The Annals of Statistics
1985, Vol. 13, No. 1, 295–314

UNIVERSAL DOMINATION AND STOCHASTIC DOMINATION: ESTIMATION SIMULTANEOUSLY UNDER A BROAD CLASS OF LOSS FUNCTIONS

BY JIUNN TZON HWANG[1]

Cornell University

In Wald's statistical decision theory, the criterion of domination (or uniform betterness) is defined with respect to a specific loss. In practice, however, the exact form of a loss function is difficult to specify. Hence, it is important to study the domination criterion simultaneously under a class of loss functions. In this paper we focus on estimation problems. We mainly investigate the possibility of domination simultaneously under the class of loss functions $L(|\theta - \delta|)$, where L is an arbitrary nondecreasing function. As usual, θ and δ (both in p-dimensional Euclidean space R^p) are, respectively, the unknown parameter of nature and the statistician's estimate. Domination simultaneously under this class of losses is called universal domination under Euclidean error.

Several theoretical questions are resolved in this paper. In particular the criterion of universal domination is shown to be equivalent to the criterion of stochastic domination that compares the estimators by the stochastic ordering of their Euclidean distances from the estimators to the true parameter.

Concrete results about universal domination relating to the usual estimator are also established. In particular when $X - \theta$ has a p-variate t distribution, and $p = 1, 2$, there exists no estimator for θ that universally dominates X; however, for $p \geq 3$, estimators (of the type of James-Stein positive part estimators) that universally dominate X are specified. When X has a p-variate normal distribution with mean θ and identity covariance matrix, we show that for any dimension p, no James-Stein positive part estimators universally dominate X. However, under slightly smaller classes of losses, some James-Stein positive part estimators are shown to simultaneously dominate X. These hitherto unstudied losses are bounded and fairly practical.

1. Introduction.

In comparing statistical decision rules, many statisticians (including Wald 1950, page 26) have proposed a domination criterion with respect to a particular loss function. A loss function $L(\theta, a)$ represents the amount by which a statistician is penalized when θ is the state of nature, and a is the statistician's action. The decision rule δ_1 is said to be as good as δ_2 if for every θ

$$E_\theta L(\theta, \delta_1(X)) \leq E_\theta L(\theta, \delta_2(X)), \tag{1.1}$$

where X is the statistician's observation and has a distribution characterized by θ. If δ_1 is as good as δ_2 and for some θ the inequality in (1.1) is strict, δ_1 is said to dominate (or to be uniformly better than) δ_2. The expectations in (1.1) are functions of θ and are called risk functions.

Received August 1983; revised August 1984.

[1] Research supported by the National Science Foundation Grant Number MCS-8301139.

AMS 1980 *subject classifications.* Primary 62C05, 62H12; secondary 62J07.

Key words and phrases. Decision theory, estimation under a class of losses, multivariate normal distribution, multivariate t distribution.

In practice, it is difficult to specify the loss function exactly. Hence it is valuable to know how robust a particular domination is with respect to the loss function it assumes, that is, over how large a class of loss functions it may hold. Even though many results have been established for domination under a particular loss, there is relatively much less work that has been done concerning a class of losses. The Rao-Blackwell Theorem, for an estimator not a function of a sufficient statistic, provides an estimator based on the sufficient statistic which is uniformly better simultaneously under the class of strict convex loss functions. Brown (1975) and Shinozaki (1980) dealt with the problem of improving upon the intuitive estimator for a normal mean simultaneously under a class of quadratic losses with variable weights. Ghosh and Auer (1983) also established similar results for exponential families. Two other related articles will also be cited in the next paragraph.

In this paper, we mainly focus on the problem of estimating a certain unknown quantity θ in R^p, the p-dimensional Euclidean space. An estimator δ_1 is said to universally dominate another (nonrandomized) estimator δ_2 under Euclidean error if for any loss $L(|\theta - \delta|)$, $L(\cdot)$ nondecreasing, δ_1 is as good as δ_2; and for one such loss, δ_1 dominates δ_2. More generally, one can similarly define universal domination under generalized Euclidean error with respect to a nonnegative definite matrix D by replacing, in the previous sentence, the Euclidean error $|\theta - \delta|$ by the generalized Euclidean error $|\theta - \delta|_D = [(\theta - \delta)^t D(\theta - \delta)]^{1/2}$. (For the general definition, see Definition 2.1 in Section 2.) Note that any reasonable loss based on $|\theta - \delta|$ (or $|\theta - \delta|_D$) should have a nondecreasing L so that a statistician would not be penalized less when his estimate δ is further away from θ. Therefore, all the reasonable loss functions based on a specific generalized Euclidean error are taken into consideration. This type of domination is hence "universal" in L. For one-parameter monotone likelihood families, Brown, Cohen and Strawdermann (1976) exhibited some complete classes (which consist of monotone procedures or procedures based on a sufficient statistic) for universal domination essentially. Rukhin (1978) discussed the prior distribution with respect to which the "universal" Bayes estimator exists, i.e., with respect to which it remains Bayes under all nondecreasing loss functions $L(|\theta - a|)$. Here we are dealing with expected loss rather than Bayes risk.

Under the generalized Euclidean error with respect to D, it is shown in Section 2 that universal domination is equivalent to another criterion called stochastic domination; i.e., δ_1 stochastically dominates δ_2 if roughly speaking $|\delta_1(X) - \theta|_D$ is stochastically smaller than $|\delta_2(X) - \theta|_D$ for all θ. (The precise definition is given in Section 2.) Therefore, stochastic domination implies domination for a large class of loss functions.

The key question concerning these two equivalent criteria of universal domination and stochastic domination is whether these criteria can reasonably distinguish estimators. In many situations, both criteria fail to distinguish two estimators. In particular, this is the case when the domination criterion under a particular loss fails to do so. For example, any two estimators admissible with respect to a particular loss cannot be compared under the domination criterion with respect to this particular loss. However, there are also many situations (discussed in Section 3) in which universal domination does occur frequently. In a linear model with a spherically distributed error, the least squares estimator is shown to universally dominate any other linear unbiased estimator with respect to any generalized Euclidean error. (An analog of the Gauss-Markov theorem.) When the unknown parameter is a p-component vector and $p = 1, 2$, it is shown that if the least squares estimator has finite fourth moments, then it is U-admissible with respect to any generalized Euclidean error, that is, there exists no other estimator that universally dominates it. When $p \geq 3$ and when the error has a spherical t distribution with arbitrary degrees of freedom, the least squares estimator is shown to be universally dominated by a modified James-Stein positive part estimator under some generalized Euclidean error. This linear model has been considered by Zellner (1976) and Thomas (1970) as a generalization of the normal linear model; since as the degrees of freedom approach infinity, the multivariate t distribution approaches the multivariate normal distribution.

For the normal case, assume without too much loss of generality that $X \sim N(\theta, I)$. Even though the normal distribution can be approximated by t distributions, we show in Section 3 that for any dimension p no James-Stein positive part estimators universally dominate (under Euclidean error) the intuitive estimator X. This phenomenon is somewhat surprising in view of the fact that Brown (1966) has shown that under practically any bounded loss function, there are estimators of the form similar to James-Stein positive part estimators that dominate X when $p \geq 3$.

However, for a slightly smaller class of losses (which consists of all the nondecreasing losses $L(|\theta - \delta|)$ that remain constant when $|\theta - \delta| > c$ for some fixed c), a class of James-Stein positive part estimators is shown to dominate X. One of these James-Stein positive part estimators has been shown by Hwang and Casella (1982a) to substantially improve upon X under a particular loss. According to their exact numerical calculations, the maximum reduction in risk is at least 60% when $p \geq 5$. These results are reported in Section 4. Section 5 consists of some unsolved problems raised by this work.

Section 2. Universal domination and stochastic domination: Definition and connection. Assume that X be a multidimensional vector whose probability density function (p.d.f.) with respect to a σ-finite measure u is $f_\theta(x)$. The parameter $\theta \in R^p$ is a p-dimensional unknown vector and is the quantity that one wants to estimate. We will focus on nonrandomized estimators which are u-measurable functions from the sample space to R^p. However, definitions and theorems discussed in this paper can be extended to randomized estimators. Consider a loss function $L(|\theta - \delta(x)|_D)$ where $L(\cdot)$ is nondecreasing and

$$|\theta - \delta(x)|_D = [(\theta - \delta(x))^t D(\theta - \delta(x))]^{1/2}$$

is the generalized Euclidean error with respect to D. The matrix D is assumed to be a known nonnegative definite matrix and is also assumed without loss of generality to be symmetric. When $D = I$, the generalized Euclidean error is reduced to Euclidean error and is denoted by $|\theta - \delta(x)|$. The risk function of an estimator $\delta(X)$

$$R_L(\theta, \delta) = E_\theta L(|\theta - \delta(X)|_D) = \int L(|\theta - \delta(x)|_D) f_\theta(x)\, du, \tag{2.1}$$

is always defined, even though it may be positively infinite.

DEFINITION 2.1. An estimator δ_1 universally dominates δ_2 (under the generalized Euclidean error with respect to D) if for every θ and every nondecreasing loss function L

$$R_L(\theta, \delta_1) \le R_L(\theta, \delta_2)$$

and for a particular loss the risk functions are not identical.

Next we discuss the notion of stochastic domination. In what follows, for any two random variables Y and Z, $Y \le_d Z$ denotes that Y is stochastically less than or equal to Z, i.e., for every real number c

$$P(Y \ge c) \le P(Z \ge c).$$

Similarly $Y =_d Z$ (or $Y \ne_d Z$) represents that Y and Z have identical (or nonidentical) distributions. Further $Y <_d Z$ iff $Y \le_d Z$ and $Y \ne_d Z$. Now we state the definition of stochastic domination.

DEFINITION 2.2. A (nonrandomized) estimator δ_1 stochastically dominates δ_2 under the generalized Euclidean error with respect to D if for every θ

$$|\delta_1(X) - \theta|_D \le_d |\delta_2(X) - \theta|_D$$

and for some θ $|\delta_1(X) - \theta|_D <_d |\delta_2(X) - \theta|_D$.

The notion of stochastic domination has attracted broad attention among statisticians. Pitman (1938) defined a similar criterion under fiducial distributions. (See also Mood, Graybill and Boes, 1974, page 289.) Pitman called an estimator the best (and argued that it is undeniably "the best" on page 401) if it has a stochastically smallest Euclidean error under the fiducial distribution. Our definition is, however, based on the sampling distribution. In estimating the mean of a normal sample, he then showed that the sample average is "the best" estimator. Later in the Bayesian context, the results are also extended by Rukhin (1978 and 1984), who characterized the distribution of X and the generalized prior distribution for which the generalized Bayes rule has a stochastically smallest Euclidean error with respect to the posterior distribution.

Savage (1954, criterion 3 on page 224) introduced a criterion which is even stronger than stochastic domination. Our definition seems, however, to be the natural modification of his one-dimensional criterion for higher dimensional problems.

Cohen and Sackrowitz (1970) had an example concerning stochastic domination. Assuming that X_1 and X_2 are two normally distributed random observations with means θ_1 and θ_2 and with known variances, they studied the problem of estimating θ_1. If it is known that $\theta_1 > \theta_2$, their Theorem 6.1 asserts that X_1 can be stochastically dominated. This result is surprisingly strong.

We now establish the equivalence between universal domination and stochastic domination.

THEOREM 2.3. (The equivalence theorem). *Under the generalized Euclidean error with respect to a nonnegative definite matrix D,* δ_1 *universally dominates* δ_2 *if and only if* δ_1 *stochastically dominates* δ_2.

The proofs of Theorem 2.3 and many other theorems in the sequel are based on the technical lemma below.

LEMMA 2.4. (i) *Suppose that* $Y_1 \leq_d Y_2$ *are two real valued random variables whose expectations exist. Then* $EY_1 \leq EY_2$. (ii) *Assume in addition to the assumptions in* (i) *that* $|EY_1| < \infty$. *Then* $Y_1 <_d Y_2$ *if and only if* $EY_1 < EY_2$.

PROOF. Statement (i) follows from, say, Lemma 1 on page 73 in Lehmann (1959). The "if" part in statement (ii) is trivial since if $Y_1 =_d Y_2$ then $EY_1 = EY_2$. The "only if" part can be proved by considering the contrapositive statement, and by modifying the proof of (i). □

Now we return to the

PROOF OF THEOREM 2.3. "If" part. Now $|\delta_1(X) - \theta|_D \leq_d |\delta_2(X) - \theta|_D$ for every θ. Hence $L(|\delta_1(X) - \theta|_D) \leq_d L(|\delta_2(X) - \theta|_D)$ for any nondecreasing function L. This, together with Lemma 2.4, implies that $R_L(\theta, \delta_1) \leq R_L(\theta, \delta_2)$. Further $|\delta_1(X) - \theta|_D <_d |\delta_2(X) - \theta|_D$ for some θ_0. Let L_0 be a strictly increasing bounded function, then $L_0(|\delta_1(X) - \theta_0|_D) <_d L_0(|\delta_2(X) - \theta_0|_D)$ which implies, by Lemma 2.4, that $R_{L_0}(\theta_0, \delta_1) < R_{L_0}(\theta_0, \delta_2)$. Hence δ_1 universally dominates δ_2.

To prove the "only if" part let $L(|\theta - \delta|_D) = 1$ if $|\theta - \delta|_D > c$ and 0 otherwise. By assumption, for every c, $R_L(\theta, \delta_1) \leq R_L(\theta, \delta_2)$ which is equivalent to $|\delta_1(X) - \theta|_D \leq_d |\delta_2(X) - \theta|_D$. To show $|\delta_1(X) - \theta|_D <_d |\delta_2(X) - \theta|_D$ for some θ, assume to the contrary $|\delta_1(X) - \theta|_D =_d |\delta_2(X) - \theta|_D$ for every θ. Then δ_1 and δ_2 would always have the same risk functions which contradicts the hypothesis that δ_1 universally dominates δ_2. □

Now we turn to the discussion of admissibility with respect to universal domination (or stochastic domination). Under a generalized Euclidean error with respect to D, a (nonrandomized) estimator $\delta(X)$ is called U-admissible with respect to D if there exists no other estimator that universally dominates $\delta(X)$. Otherwise $\delta(X)$ is called U-inadmissible with respect to D. A sufficient condition for U-admissibility is provided below.

THEOREM 2.5. *If* $\delta(X)$ *is admissible with respect to a particular loss* $L_0(|\theta - \delta|_D)$ *where D is a nonnegative definite matrix,* L_0 *is strictly increasing and* $R_{L_0}(\theta, \delta) < \infty$ *for every* θ, *then* $\delta(X)$ *is U-admissible with respect to D. Equivalently, if* $\delta(X)$ *is U-inadmissible with respect to D, then* $\delta(X)$ *is inadmissible under any strictly increasing loss* $L_0(|\theta - \delta|_D)$ *such that* $R_{L_0}(\theta, \delta) < \infty$ *for every* θ.

Theorem 2.5 will follow from the following Lemma whose proof is provided below.

LEMMA 2.6. *If $\delta^*(X)$ universally dominates $\delta(X)$ under a generalized Euclidean error with respect to D, then $\delta^*(X)$ dominates $\delta(X)$ under any strictly increasing loss $L_0(|\theta - \delta|_D)$ for which $R_{L_0}(\theta, \delta) < \infty$ for all θ.*

PROOF. Theorem 2.3 implies that $|\delta^*(X) - \theta|_D \leq_d |\delta(X) - \theta|_D$ for all θ and $|\delta^*(X) - \theta|_D <_d |\delta(X) - \theta|_D$ for some θ. Therefore the same conclusion holds when comparing $L_0(|\delta^*(X) - \theta|)$ to $L_0(|\delta(X) - \theta|)$. Lemma 2.4 then implies that $\delta^*(X)$ dominates $\delta(X)$ under L_0. □

In Theorem 2.5, the assumption that $R_{L_0}(\theta, \delta)$ is finite for all θ is sometimes implied by the admissibility of δ and in which case the assumption can be deleted. For example, when θ is a one-dimensional positive parameter and X has monotone likelihood ratio, then it can be shown that with respect to the squared error loss, all admissible estimates have finite risk functions for every θ. (Special cases are the estimation of a Poisson mean and the estimation of the noncentrality parameter of a chi-squared random observation.) However, we do not know, in general, whether this assumption can be removed.

One might ask if L_0 in Theorem 2.5 is assumed to be nondecreasing, will the same statement hold? The answer is obviously no, since one can take L_0 to be a constant loss in which case any estimator is admissible, even though many estimators are U-inadmissible as will be shown in Section 3. Aside from the trivial case that L_0 is a constant, if one takes L_0 to be nonconstant and nondecreasing, one might still ask the same question. The answer is again negative as shown in the following counterexample.

EXAMPLE 1. To be concrete, let us take a particular p.d.f.

$$f(x) = 1 - |x|, \quad |x| \leq 1.$$

Hence f has a unimode at zero. Assume that X is a one-dimensional random variable with p.d.f. $f(x - \theta)$. Compare two estimators X and $X + 0.1$. It is straightforward to show that, under Euclidean error, X universally dominates $X + 0.1$ and hence $X + 0.1$ is U-inadmissible. However if one considers the loss function $L_0(|\delta - \theta|)$, where $L_0(t) = 0$ or 1 depending on $t \leq 1.1$ or $t > 1.1$, then the risk function of $X + 0.1$ is zero for all θ and the estimator is admissible with respect to L_0. □

Example 1 can obviously be generalized to any spherically symmetric unimodal distribution with bounded support. Does the converse of Theorem 2.5 hold? Namely, does U-admissibility imply admissibility with respect to a strictly increasing loss function? The answer is no as shown in the example below.

EXAMPLE 2. Let $X - \theta$ be a one-dimensional t distributed random variable with 5 degrees of freedom. (Its p.d.f. is given in the first equation of (3.6) with N

replaced by 5). It is shown in Theorem 3.3 that X is U-admissible. However, X is inadmissible with respect to $L_0(|\delta - \theta|)$, when $L_0(t) = t^6$, since it has infinite risk for all θ. □

One question remains unanswered, however. Does U-admissibility of $\delta(X)$ imply that $\delta(X)$ is admissible under a strictly increasing L_0 with respect to which $\delta(X)$ has finite risk for every θ?

Section 3. Concrete results concerning universal domination. There are obviously questions concerning how useful is the criterion of universal domination in distinguishing estimators. Below we establish some universal domination results for specific distributions. The first result is an analog of the Gauss Markov Theorem. We assume the linear model

$$X = A\theta + \varepsilon, \quad \varepsilon \sim_{\text{p.d.f.}} |\textstyle\sum| f_m(\varepsilon^t \textstyle\sum^{-1} \varepsilon), \tag{3.1}$$

where X is the observation vector, A is a known $m \times p$ matrix of rank p, $\theta \in R^p$ is an unknown parameter that one wants to estimate, ε is an unobserved error term with elliptical distribution whose p.d.f. with respect to a σ finite measure is specified in (3.1), and $\sum$ is a positive definite symmetric matrix. Below unbiasedness refers to either expectation unbiasedness or median unbiasedness. (An estimator $(\delta_1(X), \cdots, \delta_p(X))$ is median unbiased for $(\theta_1, \cdots, \theta_p)$ if for $i = 1, 2, \cdots, p$, θ_i is a median of $\delta_i(X)$.) When referring to expectation unbiasedness, we assume the expectation of ε exists (and hence is the zero vector by symmetry).

THEOREM 3.1. *Assume that $\sum = \sigma^2 I$ where I is an identity matrix and σ^2 is (in general) unknown. Then the least square estimator $\hat{\theta}^{\text{LS}} = (A^tA)^{-1}A^tX$ universally dominates (under any generalized Euclidean error with respect to a positive definite matrix D) any other linear unbiased estimator.*

PROOF. Below we assume without loss of generality that ε is not identically zero. Otherwise the theorem is trivial, since $\hat{\theta}^{\text{LS}}$ is then the only unbiased estimator. By Theorem 2.3, we need only show that $\hat{\theta}^{\text{LS}}$ stochastically dominates any other linear unbiased estimator.

We can also assume without loss of generality that (3.1) has a canonical form, namely, the first p rows of A form a nonsingular matrix A_1 and the last $(m - p)$ rows are zero rows. (From an arbitrary matrix A, one can obtain this representation by multiplying on both sides of (3.1) by an orthogonal matrix $O = (O_1, \cdots, O_m)^t$ where O_i, $i = 1, \cdots, p$ are p m-dimensional column vectors that form an orthonormal basis for the column space of A; and O_i, $i = 1, \cdots, m$, form an othonormal basis for R^m. Such an orthogonal transformation will not change the distribution of ε.)

Now let $(\lambda_1, \cdots, \lambda_p, 0, \cdots, 0)^t = A\theta$. We first show that

$$\hat{\lambda}^{\text{LS}} =_{\text{def'n.}} (X_1, \cdots, X_p)^t$$

stochastically dominates any other linear unbiased estimator of $\lambda = (\lambda_1, \cdots, \lambda_p)^t$.

Let MX be an arbitrary unbiased estimator of λ. Unbiasedness implies that M can be decomposed as $[I, M_2]$ where I is the $p \times p$ identity matrix and M_2 is some $p \times (m-p)$ matrix. Note that the estimator $\hat{\lambda}^{\mathrm{LS}}$ is of the form MX with $M_2 = 0$. Therefore to complete the proof, it is sufficient to show that $(MX-\lambda)^t D(MX-\lambda)$ is uniquely minimized (in the sense of stochastic ordering) by letting $M_2 = 0$. Note

$$(MX - \lambda)^t D(MX - \lambda) =_d (M\varepsilon)^t DM\varepsilon. \tag{3.2}$$

Let us assume without loss of generality that D is symmetric. By Lord (1954) or Kelker (1970) the characteristic function $Ee^{is^t\varepsilon}$ is of form $\psi(s^t s)$ for some function ψ. Using this characteristic function and standard arguments, one can show that

$$(D^{1/2}MM^tD^{1/2})^{1/2}\varepsilon_p =_d D^{1/2}M\varepsilon, \tag{3.3}$$

where ε_p is the vector that consists of the first p components of ε. Hence

$$(MX - \lambda)^t D(MX - \lambda) =_d \varepsilon_p^t D^{1/2}MM^tD^{1/2}\varepsilon_p.$$

Now recalling that $M = [I, M_2]$, we have $MM^t = I + M_2M_2^t$ and consequently

$$\varepsilon_p^t(D^{1/2}MM^tD^{1/2})\varepsilon_p = |\varepsilon_p|_D^2 + \varepsilon_p^t(D^{1/2}M_2M_2^tD^{1/2})\varepsilon_p.$$

The last expression is clearly minimized by choosing $M_2 = 0$. Therefore $\hat{\lambda}^{\mathrm{LS}}$ is at least as good as any other linear unbiased estimator (in the sense of stochastic domination). However, we will next argue that $\hat{\lambda}^{\mathrm{LS}}$ actually stochastically dominates any other linear unbiased estimator.

If $\varepsilon_p^t D^{1/2}M_2M_2^tD^{1/2}\ \varepsilon_p$ is identically zero, then for any orthogonal matrix Q, $\varepsilon_p^t Q^t D^{1/2}M_2M_2^tD^{1/2}Q\ \varepsilon_p$ is also identically zero (due to the fact that ε_p and $Q\ \varepsilon_p$ have the same distribution). Choosing Q to diagonize the matrix $D^{1/2}M_2M_2^tD^{1/2}$ and using the fact that ε_p is not identically zero, one shows that all the eigenvalues of the matrix are zero. This implies that the matrix is a zero matrix. Consequently, by the fact that D is nonsingular, M_2 is a zero matrix. Therefore if $M_2 \neq 0$, then

$$|\varepsilon_p|_D^2 + \varepsilon_p^t D^{1/2}M_2M_2^tD^{1/2}\varepsilon_p >_d |\varepsilon_p|_D^2.$$

This shows that $\hat{\lambda}^{\mathrm{LS}}$ universally dominates any other linear unbiased estimator.

Returning to the proof of the theorem, we now show that $\hat{\theta}^{\mathrm{LS}}$ stochastically dominates any other linear unbiased estimator. Previous arguments have shown that for all nonsingular matrices D and all linear unbiased estimators $MX \neq \hat{\lambda}^{\mathrm{LS}}$

$$(\hat{\lambda}^{\mathrm{LS}} - \lambda)^t D(\hat{\lambda}^{\mathrm{LS}} - \lambda) <_d (MX - \lambda)^t D(MX - \lambda). \tag{3.4}$$

Using the identities $\hat{\lambda}^{\mathrm{LS}} = A_1\hat{\theta}^{\mathrm{LS}}$ and $\lambda = A_1\theta$ (recall that A_1 is the largest nonsingular submatrix of A), (3.4) is equivalent to

$$(\hat{\theta}^{\mathrm{LS}} - \theta)^t A_1^t DA_1(\hat{\theta}^{\mathrm{LS}} - \theta) <_d (A_1^{-1}MX - \theta)^t A_1^t DA_1(A_1^{-1}MX - \theta). \tag{3.5}$$

Note that any linear unbiased estimator of θ, say M_0X, can be expressed as $A_1^{-1}MX$ so that MX is unbiased for λ (by simply letting $M = A_1M_0$). Moreover, $\hat{\theta}^{\mathrm{LS}} \neq A_1^{-1}MX$ if and only if $\hat{\lambda}^{\mathrm{LS}} \neq MX$. Hence (3.5) implies that $\hat{\theta}^{\mathrm{LS}}$ stochastically dominates any other linear unbiased estimator. □

The conclusion of Theorem 3.1 is obviously stronger than that of a standard Gauss Markov theorem. The assumptions of Theorem 3.1 are usually (but not always) stronger than what is needed for the Gauss Markov theorem which assumes only the existence of second moments of ε. However, if ε has an elliptical distribution with infinite second moment (e.g., ε has a Cauchy distribution), then the Gauss Markov theorem fails while Theorem 3.1 still applies. One can generalize Theorem 3.1 to the case of estimating a linear function of θ and to the situation involving a more general covariance matrix.

COROLLARY 3.2. *Assume that $\sum = \sigma^2 \sum_0$ where $\sum_0$ is known and σ^2 is (in general) unknown. Consider the problem of estimating $B\theta$ (B a known matrix) under the generalized Euclidean error with respect to a positive definite matrix. The estimator $B\hat{\theta}^{\mathrm{LS}}$, where $\hat{\theta}^{\mathrm{LS}} = (A^t \sum_0^{-1} A)^{-1}A^t \sum_0^{-1} \underset{\sim}{X}$ is the generalized least squares estimator, universally dominates any other linear unbiased estimator.*

PROOF. Multiplying both sides of (3.1) by $\sum_0^{-1/2}$ reduces the model to that considered in Theorem 3.1. Therefore we can assume $\sum_0 = I$ without loss of generality. Still this corollary is more general than Theorem 3.1 because B is an arbitrary matrix, not necessarily the identity. To establish this result for an arbitrary B, one can follow the proof of Theorem 3.1, except for the following modification and its consequential changes. When decomposing M (where recall that MX is an arbitrary unbiased estimator), one obtains, $M = [B, M_2]$, from the unbiasedness assumption, instead of $[I, M_2]$ as in the proof of Theorem 3.1. □

An independent result which is similar to (but weaker than) Corollary 3.2 was also established by Ali and Ponnapalli (1983). We now turn to the problem concerning the U-admissibility of the least squares estimator in Corollary 3.2.

THEOREM 3.3. *Under Model (3.1), assume that $p = 1$ or 2 and that the least squares estimator $\hat{\theta}^{\mathrm{LS}}$ has finite fourth moments. Then $\hat{\theta}^{\mathrm{LS}}$ is admissible under the quadratic loss $(\theta - \delta)^t D(\theta - \delta)$ where D is an arbitrary fixed nonnegative definite matrix. Hence by Theorem 2.5, $\hat{\theta}^{\mathrm{LS}}$ is U-admissible with respect to D.*

PROOF. Without loss of generality, assume that $\sum = \sigma^2 I$. It suffices to show that $\hat{\theta}^{\mathrm{LS}}$ is admissible under the squared error loss, since then by Shinozaki's theorem (appeared in Shinozaki, 1975, as well as in Lemma 3.1 of Rao, 1976), $\hat{\theta}^{\mathrm{LS}}$ is admissible with respect to the quadratic loss $(\delta - \theta)^t D(\delta - \theta)$ for any nonnegative definite matrix D. It is also sufficient to show that $\hat{\theta}^{\mathrm{LS}}$ is admissible for any fixed known σ^2, since $\hat{\theta}^{\mathrm{LS}}$ does not depend on σ^2.

Now under a spherically symmetric distribution, the statistic $(Y_1, Y_2) = (\hat{\theta}^{\mathrm{LS}}, |X - A\hat{\theta}^{\mathrm{LS}}|^2)$ is sufficient for θ. Therefore under the squared error loss, one only need focus on the nonrandomized estimator based on (Y_1, Y_2). The problem is obviously invariant under the location transformations $\theta \to \theta + a$ and $Y_1 \to Y_1 + a$. The corresponding invariant estimator δ thus satisfies $\delta(Y_1 + a, Y_2) = \delta(Y_1, Y_2) + a$. Since Y_2 has a distribution independent of θ, the best invariant estimator is therefore the generalized Bayes estimator with respect to a uniform prior in θ. Therefore $\hat{\theta}^{\text{LS}}$ is the best invariant estimator which, by Brown and Fox (1974), is then admissible when the fourth moments are finite. □

The conditions of Theorem 3.3 are satisfied in many situations, including those where ε has a normal distribution, a t distribution with N degrees of freedom ($N > 4$) or a double exponential distribution. The probability density functions of the last two distributions are given respectively by

$$f(|\varepsilon|^2) = \text{constant}(1 + N^{-1}|\varepsilon|^2)^{-(N+p)/2}, \quad \text{and} \qquad (3.6)$$

$$f(|\varepsilon|^2) = \text{constant } e^{-k|\varepsilon|},$$

where $k > 0$ is some fixed constant.

Now we turn to the situations when $p \geq 3$. In this case, the determination of U-admissibility is very difficult. Recall that, for a particular loss function, the unique best invariant estimator is usually inadmissible when $p \geq 3$. For the normal case and the sum of squared error loss, this is the well-known result of Stein (1956). For general location problems, these were established in Brown (1966). So far, we are able to obtain some results for the case when σ^2 is known. (For this case σ can be taken to be one by a multiplicative transformation.) For some distributions (including t-distributions), we show in Corollary 3.8 and Theorem 3.9 that the least squares estimator is U-inadmissible. For the normal distribution, we are unable to determine the U-admissibility of the least square estimator. However, for the special case assumed in Theorem 3.10, no James-Stein positive part estimator δ_a in (3.7) universally dominates the usual estimator of the norm mean.

Before considering the linear model, we first deal with the simpler case when the p dimensional observation X has mean θ. We will apply Theorem 2.1 of Hwang and Chen (1983) which is quoted below for convenience. Let $\delta_a(X)$ be the James-Stein positive part estimator (1960), i.e.,

$$\delta_a(X) = (1 - a/|X|^2)_+ X \quad \text{where} \quad y_+ = \text{Max}\{y, 0\}. \qquad (3.7)$$

THEOREM 3.4. (Hwang and Chen, 1983). *Assume that* $X \sim_{\text{p.d.f.}} f(|x - \theta|^2)$ *and* $f'(s)/f(s)$ *is defined for all* s, $\alpha_0 \leq s \leq \alpha_1$, *where for some* $c > 0$ *and* $a > 0$,

$$\alpha_0 = (c - \sqrt{a})_+^2, \quad \text{and} \quad \alpha_1 = c^2 + a.$$

If

$$\inf_{\alpha_0 < s < \alpha_1} \frac{f'(s)}{f(s)} \geq \frac{-(p-2)a^{-1/2}}{2c} \ln\{[c + (c^2 + a)^{1/2}]/a^{1/2}\}, \qquad (3.8)$$

then for every θ,

$$P(|\theta - \delta_a(X)| \leq c) > P(|\theta - X| \leq c). \qquad (3.9)$$

Since $\delta_a(X)$ would stochastically (or equivalently universally) dominate X if (3.9) were satisfied for all $c > 0$, we have the following corollaries.

COROLLARY 3.5. *If, in addition to the assumptions of Theorem* 3.4, *there exists* *a fixed $a > 0$ such that* (3.8) *holds for all $c > 0$, then $\delta_a(X)$ universally dominates X under the Euclidean error.*

COROLLARY 3.6. *Assume that $f'(s)/f(s)$ is a continuous real valued function. If there exists $a = a_0 > 0$ such that for c sufficiently large and for c sufficiently close to zero* (3.8) *holds, then X is U-inadmissible under Euclidean error.*

PROOF. Let $a(c)$ be the supremum over all $a > 0$ that satisfies (3.8). (Such a exists by the argument in the next paragraph.) Since the left-hand side of (3.8) is decreasing in a and the right-hand side is increasing in a, (3.8) holds for all $a \le a(c)$. The goal here is to establish that $\inf_{0<c<\infty} a(c) > 0$. This would imply, by Corollary 3.5, that $\delta_a(X)$ universally dominates X for all a, $0 < a \le \inf_{0<c<\infty} a(c)$.

Note that as a decreases to zero, the left-hand side of (3.8) increases and the right-hand side decreases to $-\infty$, so that $a(c)$ must be positive. Now $a(c)$ is a continuous function in c, and as $c \to 0$ or $c \to \infty$, $a(c) \ge a_0 > 0$. Therefore, $\inf_{0<c<\infty} a(c) > 0$. □

Now we apply these general theorems to specific distributions.

THEOREM 3.7. *Assume that $X - \theta$ has a p-variate t distribution with N degrees of freedom whose p.d.f. is given in* (3.6). *For every N and $p \ge 3$, X is U-inadmissible under the Euclidean error. Furthermore, under the Euclidean error, the James-Stein positive part estimator $\delta_a(X)$ universally dominates X if $a > 0$ satisfies*

$$\frac{N+p}{N} \le \frac{p-2}{(N+a)^{1/2}a^{1/2}} \ln\{[(N+a)^{1/2} + (N+2a)^{1/2}]/a^{1/2}\}. \tag{3.10}$$

PROOF. In this case, condition (3.8) reduces to

$$\frac{N+p}{N+(c-a^{1/2})_+^2} \le \frac{p-2}{ca^{1/2}} \ln\{[c + (c^2+a)^{1/2}]/a^{1/2}\}. \tag{3.11}$$

To find a so that the last inequality is satisfied for all c, we consider two separate cases: (i) $c^2 \le N + a$ and (ii) $c^2 \ge N + a$.

For case (i), (3.11) is satisfied if

$$\frac{N+p}{N} \le \frac{p-2}{ca^{1/2}} \ln[(c + (c^2+a)^{1/2})/a^{1/2}]. \tag{3.12}$$

The derivative of the right-hand side of the last inequality with respect to c is

$$\frac{p-2}{a^{1/2}} \frac{1}{c^2} \left[\frac{c}{(c^2+a)^{1/2}} - \ln(c + (c^2+a)^{1/2}) + \ln a^{1/2} \right]$$

which is negative since, by the Mean Value theorem,

$$\ln[c + (c^2+a)^{1/2}] - \ln a^{1/2} > \inf_{0<t<c} c(t^2+a)^{-1/2} = c(c^2+a)^{-1/2}.$$

Therefore (3.12) is satisfied for all c, $c^2 \le N + a$ if it is satisfied for $c^2 = N + a$, which is equivalent to (3.10).

For case (ii), (3.11) is equivalent to

$$\frac{N+p}{N} \le \frac{(p-2)}{ca^{1/2}}\left(1 + \frac{(c-a^{1/2})^2}{N}\right)\ln\{[c + (c^2+a)^{1/2}]/a^{1/2}\}. \tag{3.13}$$

Direct differentiation shows that $[1 + (c - a^{1/2})^2/N]/c$ is increasing in c if $c^2 \ge N + a$. This implies that the right-hand side of (3.13) is increasing in c, when $c^2 \ge N + a$. Therefore (3.13) is satisfied for $c^2 \ge N + a$ if and only if it is satisfied for $c^2 = N + a$. However, when $c^2 = N + a$, (3.13) follows from (3.12) which is equivalent to (3.10). Therefore the proof is complete. □

It is clear that there exists an $a^* > 0$ so that for every a, $0 < a \le a^*$, (3.10) is satisfied. Therefore Theorem 3.7 is constructive. However, the values of a^* which can be found by using a programmable calculator are small and are not reported here.

Now consider the linear model $X = A\theta + \varepsilon$, where A is an $m \times p$ known design matrix with full rank p and ε/σ (when σ is known) has a t distribution with N degrees of freedom. (This is a special case of model (3.1).) Note that this model is a generalization of the usual linear model with normal error in the sense that, as N approaches infinity, the t distribution approaches a normal distribution. For this model, we establish the following result.

COROLLARY 3.8. *For every N and $p \ge 3$, the least squares estimator $\hat{\theta}^{\text{LS}} = (A^tA)^{-1}A^tX$ is U-inadmissible and is universally dominated by*

$$\delta(X) = \left(1 - \frac{a\sigma^2}{|(A^tA)^{1/2}\hat{\theta}^{\text{LS}}|^2}\right)_+ \hat{\theta}^{\text{LS}},$$

under the generalized Euclidean error with respect to the matrix A^tA, where a satisfies (3.10).

PROOF. This follows directly from Theorem 3.7 and the fact that $(A^tA)^{1/2}\hat{\theta}^{\text{LS}}/\sigma$ has a p-variate t distribution. □

When we have n replicates, we can apply Corollary 3.6 to establish U-inadmissibility results.

THEOREM 3.9. *Assume that we have n independent observations X_i,*

$$X_i = A\theta + \varepsilon_i, \quad i = 1, 2, \cdots, n$$

where A, θ are the same as in Corollary 3.8 and ε_i are i.i.d. p-variate ($p \ge 3$) t distributed random variables with N degrees of freedom. Then the least squares estimator for θ, $\hat{\theta} = 1/n \sum_{i=1}^{n} (A^tA)^{-1}A^tX_i$ is U-inadmissible under the generalized Euclidean error with respect to the matrix A^tA.

PROOF. By a linear transformation (similar to the one in Corollary 3.8) corresponding to the matrix $(A^tA)^{1/2}$, one observes that the theorem is equivalent to the following simpler assertion. Assume that $X_i - \theta$ are independently identically distributed according to a p-variate ($p \geq 3$) t distribution. Then $\bar{X} = 1/n \sum_{i=1}^{n} X_i$ is a U-inadmissible estimator for θ with respect to the Euclidean error.

To prove the simpler version, it suffices to show (and will be shown below) that $S = \sum_{i=1}^{n} X_i$ is a U-inadmissible estimator for $n\theta$ with respect to the Euclidean error. Although S has a spherical distribution, its p.d.f. $g(|s - n\theta|^2)$ can not be calculated explicitly (except for the case when the degree of freedom is one). However, it does have the following representation

$$g(u) = \int f(|u^{1/2}e_1 - x_2 \cdots - x_n|^2) f(|x_2|^2) \cdots f(|x_n|^2)\, dx_2 \cdots dx_n \tag{3.14}$$

where f is given in (3.6), $e_1 = (1, 0, \cdots, 0)^t$, and $x_2, \cdots, x_n$ are all p-dimensional vectors. To apply Corollary 3.6, we have to show that there exists an $a > 0$ such that, for c sufficiently large or sufficiently small (close to zero), (3.8) holds or in this case

$$\inf_{(c-\sqrt{a})_+^2 < u < c^2 + a} \frac{g'(u)}{g(u)} \geq \frac{-(p-2)}{2c(a)^{1/2}} \ln[(c + (c^2 + a)^{1/2})/a^{1/2}], \tag{3.15}$$

which is equivalent to

$$c \inf_{(c-\sqrt{a})_+^2 < u < c^2 + a} \frac{g'(u)}{g(u)} \geq \frac{-(p-2)}{2a^{1/2}} \ln[c + (c^2 + a)^{1/2}/a^{1/2}]. \tag{3.16}$$

We consider two separate cases: (i) $c \to \infty$, (ii) $c \to 0$.

CASE (i), $c \to \infty$. Note that the right-hand side of (3.16) approaches $-\infty$ as $c \to \infty$. If we can show that

$$g'(u)/g(u) \geq -Mu^{-1/2}, \quad \forall\, u > 0, \tag{3.17}$$

for some finite positive constant M independent of c and a, then the left-hand side of (3.16) is bounded below by $-cM/(c - a^{1/2})_+$. This approaches $-M$ as c approaches infinity. Hence (3.17) implies that for some (and in fact for every) $a > 0$ (3.16) holds under case (i) $c \to \infty$. In the following derivations, interchanging the order of differentiation and integration and passing the limit inside the integration are allowed by the Bounded Convergence theorem and by the facts that f' and f'' are uniformly bounded and the function $f(|x_2|^2) \cdots f(|x_n|^2)$ integrates to one.

Now direct calculations using (3.14), we have

$$u^{1/2} \frac{g'(u)}{g(u)} = \frac{\int D^t e_1 f'(|D|^2) f(|x_2|^2) \cdots f(|x_n|^2)\, dx_2 \cdots dx_n}{\int f(|D|^2) f(|x_2|^2) \cdots f(|x_n|^2)\, dx_2 \cdots dx_n} \tag{3.18}$$

where $D = u^{1/2}e_1 - x_2 - \cdots - x_n$. Further straightforward calculations establish that

$$f'(|D|^2) = -(N + p) f(|D|^2)/[2(N + |D|^2)],$$

and

$$\frac{-(N+p)}{2}\frac{D^te_1}{N+|D|^2} \geq \frac{-(N+p)}{2}\frac{|D|}{N+|D|^2} \geq \frac{-(N+p)}{2}\frac{\sqrt{N}}{2N}.$$

These together with (3.18) show that

$$u^{1/2}g'(u)/g(u) \geq \frac{-(N+p)}{4\sqrt{N}} \quad \text{for every} \quad u > 0.$$

Therefore (3.17) is established and we have shown that for every fixed $a > 0$, (3.16) is satisfied for sufficiently large c.

CASE (ii), $c \to 0$. We first derive a representation of $g'(u)/g(u)$ as u approaches zero. From (3.18) and some algebraic calculations, we have

$$\begin{aligned} &\lim_{u\to 0}\frac{g'(u)}{g(u)} \\ (3.19) \quad &= H_0^{-1}\int f'(|x_2+\cdots+x_n|^2)f(|x_2|^2)\cdots f(|x_n|^2)\,dx_2\cdots dx_n \\ &\quad - H_0^{-1}\lim_{u\to 0}\frac{1}{\sqrt{u}}\int (x_2+\cdots+x_n)^te_1f'(|D|^2)f(|x_2|^2) \\ &\qquad\qquad \cdots f(|x_n|^2)\,dx_2\cdots dx_n \end{aligned}$$

where

$$H_0 = \int f(|x_2+\cdots+x_n|^2)f(|x_2|^2)\cdots f(|x_n|^2)\,dx_2\cdots dx_n.$$

It is straightforward to check that

$$\lim_{u\to 0}\int (x_2+\cdots+x_n)^te_1f'(|D|^2)f(|x_2|^2)\cdots f(|x_n|^2)\,dx_2\cdots dx_n = 0$$

by taking the limit inside the integral. (Again interchanging the order is allowed by the Bounded Convergence theorem and the fact that $(x_2+\cdots+x_n)^te_1f'(|D|^2)$ is uniformly bounded.) Hence one can apply L'Hospital's rule to the second term on the right-hand side of equation (3.19), and establish that this term equals the finite quantity

$$\begin{aligned} H_1 =_{\text{def'n}} 2H_0^{-1}\int &[(x_2+\cdots+x_n)^te_1]^2f''(|x_2+\cdots+x_n|^2)f(|x_2|^2) \\ &\cdots f(|x_n|^2)\,dx_2\cdots dx_n. \end{aligned}$$

Hence, we have shown that

$$\begin{aligned} (3.20) \quad \lim_{u\to 0}\frac{g'(u)}{g(u)} = H_0^{-1}\int &f'(|x_2+\cdots+x_n|^2)f(|x_2|^2) \\ &\cdots f(|x_n|^2)\,dx_2\cdots dx_n - H_1, \end{aligned}$$

and the limit is finite. To finish the proof that as $c^2 \to 0$ there exists an $a > 0$

such that (3.15) holds, we look at those c, for which $c^2 < a$. Note that

$$\inf_{0<u<c^2+a} \frac{g'(u)}{g(u)} \to \inf_{0<u\le a} \frac{g'(u)}{g(u)} \quad \text{as} \quad c \to 0,$$

and $\inf_{0<u\le a} g'(u)/g(u)$ increases to the finite quantity in (3.20) as a decreases to zero. Furthermore, direct calculation shows that the right-hand side of (3.15) approaches $-(p-2)/2a$ as $c \to 0$. The limit then approaches $-\infty$ as $a \to 0$. These observations therefore imply that there exists an $a > 0$ such that for sufficiently small c, (3.15) holds. Case (ii) is established and the proof is now complete. □

When X has a p-variate normal distribution with mean θ and covariance matrix I, there are questions concerning the U-admissibility of X. If $p = 1$ or 2, Theorem 3.3. implies that X is U-admissible. When $p \ge 3$, Brown (1966) has shown that under any specific nonconstant loss function $L(|\theta - \delta|)$ so that X has finite risk function for all θ, X is inadmissible and is dominated by estimators having a form similar to a James-Stein positive part estimator δ_a (in (3.7)) for some sufficiently small a. This does not, however, imply that there exists an a such that $\delta_a(X)$ universally dominates X. Hence, the question concerning the U-admissibility of X remains. We do not yet know the answer to this question. However, we do know that none of the estimators (3.7) universally dominates X as shown in the following theorem.

THEOREM 3.10. *Assume X has $N(\theta, I)$ distribution. For any dimension p and for Euclidean error, δ_a, as in* (3.7), *does not universally dominate X no matter what a is.*

PROOF. Obviously, we only need to consider the case where $p \ge 3$ and $a > 0$. To prove the theorem, it suffices to show that for every fixed $a > 0$,

$$P(|\theta - \delta_a(X)| \le c) < P(|\theta - X| \le c) \tag{3.21}$$

for some $c > 0$ and some θ. (Hence $\delta_a(X)$ fails to stochastically dominate X.) Let B be the event that X satisfies both inequalities within $P(\ )$. After tedious calculations, one can show that, as $c \to \infty$,

$$\left. \frac{P(|\theta - \delta_a(X)| \le c) - P(B)}{P(|\theta - X| \le c) - P(B)} \right|_{|\theta|=c^+} \le \text{constant } c^{p-1}(e^{c\sqrt{a}/2} - 1)^{-1}.$$

(For details, see Hwang 1984, pages 26–28). This upper bound clearly approaches zero as $c \to \infty$. Hence (3.21) is established. □

Section 4. Simultaneous domination under a smaller class of loss functions when the sampling distribution is the multivariate normal distribution. Assume that $X \sim N(\theta, I)$. We have shown in the last section that every James-Stein positive part estimators δ_a (as in (3.7)) fails to universally dominate X under the Euclidean error. In a sense, this is a negative result. In this section, we succeed in proving that some δ^a can dominate X simultaneously under a smaller class of loss functions, $\mathscr{L}_c$. The class consists of all nonconstant loss functions L_c of the form

$$L_c(|\theta - \delta|) = \ell(|\theta - \delta|) \quad \text{if} \quad |\theta - \delta| \le c$$
$$= K \qquad\qquad \text{if} \quad |\theta - \delta| > c$$

where c is a fixed constant, ℓ is an arbitrary nondecreasing function, and K is an arbitrary number for which $L_c(\cdot)$ is nondecreasing. Therefore $\mathscr{L}_c$ consists of nondecreasing and nonconstant losses $L(|\theta - \delta|)$ where $L(t)$ stays constant when $t > c$.

THEOREM 4.1. *The James-Stein positive part estimator $\delta_a(X)$ dominates X simultaneously under all the loss $L_c \in \mathscr{L}_c$, provided that $a > 0$ and*

$$1 \le (p-2)\ln\{[c + (c^2 + a)^{1/2}]/a^{1/2}\}/(ca^{1/2}). \tag{4.1}$$

The last conditions are equivalent to $a \in (0, a^*]$ where a^* is the unique positive solution to (4.1) *with the inequality replaced by an equality. Furthermore, unless L_c is the trivial loss $L_c^T(t) = K_1$ or K_2, depending on whether $t = 0$ or $t > 0$, δ_a has a risk function smaller than X for all θ.*

PROOF. Consider two cases: (i) L_c is the trivial loss L_c^T; (ii) L_c is not L_c^T. For case (i), the risk of δ is $K \cdot P(|\theta - \delta(X)| > 0)$. Hence, the risk of X is K and the risk of $\delta_a(X)$ is $K \cdot P_0(|X|^2 > a)$ when $\theta = 0$ and K if $\theta \ne 0$. Therefore $\delta_a(X)$ dominates X.

For case (ii), we note first that since L_c is monotonic, the points of discontinuity of L_c are countable. Hence we can construct a sequence $t_1, \cdots, t_n, \cdots$ with different t_i's, $0 \le t_i \le c$, so that it contains all the discontinuous points. (In doing so, both the cases of infinitely many and finitely many discontinuities are unified.) Write

$$L_c(t) = L_c^*(t) + \sum_{n=1}^{\infty} (L(t_n^+) - L(t_n)) I_{(t_n,\infty)}(t)$$
$$+ \sum_{n=1}^{\infty} (L(t_n) - L(t_n^-)) I_{[t_n,\infty)}(t)$$

where $I_A(t)$ is the indicator function on A, i.e., $I_A(t) = 1$ or 0 depending on whether $t \in A$ or $t \notin A$ and L_c^* is a continuous nondecreasing function. (In the last equation, $L(0^-)$ is interpreted as $L(0)$.) Let us assume without loss of generality that $L_c^*(0) = 0$. (Otherwise, we can subtract $L^*(0)$ from both sides without changing the problem.) Hence we can write $L_c^*(t)$ as a Lebesgue-Stieltjes integral

$$L_c^*(t) = \int I_{(u,\infty)}(t)\, dL_c^*(u) = \int_{0<u<c} I_{(u,\infty)}(t)\, dL_c^*(u)$$

where the last equation follows, from the fact that $L_c(t)$ (and hence $L_c^*(t)$) remains constant for $t > c$. Now using the last two equations and Fubini's theorem, we obtain that

$$
\begin{aligned}
EL_c(|\delta(X) - \theta|) &= \int_{0<u<c} P(|\delta(X) - \theta| > u)\, dL_c^*(u) \\
&\quad + \textstyle\sum_{n=1}^{\infty} (L(t_n^+) - L(t_n))P(|\delta(X) - \theta| > t_n) \\
&\quad + \textstyle\sum_{n=1}^{\infty} [L(t_n) - L(t_n^-)]P(|\delta(X) - \theta| \geq t_n).
\end{aligned} \tag{4.2}
$$

Since L_c is not L_c^T, using (4.2) we could show that for all θ the risk function of $\delta_a(x)$ is strictly less than that of X, provided that we could prove for all θ and for all u, $0 < u \leq c$,

$$P(|\delta_a(X) - \theta| > u) < P(|X - \theta| > u), \tag{4.3}$$

and

$$P(|\delta_a(X) - \theta| \geq u) < P(|X - \theta| \geq u). \tag{4.4}$$

This can be shown to hold by applying Theorem 3.4 as follows. For this normal case, Theorem 3.4 implies that for every θ (4.3) holds, provided that

$$1 \leq \frac{(p-2)}{ua^{1/2}} \ln\{[u + (u^2 + a)^{1/2}]/a^{1/2}\}. \tag{4.5}$$

Furthermore, it can be shown that $P(|\delta_a(X) - \theta| > u)$ is continuous at all u such that $u \neq |\theta|$ and hence for such u, $P(|\delta_a(X) - \theta| > u) = P(|\delta_a(X) - \theta| \geq u)$. Therefore, (4.5) implies (4.4) for $u \neq |\theta|$. Now for $u = |\theta|$, (4.4) can be shown to be a consequence of (4.5) by using the proof of Theorem 2.1 in Hwang and Casella (1982b). Therefore in both cases (4.3) and (4.4) hold provided that (4.5) can be established.

To complete the proof, all we need to do is to show that (4.1) implies (4.5) for all u, $0 < u \leq c$. This could be accomplished if we could show that the right-hand side of (4.5) is decreasing in u. Similar to what we dealt with the right-hand side of (3.12), the derivative of the right-hand side of (4.5) with respect to u can be shown to be negative. The proof is now complete. □

For each c^2, the value of a^* can be easily found by iteratively using a programmable calculator. To give the reader some ideas as to how big a^* can be for a particular c^2, we report in Table 1 the values of a^* for selected c^2 and p. Using these James-Stein positive part estimators, the improvement in risk over the usual estimator can be quite substantial. Hwang and Casella (1982a) numerically calculated the risk of the James-Stein estimator for several a, $a \in (0, a^*]$, under the zero one loss function with turning point occurring at c. Their findings are startling in that the maximum saving in risk of the James-Stein positive part over the usual one is at least 60% when $p \geq 5$. See Table 1 in Hwang and Casella (1982a). Note that the risks are one minus the quantities they reported.

Although they focus on a single loss at a time, the results of Brandwein and Strawderman (1980) and Bock (1983) can apply simultaneously to a class of loss

TABLE 1
Upper bounds a^* for simultaneous domination

p	3	4	5	6	7	8	9	10
c^2	6.25	7.78	9.24	10.7	12.0	13.4	14.7	16.0
a^*	0.58	1.34	2.13	2.94	3.76	4.59	5.41	6.25
p	11	12	13	14	15	16	17	18
c^2	17.3	18.6	19.8	21.0	22.3	23.5	24.8	26.0
a^*	7.08	7.92	8.75	9.59	10.4	11.3	12.1	13.0
p	19	20	21	22	23	24	25	
c^2	27.2	28.4	29.6	30.8	32.0	33.2	34.4	
a^*	13.8	14.7	15.5	16.4	17.2	18.0	18.9	

functions. Their losses have to be a concave loss function of $|\theta - \delta|^2$ which may or may not be the case in practice and which could be difficult to decide. Our losses for this normal case are not exhaustive either. However, our losses can accurately approximate any bounded nondecreasing loss (which occurs very frequently in practice). Such approximation, however, will sometimes require using a very large c which yields a very small a^* and very small improvement over the usual estimator. Therefore Theorem 4.1 offers significant improvement only for the situation where moderate c has been chosen.

Section 5. Comments and conclusions. We have established in this paper that, in many situations, domination can occur simultaneously under a large class of losses. In most higher dimensional situations considered, we prove that the usual estimator can be improved simultaneously under all the nondecreasing loss functions based on the Euclidean error.

However, this paper also raised many unanswered questions. In particular, for the case $X \sim N(\theta, I)$, is the usual estimator X U-admissible for estimating θ having at least three components? Even if X is U-admissible, it would be of practical value to consider a smaller class of losses. For a wide class of losses, James-Stein positive part estimators were shown here to dominate X. However, the class of estimators exhibited here is not at all exhaustive and the optimal choice among these estimators is not discussed. It seems to be obvious that the larger is the class of losses considered, the smaller the improvement over X would be. However, the relationship between the class of losses and the improvement was not quantitatively described.

For the other distributions (primarily t-distributions), we have found estimators that universally dominate the least squares estimator when the dimension of the unknown parameter is at least 3. These estimators (namely, James-Stein positive part estimators) are not expected to perform substantially better than the least squares estimator. Therefore, the search for estimators that both universally dominate and substantially improve upon the least squares estimator is of great interest. Even if no such estimator exists, consideration of smaller class of losses (similar to what was dealt in Section 4) can lead to fruitful results.

Finally, if one simultaneously considers all the generalized Euclidean errors, then the situation is very different. Interesting domination simultaneously under all the losses based on all the generalized Euclidean errors becomes rare. In this case, an estimator is admissible if componentwise it is admissible under the corresponding one-dimensional squared error loss. See Brown (1975).

Acknowledgement. The author thanks Professor Lawrence D. Brown who, under the author's requests, has constantly explained his work and many other previous works related to this paper with an amazing clarity. Many of Professor Brown's criticisms about this paper also resulted in improvements over the author's previous draft. The author thanks Professors Robert Berk, George Casella, and Arthur Cohen for pointing out many related articles which appear in the references. He is also indebted to Professors James O. Berger, M. E. Bock, Roger Farrell, Iain Johnstone, Andrew Rukhin, Harold Sackrowitz, and William Strawderman for their comments and encouragements. The detailed editorial comments of Mark Rentmeesters are also appreciated.

Note Added in Proof. Theorem 3.1 does not follow directly from the sufficiency of $\hat{\theta}^{LS}$ and $|X - A\hat{\theta}^{LS}|$, because, in the argument, one gets involved with randomized estimators which are difficult to handle since the loss functions are not necessarily convex. Our proof uses a different approach based on the canonical form of a linear model and direct maximization to get by such difficulties.

REFERENCES

ALI, M. M. and PONNAPALLI, R. (1983). An optimum property of Gauss-Markoff estimate when errors are elliptically distributed. Tech. Rept., Univ. of Western Ontario, London, Canada. Presented in Cincinnati Annual Statistics meeting, Aug. 1982.

BOCK, M. E. (1983). Minimax estimators that shift towards a hypersphere for location vectors of spherically symmetric distributions. Tech. Rept. #81-51. Statist. Dept. Purdue U.

BRANDWEIN, A. C., and STRAWDERMAN, W. E. (1980). Minimax estimation of location parameters for spherically symmetric distributions with concave loss. *Ann. Statist.* **8** 279–284.

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. (1975). Estimation with incompletely specified loss function (the case of several location parameters). *J. Amer. Statist. Assoc.* **70** 417–427.

BROWN, L. D., COHEN, A., and STRAWDERMAN, W. E. (1976) A complete class theorem for strict monotone likelihood ratio with applications. *Ann. Statist.* **4** 712–722.

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

COHEN, A. and SACKROWITZ, H. B. (1970). Estimation of the last mean of a monotone sequence. *Ann. Math. Statist.* **41** 2021–2034.

GHOSH, M. and AUER, R. (1983). Simultaneous estimation of parameters in exponential families. *Ann. Inst. Stat. Math.* **35A**, 379–387.

HWANG, J. T. (1984). Universal domination and stochastic domination—estimation simultaneously under a broad class of loss functions. Cornell Statistical Center Tech. Rept. First revision.

HWANG, J. T. and CASELLA, G. (1982a). Minimax confidence sets for the mean of a multivariate normal distribution. *Ann. Statist.* **10** 868–881.

HWANG, J. T. AND CASELLA, G. (1982b). Improved set estimators for a multivariable normal mean. Statistical Center Technical Reports No. 82-3, Cornell University. To appear in *Statistics and Decisions.*

HWANG, J. T. and CHEN, J. (1983). How do the improved confidence sets for the mean of a multivariate normal distribution perform when the true distribution is nonnormal? Statistical Center Technical Reports No. 83-13. Cornell University. To appear in *Ann. Statist.*

JAMES, W. and STEIN, C. (1960). Estimation with quadratic loss. *Fourth Berkeley Symp. Math Statist. Probab.* **1** 361–379. Univ. of California Press.

KELKER, D. (1970). Distribution theory of spherical distribution and a location-scale parameter generalization. *Sankhya A* **32** 319–330.

LEHMANN, E. L. (1959). *Testing Statistical Hypotheses.* Wiley, New York.

LORD, R. D. (1954). The use of the Hankel transformation in statistics. I. general theory and examples. *Biometrika* **41** 44–55.

MOOD, A. M., GRAYBILL, F. A. and BOES, D. C. (1974). Introduction to the Theory of Statistics. Third edition. McGraw-Hill, New York.

PITMAN, E. J. G. (1938). The estimation of location and scale parameters of a continuous population of any given form. *Biometrika* **30** 391–421.

RAO, C. R. (1976). Estimation of parameters in a linear model. *Ann. Statist.* **6** 1023–1037.

RUKHIN, A. L. (1978). Universal Bayes estimators. *Ann. Statist.* **6** 1345–1351.

RUKHIN, A. L. (1984). Universal estimators of a vector parameter. *J. Multivariate Anal.* **14** 135–154.

SAVAGE, L. J. (1954). The Foundations of Statistics. Wiley, New York.

SHINOZAKI, N. (1975). A study of generalized inverse of matrix and estimation with quadratic loss. Ph.D. thesis submitted to the Keio University, Japan.

SHINOZAKI, N. (1980). Estimation of a multivariate normal mean with a class of quadratic loss functions. *J. Amer. Statist. Assoc.* **75** 973–976.

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. of California Press.

THOMAS, D. H. (1970). Some contributions to radial probability distributions, statistics, and the operational calculi. Ph.D. dissertation. Wayne State University.

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

ZELLNER, A. (1976). Bayesian and non-Bayesian analysis of the regression model with multivariate student-t error terms. *J. Amer. Statist. Assoc.* **71** 400–405.

DEPARTMENT OF MATHEMATICS
CORNELL UNIVERSITY
ITHACA, NEW YORK 14853