The Annals of Statistics
1978, Vol. 6, No. 4, 769-782

DERIVING UNBIASED RISK ESTIMATORS OF MULTINORMAL MEAN AND REGRESSION COEFFICIENT ESTIMATORS USING ZONAL POLYNOMIALS[1]

BY JIM ZIDEK

University of British Columbia

Unbiased risk estimators are derived for estimators in certain classes of equivariant estimators of multinormal matrix means, ξ, and regression coefficients β. In all cases the covariance matrix is unknown. The underlying method, a multivariate version of that of James and Stein (1960), uses zonal polynomial expansions for the distributions of noncentral statistics. This gives, in one case, the required generalization of the Pitman-Robbins representation of noncentral chi-square statistics including the appropriate multivariate Poisson law. In the other case, a multivariate negative binomial law emerges. The result for regression coefficients suggests a new minimax estimator and, essentially, an extension of Baranchik's result.

1. Introduction. This report shows how certain decision theoretical results concerning the parameters of the univariate normal law may be extended to the multivariate case. This is done in terms of two examples of interest in their own right.

The first problem of concern in this report is that of deriving unbiased estimators of the risks of members of certain classes of estimators of ξ, a $p \times k$ matrix of multinormal means. The class of estimators selected depends on how much is known about the covariance structure. In the most general case treated, it is assumed that the observable sufficient statistic is (X, S) where

$$\begin{aligned} & X \sim N(\xi, \Sigma \bar{\otimes} 1_k) \\ & S \sim W_p(n, \Sigma) \\ & X \quad \text{and} \quad S \quad \text{are independent} \\ & \xi \quad \text{and} \quad \Sigma \quad \text{are unknown} \\ & X = (X_1, \cdots, X_k), \qquad \xi = (\xi_1, \cdots, \xi_k) \end{aligned} \tag{1.1}$$

so that X is a normally distributed $p \times k$ $(p < k)$ matrix with independent columns $X_i \sim N_p(\xi_i, \Sigma)$. This example and the case $\Sigma = \sigma^2 1_p$ are treated in Section 2.

Received March 1977; revised October 1977.

[1] The National Research Council of Canada, Canada Council and the Office of Naval Research (N00014-67-A-0112-0085) of the United States all contributed support for this research. The work was done at the Statistics Department, Stanford University and the Division of Mathematics and Statistics, CSIRO (Canberra) both of which generously provided facilities.

AMS 1970 *subject classifications.* Primary 62C15; Secondary 62F10, 62H10.

Key words and phrases. Unbiased risk estimators, minimax estimators, multivariate Poisson, multivariate negative binomial, James-Stein estimator, multivariate regression, zonal polynomials, Pitman-Robbins representation.

The case $\Sigma = 1_p$ is considered by Stein (1973b) where the use of unbiased risk estimation was introduced.

The second problem is that of estimating the matrix of multivariate regression coefficients, $\beta: q \times p$, under a quadratic loss function. The unbiased risk estimator is found via a zonal polynomial expansion involving, implicitly, a multivariate negative binomial distribution.

It is hoped that these unbiased risk estimators will facilitate a systematic search for superior alternatives to the commonly used, i.e., maximum likelihood, estimators. Such a search is not undertaken here. However we do consider a natural generalization of the Baranchik (1973)—Stein (1960) estimator of regression coefficients and it is shown in Section 3 that this estimator dominates the usual estimator. The result is (essentially) an extension of Baranchik's result (1973).

The main results obtained are given in (2.2.12), (2.2.3) and (3.9). The method used is implicit in the work of Shorrock and Zidek (1974) where another such example is given; it has been applied by Lillestöl (1975, page 67 ff and page 88 ff) in the complex counterpart of one of the problems treated here. This method is a multivariate version of the method introduced by James and Stein (1960) to prove the superiority of their simultaneous estimator of the means of independent univariate normal laws over the usual one when loss is quadratic. Their method uses the representation given by Pitman and Robbins (1949) of the noncentral chi-squared random variable. If $\chi^2(k, \delta)$ denotes the noncentral chi-square random variable with k degrees of freedom and noncentrality parameter δ, this representation is

$$\chi^2(k, \delta) \sim \chi^2(k + 2\tilde{\kappa}) . \tag{1.2}$$

Here $\chi^2(\nu)$ denotes the central chi-squared random variable with ν degrees of freedom, $\tilde{\kappa}$ is the Poisson random variables with mean $\frac{1}{2}\delta^2$, and given $\tilde{\kappa} = \kappa$, $\chi^2(k + 2\tilde{\kappa}) = \chi^2(k + 2\kappa)$ is independent of $\tilde{\kappa}$. The multivariate version of this representation does not seem to be obtained, as one might expect, by replacing the chi-squared random variables by Wishart random variables. Rather, let $W_p(k, \delta, 1_p)$ denote the noncentral Wishart random variable, i.e., the random $p \times p$ matrix distributed as XX' where $X \sim N(\delta, 1_p \bar{\otimes} 1_k)$, and let R denote the diagonal matrix with diagonal elements its latent roots. Then the multivariate version of the Pitman–Robbins representation is

$$R[W_p(k, \delta, 1_p)] \sim \tilde{R}_p(k, 2\tilde{\kappa}) . \tag{1.3}$$

Here $\tilde{R} = \tilde{R}_p(k, 2\tilde{\kappa}) > 0$, i.e., $\tilde{R}$, positive definite, is a random diagonal matrix, say diag $\{\tilde{R}_1, \cdots, \tilde{R}_p\}$, $\tilde{R}_1 > \cdots > \tilde{R}_p > 0$, with density element proportional to

$$C_\kappa(\tilde{r}) \exp[-\tfrac{1}{2} \operatorname{tr}(\tilde{r})]|\tilde{r}|^{(k-p-1)/2} \textstyle\prod_{i<j} (\tilde{r}_i - \tilde{r}_j)\, d\tilde{r} \, ; \tag{1.4}$$

C_κ denotes the zonal polynomial of index $\kappa = (\kappa_1, \cdots, \kappa_p)$, $\kappa_1 \geqq \cdots \geqq \kappa_p \geqq 0$, the κ_i being integers. And $\tilde{\kappa} = (\tilde{\kappa}_1, \cdots, \tilde{\kappa}_p)$, $\tilde{\kappa}_1 \geqq \cdots \geqq \tilde{\kappa}_p$ is a random partition of a random integer which has what might be called a multivariate Poisson distribution, namely

$$P\{\tilde{\kappa} = \kappa\} = \exp[-\operatorname{tr} \tfrac{1}{2}\delta\delta']C_\kappa(\tfrac{1}{2}\delta\delta')/||\kappa||! = \pi_{\frac{1}{2}\delta\delta'}(\kappa)\,, \tag{1.5}$$

say, where $||\kappa|| = \kappa_1 + \cdots + \kappa_p$.

The representation just described is derived implicitly in Shorrock and Zidek (1974) and again implicitly in the derivation of equation (2.1.3). Observe that if $p = 1$ the distributions determined by (1.4) and (1.5) are, since $C_\kappa(A) = A^\kappa$ in this case, the central chi-squared with $k + 2\kappa$ degrees of freedom and the Poisson distribution with mean $\frac{1}{2}\delta^2$. It should be noted that, whatever be p, $C_\kappa(A) > 0$ (see James (1968)) for all κ if $A > 0$, i.e., A is positive definite, and hence $P\{\tilde{\kappa} = \kappa\} > 0$ for all κ; also $(\operatorname{tr} A)^K = \sum_{\kappa:||\kappa||=K} C_\kappa(A)$, so $||\tilde{\kappa}||$ has the univariate Poisson distribution with mean $\operatorname{tr} \frac{1}{2}\delta\delta'$, and therefore $\sum_\kappa P\{\tilde{\kappa} = \kappa\} = 1$.

The representation given in (1.3) is due to James (see James (1964)). It and variants of it such as that in Section 3 have been used to obtain series expansions of distributions of noncentral random variables in terms of zonal polynomials. An extensive list of references to such results may be found in Fujikoshi (1970).

Before proceeding further we review the relevant properties of zonal polynomials which are elaborated in James (1964) and Farrell (1976). These polynomials, $C_\kappa(S)$, are homogeneous symmetric polynomials in the latent roots of the $p\times p$ matrix, S indexed by the class of all ordered sequences, $\kappa=\{\kappa_1\geqq\cdots\geqq\kappa_p\}$ of p nonnegative integers. The degree of C_κ is $||\kappa||$, and C_κ is so normalized that

$$(\operatorname{tr} S)^k = \textstyle\sum_{\{\kappa:||\kappa||=k\}} C_\kappa(S)\,, \qquad k = 0, 1, 2, \cdots.$$

James (1968) proves C_κ is nonnegative on the space of nonnegative definite matrices. The essential properties of these polynomials are:

$$\int \exp\{\operatorname{tr} A\hat{O}\}\hat{\lambda}(d\hat{O}) = \textstyle\sum_\kappa \alpha_\kappa^{(n)}C_\kappa(AA^t) \tag{1.6}$$

$$\int C_\kappa(S_1\hat{O}S_2\hat{O}^t)\hat{\lambda}(d\hat{O}) = C_\kappa(S_1)C_\kappa(S_2)/C_\kappa(1_p) \tag{1.7}$$

and

$$E^0C_\kappa(UU^t) = C_\kappa(\tfrac{1}{2}1_p)/(\alpha_\kappa^{(n)}||\kappa||!) \tag{1.8}$$

where $\hat{\lambda}$ denotes the invariant probability measure on the group of $n \times n$ orthogonal matrices $\hat{O}$, $\alpha_\kappa^{(n)}$ is given in James (1964), $A: p \times n$ and $S_i: p \times p$ are arbitrary, the latter being symmetric, and $U \sim N_p(O, I_p \otimes I_n)$. Observe that $\int (\operatorname{tr} A\hat{O})^r\hat{\lambda}(d\hat{O}) = 0$ whenever r is an odd integer.

2. Multinormal means.

2.1. $\Sigma > 0$. In this case assume that in addition to $X \sim N(\xi, \Sigma \bar{\otimes} 1_k)$, $S \sim W_p(n, \Sigma)$ independent of X is to be observed, that ξ is to be estimated and that loss is determined by $L(\hat{\xi}; \Sigma, \xi) = \operatorname{tr} \Sigma^{-1}(\hat{\xi} - \xi)(\hat{\xi} - \xi)'$. This problem remains invariant under the transformation group under which

$$X \to AXO'\,, \qquad S \to ASA'\,,$$
$$\xi \to A\xi O'\,, \qquad \Sigma \to A\Sigma A'\,,$$

where A is nonsingular $p \times p$ and O is orthogonal $k \times k$. It is more convenient in subsequent calculations to choose, for the sufficient statistic, $(X, S + XX')$. Then, as shown below, if ξ is an equivariant estimator of ξ, $\hat{\xi}$ must have the form

$$\hat{\xi}(X, S + XX') = [1_p - \tilde{P}V^{\frac{1}{2}}\tilde{B}\phi(T)\tilde{B}'V^{-\frac{1}{2}}\tilde{P}']X ,$$

where ϕ is a diagonal $p \times p$ matrix and $\tilde{P}$, $\tilde{B}$ are, respectively, the $p \times p$ orthogonal matrices for which $S + XX' = \tilde{P}V\tilde{P}'$ and $V^{-\frac{1}{2}}\tilde{P}'XX'\tilde{P}V^{-\frac{1}{2}} = \tilde{B}T\tilde{B}'$ where $V = \operatorname{diag}\{V_1, \cdots, V_p\}$, $T = \operatorname{diag}\{T_1, \cdots, T_p\}$, i.e., the $\{T_i\}$ are the roots of $|XX' - T_i(S+XX')| = 0$ and with probability one $V_1 > \cdots > V_p$ and $T_1 > \cdots > T_p$. Here $\tilde{B}$ is the unique $p \times p$ orthogonal matrix with nonnegative first row while $\tilde{P}$ is chosen, for convenience in the sequel, at random from among the 2^p possible choices that work, each choice being equiprobable. To obtain this representation, note that equivariance implies

$$\begin{aligned}\hat{\xi}(X, S + XX') &= \tilde{P}V^{\frac{1}{2}}\mathrm{O}\hat{\xi}(\mathrm{O}'V^{-\frac{1}{2}}\tilde{P}'X, 1_p) \\ &= \tilde{P}V^{\frac{1}{2}}\mathrm{O}\rho(\mathrm{O}'R)\end{aligned}$$

where $R = V^{-\frac{1}{2}}\tilde{P}'X$, $\rho(A) = \hat{\xi}(A, 1_p)$ for all $A: p \times k$ and O is any orthogonal matrix. Equivariance implies $\rho(A) = \rho(A)Q$ for all reflections in the row space of A so $\rho(A)$ must also be in that space. Decompose A uniquely as $A = B[A]T^{\frac{1}{2}}[A]C[A]$, $T[A]$ being diagonal with nonnegative, decreasing diagonal elements, $B[A]$ orthogonal with nonnegative first row and $C[A]$ defined as $T^{-\frac{1}{2}}[A]B'[A]A$, so that $C[A]C'[A] = 1_p$. Choose $D[A]$ so that $(C'[A], D'[A]): k \times k$ is orthogonal. Then equivariance implies $\rho(A) = B[A]\rho\{(T^{\frac{1}{2}}[A], 0)\}(C'[A], D'[A])'$ for all A. Since $\rho(A)$ is in the row space of A, for all A, with an abuse of notation, $\rho\{(T^{\frac{1}{2}}[A], 0)\} = (\rho\{T^{\frac{1}{2}}[A]\}, 0)$. Thus $\rho(A) = B[A]\rho\{T^{\frac{1}{2}}[A]\}T^{-\frac{1}{2}}[A]B'[A]A$. With R, $\tilde{B}$, T, etc., as defined above, i.e., $\tilde{B} = B[R]$, $T = T[R]$, and with $\psi(T) = \rho\{T^{\frac{1}{2}}\}T^{-\frac{1}{2}}$, $\rho(\mathrm{O}'R) = B[\mathrm{O}'R]\psi(T)\, B'[\mathrm{O}'R]\mathrm{O}'R = \mathrm{O}'\tilde{B}\Delta\psi(T)\Delta\tilde{B}'R$ since $B[\mathrm{O}'R] = \mathrm{O}'\tilde{B}\Delta$ for $\Delta = \operatorname{diag}\{\pm 1, \cdots, \pm 1\}$ appropriately chosen to make $\mathrm{O}\tilde{B}$'s first row nonnegative. It follows that $\hat{\xi} = \tilde{P}V^{\frac{1}{2}}\tilde{B}\Delta\psi(T)\Delta\tilde{B}'V^{-\frac{1}{2}}\tilde{P}'X$ for all $\Delta = \operatorname{diag}\{\pm 1, \cdots, \pm 1\}$ since O was arbitrary. Thus $\Delta\psi(T)\Delta = \psi(T)$ for all such Δ and hence $\psi(T)$ is diagonal. The asserted equivariant representation follows.

Because throughout this section only equivariant estimators are considered, it may be assumed without loss of generality that $\Sigma = 1_p$ and the $\xi = \xi_0$ where $\xi_0\xi_0' = \operatorname{diag}\{\lambda_1, \cdots, \lambda_p\}$ and $\lambda_1 \geqq \cdots \geqq \lambda_p \geqq 0$. For any given ξ, Σ, the λ_i are, in the problem reduced by invariance, the roots of the determinantal equations

$$|\xi\xi' - \lambda\Sigma| = 0 \qquad |\xi_0\xi_0' - \lambda I| = 0 .$$

Let us turn to the problem of deriving an unbiased estimator of $E_{\xi,\Sigma} \operatorname{tr} \Sigma^{-1}(\hat{\xi} - \xi)(\hat{\xi} - \xi)'$ where $\hat{\xi}$ is equivariant. To this end consider

$$\Delta R(\hat{\xi}, \xi_0) = E_{\xi_0} \operatorname{tr} (X - \xi_0)(X - \xi_0)' - E_{\xi_0} \operatorname{tr} (\hat{\xi} - \xi_0)(\hat{\xi} - \xi_0)' . \tag{2.1.1}$$

It is readily shown that

$$(2.1.2)\qquad \Delta R(\hat{\xi}, \xi_0) = -E_{\xi_0} \operatorname{tr} WT\phi(T)[\phi(T) - 21_p] \\ -2E_{\xi_0} \operatorname{tr} \xi_0'\tilde{P}V^{\frac{1}{2}}\tilde{B}\phi(T)\tilde{B}'V^{-\frac{1}{2}}\tilde{P}'X$$

where $W = \tilde{B}'V\tilde{B}$. Consider the first term on the right of equation (2.1.2) with $r(W, T) = \operatorname{tr} WT\phi(T)[\phi(T) - 21_p]$. We find

$$E_{\xi_0} r(W, T) = \exp(-\tfrac{1}{2} \operatorname{tr} \Lambda) Er(W, T) \exp(\operatorname{tr} \xi_0'X)$$

where "E" is the expectation computed with $\xi = 0$, $\Sigma = 1_p$. In this case, the joint distribution of X and $S + XX'$ is invariant under the transformation $(X, S + XX') \to (H_p X H_k, H_p[S + XX']H_p')$ where H_p and H_k are, respectively, arbitrary $p \times p$ and $k \times k$ orthogonal matrices. Also (W, T) is invariant under these transformations. As shown below it follows from (1.6) and (1.7) that

$$(2.1.3)\qquad E_{\xi_0} r(W, T) = \textstyle\sum_\kappa \pi_{\frac{1}{2}\Lambda}(\kappa) E^{\tilde{\kappa}=\kappa} \operatorname{tr} WT\phi(T)[\phi(T) - 21_p]\,,$$

where $\pi_{\frac{1}{2}\Lambda}(\kappa)$ is the multivariate Poisson law defined in equation (1.5) and

$$E^{\tilde{\kappa}=\kappa} r(W, T) = [Er(W, T)C_\kappa(XX')][EC_\kappa(XX')]^{-1}\,.$$

To derive equation (2.1.3) change variables $X \to XH_k'$ in the above expression for $E_{\xi_0} r(W, T)$ (W, T remaining invariant under this change as does the distribution of X). This gives

$$(2.1.4)\qquad E_{\xi_0} r(W, T) = \exp(-\tfrac{1}{2} \operatorname{tr} \Lambda) Er(W, T) \exp(\operatorname{tr} \xi_0'XH_k)\,.$$

Since this identity holds for every H_k, it will hold for $H_k = \tilde{H}_k$ randomly distributed independent of X on $O(k)$ by the invariant measure on $O(k)$ (normalized to have total mass 1). By applying (1.6) to $E^X \exp(\operatorname{tr} \xi_0'X\tilde{H}_k)$, (2.1.4) becomes

$$(2.1.5)\qquad E_{\xi_0} r(W, T) = \textstyle\sum_\kappa \exp(-\tfrac{1}{2} \operatorname{tr} \Lambda) \alpha_\kappa^{(k)} Er(W, T) C_\kappa(\xi_0'XX'\xi_0)\,.$$

Since $C_\kappa(\xi_0'XX'\xi_0) = C_\kappa(\xi_0\xi_0'XX')$, the change of variables $X \to \tilde{H}_p X$, $\tilde{H}_p$ being randomly distributed, independently of X on $O(p)$, yields by equation (1.7) (with $\Lambda = \xi_0'\xi_0$)

$$E_{\xi_0} r(W, T) = \textstyle\sum_\kappa \exp(-\tfrac{1}{2} \operatorname{tr} \Lambda) C_\kappa(\Lambda) \alpha_\kappa^{(k)} / C_\kappa(1_p) \cdot Er(W, T) C_\kappa(XX')$$

i.e., by equation (1.8), equation (2.1.3).

The last term on the right of equation (2.1.2) is, with $\tilde{\phi} = \tilde{P}V^{\frac{1}{2}}\tilde{B}\phi(T)\tilde{B}'V^{-\frac{1}{2}}\tilde{P}'$,

$$(2.1.6)\qquad -2E_{\xi_0} \operatorname{tr} \xi_0'\tilde{\phi}X = -2 \exp(-\tfrac{1}{2} \operatorname{tr} \Lambda) \frac{d}{d\varepsilon} E \exp(\operatorname{tr} \xi_0'\tilde{\phi}_\varepsilon X)\Big|_{\varepsilon=0}$$

where $\tilde{\phi}_\varepsilon = 1_p + \varepsilon\tilde{\phi}$ provided that the expectation in (2.1.6) exists for sufficiently small $|\varepsilon| > 0$. This will be the case if ϕ's diagonal entries are bounded by M, say. For then, with $\tilde{A} = \tilde{P}V^{\frac{1}{2}}\tilde{B}$, $\operatorname{tr} \xi_0'\tilde{\phi}_\varepsilon X = \operatorname{tr} \xi_0'X + \varepsilon \operatorname{tr} X\xi_0'\tilde{A}\phi\tilde{A}^{-1}$ and if $\varepsilon > 0$, $\varepsilon \operatorname{tr} X\xi_0'\tilde{A}\phi\tilde{A}^{-1} \leqq \frac{1}{2}\varepsilon M[\operatorname{tr}(S + XX')^{-1}XX' + \operatorname{tr}(S + XX')\xi_0\xi_0']$. This inequality is proved using twice the fact that if B and $D - C$ are nonnegative definite matrices of the same dimension, then $\operatorname{tr} BD \geqq \operatorname{tr} BC$. The inequality implies the existence of the expectation in (2.1.6) when $\varepsilon > 0$ is sufficiently small ($I_p - \frac{1}{2}\varepsilon M\xi_0\xi_0'$ must be positive definite). If ϕ's diagonal entries are not bounded, they may be truncated at $M > 0$, say, and the subsequent analysis to equation (2.1.8) below carried out for the resulting estimator with $\tilde{\phi}$ replaced by $\tilde{\phi}^{(M)}$, say. By letting $M \to \infty$, then equation (2.1.8) will be obtained from its version where ϕ is replaced by its truncated counterpart.

Under the transformation

$$(X, S + XX') \to (H_p X H_k, H_p[S + XX']H_p'),$$

$\tilde{B}$, T and V remain invariant and $\tilde{P} \to H_p\tilde{P}$. Thus,

$$\begin{aligned}
&\exp(-\operatorname{tr} \tfrac{1}{2}\Lambda)E \exp(\operatorname{tr} \xi_0'\tilde{\phi}_\varepsilon X) \\
&\quad = \textstyle\sum_\kappa \pi_{\frac{1}{2}\Lambda}(\kappa)\{EC_\kappa(XX')\}^{-1} \\
&\qquad \times \{EC_\kappa([1_p + \varepsilon\tilde{P}V^{\frac{1}{2}}\tilde{B}\phi(T)\tilde{B}'V^{-\frac{1}{2}}\tilde{P}']X \circ X'[1_p + \varepsilon\tilde{P}V^{-\frac{1}{2}}\tilde{B}\phi(T)\tilde{B}'V^{\frac{1}{2}}\tilde{P}'])\} \\
&\quad = \textstyle\sum_\kappa \pi_{\frac{1}{2}\Lambda}(\kappa)\{EC_\kappa(XX')\}^{-1}\{EC_\kappa(\tilde{B}'V\tilde{B}[1_p + \varepsilon\phi(T)]^2T)\}.
\end{aligned}$$

Under E, $S + XX'$ and $(S + XX')^{-\frac{1}{2}}X = V^{-\frac{1}{2}}\tilde{P}'X = U$, say, are independent (see, for example, Shorrock and Zidek (1974)); it follows that V is independent of $\tilde{B}$. Also $\tilde{B}$ is independent of T. This is because the density of U is, disregarding normalizing constants, $|1_p - uu'|^{(n-p-1)/2}$ with respect to Lebesgue measure on the space $\{u: 0 \leqq uu' \leqq 1_p\}$. Since this density is spherically symmetric, an argument of James (1954) for the normal law is readily adapted to yield the conclusion. In addition, since the expression above is invariant under the transformation $\tilde{B} \to \Delta\tilde{B}$ with $\Delta = \operatorname{diag}\{\pm 1, \cdots, \pm 1\}$ we may drop the restriction that $\tilde{B}$ have nonnegative first row and conclude by appealing again to James (1954), that $\tilde{B}$ is uniformly distributed by the invariant probability measure on the group of $p \times p$ orthogonal matrices. Thus

$$EC_\kappa(\tilde{B}'V\tilde{B}[1_p + \varepsilon\phi(T)]^2T) = cEC_\kappa(V)C_\kappa([1_p + \varepsilon\phi(T)]^2T)C_\kappa^{-1}(1_p) \tag{2.1.7}$$

(see equation (1.7)). It follows that

$$\begin{aligned}
&\exp(-\tfrac{1}{2}\operatorname{tr}\Lambda)\frac{d}{d\varepsilon} E\exp(\operatorname{tr}\xi_0'\tilde{\phi}_\varepsilon X)\Big|_{\varepsilon=0} \\
&\qquad = 2c \textstyle\sum_\kappa \pi_{\frac{1}{2}\Lambda}(\kappa)\{EC_\kappa(XX')C_\kappa(1_p)\}^{-1}\{EC_\kappa(V)\operatorname{tr}[T\phi(T)\nabla^* C_\kappa(T)]\} \\
&\qquad = \textstyle\sum_\kappa \pi_{\frac{1}{2}\Lambda}(\kappa)\{EC_\kappa(XX')\}^{-1}\{EC_\kappa(\tilde{B}'V\tilde{B}T)\operatorname{tr} T\phi(T)\theta_\kappa(T)\},
\end{aligned}$$

where $\nabla^* C_\kappa(t) = \operatorname{diag}\{(\partial/\partial t_1)C_\kappa(t), \cdots, (\partial/\partial t_p)C_\kappa(t)\}$ and

$$\theta_\kappa(t) = 2\nabla^* \ln C_\kappa(t) = 2\operatorname{diag}\left\{\frac{\partial}{\partial t_1}\ln C_\kappa(t), \cdots, \frac{\partial}{\partial t_p}\ln C_\kappa(t)\right\}.$$

But $C_\kappa(\tilde{B}'V\tilde{B}T) = C_\kappa(XX')$. Thus

$$-2E_{\xi_0}\operatorname{tr}\xi_0'\tilde{\phi}X = -2\textstyle\sum_\kappa \pi_{\frac{1}{2}\Lambda}(\kappa)E^{\tilde{\kappa}=\kappa}\operatorname{tr} T\phi(T)\theta_{\tilde{\kappa}}(T)$$

and hence

$$\Delta R(\hat{\xi}, \xi_0) = -E^*_{\xi_0}\operatorname{tr}\{W\phi^2(T)T - 2W\phi(T)T + 2T\phi(T)\theta_{\tilde{\kappa}}(T)\} \tag{2.1.8}$$

where $E^*_{\xi_0}$ denotes expectation with respect to the joint distribution of W, T, and $\tilde{\kappa}$.

An unbiased risk estimator is readily derived from equation (2.1.8) by integration by parts. For this purpose an explicit formula for the density of T is required. Since T is the diagonal matrix of UU' and the density element of U (when $\xi = 0$, $\Sigma = I$) is, disregarding normalizing constants, $|1_p - uu'|^{(n-p-1)/2}\, du$, it follows from the argument used by James (1954) to derive the density of the latent roots of the Wishart matrix that the joint density of the diagonal elements of T, i.e., $T_1 > \cdots > T_p$, is, apart from normalizing constants

$$\prod (1 - t_i)^{(n-p-1)/2} \prod t_i^{(k-p-1)/2} \prod_{i<j} (t_i - t_j) \prod dt_i \, . \tag{2.1.9}$$

Observe that

$$E^*_{\xi_0} \operatorname{tr} T\phi(T)\theta_{\tilde{\kappa}}(T) = \textstyle\sum_\kappa \pi_{\frac{1}{2}\Lambda}(\kappa) E^{\tilde{\kappa}=\kappa} \operatorname{tr} T\phi(T)\theta_{\tilde{\kappa}}(T)$$

and that

$$E^{\tilde{\kappa}=\kappa} \operatorname{tr} T\phi(T)\theta_{\tilde{\kappa}}(T) = \{EC_\kappa(\tilde{B}'V\tilde{B}T) \operatorname{tr} T\phi(T)\theta_\kappa(T)\}\{EC_\kappa(XX')\}^{-1} \, . \tag{2.1.10}$$

To obtain an unbiased risk estimator, the numerator of the right-hand side of equation (2.1.10) will be evaluated using integration by parts and the density given in (2.1.9). The numerator is the product of

$$\{C_\kappa(1_p)\}^{-1}\{EC_\kappa(V)\} \qquad \text{and} \qquad \left\{\sum E2T_i\phi_i(T) \frac{\partial}{\partial T_i} C_\kappa(T)\right\} .$$

Since, assuming ϕ_i is differentiable and that the resulting integrals exist,

$$\begin{aligned} E2T_i\phi_i(T) \frac{\partial}{\partial T_i} C_\kappa(T) = ET_i C_\kappa(T)[&-2\phi_{ii}(T) - \phi_i(T)\{(k-p+1)T_i^{-1} \\ &+ 2\textstyle\sum_{j\neq i} (T_i - T_j)^{-1} - (n-p-1)(1-T_i)^{-1}\}] , \end{aligned}$$

returning along the path which led to this last expression, it is found that

$$\begin{aligned} &E^*_{\xi_0} \operatorname{tr} T\phi(T)\theta_{\tilde{\kappa}}(T) \\ &\quad = \textstyle\sum E_{\xi_0} T_i[-2\phi_{ii}(T) - \phi_i(T)\{(k-p+1)T_i^{-1} \\ &\qquad + 2\textstyle\sum_{j\neq i} (T_i - T_j)^{-1} - (n-p-1)(1-T_i)^{-1}\}] . \end{aligned} \tag{2.1.11}$$

Combining equations (2.1.8) and (2.1.11) gives

$$\begin{aligned} \Delta R(\hat{\xi}, \xi_0) = E_{\xi_0}\{&\operatorname{tr}\{-W\phi^2(T)T + 2W\phi(T)T\} \\ &+ \textstyle\sum T_i[4\phi_{ii}(T) + 2\phi_i(T)\{(k-p+1)T_i^{-1} \\ &+ 2\textstyle\sum_{j\neq i} (T_i - T_j)^{-1} - (n-p-1)(1-T_i)^{-1}\}]\} . \end{aligned} \tag{2.1.12}$$

Equation (2.1.12) yields an unbiased risk estimator of the risk of any equivariant estimator for which ϕ is differentiable.

Efron and Morris (1972) obtain a minimax estimator for this problem; in their estimator,

$$\phi_i(T) = c\left(\frac{1}{T_i} - 1\right) , \tag{2.1.13}$$

where $c = (k - p - 1)(n + p + 1)^{-1}$. It is easily verified that for $\varepsilon > 0$, the expectation in equation (2.1.6) exists for this choice. So equation (2.1.12) implies, for this choice of ϕ, that

$$\begin{aligned} \Delta R(\hat{\xi}, \xi_0) = n[&-(n + p + 1)c^2 - 4c \\ &+ 2c(k - p + 1)E_{\xi_0} \operatorname{tr} (XX')^{-1}] . \end{aligned} \tag{2.1.14}$$

The best choice for c is easily seen to be that chosen by Efron and Morris and for that choice of c we obtain the result they do, viz:

$$\Delta R(\hat{\xi}, \xi_0) = n(n + p + 1)^{-1}(k - p - 1)^2 E_{\xi_0} \operatorname{tr} (XX')^{-1} . \tag{2.1.15}$$

DISCUSSION. Were it not for the restriction of our considerations to equivariant estimators, an unbiased risk estimator might well have been derived rather more straightforwardly by an integration by parts, using, say, Green's formula. Equivariance is imposed to insure "pooling" across rows and across columns (or across replications and across equations in the regression example of Section 3), and also to reduce the difficulty of the search for a superior alternative to the maximum likelihood estimator by reducing the size of the class of contenders. Although the risk estimator would be simpler to derive without this restriction, this estimator would not so readily have suggested the sought after alternative.

2.2. $\Sigma = \sigma^2 1_p$. In this case assume $S \sim \sigma^2\chi_n^2$ is observable independently of $X \sim N(\xi, \sigma^2 1_p \otimes 1_k)$ and consider the class of estimators, $\hat{\xi}$, of ξ equivariant under the transformation group which acts as follows:

$$\begin{aligned} X &\to cPXQ , & S &\to c^2 S , \\ \xi &\to cP\xi Q , & \sigma^2 &\to c^2\sigma^2 , \end{aligned}$$

where P and Q are, respectively, $p \times p$ and $k \times k$ orthogonal matrices and $c > 0$. It is readily shown that $\hat{\xi}$ must have the form

$$\hat{\xi}(X, S) = [1_p - B\phi(LS^{-1})B']X \tag{2.2.1}$$

for some $p \times p$ diagonal matrix ϕ where B is the unique $p \times p$ orthogonal matrix with nonnegative first row for which

$$L = B'XX'B ,$$

and $L = \operatorname{diag} \{L_1, \cdots, L_p\}$, $L_1 > \cdots > L_p > 0$. Furthermore,

$$\begin{aligned} &\sigma^{-2}\{E_{\xi,\sigma^2}\|X - \xi\|^2 - E_{\xi,\sigma^2}\|\hat{\xi} - \xi\|^2\} \\ &\quad = E_\lambda{}^* \operatorname{tr} L[2\phi(LS^{-1}) - \phi^2(LS^{-1}) - 4\phi(LS^{-1})\nabla^* \ln C_{\tilde{\kappa}}(L)] \\ &\quad = \nabla R(\hat{\xi}, \lambda) , \end{aligned} \tag{2.2.2}$$

say, where $\lambda = \xi\sigma^{-2}$ and $E_\lambda{}^*$ denotes expectation with respect to the joint distribution of L, S and $\tilde{\kappa}$. The unbiased risk estimator derivable from (2.2.2) is determined by equations (2.2.3) or (2.2.4):

$$\begin{aligned} \Delta R(\hat{\xi}, \lambda) = E_\lambda\{[\textstyle\sum L_i\{&-\phi_i^2(LS^{-1}) + 4S^{-1}\phi_{ii}(LS^{-1}) \\ &+ 2(k - p + 1)L_i^{-1}\phi_i(LS^{-1})\}] \\ &+ 4 \textstyle\sum_{i<j} [L_i\phi_i(LS^{-1}) - L_j\phi_j(LS^{-1})][L_i - L_j]^{-1}\} , \end{aligned} \tag{2.2.3}$$

or equivalently,

$$\Delta R(\hat{\xi}, \lambda) = E_\lambda\{\sum L_i\{-\phi_i^2(LS^{-1}) + 4S^{-1}\phi_{ii}(LS^{-1}) + 2[(k-p+1)L_i^{-1} + 2\sum_{j\neq i}(L_i - L_j)^{-1}]\phi_i(LS^{-1})\}\}\,. \tag{2.2.4}$$

If in equation (2.2.2), ϕ is replaced by $\phi = 1_p - \psi$, then we obtain, disregarding terms which do not involve ψ,

$$\Delta R(\hat{\xi}, \lambda) = -E_\lambda \operatorname{tr} L[\psi(LS^{-1}) - \theta_{\tilde{\kappa}}(L)]^2 + \cdots\,.$$

This suggests the unbiased "predictor," ψ^*, of $\theta_{\tilde{\kappa}}(L)$ given by

$$\psi_i^*(LS^{-1}) = 1 - n^{-1}S[(k-p-1)L_i^{-1} + 2\sum_{j\neq i}(L_i - L_j)^{-1}]\,,$$

since, as an integration by parts will show,

$$E^{\tilde{\kappa}=\kappa}\theta_{\tilde{\kappa}}(L) = E^{\tilde{\kappa}=\kappa}\psi^*(LS^{-1})\,.$$

This is easily improved upon by replacing n^{-1} by $(n+2)^{-1}$. For, among estimators with ψ_i of the form

$$\psi_i(LS^{-1}) = 1 - cS[(k-p-1)L_i^{-1} + 2\sum_{j\neq i}(L_i - L_j)^{-1}]\,,$$

equation (2.2.4) implies the best one has

$$c = c_\lambda = \{n+2\}^{-1}\{1 - G(\lambda)\}$$

where $G > 0$ is given by

$$G(\lambda) = \frac{4E_\lambda \sum L_i \sum_{j\neq i}(L_i - L_j)^{-2}}{E_\lambda \sum L_i\{(k-p-1)^2 L_i^{-2} + 4\sum_{j\neq i}(L_i - L_j)^{-2}\}}\,.$$

Since $c_\lambda \leqq (n+2)^{-1} \leqq n^{-1}$ whatever be λ, and since the functional of c being minimized is quadratic in c, it follows that the choice $c = (n+2)^{-1}$ makes this functional uniformly smaller in λ than $c = n^{-1}$. It is not known whether the resulting estimator dominates the usual one. It is the counterpart for σ unknown of that proposed by Stein (1973b) for the case $\sigma = 1$.

3. Multivariate regression. Suppose $X_1, X_2, \cdots, X_n$ are independent $(q+p)$ dimensional random variables, $X_i \sim N_{q+p}(\theta, \Sigma)$. With

$$X_i = \begin{pmatrix} Y_i \\ Z_i \end{pmatrix}, \qquad \theta = \begin{pmatrix} \eta \\ \zeta \end{pmatrix} \qquad \text{and} \qquad \Sigma = \begin{pmatrix} \Gamma_Y & B' \\ B & \Gamma_Z \end{pmatrix},$$

Y_i, η being $q \times 1$, Z_i, ζ, $p \times 1$, and B, $p \times q$,

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

where

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

The parameters α, β are to be estimated so that $\hat{Y} = \hat{\alpha} + \hat{\beta}'Z_0$ will be an effective predictor of a future Y_0, more precisely, so that $\hat{Y}$ is uniformly better than the least squares estimator $\hat{Y}_0 = \hat{\alpha}_0 + \hat{\beta}_0'Z_0$ with respect to the risk function

$$\begin{aligned} &E[Y_0 - \hat{Y}]'\Gamma_{Y|Z}^{-1}[Y_0 - \hat{Y}] \\ &\quad = E\{\operatorname{tr} \Gamma_{Y|Z}^{-1}[(\alpha - \hat{\alpha}) + (\beta - \hat{\beta})'\zeta][(\alpha - \hat{\alpha}) + (\beta - \hat{\beta})'\zeta]' \\ &\qquad + \operatorname{tr} \Gamma_{Y|Z}^{-1}[\beta - \hat{\beta}]'\Gamma_Z[\beta - \hat{\beta}]\} \end{aligned} \tag{3.1}$$

where $\Gamma_{Y|Z}$ denotes the conditional covariance of Y_i given Z_i, all i. This suggests a natural loss function for the (α, β)-estimation problem, namely, the expression in equation (3.1) contained within the curly brackets, { }. Stein (1960) proposed this loss for the case $q = 1$.

An intuitively equivalent but mathematically simpler version is chosen here, namely,

$$L((\theta, \Sigma); (\hat{\alpha}, \hat{\beta})) = \operatorname{tr} \Gamma_{Y|Z}^{-1}\{[(\alpha - \hat{\alpha}) + (\beta - \hat{\beta})'\zeta][(\alpha - \hat{\alpha}) + (\beta - \hat{\beta})'\zeta]' + [\beta - \hat{\beta}]'\hat{\Gamma}_Z[\beta - \hat{\beta}]\}$$

where $\hat{\Gamma}_Z = (n-1)^{-1} \sum (Z_i - \bar{Z})(Z_i - \bar{Z})'$. Like Baranchik we restrict ourselves to the problem of estimating β under the loss obtained by setting $\zeta = 0$ (as we may do without loss of generality since only equivariant estimators are to be considered) and $\alpha = \hat{\alpha}$. That is, the loss function is

$$L((\theta, \Sigma); \hat{\beta}) = \operatorname{tr} \Gamma_{Y|Z}^{-1}[\beta - \hat{\beta}]'\hat{\Gamma}_Z[\beta - \hat{\beta}] . \tag{3.2}$$

Although this puts the problem into a form comparable to that treated by Baranchik (1973) we cannot assert as he can that if $\hat{\beta}$ dominates $\hat{\beta}_0$ with respect to the loss (3.2) then $(\bar{Y} - \hat{\beta}'\bar{Z}, \hat{\beta})$ dominates $(\hat{\alpha}_0, \hat{\beta}_0)$ with respect to the joint loss presented above for $(\hat{\alpha}, \hat{\beta})$.

The loss function in equation (3.2) is essentially that proposed by Dempster (1973) although his argument for it is different than that given here [predictions are conceived as being made at the Z_i's already observed]. Thisted (1976, page 18) comments on this loss function and cites other work where it has been introduced.

Since

$$\hat{\beta}_0 = (n-1)^{-1}\hat{\Gamma}_Z^{-1} \sum (Z_i - \bar{Z})(Y_i - \bar{Y})'$$
$$\hat{\alpha}_0 = \bar{Y} - \hat{\beta}_0'\bar{Z}$$

are based on the minimal sufficient statistic it is natural to restrict the search for a better alternative to the class of estimators based on this statistic. Furthermore, following the Stein (1960) argument, we consider only equivariant estimators with respect to the transformations under which

$$\binom{Y_i}{Z_i} \to \begin{pmatrix} A & 0 \\ 0 & C \end{pmatrix}\binom{Y_i}{Z_i} + \binom{e}{d} , \qquad i = 1, \cdots, n \tag{3.3}$$

where $A: q \times q$ and $C: p \times p$ are nonsingular and $e: q \times 1$ and $d: p \times 1$ are arbitrary.

If

$$V = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix} = \sum_i (X_i - \bar{X})(X_i - \bar{X})'$$

and

$$\hat{\theta} = \binom{\bar{Y}}{\bar{Z}} ,$$

then under the transformation (3.3),

$$V \to \begin{pmatrix} A & 0 \\ 0 & C \end{pmatrix} V \begin{pmatrix} A' & 0 \\ 0 & C' \end{pmatrix}$$
$$\hat{\theta} \to \begin{pmatrix} A & 0 \\ 0 & C \end{pmatrix}\hat{\theta} + \binom{e}{d}$$

with corresponding changes in the true parameters Γ_X and θ. Thus

$$\beta \to (C^{-1})'\beta A' , \qquad \alpha \to A\alpha - A\beta' C^{-1} d + e \tag{3.4}$$

and to require equivariance is to require (3.4) to hold for $\hat\beta$ and $\hat\alpha$ as well. Since $\Gamma_{Y|Z} \to A\Gamma_{Y|Z}A'$, the loss, (3.2), remains invariant. It follows that the risk function of equivariant estimators depends on the true parameter through the maximal parameter invariants, i.e., through the true canonical correlation coefficients. In other words, if, as will be assumed hereafter, $p > q$ it may be assumed without loss of generality that

$$\Gamma_Y = 1_q , \qquad \Gamma_Z = 1_p , \qquad B' = [\Lambda, 0] , \qquad \eta = 0 , \qquad \zeta = 0 \tag{3.5}$$

where $\Lambda : q \times q = \operatorname{diag}\{\rho_1, \cdots, \rho_q\}$ and $\rho_1 \geqq \cdots \geqq \rho_q$ are the canonical correlation coefficients. Then $\Gamma_{Y|Z} = 1_q - \Lambda^2$, $\hat\beta$ depends on the data only through V, and $\hat\alpha = \bar Y - \hat\beta' \bar Z$. Furthermore as is easily shown, equivariance implies (with symmetric square roots)

$$\hat\beta = (V_{22}^{-\frac{1}{2}})Pf(R)Q(V_{11}^{\frac{1}{2}}) \tag{3.6}$$

where $R = \operatorname{diag}\{r_1, \cdots, r_q\}$, $r_1 > \cdots > r_q$ denote the sample canonical correlations, $Q : q \times q$ is orthogonal and given by

$$(V_{11}^{-\frac{1}{2}})V_{12}(V_{22}^{-1})V_{21}(V_{11}^{-\frac{1}{2}}) = Q'R^2Q$$

and

$$P = (V_{22}^{-\frac{1}{2}})V_{21}(V_{11}^{-\frac{1}{2}})Q'R^{-1} ,$$

P being $p \times q$ so that $P'P = 1_q$. The quantity $f(R) : q \times q$ is an arbitrary diagonal matrix function of R. The least squares estimator, $\hat\beta_0$, is obtained from the choice

$$f(R) = f_0(R) = R .$$

It is readily shown that if $\hat\beta$ is equivariant,

$$\begin{aligned} E_{\theta,\Sigma} \operatorname{tr} \Gamma_{Y|Z}^{-1}[\hat\beta - \beta]'\hat\Gamma_Z[\hat\beta - \beta] \\ = E_\Lambda \operatorname{tr} [\hat\beta - \beta]'\hat\Gamma_Z[\hat\beta - \beta] = \rho(\hat\beta, \beta) , \text{ say,} \end{aligned} \tag{3.7}$$

where in E_Λ, $Y \mid Z \sim N(\beta' Z, 1_q \otimes 1_{n-1})$ and $Z \sim N(0, [1 - \beta\beta']^{-1} \otimes 1_{n-1})$. Further an argument similar to that of Section 2 shows that

$$\rho(\hat\beta, \beta) = (n-1)^{-1} E_\Lambda \operatorname{tr} Uf^2(R) - 2(n-1)^{-1} E_\Lambda \operatorname{tr} Rf(R)\theta_{\tilde\kappa}(R^2) + \operatorname{tr} \beta'\beta , \tag{3.8}$$

where $U = Q'V_{11}Q$, and $\theta_{\tilde\kappa}(l) = 2\nabla^* \ln C_{\tilde\kappa}(l)$; $\tilde\kappa = (\tilde\kappa_1, \cdots, \tilde\kappa_p)$ is a random integer partition, $\tilde\kappa_1 \geqq \cdots \geqq \tilde\kappa_p$, having the multivariate negative binomial distribution with probability mass function

$$n_{\beta'\beta}(\kappa) = \left(\frac{n-1}{2}\right)_k (\|\kappa\|!)^{-1} C_\kappa(\beta'\beta) |I - \beta'\beta|^{(n-1)/2} ,$$

for all $\kappa = (\kappa_1, \cdots, \kappa_p)$. Conditionally, given $\tilde\kappa = \kappa$, Y and Z have a density proportional to $C_\kappa(zy'yz') \exp[-\frac{1}{2}\{\operatorname{tr} zz' + \operatorname{tr} yy'\}]$. An integration by parts involving the known form of the density of R when $Z \sim N(0, 1_p \otimes 1_{n-1})$ and

$Y \sim N(0, 1_q \otimes 1_{n-1})$ yields equation (3.9) below from which an unbiased risk estimator emerges:

$$\begin{aligned}\rho(\hat\beta, \beta) = (n-1)^{-1}E_\beta \Big\{ &\operatorname{tr} Uf^2(R) + 2R^{-1}f(R) + 2\nabla^* f(R) \\ &+ 4\textstyle\sum_i R_i f_i(R)\Big[\frac{p-q-1}{2R_i^2} - \frac{n-p-q-2}{2(1-R_i^2)} \\ &+ \textstyle\sum_{j\neq i}(R_i^2 - R_j^2)^{-1}\Big]\Big\} + \operatorname{tr}\beta'\beta\,. \end{aligned} \tag{3.9}$$

As will be shown below, a superior alternative to $\hat\beta_0$ is $\hat\beta_1$ for which

$$f(R) = f_1(R) = (1+c)R - cR^{-1}\,, \qquad 0 < c < \frac{2(p-q-1)}{(n-p+q)}\,. \tag{3.10}$$

This choice is the natural extension of the estimator proposed by Stein (1960) for the case $q = 1$. Baranchik (1973) proved Stein's conjecture for the loss in (3.2) when $\hat\Gamma_Z$ is replaced by Γ_Z and $q = 1$. For the loss in (3.2) the estimator $\hat\beta_1$ given by the choice $f = f_1$, dominates $\hat\beta_0$ if $n \geqq p + 1$ and $p \geqq q + 1$. In fact, with ρ the risk function given in (3.9), it can be shown using Cochran's theorem that

$$\begin{aligned}&\rho(\hat\beta_1, \beta) - \rho(\hat\beta_0, \beta) \\ &\quad = c(n-1)^{-1}\{c - 2(p-q-1)(n-p+q)^{-1}\} \\ &\quad\quad \times (n-p-1)(n-p+q)E_\Lambda \operatorname{tr}(1_q - \Lambda^2)(V_{12}V_{22}^{-1}V_{21})^{-1}\end{aligned} \tag{3.11}$$

where, it will be recalled, $B' = [\Lambda, 0]$, $\Gamma_Z = 1_p$, $\Gamma_Y = 1_q$, and $\beta = \Gamma_Z{}^{-1}B = [\Lambda, 0]'$, all without loss of generality because of the assumed equivariance. The asserted conclusion follows since under the hypothesis $\rho(\hat\beta_1, 0) - \rho(\hat\beta_0, \beta) < 0$ for all values of the true parameters.

This result is not, strictly speaking, an extension of Baranchik's result (1973) to the case $q \geqq 1$, because the loss function differs from that of Baranchik. It should be noted in comparing the results that Baranchik assumes $n + 1$ independent X_i's are to be observed, not n as is asserted on page 312 of his article.

DISCUSSION. For the loss function considered here and the case $q = 1$, the superiority of the estimators determined by (3.10), i.e., the estimators considered by Baranchik, are very simply established by conditioning on Z, suitably transforming $\hat\beta_0$, and appealing directly to the James-Stein result (1960). This analysis is carried out by Sclove (1968). Likewise there is a James–Stein estimator for the case $q > 1$, which we have not evaluated.

The estimator determined by equation (3.10), i.e., the one suggested by the Stein (1960)–Baranchik (1973) works, is also the one determined by equation (2.1.13), i.e., by the Efron–Morris (1972) work, when for the case of fixed Z, the problem is reduced to its canonical form, the form considered by Efron and Morris (1972). In fact, an alternate argument which yields equation (3.9) directly from equation (2.1.12) may be constructed as follows. After restricting $\hat\beta$ to be equivariant, and then setting $\Gamma_Z = 1_p$, condition on Z. The resulting problem can then be reduced to that of Section 2 and equation (3.9) obtained directly [substitute p for k, q for p and note (i) that S now corresponds to the residual sum of squares so instead of n degrees of freedom, S has just $n - p - 1$]. This approach has the shortcoming, however, that it does not give the representation (3.8) which is of interest in its own right.

Our analysis of the problem preserves the algebraic character of the problem, i.e., β is regarded as a $q \times p$ matrix rather than a vector of dimension qp, and it is this which leads to the identification of $f(R)$ given in (3.10). This in turn reveals, in the next paragraph, a role for canonical variates in multivariate regression analysis.

It is clear from equation (3.11) that the best choice of c is given by

$$c = (p - q - 1)(n - p + q)^{-1}.$$

For this choice of c it is readily seen that

$$\hat{\beta}_1 = \hat{\beta}_0[a1_q - b(V_{12}V_{22}^{-1}V_{21})^{-1}V_{11}] \tag{3.12}$$

where $a = (n - 1)(n - p + q)^{-1}$ and $b = (p - q - 1)(n - p + q)^{-1}$. Alternatively this is

$$\hat{\beta}_1 = (\tilde{Z}\tilde{Z}')^{-1}\tilde{Z}[a1_q - bR^{-2})A'\tilde{Y}]'A^{-1} \tag{3.13}$$

where $\tilde{Z} = (Z_1, \cdots, Z_n) - (\bar{Z}, \cdots, \bar{Z})$, $\tilde{Y} = (Y_1, \cdots, Y_n) - (\bar{Y}, \cdots, \bar{Y})$ and $A'\tilde{Y}$ denotes the matrix of estimated canonical Y-variates. In as much as $a - bR_i^{-2}$ is small when R_i is, the effect of premultiplying $A'\tilde{Y}$ by $(a1_q - bR^{-2})$ is effectively to reduce to zero those canonical Y-variates with small canonical correlations. Post multiplying by A^{-1} then returns the analysis to Y's coordinate system. These heuristics would be more nearly correct if $(a1_q - bR^{-2}) = \operatorname{diag}\{(a - bR_i^{-2})\}$ were replaced by $(a1_q - bR^{-2})^+ = \operatorname{diag}\{(a - bR_i^{-2})^+\}$ and, indeed, these heuristics suggest this ought to be done. The rationale for $\hat{\beta}_1$ would seem to be that the estimator, $\hat{\beta}_0$, is improved when "noise" in the form of nearly uncorrelated canonical Y-variates are replaced by zero.

Observe that since $\rho(\hat{\beta}_0, \beta) \equiv pq(n - 1)^{-1}$, equation (3.11) implies when $c = (p - q - 1)(n - p + q)^{-1}$ that

$$(n - 1)\rho(\hat{\beta}_1, \beta) = pq - (p - q - 1)^2(n - p + q)^{-1}(n - p - 1)E_\beta W \tag{3.14}$$

where $W = \operatorname{tr}(V_{11} - V_{12}V_{22}^{-1}V_{21})(V_{12}V_{22}^{-1}V_{21})^{-1}$. An unbiased estimator for the risk of $\hat{\beta}_1$ is apparent from equation (3.14). Furthermore, $100(p - q - 1)^2(n - p + q)^{-1}(n - p - 1)W/(pq)$ estimates the percentage reduction in risk achieved using $\hat{\beta}_1$ instead of $\hat{\beta}_0$.

Acknowledgments. I am indebted to Professor R. H. Farrell for suggestions which led to a substantial improvement in this paper. I am grateful to Mrs. Pam Carriage for her careful preparation of the typed copy of the manuscript.

REFERENCES

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

[2] Baranchik, A. J. (1973). Inadmissibility of maximum likelihood estimators in some multiple regression problems with three or more independent variables. *Ann. Statist.* **1** 312–321.

[3] Dempster, A. P. (1973). Alternatives to least squares in multiple regression. In *Multivariate Statistical Inference* (D. Kabe and R. P. Gupta, eds.) 25–40. North-Holland, Amsterdam.

[4] Efron, B. and Morris, C. (1972). Empirical Bayes on vector observations—An extension of Stein's rule. *Biometrika* **59** 335–347.

[5] Efron, B. and Morris, C. (1976). Multivariate empirical Bayes and estimation of covariance matrices. *Ann. Statist.* **4** 22–32.

[6] Farrell, R. J. (1976). *Techniques of Multivariate Calculation.* Springer-Verlag, New York.

[7] Fujikoshi, Y. (1970). Asymptotic expansions of the distributions of test statistics in multivariate analysis. *J. Sci. Hiroshima Univ. Ser. A-I Math.* **34** 73–144.

[8] James, A. T. (1954). Normal multivariate analysis and the orthogonal group. *Ann. Math. Statist.* **25** 40–75.

[9] James, A. T. (1964). Distributions of matrix variates and latent roots derived from normal samples. *Ann. Math. Statist.* **35** 475–501.

[10] James, A. T. (1968). Calculation of zonal polynomial coefficients by the use of Laplace-Beltrami operator. *Ann. Math. Statist.* **39** 1711–1718.

[11] James, W. and Stein, C. (1960). Estimation with quadratic loss. *Proc. Fourth Berkeley Symp. Math. Statist. Probability* **1** 361–379.

[12] Lilleströl, J. Improved estimates or parameters in complex normal models with application to analysis of several stationary time series. Ph. D. dissertation, Univ. of California, Berkeley.

[13] Pitman, E. F. G. and Robbins, H. E. (1949). Application of the method of mixtures to quadratic forms in normal variates. *Ann. Math. Statist.* **20** 552–560.

[14] Sclove, S. L. (1968). Improved estimators for coefficients in linear regression. *J. Amer. Statist. Assoc.* **63** 596–606.

[15] Shorrock, R. W. and Zidek, J. V. (1976). An improved estimator of the generalized variance. *Ann. Statist.* **4** 629–638.

[16] Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. *Proc. Third Berkeley Symp. Math. Statist. Probability* **1** 197–206. Univ. California Press.

[17] Stein, C. (1960). Multiple regression. In *Contributions to Probability and Statistics.* Stanford Univ. Press.

[18] Stein, C. (1973a). Estimation of the mean of a multivariate normal distribution. *Proc. Prague Symp. Asymptotic Statist.* 345–381.

[19] Stein, C. (1973b). Estimation of the mean of a multivariate normal distribution. *Bull. Inst. Math. Statist.* Abstract # 138–24, 96.

[20] Thisted, R. (1976). Ridge regression, minimax estimation, and empirical Bayes methods. Technical Report 28, Div. of Biostatistics, Stanford Univ.

Department of Mathematics
University of British Columbia
2075 Wesbrook Mall
Vancouver, B.C.
Canada V6T 1W5