The Annals of Statistics
1973, Vol. 1, No. 2, 312-321

INADMISSIBILITY OF MAXIMUM LIKELIHOOD ESTMATORS IN SOME MULTIPLE REGRESSION PROBLEMS WITH THREE OR MORE INDEPENDENT VARIABLES[1]

BY A. J. BARANCHIK

Columbia University

1. Introduction and summary. Consider a multiple regression problem in which the dependent variable and (3 or more) independent variables have a joint normal distribution. This problem was investigated some time ago by Charles Stein, who proposed reasonable loss functions for various problems involving estimation of the regression coefficients and who obtained various minimax and admissibility results. In this paper we continue this investigation and establish the inadmissibility of the traditional maximum likelihood estimators. Inadmissibility is proved by exhibiting explicit procedures having lower risk than the corresponding maximum likelihood procedure. These results are given in Theorems 1 and 2 of Section 3.

2. Statement of the problem, invariance and a computational lemma. We begin this section by stating two estimation problems in multiple regression and by showing how invariance under an appropriate group of transformations may be used to simplify these problems. The section concludes with a computational lemma which will be needed in Section 3. The notation we use is chosen to coincide with that of Charles Stein [1].

Suppose $X_1, X_2, \cdots, X_n$ are independent $(p+1)$-dimensional random vectors, each distributed as a multivariate normal with mean θ and nonsingular covariance matrix Σ. We use the following partitions:

$$X_i = \begin{pmatrix} Y_i \\ Z_i \end{pmatrix}, \qquad \theta = \begin{pmatrix} \eta \\ \zeta \end{pmatrix}, \qquad i = 1, 2, \cdots, n,$$

and

$$\Sigma = \begin{pmatrix} A & B' \\ B & \Gamma \end{pmatrix},$$

where

$$Y_i,\ \eta \text{ and } A \text{ are } 1 \times 1, \qquad Z_i,\ \zeta \text{ and } B \text{ are } p \times 1.$$

Then

$$E(Y_i \mid Z_i) = \alpha + \beta' Z_i,$$

where

$$\beta = \Gamma^{-1}B \qquad \text{and} \qquad \alpha = \eta - \beta'\zeta.$$

Revised June 8, 1972.

[1] Research supported by National Science Foundation grant NSF-GP-9640 at Columbia University and by the Office of Naval Research under contract number NONR-225(72) at Stanford University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

In this notation Stein [1] showed that the following loss functions are meaningful analogues of squared error loss:

$$L((\theta, \Sigma); (\hat{\alpha}, \hat{\beta})) = \{[(\hat{\alpha} - \alpha) + (\hat{\beta} - \beta)'\zeta]^2 + (\hat{\beta} - \beta)'\Gamma(\hat{\beta} - \beta)\}/(A - B'\Gamma^{-1}B) \tag{2.1}$$

and

$$L((\beta, \Sigma), \hat{\beta}) = \{(\hat{\beta} - \beta)'\Gamma(\hat{\beta} - \beta)\}/(A - B'\Gamma^{-1}B) . \tag{2.2}$$

These loss functions define the problems of estimating (α, β) and β, respectively.

The maximum likelihood estimators, which are of constant risk relative to the above loss functions are, respectively,

$$\hat{\alpha}_0 = \bar{Y} - \hat{\beta}_0'\bar{Z} , \qquad \hat{\beta}_0 = v^{-1}U \tag{2.3}$$

and

$$\hat{\beta}_0 = V^{-1}U , \tag{2.4}$$

where

$$U = \sum_{i=1}^n Z_i Y_i - n\bar{Z}\bar{Y} \quad \text{and} \quad V = \sum_{i=1}^n Z_i Z_i' - n\bar{Z}\bar{Z}' .$$

In [1] it was shown that, for these problems, the maximum likelihood estimators are minimax. Moreover, it was shown that (2.3) is admissible for $p = 1$ and $n \geqq 6$, and that (2.4) is admissible both for $p = 1$, $n \geqq 4$ and for $p = 2$, $n \geqq 6$. Moreover, [1] contains the result that (2.4) is inadmissible for $p \geqq 3$, but does not exhibit a specific estimator which dominates (2.4). The first theorem of Section 3 finds such a dominating estimator, the second finds a procedure which dominates (2.3) for $p \geqq 3$, $n \geqq p + 2$ (thus establishing the inadmissibility of (2.3)).

The following transformations leave the problem of estimating β and (α, β) invariant [1]:

$$\begin{pmatrix} Y_i \\ Z_i \end{pmatrix} \to \begin{pmatrix} aY_i + b'Z_i + d \\ CZ_i + e \end{pmatrix},$$

$$\begin{pmatrix} \eta \\ \zeta \end{pmatrix} \to \begin{pmatrix} a\eta + b'\zeta + d \\ C\zeta + e \end{pmatrix},$$

$$\begin{pmatrix} A & B' \\ B & \Gamma \end{pmatrix} \to \begin{pmatrix} a^2A + 2ab'B + b'\Gamma b & (aB' + b'\Gamma)C' \\ C(\Gamma b + Ba) & C\Gamma C' \end{pmatrix},$$

and

$$\begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} \to \begin{pmatrix} a\hat{\alpha} - ae'C'^{-1}\hat{\beta} + d - b'C^{-1}e \\ aC'^{-1}\hat{\beta} + C'^{-1}b \end{pmatrix}, \tag{2.5}$$

where a $(\neq 0)$ and d are 1×1, b and e are $p \times 1$ and C is a nonsingular $p \times p$ matrix.

We shall restrict our attention to the subgroup with $b = 0$, the $p \times 1$ vector of p zeros. Under this group the estimators

$$\hat{\alpha} = \bar{Y} - \hat{\beta}'\bar{Z} , \qquad \hat{\beta} = f(R^2)V^{-1}U \tag{2.6}$$

are equivariant (i.e., they satisfy (2.5)). Here $f(R^2)$ is any measurable function of the sample multiple correlation coefficient

$$R^2 = U'V^{-1}U/T\,, \tag{2.7}$$

where

$$T = \sum_{i=1}^n Y_i^2 - n\bar{Y}^2\,.$$

Every orbit of the subgroup ($b = 0$) passes through the parameter point (θ, Σ) for which

$$(\zeta, \Gamma, A - B'\Gamma^{-1}B) = (0, I, 1) \tag{2.8}$$

and, since any estimators of the form (2.6) are equivariant, we may compute their risks assuming (2.8). Therefore, in the sequel, we shall assume (2.8).

We conclude with a computational lemma that we will need in Section 3 and which may be of independent interest. Let X be a $p \times 1$ random vector distributed normally with mean θ and identity covariance matrix. For any p-vector v we set $||v||^2 = \sum_1^p v_i^2$, where v_i is the ith component of v, e.g., for the remainder of this section X_i will denote the ith coordinate of X.

LEMMA 1.

$$E\left(\frac{\theta_i X_i}{||X||^2}\right) = \frac{\theta_i^2}{||\theta||^2} E\left(\frac{2K}{p-2+2K}\right),$$

where K has a Poisson distribution with parameter $||\theta||^2/2$.

PROOF.

$$\begin{aligned}
E\left(\frac{\theta_i X_i}{||X||^2}\right) &= \theta_i^2 E(||X||^{-2}) + E(\theta_i(X_i - \theta_i)||X||^{-2}) \\
&= \theta_i^2 E(p-2+2K)^{-1} \\
&\quad + \theta_i \frac{d}{d\theta_i}\left[(2\pi)^{-p/2} \int \cdots \int ||x||^{-2} \exp[-\Sigma(x_i - \theta_i)^2]/2\, dx_1 \cdots dx_p\right] \\
&= \theta_i^2 E(p-2-2K)^{-1} \\
&\quad + \theta_i \frac{d}{d\theta_i} \sum_0^\infty e^{-||\theta||^2/2}(||\theta||^2/2)^k/(k!\,(p-2+2k)) \\
&= \theta_i^2\{E(p-2-2K)^{-1} \\
&\quad + \sum_0^\infty e^{-||\theta||^2/2}[k(||\theta||^2/2)^{k-1} - (||\theta||^2/2)^k]/(k!\,(p-2+2k))\} \\
&= \theta_i^2 \sum_0^\infty e^{-||\theta||^2/2}[(||\theta||^2/2)^k - (||\theta||^2/2)^k + k(||\theta||^2/2)^{k-1}]/(k!\,(p-2+2k)) \\
&= \theta_i^2 \sum_0^\infty e^{-||\theta||^2/2} k(||\theta||^2/2)^{k-1}/(k!\,(p-2+2k)) \\
&= \frac{\theta_i^2}{||\theta||^2} \sum_0^\infty e^{-||\theta||^2/2}(2k)(||\theta||^2/2)^k/(k!\,(p-2+2k)) \\
&= \frac{\theta_i^2}{||\theta||^2} E\left[\frac{2K}{p-2+2K}\right],
\end{aligned}$$

as was to be proved.

3. Inadmissibility results. We first state our main results in the form of two theorems and then prove the second by showing how it follows from the first. The section is concluded by the quite lengthy proof of the first theorem.

In the notation of Section 2 and assuming henceforth that p, the number of "independent variables," is at least 3 and that the sample size $n \geqq p + 2$, we will prove:

THEOREM 1. *For the problem of estimating β with loss given by* (2.2), *the estimator*

$$\hat{\beta}_c = (1 - c(1 - R^2)R^{-2})V^{-1}U\,, \tag{3.1}$$

dominates the maximum likelihood estimator $\hat{\beta}_0 = V^{-1}U$, for any value of $c \in (0, 2(p-2)(n-p+2)^{-1})$, with $p \geqq 3$ and $n \geqq p + 2$.

THEOREM 2. *For the problem of estimating (α, β) with loss given by* (2.1), *the estimator*

$$(\hat{\alpha}_c = \bar{Y} - \hat{\beta}_c'\bar{Z},\ \hat{\beta}_c = (1 - c(1 - R^2)R^{-2}V^{-1}U) \tag{3.2}$$

dominates the maximum likelihood estimator $(\hat{\alpha}_0, \hat{\beta}_0)$ for any value of $c \in (0, 2(p-2)(n-p+2)^{-1})$, with $p \geqq 3$ and $n \geqq p+2$.

We remark that in [1] it was shown that, if the regression function is thought of as a predictor, then the risk with respect to the loss function given by (2.1) measures the mean-squared error incurred by using a regression function (formed by estimated regression coefficients) to predict the value of an unobserved dependent variable when the p independent variables associated with that dependent variable are observed. Thus, Theorem 2 implies that, in terms of this measure of accuracy, a regression function formed by estimating its parameters by (3.2) is a better predictor than one formed by estimating those parameters with their maximum likelihood estimators.

We first show how Theorem 2 follows from Theorem 1.

PROOF OF THEOREM 2. For convenience, when no confusion will obtain, we shall write the risk (excepted loss) $R((\theta, \Sigma); (\hat{\alpha}, \hat{\beta})) = \rho(\hat{\alpha}, \hat{\beta})$. Also, comparing (2.6) with (3.2), we get $1 - c(1 - R^2)R^{-2} = f(R^2)$. Since, therefore, (3.2) is equivariant relative to the subgroup given in Section 2, we may simplify the loss function (2.1) to

$$(\hat{\alpha}_c - \alpha)^2 + ||\hat{\beta}_c - \beta||^2\,, \tag{3.3}$$

by taking $\zeta = 0$, $\Gamma = I$ and $A - B'\Gamma^{-1}B = 1$ (see (2.8)). The following lemma, combined with Theorem 1, will give us Theorem 2:

LEMMA 2. *If an estimator of the form $\hat{\beta} = f(R^2)V^{-1}U$ dominates $\hat{\beta}_0 = V^{-1}U$ relative to the loss function* (2.2), *then the corresponding estimator given by* (2.6) *dominates $(\hat{\alpha}_0, \hat{\beta}_0)$ relative to the loss function* (2.1).

PROOF OF LEMMA 2. (Note that, for this lemma, f is an arbitrary measurable function of the multiple regression coefficient R^2.) Setting $g(R^2) = 1 - f(R^2)$

our risk becomes

$$\begin{aligned}
E\{[\bar{Y} - f(R^2)U'V'^{-1}\bar{Z} - \alpha]^2 + ||f(R^2)V^{-1}U - \beta||^2\} \\
= E[\bar{Y} - U'V'^{-1}\bar{Z} - \alpha]^2 + E||V^{-1}U - \beta||^2 \\
+ E[g(R^2)U'V'^{-1}\bar{Z}]^2 + E||g(R^2)V^{-1}U||^2 \\
+ 2E[g(R^2)U'V'^{-1}\bar{Z}(\bar{Y} - U'V'^{-1}\bar{Z} - \alpha) + g(R^2)U'V'^{-1}(\beta - V^{-1}U)] \\
= \rho(\hat{\alpha}_0, \hat{\beta}_0) + E[(g(R^2))^2 U'V'^{-1}(I + \bar{Z}\bar{Z}')V^{-1}U] \\
+ 2E\{E[g(R^2)U'V'^{-1}(\beta - V^{-1}U + \bar{Z}(\bar{Y} - U'V'^{-1}\bar{Z} - \alpha)) \mid \bar{Z}]\} .
\end{aligned}$$

Now the last term of this expression equals

$$2E\{g(R^2)U'V'^{-1}[\beta - V^{-1}U + \bar{Z}\bar{Z}'(\beta - V^{-1}U)]$$

which, in turn, is equal to

$$2E[g(R^2)U'V'^{-1}(I + \bar{Z}\bar{Z}')(\beta - V^{-1}U)] .$$

Therefore,

$$\begin{aligned}
(3.4) \qquad \rho(\hat{\alpha}, \hat{\beta}) = \rho(\hat{\alpha}_0, \hat{\beta}_0) + E[g(R^2)U'V'^{-1}I(2\beta - (2 - g(R^2))V^{-1}U)] \\
+ E[g(R^2)U'V'^{-1}\bar{Z}\bar{Z}'(2\beta - (2 - g(R^2))V^{-1}U)] .
\end{aligned}$$

Combining the last two terms of this expression we have, since $E(\bar{Z}) = 0$, the covariance of $\bar{Z}$ is $n^{-1}I$, and $\bar{Z}$ is independent of U, V, and R, that their sum equals

$$\begin{aligned}
\frac{n+1}{n} E[g(R^2)U'V'^{-1}(2\beta - (2 - g(R^2))V^{-1}U)] \\
= \frac{n+1}{n} E\{2g(R^2)U'V'^{-1}(\beta - V^{-1}U) + ||g(R^2)V^{-1}U||^2\} \\
= \frac{n+1}{n}\{-2E[g(R^2)\hat{\beta}_0'(\hat{\beta}_0 - \beta)] + E||g(R^2)\hat{\beta}_0||^2\} .
\end{aligned}$$

Now, upon adding and subtracting $E||\hat{\beta}_0 - \beta||^2$ to the expression in curly brackets, and recalling that $f(R^2) = 1 - g(R^2)$, we have

$$\frac{n+1}{n}\{E||f(R^2)\hat{\beta}_0 - \beta||^2 - E||\hat{\beta}_0 - \beta||^2\} ,$$

which quantity is, by hypothesis, not positive and at some point is negative. Therefore, by (3.4), $\rho(\hat{\alpha}, \hat{\beta}) \leqq \rho(\hat{\alpha}_0, \hat{\beta}_0)$ with strict inequality at some point. This proves Lemma 2 and, *a fortiori*, Theorem 2.

We now deal with Theorem 1.

PROOF OF THEOREM 1. By equivariance $\hat{\beta}_c$ (see (2.8)) we again may assume that

$$(3.5) \qquad (\zeta, \Gamma, A - B'\Gamma^{-1}B) = (0, I, 1) .$$

Then the loss function given by (2.2) becomes $||\hat{\beta} - \beta||^2$, and, when no confusion will result, we shall denote the risk of an equivariant estimator $\hat{\beta}$ by $\rho(\hat{\beta})$.

Using (2.4) and (2.7) we may rewrite the estimator given in (3.1) as

$$\hat{\beta}_c = (1 - c(T - \hat{\beta}_0' V\hat{\beta}_0)/(\hat{\beta}_0' V\hat{\beta}_0))\hat{\beta}_0 \tag{3.6}$$

where, on the parameter space defined by (3.5), $T - \hat{\beta}_0' V\hat{\beta}_0$ is independent of $\hat{\beta}_0 V\hat{\beta}_0$ and of $\hat{\beta}_0$. Then the loss

$$||\hat{\beta}_c - \beta||^2 = ||\hat{\beta}_0 - \beta||^2 - 2c\frac{T - \hat{\beta}_0' V\hat{\beta}_0}{\hat{\beta}_0' V\hat{\beta}_0}\hat{\beta}_0'(\hat{\beta}_0 - \beta) + c^2\frac{(T - \hat{\beta}_0' V\hat{\beta}_0)^2}{(\hat{\beta}_0' V\hat{\beta}_0)^2}\hat{\beta}_0'\hat{\beta}_0 .$$

Since $T - \hat{\beta}_0' V\hat{\beta}_0$ is distributed as χ^2_{n-p}, the risk

$$\begin{aligned}\rho(\hat{\beta}_c) = {} & E||\hat{\beta}_0 - \beta||^2 - 2c(n-p)E[\hat{\beta}_0'(\hat{\beta}_0 - \beta)/(\hat{\beta}_0' V\hat{\beta}_0)] \\ & + c^2(n-p)(n-p+2)E[\hat{\beta}_0'\hat{\beta}_0/(\hat{\beta}_0' V\hat{\beta}_0)^2] .\end{aligned} \tag{3.7}$$

To prove the theorem we must show that $\rho(\hat{\beta}_c) - E||\hat{\beta}_0 - \beta||^2 \leqq 0$, with strict inequality somewhere in the parameter space, for the specified values of c. The following transformation will be of use:

$$V = QDQ' , \tag{3.8}$$

where D is a diagonal matrix and Q is orthogonal. Noting that V has a Wishart distribution $W(p, n; nI_p)$ and that $\hat{\beta}_0$ given V (denoted henceforth as $\hat{\beta}_0 | V$) is normally distributed with mean β_0 and covariance matrix V^{-1}, then defining

$$Z = Q'V^{\frac{1}{2}}\hat{\beta}_0 = Q'(QD^{\frac{1}{2}}Q')\hat{\beta}_0 = D^{\frac{1}{2}}Q'\hat{\beta}_0 , \tag{3.9}$$

$Z | V$ has a normal distribution with mean $D^{\frac{1}{2}}Q'\beta$ and covariance matrix I_p. The diagonal elements of D will be denoted by $d_1, d_2, \cdots, d_p$.

This transformation enables us to rewrite some of the expressions appearing in (3.7) as

$$\hat{\beta}_0'\hat{\beta}_0 = Z'D^{-1}Z \qquad \text{and} \qquad \hat{\beta}_0' V\hat{\beta}_0 = Z'Z . \tag{3.10}$$

The relevant terms of the risk (3.7) are next evaluated in a sequence of computational lemmas.

LEMMA 3.

$$E\left[\frac{\hat{\beta}_0'\beta}{\hat{\beta}_0' V\hat{\beta}_0}\right] = h(||\beta||^2, n)\sum_0^\infty \frac{\Gamma(n/2 + k - 1)2k}{k!\,(p - 2 + 2k)}\gamma^k ,$$

where $h(||\beta||^2, n) = (2\Gamma(n/2)(1 + ||\beta||^2)^{n/2-1})^{-1}$, $\gamma = ||\beta||^2/(1 + ||\beta||^2)$, *and* $\Gamma(\cdot)$ *is the gamma function.*

PROOF OF LEMMA 3. First we will compute the conditional expectation

$$E\left\{\frac{\hat{\beta}_0'\beta}{\hat{\beta}_0' V\hat{\beta}_0}\middle| V\right\} . \tag{3.11}$$

By virtue of (3.9) and (3.10) this becomes, denoting conditional expectation given Q, D by E^*,

$$E^*\{Z'D^{-\frac{1}{2}}Q'\beta/Z'Z\}$$

which is equal to the sum of expectations

$$(3.12)\qquad E^*\{Z_{(i)}(D^{-\frac{1}{2}}Q'\beta)_i/Z'Z\} \qquad i = 1, 2, \cdots, p,$$

where $(v)_i$ of the vector v is its ith component.

Lemma 1 (Section 2) is now used to rewrite (3.12) as

$$(D^{-\frac{1}{2}}Q'\beta)_i E^*\{Z_{(i)}/Z'Z\} = (D^{-\frac{1}{2}}Q'\beta)_i \frac{E^*(Z_{(i)})}{\sum_{i=1}^p [E^*(Z_{(i)})]^2} E^*\left\{\frac{2K}{p-2+2K}\right\}$$

where K (given Q, D) is Poisson with parameter $\sum_{i=1}^p [E(Z_{(i)})]^2/2$. Since $E^*(Z_{(i)}) = (D^{\frac{1}{2}}Q'\beta)_i$, (3.12) becomes, after some algebra, $(Q'\beta)_i^2(\beta'V\beta)^{-1}E^*[2K(p-2+2K)^{-1}]$. Summing on i, we have that (3.11) equals

$$(3.13)\qquad \beta'\beta(\beta'V\beta)^{-1}E\left\{\frac{2K}{p-2+2K}\,\middle|\, V\right\},$$

where K is Poisson with parameter $\beta'V\beta/2$ $(= \sum_{i=1}^p (D^{\frac{1}{2}}Q'\beta)_i^2/2)$.

Before concluding the proof of this lemma by averaging (3.13) with respect to the distribution of V we will make an orthogonal transformation, sending Z_i, (not to be confused with $Z_{(i)}$, which is simply the ith component of the Z defined in (3.9)) into $\mathscr{O}Z_i$, 1, 2, $\cdots$, p, in such a manner as to map β into a vector all of whose components are zero save the first, β_1. This transformation maps $\eta \to \mathscr{O}\eta$, $B \to \mathscr{O}B$, $\Gamma \to \mathscr{O}\Gamma\mathscr{O}'$ and, therefore, $\beta \to \mathscr{O}\beta$, with analogous mappings of Z_i, U, V and $\hat{\beta}_0$. Since the group of these transformations is a subgroup of our original subgroup ($b = 0$) given in Section 2, an orthogonal transformation does not alter the risk of an equivariant estimator. Moreover (3.5) remains invariant. We therefore may assume, without loss of generality, that such an orthogonal transformation has been made.

Denoting the first row, first column entry of V by V_{11}, (3.13) equals (assuming the above orthogonal transformation) $V_{11}^{-1}E\{(2K/(p-2+2K))\,|\,V\}$, where the Poisson parameter may now be written as $||\beta||^2V_{11}/2$. Therefore, the unconditional expectation

$$\begin{aligned} E\left[\frac{\hat{\beta}_0'\beta}{\hat{\beta}_0'V\hat{\beta}_0}\right] &= E[V_{11}^{-1}\textstyle\sum_{k=0}^\infty \exp[-||\beta||^2V_{11}/2](||\beta||^2V_{11}/2)^k(2k/k!)(p-2+2k)^{-1}] \\ &= E[\exp[-||\beta||^2V_{11}/2]\textstyle\sum_{k=0}^\infty V_{11}^{k-1}(||\beta||^2/2)^k(2k/k!)(p-2+2k)^{-1}] \\ &= \textstyle\sum_{k=0}^\infty 2^{k-1}\Gamma(n/2+k-1)(||\beta||^2/2)^k \\ &\qquad \times 2k[k!\,(1+||\beta||^2)^{k-1+n/2}\Gamma(n/2)(p-2+2k)] \end{aligned}$$

because

$$E[\exp[-||\beta||^2V_{11}/2]V_{11}^j] = 2^j\Gamma(n/2+j)[(1+||\beta||^2)^{j+n/2}\Gamma(n/2)]^{-1}.$$

The lemma is therefore proved.

LEMMA 4.

$$E\left[\frac{\hat{\beta}_0'\hat{\beta}_0}{(\hat{\beta}_0'V\hat{\beta}_0)^2}\right] = h(||\beta||^2, n)\textstyle\sum_{k=0}^\infty l_k(||\beta||^2, n, p)\gamma^k$$

where h and γ are as in Lemma 3 and

$$l_k(\|\beta\|^2, n, p) = \frac{\Gamma(n/2+k-1)}{k!\,(2k+p)(2k+p-2)} \times \left[\frac{n-2}{n-p-1} + 2k + \frac{(n+2k-2)(p-1)}{(1+\|\beta\|^2)(n-p-1)}\right].$$

PROOF OF LEMMA 4. Proceeding as in the proof of Lemma 3 the conditional expectation given V is obtained, using the transformation (3.8) and the orthogonal transformation given after (3.13). Defining E^* as the conditional expectation given D, Q, we obtain

$$(3.14)\quad \begin{aligned} E\left\{\frac{\hat\beta_0'\hat\beta_0}{(\hat\beta_0' V\hat\beta_0)^2}\,\Big|\, V\right\} &= E^*\,[(\sum_{i=1}^p Z_{(i)}^2/d_i)/(\sum_{i=1}^p Z_{(i)}^2)^2] \\ &= E^*\left[\sum_{i=1}^p\left(\frac{2K_i+1}{d_i}\right)(\sum_{j=1}^p 2K_j+p)^{-1} \times (\sum_{j=1}^p 2K_j + p - 2)^{-1}\right], \end{aligned}$$

where $K_i\,|\,(D, Q)$ is Poisson with parameter $d_i(Q'\beta)_i^2/2$. Now this equals

$$\begin{aligned} &E^*\left\{E\left[\frac{\Sigma(2K_i+1)/d_i}{(2\Sigma K_i+p)(2\Sigma K_i+p-2)}\,\Big|\,\Sigma K_i, D, Q\right]\right\} \\ &= E^*\{(2\Sigma K_j+p)^{-1}(2\Sigma K_j+p-2)^{-1} \\ &\qquad\times[\Sigma d_i^{-1} + 2(\Sigma K_j)(\sum_{i=1}^p (Q'\beta)_i^2)(\sum_{j=1}^p d_j(Q'\beta)_j^2)^{-1}]\} \\ &= E\{(2K+p)^{-1}(2K+p-2)^{-1}[\text{trace } V^{-1} + 2K\|\beta\|^2(\beta' V\beta)^{-1}]\,|\, V\}, \end{aligned}$$

where K is Poisson with parameter $\beta' V\beta/2$. Making the same orthogonal transformation as before,

$$(3.15)\quad E\left[\frac{\hat\beta_0'\hat\beta_0}{(\hat\beta_0' V\hat\beta_0)^2}\right] = E[(2K+p)^{-1}(2K+p-2)^{-1}(\text{trace } V^{-1} + 2KV_{11}^{-1})],$$

where K is Poisson with parameter $\|\beta\|^2 V_{11}/2$. Using the fact that

$$E\{\text{trace } V^{-1}\,|\, V_{11}\} = (n-p-1)^{-1}(V_{11}^{-1}(n-2) + p - 1),$$

we have (3.15) equal to

$$\begin{aligned} &E[(2K+p)^{-1}(2K+p-2)^{-1}(E\{\text{trace } V^{-1}\,|\, V_{11}\} + 2KV_{11}^{-1})] \\ &= E\left[(2K+p)^{-1}(2K+p-2)^{-1}\left\{\left(\frac{n-2}{n-p-1}+2K\right)V_{11}^{-1} + (p-1)(n-p-1)^{-1}\right\}\right] \\ &= E\left[V_{11}^{-1}E\left\{(2K+p)^{-1}(2K+p-2)^{-1}\left(\frac{n-2}{n-p-1}+2K\right)\Big|\, V_{11}\right\}\right] \\ &\qquad + (p-1)(n-p-1)^{-1}E[(2K+p)(2K+p-2)]^{-1}. \end{aligned}$$

Evaluating the inner expectation, some arithmetic leads to the desired result.

A computation exactly paralleling that in the proof of Lemma 4 leads us to

LEMMA 5.

$$E\left[\frac{\hat{\beta}_0'\hat{\beta}_0}{\hat{\beta}_0'V\hat{\beta}_0}\right] = h(||\beta||^2, n) \sum_0^\infty (2k + p - 2)l_k(||\beta||^2, n, p)\gamma^k ,$$

where h and γ are given in Lemma 3 and l_k is given in Lemma 4.

PROOF OF THEOREM 1. (3.7) leads one to conclude that β_c will have lower risk than $\hat{\beta}_0$ provided that

$$0 < c < 2(n - p + 2)^{-1}E\left[\frac{\hat{\beta}_0'(\hat{\beta}_0 - \beta)}{\hat{\beta}_0'V\hat{\beta}_0}\right]\Big/E\left[\frac{\hat{\beta}_0'\hat{\beta}_0}{(\hat{\beta}_0'V\hat{\beta}_0)^2}\right], \tag{3.16}$$

for all possible parameter values satisfying Condition (2.8). It will now be shown that $c = 2(p - 2)(n - p + 2)^{-1}$ (and, hence, all positive c less than this) satisfies (3.16). This is equivalent to establishing

$$(p - 2)E[\hat{\beta}_0'\hat{\beta}_0(\hat{\beta}_0'V\hat{\beta}_0)^{-2}] - E[\hat{\beta}_0'(\hat{\beta}_0 - \beta)(\hat{\beta}_0'V\hat{\beta}_0)^{-1}] < 0 . \tag{3.17}$$

Using the results of Lemmas 3, 4 and 5 and dropping the positive common factor $h(||\beta||^2, n)$, (3.17) will be satisfied if

$$\sum_{k=1}^\infty t_k\{n - p - 2 - (1 - \gamma)(n + 2k - 2)\} < 0 , \tag{3.18}$$

where $t_k = (p - 1)\Gamma(n/2 + k - 1)2k\gamma^k/(k!\,(p + 2k)(p - 2 + 2k)(n - p - 1))$ and we note that $1 - \gamma = (1 + ||\beta||^2)^{-1}$. Condition (3.18) becomes, upon collecting terms,

$$\sum_{k=1}^\infty t_k(-2k - p) + \sum_{j=0}^\infty t_j\gamma(n + 2j - 2) < 0$$

which becomes, upon setting $j = k - 1$ and noting that

$$t_{k-1} = 2t_k(k - 1)(p + 2k)(n + 2k - 4)^{-1}(p + 2k - 4)^{-1}\gamma^{-1} ,$$

$$\sum_{k=1}^\infty t_k\left[(-2k - p) + (n + 2k - 4) \times \frac{2(k - 1)(p + 2k)}{(n + 2k - 4)(p + 2k - 4)}\right] < 0 . \tag{3.19}$$

The left-hand side of (3.19) is

$$\sum_{k=1}^\infty t_k(p + 2k)[-1 + 2(k - 1)(p + 2k - 4)^{-1}] = \sum_{k=1}^\infty t_k(p + 2k)(2 - p)(p + 2k - 4)^{-1} ,$$

and, since $p \geqq 3$, each term in this sum is indeed negative and thus Condition (3.19) holds and the theorem is proved.

Acknowledgment. The author wishes to thank Charles Stein both for bringing these problems to his attention and for valuable assistance in attaining their solutions.

REFERENCE

[1] STEIN, CHARLES (1960). *Multiple Regression. Contributions to Probability and Statistics*, Stanford Univ. Press.

DEPARTMENT OF MATHEMATICS
HUNTER COLLEGE
NEW YORK, NEW YORK 10021