The Annals of Statistics
1989, Vol. 17, No. 1, 252–267

UNIVERSAL DOMINATION AND STOCHASTIC DOMINATION: *U*-ADMISSIBILITY AND *U*-INADMISSIBILITY OF THE LEAST SQUARES ESTIMATOR[1]

BY LAWRENCE D. BROWN AND JIUNN T. HWANG

Cornell University

Assume the standard linear model

$$\underset{n\times 1}{X} = \underset{n\times p}{A}\ \underset{p\times 1}{\theta} + \underset{n\times 1}{\varepsilon},$$

where ε has an n-variate normal distribution with zero mean vector and identity covariance matrix. The least squares estimator for the coefficient θ is $\hat{\theta} \equiv (A'A)^{-1}A'X$. It is well known that $\hat{\theta}$ is dominated by James–Stein type estimators under the sum of squared error loss $|\theta - \hat{\theta}|^2$ when $p \geq 3$.

In this article we discuss the possibility of improving upon $\hat{\theta}$, simultaneously under the "universal" class of losses:

$$\{L(|\theta - \hat{\theta}|): L(\cdot) \text{ any nondecreasing function}\}.$$

An estimator that can be so improved is called universally inadmissible (U-inadmissible). Otherwise it is called U-admissible.

We prove that $\hat{\theta}$ is U-admissible for any p when $A'A = I$. Furthermore, if $A'A \neq I$, then $\hat{\theta}$ is U-inadmissible if p is "large enough." In a special case, $p \geq 4$ is large enough. The results are surprising. Implications are discussed.

1. Introduction. In decision theory, a single loss function is typically used. See Wald (1950). However, in practice, a loss is difficult to specify exactly and therefore it is meaningful to consider a class of loss functions. Even though there is a huge literature in single loss decision theory, there are only a few results dealing with a class of losses. See the Introduction of Hwang (1985) for a review.

The same article as well as this article study the *universal class* of loss functions based on the *Q-generalized Euclidean error* $|\delta - \theta|_Q$,

$$\{L(|\delta - \theta|_Q): L(\cdot) \text{ any nondecreasing function}\}. \tag{1.1}$$

Here, the parameter θ and its estimate δ are both p-dimensional quantities, Q is a fixed positive definite matrix and

$$|\delta - \theta|_Q \equiv \left[(\delta - \theta)^t Q(\delta - \theta)\right]^{1/2}. \tag{1.2}$$

The situation in which Q is fixed and known arises in two applications described at the end of the Introduction.

We will study for a fixed Q the theoretical question about the existence of an estimator δ that *universally dominates* the least squares estimator δ_0, i.e., δ is as good as δ_0 for every loss in the universal class and δ is better than δ_0 for at least

Received November 1986; revised May 1988.
[1]Research of both authors supported by National Science Foundation Grant DMS-85-06847.
AMS 1980 *subject classifications*. Primary 62C05, 62F11; secondary 62J07.
Key words and phrases. Decision theory under a broad class of loss functions, James–Stein positive part estimator, admissibility.

one of the losses. In the situation that such a δ exists, δ^0 is called *U-inadmissible with respect to Q*. Otherwise, if there is no estimator universally dominating δ^0, δ^0 is called *U-admissible with respect to Q*. In the discussion below, we may omit "with respect to Q," if this causes no ambiguity. Unfortunately the arguments in this article are not constructive; thus the problem of actually constructing universally better estimators remains open.

For background, we summarize some results from Hwang (1985). Only higher-dimensional cases are discussed, since for lower-dimensional cases, δ^0 is typically U-admissible.

The first example of universal domination concerns the situation in which $X - \theta$ has a p-dimensional t distribution with known degree of freedom N. It is shown in Hwang (1985), that for every fixed N and $p \geq 3$, X is U-inadmissible with respect to *Euclidean error* (i.e., Q-generalized Euclidean error with $Q = I$).

For the important normal case ($X \sim N(\theta, I)$, I the identity matrix), Hwang (1985) compared X to the positive part James–Stein estimator (1960),

$$(1.3)\qquad \delta^\alpha(X) = \left(1 - \frac{\alpha}{|X|^2}\right)^+ X \equiv \max\left(0, 1 - \frac{\alpha}{|X|^2}\right)X.$$

It was shown that, for every α, δ^α does not universally dominate X with respect to Euclidean error.

This leaves open the question (for the normal case) as to whether there exists an estimator universally dominating X for $p \geq 3$. In Section 2, we show that, for Euclidean error, there exists no such estimator; hence, for every p, X is U-admissible.

We also consider a more general problem with respect to Q-generalized Euclidean error. Let $q_1 \geq q_2 \geq \cdots \geq q_p > 0$ be the eigenvalues of Q. Under the normal model, it is shown in Section 3 that if $q_1 > q_2$, then X is U-inadmissible when the dimension p is big enough. For the special case

$$q_1 > q_2 = \cdots = q_p,$$

X is U-inadmissible if and only if $p \geq 4$. We conjecture that X is also U-inadmissible whenever $p \geq 4$ and $q_1 > q_{p-2}$.

Our studies here strengthen Stein's phenomenon (1956) for the situation $Q \neq I$. In Stein's context, only a single loss (sum of squared error loss) is considered and X can be improved when $p \geq 3$. Here for $p \geq 4$ and most Q, $Q \neq I$, we show that improvement is possible even for the universal class of losses. However, there are two notable differences between Stein's phenomenon and our results. First, it is surprising that U-admissibility depends critically on Q. To the best of our knowledge such dependence on Q has not previously been observed in the admissibility paradigm. In fact, theorems are available which show that admissibility under the loss $|\theta - \delta|_Q^2$ is equivalent to admissibility under the loss $|\theta - \delta|_{Q'}^2$, so long as Q and Q' are both nonsingular. [See Bhattacharya (1966), Berger (1979), Shinozaki (1975) and Lemma 3.1 in Rao (1976).] Our results show that these theorems cannot be extended to the U-admissibility criterion.

Another difference is the cutoff dimension. The minimum dimension for inadmissibility in the normal case (or, in general, in the location case) is 3. However, if Q is nonsingular, the minimum dimension for U-inadmissibility is at least 4. When $q_2 = q_3 = \cdots = q_p$, it is exactly 4.

We have assumed that Q is fixed and known. Below we discuss two situations in which the assumption is automatically satisfied. Consider the standard linear model

$$(1.4) \qquad \underset{n\times 1}{X} = \underset{n\times p}{A}\ \underset{p\times 1}{\theta} + \underset{n\times 1}{\varepsilon},$$

where A is the design matrix of full rank p, θ is the unknown regression parameter and ε has a $N(0, \sigma^2 I)$ distribution. The first justification relates to evaluation of the least squares estimator $\hat{\theta} = (A'A)^{-1}A'X$ with respect to any nondecreasing loss based on Euclidean distance. (This distance is natural when components of θ are of equal importance.) Using transformations, one can show that $\hat{\theta}$ is U-admissible if and only if $Y = (A'A)^{1/2}\hat{\theta}$ is U-admissible for estimating $\eta = EY$ with respect to $Q = (A'A)^{-1}$, which is a fixed and known positive definite matrix.

The second justification arises in the prediction context. Assume (1.4) and suppose that we are interested in predicting a future observation $X^* = A^*\theta + \varepsilon^*$, where ε^* is normally distributed with mean 0. A natural predictor based on the least squares estimator, $\hat{\theta}$, is $A^*\hat{\theta}$. Suppose we compare this predictor to other predictors of the form $\delta^*(Y) = A^*\delta(\hat{\theta})$, using the class of loss functions

$$(1.5) \qquad \{L(|\delta - X^*|),\ L(\cdot) \text{ nondecreasing}\}.$$

The risk function of $A^*\delta(\hat{\theta})$ is

$$E_\theta L\big(|A^*\delta(\hat{\theta}) - X^*|\big) = E_\theta L^*\big(|A^*\delta(\hat{\theta}) - A^*\theta|\big),$$

where

$$(1.6) \qquad L^*\big(|A^*\delta(\hat{\theta}) - A^*\theta|\big) = E\big\{L\big(|A^*\delta(\hat{\theta}) - A^*\theta - \varepsilon^*|\big)|\hat{\theta}\big\}.$$

The above expression is an increasing function of $|A^*\delta(\hat{\theta}) - A^*\theta|$ due to the normality of ε^*. Finally, one can write (1.6) as $L^*(|\delta(\hat{\theta}) - \theta|_{Q^*})$, with $Q^* = A^{*\prime}A^*$; this, in turn, can be written as $L^*(|\delta^*(Y) - \eta|_Q)$ with $Q = (A'A)^{-1/2}Q^*(A'A)^{-1/2}$. Hence the conclusion is that if there exists an estimator universally dominating Y with respect to this Q, then a suitably modified estimator simultaneously dominates the intuitive estimator $A^*\hat{\theta}$ for all losses in (1.5).

2. U-admissibility of the least squares estimator. For $Q = I$, we shall prove

THEOREM 1. *Assume that X has a p-dimensional $N(\theta, I)$ distribution. For any p, the least squares estimator $\delta^0(X) = X$ is U-admissible with respect to Euclidean error.*

We will first give an intuitive argument to explain why Theorem 1 stands. The rigorous proof mimics the following argument, except the losses are slightly modified. Suppose that $\delta(X)$ is an alternative estimator, which for now is assumed to be nonrandomized and continuous. Assume that $\delta(x)$ and x are not identical. Let x_0 be such that $\delta(x_0) \neq x_0$. Also take θ so that x_0 is on the line segment joining $\delta(x_0)$ and θ. (See Figure 1.) By continuity, we can find two separated open spheres S_1 and S_2 centered at x_0 and $\delta(x_0)$, respectively, so that δ maps S_1 into a subset of S_2. Let θ be on the ray from $\delta(x_0)$ through x_0. Let $c = c(\theta)$ be a positive number so that the sphere centered at θ of radius c will separate S_1 and S_2. Now considered a step loss function $\chi_{(c,\infty)}(|\theta - \delta|)$. [Throughout the paper, $\chi_S(\cdot)$ denotes the indicator function of S.] Then

$$R(\theta, \delta) = P_\theta(|\theta - \delta(X)| \geq c) \geq P_\theta(\delta(X) \in S_2) \geq P_\theta(X \in S_1).$$

Hence

$$\frac{R(\theta, \delta)}{R(\theta, \delta^0)} \geq \frac{P_\theta(X \in S_1)}{P_\theta(|X - \theta| \geq c)} \to \infty, \quad \text{a.s. } c \to \infty,$$

since X has a normal distribution with an exponential tail [Birnbaum (1955)]. Hence δ does not simultaneously dominate δ^0 for every loss in the class of step loss functions, and consequently δ does not universally dominate δ^0.

Now we turn to the formal proof of Theorem 1. In the proof, instead of the step loss functions in the above argument, we consider the class of loss $L_c(|\delta - \theta|)$, where

$$L_c(t) = (t - c)^+, \qquad 0 \leq c < \infty.$$

Theorem 1 will easily follow from Theorem 2, which is stated and proved below. (The convexity of L_c is used in the proof of Theorem 1, which follows the proof of Theorem 2.)

THEOREM 2. *Suppose $\delta(X)$ is a nonrandomized estimator such that for all θ*

$$E_\theta L_c(|\delta(X) - \theta|) \leq E_\theta L_c(|\delta_0(X) - \theta|) \quad \forall\, 0 \leq c < \infty. \tag{2.1}$$

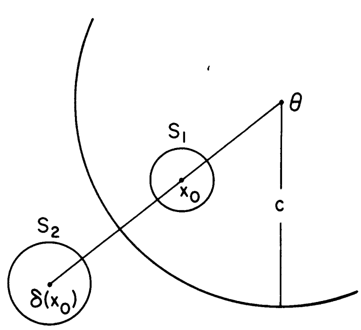

FIG. 1.

Then $\delta(X) = \delta^0(X)$ *a.e.*

PROOF. Let δ satisfy (2.1) and assume

$$E_\theta(|\delta(X) - \delta_0(X)|) > 0, \tag{2.2}$$

which contradicts the conclusion of the theorem.

We can assume without loss of generality that $\delta(X)$ is continuous due to the following argument. If $\delta(X)$ is not continuous, we can replace $\delta(X)$ by the continuous estimator $\bar{\delta}(X) = E(\delta(X - Z) + Z|X)$, where $Z \sim N(0, I)$ is independent of X. By Jensen's inequality and Fubini's theorem,

$$R(\theta, \bar{\delta}) = E_\theta L_c(|\bar{\delta}(X) - \theta|) \le E_\theta E(L_c(|\delta(X - Z) + Z - \theta|))|X$$
$$= ER(\theta - Z, \delta) \le ER(\theta - Z, \delta^0) = R(\theta, \delta^0).$$

By completeness, $\bar{\delta}(X) = \delta^0(X)$ (a.e.) holds if and only if $\delta(X) = \delta^0(X)$ a.e. Hence $\bar{\delta}$ satisfies (2.2).

Because of (2.2) there exists a point x_0 such that $\delta(x_0) \neq x_0$. Choose θ so that x_0 is between $\delta(x_0)$ and θ as in Figure 1. Obviously, $|\delta(x_0) - \theta| > |x_0 - \theta|$. Due to continuity of $\delta(X)$, there exists a positive number ε, $0 < \varepsilon < |\delta(x_0) - x_0|/2$, such that $|x - x_0| < \varepsilon$ implies

$$|\delta(x) - \theta| > |x_0 - \theta| + \varepsilon. \tag{2.3}$$

(ε can be chosen independent of θ.) Hence

$$E_\theta L_c(|\delta(X) - \theta|) \ge E_\theta \chi_{|X - x_0| < \varepsilon} L_c(|\delta(X) - \theta|). \tag{2.4}$$

Taking $c = |x_0 - \theta|$ and recalling that $|x - x_0| < \varepsilon$ implies (2.3), we note that (2.4) then implies

$$E_\theta L_c(|\delta(X) - \theta|) \ge \varepsilon P_\theta(|X - x_0| < \varepsilon).$$

(The choice of c is smaller than what was suggested by Figure 1.) Write, by using an orthogonal transformation,

$$P_\theta(|X - x_0| < \varepsilon) = P\left((Z_1 - c)^2 + \sum_{i=2}^{p} Z_i^2 < \varepsilon^2\right),$$

where the Z_i's are i.i.d. $N(0, 1)$. The probability is bounded below by

$$P(|Z_1 - c| + |Z_2| + \cdots + |Z_p| < \varepsilon)$$
$$\ge P\left(c - \frac{\varepsilon}{2} < Z_1 < c + \frac{\varepsilon}{2}, |Z_i| < \frac{\varepsilon}{2(p-1)}, i = 2, \ldots, p\right)$$
$$= \prod_{i=2}^{p} P\left(|Z_i| < \frac{\varepsilon}{2(p-1)}\right) P\left(c - \frac{\varepsilon}{2} < Z_1 < c + \frac{\varepsilon}{2}\right)$$
$$\equiv K_1 P\left(c - \frac{\varepsilon}{2} < Z_1 < c + \frac{\varepsilon}{2}\right),$$

where K_1 depends on ε but not on θ. For $c > \varepsilon/2$,

$$P\left(c - \frac{\varepsilon}{2} < Z_1 < c + \frac{\varepsilon}{2}\right) > P\left(c - \frac{\varepsilon}{2} < Z_1 < c - \frac{\varepsilon}{4}\right)$$
$$> \varepsilon\left(4\sqrt{2\pi}\right)^{-1} e^{-(c-\varepsilon/4)^2/2}.$$

In summary,

$$E_\theta L_c(|\delta(X) - \theta|) \geq \frac{K_1 \varepsilon}{4\sqrt{2\pi}} e^{-(c-\varepsilon/4)^2/2} \quad \text{as } c \to \infty. \tag{2.5}$$

Furthermore,

$$EL_c(|X - \theta|) = \Omega_p \int_c^\infty (r - c) r^{p-1} e^{-r^2/2}\, dr,$$

where Ω_p is a positive constant resulting from the spherical transformation. The last expression is bounded above by

$$K_2 c^{p+1} e^{-c^2/2} \quad \text{as } c \to \infty, \tag{2.6}$$

where K_2 is a positive constant independent of c. This and (2.5) imply that

$$\frac{E_\theta L_c(|\delta(X) - \theta|)}{EL_c(|X - \theta|)} \geq \frac{K_1 \varepsilon e^{-(c-\varepsilon/4)^2/2}}{4\sqrt{2\pi} K_2 c^{p+1} e^{-c^2/2}} \to \infty \quad \text{as } c \to \infty. \tag{2.7}$$

This contradicts (2.1), completing the proof of Theorem 2. □

PROOF OF THEOREM 1. Suppose δ is a (randomized) estimator which is as good as δ^0 under all losses of the form L_c: $0 \leq c < \infty$. Write $\delta = \delta(\cdot|X)$. Let $\delta'(x) = \int a\delta(da|x)$. Jensen's inequality then yields $R(\theta, \delta') \leq R(\theta, \delta^0)$. Furthermore, $E_\theta|\delta'(X) - \theta| < E_\theta|\delta^0(X) - \theta|$ unless

$$\delta\big(\{a: \exists\, \alpha = \alpha(X) \geq 0 \ni a = \theta + \alpha(\delta'(X) - \theta)\}|X\big) = 1 \quad \text{a.e.}$$

This condition holds for all θ if and only if δ is nonrandomized (a.e.). Hence, if δ is not equivalent to δ^0 then there is a nonrandomized estimator (δ') which satisfies (2.1) with strict inequality for some θ and some $c \geq 0$. This strict inequality means that δ' satisfies (2.2), thus contradicting the conclusion of Theorem 2. This contradiction establishes that $\delta = \delta^0$ (a.e.) and so completes the proof of Theorem 1. □

When Q is not necessarily I, we can also show that X is U-admissible in the lower dimensional case.

THEOREM 3. *Assume that Q is nonnegative definite. Then X is U-admissible if $p \leq 3$.*

PROOF. The only case that requires a proof is the situation where $p = 3$ and Q is positive definite. (All the other cases reduce to the two-dimensional or the one-dimensional case, in which X is admissible with respect to the quadratic loss

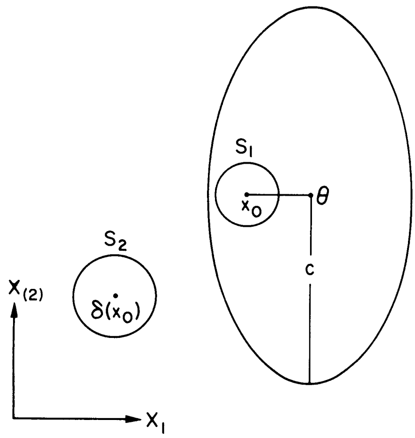

FIG. 2.

and hence is U-admissible. The detailed arguments are similar to the last paragraph of this proof.) Now we can assume without loss of generality that Q is a diagonal matrix with diagonal elements $q_1, q_2, \dots, q_p$, where $q_1 \geq q_2 \geq \cdots \geq q_p > 0$. Suppose $\delta(X)$ universally dominates X. Then the first component, $\delta_1(X)$, of $\delta(X)$ is identically X_1. The proof of this assertion is similar to the proofs of Theorem 2 and Theorem 1 and is omitted. It is based on the same type of heuristic argument, though using Figure 2 instead of Figure 1.

Note that Figure 2 is the same as Figure 1, except that the sphere that separates S_1 and S_2 in Figure 1 is now an ellipsoid defined by $\{x: |x - \theta|_Q = c\}$ for some appropriate constant so that S_1 and S_2 are separated. Furthermore the line joining θ and the center of S_1 is parallel to the X_1 axis.

Now consider the sum of squared error loss. Since $\delta_1(X) = X_1$, the problem is then reduced to the two-dimensional one. By fixing θ_1, one arrives at the conclusion that $(\delta_2(X), \delta_3(X))$ is as good as (X_2, X_3). Admissibility of (X_2, X_3) [as established in James and Stein (1960)] and strict convexity of the squared error loss imply that $\delta_2(X) = X_2$ and $\delta_3(X) = X_3$. This contradicts the fact that $\delta(X)$ universally dominates X, establishing the theorem. □

3. *U*-inadmissibility of the least squares estimator: $Q \neq I$. In this section we assume model (1.4). By an orthogonal transformation, Q can be assumed to be diagonal without loss of generality. Write $Q = \operatorname{diag}(q_1, q_2, \dots, q_p)$ and arrange so that the q_i's are decreasing in i. We assume that the largest q_i, i.e., q_1, is different from all the other q_i's. We can also rescale so that $q_2 = 1$. Consequently

$$q_1 > 1 = q_2 \geq q_3 \geq \cdots \geq q_p > 0.$$

We shall prove a general theorem (Theorem 6). When specialized to the case that q_i, $i \geq 2$, are identical, Theorem 6 implies that $\delta^0(X) = X$ is U-inadmissible if and only if $p \geq 4$ (Corollary 7).

Before we plunge into a rigorous proof of Corollary 7 (and more generally Theorem 6), it may help to indicate a calculation that led us to their discovery. Assume, for now, that $q_1 > q_2 = \cdots = q_p = 1$. Partition as

$$X^t = \left(X_1, X_{(2)}^t\right), \quad \text{where } X_{(2)}^t = \left(X_2, \ldots, X_p\right), \tag{3.1}$$
$$\theta^t = \left(\theta_1, \theta_{(2)}^t\right), \quad \text{where } \theta_{(2)}^t = \left(\theta_2, \ldots, \theta_p\right).$$

For a fixed Q, universal domination is equivalent to *stochastic domination*, i.e., simultaneous domination with respect to the step loss function $\chi_{(c,\infty)}(|\delta - \theta|_Q)$ for all $c > 0$ [Theorem 2.3 in Hwang (1985)]. Therefore a (slightly stronger) condition for universally dominating X is

$$P_\theta\left(|\delta(X) - \theta|_Q > c\right) < P\left(|X - \theta|_Q > c\right), \quad \forall\, \theta \text{ and } \forall\, c > 0. \tag{3.2}$$

Theorem 4.1 of Hwang (1985) indicates that, for any finite number c_0, there exists a James–Stein estimator that satisfies (3.2) simultaneously for all c, $0 < c < c_0$. More specifically, this is true for the estimator

$$\delta^t(X) = \left(X_1, \delta_2^t(X_{(2)})\right), \tag{3.3}$$

where δ_2 is a $(p - 1)$ dimensional James–Stein type estimator based only on $X_{(2)}$. Thus the difficulty in proving U-inadmissibility occurs as $c \to \infty$. It is exactly this case that stops X from being U-inadmissible for the symmetric loss in Section 2.

It seems natural to look at the conditional distribution of $X_{(2)}$ given $|X - \theta|_Q \geq c$. As $c \to \infty$, X will then stay very close to the boundary of the ellipsoid, $|X - \theta|_Q = c$ according to Birnbaum's argument (1955). Therefore, we study below the conditional distribution of $X_{(2)}$ given $|X - \theta|_Q = c$. After a lengthy but standard calculation one can show that the asymptotic distribution of $X_{(2)}$, given the ellipsoid, is a probability distribution! In fact, it has a $(p - 1)$-dimensional $N(\theta, q_1/(q_1 - 1)I)$ distribution.

Now, since $X_{(2)}$ has a limiting conditional distribution as $c \to \infty$, it is possible to construct a Stein type estimator (3.3) to improve upon X for large c. As we indicated earlier, by Hwang (1985) one can improve on X for small c's by using (3.3). A continuity argument therefore leads to the conjecture that for some $\delta(X)$ in (3.3), (3.2) holds for all $c > 0$; and hence X is U-inadmissible.

To prove the conjecture, we will need the following lemma which is a direct generalization of Theorem 3.3.2 in Brown (1966) to the situation involving a class of loss functions $W_c(\delta(X) - \theta)$, $c \in \mathscr{C}$. Throughout the paper, O and o are big "O" and little "o" as $a \to \infty$ uniformly in c, θ and positive b bounded away from 0 (a and b are introduced below). The major difference between this lemma and Brown's theorem is that the assumptions made and the conclusion (3.7) drawn there are uniformly in $c \in \mathscr{C}$, whereas Brown considered only one loss function. We omit the proof since it is similar to Brown's proof. [For details, see the Appendix in Brown and Hwang (1986).] The notation Sup_c denotes the supremum over $c \in \mathscr{C}$. Lemma 4 applies to the situation where X has a location probability density function $f(x - \theta)$, which is assumed to be twice continuously differentiable. Let

$$f'(x) = \left(\frac{\partial f_1}{\partial x_1}, \dots, \frac{\partial f_p}{\partial x_p} \right)$$

and let $f''(x)$ be a matrix whose (i, j)th element $f''_{ij}(x)$ is $(\partial^2/\partial x_i \, \partial x_j) f(x)$. We use E_0 to denote the expectation with respect to $f(x - \theta)$, with $\theta = 0$. Moreover, from now on and throughout the paper, E_θ and P_θ will be abbreviated as E and P, respectively.

LEMMA 4. *Assume that the observation X has a p-dimensional p.d.f. $f(x - \theta)$ with respect to Lebesgue measure. Let $\hat{W}_c(\cdot) = W_c(\cdot) - E_0 W_c(X)$. Assume that*

$$\sup_c \int |\hat{W}_c(X)| \, |X|^i |f'(X)| \, dX < \infty, \qquad i = 0, 1, 2, 3, 4, \tag{3.4}$$

and there exists a $\gamma > 0$ such that for all $|G(x)| < \gamma$,

$$\sup_c \int |\hat{W}_c(X)| \, |X|^i |f''_{jk}(X + G(X))| \, dX < \infty, \qquad i = 0, 1, \dots, 6, \; \forall j, k. \tag{3.5}$$

Also suppose that

$$\int f'(X) \hat{W}_c(X) \, dX = 0. \tag{3.6}$$

Let $\delta(X) = (I + B/(a + |X|^2))X$, where $B = b^{-1}B_0$ and B_0 is a constant matrix which does not depend on the positive scalars a and b. If $M_c B_0$ is nonnegative definite, where

$$M_c = \int \hat{W}_c(X) X f'(X) \, dX,$$

then

$$\begin{aligned} \Delta &\equiv E[W_c(X - \theta) - W_c(\delta(X) - \theta)] \\ &\geq \frac{1}{b(a + |\theta|^2)} \left(\frac{-2\theta^t M_c B_0 \theta}{|\theta|^2} + \operatorname{tr} M_c B_0 - \frac{Kp}{2b} \right) + o\left(\frac{1}{a + |\theta|^2} \right), \end{aligned} \tag{3.7}$$

where K is a positive finite number independent of a, b, c and θ.

REMARK. Typically M_c is negative definite, in which case B is negative definite and $\delta(x)$ is a James–Stein type shrinkage estimator.

Now we connect Lemma 4 to our problem. Here $X^t = (X_1, \dots, X_p) \sim N(\theta, I)$ and hence f is taken to be the p.d.f. of $N(0, I)$. We consider an estimator $\delta(X)$ as in (3.3). Since the first components of X and $\delta(X)$ are the same, it is only the choice of δ_2 that matters. By using Lemma 4, we will prove that there exists δ_2 for $\theta_{(2)}$ that has everywhere smaller risk than $X_{(2)}$ for every indicator or integral loss function $L_c^I(\delta_2 - \theta_{(2)})$, where

$$L_c^I(s) = \begin{cases} \chi_{(c,\infty)}(|s|_{Q_2}), & 0 < c \le 1, \\ \int_0^{|s|_{Q_2} \wedge c} t(1 - t^2/c^2)^{-1/2} e^{t^2/2q_1}\, dt, & c > 1. \end{cases} \tag{3.8}$$

Here $Q_2 = \text{diagonal}(q_2, \dots, q_p)$, $|\cdot|_{Q_2}$, as before, is the Q_2-generalized Euclidean distance and $a \wedge b$ denotes $\min(a, b)$ for real numbers a and b. The choice of L_c^I for $c > 1$ is inspired by the earlier heuristic argument. The integral relates to the conditional expectation of the indicator loss. The following lemma show that this implies U-inadmissibility.

LEMMA 5. *If $\delta_2(X_{(2)})$ has everywhere smaller risk than $X_{(2)}$ under L_c^I, $c > 0$, then* (3.2) *holds and $\delta(X)$ in* (3.3) *universally dominates X with respect to Q.*

PROOF. To establish (3.2) for every c, note that the left-hand side of (3.2) can be written as

$$P\left(q_1|X_1 - \theta_1|^2 + |\delta_2(X_{(2)}) - \theta_{(2)}|_{Q_2}^2 > c^2\right).$$

Writing $Z_1 = X_1 - \theta_1 \sim N(0, 1)$ and conditioning on Z_1, we can equate the last expression to

$$P\left(q_1|X_1 - \theta_1|^2 > c^2\right) + \int_0^c P\left(|\delta_2(X_{(2)}) - \theta_{(2)}|_{Q_2}^2 > t^2\right) f_R(t)\, dt, \tag{3.9}$$

where $f_R(\cdot)$ is the p.d.f. of $R = [(c^2 - q_1|X_1 - \theta_1|^2)^+]^{1/2}$ on the region $R > 0$. Hence

$$f_R(t) \propto t(1 - t^2/c^2)^{-1/2} e^{-(c^2 - t^2)/2q_1}, \qquad 0 < t < c.$$

Similarly, the right-hand side of (3.2) equals (3.9) except that $\delta_2(X_{(2)})$ is replaced by $X_{(2)}$. Note in (3.9), that the first term remains the same for X and $\delta(X)$. Hence in comparing X and $\delta(X)$, this term can be dropped. Now we focus on the second term, which is proportional to

$$\int_0^c P\left(|\delta_2(X_{(2)}) - \theta_{(2)}| > t\right) t(1 - t^2/c^2)^{-1/2} e^{t^2/2q_1}\, dt. \tag{3.10}$$

When $c \le 1$, the probability inside the integral is the risk function of $\delta_2(X_{(2)})$ with respect to L_c^I. Since by assumption the risk function of $\delta_2(X_{(2)})$ is smaller than that of $X_{(2)}$, (3.10) is smaller than the similar quantity with $\delta_2(X_{(2)})$ replaced by $X_{(2)}$, establishing (3.2) for $c \le 1$.

When $c > 1$, we apply Fubini's theorem to conclude that (3.10) equals

$$EL_c^I\left(\delta_c(X_{(2)}) - \theta_{(2)}\right),$$

where L_c^I is the integral loss in (3.8). Arguments similar to the last paragraph imply (3.2) for $c > 1$. □

Using this lemma as a bridge, we now proceed to argue the existence of $\delta_2(X_{(2)})$ dominating $X_{(2)}$ for every L_c^I. This is equivalent to domination with respect to the loss $W_c(\delta_2 - \theta_{(2)})$

$$W_c(\cdot) = L_c^I(\cdot)/\operatorname{tr} D_c, \tag{3.11}$$

where $\operatorname{tr} D_c$ is the trace of the matrix

$$D_c \equiv E_0 \hat{L}_c^I(X) X X^t, \tag{3.12}$$

and

$$\hat{L}_c^I(\cdot) = L_c^I(\cdot) - E_0 L_c^I(X). \tag{3.13}$$

It can be checked that D_c is diagonal and positive definite which implies the positivity of $\operatorname{tr} D_c$.

It is to this class $\{W_c(\cdot)\}$ that we will apply Lemma 4. In Appendix A, we show that all the assumptions of Lemma 4 are satisfied for $W_c(\cdot)$. Now we can state and prove the main theorem.

THEOREM 6. *Assume that*

$$Q = \operatorname{diagonal}(q_1, q_2, \dots, q_p)$$

and $q_1 > q_2 \geq \cdots \geq q_p > 0$ *and*

$$p > 2 \max_c \left\{ \frac{\max_i (D_c)_{(ii)}}{\min_i (D_c)_{(ii)}} \right\} + 1, \tag{3.14}$$

where $(D_c)_{(ii)}$ *is the* (i, i)*th element of the diagonal matrix* (3.12). *Then* X *is U-inadmissible with respect to Q-generalized Euclidean error.*

PROOF. We now search for δ_2 satisfying the conditions of Lemma 5. Then we apply Lemma 4, in which we identify X as $X_{(2)}$. Consequently, let

$$\delta_2(X_{(2)}) = \left(I + \frac{B}{a + |X_{(2)}|^2} \right) X_{(2)},$$

where

$$B = -b^{-1} I.$$

(That is, $B_0 = -I$ in Lemma 4.) Lemma 4 therefore gives (3.7). Here

$$\hat{W}_c(\cdot) = W_c(\cdot) - E_0 W_c(X) = \hat{L}_c^I(\cdot)/\operatorname{tr} D_c \tag{3.15}$$

and

$$M_c B_0 = \int \hat{W}_c(X) X f'(X)\, dX(-I) = \int \hat{W}_c(X) X X^t f(X)\, dX. \tag{3.16}$$

By (3.12) and (3.15) this equals $D_c/(\operatorname{tr} D_c)$. Hence $\operatorname{tr} M_c B_0 = 1$. Now (3.7) gives

$$\begin{aligned}
\Delta &\equiv E_0\Big(W_c\big(X_{(2)} - \theta_{(2)}\big) - W_c\big(\delta(X_{(2)}) - \theta_{(2)}\big)\Big) \\
&\geq \frac{1}{b\left(a + |\theta_{(2)}|^2\right)}\left(\frac{-2\theta_{(2)}^t D_c \theta_{(2)}}{|\theta_{(2)}|^2 \operatorname{tr} D_c} + 1 - \frac{K(p-1)}{2b}\right) + o\left(\frac{1}{a + |\theta_{(2)}|^2}\right) \\
&\geq \frac{1}{b\left(a + |\theta_{(2)}|^2\right)}\left(\left(\frac{-2}{p-1}\right)\max_c\left(\frac{\max_i (D_c)_{(i,i)}}{\min_i (D_c)_{(i,i)}}\right) + 1 - \frac{K(p-1)}{2b}\right) \\
&\quad + o\left(\frac{1}{a + |\theta_{(2)}|^2}\right),
\end{aligned}$$

which by (3.14) can be made positive for large a and b. Hence for sufficiently large a and b, $\delta_2(X_{(2)})$ has everywhere smaller risk than $X_{(2)}$ with respect to $\{W_c\}$. This and Lemma 5 establish the theorem. □

As a remark, the lower bound on the right-hand side of (3.14) is at least 3. It can be shown with the aid of results from the Appendix that this bound is finite. See Brown and Hwang (1986).

For the special case when

$$Q_2 = q_2 I,$$

we have the following corollary.

COROLLARY 7. *Assume that* $Q = \text{diagonal}(q_1, q_2, \ldots, q_p)$, *where* $q_1 > q_2 = q_3 = \cdots = q_p > 0$. *The estimator* X *is U-inadmissible with respect to Q-generalized Euclidean error if and only if* $p \geq 4$.

PROOF. Apply Theorem 6. In this case D_c is a multiple of the identity matrix. Hence condition (3.14) is equivalent to $p \geq 4$. Therefore, δ^0 is U-inadmissible for $p \geq 4$. For $p \leq 3$, Theorem 3 implies that X is U-admissible. □

APPENDIX A

Proof that $W_c(\cdot)$ defined in (3.8) and (3.11) satisfies all assumptions in Lemma 4. We first establish a key inequality (A.2). Consider $c > 1$. Let $u = |s|_{Q_2} \wedge c$. From (3.8),

$$\begin{aligned}
L_c^I(s) &\leq e^{u^2/2q_1}\int_0^u t\left(1 - t^2/c^2\right)^{-1/2} dt \\
&= e^{u^2/2q_1} u^2\left(1 + \left(1 - u^2/c^2\right)^{-1/2}\right)^{-1} \\
&\leq u^2 e^{u^2/2q_1}.
\end{aligned}$$

Since $q_i \le 1$ for all $i \ge 2$, $u \le |s|_{Q_2} \le |s|$,

$$L_c^I(s) \le |s|^2 e^{|s|^2/2q_1} \quad \text{for } c > 1. \tag{A.1}$$

Now for $0 < c \le 1$, $L_c^I(s)$ is bounded by 1 by the definition in (3.8). This, together with (A.1) and $q_1 > 1$, implies that

$$\sup_{0<c} E_0|X|^k L_c^I(|X|) < \infty \tag{A.2}$$

for every $k \ge 0$. This inequality is crucial in the following proofs.

Using the notation $\hat{W}_c$ of Lemma 4, we have here $\hat{W}_c(\cdot) = W_c(\cdot) - E_0 W_c(X) = \hat{L}_c^I(\cdot)/(\operatorname{tr} D_c)$, where $\hat{L}_c^I$ is defined in (3.13).

CASE 1 ($c > 1$). Condition (3.6) is equivalent to

$$\int f'(X)\hat{L}_c^I(X) = 0.$$

Since $f(X)$ is a normal p.d.f., the left-hand side equals

$$-E_0\big(L_c^I(X) - E_0 L_c^I(X)\big)X = -E_0 L_c^I(X)X = 0 \tag{A.3}$$

by symmetry. This establishes (3.6).

To check conditions (3.4) and (3.5), we first show that

$$\inf_{c>1} \operatorname{tr} D_c > 0, \tag{A.4}$$

where D_c, as defined in (3.12), equals $E_0 XX^t(L_c^I(X) - E_0 L_c^I(X))$. By symmetry, D_c is a diagonal matrix. In fact, the infimum over c of each diagonal element of D_c is positive. To see this, note that as an example the first element is

$$E_0 X_1^2\big(L_c^1(X) - E_0 L_c^I(X)\big) = E_0 X_1^2 L_c^I(X) - E_0 L_c^I(X). \tag{A.5}$$

Using inequality (A.1) and the dominated convergence theorem one can prove the continuity of the last expression in c, $1 < c < \infty$. Furthermore, for each $c > 1$, (A.5), being the covariance of two functions X_1^2 and $L_c^I(X)$ both increasing in $|X|$, is positive. As $c \to \infty$, (A.5) approaches (by the dominated convergence theorem)

$$E_0 X_1^2 L_\infty^I - E_0 L_\infty^I(X),$$

where

$$L_\infty^I(s) = \int_0^{|s|_{Q_2}} t e^{t^2/2q_1}\, dt = q_1\big(e^{|s|_{Q_2}/2q_1} - 1\big).$$

The limit is also positive. As $c \to 1$, a similar argument can be used to establish that (A.5) approaches a positive limit. Hence a continuity argument shows that the infimum of (A.5) over $c > 1$ is positive which implies (A.4). Therefore to prove conditions (3.4) and (3.5), it suffices to prove that they hold if $\hat{W}_c$ is replaced by $\hat{L}_c^I$. Now

$$
\begin{aligned}
\int |\hat{L}_c^I(X)|\,|X|^i|f'(X)|\,dX &= E_0|\hat{L}_c^I(X)|\,|X|^{i+1} \\
&\le E_0|L_c^I(X)|\,|X|^{i+1} + \left(E_0|X|^{i+1}\right)E_0|L_c^I(X)|.
\end{aligned} \tag{A.6}
$$

This and (A.2) establish condition (3.4).

Condition (3.5) can be established if one can show

$$\sup_c \int |\hat{L}_c^I(X)|\,|X|^{i+2}e^{-|X+G(X)|^2/2}\,dX < \infty$$

for every $|G(X)| < \gamma$. The left-hand side is less than or equal to

$$\sup_c \int \left(|L_c^I(X)| + E_0 L_c^I(X)\right)|X|^{i+2}e^{-|X|^2/2+\gamma|X|}\,dX,$$

which is finite by arguments similar to (A.6) and a strengthened (A.2) with $e^{\gamma|X|}$ in the expectation. All three conditions are now established.

CASE 2 ($0 < c \le 1$). Condition (3.6) holds similarly to (A.3). Conditions (3.4) and (3.5) also hold for any fixed $c > 0$ by arguments similar to those for Case 1. However, to establish (3.4) and (3.5), we still have to show as $c \to 0$, that the relevant integrals approach a positive number. This together with the fact that these integrals are continuous in $c > 0$ imply (3.4) and (3.5).

To deal with the case $c \to 0$, in the proof, we will now use O or o to denote an expression's order as $c \to 0$. Furthermore, $O_+(c^k)$ will denote a quantity such that as $c \to 0$,

$$O_+(c^k)/c^k \to \text{a positive finite number.}$$

Now, from (3.8) and (3.13), we get

$$\hat{L}_c^I(s) = -\chi_{[0,c]}\left(|s|_{Q_2}\right) + E_0\chi_{[0,c]}\left(|s|_{Q_2}\right).$$

Therefore, by (3.12),

$$D_c = -E_0\left[\chi_{[0,c]}\left(|X|_{Q_2}\right)XX^t\right] + E_0\left[\chi_{[0,c]}\left(|X|_{Q_2}\right)\right]I,$$

which is a diagonal matrix with the ith element on the main diagonal being

$$
\begin{aligned}
(D_c)_{ii} &= -E_0\chi_{[0,c]}\left(|X|_{Q_2}\right)X_i^2 + E_0\chi_{[0,c]}\left(|X|_{Q_2}\right) \\
&= -O(c^{p+1}) + O_+(c^{p-1}) = O_+(c^{p-1}).
\end{aligned} \tag{A.7}
$$

In deriving (A.7) as well as other formulas in the proof, one should keep in mind that X denotes $X_{(2)}$ which is $(p-1)$-dimensional. Thus from (A.7),

$$\hat{W}_c(s) = \frac{-\chi_{[0,c]}\left(|s|_{Q_q}\right)}{\operatorname{tr} D_c} + \frac{O(c^{p-1})}{O_+(c^{p-1})}.$$

We now deal with condition (3.4). Note that

$$\int |\hat{W}_c(X)|\, |X|^i |f'(X)|\, dX = E_0 |\hat{W}_c(X)| |X|^{i+1} \le E\frac{\chi_{[0,c]}(|X|_{Q_2})|X|^{i+1}}{O_+(c^{p-1})} + O(1)$$

$$= O(1).$$

Hence

$$\limsup_{c\to 0} \int |\hat{W}_c(X)|\, |X|^i |f'(X)|\, dX < \infty,$$

establishing (3.4) as $c \to 0$ and hence for $0 < c \le 1$.

For (3.5), note that

$$\int |\hat{W}_c(X)|\, |X|^i |f''_{jk}(X + G(X))|\, dX$$

$$\le \int |\hat{W}_c(X)|\, |X|^i |X + G(X)|^2 f(X + G(X))\, dX$$

$$\le 2\int |\hat{W}_c(X)|\, |X|^i (|X|^2 + \gamma^2) f(X + G(X))\, dX.$$

The upper bound approaches a finite number if for every $k \ge 0$ so does the expression

$$\int |\hat{W}_c(X)|\, |X|^k f(X + G(X))\, dX.$$

The last expression is bounded above by

$$\frac{e^{\gamma c}}{O_+(c^{p-1})} E_0(\chi_{[0,c]}(|X|)) + O(1) \sup_{|G(X)| < \tau} \int |X|^k f(X + G(X))\, dX = O(1)$$

as $c \to 0$. Hence,

$$\limsup_{c\to 0} \int \hat{W}_c(X) |X|^i |f''_{jk}(X + G(X)|\, dX < \infty,$$

establishing (3.5) as $c \to 0$ and hence for $0 < c \le 1$. The proof is now complete.

REFERENCES

BERGER, J. (1979). Multivariate estimation with nonsymmetric loss functions. In *Optimizing Methods in Statistics* (J. S. Rustagi, ed.). Academic, New York.

BHATTACHARYA, P. K. (1966). Estimating the mean of a multivariate normal population with general quadratic loss function. *Ann. Math. Statist.* **37** 1819–1824.

BIRNBAUM, A. (1955). Characterizations of complete classes of tests of some multiparametric hypotheses, with applications to likelihood ratio tests. *Ann. Math. Statist.* **26** 21–36.

BROWN, L. D. (1966). On the admissibility of invariant estimators of one or more location parameters. *Ann. Math. Statist.* **37** 1087–1136.

BROWN, L. D. and FOX, M. (1974). Admissibility of procedures in two-dimensional location parameter problems. *Ann. Statist.* **2** 248–266.

BROWN, L. D and HWANG, J. T. (1986). Universal domination and stochastic domination: *U*-admissibility and *U*-inadmissibility of the least squares estimator. Technical Report, Cornell Statistical Center.

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. (1960). Estimation with quadratic loss. *Proc. Fourth Berkeley Symp. Math. Statist. Probab.* **1** 361–379. Univ. California Press.

RAO, C. R. (1976). Estimation of parameters in a linear model. *Ann. Statist.* **4** 1023–1037.

SHINOZAKI, N. (1975). A study of generalized inverse of matrix and estimation with quadratic loss. Ph.D. dissertation, Keio Univ., Japan.

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.

WALD, A. (1950). *Statistical Decision Functions*. Wiley, New York.

DEPARTMENT OF MATHEMATICS
WHITE HALL
CORNELL UNIVERSITY
ITHACA, NEW YORK 14853