The Annals of Statistics
1981, Vol. 9, No. 4, 812–821

A COMPLETE CLASS THEOREM FOR THE CONTROL PROBLEM AND FURTHER RESULTS ON ADMISSIBILITY AND INADMISSIBILITY[1]

BY ASAD ZAMAN

Stanford University

The following decision problem is studied. The statistician observes a random n-vector y normally distributed with mean β and identity covariance matrix. He takes action $\delta \in \mathbb{R}^n$ and suffers the loss

$$L(\beta, \delta) = (\beta'\delta - 1)^2.$$

It is shown that this is equivalent to the linear control problem and closely related to the calibration problem. Among the invariant estimators, it is shown that the formal Bayes rules together with some of their limits include all admissible invariant rules. Other results on admissiblity and inadmissibility of some commonly used estimators for the problem are obtained.

1. Introduction. In this paper we study the following decision problem. The statistician observes y, a normal random p-vector with mean β and covariance I. He takes action $\delta \in \mathbb{R}^p$ and suffers the loss

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

The parameter space is taken to be $\mathbb{R}^p/\{0\}$. We exclude the origin since if $\beta = 0$, the decision vector has no effect on the loss.

In Section 2 we obtain a class of estimators which includes all admissible orthogonally invariant decision rules for this problem. In Sections 3 and 4 we derive results on the admissibility and inadmissibility of some commonly used procedures.

This kind of a loss function arises in a variety of problems. For example, suppose a system is described by

$$z = \gamma'x + \epsilon, \tag{1.2}$$

where z is a scalar dependent variable, γ is a vector of unknown parameters, x is a vector of nonstochastic variables directly under the control of the experimenter, and ϵ is scalar normal with mean 0 and variance 1. The experimenter desires to choose the vector x so as to make z close to some desired level z^*. He has an estimate of γ, $\hat{\gamma} \sim N(\gamma, \Lambda)$, possibly from previous observations of the same system. Assuming he is using a quadratic loss criterion, the risk of a decision rule $x(\hat{\gamma})$ is given by:

$$R(\gamma, x(\hat{\gamma})) = E[z^* - \{\gamma'x(\hat{\gamma}) + \epsilon\}]^2 = 1 + (z^*)^2 E\left\{1 - \frac{\gamma'x(\hat{\gamma})}{z^*}\right\}^2.$$

Let $y = \Lambda^{-1/2}\hat{\gamma}$, $\beta = \Lambda^{-1/2}\gamma$ and $\delta(y) = (1/z^*)\Lambda^{1/2}x(\hat{\gamma})$. Then the problem transforms to the canonical form (1.1) up to a linear transformation. In this situation we can also pose the inverse regression problem. We observe z^* and wish to know what x gave rise to it in equation (1.2). This problem can also lead to the loss function given in (1.1).

Received August, 1977; revised October, 1980.

[1] This work was supported by National Science Foundation Grant SOC73-05453-A02 at the Institute for Mathematical Studies in the Social Sciences, Stanford University.

AMS 1970 subject classification numbers. 62C07; 62C10; 62C15.

Key words and phrases. Complete class, control problem, invariance, formal Bayes, admissibility.

Finally, in the one-dimensional case, the loss function can be written as $L(\beta, \delta) = \beta^2(\delta - \beta^{-1})^2$. All results obtained here are valid for the problem of estimating $1/\beta$ with quadratic loss, given we observe $y \sim N(\beta, 1)$. This is discussed in Section 5.

2. A complete class theorem. Note that the problem is invariant under the following group of transformations: $y \to Py$, $\beta \to P\beta$, $\delta \to P\delta$, where P is an orthogonal matrix. We will obtain a complete class of invariant decision rules, i.e., rules satisfying $P\delta(y) = \delta(Py)$ for all orthogonal P. Since this is a compact group, admissibility within the class invariant rules implies admissibility (see, for example, Theorem 4.3.2 in Ferguson (1967)). Also, if an invariant rule is formal Bayes, there exists an invariant prior with respect to which it is formal Bayes. A prior measure Π is invariant if, for all orthogonal P, $\Pi\{A\} = \Pi\{PA\}$ where $PA = \{y \mid \exists\, x \in A \text{ such that } y = Px\}$.

The main result of this section is:

THEOREM 1. (A Complete Class Theorem for Spherically Symmetric Rules). *If $\delta(y)$ is invariant and admissible, there exists a probability measure μ concentrated on* $[0, \infty)$, *such that*

$$\delta(y) = \left\{\int_0^\infty \gamma^{-1} \sinh \gamma\|y\|\, d\mu(\gamma)\right\}\left\{\int_0^\infty \cosh \gamma\|y\|\, d\mu(\gamma)\right\}^{-1} \frac{y}{\|y\|}, \tag{2.1}$$

where, for $\gamma = 0$, $(1/\gamma) \sinh \gamma\|y\| = \|y\|$.

REMARKS. The strategy of the proof is the same as that used by Brown (1971) to show that the set of formal Bayes rules forms a complete class for the estimation of the multinormal mean. We first show that all admissible orthogonally invariant rules are pointwise limits of proper Bayes rules (Lemma 3). This illustrates an idea of Stein's referred to in his concluding remarks in Stein (1964). Zidek (1967) presents a general theorem of this kind which, unfortunately, does not apply to the control problem since the loss is not strictly convex. The second half of the proof demonstrates that admissible rules which are pointwise limits of proper Bayes rules can be represented as in (2.1) above.

PROOF. We shall use a series of lemmas. First we introduce some useful notation. Throughout the paper, we will use f to denote the multivariate normal density; $f(x) = (2\pi)^{-p/2} \exp(-\|x\|^2/2)$, where p is the dimension of vector x. For an invariant prior Π. we shall denote by $\tilde{\Pi}$ the marginal distribution of the first component of the measure $f(\beta)\, d\Pi(\beta)$. Thus, $\tilde{\Pi}(A) = \int_A \int_{R^{p-1}} f(\beta)\, d\Pi(\beta)$. We shall also need a version of the Laplace transform of Π, which we shall write as $F_\Pi(y'y) = \int \exp(\gamma\|y\|\, d\Pi(\gamma)$. F_Π differs from the Laplace transform as its argument is $\|y\|^2$, and not $\|y\|$. The relationship of F_Π to the original measure Π can easily be shown to be

$$F_\Pi(y'y) = \int \exp(\beta' y) f(\beta)\, d\Pi(\beta). \tag{2.2}$$

In the sequel, we shall use F' and F'' to denote the first and second derivatives of $F_\Pi(y'y)$, omitting the argument and the associated measure Π wherever these are clear from the context.

$\delta_\Pi(y)$ is defined to be a formal Bayes rule with respect to an improper prior Π if it minimizes $\int (\beta'\delta - 1)^2 f(y - \beta)\, d\Pi(\beta)$, provided that for each y, this is finite for some δ. If this integral is finite for each y, then it can easily be seen that F_Π is finite, the posterior distribution exists, and the posterior expected loss is finite for some δ.

We shall also use the customary notations $R(\beta, \delta)$ for the risk function of δ, and $B(\delta, \Pi)$ for the Bayes risk of δ with respect to the prior Π. So, $R(\beta, \delta) = \int \{\beta'\delta(y) - 1\}^2 f(y - \beta)\, dy$, and $B(\delta, \Pi) = \int R(\beta, \delta)\, \Pi(d\beta)$.

We now have the notation necessary for Lemma 1, which gives some expressions for the formal Bayes rules in this problem.

LEMMA 1. *Assume that $\delta_\Pi(y)$ is formal Bayes with respect to an invariant prior* Π (*not necessarily proper*). *Then*

A. $\delta_\Pi(y) = (E_{\Pi|y}\beta\beta')^{-1}E_{\Pi|y}\beta$
B. $\delta_\Pi(y) = (F' + 2\|y\|^2F'')^{-1}F'y$
C. $\delta_\Pi(y) = \{\int_0^\infty \gamma \sinh \gamma\|y\|\, d\tilde{\Pi}(\gamma)\}\{\int_0^\infty \gamma^2 \cosh \gamma\|y\|\, d\tilde{\Pi}(\gamma)\}^{-1}\dfrac{y}{\|y\|}$

PROOF. Minimizing the posterior expected loss by differentiating with respect to δ and setting the result equal to 0, we obtain $E_{\Pi|y}\beta\beta'\delta - E_{\Pi|y}\beta = 0$. As we shall see, for invariant priors the posterior covariance matrix is positive definite and hence A follows. We note that the posterior distribution on the parameter space is

$$\Pi\,|\,y = \left\{\int f(y-\beta)\Pi(d\beta)\right\}^{-1} f(y-\beta)\Pi(d\beta). \tag{2.3}$$

Let ∇ be the gradient operator ($\nabla = (\partial/\partial y_1, \cdots, \partial/\partial y_n)'$). From (2.2) and (2.3), we obtain

$$E_{\Pi|y}\beta = (1/F)\nabla F = (2F'/F)y, \tag{2.4}$$

and

$$E_{\Pi|y}\beta\beta' = (1/F)\nabla(\nabla F)' = (1/F)\nabla(2F'y') = (2F'/F)I + (4F''/F)yy'.$$

The inverse of this matrix is

$$(E_{\Pi|y}\beta\beta')^{-1} = \frac{F}{2F'}\left(I - \frac{2F''}{F' + 2F''y'y}yy'\right). \tag{2.5}$$

Substituting (2.4) and (2.5) into result A gives result B. To obtain result C, note that

$$F'(y'y) = (2\|y\|)^{-1}\int \gamma \exp(\gamma\|y\|)\, d\tilde{\Pi}(\gamma)$$

and

$$F''(y'y) = (4\|y\|^2)^{-1}\int \gamma^2\exp(\gamma\|y\|)\, d\tilde{\Pi}(\gamma) - (4\|y\|^3)^{-1}\int \gamma \exp(\gamma\|y\|)\, d\tilde{\Pi}(\gamma).$$

Substituting into result B gives

$$\delta(y) = \left\{\int \gamma \exp(\gamma\|y\|)\, d\tilde{\Pi}(\gamma)\right\}^{-1}\left\{\int \gamma^2 \exp(\gamma\|y\|)\, d\tilde{\Pi}(\gamma)\right\}\frac{y}{\|y\|}. \tag{2.6}$$

Since $\tilde{\Pi}$ is invariant under the change of variables $\gamma \to -\gamma$, (2.6) is easily shown to reduce to result C. □

The next lemma gives a useful expression for the difference in Bayes risk of two rules, δ and δ_Π. The proof, which follows easily from Lemma 1, is omitted.

LEMMA 2.

$$B(\delta, \Pi) - B(\delta_\Pi, \Pi) = \int (\delta - \delta_\Pi)'\left[\int \beta\beta' f(\beta - y)\Pi(d\beta)\right](\delta - \delta_\Pi)\, dy.$$

In particular, if δ and Π are orthogonally invariant, so that $\delta(y) = \psi(y'y)y$ and $\delta_\Pi(y) = \psi_\Pi(y'y)y$, we have:

$$B(\delta, \Pi) - B(\delta_\Pi, \Pi) = \int (\psi - \psi_\Pi)^2\left\{\int (\beta'y)^2 f(\beta - y)\Pi(d\beta)\right\} dy.$$

LEMMA 3. *If $\delta(y)$ is invariant and admissible then there exists a sequence of invariant proper priors $\Pi_k(y)$ and associated Bayes rules $\delta_k(y)$ converging to $\delta(y)$ pointwise almost everywhere.*

PROOF. The Stein-Le Cam theorem, as given by Farrell (1968), implies that if $\delta(y) = \psi(y'y)y$ is admissible and invariant then for any bounded continuous function α from $\mathbb{R}^p$ to $\mathbb{R}$, there exists a sequence of finite invariant measures Π_k and associated Bayes rules $d_k(y) = \psi_k(y'y)y$, such that (i) $\int \alpha(\beta)\Pi_k(d\beta) = 1$, and (ii) $\lim_{k\to\infty} B(\delta, \Pi_k) - B(d_k, \Pi_k) = 0$. For a proof that this theorem is applicable, see Berger and Zaman [1980]. Take $\alpha(\beta) = \beta_1 f(\beta)$. Then (i) and (ii) together with Lemma 2 imply that:

$$\begin{aligned} 0 &= \lim_{k\to\infty} \int (\psi - \psi_k)^2 \int (\beta' y)^2 f(\beta - y)\Pi_k(d\beta)\, dy \\ &\geq \lim_{k\to\infty} \int (\psi - \psi_k)^2 \int \beta_1^2 \|y\|^2 \exp\left(-\frac{1}{2}\|y\|^2\right) f(\beta)\Pi_k(d\beta)\, dy \\ &= \lim_{k\to\infty} \int (\psi - \psi_k)^2 \|y\|^2 \exp\left(-\frac{1}{2}\|y\|^2\right) dy \geq 0. \end{aligned}$$

The inequality in the second line comes from making an orthogonal transformation using the invariance of the Π_k, and the fact that $\|\beta - y\|^2 \leq \|\beta\|^2 + \|y\|^2$. This demonstrates that ψ_k converges to ψ in L_2 and hence a subsequence of the ψ_k converges pointwise almost everywhere to ψ. □

LEMMA 4. *If there exists an $M \geq 0$ such that $\delta(y) = 0$ for $\|y\| > M$ then δ is inadmissible.*

PROOF. Define $\tilde{\delta}(y)$ to be $\delta(y)$ if $|y_1| \leq M$ and $(1/y_1, 0, 0, \cdots, 0)$ if $|y_1| > M$. Then we have, for $\beta_1 \neq 0$,

$$\begin{aligned} R(\beta, \delta) - R(\beta, \tilde{\delta}) &= \int_{|y_1|>M} \left\{1 - \left(\frac{\beta_1}{y_1} - 1\right)^2\right\} f(\beta_1 - y_1)\, dy_1 \\ &= \int_{|z+\beta_1|>M} \left\{1 - \left(\frac{z}{z + \beta_1}\right)^2\right\} f(z)\, dz > 0. \end{aligned}$$

Since for $\beta_1 = 0$, the rules have identical risk, $\tilde{\delta}$ is uniformly superior to δ. □

Let $\delta(y)$ be invariant and admissible. Lemmas 1 and 3 imply that there exists a sequence of invariant priors Π_k such that

$$\lim_{k\to\infty} \left\{\int_0^\infty \gamma \sinh \gamma\|y\|\, d\tilde{\Pi}_k(\gamma)\right\}\left\{\int_0^\infty \gamma^2 \cosh \gamma\|y\|\, d\tilde{\Pi}_k(\gamma)\right\}^{-1} \|y\|^{-1}y = \delta(y) \text{ a.e.} \tag{2.7}$$

Since $\delta(y)$ is invariant, it can be written as

$$\delta(y) = \phi(\|y\|)\, \|y\|^{-1}y. \tag{2.8}$$

For $k = 1, 2, \cdots$ define the probability measures $d\mu_k = (\int \gamma^2\, d\tilde{\Pi}_k)^{-1}\gamma^2 d\tilde{\Pi}_k$, and the associated functions

$$\phi_k(\|y\|) = \left\{\int_0^\infty \gamma^{-1} \sinh \gamma\|y\|\, d\mu_k(\gamma)\right\}\left\{\int_0^\infty \cosh \gamma\|y\|\, d\mu_k(\gamma)\right\}^{-1}.$$

Then, (2.7) and (2.8) imply that ϕ_k converge pointwise almost everywhere to ϕ.

We will now prove

LEMMA 5. *$\int_0^\infty \cosh \gamma \|y\| \, d\mu_k(\gamma)$ is bounded in k for any fixed value of $\|y\|$.*

PROOF. Assume, toward contradiction, that there exists a y, say $y = y_0$ such that

$$\limsup_{k\to\infty} \int_0^\infty \cosh \gamma \|y_0\| \, d\mu_k(\gamma) = \infty.$$

Let $M = \|y_0\|$. Then the above must hold for all y such that $\|y\| \geq M$. Since $(1/x) \sinh x$ is monotonic increasing for x positive, we have, for any $L > 0$,

$$\phi_k(\|y\|) < \frac{\frac{1}{L}\sinh L\|y\| + \frac{1}{L}\int_L^\infty \sinh \gamma \|y\| \, d\mu_k(\gamma)}{\int_0^\infty \cosh \gamma \|y\| \, d\mu_k(\gamma)}.$$

The first term in the numerator stays fixed while the denominator is unbounded in k. Hence

$$\limsup_{k\to\infty} \phi_k(\|y\|) \leq \limsup_{k\to\infty} \frac{1}{L}\left\{\int_L^\infty \sinh\gamma\|y\| \, d\mu_k(\gamma)\right\}\left\{\int_L^\infty \cosh\gamma\|y\| \, d\mu_k(\gamma)\right\}^{-1} \leq \frac{1}{L}.$$

The second inequality follows since $\sinh x \leq \cosh x$. By taking L large, we see that the above implies that $\lim_{k\to\infty} \phi_k(\|y\|) = 0$ for $\|y\| \geq M$. Then, by Lemma 4, $\delta(y)$ is inadmissible. This contradiction completes the proof. □

Lemma 5 implies that μ_k is a tight family. Hence there exists a probability measure μ such that a subsequence μ_{k_i} converges to μ weakly. (See, for example, Theorem 4.4.3 of Chung (1972).) Lemma 5 also implies that $\lim_{i\to\infty} \int_0^\infty \gamma^{-1} \sinh \gamma\|y\| \, d\mu_{k_i} = \int_0^\infty \gamma^{-1} \sinh \gamma\|y\| \, d\mu$ and $\lim_{i\to\infty} \int_0^\infty \cosh \gamma\|y\| \, d\mu_{k_i} = \int_0^\infty \cosh \gamma\|y\| \, d\mu$. This completes the proof of Theorem 1.

It can easily be checked that the class of estimates given by Theorem 1 is strictly larger than the class of orthogonally invariant formal Bayes estimates. For any invariant measure Π, we can write it as in (2.1) by setting

$$\mu(A) = \left\{\int \gamma^2 \tilde{\Pi}(d\gamma)\right\}^{-1} \int_A \gamma^2 \tilde{\Pi}(d\gamma). \tag{2.9}$$

However, if μ has too much mass near 0, we cannot invert (2.9) to obtain a Π such that the rule in (2.1) is formal Bayes for Π. Since Theorem 1 does not give a minimal complete class, the possibility is left open that the formal Bayes rules may also form a complete class. We will show in the following section that this is not the case by showing that the rule obtained from setting μ to be point mass at 0 is admissible.

3. Admissibility of certain procedures in low dimensions. Zellner (1971) proposed the use of the rule generated by the diffuse prior (i.e., Lebesgue measure). It is well known that the analogous rule for the estimation of the multinormal mean is inadmissible in three or more dimensions. Here we shall prove that Zellner's rule is admissible in three dimensions or less. Kei Takeuchi has shown elsewhere that the diffuse prior rule is inadmissible in dimensions of six or more. Stein and Zaman (1979) have recently shown that it remains admissible in four, and is inadmissible in five dimensions.

Let τ_σ denote the multivariate normal prior distribution with mean 0 and covariance $\sigma^2 I$, and let δ_σ be the associated Bayes rule.

LEMMA 6.

$$\delta_\sigma(y) = \left(1 + \frac{\sigma^2}{\sigma^2 + 1} y'y\right)^{-1} y.$$

Also Zellner's rule is given by $\delta_z(y) = \delta_\infty(y) = (1 + y'y)^{-1}y$.

PROOF. Use of Lemma 1B leads easily to this result. □

THEOREM 2. *Zellner's rule is admissible in dimension* 1, 2 *and* 3.

PROOF. We shall consider only the case of dimension 3. It is trivial to adapt the proof to dimensions 1 and 2. Suppose δ_z is inadmissible in 3 dimensions and let $\tilde{\delta}$ be a uniform improvement. Since the risk functions of decision rules are continuous for this problem (this follows easily from the continuity of the Laplace transform), there exists an $\epsilon > 0$ and a point β_0 in the parameter space such that $\|\beta - \beta_0\| < \epsilon$ implies $R(\beta, \delta_z) > R(\beta, \tilde{\delta}) + \epsilon$. Now, note the following

$$\begin{aligned} 0 &\le \sigma^3\{B(\delta, \tau_\sigma) - B(\delta_\sigma, \tau_\sigma)\} \\ &= \sigma^3\{B(\tilde{\delta}, \tau_\sigma) - B(\delta_z, \tau_\sigma)\} + \sigma^3\{B(\delta_z, \tau_\sigma) - B(\delta_\sigma, \tau_\sigma)\}. \\ &\le -\epsilon \int_{\{\beta:\|\beta-\beta_0\|<\epsilon\}} f\left(\frac{\beta}{\sigma}\right) d\beta + \sigma^3\{B(\delta_z, \tau_\sigma) - B(\delta_\sigma, \tau_\sigma)\}. \end{aligned}$$

The first term converges to a strictly negative quantity as σ^2 approaches infinity. To complete the proof, we will show that the second term converges to zero and hence obtain a contradiction. Since this computation is the core of the proof, we present it in the form of a lemma.

LEMMA 7.

$$\lim_{\sigma\to\infty} \sigma^3\{B(\delta_z, \tau_\sigma) - B(\delta_\sigma, \tau_\sigma)\} = 0.$$

PROOF. For notational convenience, define

$$D = B(\delta_z, \tau_\sigma) - B(\delta_\sigma, \tau_\sigma), \tag{3.1}$$

and

$$\lambda = (\sigma^2 + 1)^{-1}\sigma^2, \phi_\sigma(\|y\|) = (1 + \lambda\|y\|^2)^{-1}\|y\|,$$
$$\phi_z(\|y\|) = (1 + \|y\|^2)^{-1}\|y\|.$$

The symbols K, K', K'' will represent numerical constants in the following argument.

Using Lemma 2, and the inequality $(\beta' y)^2 \le \|\beta\|^2\|y\|^2$, we can write (3.1) as

$$\begin{aligned} D &\le K\sigma^{-3} \int (\phi_z - \phi_\sigma)^2 \left\{\int \|\beta\|^2 f(\beta - y) f\left(\frac{\beta}{\sigma}\right) d\beta\right\} dy \\ &= K' \lambda^{3/2}\sigma^{-3} \int_{\mathbb{R}^3} \left\{\frac{(1-\lambda)\|y\|^3}{(1 + \|y\|^2)(1 + \lambda\|y\|^2)}\right\}^2 (\lambda^2\|y\|^2 + 3\lambda) f\left(\frac{\sqrt{\lambda}y}{\sigma}\right) dy. \end{aligned}$$

The second line is obtained by integrating out β, and writing $\phi_z - \phi_\sigma$ explicitly. Define $X = \lambda\|y\|^2/\sigma^2$. Regarding X as a χ_3^2 random variable, we can write the above as

$$\begin{aligned} D &\le K''\lambda^{3/2} E\left\{\frac{(1-\lambda)^2(\sigma^8X^4 + 3\sigma^6X^3)}{(1 + \lambda^{-1}\sigma^2X)^2(1 + \sigma^2X)^2}\right\} \\ &= K''\lambda^{3/2}(1-\lambda)^2 E\left\{\frac{1 + 3(\sigma^2X)^{-1}}{(1 + (\sigma^2X)^{-1}(\lambda^{-1} + (\sigma^2X)^{-1})}\right\}. \end{aligned}$$

Obviously the last expectation approaches 1 as σ^2 approaches infinity. Thus $0 \le D \le 2K''(1-\lambda)^2$ and hence $0 \le \lim_{\sigma\to\infty} \sigma^3 D \le \lim_{\sigma\to\infty} 2K''\sigma^3/\{(1 + \sigma^2)^2\} = 0$. □

Now we give an example of an admissible estimator which is not formal Bayes. This will demonstrate that the class of formal Bayes rules is not complete in this problem.

THEOREM 3. *In one dimension, $\delta(y) = y$ is an admissible procedure which is not formal Bayes.*

PROOF. Assume δ is formal Bayes versus a measure Π. Define $g(y) = \int \beta \exp(\beta y - \frac{1}{2}\beta^2)\, d\Pi(\beta)$. By Lemma 2A, we have $\delta(y) = (g'(y))^{-1} g(y) = y$. Solving the differential equation for g, we obtain that $g(y) = Cy$ for some constant C. Thus $g'(y) = C = \int \beta^2 \exp(\beta y - \frac{1}{2}\beta^2)\, d\Pi(\beta)$. This can only hold if Π put all mass at 0, a point excluded from the parameter space. Thus $\delta(y)$ is not formal Bayes.

We will show that no decision rule can improve on the risk function of $\delta(y)$ near 0, and if it has equal risk near 0, then it must equal $\delta(y)$. First we compute $R(\beta, \delta) = E(\beta y - 1)^2 = \beta^4 - \beta^2 + 1$.

Suppose $\tilde{\delta}$ is a uniform improvement on δ. We can pick $\tilde{\delta}$ to be spherically symmetric. Hence $\tilde{\delta}(y) = \psi(y^2)y$. We now expand the risk of $\tilde{\delta}$ near $\beta = 0$.

$$\frac{\partial}{\partial\beta} R(\beta, \tilde{\delta}) = E\{(y - \beta)(\beta\tilde{\delta} - 1)^2 + 2\tilde{\delta}(\beta\tilde{\delta} - 1)\}.$$

Since $\tilde{\delta}$ is an odd function, the above is 0 when evaluated at $\beta = 0$. Next we have

$$\frac{\partial^2}{\partial\beta^2} R(\beta, \tilde{\delta})\,|_{\beta=0} = E(2\tilde{\delta}^2 - 4y\tilde{\delta}) = 2E\{y^2\psi(\psi - 2)\}.$$

Since $(\partial^2/\partial\beta^2)R(\beta, \delta) = -2$, the above must be less than or equal to -2, since $\tilde{\delta}$ is better than δ. However, the reverse inequality must also hold: $2E\{y^2\psi(\psi - 2)\} = 2E[y^2\{(\psi - 1)^2 - 1\}] \geq -2Ey^2 = -2$. This implies that ψ must be identically 1 and hence $\tilde{\delta}(y) = y$ a.e. □

We note that this provides a fairly simple as well as natural example of an estimation problem with convex loss where the class of formal Bayes estimates is not complete. Note also that $\delta(y) = y$ results from setting $\mu\{0\} = 1$ in Theorem 1.

4. A truncation theorem. The object of this section is to prove that all admissible invariant rules $\delta(y) = \psi(y'y)y$, satisfy $0 \leq \psi \leq 1$. This is done by showing that if ψ does not lie in the interval [0, 1], a truncation which does leads to a uniformly superior rule. This theorem makes it easy to show that the space of decision rules is weakly subcompact in the sense of Farrell (1968), which is needed to show that Stein's necessary conditions for admissibility hold in this problem (see Berger and Zaman (1980)).

THEOREM 4. (The truncation theorem). *Suppose $\delta(y) = \psi(y'y)y$, and define $A = \{\lambda \mid \psi(\lambda) > 1\}$ and $B = \{\lambda \mid \psi(\lambda) < 0\}$. If A or B have Lebesgue measure greater than 0 then δ is inadmissible, and a uniform improvement is given by*

$$\tilde{\delta}(y) = \begin{cases} \delta(y) & (y'y) \notin A \cup B \\ y & (y'y) \in A \\ 0 & (y'y) \in B \end{cases}$$

PROOF. Define

$$\Delta(\beta) = E[\{\psi(y'y)\beta'y - 1\}^2 - (\beta'y - 1)^2 \mid y'y \in A] \tag{4.1}$$

$$\Gamma(\beta) = E[\{\psi(y'y)\beta'y - 1\}^2 - 1 \mid y'y \in B]. \tag{4.2}$$

Clearly, it suffices to show that Δ and Γ are positive functions of β. In (4.2), make the change of variable $v = Py$ where the first row of P is proportional to β and P is orthogonal. Then $v_1 \sim N(\|\beta\|, 1)$ and $v_2, \cdots, v_n$ are all $N(0, 1)$, with all the v_i independent. After change of variables, (4.2) becomes

$$\Delta(\beta) = E\{\psi^2(v'v)\|\beta\|^2 v_1^2 - 2\psi(v'v)\|\beta\| v_1 - \|\beta\|^2 v_1^2 + 2\|\beta\| v_1 \mid v'v \in A\}.$$

We are now going to integrate out $v_2, \cdots, v_n$. For this purpose, define $M = \sup_x\{x \in A\}$. (M may be $+\infty$). Define, for $v_1^2 \in (0, M)$,

$$h(v_1^2) = E\{\psi(v_1^2 + v_2^2 + \cdots + v_n^2) \mid v_2^2 + \cdots + v_n^2 \in A - v_1^2\}. \tag{4.3}$$

By Jensen's Inequality, we have $h^2(v_1^2) \le E\{\psi^2(v'v) \mid v_2^2 + \cdots + v_n^2 \in A - v_1^2\}$. Substituting in (4.3), we get, after integrating out $v_2, \cdots, v_n$.

$$\Delta(\beta) \ge E[\{h(v_1^2) - 1\}\|\beta\| v_1\{(h(v_1^2) + 1)\|\beta\| v_1 - 2\} \mid v_1 \in (-\sqrt{M}, \sqrt{M})]. \tag{4.4}$$

Defining $g(\beta) = f(\beta)\{\int_{\sqrt{M}}^{\sqrt{M}} f(v_1 - \|\beta\|)\, dv_1\}^{-1}$, we can write the conditional expectation as

$$\Delta(\beta) \ge \int_{\sqrt{M}}^{\sqrt{M}} \{h(v_1^2) - 1\}\|\beta\| v_1[\{h(v_1^2) + 1\}\|\beta\| v_1 - 2]\exp(\|\beta\| v_1) f(v_1) g(\beta)\, dv_1$$

$$= \int_{\sqrt{M}}^{0} \cdots dv_1 + \int_0^{\sqrt{M}} \cdots dv_1.$$

Make the change of variables $v_1 \to -v_1$ in the first integral and recombine to get

$$\frac{\Delta(\beta)}{g(\beta)} \ge \int^{\sqrt{M}} \{h(v_1^2) - 1\}\|\beta\| v_1[2\{h(v_1^2) + 1\}\|\beta\| v_1 \cosh \|\beta\| v_1 - 4 \sinh \|\beta\| v_1] f(v_1)\, dv_1.$$

Since $\psi \ge 1$ on A, we must have $h \ge 1$. We also have the inequality $x \cosh x > \sinh x$, for $x > 0$. These imply, since $g(\beta) > 0$ for all β, $\Delta(\beta) > 0$, as desired.

It remains to show that $\Gamma(\beta)$ is also positive. Using the same change of variables as before, we get from (4.2)

$$\Gamma(\beta) = K \int_{v'v \in B} \{\psi^2(v'v)\|\beta\| v_1^2 - 2\psi(v'v)\|\beta\| v_1\} \exp\{-\tfrac{1}{2}((\|\beta\| - v_1)^2 + \cdots + v_n^2)\} dv.$$

The first term is always positive; for the second term we have, $E_{v_1}\{-2\psi(v'v)\|\beta\| v_1\} > 0$, since this expectand is an odd function of v_1 which is positive for $v_1 > 0$, and $v_1 \sim N(\|\beta\|, 1)$. Therefore Γ must be positive, completing the proof of the theorem. □

5. Estimation of the reciprocal of a normal mean. In this section, we discuss the application of our results to the problem of estimation of a reciprocal of a normal mean with quadratic loss. Formally, we observe $y \sim N(\beta, 1)$, and wish to estimate $1/\beta$. (We know $\beta \ne 0$ a priori). The loss function is $L(\beta, \delta) = (\delta - 1/\beta)^2 = \beta^{-2}(\beta\delta - 1)^2$. Because of this relationship, the problem can be regarded as essentially a one-dimensional special case of the control problem. Since it is easily understood, we shall discuss it to give a feeling for the nature of the results obtained earlier.

The maximum likelihood estimate of β is $\delta_c(y) = 1/y$. This is graphed in Figure 1. Many authors (see for example, Anderson and Taylor (1976) and Basu (1974)) have noted that this has infinite risk with respect to quadratic loss, and have proposed the truncation $\delta_M(y) = 1/y$ or $\mathrm{sgn}(y)M$ according as $|y| > 1/M$ or $|y| < 1/M$. This is graphed in Figure 2. The Complete Class Theorem can be used to prove that the graphs of all symmetric admissible rules must lie in the shaded region in Figure 3. It is clear that $\delta_M(y)$ does not satisfy this condition for any M. The Truncation Theorem provides a uniformly better truncation, $\tilde{\delta}_M$, which is given in Figure 4. Even though this truncation dominates the usual truncation $\tilde{\delta}_M$, it cannot be admissible, because it has a sharp corner. All rules given by the Complete Class Theorem are $C^{(\infty)}$ and hence $\tilde{\delta}_M$ cannot be admissible. However, we have not been able to come up with a uniform improvement on $\tilde{\delta}_M$, and we suspect it may not be possible to improve on its risk substantially.

FIG. 1

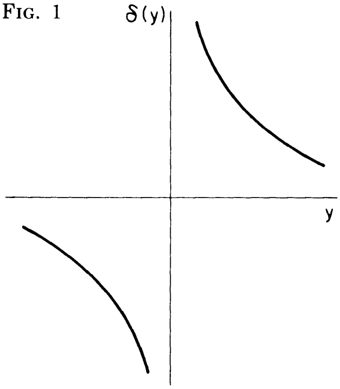

FIG. 2

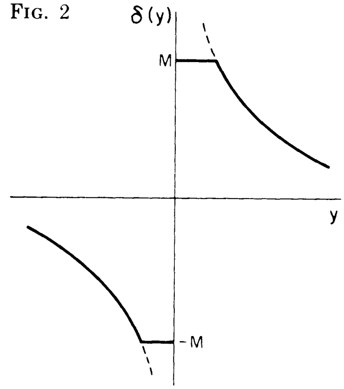

FIG. 3

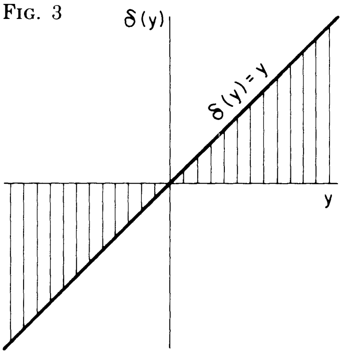

FIG. 4

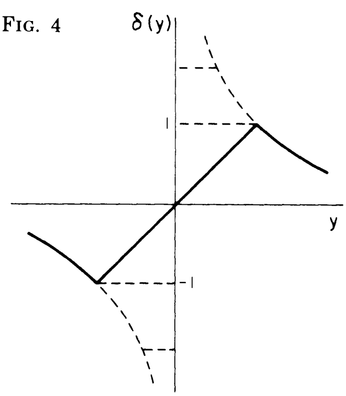

FIG. 1. *Maximum likelihood or naive certainty equivalent* $\delta_c(y)$. FIG. 2. *Usual truncation* $\delta_M(y)$. FIG. 3. *All symmetric admissible rules must lie in shaded region.* FIG. 4. *Uniform improvement on usual truncation* $\tilde{\delta}_M(y)$.

The fact that all admissible procedures go through the origin is surprising at first glance. A small value of y may well suggest that β is small and hence that $1/\beta$ is large. This conclusion would be premature for β may well be negative. An aggressively large estimate of $1/\beta$ on the order of $1/y$ will then yield a large loss. Large losses accrue as well if β happens to be too large (on the order of $2y$ or more). This latter event is almost as well-supported by small y's as is $1/\beta \sim 1/y$. The difficulty here is that if y is small there is not much information about $1/\beta$ and losses on the order of $1/\beta^2$ are inevitable. Another noteworthy fact is that not only do all admissible procedures go through the origin, but they go through tangent to the line $\delta(y) = y$. Thus for small y, all admissible estimates of $1/\beta$ satisfy $\delta(y) \approx y$. We have not been able to obtain a satisfactory intuitive explanation of this result.

For a more detailed study of this problem, see Zaman (1981).

6. Concluding remarks. We have developed a number of results about the control problem, the most fundamental being a complete class for symmetric estimators in the problem.

In Section 3 we proved that the use of Lebesgue measure as a prior leads to admissible estimators at least up to dimension three. We also proved the inadmissibility of certain commonly used procedures and provided uniform improvements.

These results are interesting not only for their own sake, but also for the contrast they provide to the "direct" problem. (We shall refer to the control problem as the "inverse" problem, and to the estimation of the multinormal mean as the "direct" problem.) In the direct problem, the formal Bayes rules form a complete class. In this problem they do not. A number of inadmissibility results, starting with Stein's indicate that the diffuse prior distribution becomes inadmissible in dimension three or more. In our problem, this does not occur until dimension five.

A very general result on inadmissibility in the control problem is given in Berger and Zaman (1979). Extensions of this framework to the case where σ^2 is unknown, and the case where the decision δ is constrained to lie in some affine subspace of $\mathbb{R}^p$ are studied in Zaman (1978).

Acknowledgments. I would like to thank Professor Charles Stein for a large number of discussions which improved my understanding of this and several other problems considerably. I am also indebted to Professor Srinivasa S. R. Varadhan for his help with the proof of the Complete Class Theorem.

I am deeply indebted to Professor T. W. Anderson for numerous comments and suggestions which have greatly improved the quality of this manuscript. The responsibility for any remaining errors is, however, entirely mine.

Thanks are also due to anonymous referees whose comments on an earlier version of this paper were very helpful.

REFERENCES

ANDERSON, T. W. and J. B. TAYLOR (1976). Some experimental results on the statistical properties of least squares estimated in control problems. *Econometrica* **44**.

BASU, A. (1974). Control level determination in regression models. Tech. Report No. 139, Economics Series, Inst. Math. Studies in Soc. Sci. Stanford Univ.

BERGER, J. and A. ZAMAN (1979). Inadmissibility in the control problem. Tech. Report. Statist. Dept. Stanford Univ.

BROWN, L. D. (1971). Admissible estimators, recurrent diffusions and insoluble boundary value problems. *Ann. Math. Statist.* **42** 855–903.

CHUNG, K. L. (1972). *A Course in Probability Theory*, Academic, New York.

FARRELL, R. H. (1968). Towards a theory of generalized Bayes tests. *Ann. Math. Statist.* **39** 1–22.

FERGUSON, T. S. (1967). *Mathematical Statistics, A Decision Theoretic Approach.* Academic, New York.

STEIN, C. (1964). Inadmissibility of the usual estimator for the variance of a normal distribution with unknown mean. *Ann. Inst. Statist. Math.* **16** 155–160.

STEIN, C. and A. ZAMAN (1980). On the admissibility of the diffuse prior for the control problem: The case of dimensions four and five. Tech. Report, Dept. Statist. Stanford Univ.

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

ZAMAN, A. (1981). Estimators without moments: The case of the reciprocal of a normal mean. *J. Econometrics* **15** 289–298.

ZAMAN, A. (1978). The single period control problem in econometrics. Unpublished Ph.D. Thesis, Stanford Univ.

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

ZIDEK, J. V. (1976). A necessary condition for the admissibility under convex loss of equivariant estimators. Tech. Report No. 113, Dept. Statist., Stanford Univ.

DIVISION OF HUMANITIES AND SOCIAL SCIENCE
BAXTER HALL 228-77
CALIFORNIA INSTITUTE OF TECHNOLOGY
PASADENA, CALIFORNIA 91125