ON SHRINKAGE M-ESTIMATORS OF LOCATION PARAMETERS

by

A.K.Md. Ehsanes Saleh
Department of Mathematics and Statistics
Carleton University, Ottawa, Canada

and

Pranab Kumar Sen
Department of Biostatistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 1479

December 1984

ON SHRINKAGE M-ESTIMATORS OF LOCATION PARAMETERS

A.K.Md.Ehsanes Saleh and Pranab Kumar Sen

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

Department of Biostatistics
University of North Carolina,
Chapel Hill,
NC 27514 , USA

Key Words and Phrases : Asymptotic risk ; James-Stein rule; local (Pitman type-) alternatives; M-estimation; multivariate location model; preliminary test estimation; shrinkage estimation.

ABSTRACT

For a general class of continuous and marginally symmetric, multivariate distributions, shrinkage estimators of locations based on suitable M-statistics are considered. These estimators are based on the James-Stein rule (incorporating the idea of preliminary test estimators too), and the main emphasis is laid on the study of their asymptotic (distributional) risk properties . Asymptotic admissibility results are also studied under fairly general regularity conditions.

1. INTRODUCTION

For a multinormal distribution (of dimension 3 or more), under a *quadratic error loss function* , the classical *maximum likelihood estimators (MLE)* of location are not generally admissible [c.f. Stein (1956)] and improved (non-linear) estimators are available in the literature [viz., James and Stein (1961) and a series of papers , referred to in Berger (1980)]. Though, in the initial stage, the covariance matrix was assumed to be of certain specific form, the case of completely unknown dispersion matrix has also been treated in generality [viz., Berger and Haff(1983) where other references have been cited]. While the assumed normality of the underlying distribution plays a vital role in the development of the so called exact dominance and admissibility results, the exact theory stumbles into difficulties when this normality assumption is dispensed with. Nevertheless, under an appropriate asymptotic setup [explained in detail in Sen (1984) and Sen and Saleh (1985)], the theory of *shrinkage estimation in multi-parameter models* extends to a wider class of statistics and a general class of distributions. Sen (1984) considered the case of *U-statistics* , while, Sen and Saleh (1985) have considered the case of *R-estimators of locations*. The object of the present study is to consider some *shrinkage M-estimators of location* where the theory of *James-Stein rule estimation* and *preliminary test estimation* are incorporated in the formulation of the prposed estimators.

Let $\underset{\sim}{X}_i = (X_{i1}, \ldots, X_{ip})'$, $i = 1, \ldots, n$ be independent and identically distributed (i.i.d.) random vectors (r.v.) having a p (> 2) variate continuous distribution function $F_{\underset{\sim}{\theta}}$, defined on the Euclidean space E^p . $F_{\underset{\sim}{\theta}}$ is assumed to have symmetric marginal distributions $F_{[j]}$ (around the location parameter θ_j), for $j = 1, \ldots, p$. Thus,

$$F_{\underset{\sim}{\theta}}(\underset{\sim}{x}) = F(\underset{\sim}{x} - \underset{\sim}{\theta}) \ , \ \underset{\sim}{x} \in E^p \ ; \ \underset{\sim}{\theta} = (\theta_1, \ldots, \theta_p)' , \qquad (1.1)$$

where $\underset{\sim}{\theta}$ is the vector of (marginal) location parameters, and our interest lies in the estimation of $\underset{\sim}{\theta}$. For an estimator $\underset{\sim}{\delta}_n$ (based on $\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_n$), we conceive of a quadratic error loss function

$$L(\underset{\sim}{\delta}_n , \underset{\sim}{\theta}) = (\underset{\sim}{\delta}_n - \underset{\sim}{\theta})'\underset{\sim}{Q}(\underset{\sim}{\delta}_n - \underset{\sim}{\theta}) , \qquad (1.2)$$

for some given positive definite (p.d.) matrix $\underset{\sim}{Q}$. Then, the *risk function* is given by

$$\rho_n(\underset{\sim}{\delta} , \underset{\sim}{\theta}) = nEL(\underset{\sim}{\delta}_n ; \underset{\sim}{\theta}) = Tr(\underset{\sim}{Q}\underset{\sim}{V}_n) , \qquad (1.3)$$

where

$$\underset{\sim}{V}_n = nE(\underset{\sim}{\delta}_n - \underset{\sim}{\theta})'(\underset{\sim}{\delta}_n - \underset{\sim}{\theta}) . \qquad (1.4)$$

For normal F, the MLE of $\underset{\sim}{\theta}$ is given by $\underset{\sim}{\delta}_n = \bar{\underset{\sim}{X}}_n = n^{-1}\Sigma_{i=1}^n \underset{\sim}{X}_i$. For p greater than 2, a detailed account of (admissible and minimax) estimation of $\underset{\sim}{\theta}$ (for normal F) , based on the James-Stein rule, is given by Berger (1980). For possibly non-normal F, particularly, for distributions with heavy tails, these estimators may not be *robust* and *efficient*. Robust shrinkage R-estimators of locations in the multivariate case have recently been considered by Sen and Saleh (1985) ; parallel results on U-statistics are due to Sen (1984) . We propose to study the *preliminary test (PT)* and *shrinkage versions* of M-estimators of location in the multivariate case and to present their *asymptotic risk properties*. Along with the preliminary notions, the proposed estimators are introduced in Section 2. Section 3 deals with the general results on their asymptotic risks . Some general remarks are made in the concluding section.

2. THE PROPOSED ESTIMATORS

First, we introduce the *score functions*. Let $\psi = (\psi_1,\ldots,\psi_p)$ with $\psi_j = \{\psi_j(x);\ x \in E\}$, be defined by

$$\psi_j(x) = \psi_{j1}(x) + \psi_{j2}(x)\ ,\ x \in E\ ,\ j=1,\ldots,p, \tag{2.1}$$

where the ψ_{jk} , k=1,2, are both nondecreasing and skew-symmetric; ψ_{j1} is absolutely continuous on any bounded interval in E and ψ_{j2} is a step-function having finitely many jumps, j=1,...,p. Further, we assume that for suitable positive constants c_j , for all $|x|>c_j$, $\psi_j(x) = \psi_j(c_j)\,\mathrm{sign}\,x$, and ψ_j is non-constant on $[-c_j,c_j]$, so that

$$0 < \sigma_{\psi_j} = \int_E \psi_j^2(x)\,dF_{[j]}(x) < \infty\ ,\ 1\le j\le p\ , \tag{2.2}$$

where $F_{[j]}$ is the $j\underline{th}$ marginal of F. We also denote by $F_{[j\ell]}$ the (j,ℓ)th (bivariate) marginal of F , for $j \neq \ell = 1,\ldots,p$, and let

$$\underset{\sim}{\Sigma}_\psi = ((\sigma_{\psi_j\psi_\ell})) = ((\iint_{E^2} \psi_j(x)\psi_\ell(y)\,dF_{[j\ell]}(x,y))). \tag{2.3}$$

In the sequel, $\underset{\sim}{\Sigma}_\psi$ will be assumed to be positive definite (p.d.). Note that by virtue of the assumed symmetry of the $F_{[j]}$ and the skew-symmetry of the ψ_j , we have

$$\bar{\psi}_j = \int_E \psi_j(x)\,dF_{[j]}(x) = 0\ ,\ \text{for } j = 1,\ldots,p. \tag{2.4}$$

Let us now define for every real t and $n \geq 1$,

$$M_{nj}(t) = \Sigma_{i=1}^{n} \psi_j(X_{ij} - t) \ , \ j = 1,\ldots,p \ . \tag{2.5}$$

Note that for each j, $M_{nj}(t)$ is nonincreasing in t, and further, by (2.4), $M_{nj}(\theta_j)$ has mean 0; this leads us to the following estimators of $\underset{\sim}{\theta}$:

$$\hat{\underset{\sim}{\theta}}_n = (\hat{\theta}_{n1} , \ldots , \hat{\theta}_{np})' \tag{2.6}$$

where

$$\hat{\theta}_{nj} = (\sup\{t: M_{nj}(t) > 0\} + \inf\{ t: M_{nj}(t) < 0\})/2 , \tag{2.7}$$

for $j = 1,\ldots,p$. In the literature, these are known as the M-estimators [viz., Huber(1981) for a detailed account of the univariate theory]. Under the assumed regularity conditions, it is well known that $\hat{\underset{\sim}{\theta}}_n$ is a robust, componentwise median-unbiased, translation-invariant and consistent estimator of $\underset{\sim}{\theta}$. Further, as $n \to \infty$,

$$n^{\frac{1}{2}}(\hat{\underset{\sim}{\theta}}_n - \underset{\sim}{\theta}) \ \underset{\sim}{\mathcal{D}} \ \mathcal{N}_p(\underset{\sim}{0} \ , \underset{\sim}{\Gamma}^{-1}\underset{\sim}{\Sigma}_{\underset{\sim}{\psi}}\underset{\sim}{\Gamma}^{-1}) , \tag{2.8}$$

where $\underset{\sim}{\Gamma} = \text{Diag}(\gamma_1 , \ldots , \gamma_p)$ is defined by letting

$$\gamma_j = \int_E \psi_j(x)\{-f'_{[j]}(x)/f_{[j]}(x)\}dF_{[j]}(x), \ j=1,\ldots,p, \tag{2.9}$$

and it is assumed that the marginal probability density functions (p.d.f) $f_{[j]}$ are all absolutely continuous with finite *Fisher infomations*. We may refer to Jurečková and Sen (1981a,b; 1982) for some detailed accounts of the asymptotic theory of M-estimators.

In a shrinkage estimation problem, one has *apriori* some reasons to believe that $\underset{\sim}{\theta}$ is likely to be in a small neighbourhood of a *pivotal point* $\underset{\sim}{\theta}_o$; by virtue of the translation-invariance of the estimators, we may set , without any loss of generality, that $\underset{\sim}{\theta}_o = \underset{\sim}{0}$. In a preliminary test estimation (PTE) problem, one has the same belief, and one sets to test for the null hypothesis that $\underset{\sim}{\theta} = \underset{\sim}{\theta}_o$, first, and on the basis of this test, the ultimate estimator is chosen. In either case, one needs a suitable test-statistic which is incorporated in a suitable form in the construction of the estimators. Towards this, we define first

$$\underset{\sim}{S}_n = \underset{\sim}{S}_n(\underset{\sim}{\psi}) = ((s_{n,j\ell})) , \tag{2.10}$$

where

$$s_{n,j\ell} = n^{-1}\Sigma_{i=1}^{n} \psi_j(X_{ij} - \hat{\theta}_{nj})\psi_\ell(X_{i\ell} - \hat{\theta}_{n\ell}) , \tag{2.11}$$

for $j, \ell = 1,\dots,p$. Then, as in Singer and Sen (1985), we may consider a test statistic

$$\mathcal{L}_n = n^{-1}(\underset{\sim}{M}_n(\underset{\sim}{0}))'\underset{\sim}{S}_n^{-}(\underset{\sim}{M}_n(\underset{\sim}{0})) , \tag{2.12}$$

where $\underset{\sim}{M}_n(\underset{\sim}{0}) = (M_{n1}(0),\dots,M_{np}(0))'$ is defined as in (2.5), and $\underset{\sim}{S}_n^{-}$ is a (reflexive) generalized inverse of $\underset{\sim}{S}_n$. Under H_0, $\mathcal{L}_n$ has asymptotically chi quare distribution with p degrees of freedom (DF) when $\underset{\sim}{\Sigma}_\psi$ is p.d. Thus, if $\chi^2_{p,\alpha}$ stands for the upper 100α% point of the chi square distribution with p DF, for $\alpha \in (0,1)$, then corresponding to a desired significance level α , the preliminary test for H_0 is based on the following :

$$\mathcal{L}_n \quad \begin{matrix} \geq \chi^2_{p,\alpha} , \text{ reject } H_0 , \\ < \chi^2_{p,\alpha} , \text{ accept } H_0 . \end{matrix} \tag{2.13}$$

The PTE of $\underset{\sim}{\theta}$ is then defined by

$$\hat{\underset{\sim}{\theta}}_n^{PT} = I(\mathcal{L}_n \geq \chi^2_{p,\alpha})\hat{\underset{\sim}{\theta}}_n + I(\mathcal{L}_n < \chi^2_{p,\alpha})\underset{\sim}{0}, \tag{2.14}$$

where I(A) stands for the indicator function of the set A. On the other hand, in the shrinkage estimation, the statistic $\mathcal{L}_n$ is directly incorporated into the modified estimator. For this purpose, we need an estimator of the dispersion matrix $\underset{\sim}{\nu} = \underset{\sim}{\Gamma}^{-1}\underset{\sim}{\Sigma}_\psi\underset{\sim}{\Gamma}^{-1}$, appearing in (2.8). Following Jurečkova and Sen (1981b), we choose some positive a and define

$$\hat{\gamma}_{nj} = (2an^{\frac{1}{2}})^{-1}\{ M_{nj}(\hat{\theta}_{nj} - n^{-\frac{1}{2}}a) - M_{nj}(\hat{\theta}_{nj} + n^{-\frac{1}{2}}a)\}, \tag{2.15}$$

for $j = 1,\dots, p$ and let

$$\hat{\underset{\sim}{\Gamma}}_n = \text{Diag}(\hat{\gamma}_{n1},\dots,\hat{\gamma}_{np}) \quad \text{and} \quad \hat{\underset{\sim}{\nu}}_n = \hat{\underset{\sim}{\Gamma}}_n^{-1}\underset{\sim}{S}_n\hat{\underset{\sim}{\Gamma}}_n^{-1} . \tag{2.16}$$

It follows from the results in Singer and Sen (1985) that under the assumed regularity conditions, $\hat{\underset{\sim}{\nu}}_n$ is p.d. in probability, and it converges to $\underset{\sim}{\nu}$, in probability, as $n \to \infty$. Let

$$d_n = ch_p(\underset{\sim}{Q}\hat{\underset{\sim}{\nu}}_n) = \text{smallest characteristic root of } \underset{\sim}{Q}\hat{\underset{\sim}{\nu}}_n , \tag{2.17}$$

where $\underset{\sim}{Q}$ is defined as in (1.2). Since for possibly non-normal F and arbitrary $\underset{\sim}{\psi}$, the behaviour of $\mathcal{L}_n^{-1}$, when $\mathcal{L}_n$ is close to 0, is not that precisely known, and hence, we use a small truncation , and as in Sen (1984), consider the following version of a shrinkage M-estimator :

$$\hat{\theta}_n^S = \begin{cases} 0 & , \text{ if } \mathcal{L}_n < \varepsilon ; \\ (I - c d_n \mathcal{L}_n^{-1} Q^{-1} \hat{\nu}_n^{-1}) \hat{\theta}_n & , \text{ if } \mathcal{L}_n \geq \varepsilon , \end{cases} \tag{2.18}$$

where ε (> 0) is an arbitrary small number and c is a positive number ($0 < c < 2(p-2)$). It is also possible to replace c by a sequence $\{c_n\}$ of positive numbers , such that c_n converges to a limit c: $0 < c < 2(p-2)$. Ideally, c may be taken as $(p-2)$. Our primary interest lies in studying the asymptotic risks of the proposed PTE and shrinkage estimators and to compare them in a meaningful way.

3. ASYMPTOTIC RISK EFFICIENCY RESULTS

In view of the boundedness of the score functions, as has been assumed, we may not need stringent moment conditions on the d.f. F. We assume that for some positive b (not necessarily ≥ 1),

$$E_F \, ||X_1||^b \leq \Sigma_{j=1}^p E_F |X_{1j}|^b < \infty . \tag{3.1}$$

Then, the following results can directly be adapted from Jurečková and Sen (1982) :

(i) For every k (> 0), there exists a positive integer $n_o(k)$, such that $E_F \, |\hat{\theta}_{nj}|^k$ exists for every $n \geq n_o(k)$ (and $j=1,\dots,p$).

(ii) For every k (> 0),

$$\lim_{n\to\infty} E_F\{ n^{k/2} (\hat{\theta}_{nj} - \theta_j)^k \} = \sigma_{\psi_j}^k EZ^k , \tag{3.2}$$

where Z has the standard normal distribution, and, in particular,

$$nE[(\hat{\theta}_n - \theta)(\hat{\theta}_n - \theta)'] \to \nu , \text{ as } n \to \infty . \tag{3.3}$$

(iii) For every $j(=1,\dots,p)$,

$$n(\hat{\theta}_{nj} - \theta_j) = \gamma_j^{-1} M_{nj}(\theta_j) + \omega_{nj} , \tag{3.4}$$

where $n^{-\frac{1}{2}}|\omega_{nj}| \to 0$ almost surely (a.s.), as $n \to \infty$, and

$$E| n^{-\frac{1}{2}} \omega_{nj} |^k \to 0 , \text{ as } n \to \infty , \text{ for every } k > 0. \tag{3.5}$$

Now, by virtue of (1.3), (1.4) and (3.3), we obtain that for the classical M-estimator,

$$\rho(\hat{\theta};Q) = \lim_{n\to\infty} \{\rho_n(\hat{\theta}_n;Q)\}$$
$$= \mathrm{Tr}(Q[\lim_{n\to\infty} nE(\hat{\theta}_n - \theta)(\hat{\theta}_n - \theta)'])$$
$$= \mathrm{Tr}(Q\nu). \tag{3.6}$$

Next, by (2.14), we have $(\hat{\theta}_n^{PT} - \hat{\theta}_n) = I(\mathcal{L}_n < \chi^2_{p,\alpha})\hat{\theta}_n$, so that

$$n(\hat{\theta}_n^{PT} - \hat{\theta}_n)'Q(\hat{\theta}_n^{PT} - \hat{\theta}_n) = n(\hat{\theta}_n'Q\hat{\theta}_n)I(\mathcal{L}_n < \chi^2_{p,\alpha}). \tag{3.7}$$

Now, note that by (2.12),

$$\mathcal{L}_n \geq \max_{1\leq j\leq p} \{ n^{-1}M^2_{nj}(0)/s_{n,jj} \}, \tag{3.8}$$

so that

$$P\{\mathcal{L}_n < \chi^2_{p,\alpha}\} \leq \min_{1\leq j\leq p} P\{n^{-1}M^2_{nj}(0) < s_{n,jj}\chi^2_{p,\alpha}\}$$
$$\leq \min_{1\leq j\leq p} P\{n^{-1}M_{nj}(0) < n^{-1/2}s^{1/2}_{n,jj}\chi_{p,\alpha}\}. \tag{3.9}$$

Note that when θ_j is the true location, $E\{M_{nj}(0)|\theta_j\}$ is a negative or positive number according as θ_j is positive or negative, while $n^{-1/2}s^{1/2}_{n,jj}\chi_{p,\alpha}$ converges to 0 (a.s.) as $n\to\infty$. Further, under θ_j, $M_{nj}(0)$ involves i.i.d.r.v.'s (bounded valued) on which the exponential inequality [viz., Hoeffding (1963)] holds. Hence, using Theorem 3.3 of Jurečková and Sen (1982) and the above facts, we arrive at the conclusion that under $\theta \neq 0$, the right hand side of (3.9) can be made $O(n^{-2})$, for n adequately large. On the other hand, by the Holder inequality, for every $q > 1$, by (3.7),

$$E\{n(\hat{\theta}_n^{PT} - \hat{\theta}_n)'Q(\hat{\theta}_n^{PT} - \hat{\theta}_n)\} \leq n[E(\hat{\theta}_n'Q\hat{\theta}_n)^q]^{q^{-1}}[P\{\mathcal{L}_n < \chi^2_{p,\alpha}]^{1-q^{-1}} \tag{3.10}$$

where

$$E(\hat{\theta}_n'Q\hat{\theta}_n)^q \leq ch_1(Q)E(\hat{\theta}_n'\hat{\theta}_n)^q \leq p^{q-1}\mathrm{Tr}(Q)\Sigma_{j=1}^{p}E|\hat{\theta}_{nj}|^{2q}, \tag{3.11}$$

and the right hand side of (3.11) is finite for every $n \geq n_0(q)$. By (3.9), (3.10) and (3.11), we conclude that for any (fixed) $\theta \neq 0$, (3.7) converges in mean to 0 as $n\to\infty$, so that $\hat{\theta}_n^{PT}$ and $\hat{\theta}_n$ are asymptotically risk-equivalent for any (fixed) $\theta \neq 0$.

Let us next consider the case of the shrinkage estimator. By (1.3), (1.4) and (2.18), we have

$$E\{n(\hat{\theta}_n^S - \hat{\theta}_n)'Q(\hat{\theta}_n^S - \hat{\theta}_n)\} = I(\mathcal{L}_n < \varepsilon)\{n(\hat{\theta}_n'Q\hat{\theta}_n)\} +$$
$$+ I(\mathcal{L}_n \geq \varepsilon)E\{d_n^2\mathcal{L}_n^{-2}c^2(n\hat{\theta}_n'\hat{\nu}_n^{-1}Q^{-1}\hat{\nu}_n^{-1}\hat{\theta}_n)\}. \tag{3.12}$$

Now, using the basic inequality in (3.10) of Sen and Saleh (1985) and proceeding as in their (3.8) through (3.11), it follows that for every (fixed) $\underset{\sim}{\theta} \neq \underset{\sim}{0}$,

$$\limsup_{n\to\infty} E\{n(\hat{\underset{\sim}{\theta}}{}^S_n - \hat{\underset{\sim}{\theta}}_n)'\underset{\sim}{Q}(\hat{\underset{\sim}{\theta}}{}^S_n - \hat{\underset{\sim}{\theta}}_n) = 0 , \tag{3.13}$$

so that for every (fixed) $\underset{\sim}{\theta} \neq \underset{\sim}{0}$, the shrinkage and the classical M-estimators are asymptotically risk-equivalent. The situation is, however, quite different for local shift alternatives, and, we plan to explore the same.

As has been mentioned earlier in Section 1 that both the PTE and shrinkage estimation works out well only for shrinking neighbourhoods of the pivotal point (here $\underset{\sim}{0}$). In the asymptotic case, this shrinking neighbourhood coincides with the usual Pitman-type (local) translation alternatives. Hence, we consider a double sequence $\{\underset{\sim}{X}_{ni}$, $i=1,\ldots,n;\ n \geq 1\}$ of (row-wise) i.i.d.r.v.'s, where for each n, the distribution of $\underset{\sim}{X}_{ni}$ is given by (1.1) with $\underset{\sim}{\theta} = \underset{\sim}{\theta}_n = n^{-\frac{1}{2}}\underset{\sim}{\lambda}$, and $\underset{\sim}{\lambda}$ belongs to a compact set C (containing $\underset{\sim}{0}$ as an inner point). By virtue of the translation-invariance of the $\hat{\underset{\sim}{\theta}}_n$, we may still work with the $\underset{\sim}{X}_{ni} - \underset{\sim}{\theta}_n$, and justify all the asymptotic results considered earlier. Therefore , (3.6) remains true for such local alternatives as well, though the other asymptotic risk expressions remain to be worked out. For later use, we denote by $\{K_n\}$, the sequence of alternatives for which (1.1) holds with $\underset{\sim}{\theta} = \underset{\sim}{\theta}_n = n^{-\frac{1}{2}}\underset{\sim}{\lambda}$, for a given $\underset{\sim}{\lambda} \neq \underset{\sim}{0}$, so that for the PTE,we have

$$\begin{aligned}\rho(\text{PTM};\underset{\sim}{\lambda}) &= \lim_{n\to\infty}\{\rho_n(\hat{\underset{\sim}{\theta}}{}^{PT};\underset{\sim}{Q}) \mid K_n\} \\ &= \lim_{n\to\infty} P\{\mathcal{L}_n < \chi^2_{p,\alpha} \mid K_n\}\ (\underset{\sim}{\lambda}'\underset{\sim}{Q}\underset{\sim}{\lambda}) + \\ &\quad + \lim_{n\to\infty} E\{I(\mathcal{L}_n \geq \chi^2_{p,\alpha})n(\hat{\underset{\sim}{\theta}}_n - \underset{\sim}{\theta}_n)'\underset{\sim}{Q}(\hat{\underset{\sim}{\theta}}_n - \underset{\sim}{\theta}_n)|K_n\}.\end{aligned} \tag{3.14}$$

Note that

$$\lim_{n\to\infty} P\{\mathcal{L}_n < \chi^2_{p,\alpha} \mid K_n\} = H_p(\chi^2_{p,\alpha};\Delta) \tag{3.15}$$

where

$$\Delta = (\underset{\sim}{\lambda}'\underset{\sim}{\nu}^{-1}\underset{\sim}{\lambda}) \tag{3.16}$$

and $H_q(x;\delta)$ stands for a noncentral chi square distribution function with q DF and noncentrality parameter δ . For the second term on the right hand side of (3.14), we make use of (3.2) (for k = 4) , (3.3), (3.4) and (2.12), and conclude that the second term on the right hand side of (3.14) converges to

$$\mathrm{Tr}(\underset{\sim}{Q}\underset{\sim}{\nu})[1 - H_{p+2}(\chi^2_{p,\alpha};\Delta)] - (\underset{\sim}{\lambda}'\underset{\sim}{Q}\underset{\sim}{\lambda})[H_p(\chi^2_{p,\alpha};\Delta) - 2H_{p+2}(\chi^2_{p,\alpha};\Delta) + H_{p+4}(\chi^2_{p,\alpha};\Delta)] \; ; \tag{3.17}$$

in this context, we have also made use of the formulale for the incomplete moments of multinormal distributions [viz. Sen and Saleh (1979)]. Thus, from (3.14) through (3.17), we obtain that

$$\rho(\mathrm{PTM};\underset{\sim}{\lambda}) = [1- H_{p+2}(\chi^2_{p,\alpha};\Delta)]\mathrm{Tr}(\underset{\sim}{Q}\underset{\sim}{\nu}) + [2H_{p+2}(\chi^2_{p,\alpha};\Delta) - H_{p+4}(\chi^2_{p,\alpha};\Delta)](\underset{\sim}{\lambda}'\underset{\sim}{Q}\underset{\sim}{\lambda}). \tag{3.18}$$

Note that (3.6) remains in tact under $\{K_n\}$ as well, and hence, for the classical M-estimator, we have

$$\rho(M;\underset{\sim}{\lambda}) = \mathrm{Tr}(\underset{\sim}{Q}\underset{\sim}{\nu}), \text{ for every (fixed) } \underset{\sim}{\lambda} \in E^p. \tag{3.19}$$

From (3.18) and (3.19), we have for every (fixed) $\underset{\sim}{\lambda}$,

$$\rho(\mathrm{PTM};\underset{\sim}{\lambda})/\rho(M;\underset{\sim}{\lambda}) = 1 - H_{p+2}(\chi^2_{p,\alpha};\Delta) + [(\underset{\sim}{\lambda}'\underset{\sim}{Q}\underset{\sim}{\lambda})/\mathrm{Tr}(\underset{\sim}{Q}\underset{\sim}{\nu})][2H_{p+2}(\chi^2_{p,\alpha};\Delta) - H_{p+4}(\chi^2_{p,\alpha};\Delta)]. \tag{3.20}$$

Since $H_q(x;\delta) \geq H_{q+2}(x;\delta)$ for every $x \in E$ and $q \geq 1$ $(\delta \geq 0)$, we obtain from (3.20) that

(i) $(\underset{\sim}{\lambda}'\underset{\sim}{Q}\underset{\sim}{\lambda})/\mathrm{Tr}(\underset{\sim}{Q}\underset{\sim}{\nu}) > 1 \Rightarrow \rho(\mathrm{PTM};\underset{\sim}{\lambda})/\rho(M;\underset{\sim}{\lambda}) > 1$, (3.20)

(ii) $(\underset{\sim}{\lambda}'\underset{\sim}{Q}\underset{\sim}{\lambda})/\mathrm{Tr}(\underset{\sim}{Q}\underset{\sim}{\nu}) < \frac{1}{2} \Rightarrow \rho(\mathrm{PTM};\underset{\sim}{\lambda})/\rho(M;\underset{\sim}{\lambda}) < 1$. (3.21)

In (3.20), the excess over 1 (for the asymptotic risk ratio) may not be very much, and this ratio converges to 1 as $\Delta \to \infty$. On the other hand, the deficit in (3.21) may be more appreciable; in particular, under H_0: $\underset{\sim}{\lambda} = \underset{\sim}{0}$, the risk ratio is $1-H_{p+2}(\chi^2_{p,\alpha};0)$, and it depends on p and α. Thus, the PTE performs better under H_0 and for small deviations from H_0 , while, in the tail, the classical M-estimator performs better. None dominates over the other over the entire range, though the PTE has a more robust performance.

Let us now consider the case of the shrinkage estimator. From (2.18), we obtain that under $\{K_n\}$ and the assumed regularity conditions,

$$\rho(SM;\lambda) = \lim_{n\to\infty} [nE\{(\hat{\theta}_n^S - \theta_n)'Q(\hat{\theta}_n^S - \theta_n) \mid K_n\}]$$
$$= (\lambda'Q\lambda)[\lim_{n\to\infty} E\{I(\mathcal{L}_n \le \varepsilon) \mid K_n\}]$$
$$+ \lim_{n\to\infty} E\{I(\mathcal{L}_n > \varepsilon)n(\hat{\theta}_n - \theta_n)'Q(\hat{\theta}_n - \theta_n) \mid K_n\}$$
$$-2c[\lim_{n\to\infty} E\{d_n\mathcal{L}_n^{-1}n(\hat{\theta}_n - \theta_n)'\hat{\nu}_n^{-1}\hat{\theta}_n I(\mathcal{L}_n > \varepsilon) \mid K_n\}]$$
$$+ c^2[\lim_{n\to\infty} E\{d_n^2\mathcal{L}_n^{-2}n\hat{\theta}_n'\hat{\nu}_n^{-1}Q^{-1}\hat{\nu}_n^{-1}\hat{\theta}_n I(\mathcal{L}_n > \varepsilon) \mid K_n\}]. \quad (3.22)$$

The first term on the right hand side (rhs) of (3.22) is equal to

$$(\lambda'Q\lambda)H_p(\varepsilon;\Delta). \quad (3.23)$$

Also, proceeding as in (3.17), the second term on the rhs of (3.22) reduces to

$$[1-H_{p+2}(\varepsilon;\Delta)]\mathrm{Tr}(Q\nu) - (\lambda'Q\lambda)[H_p(\varepsilon;\Delta)-2H_{p+2}(\varepsilon;\Delta)+H_{p+4}(\varepsilon;\Delta)]. \quad (3.24)$$

For the third term, we note that

$$|n(\hat{\theta}_n - \theta_n)'\hat{\nu}_n^{-1}\hat{\theta}_n d_n| = |n^{\frac{1}{2}}(\hat{\theta}_n - \theta_n)'\hat{\nu}_n^{-\frac{1}{2}}\hat{\nu}_n^{-\frac{1}{2}} n^{\frac{1}{2}}\hat{\theta}_n d_n|$$
$$\le \{[n(\hat{\theta}_n - \theta_n)'\hat{\nu}_n^{-1}(\hat{\theta}_n - \theta_n) d_n][d_n . n\hat{\theta}_n'\hat{\nu}_n^{-1}\hat{\theta}_n]\}^{\frac{1}{2}}$$
$$\le \{n(\hat{\theta}_n - \theta_n)'Q(\hat{\theta}_n - \theta_n) . n\hat{\theta}_n'Q\hat{\theta}_n\}^{\frac{1}{2}}$$
$$\le [n(\hat{\theta}_n - \theta_n)'Q(\hat{\theta}_n - \theta_n) + n\hat{\theta}_n'Q\hat{\theta}_n]/2, \quad (3.25)$$

while, $\mathcal{L}_n^{-1}$ is bounded from above by $\varepsilon^{-1}(< \infty)$. Hence, the integrability condition may easily be verified by using (3.2) through (3.5), and, following some routine steps (and denoting by W a r.v. having the p-variate normal distribution with mean $\omega = \nu^{-\frac{1}{2}}\lambda$ and dispersion matrix I_p), we conclude that the third term on the rhs of (3.22) is equal to

$$-2c.ch_p(Q\nu)E\{(W - \omega)'W(W'W)^{-1} - I(W'W \le \varepsilon)W'(W-\omega)(W'W)^{-1}\}. \quad (3.26)$$

For the second term under the expectation, we may proceed as in Sen and Saleh (1985) and conclude that it is $O(\varepsilon^{(p-1)/2})$, while, for the first term, we may use the classical Stein identity [viz., Appendix B of Judge and Bock (1978)], and conclude that (3.26) is

$$-2c.ch_p(Q\nu)\{1 - \Delta E(\chi_{p+2}^{-2}(\Delta))\} + O(\varepsilon^{(p-1)/2}), \quad (3.27)$$

where $W'W = \chi_p^2(\Delta)$ has the noncentral chi square distribution with p DF and noncentrality parameter $\Delta = \omega'\omega = \lambda'\nu^{-1}\lambda$. Also, noting that $d_n^2 \hat{\theta}_n' \hat{\nu}_n^{-1} Q^{-1} \hat{\nu}_n^{-1} \hat{\theta}_n \leq \hat{\theta}_n' Q \hat{\theta}_n$, we obtain on using (3.2) through (3.5) that the last term on the rhs of (3.22) is equal to

$$[c.ch_p(Q\nu)]^2 \{Tr(Q^{-1}\nu^{-1})E(\chi_{p+2}^{-4}(\Delta)) + \Delta^* E(\chi_{p+4}^{-4}(\Delta)) - E[I(W'W \leq \varepsilon)(W'W)^{-2}(W'A^*W] \} \quad (3.28)$$

where

$$\Delta^* = \lambda'\nu^{-1}Q^{-1}\nu^{-1}\lambda \quad \text{and} \quad A^* = \nu^{-1/2}Q\nu^{-1/2} . \quad (3.29)$$

As in Sen and Saleh (1985), the second term on the rhs of (3.28) is $O(\varepsilon^{p-2})$, and hence, using (3.22) through (3.28), we obtain that

$$\rho(SM;\lambda) = Tr(Q\nu) - 2c.ch_p(Q\nu)\{ 1 - \Delta E(\chi_{p+2}^{-2}(\Delta))\} + [c.ch_p(Q\nu)]^2\{ Tr(Q^{-1}\nu^{-1})E(\chi_{p+2}^{-4}(\Delta)) + \Delta^* E(\chi_{p+4}^{-4}(\Delta))\} + O(\varepsilon^{p/2}) + O(\varepsilon^{(p-1)/2}) + O(\varepsilon^{p-2}) , \quad (3.30)$$

as $H_q(\varepsilon;\delta) \leq H_q(\varepsilon;0) = O(\varepsilon^{q/2})$, for every $q \geq 1$ and $\delta \geq 0$. Since p is taken to be greater than 2 (for the shrinkage estimator), it follows that by choosing ε adequately small, the rhs of (3.30) can be approximated by the first three terms . On this approximation, the results in Berger et al.(1977) directly applies, and we obtain that for every $c : 0 < c < 2(p-2)$ and $\lambda \in E^p$,

$$\lim_{\varepsilon \downarrow 0} \rho(SM;\lambda)/\rho(M;\lambda) \leq 1 \text{ with strict } < \text{ at } \Delta = 0. \quad (3.31)$$

Thus, the shrinkage M-estimator dominates over the classical one . In particular, under H_0 (i.e., $\lambda = 0$), $c = (p-2)$ and $Q\nu = I$, the left hand side of (3.31) reduces to $4(p-1)p^{-2}$, and is < 1 for all $p > 2$; the reduction is substantial for large values of p.

Finally, from (3.18) and (3.30), we obtain that

$$\lim_{\varepsilon \downarrow 0}\rho(SM;\lambda)/\rho(PTEM;\lambda) =\{ Tr(Q\nu) - 2c.ch_p(Q\nu)[1- \Delta E(\chi_{p+2}^{-2}(\Delta))] + (c.ch_p(Q\nu))^2[Tr(Q^{-1}\nu^{-1})E(\chi_{p+2}^{-4}(\Delta))+\Delta^* E(\chi_{p+4}^{-4}(\Delta))]\}. [1-H_{p+2}(\chi_{p,\alpha}^2;\Delta)]Tr(Q\nu)+ (\lambda'Q\lambda)[2H_{p+2}(\chi_{p,\alpha}^2;\Delta)-H_{p+4}(\chi_{p,\alpha}^2;\Delta)]\}^{-1},$$
$$\lambda \in E^p , \Delta \geq 0, \Delta^* \geq 0 . \quad (3.32)$$

By (3.20) and (3.31), we conclude that for all λ , for which $\lambda'Q\lambda$ exceeds $Tr(Q\nu)$, (3.32) is less than 1 (though it converges to 1 as

λ moves away from 0, i.e., Δ or Δ^* converges to $+\infty$); however, the deficit from the upper asymptote (i.e., 1) is usually quite small for moderate to large values of Δ. On the other hand, the picture may be quite different for Δ close to 0 ; see (3.21) in this respect. Letting $\eta_j = ch_j(Q\nu)/Tr(Q\nu)$, for $j = 1,\ldots,p$ and noting that (i) $\eta_p \leq p^{-1}$ and (ii) $Tr(Q\nu)Tr(Q^{-1}\nu^{-1}) = (\Sigma\eta_j)(\Sigma\eta_j^{-1}) \geq p^2$ (where the equality sign holds when the η_j are all equal), we obtain that under H_0 (i.e., $\lambda = 0$), (3.32) reduces to

$$\{1 - 2c\eta_p + c^2\eta_p^2[Tr(Q\nu)Tr(Q^{-1}\nu^{-1})/p(p-2)]\}\{1 - H_{p+2}(\chi^2_{p,\alpha};0)\}^{-1}$$
$$\geq \{1 - 2c\eta_p + c^2\eta_p^2(p-2)^{-1}p\}/\{1 - H_{p+2}(\chi^2_{p,\alpha};0)\}. \qquad (3.33)$$

The denominator on the rhs of (3.33) is free from c, while the numerator is minimized at $c = c_o$ where $c_o = (p-2)p^{-1}\eta_p^{-1}$ ($\geq (p-2)$), so that (3.33) is bounded from below by

$$\{1 - 2c_o\eta_p + c_o^2\eta_p^2(p-2)^{-1}p\}/\{1 - H_{p+2}(\chi^2_{p,\alpha};0)\}$$
$$= (2/p)\{1 - H_{p+2}(\chi^2_{p,\alpha};0)\}^{-1}, \qquad (3.34)$$

and (3.34) exceeds one whenever $1 - H_{p+2}(\chi^2_{p,\alpha};0) < 2/p$ i.e.,

$$[H_{p+2}(\chi^2_{p,\alpha};0) > (p-2)/p] \Rightarrow [(3.33) \text{ is } > 1]. \qquad (3.35)$$

Now, looking at the central chi square distribution, we gather that (3.33) exceeds 1 for a significant range of values of p and α. For example, for $\alpha = 0.05$, for all $p \leq 21$ and for $\alpha = 0.10$, for all $p \leq 11$. In fact, the smaller is the significance level α, the larger is the range of p for which (3.33) is greater than 1. This clearly indicates that generally the shrinkage M-estimator fails to dominate over the PTE M-estimator under H_0 and for smaller values of λ. We shall elaborate this a bit more in the next section.

4. SOME GENERAL REMARKS

In comparing the PTE and shrinkage versions of M-estimators, we may note first that the PTE does not presuppose that p is greater than 2 , while the shrinkage estimator does so. Thus, the PTE may have a wider scope of applicability. On the other hand, the shrinkage estimator dominates over the classical one, while the PTE does so only for λ close to 0 (i.e., for small values of Δ). For larger values of Δ , clearly the shrinkage estimator is superior to the PTE , though both of these may behave quite closely for Δ not too small. Next, we have noted after (2.18) that the constant c may also be replaced by a sequence $\{c_n\}$ for which $\lim_{n\to\infty} c_n = c$ exists. The existence of this limit will ensure the convergence of the risk formulae. In this context, naturally, one may raise the question on the choice of ε and c_n in the definition of the shrinkage estimator in (2.18). Basically, for small n, the constant c_n should also be chosen small, while, for large n, c_n may be taken to equal to (p-2) (as was recommended for the normal mean case by Berger et al.(1977)). Since in (3.30), we have neglected terms of the $O(\varepsilon^{(p-1)/2})$, the choice of ε is naturally dependent on p as well. The larger is the value of p, the larger is the quantity ε , so that $\varepsilon^{(p-1)/2}$ is smaller than a specified $\eta > 0$. It may also be noted that in (2.18) one needs the estimator $\hat{\nu}_n$ which in turn depends on the estimators $\hat{S}_n$ and $\hat{\Gamma}_n$. For $\hat{\Gamma}_n$, one may use the stronger convergence results studied in Jurečkova and Sen (1981a,b; 1982) , while the convergence rates for S_n , generally, depends on the underlying F (through the dependence of the p characters) . We have tacitly assumed that d_n , defined in (2.17), converges in probability (or a.s.) to a positive limit. This,in fact, demands that S_n converges (in probability or a.s.) to a p.d. matrix as $n \to \infty$, and,in order that such a convergence is uniform (in λ ,under $\{K_n\}$), we may need to assume that F is non-singluar (in the sense that $ch_p(\Sigma_\psi)$ is strictly positive over a suitable class of such F's). For normal d.f.'s, it suffices to formulate this condition in terms of the covariance matrix, while, Σ_ψ takes on this role for a general F and a set of chosen score functions. Finally, we have confined ourselves to bounded score functions on which the M-estimators are defined. It is possible to relax this assumption and to incorporate (as in Jurečkova and Sen (1981 a,b)) unbounded scores functions satisfying suitable moment conditions. As we need to justify (3.2) through (3.5) (needed for the desired integrability conditions to be verfied), we may endup , in (3.1), with a value of b greater than k (> 2). Alternatively, instead of the asymptotic risk, one may consider the asymptotic distributional risk (ADR) computed directly from the asymptotic distribution of an estimator (which may or may not agree with the asymptotic risk), and , in the light of ADR, we would have the same picture as in Section 3 , valid for a broader class of score functions . However, bounded score functions are very popular (on the ground of robustness against gross errors and outliers) in M-estimation, and hence, we need not compromise on the weaker ADR results at the cost of unbounded scores and / or higher order moments of the underlying d.f. F .

ACKNOWLEDGEMENTS

This research has been supported by the Natural Sciences and Engineering Research Council of Canada, Grant No. A3088 and by the (U.S.) National Heart, Lung and Blood Institute, Contract No. NIH-NHLBI-2243-L from the National Institutes of Health.

BIBLIOGRAPHY

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

Berger, J.O., Bock, M.E., Brown, L.D., Casella, G. and Gleser, L. (1977). Minimax estimation of a normal mean vector for arbitrary quadratic loss and unknown covariance matrix. Ann. Statist. 5 , 763-771.

Berger, J.O. and Haff, L.R. (1983). A class of minimax estimators of a normal mean vector for arbitrary quadratic loss and unknown covariance matrix. Statist. Dec. 1, 105-129.

James, W. and Stein, C. (1961). Estimation with quadratic loss. Proc. 4th Berkeley Symp. Math. Statist. Probability 1, 293-325.

Judge, G.G. and Bock, M.E. (1978). The Statistical Implications of Pretest and Stein-Rule Estimators in Economics. North Holland, Amsterdam.

Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. J. Amer. Statist. Assoc. 58, 13-30.

Huber, P.J. (1981). Robust Statistics. John Wiley, New York.

Jurečková, J.J. and Sen, P.K. (1981a). Invariance principles for some stochastic processes related to M-estimators and their role in sequential statistical inference. Sankhya, Ser.A 43, 191-210.

Jurečková, J.J. and Sen, P.K. (1981b). Sequential procedures based on M-estimators with discontinuous score functions. J. Statist. Plan. Infer. 5, 253-266.

Jurečková, J.J. and Sen, P.K. (1982). M-estimators and L-estimators of location: Uniform integrability and asymptotic risk-efficient sequential versions. Sequen. Anal. 1, 27-56.

Sen, P.K. (1984). A James-Stein detour of U-statistics. Commun. Statist. Theor. Meth. A 13, 2725-2747.

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

Singer, J.M. and Sen, P.K. (1985). M-Methods in multivariate linear models. J. Multivar. Anal. 15 (in press).

Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. Proc. 3rd Berkeley Symp. Math. Statist. Probability 1, 197-206.