Estimation of the intercept parameter for linear regression model with uncertain non-sample prior information

Shahjahan Khan & **Zahirul Hoque** *
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

This paper considers alternative estimators of the intercept parameter of the linear regression model with normal error when uncertain non-sample prior information about the value of the slope parameter is available. The maximum likelihood, restricted, preliminary test and shrinkage estimators are considered. Based on their quadratic biases and mean square errors the relative performances of the estimators are investigated. Both analytical and graphical methods are explored. None of the estimators is found to be uniformly dominating the others. However, if the non-sample prior information regarding the value of the slope is not too far from its true value, the shrinkage estimator of the intercept parameter dominates the rest of the estimators.

Keywords and Phrases: Regression model; uncertain non-sample prior information; maximum likelihood, restricted, preliminary test and shrinkage estimators; bias, mean square error and relative efficiency.
AMS 2000 Mathematics Subject Classification: Primary 62F30, Secondary 62H12 and 62F10.

1 Introduction

Estimation of the slope and intercept parameter of linear regression model is a widely used statistical procedure. The use of the maximum likelihood estimator (mle) or least square estimator (lse) is very common in the literature. These estimators are

*On leave from the Department of Statistics, University of Chittagong, Bangladesh

solely based on the sample information, and disregard any other kind of non-sample prior information in their definition. The notion of inclusion of non-sample prior 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 information based unbiased estimators when judged by criteria such as the mean square error and squared error loss function. In many studies the researchers estimate the slope parameter of the regression model. However, the estimation of the intercept parameter is more difficult than that of the slope parameter. This is because the estimator of the slope parameter is required in the estimation of the intercept parameter. Khan et al (2002) studied the improved estimation of the slope parameter for the linear regression model. They introduced the coefficient of distrust on the belief of the null hypothesis, and incorporated this coefficient in the definition and analysis of the estimators. In this paper we use the unrestricted estimator of the slope parameter to define the unrestricted estimator (UE), restricted estimator (RE), preliminary test estimator (PTE) and shrinkage estimator (SE) of the intercept parameter. Statistical properties of these estimators are investigated both analytically and graphically.

A large number of studies have been conducted 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 (shrinkage) estimator for multivariate normal population that dominates the usual mle under the squared 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 intercept parameter of linear regression model has not been investigated.

Consider a linear regression model with slope and intercept parameters β and

θ respectively. Assume that *uncertain non-sample prior information* on the value of the slope parameter, β 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, $H_0: \ \beta = 0$ 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 intercept θ. Following Khan et al (2002) 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 mle. First we define the *unrestricted* mle of the unknown intercept θ 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 $H_0: \ \beta = 0$ against $H_a: \ \beta \neq 0$. 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 intercept.

Like the mle of the slope parameter the mle of the intercept parameter is unbiased. Here we attempt to search for an estimator of the intercept 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 than the popular mle. 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 prior information regarding the value of β is not too far from its true value. Usually it is expected that the non-sample prior information will not be too far from the true value.

The next section provides the specification of the model and definition of the unrestricted estimators of θ, 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. The quadratic biases of three biased estimators are analyzed in section 5. Comparative study of the relative efficiency of the estimators are included in section 6. Some concluding remarks are given in section 7.

2 The Model and Some Preliminaries

The n independently and identically distributed responses from a linear regression model can be expressed by the equation

$$\boldsymbol{y} = \theta \mathbf{1}_n + \beta \boldsymbol{x} + \boldsymbol{e} \tag{2.1}$$

where $\boldsymbol{y}$ and $\boldsymbol{x}$ are the column vectors of response and explanatory variables respectively, $\mathbf{1}_n = (1, \ldots, 1)'$ - a vector of n-tuple of 1's, θ and β 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 $E(\boldsymbol{e}) = \mathbf{0}$ and $E(\boldsymbol{e}\boldsymbol{e}') = \sigma^2 I_n$ where σ^2 is the variance of each of the error component in $\boldsymbol{e}$ and I_n is the identity matrix of order n. The *unrestricted* mle of the slope β and intercept θ are given by

$$\tilde{\beta} = (\boldsymbol{x}'\boldsymbol{x})^{-1}\boldsymbol{x}'\boldsymbol{y} \quad \text{and} \quad \tilde{\theta} = \bar{y} - \tilde{\beta}\bar{x} \tag{2.2}$$

where, $\bar{x} = \frac{1}{n}\sum_{j=1}^{n} x_j$ and $\bar{y} = \frac{1}{n}\sum_{j=1}^{n} y_j$. It is well known that, the sampling distribution of the mle of θ and β are normal with respective means, $E(\tilde{\theta}) = \theta$, $E(\tilde{\beta}) = \beta$ and variances, $E(\tilde{\theta} - \theta)^2 = \sigma^2 H$ and $E(\tilde{\beta} - \beta)^2 = \frac{\sigma^2}{S_{xx}}$ in which $S_{xx} = \sum_{j=1}^{n}(x_j - \bar{x})^2$ and $H = \left\{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right\}$. Therefore, $\tilde{\theta}$ is unbiased for θ. Here, the bias and the mse of $\tilde{\theta}$ are given by $B_1(\tilde{\theta}) = 0$ and $M_1(\tilde{\theta}) = \sigma^2 H$ respectively. The mle of σ^2 is

$$S_n^{*2} = \frac{1}{n}(\boldsymbol{y} - \hat{\boldsymbol{y}})'(\boldsymbol{y} - \hat{\boldsymbol{y}}) \tag{2.3}$$

where $\hat{\boldsymbol{y}} = \tilde{\theta}\mathbf{1}_n + \tilde{\beta}\boldsymbol{x}$. This estimator is biased for σ^2. However,

$$S_n^2 = \frac{1}{n-2}(\boldsymbol{y} - \hat{\boldsymbol{y}})'(\boldsymbol{y} - \hat{\boldsymbol{y}}) \tag{2.4}$$

is unbiased for σ^2. The above unbiased estimator of σ^2 has a scaled χ^2 distribution with d.f. $\nu = (n-2)$. It can be easily shown that the standard error of $\tilde{\beta}$ is $\frac{S_n}{\sqrt{S_{xx}}}$.

To remove the uncertainty from the non-sample prior information, we perform an appropriate statistical test on the null hypothesis, $H_0 : \beta = 0$ against the alternative hypothesis, $H_a : \beta \neq 0$. In this study, the appropriate test is the likelihood ratio test (LRT) and the test statistic is given by

$$\mathcal{L}_\nu = \frac{S_{xx}^{\frac{1}{2}}\tilde{\beta}}{S_n}. \tag{2.5}$$

Under the H_a, the above statistic $\mathcal{L}_\nu$, follows a non-central Student-t distribution with $\nu = (n-2)$ degrees of freedom (d.f.), and non-centrality parameter $\Delta^2 = \frac{S_{xx}\beta^2}{\sigma^2}$.

As per the relationship between the non-central Student-t and F distributions under the H_a, $\mathcal{L}_\nu^2$ follows the non-central F-distribution with $(1,\ \nu)$ d.f. and same non-centrality parameter. Under the null-hypothesis, $\mathcal{L}_\nu$ and $\mathcal{L}_\nu^2$ follow a central Student-t and F-distributions respectively with appropriate degrees of freedom. We use this test statistic for defining the PTE, and the shrinkage estimator by following the preliminary test approach to the shrinkage estimation.

3 Proposed Estimators of the Intercept

Consider a linear combination of $\tilde{\theta} = \bar{y} - \tilde{\beta}\bar{x}$, mle of θ under H_a and $\hat{\theta} = \bar{y}$, mle of θ under H_0 as

$$\hat{\theta}^{\mathrm{RE}}(d) = d\tilde{\theta} + (1-d)\hat{\theta}, \quad 0 \leq d \leq 1. \tag{3.1}$$

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

$$E[\hat{\theta}^{\mathrm{RE}}(d)] = \theta + (1-d)\beta\bar{x} \quad \text{and} \quad \mathrm{MSE}[\hat{\theta}^{\mathrm{RE}}(d)] = d^2 H + (1-d)^2\frac{\sigma^2}{n} \tag{3.2}$$

respectively.

Following Bancroft (1944) we define a preliminary test estimator of the intercept parameter as

$$\begin{aligned} \hat{\theta}^{\mathrm{PTE}}(d) &= \hat{\theta}^{\mathrm{RE}}(d)I(F < F_\alpha) + \tilde{\theta}I(F \geq F_\alpha) \\ &= \tilde{\theta} + \tilde{\beta}\bar{x}(1-d)I(F < F_\alpha) \end{aligned} \tag{3.3}$$

where $I(A)$ is an indicator function of the set A and F_α is the $(1-\alpha)^{th}$ quantile of a central F-distribution with $(1,\ \nu)$ degrees of freedom. For $d = 0$, the above preliminary test estimator becomes

$$\hat{\theta}^{\mathrm{PTE}}(d = 0) = \tilde{\theta} + \tilde{\beta}\bar{x}I(F < F_\alpha). \tag{3.4}$$

The PTE is a discontinuous function of $\hat{\theta}^{\mathrm{RE}}(d)$ and $\tilde{\theta}$. Also, it depends on the choice of the level of significance α of the test. To overcome the above limitations of the PTE, we define the shrinkage estimator (SE) of θ as

$$\hat{\theta}^{\mathrm{SE}}(d) = \tilde{\theta} + (1-d)\tilde{\beta}\bar{x}\frac{cS_n}{\sqrt{S_{xx}}|\tilde{\beta}|}. \tag{3.5}$$

Note that in this estimator, c is the shrinkage constant, a function of n. Unlike the preliminary test estimator, the shrinkage estimator does not depend on the choice of the level of significance.

4 Some Statistical Properties

The bias and the mean square error (mse) functions of RE, PTE and SE are derived in this section. For the RE the bias and the mse are obtained as

$$B_2[\hat{\theta}^{\mathrm{RE}}(d)] = \frac{\bar{x}\sigma}{\sqrt{S_{xx}}}(1-d)\Delta \tag{4.1}$$

$$\text{and} \quad M_2[\hat{\theta}^{\mathrm{RE}}(d)] = \sigma^2\left[d^2H + (1-d)^2\frac{\bar{x}^2\Delta^2}{S_{xx}}\right] \tag{4.2}$$

respectively, where Δ^2 is the *departure constant* from the null-hypothesis. Under the null hypothesis the value of this constant is 0 while under the alternative hypothesis it takes a positive value. The value of this constant plays an important role on the behavior of the biased estimators. The relative efficiency of the estimators change with the change in the value of this departure constant. We study this feature in a greater detail in the remainder of this paper .

4.1 The Bias and the MSE of the PTE

By definition, the bias of the PTE is given by

$$\begin{aligned} E\Big[\hat{\theta}^{\mathrm{PTE}}(d)-\theta\Big] &= E\Big[(\tilde{\theta}-\theta)+(1-d)\tilde{\beta}\bar{x}I(F<F_\theta)\Big] \qquad (4.3)\\ &= (1-d)\bar{x}\frac{\sigma}{\sqrt{S_{xx}}}E\left[\frac{\sqrt{S_{xx}}\tilde{\beta}}{\sigma}I\Big(\frac{S_{xx}\tilde{\beta}^2}{S_n^2}<F_\alpha\Big)\right]. \end{aligned}$$

Note $Z = \frac{\sqrt{S_{xx}}\tilde{\beta}}{\sigma}$ is distributed as $N(\Delta, 1)$, where $\Delta = \frac{\sqrt{S_{xx}}\beta}{\sigma}$, and $\frac{\nu S_n^2 S_{xx}}{\sigma^2}$ is distributed (independently) as a central chi-square variable with ν degrees of freedom. Evaluating the expression in (4.3) the bias function of $\hat{\beta}^{\mathrm{PTE}}(d)$ is found to be

$$B_3[\hat{\theta}^{\mathrm{PTE}}(d)] = (1-d)\bar{x}\beta G_{3,\nu}\Big(\frac{1}{3}F_\alpha;\Delta^2\Big), \tag{4.4}$$

where $G_{n_1,n_2}(\cdot;\Delta^2)$ is the c.d.f. of a non-central F-distribution with (n_1, n_2) degrees of freedom and non-centrality parameter Δ^2. 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*

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

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

$$\begin{aligned} M_3\left[\hat{\theta}^{\mathrm{PTE}}(d)\right] &= E\left[\hat{\theta}^{\mathrm{PTE}}(d) - \theta\right]^2 \\ &= E(\tilde{\theta}-\theta)^2 + (1-d)^2 E[\tilde{\beta}^2\bar{x}^2 I(F < F_\alpha)] \\ &\quad +2(1-d)E[(\tilde{\theta}-\theta)\tilde{\beta}\bar{x} I(F < F_\alpha)] \\ &= \sigma^2 H + (1-d)^2\bar{x}^2 E[\tilde{\beta}^2 I(F < F_\alpha)] \\ &\quad +2\bar{x}(1-d)E[\tilde{\beta}(\tilde{\theta}-\theta)I(F < F_\alpha)]. \end{aligned} \tag{4.7}$$

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,

$$\begin{aligned} M_3[\hat{\beta}^{\mathrm{PTE}}(d)] &= \sigma^2 H + \frac{\sigma^2\bar{x}^2}{S_{xx}}\left[\Delta^2\left\{2(1-d)G_{3,v}\left(\frac{1}{3}F_\alpha;\Delta^2\right)\right.\right. \\ &\quad \left.\left.-(1-d^2)G_{5,v}\left(\frac{1}{5}F_\alpha;\Delta^2\right)\right\} - (1-d^2)G_{3,v}\left(\frac{1}{3}F_\alpha;\Delta^2\right)\right] \end{aligned} \tag{4.8}$$

4.2 The Bias and MSE of the SE

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

$$\begin{aligned} B_4[\hat{\theta}^{\mathrm{SE}}(d)] &= (1-d)E\left[\tilde{\beta}\bar{x}\frac{cS_n}{\sqrt{S_{xx}}|\tilde{\beta}|}\right] \\ &= (1-d)\frac{c\bar{x}}{\sqrt{S_{xx}}}E[S_n]E\left[\frac{Z}{|Z|}\right] \end{aligned} \tag{4.9}$$

where $Z = \frac{\sqrt{S_{xx}}\tilde{\beta}}{\sigma} \sim \mathcal{N}(\Delta, 1)$. 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*

$$E\left[\frac{Z}{|Z|}\right] = 1 - 2\Phi(-\Delta) \tag{4.10}$$

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, $QB_4[\hat{\theta}^{\rm SE}(d)]$ is obtained as

$$QB_4[\hat{\theta}^{\rm SE}(d)] = (1-d)^2 \frac{c^2\bar{x}^2\sigma^2}{S_{xx}} K_\nu^2 \{2\Phi(\Delta)-1\}^2 \tag{4.11}$$

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

As $\Delta^2 \to 0$, $QB_4[\hat{\theta}^{\rm SE}(d)] \to 0$ and as $\Delta^2 \to \infty$, $QB_4[\hat{\theta}^{\rm SE}(d)] \to (1-d)^2 \frac{c^2\bar{x}^2\sigma^2}{S_{xx}} K_\nu^2$, 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 $\hat{\theta}^{\rm SE}(d)$ we consider

$$\begin{aligned} E[\hat{\theta}^{\rm SE}(d)-\theta]^2 &= E[\tilde{\theta}-\theta]^2 + (1-d)^2 \frac{c^2\bar{x}^2}{S_{xx}} E\left[\frac{S_n^2\tilde{\beta}^2}{|\tilde{\beta}|^2}\right] \\ &\quad + \frac{2(1-d)c\bar{x}}{\sqrt{S_{xx}}} E\left[(\tilde{\theta}-\theta)\frac{S_n\tilde{\beta}}{|\tilde{\beta}|}\right] \\ &= \sigma^2 H + (1-d)^2 \frac{c^2\bar{x}^2\sigma^2}{S_{xx}} \\ &\quad -2c(1-d)\frac{\bar{x}^2\sigma^2 K_\nu}{S_{xx}} \left\{E(|Z|) - \Delta E\left[\frac{Z}{|Z|}\right]\right\}. \end{aligned} \tag{4.12}$$

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*

$$E(|Z|) = \sqrt{\frac{2}{\pi}} e^{-\Delta^2/2} + \Delta\{2\Phi(\Delta)-1\} \tag{4.13}$$

where $\Phi(\cdot)$ is the c.d.f. of the standard normal variable.. See Khan et al (2002) for the proof of the above theorem.

Therefore, the mse of $\hat{\theta}^{\rm SE}$(d) is given by

$$M_4[\hat{\theta}^{\rm SE}(d)] = \sigma^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \left\{1 + (1-d)^2c^2 - 2(1-d)cK_\nu \sqrt{\frac{2}{\pi}} e^{-\frac{\Delta^2}{2}}\right\}\right]. \tag{4.14}$$

The value of c which minimizes (4.14) depends on Δ^2 and is given by

$$c^* = (1-d)^{-1} K_\nu \sqrt{\frac{2}{\pi}} e^{-\Delta^2/2}. \tag{4.15}$$

To make c^* independent of Δ^2, we choose $c^0 = (1-d)^{-1}\sqrt{\frac{2}{\pi}}K_\nu$. Thus, optimum $M_4[\hat{\theta}^{\rm SE}(d)]$ reduces to

$$M_4[\hat{\theta}^{\rm SE}(d)] = \sigma^2\left[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\left\{1 + \frac{2}{\pi}K_\nu^2\Big(1 - 2e^{-\frac{\Delta^2}{2}}\Big)\right\}\right]. \tag{4.16}$$

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

5 Study of Bias

Here we compare the three biased estimators by analyzing their quadratic biases analytically and graphically. Also, we propose the best performed estimator, under certain condition.

The quadratic bias of the RE, PTE and SE are respectively given by

$$QB_2[\hat{\theta}^{\rm RE}(d)] = \frac{\bar{x}^2\sigma^2}{S_{xx}}(1-d)^2\Delta^2 \tag{5.1}$$

$$QB_3[\hat{\theta}^{\rm PTE}(d)] = \frac{\bar{x}^2\sigma^2}{S_{xx}}(1-d)^2\Delta^2\Big\{G_{3,\nu}\Big(\frac{1}{3}F_\alpha;\Delta^2\Big)\Big\}^2 \tag{5.2}$$

$$QB_4[\hat{\theta}^{\rm SE}(d)] = \frac{\sigma^2\bar{x}^2}{S_{xx}}K_\nu^2\{2\Phi(\Delta)-1\}^2. \tag{5.3}$$

Note that, in the derivation of $QB_4[\hat{\theta}^{\rm SE}(\rm d)]$, the optimal value of the shrinkage constant has been used.

Under the null-hypothesis, $\Delta^2 = 0$, and hence $QB_2[\hat{\theta}^{\rm RE}(d)] = QB_3[\hat{\theta}^{\rm PTE}(d)] = QB_4[\hat{\theta}^{\rm SE}(d)] = 0$ for all d and α. It is observed that as $\Delta^2 \to \infty$, $QB_2[\hat{\theta}^{\rm RE}(d)] \to \infty$ except for $d = 1$; $QB_3[\hat{\theta}^{\rm PTE}(d)] \to 0$ for all α and d; and $QB_4[\hat{\theta}^{\rm SE}] \to \frac{\bar{x}^2\sigma^2}{S_{xx}}K_\nu^2$, a constant that does not depend on d. Therefore, in terms of quadratic bias, RE is uniformly dominated by both the PTE and SE regardless of the value of d. Also, 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 and the degree of distrust d. 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 a reasonable value of Δ^2. If there is a complete distrust on the null hypothesis, the quadratic bias of the RE and PTE become 0 for any α and Δ^2, while that of the SE remains greater than 0 except for $\Delta^2 = 0$. As the prior information is usually obtained from previous studies or expert knowledge, in practice, the chance of the non-centrality parameter to be very large is really slim and α is usually preferred to be reasonably small. Also, the quadratic bias of the

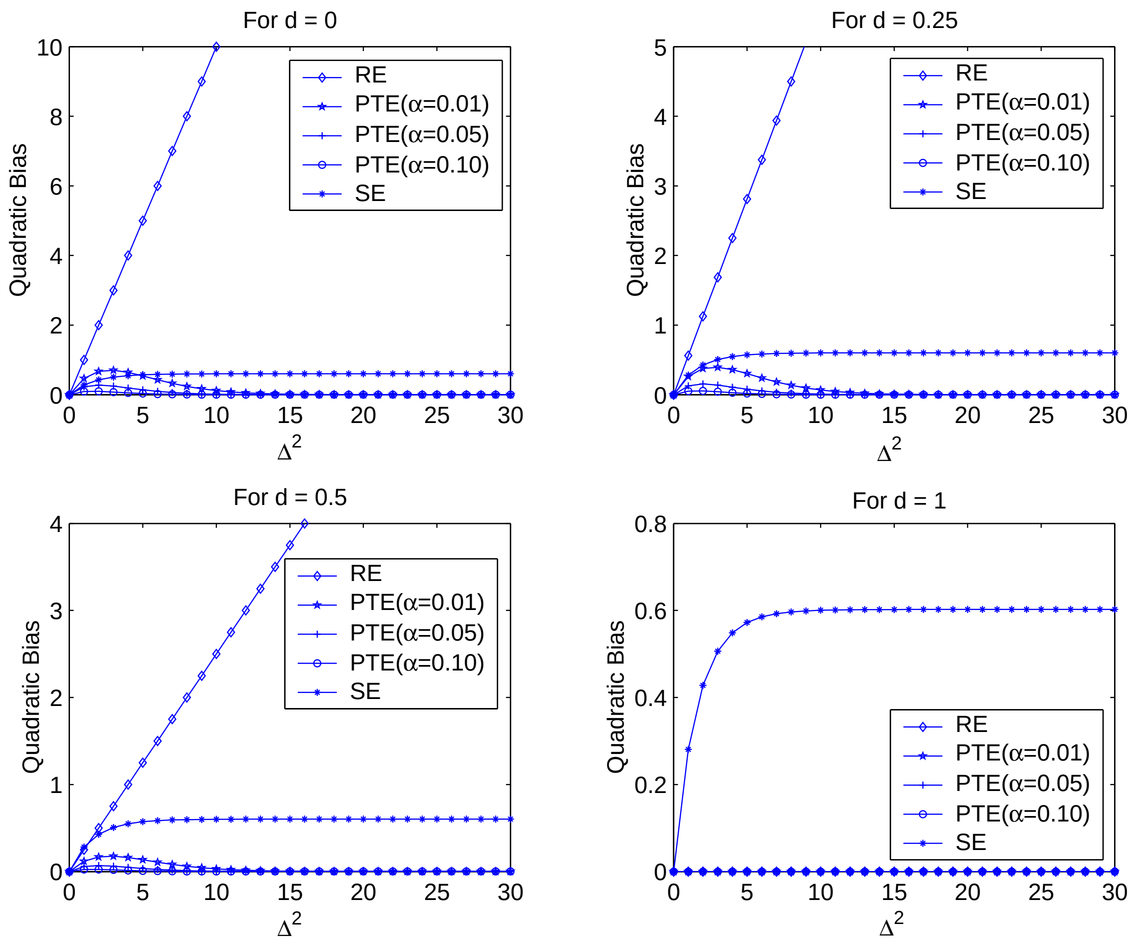

Figure 1: Graph of the quadratic bias of the RE, PTE and SE against Δ^2

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. Figure 1 is the graph of the quadratic bias of the RE, PTE and SE.

6 Study of the mean square error function

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.

6.1 Comparing RE against UE

The relative efficiency function of the RE relative to the UE is

$$\mathrm{RE}[\hat{\theta}^{\mathrm{RE}}(d) : \tilde{\theta}] = H \left[d^2 H + (1-d)^2 \frac{\bar{x}^2}{S_{xx}} \Delta^2 \right]^{-1}. \tag{6.1}$$

The relative efficiency function of the RE relative to the UE takes its highest possible value at $\Delta^2 = 0$ for $d = 0$. As Δ^2 increases, the relative efficiency function decreases for all d. It crosses the 1-line at some value of Δ^2 near zero, and finally for some moderate to large value of Δ^2 it approaches to 0. But for $d = 1$ the RE and UE are equally efficient regardless of the value of Δ^2.

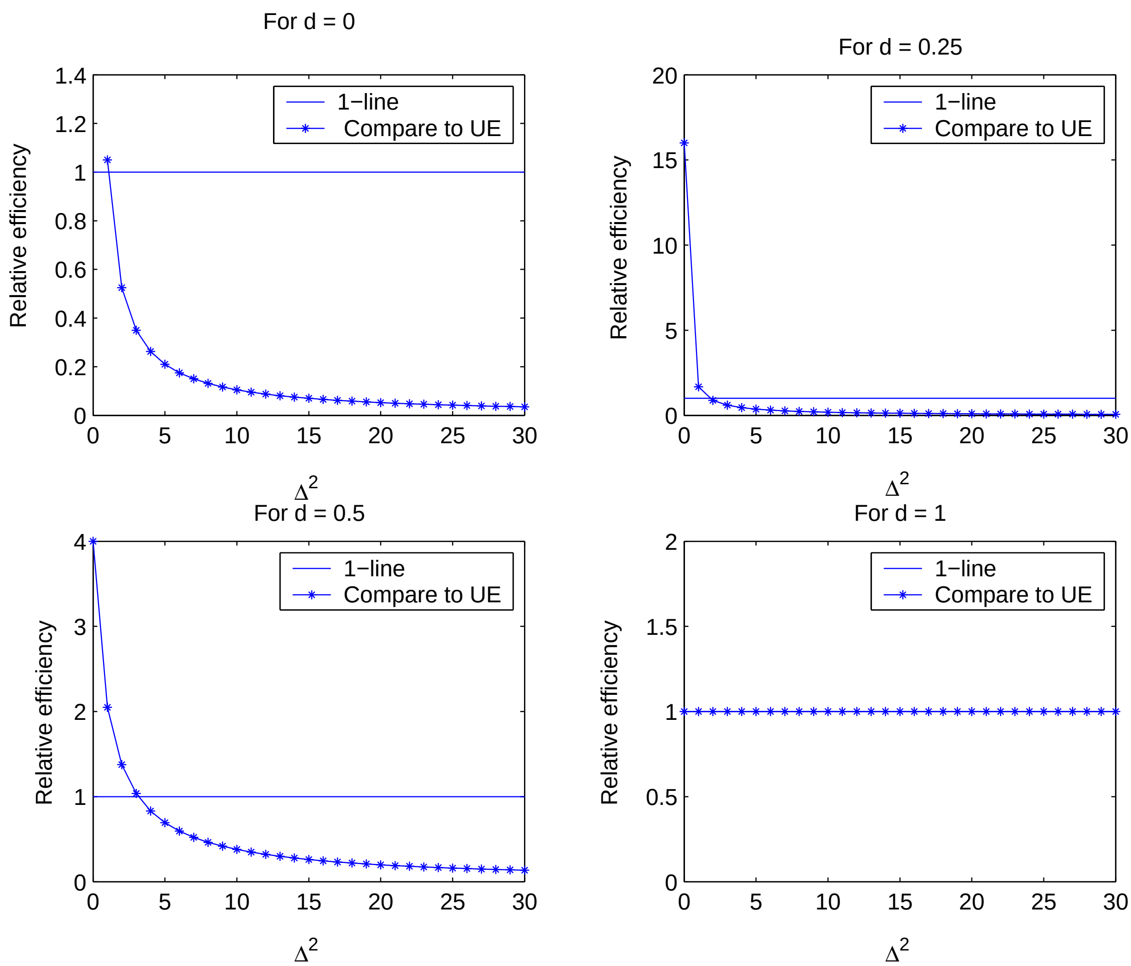

Figure 2: Graph of the relative efficiency of RE relative to UE against Δ^2.

From the expression in (6.1) we draw the following conclusions.

(i) Under H_0 $\Delta^2 = 0$, and hence $\text{RE}[\hat{\theta}^{\text{RE}}(d) : \tilde{\theta}] = d^{-2} \geq 1$. When $d = 0$, the relative efficiency function of the RE grows unboundedly. As d grows larger from 0 the relative efficiency decreases, and finally reaches to the 1-line for $d = 1$. Therefore under H_0, the RE is a better choice than the UE.

(ii) As Δ^2 grows larger, the relative efficiency function grows smaller, and finally as $\Delta^2 \to \infty$, $\text{RE}[\hat{\theta}^{\text{RE}}(d); \tilde{\theta}] \to 0$, except for $d = 1$. As $d \to 1$, $\text{RE}[\hat{\theta}^{\text{RE}}(d); \tilde{\theta}] \to 1$ from below regardless of the value of Δ^2. Therefore for very large values of Δ^2, the UE is a better choice than the RE.

In general, the relative efficiency of the RE relative to the UE is a decreasing function of Δ^2 with it's maximum value $d^{-2}(\geq 1)$ at $\Delta^2 = 0$ and minimum value 0 at $\Delta^2 = \infty$, unless $d = 1$. The relative efficiency of the RE equals 1 at $\Delta^2 = H\frac{(1+d)S_{xx}}{(1-d)\bar{x}^2}$. Thus, if $\Delta^2 \in \left[0, H\frac{(1+d)S_{xx}}{(1-d)\bar{x}^2}\right]$, the RE is more efficient than the UE, otherwise the reverse is true. However, in practice the non-sample prior information is usually obtained from some previous experience or expert knowledge, and hence it is very unlikely for Δ^2 to be very large. Therefore for $\Delta^2 = 0$ or near 0 the restricted estimator is a better choice than the unrestricted estimator.

6.2 Comparing PTE against UE and RE

The relative efficiency of the PTE relative to the UE and RE are

$$\mathrm{RE}\Big[\hat{\theta}^{\mathrm{PTE}}(d) : \tilde{\theta}\Big] = H\left[H + \frac{\bar{x}^2\sigma^2}{S_{xx}} g(\Delta^2)\right]^{-1} \tag{6.2}$$

$$\mathrm{RE}\Big[\hat{\theta}^{\mathrm{PTE}}(d) : \hat{\theta}^{\mathrm{RE}}(d)\Big] = \left[d^2 H + (1-d)^2\Delta^2\frac{\bar{x}^2}{S_{xx}}\right]\left[H + \frac{\bar{x}^2}{S_{xx}} g(\Delta^2)\right]^{-1} \tag{6.3}$$

respectively, where

$$\begin{aligned} g(\Delta^2) &= \Delta^2\Big\{2(1-d)G_{3,v}\Big(\frac{1}{3}F_\alpha;\Delta^2\Big) - (1-d^2)G_{5,v}\Big(\frac{1}{5}F_\alpha;\Delta^2\Big)\Big\} \\ &\quad -(1-d^2)G_{3,v}\Big(\frac{1}{3}F_\alpha;\Delta^2\Big). \end{aligned} \tag{6.4}$$

From the expressions in (6.2) and (6.3) we draw the following conclusions.
i) Under H_0 $\Delta^2 = 0$, and the relative efficiency functions become

$$\mathrm{RE}\Big[\hat{\theta}^{\mathrm{PTE}}(d) : \tilde{\theta}\Big] = H\left[H - \frac{\bar{x}^2\sigma^2}{S_{xx}}(1-d^2)G_{3,v}\Big(\frac{1}{3}F_\alpha;0\Big)\right]^{-1} \tag{6.5}$$

and $$\mathrm{RE}\Big[\hat{\theta}^{\mathrm{PTE}}(d) : \hat{\theta}^{\mathrm{RE}}(d)\Big] = d^2 H\left[H - \frac{\sigma^2\bar{x}^2}{S_{xx}}(1-d^2)G_{3,\nu}\Big(\frac{1}{3}F_\alpha;0\Big)\right]^{-1}. \tag{6.6}$$

Therefore, for any fixed d (< 1), the maximum relative efficiency of the PTE relative to the UE attains at $\Delta^2 = 0$, while the minimum relative efficiency of the PTE relative to the RE attains at $\Delta^2 = 0$. As d grows larger the maximum relative efficiency of the PTE relative to the UE decreases, while the minimum relative efficiency of the PTE relative to the RE increases. For $d = 1$, the efficiency of the PTE, RE and UE are same regardless of the values of α and Δ^2.

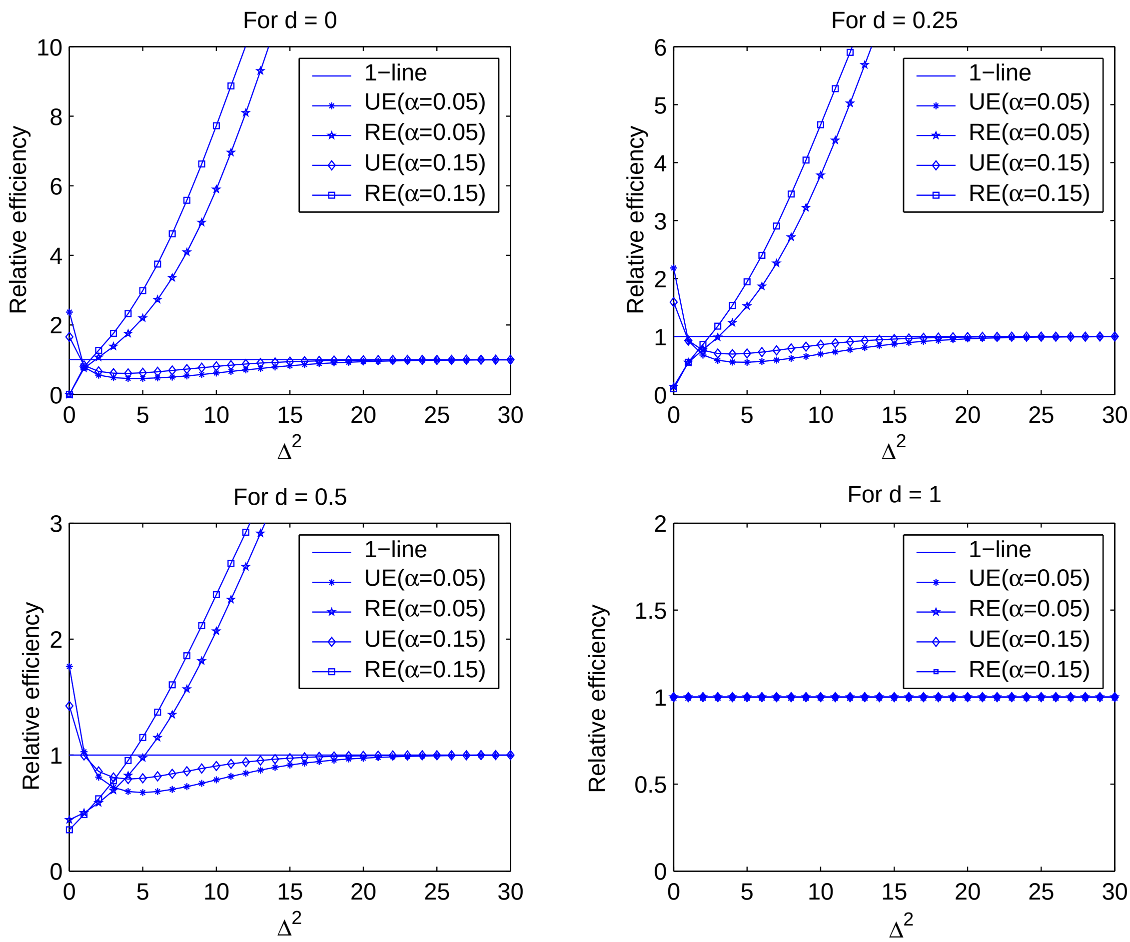

Figure 3: Graph of the relative efficiency of PTE relative to UE and RE against Δ^2.

As Δ^2 grows up, the relative efficiency of the PTE relative to the UE goes down and crosses the 1-line at

$$\Delta_*^2 = \frac{(1+d)G_{3,\nu}(\frac{1}{3}F_\alpha;\Delta^2)}{2G_{3,\nu}(\frac{1}{3}F_\alpha;\Delta^2)-(1+d)G_{5,\nu}(\frac{1}{5}F_\alpha;\Delta^2)} \tag{6.7}$$

while the relative efficiency of the PTE compare to the RE goes up and crosses the 1-line at

$$\Delta_{**}^2 = \frac{(1+d)\Big\{1-G_{3,\nu}\Big(\frac{1}{3}F_\alpha;\Delta^2\Big)\Big\}}{(1-d)\Big\{1-2G_{3,\nu}\Big(\frac{1}{3}F_\alpha;\Delta^2\Big)-(1+d)G_{5,\nu}\Big(\frac{1}{5}F_\alpha;\Delta^2\Big)\Big\}} \tag{6.8}$$

ii) Finally, as $\Delta^2 \to \infty$, $\mathrm{RE}\Big[\hat{\theta}^{\mathrm{PTE}}(d) : \tilde{\theta}\Big] \to 1$ regardless of the value of d and α, while the relative efficiency of the PTE relative to the RE grows unboundedly regardless of the value of α, unless $d = 1$.

In general, the PTE is more efficient than the UE if $0 \le \Delta^2 < \Delta_*^2$. Starting from some $\Delta^2 > \Delta_*^2$ the UE is more efficient than the PTE up to a moderate value of Δ^2, and then slowly approaches to the 1-line. On the other hand, for general $\Delta^2 > 0$, we have $\mathrm{RE}[\hat{\theta}^{\mathrm{PTE}}(d) : \hat{\theta}^{\mathrm{RE}}(d)] \lesseqqgtr 1$ according as $\Delta^2 \lesseqqgtr \Delta^2_{**}$.

6.3 Comparing SE against UE, RE and PTE

The relative efficiency of the SE relative to the UE, RE and PTE are respectively

$$\mathrm{RE}\Big[\hat{\theta}^{\mathrm{SE}}(d) : \tilde{\theta}\Big] = \left[1 + H^{-1}\frac{2\bar{x}^2 K_\nu^2 \xi}{\pi S_{xx}}\right]^{-1} \tag{6.9}$$

$$\mathrm{RE}\Big[\hat{\theta}^{\mathrm{SE}}(d) : \hat{\theta}^{\mathrm{RE}}(d)\Big] = \left[d^2 H + (1-d)^2 \frac{\bar{x}^2 K_\nu^2 \Delta^2}{S_{xx}}\right]\left[H + \frac{2\bar{x}^2\xi}{\pi S_{xx}}\right]^{-1} \tag{6.10}$$

$$\mathrm{RE}\Big[\hat{\theta}^{\mathrm{SE}}(d) : \hat{\theta}^{\mathrm{PTE}}(d)\Big] = \left[H + \frac{\bar{x}^2\sigma^2}{S_{xx}} g(\Delta^2)\right]\left[H + \frac{2\bar{x}^2 K_\nu^2 \xi}{\pi S_{xx}}\right]^{-1} \tag{6.11}$$

where $\xi = \Big\{1 - 2e^{\frac{-\Delta^2}{2}}\Big\}$, and $g(\Delta^2)$ is defined earlier.

The relative efficiency of the SE relative to the UE is a decreasing function of Δ^2 which takes its maximum value at $\Delta^2 = 0$. It falls sharply as Δ^2 moves away from 0, and approaches to some constant value from some moderate value of Δ^2. The relative efficiency of the SE relative to the RE is an increasing function of Δ^2 which takes its minimum value at $\Delta^2 = 0$. It grows unboundedly as Δ^2 increases. The relative efficiency of the SE relative to the PTE is neither increasing nor decreasing function of Δ^2. Moreover it depends on the choice of the level of significance. But from some moderate to large value of Δ^2 it approaches to a constant value regardless of the choice of α.

From the expressions in (6.9) - (6.11) we draw the following conclusions.

i) Under H_0 $\Delta^2 = 0$, and hence

$$\mathrm{RE}\Big[\hat{\theta}^{\mathrm{SE}}(d) : \tilde{\theta}\Big] = \Big[1 - 2H^{-1}VK_\nu^2\Big]^{-1} \tag{6.12}$$

$$\mathrm{RE}\Big[\hat{\theta}^{\mathrm{SE}}(d) : \hat{\theta}^{\mathrm{RE}}(d)\Big] = d^2\Big[1 - 2H^{-1}VK_\nu^2\Big]^{-1} \tag{6.13}$$

$$\mathrm{RE}\Big[\hat{\theta}^{\mathrm{SE}}(d) : \hat{\theta}^{\mathrm{PTE}}(d)\Big] = \Big[1 - H^{-1}V(1-d^2)G_3\Big]\Big[1 - 2H^{-1}VK_\nu^2\Big]^{-1} \tag{6.14}$$

where $V = \frac{H^{-1}\bar{x}^2}{\pi S_{xx}}$ and $G_3 = G_{3,v}\Big(\frac{1}{3}F_\alpha; 0\Big)$. The second term of the right hand side of (6.12) is always positive. So the maximum relative efficiency of the SE relative to the UE is always greater than 1, and at $\Delta^2 = 0$. The relative efficiency of the SE relative to the RE and PTE depends on d. When $d = 0$, the minimum relative efficiency of the SE relative to the RE is 0 at $\Delta^2 = 0$. No such minimum or maximum relative efficiency of the SE relative to the PTE exists at $\Delta^2 = 0$. For a larger value of α, $G_{3,v}\Big(\frac{1}{3}F_\alpha; 0\Big)$ is smaller than for a smaller value of α. Therefore, at $\Delta^2 = 0$ the relative efficiency of the SE relative to the PTE is higher for larger choice of α, and vice-versa. If $\alpha \to 1$, $G_{3,v}\left(\frac{1}{3}F_\alpha;0\right) \to 0$, and hence the relative efficiency of the SE relative to the PTE tends to be that of the SE relative to the UE.

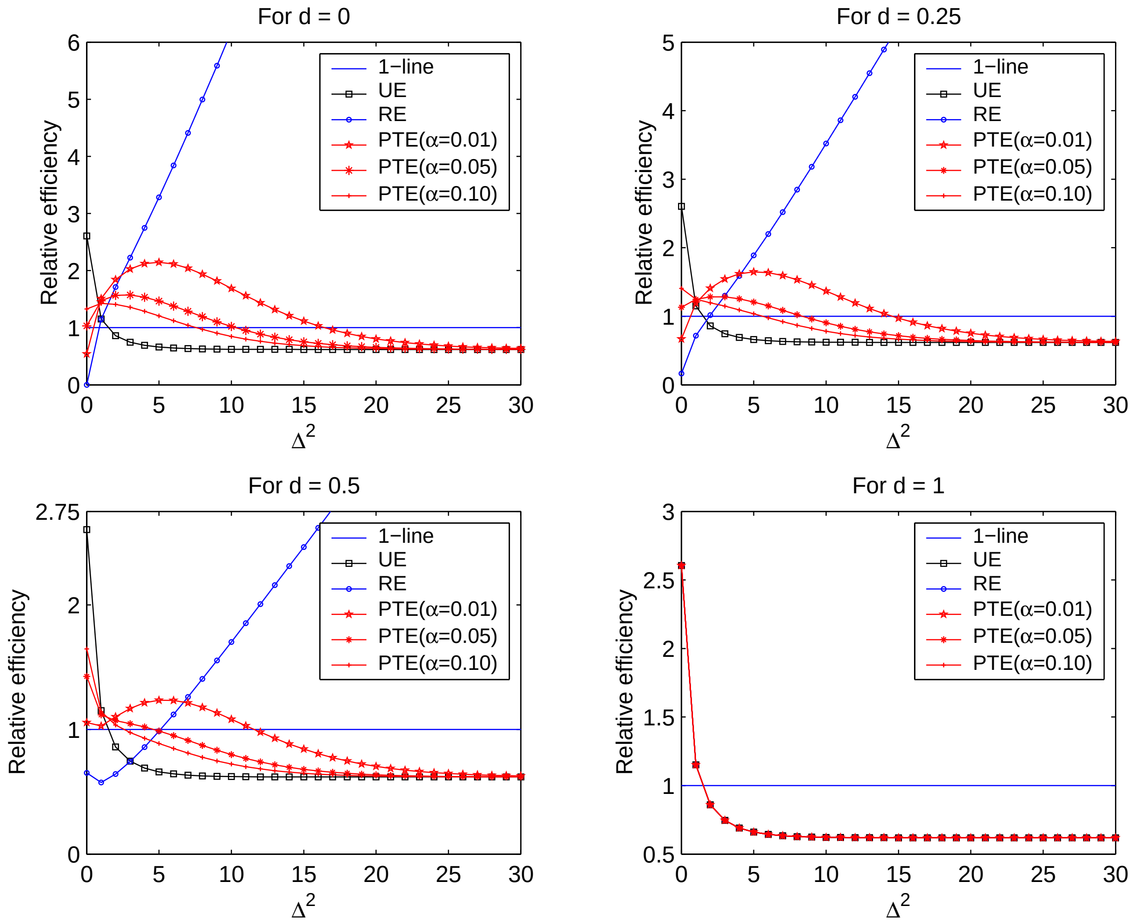

Figure 4: Graph of relative efficiency of SE relative to UE, RE and PTE against Δ^2

However, for a fixed α as d increases, the relative efficiency function also increases. When there is complete distrust on the null hypothesis the relative efficiency of the SE relative to all estimators becomes the same for all Δ^2 and all α.

ii) If Δ^2 moves away from 0, the relative efficiency of the SE relative to the UE falls sharply; that relative to the RE quickly grows up; and that relative to the PTE goes up or down depending on $2G_{3,v}\left(\frac{1}{3}F_\alpha;\Delta^2\right) \gtrless (1+d)G_{5,v}\left(\frac{1}{5}F_\alpha;\Delta^2\right)$.

As $\Delta^2 \to \infty$, $\mathrm{RE}\left[\hat{\theta}^{\mathrm{SE}}(d) : \tilde{\theta}\right] \to \left\{1+\left(1+\frac{S_{xx}}{n\bar{x}^2}\right)^{-1}\frac{2}{\pi}K_\nu^2\right\}^{-1} < 1$; $\mathrm{RE}\left[\hat{\theta}^{\mathrm{SE}}(d) : \hat{\theta}^{\mathrm{RE}}(d)\right] \to \infty$, except for $d = 1$; and $\mathrm{RE}\left[\hat{\theta}^{\mathrm{SE}}(d) : \hat{\theta}^{\mathrm{PTE}}(d)\right]$ approaches to the constant value $\left[1+\frac{2H^{-1}\bar{x}^2K_\nu^2}{\pi S_{xx}}\right]^{-1}$ which does not depend on d and α.

In general, the relative efficiency of the SE relative to the UE decreases from

$\left\{1-\left(1+\frac{S_{xx}}{n\bar{x}^2}\right)^{-1}\frac{2}{\pi}K_\nu^2\right\}^{-1}$ at $\Delta^2 = 0$, crosses the 1-line at $\Delta^2 = 2ln2$, and it approaches to a constant value as $\Delta^2 \to \infty$. Therefore, for $\Delta^2 < 2ln2$ the SE performs better than the UE, otherwise the UE performs better than the SE. On the other hand, $\text{RE}\left[\hat{\theta}^{\text{SE}}(d) : \hat{\theta}^{\text{RE}}(d)\right]$ increases as Δ^2 moves away from 0. It grows up unboundedly as $\Delta^2 \to \infty$. The general picture of the relative efficiency of SE compare to the PTE can be described as follows. The relative efficiency function begins with the value in the expression of (6.14) at $\Delta^2 = 0$ and crosses the 1-line at

$$\Delta^2 = \frac{\frac{2}{\pi}K_\nu^2\left(1-2e^{-\frac{\Delta^2}{2}}\right)-(1-d^2)G_{5,\nu}\left(\frac{1}{5}F_\alpha;\Delta^2\right)}{\sigma^2\left[2(1-d)G_{3,\nu}\left(\frac{1}{3}F_\alpha;\Delta^2\right)-(1-d^2)G_{5,\nu}\left(\frac{1}{5}F_\alpha;\Delta^2\right)\right]}. \tag{6.15}$$

Finally, as $\Delta^2 \to \infty$ the relative efficiency function approaches to the constant value $\left[1+\frac{2H^{-1}\bar{x}^2K_\nu^2}{\pi S_{xx}}\right]^{-1}$.

7 Concluding Remarks

Among the four estimators considered in this paper, the UE is the only unbiased estimator, and it is based only on the sample information. The estimators based on both the non-sample prior information and sample information are biased. However, the inclusion of non-sample prior information increases the efficiency of the estimators. The relative efficiency of the biased estimators depends on the departure constant Δ^2 and the degree of distrust d. From 0 to some moderate value of Δ^2, the SE dominates the UE for all values of d. Starting from some moderate values of Δ^2 the SE is dominated by the UE. From 0 to some moderate values of Δ^2 the SE is dominated by the RE. But starting from that moderate value of Δ^2 the SE dominates the RE. However, the increasing rate of the relative efficiency of the SE relative to the RE decreases as the value of the coefficient of distrust increases. Under the null hypothesis the SE dominates the PTE unless α or d is not too small. From some small to moderate values of Δ^2 the SE dominates the PTE if α is not too large. Starting from some moderate value of Δ^2, SE is dominated by the PTE. In practice, the non-sample prior information is obtained from expert knowledge or previous studies, and hence the value of the parameter available from prior information is expected to be close to its true value and the *degree of distrust* on the null hypothesis is very unlikely to be close to 1. Also, the level of significance is always preferred to be small. Therefore, under the above circumstances, the shrinkage estimator would be the best choice as an improved estimator of the intercept parameter among all the estimators considered in this paper.

References

1. Ahmed, S.E. and Saleh, A.K.Md.E. (1989). Pooling multivariate data. *Journal of Statistical Computation and Simulation*, **31**, 149-167

2. Bancroft, T.A. (1944). On biases in estimation due to the use of the preliminary tests of significance. *Annals of Mathematical Statistics*, **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. an Bancroft, T.A. (1968). On pooling means when variance is unknown. *Journal of American Statistical Association*, **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. *Pakistan Journal of Statistics*, **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. Khan, S., Hoque, Z. and Saleh, A K Md E (2002). Improved estimation of the slope parameter for linear regression model with normal errors and uncertain prior information. *Journal of Statistical Research*, **36**, 55-73.

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

13. 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.

14. 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.

15. 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.

16. 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.

17. Stein, C. (1981). Estimation of the mean of a multivariate normal distribution. *Annals of Statistics* **9**, 1135–1151.