ESTIMATION OF THE SLOPE PARAMETER FOR LINEAR REGRESSION MODEL WITH UNCERTAIN PRIOR INFORMATION

S. Khan [1], Z. Hoque [2] and A.K.Md. E. Saleh[3]

SUMMARY

The estimation of the slope parameter of the linear regression model with normal error is considered in this paper when uncertain prior information on the value of the slope is available. Several alternative estimators are defined to incorporate both the sample as well as the non-sample information in the estimation process. Some 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. Both analytical and graphical methods are explored. None of the estimators is found to be uniformly superior over the others. However, if the non-sample information regarding the value of the slope is close to its true value, the shrinkage estimator over performs the rest of the estimators.

Keywords: 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.

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

1 Introduction

Traditionally the classical estimators of unknown parameters are based exclusively on the sample information. 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

[1]Department of Mathematics and Computing, University of Southern Queensland, Toowoomba, Queensland, Australia, email: khans@usq.edu.au. This work was initiated when working as a visiting consultant in the Department of Mathematics and Statistics at the Sultan Quaboos University, Oman during the second half of 2001

[2]Department of Mathematics and Computing, University of Southern Queensland, Toowoomba, Queensland, Australia, email: hoque@usq.edu.au. On leave from Department of Statistics, University of Chittagong, Bangladesh

[3]School of Mathematics and Statistics, Carleton University, Ottawa, Canada

have been introduced in the literature that, under particular situation, over performs the traditional exclusive sample information based unbiased estimators when judged by criteria such as the mean square error and squared error loss.

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 term of prior distribution), in addition to the sample information. Stein (1956) introduced the Stein-rule (shrinkage) estimator for multivariate normal population that dominates the usual maximum likelihood estimator under the squared error loss criterion. 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 slope parameter of linear regression equation has not been investigated.

Consider the linear regression equation

$$y = \beta_0 + \beta_1 x + e \tag{1.1}$$

where y is the response variable; β_0 is the intercept parameter; β_1 is the slope parameter; x is the predictor variable and e is the error component associated with the response variable. Assume that the errors are independently and identically distributed as normal variables with mean 0 and variance σ^2. Then, in the conventional notation we write, $e \sim N(0,\ \sigma^2)$. Also assume that uncertain non-sample prior information on the value of the slope parameter, β_1 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, $H_0 :\ \beta_1 = \beta_{10}$ which may be true, but not sure. We wish to incorporate both the sample information and the uncertain non-sample prior information in estimating the slope β_1. Furthermore, we assign a coefficient of distrust, $0 \leq d \leq 1$, for the non-sample prior information, that represents the degree of distrust in the null hypothesis. It is assumed that the intercept parameter is unknown and estimated by the maximum likelihood estimator (mle). First we obtain the *unrestricted* mle of the unknown slope β_1 and the common variance σ^2 from the likelihood function of the sample. Based on the unrestricted and *restricted* (by the null hypothesis) mle of σ^2, we derive the likelihood ratio test for testing $H_0 :\ \beta_1 = \beta_{10}$ against $H_A :\ \beta_1 \neq \beta_{10}$. 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 slope.

It is well known that the mle of the slope parameter is unbiased. We wish to search for an alternative estimator of the slope parameter 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 to critically check various affects on the properties of the estimators. 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 β_1 is not too far from its true value. 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 β_1 and σ^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 from (1.1) in the following convenient form

$$\boldsymbol{y} = \beta_0 \mathbf{1}_n + \beta_1 \boldsymbol{x} + \boldsymbol{e} \tag{2.1}$$

where $\boldsymbol{y} = (y_1, \ldots, y_n)'$ is an $n \times 1$ vector of responses, $\mathbf{1}_n = (1, \ldots, 1)'$ – a vector of n-tuple of one's, $\boldsymbol{x}$ is the $n \times 1$ vector of explanatory variable, β_0 and β_1 are the unknown intercept and slope parameters respectively and $\boldsymbol{e} = (e_1, \ldots, e_n)'$ is a vector of errors with independent components which is distributed as $N_n(\mathbf{0},\ \sigma^2 I_n)$. So that

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

Here, σ^2 stands for the variance of each of the error component in $\boldsymbol{e}$ and I_n is the identity matrix of order n. From the exclusive sample information, the *unrestricted estimator* (UE) of the slope β_1 is the usual maximum likelihood estimator (mle) given by

$$\tilde{\beta}_1 = (\boldsymbol{x}'\boldsymbol{x})^{-1}\boldsymbol{x}'\boldsymbol{y}. \tag{2.3}$$

It is well known that, for the normal model, the sampling distribution of the mle of β_1 is normal with mean, $E(\tilde{\beta}_1) = \beta_1$ and variance, $E(\tilde{\beta}_1 - \beta_1)^2 = \frac{\sigma^2}{S_{xx}}$ in which

$S_{xx} = \sum_{j=1}^{n}(x_j - \bar{x})^2$. Therefore, $\tilde{\beta}_1$ is unbiased for β_1, and hence the mse is the same as its variance. Here, the bias and the mse of $\tilde{\beta}_1$ are given by

$$B_1(\tilde{\beta}_1) = 0 \quad \text{and} \quad M_1(\tilde{\beta}_1) = \frac{\sigma^2}{S_{xx}} \quad \text{respectively.} \tag{2.4}$$

We compare the above bias and mse functions with those of the three biased estimators, and search for a 'best' estimator that may perform better than the other estimators under some specific condition. It is well known that the mle of σ^2 is

$$S_n^{*2} = \frac{1}{n}(\boldsymbol{y} - \hat{\boldsymbol{y}})'(\boldsymbol{y} - \hat{\boldsymbol{y}}) \tag{2.5}$$

where $\hat{\boldsymbol{y}} = \tilde{\beta}_0 \mathbf{1}_n + \tilde{\beta}_1 \boldsymbol{x}$ in which $\tilde{\beta}_0$ is the mle of β_0.

This estimator of σ^2 is biased. However, an unbiased estimator of σ^2 is given by

$$S_n^2 = \frac{1}{n-2}(\boldsymbol{y} - \hat{\boldsymbol{y}})'(\boldsymbol{y} - \hat{\boldsymbol{y}}). \tag{2.6}$$

The unbiased estimator of σ^2 has a scaled χ^2 distribution with shape parameter $\nu = (n-2)$. The standard error of $\tilde{\beta}_1$ is $\frac{S_n}{\sqrt{S_{xx}}}$.

To be able to use the uncertain non-sample prior information in the estimation of the slope, 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, $H_0 : \beta_1 = \beta_{10}$ and treat it as a nuisance parameter. He proposed to conduct an appropriate statistical test on the null-hypothesis against the alternative $H_A : \beta_1 \neq \beta_{10}$ 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

$$\mathcal{L}_\nu = \frac{\sqrt{S_{xx}}(\tilde{\beta}_1 - \beta_{10})}{S_n}. \tag{2.7}$$

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

$$\Delta^2 = \frac{S_{xx}(\beta_1 - \beta_{10})^2}{\sigma^2}. \tag{2.8}$$

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 appropriate degrees of freedom. This test statistic was used by 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 of the Slope

As part of incorporating the uncertain non-sample prior information into the estimation process, first we combine the exclusive sample based estimator, $\tilde{\beta}_1$ with the non-sample prior information presented in the form of a null hypothesis, $H_0 : \beta_1 = \beta_{10}$ in some reasonable way. First, consider a simple linear combination of β_{10} and $\tilde{\beta}_1$ as

$$\hat{\beta}_1(d) = d\tilde{\beta}_1 + (1-d)\beta_{10}, \quad 0 \leq d \leq 1. \tag{3.1}$$

This estimator of β_1 is called the *restricted estimator* (RE), where d is the *degree of distrust* in the null hypothesis, $H_0 : \beta_1 = \beta_{10}$. Here, $d = 0$, means there is *no distrust* in the H_0 and we get $\hat{\beta}_1(d = 0) = \beta_{10}$, while $d = 1$ means there is *complete distrust* in the H_0 and we get $\hat{\beta}_1(d = 1) = \tilde{\beta}_1$. If $0 < d < 1$, the degree of distrust is an intermediate value which results in an interpolated value between β_{10} and $\tilde{\beta}_1$ given by (3.1). The restricted estimator, as defined above, is normally distributed with mean and mean square error given by

$$E[\hat{\beta}_1(d)] = d\beta_1 + (1-d)\beta_{10} \quad \text{and} \quad \mathrm{M}[\hat{\beta}_1(d)] = \frac{\sigma^2}{S_{xx}}[d^2 + (1-d)^2\Delta^2] \tag{3.2}$$

respectively. Following Bancroft (1944) we define a preliminary test estimator (PTE) of the slope parameter as

$$\begin{aligned} \hat{\beta}_1^{\mathrm{PTE}}(d) &= \hat{\beta}_1(d) I\big(|t_\nu| < t_{\alpha/2}) + \tilde{\beta}_1 I(|t_\nu| \geq t_{\alpha/2}\big) \\ &= \tilde{\beta}_1 - (1-d)(\tilde{\beta}_1 - \beta_{10}) I(|t_\nu| < t_{\alpha/2}) \end{aligned} \tag{3.3}$$

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

$$\hat{\beta}_1^{\mathrm{PTE}} = \beta_{10} I\big(|t_\nu| < t_{\alpha/2}\big) + \tilde{\beta}_1 I\big(|t_\nu| \geq t_{\alpha/2}\big), \tag{3.4}$$

which is a special case of (3.3) when $d = 0$. Note that, the $\hat{\beta}_1^{\mathrm{PTE}}(d)$ is a convex combination of $\hat{\beta}_1(d)$ and $\tilde{\beta}_1$, and $\hat{\beta}_1^{\mathrm{PTE}}(d = 0)$ is a convex combination of β_{10} and $\tilde{\beta}_1$. We may rewrite (3.3) as

$$\hat{\beta}_1^{\mathrm{PTE}}(d) = \tilde{\beta}_1 - (1-d)(\tilde{\beta}_1 - \beta_{10}) I(F < F_\alpha) \tag{3.5}$$

where F_α 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

$$\hat{\beta}_1^{\mathrm{PTE}}(d = 0) = \tilde{\beta}_1 - (\tilde{\beta}_1 - \beta_{10}) I(F < F_\alpha). \tag{3.6}$$

The PTE is an extreme choice between $\hat{\beta}_1(d)$ and $\tilde{\beta}_1$. Hence it does not allow any smooth transition between it's two extreme values. Also, it depends on the preselected level of significance, α of the test. To overcome these problems, we consider the shrinkage estimator (SE) of β_1 defined as follows:

$$\hat{\beta}_1^{\mathrm{SE}} = \beta_{10} + \left\{ 1 - \frac{cS_n}{|\sqrt{S_{xx}}(\tilde{\beta}_1 - \beta_{10})|} \right\} (\tilde{\beta}_1 - \beta_{10}). \tag{3.7}$$

Note that in this estimator c is a constant function of n. Now, if $|t_\nu| = \left| \frac{\sqrt{S_{xx}}(\tilde{\beta}_1 - \beta_{10})}{S_n} \right|$ is large, $\hat{\beta}_1^{\mathrm{SE}}$ tends towards $\tilde{\beta}_1$, while for small $|t_\nu|$ equaling c, $\hat{\beta}_1^{\mathrm{SE}}$ tends towards β_{10} similar to the preliminary test estimator. Unlike the preliminary test estimator, the shrinkage estimator does not depend on the level of significance.

4 Some Statistical Properties

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

First the bias and the mse of the RE, $\hat{\beta}_1(d)$ are found to be

$$\begin{aligned} B_2[\hat{\beta}_1(d)] &= -\frac{\sigma}{\sqrt{S_{xx}}}(1-d)\Delta, \quad \Delta = \frac{\sqrt{S_{xx}}(\beta_1 - \beta_{10})}{\sigma} & (4.1) \\ M_2[\hat{\beta}_1(d)] &= \frac{\sigma^2}{S_{xx}}\left[d^2 + (1-d)^2\Delta^2\right] & (4.2) \end{aligned}$$

where Δ^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 Δ. We investigate this feature in a greater detail in the forthcoming sections.

4.1 The Bias and the MSE of PTE

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

$$\begin{aligned} E[\hat{\beta}_1^{\mathrm{PTE}}(d) - \beta_1] &= E(\tilde{\beta}_1 - \beta_1) - (1-d)E\big\{(\tilde{\beta}_1 - \beta_{10})I\big(F < F_\alpha\big)\big\} & (4.3) \\ &= -(1-d)\frac{\sigma}{\sqrt{S_{xx}}}E\left\{ \frac{\sqrt{S_{xx}}(\tilde{\beta}_1 - \beta_{10})}{\sigma} I\left(\frac{S_{xx}(\tilde{\beta}_1 - \beta_{10})^2}{S_n^2} < F_\alpha \right) \right\}. \end{aligned}$$

Note $Z = \sqrt{S_{xx}}(\tilde{\beta}_1 - \beta_{10})/\sigma$ is distributed as $N(\Delta, 1)$, where $\Delta = \frac{\sqrt{S_{xx}}}{\sigma}(\beta_1 - \beta_{10})$, and $S_{xx}(n-2)S_n^2/\sigma^2$ is distributed (independently) as a central chi-square variable with $\nu = (n-2)$ degrees of freedom.

Evaluating the expression in (4.3) the bias function of $\hat{\beta}_1^{\mathrm{PTE}}(d)$ is found to be

$$\begin{aligned} B_3[\hat{\beta}_1^{\mathrm{PTE}}(d)] &= -(1-d)\frac{\sigma}{\sqrt{S_{xx}}}\Delta G_{3,\nu}\Big(\frac{1}{3}F_\alpha; \Delta^2\Big) \\ &= -(1-d)(\beta_1 - \beta_{10})G_{3,\nu}\Big(\frac{1}{3}F_\alpha; \Delta^2\Big), \end{aligned} \tag{4.4}$$

where $G_{a,b}(\cdot; \Delta^2)$ is the c.d.f. of a non-central F-distribution with (a, b) degrees of freedom and non-centrality parameter Δ^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 $Z \sim \mathcal{N}(\Delta, 1)$ and $\phi(Z^2)$ is a Borel measurable function, then*

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

To obtain the mean square error of $\hat{\beta}_1^{\mathrm{PTE}}(d)$ 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*

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

The proof of the above two theorems are given in Appendix B2 of Judge and Bock (1978). From the definition, the mse expression of the PTE is

$$\begin{aligned} M_3\left[\hat{\beta}_1^{\mathrm{PTE}}(d)\right] &= E\left[\hat{\beta}_1^{\mathrm{PTE}}(d) - \beta_1\right]^2 \\ &= E(\tilde{\beta}_1 - \beta_1)^2 + (1-d)^2 E(\tilde{\beta}_1 - \beta_{10})^2 I(F < F_\alpha) \\ &\quad -2(1-d)E[(\tilde{\beta}_1 - \beta_1)(\beta_1 - \beta_{10})]I(F < F_\alpha) \\ &= \frac{\sigma^2}{S_{xx}} + (1-d)^2 E[(\tilde{\beta}_1 - \beta_{10})^2 I(F < F_\alpha)] \\ &\quad -2(1-d)E\left[\{(\tilde{\beta}_1 - \beta_{10}) - (\beta_1 - \beta_{10})\}(\beta_1 - \beta_{10})I(F < F_\alpha)\right]. \end{aligned} \tag{4.7}$$

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,

$$\begin{aligned} M_3[\hat{\beta}_1^{\mathrm{PTE}}(d)] = \frac{\sigma^2}{S_{xx}}\Bigg[&1 - (1-d^2)G_{3,\nu}\Big(\frac{1}{3}F_\alpha; \Delta^2\Big) \\ &+(1-d)\Delta^2\Big\{2G_{3,\nu}\Big(\frac{1}{3}F_\alpha; \Delta^2\Big) - (1+d)G_{5,\nu}\Big(\frac{1}{5}F_\alpha; \Delta^2\Big)\Big\}\Bigg]. \end{aligned} \tag{4.8}$$

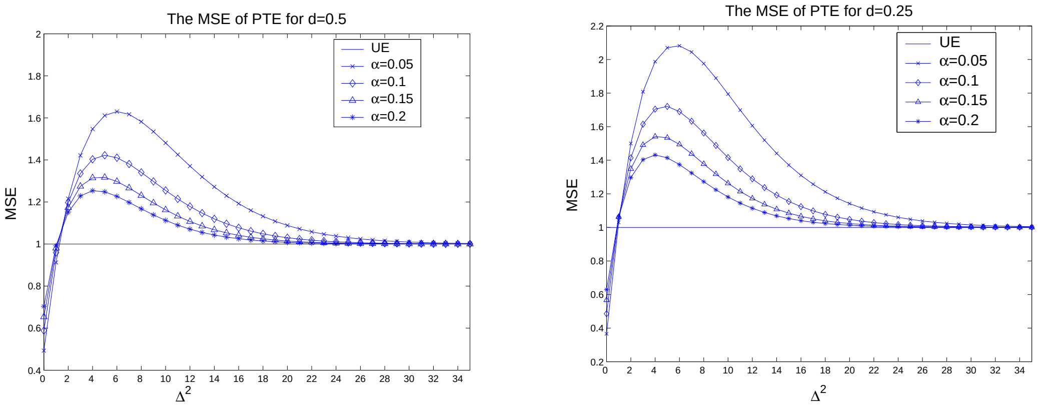

Figure 1: Graph of the mean square error of PTE as a function of Δ^2

Figure 1 displays the behavior of the mse function of the PTE for different values of α with the change in the value of Δ^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 $\hat{\beta}_1^{\text{PTE}}(d)$ equals

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

Thus, at $\Delta^2 = 0$ PTE of β_1 performs better than $\tilde{\beta}_1$, the UE. As $\alpha \to 0$, $G_{3,\nu}\left(\frac{1}{3}F_\alpha;0\right) \to 1$, then

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

which is the mse of $\hat{\beta}_1(d)$. On the other hand, if $F_\alpha \to 0$, $G_{3,\nu}\left(\frac{1}{3}F_\alpha;0\right) \to 0$, then

$$\frac{\sigma^2}{S_{xx}}\left[1-(1-d^2)G_{3,\nu}\left(\frac{1}{3}F_\alpha;0\right)\right] \to \frac{\sigma^2}{S_{xx}}, \quad \text{which is the mse of } \tilde{\beta}_1. \tag{4.11}$$

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

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

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

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

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

$$\leq \frac{\sigma^2}{S_{xx}} \quad \text{whenever} \quad \Delta^2 < \frac{1+d}{2}.$$

This means that the mse of $\hat{\beta}_1^{\mathrm{PTE}}(d)$ as a function of Δ^2 crosses the constant line $M_1(\tilde{\beta}_1) = \frac{\sigma^2}{S_{xx}}$ in the interval $\left(\frac{1+d}{2}, \frac{1+d}{1-d}\right)$.

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

4.1.1 Determination of optimum α 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 α of the test of null-hypothesis and the departure parameter Δ^2.

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

$$E(\alpha;\Delta^2) = [1+g(\Delta^2)]^{-1}, \qquad (4.14)$$

where

$$\begin{aligned}
g(\Delta^2) &= 1+(1-d)\Delta^2\Big\{2G_{3,\nu}\Big(\frac{1}{3}F_\alpha;\Delta^2\Big)-(1+d)G_{5,\nu}(\frac{1}{5}F_\alpha;\Delta^2\Big)\Big\} \qquad (4.15)\\
&\quad -(1-d^2)G_{3,\nu}\Big(\frac{1}{3}F_\alpha;\Delta^2\Big).
\end{aligned}$$

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

$$E(\alpha; 0) = \left[1 - (1 - d^2)G_{3,\nu}\left(\frac{1}{3}F_\alpha; 0\right)\right]^{-1} \geq 1. \quad (4.16)$$

As Δ^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

$$\max_\alpha E(\alpha, 0) = E(0, 0) = d^{-2}. \quad (4.17)$$

As a function of α, $E(\alpha; 0)$ decreases as α increases. On the other hand, $E(\alpha; \Delta^2)$ as a function of Δ^2 is decreasing, and the curves $E(0; \Delta^2)$ and $E(1/2; \Delta^2) = 1$ intersect at $\Delta^2 = \frac{1+d}{1-d}$. The value of Δ^2 at the intersection decreases as α increases. Therefore, for two different levels of significance say, α_1 and α_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 $0 \leq \Delta \leq \frac{1+d}{1-d}$, $\hat{\beta}_1$ is always chosen since $E(0, \Delta^2)$ is maximum for all Δ^2 in this interval. Generally, Δ^2 is unknown. In this case there is no way of choosing a uniformly best estimator of β_1. Thus, we pre-assign a tolerable relative efficiency, say, E_0. Then, consider the set

$$A_\alpha = \left\{\alpha | E(\alpha; \Delta^2) \geq E_0\right\}. \quad (4.18)$$

An estimator $\hat{\beta}_1^{\text{PTE}}(d))$ is chosen which maximizes $E(\alpha; \Delta^2)$ over all $\alpha \in A_\alpha$ and Δ^2. Thus, we solve the following equation for α

$$\max_\alpha \min_{\Delta^2} E(\alpha; \Delta^2) = E_0. \quad (4.19)$$

The solution α^* provides a *maximin rule* for the optimum level of significance of the preliminary test. A numerical procedure along with practical illustration of selecting an optimal α is provided in Khan and Saleh (2001).

4.2 The Bias and MSE of SE

Now, following Bolfarine and Zacks (1992) we compute the bias and the mse of the SE, $\hat{\beta}_1^{\text{SE}}$. The bias of the SE is given by

$$B_4(\hat{\beta}_1^{\text{SE}}) = E[\hat{\beta}_1^{\text{SE}} - \beta_1] = -cE\left[\frac{S_n(\tilde{\beta}_1 - \beta_{10})}{|\sqrt{S_{xx}}(\tilde{\beta}_1 - \beta_{10})|}\right] = -\frac{c}{\sqrt{S_{xx}}}E[S_n]E\left\{\frac{Z}{|Z|}\right\} \quad (4.20)$$

where $Z = \frac{\sqrt{S_{xx}}(\tilde{\beta}_1 - \beta_{10})}{\sigma} \sim \mathcal{N}(\Delta, 1)$. To evaluate $E\Big\{\frac{Z}{|Z|}\Big\}$ we use the theorem below.

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

$$E\left\{\frac{Z}{|Z|}\right\} = 1 - 2\Phi(-\Delta) \tag{4.21}$$

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, $QB_4(\hat{\beta}_1^{\mathrm{SE}})$ is obtained as

$$QB_4(\hat{\beta}_1^{\mathrm{SE}}) = \frac{\sigma^2}{S_{xx}} c^2 K_n^2 \{1 - 2\Phi(-\Delta)\}^2 = \frac{\sigma^2}{S_{xx}} c^2 K_n^2 \{2\Phi(\Delta) - 1\}^2 \tag{4.22}$$

where $K_n = \sqrt{\frac{2}{n-2}} \frac{\Gamma(\frac{n-1}{2})}{\Gamma(\frac{n-2}{2})}$. As $\Delta^2 \to 0$, $QB_4(\hat{\beta}_1^{\mathrm{SE}}) \to 0$ and as $\Delta^2 \to \infty$, $QB_4(\hat{\beta}_1^{\mathrm{SE}}) \to \frac{\sigma^2}{S_{xx}} K_n^2 c^2$. Therefore, $QB_4(\hat{\beta}_1^{\mathrm{SE}})$ is a non-decreasing monotonic function of Δ^2. Thus, unless Δ^2 is near the origin, the quadratic bias of the SE is significantly large.

In order to compute the mse of $\hat{\beta}_1^{\mathrm{SE}}$ we consider

$$\begin{aligned} E(\hat{\beta}_1^{\mathrm{SE}} - \beta_1)^2 &= E(\tilde{\beta}_1 - \beta_1)^2 + c^2 E(S_n^2) E\left\{\frac{(\tilde{\beta}_1 - \beta_{10})^2}{[\sqrt{S_{xx}}(\tilde{\beta}_1 - \beta_{10})]^2}\right\} \\ &\quad -2cE\Bigg\{\frac{(\tilde{\beta}_1 - \beta_1)(\tilde{\beta}_1 - \beta_{10})}{|\sqrt{S_{xx}}(\tilde{\beta}_1 - \beta_{10})|}\Bigg\} E(S_n) \\ &= \frac{\sigma^2}{S_{xx}} + \frac{c^2\sigma^2}{S_{xx}} - 2c\frac{\sigma^2 K_n}{S_{xx}}\Big\{E(|Z|) - \Delta E\Big(\frac{Z}{|Z|}\Big)\Big\}. \end{aligned} \tag{4.23}$$

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*

$$E(|Z|) = \sqrt{\frac{2}{\pi}} e^{-\Delta^2/2} + \Delta\{2\Phi(\Delta) - 1\} \tag{4.24}$$

where $\Phi(\cdot)$ is the c.d.f. of the standard normal variable. See Khan and Saleh (2001) for the proof of the above theorem.

Therefore, the mse of $\hat{\beta}_1^{\mathrm{SE}}$ is given by

$$M_4(\hat{\beta}_1^{\mathrm{SE}}) = \frac{\sigma^2}{S_{xx}}\Bigg\{1 + c^2 - 2cK_n\sqrt{\frac{2}{\pi}} e^{-\Delta^2/2}\Bigg\}. \tag{4.25}$$

The value of c which minimizes (4.26) depends on Δ^2 and is given by

$$c^* = \sqrt{\frac{2}{\pi}} K_n e^{-\Delta^2/2}. \tag{4.26}$$

To make c^* independent of Δ^2, we choose $c^0 = \sqrt{\frac{2}{\pi}} K_n$. Thus, optimum $M_4(\hat{\beta}_1^{\mathrm{SE}})$ becomes

$$M_4(\hat{\beta}_1^{\mathrm{SE}}) = \frac{\sigma^2}{S_{xx}} \Big\{ 1 - \frac{2}{\pi} K_n^2 \Big[2e^{-\Delta^2/2} - 1 \Big] \Big\}. \tag{4.27}$$

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

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

$$\begin{aligned} QB_2[\hat{\beta}_1(d)] &= \frac{\sigma^2}{S_{xx}} (1-d)^2 \Delta^2 \qquad (5.1) \\ QB_3[\hat{\beta}_1^{\mathrm{PTE}}(d)] &= \frac{\sigma^2}{S_{xx}} (1-d)^2 \Delta^2 \Big\{ G_{3,\nu}\Big(\frac{1}{3}F_\alpha; \Delta^2\Big) \Big\}^2 \\ QB_4[\hat{\beta}_1^{\mathrm{SE}}] &= \frac{\sigma^2}{S_{xx}} c^2 K_n^2 \{2\Phi(\Delta) - 1\}^2. \end{aligned}$$

Clearly, under the null-hypothesis $QB_2[\hat{\beta}_1(d)] = QB_3[\hat{\beta}_1^{\mathrm{PTE}}(d)] = QB_4[\hat{\beta}_1^{\mathrm{SE}}(d)] = 0$ for all d and α. When $\Delta^2 \to \infty$, $QB_2[\hat{\beta}_1(d)] \to \infty$ except at $d = 1$; $QB_3[\hat{\beta}_1^{\mathrm{PTE}}(d)] \to 0$ for all α and d; and $QB_4[\hat{\beta}_1^{\mathrm{SE}}] \to \frac{\sigma^2}{S_{xx}} c^2 K_n^2$, a constant that does not depend on d. Therefore, in terms of quadratic bias, the RE is uniformly dominated by both the PTE and SE. For very large values of Δ^2, the SE is dominated by the PTE regardless of the value of α. From small to moderate values of Δ^2, there is no uniform domination of one estimator over the other. In this case, domination depends on the level of significance, α. For small values of α, the PTE is dominated by the SE, and for larger values of α, the SE is dominated by the PTE. However, Chiou and Saleh (2002) suggest the value of α to be between 20% and 25%. In this interval of α, the quadratic bias of the PTE approaches to zero for not too small values of Δ^2. However, in practice, the non-centrality parameter is unlikely to be very

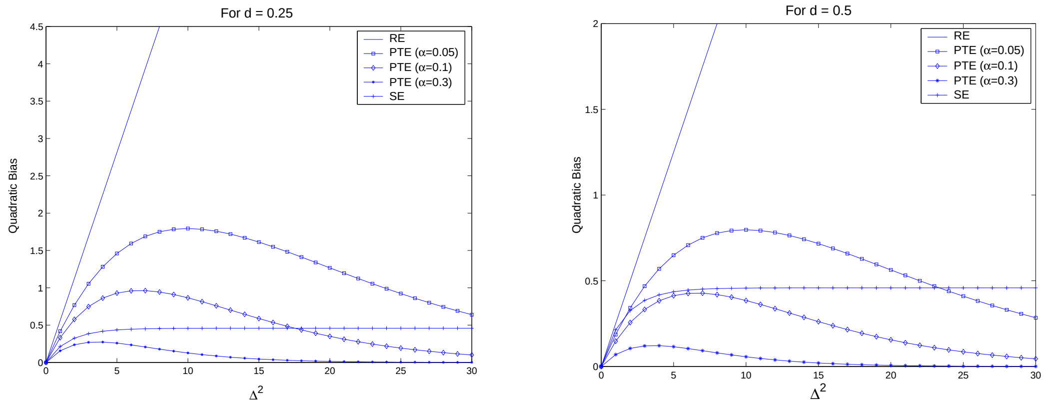

Figure 2: Graph of the quadratic bias of the RE, PTE and SE as a function of Δ^2

large (otherwise the credibility of prior information is in serious question), and α is usually preferred to be small. The quadratic bias of the SE is relatively stable and approaches to a constant value starting from some moderate value of Δ^2, and is unaffected by the choice of d and α. Therefore, the SE may be a better choice among the biased estimators considered in this paper. The graph of $QB_2[\hat{\beta}_1(d)]$, $QB_3[\hat{\beta}_1^{\mathrm{PTE}}(d)]$ and $QB_4[\hat{\beta}_1^{\mathrm{SE}}]$ are given in Figure 2.

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 $\hat{\beta}_1(d)$ compared to $\tilde{\beta}_1$ is denoted by $\mathrm{RE}[\hat{\beta}_1(d):\tilde{\beta}_1]$ and is obtained as

$$\mathrm{RE}[\hat{\beta}_1(d):\tilde{\beta}_1] = \left[d^2 + (1-d)^2\Delta^2\right]^{-1}. \tag{5.2}$$

We observe the following based on (5.2).

(i) If the non-sampling information is correct, i.e., $\Delta^2 = 0$, the $\mathrm{RE}[\hat{\beta}_1(d):\tilde{\beta}_1] = d^{-2} > 1$ and $\hat{\beta}_1(d)$ is more efficient than $\tilde{\beta}_1$. 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 (5.2) as a function of Δ^2 for a fixed d-value. As a function of Δ^2, (5.2) 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 $\Delta^2 \in \left[0, \frac{1+d}{1-d}\right)$, $\hat{\beta}_1(d)$ is more efficient than $\tilde{\beta}_1$, and outside this interval $\tilde{\beta}_1$ is more efficient than $\hat{\beta}_1(d)$. For example, if $d = \frac{1}{2}$, the interval in which $\hat{\beta}_1(d)$ is more efficient than $\tilde{\beta}_1$ is $[0, 3)$, while $\tilde{\beta}_1$ is more efficient in $[3, \infty)$ than $\hat{\beta}_1(d)$. For d= 0.5 the maximum efficiency of $\hat{\beta}_1(d)$ over $\tilde{\beta}_1$ is 4.

Comparing PTE against UE

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

$$RE[\hat{\beta}_1^{\text{PTE}}(d) : \tilde{\beta}_1] = \Big[1 - (1-d^2) G_{3,\nu}\Big(\frac{1}{3}F_\alpha; \Delta^2\Big) + (1-d)\Delta^2 \times \Big\{2G_{3,\nu}\Big(\frac{1}{3}F_\alpha; \Delta^2\Big) - (1+d) G_{5,\nu}\Big(\frac{1}{5}F_\alpha; \Delta^2\Big)\Big\}\Big]^{-1} \tag{5.3}$$

for any fixed d $(0 \leq d \leq 1)$ and at a fixed level of significance α. As $F_\alpha \to \infty$, $RE[\hat{\beta}_1^{\text{PTE}}(d) : \tilde{\beta}_1] \to [1 - (1-d^2) + (1-d)^2\Delta^2]^{-1} = [d^2 + (1-d)^2\Delta^2]^{-1}$ which is the relative efficiency of $\hat{\beta}_1(d)$ compared to $\tilde{\beta}_1$. On the other hand, as $F_\alpha \to 0$, $RE[\hat{\beta}_1^{\text{PTE}}(d) : \tilde{\beta}_1] \to 1$. This means the relative efficiency of the PTE is the same as the unrestricted estimator, $\tilde{\beta}_1$. Note that under the null hypothesis, $\Delta^2 = 0$, the relative efficiency expression (5.3) equals

$$\Big[1 - (1-d^2) G_{3,\nu}\Big(\frac{1}{3}F_\alpha; 0\Big)\Big]^{-1} \geq 1, \tag{5.4}$$

which is the maximum value of the relative efficiency. Thus the relative efficiency function monotonically decreases crossing the 1-line for Δ^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

$$\Delta_*^2 = \frac{(1+d)}{\Big\{2 - (1+d)\frac{G_{5,\nu}(\frac{1}{5}F_\alpha; \Delta^2)}{G_{3,\nu}(\frac{1}{3}F_\alpha; \Delta^2)}\Big\}}, \tag{5.5}$$

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

$$RE\Big[\hat{\beta}_1^{\text{PTE}}(d) : \tilde{\beta}_1\Big] \lesseqqgtr 1 \quad \text{according as } \Delta_*^2 \lesseqqgtr \Delta^2. \tag{5.6}$$

Finally, as $\Delta^2 \to \infty$, RE$[\hat{\beta}_1^{\text{PTE}}(d) : \tilde{\beta}_1] \to 1$. Thus, the preliminary test estimator is more efficient than the unrestricted estimator whenever $\Delta^2 < \Delta_*^2$, otherwise $\tilde{\beta}_1$ is more efficient than PTE up to a moderate value of Δ^2. As for the relative efficiency of $\hat{\beta}_1^{\text{PTE}}(d)$ compared to $\hat{\beta}_1(d)$ we have

$$\text{RE}\Big[\hat{\beta}_1^{\text{PTE}}(d) : \hat{\beta}_1\Big] = \big[d^2 + (1-d)^2\Delta^2\big]\big[1 + g(\Delta^2)\big]^{-1} \tag{5.7}$$

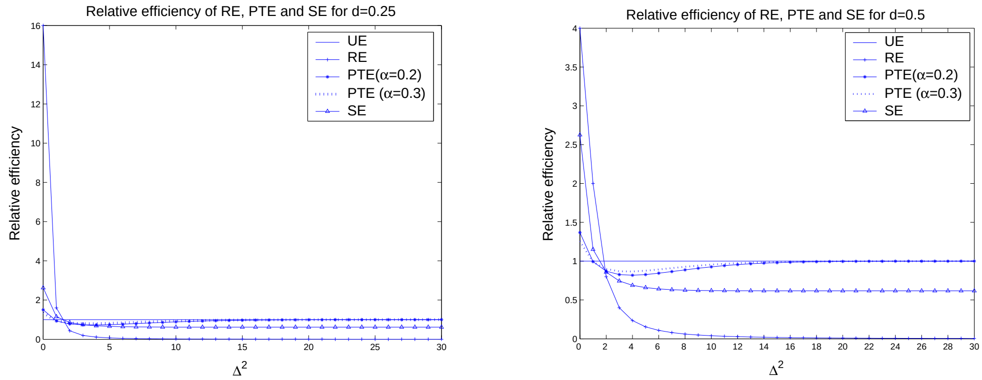

Figure 3: Graph of the relative efficiency of the RE, PTE and SE relative to UE

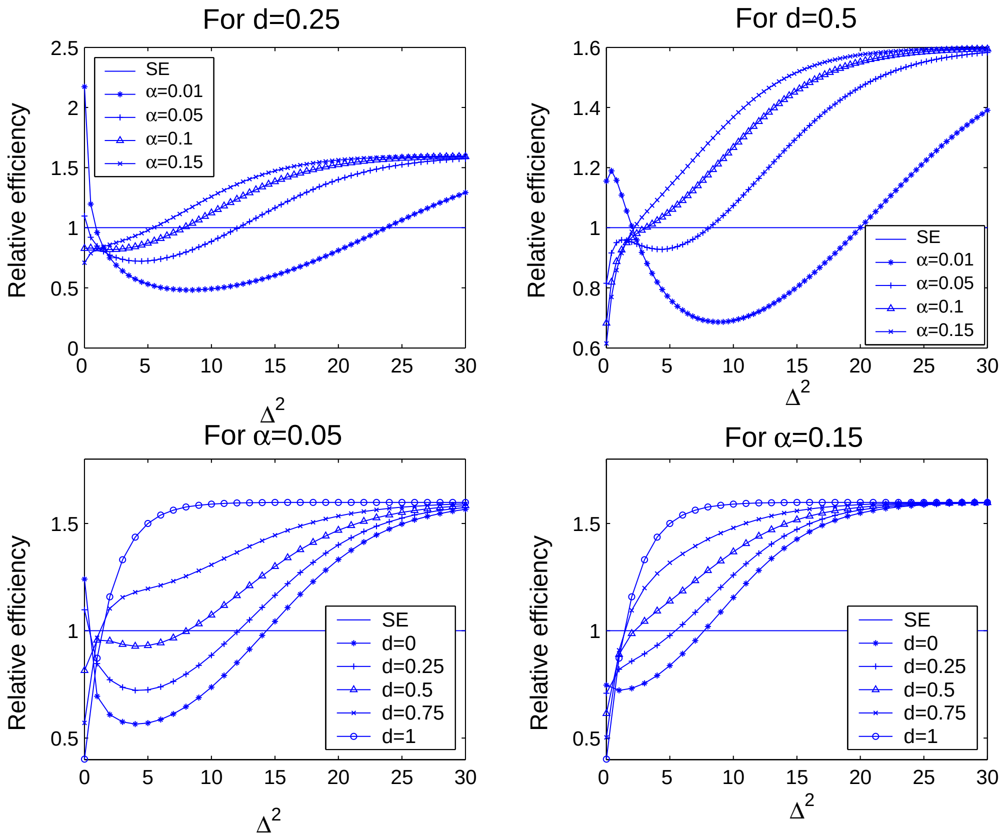

Figure 4: Graph of the relative efficiency of PTE relative to SE for selected d and α

$$\text{where} \quad g(\Delta^2) \;= (1-d)\Delta^2\Big\{2G_{3,\nu}\Big(\tfrac{1}{3}F_\alpha;\Delta^2\Big) - (1+d)G_{5,\nu}\Big(\tfrac{1}{5}F_\alpha;\Delta^2\Big)\Big\}$$
$$-(1+d^2)G_{3,\nu}\Big(\tfrac{1}{3}F_\alpha;\Delta^2\Big). \tag{5.8}$$

Under the null-hypothesis,

$$\mathrm{RE}[\hat{\beta}_1^{\mathrm{PTE}}(d) : \hat{\beta}_1(d)] = d^2\Big[1-(1-d^2)G_{3,\nu}\Big(\frac{1}{3}F_\alpha;0\Big)\Big]^{-1} \geq d^2. \tag{5.9}$$

At the same time we consider the result at (5.4). In combination, we obtain

$$d^2 \leq \mathrm{RE}[\hat{\beta}_1^{\mathrm{PTE}}(d) : \hat{\beta}_1(d)] \leq 1 \leq \mathrm{RE}[\hat{\beta}_1^{\mathrm{PTE}}(d) : \tilde{\beta}_1]. \tag{5.10}$$

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

$$\mathrm{RE}[\hat{\beta}_1^{\mathrm{PTE}}(d) : \hat{\beta}_1(d)] \begin{smallmatrix}\leq\\=\\>\end{smallmatrix} 1 \;\text{ according as} \tag{5.11}$$

$$\Delta^2 \begin{smallmatrix}\leq\\=\\>\end{smallmatrix} \frac{1+d}{1-d}\,\frac{\Big\{1-G_{3,\nu}\Big(\tfrac{1}{3}F_\alpha;\Delta^2\Big)\Big\}}{\Big\{1-2G_{3,\nu}\Big(\tfrac{1}{3}F_\alpha;\Delta^2\Big)-(1+d)G_{5,\nu}\Big(\tfrac{1}{5}F_\alpha;\Delta^2\Big)\Big\}}. \tag{5.12}$$

Finally, as $\Delta^2 \to \infty$, $\mathrm{RE}[\hat{\beta}_1^{\mathrm{PTE}}(d); \hat{\beta}_1(d)] \to 0$. Thus, except for a small interval around 0, $\hat{\beta}_1^{\mathrm{PTE}}(d)$ is more efficient than $\hat{\beta}_1(d)$.

Comparing SE against UE

The relative efficiency of $\hat{\beta}_1^{\mathrm{SE}}$ compared to $\tilde{\beta}_1$ is given by

$$\mathrm{RE}\big(\hat{\beta}_1^{\mathrm{SE}} : \tilde{\beta}_1\big) = \Big[1-\frac{2}{\pi}K_n^2\Big\{2e^{-\Delta^2/2}-1\Big\}\Big]^{-1}. \tag{5.13}$$

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

$$\mathrm{RE}\big(\hat{\beta}_1^{\mathrm{SE}} : \tilde{\beta}_1\big) = \Big[1-\frac{2}{\pi}K_n^2\Big]^{-1} \geq 1. \tag{5.14}$$

In general, $\mathrm{RE}(\hat{\beta}_1^{\mathrm{SE}} : \tilde{\beta}_1)$ decreases from $\big[1-\frac{2}{\pi}K_n^2\big]^{-1}$ at $\Delta^2 = 0$ and crosses the 1-line at $\Delta^2 = ln\,4$ and then goes to the minimum value $\Big[1+\frac{2}{\pi}K_n^2\Big]^{-1}$ as $\Delta^2 \to \infty$. Thus, the loss of efficiency of $\hat{\beta}_1^{\mathrm{SE}}$ relative to $\tilde{\beta}_1$ is $1-\Big[1+\frac{2}{\pi}K_n^2\Big]^{-1}$ while the gain in efficiency is $\Big[1-\frac{2}{\pi}K_n^2\Big]^{-1}$ respectively which is achieved at $\Delta^2 = 0$. Thus, for

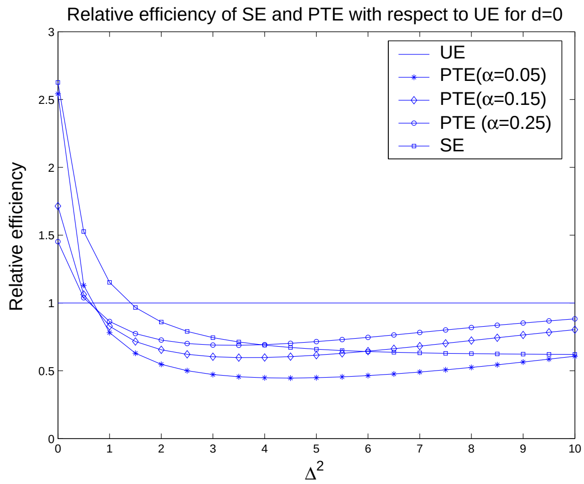

Figure 5: Graph of the relative efficiency of the SE and PTE relative to UE for different values of α.

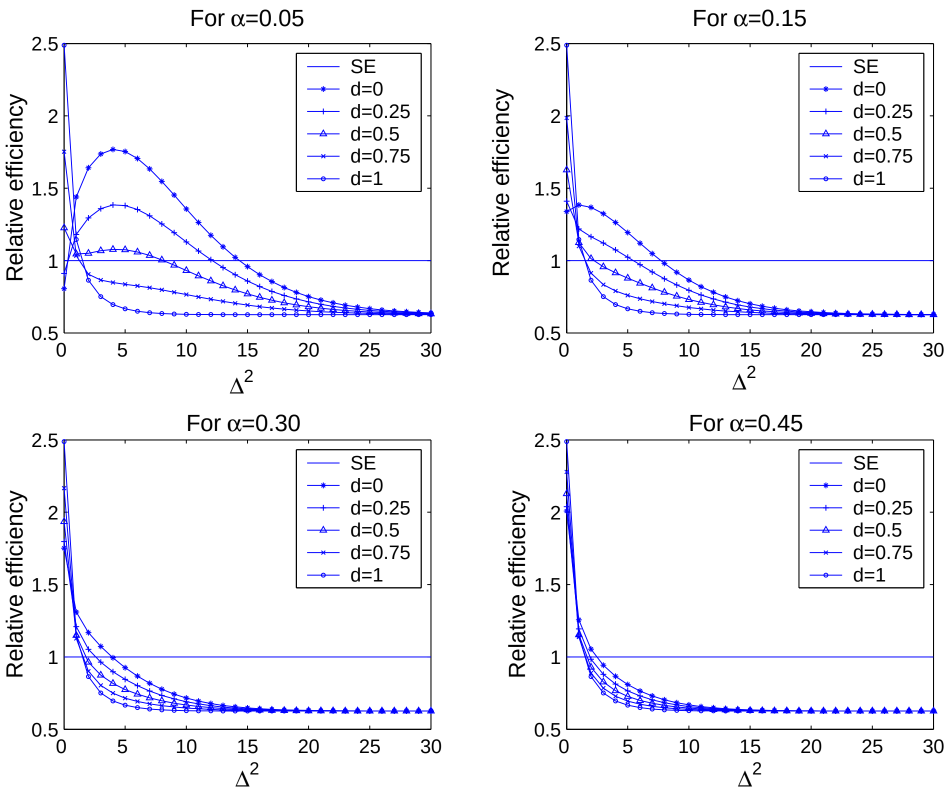

Figure 6: Graph of the relative efficiency of the SE relative to the PTE against Δ^2 for different values of d.

$\Delta^2 < ln\, 4$, $\hat{\beta}_1^{\rm SE}$ performs better than $\tilde{\beta}_1$, otherwise $\tilde{\beta}_1$ performs better. The property of $\hat{\beta}_1^{\rm SE}$ is similar to the preliminary test estimator but does not depend on the level of significance. As $\Delta^2 \to \infty$ the relative efficiency of PTE with respect to UE approaches to 1 and that of the SE with respect to UE approaches to $\left[1 + \frac{2}{\pi} K_n^2\right]^{-1}$.

Comparing SE against PTE relative to UE

To compare the relative performances of the SE and the PTE, first note that the SE is superior to the PTE when the null hypothesis is true and the level of significance, α is not too small. This is regardless of the value of the coefficient of distrust, d. However, as the value of Δ^2 increases and or α decreases the relative efficiency picture changes.

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 Δ^2 near 0. Then it slides down rapidly, and passes the curve of the unit relative efficiency (of the UE) from above. The relative efficiency of SE with respect to UE becomes constant after some moderate value of Δ^2. However, the relative efficiency of PTE approaches to 1-line as $\Delta^2 \to \infty$. For larger α the relative efficiency of PTE approaches the 1-line for relatively smaller values of Δ^2. The top two graphs in Figure 4 demonstrate the behavior of the relative efficiency curves for different values of α when $d = 0.25$ and $d = 0.50$ respectively. It is clear that as the value of α increases, the relative efficiency of the PTE with respect to the SE grows higher for some moderate value of Δ^2. When the value of α is smaller the relative efficiency of the PTE is lower, and hence the SE over performs the PTE for moderate value of Δ^2.

From figure 5, it is clear that for $\Delta^2 = 0$ the relative efficiency of both the PTE and SE relative to the UE are greater than one. As Δ^2 grows larger, the relative efficiency of both estimators decrease, but at different rates. Initially the relative efficiency of the SE relative to the UE is greater than that of the PTE. But for some larger value of Δ^2 it becomes less than that of the PTE.

There is no uniform domination of the SE over the PTE for all Δ^2 and every α. Clearly, the superior performance of the SE relative to the PTE depends on the value of Δ^2. When the value of Δ^2 is in the neighborhood of 0, the SE over performs the PTE for every value of Δ^2 close to zero. But, the value of Δ^2 is near 0 (that is, $(\beta_1 - \beta_{10}) \to 0$) 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 β_1 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 β_1 to make Δ^2 close to 0 or reasonably small, and hence the SE would normally be a preferred option over the PTE.

Comparing efficiency of SE relative to PTE

In figure 6, the maximum relative efficiency of the SE relative to PTE attained for $\Delta^2 = 0$ and $d = 1$, regardless of the value of α. At $\Delta^2 = 0$, as the coefficient of distrust, d decreases, the relative efficiency of SE also decreases, and it decreases below 1 for $d = 0$. Starting from some moderate value of Δ^2, relative efficiency of SE becomes less than 1 and converges to a stable value, below one, as $\Delta^2 \to \infty$. Except for $\Delta^2 = 0$ and near 0 the relative efficiency of SE is always higher for smaller values of d than larger values of d, before converging to a stable value. The difference between the relative efficiencies of the SE for different values of d is higher for lower value of α then it's higher values. As α increases this difference decreases. Moreover, as α increases, the relative efficiency of the SE also increases for $\Delta^2 = 0$ or near 0.

6 Concluding Remarks

The UE is based on the sample data alone and it is the only unbiased estimator among the four estimators 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 they are judged based on the mse criterion. The performance of the biased estimators depend on the value of the departure parameter Δ. 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 α is not too high. As α increases, the performance of the PTE improves when Δ is not too close to zero. At a lower level of significance, the SE performs better than the PTE more often and over a wider range of values of Δ. When the value of Δ 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 β_1 from the prior knowledge or experience that is not too far from its true value, the SE would be the best choice as an 'improved' estimator of the slope.

References

[1] Ahmed, S.E. and Saleh, A.K.Md.E. (1989). Pooling multivariate data. *Jou. Statist. Comput. Simul.* **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.,* **15**, 190–204.

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

[4] Chiou, P. and Saleh, A.K.Md.E. (2002). Preliminary test confidence sets for the mean of a multivariate normal distribution. *Journal of Propagation in Probability and Statistics*, **2**, 177-189.

[5] Han, C.P. and Bancroft, T.A. (1968). On pooling means when variance is unknown. *Jou. Amer. Statist. Assoc.*, **63**, 1333-1342.

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

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

[8] Khan, S. and Saleh, A.K.Md.E. (2001). On the comparison of the pre-test and shrinkage estimators for the univariate normal mean. *Statistical Papers*, **42**(4), 451-473.

[9] 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.

[10] 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.

[11] Maatta, J.M. and Casella, G. (1990). Developments in decision-theoretic variance estimation, *Statistical Sciences*, **5**, 90-101.

[12] 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.

[13] Saleh, A.K.Md.E. and Sen, P.K. (1978). Nonparametric estimation of location parameter after a preliminary test on regression. *Annals of Statistics*, **6**, 154-168.

[14] 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.* **43**, 1481–1490.

[15] 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.

[16] Stein, C. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.*, **9**, 1135–1151.