MINIMUM BIAS APPROXIMATION OF A GENERAL REGRESSION MODEL WITH AN APPLICATION TO RATIONAL MODELS

Robert Cote, A. R. Manson and R. J. Hader

Institute of Statistics
Mimeograph Series No. 756
Raleigh - July 1971

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

The problem of optimum design in regression has drawn the attention of many writers. In particular, Elfving (1952), Kiefer (1959, 1961), Kiefer and Wolfowitz (1959), Hoel and Levine (1964) investigated different optimality criteria applying to regression problems. Also from a different viewpoint, Folks (1958) and Box and Draper (1959, 1963) introduced bias considerations in their optimality criterion. The latter studied the situation where the model, $\eta(\underline{x})$, is a polynomial of degree d + k - 1;

$$\eta(\underline{x}) = \underline{x}_1'\underline{\beta}_1 + \underline{x}_2'\underline{\beta}_2 ,$$

is approximated by a polynomial of degree d - 1;

$$\hat{y}(\underline{x}) = \underline{x}_1'\underline{b}_1 .$$

The vector $\underline{x}_1$ is made up of terms required for the polynomial of degree d - 1; the vector $\underline{x}_2$ is made up of additional higher order terms required for the polynomial of degree d + k - 1 while $\underline{\beta}_1$ and $\underline{\beta}_2$ are corresponding vectors of regression coefficients. The estimator $\underline{b}_1$ is $\underline{b}_1 = (X_1'X_1)^{-1}X_1'\underline{y}$ where X_1 is the matrix of values taken by the variates in $\underline{x}_1'$ for the N experimental runs; $\underline{y}$ is the vector of the N uncorrelated random variables with $E(Y_i) = \eta(x_i)$ and $E[Y_i - \eta(x_i)]^2 = \sigma^2$ (i = 1, 2, ..., N). Box and Draper were primarily interested in finding designs which minimize the mean square error (MSE) of $\hat{y}(x)$ averaged over some region of interest R, namely,

$$J = \frac{N\Omega}{\sigma^2} \int_R \text{MSE}[\hat{y}(\underline{x})]d\underline{x}$$

where $\Omega^{-1} = \int_R d\underline{x}$. J can be split into two parts: $J = V + B$ where $V = \frac{N\Omega}{\sigma^2}\int_R \text{Var}\,[\hat{y}(\underline{x})]d\underline{x}$, the averaged variance of $\hat{y}(\underline{x})$ over R; and $B = \frac{N\Omega}{\sigma^2}\int_R \{E[\hat{y}(\underline{x})] - \eta(\underline{x})\}^2 d\underline{x}$, the averaged squared bias of $\hat{y}(\underline{x})$ over R. In their work they noted that unless V was many times larger than B, the minimum J designs were remarkably close to those obtained by ignoring V completely. Thus, they showed that to minimize B alone the design matrix $X = [X_1 \vdots X_2]$ should satisfy

$$(X_1'X_1)^{-1}X_1'X_2 = W_{11}^{-1}W_{12} \tag{1.1}$$

where X_2 is the matrix of values of $\underline{x}_2'$ and $W_{1j} = \Omega\int_R \underline{x}_1\underline{x}_j'\, d\underline{x}$ $j = 1, 2$. Equation (1.1) will be called the Box-Draper conditions for minimum B.

Karson, Manson and Hader (1969) presented a different approach to this problem. Accepting the Box and Draper result that B is the dominating factor in mean square error, they minimized B by choice of estimator rather than by choice of design. The minimum bias estimator of $\eta(\underline{x})$ which they obtained was $\hat{\eta}(\underline{x}) = \underline{x}_1'A(X'X)^{-}X'\underline{y}$ where A is a matrix of known constants $A = [I \vdots W_{11}^{-1}W_{12}]$. The symbol S^{-} will denote a generalized inverse of a square matrix S, which satisfies the equation $S = SS^{-}S$. The estimator, $\hat{\eta}(\underline{x})$, achieves minimum B for any design for which $A\underline{\beta}$ is estimable, where $\underline{\beta}' = (\underline{\beta}_1' \vdots \underline{\beta}_2')$. Using the design flexibility remaining, they found designs with smaller J than is given by designs which satisfy the Box-Draper conditions.

The following section will generate the Karson, Manson and Hader method to a larger class of models. Some results which hold in this more general situation will be given. This approach will be compared with the Box and Draper method. As a direct application, a brief discussion of the problem of approximating the ratio of polynomials by simple polynomials will be presented. The technique developed will be illustrated with examples. For these examples, designs satisfying the bias and variance criteria will be developed.

2. DEVELOPMENT OF THE PROBLEM

Let $(\mathfrak{X}, \mathfrak{a}, \mu)$ be a measure space where $\mathfrak{X}$ is a compact set and μ, a finite measure defined on the σ-algebra $\mathfrak{a}$ of subsets of $\mathfrak{X}$. Let $L_2(\mathfrak{X}, \mathfrak{a}, \mu)$ be the class of all square integrable, real valued functions defined a.e. on $\mathfrak{X}$, i.e.,

$$L_2(\mathfrak{X}, \mathfrak{a}, \mu) = \{f \text{ defined a.e. on } \mathfrak{X}: \int_{\mathfrak{X}} f^2(x)d\mu(x) < +\infty\}.$$

Consider n linearly independent continuous functions $f_1, f_2, \ldots, f_n$ in $L_2(\mathfrak{X}, \mathfrak{a}, \mu)$. Let $\theta_1, \theta_2, \ldots, \theta_n$ be n unknown constants in some vector space Θ over E^1, the field of real numbers. (In the sequel E^m will denote the m-fold cartesian product of the set of real numbers.) These quantities will be written as vectors:
$\underline{f}' = (f_1, f_2, \ldots, f_n)$ and $\underline{\theta}' = (\theta_1, \theta_2, \ldots, \theta_n)$.

If $\mathfrak{X}$ is a k-dimensional space, its elements are actually k-tuples $(x_1, x_2, \ldots, x_k)$ and any function defined on $\mathfrak{X}$ is a k-variable function. However, an element of the k-dimensional space $\mathfrak{X}$ shall be denoted by one letter with or without subscript, x. Over the space $\mathfrak{X}$, a hypersurface in a (k+1)-dimensional space is assumed to be described by

$$\eta(x) = \sum_{j=1}^{n} \theta_j f_j(x) = \underline{\theta}'\underline{f}(x) , \tag{2.1}$$

where the f_j $(j = 1, 2, \ldots, n)$ are known functions. The set $\mathfrak{X}$ will also be regarded as the operability region, i.e., for each $x \in \mathfrak{X}$ it is possible to obtain a value of the random variable Y(x) which has mean $\underline{\theta}'\underline{f}(x)$. The random variables $Y(x_i)$ and $Y(x_j)$ are assumed to have a known correlation apart from a constant σ^2. If several x_i's are equal then the corresponding $Y(x_i)$'s will be treated as different random variables.

A "close approximation" of the true response given in the equation (2.1) is intuitively appealing. In situations where the estimation of all parameters θ_j is expensive, difficult or even impossible, a convenient approach is to approximate the assumed true model by a "relatively simpler one". Thus, the experimenter selects a class of such models;

$$\mathcal{C} = \{\eta_0(x): \eta_0(x) = \sum_{j=1}^{s} \alpha_j g_j(x)\} \quad \text{with } s \leq n \ .$$

The g_j ($j = 1, 2, \ldots, s$) are linearly independent, continuous, real valued functions in L_2, and the parameters $\alpha_j \in \Theta$ ($j = 1, 2, \ldots, s$) are to be estimated. These quantities will be written as vectors: $\underline{g}' = (g_1, g_2, \ldots, g_s)$ and $\underline{\alpha}' = (\alpha_1, \alpha_2, \ldots, \alpha_s)$. In such an approach, the error of approximation stems from two sources: the sampling error and the bias error due to the failure of the chosen $\eta_0(x) \in \mathcal{C}$ to exactly represent the true response $\eta(x)$. Obviously, one would like to minimize both errors. However, the minimization of the sampling error must take place in the estimation space, by estimation procedures, and in the operability region, $\mathcal{X}$, by choice of design. The minimization of the bias error has to be accomplished in the parameter space, Θ, and in L_2 space by choice of $\eta_0(x) \in \mathcal{C}$. Moreover, the sampling error is defined only after the choice of $\eta_0(x) \in \mathcal{C}$ is made.

Thus, the only alternative is the conditional minimization of the sampling error, *i.e.*, the minimization of the sampling error given that the minimum bias has been achieved within $\mathcal{C}$.

2.1 Choice of Minimum Bias Model in $\mathcal{C}$

Let $B(\eta_0)$ denote the averaged squared bias over $\mathcal{X}$ due to the choice of $\eta_0(x) \in \mathcal{C}$, namely,

$$B(\eta_0) = \Omega \int_{\mathcal{X}} \{\eta_0(x) - \eta(x)\}^2 d\mu(x) \tag{2.2}$$

where $\Omega^{-1} = \int_{\mathcal{X}} d\mu(x)$. The following minimization

$$B = \underset{\eta_0 \in \mathcal{C}}{\text{minimum}} \; B(\eta_0)$$

is equivalent to

$$B = \underset{(\alpha, g) \in \Theta^s x L_2^s}{\text{minimum}} \; B(\underline{\alpha}' \underline{g})$$

where $(\underline{\alpha}, \underline{g})$ is an element of the product space $\Theta^s x L_2^s$. Θ^s is the s-fold cartesian product of Θ and L_2^s, the s-fold cartesian product of L_2.

The minimization of $B(\underline{\alpha}'\underline{g})$ over $\Theta^s x L_2^s$ is a formidable task if at all possible! Even if one reduces the domain of minimization to $\Theta^s x M^s$ where $M = M(f_1, f_2, \ldots, f_n)$ is the space spanned by the functions f_i $(i = 1, 2, \ldots, n)$, the task is not materially simplified. To reduce the minimization problem to workable size, it becomes necessary to fix a set of functions $\underline{g}' = (g_1, g_2, \ldots, g_s)$ and minimize over Θ only. Then B_g will denote the fact that the g_i $(i = 1, 2, \ldots, s)$ may not be an optimum choice. The functions η, η_0,

f_j, g_j are defined on $\mathfrak{X}$ and their integration will be over $\mathfrak{X}$ unless otherwise specified. Henceforth, the variate x and the domain of integration $\mathfrak{X}$ shall be omitted.

For a given set of functions $(g_1, g_2, \ldots, g_s)$ equation (2.2) is written as

$$B_g(\eta_0) = \Omega \int \{\underline{\alpha}'\underline{g} - \underline{\theta}'\underline{f}\}^2 d\mu$$

$$= \Omega \int \{\underline{\alpha}'\underline{g}\underline{g}'\underline{\alpha} - 2\underline{\alpha}'\underline{g}\underline{f}'\underline{\theta} + \underline{\theta}'\underline{f}\underline{f}'\underline{\theta}\} d\mu$$

$$= \underline{\alpha}' W_{gg}\underline{\alpha} - 2\underline{\alpha}' W_{gf}\underline{\theta} + \underline{\theta}' W_{ff}\underline{\theta} \tag{2.3}$$

where

$$W_{hk} = \Omega \int \underline{h}\underline{k}' d\mu \tag{2.4}$$

are real valued matrices. The following theorem is appropriate at this point.

Theorem 2.1

A necessary and sufficient condition for the matrix W_{gg} to be positive definite is that the g_i $(i = 1, 2, \ldots, s)$ be linearly independent. (For a proof, see Appendix A.1.)

Corrolary 2.1

As defined in equation (2.4) the matrix

(i) W_{ff} is positive definite

(ii) $W'_{gf} W^{-1}_{gg} W_{gf}$ is positive semi-definite

(iii) $W_{ff} - W'_{gf} W^{-1}_{gg} W_{gf}$ is positive semi-definite.

Using Theorem 2.1, equation (2.3) may be written as

$$B_g(\eta_0) = (\underline{\alpha} - W_{gg}^{-1}W_{gf}\underline{\theta})'W_{gg}(\underline{\alpha} - W_{gg}^{-1}W_{gf}\underline{\theta}) + \underline{\theta}'(W_{ff} - W'_{gf}W_{gg}^{-1}W_{gf})\underline{\theta} \ . \quad (2.5)$$

The second term of the R.H.S. of equation (2.5) is constant in $\underline{\alpha}$. Therefore, from Theorem 2.1, it follows that

$$\min B_g = \underset{\alpha \in \Theta}{\text{minimum}} \ B_g(\eta_0) = \underline{\theta}'(W_{ff} - W'_{gf}W_{gg}^{-1}W_{gf})\underline{\theta}$$

if and only if

$$\underline{\alpha} = W_{gg}^{-1}W_{gf}\underline{\theta} = A\underline{\theta}$$

where $A = W_{gg}^{-1}W_{gf}$. Moreover min B_g is unique. Choosing the set of functions g_i $(i = 1, 2, \ldots, s)$ is equivalent to selecting a subclass $\mathcal{C}_g \subset \mathcal{C}$ where

$$\mathcal{C}_g = \{\eta_0 \colon \eta_0 = \underline{\alpha}'\underline{g}, \ \ \underline{g} \text{ is given}\} \text{ with } s \leq n \ .$$

The minimum bias model (min B_g model) in $\mathcal{C}_g$ which is used to approximate $\eta = \underline{\theta}'\underline{f}$ is written as

$$\eta_0 = \underline{\alpha}'\underline{g}, \quad \underline{\alpha} = A\underline{\theta} \ . \tag{2.6}$$

The choice of functions g_i $(i = 1, 2, \ldots, s)$ as a subset of $(f_1, f_2, \ldots, f_n)$ is a natural and appealing one, and, if the set of functions g_i is a subset of $(f_1, f_2, \ldots, f_n)$ so that $g_k = f_k$ for $k = 1, 2, \ldots, s$, then the matrix A always takes the form $A = [I_s \vdots W_{11}^{-1}W_{12}]$, where $W_{11} = W_{gg}$ and $W_{12} = \Omega \int \underline{g}\underline{f}'_2 d\mu$ with $\underline{f}'_2 = (f_{s+1}, f_{s+2}, \ldots, f_n)$.

2.2 Estimation of the min B_g Model

A question of great interest is the estimability of the min B_g model given in equation (2.6). Assume that a vector of N random variables (observations) $y'(\underline{x}) = [Y(x_1), Y(x_2), \ldots, Y(x_N)]$ satisfies

(1) $E[y(\underline{x})] = F\underline{\theta}$

(2) $E\{[y(\underline{x}) - \eta(\underline{x})][y(\underline{x}) - \eta(\underline{x})]'\} = \Sigma\sigma^2.$

F is an Nxn matrix whose $(i,j)^{th}$ element is $f_j(x_i)$, Σ is a NxN positive definite matrix, of known constant elements and σ^2 is a real, positive constant. The Gauss Markoff Theorem (Scheffé, 1967) applies, i.e., $\underline{\alpha} = A\underline{\theta}$ is linearly estimable if and only if the design matrix $F \in \mathfrak{F}_{MB}$ where

$$\mathfrak{F}_{MB} = \{F: A(F'\Sigma^{-1}F)^{-}(F'\Sigma^{-1}F) = A\} .$$

This class $\mathfrak{F}_{MB}$ is also the class of designs which achieve

$$\min B_g = \underline{\theta}'(W_{ff} - W'_{gf}W^{-1}_{gg}W_{gf})\underline{\theta} .$$

An equivalent necessary and sufficient condition for estimability of $A\underline{\theta}$ is given in the following theorem.

Theorem 2.2

The vector of parameters $\underline{\alpha} = A\underline{\theta}$ is linearly estimable if and only if $A' = V_1T$. V_1 is a nxr matrix whose columns are the orthonormal characteristic vectors corresponding to the r non-zero characteristic roots of $(F'\Sigma^{-1}F)$ with $s \leq r \leq n$, and T is rxs matrix of full rank. (For a proof, see Appendix A.2.)

This theorem essentially brings out the fact that the process of linear estimation takes place in a r-dimensional subspace of the n-dimensional space spanned by the columns of $(F'\Sigma^{-1}F)$. The following theorem is then a natural one.

Theorem 2.3

If $\underline{\alpha} = A\underline{\theta}$ is estimable where $A = W_{gg}^{-1}W_{gf}$ is a matrix of full rank, then the matrix $A(F'\Sigma^{-1}F)^{-}A'$ is non-singular. (For a proof, see Appendix A.3.)

Moreover, the columns of A' are in the space generated by the columns of V_1 (Theorem 2.2) which is the space generated by the columns of $(F'\Sigma^{-1}F)^{-}$, (Rao, 1967, p. 184). Thus, it can be shown (see Graybill, 1969) that $A(F'\Sigma^{-1}F)^{-}A'$ is invariant for any generalized inverse of $(F'\Sigma^{-1}F)$.

For a fixed $F \in \mathfrak{F}_{MB}$ the BLUE of $\underline{\alpha}$ is

$$\hat{\underline{\alpha}} = A(F'\Sigma^{-1}F)^{-}F'\Sigma^{-1}y(\underline{x})$$

and the minimum variance linear estimator of η when using the min B_g model is

$$\hat{\eta}_0(F) = \underline{g}'A(F'\Sigma^{-1}F)^{-}F'\Sigma^{-1}y(\underline{x}) .$$

Now $\mathrm{Var}\,[\hat{\eta}_0(F)] = \underline{g}'A(F'\Sigma^{-1}F)^{-}A'\underline{g}\sigma^2$ when averaged over the region $\mathfrak{X}$ is denoted by V_F, namely,

$$V_F = \frac{N\Omega}{\sigma^2}\int \mathrm{Var}\,[\hat{\eta}_0(F)]d\mu$$

$$= N\cdot \mathrm{Tr}[A(F'\Sigma^{-1}F)^{-}A'W_{gg}] .$$

The notation Tr[H] will be used to denote the trace of the square matrix H. A great deal of design flexibility remains at this point. One way to use this flexibility is to achieve

$$\min V = \underset{F \in \mathfrak{Z}_{MB}}{\text{minimum}} V_F ,$$

i.e., to obtain min V having achieved min B_g.

This design flexibility may also be used to satisfy other criteria, as will be shown in Section 3.

2.3 Performance of the min B_g Approach

To approximate a given function, say $\eta(x)$, on the basis of N experimental observations various methods aimed at satisfying both bias and variance criteria can be presented. These methods could thus be compared on the basis of the compromise they offer between variance and bias.

Consider the situation where the set of functions g_i (i = 1, 2, ..., s) is a subset of the functions f_i (i = 1, 2, ..., n) so that $g_k = f_k$ for k = 1, 2, ..., s. The full model is then written as

$$\eta = \underline{\theta}_1'\underline{f}_1 + \underline{\theta}_2'\underline{f}_2 \tag{2.7}$$

where $\underline{\theta}_1'\underline{f}_1$ contains the first s terms and $\underline{\theta}_2'\underline{f}_2$, the remaining (n-s) terms. The design matrix $F = [F_1 \vdots F_2]$ is written so as to match the partitioning of the model η. The following classes of designs are defined:

$$\mathfrak{F}_{FM} = \{F: \underline{\theta} \text{ is estimable}\}$$

$$\mathfrak{F}_{MB} = \{F: A\underline{\theta} \text{ is estimable}\}$$

$$\mathfrak{F}_{RM} = \{F: \underline{\theta}_1 \text{ is estimable}\} \qquad (2.8)$$

$$\mathfrak{F}_{BD} = \{F: (F_1'F_1)^{-1}F_1'F_2 = W_{gg}^{-1}W_{gf_2}\} .$$

Also, the following notation will be used:

$$V_*(F) = \frac{N\Omega}{\sigma^2} \int \text{Var } [\hat{\eta}_0(F)]d\mu \qquad \text{for } F \in \mathfrak{F}_* \qquad (2.9)$$

and

$$V_* = \underset{F \in \mathfrak{F}_*}{\text{minimum}}\ V_*(F) . \qquad (2.10)$$

Lemma 2.1

If $A = [A_1 \vdots A_2]$ and A_1 is a sxs matrix of full rank, then a necessary condition for $A\underline{\theta}$ to be estimable is that F_1 be of full rank. (For a proof, see Appendix A.4.)

In particular, this lemma applies when $A = [I_s \vdots W_{gg}^{-1}W_{gf_2}]$. The following lemma establishes a relationship between the classes of designs defined in (2.8).

Lemma 2.2

Using the notation of equation (2.8)

(i) $\mathfrak{F}_{FM} \subseteq \mathfrak{F}_{MB} \subseteq \mathfrak{F}_{RM}$

(ii) $\mathfrak{F}_{BD} \subseteq \mathfrak{F}_{MB}$

(iii) when $N = s$, $\mathfrak{F}_{MB} = \mathfrak{F}_{BD}$.

(For a proof, see Appendix A.5).

Theorem 2.4

Using the notation defined in equations (2.9) and (2.10)

(i) $V_{FM} \geq V_{MB}$

(ii) $V_{MB} \geq V_{RM}$

(iii) $V_{MB} \leq V_{BD} = s$.

(For a proof, see Appendix A.6.)

From (iii) of Lemma 2.2, strict equality is achieved in (iii) of Theorem 2.4 when N = s and $\mathfrak{F}_{MB} \neq \emptyset$. Moreover, many examples exist where the strict inequality holds in (iii) of Theorem 2.4 (see Karson, Manson and Hader, 1969).

Consider the situation where N = s, then the min B_g model cannot contain more than s parameters. A logical method of obtaining protection against bias would be to add more terms to the fitted model. This is obviously not possible when N = s. An alternative method of obtaining additional protection against bias would be to add more terms to the assumed true model and obtain minimum bias protection via the fitted model. Unfortunately, this extra bias protection is accompanied by an increase in variance error as shown in the following theorem.

Theorem 2.5

To the model given in equation (2.7)

$$\eta_1 = \underline{\theta}_1'\underline{f}_1 + \underline{\theta}_2'\underline{f}_2 = \underline{\theta}'\underline{f}$$

add t more terms, say $\underline{\theta}_3'\underline{f}_3$. The new model is then

$$\eta_2 = \underline{\theta}'\underline{f} + \underline{\theta}_3'\underline{f}_3 .$$

Let $C_g = \{\eta_0 : \eta_0 = \underline{\alpha}'\underline{g}$ where $g_k = f_k$ for $k = 1, 2, \ldots, s\}$ be the class of simpler models. If V_i denotes the averaged variance of the min V|min B_g linear estimator of η_i for $i = 1, 2$, then $V_1 \leq V_2$. (For a proof, see Appendix A.7.)

3. APPROXIMATING RATIONAL FUNCTIONS BY POLYNOMIAL FUNCTIONS

A polynomial form in x over E^1 is an expression of the form $P_n(x:a) = \sum_{i=0}^{n} a_i x^i$ with $a_n \neq 0$ if $n > 0$. The degree of the polynomial form is n and $a_i \in E^1$ $(i = 0, 1, \ldots, n)$. A rational form is defined to be the ratio $P_n(x:a)/P_m(x:b)$ where $P_m(x:b) \neq 0$ and $P_n(x:a)$ are polynomial forms in x of degree m and n respectively. If x is regarded as a variable with a given domain, these forms are called functions, namely, polynomial functions and rational functions. These definitions can be generalized to functions of several variables. For example, a polynomial function in $(x_1, x_2, \ldots, x_n)$ is defined recursively as a function in x_n over the ring $D[x_1, x_2, \ldots, x_{n-1}]$ of polynomials. (For more details, see Birkhoff and MacLane, 1965.) The same notation will be used for a polynomial in one or in several variables, recalling that x may actually represent a product such as $\prod_{i=1}^{k} x_i^{t_i}$ when dealing with a k-variable polynomial function.

Assume that the following functional relationship between the response η and the factor x holds

$$\eta(x) = R_{n,m}(x:\theta,\gamma) = \frac{P_n(x:\theta)}{P_m(x:\gamma)}, \quad \forall\ x \in \mathfrak{X},$$

where the operability region $\mathfrak{X}$ is a closed bounded set of E^k, and

$$P_n(x:\theta) = \sum_{j=0}^{q} \theta_{t_j} x^{t_j}, \quad P_m(x:\gamma) = \sum_{j=0}^{u} \gamma_{v_j} x^{v_j}$$

are real valued polynomial functions of degree n and m respectively. The class of relatively simpler models to be considered is

$$\mathcal{C}_g = \{\eta_0 : \eta_0 = P_s(x:\alpha) = \sum_{j=0}^{s} \alpha_{s_j} x^{s_j}\} \quad \text{with } s \leq q$$

where the s_j ($j = 0, 1, \ldots, s$) are known integers and the α_{s_j} are unknown constants in E^1. No claim is made concerning any optimality criterion in the choice of $\mathcal{C}_g$: this class of models is simple and illustrates the point that the set of functions g_i ($i = 0, 1, \ldots, s$) need not be a subset of functions f_i ($i = 0, 1, \ldots, q$).

Using notation similar to that used in Section 2,

$$B_g(\eta_0) = \Omega \int \{P_s(x:\alpha) - R_{n,m}(x:\theta,\gamma)\}^2 dx ,$$

with $\Omega^{-1} = \int dx$ and where the measure μ is simply the usual Lebesgue measure defined on the Borel σ-algebra of subsets of $\mathfrak{X}$. If for appropriate range of j, one defines

$$f_j(x) = \frac{x^{t_j}}{P_m(x:\gamma)}; \ \theta_j = \theta_{t_j}; \ g_j(x) = x^{s_j}; \ \alpha_j = \alpha_{s_j}$$

then the results of Section 2 apply when restricting $\frac{x^{t_j}}{P_m(x:\gamma)} \in L_2$ ($j = 0, 1, \ldots, q$) and γ to be known. The matrix $A = W_{gg}^{-1} W_{gf}$ where W_{gg} is a $(s+1)x(s+1)$ matrix whose $(i,j)^{th}$ element is $\Omega \int x^{s_i+s_j} dx$ and W_{gf} is a $(s+1)x(q+1)$ matrix whose $(i,j)^{th}$ element is $\Omega \int \frac{x^{s_i+t_j}}{P_m(x:\gamma)} dx$. The minimum averaged squared bias is

$$\min B_g = \underline{\theta}'(W_{ff} - W'_{gf} W_{gg}^{-1} W_{gf})\underline{\theta}$$

where W_{ff} is a $(q+1)x(q+1)$ matrix whose $(i,j)^{th}$ element is

$$\Omega \int \frac{x^{t_i+t_j}}{[P_m(x:\gamma)]^2} dx.$$

Assume that the vector of the N observable responses $y'(\underline{x}) = [Y(x_1), Y(x_2), \ldots, Y(x_N)]$ satisfies the conditions

(1) $E[y(\underline{x})] = R\underline{\theta}$

(2) $E\{[(y(\underline{x}) - \eta(\underline{x})][y(\underline{x}) - \eta(\underline{x})]'\} = I_N\sigma^2$

where R is a Nx(q+1) matrix whose $(i,j)^{th}$ element is $\frac{x_i^{t_j}}{P_m(x_i : \gamma)}$ and σ^2 is a real, positive constant. For any design matrix $R \in \mathfrak{F}_{MB}$ (the class of designs for which min B_g is achieved) the BLUE of $\underline{\alpha}$ is $\hat{\underline{\alpha}} = A(R'R)^{-}R'y(\underline{x})$ and the variance of $\hat{\eta}_0$ is $\text{Var}(\hat{\eta}_0) = \underline{g}'A(R'R)^{-}A'\underline{g}\sigma^2$ with $\underline{g}' = (x^{s_0}, x^{s_1}, \ldots, x^{s_s})$. For a fixed design matrix $R \in \mathfrak{F}_{MB}$ the averaged Var $(\hat{\eta}_0)$ over $\mathfrak{X}$ is

$$V_R = N \cdot \text{Tr}[(R'R)^{-}W'_{gf}W_{gg}^{-1}W_{gf}].$$

There still remains the flexibility of choosing $R \in \mathfrak{F}_{MB}$. One way to take advantage of this flexibility is to choose R so as to achieve minimum V_R, or simply to achieve $V_R < s$ (when possible) to give $R \in \mathfrak{F}_{MB}$ improvement over those designs which satisfy the Box-Draper conditions of equation (1.1). Moreover, this flexibility may be used to minimize V_R within the class of designs which have equal spacing as is illustrated in Figure 3.2.

As illustration, let

$$R_{2,1}(\xi : \beta, \delta) = \frac{\beta_0 + \beta_1\xi + \beta_2\xi^2}{\delta_0 + \delta_1\xi} \tag{3.1}$$

be the rational function to be approximated where β_0, β_1, β_2 are unknown parameters in E^1, ξ, the controllable variable,

$-\infty < \xi_1 \le \xi \le \xi_2 < +\infty$. If $x = \dfrac{2\xi - \xi_1 - \xi_2}{\xi_2 - \xi_1}$ and $\gamma = \delta_0/\delta_1$, the equation (3.1) may be written as

$$R_{2,1}(x:\theta,\gamma) = \frac{\theta_0 + \theta_1 x + \theta_2 x^2}{\gamma + x},$$

where $x \in \mathfrak{X} = [-1, 1]$, the operability region. In order to use a linear estimation procedure and to avoid undefined terms, it is necessary to restrict γ to be known and $|\gamma| > 1$. The class of simpler models is taken to be $\mathcal{C}_g = \{\eta_0 : \eta_0 = P_1(x:\alpha) = \alpha_0 + \alpha_1 x\}$. The matrix $A = W_{gg}^{-1} W_{gf}$, where

$$W_{gg} = \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{3} \end{bmatrix} ; \quad W_{gf} = \frac{1}{2} \begin{bmatrix} Z, & (2-\gamma Z), & (\gamma^2 Z - 2\gamma) \\ (2-\gamma Z), & (\gamma^2 Z - 2\gamma), & (\frac{2}{3} + 2\gamma^2 - \gamma^3 Z) \end{bmatrix}$$

with $Z = \log_e(\gamma+1) - \log_e(\gamma-1)$. Thus, B_g depends on the specified value of γ, and its minimum may be written as

$$\min B_g(\gamma) = \underline{\theta}'(W_{ff} - W'_{gf} W_{gg}^{-1} W_{gf})\underline{\theta}$$

where

$$W_{ff} = \begin{bmatrix} \frac{1}{Z_1}, & (\frac{\gamma}{Z_1} + \frac{Z}{2}), & (1 + \frac{\gamma^2}{Z_1} + \gamma Z) \\ & (1 + \frac{\gamma^2}{Z_1} - \gamma Z), & (\frac{3\gamma^2 Z}{2} - \frac{\gamma^3}{Z_1} + 2\gamma) \\ \text{Symmetric} & & (\frac{1}{3} + 3\gamma^2 + \frac{\gamma^4}{Z_1} - 2\gamma^3 Z) \end{bmatrix}$$

with $Z_1 = \gamma^2 - 1$. Numerical values of $\min B_g(\gamma)$ can be computed for any value of $\underline{\theta}$. For example, if $\underline{\theta}' = (1, 2, 4)$, $\min B_g(\gamma)$ has been computed for different values of γ as shown in Table 3.1.

Table 3.1 $\min B_g(\gamma)$ when $\underline{\theta}' = (1,2,4)$ and $\eta_0 = \alpha_0 + \alpha_1 x$

γ	$\min B_g(\gamma)$
1.01	320.9836
1.50	1.1653
5.00	0.0510

Let the vector $y'(\underline{x}) = [Y(x_1), Y(x_2), \ldots, Y(x_N)]$ satisfy the conditions

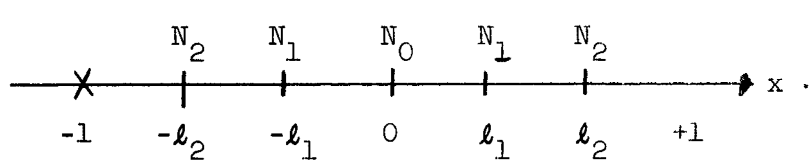
(1) $E[y(\underline{x})] = R\underline{\theta}$

(2) $E\{[y(\underline{x}) - \eta(\underline{x})][y(\underline{x}) - \eta(\underline{x})]'\} = I_N\sigma^2$,

where R is an Nx3 matrix whose $(i,j)^{th}$ element is $\frac{x_j}{\gamma + x_i}$ and σ^2 is a real, positive constant. For a given matrix R such that $\underline{\alpha}$ is estimable, the averaged Var $[\hat{\eta}_0]$ over $[-1, 1]$ is

$$V_R = N\cdot Tr[(R'R)^{-}W'_{gf}W^{-1}_{gg}W_{gf}] .$$

Let $\mathfrak{Z}_{MB}$ be the class of 3, 4 and 5 point designs which are symmetric with respect to the origin. The N observations are spaced as follows: $N = N_0 + 2N_1 + 2N_2$ where N_0 is the number of observations taken at $x = 0$, N_1 is the number of observations at $x = \pm \ell_1$, and N_2 is the number of observations at $x \pm \ell_2$. The above designs can be better described by this following schematic presentation:



(3.2)

For this class of designs, $\underset{R \in \mathfrak{Z}_{MB}}{\text{minimum}} V_R$ has been computed and min V|min B_g designs tabulated for N = 3, 4, ..., 15. These designs depend on the particular value of γ. Tables 3.2, 3.3 and 3.4 give such designs for γ = 1.01, 1.50 and 5.00 respectively.

These tables indicate that more and more design points are forced to the boundary of the operability region $\mathfrak{X}$ as the value of γ increases, i.e., as γ increases the bias decreases and therefore has less and less influence on the choice of design.

For the following class of simpler models

$$\mathcal{C}_g = \{\eta_0 : \eta_0 = P_2(x:\alpha) = \alpha_0 + \alpha_1 x + \alpha_2 x^2\} ,$$

$$W_{gg} = \begin{bmatrix} 1 & 0 & \frac{1}{3} \\ 0 & \frac{1}{3} & 0 \\ \frac{1}{3} & 0 & \frac{1}{5} \end{bmatrix} ; \; W_{gf} = \frac{1}{2} \begin{bmatrix} Z, & (2-\gamma Z) , & (\gamma^2 Z - 2\gamma) \\ & (\gamma^2 Z - 2\gamma) , & (\frac{2}{3} + 2\gamma^2 - \gamma^2 Z) \\ \text{Symmetric} & & (\gamma^4 Z - 2\gamma^3 + \frac{2\gamma}{3}) \end{bmatrix} .$$

For the same set of parameters used in the previous example, i.e., $\underline{\theta}' = (1,2,4)$, min $B_g(\gamma)$ appear in Table 3.5.

From the Tables 3.1 and 3.5 one notices that for a fixed value of γ, min $B_g(\gamma)$, when approximating η by a linear model, is larger than when using a quadratic model. In general it is easy to show that $B_g(L,\gamma) \geq B_g(Q,\gamma)$ where $B_g(L,\gamma)$ and $B_g(Q,\gamma)$ denote min $B_g(\gamma)$ when approximating η by a linear and a quadratic model respectively.

Using the class of designs described in (3.2), $\underset{R \in \mathfrak{Z}_{MB}}{\text{minimum}} V_R$ has been computed and the min V|min B_g designs appear in Tables 3.6,

Table 3.2 min $V|\min B_g$ designs when $\eta_0 = \alpha_0 + \alpha_1 x$ and $\gamma = 1.01$

N	N_0	N_1	ℓ_1	N_2	ℓ_2	min V	p
3	1	1	0.8901	0	------	1.8362	3
4	0	1	0.5788	1	0.9160	1.7157	4
5	1	1	0.7136	1	0.9245	1.7099	5
6	0	2	0.6972	1	0.9398	1.6503	4
7	1	2	0.7518	1	0.9425	1.6669	5
8	0	1	0.7395	3	0.9534	1.6206	4
9	1	3	0.7713	1	0.9543	1.6397	5
10	0	4	0.7634	1	0.9629	1.6015	4
11	1	4	0.7847	1	0.9629	1.6202	5
12	0	5	0.7802	1	0.9702	1.5873	4
13	1	5	0.7954	1	0.9698	1.6049	5
14	0	6	0.8673	1	1.0000	1.5730	4
15	1	6	0.8739	1	1.0000	1.5923	5

p = number of destinct design points

Table 3.3 min V|min B_g designs when $\eta_0 = \alpha_0 + \alpha_1 x$ and $\gamma = 1.50$

N	N_0	N_1	ℓ_1	N_2	ℓ_2	min V	p
3	1	1	0.7908	0	------	1.8748	3
4	0	1	0.3874	1	0.8428	1.8583	4
5	1	1	0.5602	1	0.8711	1.8574	5
6	0	2	0.5013	1	0.9130	1.8415	4
7	1	2	0.5846	1	0.9276	1.8453	5
8	0	3	0.5478	1	0.9681	1.8298	4
9	1	3	0.6022	1	0.9780	1.8349	5
10	0	4	0.5760	1	1.0000	1.8203	4
11	1	4	0.6155	1	1.0000	1.8263	5
12	0	5	0.5911	1	1.0000	1.8176	4
13	1	5	0.6226	1	1.0000	1.8236	5
14	0	6	0.6008	1	1.0000	1.8194	4
15	1	6	0.6272	1	1.0000	1.8246	5

p = number of distinct design points

Table 3.4 min V|min B_g designs when $\eta_0 = \alpha_0 + \alpha_1 x$ and $\gamma = 5.00$

N	N_0	N_1	ℓ_1	N_2	ℓ_2	min V	p
3	1	1	0.8106	0	------	1.8772	3
4	0	1	0.0889	1	1.0000	1.7956	4
5	1	1	0.3993	1	1.0000	1.8147	5
6	2	1	0.6156	1	1.0000	1.8255	5
7	3	1	0.8346	1	1.0000	1.8193	5
8	4	2	1.0000	0	------	1.7939	3
9	3	1	0.3656	2	1.0000	1.8040	5
10	4	1	0.6038	2	1.0000	1.8112	5
11	5	2	0.9814	1	1.0000	1.8106	5
12	6	3	1.0000	0	------	1.7939	3
13	5	1	0.3236	3	1.0000	1.8001	5
14	6	1	0.5871	3	1.0000	1.8281	5
15	7	4	1.0000	0	------	1.8034	3

p = number of distinct design points

Table 3.5 min $B_g(\gamma)$ when $\underline{\theta}' = (1,2,4)$
and $\eta_0 = \alpha_0 + \alpha_1 x + \alpha_2 x^2$

γ	min $B_g(\gamma)$
1.01	251.6719
1.50	0.0754
5.00	0.0005

3.7 and 3.8. Each table corresponds to a particular value of γ, i.e., γ = 1.01, 1.50 and 5.00 respectively.

In the Box and Draper approach when min B_g is achieved, min V = s (the number of parameters in the reduced model). The design flexibility given by the min B_g method allows one to always obtain min $V \leq s$ [Theorem 2.4 (iii)] and in many situations, min $V < s$ for infinitely many designs. As illustration, variance contours have been drawn for six designs which are shown in Figures 3.1 through 3.6. $L(N,N_0,N_1,N_2;\ \gamma)$ denotes the min V|min B_g design with a total of N observations, which has been used to approximate $R_{2,1}(x:\theta,\gamma)$ when η_0 is a linear model. Similar notation is used when η_0 is a quadratic model, i.e., $Q(N,N_0,N_1,N_2;\ \gamma)$. Another way of taking advantage of the design flexibility offered by the min B_g method is to choose the min V|min B_g design within the class of designs having equal spacing which is given by the relations $\ell_2 = 2\ell_1$ or $\ell_1 = 2\ell_2$ as shown in Figure 3.2.

Table 3.6 min V|min B_g designs when

$\eta_0 = \alpha_0 + \alpha_1 x + \alpha_2 x^2$ and $\gamma = 1.01$

N	N_0	N_1	ℓ_1	N_2	ℓ_2	min V	p
3	1	1	0.9412	0	------	3.3145	3
4	0	1	0.7201	1	0.9730	2.1807	4
5	1	1	0.7788	1	0.9768	2.0986	5
6	0	2	0.8585	1	1.0000	1.8194	4
7	1	2	0.8743	1	1.0000	1.8337	5
8	0	3	0.8623	1	1.0000	1.7000	4
9	1	3	0.8741	1	1.0000	1.7238	5
10	0	4	0.0645	1	1.0000	1.6423	4
11	1	4	0.8739	1	1.0000	1.6661	5
12	0	5	0.8659	1	1.0000	1.6088	4
13	1	5	0.8736	1	1.0000	1.6310	5
14	0	6	0.8667	1	1.0000	1.5874	4
15	1	6	0.8732	1	1.0000	1.6077	5

p = number of distinct design points

Table 3.7 min V|min B_g designs when

$\eta_0 = \alpha_0 + \alpha_1 x + \alpha_2 x^2$ and $\gamma = 1.50$

N	N_0	N_1	ℓ_1	N_2	ℓ_2	min V	p
3	1	1	0.9409	0	------	2.6938	3
4	0	1	0.2723	1	1.0000	2.3120	4
5	1	1	0.4662	1	1.0000	2.2474	5
6	0	2	0.4576	1	1.0000	2.2270	4
7	1	2	0.5385	1	1.0000	2.2559	5
8	0	3	0.5093	1	1.0000	2.2733	4
9	1	2	0.3820	2	1.0000	2.2633	5
10	0	3	0.4014	2	1.0000	2.2351	4
11	1	3	0.4613	2	1.0000	2.2322	5
12	0	4	0.4576	2	1.0000	2.2270	4
13	1	4	0.4996	2	1.0000	2.2392	5
14	0	5	0.4892	2	1.0000	2.2442	4
15	1	4	0.4199	3	1.0000	2.2384	5

p = number of distinct design points

Table 3.8 min V|min B_g designs when

$\eta_0 = \alpha_0 + \alpha_1 x + \alpha_2 x^2$ and $\gamma = 5.00$

N	N_0	N_1	ℓ_1	N_2	ℓ_2	min V	p
3	1	1	1.0000	0	------	2.4215	3
4	0	1	0.8888	1	1.0000	2.1520	4
5	1	1	0.1970	1	1.0000	2.2268	5
6	2	1	0.5122	1	1.0000	2.3253	5
7	3	2	1.0000	0	------	2.1931	3
8	4	2	1.0000	0	------	2.1452	3
9	5	2	1.0000	0	------	2.1696	3
10	4	1	0.3348	2	1.0000	2.2460	5
11	5	3	1.0000	0	------	2.1651	3
12	6	3	1.0000	0	------	2.1452	3
13	7	3	1.0000	0	------	2.1563	3
14	8	3	1.0000	0	------	2.1867	3
15	7	4	1.0000	0	------	2.1562	3

p = number of distinct design points

If the _a priori_ knowledge on γ is diffuse rather than sharp, so that only a range for γ can be specified, then one could use the design flexibility available to achieve min B_g for a grid of values of γ, spread over the range specified.

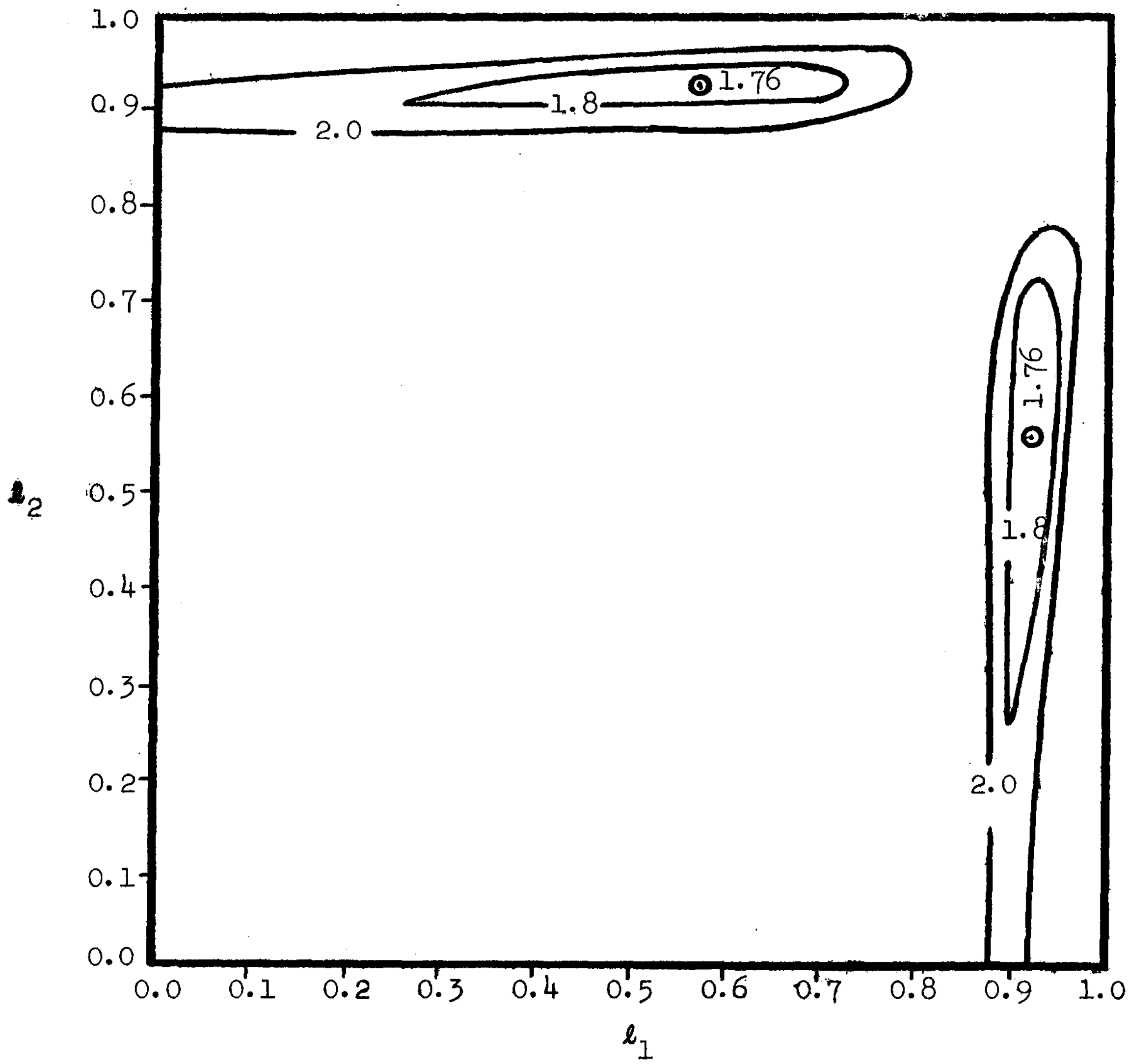


Figure 3.1 Contours of constant V for $L(N,N_0,N_1,N_2;\ \gamma) = L(4,0,1,1;\ 1.01)$

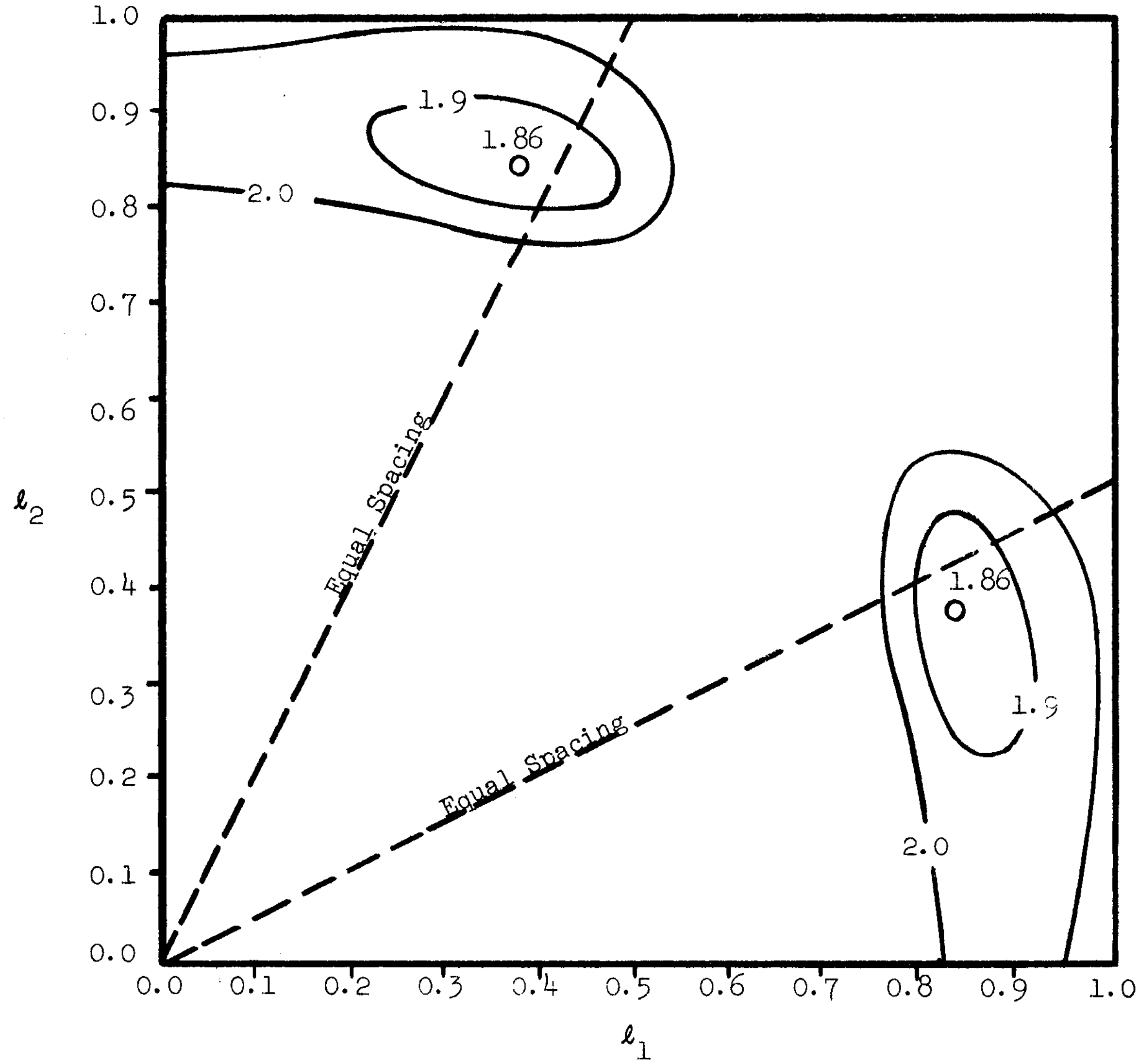


Figure 3.2 Contours of constant V for $L(N,N_0,N_1,N_2;\ \gamma) = L(4,0,1,1;\ 1.50)$

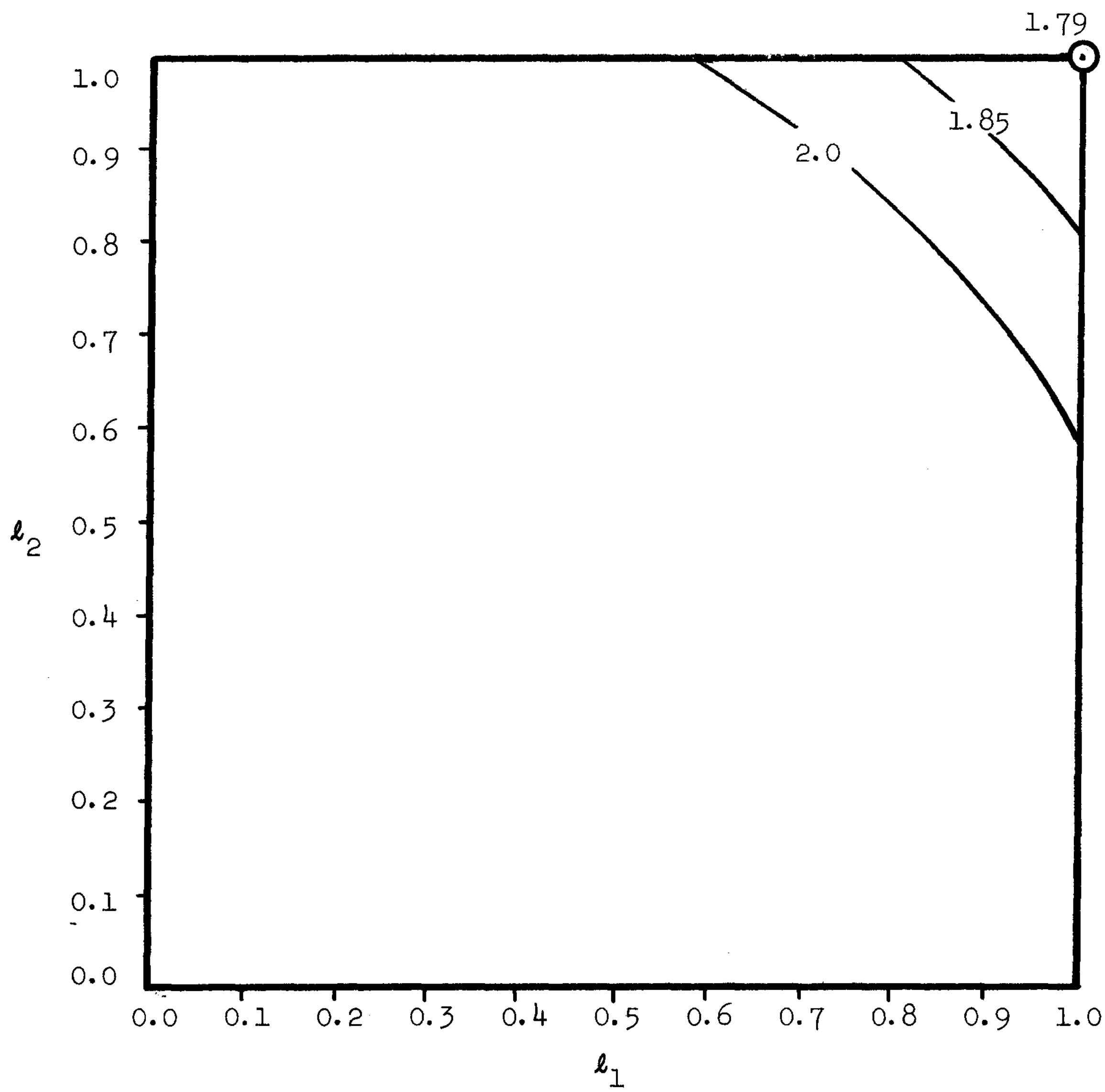


Figure 3.3 Contours of constants V for $L(N, N_0, N_1, N_2; \gamma) = L(8, 4, 1, 1; 5.00)$*

* The 5 point design collapses to a 3 point design because $\ell_1 = \ell_2$

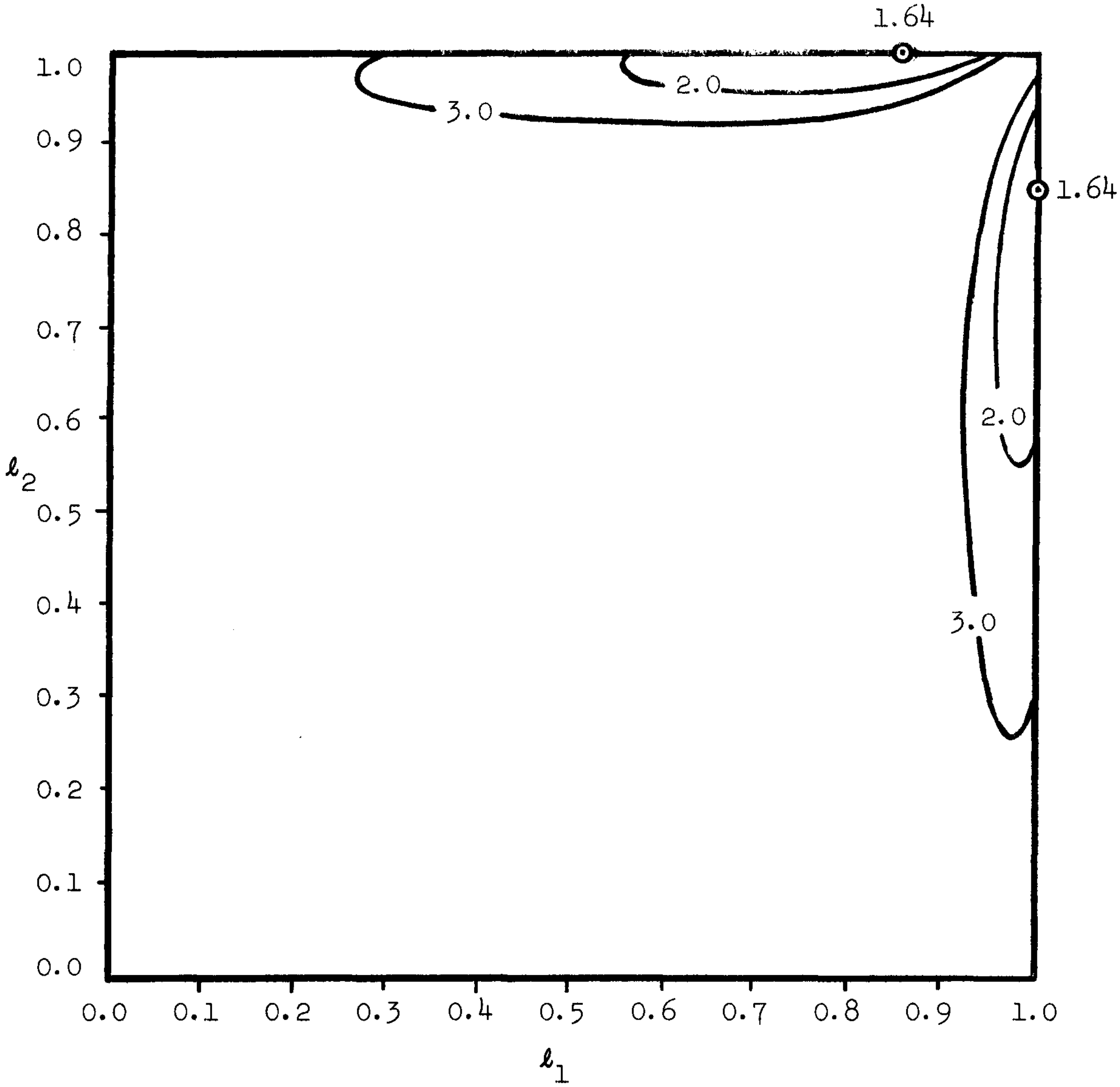


Figure 3.4 Contours of constant V for $Q(N,N_0,N_1,N_2;\ \gamma) = Q(10,0,4,1;\ 1.01)$

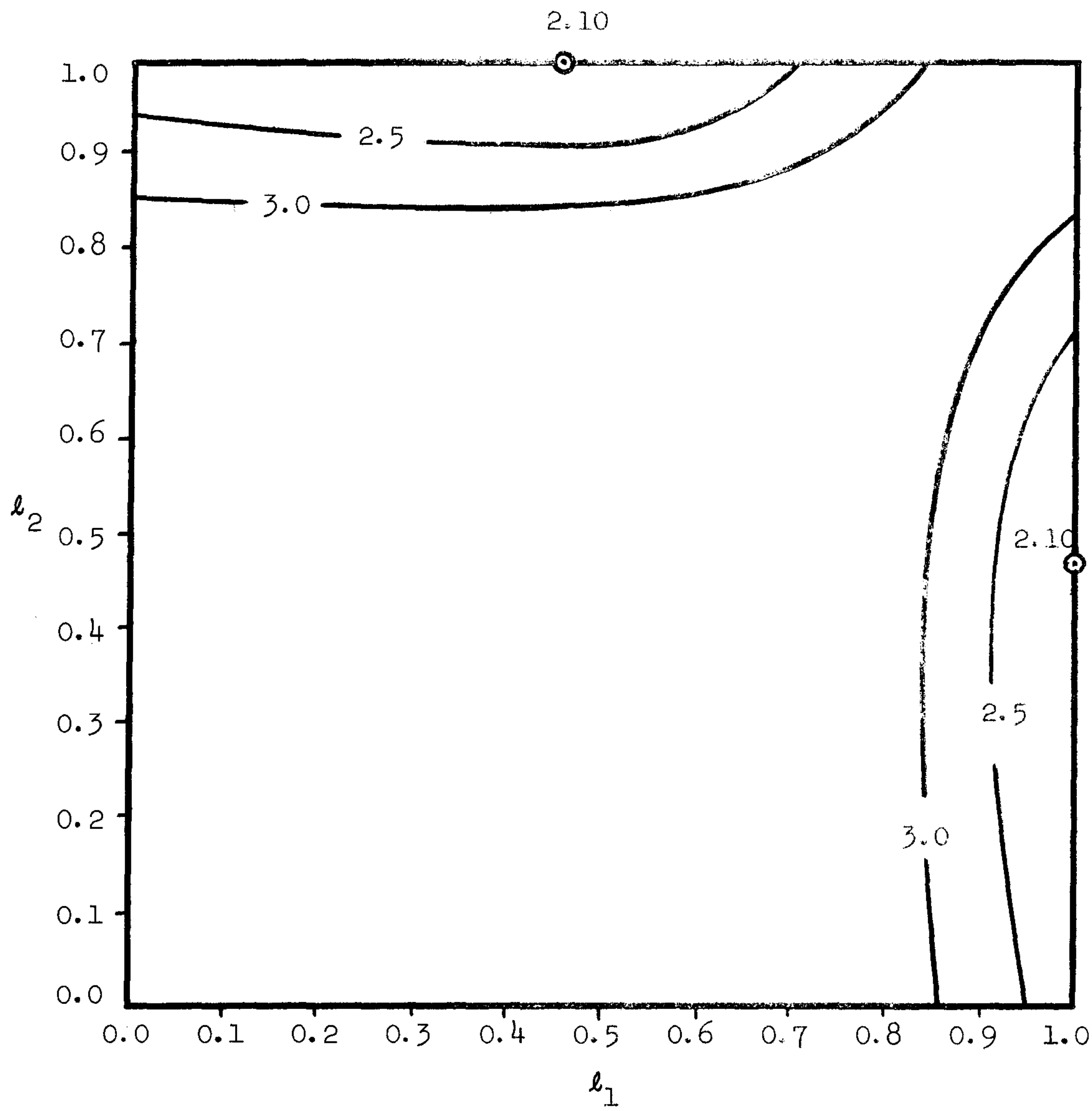


Figure 3.5 Contours of constant V for $Q(N,N_0,N_1,N_2;\ \gamma) = Q(5,1,1,1;\ 1.50)$

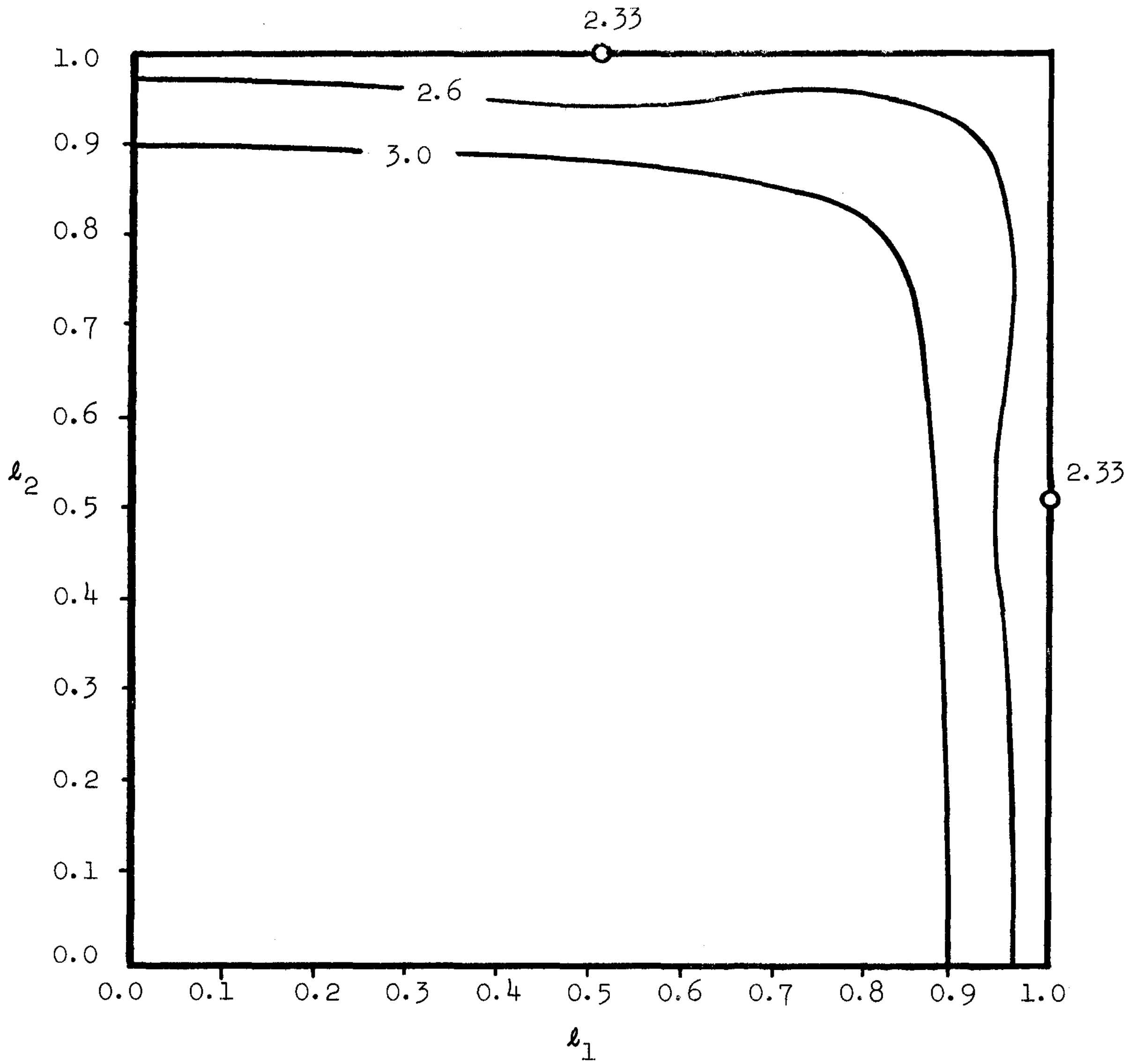


Figure 3.6 Contours of constant V for $Q(N,N_0,N_1,N_2;\ \gamma) = Q(6,2,1,1;\ 5.00)$

4. SUMMARY AND CONCLUSIONS

A general regression model $\eta(x)$ is approximated by the BLUE of a relatively simpler model $\eta_0(x)$. The choice of $\eta_0(x)$ within a fixed class is made so as to minimize the bias. The approximation $\hat{\eta}_0(x)$ subject to satisfying the bias criterion, obtains minimum variance. It is not necessary that the relatively simpler model $\eta_0(x)$ be made up of terms of the full model. The minimum bias technique has been applied to a situation where $\eta(x)$ is a rational function and $\eta_0(x)$ is a simple polynomial function. Examples have been given where $\hat{\eta}_0(x)$ satisfies both bias and variance criteria. Illustration of the great deal of design flexibility allowed by this method has been provided, in particular:

(1) An infinite number of designs achieving minimum bias also give a $V < s$ (variance obtained by designs satisfying the Box-Draper conditions when the reduced model contains s terms).

(2) An infinite number of designs achieving min B_g give $V < s$ and allow equal spacing of design levels.

(3) It is possible to obtain designs which will give min B_g for several values of γ, i.e., protection against inexact knowledge of γ may be obtained by gridwise protection over a range of γ values.

5. REFERENCES

Birkoff, G. and S. MacLane. (1965). A Survey of Modern Algebra, Third Edition, The MacMillan Company, New York.

Box, G. E. P. and N. R. Draper. (1959). A Basis for the Selection of a Response Surface Design. Jour. Amer. Statist. Assoc. 54:622-654.

Box, G. E. P. and N. R. Draper. (1963). The Choice of a Second Order Rotatable Design. Biometrika, 50:335-352.

Elfving, G. (1952). Optimum Allocation in Linear Regression Theory. Ann. Math. Statist. 23:255-262.

Folks, D. L. (1958). Comparison of Designs for Exploration of Response Relationships. Unpublished Ph.D. Thesis, Iowa State College, Microfilm, Inc., Ann Arbor, Michigan.

Gantmacher, F. R. (1960). The Theory of Matrices. Volume 1, Chelsea Publishing Company, New York.

Graybill, F. A. (1969). Introduction to Matrices with Applications in Statistics, Wadsworth Publishing Company, Belmont, California.

Hoel, P. G. and A. Levine. (1964). Optimal Spacing and Weighing in Polynomial Prediction. Ann. Math. Statist. 35:1553-1560.

Karson, M. J., A. R. Manson and R. J. Hader. (1969). Minimum Bias Estimation and Experimental Design for Response Surfaces, Technometrics. 11:461-475.

Kiefer, J. (1959). Optimum Experimental Design. J. Roy. Statist. Soc., (Series B) 21:272-319.

Kiefer, J. (1961). Optimum Designs in Regression Problems II. Ann. Math. Statist. 32:298-325.

Kiefer, J. and J. Wolfowitz. (1959). Optimum Design in Regression Problems. Ann. Math. Statist. 30:271-294.

Rao, C. R. (1967). Linear Statistical Inference and its Applications. John Wiley and Sons, New York.

Scheffé, H. (1967). The Analysis of Variance. John Wiley and Sons, London, England.

6. APPENDIX

A.1 Theorem 2.1

A necessary and sufficient condition for the matrix W_{gg} to be positive definite is that the g_i $(i = 1, 2, \ldots, s)$ be linearly independent.

Proof

Since W_{gg} is positive semi-definite, it suffices to prove that W_{gg} is not singular (Gantmacher, 1960, p. 305). To prove the non-singularity of W_{gg} is equivalent to proving the following: a necessary and sufficient condition for the matrix W_{gg} to be singular is that the g_i $(i = 1, 2, \ldots, s)$ be linearly dependent.

Assume that the g_i $(i = 1, 2, \ldots, s)$ are linearly dependent, then there exist a vector of constants $\underline{c} \neq \underline{0}$ such that $\underline{c}'\underline{g}(x) = 0$ a.e. on $\mathfrak{X}$. Therefore

$$\int [\underline{c}'\underline{g}]^2 d\mu = \Omega^{-1}\underline{c}'W_{gg}\underline{c} = 0 .$$

So W_{gg} is not a positive definite matrix. However, since W_{gg} is positive semi-definite, it must therefore be singular.

Assume now that the g_i are linearly independent. For any vector of constants $\underline{c} \neq \underline{0}$, $\underline{c}'\underline{g}(x) \neq 0$ a.e. on $\mathfrak{X}$. Hence,

$$\int [\underline{c}'\underline{g}]^2 d\mu = \Omega^{-1}\underline{c}'W_{gg}\underline{c} \neq 0 .$$

Thus W_{gg} is positive definite and is therefore non-singular.

A.2 Theorem 2.2

The vector of parameters $\underline{\alpha} = A\underline{\theta}$ is estimable if and only if $A' = V_1T$. V_1 is a nxr matrix whose columns are the orthonormal characteristic vectors corresponding to the r non-zero characteristic roots of $(F'\Sigma^{-1}F)$ with $s \leq r \leq n$, and T is a rxs matrix of full rank.

Proof

We know that $A\underline{\theta}$ is estimable if and only if the rows of A are in the columns space of $(F'\Sigma^{-1}F)$, (Scheffé, 1959). Since the characteristic vectors corresponding to the non-zero characteristic roots of $(F'\Sigma^{-1}F)$ span the same space as do the columns of $(F'\Sigma^{-1}F)$, then we have the required result.

A.3 Theorem 2.3

If $\underline{\alpha}$ is estimable and $A = W_{gg}^{-1}W_{gf}$ is of full rank, then the matrix $A(F'\Sigma^{-1}F)^{-}A'$ is non-singular.

Proof

Assume that $A(F'\Sigma^{-1}F)^{-}A'$ is singular. Then there exists a vector $\underline{d} \neq \underline{0}$ such that

$$A(F'\Sigma^{-1}F)^{-}A'\underline{d} = \underline{0} \; . \qquad (A.3.1)$$

Using the fact that $\underline{\alpha}$ is estimable and premultiplying by $\underline{d}'$, equation (A.3.1) is written as

$$\underline{d}'A(F'\Sigma^{-1}F)^{-}(F'\Sigma^{-1}F)(F'\Sigma^{-1}F)^{-}A'\underline{d} = 0 \; . \qquad (A.3.2)$$

Since Σ^{-1} is positive definite, there exists a non-singular matrix P such the $\Sigma^{-1} = PP'$. Thus equation (A.3.2) may be written as

$$[\underline{d}'A(F'\Sigma^{-1}F)^{-}F'P][P'F(F'\Sigma^{-1}F)^{-}A'\underline{d}] = 0 \; . \qquad (A.3.3)$$

From equation (A.3.3) it follows that

$$P'F(F'\Sigma^{-1}F)^{-}A'\underline{d} = \underline{0} . \qquad (A.3.5)$$

Premultiplying equation (A.3.5) by F′P, gives

$$(F'\Sigma^{-1}F)(F'\Sigma^{-1}F)^{-}A'\underline{d} = \underline{0} . \qquad (A.3.6)$$

Since $\underline{\alpha}$ is estimable, equation (A.3.6) becomes

$$A'\underline{d} = \underline{0}$$

which is a contradiction since A is of full rank. Therefore $A(F'\Sigma^{-1}F)^{-}A'$ is non-singular.

A.4 Lemma 2.1

If $A = [A_1 \vdots A_2]$ and A_1 is a sxs matrix of full rank, then a necessary condition for $A\underline{\theta}$ to be estimable is that F_1 be of full rank.

Proof

Assume that $A\underline{\theta}$ is estimable. Then A = CF where C is an sxN matrix. So $[A_1 \vdots A_2] = C[F_1 \vdots F_2]$, i.e., $A_1 = CF_1$. Therefore $A_1\underline{\theta}_1$ is estimable. Since the rank of A_1 is s, the rank of F_1 must be of at least s. Thus F_1 is of full rank.

A.5 Lemma 2.2

Using the notation of equation (2.8)

(i) $\mathfrak{F}_{FM} \subseteq \mathfrak{F}_{MB} \subseteq \mathfrak{F}_{RM}$

(ii) $\mathfrak{F}_{BD} \subseteq \mathfrak{F}_{MB}$

(iii) When N = s, $\mathfrak{F}_{MB} = \mathfrak{F}_{BD}$.

Proof

(i) If $\underline{\theta}$ is estimable then $A\underline{\theta}$ is also. A necessary condition for $A\underline{\theta}$ to be estimable is that F_1 be of full rank. (Lemma 2.1.) Therefore $\underline{\theta}_1$ is estimable. Hence $\mathfrak{F}_{FM} \subseteq \mathfrak{F}_{MB} \subseteq \mathfrak{F}_{RM}$.

(ii) The estimability condition of $A\underline{\theta}$ is $A(F'F)^{-}F'F = A$. But one can write $A(F'F)^{-}F'F$ as

$$A(F'F)^{-}F'F = [I_s \vdots\ (F_1'F_1)^{-}F_1'F_2(I_s - \Delta^{-}\Delta) + W_{gg}^{-1}W_{gf_2}\Delta^{-}\Delta] \quad (A.5.1)$$

where $\Delta^{-} = [F_2'F_2 - F_2'F_1(F_1'F_1)^{-1}F_1'F_2]^{-}$. Thus, if $F \in \mathfrak{F}_{BD}$, equation (A.5.1) becomes

$$A(F'F)^{-}F'F = [I_s \vdots\ W_{gg}^{-1}W_{gf_2}] = A .$$

Hence $F \in \mathfrak{F}_{MB}$ and $\mathfrak{F}_{BD} \subseteq \mathfrak{F}_{MB}$.

(iii) When $N = s$ and $\mathfrak{F}_{MB} = \emptyset$, the result obviously holds. However, when $N = s$ and $\mathfrak{F}_{MB} \neq \emptyset$, one can write $A(F'F)^{-}F'F$ as

$$A(F'F)^{-}F'F = [I_s \vdots\ (F_1'F_1)^{-1}F_1'F_2] .$$

Then the only way for $A\underline{\theta}$ to be estimable is for

$$(F_1'F_1)^{-1}F_1'F_2 = W_{gg}^{-1}W_{gf_2},$$

i.e., $F \in \mathfrak{F}_{BD}$. Therefore $\mathfrak{F}_{MB} = \mathfrak{F}_{BD}$ when $N = s$.

A.6 Theorem 2.4

Using the notation defined in equations (2.9) and (2.10)

(i) $V_{FM} \geq V_{MB}$

(ii) $V_{MB} \geq V_{RM}$

(iii) $V_{MB} \leq V_{BD} = s$.

Proof

(i) For any $F \in \mathfrak{F}_{FM}$,

$$V_{FM}(F) = N\cdot Tr[(F'F)^{-1}W_{ff}] \quad \text{and} \quad V_{MB}(F) = N\cdot Tr[(F'F)^{-1}W'_{gf}W_{gg}^{-1}W_{gf}] \ .$$

Therefore,

$$V_{FM}(F) - V_{MB}(F) = N\cdot Tr[(F'F)^{-1}(W_{ff}-W'_{gf}W_{gg}^{-1}W_{gf})] \ .$$

By Corollary 2.1, it follows that $V_{FM}(F) \geq V_{MB}(F)$.

Since $\mathfrak{F}_{FM} \subseteq \mathfrak{F}_{MB}$, (Lemma 2.2, (i)), $V_{FM} \geq V_{MB}$.

(ii) For any $F = [F_1 \vdots F_2] \in \mathfrak{F}_{MB}$, F_1 is of full rank (Lemma 2.1) and

$$V_{RM}(F) = N\cdot Tr[(F_1'F_1)^{-1}W_{gg}] \quad \text{and} \quad V_{MB}(F) = N\cdot Tr[A(F'F)^{-}A'W_{gg}] \ .$$

One can write $A(F'F)^{-}A'$ as

$$A(F'F)^{-}A' = (F_1'F_1)^{-1} - Q'\Delta^{-}Q$$

where $\Delta^{-} = [F_2'F_2 - F_2'F_1(F_1'F_1)^{-1}F_1'F_2]^{-}$ is a positive semi-definite matrix and $Q = [F_2'F_1(F_1'F_1)^{-1} - W'_{gf_2}W_{gg}^{-1}]$. Therefore, for any $F \in \mathfrak{F}_{MB}$

$$V_{MB}(F) - V_{RM}(F) = N\cdot Tr[Q'\Delta^{-}Q] \geq 0 \tag{A.6.1}$$

Hence Lemma 2.2 (i) gives $V_{MB} \geq V_{RM}$.

(iii) When $F \in \mathfrak{F}_{BD}$, the strict equality holds in equation (A.6.1) and

$$V_{MB}(F) = V_{BD}(F) = N\cdot Tr[(F_1'F_1)^{-1}W_{gg}] = Tr[I_s] = s .$$

From Lemma 2.2 (ii), it follows that $V_{MB} \le V_{MD} = s$.

A.7 Theorem 2.5

To the model given in equation (2.7)

$$\eta_1 = \underline{\theta}_1'\underline{f}_1 + \underline{\theta}_2'\underline{f}_2 = \underline{\theta}'\underline{f}$$

add t more terms, say $\underline{\theta}_3'\underline{f}_3$. The new model is

$$\eta_2 = \underline{\theta}'\underline{f} + \underline{\theta}_3'\underline{f}_3 .$$

Let $\mathcal{C}_g = \{\eta_0 : \eta_0 = \underline{\alpha}'\underline{g}$ where $g_k = f_k$ $(k = 1, 2, \ldots, s)\}$ be the class of simpler models. Let V_i denote the averaged variance of the min $V|$min B_g linear estimator of η_i $(i = 1, 2)$. The $V_1 \le V_2$.

Proof

Let the design matrix $F_* = [F \vdots F_3]$ be partitioned so as to match the partitioning of the model η_2. Then $V_1 = N\cdot Tr[A_1(F'F)^{-}A_1'W_{gg}]$ where $A_1 = [I_s \vdots W_{gg}^{-1}W_{gf_2}]$ and $V_2 = N\cdot Tr[A(F_*'F_*)^{-}A'W_{gg}]$ where $A = [A_1 \vdots W_{gg}^{-1}W_{gf_3}]$. The matrix $A(F_*'F_*)^{-}A'$ may be expanded as follows:

$$A(F_*'F_*)^{-}A' = A_1(F'F)^{-}A_1' - U'\Delta^{-}U$$

where $U = [A_1(F'F)^{-}F'F_3 - W_{gf_3}'W_{gg}^{-1}]$ and $\Delta^{-} = [F_3'F_3 - F_3'F(F'F)^{-}F'F_3]^{-}$ is a positive semi-definite matrix. Therefore

$$A(F_*'F_*)A' - A_1(F'F)^{-}A_1' = U'\Delta^{-}U$$

and

$$\underline{g}'A(F_*'F_*)^{-}A'\underline{g} \geq \underline{g}'A_1(F'F)^{-}A_1'\underline{g} .$$

Hence $V_2 \geq V_1$.