h 


Figure 5: Graph of the relative efficiency of the SE and PTE for different value of $\alpha$

the PTE reaches its minimum. At any smaller level of significance the value of the relative efficiency of the SE ($E_{\Delta_0}$) is larger than the minimum relative efficiency of the PTE.

This is true for all sample sizes, but the difference between the minimum relative efficiency of the PTE and the corresponding relative efficiency of the SE decreases as the sample size increases when $\alpha$ is not too large. At, and up to, $\alpha = 0.25$, the relative efficiency of the SE at $\Delta_0$ is larger than the minimum relative efficiency of the PTE. However, as $\alpha$ increases further the minimum relative efficiency of the PTE becomes larger than the relative efficiency of the SE at $\Delta_0$. Although for $\alpha > 0.25$, the relative efficiency of the SE is smaller than the minimum relative efficiency of the PTE, for a given $n$ and $\Delta$, the SE has the advantage of not involving $\alpha$ and higher maximum relative efficiency than the PTE. For example, when $n = 20$ and $\Delta_0 = 3.0110$, the relative efficiency of the SE is 0.7410 (see Table 2) which is smaller than the minimum relative efficiency of the PTE 0.8090 at $\alpha = 0.35$ (see Table 1), but the corresponding maximum relative efficiency of the SE (2.6903) is a lot larger than the maximum relative efficiency of the PTE (1.2185).

For a fixed value of $d$, the relative efficiency of the SE with respect to the PTE is above the 1-line for some value of $\Delta$ near 0. Then it sides down rapidly, and passes the curve of the unit relative efficiency (of the UE). The top two graphs in Figure 4 demonstrate the behaviour of the relative efficiency curves for different values of $\alpha$ when $d = 0.25$ and $d = 0.50$ respectively. It is clear that as the value of $\alpha$ increases, the relative efficiency of the PTE with respect to the SE grows higher. However, a larger value of $\alpha$ is not desirable. When the value of $\alpha$ is lower the relative efficiency of the PTE is also lower, and hence the SE over performs the PTE.

From the foregoing discussions and Figure 5, it is clear that the relative efficiency of the PTE relative to the UE is lower than 1 for $\Delta > 1$ and that of the SE relative to the UE is lower than 1 for $\Delta > 1.38$ when $d = 0$. Thus the SE dominates the UE over a wider interval, $[0,\ 1.38)$ than the PTE in the interval $(0,\ 1]$. Also, from Figure 5, for $\alpha < 0.25$, the SE has higher relative efficiency than the PTE over all $\Delta$. However, if $\alpha \geq 0.25$, for some small interval $(0 \leq \Delta < \Delta' < 1)$ the PTE over performs the SE; but for every $\Delta \geq \Delta'$ the SE dominates the PTE.

There is no uniform domination of the SE over the PTE for all $\Delta$ and every $\alpha$. Clearly, the superior performance of the SE relative to the PTE depends on the value of $\Delta$. When the value of $\Delta$ is in the neighborhood of 0, the SE over performs the PTE for every value of $\alpha < 0.25$. But, the value of $\Delta$ is near 0 ( that is, $\theta - \theta_0 \to 0$) only when the value of the prior non-sample information is reasonably accurate (not far from the true value). In other words, if the value of $\theta$ provided by the non-sample information is not too far from its true value then the SE dominates the PTE. Furthermore, an unreasonable (far away from the true) value of prior non-sample information is unlikely to be used by the researchers. In practice, since the prior non-sample information is based on practical experience or expert knowledge, it is expected to be close enough to the true value of $\theta$ to make $\Delta$ close to 0, and hence the SE would expected to be a preferred option over the PTE.

# 6 Concluding Remarks

The UE is based on the sample data alone and it is the only unbiased estimator among the four esimators considered in this paper. The introduction of the non-sample information in the estimation process causes the estimators to be biased.

However, the biased estimators perform better than the unbiased estimator when judged based on the mse criterion. The performance of the biased estimators depend on the value of the departure parameter $\Delta$. In case of the PTE, the performance also depends on the value of the level of significance. Under the null hypothesis, the departure parameter is zero, and the SE bits all other estimators if $\alpha$ is not too high. As $\alpha$ increases, the performance of the PTE improves. At a lower level of significance, the SE performs better than the PTE more often and over a wider range of values of $\Delta$. When the value of $\Delta$ is not far from 0, the SE always over performs the PTE and RE. Therefore, in practice if the researcher could gather a value of $\theta$ that is not far from its true value, the SE would be the best choice as an 'improved' estimator of the mean.

## References


1. Ahmed, S.E. and Saleh, A.K.Md.E. (1989). Pooling multivariate data. Jou. Statist. Comput. Simul. Vol. 31, 149–167

2. 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.

3. Bolfarine, H. and Zacks, S. (1992). Prediction Theory for Finite Populations. Springer-Verlag, New York.

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

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

6. Khan, S. (1998). On the estimation of the mean vector of Student-t population with uncertain prior information. Pak. Jou. Statist., **14**, 161-175.

7. 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, 1-17.

8. Khan, S. and Saleh, A.K.Md.E. (1995). Preliminary test estimators of the mean based on $p$-samples from multivariate Student-t populations. Bulletin of the International Statistical Institute. 50th Session of the ISI, Beijing, 599–600.

9. Maatta, J.M. and Casella, G. (1990). Developments in decision-theoretic variance estimation, Statistical Sciences, Vol. 5, 90–101.

10. Saleh, A.K.Md.E. and Sen, P.K. (1985). Shrinkage least squares estimation in a general multivariate linear model. Proceedings of the Fifth Pannonian Symposium on Mathematical Statistics, 307–325.

11. Saleh, A.K.Md.E. and Sen, P.K. (1978). Nonparametric estimation of location parameter after a preliminary test on regression. Annals of Statistics, Vol 6, 154–168.

12. Sclove, S.L., Morris, C. and Rao, C.R. (1972). Non-optimality of preliminary-test estimators for the mean of a multivariate normal distribution. Ann. Math. Statist. Vol. 43, 1481–1490.

13. 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, Vol. 1, 197–206.

14. Stein, C. (1981). Estimation of the mean of a multivariate normal distribution. Ann. Statist. Vol. 9, 1135–1151.