THE DETERMINATION OF SOME OPTIMAL EXPERIMENTAL DESIGNS

By

BILLY EARL GILLETT

Bachelor of Science
Oklahoma State University
Stillwater, Oklahoma
1961

Master of Science
Oklahoma State University
Stillwater, Oklahoma
1963

Submitted to the Faculty of the Graduate School of
the Oklahoma State University
in partial fulfillment of the requirements
for the degree of
DOCTOR OF PHILOSOPHY
May, 1964

THE DETERMINATION OF SOME OPTIMAL

EXPERIMENTAL DESIGNS

Thesis Approved:

J Leroy Folks
Thesis Adviser

Potter G. Hoffman

Robert D Morrison

John B West

David L. Weeks

J K Boggs
Dean of the Graduate School

ACKNOWLEDGMENTS

For suggesting the problem and for his aid and guidance in directing an examination of such, I am deeply indebted to Dr. Leroy Folks.

During the past three years, I have received a National Defense Education Act Fellowship which provided financial support for my education. For this fellowship, I am very grateful.

An expression of gratitude is also extended to Professor John E. Hoffman, John B. West, David L. Weeks and Robert D. Morrison for serving on my advisory committee.

TABLE OF CONTENTS

Chapter | Page

LIST OF TABLES

CHAPTER I

INTRODUCTION

1.1 Description of the Problems

Many articles have appeared in the literature concerning the problem of determining optimal designs with respect to certain criteria once a particular model has been assumed; however, hardly any work has been devoted to determining designs that are optimal for general sample size N. In determining optimal designs for an experiment, it is usually assumed that the number of design points N can be divided in any desired manner. This assumption can not always be met. Thus, we are faced with the problem of selecting the design points so that the design remains optimal in the sense of certain criteria for all values of N. Chapters II and III will be devoted primarily to this problem. Specifically, chapter II deals with the problem of trying to determine the design that will minimize the maximum variance of the estimated response when we assume the model

$$y = \beta_o + \beta_1 x_1 + \beta_2 x_2 + \varepsilon ,$$

where ε $(0, \sigma^2)$, and we only have three design points. Chapter III deals with determining exact optimal designs for general sample size N.

Another problem in the area of response relationships which is of major importance is the problem of deciding which model should be assumed. There are usually several risks involved in the selection of a model. In chapter IV some function of the bias is used as the risk function. A model is then selected from a particular class of models which will minimize this risk function.

Chapter V will consider some methods of determining an average variance of the estimated response in the two-dimensional case for any distribution of the total probability mass to the region of interest when we assume the model

$$y = \beta_o + \beta_1 x_1 + \beta_2 x_2 + \varepsilon .$$

If we have reason to believe that the variance of response will be different in one particular subregion than in another, we may want to assign a larger proportion of the probability mass to this region. The average variance of the estimated response over both a square region R and a circular region R_c will be derived in this chapter. The minimum average variance of the estimated response with respect to the division of R and R_c and the distribution of the total probability mass to these subdivisions is also determined. Other results concerning the average variance of the estimated response are obtained.

1.2 Definitions and Notation

The design points are the points in a p-dimensional space where the observations are to be taken. The p-dimensional space consisting of all possible design points will be called the factor space. An experimental design will be defined as a procedure which indicates where the design points are located and how many observations are to be taken at each design point.

In this thesis we shall consider only one and two-dimensional models. In the one-dimensional case, the response will be given by

$$y = \sum_{i=1}^{K} a_i f_i(x) + \varepsilon ,$$

where $-1 \leq x \leq 1$ and $f_i(x)$ is a real valued function of x. The design matrix for N observations will be denoted by

$$X = \begin{bmatrix} f_1(x_1) & f_2(x_1) & \cdots & f_K(x_1) \\ \vdots & & & \\ f_1(x_j) & f_2(x_j) & \cdots & f_K(x_j) \\ \vdots & & & \\ f_1(x_N) & f_2(x_N) & \cdots & f_K(x_N) \end{bmatrix} .$$

Let u be a variable point in the one-dimensional factor space. Also, let the vector $U = [f_1(u_j)\, f_2(u_j) \ldots f_K(u_j)]$ be the j-th row of X.

In the two-dimensional case, the response will be given by

$$y = \sum_{i=1}^{k} a_i f_i(x_1) + \sum_{i=k+1}^{\ell} a_i f_i(x_2) + \sum_{i=\ell+1}^{m} a_i f_i(x_1, x_2) + \varepsilon ,$$

where $-1 \leq x_1 \leq 1$, $-1 \leq x_2 \leq 1$, $f_i(x_k)$ is a real valued function of x_k $(k = 1, 2)$, and $f_i(x_1, x_2)$ is a real valued function of x_1 and x_2. Let

$$U = [f_1(u_{j1}) f_2(u_{j1}) \ldots f_k(u_{j1}) f_{k+1}(u_{j2}) \ldots f_\ell(u_{j2}) f_{\ell+1}(x_{j1}, x_{j2}) \ldots f_m(x_{j1}, x_{j2})]$$

be the j-th row of X, where

$$X = \begin{bmatrix} f_1(x_{11}) f_2(x_{11}) \ldots f_k(x_{11}) f_{k+1}(x_{12}) \ldots f_\ell(x_{12}) f_{\ell+1}(x_{11}, x_{12}) \ldots f_m(x_{11}, x_{12}) \\ \vdots \\ f_1(x_{j1}) f_2(x_{j1}) \ldots f_k(x_{j1}) f_{k+1}(x_{j2}) \ldots f_\ell(x_{j2}) f_{\ell+1}(x_{j1}, x_{j2}) \ldots f_m(x_{j1}, x_{j2}) \\ \vdots \\ f_1(x_{N1}) f_2(x_{N1}) \ldots f_k(x_{N1}) f_{k+1}(x_{N2}) \ldots f_\ell(x_{N2}) f_{\ell+1}(x_{N1}, x_{N2}) \ldots f_m(x_{N1}, x_{N2}) \end{bmatrix} .$$

The model in either case is given by $Y = X\beta + \varepsilon$, where β is a vector of the a_i's. The response at any point u in the factor space is estimated by

$$\hat{y}(u) = U\hat{\beta}$$
$$= U(X'X)^{-1}X'Y,$$

where $\hat{\beta}$ is the least squares estimate of β.

The variance of the estimated response, denoted by var $\hat{y}(u)$, is given by

$$\text{var } \hat{y}(u) = U(X'X)^{-1}U'\sigma^2.$$

In this thesis σ^2 will always be considered equal to unity. The absolute value of the bias of the estimated response, denoted by $|\text{bias }(u)|$, is defined to be

$$|\text{bias }(u)| = |E[\hat{y}(u)] - E[y(u)]| ,$$

where $y(u)$ is the response at u.

The following optimality criteria will be considered. We will want to find the design that will:

1) minimize the maximum variance of $\hat{y}(u)$, denoted by $\min_x \max_u \text{var } \hat{y}(u)$

2) minimize the average variance of $\hat{y}(u)$, denoted by $\min_x \int_R \text{var } \hat{y}(u)\, [f(u)]\, du$

3) minimize the maximum absolute value of the bias of $\hat{y}(u)$ which arises from fitting the wrong model, denoted by $\min_x \max_u |\text{bias}(u)|$

4) minimize the average absolute value of the bias of $\hat{y}(u)$, denoted by $\min_x \int_R |\text{bias}(u)| f(u)\,du$

5) minimize the average bias of $\hat{y}(u)$ squared, denoted by $\min_x \int_R [\text{bias}^2(u)]\ f(u)\,du$.

1.3 Review of the Literature

In the area of response relationships, a rather detailed review of the literature through 1958 has been presented by Folks (3). Since 1958 a number of articles have appeared which approach the optimal design problem from a probability measure standpoint. Such is the case in articles by Kiefer (4), (5), Kiefer and Wolfowitz (6), (7), and Aitchison (1). Further work in the area of optimal designs has been presented by Box and Draper (2).

Although much work has been done to try to determine optimal designs, usually one of the following conditions is assumed:

1) the number of design points can be divided in any desired manner

2) the optimal design obtained is only optimal to within a given approximation of the true theoretical optimal design.

The latter is the case in the articles by Kiefer (4), (5), Kiefer and Wolfowitz (6), (7), and Aitchison (1). In contrast, Folks (3), approaches the problem of determining optimal experimental designs for various criteria by considering two cases; namely, the case where the number of design points N is even and the case where N is odd. By this procedure, exact optimal designs were determined in the one-dimensional case for the following criteria:

(i) $\min_x \max_u \operatorname{var} \hat{y}(u)$

(ii) $\min_x \operatorname{ave}_u \operatorname{var} \hat{y}(u)$

(iii) $\min_x \operatorname{gen} \operatorname{var} \hat{y}(u)$,

where $\operatorname{var} \hat{y}(u)$ is the variance of the estimated response at u. Also, exact optimal designs for bias and mean square error considerations in the one-dimensional case and for variance and bias considerations in the two-dimensional case were determined when the number of design points was a certain multiple of four.

In the past few years, considerable effort has been put forth by Kiefer (4), (5), and Kiefer and Wolfowitz (6), (7), to determine optimal designs for existing criteria and to determine new criteria of goodness. In order to present a summary of their work, it will be necessary to introduce some definitions and notation. Until specified, all work will be concerned with the one-dimensional case. In the following situations the model is assumed to be

$$y_x = \sum_{i=1}^{k} a_i f_i(x) + e,$$

where $f_1, \ldots, f_K$ are linearly independent real - valued functions. A design is a discrete probability measure which assigns to each point in the sample space a measure equal to an integral multiple of N^{-1}. It was established by Kiefer and Wolfowitz (7) that the criteria of optimality

(1) $\min_x \max_u \operatorname{var} \hat{y}(u)$, and

$$(2) \quad \min_{x} \text{gen var } \hat{y}(u)$$

are equivalent when all probability measures are considered rather than just integral multiples of N^{-1}. Optimal designs in the sense of (1) and (2) above were determined when the inference about the regression coefficients concerned S of the K coefficients, where $S = 1, 2, \ldots, K$, and when the inference concerned the entire regression function. Necessary and sufficient conditions for a design to be optimal were established when the inference concerned the whole regression function. Still in the one-dimensional case, two other criteria of optimality were presented and were shown to be equivalent to criteria (1) and (2) above. In the q-dimensional case, Kiefer (4) considers optimal designs for quadratic regression.

The following are a few examples of the optimal designs determined by Kiefer and Wolfowitz. Consider first the case where the inference concerns S of the K regression coefficients. For $S=1$, let

$$y_x = \sum_{i=1}^{K} a_i f_i(x) + \varepsilon = \sum_{i=1}^{K} a_{i-1} x^{i-1} + \varepsilon, \quad -1 \leq x \leq 1.$$

The unique optimal design d, in the sense of (1) and (2) above, is given by

$$d(-1) = d(1) = 1/2\,(K-1)$$

$$d[\cos(j\pi/(K-1))] = 1/(K-1) \qquad (j = 1, 2, \ldots, K-2),$$

where $d(x)$ denotes the probability mass assigned to the design point x by the design d. For $S = K$, let

$$y_x = \sum_{i=1}^{K} a_i f_i(x) + \varepsilon.$$

For the sample space consisting of K points, the unique optimal design d, in the sense of (1) and (2) above is given by

$$d(x) = 1/K.$$

Consider next an example where the whole regression function is estimated. Let

$$y_x = \sum_{i=1}^{K} a_i f_i(x) + \varepsilon = \sum_{i=1}^{K} a_{i-1} x^{i-1} + \varepsilon, \quad -1 \leq x \leq 1.$$

The unique optimal design, in the sense of (1) and (2) above, assigns mass equal to $1/K$ to the points $x = -1$, $x = 1$, and the roots of $L'_h(x) = 0$, where $L'_h(x)$ is the derivative of the Legendre polynomial.

In the q-dimensional case, assume the inference concerns the estimation of the whole regression function rather than just S out of K regression coefficients. For quadratic regression with $q = 1$, an optimal design d, in the sense of (1) and (2) above, is that measure which puts equal weights on the points $x = -1, 0, 1$. For $q = 2, 3, 4, 5$, optimal designs are given by Kiefer (4). A fact worth noting is that when $q = 2$, the design which assigns measure $1/9$ to each of the nine points designated by the optimal design, yields a generalized variance of $\hat{y}(u)$ which is 15% larger and a maximum variance of $\hat{y}(u)$ which is 21% larger than does the optimal design.

Aitchison (1) constructed optimal designs which concentrated on the detection of certain specific effects while allowing at least the inspection of a wider class of effects. All of his work was done in the framework of a one-way classification model

$$Z_{jk} = \mu + y_j + e_{jk} \quad (j = 1, 2, \ldots, t; \; k = 1, 2, \ldots n),$$

where

$$Z_{jk} \sim N(\theta_j, \sigma^2).$$

Box and Draper (2) considered the problem of fitting a first degree polymonial f(X) over the region R, when the true function g(X) is quadratic, where X is a K-dimensional vector. There are two types of error which occur; namely, variance error, that due to sampling error, and bias error, the failure of f(X) to represent g(X). In the cases they considered, the optimal design in which variance error and bias error both occurred was almost identical to the design that would have been obtained if variance error were ignored completely and the experiment designed to minimize bias error alone. Also, it was proved that if the method of least squares is used to fit a polynomial of any degree d_1 over a region R when the true function is a polynomial of degree $d_2 > d_1$, then the bias averaged over R is minimized for all values of the coefficients of neglected terms by making the moments of order $d_1 + d_2$ and less of the design points equal to the corresponding moments of a uniform distribution over R. Box and Draper further indicated that the variance should be minimized if it is rather definitely known that the true function is linear. Conversely, the bias alone should be minimized if the assumption of linearity can not be made and observational errors are negligible. Another result obtained was that if bias alone had to be considered while nothing whatever were known about the true function other than it could be represented by a polynomial with infinitely many terms, then we would do best by spreading the design points evenly over the region R.

CHAPTER II

THREE POINT PROBLEM

Exact optimal designs have been determined by Folks (3) in the two-dimensional case for several criteria when the number of design points N is expressible as some multiple of four. Thus, it is desirable to determine optimal designs for all values of N. In this chapter we shall choose N equal to three and assume the model

$$y = \beta_o + \beta_1 x_1 + \beta_2 x_2 + \varepsilon,$$

where $\varepsilon \sim (0, 1)$. Let

$$R = [(u_1, u_2) \mid -1 \leq u_i \leq 1; \ i = 1, 2]$$

be the region of experimentation. We are justified in using such a region without loss of generality, due to invariance properties of optimal designs which were proved by Folks (3). The response at any point u is estimated by

$$\begin{aligned} \hat{y}(u) &= U\hat{\beta} \\ &= U(X'X)^{-1}X'Y, \end{aligned}$$

where $U = (1\ u_1 u_2)$, and

$\hat{\beta}$ is the least squares estimate of β.

The expected value of $\hat{y}(u)$ is given by

$$\begin{aligned} E[\hat{y}(u)] &= E[U\hat{\beta}] \\ &= U\beta, \end{aligned}$$

and the variance of $\hat{y}(u)$ is given by

$$\text{var } \hat{y}(u) = \text{var } [U \hat{\beta}]$$
$$= U(X'X)^{-1} U',$$

where

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} \\ 1 & x_{21} & x_{22} \\ 1 & x_{31} & x_{32} \end{bmatrix}.$$

Expanding the variance of $\hat{y}(u)$, we have

$$\text{var } \hat{y}(u) = \text{var } \hat{\beta}_o + u_1^2 \text{ var } \hat{\beta}_1 + u_2^2 \text{ var } \hat{\beta}_2 + 2u_1 \text{ cov } (\hat{\beta}_o, \hat{\beta}_1)$$
$$+ 2u_2 \text{ cov } (\hat{\beta}_o, \hat{\beta}_2) + 2u_1u_2 \text{ cov } (\hat{\beta}_1, \hat{\beta}_2).$$

As an initial investigation of optimal properties, consider the problem of trying to find the design that will minimize the maximum variance of $\hat{y}(u)$. It has been shown by Folks (3) that the $\min_x \max_u$ var $\hat{y}(u)$ is achieved by taking $N/4$ points at each corner of the square region R; however, this is an impossible task with N equal to 3. Thus, we shall investigate further to try to determine the min max var $\hat{y}(u)$ design for N equal to 3. As an aid in our investigation, consider some relevant theorems.

Every design determines the unique family of variance contours

$$\text{var } \hat{y}(u) = U(X'X)^{-1}U' = K,$$

which are ellipses. The variance contour with K equal to 1 passes through the three design points. This is pointed out in the following theorem.

Theorem 2.1 For any choice of 3 design points in the square region

$$R = [(u_1, u_2) \mid -1 \leq u_i \leq 1 ; i = 1, 2],$$

such that X is of full rank, the variance of the estimated response at each of the design points is 1. That is,

$$\text{var } \hat{y}(u) = U(X'X)^{-1} U' = 1.$$

Proof: Consider the covariance matrix of the vector of predicted responses at the design points, denoted by $\hat{Y}$.

$$\text{cov } \hat{Y} = X(X'X)^{-1} X'.$$

Since X is square and of full rank

$$\text{cov } \hat{Y} = I.$$

Therefore,

$$\text{var } \hat{y}(x_i) = 1.$$

As a further aid to establishing the $\min_x \max_u \text{var } \hat{y}(u)$ design for the three-point problem, consider the following theorems.

Theorem 2.2 For any choice of design, the maximum variance of the estimated response, $\hat{y}(u)$, occurs at one or more corners of the square region

$$R = [(u_1, u_2) \mid -1 \leq u_i \leq 1 ; i = 1, 2],$$

in the three-point problem.

Proof:

$$\text{var } \hat{y}(u) = \text{var } \hat{\beta}_o + u_1^2 \text{ var } \hat{\beta}_1 + u_2^2 \text{ var } \hat{\beta}_2 + 2u_1 \text{ cov } (\hat{\beta}_o, \hat{\beta}_1)$$
$$+ 2 u_2 \text{ cov } (\hat{\beta}_o, \hat{\beta}_2) + 2u_1u_2 \text{ cov } (\hat{\beta}_1, \hat{\beta}_2).$$

Let

$$A_i = \text{var } \hat{\beta}_i \qquad (i = 0, 1, 2)$$

$$D = 2 \text{ cov } (\hat{\beta}_o, \hat{\beta}_1)$$

$$E = 2 \text{ cov } (\hat{\beta}_o, \hat{\beta}_2)$$

$$F = 2 \text{ cov } (\hat{\beta}_1, \hat{\beta}_2).$$

Let $(1 - \varepsilon, 1 - \ell)$ be the point in R at which var $\hat{y}$ (u) attains its maximum, where $0 \leq \varepsilon \leq 2$, and $0 \leq \ell \leq 2$. Consider the class of designs specified by the signs of D, E, and F.

<u>Case 1:</u> $D, E, F \geq 0$

$$\text{var } \hat{y}\,(1-\varepsilon, 1-\ell) = A_o + (1-\varepsilon)^2 A_1 + (1-\ell)^2 A_2 + (1-\varepsilon)D + (1-\ell)E + (1-\varepsilon)(1-\ell)F$$

$$(2.1) \qquad = A_o + A_1 + A_2 + D + E + F - [(2\varepsilon - \varepsilon^2)A_1 + (2\ell - \ell^2)A_2 + \varepsilon D + \ell E + (\varepsilon + \ell - \varepsilon\ell) F].$$

Let $\varepsilon = 1 + a$, where $-1 \leq a \leq 1$, then

$$\varepsilon + \ell - \varepsilon\ell = 1 + a + \ell - (1+a)\ell$$

$$= 1 + a(1 - \ell).$$

Now $-1 \leq a \leq 1$ and $-1 \leq (1 - \ell) \leq 1$ implies $-1 \leq a(1-\ell) \leq 1$, which implies $0 \leq 1 + a(1-\ell) \leq 2$. Thus $\varepsilon + \ell - \varepsilon\ell \geq 0$. Also, since $0 \leq \varepsilon \leq 2$ and $0 \leq \ell \leq 2$, we have

$$2\varepsilon - \varepsilon^2 = \varepsilon(2 - \varepsilon) \geq 0$$

$$2\ell - \ell^2 = \ell(2 - \ell) \geq 0.$$

Therefore, since all quantities in the brackets of (2.1) are positive or zero, and zero only when $\varepsilon = \ell = 0$, we have that the var $\hat{y}$ (u) is maximum only when $\varepsilon = \ell = 0$, which implies

$$\max_u \text{var } \hat{y}(u_1, u_2) = \text{var } \hat{y}(1, 1).$$

Case 2: $D, E \geq 0, \quad F \leq 0$

In (2.1) consider the quantity

$$\gamma = \varepsilon D + \ell E + (\varepsilon + \ell - \varepsilon\ell) F. \tag{2.2}$$

(i) If D and E are greater than $|F|$, then

$$\begin{aligned}\gamma &= \varepsilon D + \ell E + (\varepsilon + \ell - \varepsilon \ell) F \\ &= \varepsilon D + \varepsilon F + \ell E + \ell F - \varepsilon \ell F \\ &= \varepsilon(D - |F|) + \ell(E - |F|) + \varepsilon\ell |F| \\ &\geq 0.\end{aligned}$$

Thus, the minimum value of γ occurs when $\varepsilon = \ell = 0$, which implies

$$\max_u \text{var } \hat{y}(u_1, u_2) = \text{var } \hat{y}(1, 1).$$

(ii) If $E \leq D \leq |F|$ or $E \leq |F| \leq D$, the minimum value of γ occurs when $\varepsilon = 0$ and $\ell = 2$, which implies

$$\max_u \text{var } \hat{y}(u_1, u_2) = \text{var } \hat{y}(1, -1).$$

(iii) If $D \leq E \leq |F|$ or $D \leq |F| \leq E$, the minimum value of γ occurs when $\varepsilon = 2$ and $\ell = 0$, which implies

$$\max_u \text{var } \hat{y}(u_1, u_2) = \text{var } \hat{y}(-1, 1).$$

Case 3: $D, F \geq 0, \quad E \leq 0$

(i) If D and F are greater than $|E|$, the minimum value of γ in (2.2) occurs when $\varepsilon = \ell = 0$, which implies

$$\max_u \text{var } \hat{y}(u_1, u_2) = \text{var } \hat{y}(1, 1).$$

(ii) If $D \leq F \leq |E|$ or $D \leq |E| \leq F$,

$$\min_{\varepsilon, \ell} \gamma(\varepsilon, \ell) = \gamma(2, 2),$$

which implies

$$\max \text{var } \hat{y}(u_1, u_2) = \text{var } \hat{y}(-1, -1).$$

(iii) If $F \leq D \leq |E|$ or $F \leq |E| \leq D$,

$$\min_{\varepsilon, \ell} \gamma(\varepsilon, \ell) = \gamma(0, 2),$$

which implies

$$\max_{u} \text{var } \hat{y}(u_1, u_2) = \text{var } \hat{y}(1, -1).$$

<u>Case 4</u>: $D \geq 0$; $E, F \leq 0$

$$\min_{\varepsilon, \ell} \gamma(\varepsilon, \ell) = \gamma(0, 2),$$

which implies

$$\max_{u} \text{var } \hat{y}(u_1, u_2) = \text{var } \hat{y}(1, -1) .$$

<u>Case 5</u>: $E, F \geq 0$, $D \leq 0$

(i) If $E \geq |D|$ and $F \geq |D|$,

$$\min_{\varepsilon, \ell} \gamma(\varepsilon, \ell) = \gamma(0, 0),$$

which implies

$$\max_{u} \text{var } \hat{y}(u_1, u_2) = \text{var } \hat{y}(1, 1).$$

(ii) If $E \leq F \leq |D|$ or $E \leq |D| \leq F$,

$$\min_{\varepsilon, \ell} \gamma(\varepsilon, \ell) = \gamma(2, 2),$$

which implies

$$\max_{u} \text{var } \hat{y}(u_1, u_2) = \text{var } \hat{y}(-1, -1) .$$

(iii) If $F \leq E \leq |D|$ or $F \leq |D| \leq E$,

$$\min_{\varepsilon, \ell} \gamma(\varepsilon, \ell) = \gamma(2, 0),$$

which implies

$$\max_u \operatorname{var} \hat{y}(u_1, u_2) = \operatorname{var} \hat{y}(-1, 1).$$

Case 6: $E \geq 0$; $D, F \leq 0$

$$\min_{\varepsilon, \ell} \gamma(\varepsilon, \ell) = \gamma(2, 0),$$

which implies

$$\max_u \operatorname{var} \hat{y}(u_1, u_2) = \operatorname{var} \hat{y}(-1, 1).$$

Case 7: $F \geq 0$; $D, E \leq 0$

$$\min_{\varepsilon, \ell} \gamma(\varepsilon, \ell) = \gamma(2, 2),$$

which implies

$$\max_u \operatorname{var} \hat{y}(u_1, u_2) = \operatorname{var} \hat{y}(-1, -1).$$

Case 8: $D, E, F \leq 0$

(i) If $|D| \leq |E| \leq |F|$ or $|D| \leq |F| \leq |E|$,

$$\min_{\varepsilon, \ell} \gamma(\varepsilon, \ell) = \gamma(0, 2),$$

which implies

$$\max_u \operatorname{var} \hat{y}(u_1, u_2) = \operatorname{var} \hat{y}(1, -1).$$

(ii) If $|E| \leq |D| \leq |F|$ or $|E| \leq |F| \leq |D|$,

$$\min_{\varepsilon, \ell} \gamma(\varepsilon, \ell) = \gamma(2, 0),$$

which implies

$$\max_u \operatorname{var} \hat{y}(u_1, u_2) = \operatorname{var} \hat{y}(-1, 1).$$

(iii) If $|F| \leq |D| \leq |E|$ or $|F| \leq |E| \leq |D|$,

$$\min_{\varepsilon, \ell} \gamma(\varepsilon, \ell) = \gamma(2, 2),$$

which implies

$$\max_u \operatorname{var} \hat{y}(u_1, u_2) = \operatorname{var} \hat{y}(-1, -1).$$

Therefore, $\operatorname{var} \hat{y}(u)$ always attains a maximum at one or more of the corners of R.

Theorem 2.3 For the two-dimensional three-point problem,

$$\min_{x} \max_{u} \operatorname{var} \hat{y}(u) \geq \min \operatorname{tr} (X'X)^{-1} .$$

Proof:

$$\operatorname{var} \hat{y}(1, 1) = \operatorname{var} \hat{\beta}_o + \operatorname{var} \hat{\beta}_1 + \operatorname{var} \hat{\beta}_2 + 2 \operatorname{cov} (\hat{\beta}_o, \hat{\beta}_1) + 2 \operatorname{cov} (\hat{\beta}_o, \hat{\beta}_2) + 2 \operatorname{cov} (\hat{\beta}_1, \hat{\beta}_2) .$$

$$\operatorname{var} \hat{y}(-1, 1) = \operatorname{var} \hat{\beta}_o + \operatorname{var} \hat{\beta}_1 + \operatorname{var} \hat{\beta}_2 - 2 \operatorname{cov} (\hat{\beta}_o, \hat{\beta}_1) + 2 \operatorname{cov} (\hat{\beta}_o, \hat{\beta}_2) - 2 \operatorname{cov} (\hat{\beta}_1, \hat{\beta}_2) .$$

$$\operatorname{var} \hat{y}(1, -1) = \operatorname{var} \hat{\beta}_o + \operatorname{var} \hat{\beta}_1 + \operatorname{var} \hat{\beta}_2 + 2 \operatorname{cov} (\hat{\beta}_o, \hat{\beta}_1) - 2 \operatorname{cov} (\hat{\beta}_o, \hat{\beta}_2) - 2 \operatorname{cov} (\hat{\beta}_1, \hat{\beta}_2) .$$

$$\operatorname{var} \hat{y}(-1, -1) = \operatorname{var} \hat{\beta}_o + \operatorname{var} \hat{\beta}_1 + \operatorname{var} \hat{\beta}_2 - 2 \operatorname{cov} (\hat{\beta}_o, \hat{\beta}_1) - 2 \operatorname{cov} (\hat{\beta}_o, \hat{\beta}_2) + 2 \operatorname{cov} (\hat{\beta}_1, \hat{\beta}_2) .$$

The average over the corners of R is given by

$$\text{ave} \operatorname{var} \hat{y}(u) = \operatorname{var} \hat{\beta}_o + \operatorname{var} \hat{\beta}_1 + \operatorname{var} \hat{\beta}_2 = \operatorname{trace} (X'X)^{-1} .$$

But,

$$\max_{u} \operatorname{var} \hat{y}(u) = \max_{(\text{corners})} \operatorname{var} \hat{y}(u) \quad \text{by theorem 2.2}$$

$$\geq \underset{(\text{corners})}{\text{average}} \operatorname{var} \hat{y}(u) ,$$

which implies

$$\max_{u} \operatorname{var} \hat{y}(u) \geq \operatorname{tr} (X'X)^{-1} .$$

Thus,

$$\min_{x} \max_{u} \operatorname{var} \hat{y}(u) \geq \min_{x} \operatorname{tr} (X'X)^{-1} .$$

This completes the proof.

Although it could not be shown algebraically, it was felt that the design points for the $\min_x \max_u \operatorname{var} \hat{y}(u)$ design should be on the boundary of the square region R. Under the assumption that the design points should fall on the boundary, an empirical investigation using the IBM 650 computer produced a design which is believed to be the min max var $\hat{y}(u)$ design; namely, the design

$$d^* = [(-1, -1), (-0.364656, 1), (1, -0.364656)], \tag{2.3}$$

which has maximum variance equal to 1.420048. The following procedure was used to determine this design.

Initially, all designs of the form

$$d = [(-1, b_1), (a_2, b_2), (a_3, b_3)],$$

where $-1 \leq b_1 \leq 0$, $-1 \leq a_2, a_3, b_2, b_3 \leq 1$, and a_i, b_i take on multiples of 0.5, were investigated. It was determined that the possible candidates for the $\min_x \max_u \operatorname{var} \hat{y}(u)$ design when all three points were taken on the boundary, were of the form

$$d = [(-1, -1), (a, 1), (1, b)],$$

where $-0.4 \leq a \leq -0.3$ and $-0.4 \leq b \leq -0.3$. Further investigation of these designs when <u>a</u> and <u>b</u> took on multiples of 0.000001 yielded the design d^* in (2.3). If it could be shown that the three design points of a $\min \max_x \operatorname{var}_u \hat{y}(u)$ design must fall on the boundary of R, then d^* would be the $\min_x \max_u \operatorname{var} \hat{y}(u)$ design. This could not be shown however.

It was thought that the vertices of the largest equilateral triangle inscribed in the square region R had possibilities of being the $\min_x \max_u \text{var } \hat{y}(u)$ design; that is, the design

$$d_1 = [(-1, -1), (3 - 2\sqrt{3}, 1), (1, 3 - 2\sqrt{3})].$$

This design was of interest because it is a rotatable design. It was found however, that d_1 yielded a maximum variance of 1.57774, which is somewhat larger than the maximum variance obtained by using the design d^* in (2.3). Thus, the design d_1 was rejected.

Although a detailed investigation was not conducted, all designs of the form

$$d = [(-1, b_1), (a_2, b_2), (a_3, b_3)],$$

where $-1 \leq b_1 \leq 0$, $-1 \leq a_2, a_3, b_2, b_3 \leq 1$ and a_i, b_i take on multiples of 0.1, were considered to try to determine the design which minimizes the average variance of $\hat{y}(u)$ and the design which minimizes the generalized variance of $\hat{y}(u)$. The design which minimized the ave var $\hat{y}(u)$ was given by

$$d_2 = [(-1, -1), (-.5, 1), (1, -.4)],$$

with average variance of $\hat{y}(u)$ equal to 2.66569 and $|X'X| = 13.69$. The design which minimized the generalized variance of $\hat{y}(u)$ was given by

$$d_3 = [(-1, -1), (-1, 1), (1, y)],$$

where $-1 \leq y \leq 1$, with $|X'X| = 16$ and average variance of $\hat{y}(u)$ equal to 2.66666. This investigation gave rise to the following theorem.

<u>Theorem 2.4</u> Consider the design

$$d = [(-1, -1), (-1, a), (1, y)]$$

in the three-point problem, where $\underline{a}$ is a fixed constant and y is a variable such that $-1 \le y \le 1$, then $|X'X| = K$ for every value of y, where K is a constant.

Proof:

$$|X'X| = \begin{vmatrix} 3 & -1 & a+y-1 \\ -1 & 3 & -a+y+1 \\ a+y-1 & -a+y+1 & a^2+y^2+1 \end{vmatrix}$$

$$= 3[3(a^2+y^2+1)-(-a+y+1)^2]+[-(a^2+y^2+1)-(-a+y+1)(a+y-1)]$$
$$+(a+y-1)[-(-a+y+1)-3(a+y-1)]$$

$$= 8(a^2+y^2+1)-3(-a+y+1)^2-2(-a+y+1)(a+y-1)-3(a+y-1)$$

$$= 8a^2+8y^2+8-3a^2-3y^2-3+6ay+6a-6y+2a^2-4a-2y^2+2-3a^2$$
$$-3y^2-3-6ay+6a+6y$$

$$= (8-3+2-3)a^2+(8-3-2-3)y^2+(6-6)ay+(6-4+6)a$$
$$+(-6+6)y+(8-3+2-3)$$

$$= 4a^2+8a+4$$

$$= 4(a+1)^2.$$

Thus, since $|X'X|$ is independent of y, the theorem is proved.

This investigation would certainly be strengthened if it could be established that the design points for the $\min_x \max_u \text{var}\ \hat{y}(u)$ design, the $\min_x \text{ave var}\ \hat{y}(u)$ design, and the $\min_x \text{gen var}\ \hat{y}(u)$ design must fall on the boundary of the square region R. All efforts in this direction produced no results however.

CHAPTER III

SOME EXACT OPTIMAL DESIGNS

Optimal designs in the sense of several criteria have been determined by Kiefer (4) for the general polynomial model

$$y = \beta_o + \sum_{i=1}^{p} \beta_i x^i + \varepsilon .$$

However, since he developed optimal designs from a probability measure standpoint, many experiments with sample size N only have optimal designs which are "within $O(N^{-1})$" of being optimal. For example, if the model

$$y = \beta_o + \beta_1 x + \varepsilon \qquad\qquad -1 \leq x \leq 1$$

is assumed, then according to Kiefer, the unique optimal design in the sense of minimizing the maximum variance of $\hat{y}$ (u) assigns probability mass equal to 1/2 to the points $x = -1$ and $x = 1$. Thus, if the sample size N is even, the optimal design assigns N/2 points to $x = -1$ and N/2 points to $x = 1$. However, if N is odd, where should we put the odd observation? This question is not answered by the unique optimal design offered by Kiefer (4). It is answered however, by Folks (3) for the above model. That is, for N odd, put

$\frac{(N - 1)}{2}$ points at $x = -1$

1 point at $x = 0$

$$\frac{(N-1)}{2} \quad \text{points at } x = 1.$$

With this example in mind, it appears reasonable to examine certain polynomial models to try to determine exact optimal designs for all values of N.

The criteria of optimality that will be considered in this chapter are as follows:

(1) $\min_x \max_u \operatorname{var} \hat{y}(u)$

(2) $\min_x \max_u |\text{bias}(u)|$

(3) $\min_x \text{ ave } |\text{bias}(u)|$

(4) $\min_x \text{ ave } \text{bias}^2(u)$.

Exact optimal designs using criterion (1) have been determined by Folks (3) for the model

$$y = \beta_o + \beta_1 x + \varepsilon \qquad -1 \leq x \leq 1. \tag{3.1}$$

In this chapter, we shall determine exact optimal designs using criteria, (2), (3), and (4) for the model in (3.1). Also, exact optimal designs will be determined using all of the criteria for several other models.

3.1 Assumed Model: $y = b_o + b_1 x + \varepsilon$, $-1 \leq x \leq 1$

Consider the bias that will arise if the true model is

$$y = d_o + d_1 x + d_2 x^2 + \varepsilon \qquad -1 \leq x \leq 1.$$

The bias function is given by

$$|\text{bias}(u)| = \left| u^2 - \frac{[N \Sigma x^3 - \Sigma x \, \Sigma x^2]}{[N \Sigma (x - \bar{x})^2]} u \right| \left| d_2 \right| .$$

Since bias functions will be needed throughout this chapter, consider the following general derivation.

Assume the relationship

$$Y_A = X_1\gamma_1 + X_2\gamma_2 + \varepsilon$$

$$= Z\xi + \varepsilon ,$$

when the true relationship is

$$Y_T = X_1 \gamma_1 + X_3 \gamma_3 + \varepsilon,$$

where $X_1 \gamma_1$ represents the k terms common to both Y_A and Y_T, out of the p possible terms in Y_A and the q possible terms in Y_T. Then,

$$\hat{\xi} = (Z'Z)^{-1} Z'Y_T \qquad W = [U_1 U_2]$$

$$\hat{Y}_A(w) = W\hat{\xi}$$

$$E[\hat{Y}_A(w)] = W(Z'Z)^{-1} Z'E(Y_T)$$

$$= W(Z'Z)^{-1} Z'(X_1 \gamma_1 + X_3\gamma_3)$$

$$E[Y_T(w)] = U_1\gamma_1 + U_3\gamma_3$$

$$|\text{bias } (u)| = |E[\hat{Y}_A(w)] - E[Y_T(w)]|$$

$$= |W(Z'Z)^{-1} Z'(X_1\gamma_1 + X_3\gamma_3) - U_1\gamma_1 - U_3\gamma_3|$$

$$= |[W(Z'Z)^{-1} Z'X_1 - U_1]\gamma_1 + [W(Z'Z)^{-1}Z'X_3 - U_3]\gamma_3| .$$

If Y_A and Y_T have no common terms,

$$X_1 = \phi$$

$$\gamma_1 = \phi$$

$$W = U_2$$

$$Z = X_2 ,$$

which implies that

$$|\text{bias } (u)| = |U_2(X_2'X_2)^{-1}X_2'X_3 - U_3||\gamma_3| .$$

If Y_T contains all of the terms that are in Y_A,

$$X_2 = \phi$$

$$\gamma_2 = \phi$$

$$W = U_1$$

$$Z = X_1,$$

which implies that

$$|\text{bias } (u)| = |[U_1(X_1'X_1)^{-1}X_1'X_1 - U_1]\gamma_1 + [U_1(X_1'X_1)^{-1}X_1'X_3 - U_3]\gamma_3|$$

$$= |U_1(X_1'X_1)^{-1}X_1'X_3 - U_3||\gamma_3| .$$

3.1.1 $\min_x \max_u |\text{bias } (u)|$ design

It has been established by Folks (3) that the $\min_x \max_u |\text{bias } (u)|$ design is one which has $\Sigma x = 0$ and $\Sigma x^2 = N/2$. When $N = 4K$ one such design has

$N/4$ points at $x = -1$

$N/2$ points at $x = 0$

$N/4$ points at $x = 1$.

Certainly the $\min_x \max_u |\text{bias } (u)|$ design is not unique. Thus, consider a design that minimizes the maximum bias when $N = 4K + 1$, $4K + 2$, and $4K + 3$.

Case 1: For $N = 4K + 1$ ($K = 1, 2, \ldots$), choose

K points at $x = -1.0$

1 point at $x = -0.5$

$(2K - 1)$ points at $x = 0.0$

1 point at $x = 0.5$

K points at $x = 1.0$.

Case 2: $N = 4K + 2$ $(K = 1, 2, \ldots)$, choose

K points at $x = -1.0$

2 points at $x = -0.5$

$(2K - 2)$ points at $x = 0.0$

2 points at $x = 0.5$

K points at $x = 1.0$.

Case 3: $N = 4K + 3$ $(K = 0, 1, 2, \ldots)$, choose

K points at $x = -1.0$

1 point at $x = -0.866$

$(2K + 1)$ points at $x = 0.0$

1 point at $x = 0.866$

K points at $x = 1.0$.

3.1.2 $\min_x$ ave $|\text{bias}(u)|$ design

It has been determined by Folks (3) that the $\min_x$ ave $|\text{bias}(u)|$ design is one which has $\Sigma x = 0$ and $\Sigma x^2 = N/4$. For $N = 8K$ one such design has

$N/8$ points at $x = -1.0$

$3N/4$ points at $x = 0.0$

$N/8$ points at $x = 1.0$.

However, exact optimal designs have not been established for N not a multiple of 8. Consider then the following exact optimal designs for all values of N.

Case 1:

(i) N = 4K (K odd), choose

$\frac{(K - 1)}{2}$ points at x = -1.0

1 point at x = -0.707

(3K - 1) points at x = 0.0

1 point at x = 0.707

$\frac{(K - 1)}{2}$ points at x = 1.0.

(ii) N = 4K (K even), choose

$\frac{K}{2}$ points at x = -1.0

3 K points at x = 0.0

$\frac{K}{2}$ points at x = 1.0.

Case 2: N = 4K + 1 (K = 1, 2, . . .), choose

(K - 1) points at x = -0.707

2 points at x = 0.559

(2K - 1) points at x = 0.0

2 points at x = 0.559

(K - 1) points at x = 0.707.

Case 3: N = 4K + 2 (K = 1, 2, . . .), choose

K points at x = -0.707

1 point at x = -0.5

2K points at x = 0.0

1 point at $x = 0.5$

K points at $x = 1.0$.

Case 4: $N = 4K + 3$ $(K = 0, 1, 2, \ldots)$, choose

1 point at $x = -0.935$

$(K - 1)$ points $x = -0.707$

$(2K + 3)$ points at $x = 0.0$

$(K - 1)$ points at $x = 0.707$

1 point at $x = 0.935$.

3.1.3 $\min_x$ ave bias$^2(u)$ design

The $\min_x$ ave bias$^2(u)$ is achieved by the design such that $\Sigma x = 0$ and $\Sigma x^2 = N/3$. The following are optimal designs for all values of N.

Case 1: $N = 4K$ $(K = 1, 2, \ldots)$, choose

K points at $x = -0.816$

2 K points at $x = 0.0$

K points at $x = 0.816$.

Case 2: $N = 4K + 1$ $(K = 1, 2, \ldots)$, choose

1 point at $x = -0.913$

$(K - 1)$ points at $x = -0.816$

$(2K + 1)$ points at $x = 0.0$

$(K - 1)$ points at $x = 0.816$

1 point at $x = 0.913$.

Case 3: $N = 4K + 2$ $(K = 1, 2, \ldots)$, choose

1 point at $x = -1.0$

(K - 1) points at $x = -0.816$

(2K + 2) points at $x = 0.0$

(K - 1) points at $x = 0.816$

1 point at $x = 1.0$.

Case 4: $N = 4K + 3$ $(K = 0, 1, 2, \ldots)$, choose

K points at $x = -0.816$

1 point at $x = -0.707$

(2K + 1) points at $x = 0.0$

1 point at $x = 0.707$

K points at $x = 0.816$.

3.2 Assumed Model: $y = c_o + c_2x^2 + \varepsilon$, $-1 \leq x \leq 1$

Consider the variance function for this model.

$$|X'X| = \begin{bmatrix} N & \Sigma x^2 \\ \Sigma x^2 & \Sigma x^4 \end{bmatrix}$$

$$(X'X)^{-1} = \frac{1}{[N\,\Sigma x^4 - (\Sigma x^2)^2]} \begin{bmatrix} \Sigma x^4 & -\Sigma x^2 \\ -\Sigma x^2 & N \end{bmatrix} .$$

$$\operatorname{var} \hat{y}(u) = U(X'X)^{-1}U'$$

$$= \frac{[Nu^4 - 2(\Sigma x^2)\,u^2 + \Sigma x^4]}{[N\,\Sigma x^4 - (\Sigma x^2)^2]} .$$

Restrict the design points to $x = -1$, $x = 0$, and $x = 1$, then

$$\operatorname{var} \hat{y}(u) = \frac{[N\,u^4 - 2(\Sigma x^2)\,u^2 + \Sigma x^2]}{[N\,\Sigma x^2 - (\Sigma x^2)^2]} .$$

3.2.1 $\min_x \max_u \operatorname{var} \hat{y}(u)$ design

Let

$$\Sigma x^2 = N - K \qquad (K = 1, 2, \ldots, N-1).$$

(i) Consider $N \geq 2K$, where $K = 1, 2, \ldots, [N/2]$.

(By $[N/2]$ we will mean the largest integer smaller than $N/2$). Thus, $N \geq 2K$ implies that

$$N - 2K \geq 0$$

$$K(N - 2K) \geq 0$$

$$2K(N - K) \geq NK$$

$$2/K \geq N/K(N - K) .$$

This is true for $K = 1, 2, \ldots, [N/2]$, which implies $\Sigma x^2 = N - [N/2], \ldots, N - 1$. Now

$$\operatorname{var} \hat{y}(u) = \frac{[Nu^4 - 2(\Sigma x^2)u^2 + \Sigma x^2]}{[\Sigma x^2 (N - \Sigma x^2)]}$$

$$= \frac{[Nu^4 - 2u^2(N-K) + (N-K)]}{[(N-K)(N - N + K)]}$$

(3.2)
$$= \frac{Nu^4}{K(N-K)} - \frac{2u^2}{K} + \frac{1}{K} .$$

But

$$u^2 \geq u^4 \qquad \text{for } -1 \leq u \leq 1$$

and

$$2/K \geq N/K(N-K) ,$$

which implies that

$$2u^2/K \geq Nu^4/K(N-K) ,$$

which implies

$$\max_u \ \text{var}\ \hat{y}(u) = \text{var}\ \hat{y}(0) = 1/K = 1/(N - \Sigma x^2) \ .$$

Thus, the $\min_x \max_u \text{var}\ \hat{y}(u)$ is achieved by the design for which Σx^2 is as small as possible under the condition $\Sigma x^2 = N - [N/2], \ldots, N - 1$. That is,

$$\min_x \max_u \text{var}\ \hat{y}(u) = 1/[N - (N - [N/2])]$$
$$= 1/[N/2] \ .$$

(ii) Consider $N < 2K$ ($K = [N/2] + 1, \ldots, N-1$), then $N < 2K$ implies

$$N - 2K < 0$$

$$2K(N - K) < NK$$

(3.3) $$2/K < N/K(N - K).$$

This is true for $K = [N/2] + 1, \ldots, N - 1$, which implies $\Sigma x^2 = 1, 2, \ldots, N - [N/2] - 1$. From (3.2) we see that

$$\text{var}\ \hat{y}(u) = \frac{Nu^2}{K(N-K)} - \frac{2u^2}{K} + \frac{1}{K} \ .$$

Since

$$\frac{N}{K(N-K)} > \frac{2}{K} \qquad \text{from (3.3)}$$

and

$$u^4 \leq u^2 \qquad \text{for } -1 \leq u \leq 1,$$

we have that the $\max_x$ var $\hat{y}(u)$ is achieved when $u = \pm 1$,

Thus,

$$\max_u \ \text{var}\ \hat{y}(u) = \text{var}\ \hat{y}(\pm 1)$$

$$= \frac{N}{K(N-K)} - \frac{2}{K} + \frac{1}{K}$$

$$= \frac{1}{(N - K)}$$

$$\max_u \operatorname{var} \hat{y}(u) = \frac{1}{\Sigma x^2} .$$

Hence, to achieve $\min_x \max_u \operatorname{var} \hat{y}(u)$, choose Σx^2 as large as possible under the condition $\Sigma x^2 = 1, 2, \ldots, N - [N/2] - 1$. That is,

$$\min_x \max_u \operatorname{var} \hat{y}(u) = \frac{1}{(N - [N/2] - 1)} .$$

Consider a $\min_x \max_u \operatorname{var} \hat{y}(u)$ design when N is odd. A $\min_x \max_u \operatorname{var} \hat{y}(u)$ design is one which has $\Sigma x^2 = (N-1)/2$ or $\Sigma x^2 = (N+1)/2$, for which

$$\max_u \operatorname{var} \hat{y}(u) = \frac{2}{(N-1)} .$$

One $\min_x \max_u \operatorname{var} \hat{y}(u)$ design has

$\frac{(N-1)}{4}$ points at $x = -1.0$

$\frac{(N+1)}{2}$ points at $x = 0.0$

$\frac{(N-1)}{4}$ points at $x = 1.0$.

Consider a $\min_x \max_u \operatorname{var} \hat{y}(u)$ design when N is even. A $\min_x \max_u \operatorname{var} \hat{y}(u)$ design is one which has $\Sigma x^2 = N/2$, for which

$$\max_u \operatorname{var} \hat{y}(u) = \frac{2}{N} .$$

One $\min_x \max_u \operatorname{var} \hat{y}(u)$ design has

$N/4$ points at $x = -1.0$

$N/2$ points at $x = 0.0$

$N/4$ points at $x = 1.0$.

3.2.2 $\min_x \max_u |\text{bias}(u)|$ design

Consider the bias that will arise if the true model is

$$y = b_o + b_1 x + \varepsilon .$$

$$|\text{bias } (u)| = \left| \frac{[N \Sigma x^3 - \Sigma x \Sigma x^2]}{[N \Sigma x^4 - (\Sigma x^2)^2]} u^2 - u + \frac{[\Sigma x \Sigma x^4 - \Sigma x^2 \Sigma x^3]}{[N \Sigma x^4 - (\Sigma x^2)^2]} \right| |b_1| .$$

Restrict the class of designs to those which have design points only at $x = -1$, $x = 0$, and $x = 1$, then

$$|\text{bias } (u)| = \left| \frac{\Sigma x(N - \Sigma x^2) u^2}{\Sigma x^2 (N - \Sigma x^2)} - u + 0 \right| |b_1|$$

$$= |u^2 \Sigma x/\Sigma x^2 - u| |b_1| .$$

To show that the $\min_x \max_u |\text{bias } (u)|$ design has

$$\max_u |\text{bias } (u)| = |b_1| ,$$

consider the following cases.

Case 1: $\Sigma x = 0$

$$|\text{bias } (u)| = |-u| |b_1| .$$

$$\max_u |\text{bias } (u)| = |b_1| .$$

Case 2: $\Sigma x < 0$

The maximum bias occurs at $u = 1$ or $u = \Sigma x^2/2 \Sigma x$.

$$|\text{bias } (1)| = |\Sigma x/\Sigma x^2 - 1| |b_1|$$

(3.4) $$= (1 - \Sigma x/\Sigma x^2) |b_1| .$$

$$|\text{bias } (\Sigma x^2/2\Sigma x)| = |-\Sigma x^2/4\Sigma x| |b_1|$$

(3.5) $$= -(\Sigma x^2/4\Sigma x) |b_1| .$$

Let $A = \Sigma x/\Sigma x^2$, then from (3.4) and (3.5)

$$1 - A = -1/4A .$$

$$4 A^2 - 4 A - 1 = 0.$$

$$A = 1/2 (1 \pm \sqrt{2}).$$

Since A must be negative, $(\Sigma x < 0)$, $A = 1/2(1 - \sqrt{2})$.

Thus, for $A = 1/2(1 - \sqrt{2})$, we have that

$$|\text{bias}(1)| = |\text{bias}(\Sigma x^2/2\Sigma x)| .$$

(i) For $A > 1/2(1-\sqrt{2})$

$$\max_u |\text{bias}(u)| = |\text{bias}(\Sigma x^2/2\Sigma x)|$$

$$= -(\Sigma x^2/4\Sigma x)\, |b_1|$$

$$> |b_1| .$$

(ii) For $A < 1/2(1 - \sqrt{2})$

$$\max_u |\text{bias}(u)| = |\text{bias}(1)|$$

$$> |b_1| .$$

Case 3: $\Sigma x > 0$

The maximum bias occurs at $u = -1$ or $u = \Sigma x^2/2\Sigma x$.

In either case,

$$\max_u |\text{bias}(u)| > |b_1| .$$

Therefore, the design such that $\Sigma x = 0$ achieves $\min_x \max_u |\text{bias}(u)|$, which has

$$\max_u |\text{bias}(u)| = |b_1| .$$

Consider a $\min_x \max_u |\text{bias}(u)|$ design when N is odd. One $\min_x \max_u |\text{bias}(u)|$ design has

$(N-1)/4$ points at $x = -1.0$

$(N+1)/2$ points at $x = 0.0$

$(N - 1)/4$ points at $x = 1.0$.

Consider a $\min_x \max_u |\text{bias}(u)|$ design when N is even. One $\min_x \max_u |\text{bias}(u)|$ design has

N/4 points at $x = -1.0$

N/2 points at $x = 0.0$

N/4 points at $x = 1.0$

3.2.3 $\min_x$ ave bias^2 (u) design

As in section 3.2.2, consider the bias that will arise if the true model is

$$y = b_o + b_1 x + \varepsilon .$$

When we consider the designs which have design points only at $x = 1$, $x = 0$, and $x = 1$, we have

$$\text{ave bias}^2(u) = \int_{-1}^{1} [\Sigma x/\Sigma x^2)u^2 - u]^2 \, b_1^2 du$$

$$= b_1^2 [\, 2/5(\Sigma x/\Sigma x^2)^2 + 2/3 \,].$$

Thus,

$$\min_x \text{ave bias}^2 (u) = 2b_1^2/3,$$

since $(\Sigma x/\Sigma x^2)^2$ is minimum when $\Sigma x = 0$. Designs for N both even and odd are given in section 3.2.2.

3.3 Assumed Model: $y = g_1 x + g_2 x^2 + \varepsilon$, $-1 \leq x \leq 1$

3.3.1 $\min_x \max_u$ var $\hat{y}$ (u) design

The variance of $\hat{y}(u)$ is given by

$$\text{var } \hat{y}\ (u) = \frac{[(\Sigma x^2)u^4 - (2\Sigma x^3)u^3 + (\Sigma x^4)u^2]}{[\Sigma x^2 \, \Sigma x^4 - (\Sigma x^3)^2 \,]} .$$

Consider the maximum variance for the different values of Σx^3.

Case 1: $\Sigma x^3 < 0$

$$\max_u \operatorname{var} \hat{y}(u) = \operatorname{var} \hat{y}(1)$$

$$= \frac{[\Sigma x^4 - 2\Sigma x^3 + \Sigma x^2]}{[\Sigma x^2 \Sigma x^4 - (\Sigma x^3)^2]} .$$

<u>Case 2:</u> $\Sigma x^3 = 0$

$$\max_u \operatorname{var} \hat{y}(u) = \operatorname{var} \hat{y}(\pm 1)$$

$$= \frac{1}{\Sigma x^4} + \frac{1}{\Sigma x^2} .$$

<u>Case 3:</u> $\Sigma x^3 > 0$

$$\max_u \operatorname{var} \hat{y}(u) = \operatorname{var} \hat{y}(-1)$$

$$= \frac{[\Sigma x^4 + 2\Sigma x^3 + \Sigma x^2]}{[\Sigma x^2 \Sigma x^4 - (\Sigma x^3)^2]} .$$

Consider a $\min_x \max_u \operatorname{var} \hat{y}(u)$ design when N is odd. As would be expected, the $\min_x \max_u \operatorname{var} \hat{y}(u)$ design is not unique.

(i) If we take $\Sigma x^3 = 0$, then $\min \max \operatorname{var} \hat{y}(u)$ is achieved by the design such that $\Sigma x^2 = N - 1$ and $\Sigma x^4 = N - 1$, for which

$$\min_u \operatorname{var} \hat{y}(u) = \frac{2}{(N-1)} .$$

One $\min_x \max_u \operatorname{var} \hat{y}(u)$ design has

$\frac{(N-1)}{2}$ points at $x = -1.0$

1 point at $x = 0.0$

$\frac{(N-1)}{2}$ points at $x = 1.0$.

(ii) If we take $\Sigma x^3 = \pm 1$ and $\Sigma x^2 = \Sigma x^4 = N$,

$$\max_u \operatorname{var} \hat{y}(u) = \frac{2}{(N-1)} .$$

One $\min_x \max_u \text{var } \hat{y}(u)$ design has

$$\frac{(N-1)}{2} \text{ points at } x = -1.0$$

$$\frac{(N+1)}{2} \text{ points at } x = 1.0.$$

Consider a $\min_x \max_u \text{var } \hat{y}(u)$ design when N is even. Choose $\Sigma x^3 = 0$ and Σx^2 as large as possible; that is, $\Sigma x^2 = \Sigma x^4 = N$. The design which minimizes the maximum variance of $\hat{y}(u)$ has

$$\max_u \text{var } \hat{y}(u) = \frac{2}{N}.$$

One such design has

$$N/2 \text{ points at } x = -1.0$$

$$N/2 \text{ points at } x = 1.0.$$

3.3.2 $\min_x \max_u |\text{bias }(u)|$ design

The bias that arises when the true model is

$$y = b_o + b_1 x + \varepsilon \qquad -1 \le x \le 1$$

is given by

$$|\text{bias }(u)| = \left| \frac{[(\Sigma x^2)^2 - \Sigma x \Sigma x^3]\, u^2}{[\Sigma x^2\, \Sigma x^4 - (\Sigma x^3)^2]} + \frac{[\Sigma x \Sigma x^4 - \Sigma x \Sigma x^3]}{[\Sigma x^2 \Sigma x^4 - (\Sigma x^3)^2]}\, u - 1 \right| |b_o|$$

Consider only the class of designs which have design points $x = -1$, $x = 0$, and $x = 1$, then

$$|\text{bias }(u)| = |u^2 - 1|\,|b_o|.$$

$$\max_u |\text{bias }(u)| = |\text{bias }(0)|$$

$$= |b_o|.$$

Since $|\text{bias }(u)|$ is independent of the design points, any design in the restricted class of designs achieves $\min_x \max_u |\text{bias}(u)|$. One such design has

$$\begin{aligned} &1 \text{ point at } x = -1.0 \\ &N-2 \text{ points at } x = 0.0 \\ &1 \text{ point at } x = 1.0. \end{aligned}$$

The same design achieves $\min_x \text{ ave } |\text{bias}(u)|$ and $\min_x \text{ ave bias}^2(u)$. For this design

$$\min_x \text{ ave } |\text{bias}(u)| = 4/3 \; |b_o|$$

$$\min_x \text{ ave bias}^2(u) = 16/15 \; b_o^2.$$

Several other exact optimal designs were derived; however, due to the similar results that were obtained, it was felt that the exact optimal designs that have been presented would be sufficient to introduce the problem of finding exact optimal designs.

CHAPTER IV

CHOOSING A MODEL

In order to determine optimal designs, we must first assume a model which we think will best represent the data. This is not an easy task, however. Of course there is always a risk involved in the selection of a model, but this risk can usually be reduced by selecting from a particular class a model which is optimal in the sense of some optimality criterion. The seriousness of fitting a model which is not the true one is certainly dependent upon what we use for a risk function. In this chapter we shall use some function of the bias as the risk function and shall determine which model should be assumed in order to minimize this risk.

Consider the particular class of models C consisting of

(1) $y = h_1(x) + \varepsilon = a_o + \varepsilon$

(2) $y = h_2(x) + \varepsilon = b_o + b_1x + \varepsilon$

(3) $y = h_3(x) + \varepsilon = c_o + c_2x^2 + \varepsilon$

(4) $y = h_4(x) + \varepsilon = d_o + d_1x + d_2x^2 + \varepsilon$

(5) $y = h_5(x) + \varepsilon = e_1x + \varepsilon$

(6) $y = h_6(x) + \varepsilon = f_2x^2 + \varepsilon$

(7) $y = h_7(x) + \varepsilon = g_1x + g_2x^2 + \varepsilon$,

where $-1 \leq x \leq 1$.

Let us fit the model $h_i(x) + \varepsilon$ in C (i arbitrary). A bias will arise if the true model $h_j(x) + \varepsilon$ in $C(j \neq i)$ contains terms not in $h_i(x) + \varepsilon$. Choose the design that will $\min_x \max_u |\text{bias}(u)|$, then for this design, let $M(i, j)$ be the maximum bias which arises from fitting $h_i(x) + \varepsilon$ in C when $h_j(x) + \varepsilon$ in C is the true model. The problem will be to select the model in C that will

$$\min_{(\text{assumed model})} \quad \max_{(\text{true model})} \quad M(i, j) .$$

Since no bias will arise when we assume the model

$$y = h_4(x) + \varepsilon = d_o + d_1 x + d_2 x^2 + \varepsilon$$

regardless of which model in C is the true model, we shall disregard model $h_4(x) + \varepsilon$ as an assumed model.

In order to select the model in C that will achieve

$$\min_x \max_u M(i, j),$$

we shall first determine $M(i, j)$ for every value of i and j except $i = j$ and $i = 4$.

4.1 Assumed Model: $y = a_o + \varepsilon$, $-1 \leq x \leq 1$

4.1.1 True model: $y = b_o + b_1 x + \varepsilon$

Consider the bias that arises from neglecting the linear term.

$$|\text{bias}(u)| = |u - \Sigma x/N| \, |b_1| .$$

Case 1: $\Sigma x < 0$

$$\begin{aligned} \max_u |\text{bias}(u)| &= |\text{bias}(1)| \\ &= |1 - \Sigma x/N| \, |b_1| \\ &> |b_1| . \end{aligned}$$

Case 2: $\Sigma x = 0$

$$\max_u |\text{bias}(u)| = |\text{bias}(\pm 1)|$$

$$= |b_1| .$$

Case 3: $\Sigma x > 0$

$$\max_u |\text{bias}(u)| = |\text{bias}(-1)|$$

$$= |-1 - \Sigma x/N| \, |b_1|$$

$$> |b_1| .$$

Therefore, $\min_x \max_u |\text{bias}(u)|$ is achieved by the design such that $\Sigma x = 0$, and

$$M(1, 2) = |b_1| .$$

4.1.2 True Model: $y = c_o + c_2 x^2 + \varepsilon$

Case 1: $\Sigma x^2/N < 1/2$

$$\max_u |\text{bias}(u)| = |\text{bias}(\pm 1)|$$

$$= |1 - \Sigma x^2/N| \, |c_2|$$

$$= |c_2| / 2 .$$

Case 2: $\Sigma x^2/N = 1/2$

$$\max_u |\text{bias}(u)| = |\text{bias}(0)|$$

$$= |\text{bias}(\pm 1)|$$

$$= |c_2| / 2 .$$

Case 3: $\Sigma x^2/N > 1/2$

$$\max_u |\text{bias}(u)| = |\text{bias}(0)|$$

$$= |-\Sigma x^2/N| \, |c_2|$$

$$\max_u |\text{bias } (u)| = |c_2|/2 .$$

Thus, $\min_x \max_u |\text{bias } (u)|$ is achieved by the design such that $\Sigma x^2 = N/2$, and

$$M(1, 3) = |c_2| / 2.$$

4.1.3 True Model: $y = d_o + d_1 x + d_2 x^2 + \varepsilon$

Assume $d_1 = d^*$ and $d_2 = Kd^*$ $(K > 0)$.

$$|\text{bias } (u)| = |(\Sigma x + K\Sigma x^2)/N - u - Ku^2| \; |d^*|$$
$$= |Ku^2 + u - (\Sigma x + K\Sigma x^2)/N| \, |d^*|.$$

Case 1: $(\Sigma x + K\Sigma x^2) < 0$

$$\max_u |\text{bias } (u)| = |\text{bias } (1)|$$
$$= |1 + K - (\Sigma x + K\Sigma x^2)/N| \, |d^*|$$
$$> (1 + K) \; |d^*|.$$

Case 2: $(\Sigma x + K\Sigma x^2) = 0$

$$\max_u |\text{bias } (u)| = |\text{bias } (1)|$$
$$= (1 + K) \, |d^*| .$$

Case 3: $(\Sigma x + K \Sigma x^2) > 0$

Let $$f(u) = Ku^2 + u - (\Sigma x + K \Sigma x^2)/N,$$

then the minimum of $f(u)$ occurs at $u = -1/2K$. In what follows, let

$$P = (\Sigma x + K \Sigma x^2)/N ,$$

(i) $P \geq K$

For this condition, we have two possibilities

$$u = -1: \; |\text{bias } (-1)| = |K - 1 - P| \, |d^*|$$

$$|\text{bias } (-1)| = (P + 1 - K)|d^*|.$$

$$u = -1/2K: \quad |\text{bias } (-1/2K)| = (P + 1/4K)|d^*|.$$

To determine which bias is the largest, assume

$$P + 1 - K > P + 1/4K, \text{ then}$$

$$1 - K > 1/4K$$

$$4K - 4K^2 > 1$$

$$4K^2 - 4K + 1 < 0$$

$$(2K - 1)^2 < 0.$$

Thus, we have a contradiction, which implies

$$P + 1/4K > P + 1 - K \qquad \text{for every P and K.}$$

Now, to minimize $P + 1/4K$ with respect to X, choose $P = K$. That is, under the assumption that $P \geq K$, $\min_x \max_u |\text{bias } (u)|$ is achieved by making P as small as possible; namely, $P = K$. Then for this design

$$\max_u |\text{bias } (u)| = |\text{bias } (-1/2K)|$$

$$= \left(\frac{4K^2 + 1}{4K}\right) |d^*|.$$

(ii) $P < K$

Similarly, for this condition we have two possibilities

$$u = 1: \quad |\text{bias } (1)| = |K + 1 - P|\,|d^*|$$

(4.1) $$= (K + 1 - P)|d^*|.$$

$$u = -1/2K: |\text{bias } (-1/2K)| = |(-1/4K) - P|\,|d^*|$$

(4.2) $$= (P + 1/4K)\,|d^*|.$$

Equating the coefficients of d^* in (4.1) and (4.2)

$$P + 1/4K = K + 1 - P$$

$$2P = K + 1 - 1/4K$$

$$P = (4K^2 + 4K - 1)/8K,$$

we have

$$|\text{bias}(1)| = |\text{bias}(-1/2K)|$$
$$= [(4K^2 + 4K + 1)/8K] \, |d^*|.$$

For $0 < P < (4K^2 + 4K - 1)/8K$

$$K + 1 - P > K + 1 - (4K^2 + 4K - 1)/8K$$
$$= (4K^2 + 4K + 1)/8K$$
$$= (4K^2 + 4K - 1)/8K + 1/4K$$
$$> P + 1/4K,$$

which implies

$$|\text{bias}(1)| > |\text{bias}(-1/2K)| \; .$$

For $P > (4K^2 + 4K - 1)/8K$

$$P + 1/4K > (4K^2 + 4K - 1)/8K + 1/4K$$
$$= (4K^2 + 4K + 1)/8K$$
$$= (K + 1) - (4K^2 + 4K - 1)/8K$$
$$> K + 1 - P,$$

which implies

$$|\text{bias}(-1/2K)| > |\text{bias}(1)|.$$

Hence, for $0 < P < (4K^2 + 4K - 1)/8K$

$$\max_u |\text{bias}(u)| = |\text{bias}(1)|$$
$$= (K + 1 - P) \, |d^*|,$$

and for $P > (4K^2 + 4K - 1)/8K$

$$\max_u |\text{bias } (u)| = |\text{bias } (-1/2K)|$$
$$= (P + 1/4K)\ |d^*| \ .$$

Thus, for $P < K$, the $\min_x \max_u |\text{bias } (u)|$ is attained by the design for which

$$(\Sigma x + K\,\Sigma x^2)/N = (4K^2 + 4K - 1)/8K,$$

and for this design

$$\max_u |\text{bias } (u)| = [(4K^2 + 4K + 1)/8K]\ |d^*| \ .$$

The next task will be to compare the maximum bias which arises when we use the $\min_x \max_u |\text{bias } (u)|$ design in (i) and the maximum bias which arises when we use the $\min_x \max_u |\text{bias } (u)|$ design in (ii).

In part (i)

$$\max_u |\text{bias } (u)| = (4K^2 + 1)/4K\ |d^*| \ ,$$

and in part (ii)

$$\max_u |\text{bias } (u)| = (4K^2 + 4K + 1)/8K\ |d^*| \ .$$

Assume

$$(4K^2 + 1)/4K < (4K^2 + 4K + 1)/8K,$$

then

$$8K^2 + 2 - 4K^2 - 4K - 1 < 0$$

$$4K^2 - 4K + 1 < 0$$

$$(2K - 1)^2 < 0 \ .$$

Thus, we have a contradiction, which implies

$$(4K^2 + 1)/4K \geq (4K^2 + 4K + 1)/8K \quad \text{for every } K.$$

Hence, for $(\Sigma x + K\,\Sigma x^2) > 0$, $\min_x \max_u |\text{bias } (u)|$ is achieved by making

$$(\Sigma x + K\, \Sigma x^2)/N = (4K^2 + 4K - 1)/8K,$$

and for this design

$$\max_u |\text{bias } (u)| = |\text{bias } (-1/2K)|$$

(4.3) $$= [(4K^2 + 4K + 1)/8K]\ |d^*| \ .$$

Next, we must compare this maximum with the maximum obtained in case 2: namely,

(4.4) $$\max_u |\text{bias } (u)| = (1 + K)\ |d^*| \ .$$

Equating the coefficients of d^* in (4.3) and (4.4)

$$(4K^2 + 4K + 1)/8K = 1 + K$$

$$K = (\sqrt{2} - 1)/2 \ .$$

The results of this section can be summarized as follows. For $0 < K < (\sqrt{2} - 1)/2$, $\min_x \max_u |\text{bias } (u)|$ is attained by the design such that

$$(\Sigma x + K \Sigma x^2)/N = 0,$$

for which

$$M(1, 4) = \max_u |\text{bias } (u)| = (1 + K)\ |d^*| \ .$$

For $K > (\sqrt{2} - 1)/2$, $\min_x \max_u |\text{bias } (u)|$ is attained by the design such that

$$(\Sigma x + K\, \Sigma x^2)/N = (4K^2 + 4K - 1)/8K,$$

for which

$$M(1, 4) = \max_u |\text{bias } (u)| = [(4K^2 + 4K + 1)/8K]\ |d^*| \ .$$

4.1.4 True Model: $y = e_1 x + \varepsilon$

$$|\text{bias}(u)| = |u - \Sigma x/N||e_1| .$$

Results are the same as those obtained in section 4.1.1. That is,

$$M(1, 5) = |e_1|$$

4.1.5 True Model: $y = f_2 x^2 + \varepsilon$

$$|\text{bias}(u)| = |u^2 - \Sigma x^2/N||f_2| .$$

Results are the same as those obtained in section 4.1.2. That is,

$$M(1, 6) = |f_2| / 2$$

4.1.6 True Model: $y = g_1 x + g_2 x^2 + \varepsilon$

Assume $g_1 = g^*$ and $g_2 = Kg^*$ $(K > 0)$.

$$|\text{bias}(u)| = |Ku^2 + u - (\Sigma x + K\Sigma x^2)/N| |g^*| .$$

Results are the same as those obtained in section 4.1.3. That is, for $0 < K \leq (\sqrt{2} - 1)/2$, $\min_x \max_u |\text{bias}(u)|$ is attained by the design such that

$$(\Sigma x + K \Sigma x^2) / N = 0,$$

for which

$$M(1, 7) = (1 + K) \; |g^*| .$$

For $K > (\sqrt{2} - 1)/2$, $\min_x \max_u |\text{bias}(u)|$ is attained by the design such that

$$(\Sigma x + K\Sigma x^2)/N = (4K^2 + 4K - 1)/8K,$$

for which

$$M(1, 7) = [(4K^2 + 4K + 1)/8K]\,|g^*| \; .$$

4.2 Assumed Model: $y = b_o + b_1 x + \varepsilon, \quad -1 \leq x \leq 1$

If the true model contains only a constant term, only a linear term, or only a constant term and a linear term, then

$$|\text{bias } (u)| = 0.$$

If the true model contains a quadratic term, then we have

$$|\text{bias } (u)| = \left| u^2 - \frac{[N\Sigma x^3 - \Sigma x \Sigma x^2]}{[N \Sigma x^2 - (\Sigma x)^2]} - \frac{[(\Sigma x^2)^2 - \Sigma x \Sigma x^3]}{[N\Sigma x^2 - (\Sigma x)^2]} \right| |\beta|,$$

where β is the coefficient of the quadratic term. It has been shown by Folks (3) that the $\min_x \max_u |\text{bias } (u)|$ is achieved by the design such that

$$\Sigma x = 0$$

$$\Sigma x^2 = N/2 ,$$

for which

$$M(2, j) = |\beta|/2 \qquad (j = 3, 4, 6, 7) .$$

Thus,

$$M(2, 3) = |c_3|/2$$

$$M(2, 4) = |d_2|/2$$

$$M(2, 6) = |f_2|/2$$

$$M(2, 7) = |g_2|/2 .$$

4.3 Assumed Model: $y = c_o + c_2 x^2 + \varepsilon, \quad -1 \leq x \leq 1$

If the true model contains only a constant term, only a quadratic term, or only a constant term and a quadratic term, then

$$|\text{bias } (u)| = 0 .$$

If the true model contains a linear term, then

$$|\text{bias } (u)| = \left| \frac{[N\Sigma x^3 - \Sigma x \Sigma x^2]}{[N\Sigma x^4 - (\Sigma x^2)^2]} u^2 - u + \frac{[\Sigma x \Sigma x^4 - \Sigma x^2 \Sigma x^3]}{[N\Sigma x^4 - (\Sigma x^2)^2]} \right| |\beta| ,$$

where β is the coefficient of the linear term. Although a $\min_x \max_u$ $|\text{bias } (u)|$ design could not be obtained in general, a solution was determined by restricting the design points to $x = -1$, $x = 0$, and $x = 1$. Under this restriction, the bias is given by

$$|\text{bias } (u)| = |u^2 \Sigma x/\Sigma x^2 - u| |\beta| .$$

For this bias, we see in chapter 3 that the $\min_x \max_u |\text{bias } (u)|$ is achieved by the design such that $\Sigma x = 0$, for which $\max_u |\text{bias } (u)| = |\beta|$. Thus, we have

$$M(3, 2) = |b_1|$$

$$M(3, 4) = |d_1|$$

$$M(3, 5) = |e_1|$$

$$M(3, 7) = |g_1| .$$

4.4 Assumed Model: $y = e_1 x + \varepsilon$, $-1 \le x \le 1$

4.4.1 True Model: $y = a_o + \varepsilon$

$$|\text{bias } (u)| = |u \Sigma x/ \Sigma x^2 - 1| |a_o| .$$

Case 1: $\Sigma x < 0$

$$\max_u |\text{bias } (u)| = |\text{bias } (1)|$$

$$\max_u |\text{bias}(u)| = |\Sigma x/\Sigma x^2 - 1||a_o|$$
$$> |a_o| .$$

<u>Case 2:</u> $\Sigma x = 0$

$$\max_u |\text{bias}(u)| = |\text{bias}(\pm 1)|$$
$$= |\text{bias}(0)|$$
$$= |a_o| .$$

<u>Case 3:</u> $\Sigma x > 0$

$$\max_u |\text{bias}(u)| = |\text{bias}(-1)|$$
$$= |-\Sigma x/\Sigma x^2 - 1||a_o|$$
$$> |a_o| .$$

Therefore, the $\min_x \max_u |\text{bias}(u)|$ is achieved by the design such that $\Sigma x = 0$, for which

$$M(5, 1) = |a_o| .$$

<u>4.4.2 True Model: $y = b_o + b_1 x + \varepsilon$</u>

$$|\text{bias}(u)| = |u\,\Sigma x/\Sigma x^2 - 1||b_o| .$$

Results are the same as those obtained in section 4.4.1.

$$M(5, 2) = |b_o| .$$

<u>4.4.3 True Model $y = c_o + c_2 x^2 + \varepsilon$</u>

Assume $c_o = c^*$ and $c_2 = Kc^*$ $(K > 0)$.

$$|\text{bias}(u)| = |Ku^2 - [(\Sigma x + K\Sigma x^3)/\Sigma x^2]\,u + 1||c^*| .$$

<u>Case 1:</u> $(\Sigma x + K\,\Sigma x^3) < 0$

$$\max_u |\text{bias}(u)| = |\text{bias}(1)|$$

$$\max_u |\text{bias } (u)| = |K - [(\Sigma x + K\Sigma x^3)/\Sigma x^2] + 1||c^*|$$
$$> (1 + K) |c^*|.$$

Case 2: $(\Sigma x + K\Sigma x^3) = 0$

$$\max_u |\text{bias } (u)| = |\text{bias } (\pm 1)|$$
$$= (1+K) |c^*| .$$

Case 3: $(\Sigma x + K \Sigma x^3) > 0$

$$\max_u |\text{bias } (u)| = |\text{bias } (-1)|$$
$$= |K + [(\Sigma x + K\Sigma x^3)/\Sigma x^2] + 1||c^*| .$$
$$> (1 + K) |c^*| .$$

Therefore, the $\min_x \max_u |\text{bias } (u)|$ is achieved by the design such that $\Sigma x + K \Sigma x^3 = 0$, for which

$$M(5, 3) = (1 + K) |c^*| .$$

4.4.4 True Model: $y = d_o + d_1 x d_2 x^2 + \varepsilon$

Assume $d_o = d^*$ and $d_2 = Kd^*$ $(K > 0)$.

$$|\text{bias } (u)| = |K u^2 - [(\Sigma x + K\Sigma x^3)/\Sigma x^2] u^2 + 1||d^*| .$$

Results are the same as those obtained in section 4.4.3.

$$M(5, 4) = (1 + K) | d^*| .$$

4.4.5 True Model: $y = f_2 x^2 + \varepsilon$

$$|\text{bias } (u)| = |u \Sigma x^3/\Sigma x^2 - u^2||f_2| ,$$

Case 1: $\Sigma x^3 < 0$

$$\max_u |\text{bias } (u)| = |\text{bias } (1)|$$

$$\max_u |\text{bias } (u)| = |\Sigma x^3/\Sigma x^2 - 1|\,|f_2|$$
$$> |f_2| .$$

Case 2: $\Sigma x^3 = 0$

$$\max_u |\text{bias } (u)| = |\text{bias } (\pm 1)|$$
$$= |f_2|.$$

Case 3: $\Sigma x^3 > 0$

$$\max_u |\text{bias } (u)| = |\text{bias } (-1)|$$
$$= |\Sigma x^3/\Sigma x^2 + 1|\,|f_2|$$
$$> |f_2| .$$

Thus, the $\min_x \max_u |\text{bias } (u)|$ is achieved by the design such that $\Sigma x^3 = 0$, for which

$$M(5, 6) = |f_2| .$$

4.4.6 True Model: $y = g_1 x + g_2 x^2 + \varepsilon$

$$|\text{bias } (u)| = |u\, \Sigma x^3/\Sigma x^2 - u^2|\,|g_2|.$$

Results are the same as those obtained in section 4.4.5

$$M(5, 7) = |g_2| .$$

4.5 Assumed Model: $y = f_2 x^2 + \varepsilon, \quad -1 \le x \le 1$

4.5.1 True Model: $y = a_o + \varepsilon$

$$|\text{bias } (u)| = |u^2\, \Sigma x^2/\Sigma x^4 - 1|\,|a_o|.$$

Case 1: $\Sigma x^2/\Sigma x^4 > 2$

$$\max_u |\text{bias } (u)| = |\text{bias } (\pm 1)|$$

$$\max_u |\text{bias}(u)| = |\Sigma x^2/\Sigma x^4 - 1||a_o|$$
$$> |a_o| .$$

Case 2: $\Sigma x^2/\Sigma x^4 \leq 2$

$$\max_u |\text{bias}(u)| = |\text{bias}(0)|$$
$$= |a_o| .$$

To $\min_x \max_u |\text{bias}(u)|$, choose the design such that

$$\Sigma x^2 \leq 2\Sigma x^4 ,$$

then

$$M(6, 1) = |a_o| .$$

4.5.2 True Model: $y = b_o + b_1 x + \varepsilon$

Assume $b_o = b^*$ and $b_1 = Kb^*$ $(K > 0)$.

$$|\text{bias}(u)| = |u^2(\Sigma x^2 + K\Sigma x^3)/\Sigma x^4 - Ku - 1||b^*| .$$

Case 1: $(\Sigma x^2 + K\Sigma x^3) < 0$

$$\max_u |\text{bias}(u)| = |\text{bias}(1)|$$
$$= |1 + K - (\Sigma x^2 + K\Sigma x^3)/\Sigma x^4| |b^*|$$
$$> (1+K)|b^*| .$$

Case 2: $(\Sigma x^2 + K\Sigma x^3) = 0$

$$\max_u |\text{bias}(u)| = |\text{bias}(1)|$$
$$= (1 + K) |b^*| .$$

Case 3: $(\Sigma x^2 + K\Sigma x^3) > 0$

Let

$$A = (\Sigma x^2 + K\Sigma x^3)/\Sigma x^4 .$$

Let

$$f(u) = A\,u^2 - Ku - 1,$$

then the minimum of $f(u)$ is attained at $u = K/2A$.

(i) $0 < A \leq 1$.

For this condition, we have 2 possibilities

$$u = 1: \; |\text{bias } (1)| = |A - K - 1|\,|b^*|$$
$$= (1 + K - A)\,|b^*|.$$

$$u = K/2A: \; |\text{bias } (K/2A)| = (K^2/4A + 1)\,|b^*|.$$

To determine which bias is the largest, assume

$$K^2/4A + 1 = 1 + K - A.$$

$$K^2 - 4AK + 4A^2 = 0$$

$$(K - 2A)^2 = 0,$$

then

$$\max_u |\text{bias } (u)| = |\text{bias } (K/2A)|$$

$$= (K^2/4A + 1)\,|b^*| \text{ for every } K > 0.$$

$$\min_{0<A\leq 1} \max_u |\text{bias } (u)| = (K^2/4 + 1)\,|b^*|.$$

(ii) $A \geq 1$

Consider the two possibilities

$$u = -1: \; |\text{bias } (-1)| = |A + K - 1|\,|b^*|$$
$$= (A + K - 1)\,|b^*|.$$

$$u = (K/2A): \; |\text{bias } (K/2A)| = (K^2/4A + 1)\,|b^*|.$$

Suppose

$$K^2/4A + 1 = A + K - 1$$

$$4A^2 + 4(K - 2)A - K^2 = 0.$$

Then

$$A = -(K - 2)/2 + [\sqrt{2K^2 - 4K + 4}]/2,$$

which implies we must consider which of the two possibilities is maximum when

$$1 \leq A \leq -(K - 2)/2 + [\sqrt{2K^2 - 4K + 4}]/2,$$

and when

$$A > -(K - 2)/2 + [\sqrt{2K^2 - 4K + 4}]/2.$$

For $1 \leq A \leq -(K - 2) + [\sqrt{2K^2 - 4K + 4}]/2$,

$$\max_u [K^2/4A + 1, \; A + K - 1] = K^2/4A + 1,$$

which is minimum when

$$A = -(K - 2)/2 + [\sqrt{2K^2 - 4K + 4}]/2.$$

For $A > -(K - 2)/2 + [\sqrt{2K^2 - 4K + 4}]/2$,

$$\max_u [K^2/4A + 1, \; A + K - 1] = A + K - 1,$$

which is minimum when

$$A = -(K - 2)/2 + [\sqrt{2K^2 - 4K + 4}]/2.$$

To summarize, we have

(i) $A < 0$

$$\max_u |\text{bias}(u)| > (1 + K)|b^*|.$$

(ii) $A = 0$

$$\max_u |\text{bias}(u)| = (1+K)|b^*|.$$

(iii) $0 < A \leq 1$

$$\min_A \max_u |\text{bias}(u)| = (K^2/4 + 1)|b^*|.$$

(iv) $A > 1$

$$\min_{A}\max_{u}|\text{bias }(u)| = [K/2 + (\sqrt{2K^2 - 4K + 4})/2\,]\,|b^*|.$$

Comparing these four quantities we find that for $0 < K < 8$, the $\min_{x}\max_{u}|\text{bias }(u)|$ is achieved by the design such that

$$(\Sigma x^2 + K\Sigma x^3) / \Sigma x^4 = -(K - 2) / 2 + (\sqrt{2K^2 - 4K + 4}) / 2 ,$$

for which

$$M(6, 2) = [K/2 + (\sqrt{2K^2 - 4K + 4})/2]\ |b^*| .$$

For $K > 8$, the $\min_{x}\max_{u}|\text{bias }(u)|$ is attained by the design such that

$$(\Sigma x^2 + K\,\Sigma x^3) / \Sigma x^4 = 0,$$

for which

$$M(6, 2) = (1 + K)\ |b^*| .$$

4.5.3 True Model: $y = c_o + c_2x^2 + \varepsilon$

$$|\text{bias }(u)| = |u^2\,\Sigma x^2 / \Sigma x^4 - 1|\,|c_o| .$$

Results are the same as those obtained in section 4.5.1.

$$M(6, 3) = |c_o| .$$

4.5.4 True Model: $y = d_o + d_1x + d_2x^2 + \varepsilon$

Assume $d_o = d_1 = d^*$, then

$$|\text{bias }(u)| = |u^2(\Sigma x^2 + \Sigma x^3)/\Sigma x^4 - u - 1|\ |d^*|.$$

Results are the same as those obtained in section 4.5.2 with $K = 1$.

$$M(6, 4) = [(1 + \sqrt{2}\,)/2]\ |d^*| .$$

4.5.5 True Model: $y = e_1 x + \varepsilon$

$$|\text{bias }(u)| = |u - u^2 \Sigma x^3/\Sigma x^4|\ |e_1| .$$

Case 1: $\Sigma x^3 < 0$

$$\max_u |\text{bias }(u)| = |\text{bias }(1)|$$
$$= |1 - \Sigma x^3/\Sigma x^4|\,|e_1|$$
$$> |e_1| .$$

Case 2: $\Sigma x^3 = 0$

$$\max_u |\text{bias }(u)| = |\text{bias }(\pm 1)|$$
$$= |e_1| .$$

Case 3: $\Sigma x^3 > 0$

$$\max_u |\text{bias }(u)| = |\text{bias }(-1)|$$
$$= |-1 - \Sigma x^3/\Sigma x^4|\,|e_1|$$
$$> |e_1| .$$

Thus, the $\min_x \max_u |\text{bias }(u)|$ is achieved by the design such that $\Sigma x^3 = 0$, for which

$$M(6, 5) = |e_1| .$$

4.5.6 True Model: $y = g_1 x + g_2 x^2 + \varepsilon$

$$|\text{bias }(u)| = |u - u^2 \Sigma x^3 / \Sigma x^4|\,|g_1| .$$

Results are the same as those obtained in section 4.5.5.

$$M(6, 7) = |g_1| .$$

4.6 Assumed Model: $y = g_1 x + g_2 x^2 + \varepsilon, \quad -1 \leq x \leq 1$

If the true model contains only linear and, or quadratic terms, then

$$|\text{bias } (u)| = 0.$$

If the true model contains a constant term, then we have

$$|\text{bias } (u)| = \left| u^2 \frac{[(\Sigma x^2)^2 - \Sigma x \Sigma x^3]}{[\Sigma x^2 \Sigma x^4 - (\Sigma x^3)^2]} + u \frac{[\Sigma x \Sigma x^4 - \Sigma x^2 \Sigma x^3]}{[\Sigma x^2 \Sigma x^4 - (\Sigma x^3)^2]} - 1 \right| |\beta|,$$

where β is the coefficient of the constant term. Consider only the class of designs which have design points $x = -1$, $x = 0$, and $x = 1$, then

$$|\text{bias } (u)| = |u^2 - 1| |\beta|.$$

Since the bias function is independent of the design points, every design achieves the $\min_x \max_u |\text{bias } (u)|$, for which

$$M(7,\ 1) = |a_o|$$

$$M(7,\ 2) = |b_o|$$

$$M(7,\ 3) = |c_o|$$

$$M(7,\ 4) = |d_o|.$$

4.7 Determining the Min Max Model

We now have the maximum biases that arise when we assume the wrong model. Before we can compare these maximum biases, we must express each of the coefficients of the various models in terms of some common element. Suppose then that we know that the true model has a function vlaue of $a_o + \varepsilon$ at $x = \delta(-1<\delta<0;\ 0<\delta \leq 1)$.

Under the assumption that $K = 1$, we have

$$b_o = b_1 = b^*$$
$$c_o = c_2 = c^*$$
$$d_o = d_1 = d_2 = d^*$$
$$g_1 = g_2 = g^* .$$

Expressing each of the coefficients in terms of a common element, we have

(1) $$b_o = b_1 = \frac{a_o}{(1 + \delta)}$$

(2) $$c_o = c_2 = \frac{a_o}{(1 + \delta^2)}$$

(3) $$d_o = d_1 = d_2 = \frac{a_o}{1 + \delta + \delta^2}$$

(4) $$e_1 = \frac{a_o}{\delta}$$

(5) $$f_2 = \frac{a_o}{\delta^2}$$

(6) $$g_1 = g_2 = \frac{a_o}{\delta + \delta^2} .$$

We will now compare the maximum biases that arise for each of the assumed models.

<u>4.7.1 Assumed model: $y = a_o + \varepsilon$</u>

$$M(1, 2) = |b_1| = \frac{a_o}{|1 + \delta|}$$

$$M(1, 3) = 1/2\ |c_2|\ \frac{a_o}{2\ |1 + \delta^2|}$$

$$M(1, 4) = 9/8\ |d^*| = \frac{9\ |a_o|}{8\ |1+\delta+\delta^2|}$$

$$M(1, 5) = |e_1| = \frac{|a_o|}{|\delta|}$$

$$M(1, 6) = 1/2\ |f_2| = \frac{|a_o|}{2\ |\delta^2|}$$

$$M(1, 7) = 9/8\ |g^*| = \frac{9\ |a_o|}{8\ |\delta+\delta^2|}$$

<u>Case 1:</u> Compare M(1, 2) with M(1, 6).

Assume M(1, 2) = M(1, 6), then

$$\frac{1}{2\delta^2} = \frac{1}{1+\delta}$$

$$2\delta^2 - \delta - 1 = 0$$

$$\delta = 1/4 \pm 3/4 .$$

For $-1 < \delta \leq -1/2$

$$\frac{|a_o|}{|1+\delta|} > \frac{|a_o|}{|2\ \delta^2|} .$$

$$M(1, 2) > M(1, 6) .$$

For $-1/2 < \delta < 0$ and $0 < \delta \leq 1$

$$\frac{|a_o|}{2\ |\delta^2|} > \frac{|a_o|}{|1+\delta|} .$$

$$M(1, 6) > M(1, 2) .$$

<u>Case 2:</u> Compare M(1, 3) with M(1, 6).

Since $\delta^2 < 1 + \delta^2$, we have

$$\frac{|a_o|}{2\ |\delta^2|} > \frac{|a_o|}{2\ |1 + \delta^2|} \quad \text{for every } \delta .$$

$$M(1, 6) > M(1, 3) .$$

<u>Case 3:</u> Compare $M(1, 4)$ with $M(1, 6)$.

Assume $M(1, 4) = M(1, 6)$, then

$$2\delta^2 = 8/9\ (1 + \delta + \delta^2)$$

$$5\delta^2 - 4\delta - 4 = 0$$

$$\delta = 2/5\ (1 \pm \sqrt{6}\ ).$$

For $\delta > 2/5\ (1 - \sqrt{6}\ )$

$$\frac{|a_o|}{2\ |\delta^2|} > \frac{9\ |a_o|}{8\ |1+\delta+\delta^2|}$$

$$M(1, 6) > M(1, 4) .$$

For $\delta < 2/5\ (1 - \sqrt{6})$

$$\frac{9\ |a_o|}{8\ |1+\delta+\delta^2|} > \frac{|a_o|}{2\ |\delta^2|}$$

$$M(1, 4) > M(1, 6) .$$

<u>Case 4:</u> Compare $M(1, 5)$ with $M(1, 6)$.

(i) $0 < \delta \leq 1$

Assume $M(1, 5) = M(1, 6)$, then

$$2\delta^2 = \delta$$

$$\delta = 0, \ 1/2.$$

For $0 < \delta \leq 1/2$, $M(1, 6) > M(1, 5)$.

For $1/2 < \delta \leq 1$, $M(1, 5) > M(1, 6)$.

(ii) $-1 < \delta < 0$

Assume $M(1, 5) = M(1, 6)$, then

$$2\,\delta^2 = -\delta$$

$$\delta = 0, \; -1/2 .$$

For $-1 < \delta \leq -1/2$, $M(1, 5) > M(1, 6)$.

For $-1/2 < \delta < 0$, $M(1, 6) > M(1, 5)$.

Case 5: Compare $M(1, 6)$ with $M(1, 7)$.

(i) $0 < \delta \leq 1$, $|\delta + \delta^2| = \delta + \delta^2$

Assume $M(1, 7) > M(1, 6)$, then

$$2\delta^2 > 8/9\,(\delta + \delta^2)$$

$$\delta > 4/5 .$$

For $0 < \delta \leq 4/5$, $M(1, 6) > M(1, 7)$.

For $4/5 < \delta \leq 1$, $M(1, 5) > M(1, 6)$.

(ii) $-1 < \delta < 0$, $|\delta + \delta^2| = -(\delta + \delta^2)$

Assume $M(1, 7) > M(1, 6)$, then

$$2\,\delta^2 > -8/9\,(\delta + \delta^2)$$

$$\delta < -8/13.$$

For $-1 < \delta \leq -8/13$, $M(1, 7) > M(1, 6)$.

For $-8/13 < \delta < 0$, $M(1, 6) > M(1, 7)$.

Case 6: Compare $M(1, 2)$ and $M(1, 5)$ in the interval $-1 < \delta < -1/2$.

Assume $M(1, 2) > M(1, 5)$, then

$$-\delta > 1 + \delta$$

$$\delta < -1/2 .$$

Therefore $M(1, 2) > M(1, 5)$ for $-1 < \delta < -1/2$.

Case 7: Compare $M(1, 2)$ with $M(1, 4)$ in the interval $-1 < \delta < -1/2$.

Assume $M(1, 2) = M(1, 4)$, then

$$8/9\,(1 + \delta + \delta^2) = 1 + \delta$$

$$8\,\delta^2 - \delta - 1 = 0$$

$$\delta = 1/16\,(1 - \sqrt{33}\,).$$

Thus, for $\delta < 1/16\,(1 - \sqrt{33}\,)$, $M(1, 2) > M(1, 4)$ which implies $M(1, 2) > M(1, 4)$ for $-1 < \delta < -1/2$.

Case 8: Compare $M(1, 2)$ with $M(1, 7)$ in the interval $-1 < \delta < -8/13$.

Note that

$$(1 + \delta) - [\,8/9(\delta + \delta^2)\,] > 0 \quad \text{for} \quad -1 < \delta < 0,$$

which implies

$$(1 + \delta) > -8/9(\delta + \delta^2).$$

Therefore

$$M(1, 7) > M(1, 2).$$

Case 9: Compare $M(1, 5)$ with $M(1, 7)$ in the interval $4/5 < \delta \leq 1$.

Assume $M(1, 5) > M(1, 7)$, then

$$8/9\,(\delta + \delta^2) > \delta$$

$$\delta > 1/8,$$

which implies that $M(1, 5) > M(1, 7)$.

To summarize, the maximum bias which arises when we assume the model

$$y = a_o + \varepsilon$$

is given by

M(1, 7) in the interval $-1 < \delta \leq -8/13$

M(1, 2) in the interval $-8/13 < \delta < -1/2$

M(1, 6) in the interval $-1/2 < \delta < 0; 0 < \delta \leq 1/2$

M(1, 5) in the interval $1/2 < \delta \leq 1$.

Similar results can be obtained using the other assumed models in C.

The following table gives the maximum bias that arises for each assumed model. The maximums are indicated for the various values of δ.

TABLE I

MAXIMUM BIAS THAT ARISES FOR EACH ASSUMED MODEL

Y_A \ δ	$-1<\delta\leq-\frac{8}{13}$	$-\frac{8}{13}<\delta\leq-\frac{1}{2}$	$-\frac{1}{2}<\delta<0$	$0<\delta\leq\frac{1}{2}$	$\frac{1}{2}<\delta\leq\frac{\sqrt{2}}{2}$	$\frac{\sqrt{2}}{2}<\delta\leq 1$
$h_1(x)+\varepsilon$	$9/8\ \lvert g^*\rvert$	$\lvert b_1\rvert$	$1/2\lvert f_2\rvert$	$1/2\lvert f_2\rvert$	$\lvert e_1\rvert$	$\lvert e_1\rvert$ **
$h_2(x)+\varepsilon$	$1/2\ \lvert g_2\rvert$ **	$1/2\ \lvert g_2\rvert$	$1/2\lvert f_2\rvert$	$1/2\lvert f_2\rvert$	$1/2\lvert f_2\rvert$	$1/2\lvert f_2\rvert$
$h_3(x)+\varepsilon$	$\lvert g_1\rvert$	$\lvert g_1\rvert$	$\lvert g_1\rvert$	$\lvert e_1\rvert$	$\lvert e_1\rvert$	$\lvert e_1\rvert$
$h_5(x)+\varepsilon$	$\lvert g_2\rvert$	$\lvert g_2\rvert$	$\lvert f_2\rvert$	$\lvert f_2\rvert$	$\lvert f_2\rvert$	$\lvert f_2\rvert$
$h_6(x)+\varepsilon$	$\lvert g_1\rvert$	$\lvert g_1\rvert$	$\lvert g_1\rvert$	$\lvert e_1\rvert$	$\lvert e_1\rvert$	$\lvert e_1\rvert$
$h_7(x)+\varepsilon$	$\lvert b_o\rvert$	$\lvert b_o\rvert$ **	$\lvert b_o\rvert$ **	$\lvert a_o\rvert$ **	$\lvert a_o\rvert$ **	$\lvert a_o\rvert$

Y_A denotes the assumed model.

** denotes the $\min_x \max_u$ model for that particular value of δ. Recall that δ is a point where the function value of the true model is known to be $a_o + \varepsilon$. Thus, for $K = 1$, the $\min_x \max_u$ model is given by

$$y = h_2(x) + \varepsilon = b_o + b_1 x + \varepsilon \qquad \text{for } -1 < \delta \leq -8/13$$

$$y = h_7(x) + \varepsilon = g_1 x + g_2 x^2 + \varepsilon \qquad \text{for } -8/13 < \delta < 0$$

$$y = h_7(x) + \varepsilon = g_1 x + g_2 x^2 + \varepsilon \qquad \text{for } 0 < \delta < \sqrt{2}/2$$

$$y = h_2(x) + \varepsilon = b_o + b_1 x + \varepsilon \qquad \text{for } \sqrt{2}/2 < \delta \leq 1 .$$

Similarly, if only the assumed models with one term are considered, the $\min_x \max_u$ model is given by

$$y = h_6(x) + \varepsilon = f_2 x^2 + \varepsilon \qquad \text{for every } \delta .$$

For $K > 0$, the $\min_x \max_u$ models were determined for $0 < \delta \leq 1$; however, because of the large number of special cases involving different values of K and δ, the solutions will not be given here.

In this chapter, we assumed a model $h_i(x) + \varepsilon$ (i arbitrary), when the true model was $h_j(x) + \varepsilon$ ($i \neq j$). For this combination we determined the $\min_x \max_u |\text{bias } (u)|$ design and the maximum bias $M(i, j)$ for this design. This was done for each $i, j = 1, 2, \ldots, 7$, except for $i = j$ and $i = 4$. Then for each assumed model $h_k(x) + \varepsilon$ (k arbitrary) we obtained the maximum $M(k, j)$ with respect to j. Thus, we obtained six $M(i, j)$ values. The assumed model associated with the smallest of these $M(i, j)$ values was determined to be the $\min_x \max_u$ model.

CHAPTER V

AVERAGE VARIANCE OF THE ESTIMATED RESPONSE

In this chapter we will be concerned with determining the average variance of the estimated response $\hat{y}(u_1, u_2)$, in the two-dimensional case, for any distribution of the total probability mass to the region of interest; namely, the square region

$$R = [(u_1, u_2) \mid -1 \leq u_i \leq 1; \quad i = 1, 2],$$

and the circular region

$$R_c = [(u_1, u_2) \mid u_1^2 + u_2^2 \leq 1] .$$

In either case, assume the model

$$y = \beta_o + \beta_1 x_1 + \beta_2 x_2 + \varepsilon .$$

Although the average variance of $\hat{y}(u_1, u_2)$ has been determined previously by Folks (3), it was determined under the assumption that every point in the region of interest was assigned equal probability mass. That is, the density function of (u_1, u_2) in the square region R was given by

$$\begin{aligned} f(u_1, u_2) &= 1/4 \qquad (u_1, u_2) \in R \\ &= 0 \qquad \text{otherwise.} \end{aligned}$$

In the circular region R_c, the density function was given by

$$f(u_1, u_2) = 1/\Pi \qquad (u_1, u_2) \in R_c$$

$$f(u_1, u_2) = 0 \qquad \text{otherwise.}$$

Suppose, however, that we are more interested in making predictions in one subregion of R or R_c than in others. It may be that we suspect the response will vary greater in one particular region; in which case, we might want to assign a larger proportion of the probability mass to this region. As an example, assume the design points can be chosen anywhere in the square region R, but we are more interested in making predictions in the region about the origin, say the region

$$R_1 = [(u_1, u_2) | -1/2 \leq u_i \leq 1/2; \ i = 1, 2] .$$

Thus, we might assign probability mass equal to $1/2$ to this region and probability mass equal to $1/2$ to the region

$$R_2 = R - R_1 ,$$

where $R - R_1$ denotes the points of R which are not in R_1. Hence, we would have

$$\begin{aligned} f(u_1, u_2) &= \text{mass/area} \\ &= (1/2)/1 \\ &= 1/2 \qquad (u_1, u_2) \ \varepsilon \ R_1 \\ f(u_1, u_2) &= (1/2)/3 \\ &= 1/6 \qquad (u_1, u_2) \ \varepsilon \ R_2 \\ f(u_1, u_2) &= 0 \qquad \text{otherwise.} \end{aligned}$$

Consider now some questions which might arise in connection with the distribution of the probability mass. How does the distribution of the probability mass to R or R_c effect the average variance of $\hat{y}(u_1, u_2)$? Can the average variance of $\hat{y}(u_1, u_2)$ be minimized or maximized with respect to the distribution of the probability mass to subregions of R or R_c? Is the minimum average variance of $\hat{y}(u_1, u_2)$ with respect to the choice of design a function of the distribution of the probability mass? These and other questions will be of interest in this chapter.

To answer the question of how the distribution of the probability mass to R and R_c effects the average variance of $\hat{y}(u_1, u_2)$, consider the following theorems and corollaries.

5.1 Average Variance of $\hat{y}(u_1, u_2)$ in the Square Region $R=[(u_1, u_2) \mid -1 \leq u_i \leq 1; i = 1, 2]$ and in the Circular Region $R_c=[(u_1,u_2) \mid u_1^2+u_2^2 \leq 1]$

Theorem 5.1 If the total probability mass $M = 1$ is assigned to the n subregions

$$R_i = [u_1, u_2) \mid -a_i \leq u_j \leq a_i; \; j = 1, 2] - R_{i-1}$$

$(i = 1, 2, \ldots, n;\ R_o = \phi,\ a_o = 0)$, of the region R, then the average variance of $\hat{y}(u_1, u_2)$ over R is given by

$$\text{ave var } \hat{y}(u_1,u_2) = \text{var } \hat{\beta}_o+[(1/3)(\text{var } \hat{\beta}_1+\text{var } \hat{\beta}_2)] \sum_{i=1}^{n} M_i(a_i^2+a_{i-1}^2),$$

where M_i denotes the probability mass assigned to R_i and $\sum_{i=1}^{n} M_i = 1$.

Proof: Let

$$A_i = (2a_i)^2 - (2a_{i-1})^2$$

be the area of $R_i (i = 1, 2, \ldots, n;\ a_o = 0)$, then

$$\begin{aligned} f_i(u_1, u_2) &= M_i/A_i \\ &= K_i \qquad (u_1, u_2) \in R_i \\ &= 0 \qquad \text{otherwise} \end{aligned}$$

is the density function of (u_1, u_2) in R. Thus,

$$\text{ave var } \hat{y}(u_1, u_2) = \sum_{i=1}^{n} \Big[\int_{-a_i}^{a_i} \int_{-a_i}^{a_i} \text{var } \hat{y}(u_1, u_2) f_i(u_1, u_2)\, du_1 du_2$$

$$- \int_{-a_{i-1}}^{a_{i-1}} \int_{-a_{i-1}}^{a_{i-1}} \text{var } \hat{y}(u_1, u_2)\, f_i(u_1, u_2)\, du_1 du_2 \Big]$$

$$= \sum_{i=1}^{n} K_i \int_{-a_i}^{a_i} \int_{-a_i}^{a_i} [\text{var } \hat{\beta}_o + u_1^2 \text{ var } \hat{\beta}_1 + u_2^2 \text{ var } \hat{\beta}_2 + 2u_1 \text{ cov}(\hat{\beta}_o, \hat{\beta}_1)$$

$$+ 2u_2 \text{ cov}(\hat{\beta}_o, \hat{\beta}_2) + 2 u_1 u_2 \text{ cov}(\hat{\beta}_1, \hat{\beta}_2)]\, du_1 du_2$$

$$- \int_{-a_{i-1}}^{a_{i-1}} \int_{-a_{i-1}}^{a_{i-1}} [\text{var } \hat{\beta}_o + u_1^2 \text{var } \hat{\beta}_1 + u_2^2 \text{ var } \hat{\beta}_2 + 2u_1 \text{ cov}(\hat{\beta}_o, \hat{\beta}_1)$$

$$+ 2 u_2 \text{ cov}(\hat{\beta}_o, \hat{\beta}_2) + 2u_1 u_2 \text{ cov}(\hat{\beta}_1, \hat{\beta}_2)]\, du_1 du_2$$

$$= \sum_{i=1}^{n} K_i \int_{-a_i}^{a_i} [2a_i \text{ var } \hat{\beta}_o + (2/3)a_i^3 \text{ var } \hat{\beta}_1 + 2a_i u_2^2 \text{ var } \hat{\beta}_2 + 4a_i u_2 \text{cov}(\hat{\beta}_o, \hat{\beta}_2)]\, du_2$$

$$- \int_{-a_{i-1}}^{a_{i-1}} [2a_{i-1} \text{var } \hat{\beta}_o + (2/3)a_{i-1}^3 \text{var } \hat{\beta}_1 + 2a_{i-1} u_2 \text{var } \hat{\beta}_2$$

$$+ 4a_{i-1} u_2 \text{ cov}(\hat{\beta}_o, \hat{\beta}_2)]\, du_2$$

$$= \sum_{i=1}^{n} K_i \; [4a_i^2 \text{ var } \hat{\beta}_o + (4/3)a_i^4 \text{var } \hat{\beta}_1 + (4/3)a_i^4 \text{var } \hat{\beta}_2] - [4a_{i-1}^2 \text{ var } \hat{\beta}_o$$

$$+ (4/3)a_{i-1}^4 \text{ var } \hat{\beta}_1 + (4/3)a_{i-1}^4 \text{ var } \hat{\beta}_2]$$

$$= \sum_{i=1}^{n} K_i [4 \text{ var } \hat{\beta}_o (a_i^2 - a_{i-1}^2) + (4/3)(\text{var } \hat{\beta}_1 + \text{var } \hat{\beta}_2)(a_i^4 - a_{i-1}^4)]$$

$$= \sum_{i=1}^{n} \frac{M_i}{4(a_i^2 - a_{i-1}^2)} [4 \text{ var } \hat{\beta}_o (a_i^2 - a_{i-1}^2) + (4/3)(\text{var } \hat{\beta}_1 + \text{var } \hat{\beta}_2)(a_i^4 - a_{i-1}^4)]$$

$$= \sum_{i=1}^{n} M_i \left[\text{var } \hat{\beta}_o + (1/3)(\text{var } \hat{\beta}_1 + \text{var } \hat{\beta}_2)(a_i^2 + a_{i-1}^2)\right]$$

$$= \left(\sum_{i=1}^{n} M_i\right)(\text{var } \hat{\beta}_o) + (1/3)(\text{var } \hat{\beta}_1 + \text{var } \hat{\beta}_2) \sum_{i=1}^{n} M_i \,(a_i^2 + a_{i-1}^2)$$

$$\text{ave var } \hat{y}\,(u_1, u_2) = \text{var } \hat{\beta}_o + (1/3)(\text{var } \hat{\beta}_1 + \text{var } \hat{\beta}_2) \sum_{i=1}^{n} M_i \,(a_i^2 + a_{i-1}^2) .$$

This completes the proof.

Corollary 5. 1. 1. For any subdivision of the region R and for any distribution of the total probability mass to these subdivisions, the the minimum average variance of $\hat{y}(u_1, u_2)$ with respect to design is attained by taking $N/4$ points at each corner of R.

Proof: From theorem 5. 1,

$$\text{ave var } \hat{y}\,(u_1, u_2) = \text{var } \hat{\beta}_o + (1/3)(\text{var } \hat{\beta}_1 + \text{var } \hat{\beta}_2) \sum_{i=1}^{n} M_i (a_i^2 + a_{i-1}^2).$$

To minimize this expression with respect to design, the quantity $\sum_{i=1}^{n} M_i (a_i^2 + a_{i-1}^2)$ can be regarded as a constant. Thus, the average var $\hat{y}\,(u_1, u_2)$ is minimized by the design which gives simultaneous maximum precision on the β's; that is, one which simultaneously maximizes all of the diagonal elements of $X'X$ and makes the off-diagonal elements zero. Proof of this is found in Tocher (8). Such a design is given by Folks (3). That is, take $N/4$ points at each corner of R.

Since $\sum_{i=1}^{n} M_i (a_i^2 + a_{i-1}^2)$ is a constant with respect to determining the design which minimizes the average variance of $\hat{y}\,(u_1, u_2)$, the corollary is proved.

Theorem 5.2 If the total probability mass $M = 1$ is assigned to the n subregions

$$R_i = [(u_1, u_2) \mid a_{i-1}^2 \leq u_1^2 + u_2^2 \leq a_i^2],$$

$(i = 1, 2, \ldots, n; a_o = 0)$, of the region R_c, then the average variance of $\hat{y}(u_1, u_2)$ over R_c is given by

$$\text{ave var } \hat{y}(u_1, u_2) = \text{var } \hat{\beta}_o + [(1/4)(\text{var } \hat{\beta}_1 + \text{var } \hat{\beta}_2)] \sum_{i=1}^{n} M_i(a_i^2 + a_{i-1}^2),$$

where M_i denotes the probability mass assigned to R_i and $\sum_{i=1}^{n} M_i = 1$.

Proof: Let

$$A_i = \Pi (a_i^2 - a_{i-1}^2)$$

be the area of R_i, then

$$\begin{aligned} f_i(u_1, u_2) &= M_i/A_i \\ &= K_i && (u_1, u_2) \in R_i \\ &= 0 && \text{otherwise} \end{aligned}$$

is the density function of (u_1, u_2) in R_c. Thus,

$$\text{ave var } \hat{y}(u_1, u_2) = \sum_{i=1}^{n} 4 [\int_o^{a_i} \int_o^{\sqrt{a_i^2 - u_2^2}} f_i(u_1, u_2) \text{ var } \hat{y}(u_1, u_2) \, du_1 du_2$$

$$- \int_o^{a_i} \int_o^{\sqrt{a_{i-1}^2 - u_2^2}} f_i(u_1, u_2) \text{ var } \hat{y}(u_1, u_2) \, du_1 du_2]$$

$$= \sum_{i=1}^{n} 4K_i \int_o^{a_i} \int_o^{\sqrt{a_i^2 - u_2^2}} [\text{var } \hat{\beta}_o + u_1^2 \text{ var } \hat{\beta}_1 + u_2^2 \text{ var } \hat{\beta}_2 + 2u_1 \text{ cov}(\hat{\beta}_o, \hat{\beta}_1)$$

$$+ 2 u_2 \text{ cov } (\hat{\beta}_o, \hat{\beta}_2) + 2u_1 u_2 \text{ cov } (\hat{\beta}_1, \hat{\beta}_2)] \, du_1 \, du_2$$

$$- \int_o^{a_i} \int_o^{\sqrt{a_{i-1}^2 - u_2^2}} [\text{var } \hat{\beta}_o + u_1^2 \text{ var } \hat{\beta}_1 + u_2^2 \text{ var } \hat{\beta}_2 + 2u_1 \text{cov}(\hat{\beta}_o, \hat{\beta}_1)$$

$$+ 2u_2 \operatorname{cov}(\hat{\beta}_o, \hat{\beta}_2) + 2u_1u_2 \operatorname{cov}(\hat{\beta}_1, \hat{\beta}_2)] \, du_1 \, du_2 \quad .$$

Make the following transformation to polar coordinates.

Let

$$u_1 = r \cos \theta$$

$$u_2 = r \sin \theta$$

$$J = \begin{vmatrix} \partial u_1/\partial r & \partial u_1/\partial \theta \\ \partial u_2/\partial r & \partial u_2/\partial \theta \end{vmatrix} = \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix}$$

$$= r(\cos^2\theta + \sin^2\theta)$$

$$= r.$$

Then we have

$$\text{ave var } \hat{y}(u_1, u_2) = \sum_{i=1}^{n} K_i \int_o^{2\Pi} \int_{a_{i-1}}^{a_i} [\operatorname{var}\hat{\beta}_o + r^2\cos^2\theta(\operatorname{var}\hat{\beta}_1)$$

$$+ r^2\sin^2\theta(\operatorname{var}\hat{\beta}_2) + 2r\cos\theta(\operatorname{cov}(\hat{\beta}_o, \hat{\beta}_1)) + 2r\sin\theta(\operatorname{cov}(\hat{\beta}_o, \hat{\beta}_2))$$

$$+ 2r^2\cos\theta\sin\theta(\operatorname{cov}(\hat{\beta}_1, \hat{\beta}_2))] \; r \, dr \, d\theta$$

$$= \sum_{i=1}^{n} K_i \int_o^{2\Pi} [(r^2/2)(\operatorname{var}\hat{\beta}_o) + (r^4/4)\cos^2\theta(\operatorname{var}\hat{\beta}_1) + (r^4/4)\sin^2\theta(\operatorname{var}\hat{\beta}_2)$$

$$+ (2r^3/3)\cos\theta(\operatorname{cov}(\hat{\beta}_o, \hat{\beta}_1)) + (2r^3/3)\sin\theta(\operatorname{cov}(\hat{\beta}_o, \hat{\beta}_2))$$

$$+ (2r^4/4)\cos\theta\sin\theta(\operatorname{cov}(\hat{\beta}_1, \hat{\beta}_2))]_{a_{i-1}}^{a_i} \; d\theta$$

$$= \sum_{i=1}^{n} K_i \int_o^{2\Pi} [(1/2)\operatorname{var}\hat{\beta}_o(a_i^2 - a_{i-1}^2) + (1/4)\operatorname{var}\hat{\beta}_1(a_i^4 - a_{i-1}^4)\cos^2\theta$$

$$+ (1/4)\operatorname{var}\hat{\beta}_2(a_i^4 - a_{i-1}^4)\sin^2\theta + (2/3)\operatorname{cov}(\hat{\beta}_o, \hat{\beta}_1)(a_i^3 - a_{i-1}^3)\cos\theta$$

$$+ (2/3)\operatorname{cov}(\hat{\beta}_o, \hat{\beta}_2)(a_i^3 - a_{i-1}^3)\sin\theta + (1/2)\operatorname{cov}(\hat{\beta}_1, \hat{\beta}_2)(a_i^4 - a_{i-1}^4)\cos\theta\sin\theta] \, d\theta$$

$$= \sum_{i=1}^{n} K_i \, [\Pi(a_i^2 - a_{i-1}^2) \text{ var } \hat{\beta}_o + (\Pi/4)(a_i^4 - a_{i-1}^4)(\text{var } \hat{\beta}_1 + \text{var } \hat{\beta}_2)\,]$$

$$= \sum_{i=1}^{n} (M_i/\Pi(a_i^2 - a_{i-1}^2))[\,\Pi(a_i^2 - a_{i-1}^2)\text{var } \hat{\beta}_o + (\Pi/4)(a_i^2 - a_{i-1}^2)(a_i^2 + a_{i-1}^2)$$

$$(\text{var } \hat{\beta}_1 + \text{var } \hat{\beta}_2)\,]$$

$$\text{ave var } \hat{y}(u_1, u_2) = \text{var } \hat{\beta}_o \sum_{i=1}^{n} M_i + (1/4)(\text{var } \hat{\beta}_1 + \text{var } \hat{\beta}_2) \sum_{i=1}^{n} M_i(a_i^2 + a_{i-1}^2)$$

$$= \text{var } \hat{\beta}_o + (1/4)(\text{var } \hat{\beta}_1 + \text{var } \hat{\beta}_2) \sum_{i=1}^{n} M_i(a_i^2 + a_{i-1}^2)\,.$$

Thus, the proof is established.

Consider now an upper bound on the average variance of $\hat{y}(u_1, u_2)$ with respect to the distribution of the probability mass.

5.2 Upper Bound on the Average Variance of $\hat{y}(u_1, u_2)$ in the Square Region R and the Circular Region R_c

Theorem 5.3 For any division of R into subregions

$$R_i = [(u_1, u_2) \mid -a_i \leq u_j \leq a_i \,;\; j = 1, 2] - R_{i-1}$$

$(i = 1, 2, \ldots, n;\; R_o = \phi,\; a_o = 0)$, and for any distribution of the total probability mass to these subregions,

$$\max_{M_i} \text{ ave var } \hat{y}(u_1, u_2) < \text{var } \hat{\beta}_o + (2/3)(\text{var } \hat{\beta}_1 + \text{var } \hat{\beta}_2),$$

where M_i is the probability mass assigned to the region R_i and $\sum_{i=1}^{n} M_i = 1$.

Proof: By theorem 5.1, we have

$$\text{ave var } \hat{y}(u_1, u_2) = \text{var } \hat{\beta}_o + (1/3)(\text{var } \hat{\beta}_1 + \text{var } \hat{\beta}_2) \sum_{i=1}^{n} M_i(a_i^2 + a_{i-1}^2) .$$

In order to maximize the average var $\hat{y}(u_1, u_2)$, we must maximize $\sum_{i=1}^{n} M_i(a_i^2 + a_{i-1}^2)$ with respect to M_i. Thus, the proof will consist of showing that

$$\max_{\substack{0 \le M_i \le 1 \\ \Sigma M_i = 1}} \left[\sum_{i=1}^{n} M_i (a_i^2 + a_{i-1}^2) \right] < 2.$$

Since each $a_i \le 1$, we have

$$a_i^2 + a_{i-1}^2 < 2 \text{ for every } i = 1, 2, \ldots, n.$$

Let

$$b_K = \max_{1 \le i \le n} (a_i^2 + a_{i-1}^2),$$

then the maximum of $\sum_{i=1}^{n} M_i(a_i^2 + a_{i-1}^2)$ is attained by assigning probability mass equal to 1 to M_K and probability mass equal to 0 to each $M_i (i \neq K)$, since every convex combination of a set of numbers is less than or equal to the largest number in the set; that is,

$$\max_{0 \le M_i \le 1} \left[\sum_{i=1}^{n} M_i (a_i^2 + a_{i-1}^2) \right] \le b_K < 2.$$

Thus,

$$\max_{M_i} \text{ ave var } \hat{y}(u_1, u_2) < \text{var } \hat{\beta}_o + (2/3)(\text{var } \hat{\beta}_1 + \text{var } \hat{\beta}_2) .$$

Corollary 5.3.1 There exists no subdivision of the region R or distribution of the total probability mass to these subregions such that

$$\text{ave var } \hat{y}(u_1, u_2) = \text{var } \hat{\beta}_o + \text{var } \hat{\beta}_1 + \text{var } \hat{\beta}_2$$

$$\text{ave var } \hat{y}(u_1, u_2) = \text{trace } (X'X)^{-1}.$$

<u>Proof:</u> Assume

$$\text{ave var } \hat{y}(u_1, u_2) = \text{trace } (X'X)^{-1},$$

then by theorem 5.1, this implies that

$$(1/3) \sum_{i=1}^{n} M_i (a_i^2 + a_{i-1}^2) = 1$$

or

$$\sum_{i=1}^{n} M_i (a_i^2 + a_{i-1}^2) = 3.$$

But by theorem 5.3

$$\max_{0 \leq M_i \leq 1} \left[\sum_{i=1}^{n} M_i (a_i^2 + a_{i-1}^2) \right] < 2.$$

Hence, we always have

$$\text{ave var } \hat{y}(u_1, u_2) \neq \text{trace } (X'X)^{-1}.$$

<u>Theorem 5.4</u> For any division of the region R_c into subregions

$$R_i = [(u_1, u_2) \mid a_{i-1}^2 \leq u_1^2 + u_2^2 \leq a_i^2],$$

$(i = 1, 2, \ldots, n;\ a_o = 0)$, and for any distribution of the total probability mass to these subregions

$$\max_{M_i} \text{ ave var } \hat{y}(u_1, u_2) < \text{var } \hat{\beta}_o + (1/2)(\text{var } \hat{\beta}_1 + \text{var } \hat{\beta}_2),$$

where M_i denotes the probability mass assigned to R_i and $\Sigma M_i = 1$.

<u>Proof:</u> The proof of this theorem follows directly from the proof of theorem 5.3 after noting in theorem 5.2 that

$$\text{ave var } \hat{y}(u_1, u_2) = \text{var } \hat{\beta}_o + (1/4)(\text{var } \hat{\beta}_1 + \text{var } \hat{\beta}_2) \sum_{i=1}^{n} M_i (a_i^2 + a_{i-1}^2)$$

for the circular region R_c.

Corollary 5.4.1 There exists no subdivision of the region R_c or distribution of the total probability mass to these subregions such that

$$\text{ave var } \hat{y}(u_1, u_2) = \text{trace } (X'X)^{-1}.$$

Proof: The proof of this corollary follows from the proof of corollary 5.3.1.

Consider now the minimum average variance of $\hat{y}(u_1, u_2)$ with respect to the division of the region of interest and the distribution of the total probability mass to these subdivisions.

5.3 Lower Bound on the Average Variance of $\hat{y}(u_1, u_2)$ in the Square Region R and the Circular Region R_c

Theorem 5.5 For the square region R, there exists a division of R into subregions

$$R_i = [(u_1, u_2) \mid -a_i \leq u_j \leq a_i;\; j = 1, 2] - R_{i-1}$$

$(i = 1, 2, \ldots, n;\; R_o = 0)$, and a distribution of the total probability mass to these subregions such that

$$\min_{M_i} \text{ ave var } \hat{y}(u_1, u_2) = \text{var } \hat{\beta}_o,$$

where M_i denotes the probability mass assigned to R_i and $\sum_{i=1}^{n} M_i = 1$.

Proof: From theorem 5.1, we have

$$\text{ave var } \hat{y}(u_1, u_2) = \text{var } \hat{\beta}_o + (1/3)(\text{var } \hat{\beta}_1 + \text{var } \hat{\beta}_2) \sum_{i=1}^{n} M_i(a_i^2 + a_{i-1}^2).$$

The proof will consist of showing that there exists a division of R and a distribution of the probability mass to R such that

$$\sum_{i=1}^{n} M_i (a_i^2 + a_{i-1}^2) = 0.$$

Let

$$R_1 = [(u_1, u_2) \mid -\varepsilon \leq u_i \leq \varepsilon; \; i = 1, 2; \varepsilon > 0]$$

$$R_2 = R - R_1$$

then $a_o = 0$, $a_1 = \varepsilon$, and $a_2 = 1$.

Assign mass $M_1 = 1 - \varepsilon$ to R_1, and mass $M_2 = \varepsilon$ to R_2. Then,

$$\sum_{i=1}^{2} M_i(a_i^2 + a_{i-1}^2) = (1 - \varepsilon)(\varepsilon^2) + \varepsilon (1 + \varepsilon^2)$$

$$= \varepsilon(1 + \varepsilon) .$$

Thus,

$$\lim_{\varepsilon \to 0} \sum_{i=1}^{2} M_i(a_i^2 + a_{i-1}^2) = \lim_{\varepsilon \to 0} \varepsilon (1+\varepsilon)$$

$$= 0,$$

which implies

$$\min_{M_i} \text{ave var } \hat{y} (u_1, u_2) = \text{var } \hat{\beta}_o.$$

Theorem 5.6 For the circular region R_c, there exists a division of R_c into subregions

$$R_i = [(u_1, u_2) \mid a_{i-1}^2 \leq u_1^2 + u_2^2 \leq a_i^2] ,$$

and a distribution of the total probability mass to these subdivisions such that

$$\min_{M_i} \text{ ave var } \hat{y} (u_1, u_2) = \text{var } \hat{\beta}_o,$$

where M_i denotes the probability mass assigned to R_i and $\sum_{i=1}^{n} M_i = 1$.

Proof: From theorem 5.2, we have

$$\text{ave var } \hat{y}(u_1, u_2) = \text{var } \hat{\beta}_o + (1/4)(\text{var } \hat{\beta}_1 + \text{var } \hat{\beta}_2) \sum_{i=1}^{n} M_i(a_i^2 + a_{i-1}^2).$$

The proof follows from the proof of theorem 5.5.

Although we have established several properties of the average variance of $\hat{y}(u_1, u_2)$ for any distribution of the total probability mass to the region of interest, everything was done under the assumption that the true model was

$$y = \beta_o + \beta_1 x_1 + \beta_2 x_2 + \varepsilon .$$

However, similar results for more complicated models can be obtained rather easily.

CHAPTER VI

SUMMARY

Determining optimal experimental designs for particular assumed models and choosing an optimal model represent the essence of this study. Although optimal designs had already been determined by Kiefer (4) for a general polynomial model, the designs were only "within $O(N^{-1})$" of being optimal. Therefore, in Chapter III several different polynomial models were assumed, and using each assumed model, exact optimal designs were determined for each value of N. For the models which were assumed, the $\min_x \max_u \operatorname{var} \hat{y}(u)$ design always yielded a maximum variance of $\hat{y}(u)$, say M_1, that was the same for all multiples of N when N is odd and a maximum variance of $\hat{y}(u)$, say M_2, that was the same for all multiples of N when N is even. That is, for N even

$$\max_u \operatorname{var} \hat{y}(u) = 2/N,$$

and for N odd

$$\max_u \operatorname{var} \hat{y}(u) = 2/(N-1).$$

When N was even, the min max var $\hat{y}(u)$ design was always the same; namely, take N/2 points at $x = -1$ and N/2 points at $x = 1$. However, for N odd the min max var $\hat{y}(u)$ design was not always the same.

Since the bias function depended upon the true model, there were many designs which achieved $\min_x \max_u |\text{bias }(u)|$. Usually, for a given assumed model and a given true model, various multiples of N produced different $\min_x \max_u |\text{bias }(u)|$ designs.

Since only a small number of models were assumed, determining exact optimal designs for an expanded number of models would be worthy of future work.

Chapter II was devoted to the specific problem of trying to determine the $\min_x \max_u \text{var } \hat{y}(u)$ design, the $\min_x \text{ ave var } \hat{y}(u)$ design, and the $\min_x$ generalized var $\hat{y}(u)$ design, using the assumed model

$$y = \beta_o + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

in the region $R = [(x_1, x_2) | -1 \leq x_i \leq 1; i = 1, 2]$, with N equal to 3. It was determined that the variance of the estimated response at each of the design points is 1. Also determined was the fact that the maximum variance of the estimated response occurs at one or more corners of the square region R. However, there still remains the problem of showing that the 3 design points must be on the boundary of the square region R in order for the design to achieve $\min_x \max_u \text{var } \hat{y}(u)$. Under the assumption that the design points had to be on the boundary of R, an empirical investigation yielded a design that achieved $\min_x \max_u \text{var } \hat{y}(u)$, with

$$\max_u \text{ var } \hat{y}(u) = 1.42.$$

The design consisting of the 3 corners of the largest equilateral triangle inscribed in the square region R yielded a maximum variance of $\hat{y}(u)$ equal to 1.57. Thus, this design was rejected in favor of the design which produced a maximum variance of $\hat{y}(u)$

equal to 1.42.

Since there is always some sort of risk involved in choosing a model, an attempt was made in Chapter IV to try to determine which model should be fit in order to minimize some form of the bias. A restricted class C of models was assumed and for each of these models, the maximum biases were determined using the other models in C as the true model. For each of the assumed models, one of the true models had the largest maximum bias. These maximum biases were compared to determine the minimum one. The assumed model associated with this bias was chosen as the $\min_x \max_u$ model. Certain assumptions had to be made in order to compare the maximum biases. The first assumption was that we knew that the true model had a function value equal to $a_o + \varepsilon$ at $x = \ell$, where $-1 < \ell < 0$; $0 < \ell \leq 1$. The second assumption was that the coefficients in each assumed model could be expressed as a multiple of one another. Under the assumptions

1) $-1 < \ell < 0$; $0 < \ell \leq 1$

2) coefficients in assumed model are equal,

the details of determining the $\min_x \max_u$ model were shown. Also, under the assumptions

1) $0 < \ell \leq 1$

2) coefficients in assumed model can be expressed as a multiple of one another,

the details of determining the $\min_x \max_u$ model were worked out but not shown due to the large number of special cases created by ℓ and

K, the constant involved in expressing the coefficients of the assumed model as multiples of one another. The details were also worked out for all but a few cases when $-1 < \ell < 0$, $0 < \ell \leq 1$, and the coefficients in the assumed model were expressed as a multiple of one another. It appears feasible that future work could produce a more realistic $\min_x \max_u$ model by increasing the number of models in the class C and by relaxing the assumptions somewhat.

In chapter V several properties were determined concerning the average variance of the estimated response in the two-dimensional case. The model

$$y = \beta_o + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

was assumed. The first property established was the average variance of the estimated response over both a square region R and a circular region R_c for any distribution of the total probability mass to n sub-regions of R and R_c. Next, the $\min_x$ ave var $\hat{y}$ (u) with respect to design was determined for any division of R and for any distribution of the total probability mass to these subdivisions. It was shown that the $\min_x$ ave var $\hat{y}$ (u) is achieved by taking N/4 points at each corner of R. Also determined was an upper and lower bound on the average variance of $\hat{y}(u_1, u_2)$ with respect to the distribution of the probability mass. For the square region R

$$\max_{Mi} \text{ave var } \hat{y}(u) < \text{var } \hat{\beta}_o + (2/3)(\text{var } \hat{\beta}_1 + \text{var } \hat{\beta}_2)$$

$$\min_{Mi} \text{ave var } \hat{y}(u) = \text{var } \hat{\beta}_o ,$$

where Mi denotes the probability mass assigned to R_i. For the circular region R_c

$$\max_{Mi} \text{ ave var } \hat{y}(u) < \text{var } \beta_o + (1/2)(\text{var } \hat{\beta}_1 + \text{var } \hat{\beta}_2)$$

$$\min_{Mi} \text{ ave var } \hat{y}(u) = \text{var } \hat{\beta}_o .$$

Since all properties established were under the assumption that the true model was

$$y = \beta_o + \beta_1 x_1 + \beta_2 x_2 + \varepsilon,$$

future work could be devoted to establishing similar properties for a larger class of models.

Kiefer (4) indicates that for the two-dimensional case, the $\min_x \max_u \text{var } \hat{y}(u)$ design assigns measure α to each corner of the square region R, measure β to the midpoint of each of the 4 edges of R, and measure γ to the center of R, where

$$\alpha = 0.1458$$

$$\beta = 0.0802$$

$$\gamma = 0.0962.$$

Since it would be an impossible task to divide up the sample in this manner for reasonable sample size, an investigation was made to determine how to divide up the sample among the 4 corners of R in order to achieve $\min_x \max_u \text{var } \hat{y}(u)$, $\min_x \text{ ave var } \hat{y}(u)$, and $\min_x$ generalized var $\hat{y}(u)$. All of these optimality criteria are satisfied by the following designs. For sample size $N=4K (K = 1, 2, \ldots)$, K points should be at each corner of R. For sample size $N = 4K + 1$ $(K = 1, 2, \ldots)$, K points should be at any 3 corners of R, and $(K + 1)$ points should be at the remaining corner of R. For $N = 4K+2$ $(K = 1, 2, \ldots)$, one design indicates that K points should be at

$x = (-1, 1)$ and $x = (1, -1)$, and $(K + 1)$ points should be at $x=(-1, -1)$ and $x = (1, 1)$. For $N = 4K + 3$ $(K = 0, 1, 2, \ldots)$, $(K + 1)$ points should be at any 3 corners of R and K points should be at the remaining corner. Relaxing the restrictions somewhat, an attempt was made to determine optimal designs using the 4 corners of R, the midpoint of the edges of R, and the center of R as design points, but no results could be obtained.

BIBLIOGRAPHY

(1) Aitchison, J. "The construction of optimum designs for the the one-way classification analysis of variance." Journal of the Royal Statistical Society, XXIII (Series B), (1961), 352-67.

(2) Box, G. E. P. and Draper, Norman R. "A basis for the selection of a response surface design." Journal of the American Statistical Association, CIV (1959), 622-54.

(3) Folks, J. Leroy "Comparison of designs for exploration of response relationships." Doctorial dissertation (1958).

(4) Kiefer, J. "Optimum designs in regression problems II." Annals of Mathematical Statistics, XXXII (1961), 298-325.

(5) __________. "Two more criteria equivalent to D-optimality of designs." Annals of Mathematical Statistics, XXXIII (1962), 792-96.

(6) Kiefer, J. and Wolfowitz, J. "Optimum designs in regression problems." Annals of Mathematical Statistics, XXX (1959), 271-94.

(7) __________. "The equivalence of two extremum problems." Canadian Journal of Mathematics, XII (1960), 363-66.

(8) Tocher, K. D. "A note on the design problem." Biometrika, XXXIX (1952), 189.

VITA

Billy E. Gillett

Candidate for the Degree of

Doctor of Philosophy

Thesis: THE DETERMINATION OF SOME OPTIMAL EXPERIMENTAL DESIGNS

Major Field: Statistics (Mathematics)

Biographical:

Personal Data: Born in Guthrie, Oklahoma, June 7, 1935, the son of Vernon H. and Helen C. Gillett.

Education: Attended grade school at Guthrie and Crescent, Oklahoma; graduated from Crescent High School in 1953; received the Bachelor of Science degree from Oklahoma State University, Stillwater, Oklahoma, in May, 1961; received the Master of Science degree from Oklahoma State University in May, 1963; completed requirements for the Doctor of Philosophy degree in May, 1964.

Professional Experience: Entered the United States Army in 1953; released from active duty in 1956. Employed by Continental Oil Company, Ponca City, Oklahoma, from August, 1956, to September, 1957. Employed by Oklahoma State University from September, 1957, to September, 1961, in the following departments and capacities:

IBM operator, Data Processing, from September, 1957, to May, 1958.

Student Assistant, Computing Center, Department of Mathematics, from September, 1958, to September, 1961.

Employed by Pan American Oil Company, Oklahoma City, Oklahoma, as an IBM computer programmer during the summer of 1962. Employed by Oklahoma State University as a statistical assistant during the summer of 1963.