STATISTICAL ANALYSIS OF NONESTIMABLE FUNCTIONALS

DIBYEN MAJUMDAR

Thesis submitted to the Indian Statistical Institute in partial fulfilment of the requirements for the award of Doctor of Philosophy

NEW DELHI
1979

ACKNOWLEDGEMENTS

I am deeply indebted to Professor Sujit Kumar Mitra, my thesis supervisor, for his masterly guidance, for the interest in the subject that he aroused in me and for the training in Statistics that I received from him.

I am grateful to Dr. Yadolah Dodge for the benefit I derived from my association with him. I thank Professor Mitra and Dr. Dodge for their kind permission to include in this thesis some of the work I did jointly with them.

I had some very helpful discussions with Sri S. Ramasubramanian. I record my gratitude for him and for all who helped and encouraged me in this effort.

I thank the authorities of the Indian Statistical Institute for providing the facilities for carrying out this work.

I am thankful to all the members of our Computer Centre, especially Sri R.N. Kar, for their cooperation while doing the numerical exercises. Finally, I thank Sri V.P. Sharma for his elegant typing of the manuscript.

CONTENTS

		Page No.
	NOTATIONS	(i)
CHAPTER 1 :	INTRODUCTION	1
CHAPTER 2 :	ESTIMATING NONESTIMABLE FUNCTIONALS	11
	1. Some approaches to estimating nonestimable functionals	11
	2. A bounded parameter space for β	18
	3. Comparison of rival estimators	30
CHAPTER 3 :	BHLE AND ALE	44
	1. BHLE	44
	2. ALE	51
CHAPTER 4 :	TESTING NONTESTABLE FUNCTIONALS	59
	1. The Test	60
	2. The best test for certain hypotheses	67
	3. The best test	76
CHAPTER 5 :	ANALYSING DATA FROM CLASSIFICATION MODELS WITH ARBITRARY PATTERNS	88
	1. Preliminaries	90
	2. Least square generalized inverses for additive two way models	92
	3. Least square generalized inverses for L-way classification models	102
	REFERENCES	109

NOTATIONS

R^n = n-dimensional Euclidean space.

$R^{n\times m}$ = set of all matrices with n rows and m columns, having real entries

For a subset S of a linear space,

$\mathcal{L}(S)$ = smallest subspace in which the whole of S can be embedded.

For a linear space V,

$\dim(V)$ = dimension of V.

For a matrix A in $R^{n\times m}$

$R(A)$ = rank of A,

$\mathcal{M}(A)$ = $\{Ax : x\epsilon R^m\}$, the column span of A,

$\mathcal{N}(A)$ = $\{y \epsilon R^m : Ay = 0\}$, the null space of A.

For a positive definite matrix N in $R^{n\times n}$,

$P_{A,N}$ = $A(A'NA)^- A'N$, the orthogonal projection operator onto $\mathcal{M}(A)$ when the inner product in R^n is induced by N; $(A'NA)^-$ being a generalized inverse (g-inverse) of A'NA.

P_A = $P_{A,I}$; I being the identity matrix.

For generalized inverses of matrices, the notations of Rao and Mitra [33] will be adopted. $A^+_{N\theta M}$ will denote an optimal inverse of a matrix as defined by Mitra [23].

CHAPTER 1

INTRODUCTION

Our interest will be centred around the Gauss Markov model $(Y, X\beta, \sigma^2\Lambda)$, where Y is a random variable assuming values in R^n with expectation and dispersion matrix given by

$$E(Y) = X\beta \tag{1.1}$$

$$D(Y) = \sigma^2\Lambda \tag{1.2}$$

X in $R^{n\times m}$ and Λ in the subset of nonnegative definite (n.n.d.) matrices of $R^{n\times n}$ are known. Unless specified to the contrary, Λ will be assumed to be positive definite (p.d.). The unknown parameter vector β varies in Ω_β, a subset of R^m and σ^2 in Ω_σ. Ω_σ will always be the positive half of the real line, unless otherwise specified, and likewise Ω_β satisfies the minimum requirement dim $\mathcal{L}(\Omega_\beta) = m$.

Historically, the first contribution towards estimating linear functionals of β was by Gauss [16] who in 1821 showed that the method of least squares provides the BLUE (best linear unbiased estimator) of β, when $R(X) = m$ and $\Lambda = I$. Markov [22] in 1912 and David and Neyman [12] in 1938 gave a systematic presentation of the theory under the same conditions. In 1934 Aitken [1] considered the setup where $R(X) = m$, but Λ could be any positive definite matrix. Bose [6] in 1944 was the first to consider deficiencies in $R(X)$. In his model $R(X) = r < m$ and $\Lambda = I$, while Rao [28] in 1945 generalized this to any positive definite Λ. In all of these Ω_β was R^m.

Bose [6] defined a linear parametric functional $p'\beta$ as being estimable if it posesses an unbiased estimator which is linear in Y. Any parametric functional without this property was nonestimable. If $\dim \mathcal{L}(\Omega_\beta) = m$, then it is not difficult to see that $p'\beta$ is (linearly) estimable if and only if there exists a $b \in R^n$, such that $p = X'b$, $b'Y$ being an unbiased estimator of $p'\beta$ in such cases. Clearly nonestimable functionals are precisely those for which $p \notin \mathcal{M}(X')$. For simultaneous estimation of several linear parametric functionals, $P\beta$ (P being a matrix) is estimable if $\mathcal{M}(P') \subset \mathcal{M}(X')$, violation of which will make it nonestimable. If $R(X) < m$, then, in particular β is nonestimable.

While estimable functionals have been studied in depth, very few theoretical investigations are concerned with nonestimable linear parametric functionals. However, experimenters using a fractional replicate of a factorial design are often required to estimate important factorial effects which are nonestimable on account of incomplete replication. Even if one starts with a full rank design matrix (X), or at least one which allows for unbiased estimation of all the parametric functionals of interest, due to factors beyond one's control, data might only be available for a subset of the experimental points at the end of the experiment, to render important parametric functionals nonestimable. This is the well known problem of missing observations. Since it is generally impossible to repeat the experiment(with the hope of better outcomes), it is of great importance to device ways to infer about nonestimable functionals. This work will be directed towards this goal. But, before launching upon the task, let us examine the difficulties which may arise, and how we might be able to overcome them.

Let $\mathcal{P}_{\beta,\sigma^2}$ denote the probability distribution of Y. The parametric functional $p'\beta$ is said to be identifiable by distribution if for distinct parameter points (β_1,σ_1^2) and (β_2,σ_2^2),

$$\mathcal{P}_{\beta_1,\sigma_1^2} = \mathcal{P}_{\beta_2,\sigma_2^2} \Rightarrow p'\beta_1 = p'\beta_2$$

Consider now the case where $\mathcal{P}_{\beta,\sigma^2}$ depends on β only through $X\beta$. This would be the situation, for example when $Y \sim N_n(X\beta,\sigma^2\Lambda)$. If $p'\beta$ is (linearly) estimable, then there exists a $b \in R^n$, such that $p = X'b$. Hence $\mathcal{P}_{\beta_1,\sigma_1^2} = \mathcal{P}_{\beta_2,\sigma_2^2} \Rightarrow X\beta_1 = X\beta_2 \Rightarrow p'\beta_1 = p'\beta_2$. Conversely if $X\beta_1 = X\beta_2 \Rightarrow p'\beta_1 = p'\beta_2$, then surely $p \in \mathcal{M}(X')$. More generally, if there is a function of Y, f(Y) say, for which $E(f(Y)/\mathcal{P}_{\beta,\sigma^2}) = p'\beta$ for all $\beta \in \Omega_\beta$ then $\mathcal{P}_{\beta_1,\sigma_1^2} = \mathcal{P}_{\beta_2,\sigma_2^2} \Rightarrow E(f(Y)/\mathcal{P}_{\beta_1,\sigma_1^2}) = E(f(Y)/\mathcal{P}_{\beta_2,\sigma_2^2}) \Rightarrow p'\beta_1 = p'\beta_2$. Thus we arrive at the following theorem, due to Bunke and Bunke [8]:

<u>Theorem 1.1</u> If the probability distribution of a random variable Y depends on a parameter β only through $X\beta$, then the following are equivalent:

(1) $p'\beta$ is linearly estimable, i.e. $p \in \mathcal{M}(X')$

(2) $p'\beta$ is identifiable by distribution

(3) $p'\beta$ has an unbiased estimator, not necessarily linear.

Theorem 1.1 strengthens Bose's definition of nonestimability, since if $q \notin \mathcal{M}(X')$, $q'\beta$ is not the expectation of any function of the observations Y, linear or otherwise. On the other hand, the fact that such functionals are nonidentifiable by distribution is alarming, since it implies that even when we have complete knowledge of $\mathcal{P}_{\beta,\sigma^2}$, we are not sure of the value assumed by a nonestimable $q'\beta$. Evidently, the worst would happen when the probability distribution of the estimator we propose for $q'\beta$ is insensative to the value assumed by $q'\beta$, and this could happen unless one has some prior information on the parameters to supplement the information contained in the observations Y. This prior information could either be in the form of an assumed prior distribution for the parameters or at least as bounds on values the parameters could assume. The latter assumption sometimes arises quite naturally, like in the linear consumption model $E(y) = \alpha+\beta x$, where x is the income, y the consumption, α the threshold consumption and β the marginal propensity to consume, which necessarily varies in the interval [0,1]. Even otherwise, the assumption of a bounded parameter space is usually quite realistic and indeed in several applications the experimenter on the basis of past experience would be able to place appropriate bounds on the parameters that are universally acceptable. We shall consider only one kind of bounded region for β - the ellipsoid in R^m. If some other region seems more appropriate, then one can obtain approximate inferences on the parameters by enclosing the region in the smallest ellipsoid and then using the methods which we shall develop. Thus in a major part of this work we shall assume that

$$\Omega_\beta = \{\beta : \beta'H\beta \leq \delta^2\} \tag{1.3}$$

If the actual centre of the ellipsoid is at a known point $\alpha(\neq 0) \in R^m$, the model can be transformed to $(Y-X\alpha, X\beta, \sigma^2\Lambda)$ with $\beta \in \Omega_\beta$ as defined in (1.3). Hence choosing $\alpha = 0$ involves no loss of generality.

Let us examine the nonidentifiability of a nonestimable $q'\beta$ in this new context. To keep our discussions simple we shall restrict ourselves to the case where the matrix H in (1.3) is p.d. Writing $q = X'b + s$, we notice that the part $p'\beta = b'X\beta$ of $q'\beta$ being estimable is identifiable by distribution, while, by Cauchy Schwarz inequality, $s'\beta$ can vary in

$[-\delta\sqrt{s'H^{-1}s},\ \delta\sqrt{s'H^{-1}s}\ ]$. This interval of uncertainity of $s'\beta$ is shortest when one chooses $p = P_{X',H^{-1}}\, q = p_o$ (say) and $s = (I-P_{X',H^{-1}})q = s_o$, and accordingly one could propose $\delta\sqrt{s_o'H^{-1}s_o}$ as a measure of nonidentifiability of a nonestimable $q'\beta$. In fact one can do even better noting that a measure of nonidentifiability should indicate the extent to which $q'\beta$ is indeterminable given $\mathcal{P}_{\beta,\sigma^2}$. Thus given $X\beta = \mu$, $\beta \in \Omega_\beta$, and writing $G = H^{-1}X'(XH^{-1}X')^- = X^-_{m(H)}$ it is easy to see from the known properties of minimum H-norm generalized inverses that

$$q'\beta = q'G\mu + q'(I-GX)z, \text{ where } z'H(I-GX)z \leq$$

$$\delta^2 - \mu'G'HG\mu. \text{ Since } GX = P'_{X',H^{-1}},\ q'(I-GX)z = s_o'z$$

$$\leq \sqrt{[(\delta^2-\mu'G'HG\mu)s_o'H^{-1}s_o]} \leq \delta\sqrt{s_o'H^{-1}s_o}$$

Moreover, if one considers the extreme situation

$$X\beta = (\delta/\sqrt{t'Ht})f \qquad (1.4)$$

where $f \in R^n$ and $t = X^-_{m(H)}f$, then the minimum H-norm solution of (1.4) lies on the surface of Ω_β and thus there is a unique β satisfying (1.4) and (1.3). Thus here $q'\beta$ is unique and is as identifiable as any estimable parametric functional. In problems of inference if one considers an estimable $p'\beta$ in place of the nonestimable $q'\beta$, then one at least has an idea of the (finite) maximum error $(\delta\sqrt{s'H^{-1}s})$ that one can commit in the process of this approximation.

The situation is more revealing when one considers the testing problem. Let $Y \sim N_n(X\beta,\sigma^2\Lambda)$ and $\Omega_\beta = R^m$. The following theorem is well known to all students of Mitra, since he has been teaching it for the last twenty years, and possibly to many others also. However, the author has not seen it anywhere in print.

<u>Theorem 1.2</u> Any test for $H_o : q'\beta = c$, where $q \notin \mathcal{M}(X')$, has a constant power function.

<u>Proof</u> Under H_o one has to solve

$$\begin{aligned} X\beta &= \mu \\ q'\beta &= c \end{aligned} \qquad (a)$$

and under an alternative $q'\beta = d \neq c$,

$$\begin{aligned} X\beta &= \mu \\ q'\beta &= d \end{aligned} \qquad (b)$$

If S_a and S_b are the solution sets for β for (a) and (b) respectively,

$$\{X\beta : \beta \in S_a\} = \{X\beta : \beta \in S_b\} , \qquad (1.5)$$

since values assumed by nonestimable functionals do not affect the range of variation of estimable functionals. Under normality assumption, the power of a test depends on β through $X\beta$ only, and hence the power at each point is α, the preassigned size of the test.

q.e.d.

Thus here the hypothesis and the alternatives are indistinguishable on the basis of observations on Y. On the other hand if Ω_β is bounded as in (1.3), then (1.5) need no longer hold, and using an approximation to $q'\beta$ by an estimable $p'\beta$ (details of which will be considered in Chapter 4) it is possible to construct a test function ϕ such that

$$\sup_{\beta : q'\beta = c} E\phi = \alpha \text{ , and for } d \neq c \text{ ,}$$

$$\sup_{\beta : q'\beta = d} E\phi > \alpha. \text{ In fact for d sufficiently}$$

removed from c, even $\inf_{\beta : q'\beta = d} E\phi > \alpha$. The only exception to this, for the class of test functions we consider in chapter 4, is possibly when $||P_{X', H^{-1}} q||$ is zero or very close to it.

We shall close our discussion of the problem with a consideration of the admissible class of linear estimators for $q'\beta$ for bounded and unbounded Ω_β. Our criterion will be the mean square error, or in other words, if $g'Y$ is an estimator of $q'\beta$ then the risk function is

$$R(g'Y,q'\beta) = E(g'Y-q'\beta)^2 = \sigma^2 g'\Lambda g + [(X'g-q)'\beta]^2 \qquad (1.6)$$

An estimator $g'Y$ will be said to be admissible for $q'\beta$ if there does not exist another estimator $h'Y$ such that

$$R(h'Y,q'\beta) \leq R(g'Y,q'\beta) \text{ for all } \beta \in \Omega_\beta \qquad (1.7)$$

with inequality strict for at least one β in Ω_β. Firstly note that for any g, $h = \Lambda^{-1}P_{X,\Lambda^{-1}}\Lambda g = (P_{X,\Lambda^{-1}})'g$ satisfies (1.7) since $h'\Lambda h \leq g'\Lambda g$ and $X'h = X'g$. The only exception to this rule is when $\Lambda g \in \mathcal{M}(X)$. Here $h = (P_{X,\Lambda^{-1}})'g = g$. Disregarding such estimators $g'Y$, we shall henceforth consider estimators only from the class

$$\mathcal{G} = \{g'Y : \Lambda g \in \mathcal{M}(X)\} \qquad (1.8)$$

whilst discussing admissibility.

Let $\mathcal{A}^{q'\beta}_{\Omega_\beta}$ be the class of admissible linear estimators for $q'\beta$ in the Gauss-Markov model $(Y,X\beta,\sigma^2\Lambda)$ when β assumes values in Ω_β. Rao [31] characterized $\mathcal{A}^{q'\beta}_{R^m}$ for estimable parametric functionals. Part (i) of theorem 1.3 is a special case of his Theorem 6.6 when only a single parametric functional is considered instead of several of them, while part (ii) is compatible with a remark he makes, viz. if $q'\beta$ is nonidentifiable then there is a risk that every function might be admissible.

<u>Theorem 1.3</u> (i) If $q \in \mathcal{M}(X')$ then

$$\mathcal{A}^{q'\beta}_{R^m} = \{g'Y \in \mathcal{C} : g'\Lambda g \leq g'\Lambda b\}$$

where b is any vector satisfying $X'b = q$.

(ii) If $q \notin \mathcal{M}(X')$ then

$$\mathcal{A}^{q'\beta}_{R^m} = \mathcal{C}$$

<u>Proof</u> Observe that $g'\Lambda b$ is invariant under choice of b satisfying $X'b = q$ and hence $\mathcal{A}^{q'\beta}_{R^m}$ in (i) is well defined. To prove (ii) observe that for any $g'Y \in \mathcal{C}$, and h satisfying (1.7),

$$[(X'h-q)'\beta]^2 \leq [(X'g-q)'\beta]^2 \;\forall \beta \in R^m \qquad (1.9)$$

and

$$h'\Lambda h \leq g'\Lambda g \qquad (1.10)$$

From (1.9) we have $X'h - q = c(X'g-q)$, $c \in [-1,1]$. Since $q \notin \mathcal{M}(X')$, surely $c = 1$. Hence $h = g+\alpha$, where α is some vector such that $X'\alpha = 0$. From (1.10), we find that $h'\Lambda h = g'\Lambda g + \alpha'\Lambda\alpha + 2g'\Lambda\alpha \leq g'\Lambda g$. But $g'\Lambda\alpha = 0$, since $g'Y \in \mathcal{C}$. Thus $\alpha'\Lambda\alpha \leq 0$, which is impossible unless $\Lambda\alpha = 0$. Hence (1.7) is an identity, with strict inequality nowhere in Ω_β. Infact $h = g$ if Λ is p.d., and $h'Y = g'Y$ almost everywhere even otherwise. Thus no estimator in $\mathcal{C}$ can be inadmissible.

For $\Omega_\beta = R^m$ the class of admissible estimators for a nonestimable $q'\beta$ is the whole of $\mathcal{C}$, leaving no scope for uniformly improving on any estimator, however bad it might be from other considerations. On the other hand, if Ω_β is of the form as in (1.3) then

$$\max_{\beta\in\Omega_\beta} R(h'Y,q'\beta) = \sigma^2 h'\Lambda h + \delta^2(X'h-q)'H^{-1}(X'h-q)$$

and any $g'Y$ such that

$$\sigma^2 g'\Lambda g > \sigma^2 h'\Lambda h + \delta^2 (X'h-q)'H^{-1}(X'h-q)$$

is inadmissible since

$$R(h'Y,q'\beta) \leq \underset{\beta\epsilon\Omega_\beta}{\text{Max}}\ R(h'Y,q'\beta) < \underset{\beta\epsilon\Omega_\beta}{\text{Min}}\ R(g'Y,q'\beta) \leq R(g'Y,q'\beta)$$

for all β in Ω_β. In fact a necessary condition for $g'Y$ to be admissible is

$$\sigma^2 g'\Lambda g \leq \sigma^2 g_1'\Lambda g_1 + \delta^2 (X'g_1-q)'H^{-1}(X'g_1-q) \qquad (1.11)$$

where $g_1'Y$ is the minimax estimator (i.e. it minimizes $\underset{\beta\epsilon\Omega_\beta}{\text{Max}}\ R(h'Y,q'\beta)$ among all $h \epsilon R^n$), which we shall discuss in chapter 2.

Possibly the first attempt towards estimation of nonestimable functionals was by Chipman [11] in 1964. His approach, and the subsequent ones will be elaborated upon in Chapter 2. In this chapter we shall also develop several estimators for the bounded parameter case. These estimators will be compared on the basis of their biases and mean square errors. Numerical comparison based on simulation will be resorted to whenever theoretical expressions are difficult to obtain.

Optimal estimators for $q'\beta$, when β has an assumed prior distribution will be studied in chapter 3, along with comparisons among some such estimators. This chapter also includes some results on admissible linear estimators, especially in the class of the Bayes estimators considered.

Tests for hypotheses specifying the value of a nonestimable $q'\beta$, when Ω_β satisfies (1.3) will be developed in chapter 4, along the lines hinted earlier.

We have already observed that data with many missing observations are liable to give rise to nonestimable functionals in the model. Analysis of some such data, more precisely those arising from classification models with arbitrary patterns will be considered in chapter 5. A simple algorithm without rounding off errors, applicable to any classification model data will be developed in this chapter.

Each chapter will have an introduction of its own.

We shall not be citing specific references for the well known results in generalized inverses and matrix algebra which will be used. These may be obtained from Rao [29] and Rao and Mitra [33].

CHAPTER 2

ESTIMATING NONESTIMABLE FUNCTIONALS

In section 1 we consider some approaches to estimating nonestimable functionals. These have been studied by various authors, may be not in the same form or to a similar depth as we shall consider them. These approaches either assume $\Omega_\beta = R^m$ or make no specific assumption in this regard. In section 2 we consider various methods of estimation when Ω_β is bounded as in (1.3). A study of their relative merits and demerits will be taken up in section 3.

1. Some approaches to estimating nonestimable functionals

(i) Best Linear Minimum Bias Estimator (BLIMBE)

This approach has been studied by Chipman [11], Drygas [13], Rao and Mitra [33,p.139], Schonfeld [35] and several other authors. We note that a nonestimable linear parametric functional $q'\beta$ has no linear unbiased estimator and that the actual bias $(X'b-q)'\beta$ of the estimator $b'Y$ is a function of unknown parameters. A sensible way of controlling bias seems to be to choose b such that the coefficient vector $X'b-q$ of the expression determining bias is as close to the null vector as possible in some acceptable sense. With the norm $||X'b-q||$ of the vector as induced by a p.d. $M \in R^{m \times m}$, that is

$$||X'b-q|| = \sqrt{(X'b-q)'M(X'b-q)} \qquad (2.1.1)$$

an optimal choice of b is given by $b = (X')^-_{\ell(M)}q$. The corresponding estimator may be called a minimum bias estimator. Since $V(b'Y) = \sigma^2 b'\Lambda b$, the minimum variance minimum bias estimator uses as coefficient vector a least squares solution with the least seminorm. The BLIMBE is accordingly given by

$$b_o'Y = q'[(X')^+_{M\Lambda}]'Y \qquad (2.1.2)$$

which simplifies to $q'X^+_{\Lambda^{-1}M^{-1}}Y$ when Λ is invertible.

We note that the vector $X'b = X'(X')^-_{\ell(M)}q = P_{X',M}\,q$ is unique for every least square solution b. This implies in particular that every minimum bias estimator has precisely the same bias. The BLIMBE is thus automatically the minimum mean square error minimum bias estimator.

(ii) Conditionally Unbiased Estimators.

Consider the normal equations

$$X'\Lambda^{-1}X\beta = X'\Lambda^{-1}Y \qquad (2.1.3)$$

associated with the model $(Y, X\beta, \sigma^2\Lambda)$. It is well known that equations (2.1.3) have a multiplicity of solutions when $R(X) = r < m$. One could however get a unique solution requiring for example that the solution $\hat{\beta}$ of (2.1.3) belongs to $\mathcal{M}(K)$, a given subspace of R^m, and that $R(XK) = R(K) = r$ which will guarantee this uniqueness. If $K^{\perp} \varepsilon R^{m-r \times m}$ is a matrix of rank m-r such that $K^{\perp}K = 0$ (the null matrix), then $K^{\perp}\beta = 0$ in addition to (2.1.3) and the unique solution is $\hat{\beta}=(X'\Lambda^{-1}X+K^{\perp\prime}K^{\perp})^{-1}X'\Lambda^{-1}Y$. Alternatively $\hat{\beta} = K(K'X'\Lambda^{-1}XK)^-K'X'\Lambda^{-1}Y = K(XK)^-_{\ell(\Lambda^{-1})}Y$. It is not difficult to see that $E(\hat{\beta}) = \beta$ if

$$\mathcal{L}(\Omega_\beta) = \mathcal{M}(K) \qquad (2.1.4)$$

In this sense $\hat{\beta}$ may be called a conditionally unbiased estimator of β. In choosing this solution one is thus playing safe in that if $p'\beta$ is estimable then $p'\hat{\beta}$ is still the BLUE of $p'\beta$. If $q'\beta$ is nonestimable $q'\hat{\beta}$ is conditionally the BLUE of $q'\beta$ if (2.1.4) holds. This observation is due to Plackett [27] (see also Scheffe [34, p.19]).

In general one may enquire if it is possible to find a matrix G such that GY is conditionally an unbiased estimator of β given that (2.1.4) holds. This is equivalent to requiring that G satisfies the equation

$$GXK = K \tag{2.1.5}$$

for consistency of which we demand that

$$R(XK) = R(K) \leq R(X) \tag{2.1.6}$$

Defining $\tilde{G} = K(XK)^{-}_{\ell(\Lambda^{-1})} = K(XK)^{+}_{\Lambda^{-1}I}$, it is easy to establish the following:

<u>Theorem 2.1.1</u> $\tilde{G}$ satisfies (2.1.5). Moreover for any G satisfying (2.1.5), $D(GY) - D(\tilde{G}Y)$ is nonnegative definite (n.n.d.)

Proof of the first statement is immediate. To prove the second part observe that

$$D(GY) - D(\tilde{G}Y) = + D(GY-\tilde{G}Y) + 2Cov(GY-\tilde{G}Y, \tilde{G}Y).$$

The covariance term vanishes since $(G-\tilde{G})\Lambda\,\tilde{G}' = 0$ as $\mathcal{M}(\tilde{G}') = \mathcal{M}(\Lambda^{-1}XK)$ Hence the theorem.

In the sense of theorem 2.1.1 $\tilde{\beta} = \tilde{G}Y$ may be called the optimal conditionally unbiased estimator of β.

When one is sufficiently convinced about the validity of (2.1.4) one could actually reformulate the model as $(Y, XK\gamma, \sigma^2\Lambda)$ noting that $\beta = K\gamma$ for some γ. If further (2.1.6) holds $\beta = K\gamma$ is indeed estimable and $K\hat{\gamma} = K(XK)^{-}_{\ell(\Lambda^{-1})}\,Y = \tilde{G}Y = \tilde{\beta}$ is conditionally the BLUE of β if (2.1.4) holds. If equality holds all through in (2.1.6), $K\hat{\gamma}$ is infact the unique solution $\hat{\beta}$ of (2.1.3) to which a reference was made earlier in this section. In general if $R(K) = R(XK) < r$, then $\tilde{\beta}$ may not even satisfy (2.1.3).

Observe that $E(p'\tilde{\beta}) = p'\beta$ can hold for all β in R^m if and only if $X'\Lambda^{-1}XK(K'X'\Lambda^{-1}XK)^-K'p = p$, which is equivalent to saying that $p \in \mathcal{M}(X'\Lambda^{-1}XK)$. The same condition is also necessary and sufficient for $p'\tilde{\beta} = p'\hat{\beta}$, the BLUE of $p'\beta$, though in general if $p = X'b = X'\Lambda^{-1/2}b_1$, then $V(p'\tilde{\beta}) = \sigma^2 b'\tilde{G}\Lambda\tilde{G}'b = \sigma^2 b_1' P_{\Lambda^{-1/2}XK} b_1 \leqq \sigma^2 b_1' P_{\Lambda^{-1/2}X} b_1 = V(p'\hat{\beta})$. Since $\mathcal{M}(X'\Lambda^{-1}XK) \subsetneq \mathcal{M}(X')$, with equality if and only if $R(K) = R(XK) = R(X)$, any estimable functional $p'\beta$ with p outside $\mathcal{M}(X'\Lambda^{-1}XK)$ would be estimated with a bias unless (2.1.4) holds. For such a p, the use of $p'\tilde{\beta}$, which is not globally unbiased would be justified on the grounds of a lower variance if the a priori evidence about the truth of (2.1.4) is sufficiently strong.

In other cases in addition to (2.1.5), one may bring in global unbiasedness of estimable functionals as an additional condition and note that the latter condition is equivalent to requiring

$$XGX = X \qquad (2.1.7)$$

A general solution to (2.1.7) is (see Rao and Mitra [33,p.26])

$$G = X^+_{\Lambda^{-1}I} + V(I-XX^+_{\Lambda^{-1}I}) + (I-X^+_{\Lambda^{-1}I}X)W,$$

W and V being arbitrary. If additionally G satisfies (2.1.5), then

$$(I-X^+_{\Lambda^{-1}I}X)WXK = (I-X^+_{\Lambda^{-1}I}X)K ,$$

so that a general solution to (2.1.5) and (2.1.7) is

$$G = X^+_{\Lambda^{-1}I} + V(I-XX^+_{\Lambda^{-1}I}) + (I-X^+_{\Lambda^{-1}I}X)K(XK)^+_{\Lambda^{-1}I}$$

$$+ (I-X^+_{\Lambda^{-1}I}X)Z(I-XK(XK)^+_{\Lambda^{-1}I}) \qquad (2.1.8)$$

with V and Z arbitrary. Observe that if G satisfies (2.1.5) and (2.1.7), then so does $GP_{X,\Lambda^{-1}}$. Moreover $E(GY) = E(GP_{X,\Lambda^{-1}}Y)$ and $D(GY)-D(GP_{X,\Lambda^{-1}}Y)$ is n.n.d. This shows that in arriving at an optimal choice of G in the class (2.1.8), one may with no loss of generality restrict oneself to the subclass

$$G = X^{+}_{\Lambda^{-1}I} + (I-X^{+}_{\Lambda^{-1}I}X)K(XK)^{+}_{\Lambda^{-1}I} + (I-X^{+}_{\Lambda^{-1}I}X)Z\,(I-XK(XK)^{+}_{\Lambda^{-1}I})P_{X,\Lambda^{-1}} \tag{2.1.9}$$

Let

$$G_1 = X^{+}_{\Lambda^{-1}I},\ G_2 = (I-X^{+}_{\Lambda^{-1}I}X)K(XK)^{+}_{\Lambda^{-1}I},$$

$$G_3 = (I-X^{+}_{\Lambda^{-1}I}X)Z\,(I-XK(XK)^{+}_{\Lambda^{-1}I})P_{X,\Lambda^{-1}},\ G_o = G_1 + G_2,$$

and let $\mathcal{G}$ denote the entire class of matrices determined by (2.1.9) through an arbitrary choice of the matrix Z. Then we have the following:

Theorem 2.1.2

$$\operatorname{tr} D(G_oY) \leq \operatorname{tr} D(GY)\ \forall\, G \in \mathcal{G} \tag{2.1.10}$$

with equality if and only if $G = G_o$.

Proof : The theorem follows once it is noticed that $G_3\Lambda G_2' = 0$ and $G_1'G_3 = 0$.

Since for an arbitrary g-inverse G of X, $GP_{X,\Lambda^{-1}}$ is a reflexive Λ^{-1} least square generalized inverse of X and every reflexive least square g-inverse $X^{-}_{\ell r(\Lambda^{-1})}$ of X can be so determined, the class $\mathcal{G}$ consists precisely of such inverses $X^{-}_{\ell r(\Lambda^{-1})}$ which in addition satisfy (2.1.5) or equivalently

$$\mathcal{M}(K) \subset \mathcal{M}(X^{-}_{\ell r(\Lambda^{-1})}X) = \mathcal{M}(X^{-}_{\ell r(\Lambda^{-1})}) \tag{2.1.11}$$

G_o is the unique member of this class for which

$$\operatorname{tr} D(GY) = \operatorname{tr} G\Lambda G' = \operatorname{tr} \Lambda GG' \tag{2.1.12}$$

is a minimum.

For a matrix $X \in R^{n\times m}$ and p.d. matrices $N \in R^{n\times n}$, $M \in R^{m\times m}$, the Moore Penrose inverse X^+_{NM} is the unique N least square inverse of X such that if G is any other such inverse then

$$G'MG - (X^+_{NM})' MX^+_{NM} \tag{2.1.13}$$

is n.n.d. The uniqueness achieved above is through a different principle that incorporates the principle of conditional unbiasedness due to Plackett and Scheffe. Accordingly we propose to call G_oY the optimal Plackett-Sheffe estimator and G_o the Plackett-Scheffe inverse of X.

Observe that

$$G_o = \tilde{G} + X^+_{\Lambda^{-1}I} [I-XK(XK)^+_{\Lambda^{-1}I}] \tag{2.1.14}$$

and

$$D(G_oY) - D(\tilde{G}Y) = D[X^+_{\Lambda^{-1}I}(I-XK(XK)^+_{\Lambda^{-1}I})Y] \tag{2.1.15}$$

is n.n.d., the right hand side of (2.1.15) vanishing if and only if $R(XK) = R(X)$ when $G_o = \tilde{G}$. This gives us an idea about the loss of accuracy that results, when on account of lack of conviction on the truth of (2.1.4) one insists on global unbiasedness of estimable functionals as an additional condition to be fulfilled in addition to conditional unbisedness as enunciated in (2.1.5).

(iii) Hyperestimators.

Consider the expectational relation

$$X\beta = \mu \quad (=E(Y)) \tag{2.1.16}$$

in the Gauss-Markov model $(Y, X\beta, \sigma^2\Lambda)$. When $R(X)=r< m$, β does not have an unbiased estimator.We may, nevertheless look for an unbiased estimator of the set of points (2.1.16). Thus we try to estimate the entire hyperplane (2.1.16) rather than the single point (the true parameter point) which belongs to this hyperplane. This approach is due to Bunke and Bunke [8], Bjerhammar [4, Chapter 10], [5] and Sjoberg [36]. The name hyperestimator was suggested by Bjerhammar.

The general solution to equation (2.1.16) is given by

$$\beta = X^{-}\mu + (I-X^{-}X)z \tag{2.1.17}$$

where X^{-} is an arbitrary but fixed g-inverse of X and $z \in R^{m}$ is arbitrary. As z varies over R^{m}, (2.1.17) gives the set of points we propose to estimate. Since $(I-X^{-}X)z$ is parameter free the problem thus reduces to estimating a single point $X^{-}\mu$.

$X^{-}\mu$ is estimable, and its BLUE is given by $X^{-}XX^{-}_{\ell(\Lambda^{-1})}Y$ with dispersion matrix $\sigma^2 D$. Observe that $X^{-}XX^{-}_{\ell(\Lambda^{-1})} \in \{X^{-}_{\ell r(\Lambda^{-1})}\}$ and that

$$X^{-}XX^{-}_{\ell(\Lambda^{-1})}\mu = X^{-}\mu \tag{2.1.18}$$

Without any loss of generality we can assume that the single point $X^{-}\mu$ that we shall be estimating is defined in terms of a reflexive least squares inverse of X. By theorem 3.2.2 of Rao and Mitra [33], such a g-inverse always has the representation

$$(X'\Lambda^{-1}X)^{-}\ X'\Lambda^{-1} \tag{2.1.19}$$

for some g-inverse $(X'\Lambda^{-1}X)^{-}$ of $(X'\Lambda^{-1}X)$ and the problem reduces to choosing a g-inverse $(X'\Lambda^{-1}X)^{-}$ of $(X'\Lambda^{-1}X)$ suitably. Noting that

$$D = (X'\Lambda^{-1}X)^{-}\ X'\Lambda^{-1}X[(X'\Lambda^{-1}X)^{-}]' \tag{2.1.20}$$

we establish the following theorem:

<u>Theorem 2.1.3</u> tr Δ D with Δ p.d. in $R^{m\times m}$ is minimized when $(X'\Lambda^{-1}X)^{-}$ in (2.1.20) is chosen as $(X'\Lambda^{-1}X)^{-}_{m(\Delta)}$.

Proof Let T be a nonsingular matrix such that $\Delta = T'T$ and

$X'\Lambda^{-1}X = T'\begin{pmatrix} A & 0 \\ 0 & 0 \end{pmatrix}T$, where $A \in R^{r\times r}$ is diagonal p.d. Any g-inverse of $X'\Lambda^{-1}X$ can be expressed as $(X'\Lambda^{-1}X)^{-} = T^{-1}\begin{pmatrix} A^{-1} & E \\ F & G \end{pmatrix}(T')^{-1}$ for arbitrary E, F, G. $\operatorname{tr} \Delta D = \operatorname{tr} A^{-1} + \operatorname{tr} F A F'$ is minimized when $F = 0$, and hence the optimal choice of $(X'\Lambda^{-1}X)^{-}$ is $T^{-1}\begin{pmatrix} A^{-1} & E \\ 0 & G \end{pmatrix}(T')^{-1} = (X'\Lambda^{-1}X)^{-}_{m(\Delta)}$. Observe that the same choice is also optimal in the sense of minimizing $\operatorname{tr} D'\Delta D$.

Thus the optimal hyperestimator is

$$(X'\Lambda^{-1}X)^{-}_{m(\Delta)} X'\Lambda^{-1}Y + \alpha = X^{+}_{\Lambda^{-1}\Delta} Y + \alpha \tag{2.1.21}$$

where α is an arbitrary vector in $\mathcal{N}(X)$.

2. A bounded parameter space for β

In this section we shall consider the case where the parameter space Ω_β of β is given by

$$\Omega_\beta = \{\beta \in R^m : \beta' H \beta \le \delta^2\} \tag{2.2.1}$$

In this setup the BLIMBE we introduced in section 2 has a more natural interpretation.

(i) BLIMBE revisited.

If H is positive semidefinite, we note that an estimator $b'Y$ of $p'\beta$ has a finite maximum bias over Ω_β if and only if

$$X'b - p \in \mathcal{M}(H) \tag{2.2.2}$$

Thus the parametric functional $p'\beta$ will admit an estimator with a finite maximum bias over Ω_β if and only if

$$p \in \mathcal{M}(X' : H) \tag{2.2.3}$$

Further if $p = X'b + Ha$, the maximum bias of the estimator $b'Y$ on Ω_β, is given by $\delta\sqrt{a'Ha}$. Hence determining a minimax bias estimator for $p'\beta$ requires finding a minimum H_e seminorm solution of the consistent equation

$$(X':H) \begin{pmatrix} b \\ a \end{pmatrix} = p \tag{2.2.4}$$

where $H_e = \text{diag}\,(0,H)$. The following theorem establishes the nature of the solutions b in terms of minimum H_e seminorm solutions of (2.2.4).

<u>Theorem 2.2.1</u> $\begin{pmatrix} G_1 \\ G_2 \end{pmatrix}$ is $(X':H)^-_{m(H_e)}$ if and only if G_1' is $X^-_{m(H)}$.

<u>Proof</u> $\begin{pmatrix} G_1 \\ G_2 \end{pmatrix} = (X':H)^-_{m(H_e)}$ is equivalent to

$$\begin{aligned} X'G_1X' + HG_2X' &= X' \\ X'G_1H + HG_2H &= H \end{aligned} \tag{2.2.5}$$

and

$$H_e \begin{pmatrix} G_1 \\ G_2 \end{pmatrix} (X':H) = \begin{pmatrix} 0 & 0 \\ HG_2X' & HG_2H \end{pmatrix} \text{is symmetric} \tag{2.2.6}$$

From (2.2.6) $HG_2X' = 0$ and HG_2H is symmetric, and hence from (2.2.5) $X'G_1X' = X'$ and $X'G_1H$ is symmetric. Thus $G_1' = X^-_{m(H)}$.

Conversely if $G_1' = X^-_{m(H)}$, then choosing $G_2 = H^-(I-X'G_1)$, it is easily seen that G_1, G_2 satisfies (2.2.5) and (2.2.6) and hence $\begin{pmatrix} G_1 \\ G_2 \end{pmatrix}$ is one choice of $(X':H)^-_{m(H_e)}$.

Hence if $b'Y$ is a minimax bias estimator of $p'\beta$, $b = (X^-_{m(H)})'p = (X')^-_{\ell(H^{-1})}p$ when H is p.d. In this particular case the minimum variance minimax bias estimator is given by $b'Y$ where

$$b = (X')^+_{H^{-1}\Lambda}\, p \tag{2.2.7}$$

which indeed leads to BLIMBE as introduced in section 2, if H^{-1} is used instead of M.

When H is singular it was shown in Rao and Mitra [32] that if M denotes a n.n.d. g-inverse of H + X'X, then

$$\{(X')^-_{\ell(M)}\} \subset \{[X^-_{m(H)}]'\} \qquad (2.2.8)$$

Nonnegative definiteness of M is not essential but will nevertheless be preferred to avoid conceptual difficulties that may otherwise arise in using M to define a seminorm. Generally the set inclusion in (2.2.8) will be a 'proper' one. So the class of minimax bias estimators obtained through arbitrary choice of a M semileast squares inverse of X' cannot be expected to exhaust all minimax bias estimators. Hence the estimator $p'[(X')^+_{M\Lambda}]'Y$ will have the minimum variance property in a smaller subclass of minimax bias estimators. In the present case however, since the general solution to a minimum seminorm g-inverse $X^-_{m(H)}$ of X is given by (see Rao and Mitra [32]),

$$G + W(I-XG) + (I-GX)V \qquad (2.2.9)$$

where G is any particular $X^-_{m(H)}$, W is arbitrary, and V is an arbitrary solution of H(I-GX)V = 0, it is seen that for $p \in \mathcal{M}(X':H)$,

$$\{[X^-_{m(H)}]'p\} = \{(X')^-_{\ell(M)}p\} \qquad (2.2.10)$$

Hence $p'[(X')^+_{M\Lambda}]'Y$ is the minimum variance minimax bias estimator of $p'\beta$.

Finally, let us determine the class of models which allow for a finite maximum bias (over Ω_β) estimator of the type p'GY for $p'\beta$ whatever $p \in R^m$. Defining

$$K = I-HH^+ \qquad (2.2.11)$$

it is seen that for p'GY to estimate $p'\beta$ with a finite maximum bias, K(X'G'p-p) = 0, and since this should hold for all p in R^m,

$$GXK = K \tag{2.2.12}$$

which is the same as (2.1.5). Thus we arrive at the interesting fact that for $p'GY$ to have finite maximum bias for $p'\beta$ for all $p \in R^m$ the model must satisfy the condition

$$R(XK) = R(K) \tag{2.2.13}$$

and thus in such situations the estimator $p'GY$ is conditionally unbiased for $p'\beta$. Since $V(p'GY) = \sigma^2 p'G\Lambda G'p$, by theorem 2.1.1 the minimum variance estimator of finite maximum bias is $p'\tilde{G}Y$, while the minimum tr $D(GY)$ finite maximum bias estimator, which in addition estimates estimable functionals without bias is an optimal Plackett-Scheffe estimator.

(ii) Minimax estimator

Let us assume that the parameter space for (β, σ^2) is $\Omega_\beta \times \Omega_\sigma = \Omega$, where Ω_β is defined in (2.2.1) and Ω_σ is a subset of the positive half of the real line with a known finite maximal element σ_u^2. The estimator $b_o'Y$ of $p'\beta$ is called a minimax (linear) estimator (Rao [31]) if

$$\max_{\beta, \sigma^2 \in \Omega} E(b_o'Y - p'\beta)^2 = \min_{b \in R^n} \max_{\beta, \sigma^2 \in \Omega} E(b'Y - p'\beta)^2 \tag{2.2.14}$$

As in the earlier section we observe that $p'\beta$ will admit an estimator with a finite maximum mean square error if and only if $p \in \mathcal{M}(X':H)$, and that if $p = X'b + Ha$,

$$\max_{\beta, \sigma^2 \in \Omega} E(b'Y - p'\beta)^2 = \sigma_u^2 \, b'\Lambda b + \delta^2 a'Ha = \delta^2(\theta^2 b'\Lambda b + a'Ha) \tag{2.2.15}$$

where

$$\theta^2 = \sigma_u^2 / \delta^2 \tag{2.2.16}$$

Hence if $\begin{pmatrix} G_1 \\ G_2 \end{pmatrix}$ is a minimum diag $(\theta^2\Lambda, H)$ seminorm g-inverse f $(X':H)$, $b_o'Y = p'G_1'Y$ is a minimax estimator of $p'\theta$.

Let us denote

$$F = \begin{bmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{bmatrix} = \begin{bmatrix} \theta^2\Lambda + XX' & XH \\ HX' & H+H^2 \end{bmatrix}^{-} \qquad (2.2.17)$$

ence using formula (3.1.6) of Rao and Mitra [33], one choice of G_1 is

$$G_1 = (F_{11}X + F_{12}H)[X'F_{11}X + X'F_{12}H + HF_{21}X + HF_{22}H]^{-} \qquad (2.2.18)$$

If F is chosen to be n.n.d. then G_1 is one choice of $(A_L^{\dagger})'$ as defined by Mitra [23], where $L = [\text{diag}(I:H)]\, F[\text{diag}(I:H)]$. It is easily seen that when Λ is p.d. one may use formula (3.1.7) of Rao and Mitra [33] to obtain $b_o'Y = p'G_1'Y = p'(X'\Lambda^{-1}X + \theta^2 H)^{-}\, X'\Lambda^{-1}Y = p'X^{\dagger}_{\Lambda^{-1}\oplus\theta^2 H}\, Y$. When H is also p.d. $(X'\Lambda^{-1}X + \theta^2 H)^{-}$ may be replaced by $(X'\Lambda^{-1}X + \theta^2 H)^{-1}$ to obtain the expression for the minimax estimator given by Rao [31]. If K denotes a g-inverse of the matrix $X'F_{11}X + X'F_{12}H + HF_{21}X + HF_{22}H$, the maximum mean square error for the minimax estimator (the L.H.S. of (2.2.14)) is given by

$$\delta^2 p'[K-I]p \qquad (2.2.19)$$

Observe that $b_o'Y$ is also minimax in the wider class of estimators of the type $b'Y + c$. Since $p \in \mathcal{M}(X':H)$, if $p = X'b + Ha$, the square of the bias of $b'Y + c$ is

$$(a'H\beta)^2 + c^2 - 2c(a'H\beta) \leq (a'H\beta)^2 + c^2 + 2|c|\ |a'H\beta|$$

$$\leq\ (\delta\sqrt{a'Ha} + |c|)^2\ .$$

Thus the optimal choice for c is 0 and hence $b_o'Y$ is minimax in the wider class of nonhomogeneous linear estimators. This remark obviously applies to the minimum variance minimax bias estimator (BLIMBE) also.

We noted in Chapter 1 that there is no loss of generality in considering ellipsoids of the form (1.3) with centre at the origin. If our actual model is $(Y, X\beta_*, \sigma^2\Lambda)$ with $(\beta_*-\alpha)'H\ (\beta_*-\alpha) \leq \delta^2$, then the only modification is that the minimax (BLIMBE) estimator of $p'\beta_*$ is $p'\hat{\beta} + p'\alpha$, where $p'\hat{\beta}$ is the minimax (BLIMBE) estimator for $p'\beta$ in $(Y-X\alpha,\ X\beta, \sigma^2\Lambda)$ with $\beta \in \Omega_\beta$ as in (1.3).

For simultaneous estimation of several parametric functionals, say $P\beta$ (P being a matrix), one defines the minimax linear estimator (MILE) as $C_*Y + d_*$, where

$$\underset{\beta,\sigma^2\epsilon\Omega}{\mathrm{Max}}\ E(C_*Y+d_*-P\beta)'\Lambda(C_*Y+d_*-P\beta)$$

$$= \underset{C,d}{\mathrm{Min}}\ \underset{\beta,\sigma^2\epsilon\Omega}{\mathrm{Max}}\ E(CY+d-P\beta)'\Lambda(CY+d-P\beta) \qquad (2.2.20)$$

Λ being a given n.n.d. matrix. In this definition if one restricts the class to estimators of the form CY then one obtains the minimax homogeneous linear estimator (MIHLE).

$$\beta^{(m)} = [(X')^{+}_{\theta^{-2}F^{-1} \oplus \Lambda}]'Y \quad \text{is MILE of } \beta$$

when $A = pp'$, as was noted by Kuks [19]. Considering the matrix loss function $(CY-\beta)(CY-\beta)'$ Bunke [9] showed that $\beta^{(m)}$ is minimax. Lauter [21] showed that when $R(X) = m$ and $H = \Lambda = I$, the shrunken least square estimator

$$\frac{1}{1 + \theta^2 \text{ tr } (X'\Lambda^{-1}X)^{-1}} (X'\Lambda^{-1}X)^{-1} X'\Lambda^{-1}Y \tag{2.2.21}$$

is the MILE. But his expression for the more general situation cannot be used due to the absence of a computational algorithm. However for certain situations there is an iterative procedure for computing MILE suggested by Kuks and Olman [20].

We shall not consider this problem here except for establishing a somewhat disconcerting result that in many important situations the MIHLE is 0. In the following theorem we shall consider (2.2.20) with $d = 0$, $P = I$ and $A = I$, or in other words the loss function is $(CY-\beta)'(CY-\beta)$.

<u>Theorem 2.2.2</u> If $H = I$ and $R(X) = r < m$, then the minimax homogeneous linear estimator of β is 0.

<u>Proof</u> Observe that

$$\max_{\beta'\beta \leq \delta^2, \sigma^2 \epsilon \Omega_\sigma} E(CY-\beta)'(CY-\beta) = \max_{\beta'\beta \leq \delta^2, \sigma^2 \epsilon \Omega_\sigma} [\sigma^2 \text{ tr } CAC' + \beta'(CX-I)'(CX-I)\beta]$$

$$= \sigma_u^2 \text{ tr } CAC' + \delta^2 e(C)$$

where $e(C)$ is the largest eigenvalue of $(CX-I)'(CX-I)$

$$e(C) = \max_{\beta \epsilon R^m} \frac{\beta'(CX-I)'(CX-I)\beta}{\beta'\beta}$$

$$\geq \frac{\beta_o'(CX-I)'(CX-I)\beta_o}{\beta_o'\beta_o} = 1$$

where β_o is a non null vector in $\mathcal{N}(X) \cap \Omega$.

Hence

$$\underset{\beta'\beta\leq\delta^2,\sigma^2\epsilon\Omega_\sigma}{\text{Max}} \; E(CY-\beta)'(CY-\beta) \;\geq\; \sigma_u^2 \text{ tr } CAC' + \delta^2 \geq \delta^2 \qquad (2.2.22)$$

The theorem is established once it is noted that equality holds in (2.2.22) if and only if C=0.

Working on the same lines, it is easy to see that theorem 2.2.2 can be generalized at least to the situation where $A = (I-P_{X'})A(I-P_{X'})$ is n.n.d. In this case the class of MIHLE is $\{C_*Y: \mathcal{M}(C_*) \subset \mathcal{N}(A)\}$ and 0 is one choice. This in particular (choosing $A = pp'$) shows that the minimax estimator for $p'\beta$ where $Xp = 0$ is 0. This result we shall encounter once more in the next section.

Let us now consider a more general risk function

$$E(CY-P\beta)'A(CY-P\beta) = \sigma^2 \text{ tr } CAC'A + \beta'(CX-P)'A(CX-P)\beta \qquad (2.2.23)$$

Write $A = E'E$, $F = ECX-EP$, $Z = EC$, $L = EP$.
If $E \in R^{s\times t}$, then partition,

$$F' = (f_1, f_2, \ldots, f_s),\; Z'=(z_1,z_2,\ldots,z_s), L' = (\ell_1,\ell_2,\ldots,\ell_s),$$

and observe that (2.2.23) can be written as

$$\sigma^2 \sum_{i=1}^{s} z_i' \, \Lambda \, z_i + \sum_{i=1}^{s} (f_i'\beta)^2 \qquad (2.2.24)$$

For simplicity let us restrict ourselves to the situation where H is p.d. By Cauchy Schwarz inequality applied to each term of the second sum in (2.2.24),

$$\underset{\beta,\sigma^2\epsilon\Omega}{\text{Max}} \; E(CY-P\beta)'A(CY-P\beta) \leq \sigma_u^2 \sum_{i=1}^{s} z_i'\Lambda z_i + \delta^2 \sum_{i=1}^{s} f_i' \, H^{-1} f_i \,(2.2.25)$$

The inequality is strict except in certain special cases, so that the R.H.S. of (2.2.25), viz.

$$\sigma_u^2 [\sum_{i=1}^{s} z_i'\Lambda z_i + \frac{1}{\theta^2} \sum_{i=1}^{s} (X'z_i-\ell_i)'H^{-1}(X'z_i-\ell_i)] \qquad (2.2.26)$$

can be only an approximation to the required maximum risk over Ω.

Minimizing (2.2.26) we may obtain an approximate MIHLE. By Mitra [23], the minimum of (2.2.26) is attained when

$$z_i = (X')^{\dagger}_{\frac{1}{\theta^2}H^{-1} \oplus \Lambda} \ell_i$$

i.e.

$$C_*'E' = (X')^{\dagger}_{\frac{1}{\theta^2}H^{-1} \oplus \Lambda} P'E'$$

i.e.

$$E[C_*-P[(X')^{\dagger}_{\frac{1}{\theta^2}H^{-1} \oplus \Lambda}]'] = 0 \qquad (2.2.27)$$

Hence an approximate MIHLE for $P\beta$ is given by (2.2.27) and one choice is

$$P[(X')^{\dagger}_{\frac{1}{\theta^2}H^{-1} \oplus \Lambda}]'Y = PX^{\dagger}_{\Lambda^{-1} \oplus \theta^2 H} Y \qquad (2.2.28)$$

In situations where theorem 2.2.2 apply, the minimax estimator is zero. However the approximate minimax estimator $\hat{\beta}_1 = X^{\dagger}_{\Lambda^{-1}\oplus\theta^2 H} Y$ could still be an useful estimator for β for many purposes especially since it gives the MILE for parametric functionals $p'\beta$. In the next section we shall compare the relative performances of various estimators for $p'\beta$, irrespective of the choice of p. Writing an estimator of $p'\beta$ as $p'\hat{\beta}$, our criteria shall be $[E(\hat{\beta})-\beta]'H[E(\hat{\beta})-\beta]$ and $E(\hat{\beta}-\beta)'H(\hat{\beta}-\beta)$. Obviously for minimax, $\hat{\beta}_1$ is more appropriate. In Chapter 3 it will be shown that $\hat{\beta}_1$ is also an optimal Bayes estimator of β when it has a prior distribution with $E(\beta\beta') = \delta^2H^{-1}$. When H = I, $\hat{\beta}_1$ is a ridge type estimator

(Hoerl and Kennard [17], [18]).

In conclusion observe that when H is p.d. by Theorem 5.1 of Mitra [23]

$$\lim_{\delta^2 \to \infty} p' X^{+}_{\Lambda^{-1} \oplus \theta^2 H} Y = p' X^{+}_{\Lambda^{-1} H} Y ;$$

thus BLIMBE can be looked upon as the MILE of $p'\beta$ when Ω_β is unbounded in every direction. If $p \in \mathcal{M}(X')$, BLIMBE is infact the BLUE of $p'\beta$, and does not depend on H, a conclusion which is no longer valid if $p \notin \mathcal{M}(X')$. This observation is due to Mitra [24], who uses this as a pointer to the futility of estimating nonestimable functionals in unbounded parameter spaces.

(iii) Restricted least squares estimator (maximum likelihood estimator)

If $Y \sim N_n(X\beta, \sigma^2\Lambda)$ and Λ is p.d. then the maximum likelihood estimate $\beta_{m\ell}$ minimizes

$$G(\beta) = (Y-X\beta)'\Lambda^{-1}(Y-X\beta) \tag{2.2.29}$$

subject to the condition $\beta'H\beta \leq \delta^2$. Even without the distributional assumption the same method yields a least square estimate subject to the quadratic constraint $\beta'H\beta \leq \delta^2$.

This and similar optimization problems were considered by Balakrishnan [2], Forsythe and Golub [15]. Let us write $\beta_o = X^{+}_{\Lambda^{-1}H} Y$, the minimum H seminorm Λ^{-1} least square solution of $X\beta = Y$.

If $\hat{\beta}_o'H\hat{\beta}_o \leq \delta^2$ *then* $\hat{\beta}_o$ *is one choice of* $\hat{\beta}_{m\ell}$ *and infact any solution of the* normal equation

$$X'\Lambda^{-1}X\beta = X'\Lambda^{-1}Y \tag{2.2.30}$$

which belongs to Ω_β could serve as the ML estimate. To avoid confusion in such cases we shall choose and fix $\hat{\beta}_{m\ell} = \beta_o$. If $\hat{\beta}_o'H\beta_o > \delta^2$ the ML estimate $\hat{\beta}_{m\ell}$ will obviously lie on the surface of the ellipsoid Ω_β. To compute $\hat{\beta}_{m\ell}$ in such cases one has to minimize $G(\beta)$ subject to the condition $\beta' H\beta = \delta^2$. By Lagrange's method of constrained optimization, the solution β_λ will satisfy the equation

$$(X'\Lambda^{-1}X + \lambda H)\beta = X'\Lambda^{-1}Y \qquad (2.2.31)$$

where λ is a root of the equation

$$F(\lambda) = Y'\Lambda^{-1}X(X'\Lambda^{-1}X+\lambda H)^{-} H(X'\Lambda^{-1}X+\lambda H)^{-}X'\Lambda^{-1}Y = \delta^2 \qquad (2.2.32)$$

Let W be a nonsingular matrix such that

$$H = W' \begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix} W \qquad (2.2.33)$$

and

$$X'\Lambda^{-1}X = W' \begin{pmatrix} D_1 & 0 \\ 0 & D_2 \end{pmatrix} W \qquad (2.2.34)$$

where $D_1 = \text{diag}\ (d_1, d_2, \ldots, d_k)$, $D_2 = \text{diag}(d_{k+1}, d_{k+2}, \ldots, d_m)$. It is well known that such a matrix W exists for every pair of n.n.d. matrices (see e.g. Rao and Mitra [33,p.122]). Assume without any loss of generality that $d_1 \geq d_2 \geq \ldots \geq d_u > 0$ and $d_{u+1} = \ldots = d_k = 0$. Put $W\hat{\beta}_o = \gamma = (\gamma_1, \ldots, \gamma_m)'$. Then

$$F(\lambda) = \hat{\beta}_o X'\Lambda^{-1}X(X'\Lambda^{-1}X + \lambda H)^{-} H(X'\Lambda^{-1}X + \lambda H)^{-}X'\Lambda^{-1}X\hat{\beta}_o$$

$$= \sum_{i=1}^{u} \frac{d_i^2\ \gamma_i^2}{(d_i + \lambda)^2} \qquad (2.2.35)$$

and $F(0) = \sum_{i=1}^{u} \gamma_i^2 = \hat{\beta}_o' H\hat{\beta}_o > \delta^2$. Clearly $F(\lambda)$ is monotonically decreasing in $(0,\infty)$ and $F(\lambda) \to 0$ as $\lambda \to \infty$. Thus there exists a unique positive root λ_o of the equation

$$F(\lambda) = \delta^2 \qquad (2.2.36)$$

If $\hat{\beta}_\lambda$ satisfies (2.2.31) with $\lambda \geq 0$, clearly $\forall \beta \varepsilon R^m$,

$$(Y-X\beta)'\Lambda^{-1}(Y-X\beta)+\lambda\beta'H\beta = (Y-X\hat{\beta}_\lambda)'\Lambda^{-1}(Y-X\hat{\beta}_\lambda)+\lambda\hat{\beta}_\lambda'H\hat{\beta}_\lambda+ (\beta-\hat{\beta}_\lambda)'[X'\Lambda^{-1}X+\lambda H](\beta-\beta_\lambda)$$

$$\geq (Y-X\hat{\beta}_\lambda)'\Lambda^{-1}(Y-X\hat{\beta}_\lambda)+\lambda\hat{\beta}_\lambda'H\hat{\beta}_\lambda \qquad (2.2.37)$$

If β satisfies $\beta'H\beta = \delta^2$,

$$(Y-X\beta)'\Lambda^{-1}(Y-X\beta) \geq (Y-X\hat{\beta}_\lambda)'\Lambda^{-1}(Y-X\hat{\beta}_\lambda) \qquad (2.2.38)$$

This shows that $\hat{\beta}_{\lambda_o}$ in fact gives the required minimum. On the other hand if $\lambda \leq -d_1$, then writing $W(\beta-\beta_\lambda) = \theta = (\theta_1,\ldots,\theta_m)'$,$(\beta-\hat{\beta}_\lambda)'[X'\Lambda^{-1}X+\lambda H](\beta-\beta_\lambda)$ $= \sum_{i=1}^{u} (d_i+\lambda)\theta_i^2 \leq 0$, and hence for such a λ, $(Y-X\beta)'\Lambda^{-1}(Y-X\beta)$ $\leq (Y-X\beta_\lambda)'\Lambda^{-1}(Y-X\hat{\beta}_\lambda)$. Since $F(\lambda)$is monotonically increasing in $(-\infty, -d_1)$, $F(-d_1) >> 0$, and $F(\lambda) \to 0$ as $\lambda \to -\infty$, there exists a unique $\lambda_- < -d_1$ satisfying (2.2.36) and with this λ_-, $\hat{\beta}_{\lambda_-}$ gives the maximum of $G(\beta)$ on the surface of Ω_β. Other real roots, if any, belong to the open interval $(-d_1, -d_u)$ and can similarly be shown to correspond to saddle points of $G(\beta)$ on Ω_β. Note that since $F(\lambda) \to \infty$ as $\lambda \to -d_i$, real roots (if they exist) in $(-d_i, -d_{i+1})$ must be even in number.

Writing explicitly,

$$\hat{\beta}_{\lambda_o} = (X'\Lambda^{-1}X + \lambda_o H)^- X'\Lambda^{-1}Y = X^{\dagger}_{\Lambda^{-1} \oplus \lambda_o H} Y \qquad (2.2.39)$$

Observe that if both $X'\Lambda^{-1}X$ and H are singular matrices then the choice of $\hat{\beta}_{\lambda_o}$ may not be unique in all situations. In such an eventuality some a priori knowledge or a secondary criterion may be used to make a choice. Otherwise one may, for example take the ML estimate as a minimum norm solution of (2.2.31) with λ replaced by λ_o. Put $W\hat{\beta}_{\lambda_o} = \psi = (\underline{\psi}_1' : \underline{\psi}_2')'$, where $\underline{\psi}_1 \varepsilon R^k$, $\underline{\psi}_2 \varepsilon R^{m-k}$, and partition Y accordingly as $(\underline{Y}_1' : \underline{Y}_2')$.

Observe that equations (2.2.31) are equivalent to $(D_1+\lambda_o I)\underline{\psi}_1 = D_1\underline{\gamma}_1$, $D_2\,\underline{\psi}_2 = 0$. Thus the nonuniqueness will arise only in those co-ordinates of $\underline{\psi}_2$ which correspond to a zero diagonal entry in D_2. These could be taken as zero and for this choice the resulting $\hat{\beta}_{\lambda_o}$ is a minimum norm solution of (2.2.31).

The ML estimator is thus given by

$$\hat{\beta}_{m\ell} = \begin{cases} \hat{\beta}_o & \text{if} \quad \hat{\beta}_o' H \hat{\beta}_o \leq \delta^2 \\ \hat{\beta}_{\lambda_o} & \text{if} \quad \hat{\beta}_o' H \hat{\beta}_o > \delta^2 \end{cases} \tag{2.2.40}$$

3. Comparison of rival estimators

The alternative interpretation for BLIMBE given in section 3 is based on the well known fact that an inner product defined on a real (or complex) vector space V induces a dual inner product on the vector space of linear functionals V* of V. For example let $V = R^m$, so that V* is isomorphic to R^m. Let $<\cdot,\cdot>$, $<\cdot,\cdot>_*$ denote inner products in V and V* and $||\cdot||$, $||\cdot||_*$ denote the corresponding norms; further $<x,y> = \underline{x}'H\underline{y}$, for H p.d. where $\underline{x} \in R^m$ are the coordinates of x in terms of a basis If $\alpha_1, \ldots, \alpha_m$ is a basis in $V(=R^m)$ then define a basis $\gamma_1,\ldots,\gamma_m$ in V* by the relation $\gamma_i(\alpha_j) = \delta_{ij}$, the Kronecker symbol. If $x = \sum_{j=1}^m x_j\alpha_j$ and $f = \sum_{i=1}^m f_i\gamma_i$, then the norm of f is

$$||f||_* = \sup_{x \in V} \frac{|f(x)|}{||x||} \tag{2.3.1}$$

However $f(x) = \sum_{i=1}^m f_i\gamma_i\left(\sum_{j=1}^m x_j\,\alpha_j\right) = \sum_{i=1}^m f_i\,x_i = \underline{f}'\underline{x}$, so that (2.3.1) becomes

$$\sup_{\underline{x} \in R^m} \frac{|\underline{f}'\underline{x}|}{\sqrt{(\underline{x}'H\underline{x})}} = \underline{f}'H^{-1}\underline{f}$$

by Cauchy Schwarz inequality. Thus

$$< f,g >_* = (||f+g||_*^2 - ||f-g||_*^2)/4 = \underset{\sim}{f}'H^{-1}\underset{\sim}{g} \qquad (2.3.2)$$

Hence for such a choice of the basis in V^* the dual inner product is induced by H^{-1}, if H induces the inner product in $V(=R^m)$.

Consider now the resolution of the vector space of linear parametric functionals into the vector subspace of estimable functionals and its orthogonal complement under the dual inner product. A parametric functional in the orthogonal complement may be called totally nonestimable. We have seen that $p'\beta$ is estimable if $p \in \mathcal{M}(X')$. The above resolution corresponds to the following resolution of p

$$p = p_e + p_{ne} \quad ,$$

where $p_e = P_{X',H^{-1}}\, p$ and $p_{ne} = (I-P_{X',H^{-1}})p$. Thus $p_e'\beta$ is the estimable part of $p'\beta$ and $p_{ne}'\beta$ is its totally nonestimable part. The above resolution also corresponds to the following resolution of β

$$\beta = \beta_e + \beta_{ne} \quad ,$$

where $\beta_e = P'_{X',H^{-1}}\beta$ and $\beta_{ne} = (I-P'_{X',H^{-1}})\beta$, and this indeed is an orthogonal resolution under the original inner product i.e. $\beta_e'H\beta_{ne} = 0$. Clearly each co-ordinate of β_e is estimable. Hence β_e may be called the estimable part of β. We may call β_{ne} the totally nonestimable part of β; the justification is seen from the fact that if β is totally nonestimable, i.e. $\beta = (I-P'_{X',H^{-1}})\beta$, then every linear functional in Y has identically a zero expectation. The name totally nonestimable given to $p_{ne}'\beta = p'\beta_{ne}$ is therefore like a transferred epithet which seems to be quite appropriate.

The interesting fact to note is that a totally nonestimable parametric functional is estimated by zero by the BLIMBE, MLE or the minimax estimator as given in section 2. Equivalently the result could be stated thus:

Theorem 2.3.1 Write $\hat{\beta}_o$ for $X^+_{\Lambda^{-1}H}Y$ which provides BLIMBE for $p'\beta$, β_1 for $X^+_{\Lambda^{-1}\oplus\theta^2H}Y$ which provides the minimax estimator for $p'\beta$ and let β_2 denote the MLE as defined in (2.2.40). Then

$$(I - P'_{X',H^{-1}})\hat{\beta}_i = 0, \quad i = 0,1,2 \tag{2.3.3}$$

Proof From the representation of the minimum norm least square inverse given in (3.3.11) of Rao and Mitra [33] and from Theorem 3.2 (V) of Mitra [23] it follows that $\hat{\beta}_i \in \mathcal{M}(H^{-1}X')$. Theorem 2.3.1 is an immediate consequence since $(I-P'_{X',H^{-1}})\,H^{-1}X' = 0$.

We now proceed to compare $\hat{\beta}_o$, $\hat{\beta}_1$ and $\hat{\beta}_2$ on the basis of their biases and mean square errors at $\beta \in \Omega_\beta$. Our criterion will be norm (induced by H) of the bias vector and trace $H\Delta$ where Δ is the mean square and product error (m.s.p.e.) matrix. Put

$$B_i = [E(\hat{\beta}_i) - \beta]'\,H[E(\hat{\beta}_i) - \beta] \tag{2.3.4}$$

$$M_i = E(\hat{\beta}_i-\beta)'\,H(\hat{\beta}_i - \beta) \tag{2.3.5}$$

Theorem 2.3.2 gives certain inequalities among the B_i's and M_i's that hold when σ^2 is known ($\theta^2=\sigma^2/\delta^2$ in the expression for $\hat{\beta}_1$).

Theorem 2.3.2

(a) $$B_o \leq \min(B_1, B_2) \tag{2.3.6}$$

(b) $$M_1 \leq M_o \tag{2.3.7}$$

Proof Let $\hat{\beta}$ denote an estimator which is such that

$$P'_{X',H^{-1}}\,\hat{\beta} = \hat{\beta} \tag{2.3.8}$$

then

$$[E(\hat{\beta}) - \beta]' H[E(\hat{\beta}) - \beta] = [E(\hat{\beta}) - \beta]' H[P'_{X',H^{-1}} + (I-P'_{X',H^{-1}})][E(\hat{\beta}) - \beta]$$

$$= [E(P'_{X',H^{-1}}\hat{\beta}) - \beta_e]' H[E(P'_{X',H^{-1}}\hat{\beta}) - \beta_e] + \beta'_{ne} H\beta_{ne}$$

$$= [E(\hat{\beta}) - \beta_e]' H[E(\hat{\beta}) - \beta_e] + \beta'_{ne} H\beta_{ne}$$

$$\geq \beta'_{ne} H\beta_{ne} \qquad (2.3.9)$$

using known properties of such projections, viz. $P_{X',H^{-1}} H[I-P'_{X',H^{-1}}] = 0$ (see e.g. Theorem 5.2.1 of Rao and Mitra [33]) and the facts

$$P'_{X',H^{-1}} E(\hat{\beta}) = E(P'_{X',H^{-1}}\hat{\beta}) = E(\hat{\beta}),\ (I-P'_{X',H^{-1}})E(\hat{\beta}) = E[(I-P'_{X',H^{-1}})\hat{\beta}] = 0.$$

Since β_i satisfies (2.3.8) we have from (2.3.9),

$$B_i \geq \beta'_{ne} H \beta_{ne}$$

However, for $i = 0$, equality holds since

$$E(\hat{\beta}_o) = X^{+}_{\Lambda^{-1}H} X\beta = [X'(X')^{+}_{H^{-1}\Lambda}]'\beta = P'_{X',H^{-1}}\beta = \beta_e \qquad (2.3.10)$$

This establishes (2.3.8).

To prove (b) it can similarly be seen that

$$M_i = E(\hat{\beta}_i - \beta_e)' H(\hat{\beta}_i - \beta_e) + \beta'_{ne} H\beta_{ne} \qquad (2.3.11)$$

By (2.3.10), $M_o = \text{tr } HD(\hat{\beta}_o) + \beta'_{ne} H\beta_{ne}$

Let T be a nonsingular matrix such that $X'\Lambda^{-1}X = T'DT$ and $H = T'T$, where $D = \text{diag}(d_1,\ldots,d_r,0,\ldots,0)$. Let T be partitioned as $T' = (T'_1 : T'_2)$ where $T_1 \in R^{r\times m}$, $T_2 \in R^{(m-r)\times m}$, and $T^{-1} = S = (S'_1 : S'_2)$ where $S_1 \in R^{r\times m}$, $S_2 \in R^{(m-r)\times m}$ Then $H = T'_1T_1 + T'_2T_2$, $\mathcal{M}(T'_1) = \mathcal{M}(X'\Lambda^{-1}X) = \mathcal{M}(X')$, $P_{X',H^{-1}} = T'_1S_1$, $I-P_{X',H^{-1}} = T'_2 S_2$, $T_1S'_1 = I$, $T_2S'_2 = I$, $T_1S'_2 = 0$, $T_2S'_1 = 0$.

$$D(\hat{\beta}_o) = D(P'_{X',H^{-1}}\hat{\beta}_o) = \sigma^2 P'_{X',H^{-1}}(X'\Lambda^{-1}X)^- P_{X',H^{-1}}$$

$$= \sigma^2 S_1'T_1(S_1':S_2')D^- \begin{pmatrix} S_1 \\ S_2 \end{pmatrix} T_1'S_1$$

$$= \sigma^2(S_1' : 0)D^- \begin{pmatrix} S_1 \\ 0 \end{pmatrix} = \sigma^2 T^{-1}D^+(T^{-1})'$$

where $D^+ = \text{diag}(d_1^{-1}, \ldots, d_r^{-1}, 0, \ldots, 0)$.

Hence $$M_o = \sigma^2 \sum_{i=1}^{r} \frac{1}{d_i} + \beta'_{ne} H \beta_{ne}$$

Similarly $$M_1 = \text{tr}\, HD(\hat{\beta}_1) + [E(\hat{\beta}_1) - \beta_e]'H[E(\hat{\beta}_1)-\beta_e] + \beta'_{ne} H\beta_{ne}$$

Observe that $E(\hat{\beta}_1) = X^{\dagger}_{\Lambda^{-1}\oplus\theta^2 H} X\beta = X^{\dagger}_{\Lambda^{-1}\oplus\theta^2 H} X\beta_e$, and writing

$T\beta_e = \underline{n} = \left(\frac{\underline{n}_1}{\underline{n}_2}\right)$, where $\underline{n}_1 \in R^r$, $\underline{n}_2 \in R^{m-r}$, $\underline{n}_2 = T_2\beta_e = T_2P'_{X',H^{-1}}\beta_e =$

$T_2S_1'T_1\beta_e = 0$. Hence

$$[E(\hat{\beta}_1)-\beta_e]'H[E(\hat{\beta}_1)-\beta_e] = \underline{n}'[(D+\theta^2 I)^{-1}D-I]^2\underline{n} = \theta^4 \sum_{i=1}^{r} \frac{n_i^2}{(d_i+\theta^2)^2}$$

Thus

$$M_1 = \sigma^2 \,\text{tr}\,(D+\theta^2 I)^{-1} D(D+\theta^2 I)^{-1} + \theta^4 \sum_{i=1}^{r} \frac{n_i^2}{(d_i+\theta^2)^2} + \beta'_{ne} H\beta_{ne}$$

$$= \sigma^2 \sum_{i=1}^{r} \frac{1}{(d_i+\theta^2)^2}\left(d_i + \frac{\sigma^2}{\delta^4} n_i^2\right) + \beta'_{ne} H\beta_{ne}$$

Since

$$\beta'H\beta = \beta_e'H\beta_e + \beta'_{ne} H\beta_{ne},$$

$$\beta'H\beta \le \delta^2 \Rightarrow \beta_e'H\beta_e \le \delta^2 \Leftrightarrow \underline{n}_1'\underline{n}_1 \le \delta^2 \Rightarrow n_i^2 \le \delta^2,\ i=1,\ldots,r$$

Hence

$$M_1 \le \sigma^2 \sum_{i=1}^{r} \frac{1}{(d_i+\theta^2)^2}(d_i+\theta^2) + \beta'_{ne} H \beta_{ne}$$

$$= \sigma^2 \sum_{i=1}^{r} \frac{1}{(d_i+\theta^2)} + \beta'_{ne} H\beta_{ne}$$

$$\le \sigma^2 \sum_{i=1}^{r} \frac{1}{d_i} + \beta'_{ne} H\beta_{ne} = M_o \text{ since } \theta^2 \ge 0.$$

Since the MLE is nonlinear it is difficult to carry out a similar exercise on $\hat{\beta}_2$. Simulation was therefore resorted to for a comparison of these estimators. For simplicity, Ω_β was taken to be a sphere rather than an ellipsoid, and the dispersion matrix of Y was taken to be $\sigma^2 I$. Also since the nonestimable part affects B_i and M_i values of all the three estimators in the same way (see equations (2.3.9) and (2.3.11)) we shall compare them only on a full rank model (i.e. $R(X) = m$). Further, keeping in view the possibility of a singular value decompostion of X,

$$X = U' \binom{A}{0} V,$$

where $A = \text{diag}(\alpha_1,\ldots,\alpha_m)$, U and V are orthogonal matrices, and noting that a parameter transformation $\beta \to \tau = VB$ replaces a sphere in β by a sphere in τ of equal radius and an orthogonal transformation of the observations $Y \to Z = UY$ keeps the dispersion matrix $(\sigma^2 I)$ invariant, one could without any loss of generality restrict oneself to a simpler model that holds for Z, $(Z,\binom{A}{0}\tau,\sigma^2 I)$, and even omit the last n-m observations as they contribute only to the estimation of σ^2. For the simulation the following canonical model was considered

$$\begin{aligned} E(y_i) &= \alpha_i\, \beta_i \\ \text{Cov}(y_i,y_j) &= \sigma^2 \delta_{ij} \end{aligned} \qquad (2.3.12)$$

α_i given positive numbers, $\sum_{i=1}^{m} \beta_i^2 \leq \delta^2$, δ_{ij} the Kronecker symbol, $i,j = 1,2,\ldots,m$.

In the computations m was chosen to be 4. Four different design matrices were chosen. One had large α_i values, one moderate but $\alpha_i \geq 1$ the third had $\alpha_i \leq 1$ and the last one had α_i on either side of 1. For each design matrix, three points in Ω_β, one near the circumference, one well inside and the third near the origin were chosen. The entire experiment was done with three values of $\sigma, \sigma = 1,2,3$. Since σ was varied,

δ^2 was kept fixed at 17, their ratio only matters. Random normal deviates were obtained from Herman Wold's Random Normal Deviates: Tracts for Computers No. XXV, Cambridge University Press, 1954.

Tables 1,2 and 3 give the B_i and M_i values for $i = 0,1,2$ and $\sigma = 1,2$ and 3. (B_i is called bias and M_i mean square error). Each table corresponds to one particular value of σ. Besides these the frequency of cases where $\lambda = 0$ and $\lambda > 0$ are also tabulated, where to avoid suffixing we denote λ_o of section 2 by λ. Note that when $\lambda = 0$, MLE is same as BLIMBE. For BLIMBE and the minimax estimator, theoretical expressions for B_i and M_i are available. Theoretical values computed from these expressions are shown in parenthesis in addition to values estimated by simulation. A comparison of these two entries will indicate the accuracy attained by simulation.

Table 4 gives a ranking of these procedures from the point of view of bias and also from the point of view of mean square error. Biaswise BLIMBE appear to be distinctly superior to the other two methods. The second place is shared by minimax and MLE about equally frequently. However, interestingly enough one that does better on bias fares poorly on m.s.e. In the whole table there are only four exceptions to this rule. We have already noted that when $\lambda = 0$, MLE is same as BLIMBE. Hence in those cases where $\lambda = 0$ has a large frequency it was to be expected that minimax would have a lower m.s.e. compared to MLE. However there are cases where minimax has a lower m.s.e. compared to MLE even when the frequency at $\lambda = 0$ is small.

These happen to be cases where the model itself does not allow for a precise estimation of parameters either on account of small singular values of the design matrix or large values of σ or both. One could, for example, take $M_o (= \sigma^2 \sum_{i=1}^{m} \frac{1}{\alpha_i^2})$ as a measure of this characterestic

Table 1 : Showing the bias and mean square error of

$(n = m = 4,\ \delta^2 = 17,\ \sigma = 1)$

Sl. No.	β/DR				BLIMBE BIAS	BLIMBE MSE	MINIMAX BIAS	MINIMAX MSE	MLE BIAS	MLE MSE	Frequency of λ at 0	Frequency of λ at >0
1	2	2	2	2	$.0^{3}15$	.1363	$.0^{3}40$	.1359	.0020	.1185	738	262
	4	5	7	10	(0)	(.1329)	$(.0^{4}82)$	(.1323)				
2	1	2	1	2	$.0^{4}91$	.1352	$.0^{3}16$	.1345	$.0^{4}91$	.1352	1000	0
	4	5	7	10	(0)	(.1329)	$(.0^{4}38)$	(.1322)				
3	1	-1	-1	1	$.0^{4}49$	.1337	$.0^{4}76$	.1331	$.0^{4}49$	.1337	1000	0
	4	5	7	10	(0)	(.1329)	$(.0^{4}21)$	(.1322)				
4	2	2	2	2	.0017	1.2522	.0023	1.1614	.1114	.8499	559	441
	1.0	3.0	3.5	5.0	(0)	(1.2327)	(.0126)	(1.1349)				
5	1	2	1	2	$.0^{3}25$	1.2662	.0031	1.1559	$.0^{3}43$	1.1867	950	50
	1.0	3.0	3.5	5.0	(0)	(1.2327)	(.0033)	(1.1256)				
6	1	-1	-1	1	.0011	1.2757	.0076	1.1671	.0012	1.2724	995	5
	1.0	3.0	3.5	5.0	(0)	(1.2327)	(.0032)	(1.1255)				
7	2	2	2	2	.0082	23.4844	1.1405	10.6609	1.9740	8.8110	207	793
	0.25	0.50	0.75	1.00	(0)	(22.7777)	(1.1336)	(10.3501)				
8	1	2	1	2	.0391	22.7410	.26[illegible]6	9.5188	.4959	8.2916	282	718
	0.25	0.50	0.75	1.00	(0)	(22.7777)	(.4015)	(9.6180)				
9	1	-1	-1	1	.0040	23.1457	.2908	9.5572	.2618	10.7166	445	555
	0.25	0.50	0.75	1.00	(0)	(22.7777)	(.2834)	(9.4999)				
10	2	2	2	2	.0258	19.1109	1.1617	9.2224	1.8065	6.7288	250	750
	0.25	0.75	1.00	5.00	(0)	(18.8177)	(.9885)	(7.6235)				
11	1	2	1	2	.0047	19.3796	.2719	7.1004	.3333	6.6350	389	611
	0.25	0.75	1.00	5.00	(0)	(18.8177)	(.2740)	(6.9090)				
12	1	-1	-1	1	.0177	19.3637	.3211	7.0908	.2695	9.0579	547	453
	0.25	0.75	1.00	5.00	(0)	(18.8177)	(.2471)	(6.8821)				

Note Theoretical values are shown in parenthesis. Other values are based on simulation using a sample of 1000 from $N(X\beta,\sigma^2 I)$. The first column with the heading β/DR has two entries in each cell. The upper entries are the values of β_1, β_2, β_3 and β_4. The lower entries are the singular values of the X matrix.

Table 2 : Showing the bias and mean square error of BLIMBE, MINIMAX and MLE

$(n = m = 4,\ \delta^2 = 17,\ \sigma = 2)$

Sl. No.	β/DR				BLIMBE BIAS	BLIMBE MSE	MINIMAX BIAS	MINIMAX MSE	MLE BIAS	MLE MSE	Frequency of λ at 0	Frequency of λ at >0
1	2	2	2	2	$.0^{3}41$	.4971	$.0^{3}71$	.4872	.0142	.3877	572	428
	4	5	7	10	(0)	(.5316)	(.0013)	(.5218)				
2	1	2	1	2	$.0^{3}36$	.5405	.0012	.5299	$.0^{3}38$	.5379	994	6
	4	5	7	10	(0)	(.5316)	$(.0^{3}60)$	(.5211)				
3	1	-1	-1	1	$.0^{3}19$	.5348	$.0^{3}56$	.5240	$.0^{3}19$	.5348	1000	0
	4	5	7	10	(0)	(.5316)	$(.0^{3}33)$	(.5208)				
4	2	2	2	2	.0069	5.0090	.2035	3.7745	.5756	3.0010	432	568
	1.0	3.0	3.5	5.0	(0)	(4.9310)	(.1495)	(3.6643)				
5	1	2	1	2	.0010	5.0644	.0387	3.6519	.0367	3.4938	720	280
	1.0	3.0	3.5	5.0	(0)	(4.9310)	(.0396)	(3.5543)				
6	1	-1	-1	1	.0047	5.0646	.0616	3.6608	.0216	4.3777	894	106
	1.0	3.0	3.5	5.0	(0)	(4.9310)	(.0374)	(3.5521)				
7	2	2	2	2	.0327	93.9208	4.0281	17.7834	4.7617	16.4960	48	952
	0.25	0.50	0.75	1.00	(0)	(91.1109)	(3.9305)	(17.1521)				
8	1	2	1	2	.1514	90.9601	1.4377	14.8443	1.9268	14.2518	73	927
	0.25	0.50	0.75	1.00	(0)	(91.1109)	(1.7967)	(15.0183)				
9	1	-1	-1	1	.0161	92.5692	1.0389	14.2610	.9946	15.5000	112	888
	0.25	0.50	0.75	1.00	(0)	(91.1109)	(.9826)	(14.2042)				
10	2	2	2	2	.1031	76.4459	3.2608	12.6214	3.7257	12.2599	104	896
	0.25	0.75	1.00	5.00	(0)	(75.2709)	(2.9906)	(12.1231)				
11	1	2	1	2	.0189	77.5149	1.0684	10.4787	1.1729	10.9519	136	864
	0.25	0.75	1.00	5.00	(0)	(75.2709)	(1.0089)	(10.1414)				
12	1	-1	-1	1	.0709	77.4561	.8565	10.0552	.7832	13.7544	185	815
	0.25	0.75	1.00	5.00	(0)	(75.2709)	(.7476)	(9.8802)				

Note : Theoretical values are shown in parenthesis. Other values are based on simulation using a sample of 1000 from $N(X\beta, \sigma^2 I)$. The first column with the heading β/DR has two entries in each cell. The upper entries are the values of β_1, β_2, β_3 and β_4. The lower entries are the singular values of the X matrix.

Table 3 : Showing the bias and mean square error of BLIMBE, MINIMAX and MLE

($n = m = 4$, $\delta^2 = 17$, $\sigma = 3$)

Sl. No.	β/DR				BLIMBE BIAS	BLIMBE MSE	MINIMAX BIAS	MINIMAX MSE	MLE BIAS	MLE MSE	Frequency of λ at 0	Frequency of λ at >0
1	2	2	2	2	$.0^{3}89$	1.1185	.0038	1.0706	.0485	.8297	508	492
	4	5	7	10	(0)	1.1962	.0064	(1.1475)				
2	1	2	1	2	$.0^{3}81$	1.2162	.0045	1.1635	.0013	1.1723	941	59
	[illegible]	5	7	10	(0)	(1.1962)	(.0030)	(1.1441)				
3	1	-1	-1	1	$.0^{3}40$	1.2023	.0022	1.1485	$.0^{3}40$	1.2023	1000	0
	4	5	7	10	(0)	(1.1962)	(.0016)	(1.1427)				
4	2	2	2	2	.0155	11.2687	.6173	6.4763	1.1946	5.4038	333	667
	1.0	3.0	3.5	5.0	(0)	(11.0947)	(.5002)	(6.2601)				
5	1	2	1	2	.0052	11.5097	.1193	6.0942	.1463	5.7310	494	506
	1.0	3.0	3.5	5.0	(0)	(11.0947)	(.1356)	(5.8955)				
6	1	-1	-1	1	.0102	11.3974	.1761	6.0578	.1139	7.5504	702	298
	1.0	3.0	3.5	5.0	(0)	(11.0947)	(.1251)	(5.0850)				
7	2	2	2	2	.0736	211.3267	6.6153	20.0562	6.9253	20.8978	15	985
	0.25	0.50	0.75	1.00	(0)	(204.9996)	(6.4650)	(19.8380)				
8	1	2	1	2	.3404	204.6011	2.9140	15.5596	3.2267	17.5061	24	976
	0.25	0.50	0.75	1.00	(0)	(204.9996)	(3.3598)	(16.7629)				
9	1	-1	-1	1	.0361	208.2756	1.7295	15.1542	1.6268	17.3724	43	957
	0.25	0.50	0.79	1.00	(0)	(204.9996)	(1.6162)	(15.0193)				
10	2	2	2	2	.2318	171.9734	4.9235	15.2234	5.2546	15.7386	41	959
	0.25	0.75	1.00	5.00	(0)	(169.3596)	(4.621[illegible]	(14.6357)				
11	1	2	1	2	.0426	174.3936	1.9919	12.3709	2.1188	13.5888	54	943
	0.25	0.75	1.00	5.00	(0)	(169.3596)	(1.8618)	(11.9363)				
12	1	-1	-1	1	.1597	173.6389	1.2727	11.2667	1.1798	15.6814	82	918
	0.25	0.75	1.00	5.00	(0)	(169.3596)	(1.1553)	(11.1998)				

Note : Theoretical values are shown in parenthesis. Other values are based on simulation using a sample of 1000 from $N(X\beta, \sigma^2 I)$. The first column with the heading β/DR has two entries in each cell. The upper entries are the values of β_1, β_2, β_3 and β_4. The lower entries are the singular values of the X matrix.

Table 4 Showing ranking by order of magnitude of bias and mean square error of BLIMBE (1), MINIMAX (2) and MLE (3), based on simulation results

β \ DR					4	5	7	10	1.0	3.0	3.5	5.0	.25	.50	.75	1.00	.25	.75	1.00	5.00
				σ	BIAS															
2	2	2	2	1	1	2	3	(738)	1	2	3	(559)	1	2	3	(207)	1	2	3	(250)
				2	1	2	3	(572)	1	2	3	(432)	1	2	3	(48)	1	2	3	(104)
				3	1	2	3	(508)	1	2	3	(333)	1	2	3	(15)	1	2	3	(41)
1	2	1	2	1	1	3	2	(1000)	1	3	2	(950)	1	2	3	(282)	1	2	3	(389)
				2	1	3	2	(994)	1	3	2	(720)	1	2	3	(73)	1	2	3	(136)
				3	1	3	2	(941)	1	2	3	(494)	1	2	3	(24)	1	2	3	(54)
1	-1	-1	1	1	1	3	2	(1000)	1	3	2	(995)	1	3	2	(445)	1	3	2	(547)
				2	1	3	2	(1000)	1	3	2	(894)	1	3	2	(112)	1	3	2	(185)
				3	1	3	2	(1000)	1	3	2	(702)	1	3	2	(43)	1	3	2	(82)
				σ	MSE															
2	2	2	2	1	3	2	1	(738)	3	2	1	(559)	3	2	1	(207)	3	2	1	(250)
				2	3	2	1	(572)	3	2	1	(432)	3	2	1	(48)	3	2	1	(104)
				3	3	2	1	(508)	3	2	1	(333)	2	3	1	(15)	2	3	1	(41)
1	2	1	2	1	2	3	1	(1000)	2	3	1	(950)	3	2	1	(282)	3	2	1	(389)
				2	2	3	1	(994)	3	2	1	(720)	3	2	1	(73)	2	3	1	(136)
				3	2	3	1	(941)	3	2	1	(494)	2	3	1	(24)	2	3	1	(54)
1	-1	-1	1	1	2	3	1	(1000)	2	3	1	(995)	2	3	1	(445)	2	3	1	(547)
				2	2	3	1	(1000)	2	3	1	(894)	2	3	1	(112)	2	3	1	(185)
				3	2	3	1	(1000)	2	3	1	(702)	2	3	1	(43)	2	3	1	(82)

Note Figures in parenthesis indicate the frequency of λ at 0, out of a total frequency of 1000.

Table 5 : Showing a possible analysis of a lower mean square error (2 indicates MINIMAX and 3 MLE)

f_o \ M_o	.133	.532	1.196	1.233	4.931	11.095	18.818	22.778	75.271	91.111	169.360	205.00
.015												2
.024												2
.041											2	
.043												2
.048										3		
.054											2	
.073										3		
.082											2	
.104									3			
.112										2		
.136									2			
.185									2			
.207								3				
.250							3					
.282								3				
.333						3						
.389							3					
.432					3							
.445								2				
.494						3						
.508			3									
.547							2					
.559				3								
.572		3										
.702						2						
.720					3							
.738	3											
.894					2							
.941			2									
.950				2								
.994		2										
.995				2								
1.000	2,2	2	2									

Note : f_o denotes the relative frequency of λ at zero.

of the model. Subject to this reservation for parameter values near the surface of the sphere the MLE seems to have a lower mean square error compared to minimax while for parameter values closer to the centre of the sphere minimax is better on this score. These facts are further brought out in Table 5.

Another fact worth recording is that BLIMBE has the largest m.s.e. of all three methods. That $M_1 \leq M_o$ was proved in theorem 2.3.2. It would be nice if one could similarly show that $M_2 \leq M_o$. A special case where $\alpha_1 = \alpha_2 = \ldots = \alpha_m = \alpha$, is established in the following theorem:

Theorem 2.3.3 For the model $(Y, \alpha\beta, \sigma^2 I), \beta \in \Omega_\beta = \{\beta \in R^m : \beta'\beta \leq \delta^2\}$, $M_2 \leq M_o$.

Proof Clearly

$$\hat{\beta}_2 = \begin{cases} \hat{\beta}_o & \text{if} \quad \hat{\beta}_o \in \Omega_\beta \\ \hat{\beta}_\lambda = \left(\dfrac{\alpha^2}{\alpha^2+\lambda}\right)\hat{\beta}_o & \text{if } \hat{\beta}_o \in \Omega_\beta^c \cap R^m . \end{cases}$$

where $\lambda > 0$ is such that

$$\hat{\beta}'_\lambda \hat{\beta}_\lambda = \delta^2 \tag{2.3.13}$$

If $P(\hat{\beta}_o)$ denotes the distribution function of $\hat{\beta}_o$,

$$M_2 = \int_{\Omega_\beta} (\hat{\beta}_o-\beta)'(\hat{\beta}_o-\beta)dP(\hat{\beta}_o) + \int_{\Omega_\beta^c \cap R^m} (\hat{\beta}_\lambda-\beta)'(\hat{\beta}_\lambda-\beta)dP(\hat{\beta}_o)$$

To show $M_2 \leq M_o$ it is sufficient to show that

$$(\hat{\beta}_\lambda-\beta)'(\hat{\beta}_\lambda-\beta) \leq (\hat{\beta}_o-\beta)'(\hat{\beta}_o-\beta) \tag{2.3.14}$$

for all $\beta \in \Omega_\beta$, when

$$\hat{\beta}'_o \hat{\beta}_o > \delta^2 \tag{2.3.15}$$

Let $S = \{c\hat{\beta}_o : c \in R\}$ and P_S the orthogonal projection operator onto S under inner product induced by I_m. If $\beta \in \Omega_\beta$, so does $P_S\beta$ and hence

$$(\hat{\beta}_\lambda - P_S\beta)'(\hat{\beta}_\lambda - P_S\beta) \leq (\hat{\beta}_o - P_S\beta)'(\hat{\beta}_o - P_S\beta) \qquad (2.3.16)$$

using (2.3.13), (2.3.15) and the fact that $\hat{\beta}_\lambda \in S$. Adding $\beta'(I-P_S)\beta$ to both sides of (2.3.16) we obtain (2.3.14) and hence the theorm.

In the setup of theorem 2.3.3, if additionally Y follows a multivariate normal distribution, then it is easily seen that

$$\lambda = \alpha[\frac{\sigma\chi}{\delta} - \alpha] \qquad (2.3.17)$$

where χ^2 is a noncentral chisquare variate. In the general situation however the distribution of λ is likely to be very complicated. To study this distribution we tabulated λ in intervals of width 0.1, but we shall not report on our findings since we suspect that definite conclusions, if any, can only be obtained through a much more detailed study.

C H A P T E R 3

BHLE AND ALE

If the parameter β in $(Y,X\beta,\sigma^2\Lambda)$ has a known prior distribution, then one way to obtain optimal linear estimators for parametric functionals is to minimize the expected mean square error. These estimators are known as Bayes homogeneous linear estimators (BHLE). In section 1 we state explicit expressions for BHLE's, and also present numerical comparisons in a class of such estimators. In section 2 we establish some properties of the class of admissible linear estimators (ALE) in bounded parameter spaces under the quadratic loss function defined in Chapter 1. In particular, we examine the admissibility of the linear estimators we have studied.

In this chapter, the parameter space Ω_σ for σ^2 in the model $(Y,X\beta,\sigma^2\Lambda)$ will be the entire positive half of the real line. While considering bounded β-spaces of the form $\Omega_\beta = \{\beta \in R^m : \beta' H\beta \leq \delta^2\}$ we shall assume a knowledge of $\theta^2 = \sigma^2/\delta^2$. This type of bounds have been considered by Rao [31]. Thus, for example, when both H and Λ are p.d., the minimax estimator for $p'\beta$ is $p'(X'\Lambda^{-1}X + \theta^2 H)^{-1}X'\Lambda^{-1}Y$.

1. BHLE

Consider a specific prior distribution for β over Ω_β, and denote $E(\beta)$, $D(\beta)$ and $E(\beta\beta')$ by a, $\sigma^2\Delta$ and $\sigma^2 E$ respectively. Clearly

$$\sigma^2 E = \sigma^2\Delta + aa' \qquad (3.1.1)$$

If $b'Y$ is an estimator of $p'\beta$, then the expected mean square error is

$$E(b'Y-p'\beta)^2 = \sigma^2[b'\Lambda b + (X'b-p)'E(X'b-p)] \quad (3.1.2)$$

The BHLE $b_*'Y$ is defined as

$$b_*'\Lambda b_* + (X'b_*-p)'E(X'b_*-p) = \underset{b \in R^n}{\text{Min}} \; [b'\Lambda b+(X'b-p)'E(X'b-p)] \quad (3.1.3)$$

Study of such estimators is essentially due to Chipman [11], Rao [31].

By Mitra [23], b_* stated in terms of an optimal inverse of the design matrix is

$$b_* = (X')^{\dagger}_{E \oplus \Lambda}\, p \quad (3.1.4)$$

Thus the BHLE is $b_*'Y = p'[(X')^{\dagger}_{E\oplus\Lambda}]'Y = p'(X'\Lambda^{-1}X + E^{-1})^{-1}X'\Lambda^{-1}Y$ (3.1.5)

when both Λ and E are p.d. The expected bias of $b_*'Y$ is

$$p'[[(X')^{\dagger}_{E\oplus\Lambda}]'X-I]a \quad (3.1.6)$$

and, by corollary 4.1(b) of Mitra [23], the expected mean square error is

$$\sigma^2 p'[E-EX'(X')^{\dagger}_{E\oplus\Lambda}]p \quad (3.1.7)$$

The Bayes linear estimator (BLE), $b_*'Y + c_*$ similarly minimizes

$$E(b'Y+c-p'\beta)^2 = \sigma^2[b'\Lambda b+(X'b-p)'\Delta(X'b-p)] + [(X'b-p)'a+c]'[(X'b-p)'a+c] \quad (3.1.8)$$

and is given by the choice

$$b_* = (X')^{\dagger}_{\Delta\oplus\Lambda}p, \quad c_* = p'[I-[(X')^{\dagger}_{\Delta\oplus\Lambda}]'X]a \quad (3.1.9)$$

The expected bias of the BLE is the null vector, while the expected mean square error is

$$\sigma^2 p'[\Delta-\Delta X'(X')^{\dagger}_{\Delta\oplus\Lambda}]p \quad (3.1.10)$$

Since $E-\Delta$ is n.n.d so is $E \oplus \Lambda - \Delta \oplus \Lambda$. Corollary 4.1(b) of Mitra [23] therefore implies that

$$[E-EX'(X')^{\dagger}_{E\oplus\Lambda}] - [\Delta-\Delta X'(X')^{\dagger}_{\Delta\oplus\Lambda}] \tag{3.1.11}$$

is n.n.d., as is to be even otherwise expected. (3.1.11) is a null matrix when a = 0, in which case the BLE is the same as the BHLE.

For simultaneous estimation of several parametric functionals, the BHLE L_*Y for $P\beta$ minimizes

$$E(LY-P\beta)'A(LY-P\beta) = \sigma^2 \text{ tr}[L'ALA + (LX-P)'A(LX-P)E] \tag{3.1.12}$$

where A is n.n.d. The problem of minimizing this expression was solved in chapter 2, section 2, while obtaining the approximate minimax estimator. Thus

$$L_* = P[(X')^{\dagger}_{E\oplus\Lambda}]' \tag{3.1.13}$$

leads to one choice of BHLE. The entire class of BHLE's for $P\beta$ is given by L_*Y+ZY, where Z is any matrix such that $\mathcal{M}(Z)\subset\mathcal{N}(\Lambda)$. When Λ and E are p.d., the BHLE for β is

$$\hat{\beta}_* = (X'\Lambda^{-1}X + E^{-1})^{-1} X'\Lambda^{-1}Y \tag{3.1.14}$$

Observe that this establishes the assertion made in chapter 2, viz. the approximate minimax estimator over $\Omega_\beta = \{\beta:\beta'H\beta \leq \delta^2\}$ is a BHLE with $E = \theta^{-2}H^{-1}$ On the same lines as in theorem 2.3.1 one can show $(I-P'_{X',E})\hat{\beta}_* = 0$. Also, defining $B_* = (E(\hat{\beta}_*)-\beta)'E^{-1}(E(\hat{\beta}_*)-\beta), \hat{\beta}_o = X^+_{\Lambda^{-1}E^{-1}}Y$, $B_o = (E(\hat{\beta}_o)-\beta)'E^{-1}(E(\hat{\beta}_o)-\beta)$, we have $B_o \leq B_*$, which corresponds to theorem 2.3.2 (a). However, it will be seen through numerical examples that a result corresponding to theorem 2.3.2 (b) need not always hold. One situation where this is valid is when $\Omega_\beta = \{\beta\varepsilon R^m:\beta'E^{-1}\beta \leq \sigma^2\}$, when $\hat{\beta}_*$ is the approximate minimax estimator.

Let us now study the relative performances of BHLE's with different prior distributions in the model $(Y,X\beta,\sigma^2I)$. If β is uniform over $\Omega_\beta = \{\beta\varepsilon R^m:\beta'\beta\leq\delta^2\}$ then a = 0, $\Delta = E = [(m+2)\theta^2]^{-1}I$. Keeping this Ω_β in mind, another class of priors on R^m could be those with $a = 0$, and with a density which is constant on spheres around 0, but decreasing in magnitude as one moves away from the origin. Thus one expresses maximum faith in the origin, and this faith diminishes as $||\beta|| = \sqrt{\beta'\beta}$ increases; however, there is no preference for any particular direction. If one assumes normality of the underlying β-distribution, then it is not difficult to see that, $E = \Delta = [\lambda\theta^2]^{-1}I$, for some $\lambda > 0$. Here the BHLE assumes the form

$$\hat{\beta}_\lambda = (X'X + \lambda\theta^2 I)^{-1}X'Y \qquad (3.1.15)$$

Observe that $\lambda = 1$ gives the minimax estimator, and $\lambda = m+2$ the BHLE with uniform prior; $\hat{\beta}_\lambda \to \hat{\beta}_o$, the BLIMBE, as $\lambda \downarrow 0$ ($\hat{\beta}_2$ is not to be confused with the MLE ($\hat{\beta}_2$) of chapter 2, section 3).

We shall compare $\hat{\beta}_\lambda$'s for various λ's on the basis of their biases and mean square errors. Observe that these priors are invariant under an orthogonal transformation of the variables. Hence, as in chapter 2, section 3, we can take, without loss of generality, $X = \begin{pmatrix} A & 0 \\ 0 & 0 \end{pmatrix}$, where $A = \mathrm{diag}(\alpha_1,\ldots,\alpha_r)$. Defining $d_i = \alpha_i^2$,

$$B_\lambda = (E(\hat{\beta}_\lambda)-\beta)'(E(\hat{\beta}_\lambda)-\beta) = \lambda^2\theta^4 \sum_{i=1}^{r} \frac{\beta_i^2}{(d_i+\lambda\theta^2)^2} + \sum_{i=r+1}^{m} \beta_i^2$$

$$M_\lambda = E(\hat{\beta}_\lambda-\beta)'(\hat{\beta}_\lambda-\beta) = \sigma^2 \sum_{i=1}^{r} \frac{d_i}{(d_i + \lambda\theta^2)^2} + B_\lambda \qquad (3.1.16)$$

Clearly B_λ increases with λ for all β. M_λ decreases with increasing λ at $\beta = 0$, and hence we may expect such a behaviour in a neighbourhood of the origin. Hence, if $\Omega_\beta = R^m$, then for $\lambda_1 \neq \lambda_2$ the inequality $M_{\lambda_1} \leq M_{\lambda_2}$ cannot be valid for all β. However, if Ω_β is a bounded set containing the origin, then one expects $M_{\lambda_1} \leq M_{\lambda_2}$ for a nondegenerate subset of Ω_β if $\lambda_1 \geq \lambda_2$, and as one moves towards the boundary of Ω_β, it may happen that $M_{\lambda_1} > M_{\lambda_2}$.

These observations were mostly confirmed by the numerical examples studied. We consider the examples of chapter 2, section 3, with a few additional points in Ω_β for each of the 12 combinations of a design matrix and a σ. Recall that the model is

$$E(y_i) = \alpha_i\beta_i,\ Cov(y_i,y_j) = \sigma^2\delta_{ij},\ \Omega_\beta = \{\beta\varepsilon R^4 : \sum_{i=1}^{4} \beta_i^2 \leq 17\} \qquad (3.1.17)$$

for i,j = 1,2,3,4. Clearly (see equations(3.1.16)), it is sufficient to compare M_λ's on full rank models of this type. However, if λ is a function of m, then one will have to consider different $\hat{\beta}_\lambda$'s for full and nonfull rank models. Thus for the full rank model (m = 4), the BHLE with uniform prior is $\hat{\beta}_6$ while for the model with m=5, r=4 it is $\hat{\beta}_7$. In the next section it will be seen that $p'\hat{\beta}_\lambda$ is admissible for $p'\beta$ whenever $\lambda \geq 1$. Moreover, if p is an eigenvector of X'X corresponding to a positive eigenvalue then $p'\hat{\beta}_\lambda$ is inadmissible for $p'\beta$, whenever $\lambda < 1$; even otherwise this result holds at least for λ in a neighbourhood of the origin except possibly when $p'P_{X'}p$ is small with respect to p'p. Hence we compare M_λ's for $\lambda \geq 1$ only. In fact we consider $M_1, M_2, \ldots, M_7$ and also M_o, the m.s.e. of $\hat{\beta}_o = (X'X)^+X'Y$, which gives the BLIMBE for the model (3.1.17). Ranking of these M_λ's in increasing order of magnitude is shown in table 6.

From the tables it is seen that the natural ranking at $\beta = 0$ (i.e. $M_{\lambda_1} \leq M_{\lambda_2}$ if $\lambda_1 \geq \lambda_2$) is also prevelant in a neighbourhood of the origin. It tends to be disrupted as one moves towards the surface of the sphere. In this region, this phenomenon is more noticable if the numerically larger coordinates of β correspond to smaller singular values of the design matrix. (Note that the singular values α_i are arranged in an increasing order i.e. $\alpha_1 < \alpha_2 < \alpha_3 < \alpha_4$). This is especially true when

Table 6: Showing ranking by increasing order of magnitude of mean square error of BHLE's

β'β	β (Mₒ / D,σ)				.133 D_1,1	.532 D_1,2	1.196 D_1,3	1.233 D_2,1	4.931 D_2,2	11.095 D_2,3
0	0	0	0	0	76543210	76543210	76543210	76543210	76543210	76543210
1	1	0	0	0	76543210	76543210	76543210	76543210	76543210	76543210
	0	1	0	0	76543210	76543210	76543210	76543210	76543210	76543210
	0	0	1	0	76543210	76543210	76543210	76543210	76543210	76543210
	0	0	0	1	76543210	76543210	76543210	76543210	76543210	76543210
4	-1	-1	-1	1	76543210	76543210	76543210	76543210	76543210	76543210
7	2	1	1	1	65743210	65743210	67543210	45367210	54673210	67543210
	1	1	1	2	76543210	76543210	76543210	76543210	76543210	76543210
10	2	1	2	1	56473210	56473210	65743210	45367210	54637210	56473210
	1	2	1	2	76543210	76543210	76543210	76543210	76543210	76543210
12	3	1	1	1	32415067	32415067	32415607	21340567	23145670	23415670
	1	1	1	3	76543210	76543210	76543210	76543210	76543210	76543210
14	3	2	1	0	23140567	23145067	32415067	21340567	23145670	23415670
	0	1	2	3	76543210	76543210	76543210	76543210	76543210	76543210
15	3	1	1	2	32415067	32415067	32415607	21340567	23145670	23415670
16	2	2	2	2	45362710	45362710	45362710	45367210	45367210	45637210
	4	0	0	0	21304567	21304567	21304567	12034567	12034567	12345607
	0	4	0	0	43526170	43526170	43526170	76543210	76543210	56743210
	0	0	4	0	76543210	76543210	76543210	76543210	76543210	76543210
	0	0	0	4	76543210	76543210	76543210	76543210	76543210	76543210

Note 1. D_1 denotes the design matrix with singular values 4,5,7,10; D_2 with 1.0, 3.0, 3.5, 5.0; D_3 with .25, .50, .75, 1.00 and D_4 with .25, .75, 1.00, 5.00. In D_i,i, σ assumes the value i.

Note 2. In each cell i denotes BHLE with prior dispersion matrix $(\delta^2/i)I$ (i= 1,2,3,4,5,6,7). 0 denotes BLIMBE. The first entry corresponds to least m.s.e. and so on.

M_o is small, or in other words, the model does allow precise estimation of the parameters. When the larger coordinates of β correspond to larger α_i, the natural order tends to be disrupted for larger M_o, though not as frequently, (compare, for example, the two rows under $\beta'\beta = 7$). However, there are some exceptions to this . For example, a study of the column under $M_o = 169.360$ shows that even for large M_o, the natural order need not be disturbed very much even near the surface of Ω_β if the large β-coordinates correspond to α_i's which are sufficiently large in comparison to the other singular values.

$\hat{\beta}_o$ is generally the worst estimator with respect to the m.s.e. However its relative performance improves near the surface of Ω_β. In fact, on the lines of theorem 2.3.2(b) one can show that

$$M_\lambda \leq M_o \quad \text{if} \quad \sum_{i=1}^{m} \beta_i^2 \leq \frac{\delta^2}{\lambda}$$

Thus if $\lambda \leq 1$, $M_\lambda \leq M_o$ for all β in Ω_β. On the other hand, if $\lambda > 1$, $\frac{\delta^2}{\lambda} < \delta^2$, and it is not surprising if M_λ exceeds M_o near the surface of the sphere Ω_β.

In conclusion, we recommend the use of β_λ with the largest λ attainable under a priori contraints, especially if the true parameter point is not very close to the surface of the sphere. Otherwise, if $\beta'\beta$ is close to δ^2, and one has reasons to believe that the larger β-coordinates correspond to the smaller α_i, and additionally M_o is not very large, then one can safely opt for a smaller λ. Use of $\hat{\beta}_o$ is discouraged unless reduction of bias is of prime importance.

2. ALE

In this section we study some properties of the class of admissible linear estimators for parametric functionals in the Gauss Markov model $(Y, X\beta, \sigma^2\Lambda)$ with $\beta \in \Omega_\beta = \{\beta \in R^m : \beta' H\beta \leq \delta^2\}$, where Λ and H are p.d. Observe that g'Y is admissible for $q'\beta$ if and only if $g_1'Y_1$ is admissible for $q_1'\beta_1$ in the model $(Y_1, X_1\beta_1, \sigma^2 I)$, $\Omega_{\beta_1} = \{\beta_1 \in R^m : \beta_1' \beta_1 \leq \delta^2\}$, where $\Lambda = F'F$, $H = T'T$, $Y_1 = (F')^{-1}Y$, $X_1 = (F')^{-1}XT^{-1}$, $\beta_1 = T\beta$, $g_1 = Fg$, $q_1 = (T')^{-1}q$. Thus it is sufficient to consider the model $(Y, X\beta, \sigma^2 I)$, $\beta \in \Omega_\beta$ and $\sigma^2 \in \Omega_\sigma = (0, \infty)$, where

$$\Omega_\beta = \{\beta \in R^m : \beta'\beta \leq \delta^2\} \tag{3.2.1}$$

Let us recall the relevant definitions of chapter 1. If g'Y estimates $q'\beta$, then the risk is

$$R(g'Y, q'\beta) = E(g'Y - q'\beta)^2 = \sigma^2 g'g + [(X'g-q)'\beta]^2 \tag{3.2.2}$$

g'Y is said to be inadmissible for $q'\beta$ if there exists a h'Y with the property

$$R(h'Y, q'\beta) \leq R(g'Y, q'\beta) \tag{3.2.3}$$

for all $\beta \in \Omega_\beta$, with strict inequality somewhere in Ω_β. Otherwise g'Y is admissible for $q'\beta$. If

$$\mathcal{C} = \{g'Y : g \in \mathcal{M}(X)\} \tag{3.2.4}$$

$\mathcal{A}_\Omega^{q'\beta}$ = {all admissible linear estimators for $q'\beta$ in

$$(Y, X\beta, \sigma^2 I), \text{ when } \beta \in \Omega\} \tag{3.2.5}$$

then it was shown in chapter 1 that

$$\mathcal{A}_{\Omega\beta}^{q'\beta} \subset \mathcal{C} \tag{3.2.6}$$

Also denote $p = P_X q$ and $s = q - p$.

Some characterestics of $\mathcal{A}_{\Omega}^{q'\beta}$ are established in the following theorem.

Theorem 3.2.1 (i) $\mathcal{A}_{\Omega_1}^{q'\beta} \subset \mathcal{A}_{\Omega_2}^{q'\beta}$ if $0 \in \Omega_1 \subset \Omega_2$ and $\dim[\mathcal{L}(\Omega_1)] = m$.

(ii) $\mathcal{A}_{\Omega_\beta}^{p_o'\beta} \subset \mathcal{A}_{\Omega_\beta}^{q'\beta}$

Proof: (i) If $g'Y \notin \mathcal{A}_{\Omega_2}^{q'\beta}$, then (3.2.3) is satisfied with some h, for β in Ω_2. Inequality in (3.2.3) is strict somewhere in Ω_1, since otherwise the conditions imply $h'h = g'g$, $(X'h-q) = \pm (X'g-q)$. Hence $Y \notin \mathcal{A}_{\Omega_1}^{q'\beta}$. (ii) If $g'Y \notin \mathcal{A}_{\Omega_\beta}^{q'\beta}$, then considering (3.2.3) and writing $\beta=\beta_o+\beta_1$, where $\beta_o = P_{X'}\beta \in \Omega_\beta$, we obtain

$$\sigma^2 h'h + [(X'h-p_o)'\beta_o]^2 \leq \sigma^2 g'g + [(X'g-p_o)'\beta_o]^2$$

$$\Leftrightarrow \sigma^2 h'h + [(X'h-p_o)'\beta]^2 \leq \sigma^2 g'g + [(X'g-p_o)'\beta]^2$$

since $(X'h-p_o)'\beta_1 = 0 = (X'g-p_o)'\beta_1$. It is now easily seen that

$g'Y \notin \mathcal{A}_{\Omega_\beta}^{p_o'\beta}$

Considering specific estimators, we examine the admissibility of some important ones in theorem 3.2.2.

Theorem 3.2.2 (i) $0 \in \mathcal{A}_{\Omega}^{q'\beta}$ whenever Ω contains the origin.

(ii) The minimax estimator $g_1'Y \in \mathcal{A}_{\Omega}^{q'\beta}$, whenever $\Omega=\{\beta\in R^m : \beta'H\beta\leq\delta^2\}$, H p.d.

(iii) The BHLE $g_*^{*\prime}Y \in \mathcal{A}_{\Omega_\beta}^{q'\beta}$, whenever the maximum eigenvalue of $E(\beta\beta') = \sigma^2 E$ does not exceed δ^2.

(iv) If $p'\beta$ is estimable, then its BLUE $g_o'Y \notin \mathcal{A}_{\Omega_\beta}^{p'\beta}$, whenever $p \neq 0$.

Proof (i) If $g'Y = 0 \notin \mathcal{A}_{\Omega}^{q'\beta}$, then there exists a $h'Y$ satisfying (3.2.3) for all β in Ω. Since $0 \in \Omega$ it is easy to see that $h'h \leq g'g = 0$. Hence $h'Y = 0$.

(ii) Suppose, to the contrary, that there exists a $h'Y$ satisfying (3.2.3) for all β in Ω. Further, let $R(h'Y,q'\beta)$ be maximized at $\beta = \beta_m$ in Ω. Clearly

$$\max_{\beta \in \Omega} R(h'Y,q'\beta) = \sigma^2 h'h + [(X'h-q)'\beta_m]^2 \leq \sigma^2 g_1'g_1 + [(X'g_1-q)'\beta_m]^2$$

$$\leq \max_{\beta \in \Omega} R(g_1'Y, q'\beta) .$$

This contradicts the fact that $g_1'Y$ is minimax for $q'\beta$ over Ω.

(iii) Since the BHLE $g_*'Y$ minimizes $\sigma^2 g'g + \theta^2(X'g-q)'E(X'g-q)$, it is minimax for $q'\beta$ over $\Psi = \{\beta \in R^m : \beta'H\beta \leq \delta^2\}$, where $H = \theta^{-2}E^{-1}$. Hence by (ii) $g_*'Y \in \mathcal{A}_\Psi^{q'\beta}$. If the maximum eigenvalue of $\sigma^2 E$ is no greater than δ^2, then $\beta'H\beta \geq \beta'\beta$, whatever β. Thus $\Psi \subset \Omega_\beta$ and an appeal to theorem 3.2.1(i) establishes (iii).

(iv) If $g_1'Y$ is minimax for $p'\beta$ over Ω_β, then

$$R(g_1'Y,p'\beta) \leq \max_{\beta\in\Omega_\beta} R(g_1'Y,p'\beta) \leq \max_{\beta\in\Omega_\beta} R(g_o'Y, p'\beta) = \sigma^2 g_o'g_o$$

$$= R(g_o'Y, p'\beta),$$

whatever β in Ω_β, since $X'g_o = p$. Hence (3.2.3) is satisfied with $q=p$, $h=g_1$ and $g=g_o$, for all β in Ω_β.

<u>Note 3.2.1</u> Theorems 3.2.1 and 3.2.2 (ii), (iii) remain valid in the model $(Y,X\beta,\sigma^2\Lambda)$, where Λ is positive semidefinite (p.s.d.) while theorems 3.2.2(i) and 3.2.2 (iv) need some modifications. Observe that the risk function is $R(h'Y,q'\beta) = \sigma^2 h'\Lambda h + [(X'h-q)'\beta]^2$. Hence if $R(h'Y,q'\beta) \leq R(0,q'\beta)$ for all β in Ω, then the conditions of theorem 3.2.2(i) imply $h'\Lambda h = 0$, or in other words $h'Y = h'X\beta$ almost surely. Thus $[(X'h-q)'\beta]^2 \leq (q'\beta)^2$ for all β in Ω. Under the additional assumption $\dim[\mathcal{C}(\Omega)] = m$, one obtains $X'h-q = cq$, for some c in $[-1,1]$. If $q'\beta$ is nonestimable, then surely $c = -1$, $X'h = 0$ and hence $h'Y = 0$,

almost surely. Thus, in this situation $0 \in \mathcal{A}_{\Omega}^{q'\beta}$. However if $q'\beta = p'\beta$ is estimable, it may happen that there exists a 'h' with the properties $Ah = 0$, $X'h = p$. In such cases 0 will not be admissible. Therefore in general, 0 may be in $\mathcal{A}_{\Omega}^{q'\beta} - \mathcal{A}_{\Omega}^{p_o'\beta}$ (the set theoretic difference). It is not difficult to see that the BLUE $h'Y$ will be admissible for $p'\beta$ when a 'h' as specified above exists. In all other situations it will be inadmissible.

Consider now, the class of BHLE's $\{q'\hat{\beta}_\lambda, \lambda > 0\}$ for $q'\beta$ in the model $(Y, X\beta, \sigma^2 I)$, $\Omega_\beta = \{\beta \in R^m : \beta'\beta \leq \delta^2\}$, where $\hat{\beta}_\lambda = (X'X + \lambda\theta^2 I)^{-1}X'Y$. We observed in section 1 that this class arises from a natural choice of prior distributions for β, and in particular includes the minimax estimator. Moreover, the BLIMBE is obtained as a limit when $\lambda \downarrow 0$. Let us examine the admissibility of $q'\hat{\beta}_\lambda$ for each λ in $(0,\infty)$. If $q'\beta$ is totally nonestimable, then using theorem 3.2.2(i), $q'\hat{\beta}_\lambda = 0 \in \mathcal{A}_{\Omega_\beta}^{q'\beta}$, whatever λ. In general by theorem 3.2.2 (iii), $q'\hat{\beta}_\lambda$ is admissible whenever $\lambda \geq 1$. The case $\lambda < 1$ is considered in theorem 3.2.3. To prove this theorem we need the following lemma.

<u>Lemma 3.2.1</u> $\quad \underset{x \in R^m}{\text{Max}} \dfrac{(a'x)(b'x)}{x'x} = \dfrac{b'a + \sqrt{a'a}\sqrt{b'b}}{2}$

<u>Proof</u> Since $\dfrac{(a'x)(b'x)}{x'x} = \dfrac{x'(ab' + ba')x}{2x'x}$, its maximum is the maximum eigenvalue of $\dfrac{ab' + ba'}{2} = C$ (say). Observe that C can have at most two nonzero eigenvalues since $R(C) \leq 2$. The lemma now follows once it is verified that

$$C\left(\frac{a}{\sqrt{a'a}} \pm \frac{b}{\sqrt{b'b}}\right) = \frac{1}{2}\left(b'a \pm \sqrt{a'a}\sqrt{b'b}\right)\left(\frac{a}{\sqrt{a'a}} \pm \frac{b}{\sqrt{b'b}}\right).$$

Consider the spectral decomposition, $X'X = P'DP$, where P is orthogonal and $D = \text{diag}(d_1,\ldots,d_m)$. If $R(X) = r \leq m$, $d_{r+1} = d_{r+2} = \ldots = d_m = 0$. Denote $Pq = f = (f_1,\ldots,f_m)'$, $P\beta = \gamma = (\gamma_1,\ldots,\gamma_m)'$, and note that

$$\beta'\beta \leq \delta^2 \iff \sum_{i=1}^{m} \gamma_i^2 \leq \delta^2 \qquad (3.2.7)$$

<u>Theorem 3.2.3</u> $q'\hat{\beta}_\lambda \notin \mathcal{A}^{q'\beta}_{\Omega_\beta}$, whenever

$$\lambda\{\sum_{i=1}^{m} \frac{f_i^2}{(d_i+\lambda\theta^2)^2}\}^{1/2} \{\sum_{i=1}^{m} \frac{d_i^2 f_i^2}{(d_i+\lambda\theta^2)^4}\}^{1/2} < (2-\lambda)\{\sum_{i=1}^{m} \frac{d_i f_i^2}{(d_i+\lambda\theta^2)^3}\} \tag{3.2.8}$$

<u>Proof</u> $R(q'\hat{\beta}_\lambda, q'\beta) = \sigma^2 q'(X'X+\lambda\theta^2 I)^{-1}X'X(X'X+\lambda\theta^2 I)^{-1}q$

$$+ [q'\{(X'X+\lambda\theta^2 I)^{-1}X'X-I\}\beta]^2$$

$$= \sigma^2 \sum_{i=1}^{m} \frac{d_i f_i^2}{(d_i+\lambda\theta^2)^2} + \theta^4[\sum_{i=1}^{m} \frac{f_i \gamma_i}{(\frac{d_i}{\lambda}+\theta^2)}]^2$$

$$= G_q(\lambda,\gamma), \text{ say.}$$

$$\frac{\partial G_q(\lambda,\gamma)}{\partial\lambda} = -2\sigma^2\theta^2 \sum_{i=1}^{m} \frac{d_i f_i^2}{(d_i+\lambda\theta^2)^3} + 2\theta^4\lambda[\sum_{i=1}^{m} \frac{f_i\gamma_i}{(d_i+\lambda\theta^2)}][\sum_{i=1}^{m} \frac{d_i f_i \gamma_i}{(d_i+\lambda\theta^2)^2}]$$

Denoting $G_q(\lambda) = \underset{\gamma:\sum_{i=1}^{m}\gamma_i^2 \le \delta^2}{\text{Max}} \frac{\partial G_q(\lambda,\gamma)}{\partial\lambda}$, and using lemma 3.2.1 and the relation $\theta^4\delta^2 = \sigma^2\theta^2$, one observes

$$G_q(\lambda) = -2\sigma^2\theta^2 \sum_{i=1}^{m} \frac{d_i f_i^2}{(d_i+\lambda\theta^2)^3} + \lambda\sigma^2\theta^2[\sum_{i=1}^{m} \frac{d_i f_i^2}{(d_i+\lambda\theta^2)^3} + \{\sum_{i=1}^{m} \frac{f_i^2}{(d_i+\lambda\theta^2)^2}\}^{1/2}\{\sum_{i=1}^{m} \frac{d_i^2 f_i^2}{(d_i+\lambda\theta^2)^4}\}^{1/2}]$$

< 0, by (3.2.8).

Since $G_q(\lambda)$ is continuous in $\lambda \in [0,\infty)$, there exists $\eta_1 > 0$ such that $G_q(\lambda+\eta) < 0$ whenever $0 < \eta < \eta_1$. Hence, by the Mean Value Theorem,

$$G_q(\lambda+\eta,\gamma) - G_q(\lambda,\gamma) = \frac{\partial G_q(\lambda,\gamma)}{\partial\lambda}\Big/_{\lambda+\xi}\ \eta,\ 0 < \xi < \eta,$$

$$\leq G_q(\lambda+\xi)\eta$$

$$< 0,$$

whatever γ satisfying $\sum_{i=1}^{m} \gamma_i^2 \leq \delta^2$. Thus

$$R(q'\hat{\beta}_{\lambda+\eta}, q'\beta) < R(q'\hat{\beta}_\lambda, q'\beta),$$

for all β in Ω_β and some $\eta > 0$, whenever λ satisfies (3.2.8). This establishes the theorem.

Note 3.2.2 The inequality (3.2.8) is violated whenever $\lambda \geq 1$. Defining $\lambda_q = \text{Inf}\{\lambda : G_q(\lambda) \geq 0\}$, this may be equivalently stated as $\lambda_q \leq 1$.

Note 3.2.3 Inequality (3.2.8) may be written as

$$\left\{\sum_{i=1}^{r} \frac{\lambda^2 f_i^2}{(d_i + \lambda\theta^2)^2} + \sum_{i=r+1}^{m} \frac{f_i^2}{\theta^4}\right\}^{1/2} \left\{\sum_{i=1}^{r} \frac{d_i^2 f_i^2}{(d_i + \lambda\theta^2)^4}\right\}^{1/2} < (2-\lambda)\left\{\sum_{i=1}^{r} \frac{d_i f_i^2}{(d_i+\lambda\theta^2)^3}\right\} \tag{3.2.9}$$

since $d_{r+1} = \ldots = d_m = 0$. Hence $G_q(0) < 0$, if and only if ,

$$\left\{\sum_{i=r+1}^{m} f_i^2\right\}^{1/2} < 2\theta^2 \left\{\sum_{i=1}^{r} \frac{f_i^2}{d_i^2}\right\}^{1/2} \tag{3.2.10}$$

In this case $\lambda_q > 0$. When $q = p_o$, $f_{r+1} = \ldots = f_m = 0$; hence, in particular $\lambda_{p_o} > 0$. Moreover by (3.2.9), $\lambda_q \leq \lambda_{p_o}$ In fact if λ satisfies (3.2.8), i.e. $q'\hat{\beta}_\lambda \notin \mathcal{A}_{\Omega_\beta}^{q'\beta}$, then λ satisfies (3.2.8) with q replaced by p_o; i.e. $q'\hat{\beta}_\lambda \notin \mathcal{A}_{\Omega_\beta}^{p_o'\beta}$ (note that $q'\hat{\beta}_\lambda = p_o'\hat{\beta}_\lambda$, whatever $\lambda > 0$). This is compatible with theorem 3.2.1(ii).

Suppose $q = p_o$ is an **eigenvector** of $X'X$ corresponding to the eigenvalue $d_o > 0$. If k is the multiplicity of this eigenvalue, then one may assume, with no loss of generality, $d_1 = d_2 = \ldots = d_k = d_o$. Thus, $f_i = 0$ whenever $i > k$, so that (3.2.8) simplifies to

$$\lambda < 1 \qquad (3.2.11)$$

or $\lambda_q = 1$. Hence, in this situation $q'\hat{\beta}_\lambda \notin \mathcal{A}^{q'\beta}_{\Omega_\beta}$ whenever $\lambda < 1$. In general, however, there could exist a λ in $(0,1)$ which violates (3.2.8). The author feels that in such situations $q'\beta_\lambda$ will be admissible for $q'\beta$. However, attemps to prove or disprove this have been unsuccessful so far.

<u>Note 3.2.4</u> Since $G_q(\lambda)$ is continuous at $\lambda = 0$, the BLIMBE $q'(X'X)^{+}X'Y$ is inadmissible whenever (3.2.10) holds. This generalizes theorem 3.2.2(iv). However, if (3.2.10) is violated, the BLIMBE could be admissible for $q'\beta$, or in other words it could be in $\mathcal{A}^{q'\beta}_{\Omega_\rho} - \mathcal{A}^{p_o'\beta}_{\Omega_\beta}$. We conclude this chapter with an illustration of such a situation.

Consider a single observation y with $E(y) = \beta_1$, $V(y) = 1$, $\Omega_\beta = \{\beta \in R^2 : \beta_1^2 + \beta_2^2 \leq 1\}$. Thus $n = r = 1$, $m = 2$, $\sigma^2 = \delta^2 = 1$, $X = (1,0)$. Consider the problem of estimating $q'\beta = \beta_1 + c\beta_2$, where

$$c > 2 \qquad (3.2.12)$$

Observe that for such a choice of c, (3.2.10) is violated. The BLIMBE is y. If there exists an estimator hy with the property

$$R(hy, q'\beta) \leq R(y, q'\beta)$$

i.e.

$$h^2 + (h-1)^2\beta_1^2 - 2c(h-1)\beta_1\beta_2 \leq 1 \ . \qquad (3.2.13)$$

for all $\beta = \begin{pmatrix}\beta_1 \\ \beta_2\end{pmatrix}$ in Ω_β, then surely

$$h^2 < 1 \tag{3.2.14}$$

The inequality (3.2.14) is strict, since if $h^2 = 1$, then considering (3.2.13) with $\beta_1 = 1$, $\beta_2 = 0$ it is easy to see that $h = 1$; thus in this case (3.2.13) is an identity. Moreover, straightforward computations show that when (3.2.13) is satisfied with $\beta_1 = \beta_2 - \frac{1}{\sqrt{2}}$,

$$h \in [1, \frac{2c-1}{3}] \tag{3.2.15}$$

a nondegenerate interval by virtue of (3.2.12). Since (3.2.15) contradicts (3.2.14), the BLIMBE is admissible for $\beta_1 + c\beta_2$.

CHAPTER 4

TESTING NONTESTABLE FUNCTIONALS

Consider the problem of testing $H_o:q'\beta = c$ in the Gauss-Markov model $(Y,X\beta,\sigma^2 A)$, where Y follows a multivariate normal distribution. It was seen in theorem 1.2 that if $q'\beta$ is nonestimable and the parameter space is unbounded, then H_o cannot possess any nontrivial test. In this sense nonestimable parametric functionals are nontestable. If the parameter space is only partially bounded, for example, $\Omega_\beta = \{\beta \in R^m:\beta'H\beta \leq \delta^2\}$, with H p.s.d., then the theorem remains valid for parametric functionals with coefficient vectors outside $\mathcal{M}(X':H)$. For all other functionals, i.e. $q'\beta$ with $q \in \mathcal{M}(X':H)$ it is possible to develop a procedure to test $H_o:q'\beta=c$ by approximating $q'\beta$ with suitable estimable parametric functionals.

For simplicity we shall restrict ourselves to the situation where H is p.d., when all parametric functionals can be tested. For such models, we develop a test for H_o in section 1, by replacing the hypothesis by one stated in terms of a (fixed) estimable parametric functional. In sections 2 and 3 we search the 'best' approximating estimable functional, or the one that, in some sense, maximizes the power at alternatives.

In this chapter it will be assumed that the underlying distribution of Y is multivariate normal. Moreover, the parameter space for (β,σ^2) will be $\Omega_\beta \times \Omega_\sigma$, the Cartesan product of the parameter spaces of β and σ^2.

1. The test

The testing problem in $(Y, X\beta, \sigma^2\Lambda)$, $\beta\epsilon\Omega_\beta = \{\beta\epsilon R^m : \beta' H\beta \leq \delta^2\}$, Λ and H p.d., can be reduced to an equivalent problem in $(Y, X\beta, \sigma^2 I)$, $\beta\epsilon\Omega_\beta = \{\beta\epsilon R^m : \beta'\beta \leq \delta^2\}$ by nonsingular transformations similar to those discussed in chapter 3, section 2. Writing explicitly, the model is,

$$E(Y) = X\beta \tag{4.1.1}$$

$$D(Y) = \sigma^2 I \tag{4.1.2}$$

$$\beta\epsilon\Omega_\beta = \{\beta\epsilon R^m : \beta'\beta \leq \delta^2\} \tag{4.1.3}$$

Our problem is to test the hypothesis

$$H_o : q'\beta = c \tag{4.1.4}$$

Equation (4.1.4) will have a solution in Ω_β if and only if $c^2 \leq \delta^2(q'q)$; if this is violated then H_o can be rejected without carrying out any statistical test. Thus, we shall only be concerned with situations where

$$\Delta^2 = \delta^2 - \frac{c^2}{q'q} \geq 0 \tag{4.1.5}$$

Observe that $q'\beta = c$, $\beta'\beta \leq \delta^2$ implies

$$\beta = \frac{cq}{q'q} + (I - \frac{cq'}{q'q})z \tag{4.1.6}$$

where $z \epsilon R^m$ is arbitrary except for

$$z'(I - \frac{qq'}{q'q})z \leq \Delta^2 \tag{4.1.7}$$

Thus, for any $p \epsilon R^m$,

$$p'\beta = \frac{cp'q}{q'q} + p'(I - \frac{qq'}{q'q})z \quad ,$$

where z satisfies (4.1.7). By Cauchy Schwarz inequality

$$\frac{cp'q}{q'q} - \Delta\sqrt{\{p'(I-\frac{qq'}{q'q})p\}} \leq p'\beta \leq \frac{cp'q}{q'q} + \Delta\sqrt{\{p'(I-\frac{qq'}{q'q})p\}} .$$

Writing

$$\lambda^2 = p'(I - \frac{qq'}{q'q})p = p'p - \frac{(p'q)^2}{q'q} \qquad (4.1.8)$$

We have thus arrived at the following theorem .

Theorem 4.1.1 If $\beta'\beta \leq \delta^2$, $q'\beta = c$ =>

$$\frac{cp'q}{q'q} - \Delta\lambda \leq p'\beta \leq \frac{cp'q}{q'q} + \Delta\lambda \qquad (4.1.9)$$

When $q'\beta$ is nonestimable H_o cannot be tested directly by conventional methods. However, if there is a testable hypothesis H_o' with the property $H_o \Rightarrow H_o'$, then considering H_o' instead of H_o one stays on the safer side since rejection of H_o' will surely imply the falsity of H_o. As methods for testing hypothesis involving estimable functionals are well known, theorem 4.1.1 with $p \in \mathcal{M}(X')$, gives a plausible choice for H_o'. Accordingly, we shall test the hypothesis

$$H_o' : p'\beta \in [(p'\beta)_1, (p'\beta)_2] \qquad (4.1.10)$$

where

$$(p'\beta)_1 = \frac{cp'q}{q'q} - \Delta\lambda \qquad (4.1.11)$$

$$(p'\beta)_2 = \frac{cp'q}{q'q} + \Delta\lambda \qquad (4.1.12)$$

Since $p \in \mathcal{M}(X')$, we shall always use the representation

$$p = X'b \qquad (4.1.13)$$

for some $b \in R^n$.

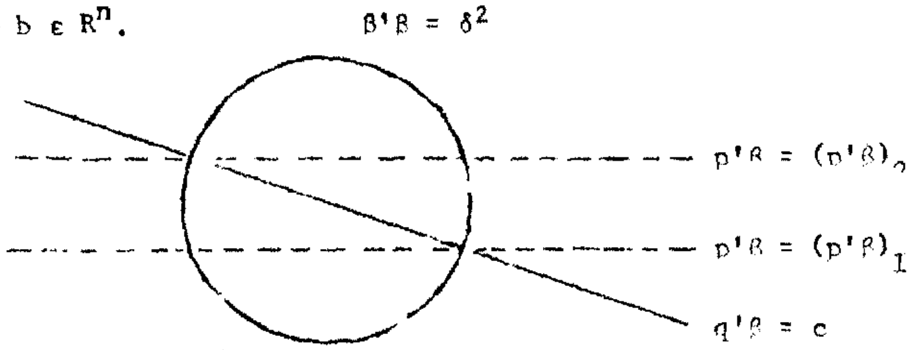

Figure 4.1.1

The replacement of H_o by H_o' is brought out clearly in figure 4.1.1. Instead of testing for the plane $q'\beta = c$, we shall test for the strip between $p'\beta = (p'\beta)_1$ and $p'\beta = (p'\beta)_2$. Observe that if $\Lambda^2 = 0$, $q'\beta$ is a tangent plane to the sphere $\beta'\beta = \delta^2$. In this case H_o is equivalent to $H_o^* : \beta = \frac{cq}{q'q}$.

To obtain a test for H_o' we use the BLUE of $p'\beta$ to test $p'\beta = \frac{cp'q}{q'q}$ with the critical value so determined that the test is of size α, i.e. the maximum power of the test under H_o' is α. If $\hat{\beta}$ is a solution of the normal equations

$$X'X\beta = X'Y \tag{4.1.14}$$

then the BLUE of $p'\beta$ is $p'\hat{\beta}$. In case $\sigma^2 = \sigma_o^2$ is known, the test is

$$\text{Reject if} \quad \tau = \frac{p'\hat{\beta} - cp'q/q'q}{\sigma_o v} > \tau_\varepsilon$$

$$\text{or} \quad \ldots\ldots\ldots < -\tau_\varepsilon \tag{4.1.15}$$

where

$$v^2 = p'(X'X)^- p = b'P_X b \tag{4.1.16}$$

τ is normal with

$$E(\tau) = \frac{p'\beta - cp'q/q'q}{\sigma_o v}, \; V(\tau) = 1 \tag{4.1.17}$$

Hence the power function of the test (4.1.15) is symmetric about $\frac{cp'q}{q'q}$ and is monotonically increasing as one moves away from $\frac{cp'q}{q'q}$ in either direction. Thus if one chooses τ_ε so that the power at $(p'\beta)_2$ is α then the power at $(p'\beta)_1$ is also α, and the resultant test is size α for H_o'. Since $E(\tau) = \frac{\Delta\lambda}{\sigma_o v}$ when $p'\beta = (p'\beta)_2$, τ_ε is determined from

$$1 - \Phi(\tau_\varepsilon - \frac{\Delta\lambda}{\sigma_o v}) + \Phi(-\tau_\varepsilon - \frac{\Delta\lambda}{\sigma_o v}) = \alpha \tag{4.1.18}$$

where $\Phi(\cdot)$ is the distribution function of a $N(0,1)$ variate. Writing ξ for the test function of this test, note that this means $\sup_{\beta:q'\beta=c} E(\xi) = \alpha$, so that the test is size α for H_o.

Equations (4.1.15) and (4.1.18) describe the test procedure for a fixed choice of the approximating functional $p'\beta$. Since p can be any vector in $\mathcal{M}(X')$ (equivalently, by equation (4.1.13), b can be any vector in R^n) one has to obtain an optimal test procedure. Obviously the test which has maximum power at alternatives will be best in an acceptable sense. If

$$H_1 : q'\beta = c_1 \tag{4.1.19}$$

denotes an alternative of interest, then by theorem 4.1.1, the corresponding hypothesis stated in terms of $p'\beta$ is

$$H_1' : p'\beta \in [c_1 \frac{p'q}{q'q} - \Delta_1\lambda \ , \quad c_1 \frac{p'q}{q'q} + \Delta_1\lambda] \tag{4.1.20}$$

where

$$\Delta_1^2 = \delta^2 - \frac{c_1^2}{q'q} \tag{4.1.21}$$

Clearly, one has to consider only those alternatives for which $\Delta_1^2 \geq 0$, since otherwise H_1 is not a plausible alternative.

If $\Delta_1^2 = 0$, the alternative H_1 is equivalent to $\beta = \frac{c_1}{q'q} q$. Here (4.1.20) collapses to the point $c_1 \frac{p'q}{q'q}$, so that the power of the test at the alternative H_1 is the power at $p'\beta = c_1 \frac{p'q}{q'q}$. However, in general H_1 is not equivalent to a single β - point, and hence (4.1.20) will be a nondegenerate interval. To compare powers of various tests, i.e. for different $p \in \mathcal{M}(X')$, one has thus to define a measure for the power of a test under H_1. One such measure can be the maximum power attained by the test under H_1, where it achieves maximum discrimination between H_o and H_1. Note that this is the maximum power attained by the test (4.1.15) in the interval (4.1.20). Observe that the maximum power is attained at $c_1 \frac{p'q}{q'q} + \Delta_1\lambda$ if $c_1 \frac{p'q}{q'q} > c \frac{p'q}{q'q}$, and at $c_1 \frac{p'q}{q'q} - \Delta_1\lambda$, otherwise. Hence the maximum power in (4.1.20) is

$$g_1(p,c_1) = 1 - \Phi(\tau_\varepsilon - \left|\frac{c_1-c}{\sigma_o\sqrt{q'q}} p'q\right| - \Delta_1\lambda) + \Phi(-\tau_\varepsilon - \left|\frac{c_1-c}{\sigma_o\sqrt{q'q}} p'q\right| - \Delta_1\lambda) \tag{4.1.22}$$

To obtain the best test one has to maximize (4.1.22) subject to the restriction (4.1.18) among all $p \in \mathcal{M}(X')$, i.e. $b \in R^n$. A test which is optimal in this sense will be called the 'best test of type I' for the alternative c_1.

A point to note is that there can be situations where the maximum power of a test in (4.1.20) is less than the size α. By equations (4.1.18) and (4.1.22) this happens when

$$\frac{|c_1-c|}{q'q} |p'q| + \Delta_1\lambda < \Delta\lambda \tag{4.1.23}$$

Clearly this can hold only when $\Delta_1 < \Delta$, i.e. $c_1 \notin (-|c|, |c|)$. In this situation, (4.1.23) holds for all $p \in \mathcal{M}(X')$, if and only if

$$XX' - (\frac{a_1^2}{a_2^2} + \frac{1}{q'q})Xqq'X' \text{ is p.d.} \tag{4.1.24}$$

where $a_1 = \frac{|c_1-c|}{q'q}$ and $a_2 = \Delta - \Delta_1$. When (4.1.24) holds there exists no test of the form $p'\hat{\beta}$ which discriminates between H_o and H_1. In particular, this would be the case when $Xq = 0$, i.e. $q'\beta$ is totally nonestimable. Writing $p = p_o = P_{X'}q$, $s_o = q - p_o$ in (4.1.23) one obtains

$$\sqrt{s_o's_o} > \frac{a_1 \sqrt{p_o'p_o} \sqrt{q'q}}{a_2} \tag{4.1.25}$$

Thus if $c_1 \notin (-|c|, |c|)$ and $q'\beta$ is totally nonestimable, or close to being totally nonestimable (in the sense of small $p_o'p_o$), then our test procedures may not be able to discriminate between H_o and H_1. However in all other situations, one expects the maximum power in (4.1.20) for some $p'\hat{\beta}$ to exceed α. In this sense, there exists some test function which discriminates between H_o and H_1. Even the minimum power of $p'\hat{\beta}$ in (4.1.20) can exceed α if $|c_1-c|$ is large.

Another approach to finding a best test could be to maximize power under the alternative $\beta = \frac{c_1}{q'q} q$, which is the minimum norm solution of (4.1.19). Thus, here there is a common basis for comparing power of tests constructed from $p'\hat{\beta}$, p varying in $\mathcal{M}(X')$. Observe that this corresponds to maximizing power of tests (4.1.15) at $p'\beta = c_1 \frac{p'q}{q'q}$, which is the central plane of (4.1.20). The best test obtained by this method will be called the 'best test of type II' for the alternative c_1. It is easy to see that such a test is obtained by maximizing

$$g_2(p,c_1) = 1-\Phi(\tau_\varepsilon - \frac{c_1-c}{\sigma_o vq'q} p'q) + \Phi(-\tau_\varepsilon - \frac{c_1-c}{\sigma_o vq'q} p'q) \qquad (4.1.26)$$

subject to (4.1.18).

When σ^2 is unknown, one can use the BLUE of $p'\hat{\beta}$ and construct a t-statistic to test H_1'. Let $X \in R^{n\times m}$ be of rank r. Denoting $\nu = n-r$, $R_o^2 = (Y-X\hat{\beta})'(Y-X\hat{\beta})$, the test is

$$\text{Reject if} \qquad t = \frac{p'\hat{\beta} - cp'q/q'q}{v\sqrt{(R_o^2/\nu)}} > t_\varepsilon$$

$$\text{or} \qquad \ldots\ldots \qquad < -t_\varepsilon \qquad (4.1.27)$$

t follows a noncentral t distribution with ν degrees of freedom and noncentrality parameter (n.c.p.)

$$\mu(p'\beta) = \frac{p'\beta - cp'q/q'q}{\sigma v} \qquad (4.1.28)$$

Observe that the power of a t-test increases with $|\mu(p'\beta)|$ and the power function is symmetric about $cp'q/q'q$. To obtain a size α test for H_o' one has to determine t_ε such that

$$\sup_{\substack{p'\beta \varepsilon \text{ interval } (4.1.20) \\ \sigma^2 \varepsilon \Omega_\sigma}} \{1-F(t_\varepsilon / \nu, \mu(p'\beta)) + F(-t_\varepsilon / \nu, \mu(p'\beta))\} = \alpha$$

where $F(\cdot/\nu,\mu(p'\beta))$ denotes the distribution function of a noncentral t-variate. If $\Omega_\sigma = [\sigma_o^2, \infty)$ then this equation can be written as

$$1 - F(t_\varepsilon / \nu, \frac{\Delta\lambda}{\sigma_o v}) + F(-t_\varepsilon / \nu, \frac{\Delta\lambda}{\sigma_o v}) = \alpha \qquad (4.1.29)$$

Clearly, the resulting test is also size α for H_o. Observe that this procedure fails if $\sigma_o^2 = 0$. If $\Omega_\sigma = (\sigma_o^2, \infty)$ with $\sigma_o^2 > 0$, then (4.1.29) assures only a level of significance α, while the size of the test may be less than α. Also note that, this test can be used only when a knowledge of σ_o^2 is available, possibly from past experience.

The best test of type I for an alternative c_1 maximizes the maximum power under H_1, i.e. it maximizes

$$f_1(p,c_1) = 1 - F(t_\varepsilon / \nu, \left|\frac{(c_1-c)p'q}{\sigma_o v\, q'q}\right| + \frac{\Delta_1\lambda}{\sigma_o v}) + F(-t_\varepsilon / \nu, \left|\frac{(c_1-c)p'q}{\sigma_o vq'q}\right| + \frac{\Delta_1\lambda}{\sigma_o v})$$

(4.1.30)

subject to (4.1.29). The best test of type II for an alternative c_1 may be defined as the test which maximizes the maximum power under the alternative $\beta = \frac{c_1}{q'q} q$, i.e. it maximizes

$$f_2(p,c_1) = 1 - F(t_\varepsilon/\nu, \frac{(c_1-c)p'q}{\sigma_o vq'q}) + F(-t_\varepsilon/\nu, \frac{(c_1-c)p'q}{\sigma_o vq'q})$$

(4.1.31)

subject to (4.1.29).

The process of obtaining best tests involve difficult optimization problems and it is not easy to arrive at analytic solutions. In section 2 we obtain the best tests when q satisfies some conditions. Some important characterestics of the best tests for arbitrary q and some convenient approximations will be discussed in section 3. For convenience we shall denote

$$\mu_1 = \frac{\Delta\lambda}{\sigma_o v}, \quad \mu_2 = \frac{(c_1-c)p'q}{\sigma_o v\, q'q}, \quad \mu_2^* = |\mu_2| + \frac{\Lambda_1 \lambda}{\sigma_o v} \qquad (4.1.32)$$

Though λ, v, τ_ε, t_ε, μ_1, μ_2, μ_2^* are all functions of p, for notational simplicity, we avoid suffixes.

2. The best test for certain hypotheses

Let us fix a $c_1 \neq c$, and consider the problem of obtaining the best test of type II for the alternative $H_1: q'\beta = c_1$. If $\sigma = \sigma_o$ is known, then it is easily seen from equation (4.1.18) that τ_ε is an increasing function of λ^2/v^2. One arrives at a similar conclusion even when σ is unknown, with a knowledge of the lower bound σ_o^2 of Ω_σ only available. This follows from equation (4.1.29), using the fact that the power of a t-test increases with the magnitude of the n.c.p. A study of $g_2(p,c_1)$ and $f_2(p,c_1)$ now reveals

(i) $g_2(p,c_1)$, $f_2(p,c_1)$ increases as $\frac{(p'q)^2}{v^2}$ increases, $\frac{\lambda^2}{v^2}$ remaining fixed.

(ii) $g_2(p,c_1)$, $f_2(p,c_1)$ increases as $\frac{\lambda^2}{v^2}$ decreases, $\frac{(p'q)^2}{v^2}$ remaining fixed (since then $\tau_\varepsilon(t_\varepsilon)$ decreases and the critical region expands).

If there exists a p in $\mathcal{M}(X')$ which simultaneously maximizes $\frac{(p'q)^2}{v^2}$ and minimizes $\frac{\lambda^2}{v^2}$, then clearly $p'\hat{\beta}$ will give the best test of type II, irrespective of whether σ_o is the value assumed by σ or the lower bound of Ω_σ. Let us determine the necessary and sufficient conditions for the existence of such a test.

Firstly, we define certain useful notations. Let

$$p_o = P_{X'}q$$

$$s_o = q - p_o = (I-P_{X'})q \qquad (4.2.1)$$

Note that $p_o = X'b_o$, where

$$b_o = (X^+)'q \qquad (4.2.2)$$

Moreover $q = p_o + s_o$ and

$$q'q = p_o'p_o + s_o's_o \qquad (4.2.3)$$

Also define

$$b_* = Xq = Xp_o \qquad (4.2.4)$$

The positive eigenvalues of XX'(same as the positive eigenvalues of X'X) will be denoted by $d_1 \leq d_2 \leq \ldots \leq d_r$. Write

$$A = X(I- \frac{qq'}{q'q})X' = X(I- \frac{p_op_o'}{q'q})X' \qquad (4.2.5)$$

$$= XX' - \frac{1}{q'q}(Xp_o)(Xp_o)' \qquad (4.2.6)$$

$$= \frac{s_o'\,s_o}{q'q}XX' + \frac{p_o'p_o}{q'q}\;X(I- \frac{p_op_o'}{p_o'p_o})X' \qquad (4.2.7)$$

Observe that $\mathcal{M}(A) = \mathcal{M}(X\ )$ when $q \notin \mathcal{M}(X')$. Since any p in $\mathcal{M}(X')$ can be written as

$$p = X'b \qquad (4.1.13)$$

for some b in R^n, the optimization problem can be expressed as one with the entire R^n as its domain. In other words, one has to determine a 'b' in R^n which simultaneously minimizes $\frac{\lambda^2}{v^2} = \frac{b'Ab}{b'P_Xb}$ and maximizes $\frac{(p'q)^2}{v^2} = \frac{(b'Xq)^2}{b'P_Xb}$ If $p_o = 0$ then $Xq = Xp_o = 0$, $\frac{\lambda^2}{v^2} = \frac{b'XX'b}{b'P_Xb}$ Obviously the best test of type II would be given by $b_1'X\hat{\beta}$, where b_1 is an eigenvector of XX' corresponding to the eigenvalue d_1. However, in this situation $g_2(p,c_1)$, $f_2(p,c_1)$ is less than α at all alternatives. On the other hand if $q'\beta$ is estimable, then surely $q'\hat{\beta}$ will give the best test. Having settled these two extreme cases, henceforth we shall be interested only in situations where $p_o'p_o \neq 0$, $s_o's_o \neq 0$.

Observe that $\frac{(b'Xq)^2}{b'P_Xb} \leq q'X'Xq = b_*'b_*$, with equality if and only if $b = ab_*$ for some nonzero 'a'. Thus the best approximating estimable functional, if it exists, must be a multiple of $X'b_*$. Moreover $\frac{\lambda^2}{v^2} = \frac{b'Ab}{b'P_Xb}$ is minimized if and only if b is a proper eigenvector of A with respect to P_X, corresponding to its smallest proper eigenvalue (see, for example, Rao and Mitra [33, p 124] for a definition of proper eigenvectors and eigenvalues). Since $\mathcal{M}(A) = \mathcal{M}(X)$, $(A-\lambda P_X)b = 0 \Leftrightarrow (A-\lambda I)P_Xb = 0$; i.e. b is a proper eigenvector of A with respect to P_X if and only if P_Xb is an eigenvector of A. Hence one has to obtain the necessary and sufficient conditions under which $ab_* = a P_Xb_*$ is an eigenvector of A corresponding to its minimum positive eigenvalue.

<u>Lemma 4.2.1</u> (i) b_* is an eigenvector of A if and only if it is an eigenvector of XX'.

(ii) b_o is an eigenvector of A if and only if it is an eigenvector of XX'.

(iii) $b_* = d_o\, b_o$ for some $d_o > 0$ if and only if b_* (or b_o) is an eigenvector of XX'.

<u>Proof (i)</u>

$$Ab_* = \lambda b_* \quad \text{for some } \lambda$$

$$\Rightarrow \quad XX'b_* = \left(\lambda + \frac{b_*'b_*}{q'q}\right)b_* \quad \text{from (4.2.6)}$$

Conversely,

$$XX'b_* = db_* \quad \text{for some } d$$

$$\Rightarrow \quad Ab_* = \left(d - \frac{b_*'\,b_*}{q'q}\right)b_*$$

(ii) follows similarly from (4.2.7).

(iii)

$$b_* = Xp_o = XX'b_o = d_ob_o$$

if and only if b_o is an eigenvectors of XX' corresponding to the eigenvalue d_o. The lemma is established once it is noted that b_o is an eigenvector of XX' if and only if b_* is an eigenvector of XX'. Note that $d_o > 0$, whenever $p_o'p_o \neq 0$.

An interesting fact to note from equations (4.1.18), (4.1.26) and (4.1.29), (4.1.31) is that best tests of type II are invariant under scalar multiplication of p by nonzero scalars. For the problem under consideration, lemma 4.2.1 shows that $b_*'X\hat{\beta}$ or any nonzero multiple of it will give the best test if and only if $b_o'X\beta$ is the best test. $b_o'X\hat{\beta}=p_o'\beta$ is our familiar estimator BLIMBE. In the following theorem we find conditions for b_o to minimize λ^2/v^2, or in other words, some conditions for BLIMBE to be the best test of type II.

<u>Theorem 4.2.1</u> b_o minimizes λ^2/v^2 if and only if it is an eigenvector of XX' corresponding to an eigenvalue $d_o \epsilon [d_1, \frac{q'q}{s_o's_o} d_1]$.

<u>Proof</u> Necessity : Let b_o minimize λ^2/v^2. Then $b_o = P_X b_o$ is an eigenvector of A and hence of XX' (by lemma 4.2.1) corresponding to the eigenvalue $\frac{b_o'XX'b_o}{b_o'b_o} = \frac{p_o'p_o}{b_o'b_o} = d_o$ of XX'. When $b=b_o$, $\frac{\lambda^2}{v^2} = \frac{s_o's_o}{q'q} d_o$, from (4.2.7). For any other b in R^n,

$$\frac{\lambda^2}{v^2} = \frac{b'XX'b}{b'P_Xb} - \frac{(b'Xp_o)^2}{q'q(b'P_Xb)} \geq \frac{s_o's_o}{q'q} d_o$$

$$\Rightarrow \frac{b'XX'b}{b'P_Xb} \geq \frac{s_o's_o}{q'q} d_o \text{ , for all b in } R^n$$

$$\Rightarrow d_1 = \min_{b \,\epsilon\, R^n} \frac{b'XX'b}{b'P_Xb} \geq \frac{s_o's_o}{q'q} d_o$$

$$\Rightarrow d_1 \leq d_o \leq \frac{q'q}{s_o's_o} d_1 .$$

Sufficiency : Let b_o be an eigenvector of XX' corresponding to the eigenvalue $d_o \in [d_1, \frac{q'q}{s_o's_o} d_1]$. By lemma 4.2.1, b_o is also an eigenvector of A corresponding to the eigenvalue

$$\frac{s_o's_o}{q'q} d_o \leq d_1 \tag{4.2.8}$$

Suppose b_o does not minimize λ^2/v^2, but $\tilde{b}$ does. Then $P_X\tilde{b}$ is an eigenvector of A corresponding to the eigenvalue

$$\eta = \underset{b \in R^n}{\text{Min}} \frac{\lambda^2}{v^2} < \left.\frac{\lambda^2}{v^2}\right|_{b=b_o} = \frac{s_o's_o}{q'q} d_o \tag{4.2.9}$$

Note that when $b = \tilde{b}$,

$$\frac{\lambda^2}{v^2} = \frac{\tilde{b}'XX'\tilde{b}}{\tilde{b}'P_X\tilde{b}} = \eta \tag{4.2.10}$$

from (4.2.5), since $\tilde{b}'XX'b_o = d_o\tilde{b}'P_Xb_o = 0$, because $P_X\tilde{b}$ and P_Xb_o are eigenvectors of A corresponding to distinct eigenvalues. From (4.2.8) and (4.2.9), $\eta < d_1$, which is a contradiction since from (4.2.10) we have

$$\eta = \frac{\tilde{b}'XX'\tilde{b}}{\tilde{b}'P_X\tilde{b}} \geq \underset{b \in R^n}{\text{Min}} \frac{b'XX'b}{b'P_Xb} = d_1 .$$

Thus b_o minimizes λ^2/v^2.

Note 4.2.1 The condition of the theorem can be stated equivalently as "p_o is an eigenvector of X'X corresponding to an eigenvalue $d_o \in [d_1, \frac{q'q}{s_o's_o} d_1]$".

To carry out a similar exercise for the best test of type I for an alternative $c_1 \neq c$, observe that (see equations (4.1.18), (4.1.22) and (4.1.29), (4.1.30)), one has to obtain a 'b' in R^n which simultaneously minimizes λ^2/v^2 and maximizes

$$\frac{h}{v} = \frac{1}{v}\left[\frac{|c_1 - c|}{q'q} |p'q| + \Delta_1 \lambda \right] \qquad (4.2.11)$$

Note that best tests of type I are also invariant under scalar multiplication by nonzero scalars. If b_m is the solution, then clearly b_m has to be a multiple of b_*. Otherwise $db_*(d \neq 0)$, which maximizes $\frac{|p'q|}{v}$ (and also has a larger $\frac{\lambda^2}{v^2}$) will lead to a larger value for h/v. When $b_m = d\, b_*$ minimizes λ^2/v^2, it has to be an eigenvector of A, and hence (by lemma 4.2.1) of XX'. In such a case $d = \frac{b_o'b_o}{p_o'p_o}$, clearly leads to the choice b_o for b_m. Thus a b_m which simultaneously minimizes λ^2/v^2 and maximizes h/v, if it exists, has to be a multiple of b_o. The conditions of theorem 4.2.1 are obviously necessary since they guarantee that b_o minimizes λ^2/v^2. However, they may not be sufficient to ensure that b_o maximizes h/v also. To obtain complete necessary and sufficient conditions, one has to determine when b_o maximizes h/v, given that b_o is an eigenvector of XX' corresponding to an eigenvalue d_o in $[d_1, \frac{q'q}{s_o's_o} d_1]$.

Let $t_1 = b_o/\sqrt{b_o'b_o}$, $t_2, \ldots, t_r$, t_{r+1}, ... , t_n be a set of orthonormal eigenvectors of A, corresponding to the eigenvalues $\eta_1, \ldots, \eta_n$, where $0 < \eta_1 \leq \eta_2 \leq \cdots \leq \eta_r$ and $\eta_{r+1} = \cdots = \eta_n = 0$ (the smallest positive eigenvalue corresponds to t_1 since b_o minimizes $\frac{b'Ab}{b'P_Xb}$). From (4.2.7),

$$\eta_1 = \frac{s_o's_o}{q'q} \frac{p_o'p_o}{b_o'b_o} = \frac{s_o's_o}{q'q} d_o \leq d_1 \qquad (4.2.12)$$

and from (4.2.6)

$$\eta_i = d_i \qquad (4.2.13)$$

for $i = 2, \ldots, r$. Writing $T = (t_1, \ldots, t_n)$, $G = \text{diag}(\eta_1, \ldots, \eta_n)$, observe that A has the spectral decomposition

$$A = T\, G\, T' \qquad (4.2.14)$$

Transforming b to $T'b = x = (x_1,\ldots,x_n)'$ note that since
$\mathcal{M}(A) = \mathcal{M}(X)$, $v^2 = b'P_X b = b'P_A b = \sum_{i=1}^{r} x_i^2$

Moreover

$$\frac{h}{v} = \frac{1}{v}[a|x_1| + \Delta_1 \sqrt{\sum_1^r \eta_i x_i^2}] \tag{4.2.15}$$

where $a = \frac{|c_1-c|}{q'q} d_o \sqrt{b_o' b_o} > 0$, $\Delta_1^2 = \delta^2 - \frac{c_1^2}{q'q} \geq 0$. b_o maximizes h/v if and only if $|x_1| = 1$, $x_2 = \ldots = x_r = 0$ maximizes h under the restriction $\sum_{i=1}^{r} x_i^2 = 1$. Fixing the value of $|x_1|$ at $\theta \in [0,1]$, i.e. $|x_1|=\theta$, $\sum_{i=2}^{r} x_i^2=1-\theta^2$, one observes that h is maximized when $x_2 = \ldots = x_{r-1} = 0$ and $x_r^2 = 1 - \theta^2$. Hence it is enough to determine the necessary and sufficient conditions for $\theta = 1$ to maximize

$$\begin{aligned} h(\theta) &= a\theta + \Delta_1 \sqrt{\{\eta_1\theta^2 + \eta_r(1-\theta^2)\}} \\ &= a\theta + \Delta_1 \sqrt{\{\eta_r + (\eta_1-\eta_r)\theta^2\}} \end{aligned} \tag{4.2.16}$$

when θ varies in $[0,1]$. Differentiating $h(\theta)$

$$h_1(\theta) = \frac{dh(\theta)}{d\theta} = a + \frac{\Delta_1(\eta_1-\eta_r)\theta}{\{\eta_r + (\eta_1-\eta_r)\theta^2\}^{1/2}} \tag{4.2.17}$$

$$h_2(\theta) = \frac{d^2h(\theta)}{d\theta^2} = \frac{\Delta_1(\eta_1-\eta_r)}{\{\eta_r+(\eta_1-\eta_r)\theta^2\}^{1/2}} - \frac{\Delta_1(\eta_1-\eta_r)^2\theta^2}{\{\eta_r+(\eta_1-\eta_r)\theta^2\}^{3/2}} \tag{4.2.18}$$

Let $\eta_o^2 = \frac{\eta_r}{\eta_r-\eta_1} > 1$. $h_1(\theta)$ and $h_2(\theta)$ are well defined in $(-\eta_o, \eta_o) \supset [0,1]$, and in this interval $h_2(\theta) < 0$. Hence $h(\theta)$ is strictly concave in $(-\eta_o,\eta_o)$. Thus $h(1) \geq h(\theta)$, for all θ in $[0,1]$ if and only if $h(\theta)$ is increasing in a neighbourhood of $\theta = 1$, or attains its maximum at this point. An equivalent way of stating this is

$$h_1(1) \geq 0$$

$$\Leftrightarrow \quad \Delta_1(\eta_r - \eta_1) \leq a\sqrt{\eta_1}$$

$$\text{i.e.} \quad \Delta_1(d_r \quad \frac{s_o' s_o}{q'q} d_o) \leq \frac{d_o|c_1-c|}{q'q} \surd \{ \frac{(s_o's_o)(p_o'p_o)}{q'q} \} \tag{4.2.19}$$

which is the required condition.

Summarizing these results we state the main theorem of this section.

<u>Theorem 4.2.2</u> The BLIMBE $p_o'\beta$ gives the best test of type II for the hypothesis $H_o: q'\beta = c$ when

(i) p_o is an eigenvector of X'X corresponding to an eigenvalue $d_o \in [d_1, \frac{q'q}{s_o's_o} d_1]$, $d_1 \leq d_2 \leq \ldots \leq d_r$ being the positive eigenvalues of X'X, $p_o = P_{X'}q$ and $s_o = q - p_o$.

The BLIMBE gives the best test of type I for H_o for an alternative c_1 when in addition to (i) the following holds

$$\text{(ii)} \quad \Delta_1(d_r - \frac{s_o' s_o}{q'q} \; d_o) < \frac{|c_1-c|}{q'q} d_o \surd \{ \frac{(s_o' s_o)(p_o' p_o)}{q'q} \}.$$

<u>Note 4.2.2</u> Under the conditions of the theorem $p_o'\beta$ is the best test of type II for H_o whatever c_1. Hence it is uniformly best. However, this is not true for the best test of type I.

Before concluding this section, we establish a result, which though we do not have an occasion to use, deserves attention due to its own merit. Theorem 4.2.3 is complementary to theorem 4.2.1; it investigates the situation when b_o maximizes λ^2/v^2, given $p_o'p_o \neq 0$.

<u>Theorem 4.2.3</u> If d_r is of multiplicity one as an eigenvalue of XX', then b_o maximizes λ^2/v^2 if and only if it is an eigenvector of XX' corresponding to d_r, and $d_{r-1} \leq \frac{s_o' s_o}{q'q} d_r$. However, if d_r is of multiplicity greater than one, then b_o cannot maximize λ^2/v^2.

<u>Proof</u> <u>d_r is of multiplicity one</u>

Necessity : If b_o maximizes λ^2/v^2, then it is an eigenvector of A and hence of XX' corresponding to the eigenvalue d_o (say). Moreover $\frac{b_o'Ab_o}{b_o'P_Xb_o} = \frac{s_o's_o}{q'q} d_o$. From (4.2.7), $\frac{b'XX'b}{b'P_Xb} \leq d_o$, whatever $b \in R^n$, so that $d_o = d_r$. Let b_1 be an eigenvector of XX' corresponding to the eigenvalue d_{r-1}. Since $b_1'Xq = b_1'XX'b_o = d_r b_1'b_o = 0$,

$$d_{r-1} = \frac{b_1'XX'b_1}{b_1' b_1} = \frac{b_1'Ab_1}{b_1'P_Xb_1} \leq \frac{s_o's_o}{q'q} d_r .$$

Sufficiency : Under conditions of the theorem if b_2, and not b_o, maximizes λ^2/v^2, then P_Xb_2 is an eigenvector of A with $b_o'b_2 = b_o'P_Xb_2 = 0$. It is not difficult to see that this implies

$$\frac{b_2' Ab_2}{b_2'P_Xb_2} = \frac{b_2'XX'b_2}{b_2'P_Xb_2} \leq d_{r-1} \leq \frac{s_o's_o}{q'q} d_r = \frac{b_o'Ab_o}{b_o'P_Xb_o} ,$$

since d_r is of multiplicity one. This is a contradiction.

<u>d_r is of multiplicity strictly greater than one</u>

Let b_3 be an eigenvector of XX' corresponding to the eigenvalue d_r. Further let $b_3'b_o = 0$. Clearly,

$$\frac{b_3'Ab_3}{b_3'P_Xb_3} = d_r \geq \frac{p_o'p_o}{b_o'b_o} > \frac{s_o's_o}{q'q} \frac{p_o'p_o}{b_o'b_o} = \frac{b_o'Ab_o}{b_o'P_Xb_o} ,$$

since $p_o'X'b_3 = b_o'XX'b_3 = d_r b_o'b_3 = 0$.

3. <u>The best test.</u>

In general, if an analytic expression for τ_ε in terms of b is available, then an equation **satisfied by** the best test of type II may be obtained by setting the derivatives of $g_2(p, c_1)$ with respect to b, equal to zero. Using the notation defined by equation (4.1.32) this equation is

$$[1 + e^{-2\tau_\varepsilon \mu_2}] \frac{d\tau_\varepsilon}{db} = [1 - e^{-2\tau_\varepsilon \mu_2}] \frac{d\mu_2}{db} \tag{4.3.1}$$

where

$$\frac{d\mu_2}{db} = \frac{\mu_2}{p'q} Xq - \frac{\mu_2}{v^2} P_X b \tag{4.3.2}$$

Unfortunately, the desired analytic expression is difficult to obtain from equation (4.1.18). However, several approximations are available in literature, and using these one may arrive at an approximate best test of type II for an alternative c_1. By Patnaik's [26] (see also Zelen and Severo [39], formula 26.4.31) normal approximation to the noncentral chisquare distribution,

$$\tau_\varepsilon \simeq \frac{\tau_\alpha \sqrt{1+2\mu_1^2} + \sqrt{1+2\mu_1^2 + 2\mu_1^4}}{\sqrt{2(1+\mu_1^2)}} \tag{4.3.3}$$

where τ_α is the upper α % point of the standard normal distribution and μ_1 is defined by equation (4.1.32). Hence

$$\frac{d\tau_\varepsilon}{db} = \frac{\sqrt{2}}{v^2\sqrt{1+\mu_1^2}} [\frac{\tau_\alpha}{\sqrt{1+2\mu_1^2}} + \frac{1+2\mu_1^2}{\sqrt{1+2\mu_1^2+2\mu_1^4}} + \frac{\tau_\varepsilon}{\sqrt{2(1+\mu_1^2)}}][\frac{\Delta^2}{\sigma_o^2} A - \mu_1^2 I] P_X b \tag{4.3.4}$$

A being the matrix defined by equation (4.2.5).

Observe that a best test of type I for an alternative c_1 is also obtained by maximizing

$$1 - \Phi(\tau_\varepsilon - \mu_2^{**}) + \Phi(-\tau_\varepsilon - \mu_2^{**}) \qquad (4.3.5)$$

subject to (4.1.18), where

$$\mu_2^{**} = \frac{|c_1 - c|}{\sigma_o q'q} \frac{p'q}{v} + \frac{\Delta_1}{\sigma_o} \frac{\lambda}{v} \qquad (4.3.6)$$

This is due to the fact that if the maximum of (4.3.5) is attained at $p = \tilde{p}$ then surely $\tilde{p}'q > 0$ and hence $\tilde{p}$ maximizes $\sigma_1(p, c_1)$ (see equation (4.1.22)) subject to (4.1.18). Thus an equation determining an approximate best test of type I would be given by (4.3.1) with μ_2 replaced by μ_2^{**} and $\frac{d\mu_2}{db}$ replaced by

$$\frac{d\mu_2^{**}}{db} = \frac{|c_1 - c|}{\sigma_o q'qv} [Xq - \frac{p'q}{v^2} P_X b] + \frac{\Delta_1}{\sigma_o \lambda v} [A\ b - \frac{\lambda^2}{v^2} P_X b] \qquad (4.3.7)$$

Equation (4.3.1) is quite complicated in structure and computations required for obtaining a solution would be laborious. We do not attempt a similar exercise for the situation when σ is unknown since the corresponding equation is likely to be even more complicated. Another important observation is that the best test apparently depends on the value assumed by the alternative c_1. A test which is not uniformly best will have limited use in practice, since an experimenter is generally interested in testing a hypothesis H_o against all possible alternatives. In such a case, the best that one can hope to obtain is possibly a test which is independent of the alternative and is close to the best test, in some sense, whatever c_1.

To get some idea about the nature of the best tests we shall study a numerical example. One of the principal aims is to determine whether the best test for an alternative c_1, is uniformly best or not. Recall that, even when conditions of theorem 4.2.2 hold, the best test of type I, unlike the best test of type II, need not be uniformly best.

Using the singular value decomposition of the design matrix, in a manner demonstrated in chapter 2, section 3, observe that it is sufficient to consider models of the type

$$E(Y_i) = \alpha_i \beta_i, \quad i = 1,\dots,r$$

$$E(Y_i) = 0 \quad , \quad i=r+1,\dots,n$$

$$\mathrm{Cov}(Y_i,Y_j) = \sigma^2\delta_{ij} \tag{4.3.8}$$

where $\sum_{i=1}^{m} \beta_i^2 \leq \delta^2$, $m \geq r$, and δ_{ij} is the Kronecker symbol, $i,j=1,\dots,n$. The hypothesis to be tested is

$$\sum_{i=1}^{m} q_i \beta_i = c \tag{4.3.9}$$

We consider an example where $r = 2$, $d_1 = \alpha_1^2 = 1$, $d_2 = \alpha_2^2 = 2$, $\nu = n-r=6$, $q_1 = q_2 = 1$, $\sum_{i=1}^{m} q_i^2 = 3$ and $\delta^2 = 9$. For the case of known σ, we take $\sigma^2 = \sigma_o^2 = 1$. When σ is unknown, we take $\sigma_o^2 = 1$ as the lower bound of the parameter space Ω_σ for σ^2. We test H_o at the 5% level of significance, i.e. $\alpha = .05$.

To find the best test we first obtain the test $p_d'\beta$ which maximizes power subject to the constraint $\frac{p'p}{v^2} = d$, for each d in $[d_1,d_2]$ and then choose the one with largest power in the class $\{p_d'\beta: d\varepsilon[d_1,d_2]\}$. If $p_*'\hat{\beta}$ is the best test, and $v_*^2 = p_*'(X'X)^- p_*$, then

$$d_* = \frac{p_*'p_*}{v_*^2} \tag{4.3.10}$$

will denote the 'optimal' d. Writing $p' = (\alpha_1 b_1, \alpha_2 b_2, 0, \ldots, 0)$, $e_1^2 = \frac{d-d_1}{d_2-d_1}$, $e_2^2 = 1-e_1^2$, it is easy to see that two solutions to

$$d = \frac{p'p}{v^2} = \frac{d_1 b_1^2 + d_2 b_2^2}{b_1^2 + b_2^2} \tag{4.3.11}$$

are

$$\begin{aligned} &(a) \qquad b_1 = e_1, \quad b_2 = e_2 \\ &(b) \qquad b_1 = e_1, \quad b_2 = -e_2 \end{aligned} \tag{4.3.12}$$

All other solutions are nonzero multiples of (a) or (b). Thus it is enough to consider only these two alternatives at each d in $[d_1, d_2]$. In fact for tests of type II, in this example, (a) has the highest $\frac{(p'q)^2}{v^2}$, and hence maximizes power subject to the constraint $\frac{p'p}{v^2} = d$. Moreover, it turned out that, in all the cases considered, (a) lead to the maximum power for tests of type I also. Clearly, it is impossible to exhaust all points d in $[d_1, d_2] = [1, 2]$. Hence $d = 1(0.1)2$ only were considered, except for some additional points whenever necessary. The powers of $q'X'X\hat{\beta}$ and $p_o'\hat{\beta}$ (BLIMBE) were also computed. The values of $p_1 = \alpha_1 b_1$ and $p_2 = \alpha_2 b_2$ computed according to (a) of (4.3.12) for all the estimable functionals considered are displayed in each table. The critical values τ_ε and t_ε were computed by iterative procedures. They are shown in the tables along with the exact size attained, which are very close to $\alpha = .05$. While interpreting the results of the computations, we shall assume that the power of the test $p_d'\hat{\beta}$ is a concave function of d. This seems quite reasonable - though, unfortunately, we could not establish it theoretically.

Tables 7 and 8 relate to the situation where σ is known. Table 7 gives the best test of type II (columns under 'Power') and type I (columns under 'Max. Power') for the hypothesis $c = -1.0$ against the alternatives $c_1 = 2.0$, 3.0 and 5.0. More explicitly, 'Power' denotes power of the test at $c_1 p'q/q'q$ and 'Max. Power' denotes the maximum power in the interval (4.1.20). For $c_1 = 2.0$ the best test of type II has an optimal d between 1.30 and 1.33, i.e. $d_* \in (1.30, 1.33)$. However for $c_1 = 3.0$, $d_* \in (1.30, 1.40)$ and for $c_1 = 5.0$, $d_* \in (1.33, 1.50)$ indicating thereby the dependence of the best test on the alternative c_1. Note that though the power at $c_1 p'q/q'q$ is often less that .05 (the size), nowhere does the maximum power under H_1' fall below .05. Thus all the tests are able to discriminate between the hypothesis and the alternatives considered. Not surprisingly, the dependence of the best test of type I for the alternative c_1 seems more pronounced. For example $d_* > 1.5$ when $c_1 = 2.0$ and $d_* < 1.5$ when $c_1 = 5.0$. An interesting fact observed from table 7 is that the best test of type II may not have the highest power under the alternative H_1'. Thus the two criteria I and II, can lead to entirely different tests. Another fact to note is that the maximum power of a test can actually decrease even when $|c_1|$ increases. This is because the length of the interval (4.1.20) decreases as $|c_1|$ increases.

In table 8 we study the best test of type II for two more hypotheses $c = -3.0$ and $c = -5.0$. Observe that in table 7, $c = -1.0$ which means that Δ^2 was close to δ^2. (Note that $0 \leq \Delta^2 \leq \delta^2$). For $c = -5.0$, Δ^2 is close to 0, while for $c = -3.0$, Δ^2 lies between these two extremes. When $c = -3.0$, $d_* < 1.33$ for $c_1 = -2.0$, $d_* \in (1.33, 1.50)$ for $c_1 = 2.0$, $d_* \in (1.40, 1.60)$ for $c_1 = 5.0$. When $c = -5.0$, $d_* < 1.40$ for $c_1 = -4.0$, $d_* \in (1.40, 1.60)$ for $c_1 = 0$ and $d_* \in (1.50, 1.67)$ for $c_1 = 5.0$. Thus the best test of type II need not be uniformly best

Table 7: Showing best tests of types I and II for the hypothesis $c=-1.0$ (σ known)

d	p_1	p_2	τ_ε	SIZE	$c_1 = 2.0$ POWER	$c_1 = 2.0$ MAX POWER	$c_1 = 3.0$ POWER	$c_1 = 3.0$ MAX POWER	$c_1 = 5.0$ POWER	$c_1 = 5.0$ MAX POWER
1.00	1.0000	0	4.0488	.0499	.0011	.2153	.0033	.2370	.0202	.0835
1.10	.9487	.4472	3.6206	.0499	.0131	.3572	.0315	.4149	.2035	.3894
1.20	.8944	.6325	3.5593	.0499	.0220	.4083	.0638	.5275	.3064	.5100
1.30	.8367	.7746	3.5856	.0499	.0242	.4408	.0754	.5706	.3583	.5695
1.40	.7746	.8944	3.6660	.0499	.0229	.4618	.0749	.5952	.3715	.5921
1.50	.7071	1.0000	3.7850	.0499	.0189	.4741	.0656	.6068	.3553	.5882
1.60	.6325	1.0954	3.9339	.0499	.0137	.4789	.0516	.6083	.3163	.5624
1.70	.5477	1.1832	4.1098	.0499	.0087	.4757	.0359	.5983	.2585	.5144
1.80	.4472	1.2649	4.3150	.0499	.0046	.4638	.0211	.5749	.1865	.4404
1.90	.3162	1.3416	4.5649	.0499	.0018	.4361	.0093	.5299	.1058	.3303
2.00	0	1.4142	5.0441	.0499	.0002	.3325	.0008	.3707	.0134	.1015
1.33 (BLIMBE)	1	1	3.6073	.0499	.0242	.4491	.0764	.5804	.3662	.5804
1.67 ($p = X'Xq$)	1	2	4.0488	.0499	.0103	.4777	.0420	.6029	.2796	.5327

Note : 'POWER' denotes power of the test at $c_1 p'q/q'q$, 'MAX POWER' denotes the maximum power under the alternative.

Table 9: Showing the best test of type II for the hypotheses c = - 3.0 and c = - 5.0 (σ known)

			c = - 3.0					c = - 5.0				
d	p_1	p_2	τ_ε	SIZE	POWER AT c_1			τ_ε	SIZE	POWER AT c_1		
					$c_1=-2.0$	$c_1=2.0$	$c_1=5.0$			$c_1=-4.0$	$c_1=0.0$	$c_1=5.0$
1.00	1.0000	0	3.6448	.0499	.0005	.0240	.1641	2.3300	.0495	.0268	.2533	.8421
1.10	.9487	.4472	3.2889	.0499	.0025	.1680	.6679	2.2200	.0491	.0433	.5426	.9925
1.20	.8944	.6325	3.2376	.0499	$.0033^{-}$	.2442	.7979	2.2100	.0496	.0477	.6312	.9980
1.30	.8367	.7746	3.2596	.0499	$.0033^{+}$	.2830	.8501	2.2200	.0492	.0491	.6794	.9992
1.40	.7746	.8944	3.3267	.0499	.0029	.2929	.8697	2.2400	.0491	.0487	.7061	.9996
1.50	.7071	1.0000	3.4257	.0499	.0022	.2806	.8699	2.2600	.0500	.0477	.7207	$.9997^{-}$
1.60	.6325	1.0954	3.5497	.0499	.0015	.2514	.8550	2.3000	.0496	.0444	.7190	$.9997^{+}$
1.70	.5477	1.1832	3.6961	.0499	.0009	.2087	.8212	2.3400	.0501	.0407	.7071	$.9997^{-}$
1.80	.4472	1.2649	3.8668	.0499	.0005	.1555	.7577	2.4000	.0494	.0352	.6751	.9995
1.90	.3162	1.3416	4.0744	.0499	.0002	.0949	.6353	2.4600	.0500	.0296	.6198	.9989
2.00	0	1.4142	4.4732	.0499	$\simeq 0$	.0712	.2413	2.5900	.0499	.0182	.4079	.9832
1.33 (BLIMBE)	1	1	3.2778	.0499	.0032	.2891	.8593	2.2250	.0492	.0492	.6904	.9994
1.67 ($p = X'Xq$)	1	2	3.6448	.0499	.0011	.2242	.8350	2.3300	.0495	.0416	.7112	.9997

Note : Superscript (+) denotes rounding off from above and (-) rounding off from below.

Table 9: Showing the best test of type II for the hypothesis $c=-5.0$ (σ unknown)

d	p_1	p_2	t_ε	SIZE	POWER AT c_1: $c_1=-3.0$	$c_1=-1.0$	$c_1=3.0$	$c_1=5.0$
1.00	1.0000	0	2.924	.0498	.0498	.1277	.4630	.6623
1.10	.9487	.4472	2.790	.0495	.0849	.2636	.7954	.9377
1.20	.8944	.6325	2.770	.0495	.0981	.3157	.8667	.9703
1.30	.8367	.7746	2.778	.0495	.1051	.3460	.8990	.9816
1.40	.7746	.8944	2.408	.0494	.1057	.3620	.9146	.9863
1.49	.7141	.9899	2.838	.0494	.1073+	.3670	.9206	.9880
1.50	.7071	1.0000	2.842	.0494	.1072-	.3671	.9209	.9881
1.55	.6708	1.0488	2.864	.0495	.1059	.3663	.9216	.9883
1.60	.6325	1.0954	2.889	.0496	.1040	.3631	.9207	.9882
1.70	.5477	1.1832	2.942	.0499	.0987	.3505	.9143	.9866
1.80	.4472	1.2649	3.022	.0496	.0892	.3238	.8960	.9816
1.90	.3162	1.3416	3.115	.0495	.0769	.2832	.8584	.9692
2.00	0	1.4142	3.300	.0495	.0495	.1751	.6759	.8703
1.33 (BLIMBE)	1	1	2.785	.0495	.1064	.3527	.9055	.9836
1.67 (p= X'Xq)	1	2	2.924	.0498	.1006	.3557	.9171	.9873

Table 10; Showing the best test of type I for the hypothesis c=-5.0 (σ unknown)

d	p_1	p_2	MAX POWER AT c_1			
			$c_1=-3.0$	$c_2=-1.0$	$c_1=3.0$	$c_1=5.0$
1.00	1.0000	0	.4629	.7658	.9224	.8821
1.10	.9487	.4472	.4731	.8196	.9809	.9743
1.20	.8944	.6325	.4904	.8456	.9897	.9888
1.30	.8367	.7746	.5125	.8681	.9937	.9936
1.40	.7746	.8944	.5377	.8877	.9957	.9956
1.50	.7071	1.0000	.5647	.9047	.9969	.9965
1.60	.6325	1.0954	.5924	.9191	.9975	.9968
1.70	.5477	1.1832	.6206	.9311	$.9979^+$	.9966
1.80	.4472	1.2649	.6427	.9386	$.9979^-$	.9956
1.90	.3162	1.3416	.6638	.9436	.9975	.9927
2.00	0	1.4142	.6759	.9368	.9925	.9630
1.33 (BLIMBE)	1	1	.5207	.8749	.9945	.9945
1.67 ($p=X'Xq$)	1	2	.6111	.9273	.9978	.9967

in general. In fact, d_* seems to be nondecreasing in $|c_1-c|$, for each c.

When σ is unknown exact computations are time consuming - even if an electronic computing system is available. Hence only one hypothesis (c = - 5.0) was considered. Table 9 gives the best test of type II for the alternatives c_1 = - 3.0, -1.0, 3.0 and 5.0, while table 10 gives the maximum power at these alternatives. The conclusions here are similar to those for the case when σ was known. For example, in table 9, $d_* < 1.50$ when $c_1 = -3.0$, but for all other alternatives $d_* \geq 1.50$; while in table 10, $d_* > 1.8$ when $c_1 = -1.0$ and $d_* < 1.7$ when $c_1 = 5.0$. Here also the two criteria I and II do not always coincide. Moreover the maximum power decreases when c_1 increases from 3.0 to 5.0.

It may be observed from these examples that both $q'X'X\hat{\beta}$ and $p_o'\hat{\beta}$ (BLIMBE) generally have powers close to that of the best test of type II.

Observe that (see equations (4.1.18),(4.1.26) and (4.1.29), (4.1.31)) a necessary condition for $p_*'\hat{\beta}$ to give the best test of type II is that there does not exist a functional $p'\beta$ with $\frac{\lambda^2}{v^2}\Big|_p < \frac{\lambda^2}{v^2}\Big|_{p_*}$ and $\frac{(p'q)^2}{v^2}\Big|_p > \frac{(p'q)^2}{v^2}\Big|_{p_*}$. In general (whatever the model and the hypothesis) there will be no sufficient condition independent of c_1, since in such a case a uniformly best test must exist. A $p \in \mathcal{M}(X')$, which maximizes $\frac{(p'q)^2}{\lambda^2}$ satisfies this necessary condition, and it is not difficult to see that the required maximum is attained at $p = p_o$ (or any nonzero multiple of it). When conditions of theorem 4.2.2 hold, the necessary condition is also sufficient, and the BLIMBE $p_o'\hat{\beta}$ in fact gives the uniformly best test of type II. Even otherwise, one can expect BLIMBE to have some nice properties - especially a fairly high power at alternatives. Also note that in situations where $s_o's_o = 0$, $p_o'\beta$ is the best test. Hence if $s_o's_o$

is small compared to $q'q$ BLIMBE would be expected to give a very good test.

On the other hand, if $\Delta^2 = 0$ then surely $q'X'X\hat{\beta}$ gives the best test of type II, and hence it would perform very well for Δ^2 in some neighbourhood of zero also. Moreover if $X'Xq$ has a small λ^2/v^2 value then also $q'X'X\hat{\beta}$ will have a high power, since only those p in $\mathcal{M}(X')$ with a smaller λ^2/v^2 can possible lead to a better test. Since the best test should have a fairly large $(p'q)^2/v^2$ and small λ^2/v^2, $X'Xq$ which maximizes $(p'q)^2/v^2$ cannot have a very poor performance in most situations.

Apparently, it is difficult to obtain similar approximations to the best test of type I. The nonmonotonicity of $g_1(p,c_1)$ and $f_1(p,c_1)$ is also disturbing. Thus, on the whole, one may prefer criterion II to I. By our present state of knowledge we recommend the use of $p_o'\hat{\beta}$ or $q'X'X\hat{\beta}$. In fact, since it satisfies a necessary condition to be the best test of type II, $p_o'\hat{\beta}$ may be preferred over $q'X'X\hat{\beta}$, unless situations, where the latter is known to perform very well, prevail.

When σ is unknown, determination of the critical value t_ϵ of the test (4.1.27), from equation (4.1.29), usually through iterations is a time consuming process. If one does not have access to an electronic computer this would almost be an impossible task. In such situations, one can use an approximation to the exact distribution of the test statistic to obtain an approximate critical value. For example by Patnaik's [26] central F approximation to the noncentral F variate,

$$t_\epsilon^2 \simeq (1 + \mu_1^2)F_{*\alpha} \qquad (4.3.13)$$

where $F_{*\alpha}$ is the upper α % point of F_*, a central F with $\nu_* = \dfrac{(1+\mu_1^2)^2}{1+2\mu_1^2}$ and ν degrees of freedom. When ν_* is a fraction one has to interpolate from the F-ratio tables to obtain $F_{*\alpha}$.

Before concluding, we note that a likelihood ratio test may be constructed using the maximum likelihood estimator developed in chapter 2, section 2. Consider the general situation where σ^2 is unknown and $\Omega_\sigma = (0,\infty)$. The likelihood ratio is

$$L = \left\{ \frac{(Y-X\hat{\beta}_2)'(Y-X\hat{\beta}_2)}{(Y-X\hat{\beta}_{o2})'(Y-X\hat{\beta}_{o2})} \right\}^{n/2} \qquad (4.3.14)$$

where $\hat{\beta}_2$ and $\hat{\beta}_{o2}$ denote the unconditional and conditional (subject to H_o) maximum likelihood estimators of β respectively. Observe that, by equation (4.1.6),

$$\hat{\beta}_{o2} = \frac{c\,q}{q'q} + \left(I - \frac{qq'}{q'q}\right)\hat{z}_2 \qquad (4.3.15)$$

where z_2 is the MLE of z in the model $(Y - \frac{c}{q'q}Xq,\ X(I - \frac{qq'}{q'q})z,\ \sigma^2 I)$ when $z'(I - \frac{qq'}{q'q})z \leq \Delta^2$. In large samples, H_o can be tested by using the chisquare approximation to $-2\log_e L$, whenever $\Delta^2 > 0$.

C H A P T E R 5

ANALYSING DATA FROM CLASSIFICATION MODELS WITH ARBITRARY PATTERNS

In classification models when there are no missing observations, the data can be analysed in a straightforward manner by the usual Analysis of Variance techniques. However, if there are missing observations, all the contrasts within each factor might still be estimable, in which case the design matrix is said to be of maximal rank. In the event that enough observations are missing so that the design matrix is not of maximal rank, some parametric functionals of interest could become nonestimable. Such models may become so complicated in structure, that it could be difficult even to identify the estimable functionals and the confounded effects. Clearly, analysis of such data will not be easy. However, if one has access to a least square generalized inverse of the design matrix, then one can obtain BLUE's for the estimable functionals and unbiased estimates of their variances. Additionally, this inverse could be used to identify the estimable functionals since $p'\beta$ is estimable if and only if $p'X^{-}X = p'$, and one may use the X_{ℓ}^{-} for X^{-}.

Computational aspects of generalized inverses have not been dealt with very extensively in literature (see e.g. Rao and Mitra [33], Chapter 11, for a brief description of some available methods), and at any event it seems to us that such computations are subject to rounding off errors, which might even lead to erroneous conclusions. In this chapter we shall give an algorithm for finding a least square generalized inverse for the design matrix of a data set following a multiple way classification model with completely arbitrary pattern. The algorithm gives the g-inverse without any rounding off error", and can be programmed easily for electronic computers.

Different approaches to analysing a data set following an arbitrary pattern have been considered by many authors. Bose [7] introduced the notion of connectedness in additive two way classification models and gave a necessary and sufficient condition for the design matrix to be of maximal rank. Weeks and Williams [38], Srivastava and Anderson [37], Eccleston and Hedayat [14] discussed concepts of connectedness in additive multiple way classification models.

Calinski [10] gave an iterative formula for analysing data from two-way classification models possessing any arbitrary pattern. But it appears that one is uncertain about the number of steps needed for the required accuracy.

Birkes, Dodge and Seely [3] presented complete results for estimability in classification models. They introduced the R-process which determine what cell expectations are estimable and also an algorithm for finding a basis for each effect in an additive two way classification model.

Our algorithm will make use of the R-process and some theorems on generalized inverses of partitioned matrices established by Mitra and Bhimasankaram [25]. We state these in section 1, where we also write the model explicitly. In the second section we give the algorithm for additive two way classification models. This is extended to the most general classification models in section 3. Everywhere examples will illustrate the algorithm.

1. Preliminaries.

Let $y_{ijk\ldots wh}$ be a collection of independant random variables with a common unknown variance σ^2 and each having expectation of the form

$$E(y_{ijk\ldots wh}) = \alpha_i + \gamma_j + \theta_k + \ldots + \eta_w + (\alpha\gamma)_{ij} + \ldots + (\alpha\gamma\theta\ldots\eta)_{ij\ldots w} \tag{5.1.1}$$

where $i = 1,2,\ldots,a$; $j = 1,2,\ldots,b;\ldots;w = 1,2,\ldots,t$. The index h is an integer between 1 and $n_{ijk\ldots w}$. If $n_{ijk\ldots w} = 0$ then no random variable with subscripts $ijk\ldots w$ occurs in the collection. Thus we are working with a fixed effect multiple way classification model with arbitrary pattern. If X is the design matrix then it may be partitioned as

$$X = (A : B : \ldots T : X_1) \tag{5.1.2}$$

where $A,B,\ldots,T$ are the submatrices associated with the α-, γ-,...,η-effects respectively, and X_1 corresponds to the interactions. Thus if Y is the vector of observations then

$$E(Y) = X\beta \tag{5.1.3}$$

where $\beta' = (\alpha_1,\ldots,\alpha_a,\gamma_1,\ldots,\gamma_b,\ldots, \eta_1,\ldots,\eta_t , (\alpha\gamma)_{11},\ldots,(\alpha\gamma\ldots\eta)_{ab\ldots t})$.

For completeness we state the R-process of Birkes, Dodge and Seely [3], and theorems on least square g-inverses due to Mitra and Bhimasankaram [25] (see also pp 64-66 of Rao and Mitra [33]), as we shall use them extensively later on.

R-process Let N be the incidence matrix of a two way classification model data. Then the R-process is applied on N to obtain the final matrix M as follows:

1) Set m_{ij} equal to zero or one according as n_{ij} is zero or non-zero.

2) Change any zero m_{ij} to one if there exists k and ℓ such that $m_{i\ell} = m_{k\ell} = m_{kj} = 1$. (Pictorially, add the fourth corner whenever three corners of a rectangle appear in the matrix).

3) Continue step 2, using both the original and the new nonzero m_{ij}'s as corners of new rectangles, until no new entries can be changed.

Theorems on least square g-inverses

Theorem A1 : Let (G:b) be the least square g-inverse of $\binom{A}{a'}$, where a,b are column vectors and A,G are matrices of appropriate dimensions. Further let $a \in \mathcal{M}(A')$ and $b'a \neq 1$. Then $Y = (I + \frac{ba'}{1-b'a})G$ is a least square g-inverse of A.

Theorem A2 : Let G be a least square g-inverse of A, and let a be a column vector such that $a \in \mathcal{M}(A')$. Let $d = G'a$ and $b = \frac{GG'a}{1+a'GG'a}$. Then $X = (G-bd':b)$ is a least square g-inverse of $\binom{A}{a'}$.

Theorem A3 : Let G be a least square g-inverse of A, and let a be a column vector such that $a \in \mathcal{M}(A)$. Let $d = Ga$. Then $X = \binom{G-db'}{b'}$, with b arbitrary is a least square g-inverse of (A : a).

Theorem A4 : Let G be a least square g-inverse of A and let a be a column vector such that $a \notin \mathcal{M}(A)$. Let $d = Ga$, $c = (I-AG)a$ and $b = c/c'a$. Then $X = \binom{G-db'}{b'}$ is a least square g-inverse of (A : a).

Theorem A5 : Let G be a least square g-inverse of A and let a be a column vector such that $a \notin \mathcal{M}(A')$. Let $d = G'a$, $c = (I-GA)(I-GA)'a$ and $b = c/c'a$. Then $X = (G-bd' : b)$ is a least square g-inverse of $\binom{A}{a'}$.

2. Least square generalized inverses for additive two way models.

In additive models interaction terms are absent and hence here the model is

$$E(y_{ijh}) = \alpha_i + \gamma_j \tag{5.2.1}$$

The design matrix is $X = (A:B)$ and the incidence matrix $N = (n_{ij}) = A'B$. We assume that,

$$\begin{aligned} \sum_j n_{ij} &\neq 0 \quad \text{for} \quad i = 1,2,\dots,a \\ \sum_i n_{ij} &\neq 0 \quad \text{for} \quad j = 1,2,\dots,b \end{aligned} \tag{5.2.2}$$

i.e. there is at least one observation in each row and column.

(i) The algorithm.

The algorithm for finding a least square generalized inverse for such models consists of the following steps.

Step 1 : Apply the R-process to the incidence matrix N in order to obtain a final matrix M as introduced by Birkes, Dodge and Seely [3].

Step 2 : Construct the design matrix X* corresponding to M.

Note that X* differs from the original design matrix X in having some extra rows corresponding to those cells which were filled by the R-process and with no row being repeated. We can partition the final matrix M into s sets of connected portions as defined by Bose [7], such that the final matrix will be of the following form :

$$M = \begin{bmatrix} M_1 & 0 \ldots & 0 \\ 0 & M_2 \ldots & 0 \\ 0 & 0 & M_s \end{bmatrix}$$

This induces a corresponding partition on X*.

$$X^* = \begin{bmatrix} X_1^* & 0 \ \ldots & 0 \\ 0 & X_2^* \ \ldots & 0 \\ 0 & 0 \ \ldots & X_s^* \end{bmatrix}$$

where $X_i^* = (A_i^* : B_i^*)$ is the design matrix associated with $M_i = (m_{jh}^{(i)})(i = 1,2,\ldots,s)$. Let M_i use rows $j_1,\ldots,j_{s_i}$ and columns $k_1,\ldots,k_{t_i}$ of N. Then define $N_i = (n_{jh}^{(i)})$ as the submatrix using these rows and columns of N. Also define

$$T_i = \{(j,h) \mid n_{jh}^{(i)} = k_\varepsilon \geq 2\}$$

$$S_i = \{(j,h) \mid m_{jh}^{(i)} = 1, n_{jh}^{(i)} = 0\},\ i = 1,2,\ldots,s.$$

If there are d_i elements in T_i then $\varepsilon = 1,2,\ldots,d_i$. For each connected portion (i.e. for each i, i = 1,2,...,s), execute steps 3 through 8.

<u>Step 3</u> : In $X_i^{*\prime}$ replace 0's by -1 and 1's by $2a_i - 1$ for $A_i^{*\prime}$ and $2b_i - 1$ for $B_i^{*\prime}$ where a_i is the number of levels of the α-effects in M_i and b_i the number of levels of γ-effects in M_i. Multiply the resultant matrix by the scalar ($\frac{1}{2a_i b_i}$) to obtain Z_{i1}. A justification that Z_{i1} is a $(X_i^*)_\ell^-$ is given in theorem 5.2.1.

In the cells for which $n_{jh} = k \geq 2$ in the original incidence matrix, they correspond to k repetitions of the same row in the design matrix.

Step 4 : Choose (j_1, h_1) from T_i. Let $a_1' = f_1'X^*$. be the row of X_i^* corresponding to (j_1, h_1) where f_1 is a vector of 0's except for one 1 suitably placed. Set $f = f_1$, $k = k_1-1$, $a = a_1$, $U = Z_{i1}$, and compute

$$g = Uf, \quad d = U'a, \quad c_k = \frac{1}{1 + kg'a}$$

Step 5 : Compute the least square g-inverse

$$V = (U-k\, c_k gd' : c_k g : \ldots : c_k g)$$

where the column $c_k g$ is repeated k times. (In fact one column of $U-k\, c_k gd'$ will also be $c_k g$).

Step 6 : Choose (j_2, h_2) from T_i. Set $k = k_2-1$, $a = a_2(=f_2'\, X_i^*)$, $f = f_2$, $U = V$ and compute steps 4 and 5. Continue till T_i is exhausted. Call final V as Z_{i2}.

Remark 5.2.1 The above steps 4,5 and 6 have been obtained by applying Theorem A2 repeatedly. To arrive at this form one has to use the fact that the U's will be reflexive g-inverses. This follows from the property that Z_{i1} is reflexive, a fact which is proved in corollary 5.2.1.

Now suppose there are ℓ elements in S_i. Let the rows of X_i^* corresponding to these be a_1', a_2', ... a_ℓ' and the columns of Z_{i2} be $g_1, g_2, \ldots, g_\ell$.

Let $X_i^{*'} = [A_i' : a_\ell , \ldots , a_1]$

and $Z_{i2} = [Q : g_\ell , \ldots , g_1]$

Note that $a_j \in \mathcal{M}(A_i')$ for each j, $j = 1,2,\ldots,\ell$. By repeated application of Theorem A1 we get the least square generalized inverse of A_i in the following two steps.

Steps 7 : Set $C_1 = I$, $b_1 = C_1 g_1$

Compute
$$C_i = I + \frac{b_{i-1}\, a'_{i-1}}{1-b'_{i-1} a_{i-1}} C_{i-1}, \quad b_i = C_i g_i$$
$$i = 2, 3, \ldots$$

Continue till $C_{\ell+1}$ is obtained.

Step 8 : Compute $Y = C_{\ell+1}\, Q$.

Remark 5.2.2 : In case one is willing to sacrifice some accuracy of this procedure one can substitute steps 7 and 8 by the following:

Let $X_i^* = \begin{pmatrix} A_1 \\ A_2 \end{pmatrix}$, where A_2 corresponds to the missing cells of N_i. Corresponding to this let $Z_{i2} = (Q_1 : Q_2)$. Then $Y = Q_1 + Q_2(I-A_2Q_2)^{-1}A_2Q_1$ is a least square generalized inverse of A_1 (see theorem 5.2.2). This involves the inversion of a matrix whose dimension is the number of missing observations in N_i.

Step 9 : If the resultant matrices at step 8 were $G_i (i = 1, 2, \ldots, s)$ then form the matrix

$$G = \begin{bmatrix} G_1 & 0 & \cdots\cdots & 0 \\ 0 & G_2 & \cdots\cdots & 0 \\ \vdots & & & \\ 0 & 0 & & G_s \end{bmatrix}$$

The least square g-inverse X_ℓ^- of X is obtained from G by suitably permuting rows and columns.

Remark 5.2.3 In many situations especially if the number of missing cells in any connected portion is greater than the number of occupied cells it might be easier to adopt a different strategy.

Consider an 8×8 design with 48 missing cells having the following pattern :

	γ_1	γ_2	γ_3	γ_4	γ_5	γ_6	γ_7	γ_8
α_1	1							
α_2		1	1					
α_3			1	1			1	
α_4				1	1			
α_5					1	1		
α_6						1	1	
α_7							1	1
α_8	1							1

First find the least square g-inverses for the design matrices *corresponding to boxes 1,2,3* and 4, where box 1 consists of cells {(2,2), (2,3), (3,2), (3,3)}, box 2 of {(4,4), (4,5), (5,4) (5,5)} box 3 of {(6,6), (6,7), (7,6), (7,7)} and finally box 4 {(1,1), (1,8), (8,1), (8,8)}, by steps 1 through 8, for each box separately. Connect these g-inverses by step 9. Add the rows corresponding to cells (3,4), (5,6), (7,8) to the design matrix for the boxes and find the least square g-inverse by Theorem A5, since these additional rows do not belong to the row span of the existing design matrix. [illegible], the row corresponding to cell (3,[illegible]) now belongs to the row [illegible]. So, [illegible] this row and find the g-inverse by Theorem A2. At any stage if the R process applied to the incidence matrix of the existing design matrix can fill a cell then the corresponding row is in the row span of the existing design matrix, otherwise not.

We now prove the assertions made in the algorithm.

<u>Theorem 5.2.1</u> Let $N = (n_{ij})$ denote the incidence matrix of an $a \times b$ additive 2 way classification model, with $n_{ij} = 1$ for all i and j. Let $X = (A:B)$ denote the corresponding design matrix. Then a least square generalized inverse of X is obtained from $(1/2ab)X'$ by replacin all 0's by -1's and 1's by (2a-1) in A' and (2b-1) in B'.

<u>Proof</u> Let $X = [x_1,\ldots,x_a,x_{a+1},\ldots,x_{a+b}]$, where $x_i \in R^{ab}$, $R_i = x_i'Y$, $C_j = x_{a+j}'Y$, $i = 1,\ldots,a; j = 1,\ldots,b$, and

$$T = \sum_{i=1}^{a} R_i = \sum_{j=1}^{b} C_j$$

The normal equations are

$$\begin{aligned} b\alpha_i + \gamma_1 + \ldots + \gamma_b &= R_i, \quad i = 1, \ldots, a \\ \alpha_1 + \ldots + \alpha_a + a\gamma_j &= C_j, \quad j = 1, \ldots, b \end{aligned} \tag{5.2.3}$$

Solving these under the restriction $\sum_{j=1}^{b} \gamma_j = T/2a$,

$$\begin{aligned} \hat{\alpha}_i &= \frac{1}{2ab} [(2a-1)R_i - (T-R_i)], \quad i = 1,\ldots,a \\ \hat{\gamma}_j &= \frac{1}{2ab} [(2b-1)C_j - (T-C_j)], \quad j = 1,\ldots,b \end{aligned} \tag{5.2.4}$$

Since these are least square solutions, the theorem follows.

Now we establish a property of the g-inverse obtained in theorem 5.2.1.

<u>Corollary 5.2.1</u> The g-inverse G obtained in theorem 5.2.1 is reflexive

<u>Proof</u> Let $G = (G_\alpha : G_\gamma)$, where G_α and G_γ correspond to α- and γ-effects respectively. Observe that $R(G) \geq R(X) = a+b-1$. However if a i

a vector of unities, $bG_\alpha e = aG_\gamma e$. Thus $R(G) = a+b-1$ and hence by Lemma 2.5.1 of Rao and Mitra [33], G is reflexive.

The following is a generalization of theorem A1 of Mitra and Bhimasankaram [25]:

Theorem 5.2.2 Let $A = \begin{pmatrix} A_1 \\ A_2 \end{pmatrix}$ be such that $\mathcal{M}(A_2') \subset \mathcal{M}(A_1')$. Further suppose $G = (Q_1 : Q_2)$ is a A_ℓ^-. Then $Y = Q_1 + Q_2(I-A_2Q_2)^{-1}A_2Q_1$ is a least square g-inverse of A_1.

Proof Since G is A_ℓ^-, we have

$$A_1Q_1A_1 + A_1Q_2A_2 = A_1 \tag{5.2.5}$$

$$A_2Q_1A_1 + A_2Q_2A_2 = A_2 \tag{5.2.6}$$

and

$$\begin{bmatrix} A_1Q_1 & A_1Q_2 \\ A_2Q_1 & A_2Q_2 \end{bmatrix} = \begin{bmatrix} Q_1'A_1' & Q_1'A_2' \\ Q_2'A_1' & Q_2'A_2' \end{bmatrix} \tag{5.2.7}$$

Observe that $(I-A_2Q_2)x = 0$

$\Rightarrow$ $x = A_2Q_2x = A_2Q_2A_2Q_2x = (A_2 - A_2Q_1A_1)Q_2x$ from (5.2.6)

$\Rightarrow$ $A_2Q_1A_1Q_2x = 0$

$\Rightarrow$ $A_1Q_2x = 0$ from (5.2.7)

$\Rightarrow$ $x = A_2Q_2x = 0$ since $\mathcal{M}(A_2') \subset \mathcal{M}(A_1')$.

Thus $(I-A_2Q_2)$ is nonsingular. From equation (5.2.6) one obtains

$$A_2 = (I-A_2Q_2)^{-1} A_2Q_1A_1 \tag{5.2.8}$$

Substituting (5.2.8) in (5.2.5) one obtains $A_1YA_1 = A_1$. Moreover, using (5.2.7) $A_1Y = A_1Q_1 + A_1Q_2(I-A_2Q_2)^{-1}A_2Q_1$ is symmetric; hence Y is a least square g-inverse of A_1.

In step 7 of the algorithm, where theorem A1 is being repeatedly applied, we have to check at each stage that $b_i' a_i \neq 1$. This is equivalent to saying that $|I-A_2Q_2| \neq 0$, a proof of which has just been given.

Working of the algorithm will be clear from a numerical illustration given below:

(ii) An example.

We consider an example in which some $n_{ij} \geq 2$ and there are many missing observations. Here we also demonstrate the advantage of finding least square g-inverses for different connected portions and then connecting the disconnected parts. Suppose we have an additive 4×5 classification model with the following pattern :

$$
N = \begin{array}{c} \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{array}
\begin{array}{c} \begin{array}{ccccc} \gamma_1 & \gamma_2 & \gamma_3 & \gamma_4 & \gamma_5 \end{array} \\
\begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 2 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix} \end{array}
$$

the corresponding design matrix is

$$
X = \begin{bmatrix}
1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0
\end{bmatrix}
$$

We briefly sketch the working of the algorithm as follows:

Steps (1) and (2) Applying the R-process to N, the final matrix M(after a permutation of rows) comes out as,

$$M = \begin{array}{c} \\ \alpha_1 \\ \\ \alpha_4 \\ \alpha_2 \end{array} \begin{array}{c} \begin{array}{ccccc} \gamma_1 & \gamma_2 & \gamma_3 & \gamma_4 & \gamma_5 \end{array} \\ \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \end{array} = \begin{bmatrix} M_1 & 0 \\ 0 & M_2 \end{bmatrix}$$

Thus there are two connected portions. Here

$$T_1 = \phi \text{ (the null set)} \qquad T_2 = \{(1,2)\}$$

$$S_1 = \{(1,3), (2,2), (3,1), (3,3)\} \quad S_2 = \phi \ .$$

The corresponding partition $X^* = \begin{bmatrix} X_1^* & 0 \\ 0 & X^*_2 \end{bmatrix}$ can be easily performed.

$$Z_{11} = \frac{1}{18}\begin{bmatrix} 5 & 5 & 5 & -1 & -1 & -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & 5 & 5 & 5 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 & -1 & -1 & 5 & 5 & 5 \\ 5 & -1 & -1 & 5 & -1 & -1 & 5 & -1 & -1 \\ -1 & 5 & -1 & -1 & 5 & -1 & -1 & 5 & -1 \\ -1 & -1 & 5 & -1 & -1 & 5 & -1 & -1 & 5 \end{bmatrix} \qquad Z_{21} = \frac{1}{4}\begin{vmatrix} 1 & 1 \\ 3 & -1 \\ -1 & 3 \end{vmatrix}$$

Steps (4), (5), (6) : Here Z_{11} remains unaltered since $T_1 = \phi$. There is only one element in T_2, and Z_{21} becomes

$$Z_{22} = \frac{1}{8}\begin{bmatrix} 2 & 1 & 1 \\ 6 & -1 & -1 \\ -2 & 3 & 3 \end{bmatrix}$$

<u>Steps (7), (8)</u> For the first connected portion, S_1 has four elements, i.e. $\ell = 4$. Z_{11} becomes

$$Y = \frac{1}{6}\begin{bmatrix} 3 & 2 & -2 & 1 & -1 \\ -3 & 2 & 4 & 1 & -1 \\ 3 & -4 & -2 & 1 & 5 \\ 3 & -2 & 2 & -1 & 1 \\ -3 & 4 & 2 & -1 & 1 \\ 3 & -2 & -4 & 5 & 1 \end{bmatrix}$$

It can be verified that the same result is obtained by doing steps (7) and (8) by the methods proposed in Remark 5.2.2 or Remark 5.2.3. In Remark 5.2.3,

N_1 =

	1	2	3
1	1	1	0
2	1	0	1
3	0	1	0

Take Box 1 = {(1,1), (1,2), (3,1), (3,2)} and Box 2 = {(2,3)}.

After computing the least square g-inverses for the two boxes and connecting them one has to add the row corresponding to the cell (2,1) of N_1 by using Theorem A5. Note that while finding the g-inverse for Box 1 one has to apply Theorem A1 only once. This method is thus very quick.

Z_{22} remains unaltered here since $S_2 = \phi$.

<u>Step (9)</u> The matrix G is

$$G = \begin{bmatrix} Y & 0 \\ 0 & Z_{22} \end{bmatrix}$$

The least square g-inverse of X is obtained from G by appropriate permutations of rows and columns. Thus,

$$X_{\ell}^{-} = \frac{1}{24}\begin{bmatrix} 12 & 8 & 0 & 0 & 0 & -8 & 4 & -4 \\ 0 & 0 & 6 & 3 & 3 & 0 & 0 & 0 \\ -12 & 8 & 0 & 0 & 0 & 16 & 4 & -4 \\ 12 & -16 & 0 & 0 & 0 & -8 & 4 & 20 \\ 12 & -8 & 0 & 0 & 0 & 8 & -4 & 4 \\ -12 & 16 & 0 & 0 & 0 & 8 & -4 & 4 \\ 12 & -8 & 0 & 0 & 0 & -16 & 20 & 4 \\ 0 & 0 & 18 & - & -3 & 0 & 0 & 0 \\ 0 & 0 & -6 & 9 & 9 & 0 & 0 & 0 \end{bmatrix}$$

3. Least square generalized inverses for L-way classification models

(i) Additive L-way models.

Consider an L-way classification model of the form (5.1.1) without interactions i.e. X = (A:B: ... :T). Without loss of generality assume that $\sum_{jk\ldots w} n_{ijk\ldots w} \neq 0$, $i = 1,\ldots,a;\ldots;$ $\sum_{ijk\ldots} n_{ijk\ldots w} \neq 0$, $w = 1,\ldots,t.$

When the design is such that all contrasts within each factor are estimable, i.e. the design matrix is of maximal rank, we find the least square g-inverse for the design matrix by using theorem 5.3.1. The situation when the design matrix is not of maximal rank, will be treated as a separate case.

Case I : The design matrix is of maximal rank. Here the rows corresponding to the missing observations are in the row space of the rows corresponding to the available observations. Theorem 5.3.1 can be proved exactly on the same lines as theorem 5.2.1, and hence we omit the proof.

Theorem 5.3.1 Let X be the design matrix for an additive L-way classification model with no missing observations, and no repeated observations. Then a least square generalized inverse of X is obtained from $\frac{1}{Lab\ldots t}$ X by replacing 0's by -(L-1), and 1's by La-(L-1) in A', Lb-(L-1) in B',...,Lt-(L-1) in T'.

After obtaining the least square g-inverse for the complete design matrix we drop the rows corresponding to the missing observations and find the g-inverse by applying theorem A1. If any row is repeated in X then we apply theorem A2 to find the least square generalized inverse.

Corollary 5.3.1 For a complete fractional design, say $\frac{1}{q}$ of a q^L design the least square g-inverse of X is obtained from $\frac{q}{Lq^L}$ X' by replacing 0's by-(L-1) and 1's by Lq-(L-1).

This corollary can be very useful for latin squares with no missing observations. Consider, for example a 3×3 latin square design. This is a 1/3 replicate of a 3^3 factorial design. If its design matrix is X, then X_ℓ^- is obtained from $\frac{1}{27}$ X' by replacing 0's by - 2 and 1's by 7.

Case II : The design matrix is not of maximal rank. In this case some effects of one factor are confounded with the effects of some other factor(s).

In such situations when there are confounded effects the method of case I cannot be applied directly. However, the least square generalized inverse can be obtained by modifying our technique. Consider a 3 way classification model with design matrix X = (A : B : C). Suppose some of the α-effects are confounded with some γ and θ-effects. Then we first find the least square generalized inverse G_o for X_o = (B : C) by considering it as a two way classification model. If this is a case of total confounding i.e. all α-effects are confounded with γ and θ-effects then $\mathcal{M}(A) \subset \mathcal{M}(X_o)$. Here the least square generalized inverse of X is obtained by adding 'a' null rows to G_0 (see Theorem A3). However, if the α-effects are only partially confounded with the γ and θ-effects then some columns of A will not be in $\mathcal{M}(X_o)$. Here also we first compute the least square g-inverse of X_o. When we add a column of A to X_o, we find the corresponding g-inverse by the formula in Theorem A4. If this column is not in the column span then this formula will go through. Otherwise c = 0. In this case put b = 0. This conforms with Theorem A3. In this way we add the columns of A to X_o successively and find the least square g-inverse of X.

Since in some designs with arbitrary patterns it is difficult to find its nature (for example, whether the design matrix is of maximal rank or not), the algorithm is such that one can avoid such complications. Thus, we first compute the least square g-inverse for X_o by the two-way techniques and then add the columns of A one by one and find the corresponding least square g-inverses by Theorem A4, putting b = 0 whenever c = 0.

The L-way classification model can be treated similarly. Thus, in general if $X = (A:B:C: \cdots :T)$, one can obtain $(X_o)^-_\ell$, where $X_o = (A:B)$ and then add the remaining columns of X in a manner we have just now demonstrated.

Our methods will be clear from the following examples:

Example 5.3.1 Consider the following 2×2×2 design where the α-effects are totally confounded with the γ-effects.

	θ_1	
	γ_1	γ_2
α_1	1	0
α_2	0	1

	θ_2	
	γ_1	γ_2
α_1	1	0
α_2	0	1

The design matrix is

$$X = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix} = (A : B : C)$$

Observe that B=A. So we find the least square g-inverse for $X_o = (A:C)$ by the usual techniques for 2 way and add 2 null rows as proposed in Theorem A3. Thus we obtain

$$X_\ell^- = \frac{1}{8}\begin{bmatrix} 3 & -1 & 3 & -1 \\ -1 & 3 & -1 & 3 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 3 & 3 & -1 & -1 \\ -1 & -1 & 3 & 3 \end{bmatrix}$$

<u>Example 5.3.2</u> Consider the following Latin Square design with two missing observations :

-	2	3
2	3	1
3	1	-

$$X = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \end{bmatrix}$$

Write $X = (A : B : C)$, where A corresponds to rows, B to columns and C to treatment effects. The least square g-inverse for $X_o = (A : B)$ is

$$(X_o)^-_\ell = \frac{1}{30} \begin{bmatrix} 14 & 11 & -3 & -6 & 4 & 1 & -6 \\ -2 & -3 & 9 & 8 & 8 & -3 & -2 \\ -6 & 1 & -3 & 4 & -6 & 11 & 14 \\ 4 & 1 & -3 & -6 & 14 & 11 & -6 \\ 8 & -3 & 9 & -2 & -2 & -3 & 8 \\ -6 & 11 & -3 & 14 & -6 & 1 & 4 \end{bmatrix}$$

Then we add the columns of C one by one to X_o and find the least square g-inverses by Theorem A4. For the first column $c \neq 0$ (hence it is not in $\mathcal{M}(X_o)$). But for the next two, $c = 0$ and thus these columns are in the column span of the existing design matrices. For both of these we take $b = 0$.

The final least square g-inverse comes out as

$$X_{\ell}^{-} = \frac{1}{8}\begin{bmatrix} 4 & 1 & -3 & 0 & 2 & -1 & 0 \\ -1 & 0 & 3 & 1 & 1 & 0 & -1 \\ -3 & 2 & 3 & -1 & -3 & 4 & 1 \\ 2 & -1 & -3 & 0 & 4 & 1 & 0 \\ 1 & 0 & 3 & -1 & -1 & 0 & 1 \\ -3 & 4 & 3 & 1 & -3 & 2 & -1 \\ 3 & -3 & -6 & 3 & 3 & -3 & 3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

(ii) The most general classification models.

Consider the model (5.1.1) in its utmost generality. Rewrite (5.1.2) as

$$X = (X_o : X_1) \qquad (5.3.1)$$

where X_o = (A:B: ... : T) is the design matrix under the assumption of additivity, and X_1 corresponds to the interactions which we believe might be present.

To get X_{ℓ}^{-} we first find $(X_o)_{\ell}^{-}$ by methods already discussed. Then we append the columns of X_1, one by one, to X_o and find the least square g-inverses by Theorem A4, putting b = 0 whenever c = 0 (this conforms with Theorem A3).

If s is the total number of observations, then c can be non null at most $[s-R(X_o)]$ times. When c = 0, we only add a null row to the existing g-inverse and the "AG" of Theorem A4 remains unchanged. Thus the procedure is expected to be very fast. The algorithm can be programmed for electronic computers without much difficulty. Even without access to such a system computations are not much laborious and are quite speedy.

Observe that though the algorithm gives the BLUE for estimable functionals, the estimator need not be BLIMBE. In fact X_{ℓ}^{-} of theorem 5.2.1 will be a Moore Penrose inverse if an only if a = b. If X = (A:B) is the design matrix of theorem 5.2.1, then it can be verified in a straightforward manner that the Moore Penrose inverse of X is obtained from $\frac{1}{(a+b)}$ X' by replacing 0's by - (1/b) and 1's by (a+b-1)/b in A' and 0's by -(1/a) and 1's by (a+b-1)/a in B'. Note that if G is a least square g-inverse of X with $(I-X^{+}X)G = 0$ then $G = X^{+}$. Hence an alternative way of obtaining the result we stated just now is to solve the normal equations (5.2.3) under the restriction that the totally nonestimable functionals vanish, or in other words, $\sum_{i} \alpha_i - \sum_{j} \gamma_j = 0$.

To obtain BLIMBE one may start with X^{+} and work on the same lines as in our algorithm using theorems giving Moore Penrose inverses from Mitra and Bhimasankaram [25], instead of theorems giving some choices of least square g-inverses as we have done. However, much of the inherent simplicity and speed of the algorithm we described is likely to be lost in the process.

REFERENCES

1. Aitken, A.C. On least squares and linear combination of observations. Proc. Roy. Soc. Edinb. A, 55 (1934), 42-47.

2. Balakrishnan, A.V. An operator theoretic formulation of a class of control problems and a steepest descent method of solution. J. Soc. Indust. Appl. Math. Ser. A: Control 1 (1963), 109-127.

3. Birkes, D., Dodge, Y. and Seely, J. Spanning sets for estimable contrasts in classification models. Ann. Statist. 4 (1976), 86-107.

4. Bjerhammar, A. Theory of errors and generalized matrix inverses. Elsevier, Amsterdam, 1973.

5. Bjerhammar, A. Hyper-estimators. Paper presented at the 41st session of the International Statistical Institute, New Delhi, 1977.

6. Bose, R.C. The fundamental theorem of linear estimation [Abstract]. Proc. 31st Indian Science Congress. 4 (1944), 2-3.

7. Bose, R.C. Least square aspects of analysis of variance. Inst. Stat. Mimeo series 9 (1949), Chapel Hill, N.C.

8. Bunke, H. and Bunke, O. Identifiability and estimability. Math. Operationsforsch Statist. 5 (1974), 223-233.

9. Bunke, O. Minimax linear, ridge and shrunken estimators for linear parameters. Math. Operationsforsch. Statist. 6 (1975), 817-829.

10. Calinski, T. On some desirable patterns in block designs. Biometrics 27 (1971), 275-292.

11. Chipman, J.S. On least squares with insufficient observations. J. Amer. Statist. Assoc. 59 (1964), 1078-1111.

12. David, F.N. and Neyman J. Extensions of the Markoff theorem on least squares. Statist. Res. Mem. 2 (1938), 105-116.

13. Drygas, H. Gauss-Markov estimation and best linear minimum bias estimation. Unpublished Technical Report, University of Heidelberg, May 1969.

14. Eccleston, J.A. and Hedayat, A. On the theory of connected designs: characterisation and optimality. Ann. Statist. 2 (1974), 1238-1255.

15. Forsythe, G.E. and Golub, G.H. On the stationary values of a second degree polynomial on the unit sphere. Siam J. Appl. Math. 13 (1965), 1050-1068.

16. Gauss, C.F. Methode des moindres carres. (Trans. J. Bertrand) Paris 1865

17. Hoerl, A.E. and Kennard, R.W. Ridge regression : biased estimation for nonorthogonal problems. Technometrics 12(1970), 55-67.

18. Hoerl, A.E. and Kennard, R.W. Ridge regression : applications to nonorthogonal problems. Technometrics 12(1970), 69-82.

19. Kuks, J. A minimax estimator of regression coefficients (in Russian) Izv. Akad. Nauk. Eston SSR 21 (1972), 73 - 78.

20. Kuks, J. and Olman, W. Minimax linear estimation of regression coefficients (in Russian) Izv. Akad. Nauk. Eston SSR 21 (1972), 66-72.

21. Lauter, H. A minimax linear estimator for linear parameters under restrictions in form of inequalities. Math Operationsforsch. Statist. 5 (1975), 689-696.

22. Markov, A.A. Warscheinlichkeitsrechnung. (Trans. H. Liebmann) 2nd Edition. Leipzig and Berlin, 1912.

23. Mitra, S.K. Optimal inverse of a matrix. Sankhyā A 37(1975), 550-563.

24. Mitra, S.K. Generalized inverse of matrices and applications to linear models. Handbook of Statistics (Ed. P.R. Krishnaiah) North Holland, Amsterdam, 1979.

25. Mitra, S.K. and Bhimasankaram, P. Generalized inverse of partitioned matrices and recalculation of least square estimates for data or model changes. Sankhyā A 33 (1971), 395-410.

26. Patnaik, P.B. The noncentral χ^2 - and F-distributions and their applications. Biometrika 36(1949), 202-232.

27. Plackett, R.L. Some theorems in least squares Biometrika 37(1950), 149-157.

28. Rao, C.R. Generalizations of Markoff's theorem and tests of linear hypotheses. Sankhyā 7 (1949), 9-16.

29. Rao, C.R. Linear statistical inference and its applications. Wiley, New York, 1965.

30. Rao, C.R. Unified theory of linear estimation. Sankhyā A 33(1971), 371-394.

31. Rao, C.R. Estimation of parameters in a linear model, Ann. Statist. 4(1976), 1023-1037.

32. Rao, C.R. and Mitra, S.K. Further contributions to the theory of generalized inverse of matrices and its applications. Sankhyā A 33(1971), 289-300.

33. Rao, C.R. and Mitra, S.K. Generalized inverse of matrices and its applications. Wiley, New York, 1971.

34. Scheffe, H. Analysis of variance, Wiley, New York, 1959.

35. Schonfeld, P. Best linear minimum bias estimation in linear regression. Econometrika 39(1971), 531-544.

36. Sjoberg, L. Are least square estimators biased? Technical Report. Division of Geodesy, Royal Institute of Technology, Stockholm, 1976.

37. Srivastava, J.N. and Anderson, D.A. Some basic properties of multidimensional partially balanced designs. Ann. Math. Statist. 41(1970), 1438-1445.

38. Weeks, D.L. and Williams, D.R. A note on determination of connectedness in an N-way cross classification. Technometrics 6(1964), 319-342, errata Technometrics 7(1964), 281.

39. Zelen, M. and Severo, N.C. Probability Functions. Handbook of Mathematical Functions (Ed. M.Abramowitz and IrA. Stegun.) Dover, New York, 1965, 925-996.