A NECESSARY CONDITION FOR THE
ADMISSIBILITY UNDER CONVEX LOSS OF
EQUIVARIANT ESTIMATORS

BY

J. V. ZIDEK

TECHNICAL REPORT NO. 113
AUGUST 15, 1976

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT
MPS 7509450-000
Ingram Olkin, Project Director

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
A NECESSARY CONDITION FOR THE
ADMISSIBILITY UNDER CONVEX LOSS OF
EQUIVARIANT ESTIMATORS

BY

J. V. ZIDEK

TECHNICAL REPORT NO. 113
AUGUST 15, 1976

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT
MPS 7509450-000

Ingram Olkin, Project Director

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
Abstract

Whatever be the underlying distribution or (convex) loss function, an admissible equivariant estimator must be the limit in measure of some collection of Bayes estimators relative to the class containing all invariant estimators. This result is proved in the general setting provided by Le Cam's extension of Wald's decision problem. The direct argument is carried out for a compact action space and it is then shown how a noncompact action space may be suitably compactified. Using this result it is shown that the best fully equivariant estimator of the generalized variance of the multivariate normal law with unknown mean is inadmissible if the loss function is (neglecting regularity conditions) strictly convex.
1. **INTRODUCTION.** This paper presents a usable necessary condition for the admissibility of equivariant estimators when loss is strictly convex. The condition is not new and was first observed by Stein (1964). But there is no published proof of its necessity. A heuristic argument is given below in this introduction for the special case of quadratic loss in order to indicate why this condition is necessary. A rigorous version of this argument for a fairly general convex loss function appeared in the original version of this manuscript but after its distribution, L. D. Brown communicated to the author another proof which, in the same setting, leads to a somewhat more general result by a shorter route. His alternate proof will be developed here in Section 2. An application of the result appears in Section 3, where an important special case of the Shorrock-Zidek (1976) results are extended in another direction to the case of more general loss functions.

To describe the condition we adopt Lehman's language and notation (see Lehman (1959), pp. 213-218) and suppose given an invariant estimation problem where loss is quadratic. Let \( \omega(\theta) \) represent the quantity being estimated and suppose \( \hat{\omega}(X) \) is an invariant estimator with risk function denoted by \( r(\hat{\omega}, \theta) = c(\theta) E_\theta (\hat{\omega}(X) - \omega(\theta))^2 \). Since \( \hat{\omega} \) is invariant under \( G \), \( r(\hat{\omega}, \theta) \) really depends on \( \theta \) only through \( \nu(\theta) \), the \( G \)-maximal invariant in \( \theta \)-space. In many special cases, \( \hat{\omega} \) may be represented as \( \hat{\omega}(X) = G[X] \hat{\omega}^*(T(X)) \) where \( Z = T[X] \) is the \( G \)-maximal invariant in \( X \)-space and \( \hat{g} \) maps \( X \)-space onto a group which is homomorphic to \( G \) and such that if \( \hat{g} \) is the homomorphic image of \( g \), \( c(\theta)(\hat{g} \omega - \omega(\theta))^2 = c(\theta)(\omega - \omega(\theta))^2 \) for all \( g \in G \). Thus \( r(\hat{\omega}, \theta) = E_\theta c(\theta) \omega^*(Z) - \omega(Y_\theta))^2 \) where \( Y = (G[X])^{-1} \). If \( \hat{\omega} \) is admissible then according to the well-known
result of Stein (1955), it is necessary that for each fixed parameter value, \( \theta_0 \), and \( \epsilon > 0 \), there exist a prior distribution \( \xi \) and \( \delta > 0 \) such that \( \hat{\omega} \) is \( \delta \)-Bayes with respect to \( \pi = (1-\delta)\xi + \delta[\theta_0] \) where \( [\theta_0] \) denotes the probability distribution degenerate at \( \theta_0 \). From this it follows that

\[
E_{\theta_0} c^*(\theta_0, Z)(\hat{\omega}(Z) - \hat{\omega}_1(Z))^2 < \epsilon
\]

where \( \hat{\omega}(X) \hat{\omega}_1(T(X)) \) is the Bayes rule relative to the class of invariant rules with respect to the measure induced on the range of the \( \bar{G} \)-maximal invariant \( \nu \) by \( \pi \) and \( c^*(\theta_0, Z) = E_{\theta_0} [c(Y_{\theta_0})|Z] \).

This observation implies that \( \hat{\omega} \) must be the limit in \( P_{\theta_0} \)-measure (for each \( \theta_0 \)) of a sequence of rules each being Bayes relative to the class of invariant rules. This is the necessary condition for admissibility which is deduced in the more general setting of Section 2.

2. NECESSARY CONDITION FOR ADMISSIBILITY.

2.1 Basic Model. Let \( X \) represent an observable random variable whose range is a measurable space \( (\mathcal{X}, \mathcal{E}) \). We suppose \( X \) is distributed by a unique but unknown probability measure which is an element of \( \{P_{\theta}: \theta \in \Theta\} \). After \( X \) is observed an action is to be selected from the action space, \( \mathcal{A} \), which we assume, is a topological space and a subset, with convex interior, of \( [-\infty, \infty]^m \) for some \( m \). Then a loss, \( 0 \leq L(a, \theta) \leq \infty \) is incurred if \( a \in \mathcal{A} \) is selected when \( X \) is distributed by \( P_{\theta} \). Assume for each \( \theta \), that \( L(\cdot, \theta) \) is strictly convex on the interior of \( \mathcal{A} \), is lower semi-continuous and satisfies
\( L(a, \theta) \geq \alpha \|a\| + \beta \) for some functions \( \alpha \) and \( \beta \) of \( \theta \) satisfying 
\( \alpha > 0, \ -\infty < \beta < \infty \) whenever \( a \in (\mathbb{R}^d)^m \).

The action is selected by means of a decision rule \( \delta : \mathcal{B} \times \mathcal{X} \rightarrow [0,1] \). Here \( \mathcal{B} \) denotes the class of Borel subsets of \( \mathcal{A} \) and for each \( A \in \mathcal{A}, \delta(A, \cdot) \) is measurable while for each \( x \in \mathcal{X}, \delta(\cdot, x) \) is a probability measure. If \( X = x \) were observed, \( a \in A \) would be chosen at random according to the probability distribution \( \delta(\cdot, x) \). The risk function of \( \delta, r(\delta, \theta) \), is defined as the overall expected loss incurred using \( \delta \) if \( X \) is distributed by \( P_\theta \). The collection of all decision rules is denoted by \( \mathcal{D} \) while \( \mathcal{D}_N \) denotes the class of all nonrandomized rules, that is, measurable mappings of \( \mathcal{X} \) into \( \mathcal{A} \).

Assume \( \mu \) is \( \sigma \)-finite and that each \( P_\theta \) is absolutely continuous with respect to \( \mu \). Let \( L_1 = L_1(\mathcal{C}, \mu) \). Assume \( L_1 \) is separable (Professor Shorrock pointed out in private communication that this condition could be omitted at the expense of a lengthier argument).

Suppose that \( G \) is a topological transformation group acting on \( \mathcal{X} \) and that the problem remains invariant under \( G \). This assumptions entails measurable transformation groups (see Zidek (1969)) \( \mathcal{G} \) and \( \hat{G} \) homomorphic to \( G \) and acting on \( \Theta \) and \( A \), respectively. These groups are required to act in such a way that if \( \bar{g} \) and \( \hat{g} \) denote, respectively, the homorphic images of \( g \in G \) in \( \bar{G} \) and \( \hat{G} \), then \( P_{\bar{g} \Theta}(gA) = P_\Theta(A) \) and \( L(\hat{g}a, \hat{g} \Theta) = L(a, \Theta), \ g \in G, \ a \in A, \ \Theta \in \Theta, \ \text{and} \ A \in \mathcal{C} \). In addition we assume that \( \hat{G} \) is a subgroup of the full affine group.

Assume the group's action on \( \mathcal{A} \) is continuous. This insures that \( c \circ \hat{g} \) is continuous if \( c \) is.
A procedure \( \delta \in \mathcal{Y} \) is called invariant if \( \delta(gA, gx) = \delta(A, x) \) for all measurable \( A, x \in \mathcal{X} \) and \( g \in G \). A nonrandomized procedure \( \zeta: \mathcal{X} \Rightarrow \mathcal{A} \) is called invariant (or equivariant) if \( \hat{g} \zeta(x) = \zeta(gx) \). The class of all invariant (invariant, nonrandomized) procedures will be denoted by \( \mathcal{D}_I \) (\( \mathcal{D}_{IN} \)). Bayes procedures relative to \( \mathcal{D}_I \) will be called Bayes-invariant.

If \( \delta \in \mathcal{D}_I \), \( r(\delta, \theta) \) depends on \( \delta \) only through \( \theta^* \) defined as the label of the \( G \)-orbit of \( \theta \) in \( \theta \), that is, of \( \{ \tilde{g}\theta : g \in G \} \). Let \( \theta^* = \{ \theta^*: \theta \in \theta \} \) and, with an abuse of notation, \( r(\delta, \theta^*) = r(\delta, \theta) \) when \( \delta \in \mathcal{D}_I \).

Our assumptions imply (see Hodges and Lehman (1950)) that \( \mathcal{D}_N \) is a complete class relative to \( \mathcal{D} \). Moreover, because \( \hat{G} \) is a subset of the full affine group, \( \mathcal{D}_{IN} \) is a complete class relative to \( \mathcal{D}_I \). The proof of this assertion requires, for each \( \delta \in \mathcal{D}_I \), the construction of a superior procedure (as in Hodges and Lehman (1950)) which is not only nonrandomized but invariant as well, and this requires the special form for \( \hat{G} \). It follows that every admissible procedure in \( \mathcal{D}_I \) is in \( \mathcal{D}_{IN} \) and that every Bayes-invariant procedure may be chosen from \( \mathcal{D}_{IN} \).

2.2 Compact Action Space. Assume \( \mathcal{A} \) is a compact metric space. Let \( \mathcal{C}(\mathcal{A}) \) denote the class of all continuous functions on \( \mathcal{A} \). Under the topology of uniform convergence, \( \mathcal{C}(\mathcal{A}) \) is a separable metric space. Let \( \mathcal{F} \) denote the space of all continuous real bilinear forms on \( L^1 \times \mathcal{C}(\mathcal{A}) \), and \( \| \cdot \| \) the norm on \( \mathcal{F} \). Consider the subset, \( \mathcal{F}(\mathcal{D}) \), of \( \mathcal{F} \) consisting of all bilinear forms \((\cdot, \cdot)\) for which \((f, g) \geq 0 \ldots \)
when \( f \in L_1, g \in C(\mathcal{A}) \) and \( f \geq 0, g \geq 0 \), and \((f,1) = \int f(x) \mu(dx), f \in L_1 \). In the topology of weak convergence in \( \mathcal{F} \), the unit ball of \( \mathcal{F}, \{h: h \in \mathcal{F}, ||h|| \leq 1\} \), is sequentially compact. In the same topology, \( \mathcal{F}(\mathcal{D}) \) is a convex, closed, and hence sequentially compact subset of the unit ball [See Farrell (1967)].

Two decision rules \( \delta \) and \( \delta' \) are called equivalent if they determine the same element of \( \mathcal{F}(\mathcal{D}) \), that is if

\[
\int \int g(y)f(x) \delta(dy, x) \mu(dx) = \int \int g(y)f(x) \delta'(dy, x) \mu(dx), f \in L_1,
\]

\( g \in C(\mathcal{A}) \). Then \( \mathcal{F}(\mathcal{D}) \) is in 1:1 correspondence with the quotient space, under this equivalence, of the space of all decision rules.

Denote this quotient space by \( \mathcal{D} \). In the Le Cam (1955) extension of the Wald theory, \( \mathcal{D} \) is equipped with the topology it inherits from \( \mathcal{F}(\mathcal{D}) \) when \( \mathcal{F} \) is equipped with the topology of weak convergence.

**Lemma 2.1.** Let \( \{\delta_\alpha: \alpha \in \Delta\} \), \( \delta_\alpha \in \mathcal{D}_N \) be a convergent sequence in \( \mathcal{D} \) with \( \lim \delta_\alpha = \delta \in \mathcal{D}_N \) and \( \delta_\alpha((d_\alpha(x)), x) = \delta((d(x)), x) = 1 \). Then there exists a subsequence \( \{\delta_{\alpha'}: \alpha' \in \Delta'\}, \Delta' \subset \Delta \) such that \( d_{\alpha} \Rightarrow d \) in \( \mu \)-measure.

**Proof.** Denote the metric on \( \mathcal{A} \) by \( \rho \). Let \( S \) and \( 2S \) denote respectively, \( \{y: \rho(a,y) < \varepsilon\} \) and \( \{y: \rho(a,y) < 2\varepsilon\} \), where \( a \) is any fixed element of \( \mathcal{A} \). Then if we let \( T = d^{-1}(S) \subset \mathcal{K} \),

\[
\lim \inf \mu[T \cap d_{\alpha}^{-1}(2S)] = \mu(T).
\]

(2.2.1)

For suppose not. Choose \( c \in C(\mathcal{A}) \) so that \( c \) is 1 on \( S \), 0 on...
the complement of $2S$, and $0 \leq c \leq 1$. Then if $f$ denotes the indicator function of $T$,

$$\liminf \int c[d_\alpha(x)]f(x)\mu(dx)$$

$$< \mu(T) = \int c[d(x)]f(x)\mu(dx)$$

which is a contradiction. So equation (2.1) holds for every open ball, $S$. The conclusion of the lemma is a consequence of this observation and the compactness of $A$. ||

Now in preparation for the next lemma we add the further assumption that $G$ is equipped with a $\sigma$-finite measure $\nu$ for which

$$\nu(B) = 0 \implies \nu(Bg) = 0$$

for all $g \in G$. The proof of the lemma is straightforward. It requires, in particular, an extension to the present context, of Theorem 4 of Lehman (1959), p. 225. This extension asserts that any almost invariant procedure is invariant and its proof uses the metrizability and separability of $A$.

**LEMMA 2.2.** $\mathcal{U}_I$ is closed and hence compact in $\mathcal{D}$. ||

For $\xi^* = (\xi_1^*, \ldots, \xi_m^*) \in \mathcal{E}_m^*$, define $\mathcal{R}_m[\xi^*] = \mathcal{R}_m$ by

$$\mathcal{R}_m = \{ z \in [0,\infty)^m : \xi^* \in \mathcal{D}_I, \exists z_i \geq r(\xi_i^*) \text{ for all } i \}. $$

Since $L(\cdot, \xi)$ is lower semicontinuous it follows that if a net $\{ \xi_\alpha : \alpha \in \mathcal{D} \}$ converges to a procedure $\xi$ then
\[ \lim_{\delta} \inf r(\delta, \theta) \geq r(\delta, \theta) \]

(see, for example, Farrell (1967), p. 109). Thus Lemma 2.2 implies that \( \mathcal{R}_m \) is closed. It is convex because \( \mathcal{D}_1 \) is. As a consequence of the foregoing we have

**Lemma 2.3.** For each probability measure, \( \xi \), with finite support in \( \Theta \), there exists a Bayes-invariant procedure \( \delta_\xi \in \mathcal{D}_{\text{IN}} \).

**Lemma 2.4.** If \( \mathbf{r} = (r_1, \ldots, r_m) \in \mathcal{R}_m \left[ (\theta_1^*, \ldots, \theta_m^*) \right] \), there exists a prior probability measure \( \xi \) on \( \{\theta_1^*, \ldots, \theta_m^*\} \) and associated Bayes-invariant procedure \( \delta \) such that \( r_i \geq r(\delta, \theta_i^*) \) for all \( i \).

**Proof.** Let \( \mathbf{r}'(0) = (r_1(0), \ldots, r_m(0)) \) denote that point of \( \mathcal{R}_m \) which is on the line joining \( \mathcal{Q} \in \mathcal{R}_m \) and \( \mathbf{r} \) and is the closest point in \( \mathcal{R}_m \) to \( \mathcal{Q} \). This point exists since \( \mathcal{R}_m \) is closed. And because \( \mathbf{r}'(0) \) is not an interior point of the convex set \( \mathcal{R}_m \), there is a supporting hyperplane \( \mathbf{r}'(0) \). Represent it by \( \{ \mathbf{x} \in \mathcal{R}_m : \sum \xi_i (x_i - r_i(0)) = 0 \} \). Then \( \sum \xi_i (x_i - r_i(0)) \geq 0 \), \( \mathbf{x} \in \mathcal{R}_m \). Because \( \mathcal{R}_m \) contains \( \{ \mathbf{x} \in \mathcal{R}_m : x_i \geq r_i \} \), \( \xi_i \geq 0 \). Without loss of generality take \( \sum \xi_i = 1 \). Since \( \mathbf{r}'(0) \in \mathcal{R}_m \), there exists \( \delta \in \mathcal{D}_1 \) and hence, by the completeness of \( \mathcal{D}_{\text{IN}} \), \( \delta \in \mathcal{D}_{\text{IN}} \) for which \( r(\delta, \theta_i^*) \leq r_i(0) \) all \( i \), so \( \sum \xi_i r(\delta, \theta_i^*) = \sum \xi_i r_i(0) = \inf \sum \xi_i x_i \). Thus \( \delta \) is a Bayes-invariant rule with respect to \( \xi \) while \( r(\delta, \theta_i^*) \leq r_i(0) \leq r \) as required.

**Theorem 2.1.** If \( \delta \in \mathcal{D}_1 \) is admissible, has a finite risk function, and \( \delta((d(x)), x) = 1 \), then there exists a sequence of Bayes-invariant estimators, \( \{\delta_\alpha : \alpha \in \mathcal{A}\} \) with \( \delta_\alpha \in \mathcal{D}_{\text{IN}} \) and say, \( \delta_\alpha((d_\alpha(x)), x) = 1 \), such that \( d_\alpha \Rightarrow d \) in \( \mu \)-measure.
PROOF. Let $M$ be the direct set of all finite subsets $Q \subset \Theta^*$, directed by inclusion. Lemma 2.4 implies that there is a prior measure $\xi_Q$ on $Q$ and associated Bayes-invariant rule $\xi_{Q^*} \in \mathcal{D}_{IN}$ such that $r(\xi_{Q^*}, \theta^*) \leq r(\xi_Q, \theta^*)$ for each $\theta^* \in Q$. Now $\{\xi_{Q^*}\}_{Q^*} \in \mathcal{D}$ is a net and $\mathcal{D}$ is sequentially compact. Hence there exists a convergent subsequence $\{\xi_{Q^*_k}\}_{Q^*_k}$ and $\xi' \in \mathcal{D}$ to which this subsequence converges, and

$$r(\xi, \theta) \geq \liminf_{Q^*_k} r(\xi_{Q^*_k}, \theta) \geq r(\xi', \theta), \ \theta \in \Theta.$$  

[see (2.2.3)].

Since $\xi$ is admissible, it follows that $r(\xi, \cdot) = r(\xi', \cdot)$. But $L(\cdot, \theta)$ is strictly convex. So $\xi$ is equivalent to $\xi'$ for otherwise $\frac{1}{2}(\xi + \xi')$ would be better than $\xi$ for all $\theta$ and strictly better for some.

The conclusion of the theorem is now seen as a consequence of Lemma 2.1. \|  

2.3. Noncompact Action Space. Even if $\mathcal{A}$ is not compact, the conclusion of Theorem 2.1 may nevertheless be true. It will be true if the problem can be suitably imbedded in another whose action space, $\mathcal{A}^*$, is compact. We will not describe how this imbedding may be attempted, acting under the assumptions of section 2.1 and that imposed in (2.2.2). Discussion related to this imbedding may be found in the work of Le Cam (1955), Kudo (1966), Farrell (1967), and more recently, Portnoy (1972). A very extensive discussion is given by Brown (1971).
The imbedding consists of appropriately extending $\mathcal{A}, \mathcal{D}, \hat{\mathcal{G}},$ and $L$ to, say $\mathcal{A}^*, \mathcal{D}^*, \hat{\mathcal{G}}^*$, and $L^*$, respectively. Here $\mathcal{A}^*$ is a compact metric space and the Borel sets of $\mathcal{A}$ are the restrictions to $\mathcal{A}$ of the Borel subsets of $\hat{\mathcal{A}}$. Each $\delta \in \mathcal{D}$ is extended to $\delta^* \in \mathcal{D}^*$ with $\delta^* (\mathcal{A}^* - \mathcal{A}, x) = 0$, and, in fact, $\mathcal{D}^*$ consists of all decision rules for the new problem with sample space $\mathcal{X}$ and action space $\mathcal{A}^*$. Each $\hat{g} \in \hat{\mathcal{G}}$ is extended to a 1:1 mapping $\hat{g}^*: \mathcal{A}^* \to \mathcal{A}^*$ in such a way that $\hat{g}^*$ is continuous and $\hat{\mathcal{G}}^* = \{ \hat{g}^* : g \in \mathcal{G} \}$ is a measurable transformation group which is homomorphic to $\mathcal{G}$. Finally $L^*$ is chosen so that each $\theta \in \Theta$, $L^*(\cdot, \theta)$ is lower semicontinuous and equal to $L(\cdot, \theta)$ on $\mathcal{A}$. Assume these extensions are possible and that the extended problem satisfies the assumptions of sections 2.1 and 2.2.

There is an additional requirement; $L^*$ must be chosen so that

$[\delta \in \mathcal{D}^* : \delta (\mathcal{A}^* - \mathcal{A}, x) = 0]$ is essentially complete in the extended problem. This condition will hold if $L^*$ is chosen so that for each $a^* \in \mathcal{A}^*$ there is an $a \in \mathcal{A}$ such that

\[(2.3.1) \quad L^*(a^*, \theta) > L(a, \theta), \text{ all } \theta.\]

This requirement implies first that an admissible $\mathcal{D}$-rule is an admissible $\mathcal{D}^*$-rule and second that a Bayes-invariant rule relative to $\mathcal{D}^*$ may be taken as the image, under $\delta \mapsto \delta^*$, of a Bayes-invariant rule relative to $\mathcal{D}$.

The requirements stated above for the imbedding imply that Theorem 2.1 is applicable to the extended problem. But the requirement and remark of the last paragraph imply the conclusion of Theorem 2.1 for the original problem.
2.4 REMARKS. The condition in the above that \( \hat{G} \) act continuously on \( \mathcal{A} \) can be replaced by the condition \( L(a, \varepsilon) \geq \beta_1 \| a \|^{1 + \alpha} (\varepsilon) \) whenever \( \| a \| > \beta_2 \) for some function \( \alpha > 0 \) and constants \( \beta_1 > 0 \). The proof is somewhat more involved, however, and for brevity is omitted.

Perhaps an alternative approach to the subject of this paper may be found in the work of Farrell (1967). However, the condition required there, that the underlying distributions have a common support would have to be avoided since it would be unnecessarily restrictive in the present problem.

The character of the proof given in section 2.1 reveals a shortcoming of the method implicit in Theorem 2.1 for proving inadmissibility. The method may establish the inadmissibility of a given estimator without providing a superior alternative. In fact, in each case where the method has been successfully applied, including that to which we turn in section 3, it has been possible to find, subsequently, by different methods, a superior alternative. Nevertheless the method, because of its generality and simplicity, does seem likely to be useful as a tool for preliminary analysis.

3. APPLICATION: ESTIMATING THE GENERALIZED VARIANCE OF THE NORMAL LAW.

Data consists of \( \mathbf{T} \), a \( p \times 1 \) multivariate normal random variable with mean \( \mu \) and covariance \( \Sigma (\Sigma > 0) \) and \( S \), a Wishart random variable with \( n \) degrees of freedom and parameter \( \Sigma \), which is independent of \( \mathbf{T} \). An estimate of \( \sigma^2 = |\hat{\Sigma}| \) is required when loss is given by \( L(\hat{\Sigma}^2 \sigma^{-2}) \). It is assumed that \( L > 0 \), that \( L \) is strictly
convex and, to avoid complication, that $L(u)$ is differentiable except possibly at $u = 1$ where its minimum is achieved. Assume that 
$L(u) \geq \alpha u + b$, $u > 0$ for some constants $\alpha > 0$, and $b$ so that in particular, the tacit reduction by sufficiency is justified. Finally, assume that if $\mu = 0$ and $\Sigma = I$,

\begin{equation}
(3.1) \quad \mathbb{E}(c \mid \Sigma) < \infty, \quad c > 0.
\end{equation}

The problem remains invariant under the full affine group, $H$, whose typical element, $(A, b), A : p \times p$ nonsingular and $b \in (-\infty, \infty)^p$ sends $(\Sigma, T), (\Sigma, \Sigma)$ and $\sigma^2$ into $(A \Sigma A', A T + b), (A \Sigma A', A \Sigma + b)$ and $|A| \sigma^2$, respectively. If a nonrandomized estimator of $\sigma^2$, $\hat{\sigma}^2$ is $H$-equivariant, it must have the form $\hat{\sigma}^2(\Sigma, T) = c \mid \Sigma \mid$ for some constant $c$. There is an optimum choice for $c$ which, by an adaptation of the argument of De Groot and Rao (1963) that uses (3.1), is that value, 
$c_0(n) = c_0$ which satisfies, when $\mu = 0$ and $\Sigma = I$,

\begin{equation}
E \mid \Sigma \mid L' \{c_0 \mid \Sigma \mid\} = 0.
\end{equation}

Is $c_0 \mid \Sigma \mid$ admissible? In the corresponding problem where $\mu$ is known to be zero and $L(u) = (u-1)^2$, Selliah (1964) shows that the optimum equivariant estimator, $c_0(n+1) \mid \Sigma + T T' \mid$, is minimax. On the other hand, in the problem obtained by setting $\Sigma = \sigma^2 I$ and $L(u) = (u-1)^2$, Stein (1964) shows that the optimum fully affine equivariant estimator is inadmissible and Stein's conclusion is extended by Brewster and Zidek (1974)
and, when $p = 1$, in Brown (1968), to include a more general choice of $L$, which is not required to be convex but only 'bowl-shaped' (see Brown (1968), for a definition). The last cited references do prove, in particular, that if in the present problem, $p = 1$, then $c_0 |S|$ is inadmissible for a large variety of loss functions. We now extend this conclusion using the method entailed in Theorem 2.1, to general $p$ with restriction to the loss structure described above. The method used here differs from those used in the last three cited references although Stein (1964) introduces this method in the course of his discussion.

The problem of concern here also remains invariant under the full linear group, $G$, whose typical element, $A; p \times p$ and nonsingular, operates like the element $(A, Q)$ of $H$. Again by a straightforward argument, the form of $G$-equivariant estimators may be found; it is

$$\varphi(z) |S|$$

for some $\varphi : [0, \infty) \rightarrow [0, \infty)$, where $Z = T \cdot S^{-1} \cdot T$. Since $c_0 |S|$ is also $G$-invariant, a method of applying Theorem 2.1 to establish the inadmissibility of $c_0 |S|$ readily suggests itself. It will be shown below that the present problem falls in the domain of Theorem 2.1 (as extended in section 2.2), and also that $c_0 |S|$ cannot be the probability limit of any sequence of Bayes-equivariant estimators (that is, Bayes estimators relative the class specified in 3.1). The required result will then have been achieved.
For the problem under consideration it is natural to let $A = (0, \infty)$. Two possible metrizable compactifications suggest themselves. The first is the usual one point compactification, $(0, \infty]$ where the open sets consist of those of $(0, \infty)$ together with all subsets of $(0, \infty)$ containing $\infty$, whose complement in $(0, \infty]$ is compact in the topology of $(0, \infty)$. But this compactification is unsatisfactory; both 
$\{n^{-1}: n = 1, 2, \ldots\}$ and $\{n: n = 1, 2, \ldots\}$ converge to $\infty$ and so to achieve a lower semicontinuous extension of $L$ when $L(u) = (u-1)^2$, for example, requires that $L^*(\infty)$ be defined so that $L^*(\infty) \leq 1$. But then it is clear that in the extended problem so obtained, there are Bayes estimators, $\hat{\sigma}^2$ such that $\hat{\sigma} = \infty$ with positive probability whatever be the true underlying distribution and this cannot be allowed.

A more natural compactification of $A$ is given by $A^* = [0, \infty]$ where the open sets are those of $(0, \infty)$ together with all subsets containing 0, $\infty$, or both whose complements in $[0, \infty)$, $(0, \infty]$ or $(0, \infty)$, respectively, are compact in $(0, \infty)$. If

$$\lim_{u \to 0} L(u) = \lim_{u \to \infty} L(u) = \infty,$$

we may define $L^*(u) = \infty$ for $u = 0$ and $\infty$ and this compactification is satisfactory in as much as the class of nonrandomized estimators, $\hat{\sigma}^2$, with $0 < \hat{\sigma} < \infty$ is then essentially complete in the extended problem. However, if $\lim_{u \to 0} L(u) < \infty$ difficulties arise because now we are required to define $L^*(0) \leq \lim_{u \to 0} L(u)$ and it is no longer easy to see whether or not the class of estimators, $\hat{\sigma}^2$, with $0 < \hat{\sigma} < \infty$ is
essentially complete. As an expedient in this case we redefine $\hat{A}$ as $[0, \infty)$ and $L(0) = \lim_{u \to 0} L(u)$; but in doing so we weaken the conclusion of inadmissibility reached below, for now this means that there exists a dominating estimator, $\hat{\sigma}$, with $0 \leq \hat{\sigma} < \infty$ rather than $0 < \hat{\sigma} < \infty$.

The stronger conclusion is valid for squared error loss (see Shorrock and Zidek (1976)), but we have not earnestly attempted to determine whether it can be achieved by the present method since, in any case, the difference is of little importance. At any rate with $\hat{A} = [0, \infty)$ we may take $\hat{A}^*$ as the one point compactification, $[0, \infty]$, and define $L^*(u) = \infty$ so that class of estimators, $\hat{\sigma}$, with $0 \leq \hat{\sigma} < \infty$ is essentially complete.

To complete the extension outlined in section 2.3, we need only extend $\hat{\sigma}$ to $\hat{\sigma}^*$ and this is trivial.

Theorem 2.1 now implies that $c_0 |\bar{S}|$ must be the limit in probability with respect to each underlying distribution of a sequence of Bayes estimators relative to the class, given in (3.2) of estimators of the form $\phi(Z) |\bar{S}|$.

The risk function of any member of this class depends on $(\Sigma, \mu)$ only through $\lambda = \mu' \Sigma^{-1} \mu$. Thus Bayes-equivariant estimators are determined from prior probability measures on $(0, \infty)$, the range of $\lambda$, and as is clear from the proof of Theorem 2.1, we may assume $\Pi$ has finite support. Such a Bayes estimator is given by $\phi_{\Pi}(Z) |\bar{S}|$ where $\phi_{\Pi}(Z)$ satisfies with $E_{\Pi}(\cdot) = \int E_{\lambda}(\cdot) \Pi(\lambda \ d\lambda)$,

$$E_{\Pi} |\bar{S}| \bar{L}'(\phi_{\Pi}(Z) |\bar{S}|) = 0.$$  

To obtain (3.3) we first observe that
for all $z, \lambda, c$ for otherwise condition (3.1) would be violated. Then we apply again the DeGroot-Rao (1963) argument to achieve (3.2).

It is easy to show that $\sup_{\lambda} \phi_\lambda(z) = \sup_{\lambda} \phi_\lambda(z)$ where the first supremum is taken overall $\Pi$ of finite support and $\phi_\lambda(z)|_{\lambda}$ denotes the Bayes-equivariant rule with respect to the degenerate prior at $\lambda$. Furthermore, by extending an argument given by Brewster and Zidek (1974),

$$\sup_{\lambda} \phi_\lambda(z) < c(z,n)$$

for $0 \leq z \leq K$ and some constant $K > 0$ and the required result is now immediate.
REFERENCES


