AD-767 016

GENERALIZED BAYES MINIMAX ESTIMATORS OF A MULTIPLE REGRESSION COEFFICIENT VECTOR WITH THREE OR MORE PREDICTORS

Pi-Erh Lin, et al

Florida State University

Prepared for:

Office of Naval Research

August 1973

The Florida State University

Department of Statistics

Tallahassee, Florida

GENERALIZED BAYES MINIMAX ESTIMATORS OF A MULTIPLE REGRESSION COEFFICIENT VECTOR WITH THREE OR MORE PREDICTORS

by

Pi-Erh Lin[1] and Erwin P. Bedo[2]

FSU Statistics Report M270
ONR Technical Report No. 80

August, 1973
The Florida State University
Department of Statistics
Tallahassee, Florida

[1]Research supported in part by the Florida State University Grant No. 011205-413 and in part by the National Institute of General Medical Sciences through Training Grant 5T1 GM-913.

[2]Research supported in part by the Army, Navy and Air Force under Office of Naval Research Contract No. N000 14-67-A-0235-0006 and in part by the National Institute of General Medical Sciences through Training Grant 5T1 GM-913. Reproduction in whole or in part is permitted for any purpose of the United States Government.

Unclassified

SECURITY CLASSIFICATION

DOCUMENT CONTROL DATA - R & D

1. Originating Activity The Florida State University Department of Statistics Tallahassee, Florida	2a. Report Security Classification Unclassified 2b. Group

3. Report Title
Generalized Bayes Minimax Estimators of a Multiple Regression Coefficient Vector with Three or More Predictors

4. Author
Pi-Erh Lin and Erwin P. Bodo

6. Report Date August, 1973	7a. Total No. of Pages 22	7b. No. of References 9

8a. Contract of Grant No. N000 14-67-A-0235-0006	9a. Originator's Report Number ONR Technical Report No. 80
b. Project No. c. Task No.	9b. Other Report No(s). FSU Statistics Report M270

10. Availability/Limitation Notices

Releasable without limitations on dissemination

11. Supplementary Notes	12. Sponsoring Military Activity Logistics and Mathematical Statistics Branch Office of Naval Research Washington, D.C. 20360

13. Abstract

Consider a multiple regression problem in which the dependent variable and (3 or more) independent variables have a joint normal distribution with unknown mean vector and covariance matrix. A family of minimax estimators based on the maximum likelihood estimator and the sample multiple correlation is obtained for the regression coefficient vector. It is shown that there are minimax estimators of the same form as the ones mentioned above, which are also generalized Bayes. The problem of estimating both the intercept and regression coefficient vector is also investigated.

14. Key Words
Regression coefficient vector, intercept, multivariate normal distribution, maximum likelihood estimator, prior, generalized Bayes, minimax, invariant, equivariant, conditional expectation.

S U M M A R Y

GENERALIZED BAYES MINIMAX ESTIMATORS OF A MULTIPLE REGRESSION COEFFICIENT VECTOR WITH THREE OR MORE PREDICTIORS

by

Pi-Erh Lin and Erwin P. Bodo

The Florida State University

Consider a multiple regression problem in which the dependent variable and (3 or more) independent variables have a joint normal distribution with unknown mean vector and covariance matrix. A family of minimax estimators based on the maximum likelihood estimator and the sample multiple correlation is obtained for the regression coefficient vector. It is shown that there are minimax estimators of the same form as the ones mentioned above, which are also generalized Bayes. The problem of estimating both the intercept and regression coefficient vector is also investigated.

AMS 1970 Subject Classifications. Primary 62F10; Secondary 62C10.

Some key words and phrases: Regression coefficient vector, intercept, multivariate normal distribution, maximum likelihood estimator, prior, generalized Bayes, minimax, invariant, equivariant, conditional expectation.

GENERALIZED BAYES MINIMAX ESTIMATORS OF A MULTIPLE REGRESSION COEFFICIENT VECTOR WITH THREE OR MORE PREDICTORS

by

Pi-Erh Lin[1] and Erwin P. Bodo[2]

The Florida State University

1. Introduction.

The problem under consideration is that of estimating the regression coefficients when the predictors and the dependent variable have a joint normal distribution with unknown mean vector and covariance matrix. Stein (1960) showed that the maximum likelihood estimator (m.l.e.) of the regression coefficient vector is minimax and inadmissible relative to a reasonable loss function if there are three or more predictors. Baranchik (1973) obtained a class of minimax estimators of the form

$$(1.1) \qquad \hat{\alpha}_c = \bar{X}^{(1)} - \hat{\beta}_c' \bar{X}^{(2)}, \qquad \hat{\beta}_c = [1 - c(1 - R^2)/R^2]\hat{\beta}_0,$$

where $\bar{X}^{(1)}$ and $\bar{X}^{(2)}$ are the sample means for the dependent and independent variables, respectively, $\hat{\beta}_0$ is the m.l.e. of the regression coefficient vector, R is the sample multiple correlation coefficient, and c is a nonnegative real number bounded by $2(p - 2)/(n - p + 2)$ with the number of predictors $p \geq 3$ and sample size $n + 1$. Each of

---

[1]Research supported in part by the Florida State University Grant No. 011205-413 and in part by the National Institute of General Medical Sciences through Training Grant 5T01 GM-913.

[2]Research supported in part by the Army, Navy and Air Force under Office of Naval Research Contract No. N000-67-A-0235-0006 and in part by the National Institute of General Medical Sciences through Training Grant 5T01 GM-913. Reproduction in whole or in part is permitted for any purpose of the United States Government.

these estimators has a lower risk than that of the m.l.e.. In this paper we extend Baranchik's result to a class of minimax estimators of the form

$$(1.2) \qquad \hat{\alpha}_t = \bar{X}^{(1)} - \hat{\beta}_t' \bar{X}^{(2)}, \qquad \hat{\beta}_t = [1 - t(F)/F]\hat{\beta}_0,$$

where $F = R^2/(1 - R^2)$ and $t(F)$ is a nonnegative, nondecreasing function. Thus Baranchik's estimators form a subclass of (1.2). We further exhibit a class of generalized Bayes estimators which are of the form (1.2) and hence minimax.

Section 2 consists of the formulation of the problem. A group of transformations leaving the problem invariant is introduced in Section 3. Sections 4 and 5 are devoted to estimating the regression coefficient vector β alone. The minimaxity of $\hat{\beta}_t$ is established in Theorem 4.4. A family of generalized Bayes minimax estimators is obtained in Theorem 5.1. Section 6 extends the results of Sections 4 and 5 to the case of estimating both the intercept α and the regression coefficient vector β. Some concluding remarks are given in Section 7.

2. Formulation of the Problem.

Let $X_1, \ldots, X_{n+1}$ be independently normally distributed $(p + 1)$-dimensional random vectors $(p \geq 3)$, with common mean vector μ and common covariance matrix Σ, where both μ and Σ are unknown. The X_i, μ, and Σ are partitioned as follows:

$$(2.1)\qquad X_i = \begin{bmatrix} X_i^{(1)} \\ X_i^{(2)} \end{bmatrix}, \qquad \mu = \begin{bmatrix} \mu^{(1)} \\ \mu^{(2)} \end{bmatrix}, \qquad i = 1, \ldots, n+1,$$

and

$$(2.2)\qquad \Sigma = \begin{bmatrix} \sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix},$$

where $X_i^{(1)}$ and $\mu^{(1)}$ are 1×1, $X_i^{(2)}$, $\mu^{(2)}$ and Σ_{21} $(= \Sigma_{12}')$ are $p \times 1$, and Σ_{22} is $p \times p$. The regression of $X_i^{(1)}$ on $X_i^{(2)}$ is given by

$$(2.3)\qquad E(X_i^{(1)} | X_i^{(2)}) = \alpha + \beta' X_i^{(2)},$$

where

$$(2.4)\qquad \beta = \Sigma_{22}^{-1} \Sigma_{21},$$

and

$$(2.5)\qquad \alpha = \mu^{(1)} - \beta'\mu^{(2)}.$$

We are interested in estimating both α and β in one case and β alone in the other with loss functions

$$(2.6)\qquad L[(\hat{\alpha}, \hat{\beta}); (\mu, \Sigma)]$$

$$= \{[(\hat{\alpha} - \alpha) + (\hat{\beta} - \beta)'\mu^{(2)}]^2 + (\hat{\beta} - \beta)'\Sigma_{22}(\hat{\beta} - \beta)\}/\sigma_{11.2}$$

and

$$(2.7)\qquad L[\hat{\beta}; (\mu, \Sigma)] = (\hat{\beta} - \beta)'\Sigma_{22}(\hat{\beta} - \beta)/\sigma_{11.2},$$

respectively, where $\hat{\alpha}$ and $\hat{\beta}$ denote any estimators of α and β, and

(2.8) $$\sigma_{11.2} = \sigma_{11} - \Sigma_{12}\,\Sigma_{22}^{-1}\,\Sigma_{21}.$$

Loss functions (2.6) and (2.7) were first introduced by Stein (1960), and subsequently used by Baranchik (1973). The m.l.e. $(\hat{\alpha}_0, \hat{\beta}_0)$ of (α, β) is given by

(2.9) $$\hat{\beta}_0 = A_{22}^{-1}\,A_{21},$$

and

(2.10) $$\hat{\alpha}_0 = \bar{X}^{(1)} - \hat{\beta}_0'\,\bar{X}^{(2)},$$

where

(2.11) $$\bar{X} = \begin{pmatrix} \bar{X}^{(1)} \\ \bar{X}^{(2)} \end{pmatrix} = \frac{1}{n+1}\sum_{i=1}^{n+1} X_i,$$

and

(2.12) $$A = \begin{pmatrix} a_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} = \sum_{i=1}^{n+1} (X_i - \bar{X})(X_i - \bar{X})'.$$

The partitioning of $\bar{X}$ and A is similar to that of μ and Σ.

3. Transformations Leaving the Problem Invariant.

Stein (1960) showed that the problem defined in Section 2 is invariant under the following group of transformations:

(3.1) $$\begin{pmatrix} X_i^{(1)} \\ X_i^{(2)} \end{pmatrix} \to \begin{pmatrix} a & b' \\ 0 & C \end{pmatrix}\begin{pmatrix} X_i^{(1)} \\ X_i^{(2)} \end{pmatrix} + \begin{pmatrix} d \\ e \end{pmatrix} = \begin{pmatrix} aX_i^{(1)} + b'X_i^{(2)} + d \\ CX_i^{(2)} + e \end{pmatrix},$$

$$(3.2)\qquad \begin{pmatrix} \mu^{(1)} \\ \mu^{(2)} \end{pmatrix} \to \begin{pmatrix} a & b' \\ 0 & C \end{pmatrix} \begin{pmatrix} \mu^{(1)} \\ \mu^{(2)} \end{pmatrix} + \begin{pmatrix} d \\ e \end{pmatrix} = \begin{pmatrix} a\mu^{(1)} + b'\mu^{(2)} + d \\ C\mu^{(2)} + e \end{pmatrix},$$

$$(3.3)\qquad \begin{pmatrix} \sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \to \begin{pmatrix} a & b' \\ 0 & C \end{pmatrix} \begin{pmatrix} \sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \begin{pmatrix} a & 0' \\ b & C' \end{pmatrix}$$

$$= \begin{pmatrix} a^2\sigma_{11} + 2ab'\Sigma_{21} + b'\Sigma_{22}b & (a\Sigma_{12} + b'\Sigma_{22})C' \\ C(\Sigma_{21}a + \Sigma_{22}b) & C\Sigma_{22}C' \end{pmatrix},$$

$$(3.4)\qquad \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} \to \begin{pmatrix} a\hat{\alpha} - ae'C'^{-1}\hat{\beta} + d - b'C^{-1}e \\ aC'^{-1}\hat{\beta} + C'^{-1}b \end{pmatrix},$$

where $a\ (\neq 0)$ and d are real numbers, b and e are $p \times 1$ vectors 0 is a $p \times 1$ vector of zeros, and C is a $p \times p$ nonsingular matrix. As in Baranchik (1973), if we restrict our attention to the subgroup with $b = 0$ then the estimators

$$(3.5)\qquad \hat{\alpha} = \bar{X}^{(1)} - \hat{\beta}'\,\bar{X}^{(2)}, \qquad \hat{\beta} = f(R^2)\hat{\beta}_0$$

are equivariant; that is, they satisfy (3.4) with $b = 0$. Here $f(R^2)$ is a measurable function of the sample multiple correlation coefficient R, where

$$(3.6)\qquad R^2 = A_{12}\,A_{22}^{-1}\,A_{21}/a_{11}.$$

Since every orbit of the subgroup ($b = 0$) passes through the parameter point (μ, Σ), the risk of the equivariant estimators given by (3.5) may be computed at any convenient parameter point on the same orbit as (μ, Σ). If the parameter point is chosen properly, the calculation of the risk is usually simplified. For example, Baranchik (1973) computed the risk of $\hat{\beta}_c$ at the point $(\mu^{(2)}, \Sigma_{22}, \sigma_{11.2}) = (0, I_p, 1)$, where I_p is the identity matrix of order p.

4. A Class of Minimax Estimators of β.

In this and the following section we will consider the estimation of β alone. The following three lemmas will be used in the proof of Theorem 4.4. Lemma 4.1 can be found in Baranchik (1970) and Lemmas 4.2 and 4.3 in Stein (1969, pp. 29-32, 61-62). They are presented here without proofs.

LEMMA 4.1. Let $Z \sim N(\theta, I_p)$. Let $g(T)$ be a measurable function of $T = Z'Z/\chi_n^2$, where χ_n^2 is independent of Z. Then

$$\text{i)}\quad E[Z'Zg^2(T)] = e^{-\frac{1}{2}\|\theta\|^2} \sum_{k=0}^{\infty} \frac{(\frac{1}{2}\|\theta\|^2)^k}{k!} E[\chi_{p+2k}^2\, g^2(\chi_{p+2k}^2/\chi_n^2)] ,$$

and

$$\text{ii)}\quad E[\theta'Zg(T)] = e^{-\frac{1}{2}\|\theta\|^2} \sum_{k=0}^{\infty} \frac{(\frac{1}{2}\|\theta\|^2)^k}{k!} 2k\, E[g(\chi_{p+2k}^2/\chi_n^2)] ,$$

where $\|\theta\|^2 = \theta'\theta$.

LEMMA 4.2. Let $A \sim W(\Sigma, n)$. Partition A and Σ as in (2.12) and (2.2). Then

i) $a_{11.2} \equiv a_{11} - A_{12} A_{22}^{-1} A_{21} \sim \sigma_{11.2} \chi^2_{n-p}$,

ii) $A_{22}^{-1} A_{21} \mid A_{22} \sim N(\beta, \sigma_{11.2} A_{22}^{-1})$,

iii) $A_{22} \sim W(\Sigma_{22}, n)$, and

iv) $a_{11.2}$ is independent of (A_{12}, A_{22}).

LEMMA 4.3. Conditional on A_{22}, the distribution of $F = \hat{\beta}_0' A_{22} \hat{\beta}_0 / a_{11.2}$ is that of $\chi^2_{p+2K}/\chi^2_{n-p}$, where χ^2_{p+2K} and χ^2_{n-p}, given A_{22}, are independent, and K given A_{22} is distributed as Poisson with parameter $\frac{1}{2}\beta' A_{22} \beta$.

The main result of this section is given by Theorem 4.4 in which a class of minimax estimators for β is obtained. It is evident that this class includes both Baranchik's (1973) estimator $\hat{\beta}_c$ as well as the m.l.e. $\hat{\beta}_0$.

THEOREM 4.4. Let F be as defined in Lemma 4.3. Then

$$\hat{\beta}_t = [1 - t(F)/F]\hat{\beta}_0 \tag{4.1}$$

is a minimax estimator of β, relative to the loss function (2.7), if

i) $t(F)$ is nondecreasing in F, and

ii) $0 \le t(F) \le 2(p - 2)/(n - p + 2)$.

PROOF. Let $R(\hat{\beta}, \beta)$ denote the risk, or expected loss, incurred when estimating β by $\hat{\beta}$. Stein (1960) showed that $\hat{\beta}_0$ is minimax with risk equal to $p/(n - p - 1)$. (Note that our n is Stein's and

Baranchik's $n - 1$). The theorem will be proved if we can show that

$$(4.2) \qquad R(\hat{\beta}_t, \beta) - R(\hat{\beta}_0, \beta) \le 0$$

for all parameter points (μ, Σ). Writing the risk in terms of expectations, (4.2) is equivalent to

$$(4.3) \qquad E\{[(\hat{\beta}_t - \beta)'\Sigma_{22}(\hat{\beta}_t - \beta) - (\hat{\beta}_0 - \beta)'\Sigma_{22}(\hat{\beta}_0 - \beta)]/\sigma_{11.2}\} \le 0.$$

Let E^* denote the conditional expectation, given A_{22}. Then (4.3) can be written as

$$(4.4) \qquad E\{E^*\{[(\hat{\beta}_t-\beta)'\Sigma_{22}(\hat{\beta}_t-\beta) - (\hat{\beta}_0-\beta)'\Sigma_{22}(\hat{\beta}_0-\beta)]/\sigma_{11.2}\}\} \le 0.$$

Thus (4.4) holds if

$$(4.5) \qquad E^*\{[(\hat{\beta}_t - \beta)'\Sigma_{22}(\hat{\beta}_t - \beta) - (\hat{\beta}_0 - \beta)'\Sigma_{22}(\hat{\beta}_0 - \beta)]/\sigma_{11.2}\} \le 0.$$

By the remark made at the end of Section 3 we may evaluate the conditional expectation on the LHS of (4.5) at $(\mu^{(2)}, \Sigma_{22}, \sigma_{11.2}) = (0, A_{22}, 1)$. Then (4.5) becomes

$$(4.6) \qquad E^*[(\hat{\beta}_t - \beta)'A_{22}(\hat{\beta}_t - \beta) - (\hat{\beta}_0 - \beta)'A_{22}(\hat{\beta}_0 - \beta)] \le 0.$$

Substituting (4.1) into (4.6) we obtain

$$(4.7) \qquad E^*[t^2(F)\hat{\beta}_0'A_{22}\hat{\beta}_0/F - 2t(F)\hat{\beta}_0'A_{22}\hat{\beta}_0/F + 2t(F)\hat{\beta}_0'A_{22}\beta/F] \le 0.$$

Denote by E^{**} the conditional expectation given K and A_{22}, where K given A_{22} is distributed as Poisson $(\frac{1}{2}\beta'A_{22}\beta)$. Then we may write the LHS of (4.7) with the use of Lemmas 4.1 and 4.2 as

(4.8) $$E^*\{E^{**}\{t(F)[t(F)\chi^2_{n-p}/F - 2\chi^2_{n-p} + 4K/F]\}\}.$$

Since $0 \le t(F) \le 2(p - 2)/(n - p + 2)$ the inner conditional expectation of (4.8) is less than or equal to

$$E^{**}\{t(F)\{2(p-2)(\chi^2_{n-p})^2/[(n-p+2)\chi^2_{p+2K}] - 2\chi^2_{n-p} + 4K\chi^2_{n-p}/\chi^2_{p+2K}\}\},$$

which in turn is equal to

(4.9) $$\mathrm{cov}^{**}\{t(F),\ 2(p-2)(\chi^2_{n-p})^2/[(n-p+2)\chi^2_{p+2K}] - 2\chi^2_{n-p} + 4K\chi^2_{n-p}/\chi^2_{p+2K}\},$$

where cov^{**} denotes the conditional covariance given K and A_{22}. It remains to show that for all K and A_{22} (4.9) is less than or equal to zero. Conditioning on $\chi^2_{n-p} = s$, (4.9) becomes

(4.10) $$E^{**}\{\mathrm{cov}^{***}\{t(\chi^2_{p+2K}/s),\ 2(p-2)s^2/[(n-p+2)\chi^2_{p+2K}] - 2s + 4Ks/\chi^2_{p+2K}\}\},$$

where cov^{***} denotes the conditional covariance given K, A_{22} and χ^2_{n-p}. By hypothesis $t(\chi^2_{p+2K}/s)$ is nondecreasing in χ^2_{p+2K}. Furthermore, the quantity

$$2(p-2)s^2/[(n-p+2)\chi^2_{p+2K}] - 2s + 4Ks/\chi^2_{p+2K}$$

is clearly decreasing in χ^2_{p+2K}. Hence the conditional covariance in (4.10) is nonpositive. This completes the proof.

5. Bayes Minimax Estimators for β.

The estimator (4.1) is minimax but not necessarily admissible. In this section we derive a class of generalized Bayes estimators for β

and show that they satisfy the conditions of Theorem 4.4 and hence are minimax. The class of generalized Bayes minimax estimators is of course essentially complete.

For convenience, we write X for $(X_1, \ldots, X_{n+1})$, $f(u|v)$ for the density function of u, given v, in general, and $n(z|\theta, \Gamma)$ for the density function of a normal vector z with mean θ and covariance matrix Γ in particular.

In estimating β, the parameters involved are elements of Σ, or equivalently β, $\sigma_{11.2}$ and Σ_{22}. Assume that a family of generalized prior distributions of β, $\sigma_{11.2}$ and Σ_{22} is given by the densities

$$\tau(\beta, \sigma_{11.2}, \Sigma_{22}) = \int_0^1 \int \{f(\beta|\lambda, \Phi)\, w(A_{22}|\Sigma_{22}, n) \times g(\Sigma_{22}^{-1}, \sigma_{11.2}^{-1})\, h(\lambda)\}\, dA\, d\lambda,$$

where the region of integration with respect to A is the subspace of the $p(p+1)/2$ - dimensional Euclidean space for which A is positive definite, and

(5.1) $$f(\beta|\lambda, \Phi) = n(\beta|0, \frac{1-\lambda}{\lambda}\Phi), \qquad \Phi = \sigma_{11.2}A_{22}^{-1},$$

(5.2) $$w(A_{22}|\Sigma_{22}, n) \propto |A_{22}|^{-\frac{1}{2}(n-p-1)}|\Sigma_{22}|^{-\frac{1}{2}n} \exp(-\tfrac{1}{2}\mathrm{tr}\, \Sigma_{22}^{-1} A_{22}),$$

(5.3) $$g(\Sigma_{22}^{-1}, \sigma_{11.2}^{-1}) \propto |\Sigma_{22}|^{\frac{1}{2}(p+1)}\, \sigma_{11.2}, \qquad \text{and}$$

(5.4) $$h(\lambda) \propto \lambda^{-a}, \qquad -\infty < a < \tfrac{1}{2}p + 1.$$

Since the sampling distribution of $\hat{\beta}_0$, given A_{22}, is $N(\beta, \sigma_{11.2}A_{22})$, the prior of β is taken to be a p-variate normal distribution with covariance matrix the same as that of $\hat{\beta}_0$ (given A_{22}) multiplied by a positive constant. Strawderman (1971) and Lin and Tsai (1973) have used similar priors in obtaining Bayes estimators for a multivariate normal mean. The prior of β given by (5.1) enables us to write $E(\beta|X) = [1 - E(\lambda|X)]\hat{\beta}_0$ which eventually will reduce to the form of (4.1) and hence minimax. Although the covariance matrix of β does depend on X, the prior of β given by (5.1) should not be confused with the posterior distribution of β, given X. In a similar situation when estimating the regression coefficient vectors of m populations Jackson, Novick and Thayer (1971, p. 148) in supporting the use of this prior stated: "When a regression model is used, the predictor variables are not <u>data</u> in the sense of random variables with probability distribution. In considering the form of his prior distributions and their mutual independence or lack of it, a Bayesian is entitled to make use of any information he may have about the predictors, whether this be <u>external</u> information about a distribution from which the predictor values may be supposed randomly drawn, or <u>internal</u> knowledge of the particular values the predictors take in the observations to hand." In view of this argument A_{22} is regarded as a parameter matrix and its prior distribution given by (5.2) is a natural consequence of Lemma 4.2.

The priors for Σ_{22}^{-1} and $\sigma_{11.2}^{-1}$ given by (5.3) have been used by many authors, e.g., Geisser & Cornfield (1963), Lin & Tsai (1973), and Tiao & Zellner (1964), among others. Of course (5.3) is not a proper prior. Combining the prior (5.4) with $0 < a < 1$ and the prior (5.1) Strawderman (1971) was able to obtain a family of proper Bayes minimax estimators for a multivariate normal mean when the covariance matrix is the identity and the dimension $p \geq 5$. Strawderman credited the use of these priors to Charles Stein. In (5.4) the condition on a is now relaxed to $-\infty < a < \frac{1}{2}p + 1$.

The main result of this section is presented in the following theorem where a family of generalized Bayes minimax estimators is obtained.

THEOREM 5.1. For $p \geq 3$ with

i) $n > p - 2$,

ii) $-\infty < a < \frac{1}{2}p + 1$,

iii) $2(p-2)/(n-p+2) \geq (p-2a+2)/(n-p+2a-4)$,

an estimator of the form

$$\hat{\beta}_s = [1 - s(F)/F]\hat{\beta}_0 \tag{5.5}$$

is generalized Bayes minimax with respect to the priors (5.1) - (5.4) and loss function (2.7), where $F = \hat{\beta}_0' A_{22} \hat{\beta}_0 / a_{11.2}$ and

$$s(F) = \frac{p-2a+2}{n-p+2a-4} - \frac{2(1+F)^{-\frac{1}{2}(n-2)}}{(n-p+2a-4)\int_0^1 \lambda^{\frac{1}{2}p-a}(1+\lambda F)^{-\frac{1}{2}n}\, d\lambda} . \tag{5.6}$$

The theorem will be proved by the following computational lemmas.

Note that $(a_{11.2}, \hat{\beta}_0, A_{22})$ is a sufficient statistic of $(\sigma_{11.2}, \beta, \Sigma_{22})$. In the remainder of this section X will denote $(a_{11.2}, \hat{\beta}_0, A_{22})$ when there is no confusion.

LEMMA 5.2. $E(\beta|X) = [1 - E(\lambda|X)]\hat{\beta}_0$.

PROOF. The joint probability density function (jpdf) of $\hat{\beta}_0$ and β, given $a_{11.2}$, A_{22}, $\sigma_{11.2}$, Σ_{22}, and λ is

$$\begin{aligned} f(\hat{\beta}_0, \beta|a_{11.2}, A_{22}, \sigma_{11.2}, \Sigma_{22}, \lambda) &= f(\hat{\beta}_0|\beta, a_{11.2}, A_{22}, \sigma_{11.2}, \Sigma_{22}, \lambda) \\ &\quad \times f(\beta|a_{11.2}, A_{22}, \sigma_{11.2}, \Sigma_{22}, \lambda) \\ &= n(\hat{\beta}_0|\beta, \Phi) \times n(\beta|0, \frac{1-\lambda}{\lambda}\Phi) \\ &= n(\hat{\beta}_0|0, \frac{1}{\lambda}\Phi) \times n[\beta|(1-\lambda)\hat{\beta}_0, (1-\lambda)\Phi], \end{aligned}$$

where Φ is defined in (5.1). This implies that the conditional distribution of β, given $a_{11.2}$, A_{22}, $\sigma_{11.2}$, Σ_{22}, λ, and $\hat{\beta}_0$ is $N[(1-\lambda)\hat{\beta}_0, (1-\lambda)\Phi]$ which does not depend on Σ_{22}. Thus

$$\begin{aligned} E(\beta|X) &= E[E(\beta|\sigma_{11.2}, \Sigma_{22}, \lambda, X)|X] \\ &= E[(1-\lambda)\hat{\beta}_0|X] \\ &= [1 - E(\lambda|X)]\hat{\beta}_0, \end{aligned}$$

as was to be proven.

LEMMA 5.3. The Bayes estimator of β with respect to the priors (5.1) - (5.4) and loss function (2.7) is given by

$$\hat{\beta}_\tau = [1 - E(\lambda|X)]\hat{\beta}_0. \tag{5.7}$$

PROOF. A Bayes rule minimizes the posterior expected loss

$$E[(\hat{\beta} - \beta)'\Sigma_{22}(\hat{\beta} - \beta)/\sigma_{11.2}|\ X], \tag{5.8}$$

where $\hat{\beta}$ is an estimator of β. To minimize (5.8) we solve the system of equations

$$\frac{\partial}{\partial\hat{\beta}} E[(\hat{\beta} - \beta)'\Sigma_{22}(\hat{\beta} - \beta)/\sigma_{11.2}|\ X] = 0. \tag{5.9}$$

Passing the differential operator under the integral sign, a solution $\hat{\beta}_\tau$ to (5.9) satisfies

$$E[(\hat{\beta}_\tau - \beta)'\Sigma_{22}/\sigma_{11.2}|\ X] = 0. \tag{5.10}$$

The LHS of (5.10) is equal to

$$\begin{aligned} &E\{E[(\hat{\beta}_\tau - \beta)'\Sigma_{22}/\sigma_{11.2}|\sigma_{11.2}, \Sigma_{22}, \lambda, X]|X\} \\ &= E\{\{E[(\hat{\beta}_\tau - \beta)'|\sigma_{11.2}, \Sigma_{22}, \lambda, X]\}\ \Sigma_{22}/\sigma_{11.2}|X\} \\ &= E[\hat{\beta}'_\tau - E(\beta'|\sigma_{11.2}, \Sigma_{22}, \lambda, X)|X]\ E(\Sigma_{22}/\sigma_{11.2}|X) \\ &= [\hat{\beta}'_\tau - E(\beta'|X)]\ E(\Sigma_{22}/\sigma_{11.2}|X). \end{aligned}$$

Therefore the Bayes rule is given by $\hat{\beta}_\tau = E(\beta|X)$ which together with Lemma 5.2 establishes (5.7).

<u>LEMMA 5.4</u>. The conditional expectation of λ given X is

$$(5.11) \qquad E(\lambda|X) = \frac{\int_0^1 \lambda^{\frac{1}{2}p-a+1} (1+\lambda F)^{-\frac{1}{2}n} \, d\lambda}{\int_0^1 \lambda^{\frac{1}{2}p-a} (1+\lambda F)^{-\frac{1}{2}n} \, d\lambda},$$

where F is defined in Lemma 4.3.

<u>PROOF</u>. With priors (5.1) - (5.4) the jpdf of $\hat{\beta}_0$, β, $a_{11.2}$, A_{22}, $\sigma_{11.2}^{-1}$, Σ_{22}^{-1}, and λ is

$$f(\hat{\beta}_0, \beta, a_{11.2}, A_{22}, \sigma_{11.2}^{-1}, \Sigma_{22}^{-1}, \lambda)$$

$$= f(\hat{\beta}_0, \beta | a_{11.2}, A_{22}, \sigma_{11.2}^{-1}, \Sigma_{22}^{-1}, \lambda)$$

$$\times f(a_{11.2}, A_{22}, \sigma_{11.2}^{-1}, \Sigma_{22}^{-1}, \lambda)$$

$$\propto \lambda^{\frac{1}{2}p} |\Phi|^{-\frac{1}{2}} \exp(-\tfrac{1}{2}\lambda\hat{\beta}_0'\Phi^{-1}\hat{\beta}_0)$$

$$\times \{|(1-\lambda)\Phi|^{-\frac{1}{2}} \exp\{-\frac{1}{2(1-\lambda)} [\beta-(1-\lambda)\hat{\beta}_0]'\Phi^{-1}[\beta-(1-\lambda)\hat{\beta}_0]\}\}$$

$$\times (a_{11.2}/\sigma_{11.2})^{\frac{1}{2}(n-p-2)} \sigma_{11.2}^{-1} \exp(-\frac{a_{11.2}}{2\sigma_{11.2}})$$

$$\times |A_{22}|^{\frac{1}{2}(n-p-1)} |\Sigma_{22}|^{-\frac{1}{2}n} \exp(-\tfrac{1}{2}\mathrm{tr}\, \Sigma_{22}^{-1} A_{22})$$

$$\times |\Sigma_{22}|^{\frac{1}{2}(p+1)} \sigma_{11.2} \lambda^{-a}.$$

After some simplifications and noting $\Phi = \sigma_{11.2}A_{22}^{-1}$ this reduces to

$$\lambda^{\frac{1}{2}p-a} |A_{22}|^{\frac{1}{2}(n-p)} |\Sigma_{22}|^{-\frac{1}{2}(n-p-1)} \exp(-\tfrac{1}{2}\mathrm{tr}\, \Sigma_{22}^{-1} A_{22})$$

$$\times (a_{11.2})^{\frac{1}{2}(n-p-2)} (\sigma_{11.2})^{-\frac{1}{2}(n-2)} \exp[-\tfrac{1}{2}(a_{11.2}+\lambda\hat{\beta}_0'A_{22}\hat{\beta}_0)/\sigma_{11.2}]$$

$$\times\ n[\beta|(1-\lambda)\hat{\beta}_0,\ (1-\lambda)\Phi]\ (2\pi)^{\frac{1}{2}p}.$$

Thus the jpdf of λ and X is

$$g(\lambda, X) = \iiint f(\hat{\beta}_0, \beta, a_{11.2}, A_{22}, \sigma^{-1}_{11.2}, \Sigma^{-1}_{22}, \lambda)\ d\beta\ d\Sigma^{-1}_{22}\ d\sigma^{-1}_{11.2}$$

$$(5.12) \qquad \propto \lambda^{\frac{1}{2}p-a}\ |A_{22}|^{-\frac{1}{2}p}\ (a_{11.2})^{\frac{1}{2}(n-p-2)}\ (a_{11.2} + \lambda\hat{\beta}_0'A_{22}\hat{\beta}_0)^{-\frac{1}{2}n}.$$

From (5.12) we have

$$E(\lambda|X) = \frac{\int_0^1 \lambda^{\frac{1}{2}p-a+1}\ (a_{11.2} + \lambda\hat{\beta}_0'A_{22}\hat{\beta}_0)^{-\frac{1}{2}n}\ d\lambda}{\int_0^1 \lambda^{\frac{1}{2}p-a}\ (a_{11.2} + \lambda\hat{\beta}_0'A_{22}\hat{\beta}_0)^{-\frac{1}{2}n}\ d\lambda}$$

$$= \frac{\int_0^1 \lambda^{\frac{1}{2}p-a+1}\ (1 + \lambda F)^{-\frac{1}{2}n}\ d\lambda}{\int_0^1 \lambda^{\frac{1}{2}p-a}\ (1 + \lambda F)^{-\frac{1}{2}n}\ d\lambda},$$

as was to be proven.

<u>LEMMA 5.5</u>. $E(\lambda|X) = s(F)/F$, where $s(F)$ is defined by (5.6).

<u>PROOF</u>. The desired result can be obtained by means of algebra after integrating by parts the numerator of (5.11).

<u>LEMMA 5.6</u>. Let $s(F)$ be defined by (5.6), then for all $F \geq 0$ and $n > p - 2$

i) $s(F)$ is nondecreasing in F, and

ii) $0 \leq s(F) \leq (p - 2a + 2)/(n - p + 2a - 4)$.

PROOF. That $s(F)$ is nondecreasing follows immediately from Lemma 3.1 of Lin and Tsai (1973) with $\nu = p + 2$. From Lemmas 5.4 and 5.5 it is clear that $s(F)$ is nonnegative. The fact that

$$(n-p+2a-4)(1 + F)^{\frac{1}{2}(n-2)} \int_0^1 \lambda^{\frac{1}{2}p-a} (1 + \lambda F)^{-\frac{1}{2}n} \, d\lambda \geq 0$$

and (5.6) yields $s(F) \leq (p-2a+2)/(n-p+2a-4)$. Hence the lemma follows.

PROOF OF THEOREM 5.1. The assumptions of Theorem 4.4 are satisfied by virtue of Lemma 5.6 along with Condition iii). This establishes the minimaxity of $\hat{\beta}_s$ given by (5.5). Lemmas 5.3 and 5.5 show that $\hat{\beta}_s$ is also generalized Bayes.

It should be noted that the estimators (5.5) are generalized Bayes for any dimension p. If the estimators are to be minimax, the set of values of n, a, and p must satisfy Conditions i) - iii) of Theorem 5.1. As a result p must be greater than or equal to 3. It is further noted that the integral expression in (5.6) may be replaced by an incomplete beta function, i.e.,

$$(5.13) \qquad \int_0^1 \lambda^{\frac{1}{2}p-a} (1 + \lambda F)^{-\frac{1}{2}n} \, d\lambda$$

$$= F^{-(\frac{1}{2}p-a+1)} \int_0^{F/(1+F)} y^{\frac{1}{2}p-a} (1 - y)^{\frac{1}{2}(n-p)+a-2} \, dy$$

$$= [(1-R^2)/R^2]^{\frac{1}{2}p-a+1} B[\tfrac{1}{2}p - a + 1, \tfrac{1}{2}(n - p) + a - 1]$$

$$\times I_{R^2}[\tfrac{1}{2}p - a + 1, \tfrac{1}{2}(n - p) + a - 1],$$

where $B(.,.)$ and $I_t(.,.)$ are beta and incomplete beta functions, respectively, and R is the sample multiple correlation.

6. Bayes Minimax Estimators for α and β.

In Sections 4 and 5 a class of generalized Bayes minimax estimators for β has been obtained. This result will now be extended to a class of estimators for both α and β. The loss function under consideration is now (2.6).

Since $X_1, \ldots, X_{n+1}$ are i.i.d. $N(\mu, \Sigma)$, $\bar{X}$ is distributed as $N(\mu, \frac{1}{n+1}\Sigma)$. Assume that the prior distribution of μ is given by

$$(6.1) \qquad dH(\mu) = d\mu.$$

This noninformative prior has been used by Geisser and Cornfield (1963) and numerous other authors. We also assume the priors (5.1) - (5.4) for β, $\sigma_{11.2}$, and Σ_{22} for the remainder of this section.

The main result is given in Theorem 6.1 to be followed by two computational lemmas which together with Theorem 5.1 and a result of Baranchik (1973) will prove the theorem.

THEOREM 6.1. If Conditions i) - iii) of Theorem 5.1 hold, then an estimator of the form

$$(6.2) \qquad \hat{\alpha}_s = \bar{X}^{(1)} - \hat{\beta}_s' \bar{X}^{(2)}, \qquad \hat{\beta}_s = [1 - s(F)/F]\hat{\beta}_0$$

is generalized Bayes minimax with respect to the priors (5.1) - (5.4) and (6.1) relative to the loss function (2.6) where $s(F)$ is given by (5.6).

Lemma 6.2 derives the distribution of μ, given $\bar{X}$ and Σ. Its proof is straight forward and is omitted. Lemma 6.3 obtains a class of generalized Bayes estimators of α and β.

<u>LEMMA 6.2</u>. With the prior (6.1), μ given $\bar{X}$ and Σ is distributed as $N(\bar{X}, \frac{1}{n+1}\Sigma)$.

<u>LEMMA 6.3</u>. With the priors (5.1) - (5.4) and (6.1) a generalized Bayes estimator of α and β with respect to loss function (2.6) is given by

$$\hat{\alpha}_\tau = \bar{X}^{(1)} - \hat{\beta}_\tau' \bar{X}^{(2)}, \tag{6.3}$$

and

$$\hat{\beta}_\tau = [1 - E(\lambda|X)]\hat{\beta}_0, \tag{6.4}$$

where X denotes $(\hat{\beta}_0, \varepsilon_{11.2}, A_{22})$.

<u>PROOF</u>. Proceeding as in Lemma 5.3 we seek a solution to the system of equations

$$\frac{\partial}{\partial\hat{\alpha}} E\{\{[(\hat{\alpha}-\alpha) + (\hat{\beta}-\beta)'\mu^{(2)}]^2 + (\hat{\beta}-\beta)'\Sigma_{22}(\hat{\beta}-\beta)\}/\sigma_{11.2}|X, \bar{X}\} = 0 \tag{6.5}$$

and

$$\frac{\partial}{\partial\hat{\beta}} E\{\{[(\hat{\alpha}-\alpha) + (\hat{\beta}-\beta)'\mu^{(2)}]^2 + (\hat{\beta}-\beta)'\Sigma_{22}(\hat{\beta}-\beta)\}/\sigma_{11.2}|X, \bar{X}\} = 0, \tag{6.6}$$

where $\hat{\alpha}$ and $\hat{\beta}$ are any estimators of α and β. It is easy to verify that

$$\hat{\alpha}_\tau = E(\alpha|\bar{X}, X). \tag{6.7}$$

and

$$\hat{\beta}_\tau = E(\beta|\bar{X}, X) \tag{6.8}$$

satisfy (6.5) and (6.6) and hence are generalized Bayes. By the independence of A and $\bar{X}$ we have $E(\beta|\bar{X}, X) = E(\beta|X)$. With the aid of Lemma 6.2 and Lemma 5.3, the results (6.3) and (6.4) follow directly from (6.7) and (6.8), respectively.

The following lemma of Baranchik (1973) will be used to prove the minimaxity of (6.2) in Theorem 6.1.

LEMMA 6.4. If an estimator of the form $f(R^2)\hat{\beta}_0$ dominates $\hat{\beta}_0$ relative to the loss function (2.7), then the corresponding estimator

(6.9) $$\hat{\alpha}_f = \bar{X}^{(1)} - \hat{\beta}_f' \bar{X}^{(2)}, \qquad \hat{\beta}_f = f(R^2)\hat{\beta}_0$$

dominates $\hat{\alpha}_0$ and $\hat{\beta}_0$ relative to the loss funciton (2.6).

PROOF OF THEOREM 6.1. The minimaxity of $(\hat{\alpha}_s, \hat{\beta}_s)$ given by (6.2) follows from Theorem 5.1 and Lemma 6.4. Lemma 6.3 shows that (6.2) is also generalized Bayes.

7. Concluding Remarks.

Theorem 4.4 is an extension of Baranchik's (1973) Theorem 1 which appears to be incorrect unless the sample size is $n + 1$ rather than n. Furthermore, the proofs of Baranchik's results may be simplified using the invariance argument of Section 4.

It is noted that the modified estimator

(7.1) $$\hat{\alpha}_\psi = \bar{X}^{(1)} - \hat{\beta}_\psi' \bar{X}^{(2)}, \qquad \hat{\beta}_\psi = \psi(F)\hat{\beta}_0,$$

where

(7.2) $$\psi(F) = \max\{0, [1 - s(F)/F]\},$$

is better than (6.2). Similar result also holds for the estimation of β alone.

This paper partially answers Stein's (1960) question on the existence of admissible minimax estimators of the form $\phi(R^2)\hat{\beta}_0$ for β. While we have not shown the admissibility, we have obtained a class of minimax estimators of the form $\phi(R^2)\hat{\beta}_0$, and in particular exhibited a subclass that is generalized Bayes. The admissibility of these estimators is yet to be shown.

REFERENCES

BARANCHIK, A.J. (1970). A family of minimax estimators of the mean of a multivariate normal distribution. Ann. Math. Statist. 41, 642-45.

BARANCHIK, A.J. (1973). Inadmissibility of maximum likelihood estimators in some multiple regression problems with three or more independent variables. Ann. Statist. 1, 312-21.

GEISSER, S. and CORNFIELD, J. (1963). Posterior distributions for multivariate normal parameters. J. Roy. Statist. Soc. Ser. B 25, 368-76.

JACKSON, P., NOVICK, M., and THAYER, D. (1971). Estimating regression in groups. Br. J. Math. Statist. Psychol. 24, 129-53.

LIN, P.E. and TSAI, H.L. (1973). Generalized Bayes minimax estimators of the multivariate normal mean with unknown covariance matrix. Ann. Statist. 1, 142-45.

STEIN, C. (1960). Multiple regression. Contributions to Probability and Statistics. (Olkin ed.) Stanford University Press.

STEIN, C. (1969). Multivariate analysis I. (Prepared by M. L. Eaton). Technical Report No. 42, Stanford University.

STRAWDERMAN, W.E. (1971). Proper Bayes minimax estimator of the multivariate normal mean. Ann. Math. Statist. 42, 385-88.

TIAO, G.C. and ZELLNER, A. (1964). On the Bayesian estimation of multivariate regression. J. Roy. Statist. Soc. Ser. B 26, 277-85.