The Annals of Statistics
1989, Vol. 17, No. 3, 1375–1386

THE STEIN PARADOX IN THE SENSE OF THE PITMAN MEASURE OF CLOSENESS

BY PRANAB KUMAR SEN,[1] TATSUYA KUBOKAWA AND A. K. MD. EHSANES SALEH[2]

University of North Carolina at Chapel Hill, University of Tsukuba and Carleton University

The dominance and related optimality properties of the usual Stein-rule or shrinkage estimators are typically developed for quadratic error loss functions. It is shown that under the classical Pitman closeness criterion the Stein-rule estimators possess a similar dominance property when the "closeness" measure is based on suitable quadratic norms.

1. Introduction. Suppose that for some positive integer p, $\mathbf{X}$ has a p-variate normal distribution with mean vector $\boldsymbol{\theta}$ and dispersion matrix $\sigma^2\mathbf{V}$ where $\mathbf{V}$ is known and positive definite (p.d.) while $\boldsymbol{\theta}$ and σ^2 are both unknown. We also assume that there exists another statistic S distributed independently of $\mathbf{X}$ as $m^{-1}\sigma^2\chi_m^2$, where χ_m^2 stands for a random variable (r.v.) having the central chi square distribution with m degrees of freedom (DF) and m is a positive integer. Based on $(\mathbf{X}, S)$, the problem is to estimate $\boldsymbol{\theta}$ in an optimal manner. Usually, a quadratic loss function is incorporated in the formulation of an optimality criterion, and $\mathbf{X}$ is optimal (when $p = 1$ or 2).

Stein (1956) showed that for $p \geq 3$, $\mathbf{X}$ is inadmissible under a quadratic loss, and James and Stein (1961) constructed a shrinkage version which dominates $\mathbf{X}$ in quadratic error loss. The past 25 years have witnessed a phenomenal growth in the literature on this Stein-rule estimation theory in its diverse tributaries; we may refer to Arnold (1981), Anderson (1984) and Berger (1985) for some systematic accounts of the related developments.

Pitman (1937) laid down the foundation of an important concept of nearness or closeness of an estimator. In the current context, for two estimators $\boldsymbol{\delta}_1$ and $\boldsymbol{\delta}_2$ of $\boldsymbol{\theta}$, and for a given p.d. matrix $\mathbf{Q}$, defining the norm $\|\mathbf{x} - \mathbf{y}\|_{\mathbf{Q}}^2$ as $(\mathbf{x} - \mathbf{y})'\mathbf{Q}(\mathbf{x} - \mathbf{y})$, we say that $\boldsymbol{\delta}_1$ is "closer" to $\boldsymbol{\theta}$ than $\boldsymbol{\delta}_2$ (in the norm $\|\cdot\|_{\mathbf{Q}}$) in the Pitman sense if

$$P_{\omega}\{\|\boldsymbol{\delta}_1 - \boldsymbol{\theta}\|_{\mathbf{Q}} \leq \|\boldsymbol{\delta}_2 - \boldsymbol{\theta}\|_{\mathbf{Q}}\} \geq \tfrac{1}{2} \quad \text{for all } \boldsymbol{\omega} = (\boldsymbol{\theta}, \sigma^2). \tag{1.1}$$

Although the relationship of this Pitman closeness criterion with other conventional measures of efficiency has been explored by a number of workers [viz., Sen (1986a) and Peddada and Khattree (1986)], the impact of the Pitman measure of

Received March 1987; revised November 1988.
[1]Work supported by the Office of Naval Research, Contract N00014-83-K-0387.
[2]Work supported by NSERC Canada, Grant A-3088.
AMS 1980 *subject classifications*. Primary 62C15; secondary 62H12.
Key words and phrases. Inadmissibility, loss functions, Pitman closeness, quadratic norm, shrinkage estimators, Stein phenomenon.

closeness in the Stein-rule estimation theory has not yet been fully assessed. However, there are some striking points worth pondering:

(i) In the light of (1.1), $\mathbf{X}$ may not be the closest estimator of $\boldsymbol{\theta}$ even for $p = 1$ or 2. For the univariate normal mean (variance known), Efron (1975) has constructed such an estimator (based on a sample of size n):

$$(1.2)\quad \delta_1 = \overline{X} - \Delta(\overline{X}) \quad \text{where } \Delta(x) = \left(\tfrac{1}{2}\right)\min\left(x, n^{1/2}\Phi(-xn^{1/2})\right),\ x \geq 0,$$

$\Delta(-x) = -\Delta(x)$ and $\Phi(x)$ is the standard normal distribution function (d.f.). Since $p = 1$, in (1.1), we may eliminate $\mathbf{Q}$ altogether, and the dominance of δ_1 over the classical estimator $\overline{X}$ in the sense of (1.1) holds for all $\mathbf{Q}$. Efron (1975) has also discussed the case of $p \geq 2$; however, no dominance result has been established in an analytical sense. Moreover, though δ_1 in (1.2) is a shrinkage estimator (in a sense), structurally it is quite different from the usual Stein-rule estimators (as we shall see later on). For the proper Stein-rule estimators, in the multivariate case, this dominance (in the Pitman closeness sense) deserves an analytical treatment.

(ii) Rao (1981) has considered some simple shrinkage estimators and showed that they need not be the Pitman closest ones. This led him to the query: Is a Stein-rule estimator closer than $\mathbf{X}$ in the Pitman sense [i.e., (1.1)]? Rao has argued that the quadratic error loss function places undue emphasis on large deviations which may occur with small probability, and minimizing the mean square error may insure against large errors in estimation occurring more frequently rather than providing greater concentration of an estimator in neighborhoods of the true value. This criticism is more appropriate for the Stein-rule estimators which generally do not have (multi-) normal distributions (even asymptotically). Recently, Rao, Keating and Mason (1986) and Keating and Mason (1988) have shown by extensive numerical studies that the James–Stein estimator is closer than $\mathbf{X}$, for $p \geq 2$.

(iii) Hwang (1985) has introduced another measure, stochastic dominance, which is based solely on the two marginal d.f.'s of $\|\boldsymbol{\delta}_1 - \boldsymbol{\theta}\|_{\mathbf{Q}}$ and $\|\boldsymbol{\delta}_2 - \boldsymbol{\theta}\|_{\mathbf{Q}}$. Neither of the Pitman closeness dominance and stochastic dominance implies the other. The situation is more complicated for the Pitman closeness dominance as in (1.1); one needs to consider the joint distribution of the two norms $\|\boldsymbol{\delta}_1 - \boldsymbol{\theta}\|_{\mathbf{Q}}$ and $\|\boldsymbol{\delta}_2 - \boldsymbol{\theta}\|_{\mathbf{Q}}$ (whereas the dominance of their marginal d.f.'s leads to the stochastic dominance measure).

In the current study, we consider shrinkage estimators of the form

$$(1.3)\qquad \boldsymbol{\delta}_\phi = \mathbf{X} - \phi(\mathbf{X}, S)S\|\mathbf{X}\|^{-2}_{\mathbf{Q},\mathbf{V}}\mathbf{Q}^{-1}\mathbf{V}^{-1}\mathbf{X},$$

where ϕ is a nonnegative function bounded from above by $(p-1)(3p+1)/(2p)$ and $\|\mathbf{X}\|^2_{\mathbf{Q},\mathbf{V}} = \mathbf{X}'\mathbf{V}^{-1}\mathbf{Q}^{-1}\mathbf{V}^{-1}\mathbf{X}$. Estimators of this type with a different bound for ϕ were considered by Stein (1981), for $p \geq 3$, and hence, we shall term these as Stein-rule estimators. It is shown here analytically that for $p \geq 2$, $\boldsymbol{\delta}_\phi$ is closer than $\mathbf{X}$ in the Pitman sense in (1.1). This demonstrates that the numerical results obtained earlier by Rao, Keating and Mason (1986) and Keating and Mason (1988) were in the right direction. It is also shown that this dominance in the Pitman closeness sense holds when the dispersion matrix of $\mathbf{X}$ is arbitrary (p.d.) and unknown. In the case of a known covariance matrix, additional dominance results are considered too. The asymptotic case is discussed briefly in the last section.

2. The main results. Toward the inadmissibility of the classical estimator $\mathbf{X}$ under the Pitman measure of closeness, we have the following.

THEOREM 1. *Assume that* $p \geq 2$ *and*

$$0 \leq \phi(X, s) \leq (p-1)(3p+1)/(2p) \quad \textit{for every } (\mathbf{X}, S) \textit{ a.e.} \tag{2.1}$$

Then $\boldsymbol{\delta}_\phi$, *given by* (1.3), *is closer than* $\mathbf{X}$ *in the Pitman sense in* (1.1).

REMARK 1. If σ^2 is known, then in (1.3), we may replace S by σ^2 and Theorem 1 remains true. If further, $\sigma^2 = 1$ and $\mathbf{V} = \mathbf{Q} = \mathbf{I}$, then (1.3) includes as special cases the following:

$$\boldsymbol{\delta}_{JS} = [1 - (p-2)/\mathbf{X}'\mathbf{X}]\mathbf{X} \quad \text{[James and Stein (1961)]}, \tag{2.2}$$

$$\boldsymbol{\delta}_{KM} = [1 - (p-1)/\mathbf{X}'\mathbf{X}]\mathbf{X} \quad \text{[Keating and Mason (1988)]}. \tag{2.3}$$

In (2.2), p needs to be greater than 2, while in (2.3) also, Keating and Mason (1988) considered the case of $p \geq 3$. This shows that our analytical results are applicable to the special cases studied numerically by Rao, Keating and Mason (1986) and Keating and Mason (1988). The case of $p = 1$ is left out in (2.1). However, the Efron (1975) estimator in (1.2) exhibits the existence of some shrinkage estimator, even in the univariate case, which dominates the classical estimator X. On the other hand, structurally (1.2) and (1.3) are not that similar. Note that for $p = 1$ and S replaced by σ^2 $(= 1)$, (1.3) may be taken as $[1 - \phi(X, 1)/X^2]X$, while (1.2) may be taken as $[1 - \Delta(X)/X]X$. Thus, here allowing $\phi(X, 1)$ to depend on $\Delta(X)$ by letting $\phi(x, 1) = x\Delta(x)$, we see that (1.2) and (1.3) are related to each other. However, in this setup, $x\Delta(x)$ may not have the usual "pretest" interpretation of the shrinkage factor of the James–Stein (1961) estimator. Moreover, in the multivariate case, such a correspondence between (1.2) and (1.3) may be more difficult to establish.

REMARK 2. It is well known that the usual Stein-rule estimator can be improved on by the positive-rule version. The same dominance holds in the Pitman closeness sense too. For example, we may consider the estimators

$$\boldsymbol{\delta} = \mathbf{X} - aS\|\mathbf{X}\|_{\mathbf{Q},\mathbf{V}}^{-2}\mathbf{Q}^{-1}\mathbf{V}^{-1}\mathbf{X}, \tag{2.4}$$

$$\boldsymbol{\delta}^{+} = \mathbf{X} - \min\{aS\|\mathbf{X}\|_{\mathbf{Q},\mathbf{V}}^{-2}, \mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\|\mathbf{X}\|_{\mathbf{Q},\mathbf{V}}^{-2}\}\mathbf{Q}^{-1}\mathbf{V}^{-1}\mathbf{X}. \tag{2.5}$$

On the set $\{aS > \mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\}$, we see that the inequality $\|\boldsymbol{\delta}^{+} - \boldsymbol{\theta}\|_{\mathbf{Q}} \leq \|\boldsymbol{\delta} - \boldsymbol{\theta}\|_{\mathbf{Q}}$ [equivalent to $aS + \mathbf{X}'\mathbf{V}^{-1}\mathbf{X} \geq 2(\mathbf{X} - \boldsymbol{\theta})'\mathbf{V}^{-1}\mathbf{X}$] holds if $\boldsymbol{\theta}'\mathbf{V}^{-1}\mathbf{X} \geq 0$. Let then $\mathbf{H}$ be an orthogonal matrix, such that $\mathbf{H}\mathbf{V}^{-1/2}\boldsymbol{\theta} = (\eta, 0, \dots, 0)'$ and $\eta =$

$(\boldsymbol{\theta}'\mathbf{V}^{-1}\boldsymbol{\theta})^{1/2}$ and let $\mathbf{Y} = \mathbf{H}\mathbf{V}^{-1/2}\mathbf{X}$. Then, we have that

$$\begin{aligned} &P_\omega\{\|\boldsymbol{\delta}^+ - \boldsymbol{\theta}\|_{\mathbf{Q}}^2 \le \|\boldsymbol{\delta} - \theta\|_{\mathbf{Q}}^2\} \\ &\quad \ge P_\omega\{aS \le \mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\} + P_\omega\{\{as > \mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\} \cap \{\boldsymbol{\theta}'\mathbf{V}^{-1}\mathbf{X} \ge 0\}\} \\ &\quad = P_\omega\{aS \le \mathbf{Y}'\mathbf{Y}\} + P_\omega\{\{aS > \mathbf{Y}'\mathbf{Y}\} \cap \{Y_1 \ge 0\}\}. \end{aligned} \tag{2.6}$$

Next, we note that for any nonnegative $f(\cdot)$ with respect to the normal density $p(x) = \sigma^{-1}(2\pi)^{-1/2}\exp\{-x^2/2\sigma^2\}$ and any nonnegative η,

$$\begin{aligned} E\left[f(Y_1^2)I(Y_1 \le 0)\right] &= \int f(y^2)I(y < 0)p(y - \eta)\,dy \\ &= \int f(y^2)I(y \ge 0)p(y + \eta)\,dy \\ &\le \int f(y^2)I(y \ge 0)p(y - \eta)\,dy = E\left[f(Y_1^2)I(Y_1 \ge 0)\right], \end{aligned} \tag{2.7}$$

and this implies that $E[f(Y_1^2)I(Y_1 \ge 0)] \ge \frac{1}{2}E[f(Y_1^2)]$, so that the right-hand side of (2.6) is $\ge P_\omega\{aS \le \mathbf{Y}'\mathbf{Y}\} + \frac{1}{2}P_\omega\{aS > \mathbf{Y}'\mathbf{Y}\} \ge \frac{1}{2}$. This establishes the dominance of $\boldsymbol{\delta}^+$ in (2.5) over $\boldsymbol{\delta}$ in (2.4) in the light of (1.1). In particular, if we let $\mathbf{V} = \mathbf{Q} = \mathbf{I}$ and $a = p - 2$ or $p - 1$, then (2.5) corresponds to the positive-rule version of (2.2) or (2.3), and hence, the aforesaid dominance result applies to these particular cases as well.

REMARK 3. Sclove, Morris and Radhakrishnan (1972) considered shrinking $\mathbf{X}$ toward a linear subspace $\Lambda \subseteq R^p$ of dimension r $(\le p)$. Let $\mathbf{PX}$ denote the projection of $\mathbf{X}$ onto Λ, defined by $\|\mathbf{X} - \mathbf{PX}\|_{\mathbf{Q}} = \inf_{\boldsymbol{\lambda} \in \Lambda}\|\mathbf{X} - \boldsymbol{\lambda}\|_{\mathbf{Q}}$. Then the shrinkage estimator is given by

$$\boldsymbol{\delta}_\phi(\Lambda) = \mathbf{X} - \phi(\mathbf{X}, S)S\|\mathbf{X} - P\mathbf{X}\|_{\mathbf{Q}}^{-2}\mathbf{Q}^{-1}\mathbf{V}^{-1}(\mathbf{X} - \mathbf{PX}). \tag{2.8}$$

Using the fact that $\mathbf{V}^{-1/2}\mathbf{P}\mathbf{V}^{-1/2}$ is idempotent, it can be easily verified that $\boldsymbol{\delta}_\phi(\Lambda)$ is closer than $\mathbf{X}$ [in the sense of (1.1)] if $p - r \ge 2$ and $0 < \phi(\mathbf{X}, S) \le (p - r - 1)(3p + 3r + 1)/(2(p - r))$, and this provides a natural extension of Theorem 1.

Before we consider the proof of Theorem 1, we present the following lemmas which are needed in this proof. Consider the confluent hypergeometric function

$$M(a, b, z) = 1 + az/b + \cdots + (a)_j z^j/(b)_j j! + \cdots, \tag{2.9}$$

where $(p)_j = p(p + 1)\cdots(p + j - 1)$, $j \ge 1$, and $(p)_0 = 1$.

LEMMA 2.1. *For real a, b, ν and z,*

$$\begin{aligned} &\sum_{j=0}^{\infty} (z^2/4)^j/\{j!\Gamma(\nu + j + 1)\} \\ &\qquad = \exp(-z)(\Gamma(\nu + 1))^{-1}M\left(\nu + \tfrac{1}{2}, 2\nu + 1, 2z\right), \end{aligned} \tag{2.10}$$

$$zM(a, b + 1, z) = bM(a, b, z) - bM(a - 1, b, z), \tag{2.11}$$

$$\begin{aligned} aM(a + 1, b, z) &= (1 + a - b)M(a, b, z) \\ &\quad + (b - 1)M(a, b - 1, z) \end{aligned} \tag{2.12}$$

and, for $a > 0$, $b > 1$, $b \geq a$ *and* $z > 0$,

$$M(a, b-1, z) \leq (1 + z/(b-1))M(a, b, z). \tag{2.13}$$

PROOF. First, (2.10) follows from (9.6.10) and (9.6.47) of Abramowitz and Stegun (1964), while their formulas (13.4.4) and (13.4.3) are rewritten as (2.11) and (2.12). Thus, we need to prove only (2.13). For this $M(a, b-1, z)$ is written as

$$\begin{aligned} &M(a, b-1, z) \\ &\quad = \sum_{j=0}^{\infty} \{(a)_j z^j (b - 1 + j)\}/\{(b)_j j!(b-1)\} \\ &\quad = M(a, b, z) + \{az/b(b-1)\} M(a+1, b+1, z). \end{aligned} \tag{2.14}$$

From Lemma 2.1 of Alam (1973), we conclude that $M(a+1, b+1, z)/M(a, b, z)$ is nondecreasing in z for $b \geq a$. Also, by (13.1.4) of Abramowitz and Stegun (1964),

$$M(a, b, z) = \{\Gamma(b)/\Gamma(a)\} e^z z^{a-b}\{1 + O(|z|^{-1})\} \quad \text{as } |z| \to \infty, \tag{2.15}$$

so that

$$\begin{aligned} M(a+1, b+1, z)/M(a, b, z) &\leq \lim_{z \to \infty} \{M(a+1, b+1, z)/M(a, b, z)\} \\ &= b/a, \end{aligned}$$

for every $z < \infty$. Hence, (2.13) follows from (2.14) and (2.15). □

LEMMA 2.2. *Let* $g_q^{(\lambda)}(x)$ *be the density function corresponding to the noncentral chi square d.f.* $G_q^{(\lambda)}(x)$ *with* q *DF and noncentrality parameter* λ. *Then, for any* $\lambda > 0$, $(\partial/\partial\lambda) g_{p+2}^{(\lambda)}(\lambda + a)$ *is nonnegative if* $0 < a < (p-1)(3p+1)/(4p)$.

PROOF. Let $g_q^{(0)}(x) = g_q(x)$ be the central chi square density function with q DF and let $f(\lambda) = (\partial/\partial\lambda) g_{p+2}^{(\lambda)}(\lambda + a)$. Then, we have

$$f(\lambda) = \tfrac{1}{2} \sum_{j=0}^{\infty} (\lambda/2)^j (j!)^{-1} e^{-\lambda/2} \{g_{p+2j}(a + \lambda) - g_{p+2+2j}(a + \lambda)\}. \tag{2.16}$$

Note that by virtue of the unimodality of the central chi square density,

$$\begin{aligned} g_{p+2j}(a+\lambda) - g_{p+2j+2}(a+\lambda) &\gtreqless 0 \text{ according as} \\ a + \lambda &\lesseqgtr p + 2j, \; j \geq 0. \end{aligned} \tag{2.17}$$

Therefore, $f(\lambda) \geq 0$ for any $\lambda \leq p - a$. Hence, we need to consider only the case of $\lambda > p - a$. By using (2.10) and (2.16), we obtain that $f(\lambda) \geq 0$ is equivalent to

$$\begin{aligned} pM(\tfrac{1}{2}(p-1), p-1, \delta) - (a + \lambda) M(\tfrac{1}{2}(p+1), p+1, \delta) &\geq 0, \\ \delta &= 2\sqrt{\lambda(a+\lambda)}. \end{aligned} \tag{2.18}$$

Making use of (2.11) and (2.12) on (2.18), we need to show that

$$(2.19)\quad 2\eta\, M\big(\tfrac{1}{2}(p-1), p, \delta\big) - (2\eta - 1)M\big(\tfrac{1}{2}(p-1), p-1, \delta\big) \geq 0, \qquad \forall\, \delta,$$

where $2\eta = (a+\lambda)^{1/2\lambda - 1/2} > 1$, for every $a > 0$. At this stage, we make use of (2.13) and conclude that (2.19) holds if $(2\eta - 1)\{1 + \delta/(p-1)\} \leq 2\eta$ or, equivalently,

$$(2.20)\quad \begin{aligned} &(a + \lambda - (p-1)/2)^2 \leq \big(\sqrt{\lambda}\sqrt{a+\lambda}\big)^2 = \lambda(a+\lambda) \\ &\Leftrightarrow (p - 1 - a)\lambda \geq \{a - (p-1)/2\}^2. \end{aligned}$$

Since $\lambda > p - a$ and (2.20) is to hold for all $\lambda > p - a$, we must have $(p - 1 - a)(p - a) \geq \{a - (p-1)/2\}^2$, and this is always guaranteed for $0 < a < (p-1)(3p+1)/(4p)$. □

LEMMA 2.3. *$(d/dx)\log g_{p+2}^{(\lambda)}(x)$ is nonincreasing and convex in x, for all $x \geq 0$; $\lambda \geq 0$.*

The proof of the lemma is relegated to the Appendix.

Let us now return to the proof of Theorem 1. Defining $\boldsymbol{\delta}_\phi$ as in (1.3) and writing ϕ for $\phi(\mathbf{X}, S)$, we may note that $[\|\boldsymbol{\delta}_\phi - \boldsymbol{\theta}\|_{\mathbf{Q}} \leq \|\mathbf{X} - \boldsymbol{\theta}\|_{\mathbf{Q}}]$ is equivalent to $[2(\mathbf{X} - \boldsymbol{\theta})'\mathbf{V}^{-1}\mathbf{X} \geq \phi S]$. Hence, by reference to (2.1), it suffices to show that

$$(2.21)\quad P_\omega\{cmS \leq (\mathbf{X} - \boldsymbol{\theta})'\mathbf{V}^{-1}\mathbf{X}\} \geq \tfrac{1}{2} \quad \text{for } c = (p-1)(3p+1)/(4pm), \forall\, \omega.$$

Define $T = mS/\sigma^2$, $\lambda = (\boldsymbol{\theta}'\mathbf{V}^{-1}\boldsymbol{\theta})/4$ and let $\mathbf{P}$ be an orthogonal matrix, such that $\mathbf{P}\mathbf{V}^{-1/2}\boldsymbol{\theta} = 4\lambda(1, \mathbf{0}')'$. Let then $\mathbf{Z} = \sigma^{-1}\mathbf{P}\mathbf{V}^{-1/2}(\mathbf{X} - \boldsymbol{\theta})$, so that $\mathbf{Z}$ has the normal d.f. with null mean vector and dispersion matrix $\mathbf{I}_p$ and T has the central chi square d.f. with m DF; T and $\mathbf{Z}$ are mutually independent too. Then

$$(2.22)\quad \begin{aligned} &P_\omega\{cmS \leq (\mathbf{X} - \boldsymbol{\theta})'\mathbf{V}^{-1}\mathbf{X}\} \\ &\quad = P_\lambda\{(cT - \|\mathbf{Z}\|^2)/(2\lambda) \leq Z_1\} = h(\lambda), \quad \text{say.} \end{aligned}$$

Through the distributional results given in Rao, Keating and Mason (1986) [see also Efron (1975)], $h(\lambda)$ has been numerically studied; the task remains to show analytically that $h(\lambda) \geq \frac{1}{2}$, for every finite $\lambda \geq 0$. By the dominated convergence theorem, we claim that as $\lambda \to \infty$, $h(\lambda) \to P\{Z_1 \geq 0\} = \frac{1}{2}$ and, hence, it suffices to show that

$$(2.23)\quad h'(\lambda) = (d/d\lambda)h(\lambda) \leq 0 \quad \text{for every finite (positive) } \lambda.$$

Defining the g_m, $g_p^{(\lambda)}$ as in Lemma 2.2, we note that

$$(2.24)\quad \begin{aligned} h(\lambda) &= P_\omega\{(\mathbf{X} - \tfrac{1}{2}\boldsymbol{\theta})'\mathbf{V}^{-1}(\mathbf{X} - \tfrac{1}{2}\boldsymbol{\theta})/\sigma^2 \geq \lambda + cT\} \\ &= P_\lambda\{\chi^2_{p,\lambda} \geq \lambda + c\chi^2_m\} \\ &= \int_0^\infty \int_{\lambda + ct}^\infty g_p^{(\lambda)}(u) g_m(t)\, du\, dt. \end{aligned}$$

Thus,

$$h'(\lambda) = \int_0^\infty \left\{ -g_p^{(\lambda)}(\lambda + ct) + \int_{\lambda+ct}^\infty (d/d\lambda) g_p^{(\lambda)}(u)\, du \right\} g_m(t)\, dt$$

$$(2.25) \qquad = \int_0^\infty \left\{ g_{p+2}^{(\lambda)}(\lambda + ct) - g_p^{(\lambda)}(\lambda + ct) \right\} g_m(t)\, dt$$

$$= -2 \int_0^\infty \psi(t) \phi_m(t)\, dt,$$

where $\psi(t) = (d/dt)\log g_{p+2}^{(\lambda)}(\lambda + ct)$ and $\phi_m(t) = g_{p+2}^{(\lambda)}(\lambda + ct) g_m(t)$. Let $A_m = \int_0^\infty \phi_m(t)\, dt$ and let $E_*[\cdot]$ stand for the expectation with respect to the probability law given by $P(B) = \int_B \phi_m(t)\, dt / A_m$ for a set B. Then, by the convexity property in Lemma 2.3, and the Jensen inequality, we have

$$(2.26) \qquad \begin{aligned} h'(\lambda) &= -2A_m E_*[\psi(T)] \le -2A_m \psi(E_*[T]), \\ E_*[T] &= mA_{m+2}/A_m. \end{aligned}$$

Hence, from Lemma 2.2, it follows that $h'(\lambda) \le 0$ if $cmA_{m+2}/A_m \le (p-1) \cdot (3p+1)/(4p)$ or $A_{m+2} \le A_m$. A different proof is needed for the case when $A_{m+2} \ge A_m$. Here

$$A_{m+2} - A_m = \int_0^\infty g_{p+2}^{(\lambda)}(\lambda + ct) \{ g_{m+2}(t) - g_m(t) \}\, dt$$

$$= -2 \int_0^\infty g_{p+2}^{(\lambda)}(\lambda + ct) \{ (d/dt) g_{m+2}(t) \}\, dt$$

$$(2.27) \qquad = 2 \int_0^\infty g_{m+2}(t) \left\{ (d/dt) g_{p+2}^{(\lambda)}(\lambda + ct) \right\} dt$$

$$= 2m^{-1} \int_0^\infty t g_m(t) \left\{ (d/dt) g_{p+2}^{(\lambda)}(\lambda + ct) \right\} dt$$

$$= 2m^{-1} A_m E_*[T\psi(T)]$$

$$\le 2m^{-1} A_m E_*[T] E_*[\psi(T)],$$

as $\psi(t)$ is nondecreasing in t [see also Das Gupta and Sarkar (1984) in this respect]. Therefore, $E_*[\psi(T)]$ is nonnegative when $A_{m+2} - A_m \ge 0$, and hence, by (2.25) and (2.26) we conclude that $h'(\lambda) \le 0$, for every $\lambda > 0$ when $A_{m+2} - A_m \ge 0$. This shows that (2.23) holds for all $\lambda \ge 0$ and the proof of the theorem is complete.

Some additional results of considerable interest are presented in the next section.

3. Some related dominance results. First, we consider the case of a multivariate normal population with unknown mean vector $\boldsymbol{\theta}$ and arbitrary (p.d.) covariance matrix $\boldsymbol{\Sigma}$. As in Berger, Bock, Brown, Casella and Gleser (1977) and

Stein (1981), we consider a shrinkage estimator of the form

$$\begin{aligned} \delta_\phi^* &= \mathbf{X} - (m-p+1)^{-1}\phi(\mathbf{X},\mathbf{S})d_m(\mathbf{X}'\mathbf{S}^{-1}\mathbf{X})^{-1}\mathbf{Q}^{-1}\mathbf{S}^{-1}\mathbf{X}, \\ d_m &= d = \mathrm{ch}_{\min}(\mathbf{QS}), \qquad m \geq p, \end{aligned} \tag{3.1}$$

where $\mathbf{X} \sim \mathscr{N}_p(\boldsymbol{\theta}, \boldsymbol{\Sigma})$, $\mathbf{S}$ (independently of $\mathbf{X}$) has a Wishart $(\boldsymbol{\Sigma}, p, m)$ distribution and $\mathrm{ch}_{\min}(\mathbf{A})$ [or $\mathrm{ch}_{\max}(\mathbf{A})$] stands for the smallest (or largest) characteristic root of $\mathbf{A}$. Then, we have the following.

THEOREM 2. *Assume that* $p \geq 2$ *and* (2.1) *holds when the scalar r.v.* S *is replaced by the Wishart matrix* $\mathbf{S}$. *Then* δ_ϕ^*, *given by* (3.1), *is closer than* $\mathbf{X}$ *in the Pitman sense in* (1.1).

OUTLINE OF THE PROOF. Proceeding as in Section 2, we need to show that

$$\begin{aligned} P_\omega\big\{2(\mathbf{X}-\boldsymbol{\theta})'\mathbf{S}^{-1}\mathbf{X} &\geq (m-p+1)^{-1}\phi(\mathbf{X},\mathbf{S})d \\ &\times(\mathbf{X}'\mathbf{S}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{S}^{-1}\mathbf{Q}^{-1}\mathbf{S}^{-1}\mathbf{X}\big\} \geq \tfrac{1}{2}, \qquad \forall\, \omega = (\boldsymbol{\theta}, \boldsymbol{\Sigma}). \end{aligned} \tag{3.2}$$

Since $\mathbf{X}'\mathbf{S}^{-1}\mathbf{Q}^{-1}\mathbf{S}^{-1}\mathbf{X}/(\mathbf{X}'\mathbf{S}^{-1}\mathbf{X}) \leq \mathrm{ch}_{\max}(\mathbf{S}^{-1}\mathbf{Q}^{-1}) = d^{-1}$, it suffices to show that for all ω,

$$\begin{aligned} P_\omega\big\{(\mathbf{X}-\boldsymbol{\theta})'\mathbf{S}^{-1}\mathbf{X} \geq c\big\} &\geq \tfrac{1}{2} \\ \text{for } c = (p-1)(3p+1)/\{4p(m-p+1)\}&. \end{aligned} \tag{3.3}$$

Letting $\mathbf{Y} = \boldsymbol{\Sigma}^{-1/2}(\mathbf{X}-\boldsymbol{\theta})$, $\mathbf{W} = \boldsymbol{\Sigma}^{-1/2}\mathbf{S}\boldsymbol{\Sigma}^{-1/2}$, $\boldsymbol{\gamma} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\theta}$ and $\lambda = (\boldsymbol{\theta}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\theta})/4$, it can be shown by some standard steps that

$$\begin{aligned} (\mathbf{X}-\boldsymbol{\theta})'\mathbf{S}^{-1}\mathbf{X} &= \mathbf{Y}'\mathbf{W}^{-1}\mathbf{Y} + \boldsymbol{\gamma}'\mathbf{W}^{-1}\mathbf{Y} \\ &= (\mathbf{Y}-\boldsymbol{\gamma})'\mathbf{W}^{-1}(\mathbf{Y}-\boldsymbol{\gamma}) - \left(\tfrac{1}{4}\right)\boldsymbol{\gamma}'\mathbf{W}^{-1}\boldsymbol{\gamma} \\ &= \left\{\chi^2_{p,\lambda} - \lambda\right\}/\chi^2_{m-p+1}, \end{aligned} \tag{3.4}$$

so that for (3.3), we need to verify that

$$P_\lambda\left\{\chi^2_{p,\lambda} \geq \lambda + c\chi^2_{m-p+1}\right\} \geq \tfrac{1}{2} \quad \text{for all } \lambda \geq 0, \tag{3.5}$$

and at this stage the proof of Theorem 1 can be called on to complete the task. □

In the context of robust and nonparametric estimation, Stein-rule estimators have been considered and the asymptotic dominance has been established when the true parameter point belongs to a Pitman neighborhood of the assumed pivot; in this setup the normality assumption on the underlying d.f. has been dispensed with and a general class of shrinkage estimators has been incorporated in the formulation of Stein-type estimators. We may consider the same setup in the light of the Pitman closeness.

Let $\{\mathbf{X}_n;\ n \geq n_0\}$ be a sequence of estimators such that asymptotically $n^{1/2}(\mathbf{X}_n - \boldsymbol{\theta})$ has a normal distribution with null mean vector and covariance matrix $\boldsymbol{\Sigma}$. Also, let $\{\mathbf{S}_n;\ n \geq n_0\}$ be a sequence of stochastic matrices such that $\mathbf{S}_n \to \boldsymbol{\Sigma}$, in probability, as $n \to \infty$. Further, corresponding to the null pivot for $\boldsymbol{\theta}$, consider a sequence $\{K_n\}$ of local (Pitman) alternatives

$$K_n\colon \boldsymbol{\theta} = \boldsymbol{\theta}_{(n)} = n^{-1/2}\boldsymbol{\xi}, \qquad \boldsymbol{\xi}\ (\text{fixed}) \in R^p. \tag{3.6}$$

Then, typically a Stein-rule version of $\mathbf{X}_n$ is given by

$$\boldsymbol{\delta}_n = \mathbf{X}_n - ad_n\left(\mathbf{X}_n'\mathbf{S}_n^{-1}\mathbf{X}_n\right)^{-1}\mathbf{Q}^{-1}\mathbf{S}_n^{-1}\mathbf{X}_n, \qquad d_n = \mathrm{ch}_{\min}(\mathbf{Q}\mathbf{S}_n), \tag{3.7}$$

where a is a positive constant. In terms of the asymptotic distributional risk of $n^{1/2}(\boldsymbol{\delta}_n - \boldsymbol{\theta}_{(n)})$ {under K_n}, the dominance results have been studied earlier by Sen (1984), Sen and Saleh (1985, 1987) and Saleh and Sen (1985, 1986), among others. By an appeal to Theorem 2 and some standard arguments for the incorporation of the asymptotic theory, it follows that under $\{K_n\}$, for every a: $0 < a < (p-1)(3p+1)/(2p)$, $p \geq 2$, $\boldsymbol{\delta}_n$ dominates $\mathbf{X}_n$ in the light of the Pitman closeness in (1.1) when n is large. Thus, the usual robust and nonparametric Stein-rule estimators enjoy the Pitman closeness dominance property in the asymptotic case under less restrictive regularity conditions.

APPENDIX

Proof of Lemma 2.3. We provide a broad outline of the proof. Note that on letting

$$a_r = \{2^r r!\,\Gamma(r+1+p/2)/\Gamma(1+p/2)\}^{-1}, \qquad r = 0, 1, \dots,$$

we have

$$g_{p+2}^{(\lambda)}(x) = \text{const.}\, e^{-x/2}x^{p/2} \sum_{r \geq 0} (\lambda x/2)^r a_r. \tag{A.1}$$

Thus, writing $\theta = (\lambda x)/2$ and $q(\theta) = \sum_{r \geq 0} a_r \theta^r$, $g(\theta) = q'(\theta)/q(\theta)$, we have

$$(d/dx)\log g_{p+2}^{(\lambda)}(x) = (p/x - 1)/2 + (\lambda/2)g(\lambda x/2), \qquad x \geq 0. \tag{A.2}$$

Hence to show that $(d/dx)\log g_{p+2}^{(\lambda)}(x)$ is $\downarrow$ in x (≥ 0), it suffices to show that $g'(\theta) \leq 0$, for every $\theta \geq 0$. Toward this, note that

$$\begin{aligned} g'(\theta) &= -\{q'(\theta)/q(\theta)\}\{q'(\theta)/q(\theta) - q''(\theta)/q'(\theta)\} \\ &= -g^2(\theta) + (2\theta)^{-1} - \{(p+2)/2\}\theta^{-1}g(\theta) \\ &= -(2\theta)^{-1}\{\nu(\theta)g(\theta) - 1\}, \end{aligned} \tag{A.3}$$

$$\nu(\theta) = p + 2 + 2l(\theta) \quad \text{and} \quad l(\theta) = \theta g(\theta). \tag{A.4}$$

At this stage we note that

$$g(\theta) = E\left[(p+2+2K)^{-1}\right] \qquad \left[\geq (p+2+2E(K))^{-1}\right], \tag{A.5}$$

where K is a r.v. having the power series distribution with $P(K = r) =$

$a_r\theta^r/q(\theta)$, $r = 0, 1, \dots,$ so that $E(K) = l(\theta)$. Thus, $g(\theta)\nu(\theta) \geq 1$ and, hence, (A.3) is nonpositive.

To establish the desired convexity property, note that

$$(d^3/dx^3)\log g_{p+2}^{(\lambda)}(x) = (p/x^3) + (\lambda/2)^3 g''(\lambda x/2) = (\lambda^3/8)\{p/\theta^3 + g''(\theta)\}, \tag{A.6}$$

where, by (A.3),

$$\begin{aligned} g''(\theta) &= (2\theta^2)^{-1}(\nu(\theta)g(\theta) - 1) - (2\theta)^{-1}(\nu(\theta)g'(\theta) + \nu'(\theta)g(\theta)) \\ &= (4\theta^2)^{-1}\{[\nu(\theta)g(\theta) - 1][\nu(\theta) + 2l(\theta) + 2] - 4l(\theta)g(\theta)\}. \end{aligned} \tag{A.7}$$

Thus, in order that (A.6) is nonnegative for all $\theta \geq 0$, it suffices to show that

$$\nu(\theta)g(\theta) - 1 \geq 4\{l^2(\theta) - p\}/\{\theta(2\nu(\theta) - p)\}, \qquad \forall\, \theta \geq 0. \tag{A.8}$$

Since $\nu(\theta)g(\theta) \geq 1$, (A.8) holds automatically for all θ: $l^2(\theta) \leq$ p. Thus, we need to consider only the case of θ: $l^2(\theta) > p$. Toward this note that

$$\begin{aligned} (p + 2 + 2K)^{-1} &= 1/\nu(\theta) - 2[K - l(\theta)]/\nu^2(\theta) + 4[K - l(\theta)]^2/\nu^3(\theta) \\ &\quad + \{8[l(\theta) - K]^3/\nu^3(\theta)\}/\{p + 2 + 2K\}, \end{aligned} \tag{A.9}$$

where $[l(\theta) - K]^3$ and $(p + 2K + 2)^{-1}$ are both $\downarrow$ in K (and hence, concordant), so that

$$g(\theta)\nu(\theta) - 1 \geq \{4V(K)\}/\nu^2(\theta) - \{8E[K - l(\theta)]^3\}g(\theta)/\nu^2(\theta). \tag{A.10}$$

Using the recursion relations among the cumulants of a power series distribution [viz., Johnson and Kotz (1969), page 34], we obtain that

$$V(K) = \theta[(\partial/\partial\theta)l(\theta)] = l(\theta) - [g(\theta)\nu(\theta) - 1]\theta/2, \tag{A.11}$$

$$\begin{aligned} E[K - l(\theta)]^3 &= l(\theta) - l^2(\theta) + (\theta/4) \\ &\quad \times [g(\theta)\nu(\theta) - 1][\nu(\theta) + 2l(\theta) - 4]. \end{aligned} \tag{A.12}$$

From (A.10), (A.11) and (A.12), we obtain that

$$\begin{aligned} \nu(\theta)g(\theta) - 1 &\geq \{4l(\theta)/\nu^2(\theta)\}\{1 + 2g(\theta)[l(\theta) - 1]\} \\ &\quad \div \{1 + 2\theta[1 + g(\theta)\{\nu(\theta) + 2l(\theta) - 4\}]/\nu^2(\theta)\}. \end{aligned} \tag{A.13}$$

Letting $a(\theta) = g(\theta)\nu(\theta) - 1$ and noting that

$$g(\theta) = [1 + a(\theta)]/\nu(\theta) \tag{A.14}$$

and

$$2\theta/\nu^2(\theta) = [2l(\theta)/\nu(\theta)][\nu(\theta)g(\theta)]^{-1} = [2l(\theta)/\nu(\theta)][1 + a(\theta)]^{-1}, \tag{A.15}$$

we obtain from (A.13) that

$$\begin{aligned} a(\theta) &\geq 4l(\theta)\{1 + [2/\nu(\theta)][1 + a(\theta)][l(\theta) - 1]\} \\ &\quad \div \nu^2(\theta)\{1 + [2l(\theta)/\nu(\theta)\{1 + a(\theta)\}] \\ &\quad \times [1 + [1 + a(\theta)][\nu(\theta) + 2l(\theta) - 4]]\} \\ &= \frac{4l(\theta)\{\nu(\theta) + 2l(\theta) - 2 + 2a(\theta)[l(\theta) - 1]\}}{\nu(\theta)\{\nu(\theta)[\nu(\theta) - 2l(\theta) + 4l^2(\theta)/\nu(\theta) - 8l(\theta)/\nu(\theta)]\}}. \end{aligned} \tag{A.16}$$

Since $a(\theta) \geq 0$, for $l(\theta) > 1$, the right-hand side of (A.16) is greater than

$$\begin{aligned} &\frac{4l(\theta)[\nu(\theta) - 2][2\nu(\theta) - p - 4]}{\{\nu(\theta) - 2\}\{\nu^2(\theta)[2\nu(\theta) - p - 4 + 4\{l(\theta)/\nu(\theta)\}\{l(\theta) - 2\}]\}} \\ &= \frac{4l(\theta)}{\nu(\theta) - 2} \frac{1 - p/l^2(\theta)}{2\nu(\theta) - p} \frac{\alpha(\theta)}{\beta(\theta)}, \end{aligned} \tag{A.17}$$

where

$$\alpha(\theta) = [2\nu(\theta) - p][\nu(\theta) - 2][2\nu(\theta) - p - 2], \tag{A.18}$$

$$\begin{aligned} \beta(\theta) = \nu^2(\theta)[1 - p/l^2(\theta)][2\nu(\theta) - p - 4\{1 - l^2(\theta)/\nu(\theta) \\ + 2l(\theta)/\nu(\theta)\}]. \end{aligned} \tag{A.19}$$

Direct computations yield that for every $l^2(\theta) \geq p$,

$$\alpha(\theta)/\beta(\theta) \geq 1 \quad \text{and} \quad 4l^2(\theta)/\theta \leq 4l(\theta)/[\nu(\theta) - 2]; \tag{A.20}$$

hence, (A.8) follows from (A.13), (A.17) and (A.20). □

Acknowledgments. The authors are grateful to the Editor, Associate Editor and the referees for making concrete and useful suggestions toward this combined writeup.

REFERENCES

ABRAMOWITZ, M. and STEGUN, I. A. (1964). *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables. Appl. Math. Ser.* **55**. National Bureau of Standards, Washington.

ALAM, K. (1973). A family of admissible minimax estimators of the mean of a multivariate normal distribution. *Ann. Statist.* **1** 517–525.

ANDERSON, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*, 2nd ed. Wiley, New York.

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

BERGER, J. O. (1985). *Statistical Decision Theory*, 2nd ed. Springer, New York.

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.

DAS GUPTA, S. and SARKAR, S. K. (1984). On TP_2 and log-concavity. In *Inequalities in Statistics and Probability* (Y. L. Tong, ed.) 54–58. IMS, Hayward, Calif.

EFRON, B. (1975). Biased versus unbiased estimation. *Adv. in Math.* **16** 259–277.

HWANG, J. T. (1985). Universal domination and stochastic domination: Estimation simultaneously under a broad class of loss functions. *Ann. Statist.* **13** 295–314.

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

JOHNSON, N. L. and KOTZ, S. (1969). *Distributions in Statistics: Discrete Distributions.* Wiley, New York.

KEATING, J. P. and MASON, R. L. (1988). James–Stein estimation from an alternative perspective. *Amer. Statist.* **42** 160–164.

LEE, C.M.-S. (1987). On the characterization of Pitman measures of nearness. Technical Report, Dept. Mathematics, Central Michigan Univ.

PEDDADA, S. D. and KHATTREE, R. (1986). On Pitman nearness and variance of estimators. *Comm. Statist. A—Theory Methods* **15** 3005–3017.

PITMAN, E. J. G. (1937). The closest estimates of statistical parameters. *Proc. Cambridge Philos. Soc.* **33** 212–222.

RAO, C. R. (1981). Some comments on the minimum mean square error as a criterion of estimation. In *Statistics and Related Topics* (M. Csörgő, D. A. Dawson, J. N. K. Rao and A. K. Md. E. Saleh, eds.) 123–143. North-Holland, Amsterdam.

RAO, C. R., KEATING, J. P. and MASON, R. L. (1986). The Pitman nearness criterion and its determination. *Comm. Statist. A—Theory Methods* **15** 3173–3191.

SALEH, A. K. MD. E. and SEN, P. K. (1985). On shrinkage *M*-estimators of location parameters. *Comm. Statist. A—Theory Methods* **14** 2313–2329.

SALEH, A. K. MD. E. and SEN, P. K. (1986). On shrinkage *R*-estimation in a multiple regression model. *Comm. Statist. A—Theory Methods* **15** 2229–2244.

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

SEN, P. K. (1984). A James–Stein detour of *U*-statistics. *Comm. Statist. A—Theory Methods* **13** 2725–2747.

SEN, P. K. (1986a). Are BAN estimators the Pitman closest ones too? *Sankhyā Ser. A* **48** 51–58.

SEN, P. K. (1986b). On the asymptotic distributional risks of shrinkage and preliminary test versions of maximum likelihood estimators. *Sankhyā Ser. A* **48** 354–372.

SEN, P. K. (1989). The mean-median-mode inequality and noncentral chi square distributions. *Sankhyā Ser. A* **51**.

SEN, P. K. and SALEH, A. K. MD. E. (1985). On some shrinkage estimators of multivariate location. *Ann. Statist.* **13** 272–281.

SEN, P. K. and SALEH, A. K. MD. E. (1987). On the preliminary test and shrinkage *M*-estimation in linear models. *Ann. Statist.* **15** 1580–1592.

SEN, P. K. and SALEH, A. K. MD. E. (1989). Some recent developments in nonparametric and robust shrinkage estimation theory. In *Proc. Korea–Ohio State Univ. Statistics Conference* (J. C. Lee and J. P. Klein, eds.). To appear.

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

STEIN, C. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* **9** 1135–1151.

PRANAB KUMAR SEN
DEPARTMENTS OF BIOSTATISTICS AND STATISTICS
UNIVERSITY OF NORTH CAROLINA
CHAPEL HILL, NORTH CAROLINA 27599–7400

TATSUYA KUBOKAWA
INSTITUTE OF MATHEMATICS
UNIVERSITY OF TSUKUBA
TSUKUBA, IBARAKI 305
JAPAN

A. K. MD. EHSANES SALEH
DEPARTMENT OF MATHEMATICS AND STATISTICS
CARLETON UNIVERSITY
OTTAWA, ONTARIO
CANADA K1S 5B6