The Annals of Statistics
1985, Vol. 13, No. 1, 272–281

ON SOME SHRINKAGE ESTIMATORS OF MULTIVARIATE LOCATION[1]

BY PRANAB KUMAR SEN AND A.K.MD. EHSANES SALEH

University of North Carolina, Chapel Hill, and Carleton University, Ottawa

For a continuous and diagonally symmetric multivariate distribution, incorporating the idea of preliminary test estimators, a variant form of the James–Stein type estimation rule is used to formulate some shrinkage estimators of location based on rank statistics and U-statistics. In an asymptotic setup, the relative risks for these shrinkage estimators are shown to be smaller than their classical counterparts.

1. Introduction. Let $\mathbf{X}_i = (X_{i1}, \cdots, X_{ip})'$, $i = 1, \cdots, n$, be n independent and identically distributed random vectors (i.i.d.r.v) having a $p(\geq 1)$-variate continuous distribution function (d.f.) F_θ, defined on the Euclidean space E^p. F_θ is assumed to be *diagonally symmetric* about its *location* $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_p)'$, i.e.,

$$F_\theta(\mathbf{x}) = F(\mathbf{x} - \boldsymbol{\theta}), \quad \mathbf{x} \in E^p, \tag{1.1}$$

where F is diagonally symmetric about $\mathbf{0}$. Based on $\mathbf{X}_1, \cdots, \mathbf{X}_n$, let $\boldsymbol{\delta}_n = (\delta_{n1}, \cdots, \delta_{np})'$ be an estimator of $\boldsymbol{\theta}$, and consider a *quadratic loss function*

$$L(\boldsymbol{\delta}_n, \boldsymbol{\theta}) = n(\boldsymbol{\delta}_n - \boldsymbol{\theta})'\mathbf{Q}(\boldsymbol{\delta}_n - \boldsymbol{\theta}), \tag{1.2}$$

for some given positive definite (p.d.) matrix $\mathbf{Q}$. The *risk* is then given by

$$\rho_n(\boldsymbol{\delta}, \boldsymbol{\theta}) = EL(\boldsymbol{\delta}_n, \boldsymbol{\theta}) = \mathrm{Tr}(\mathbf{Q}\mathbf{V}_n), \quad \text{where} \quad \mathbf{V}_n = nE(\boldsymbol{\delta}_n - \boldsymbol{\theta})(\boldsymbol{\delta}_n - \boldsymbol{\theta})'. \tag{1.3}$$

For normal F and $p \geq 3$, in view of the inadmissibility of the sample mean $\bar{\mathbf{X}}_n = n^{-1} \sum_{i=1}^n \mathbf{X}_i$ (cf. Stein, 1956), a simple (nonlinear) admissible estimator was proposed by James and Stein (1961). Since then, this theory has been extensively studied (in a parametric setup) by a host of workers; a detailed account of these developments is given by Berger (1980). For possible nonnormal F, the sample mean $\bar{\mathbf{X}}_n$ may not be very robust, and may even be quite inefficient for distributions with heavy tails. Robust rank based (R-) estimators of location in the multivariate case have been studied by Sen and Puri (1969), Puri and Sen (1971), and others. Also, preliminary test R-estimators were studied by Saleh and Sen (1978) and Sen and Saleh (1979), among others. The object of the present study is to consider suitable shrinkage R-estimators of $\boldsymbol{\theta}$, and also to present briefly the shrinkage U-statistics which contain the sample mean as a special case.

Received May 1983; revised August 1984.

[1] Work supported partially by the (U.S.) National Heart, Lung and Blood Institute, Contract NIH-NHLBI-71-2243-L from the N.I.H., and partially by the Natural Sciences and Engineering Research Council of Canada, Grant No. A3088.

AMS 1980 *subject classifications.* Primary 62C15, 62G05; secondary 62G99.

Key words and phrases. Asymptotic risk, James–Stein rule, local alternatives, preliminary test estimator, rank estimator, robustness, shrinkage estimates, U-statistics.

Along with the preliminary notions, the proposed shrinkage R-estimators are introduced in Section 2. The general results on the asymptotic risks of these estimators are presented in Section 3. The case of shrinkage estimators based on U-statistics (not necessarily pertaining to the location model) is treated briefly in the concluding section.

2. Shrinkage R-estimators. For every $n(\geq 1)$ and $j(=1, \cdots, p)$, we define a set of scores by letting

$$a_{nj}^{+}(k) = E\phi_j^{+}(U_{nk}) \quad \text{or} \quad \phi_j^{+}(EU_{nk}), \quad \text{for} \quad k = 1, \cdots, n, \tag{2.1}$$

where $U_{n1} < \cdots < U_{nn}$ are the ordered r.v. of a sample of size n from the uniform (0, 1) d.f., and, for every $u \in (0, 1)$,

$$\phi_j^{+}(u) = \phi_j((1 + u)/2), \quad \phi_j(u) + \phi_j(1 - u) = 0; \tag{2.2}$$

the ϕ_j are all assumed to be nondecreasing and square integrable. For every real b, let $R_{ij}^{+}(b)$ be the rank of $|X_{ij} - b|$ among $|X_{1j} - b|, \cdots, |X_{nj} - b|$, for $i = 1, \cdots, n$, $j = 1, \cdots, p$. Consider the statistics

$$T_{nj}(b) = n^{-1} \textstyle\sum_{i=1}^{n} \operatorname{sgn}(X_{ij} - b) a_{nj}^{+}(R_{ij}^{+}(b)), \quad j = 1, \cdots, p. \tag{2.3}$$

Note that $T_{nj}(b)$ is $\searrow$ in b (viz., Puri and Sen, 1971, Chapter 6), and we set

$$\hat{\theta}_{nj} = \tfrac{1}{2}(\sup\{b: T_{nj}(b) > 0\} + \inf\{b: T_{nj}(b) < 0\}), \quad j = 1, \cdots, p; \tag{2.4}$$

$$\hat{\boldsymbol{\theta}}_n = (\hat{\theta}_{n1}, \cdots, \hat{\theta}_{np})'. \tag{2.5}$$

$\hat{\boldsymbol{\theta}}_n$ is an R-estimator and is known to be a robust, translation-invariant and consistent estimator having the (coordinatewise) median-unbiasedness and other desirable properties too. Let $F_{[j]}$ be the jth marginal d.f. corresponding to the d.f. F $(1 \leq j \leq p)$ and let $F_{[j\ell]}$, $j \neq \ell = 1, \cdots, p$ be the bivariate marginal d.f.s. We assume that $F_{[j]}$ possesses an absolutely continuous probability density function $f_{[j]}$ and set

$$\psi_j(u) = -f'_{[j]}(F_{[j]}^{-1}(u))/f_{[j]}(F_{[j]}^{-1}(u)), \quad u \in (0, 1), \quad j = 1, \cdots, p.$$

Let then

$$\nu_{j\ell} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi_j(F_{[j]}(x))\phi_\ell(F_{[\ell]}(y))\, dF_{[j\ell]}(x, y), \quad j, \ell = 1, \cdots, p; \tag{2.6}$$

$$\xi_j = \int_0^1 \phi_j(u)\psi_j(u)\, du, \quad j = 1, \cdots, p; \tag{2.7}$$

$$\boldsymbol{\nu} = ((\nu_{j\ell})), \quad \boldsymbol{\xi} = \operatorname{Diag}(\xi_1, \cdots, \xi_p); \tag{2.8}$$

$$\boldsymbol{\Gamma} = ((\gamma_{j\ell})) = \boldsymbol{\xi}^{-1}\boldsymbol{\nu}\boldsymbol{\xi}^{-1}. \tag{2.9}$$

Then, it is known (cf. Puri and Sen, 1971, Chapter 6) that

$$n^{1/2}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \to_{\mathscr{D}} \mathscr{N}_p(\mathbf{0}, \boldsymbol{\Gamma}). \tag{2.10}$$

In a shrinkage estimation problem, one has, a priori, some reason to believe that θ is likely to lie in a small region containing a specified point θ_0; without any loss of generality, we may set $\theta_0 = \mathbf{0}$. In a preliminary test estimation (PTE) problem, one sets to test for this hypothesis first, before choosing the estimator, while in a shrinkage estimation, the test statistic is itself incorporated in the estimator to adjust for possible shifts from the specified θ_0. In either case, we need a suitable test statistic, and towards this, we define $\mathbf{M}_n^* = ((m_{n,j\ell}^*))$ by letting

$$m_{n,j\ell}^* = n^{-1} \textstyle\sum_{i=1}^n a_{nj}^+(R_{ij}^+(0))a_{n\ell(R_{i\ell}^+}(0))\text{sgn } X_{ij}\text{sgn } X_{i\ell}, \tag{2.11}$$

for $j, \ell = 1, \cdots, p$. Then, as in Sen and Puri (1967), the test-statistic used is

$$\mathscr{L}_n = n(\mathbf{T}_n)'(\mathbf{M}_n^*)^-(\mathbf{T}_n) \tag{2.12}$$

where $\mathbf{T}_n = (T_{n1}(0), \cdots, T_{np}(0))'$ and $(\mathbf{M}_n^*)^-$ is a (reflexive) generalized inverse of $\mathbf{M}_n^*$. Following James and Stein (1961), one may consider then the estimator

$$\theta_n^{\text{JS}} = \{1 - a/\mathscr{L}_n\}\hat{\theta}, \tag{2.13}$$

where a is an appropriate constant. Since $\mathscr{L}_n$ may assume the value 0 with a positive probability, there may be a small technical problem with the computation of the risk of $\hat{\theta}_n^{\text{JS}}$. Moreover, $\hat{\theta}_n^{\text{JS}}$ may not dominate over $\hat{\theta}_n$ unless in (1.2), $\mathbf{Q} = \mathbf{\Gamma}^{-1}$. To eliminate these problems, we consider the following formulation of shrinkage R-estimators of multivariate locations.

Parallel to (2.1)–(2.2), we let $a_{nj}(k) = E\phi_j(U_{nk})$ or $\phi_j(EU_{nk})$, for $k = 1, \cdots, n$, $j = 1, \cdots, p$, and define $\mathbf{M}_n = ((m_{n,j\ell}))$ by letting

$$m_{n,j\ell} = n^{-1} \textstyle\sum_{i=1}^n a_{nj}(R_{ij})a_{n\ell}(R_{i\ell}), \quad j, \ell = 1, \cdots, p, \tag{2.14}$$

where R_{ij} is the rank of X_{ij} among $X_{1j}, \cdots, X_{nj}$, for $i = 1, \cdots, n$; $j = 1, \cdots, p$. Then $\mathbf{M}_n$ is a translation-invariant, consistent and robust estimator of $\boldsymbol{\nu}$ (cf. Puri and Sen, 1971, Chapter 5). Also, we define $\hat{\xi}_n = \text{Diag}(\hat{\xi}_{n1}, \cdots, \hat{\xi}_{np})$ by letting

$$\hat{\xi}_{nj} = n^{1/2}\{T_{nj}(\hat{\theta}_{nj} - n^{-1/2}a) - T_{nj}(\hat{\theta}_{nj} + n^{-1/2}a)\}/(2a), \quad j = 1, \cdots, p, \tag{2.15}$$

where a is some pre-fixed positive number. Let then

$$\hat{\mathbf{\Gamma}}_n = \hat{\xi}_n^{-1}\mathbf{M}_n\hat{\xi}_n^{-1}; \quad d_n = \text{smallest characteristic root of } \mathbf{Q}\hat{\mathbf{\Gamma}}_n. \tag{2.16}$$

Analogous to the normal case, treated in Berger et al. (1977), one may consider a constant c: $0 < c < 2(p - 2)$ and (for $p \geq 3$) consider an estimator of the form $(\mathbf{I} - cd_n\mathscr{L}_n^{-1}\mathbf{Q}^{-1}\hat{\mathbf{\Gamma}}_n^{-1})\hat{\theta}_n$. However, in view of the fact that $\mathscr{L}_n$ may be equal to 0 with a positive probability, we consider the following shrinkage estimator:

$$\hat{\theta}_n^S = \begin{cases} \mathbf{0}, & \text{if } \mathscr{L}_n < \varepsilon, \\ (\mathbf{I} - cd_n\mathscr{L}_n^{-1}\mathbf{Q}^{-1}\hat{\mathbf{\Gamma}}_n^{-1})\hat{\theta}_n, & \text{if } \mathscr{L}_n \geq \varepsilon, \end{cases} \tag{2.17}$$

where ε (>0) is an arbitrarily small number, and $\hat{\theta}_n$ and $\mathscr{L}_n$ are defined as before. Note that in a PTE case, one takes the estimator θ_n^{PT} which is equal to $\mathbf{0}$ or $\hat{\theta}_n$, according as $\mathscr{L}_n$ is $\leq$ or $>$ $\ell_{n,\alpha}$, the upper $100\alpha\%$ point of the null distribution of $\mathscr{L}_n$. Thus, the proposed shrinkage estimator adapts the James–Stein rule along with the PTE, though ε need not be equal to $\ell_{n,\alpha}$. We actually recommend ε to be small, while $\ell_{n,\alpha}$ is not so.

3. Asymptotic risk of shrinkage R-estimators. With respect to the loss (risk) function in (1.2) ((1.3)), we intend to study the risk of the shrinkage R-estimators, and to compare the same with that of the classical R-estimator $\hat{\theta}_n$ in (2.5). We shall confine ourselves to some asymptotic setup where simple and meaningful results can be derived under quite general regularity conditions. In this context, we need some moment convergence results on $\hat{\theta}_n$, which are presented first. We assume that for some positive b (not necessarily ≥ 1),

$$E_F |X_{ij}|^b < \infty, \quad \text{for} \quad j = 1, \cdots, p. \tag{3.1}$$

Further, as in Sen (1980b), we assume that for each $j(=1, \cdots, p)$, $\phi_j^{(r)}(u) = (d^r/du^r)\phi_j(u)$, $r = 0, 1, 2$, $u \in (0, 1)$ exist almost everywhere, and there exist positive constants K and $\delta(<\frac{1}{4})$, such that

$$|\phi_j^{(r)}(u)| \leq K\{u(1-u)\}^{-\delta-r}, \quad 0 < u < 1, \quad r = 0, 1, 2. \tag{3.2}$$

Finally, we assume that the derivative $f'_{[j]}$ is bounded almost everywhere and

$$\sup_x f_{[j]}(x)\{F_{[j]}(x)[1 - F_{[j]}(x)\}^{-\delta-\eta} < \infty, \quad 1 \leq j \leq p, \tag{3.3}$$

where δ is defined in (3.2) and $\eta > 0$. Then, from Theorem 2.2 and (2.49)–(2.50) of Sen (1980b), we conclude that for each $j(=1, \cdots, p)$, as $n \to \infty$,

$$n^{1/2}\{(\hat{\theta}_{nj} - \theta_j) - \xi_j^{-1}T_{nj}(\theta_j)\} = \omega_{nj} \to 0 \quad \text{almost surely (a.s.);} \tag{3.4}$$

$$E|\omega_{nj}|^k \to 0, \quad \forall k\colon k < (1 - 2\delta)/\delta \quad (>2); \tag{3.5}$$

$$nE(\hat{\theta}_n - \theta)(\hat{\theta}_n - \theta)' \to \mathbf{\Gamma} = \xi^{-1}\nu\xi^{-1}. \tag{3.6}$$

In passing, we may remark that (3.2) holds for the Wilcoxon scores ($\delta = 0$), Normal scores (δ arbitrarily close to 0) and all the other commonly used scores.

Note that by (1.2), (1.3) and (3.6), for the classical R-estimator $\hat{\theta}_n$,

$$\lim_{n\to\infty} \rho_n(\hat{\theta}, \theta) = \operatorname{Tr}(\mathbf{Q} \lim_{n\to\infty} nE(\hat{\theta}_n - \theta)(\hat{\theta}_n - \theta)') = \operatorname{Tr}(\mathbf{Q\Gamma}). \tag{3.7}$$

We intend to study the asymptotic risk of the shrinkage R-estimator $\hat{\theta}_n^S$ and compare the same with (3.7). First, we discuss briefly the case when $\theta(\neq \mathbf{0})$ is held fixed in this asymptotic setup. Note that by (2.17),

$$\begin{aligned} n(\hat{\theta}_n^S - \hat{\theta}_n)'\mathbf{Q}(\hat{\theta}_n^S - \hat{\theta}_n) \\ = I(\mathscr{L}_n < \varepsilon)n\hat{\theta}_n'\mathbf{Q}\hat{\theta}_n + I(\mathscr{L}_n \geq \varepsilon)c^2 d_n^2 \mathscr{L}_n^{-2} n\hat{\theta}_n'\hat{\mathbf{\Gamma}}_n^{-1}\mathbf{Q}^{-1}\hat{\mathbf{\Gamma}}_n^{-1}\hat{\theta}_n, \end{aligned} \tag{3.8}$$

where $I(A)$ stands for the indicator function of the set A. Since, by (2.12),

$$P\{\mathscr{L}_n < k\} \leq \min_{1\leq j\leq p} P\{|T_{nj}| \leq n^{-1/2}k^{1/2}\}, \quad \text{for every} \quad k > 0, \tag{3.9}$$

it follows from Theorems 1 and 3 of Sen (1970) that under (3.2), whenever $\theta \neq \mathbf{0}$, the right-hand side of (3.9) is $O(n^{-2})$, while, by (3.4)–(3.6), $\hat{\theta}_n'\mathbf{Q}\hat{\theta}_n$ has a bounded expectation. Thus, by some standard analysis, we conclude that the first term on the right-hand side of (3.8), for any (fixed) $\theta \neq \mathbf{0}$, converges in the first mean to 0, as $n \to \infty$. For the second term, we note that

$$nd_n^2\hat{\theta}_n'\hat{\mathbf{\Gamma}}_n^{-1}\mathbf{Q}^{-1}\hat{\mathbf{\Gamma}}_n^{-1}\hat{\theta}_n = (n\hat{\theta}_n'\mathbf{Q}\hat{\theta}_n)\{d_n^2(\hat{\theta}_n'\hat{\mathbf{\Gamma}}_n^{-1}\mathbf{Q}^{-1}\hat{\mathbf{\Gamma}}_n^{-1}\hat{\theta}_n)/(\hat{\theta}_n'Q\hat{\theta}_n)\} \leq n\hat{\theta}_n'\mathbf{Q}\hat{\theta}_n. \tag{3.10}$$

As such, using (3.4)–(3.6) and (3.10), it follows that for every fixed $\theta \neq \mathbf{0}$, the second term on the right-hand side of (3.8) converges in the first mean to 0, as $n \to \infty$. Hence, *for every fixed* $\theta \neq \mathbf{0}$, as $n \to \infty$,

$$\lim\ \sup_{n\to\infty} E\{n(\hat{\theta}_n^S - \hat{\theta}_n)'\mathbf{Q}(\hat{\theta}_n^S - \hat{\theta}_n) \mid \theta \neq \mathbf{0}\} = 0, \tag{3.11}$$

so that they are *asymptotically risk-equivalent.* The situation is, however, different for local translation alternatives, as will be discussed now.

Note that shrinkage estimators work out well only for shrinking neighbourhoods of a specified point (here $\mathbf{0}$). In the asymptotic case, this shrinking neighbourhood coincides with the usual Pitman-type local translation alternatives. For this, we conceive of a triangular array $\{\mathbf{X}_{ni},\ 1 \leq i \leq n;\ n \geq 1\}$ of row-wise i.i.d.r.v. with the d.f. $\{F_{(n)}\}$, and, by reference to (1.1), we consider an alternative hypothesis K_n that for the given n, $\theta = \theta_n = n^{-1/2}\lambda$, where λ belongs to a compact subset C (containing $\mathbf{0}$ as an inner point). Note that by virtue of the translation-invariance of $\hat{\theta}_n$ and the fact that $\mathbf{X}_{ni} - n^{-1/2}\lambda$, has, under K_n, the d.f. F in (1.1), which does not depend on n, we conclude that (3.7) holds under $\{K_n\}$ as well. The main theorem of the paper is the following.

THEOREM 3.1. *Under the assumed regularity conditions, for ε chosen adequately small, for every c: $0 < c < 2(p-2)$, the shrinkage estimator $\hat{\theta}_n^S$ dominates over the classical R-estimator $\hat{\theta}_n$, for local translation alternatives $\{K_n\}$, uniformly in λ in any compact set C (containing $\mathbf{0}$ as an inner point).*

PROOF. Note that by (2.17), we have under K_n,

$$\begin{aligned} n(\hat{\theta}_n^S - \theta_n)'\mathbf{Q}(\hat{\theta}_n^S - \theta_n) &= I(\mathscr{L}_n < \varepsilon)(\lambda'\mathbf{Q}\lambda) \\ &\quad + I(\mathscr{L}_n > \varepsilon)\{n(\hat{\theta}_n - \theta_n)'\mathbf{Q}(\hat{\theta}_n - \theta_n) - 2cd_n\mathscr{L}_n^{-1}n(\hat{\theta}_n - \theta_n)'\hat{\Gamma}_n^{-1}\hat{\theta}_n \\ &\quad + c^2 d_n^2 \mathscr{L}_n^{-2} n\hat{\theta}_n'\hat{\Gamma}_n^{-1}\mathbf{Q}^{-1}\hat{\Gamma}_n^{-1}\hat{\theta}_n\}. \end{aligned} \tag{3.12}$$

If we denote by $H_p(\cdot\,; \Delta)$ the noncentral chi-squared d.f. with p degrees of freedom (DF) and noncentrality parameter $\Delta(\geq 0)$, then, from the results in Sen and Puri (1967), we have

$$\lim_{n\to\infty} P\{\mathscr{L}_n \leq x \mid K_n\} = H_p(x;\ \lambda'\Gamma^{-1}\lambda), \quad \forall x \geq 0. \tag{3.13}$$

Therefore, the first term on the right-hand side of (3.12) converges to $(\lambda'\mathbf{Q}\lambda)H_p(\varepsilon;\ \lambda'\Gamma^{-1}\lambda)$, uniformly in λ in any compact subset C; the latter result follows from Hušková (1971). Also, proceeding as in Section 4 of Sen and Saleh (1979), with adaptations from Hušková (1971), we claim that uniformly in λ in C,

$$\begin{aligned} \lim_{n\to\infty} E\{I(\mathscr{L}_n \geq \varepsilon) n(\hat{\theta}_n - \theta_n)'\mathbf{Q}(\hat{\theta}_n - \theta_n) \mid K_n\} \\ = \{1 - H_{p+2}(\varepsilon;\ \Delta)\}\mathrm{Tr}(\mathbf{Q}\Gamma) \\ - (\lambda'\mathbf{Q}\lambda)\{H_p(\varepsilon;\ \Delta) - 2H_{p+2}(\varepsilon;\ \Delta) + H_{p+4}(\varepsilon;\ \Delta)\}, \end{aligned} \tag{3.14}$$

where $\Delta = \lambda'\Gamma^{-1}\lambda$. Further,

$$
\begin{aligned}
&|\, n(\hat{\theta}_n - \theta_n)'\hat{\Gamma}_n^{-1}\hat{\theta}_n \cdot d_n \,| \\
&\quad = |\, n^{1/2}(\hat{\theta}_n - \theta_n)'\hat{\Gamma}_n^{-1/2} \cdot \hat{\Gamma}_n^{-1/2} n^{1/2}\hat{\theta}_n \cdot d_n \,| \\
&\quad \le \{[n(\hat{\theta}_n - \theta_n)'\hat{\Gamma}_n^{-1}(\hat{\theta}_n - \theta_n) \cdot d_n][d_n n\hat{\theta}_n'\hat{\Gamma}_n^{-1}\hat{\theta}_n]\}^{1/2} \\
&\quad \le \{n(\hat{\theta}_n - \theta_n)'\mathbf{Q}(\hat{\theta}_n - \theta_n) \cdot n\hat{\theta}_n'\mathbf{Q}\hat{\theta}_n\}^{1/2} \\
&\quad \le \tfrac{1}{2}\{n(\hat{\theta}_n - \theta_n)'\mathbf{Q}(\hat{\theta}_n - \theta_n) + n\hat{\theta}_n'\mathbf{Q}\hat{\theta}_n\}.
\end{aligned}
\tag{3.15}
$$

Notice that because of the translation invariance of $\hat{\theta}_n$, (3.4)–(3.6) hold under $\{K_n\}$ as well (with θ being replaced by θ_n), so that, for every $k(>0)$, under K_n, with n adequately large, $(n\hat{\theta}_n'\mathbf{Q}\hat{\theta}_n)^k$ and $(n(\hat{\theta}_n - \theta_n)'\mathbf{Q}(\hat{\theta}_n - \theta_n))^k$ are integrable (uniformly in λ). Hence, by (3.4)–(3.6), we obtain that under $\{K_n\}$, uniformly in λ belonging to a compact C, as $n \to \infty$,

$$
\mathscr{L}_n = n\hat{\theta}_n'\Gamma^{-1}\hat{\theta}_n + \eta_n; \quad \eta_n \to 0, \quad \text{in probability (or a.s.).} \tag{3.16}
$$

Therefore, on the set $\{\mathscr{L}_n > \varepsilon\}$, $\varepsilon > 0$, under $\{K_n\}$, uniformly in $\lambda \in C$,

$$
\mathscr{L}_n^{-1} - (n\hat{\theta}_n'\Gamma^{-1}\hat{\theta}_n)^{-1} \to_p 0, \quad \text{as} \quad n \to \infty. \tag{3.17}
$$

Finally, under $\{K_n\}$, by (3.4)–(3.6), uniformly in $\lambda \in C$,

$$
n^{1/2}\Gamma^{-1/2}\hat{\theta}_n \to_{\mathscr{D}} \mathscr{N}(\Gamma^{-1/2}\lambda, \mathbf{I}_p), \quad \text{as} \quad n \to \infty; \tag{3.18}
$$

the uniformity result (on the $\mathbf{T}_n$) due to Hušková (1971) is utilized here too.

Let us now denote by $\mathbf{W}$ a p-vector having the multinormal distribution with mean vector $\omega = \Gamma^{-1/2}\lambda$ and dispersion matrix $\mathbf{I}_p$. Also, let $\chi^2_{q,\delta}$ be a r.v. having the noncentral chi-squared d.f. $H_q(\cdot\,; \delta)$. Therefore, using (3.10), (3.14)–(3.18) along with the uniform integrability results (considered above), we conclude that for every $\varepsilon > 0$, uniformly in $\lambda \in C$,

$$
\begin{aligned}
&\lim_{n\to\infty} E\{I(\mathscr{L}_n \ge \varepsilon) d_n \mathscr{L}_n^{-1} n(\hat{\theta}_n - \theta_n)'\hat{\Gamma}_n^{-1}\hat{\theta}_n \mid K_n\} \\
&\qquad = ch_p(\mathbf{Q}\Gamma)\{E(\chi^{-2}_{p+2,\Delta}) - E[I(\mathbf{W}'\mathbf{W} < \varepsilon)(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\omega]\},
\end{aligned}
\tag{3.19}
$$

$$
\begin{aligned}
&\lim_{n\to\infty} E\{I(\mathscr{L}_n \ge \varepsilon) n d_n^2 \mathscr{L}_n^{-2}\hat{\theta}_n'\hat{\Gamma}_n^{-1}\mathbf{Q}^{-1}\hat{\Gamma}_n^{-1}\hat{\theta}_n \mid K_n\} \\
&\qquad = \{ch_p(\mathbf{Q}\Gamma)\}^2 \cdot \{\mathrm{Tr}(\mathbf{Q}^{-1}\Gamma^{-1})E(\chi^{-4}_{p+2,\Delta}) + \Delta^* E(\chi^{-4}_{p+4,\Delta}) \\
&\qquad\qquad - E[I(\mathbf{W}'\mathbf{W} < \varepsilon)(\mathbf{W}'\mathbf{W})^{-2}\mathbf{W}'\mathbf{A}\mathbf{W}]\},
\end{aligned}
\tag{3.20}
$$

where $ch_p(\mathbf{A})$ stands for the smallest characteristic root of a $p \times p$ matrix $\mathbf{A}$ and where $\Delta = \lambda'\Gamma^{-1}\lambda$, $\Delta^* = \lambda'\Gamma^{-1}\mathbf{Q}^{-1}\Gamma^{-1}$ and $\mathbf{A} = \Gamma^{-1/2}\mathbf{Q}^{-1}\Gamma^{-1/2}$. Now

$$
\begin{aligned}
&|\, E\{I(\mathbf{W}'\mathbf{W} < \varepsilon)(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\omega\} \,| \\
&\qquad \le E\{I(\mathbf{W}'\mathbf{W} < \varepsilon)(\mathbf{W}'\mathbf{W})^{-1/2}(\omega'\omega)^{1/2}\} = \Delta^{1/2}\int_0^{\varepsilon} x^{-1/2}\, dH_p(x; \Delta) \\
&\qquad \le (\Gamma(p/2)2^{p/2})^{-1}\Delta^{1/2}\varepsilon^{(p-1)/2}e^{-(1/2)\Delta(1-\varepsilon)}, \quad p \ge 2, \quad 0 < \varepsilon < 1.
\end{aligned}
\tag{3.21}
$$

Also, for every $p \geq 2$ and $0 < \varepsilon < 1$,

$$
\begin{aligned}
E\{I(\mathbf{W}'\mathbf{W} < \varepsilon)(\mathbf{W}'\mathbf{W})^{-2}\mathbf{W}'\mathbf{A}\mathbf{W}\} &\leq E\{I(\mathbf{W}'\mathbf{W} < \varepsilon)(\mathbf{W}'\mathbf{W})^{-1}\mathrm{Tr}(\mathbf{A})\} \\
&= \mathrm{Tr}(\mathbf{A}) \cdot E\{I(\mathbf{W}'\mathbf{W} < \varepsilon)(\mathbf{W}'\mathbf{W})^{-1}\} \\
&\leq \mathrm{Tr}(\mathbf{A})(\Gamma(p/2)2^{p/2})^{-1}\varepsilon^{(p-2)}e^{-(1/2)\Delta(1-\varepsilon)}.
\end{aligned} \tag{3.22}
$$

Therefore, whenever $p \geq 3$, from (3.12) through (3.22), we obtain that

$$
\begin{aligned}
&\rho^*(\hat{\theta}^S, \lambda) \\
&\quad = \lim_{n\to\infty} E\{n(\hat{\theta}_n^S - \theta_n)'\mathbf{Q}(\hat{\theta}_n^S - \theta_n) \mid K_n\} \\
&\quad = \mathrm{Tr}(\mathbf{Q\Gamma})\{1 - H_{p+2}(\varepsilon; \Delta)\} + (\lambda'\mathbf{Q}\lambda)\{2H_{p+2}(\varepsilon; \Delta) - H_{p+4}(\varepsilon; \Delta)\} \\
&\quad\quad - 2c(ch_p(\mathbf{Q\Gamma}))E(\chi_{p+2,\Delta}^{-2}) + c^2(ch_p(\mathbf{Q\Gamma}))^2 \\
&\quad\quad \cdot \{\mathrm{Tr}(\mathbf{Q}^{-1}\mathbf{\Gamma}^{-1})E(\chi_{p+2,\Delta}^{-4}) + \Delta^* E(\chi_{p+4,\Delta}^{-4})\} \\
&\quad\quad + 2c \cdot ch_p(\mathbf{Q\Gamma})E\{I(\mathbf{W}'\mathbf{W} < \varepsilon)(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\omega\} \\
&\quad\quad - c^2(ch_p(\mathbf{Q\Gamma}))^2 E\{I(\mathbf{W}'\mathbf{W} < \varepsilon)(\mathbf{W}'\mathbf{W})^{-2}\mathbf{W}'\mathbf{A}\mathbf{W}\},
\end{aligned} \tag{3.23}
$$

uniformly in $\lambda \in C$. Note that the last term is nonnegative, so that using (3.21), we obtain after some rearrangements of terms that (uniformly in $\lambda \in C$),

$$
\begin{aligned}
\rho^*(\hat{\theta}^S, \lambda) \leq \{&\mathrm{Tr}(\mathbf{Q\Gamma}) - 2c \cdot ch_p(\mathbf{Q\Gamma})E(\chi_{p+2,\Delta}^{-2}) \\
&+ c^2(ch_p(\mathbf{Q\Gamma}))^2[\Delta^* E(\chi_{p+4,\Delta}^{-4}) + \mathrm{Tr}(\mathbf{Q}^{-1}\mathbf{\Gamma}^{-1})E(\chi_{p+2,\Delta}^{-4})]\} \\
&+ O(\varepsilon^{(p-1)/2}) + O(\varepsilon^{p/2}).
\end{aligned} \tag{3.24}
$$

Now, by the results of Section 2 of Sclove, Morris and Radhakrishnan (1972), the leading term on the right-hand side of (3.24) is $< \mathrm{Tr}(\mathbf{Q\Gamma}) = \rho^*(\hat{\theta}, \lambda)$, for every c: $0 < c < 2(p-2)$, and uniformly in λ in any compact set C. This completes the proof of the theorem.

Note that in the definition of the shrinkage estimator in (2.17), one has to decide on the choice of c and ε. Though c may belong to the interval $(0, 2(p-2))$, the choice of $c = (p-2)$ has been found to be better in the normal theory case and may also be recommended here. Further, from the results of Berger et al. (1977), we are tempted to make c dependent on n (i.e., $c = c_n$), where c_n is nondecreasing in n and $\lim_{n\to\infty} c_n = c$ exists (and belongs to the interval $(0, 2(p-2))$). The existence of the limit c of c_n ensures the validity of the formulae in (3.23)–(3.24), and hence, the conclusion of the theorem remains true as well. Typically, for small values of n, c_n should be taken small, while, for larger values of n, it may be taken closer to $(p-2)$. Also, we recommend the use of a small value of ε (e.g., $\varepsilon = 0.05$), though the choice of ε may depend on the value of p at hand. Looking at (3.22) and (3.24) we observe that the higher is the value of p, the greater is the range of the admissible values of ε (corresponding to a given margin of the residual term): Basically, $\varepsilon^{(p-1)/2}$ should be chosen small. Finally, the dominance result in Theorem 3.1 is of an asymptotic nature. The question may arise whether the size of ε and n needed to ensure the dominance of $\hat{\theta}_n^S$

depends on the underlying d.f. F. The answer to this query depends on whether (3.17) and (3.18) hold uniformly in a class of F. The answer is in the affirmative, and such uniformity results can be established through the uniform linearity results on rank statistics, as have been studied in detail in Jurečková (1983, 1985). Since we are dealing here with the multivariate case, we need here the additional condition that for F belonging to the given class, the convergence of $\mathbf{M}_n^{*-1}$ to $\mathbf{\Gamma}^{-1}$ is also uniform, and for this, apart from the assumed regularity conditions in Section 2, it suffices to assume that $ch_p(\mathbf{\Gamma})$ is bounded away from 0, uniformly in the class of d.f. Essentially, (nearly) singular multivariate distributions are excluded from this class, and otherwise, the distributions are absolutely continuous with bounded and continuous density functions having finite Fisher informations.

We conclude this section with a remark on the choice of the particular estimator in (2.17). It is clear from (3.17) that in (2.17) one may also replace $\mathscr{L}_n$ by $(n\hat{\boldsymbol{\theta}}_n'\hat{\mathbf{\Gamma}}_n^{-1}\hat{\boldsymbol{\theta}}_n)$ and the asymptotic results would continue to hold. In the normal theory case, these two forms are the same, while in the nonparametric case, they are only asymptotically the same. As has been stressed after (2.17), the use of $\mathscr{L}_n$ makes clear the relationship between the PTE and the shrinkage estimator. Further, in the nonparametric case, though $\mathscr{L}_n$ can be justified on the ground of permutational distribution-freeness, $(n\hat{\boldsymbol{\theta}}_n'\hat{\mathbf{\Gamma}}_n^{-1}\hat{\boldsymbol{\theta}}_n)$ is only asymptotically distribution-free (cf. Sen and Puri, 1967). Actually, the rank estimates are derived from the associated signed rank statistics, and hence, it is more natural to use these signed rank statistics in the construction of the test statistic. On these grounds, we prefer to prescribe the use of $\mathscr{L}_n$ in (2.17).

4. Shrinkage U-statistics. The sample mean $\bar{\mathbf{X}}_n$ is a particular case of U-statistics. It may be of some interest to construct shrinkage estimators of general parameters based on U-statistics and their (jackknifed) dispersion matrices. In this context, we may not need the diagonal symmetry of the d.f. F, but other moment conditions, not needed with the R-estimators, may be needed here.

Let $\mathscr{F}$ be a space of all d.f.'s belonging to a class, and for every $F \in \mathscr{F}$, consider a vector $\boldsymbol{\theta} = \boldsymbol{\theta}(F) = (\theta_1(F), \cdots, \theta_p(F))'$ of estimable parameters, where

$$(4.1) \qquad \theta_j(F) = E_F\{\phi_j(X_1, \cdots, X_{m_j})\}, \quad F \in \mathscr{F}, \quad j = 1, \cdots, p(\geq 1);$$

$\phi_j(x_1, \cdots, x_{m_j})$ is a *kernel* of *degree* $m_j(\geq 1)$, symmetric in its m_j arguments, for $j = 1, \cdots, p$. For $n \geq m^* = \max\{m_1, \cdots, m_p\}$, we may then define the vector of U-statistics $\mathbf{U}_n = (U_{n1}, \cdots, U_{np})'$, by letting

$$(4.2) \qquad U_{nj} = \binom{n}{m_j}^{-1} \sum_{1\leq i_1<\cdots<i_m\leq n} \phi_j(X_{i_1}, \cdots, X_{i_{m_j}}), \quad j = 1, \cdots, p.$$

$\mathbf{U}_n$ is a symmetric, unbiased and optimal estimator of $\boldsymbol{\theta}$. We assume that the kernels ϕ_j are all square integrable, and define

$$(4.3) \quad \phi_{j,c}(x_1, \cdots, x_c) = E\phi_j(x_1, \cdots, x_c, X_{c+1}, \cdots, X_{m_j}), \quad c = 0, 1, \cdots, m_j;$$

$$(4.4) \qquad \zeta_{j\ell,c} = E_F\{\phi_{j,c}(X_1, \cdots, X_c)\phi_{\ell,c}(X_1, \cdots, X_c)\} - \theta_j(F)\theta_\ell(F),$$

for $c = 0, \cdots, \min(m_j, m_\ell)$ and $j, \ell = 1, \cdots, p$. Then (cf. Hoeffding, 1948)

$$nE\{(\mathbf{U}_n - \theta)(\mathbf{U}_n - \theta)'\} \tag{4.5}$$
$$= n\left(\left(\binom{n}{m_j}^{-1} \sum_{c=1}^{m_j} \binom{m_\ell}{c}\binom{n-m_\ell}{m_j - c}\zeta_{j\ell,c}\right)\right) = \mathbf{\Gamma} + \mathbf{O}(n^{-1}),$$

where

$$\mathbf{\Gamma} = ((\gamma_{j\ell})) = ((m_j m_\ell \zeta_{j\ell,1})). \tag{4.6}$$

Consider then the loss and risk function as in (1.2) and (1.3). We intend to construct some shrinkage estimators which dominates over $\mathbf{U}_n$, at least asymptotically. For U-statistics, one may use the jackknifing to obtain a convenient estimator of $\mathbf{\Gamma}$ (viz., Sen, 1960, 1981). We write $\mathbf{U}_n = \mathbf{U}(X_1, \cdots, X_n)$, and, for every i: $1 \le i \le n$, let

$$\mathbf{U}_{n-1}^{(i)} = \mathbf{U}(X_1, \cdots, X_{i-1}, X_{i+1}, \cdots, X_n), \quad \mathbf{U}_{n,i} = n\mathbf{U}_n - (n-1)\mathbf{U}_{n-1}^{(i)}. \tag{4.7}$$

Then, the jackknifed estimator of $\mathbf{\Gamma}$ is

$$\hat{\mathbf{\Gamma}}_n = (n-1)^{-1} \sum_{i=1}^{n} (\mathbf{U}_{n,i} - \mathbf{U}_n)(\mathbf{U}_{n,i} - \mathbf{U}_n)'. \tag{4.8}$$

If we have reason to believe that θ lies in a small neighbourhood of some specified θ_0 (which, without any loss of generality, we may take as $\mathbf{0}$), then a test statistic for this problem is $\mathscr{L}_n = n\mathbf{U}_n'\hat{\mathbf{\Gamma}}_n^{-1}\mathbf{U}_n$, and, as in (2.17), we may consider the shrinkage estimator:

$$\mathbf{U}_n^S = \begin{cases} \mathbf{0}, & \text{if } \mathscr{L}_n < \varepsilon; \\ (\mathbf{I} - cd_n\mathscr{L}_n^{-1}\mathbf{Q}^{-1}\hat{\mathbf{\Gamma}}_n^{-1})\mathbf{U}_n, & \text{if } \mathscr{L}_n \ge \varepsilon, \end{cases} \tag{4.9}$$

where c and ε are positive numbers, defined as in (2.17) and d_n is the smallest characteristic root of $\mathbf{Q}\hat{\mathbf{\Gamma}}_n$. In particular, if the $\mathbf{X}_i$ are p-vectors, defined as in Section 1, and if $m_1 = \cdots = m_p = 1$, $\phi_j(\mathbf{X})$ is the jth component of $\mathbf{X}$, $j = 1, \cdots, p$, then $\mathbf{U}_n = \bar{\mathbf{X}}_n$, so that (4.9) is a natural extension of the shrinkage estimator considered in Berger et al. (1977). For higher order moments (and product moments) of F, similar U-statistics can be constructed, so that (4.9) provides shrinkage versions for these estimates. There is an abundance of use of U-statistics in nonparametric estimation problems, and (4.9) would provide usable shrinkage estimators in these situations as well.

As in Section 3, for any fixed $\theta(\neq\mathbf{0})$, asymptotically, $\mathbf{U}_n^S$ and $\mathbf{U}_n$ are risk-equivalent. Also, for local alternatives: K_n: $\theta = \theta_{(n)} = n^{-1/2}\boldsymbol{\lambda}$, $\boldsymbol{\lambda}$ fixed, the conclusion of Theorem 3.1 applies to (4.9) as well. Though the derivations of these results follow an entirely different track, since the conclusions are similar, we omit these details. However, we may refer to Sen (1984) for some of these mathematical treatments.

Acknowledgements. Thanks are due to the Associate Editor and the referees for their valuable comments on the manuscript.

REFERENCES

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.

JAMES, W. and STEIN, C. (1961). Estimation with quadratic loss. *Proc. 4th Berkeley Symp. Math. Statist. Probab.* **1** 361–379.

HOEFFDING, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statist.* **19** 293–325.

HUŠKOVÁ, M. (1971). Asymptotic distribution of rank statistics used for multivariate testing symmetry. *J. Multivariate Anal.* **1** 461–484.

JUREČKOVÁ, J. (1983). Robust estimators of location and regression parameters and their second order asymptotic relations. *Trans. 9th Prague Confer. Infor. Th. Statist. Dec. Func. Rand. Proc.* 19–32. Reidel, Dordrecht.

JUREČKOVÁ, J. (1985). Robust estimators of location and their second order asymptotic relations. *Internat. Statist. Rev.* (Centennial Vol), in press.

PURI, M. L. and SEN, P. K. (1971). *Nonparametric Methods in Multivariate Analysis.* Wiley, New York.

SALEH, A.K.MD. E. and SEN, P. K. (1978). Nonparametric estimation of location parameter after a preliminary test on regression. *Ann. Statist.* **6** 154–168.

SCLOVE, S. L., MORRIS, C. and RADHAKRISHNAN, R. (1972). Non-optimality of preliminary test estimators for the mean of a multivariate normal distribution. *Ann. Math. Statist.* **43** 1481–1490.

SEN, P. K. (1960). On some convergence properties of *U*-statistics. *Calcutta Statist. Asso. Bull.* **10** 1–18.

SEN, P. K. (1970). On some convergence properties of one-sample rank order statistics. *Ann. Math. Statist.* **41** 2140–2143.

SEN, P. K. (1980a). On almost sure linearity theorems for signed rank order statistics. *Ann. Statist.* **8** 313–321.

SEN, P. K. (1980b). On nonparametric sequential point estimation of location based on general rank order statistics. *Sankhya, Ser. A* **42** 201–218.

SEN, P. K. (1981). *Sequential Nonparametrics: Invariance Principles and Statistical Inference.* Wiley, New York.

SEN, P. K. (1984). A James–Stein type detour of *U*-statistics. *Commun. Statist. Theor. Meth.* **A13**, in press.

SEN, P. K. and PURI, M. L. (1967). On the theory of rank order tests for location in the multivariate one-sample problem. *Ann. Math. Statist.* **38** 1216–1228.

SEN, P. K. and PURI, M. L. (1969). On robust nonparametric estimation in some multivariate linear models. In *Multivariate Analysis*, II 33–52. (Ed: P. R. Krishnaiah), Academic, New York.

SEN, P. K. and SALEH, A.K.MD. E. (1979). Nonparametric estimation of location parameters after a preliminary test on regression in the multivariate case. *J. Multivariate Anal.* **9** 322–331.

STEIN, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. *Proc. 3rd Berkeley Symp. Math. Statist. Probab.* **1** 197–206.

DEPARTMENT OF BIOSTATISTICS 201H
UNIVERSITY OF NORTH CAROLINA
CHAPEL HILL, NORTH CAROLINA 27514

DEPARTMENT OF MATHEMATICS
AND STATISTICS
CARLETON UNIVERSITY
OTTAWA, ONTARIO, CANADA K1S 5B6