Journal of Statistical Research *ISSN 0256 - 422 X*
2005, Vol. 39, No. 2, pp. 29–50
Bangladesh

SOME IMPROVED ESTIMATORS IN LOGISTIC REGRESSION MODEL

M A Matin
Department of Statistics, School of Mathematics
University of Leeds, Leeds, LS2 9JT, UK
Email: matin@maths.leeds.ac.uk

A. K. MD. Ehsanes Saleh
School of Mathematics and Statistics
Carleton University, Ottawa, K1S 5B6, Canada
Email: esaleh@math.carleton.ca

summary

The problem of estimating the parameters of logistic regression model is considered when it is known from extraneous sources that the *uncertain prior information* in the form of the hypothesis $H_0 : \beta_0 = \ldots = \beta_{k-1} = \beta^0$ (pivot) may hold. Five estimators, namely, the unrestricted maximum likelihood estimator (UMLE), the shrinkage restricted estimator (SRE), the shrinkage preliminary test estimator (SPTE), the shrinkage estimator (SE) and the positive-rule shrinkage estimator (SE^+) are considered. The SE and SE^+ are the Stein-type estimators based on the preliminary test approach of Saleh and Sen. In the light of derived MSE matrices and distributional risks, the relative performance of the five estimators under local alternatives are studied in detail. These analyses reveal that when $k \geq 3$, we should use the SE or SE^+ and for $k \leq 2$ it is advisable to use the preliminary test estimator (PTE).

Keywords and phrases: Preliminary test estimator, Shrinkage estimator, Stein-type estimator, Quadratic risk, Pitman alternatives, Wald test.

AMS Classification: 62Jxx.

1 Introduction

In problems of statistical inference, the use of prior information on some (all) of the parameters in a statistical model usually leads to an improved inference procedure for other (all) parameters of interest. The prior information may be known or uncertain. The known prior information is generally incorporated in the model in the form of a constraint, giving rise to *restricted* models. The analysis of such restricted models leads to improved statistical procedures compared with the *unrestricted* model case when such constraints hold. (The estimators resulting from the restricted (unrestricted) model is known as the restricted (unrestricted) estimators of the parameters of the model.) The validity and efficiency of the restricted model analysis retains its properties over the restricted parameter space induced by the constraint, while the same holds for the unrestricted model analysis over the entire parameter space. Therefore, the results of an analysis of the restricted and unrestricted models need to weigh against loss of efficiency and validity of the constraints in order to choose between two extreme inference techniques, when a full confidence *may not* be put in the prior information.

When we encounter problems with *uncertain prior information* in a statistical model, we may impose some prior information (which may be available from some extraneous considerations) on the model. However, there may be reasons to suspect their validity as a recommendation for restricted model analysis. This uncertain prior information considered in the form of a constraint may be regarded as "nuisance parameter" in the statistical inference of the model. To ameliorate this uncertainty of the prior information in the model one could naturally follow "Fisher's recipe" of elimination of the nuisance parameter by a preliminary test on the validity of the uncertain prior information in the form of a parametric constraint and then choose either the restricted or the unrestricted inference depending on the fate of the preliminary test. This theme brings home a compromised inference procedure between two extremes. (It could be looked upon as a happy marriage between two extremes.) This line of thought was pioneered by Bancroft (1944) in the study of the consequences of incorporating a preliminary test in pooling several estimates of variances in an analysis of variance table. This idea of preliminary test estimation has been developed further in a series of papers by Bancroft, Bancroft and Han (see Bancroft and Han (1980) for an excellent review) among others.

In 1956, Charles Stein (Stein, 1956) discovered an unexpected but surprising result (in the form of a counter example) that the sample mean was an inadmissible estimator of the mean of a normal distribution under squared error loss in three or higher dimension. Afterwards, James and Stein (1961) provided an explicit estimator dominating the sample mean. This work was like a shock for the whole statistical world and profoundly affected the course of statistical theory. Virtually, the James-Stein work attacked all good properties of statistical estimators viz: least squares, maximum likelihood, unbiasedness, invariance and the minimax principle. The usual estimator of the sample mean is uniformly minimum covariance unbiased, maximum likelihood, best invariant, and minimax, and yet its performance in terms of expected squared error loss is demonstratively inferior to that of the James-Stein estimator. The last fifty years or so of statistical science witnessed a fundamental growth of the literature in this fruitful area of research. See Berger (1985) for a detail account of some of these developments, mostly related to the classical multinomial and some specific types of exponential families of distributions. For further development in multivariate Stein-type estimation see Ghosh and Lin (1986).

The estimators which uniformly improve on standard estimators are usually called Stein-type estimators as a gesture of honor to Professor Charles Stein. Basically, the Stein-type estimators may be obtained via a *decision theoretic* approach for the normal theory models and *empirical Bayes* approach as propounded by Efron and Morris (1973). However, Casella (1985) pointed out that the Stein-type estimators need appropriate test statistics for testing out the adequacy of the uncertain prior information that is incorporated in the actual formation of the estimators. Stein-type estimators adjust the unrestricted estimator by an amount equal to the difference (= the restricted estimator - the unrestricted estimator) multiplied by a shrinkage factor times the reciprocal of the relevant test-statistics (for testing the uncertain prior information). Generally, these test-statistics measure the normalized distance between the restricted and unrestricted estimators and follow a noncentral chi-squared (or noncentral F) distribution with appropriate degrees of freedom. The risks of Stein-type estimators (including other estimators) depend on the noncentrality parameter of the chi-square (or F) distribution. In an asymptotic setup the computation of the risk of the Stein-type estimators (as well as the other estimators) may be obtained from their non-degenerate asymptotic distributions. The derivation of these non-degenerate asymptotic distributions depends on the sequence of local alternatives (known as Pitman alternatives). Note that the local alternatives mean the alternatives which are in close proximity of the null hypothesis. In other words, we may concieve of a sphere of radius, r_n with the null hypothesis as the centre such that the sphere reduces to a point in k-dimension as $n \to \infty$. With such alternatives the aim is to have a steeper power function of the preliminary test near the null hypothesis. The risks obtained via asymptotic technique under the local alternatives are termed as the asymptotic distributional risks (ADRs). The ADR provides an easy and meaningful access to the study of the Stein-type estimators (as well as the other estimators) in a much broader setup. (For more on the ADR see Sen (1979, 1986), and Saleh and Sen (1987).)

Though Stein-type estimators achieved a big theoretical success they are really under-achievers in a number of practical applications and least affected by the enriched development of state-of-art computer facilities. So, it is ripe for a computer-intensive treatment that brings the substantial benefits of James-Stein estimation to bear on complicated, realistic problems (Efron, 1995).

It is not hard for one to find a close relation between Robbins (1956)'s empirical Bayes theory and the James-Stein phenomenon. The ultimate gain (whatever theory we use) is to have considerably better inference than a classical one. Moreover, in both the cases, the statisticians get advantage of using a Bayes estimation rule, without trouble of choosing a Bayes prior distribution. The data effectively choose the correct prior. For a better understanding of the empirical Bayes interpretation of the James-Stein estimator see Efron and Morris (1973). In his lucid introduction written for the James-Stein (James and Stein, 1961) article Efron (1992) described how the Bayesian theme works behind the James-Stein theorem and the exactness of the empirical Bayes estimator of the sample mean to the James-Stein estimator.

Given the benefit of the above discussion, it is a legitimate question to ask (or the same question frequently asked by the Bayesians, empirical Bayesians and mostly by practitioners in these areas of research): how much do we gain from the Stein-type phenomenon compared to the Bayes and empirical Bayes? We may recall here the mathematical complications in obtaining the posterior distribution in the Bayes approach and the numerical difficulty and computer-intensiveness (e.g., Gibbs sampling) in obtaining Stein-type estimators via empirical Bayes approach in non-standard problems. The reader may see Albert and Chib (1993)'s article to have a glimpse of how much mathematical derivation they have done and what is the ultimate gain?

Saleh and Sen (Saleh and Sen, 1978, 1983, 1984a,b, 1985a,b,c,d, 1986a,b) took the frequentist path of improving standard estimators regarding non-parametric problems on location and linear regression parameters by developing Stein-type estimators. (The path they have chosen actually combined the idea of the preliminary test approach and the concept of shrinkage in the James-Stein estimation setup. See Saleh and Sen (1984b) for an overview.) Using their approach, it is possible to obtain explicit expressions for the Stein-type estimators which is usually not possible by the empirical Bayes approach in non-standard problems as mentioned above. However, this type of approach has never been used in the development of the analysis of binary data based on logistic regression. This work is the first attempt to apply the developments in Stein-type estimators due to the aforesaid authors and see the consequences of positive or negative results.

Section 2 introduces the preliminaries on the inference of logistic regression model parameters. In section 3 we proposes new estimators. Section 4 contains the asymptotic bias, MSE matrix and risk expression for the newly proposed estimators in section 3. In section 5, risk and MSE analyses of the estimators are given side by side. Section 6 includes illustrative example and graph analysis. Section 7 concludes with some remarks.

2 Preliminaries on the Estimation of Logistic Regression Model Parameters

Let $Y_i \in \{0, 1\}$ denote a dichotomous dependent variable, and let $\mathbf{x}_i = (1, x_1, x_2, \ldots, x_{k-1})$ be a k dimensional vector of explanatory variables, for the ith observation. The probability that $Y_i = 1$, given the value of $\mathbf{x}_i$, is assumed to be $P(Y_i = 1) = \pi(\mathbf{x}_i)$ and is defined by the logistic regression model as

$$\pi(\mathbf{x}) = [1 + e^{-\beta^T \mathbf{x}}]^{-1} \tag{2.1}$$

where $\beta^T = (\beta_0, \beta_1, \ldots, \beta_{k-1})$ is a vector of k parameters of interest. The logit transformation in terms of $\pi(\mathbf{x})$ is given by

$$log \frac{\pi}{1 - \pi} = \beta_0 + \beta_1 x_1 + \ldots + \beta_{k-1} x_{k-1}. \tag{2.2}$$

The maximum likelihood estimate of β can be obtained from the likelihood equation

$$X^T (Y - \pi) = \mathbf{0} \tag{2.3}$$

where X is an $n \times k$ matrix of explanatory variables, Y is an $n \times 1$ vector values of dependent variable, π is an $n \times 1$ vector of π_i's and $\mathbf{0}$ is a $k \times 1$ vector of zeros. The likelihood equation is non-linear in $\beta_0, \beta_1, \ldots, \beta_{k-1}$ and is solved by suitable iterative methods (Hosmer and Lemeshow, 1989). The formula used for $(t+1)$-th iteration based on the t-th result is given by

$$\beta^{(t+1)} = \beta^{(t)} + \{X^T Diag[\pi_i^{(t)}(1-\pi_i^{(t)})]X\}^{-1} X^T (Y - m^{(t)}) \tag{2.4}$$

where $m^{(t)} = (\pi_1^{(t)}, \pi_2^{(t)}, \ldots, \pi_k^{(t)})$.

The estimated information matrix can be formulated as

$$\widehat{\mathbf{I}(\beta)} = X^T V X \tag{2.5}$$

where V is an $n \times n$ diagonal matrix with element $\widehat{\pi}_i(1-\widehat{\pi}_i)$ in the main-diagonal and 0 otherwise. The variances and covariances of the estimated coefficients are obtained from the inverse of the matrix defined in (2.5).

In order to test the significance of the parameter vector $\beta = (\beta_0, \beta_1, \ldots, \beta_{k-1})^T$ the usual test procedures like the likelihood ratio, score and Wald are used in practice. These procedures have optimal asymptotic properties, but the small-sample behavior is less well known (see some detail results in Matin (2005)).

3 The Proposed Estimators

Our primary objective is to estimate the parameter vector $\beta = (\beta_0, \beta_1, \ldots, \beta_{k-1})^T$ when it is known from extraneous sources that the *uncertain prior information* in the form of the hypothesis $H_0 : \beta_0 = \ldots = \beta_{k-1} = \beta^0$ (pivot) may hold. For example, if $\beta^0 = 0$, then the probabilities are equal to 1/2. Under this situation, we first consider (begin with) the maximum likelihood estimator (MLE) which is already described in section 2. We designated this as the *unrestricted maximum likelihood estimator* (UMLE), $\widetilde{\beta}_n$, of β.

If the hypothesis H_0 is true, then it is worthwhile to consider $\beta^0\mathbf{1_k}$ as the estimate of β. However, the presence of sample data dictates that we combine $\beta^0\mathbf{1_k}$ with the UMLE $\widetilde{\beta}_n$ as

$$\widehat{\beta}_n^{(c)} = c\beta^0\mathbf{1_k} + (1-c)\widetilde{\beta}_n \tag{3.1}$$

$0 < c \leq 1$ where c is appropriately determined from some conviction on the degree of confidence on H_0. If $c = 1$, we arrive at the usual estimator $\beta^0\mathbf{1_k}$ (not based on the sample). We designate the estimator in (3.1) as the *shrinkage restricted estimator* (SRE) of β.

Now, for fixed c, the performance of $\widehat{\beta}_n^{(c)}$ is optimal when H_0 is true but disastrous (as the risk of the estimator increases unboundedly) when H_0 does not hold. The SRE will perform poorly in this event and may even be inconsistent.

Thus, there is compelling reason to combine $\beta^0\mathbf{1_k}$ and $\widetilde{\beta}_n$ in some other way. One such method is due to Bancroft (1944) and Han and Bancroft (1968) known as the preliminary test estimation. Treating the null hypothesis $H_0 : \beta_0 = \ldots = \beta_{k-1} = \beta^0$ (pivot) as the nuisance parameter for the problem one applies the "Fisher's recipie" to test it out and on the result of the test chooses $\beta^0\mathbf{1_k}$ or $\widetilde{\beta}_n$ at a certain level α of significance. This estimator via the preliminary test is known as the preliminary test estimator (PTE). The result is that we obtain a simple estimator which controls the problem (regarding risk) inherited by the shrinkage restricted estimator $\widehat{\beta}_n^{(c)}$.

In order to test the null hypothesis we consider the test statistic W_n defined by

$$W_n = n(\widetilde{\beta}_n - \beta^0\mathbf{1_k})^T(X^TVX)(\widetilde{\beta}_n - \beta^0\mathbf{1_k}). \tag{3.2}$$

This allows us to combine the SRE and UMLE as follows:

$$\widehat{\beta}_n^{SPT} = \widehat{\beta}_n^{(c)}I(W_n < w_{n,\alpha}) + \widetilde{\beta}_n I(W_n \geq w_{n,\alpha}) = \widetilde{\beta}_n - c(\widetilde{\beta}_n - \beta^0\mathbf{1_k})I(W_n < w_{n,\alpha}) \tag{3.3}$$

where $w_{n,\alpha}$ is the α level critical value using the distribution of W_n and $I(A)$ is the indicator function for the set A ($I(A) = 1$ if A happens and 0 otherwise). The expression (3.3) is a convex combination of $\widehat{\beta}_n^{(c)}$ and $\widetilde{\beta}_n$. The estimator is called the *shrinkage preliminary test estimator* (SPTE). For c = 1 in (3.3), we obtain the ordinary preliminary test estimator (PTE) of β, $\widehat{\beta}_n^{PT}$.

Note that the test statistic in (3.3) is the Wald statistic. One can use the score statistic as well as the likelihood ratio statistic. These three test statistics are asymptotically equivalent, while in small samples they behave differently. (For their detail use see Matin and Saleh (2005)).

The preliminary test estimator has its own problem. Though it combines the SRE and UMLE convexly, it depends on the level α ($0 < \alpha < 1$) of significance. Thus, what would be the optimal value of α remains a burning question. On

the other hand, the mean square error of the estimator at certain points of the parameter space may not be acceptable being too high even with optimal α.

The SPTE has the same disadvantage as the PTE in that it depends on the level of significance α ($0 < \alpha < 1$) and also results in two extreme estimators, namely, $\widehat{\beta}_n^{(c)}$ and $\widetilde{\beta}_n$ depending on the outcome of the preliminary test (PT).

This encourages one to find alternative estimators. One may try to develop Stein-type estimators based on the preliminary test approach of Saleh and Sen (Saleh and Sen, 1978, 1983, 1984a,b, 1985a,b,c,d, 1986a,b). There are two types of Stein-type estimators we shall consider, namely the ordinary shrinkage estimator (SE) and the positive-rule shrinkage estimator (SE$^+$).

Thus, a *shrinkage estimator* (SE) of β is defined by

$$\widehat{\beta}_n^S = \beta^0\mathbf{1_k} + \{1-(k-2)W_n^{-1}\}(\widetilde{\beta}_n - \beta^0\mathbf{1_k}) = \widetilde{\beta}_n - (k-2)(\widetilde{\beta}_n - \beta^0\mathbf{1_k})W_n^{-1}, \text{ for } k \geq 3. \tag{3.4}$$

Comparing (3.3) and (3.4) we note that they are similar except that c has been replaced by $k-2$ and $I(W_n < w_{n,\alpha})$ has been replaced by a smooth function W_n^{-1}. For large values of W_n, $W_n^{-1} \to 0$ giving the estimate $\widetilde{\beta}_n$ as in the case of PTE, $\widehat{\beta}_n^{PT}$. For small values of W_n, $\widehat{\beta}_n^{PT} = \widehat{\beta}_n^{(c)}$ but $\widehat{\beta}_n^S$ remains unsettled as $W_n^{-1} \to \infty$ the estimator $\widehat{\beta}_n^S \to -\infty$. Thus, near the origin of W_n, the two estimators $\widehat{\beta}_n^{PT}$ and $\widehat{\beta}_n^{S}$ behave differently. The SE has been defined for $k \geq 3$ while the PTE has no such restriction. However, we note that if $W_n = k-2$, then $\widehat{\beta}_n^S = \beta^0 \mathbf{1_k}$ which is the result for small W_n. This means that a better situation is to define a truncated shrinkage estimator known as the *positive-rule shrinkage estimator* (SE$^+$) defined by

$$\widehat{\beta}_n^{S^+} = \beta^0 \mathbf{1_k} + \{1 - (k-2)W_n^{-1}\} I(W_n > k-2)(\widetilde{\beta}_n - \beta^0 \mathbf{1_k}). \tag{3.5}$$

Thus, we have proposed five estimators:

Estimator	Full name	Defined in
UMLE	Unrestricted maximum likelihood estimator	Section 2
SRE	Shrinkage restricted estimator	Equation 3.1
SPTE	Shrinkage preliminary test estimator	Equation 3.3
SE	Shrinkage estimator	Equation 3.4
SE$^+$	Positive-rule shrinkage estimator	Equation 3.5

4 Asymptotic Bias, MSE Matrix and Risk of the Estimators

In this section, first we formulate the asymptotic distributional risk and local alternatives which we need in the sequel and then, we move on to the asymptotic distributional properties of the estimators.

Let β_n^* be any estimator of the vector parameter β, Q be a positive semi-definite matrix, and consider the quadratic loss function

$$L_n(\beta_n^*, \beta) = n(\beta_n^* - \beta)^T Q(\beta_n^* - \beta) = ntr(Q(\beta_n^* - \beta)(\beta_n^* - \beta)^T). \tag{4.1}$$

Then, the risk of β_n^* is given by

$$R_n(\beta_n^*, \beta) = EL_n(\beta_n^*, \beta) = tr(QD_n) \tag{4.2}$$

where $D_n = ntrE(\beta_n^* - \beta)(\beta_n^* - \beta)^T$. Note that if $Q = I$ (where I is an identity matrix) then $R_n(\beta_n^*, \beta)$ equals the usual summed mean squared error.

It is well-known that the W_n-statistic defined in (4.2) is a consistent test statistic since, for any fixed β ($\neq \beta^0 \mathbf{1_k}$) the power of W_n converges in probability to 1 as $n \to \infty$. Now, by virtue of the consistency of the W_n, the estimators $\widetilde{\beta}_n$, $\widehat{\beta}_n^{SPT}$, $\widehat{\beta}_n^S$ and $\widehat{\beta}_n^{S^+}$ are equivalent in probability as $n \to \infty$ for any fixed alternative while $\widehat{\beta}_n^{(c)}$ will have unbounded risk or MSE. (Hence, in the asymptotic setup, we need not consider a fixed alternative.) So, following Saleh and Sen (1987), this is determined in the setting of a more sensible and interesting sequence of local alternatives to avoid the asymptotic degeneracy. (For more on the local alternatives see Puri and Sen (1985), Agresti (1991).) We consider a sequence $\{K_{(n)}\}$ of local alternatives to H_0 :

$$K(n) : \beta_{(n)} = \beta^0 \mathbf{1_k} + n^{-1/2} \delta \tag{4.3}$$

where $\delta = (\delta_1, \delta_2, \ldots, \delta_k)$ and δ_i's are fixed numbers for $i = 1, 2, \ldots, k$. This sequence of alternatives allows one to derive the non-degenerate asymptotic distribution of the various estimators in question. Note that, for $\delta = \mathbf{0}$, $K_{(n)}$ reduces to the null hypothesis, $H_0 : \beta = \beta^0 \mathbf{1_k}$. Further, the sequence $\{K_{(n)}\}$ of alternatives converges to the null hypothesis as $n \to \infty$. Next, we note that under $\{K_{(n)}\}$,

$$\sqrt{n}(\widetilde{\beta}_n - \beta_{(n)}) \sim N_k(\mathbf{0}, (X^T V X)). \tag{4.4}$$

In general, the asymptotic distribution function of β^*_n is given by

$$G^*_{(\beta^*)}(x) = \lim_{n\to\infty} P\{n^{1/2}(\beta^*_n - \beta_{(n)}) \leq x | K_{(n)}\} \tag{4.5}$$

whenever the limit exists; while the asymptotic dispersion matrix is given by

$$V^*_{(\beta^*)}(x) = \int_{\cdots} \int x x^T dG^*_{(\beta^*)}(x).$$

Then, the asymptotic distributional risk of β^* is defined by

$$R^*(\beta^*, \delta) = tr(QV^*_{(\beta^*)}). \tag{4.6}$$

On the basis of the above one may determine the asymptotic distribution of $\widehat{\beta}^{SPT}_n$ and the asymptotic representation of $\widehat{\beta}^{S}_n$ and $\widehat{\beta}^{S^+}_n$ under $\{K_{(n)}\}$ routinely follows from Saleh and Sen (1987). In our problem, the following theorem details out the results.

Theorem 1 Under $\{K_{(n)}\}$ in (4.3) and usual regularity conditions, the following holds:

(a)

$$\lim_{n\to\infty} P\{W_n \leq x | K_{(n)}\} = H_k(x; \Delta), \Delta = \delta^T D_1^{-1} \delta \tag{4.7}$$

where $D_1 = (X^T V X)^{-1}$ and $H_m(x; \Delta)$ stands for the CDF of a non-central chi-squared distribution with m degrees of freedom;

(b)

$$\lim_{n\to\infty} P\{n^{1/2}(\widehat{\beta}^{(c)}_n - \beta_{(n)}) \leq x\} = G_k(x; -c\delta, (1-c)^2 D_1) \tag{4.8}$$

where $G_k(x; \mu, \sum)$ stands for the CDF of a k-variate normal distribution with mean μ and covariance matrix $\sum$;

(c)

$$\begin{aligned} \lim_{n\to\infty} P\{n^{1/2}(\widehat{\beta}^{SPT}_n - \beta_{(n)}) \leq x\} &= G_k(x - c\delta; \mathbf{0}, (1-c)^2 D_1) H_k(\chi^2_{k,\alpha}) \\ &+ \int_{E(\delta)} G_k(x; -c\mathbf{z}, \mathbf{0}, (1-c)^2 D_1) dG_k(\mathbf{z}; \mathbf{0}, D_1) \end{aligned} \tag{4.9}$$

where $E(\delta) = \{\mathbf{z} : (\mathbf{z} + \delta)^T D_1^{-1} (\mathbf{z} + \delta) \geq \chi^2_{k,\alpha}\}$;

(d) asymptotic distributional representation of $n^{1/2}(\widehat{\beta}_n^S - \beta_{(n)})$ is given by

$$\mathbf{z} - (k-2)\frac{(\mathbf{z}+\delta)}{(\mathbf{z}+\delta)^T D_1^{-1}(\mathbf{z}+\delta)} \tag{4.10}$$

where $\mathbf{z} \sim N_k(\mathbf{0}, D_1^{-1})$ and

(e) that of $n^{1/2}(\widehat{\beta}_n^{S+} - \beta_{(n)})$ is given by

$$\begin{aligned}\mathbf{z} - (k-2)&\frac{(\mathbf{z}+\delta)}{(\mathbf{z}+\delta)^T D_1^{-1}(\mathbf{z}+\delta)} - (\mathbf{z}+\delta)I\{(\mathbf{z}+\delta)^T D_1^{-1}(\mathbf{z}+\delta) < k-2\} \\ + \ (k-2)&\frac{(\mathbf{z}+\delta)I\{(\mathbf{z}+\delta)^T D_1^{-1}(\mathbf{z}+\delta) \ < \ k-2\}}{(\mathbf{z}+\delta)^T D_1^{-1}(\mathbf{z}+\delta)}.\end{aligned} \tag{4.11}$$

Using the above theorem and some of the results (in Appendix B) of Judge and Bock (1978) we derive the asymptotic distributional biases, MSE matrices and quadratic risks (based on the loss function defined in (4.1)) for the five estimators considered. They are as follows.

First, the biases of the estimators are

$$E\sqrt{n}(\widetilde{\beta}_n - \beta) = \mathbf{0}. \tag{4.12}$$

$$E\sqrt{n}(\widehat{\beta}_n^{(c)} - \beta) = -\ c\delta. \tag{4.13}$$

$$E\sqrt{n}(\widehat{\beta}_n^{SPT} - \beta) = -\ c\delta H_{k+2}(\chi^2_{k,\alpha};\Delta). \tag{4.14}$$

$$E\sqrt{n}(\widehat{\beta}_n^{S} - \beta) = -\ (k-2)\delta E(\chi^{-2}_{k+1}(\Delta)). \tag{4.15}$$

$$\begin{aligned}E\sqrt{n}(\widehat{\beta}_n^{S^+} - \beta) = -\ (k-2)\delta\{&E(\chi^{-2}_{k+1}(\Delta)) + H_{k+2}(\chi^2_{k,\alpha};\Delta) \\ -\ &E[(\chi^{-2}_{k+1}(\Delta))I((\chi^2_{k+2}(\Delta)) < k-2)]\}.\end{aligned} \tag{4.16}$$

The biases in 4.12-4.16 can be written in terms of the non-centrality parameter Δ to facilitate the comparisons among them. These are known as normalized biases of the estimators. They are given below:

$$B_1 = \mathbf{0}. \tag{4.17}$$

$$B_2 = c^2\Delta. \tag{4.18}$$

$$B_3 = c^2\Delta\{H_{k+2}(\chi^2_{k,\alpha};\Delta)\}^2. \tag{4.19}$$

$$B_4 = (k-2)^2\Delta\{E(\chi^{-2}_{k+2,\alpha}(\Delta))\}^2. \tag{4.20}$$

$$\begin{aligned}B_5 = (k-2)^2\Delta[\{&E(\chi^{-2}_{k+2,\alpha}(\Delta)) + H_{k+2}(\chi^2_{k,\alpha};\Delta) \\ -\ &E[(\chi^{-2}_{k+2,\alpha}(\Delta))I((\chi^2_{k+2}(\Delta)) < k-2)]\}]^2.\end{aligned} \tag{4.21}$$

Now, the MSE matrices of the estimators are

$$D_1 = (X^T V X)^{-1}. \tag{4.22}$$

$$D_2 = (1-c)^2 D_1 + c^2 \delta\delta^{\mathbf{T}}. \tag{4.23}$$

$$D_3 = D_1 - c(2-c) D_1 H_{k+2}(\chi^2_{k,\alpha};\Delta) + c\delta\delta^{\mathbf{T}}\{2H_{k+2}(\chi^2_{k,\alpha};\Delta) - (2-c)H_{k+4}(\chi^2_{k,\alpha};\Delta)\}. \tag{4.24}$$

$$D_4 = D_1 - (k-2)D_1\{E(\chi^{-2}_{k+2}(\Delta)) + \Delta E(\chi^{-4}_{k+4}(\Delta))\} + (k^2-4)\delta\delta^{\mathbf{T}} E(\chi^{-4}_{k+4}(\Delta)). \tag{4.25}$$

$$\begin{aligned} D_5 = {} & D_4 + (k-2)D_1[2E\{(\chi^{-2}_{k+2}(\Delta))I(\chi^2_{k+2}(\Delta) < k-2)\} \\ - {} & (k-2)E(\chi^{-4}_{k+2}(\Delta))I(\chi^2_{k+2}(\Delta) < k-2)\}] \\ - {} & [D_1 H_{k+2}(k-2;\Delta) - \delta\delta^{\mathbf{T}}\{2H_{k+2}(k-2;\Delta) - H_{k+4}(k-2;\Delta)\}] \\ - {} & (k-2)\delta\delta^{\mathbf{T}}[2E\{(\chi^{-2}_{k+2}(\Delta))I(\chi^2_{k+2}(\Delta) < k-2)\} \\ - {} & 2E(\chi^{-4}_{k+2}(\Delta))I(\chi^2_{k+2}(\Delta) < k-2)\} \\ + {} & (k-2)E(\chi^{-4}_{k+4}(\Delta))I(\chi^2_{k+4}(\Delta) < k-2)\}]. \end{aligned} \tag{4.26}$$

Finally, the risks of the estimators are

$$R(\widetilde{\beta}_n, Q) = tr(QD_1). \tag{4.27}$$

$$R(\widehat{\beta}^{(c)}_n, Q) = (1-c)^2 tr(QD_1) + c^2 \delta^{\mathbf{T}} Q\delta. \tag{4.28}$$

$$\begin{aligned} R(\widehat{\beta}^{SPT}_n, Q) = {} & tr(QD_1) - c(2-c)tr(QD_1)H_{k+2}(\chi^2_{k,\alpha};\Delta) \\ + {} & c(\delta^{\mathbf{T}}Q\delta)\{2H_{k+2}(\chi^2_{k,\alpha};\Delta) - (2-c)H_{k+4}(\chi^2_{k,\alpha};\Delta)\}. \end{aligned} \tag{4.29}$$

$$\begin{aligned} R(\widehat{\beta}^{S}_n, Q) = {} & tr(QD_1) - (k-2)tr(QD_1)\{E(\chi^{-2}_{k+2}(\Delta)) + \Delta E(\chi^{-4}_{k+4}(\Delta))\} \\ + {} & (k^2-4)E(\chi^{-4}_{k+4}(\Delta)). \end{aligned} \tag{4.30}$$

$$\begin{aligned} R(\widehat{\beta}^{S^+}_n, Q) = {} & tr(QD_4) + (k-2)tr(QD_1)[D_4 \\ + {} & (k-2)D_1[2E\{(\chi^{-2}_{k+2}(\Delta))I(\chi^2_{k+2}(\Delta) < k-2)\} \\ - {} & (k-2)E\{(\chi^{-4}_{k+2}(\Delta))I(\chi^2_{k+2}(\Delta) < k-2)\}] \\ - {} & [D_1 H_{k+2}(k-2;\Delta) - \delta\delta^{\mathbf{T}}\{2H_{k+2}(k-2;\Delta) - H_{k+4}(k-2;\Delta)\}] \\ - {} & (k-2)\delta\delta^{\mathbf{T}}[2E\{(\chi^{-2}_{k+2}(\Delta))I(\chi^2_{k+2}(\Delta) < k-2)\} \\ - {} & 2E\{(\chi^{-4}_{k+2}(\Delta))I(\chi^2_{k+2}(\Delta) < k-2)\} \\ + {} & (k-2)E\{(\chi^{-4}_{k+4}(\Delta))I(\chi^2_{k+4}(\Delta) < k-2)\}]] \\ - {} & [tr(QD_1)H_{k+2}(k-2;\Delta) - (\delta^{\mathbf{T}}Q\delta)\{2H_{k+2}(k-2;\Delta) \\ - {} & H_{k+2}(k-2;\Delta)\}] - (k-2)(\delta^{\mathbf{T}}Q\delta)[D_4 \\ + {} & (k-2)D_1[2E\{(\chi^{-2}_{k+2}(\Delta))I(\chi^2_{k+2}(\Delta) < k-2)\} \\ - {} & (k-2)E\{(\chi^{-4}_{k+2}(\Delta))I(\chi^2_{k+2}(\Delta) < k-2)\}] \\ - {} & [D_1 H_{k+2}(k-2;\Delta) - \delta\delta^{\mathbf{T}}\{2H_{k+2}(k-2;\Delta) - H_{k+4}(k-2;\Delta)\}] \\ - {} & (k-2)\delta\delta^{\mathbf{T}}[2E\{(\chi^{-2}_{k+2}(\Delta))I(\chi^2_{k+2}(\Delta) < k-2)\} \\ - {} & 2E\{(\chi^{-4}_{k+2}(\Delta))I(\chi^2_{k+2}(\Delta) < k-2)\} \\ + {} & (k-2)E\{(\chi^{-4}_{k+4}(\Delta))I(\chi^{-2}_{k+4}(\Delta) < k-2)\}]]. \end{aligned} \tag{4.31}$$

If $Q = D_1^{-1}$ in 4.27-4.31, we have

$$R(\widetilde{\beta}_n, Q) = k. \tag{4.32}$$

$$R(\widehat{\beta}_n^{(c)}, Q) = (1-c)^2 k + c^2\Delta. \tag{4.33}$$

$$\begin{aligned} R(\widehat{\beta}_n^{SPT}, Q) = & \{1 - c(2-c)H_{k+2}(\chi^2_{k,\alpha}; \Delta)\}k + c\Delta\{2H_{k+2}(\chi^2_{k,\alpha}; \Delta) \\ - & \ (2-c)H_{k+4}(\chi^2_{k,\alpha}; \Delta)\}. \end{aligned} \tag{4.34}$$

$$R(\widehat{\beta}_n^{S}, Q) = k - k(k-2)\{E(\chi^{-2}_{k+2}(\Delta)) + \Delta E(\chi^{-4}_{k+4}(\Delta))\} + (k^2-4)E(\chi^{-4}_{k+4}(\Delta)). \tag{4.35}$$

$$\begin{aligned} R(\widehat{\beta}_n^{S^+}, Q) = & R(\widehat{\beta}_n^{S}, Q) - [kH_{k+2}(k-2; \Delta) \\ - & \ D\{2H_{k+2}(k-2;\Delta) - H_{k+4}(k-2;\Delta)\}] \\ - & \ (k-2)\Delta[2E\{(\chi^{-2}_{k+2}(\Delta))I((\chi^2_{k+2}(\Delta)) < k-2)\} \\ - & \ 2E\{(\chi^{-2}_{k+4}(\Delta))I(\chi^2_{k+2}(\Delta) < k-2)\} \\ + & \ E\{(\chi^{-4}_{k+4}(\Delta))I((\chi^2_{k+2}(\Delta) < k-2)\}] \\ + & \ k(k-2)[2E\{(\chi^{-2}_{k+2}(\Delta))I((\chi^2_{k+2}(\Delta)) < k-2)\} \\ - & \ (k-2)E\{(\chi^{-4}_{k+2}(\Delta))I(\chi^2_{k+2}(\Delta) < k-2)\}], \end{aligned} \tag{4.36}$$

respectively.

5 Asymptotic Distributional Risk and MSE Analysis for the Estimators

In this section, we carry out the analysis of the asymptotic distributional risk (ADR) and MSE to compare the five estimators proposed in section 3. First, we compare $\widetilde{\beta}_n$ (unrestricted estimator) with the rest.

Comparison of $\widetilde{\beta}_n$ and $\widehat{\beta}_n^{(c)}$: The risk difference is given by

$$R(\widetilde{\beta}_n, Q) - R(\widehat{\beta}_n^{(c)}, Q) = tr(QD_1) - (1-c)^2 tr(QD_1) - c^2\delta^T Q\delta. \tag{5.1}$$

The expression $R(\widehat{\beta}_n^{(c)}, Q) - R(\widetilde{\beta}_n, Q) \gtreqless 0$ according as

$$\delta^T Q\delta \lesseqgtr (\frac{2}{c} - 1)tr(QD_1). \tag{5.2}$$

In terms of the noncentrality parameter, we obtain

$$R(\widetilde{\beta}_n, Q) - R(\widehat{\beta}_n^{(c)}, Q) > 0 \quad \text{if } \Delta < (\frac{2}{c} - 1)\frac{tr(QD_1)}{Ch_{max}(QD_1)}$$

and

$$R(\widetilde{\beta}_n, Q) - R(\widehat{\beta}_n^{(c)}, Q) < 0 \quad \text{if } \Delta > (\frac{2}{c} - 1)\frac{tr(QD_1)}{Ch_{min}(QD_1)}$$

where $Ch_{max}(A)$ ($Ch_{min}(A)$) stands for the maximum (minimum) characteristic root of a given matrix A. If $Q = D_1^{-1}$, then $tr(QD_1) = k$ and $Ch_{max}(A) = Ch_{min}(A) = 1$. Hence

$$R(\widetilde{\beta}_n, Q) - R(\widehat{\beta}_n^{(c)}, Q) \gtreqless 0 \text{ whenever } \Delta \lesseqgtr (\frac{2}{c} - 1)k. \tag{5.3}$$

The analysis means that the restricted estimator $\widehat{\beta}_n^{(c)}$ is superior to the unrestricted estimator $\widetilde{\beta}_n$ whenever $\Delta < (\frac{2}{c} - 1)\frac{tr(QD_1)}{Ch_{max}(QD_1)}$ and inferior whenever $\Delta > (\frac{2}{c} - 1)\frac{tr(QD_1)}{Ch_{min}(QD_1)}$. For $Q = D_1^{-1}$ the same result hold whenever $\Delta \lesseqgtr (\frac{2}{c} - 1)k$. Finally, for $c = 1$, we obtain a smaller range on the superiority of $\widehat{\beta}_n^{(1)}$. Thus, $\widehat{\beta}_n^{(c)}$ is preferred over $\widehat{\beta}_n^{(1)}$ compared to $\widetilde{\beta}_n$.

For the MSE comparison of $\widetilde{\beta}_n$ and $\widehat{\beta}_n^{(c)}$ we consider the MSE difference

$$D_1 - D_2 = D_1 - (1-c)^2 D_1 - c^2 \delta\delta^T. \tag{5.4}$$

This difference is positive semi-definite whenever for any non-zero vector $\mathbf{l} = (l_1, \ldots, l_k)^T$, we have $\mathbf{l}^T(D_1 - D_2)\mathbf{l} \geq 0$. That is to say,

$$(\frac{2}{c} - 1)\mathbf{l}^T D_1 \mathbf{l} \geq \mathbf{l}^T \delta\delta^T \mathbf{l}. \tag{5.5}$$

Since D_1 is positive definite, (5.5) is equivalent to the requirement that

$$\frac{\mathbf{l}^T \delta\delta^T \mathbf{l}}{(\frac{2}{c} - 1)\mathbf{l}^T D_1 \mathbf{l}} \leq 1. \tag{5.6}$$

Now, $\max_l \frac{\mathbf{l}^T \delta\delta^T \mathbf{l}}{\mathbf{l}^T D_1 \mathbf{l}} = \delta^T D_1^{-1} \delta = \Delta$. Hence, (5.6) is equivalent to

$$\Delta \leq (\frac{2}{c} - 1). \tag{5.7}$$

Thus, the restricted estimator $\widehat{\beta}_n^{(c)}$ is superior to the unrestricted estimator $\widetilde{\beta}_n$ whenever $\Delta \leq (\frac{2}{c} - 1)$, otherwise $\widetilde{\beta}_n$ is superior. Notice that the range of Δ for the superiority of $\widehat{\beta}_n^{(c)}$ is smaller here compared to the range of Δ based on ADR analysis.

Comparison of $\widetilde{\beta}_n$ and $\widehat{\beta}_n^{SPT}$: Considering the risk difference we have

$$\begin{aligned} R(\widetilde{\beta}_n, Q) - R(\widehat{\beta}_n^{SPT}, Q) &= c(2-c)tr(QD_1)H_{k+2}(\chi^2_{k,\alpha}; \Delta) \\ &- \; c(\delta^T Q \delta)\{2H_{k+2}(\chi^2_{k,\alpha}; \Delta) - (2-c)H_{k+4}(\chi^2_{k,\alpha}; \Delta)\}. \end{aligned}$$

This risk difference is positive, whenever

$$\Delta < \frac{(2-c)kH_{k+2}(\chi^2_{k,\alpha}; \Delta)}{Ch_{max}(QD_1)\{2H_{k+2}(\chi^2_{k,\alpha}; \Delta) - (2-c)H_{k+4}(\chi^2_{k,\alpha}; \Delta)\}} \leq (\frac{2}{c} - 1)\frac{tr(QD_1)}{Ch_{max}(QD_1)} \tag{5.8}$$

and negative whenever

$$\Delta > \frac{(2-c)kH_{k+2}(\chi^2_{k,\alpha}; \Delta)}{Ch_{min}(QD_1)\{2H_{k+2}(\chi^2_{k,\alpha}; \Delta) - (2-c)H_{k+4}(\chi^2_{k,\alpha}; \Delta)\}} \geq (\frac{2}{c} - 1)\frac{tr(QD_1)}{Ch_{min}(QD_1)} \tag{5.9}$$

That is to say, $\widehat{\beta}_n^{SPT}$ is superior to $\widetilde{\beta}_n$ whenever (5.8) holds and inferior whenever (5.9) holds. Under the null hypothesis H_0,

$$R(\widetilde{\beta}_n, Q) - R(\widehat{\beta}_n^{SPT}, Q) = c(2-c)tr(QD_1)H_{k+2}(\chi^2_{k,\alpha}; 0) > 0, \forall\ \alpha \in (0,1).$$

Further, $R(\widehat{\beta}_n^{SPT}, Q)$ takes the smallest value $tr(QD_1)\{1 - c(2-c)H_{k+2}(\chi^2_{k,\alpha}; 0)\} > 0$, then increases crossing the ADR of $\widetilde{\beta}_n$ to a maximum at $\Delta = \Delta_\alpha^{max}$ thereafter decreases gradually towards the ADR of $\widetilde{\beta}_n$ as $\Delta \to \infty$. This feature of $\widehat{\beta}_n^{SPT}$ and $\widetilde{\beta}_n$ indicates that neither $\widetilde{\beta}_n$ nor $\widehat{\beta}_n^{SPT}$ is admissible with respect to each other. Notice further that the range of Δ (5.8) with $c = 1$ is smaller than (5.8) for any $c \in (0,1)$. Thus SPTE, $\widehat{\beta}_n^{(SPT)}$ is superior to PTE($c = 1$) under ADR criterion.

Similarly, we obtain the MSE difference

$$\begin{aligned} D_1 - D_3 = &\ c(2-c)D_1 H_{k+2}(\chi^2_{k,\alpha}; \Delta) \\ &- \ c(\delta^T\delta)\{2H_{k+2}(\chi^2_{k,\alpha}; \Delta) - (2-c)H_{k+4}(\chi^2_{k,\alpha}; \Delta)\}. \end{aligned} \tag{5.10}$$

This difference is positive semi-definite whenever for any nonzero vector $\mathbf{l} = (l_1, \ldots, l_k)^T$, we have

$$c(2-c)\mathbf{l}^T D_1 \mathbf{l} H_{k+2}(\chi^2_{k,\alpha}; \Delta) - c\mathbf{l}^T\delta\delta^{\mathbf{T}}\mathbf{l}\{2H_{k+2}(\chi^2_{k,\alpha}; \Delta) - (2-c)H_{k+4}(\chi^2_{k,\alpha}; \Delta)\} \geq 0 \tag{5.11}$$

which implies that the MSE difference is positive semi-definite whenever

$$\Delta \leq \frac{(2-c)kH_{k+2}(\chi^2_{k,\alpha}; \Delta)}{2H_{k+2}(\chi^2_{k,\alpha}; \Delta) - (2-c)H_{k+4}(\chi^2_{k,\alpha}; \Delta)} \leq (\frac{2}{c} - 1). \tag{5.12}$$

Thus, $\widehat{\beta}_n^{SPT}$ is superior to $\widetilde{\beta}_n$ in this range of Δ, otherwise $\widetilde{\beta}_n$ is superior. Notice the difference between (5.8) and (5.12).

Comparison of $\widetilde{\beta}_n$ and $\widehat{\beta}_n^S$: Here, the risk difference is given by

$$\begin{aligned} R(\widetilde{\beta}_n, Q) - R(\widehat{\beta}_n^S, Q) = &\ (k-2)tr(QD_1)\{E(\chi^{-2}_{k+2}; (\Delta)) \\ &+ \ \Delta E(\chi^{-4}_{k+4}; (\Delta))\} - (\delta^T Q\delta)(k^2-4)E(\chi^{-4}_{k+4}(\Delta)) \end{aligned} \tag{5.13}$$

$$= (k-2)tr(QD_1)[2(k-2)E(\chi^{-4}_{k+2}(\Delta)) + 2\Delta E(\chi^{-4}_{k+4}(\Delta))\{1 - \frac{\delta^T Q\delta(k-2)}{2\Delta tr(QD_1)}\}].$$

Thus, $R(\widetilde{\beta}_n, Q) - R(\widehat{\beta}_n^S, Q) \geq 0$ whenever for *all* Q satisfying

$$\frac{tr(QD_1)}{Ch_{max}(QD_1)} \geq \frac{k+2}{2}\ \forall\ \Delta \tag{5.14}$$

holds. For $Q = D_1^{-1}$, $tr(QD_1) = k$ and (5.14) holds. Therefore, $\widehat{\beta}_n^S$ dominates $\widetilde{\beta}_n$ uniformly whenever (5.14) is satisfied.

Regarding the MSE difference, we obtain

$$D_1 - D_4 = (k-2)D_1\{E(\chi_{k+2}^{-2}(\Delta)) + \Delta E(\chi_{k+4}^{-4}(\Delta))\} - (\delta\delta^{\mathbf{T}})(k^2-4)E(\chi_{k+4}^{-4}(\Delta)). \quad (5.15)$$

This difference is positive semi-definite whenever for any nonzero vector $\mathbf{l} = (l_1, \ldots, l_k)^T$, we have

$$\mathbf{l}^T D_1 \mathbf{l}\{E(\chi_{k+2}^{-2}(\Delta)) + \Delta E(\chi_{k+4}^{-4}(\Delta))\} - (\mathbf{l}^T \delta\delta^T \mathbf{l})(k+2)E(\chi_{k+4}^{-4}(\Delta)) \geq 0. \quad (5.16)$$

This implies that we must have (after simplification)

$$(k+1)(k-2)E(\chi_{k+2}^{-4}(\Delta)) \geq kE(\chi_{k+2}^{-2}(\Delta)) \quad (5.17)$$

which holds for all $\Delta \in (0, \infty)$. Thus, $\widehat{\beta}_n^S$ is uniformly better than $\widetilde{\beta}_n$ when we consider the MSE criterion.

Comparison of $\widetilde{\beta}_n$ and $\widehat{\beta}_n^{S^+}$: In this case, the risk difference turns out to be

$$\begin{aligned}
&R(\widehat{\beta}_n^S, Q) - R(\widehat{\beta}_n^{S^+}, Q) = tr(QD_1)[2H_{k+2}(k-2;\Delta) \\
&- \ (k-2)\{2\sum_{r=0}^{\infty} e^{-1/2\Delta}(1/2\Delta)^r \frac{1}{r!}\frac{1}{k+2r} H_{k+2r}(k-2;0)\} \\
&- \ (k-2)^2\{\sum_{r=0}^{\infty} e^{-1/2\Delta}(1/2\Delta)^r \frac{1}{r!}\frac{1}{(k+2r)(k+2+2r)} H_{k+2r}(k-2;0)\}] \\
&- \ \delta^T Q\delta[2H_{k+2r}(k-2;\Delta) - 2(k-2)\{\sum_{r=0}^{\infty} e^{-1/2\Delta}(1/2\Delta)^r \frac{1}{r!}\frac{1}{k+2r} H_{k+2r}(k-2;0)\} \\
&- \ \{H_{k+4}(k-2;\Delta) - 2(k-2)(\sum_{r=0}^{\infty} e^{-1/2\Delta}(1/2\Delta)^r \frac{1}{r!}\frac{1}{(k+2r)(k+2+2r)} H_{k+2r}(k-2;0)) \\
&+ \ 2(k-2)(\sum_{r=0}^{\infty} e^{-1/2\Delta}(1/2\Delta)^r \frac{1}{r!}\frac{1}{(k+2+2r)(k+4+2r)} H_{k+2+2r}(k-2;0))\}].
\end{aligned} \quad (5.18)$$

The right hand side of (5.18) is non-negative for all Δ. Hence, $\widehat{\beta}_n^{S^+}$ is uniformly superior to $\widehat{\beta}_n^S$. Therefore, when we consider the three estimators $\widetilde{\beta}_n$, $\widehat{\beta}_n^S$ and $\widehat{\beta}_n^{S^+}$ we find the risk ordering as

$$R(\widehat{\beta}_n^{S^+}, Q) < R(\widehat{\beta}_n^S, Q) < R(\widetilde{\beta}_n, Q) \quad (5.19)$$

for all Δ. The risk of $R(\widehat{\beta}_n^S, Q)$ ($R(\widehat{\beta}_n^{S^+}, Q)$) tends to $R(\widetilde{\beta}_n, Q)$ as $\Delta \to \infty$ from below while the risk of $R(\widehat{\beta}_n^{SPT}, Q) \to R(\widetilde{\beta}_n, Q)$ as $\Delta \to \infty$ from above. Similar conclusion holds with respect to the MSE criterion since

$$D_4 - D_5 = D_1[2H_{k+2}(k-2;\Delta) - (k-2)\{2\sum_{r=0}^{\infty} e^{-1/2\Delta}(1/2\Delta)^r \frac{1}{r!}\frac{1}{k+2r} H_{k+2r}(k-2;0)\}$$
$$-\ (k-2)^2\{\sum_{r=0}^{\infty} e^{-1/2\Delta}(1/2\Delta)^r \frac{1}{r!}\frac{1}{(k+2r)(k+2+2r)} H_{k+2r}(k-2;0)\}]$$
$$-\ \delta\delta^{\mathbf{T}}[2H_{k+2}(k-2;\Delta) - 2(k-2)\{\sum_{r=0}^{\infty} e^{-1/2\Delta}(1/2\Delta)^r \frac{1}{r!}\frac{1}{k+2r} H_{k+2r}(k-2;0)\}$$
$$-\ \{H_{k+4}(k-2;\Delta) - 2(k-2)(\sum_{r=0}^{\infty} e^{-1/2\Delta}(1/2\Delta)^r \frac{1}{r!}\frac{1}{(k+2r)(k+2+2r)} H_{k+2r}(k-2;0))$$
$$+\ 2(k-2)(\sum_{r=0}^{\infty} e^{-1/2\Delta}(1/2\Delta)^r \frac{1}{r!}\frac{1}{(k+2r)(k+2+2r)} H_{k+4+2r}(k-2;0))\}]. \tag{5.20}$$

We know $\underset{\alpha}{max}\frac{\mathbf{1}^T\delta\delta^T\mathbf{1}}{\mathbf{1}^T D_1\mathbf{1}} = \Delta$. Therefore, $\mathbf{l}^T(D_1 - D_2)\mathbf{l} \geq 0$ for all Δ as before. Hence, $D_1 - D_4 \geq \mathbf{0}$ and $D_4 - D_5 \geq \mathbf{0}$ and we can order the MSE matrices as

$$D_1 \geq D_4 \geq D_5$$

where $\geq$ stands for $D_1 - D_4 \geq \mathbf{0}$ and $D_4 - D_5 \geq \mathbf{0}$.

Comparison of $\widehat{\beta}_n^{(c)}$ and $\widehat{\beta}_n^{SPT}$: The risk difference and MSE difference are given by

$$\begin{aligned} R(\widehat{\beta}_n^{(c)}, Q) - R(\widehat{\beta}_n^{SPT}, Q) =& -\ c(2-c)tr(QD_1)\{1 - H_{k+2}(\chi^2_{k,\alpha};\Delta)\} \\ &+\ c(\delta^T Q\delta)\{c - 2H_{k+2}(\chi^2_{k,\alpha};\Delta) - (2-c)H_{k+4}(\chi^2_{k,\alpha};\Delta)\} \end{aligned} \tag{5.21}$$

and

$$\begin{aligned} D_2 - D_3 =& -c(2-c)D_1\{1 - H_{k+2}(\chi^2_{k,\alpha};\Delta)\} \\ &+\ c\delta\delta^{\mathbf{T}}\{c - 2H_{k+2}(\chi^2_{k,\alpha};\Delta) - (2-c)H_{k+4}(\chi^2_{k,\alpha};\Delta) \end{aligned} \tag{5.22}$$

respectively. Now, $R(\widehat{\beta}_n^{(c)}, Q) - R(\widehat{\beta}_n^{SPT}, Q) \gtreqless 0$ whenever

$$(\delta^{\mathbf{T}} Q\delta) \gtreqless \frac{(2-c)\{1 - H_{k+2}(\chi^2_{k,\alpha};\Delta)\}tr(QD_1)}{\{c - 2H_{k+2}(\chi^2_{k,\alpha};\Delta) - (2-c)H_{k+4}(\chi^2_{k,\alpha};\Delta)\}}. \tag{5.23}$$

In terms of Δ, we have $R(\widehat{\beta}_n^{(c)}, Q) - R(\widehat{\beta}_n^{SPT}, Q) > 0$ whenever

$$\Delta > \frac{(\frac{2}{c} - 1)tr(QD_1)\{1 - H_{k+2}(\chi^2_{k,\alpha};\Delta)\}}{Ch_{min}(QD_1)\{1 - \frac{2}{c}H_{k+2}(\chi^2_{k,\alpha};\Delta) - (\frac{2}{c} - 1)H_{k+4}(\chi^2_{k,\alpha};\Delta)\}} \tag{5.24}$$

and $R(\widehat{\beta}_n^{(c)}, Q) - R(\widehat{\beta}_n^{SPT}, Q) < 0$ whenever

$$\Delta < \frac{(\frac{2}{c} - 1)tr(QD_1)\{1 - H_{k+2}(\chi^2_{k,\alpha};\Delta)\}}{Ch_{max}(QD_1)\{1 - \frac{2}{c}H_{k+2}(\chi^2_{k,\alpha};\Delta) - (\frac{2}{c} - 1)H_{k+4}(\chi^2_{k,\alpha};\Delta)\}}. \tag{5.25}$$

For $Q = D_1^{-1}$, (5.24) and (5.25) equal each other, and for $c = 1$, we get the analysis for usual PTE. This analysis implies that $\widehat{\beta}_n^{(c)}$ is superior to $\widehat{\beta}_n^{SPT}$ whenever (5.25) holds, otherwise (5.24) implies that $\widehat{\beta}_n^{SPT}$ is superior. So near the null hypothesis $\widehat{\beta}_n^{(c)}$ dominates $\widehat{\beta}_n^{SPT}$ and outside the interval $\widehat{\beta}_n^{SPT}$ dominates $\widehat{\beta}_n^{(c)}$. Under H_0, we note that $R(\widehat{\beta}_n^{(c)}, Q) - R(\widehat{\beta}_n^{SPT}, Q) < 0$ since

$$R(\widehat{\beta}_n^{(c)}, Q) - R(\widehat{\beta}_n^{SPT}, Q) = -\ c(2-c)tr(QD_1)\{1 - H_{k+2}(\chi_{k,\alpha}; 0)\} < 0. \tag{5.26}$$

Similarly, the MSE analysis tells us that (5.22) is positive semi-definite and for any nonzero vector $\mathbf{l} = (l_1, \dots, l_k)^T$, we have

$$\begin{aligned} &c\mathbf{l}^T\delta\delta^T\mathbf{l}\{c - 2H_{k+2}(\chi^2_{k,\alpha}; \Delta) - (2-c)H_{k+4}(\chi^2_{k,\alpha}; \Delta) \\ &-\ c(2-c)\mathbf{l}^T D_1\mathbf{l}\{1 - H_{k+2}(\chi^2_{k,\alpha}; \Delta)\} \geq 0 \end{aligned} \tag{5.27}$$

which implies that (5.27) holds whenever

$$\Delta > \frac{(\frac{2}{c} - 1)\{1 - H_{k+2}(\chi^2_{k,\alpha}; \Delta)\}}{\{1 - \frac{2}{c}H_{k+2}(\chi^2_{k,\alpha}; \Delta) - (\frac{2}{c} - 1)H_{k+4}(\chi^2_{k,\alpha}; \Delta)\}} \tag{5.28}$$

since $\underset{\alpha}{max} \frac{\mathbf{l}^T\delta\delta^T\mathbf{l}}{\mathbf{l}^T D_1\mathbf{l}} = \Delta$. If $\Delta <$ the right hand side of (5.28) the expression (5.27) is negative definite. Hence, $\widehat{\beta}_n^{(c)}$ is superior to $\widehat{\beta}_n^{SPT}$ whenever (5.28) holds and inferior otherwise, compared to $\widehat{\beta}_n^{(c)}$. Naturally, under the null hypothesis $\widehat{\beta}_n^{(c)}$ is superior to $\widehat{\beta}_n^{SPT}$.

Comparison of $\widehat{\beta}_n^{(c)}$ and $\widehat{\beta}_n^{S}$ ($\widehat{\beta}_n^{S+}$): Considering the risk difference and MSE difference we get

$$\begin{aligned} R(\widehat{\beta}_n^{(c)}, Q) - R(\widehat{\beta}_n^{S}, Q) = &-\ c(2-c)tr(QD_1) + (k-2)tr(QD_1)\{E(\chi^{-2}_{k+2}(\Delta)) \\ &+\ \Delta E(\chi^{-4}_{k+4}(\Delta))\} - (\delta^T Q\delta)\{c^2 - (k^2-4)E(\chi^{-4}_{k+4}(\Delta))\} \end{aligned} \tag{5.29}$$

and

$$\begin{aligned} D_2 - D_4 = &-c(2-c)D_1 + (k-2)D_1\{E(\chi^{-2}_{k+2}(\Delta)) \\ &+\ \Delta E(\chi^{-4}_{k+4}(\Delta))\} - (\delta\delta^{\mathbf{T}})\{c^2 - (k^2-4)E(\chi^{-4}_{k+4}(\Delta))\} \end{aligned} \tag{5.30}$$

respectively.

The risk difference is $\gtreqless 0$ according as

$$(\delta^T Q\delta) \lesseqgtr \frac{tr(QD_1)[c(2-c) - (k-2)\{E(\chi^{-2}_{k+2}(\Delta)) + \Delta E(\chi^{-4}_{k+4}(\Delta))\}]}{\{c^2 - (k^2-4)E(\chi^{-4}_{k+4}(\Delta))\}}. \tag{5.31}$$

In terms of Δ, we get $R(\widehat{\beta}_n^{(c)}, Q) - R(\widehat{\beta}_n^{S}, Q) > 0$ whenever

$$\Delta > \frac{tr(QD_1)[c(2-c) - (k-2)\{E(\chi^{-2}_{k+2}(\Delta)) + \Delta E(\chi^{-4}_{k+4}(\Delta))\}]}{Ch_{min}(QD_1)\{c^2 - (k^2-4)E(\chi^{-4}_{k+4}(\Delta))\}} \tag{5.32}$$

and < 0 whenever

$$\Delta < \frac{tr(QD_1)[c(2-c)-(k-2)\{E(\chi_{k+2}^{-2}(\Delta))+\Delta E(\chi_{k+4}^{-4}(\Delta))\}]}{Ch_{max}(QD_1)\{c^2-(k^2-4)E(\chi_{k+4}^{-4}(\Delta))\}}. \tag{5.33}$$

Thus, $\widehat{\beta}_n^{(c)}$ is superior to $\widehat{\beta}_n^S$ whenever (5.33) is satisfied and is inferior to $\widehat{\beta}_n^S$ whenever (5.32) is satisfied. Under the null hypothesis, the risk difference equals

$$\{(1-\frac{2}{k})-c(2-c)\}tr(QD_1) < 0 \quad \text{for } c = 1. \tag{5.34}$$

Thus, the usual restricted estimator is better than the shrinkage estimator. Similar conclusion holds for the positive-rule shrinkage estimator $\widehat{\beta}_n^{S^+}$. Thus, ordering of the risks under the null hypothesis is given by

$$R(\widehat{\beta}_n^{(c)}, Q) < R(\widehat{\beta}_n^{S^+}, Q) < R(\widehat{\beta}_n^S, Q) < R(\widetilde{\beta}_n, Q). \tag{5.35}$$

The position of the preliminary test estimator may shift from "in between" $R(\widehat{\beta}_n^{(c)}, Q)$ and $R(\widehat{\beta}_n^{S^+}, Q)$ to "in between" $R(\widehat{\beta}_n^S, Q)$ and $R(\widetilde{\beta}_n, Q)$. (See the graph of risks of the five estimators in Figure 1, subsetion 6.2.)

Similar conclusion holds for the MSE matrix analysis. Thus, from these analyses we see that when $k \geq 3$, we should use the shrinkage or positive-rule shrinkage estimator and for $k \leq 2$, it is advisable to use the PTE with $\alpha^* = P(\chi_k^2 > 2)$ as optimum value of α by the Akaike Information Criterion (AIC).

6 Illustration

6.1 Numerical Example

We want to illustrate (provide some idea of) how the proposed estimators actually act in practice. We show three replicates, each of sample size 40, of simulation experiments (see for detail Matin and Saleh (2005)). The parameter values for β_i $i = 0, 1, 2, 3, 4$ were taken to be equal to 2. Further, in order to test the null hypothesis $H_0 : \beta_0 = \beta_1 = \beta_2 = \beta_3 = \beta_4 = 2$ we considered the Wald test statistic. The logistic regression model parameters were estimated by the ML method and then the SRE($c = .50$), SPTE($c = .50$), SPTE($c = 1.00$), SE and SE$^+$ were computed. In order to assess the performance of the various estimators the squared error loss ($= (\widetilde{\beta}_* - 2)^2/var(\widetilde{\beta}_*)$, $\widetilde{\beta}_*$ standing for any estimator) for each of the estimators were also computed. Results are given in Table 1.

For sample 1, the test statistic Wald = 0.88 and $\chi^2_{5,.10} = 9.24$. Thus we are unable to reject the null hypothesis. It is clear that SPTE($c = 1.00$) and SE$^+$ equal the hypothesized value. Furthermore, SRE($c = .50$) and SPTE($c = .50$) are exactly equal. For sample 2, the test statistic Wald = 4.20 and $\chi^2_{5,.10} = 9.24$. Again we are unable to reject the null hypothesis. It is clear that SPTE($c = 1.00$) is equal to the hypothesized value. Furthermore, SRE($c = .50$) and SPTE($c = .50$) are exactly equal. Also, the estimator SE and SE$^+$ are exactly equal. For sample 3, the test statistic Wald = 12.38 and $\chi^2_{5,.10} = 9.24$. We reject the null hypothesis. It is clear that SPTE($c = .50$) and SPTE ($c = 1.00$) are exactly equal. Also, the estimators SE and SE$^+$ are exactly equal.

Table 1: Estimator with Squared Error Loss

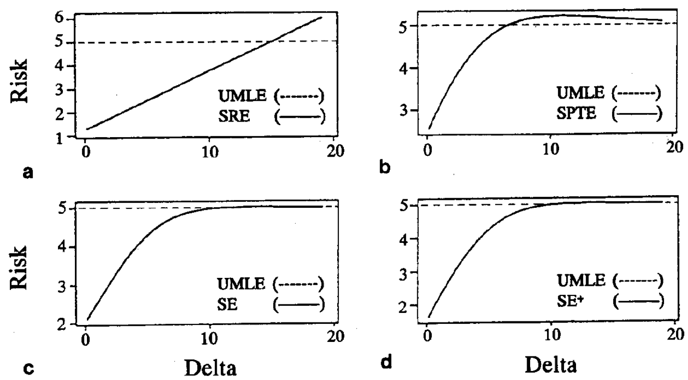
	Estimator						Squared error loss					
	UMLE	SRE (c=.5)	SPTE (c=.5)	SPTE (c=1)	SE	SE$^+$	UMLE	SRE (c=.5)	SPTE (c=.5)	SPTE (c=1)	SE	SE$^+$
Sample 1 $k = 5$ Wald = 0.88 $\chi^2_{5,.10} = 9.24$ H_0 accepted												
β_0	2.302	2.150	2.150	2.000	1.330	2.000	0.121	0.030	0.030	0.000	0.597	0.000
β_1	1.974	1.987	1.987	2.000	2.059	2.000	0.001	0.0002	0.0002	0.000	0.004	0.000
β_2	2.499	2.249	2.249	2.000	0.892	2.000	0.217	0.054	0.054	0.000	1.068	0.000
β_3	1.607	1.803	1.803	2.000	2.873	2.000	0.233	0.058	0.058	0.000	1.148	0.000
β_4	1.882	1.941	1.941	2.000	2.261	2.000	0.017	0.004	0.004	0.000	0.082	0.000
Sample 2 $k = 5$ Wald = 4.20 $\chi^2_{5,.10} = 9.24$ H_0 accepted												
β_0	2.942	2.471	2.471	2.000	2.305	2.305	0.600	0.150	0.150	0.000	0.063	0.063
β_1	2.161	2.080	2.080	2.000	2.052	2.052	0.018	0.004	0.004	0.000	0.002	0.002
β_2	4.882	3.441	3.441	2.000	2.934	2.934	1.329	0.332	0.332	0.000	0.140	0.140
β_3	4.738	3.369	3.369	2.000	2.888	2.888	1.337	0.334	0.334	0.000	0.141	0.141
β_4	4.951	3.475	3.475	2.000	2.957	2.957	0.150	0.487	0.487	0.000	0.205	0.205
Sample 3 k = 5 Wald = 12.38 $\chi^2_{5,.10} = 9.24$ H_0 rejected												
β_0	1.257	1.628	1.257	1.257	1.427	1.427	0.977	0.245	0.977	0.977	0.581	0.581
β_1	1.576	1.788	1.576	1.576	1.673	1.673	0.238	0.060	0.238	0.238	0.142	0.142
β_2	2.168	2.084	2.168	2.168	1.673	1.673	0.034	0.008	0.034	0.034	0.020	0.020
β_3	1.989	1.994	1.989	1.989	1.993	1.993	0.0002	0.0006	0.0002	0.0002	0.0001	0.0001
β_4	0.693	1.374	0.693	0.693	0.993	0.993	5.564	1.389	5.564	5.564	3.303	3.303

It is clear that in sample 1 with a small value of the Wald statistic, the squared error loss is better for the SRE (= SPTE) for $c = .50$ compared to the SE. In sample 2, with a moderately large value of the Wald statistic the SE (= SE$^+$) is better compared to the SRE (= SPTE) for $c = .50$. In sample 3, with a large value of the Wald statistic the SE (= SE$^+$) is better compared to the SPTE but not to the SRE. In general, in all three samples the unrestricted estimator has the higher loss compared to the other estimators.

6.2 Graph Analysis

From the formulas (4.27-4.31) of the risks of the proposed estimators it is clear that their values depend on the two matrices Q and D_1^{-1}. For an ideal situation, we let $Q = D_1^{-1}$ so that $tr(QD_1) = tr(\mathbf{I}_k) = k$. Thus the risk of the unrestricted estimator becomes k and the other formulas in 4.28-4.31 changed accordingly. (These risk formulas are given in 4.32-4.36.) Now, it is easier to compare the other risks with this fixed value k. In doing so the

Figure 1: Risk of the Five Estimators



graph of risks of the five estimators are given in Figure 1.

The risk graphs are drawn as a function of the non-centrality parameter Δ when $k = 5$. For the risk of the unrestricted estimator we observe a parallel line to the X-axis which cuts the Y-axis at $y = 5$ (that is, the risk possesses a constant value 5 for all Δ). Let us call this line the risk line. The graph in (a) portrays the risks for SRE($c = .50$). This risk curve starts at zero, however with the increase in Δ it increases unboundedly. The graph in (b) displays the risk for SPTE($c = .50$). Note that the SPTE($c = .50$) begins at a risk level above 2 and increases sharply as the Δ level increases; crosses the risk line at a certain level of Δ and moves slowly along the risk line but never crosses the line again. The risk curves for SE and SE$^+$ are given in (c) and (d) respectively. We observe that the risk for SE (SE$^+$) initiates at 2 (below 2, actually starts almost from zero) and slowly increases as the Δ level increases but never crosses the risk line at 5.

7 Conclusions

We have presented and discussed (under quadratic loss with local alternatives) five estimators for estimating a parameter vector of interest in logistic regression model in presence of uncertain prior information. Our recommendations as to the choice of estimator are:

(i) if the dimension of the parameter vector is greater than or equal to 3 one should use the shrinkagre estimator or positive-rule shrinkage estimator;

(ii) if the dimension of the parameter vector is less than or equal to 2 the use of the preliminary test estimator is advisable with the optimal level of significance.

A small sample study Matin and Saleh (2005) follows to confirm the same preferences towards the application of the estimators.

Acknowledgement

The work of both the authors sponsored by the NSERC Grant No. A3088, Canada. The first author was also partially funded by the Department of Statistics, Uppsala University, Sweden. Authors would like to thank Reinhold Bergström for his valuable comments on an earlier version of the paper.

References

[1] Agresti, A. (1991). *Categorical data analysis*. John Wiley & Sons.

[2] Albert, J. H. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data, *Journal of the American Statistical Association,* **88***, 669-679.*

[3] Bancroft, T. A. (1944). On biases in estimation due to the use of preliminary tests of significance, *Annals of Mathematical Statisitcs,* **15***, 195-204.*

[4] Bancroft, T. A. and Han, C-P. (1980). Inference based on conditionally specified ANOVA models incorporating preliminary testing. In P. R. Krishnaiah (Ed.), *Handbook of statistics,* **6***, pp. 407-411*, North Holland, New York.

[5] Berger, J. O. (1985). *Statistical decision theory*. Springer-Verlag.

[6] Casella, G. (1985). An introduction to empirical Bayes data analysis, *The American Statistician,* **40***, 124-128.*

[7] Efron, B. (1992). Introduction to James and Stein (1961) estimators with quadratic loss. In S. Kotz and N. L. Johnson

(Eds.), *Breakthrough in statistics, Vol.* **I***, pp. 437-442*, Springer-Verlag, New York.

[8] Efron, B. (1995). The statistical century, *RSS News, Vol.* **22***, No. 5, 1-2.*

[9] Efron, B. and Morris, C. (1973). Stein's estimation rule and its competitors - an empirical Bayes approach, *Journal of the American Statistical Association,* **68***, 117-130.*

[10] Ghosh, M. and Lin, P-E. (1986). Stein-type multivariate estimation, *Communication in Statistics, Theory and Method,* **15***, 2005-2266.*

[11] Han, C-P. and Bancroft, T. A. (1968). On pooling mean when variance is unknown, *Journal of the American Statistical Association,* **63***, 1333-1342.*

[12] Hosmer, D. W. and Lemeshow, S. (1989). *Applied logistic regression*. Wiley Inter Science.

[13] James, W. and Stein, C. (1961). Estimation with quadratic loss. In *Proceedings of Fourth Berkeley Symposium on Mathematical Statistics and Probability,* **1***, pp. 361-379.*

[14] Judge, G. G. and Bock, M. E. (1978). *The statistical implications of pre-test and Stein-rule estimators in econometrics.* North-Holland, Amsterdam.

[15] Matin, M. A. (2005). Small-sample properties of estimators and tests in logistic regression model with skew-normally distributed explanatory variables. *Manuscript.*

[16] Matin, M. A. and Saleh, A. K. Md. E. (2005). Small-sample properties of some improved estimators in logistic regression model with skew-normally distributed explanatory variables. *Manuscript.*

[17] Puri, M. L. and Sen, P. K. (1985). *Nonparametric methods in general linear models.* John Wiley & Sons.

[18] Robbins, H. (1956). An empirical Bayes approach to statistics. In *Proceedings of Third Berkeley Symposium on Mathematical Statistics and Probability,* **1**, *pp. 157-164.*

[19] 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.*

[20] Saleh, A. K. Md. E. and Sen, P. K. (1983). Nonparametric tests of location after a preliminary test on regression in the multivariate case, *Communication in Statistics, Theory and Method,* **12**, *1855-1872.*

[21] Saleh, A. K. Md. E. and Sen, P. K. (1984a). Least squares and rank order preliminary test estimation in general multivariate linear models. In J. K. Ghosh and J. Roy (Eds.), *Proceedings of Indian Statistical Institute, Golden Jubilee Conference in Statistical Applications and New Directions, pp. 237-253.*

[22] Saleh, A. K. Md. E. and Sen, P. K. (1984b). Nonparametric preliminary test inference. In P. R. Krishnaiah and P. K. Sen (Eds.), *Handbook of statistics,* **4**, *pp. 275-297.* North-Holland, Amsterdam.

[23] Saleh, A. K. Md. E. and Sen, P. K. (1985a). Nonparametric shrinkage estimation in a parallelism problem, *Sankhya,* **A47**, *156-165.*

[24] Saleh, A. K. Md. E. and Sen, P. K. (1985b). On shrinkage M-estimator of location parameters, *Communication in Statistics, Theory and Method,* **14**, *2313-2329.*

[25] Saleh, A. K. Md. E. and Sen, P. K. (1985c). Preliminary test predicted in general multivariate linear models, In *Proceedings of the Pacific Area Statistical Conference, pp. 619-638.* North-Holland, Amsterdam.

[26] Saleh, A. K. Md. E. and Sen, P. K. (1985d). Shrinkage least squares estimation in a general multivariate linear model. In *Proceedings of the Fifth Pannonian Symposium on Mathematical Statistics, pp. 307-325.*

[27] Saleh, A. K. Md. E. and Sen, P. K. (1986a). On shrinkage least squares estimation in a parallelism problem, *Communications in Statistics, Theory and Method,* **15***, 1451-1466.*

[28] Saleh, A. K. Md. E. and Sen, P. K. (1986b). On shrinkage R-estimation in a multiple regression model, *Communications in Statistics, Theory and Method,* **15***, 2229-2244.*

[29] Saleh, A. K. Md. E. and Sen, P. K. (1987). Estimation of regression parameters under uncertain prior restriction in a subspace. In *Recent development in statistics and actuarial science, pp. 171-187.* New York, SCTEX Publications.

[30] Sen, P. K. (1979): Asymptotic properties of maximum likelihood estimators based on conditional specifications, *Annals of Statistics,* **7***, 1019-1033.*

[31] Sen, P. K. (1986): On the asymptotic distributional risks of shrinkage and preliminary test versions of maximum likelihood estimators, *Sankhya, A48, 354-371.*

[32] Stein, C. (1956). Inadmissibility of the usual estimator for mean of multivariate normal distribution, In *Proceedings of Third Berkeley Symposium on Mathematical Statistics and Probability,* **1***, pp. 197-206.*