OUTLIERS AND ROBUST RESPONSE SURFACE DESIGNS

by

Mary Ann O'Gorman

Dissertation submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Statistics

APPROVED:

Raymond H. Myers, Chairman

Jesse C. Arnold

Klaus H. Hinkelmann

Jeffrey B. Birch

Marion R. Reynolds

December, 1984

Blacksburg, Virginia

OUTLIERS AND ROBUST RESPONSE SURFACE DESIGNS

by

Mary Ann O'Gorman

Committee Chairman: Raymond H. Myers

Statistics

(ABSTRACT)

A commonly occurring problem in response surface methodology is that of inconsistencies in the response variable. These inconsistencies, or maverick observations, are referred to here as outliers. Many models exist for describing these outliers. Two of these models, the mean shift and the variance inflation outlier models, are employed in this research.

Several criteria are developed for determining when the outlying observation is detrimental to the analysis. These criteria all lead to the same condition which is used to develop statistical tests of the null hypothesis that the outlier is not detrimental to the analysis. These results are extended to the multiple outlier case for both models.

The robustness of response surface designs is also investigated. Robustness to outliers, missing data and errors in control are examined for first order models. The orthogonal designs with large second moments, such as the 2^k factorial designs, are optimal in all three cases.

In the second order case, robustness to outliers and to missing data are examined. Optimal design parameters are obtained by computer for the central composite, Box-Behnken, hybrid, small composite and equiradial designs. Similar results are seen for both robustness to outliers and to missing data. The central composite turns out to be the optimal design type and of the two economical design types the small composite is preferred to the hybrid.

ACKNOWLEDGMENTS

The author wishes to express her sincere appreciation to the following:

Mrs. Theresa O'Gorman, her mother, and the rest of her family for their understanding and support.

Dr. Raymond H. Myers for directing this research and especially for his advice, concern and encouragement.

Mrs. Sharon Myers for her assistance academically and more importantly for her friendship.

Drs. Jesse C. Arnold, Jeffrey B. Birch, Klaus H. Hinkelmann and Marion R. Reynolds for graciously serving as committee members.

Mrs. Faye Roop for her skillful typing of this manuscript.

TABLE OF CONTENTS

TABLE OF CONTENTS (cont.)

TABLE OF CONTENTS (cont.)

TABLE OF CONTENTS (cont.)

TABLE OF CONTENTS (cont.)

LIST OF TABLES

LIST OF TABLES (cont.)

Table | Page

LIST OF FIGURES

I. INTRODUCTION

Response Surface Methodology (RSM) is the term applied to a group of mathematical and statistical techniques used to solve particular types of problems often encountered in the sciences and engineering. In these problems the experimenter is interested in the *response* of some system which is influenced by several inputs or *independent variables*. The response and the independent variables are usually measured on a continuous scale with the independent variables being controlled by the experimenter. Of primary importance to the experimenter are prediction of the response and optimization of the system. Optimization is finding combinations of the independent variables which give the optimum response. In some cases the experimenter may also be interested in selecting the independent variables which best describe the system.

It is usually assumed that the system can be approximated by a low order polynomial in the independent variables. Normally, a first or second order model is used. The first order model has the form

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon \ ,$$

and the second order model is of the form

$$y = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k + \beta_{11} X_1^2 + \dots + \beta_{kk} X_k^2$$

$$+ \beta_{12} X_1 X_2 + \dots + \beta_{k-1,k} X_{k-1} X_k + \varepsilon \ .$$

These models fall in the category of the general linear model,

$$\underline{y} = X\underline{\beta} + \underline{\varepsilon} \ ,$$

where $\underline{y}$ is an $N \times 1$ vector of responses, X is the $N \times p$ matrix of independent variables which takes into account the form of the model, $\underline{\beta}$ is a $p \times 1$ vector of unknown constants to be estimated and $\underline{\varepsilon}$ is an $N \times 1$ vector of unobservable random error terms.

The response surface models presented are special cases of linear regression models. The important distinction is that the experimenter has control of the independent variables in the response surface model. Thus, he can select the levels of these variables to be used in the experiment.

The control of the independent variables by the experimenter does not, however, prevent the occurrence of unusual, or maverick data points from these experiments. The term *outlier* has been used to refer to any observation which does not seem to be consistent with the bulk of the data. An observation can only be detected as an outlier, then, with respect to some baseline model used to describe the data. A more rigorous definition of what is meant by an outlier will be given later.

Many techniques exist in the literature for identifying outliers. Identification of a point as an outlier may lead to its rejection from the analysis or replacement of the model being used with another which is more appropriate for the data, or the experimenter and collaborating scientists may decide to take no action at all in regard to the outlier. The goal of this research on outliers in RSM is to determine when it is appropriate to delete the observation from the analysis and to give diagnostic tools to aid in making this decision in practical situations.

Outliers can be modeled in a number of ways. In Chapter III a mean shift model and a variance inflation model will be used to describe the presence of an outlier. For more details on these models refer to Section 2.2. These models will be used in developing criteria for determining when an outlier is detrimental to the analysis and should be discarded. Statistical test procedures based on these criteria will be developed which are intended for use as an aid to the experimenter in deciding whether or not to reject an observation. The criteria and test procedures will then be generalized to the case of multiple outliers for both outlier models.

An important consideration in response surface work is the selection of the levels of the independent variables to be used in the experiment. The combinations of levels of the independent variables used is called the *experimental design*. The experimenter should choose the levels in order to achieve certain design properties that are desirable for his purposes. It is common to choose designs which provide a particular pattern for the variances of $\hat{\underline{y}}$.

Another design philosophy is that the design should not only have desirable properties when the usual least squares assumptions are met, but should also perform well when one or more of these assumptions are violated. Designs selected in this manner are called *robust designs*. For example, Box and Draper (1959) discuss designs which are robust to violations in the assumption that the model is correct. In particular, they consider the case where the model that is fit does not include all of the relevant terms. If the model is misspecified in this way then $E(\varepsilon)$ is nonzero and bias is induced into the estimates. The designs Box and Draper recommend minimize the effect of this bias on the estimated response. (Chapter II gives more details on model misspecification.)

This research will consider several types of design robustness. Specifically, designs which are robust to:

(i) Outliers. Designs will be found which are insensitive to the occurrence of an outlier in the data.

(ii) Missing data. Losing a data point alters the intended properties of the design. Designs will be found which minimize the effect of losing a single data point.

(iii) Errors in control. Often the experimenter is unable to attain the exact level of the independent variables chosen for the design. The effects of these errors will be investigated in an attempt to find designs which are robust to them.

The criteria used for evaluating these types of design robustness will be introduced as each type is discussed. The necessary background for understanding each problem is presented in Chapter II.

Designs will be investigated for both first and second order models. Theoretical results will be obtained in the first order case. Empirical results obtained by computer will be presented for several classes of second order designs. The second order designs to be considered are the central composite, small composite, Box-

Behnken, hybrid and equiradial designs. These designs will be described in detail in Chapter II. In the final chapter a summary of the results will be given along with practical recommendations for the use of these designs.

II. LITERATURE REVIEW

2.1 Response Surface Methodology

Recently response surface methodology has become more and more popular in industrial applications, as well as in fields such as chemistry, biology, nutrition and engineering. Researchers are becoming more interested in studying the relationships between variables in a system and in being able to optimize the system. The techniques of RSM are frequently used for these purposes.

A brief review of response surface methodology will be given in the following sections. The necessary background material will be given in the next section along with a brief review of the relevant RSM literature. Section 2.1.2 will give some important response surface design considerations.

2.1.1 Response Surface Methodology Review

Response surface methodology is the term which refers to a group of techniques used to explore the relationships between variables in some region of interest, R. In particular, it is desirable to find conditions for the variables which lead to some type of optimality. These techniques were originally presented in a paper by Box and Wilson (1951). Davies (1956) devoted a chapter in a book on the design and analysis of industrial experiments to the determination of optimum conditions. In that chapter Davies presents an exposition on the techniques of Box and Wilson, along with new designs which allowed the search for the optimum conditions to be done in an efficient, economical manner. Hill and Hunter (1966) give an overview of the RSM techniques in use up to that time. The text by Myers (1976) provides a comprehensive account of the RSM techniques as well as some of the theoretical developments.

In RSM the experimenter is interested in investigating a system which can be represented as follows:

$$\eta = f(\xi_1, \xi_2, \ldots, \xi_k) \quad . \tag{2.1}$$

η is the true response of the system, $\xi_1, \ldots, \xi_k$ are the inputs to the system (which are controlled by the experimenter) and f is the function relating the input and output variables. Note that η is the true response which would be observed in the absence of experimental error. The observed response, which is measured with error, will be denoted by y.

In experimental situations the function, f, is unknown and therefore must be approximated, usually by a low order polynomial. Rather than expressing the polynomial in terms of $\xi_1, \ldots, \xi_k$ we will instead express it in terms of the variables $X_1, \ldots, X_k$, called the *design variables*. The design variables are obtained from the natural variables, i.e., the ξ's, by linear transformations. These transformations amount to a location and scale change for each of the variables. Generally, we center each variable at 0, and scale each to the interval (-1,1). The model, then, is expressed as a polynomial in $X_1, \ldots, X_k$. Usually a first order or second order model is used. This research will deal exclusively with first and second order models, although many other polynomial models could be considered. The first order model is given by

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i X_i + \varepsilon \tag{2.2}$$

and the second order model by

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i X_i + \sum_{i=1}^{k} \beta_{ii} X_i^2 + \sum_{i<j}^{k-1} \sum^{k} \beta_{ij} X_i X_j + \varepsilon \ . \tag{2.3}$$

In matrix notation these models are written:

$$\underline{y} = X\underline{\beta} + \underline{\varepsilon} \quad . \tag{2.4}$$

In this notation $\underline{y}$ is an $N \times 1$ vector of responses, X is an $N \times p$ matrix of independent variables taking into account the form of the model, $\underline{\beta}$ is the $p \times 1$ vector of coefficients to be estimated and $\underline{\varepsilon}$ is the $N \times 1$ vector of random error terms. The usual assumptions made on the ε's are that they are i.i.d. with $E(\underline{\varepsilon}) = \underline{0}$ and $Var(\underline{\varepsilon}) = \sigma^2 I$.

The techniques of RSM are based on least squares estimation of the coefficients. There are many other estimation methods available (see Draper and Smith (1981)), but only least squares estimation will be dealt with in this research. The least squares estimator of $\underline{\beta}$ is given by

$$\underline{\hat{\beta}} = (X'X)^{-1}X'\underline{y} \tag{2.5}$$

and the estimated responses given by

$$\hat{\underline{y}} = X\hat{\underline{\beta}} = X(X'X)^{-1}X'\underline{y} \quad . \tag{2.6}$$

Under the assumptions stated above $\hat{\underline{\beta}}$ has the following properties:

$$E(\hat{\underline{\beta}}) = \underline{\beta} \tag{2.7}$$

$$\mathrm{Var}(\hat{\underline{\beta}}) = \sigma^2(X'X)^{-1} \quad . \tag{2.8}$$

The properties of $\hat{\underline{y}}$ are

$$E(\hat{\underline{y}}) = X\underline{\beta} \tag{2.9}$$

$$\mathrm{Var}(\hat{\underline{y}}) = \sigma^2 X(X'X)^{-1}X' \quad . \tag{2.10}$$

The variance of $\hat{y}$ is referred to as the prediction variance. If $\underline{x}$ is an arbitrary point in the design region then the estimated response at $\underline{x}$ is given by

$$\hat{y} = \underline{x}'\hat{\underline{\beta}} \tag{2.11}$$

and the prediction variance is

$$\mathrm{Var}(\hat{y}) = \sigma^2\underline{x}'(X'X)^{-1}\underline{x} \quad . \tag{2.12}$$

2.1.2 Design Considerations

Since one of the primary goals of the experimenter is estimation of the response, he will be concerned with the quality of the predicted response, $\hat{y}$. With the usual assumptions stated above the quality of this estimate is reflected in the prediction variance, which is determined by the experimental design chosen. This can be seen from (2.10). Thus, the experiment should be well planned and a design selected which will have desirable properties and allow for good estimation of the response.

Some of the most commonly used design properties are those dealing with the prediction variance, such as *rotatability, uniform precision* and *minimum integrated prediction variance*. These properties will be discussed in detail, but first it will be helpful to introduce the notion of design moments (see Myers (1976)), which will be used to establish the design properties.

The design moments of interest are given in the moment matrix, $N^{-1}(X'X)$, where N is the number of runs in the experiment. In comparing this matrix with (2.8) we see that the variances of the coefficients are determined by these design moments. For first order designs the moment matrix involves first order moments, or simply *first moments* and *second moments*. The first moments are given by

$$[i] = \frac{1}{N} \sum_{u=1}^{N} X_{iu} \quad . \tag{2.13}$$

The second pure moments and second mixed moments are given, respectively, by

$$[ii] = \frac{1}{N} \sum_{u=1}^{N} X_{iu}^2 \tag{2.14}$$

and

$$[ij] = \frac{1}{N} \sum_{u=1}^{N} X_{iu} X_{ju} \quad . \tag{2.15}$$

The second order design moment matrix will contain moments through order four. A similar notation is used to refer to third and fourth moments. For example,

$$[iii] = \frac{1}{N} \sum_{u=1}^{N} X_{iu}^{3} \quad ,$$

$$[iiii] = \frac{1}{N} \sum_{u=1}^{N} X_{iu}^{4} \quad ,$$

$$[iiij] = \frac{1}{N} \sum_{u=1}^{N} X_{iu}^{3} X_{ju} \quad .$$

The design properties mentioned above can now be defined and discussed in terms of the design moments. Let

$$\underline{x}' = (1, x_1, x_2, \ldots, x_k) \tag{2.16}$$

be a point in the design region, R, and $\hat{y}(\underline{x})$ be the predicted response at the point $\underline{x}$. Let ρ be the distance of the point $\underline{x}$ from the center of the design. With the variables centered at zero,

$$\rho = \left(\sum_{i=1}^{k} x_i^2 \right)^{1/2} \quad . \tag{2.17}$$

A design is said to be rotatable if the variance of $\hat{y}(\underline{x})$ is a function only of ρ and not of the direction of $\underline{x}$ from the center of the design. Then the contours of constant prediction variance will be spheres, or hyperspheres. So, two points equidistant from the center will have the same prediction variances.

The moment conditions for achieving a rotatable design, as derived by Box and Hunter (1957) differ for first and second order designs. For first order designs the conditions which yield rotatability are as follows:

1) [i] = 0 for all i

2) [ij] = 0 $\quad i \neq j$

and (2.18)

3) All [ii] must be equal .

Notice that first order rotatable designs are also orthogonal. For second order designs the conditions are

1) All odd moments through order four must be zero

2) All [ii] are equal

and (2.19)

3) [iiii] = 3[iijj] $\quad i \neq j$.

Another design property presented in Box and Hunter (1957) referred to as uniform precision results in equal prediction variances for all points within the unit sphere, i.e., all points for which $\rho \le 1$. This property is intuitively appealing since in most cases the experimenter has little or no knowledge a priori of the locations within the design region at which he will want to predict well. Thus, from this point of view all points are equally important and should be estimated with equal precision. Uniform precision is determined by the mixed fourth moment, [iijj]. Box and Hunter give values for [iijj] which give uniform precision designs. Designs can be found which are both rotatable and uniform precision.

The final design property to be discussed here is based on a criterion presented by Box and Draper (1959). It is desirable to use a design whose average prediction variance is a minimum. The average is obtained by integrating the prediction variance over the region of interest, R. Box and Draper standardize the integrated variance to account for differences in sample size and error variance. The integrated prediction variance is then given by

$$V = \frac{NK}{\sigma^2} \int_R \mathrm{Var}(\hat{y})\, d\underline{x} \qquad (2.20)$$

where, $d\underline{x}$ denotes $dx_1, dx_2, \ldots, dx_k$ and,

$$K^{-1} = \int_R d\underline{x} \quad . \qquad (2.21)$$

From (2.12) it can be seen that

$$V = NK \int_R \underline{x}'(X'X)^{-1}\underline{x}\, d\underline{x} \quad . \qquad (2.22)$$

It is easiest to compute the integrated variance indirectly by using *region moments* (see Myers (1976)). In general, define the region moment matrix to be

$$\mu = K \int_R \underline{x}\underline{x}'d\underline{x} \qquad (2.23)$$

where, for first order designs $\underline{x}$ is defined as in (2.16) and for second order designs $\underline{x}$ is defined as

$$\underline{x}' = (1,x_1,\ldots,x_k,x_1^2,\ldots,x_k^2,x_1x_2,\ldots,x_{k-1}x_k) \quad . \qquad (2.24)$$

So, a general region moment looks like

$$K \int_R x_1^{\delta_1} x_2^{\delta_2} \ldots x_k^{\delta_k} d\underline{x} \quad . \qquad (2.25)$$

Now it can be seen that

$$V = N \, \mathrm{Trace}(\mu(X'X)^{-1}) \ . \qquad (2.26)$$

In practice usually designs are not found which minimize V, rather V is used as a criterion for comparing designs. An experimenter would select a design for which V was smallest when compared with the competing designs.

The design properties discussed here are only a few of the possible properties to be attained by a design. In practice the experimenter will select one or two properties which are important to him based on the goals of his research and then select a design having these properties. It is generally not possible to achieve all of the properties one would like to have with a particular design and thus, some compromise must be made.

2.2 Outliers

In RSM, as in any situation where data is observed with error, there is always a possibility that discordant observations will appear. In this paper the only type of discordant observations to be addressed are those which result from inconsistency in the response variable. These observations will be referred to as outliers.

Many models exist for describing outlying observations. In this section two such models will be presented, the mean shift model and the variance inflation model. These two models will be dealt with exclusively in this research. The effects of the outlier on least squares estimation will be described and a brief discussion of the basic approaches to dealing with outliers will be presented.

2.2.1 Mean Shift Outlier Model

The mean shift outlier model is defined by Cook and Weisberg (1982) as the following:

$$\underline{y} = X\underline{\beta} + \underline{d}_i\phi + \underline{\varepsilon} \quad , \tag{2.27}$$

where $\underline{d}_i$ is an index vector with a 1 in the i^{th} position and zeros elsewhere, and ϕ is the amount of the mean shift. Therefore, if ϕ is nonzero the i^{th} point is an outlier. The usual assumptions are made on $\underline{\varepsilon}$ and least squares estimation will be employed.

The outlier in this case induces bias into the least squares estimator, $\hat{\underline{\beta}}$, and thus into the estimated response, $\hat{\underline{y}}$, however the variances remain the same as in (2.8) and (2.10). The biases are

$$\text{Bias}(\underline{\hat{\beta}}) = (X'X)^{-1}\underline{x}_i\phi \tag{2.28}$$

$$\text{Bias}(\hat{\underline{y}}) = X(X'X)^{-1}\underline{x}_i\phi \quad . \tag{2.29}$$

The effect of the experimental design on the biases is seen by the role of $(X'X)^{-1}$.

To gain further insight into the effects of the outlier, consider the bias in prediction at the design points due to the outlier occurring at the i^{th} point. From (2.28) it can be seen that

$$\text{Bias}(\hat{y}_i) = \underline{x}_i'(X'X)^{-1}\underline{x}_i\phi \tag{2.30}$$

and

$$\text{Bias}(\hat{y}_j) = \underline{x}_j'(X'X)^{-1}\underline{x}_i\phi \qquad j \neq i \; . \tag{2.31}$$

From this we see that the bias in prediction at the point where the outlier occurs is a function of the i^{th} diagonal element of the following matrix:

$$H = X(X'X)^{-1}X' \; . \tag{2.32}$$

Also notice that the biases at the other design points are functions of the off-diagonal elements of H. The elements of H will be denoted by h_{ij}.

$$h_{ij} = \underline{x}_i'(X'X)^{-1}\underline{x}_j = h_{ji} \; . \tag{2.33}$$

If $i = j$ then we have the diagonal elements of H. Hoaglin and Welsch (1978) refer to H as the hat matrix and the diagonal elements of H as the hat diagonals. They discuss the importance of the hat matrix in regression which can be seen from (2.6) and (2.10) where

$$\hat{\underline{y}} = X(X'X)^{-1}X'\underline{y} = H\underline{y} \tag{2.34}$$

and

$$\mathrm{Var}(\hat{\underline{y}}) = \sigma^2 X(X'X)^{-1}X' = \sigma^2 H \ . \tag{2.35}$$

The role of the hat matrix in identifying "high leverage points" (points exerting a great deal of influence on $\hat{y}$) is discussed in Hoaglin and Welsch (1978) and Belsley, Kuh and Welsch (1980).

Several properties of the hat matrix and hat diagonals which will be useful in this reseach are given now without proof. (For details see Hoaglin and Welsch (1978).)

1) H is symmetric and idempotent.

2) $0 \leq h_{ii} \leq 1$.

3) $\sum_{i=1}^{N} h_{ii} = p$ (the number of parameters to be estimated).

The hat diagonal for a particular point is a measure of the standardized distance from the point to the center of the data. So if h_{ii} is large (h_{ii} near 1), x_i is far from the center thus indicating that x_i is a high leverage point. The hat matrix (and hat diagonals in particular) will play a key role in determining the severity of an outlying observation.

2.2.2 Variance Inflation Model

A second model for describing an outlier is a variance inflation model described by Cook, Holschuh and Weisberg (1982). In this case the model is that given in (2.4), but now the assumptions on the $\underline{\varepsilon}$ differ. We still assume that $E(\underline{\varepsilon}) = \underline{0}$ and that the elements of $\underline{\varepsilon}$ are independent of each other. Now, however, the variance structure for $\underline{\varepsilon}$ is somewhat altered. Instead of having constant variance, σ^2, the variance-covariance matrix is given by

$$\mathrm{Var}(\underline{\varepsilon}) = \begin{bmatrix} \sigma^2 & & & & & 0 \\ & \sigma^2 & & & & \\ & & \ddots & & & \\ & & & \sigma^2+\sigma^2_\Delta & & \\ & 0 & & & \ddots & \\ & & & & & \sigma^2 \end{bmatrix} = V \; . \tag{2.36}$$

That is, all of the errors have variance, σ^2 except for the i^{th} error which has variance that is inflated by an amount σ^2_Δ.

The least squares estimators, $\hat{\underline{\beta}}$ and $\hat{\underline{y}}$, will still be used to estimate the coefficients and responses. The properties of these estimators under the variance inflation model are as follows:

$$E(\hat{\underline{\beta}}) = \underline{\beta} \tag{2.37}$$

$$\mathrm{Var}(\hat{\underline{\beta}}) = \sigma^2(X'X)^{-1} + \sigma^2_\Delta(X'X)^{-1}\underline{x}_i\underline{x}_i'(X'X)^{-1} \tag{2.38}$$

$$E(\hat{\underline{y}}) = X\underline{\beta} \tag{2.39}$$

$$\mathrm{Var}(\hat{\underline{y}}) = \sigma^2 X(X'X)^{-1}X' + \sigma_\Delta^2 X(X'X)^{-1}\underline{x}_i\underline{x}_i'(X'X)^{-1}X' \ . \tag{2.40}$$

Therefore,

$$\mathrm{Var}(\hat{y}_i) = \sigma^2 h_{ii} + \sigma_\Delta^2 h_{ii}^2 \tag{2.41}$$

and

$$\mathrm{Var}(\hat{y}_j) = \sigma^2 h_{jj} + \sigma_\Delta^2 h_{ij}^2 \ . \tag{2.42}$$

2.2.3 How to Treat Outliers

In this research the outlier problem will be addressed in two ways. One is to develop techniques for identifying potentially damaging outliers. Once an outlier has been identified as possibly being harmful to the analysis it will be up to the analyst and scientists to decide whether or not to eliminate the point. The second approach is to employ experimental designs which are insensitive to outliers. Then if outliers should exist their impact on the analysis will be minimized. These designs will be discussed in the next section.

2.3 Robust Design

In this section we will deal with the concept of robust design and present results that are necessary to this research. Robustness in general refers to an insensitivity of a procedure to violations in assumptions, or to nonstandard conditions. Robustness can be achieved in one of two ways. Either an estimation technique can be chosen which provides this insensitivity or, in the context of RSM, an experimental design can be selected which will give robustness. Since we will be using least squares estimation we will look for designs which provide some type of robustness.

The following types of design robustness will be discussed in the next two sections:

1) Robustness to model misspecifications

2) Robustness to outliers.

2.3.1 Robustness to Model Misspecification

The original work on robust designs in RSM appeared in a paper by Box and Draper (1959). Their work dealt with designs which are robust to the assumption that the model is correct, i.e., that $E(\underline{\varepsilon}) = \underline{0}$. They deal with a specific type of model misspecification in which the model that is fit is of order d and the "true model" is of order $d+1$. The estimated coefficients of the fitted model are biased due to the misspecification. An integrated mean square error (MSE) criterion is used to find designs which are insensitive to this type of misspecification. Three cases are considered: minimization of bias only, minimization of variance only and minimization of integrated MSE.

In the case where the fitted model is first order, as in (2.2), and the true model is second order, as in (2.3), the moment conditions for a spherical design region are summarized as follows:

1) Minimum bias:

a) $[ij] = 0$

b) All third moments equal zero (2.43)

c) $[ii] = \frac{1}{k+2}$

Notice that these conditions give a symmetric, orthogonal design.

2) Minimum variance:

a) $[ij] = 0$ (2.44)

b) $[ii]$ as large as possible

This design will also be orthogonal.

3) Minimum integrated MSE:

It turns out that an optimal design in terms of minimum integrated MSE can not be found without knowledge of the unfit coefficients. However, the point is made that a small amount of bias can greatly effect the least squares estimates, and that bias may be more of a problem than variance. In fact, Box and Draper show that in cases where variance and bias are contributing equally to the MSE, the optimal design is very close to the minimum bias design. Thus, the recommendation when both variance and bias are involved is to compromise. Select a symmetric, orthogonal design with second moments between the second moments for the minimum bias design and the minimum variance design, but closer to the minimum bias design.

Similar results are given also for a spherical region for the case where the fitted model is second order and the "true model" is third order. The conditions for the minimum bias designs are

$$
\begin{aligned}
&\text{a)}\quad \text{All odd moments through order five are zero} \\
&\text{b)}\quad [ii] = \frac{1}{k+2} \\
&\text{c)}\quad [iijj] = \frac{1}{(k+2)(k+4)} \\
&\text{d)}\quad [iiii] = \frac{3}{(k+2)(k+4)} \ .
\end{aligned}
\tag{2.45}
$$

Notice that these conditions result in a rotatable design.

2.3.2 Robustness to Outliers

The first research on designs which are robust to outliers was also done by Box and Draper (1975). They define criteria for measuring insensitivity of the design to outliers in terms of the hat diagonals, as defined in (2.33). They indicate that the hat diagonals should be made equal in order to achieve robustness to outliers. In the first order case this can be achieved. For example the 2^k factorials and the first order orthogonal fractions have equal hat diagonals. These designs will be discussed in more detail in Section 2.4.1. In the second order case, however, equal hat diagonals can not be achieved and thus, they recommend minimization of the sample variance of the hat diagonals. They give results for a specific class of second order designs called the central composite design, which will be discussed in Section 2.4.2.

Draper and Herzberg (1979) use an average integrated MSE criterion to find outlier robust designs in the presence or absence of model misspecification. They do not recommend the use of this criterion alone in evaluating designs. The criterion is set forth as an additional criterion to help evaluate competing designs in terms of the fourteen design properties listed in Box and Draper (1975).

2.4 Design Classes to be Considered

This section will describe the types of designs to be used in this research. Both first and second order designs will be described. In the first order case the 2^k factorial and fractional factorial designs and Plackett-Burman designs will be investigated. The following second order designs will be discussed in the following sections:

1) Central composite
2) Box-Behnken
3) Hybrid
4) Small composite
5) Equiradial.

2.4.1 2^k Factorial Plans

The 2^k factorial arrays involve variables at 2 levels (usually +1 and -1). For example, the following is a 2^3 design:

$$\begin{array}{c} \begin{array}{ccc} x_1 & x_2 & x_3 \end{array} \\ \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & -1 & -1 \\ -1 & 1 & 1 \\ -1 & 1 & -1 \\ -1 & -1 & 1 \\ -1 & -1 & -1 \end{bmatrix} \end{array} . \qquad (2.48)$$

A fractional factorial array is obtained when not all of the 2^k combinations of design levels are run. In this research only fractional factorials which are first order orthogonal and of Resolution III or higher will be considered.

The first order orthogonality of these designs makes them optimal in many ways, some of which have already been noted. The 2^k factorials will also turn out to be optimal in terms of the criterion developed in this research. For details on the 2^k factorials and fractional factorials see Box, Hunter and Hunter (1978).

2.4.2 Plackett-Burman Design

Plackett-Burman designs are two level orthogonal designs which are saturated or nearly so, that is, designs in which the number of observations is nearly equal to the number of parameters being fit. An example will illustrate the construction of these designs. For a sample size of 8 Plackett and Burman (1946) give the following levels for the first seven levels of the first variable

$$+1 \quad +1 \quad +1 \quad -1 \quad +1 \quad -1 \quad -1$$

The levels for other variables (up to seven) are obtained by taking cyclic permutations of the levels of the first variable. The eighth run is taken with all variables at the -1 setting. Thus, the eight run design for four variables is as follows:

$$\begin{array}{cccc} x_1 & x_2 & x_3 & x_4 \end{array}$$
$$\begin{bmatrix} +1 & +1 & +1 & -1 \\ +1 & +1 & -1 & +1 \\ +1 & -1 & +1 & -1 \\ -1 & +1 & -1 & -1 \\ +1 & -1 & -1 & +1 \\ -1 & -1 & +1 & +1 \\ -1 & +1 & +1 & +1 \\ -1 & -1 & -1 & -1 \end{bmatrix} .$$

Since these designs are first order orthogonal they are also rotatable. These designs will be used later in the construction of small composite designs.

2.4.3 Central Composite Design

A central composite design (ccd) (see Myers (1976)) is obtained by augmenting a 2^k factorial or fractional factorial with the following:

$$\begin{array}{ccccc} x_1 & x_2 & x_3 & \cdots & x_k \end{array}$$
$$\begin{bmatrix} \alpha & 0 & 0 & \cdots & 0 \\ -\alpha & 0 & 0 & \cdots & 0 \\ 0 & \alpha & 0 & \cdots & 0 \\ 0 & -\alpha & 0 & \cdots & 0 \\ 0 & 0 & \alpha & \cdots & 0 \\ 0 & 0 & -\alpha & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \alpha \\ 0 & 0 & 0 & \cdots & -\alpha \end{bmatrix} . \qquad (2.49)$$

These points are called axial points and the value of α is referred to as the *axial value*. The value of α is chosen by the experimenter to achieve certain design properties. Center points, that is observations taken at the center of the design, may also be included. The number of center points is also left to the discretion of the experimenter. The choice of α and the number of center points makes this a very flexible class of designs. We will take advantage of this flexibility to obtain robustness.

An example of a 3 factor ccd is given below.

$$\begin{array}{c} \begin{array}{ccc} x_1 & x_2 & x_3 \end{array} \\ \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & -1 & -1 \\ -1 & 1 & 1 \\ -1 & 1 & -1 \\ -1 & -1 & 1 \\ -1 & -1 & -1 \\ 1.682 & 0 & 0 \\ -1.682 & 0 & 0 \\ 0 & 1.682 & 0 \\ 0 & -1.682 & 0 \\ 0 & 0 & 1.682 \\ 0 & 0 & -1.682 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \end{array} . \qquad (2.50)$$

In this example α = 1.682 makes the design rotatable.

2.4.4 Box-Behnken Design

These designs are formed by combining 2^k factorials with the notion of balanced incomplete block structures. (For details see Box and Behnken (1960).) A 2^2 factorial is formed for each pair of variables and the level of the remaining variables is zero. There will be $\binom{k}{2} \cdot 4$ points in addition to center points. An example will help to illustrate the construction of these designs. The three variable Box-Behnken design is

$$\begin{array}{ccc} x_1 & x_2 & x_3 \end{array}$$
$$\begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \\ -1 & 1 & 0 \\ -1 & -1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & -1 \\ -1 & 0 & 1 \\ -1 & 0 & -1 \\ 0 & 1 & 1 \\ 0 & 1 & -1 \\ 0 & -1 & 1 \\ 0 & -1 & -1 \\ \underline{0} & \underline{0} & \underline{0} \end{bmatrix} . \qquad (2.51)$$

For $k \geq 6$, 2^3 factorials are formed for combinations of three variables. For example, the six variable design is

$$\begin{array}{c} \begin{array}{cccccc} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 \end{array} \\ \begin{bmatrix} \pm1 & \pm1 & 0 & \pm1 & 0 & 0 \\ 0 & \pm1 & \pm1 & 0 & \pm1 & 0 \\ 0 & 0 & \pm1 & \pm1 & 0 & \pm1 \\ \pm1 & 0 & 0 & \pm1 & \pm1 & 0 \\ 0 & \pm1 & 0 & 0 & \pm1 & \pm1 \\ \pm1 & 0 & \pm1 & 0 & 0 & \pm1 \\ \underline{0} & \underline{0} & \underline{0} & \underline{0} & \underline{0} & \underline{0} \end{bmatrix} \end{array} .$$

The notation $(\pm1,\pm1,\pm1)$ refers to the full 2^3 factorial. Notice that center points are added to the basic design.

In this research the optimum number of center points for achieving certain types of robustness will be determined.

2.4.5 Hybrid Design

Roquemore (1976) develops hybrid designs in order to obtain designs which have nearly the minimum number of points for estimating the parameters in the model. For k variables the designs are constructed as follows:

1) At one level of the k^{th} variable, construct a $2^{(k-1)}$ factorial in the remaining variables.

2) At a second level of the k^{th} variable, construct the axial points for a (k-1) ccd as in (2.49).

3) Construct two axial points for the k^{th} variable.

4) Possibly include a number of center points.

The hybrid design in three variables is

$$\begin{array}{ccc} x_1 & x_2 & x_3 \\ \end{array}$$

$$\begin{bmatrix} 0 & 0 & \alpha_1 \\ 0 & 0 & \alpha_2 \\ 1 & 1 & \alpha_3 \\ 1 & -1 & " \\ -1 & 1 & " \\ -1 & -1 & " \\ \alpha & 0 & \alpha_4 \\ -\alpha & 0 & " \\ 0 & \alpha & " \\ 0 & -\alpha & " \\ \underline{0} & \underline{0} & \underline{0} \end{bmatrix} . \qquad (2.53)$$

So, α, α_1, α_2, α_3, α_4 and the number of center points are chosen by the experimenter. Notice the ccd in the first two variables. The most attractive feature of these designs is that they are extremely economical, especially for large k.

2.4.6 Small Composite Design

These designs are very similar to the ccd's described in Section 2.4.2. However, in small composite designs the factorial points are formed using Plackett-Burman designs rather than 2^k factorial designs. (See Draper (1983) for a discussion of small composite designs.) This class of designs is also very economical. An example of a small composite design for four variables follows:

$$\begin{array}{cccc} x_1 & x_2 & x_3 & x_4 \end{array}$$
$$\begin{bmatrix} -1 & -1 & -1 & -1 \\ -1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \\ -1 & -1 & 1 & 1 \\ 1 & 1 & 1 & -1 \\ 1 & 1 & -1 & 1 \\ -1 & 1 & 1 & 1 \\ \alpha & 0 & 0 & 0 \\ -\alpha & 0 & 0 & 0 \\ 0 & \alpha & 0 & 0 \\ 0 & -\alpha & 0 & 0 \\ 0 & 0 & \alpha & 0 \\ 0 & 0 & -\alpha & 0 \\ 0 & 0 & 0 & \alpha \\ 0 & 0 & 0 & -\alpha \\ 0 & 0 & 0 & 0 \end{bmatrix} . \qquad (2.54)$$

The factorial portion of this design is a $\frac{1}{2}$ fraction with defining contrast I = ACD. This small composite design for four variables has N = 18 while the four variable ccd has N = 26, a significant savings when obtaining observations is expensive.

2.4.7 Equiradial Designs

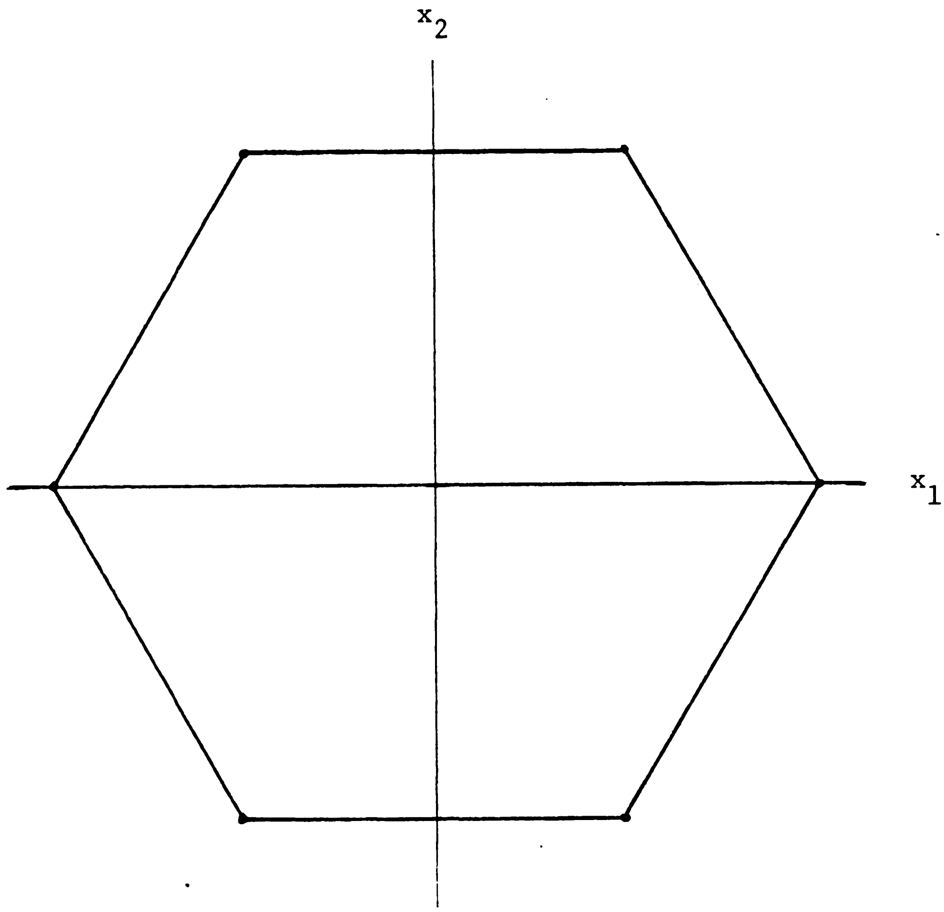
This is a class of designs for two variables, which are all rotatable. The levels of X_1 and X_2 are obtained as follows:

$$\begin{aligned} X_1 &= \rho \cos[\theta + (2\pi u)/N_1] \quad u=0,1,\ldots,N_1-1 \\ X_2 &= \rho \sin[\theta + (2\pi u)/N_1] \ , \end{aligned} \qquad (2.55)$$

where ρ is the distance of the points from the center of the design, θ is the angle the first point makes with the real axis and N_1 is the number of non-center points in the design. (See Myers (1976) for more details.) These designs result in points which are equally-spaced on a circle and therefore must include center points to avoid singularity of the design matrix. The following example is a design with six points on the circle (a hexagon) and three center points.

$$\begin{array}{c} \begin{array}{cc} x_1 & x_2 \end{array} \\ \begin{bmatrix} 1 & 0 \\ 0.5 & \sqrt{.75} \\ -0.5 & \sqrt{.75} \\ -1 & 0 \\ -0.5 & -\sqrt{.75} \\ 0.5 & -\sqrt{.75} \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \end{array} . \tag{2.56}$$

As well as being rotatable the three center points make this design nearly uniform precision. The following diagram illustrates the geometry of the design:



In the next chapter the concept of outliers will be further developed. In particular, it will be determined when an outlier is detrimental to the analysis. Recommendations will be given for diagnosing the effects of discordant observations in an RSM framework.

III. OUTLIERS

In response surface studies, as in any data analysis situation, it is not unusual for discordant observations to occur. The observations can only be judged to be discordant relative to some model of interest, which in this case is the model given in (2.4). In this chapter two techniques for modeling outliers will be investigated, the mean shift outlier model and the variance inflation outlier model.

The effects of the outlier on least squares estimation will be investigated in order to aid in determining the appropriate action to be taken. In some instances it may be determined to be appropriate to discard the outlying observation rather than include it in the analysis and in other cases it may be more beneficial to include the point even though an outlier has occurred. There are trade-offs, then, between the effects of the outlier and the effects of discarding a data point. Thus, the decision to eliminate an observation should only be made by the analyst in close collaboration with the scientists who are experts in the field of experimentation. To aid in the decision-making process, criteria will be developed for determining when the effects of the outlier outweigh the effects of eliminating the observation and thus the outlier should be considered detrimental to the analysis. Statistical tests will be developed based on these criteria for both outlier models.

It is not the intent of this research to recommend the elimination of observations based solely on a statistical test. These test procedures are being set forth as diagnostic tools to assist the analyst and the scientists in making the decision. However, if a formal statistical test is desired, it will be demonstrated in this research that the tests developed here are the appropriate tests to be made.

3.1 Mean Shift Outlier Model

One approach to modeling a single outlier in a set of data is to assume that the model given in (2.4) is correct except that at one point (the i^{th} point) the mean has shifted. The data can then be modeled using the mean shift outlier model given in (2.27). As stated in Chapter II, the mean shift outlier will induce bias into the estimates of the coefficients and into the predicted responses.

To determine the effects of eliminating the outlier, let X_{-i} denote the $(N-1) \times p$ matrix obtained from X by deleting the i^{th} row. It can be shown that

$$(X'_{-i}X_{-i})^{-1} = (X'X)^{-1} + \frac{(X'X)^{-1}\underline{x}_i\underline{x}'_i(X'X)^{-1}}{1-h_{ii}} \tag{3.1}$$

where h_{ii} is defined as in (2.33) (see Rao (1965), pg. 29). This is, apart from σ^2, the variance of the coefficients estimated with only the (N-1) "good" points. Since the diagonal elements in the second term of (3.1) are obviously positive, the variances of the coefficients are inflated by eliminating the outlier, as one would expect. In fact, the increase in variance will be large if the outlier is also a high leverage point (i.e., h_{ii} is close to unity). Thus, the bias induced by the outlier must be weighed against the increase in variance occurring from eliminating the point.

In the next section several criteria will be developed for determining when the outlier is detrimental to the analysis. The outlier is considered detrimental to the analysis if the bias incurred is greater than the increase in variance which occurs when the point is eliminated.

3.2 Criteria of Interest

In this section criteria will be developed for determining when the outlier is degrading. The criteria to be considered are as follows:

1) Sum of the mean square errors of coefficients
2) Sum of the mean square errors of prediction
3) Mean square error of $\hat{y}_i$
4) Integrated mean square error of $\hat{y}$.

3.2.1 Sum of the Mean Square Errors of Coefficients

In general the mean square error is defined as

$$\text{MSE} = \text{Variance} + \text{Bias}^2 \tag{3.2}$$

From this we define the mean square error matrix for $\underline{\hat{\beta}}$ to be

$$\begin{aligned} \text{MSE}(\underline{\hat{\beta}}) &= \text{Var}(\underline{\hat{\beta}}) + \text{Bias}(\underline{\hat{\beta}}) \cdot \text{Bias}(\underline{\hat{\beta}})' \\ &= \sigma^2(X'X)^{-1} + \phi^2(X'X)^{-1}\underline{x}_i\underline{x}_i'(X'X)^{-1} \ , \end{aligned} \tag{3.3}$$

where $\text{Bias}(\underline{\hat{\beta}})$ is that given in (2.28). The first term in (3.3) is the variance-covariance matrix for $\underline{\hat{\beta}}$, which has variances of the coefficients on the diagonal and covariances between coefficients on the off-diagonals. The second term has squares of biases on the diagonal and products of biases on the off-diagonals. Thus the diagonal elements of the matrix given in (3.3) are the mean square errors of the estimated coefficients.

An obvious norm to be used in developing a criterion is the sum of the mean square errors of the coefficients, which is obtained from $MSE(\hat{\underline{\beta}})$ by taking the trace of the matrix. It can be seen that

$$\begin{aligned}\text{Trace}[MSE(\hat{\underline{\beta}})] &= \text{Trace}[\sigma^2(X'X)^{-1} + \phi^2(X'X)^{-1}\underline{x}_i\underline{x}_i'(X'X)^{-1}] \\ &= \sigma^2\ \text{Trace}(X'X)^{-1} + \phi^2\ \text{Trace}[(X'X)^{-1}\underline{x}_i\underline{x}_i'(X'X)^{-1}] \\ &= \sigma^2\ \text{Trace}(X'X)^{-1} + \phi^2\underline{x}_i'(X'X)^{-2}\underline{x}_i \quad . \qquad (3.4)\end{aligned}$$

It is clear from (3.4) that the effect of the outlier on the sum of the mean square errors of the coefficients is determined not only by the magnitude of the outlier but also by the location of the i^{th} point in the design space, as measured by $\underline{x}_i'(X'X)^{-2}\underline{x}_i$. This quantity, which will be referred to as a "pseudo-hat diagonal", plays an important role in the developments of many of the criteria, as will be seen later. Although similar to a hat diagonal in that it is a distance measure, the pseudo-hat diagonal is not scale free as the hat diagonal is. However, this quantity can be used as a diagnostic tool in determining the affect of an outlying observation.

The effects of eliminating the outlier, the i^{th} point, must also be considered. Let X_{-i} denote the X matrix with the i^{th} row deleted, $\underline{y}_{-i}$ denote the response vector with the i^{th} observation deleted and

$\hat{\underline{\beta}}_{-i}$ denote the least squares estimator obtained with the i^{th} observation deleted, i.e.,

$$\hat{\underline{\beta}}_{-i} = (X'_{-i}X_{-i})^{-1}X'_{-i}\underline{y}_{-i} , \tag{3.5}$$

and we will make the usual least squares assumptions. Then as long as $N - 1 \geq p$ it is obvious that

$$E(\hat{\underline{\beta}}_{-i}) = \underline{\beta} \tag{3.6}$$

and

$$\mathrm{Var}(\hat{\underline{\beta}}_{-i}) = \sigma^2(X'_{-i}X_{-i})^{-1} . \tag{3.7}$$

Since $\hat{\underline{\beta}}_{-i}$ is unbiased for $\underline{\beta}$, its mean square error matrix is simply its variance-covariance matrix. Then from (3.7) and the relation given in (3.1)

$$\mathrm{MSE}(\hat{\underline{\beta}}_{-i}) = \sigma^2[(X'X)^{-1} + \frac{(X'X)^{-1}\underline{x}_i\underline{x}_i'(X'X)^{-1}}{1-h_{ii}}] . \tag{3.8}$$

The mean square errors of the coefficients with the i^{th} observation deleted are found on the diagonal of the matrix given in (3.8). The sum of the mean square errors can then be found by taking the trace of the matrix as follows:

$$\text{Trace}[\text{MSE}(\hat{\underline{\beta}}_{-i})] = \sigma^2 \, \text{Trace}[(X'X)^{-1} + \frac{(X'X)^{-1}\underline{x}_i\underline{x}_i'(X'X)^{-1}}{1-h_{ii}}]$$

$$= \sigma^2[\text{Trace}(X'X)^{-1} + \frac{1}{1-h_{ii}} \, \text{Trace}[\underline{x}_i'(X'X)^{-2}\underline{x}_i]]$$

$$= \sigma^2[\text{Trace}(X'X)^{-1} + \frac{\underline{x}_i'(X'X)^{-2}\underline{x}_i}{1-h_{ii}}] \quad . \tag{3.9}$$

It can be seen from the relationship in (3.9) that the pseudo-hat diagonal, $\underline{x}_i'(X'X)^{-2}\underline{x}_i$, not only plays a role in determining the effect of the outlier (as seen in (3.4)), but also in determining the effect of deleting the observation.

Recall that it is desirable to choose the estimator which is more accurate, in this case either $\hat{\underline{\beta}}$ or $\hat{\underline{\beta}}_{-i}$. If $\hat{\underline{\beta}}_{-i}$ is more accurate than $\hat{\underline{\beta}}$, this indicates that the outlier is causing undue harm and should be discarded. Thus, if the sum of the mean square errors for $\hat{\underline{\beta}}$ is larger than the sum of the mean square errors for $\hat{\underline{\beta}}_{-i}$, the outlier is degrading the analysis. The following theorem provides the condition under which inclusion of the point is degrading to the analysis.

Theorem 3.1

$$\text{Trace}[\text{MSE}(\hat{\underline{\beta}})] > \text{Trace}[\text{MSE}(\hat{\underline{\beta}}_{-i})] \quad \text{iff}$$

$$\frac{\phi^2}{\sigma^2} > \frac{1}{1-h_{ii}} \quad . \tag{3.10}$$

Proof

$$\text{Trace}[\text{MSE}(\hat{\underline{\beta}})] > \text{Trace}[\text{MSE}(\hat{\underline{\beta}}_{-i})]$$

$$\Longleftrightarrow \sigma^2 \text{ Trace}(X'X)^{-1} + \phi^2 \underline{x}_i'(X'X)^{-2}\underline{x}_i > \sigma^2[\text{Trace}(X'X)^{-1} + \frac{\underline{x}_i'(X'X)^{-2}\underline{x}_i}{1-h_{ii}}]$$

$$\Longleftrightarrow \phi^2 \underline{x}_i'(X'X)^{-2}\underline{x}_i > \sigma^2 \frac{\underline{x}_i'(X'X)^{-2}\underline{x}_i}{1-h_{ii}}$$

$$\Longleftrightarrow \phi^2 > \frac{\sigma^2}{1-h_{ii}}$$

$$\Longleftrightarrow \frac{\phi^2}{\sigma^2} > \frac{1}{1-h_{ii}} \quad .$$

If the relationship stated in Theorem 3.1 holds, then the outlying observation should be discarded. However, in practice, the experimenter will not know for certain whether or not this relationship holds and thus must collaborate with the expert scientists in determining when to delete the observation in question. The statistical test to be developed in Section 3.3 is intended as an aid in providing a probabilistic basis for a decision regarding the outlier.

3.2.2 Sum of the Mean Square Errors of Prediction

In doing RSM, predicting the response is generally of more interest to the experimenter than estimation. Thus, a criterion similar to that developed in the previous section will be developed for the vector of predicted values, $\hat{\underline{y}}$.

In determining the severity of the outlier, the sum of the mean square errors for the predicted values will be investigated for the cases where the outlier is included in the analysis and where the outlier is deleted. The more accurate prediction is the one whose sum of mean square errors is smaller. Recall that the mean square error matrix of $\hat{\underline{y}}$ is defined from (2.10) and (2.29) as follows:

$$\mathrm{MSE}(\hat{\underline{y}}) = \mathrm{Variance}(\hat{\underline{y}}) + \mathrm{Bias}(\hat{\underline{y}}) \cdot \mathrm{Bias}(\hat{\underline{y}})'$$

$$= \sigma^2 X(X'X)^{-1}X' + \phi^2 X(X'X)^{-1}\underline{x}_i\underline{x}_i'(X'X)^{-1}X' . \qquad (3.11)$$

The first term in (3.11) contains variances and covariances of the elements of $\hat{\underline{y}}$ and the second term contains squares of biases of $\hat{y}$'s on the diagonal and products of biases as off-diagonal elements.

The sum of the mean square errors of prediction is obtained by taking the trace of the matrix given in (3.11) as follows:

$$\mathrm{Trace}[\mathrm{MSE}(\hat{\underline{y}})] = \sigma^2\ \mathrm{Trace}[X(X'X)^{-1}X'] + \phi^2\ \mathrm{Trace}[X(X'X)^{-1}\underline{x}_i\underline{x}_i'(X'X)^{-1}X']$$

$$= \sigma^2\ \mathrm{Trace}[X'X(X'X)^{-1}] + \phi^2\ \mathrm{Trace}[\underline{x}_i'(X'X)^{-1}X'X(X'X)^{-1}\underline{x}_i]$$

$$= \sigma^2\ \mathrm{Trace}(I_p) + \phi^2\underline{x}_i'(X'X)^{-1}\underline{x}_i$$

$$= \sigma^2 p + \phi^2 h_{ii} \quad , \qquad (3.12)$$

where p is the number of parameters being estimated. The effect of high leverage (i.e. h_{ii} near unity) on the bias induced by the outlier can be clearly seen in the relationship given in (3.12).

The effects on prediction of deleting the outlying observation can be determined by investigating $\hat{\underline{y}}_{-i}$, the N-dimensional vector of predicted values obtained by deleting the i^{th} observation. Then,

$$\hat{\underline{y}}_{-i} = X\hat{\underline{\beta}}_{-i} , \tag{3.13}$$

where $\hat{\underline{\beta}}_{-i}$ is as previously defined. Clearly, under the usual least squares assumptions

$$E(\hat{\underline{y}}_{-i}) = X\underline{\beta} \tag{3.14}$$

and

$$\mathrm{Var}(\hat{\underline{y}}_{-i}) = \sigma^2 X(X'_{-i}X_{-i})^{-1}X' \quad . \tag{3.15}$$

The mean square error matrix for $\hat{\underline{y}}_{-i}$ is simply the variance-covariance matrix since $\hat{\underline{y}}_{-i}$ is unbiased for $X\underline{\beta}$. Using the relation given in (3.1),

$$\begin{aligned} \mathrm{MSE}(\hat{\underline{y}}_{-i}) &= \sigma^2 X[(X'X)^{-1} + \frac{(X'X)^{-1}\underline{x}_i\underline{x}_i'(X'X)^{-1}}{1-h_{ii}}]X' \\ &= \sigma^2[X(X'X)^{-1}X' + \frac{X(X'X)^{-1}\underline{x}_i\underline{x}_i'(X'X)^{-1}X'}{1-h_{ii}}] \ . \end{aligned} \tag{3.16}$$

This matrix has mean square errors (i.e. variances) of the elements of $\hat{\underline{y}}_{-i}$ on the diagonal and covariances of these elements as off-diagonals. The trace of the matrix in (3.16) gives the sum of the mean square errors for $\hat{\underline{y}}_{-i}$ as follows:

$$\text{Trace}\,[\text{MSE}(\hat{\underline{y}}_{-i})]$$

$$= \sigma^2 \,\text{Trace}[X(X'X)^{-1}X' + \frac{X(X'X)^{-1}\underline{x}_i\underline{x}_i'(X'X)^{-1}X'}{1-h_{ii}}]$$

$$= \sigma^2[\text{Trace}[X'X(X'X)^{-1}] + \frac{\text{Trace}[\underline{x}_i'(X'X)^{-1}X'X(X'X)^{-1}\underline{x}_i}{1-h_{ii}}]$$

$$= \sigma^2 p + \frac{\sigma^2 h_{ii}}{1-h_{ii}} \quad . \tag{3.17}$$

As would be expected, the hat diagonal also determines the effect of discarding the outlier.

Once again the trade-off between the effect of the outlier and the effect of discarding an observation must be assessed to determine the appropriate action. If the sum of the mean square errors for $\hat{\underline{y}}$ is larger than the sum of mean square errors for $\hat{\underline{y}}_{-i}$, then this indicates that the effect of the outlier outweighs the effect of deleting an observation and the point should be discarded. The following theorem gives the condition under which the outlier is having a degrading effect on the analysis.

<u>Theorem 3.2</u>

$$\text{Trace}[\text{MSE}(\hat{\underline{y}})] > \text{Trace}[\text{MSE}(\hat{\underline{y}}_{-i})] \quad \text{iff}$$

$$\frac{\phi^2}{\sigma^2} > \frac{1}{1-h_{ii}} \quad . \tag{3.18}$$

Proof

$$\text{Trace}[\text{MSE}(\hat{\underline{y}})] > \text{Trace}[\text{MSE}(\hat{\underline{y}}_{-i})]$$

$$\Longleftrightarrow \sigma^2 p + \phi^2 h_{ii} > \sigma^2 p + \frac{\sigma^2 h_{ii}}{1-h_{ii}}$$

$$\Longleftrightarrow \phi^2 h_{ii} > \frac{\sigma^2 h_{ii}}{1-h_{ii}}$$

$$\Longleftrightarrow \frac{\phi^2}{\sigma^2} > \frac{1}{1-h_{ii}} \quad .$$

This theorem gives the same condition for determining when the outlier is detrimental as does the criterion for MSE($\hat{\underline{\beta}}$). Thus, whether estimation of coefficients or prediction is the main concern of the experimenter, the same condition is reached. It is obvious from the condition that the leverage of the point determines whether or not the outlier is severe enough to require elimination. Recall, however, that in practice the experimenter can not know whether this condition holds or not.

3.2.3 Mean Square Error of $\hat{y}_i$

In certain situations in response surface work the experimenter may be interested in the predicted response at the point where the outlier occurs. In these cases the quality of $\hat{y}_i = \underline{x}_i'\hat{\underline{\beta}}$, the estimated response at the i^{th} point, is of interest. The effect of the outlier on the accuracy of this prediction will be investigated along with the effect of deleting the i^{th} point. These results will be used to develop a criterion similar to those developed above.

From the development in the previous section it is clear that the mean square error of $\hat{y}_i$ is the i^{th} diagonal element of the mean square error matrix given in (3.11). That is,

$$\begin{aligned} MSE(\hat{y}_i) &= \sigma^2 \underline{x}_i'(X'X)^{-1}\underline{x}_i + \phi^2 \underline{x}_i'(X'X)^{-1}\underline{x}_i\underline{x}_i'(X'X)^{-1}\underline{x}_i \\ &= \sigma^2 h_{ii} + \phi^2 h_{ii}^2 \quad . \end{aligned} \tag{3.19}$$

Once again the leverage of the outlying point is seen to determine the effect of the outlier.

Let the predicted response at the i^{th} point obtained by deleting the i^{th} point be denoted by $\hat{y}_{-i}$. Then,

$$\hat{y}_{-i} = \underline{x}_i'\hat{\underline{\beta}}_{-i} \quad . \tag{3.20}$$

From previous results it is obvious that

$$E(\hat{y}_{-i}) = \underline{x}_i'\underline{\beta} \tag{3.21}$$

and

$$Var(\hat{y}_{-i}) = \sigma^2 \underline{x}_i'(X_{-i}'X_{-i})^{-1}\underline{x}_i \quad . \tag{3.22}$$

The mean square error of $\hat{y}_{-i}$ is then equal to the variance of $\hat{y}_{-i}$ since it is unbiased for $\underline{x}_i'\underline{\beta}$. Using the relationship in (3.1) gives

$$\begin{aligned} MSE(\hat{y}_{-i}) &= \sigma^2 [\underline{x}_i'(X'X)^{-1}\underline{x}_i + \frac{\underline{x}_i'(X'X)^{-1}\underline{x}_i\underline{x}_i'(X'X)^{-1}\underline{x}_i}{1-h_{ii}}] \\ &= \sigma^2 [h_{ii} + \frac{h_{ii}^2}{1-h_{ii}}] \quad . \end{aligned} \tag{3.23}$$

In terms of the prediction at the i^{th} point, the outlier is damaging to the analysis if the mean square error of prediction at the i^{th} point is larger with the point included in the analysis than with the point deleted. This occurs when the condition stated in the following theorem holds.

Theorem 3.3

$$\mathrm{MSE}(\hat{y}_i) > \mathrm{MSE}(\hat{y}_{-i}) \quad \text{iff}$$

$$\frac{\phi^2}{\sigma^2} > \frac{1}{1-h_{ii}} \quad . \tag{3.24}$$

Proof

$$\mathrm{MSE}(\hat{y}_i) > \mathrm{MSE}(\hat{y}_{-i})$$

$$\Longleftrightarrow \sigma^2 h_{ii} + \phi^2 h_{ii}^2 > \sigma^2 [h_{ii} + \frac{h_{ii}^2}{1-h_{ii}}]$$

$$\Longleftrightarrow \phi^2 h_{ii}^2 > \frac{\sigma^2 h_{ii}^2}{1-h_{ii}}$$

$$\Longleftrightarrow \frac{\phi^2}{\sigma^2} > \frac{1}{1-h_{ii}} \quad .$$

The same condition then is obtained with this criterion as with the previous criteria. It is interesting to note that whether one is concerned with estimation at all of the design points or estimation only at the outlying point the same condition is obtained.

3.2.4 Integrated Mean Square Error of $\hat{y}(x)$

The criteria developed in the previous two sections, although prediction-oriented, have a shortcoming which may be of concern to experimenters involved in RSM. The experimenter is interested in predicting the response throughout the region of interest, R. The criteria in the previous sections take into account only the prediction occurring at the design points, which may be of little or no interest to the experimenter. A more appropriate criterion, then, is the integrated mean square error of prediction, where the integration is as described in Chapter II.

Recall that if $\underline{x}$ is an arbitrary point in R, then the estimated response at $\underline{x}$ is $\hat{y}(\underline{x}) = \underline{x}'\hat{\underline{\beta}}$ and the prediction variance (apart from σ^2) is

$$\frac{\mathrm{Var}(\hat{y}(\underline{x}))}{\sigma^2} = \underline{x}'(X'X)^{-1}\underline{x} \quad . \tag{3.25}$$

The predicted response at $\underline{x}$ will be biased as seen in the following:

$$\begin{aligned} E(\hat{y}(\underline{x})) &= E(\underline{x}'\hat{\underline{\beta}}) \\ &= \underline{x}'E(\hat{\underline{\beta}}) \\ &= \underline{x}'[\underline{\beta} + (X'X)^{-1}\underline{x}_i\phi] \text{ from (2.28)} \\ &= \underline{x}'\underline{\beta} + \underline{x}'(X'X)^{-1}\underline{x}_i\phi \quad . \end{aligned} \tag{3.26}$$

Thus, the bias in prediction at $\underline{x}$ is given by

$$\mathrm{Bias}_i(\hat{y}(\underline{x})) = \underline{x}'(X'X)^{-1}\underline{x}_i\phi \quad . \tag{3.27}$$

The integrated mean square error of prediction at $\underline{x}$ (denoted by IMSE($\hat{y}$(x))) is defined to be

$$\text{IMSE}(\hat{y}(\underline{x})) = \frac{NK}{\sigma^2}\int_R \text{MSE}(\hat{y}(\underline{x}))d\underline{x}$$

$$= \frac{NK}{\sigma^2}\int_R \text{Var}(\hat{y}(\underline{x}))d\underline{x} + \frac{NK}{\sigma^2}\int_R \text{Bias}_i^2\,(\hat{y}(\underline{x}))d\underline{x}$$

$$= NK\int_R \underline{x}'(X'X)^{-1}\underline{x}d\underline{x} + \frac{NK\phi^2}{\sigma^2}\int_R \underline{x}_i'(X'X)^{-1}\underline{xx}'(X'X)^{-1}\underline{x}_i d\underline{x}\ , \tag{3.28}$$

where K and $d\underline{x}$ are as defined in Chapter II. The first term in (3.28), the integrated variance, is given in (2.26) as

$$N\ \text{Trace}[\mu(X'X)^{-1}]\ , \tag{3.29}$$

where μ is the region moment matrix defined in (2.23). A development similar to that given in Chapter II for the integrated variance will be given for the bias portion of the integrated mean square error.

$$\frac{NK}{\sigma^2}\int_R \text{Bias}_i^2(\hat{y}(\underline{x}))d\underline{x} = \frac{NK\phi^2}{\sigma^2}\int_R \underline{x}_i'(X'X)^{-1}\underline{xx}'(X'X)^{-1}\underline{x}_i d\underline{x}$$

$$= \frac{N\phi^2}{\sigma^2}\text{Trace}[K\int_R \underline{x}_i'(X'X)^{-1}\underline{xx}'(X'X)^{-1}\underline{x}_i d\underline{x}]$$

$$= \frac{N\phi^2}{\sigma^2}\text{Trace}[K\int_R \underline{xx}'d\underline{x}(X'X)^{-1}\underline{x}_i\underline{x}_i'(X'X)^{-1}]$$

$$= \frac{N\phi^2}{\sigma^2}\text{Trace}[\mu(X'X)^{-1}\underline{x}_i\underline{x}_i'(X'X)^{-1}]\ . \tag{3.30}$$

Combining (3.29) and (3.30) gives

$$\mathrm{IMSE}(\hat{y}(\underline{x})) = N\ \mathrm{Trace}[\mu(X'X)^{-1}]$$

$$+ \frac{N\phi^2}{\sigma^2}\ \mathrm{Trace}[\mu(X'X)^{-1}\underline{x}_i\underline{x}_i'(X'X)^{-1}] \quad . \qquad (3.31)$$

To determine the effects of deleting the outlying observation let X_{-i}, $\underline{y}_{-i}$ and $\hat{\underline{\beta}}_{-i}$ be as previously defined. Then, let $\hat{y}_{-i}(\underline{x})$ denote the predicted response at the point $\underline{x}$ obtained by deleting the i^{th} observation, i.e.,

$$\hat{y}_{-i}(\underline{x}) = \underline{x}'\hat{\underline{\beta}}_{-i} \quad . \qquad (3.32)$$

It is obvious from previous results that

$$E(\hat{y}_{-i}(\underline{x})) = \underline{x}'\underline{\beta} \qquad (3.33)$$

and

$$\frac{\mathrm{Var}(\hat{y}_{-i}(\underline{x}))}{\sigma^2} = \underline{x}'(X_{-i}'X_{-i})^{-1}\underline{x} \quad . \qquad (3.34)$$

Thus, the integrated mean square error of $\hat{y}_{-i}(\underline{x})$ is equal to the integrated variance for $\hat{y}_{-i}(\underline{x})$. Clearly, from the development in Chapter II

$$\text{IMSE}(\hat{y}_{-i}(\underline{x}))$$

$$= N\ \text{Trace}[\mu(X'_{-i}X_{-i})^{-1}]$$

$$= N\ \text{Trace}\{\mu[(X'X)^{-1} + \frac{(X'X)^{-1}\underline{x}_i\underline{x}_i'(X'X)^{-1}}{1-h_{ii}}]\}$$

$$= N\ \text{Trace}[\mu(X'X)^{-1}] + \frac{N}{1-h_{ii}}\ \text{Trace}[\mu(X'X)^{-1}\underline{x}_i\underline{x}_i'(X'X)^{-1}]\ . \quad (3.35)$$

Now, the two integrated mean square errors must be compared to determine under what condition the outlier is deteriorating the analysis. This condition is given by the following theorem:

Theorem 3.4

$$\text{IMSE}(\hat{y}(\underline{x})) > \text{IMSE}(\hat{y}_{-i}(\underline{x})) \quad \text{iff}$$

$$\frac{\phi^2}{\sigma^2} > \frac{1}{1-h_{ii}} \quad . \quad (3.36)$$

Proof

$$\text{IMSE}(\hat{y}(\underline{x})) > \text{IMSE}(\hat{y}_{-i}(\underline{x}))$$

$$\iff N\ \text{Trace}[\mu(X'X)^{-1}] + \frac{N\phi^2}{\sigma^2}\ \text{Trace}[\mu(X'X)^{-1}\underline{x}_i\underline{x}_i'(X'X)^{-1}]$$

$$> N\ \text{Trace}[\mu(X'X)^{-1}] + \frac{N}{1-h_{ii}}\ \text{Trace}[\mu(X'X)^{-1}\underline{x}_i\underline{x}_i'(X'X)^{-1}]$$

$$\iff \frac{N\phi^2}{\sigma^2}\ \text{Trace}[\mu(X'X)^{-1}\underline{x}_i\underline{x}_i'(X'X)^{-1}]$$

$$> \frac{N}{1-h_{ii}}\ \text{Trace}[\mu(X'X)^{-1}\underline{x}_i\underline{x}_i'(X'X)^{-1}]$$

$$\Longleftrightarrow \frac{\phi^2}{\sigma^2} > \frac{1}{1-h_{ii}} \quad .$$

Thus, all of the criteria considered, MSE($\hat{\underline{\beta}}$), MSE($\hat{\underline{y}}$), MSE($\hat{y}_i$) and IMSE($\hat{y}(\underline{x})$), yield the same result. Whether the experimenter is interested in estimation of coefficients or prediction of the response, the same condition should be used in determining when the outlier is detrimental to the analysis. As pointed out earlier, the experimenter in practice will not know whether or not the condition holds. In the next section a statistical test, intended for use as a diagnostic tool, will be developed, which can aid the experimenter in the decision-making process.

3.3 Test Procedure for Mean Shift Outlier

3.3.1 Test Statistic and Distribution

From the criteria developed in Section 3.2 it is clear that the experimenter need not be concerned with the occurrence of a mean shift (i.e., $\phi \neq 0$), but there should be concern with whether the shift is large enough to be damaging (i.e., $\phi^2/\sigma^2 > 1/1-h_{ii}$). If the outlier does not represent a "negative impact" on the results, there is no reason to exclude the point from the analysis.

The data presented in Table 1 was obtained in a consulting project through the Statistical Consulting Lab and will be used to illustrate the test procedures. The data is from a chemical experiment in which the effects of fluidizing gas flow rate (x_1), supernatent gas flow rate (x_2), and supernatent gas inlet nozzle opening (x_3)

on a heat transfer coefficient (y) were investigated. The design used was a central composite design with $\alpha = 1.73$ and two center points.

A full second order model was fit to the data and the residuals (i.e., $\hat{\varepsilon}_i = y_i - \hat{y}_i$) obtained are plotted in Figure 1. It can be seen from this plot that the residual for observation 13 appears to be larger than the other residuals. Although this is by no means conclusive evidence, it might appear that this point is a maverick observation. Cook and Weisberg (1982) give a statistical test for determining whether a mean shift has occurred at a given point. The test is based on the studentized residual, commonly referred to as RSTUDENT, which is defined as

$$\text{RSTUDENT}_i = \frac{\hat{\varepsilon}_i}{s_{-i}(1-h_{ii})^{1/2}} , \tag{3.37}$$

where $\hat{\varepsilon}_i$ is the residual at the i^{th} point and s^2_{-i} is the residual mean square computed without the i^{th} point. Cook and Weisberg show

$$s^2_{-i} = \frac{\sum\limits_{j \neq i} (y_i - \underline{x}_j'\hat{\underline{\beta}}_{-i})^2}{N-p-1}$$

$$= \frac{(N-p)s^2 - \hat{\varepsilon}_i^2/(1-h_{ii})}{N-p-1} . \tag{3.38}$$

The residuals, hat diagonals and RSTUDENT values for the chemical data are listed in Table 2 and a plot of the RSTUDENT values is given in Figure 2. It can be seen from the plot in Figure 2 that

Table 1

Chemical Data

obs	x_1	x_2	x_3	y
1	1.00	1.00	1.00	308.299
2	1.00	1.00	-1.00	537.233
3	1.00	-1.00	1.00	180.967
4	1.00	-1.00	-1.00	362.379
5	-1.00	1.00	1.00	159.955
6	-1.00	1.00	-1.00	293.042
7	-1.00	-1.00	1.00	86.043
8	-1.00	-1.00	-1.00	142.467
9	-1.73	0.00	0.00	57.907
10	1.73	0.00	0.00	367.200
11	0.00	-1.73	0.00	91.335
12	0.00	1.73	0.00	336.102
13	0.00	0.00	-1.73	528.764
14	0.00	0.00	1.73	187.823
15	0.00	0.00	0.00	184.368
16	0.00	0.00	0.00	190.829

x_1 = fluidizing gas flow rate

x_2 = supernatent gas flow rate

x_3 = supernatent gas nozzle inlet opening

y = heat transfer coefficient

Table 2

Chemical Data Diagnostics

obs	Residual	RSTUDENT	Hat Diagonal
1	9.9688	0.8911	0.6611
2	-17.4619	-1.9047	0.6611
3	6.8917	0.5920	0.6611
4	-5.9935	-0.5105	0.6611
5	3.5207	0.2949	0.6611
6	-9.3645	-0.8294	0.6611
7	14.9891	1.4982	0.6611
8	-12.4416	-1.1649	0.6611
9	0.6991	0.0547	0.6186
10	2.6058	0.2049	0.6186
11	-3.1980	-0.2519	0.6186
12	6.5029	0.5227	0.6186
13	24.9564	4.0140	0.6186
14	-21.6515	-2.5979	0.6186
15	-3.2422	-0.2228	0.5000
16	3.2188	0.2211	0.5000

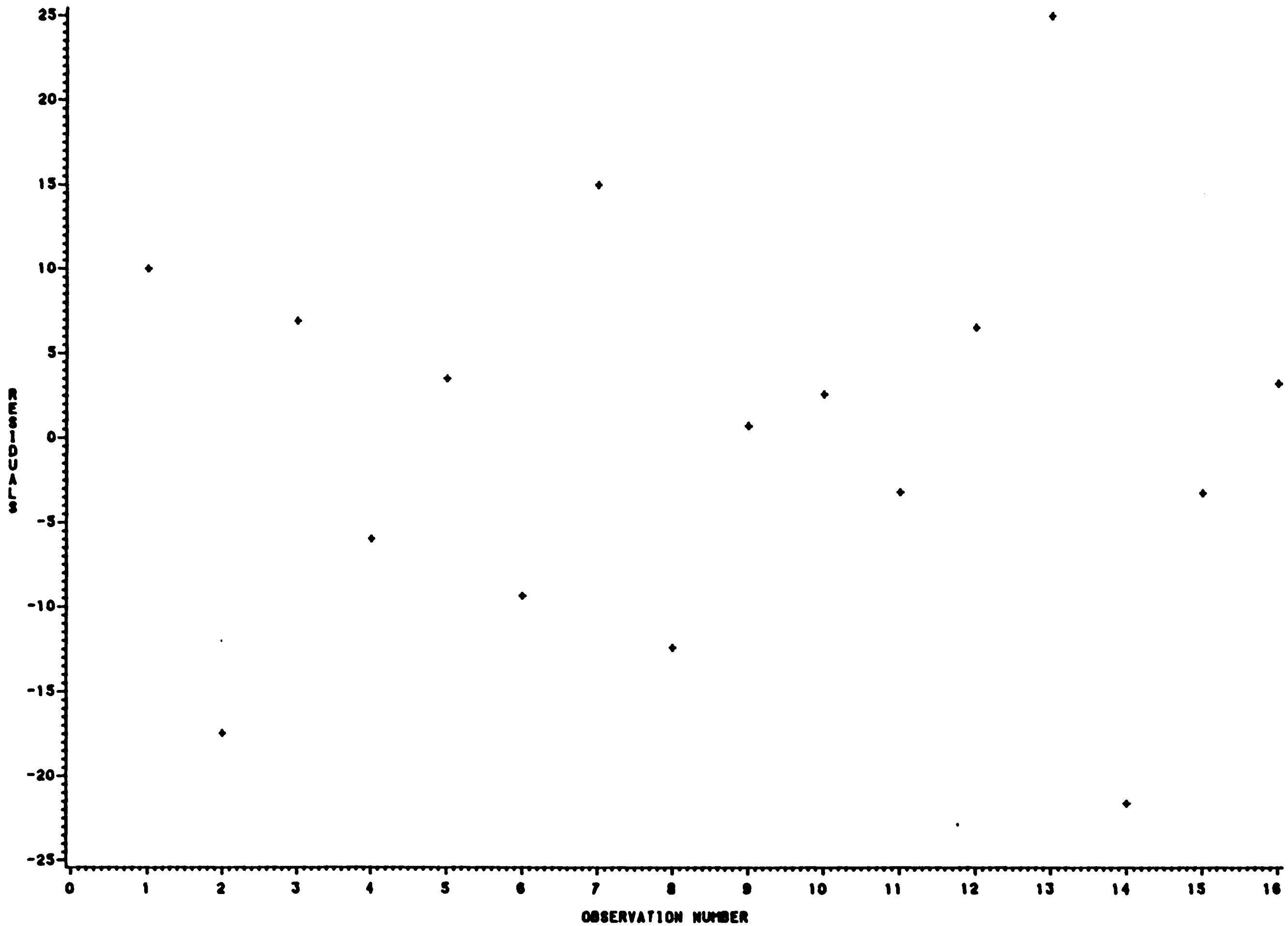


Figure 1

Chemical Data Residuals

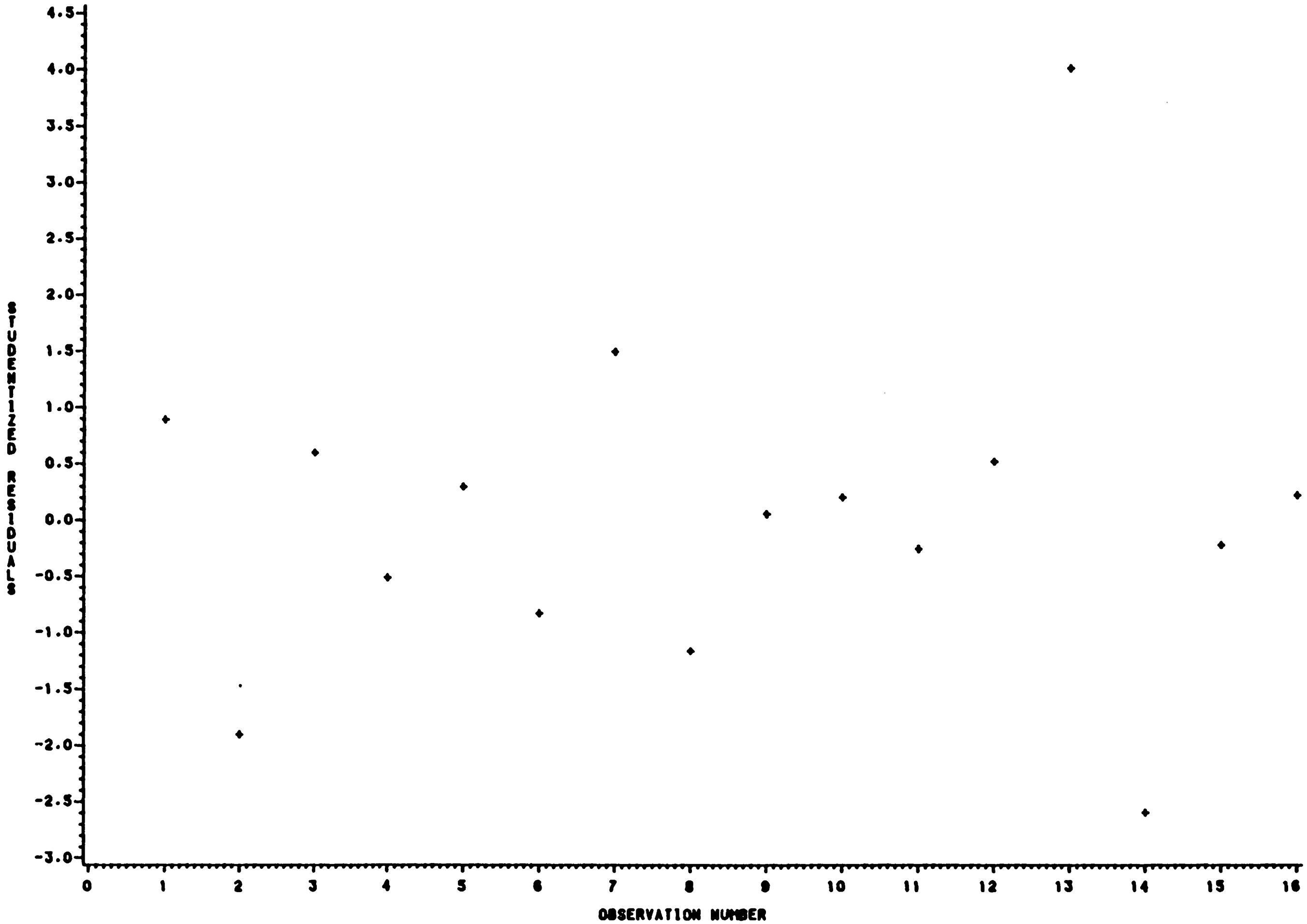


Figure 2

Chemical Data Studentized Residuals

the RSTUDENT value for observation 13 is much larger than the values for the remaining points, thus lending additional evidence for this observation being a maverick.

Cook and Weisberg also show that under the null hypothesis H_0: $\phi = 0$, RSTUDENT_i has a t-distribution with N-p-1 degrees of freedom. For convenience $(\text{RSTUDENT}_i)^2$ will be used here which can be shown to have an F distribution with 1 and N-p-1 degrees of freedom under the null hypothesis stated above (see Graybill (1976)). Then, if $(\text{RSTUDENT}_i)^2$ is greater than or equal to the $1-\alpha$ percentage point for the F distribution with 1 and N-p-1 degrees of freedom there is evidence that a mean shift has occurred at the i^{th} point.

For the chemical data $(\text{RSTUDENT}_{13})^2 = 16.11$ and for $\alpha = .01$, $F_{1,5,.99} = 10.0$. Thus, it can be concluded that a shift in the mean has occurred at this point. However, this in no way implies that the mean shift is sufficient to warrant elimination of the point. So, based on this test the experimenter might decide to discard the outlying observation even though, as will be seen later, it is most likely not harmful to the analysis.

As an alternative, a test will be developed based on the same statistic, RSTUDENT, which will test the null hypothesis that is directed at the performance of the regression:

$$H_0:\ \frac{\phi^2}{\sigma^2} \le \frac{1}{1-h_{ii}} \qquad H_1:\ \frac{\phi^2}{\sigma^2} > \frac{1}{1-h_{ii}}\ . \tag{3.39}$$

The following theorem gives the distribution of the statistic.

Theorem 3.5

Under the mean shift outlier model given in (2.27) with the usual least squares assumptions on $\underline{\varepsilon}$ and with the additional assumption that each term in the $\underline{\varepsilon}$ vector has a normal distribution, then

$$(\text{RSTUDENT}_i)^2 = \frac{\hat{\varepsilon}_i^2}{s_{-i}^2(1-h_{ii})} \sim F'_{1,n-p-1,\lambda} \;, \tag{3.40}$$

where

$$\lambda = \frac{\phi^2(1-h_{ii})}{2\sigma^2} \tag{3.41}$$

is the non-centrality parameter.

Proof

Given in Appendix A.

Then to test the hypothesis that the mean shift is not harmful, $(\text{RSTUDENT}_i)^2$ will be compared to the appropriate percentage point from the non-central F distribution with 1 and N-p-1 degrees of freedom. Under the null hypothesis stated in (3.39) the noncentrality parameter, λ, will be equal to 1/2. To see this note that under H_0,

$$\frac{\phi^2}{\sigma^2} = \frac{1}{1-h_{ii}} \quad . \tag{3.42}$$

Substituting (3.42) into (3.41) gives $\lambda = 1/2$. The appropriate decision rule for testing the hypothesis in (3.39) is to reject H_0 if $(\text{RSTUDENT}_i)^2$ is greater thán or equal to the $1-\alpha$ percentage point for the noncentral F distribution with degrees of freedom 1 and N-p-1 and non-centrality parameter, $\lambda = 1/2$.

In the chemical data example $(\text{RSTUDENT}_{13})^2$ turned out to be 16.11 and it turns out that $F'_{1,5,1/2,.99} \doteq 29.78$ (see Patnaik (1949)). In this case then it should be concluded that the mean shift is not of sufficient magnitude to warrant elimination of the point. On the basis of the test of the hypothesis that $\phi = 0$ the experimenter might decide to discard the point in question. However, it is not clear from this test that the point is severely degrading the analysis and should likely be retained.

It can be seen from this example that the test described by Cook and Weisberg is not the appropriate test to use if a mean shift outlier is feared. This test can lead to an incorrect decision regarding the elimination of an outlying observation. The test of the hypothesis given in (3.39) is the appropriate test to use in this situation.

As stated previously, the purpose of this research is not to advocate the elimination of data based solely on any statistical test. Rather, the test developed here can and should be used as a diagnostic tool by the scientists and analysts. If the analyst chooses to do a formal statistical test, however, the test developed in this research is the appropriate test to be made.

3.3.2 Power of Test

The power of the test developed in the previous section is given by

$$1-\beta = P(\text{Rejecting } H_0 \mid H_0 \text{ is false})$$

$$= P((\text{RSTUDENT}_i)^2 \geq F'_{1,n-p-1,1/2,1-\alpha} \mid \frac{\phi^2}{\sigma^2} > \frac{1}{1-h_{ii}})$$

$$= P(F'_{1,n-p-1,\lambda} \geq F'_{1,n-p-1,1/2,1-\alpha} \mid \lambda > \tfrac{1}{2}) . \tag{3.43}$$

It can be seen from the expression for λ given in (3.41) that λ (and thus the power) is an increasing function of ϕ^2 and a decreasing function of the leverage, measured by h_{ii}.

An approximation to the power given in (3.43) can be made using the approximation to the non-central F distribution given in Patnaik (1949). The approximation states that

$$\int_0^{F'} P_{v_1,v_2}(F'/\lambda)dF' \tag{3.44}$$

can be approximated by

$$\int_0^{X} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz , \tag{3.45}$$

where

$$X = \frac{\left(\frac{v_1F'}{v_1+\lambda}\right)^{1/3}\left(1-\frac{2}{9v_2}\right)-\left(1-\frac{2(v_1+2\lambda)}{9(v_1+\lambda)^2}\right)}{\left(\frac{2(v_1+2)}{9(v_1+\lambda)^2}+\frac{2}{9v_2}\left(\frac{v_1F'}{v_1+\lambda}\right)^{2/3}\right)^{1/2}} \tag{3.46}$$

and $P_{v_1,v_2}(F'|\lambda)$ is the non-central F density function.

3.4 Variance Inflation Outlier Model

An alternative approach to modeling a single outlier is the variance inflation outlier model. Rather than one point (the i^{th} point) having a shift in the mean, in this case the point has an error variance which is larger than the remaining points. The effects of this type of outlier on least squares estimation will be investigated. The goal of the investigation is the same as that of the previous section; that is, to determine under what conditions the outlier is of large enough magnitude to be considered detrimental to the analysis. Criteria similar to those developed in Section 3.2 will be developed for the variance inflation outlier model. It will be seen that the usual least squares estimators of the coefficients and the response remain unbiased under this model and thus the criteria to be developed will be variance criteria rather than criteria based on mean square errors.

3.5 Criteria of Interest

In this section the criteria will be developed for determining when the variance inflation outlier is damaging to the analysis. The criteria to be developed are the following:

1) Sum of the variances of coefficients

2) Sum of the prediction variances

3) Variance of $\hat{y}_i$

4) Integrated variance of $\hat{y}$

3.5.1 Sum of the Variances of Coefficients

Recall that the model to be fit is

$$\underline{y} = X\underline{\beta} + \underline{\varepsilon} \tag{3.47}$$

with the elements of $\underline{\varepsilon}$ independent of each other, each with expectation zero and the variance-covariance matrix for $\underline{\varepsilon}$ is that given in (2.36).

It is obvious that the usual least squares estimator, $\hat{\underline{\beta}}$, is unbiased for $\underline{\beta}$, i.e., $E(\hat{\underline{\beta}}) = \underline{\beta}$. The variance of $\hat{\underline{\beta}}$ is found as follows:

$$\begin{aligned} \mathrm{Var}(\hat{\underline{\beta}}) &= \mathrm{Var}[(X'X)^{-1}X'\underline{y}] \\ &= (X'X)^{-1}X'\mathrm{Var}(\underline{y})X(X'X)^{-1} \end{aligned}$$

$$= (X'X)^{-1}X' \begin{bmatrix} \sigma^2 & & & & & \phi \\ & \sigma^2 & & & & \\ & & \ddots & & & \\ & & & \sigma^2+\sigma^2_\Delta & & \\ & & & & \ddots & \\ \phi & & & & & \sigma^2 \end{bmatrix} X(X'X)^{-1}$$

$$= \sigma^2(X'X)^{-1}X'X(X'X)^{-1} + \sigma^2_\Delta(X'X)^{-1}\underline{x}_i\underline{x}_i'(X'X)^{-1}$$

$$= \sigma^2(X'X)^{-1} + \sigma^2_\Delta(X'X)^{-1}\underline{x}_i\underline{x}_i'(X'X)^{-1} \quad . \tag{3.48}$$

Since the diagonal elements of the second term in (3.48) are positive, the variances of the coefficients are inflated by the presence of the outlier.

Recall, also, that if the outlying observation is deleted then the variances of the coefficients are also inflated. Thus, the trade-off that exists in this case is between the increase in variance caused by the outlier and the increase in variance caused by deleting a data point. This trade-off will be evaluated by comparing the sums of the variances of coefficients to determine when the outlier is severe enough to warrant elimination.

The sum of the variances of the coefficients is found by taking the trace of the variance-covariance matrix given in (3.48) as follows:

$$\text{Trace}[\text{Var}(\hat{\underline{\beta}})] = \text{Trace}[\sigma^2(X'X)^{-1} + \sigma^2_\Delta(X'X)^{-1}\underline{x}_i\underline{x}_i'(X'X)^{-1}]$$

$$= \sigma^2 \ \text{Trace}(X'X)^{-1} + \sigma^2_\Delta \ \text{Trace}[(X'X)^{-1}\underline{x}_i\underline{x}_i'(X'X)^{-1}]$$

$$= \sigma^2 \ \text{Trace}(X'X)^{-1} + \sigma^2_\Delta\underline{x}_i'(X'X)^{-2}\underline{x}_i \quad . \tag{3.49}$$

The similarity between the sum of variances of coefficients in this case and the sum of mean square errors in the mean shift case can be seen by comparing (3.49) and (3.4). The pseudo-hat diagonal is again determining the effect of the outlier on the variances in the same way that it determined the effect of the mean shift outlier on the mean square errors.

The effect of deleting the outlier will be the same as in the case of the mean shift outlier model. The sum of the variances of coefficients found with the i^{th} point deleted is then the trace of the variance-covariance matrix of $\hat{\underline{\beta}}_{-i}$ given in (3.7). This sum was found in (3.9) to be

$$\text{Trace}[\text{Var}(\hat{\underline{\beta}}_{-i})] = \sigma^2[\text{Trace}(X'X)^{-1} + \frac{\underline{x}_i'(X'X)^{-2}\underline{x}_i}{1-h_{ii}}] \; . \tag{3.50}$$

The sums of variances of coefficients for the situation in which the outlier is included, (3.49), and for the outlier deleted, (3.50), will be compared to determine when the inflation in variance is large enough to be destructive to the analysis. The following theorem gives the condition under which inclusion of the outlier is detrimental to the analysis.

<u>Theorem 3.6</u>

$$\text{Trace}[\text{Var}(\hat{\underline{\beta}})] > \text{Trace}[\text{Var}(\hat{\underline{\beta}}_{-i})] \quad \text{iff}$$

$$\frac{\sigma_\Delta^2}{\sigma^2} > \frac{1}{1-h_{ii}} \quad . \tag{3.51}$$

Proof

$$\text{Trace}[\text{Var}(\hat{\underline{\beta}})] > \text{Trace}[\text{Var}(\hat{\underline{\beta}}_{-i})]$$

$$\Longleftrightarrow \sigma^2 \ \text{Trace}(X'X)^{-1} + \sigma_\Delta^2 \underline{x}_i'(X'X)^{-2}\underline{x}_i$$

$$> \sigma^2[\text{Trace}(X'X)^{-1} + \frac{\underline{x}_i'(X'X)^{-2}\underline{x}_i}{1-h_{ii}}]$$

$$\Longleftrightarrow \sigma_\Delta^2 \underline{x}_i'(X'X)^{-2}\underline{x}_i > \frac{\underline{x}_i'(X'X)^{-2}\underline{x}_i}{1-h_{ii}}$$

$$\Longleftrightarrow \frac{\sigma_\Delta^2}{\sigma^2} > \frac{1}{1-h_{ii}} \quad .$$

It can be seen from this theorem that if the hat diagonal at the point where the outlier occurs is large (near unity) then a large inflation in the variance can be tolerated without being damaging to the analysis. As in the case of the mean shift outlier, the experimenter will obviously not know whether this condition holds or not and must rely on a statistical test to give information about the magnitude of the outlier.

3.5.2 Sum of Prediction Variances

The development of the criterion for the sum of prediction variances parallels that for the sum of mean square errors of prediction given in Section 3.2.3, where ϕ^2 is replaced by σ_Δ^2. In this case, however, there is no bias involved, only variance. The details will not be given here.

Theorem 3.7

$$\mathrm{Trace}[\mathrm{Var}(\hat{\underline{y}})] > \mathrm{Trace}[\mathrm{Var}(\hat{\underline{y}}_{-i})] \quad \text{iff}$$

$$\frac{\sigma_\Delta^2}{\sigma^2} > \frac{1}{1-h_{ii}} \quad . \tag{3.52}$$

Proof

The proof of this theorem is identical to that of Theorem 3.2 with σ_Δ^2 replacing ϕ^2.

As might be expected at this point, the criteria being developed for this case lead to the same condition for an outlier to be counter-productive.

3.5.3 Variance of $\hat{y}_i$

The development of this criterion will not be included here since it follows that given in Section 3.2.3 for the mean square error of $\hat{y}_i$. As in the case of the sum of prediction variances, simply replace ϕ^2 by σ_Δ^2 in that development.

Theorem 3.8

$$\mathrm{Var}(\hat{y}_i) > \mathrm{Var}(\hat{y}_{-i}) \quad \text{iff}$$

$$\frac{\sigma_\Delta^2}{\sigma^2} > \frac{1}{1-h_{ii}} \quad . \tag{3.53}$$

Proof

In the proof of Theorem 3.3 replace ϕ^2 by σ_Δ^2.

3.5.4 Integrated Variance of $\hat{y}(x)$

The development of the integrated prediction variance will not be given here since it can be obtained directly from Section 3.2.4 by replacing ϕ^2 by σ_Δ^2. Recall that there is no bias in this case and the criterion is a variance criterion rather than a mean square error criterion.

Theorem 3.9

$$\mathrm{IVAR}(\hat{y}(x)) > \mathrm{IVAR}(\hat{y}_{-i}(x)) \quad \text{iff}$$

$$\frac{\sigma_\Delta^2}{\sigma^2} > \frac{1}{1-h_{ii}} \quad . \tag{3.54}$$

Proof

Replace ϕ^2 by σ_Δ^2 in the proof of Theorem 3.4.

Thus, we see that the four criteria developed for the variance inflation outlier model lead to the same condition for an outlier to be causing undue harm. It also turns out that this condition is the same as that found in the mean square error counterpart in the mean shift case, with, of course, σ_Δ^2 replacing ϕ^2. Since there is no way to ascertain the type of outlier occurring, it is reassuring to know that, in terms of the damage being done by the outlier, it makes no difference which model is chosen.

The condition derived here, that the outlier is causing undue harm if $\sigma_\Delta^2/\sigma^2 > 1/(1-h_{ii})$, will be used to develop a statistical test for the variance inflation model similar to that developed in Section 3.3 for the mean shift model.

3.6 Test Procedure for Variance Inflation Outlier

3.6.1 Test Statistic and Distribution

A test procedure similar to that of Section 3.3 will be developed for determining when a variance inflation outlier is harmful. The chemical data in Table 1 will again be used to illustrate the test procedure. The test statistic to be used will be based on the PRESS residual

$$\begin{aligned} \delta_i &= y_i - \hat{y}_{-i} \\ &= y_i - \underline{x}_i'\hat{\underline{\beta}}_{-i} \; . \end{aligned} \tag{3.55}$$

To determine the distribution of the PRESS residual, recall the assumptions for the variance inflation outlier model:

$$E(\underline{\varepsilon}) = \underline{0}$$

$$\mathrm{Var}(\underline{\varepsilon}) = \begin{bmatrix} \sigma^2 & & & & & 0 \\ & \sigma^2 & & & & \\ & & \ddots & & & \\ & & & \sigma^2+\sigma_\Delta^2 & & \\ & & & & \ddots & \\ 0 & & & & & \sigma^2 \end{bmatrix} \tag{3.56}$$

and also assume normality of the error terms.

<u>Theorem 3.10</u>

Under the assumptions stated above

$$\delta_i \sim N(0,\ \sigma_\Delta^2 + \frac{\sigma^2}{1-h_{ii}}) \quad . \tag{3.57}$$

Proof

Given in Appendix A.

Since the square of a normal variable is a χ^2 variable with one degree of freedom (see Graybill (1976)), it turns out that

$$\frac{\delta_i^2}{\sigma_\Delta^2 + \frac{\sigma^2}{1-h_{ii}}} = \frac{\delta_i^2(1-h_{ii})}{\sigma^2 + \sigma_\Delta^2(1-h_{ii})} \sim \chi_1^2 \quad . \tag{3.58}$$

In the proof of theorem 3.5 it was shown that

$$(n-p-1) \cdot \frac{s_{-i}^2}{\sigma^2} \sim \chi_{N-p-1}^2 \quad . \tag{3.59}$$

This is also true for the variance inflation outlier model.

Theorem 3.11

$$\frac{\delta_i^2(1-h_{ii})}{s_{-i}^2} \sim \left(1 + \frac{\sigma_\Delta^2(1-h_{ii})}{\sigma^2}\right) \cdot F_{1,N-p-1} \quad . \tag{3.60}$$

Proof

It is known that the ratio of independent χ^2-variables, each divided by its degrees of freedom, is distributed as an F variable with degrees of freedom, n_1 which is the degrees of freedom of the numerator χ^2 variable and n_2, which is the denominator degrees of freedom. (See Graybill (1976).) Ther if δ_i and s_{-i}^2 are independent,

$$\frac{\delta_i^2(1-h_{ii})}{\sigma^2+\sigma_\Delta^2(1-h_{ii})} \cdot \frac{\sigma^2}{s_{-i}^2}$$

$$= \frac{\delta_i^2(1-h_{ii})}{s_{-i}^2} \cdot \frac{\sigma^2}{\sigma^2+\sigma_\Delta^2(1-h_{ii})} \sim F_{1,N-p-1} \text{ ,}$$

which can be written as

$$\frac{\delta_i^2(1-h_{ii})}{s_{-i}^2} \sim \frac{\sigma^2+\sigma_\Delta^2(1-h_{ii})}{\sigma^2} F_{1,N-p-1}$$

$$= (1+\frac{\sigma_\Delta^2}{\sigma^2}(1-h_{ii})) \cdot F_{1,N-p-1} \text{ .} \tag{3.61}$$

It was shown in the proof of Theorem 3.5 that the PRESS residual, δ_i, and s_{-i}^2 are independent and thus, the result holds.

The statistic given in (3.61) can then be used to test the null hypothesis, H_0: $\sigma_\Delta^2 = 0$. It can be shown that this statistic is equal to $(\text{RSTUDENT})^2$. To see this, note that the PRESS residual, δ_i, can be written as

$$\frac{\hat{\varepsilon}_i}{1-h_{ii}} \text{ ,} \tag{3.62}$$

where $\hat{\varepsilon}_i$ is the ordinary residual. Substituting (3.62) into the statistic in (3.61) gives

$$\frac{\delta_i^2(1-h_{ii})}{s_{-i}^2} = \frac{\hat{\varepsilon}_i^2(1-h_{ii})}{(1-h_{ii})^2 s_{-i}^2}$$

$$= \frac{\hat{\varepsilon}_i^2}{s_{-i}^2(1-h_{ii})}$$

$$= (\text{RSTUDENT}_i)^2 \quad . \tag{3.63}$$

Under H_0: $\sigma_\Delta^2 = 0$, $(\text{RSTUDENT}_i)^2$ has an F distribution with 1 and N-p-1 degrees of freedom. This test is equivalent to the test of H_0: $\phi = 0$ given in Section 3.3. Thus, for the chemical data, there is again evidence that an outlier has occurred at observation 13.

However, it is more appropriate to test the null hypothesis that the outlier is not "harmful". The appropriate hypotheses are, then:

$$H_0: \frac{\sigma_\Delta^2}{\sigma^2} \le \frac{1}{1-h_{ii}} \qquad H_1: \frac{\sigma_\Delta^2}{\sigma^2} > \frac{1}{1-h_{ii}} \ . \tag{3.64}$$

From (3.61) it can be seen that under the null hypothesis just stated,

$$(\text{RSTUDENT}_i)^2 \sim 2 \cdot F_{1,N-p-1} \ . \tag{3.65}$$

The null hypothesis in (3.64) should be rejected then if $(\text{RSTUDENT}_i)^2$ is greater than or equal to twice the $1-\alpha$ percentage point for the F distribution with 1 and N-p-1 degrees of freedom. For the chemical data $2 \cdot F_{1,5,.99} = 2(10) = 20$ and $(\text{RSTUDENT}_{13})^2 = 16.11$. Thus, there is again no evidence that the magnitude of the outlier is large enough to be considered counterproductive.

As in the case of the mean shift outlier, it has been demonstrated that a test to determine whether the outlier has occurred is not the appropriate test to make and can, in fact, lead to incorrect decisions on the part of the experimenter.

Again, it is not being suggested that the decision to eliminate any data point be made on the basis of a statistical test alone. From this research, however, we see that in situations where a formal test is desired either the test developed in Section 3.3 for the mean shift model or the test developed here for the variance inflation model are the appropriate ones to be used.

It is also clear from the results presented here that the rules-of-thumb currently suggested in the literature as cut-offs for RSTUDENT (such as the + or −2 recommended by Belsley, Kuh and Welsch (1980)) are inappropriate. A more reasonable guideline would be that if RSTUDENT is 3 or larger the data point should be seriously investigated.

3.6.2 Power of Test

The power of the test developed in Section 3.6.1 is

$$1 - \beta = P(\text{Rejecting } H_0 \mid H_0 \text{ is false})$$

$$= P\left((\text{RSTUDENT}_i)^2 \geq 2 \cdot F_{1,N-p-1,1-\alpha} \;\middle|\; \frac{\sigma^2_\Delta}{\sigma^2} > \frac{1}{1-h_{ii}}\right)$$

$$= P[(1+\frac{\sigma_\Delta^2(1-h_{ii})}{\sigma^2})\cdot F_{1,N-p-1} \geq 2\ F_{1,N-p-1,1-\alpha} \left| \frac{\sigma_\Delta^2}{\sigma^2} > \frac{1}{1-h_{ii}}\right.]$$

$$= P[F_{1,N-p-1} \geq (2\ F_{1,N-p-1,1-\alpha})/(1+\frac{\sigma_\Delta^2(1-h_{ii})}{\sigma^2})] \ , \qquad (3.66)$$

where $\sigma_\Delta^2/\sigma^2 > 1/(1-h_{ii})$.

As in the case of the mean shift outlier test, it can be seen that the power of this test is an increasing function of σ_Δ^2 and a decreasing function of h_{ii}.

3.7 Multiple Outliers

The results of the preceeding sections will now be extended to the case in which multiple outliers may occur. Several criteria which are natural extensions of those developed in Sections 3.2 and 3.5 will be developed for both the variance inflation and the mean shift outlier models. In this case the experimenter is interested in determining if the set of outliers as a whole is damaging to the analysis rather than determining if each point, individually, is damaging.

At this point it is necessary to define some notation to be used. Let m be the number of outliers in the data set and for simplicity it will be assumed that m is known. Let I be an m-dimensional vector of subscripts used to index the outlying observations. The matrix, X, will be partitioned such that the first m points are the points where the outliers occur. So, X_I denotes the submatrix of X where the outliers occur and $X_{(-I)}$ will denote the remaining $N-m$ points in X. The response vector, $\underline{y}$, will be similarly partitioned. The following notation will also be used:

$$\hat{\underline{\beta}}_{(-I)} = (X'_{(-I)}X_{(-I)})^{-1}X'_{(-I)}\underline{y}_{(-I)} \tag{3.67}$$

$$\hat{\underline{y}}_{(-I)} = X\hat{\underline{\beta}}_{(-I)} \tag{3.68}$$

$$H_I = X_I(X'X)^{-1}X'_I \tag{3.69}$$

and

$$H_{(-I)} = X_I(X'_{(-I)}X_{(-I)})^{-1}X'_I \quad . \tag{3.70}$$

In this notation, then, $\hat{\underline{\beta}}_{(-I)}$ is the vector of coefficients estimated with the m outlying observations deleted. The N-dimensional vector, $\hat{\underline{y}}_{(-I)}$, is the vector of estimated responses found with the outliers deleted. H_I is the $m \times m$ submatrix of H corresponding to the intersection of the rows and columns for the m outlying points. The $m \times m$ matrix $H_{(-I)}$ contains variances and covariances (apart from σ^2) for the m predicted responses indexed by I found with the m outlying observations deleted. Clearly, these quantities are the natural extensions of the special case of m equal to one.

3.7.1 Variance Inflation Outlier Model

In the general case of the variance inflation outlier model m points will have variances inflated by the same amount, σ^2_Δ. The model to be fit in this case is the usual general linear model

$$\underline{y} = X\underline{\beta} + \underline{\varepsilon} \quad . \tag{3.71}$$

The assumptions on the error vector, $\underline{\varepsilon}$, are

$$E(\underline{\varepsilon}) = \underline{0}$$

$$\mathrm{Var}(\underline{\varepsilon}) = V = \begin{bmatrix} \sigma^2+\sigma_\Delta^2 & & & & & 0 \\ & \ddots & & & & \\ & & \sigma^2+\sigma_\Delta^2 & & & \\ & & & \sigma^2 & & \\ & & & & \ddots & \\ 0 & & & & & \sigma^2 \end{bmatrix} \begin{array}{l} \left.\vphantom{\begin{matrix} a \\ a \\ a \end{matrix}}\right\} \text{m points} \\ \left.\vphantom{\begin{matrix} a \\ a \\ a \end{matrix}}\right\} \text{N-m points} \end{array} \tag{3.72}$$

The least squares estimator, $\hat{\underline{\beta}}$, is obviously unbiased for $\underline{\beta}$ with variance-covariance matrix given as follows:

$$\begin{aligned} \mathrm{Var}(\hat{\underline{\beta}}) &= \mathrm{Var}[(X'X)^{-1}X'\underline{y}] \\ &= (X'X)^{-1}X' \mathrm{Var}(\underline{y})X(X'X)^{-1} \\ &= (X'X)^{-1}X'VX(X'X)^{-1} \\ &= (X'X)^{-1}[\sigma^2(X'X) + \sigma_\Delta^2(X_I'X_I)](X'X)^{-1} \\ &= \sigma^2(X'X)^{-1}(X'X)(X'X)^{-1} + \sigma_\Delta^2(X'X)^{-1}X_I'X_I(X'X)^{-1} \\ &= \sigma^2(X'X)^{-1} + \sigma_\Delta^2(X'X)^{-1}X_I'X_I(X'X)^{-1} \quad . \end{aligned} \tag{3.73}$$

Two of the criteria developed in the single outlier case will be used here to determine under what conditions the inflation in variance is large enough to be damaging to the analysis. The criteria are

- Sum of the variances of coefficients
- Sum of the prediction variances .

The development here is completely analogous to the single outlier case.

As in the single outlier case the sum of the variances of the coefficients is obtained by taking the trace of the variance-covariance matrix in (3.73) which gives

$$\begin{aligned}\mathrm{Trace}[\mathrm{Var}(\hat{\underline{\beta}})] &= \mathrm{Trace}[\sigma^2(X'X)^{-1} + \sigma_\Delta^2(X'X)^{-1}X_I'X_I(X'X)^{-1}] \\ &= \sigma^2\ \mathrm{Trace}(X'X)^{-1} + \sigma_\Delta^2\ \mathrm{Trace}[(X'X)^{-1}X_I'X_I(X'X)^{-1}] \\ &= \sigma^2\ \mathrm{Trace}(X'X)^{-1} + \sigma_\Delta^2\ \mathrm{Trace}[X_I(X'X)^{-2}X_I'] \ . \qquad (3.74)\end{aligned}$$

Notice that the trace of the matrix in the second term of (3.74) is the analog to the pseudo-hat diagonal found in the single outlier case.

The least squares estimator found with the outlying points deleted, as given in (3.67) is, clearly, unbiased with variance-covariance matrix

$$\mathrm{Var}(\hat{\underline{\beta}}_{(-I)}) = \sigma^2(X_{(-I)}'X_{(-I)})^{-1} \ . \qquad (3.75)$$

The multivariate extension of the deletion formula given in (3.1) is

$$(X_{(-I)}'X_{(-I)})^{-1} = (X'X)^{-1} + (X'X)^{-1}X_I'(I-H_I)^{-1}X_I(X'X)^{-1} \ , \qquad (3.76)$$

where X_I, $X_{(-I)}$ and H_I are as previously defined. (For details regarding the derivation of this formula see Belsley, Kuh and Welsch (1980).) The trace of the matrix in (3.75) gives the sum of the variances of the coefficients estimated with the $N-m$ "good" points. Using the relationship given in (3.76) that sum can be found as follows:

$$\text{Trace}[\text{Var}(\hat{\underline{\beta}}_{(-I)})]$$

$$= \sigma^2 \text{ Trace}[(X'_{(-I)}X_{(-I)})^{-1}]$$

$$= \sigma^2 \text{ Trace}[(X'X)^{-1} + (X'X)^{-1}X_I'(I-H_I)^{-1}X_I(X'X)^{-1}]$$

$$= \sigma^2 \text{ Trace}(X'X)^{-1} + \sigma^2 \text{ Trace}[(X'X)^{-1}X_I'(I-H_I)^{-1}X_I(X'X)^{-1}]$$

$$= \sigma^2 \text{ Trace}(X'X)^{-1} + \sigma^2 \text{ Trace}[X_I(X'X)^{-2}X_I'(I-H_I)^{-1}] \quad . \tag{3.77}$$

Note that the analog to the pseudo-hat diagonal occurs in this sum as it does in the sum given in (3.74). It is evident that the sum of the variances of the coefficients is inflated by the outliers in the data. It is also clear that deleting the outlying observations increases the sum of the variances. Thus, the two sums must be compared as in the single outlier case to determine when it is best to delete the outliers from the analysis. This comparison is made in the following theorem.

Theorem 3.12

$$\text{Trace}[\text{Var}(\hat{\underline{\beta}})] > \text{Trace}[\text{Var}(\hat{\underline{\beta}}_{(-I)})] \text{ iff}$$

$$\frac{\sigma_\Delta^2}{\sigma^2} > \frac{\text{Trace}[X_I(X'X)^{-2}X_I'(I-H_I)^{-1}]}{\text{Trace}[X_I(X'X)^{-2}X_I']} \quad . \tag{3.78}$$

Proof

$$\text{Trace}[\text{Var}(\hat{\underline{\beta}})] > \text{Trace}[\text{Var}(\hat{\underline{\beta}}_{(-I)})]$$

$$\iff \sigma^2 \text{ Trace}(X'X)^{-1} + \sigma_\Delta^2 \text{ Trace}[X_I(X'X)^{-2}X_I']$$

$$> \sigma^2 \text{ Trace}(X'X)^{-1} + \sigma^2 \text{ Trace}[X_I(X'X)^{-2}X_I'(I-H_I)^{-1}]$$

$$\iff \sigma_\Delta^2 \text{ Trace}[X_I(X'X)^{-2}X_I'] > \sigma^2 \text{ Trace}[X_I(X'X)^{-2}X_I'(I-H_I)^{-1}]$$

$$\iff \frac{\sigma_\Delta^2}{\sigma^2} > \frac{\text{Trace}[X_I(X'X)^{-2}X_I'(I-H_I)^{-1}]}{\text{Trace}[X_I(X'X)^{-2}X_I']} \quad .$$

It can be easily demonstrated that the condition stated in (3.78) reduces to that given in (3.51) in the special case of m equal to one. As in the single outlier case, the experimenter will not know if this condition holds in practice. However, a statistical test can be developed which will aid the experimenter in the decision-making process.

It is obvious from the preceding development that the estimated response vector, $\hat{\underline{y}}$, is unbiased and has variance-covariance matrix

$$\text{Var}(\hat{\underline{y}}) = \sigma^2 X(X'X)^{-1}X' + \sigma_\Delta^2 X(X'X)^{-1}X_I'X_I(X'X)^{-1}X' \quad . \qquad (3.79)$$

The sum of the prediction variances is given by the trace of the matrix in (3.79) as follows:

$\text{Trace}[\text{Var}(\hat{\underline{y}})]$

$$= \text{Trace}[\sigma^2 X(X'X)^{-1}X' + \sigma_\Delta^2 X(X'X)^{-1}X_I'X_I(X'X)^{-1}X']$$

$$= \sigma^2 \text{ Trace}[X(X'X)^{-1}X'] + \sigma_\Delta^2 \text{ Trace}[X(X'X)^{-1}X_I'X_I(X'X)^{-1}X']$$

$$= \sigma^2 \text{ Trace}[X'X(X'X)^{-1}] + \sigma_\Delta^2 \text{ Trace}[X_I(X'X)^{-1}X'X(X'X)^{-1}X_I']$$

$$= \sigma^2 \text{ Trace}[I_p] + \sigma_\Delta^2 \text{ Trace}[X_I(X'X)^{-1}X_I']$$

$$= \sigma^2 p + \sigma_\Delta^2 \text{ Trace}[H_I] \quad . \tag{3.80}$$

Note that the $\text{Trace}[H_I]$ is the multivariate analog of the hat diagonal, h_{ii}. It is a measure of the joint leverage exerted by the m outlying observations. If m is one then obviously, $\text{Trace}[H_I]$ is equal to the hat diagonal.

The vector of estimated responses found with the m points indexed by I deleted is given in (3.68). This vector is unbiased and we see from (3.75) that it has variance-covariance matrix

$$\text{Var}(\hat{\underline{y}}_{(-I)}) = \sigma^2 X(X_{(-I)}'X_{(-I)})^{-1}X' \ . \tag{3.81}$$

Substituting (3.76) into (3.81) and taking the trace gives the sum of the variances as follows:

$\text{Trace}[\text{Var}(\hat{\underline{y}}_{(-I)})]$

$$= \sigma^2 \text{ Trace}[X(X'X)^{-1}X' + X(X'X)^{-1}X_I'(I-H_I)^{-1}X_I(X'X)^{-1}X']$$

$$= \sigma^2 \text{ Trace}[X(X'X)^{-1}X'] + \sigma^2 \text{ Trace}[X(X'X)^{-1}X_I'(I-H_I)^{-1}X_I(X'X)^{-1}X']$$

$$= \sigma^2 \text{ Trace}[X'X(X'X)^{-1}] + \sigma^2 \text{ Trace}[X_I(X'X)^{-1}X'X(X'X)^{-1}X_I'(I-H_I)^{-1}]$$

$$= \sigma^2 \text{ Trace}[I_p] + \sigma^2 \text{ Trace}[X_I(X'X)^{-1}X_I'(I-H_I)^{-1}]$$

$$= \sigma^2 p + \sigma^2 \text{ Trace}[H_I(I-H_I)^{-1}] \quad . \tag{3.82}$$

The leverage measure, $X_I(X'X)^{-1}X_I'$, also affects the sum of the variances of the estimated responses found with the m points deleted, as it does the sum of variances of the usual least squares estimator of the response.

The outliers are considered to be causing undue harm if the sum of prediction variances found with all N points is larger than that found with only the $N-m$ non-outlying points. The following theorem gives the conditions under which this occurs:

Theorem 3.13

$$\text{Trace}[\text{Var}(\hat{\underline{y}})] > \text{Trace}[\text{Var}(\hat{\underline{y}}_{(-I)})] \text{ iff}$$

$$\frac{\sigma_\Delta^2}{\sigma^2} > \frac{\text{Trace}[H_I(I-H_I)^{-1}]}{\text{Trace}[H_I]} \quad . \tag{3.83}$$

Proof

$$\text{Trace}[\text{Var}(\hat{\underline{y}})] > \text{Trace}[\text{Var}(\hat{\underline{y}}_{(-I)})]$$

$$\Longleftrightarrow \sigma^2 p + \sigma_\Delta^2 \,\text{Trace}[H_I] > \sigma^2 p + \sigma^2 \,\text{Trace}[H_I(I-H_I)^{-1}]$$

$$\Longleftrightarrow \sigma_\Delta^2 \,\text{Trace}[H_I] > \sigma^2 \,\text{Trace}[H_I(I-H_I)^{-1}]$$

$$\Longleftrightarrow \frac{\sigma_\Delta^2}{\sigma^2} > \frac{\text{Trace}[H_I(I-H_I)^{-1}]}{\text{Trace}[H_I]} \quad .$$

Note that in the special case of m equal to 1 (a single outlier), the condition in (3.83) and that given in (3.78) for the sum of variances of coefficients do reduce to the same condition, which is the condition found in Section 3.5.

3.7.2 Test Procedure for Variance Inflation Model

A test procedure similar to that developed in Section 3.6 will be developed which can be applied using either condition (3.78) or (3.83). Since the procedure is being set forth as a diagnostic aid, it would be best to perform the test for both conditions and compare results. This will be demonstrated by means of an example using the data in Table 3. This data was received as part of a consulting project in the Statistical Consulting Laboratory.

The data in this example is from blood samples in which the content of a particular enzyme in the sample was being studied. This enzyme occurs in the blood after a heart attack has occurred. The experiment involved the control of pH, denoted x_1, and the activity coefficient (which is related to the concentration of the enzyme), denoted by x_2. In one case the response variable, y_1, is the range of the peaks on the response spectrum. That is, the difference between the highest and lowest peaks. The second response, y_2, is the correlation coefficient for the calibration curve for the data. The chemist is interested in maximizing both the range and the correlation. Notice that this data is from a rotatable central composite design with two center points.

The test statistic for this procedure will be based on PRESS-type residuals for the points indexed by I. These residuals are

$$\underline{\delta}_I = \underline{y}_I - X_I\hat{\underline{\beta}}_{(-I)} \quad . \tag{3.84}$$

The expectation of these residuals is clearly 0 and the variance-covariance matrix for $\underline{\delta}_I$ is found as follows:

$$\begin{aligned}
\mathrm{Var}(\underline{\delta}_I) &= \mathrm{Var}[\underline{y}_I - X_I\hat{\underline{\beta}}_{(-I)}] \\
&= \mathrm{Var}(\underline{y}_I) + X_I \; \mathrm{Var}(\hat{\underline{\beta}}_{(-I)})X_I' \\
&= (\sigma^2 + \sigma_\Delta^2)I_m + \sigma^2 X_I(X_{(-I)}'X_{(-I)})^{-1}X_I \\
&= (\sigma^2 + \sigma_\Delta^2)I_m + \sigma^2 H_{(-I)} \\
&= \sigma^2(I_m + H_{(-I)}) + \sigma_\Delta^2 I_m \quad .
\end{aligned} \tag{3.85}$$

Notice that if m is one this reduces to the variance of the PRESS residual given in (3.57), as one would expect.

Table 3

Blood Data

obs	x_1	x_2	y_1	y_2
1	-1.000	-1.000	0.1718	0.214
2	-1.000	1.000	0.0815	0.639
3	1.000	-1.000	0.0368	0.857
4	1.000	1.000	0.0183	0.554
5	0.000	1.404	0.1082	0.045
6	0.000	-1.404	0.0521	0.988
7	1.404	0.000	0.0116	0.809
8	-1.404	0.000	0.1369	0.477
9	0.000	0.000	0.0226	0.711
10	0.000	0.000	0.0151	0.740

x_1 = pH of blood sample

x_2 = Activity coefficient of enzyme

y_1 = Range of the enzyme

y_2 = Correlation coefficient for calibration curve

For ease of notation let

$$R_1 = \frac{\text{Trace}[X_I(X'X)^{-2}X_I'(I-H_I)^{-1}]}{\text{Trace}[X_I(X'X)^{-2}X_I']} \tag{3.86}$$

and

$$R_2 = \frac{\text{Trace}[H_I(I-H_I)^{-1}]}{\text{Trace}[H_I]} \quad . \tag{3.87}$$

R_1 and R_2 are the bounds for the conditions given in (3.78) and (3.83), respectively. The experimenter would like to discover whether or not there is evidence that the ratio of variances, σ_Δ^2/σ^2, is larger than the bounds given in (3.86) and (3.87). If there is evidence of this, based on the limited experimental data, the m outliers can be considered detrimental to the analysis, and elimination of the points should be considered. Thus, the hypotheses of interest are

$$H_0: \frac{\sigma_\Delta^2}{\sigma^2} \le R_i \qquad H_1: \frac{\sigma_\Delta^2}{\sigma^2} > R_i \qquad i = 1, 2 \tag{3.88}$$

The test statistic given in the following theorem will be used to test these hypotheses.

Theorem 3.14

Assume that the correct model for the data to be fit is the general linear model, $\underline{y} = X\underline{\beta} + \underline{\varepsilon}$, with the assumptions on $\underline{\varepsilon}$ as given in (3.72). Also assume normality of the error terms. Let $s^2_{(-I)}$ be the error mean square obtained by fitting the data with the m outliers deleted. Then, under the null hypotheses stated in (3.88)

$$T_i = \frac{\underline{\delta}_I'[(I_m + H_{(-I)}) + R_i(I_m)]^{-1}\underline{\delta}_I}{ms^2_{(-I)}} \qquad i = 1, 2 \tag{3.89}$$

has an F distribution with m and N-p-m degrees of freedom.

Proof

Given in Appendix A.

Notice that the test of

$$H_0: \ \sigma^2_\Delta = 0 \quad \text{vs} \quad H_1: \ \sigma^2_\Delta \neq 0 \tag{3.90}$$

can be made with the same test statistic. Under the null hypothesis given in (3.90) the test statistic reduces to

$$T_0 = \frac{\underline{\delta}_I'[I_m + H_{(-I)}]^{-1}\underline{\delta}_I}{ms^2_{(-I)}} \tag{3.91}$$

which also has an F distribution with m and N-p-m degrees of freedom.

As a diagnostic tool, the test statistic can be calculated for all $\binom{N}{m}$ subsets of data points. However, if N is large, this can become too time consuming and costly to be practical. Frequently the experimenter has prior knowledge of areas where outliers may occur and can limit the diagnostics to those points. For instance, due to limitations in equipment, often outliers are a problem at the axial points of a central composite design. In that case the experimenter knows which two points to check as a possible problem set of points.

Without such prior knowledge, multiple point diagnostics can be burdensome and their interpretation confusing.

The situation just described is the case for the data in Table 3. The experimenter suspected that the axial points for x_2 would be susceptible to outliers in both y_1, the range data, and y_2, the correlation data. In this case, then, m is equal to two.

The residuals, hat diagonals and RSTUDENT values for the range data and the correlation data are given in Tables 4 and 5, respectively. It can be observed from the RSTUDENT values that none of the points, individually, seems to warrant elimination. In particular, notice the RSTUDENT values for observations 5 and 6, the axial points for x_2. For both y_1 and y_2 the RSTUDENT values for these observations are between two and three, which are not large enough to indicate a harmful outlier.

For the range data it turns out that $R_1 = 3.6452$ and $R_2 = 3.16162$. The hypotheses of interest for the coefficient criterion are

$$H_0\colon\ \frac{\sigma_\Delta^2}{\sigma^2} \le 3.6452 \qquad H_1\colon\ \frac{\sigma_\Delta^2}{\sigma^2} > 3.6452 \tag{3.92}$$

and for the prediction criterion

$$H_0\colon\ \frac{\sigma_\Delta^2}{\sigma^2} \le 3.16162 \qquad H_1\colon\ \frac{\sigma_\Delta^2}{\sigma^2} > 3.16162\ . \tag{3.93}$$

The null hypotheses in (3.92) and (3.93) should be rejected if the test statistic (T_i) exceeds $F_{2,2,.95} = 19.0$. The test statistic for the hypotheses in (3.92) is $T_1 = 10.2141$ and for the hypotheses in

Table 4

Diagnostics for Range Data

obs	Residual	RSTUDENT	Hat Diagonal
1	.025395	1.3737	0.6286
2	-.021441	-1.0675	0.6286
3	.020503	1.0046	0.6286
4	-.026333	-1.4594	0.6286
5	.033835	2.4811	0.6214
6	-.032884	-2.2848	0.6214
7	.00396	0.1670	0.6214
8	-.003001	-0.1266	0.6214
9	.003736	0.1369	0.4999
10	-.003764	-0.1379	0.4999

Table 5

Diagnostics for Correlation Data

obs	Residual	RSTUDENT	Hat Diagonal
1	-.198117	-1.3976	0.6286
2	.165554	1.0677	0.6286
3	-.177005	-1.1739	0.6286
4	.186667	1.2712	0.6286
5	-.253216	-2.2692	0.6214
6	.264834	2.5912	0.6214
7	-.009228	-0.0502	0.6214
8	.020846	0.1136	0.6214
9	-.014667	-0.0694	0.4999
10	.014333	0.0679	0.4999

(3.93) is $T_2 = 11.1735$. These tests indicate that the magnitude of the variance inflation is not large enough to be damaging and thus, the two points should remain in the analysis. However, it turns out that T_0, the test statistic for the hypotheses given in (3.90) is 28.9596. Since this does exceed the critical value, 19.0, there is evidence that $\sigma_\Delta^2 > 0$; however, there is no evidence that σ_Δ^2 is large enough to be detrimental to the analysis. Thus, it is clear in this case as in the single outlier case that the test of H_0: $\sigma_\Delta^2 = 0$ is inappropriate and can lead to incorrect decisions regarding the elimination of data.

The correlation data will demonstrate another phenomenon which can occur in studying outliers. As mentioned previously, none of the observations when considered individually seem to be problematic, including observations 5 and 6 which are of interest here. Since the same design was used for this response as for the range data, the hypotheses are the same as stated in (3.92) and (3.93). The critical value, 19.0, remains unchanged as well. It turns out that the test statistics for the hypotheses in (3.92) is $T_1 = 35.6838$ and for the hypotheses in (3.93) is $T_2 = 39.0349$. Thus, in this case for both the coefficient and prediction criteria the test statistics exceed the critical value for the test. (It turns out that the test statistic for H_0: $\sigma_\Delta^2 =0$ is 101.145.) There is evidence in this case that the two points together are damaging to the analysis, even though singly they were not. The experimenter, then, should at least examine observations 5 and 6 to determine if further action is warranted.

As in the single outlier case, the test procedure described here is being recommended for diagnostic purposes. It is clear, at this point, that the appropriate hypotheses to test, should formal tests be desired, are those given in (3.88). To apply these tests, even as a diagnostic aid, to all $\binom{N}{m}$ subsets of points is, in most circumstances, impractical. Thus, their use should be limited to situations described in the example. In that case, not only was m known, but the experimenter also had prior knowledge of the points that should be placed under scrutiny.

The non-null distribution of the statistic given in Theorem 3.14 has not been determined. However, it can likely be approximated by a non-central F variable. The power of the test then could likely be found, at least approximately.

3.7.3 Mean Shift Outlier Model

In the general case of the mean shift outlier model m points have mean shifts, $\phi_1, \phi_2, \ldots, \phi_m$. In the development to follow it will be assumed that the m shifts may all be different. However, a statistical test will be developed for the special case of all shifts being equal.

The general mean shift outlier model (as given in Cook and Weisberg (1982)) is

$$\underline{y} = X\underline{\beta} + D\underline{\phi} + \underline{\varepsilon} \ , \qquad (3.94)$$

where D is an $N \times m$ matrix in which the i^{th} column is an index vector, $\underline{d}_i$, with a one in the i^{th} position and zeros elsewhere and $\underline{\phi}$ is the m-dimensional vector of mean shifts, $\phi_1, \ldots, \phi_m$. The usual least squares assumptions are made on $\underline{\varepsilon}$, i.e.,

$$E(\underline{\varepsilon}) = \underline{0}$$

and (3.95)

$$\mathrm{Var}(\underline{\varepsilon}) = \sigma^2 I \quad .$$

The usual least squares estimator of the coefficients, $\underline{\hat{\beta}}$, has expectation as follows:

$$\begin{aligned} E(\underline{\hat{\beta}}) &= E[(X'X)^{-1}X'\underline{y}] \\ &= (X'X)^{-1}X'E(\underline{y}) \\ &= (X'X)^{-1}X'[X\underline{\beta} + D\underline{\phi}] \\ &= (X'X)^{-1}X'X\underline{\beta} + (X'X)^{-1}X'D\underline{\phi} \\ &= \underline{\beta} + (X'X)^{-1}X_I'\underline{\phi} \quad . \end{aligned} \tag{3.96}$$

The estimated coefficients have bias induced by the outliers, as one would expect. The bias vector,

$$\mathrm{Bias}(\underline{\hat{\beta}}) = (X'X)^{-1}X_I'\underline{\phi} \quad , \tag{3.97}$$

is easily seen to be an extension of the bias induced by a single outlier which is given in (2.28). The variance-covariance matrix for $\underline{\hat{\beta}}$ is not affected by the outliers and thus is given by

$$\mathrm{Var}(\underline{\hat{\beta}}) = \sigma^2(X'X)^{-1} \quad . \tag{3.98}$$

In order to determine when these mean shift outliers have a detrimental affect on the analysis, two of the criteria developed in Section 3.2 will be extended to take into account the multiple outliers. These criteria are analogous to the variance criteria developed for the general variance inflation model. The criteria in this case are

- Sum of the mean square errors of coefficients
- Sum of the mean square errors of prediction .

As in the case of the single outlier, these criteria will be used to weigh the effects of the outliers against the effects of deleting the points.

The mean square error matrix for $\underline{\hat{\beta}}$ in the case of multiple outliers is defined as in (3.3) for the single outlier case. From (3.97) and (3.98) it turns out that

$$\mathrm{MSE}(\underline{\hat{\beta}}) = \sigma^2(X'X)^{-1} + (X'X)^{-1}X_I'\underline{\phi}\underline{\phi}'X_I(X'X)^{-1} \quad . \tag{3.99}$$

The sum of the mean square errors will again be obtained as the trace of the mean square error matrix as follows:

$$\begin{aligned}
\mathrm{Trace}[\mathrm{MSE}(\underline{\hat{\beta}})] &= \mathrm{Trace}[\sigma^2(X'X)^{-1} + (X'X)^{-1}X_I'\underline{\phi}\underline{\phi}'X_I(X'X)^{-1}] \\
&= \sigma^2\ \mathrm{Trace}(X'X)^{-1} + \mathrm{Trace}[(X'X)^{-1}X_I'\underline{\phi}\underline{\phi}'X_I(X'X)^{-1}] \\
&= \sigma^2\ \mathrm{Trace}(X'X)^{-1} + \mathrm{Trace}[\underline{\phi}'X_I(X'X)^{-2}X_I'\underline{\phi}] \\
&= \sigma^2\ \mathrm{Trace}(X'X)^{-1} + \underline{\phi}'X_I(X'X)^{-2}X_I'\underline{\phi} \quad . \qquad (3.100)
\end{aligned}$$

The multiple outlier analog to the pseudo-hat diagonal, $X_I(X'X)^{-2}X_I'$, is again playing a role in determining the effect of the outliers.

The effect of deleting the m outlying observations will be the same as in the variance inflation case. Thus, the sum of the mean square errors of coefficients for the estimator with the m points deleted is that given in (3.77), which is

$$\text{Trace}[\text{MSE}(\hat{\underline{\beta}}_{(-I)})]$$
$$= \sigma^2 \text{ Trace}(X'X)^{-1} + \sigma^2 \text{ Trace}[X_I(X'X)^{-2}X_I'(I-H_I)^{-1}] \; . \quad (3.101)$$

The conditions under which the outliers are damaging to the analysis are described in the following theorem.

Theorem 3.15

$$\text{Trace}[\text{MSE}(\hat{\underline{\beta}})] > \text{Trace}[\text{MSE}(\hat{\underline{\beta}}_{(-I)})] \quad \text{iff}$$

$$\frac{\underline{\phi}'X_I(X'X)^{-2}X_I'\underline{\phi}}{\sigma^2} > \text{Trace}[X_I(X'X)^{-2}X_I'(I-H_I)^{-1}] \; . \qquad (3.102)$$

Proof

$$\text{Trace}[\text{MSE}(\hat{\underline{\beta}})] > \text{Trace}[\text{MSE}(\hat{\underline{\beta}}_{(-I)})]$$

$$\Longleftrightarrow \sigma^2 \text{ Trace}(X'X)^{-1} + \underline{\phi}'X_I(X'X)^{-2}X_I'\underline{\phi}$$

$$> \sigma^2 \text{ Trace}(X'X)^{-1} + \sigma^2 \text{ Trace}[X_I(X'X)^{-2}X_I'(I-H_I)^{-1}]$$

$$\Longleftrightarrow \frac{\underline{\phi}'X_I(X'X)^{-2}X_I'\underline{\phi}}{\sigma^2} > \text{Trace}[X_I(X'X)^{-2}X_I'(I-H_I)^{-1}] \quad .$$

Notice that here again if there is only one outlier in the data this condition reduces to that given in Theorem 3.1. If this condition holds, then, the m outliers are deteriorating the analysis and should be removed. The test to be developed for the special case of $\phi_1 = \phi_2 = \ldots = \phi_m$ will aid the experimenter in deciding whether this is the case or not.

The mean square error matrix for $\hat{\underline{y}}$ can easily be obtained from the mean square error matrix for $\hat{\underline{\beta}}$ obtained above. From (3.96) and (3.98) it is obvious that the expectation and variance of $\hat{\underline{y}}$ are as follows:

$$E(\hat{\underline{y}}) = X\underline{\beta} + X(X'X)^{-1}X_I'\underline{\phi} \tag{3.103}$$

and

$$\mathrm{Var}(\hat{\underline{y}}) = \sigma^2 X(X'X)^{-1}X' \ . \tag{3.104}$$

The bias in prediction induced by the m outliers is then

$$\mathrm{Bias}(\hat{\underline{y}}) = X(X'X)^{-1}X_I'\underline{\phi} \ . \tag{3.105}$$

From (3.104) and (3.105) the mean square error matrix for $\hat{\underline{y}}$ is

$$\mathrm{MSE}(\hat{\underline{y}}) = \sigma^2 X(X'X)^{-1}X' + X(X'X)^{-1}X_I'\underline{\phi}\underline{\phi}'X_I(X'X)^{-1}X' \ . \tag{3.106}$$

Taking the trace of this matrix to produce the sum of the mean square errors gives

Trace[MSE($\hat{\underline{y}}$)]

$$= \text{Trace}[\sigma^2 X(X'X)^{-1}X' + X(X'X)^{-1}X_I'\underline{\phi}\underline{\phi}'X_I(X'X)^{-1}X']$$

$$= \sigma^2\ \text{Trace}[X'X(X'X)^{-1}] + \text{Trace}[\underline{\phi}'X_I(X'X)^{-1}X'X(X'X)^{-1}X_I'\underline{\phi}]$$

$$= \sigma^2\ \text{Trace}[I_p] + \underline{\phi}'X_I(X'X)^{-1}X_I'\underline{\phi}$$

$$= \sigma^2 p + \underline{\phi}'H_I\underline{\phi} \quad . \tag{3.107}$$

Notice here again the hat diagonals and off diagonals for the m outliers are determining their affect on the analysis.

The estimated response vector for the case with the m points deleted, $\hat{\underline{y}}_{(-I)}$, has the same properties as in the variance inflation model. Thus the sum of mean square errors of prediction for $\hat{\underline{y}}_{(-I)}$ is given in (3.82). The damaging effect of the outliers is greater than that of deleting the points if the sum of the variances in (3.107) is greater than that given in (3.82). The theorem to follow gives the conditions for which the analysis is benefitted from elimination of the outliers.

Theorem 3.16

$$\text{Trace}[\text{MSE}(\hat{\underline{y}})] > \text{Trace}[\text{MSE}(\hat{\underline{y}}_{(-I)})] \quad \text{iff}$$

$$\frac{\underline{\phi}'H_I\underline{\phi}}{\sigma^2} > \text{Trace}[H_I(I-H_I)^{-1}] \quad . \tag{3.108}$$

Proof

$$\text{Trace}[\text{MSE}(\hat{\underline{y}})] > \text{Trace}[\text{MSE}(\hat{\underline{y}}_{(-I)})]$$

$$\Longleftrightarrow \sigma^2 p + \underline{\phi}' H_I \underline{\phi} > \sigma^2 p + \sigma^2 \text{ Trace}[H_I (I - H_I)^{-1}]$$

$$\Longleftrightarrow \underline{\phi}' H_I \underline{\phi} > \sigma^2 \text{ Trace}[H_I (I - H_I)^{-1}]$$

$$\Longleftrightarrow \frac{\underline{\phi}' H_I \underline{\phi}}{\sigma^2} > \text{Trace}[H_I (I - H_I)^{-1}] \quad .$$

Again, it is seen that this condition is a natural extension of the single outlier case. Also, note that if there is only one outlier then the conditions given in (3.102) and (3.108) are the same, as was the case in the variance inflation model. It can also be seen that if the m mean shifts are equal, i.e., $\phi_1 = \phi_2 = \ldots = \phi_m = \phi$, then, the conditions in (3.102) and (3.108) reduce to

$$\frac{\phi^2}{\sigma^2} > \frac{\text{Trace}[X_I (X'X)^{-2} X_I' (I - H_I)^{-1}]}{\underline{j}' X_I (X'X)^{-2} X_I' \underline{j}} = R_1 \tag{3.109}$$

and

$$\frac{\phi^2}{\sigma^2} > \frac{\text{Trace}[H_I (I - H_I)^{-1}]}{\underline{j}' H_I \underline{j}} = R_2 \; , \tag{3.110}$$

respectively. The vector denoted by $\underline{j}$ above is an m-dimensional vector of ones. The test to be developed will be for this special case of equal mean shifts.

As in the previous cases examined, the experimenter will not know the magnitude of ϕ. In particular, he will not know if the magnitude of ϕ is sufficient to consider the outliers damaging to the analysis, i.e., if $\phi^2/\sigma^2 > R_1$ or R_2. The test procedure is designed to give evidence as to the magnitude of the shift. The hypotheses of interest, as in the variance inflation case, are

$$H_0\colon\ \frac{\phi^2}{\sigma^2} \le R_i \qquad H_1\colon\ \frac{\phi^2}{\sigma^2} > R_i \qquad i=1,2\ . \tag{3.111}$$

The test statistic will again be based on the PRESS residuals given in (3.84). In this case the variance-covariance matrix is

$$\mathrm{Var}(\underline{\delta}_I) = \sigma^2(I_m + H_{(-I)}) \tag{3.112}$$

as seen from the development of (3.85) and the expectation is found as follows:

$$\begin{aligned} E(\underline{\delta}_I) &= E(\underline{y}_I - X_I\underline{\hat{\beta}}_{(-I)}) \\ &= E(\underline{y}_I) - X_I E(\underline{\hat{\beta}}_{(-I)}) \\ &= X_I\underline{\beta} + \phi\underline{j} - X_I\underline{\beta} \\ &= \phi\underline{j}\quad , \end{aligned} \tag{3.113}$$

where $\underline{j}$ is as previously defined. The following theorem gives the statistic for testing the null hypotheses given in (3.111).

Theorem 3.17

Assume that the correct model is

$$\underline{y} = X\underline{\beta} + \phi D\underline{j} + \underline{\varepsilon} \quad , \tag{3.114}$$

with the usual least squares assumptions on $\underline{\varepsilon}$. Also assume normality of the error terms. Let $s^2_{(-I)}$ be as previously defined. Then, under the null hypotheses given in (3.111)

$$\frac{\underline{\delta}_I'[I_m + H_{(-I)}]^{-1}\underline{\delta}_I}{ms^2_{(-I)}} \tag{3.115}$$

is distributed as a non-central F variable with m and N-p-m degrees of freedom and non-centrality parameter

$$\lambda_i = \frac{R_i}{2}\,\underline{j}'[I_m + H_{(-I)}]^{-1}\underline{j} \qquad i = 1, 2 \ . \tag{3.116}$$

Proof

Given in Appendix A.

It is suggested that this test be performed for both hypotheses given in (3.111) as described in the variance inflation case. As seen in the example given in that case, the results for the two conditions will in most instances be nearly the same. Thus, the same conclusion would be reached whether using the condition for the coefficients or the predictions.

The non-null distribution of this test statistic is also a non-central F. The normal approximation to the non-central F distribution can then be used to approximate the power of this test.

The procedures developed in this chapter are intended for use as aids in detecting outliers which have a detrimental affect on the analysis. It has been demonstrated that procedures for testing that there is no outlier are not geared to the goals of the analysis and thus are inappropriate and can be misleading.

It has been suggested in the case of a single outlier to use a cut-off of at least + or -3 for RSTUDENT values. That is, if RSTUDENT is larger than three in magnitude, it may indicate that the observation is counterproductive and needs further investigation. The same rule-of-thumb can be used regardless of the model chosen for the outlier.

It is obviously impractical to apply the multiple outlier procedures to all subsets of the data. Thus, it is most practical to apply them when there is some knowledge of how many and where the outliers might occur. In this case the two models give similar results and lead to the same conclusions in most cases. Since the experimenter will not know which model is truly appropriate, he should select the one which is least complicated to apply. The variance inflation model requires the use of only central F values and not non-central F values, and thus fewer complications occur.

In both the single outlier case and in the multiple outlier case, these procedures should not be used as the final decision maker in determining whether or not to discard data. That decision should only be made in collaboration with those who are the experts involved in the experimentation.

Another approach to dealing with outliers will be addressed in the next chapter. In RSM, the experimenter can take advantage of the design capabilities to safeguard against outliers. That is, the experimenter can choose a design which will be insensitive to outliers should they occur. Then, if outliers do occur the analysis is not greatly affected. Chapter IV presents results for first order designs and second order designs will be presented in Chapter V.

IV. ROBUSTNESS AND FIRST ORDER DESIGNS

In the preceeding chapter the problem of outliers was addressed from the standpoint of diagnosing and dealing with them as they occur in the data. The detrimental effects of outliers can be limited by proper choice of design. The analysis is rendered relatively insensitive to the occurrence of outliers. In this chapter first order designs which are robust to outliers will be investigated.

A second problem of concern to the analyst is that of missing data. It is not unusual in design situations to be left with an incomplete set of data to be analyzed at the end of the experiment. Data can be lost for any number of reasons; for instance, samples can be destroyed or contaminated and mechanical malfunctions or human error can render the data unusable. The analyst is then left with a "design" which is different from the one originally intended. Depending upon the original design choice and the particular design points which are lost, the analysis may be substantially affected. In many circumstances, then, it would be wise for the experimenter to choose a design which is robust to missing data. Of course if it turns out that no data is lost in the course of the experiment, then the designs should still have desirable properties. Such designs will be found in this chapter for the case of first order models.

Finally, a problem which occurs more often than not in response surface studies is that of errors in control. This occurs when the intended levels of the design variables can not be attained exactly and, in fact, the true levels actually reached are unknown. These errors can occur as the result of such things as mechanical limitations in equipment, or human limitations in reading instruments. Box (1963) studied the effects of these errors on the usual least squares estimates. These results will be used in this chapter to obtain first order designs which are robust to errors in control.

The next section will include these three types of design robustness for the first order model involving a single design variable. The last section will extend these results to the case of k variables.

4.1 First Order Designs for $k=1$

The first order model for the special case of a single design variable is

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \qquad (i=1,2,\ldots,N) \tag{4.1}$$

which is in the form of the general linear model of (2.4). In this case the X matrix is an $N \times 2$ array of the form

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ \vdots & \vdots \\ 1 & x_N \end{bmatrix} . \tag{4.2}$$

The usual assumptions, excluding normality, are made on the error terms.

4.1.1 Robustness to Outliers

Let us assume that a single outlier may occur at any one of the design points. Since the experimenter does not know a priori where the outlier will occur (or in fact, even if it will occur) he would like to choose a design which limits the effect of the outlier no matter where it occurs.

The mean shift outlier model given in (2.27) will be used in representing the occurrence of an outlier. Recall, that under this model the prediction at an arbitrary point in the design space is biased. A design will be considered robust to outliers if the average prediction bias, as given in (3.31), does not become too large no matter where the outlier might occur. Let ISB_i denote the integrated squared bias in prediction due to an outlier at the i^{th} design point. Then, from (3.31)

$$ISB_i = \frac{NK}{\sigma^2} \int_R \text{Bias}^2(\hat{y}(\underline{x}))d\underline{x}$$

$$= \frac{N\phi^2}{\sigma^2} \text{ Trace}[\mu(X'X)^{-1}\underline{x}_i\underline{x}_i'(X'X)^{-1}] \; . \tag{4.3}$$

For a given design there will then be N possible ISB_i's, one for each of the points at which the outlier could occur. The experimenter would like to ensure that none of the N ISB_i's is large. This can be achieved by minimizing over designs the maximum ISB_i, where the maximum is over locations in the design. The ISB_i for the case of one design variable is given in the following theorem:

Theorem 4.1

Given the mean shift outlier model for a single independent variable and design region, R = (-1,1), then

$$ISB_i = \frac{N\phi^2}{\sigma^2}\left[\frac{1}{N^2} + \frac{x_i^2}{3\{N[11]\}^2}\right] \quad . \tag{4.4}$$

Proof

Without loss of generality assume the first moment, [1] = 0.

$$ISB_i = \frac{N\phi^2}{\sigma^2}\,\mathrm{Trace}[\mu(X'X)^{-1}\underline{x}_i\underline{x}_i'(X'X)^{-1}]$$

$$= \frac{N\phi^2}{\sigma^2}\,\underline{x}_i'(X'X)^{-1}\mu(X'X)^{-1}\underline{x}_i$$

Note that

$$\underline{x}\underline{x}' = \begin{bmatrix} 1 & x \\ x & x^2 \end{bmatrix}$$

and thus

$$\mu = K^{-1}\int_{-1}^{1} \underline{x}\underline{x}'d\underline{x} = \begin{bmatrix} 1 & 0 \\ 0 & 1/3 \end{bmatrix} \quad .$$

Also note that

$$(X'X)^{-1} = \begin{bmatrix} \frac{1}{N} & 0 \\ 0 & \frac{1}{N[11]} \end{bmatrix}$$

and thus

$$(X'X)^{-1}\mu(X'X)^{-1} = \begin{bmatrix} \frac{1}{N^2} & 0 \\ 0 & \frac{1}{3\{N[11]\}^2} \end{bmatrix}$$

Since $\underline{x}_i = (1\ x_i)$ then

$$ISB_i = \frac{N\phi^2}{\sigma^2}(1\ x_i)\begin{bmatrix} \frac{1}{N^2} & 0 \\ 0 & \frac{1}{3\{N[11]\}^2} \end{bmatrix}\begin{bmatrix} 1 \\ x_i \end{bmatrix}$$

$$= \frac{N\phi^2}{\sigma^2}\{\frac{1}{N^2} + \frac{x_i^2}{3\{N[11]\}^2}\} \quad .$$

The design which is outlier robust, then, is that giving $\underset{D}{\text{minimax}}(\underset{i}{ISB_i})$. The optimal design is described in the following theorem.

Theorem 4.2

For a first order model in one variable and design region, $R = (-1,1)$, $\underset{D}{\text{minimax}}(\underset{i}{ISB_i})$ is obtained for N even with N/2 points at 1 and N/2 points at -1 and for N odd with (N-1)/2 points at 1, (N-1)/2 points at -1 and one point at 0.

Proof

Assume N is even. It is obvious from (4.4) that ISB_i is a maximum with respect to locations in the design when x_i is the point in the design with the largest magnitude. Let x_{max} represent that point. Then,

$$\max(ISB_i) = \frac{N\phi^2}{\sigma^2}\left[\frac{1}{N^2} + \frac{x^2_{max}}{3\{N[11]\}^2}\right] .$$

This expression is minimized with respect to design by minimizing $x^2_{max}/3\{N[11]\}^2$ w.r.t. [11]. Recall that $N[11] = \sum_{u=1}^{N} x^2_u$. Thus,

$$\frac{x^2_{max}}{3\{N[11]\}^2} = \frac{x^2_{max}}{3\{\sum_{u=1}^{N} x^2_u\}^2}$$

and $x^2_{max} \geq x^2_u$ for all u. Therefore,

$$\frac{x^2_{max}}{3\{\sum_{u=1}^{N} x^2_u\}^2} \geq \frac{x^2_{max}}{3\{Nx^2_{max}\}^2} = \frac{1}{3N^2x^2_{max}} .$$

Thus, we see that

$$\frac{x^2_{max}}{3\{N[11]\}^2} \geq \frac{1}{3N^2x^2_{max}} ,$$

with exact equality being obtained when all x^2_u are equal and thus the minimum occurs when all x^2_u equal one. The design giving $\underset{D}{\text{minimax}}(ISB_i)$, then, is the design with N/2 points at +1 and N/2 points at -1.

The proof for N odd is similar to that for N even and thus will not be given here.

The choice of the design levels at the endpoints of the design region, that is, at +1 and -1, is a somewhat arbitrary choice. From the proof above it is obvious that the best design has the levels as far from the center as possible. In practice, however, there is a reasonable limit on the levels based on the appropriateness of the first order model and thus the endpoints of the design region are usually used.

Myers (1976) shows that this design also minimizes the integrated prediction variance and therefore, minimizes the *integrated mean square error* of prediction. Since this design is optimal in terms of its variance properties (it also minimizes the variances of coefficients), which are unaffected by the presence of an outlier, then the outlier robust design is also optimal if no outlier occurs. It turns out that this design is also the outlier robust design if the variance inflation model is used to represent the occurrence of an outlier rather than the mean shift outlier model. The details are similar to that of the mean shift model and thus will not be given here.

4.1.2 Robustness to Missing Data

As seen in Chapter III, when data is dropped from the analysis for some reason, it causes the variances of the coefficients and predictions to become inflated. In choosing a design robust to missing data the amount of the variance inflation is reduced. In this case the effects of losing (or dropping) a single data point will be investigated and only the effect on prediction variances will be considered in finding robust designs.

In (3.26) it is seen that the prediction variance (apart from σ^2) at an arbitrary point $\underline{x}$ is

$$\frac{\mathrm{Var}(\hat{y}(x))}{\sigma^2} = \underline{x}'(X'X)^{-1}\underline{x} \quad . \tag{4.5}$$

From (3.35) and (3.1) the corresponding variance with the i^{th} point deleted from the analysis is (apart from σ^2)

$$\frac{\mathrm{Var}_{-i}(\hat{y}(\underline{x}))}{\sigma^2} = \underline{x}'(X'X)^{-1}\underline{x} + \frac{\underline{x}'(X'X)^{-1}\underline{x}_i\underline{x}_i'(X'X)^{-1}\underline{x}}{1-h_{ii}} \quad . \tag{4.6}$$

As mentioned previously since the second term in (4.6) will be positive, the prediction variance at an arbitrary point in the design space is inflated by losing or dropping a point.

As in the outlier case it will not be known prior to experimentation which point will be lost during the course of the experiment. So the experimenter should choose a design in this case which will result in the analysis being minimally affected if a point is lost, no matter which point in the design it might be. As in the outlier case the design should have good properties even if no data is lost and the entire design is used in the analysis.

The optimality criterion to be used will be based on the average prediction variance with the i^{th} point removed, where the average is obtained by integrating over the design region. Let V_{-i} denote the integrated prediction variance with the i^{th} point deleted. Then,

$$\begin{aligned} V_{-i} &= \frac{NK}{\sigma^2}\int_R \mathrm{Var}_{-i}(\hat{y}(x))d\underline{x} \\ &= \frac{NK}{\sigma^2}\int_R \underline{x}'(X'X)^{-1}\underline{x}d\underline{x} + \frac{NK}{\sigma^2(1-h_{ii})}\int_R \underline{x}'(X'X)^{-1}\underline{x}_i\underline{x}_i'(X'X)^{-1}\underline{x}d\underline{x} \quad . \end{aligned} \tag{4.7}$$

Recall that K^{-1} is the volume of the design region. From (3.36), then,

$$V_{-i} = N \text{ Trace}[\mu(X'X)^{-1}] + \frac{N}{1-h_{ii}} \text{ Trace}[\mu(X'X)^{-1}\underline{x}_i\underline{x}_i'(X'X)^{-1}] \text{ , } \quad (4.8)$$

where μ is the region moment matrix defined in Chapter II.

It is obvious that there will be a set of N V_{-i}'s, one for each point in the design. The optimal design will be the design which minimizes the maximum V_{-i} in the design. The following theorem gives V_{-i} for the special case under consideration here, $k = 1$ and a first order model:

Theorem 4.3

Given a first order model with a single design variable and design region $R = (-1,1)$, then

$$V_{-i} = 1 + \frac{1}{3[11]} + \frac{N}{1-h_{ii}} [\frac{1}{N^2} + \frac{x_i^2}{3\{N[11]\}^2}] \quad . \qquad (4.9)$$

Proof

Without loss of generality assume $[1] = 0$. In the proof of Theorem 4.1 the region moment matrix was found to be

$$\mu = \begin{bmatrix} 1 & 0 \\ 0 & 1/3 \end{bmatrix}$$

and

$$(X'X)^{-1} = \begin{bmatrix} \frac{1}{N} & 0 \\ 0 & \frac{1}{N[11]} \end{bmatrix} .$$

So, the first term in (4.8) is found as follows:

$$N \text{ Trace}[\mu(X'X)^{-1}] = N \text{ Trace} \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1/3 \end{bmatrix} \begin{bmatrix} \frac{1}{N} & 0 \\ 0 & \frac{1}{N[11]} \end{bmatrix} \right\}$$

$$= 1 + \frac{1}{3[11]} . \tag{4.10}$$

The second term in (4.8) is found as follows:

$$\frac{N}{1-h_{ii}} \text{ Trace}[\mu(X'X)^{-1}\underline{x}_i\underline{x}_i'(X'X)^{-1}]$$

$$= \frac{N}{1-h_{ii}} \text{ Trace} \left\{ \begin{bmatrix} \frac{1}{N} & 0 \\ 0 & \frac{1}{3N[11]} \end{bmatrix} \begin{bmatrix} 1 & x_i \\ x_i & x_i^2 \end{bmatrix} \begin{bmatrix} \frac{1}{N} & 0 \\ 0 & \frac{1}{N[11]} \end{bmatrix} \right\}$$

$$= \frac{N}{1-h_{ii}} \left[\frac{1}{N^2} + \frac{x_i^2}{3\{N[11]\}^2}\right] . \tag{4.11}$$

Combining (4.10) and (4.11) gives the result stated in the theorem.

The design which is robust to missing data is that which gives minimax(V_{-i}), where the maximum is taken over locations in the design and the minimum is with respect to designs. This design is found in the following theorem:

Theorem 4.4

The first order design for one independent variable with design region R = (-1,1) which gives minimax(V_{-i}) is for N even the design with N/2 points at 1 and N/2 points at -1 and for N odd has (N-1)/2 points at 1, (N-1)/2 points at -1 and one point at 0.

Proof

Since the proofs for N even and N odd are similar only the proof for N even will be given. To find the design giving minimax(V_{-i}) first consider only the last term in (4.9):

$$\frac{N}{1-h_{ii}}\left[\frac{1}{N^2} + \frac{x_i^2}{3\{N[11]\}^2}\right] \quad . \tag{4.12}$$

The design which minimizes the maximum value of (4.12) can be found in two stages. First the design must be found which minimizes the maximum value of the term in brackets. It will, then, be shown that the same design minimizes the maximum h_{ii}. (Recall that $0 \leq h_{ii} \leq 1$.) The design which minimizes the maximum value of the term in brackets was found in Theorem 4.2 to be the design with N/2 points at 1 and N/2 points at -1. To find the design which gives minimax(h_{ii}) notice that in this case

$$h_{ii} = \frac{1}{N} + \frac{x_i^2}{N[11]} \quad . \tag{4.13}$$

The maximum h_{ii} for a given design occurs when x_i is the endpoint of the design. Let this point be denoted by x_{max}. So,

$$\mathrm{Max}(h_{ii}) = \frac{1}{N} + \frac{x^2_{max}}{N[11]} \quad .$$

To minimize this with respect to designs only the second term need be considered. Recall that $N[11] = \sum_{u=1}^{N} x_u^2$. So,

$$\frac{x^2_{max}}{N[11]} = \frac{x^2_{max}}{\sum_{u=1}^{N} x_u^2} \quad .$$

Since x_{max} is the endpoint of the design $x_u^2 \le x^2_{max}$ for all u, and thus

$$\frac{x^2_{max}}{\sum_{u=1}^{N} x_u^2} \geq \frac{x^2_{max}}{Nx^2_{max}} = \frac{1}{N} \quad ,$$

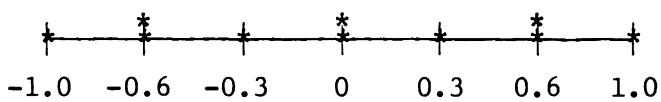
with exact equality (and thus the minimum) occurring when all x_u^2 are equal. So, the design which gives minimax(h_{ii}) is the design with all x_u^2 equal, i.e., a two-level design, regardless of the second moment, [11]. Thus, the design with N/2 points at 1 and N/2 points at -1 minimizes the maximum value of (4.12).

Observing the first part of (4.9), 1 + 1/3[11], it is clear that this term is independent of locations in the design and Myers (1976) shows that this is minimized by placing half of the points at each endpoint. Then, the design which is robust to missing data is the design with N/2 points at 1 and N/2 points at -1.

The design which gives robustness to missing data, then, is the same design that gives robustness to outliers. This is also the design which is optimal in terms of variances of coefficients and integrated prediction variance.

4.1.3 Example

An example will illustrate the potential benefits of using the robust design. For a sample size of 10 the optimal design for the region (-1,1) will in this case have five points at +1 and five at -1. The +1 and -1 are arbitrary and in practice would be taken to be as large as the first order model would allow. This design will be referred to as Design 1. Design 2, which will be compared with the optimal design, is given by the following diagram, in which the design points are indicated:



So, one observation is made at -1.0, -0.3, 0.3 and 1.0 and two observations are made at each of -0.6, 0 and 0.6. The integrated prediction variances with one point at a time removed, the V_{-i}, were calculated (apart from σ^2) for the two designs. All V_{-i} (and thus the maximum) for the optimal design are equal to 1.5. The maximum V_{-i} for Design 2 is 2.49 and the minimum is 2.06. Thus, no matter which point is lost the optimal design always performs better than Design 2. Also, the optimal design is better than Design 2 even if no points are lost. The integrated prediction variance (apart from σ^2) with all ten points included in the analysis is 1.3 for Design 1 and 1.9 for Design 2. If a point is lost from the optimal design then there is a 15% increase in the integrated prediction variance. If a point is lost from Design 2 the increase is between 8% and 31% depending upon the point lost. It is clear that Design 1 is always preferred to Design 2 and Design 2 can be much worse than the optimal design.

4.1.4 Center Points

In most practical situations the experimenter will want to include several center points in the design. From the results above it is clear that a design with center points will not be optimal in terms of its robustness. To determine the effect of center points consider the integrated squared bias given in (4.4). The optimal design will have N/2 points at +1 and N/2 points at -1. The second moment for this design will be one and all ISB_i will be equal. It is easily seen that ISB_i will equal 4/3N (apart from ϕ^2/σ^2) for the optimal design.

Now, rather than putting all N points at the extremes of the design, suppose we allow $N_1/2$ points at +1 and -1 and N_0 points at the center. The total sample size is still $N_1 + N_0 = N$. In this case the second moment is N_1/N and the maximum ISB_i will occur at the endpoints. It turns out, then, that the maximum ISB_i (apart from ϕ^2/σ^2) is given by

$$\text{Max}(ISB_i) = \frac{1}{N} + \frac{N}{3N_1^2} \qquad (4.14)$$

which is larger than 4/3N. The difference between (4.14) and the optimum, 4/3N, is

$$D = \frac{1}{N} + \frac{N}{3N_1^2} - \frac{4}{3N}$$

$$= \frac{N}{3N_1^2} - \frac{1}{3N}$$

$$= \frac{N^2 - N_1^2}{3N_1^2 N} \quad . \tag{4.15}$$

Obviously as long as N_1^2 is close to N^2, that is, few center points are used, the difference between the robustness of the two designs will be small.

To illustrate the effect of center points on robustness several examples will be given.

Example 1: For a sample size of ten, minimax(ISB_i) = 4/30 = 0.1333, apart from ϕ^2/σ^2. If two center points are used in the design ($N_1 = 8$), then, the max(ISB_i) in (4.14) turns out to be 0.1521 resulting in a 14% increase over the optimal design.

If four center points are included ($N_1 = 6$), then the Max(ISB_i) is 0.1926 and the increase is 44% over the optimum.

Example 2: If the sample size is sixteen then minimax(ISB_i) = 0.0833, apart from ϕ^2/σ^2. If two center points are used in this case $N_1 = 14$, and Max(ISB_i) is 0.0897, which is an increase of 7.6% over the optimum. For $N_1 = 12$ there is a 19% difference between the two designs.

Thus, we see that the use of center points can seriously affect the robustness of the design to outliers. However, if the number of center points is small relative to the total sample size the robustness of the design will not be unduly affected.

The same is true of robustness to missing data. For the optimal design $h_{ii} = p/N$ which in this case is $2/N$ and, again, $[11] = 1$. Then, from (4.9) it turns out that

$$V_{-i} = 1 + \frac{1}{3} + \frac{N^2}{N-2}\left[\frac{1}{N^2} + \frac{1}{3N^2}\right]$$

$$= \frac{4}{3}\left[1 + \frac{1}{N-2}\right]$$

$$= \frac{4}{3}\left[\frac{N-1}{N-2}\right] \quad . \tag{4.16}$$

If center points are used in the design then recall that $[11] = N_1/N$ and from (4.13) the maximum h_{ii} is

$$\max(h_{ii}) = \frac{1}{N} + \frac{1}{N_1} \quad . \tag{4.17}$$

Then, for a design with center points the maximum V_{-i}, which occurs at the endpoints, is

$$\max(V_{-i}) = 1 + \frac{N}{3N_1} + \frac{N^2N_1}{N_1N-N_1-N}\left[\frac{1}{N^2} + \frac{N^2}{3N^2N_1}\right]$$

$$= 1 + \frac{N}{3N_1} + \frac{N_1}{N_1N-N_1-N}\left[1 + \frac{N^2}{3N_1^2}\right] \quad . \tag{4.18}$$

For example 1 above in which a sample size of ten is used,

$$\text{minimax}(V_{-i}) = \frac{4}{3}\left[\frac{9}{8}\right] = 1.5$$

and if two center points are used $\max(V_{-i})$ turns out to be 1.6129. This is an increase of 7.5% over the optimal design. For the case of four center points $\max(V_{-i})$ is 1.6819, a 12% increase over the optimum.

For the sample size of sixteen given in example 2 $\text{minimax}(V_{-i})$ = 1.4286. In this case if two center points are used there is a 3.7% increase in the $\max(V_{-i})$ and a 9.3% increase if four center points are used in the design.

Thus, the same trends are seen as in the outlier case. However, the effect of the use of center points is somewhat less severe in the case of missing data. We see, then, that center points can be included in the design without unduly affecting the robustness as long as the number of center points is small relative to the total sample size.

4.1.5 Robustness to Errors in Control

Although in a response surface study the levels of the independent variables can be controlled, many times the desired levels of the variables can not be obtained exactly. In that case the u^{th} observation for the first order model with a single independent variable can be defined as:

$$y_u = \beta_0 + \beta_1\theta_u + \varepsilon_u \quad . \tag{4.19}$$

The θ_u are the true, but unknown, values of the independent variable obtained in the experiment. Let X_u denote the aimed at or intended levels of the variable. Then,

$$X_u = \theta_u + \delta_u , \tag{4.20}$$

where δ_u represents the error in control of the independent variable for the u^{th} run.

The model in (4.19) can then be rewritten as follows:

$$\begin{aligned} y_u &= \beta_0 + \beta_1(X_u - \delta_u) + \varepsilon_u \\ &= \beta_0 + \beta_1 X_u + (\varepsilon_u - \beta_1 \delta_u) \\ &= \beta_0 + \beta_1 X_u + Z_u \quad . \end{aligned} \tag{4.21}$$

The error term, Z_u, includes the random, or residual, error and the error due to the inability to obtain the specified levels of the design variable.

In the results to be developed it will be assumed that $E(\delta_u) = 0$, $Var(\delta_u) = \sigma_\delta^2$ and the errors in control, the δ's, are uncorrelated from run to run. The usual least squares assumptions will be made on the residual error terms, the ε_u's, and also assume that $E(\delta_u \varepsilon_u) = \rho\sigma\sigma_\delta$.

Box (1963) states that the error term, Z_u, has expectation zero and its variance is constant from run to run. This can easily be seen from the following:

$$
\begin{aligned}
E(Z_u) &= E(\varepsilon_u - \beta_1 \delta_u) \\
&= E(\varepsilon_u) - \beta_1 E(\delta_u) \\
&= 0 - \beta_1 \cdot 0 \\
&= 0 \qquad (4.22)
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{Var}(Z_u) &= \mathrm{Var}(\varepsilon_u - \beta_1 \delta_u) \\
&= \mathrm{Var}(\varepsilon_u) + \beta_1^2 \,\mathrm{Var}(\delta_u) - 2\beta_1 \,\mathrm{Cov}(\varepsilon_u, \delta_u) \\
&= \sigma^2 + \beta_1^2 \sigma_\delta^2 - 2\beta_1 \rho \sigma \sigma_\delta \ . \qquad (4.23)
\end{aligned}
$$

From (4.23) it is clear that $\mathrm{Var}(Z_u)$ is independent of u and thus remains constant across runs. So, although the error variance is inflated by the errors in control it does not affect the estimation since the inflated variance is homogeneous across runs and since no bias is induced by the errors. The variance-covariance matrix of the usual least squares estimator of the coefficiensts is then equal to

$$\mathrm{Var}(\hat{\underline{\beta}}) = \sigma_\Delta^2 (X'X)^{-1} \ , \qquad (4.24)$$

where, $\sigma_\Delta^2 = \sigma^2 + \beta_1^2 \sigma_\delta^2 - 2\beta_1 \rho \sigma \sigma_\delta$. The variance-covariance matrix for $\hat{\underline{y}}$, the vector of predicted responses, is given by

$$\mathrm{Var}(\hat{\underline{y}}) = \sigma_\Delta^2 X(X'X)^{-1} X' \ , \qquad (4.25)$$

which is the usual variance-covariance matrix with the residual variance, σ^2, replaced by the variance of the overall error term.

The prediction variance at an arbitrary point, $\underline{x}$, in the design space is therefore

$$\mathrm{Var}(\hat{y}(\underline{x})) = \sigma_\Delta^2 \underline{x}'(X'X)^{-1}\underline{x} \quad . \tag{4.26}$$

It is clear from this development that the design chosen has no affect on the amount that the error variance is inflated due to the errors in control. Since no bias is incurred from the errors, the design criterion of interest is the minimum integrated prediction variance. It should be obvious at this point that the design which gives robustness to errors in control is the usual minimum variance design. This can be seen in the following development where V denotes the integrated prediction variance.

$$\begin{aligned} V &= \frac{NK}{\sigma_\Delta^2} \int_R \mathrm{Var}(\hat{y}(\underline{x}))d\underline{x} \\ &= NK \int_R \underline{x}'(X'X)^{-1}\underline{x}d\underline{x} \quad . \end{aligned} \tag{4.27}$$

From the proof of Theorem 4.3 it can be seen that

$$V = 1 + \frac{1}{3[11]} \quad . \tag{4.28}$$

The design which minimizes V is the design with N/2 points at 1 and N/2 points at -1 for N even and (N-1)/2 points at 1, (N-1)/2 points at -1 and one point at 0 for N odd.

Therefore, if errors in control are potentially a problem,as they usually are in response surface experiments, the optimal design to use for a first order model with one independent variable is the minimum variance design obtained by making [11] as large as possible.

Thus, it has been demonstrated that for the design region (-1,1), a single independent variable and a first order model robustness to outliers, to missing data and to errors in control are obtained with a two-level design symmetric about zero in which the two levels are taken as far from the origin as the first order model will allow. (If N is odd the extra point should be placed in the center.) It was also seen that even under ideal conditions in which none of these problems occur this design is still optimal.

4.2 First Order Designs for $k > 1$

The results of the previous section will now be generalized to the case of multiple independent variables. The first order model for k independent variables was given in (2.2) and can also be written in matrix notation as in (2.4). First order designs for multiple variables will be found which are robust to outliers, to missing data and to errors in control. As in the single variable case it will turn out that the same conditions will give all three types of robustness.

4.2.1 Robustness to Outliers

The mean shift outlier model will again be used to model the occurrence of a single outlier which may occur at any one of the design points. The goal, as before, is to limit the effect of the outlier no matter where it might occur.

The design criterion to be used is that used in the single variable case, the minimax integrated squared prediction bias. Recall that the integrated squared bias in prediction due to an outlier occurring at the i^{th} point, as given in (4.3) is

$$ISB_i = \frac{N\phi^2}{\sigma^2} \text{Trace}[\mu(X'X)^{-1}\underline{x}_i\underline{x}_i'(X'X)^{-1}] \ . \tag{4.29}$$

The design which minimizes the maximum ISB_i is given in the following theorem:

Theorem 4.5

Given a first order model and a spherical design region, then the maximum ISB_i is minimized by an orthogonal design with second moments as large as possible.

Proof

Without loss of generality assume [i] = 0. Myers (1976) gives the region moments for a spherical design region. The region moment matrix, μ, for a first order model turns out to be

$$\mu = \begin{bmatrix} 1 & & & 0 \\ & \frac{1}{k+2} & & \\ & & \ddots & \\ 0 & & & \frac{1}{k+2} \end{bmatrix} \ . \tag{4.30}$$

Since this region moment matrix is a diagonal matrix of constants, finding the design which gives minimax(ISB_i) is equivalent to finding the design which gives

$$\text{minimax}\{\text{Trace}[(X'X)^{-1}\underline{x}_i\underline{x}_i'(X'X)^{-1}]\} \quad , \qquad (4.31)$$

where the minimum is again over designs and the maximum is over locations within a particular design. Note that the portion in braces in (4.31) can be reduced as follows:

$$\begin{aligned}\text{Trace}[(X'X)^{-1}\underline{x}_i\underline{x}_i'(X'X)^{-1}] &= \text{Trace}[\underline{x}_i'(X'X)^{-2}\underline{x}_i] \\ &= \underline{x}_i'(X'X)^{-2}\underline{x}_i \quad . \qquad (4.32)\end{aligned}$$

Notice that this quantity is the pseudo-hat diagonal which was seen repeatedly in the results developed in Chapter 3. Thus, the design which results in minimax(ISB_i) is equivalent to the design resulting in minimax($\underline{x}_i'(X'X)^{-2}\underline{x}_i$).

To find this design first consider the sum of the pseudo-hat diagonals.

$$\begin{aligned}\sum_{i=1}^{N} \underline{x}_i'(X'X)^{-2}\underline{x}_i &= \sum_{i=1}^{N} \text{Trace}[\underline{x}_i'(X'X)^{-2}\underline{x}_i] \\ &= \sum_{i=1}^{N} \text{Trace}[\underline{x}_i\underline{x}_i'(X'X)^{-2}] \\ &= \text{Trace} \sum_{i=1}^{N} [\underline{x}_i\underline{x}_i'(X'X)^{-2}] \ .\end{aligned}$$

Recall that $\sum_{i=1}^{N} \underline{x}_i \underline{x}_i' = (X'X)$. Thus,

$$\sum_{i=1}^{N} \underline{x}_i'(X'X)^{-2}\underline{x}_i = \text{Trace}[(X'X)(X'X)^{-2}]$$

$$= \text{Trace}(X'X)^{-1}$$

$$= \sum_{j=0}^{k} \frac{\text{Var}(\hat{\beta}_j)}{\sigma^2} \quad . \tag{4.33}$$

It has been shown, then, that the sum of the pseudo-hat diagonals is equal to the sum of the variances of the estimated coefficients (apart from σ^2).

It is known (see Myers (1976)) that the variances of the estimated coefficients are minimized by an orthogonal design with [ii] as large as possible. So, $\sum_{i=1}^{N} ISB_i$, as well as the average ISB_i, are minimized by the same design. Note, also, that with this design all ISB_i are equal to the average ISB_i, and in particular the maximum ISB_i is equal to the average. Now, let $M_1 = \max(ISB_i)$ for the orthogonal design with second moments as large as possible and let $M_2 = \max(ISB_i)$ for any other design. If $M_2 < M_1$ then all ISB_i for the other design are smaller than M_1. Therefore, the sum of the ISB_i for the other design is smaller than the sum for the orthogonal design with large second moments, which is impossible based on the development above. Thus, M_2 must be at least as large as M_1. Therefore, the design which minimizes the maximum ISB_i is the orthogonal design with second moments as large as possible.

Examples of such optimal designs are the 2^k factorial designs, fractions of factorials that are first order orthogonal and Plackett-Burman designs.

Recall that the conditions stated in Theorem 4.5 are the same conditions that give minimum variances of coefficients. Myers (1976) also shows that these conditions give the minimum integrated prediction variance. Therefore, these optimal designs also minimize the *integrated mean square error* of prediction, as was found in the single variable case. It is also clear that these designs are optimal based on the variance properties even if no outliers occur in the data. Nothing is sacrificed, then, in obtaining robustness to outliers. It also turns out that the same designs are optimal if the outlier is modeled as a variance inflation outlier rather than a mean shift outlier.

4.2.2 Robustness to Missing Data

For the case of multiple independent variables the effects of losing a single data point will now be examined. As in the single variable case it is unknown which, if any, of the data points will be lost and so the aim will be to find designs which will minimize the effect of losing a point no matter which point in the design it might be. The optimal design will again be that which minimizes the maximum integrated prediction variance with a point deleted. The maximum is, as before, taken over locations within a particular design.

The integrated prediction variance with the i^{th} point deleted is given by V_{-i} in (4.8). The design that minimizes the maximum of the $N\,V_{-i}$ is given in the following theorem:

Theorem 4.6

Given a first order model and a spherical design region, then the maximum V_{-i} is minimized by an orthogonal design with second moment as large as possible.

Proof

Without loss of generality assume [i] = 0. The integrated prediction variance with the i^{th} data point deleted is

$$V_{-i} = N\ \text{Trace}[\mu(X'X)^{-1}] + \frac{N}{1-h_{ii}}\ \text{Trace}[\mu(X'X)^{-1}\underline{x}_i\underline{x}_i'(X'X)^{-1}]. \quad (4.34)$$

To find the design giving minimax(V_{-i}), each term in (4.34) will be considered individually. The first term is obviously independent of locations in the design. This term is the integrated prediction variance with all design points included in the analysis. As mentioned previously, Myers (1976) shows that this is minimized by an orthogonal design with second moments as large as possible.

Notice that the numerator of the second term is similar to the integrated squared bias in prediction given in (4.29). Thus, the design which minimizes the maximum value of this numerator is the design just described, the orthogonal design with large second moments. It remains to be shown, then, that the same design minimizes the maximum hat diagonal, h_{ii}. Since the hat diagonals sum to a constant, the number of model parameters, the maximum h_{ii} is minimized when they are all equal. For the first order model this occurs when the design is orthogonal.

Therefore, the design which is robust to missing data is the orthogonal design with second moments, [ii], as large as possible.

As previously mentioned the 2^k factorials, first order orthogonal fractions and Plackett-Burman designs are examples of designs that meet these criteria and thus, are also robust to missing data.

4.2.3 Example

The 2^3 factorial design has design matrix as follows:

$$D_1 = \begin{array}{c} \begin{array}{ccc} x_1 & x_2 & x_3 \end{array} \\ \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & -1 & -1 \\ -1 & 1 & 1 \\ -1 & 1 & -1 \\ -1 & -1 & 1 \\ -1 & -1 & -1 \end{bmatrix} \end{array} . \qquad (4.35)$$

The levels of +1 and -1 are arbitrarily used in this example. The actual levels used in practice should be as far from the origin as allowed by the first order model. This optimal design has integrated prediction variance (apart from σ^2), with all 8 points included, of 1.6 and the maximum V_{-i} of 2.0. In fact, in this case all V_{-i} are equal. Thus, if a point is lost from the optimal design, no matter which point it might be, the integrated prediction variance is inflated by 25%. Contrast this with a similar, but non-optimal design given by the following design matrix:

$$D_2 = \begin{array}{c} \begin{array}{ccc} x_1 & x_2 & x_3 \end{array} \\ \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ 0 & -1 & -1 \\ 0 & 1 & 1 \\ -1 & 1 & -1 \\ -1 & -1 & 1 \\ -1 & -1 & -1 \end{bmatrix} \end{array} . \tag{4.36}$$

This design has integrated prediction variance with all 8 points of 1.76, an increase of 10% over that of the optimal design. The second design has maximum V_{-i} equal to 2.34. This is an increase of 33% over its integrated prediction variance with all 8 points included in the analysis. The second design has maximum V_{-i} 17% larger than that of the optimal design. In fact, the minimum V_{-i} for the second design is 3% larger than the V_{-i} for the optimal design. Thus, no matter which point is lost the optimal design has a smaller integrated prediction variance. Depending upon exactly which point is lost the difference may be large. If no points are lost, the optimal design still has integrated prediction variance significantly smaller than that of the second design.

It has been seen that the design which is robust to missing data is the same design which is robust to outliers. The same design is also optimal in terms of the usual integrated prediction variance.

4.2.4 Center Points

As in the single variable case, center points are usually included in the experimental design. To investigate their effect, the optimal design will be compared to 2^k factorial designs augmented by N_0 center points.

For the optimal design $(X'X) = N \cdot I$ and the ISB_i, apart from ϕ^2/σ^2, can be found as follows:

$$ISB_i = N \,\mathrm{Trace}[\mu(X'X)^{-1}\underline{x}_i\underline{x}_i'(X'X)^{-1}]$$

$$= N\, \underline{x}_i'(X'X)^{-1}\mu(X'X)^{-1}\underline{x}_i$$

$$= \frac{1}{N}\, \underline{x}_i'\mu\underline{x}_i$$

$$= \frac{1}{N}\, \underline{x}_i' \begin{bmatrix} 1 & & & \\ & \frac{1}{k+2} & & \\ & & \ddots & \\ & & & \frac{1}{k+2} \end{bmatrix} \underline{x}_i$$

$$= \frac{1}{N}\left(1 + \frac{k}{k+2}\right) \quad . \tag{4.37}$$

Therefore, since all ISB_i are equal, minimax(ISB_i) is given by (4.37).

For the factorial design with center points, let N_1 be the number of factorial points. The total sample size is $N_1 + N_0 = N$. It is easily seen that in this case

$$(X'X)^{-1} = \begin{bmatrix} \frac{1}{N} & & & \\ & \frac{1}{N_1} & & \\ & & \ddots & \\ & & & \frac{1}{N_1} \end{bmatrix} . \qquad (4.38)$$

Apart from ϕ^2/σ^2 the ISB_i is found as follows:

$$ISB_i = N\,\underline{x}_i' \begin{bmatrix} \frac{1}{N} & & & \\ & \frac{1}{N_1} & & \\ & & \ddots & \\ & & & \frac{1}{N_1} \end{bmatrix} \begin{bmatrix} 1 & & & \\ & \frac{1}{k+2} & & \\ & & \ddots & \\ & & & \frac{1}{k+2} \end{bmatrix} \begin{bmatrix} \frac{1}{N} & & & \\ & \frac{1}{N_1} & & \\ & & \ddots & \\ & & & \frac{1}{N_1} \end{bmatrix} \underline{x}_i$$

$$= \underline{x}_i' \begin{bmatrix} \frac{1}{N} & & & \\ & \frac{N}{N_1^2(k+2)} & & \\ & & \ddots & \\ & & & \frac{N}{N_1^2(k+2)} \end{bmatrix} \underline{x}_i$$

$$= \frac{1}{N} + \frac{N}{N_1^2(k+2)} \sum_{j=1}^{k} x_{ji}^2 \quad . \qquad (4.39)$$

Obviously the maximum ISB_i for this design will occur at the factorial points rather than the center points and thus,

$$\max(ISB_i) = \frac{1}{N} + \frac{Nk}{N_1^2(k+2)} \quad . \tag{4.40}$$

The difference in the maximum ISB_i for the two designs is

$$\begin{aligned} D &= \frac{Nk}{N_1^2(k+2)} - \frac{k}{N(k+2)} \\ &= \frac{k(N^2-N_1^2)}{NN_1^2(k+2)} \quad . \end{aligned} \tag{4.41}$$

Notice that for $k=1$ this difference reduces to that found in (4.15). This difference will be close to zero when N_1^2 is close to N^2.

The difference in integrated squared bias has been calculated for several 2^k factorial designs plus center points. The results are given in Table 6 for $k=3$, 4 and 5 along with the percentage increase in integrated bias over the optimal design and the ratio of center points to total sample size. From these results we see that adding center points to the design can significantly reduce its robustness, especially for smaller values of k. However, as k increases (and thus N increases) center points are less detrimental to the robustness of the design.

For the missing data case, the optimal design will have

$$\begin{aligned} N\,\mathrm{Trace}[\mu(X'X)^{-1}] &= \mathrm{Trace}(\mu) \\ &= 1 + \frac{k}{k+2} \quad . \end{aligned} \tag{4.42}$$

Table 6

Differences in $\max(ISB_i)$ as a
Function of Center Points

2^3 + N_0	D	% INC	N_0/N
2	0.0338	21	0.2
4	0.0625	47	0.3333
8	0.1125	113	0.5

2^4 + N_0	D	% INC	N_0/N
2	0.0098	11	0.1111
4	0.0187	22	0.2
8	0.0347	50	0.3333

2^5 + N_0	D	% INC	N_0/N
2	0.0027	5	0.0588
4	0.0053	11	0.1111
8	0.0100	23	0.2

This design has all hat diagonals equal to p/N and thus from (4.34) it turns out that

$$\text{minimax}(V_{-i}) = 1 + \frac{k}{k+2} + \frac{1}{N(1-p/N)}\left(1 + \frac{k}{k+2}\right)$$

$$= \left(1 + \frac{k}{k+2}\right)\left(1 + \frac{1}{N-p}\right)$$

$$= \left(\frac{2(k+1)}{k+2}\right)\left(\frac{N-p+1}{N-p}\right) \quad . \qquad (4.43)$$

For a design with center points we see that

$$N \text{ Trace}[\mu(X'X)^{-1}] = \text{Trace}\begin{bmatrix} 1 & & & \\ & \frac{1}{k+2} & & \\ & & \ddots & \\ & & & \frac{1}{k+2} \end{bmatrix}\begin{bmatrix} 1 & & & \\ & \frac{N}{N_1} & & \\ & & \ddots & \\ & & & \frac{N}{N_1} \end{bmatrix}$$

$$= 1 + \frac{Nk}{N_1(k+2)} \quad . \qquad (4.44)$$

The hat diagonals for this design turn out to be

$$h_{ii} = \frac{1}{N} + \frac{1}{N_1}\sum_{j=1}^{k} x_{ji}^2 \qquad (4.45)$$

and thus the maximum hat diagonal (which occurs at the factorial points) is

$$\max(h_{ii}) = \frac{1}{N} + \frac{k}{N_1} \quad . \qquad (4.46)$$

From (4.40), (4.44) and (4.46), the maximum V_{-i} for a factorial design with center points is

$$\max(V_{-i}) = 1 + \frac{Nk}{N_1(k+2)} + \frac{NN_1}{NN_1-N_1-Nk}\left[\frac{1}{N} + \frac{Nk}{N_1^2(k+2)}\right] . \qquad (4.47)$$

For k = 3, 4 and 5 Table 7 gives the difference between (4.47) and $\text{minimax}(V_{-i})$ given in (4.43) for several values of N_0. The percentage increase and N_0/N are also given. The same trends are seen as in the outlier case, although the effect of the use of center points is less pronounced for the missing data case.

If center points are to be added to the basic design, only a relatively small number can be used without severely diminishing the robustness of the design. However, as k increases, a greater number of center points can be tolerated.

4.2.5 Robustness to Errors in Control

The u^{th} observation for the first order model in which there are errors in control can be written as

$$y_u = \beta_0 + \beta_1\theta_{1u} + \beta_2\theta_{2u} + \dots + \beta_k\theta_{ku} + \varepsilon_u , \qquad (4.48)$$

where θ_{ju} are the true levels of the design variables obtained in the experiment. Recall, however, that the θ_j's are unknown. X_{ju} will represent the intended levels of the independent variables. So

$$X_{ju} = \theta_{ju} + \delta_{ju} , \qquad (4.49)$$

Table 7

Differences in Integrated Prediction Variance
as a Function of Center Points

2^3

N_0	D	% INC	N_0/N
2	0.2143	11	0.2
4	0.4154	23	0.3333
8	0.8	45	0.5

2^4

N_0	D	% INC	N_0/N
2	0.0975	5	0.1111
4	0.1935	11	0.20
8	0.3824	22	0.3333

2^5

N_0	D	% INC	N_0/N
2	0.0480	2.7	0.0588
4	0.0958	5	0.1111
8	0.1908	11	0.2

where δ_{ju} represents the error associated with the j^{th} design variable at the u^{th} run. The following assumptions are made on the error terms:

$$E(\varepsilon_u) = E(\delta_{ju}) = 0 \quad \text{for all } j$$

$$Var(\varepsilon_u) = \sigma^2$$

$$Var(\delta_{ju}) = \sigma_j^2 \quad \text{for all } u \tag{4.50}$$

$$E(\delta_{iu}\delta_{ju}) = \rho_{ij}\sigma_i\sigma_j$$

$$E(\varepsilon_u\delta_{ju}) = \rho_{0j}\sigma\sigma_j \quad .$$

So, the usual least squares assumptions are made on the residual error terms, ε_u, the δ_j have zero expectation and constant variance for each j. It is also assumed that the errors in control may be correlated within a run, but are uncorrelated from run to run.

From (4.48) and (4.49) the model can be rewritten in terms of the aimed-at levels of the independent variables as follows:

$$y_u = \beta_0 + \beta_1(X_{1u}-\delta_{1u}) + \beta_2(X_{2u}-\delta_{2u}) + \ldots + \beta_k(X_{ku}-\delta_{ku}) + \varepsilon_u$$

$$= \beta_0 + \beta_1X_{1u} + \ldots + \beta_kX_{ku} - \beta_1\delta_{1u} - \ldots - \beta_k\delta_{ku} + \varepsilon_u$$

$$= \beta_0 + \sum_{j=1}^{k} \beta_jX_{ju} + (\varepsilon_u - \sum_{j=1}^{k} \beta_j\delta_{ju})$$

$$= \beta_0 + \sum_{j=1}^{k} \beta_jX_{ju} + Z_u \quad . \tag{4.51}$$

Thus, as in the single variable case, the overall error term is a function of the residual error, the unknown coefficients and the errors due to missing the intended levels of the variables.

It is obvious that the Z_u have expectation zero and the variance of Z_u can be found as follows:

$$\begin{aligned}
\mathrm{Var}(Z_u) &= \mathrm{Var}[\varepsilon_u - \sum_{j=1}^{k} \beta_j \delta_{ju}] \\
&= \mathrm{Var}(\varepsilon_u) + \sum_{j=1}^{k} \beta_j^2 \,\mathrm{Var}(\delta_{ju}) - 2 \sum_{j=1}^{k} \beta_j \,\mathrm{Cov}(\varepsilon_u, \delta_{ju}) \\
&\quad + 2 \sum_{i<j}\sum \beta_i \beta_j \,\mathrm{Cov}(\delta_{iu}, \delta_{ju}) \\
&= \sigma^2 + \sum_{j=1}^{k} \beta_j^2 \sigma_j^2 - 2 \sum_{j=1}^{k} \beta_j \rho_{0j} \sigma\sigma_j + 2 \sum_{i<j}\sum \rho_{ij}\sigma_i\sigma_j \,.
\end{aligned} \tag{4.52}$$

As in the single variable case, then, the variance of the overall error term, Z_u, is not a function of u. Thus,

$$\mathrm{Var}(y) = \sigma_\Delta^2 \, I_N \,, \tag{4.53}$$

where σ_Δ^2 is the expression given for $\mathrm{Var}(Z_u)$ in (4.52). As in the case of one independent variable the error variance is inflated by the errors in control and is a function of the unknown coefficients, but the estimation is unaffected by this.

The variance-covariance matrices for the estimated coefficients and predicted responses are given by (4.24) and (4.25), respectively, where in this case σ_Δ^2 is given in (4.52). Again it is clear that the amount of the inflation in variance due to the errors in control is unaffected by the design chosen.

The design criterion to be used, the integrated prediction variance, is given in (4.27), where, again, σ_Δ^2 is given in (4.52). Recall that the design which minimizes the integrated prediction variance is the orthogonal design with second moments as large as possible. The 2^k factorials, first order orthogonal fractions and the Plackett-Burman designs are again found to be robust, this time to errors in control.

Thus, it has been demonstrated that for a single independent variable the design that is robust to outliers, missing data and errors in control for the design region $R = (-1,1)$ is obtained by using two levels symmetric about zero and as far from the origin as practicable. These designs were also seen to be optimal under the ideal conditions in which none of these problems occur. For more than one independent variable the conditions that give robustness to these three problems are orthogonality and second moments as large as possible. The 2^k factorials, first order orthogonal fractions and Plackett-Burman designs are examples of robust designs. These, too, are optimal even if none of the problems described actually occur. Thus, nothing is sacrificed in obtaining robustness to these problems.

In the next chapter these results will be extended to the case of second order models. The same types of robustness will be investigated: robustness to outliers, to missing data and to errors in control. Results will be obtained by computer for the case of a single independent variable and for several types of second order designs for more than one variable. The design types to be examined are the central composite, the Box-Behnken, the small composite, the hybrid and the equiradial design.

V. ROBUSTNESS AND SECOND ORDER DESIGNS

In Chapter IV designs were found for the first order model that are robust to outliers, to missing data and to errors in control. In this chapter the same types of robustness will be investigated for the second order model given in (2.3). The criteria developed in the last chapter will be used to evaluate the second order designs. Results are obtained by computer for outlier robustness and robustness to missing data for the central composite, Box-Behnken, small composite, hybrid and equiradial designs. For each of the design classes, recommendations will be made as to the appropriate design parameters to be used to achieve robustness. Comparisons will also be made across design types where appropriate.

In examining errors in control in a second order response surface model, the u^{th} observation can be written as

$$y_u = \beta_0 + \sum_{i=1}^{k} \beta_i \theta_{iu} + \sum_{i=1}^{k} \beta_{ii} \theta_{iu}^2 + \sum_{i<j}\sum \beta_{ij} \theta_{iu} \theta_{ju} + \varepsilon_u , \qquad (5.1)$$

where, as before, the θ_{iu} represent the true levels of the independent variables obtained in the experiment. The intended levels of these variables will again be denoted by X_{iu} and differ from the actual levels obtained by an unknown error, δ_{iu}. That is,

$$X_{iu} = \theta_{iu} + \delta_{iu} \quad . \qquad (5.2)$$

The assumptions stated in (4.33) will again be made on the errors terms.

Using the relationship given in (5.2) the model in (5.1) can be rewritten as follows:

$$y_u = \beta_0 + \sum_{i=1}^{k} \beta_i (X_{iu} - \delta_{iu}) + \sum_{i=1}^{k} \beta_{ii} (X_{iu} - \delta_{iu})^2$$

$$+ \sum_{i<j}\sum \beta_{ij} (X_{iu} - \delta_{iu})(X_{ju} - \delta_{ju}) + \varepsilon_u$$

$$= \beta_0 + \sum_{i=1}^{k} \beta_i X_{iu} - \sum_{i=1}^{k} \beta_i \delta_{iu} + \sum_{i=1}^{k} \beta_{ii} (X_{iu}^2 - 2X_{iu}\delta_{iu} + \delta_{iu}^2)$$

$$+ \sum_{i<j}\sum \beta_{ij} (X_{iu}X_{ju} - X_{iu}\delta_{ju} - X_{ju}\delta_{iu} + \delta_{iu}\delta_{ju}) + \varepsilon_u$$

$$= \beta_0 + \sum_{i=1}^{k} \beta_i X_{iu} + \sum_{i=1}^{k} \beta_{ii} X_{iu}^2 + \sum_{i<j}\sum \beta_{ij} X_{iu} X_{ju}$$

$$+ \varepsilon_u - \sum_{i=1}^{k} \beta_i \delta_{iu} - 2\sum_{i=1}^{k} \beta_{ii} X_{iu} \delta_{iu} + \sum_{i=1}^{k} \beta_{ii} \delta_{iu}^2$$

$$- \sum_{i<j}\sum \beta_{ij} X_{iu} \delta_{ju} - \sum_{i<j}\sum \beta_{ij} X_{ju} \delta_{iu} + \sum_{i<j}\sum \beta_{ij} \delta_{iu} \delta_{ju}$$

$$= \beta_0 + \sum_{k=1}^{k} \beta_i X_{iu} + \sum_{i=1}^{k} \beta_{ii} X_{iu}^2 + \sum_{i<j}\sum \beta_{ii} X_{iu} X_{ju} + Z_u \,. \tag{5.3}$$

The over all error term, Z_u, then is a function of the residual error, ε_u, the errors in control, δ_{iu}, the unknown coefficients and is also a function of the design chosen, through the X_{ju}.

Unlike the first order case in which the Z_u had expectation zero, in this case the expectation of the overall error term is non-zero, as seen in the following:

$$E(Z_u) = E(\varepsilon_u) - \sum_{i=1}^{k} \beta_i E(\delta_{iu}) - 2\sum_{i=1}^{k} \beta_{ii} X_{iu} E(\delta_{iu}) + \sum_{i=1}^{k} \beta_{ii} E(\delta_{iu}^2)$$

$$- \sum_{i<j}\sum \beta_{ij} X_{iu} E(\delta_{ju}) - \sum_{i<j}\sum \beta_{ij} X_{ju} E(\delta_{iu}) + \sum_{i<j}\sum \beta_{ij} E(\delta_{iu}\delta_{ju})$$

$$= \sum_{i=1}^{k} \beta_{ii} E(\delta_{iu}^2) + \sum_{i<j}\sum \beta_{ij} E(\delta_{iu}\delta_{ju})$$

$$= \sum_{i=1}^{k} \beta_{ii}\sigma_j^2 + \sum_{i<j}\sum \beta_{ij}\sigma_i\sigma_j\rho_{ij}$$

$$= \sum_{i\le j}\sum \beta_{ij}\sigma_i\sigma_j\rho_{ij} \quad . \tag{5.4}$$

The expected value of the error term, then, is a function of the unknown coefficients; however, it is not a function of the design chosen. To examine the variance structure of the error terms, consider the simple case of one independent variable. In that case

$$Z_u = \varepsilon_u - \beta_1\delta_{1u} - 2\beta_{11}X_{1u}\delta_{1u} + \beta_{11}\delta_{1u}^2 \quad . \tag{5.5}$$

The variance of this error term is found as follows:

$$\begin{aligned}
\mathrm{Var}(Z_u) &= \mathrm{Var}(\varepsilon_u - \beta_1\delta_{1u} - 2\beta_{11}X_{1u}\delta_{1u} + \beta_{11}\delta_{1u}^2) \\
&= \mathrm{Var}(\varepsilon_u) + \beta_1^2\ \mathrm{Var}(\delta_{1u}) + 4\beta_{11}^2X_{1u}^2\ \mathrm{Var}(\delta_{1u}) \\
&\quad + \beta_{11}^2\ \mathrm{Var}(\delta_{1u}^2) - 2\beta_1\ \mathrm{Cov}(\varepsilon_u,\delta_{1u}) - 4\beta_{11}X_{1u}\ \mathrm{Cov}(\varepsilon_u,\delta_{1u}) \\
&\quad + 2\beta_{11}\ \mathrm{Cov}(\varepsilon_u,\delta_{1u}^2) + 4\beta_1\beta_{11}X_{1u}\ \mathrm{Var}(\delta_{1u}) \\
&\quad - 2\beta_1\beta_{11}\ \mathrm{Cov}(\delta_{1u},\delta_{1u}^2) - 4\beta_{11}^2X_{1u}\ \mathrm{Cov}(\delta_{1u},\delta_{1u}^2) \\
&= \sigma^2 + \beta_1^2\sigma_1^2 + 4\beta_{11}^2X_{1u}^2\sigma_1^2 + \beta_{11}^2\ \mathrm{Var}(\delta_{1u}^2) \\
&\quad - 2\beta_1\rho_{01}\sigma\sigma_1 - 4\beta_{11}X_{1u}\rho_{01}\sigma\sigma_1 + 4\beta_1\beta_{11}X_{1u}\sigma_1^2\ . \qquad (5.6)
\end{aligned}$$

If it is assumed that the δ_{1u}'s are normally distributed, then it can be shown that $\mathrm{Var}(\delta_{1u}^2) = 2\sigma_1^4$. Therefore, $\mathrm{Var}(Z_u)$ becomes

$$\begin{aligned}
\mathrm{Var}(Z_u) &= \sigma^2 + \beta_1^2\sigma_1^2 + 4\beta_{11}^2X_{1u}^2\sigma_1^2 + 2\beta_{11}^2\sigma_1^4 \\
&\quad - 2\beta_1\rho_{01}\sigma\sigma_1 - 4\beta_{11}X_{1u}\rho_{01}\sigma\sigma_1 + 4\beta_1\beta_{11}X_{1u}\sigma_1^2\ . \qquad (5.7)
\end{aligned}$$

From this it can be seen that the variance of the overall error term is a function of the coefficients, which are unknown, and also differs from run to run. The problem of finding second order designs that are robust to errors in control then becomes a problem of heterogeneous variance. The fact that the variances are functions of the unknown coefficients, however, makes solving the problem intractable. Thus, robustness to errors in control in second order designs will not be examined further.

In the next two sections, designs will be found which are robust to outliers and to missing data. In the third section, comparisons will be made between several of the design classes considered. The fourth section will examine robustness in the central composite and small composite designs when the region of interest extends beyond that defined by the factorial points. A summary of the recommendations will be given in Chapter VI.

5.1 Robustness to Outliers

The basis for examining robustness to outliers in second order designs is the same as in the first order case. Recall that the criterion developed for that case was the integrated squared bias in prediction due to an outlier occurring at the i^{th} design point. From (4.3) the integrated bias apart from ϕ^2/σ^2 is

$$ISB_i \cdot \frac{\sigma^2}{\phi^2} = N \cdot \mathrm{Trace}[\mu(X'X)^{-1}\underline{x}_i\underline{x}_i'(X'X)^{-1}] \ . \qquad (5.8)$$

The region of integration, the unit sphere, will be discussed in more detail for each design type, individually. For each type of design the optimal choice of design parameters will be found; that is, the design parameters which minimize the maximum ISB_i. To do this, the maximum ISB_i is calculated (apart from ϕ^2/σ^2) for each combination of design parameters. The special case of a single independent variable is addressed in the following section.

5.1.1 One Independent Variable

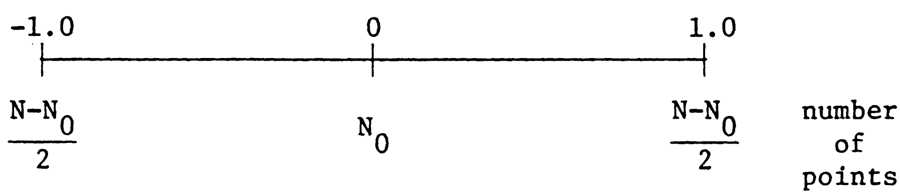
In this section the case of a quadratic model in one variable will be considered. As in the first order model, the region of interest is assumed to have been scaled to the interval (-1,1). The prediction function for this case is

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x + \hat{\beta}_{11} x^2 \quad . \tag{5.9}$$

The X-matrix in general is as follows:

$$X = \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_N & x_N^2 \end{bmatrix} \quad . \tag{5.10}$$

Two types of designs will be evaluated and compared in this case. One is the three level design with levels, -1, 0 and +1. Sample sizes of eight to twenty-five will be examined and for a given sample size the number of center points, N_0, will be varied and the remaining $N - N_0$ points will be evenly distributed between the two endpoints. The following diagram illustrates this design.



The second design type is a five-level design with levels taken to be -1, -D, 0, +D and +1, where D is allowed to vary from .1 to .9. As well as varying D for a given sample size, the distribution of the points at the five levels will also be varied. The number of center points and the total number of points at the endpoints, N_1, will be varied, with the remaining $N - N_0 - N_1$ points distributed evenly between the two interior levels. This design is illustrated in the following diagram, where N_2 denotes $N - N_0 - N_1$.

-1.0	-D	0	D	1.0	
$\frac{N_1}{2}$	$\frac{N_2}{2}$	N_0	$\frac{N_2}{2}$	$\frac{N_1}{2}$	number of points

The maximum integrated squared bias in prediction is calculated using (5.8) and the X-matrix given in (5.10) for sample sizes of 8 through 25. Sample results for the 3-level designs for $N = 8$ through 20 are given in Table 8. From these results it can be seen that the design which is most robust has approximately half of the points in the center. It is also obvious from these results that the damage due to the occurrence of an outlier can be quite severe if less than three center points are used. The same results hold for the larger sample sizes.

For the 5-level designs results are given in Table 9 for sample sizes of 8 through 15. It is impossible in this case to give specific recommendations as to the distribution of points. However, the following generalizations can be made:

1) N_0 should be less than half the total number of points;

2) N_0 should be approximately equal to the number of points at each endpoint;

3) the number of points at the extremes should be larger than the number of points at the interior levels, i.e., the non-center levels;

4) D should generally be in the range of .4 to .6.

The same computations were made for sample sizes of 16 through 25 and similar results were seen.

In comparing the 3-level and the 5-level designs, it can be seen that neither is clearly the better choice. For a given sample size the best 3-level design is at least competitive with the best 5-level design in terms of robustness to outliers, and in some cases the 3-level design is best. Table 10 gives the most robust design for a given sample size of those 3 and 5-level designs considered. These designs can be seen to follow the generalizations made above.

The generalizations made regarding these designs are intended as broad rules-of-thumb and not as hard-and-fast rules for choosing a design. In designing an experiment, then, for a quadratic regression model with a single independent variable these principles for achieving robustness should be considered and adapted to the specific situation. There are many other types of designs that could be considered for this case; however, since this is less interesting and less practical than the multiple variable case, only these two design types will be considered. In the following sections the most

Table 8

Maximum Integrated Squared Bias

k=1, 3-Level Designs

N	NO	MAXBIAS
8	2	1.0667
8	4	0.2667
8	6	1.0667
9	1	4.8000
9	3	0.5333
9	5	0.3000
9	7	1.2000
10	2	1.3333
10	4	0.3333
10	6	0.3333
10	8	1.3333
11	1	5.8667
11	3	0.6519
11	5	0.2347
11	7	0.3667
11	9	1.4667
12	2	1.6000
12	4	0.4000
12	6	0.1778
12	8	0.4000
12	10	1.6000
13	1	6.9333
13	3	0.7704
13	5	0.2773
13	7	0.1926
13	9	0.4333
13	11	1.7333
14	2	1.8667
14	4	0.4667
14	6	0.2074
14	8	0.2074
14	10	0.4667
14	12	1.8667

Table 8 (cont.)

N	NO	MAXBIAS
15	1	8.0000
15	3	0.8889
15	5	0.3200
15	7	0.1633
15	9	0.2222
15	11	0.5000
15	13	2.0000
16	2	2.1333
16	4	0.5333
16	6	0.2370
16	8	0.1333
16	10	0.2370
16	12	0.5333
16	14	2.1333
17	1	9.0667
17	3	1.0074
17	5	0.3627
17	7	0.1850
17	9	0.1417
17	11	0.2519
17	13	0.5667
17	15	2.2667
18	2	2.4000
18	4	0.6000
18	6	0.2667
18	8	0.1500
18	10	0.1500
18	12	0.2667
18	14	0.6000
18	16	2.4000
19	1	10.1333
19	3	1.1259
19	5	0.4053
19	7	0.2068
19	9	0.1251
19	11	0.1583
19	13	0.2815
19	15	0.6333
19	17	2.5333

Table 8 (cont.)

N	NO	MAXBIAS
20	2	2.6667
20	4	0.6667
20	6	0.2963
20	8	0.1667
20	10	0.1067
20	12	0.1667
20	14	0.2963
20	16	0.6667
20	18	2.6667

Table 9

Maximum Integrated Squared Bias

k=1, 5-Level Designs

N	NO	N1	D	MAXBIAS
8	2	2	0.2	0.9571
8	2	2	0.4	0.7219
8	2	2	0.6	0.4807
8	2	2	0.8	0.7334
8	2	4	0.2	0.2864
8	2	4	0.4	0.3589
8	2	4	0.6	0.5287
8	2	4	0.8	0.8309
8	4	2	0.2	1.0087
8	4	2	0.4	0.8577
8	4	2	0.6	0.6529
8	4	2	0.8	0.4329
9	1	2	0.2	1.0276
9	1	2	0.4	0.7174
9	1	2	0.6	0.5282
9	1	2	0.8	2.3242
9	1	4	0.2	0.2812
9	1	4	0.4	0.3208
9	1	4	0.6	0.7303
9	1	4	0.8	2.3818
9	1	6	0.2	0.5904
9	1	6	0.4	0.8159
9	1	6	0.6	1.4536
9	1	6	0.8	3.0807
9	3	2	0.2	1.0786
9	3	2	0.4	0.8147
9	3	2	0.6	0.5309
9	3	2	0.8	0.3992
9	3	4	0.2	0.2903
9	3	4	0.4	0.2634
9	3	4	0.6	0.3210
9	3	4	0.8	0.4437
9	5	2	0.2	1.1358
9	5	2	0.4	0.9678

Table 9 (cont.)

N	NO	N1	D	MAXBIAS
9	5	2	0.6	0.7375
9	5	2	0.8	0.4895
10	2	2	0.2	1.1440
10	2	2	0.4	0.7958
10	2	2	0.6	0.4982
10	2	2	0.8	0.8516
10	2	4	0.2	0.3132
10	2	4	0.4	0.2631
10	2	4	0.6	0.4093
10	2	4	0.8	0.8831
10	2	6	0.2	0.3590
10	2	6	0.4	0.4517
10	2	6	0.6	0.6636
10	2	6	0.8	1.0399
10	4	2	0.2	1.2000
10	4	2	0.4	0.9078
10	4	2	0.6	0.5861
10	4	2	0.8	0.3012
10	4	4	0.2	0.3230
10	4	4	0.4	0.2940
10	4	4	0.6	0.2506
10	4	4	0.8	0.2869
10	6	2	0.2	1.2629
10	6	2	0.4	1.0777
10	6	2	0.6	0.8220
10	6	2	0.8	0.5461
11	1	2	0.2	1.2055
11	1	2	0.4	0.7964
11	1	2	0.6	0.5568
11	1	2	0.8	2.6338
11	1	4	0.2	0.3353
11	1	4	0.4	0.2653
11	1	4	0.6	0.5564
11	1	4	0.8	2.4322
11	1	6	0.2	0.2642
11	1	6	0.4	0.3947

Table 9 (cont.)

N	NO	N1	D	MAXBIAS
11	1	6	0.6	0.8737
11	1	6	0.8	2.8001
11	1	8	0.2	0.7227
11	1	8	0.4	1.0001
11	1	8	0.6	1.7751
11	1	8	0.8	3.7491
11	3	2	0.2	1.2602
11	3	2	0.4	0.8754
11	3	2	0.6	0.5230
11	3	2	0.8	0.4609
11	3	4	0.2	0.3452
11	3	4	0.4	0.2908
11	3	4	0.6	0.2706
11	3	4	0.8	0.4802
11	3	6	0.2	0.2486
11	3	6	0.4	0.2966
11	3	6	0.6	0.3952
11	3	6	0.8	0.5449
11	5	2	0.2	1.3214
11	5	2	0.4	1.0010
11	5	2	0.6	0.6434
11	5	2	0.8	0.3307
11	5	4	0.2	0.3556
11	5	4	0.4	0.3244
11	5	4	0.6	0.2770
11	5	4	0.8	0.2186
11	7	2	0.2	1.3899
11	7	2	0.4	1.1875
11	7	2	0.6	0.9066
11	7	2	0.8	0.6026
12	2	2	0.2	1.3171
12	2	2	0.4	0.8642
12	2	2	0.6	0.5283
12	2	2	0.8	0.9736
12	2	4	0.2	0.3666
12	2	4	0.4	0.2902

Table 9 (cont.)

N	NO	N1	D	MAXBIAS
12	2	4	0.6	0.3447
12	2	4	0.8	0.9458
12	2	6	0.2	0.1957
12	2	6	0.4	0.2691
12	2	6	0.6	0.4872
12	2	6	0.8	1.0439
12	2	8	0.2	0.4314
12	2	8	0.4	0.5439
12	2	8	0.6	0.7982
12	2	8	0.8	1.2489
12	4	2	0.2	1.3765
12	4	2	0.4	0.9556
12	4	2	0.6	0.5572
12	4	2	0.8	0.2981
12	4	4	0.2	0.3771
12	4	4	0.4	0.3185
12	4	4	0.6	0.2425
12	4	4	0.8	0.3114
12	4	6	0.2	0.1863
12	4	6	0.4	0.2148
12	4	6	0.6	0.2698
12	4	6	0.8	0.3465
12	6	2	0.2	1.4427
12	6	2	0.4	1.0942
12	6	2	0.6	0.7016
12	6	2	0.8	0.3610
12	6	4	0.2	0.3882
12	6	4	0.4	0.3547
12	6	4	0.6	0.3033
12	6	4	0.8	0.2395
12	8	2	0.2	1.5169
12	8	2	0.4	1.2972
12	8	2	0.6	0.9911
12	8	2	0.8	0.6590
13	1	2	0.2	1.3712
13	1	2	0.4	0.8708

Table 9 (cont.)

N	NO	N1	D	MAXBIAS
13	1	2	0.6	0.6203
13	1	2	0.8	2.9512
13	1	4	0.2	0.3872
13	1	4	0.4	0.2921
13	1	4	0.6	0.4634
13	1	4	0.8	2.5245
13	1	6	0.2	0.1803
13	1	6	0.4	0.2481
13	1	6	0.6	0.6281
13	1	6	0.8	2.6733
13	1	8	0.2	0.3131
13	1	8	0.4	0.4683
13	1	8	0.6	1.0223
13	1	8	0.8	3.2388
13	1	10	0.2	0.8548
13	1	10	0.4	1.1840
13	1	10	0.6	2.0970
13	1	10	0.8	4.4193
13	3	2	0.2	1.4286
13	3	2	0.4	0.9334
13	3	2	0.6	0.5320
13	3	2	0.8	0.5246
13	3	4	0.2	0.3977
13	3	4	0.4	0.3154
13	3	4	0.6	0.2403
13	3	4	0.8	0.5206
13	3	6	0.2	0.1842
13	3	6	0.4	0.1993
13	3	6	0.6	0.3196
13	3	6	0.8	0.5645
13	3	8	0.2	0.2943
13	3	8	0.4	0.3519
13	3	8	0.6	0.4689
13	3	8	0.8	0.6458
13	5	2	0.2	1.4926
13	5	2	0.4	1.0361

Table 9 (cont.)

N	NO	N1	D	MAXBIAS
13	5	2	0.6	0.5958
13	5	2	0.8	0.2543
13	5	4	0.2	0.4089
13	5	4	0.4	0.3461
13	5	4	0.6	0.2633
13	5	4	0.8	0.2234
13	5	6	0.2	0.1883
13	5	6	0.4	0.1761
13	5	6	0.6	0.2000
13	5	6	0.8	0.2455
13	7	2	0.2	1.5640
13	7	2	0.4	1.1875
13	7	2	0.6	0.7604
13	7	2	0.8	0.3918
13	7	4	0.2	0.4208
13	7	4	0.4	0.3849
13	7	4	0.6	0.3296
13	7	4	0.8	0.2604
13	9	2	0.2	1.6438
13	9	2	0.4	1.4069
13	9	2	0.6	1.0756
13	9	2	0.8	0.7154
14	2	2	0.2	1.4785
14	2	2	0.4	0.9300
14	2	2	0.6	0.5663
14	2	2	0.8	1.0976
14	2	4	0.2	0.4177
14	2	4	0.4	0.3148
14	2	4	0.6	0.3057
14	2	4	0.8	1.0146
14	2	6	0.2	0.1946
14	2	6	0.4	0.1871
14	2	6	0.6	0.3912
14	2	6	0.8	1.0646
14	2	8	0.2	0.2290
14	2	8	0.4	0.3156

Table 9 (cont.)

N	NO	N1	D	MAXBIAS
14	2	8	0.6	0.5666
14	2	8	0.8	1.2081
14	2	10	0.2	0.5037
14	2	10	0.4	0.6359
14	2	10	0.6	0.9327
14	2	10	0.8	1.4580
14	4	2	0.2	1.5401
14	4	2	0.4	1.0035
14	4	2	0.6	0.5493
14	4	2	0.8	0.3370
14	4	4	0.2	0.4289
14	4	4	0.4	0.3406
14	4	4	0.6	0.2396
14	4	4	0.8	0.3383
14	4	6	0.2	0.1987
14	4	6	0.4	0.1753
14	4	6	0.6	0.2306
14	4	6	0.8	0.3633
14	4	8	0.2	0.2177
14	4	8	0.4	0.2517
14	4	8	0.6	0.3163
14	4	8	0.8	0.4058
14	6	2	0.2	1.6088
14	6	2	0.4	1.1169
14	6	2	0.6	0.6366
14	6	2	0.8	0.2716
14	6	4	0.2	0.4407
14	6	4	0.4	0.3738
14	6	4	0.6	0.2843
14	6	4	0.8	0.1883
14	6	6	0.2	0.2030
14	6	6	0.4	0.1901
14	6	6	0.6	0.1699
14	6	6	0.8	0.1863
14	8	2	0.2	1.6853
14	8	2	0.4	1.2807

Table 9 (cont.)

N	NO	N1	D	MAXBIAS
14	8	2	0.6	0.8195
14	8	2	0.8	0.4228
14	8	4	0.2	0.4533
14	8	4	0.4	0.4152
14	8	4	0.6	0.3558
14	8	4	0.8	0.2812
14	10	2	0.2	1.7708
14	10	2	0.4	1.5166
14	10	2	0.6	1.1600
14	10	2	0.8	0.7718
15	1	2	0.2	1.5265
15	1	2	0.4	0.9429
15	1	2	0.6	0.6872
15	1	2	0.8	3.2734
15	1	4	0.2	0.4369
15	1	4	0.4	0.3166
15	1	4	0.6	0.4074
15	1	4	0.8	2.6405
15	1	6	0.2	0.2047
15	1	6	0.4	0.1773
15	1	6	0.6	0.4973
15	1	6	0.8	2.6242
15	1	8	0.2	0.1854
15	1	8	0.4	0.2874
15	1	8	0.6	0.7092
15	1	8	0.8	2.9579
15	1	10	0.2	0.3619
15	1	10	0.4	0.5416
15	1	10	0.6	1.1729
15	1	10	0.8	3.6865
15	1	12	0.2	0.9868
15	1	12	0.4	1.3677
15	1	12	0.6	2.4190
15	1	12	0.8	5.0903
15	3	2	0.2	1.5858
15	3	2	0.4	0.9908

Table 9 (cont.)

N	NO	N1	D	MAXBIAS
15	3	2	0.6	0.5514
15	3	2	0.8	0.5896
15	3	4	0.2	0.4481
15	3	4	0.4	0.3378
15	3	4	0.6	0.2301
15	3	4	0.8	0.5633
15	3	6	0.2	0.2088
15	3	6	0.4	0.1754
15	3	6	0.6	0.2729
15	3	6	0.8	0.5887
15	3	8	0.2	0.1774
15	3	8	0.4	0.2314
15	3	8	0.6	0.3691
15	3	8	0.8	0.6497
15	3	10	0.2	0.3398
15	3	10	0.4	0.4071
15	3	10	0.6	0.5424
15	3	10	0.8	0.7466
15	5	2	0.2	1.6516
15	5	2	0.4	1.0741
15	5	2	0.6	0.5733
15	5	2	0.8	0.2396
15	5	4	0.2	0.4600
15	5	4	0.4	0.3659
15	5	4	0.6	0.2557
15	5	4	0.8	0.2423
15	5	6	0.2	0.2131
15	5	6	0.4	0.1885
15	5	6	0.6	0.1769
15	5	6	0.8	0.2586
15	5	8	0.2	0.1701
15	5	8	0.4	0.1920
15	5	8	0.6	0.2320
15	5	8	0.8	0.2846
15	7	2	0.2	1.7249
15	7	2	0.4	1.1977

Table 9 (cont.)

N	NO	N1	D	MAXBIAS
15	7	2	0.6	0.6788
15	7	2	0.8	0.2899
15	7	4	0.2	0.4725
15	7	4	0.4	0.4013
15	7	4	0.6	0.3053
15	7	4	0.8	0.2023
15	7	6	0.2	0.2176
15	7	6	0.4	0.2041
15	7	6	0.6	0.1826
15	7	6	0.8	0.1546
15	9	2	0.2	1.8065
15	9	2	0.4	1.3739
15	9	2	0.6	0.8788
15	9	2	0.8	0.4538
15	9	4	0.2	0.4859
15	9	4	0.4	0.4453
15	9	4	0.6	0.3820
15	9	4	0.8	0.3020
15	11	2	0.2	1.8978
15	11	2	0.4	1.6262
15	11	2	0.6	1.2444
15	11	2	0.8	0.8281

Table 10

Most Robust Designs for N = 8 to N = 20

N	N_0	N_1	N_2	D	MAXBIAS
8	4	4	-	-	0.2667
9	1	4	4	.3	0.2603
10	4	4	2	.6	0.2506
11	5	4	2	.8	0.2186
12	6	6	-	-	0.1778
13	3	6	4	.3	0.1745
14	6	6	2	.6	0.1699
15	7	6	2	.8	0.1546
16	8	8	-	-	0.1333
17	5	8	4	.3	0.1314
18	8	8	2	.6	0.1287
19	9	8	2	.8	0.1197
20	10	10	-	-	0.1067

commonly used second order designs for multiple variables will be evaluated.

5.1.2 Central Composite Design

The central composite design (ccd) described in Section 2.4.3 will be evaluated in this section and optimal design parameters will be found. An analytical expression is given in Appendix B for the integrated squared prediction bias. From this expression, it can be seen that ISB_i is a function of α and the number of center points.

The maximum ISB_i was calculated using (5.8) and the $(X'X)^{-1}$ matrix given in Appendix B for varying α and the number of center points. The values of α and N_0 which minimize the maximum ISB_i will be found. The region of interest is taken to be a sphere of radius one defined by the factorial points. Thus, it is assumed that the experimenter is interested in using the model for prediction only in the area inside the factorial points. The design will be scaled so that the factorial points lie on the unit sphere.

The axial value will be varied from 1 to 3 and the center points from 1 to 5. Note that with the region of interest as defined, if $\alpha \leq \sqrt{k}$ then the axial points are inside the sphere. Otherwise, the axial points are on the exterior of the sphere. Results will be obtained for $k = 2$ through 7. For $k = 5$ two cases will be considered, the first in which the factorial portion of the design is a full 2^5 factorial design and in the second a $\frac{1}{2}$ fraction of a 2^5 factorial will be used. For $k = 6$ and 7 only the $\frac{1}{2}$ fraction will be used in the evaluation.

In comparing the robustness of these designs it will be useful to compare the rotatable value of α with other axial values. For the ccd the value of α which gives a rotatable design is $\alpha = \sqrt[4]{F}$, where F is the number of factorial points in the design. The axial values are given in Table 11 for the relevant designs.

The maximum ISB_i are summarized in Table 12 for selected values of α for each value of k. The rotatable (or nearly rotatable) value of α is indicated with an asterisk. From these results it is seen that large values of α give robustness to outliers. Thus, for the region of interest defined by the factorial points α should be chosen as large as practicable based on the suitability of the second order model and the constraints of the particular system under investigation. Secondly, it is apparent that using only one center point can result in a design which is highly sensitive to outliers, unless α is very large. Including several center points eliminates this problem for reasonably large values of α. However, more than two or three center points gives only a marginal gain in robustness in most cases and in some cases gives no improvement at all.

It turns out that the rotatable designs are not robust, particularly if only one or two center points are used. Therefore, if rotatability is desired, the experimenter should seriously consider a compromise in which the α-level chosen is slightly larger than the rotatable value. If at least three center points are used, this design will be reasonably robust along with being nearly rotatable.

Table 11

Rotatable α-Values for ccd

k	2	3	4	5	5 $\frac{1}{2}$ frac.	6 $\frac{1}{2}$ frac.	7 $\frac{1}{2}$ frac.
α	1.414	1.682	2.0	2.378	2.0	2.378	2.828

Table 12

Maximum Integrated Squared Bias for ccd's

k=2

α

N_0	1.0	1.41*	2.0	3.0
1	0.7778	2.9907	1.4444	0.6288
2	0.6746	0.8308	0.6633	0.4118
3	0.7051	0.4062	0.3961	0.3167
4	0.75	0.3769	0.2917	0.2907
5	0.8088	0.4082	0.2956	0.2779

k=3

α

N_0	1.0	1.682*	2.0	3.0
1	1.4950	3.2321	2.5979	0.5113
2	1.5521	0.8719	0.8768	0.3808
3	1.6256	0.4133	0.4508	0.3622
4	1.7065	0.4315	0.3863	0.3558
5	1.7914	0.4560	0.4028	0.3561

k=4

α

N_0	1.0	1.5	2.0*	3.0
1	3.2563	0.8366	4.1667	0.4594
2	3.3651	0.8132	1.0833	0.3423
3	3.4804	0.8200	0.5	0.3367
4	3.5995	0.8372	0.3727	0.3365
5	3.7209	0.8590	0.3860	0.3394

* Rotatable design

Table 12 (cont.)

k=5 (full factorial)

N_0	α = 1.0	2.0	2.38*	3.0
1	6.9168	2.7748	4.6892	0.5555
2	7.0637	0.9271	1.3460	0.3702
3	7.2142	0.5633	0.6367	0.2723
4	7.3671	0.5674	0.3736	0.2728
5	7.5215	0.5751	0.3449	0.2748

k=5 ($\frac{1}{2}$ fraction)

N_0	α = 1.0	1.5	2.0*	3.0
1	4.3866	1.0184	1.7500	0.7527
2	4.5407	1.0352	0.6323	0.5939
3	4.6967	1.0605	0.6562	0.5951
4	4.8540	1.0898	0.6798	0.6036
5	5.0122	1.1211	0.7032	0.6157

k=6 ($\frac{1}{2}$ fraction)

N_0	α = 1.0	2.0	2.38*	3.0
1	8.5032	0.8000	4.0731	1.0414
2	8.6858	0.6811	1.0700	0.5585
3	8.8697	0.6786	0.4904	0.4234
4	9.0546	0.6840	0.4462	0.4274
5	9.2402	0.6928	0.4557	0.4330

* Rotatable design

Table 12 (cont.)

k=7 ($\frac{1}{2}$ fraction)

α

N_0	1.0	2.0	2.8*	3.0
1	16.8088	1.1899	5.1633	2.3284
2	17.0160	1.1706	1.5058	0.9895
3	17.2242	1.1674	0.7120	0.5479
4	17.4333	1.1712	0.4158	0.3491
5	17.6429	1.1786	0.3720	0.3102

* Rotatable design

Thus, for the central composite design the following recommendations should be followed to achieve robustness to outliers:

- choose α as large as possible;
- include 2-3 center points .

In a later section the ccd's will be evaluated for an alternative design region.

5.1.3 Box-Behnken Design

In this section the Box-Behnken design, which was described in Section 2.4.4 will be evaluated to determine the optimum number of center points for achieving robustness to outliers. The region of interest is taken to be a unit sphere and the design scaled so that the design points lie on the sphere.

These designs will be evaluated for $k=3$ through 7. For $k=6$ two cases will be examined, the first is the design based on the 2^2 factorial and the second is based on the 2^3 factorial as illustrated in (2.52). For the seven variable case only the design based on the 2^3 factorial will be investigated. The number of center points, N_0, will range from one to five. Since only the number of center points can be varied in this design, it is obviously much less flexible than the ccd.

An analytical expression for the integrated squared bias in prediction for the designs based on a 2^2 factorial is found in Appendix B. The maximum ISB_i were computed using equation (5.8), where $(X'X)^{-1}$ is as given in Appendix B, for the 2^2 factorial based designs. The results are given in Table 13. The six variable design based on the 2^2 factorial is denoted by 6_1 and the design based on the 2^3 factorial is denoted by 6_2. As with the ccd, adding center points improves the robustness of the design but more than three or four center points gives only marginal improvement, if any at all. As one would expect, using only one center point is not advisable since the bias can be very large if an outlier does occur. Thus, for the Box-Behnken designs to be outlier robust, three or four center points should be included in the design.

5.1.4 Hybrid Design

In this section the optimal parameters for the hybrid design described in Section 2.4.5 will be determined. The parameters to be determined are α, the axial value for the ccd portion of the design, and the four levels of the k^{th} variable, denoted α_1, α_2, α_3 and α_4. The first three levels of the k^{th} variable, α_1, α_2 and α_3 will be varied and the fourth level, α_4, will be determined from the first three so that [k], the first moment, is zero. The design will be scaled to the unit sphere by multiplying the original design matrix by $1/\sqrt{k}$. Thus, when $\alpha_3 = 1$, the factorial points in the design will lie on the sphere. The designs to be evaluated are for three, four and six variables, which are the cases originally presented in Roquemore (1976).

For each combination of the design parameters the maximum integrated squared bias was computed using (5.8) and the design matrix described in Section 2.4.5. From the results in Table 14, it is apparent that unless the parameters are chosen properly the resulting analysis can be extremely sensitive to outliers. In general the most robust designs will have the α-value in the ccd at least two, α_1 should be no larger than $|\alpha_2|$, which should also be at least two and α_3 should not be too large. Although in some cases adding center points improves the robustness of the design, center points alone will not make the designs with large biases robust. The most sensitive designs, and thus, the designs to be avoided, have α, α_1 and $|\alpha_2|$ all small and α_3 large.

The flexibility of these designs allows a reasonable level of robustness to be obtained. However, if the parameters are not chosen in accordance with the above guidelines, the analysis can be severely biased by the occurrence of an outlier.

5.1.5 Small Composite Design

The evaluation of the small composite designs (scd) will be similar to that done for the central composite designs in Section 5.1.2. (For a discussion of the scd see Section 2.4.6.) As in the case of the ccd, the axial value and number of center points will be varied. However, unlike the ccd there is no unique scd for a specific k. Thus sample scd's were chosen for this study. The region of interest will again be the unit sphere defined by the factorial points.

These designs will be examined for 4, 5, 6 and 7 variables. For $k=4$, the design matrix is given in (2.54). The factorial portion of the five variable design is the first five columns of a twelve run Plackett-Burman design. For the six variable design the factorial portion is the first six columns of a twenty run Plackett-Burman design, and for the seven variable design the first seven columns of the twenty-eight run Plackett-Burman design form the factorial portion. In practice any of the columns of the Plackett-Burman designs could be used. The first k columns were used here for simplicity.

The small composite design is made rotatable with $\alpha = \sqrt[4]{F}$, as is the ccd. Table 15 gives the rotatable axial values for the designs to be investigated here. The results for these values will be compared with the results for other axial values in order to evaluate the robustness of the rotatable scd.

The maximum ISB_i computed from (5.8) for these designs are given in Table 16 where the rotatable designs are indicated by an asterisk. As might be expected the designs with large axial values are robust and when α is small (i.e., near unity) the bias induced by an outlier can be very large. Notice also that the addition of center points has little affect on robustness and in some cases even slightly reduces the robustness. However, in addition to a large axial value, using several center points is recommended based on results to be discussed later for the missing data case. In Section 5.4 the small composite design will be re-examined, along with the ccd, for a design region which extends to the outermost points of the design.

It would appear as if the conclusions drawn here concerning the scd are specific for these designs chosen for the study. However, for a specific k and sample size, the structure of X'X will be either the same or very similar for all scd's. Thus, one should consider the conclusions as being quite general.

Table 13

Maximum Integrated Squared Bias
for Box-Behnken Designs

k	N_0	Maxbias
3	1	2.9714
	2	0.8000
	3	0.4821
	4	0.5143
	5	0.5464
4	1	4.1667
	2	1.0833
	3	0.5000
	4	0.3727
	5	0.3860
5	1	5.2064
	2	1.3333
	3	0.6067
	4	0.3492
	5	0.3423
6_1	1	6.1000
	2	1.5500
	3	0.7000
	4	0.4000
	5	0.3291
6_2	1	4.9000
	2	1.2500
	3	0.5667
	4	0.4325
	5	0.4408
7	1	4.6061
	2	1.1717
	3	0.5297
	4	0.4451
	5	0.4525

Table 14

Maximum Integrated Squared Bias

Hybrid Designs

k=3

α	α1	α2	α3	NO	MAXBIAS
1	1	-3	1	1	2.1549
1	1	-3	1	3	2.5554
1	1	-3	1	5	2.9587
1	1	-3	2	1	1.9097
1	1	-3	2	3	2.2379
1	1	-3	2	5	2.5764
1	1	-3	3	1	9.5586
1	1	-3	3	3	10.5674
1	1	-3	3	5	11.9281
1	1	-2	1	1	1.8555
1	1	-2	1	3	2.1861
1	1	-2	1	5	2.5247
1	1	-2	2	1	9.6180
1	1	-2	2	3	9.3879
1	1	-2	2	5	10.2261
1	1	-2	3	1	77.4567
1	1	-2	3	3	66.0681
1	1	-2	3	5	70.0095
1	1	-1	1	1	6.2857
1	1	-1	1	3	5.0776
1	1	-1	1	5	5.8587
1	1	-1	2	1	57.1627
1	1	-1	2	3	60.1753
1	1	-1	2	5	67.3873
1	1	-1	3	1	190.4571
1	1	-1	3	3	225.0857
1	1	-1	3	5	259.7143
1	2	-3	1	1	1.6980
1	2	-3	1	3	1.9385
1	2	-3	1	5	2.2230

Table 14 (cont.)

α	$\alpha 1$	$\alpha 2$	$\alpha 3$	NO	MAXBIAS
1	2	-3	2	1	1.6536
1	2	-3	2	3	1.8292
1	2	-3	2	5	2.0776
1	2	-3	3	1	8.9255
1	2	-3	3	3	6.7700
1	2	-3	3	5	7.0733
1	2	-2	1	1	1.4684
1	2	-2	1	3	1.6957
1	2	-2	1	5	1.9511
1	2	-2	2	1	8.2304
1	2	-2	2	3	3.8460
1	2	-2	2	5	4.4377
1	2	-2	3	1	28.2038
1	2	-2	3	3	21.8097
1	2	-2	3	5	24.3315
1	2	-1	1	1	1.3536
1	2	-1	1	3	1.5267
1	2	-1	1	5	1.7619
1	2	-1	2	1	5.1573
1	2	-1	2	3	6.0619
1	2	-1	2	5	6.9833
1	2	-1	3	1	19.4910
1	2	-1	3	3	23.5846
1	2	-1	3	5	27.4179
1	3	-3	1	1	1.4669
1	3	-3	1	3	1.6572
1	3	-3	1	5	1.9013
1	3	-3	2	1	3.0526
1	3	-3	2	3	1.5187
1	3	-3	2	5	1.7420
1	3	-3	3	1	9.6381
1	3	-3	3	3	4.0344
1	3	-3	3	5	4.6551
1	3	-2	1	1	2.6116
1	3	-2	1	3	1.5023
1	3	-2	1	5	1.7335

Table 14 (cont.)

α	α1	α2	α3	NO	MAXBIAS
1	3	-2	2	1	3.3129
1	3	-2	2	3	1.4402
1	3	-2	2	5	1.6350
1	3	-2	3	1	5.7854
1	3	-2	3	3	6.3809
1	3	-2	3	5	7.2334
1	3	-1	1	1	1.1976
1	3	-1	1	3	1.3966
1	3	-1	1	5	1.6109
1	3	-1	2	1	1.4245
1	3	-1	2	3	1.5835
1	3	-1	2	5	1.8318
1	3	-1	3	1	5.1351
1	3	-1	3	3	6.2070
1	3	-1	3	5	7.2137
2	1	-3	1	1	2.9188
2	1	-3	1	3	0.6848
2	1	-3	1	5	0.7856
2	1	-3	2	1	1.9986
2	1	-3	2	3	0.4835
2	1	-3	2	5	0.5353
2	1	-3	3	1	2.3664
2	1	-3	3	3	0.8409
2	1	-3	3	5	0.9804
2	1	-2	1	1	2.6507
2	1	-2	1	3	0.6897
2	1	-2	1	5	0.7947
2	1	-2	2	1	2.8032
2	1	-2	2	3	1.0973
2	1	-2	2	5	1.2587
2	1	-2	3	1	4.2146
2	1	-2	3	3	4.3418
2	1	-2	3	5	5.0502
2	1	-1	1	1	6.2857
2	1	-1	1	3	3.4195
2	1	-1	1	5	3.9456

Table 14 (cont.)

α	$\alpha 1$	$\alpha 2$	$\alpha 3$	NO	MAXBIAS
2	1	-1	2	1	16.9916
2	1	-1	2	3	15.7082
2	1	-1	2	5	17.5061
2	1	-1	3	1	46.2000
2	1	-1	3	3	54.6000
2	1	-1	3	5	63.0000
2	2	-3	1	1	3.9246
2	2	-3	1	3	0.5988
2	2	-3	1	5	0.6865
2	2	-3	2	1	4.8463
2	2	-3	2	3	0.7177
2	2	-3	2	5	0.5614
2	2	-3	3	1	5.2380
2	2	-3	3	3	1.2106
2	2	-3	3	5	1.3478
2	2	-2	1	1	2.6188
2	2	-2	1	3	0.6620
2	2	-2	1	5	0.7481
2	2	-2	2	1	8.2304
2	2	-2	2	3	1.0808
2	2	-2	2	5	1.1822
2	2	-2	3	1	15.4708
2	2	-2	3	3	5.4962
2	2	-2	3	5	6.0518
2	2	-1	1	1	1.0870
2	2	-1	1	3	0.8503
2	2	-1	1	5	0.9627
2	2	-1	2	1	5.5736
2	2	-1	2	3	4.4876
2	2	-1	2	5	4.6993
2	2	-1	3	1	43.2618
2	2	-1	3	3	22.8469
2	2	-1	3	5	23.8613
2	3	-3	1	1	2.1249
2	3	-3	1	3	0.6555
2	3	-3	1	5	0.7245

Table 14 (cont.)

α	$\alpha 1$	$\alpha 2$	$\alpha 3$	NO	MAXBIAS
2	3	-3	2	1	4.9889
2	3	-3	2	3	0.7355
2	3	-3	2	5	0.6192
2	3	-3	3	1	9.6381
2	3	-3	3	3	1.2656
2	3	-3	3	5	1.1511
2	3	-2	1	1	1.2362
2	3	-2	1	3	0.7475
2	3	-2	1	5	0.8251
2	3	-2	2	1	3.0317
2	3	-2	2	3	0.6781
2	3	-2	2	5	0.7489
2	3	-2	3	1	9.7817
2	3	-2	3	3	2.6081
2	3	-2	3	5	2.5846
2	3	-1	1	1	1.1059
2	3	-1	1	3	0.8752
2	3	-1	1	5	0.9966
2	3	-1	2	1	1.5038
2	3	-1	2	3	1.0265
2	3	-1	2	5	1.1361
2	3	-1	3	1	5.9727
2	3	-1	3	3	6.0730
2	3	-1	3	5	6.6729
3	1	-3	1	1	6.3124
3	1	-3	1	3	0.8895
3	1	-3	1	5	0.8048
3	1	-3	2	1	3.0791
3	1	-3	2	3	0.6825
3	1	-3	2	5	0.5364
3	1	-3	3	1	2.6998
3	1	-3	3	3	0.7169
3	1	-3	3	5	0.5052
3	1	-2	1	1	5.1773
3	1	-2	1	3	0.7179
3	1	-2	1	5	0.8095

Table 14 (cont.)

α	α1	α2	α3	NO	MAXBIAS
3	1	-2	2	1	3.0919
3	1	-2	2	3	0.7064
3	1	-2	2	5	0.6612
3	1	-2	3	1	2.9435
3	1	-2	3	3	1.0575
3	1	-2	3	5	1.2674
3	1	-1	1	1	6.2857
3	1	-1	1	3	3.9361
3	1	-1	1	5	4.5416
3	1	-1	2	1	4.4567
3	1	-1	2	3	4.8233
3	1	-1	2	5	5.6752
3	1	-1	3	1	7.4195
3	1	-1	3	3	7.1134
3	1	-1	3	5	8.7544
3	2	-3	1	1	3.2755
3	2	-3	1	3	0.6636
3	2	-3	1	5	0.6870
3	2	-3	2	1	7.0812
3	2	-3	2	3	0.9571
3	2	-3	2	5	0.5579
3	2	-3	3	1	5.3973
3	2	-3	3	3	0.9957
3	2	-3	3	5	0.5275
3	2	-2	1	1	1.9934
3	2	-2	1	3	0.6737
3	2	-2	1	5	0.7412
3	2	-2	2	1	8.2304
3	2	-2	2	3	1.0808
3	2	-2	2	5	0.7001
3	2	-2	3	1	7.5522
3	2	-2	3	3	1.4637
3	2	-2	3	5	1.6791
3	2	-1	1	1	1.8643
3	2	-1	1	3	1.4210
3	2	-1	1	5	1.3983

Table 14 (cont.)

α	α1	α2	α3	NO	MAXBIAS
3	2	-1	2	1	8.5413
3	2	-1	2	3	2.2132
3	2	-1	2	5	2.7862
3	2	-1	3	1	24.6016
3	2	-1	3	3	12.8687
3	2	-1	3	5	14.5399
3	3	-3	1	1	1.2251
3	3	-3	1	3	0.7003
3	3	-3	1	5	0.7380
3	3	-3	2	1	4.6548
3	3	-3	2	3	0.8421
3	3	-3	2	5	0.6025
3	3	-3	3	1	9.6381
3	3	-3	3	3	1.2656
3	3	-3	3	5	0.5519
3	3	-2	1	1	0.8223
3	3	-2	1	3	0.7565
3	3	-2	1	5	0.8114
3	3	-2	2	1	2.9280
3	3	-2	2	3	0.7214
3	3	-2	2	5	0.6837
3	3	-2	3	1	11.1424
3	3	-2	3	3	1.5841
3	3	-2	3	5	1.3079
3	3	-1	1	1	1.7708
3	3	-1	1	3	1.0863
3	3	-1	1	5	0.9983
3	3	-1	2	1	1.5727
3	3	-1	2	3	0.7793
3	3	-1	2	5	0.8716
3	3	-1	3	1	6.6747
3	3	-1	3	3	4.3674
3	3	-1	3	5	4.3365

Table 14 (cont.)

α	α1	α2	α3	NO	MAXBIAS
			k=4		
1	1	-3	1	1	3.2284
1	1	-3	1	3	3.6382
1	1	-3	1	5	4.0387
1	1	-3	2	1	12.7263
1	1	-3	2	3	12.1125
1	1	-3	2	5	12.7091
1	1	-3	3	1	193.1241
1	1	-3	3	3	100.3783
1	1	-3	3	5	115.4891
1	1	-2	1	1	3.1425
1	1	-2	1	3	3.4796
1	1	-2	1	5	3.8366
1	1	-2	2	1	86.3466
1	1	-2	2	3	62.6771
1	1	-2	2	5	68.2712
1	1	-2	3	1	146.2662
1	1	-2	3	3	184.0102
1	1	-2	3	5	211.0609
1	1	-1	1	1	11.0990
1	1	-1	1	3	12.0440
1	1	-1	1	5	13.2191
1	1	-1	2	1	45.5906
1	1	-1	2	3	57.2118
1	1	-1	2	5	65.7667
1	1	-1	3	1	103.6565
1	1	-1	3	3	132.2186
1	1	-1	3	5	153.5508
1	2	-3	1	1	2.9381
1	2	-3	1	3	3.2549
1	2	-3	1	5	3.5914
1	2	-3	2	1	7.6008
1	2	-3	2	3	5.0276
1	2	-3	2	5	5.8546

Table 14 (cont.)

α	$\alpha 1$	$\alpha 2$	$\alpha 3$	NO	MAXBIAS
1	2	-3	3	1	38.9546
1	2	-3	3	3	33.1922
1	2	-3	3	5	36.0294
1	2	-2	1	1	2.7712
1	2	-2	1	3	3.0793
1	2	-2	1	5	3.4008
1	2	-2	2	1	9.4482
1	2	-2	2	3	9.9932
1	2	-2	2	5	10.9491
1	2	-2	3	1	25.7040
1	2	-2	3	3	30.9036
1	2	-2	3	5	34.9162
1	2	-1	1	1	2.6361
1	2	-1	1	3	2.9459
1	2	-1	1	5	3.2565
1	2	-1	2	1	7.7715
1	2	-1	2	3	9.2190
1	2	-1	2	5	10.3861
1	2	-1	3	1	23.4423
1	2	-1	3	3	28.6405
1	2	-1	3	5	32.6390
1	3	-3	1	1	2.7756
1	3	-3	1	3	3.0596
1	3	-3	1	5	3.3748
1	3	-3	2	1	5.6238
1	3	-3	2	3	2.8034
1	3	-3	2	5	3.0981
1	3	-3	3	1	10.7044
1	3	-3	3	3	10.2943
1	3	-3	3	5	11.2727
1	3	-2	1	1	2.6743
1	3	-2	1	3	2.9442
1	3	-2	1	5	3.2542
1	3	-2	2	1	3.3069
1	3	-2	2	3	2.8307
1	3	-2	2	5	3.1239

Table 14 (cont.)

α	α1	α2	α3	NO	MAXBIAS
1	3	-2	3	1	8.1822
1	3	-2	3	3	9.6357
1	3	-2	3	5	10.8176
1	3	-1	1	1	2.5561
1	3	-1	1	3	2.8534
1	3	-1	1	5	3.1532
1	3	-1	2	1	2.4088
1	3	-1	2	3	2.7832
1	3	-1	2	5	3.1108
1	3	-1	3	1	7.9793
1	3	-1	3	3	9.4540
1	3	-1	3	5	10.6543
2	1	-3	1	1	1.8496
2	1	-3	1	3	0.5954
2	1	-3	1	5	0.6587
2	1	-3	2	1	3.5520
2	1	-3	2	3	1.8495
2	1	-3	2	5	1.9816
2	1	-3	3	1	12.6748
2	1	-3	3	3	12.7096
2	1	-3	3	5	13.6027
2	1	-2	1	1	2.8761
2	1	-2	1	3	0.9357
2	1	-2	1	5	1.0130
2	1	-2	2	1	17.4359
2	1	-2	2	3	11.6403
2	1	-2	2	5	12.1337
2	1	-2	3	1	99.5579
2	1	-2	3	3	75.1600
2	1	-2	3	5	78.0043
2	1	-1	1	1	11.4261
2	1	-1	1	3	7.2885
2	1	-1	1	5	7.4993
2	1	-1	2	1	100.6902
2	1	-1	2	3	84.1195
2	1	-1	2	5	86.0793

Table 14 (cont.)

α	$\alpha1$	$\alpha2$	$\alpha3$	NO	MAXBIAS
2	1	-1	3	1	485.8557
2	1	-1	3	3	443.9383
2	1	-1	3	5	463.1961
2	2	-3	1	1	3.9884
2	2	-3	1	3	0.5845
2	2	-3	1	5	0.6466
2	2	-3	2	1	9.3665
2	2	-3	2	3	1.5497
2	2	-3	2	5	1.6258
2	2	-3	3	1	30.2341
2	2	-3	3	3	9.2119
2	2	-3	3	5	9.6703
2	2	-2	1	1	3.1862
2	2	-2	1	3	0.6183
2	2	-2	1	5	0.6786
2	2	-2	2	1	13.7793
2	2	-2	2	3	4.8563
2	2	-2	2	5	4.8558
2	2	-2	3	1	46.5225
2	2	-2	3	3	28.3475
2	2	-2	3	5	28.6985
2	2	-1	1	1	1.2244
2	2	-1	1	3	1.0001
2	2	-1	1	5	1.0885
2	2	-1	2	1	8.7696
2	2	-1	2	3	9.2066
2	2	-1	2	5	9.9686
2	2	-1	3	1	38.9738
2	2	-1	3	3	42.0099
2	2	-1	3	5	45.8777
2	3	-3	1	1	2.9256
2	3	-3	1	3	0.6077
2	3	-3	1	5	0.6650
2	3	-3	2	1	6.3338
2	3	-3	2	3	0.9474
2	3	-3	2	5	0.9168

Table 14 (cont.)

α	$\alpha 1$	$\alpha 2$	$\alpha 3$	NO	MAXBIAS
2	3	-3	3	1	15.6702
2	3	-3	3	3	5.3252
2	3	-3	3	5	5.3020
2	3	-2	1	1	1.5299
2	3	-2	1	3	0.6497
2	3	-2	1	5	0.7089
2	3	-2	2	1	3.1528
2	3	-2	2	3	2.0031
2	3	-2	2	5	2.0589
2	3	-2	3	1	10.3339
2	3	-2	3	3	9.7462
2	3	-2	3	5	10.2202
2	3	-1	1	1	1.0497
2	3	-1	1	3	0.6789
2	3	-1	1	5	0.7470
2	3	-1	2	1	1.9678
2	3	-1	2	3	2.1515
2	3	-1	2	5	2.3603
2	3	-1	3	1	8.8012
2	3	-1	3	3	9.7421
2	3	-1	3	5	10.7323
3	1	-3	1	1	6.3374
3	1	-3	1	3	0.9242
3	1	-3	1	5	0.6727
3	1	-3	2	1	4.8595
3	1	-3	2	3	1.0117
3	1	-3	2	5	0.5564
3	1	-3	3	1	6.2751
3	1	-3	3	3	2.0915
3	1	-3	3	5	2.3172
3	1	-2	1	1	6.7167
3	1	-2	1	3	0.9065
3	1	-2	1	5	0.7950
3	1	-2	2	1	8.0246
3	1	-2	2	3	2.3715
3	1	-2	2	5	2.5743

Table 14 (cont.)

α	$\alpha 1$	$\alpha 2$	$\alpha 3$	NO	MAXBIAS
3	1	-2	3	1	12.6998
3	1	-2	3	3	10.0181
3	1	-2	3	5	11.0501
3	1	-1	1	1	13.3910
3	1	-1	1	3	5.3338
3	1	-1	1	5	6.0802
3	1	-1	2	1	52.2694
3	1	-1	2	3	28.8684
3	1	-1	2	5	30.5973
3	1	-1	3	1	111.4507
3	1	-1	3	3	117.3334
3	1	-1	3	5	127.4927
3	2	-3	1	1	4.6091
3	2	-3	1	3	0.8035
3	2	-3	1	5	0.6431
3	2	-3	2	1	11.5228
3	2	-3	2	3	1.4540
3	2	-3	2	5	0.5804
3	2	-3	3	1	15.0365
3	2	-3	3	3	2.6060
3	2	-3	3	5	2.7539
3	2	-2	1	1	2.4210
3	2	-2	1	3	0.6171
3	2	-2	1	5	0.6686
3	2	-2	2	1	15.3133
3	2	-2	2	3	2.0358
3	2	-2	2	5	1.6773
3	2	-2	3	1	40.6211
3	2	-2	3	3	9.4799
3	2	-2	3	5	10.1682
3	2	-1	1	1	2.0144
3	2	-1	1	3	0.9998
3	2	-1	1	5	0.9748
3	2	-1	2	1	9.3273
3	2	-1	2	3	7.5601
3	2	-1	2	5	7.5219

Table 14 (cont.)

α	$\alpha 1$	$\alpha 2$	$\alpha 3$	NO	MAXBIAS
3	2	-1	3	1	56.5888
3	2	-1	3	3	41.2389
3	2	-1	3	5	40.6759
3	3	-3	1	1	1.3084
3	3	-3	1	3	0.6261
3	3	-3	1	5	0.6639
3	3	-3	2	1	5.9209
3	3	-3	2	3	1.1014
3	3	-3	2	5	0.5960
3	3	-3	3	1	17.7231
3	3	-3	3	3	2.3562
3	3	-3	3	5	1.6790
3	3	-2	1	1	1.2277
3	3	-2	1	3	0.7062
3	3	-2	1	5	0.6973
3	3	-2	2	1	3.1804
3	3	-2	2	3	1.2307
3	3	-2	2	5	1.0682
3	3	-2	3	1	12.6345
3	3	-2	3	3	5.4195
3	3	-2	3	5	4.9578
3	3	-1	1	1	1.9982
3	3	-1	1	3	0.8937
3	3	-1	1	5	0.7469
3	3	-1	2	1	1.9588
3	3	-1	2	3	1.5908
3	3	-1	2	5	1.6730
3	3	-1	3	1	9.9751
3	3	-1	3	3	9.6010
3	3	-1	3	5	10.0794

Table 14 (cont.)

α	$\alpha1$	$\alpha2$	$\alpha3$	NO	MAXBIAS
			k=6		
1	1	-3	1	1	6.9319
1	1	-3	1	3	7.4398
1	1	-3	1	5	7.9356
1	1	-3	2	1	92.3242
1	1	-3	2	3	72.5047
1	1	-3	2	5	80.3930
1	1	-3	3	1	154.4624
1	1	-3	3	3	201.3137
1	1	-3	3	5	227.6600
1	1	-2	1	1	7.5400
1	1	-2	1	3	9.6349
1	1	-2	1	5	10.8797
1	1	-2	2	1	75.3578
1	1	-2	2	3	93.4695
1	1	-2	2	5	103.8710
1	1	-2	3	1	118.1780
1	1	-2	3	3	117.1192
1	1	-2	3	5	131.7421
1	1	-1	1	1	18.4898
1	1	-1	1	3	21.0013
1	1	-1	1	5	22.7306
1	1	-1	2	1	58.9810
1	1	-1	2	3	72.6336
1	1	-1	2	5	81.4136
1	1	-1	3	1	130.3614
1	1	-1	3	3	158.0330
1	1	-1	3	5	176.6962
1	2	-3	1	1	6.6938
1	2	-3	1	3	7.1502
1	2	-3	1	5	7.6113
1	2	-3	2	1	23.8639
1	2	-3	2	3	18.0400
1	2	-3	2	5	19.1380

Table 14 (cont.)

α	$\alpha 1$	$\alpha 2$	$\alpha 3$	NO	MAXBIAS
1	2	-3	3	1	40.9527
1	2	-3	3	3	49.3318
1	2	-3	3	5	54.3018
1	2	-2	1	1	6.5574
1	2	-2	1	3	7.0057
1	2	-2	1	5	7.4571
1	2	-2	2	1	14.9725
1	2	-2	2	3	17.2420
1	2	-2	2	5	18.7309
1	2	-2	3	1	33.7134
1	2	-2	3	3	41.6433
1	2	-2	3	5	46.5222
1	2	-1	1	1	6.4512
1	2	-1	1	3	6.8959
1	2	-1	1	5	7.3409
1	2	-1	2	1	13.9873
1	2	-1	2	3	16.3111
1	2	-1	2	5	17.8722
1	2	-1	3	1	38.5566
1	2	-1	3	3	45.9739
1	2	-1	3	5	50.8685
1	3	-3	1	1	6.5566
1	3	-3	1	3	7.0000
1	3	-3	1	5	7.4512
1	3	-3	2	1	7.9370
1	3	-3	2	3	6.5772
1	3	-3	2	5	7.0024
1	3	-3	3	1	15.0543
1	3	-3	3	3	17.4165
1	3	-3	3	5	18.9436
1	3	-2	1	1	6.4557
1	3	-2	1	3	6.9014
1	3	-2	1	5	7.3467
1	3	-2	2	1	6.1149
1	3	-2	2	3	6.5370
1	3	-2	2	5	6.9593

Table 14 (cont.)

α	$\alpha1$	$\alpha2$	$\alpha3$	NO	MAXBIAS
1	3	-2	3	1	13.7064
1	3	-2	3	3	16.2190
1	3	-2	3	5	17.8311
1	3	-1	1	1	6.3755
1	3	-1	1	3	6.8127
1	3	-1	1	5	7.2516
1	3	-1	2	1	6.0826
1	3	-1	2	3	6.5030
1	3	-1	2	5	6.9229
1	3	-1	3	1	14.7899
1	3	-1	3	3	17.0375
1	3	-1	3	5	18.6020
2	1	-3	1	1	0.8884
2	1	-3	1	3	0.8915
2	1	-3	1	5	0.9457
2	1	-3	2	1	26.5718
2	1	-3	2	3	22.1116
2	1	-3	2	5	21.9780
2	1	-3	3	1	411.3236
2	1	-3	3	3	120.3823
2	1	-3	3	5	124.3092
2	1	-2	1	1	5.3799
2	1	-2	1	3	4.4137
2	1	-2	1	5	4.3994
2	1	-2	2	1	160.7625
2	1	-2	2	3	96.9726
2	1	-2	2	5	97.8827
2	1	-2	3	1	389.6160
2	1	-2	3	3	429.8369
2	1	-2	3	5	462.3593
2	1	-1	1	1	22.2165
2	1	-1	1	3	21.8711
2	1	-1	1	5	22.7448
2	1	-1	2	1	101.5390
2	1	-1	2	3	117.6351
2	1	-1	2	5	128.8431

Table 14 (cont.)

α	α1	α2	α3	NO	MAXBIAS
2	1	-1	3	1	245.4424
2	1	-1	3	3	293.1898
2	1	-1	3	5	325.5891
2	2	-3	1	1	2.1541
2	2	-3	1	3	0.8030
2	2	-3	1	5	0.8505
2	2	-3	2	1	24.8726
2	2	-3	2	3	7.8752
2	2	-3	2	5	8.2476
2	2	-3	3	1	67.3602
2	2	-3	3	3	59.1895
2	2	-3	3	5	59.9077
2	2	-2	1	1	3.6858
2	2	-2	1	3	1.6336
2	2	-2	1	5	1.6996
2	2	-2	2	1	18.4493
2	2	-2	2	3	17.8853
2	2	-2	2	5	18.5165
2	2	-2	3	1	56.1401
2	2	-2	3	3	61.5770
2	2	-2	3	5	66.1060
2	2	-1	1	1	2.1550
2	2	-1	1	3	2.2778
2	2	-1	1	5	2.4165
2	2	-1	2	1	14.9183
2	2	-1	2	3	16.6333
2	2	-1	2	5	17.9606
2	2	-1	3	1	48.6524
2	2	-1	3	3	55.8030
2	2	-1	3	5	60.8957
2	3	-3	1	1	2.9823
2	3	-3	1	3	0.7990
2	3	-3	1	5	0.8505
2	3	-3	2	1	7.9791
2	3	-3	2	3	4.2040
2	3	-3	2	5	4.2607

Table 14 (cont.)

α	$\alpha 1$	$\alpha 2$	$\alpha 3$	NO	MAXBIAS
2	3	-3	3	1	19.5590
2	3	-3	3	3	18.9049
2	3	-3	3	5	19.5547
2	3	-2	1	1	1.8812
2	3	-2	1	3	0.8143
2	3	-2	1	5	0.8653
2	3	-2	2	1	4.7640
2	3	-2	2	3	4.8847
2	3	-2	2	5	5.1353
2	3	-2	3	1	16.4728
2	3	-2	3	3	18.0019
2	3	-2	3	5	19.3023
2	3	-1	1	1	0.9668
2	3	-1	1	3	0.8221
2	3	-1	1	5	0.8744
2	3	-1	2	1	4.2838
2	3	-1	2	3	4.7100
2	3	-1	2	5	5.0613
2	3	-1	3	1	15.4951
2	3	-1	3	3	17.3550
2	3	-1	3	5	18.7726
3	1	-3	1	1	5.0639
3	1	-3	1	3	0.8207
3	1	-3	1	5	0.8388
3	1	-3	2	1	12.5662
3	1	-3	2	3	5.1391
3	1	-3	2	5	5.2266
3	1	-3	3	1	45.0997
3	1	-3	3	3	35.2587
3	1	-3	3	5	35.8587
3	1	-2	1	1	10.4179
3	1	-2	1	3	2.2359
3	1	-2	1	5	2.3077
3	1	-2	2	1	69.1897
3	1	-2	2	3	27.1403
3	1	-2	2	5	27.2573

Table 14 (cont.)

α	α1	α2	α3	NO	MAXBIAS
3	1	-2	3	1	406.1062
3	1	-2	3	3	164.7839
3	1	-2	3	5	166.5846
3	1	-1	1	1	22.3790
3	1	-1	1	3	16.8708
3	1	-1	1	5	16.2757
3	1	-1	2	1	198.0729
3	1	-1	2	3	167.1567
3	1	-1	2	5	165.8374
3	1	-1	3	1	815.3151
3	1	-1	3	3	778.6275
3	1	-1	3	5	799.6303
3	2	-3	1	1	6.9120
3	2	-3	1	3	0.9025
3	2	-3	1	5	0.8339
3	2	-3	2	1	24.7957
3	2	-3	2	3	2.9513
3	2	-3	2	5	3.0948
3	2	-3	3	1	84.0530
3	2	-3	3	3	16.1727
3	2	-3	3	5	17.4114
3	2	-2	1	1	3.6349
3	2	-2	1	3	1.1636
3	2	-2	1	5	1.0506
3	2	-2	2	1	19.6463
3	2	-2	2	3	11.3615
3	2	-2	2	5	10.6034
3	2	-2	3	1	79.3694
3	2	-2	3	3	61.4479
3	2	-2	3	5	59.5433
3	2	-1	1	1	1.9322
3	2	-1	1	3	1.9372
3	2	-1	1	5	2.0185
3	2	-1	2	1	16.3424
3	2	-1	2	3	16.5358
3	2	-1	2	5	17.2705

Table 14 (cont.)

α	$\alpha 1$	$\alpha 2$	$\alpha 3$	NO	MAXBIAS
3	2	-1	3	1	71.3618
3	2	-1	3	3	74.5307
3	2	-1	3	5	78.7066
3	3	-3	1	1	2.2223
3	3	-3	1	3	0.7983
3	3	-3	1	5	0.8428
3	3	-3	2	1	7.8861
3	3	-3	2	3	2.5056
3	3	-3	2	5	2.1781
3	3	-3	3	1	21.9995
3	3	-3	3	3	12.6289
3	3	-3	3	5	11.7581
3	3	-2	1	1	1.4241
3	3	-2	1	3	0.8101
3	3	-2	1	5	0.8574
3	3	-2	2	1	4.5420
3	3	-2	2	3	3.8773
3	3	-2	2	5	3.8151
3	3	-2	3	1	19.8051
3	3	-2	3	3	18.2839
3	3	-2	3	5	18.5226
3	3	-1	1	1	1.7537
3	3	-1	1	3	0.8168
3	3	-1	1	5	0.8680
3	3	-1	2	1	3.6706
3	3	-1	2	3	3.8279
3	3	-1	2	5	4.0393
3	3	-1	3	1	16.6734
3	3	-1	3	3	17.6859
3	3	-1	3	5	18.7762

Table 15

Rotatable α-Values for scd

k	4	5	6	7
α	1.68	1.86	2.12	2.30

Table 16

Maximum Integrated Squared Bias for scd

k=4

N_0	α 1.0	1.68*	2.0	3.0
1	6.4311	1.8681	2.8333	0.7778
2	6.7957	1.9394	1.3008	0.7186
3	7.1641	2.0344	1.3731	0.7136
4	7.5347	2.1352	1.4453	0.7244
5	7.9066	2.2383	1.5176	0.7430

k=5

N_0	α 1.0	1.86*	2.0	3.0
1	10.6474	2.434	2.0782	1.3461
2	11.1033	2.5073	2.1397	1.3476
3	11.5609	2.6001	2.2201	1.3761
4	12.0194	2.6983	2.3048	1.4152
5	12.4787	2.7986	2.391	1.4593

k=6

N_0	α 1.0	2.1*	2.5	3.0
1	16.4336	2.8006	3.3754	1.3173
2	16.9271	2.8525	1.9578	1.3461
3	17.4215	2.9250	2.0166	1.3837
4	17.9165	3,0030	2.0748	1.4235
5	18.4119	3.0831	2.1329	1.4639

* Rotatable design

Table 16 (cont.)

k=7

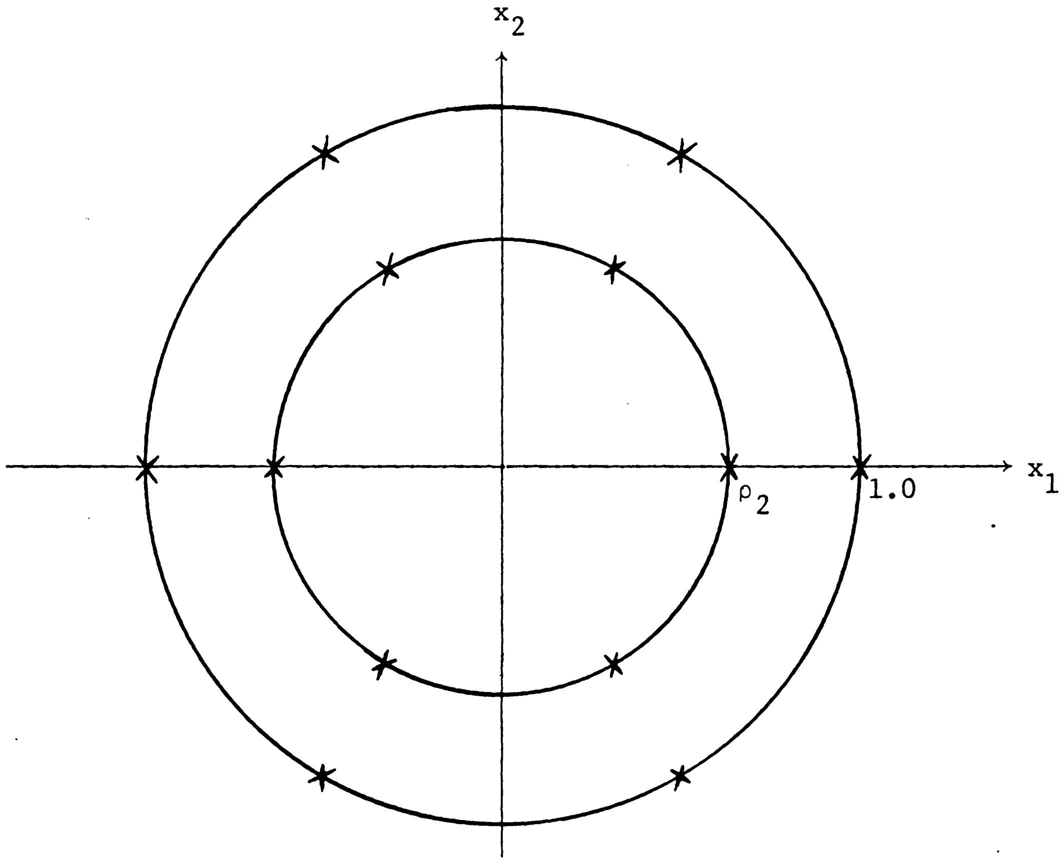
α

N_0	1.0	2.0	2.3*	3.0
1	24.3828	6.1499	4.9407	3.0238
2	24.9645	6.3660	5.0989	3.1357
3	25.545	6.557	5.2363	3.2272
4	26.1244	6.7364	5.3656	3.3109
5	26.7029	6.9074	5.4909	3.3910

* Rotatable design

5.1.6 Equiradial Design

These two variable designs, which were described in Section 2.4.6, will be examined for two cases. The first will be the design described in Chapter II in which the points are equally spaced on a circle and augmented with center points. In this case the circle will be of radius one, the number of points on the circle will vary from five to ten and the number of center points will range from one to five. The second case will combine two equiradial designs having different radii. The first will have its points on a unit circle and the second will have its points on a smaller circle, the exact radius of which will be varied. In using the combined equiradial design rotatability is preserved but the singularity that exists with a single equiradial design is eliminated. Thus, there is no need to include center points, although it may be desirable to do so. The following diagram will illustrate the combined design:



The two types of equiradial designs will be evaluated individually to determine the design parameters which give robustness. Along with investigating the designs individually, the two will be compared to determine whether it is more advantageous to use a single design augmented with center points or to use the combined design.

The results computed from (5.8) for the first case are given in Table 17, where N_1 denotes the number of points on the circle. It turns out that no matter what N_1 is chosen, approximately three center points is optimal. Table 18 gives the same results by sample size. This illustrates that for a given sample size, the points should be distributed with roughly three points in the center and

Table 17

Maximum Integrated Squared Bias
for Equiradial Designs

N_1	N_0	Maxbias
5	1	2.0
	2	0.5833
	3	0.6400
	4	0.7200
	5	0.8000
6	1	2.3333
	2	0.6667
	3	0.5000
	4	0.5556
	5	0.6111
7	1	2.6667
	2	0.7500
	3	0.4082
	4	0.4490
	5	0.4898
8	1	3.0
	2	0.8333
	3	0.4074
	4	0.375
	5	0.4063
9	1	3.3333
	2	0.9167
	3	0.4444
	4	0.3210
	5	0.3457
10	1	3.6667
	2	1.0000
	3	0.4815
	4	0.2917
	5	0.3000

Table 18

Maximum Integrated Squared Bias
for Equiradial Designs - By Sample Size

N	N_1	N_0	Maxbias
6	5	1	2.0
7	5	2	0.5833
	6	1	2.3333
8	5	3	0.6400
	6	2	0.6667
	7	1	2.6667
9	5	4	0.7200
	6	3	0.5000
	7	2	0.7500
	8	1	3.0000
10	5	5	0.8000
	6	4	0.5556
	7	3	0.4082
	8	2	0.8333
	9	1	3.3333
11	6	5	0.6111
	7	4	0.4490
	8	3	0.4074
	9	2	0.9167
	10	1	3.6667
12	7	5	0.4898
	8	4	0.3750
	9	3	0.4444
	10	2	1.0000
13	8	5	0.4063
	9	4	0.3210
	10	3	0.4815
14	9	5	0.3457
	10	4	0.2917
15	10	5	0.3000

the remaining points on the circle. Thus, having N_1 as large as possible is not optimal and, in fact, can be a very poor choice of designs. As the sample size gets large (greater than about 12), additional center points should be added to achieve optimality.

The results for the combined design appear in Table 19, where N is the total sample size, N_1 is the number of points on each circle and ρ_2 is the radius of the inner circle. From these results, it is seen that the second circle of points should be placed about halfway between the origin and the outer circle, at a radius of approximately .4 to .6. As one would expect, increasing the number of points on the circles improves the robustness of the design.

When the results for the two types of designs are compared, one sees that for a particular sample size it is better to use a single design with center points than to use the combined design, provided the points are distributed properly.

5.1.7 Mean Square Error

In the first order design case, it turned out that the designs that gave minimax(ISB_i) were also optimal in terms of the mean square error of prediction. The same result holds true in the second order case. Based on results in Wardrop (1984), it can be shown that the design parameters found to be optimal in this case also minimize the integrated prediction variance for that design class. Since the integrated variance is not affected by outliers, it will be constant for a given design. Therefore, the designs that minimize the maximum integrated mean square error for a particular design class will be those designs recommended above.

Table 19

Maximum Integrated Squared Bias
for Combined Equiradial Designs

N	N_1	ρ_2	Maxbias
10	5	.1	0.7907
		.3	0.7205
		.5	0.5959
		.7	0.6425
		.9	3.8561
12	6	.1	0.6589
		.3	0.6004
		.5	0.4966
		.7	0.5335
		.9	3.2134
14	7	.1	0.5648
		.3	0.5147
		.5	0.4257
		.7	0.4590
		.9	2.7544
16	8	.1	0.4942
		.3	0.4503
		.5	0.3725
		.7	0.4016
		.9	2.4101
18	9	.1	0.4393
		.3	0.4003
		.5	0.3311
		.7	0.3570
		.9	2.1423
20	10	.1	0.3954
		.3	0.3603
		.5	0.2980
		.7	0.3213
		.9	1.9281

In the next section the robustness of these designs to missing data will be examined. Recommendations will be made for achieving robustness to missing data. Then in Section 5.3 comparisons will be made across design types to determine which types are most robust to outliers and to missing data.

5.2 Robustness to Missing Data

The philosophy developed in Chapter IV for determining the robustness of first order designs to missing data will now be applied to second order designs. The integrated prediction variance, V_{-i}, defined in (4.7) will again be the basis for this development. The maximum V_{-i} was computed for each design and the parameters found which minimize the maximum V_{-i}.

The notion of robustness to missing data will be illustrated with an example using the five variable ccd with a full 2^5 in the factorial portion. Two designs will be compared, one with $\alpha = 2.0$ and the other with $\alpha = 3.0$ and both having one center point. In the following table the integrated prediction variance as defined in equation (2.26) and the maximum V_{-i} as defined in (4.8) are given for the two designs.

α	Int. Variance	Max V_{-i}
2.0	17.76	28.86
3.0	11.94	12.67

We see that the design with $\alpha = 3.0$ has a maximum increase in prediction variance of only 6% when a point is lost as compared with a maximum increase of 63% for the design with $\alpha = 2.0$. In addition to being a more desirable design in terms of prediction variance with the complete data set, the $\alpha = 3.0$ design is more robust to missing data.

5.2.1 One Independent Variable

The maximum V_{-i} computed from (4.7) for the single variable second order designs are shown in Table 20 for the three-level designs and in Table 21 for the five-level designs. The results for the three-level designs show that, as with the outlier robust designs, the optimal three-level design has approximately half of the points at the origin and the remaining points distributed between the two endpoints. In general the best five-level designs are nearly the same as those found in the outlier case. That is, the number of center points should be less than half the total number of observations and approximately equal to the number of points at each endpoint and the number of points at the extremes should be larger than the number at the interior levels. In this case the location of the interior level, D, makes very little difference if the other specifications are met.

The most robust design for each sample size is given in Table 22. As in the outlier case, neither design type is best in all situations, although for larger sample sizes the three-level design is always best. However, the best three-level design and the best five-

Table 20

Maximum Integrated Prediction Variance -i

k=1, 3-Level Designs

N	NO	MAXIVI
8	2	4.9778
8	4	2.6667
8	6	2.9867
9	3	3.2000
9	5	2.7600
9	7	3.2000
10	2	6.0000
10	4	2.6667
10	6	2.8889
10	8	3.4286
11	3	3.6667
11	5	2.4444
11	7	3.0381
11	9	3.6667
12	2	7.0400
12	4	2.9333
12	6	2.4000
12	8	3.2000
12	10	3.9111
13	3	4.1600
13	5	2.6000
13	7	2.4349
13	9	3.3704
13	11	4.1600
14	2	8.0889
14	4	3.2356
14	6	2.4267
14	8	2.4889
14	10	3.5467
14	12	4.4121

Table 20 (cont.)

N	NO	MAXIVI
15	3	4.6667
15	5	2.8000
15	7	2.3333
15	9	2.5556
15	11	3.7273
15	13	4.6667
16	2	9.1429
16	4	3.5556
16	6	2.5600
16	8	2.3111
16	10	2.6311
16	12	3.9111
16	14	4.9231
17	3	5.1810
17	5	3.0222
17	7	2.4178
17	9	2.3296
17	11	2.7131
17	13	4.0974
17	15	5.1810
18	2	10.2000
18	4	3.8857
18	6	2.7200
18	8	2.3314
18	10	2.3600
18	12	2.8000
18	14	4.2857
18	16	5.4400
19	3	5.7000
19	5	3.2571
19	7	2.5333
19	9	2.2800
19	11	2.3990
19	13	2.8906
19	15	4.4756
19	17	5.7000

Table 20 (cont.)

N	NO	MAXIVI
20	2	11.2593
20	4	4.2222
20	6	2.8952
20	8	2.4127
20	10	2.2667
20	12	2.4444
20	14	2.9841
20	16	4.6667
20	18	5.9608

Table 21

Maximum Integrated Prediction Variance -i

k=1, 5-Level Designs

N	NO	N1	D	MAXIVI
8	2	2	0.2	28.2111
8	2	2	0.4	8.0726
8	2	2	0.6	4.3010
8	2	2	0.8	4.0044
8	2	4	0.2	2.6634
8	2	4	0.4	2.7387
8	2	4	0.6	3.2322
8	2	4	0.8	4.1925
8	4	2	0.2	53.7986
8	4	2	0.4	13.6375
8	4	2	0.6	6.1224
8	4	2	0.8	3.6232
9	1	2	0.2	21.9854
9	1	2	0.4	6.9264
9	1	2	0.6	4.4381
9	1	2	0.8	13.0384
9	1	4	0.2	2.7285
9	1	4	0.4	2.7281
9	1	4	0.6	3.9734
9	1	4	0.8	12.8925
9	1	6	0.2	3.3823
9	1	6	0.4	4.1289
9	1	6	0.6	6.6339
9	1	6	0.8	21.0479
9	3	2	0.2	31.5485
9	3	2	0.4	8.8268
9	3	2	0.6	4.4433
9	3	2	0.8	3.0378
9	3	4	0.2	2.7418
9	3	4	0.4	2.7028
9	3	4	0.6	2.7053
9	3	4	0.8	3.0124

Table 21 (cont.)

N	NO	N1	D	MAXIVI
9	5	2	0.2	60.3508
9	5	2	0.4	15.1678
9	5	2	0.6	6.7076
9	5	2	0.8	3.8727
10	2	2	0.2	24.2599
10	2	2	0.4	7.4015
10	2	2	0.6	4.2076
10	2	2	0.8	4.5028
10	2	4	0.2	2.8363
10	2	4	0.4	2.7517
10	2	4	0.6	3.0439
10	2	4	0.8	4.4825
10	2	6	0.2	2.7460
10	2	6	0.4	3.0331
10	2	6	0.6	3.6991
10	2	6	0.8	4.9548
10	4	2	0.2	34.8997
10	4	2	0.4	9.6152
10	4	2	0.6	4.6781
10	4	2	0.8	3.0113
10	4	4	0.2	2.8604
10	4	4	0.4	2.7874
10	4	4	0.6	2.6933
10	4	4	0.8	2.6315
10	6	2	0.2	66.9141
10	6	2	0.4	16.7134
10	6	2	0.6	7.3143
10	6	2	0.8	4.1510
11	1	2	0.2	20.7139
11	1	2	0.4	6.8548
11	1	2	0.6	4.6211
11	1	2	0.8	13.9368
11	1	4	0.2	2.9411
11	1	4	0.4	2.8108
11	1	4	0.6	3.5840
11	1	4	0.8	11.8940

Table 21 (cont.)

N	NO	N1	D	MAXIVI
11	1	6	0.2	2.5347
11	1	6	0.4	2.9398
11	1	6	0.6	4.4746
11	1	6	0.8	14.3330
11	1	8	0.2	3.8977
11	1	8	0.4	4.8297
11	1	8	0.6	7.8827
11	1	8	0.8	24.9697
11	3	2	0.2	26.5455
11	3	2	0.4	7.9218
11	3	2	0.6	4.2161
11	3	2	0.8	3.3218
11	3	4	0.2	2.9694
11	3	4	0.4	2.8270
11	3	4	0.6	2.7449
11	3	4	0.8	3.2492
11	3	6	0.2	2.4868
11	3	6	0.4	2.6325
11	3	6	0.6	2.9302
11	3	6	0.8	3.3749
11	5	2	0.2	38.2599
11	5	2	0.4	10.4236
11	5	2	0.6	4.9580
11	5	2	0.8	3.0747
11	5	4	0.2	3.0017
11	5	4	0.4	2.9031
11	5	4	0.6	2.7621
11	5	4	0.8	2.5894
11	7	2	0.2	73.4847
11	7	2	0.4	18.2687
11	7	2	0.6	7.9338
11	7	2	0.8	4.4460
12	2	2	0.2	22.4672
12	2	2	0.4	7.2193
12	2	2	0.6	4.2952
12	2	2	0.8	5.0189

Table 21 (cont.)

N	NO	N1	D	MAXIVI
12	2	4	0.2	3.0828
12	2	4	0.4	2.8764
12	2	4	0.6	3.0077
12	2	4	0.8	4.8132
12	2	6	0.2	2.4140
12	2	6	0.4	2.6213
12	2	6	0.6	3.2963
12	2	6	0.8	5.0800
12	2	8	0.2	3.0358
12	2	8	0.4	3.3997
12	2	8	0.6	4.2224
12	2	8	0.8	5.7548
12	4	2	0.2	28.8386
12	4	2	0.4	8.4686
12	4	2	0.6	4.3242
12	4	2	0.8	2.9309
12	4	4	0.2	3.1179
12	4	4	0.4	2.9308
12	4	4	0.6	2.7522
12	4	4	0.8	2.7995
12	4	6	0.2	2.4059
12	4	6	0.4	2.4531
12	4	6	0.6	2.6106
12	4	6	0.8	2.8204
12	6	2	0.2	41.6262
12	6	2	0.4	11.2447
12	6	2	0.6	5.2635
12	6	2	0.8	3.1790
12	6	4	0.2	3.1570
12	6	4	0.4	3.0374
12	6	4	0.6	2.8593
12	6	4	0.8	2.6377
12	8	2	0.2	80.0603
12	8	2	0.4	19.8305
12	8	2	0.6	8.5617
12	8	2	0.8	4.7515

Table 21 (cont.)

N	NO	N1	D	MAXIVI
13	1	2	0.2	20.1007
13	1	2	0.4	6.9399
13	1	2	0.6	4.8627
13	1	2	0.8	15.0406
13	1	4	0.2	3.1981
13	1	4	0.4	2.9343
13	1	4	0.6	3.4567
13	1	4	0.8	11.7499
13	1	6	0.2	2.4384
13	1	6	0.4	2.6255
13	1	6	0.6	3.8262
13	1	6	0.8	12.4150
13	1	8	0.2	2.7203
13	1	8	0.4	3.2320
13	1	8	0.6	5.0463
13	1	8	0.8	16.0786
13	1	10	0.2	4.4382
13	1	10	0.4	5.5521
13	1	10	0.6	9.1522
13	1	10	0.8	28.9681
13	3	2	0.2	24.2270
13	3	2	0.4	7.6193
13	3	2	0.6	4.2151
13	3	2	0.8	3.6253
13	3	4	0.2	3.2355
13	3	4	0.4	2.9691
13	3	4	0.6	2.8306
13	3	4	0.8	3.4981
13	3	6	0.2	2.4354
13	3	6	0.4	2.4709
13	3	6	0.6	2.8328
13	3	6	0.8	3.5688
13	3	8	0.2	2.6570
13	3	8	0.4	2.8485
13	3	8	0.6	3.2274
13	3	8	0.8	3.7859

Table 21 (cont.)

N	NO	N1	D	MAXIVI
13	5	2	0.2	31.1370
13	5	2	0.4	9.0321
13	5	2	0.6	4.4834
13	5	2	0.8	2.9009
13	5	4	0.2	3.2765
13	5	4	0.4	3.0520
13	5	4	0.6	2.8014
13	5	4	0.8	2.5953
13	5	6	0.2	2.4343
13	5	6	0.4	2.4360
13	5	6	0.6	2.4587
13	5	6	0.8	2.5638
13	7	2	0.2	44.9969
13	7	2	0.4	12.0744
13	7	2	0.6	5.5847
13	7	2	0.8	3.3072
13	7	4	0.2	3.3214
13	7	4	0.4	3.1835
13	7	4	0.6	2.9739
13	7	4	0.8	2.7112
13	9	2	0.2	86.6395
13	9	2	0.4	21.3969
13	9	2	0.6	9.1954
13	9	2	0.8	5.0640
14	2	2	0.2	21.5405
14	2	2	0.4	7.2341
14	2	2	0.6	4.4550
14	2	2	0.8	5.5445
14	2	4	0.2	3.3531
14	2	4	0.4	3.0165
14	2	4	0.6	3.0318
14	2	4	0.8	5.1646
14	2	6	0.2	2.4772
14	2	6	0.4	2.5080
14	2	6	0.6	3.1332
14	2	6	0.8	5.2724

Table 21 (cont.)

N	NO	N1	D	MAXIVI
14	2	8	0.2	2.5008
14	2	8	0.4	2.7953
14	2	8	0.6	3.6180
14	2	8	0.8	5.7212
14	2	10	0.2	3.3595
14	2	10	0.4	3.7960
14	2	10	0.6	4.7698
14	2	10	0.8	6.5719
14	4	2	0.2	25.9915
14	4	2	0.4	8.0416
14	4	2	0.6	4.2433
14	4	2	0.8	3.1030
14	4	4	0.2	3.3957
14	4	4	0.4	3.0787
14	4	4	0.6	2.8266
14	4	4	0.8	2.9807
14	4	6	0.2	2.4795
14	4	6	0.4	2.4746
14	4	6	0.6	2.6084
14	4	6	0.8	2.9829
14	4	8	0.2	2.4620
14	4	8	0.4	2.5772
14	4	8	0.6	2.7899
14	4	8	0.8	3.0704
14	6	2	0.2	33.4394
14	6	2	0.4	9.6070
14	6	2	0.6	4.6723
14	6	2	0.8	2.9242
14	6	4	0.2	3.4421
14	6	4	0.4	3.1847
14	6	4	0.6	2.8755
14	6	4	0.8	2.5791
14	6	6	0.2	2.4835
14	6	6	0.4	2.4697
14	6	6	0.6	2.4499
14	6	6	0.8	2.4340

Table 21 (cont.)

N	NO	N1	D	MAXIVI
14	8	2	0.2	48.3708
14	8	2	0.4	12.9102
14	8	2	0.6	5.9164
14	8	2	0.8	3.4504
14	8	4	0.2	3.4923
14	8	4	0.4	3.3378
14	8	4	0.6	3.0998
14	8	4	0.8	2.8005
14	10	2	0.2	93.2214
14	10	2	0.4	22.9667
14	10	2	0.6	9.8331
14	10	2	0.8	5.3814
15	1	2	0.2	19.8173
15	1	2	0.4	7.1041
15	1	2	0.6	5.1341
15	1	2	0.8	16.2470
15	1	4	0.2	3.4701
15	1	4	0.4	3.0719
15	1	4	0.6	3.4313
15	1	4	0.8	11.9468
15	1	6	0.2	2.5270
15	1	6	0.4	2.5521
15	1	6	0.6	3.5513
15	1	6	0.8	11.6945
15	1	8	0.2	2.4095
15	1	8	0.4	2.7637
15	1	8	0.6	4.1575
15	1	8	0.8	13.3570
15	1	10	0.2	2.9470
15	1	10	0.4	3.5572
15	1	10	0.6	5.6480
15	1	10	0.8	17.9475
15	1	12	0.2	4.9915
15	1	12	0.4	6.2855
15	1	12	0.6	10.4322
15	1	12	0.8	33.0051

Table 21 (cont.)

N	NO	N1	D	MAXIVI
15	3	2	0.2	22.9845
15	3	2	0.4	7.5584
15	3	2	0.6	4.3045
15	3	2	0.8	3.9394
15	3	4	0.2	3.5140
15	3	4	0.4	3.1154
15	3	4	0.6	2.9247
15	3	4	0.8	3.7541
15	3	6	0.2	2.5321
15	3	6	0.4	2.5156
15	3	6	0.6	2.8058
15	3	6	0.8	3.7733
15	3	8	0.2	2.3817
15	3	8	0.4	2.5676
15	3	8	0.6	3.0302
15	3	8	0.8	3.9305
15	3	10	0.2	2.8699
15	3	10	0.4	3.1023
15	3	10	0.6	3.5555
15	3	10	0.8	4.2190
15	5	2	0.2	27.7595
15	5	2	0.4	8.4789
15	5	2	0.6	4.3289
15	5	2	0.8	2.8908
15	5	4	0.2	3.5614
15	5	4	0.4	3.1996
15	5	4	0.6	2.8595
15	5	4	0.8	2.7270
15	5	6	0.2	2.5385
15	5	6	0.4	2.5066
15	5	6	0.6	2.5010
15	5	6	0.8	2.6925
15	5	8	0.2	2.3564
15	5	8	0.4	2.4303
15	5	8	0.6	2.5609
15	5	8	0.8	2.7205

Table 21 (cont.)

N	NO	N1	D	MAXIVI
15	7	2	0.2	35.7447
15	7	2	0.4	10.1898
15	7	2	0.6	4.8799
15	7	2	0.8	2.9784
15	7	4	0.2	3.6126
15	7	4	0.4	3.3254
15	7	4	0.6	2.9655
15	7	4	0.8	2.6057
15	7	6	0.2	2.5464
15	7	6	0.4	2.5207
15	7	6	0.6	2.4796
15	7	6	0.8	2.4142
15	9	2	0.2	51.7472
15	9	2	0.4	13.7504
15	9	2	0.6	6.2553
15	9	2	0.8	3.6038
15	9	4	0.2	3.6678
15	9	4	0.4	3.4978
15	9	4	0.6	3.2337
15	9	4	0.8	2.9006
15	11	2	0.2	99.8053
15	11	2	0.4	24.5390
15	11	2	0.6	10.4739
15	11	2	0.8	5.7025

Table 22

Best Designs for k = 1, N = 8 to 20

N	N_0	N_1	N_2	D	MAXIVI
8	2	4	2	.2	2.6630
9	3	4	2	.5	2.6860
10	4	4	2	.8	2.6310
11	5	6	-	-	2.4444
12	6	6	-	-	2.4000
13	5	6	2	.2	2.4343
14	6	8	-	-	2.4267
15	7	8	-	-	2.3333
16	8	8	-	-	2.3111
17	9	8	-	-	2.3296
18	8	10	-	-	2.3314
19	9	10	-	-	2.2800
20	10	10	-	-	2.2667

level design for a particular sample size turn out to be nearly equivalent in terms of their robustness to missing data. Therefore, it makes little difference in this case whether a three-level or a five-level design is used as long as the points are distributed appropriately.

5.2.2 Central Composite Design

The maximum integrated prediction variances for the central composite design are given in Table 23. Robustness is achieved with large values of α, as was the case with outlier robustness. Center points seem to have less affect on these results than they did in the outlier case. Almost no difference exists when more than two center points are used. For the rotatable designs (indicated by *), at least two center points must be included to achieve some degree of robustness. However, even with additional center points the rotatable designs are not very robust in most cases. As in the outlier case, if missing data is anticipated then using an axial value somewhat larger than the rotatable value (preferably as large as possible) is recommended.

5.2.3 Box-Behnken Design

Results for these designs are given in Table 24. The two designs for six variables, 6_1 and 6_2, refer to the designs based on the 2^2 and the 2^3 factorial, respectively. (Only results for multiple center points are given since losing a single center point results in a singular $X'X$ matrix.) From the results in Table 24 it is seen that including more than two center points does little to improve the

Table 23

Maximum Integrated Prediction Variance -i for ccd

k=2

α

N_0	1.0	1.41*	2.0	3.0
1	7.7143	-	7.2188	3.9129
2	7.8922	5.8344	4.4097	3.7613
3	8.3475	4.9046	4.1445	3.7483
4	8.9000	5.0185	4.1750	3.8050
5	9.4977	5.2209	4.2814	3.9018

k=3

α

N_0	1.0	1.68*	2.0	3.0
1	16.7721	266.267	18.9911	5.8089
2	17.4192	8.6815	7.7873	5.5912
3	18.2168	8.0043	6.9573	5.6040
4	19.0869	8.1494	6.9739	5.6833
5	19.9974	8.3955	7.1054	5.8033

k=4

α

N_0	1.0	1.5	2.0*	3.0
1	40.2134	15.3794	-	8.4887
2	41.4375	15.3271	12.6389	8.3734
3	42.7562	15.5670	10.9875	8.3862
4	44.1314	15.9250	11.0056	8.4667
5	45.5429	16.3424	11.1569	8.5897

* Rotatable Design

Table 23 (cont.)

k=5 (full factorial)

	α			
N_0	1.0	2.0	2.38*	3.0
1	93.7908	28.8586	58.7143	12.6698
2	95.5592	18.1724	17.208	12.4176
3	97.4115	17.4481	15.0	12.3468
4	99.3205	17.4480	14.6515	12.3672
5	101.269	17.5813	14.6636	12.4435

k=5 ($\frac{1}{2}$ fraction)

	α			
N_0	1.0	1.5	2.0*	3.0
1	78.1817	31.1509	24.1875	16.1241
2	80.9396	32.0069	21.1739	15.7935
3	83.7194	32.9552	21.4969	15.8065
4	86.5145	33.9526	22.0012	15.9820
5	89.3208	34.9794	22.5831	16.2470

k=6 ($\frac{1}{2}$ fraction)

	α			
N_0	1.0	2.0	2.38*	3.0
1	135.6850	24.9555	170.9700	19.6672
2	138.378	24.8101	22.6336	18.7043
3	141.123	24.9945	21.4189	18.6376
4	143.906	25.3098	21.4989	18.7243
5	146.716	25.6900	21.7159	18.8944

* Rotatable Design

Table 23 (cont.)

k=7 ($\frac{1}{2}$ fraction)

N_0	α = 1.0	2.0	2.8*	3.0
1	289.475	38.3499	66.4438	30.9339
2	292.715	38.3394	29.0038	26.1583
3	296.017	38.5140	26.5254	24.9387
4	299.366	38.7824	26.057	24.7669
5	302.752	39.1049	26.0354	24.7698

* Rotatable Design

Table 24

Maximum Integrated Prediction Variance -i
for Box-Behnken

k	N_0	MAXIVI
3	2	8.8000
	3	8.8571
	4	9.1429
	5	9.5200
4	2	12.6389
	3	10.9875
	4	11.0056
	5	11.1569
5	2	18.1111
	3	15.8122
	4	15.4524
	5	15.5179
6_1	2	25.0325
	3	22.2862
	4	21.6291
	5	21.6420
6_2	2	24.9609
	3	23.0297
	4	23.0479
	5	23.2261
7	2	28.7803
	3	27.2652
	4	27.3232
	5	27.5322

robustness. In this case, then, the best design is achieved with two or three center points.

5.2.4 Hybrid Design

Results for the robustness of the hybrid designs to missing data are given in Table 25 for k of 3, 4 and 6. The same trends appear in all three cases and are the same trends seen in the outlier case. The most robust designs have α, α_1 and $|\alpha_2|$ large and α_3 small. As in the other designs, several center points are required for optimality. It should be noted that if the parameters are not chosen in this manner the design can be extremely poor in terms of its robustness to missing data.

5.2.5 Small Composite Design

Results for selected axial values for the small composite design are given in Table 26, where the rotatable designs are indicated by an asterisk and $\alpha = \sqrt{k}$ is indicated by Δ. In this case there is little improvement in robustness for α greater than $\sqrt{k}$ and in some cases the robustness decreases for larger α-values. Adding more than two center points does not improve the robustness in most cases. Therefore, the recommended design has α approximatley equal to $\sqrt{k}$ and two to three center points.

5.2.6 Equiradial Designs

Results for the single equiradial design with center points appear in Table 27. Here again, as in the Box-Behnken design, only results for more than one center point are given. These results

Table 25

Maximum Integrated Prediction Variance -i

Hybrid Designs

k=3

α	α1	α2	α3	NO	MAXIVI
1	1	-3	1	1	108.5563
1	1	-3	1	3	80.9366
1	1	-3	1	5	82.4598
1	1	-3	2	1	256.0201
1	1	-3	2	3	195.7498
1	1	-3	2	5	201.2146
1	1	-3	3	1	597.0030
1	1	-3	3	3	454.5803
1	1	-3	3	5	466.5999
1	1	-2	1	1	97.2073
1	1	-2	1	3	78.2212
1	1	-2	1	5	81.7951
1	1	-2	2	1	191.6460
1	1	-2	2	3	157.9572
1	1	-2	2	5	166.4429
1	1	-2	3	1	520.4829
1	1	-2	3	3	377.0993
1	1	-2	3	5	394.3875
1	1	-1	1	1	.
1	1	-1	1	3	102.2048
1	1	-1	1	5	116.7857
1	1	-1	2	1	2186.32
1	1	-1	2	3	1914.76
1	1	-1	2	5	2054.94
1	1	-1	3	1	.
1	1	-1	3	3	.
1	1	-1	3	5	.
1	2	-3	1	1	50.7364
1	2	-3	1	3	42.7559
1	2	-3	1	5	45.3633

Table 25 (cont.)

α	α1	α2	α3	NO	MAXIVI
1	2	-3	2	1	61.9637
1	2	-3	2	3	58.4297
1	2	-3	2	5	64.0035
1	2	-3	3	1	97.0050
1	2	-3	3	3	93.9295
1	2	-3	3	5	105.2928
1	2	-2	1	1	42.6741
1	2	-2	1	3	44.5056
1	2	-2	1	5	49.9848
1	2	-2	2	1	.
1	2	-2	2	3	91.5842
1	2	-2	2	5	104.1777
1	2	-2	3	1	337.0557
1	2	-2	3	3	293.1408
1	2	-2	3	5	318.7516
1	2	-1	1	1	184.3263
1	2	-1	1	3	181.0619
1	2	-1	1	5	200.4303
1	2	-1	2	1	92591.8
1	2	-1	2	3	79630.9
1	2	-1	2	5	85005.8
1	2	-1	3	1	20955.2
1	2	-1	3	3	17437.1
1	2	-1	3	5	18428.6
1	3	-3	1	1	38.8492
1	3	-3	1	3	37.1451
1	3	-3	1	5	40.8365
1	3	-3	2	1	47.9083
1	3	-3	2	3	50.3395
1	3	-3	2	5	57.2377
1	3	-3	3	1	.
1	3	-3	3	3	90.6492
1	3	-3	3	5	102.8429
1	3	-2	1	1	1985.43
1	3	-2	1	3	51.5682
1	3	-2	1	5	59.0266

Table 25 (cont.)

α	α1	α2	α3	NO	MAXIVI
1	3	-2	2	1	121.8078
1	3	-2	2	3	125.0415
1	3	-2	2	5	139.9140
1	3	-2	3	1	701.5600
1	3	-2	3	3	633.1500
1	3	-2	3	5	685.3347
1	3	-1	1	1	2352.44
1	3	-1	1	3	2180.92
1	3	-1	1	5	2378.16
1	3	-1	2	1	8337.94
1	3	-1	2	3	7079.6
1	3	-1	2	5	7528.54
1	3	-1	3	1	6085.39
1	3	-1	3	3	5040.2
1	3	-1	3	5	5319.08
2	1	-3	1	1	176.2554
2	1	-3	1	3	188.4923
2	1	-3	1	5	212.9196
2	1	-3	2	1	21132.7
2	1	-3	2	3	18447.8
2	1	-3	2	5	19779.6
2	1	-3	3	1	1525.41
2	1	-3	3	3	1297.16
2	1	-3	3	5	1380.05
2	1	-2	1	1	61.1199
2	1	-2	1	3	64.9691
2	1	-2	1	5	73.2882
2	1	-2	2	1	2600.24
2	1	-2	2	3	2255.22
2	1	-2	2	5	2413.46
2	1	-2	3	1	7343.11
2	1	-2	3	3	6196.13
2	1	-2	3	5	6576.59
2	1	-1	1	1	.
2	1	-1	1	3	42.1726
2	1	-1	1	5	47.5179

Table 25 (cont.)

α	$\alpha 1$	$\alpha 2$	$\alpha 3$	NO	MAXIVI
2	1	-1	2	1	589.1040
2	1	-1	2	3	504.0615
2	1	-1	2	5	537.2665
2	1	-1	3	1	1025.16
2	1	-1	3	3	750.9821
2	1	-1	3	5	760.2321
2	2	-3	1	1	108.7736
2	2	-3	1	3	18.3942
2	2	-3	1	5	20.5136
2	2	-3	2	1	63.5560
2	2	-3	2	3	36.5543
2	2	-3	2	5	40.1927
2	2	-3	3	1	115.6398
2	2	-3	3	3	97.7055
2	2	-3	3	5	103.7464
2	2	-2	1	1	20.6621
2	2	-2	1	3	17.3362
2	2	-2	1	5	19.2221
2	2	-2	2	1	.
2	2	-2	2	3	33.4324
2	2	-2	2	5	37.0795
2	2	-2	3	1	177.5361
2	2	-2	3	3	82.1583
2	2	-2	3	5	88.3239
2	2	-1	1	1	42.9395
2	2	-1	1	3	36.3009
2	2	-1	1	5	38.5520
2	2	-1	2	1	85.9124
2	2	-1	2	3	66.8722
2	2	-1	2	5	69.1616
2	2	-1	3	1	390.1954
2	2	-1	3	3	138.3554
2	2	-1	3	5	142.5356
2	3	-3	1	1	15.5353
2	3	-3	1	3	14.1412
2	3	-3	1	5	15.3432

Table 25 (cont.)

α	α1	α2	α3	NO	MAXIVI
2	3	-3	2	1	67.4578
2	3	-3	2	3	18.8812
2	3	-3	2	5	20.8438
2	3	-3	3	1	.
2	3	-3	3	3	32.8456
2	3	-3	3	5	36.1464
2	3	-2	1	1	19.5381
2	3	-2	1	3	16.6597
2	3	-2	1	5	17.7387
2	3	-2	2	1	28.0869
2	3	-2	2	3	24.2575
2	3	-2	2	5	26.3233
2	3	-2	3	1	75.4948
2	3	-2	3	3	39.4199
2	3	-2	3	5	43.0992
2	3	-1	1	1	68.6105
2	3	-1	1	3	49.3499
2	3	-1	1	5	49.6187
2	3	-1	2	1	137.0687
2	3	-1	2	3	101.1345
2	3	-1	2	5	102.6501
2	3	-1	3	1	254.2625
2	3	-1	3	3	183.0910
2	3	-1	3	5	184.7794
3	1	-3	1	1	719.7696
3	1	-3	1	3	830.8272
3	1	-3	1	5	954.0754
3	1	-3	2	1	350.3889
3	1	-3	2	3	306.2097
3	1	-3	2	5	328.4220
3	1	-3	3	1	1080.73
3	1	-3	3	3	875.2305
3	1	-3	3	5	917.1131
3	1	-2	1	1	232.6593
3	1	-2	1	3	267.6975
3	1	-2	1	5	307.2055

Table 25 (cont.)

α	α1	α2	α3	NO	MAXIVI
3	1	-2	2	1	350.4531
3	1	-2	2	3	300.4116
3	1	-2	2	5	320.3764
3	1	-2	3	1	1257.18
3	1	-2	3	3	984.8151
3	1	-2	3	5	1020.72
3	1	-1	1	1	.
3	1	-1	1	3	114.6734
3	1	-1	1	5	131.1726
3	1	-1	2	1	479.6251
3	1	-1	2	3	441.5344
3	1	-1	2	5	480.5485
3	1	-1	3	1	1959.24
3	1	-1	3	3	1416.89
3	1	-1	3	5	1427.51
3	2	-3	1	1	28.6812
3	2	-3	1	3	26.4844
3	2	-3	1	5	28.8486
3	2	-3	2	1	338.6342
3	2	-3	2	3	49.5817
3	2	-3	2	5	55.4155
3	2	-3	3	1	32474.7
3	2	-3	3	3	32530.4
3	2	-3	3	5	36185.3
3	2	-2	1	1	27.4001
3	2	-2	1	3	23.9088
3	2	-2	1	5	25.6317
3	2	-2	2	1	.
3	2	-2	2	3	44.1207
3	2	-2	2	5	49.4120
3	2	-2	3	1	3241.59
3	2	-2	3	3	3202.67
3	2	-2	3	5	3550.4
3	2	-1	1	1	51.3209
3	2	-1	1	3	39.5806
3	2	-1	1	5	40.8073

Table 25 (cont.)

α	α1	α2	α3	NO	MAXIVI
3	2	-1	2	1	62.9658
3	2	-1	2	3	52.4011
3	2	-1	2	5	57.8753
3	2	-1	3	1	691.4537
3	2	-1	3	3	669.8648
3	2	-1	3	5	738.9268
3	3	-3	1	1	27.1480
3	3	-3	1	3	20.7187
3	3	-3	1	5	21.2834
3	3	-3	2	1	29.1235
3	3	-3	2	3	24.3895
3	3	-3	2	5	26.6876
3	3	-3	3	1	.
3	3	-3	3	3	42.9052
3	3	-3	3	5	47.7536
3	3	-2	1	1	30.8599
3	3	-2	1	3	22.4016
3	3	-2	1	5	22.6012
3	3	-2	2	1	31.4577
3	3	-2	2	3	27.3602
3	3	-2	2	5	29.3040
3	3	-2	3	1	205.7462
3	3	-2	3	3	42.6528
3	3	-2	3	5	47.2773
3	3	-1	1	1	69.8939
3	3	-1	1	3	49.2269
3	3	-1	1	5	49.0984
3	3	-1	2	1	89.2851
3	3	-1	2	3	69.3707
3	3	-1	2	5	71.7012
3	3	-1	3	1	87.8317
3	3	-1	3	3	77.1859
3	3	-1	3	5	82.9186

Table 25 (cont.)

α	$\alpha1$	$\alpha2$	$\alpha3$	NO	MAXIVI
			k=4		
1	1	-3	1	1	338.0324
1	1	-3	1	3	224.8627
1	1	-3	1	5	214.7250
1	1	-3	2	1	496.4188
1	1	-3	2	3	383.8750
1	1	-3	2	5	386.4949
1	1	-3	3	1	6180.15
1	1	-3	3	3	587.5075
1	1	-3	3	5	629.0958
1	1	-2	1	1	209.3591
1	1	-2	1	3	163.4041
1	1	-2	1	5	165.0015
1	1	-2	2	1	4668.19
1	1	-2	2	3	750.4416
1	1	-2	2	5	802.9840
1	1	-2	3	1	5373.53
1	1	-2	3	3	4621.89
1	1	-2	3	5	4802.51
1	1	-1	1	1	2933.71
1	1	-1	1	3	2785.2
1	1	-1	1	5	2969.26
1	1	-1	2	1	4560.66
1	1	-1	2	3	3266.27
1	1	-1	2	5	3205.35
1	1	-1	3	1	22995.7
1	1	-1	3	3	15226.4
1	1	-1	3	5	14513.7
1	2	-3	1	1	122.1572
1	2	-3	1	3	91.6429
1	2	-3	1	5	91.3675
1	2	-3	2	1	142.7874
1	2	-3	2	3	139.6477
1	2	-3	2	5	149.9400

Table 25 (cont.)

α	$\alpha 1$	$\alpha 2$	$\alpha 3$	NO	MAXIVI
1	2	-3	3	1	1121.5
1	2	-3	3	3	1068.18
1	2	-3	3	5	1139.67
1	2	-2	1	1	92.9623
1	2	-2	1	3	95.1026
1	2	-2	1	5	103.1689
1	2	-2	2	1	2457.74
1	2	-2	2	3	2314.79
1	2	-2	2	5	2462.93
1	2	-2	3	1	2858.58
1	2	-2	3	3	2441.54
1	2	-2	3	5	2532.01
1	2	-1	1	1	32568
1	2	-1	1	3	29487.8
1	2	-1	1	5	31063.9
1	2	-1	2	1	3635.04
1	2	-1	2	3	2771.24
1	2	-1	2	5	2777.47
1	2	-1	3	1	13445.7
1	2	-1	3	3	9608.03
1	2	-1	3	5	9421.41
1	3	-3	1	1	83.5155
1	3	-3	1	3	72.6664
1	3	-3	1	5	75.7454
1	3	-3	2	1	9293.72
1	3	-3	2	3	127.9373
1	3	-3	2	5	140.4544
1	3	-3	3	1	2537.82
1	3	-3	3	3	2384.03
1	3	-3	3	5	2534.98
1	3	-2	1	1	1172.89
1	3	-2	1	3	103.5606
1	3	-2	1	5	113.9945
1	3	-2	2	1	25839
1	3	-2	2	3	23853.1
1	3	-2	2	5	25253

Table 25 (cont.)

α	$\alpha 1$	$\alpha 2$	$\alpha 3$	NO	MAXIVI
1	3	-2	3	1	1837.07
1	3	-2	3	3	1559.66
1	3	-2	3	5	1614.74
1	3	-1	1	1	1754.35
1	3	-1	1	3	1535.7
1	3	-1	1	5	1603.4
1	3	-1	2	1	3567.11
1	3	-1	2	3	2773.32
1	3	-1	2	5	2797.01
1	3	-1	3	1	11272.3
1	3	-1	3	3	8331.06
1	3	-1	3	5	8264.68
2	1	-3	1	1	114.5532
2	1	-3	1	3	97.5363
2	1	-3	1	5	101.0625
2	1	-3	2	1	380.5673
2	1	-3	2	3	302.3639
2	1	-3	2	5	307.0076
2	1	-3	3	1	1100.11
2	1	-3	3	3	824.8131
2	1	-3	3	5	822.1707
2	1	-2	1	1	84.3699
2	1	-2	1	3	77.2669
2	1	-2	1	5	81.6361
2	1	-2	2	1	299.7108
2	1	-2	2	3	248.2706
2	1	-2	2	5	255.2390
2	1	-2	3	1	956.1440
2	1	-2	3	3	715.6123
2	1	-2	3	5	712.9040
2	1	-1	1	1	116.0366
2	1	-1	1	3	85.5537
2	1	-1	1	5	91.5692
2	1	-1	2	1	860.7003
2	1	-1	2	3	615.4854
2	1	-1	2	5	603.6844

Table 25 (cont.)

α	$\alpha 1$	$\alpha 2$	$\alpha 3$	NO	MAXIVI
2	1	-1	3	1	7998.1
2	1	-1	3	3	5808.13
2	1	-1	3	5	5727.41
2	2	-3	1	1	697.2586
2	2	-3	1	3	27.1452
2	2	-3	1	5	29.2726
2	2	-3	2	1	231.6929
2	2	-3	2	3	57.9081
2	2	-3	2	5	61.4666
2	2	-3	3	1	764.2391
2	2	-3	3	3	123.8198
2	2	-3	3	5	129.1429
2	2	-2	1	1	34.8240
2	2	-2	1	3	29.2783
2	2	-2	1	5	31.5267
2	2	-2	2	1	132.7980
2	2	-2	2	3	68.0562
2	2	-2	2	5	72.2299
2	2	-2	3	1	368.4634
2	2	-2	3	3	180.2968
2	2	-2	3	5	175.2104
2	2	-1	1	1	128.1401
2	2	-1	1	3	101.6641
2	2	-1	1	5	103.1805
2	2	-1	2	1	1670.13
2	2	-1	2	3	1199.02
2	2	-1	2	5	1177.66
2	2	-1	3	1	18477.9
2	2	-1	3	3	13453.2
2	2	-1	3	5	13278
2	3	-3	1	1	28.7895
2	3	-3	1	3	20.8506
2	3	-3	1	5	22.2587
2	3	-3	2	1	65.1466
2	3	-3	2	3	35.3093
2	3	-3	2	5	37.6432

Table 25 (cont.)

α	$\alpha 1$	$\alpha 2$	$\alpha 3$	NO	MAXIVI
2	3	-3	3	1	149.7796
2	3	-3	3	3	66.7707
2	3	-3	3	5	70.4634
2	3	-2	1	1	35.6699
2	3	-2	1	3	30.7763
2	3	-2	1	5	32.0066
2	3	-2	2	1	81.6705
2	3	-2	2	3	68.3265
2	3	-2	2	5	70.4452
2	3	-2	3	1	345.8250
2	3	-2	3	3	241.9929
2	3	-2	3	5	235.5198
2	3	-1	1	1	205.6026
2	3	-1	1	3	150.1147
2	3	-1	1	5	148.3035
2	3	-1	2	1	3947.32
2	3	-1	2	3	2845.27
2	3	-1	2	5	2798.51
2	3	-1	3	1	68065.6
2	3	-1	3	3	49677.3
2	3	-1	3	5	49071.6
3	1	-3	1	1	114.4731
3	1	-3	1	3	109.0013
3	1	-3	1	5	116.2886
3	1	-3	2	1	1879.97
3	1	-3	2	3	1575.08
3	1	-3	2	5	1624.6
3	1	-3	3	1	97321.8
3	1	-3	3	3	79099.1
3	1	-3	3	5	80866.2
3	1	-2	1	1	123.1117
3	1	-2	1	3	89.8796
3	1	-2	1	5	96.6418
3	1	-2	2	1	939.5534
3	1	-2	2	3	777.6336
3	1	-2	2	5	799.2628

Table 25 (cont.)

α	α1	α2	α3	NO	MAXIVI
3	1	-2	3	1	318730
3	1	-2	3	3	257353
3	1	-2	3	5	262587
3	1	-1	1	1	292.6238
3	1	-1	1	3	82.1920
3	1	-1	1	5	88.5823
3	1	-1	2	1	601.3748
3	1	-1	2	3	491.5091
3	1	-1	2	5	503.3214
3	1	-1	3	1	25565.6
3	1	-1	3	3	20513.5
3	1	-1	3	5	20891.1
3	2	-3	1	1	33.0286
3	2	-3	1	3	30.1267
3	2	-3	1	5	31.7976
3	2	-3	2	1	980.1622
3	2	-3	2	3	65.8369
3	2	-3	2	5	70.3820
3	2	-3	3	1	182.5246
3	2	-3	3	3	159.8034
3	2	-3	3	5	166.8554
3	2	-2	1	1	37.6241
3	2	-2	1	3	31.4636
3	2	-2	1	5	32.4353
3	2	-2	2	1	327.0710
3	2	-2	2	3	63.0632
3	2	-2	2	5	67.3428
3	2	-2	3	1	1987.47
3	2	-2	3	3	151.9733
3	2	-2	3	5	159.1336
3	2	-1	1	1	109.0945
3	2	-1	1	3	78.2346
3	2	-1	1	5	76.8111
3	2	-1	2	1	173.3692
3	2	-1	2	3	140.0563
3	2	-1	2	5	142.9265

Table 25 (cont.)

α	α1	α2	α3	NO	MAXIVI
3	2	-1	3	1	362.5471
3	2	-1	3	3	259.0816
3	2	-1	3	5	265.5809
3	3	-3	1	1	33.1723
3	3	-3	1	3	24.4876
3	3	-3	1	5	24.2827
3	3	-3	2	1	39.1261
3	3	-3	2	3	34.4416
3	3	-3	2	5	36.0140
3	3	-3	3	1	377.4439
3	3	-3	3	3	61.4755
3	3	-3	3	5	65.2243
3	3	-2	1	1	47.1581
3	3	-2	1	3	32.4371
3	3	-2	1	5	31.3710
3	3	-2	2	1	63.4750
3	3	-2	2	3	49.9123
3	3	-2	2	5	50.5174
3	3	-2	3	1	95.9544
3	3	-2	3	3	82.1544
3	3	-2	3	5	85.2563
3	3	-1	1	1	150.2805
3	3	-1	1	3	100.9544
3	3	-1	1	5	96.7693
3	3	-1	2	1	276.9842
3	3	-1	2	3	199.2799
3	3	-1	2	5	195.8767
3	3	-1	3	1	437.9219
3	3	-1	3	3	315.1347
3	3	-1	3	5	314.5373

Table 25 (cont.)

α	$\alpha 1$	$\alpha 2$	$\alpha 3$	NO	MAXIVI
			k=6		
1	1	-3	1	1	11603.8
1	1	-3	1	3	6789.04
1	1	-3	1	5	6031.59
1	1	-3	2	1	6628.87
1	1	-3	2	3	6724.71
1	1	-3	2	5	7081.63
1	1	-3	3	1	51003.7
1	1	-3	3	3	36619.9
1	1	-3	3	5	35171.2
1	1	-2	1	1	5339.31
1	1	-2	1	3	4378.87
1	1	-2	1	5	4378.49
1	1	-2	2	1	40510.4
1	1	-2	2	3	30770.9
1	1	-2	2	5	30087.8
1	1	-2	3	1	361846
1	1	-2	3	3	229228
1	1	-2	3	5	210469
1	1	-1	1	1	34804.6
1	1	-1	1	3	28436.2
1	1	-1	1	5	28404
1	1	-1	2	1	613076
1	1	-1	2	3	395434
1	1	-1	2	5	365607
1	1	-1	3	1	.
1	1	-1	3	3	.
1	1	-1	3	5	.
1	2	-3	1	1	4357.99
1	2	-3	1	3	2749.78
1	2	-3	1	5	2520.81
1	2	-3	2	1	6071.58
1	2	-3	2	3	6437.1
1	2	-3	2	5	6841.07

Table 25 (cont.)

α	α1	α2	α3	NO	MAXIVI
1	2	-3	3	1	36366.9
1	2	-3	3	3	28575.7
1	2	-3	3	5	28226.6
1	2	-2	1	1	3658.15
1	2	-2	1	3	3758.55
1	2	-2	1	5	3968.7
1	2	-2	2	1	33791.7
1	2	-2	2	3	27561.5
1	2	-2	2	5	27517
1	2	-2	3	1	185089
1	2	-2	3	3	127233
1	2	-2	3	5	120407
1	2	-1	1	1	35160.5
1	2	-1	1	3	29613.4
1	2	-1	1	5	29826.7
1	2	-1	2	1	350635
1	2	-1	2	3	246234
1	2	-1	2	5	234745
1	2	-1	3	1	.
1	2	-1	3	3	964174
1	2	-1	3	5	901432
1	3	-3	1	1	3065.78
1	3	-3	1	3	2262.17
1	3	-3	1	5	2192.01
1	3	-3	2	1	6971.64
1	3	-3	2	3	7216.63
1	3	-3	2	5	7632.01
1	3	-3	3	1	33611.4
1	3	-3	3	3	27403.8
1	3	-3	3	5	27356.7
1	3	-2	1	1	4119.92
1	3	-2	1	3	4398.17
1	3	-2	1	5	4680.67
1	3	-2	2	1	33296.8
1	3	-2	2	3	27780.7
1	3	-2	2	5	27909.7

Table 25 (cont.)

α	α1	α2	α3	NO	MAXIVI
1	3	-2	3	1	145760
1	3	-2	3	3	105480
1	3	-2	3	5	101570
1	3	-1	1	1	38161.3
1	3	-1	1	3	32248.9
1	3	-1	1	5	32510.3
1	3	-1	2	1	287850
1	3	-1	2	3	209844
1	3	-1	2	5	202548
1	3	-1	3	1	1040320
1	3	-1	3	3	720978
1	3	-1	3	5	684229
2	1	-3	1	1	1106.45
2	1	-3	1	3	842.5230
2	1	-3	1	5	824.4430
2	1	-3	2	1	1883.81
2	1	-3	2	3	1652.4
2	1	-3	2	5	1682.08
2	1	-3	3	1	57890.7
2	1	-3	3	3	2210.74
2	1	-3	3	5	2288.85
2	1	-2	1	1	628.1181
2	1	-2	1	3	586.0938
2	1	-2	1	5	605.7365
2	1	-2	2	1	2165
2	1	-2	2	3	1639.64
2	1	-2	2	5	1689.36
2	1	-2	3	1	201547
2	1	-2	3	3	148263
2	1	-2	3	5	143525
2	1	-1	1	1	2251.53
2	1	-1	1	3	1824.17
2	1	-1	1	5	1817.81
2	1	-1	2	1	26828.2
2	1	-1	2	3	17711
2	1	-1	2	5	16518.6

Table 25 (cont.)

α	α1	α2	α3	NO	MAXIVI
2	1	-1	3	1	128876
2	1	-1	3	3	80468.6
2	1	-1	3	5	73461.7
2	2	-3	1	1	232.7542
2	2	-3	1	3	223.6314
2	2	-3	1	5	232.6994
2	2	-3	2	1	6420.02
2	2	-3	2	3	598.6820
2	2	-3	2	5	632.7910
2	2	-3	3	1	1862.11
2	2	-3	3	3	1603.11
2	2	-3	3	5	1624.05
2	2	-2	1	1	332.0318
2	2	-2	1	3	337.8605
2	2	-2	1	5	356.0237
2	2	-2	2	1	2209.87
2	2	-2	2	3	1774.96
2	2	-2	2	5	1764.43
2	2	-2	3	1	49665.1
2	2	-2	3	3	36540.2
2	2	-2	3	5	35374.1
2	2	-1	1	1	4693.39
2	2	-1	1	3	3532.42
2	2	-1	1	5	3444.23
2	2	-1	2	1	21842.9
2	2	-1	2	3	15174
2	2	-1	2	5	14412.4
2	2	-1	3	1	78003.1
2	2	-1	3	3	52185.4
2	2	-1	3	5	48910.2
2	3	-3	1	1	72363.6
2	3	-3	1	3	174.8358
2	3	-3	1	5	185.6668
2	3	-3	2	1	647.3113
2	3	-3	2	3	603.3033
2	3	-3	2	5	623.3522

Table 25 (cont.)

α	α1	α2	α3	NO	MAXIVI
2	3	-3	3	1	2209.41
2	3	-3	3	3	1770.15
2	3	-3	3	5	1758.39
2	3	-2	1	1	456.7377
2	3	-2	1	3	413.7099
2	3	-2	1	5	424.5338
2	3	-2	2	1	2683.55
2	3	-2	2	3	2064.25
2	3	-2	2	5	2026.18
2	3	-2	3	1	22605.6
2	3	-2	3	3	16635.2
2	3	-2	3	5	16105.4
2	3	-1	1	1	37643.9
2	3	-1	1	3	28105.2
2	3	-1	1	5	27334.9
2	3	-1	2	1	20906.3
2	3	-1	2	3	14773.3
2	3	-1	2	5	14113.8
2	3	-1	3	1	64985.6
2	3	-1	3	3	44757.3
2	3	-1	3	5	42384
3	1	-3	1	1	464.9516
3	1	-3	1	3	423.3007
3	1	-3	1	5	434.9159
3	1	-3	2	1	1658.4
3	1	-3	2	3	1411.45
3	1	-3	2	5	1425.58
3	1	-3	3	1	3312.07
3	1	-3	3	3	2773.65
3	1	-3	3	5	2789.33
3	1	-2	1	1	339.9042
3	1	-2	1	3	337.8211
3	1	-2	1	5	354.1817
3	1	-2	2	1	1093.52
3	1	-2	2	3	1036.39
3	1	-2	2	5	1075.03

Table 25 (cont.)

α	α1	α2	α3	NO	MAXIVI
3	1	-2	3	1	8218.14
3	1	-2	3	3	2143.41
3	1	-2	3	5	2194.15
3	1	-1	1	1	597.8011
3	1	-1	1	3	503.8715
3	1	-1	1	5	507.6040
3	1	-1	2	1	3188.44
3	1	-1	2	3	2419.37
3	1	-1	2	5	2364.91
3	1	-1	3	1	39968.1
3	1	-1	3	3	27270
3	1	-1	3	5	25739.1
3	2	-3	1	1	121.7485
3	2	-3	1	3	108.4789
3	2	-3	1	5	113.2626
3	2	-3	2	1	15657
3	2	-3	2	3	291.4170
3	2	-3	2	5	306.3161
3	2	-3	3	1	80926.8
3	2	-3	3	3	601.7760
3	2	-3	3	5	627.6181
3	2	-2	1	1	187.2890
3	2	-2	1	3	151.2029
3	2	-2	1	5	150.5251
3	2	-2	2	1	577.5724
3	2	-2	2	3	476.9389
3	2	-2	2	5	477.8035
3	2	-2	3	1	1485.78
3	2	-2	3	3	1173.47
3	2	-2	3	5	1160.87
3	2	-1	1	1	1073.52
3	2	-1	1	3	724.5442
3	2	-1	1	5	681.2280
3	2	-1	2	1	4407.11
3	2	-1	2	3	3042.48
3	2	-1	2	5	2883.52

Table 25 (cont.)

α	$\alpha1$	$\alpha2$	$\alpha3$	NO	MAXIVI
3	2	-1	3	1	77722.3
3	2	-1	3	3	53135.3
3	2	-1	3	5	50187.5
3	3	-3	1	1	124.4742
3	3	-3	1	3	93.7365
3	3	-3	1	5	91.4122
3	3	-3	2	1	275.2883
3	3	-3	2	3	224.8445
3	3	-3	2	5	224.5689
3	3	-3	3	1	581.0763
3	3	-3	3	3	476.2570
3	3	-3	3	5	476.1349
3	3	-2	1	1	277.4631
3	3	-2	1	3	184.9649
3	3	-2	1	5	173.1310
3	3	-2	2	1	818.6944
3	3	-2	2	3	576.5851
3	3	-2	2	5	550.2174
3	3	-2	3	1	1924.79
3	3	-2	3	3	1367.05
3	3	-2	3	5	1308.24
3	3	-1	1	1	1428.12
3	3	-1	1	3	934.6901
3	3	-1	1	5	868.9713
3	3	-1	2	1	6729.37
3	3	-1	2	3	4490.47
3	3	-1	2	5	4204.7
3	3	-1	3	1	200841
3	3	-1	3	3	137571
3	3	-1	3	5	130027

Table 26

Maximum Integrated Prediction Variance -i

Small Composite Designs

k=4

α

N_0	1.0	1.68*	1.75	2.0^{Δ}	2.5	3.0
1	335.467	51.2184	45.2476	-	48.3044	85.6458
2	232.200	47.5259	43.4046	36.0	41.7727	62.7949
3	201.822	47.8069	44.2310	37.4722	40.7954	56.4707
4	189.667	49.0814	45.7248	39.1667	41.2067	54.2783
5	184.800	50.7513	47.4855	40.9500	42.1735	53.7385

k=5

α

N_0	1.0	1.86*	2.0	2.24^{Δ}	2.5	3.0
1	3552.39	126.422	89.2340	-	63.3692	84.2867
2	2166.27	112.832	86.4709	65.6607	59.9529	65.5457
3	1721.61	110.906	87.7673	67.8624	60.3081	60.4970
4	1512.31	111.897	90.0786	70.2991	61.6061	58.8715
5	1397.16	114.053	92.7958	72.8299	63.2813	58.6151

k=6

α

N_0	1.0	2.1*	2.4^{Δ}	2.5	2.7	3.0
1	334.742	71.7304	290.589	308.103	55.3586	55.1799
2	331.466	70.2669	58.5152	56.4828	54.3652	53.4320
3	331.279	70.8578	59.646	57.5284	54.9616	53.3988
4	333.042	72.0703	61.0461	58.855	55.9979	53.9800
5	336.112	73.5525	62.5542	60.2942	57.2177	54.8509

* Rotatable design

Δ $\alpha = \sqrt{k}$

Table 26 (cont.)

k=7

α

N_0	1.0	2.0	2.3*	2.6^{Δ}	2.8	3.0
1	549.962	164.504	137.673	397.280	115.790	116.493
2	555.267	162.929	136.392	120.395	115.899	114.867
3	561.826	163.882	137.652	122.508	117.616	115.712
4	569.284	165.776	139.697	124.912	119.753	117.313
5	577.409	168.161	142.085	127.431	122.062	119.245

* Rotatable design

$\Delta\ \alpha = \sqrt{k}$

Table 27

Maximum Integrated Prediction Variance -i

Single Equiradial Designs

N	NO	MAXIVI
6	2	6.6667
6	3	7.0000
6	4	7.5000
6	5	8.0667
7	2	5.5714
7	3	5.3968
7	4	5.6310
7	5	5.9429
8	2	5.8333
8	3	4.8889
8	4	5.0000
8	5	5.2000
9	2	6.1111
9	3	4.6667
9	4	4.6944
9	5	4.8222
10	2	6.4000
10	3	4.7667
10	4	4.5267
10	5	4.6000

turn out to be quite similar to those obtained for the outlier robustness case. For a given number of points on the circle, about three center points turn out to be optimal. When these results are examined by sample size (see Table 28), the same trend appears as in the outlier case. That is, for smaller sample sizes, two to three center points are best with the remaining points placed equidistant on the circle. For sample sizes larger than about twelve, additional center points are required to obtain optimality.

The results for the combined equiradial design are given in Table 29. It is seen from these results that for sample sizes larger than twelve the radius of the second circle of points has little affect on the robustness as long as it is less than about .8. For sample sizes less than twelve the second radius should be approximately between .5 and .7.

Comparing the two types of equiradial designs shows that the single design with center points is somewhat more robust than the combined design for a given sample size. There is, however, a great deal of flexibility in the choice of designs all of which lead to comparable robustness.

It has been observed that for each of the design types, the designs that are robust to outliers are also robust to missing data. In the next section these design types will be compared to find the most robust classes of designs.

Table 28

Maximum Integrated Prediction Variance -i

Single Equiradial Designs by Sample Size

N	N1	NO	MAXIVI
8	6	2	6.6667
9	6	3	7.0000
9	7	2	5.5714
10	6	4	7.5000
10	7	3	5.3968
10	8	2	5.8333
11	6	5	8.0667
11	7	4	5.6310
11	8	3	4.8889
11	9	2	6.1111
12	7	5	5.9429
12	8	4	5.0000
12	9	3	4.6667
12	10	2	6.4000
13	8	5	5.2000
13	9	4	4.6944
13	10	3	4.7667
14	9	5	4.8222
14	10	4	4.5267
15	10	5	4.6000

Table 29

Maximum Integrated Prediction Variance -i

Combined Equiradial Designs

N	N1	RHO2	MAXIVI
10	5	0.1	202.3172
10	5	0.3	24.4509
10	5	0.5	10.3147
10	5	0.7	7.9279
10	5	0.9	38.6553
12	6	0.1	8.5297
12	6	0.3	7.6197
12	6	0.5	6.5219
12	6	0.7	6.8004
12	6	0.9	36.1386
14	7	0.1	6.6109
14	7	0.3	6.2214
14	7	0.5	5.7421
14	7	0.7	6.4134
14	7	0.9	34.7889
16	8	0.1	5.9628
16	8	0.3	5.7023
16	8	0.5	5.4056
16	8	0.7	6.2178
16	8	0.9	33.9474
18	9	0.1	5.6371
18	9	0.3	5.4314
18	9	0.5	5.2182
18	9	0.7	6.0997
18	9	0.9	33.3725
20	10	0.1	5.4412
20	10	0.3	5.2651
20	10	0.5	5.0986
20	10	0.7	6.0260
20	10	0.9	32.9548

5.3 Comparisons of Design Types

In this section comparisons will be made of the robustness of the various design classes considered. Recall that the criteria used for evaluating both types of robustness involved a multiplication by the sample size in order to account for differences in size as much as possible. Therefore comparisons can be made between the different design types, even those that are of quite different sizes. The case of two independent variables will be treated in the next section. Comparisons of the central composite, Box-Behnken, hybrid and small composite designs will be made in subsequent sections.

5.3.1 Two Independent Variables

The two variable central composite and the equiradial designs will be compared in this section. From the results given in Tables 12 and 18 for outlier robustness and in Tables 23, 28 and 29 for robustness to missing data, it can be seen that the ccd can be made more robust than the equiradial design provided α can be chosen to be at least two. (The differences between the two designs are somewhat more pronounced in the missing data case.) However, if it is not reasonable to set the axial value larger than two, then the equiradial and the ccd are nearly the same in terms of robustness, both to outliers and to missing data. In particular, the rotatable ccd and the equiradial design (which is also rotatable), are nearly equivalent for both types of robustness.

Notice that if α is equal to one, then the ccd for $k=2$ is a 3^2 factorial design. It turns out that the factorial design is much less robust than either the ccd or the equiradial design.

5.3.2 Central Composite vs. Box-Behnken

In comparing the central composite designs with the Box-Behnken designs for a given k (see Tables 12, 13, 23 and 24), it turns out that the robustness of the Box-Behnken designs is nearly equivalent to that of the rotatable ccd's. In fact, for $k=4$ the two are identical, as one would expect, since the four variable Box-Behnken is simply a rotation of the ccd. (This is not true for other values of k, however.) The ccd can be made more robust than the Box-Behnken design by extending the axial value beyond the rotatable value. Thus, the ccd allows more flexibility in achieving robustness than the Box-Behnken designs. Since the two designs have nearly the same sample sizes, there is no advantage in using the Box-Behnken over the ccd. However, if α can be taken larger than the rotatable value, the ccd is more robust than the Box-Behnken and is the preferred design.

5.3.3 Central Composite vs. Small Composite

The results in Tables 12 and 23 for the ccd and in Tables 16 and 26 for the scd show that the ccd is far more robust than scd. Keep in mind that the differences in sample size have been accounted for by multiplying by N and so even though the ccd has a much larger sample size, it is clearly preferable to the smaller design. Even the smaller ccd for five variables in which the 1/2 fraction is used for the factorial portion performs much better than the five variable scd. These trends hold for both types of robustness but the differences between the two designs are more pronounced in the missing data case, as might be expected. It is clear that the near saturated nature and thus the more complicated form of $X'X$ is responsible for the lack of robustness of the scd in comparison to the ccd.

5.3.4 Central Composite vs. Hybrid

The results for the hybrid design seen in Tables 14 and 25 show that this design is extremely poor in terms of its robustness when compared with the central composite design (see Tables 12 and 23). The best hybrid designs are never even close to being as robust as the ccd, especially with respect to robustness to missing data. Thus, even though there is a great deal of flexibility in choosing the parameters of the hybrid design, it still can not be made robust.

5.3.5 Small Composite vs. Hybrid

The appealing feature of these two types of designs is the small sample size. Both are nearly saturated designs. From the above discussions, one sees that neither design is robust. However, situations will arise in which an economical design is required.

For $k = 4$ and $k = 6$, comparisons can be made between these types of designs. The results appear in Tables 14 and 25 for the hybrid design and in Tables 16 and 26 for the small composite design. From these results, it can be seen that for four variables the hybrid design can be made somewhat more robust than the scd. However, for the missing data case the optimal range of parameters is very limited.

If these levels can not be obtained, the design can be much worse than the scd. For six variables the scd is clearly the better choice.

When an economical design is required, the small composite is the best choice. In terms of its robustness it is in many cases better than the hybrid and in the cases where it is not better, it is at least comparable. For all but a small region of the design parameters the hybrid is far worse than the scd. Choosing the appropriate design parameters is much simpler for the scd than for the hybrid and the scd can easily be made rotatable and still be more robust than the hybrid. Thus, although neither of these designs is recommended for general use, in situations requiring an economical design the small composite is recommended.

5.4 Extended Design Region

In sections 1 and 2, above, it was assumed that the region of interest for the composite designs was a sphere defined by the location of the factorial points. The region of interest is the region in which the experimenter intends to use the model for prediction. In many cases the region of interest can not be localized to this extent, and the experimenter may wish to define the region to be as large as possible. In this case, the region is defined by the design points that are the largest distance from the center of the design.

A simple example in two variables will illustrate the difference in the two regions. The inner circle in Figure 3 is the design region defined by the factorial points and the outer circle is the extended design region. Notice that if $\alpha \leq \sqrt{k}$ the two regions are identical. However, for $\alpha > \sqrt{k}$ the extended region is larger than the region described by the factorial points.

In this section the robustness of the central composite and small composite designs will be evaluated for the extended region. These results will be compared with the results for the region originally defined. The same design parameters will be investigated and the design will again be scaled so that the sphere has radius one.

5.4.1 Central Composite Design

The maximum integrated squared bias in prediction for the ccd is given in Table 30 for the extended design region. For $\alpha \leq \sqrt{k}$ the results are the same as for the original region. The designs turn out to be less robust for large values of α. In fact, the optimal design (assuming one is free to choose N_0) always includes $\alpha = \sqrt{k}$, that is, the case in which all of the non-center points in the design lie on the sphere. Thus, the *rotatable designs* are more attractive than in the case of the previous region of interest.

For one center point the rotatable axial value and $\alpha = \sqrt{k}$ should be avoided. For less than three center points, α should be taken somewhat larger than $\sqrt{k}$. For large numbers of center points the rotatable value of α or $\alpha = \sqrt{k}$ are optimal. The best designs include at least four center points.

The results for the missing data case are given in Table 31. It again turns out that the optimal designs have $\alpha = \sqrt{k}$. The

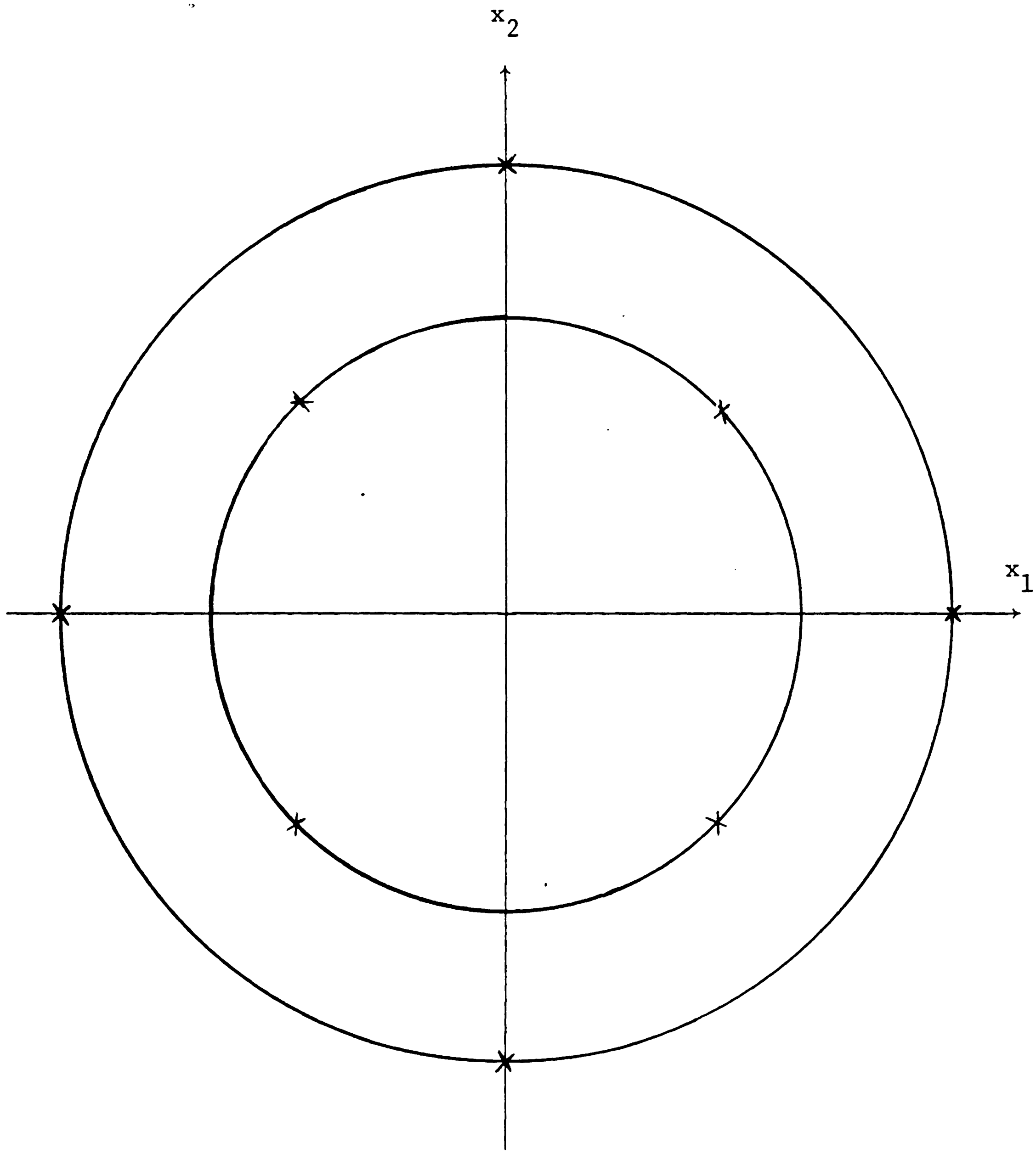


Figure 3

Comparison of Design Regions

Table 30

Maximum Integrated Squared Bias

Extended Design Region - ccd

k=2

α

N_0	1.0	1.41*Δ	1.5	2.0	2.5	3.0
1	0.7778	2.9907	2.6930	0.7778	1.1990	2.1445
2	0.6746	0.8308	0.7685	0.6746	1.2531	2.3172
3	0.7051	0.4062	0.3826	0.7051	1.3394	2.5128
4	0.7500	0.3769	0.4144	0.7500	1.4383	2.7184
5	0.8008	0.4082	0.4471	0.8008	1.5432	2.9293

k=2

α

N_0	1.0	1.68*	1.73Δ	2.0	2.5	3.0
1	1.4950	3.2216	3.4232	1.7090	1.0167	1.8254
2	1.5521	0.8699	0.9129	0.5768	1.0508	1.9253
3	1.6256	0.4125	0.4311	0.5785	1.1004	2.0326
4	1.7065	0.4320	0.4220	0.6070	1.1559	2.1435
5	1.7914	0.4565	0.4455	0.6375	1.2142	2.2565

k=4

α

N_0	1.0	1.5	2.0*Δ	2.25	2.75	3.0
1	3.2563	0.8366	4.1667	1.8605	0.8757	1.1580
2	3.3651	0.8132	1.0833	0.6237	0.9002	1.1973
3	3.4805	0.8200	0.5000	0.5064	0.9295	1.2392
4	3.5995	0.8372	0.3727	0.5225	0.9608	1.2824
5	3.7209	0.8590	0.3860	0.5396	0.9931	1.3264

* Rotatable design

Δ $\alpha = \sqrt{k}$

Table 30 (cont.)

k=5 (full factorial)

N_0	α = 1.0	2.0	2.24^{Δ}	2.38*	2.5	3.0
1	6.9168	2.7748	5.4229	3.4549	1.9052	0.6720
2	7.0637	0.9271	1.3874	0.9917	0.6712	0.6854
3	7.2142	0.5633	0.6307	0.4691	0.4396	0.6998
4	7.3671	0.5674	0.4001	0.4214	0.4476	0.7148
5	7.5215	0.5751	0.4089	0.4309	0.4567	0.7300

k=5 ($\frac{1}{2}$ fraction)

N_0	α = 1.0	1.5	2.0*	2.24^{Δ}	2.5	3.0
1	4.3866	1.0184	1.7500	3.4096	1.4918	1.5494
2	4.5407	1.0352	0.6323	0.8840	0.8696	1.5972
3	4.6967	1.0605	0.6562	0.6301	0.8955	1.6495
4	4.8540	1.0898	0.6798	0.6518	0.9238	1.7037
5	5.0122	1.1211	0.7032	0.6735	0.9532	1.7587

k=6 ($\frac{1}{2}$ fraction)

N_0	α = 1.0	2.0	2.38*	2.45^{Δ}	2.75	3.0
1	8.5032	0.8000	4.0731	4.4959	1.3228	0.8142
2	8.6858	0.6811	1.0700	1.1490	0.6226	0.8281
3	8.8697	0.6786	0.4904	0.5217	0.6336	0.8444
4	9.0546	0.6840	0.4462	0.4432	0.6459	0.8615
5	9.2402	0.6928	0.4557	0.4524	0.6586	0.8789

* Rotatable design

Δ $\alpha = \sqrt{k}$

Table 30 (cont.)

k=7 ($\frac{1}{2}$ fraction)

α

N_0	1.0	2.0	2.6^{Δ}	2.7	2.8*	3.0
1	16.8088	1.1899	6.0838	5.4772	3.5229	1.1554
2	17.0160	1.1706	1.5612	1.4134	1.0274	0.4916
3	17.2242	1.1674	0.7057	0.6401	0.4858	0.4883
4	17.4333	1.1712	0.4629	0.4413	0.4534	0.4901
5	17.6429	1.1786	0.4680	0.4470	0.4589	0.4938

* Rotatable design

Δ $\alpha = \sqrt{k}$

Table 31

Maximum Integrated Prediction Variance -i

Extended Design Region - ccd

k=2

α

N_0	1.0	1.41*Δ	1.5	2.0	2.5	3.0
1	7.7143	-	105.541	7.7143	11.6319	17.9030
2	7.8922	5.8344	5.8078	7.8922	12.5144	19.5789
3	8.3475	4.9046	5.2327	8.3475	13.5195	21.3288
4	8.9000	5.0185	5.4015	8.9000	14.5791	23.1161
5	9.4977	5.2209	5.6517	9.4977	15.6679	24.9247

k=3

α

N_0	1.0	1.68*	1.73Δ	2.0	2.5	3.0
1	16.7721	266.267	-	16.1786	14.263	22.9804
2	17.4192	8.6815	8.6630	9.5095	17.7403	24.1533
3	18.2168	8.0043	7.7939	9.6357	15.4208	25.4222
4	19.0869	8.1494	7.9101	9.9410	16.1802	26.7423
5	19.9974	8.3955	8.1328	10.3227	16.9720	28.0929

k=4

α

N_0	1.0	1.5	2.0*Δ	2.25	2.75	3.0
1	40.2134	15.3794	-	20.7326	20.6470	26.0742
2	41.4375	15.3271	12.6389	13.4927	21.1732	26.9507
3	42.7562	15.5670	10.9875	13.5435	21.8108	27.8700
4	44.1314	15.9250	11.0056	13.7858	22.5052	28.8144
5	45.5429	16.3424	11.1569	14.1111	23.2245	29.7746

* Rotatable design

Δ $\alpha = \sqrt{k}$

Table 31 (cont.)

k=5 (full factorial)

N_0	α 1.0	2.0	2.24^{Δ}	2.38*	2.5	3.0
1	93.7908	28.8586	-	49.0319	25.3067	29.4338
2	95.5592	18.1724	18.1747	18.5856	19.3035	29.8479
3	97.4115	17.4481	15.7500	17.3444	19.1371	30.3459
4	99.3205	17.4480	15.5257	17.3655	19.2815	30.8904
5	101.2690	17.5813	15.5669	17.5165	19.5209	31.4635

k=5 ($\frac{1}{2}$ fraction)

N_0	α 1.0	1.5	2.0*	2.24^{Δ}	2.5	3.0
1	78.1817	31.1509	24.1875	-	25.3211	38.8913
2	80.9396	32.0069	21.1739	19.2025	24.2514	39.7681
3	83.7194	32.9552	21.4969	19.2773	24.6823	40.8536
4	86.5145	33.9526	22.0012	19.6260	25.2946	42.0378
5	89.3208	34.9794	22.5831	20.0842	25.9849	43.2764

k=6 ($\frac{1}{2}$ fraction)

N_0	α 1.0	2.0	2.38*	2.45^{Δ}	2.75	3.0
1	135.685	24.9555	170.9700	-	30.0622	34.4621
2	138.378	24.8101	22.6336	22.5157	27.5431	34.9007
3	141.123	24.9945	21.4189	21.0405	27.7749	35.4714
4	143.906	25.3098	21.4989	21.0886	28.1578	36.1030
5	146.716	25.6900	21.7159	21.2831	28.6093	36.7678

* Rotatable design

Δ $\alpha = \sqrt{k}$

Table 31 (cont.)

k=7 ($\frac{1}{2}$ fraction)

α

N_0	1.0	2.0	2.6^{Δ}	2.7	2.8*	3.0
1	289.475	38.3499	481.5070	318.1800	57.5738	38.3514
2	292.715	38.3394	30.7156	31.0638	32.1833	36.6656
3	296.017	38.5140	27.9595	28.6171	30.9217	36.7234
4	299.366	38.7824	27.6799	28.4644	30.9277	36.9424
5	302.752	39.1049	27.6899	28.5157	31.0725	37.2384

* Rotatable design

Δ $\alpha = \sqrt{k}$

larger axial values are extremely poor in this case. As in the outlier case, for one center point the rotatable axial value and $\alpha = \sqrt{k}$ should be avoided; instead, a somewhat larger value should be used. Including more than two center points has little affect on the robustness to missing data.

For this region the design that is optimal both in terms of robustness to outliers and to missing data has the non-center points approximately equidistant from the center and has at least four center points. If fewer center points are used, an axial value somewhat larger than $\sqrt{k}$ should be chosen. Except for the case of one or two center points, the rotatable designs turn out to be quite robust with this design region.

5.4.2 Small Composite Design

The results for outlier robustness and for robustness to missing data for these designs appear in Tables 32 and 33, respectively. It is clear from these results that as long as more than one center point is used the axial value should be approximately equal to $\sqrt{k}$ to achieve both types of robustness. If only one center point is used, $\sqrt{k}$ should not be chosen for the axial value since this will lead to near singularity of the X'X matrix. More than two center points affects the robustness of the designs only slightly. It turns out that the rotatable scd's are not very robust for this design region. It can be observed in this case as in the case of the original design region that the ccd is far superior to the scd and thus, the smaller design

Table 32

Maximum Integrated Squared Bias

Extended Design Region - scd

k=4

α

N_0	1.0	1.68*	1.75	2.0^{Δ}	2.5	3.0
1	6.4311	1.8681	1.7022	2.8333	1.4562	2.8072
2	6.7957	1.9394	1.7678	1.3008	1.5392	2.9471
3	7.1641	2.0344	1.8554	1.3731	1.6253	3.0978
4	7.5347	2.1352	1.9480	1.4453	1.7119	3.2530
5	7.9066	2.2383	2.0424	1.5176	1.7984	3.4104

k=5

α

N_0	1.0	1.86*	2.0	2.24^{Δ}	2.5	3.0
1	10.6474	2.4340	2.0782	2.9063	1.9441	3.9587
2	11.1033	2.5073	2.1397	1.6761	2.0073	4.1152
3	11.5609	2.6001	2.2201	1.7461	2.0844	4.2792
4	12.0194	2.6983	2.3048	1.8160	2.1646	4.4461
5	12.4787	2.7986	2.3910	1.8859	2.2460	4.6144

k=6

α

N_0	1.0	2.1*	2.4^{Δ}	2.5	2.7	3.0
1	16.4336	2.8006	3.0808	2.9869	2.2122	2.5468
2	16.9271	2.8525	2.1376	2.0873	2.2744	2.6148
3	17.4215	2.9250	2.1989	2.1497	2.3413	2.6879
4	17.9165	3.0030	2.2610	2.2116	2.4088	2.7627
5	18.4119	3.0831	2.3233	2.2734	2.4762	2.8382

* Rotatable design

Δ $\alpha = \sqrt{k}$

Table 32 (cont.)

k=7

N_0	α 1.0	2.0	2.3*	2.6^{Δ}	2.8	3.0
1	24.3828	6.1499	4.9407	4.0770	4.3045	4.7564
2	24.9645	6.3660	5.0989	4.1696	4.4270	4.9252
3	25.5450	6.5577	5.2363	4.2637	4.5363	5.0641
4	26.1244	6.7364	5.3656	4.3581	4.6417	5.1922
5	26.7029	6.9074	5.4909	4.4526	4.7456	5.3154

* Rotatable design

Δ $\alpha = \sqrt{k}$

Table 33

Maximum Integrated Prediction Variance -i

Extended Design Region - scd

k=4

α

N_0	1.0	1.68*	1.75	2.0^{Δ}	2.5	3.0
1	335.467	51.2184	45.2476	-	91.4343	290.246
2	232.200	47.5259	43.4046	36.0000	80.6263	215.820
3	201.822	47.8069	44.2310	37.4722	79.4103	195.615
4	189.667	49.0814	45.7248	39.1667	80.5923	188.966
5	184.800	50.7531	47.4855	40.9500	82.7335	187.740

k=5

α

N_0	1.0	1.86*	2.0	2.24^{Δ}	2.5	3.0
1	3552.39	126.422	89.2340	-	92.4201	241.753
2	2166.27	112.832	86.4709	66.0723	89.2598	192.835
3	1721.61	110.906	87.7673	68.2958	90.4864	180.235
4	1512.31	111.897	90.0786	70.7523	92.8097	176.714
5	1397.16	114.053	92.7958	73.3020	95.5717	176.825

k=6

α

N_0	1.0	2.1*	2.4^{Δ}	2.5	2.7	3.0
1	334.742	71.7304	290.5890	279.0480	74.4478	106.575
2	331.466	70.2669	58.5152	60.1798	74.2820	104.750
3	331.279	70.8578	59.6460	61.4010	75.5576	105.460
4	333.042	72.0703	61.0461	62.8726	77.2322	107.077
5	336.112	73.5525	62.5542	64.4447	79.0732	109.123

* Rotatable design

Δ $\alpha = \sqrt{k}$

Table 33 (cont.)

k=7

α

N_0	1.0	2.0	2.3*	2.6$^{\Delta}$	2.8	3.0
1	549.962	164.504	137.673	397.280	139.531	178.637
2	555.267	162.929	136.392	120.395	140.562	177.574
3	561.826	163.882	137.652	122.508	142.971	179.485
4	569.284	165.776	139.697	124.912	145.740	182.306
5	577.409	168.161	142.085	127.431	148.656	185.525

* Rotatable design

Δ $\alpha = \sqrt{k}$

should be used only when the economics of the particular situation forbids the use of the larger design.

Whether the central composite or the small composite design is used, the optimal design has the axial value near $\sqrt{k}$ and includes several center points when the region of interest extends to the extremes of the design.

As in the case of the original design region it can be shown, based on results in Wardrop (1984), that these optimal designs are also optimal in terms of mean square error.

VI. CONCLUSIONS

6.1 Summary

In this research the effects of an outlier on least squares estimation were investigated for both the mean shift and variance inflation outlier models. Conditions were found for determining when the outlier is of sufficient magnitude to be considered detrimental to the analysis. Based on these conditions, statistical tests were developed which can be used for diagnostic purposes. From this, a general rule-of-thumb can be given that if RSTUDENT is greater than three the outlier may be large enough to warrant deletion of the observation from the analysis.

Criteria were developed for evaluating the robustness of an experimental design to outliers, missing data and errors in control. It turns out that for the first order model the same designs are robust to all three of the conditions studied. The orthogonal designs with second moments as large as possible are optimal.

The robustness of several commonly used classes of second order designs was assessed by computer. The aim was to find optimal design parameters and optimal design classes. The recommendations for the various design types can be summarized as follows:

Central Composite Design - Factorial Design Region

- α as large as practicable
- 2 - 3 center points

Central Composite Design - Extended Design Region

- α near $\sqrt{k}$
- at least 3 center points

Box-Behnken

- 2 - 3 center points

Hybrid Design

- α, α_1, $|\alpha_2|$ large
- α_3 small
- 2 - 3 center points

Small Composite Design - Factorial Region

- α as large as possible - for outlier case
- α near $\sqrt{k}$ - for missing data case
- 2 - 3 center points

Small Composite Design - Extended Design Region

- α near $\sqrt{k}$
- 2 - 3 center points

Equiradial Design

- Single equiradial design
- 3 center points

The optimal design class of those considered is the central composite design. The ccd is somewhat more robust than the Box-Behnken design (and the equiradial design for $k = 2$) and a great deal better than the other types of designs considered. Of the two small designs considered, the small composite is preferred to the hybrid design.

6.2 Further Research

An obvious extension of these results is to re-evaluate these designs based on a cuboidal design region. It was seen in the case of the composite designs that the region of interest did affect the results. Although the optimal design types would probably not change, the optimal parameters for a particular type might possibly change for the cuboidal region.

Additional optimality criteria could also be investigated. A criterion based on the bias induced on the coefficients by an outlier or the inflation in variance due to missing data could easily be developed. Two possibilities for the criterion are the trace of the mean square error matrix of the coefficients and the determinant of this matrix. Of more interest in response surface methodology would be the affect of the outlier or missing data on the estimated slopes of the response function. A criterion could be developed which takes into account the bias or increased variance in the slopes. Since the slopes differ from point to point in the design region, an integrated criterion could be developed as was done in this work.

Another possible extension of this work is to consider robustness to multiple outliers or multiple missing observations. The criterion developed here can be extended to account for multiple outliers and missing observations. Since the results are obtained by computer, these criteria can easily be evaluated.

REFERENCES

Belsley, D.A., Kuh, E. and Welsch, R.E. (1980). *Regression Diagnostics*, John Wiley and Sons, New York.

Box, G.E.P. (1963). "The Effects of Errors in the Factor Levels and Experimental Design," *Technometrics*, 5, 247-262.

Box, G.E.P. and Behnken, D.W. (1960). "Some New Three Level Designs for the Study of Quantitative Variables," *Technometrics*, 2, 455-475.

Box, G.E.P. and Draper, N.R. (1959). "A Basis for the Selection of Response Surface Designs," *J. Amer. Statist. Assoc.*, 54, 622-654.

Box, G.E.P. and Draper, N.R. (1975). "Robust Designs," *Biometrika*, 62, 347-352.

Box, G.E.P. and Hunter, J.S. (1957). "Multifactor Designs for Exploring Response Surfaces," *Annals Math. Statist.*, 28, 195-241.

Box, G.E.P., Hunter, W.G. and Hunter, J.S. (1978). *Statistics for Experimenters, An Introduction to Design, Data Analysis, and Model Building*, John Wiley and Sons, New York.

Box, G.E.P. and Wilson, K.B. (1951). "On the Experimental Attainment of Optimum Conditions," *J. Roy. Statist. Soc., B*, 13, 1-34.

Cook, R.D., Holschuh, N. and Weisberg, S. (1982). "A Note on an Alternative Outlier Model," *J. Roy. Statist. Soc., B*, 44, 370-376.

Cook, R.D. and Weisberg, S. (1982). *Residuals and Influence in Regression*, Chapman-Hall, Inc., New York.

Davies, O.L. (1956). *The Design and Analysis of Industrial Experiments*, 2nd Edition, Hafner Publishing Co., New York.

Draper, N.R. (1983). "On Small Composite Design," Unpublished Manuscript.

Draper, N.R. and Herzberg, A.M. (1979). "Designs to Guard Against Outliers in the Presence or Absence of Model Misspecification," *Canadian J. Statist.*, 7, 127-135.

Draper, N.R. and Smith, H. (1981). *Applied Regression Analysis*, 2nd Edition, John Wiley and Sons, New York.

Graybill, F.A. (1976). *Theory and Application of the Linear Model,* Duxbury Press, Boston, Massachusetts.

Hill, W.J. and Hunter, W.G. (1966). "A Review of Response Surface Methodology: A Literature Survey," *Technometrics,* 8, 571-590.

Hoaglin, D.C. and Welsch, R.E. (1978). "The Hat Matrix in Regression and ANOVA," *The American Statistician,* 32, 17-22.

Myers, R.H. (1976). *Response Surface Methodology,* Distributed by Edwards Brothers, Inc., Ann Arbor, Michigan.

Patnaik, P.B. (1949). "The Non-Central χ^2 and F Distributions and Their Applications," *Biometrika,* 36, 202-232.

Plackett, R.L. and Burman, J.P. (1946). "The Design of Optimum Multifactorial Experiments," *Biometrika,* 33, 305-325.

Roquemore, K.G. (1976). "Hybrid Designs for Quadratic Response Suffaces," *Technometrics,* 18, 419-423.

Wardrop, D.M (1984). Research Notes from Forthcoming PhD Dissertation, Virginia Polytechnic Institute and State University, Blacksburg, Virginia.

APPENDIX A

A.1 Proof of Theorem 3.5

For convenience Theorem 3.5 is restated here.

Theorem 3.5

Under the mean shift outlier model with the usual least squares assumptions on $\underline{\varepsilon}$ and with the assumption that each ε_i has a normal distribution, then

$$(\text{RSTUDENT}_i)^2 = \frac{\hat{\varepsilon}_i^2}{s_{-i}^2(1-h_{ii})} \sim F'_{1,N-p-1,\lambda}$$

where

$$\lambda = \frac{\phi^2(1-h_{ii})}{2\sigma^2}$$

Proof

Given the model in (2.27) and $\underline{\varepsilon} \sim N(\underline{0},\sigma^2 I)$ then

$$\underline{y} \sim N(X\underline{\beta} + \underline{d}_i\phi, \sigma^2 I) \tag{A1.1}$$

since $\underline{y}$ is a linear combination of $\underline{\varepsilon}$ (see Graybill (1976), pg. 101).

Let

$$\begin{aligned}\hat{\underline{\varepsilon}} &= \underline{y} - X\hat{\underline{\beta}} = \underline{y} - X(X'X)^{-1}X'\underline{y} \\ &= [I - X(X'X)^{-1}X']\underline{y} \quad . \end{aligned} \tag{A1.2}$$

From (A1.1) then

$$E(\underline{\hat{\varepsilon}}) = [I - X(X'X)^{-1}X'][X\underline{\beta} + \underline{d}_i\phi]$$

$$= \underline{d}_i\phi - X(X'X)^{-1}\underline{x}_i\phi \qquad (A1.3)$$

and

$$\mathrm{Var}(\underline{\hat{\varepsilon}}) = [I - X(X'X)^{-1}X'][\sigma^2 I][I - X(X'X)^{-1}X']$$

$$= \sigma^2[I - X(X'X)^{-1}X'] \qquad (A1.4)$$

Then, based on these results and the theorem referred to above in Graybill (1976)

$$\underline{\hat{\varepsilon}} \sim N(\underline{d}_i\phi - X(X'X)^{-1}\underline{x}_i\phi,\ \sigma^2[I - X(X'X)^{-1}X']) \ . \qquad (A1.5)$$

In particular,

$$\hat{\varepsilon}_i \sim N(\phi(1-h_{ii}),\ \sigma^2(1-h_{ii}))\ . \qquad (A1.6)$$

It is known that (see Graybill (1976), pg. 127)

$$\frac{\hat{\varepsilon}_i^2}{\sigma^2(1-h_{ii})} \sim \chi^2_{1,\lambda}\ , \qquad (A1.7)$$

where

$$\lambda = \frac{\phi^2(1-h_{ii})}{2\sigma^2}\ . \qquad (A1.8)$$

Now, consider deletion of the i^{th} observation. From (A1.1), clearly

$$\underline{y}_{-i} \sim N(X_{-i}\underline{\beta}, \sigma^2 I_{N-1}) \quad . \tag{A1.9}$$

It can be shown from Graybill (1976, pg. 135) that

$$\frac{(N-p-1)s^2_{-i}}{\sigma^2} = \frac{\underline{y}'_{-i}[I - X_{-i}(X'_{-i}X_{-i})^{-1}X'_{-i}]\underline{y}_{-i}}{\sigma^2} \sim \chi^2_{N-p-1} \tag{A1.10}$$

Notice that the matrix of the quadratic form in (A1.10) is idempotent of rank N-p-1 and the non-centrality parameter for the variable is found as follows:

$$\lambda = \frac{1}{2}\, \underline{\beta}'X'_{-i}[I - X_{-i}(X'_{-i}X_{-i})^{-1}X'_{-i}]X_{-i}\underline{\beta}$$

$$= \frac{1}{2}\, [\underline{\beta}'X'_{-i}X_{-i}\underline{\beta} - \underline{\beta}'X'_{-i}X_{-i}\underline{\beta}] = \phi \; .$$

Therefore, the result in (A1.10) is true.

From the definition of a non-central F variable it is known that the ratio of a non-central chi-square variable and a central chi-square variable each divided by its degrees of freedom is a non-central F variable provided the two are independent. Therefore

$$\frac{\dfrac{\hat{\varepsilon}_i}{\sigma^2(1-h_{ii})}}{\dfrac{s^2_{-i}}{\sigma^2}} = \frac{\hat{\varepsilon}_i}{s^2_{-i}(1-h_{ii})} \sim F'_{1,N-p-1,\lambda} \; , \tag{A1.11}$$

provided that the numerator and denominator are independent. To see this note that

$$\frac{\hat{\varepsilon}_i}{1-h_{ii}} = y_i - \underline{x}_i'\hat{\underline{\beta}}_{-i} \quad , \tag{A1.12}$$

which is the PRESS residual. Obviously, s^2_{-i} and y_i are independent and thus it remains to be shown that s^2_{-i} and $\hat{\underline{\beta}}_{-i}$ are independent. Since s^2_{-i} is a quadratic form in $\underline{y}_{-i}$ and $\hat{\underline{\beta}}_{-i}$ is a linear form in $\underline{y}_{-i}$ the two are proven to be independent if the product of their corresponding matrices is a null matrix, (see Graybill (1976), pg. 137). This is shown as follows:

$$[(X'_{-i}X_{-i})^{-1}X'_{-i}][I - X_{-i}(X'_{-i}X_{-i})^{-1}X'_{-i}] =$$

$$(X'_{-i}X_{-i})^{-1}X'_{-i} - (X'_{-i}X_{-i})^{-1}X'_{-i}X_{-i}(X'_{-i}X_{-i})^{-1}X'_{-i} =$$

$$(X'_{-i}X_{-i})^{-1}X'_{-i} - (X'_{-i}X_{-i})^{-1}X'_{-i} = \phi \quad .$$

This completes the proof of the theorem.

A.2 Proof of Theorem 3.10

For convenience Theorem 3.10 will be repeated here.

Theorem 3.10

Given $\underline{y} = X\underline{\beta} + \underline{\varepsilon}$, where $E(\underline{\varepsilon}) = \underline{0}$,

$$\operatorname{Var}(\underline{\varepsilon}) = \begin{bmatrix} \sigma^2 & & & & & \\ & \sigma^2 & & & \phi & \\ & & \ddots & & & \\ & & & \sigma^2+\sigma^2_\Delta & & \\ & \phi & & & \ddots & \\ & & & & & \sigma^2 \end{bmatrix}$$

and the error terms are normally distributed. Then,

$$\delta_i \sim N(0,\ \sigma^2_\Delta + \frac{\sigma^2}{1-h_{ii}}) \ .$$

Proof

$$\begin{aligned} E(\delta_i) &= E(y_i - \underline{x}_i'\underline{\hat{\beta}}_{-i}) \\ &= E(y_i) - \underline{x}_i' E(\underline{\hat{\beta}}_{-i}) \\ &= \underline{x}_i'\underline{\beta} - \underline{x}_i'\underline{\beta} = 0 \qquad \text{(A2.1)} \end{aligned}$$

$$\begin{aligned}
\mathrm{Var}(\delta_i) &= \mathrm{Var}(y_i - \underline{x}_i'\hat{\underline{\beta}}_{-i}) \\
&= \mathrm{Var}(y_i) + \underline{x}_i' \,\mathrm{Var}(\hat{\underline{\beta}}_{-i})\underline{x}_i \\
&= \sigma^2 + \sigma_\Delta^2 + \sigma^2\underline{x}_i'(X_{-i}'X_{-i})^{-1}\underline{x}_i \qquad \text{(from (3.7))} \\
&= \sigma^2 + \sigma_\Delta^2 + \sigma^2\underline{x}_i'[(X'X)^{-1} + \frac{(X'X)^{-1}\underline{x}_i\underline{x}_i'(X'X)^{-1}}{1-h_{ii}}]\underline{x}_i \quad \text{(from (3.1))} \\
&= \sigma^2 + \sigma_\Delta^2 + \sigma^2 h_{ii} + \frac{\sigma^2 h_{ii}^2}{1-h_{ii}} \\
&= \sigma_\Delta^2 + \sigma^2(1 + h_{ii} + \frac{h_{ii}^2}{1-h_{ii}}) \\
&= \sigma_\Delta^2 + \frac{\sigma^2}{1-h_{ii}} \quad . \qquad \text{(A2.2)}
\end{aligned}$$

Since δ_i is a linear combination of normal variables, it too has a normal distribution. Therefore,

$$\delta_i \sim N(0,\ \sigma_\Delta^2 + \frac{\sigma^2}{1-h_{ii}}) \ .$$

A.3 Proof of Theorem 3.14

For convenience, Theorem 3.14 is restated here.

Theorem 3.14

Assume that the correct model for the data to be fit is the general linear model, $\underline{y} = X\underline{\beta} + \underline{\varepsilon}$, with the assumptions on $\underline{\varepsilon}$ as given in (3.72). Also assume normality of the error terms. Let $s^2_{(-I)}$ be the error mean square obtained by fitting the data with the m outliers deleted. Then, under the null hypothesis stated in (3.88)

$$T_i = \frac{\underline{\delta}_I'[(I_m + H_{(-I)}) + R_i(I_m)]^{-1}\underline{\delta}_I}{ms^2_{(-I)}} \quad i=1,2$$

has an F distribution with m and N-p-m degrees of freedom.

Proof

Under normality of the error terms, clearly, the PRESS residuals, the δ's, are also normally distributed. Thus, from (3.85) and under the null hypothesis in (3.88)

$$\underline{\delta}_I \sim N(\underline{0},\ \sigma^2(I_m + H_{(-I)}) + \sigma^2 R_i I_m) \ . \tag{A3.1}$$

Using a theorem from Graybill (1976, pg. 127) regarding quadratic forms of normal variables,

$$\frac{\underline{\delta}_I'[(I_m + H_{(-I)}) + R_i(I_m)]^{-1}\underline{\delta}_I}{\sigma^2} \sim \chi^2_m \ . \tag{A3.2}$$

The error mean square, $s^2_{(-I)}$, can be written as a quadratic form in $\underline{y}_{(-I)}$ as follows:

$$(N-p-m)s^2_{(-I)} = \underline{y}_{(-I)}[I_{N-m} - X_{(-I)}(X'_{(-I)}X_{(-I)})^{-1}X'_{(-I)}]\underline{y}_{(-I)} \quad . \tag{A3.3}$$

Clearly $\underline{y}_{(-I)} \sim N(X_{(-I)}\underline{\beta}, \sigma^2 I_{N-m})$ and thus it can be shown that (see Graybill (1976) pg. 135)

$$\frac{(N-p-m)s^2_{(-I)}}{\sigma^2} \sim \chi^2_{N-p-m} \; . \tag{A3.4}$$

Using the same logic as in the proof of Theorem 3.5, it can be shown that the ratio of (A3.2) divided by m and (A3.4) divided by N-p-m is an F variable with m and N-p-m degrees of freedom. It can easily be shown that $\underline{\delta}_I$ and $s^2_{(-I)}$ are independent using a proof very similar to that given in the proof of Theorem 3.5.

A.4 Proof of Theorem 3.17

For convenience, the theorem will be restated here.

Theorem 3.17

Assume that the correct model is

$$\underline{y} = X\underline{\beta} + \phi D\underline{j} + \underline{\varepsilon} \; ,$$

with the usual least squares assumptions on $\underline{\varepsilon}$. Also assume normality of the error terms. Let $s^2_{(-I)}$ be as previously defined. Then, under the null hypotheses given in (3.111)

$$\frac{\underline{\delta}_I'[I_m + H_{(-I)}]^{-1}\underline{\delta}_I}{ms^2_{(-I)}} \sim F'_{m,N-p-m,\lambda_i}$$

where

$$\lambda_i = \frac{R_i}{2}\,\underline{j}'[I_m + H_{(-I)}]^{-1}\underline{j} \qquad i=1,2 \quad .$$

Proof

Since the δ's are linear combinations of normal variables, they also have normal distributions. From (3.112) and (3.113),

$$\underline{\delta}_I \sim N(\phi\underline{j},\ \sigma^2[I_m + H_{(-I)}]) \quad . \tag{A4.1}$$

From Graybill (1976) it can be shown that under the null condition $\frac{\phi^2}{\sigma^2} = R_i$

$$\frac{\underline{\delta}_I'[I_m + H_{(-I)}]^{-1}\underline{\delta}_I}{\sigma^2} \sim \chi'^2_{m,\lambda_i} \tag{A4.2}$$

where

$$\lambda_i = \frac{R_i}{2}\,\underline{j}'[I_m + H_{(-I)}]^{-1}\underline{j} \quad . \tag{A4.3}$$

In the proof of Theorem 3.14 it was shown that

$$\frac{(N-p-m)s^2_{(-I)}}{\sigma^2} \sim \chi^2_{N-p-m} \tag{A4.4}$$

and that $s^2_{(-I)}$ and $\underline{\delta}_I$ are independent. Therefore

$$\frac{\underline{\delta}_I'[I_m + H_{(-I)}]^{-1}\underline{\delta}_I}{ms^2_{(-I)}} \sim F'_{m,N-p-m,\lambda_i} \quad .$$

APPENDIX B

B.1 Integrated Squared Bias for ccd

Recall that

$$ISB_i = \frac{NK\phi^2}{\sigma^2} \int_R \underline{x}_i'(X'X)^{-1}\underline{\underline{xx}}'(X'X)^{-1}\underline{x}_i d\underline{x} \ .$$

For the central composite design with design matrix as described in Section 2.4.3, $(X'X)^{-1}$ can be written as

$$(X'X)^{-1} = \left[\begin{array}{c|c|c|c} a & 0 & b\underline{1}_k' & 0 \\ \hline 0 & cI_k & 0 & 0 \\ \hline b\underline{1}_k & 0 & (d-e)I_k + e\underline{1}_k\underline{1}_k' & 0 \\ \hline 0 & 0 & 0 & fI_\ell \end{array}\right]$$

where

$$\ell = \binom{k}{2}$$

$$a = (kF + 2\alpha^4)/D$$

$$b = -(F + 2\alpha^2)/D$$

$$c = 1/(F + 2\alpha^2)$$

$$d = (D + J)/2\alpha^4 D$$

$$e = J/2\alpha^4 D$$

$$f = 1/F$$

$$D = N(kF + 2\alpha^4) - k(F + 2\alpha^2)^2$$

$$J = (F + 2\alpha^2)^2 - NF \quad .$$

Let the region moment matrix be written as

$$\mu = \left[\begin{array}{c|c|c|c} 1 & 0 & \mu^2 \underline{1}'_k & 0 \\ \hline 0 & \mu^2 I_k & 0 & 0 \\ \hline \mu^2 \underline{1}_k & 0 & (\mu^4 - \mu^2_{IJ}) I_k + \mu^2_{IJ} \underline{1}_k \underline{1}'_k & 0 \\ \hline 0 & 0 & 0 & \mu^2_{IJ} I_\ell \end{array}\right] .$$

Then, it turns out that

$$ISB_i = \frac{\phi^2}{\sigma^2} \Bigg\{ [a + b \sum_{j=1}^{k} x^2_{ji}][a + b \sum_{j=1}^{k} x^2_{ji} + \mu^2 \sum_{\ell=1}^{k} (b + dx^2_{\ell i} + e \sum_{j \neq \ell} x^2_{ji})]$$

$$+ c^2 \mu^2 \sum_{j=1}^{k} x^2_{ji} + \sum_{n=1}^{k} [b + dx^2_{ni} + e \sum_{j \neq n} x^2_{ji}][\mu^2 (a + b \sum_{j=1}^{k} x^2_{ji})$$

$$+ \mu^4 (b + dx^2_{ni} + e \sum_{j \neq n} x^2_{ji}) + \mu^2_{IJ} \sum_{\ell \neq n} (b + dx^2_{\ell i} + e \sum_{j \neq \ell} x^2_{ji})]$$

$$+ f^2 \mu^2_{IJ} \sum_{j=1}^{k-1} \sum_{m>j} x^2_{ji} x^2_{mi} \Bigg\} \quad .$$

B.2 Integrated Squared Bias for Box-Behnken Design

The (X'X) matrix for the Box-Behnken designs for $k = 3$ through 6 (see Section 2.4.4) is of the form

$$\left[\begin{array}{c|c|c|c} N & 0 & A\underline{1}_k' & 0 \\ \hline 0 & AI_k & 0 & 0 \\ \hline A\underline{1}_k & 0 & (A-B)I_k + B\underline{1}_k\underline{1}_k' & 0 \\ \hline 0 & 0 & 0 & BI_\ell \end{array}\right]$$

where,

A = the number of non-zero elements for each main effect in the X matrix

B = the number of non-zero elements for each interaction term in the X matrix.

From this it can be shown that

$$(X'X)^{-1} = \left[\begin{array}{c|c|c|c} a & 0 & c\underline{1}_k' & 0 \\ \hline 0 & cI_k & 0 & 0 \\ \hline b\underline{1}_k & 0 & (d-e)I_k + e\underline{1}_k\underline{1}_k' & 0 \\ \hline 0 & 0 & 0 & fI_\ell \end{array}\right],$$

where

$$a = [A + (k-1)B]/D$$

$$b = -A/D$$

$$c = 1/A$$

$$d = [D + (A^2 - NB)]/D(A - B)$$

$$e = (A^2 - NB)/D(A - B)$$

$$f = 1/B$$

$$D = N[A + (k-1)B] - kA^2 .$$

Since the $(X'X)^{-1}$ matrix has the same form as for the ccd, the integrated squared bias for these Box-Behnken designs has the same form as that given above for the ccd, with $a - f$ as given here.

The vita has been removed from the scanned document