The Annals of Statistics
1983, Vol. 11, No. 3, 814–826

IMPROVING ON INADMISSIBLE ESTIMATORS IN THE CONTROL PROBLEM[1]

BY L. MARK BERLINER

The Ohio State University

Let X have a p-variate normal distribution with unknown mean θ and identity covariance matrix. The following transformed version of a control problem (Zaman, 1981) is considered: estimate θ by d subject to incurring a loss $L(d, \theta) = (\theta^t d - 1)^2$. The comparison of decision rules in terms of expected loss is reduced to the study of differential inequalities. Results establishing the minimaxity of a large class of estimators are obtained. Special attention is given to the proposition of admissible, generalized Bayes rules which dominate the uniform prior, generalized Bayes controller when $p \geq 5$.

1. Introduction.

1.1. *Statement of the problem.* The control problem arises when it is desired to choose design variables in a standard linear regression model so that the resulting dependent (scalar) random variable will be close to a desired value. Presentations of this problem may be found in Zellner (1971), Lindley (1968), and Dunsmore (1969). Our discussion focuses on a transformed version of the problem (see Zaman, 1981). In particular, let $X = (X_1, \cdots, X_p)^t$ have a multivariate normal distribution with unknown mean $\theta = (\theta_1, \cdots, \theta_p)^t$ and identity covariance matrix. The problem is to choose a (non-randomized) rule $\delta(x) = (\delta_1(x), \cdots, \delta_p(x))^t$, subject to a loss $L(\delta, \theta)$ given by

$$L(\delta, \theta) = (\theta^t \delta - 1)^2. \tag{1.1}$$

Such a decision rule is to be evaluated in terms of its risk or expected loss function, $R(\delta, \theta) = E_\theta L(\delta(X), \theta)$.

Define a spherically symmetric (s.s.) rule to be of the form

$$\delta(x) = \psi(|x|)x \tag{1.2}$$

where $|x|^2 = \sum_{i=1}^p x_i^2$. Virtually all previous results, as well as the results presented here, apply only to s.s. rules.

In addition to proper Bayes rules, two control procedures stand out in the literature. The first is the uniform measure generalized Bayes rule δ_u given by

$$\delta_u(x) = (1 + |x|^2)^{-1}x. \tag{1.3}$$

This rule has been shown to be admissible for $p \leq 4$ (see Zaman, 1981, and Stein and Zaman, 1980). Furthermore, δ_u is inadmissible when $p \geq 5$ (see Takeuchi, 1968, and Stein and Zaman, 1980).

The second major rule often considered is δ_m defined by

$$\delta_m(x) = |x|^{-2}x. \tag{1.4}$$

This is perhaps best thought of as the maximum likelihood estimator of the quantity $|\theta|^{-2}\theta$ (see the loss function (1.1)). This rule is also known as the certainty equivalent estimator in control theory literature (Aoki, 1967). However, a complete class theorem of

Received April 1981; revised December 1982.

[1] Research supported in part by the National Science Foundation under Grant MCS-7802300. This manuscript is a revised version of a portion of the author's Ph.D. thesis presented to Purdue University.

AMS 1970 *subject classifications.* Primary 62F10; secondary 62C15, 62H99.

Key words and phrases. Admissibility, inadmissibility, generalized Bayes rules, control problem, differential inequalities, multivariate normal distribution.

Zaman (1981) demonstrates that if a s.s. estimator $\delta(x) = \psi(|x|)x$ is admissible then

$$0 \le \psi(r) \le 1, \quad \forall r \ge 0. \tag{1.5}$$

Hence, δ_m is inadmissible in all dimensions.

The main purpose of this paper is to present methods for finding rules which dominate inadmissible control procedures. Particular attention is given to the domination (when possible) of δ_u. Berliner (1982), employing the theory developed here, considers the domination of δ_m.

In general, to compare two competing estimators, δ^0 and δ^1, in terms of risk, the quantity $\Delta(\theta) = R(\delta^0, \theta) - R(\delta^1, \theta)$ is considered. By definition, if $\Delta(\theta) \ge 0$ for all θ with strict inequality for some θ, then δ^1 dominates δ^0. Unfortunately, in most cases $\Delta(\theta)$ is a difficult object to study. However, Stein (1973), while considering the estimation of a multivariate normal mean, introduced a method whereby the integrand of $\Delta(\theta)$ is manipulated in order to obtain a more manageable quantity for analysis. Stein's original procedure was to derive an expression of the form $\Delta(\theta) = E_\theta \Delta^*(X)$, where $\Delta^*(X)$ is not a function of θ. The study of the sign of $\Delta^*(X)$ may then lead to the desired dominance assertion. The quantity $\Delta^*(X)$ is usually obtained via integration by parts and, therefore, involves various (partial) derivatives of δ^0 and δ^1. Hence, the expression $\Delta^*(X) \ge 0$ is called a differential inequality. Note that $\Delta^*(X)$ is an unbiased estimate of the difference of risks, $\Delta(\theta)$.

This general technique of analysis has been applied in many settings. Examples include Berger (1980a), Haff (1977), Hudson (1978), Zidek (1978), and references given in these papers. The basic technique used here is a variant of this type of argument.

1.2. *Preliminary results.* Some general background is now presented. First, inspection of the loss function given in (1.1) justifies defining the parameter space Θ to be $\mathbb{R}^p - \{0\}$, where $\{0\}$ denotes the origin. Second, Berger, Berliner, and Zaman (1982) have shown that non-randomized rules form a complete class for the problem considered here.

Two other previous results will play important roles below. First, Zaman (1981) has shown that if $\pi(\theta)$ is a s.s. prior measure on Θ (i.e., $d\pi(\theta) = g(|\theta|)d\theta$), then the corresponding s.s. (generalized) Bayes rule δ^π is unique and is given by

$$\delta^\pi(x) = [E^\pi(\theta\theta^t | X = x)]^{-1} E^\pi(\theta | X = x). \tag{1.6}$$

As an example, consider the multivariate normal prior with mean $(0, \cdots, 0)^t$ and covariance $\tau^2 I_p$. Then the resulting Bayes rule δ_τ is given by

$$\delta_\tau(x) = [1 + \{\tau^2/(1+\tau^2)\}|x|^2]^{-1}x. \tag{1.7}$$

Second, employing invariance arguments, Berger, Berliner, and Zaman (1982) have obtained a useful formula for the (finite) Bayes risk of an arbitrary, finite risk s.s. estimator δ against a given s.s. prior measure π. Let $r(\delta, \pi)$ denote the indicated Bayes risk. Also, let $\theta^* = (\theta_2, \cdots, \theta_p)^t$. Let $S_{(p)}$ denote the surface area of a p-sphere of radius 1. Then the results of Berger, Berliner, and Zaman yield the equation

$$r(\delta, \pi) = S_{(p)}(2\pi)^{-p/2} \cdot \int_\Theta \left[\int_0^\infty r^{(p-1)}\{r\psi(r)\theta_1 - 1\}^2 \exp\{-\tfrac{1}{2}(\theta_1 - r)^2\}\exp\{-\tfrac{1}{2}|\theta^*|^2\}\, dr \right] d\pi(\theta). \tag{1.8}$$

The remainder of this section briefly reviews previous admissibility results for the control problem. These results motivate much of the discussion below. The most general inadmissibility results are found in Berger, Berliner and Zaman (1982). Their consideration of rules of the form

$$\delta(x) = (c + |x|^2)^{-1}x + |x|^{-4}w(|x|)x, \tag{1.9}$$

where c is a constant and $w(|x|) = o(1)$ as $x \to \infty$, is of particular interest here. It was shown that if $c > 5 - p$, then δ is inadmissible. In related work Zaman (1977) considered rules of the form

$$\delta(x) = (c + |x|^2)^{-1}x \tag{1.10}$$

for $|x|$ large. He showed that if $c = 5 - p$, the corresponding rule δ_a is *asymptotically optimal* (i.e., for all $|\theta|$ sufficiently large, δ_a has the smallest risk of rules of the form (1.10)).

Berger, Berliner, and Zaman (1982) also considered the admissibility of generalized Bayes rules. Let $g(|\theta|)$ be a bounded, generalized prior density and let δ_g denote the corresponding generalized Bayes rule. Under suitable regularity conditions it is shown that if, for a constant K,

$$g(|\theta|) \le K|\theta|^{4-p}$$

when $|\theta|$ is large, then δ_g is admissible.

Next, note that if $g(|\theta|) = |\theta|^{c-1}$, g can be approximated as in (1.9) (see Berger, Berliner, and Zaman, 1982). Hence, combining the above results, we see that (generalized Bayes) rules which behave like δ_a are on the "boundary of admissibility." That is, they correspond to generalized prior densities with tails as flat as possible, while preserving admissibility. Such rules are therefore attractive for use in the absence of complete prior information. The reader can find a complete discussion of this reasoning in Berger (1982).

2. The main theorem and related results.

In this chapter a differential inequality for the control problem is obtained. This relation is then used to establish the minimaxity of a large class of estimators. Finally, classes of rules which dominate δ_u are presented.

2.1 *A differential inequality.* The derivation of the differential inequality requires some preliminary computations. First, by employing spherical symmetry, and the implied invariance, some useful representations of risks can be obtained. Since much of the argument here is based on or overlaps with that given in Berger, Berliner, and Zaman (1982), our discussion will be brief.

For an arbitrary, finite risk estimator $\delta(x) = \psi(|x|)x$ and a fixed θ, consider the (artificial) prior π_0 which assigns a uniform probability measures on the p-sphere of radius $|\theta|$. Then $R(\delta, \theta)$ may be viewed as a Bayes risk;

$$R(\delta, \theta) = \int_{\{\eta:|\eta|=|\theta|\}} R(\delta, \eta)\, d\pi_0(\eta) = r(\delta, \pi_0). \tag{2.1}$$

Let $K_0 = (2\pi)^{-p/2}$; also, let $S_{(p)} = 2\pi^{p/2}[\Gamma(\frac{1}{2}p)]^{-1}$ be the surface area of a p-sphere of radius 1. Hence,

$$d\pi_0(\eta) = [S_{(p)}|\theta|^{p-1}]^{-1}\, d\sigma(\eta)$$

where $d\sigma(\eta)$ is the uniform measure on the sphere.

For notational convenience assume $p > 1$. For $p = 1$ the analysis is parallel to that given below. Define $Z(|\theta|)$ by

$$Z(|\theta|) = K_0 S_{(p-1)}|\theta|^{(2-p)}\exp(-\tfrac{1}{2}|\theta|^2).$$

Then applying Formula (1.8) in (2.1) and performing the integration over $(\eta_2, \cdots, \eta_p)$ yield

$$R(\delta, \theta) = Z(|\theta|) \int_{-|\theta|}^{|\theta|} \left[\int_0^\infty r^{(p-1)}\{r\psi(r)\eta_1 - 1\}^2 \exp(r\eta_1) \exp\left(-\frac{1}{2}r^2\right) dr \right] (|\theta|^2 - \eta_1^2)^{(p-3)/2}\, d\eta_1. \tag{2.2}$$

Define the function $\xi[\psi(r)]$ by

$$\xi[\psi(r)] = r^{(p+1)}\psi^2(r)\exp(-\tfrac{1}{2}r^2)$$

and the quantity $\zeta(\psi)$ by

$$\zeta(\psi) = \xi[\psi(r)]\exp(r\eta)|_0^\infty.$$

Also, define the differential operator $\mathscr{D}$ by

$$\mathscr{D}(\psi(r)) = \psi(r)\{2r\psi'(r) + (p + 1 - r^2)\psi(r) + 2\}$$

where

$$\psi'(r) = \frac{d\psi(r)}{dr}.$$

THEOREM 2.1. *Let $\delta_0(x) = \psi_0(|x|)x$ and $\delta_1(x) = \psi_1(|x|)x$. Suppose that both ψ_0 and ψ_1 are continuous, piecewise differentiable functions on* (0, ∞) *such that, for all real η,*

$$\int_0^\infty |\xi'[\psi_0(r)]|\exp(r\eta)\,dr < \infty, \quad \int_0^\infty |\xi'[\psi_1(r)]|\exp(r\eta)\,dr < \infty,$$

and

$$|\zeta(\psi_0)| < \infty, \ |\zeta(\psi_1)| < \infty.$$

If

$$\mathscr{D}(\psi_1(r)) \ge \mathscr{D}(\psi_0(r)), \quad \forall\, r > 0 \tag{2.3}$$

then

$$R(\delta_1, \theta) \le R(\delta_0, \theta), \quad \forall\, \theta.$$

Furthermore, if the inequality (2.3) *holds and is strict on a non-degenerate interval, then δ_1 dominates δ_0.*

PROOF. First, consider the inner integral of the risk representation (2.2); namely,

$$\int_0^\infty r^{(p-1)}\{r^2\psi^2(r)\eta_1^2 - 2r\psi(r)\eta_1 + 1\}\exp(r\eta_1)\exp\left(-\frac{1}{2}r^2\right)dr.$$

Note that the first term of this integral is

$$\eta_1 \int_0^\infty \xi[\psi(r)]\eta_1\exp(r\eta_1)\,dr.$$

When integration by parts is valid, we obtain the equivalent expression

$$\eta_1\left[\zeta(\psi) - \int_0^\infty \left(\frac{d}{dr}\xi[\psi(r)]\right)\right]\exp(r\eta_1)\,dr.$$

Now consider the quantity $\Delta(\theta)$ where $\Delta(\theta) = R(\delta_1, \theta) - R(\delta_0, \theta)$. Since the above integration by parts argument is valid for both ψ_0 and ψ_1, $\Delta(\theta)$ can be written as (after some algebra)

$$\Delta(\theta) = -Z(|\theta|)\int_{-|\theta|}^{|\theta|} \eta_1(|\theta|^2 - \eta_1^2)^{(p-3)/2}$$

$$\left[\int_0^\infty r^p[\mathscr{D}(\psi_1(r)) - \mathscr{D}(\psi_0(r))]\exp(r\eta_1)\exp\left(-\frac{1}{2}r^2\right)dr + \zeta(\psi_0) - \zeta(\psi_1)\right]d\eta_1.$$

Observe that $\zeta(\psi_0)$ and $\zeta(\psi_1)$ are independent of η_1 and that

$$\int_{-|\theta|}^{|\theta|} \eta_1(|\theta|^2 - \eta_1^2)^{(p-3)/2}\, d\eta_1 = 0.$$

Then, clearly, $\Delta(\theta)$ is given by

$$\Delta(\theta) = -Z(|\theta|) \int_0^{|\theta|} \eta_1(|\theta|^2 - \eta_1^2)^{(p-3)/2} \int_0^\infty r^p[\exp(r\eta_1) - \exp(-r\eta_1)]\exp\left(-\frac{1}{2} r^2\right)$$

$$[\mathscr{D}(\psi_1) - \mathscr{D}(\psi_0)]\, dr\, d\eta_1.$$

Since $\exp(r\eta_1) \geq \exp(-r\eta_1)$ for all non-negative r and η_1, it follows that $\Delta(\theta) \leq 0$ if (2.3) holds.

The dominance assertion is obvious. □

It is interesting to note that the inequality (2.3) is not the only relation obtainable by the above type of argument. For example, the unbiased estimate of risk for the control problem has been obtained by Zaman (1977). However, its derivation requires integration by parts twice. Hence, the resulting differential inequality involves both first and second derivatives of ψ_0 and ψ_1.

For most estimators considered below, the integration by parts conditions are easily verified. An exception occurs when $\psi(r) \approx Kr^{-2}$ for r near zero. In this case the integration by parts requirements are violated when $p = 1, 2$. However, this is not a severe restriction since the associated estimator clearly has infinite risk.

2.2. *Minimax control procedures.* Note that for the control problem, the minimax risk is identically equal to one for all p. A proof goes as follows. First, note that the control problem loss function (1.1) is a polynomial in the coordinates of θ. Then, since the normal distribution is a member of the exponential family, the control risk function is continuous in θ (by continuity of the Laplace transform). Observing that for any δ, $R(\delta, \theta) = 1$ for $\theta = (0, \cdots, 0)^t$, it is clear that the risk can never be bounded away from one.

THEOREM 2.2. *Suppose $\delta(x) = \psi(|x|)x$ where ψ satisfies the conditions of Theorem* 2.1. *Define the differential operator $\mathscr{D}^M$ by*

$$\mathscr{D}^M(\psi(r)) = 2r\psi'(r) + (p + 1 - r^2)\psi(r) + 2.$$

If $\psi(r) \geq 0$, $\forall\, r \geq 0$, and $\mathscr{D}^M(\psi(r)) \geq 0$, $\forall\, r \geq 0$, then δ is minimax.

PROOF. The proof follows directly from Theorem 2.1, after noting that $\delta(x) \equiv 0$ is minimax. □

The minimaxity of an interesting class of estimators is considered in the following theorem.

THEOREM 2.3. *Let a and b be constants such that* $0 \leq a \leq 2$, $b \geq 0$. *Let δ be defined by $\delta(x) = \psi(|x|)x = a(b + |x|^2)^{-1}x$. Then δ is minimax in the following cases*: (i) $p = 1$, $b > 0$, $a \leq 4b(2 + b)^{-1}$, (ii) $p = 2$, $b > 0$, $a \leq 4b(1 + b)^{-1}$, *and* (iii) $p \geq 3$, $b \geq 0$.

PROOF. The proof is a direct application of Theorem 2.2 (note that for $p = 1, 2$; $b > 0$). Hence, we must verify that $\mathscr{D}^M(\psi(r)) \geq 0$, $\forall\, r$. Computation and simplification yield the equivalent inequality

$$(2 - a)r^4 + [(p - 3)a + (4 - a)b]r^2 + (p + 1)ab + 2b^2 \geq 0; \quad \forall\, r \geq 0. \tag{2.4}$$

Case (iii) follows immediately. Next, let $p = 1$. Then (2.4) holds when $-2a + (4 - a)b \geq 0$ or $a \leq 4b(2 + b)^{-1}$. Case (ii) is established in the same fashion. □

Note that Cases (i) and (ii) are not the most general statements possible, based on the sufficient condition (2.4). They do, however, suit our purposes. In particular, note that for all p, minimaxity is implied when $0 < a = b \leq 2$. Hence, the uniform measure, generalized Bayes rule δ_u (see (1.3)) is minimax. Theorem 2.3 also implies an interesting result concerning the normal prior Bayes rule δ_τ (see (1.7)). It is easy to check that, for all p, if $\tau \geq 1$, δ_τ is minimax.

Also, note that is clear from Case (iii) above that if $p \geq 3$, δ_m (see (1.4)) is minimax.

2.3. *Domination of* δ_u. The first theorem below is a generalization of the results of Takeuchi (1968) and Stein and Zaman (1980). Recall that δ_u is admissible when $p \leq 4$.

THEOREM 2.4. *Let* $\delta_m(x) = \psi(|x|)x = |x|^{-2}x$. *Define* $\delta_c(x) = \psi_c(|x|)x = (c + |x|^2)^{-1}x$, *for* $c > 0$, *a constant. Assume that* $p \geq 3$. *If* $2(p - 5) + c \geq 0$, *then* δ_m *dominates* δ_c.

PROOF. Substitution of ψ and ψ_c into the differential inequality (2.3) yields

$$\begin{aligned} & r^{-2}[2r(r^{-2})' + (p + 1 - r^2)r^{-2} + 2] \\ & \qquad \geq (c + r^2)^{-1}[2r[(c + r^2)^{-1}]' + (p + 1 - r^2)(c + r^2)^{-1} + 2] \end{aligned} \tag{2.5}$$

or, equivalently,

$$[2(p - 5) + c]r^4 + c[3(p - 3) + c]r^2 + (p - 3)c^2 \geq 0. \tag{2.6}$$

Clearly, (2.6) is true for all $r \geq 0$ if $2(p - 5) + c \geq 0$. Since $c > 0$, the inequality (2.6) is in fact strict. Theorem 2.1 then implies the result. □

Theorem 2.4 implies that δ_m dominates δ_c for all $c > 0$ when $p \geq 5$. Takeuchi (1968) gave the same result but for $p \geq 6$. Stein and Zaman (1980) proved that δ_m dominates δ_c for $c = 1$ (i.e., $\delta_c = \delta_u$) when $p = 5$. The results that δ_m dominates δ_c if (i) $p = 4$ and $c \geq 2$, or (ii) $p = 3$ and $c \geq 4$ are new, but not surprising in light of the inadmissibility results of Berger, Berliner, and Zaman (1982) discussed in Section 1.2.

THEOREM 2.5. *Assume* $p \geq 5$. *Let* $\delta(x) = \psi(|x|)x$ *where* $\psi(r) = (1 + r - g(r))^{-1}$ *Suppose that* $\psi(r) \geq 0$ *for all* $r \geq 0$ *and that* ψ *satisfies the conditions of Theorem* 2.1. *If*

$$\begin{aligned} & 2rg'(r) + (1 + r^2)^{-3}g(r)\{[2(p - 4) - g(r)]r^6 \\ & \qquad + [g^2(r) - 3(p - 1)g(r) + 6(p - 2)]r^4 \\ & \qquad + [pg^2(r) - 3(2p + 1)g(r) + 6p]r^2 \\ & \qquad + [(p + 3)g^2(r) - (3p + 7)g(r) + 2(p + 2)]\} \geq 0, \end{aligned} \tag{2.7}$$

for all $r > 0$, *with strict inequality on a non-degenerate interval, then* δ *dominates* δ_u.

PROOF. The proof is a direct application of Theorem 2.1 and is therefore omitted. □

Though Theorem 2.5 is rather general, the key inequality (2.7) is quite complex. Hence, the following more readily applicable, though less general, result is presented.

THEOREM 2.6. *Assume* $p \geq 5$. *Let* $\delta(x) = (1 + |x|^2 - g(|x|))^{-1}x$. *Define the quantity* $T(p)$ *by*

$$\begin{aligned} T(p) = {} & (p - 4)^{-2}\{2p^2 + 9.5 \\ & - p[10 + \{6 - (2p)^{-1}[3p + 7 - 2[2(p^2 + 5p + 6)]^{1/2}]\}^{1/2}]\}. \end{aligned} \tag{2.8}$$

If (i) $g(r)$ *is continuous, piece-wise differentiable,* (ii) $g(r)$ *is not identically zero, and* $g'(r) \geq 0$, *and* (iii) $0 \leq g(r) \leq \min\{p-4; (\max\{T(p); 2\})^{-1}r^2\}$, *for all* $r > 0$, *then* δ *dominates* δ_u.

PROOF. The proof is an application of Theorem 2.5. Since g is non-decreasing and not identically 0, the inequality (2.7) reduces to (suppressing the dependence of g on r).

$$\begin{aligned}&[2(p-4)-g]r^6 + [g^2 - 3(p-1)g + 6(p-2)]r^4 \\ &\quad + [pg^2 - 3(2p+1)g + 6p]r^2 + [(p+3)g^2 - (3p+7)g + 2(p+2)] \geq 0.\end{aligned} \tag{2.9}$$

Let $P(r)$ denote the quantity on the L.H.S. of (2.9). Let ℓ and k be non-negative constants. Also let $c = \max\{T(p); 2\}$.

Assumption (iii) implies that $r \geq cg \geq 2g$. Simple manipulation then implies that

$$\begin{aligned}P(r) \geq \{&[2(p-4)-g]r^2 + g^2 - 3(p-1)g + (6-\ell)(p-2)\}r^4 \\ &+ \{pg^2 - [3(2p+1) - 2(p-2)\ell]g + (6-k)p\}r^2 \\ &+ \{(p+3)g^2 - [3p+7-2kp]g + 2(p+2)\}.\end{aligned} \tag{2.10}$$

Let the three terms in brackets ($\{\}$) be denoted $A(\ell, k)$, $B(\ell, k)$, and $C(k)$ respectively. Values of ℓ and k will be found such that these three terms are non-negative.

First, define k_0 by

$$k_0 = (2p)^{-1}[3p + 7 - 2[2(p+2)(p+3)]^{1/2}].$$

Note that it is easy to check that $0 < k_0 < 6$ for all $p \geq 5$. Then, clearly,

$$C(k_0) = (p+3)[g - \{2(p+2)/(p+3)\}^{1/2}]^2 \geq 0.$$

Next, define ℓ_0 by

$$\ell_0 = [2(p-2)]^{-1}\{3(2p+1) - 2p(6-k_0)^{1/2}\}.$$

Again, it is easy to check that $0 < \ell_0 < 6$ for all $p \geq 5$. Then

$$B(\ell_0, k_0) = p[g - (6-k_0)^{1/2}]^2 \geq 0.$$

Finally, consider $A(\ell_0, k_0)$. Since $r^2 \geq cg$ and $g \leq p-4$, it is clear that

$$A(\ell_0, k_0) \geq (1-c)g^2 + [2(p-4)c - 3(p-1)]g + (6-\ell_0)(p-2). \tag{2.11}$$

Let $Q(g)$ denote the quantity on the R.H.S. of (2.11). Since $Q(g)$ is quadratic in g and $c \geq 2$, it suffices to show that $Q(0)$ and $Q(p-4)$ are non-negative. Clearly, $Q(0) = (6-\ell_0)(p-2) > 0$. Next

$$Q(p-4) = (c+1)(p-4)^2 - 3(p-1)(p-4) + (6-\ell_0)(p-2).$$

Simple algebra reduces $Q(p-4) \geq 0$ to the inequality $c \geq T(p)$ where $T(p)$ is defined in (2.8). Hence $P(r) \geq 0$ for all $r \geq 0$.

Inspection of $P(r)$ (see (2.9)) implies that $P(r) > 0$ for all r sufficiently large. Hence, δ dominates δ_u. □

For convenience, the function $\max\{T(p); 2\}$ is given in Table 1.

2.4 *Remarks.* 1. It is interesting to compare the domination results presented here with analogous results for the problem of estimating a p-variate normal mean under quadratic loss. Since the work of Stein (1955) and James and Stein (1960), a large body of literature has been devoted to the domination of the usual maximum likelihood (or least squares) estimator δ_0 when $p \geq 3$. For our purposes, consider spherically symmetric estimators of the form $\delta(x) = \{1 - \tau(|x|^2)|x|^{-2}\}x$. Under regularity conditions permitting

TABLE 1

Values of max $\{T(p); 2\}$

p	max $\{T(p); 2\}$	p	max $\{T(p); 2\}$
5	2.000000	13	2.242694
6	2.000000	14	2.229505
7	2.000000	15	2.216882
8	2.199960	20	2.166817
9	2.259628	30	2.111882
10	2.271731	50	2.066866
11	2.266819	∞	2.000000
12	2.255721		

integration by parts, it can be shown that if $p \geq 3$ and

$$4\tau'(|x|^2) + \tau(|x|^2)|x|^{-2}[2(p-2) - \tau(|x|^2)] \geq 0, \tag{2.12}$$

with strict inequality on an interval, then δ dominates δ_0. This result is due to Efron and Morris (1976). (Their result is actually more general since their proof does not use integration by parts.) The analogy between (2.12) and (2.7) of Theorem 2.5 is clear. It is interesting to compare the relative complexity of conditions (2.7) and (2.12). Also, note that if $\tau' \geq 0$, then (2.12) implies that if $0 \leq \tau(|x|^2) \leq 2(p-2)$, with strict inequality on an interval, then δ dominates δ_0 (Baranchik, 1970). This case is analogous to our Theorem 2.6.

2. It is suspected that the requirement that $g(r) \leq p - 4$ implicit in Theorem 2.6 is stronger than necessary. However, the result has been tailored for use in proposing control procedures. Therefore, this bound does not impose a serious limitation in that rules which attain this bound are asymptotically optimal (see Section 1.2).

3. *New control procedures.* The results of Chapter 2 are now employed to propose control procedures which dominate $\delta_u (p \geq 5)$. It is assumed that the decision maker is unwilling to assume perfect prior information, ruling out a standard Bayesian solution. It is then common to evaluate proposed rules in terms of minimax criteria and admissibility. However, the results of Section 2.2 indicate that the minimax principle is not a very discriminating criterion in the control problem. (All rules proposed below are minimax.) Attention is given to admissibility and asymptotic optimality (see Section 1.2). It is also reasonable to require that a rule be relatively simple to compute. This is the case for the rules below.

The method of construction of estimators is Bayesian. However, the particular generalized prior measure employed may be considered as merely a tool. Within the scope of this paper, it is not necessary to consider, in depth, Bayesian justifications of the prior or the resulting rules. The prior (or versions of it) has been useful in the problem of estimating a multivariate normal mean. In particular, see Strawderman (1971), Berger (1976), (1980b). For convenience, results from Berger (1980b), required in the analysis here, are presented in the Appendix.

Define a generalized prior density $g_n(\theta)$ by

$$\begin{aligned} g_n(\theta) = (2\pi)^{-p/2} \int_0^1 & [(\rho(1+a)/\lambda - 1)]^{-p/2} \\ & \exp\left\{-\frac{1}{2}[(\rho(1+a)/\lambda - 1)]^{-1}|\theta|^2\right\}\lambda^{(n-1-p/2)}\, d\lambda, \end{aligned} \tag{3.1}$$

where $\rho > 0$, $a > 0$, and $n \geq \frac{1}{2}$. Also, define the function r_n by

$$r_n(v) = \frac{v \int_0^1 \lambda^n \exp\left(-\frac{1}{2} v\lambda\right) d\lambda}{\int_0^1 \lambda^{(n-1)} \exp\left(-\frac{1}{2} v\lambda\right) d\lambda}. \tag{3.2}$$

Let $\| x \|^2 = [\rho(1 + a)]^{-1} | x |^2$. In Lemma A.3 of the Appendix, it is shown that the generalized Bayes control rule δ^n corresponding to g_n is given by

$$\delta^n(x) = \{1 + | x |^2 - s_n(\| x \|^2)\}^{-1} x \tag{3.3}$$

where

$$s_n(v) = r_n(v)[\rho(1 + a)v - r_{n+1}(v)]/[\rho(1 + a)v - r_n(v)]. \tag{3.4}$$

We now find values of the constants n, ρ, and a such that the resulting rule satisfies the desired properties. First, Lemma A.2, iii, implies that

$$\delta^n(x) \approx \{1 + | x |^2 - 2n\}^{-1} x$$

for $| x |$ large. Therefore, if $n = \frac{1}{2}(p - 4)$ the resulting rule, denoted δ^*, is asymptotically optimal. Furthermore:

THEOREM 3.1. *If $p \geq 5$, then δ^* is admissible.*

The proof is a direct application of the results of Berger, Berliner, and Zaman (1982) and is, therefore, omitted. It is interesting to note that for $n = \frac{1}{2}(p - 4)$, $g^n = g^*$ is given by

$$g^*(\theta) \approx K | \theta |^{(4-p)}$$

for $| \theta |$ large and K a constant (Berger, 1980b). Hence, δ^* is on the "boundary of admissibility" alluded to in Section 1.2.

The next step is to find conditions under which δ^* dominates δ_u. This is done by applying Theorem 2.6.

THEOREM 3.2. *Assume $p \geq 5$. If*

$$\rho(1 + a) \geq (p - 4)(p - 2)^{-1} \max\{T(p); 2\}, \tag{3.5}$$

where $T(p)$ is defined by (2.8), *then δ^* dominates δ_u.*

PROOF. Let $c = \rho(1 + a)$. Let $s^* = s_n$ and $r^* = r_n$ when $n = (p - 4)/2$. First, it is clear that (see Lemma A.2) s^* is continuous, differentiable, and non-negative. To verify that s^* satisfies the conditions of Theorem 2.6, it is sufficient to check that: (i) $s^* \leq \min\{p - 4; (\max\{T(p); 2\})^{-1} | x |^2\}$, and (ii) $s^{*\prime}(v) \geq 0$.

(i) First, since $r_{n+1}(v) - r_n(v) > 0$ (Lemma A.2vii), it follows that

$$[cv - r_{n+1}(v)]/[cv - r_n(v)] \leq 1.$$

Therefore, it is sufficient to verify that (i) holds with s^* replaced by r^*. Lemma A.2i, implies immediately that $r^* \leq p - 4$. Also, Lemma A.2v, implies that

$$r^*(v)/v \leq (p - 4)/(p - 2).$$

A simple manipulation and (3.5) then yield the result.

(ii) Note that $s_n'(v)$ is given by (suppressing the dependence on v)

$$\begin{aligned} s_n' &= r_n'(cv - r_{n+1})/(cv - r_n) \\ &\quad + r_n\{(c - r_{n+1}')/(cv - r_n) - (c - r_n')(cv - r_{n+1})/(cv - r_n)^2\}. \end{aligned} \tag{3.6}$$

Employing Lemma A.2vi, (3.6) can be written as

$$s_n' = (s_n/v)\{1 - (r_{n+1} - r_n)/2 + 1 + (r_{n+1}/2)((r_{n+2} - r_{n+1})/(cv - r_{n+1})) - 1 - (r_n/2)((r_{n+1} - r_n)/(cv - r_n))\}$$

or

$$s_n' = (s_n/v)\{1 - (s_{n+1} - s_n)/2\}.$$

Hence, it suffices to show that $s_{n+1} - s_n \leq 2$. It is easy to check that $s_{n+1} - s_n \leq 2$ is equivalent to

$$2 - (3r_{n+1}/cv) - (r_n/cv) + (r_{n+1}/(cv)^2)(r_{n+2} + r_n) \geq 0. \tag{3.7}$$

Lemma A.2v, implies that (3.7) holds if

$$2 - 3(n+1)/c(n+2) - n/c(n+1) \geq 0.$$

Hence, it is enough if

$$2(n+1)/c(n+2) \leq 1. \tag{3.8}$$

For $n = (p-4)/2$, (3.8) reduces to $c \geq 2(p-2)/p$, which is trivially satisfied by hypothesis. □

Theorem 3.2 only provides a constraint on the values of a and ρ under which δ^* dominates δ_u. One possible choice for these constants is now suggested. This discussion is based on Berger's (1980b) arguments concerning the prior measure used here. Note that by Lemma A.2iv, δ^n is given by

$$\delta^n(x) \approx \{1 + [1 - (n/(\rho(1+a)(n+1)))\,|x|^2]\}^{-1}x \tag{3.9}$$

for $|x|$ small. Let $\delta_a^*(x)$ denote $\delta^n(x)$ when $n = (p-4)/2$ and $\rho = n/(n+1)$. For the remainder of the discussion, assume that $a \geq \max\{T(p); 2\} - 1$ so that, by Theorem 3.2, δ_a^* dominates δ_u. Note that (3.9) implies that, for $|x|$ small,

$$\delta_a^*(x) \approx \{1 + [1 - (a+1-d)/(a(1+a))]\,|x|^2\}^{-1}x,$$

where $d = (p-2)^2/(p(p-4))$. Now let the function $\tau^2(a)$ be implicitly defined by the equation

$$(1 + \tau^2(a))^{-1} = (1+a)^{-1}(1 + (1-d)/a). \tag{3.10}$$

Then for a fixed a and for small $|x|$, δ_a^* is approximately equal to the normal prior Bayes rule with prior mean $(0, \cdots, 0)^t$ and prior convariance $\tau^2(a)I_p$ (see (1.7)). (The prior g^* is in some sense a "robust" or flat-tailed version of the above normal prior.) (Note that $\tau^2(a)$ is increasing in a if $p \geq 6$. If $p = 5$ $\tau^2(a)$ is decreasing in a for $1 \leq a < 2$ and increasing for $a > 2$. This anomaly could be avoided by a suitable redefinition of ρ when $p = 5$.)

Intuitively this Bayesian interpretation suggests two possibilities. First, we might expect the greatest improvement in risk of δ_a^* over δ_u to occur when $|\theta|$ is small. This is not the case, however, since all risk functions are driven to 1 as $|\theta| \to 0$ (see Section (2.2)). Second, as a and hence $\tau^2(a)$ increase (moderated by the above remark if $p = 5$ and $1 \leq a < 2$), the "size" of the region of Θ in which δ_a^* offers significant improvement over δ_u is expected to increase, but the maximum amount of improvement should decrease. This observation is true, as can easily been seen from the analysis in Chapter 2.

Table 2 presents some risk computations for the case $p = 5$. First, the risk of δ_u, $R(\delta_u, |\theta|)$ is given for various values of $|\theta|$. Then the quantity PIR, the percent of improvement in risk, defined by

$$\mathrm{PIR}(\delta_a^*, \delta_u) = 1 - R(\delta_a^*, |\theta|)/R(\delta_u, |\theta|)$$

is given for several values of a. All entries were computed numerically.

TABLE 2

$R(\delta_u, |\theta|)$ and PIR (δ_a^*, δ_u): p = 5

$\lvert\theta\rvert$	$R(\delta_u, \lvert\theta\rvert)$	PIR (δ_a^*, δ_u) $a = 1.0$	$a = 1.5$	$a = 2.0$	$a = 2.5$	$a = 3.0$
.25	.98440	.00147	.00156	.00159	.00157	.00153
.50	.93364	.00590	.00626	.00635	.00629	.00615
.75	.85584	.01333	.01410	.01430	.01416	.01384
1.00	.75976	.02376	.02504	.02537	.02514	.02459
1.25	.65519	.03710	.03890	.03936	.03903	.03823
1.50	.55113	.05300	.05524	.05583	.05541	.05436
1.75	.45449	.07084	.07335	.07403	.07352	.07229
2.00	.36934	.08962	.09215	.09285	.09232	.09099
2.25	.29714	.10817	.11042	.11106	.11055	.10925
2.50	.23752	.12520	.12688	.12738	.12696	.12585
2.75	.18933	.13932	.14022	.14052	.14024	.13946
3.00	.15137	.14875	.14879	.14885	.14875	.14837
3.25	.12256	.15126	.15052	.15035	.15042	.15047
3.50	.10170	.14526	.14396	.14362	.14383	.14425
3.75	.08705	.13185	.13031	.12988	.13018	.13083
4.00	.07640	.11549	.11401	.11358	.11390	.11464
4.25	.06748	.10144	.10018	.09980	.10009	.10083
4.50	.05874	.09248	.09150	.09120	.09144	.09211
4.75	.04983	.08788	.08715	.08691	.08712	.08770
5.00	.04156	.08371	.08319	.08301	.08318	.08368
5.25	.03522	.07426	.07392	.07380	.07393	.07432

Note that the maximum PIR in Table 2 is about 15%. At first glance this improvement seems small (especially when compared to the success of Stein-type estimators in the problem of the estimation of a multivariate normal mean). However, two points should be raised. First, employing standard approximation techniques and Lemma A.2viii, ix, it can be shown that the PIR defined above is approximately given by

$$\text{PIR} \simeq 1 - (p + |\theta|^2)^2/(p|\theta|)^2$$

for p very large. The maximum indicated PIR is then $1 - 4/p$, attained at $|\theta|^2 = p$. Hence, substantial improvement is obtainable.

The second key point concerns the limitations of spherical symmetry. The transformation (alluded to in Section 1.1) yielding the problem considered here defines the normally distributed random vector X and its mean θ by

$$X = \Lambda^{-1/2}\hat{\beta}; \quad \theta = \Lambda^{-1/2}\beta$$

where $\hat{\beta} \sim N_p(\beta, \Lambda)$ is a estimate of the regression coefficients of some underlying linear model (whose output is to be controlled). See Zaman (1981) for details. Hence, any spherically symmetric Bayes rule incorporates rather unrealistic (perhaps silly) prior information. First, if $|\theta|$ was really believed to be near zero, one would probably not be attempting control. Second, a prior covariance matrix for θ proportional to the identity (corresponding to a prior covariance matrix for β that is a function of the design matrix Λ) is quite unlikely. It then seems that the incorporation of realistic prior information would usually require consideration of non-symmetric priors. Non-symmetric versions of g^* and δ_a^* are readily obtainable. See Berger (1980b) for non-symmetric analogs of g^* and Lemma A.1. Minor modifications of Lemma A.3 then yield the corresponding control rules. However, obtaining extensions of the mathematical results of Section 2 in such cases appears to be quite difficult.

APPENDIX

LEMMA A.1. (Berger, 1980b). *Assume* $X \sim N_p(\theta, I)$. *Let* $E(\theta \mid X = x)$ *and* $\operatorname{Cov}(\theta \mid X = x)$ *denote the posterior mean and covariance of* θ w.r.t. *the prior measure defined by* (3.1). *Then*

$$E(\theta \mid x) = (1 - r_n(\| x \|^2) | x |^{-2}) x \tag{A.1}$$

and

$$\begin{aligned} \operatorname{Cov}(\theta \mid X = x) &= (1 - r_n(\| x \|^2) | x |^{-2}) I \\ &\quad + \{ r_n(\| x \|^2)[r_{n+1}(\| x \|^2) - r_n(\| x \|^2)] | x |^{-4} \} x x^t. \end{aligned} \tag{A.2}$$

LEMMA A.2. (Berger, 1980b). *If* $n > 0$, *then* (i) $0 < r_n(v) < 2n$, (ii) $r_n'(v) > 0$, (iii) $\lim_{v \to \infty} r_n(v) = 2n$, (iv) $\lim_{v \to 0} r_n(v)/v = n/(n + 1)$, (v) $r_n(v)/v \le n/(n + 1)$, (vi) $r_n'(v) = \{r_n(v)/v\}[1 - \frac{1}{2}\{r_{n+1}(v) - r_n(v)\}]$, (vii) $0 < r_{n+1}(v) - r_n(v) < 2$, (viii) $\lim_{n \to \infty} r_n(v) = v$, and (ix) $\lim_{n \to \infty}[r_n(2nc)/(2n\{\min(1, c)\})] = 1$.

LEMMA A.3. *The generalized Bayes control rule* δ^n *corresponding to* g_n *is given in Equation* (3.3).

PROOF. Combining (1.6) and Lemma A.1, δ^n can be written as

$$\delta^n(x) = \{I + W(| x |) x x^t\}^{-1} x \tag{A.3}$$

where

$$\begin{aligned} W(| x |) &= 1 - r_n(\| x \|^2) | x |^{-2} \\ &\quad + \{ r_n(\| x \|^2)[r_{n+1}(\| x \|^2) - r_n(\| x \|^2)]\} / \{(1 - r_n(\| x \|^2) | x |^{-2}) | x |^4\}. \end{aligned} \tag{A.4}$$

Consider the following well-known identity for a non-singular, $p \times p$ matrix B and p dimensional column vectors E and F:

$$(B + EF^t)^{-1} = B^{-1} - [1 + F^t B^{-1} E]^{-1} (B^{-1} E)(F^t B^{-1}).$$

Application of this identity in (A.4) yields

$$\delta^n(x) = \{1 + W(| x |) | x |^2\}^{-1} x.$$

Simplifying W (suppressing the dependence on $\| x \|^2$) implies

$$\begin{aligned} \delta^n(x) &= \{1 - r_n + | x |^2 + [r_n(r_{n+1} - r_n)/(| x |^2 - r_n)]\}^{-1} x \\ &= \{1 + [r_n r_{n+1} + | x |^4 - 2 | x |^2 r_n]/(| x |^2 - r_n)\}^{-1} x \\ &= \{1 + | x |^2 - r_n(| x |^2 - r_{n+1})/(| x |^2 - r_n)\}^{-1} x \end{aligned}$$

as was to be shown. □

Acknowledgments. I wish to express my most sincere appreciation to Professor James Berger, whose aid and direction were invaluable in the course of this research. I am grateful to Sukhoon Lee for assistance in the numerical work. My thanks also go to an associate editor and a referee for their very useful comments.

REFERENCES

AOKI, M. (1967). *Optimization of Stochastic Systems.* Academic, New York.

BARANCHIK, A. J. (1970). A family of minimax estimators of the mean of a multivariate normal distribution. *Ann. Math. Statist.* **41** 642–645.

BASU, A. (1974). Control level determination in regression models. Technical Report No. 139, Economic Series, Institute for Mathematical Studies in the Social Sciences, Stanford University.

BERGER, J. (1976). Admissible minimax estimation of a multivariate normal mean with arbitrary quadratic loss. *Ann. Statist.* **4** 223–226.

BERGER, J. (1980a). Improving on inadmissible estimators in continuous exponential families with applications to simultaneous estimation of gamma scale parameters. *Ann. Statist.* **8** 545–571.

BERGER, J. (1980b). A robust generalized Bayes estimator and confidence region for a multivariate normal mean. *Ann. Statist.* **8** 716–761.

BERGER, J. (1982). The robust Bayesian viewpoint. Mimeograph Series #82-27, Statistics Department, Purdue University.

BERGER, J., BERLINER, L. M., and ZAMAN, A. (1982). General admissibility and inadmissibility results for estimation in a control problem. *Ann. Statist.* **10** 838–856.

BERLINER, L. M. (1981). Uniform improvements on the certainty equivalent rule in a statistical control problem. S. S. Gupta and J. Berger (Eds.). *Statistical Decision Theory and Related Topics III.* Academic, New York.

DUNSMORE, I. R. (1969). Regulation and optimization. *J. Roy Statist. Soc. Ser. B* **31** 160–170.

EFRON, B. and MORRIS, C. (1976). Families of minimax estimators of the mean of a multivariate normal distribution. *Ann. Statist.* **4** 11–21.

HAFF, L. (1977). Minimax estimators for a multinormal precision matrix. *J. Multivariate Anal.* **7** 374–385.

HUDSON, H. (1978). A natural identity for exponential families with applications in multiparameter estimation. *Ann. Statist.* **6** 473–484.

JAMES, W. and STEIN, C. (1960). Estimation with quadratic loss. *Proc. Fourth Berkeley Symp. Math. Statist. Probab.* **1** 361–380. Univ. of California Press.

LINDLEY, D. V. (1968). The choice of variables in multiple regression. *J. Roy. Statist. Soc. Ser. B* **30** 31–66.

STEIN, C. (1955). 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.

STEIN, C. (1973). Estimation of a mean of a multivariate distribution. *Proc. Prague Symp. Asymptotic Statist.* 345–381.

STEIN, C. and ZAMAN, A. (1980). On the admissibility of the uniform prior estimator for the control problem: The case of dimensions four and five. Technical Report, Department of Statistics, Stanford University.

STRAWDERMAN, W. E. (1971). Proper Bayes minimax estimators of the multivariate normal mean. *Ann. Math. Statist.* **42** 385–388.

TAKEUCHI, K. (1968). On the problem of fixing the level of independent variables in a linear regression function. IMM 367, Courant Institute of Math. Sciences, New York University.

ZAMAN, A. (1977). An application of Stein's method to the problem of single period control of regression models. Technical Report 231, Economic Series, Stanford University.

ZAMAN, A. (1981). A complete class theorem for the control problem, and further results on admissibility and inadmissibility. *Ann. Statist.* **9** 812–821.

ZELLNER, A. (1971). *An Introduction to Bayesian Inference in Econometrics*, Wiley, New York.

ZIDEK, J. (1978). Deriving unbiased risk estimators of multinormal mean and regression coefficient estimators using zonal polynomials. *Ann. Statist.* **6** 769–782.

DEPARTMENT OF STATISTICS
THE OHIO STATE UNIVERSITY
COLUMBUS, OHIO 43210