The Annals of Statistics
1986, Vol. 14, No. 2, 444–460

IMPROVED CONFIDENCE SETS FOR THE COEFFICIENTS OF A LINEAR MODEL WITH SPHERICALLY SYMMETRIC ERRORS

BY JIUNN TZON HWANG[1] AND JEESEN CHEN

Cornell University and University of Cincinnati

Under the assumption of normal errors, confidence spheres for p ($p \geq 3$) coefficients of a linear model centered at the positive part James–Stein estimators were recently proved, by Hwang and Casella, to dominate the usual confidence set with the same radius. In this paper, the same domination results are established under various spherically symmetric distributions. These distributions include uniform distributions, double exponential distributions, and multivariate t distributions.

1. Introduction. For a standard linear model

$$X_{n\times 1} = A_{n\times p}\,\theta_{p\times 1} + \sigma\,\varepsilon_{n\times 1}, \tag{1.1}$$

assume that the design matrix A has a full rank p, ε has an n-variate normal distribution $N(0, I)$, and hence $\sigma\varepsilon \sim N(0, \sigma^2 I)$. For the confidence set problem, the unequal variance case can be transformed to this equal variance case when the ratios of variances are known. We focus here on the simpler situation where the variance σ^2 is known. The usual $1-\alpha$ confidence set for θ is

$$C_{X,\sigma} = \{\theta: (\theta - \hat{\theta})(A'A)(\theta - \hat{\theta}) \leq c^2\sigma^2\}, \tag{1.2}$$

where $\hat{\theta} = (A'A)^{-1} A'X$ is the least squares estimator, and c^2 is chosen so that $P(\chi_p^2 \leq c^2) = 1 - \alpha$.

Even though $C_{X,\sigma}$ enjoys many optimal properties (i.e., best invariant; minimax; admissible for $p \leq 2$), it is inadmissible when $p \geq 3$. Thus there exists a confidence set that dominates $C_{X,\sigma}$, i.e., which has the same volume but has higher coverage probabilities than $C_{X,\sigma}$ for every θ. In fact, the James–Stein confidence sets $C_{\mathrm{JS},\sigma}$ are such confidence sets, where

$$C_{\mathrm{JS},\sigma} = \{\theta: (\theta - \delta_{\mathrm{JS}})'(A'A)(\theta - \delta_{\mathrm{JS}}) \leq c^2\sigma^2\} \tag{1.3}$$

and

$$\delta_{\mathrm{JS}} = \theta_0 + \left\{1 - \frac{a\sigma^2}{(\hat{\theta} - \theta_0)'(A'A)(\hat{\theta} - \theta_0)}\right\}^+ (\hat{\theta} - \theta_0). \tag{1.4}$$

The point estimator δ_{JS}, with a being some positive number, is the positive part

Received March 1983; revised October 1985.
[1] Research supported by the NSF Grant Number MCS-8003568.
AMS 1980 *subject classifications*. Primary 62F25, 62F35; secondary 62J07, 62C20.
Key words and phrases. Coverage probability, uniform distribution, double exponential distribution, multivariate t distribution, spherically symmetric distributions.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Statistics*. www.jstor.org

James–Stein estimator (1961) shrinking toward a prior guess θ_0 of θ. The existence of confidence sets dominating $C_{X,\sigma}$ was proved independently by Brown (1966) and Joshi (1967). Constructive results were given in Hwang and Casella (1984), which imply domination for any positive a less than a specific bound. The upper bound that is derived analytically depends on c^2 and p but is about $0.8(p-2)$ when $C_{X,\sigma}$ has 0.9 coverage probability and $p = 20$. The largest a so that domination maintains is analytically proved to be no greater than $2(p-2)$ but is close to being $2(p-2)$ according to their numerical study. The numerical study in an earlier article by Hwang and Casella (1982) shows that the largest improvement in probability (which occurs at $\theta = \theta_0$) can be substantial. The history of this problem was discussed therein.

In this paper, we consider the same model as in (1.1), except that ε is assumed to have a spherically symmetric distribution. That is,

$$X = A\theta + \sigma\varepsilon, \qquad \varepsilon \sim f(|\varepsilon|^2) \quad \text{p.d.f.} \tag{1.5}$$

(Normally the variance of any component of ε is taken to be 1.) We establish results for a general f with special emphasis on t distributions, double exponential distributions, and spherical uniform distributions. For a t distribution (with N degrees of freedom),

$$f(|\varepsilon|^2) = \text{constant}\left(1 + \frac{1}{N}|\varepsilon|^2\right)^{-(N+p)/2}; \tag{1.6}$$

for an exponential distribution (with parameter k)

$$f(|\varepsilon|^2) = \text{constant } e^{-k|\varepsilon|}. \tag{1.7}$$

The specific probability density function of the spherical uniform distribution is given in (2.6). Model (1.5) has been considered in the literature. In particular, when ε has a t distribution, it was used by authors cited in Zellner (1976) to model some practical situations. Under model (1.5), the usual least squares estimator and its corresponding tests were studied to some extent by Thomas (1970), Zellner (1976), and Box (1952 and 1953) and others. See Chmielewski (1981) for an excellent survey.

Note, for virtually any distribution having θ as a location parameter, Brown (1966) has proved, for $p \geq 3$, the existence of confidence sets dominating $C_{X,\sigma}$. However, no constructive results have been obtained.

In this paper, we prove under (1.5) the domination of $C_{JS,\sigma}$ over $C_{X,\sigma}$ for a less than a given bound. The upper bound, in general, depends on the underlying distribution. However, for many classes of flat tailed distributions that include multivariate t distributions, double exponential distributions, and uniform distributions, the corresponding upper bounds are greater or approximately equal to a_{NL}, the upper bound for the normal case. Therefore, the superiority of $C_{JS,\sigma}$ over $C_{X,\sigma}$ as proved by Hwang and Casella (1984) for the normal case, has been broadened to these classes of distributions.

In Section 2, we prove a general theorem (Theorem 2.2) that relates the domination result of $C_{JS,\sigma}$ over $C_{X,\sigma}$ to a measure of flatness of the assumed distribution. See also Corollary 5.2. Sections 3 and 4 contain some stronger results for the t distributions and double exponential distributions. Section 5 provides conditions on a which are necessary and are nearly sufficient for the domination of $C_{\mathrm{JS},\sigma}$ over $C_{X,\sigma}$. This is based on the study of the coverage probability of $C_{\mathrm{JS},\sigma}$ when $|\theta|$ is large. In many cases, the largest possible a for domination is at least $2(p-2)$.

Note that our theorems also establish the minimaxity of $C_{\mathrm{JS},\sigma}$ for a wide class of spherically symmetric distributions since $C_{\mathrm{JS},\sigma}$ is superior to $C_{X,\sigma}$, which is proved to be minimax by Hooper (1982).

Previously, in the context of point estimation, the point estimators of the form $\delta_{\mathrm{JS},\sigma}$ have been shown in Strawderman (1974), Berger (1975), Brandwein (1979), and Brandwein and Strawderman (1978 and 1980) to dominate $\hat{\theta}$ for various spherically symmetric distributions and various losses. Our results here can also be considered to be their counterpart, namely the establishment of improved confidence sets associated with their point estimators.

2. Domination results and their implications. To study the model (1.5), we can assume without loss of generality that it has a canonical form representation, i.e., $A = \begin{pmatrix} A_1 \\ 0 \end{pmatrix}$, where A_1 is a $p \times p$ nonsingular matrix. Furthermore, we apply a linear transformation

$$\hat{\theta} \to \hat{\theta}^* = (A_1'A_1)^{1/2}(\hat{\theta} - \theta_0) \quad \text{and} \quad \theta \to \theta^* = (A_1'A_1)^{1/2}(\theta - \theta_0)$$

and note that $\hat{\theta}^*$ has the same distribution as $\varepsilon_p + \theta^*$ where ε_p is the vector of the first p components of ε. The distribution of ε_p is also spherical. Even though the p.d.f. of ε_p is obviously different from that of ε, it will be denoted as $f(\cdot)$ below. However, if ε has a n-dimensional t distribution with N degrees of freedom, then ε_p has a p-dimensional t distribution with the same degree N. By using the new variables $\hat{\theta}^*$ and θ^*, we can now assume without loss of generality that model (1.5) is

$$X_p = \theta + \varepsilon_p,$$

where X_p is the first p components of X. Also, since σ^2 is known, we can assume without loss of generality that it is 1. Suppressing the subscripts p in X_p and ε_p leads us to write the model (1.5) as

$$X = \theta + \varepsilon, \qquad \varepsilon \sim f(|\varepsilon|^2) \quad \text{p.d.f.} \tag{2.1}$$

Now $C_{X,\sigma}$ and $C_{\mathrm{JS},\sigma}$ are reduced, respectively, to

$$C_X = \{\theta: |\theta - X|^2 \le c^2\}, \tag{2.2}$$

and

$$C_{\delta^a} = \{\theta: |\theta - \delta^a(X)|^2 \le c^2\}, \tag{2.3}$$

where

$$\delta^a(X) = \left(1 - \frac{a}{|X|^2}\right)^+ X.$$

Note that in all the examples considered in this paper, f is decreasing. (Throughout all the paper, decreasing means nonincreasing. Similarly, increasing is equivalent to nondecreasing.) Hence C_X is best among the invariant set estimators in the sense that C_X has minimum volume among the invariant set estimators with coverage probability at least $1 - \alpha = P(\theta \in C_X)$. Furthermore, C_X is minimax. That is C_X minimizes the maximum of the volumes among all the set estimators with confidence coefficient (or minimum coverage probability) at least $1 - \alpha$. See Hooper (1982, Theorems 1 and 2).

In this section, we will develop theorems that relate the domination of C_{δ^a} over C_X to a quantity that measures the flatness of f. Readers more interested in t or exponentially distributed error may read the next sections first without much difficulty.

DEFINITION 2.1. The quantity $f'(s)/f(s)$, when it is defined, is called the relative increasing rate (RIR) of f at s.

The RIR of f measures the rate of increase of f relative to f and is usually negative. If f has a large RIR, f dies out to zero slowly and consequently f has a heavy tail. On the other hand, if f is very small (or very negative), f has a sharp tail that dies out to zero quickly.

In the special case that $X \sim N(\theta, \sigma^2 I)$, the RIR is a constant function equal to $-(2\sigma^2)^{-1}$. The following theorem states the domination results and will be proved at the end of this section.

THEOREM 2.2. *Assume that the* RIR *of $f(s)$ is defined for every s, $\alpha_0 < s < \alpha_1$, where*

$$\alpha_0 = \left[(c - \sqrt{a})^+\right]^2 \quad \text{and} \tag{2.4}$$

$$\alpha_1 = c^2 + a.$$

If $a > 0$ is such that

$$\inf_{\alpha_0 < s < \alpha_1} \frac{f'(s)}{f(s)} \geq \frac{-(p-2)}{2c\sqrt{a}} \ln\left[\frac{c + \sqrt{c^2 + a}}{\sqrt{a}}\right], \tag{2.5}$$

then the coverage probability of C_{δ^a} is higher than C_X for every θ. Since C_{δ^a} has the same volume as C_X, C_{δ^a} dominates C_X.

The following corollary gives some insight about the a's that satisfy (2.5). The proof is straightforward and is omitted.

COROLLARY 2.3. *When $p > 2$, the solutions of a to the inequality* (2.5) *form an interval $(0, a_0]$ where $a_0 > 0$. If the left-hand side of* (2.5) *is continuous in a, then a_0 is the unique solution to* (2.5) *with the inequality replaced by an equality.*

When $X \sim N(\theta, \sigma^2 I)$, the left-hand side of (2.5) is then the RIR, $-(2\sigma^2)^{-1}$. Theorem 2.2 and Corollary 2.3 then reduce to Theorem 2.1 of Hwang and Casella (1984) which is stronger than the domination result in Hwang and Casella (1982). Using a programmable calculator, one can calculate the upper bound a_0. For the normal case with $\sigma^2 = 1$, the numerical values of a_0, denoted as a_{NL}, were reported for selected values of c^2 and p in Hwang and Casella (1984) and are reported in Table 1 ($N = \infty$) for convenience of further discussions.

Theorem 2.2 and Corollary 2.3 apply to virtually any spherically symmetric distributions. In applying these theorems, the calculations are usually straightforward. See Hwang and Chen (1983) for results concerning other distributions.

TABLE 1

Values of bounds a_t for domination under multivariate t distribution with N degrees of freedom. For convenience of comparing to the normal case, c^2 were chosen so that $P(\chi_p^2 \le c^2) = 0.90$. Values of c^2 are given in Table 2.

N	$p = 3$	$p = 5$	$p = 7$	$p = 9$
1	0.866	1.887	2.702	3.438
3	0.792	2.081	3.003	3.811
5	0.746	2.220	3.240	4.123
7	0.716	2.324	3.431	4.384
9	0.695	2.401	3.586	4.605
10	0.687	2.396	3.653	4.703
15	0.659	2.345	3.913	5.098
20	0.642	2.310	4.019	5.380
25	0.632	2.284	3.990	5.591
35	0.618	2.250	3.947	5.652
45	0.614	2.228	3.917	5.621
55	0.605	2.213	3.895	5.596
90	0.596	2.184	3.850	5.540
∞	0.580	2.132	3.760	5.413

N	$p = 11$	$p = 15$	$p = 20$	$p = 25$
1	4.132	5.449	7.018	8.539
3	4.556	5.940	7.560	9.112
5	4.924	6.385	8.065	9.657
7	5.242	6.787	8.535	10.172
9	5.519	7.150	8.971	10.658
10	5.644	7.318	9.117	10.890
15	6.162	8.043	10.097	11.954
20	6.547	8.616	10.863	12.870
25	6.843	9.076	11.503	13.661
35	7.265	9.764	12.508	14.946
45	7.323	10.251	13.252	15.937
55	7.229	10.613	13.824	16.718
90	7.240	10.645	14.882	18.489
∞	7.079	10.434	14.653	18.890

For double exponential distributions and t distributions, the a_0 are small. Fortunately, these upper bounds are enlarged in Sections 3 and 4. Even though Theorem 2.2 is weak for these special cases, it reveals clearly the relationship between the domination result and the RIR. For another perhaps more significant connection, see the first paragraph after Corollary 5.2.

Theorem 2.2 asserts that if the RIR of f is uniformly bounded below by a certain bound, depending on p, c, and a, C_{δ^a} dominates C_X. Therefore, the message is clear: Stein's set estimator dominates C_X if the tail is heavy enough. This is probably due to the fact that James–Stein estimator is a shrinkage estimator.

We can also apply Theorem 2.2 to a spherical uniform distribution.

COROLLARY 2.4 ($p \geq 2$) (Uniform distribution over a sphere centered at the origin with known radius). *Suppose that the* p.d.f. *of* $\varepsilon = X - \theta$ *is*

$$f(|\varepsilon|^2) = \text{constant} \quad \textit{if } |\varepsilon| \leq R, \qquad = 0 \quad \textit{otherwise}. \tag{2.6}$$

Then C_{δ^a} *dominates* C_X *if* $0 < a \leq (R^2 - c^2)$.

In deriving Corollary 2.4, note that (2.5) is automatically satisfied for $p \geq 2$ as long as the RIR is well defined. This is equivalent to $a \leq R^2 - c^2$. Also note that if the true distribution is uniform and if C_X has coverage probability $1 - \alpha < 1$ then $c^2 < R^2$ and hence the condition on a is not vacuous. A striking feature is that even for $p = 2$, C_X can be improved by Corollary 2.4. This is not very surprising in light of the fact that the best location invariant set estimator is not unique which implies that C_X can be uniformly improved even for $p = 1$ as shown in Farrell (1964).

The remainder of this section will be devoted to the proof of Theorem 2.2. We will need the following two lemmas which will also be useful in dealing with the t distributions and double exponential distributions in Sections 3 and 4.

Assume as in (2.1) that the p.d.f. of X is $f(|x - \theta|^2)$. To prove the domination of C_{δ^a} over C_X, we follow the technique developed in Hwang and Casella (1984). We consider two regions of θ: $|\theta| \leq c$ and $|\theta| > c$. For the first case, we have the following lemma.

LEMMA 2.5. *For* $|\theta| \leq c$ *and* $a > 0$,

$$P(\theta \in C_{\delta^a}) > P(\theta \in C_X). \tag{2.7}$$

PROOF. The proof is similar to the proof of Theorem 2.1 in Hwang and Casella (1982) and is hence omitted. □

Below we need only focus on the situation $|\theta| > c$. For such a region a formula for $\partial/\partial a\, P(\theta \in C_{\delta^a})$ is established in Lemma 2.6. Note that the domination results can be proved for $a \in (0, a_0]$ if one can show that for every $a \in (0, a_0]$,

$$\frac{\partial}{\partial a} P(\theta \in C_{\delta^a}) > 0,$$

since this implies that for all $a \in (0, a_0]$

$$P(\theta \in C_{\delta^a}) > \lim_{a \to 0^+} P(\theta \in C_{\delta^a}) = P(\theta \in C_X),$$

due to the continuity of $P(\theta \in C_{\delta^a})$ as a function of a.

Define

$$u(r) = \left(1 - \frac{a}{r^2}\right)^+,$$

$$\alpha(r) = \alpha(r, \beta) = r^2 - 2r|\theta|\cos\beta + |\theta|^2,$$

$$\Omega = 2 \prod_{i=1}^{p-2} \int_0^\pi \sin^{i-1}(t)\, dt, \tag{2.8}$$

and

$$\beta_0 = \sin^{-1}\left(\frac{c}{|\theta|}\right) < \frac{\pi}{2}.$$

Also let $r_{\pm}$ be solutions to

$$r_{\pm} u(r_{\pm}) = |\theta|\cos\beta \pm \sqrt{c^2 - |\theta|^2 \sin^2\beta} \overset{\text{def}}{=} r_{\pm}^0; \tag{2.9}$$

i.e.,

$$r_{\pm}(a, \theta, \beta) = \left(r_{\pm}^0 + \sqrt{\left(r_{\pm}^0\right)^2 + 4a}\right)/2. \tag{2.10}$$

Using these notations and a spherical transformation, one can write

$$P(\theta \in C_{\delta^a}) = \Omega \int_0^{\beta_0} \int_{r_-}^{r^+} r^{p-1} \sin^{p-2}\beta f(\alpha(r))\, dr\, d\beta. \tag{2.11}$$

Now the derivative formula of Hwang and Casella (1984) can be generalized to this case. The straightforward proof, which is omitted, is based on interchanging the order of differentiation and integration, and the fundamental theorem of calculus.

LEMMA 2.6. *Assume that $|\theta| > c$ and that $f(\alpha(r))$ is a continuous function on the set of (r, β) such that $r_- \le r \le r_+$ and $0 \le \beta < \beta_0$. Then*

$$\frac{\partial}{\partial a} P_\theta(\theta \in C_{\delta^a}) = \Omega \int_0^{\beta_0} m(a, \theta, \beta)\, d\beta, \tag{2.12}$$

where

$$m(a, \theta, \beta) = \sin^{p-2}\beta \left[\frac{r_+^p f(\alpha(r_+))}{r_+^2 + a} - \frac{r_-^p f(\alpha(r_-))}{r_-^2 + a}\right]. \tag{2.13}$$

PROOF OF THEOREM 2.2. By Lemma 2.6., we need only show that for all β, $0 < \beta < \beta_0$, and θ, $|\theta| > c$,

$$m(a, \theta, \beta) > 0,$$

which is clearly equivalent to

$$\left(\frac{r_+}{r_-}\right)^{p-2} \frac{f(\alpha(r_+))}{f(\alpha(r_-))} > \frac{1 + ar_+^{-2}}{1 + ar_-^{-2}}. \tag{2.14}$$

Since $r_+ > r_-$ for all $0 < \beta < \beta_0$, (2.14) could be established if one could show

$$\left(\frac{r_+}{r_-}\right)^{p-2} \frac{f(\alpha(r_+))}{f(\alpha(r_-))} \geq 1, \tag{2.15}$$

or equivalently,

$$g(\alpha(r_+)) - g(\alpha(r_-)) \geq -(p-2)\ln\frac{r_+}{r_-}, \tag{2.16}$$

where $g(s) = \ln f(s)$. By the mean value theorem, $g(\alpha(r_+)) - g(\alpha(r_-))$ equals $g'(s)(\alpha(r_+) - \alpha(r_-))$ for some number s between $\alpha(r_-)$ and $\alpha(r_+)$. Now

$$\alpha(r_+) - \alpha(r_-) = (r_+ - r_-)(r_+ + r_- - 2|\theta|\cos\beta). \tag{2.17}$$

From (2.9),

$$2|\theta|\cos\beta = r_+^0 + r_-^0.$$

Since (2.9) and (2.10) imply $r_+ > r_- > \sqrt{a}$,

$$u(r_\pm) = \left(1 - \frac{a}{r_\pm^2}\right)^+ = 1 - \frac{a}{r_\pm^2},$$

which, together with (2.9), imply that

$$2|\theta|\cos\beta = r_+ - \frac{a}{r_+} + r_- - \frac{a}{r_-}.$$

Substituting this expression for $2|\theta|\cos\beta$ in (2.17) shows that

$$\alpha(r_+) - \alpha(r_-) = a\left(\frac{r_+}{r_-} - \frac{r_-}{r_+}\right) > 0 \tag{2.18}$$

and that (2.16) is equivalent to

$$ag'(s)\left(t - \frac{1}{t}\right) \geq -(p-2)\log t,$$

where t denotes r_+/r_-. The last inequality can be established if we require that

$$\inf_{\alpha(r_-) < s < \alpha(r_+)} g'(s) \geq -\frac{p-2}{a} \inf_{\substack{|\theta| > c \\ 0 < \beta < \beta_0}} \frac{\log t}{t - 1/t}. \tag{2.19}$$

Differentiating the function $(\log t)/(t - 1/t)$, dropping the denominator, and differentiating the numerator again show that

$$\frac{\log t}{t - 1/t} \text{ is decreasing in } t, \qquad t > 1. \tag{2.20}$$

Hence the right-hand side of (2.19) attains its infimum at $t = t^*$, where

$$t^* \overset{\text{def}}{=} \sup_{\substack{|\theta| > c \\ 0 < \beta < \beta_0}} t .$$

It can be shown as on page 9 of Hwang and Casella (1984) that t is decreasing in β and $|\theta|$, and consequently

$$t^* = t\big|_{\substack{|\theta| = c \\ \beta = 0}} = \frac{c + \sqrt{c^2 + a}}{\sqrt{a}} \tag{2.21}$$

and

$$t^* - \frac{1}{t^*} = \frac{2c}{\sqrt{a}}.$$

Therefore (2.19) is equivalent to

$$\inf_{\alpha(r_-) < s < \alpha(r_+)} g'(s) \geq \frac{-(p-2)}{2c\sqrt{a}} \log\left[\frac{c + \sqrt{c^2 + a}}{\sqrt{a}}\right].$$

Comparing this inequality with (2.5) and noting $g'(s) = f'(s)/f(s)$, we would have established Theorem 2.2 if we could show

$$\alpha_0 = \inf_{\substack{|\theta| > c \\ 0 < \beta < \beta_0}} \alpha(r_-) \tag{2.22}$$

and

$$\alpha_1 = \sup_{\substack{|\theta| > c \\ 0 < \beta < \beta_0}} \alpha(r_+), \tag{2.23}$$

where α_0 and α_1 were given in the statement of Theorem 2.2. These two equations are established in Lemmas A.3 and A.4 of Hwang and Chen (1983) and Theorem 2.2 is proved. □

For a multivariate t distribution and a double exponential distribution, Theorem 2.2 can be strengthened, which will be the goals of the next two sections.

3. Refinement of the domination results for the multivariate t distribution. In many distributions including normal distributions, t distributions, and the exponential distributions, the RIR $f'(s)/f(s)$ is an increasing function. (In other words, ln f is convex or f is log convex.) In this section we take advantage of such a fact and derive some stronger theorems. Even though we mainly focus on t distributions, the ideas can be grasped more easily if the general results are presented first.

THEOREM 3.1. *Assume that* $X \sim f(|x-\theta|^2)$ *and* f *is log convex. Let* α_0 *and* t^* *be as in* (2.4) *and* (2.21). *Then* C_{δ^a} *dominates* C_X *provided that*

$$\inf_{1 \le t \le t^*} h(t) \ge -(p-2), \tag{3.1}$$

where

$$h(t) = \frac{g\big(a(t-t^{-1})+\alpha_0\big)-g(\alpha_0)}{\ln t} \quad \textit{and} \quad g(s) = \ln f(s).$$

PROOF. Applying (2.18) and using the notation $t = r_+/r_-$ show that (2.16) is equivalent to

$$\frac{g\big(\alpha(r_-)+a(t-t^{-1})\big)-g\big(\alpha(r_-)\big)}{\ln t} \ge -(p-2). \tag{3.2}$$

Since $\alpha(r_-) \ge \alpha_0$ and g is convex, the left-hand side is greater than or equal to

$$\frac{g\big(a(t-t^{-1})+\alpha_0\big)-g(\alpha_0)}{\ln t}.$$

This together with (3.1) imply the theorem. □

Minimizing $h(t)$ is usually quite difficult. However, with the help of the following Lemma 3.2, it can be solved for t distributions and double exponential distributions. An empty set and a single point set are called degenerate intervals.

LEMMA 3.2. *Assume that* g *is differentiable. Suppose the set of* $t \ge 1$ *such that*

$$\frac{p-2}{t} + ag'\big(a(t-t^{-1})+\alpha_0\big)(1+t^{-2}) \ge 0 \tag{3.3}$$

is an interval (*possibly degenerate or possibly with infinite length*). *Then* (3.1) *is equivalent to* (i) $h(1^+) \ge -(p-2)$ *and* (ii) $h(t^*) \ge -(p-2)$.

PROOF. Clearly (i) and (ii) are necessary for (3.1). To prove that (i) and (ii) are sufficient, note that (3.1) is equivalent to

$$(p-2)\ln t + g\big(a(t-t^{-1})+\alpha_0\big) - g(\alpha_0) \ge 0, \quad \forall t, 1 \le t \le t^*. \tag{3.4}$$

The derivative of the left-hand side of (3.4) is exactly the left-hand side of (3.3) which by assumption has an interval solution. If this interval is degenerate, then (i) and (ii) clearly imply (3.4) and this lemma is proved. If this interval is not degenerate, let $\lambda_1, \lambda_2, 1 \le \lambda_1 < \lambda_2 < \infty$ be the endpoints. Below, we show that $\lambda_1 = 1$ and hence the left-hand side of (3.4) is increasing for $t \in [1, \lambda_2]$ and is decreasing for $t \ge \lambda_2$. This should imply that (3.4) holds for t, $1 \le t \le t^*$, since

(3.4) holds for $t = 1$ by trivial observation and also holds for $t = t^*$ by condition (ii). This would have established this lemma.

To show $\lambda_1 = 1$, all we need to do is to show that (3.3) is satisfied for $t = 1$. Now condition (i) and l'Hospital's rule imply

$$-(p-2) \le h(1^+) = atg'(a(t-t^{-1}) + \alpha_0)(1+t^{-2})|_{t=1},$$

which is equivalent to (3.3) for $t = 1$. □

Theorem 3.1 and Lemma 3.2 can be applied to the t distributions and the double exponential distributions and yield domination results stronger than Theorem 2.2. Here, we concentrate on t distributions, since the results for the double exponential distributions can be further improved in the next section.

COROLLARY 3.3 (Multivariate t distribution). *For $p > 2$, C_{δ^a} uniformly dominates C_X provided $0 < a \le a_t$, where $a_t = \min(a_1, a_2)$,*

$$a_1 = \min\left\{c^2, \left(\frac{p-2}{N+2}\right)^2\left(-c + \sqrt{c^2 + \frac{N+c^2}{p-2}(N+2)}\right)^2\right\} \quad \text{if } \frac{N}{N+p}(p-2) \le c^2,$$

$$= \frac{N}{N+p}(p-2) \quad \text{otherwise},$$

and a_2 is the unique solution to

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

PROOF. For multivariate t distributions, (3.3) is equivalent to

$$\frac{p-2}{t} - a(1+t^{-2})\frac{N+p}{2(N + a(t-t^{-1}) + \alpha_0)} \ge 0 \tag{3.5}$$

or

$$u(t) \overset{\text{def}}{=} \left(t + \frac{1}{t}\right) - m\left(t - \frac{1}{t}\right) \le \frac{m}{a}(N + \alpha_0),$$

where $m = 2(p-2)/n + p$. Clearly the derivative of $u(t)$ is increasing in t and hence $u(t)$ is convex. Therefore the solutions to (3.5) form an interval. Applying Theorem 3.1 and Lemma 3.2, and solving conditions (i) and (ii) yield this corollary. □

Numerical values of a_t for the multivariate t distribution are reported in Table 1. Note in Table 1, that the a_t are, in many cases, comparable with a_{NL}, whose values were given in the same table under $N = \infty$. Hence the domination results for the normal case as established in Hwang and Casella (1984) hold for many multivariate t distributions by Corollary 3.3. Numerical studies in Hwang (1983) and the asymptotic results in Section 5 show that under multivariate t distributions C_{δ^a} usually dominates C_X even when $a = 2(p-2)$.

4. Refinements of the domination results for the double exponential distribution. Even though the technique developed in previous sections applies also to the double exponential distribution and yields larger a's than Theorem 2.2, we can do even better by another approach to be described here. In establishing the domination results, as before, all we have to do is to establish a sufficient condition for (2.16) or equivalently

$$L(|\theta|, \beta) \overset{\text{def}}{=} \frac{g(\alpha(r_+)) - g(\alpha(r_-))}{\log t} \geq -(p-2). \tag{4.1}$$

The difficult in deriving a sufficient condition for (4.1) is that L depends on $|\theta|$ and β in a fairly complicated manner and consequently the minimum over all $|\theta|$ and β is hard to find. Under the condition of the following lemma, we were able to show that L is minimized at $\beta = 0$ and hence the remaining minimization problem involves only $|\theta|$ which is considerably simpler.

LEMMA 4.1. *Assume that $g(t)$ is convex and decreasing and g' is concave. Then for every $|\theta|$, $L(|\theta|, \beta)$ is increasing in β and consequently for every θ, $L(|\theta|, \beta)$ is minimized at $\beta = 0$.*

PROOF. Write

$$\begin{aligned} L(|\theta|, \beta) &= \left[\frac{g(\alpha(r_+)) - g(\alpha(r_-))}{\alpha(r_+) - \alpha(r_-)}\right]\left[\frac{\alpha(r_+) - \alpha(r_-)}{\log t}\right] \\ &\overset{\text{def}}{=} R_1 R_2. \end{aligned} \tag{4.2}$$

From (2.18), $R_2 = a(t - t^{-1})/\log t$ and by (2.20), R_2 increases as t increases. Since t decreases as β increases, so does R_2. Now because R_1 is nonpositive, to establish the lemma, it suffices to show that $-R_1$ is decreasing or R_1 is increasing in β. This can be shown as in Lemma 3.5 of Hwang and Chen (1983). The arguments are fairly technical and are omitted. □

The assumptions of Lemma 4.1 are satisfied for multivariate t distributions and double exponential distributions. Now under the assumptions of Lemma 4.1, a sufficient condition for domination is

$$L(|\theta|, 0) \geq -(p-2) \quad \text{for every } |\theta| > c,$$

which by (2.9) and (2.10), is equivalent to

$$\left.\frac{g((r_+ - |\theta|)^2) - g((r_- - |\theta|)^2)}{\log(r_+/r_-)}\right|_{\beta=0} \geq -(p-2), \tag{4.3}$$

where

$$r_{\pm}|_{\beta=0} = \tfrac{1}{2}\left[|\theta| \pm c + \sqrt{(|\theta| \pm c)^2 + 4a}\right].$$

If one could find the minimum of the left-hand side of (4.3), in general, one would have established domination results stronger than Theorem 2.2 and Corollary 3.3. However, solving this minimization problem, in general, is very difficult. So far, we have only been successful for the double exponential distribution. The result is reported in the following theorem.

THEOREM 4.2 (Double exponential distribution). *For* $p > 2$, C_{δ^a} *uniformly dominates* C_X *provided* $0 < a \le \min(c^2, a_3) \overset{\text{def}}{=} a_E$ *where* a_3 *is the unique solution to*

$$\left(\frac{c + \sqrt{c^2 + a}}{\sqrt{a}}\right)^{(p-2)/k} \cdot \exp\{-\sqrt{c^2 + a} + c - \sqrt{a}\} = 1 \tag{4.4}$$

PROOF. Note that $g(s) = -k\sqrt{s}$. To establish (4.3), we show that under the condition $a \le c^2$, $L(|\theta|, 0)$ is minimized at $|\theta| = c$. Since the condition $L(c,0) \ge -(p-2)$ is equivalent to $0 < a \le a_3$, we will have established the theorem. Now similar to (4.2), write

$$L(|\theta|, 0) = \frac{a(t - t^{-1})}{\log t}\,\frac{g(\alpha(r_+)) - g(\alpha(r_-))}{\alpha(r_+) - \alpha(r_-)}\bigg|_{\beta=0}$$

$$= \frac{a(t - t^{-1})}{\log t}\,\frac{(-K)}{|r_+ - |\theta|| + |r_- - |\theta||}\bigg|_{\beta=0}.$$

Note that $r_+ > |\theta|$ and since, $a \le c^2$, that $r_- \le |\theta|$. Therefore

$$|r_+ - |\theta|| + |r_- - |\theta|||_{\beta=0} = -r_-|_{\beta=0},$$

which equals

$$\phi(|\theta| + c) - \phi(|\theta| - c), \tag{4.5}$$

where $\phi(s) = [s + \sqrt{s^2 + 4a}]/2$. Since $\phi(s)$ is convex, (4.5) increases in $|\theta|$. The function $-L(|\theta|, 0)$ thus decreases in $|\theta|$, since $a(t - t^{-1})/\log t$ increases in t by (2.20) and t decreases in $|\theta|$. This implies that $L(|\theta|, 0)$ increases in $|\theta|$ and the theorem follows. □

Note that Theorem 4.2 is very strong in that it specifies a very large upper bound on a. This bound, a_E, is probably very close to the best bound that one can establish using the technique of Hwang and Casella (1984). As in Table 2, for $k = 1$, a_E is larger than $p - 2$, the traditional choice of a in the point estimation problem. If one considers the k so that the common variance is 1, then $k = \sqrt{p+1}$. (In general the variance equals $(p+1)/k^2$.) For $k = \sqrt{p+1}$, a_E is also reported in Table 2 and is less than but close to a_{NL} as reported in the

TABLE 2

Values of bounds for domination. The c^2 are chosen so that $P(\chi_p^2 \le c^2) = 0.9$.

	c^2	$N(\theta, I)$ a_{NL}	$k = 1$ a_E	$k = \sqrt{p+1}$ a_E
3	6.251	0.580	1.448	0.643
4	7.779	1.339	3.408	1.401
5	9.236	2.132	5.650	2.158
6	10.645	2.942	8.129	2.906
7	12.017	3.760	10.804	3.646
8	13.362	4.585	[a]	4.378
9	14.684	5.413	[a]	5.103
10	15.987	6.245	[a]	5.822
11	17.275	7.079	[a]	6.537
12	18.549	7.915	[a]	7.246
13	19.812	8.754	[a]	7.952
14	21.064	9.593	[a]	8.654
15	22.307	10.434	[a]	9.353
16	23.542	11.276	[a]	10.050
17	24.769	12.119	[a]	10.743
18	25.989	12.963	[a]	11.435
19	27.204	13.808	[a]	12.124
20	28.412	14.653	[a]	12.811
21	29.615	15.500	[a]	13.496
22	30.813	16.346	[a]	14.180
23	32.007	17.194	[a]	14.862
24	33.196	18.042	[a]	15.542
25	34.382	18.890	[a]	16.211

[a] The value is the same as the value of c^2 in the first column and the same row.

same table. Hence most of the domination results for the normal case established in Theorem 2.1 of Hwang and Casella (1984) stand under the double exponential distribution with common variance 1.

It is unfortunate that we failed to establish a theorem similar to Theorem 4.2 for the multivariate t distribution. The corresponding expression on the left-hand side of (4.3) becomes very messy and we cannot find the minimum. If we were able to show that the minimum occurs at $|\theta| = c$, we would have shown that domination results hold for $a \le a_2$ rather than $a \le \min(a_1, a_2)$ as needed in Corollary 3.3. This would establish a larger interval of a for domination. However, numerical study also shows that $|\theta| = c$ is not the minimum point of the left-hand side of (4.3) unless there are further conditions on a and c^2.

5. A necessary and nearly sufficient condition for the domination of C_{δ^a} over C_X. In the previous sections, we provided sufficient conditions for C_{δ^a} to dominate C_X. In most cases, a was less than $p - 2$, the traditional choice of a in the James–Stein point estimator. Here, we provide some evidence that the range of a is probably at least twice as large as what was given earlier.

In this section, we use an asymptotic formula (5.1) (as $|\theta| \to \infty$) to derive necessary conditions (Corollary 5.2). We provide some evidence that these necessary conditions are close to be sufficient. Theorem 5.1, generalizing Theorem 3.1 of Hwang and Casella (1984) for the normal case, can be proved by using Taylor expansions and some tricky applications of integration by parts. For the details, see Hwang (1983).

THEOREM 5.1 ($p \geq 2$). *Assume that $f''(t)$ exists and is continuous in t, $0 < t \leq c^2$. If*

(i) $$\left| \int_{|Y| \leq c} f'(|Y|^2)\, dY \right| < \infty,$$

(ii) $$\left| \int_{|Y| \leq c} |Y|^2 f''(|Y|^2)\, dY \right| < \infty,$$

and

(iii) $$\lim_{t \to 0} t^p f(t^2) = \lim_{t \to 0} t^p f'(t^2) = 0,$$

then, as $|\theta| \to \infty$

$$P(\theta \in C_{\delta^a}) = 1 - \alpha + \frac{\Omega a c^p}{p|\theta|^2}\left[(p-2)f(c^2) + af'(c^2)\right] + o(|\theta|^{-2}), \tag{5.1}$$

where $1 - \alpha = P(\theta \in C_X)$ and Ω is as in (2.8).

The assumptions in Theorem 5.1 are not restrictive and are satisfied by normal, double exponential, and multivariate t distributions. Theorem 5.1 can be used to provide conditions necessary for the domination of C_{δ^a} over C_X.

COROLLARY 5.2. *Under the assumptions of Theorem* 5.1, *necessary conditions for the domination of C_{δ^a} over C_X are $a > 0$ and*

$$af'(c^2) + (p-2)f(c^2) \geq 0. \tag{5.2}$$

PROOF. Obviously if C_{δ^a} dominates C_X then $a > 0$. (Otherwise if $a < 0$, C_{δ^a} has smaller coverage probability than C_X at $|\theta| = 0$ and if $a = 0$, C_{δ^a} and C_X are identical.) Inequality (5.2) follows directly from (5.1). □

If $f(c^2) > 0$, condition (5.2) is then equivalent to

$$\frac{f'(c^2)}{f(c^2)} \geq -\frac{p-2}{a}. \tag{5.3}$$

Note that the left-hand side is exactly the RIR. Therefore, (5.3) implies domination for large $|\theta|$ if the tail of the underlying distribution is flat enough.

Due to the shrinkage nature of δ^a, one expects that for small $|\theta|$, δ^a and C_{δ^a} perform better than X and C_X, respectively. In fact, by Lemma 2.5, the coverage probability of C_{δ^a} is higher than C_X for $|\theta| \le c$ as long as $a > 0$. Therefore, it is moderate $|\theta|$ that are of concern.

For moderate $|\theta|$, one can expect the coverage probability to be reasonable. Exact numerical computations of the coverage probabilities of C_{δ^a} performed in Hwang (1983), using (2.11) show that this is the case. It turns out that if a satisfies the necessary conditions in Corollary 5.2, then C_{δ^a} is close to dominating C_X. In fact, with the addition of the condition

$$\begin{aligned} 1 - \alpha &\le P_\theta(\theta \in C_{\delta^a})\big|_{\theta = c}, \\ &= (2.11) \quad \text{with } |\theta| \text{ replaced by } c \text{ in the definition} \\ &\qquad\qquad \text{of } \beta_0 \text{ and } r_\pm, \end{aligned} \tag{5.4}$$

these would become sufficient according to the numerical studies in Hwang (1983), and, given (5.2), (5.4) is not much of a restriction.

Next we apply Corollary 5.2 to various distributions. The necessary condition is $0 < a < 2(p-2)$ for $N(\theta, I)$; $0 < a \le 2(p-2)c/k$ for the double exponential distribution (2.4); and $0 < a < 2(p-2)(N + c^2)/(N + p)$ for a multivariate t distribution. If we have chosen c according to a normal distribution (i.e., $P(\chi_p^2 \le c) = 1 - \alpha$), c^2 is larger than p (unless $1 - \alpha$ is smaller than 0.6) and hence the upper bound for the multivariate t distribution is larger than $2(p-2)$, i.e., the bound for the $N(\theta, I)$ distribution. If $k = 1$, c is usually larger than 1 (unless $1 - \alpha$ is less than 0.01); hence similar conclusion holds for this double exponential distribution. Even if $k = \sqrt{p+1}$ (so that the component variance is 1), the upper bound for the exponential distribution is usually larger than $2(p-2)$ (unless $1 - \alpha$ is less than 0.75). Hence, again, the domination results for the normal case usually hold for the multivariate t distribution and the double exponential distribution with $k = 1$ or $k = \sqrt{p+1}$.

Acknowledgments. The authors wish to thank Professors James O. Berger, Lawrence D. Brown, and George Casella, and the referees for their valuable comments which helped to substantially improve this paper over the previous draft. The authors thank Professor Peter Hooper for pointing out some references. The help of Jing-Huei Chen in many calculations is also highly appreciated.

REFERENCES

BERGER, J. O. (1975). Minimax estimation of location vectors for a wide class of densities. *Ann. Statist.* **3** 1318–1328.

BOX, G. E. P. (1952). Multi-factor designs of first order. *Biometrika* **39** 49–57.

BOX, G. E. P. (1953). Spherical Distributions (Abstract). *Ann. Math. Statist.* **24** 687–688.

BRANDWEIN, A. C. (1979). Minimax estimation of the mean of spherically symmetric distributions under general quadratic loss. *J. Multivariate Anal.* **9** 579–588.

BRANDWEIN, A. C. and STRAWDERMAN, W. E. (1978). Minimax estimation of location parameters for spherically symmetric unimodal distributions under quadratic loss. *Ann. Statist.* **6** 377–416.

BRANDWEIN, A. C. and STRAWDERMAN, W. E. (1980). Minimax estimators 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.

CASELLA, G. and HWANG, J. T. (1982). Employing vague prior information in the construction of confidence sets. Technical Report, Biometrics Unit, Cornell University.

CASELLA, G. and HWANG, J. T. (1983). Empirical Bayes confidence sets for the mean of a multivariate normal distribution. *J. Amer. Statist. Assoc.* **78** 688–698.

CHMIELEWSKI, M. A. (1981). Spherically symmetric distributions: A review and bibliography. *Internat. Statist. Rev.* **49** 67–74.

FARRELL, R. H. (1964). Estimators of a location parameter in the absolutely continuous case. *Ann. Math. Statist.* **35** 949–998.

HOOPER, P. (1982). Invariant confidence sets with smallest expected measure. *Ann. Statist.* **10** 1283–1294.

HWANG, J. T. (1983). Robust improved confidence sets. Technical Report, Cornell Statistical Center.

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

HWANG, J. T. and CASELLA, G. (1984). Improved set estimators for a multivariate normal mean. *Statist. Decisions* **1** 3–16.

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? Technical Report, Cornell Statistical Center (original version).

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

JOSHI, V. M. (1967). Inadmissibility of the usual confidence sets for the mean of a multivariate normal population. *Ann. Math. Statist.* **38** 1868–1875.

STRAWDERMAN, W. E. (1974). Minimax estimation of location parameters for certain spherically symmetric distributions. *J. Multivariate Anal.* 4 255–264.

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

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

DEPARTMENT OF MATHEMATICAL SCIENCES
UNIVERSITY OF CINCINNATI
CINCINNATI, OHIO 45221