Statistica Sinica 1(1991), 271-294

ASYMPTOTIC THEORY FOR SIMULTANEOUS ESTIMATION OF BINOMIAL MEANS

Abdunnabi M. Ali and A. K. Md. Ehsanes Saleh

Carleton University

Abstract: Efron and Morris (1975) considered a James-Stein (1961) estimator to predict the batting averages of major league players after using an arc sine transformation of the batting average. In this paper, the relevant asymptotic theory assuming unequal Bernoulli trials without transformation is considered. Under quadratic loss the unrestricted, restricted, preliminary test and Stein-rule estimators are compared. It is shown that although the Stein-rule estimator dominates the unrestricted estimator uniformly, it does not dominate the preliminary test estimator except for large dimensions and a range of significance levels, while both the Stein-rule and the PTE perform well relative to the unrestricted and restricted estimators.

Key words and phrases: Preliminary test estimator, Stein-rule estimator, asymptotic distributional risk, Bernoulli model.

1. Introduction

Let $x_{i1}, x_{i2}, \dots, x_{in_i}$ $(i = 1, 2, \dots, p)$ be a random sample of size n_i from the ith set of Bernoulli trials with parameters $(1, \theta_i)$. The likelihood function for the parameters given the vector $\boldsymbol{x} = (x_{11}, \dots, x_{1n_1}, \dots, x_{p1}, \dots, x_{pn_p})$ is

$$L(\boldsymbol{x}|\boldsymbol{\theta}) = \prod_{i=1}^{p} \prod_{j=1}^{n_i} \theta_i^{x_{ij}} (1 - \theta_i)^{1 - x_{ij}}, \tag{1.1}$$

where $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)'$. If we define $n = n_1 + n_2 + \cdots + n_p$ and the null-hypothesis H_0 by

$$H_0 : \theta_1 = \theta_2 = \cdots = \theta_p = \theta_0 \text{ (unknown)}, \tag{1.2}$$

then the likelihood function under H_0 is

$$L(\boldsymbol{x}|\boldsymbol{\theta}) = \theta_0^{\sum_{i=1}^{p} y_i} (1 - \theta_0)^{n - \sum_{i=1}^{p} y_i}, \tag{1.3}$$

where

$$y_i = \sum_{j=1}^{n_i} x_{ij}, \quad i = 1, 2, \dots, p. \tag{1.4}$$

We are primarily interested in the estimation of $\boldsymbol{\theta}$ when we may have uncertain prior information in (1.2).

Let us define

$$\tilde{\boldsymbol{\theta}}_n = (\tilde{\theta}_1, \ldots, \tilde{\theta}_p)', \quad \tilde{\theta}_i = \frac{y_i}{n_i}, \quad i = 1, 2, \ldots, p; \tag{1.5}$$

$$\hat{\boldsymbol{\theta}}_n = (\hat{\theta}_n, \ldots, \hat{\theta}_n)' = \hat{\theta}_n \mathbf{1}_p; \tag{1.6}$$

where

$$\hat{\theta}_n = \frac{1}{n} \sum_{i=1}^{p} y_i, \quad \mathbf{1}_p = (1, 1, \ldots, 1)'. \tag{1.7}$$

Then, for the global model (1.1), the unrestricted maximum likelihood estimator (UMLE) of $\boldsymbol{\theta}$ is given by $\tilde{\boldsymbol{\theta}}_n$, while under the model (1.3), the restricted MLE (RMLE) is given by $\hat{\boldsymbol{\theta}}_n$. When H_0 holds, $\hat{\boldsymbol{\theta}}_n$ has a smaller risk (under quadratic loss) than $\tilde{\boldsymbol{\theta}}_n$. On the other hand, when H_0 does not hold, $\tilde{\boldsymbol{\theta}}_n$ may perform better than $\hat{\boldsymbol{\theta}}_n$. As a result, when the prior information on H_0 is rather uncertain it may be desirable to have either a preliminary test estimator (PTE) denoted by $\hat{\boldsymbol{\theta}}_n^{\mathrm{PT}} = (\hat{\theta}_1^{\mathrm{PT}}, \ldots, \hat{\theta}_p^{\mathrm{PT}})'$ or a shrinkage estimator (SE) denoted by $\hat{\boldsymbol{\theta}}_n^{\mathrm{S}} = (\hat{\theta}_1^{\mathrm{S}}, \ldots, \hat{\theta}_p^{\mathrm{S}})'$ of $\boldsymbol{\theta}$. In the case of a preliminary test the estimator $\hat{\boldsymbol{\theta}}_n$ or $\tilde{\boldsymbol{\theta}}_n$ is chosen according as H_0 is accepted or rejected. The SE is based on the classical Stein-rule. Unlike the case of $\hat{\boldsymbol{\theta}}_n$, both the PTE and SE behave robustly against departures from the null hypothesis in (1.2) (see e.g. Sen (1986)) and have bounded risk even when n is very large. From these four basic estimators two more improved estimators may be defined; namely, Stein's positive-rule shrinkage estimator (PRSE) denoted by $\hat{\boldsymbol{\theta}}_n^{\mathrm{S+}} = (\hat{\theta}_1^{\mathrm{S+}}, \ldots, \hat{\theta}_p^{\mathrm{S+}})'$ and the modified PTE (MPTE) denoted by $\hat{\boldsymbol{\theta}}_n^{\mathrm{PT+}} = (\hat{\theta}_1^{\mathrm{PT+}}, \ldots, \hat{\theta}_p^{\mathrm{PT+}})'$. The primary objective of this paper is to focus on the PTE and SE and to compare them with the UMLE and RMLE.

We shall find it convenient to formulate this problem in an asymptotic setup. For general linear models, the asymptotic theory of the PTE has been treated in a nonparametric setup in Saleh and Sen (1978, 1983, 1984a, 1984b, 1985a, 1985b, 1985c) and Sen and Saleh (1979, 1985, 1987). Also, the asymptotic theory of shrinkage estimation has been developed in a nonparametric setup by the same authors among others. Finite sample studies on the subject have been carried out by Albert (1984), Albert and Gupta (1981) among others. For some nice accounts of the parametric theory of the PTE and the SE in the finite sample case, we refer to Judge and Bock (1978), Anderson (1984), Arnold (1981) and Berger (1980) among others.

Along with preliminary notions, the proposed estimators are all presented in Section 2. The concept of asymptotic risk (AR) and its relation to asymptotic distributional risk (ADR) is outlined in Section 3. Asymptotic bias (AB) and ADR results for the various estimators of $\boldsymbol{\theta}$ are presented in Section 4; and the related asymptotic dominance results and efficiencies are considered in Section 5. Data analysis of Efron and Morris (1975) baseball data are studied in Section 6.

2. The Proposed Estimators

For the preliminary test (PT) on H_0 in (1.2), we consider the chi-square test statistic

$$L_n = n(\tilde{\boldsymbol{\theta}}_n - \hat{\theta}_n \mathbf{1}_p)'\hat{\boldsymbol{\Sigma}}_n^{-1}(\tilde{\boldsymbol{\theta}}_n - \hat{\theta}_n \mathbf{1}_p), \tag{2.1}$$

where

$$\hat{\boldsymbol{\Sigma}}_n^{-1} = \frac{1}{\hat{\theta}_n(1-\hat{\theta}_n)}\boldsymbol{\Lambda}, \quad \boldsymbol{\Lambda} = \mathrm{Diag}(\lambda_i),$$
$$\lambda_{ni} = \frac{n_i}{n} \quad \text{and} \quad \lim_{n\to\infty} \lambda_{ni} = \lambda_i. \tag{2.2}$$

Under H_0, for large n, L_n closely follows the central chi-square distribution with $(p-1)$ degrees of freedom (DF). Thus, for a given *level of significance* α $(0 < \alpha < 1)$, the critical value of L_n may be approximated by $\chi^2_{p-1,\alpha}$, the upper $100\alpha\%$ point of the chi-square distribution with $(p-1)$ DF. Then, the PTE of $\boldsymbol{\theta}$ is defined by

$$\hat{\boldsymbol{\theta}}_n^{\mathrm{PT}} = \hat{\theta}_n \mathbf{1}_p + [1 - I(L_n \le \chi^2_{p-1,\alpha})](\tilde{\boldsymbol{\theta}}_n - \hat{\theta}_n \mathbf{1}_p), \tag{2.3}$$

where $I(A)$ is the indicator function for the set A.

To introduce the SE we follow Arnold (1981) and propose

$$\hat{\boldsymbol{\theta}}_n^{\mathrm{S}} = \hat{\theta}_n \mathbf{1}_p + [1 - (p-3)L_n^{-1}](\tilde{\boldsymbol{\theta}}_n - \hat{\theta}_n \mathbf{1}_p), \tag{2.4}$$

where we need to assume that $p > 3$.

Note that whereas in (2.3) we have a dichotomous indicator function, in (2.4), $(p-3)L_n^{-1}$ may be regarded as a smoother version. Thus, the SE might be considered from a "pre-test" point of view. Also, (2.3) is a convex combination of $\tilde{\boldsymbol{\theta}}_n$ and $\hat{\boldsymbol{\theta}}_n$, while (2.4) need not be especially because $(p-3)L_n^{-1}$ may be greater than 1. In the spirit of Sclove et al. (1972), we may introduce the following *positive-rule* SE (PRSE):

$$\hat{\boldsymbol{\theta}}_n^{\mathrm{S}+} = \hat{\theta}_n \mathbf{1}_p + [1 - (p-3)L_n^{-1}]^{+}(\tilde{\boldsymbol{\theta}}_n - \hat{\theta}_n \mathbf{1}_p), \tag{2.5}$$

where $y^+ = \max\{0, y\}$. Note that (2.3) and (2.5) may not agree even if we let $\chi^2_{p-1,\alpha} = p-3$. The justification of (2.3) lies in the preliminary test of H_0, whereas (2.5) is intended to control possible overshrinking in (2.4).

Finally, we propose the MPTE of $\boldsymbol{\theta}$ as follows:

$$\hat{\boldsymbol{\theta}}_n^{\mathrm{PT}+} = \hat{\theta}_n \mathbf{1}_p + [1 - (p-3)L_n^{-1}]I(L_n \geq \chi^2_{p-1,\alpha})(\tilde{\boldsymbol{\theta}}_n - \hat{\theta}_n \mathbf{1}_p), \tag{2.6}$$

where the restriction $\chi^2_{p-1,\alpha} < (p-3)$ is required. This estimator uniformly dominates $\hat{\boldsymbol{\theta}}_n^{\mathrm{PT}}$ but now we have the restriction $p \geq 4$. For $p < 4$ we are left with the PTE to improve over $\hat{\boldsymbol{\theta}}_n$ or $\tilde{\boldsymbol{\theta}}_n$. It is well-known (see e.g. Sclove et al. (1972)) that the positive-rule shrinkage estimator (PRSE) dominates the usual shrinkage estimator (SE), while the MPTE dominates the usual PTE.

In this context, one may also mention the Bayes and empirical Bayes estimators of $\boldsymbol{\theta}$ with a mixture of a product of Beta distributions due to Albert (1984) and Albert and Gupta (1981,1983).

3. Asymptotic Risks and Pitman Alternatives

Let $\boldsymbol{\delta}_n$ be an estimator of $\boldsymbol{\theta}$, $\boldsymbol{W}$ be a positive semi-definite (p.s.d.) matrix, and consider the quadratic loss function

$$L_n(\boldsymbol{\delta}_n, \boldsymbol{\theta}) = n(\boldsymbol{\delta}_n - \boldsymbol{\theta})'\boldsymbol{W}(\boldsymbol{\delta}_n - \boldsymbol{\theta}) = n\mathrm{Tr}[\boldsymbol{W}(\boldsymbol{\delta}_n - \boldsymbol{\theta})(\boldsymbol{\delta}_n - \boldsymbol{\theta})'], \tag{3.1}$$

where $\mathrm{Tr}(\boldsymbol{A})$ =trace of $\boldsymbol{A}$. Then the risk of $\boldsymbol{\delta}_n$ is given by

$$R_n(\boldsymbol{\delta}_n, \boldsymbol{\theta}) = EL_n(\boldsymbol{\delta}_n, \boldsymbol{\theta}) = \mathrm{Tr}(\boldsymbol{W}\boldsymbol{V}_n), \tag{3.2}$$

where

$$\boldsymbol{V}_n = nE(\boldsymbol{\delta}_n - \boldsymbol{\theta})(\boldsymbol{\delta}_n - \boldsymbol{\theta})'.$$

For the p-dimensional unit cube Ω, let $\omega(\subset \Omega)$ be the subspace for which $\boldsymbol{\theta}$ satisfies the hypothesis in (1.2). Then, by virtue of the consistency of the L_n test, we note that for any fixed $\boldsymbol{\theta} \notin \omega$, $L_n \xrightarrow{\mathrm{P}} +\infty$, as $n \to \infty$ ($\xrightarrow{\mathrm{P}}$ means convergence in probability). Thus, for any fixed $\boldsymbol{\theta} \notin \omega$, as n increases $\hat{\boldsymbol{\theta}}_n^{\mathrm{PT}}, \hat{\boldsymbol{\theta}}_n^{\mathrm{S}}, \hat{\boldsymbol{\theta}}_n^{\mathrm{S}+}$ and $\hat{\boldsymbol{\theta}}_n^{\mathrm{PT}+}$ are equivalent in probability to the UMLE $\tilde{\boldsymbol{\theta}}_n$, while $\hat{\boldsymbol{\theta}}_n$ will have unbounded risk. The situation is different, however, when $\boldsymbol{\theta} \in \omega$, i.e., $\boldsymbol{\theta}$ belongs to a shrinking neighborhood of ω. For this reason, we consider a sequence $\{K_{(n)}\}$ of Pitman-alternatives:

$$K_{(n)}: \quad \boldsymbol{\theta}_{(n)} = \theta_0 \mathbf{1}_p + n^{-1/2}\boldsymbol{\xi}; \quad \boldsymbol{\xi} = (\xi_1, \cdots, \xi_p)', \quad \theta_0 \mathbf{1}_p \in \omega, \tag{3.3}$$

where the ξ_i are fixed numbers. Note that for $\boldsymbol{\xi} = \mathbf{0}$, $\boldsymbol{\theta}_{(n)} = \theta_0 \mathbf{1}_p \in \omega$, so that (1.2) is a particular case of (3.3).

For the estimator $\boldsymbol{\delta}_n$, the asymptotic risk (AR) under $K_{(n)}$ is defined by

$$R(\boldsymbol{\delta}; \boldsymbol{\xi}) = \lim_{n\to\infty} R_n(\boldsymbol{\delta}; \boldsymbol{\theta}_{(n)}), \tag{3.4}$$

whenever the limit exists. We introduce the asymptotic distribution function (ADF) of $\boldsymbol{\delta}_n$ by

$$G^*_{(\boldsymbol{\delta})}(\boldsymbol{x}) = \lim_{n\to\infty} P\{n^{\frac{1}{2}}(\boldsymbol{\delta}_n - \boldsymbol{\theta}_{(n)}) \leq \boldsymbol{x}|K_{(n)}\}, \tag{3.5}$$

whenever the limit exists. We let

$$\boldsymbol{V}^*_{(\boldsymbol{\delta})} = \int \cdots \int \boldsymbol{x}\boldsymbol{x}' dG^*_{(\boldsymbol{\delta})}(\boldsymbol{x}). \tag{3.6}$$

Then, the *asymptotic distributional risk* (ADR) of $\boldsymbol{\delta}_n$ is defined by

$$R^*(\boldsymbol{\delta}, \boldsymbol{\xi}) = \mathrm{Tr}(\boldsymbol{W}\boldsymbol{V}^*_{(\boldsymbol{\delta})}). \tag{3.7}$$

One may need extra regularity conditions to evaluate (3.4) rather than (3.7). This point has been explained in detail in various other contexts, e.g. in Sen (1984), Sen and Saleh (1985) and Saleh and Sen (1985), among others. Fortunately, in this specific contexts, e.g., we shall see that both (3.4) and (3.7) can be evaluated under the same set of regularity conditions, although it is a lot simpler to work with (3.7). As such, in Section 4 we shall study the ADR results for the various estimators.

4. Asymptotic Risks and Bias

First, we consider the case of *fixed alternatives.* Note that by (2.3),

$$\begin{aligned} n(\hat{\boldsymbol{\theta}}_n^{\mathrm{PT}} - \tilde{\boldsymbol{\theta}}_n)'\boldsymbol{W}(\hat{\boldsymbol{\theta}}_n^{\mathrm{PT}} - \tilde{\boldsymbol{\theta}}_n) &\\ = n(\tilde{\boldsymbol{\theta}}_n - \hat{\theta}_n \mathbf{1}_p)'\boldsymbol{W}(\tilde{\boldsymbol{\theta}}_n - \hat{\theta}_n \mathbf{1}_p) I(L_n < \chi^2_{p-1,\alpha}) &\\ = \{L_n I(L_n < \chi^2_{p-1,\alpha})\}\{n(\tilde{\boldsymbol{\theta}}_n - \hat{\theta}_n \mathbf{1}_p)'\boldsymbol{W}(\tilde{\boldsymbol{\theta}}_n - \hat{\theta}_n \mathbf{1}_p) L_n^{-1}\} &\\ \leq \{L_n I(L_n < \chi^2_{p-1,\alpha})\}\mathrm{ch}_{\max}(\boldsymbol{W}\boldsymbol{\Lambda}^{-1}) &\\ \leq \{L_n I(L_n < \chi^2_{p-1,\alpha})\}\mathrm{Tr}(\boldsymbol{W}\boldsymbol{\Lambda}^{-1}) &\\ \leq \{L_n I(L_n < \chi^2_{p-1,\alpha})\} \sum_{j=1}^{p} (w_{jj}\lambda_j^{-1}), & \end{aligned} \tag{4.1}$$

where $\mathrm{ch}_{\max}(\boldsymbol{A})$ = largest characteristic root of $\boldsymbol{A}$. Also, note that

$$E\{L_n I(L_n < \chi^2_{p-1,\alpha})|\boldsymbol{\theta} \notin \omega\} \leq \chi^2_{p-1,\alpha} P\{L_n < \chi^2_{p-1,\alpha}|\boldsymbol{\theta} \notin \omega\}.$$

Therefore, by the consistency of the test based on L_n,

$$E\{L_n I(L_n < \chi^2_{p-1,\alpha})|\boldsymbol{\theta} \notin \omega\} \to 0, \quad \text{as} \quad n \to \infty. \tag{4.2}$$

Thus, for fixed $\boldsymbol{\theta} \notin \omega$, $\hat{\boldsymbol{\theta}}_n^{\mathrm{PT}}$ and $\tilde{\boldsymbol{\theta}}_n$ are asymptotically risk-equivalent. A very similar treatment holds for $\hat{\boldsymbol{\theta}}_n^{\mathrm{S}}$, $\hat{\boldsymbol{\theta}}_n^{\mathrm{S}+}$ and $\hat{\boldsymbol{\theta}}_n^{\mathrm{PT}+}$. For $\hat{\boldsymbol{\theta}}_n^{\mathrm{S}}$ we note that on the set $\{L_n > 0\}$

$$
\begin{aligned}
n(\hat{\boldsymbol{\theta}}_n^{\mathrm{S}} - \tilde{\boldsymbol{\theta}}_n)'\boldsymbol{W}(\hat{\boldsymbol{\theta}}_n^{\mathrm{S}} - \tilde{\boldsymbol{\theta}}_n) &= (p-3)^2 L_n^{-2}\{n(\tilde{\boldsymbol{\theta}}_n - \hat{\theta}_n \mathbf{1}_p)'\boldsymbol{W}(\tilde{\boldsymbol{\theta}}_n - \hat{\theta}_n \mathbf{1}_p)\} \\
&\leq (p-3)^2 L_n^{-1} \mathrm{ch}_{\max}(\boldsymbol{W}\boldsymbol{\Lambda}^{-1}),
\end{aligned} \tag{4.3}
$$

while on the set $\{L_n = 0\}$, we have $\hat{\boldsymbol{\theta}}_n^{\mathrm{S}} = \tilde{\boldsymbol{\theta}}_n = \hat{\theta}_n \mathbf{1}_p$. Thus, if we can show that

$$
E\{L_n^{-1} I(L_n > 0) | \boldsymbol{\theta} \notin \omega\} \to 0, \quad \text{as} \quad n \to \infty, \tag{4.4}
$$

then $\hat{\boldsymbol{\theta}}_n^{\mathrm{S}}$ and $\tilde{\boldsymbol{\theta}}_n$ become asymptotically risk-equivalent for every $\boldsymbol{\theta} \notin \omega$. Now L_n is non-negative and for every $\boldsymbol{\theta} \notin \omega$, $L_n \xrightarrow{\mathrm{P}} +\infty$, as $n \to \infty$; hence, to show (4.4) holds, it suffices to show that for every positive $\epsilon > 0$, we have

$$
E\{L_n^{-1} I(0 < L_n \leq \epsilon) | \boldsymbol{\theta} \notin \omega\} \to 0, \quad \text{as} \quad n \to +\infty. \tag{4.5}
$$

We prove (4.5) in the Appendix.

Next, note that for any $\boldsymbol{\theta} \notin \omega$, $\hat{\theta}_n \mathbf{1}_p - \boldsymbol{\theta} \xrightarrow{\mathrm{a.s.}} \eta(\neq 0)$, as $n \to \infty$ ($\xrightarrow{\mathrm{a.s.}}$ means convergence almost surely), so that as $n \to \infty$

$$
n(\hat{\theta}_n \mathbf{1}_p - \boldsymbol{\theta})'\boldsymbol{W}(\hat{\theta}_n \mathbf{1}_p - \boldsymbol{\theta}) \xrightarrow{\mathrm{P}} +\infty, \tag{4.6}
$$

consequently, the asymptotic risk of $\hat{\theta}_n \mathbf{1}_p$, for any $\boldsymbol{\theta} \notin \omega$, approaches $+\infty$. Also, note that

$$
nE(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta})(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta})' = \mathrm{Diag}\left\{ \left(\frac{\theta_1(1-\theta_1)}{\lambda_1}, \ldots, \frac{\theta_p(1-\theta_p)}{\lambda_p} \right) \right\},
$$

so that the asymptotic risk of UMLE is bounded for every $\boldsymbol{\theta} \in \Omega$; hence, we arrive at the following theorem.

Theorem 4.1. *For any fixed $\boldsymbol{\theta} \notin \omega$, $\hat{\theta}_n \mathbf{1}_p$ has asymptotic risk $+\infty$, while $\hat{\boldsymbol{\theta}}_n^{\mathrm{S}}, \hat{\boldsymbol{\theta}}_n^{\mathrm{S}+}, \hat{\boldsymbol{\theta}}_n^{\mathrm{PT}}, \hat{\boldsymbol{\theta}}_n^{\mathrm{PT}+}$ and $\tilde{\boldsymbol{\theta}}_n$ are asymptotically risk-equivalent, with bounded risk.*

Next, we consider case of Pitman alternatives in (3.3). As we pointed out earlier we shall mainly study ADR properties though similar AR properties hold. Note that $\hat{\theta}_n \mathbf{1}_p$ and $\tilde{\boldsymbol{\theta}}_n$ both have non-negative elements bounded by 1, so finite moments of $\sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta})$ and $\sqrt{n}(\hat{\theta}_n \mathbf{1}_p - E\hat{\theta}_n \mathbf{1}_p)$ of all finite orders exist and have finite limit as $n \to \infty$. Also, under $\{K_{(n)}\}$ in (3.3),

$$
E[\hat{\theta}_n \mathbf{1}_p | K_{(n)}] - \boldsymbol{\theta} = O(n^{-\frac{1}{2}}),
$$

so that in this case, $n\|\tilde{\boldsymbol{\theta}}_n - \hat{\theta}_n \mathbf{1}_p\|^2 = O(1)$, in L_∞ norm. Thus, for (2.3)–(2.6), the convergence in distribution will ensure the convergence in second moment, and hence (3.7) will ensure (3.4). For the SE in (2.4), PRSE in (2.5) and MPTE in (2.6) we need a more refined treatment. Note that (4.3) applies as well for $\{K_{(n)}\}$, and hence, if we are able to show (4.4) holds under $\{K_{(n)}\}$ (as well as H_0), then again (3.7) would ensure (3.4). Thus for our purpose it suffices to consider the ADR in (3.7) through the asymptotic distribution in (3.5). We work out the ADR results in this section. For the study of the asmyptotic distribution theory and the ADR results, we assume the usual regularity conditions hold. Further, we define $\theta_0 \mathbf{1}_p$ as in (3.3) so that $\theta_0 \mathbf{1}_p \in \omega$, then we have the following theorem.

Theorem 4.2. *Under $\{K_{(n)}\}$ in (3.3) and assumed regularity conditions, the following hold:*

(i) $\boldsymbol{X}_{(n)} = \sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \theta_0 \mathbf{1}_p) \sim N_p(\boldsymbol{\xi}, \boldsymbol{\Sigma}_0)$, *where* $\boldsymbol{\Sigma}_0 = \theta_0(1-\theta_0)\boldsymbol{\Lambda}^{-1}$ (*$\boldsymbol{\Lambda}$ defined by* (2.2)); (4.7)

(ii) $\boldsymbol{Y}_{(n)} = \sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \hat{\theta}_n \mathbf{1}_p) \sim N_p(\boldsymbol{J\xi}, \boldsymbol{\Sigma}_0 \boldsymbol{J}')$, *where* $\boldsymbol{J} = \boldsymbol{I}_p - \mathbf{1}_p \mathbf{1}_p' \boldsymbol{\Lambda}$ (*$\mathbf{1}_p$ defined by* (1.7)); (4.8)

(iii) $\boldsymbol{Z}_{(n)} = \sqrt{n}(\hat{\theta}_n \mathbf{1}_p - \theta_0 \mathbf{1}_p) \sim N_p(\mathbf{0}, \boldsymbol{B})$, *where* $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_p)'$, $\boldsymbol{B} = \theta_0(1-\theta_0)\mathbf{1}_p\mathbf{1}_p'$, *and we assume* $\boldsymbol{\lambda}'\boldsymbol{\xi} = \mathbf{0}$; (4.9)

(iv)
$$\begin{bmatrix} \boldsymbol{X}_{(n)} \\ \boldsymbol{Y}_{(n)} \end{bmatrix} \sim N_{2p}\left\{ \begin{bmatrix} \boldsymbol{\xi} \\ \boldsymbol{J\xi} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_0 & \boldsymbol{\Sigma}_0 \boldsymbol{J}' \\ \boldsymbol{J\Sigma}_0 & \boldsymbol{\Sigma}_0 \boldsymbol{J}' \end{bmatrix} \right\}; \tag{4.10}$$
$$\begin{bmatrix} \boldsymbol{Z}_{(n)} \\ \boldsymbol{Y}_{(n)} \end{bmatrix} \sim N_{2p}\left\{ \begin{bmatrix} \mathbf{0} \\ \boldsymbol{J\xi} \end{bmatrix}, \begin{bmatrix} \boldsymbol{B} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_0 \boldsymbol{J}' \end{bmatrix} \right\}; \tag{4.11}$$

(v) *L_n is distributed asymptotically as a noncentral chi-square with* $(p-1)$ DF, *and noncentrality parameter*
$$\triangle = \boldsymbol{\delta}' \boldsymbol{\Sigma}_0^{-1} \boldsymbol{\delta}, \quad \boldsymbol{\delta} = \boldsymbol{J\xi}. \tag{4.12}$$

We denote a multinormal distribution function with mean vector $\boldsymbol{\mu}$ and dispersion matrix $\boldsymbol{\Sigma}$ by $\Phi(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ and the corresponding density function by $\phi_p(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$. Also, let $H_p(\boldsymbol{x}; \triangle)$ denote the noncentral chi-square distribution function with p DF and noncentrality parameter $\triangle (\geq 0)$ and $E(\chi_m^{-2r}(\triangle)) = \int_0^\infty x^{-r} dH_m(x; \triangle)$. Note that by (2.3) and Theorem 4.2 we have
$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n^{\mathrm{PT}} - \theta_0 \mathbf{1}_p) = \boldsymbol{X}_{(n)} - \boldsymbol{Y}_{(n)} I(L_n \leq \chi^2_{p-1,\alpha}). \tag{4.13}$$

By (4.10)–(4.13) and the general formulation of Saleh and Sen (1984a), we arrive at the following theorem:

Theorem 4.3. *Under $\{K_{(n)}\}$ in (3.3) and the assumed regularity conditions, the asymptotic distribution of $\sqrt{n}(\hat{\boldsymbol{\theta}}_n^{\mathrm{PT}} - \theta_0 \mathbf{1}_p)$ is given by*

$$G_p^{\mathrm{PT}}(\boldsymbol{x}) = \Phi_p(\boldsymbol{x}; \mathbf{0}, \boldsymbol{B}) H_{p-1}(\chi^2_{p-1,\alpha}; \Delta) + \int \cdots \int_{M(\boldsymbol{\delta})} \Phi_p(\boldsymbol{x} - \boldsymbol{\xi} - \boldsymbol{z}; \mathbf{0}, \boldsymbol{B}) d\Phi_p(\boldsymbol{z}; \mathbf{0}, \Sigma_0 \boldsymbol{J}'), \tag{4.14}$$

where

$$M(\boldsymbol{\delta}) = \{\boldsymbol{z}; (\boldsymbol{z} + \boldsymbol{J\xi})'(\boldsymbol{z} + \boldsymbol{J\xi}) \geq \chi^2_{p-1,\alpha}\}, \tag{4.15}$$

and $\boldsymbol{J}$, $\boldsymbol{B}$ and Δ are all defined as in Theorem 4.2.

Note that the density function corresponding to $G_p^{\mathrm{PT}}(\boldsymbol{x})$ is

$$g_p^{\mathrm{PT}}(\boldsymbol{x}) = \phi_p(\boldsymbol{x}; \mathbf{0}, \boldsymbol{B}) H_{p-1}(\chi^2_{p-1,\alpha}; \Delta) + \int \cdots \int_{M(\boldsymbol{\delta})} \phi_p(\boldsymbol{x} - \boldsymbol{\xi} - \boldsymbol{z}; \mathbf{0}, \boldsymbol{B}) d\phi_p(\boldsymbol{z}; \mathbf{0}, \Sigma_0 \boldsymbol{J}'). \tag{4.16}$$

Next, by virtue of (2.1), (2.4) and Theorem 4.2, we obtain the following theorem:

Theorem 4.4. *Under $\{K_{(n)}\}$ in (3.3) and the assumed regularity conditions,*

$$\begin{aligned} \sqrt{n}(\hat{\boldsymbol{\theta}}_n^{\mathrm{S}} - \theta_0 \mathbf{1}_p) &= \boldsymbol{X}_{(n)} - (p-3)\boldsymbol{Y}_{(n)}\{\boldsymbol{Y}'_{(n)}\Sigma_0^{-1}\boldsymbol{Y}_{(n)}\}^{-1} \\ &\xrightarrow{\mathrm{D}} (\boldsymbol{X} + \boldsymbol{\xi}) - (p-3)(\boldsymbol{Y} + \boldsymbol{J\xi}) \times \\ &\qquad \{(\boldsymbol{Y} + \boldsymbol{J\xi})'\Sigma_0^{-1}(\boldsymbol{Y} + \boldsymbol{J\xi})\}^{-1}, \end{aligned} \tag{4.17}$$

where

$$\begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{Y} \end{bmatrix} \sim N_{2p}\left\{ \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \Sigma_0 & \Sigma_0 \boldsymbol{J}' \\ \boldsymbol{J}\Sigma_0 & \Sigma_0 \boldsymbol{J}' \end{bmatrix} \right\}$$

and $\xrightarrow{\mathrm{D}}$ means convergence in distribution.

Very similar representations hold for the positive rule estimator in (2.5). Under $\{K_{(n)}\}$ in (3.3), and by virtue of the general equivalence result in (3.4) and (3.7) and by Theorems 4.2, 4.3 and 4.4, we arrive at following expressions for the asymptotic bias and risks. Since the derivation of these formula are very similar to the general case treated in Sen and Saleh (1985), we omit these details of the derivation. The asymptotic bias expressions for the first four estimators in (1.5), (1.6), (2.3) and (2.4) are given by

$$\boldsymbol{B}_1(\boldsymbol{\delta}) = \lim_{n\to\infty} E\{\sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{(n)})\} = \mathbf{0}, \tag{4.18}$$

$$\boldsymbol{B}_2(\boldsymbol{\delta}) = \lim_{n\to\infty} E\{\sqrt{n}(\hat{\theta}_n \mathbf{1}_p - \boldsymbol{\theta}_{(n)})\} = -\boldsymbol{\delta}, \tag{4.19}$$

$$\boldsymbol{B}_3(\boldsymbol{\delta}) = \lim_{n\to\infty} E\{\sqrt{n}(\hat{\boldsymbol{\theta}}_n^{\mathrm{PT}} - \boldsymbol{\theta}_{(n)})\} = -\boldsymbol{\delta} H_{p+1}(\chi^2_{p-1,\alpha}; \Delta), \tag{4.20}$$

$$\boldsymbol{B}_4(\boldsymbol{\delta}) = \lim_{n\to\infty} E\{\sqrt{n}(\hat{\boldsymbol{\theta}}_n^{\mathrm{S}} - \boldsymbol{\theta}_{(n)})\} = -(p-3)\boldsymbol{\delta} E(\chi^{-2}_{p+1}(\Delta)), \tag{4.21}$$

where $\triangle$ is defined by (4.12). Also, the *asymptotic distributional covariance matrices* of the first four estimators in (1.5), (1.6), (2.3) and (2.4) are given by

$$\Gamma_1(\boldsymbol{\delta}) = \lim_{n\to\infty} E\{n(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{(n)})(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{(n)})'\} = \boldsymbol{\Sigma}_0, \tag{4.22}$$

$$\Gamma_2(\boldsymbol{\delta}) = \lim_{n\to\infty} E\{n(\hat{\theta}_n \mathbf{1}_p - \boldsymbol{\theta}_{(n)})(\hat{\theta}_n \mathbf{1}_p - \boldsymbol{\theta}_{(n)})'\} = \boldsymbol{B} + \boldsymbol{\delta}\boldsymbol{\delta}', \tag{4.23}$$

$$\begin{aligned}\Gamma_3(\boldsymbol{\delta}) &= \lim_{n\to\infty} E\{n(\hat{\boldsymbol{\theta}}_n^{\mathrm{PT}} - \boldsymbol{\theta}_{(n)})(\hat{\boldsymbol{\theta}}_n^{\mathrm{PT}} - \boldsymbol{\theta}_{(n)})'\} \\ &= \boldsymbol{\Sigma}_0 - \boldsymbol{\Sigma}_0 H_{p+1}(\chi^2_{p-1,\alpha};\triangle) \\ &\quad + \boldsymbol{\delta}\boldsymbol{\delta}'\left\{2H_{p+1}(\chi^2_{p-1,\alpha};\triangle) - H_{p+3}(\chi^2_{p-1,\alpha};\triangle)\right\},\end{aligned} \tag{4.24}$$

$$\begin{aligned}\Gamma_4(\boldsymbol{\delta}) &= \lim_{n\to\infty} E\{n(\hat{\boldsymbol{\theta}}_n^{\mathrm{S}} - \boldsymbol{\theta}_{(n)})(\hat{\boldsymbol{\theta}}_n^{\mathrm{S}} - \boldsymbol{\theta}_{(n)})'\} \\ &= \boldsymbol{\Sigma}_0 - (p-3)\boldsymbol{\Sigma}_0\left\{2E(\chi^{-2}_{p+1}(\triangle)) - (p-3)E(\chi^{-4}_{p+1}(\triangle))\right\} \\ &\quad + (p-3)(p+1)\boldsymbol{\delta}\boldsymbol{\delta}' E(\chi^{-4}_{p+3}(\triangle)).\end{aligned} \tag{4.25}$$

Consequently, the *asymptotic distributional risk function* under quadratic loss in (3.1) for the first four estimators are given by

$$R(\tilde{\boldsymbol{\theta}}_n, \boldsymbol{W}) = \mathrm{Tr}(\boldsymbol{W}\boldsymbol{\Sigma}_0), \tag{4.26}$$

$$R(\hat{\theta}_n \mathbf{1}_p, \boldsymbol{W}) = \mathrm{Tr}(\boldsymbol{W}\boldsymbol{B}) + \boldsymbol{\delta}'\boldsymbol{W}\boldsymbol{\delta}, \tag{4.27}$$

$$\begin{aligned}R(\hat{\boldsymbol{\theta}}_n^{\mathrm{PT}}, \boldsymbol{W}) &= \mathrm{Tr}(\boldsymbol{W}\boldsymbol{\Sigma}_0) - \mathrm{Tr}(\boldsymbol{W}\boldsymbol{\Sigma}_0) H_{p+1}(\chi^2_{p-1,\alpha};\triangle) \\ &\quad + \boldsymbol{\delta}'\boldsymbol{W}\boldsymbol{\delta}\left\{2H_{p+1}(\chi^2_{p-1,\alpha};\triangle) - H_{p+3}(\chi^2_{p-1,\alpha};\triangle)\right\},\end{aligned} \tag{4.28}$$

$$\begin{aligned}R(\hat{\boldsymbol{\theta}}_n^{\mathrm{S}}, \boldsymbol{W}) &= \mathrm{Tr}(\boldsymbol{W}\boldsymbol{\Sigma}_0) - (p-3)\mathrm{Tr}(\boldsymbol{W}\boldsymbol{\Sigma}_0)\left\{2E(\chi^{-2}_{p+1}(\triangle)) \right. \\ &\quad \left. -(p-3)E(\chi^{-4}_{p+1}(\triangle))\right\} \\ &\quad + (p-3)(p+1)\boldsymbol{\delta}'\boldsymbol{W}\boldsymbol{\delta} E(\chi^{-4}_{p+3}(\triangle)).\end{aligned} \tag{4.29}$$

For some numerical work and graphs we consider the particular case $\boldsymbol{W} = \boldsymbol{\Sigma}_0^{-1}$. Then the expressions in (4.26)–(4.29) reduce to

$$R(\tilde{\boldsymbol{\theta}}_n, \boldsymbol{\Sigma}_0^{-1}) = p, \tag{4.30}$$

$$R(\hat{\theta}_n \mathbf{1}_p, \boldsymbol{\Sigma}_0^{-1}) = 1 + \triangle, \tag{4.31}$$

$$\begin{aligned}R(\hat{\boldsymbol{\theta}}_n^{\mathrm{PT}}, \boldsymbol{\Sigma}_0^{-1}) &= p - pH_{p+1}(\chi^2_{p-1,\alpha};\triangle) \\ &\quad + \triangle\left\{2H_{p+1}(\chi^2_{p-1,\alpha};\triangle) - H_{p+3}(\chi^2_{p-1,\alpha};\triangle)\right\},\end{aligned} \tag{4.32}$$

$$\begin{aligned}R(\hat{\boldsymbol{\theta}}_n^{\mathrm{S}}, \boldsymbol{\Sigma}_0^{-1}) &= p - p(p-3)\left\{2E(\chi^{-2}_{p+1}(\triangle)) - (p-3)E(\chi^{-4}_{p+1}(\triangle))\right\} \\ &\quad + (p-3)(p+1)\triangle E(\chi^{-4}_{p+3}(\triangle)).\end{aligned} \tag{4.33}$$

5. ADR and Efficiency Analysis for Various Estimators

We now proceed with an analysis of the risks of the estimators and determine their asymptotic dominance properties. Firstly, we note that

$$\frac{R(\hat{\boldsymbol{\theta}}_n \mathbf{1}_p, \boldsymbol{W})}{R(\tilde{\boldsymbol{\theta}}_n, \boldsymbol{W})} \lesseqqgtr 1 \qquad \text{according as} \qquad \boldsymbol{\delta}'\boldsymbol{W}\boldsymbol{\delta} \lesseqqgtr \operatorname{Tr}(\boldsymbol{W}\boldsymbol{A}), \tag{5.1}$$

where $\boldsymbol{A} = \boldsymbol{\Sigma}_0 - \boldsymbol{B}$. Secondly,

$$\begin{aligned} &\frac{R(\hat{\boldsymbol{\theta}}_n^{\mathrm{PT}}, \boldsymbol{W})}{R(\tilde{\boldsymbol{\theta}}_n, \boldsymbol{W})} \lesseqqgtr 1 \qquad \text{according as} \\ &\qquad \boldsymbol{\delta}'\boldsymbol{W}\boldsymbol{\delta}\left\{2H_{p+1}(\chi^2_{p-1,\alpha};\Delta) - H_{p+3}(\chi^2_{p-1,\alpha};\Delta)\right\} \\ &\qquad\qquad \lesseqqgtr \operatorname{Tr}(\boldsymbol{W}\boldsymbol{\Sigma}_0)H_{p+1}(\chi^2_{p-1,\alpha};\Delta), \end{aligned} \tag{5.2}$$

and

$$\begin{aligned} &\frac{R(\hat{\boldsymbol{\theta}}_n^{\mathrm{PT}}, \boldsymbol{W})}{R(\hat{\boldsymbol{\theta}}_n \mathbf{1}_p, \boldsymbol{W})} \lesseqqgtr 1 \qquad \text{according as} \\ &\qquad \boldsymbol{\delta}'\boldsymbol{W}\boldsymbol{\delta}\left\{1 - 2H_{p+1}(\chi^2_{p-1,\alpha};\Delta) + H_{p+3}(\chi^2_{p-1,\alpha};\Delta)\right\} \\ &\qquad\qquad \lesseqqgtr \operatorname{Tr}(\boldsymbol{W}\boldsymbol{A}) - \operatorname{Tr}(\boldsymbol{W}\boldsymbol{\Sigma}_0)H_{p+1}(\chi^2_{p-1,\alpha};\Delta). \end{aligned} \tag{5.3}$$

Finally, note that $\boldsymbol{J}$ is an idempotent matrix with rank $(p-1)$, and the rank of $\boldsymbol{\Sigma}_0$ is greater than $(p-1)$. Thus, by the Courant theorem (Rao (1973)),

$$\frac{\boldsymbol{\delta}'\boldsymbol{W}\boldsymbol{\delta}}{\boldsymbol{\delta}'\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\delta}} \le \operatorname{ch}_{\max}(\boldsymbol{W}\boldsymbol{\Sigma}_0) \quad \text{for all} \quad \boldsymbol{\delta},$$

where $\operatorname{ch}_{\max}(\boldsymbol{W})$ stands for the largest characteristic root of $\boldsymbol{W}$. For convenience, we characterize a class $\mathcal{W}$ by

$$\mathcal{W} = \left\{\boldsymbol{W}_{p.s.d.} : \operatorname{Tr}(\boldsymbol{W}\boldsymbol{\Sigma}_0)/\operatorname{ch}_{\max}(\boldsymbol{W}\boldsymbol{\Sigma}_0) \ge \frac{p+1}{2}\right\}. \tag{5.4}$$

Then, we obtain from (4.26), (4.29) and (5.4) the result:

$$\frac{R(\hat{\boldsymbol{\theta}}_n^{\mathrm{S}}, \boldsymbol{W})}{R(\tilde{\boldsymbol{\theta}}_n, \boldsymbol{W})} \le 1 \quad \forall\ \boldsymbol{\delta} \quad \text{and} \quad \boldsymbol{W} \in \mathcal{W}. \tag{5.5}$$

Note that for the dominance of $\hat{\boldsymbol{\theta}}_n^{\mathrm{S}}$ over $\tilde{\boldsymbol{\theta}}_n$ for all $p \ge 4$, we need the typical choice of $\boldsymbol{W}$ given by (5.4). For an arbitrary choice of $\boldsymbol{W}$, the inequality in (5.4) may not hold simultaneously for all $\boldsymbol{\Sigma}_0$. In such a case, to induce the dominance of $\hat{\boldsymbol{\theta}}_n^{\mathrm{S}}$ over $\tilde{\boldsymbol{\theta}}_n$, one may introduce the modified Stein-estimator, $\hat{\boldsymbol{\theta}}_n^{\mathrm{MS}}$, given by

$$\hat{\boldsymbol{\theta}}_n^{\mathrm{MS}} = \hat{\theta}_n \mathbf{1}_p + [\boldsymbol{I}_p - c d_n L_n^{-1} \boldsymbol{W}^{-1} \hat{\boldsymbol{\Sigma}}_n^{-1}](\tilde{\boldsymbol{\theta}}_n - \hat{\theta}_n \mathbf{1}_p), \tag{5.6}$$

where c is the shrinkage constant and $d_n = \mathrm{ch}_{\min}(\boldsymbol{W}^{-1}\hat{\boldsymbol{\Sigma}}_n^{-1})$. The modified Stein-estimator is similar to the normal theory case treated by Berger et al. (1977) and to the nonparametric case discussed by Sen and Saleh (1985). Sen (1986) provides a detailed discussion of a general form of the shrinkage estimator, with arbitrary $\boldsymbol{W}$, which is also quite applicable to categorical data. In the sequel we assume $\boldsymbol{W}$ belongs to $\mathcal{W}$ given by (5.4) and the remaining discussion follows.

In the above we may note that $\mathrm{Tr}(\boldsymbol{WA}) > 0$ and

$$\mathrm{ch}_{\min}(\boldsymbol{W}\boldsymbol{\Sigma}_0) \leq \frac{\boldsymbol{\delta}'\boldsymbol{W}\boldsymbol{\delta}}{\triangle} \leq \mathrm{ch}_{\max}(\boldsymbol{W}\boldsymbol{\Sigma}_0),$$

where $\mathrm{ch}_{\min}(\boldsymbol{W})$ stands for the minimum characteristic root of $\boldsymbol{W}$. Hence, (5.1) implies that $\hat{\theta}_n \mathbf{1}_p$ has a smaller ADR than that of the ADR of $\tilde{\boldsymbol{\theta}}_n$ if $\triangle < \frac{\mathrm{Tr}(\mathbf{WA})}{\mathrm{ch}_{\max}(\mathbf{W\Sigma_0})}$, i.e., $\hat{\theta}_n \mathbf{1}_p$ performs better than $\tilde{\boldsymbol{\theta}}_n$ in the interval $[0, \frac{\mathrm{Tr}(\mathbf{WA})}{\mathrm{ch}_{\max}(\mathbf{W\Sigma_0})})$. Alternatively, when $\triangle$ deviates from the null hypothesis beyond $\frac{\mathrm{Tr}(\mathbf{WA})}{\mathrm{ch}_{\min}(\mathbf{W\Sigma_0})}$, the ADR of $\hat{\theta}_n \mathbf{1}_p$ grows and becomes unbounded whereas the ADR of $\tilde{\boldsymbol{\theta}}_n$ remains bounded. Hence departure from the null hypothesis, i.e., $\triangle \in (\frac{\mathrm{Tr}(\mathbf{WA})}{\mathrm{ch}_{\min}(\mathbf{W\Sigma_0})}, \infty)$ is fatal to $\hat{\theta}_n \mathbf{1}_p$ but is of very little concern to $\tilde{\boldsymbol{\theta}}_n$. For $\boldsymbol{W} = \boldsymbol{\Sigma}_0^{-1}$, we obtain $\frac{\mathrm{Tr}(\mathbf{WA})}{\mathrm{ch}_{\max}(\mathbf{W\Sigma_0})} = \frac{\mathrm{Tr}(\mathbf{WA})}{\mathrm{ch}_{\min}(\mathbf{W\Sigma_0})} = p - 1$.

Similarly, we note that $H_m(x; \triangle)$ is a decreasing function of m (DF) and $\triangle$. In fact for fixed x we have

$$\lim_{m \to \infty} H_m(x; \triangle) = \lim_{\triangle \to \infty} H_m(x; \triangle) = 0,$$

and in particular, we have

$$H_{p+1}(\chi^2_{p-1,\alpha}; \triangle) < H_{p-1}(\chi^2_{p-1,\alpha}; \triangle) \leq H_{p-1}(\chi^2_{p-1,\alpha}; 0) = 1 - \alpha \tag{5.7}$$

for every $p \geq 2$, $0 < \alpha < 1$ and $\triangle > 0$. Using these results in (4.28), we may point out some important characteristics of the ADR of $\hat{\boldsymbol{\theta}}_n^{\mathrm{PT}}$. For $\boldsymbol{\delta} = \mathbf{0}$, i.e., under the null hypothesis, the ADR of $\hat{\boldsymbol{\theta}}_n^{\mathrm{PT}}$ reduces to $\mathrm{Tr}(\boldsymbol{W}\boldsymbol{\Sigma}_0)\{1 - H_{p+1}(\chi^2_{p-1,\alpha}; 0)\}$ which is smaller than the ADR of $\tilde{\boldsymbol{\theta}}_n$. Also, for large deviations of $\boldsymbol{\delta}$ from $\mathbf{0}$ (null hypothesis), i.e., as $\triangle \to \infty$ the ADR of $\hat{\boldsymbol{\theta}}_n^{\mathrm{PT}}$ approaches the ADR of $\tilde{\boldsymbol{\theta}}_n$ from above. Similar results hold when $p \to \infty$. Further, as $\boldsymbol{\delta}$ departs from $\mathbf{0}$, i.e., as $\triangle$ grows, the value of ADR of $\hat{\boldsymbol{\theta}}_n^{\mathrm{PT}}$ increases to a maximum after crossing the

ADR of $\tilde{\boldsymbol{\theta}}_n$, then decreases towards the ADR of $\tilde{\boldsymbol{\theta}}_n$. Furthermore, we find that the ADR of the $\hat{\boldsymbol{\theta}}_n^{\rm PT}$ is smaller than the ADR of $\tilde{\boldsymbol{\theta}}_n$ when

$$\Delta < \frac{\operatorname{Tr}(\boldsymbol{W}\boldsymbol{\Sigma}_0)}{\operatorname{ch}_{\max}(\boldsymbol{W}\boldsymbol{\Sigma}_0)(2-\nu_0)}, \qquad \nu_0 = \frac{H_{p+3}(\chi^2_{p-1,\alpha};\Delta)}{H_{p+1}(\chi^2_{p-1,\alpha};\Delta)} \le 1$$

and for $\Delta > \frac{\operatorname{Tr}(\mathbf{W}\boldsymbol{\Sigma}_0)}{\operatorname{ch}_{\min}(\mathbf{W}\boldsymbol{\Sigma}_0)(2-\nu_0)}$, the opposite conclusion holds, but as soon as $\Delta \to \infty$, the ADR of $\hat{\boldsymbol{\theta}}_n^{\rm PT}$ converges to that of $\tilde{\boldsymbol{\theta}}_n$. For $\boldsymbol{W} = \boldsymbol{\Sigma}_0^{-1}$, we obtain $\frac{\operatorname{Tr}(\mathbf{W}\boldsymbol{\Sigma}_0)}{\operatorname{ch}_{\min}(\mathbf{W}\boldsymbol{\Sigma}_0)} = \frac{\operatorname{Tr}(\mathbf{W}\boldsymbol{\Sigma}_0)}{\operatorname{ch}_{\max}(\mathbf{W}\boldsymbol{\Sigma}_0)} = p$. In this case, for $\Delta \in [0, p(2-\nu_0)^{-1})$, the PTE $\hat{\boldsymbol{\theta}}_n^{\rm PT}$ has a smaller ADR than that of $\tilde{\boldsymbol{\theta}}_n$, whereas the opposite is true for $\Delta \in (p(2-\nu_0)^{-1}, \infty)$.

Now, we compare $\hat{\boldsymbol{\theta}}_n^{\rm PT}$ and $\hat{\theta}_n \mathbf{1}_p$. First, under the null hypothesis, i.e., $\boldsymbol{\delta} = \mathbf{0}$ the ADR of $\hat{\theta}_n \mathbf{1}_p$ is $\operatorname{Tr}(\boldsymbol{W}\boldsymbol{\Sigma}_0) - \operatorname{Tr}(\boldsymbol{W}\boldsymbol{A})$ and $R(\hat{\boldsymbol{\theta}}_n^{\rm PT}, \boldsymbol{W}) - R(\hat{\theta}_n \mathbf{1}_p, \boldsymbol{W}) > 0$ whenever

$$H_{p+1}(\chi^2_{p-1,\alpha};0) < \frac{\operatorname{Tr}(\boldsymbol{W}\boldsymbol{A})}{\operatorname{Tr}(\boldsymbol{W}\boldsymbol{\Sigma}_0)}. \tag{5.8}$$

Hence, we conclude that under the null hypothesis $\hat{\theta}_n \mathbf{1}_p$ performs better than $\hat{\boldsymbol{\theta}}_n^{\rm PT}$ for a range of α for which (5.8) is satisfied. However, for a fixed $\boldsymbol{\delta}$ as $\alpha \to 0$, the ADR of $\hat{\boldsymbol{\theta}}_n^{\rm PT}$ approaches $\boldsymbol{\delta}'\boldsymbol{W}\boldsymbol{\delta}$ while the ADR of $\hat{\theta}_n \mathbf{1}_p$ remains unchanged. In such a case $\hat{\theta}_n \mathbf{1}_p$ performs poorer than $\hat{\boldsymbol{\theta}}_n^{\rm PT}$. In the sequel we assume α to be bounded away from 0 satisfying (5.8) and the remaining discussion follows. Then, $\hat{\theta}_n \mathbf{1}_p$ performs better than $\hat{\boldsymbol{\theta}}_n^{\rm PT}$ for

$$\Delta < \frac{\operatorname{Tr}(\boldsymbol{W}\boldsymbol{A}) - \operatorname{Tr}(\boldsymbol{W}\boldsymbol{\Sigma}_0)H_{p+1}(\chi^2_{p-1,\alpha};\Delta)}{\operatorname{ch}_{\max}(\boldsymbol{W}\boldsymbol{\Sigma}_0)\{2(1 - H_{p+1}(\chi^2_{p-1,\alpha};\Delta)) - (1 - H_{p+3}(\chi^2_{p-1,\alpha};\Delta))\}}$$

and for

$$\Delta > \frac{\operatorname{Tr}(\boldsymbol{W}\boldsymbol{A}) - \operatorname{Tr}(\boldsymbol{W}\boldsymbol{\Sigma}_0)H_{p+1}(\chi^2_{p-1,\alpha};\Delta)}{\operatorname{ch}_{\min}(\boldsymbol{W}\boldsymbol{\Sigma}_0)\{2(1 - H_{p+1}(\chi^2_{p-1,\alpha};\Delta)) - (1 - H_{p+3}(\chi^2_{p-1,\alpha};\Delta))\}}$$

the opposite conclusion holds. For $\boldsymbol{W} = \boldsymbol{\Sigma}_0^{-1}$, the right-hand quantities equal

$$\frac{p\{1 - H_{p+1}(\chi^2_{p-1,\alpha};\Delta)\} - 1}{2(1 - H_{p+1}(\chi^2_{p-1,\alpha};\Delta)) - (1 - H_{p+3}(\chi^2_{p-1,\alpha};\Delta))}.$$

For large p, this expression approaches $p - 1$. Considering the discussion above we conclude that none of the three estimators $\tilde{\boldsymbol{\theta}}_n, \hat{\theta}_n \mathbf{1}_p$ and $\hat{\boldsymbol{\theta}}_n^{\rm PT}$ is asymptotically inadmissible with respect to the other two.

The picture is somewhat different with respect to the shrinkage estimator. First, note from (5.5) that the ADR of $\hat{\boldsymbol{\theta}}_n^{\rm S}$ is uniformly smaller than the ADR of

$\tilde{\boldsymbol{\theta}}_n$, where the upper limit is attained when $\Delta \to \infty$. This shows the asymptotic inadmissibility of $\tilde{\boldsymbol{\theta}}_n$ (under $K_{(n)}$) relative to $\hat{\boldsymbol{\theta}}_n^{\mathrm{S}}$. To compare $\hat{\boldsymbol{\theta}}_n^{\mathrm{S}}$ and $\hat{\theta}_n \mathbf{1}_p$ under the null-hypothesis in (1.2) ($\boldsymbol{\delta} = \mathbf{0}$) we have

$$\begin{aligned} R(\hat{\boldsymbol{\theta}}_n^{\mathrm{S}}, \boldsymbol{W}) &= R(\hat{\theta}_n \mathbf{1}_p, \boldsymbol{W}) + \mathrm{Tr}(\boldsymbol{W}\boldsymbol{A}) - \frac{p-3}{p-1}\mathrm{Tr}(\boldsymbol{W}\boldsymbol{\Sigma}_0) \\ &> R(\hat{\theta}_n \mathbf{1}_p, \boldsymbol{W}). \end{aligned} \tag{5.9}$$

Thus, the asymptotic risk of $\hat{\boldsymbol{\theta}}_n^{\mathrm{S}}$ is greater than that of $\hat{\theta}_n \mathbf{1}_p$ for $\boldsymbol{\delta} = \mathbf{0}$. However, as $\boldsymbol{\delta}$ moves away from $\mathbf{0}, \boldsymbol{\delta}'\boldsymbol{W}\boldsymbol{\delta}$ increases while $E(\chi_{p+1}^{-4}(\Delta))$ and $E(\chi_{p+3}^{-4}(\Delta))$ decrease, so the opposite conclusion holds. In general, $\hat{\boldsymbol{\theta}}_n^{\mathrm{S}}$ does not dominate $\hat{\theta}_n \mathbf{1}_p$ for Δ in the interval $[0, \Delta_*)$, where

$$\Delta_* = \frac{\mathrm{Tr}(\boldsymbol{W}\boldsymbol{A}) - (p-3)\mathrm{Tr}(\boldsymbol{W}\boldsymbol{\Sigma}_0)\left\{E\left(\chi_{p+1}^{-2}(\Delta)\right) + \Delta E\left(\chi_{p+3}^{-4}(\Delta)\right)\right\}}{\mathrm{ch}_{\max}(\boldsymbol{W}\boldsymbol{\Sigma}_0)\left\{1 - (p-3)(p+1)E\left(\chi_{p+3}^{-4}(\Delta)\right)\right\}},$$

however, it does so for $\Delta \in (\Delta^*, \infty)$ where

$$\Delta^* = \frac{\mathrm{Tr}(\boldsymbol{W}\boldsymbol{A}) - (p-3)\mathrm{Tr}(\boldsymbol{W}\boldsymbol{\Sigma}_0)\left\{E\left(\chi_{p+1}^{-2}(\Delta)\right) + \Delta E\left(\chi_{p+3}^{-4}(\Delta)\right)\right\}}{\mathrm{ch}_{\min}(\boldsymbol{W}\boldsymbol{\Sigma}_0)\left\{1 - (p-3)(p+1)E\left(\chi_{p+3}^{-4}(\Delta)\right)\right\}}.$$

Hence neither $\hat{\boldsymbol{\theta}}_n^{\mathrm{S}}$ nor $\hat{\theta}_n \mathbf{1}_p$ asymptotically dominates the other (under $K_{(n)}$).

Now, we compare $\hat{\boldsymbol{\theta}}_n^{\mathrm{S}}$ and $\hat{\boldsymbol{\theta}}_n^{\mathrm{PT}}$. First, consider the case when $\boldsymbol{\delta} = \mathbf{0}$. In this case $R(\hat{\boldsymbol{\theta}}_n^{\mathrm{S}}, \boldsymbol{W}) = 2(p-1)^{-1}\mathrm{Tr}(\boldsymbol{W}\boldsymbol{\Sigma}_0)$. Hence we have

$$\begin{aligned} &R(\hat{\boldsymbol{\theta}}_n^{\mathrm{S}}, \boldsymbol{W}) - R(\hat{\boldsymbol{\theta}}_n^{\mathrm{PT}}, \boldsymbol{W}) \\ &\qquad = \mathrm{Tr}(\boldsymbol{W}\boldsymbol{\Sigma}_0)\left\{H_{p+1}(\chi_{p-1,\alpha}^2; 0) - \frac{p-3}{p-1}\right\}, \quad p \geq 4. \end{aligned} \tag{5.10}$$

Thus the PTE $\hat{\boldsymbol{\theta}}_n^{\mathrm{PT}}$ has a smaller ADR than that of the ADR of $\hat{\boldsymbol{\theta}}_n^{\mathrm{S}}$ at $\boldsymbol{\delta} = \mathbf{0}$ whenever

$$H_{p+1}(\chi_{p-1,\alpha}^2; 0) > \frac{p-3}{p-1}, \tag{5.11}$$

otherwise the opposite conclusion holds. Thus, $\hat{\boldsymbol{\theta}}_n^{\mathrm{S}}$ does not dominate $\hat{\boldsymbol{\theta}}_n^{\mathrm{PT}}$ when the null hypothesis holds provided that $H_{p+1}(\chi_{p-1,\alpha}^2; 0) > (p-3)/(p-1)$. Under this situation we can order dominance of the estimators under the null-hypothesis in (1.2) as follows

$$\hat{\theta}_n \mathbf{1}_p \succ \hat{\boldsymbol{\theta}}_n^{\mathrm{PT}} \succ \hat{\boldsymbol{\theta}}_n^{\mathrm{S}} \succ \tilde{\boldsymbol{\theta}}_n, \tag{5.12}$$

where the notation $\succ$ means dominates. However, the dominance picture changes to $\hat{\theta}_n \mathbf{1}_p \succ \hat{\theta}_n^{\mathrm{S}} \succ \hat{\theta}_n^{\mathrm{PT}} \succ \tilde{\theta}_n$, whenever $H_{p+1}(\chi^2_{p-1,\alpha};0) < (p-3)/(p-1)$. Further, we can specify the condition in (5.11) equivalently as

$$2h_{p+1}(\chi^2_{p-1,\alpha};0) \leq \frac{2}{p-1} + \alpha, \tag{5.13}$$

here $h_{p+1}(\cdot;0)$ is the p.d.f. of a chi-square variable with $(p+1)$ DF and α is the level of significance of the test of the null-hypothesis in (1.2). Thus, (5.13) specifies a range of the values of α for given $p \geq 4$ for which the PTE $\hat{\theta}_n^{\mathrm{PT}}$ dominates $\hat{\theta}_n^{\mathrm{S}}$. The picture changes as $\boldsymbol{\delta}$ moves away from the origin (i.e., under the null-hypothesis). Note that $\tilde{\theta}_n$ has a constant risk equal to $\mathrm{Tr}(\boldsymbol{W}\boldsymbol{\Sigma}_0)$ and the risk of $\hat{\theta}_n \mathbf{1}_p$ depends on $\boldsymbol{\delta}$; its risk becomes unbounded as $\boldsymbol{\delta}$ moves further away from $\mathbf{0}$. As for $\hat{\theta}_n^{\mathrm{PT}}$, the risk increases from the initial value $\mathrm{Tr}(\boldsymbol{W}\boldsymbol{\Sigma}_0)\{1 - H_{p+1}(\chi^2_{p-1,\alpha};0)\}$ to a maximum after crossing the risk of $\tilde{\theta}_n$, then decreases to the value $\mathrm{Tr}(\boldsymbol{W}\boldsymbol{\Sigma}_0)$ which is the ADR of $\tilde{\theta}_n$. Similarly, the risk of $\hat{\theta}_n^{\mathrm{S}}$ with the initial value $\frac{2}{p-1}\mathrm{Tr}(\boldsymbol{W}\boldsymbol{\Sigma}_0)$ increases monotonically towards $\mathrm{Tr}(\boldsymbol{W}\boldsymbol{\Sigma}_0)$ as Δ moves away from 0. The ADR of $\hat{\theta}_n^{\mathrm{S}}$ and $\hat{\theta}_n^{\mathrm{PT}}$ intersect at the point $\Delta = \Delta_\alpha$ for each α $(0 < \alpha < 1)$, if the condition in (5.11) is satisfied, otherwise there is no intersection. If $\Delta \in [0, \Delta_\alpha)$, then the PTE dominates the SE while outside this interval the SE dominates the PTE. Thus under the condition in (5.11) neither the PTE $\hat{\theta}_n^{\mathrm{PT}}$ nor the SE $\hat{\theta}_n^{\mathrm{S}}$ dominates each other under $\{K_{(n)}\}$. The SE $\hat{\theta}_n^{\mathrm{S}}$ dominates the PTE $\hat{\theta}_n^{\mathrm{PT}}$ uniformly if the condition in (5.11) is not satisfied. However, $\hat{\theta}_n^{\mathrm{PT}}$ and $\hat{\theta}_n^{\mathrm{S}}$ share a common property as $\Delta \to \infty$, namely, their ADR's converge to $\mathrm{Tr}(\boldsymbol{W}\boldsymbol{\Sigma}_0)$ which is the risk of $\tilde{\theta}_n$. But the risk function of $\hat{\theta}_n^{\mathrm{S}}$ is always below this limit, whereas the risk function of $\hat{\theta}_n^{\mathrm{PT}}$ exceeds this limit for some intermediate values of Δ depending on α, the level of significance of the test of (1.2).

To improve over $\hat{\theta}_n^{\mathrm{PT}}(\hat{\theta}_n^{\mathrm{S}})$ one may use $\hat{\theta}_n^{\mathrm{PT}+}(\hat{\theta}_n^{\mathrm{S}+})$ as studied by Sclove, Morris and Radhakrishnan (1972) in the multivariate normal mean problem. This relative dominance of (2.5) over (2.4) remains intact under our asymptotic setup, hence, we may advocate the use of this modified SE. In terms of ordering, MPTE comes in between the usual PTE in (2.3) and SE, while in general the relative ADR picture with PTE and SE remains applicable to this case as well. Finally, from the point of robust-efficiency both $\hat{\theta}_n^{\mathrm{PT}}$ and $\hat{\theta}_n^{\mathrm{S}}$ may be advocated, leaning more towards $\hat{\theta}_n^{\mathrm{S}}$, since the size of Δ is generally unknown and unlikely to be small.

Next, we consider the relative efficiency analysis of the estimators for the specific case $\boldsymbol{W} = \boldsymbol{\Sigma}_0^{-1}$. Using (4.30)–(4.33), the relative efficiency (R.E.) of the RMLE $\hat{\theta}_n \mathbf{1}_p$ to the UMLE $\tilde{\theta}_n$ is defined to be

$$\text{R.E.}(\hat{\theta}_n \mathbf{1}_p : \tilde{\boldsymbol{\theta}}_n) = \frac{R(\tilde{\boldsymbol{\theta}}_n, \boldsymbol{\Sigma}_0^{-1})}{R(\hat{\theta}_n \mathbf{1}_p, \boldsymbol{\Sigma}_0^{-1})} = \frac{p}{1+\triangle}. \tag{5.14}$$

Similarly, the relative efficiency of $\hat{\boldsymbol{\theta}}_n^{\text{PT}}$ and the SE $\hat{\boldsymbol{\theta}}_n^{\text{S}}$ to the UMLE $\tilde{\boldsymbol{\theta}}_n$ are given by (5.15)–(5.16) respectively:

$$\text{R.E.}(\hat{\boldsymbol{\theta}}_n^{\text{PT}} : \tilde{\boldsymbol{\theta}}_n) = \frac{R(\tilde{\boldsymbol{\theta}}_n, \boldsymbol{\Sigma}_0^{-1})}{R(\hat{\boldsymbol{\theta}}_n^{\text{PT}}, \boldsymbol{\Sigma}_0^{-1})} = \frac{1}{1+g_1(\alpha,\triangle)}, \tag{5.15}$$

where

$$g_1(\alpha,\triangle) = \frac{\triangle}{p}\left\{2H_{p+1}(\chi^2_{p-1,\alpha};\triangle) - H_{p+3}(\chi^2_{p-1,\alpha};\triangle)\right\} - H_{p+1}(\chi^2_{p-1,\alpha};\triangle);$$

$$\text{R.E.}(\hat{\boldsymbol{\theta}}_n^{\text{S}} : \tilde{\boldsymbol{\theta}}_n) = \frac{R(\tilde{\boldsymbol{\theta}}_n, \boldsymbol{\Sigma}_0^{-1})}{R(\hat{\boldsymbol{\theta}}_n^{\text{S}}, \boldsymbol{\Sigma}_0^{-1})} = \frac{1}{1+g_2(\triangle)}, \tag{5.16}$$

where

$$\begin{aligned} g_2(\triangle) =& \frac{\triangle}{p}(p-3)(p+1)E(\chi^{-4}_{p+3}(\triangle)) \\ &- (p-3)\left\{2E(\chi^{-2}_{p+1}(\triangle)) - (p-3)E(\chi^{-4}_{p+1}(\triangle))\right\}. \end{aligned}$$

Tables 5.1–5.2 provide the maximum relative efficiencies (at $\triangle = 0$) for $\hat{\theta}_n \mathbf{1}_p(E_1), \hat{\boldsymbol{\theta}}_n^{\text{PT}}(E_3)$ and $\hat{\boldsymbol{\theta}}_n^{\text{S}}(E_4)$ relative to $\tilde{\boldsymbol{\theta}}_n$, the intersecting relative efficiencies denoted by $E_{\triangle_\alpha}$ of $\hat{\boldsymbol{\theta}}_n^{\text{S}}$ and $\hat{\boldsymbol{\theta}}_n^{\text{PT}}$ (which is the common efficiency of the PTE and the SE) and the $\triangle_\alpha$-values at which the intersection occurs. For example, using Table 5.1, with $p = 6, \alpha = 0.10$ and $\triangle \in [0, 1.1514]$, it appears that $\hat{\boldsymbol{\theta}}_n^{\text{PT}}$ dominates $\hat{\boldsymbol{\theta}}_n^{\text{S}}$, but outside the interval domination is reversed. While, if the hypothesis in (1.2) is correct, i.e., $\triangle = 0$, then $\hat{\theta}_n \mathbf{1}_p$ has larger efficiency ($E_1 = 6.0000$) than $\hat{\boldsymbol{\theta}}_n^{\text{PT}}(E_3 = 4.2350)$ and $\hat{\boldsymbol{\theta}}_n^{\text{S}}(E_4 = 2.5000)$ relative to $\tilde{\boldsymbol{\theta}}_n$. Note that $\triangle_\alpha = 0$ in Tables 5.1–5.2 means that no intersection between $\hat{\boldsymbol{\theta}}_n^{\text{PT}}$ and $\hat{\boldsymbol{\theta}}_n^{\text{S}}$ occurs, and the corresponding relative efficiencies $E_{\triangle_\alpha}$ are equal to E_3, i.e., the SE $\hat{\boldsymbol{\theta}}_n^{\text{S}}$ in (2.4) dominates the PTE $\hat{\boldsymbol{\theta}}_n^{\text{PT}}$ in (2.3) when p-values or α-values are large. Figure 5.1 exhibits plots of the relative efficiency (R.E.) curves for $\hat{\theta}_n \mathbf{1}_p, \hat{\boldsymbol{\theta}}_n^{\text{PT}}$ and $\hat{\boldsymbol{\theta}}_n^{\text{S}}$ relative to $\tilde{\boldsymbol{\theta}}_n$ versus $\triangle$ for various values of α and $p = 5$. Apparently, $\hat{\boldsymbol{\theta}}_n^{\text{PT}}$ dominates $\hat{\boldsymbol{\theta}}_n^{\text{S}}$ for a range of α $(0 < \alpha < 1)$ and for small values of $\triangle$. Figure 5.2 exhibits plots of the relative efficiency (R.E.) curves for $\hat{\theta}_n \mathbf{1}_p, \hat{\boldsymbol{\theta}}_n^{\text{PT}}$ and $\hat{\boldsymbol{\theta}}_n^{\text{S}}$ relative to $\tilde{\boldsymbol{\theta}}_n$ versus $\triangle$ for various values of α and $p = 20$. Apparently, $\hat{\boldsymbol{\theta}}_n^{\text{S}}$ dominates $\hat{\boldsymbol{\theta}}_n^{\text{PT}}$ for large values of p.

6. Baseball Data Analysis

Now, we illustrate the above theory with an analysis of the famous baseball data of Efron and Morris (1975). In this data a James-Stein estimator is used to predict batting averages of 18 major league players in the remainder of the 1970 season. The number of hits y_i in the first 45 bats is observed for each player i $(i = 1, 2, \dots, 18)$. The problem is to estimate $\boldsymbol{\theta} = (\theta_1, \dots, \theta_{18})'$, where θ_i denotes the final season batting average of player i. Efron and Morris (1975) used an arc sine transformation on each y_i to obtain approximate normality and then used a James-Stein estimator on the transformed counts. This will be denoted by $\hat{\boldsymbol{\theta}}_n^{\mathrm{EM}}$. Albert (1984) proposed an empirical Bayes estimator denoted by $\hat{\boldsymbol{\theta}}_n^{\mathrm{EB}}$ to estimate $\boldsymbol{\theta}$. We present in Table 5.3 the true batting averages $(\boldsymbol{\theta}^{\mathrm{T}})$, the UMLE $(\tilde{\boldsymbol{\theta}}_n), \hat{\boldsymbol{\theta}}_n^{\mathrm{EM}}, \hat{\boldsymbol{\theta}}_n^{\mathrm{EB}}, \hat{\boldsymbol{\theta}}_n^{\mathrm{PT}}$ and $\hat{\boldsymbol{\theta}}_n^{\mathrm{S}}$. The true batting averages, obtained from Morris (1983), have mean 0.267 and s.d. 0.037. In order to assess the performance of the various estimators, loss defined by $(\hat{\theta}_i - \theta_i^{\mathrm{T}})^2/(0.037)^2$ for each estimator $\hat{\theta}_i (i = 1, \dots, 18)$ along with the average loss (due to the estimators) have been tabulated in Table 5.4. The PTE $\hat{\boldsymbol{\theta}}_n^{\mathrm{PT}} = (.265, \dots, .265)'$ is the result of testing the null hypothesis $H_0 : \theta_1 = \theta_2 = \cdots = \theta_p$ at 15% level of significance. In Table 5.4 we have denoted by $L_{\mathrm{MLE}}, L_{\mathrm{EM}}, L_{\mathrm{EB}}, L_{\mathrm{PT}}$ and L_{S} the losses due to the estimators, $\tilde{\boldsymbol{\theta}}_n$ (UMLE),$\hat{\boldsymbol{\theta}}_n^{\mathrm{EM}}, \hat{\boldsymbol{\theta}}_n^{\mathrm{EB}}, \hat{\boldsymbol{\theta}}_n^{\mathrm{PT}}$ and $\hat{\boldsymbol{\theta}}_n^{\mathrm{S}}$ respectively. A comparison of the average loss of various estimators show that our estimator $\hat{\boldsymbol{\theta}}_n^{\mathrm{S}}$ is quite comparable to that of $\hat{\boldsymbol{\theta}}_n^{\mathrm{EM}}$ and $\hat{\boldsymbol{\theta}}_n^{\mathrm{EB}}$ (and $\hat{\boldsymbol{\theta}}_n^{\mathrm{PT}}$ is not far behind). If we apply Morris's ((1983), (3.8) page 32) approximation formula, $\hat{\theta}_i^{\mathrm{M}} = .208\tilde{\theta}_i + 0.2105$ obtained by an empirical Bayes approach, we get exactly $\hat{\theta}_i^{\mathrm{S}} (i = 1, \dots, 18)$, which supports the validity of our asymptotic theory as well.

Acknowledgement

This research has been supported by the NSERC grant no. A3088. Thanks are due to the Editor and the referees for their constructive suggestions to bring the paper in the present form.

Appendix

Here, we prove (4.5). First, define

$$k_n = \max_{1 \le j \le p} n|\tilde{\theta}_j - \hat{\theta}_n|. \tag{A.1}$$

Then

$$L_n \ge \frac{4}{n} k_n^2 \qquad \text{with probability 1.} \tag{A.2}$$

Now $k_n = 0$ yields $L_n = 0$ and hence, $\hat{\boldsymbol{\theta}}_n^{\mathrm{S}} = \tilde{\boldsymbol{\theta}}_n = \hat{\theta}_n \mathbf{1}_p$. Thus, we exclude the case $k_n = 0$. In this case $k_n > 0$ and it can be less than 1, with probability 1; in any case $k_n > \frac{1}{n}$, with probability 1 (excluding the case $k_n = 0$). Thus, on the set $\{L_n > 0\}$, we may write

$$L_n^{-1} I(0 < L_n \le \epsilon) = L_n^{-1} I(0 < L_n < \frac{4}{n}) + L_n^{-1} I(\frac{4}{n} \le L_n \le \epsilon). \qquad (A.3)$$

Let

$$k_{0n} = \frac{1}{2}\sqrt{n\epsilon}. \qquad (A.4)$$

Then, using (A.2) and (A.4) we obtain for every $\boldsymbol{\theta}$ (under $K_{(n)}$ as well)

$$\begin{aligned} &E_{\boldsymbol{\theta}}\{L_n^{-1} I(0 < L_n \le \epsilon)\} \\ &\quad \le \frac{n}{4}\Big\{E_{\boldsymbol{\theta}}\big[k_n^{-2} I(\frac{1}{n} < k_n < 1)\big] + E_{\boldsymbol{\theta}}\big[k_n^{-2} I(1 \le k_n \le k_{0n})\big]\Big\}. \qquad (A.5) \end{aligned}$$

Replacing k_n by $[k_n]$, the largest integer less than or equal to k_n, the second term in (A.5) (for $k_n \ge 1$) satisfies the inequality

$$\begin{aligned} &E_{\boldsymbol{\theta}}\{k_n^{-2} I(1 \le k_n \le k_{0n})\} \\ &\quad \le E_{\boldsymbol{\theta}}\{k_n^{-2} I(1 \le [k_n] \le k_{0n})\} = \sum_{k=1}^{[k_{0n}]} k^{-2} P_{\boldsymbol{\theta}}\{[k_n] = k\} \\ &\quad \le 16 \sum_{k=1}^{[k_{0n}]} k^{-3} P_{\boldsymbol{\theta}}\{[k_n] \le k\} + [k_{0n}]^{-2} P_{\boldsymbol{\theta}}\{[k_n] \le [k_{0n}]\}. \qquad (A.6) \end{aligned}$$

Now, the degrees of freedom for L_n is $p-1$, hence there are k^{p-1} configuration for which $k_n \le k$. For each configuration the probability is $O(n^{-(p-1)/2})$. Thus, $P_{\boldsymbol{\theta}}\{k_n \le k\} = O(k^{p-1} n^{-(p-1)/2})$ for every $k \le k_{0n}$. Hence, by (A.5) and (A.6),

$$\begin{aligned} &\frac{n}{4} E_{\boldsymbol{\theta}}\{k_n^{-2} I(1 \le k_n \le k_{0n})\} \\ &\quad \le \frac{n}{4}\Big\{\sum_{k=1}^{[k_{0n}]} O(k^{p-4} n^{-(p-1)/2}) + O(k_{0n}^{p-3} n^{-(p-1)/2})\Big\} \\ &\quad = O(n^{-(p-3)/2})\Big\{\sum_{k=1}^{[k_{0n}]} O(k^{p-4}) + O(k_{0n}^{p-3})\Big\} \\ &\quad = O(n^{-(p-3)/2} k_{0n}^{p-3}) \\ &\quad = O(\epsilon^{(p-3)/2} n^{-(p-3)/2} n^{(p-3)/2}) \\ &\quad = O(\epsilon^{(p-3)/2}), \quad p \ge 4. \qquad (A.7) \end{aligned}$$

Since ϵ is arbitrarily small (A.7) may be made small. For the first term we set $[nk] = k_n^*$ so that on $\frac{1}{n} \le k_n < 1$, we obtain $1 \le k_n^* < n$; thus,

$$\frac{n}{4}E_{\theta}\{k_n^{-2}I(\frac{1}{n} \le k_n < 1)\} \le \frac{n^3}{4}E_{\theta}\{(k_n^*)^{-2}I(1 \le k_n^* \le n-1)\}$$
$$= \frac{n^3}{4}\sum_{k^*=1}^{n-1}(k^*)^{-2}P_{\theta}(k_n^* = k^*).$$

By a similar argument utilized before, one may show that the last term is $O(n^{-(p-3)/2})$, $p \ge 4$. Hence, we have proved that $E_{\theta}\{L_n^{-1}I(0 < L_n \le \epsilon)\} \to 0$, as $n \to \infty$.

Table 5.1. Maximum relative efficiencies of the RMLE, PTE, and SE and the intersecting efficiencies for the PTE and SE for each α with corresponding Δ_{α}-values for p-values

$\alpha\backslash p$		4	6	8	10	12	14	16
.05	E_1	4.0000	6.0000	8.0000	10.0000	12.0000	14.0000	16.0000
	E_3	5.9971	7.3761	8.3363	9.0666	9.6514	10.1362	10.5481
	E_4	1.5000	2.5000	3.5000	4.5000	5.5000	6.5000	7.5000
	$E_{\Delta_{.05}}$	1.2617	1.8128	2.4102	3.0680	3.8010	4.6266	5.5660
	$\Delta_{.05}$	2.0988	1.8974	1.7020	1.5048	1.3048	1.1025	.8983
.10	E_1	4.0000	6.0000	8.0000	10.0000	12.0000	14.0000	16.0000
	E_3	3.5396	4.2350	4.7092	5.0648	5.3466	5.5783	5.7739
	E_4	1.5000	2.5000	3.5000	4.5000	5.5000	6.5000	7.5000
	$E_{\Delta_{.10}}$	1.2913	1.9996	2.9013	4.0761	5.3466	5.5783	5.7739
	$\Delta_{.10}$	1.6932	1.1514	.7081	.3048	.0000	.0000	.0000
.15	E_1	4.0000	6.0000	8.0000	10.0000	12.0000	14.0000	16.0000
	E_3	2.6425	3.1003	3.4085	3.6375	3.8177	3.9652	4.0891
	E_4	1.5000	2.5000	3.5000	4.5000	5.5000	6.5000	7.5000
	$E_{\Delta_{.15}}$	1.3199	2.2163	3.4085	3.6375	3.8177	3.9652	4.0891
	$\Delta_{.15}$	1.3593	.5482	.0000	.0000	.0000	.0000	.0000
.20	E_1	4.0000	6.0000	8.0000	10.0000	12.0000	14.0000	16.0000
	E_3	2.1684	2.5038	2.7274	2.8925	3.0218	3.1272	3.2155
	E_4	1.5000	2.5000	3.5000	4.5000	5.5000	6.5000	7.5000
	$E_{\Delta_{.20}}$	1.3511	2.4973	2.7274	2.8925	3.0218	3.1272	3.2155
	$\Delta_{.20}$	1.0462	.0044	.0000	.0000	.0000	.0000	.0000
.25	E_1	4.0000	6.0000	8.0000	10.0000	12.0000	14.0000	16.0000
	E_3	1.8729	2.1328	2.3049	2.4314	2.5300	2.6101	2.6771
	E_4	1.5000	2.5000	3.5000	4.5000	5.5000	6.5000	7.5000
	$E_{\Delta_{.25}}$	1.3873	2.1328	2.3049	2.4314	2.5300	2.6101	2.6771
	$\Delta_{.25}$	.7340	.0000	.0000	.0000	.0000	.0000	.0000

Table 5.2. Maximum relative efficiencies of the RMLE, PTE, and SE and the intersecting efficiencies for the PTE and SE for each α with corresponding Δ_α-values for p-values

$\alpha \backslash p$		4	6	8	10	12	14	16
.30	E_1	4.0000	6.0000	8.0000	10.0000	12.0000	14.0000	16.0000
	E_3	1.6706	1.8788	2.0161	2.1165	2.1946	2.2578	2.3106
	E_4	1.5000	2.5000	3.5000	4.5000	5.5000	6.5000	7.5000
	$E_{\Delta_{.30}}$	1.4314	1.8788	2.0161	2.1165	2.1946	2.2578	2.3106
	$\Delta_{.30}$	.4109	.0000	.0000	.0000	.0000	.0000	.0000
.35	E_1	4.0000	6.0000	8.0000	10.0000	12.0000	14.0000	16.0000
	E_3	1.5234	1.6938	1.8057	1.8873	1.9507	2.0019	2.0445
	E_4	1.5000	2.5000	3.5000	4.5000	5.5000	6.5000	7.5000
	$E_{\Delta_{.35}}$	1.4875	1.6938	1.8057	1.8873	1.9507	2.0019	2.0445
	$\Delta_{.35}$	.0678	.0000	.0000	.0000	.0000	.0000	.0000
.40	E_1	4.0000	6.0000	8.0000	10.0000	12.0000	14.0000	16.0000
	E_3	1.4119	1.5531	1.6457	1.7130	1.7652	1.8072	1.8422
	E_4	1.5000	2.5000	3.5000	4.5000	5.5000	6.5000	7.5000
	$E_{\Delta_{.40}}$	1.4119	1.5531	1.6457	1.7130	1.7652	1.8072	1.8422
	$\Delta_{.40}$	.0000	.0000	.0000	.0000	.0000	.0000	.0000
.45	E_1	4.0000	6.0000	8.0000	10.0000	12.0000	14.0000	16.0000
	E_3	1.3248	1.4427	1.5199	1.5760	1.6194	1.6543	1.6833
	E_4	1.5000	2.5000	3.5000	4.5000	5.5000	6.5000	7.5000
	$E_{\Delta_{.45}}$	1.3248	1.4427	1.5199	1.5760	1.6194	1.6543	1.6833
	$\Delta_{.45}$	.0000	.0000	.0000	.0000	.0000	.0000	.0000
.50	E_1	4.0000	6.0000	8.0000	10.0000	12.0000	14.0000	16.0000
	E_3	1.2554	1.3541	1.4187	1.4656	1.5019	1.5310	1.5552
	E_4	1.5000	2.5000	3.5000	4.5000	5.5000	6.5000	7.5000
	$E_{\Delta_{.50}}$	1.2554	1.3541	1.4187	1.4656	1.5019	1.5310	1.5552
	$\Delta_{.50}$	.0000	.0000	.0000	.0000	.0000	.0000	.0000

Table 5.3. The true values (θ_i^{T}), the maximum likelihood estimator MLE($\tilde{\theta}_i$), Efron and Morris estimators ($\hat{\theta}_i^{\mathrm{EM}}$), Albèrt empirical Bayes estimator ($\hat{\theta}_i^{\mathrm{EB}}$) and the proposed estimators ($\hat{\theta}_i^{\mathrm{PT}}$) and ($\hat{\theta}_i^{\mathrm{S}}$)

Table 5.4. Computed values of the losses of the MLE($\tilde{\theta}_i$), Efron and Morris estimators ($\hat{\theta}_i^{\mathrm{EM}}$), Albert empirical Bayes estimator ($\hat{\theta}_i^{\mathrm{EB}}$) and the proposed estimators ($\hat{\theta}_i^{\mathrm{PT}}$) and ($\hat{\theta}_i^{\mathrm{S}}$)

Table 5.3

i	θ_i^{T}	$\tilde{\theta}_i$	$\hat{\theta}_i^{\mathrm{EM}}$	$\hat{\theta}_i^{\mathrm{EB}}$	$\hat{\theta}_i^{\mathrm{PT}}$	$\hat{\theta}_i^{\mathrm{S}}$
1	0.346	0.400	0.290	0.279	0.265	0.294
2	0.300	0.378	0.286	0.277	0.265	0.289
3	0.279	0.356	0.281	0.274	0.265	0.245
4	0.223	0.333	0.277	0.272	0.265	0.280
5	0.276	0.311	0.273	0.270	0.265	0.275
6	0.273	0.311	0.273	0.270	0.265	0.275
7	0.266	0.289	0.268	0.268	0.265	0.270
8	0.211	0.267	0.264	0.265	0.265	0.266
9	0.271	0.244	0.259	0.263	0.265	0.261
10	0.232	0.244	0.259	0.263	0.265	0.261
11	0.266	0.222	0.254	0.261	0.265	0.256
12	0.258	0.222	0.254	0.261	0.265	0.256
13	0.306	0.222	0.254	0.261	0.265	0.256
14	0.267	0.222	0.254	0.261	0.265	0.256
15	0.228	0.222	0.254	0.261	0.265	0.256
16	0.288	0.200	0.249	0.259	0.265	0.252
17	0.318	0.178	0.233	0.257	0.265	0.247
18	0.200	0.156	0.208	0.254	0.265	0.242
mean	0.267	0.265	0.261	0.265	0.265	0.263
s.d.	0.037	0.068	0.019	0.007	0.000	0.014

Table 5.4

i	L_{MLE}	L_{EM}	L_{EB}	L_{PT}	L_{S}
1	2.130	2.291	3.279	4.793	1.975
2	4.444	0.143	0.386	0.895	0.088
3	4.331	0.003	0.018	0.143	0.844
4	8.839	2.130	1.754	1.289	2.373
5	0.895	0.007	0.026	0.088	0.001
6	1.055	0.000	0.007	0.047	0.003
7	0.386	0.003	0.003	0.001	0.012
8	2.291	2.052	2.130	2.130	2.210
9	0.533	0.105	0.047	0.026	0.073
10	0.105	0.533	0.702	0.795	0.614
11	1.414	0.105	0.018	0.001	0.073
12	0.947	0.012	0.007	0.036	0.003
13	5.154	1.975	1.479	1.228	1.826
14	1.479	0.123	0.026	0.003	0.088
15	0.026	0.494	0.795	1.000	0.573
16	5.657	1.111	0.614	0.386	0.947
17	14.317	5.278	2.718	2.052	3.682
18	1.414	0.047	2.130	3.086	1.289
mean	3.079	0.912	0.897	1.000	0.926
s.d.	3.565	1.346	1.045	1.270	1.051

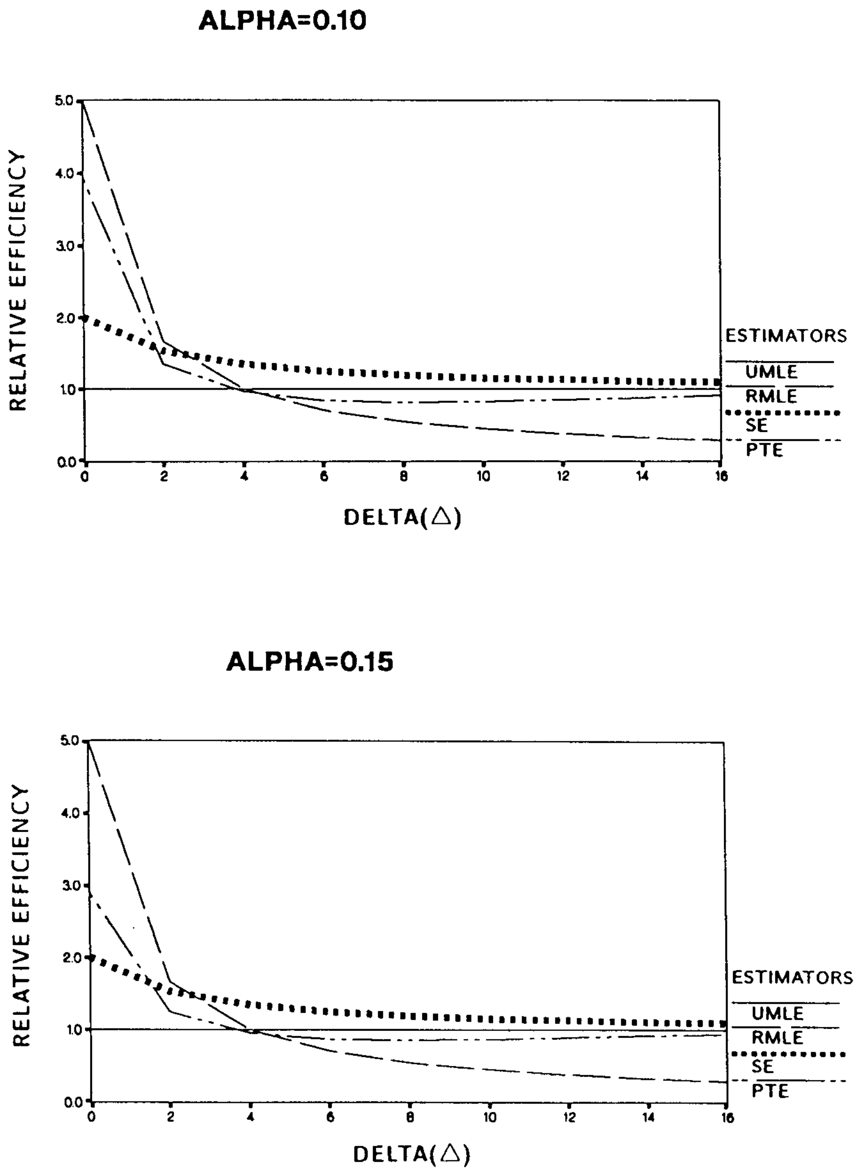
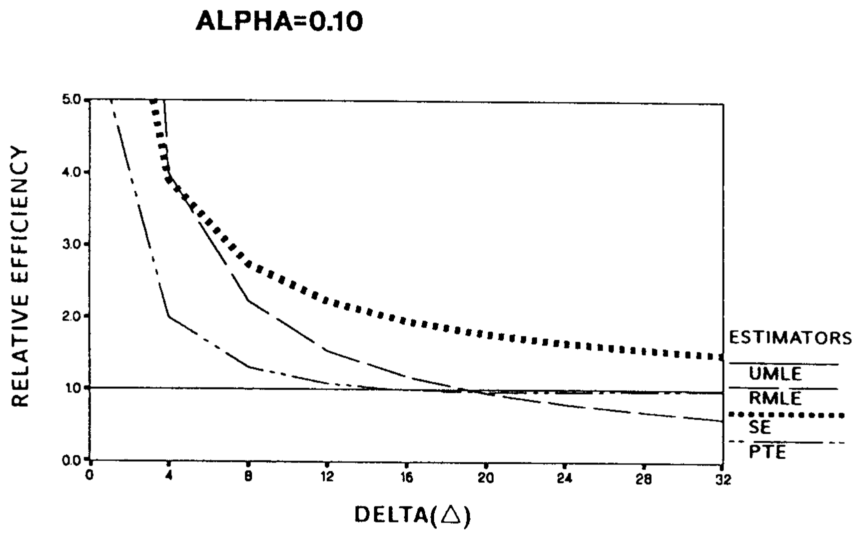


Figure 5.1. Dominance picture for UMLE, RMLE, PTE and SE for $p = 5$.



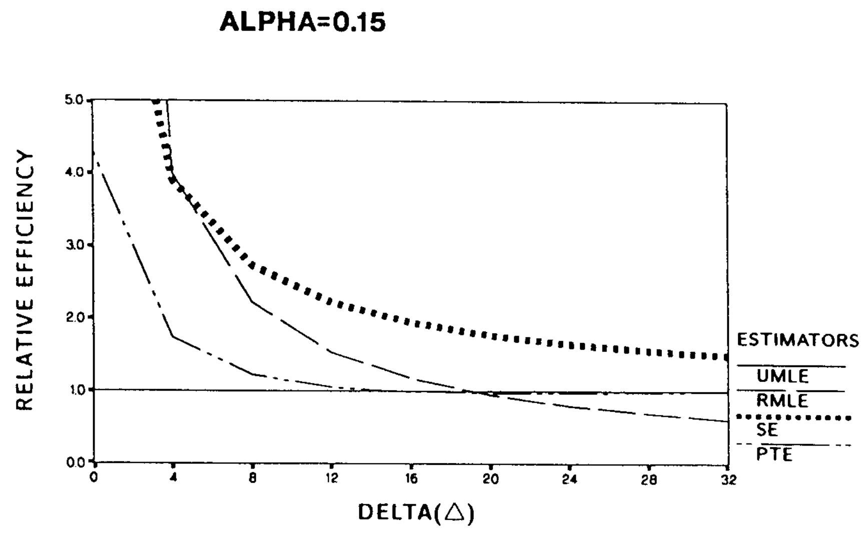


Figure 5.2. Dominance picture for UMLE, RMLE, PTE and SE for $p = 20$.

References

Albert, J. H. (1984). Empirical Bayes estimation of a set of binomial prababilities. *J. Statist. Comput. Simulation* **20**, 129-144.

Albert, J. H. and Gupta, A. K. (1981). Bayesian methods for binomial data with applications to a nonresponse problem. Technical report, Department of Mathematics and Statistics, Bowling Green State University.

Albert, J. H. and Gupta, A. K. (1983). Estimation in contingency tables using prior information. *J. Roy. Statist. Soc. Ser.B* **45**, 60-69.

Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*, 2nd edition. John Wiley, New York.

Arnold, S. F. (1981). *The Theory of Linear Models and Multivariate Analysis.* John Wiley, New York.

Berger, J. O. (1980). A robust generalized Bayes estimator and confidence region for a multivariate normal mean. *Ann. Statist.* **8**, 716-761.

Efron, B. and Morris, C. N. (1975). Data analysis using Stein's estimator and its generalizations. *J. Amer. Statist. Assoc.* **70**, 311-319.

James, W. and Stein, C. (1961). Estimation with quadratic loss. *Proc. 4th Berkeley Symp. on Math. Statist. and Probab.* 1, 361-379, Ulcla Press, Berkeley, California.

Judge, G. G. and Bock, M. E. (1978). *The Statistical Implications of Pre-test and Stein-Rule Estimators in Econometrics.* North-Holland, Amsterdam.

Morris, C. N. (1983). Parametric empirical Bayes confidence intervals. *Scientific Inference, Data Analysis, and Robustness* (Edited by G. E. P. Box, T. Leonard and Chien-Fu Wu), 25-50, Academic Press.

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*, 2nd edition. John Wiley, New York.

Saleh, A. K. Md. Ehsanes and Sen, P. K. (1978). Non-parametric estimation of location parameter after a preliminary test on regression. *Ann. Statist.* **6**, 154-168.

Saleh, A. K. Md. Ehsanes and Sen, P. K. (1983). Nonparametric tests of location after a preliminary test on regression in the multivariate case. *Comm. Statist. Theory Methods* **12**(16), 1855-1872.

Saleh, A. K. Md. Ehsanes and Sen, P. K. (1984a). Least squares and rank order preliminary test estimation in general multivariate linear models. *Proc. of ISI, Golden Jubilee Conference on Statistics: Application and New Directions* (Dec. 16–19, 1981), 237-253.

Saleh, A. K. Md. Ehsanes and Sen, P. K. (1984b). Nonparametric preliminary test inference. *Handbook of Statistics* Vol.4 (Edited by P. R. Krishnaiah and P. K. Sen), 275-297, North-Holland Publishers.

Saleh, A. K. Md. Ehsanes and Sen, P. K. (1985a). On shrinkage M-estimators of location parameters. *Comm. Statist. Theory Methods* **14**(10), 2313-2329.

Saleh, A. K. Md. Ehsanes and Sen, P. K. (1985b). Preliminary test predicted in general multivariate linear models. *Proc. of the Pacific Area Statistical Conference* (Edited by M. Matusita; Held Dec. 15-19, 1982), 619-638, North-Holland Publishers.

Saleh, A. K. Md. Ehsanes and Sen, P. K. (1985c). Shrinkage least squares estimation in a general multivariate linear model. *Proc. of the 5th Pannonian Symp. on Math. Statistics* (Edited by J. Mogyorodi, I. Vincze and W. Wertz), 307-325.

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

Sen, P. K. (1984). A James-Stein type detour of U-Statistics. *Comm. Statist. Theory Methods* **13**(22), 2725-2747.

Sen, P. K. (1986). On the asymptotic distributional risks of shrinkage and preliminary test versions of maximum likelihood estimators. *Sankhyā Ser.A* **48**, 354-371.

Sen, P. K. and Saleh, A. K. Md. Ehsanes (1979). Nonparametric estimation of location parameter after a preliminary test on regression in multivariate case. *J. Multivariate Anal.* **9**(2), 322-331.

Sen, P. K. and Saleh, A. K. Md. Ehsanes (1985). On some shrinkage estimators of multivariate location. *Ann. Statist.* **13**, 272-281.

Sen, P. K. and Saleh, A. K. Md. Ehsanes (1987). On preliminary test and shrinkage M-estimation in linear models. *Ann. Statist.* **15**(4), 1580-1592.

Department of Mathematics and Statistics, Carleton University, Ottawa, K1S 5B6, Canada.

(Received April 1989; accepted March 1990)