TESTING LACK OF FIT IN A MIXTURE MODEL

BY

JOHN THOMAS SHELTON

A DISSERTATION PRESENTED TO THE GRADUATE COUNCIL
OF THE UNIVERSITY OF FLORIDA IN
PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1982

To Nydra
and
My Parents

ACKNOWLEDGEMENTS

I would like to express my deepest appreciation to Drs. Andre' Khuri and John Cornell for suggesting this topic to me and for providing constant guidance and assistance. They have made this project not only a rewarding educational experience but an enjoyable one as well. A special word of thanks goes to Mrs. Carol Rozear for her diligence in transforming my handwritten draft into an expertly typed manuscript.

TABLE OF CONTENTS

	page
ACKNOWLEDGEMENTS.....	iii
ABSTRACT.....	vii
CHAPTER	
ONE INTRODUCTION.....	1
1.1 The Response Surface Problem.....	1
1.2 The Mixture Problem.....	5
1.2.1 Mixture Models.....	6
1.2.2 Experimental Designs for Mixtures..	12
1.3 The Purpose of this Research: Investigation of Procedures for Testing a Model Fitted in a Mixture System for Lack of Fit.....	17
TWO LITERATURE REVIEW--TESTING FOR LACK OF FIT.....	19
2.1 Introduction.....	19
2.2 Partitioning the Residual Sum of Squares..	21
2.3 Testing for Lack of Fit Without Replicated Observations--Near Neighbor Procedures.....	26
2.4 Testing for Lack of Fit with Check Points.	33
THREE AN OPTIMAL CHECK POINT METHOD FOR TESTING LACK OF FIT IN A MIXTURE MODEL.....	40
3.1 Introduction.....	40
3.2 Testing for Lack of Fit in the Presence of an External Estimate of Experimental Error Variation.....	41
3.2.1 The Test Statistic.....	41
3.2.2 The Testing Procedure and an Expression for the Power of The Test.....	45
3.2.3 A Method for Locating Optimal Check Points.....	47
3.3 Testing for Lack of Fit When MSE Is Used to Estimate Experimental Error Variation.....	51
3.3.1 The Test Statistic.....	51
3.3.2 The Rejection Region and its Relation to the Power of the Test..	53

3.3.3	A Method for Locating Optimal Check Points.....	56
3.3.4	Determining Whether the Test Is Upper Tailed or Lower Tailed.....	58
3.4	Examples.....	67
3.4.1	Theoretical Examples.....	67
3.4.2	Numerical Examples.....	83
3.5	Discussion.....	95
FOUR	USE OF NEAR NEIGHBOR OBSERVATIONS FOR TESTING LACK OF FIT.....	99
4.1	Introduction.....	99
4.2	Notation.....	101
4.3	Shillington's Procedure.....	106
4.3.1	Development of MSE_B	109
4.3.2	Development of MSE_W	110
4.4	Development of SSE_W (weighted).....	112
4.5	Equivalence of SSE_W and SSE_W (weighted)....	116
4.6	The Test Statistic.....	118
4.7	The Testing Procedure and its Power.....	122
4.8	Selection of Near Neighbor Groupings.....	125
4.8.1	Example 1--Stack Loss Data.....	129
4.8.2	Example 2--Glass Leaching Data.....	134
4.9	Discussion.....	142
FIVE	CONCLUSIONS AND RECOMMENDATIONS.....	145
APPENDICES		
1	INFLUENCE OF λ_1 ON $P\{F''_{v_1, v_2; \lambda_1, \lambda_2} > F_{\alpha; v_1, v_2}\}$	156
2	A CONTROLLED RANDOM SEARCH PROCEDURE FOR GLOBAL OPTIMIZATION.....	159
3	STATISTICAL INDEPENDENCE OF $\underline{d}'V_0^{-1}\underline{d}/\sigma^2$ AND SSE/σ^2	164
4	THEOREM 3.1.....	168
5	THEOREM 3.2.....	169
6	AN APPROXIMATION TO THE DOUBLY NONCENTRAL F DISTRIBUTION.....	171
7	EQUIVALENCE OF SSE_B AND SS_{LOF} WHEN REPLICATES REPLACE NEAR NEIGHBOR OBSERVATIONS.....	172

8	LEMMA 4.1.....	175
9	PROOF OF THEOREM 4.1(i).....	178
10	PROOF OF THEOREM 4.1(ii).....	182
11	PROOF OF THEOREM 4.1(iii).....	185
12	PROOF OF THEOREM 4.2.....	191
13	PROOF OF THE EQUALITY $SSE_A = \underline{d}'V_0^{-1}\underline{d} + SSE$	193
	REFERENCES.....	198
	BIOGRAPHICAL SKETCH.....	202

Abstract of Dissertation Presented to the
Graduate Council of the University of Florida in
Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy

TESTING LACK OF FIT IN A MIXTURE MODEL

By

John Thomas Shelton

May 1982

Chairman: Andre' I. Khuri
Cochairman: John A. Cornell
Major Department: Statistics

A common problem in modeling the response surface in most systems, and in particular in a mixture system, is that of detecting lack of fit, or inadequacy, of a fitted model of the form $E(\underline{Y}) = \underline{X}\underline{\beta}_1$ in comparison to a model of the form $E(\underline{Y}) = \underline{X}\underline{\beta}_1 + \underline{X}_2\underline{\beta}_2$ postulated as the true model. One method for detecting lack of fit involves comparing the value of the response observed at certain locations in the factor space, called "check points," with the value of the response that the fitted model predicts at these same check points. The observations at the check points are used only for testing lack of fit and are not used in fitting the model.

It is shown that under the usual assumptions of independent and normally distributed errors, the lack of fit test statistic which uses the data at the check points is an

F statistic. When no lack of fit is present the statistic possesses a central F distribution, but in general, in the presence of lack of fit, the statistic possesses a doubly noncentral F distribution. The power of this F test depends on the location of the check points in the factor space through its noncentrality parameters. A method of selecting check points that maximize the power of the test for lack of fit through their influence on the numerator noncentrality parameter is developed.

A second method for detecting lack of fit relies on replicated response observations. The residual sum of squares from the fitted model is partitioned into a pure error variation component and into a lack of fit variation component. Lack of fit is detected if the lack of fit variation is large in comparison to the pure error variation. This method can be generalized when "near neighbor" observations must be substituted for replicates. In this case, the test statistic (assuming independent and normally distributed errors) has a central F distribution when the fitted model is adequate and a doubly noncentral F distribution under lack of fit. The arrangement of near neighbors is seen to affect the testing procedure and its power.

CHAPTER ONE INTRODUCTION

1.1 The Response Surface Problem

A mixture problem is a special type of a response surface problem. First we shall define the general response surface problem and indicate the basic objectives sought in its analysis, and follow this development with a discussion of the mixture problem.

In the general response surface problem, we are interested in studying the relationship between an observable response, Y , and a set of q independent variables or factors, $x_1, x_2, \dots, x_q$, whose levels are assumed controlled by the experimenter. The independent variables are quantitative and continuous. We express this relationship in terms of a continuous response function, ϕ , as

$$Y_u = \phi(x_{u1}, x_{u2}, \dots, x_{uq}) + \varepsilon_u$$

where Y_u is the u th of N observations of the response collected in an experiment, and x_{ui} represents the u th level of the i th independent variable, $u = 1, 2, \dots, N$, $i = 1, 2, \dots, q$. The exact functional relationship, ϕ , is unknown. The term ε_u is the experimental error of the u th

observation. It is assumed that $E(\varepsilon_u) = 0$, $E(\varepsilon_u \varepsilon_{u'}) = 0$, for $u \neq u'$, and $E(\varepsilon_u^2) = \sigma^2$, for $u = 1, 2, \dots, N$.

As the form of ϕ is unknown and may be quite complex, a low order polynomial (usually first or second order) in the independent variables $x_1, x_2, \dots, x_q$ is generally used to approximate ϕ . This may be justified by noting that such polynomials constitute low order terms of a Taylor series expansion of ϕ about the point $x_1 = x_2 = \dots = x_q = 0$, (Myers, 1971, p. 62). Cochran and Cox (1957, p. 336) point out that these low order polynomials may give a poor approximation to ϕ when extrapolated beyond the experimental region, and thus should not be used for this purpose.

A linear response surface model may be written in matrix notation as

$$\underline{Y} = X\underline{\beta} + \underline{\varepsilon} \quad (1.1)$$

where $\underline{Y}$ is an $N \times 1$ vector of observable response values, X is an $N \times p$ matrix of known constants, $\underline{\beta}$ is a $p \times 1$ vector of unknown parameters (regression coefficients), and $\underline{\varepsilon}$ is the $N \times 1$ vector of random errors. When the model is a first or a second degree polynomial, the columns of X correspond to the first or second degree powers of the independent variables $x_1, x_2, \dots, x_q$, or their cross products. If the model contains a constant term, β_0 , the first column of X will correspond to this term, and will consist of N ones. Since $E(\underline{\varepsilon}) = \underline{0}$, an alternative representation for the response

surface model of (1.1) is

$$E(\underline{Y}) = \underline{X}\underline{\beta} .$$

Once the form of the model that will be used to approximate $\phi(x_1, x_2, \dots, x_q)$ is chosen, the next step is to estimate the regression coefficients, $\underline{\beta}$, and then use the estimated model to make inferences about the true response function, ϕ . The estimation of the elements of $\underline{\beta}$ is usually accomplished by ordinary least squares techniques. For the purpose of testing hypotheses concerning the regression coefficients, $\underline{\beta}$, it is assumed that $\underline{\epsilon}$ has a normal distribution, that is, $\underline{\epsilon} \sim N_N(\underline{0}, \sigma^2 \underline{I}_N)$.

Perhaps the most common objective in the exploration of a response system is the determination of its optimum operating conditions. By this we mean that it is desired to find the settings of $x_1, x_2, \dots, x_q$ that optimize ϕ , which in some applications may be interpreted as maximizing ϕ , while in other applications a minimum value of ϕ may be of interest. It is also often desirable to determine the behavior of the response function in the neighborhood of the optimum. For second order response models, such an investigation can be carried out by performing a canonical analysis of the second order surface as discussed in Myers (1971).

For simple systems having only one or two independent variables, the response surface may be explored by just plotting the fitted response values against values taken by

the independent variables. If $q = 1$, implying only one independent variable, say x_1 , then a two-dimensional plot of the fitted response against x_1 may be used to locate the optimum, as well as to investigate the response behavior in other parts of the experimental range of x_1 . If $q = 2$, and the two independent variables are x_1 and x_2 , then a plot of the contours of constant response over the region specified by the ranges of the values for x_1 and x_2 can be used to describe the response surface.

The properties that the fitted model possesses in terms of its ability to represent the true surface, ϕ , depend on the settings of $x_1, x_2, \dots, x_q$ at which values of Y are observed. Thus the experimental design is of great importance. Much work has been done on the construction of designs that are optimal with respect to one criterion or another involving the fitted response and/or the true unfitted model. Box and Draper (1975) list fourteen criteria to consider when choosing a design for investigating response surfaces. Myers (1971) gives several designs for fitting first and second order polynomial models. A discussion of specific design considerations will not be attempted here, as such a discussion is not the focus of this dissertation, and would necessarily be lengthy.

The initial steps in the analysis of a response system may be described as follows: First an attempt is made to approximate the true response function, $\phi(x_1, x_2, \dots, x_q)$, usually with a low order polynomial in $x_1, x_2, \dots, x_q$.

After the form of the model has been chosen, then comes the selection of an appropriate experimental design, which specifies the settings of the independent variables at which observed values of the response will be collected. The observed values of the response are used in estimating the regression coefficients in the model, using, in general, ordinary least squares. After a test for "goodness of fit" of the model verifies the fitted model is adequate, the fitted model is used in determining optimum operating conditions for the response system.

1.2 The Mixture Problem

A mixture system is a response system in which the response depends only on the relative proportions of the components or ingredients present in a mixture, and not on the total amount of the mixture. For example, the response might be the octane rating of a blend of gasolines where the rating is a function only of the relative percentages of the gasoline types present in the blend. The proportion of each ingredient in the mixture, denoted by x_i , must lie between zero and unity, $i = 1, 2, \dots, q$. The sum of the proportions of all the components will equal unity, that is,

$$0 < x_i < 1, i = 1, 2, \dots, q, \quad \sum_{i=1}^q x_i = 1. \quad (1.2)$$

The factor space containing the q components is represented by a $(q - 1)$ -dimensional simplex. For $q = 2$ components, the factor space is a straight line, whereas for $q = 3$

components, the factor space is an equilateral triangle, and for $q = 4$ components, the factor space is represented by a regular tetrahedron.

The objectives in the analysis of a mixture response system are, in general, the same as in any response surface exploration. That is, one seeks to approximate the surface with a model equation by fitting an equation to observations taken at preselected combinations of the mixture components. Another objective is to determine the roles played by the individual components. We shall not concern ourselves with this but rather concentrate on the empirical model fit. Once the model equation is deemed adequate an attempt is made to determine which of the component combinations yield the optimal response. The models used to represent the response in a mixture system are in most cases different in form from the standard polynomial models. The first type of model form that we discuss is the canonical polynomial suggested by Scheffé.

1.2.1 Mixture Models

Scheffé (1958) introduced a canonical form of the polynomial model for representing the response in a mixture system. These canonical polynomial models are derived from the standard polynomials using the restrictions on the x_i shown in (1.2). With $q = 2$ mixture components, for example, the standard second order polynomial model is of the form

$$E(Y) = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_{12} x_1 x_2 + \alpha_{11} x_1^2 + \alpha_{22} x_2^2. \quad (1.3)$$

Restrictions (1.2) imply that $\alpha_0 = \alpha_0(x_1 + x_2)$, $x_1^2 = x_1(1 - x_2)$, and $x_2^2 = x_2(1 - x_1)$, thus (1.3) can be written in the canonical form

$$E(Y) = \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2,$$

where $\beta_1 = \alpha_0 + \alpha_1 + \alpha_{11}$, $\beta_2 = \alpha_0 + \alpha_2 + \alpha_{22}$, and $\beta_{12} = \alpha_{12} - \alpha_{11} - \alpha_{22}$. There is no constant term in the above canonical form and the pure quadratic terms in equation (1.3) have been absorbed in the $x_i x_j$ terms.

The general form of the canonical polynomial of degree d in q mixture components can be written as

$$\begin{aligned} E(Y) &= \sum_{i=1}^q \beta_i x_i, \quad \text{for } d = 1, \\ E(Y) &= \sum_{i=1}^q \beta_i x_i + \sum_{1 \leq i < j}^q \beta_{ij} x_i x_j, \quad \text{for } d = 2, \text{ and} \\ E(Y) &= \sum_{i=1}^q \beta_i x_i + \sum_{1 \leq i < j}^q \beta_{ij} x_i x_j + \sum_{1 \leq i < j}^q \delta_{ij} x_i x_j (x_i - x_j) \\ &\quad + \sum_{1 \leq i < j < k}^q \beta_{ijk} x_i x_j x_k, \quad \text{for } d = 3. \end{aligned} \tag{1.4}$$

The fourth degree canonical polynomial in q components is given in Cornell (1981, p. 64). The general canonical polynomial of degree $d > 4$ in q components does not explicitly appear in the literature, but is mentioned in Scheffé (1958). If terms of the form $\delta_{ij} x_i x_j (x_i - x_j)$ are removed from the full cubic model (1.4), then the remaining terms

make up what is referred to as the special cubic model. For example, for $q = 3$ components, the special cubic model is

$$E(Y) = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 \\ + \beta_{23} x_2 x_3 + \beta_{123} x_1 x_2 x_3 .$$

Scheffé's canonical polynomial models are used for approximating the response surface in many mixture systems. Their popularity stems from the ease in interpreting the coefficient estimates, especially when the models are fitted to data collected at the points of the associated designs (see Section 1.2.2). However, other models have been introduced which seem to better represent the response when the components have strictly additive blending effects. We present some of them now.

Becker (1968) introduced three forms of homogeneous models of degree one which he recommends instead of the polynomial models when one or more of the mixture components have an additive effect or when one or more components are inert. A function $f(x, y, \dots, z)$ is said to be homogeneous of degree n when $f(tx, ty, \dots, tz) = t^n f(x, y, \dots, z)$, for every positive value of t and $(x, y, \dots, z) \neq (0, 0, \dots, 0)$. These models, which Becker refers to as models H1, H2, and H3, are of the form

$$H1: E(Y) = \sum_{i=1}^q \beta_i x_i + \sum_{1 \leq i < j}^q \beta_{ij} \min(x_i, x_j) + \dots$$

$$+ \beta_{12\dots q} \min(x_1, x_2, \dots, x_q),$$

$$H2: E(Y) = \sum_{i=1}^q \beta_i x_i + \sum_{1 \leq i < j}^q \beta_{ij} x_i x_j / (x_i + x_j)^{2-1} + \dots$$

$$+ \beta_{12\dots q} x_1 x_2 \dots x_q / (x_1 + x_2 + \dots + x_q)^{q-1},$$

$$H3: E(Y) = \sum_{i=1}^q \beta_i x_i + \sum_{1 \leq i < j}^q \beta_{ij} (x_i x_j)^{1/2} + \dots$$

$$+ \beta_{12\dots q} (x_1 x_2 \dots x_q)^{1/q}.$$

Each term in the H2 model is defined to be zero when the denominator of the term is zero.

Draper and St. John (1977) suggest a model which includes inverse terms, $1/x_i$, in addition to terms in the Scheffé polynomials. Such a term is used to model an extreme change in the response as x_i approaches zero. The experimental region of interest is assumed to include the region near the zero boundary ($x_i = 0$), but does not include the boundary itself. One example of this type of model is the Scheffé linear polynomial model with inverse terms

$$E(Y) = \sum_{i=1}^q \beta_i x_i + \sum_{i=1}^q \beta_{-i} x_i^{-1}.$$

Another model form that is useful in the study of the response in a mixture system is the model containing ratios of the component proportions. A term such as x_i/x_j measures the relationship of x_i to x_j rather than the percentage of each in the blends. Snee (1973) points out that the ratio model presents a useful alternative to the Scheffé and Becker models in that the ratio model describes a different type of curvature. He notes that the curvilinear terms for the Scheffé and Becker models, when plotted as a function of x_i , are symmetric functions about $x_i = 1/2$, whereas the ratio term x_i/x_j is a monotone function when plotted against x_i .

The terms in the ratio models may also contain sums of the components. For example, with $q = 3$ components, we might express the second order model

$$E(Y) = \beta_0 + \sum_{i=1}^{q-1} \beta_i z_i + \sum_{1 \leq i < j}^{q-1} \beta_{ij} z_i z_j + \sum_{i=1}^{q-1} \beta_{ii} z_i^2$$

(note the constant term) where z_1 and z_2 are defined as $z_1 = x_1/(x_2 + x_3)$ and $z_2 = x_2/x_3$. Some terms will be undefined if points from the boundary of the experimental simplex are included in the design, for example, if $x_3 = 0$, then $z_2 = x_2/x_3$ is not defined. Snee (1973) suggests adding a small positive quantity, c , to each x_i in this case. This, of course, will not be of concern if the experimental region is entirely inside of the simplex.

When one or more of the components is inactive, Becker (1978) suggests that a ratio model that is homogeneous of degree zero in the remaining components is appropriate. In three components, such a model is of the form

$$\begin{aligned}
 E(Y) = & \beta_0 + \beta_1 x_1 / (x_1 + x_2) + \beta_2 x_2 / (x_2 + x_3) \\
 & + \beta_3 x_3 / (x_1 + x_3) + \sum_{1 \leq i < j}^3 \sum \beta_{ij} h_{ij}(x_i, x_j) \\
 & + \beta_{123} h_{123}(x_1, x_2, x_3),
 \end{aligned} \tag{1.5}$$

where h_{ij} and h_{123} are specified functions that are homogeneous of degree zero. The function h_{123} is intended to represent the joint effect of all three components simultaneously. If in fitting a model of the form (1.5) we determine the model should be

$$E(Y) = \beta_0 + \beta_1 x_1 / (x_1 + x_2) + \beta_{12} h_{12}(x_1, x_2)$$

then component three is said to be inactive and is removed from further consideration. The model of equation (1.5) may produce an extreme value near the vertices of the simplex factor space when there are no inactive components. In this case it is suggested that a model of the form (1.5) be used only when the proportions are restricted so that no two of the x_i are simultaneously very close to zero. Becker notes that other authors who have suggested ratio models have also

used them primarily over a subregion inside the simplex factor space. Apparently this is where they are most appropriate.

1.2.2 Experimental Designs for Mixtures

As in the general response surface problem, one of the major concerns in exploring a mixture system is that of choosing the experimental design for collecting observed values of the response that will be used in fitting the model. Scheffé (1958) proposed the $\{q,m\}$ simplex lattice designs for exploring the entire q -component simplex factor space. In these designs, the proportions used for each component have the $m + 1$ values spaced equally from zero to one, $x_i = 0, 1/m, 2/m, \dots, (m - 1)/m, 1$, and all possible mixtures with these proportions for each component are used. The number of design points in the $\{q,m\}$ simplex lattice design is $\binom{m+q}{m} - 1$. The main appeal of these designs is that they provide a uniform coverage of the factor space. Another feature, which Scheffé (1958) demonstrates, is that the parameters of the canonical polynomial of degree m in q components are expressible as simple linear combinations of the true response values at the design points of the $\{q,m\}$ simplex lattice. The $\{3,2\}$ simplex lattice, which consists of six design points, is represented on the two dimensional simplex in Figure 1 along with the triangular coordinates (x_1, x_2, x_3) .

Scheffé (1963) also developed the simplex centroid designs consisting of $2^q - 1$ points, where the only mixtures

considered are the ones in which the components present appear in equal proportions. That is, in a q -component simplex centroid design, the design points correspond to the q permutations of $(1, 0, 0, \dots, 0)$, the $\binom{q}{2}$ permutations of $(1/2, 1/2, 0, \dots, 0)$, the $\binom{q}{3}$ permutations of $(1/3, 1/3, 1/3, 0, \dots, 0)$, ..., and the point $(1/q, 1/q, \dots, 1/q)$. This design alleviates the problem inherent in the $\{q, m\}$ simplex lattice designs of observing responses at mixtures containing at most m components. To give an example, the $q = 3$ simplex centroid design is made up of $2^3 - 1 = 7$ design points, and is equivalent to the $\{3, 2\}$ simplex lattice design augmented by the center point $(x_1, x_2, x_3) = (1/3, 1/3, 1/3)$. This design is represented in Figure 2.

Scheffé (1963) mentions that the number of parameters in the polynomial model of the form

$$\begin{aligned}
 E(Y) = & \sum_{i=1}^q \beta_i x_i + \sum_{1 \leq i < j}^q \beta_{ij} x_i x_j + \sum_{1 \leq i < j < k}^q \beta_{ijk} x_i x_j x_k \\
 & + \dots + \beta_{12\dots q} x_1 x_2 \dots x_q
 \end{aligned} \tag{1.6}$$

is $2^q - 1$ and therefore these models have a special relationship with the simplex centroid design in q components. This relationship is that the number of terms in the model equals the number of points in the design and as a result the parameters in model (1.6) are expressible as simple functions of the responses at the $2^q - 1$ points of the simplex centroid design. Polynomial models of the form (1.6)

Figure 1. The $\{3,2\}$ simplex lattice design.

Figure 2. The $q = 3$ simplex centroid design.

therefore are natural models to fit using the simplex centroid design.

Ratio models may be desirable when the interest in one or more of the mixture components is in terms of their relationship to one another, rather than in terms of their percentages in blends. Kenworthy (1963) proposed factorial arrangements for ratio variables. An example of the use of ratios is the following three component system in which the mixture components are constrained by the upper and lower bounds:

$$.2 \leq x_1 \leq .4, \quad .2 \leq x_2 \leq .4, \quad .3 \leq x_3 \leq .5. \quad (1.7)$$

The ratio variables of interest are $z_1 = x_2/x_1$ and $z_2 = x_2/x_3$, and it is desired to fit either a first or a second order polynomial model in z_1 and z_2 . For such a problem, we can define a 2^2 and a 3^2 factorial design that can be used for fitting the first and second order polynomial models, respectively, by taking as design points the intersection of rays passing from two of the three vertices of the two-dimensional simplex through the region of interest defined by the constraints (1.7). Kenworthy's 2^2 factorial design is shown in Figure 3.

Becker (1978) uses rays extending from one or more vertices of the simplex factor space to the opposite boundaries in developing "radial designs." These designs are suggested for detecting the presence of an inactive

Figure 3. Kenworthy's 2^2 factorial design.

component or in another case a component which has an additive effect, when models containing ratio terms that are homogeneous of degree zero are fitted.

McLean and Anderson (1966) suggest an algorithm for locating the vertices of a restricted region of the simplex factor space which is defined by the placing of upper and lower bounds on the mixture component proportions. The vertices of the factor space and convex combinations of the vertices are the candidates for design points for fitting a first or second degree polynomial model in the mixture components. One drawback of the "extreme vertices" design is that the design points are not uniformly distributed over the factor space resulting in an imbalance in the variances of $\hat{Y}(\underline{x})$, see Cornell (1973).

Another method that has been suggested for studying the response over a sub-region of the simplex mixture space is to transform the q mixture components into $q - 1$ independent variables. Transforming to an independent variable system was first suggested by Claringbold (1955) and later proposed by Draper and Lawrence (1965a, 1965b) and Thompson and Myers (1968). Standard response surface polynomial models in the transformed variables can be fitted to data values collected on standard designs and a design criterion such as the average mean square error of the response can be employed when distinguishing between designs. Thompson and Myers (1968) suggest the use of rotatable designs (see also Cornell and Good, 1970).

Designs other than rotatable designs, such as multiple lattices and symmetric-simplex designs, to name a few, have been suggested in the literature for fitting models to a mixture system which may be appropriate depending on particular experimental situations. However, as the intent here is not to give an exhaustive list but only a sampling of available designs, we shall not discuss designs further but instead state the purpose of this work.

1.3 The Purpose of this Research: Investigation of Procedures for Testing a Model Fitted in A Mixture System for Lack of Fit

A common problem in modeling the response in a mixture system is that of detecting lack of fit, or inadequacy, of a fitted model of the form $E(\underline{Y}) = X\underline{\beta}_1$ when the true model is of the form $E(\underline{Y}) = X\underline{\beta}_1 + X_2\underline{\beta}_2$. The statistical literature

suggests several procedures for testing lack of fit, which will be described in Chapter Two. In general, the authors of these procedures are not specific in stating hypotheses to be tested and do not adequately discuss the power of their procedures.

The major purpose of this research is to investigate the power of two of the testing procedures appearing in the literature in detecting the inadequacy of a fitted model when the general form of the true model is specified. Our findings for a "check points" lack of fit testing procedure are presented in Chapter Three while Chapter Four contains findings for a "near neighbor" lack of fit testing procedure. For both procedures, explicit formulas for the power of the test are given in terms of cumulative probabilities of either the noncentral F or doubly noncentral F distribution, which are derived by assuming that the response observations are independent and normally distributed. Additionally, we propose methods for maximizing the power of the testing procedures. In the final chapter, we make some concluding comments concerning both of these procedures.

CHAPTER TWO
LITERATURE REVIEW--TESTING FOR LACK OF FIT

2.1 Introduction

Let us return to the general response surface problem and assume the true response is to be approximated by fitting a model of the form

$$E(\underline{Y}) = X\underline{\beta}_1 \quad (2.1)$$

where $\underline{Y}$ is an $N \times 1$ vector of observable response values, X is an $N \times p$ matrix of known constants, and $\underline{\beta}_1$ is a $p \times 1$ vector of unknown regression coefficients. We wish to consider the situation in which the true model contains terms in addition to those in the fitted model. Then the true model has the form

$$E(\underline{Y}) = X\underline{\beta}_1 + X_2\underline{\beta}_2 \quad (2.2)$$

where X_2 is an $N \times p_2$ matrix of known constants, and $\underline{\beta}_2$ is a $p_2 \times 1$ vector of unknown regression coefficients. We assume that the vector $\underline{Y}$ has the normal distribution with $\text{var}(\underline{Y}) = \sigma^2 I_N$.

It is desirable to determine the suitability of the fitted model given by Eq. (2.1) when in reality the true model is of the form given by Eq. (2.2). The process of

making this determination is referred to as testing for lack of fit of the fitted model.

There are three general approaches to testing for lack of fit. The first approach requires that there be replicate observations of the response at one or more design points, and involves partitioning the residual sum of squares from the fitted model into a sum of squares due to lack of fit and a sum of squares due to pure error. A large value for the ratio of the mean square due to lack of fit to the mean square due to pure error provides evidence for lack of fit.

If replicate observations are not available then the above approach to testing for lack of fit cannot be used. Green (1971), Daniel and Wood (1971), and Shillington (1979) have proposed alternative methods that are applicable in this case. Their approach is to group values of the response which are observed at similar settings of the independent variables and to call these grouped values "pseudoreplicates" or "near neighbor observations." They then treat these pseudoreplicates as they would treat true replicates to form statistics for lack of fit testing, although arriving at their respective statistics through different approaches.

A third approach to testing for lack of fit involves the use of "check points." In this method a model of the form (2.1) is fitted to data at the design points and additional observations are collected at other points in the experimental region. The additional points other than the

design points are called check points, and the data at these check points are not used in fitting the model. Lack of fit is tested by comparing the values of the response observed at the check points to the values of the response which the fitted model predicts at these same check points.

We now discuss the first method mentioned above of testing for lack of fit which involves partitioning the residual sum of squares.

2.2 Partitioning the Residual Sum of Squares

The method for testing lack of fit which makes use of a partitioning of the residual sum of squares from the fitted model requires there be replicate observations of the response at some of the design points (Draper and Smith, 1981, p. 120). When a model of the form (2.1) is fitted, the residual sum of squares is defined as

$$\begin{aligned} \text{SSE} &= \sum_{i=1}^n \sum_{j=1}^{n_i} (Y_{ij} - \hat{Y}_i)^2 \\ &= \underline{Y}'(\underline{I}_N - \underline{X}(\underline{X}'\underline{X})^{-1}\underline{X}')\underline{Y} \end{aligned}$$

where n is the number of distinct design points, $n_i > 1$ is the number of replicate observations at the i th design point, Y_{ij} is the j th observed value of the response at the i th design point, $\hat{Y}_i$ is the value which the model of the form in Eq. (2.1), fitted by ordinary least squares techniques, predicts for the response at the i th design point, and $N = \sum_{i=1}^n n_i$. Using the replicated observations

only, a pure error sum of squares can be calculated as

$$SSE_{\text{pure}} = \sum_{i=1}^n \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{i.})^2,$$

where $\bar{Y}_{i.}$ is the average of the values of the response observed at the i th design point. The sum of squares due to lack of fit can be obtained by taking the difference

$$SS_{\text{LOF}} = SSE - SSE_{\text{pure}}.$$

This partitioning of the residual sum of squares is displayed in the analysis of variance table in Table 1.

Table 1. Analysis of Variance--
Partitioning the Residual Sum of Squares.

<u>Source of Variation</u>	<u>Sum of Squares</u>	<u>Degrees of Freedom</u>	<u>Mean Square</u>
Regression	$\underline{b}_1' \underline{X}' \underline{Y} - (\underline{1}' \underline{Y})^2 / N$	$p - 1$	
Residual	SSE	$N - p$	MSE
Pure Error	SSE_{pure}	$N - n$	MSE_{pure}
Lack of Fit	SS_{LOF}	$n - p$	MS_{LOF}
Total(corrected)	$\underline{Y}' \underline{Y} - (\underline{1}' \underline{Y})^2 / N$	$N - 1$	

$\underline{b}_1$ represents the ordinary least squares estimator of $\underline{\beta}$ in model (2.1), $\underline{b}_1 = (\underline{X}' \underline{X})^{-1} \underline{X}' \underline{Y}$, and $\underline{1}$ is an $N \times 1$ vector of ones.

To test the hypothesis of zero lack of fit, that is

H_0 : lack of fit = 0 or $E(\underline{Y}) = X\underline{\beta}_1$, an F statistic is formed

$$F = \frac{MS_{LOF}}{MSE_{pure}} \quad (2.3)$$

which possesses a central F distribution if the true model is of the form (2.1), but has a noncentral F distribution if the true model is of the form (2.2). In other words

$$F \sim F_{n-p, N-n}$$

under H_0 : $E(\underline{Y}) = X\underline{\beta}_1$, and

$$F \sim F'_{n-p, N-n; \lambda_2}$$

under H_a : $E(\underline{Y}) = X\underline{\beta}_1 + X_2\underline{\beta}_2$, where λ_2 is the noncentrality parameter $\lambda_2 = \underline{\beta}_2'(X_2 - XA)'(X_2 - XA)\underline{\beta}_2/2\sigma^2$, and $A = (X'X)^{-1}X'X_2$.

Under H_a , $E(MS_{LOF}) = \sigma^2 + \underline{\beta}_2'(X_2 - XA)'(X_2 - XA)\underline{\beta}_2/(n-p)$ and $E(MSE_{pure}) = \sigma^2$ (Draper and Smith, 1981, p. 120), hence H_0

is rejected in favor of H_a if the value of F in (2.3) exceeds the upper 100α percentage point of the central F distribution, $F_{\alpha; n-p, N-n}$. When H_0 is rejected, we conclude that a significant lack of fit is present.

Draper and Herzberg (1971) demonstrated that the lack of fit sum of squares can be partitioned into two statistically independent sums of squares, SS_{L1} and SS_{L2} , when there are replicate observations at the center of the

response surface design and when the basic design without center points is nonsingular. If the true model and the fitted model are of the same form as in equation (2.1) then the two F ratios $F_{L1} = [SS_{L1}/(n - p - 1)]/MSE_{\text{pure}}$ and $F_{L2} = SS_{L2}/MSE_{\text{pure}}$ are both distributed as central F random variates, with respective numerator and denominator degrees of freedom $(n - p - 1)$, $(N - n)$ for F_{L1} and 1, $(N - n)$ for F_{L2} . If the true model is of the form shown in equation (2.2), then F_{L1} and F_{L2} are both distributed as noncentral F random variates. The expected values of SS_{L1} and SS_{L2} are used to show what functions of β_2 are testable with F_{L1} and F_{L2} .

Two examples are presented by Draper and Herzberg to illustrate this testing for lack of fit. The first example makes use of a first order orthogonal design in k factors augmented with center point replicates for fitting a first order polynomial model. If the true model is of the second order, then F_{L2} can be used to test a hypothesis concerning the parameters associated with the pure quadratic terms in the model. If all such parameters are zero, then F_{L1} provides a check on interaction terms. The second example illustrates the fitting of a second order polynomial model to a second order design with all odd design moments of order six or less zero. If the true model is third degree, then F_{L1} can be used to test the significance of the third order terms, while F_{L2} tests terms of order greater than three. The partitioning of SS_{LOF} into SS_{L1} and SS_{L2} and the

corresponding tests of hypotheses are also given in Myers (1971, p. 114-119), for the special case of fitting a first order polynomial model to a 2^q factorial or a fraction of a 2^q factorial design augmented with center point replicates and the true model is of the second degree.

A more complete partitioning of the lack of fit sum of squares in an attempt to obtain a more detailed diagnosis of the lack of fit of the fitted model is given in a technical report written by Khuri and Cornell (1981). The lack of fit sum of squares, which has $n - p$ degrees of freedom, is partitioned into $n - p$ independent sums of squares, each having one degree of freedom. The expected values of these single degree-of-freedom sums of squares are used to identify at most $n - p$ linearly independent causes for the lack of fit variation. Tests of significance are performed on the assumed contributing causes. This method enables the screening of all subsets of $\mathcal{S}_2$ in order to identify those subsets which are most responsible for lack of fit of the fitted model.

We shall now discuss the second general approach used in lack of fit testing, which is to test for lack of fit by making use of response values observed at points which are near neighbors in the factor space when true replicate observations are not available.

2.3 Testing for Lack of Fit Without Replicated Observations--Near Neighbor Procedures

Green (1971) suggests the following approach when testing for lack of fit if there are no design points at which replicate observations of the response are available. The N observed values of a response, Y , considered a function of only one variable, x , are divided into g groups, by grouping observations which have similar values of x . Green hypothesizes a model of the form $\underline{Y} = H\underline{\beta} + \underline{\varepsilon}$, where $\underline{Y}$ is an $N \times 1$ vector of observable responses, H is an $N \times m$ matrix whose columns correspond to known functions of the variable, x , $\underline{\beta}$ is an $m \times 1$ vector of unknown regression coefficients, and $\underline{\varepsilon}$ is the $N \times 1$ vector of random errors, $\underline{\varepsilon} \sim N_N(\underline{0}, \sigma^2 I_N)$.

Green's method assumes that the vector of differences $(E\underline{Y} - H\underline{\beta})$ can be well approximated by a d th order polynomial in x within each of the g groups, $d \geq 1$. An alternative model of the form

$$\underline{Y} = H_1 \underline{v} + \underline{n} + \underline{\xi}$$

is given, where $\underline{\xi}$ is distributed as $N_N(\underline{0}, \sigma^2 I_N)$, H_1 is an $N \times [g(d + 1) + m_1]$ matrix of known constants, $\underline{v}$ is a $[g(d + 1) + m_1] \times 1$ vector of regression coefficients, and $\underline{n}$, as Green states is "a small vector." The first $g(d + 1)$ columns of H_1 correspond to the polynomial terms for the g groups (with $(d + 1)$ terms for each group), the rightmost $m_1 \leq m$ columns in H_1 correspond to terms that are in the

fitted model, but are not represented among the $g(d + 1)$ polynomial terms in the alternative model.

Under the assumption that $\eta = 0$, the presence of lack of fit is tested by using the test statistic:

$$F = \frac{\underline{Y}'[H_1(H_1'H_1)^{-1}H_1' - H(H'H)^{-1}H']\underline{Y}/[g(d + 1) + m_1 - m]}{\underline{Y}'[I - H_1(H_1'H_1)^{-1}H_1']\underline{Y}/[N - g(d + 1) - m_1]} . \quad (2.4)$$

This statistic is of the same form as the F statistic used in the standard multiple regression test of a postulated model against a more general one which includes the postulated model as a special case. Lack of fit is suspected if the calculated F ratio in (2.4) is greater than $F_{\alpha; g(d+1)+m_1-m, N-g(d+1)-m_1}$ where this latter quantity is the upper 100α percentage point of the central F distribution.

Green notes that when there is no lack of fit, the quadratic forms $\underline{Y}'[H_1(H_1'H_1)^{-1}H_1' - H(H'H)^{-1}H']\underline{Y}$ and $\underline{Y}'[I - H_1(H_1'H_1)^{-1}H_1']\underline{Y}$ are distributed independently as $\sigma^2\chi^2$ with $g(d + 1) + m_1 - m$ and $N - g(d + 1) - m_1$ degrees of freedom, respectively. In this case the F ratio in (2.4) possesses a central F distribution. If there is lack of fit on the other hand, then these two quadratic forms are distributed as noncentral chi-squares, multiplied by σ^2 , with respective noncentrality parameters

$\zeta_1 = [H_1 \underline{v} + \underline{\eta}]' [H_1 (H_1' H_1)^{-1} H_1' - H (H' H)^{-1} H'] [H_1 \underline{v} + \underline{\eta}]$
 and $\zeta_2 = \underline{\eta}' [I - H_1 (H_1' H_1)^{-1} H_1'] \underline{\eta}$. Thus the assumption that $\underline{\eta} = \underline{0}$ can affect the power of the test, since if $\underline{\eta} \neq \underline{0}$, the expected value of MSE is greater than σ^2 , where MSE is the quadratic form in the denominator of the F ratio. Hence if $\underline{\eta} \neq \underline{0}$, the probability of calculating a large F value is reduced, and we are less likely to detect lack of fit using an upper tailed rejection region.

Daniel and Wood (1971) suggest another method for lack of fit testing when replicated observations of the response are not available. They make use of "near replicates" to obtain an estimate of σ , which is the standard deviation of the observable responses in the true model. The value of the estimate $\hat{\sigma}$ is compared to the square root of the residual mean square from the analysis of the fitted model. Lack of fit is indicated if the square root of the residual mean square is large compared to the estimate $\hat{\sigma}$. To determine when observations are near replicates so that an estimate of σ can be found, they define the squared distance between any two data points, j and j' , to be measured by

$$D_{jj'}^2 = \sum_{i=1}^K [b_i (x_{ij} - x_{ij'}) / s_y]^2,$$

where x_{ij} and $x_{ij'}$ are the values of the i th independent variable corresponding to the observations y_j and $y_{j'}$, respectively, $i = 1, 2, \dots, K$, and b_i is the ordinary least

squares estimate of the i th regression coefficient. In the denominator, s_y is the square root of the residual mean square for the fitted model.

To obtain an estimate of σ from near replicates, let $\Delta_n d = |d_j - d_{j'}|$, $n = 1, 2, \dots, \binom{N}{2}$, where d_j and $d_{j'}$ are the residuals at points j and j' , respectively, and where there are N data observations in the experiment. Since the expected value of the range for pairs of independent observations from a normal distribution is 1.128σ , a running average of the $\Delta_n d$'s is calculated and their average is multiplied by $.886 = (1/1.128)$ to get a running estimate, s_n , of σ . That is, $s_n = .886 \sum_n \Delta_n d / n$. The closest pair of observations as judged by $D_{jj'}^2$, is used to begin the running estimate, the next closest pair (next "nearest neighbors") is used for $\Delta_2 d$, and the procedure continues until s_n "stabilizes." The stabilized value of s_n is used to estimate σ .

A third method for testing for lack of fit without replication is given by Shillington (1979). The fitted model is of the form

$$\underline{y} = X\underline{\beta} + \underline{\varepsilon} \quad (2.5)$$

where $\underline{y}$ ($N \times 1$), X ($N \times p$), and $\underline{\beta}$ ($p \times 1$) are defined as in (1.2) and $\underline{\varepsilon} \sim N_N(\underline{0}, \sigma^2 I_N)$. The test for lack of fit of the fitted model is a test for whether the true model has the form

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

where $\underline{\delta}$ ($N \times 1$) is a fixed effect quantifying the departure of (2.5) from the true model.

Shillington assumes that the data can be grouped into g cells, with n_j observations in the j th cell, determined in advance. Letting C_j refer to the j th cell, $j = 1, 2, \dots, g$, a vector of cell averages is written $\bar{\underline{Y}}_C$ ($g \times 1$), where the j th element of $\bar{\underline{Y}}_C$ is the average of the observed responses in C_j . The matrix X_C of independent variables associated with $\bar{\underline{Y}}_C$ is the $g \times p$ matrix where the elements in the j th row are $\bar{\underline{x}}'_{.j} = \frac{\sum_{i=1}^{n_j} x_{ij}}{n_j}$, that is, row j of X_C is the row vector $\bar{\underline{x}}'_{.j}$. The matrix X_C is assumed to be of full rank $p < g$. Also within each cell are defined the differences $W_{ij} = Y_{ij} - \bar{Y}_{.j}$, $i \in C_j$, $j = 1, 2, \dots, g$, where $\bar{Y}_{.j}$ is the j th element of $\bar{\underline{Y}}_C$.

The two independent data sets, $\bar{\underline{Y}}_C$ and $\{W_{ij}\}$ with g and $N - g$ degrees of freedom, respectively, are used to find two independent estimates of σ^2 . The first estimate is written as

$$MSE_B = \sum_{j=1}^g n_j (\bar{Y}_{.j} - \bar{\underline{x}}'_{.j} \hat{\underline{\beta}}_B)^2 / (g - p),$$

where $\hat{\underline{\beta}}_B$ is the weighted least squares estimate of $\underline{\beta}$ using the regression of cell means, $\bar{\underline{Y}}_C$, on X_C . The second estimate of σ^2 uses the within cell deviations on cell means, $\{W_{ij}\}$, and is

$$MSE_W = \sum_{j=1}^g \sum_{i=1}^{n_j} (W_{ij} - \hat{W}_{ij})^2 / (N - g - r),$$

where r is the rank of an $N \times p$ matrix with rows equal to $\underline{x}'_{ij} - \bar{\underline{x}}'_{.j}$, $i \in C_j$, $j = 1, 2, \dots, g$. If the matrix of independent variables, corrected for cell means, is of full rank, then $r = p$. Here $\hat{W}_{ij}$ is the estimate of W_{ij} from the regression of cell residuals $\{W_{ij}\}$ on the associated vectors of independent variates, $\underline{x}'_{ij} - \bar{\underline{x}}'_{.j}$.

If the fitted model is the correct model, then MSE_B and MSE_W are independent estimates of σ^2 and the ratio MSE_B/MSE_W is an F statistic with $g - p$ and $N - g - r$ degrees of freedom. When all observations in a cell have the same settings of the independent variables, that is, the observations are truly replicates for all cells, then this F statistic is identical to the F statistic in the usual lack of fit test in which the residual sum of squares is partitioned into lack of fit and pure error sums of squares, as given in Draper and Smith (1981, p. 120).

If the true model is $\underline{y} = \underline{X}\underline{\beta} + \underline{\delta} + \underline{\varepsilon}$, however, and if we let $\underline{X}'\underline{\delta} = \underline{0}$ and $\sigma_{\delta}^2 = \underline{\delta}'\underline{\delta}/N$, then

$$E(MSE_B) = \sigma^2 + \bar{\underline{\delta}}'_B [I - \underline{X}_C(\underline{X}'_C \underline{X}_C)^{-1} \underline{X}'_C] \bar{\underline{\delta}}_B / (g-p)$$

where $\bar{\underline{\delta}}_B$ ($g \times 1$) has j th component equal to $\sum_{i=1}^{n_j} \delta_{ij} / n_j$.

Furthermore, with this latter true model form

$$E(MSE_W) = \sigma^2 + \underline{\delta}'_W (I - \underline{X}_W(\underline{X}'_W \underline{X}_W)^{-1} \underline{X}'_W) \underline{\delta}_W / (N - g - r)$$

where $\hat{\delta}_W$ has the components $\delta_{ij} - \bar{\delta}_{.j}$, $i \in C_j$, $j = 1, 2, \dots, g$. The matrix X_W ($N \times p$) has the rows $\underline{x}_{ij}' - \bar{\underline{x}}_{.j}'$, $i \in C_j$, $j = 1, 2, \dots, g$. The power of the F test, $F = \text{MSE}_B / \text{MSE}_W$, depends on the relative bias of the estimates of σ^2 , that is, the biases in MSE_B and MSE_W .

Shillington states that the power of the F test which makes use of $F = \text{MSE}_B / \text{MSE}_W$ is maximized by forming cells so that the bias of $E(\text{MSE}_W)$ is minimized. This is the same as forming cells so that the within cell variation in $\hat{\delta}$ is minimized. Shillington (1979, p. 141) also states, "Observations with near covariate (independent variable) values might be expected to have similar $\hat{\delta}$ values, since we assume that $\hat{\delta}$ varies in some continuous but unknown fashion with X. This justifies the usual procedure of forming groups by collapsing observations with adjacent covariate values. Indeed, if covariates do not vary within cells we have the usual lack of fit test and maximum power."

By imposing a further structure on the form of $\hat{\delta}$, it is shown that if the F test has an upper tailed rejection region, the power is maximized by selecting the group sizes as $n_j = 2$, $j = 1, 2, \dots, g$. Finally, Shillington suggests that in the presence of more than one independent variable problems in grouping may arise, and in this case it may be wise to perform a different lack of fit test for each parameter. Following this approach, an example is given which suggests testing lack of fit for each of the p independent variables separately may be more powerful than

trying to form groups based on all independent variables at once.

In summary, all the approaches we have discussed for testing for lack of fit when replicate observations of the response are not available at any of the settings of the independent variables make use of grouping the observed response values according to similar values of the independent variables. The observations falling in such groups are referred to as "pseudoreplicates" or "near neighbor observations." These pseudoreplicates are used to estimate the true variance of the observations, σ^2 , but a completely unbiased estimate of σ^2 cannot be attained unless true replicate observations are available. In each case, the power of the lack of fit testing procedure is reduced because an unbiased estimate of σ^2 is not attainable. We now turn to the use of check points for lack of fit testing.

2.4 Testing for Lack of Fit with Check Points

An alternative to the two approaches to lack of fit testing already discussed is the method which makes use of check points. We assume a model of the form $E(\underline{Y}) = \underline{X}\underline{\beta}_1$, as given in (2.1), is fitted in a response surface system, but that the true model is of the form $E(\underline{Y}) = \underline{X}\underline{\beta}_1 + \underline{X}_2\underline{\beta}_2$ as given in (2.2). The parameters, $\underline{\beta}_1$, in the fitted model are estimated by ordinary least squares techniques, making use of the values of the response observed at the design points. After the model is fitted, values of the response are observed at additional points in the experimental region

called "check points." The observed response values at the check points are compared to the values which the fitted model predicts at these same check points. It is important to note that the observed values of the response at the check points are not used in fitting the model initially.

Snee (1977) gives four methods of validating regression models, one of which is the collection of new data to check predictions from a previously fitted model. In a designed experiment these new data take the form of check points. Snee suggests that the inclusion of a small number of check points in any designed experiment is a "worthwhile" procedure.

Scheffé (1958) proposed a test for lack of fit when the {3,2} simplex lattice design is used for fitting a second order canonical polynomial model in three mixture components. It is desired to use the observed value of the response at (1/3, 1/3, 1/3) as a check point blend. The test statistic proposed is the t statistic of the form

$$t = \frac{Y - \hat{Y}}{[\hat{\text{var}}(Y - \hat{Y})]^{1/2}} \quad (2.6)$$

where Y is the observed value of the response at the check point, and $\hat{Y}$ is the value of the response predicted at the same point by the second order model which is fitted by ordinary least squares techniques to the observed response values at the six design points of the {3,2} simplex lattice. The response value observed at the point

$(1/3, 1/3, 1/3)$ is not used in fitting the model. Lack of fit is inferred if the absolute value of the calculated t value in equation (2.6) is larger than the corresponding tabled t value.

In the denominator of the t test of equation (2.6), the variance of the difference $Y - \hat{Y}$ is shown to be

$$\begin{aligned}\text{var}(Y - \hat{Y}) &= \text{var}(Y) + \text{var}(\hat{Y}) \\ &= (44/27r)\sigma^2 ,\end{aligned}$$

when r replicates are taken at each design point. The estimate of the variance of $Y - \hat{Y}$ is $(44/27r)\hat{\sigma}^2$, where $\hat{\sigma}^2$ is calculated from the replicated response values at the design points.

Scheffé (1958) also alludes to a test for lack of fit when several check points are used simultaneously. When there are k check points, the test for lack of fit is an F statistic of the form

$$F = \frac{\underline{d}' V_0^{-1} \underline{d}}{k \hat{\sigma}^2} \quad (2.7)$$

where $\underline{d}' = (Y_1 - \hat{Y}_1, Y_2 - \hat{Y}_2, \dots, Y_k - \hat{Y}_k)$, and $V = \sigma^2 V_0 = \text{var}(\underline{d})$. Formulas are given for the elements of V_0 in the special case when the check points are the design points of the $\{3,2\}$ simplex lattice. Lack of fit is suspected if the calculated value of the F statistic given in (2.7) is larger than the corresponding tabled F value.

Gorman and Hinman (1962) suggest the same t test in equation (2.6) that Scheffé (1958) suggested for a check point taken at $(1/3, 1/3, 1/3)$ to test for lack of fit in a second order polynomial model fitted from a $\{3,2\}$ simplex lattice design. They suggest using $(1/3, 1/3, 1/3)$ as the location of the check point because the observation at this point may later be used to fit the next more complex model, the special cubic, if the second order model is found to be inadequate. They state that in general for the second order polynomial model as well as higher order models, check points should be taken in regions of particular interest, of which there are usually many in any blending study. Further, they suggest that the number of check points depends on individual experimental situations--technical background, precision required, cost of materials and analyses, and probability of requiring a more complex model. However, no specific criterion is given by Gorman and Hinman for selecting the location of the check points.

Gorman and Hinman (1962) indicate that a t test at a check point other than at $(1/3, 1/3, 1/3)$ takes the same form as the statistic of equation (2.6),

$$t = \frac{Y - \hat{Y}}{[\hat{\text{var}}(Y) + \hat{\text{var}}(\hat{Y})]^{1/2}}$$

with the additional condition that if several check points are taken, say for example k points, the method of checking

the fit is to compute the t value at each location and refer these calculated t values to the $100(\alpha/2k)$ percentage point of the central t distribution rather than the $100(\alpha/2)$ percentage point.

Kurotori (1966) gives an example of a mixture experiment where the response is the modulus of elasticity of a rocket fuel, which is a mixture of three components, binder (x_1), oxidizer (x_2), and fuel (x_3). The factor space of feasible mixtures is a subspace inside the two-dimensional simplex or triangle where all three components are present simultaneously. "Pseudocomponents" are defined and in the pseudocomponent system a special cubic model is fitted to data collected at the points of the $q = 3$ simplex centroid design (Figure 4). A check for adequacy of fit is made by using three check points and the response values at the check points are used only for testing the fit of the model and not for fitting the model initially.

The reason for the choice of the particular check point locations by Kurotori is that, as he states, "They are the most remote mixtures from the seven design points." The lack of fit test is an F statistic of the form

$$F = \frac{s^2}{\hat{\sigma}^2} \quad (2.8)$$

where $s^2 = \sum_{i=1}^3 (Y_i - \hat{Y}_i)^2$, for the $i = 1, 2, 3$ check points and $\hat{\sigma}^2$ is an estimate of measurement error from a previous analysis. Kurotori admits that the use of the F statistic

Figure 4. Kurotori's rocket fuel example, x'_1 , x'_2 , and x'_3 represent pseudocomponents.

in Eq. (2.8) for lack of fit testing may be risky because the predicted values at the check points are correlated (correlation of .5), although the observed values are not correlated. Kurotori suggests individual t tests as proposed by Scheffé (1958) might be the preferred procedure.

Snee (1971) repeats Kurotori's rocket fuel example using the same F test for lack of fit as Kurotori and makes the comment that the $\hat{Y}_i$'s at the check points are correlated. In stating that the F test is not an exact test, he nevertheless offers no solution in the form of an exact test.

In summary, only Scheffé refers to an exact F test when several check points are considered simultaneously for testing for possible lack of fit of a model fitted in a mixture space, and his development is limited to the special case where the check points are the design points used to fit the model initially. No criterion is proposed by Scheffé for selecting other locations for the check points.

CHAPTER THREE
AN OPTIMAL CHECK POINT METHOD FOR TESTING
LACK OF FIT IN A MIXTURE MODEL

3.1 Introduction

In Chapter Three we investigate the problem of testing for lack of fit of a linear model fitted in a mixture space. The testing is to be accomplished with the use of check points. We assume that an experimental design is specified, and that the fitted model is of the form

$$E(\underline{Y}) = X\underline{\beta}_1 \quad (3.1)$$

where $\underline{Y}$ is an $N \times 1$ vector of observable response values, X is an $N \times p$ matrix of known constants and rank p , and $\underline{\beta}_1$ is a vector of p unknown regression coefficients. The true model is assumed to be of the form

$$E(\underline{Y}) = X\underline{\beta}_1 + X_2\underline{\beta}_2 \quad (3.2)$$

where X_2 is an $N \times p_2$ matrix of known constants and $\underline{\beta}_2$ is a vector of p_2 unknown regression coefficients. Throughout our development, we will assume that the random vector $\underline{Y}$ has the normal distribution with variance-covariance matrix equal to $\sigma^2 I_N$.

In our investigation we wish to determine the proper testing procedure to follow in deciding whether the fitted model exhibits lack of fit. In order to optimize the lack of fit testing procedure, we will determine the location of the check points so that the power of the test is maximized.

3.2 Testing for Lack of Fit in the Presence of an External Estimate of Experimental Error Variation

3.2.1 The Test Statistic

We wish to test the performance or fit of a fitted model in a mixture space when the true model possibly contains terms in addition to those in the fitted model. The fit of the model is to be tested by a test which makes use of the response values observed at certain locations called "check points" in the experimental region, by comparing them to the values which the fitted model predicts at the same check points. The observed values at the check points are not used for estimating the coefficients in the fitted model and are assumed to represent the values of the true surface at the check points.

Let us define the vector of differences

$$\begin{aligned}\underline{d} &= (\underline{y}^* - \hat{\underline{y}}^*) \\ &= (y_1^* - \hat{y}_1^*, y_2^* - \hat{y}_2^*, \dots, y_k^* - \hat{y}_k^*)'\end{aligned}$$

where y_i^* , $i = 1, 2, \dots, k$ are observed response values at k check points and $\hat{y}_i^*$, $i = 1, 2, \dots, k$ are response values predicted at the k check points by the fitted model,

$\hat{y}_i^* = \underline{x}_i^{*'} \underline{b}_1$, where $\underline{b}_1$ is the ordinary least squares estimator of $\underline{\beta}_1$, and where $\underline{x}_i^{*'}$ is the i th row of X^* , the $k \times p$ matrix whose columns are of the same form as the columns of X but with its rows evaluated at the k check points. Note that if $\underline{\beta}_2 = \underline{0}$, then $E(\underline{d}) = \underline{0}$ and if $\underline{\beta}_2 \neq \underline{0}$, then $E(\underline{d}) = (X_2^* - X^*(X'X)^{-1}X'X_2)\underline{\beta}_2$. Let V represent the variance-covariance matrix of the random vector $\underline{d}$. Then $V = \sigma^2 V_0$ where

$$V_0 = I_k + X^*(X'X)^{-1}X^{*'}.$$

and where I_k is the identity matrix of order $k \times k$.

We assume that an unbiased estimate of σ^2 is available and we denote this estimate by $\hat{\sigma}_{\text{ext}}^2$, where the subscript ext stands for external, and $\hat{\sigma}_{\text{ext}}^2$ is independent of the model being fitted. The test statistic for the hypothesis of zero lack of fit $H_0: E(\underline{d}) = \underline{0}$ is

$$F = \frac{\underline{d}' V_0^{-1} \underline{d} / k}{\hat{\sigma}_{\text{ext}}^2} \quad (3.3)$$

(see Scheffé', 1958, p.358). It will be shown later in this section that the F ratio in Eq. (3.3) possesses either a central F distribution or a noncentral F distribution, depending upon whether the true model is represented by Eq. (3.1) or Eq. (3.2).

The variance estimate $\hat{\sigma}_{\text{ext}}^2$ that appears in equation (3.3) is ordinarily generated from replicated observations

at some of the design points in the experiment. We assume that $\hat{\sigma}_{\text{ext}}^2$ is a constant multiple of a central chi-square random variable with ν degrees of freedom. This is written as

$$\begin{aligned}\hat{\sigma}_{\text{ext}}^2 &= \text{SSE}_{\text{pure}}/\nu \\ &= (\sigma^2/\nu)(\text{SSE}_{\text{pure}}/\sigma^2)\end{aligned}$$

where $\text{SSE}_{\text{pure}}/\sigma^2 \sim \chi_{\nu}^2$. Note that SSE_{pure} denotes the portion of the residual sum of squares due to replication variation from the fitted model. The residual sum of squares from the fitted model may be partitioned into SSE_{pure} and SS_{LOF} only if replicated observations are collected at one or more design points. For the case where replicate observations are collected at all of the design points

$$\text{SSE}_{\text{pure}} = \sum_{i=1}^n \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{i.})^2,$$

where n is the number of distinct design points, $n_i > 2$ is the number of replicates at the i th design point, Y_{ij} is the j th observation at the i th design point, and $\bar{Y}_{i.}$ is the average of the n_i observations at the i th design point. Here SSE_{pure} has $\nu = \sum_{i=1}^n (n_i - 1)$ degrees of freedom.

When the fitted model and the true model are of the same form as defined by Eq. (3.1), the quantity $\underline{d}'V_0^{-1}\underline{d}/\sigma^2$

possesses a central chi-square distribution (Searle, 1971, p.57, Theorem 2). However, when the true model is of the form specified by Eq. (3.2), $\underline{d}'V_0^{-1}\underline{d}/\sigma^2$ possesses a noncentral chi-square distribution. Thus when the true model is of the form in Eq. (3.1),

$$\underline{d}'V_0^{-1}\underline{d}/\sigma^2 \sim \chi_k^2,$$

but when the true model is of the form in Eq. (3.2),

$$\underline{d}'V_0^{-1}\underline{d}/\sigma^2 \sim \chi_{k,\lambda_1}^2,$$

where in the second case the noncentrality parameter λ_1 has the form

$$\begin{aligned}\lambda_1 &= E(\underline{d})'V_0^{-1}E(\underline{d})/2\sigma^2 \\ &= \underline{\beta}_2'(X_2^* - X^*A)'V_0^{-1}(X_2^* - X^*A)\underline{\beta}_2/2\sigma^2.\end{aligned}$$

The matrix $A = (X'X)^{-1}X'X_2$ is called the alias matrix and is of order $p \times p_2$. In λ_1 , the matrix X_2^* is of order $k \times p_2$ and has the same relationship to X_2 as X^* has to X .

Since $SSE_{\text{pure}}/\sigma^2$ is statistically independent of $\underline{d}'V_0^{-1}\underline{d}/\sigma^2$, then under model (3.1) the test statistic

$$F = \frac{\underline{d}'V_0^{-1}\underline{d}/k\sigma^2}{SSE_{\text{pure}}/v\sigma^2}$$

$$= \frac{\underline{d}'V_0^{-1}\underline{d}/k}{\hat{\sigma}_{\text{ext}}^2}$$

will have a central F distribution. When the true model contains terms in addition to those in the fitted model then F will have a noncentral F distribution. We write these two cases as

$$F \sim F_{k,v}$$

under model (3.1), and

$$F \sim F'_{k,v;\lambda_1}$$

under model (3.2), where the noncentrality parameter is

$$\lambda_1 = \underline{\beta}_2'(X_2^* - X^*A)'V_0^{-1}(X_2^* - X^*A)\underline{\beta}_2/2\sigma^2.$$

3.2.2 The Testing Procedure and an Expression for the Power of the Test

Given that the form of the fitted model is defined as Eq. (3.1), the expected value of the numerator of the F statistic in Eq. (3.3) will depend on the form of the true model. For the case where the true model is expressed as

Eq. (3.2),

$$\begin{aligned}
 E(\text{numerator}) &= E(\underline{d}' V_0^{-1} \underline{d} / k) \\
 &= (\sigma^2 / k) E \chi_{k, \lambda_1}^2 \\
 &= (\sigma^2 / k) (k + 2\lambda_1) \\
 &= \sigma^2 + 2\sigma^2 \lambda_1 / k \\
 &= \sigma^2 + \underline{\beta}_2' A_1 \underline{\beta}_2 / k, \quad (3.4)
 \end{aligned}$$

where $A_1 = (X_2^* - X^*A)' V_0^{-1} (X_2^* - X^*A)$. However, when the true model is Eq. (3.1), $\underline{\beta}_2 = \underline{0}$ and in this case $\lambda_1 = 0$ so that $E(\text{numerator}) = \sigma^2$. Also $\hat{\sigma}_{\text{ext}}^2$ is an unbiased estimator of σ^2 and

$$E(\hat{\sigma}_{\text{ext}}^2) = \sigma^2. \quad (3.5)$$

Therefore the ratio $E(\text{numerator})/E(\text{denominator})$ where the denominator is $\hat{\sigma}_{\text{ext}}^2$, will equal unity under model (3.1), that is, when there is no lack of fit. Under model (3.2), the ratio will be greater than or equal to unity so lack of fit should be suspected if the calculated F ratio in equation (3.3) is large. We can thus use an upper tailed rejection region to reject the hypothesis of zero lack of fit. The power of the test is

$$P\{F'_{k,v;\lambda_1} > F_{\alpha;k,v}\}$$

where $F_{\alpha;k,v}$ is the upper 100α percentage point of the central F distribution with k numerator degrees of freedom and v denominator degrees of freedom.

It is worth noting that from Eq. (3.4) and Eq. (3.5) testing the hypothesis that $\beta_2 = 0$ is equivalent to testing the hypothesis that $\lambda_1 = 0$, assuming A_1 is positive definite. Thus testing a null hypothesis of zero lack of fit using the proposed testing procedure involving the F ratio in (3.3) may be expressed as a test of the hypotheses:

$$H_0: \lambda_1 = 0$$

$$H_a: \lambda_1 > 0.$$

3.2.3 A Method for Locating Optimal Check Points

Once a design for fitting model (3.1) in a mixture space is chosen and the number of simultaneous check points is decided on, say $k > 1$, the next step is to determine where in the mixture space we should place the k check points so as to maximize the power of the test for lack of fit. The location of the check points is to be made independently of the value of β_2 .

The power of the upper tailed F test for lack of fit is an increasing function of λ_1 (see Appendix 1 for proof, with $\lambda_2 = 0$). Therefore, to maximize the power of the test we maximize the value of λ_1 defined as

$$\lambda_1 = \underline{\beta}_2' A_1 \underline{\beta}_2 / 2\sigma^2$$

where $A_1 = (X_2^* - X^*A)' V_0^{-1} (X_2^* - X^*A)$, by properly selecting the k check points whose coordinates are defined in X^* . To maximize the value of λ_1 , we shall concentrate on the matrix A_1 .

The matrix A_1 is a square matrix of order $p_2 \times p_2$ and is a scalar quantity when $p_2 = 1$. By maximizing the scalar quantity A_1 with respect to the k check points, the power is maximized no matter what the value of $\underline{\beta}_2$. Maximizing the scalar A_1 can be accomplished by using The Controlled Random Search Procedure given by Price (1977). This procedure is described in Appendix 2. As a computational aid, A_1 can be expressed as

$$A_1 = \frac{|V_0 + (X_2^* - X^*A)(X_2^* - X^*A)'|}{|V_0|} - 1 \quad (3.6)$$

when $p_2 = 1$, where the symbol $|B|$ denotes the determinant of the square matrix B . Thus the computations reduce to evaluating two determinants rather than inverting V_0 (see Scheffé, 1959, Appendix V, p.417).

When $p_2 > 1$ and A_1 is no longer a scalar, maximizing λ_1 (and thus maximizing the power of the test) cannot be accomplished without specifying β_2 . In this case we make use of a lower bound for λ_1 (Graybill, 1969, p.330, Theorem 12.2.14(9)) defined as

$$\mu_{\min} \beta_2' \beta_2 / 2\sigma^2 \leq \lambda_1$$

(where $\mu_{\min}$ is the smallest eigenvalue of A_1) to be used in place of λ_1 . Hence an approximate solution to the maximization of λ_1 will be achieved by finding the k simultaneous check points (using Price's procedure) that maximize $\mu_{\min}$, the smallest eigenvalue of A_1 . In other words when $p_2 > 1$, and in order to avoid specifying β_2 , we seek to maximize a lower bound value for λ_1 . This maximization does not depend on the value of β_2 .

There are cases where the matrix A_1 is of less than full rank (less than rank p_2) or equivalently where the matrix A_1 is positive semi-definite so that $\mu_{\min}$ will be equal to zero no matter which check points are selected. One such case occurs when $k < p_2$ (when the number of check points is less than the number of parameters in the true model which are not in the fitted model) since when $k < p_2$

$$\begin{aligned} \text{rank}(A_1) &= \text{rank}[V_0^{-1/2}(X_2^* - X^*A)] \\ &= \text{rank}(X_2^* - X^*A), \end{aligned}$$

and so $\text{rank}(A_1) < \min(k, p_2)$ because the matrix $(X_2^* - X^*A)$ is of order $k \times p_2$. Therefore when $k < p_2$, the rank of A_1 is at most k so that A_1 is of less than full rank. Since $\mu_{\min}$ must be equal to zero when A_1 is positive semi-definite, an alternative method to that of maximizing $\mu_{\min}$ to select optimal check points must be found when A_1 is positive semi-definite in order to produce a positive lower bound for λ_1 . In this pursuit, let us write λ_1 as

$$\begin{aligned}
 \lambda_1 &= \underline{g}_2' A_1 \underline{g}_2 / 2\sigma^2 \\
 &= \underline{g}_2' P \Lambda P' \underline{g}_2 / 2\sigma^2 \\
 &= \underline{g}_2' [P_1 : P_2] \text{diag}[\Lambda_1, \Lambda_2=0] [P_1 : P_2]' \underline{g}_2 / 2\sigma^2 \\
 &= \underline{g}_2' P_1 \Lambda_1 P_1' \underline{g}_2 / 2\sigma^2
 \end{aligned}$$

where Λ is a diagonal matrix with elements equal to the eigenvalues of A_1 , P is an orthogonal matrix whose columns are orthonormal eigenvectors of A_1 , Λ_1 and P_1 correspond to the positive eigenvalues of A_1 , while $\Lambda_2 = 0$ and P_2 correspond to zero eigenvalues of A_1 . Then by Theorem 12.2.14(9) in Graybill (1969) we can write

$$\mu_{\min}^+ \underline{z}' \underline{z} / 2\sigma^2 < \lambda_1 \quad (3.7)$$

where $\mu_{\min}^+$ is the smallest positive eigenvalue of A_1 , and

$\underline{z} = \underline{P}_1' \underline{\beta}_2$. Thus by Eq. (3.7), an approach to maximizing a positive lower bound for λ_1 when A_1 is positive semi-definite is to select check points that maximize the smallest positive eigenvalue of A_1 . It must be noted, however, that this method can only be used when $\underline{\beta}_2 \in \cap C(P_1)$, where $C(P_1)$ denotes the column space of P_1 and $\cap C(P_1)$ denotes the intersection of all such spaces which can be obtained at all possible check points locations. This is because, in general, $\underline{z}'\underline{z}$ in (3.7) depends on the location of the check points through its dependency on P_1 . If, however, $\underline{\beta}_2 \in \cap C(P_1)$, then $\underline{z}'\underline{z} = \underline{\beta}_2' P_1 P_1' \underline{\beta}_2 = \underline{\beta}_2' P P' \underline{\beta}_2 = \underline{\beta}_2' \underline{\beta}_2$, since $\underline{\beta}_2' P_2 = 0$. It follows that when $\underline{\beta}_2 \in \cap C(P_1)$, $\mu_{\min}^+ \underline{z}'\underline{z} / 2\sigma^2 = \mu_{\min}^+ \underline{\beta}_2' \underline{\beta}_2 / 2\sigma^2$ and only $\mu_{\min}^+$ depends on the location of the check points.

3.3 Testing for Lack of Fit When MSE Is Used to Estimate Experimental Error Variation

3.3.1 The Test Statistic

In this section we shall show that when an external estimate of σ^2 is not available and the residual mean square (MSE) from the fitted model of the form (3.1) must be used as an estimate of σ^2 , the test statistic

$$F = \frac{\underline{d}' V_0^{-1} \underline{d} / k}{\text{MSE}} \quad (3.8)$$

possesses a central F distribution when the true model is

Eq. (3.1), but possesses a doubly noncentral F distribution when the true model is Eq. (3.2).

In the initial section of this chapter, the quantity $\underline{d}'\underline{V}_0^{-1}\underline{d}/\sigma^2$ was said to possess a central chi-square distribution or to possess a noncentral chi-square distribution, depending on whether the true model was specified by Eq. (3.1) or Eq. (3.2). Now, the residual sum of squares from the fitted model is defined as

$$\begin{aligned} \text{SSE} &= \sum_{i=1}^N (Y_i - \hat{Y}_i)^2 \\ &= \underline{Y}'(\underline{I}_N - \underline{X}(\underline{X}'\underline{X})^{-1}\underline{X}')\underline{Y} \end{aligned}$$

and it is easy to show (Searle, 1971, p.57, Theorem 2) that SSE/σ^2 possesses a central chi-square distribution if the true model is Eq. (3.1), but under model (3.2), SSE/σ^2 possesses a noncentral chi-square distribution. This is expressed as

$$\text{SSE}/\sigma^2 \sim \chi_{N-p}^2$$

under model (3.1), and

$$\text{SSE}/\sigma^2 \sim \chi_{N-p, \lambda_2}^2$$

under model (3.2), where the noncentrality parameter λ_2 is

$$\lambda_2 = \underline{\beta}_2'(X_2 - XA)'(X_2 - XA)\underline{\beta}_2/2\sigma^2.$$

The distributional form of the test statistic in Eq. (3.8) is derived by knowing that the quantities $\underline{d}'V_0^{-1}\underline{d}/\sigma^2$ and SSE/σ^2 are statistically independent (see Appendix 3), so that

$$\begin{aligned} F &= \frac{\underline{d}'V_0^{-1}\underline{d}/k\sigma^2}{MSE/\sigma^2} \\ &= \frac{\underline{d}'V_0^{-1}\underline{d}/k}{MSE} \end{aligned}$$

is distributed as a central F when the true model is Eq. (3.1), but when the true model is Eq. (3.2) the F ratio is a doubly noncentral F, that is, under model (3.1),

$$F \sim F_{k, N-p}$$

and under model (3.2),

$$F \sim F_{k, N-p; \lambda_1, \lambda_2}''.$$

3.3.2 The Rejection Region and its Relation to the Power of the Test

In Appendix 1 it is shown that if k , $N-p$, and λ_2 are fixed, then the power of the F test using the ratio (3.8) is

a function of the location of the rejection region (upper tailed or lower tailed) of the test. The power increases with increasing values of the numerator noncentrality parameter, λ_1 , when the test is an upper tailed test. The power decreases with increasing values of λ_1 when the test is a lower tailed test. This means that to study ways of increasing the power of the test, we have to determine whether the test is an upper tailed test or a lower tailed test. Similarly, for fixed values of k , $N-p$, and λ_1 , the power of the F test is a decreasing function of λ_2 for an upper tailed test, and is an increasing function of λ_2 when the F test is a lower tailed test (Scheffé, 1959, p. 136-137).

To decide if the test is an upper tailed test or a lower tailed test, we recall from Section 3.2.2 that if the true model is Eq. (3.1) then the expected value of the numerator of the F statistic in (3.8) can be written as

$$E(\text{numerator}) = \sigma^2,$$

and if the true model is Eq. (3.2),

$$E(\text{numerator}) = \sigma^2 + 2\sigma^2 \lambda_1 / k \quad (3.9)$$

$$= \sigma^2 + \beta_2' A_1 \beta_2 / k$$

where the $p_2 \times p_2$ matrix A_1 is $A_1 = (X_2^* - X^*A)'V_0^{-1}(X_2^* - X^*A)$. Similarly, it can be shown that if the true model is Eq. (3.1), the expected value of the denominator of the F statistic in (3.8), where the denominator equals MSE, is

$$E(\text{denominator}) = E(\text{MSE})$$

$$= \sigma^2,$$

but if the true model is Eq.(3.2),

$$E(\text{denominator}) = E(\text{MSE})$$

$$= [\sigma^2/(N - p)] E\chi_{N-p, \lambda_2}^2$$

$$= [\sigma^2/(N - p)] [N - p + 2\lambda_2]$$

$$= \sigma^2 + 2\sigma^2\lambda_2/(N - p) \quad (3.10)$$

$$= \sigma^2 + \beta_2' A_2 \beta_2 / (N - p)$$

where the $p_2 \times p_2$ matrix A_2 is $A_2 = (X_2 - XA)'(X_2 - XA)$. Thus the ratio $E(\text{numerator})/E(\text{denominator})$ will equal unity if the true model is Eq. (3.1), but if the true model is Eq. (3.2), the ratio is greater than unity if $\beta_2' A_1 \beta_2 / k > \beta_2' A_2 \beta_2 / (N - p)$. In this latter case we reject the null

hypothesis of zero lack of fit if the calculated value of the F ratio in (3.8) is large. An upper tailed rejection region seems reasonable for this test. When the true model is Eq. (3.2), and if $\beta_2' A_1 \beta_2 / k < \beta_2' A_2 \beta_2 / (N - p)$, then a lower tailed rejection region is preferred.

3.3.3. A Method for Locating Optimal Check Points

Given a design for fitting a model of the form in Eq. (3.1) in a mixture space (note that fixing the design fixes A_2 and $(N - p)$), and given the number of simultaneous check points desired, $k > 1$, we now wish to determine where in the mixture space the k check points should be located so as to maximize the power of the F test for lack of fit, where the test statistic is given in Eq. (3.8). We also wish to position the optimal check points in a manner that is independent of the values of the elements in β_2 .

The case of an upper tailed test. To help us find k simultaneous check points that maximize the power of an upper tailed test, we shall make use of the fact that the power is an increasing function of λ_1 . Therefore to maximize the power of the upper tailed F test, we shall seek the locations of the k check points that maximize λ_1 .

As in the case considered in Section 3.2.3, where the test statistic had a noncentral F distribution, if the number of extra terms in the true model is $p_2 = 1$, then maximizing λ_1 is equivalent to maximizing the scalar A_1 . However, as before, if $p_2 > 1$, then the $p_2 \times p_2$ matrix A_1 is not a scalar and we will have to approximate the

maximization of λ_1 by maximizing a lower bound for λ_1 . This is done by finding the maximum value of $\mu_{\min}$, the smallest eigenvalue of A_1 , since

$$\mu_{\min} \beta_2' \beta_2 / 2\sigma^2 \leq \lambda_1.$$

When the number of check points is less than the order of the square matrix A_1 , that is, $k < p_2$, then $\text{rank}(A_1) < \min(k, p_2)$, and A_1 will have $\mu_{\min} = 0$. For this case, we again try to maximize the smallest positive eigenvalue of A_1 which we denote by $\mu_{\min}^+$, while remembering from Section 3.2.3 that this technique is limited to situations where $\beta_2 \in nC(P_1)$.

The case of a lower tailed test. To find k check points to maximize the power of a lower tailed test, we make use of the fact that the power of the lower tailed F test increases as λ_1 decreases. Then if $p_2 = 1$ and A_1 is a scalar quantity, λ_1 can be minimized with respect to the k check points by finding the check points that minimize A_1 . If $p_2 > 1$, then by Theorem 12.2.14(9) in Graybill (1969), we see that an upper bound for λ_1 is

$$\lambda_1 \leq \mu_{\max} \beta_2' \beta_2 / 2\sigma^2, \quad (3.11)$$

where $\mu_{\max}$ is the largest eigenvalue of A_1 . An approximate solution to minimizing λ_1 in (3.11) can be achieved by minimizing $\mu_{\max}$. It is not necessary to treat the case

of $k < p_2$ separately here, although λ_1 will equal zero if β_2 is in the column space of P_2 , where P_2 is the matrix whose columns are orthonormalized eigenvectors corresponding to the zero eigenvalues of the matrix A_1 .

3.3.4 Determining Whether the Test Is Upper Tailed or Lower Tailed

The procedures outlined in Section 3.3.3 produce a set of k check points that simultaneously maximize the power of the upper tailed test as well as a second set of k check points that simultaneously maximize the power of the lower tailed test. The check points that are selected maximize the power, given A_2 , k , and $N - p$ without specification of β_2 , except that when A_1 is positive semi-definite we require that $\beta_2 \in {}^n C(P_1)$.

It is now necessary to decide which of our two candidates will be used for a lack of fit test. To choose between the upper tailed test and the lower tailed test, let us consider the quantity

$$R = [A_1/k] - [A_2/(N - p)].$$

If R is positive definite when the true model is Eq. (3.2), then no matter what the value of β_2 is, the ratio $E(\text{numerator})/E(\text{denominator})$ will be greater than unity, implying an upper tailed test is to be used. Similarly, if R is negative definite, then a lower tailed test should be used. Finally, if R is not definite, then neither an upper nor a lower tailed test is implicated and further

investigation is necessary. The criterion of $R = [A_1/k] - [A_2/(N - p)]$ may yield any of the four following cases.

Case 1. If $R = [A_1/k] - [A_2/(N - p)]$ is positive definite when A_1 is generated by the k optimal upper tailed test check points, and R is not negative definite when A_1 is generated by the k optimal lower tailed test check points, then we recommend that the check points be used that yield the optimal upper tailed test with an upper tailed rejection region.

For Case 1 it is necessary for A_1 to be positive definite (see Appendix 4). Since A_1 is a square matrix of order $p_2 \times p_2$ with $\text{rank}(A_1) < \min(k, p_2)$, then A_1 can be positive definite only if $k > p_2$. Thus, there must be at least p_2 check points for Case 1 to hold, where p_2 is the number of terms in the model of Eq. (3.2) that are not in the model of Eq. (3.1).

From inspection of equations (3.9) and (3.10), it is apparent that the testing for lack of fit in Case 1 is equivalent to testing the hypothesis

$$H_0: \frac{\lambda_1}{k} - \frac{\lambda_2}{N - p} = 0 \quad (3.12)$$

against the alternative

$$H_a: \frac{\lambda_1}{k} - \frac{\lambda_2}{N - p} > 0$$

since $R = [A_1/k] - [A_2/(N-p)]$ is positive definite when the true model is Eq. (3.2). In Appendix 5(a) it is shown that under Case 1, the hypothesis given by (3.12) is equivalent to the hypothesis

$$H_0: \lambda_1 = \lambda_2 = 0.$$

Case 2. In Case 2 we assume that $R = [A_1/k] - [A_2/(N - p)]$ is not positive definite for the k optimal upper tailed test check points, but that R is negative definite for the k optimal lower tailed test check points. Here we recommend that the lower tailed test check points be used with a lower tailed rejection region.

It is necessary for A_2 to be positive definite for Case 2 to occur (see Appendix 4). However, A_1 need not be positive definite, and so k need not be greater than p_2 . In Case 2 then, it is possible that lack of fit may be tested with only one check point.

By inspection of equations (3.9) and (3.10), a hypothesis of no lack of fit is equivalent to

$$H_0: \frac{\lambda_1}{k} - \frac{\lambda_2}{N - p} = 0 \quad (3.13)$$

while the alternative hypothesis that lack of fit is present is equivalent to

$$H_a: \frac{\lambda_1}{k} - \frac{\lambda_2}{N - p} < 0$$

since $R = [A_1/k] - [A_2/(N - p)]$ is negative definite. In Appendix 5(b) it is shown that the hypothesis given by (3.13) is equivalent to the hypothesis

$$H_0: \lambda_1 = \lambda_2 = 0.$$

Case 3. We assume R is positive definite for the k optimal upper tailed test check points, and R is negative definite for the k optimal lower tailed test check points. Hence either an upper or lower tailed test may be considered as a possible test for lack of fit. If the quantity $\beta_{-2}^* \beta_{-2} / \sigma^2$ can be specified, then the minimum power for both the optimal upper and optimal lower tailed tests can be approximated, and the test with the greater minimum power is recommended. In Appendix 4 it is shown that Case 3 can occur only when A_1 is positive definite for the upper tailed test. Thus Case 3 can only occur when there are at least p_2 check points.

The minimum power of the upper tailed test may be found by calculating

$$P \{ F_{k, N-p; \lambda_{1L}, \lambda_{2U}} > F_{\alpha; k, N-p} \}, \quad (3.14)$$

where $F_{\alpha; k, N-p}$ is the upper 100α percentage point of the central F distribution,

$$\lambda_{1L} = \mu_{\min} \beta_2' \beta_2 / 2\sigma^2,$$

and

$$\lambda_{2U} = \delta_{\max} \beta_2' \beta_2 / 2\sigma^2,$$

where $\mu_{\min}$ is the smallest eigenvalue of A_1 and $\delta_{\max}$ is the largest eigenvalue of A_2 . Formula (3.14) yields a conservative lower bound for the power of the optimal upper tailed test. Note that A_1 is generated using the optimal upper tailed test check points. The cumulative distribution function of F'' can be approximated by multiplying the cumulative probabilities of the central F distribution by a constant (Johnson and Kotz, 1970, p.197). This approximation is described in Appendix 6. Other approximations for F'' (such as the Edgeworth series approximation suggested by Mudholkar, Chaubey, and Lin, 1976) exist which are generally more accurate, but we chose to use the approximation given in Johnson and Kotz (1970, p.197) due to its simplicity. Additionally, the approximation of Mudholkar, Chaubey, and Lin (1976) produced negative probabilities when only one degree of freedom was available in either the numerator or denominator of F'' . This problem was avoided by using the approximation given by Johnson and Kotz (1970).

The minimum power of the optimal lower tailed test can be approximated similarly (if $\beta_2' \beta_2 / \sigma^2$ is specified) by

calculating

$$P \{ F_{k, N-p; \lambda_{1U}, \lambda_{2L}}'' < F_{(1-\alpha); k, N-p} \}$$

where

$$\lambda_{1U} = \mu_{\max} \beta_2' \beta_2 / 2\sigma^2,$$

and

$$\lambda_{2L} = \delta_{\min} \beta_2' \beta_2 / 2\sigma^2,$$

with $\mu_{\max}$ equal to the largest eigenvalue of A_1 and $\delta_{\min}$ equal to the smallest eigenvalue of A_2 . Note that A_1 is generated by using the optimal lower tailed test check points. For the lower tailed test, A_1 may be positive semi-definite, and if β_2 is in the column space of P_2 then $\lambda_1 = 0$.

In Case 3, the upper tailed test is a test of

$$H_0: \lambda_1 = \lambda_2 = 0$$

$$H_a: \frac{\lambda_1}{k} - \frac{\lambda_2}{N-p} > 0,$$

while the lower tailed test is a test of

$$H_0: \lambda_1 = \lambda_2 = 0$$

$$H_a: \frac{\lambda_1}{k} - \frac{\lambda_2}{N-p} < 0.$$

Case 4. In Case 4 we assume that $R = [A_1/k] - [A_2/(N-p)]$ is not positive definite for the k optimal upper tailed test check points and R is not negative definite for the k optimal lower tailed test check points. Here it is useful to write the difference between the expected value of the numerator and the expected value of the denominator of the F ratio in (3.8) as

$$\begin{aligned} \beta_2' [A_1/k - A_2/(N-p)] \beta_2 &= \beta_2' S \Omega S' \beta_2 \\ &= \beta_2' [S_1 : S_2 : S_3] \text{diag}[\Omega_1, \Omega_2=0, \Omega_3] [S_1 : S_2 : S_3]' \beta_2 \\ &= \beta_2' S_1 \Omega_1 S_1' \beta_2 + \beta_2' S_3 \Omega_3 S_3' \beta_2 \end{aligned}$$

where $\Omega = \text{diag}(\Omega_1, \Omega_2, \Omega_3)$ is a diagonal matrix consisting of the eigenvalues of R , Ω_1 is a diagonal matrix of the positive eigenvalues of R , Ω_2 is a diagonal matrix of the zero eigenvalues of R , and Ω_3 is a diagonal matrix of the negative eigenvalues of R . The orthogonal matrix S can be expressed as $S = [S_1 : S_2 : S_3]$, where the matrices S_1 , S_2 , and S_3 have columns which are orthonormalized eigenvectors corresponding to Ω_1 , Ω_2 , and Ω_3 , respectively.

In Case 4, neither the optimal upper tailed test nor the optimal lower tailed test is applicable for all values of β_2 . For completeness, we note that Case 4 actually consists of nine subcases, where R may be positive semi-definite, negative semi-definite, or indefinite for either the optimal upper tailed test or lower tailed test check points. These subcases are listed in Table 2.

Table 2. Nine Subcases of Case 4.

<u>Subcase</u>	<u>R--Upper Tailed Test</u>	<u>R--Lower Tailed Test</u>
1	PSD	PSD
2	PSD	NSD
3	PSD	I
4	NSD	PSD
5	NSD	NSD
6	NSD	I
7	I	PSD
8	I	NSD
9	I	I

PSD = positive semi-definite, NSD = negative semi-definite, I = indefinite.

If β_2 lies in the column space of S_2 , then $\beta_2' [A_1/k - A_2/(N - p)] \beta_2$ is zero, and therefore lack of fit is not testable with either an upper or lower tailed test. A sufficient condition for the test for lack of fit to be upper tailed in Case 4 is that β_2 be in the column space of $[S_1:S_2]$, but not entirely in the column space of S_2 . In this case

$$\begin{aligned}
\beta_2' [A_1/k - A_2/(N - p)] \beta_2 &= \beta_2' S_1 \Omega_1 S_1' \beta_2 + \beta_2' S_3 \Omega_3 S_3' \beta_2 \\
&= \beta_2' S_1 \Omega_1 S_1' \beta_2 + 0 \\
&= \beta_2' S_1 \Omega_1 S_1' \beta_2,
\end{aligned}$$

and $\beta_2' [A_1/k - A_2/(N - p)] \beta_2$ will be greater than zero, indicating an upper tailed test. Similarly, a sufficient condition for the test for lack of fit to be lower tailed is that β_2 be in the column space of $[S_2:S_3]$, but not entirely in the column space of S_2 . Then

$$\begin{aligned}
\beta_2' [A_1/k - A_2/(N - p)] \beta_2 &= 0 + \beta_2' S_3 \Omega_3 S_3' \beta_2 \\
&= \beta_2' S_3 \Omega_3 S_3' \beta_2
\end{aligned}$$

which makes $\beta_2' [A_1/k - A_2/(N - p)] \beta_2$ less than zero, indicating a lower tailed test.

To determine whether β_2 is in the column space of $[S_1:S_2]$, let us define the augmented matrix $Q_1 = [\beta_2:S_1:S_2]$. If $Q_1'Q_1$ has a zero eigenvalue, then β_2 is in the column space of $[S_1:S_2]$. Similarly, if we define $Q_2 = [\beta_2:S_2]$ and $Q_3 = [\beta_2:S_2:S_3]$, then β_2 is in the column space of S_2 if $Q_2'Q_2$ has a zero eigenvalue, and β_2 is in the column space of $[S_2:S_3]$ if $Q_3'Q_3$ has a zero eigenvalue.

Given that we are in a particular subcase of the nine subcases described in Table 2, we recommend that lack of fit be tested with the upper tailed test check points if it is determined that β_2 is such that $\beta_2'[A_1/k - A_2/(N - p)]\beta_2$ is positive when A_1 is generated from the upper tailed test check points. Likewise, for the same given subcase, if the value of β_2 of interest is determined to produce a negative value for $\beta_2'[A_1/k - A_2/(N - p)]\beta_2$ when A_1 is generated from the lower tailed test check points, then we recommend that lack of fit be tested with the lower tailed test.

We see then that Case 4 is an undesirable situation in practice, since, in order to test for lack of fit, we must assume a priori that any lack of fit is due to a nonzero value of β_2 that produces an upper tailed or lower tailed rejection region. However, it would seem rare that such knowledge would be available.

3.4 Examples

We now present several examples to illustrate the technique for locating optimal check points to be used in testing for lack of fit in a mixture model.

3.4.1 Theoretical Examples

Example 1. In this example a second order canonical polynomial model is fitted in three mixture components using the $\{3,2\}$ simplex lattice design, which is presented in Figure 1 of Chapter 1. The true model is assumed to be the special cubic model containing the term $\beta_{123}x_1x_2x_3$ in addition to the six terms of the fitted model. The expected

values of the response at the six design points are assumed to be represented by the fitted model in the form

$$E(\underline{Y}) = X\underline{\beta}_1,$$

but with the true model the expectations are written as

$$E(\underline{Y}) = X\underline{\beta}_1 + X_2\underline{\beta}_2,$$

where X is a 6×6 matrix with rows that define the coordinates of the six design points and columns that correspond to the six terms in the fitted model ($x_1, x_2, x_3, x_1x_2, x_1x_3, x_2x_3$), $\underline{\beta}_1$ is the 6×1 vector of regression coefficients ($\beta_1, \beta_2, \beta_3, \beta_{12}, \beta_{13}, \beta_{23}$), X_2 is a 6×1 column vector containing the values of the term $x_1x_2x_3$ at the design points, and $\underline{\beta}_2$ is the single regression coefficient β_{123} .

The $\{3,2\}$ simplex lattice design consists of only six design points, and since six parameters are estimated in the second order fitted model, there are no degrees of freedom remaining for obtaining an estimate of the experimental error, σ^2 . We assume therefore that an external estimate of σ^2 is available, $\hat{\sigma}_{\text{ext}}^2$, which will be used in the denominator of the lack of fit F statistic given in Eq. (3.3).

Since there is one term in the true model in addition to those in the fitted model, that is $p_2 = 1$, we know that in order to locate k simultaneous check points that maximize

the power of the test for lack of fit it is necessary to maximize the scalar quantity

$$A_1 = (X_2^* - X^*A)'V_0^{-1}(X_2^* - X^*A)$$

with respect to the coordinates of the k check points. Here X_2^* is a k -element column vector with i th element equal to the value of $x_{i1}^*x_{i2}^*x_{i3}^*$ at the i th check point, X^* is a $k \times 6$ matrix with i th row equal to the value of $(x_{i1}^*, x_{i2}^*, x_{i3}^*, x_{i1}^*x_{i2}^*, x_{i1}^*x_{i3}^*, x_{i2}^*x_{i3}^*)$ at the i th check point, $A = (X'X)^{-1}X'X_2$ is the 6×1 alias vector, and $V_0 = I_k + X^*(X'X)^{-1}X^{*'}.$ This maximization is accomplished by use of the Controlled Random Search Procedure (Price, 1977), which is described in Appendix 2.

When only a single ($k = 1$) optimal check point is desired the Controlled Random Search Procedure locates a point (x_1^*, x_2^*) which maximizes

$$A_1 = (X_2^* - X^*A)'V_0^{-1}(X_2^* - X^*A),$$

where

$$X_2^* = x_1^*x_2^*x_3^* = x_1^*x_2^*(1 - x_1^* - x_2^*),$$

$$X^* = (x_1^*, x_2^*, x_3^*, x_1^*x_2^*, x_1^*x_3^*, x_2^*x_3^*)$$

$$= (x_1^*, x_2^*, (1 - x_1^* - x_2^*), x_1^*x_2^*, x_1^*(1 - x_1^* - x_2^*),$$

$$x_2^*(1 - x_1^* - x_2^*)),$$

and $V_0 = 1 + X^*(X'X)^{-1}X^{*'}.$ The value of A_1 is calculated using the formula of Eq. (3.6). Following this procedure, we find that the single check point that maximizes A_1 , and thus maximizes the power of the test, is the centroid of the triangular factor space $(1/3, 1/3, 1/3)$. The value of A_1 at this centroid point is $A_1 = 0.00084$.

When the Controlled Random Search Procedure is used to locate $k = 2$ simultaneous check points that maximize A_1 , the centroid $(1/3, 1/3, 1/3)$ is selected twice, and $A_1 = 0.00121$. For three simultaneous optimal check points, the centroid is selected three times, and $A_1 = 0.00142$.

To test whether the second order model exhibits lack of fit, when we suspect the special cubic model is the true model, we form the F ratio

$$F = \frac{\underline{d}'V_0^{-1}\underline{d}/k}{\hat{\sigma}_{\text{ext}}^2}$$

with the single check point $(1/3, 1/3, 1/3)$ where $\underline{d} = y_1^* - \hat{y}_1^*$, y_1^* is the observed response, $\hat{y}_1^*$ is the response

predicted by the second order fitted model at $(1/3, 1/3, 1/3)$, and $V_0 = 1 + (1/3, 1/3, 1/3, 1/9, 1/9, 1/9)(X'X)^{-1}(1/3, 1/3, 1/3, 1/9, 1/9, 1/9)'$. If the calculated value of the F ratio exceeds $F_{\alpha;1,v}$ where v equals the number of degrees of freedom associated with $\hat{\sigma}_{\text{ext}}^2$ then we reject the null hypothesis that the second order model is the true model in favor of the alternative hypothesis that the special cubic model is the true model. Equivalently, we reject $H_0: \lambda_1 = 0$ in favor of $H_a: \lambda_1 > 0$. For $k = 2$ or $k = 3$ check points, the value of the F ratio is calculated using the observed and predicted responses at the two or three replicates at the centroid. The hypothesis $H_0: \lambda_1 = 0$ is rejected in favor of $H_a: \lambda_1 > 0$ if F exceeds $F_{\alpha;k,v}$.

Example 2. In Example 2 we illustrate the second of the four cases that could arise when MSE is used as an estimate of σ^2 in the lack of fit test statistic (see Section 3.3.4). We again fit a second order canonical polynomial model in three mixture components, and assume the true model is special cubic. The design to be used is the $q = 3$ simplex centroid design, which consists of seven design points, and is illustrated in Figure 2 of Chapter 1.

There are six parameters to be estimated and seven design points hence one degree of freedom can be used to calculate MSE. We shall use MSE to estimate σ^2 . Optimal upper and lower tailed test check points must be located, and then a decision is made as to which test should

be used. The actual testing for lack of fit involves the F statistic in (3.8).

As in Example 1, $p_2 = 1$, since there is one term in the true model in addition to those in the fitted model. Thus A_1 is a scalar whose value we seek to optimize with respect to the desired number of check points, k . When only a single check point is sought for the purpose of testing lack of fit, the Controlled Random Search Procedure has two functions. First, the procedure is used to locate the optimal candidate check point for an upper tailed test by locating the check point that maximizes the scalar A_1 . Secondly, the procedure is used to locate the optimal candidate check point for a lower tailed test, which is accomplished by locating the point that minimizes A_1 . The quantity $R = [A_1/k] - [A_2/(N - p)]$ is then calculated to determine whether the upper or lower tailed test will be used. If R is positive for the candidate check point for an upper tailed test, then the test is upper tailed, and the test is lower tailed if the candidate check point for a lower tailed test produces a negative value for R . Note that $A_2 = (X_2 - XA)'(X_2 - XA)$ is fixed once the design is specified, since A_2 does not depend on the check points. Using the Controlled Random Search Procedure it is found that the maximum value of A_1 occurs at $(x_1^*, x_2^*, x_3^*) = (1/3, 1/3, 1/3)$, which will be the location for the check point for the upper tailed test. Calculating A_1 at this centroid

point, we find that $R = [A_1/k] - [A_2/(N - p)] = [(3.7258 \times 10^{-4})/1] - [(8.4175 \times 10^{-4})/1] = -4.6917 \times 10^{-4}$. Since R is negative, the test is not upper tailed.

Using the Controlled Random Search Procedure to minimize A_1 , we find that a subregion of the factor space exists in which all points yield a near minimum value for A_1 . We choose the point (0.0189, 0.9269, 0.0542) at random from this subregion to be used as the optimal candidate for a lower tailed test. Here $R = 0 - [(8.4175 \times 10^{-4})/1] = -8.4175 \times 10^{-4}$.

Since R is negative for both the optimal upper tailed test check point and for the optimal lower tailed test check point, we have Case 2 of Section 3.3.4. The upper tailed test check point is disregarded, and the lower tailed test check point (0.0189, 0.9269, 0.0542) is used to test for lack of fit. If the calculated F ratio,

$$F = \frac{\underline{d}'V_0^{-1}\underline{d}}{MSE}$$

is less than $F_{(1-\alpha);1,1}$ then $H_0: \lambda_1 = \lambda_2 = 0$ is rejected in favor of $H_a: [\lambda_1/1] - [\lambda_2/1] < 0$, that is we conclude that the second order model exhibits lack of fit, and the true model is special cubic.

When two simultaneous check points are desired for testing lack of fit, we can again use the Controlled Random Search Procedure to locate the optimal settings. To maximize the scalar A_1 , we find that both check points

should be selected at $(1/3, 1/3, 1/3)$, for an upper tailed test. With our calculations $R = [(5.8275 \times 10^{-4})/2] - [(8.4175 \times 10^{-4})/1] = -5.5038 \times 10^{-4}$, but since R is negative, the test is not upper tailed.

Minimizing A_1 to locate two optimal lower tailed test check points yields a subregion in the factor space of optimal check points. The pair of check points $(0.3749, 0.5752, 0.0499)$ and $(0.5332, 0.4169, 0.0499)$ is selected at random from this subregion, and these check points yield $R = 0 - [(8.4175 \times 10^{-4})/1] = -8.4175 \times 10^{-4}$.

Since R is negative for the upper tailed test points and the lower tailed test points, we have Case 2 of Section 3.3.4 again and the lower tailed test check points are used to test for lack of fit. The hypothesis $H_0: \lambda_1 = \lambda_2 = 0$ is rejected in favor of $H_a: [\lambda_1/2] - [\lambda_2/1] < 0$ if the calculated value of $F = (\underline{d}'V_0^{-1}\underline{d}/2)/MSE$ is less than $F_{(1-\alpha);2,1}$, in which case we say lack of fit of the model is present.

If an external estimate $\hat{\sigma}_{\text{ext}}^2$ had been available for this example, then the optimal upper tailed test check points could have been used in the F ratio, $F = (\underline{d}'V_0^{-1}\underline{d}/k)/\hat{\sigma}_{\text{ext}}^2$, and lack of fit would then be detected if the calculated value of F exceeded $F_{\alpha;k,v}$.

Example 3. Example 3 illustrates the procedure for locating optimal check points when there are two terms in the true model in addition to those in the fitted model. A second order canonical polynomial model in three mixture

components is fitted using a $q = 3$ simplex centroid design. The true model is assumed to contain eight terms, six of which are the same terms as in the fitted model, with the additional two terms being the third order terms $\delta_{12}x_1x_2(x_1 - x_2)$ and $\beta_{123}x_1x_2x_3$. As in Example 2, there is one degree of freedom for MSE which is used to estimate σ^2 . The test statistic, $F = (\underline{d}'V_0^{-1}\underline{d}/k)/\text{MSE}$, is given in equation (3.8).

Since $p_2 = 2$ and A_1 is a 2×2 matrix, locating the optimal upper tailed test check points by the procedure of maximizing λ_1 is assisted by the maximizing of a lower bound for λ_1 , namely maximizing $\mu_{\min}\beta_2/2\sigma^2$, where $\mu_{\min}$ is the smallest eigenvalue of A_1 . Since β_2 and σ^2 are unknown, this is equivalent to maximizing $\mu_{\min}$. For $\mu_{\min}$ to exceed zero, it is necessary that A_1 be of full rank, and since $\text{rank}(A_1) \leq \min(k, p_2)$, it is necessary to select $k \geq 2$ check points. If A_1 is less than full rank, and thus is positive semi-definite, only a subset of possible values of β_2 could be considered to make it possible to test for lack of fit with an upper tailed test.

Using the Controlled Random Search Procedure, the points that maximize $\mu_{\min}$ are found to be (0.418, 0.277, 0.305) and (0.277, 0.418, 0.305). These points are thus optimal candidates for upper tailed test check points. At these check points we have $\mu_{\min} = 5.1623 \times 10^{-4}$, $A_1 = \text{diag}[5.1623 \times 10^{-4}, 5.1916 \times 10^{-4}]$, $A_2 = \text{diag}[0, 8.4175 \times 10^{-4}]$, and $R = [A_1/2] - [A_2/1] = \text{diag}[2.5811 \times 10^{-4}, -5.8217$

$\times 10^{-4}]$. Since the eigenvalues of R are -5.8217×10^{-4} and 2.5811×10^{-4} , R is indefinite. Following the suggested procedure for Case 4 of Section 3.3.4, we note that an upper tailed test for lack of fit exists if the value of $\underline{\beta}_2 = [\delta_{12}, \beta_{123}]'$ is in the column space of $[S_1:S_2]$ but not entirely in the column space of S_2 , where S_1 is the matrix whose columns are the orthonormalized eigenvectors of R corresponding to the positive eigenvalues of R , and S_2 is the matrix whose columns are the orthonormalized eigenvectors of R corresponding to the zero eigenvalues of R . Since R has no zero eigenvalues in this example, S_2 does not exist, but S_1 is the column vector, $S_1 = [1, 0]'$. Thus if $\underline{\beta}_2$ is of the form $\underline{\beta}_2 = [\delta_{12}, 0]'$, where $\delta_{12} \neq 0$, then $\underline{\beta}_2$ is in the column space of S_1 and the test is upper tailed.

The matrix A_2 has rank one and therefore is positive semi-definite. Hence it is impossible to locate two check points that minimize $\mu_{\max}$ and also make $R = [A_1/2] - [A_2/1]$ negative definite (see Appendix 4), that is, it is impossible to find a lower tailed test that is capable of testing lack of fit for all values of $\underline{\beta}_2$. However, if we use the Controlled Random Search Procedure to locate two check points that minimize an upper bound for λ_1 which is $\mu_{\max} \beta_{-2} \beta_{-2} / 2\sigma^2$, then by minimizing $\mu_{\max}$, we find that any of the check points in a particular subregion of the factor space yield a near minimum for $\mu_{\max}$. One pair of points in this subregion is selected as the points to be used as

optimal lower tailed test check points, namely the pair consisting of the point (0.053, 0, 0.947) replicated twice. Replicating this check point, we find $\mu_{\max} = 7.3900 \times 10^{-11}$, $A_1 = \text{diag}[0, 7.3900 \times 10^{-11}]$, $A_2 = \text{diag}[0, 8.4175 \times 10^{-4}]$, and $R = [A_1/2] - [A_2/1] = \text{diag}[0, -8.4175 \times 10^{-4}]$. The eigenvalues of R are 0 and -8.4175×10^{-4} implying that R is negative semi-definite. The values of $\underline{\beta}_2$ that are in the column space of $[S_2:S_3]$ but not entirely in the column space of S_2 will provide a lower tailed test. Here, $[S_2:S_3] = \text{diag}[1,1]$ and $S_2 = [1,0]'$. Thus, the test for lack of fit is lower tailed if $\beta_{123} \neq 0$.

For values of $\underline{\beta}_2$ that produce an upper tailed test we use the check points (0.418, 0.277, 0.305) and (0.277, 0.418, 0.305) with the F ratio

$$F = \frac{\underline{d}'V_0^{-1}\underline{d}/2}{\text{MSE}}$$

and conclude there is lack of fit if the calculated value of F exceeds $F_{\alpha;2,1}$. For values of $\underline{\beta}_2$ that produce a lower tailed test, we use two replicates of the check point (0.053, 0, 0.947), and conclude there is lack of fit if F is less than $F_{(1-\alpha);2,1}$, where again F is calculated by $F = (\underline{d}'V_0^{-1}\underline{d}/2)/\text{MSE}$.

Example 4. Example 4 illustrates Case 3 of Section 3.3.4 in which MSE is used to estimate σ^2 in the lack of fit test statistic. A second order canonical polynomial model

in three mixture components is fitted using the $\{3,3\}$ simplex lattice design, which appears in Figure 5. The true model is assumed to be special cubic, thus $p_2 = 1$ and A_1 is a scalar. The $\{3,3\}$ design consists of ten design points and since there are six parameters to be estimated in the fitted model, σ^2 can be estimated by MSE with $N - p = 10 - 6 = 4$ degrees of freedom.

We first suppose that a single check point is to be used to test for lack of fit. Using the Controlled Random Search Procedure we find the single check point that maximizes the scalar

$$A_1 = (X_2^* - X^*A)'V_0^{-1}(X_2^* - X^*A)$$

is located at the centroid of the simplex factor space.

Thus $(x_1^*, x_2^*, x_3^*) = (1/3, 1/3, 1/3)$ is the optimal candidate for an upper tailed test check point. At this centroid point, $A_1 = 4.9076 \times 10^{-4}$. For the $\{3,3\}$ design the scalar quantity $A_2 = (X_2 - XA)'(X_2 - XA)$ is fixed and is equal to $A_2 = 9.4062 \times 10^{-4}$ and thus, $R = [A_1/k] - [A_2/(N - p)] = [(4.9076 \times 10^{-4})/1] - [(9.4062 \times 10^{-4})/4] = 2.5560 \times 10^{-4}$.

The point that is the optimal candidate for a lower tailed test check point is chosen randomly from a subregion of points in the factor space, in which all points minimize A_1 . The point selected has the value $(x_1^*, x_2^*, x_3^*) = (0.560, 0.410, 0.030)$. Here $A_1 = 9.6590 \times 10^{-7}$ and $R = [(9.6590 \times 10^{-7})/1] - [(9.4062 \times 10^{-4})/4] = -2.3419 \times 10^{-4}$.

Figure 5. The $\{3,3\}$ simplex lattice design.

Since R is positive for the optimal upper tailed test check point $(1/3, 1/3, 1/3)$ and R is negative for the optimal lower tailed test check point $(0.560, 0.410, 0.030)$ we are in Case 3 of Section 3.3.4. Either the upper or lower tailed test could be used to test for lack of fit, but if the quantity $\beta_{-2}'\beta_{-2}/\sigma^2$ can be specified, then we will choose to use the test that has greater minimum power, since greater power means that we are more likely to detect lack of fit when in fact lack of fit exists. In this example $\beta_{-2} = \beta_{123}$.

For illustrative purposes, we arbitrarily choose $\beta_{-2}'\beta_{-2}/\sigma^2 = 2000$, so that an approximate conservative lower bound for the power of the upper tailed test is found by

calculating

$$P \{ F''_{k, N-p; \lambda_{1L}, \lambda_{2U}} > F_{\alpha; k, N-p} \}$$

where $F_{\alpha; k, N-p}$ is the upper 100α percentage point of the central F distribution, k is the number of check points, N is the total number of response observations, p is the number of parameters in the fitted model,

$\lambda_{1L} = \mu_{\min} \beta_2' \beta_2 / 2\sigma^2$, and $\lambda_{2U} = \delta_{\max} \beta_2' \beta_2 / 2\sigma^2$. The quantity $\mu_{\min}$ is the smallest eigenvalue of A_1 , where A_1 is evaluated at the optimal upper tailed test check point. Since A_1 is a scalar, $\mu_{\min} = A_1$. Likewise, $\delta_{\max}$ is the largest eigenvalue of A_2 , and since in this example A_2 is a scalar, $\delta_{\max} = A_2$. In this example we have $k = 1$, $N - p = 10 - 6 = 4$, $\lambda_{1L} = \mu_{\min} \beta_2' \beta_2 / 2\sigma^2 = (4.9076 \times 10^{-4})(2000/2) = 4.9076 \times 10^{-1}$, and $\lambda_{2U} = \delta_{\max} \beta_2' \beta_2 / 2\sigma^2 = (9.4062 \times 10^{-4})(2000/2) = 9.4062 \times 10^{-1}$. Using the approximation to the cumulative probabilities of the doubly noncentral F distribution given by Johnson and Kotz (1970, p.197) which is described in Appendix 6, and taking $\alpha = .05$, we find that a conservative lower bound for the power of the optimal upper tailed test is approximately equal to .0649.

The minimum power for the optimal lower tailed test is approximated (assuming $\beta_2' \beta_2 / \sigma^2 = 2000$) by calculating

$$P \{ F''_{k, N-p; \lambda_{1U}, \lambda_{2L}} < F_{(1-\alpha); k, N-p} \}.$$

The quantities λ_{1U} and λ_{2L} are taken as $\lambda_{1U} = \mu_{\max} \beta_2' \beta_2 / 2\sigma^2$ and $\lambda_{2L} = \delta_{\min} \beta_2' \beta_2 / 2\sigma^2$ where $\mu_{\max}$ is the largest eigenvalue of A_1 with A_1 calculated using the optimal lower tailed test check point, and where $\delta_{\min}$ is the smallest eigenvalue of A_2 . Since A_1 and A_2 are scalars, $\mu_{\max} = A_1$ and $\delta_{\min} = A_2$. In this example, $k = 1$, $N - p = 4$,

$\lambda_{1U} = (9.6590 \times 10^{-7})(2000/2) = 9.6590 \times 10^{-4}$, and
 $\lambda_{2L} = (9.4062 \times 10^{-4})(2000/2) = 9.4062 \times 10^{-1}$. Again if the approximation to the doubly noncentral F distribution given in Johnson and Kotz is used, an approximate conservative lower bound for the power of the optimal lower tailed test is .0555.

Having specified $\beta_2' \beta_2 / \sigma^2 = 2000$, the optimal upper tailed test is chosen over the optimal lower tailed test, because the approximate minimum power of the upper tailed test is greater than the approximate minimum power of the lower tailed test. Using the optimal upper tailed test check point $(1/3, 1/3, 1/3)$ in the test statistic

$$F = \frac{\underline{d}' V_0^{-1} \underline{d}}{\text{MSE}}$$

we conclude that lack of fit is significant if the calculated value of F exceeds $F_{\alpha; 1, 4}$, in which case we reject $H_0: \lambda_1 = \lambda_2 = 0$ in favor of $H_a: \lambda_1/1 - \lambda_2/4 > 0$.

When two simultaneous check points are used for testing lack of fit, the Controlled Random Search Procedure locates the optimal upper tailed test and optimal lower tailed test

check points. It turns out that two replicates at $(1/3, 1/3, 1/3)$ maximize A_1 , and are used as optimal check points for an upper tailed test. The value of $R = [A_1/2] - [A_2/4]$ is $[(7.9210 \times 10^{-4})/2] - [(9.4062 \times 10^{-4})/4] = 1.6090 \times 10^{-4}$.

In searching for two optimal lower tailed test check points, again a subregion of the factor space is found in which any of the points nearly minimize A_1 . From this subregion are chosen the points $(0.6386, 0.3263, 0.0351)$ and $(0.7257, 0.2421, 0.0322)$ resulting in a value of $R = [A_1/2] - [A_2/4]$ of $[(1.5216 \times 10^{-9})/2] - [(9.4062 \times 10^{-4})/4] = -2.3516 \times 10^{-4}$.

In conclusion, when two simultaneous check points are used in the test for lack of fit in this example, R is positive for the optimal upper tailed test and R is negative for the optimal lower tailed test, and we have Case 3 of Section 3.3.4. Selecting $\underline{\beta}'_2 \underline{\beta}_2 / \sigma^2 = 2000$ arbitrarily, we found the approximate lower bound for the power of the upper tailed test to be .0504, and the approximate lower bound for the power of the lower tailed test to be .0612. Since the power is higher with the lower tailed test it is our choice for testing lack of fit when two check points are used simultaneously. Lack of fit is detected and we reject $H_0: \lambda_1 = \lambda_2 = 0$ in favor of $H_a: [\lambda_1/2] - [\lambda_2/4] < 0$ if the F ratio, $F = (\underline{d}' V_0^{-1} \underline{d} / 2) / \text{MSE}$, using the optimal lower tailed test check points $(0.6386, 0.3263, 0.0351)$ and $(0.7257, 0.2421, 0.0322)$ is calculated to be less than $F_{(1-\alpha); 2, 4}$.

3.4.2 Numerical Examples

Numerical Example 1. In this example we illustrate numerically some of the findings in the first theoretical example of Section 3.4.1. Data that were collected in a rocket fuel experiment (Kurotori, 1966) will be used to investigate the power of the lack of fit F test. The test is set up to detect the inadequacy of a fitted second order canonical polynomial model when the true model is special cubic. Calculated values of the power of the test which detects lack of fit through large values of

$$F = \frac{\underline{d}' V_0^{-1} \underline{d} / k}{\hat{\sigma}_{\text{ext}}^2}$$

will be compared for several check point locations, including the location (1/3, 1/3, 1/3) at which the power was found to be maximum in Example 1 of Section 3.4.1.

In Kurotori's experiment the modulus of elasticity (Y) of a rocket fuel is expressed as a function of the proportions of three components--binder (x_1), oxidizer (x_2), and fuel (x_3). Since lower bounds are placed on the component proportions x_1 , x_2 , and x_3 , in the form of $0.20 < x_1$, $0.40 < x_2$, and $0.20 < x_3$, pseudocomponents (x'_i) are defined in terms of the original components in the form of $x'_1 = (x_1 - 0.20)/(1 - .80)$, $x'_2 = (x_2 - 0.40)/(1 - .80)$, and $x'_3 = (x_3 - 0.20)/(1 - .80)$. The true special cubic model in the pseudocomponents, which is obtained by using the data at the seven points of the simplex centroid design

in the pseudocomponent system, is

$$E(Y) = 2350x'_1 + 2450x'_2 + 2650x'_3 + 0x'_1x'_2 \\ + 1000x'_1x'_3 + 1600x'_2x'_3 + 6150x'_1x'_2x'_3.$$

The second order canonical polynomial model that is fitted to the six boundary points only, and which will be tested for lack of fit, is given by

$$\hat{Y} = 2350x'_1 + 2450x'_2 + 2650x'_3 \\ + 1000x'_1x'_3 + 1600x'_2x'_3.$$

The configuration of the experimental design as well as the check point locations are depicted in Figure 4 of Chapter 2 and the observed response values are given in Table 3 of this chapter.

A value of the ratio $F = [\underline{d}'V_0^{-1}\underline{d}]/\hat{\sigma}_{\text{ext}}^2$ is calculated at each of the four individual check points $(1/3, 1/3, 1/3)$, $(2/3, 1/6, 1/6)$, $(1/6, 2/3, 1/6)$, and $(1/6, 1/6, 2/3)$ where $\hat{\sigma}_{\text{ext}}^2$ is assumed to have the value $\hat{\sigma}_{\text{ext}}^2 = 144$ as suggested by Kurotori (1966). We also assume without loss of generality that the degrees of freedom associated with $\hat{\sigma}_{\text{ext}}^2$ are $\nu = 10$. The power of the F test is calculated at each of the four check points by using the approximation to the cumulative probabilities of the noncentral F

Table 3. Observed Response Values at the Pseudocomponent Settings for Kurotori's Rocket Fuel Experiment--Numerical Example 1.

Observation Number	Binder x'_1	Oxidizer x'_2	Fuel x'_3	Modulus of Elasticity Y
1	1	0	0	2350
2	0	1	0	2450
3	0	0	1	2650
4	1/2	1/2	0	2400
5	1/2	0	1/2	2750
6	0	1/2	1/2	2950
7*	1/3	1/3	1/3	3000
8*	2/3	1/6	1/6	2690
9*	1/6	2/3	1/6	2770
10*	1/6	1/6	2/3	2980

*Check Points.

distribution given by Johnson and Kotz (1970, p. 197) to evaluate

$$\text{Power} = P\{F'_{1,10;\lambda_1} > F_{.05;1,10}\}$$

where $\lambda_1 = A_1 \beta_{123}^2 / 2\sigma_{\text{ext}}^2$. The value of

$A_1 = (X_2^* - X^*A)'V_0^{-1}(X_2^* - X^*A)$ is calculated for each check point using the values of X_2^* , X^* , V_0^{-1} and the value of

$A = (X'X)^{-1}X'X_2$ which is fixed by the {3,2} simplex lattice design. Since the {3,2} simplex lattice consists of points

only on the boundaries of the triangle (and therefore at each point at least one of the x_i^1 values is equal to zero), then $X_2 = 0$ and $A = 0$. From the true special cubic model, $\beta_{123} = 6150$.

The calculated value of F as well as the approximate value for the power at each of the four check points is given in Table 4. The check point $(1/3, 1/3, 1/3)$ produced the highest power of the four check points investigated, supporting the previous results of Example 1 in Section 3.4.1 where $(1/3, 1/3, 1/3)$ was selected as the check point location with the maximum power when a second order canonical polynomial was fitted using the $\{3,2\}$ simplex lattice design, but the true model was assumed to be special cubic. Additional support for the point $(1/3, 1/3, 1/3)$ being optimal is given by the contour plot of values of A_1 in Figure 6(d). The highest values of A_1 appear near the centroid $(1/3, 1/3, 1/3)$ where high A_1 values translate into high λ_1 values, since $\lambda_1 = A_1 \beta_{123}^2 / 2\sigma^2$, which in turn implies high power since we know the power is an increasing function of λ_1 .

As a second part of this example the power of the F test that is obtained when three replicates are taken at $(1/3, 1/3, 1/3)$ is compared to the power of the F test that is obtained when one replicate is taken at the three check points $(2/3, 1/6, 1/6)$, $(1/6, 2/3, 1/6)$, and $(1/6, 1/6, 2/3)$. These latter three point locations were suggested by Kurotori for testing lack of fit of his fitted special

Table 4. Effect of Check Point Location on the Power of the Lack of Fit F Test--Numerical Example 1.

Check Point**	Numerator d.f. (k)	Denominator d.f. (v)	A ₁	λ ₁	F	Power
(1/3,1/3,1/3)	1	10	.00084	110.32	221.1*	1.000
(2/3,1/6,1/6)	1	10	.00023	30.21	65.0*	.999
(1/6,2/3,1/6)	1	10	.00023	30.21	44.8*	.999
(1/6,1/6,2/3)	1	10	.00023	30.21	72.6*	.999
(1/3,1/3,1/3)	3	10	.00142	186.49	-----	1.000
(1/3,1/3,1/3)						
(1/3,1/3,1/3)						
(2/3,1/6,1/6)	3	10	.00051	66.98	45.3*	.999
(1/6,2/3,1/6)						
(1/6,1/6,2/3)						

* F > F_{.05;k,v}.

**Check point coordinates are in the pseudocomponent system.

--F ratio cannot be calculated since only one response observation was collected at (1/3,1/3,1/3).

cubic model. The result of this comparison, see Table 4, is that the three replicates at $(1/3, 1/3, 1/3)$ produce the test with greater power which again supports the findings of Example 1 of Section 3.4.1.

All of the check point locations listed in Table 4 produce very high power values ($> .999$) which is due in part to the high value of β_{123} ($\beta_{123} = 6150$). If β_{123} were of lower magnitude, then the three replicates at $(1/3, 1/3, 1/3)$ would show a still greater superiority in the power value compared to the power using the other check points. This superiority is demonstrated in Table 5 where values of β_{123} are listed as 3000 and 1500 and the comparative power values are listed as 0.998 compared to 0.795 and 0.662 compared to 0.249, respectively. Table 5 also demonstrates the superior power value for the point $(1/3, 1/3, 1/3)$ when $\beta_{123} = 3000$ or $\beta_{123} = 1500$ and each of the four check points is used one at a time.

Finally, $(1/3, 1/3, 1/3)$ being the optimal check point location is seen in Figure 6(c), where contour plots of the expected difference in the heights of the surfaces are drawn. The differences in the heights are found by subtracting the estimated height of the surface obtained with the fitted second order model from the estimated height of the surface obtained with the true special cubic model. The expected difference between the true and fitted surfaces approaches a maximum the closer one moves to the centroid of the simplex factor space, so that the optimal check point

Table 5. Effect of the Magnitude of β_{123} on the Power of the Lack of Fit F Test--Numerical Example 1.

Check Points**	$\beta_{123} = 6150$		$\beta_{123} = 3000$		$\beta_{123} = 1500$	
	λ_1	Power	λ_1	Power	λ_1	Power
(1/3,1/3,1/3)	110.32	1.000	26.25	.998	6.56	.638
(2/3,1/6,1/6)	30.21	.999	7.19	.680	1.80	.220
(1/6,2/3,1/6)	30.21	.999	7.19	.680	1.80	.220
(1/6,1/6,2/3)	30.21	.999	7.19	.680	1.80	.220
(1/3,1/3,1/3)						
(1/3,1/3,1/3)						
(1/3,1/3,1/3)	186.49	1.000	44.38	.998	11.09	.662
(2/3,1/6,1/6)						
(1/6,2/3,1/6)						
(1/6,1/6,2/3)	66.98	.999	15.94	.795	3.98	.249

**Check point coordinates are in the pseudocomponent system.

(a) True special cubic surface.

(b) Fitted second order surface.

(c) Expected difference between the true special cubic surface and the fitted second order surface.

(d) $A_1 = (X_2^* - X^*A)' V_0^{-1} (X_2^* - X^*A)$.

Figure 6. Contour plots for Numerical Example 1.

location $(1/3, 1/3, 1/3)$ coincides with the point where the expected difference between the true special cubic surface and the fitted second order surface is maximum.

Numerical Example 2. In this second numerical example, we investigate the power of the F test for detecting lack of fit when a second order canonical polynomial model is fitted in a mixture system which is in truth represented by a special cubic model. The true model is assumed to be

$$E(Y) = 2350x_1 + 2450x_2 + 2650x_3 \\ + 1000x_1x_3 + 1600x_2x_3 + 6150x_1x_2x_3$$

which is used to generate hypothetical response observations at the seven points of the $q = 3$ simplex centroid design as well as at three check points. The values of the response are obtained by adding the value of a pseudorandom normal variate with mean 0 and variance 144 to each true predicted response value. The data are given in Table 6.

The response values at the seven points of the simplex centroid design are used in the least squares normal equations to obtain the fitted second order model

$$\hat{Y} = 2341x_1 + 2438x_2 + 2630x_3 \\ + 310x_1x_2 + 1304x_1x_3 + 1970x_2x_3$$

Table 6. Generated Response Values--Numerical Example 2.

x_1	x_2	x_3	Y
1	0	0	2357
0	1	0	2454
0	0	1	2646
1/2	1/2	0	2403
1/2	0	1/2	2747
0	1/2	1/2	2962
1/3	1/3	1/3	3013
* 1/3	1/3	1/3	2993
* 2/3	1/6	1/6	2693
* .02	.93	.05	2550

*Check points.

which is to be tested for lack of fit using the test statistic $F = \underline{d}'V_0^{-1}\underline{d}/MSE$. The F statistic will be evaluated at each of the three check points $(1/3, 1/3, 1/3)$, $(2/3, 1/6, 1/6)$, and $(.02, .93, .05)$, taken one at a time, and the power of the test at the three check point locations will be calculated and compared. The test is lower tailed for all check point locations (since $R = A_1 - A_2$ is negative for all check point locations) and thus the power is defined as

$$P\{F_{1,1;\lambda_1,\lambda_2}'' < F_{.95;1,1}\}.$$

Figure 7. Contours of $R = A_1 - A_2$ for Numerical Example 2.

The values of $\lambda_1 = \beta_{123}^2 A_1 / 2\sigma^2$ and $\lambda_2 = \beta_{123}^2 A_2 / 2\sigma^2$ are found by taking $\beta_{123} = 6150$ and $\sigma^2 = 144$. The results of this power investigation are given in Table 7.

Since the check point $(.02, .93, .05)$ produces the greatest power of the three check points investigated, this supports the result in Example 2 of Section 3.4.1, where we saw that the point $(.02, .93, .05)$ yielded the maximum power of all points for detecting lack of fit of a fitted second order canonical polynomial model, using the $q = 3$ simplex centroid design in the presence of a true special cubic surface. Additional evidence for the point $(.02, .93, .05)$ being an optimal check point is shown in Figure 7, where contours of the values of $R = A_1 - A_2$ are presented. The point $(.02, .93, .05)$ is seen to belong to an area of the

Table 7. Effect of Check Point Location on the Power of the Lack of Fit F Test--Numerical Example 2.

Check Point	Numerator d.f.(k)	Denominator d.f.(N - p)	A ₁	A ₂	R	λ_1	λ_2	F	Power**
(1/3,1/3,1/3)	1	1	3.70×10^{-4}	.00084	-.00047	48.59	110.3	.333	.000
(2/3,1/6,1/6)	1	1	5.20×10^{-5}	.00084	-.00079	6.84	110.3	.063	.012
(.02,.93,.05)	1	1	8.12×10^{-9}	.00084	-.00084	.001	110.3	.0005*	.590

*Significant at $\alpha = .05$ ($F < F_{.95;1,1}$).

**Power is approximated using an approximation found in Johnson and Kotz (1970, p.197).

simplex factor space where R is minimum, which implies that A_1 (and in turn λ_1) is also minimum in this area, since $R = A_1 - A_2$ and A_2 has the fixed value of $A_2 = .00084$ for the simplex centroid design. Thus the check point $(.02, .93, .05)$ produces a minimal λ_1 value and maximum power, since the power increases with decreasing values of λ_1 .

3.5 Discussion

When check points are used for testing lack of fit in a mixture model, the appropriate testing procedure, assuming a normally distributed response, involves an F statistic. If an external estimate, $\hat{\sigma}_{\text{ext}}^2$, of the experimental error variance is available so that the test statistic is given by

$$F = \frac{\underline{d}' \underline{V}_0^{-1} \underline{d} / k}{\hat{\sigma}_{\text{ext}}^2}$$

then the power of the test for lack of fit is maximized by choosing k check points that maximize the value of the non-centrality parameter λ_1 . When $p_2 = 1$, maximizing λ_1 is achieved without knowing the value of the elements of $\underline{\beta}_2$ by selecting check points that maximize the scalar A_1 . When $p_2 > 1$, the maximization of λ_1 is approximated by maximizing a lower bound for λ_1 . This is achieved also without knowing the values of the elements of $\underline{\beta}_2$ by selecting check points that maximize the smallest eigenvalue of the matrix A_1 . The test is upper tailed, and for given values of λ_1 , the actual

power of the test can be calculated from the cumulative probabilities of the noncentral F distribution. A problem arises when A_1 is positive semi-definite and its smallest eigenvalue is equal to zero. In this case check points that maximize the smallest positive eigenvalue of A_1 are selected, and lack of fit is only detectable for a subset of the possible values of the elements of $\underline{\beta}_2$.

When an external estimate of σ^2 is not available, testing lack of fit at the check points is further complicated. The F statistic is

$$F = \frac{(\underline{d}' V_0^{-1} \underline{d} / k)}{MSE}$$

and the rejection region for the lack of fit test can be upper tailed or lower tailed. The power of the test is determined by using the doubly noncentral F distribution, which depends on the parameters k , $N - p$, λ_1 , and λ_2 . Of these four parameters, only k and λ_1 are influenced by check points, and if the value of k is fixed, the power of the test is maximized by choosing check points that affect the value of λ_1 . Regardless of the values of the elements of $\underline{\beta}_2$, check points that maximize λ_1 are selected for maximizing the power of an upper tailed test, and check points that minimize λ_1 are selected for maximizing the power of a lower tailed test.

Lack of fit can be tested with the upper tailed test for all nonzero values of the elements of β_2 if the check points are selected so that $[A_1/k] - [A_2/(N - p)]$ is positive definite, since this forces the expected value of the numerator mean square in the F ratio to be greater than the expected value of the denominator mean square. Similarly, lack of fit can be tested with a lower tailed test if check points are selected which make $[A_1/k] - [A_2/(N - p)]$ negative definite. When it is not possible to select check points that make $[A_1/k] - [A_2/(N - p)]$ either positive definite or negative definite then detection of lack of fit is only possible for a subset of all nonzero values of the elements of β_2 .

The power of the test for lack of fit using the F statistic in (3.8) is a function of λ_1 and λ_2 . Since the magnitudes of λ_1 and λ_2 are influenced by the experimental design, an area for future study is the investigation of the effect of the experimental design on the power of the lack of fit test. In the presence of an external estimate of σ^2 , Atkinson (1972) suggested designs that maximize the determinant of A_2 , $|A_2|$, when lack of fit is to be detected by a large value of λ_2 using a procedure which is, in general, equivalent to the lack of fit testing procedure that partitions the residual sum of squares into pure error and lack of fit sums of squares. It might be useful to apply Atkinson's (1972) methodology not only to $|A_2|$, but to $|A_1|$ or $|A_1/k - A_2/(N - p)|$ in order to find an

appropriate design when testing lack of fit with the F ratio in Eq. (3.8). Since the power of the test in Eq. (3.8) is also affected by the values of k and $N - p$, which are the numerator and denominator degrees of freedom of the doubly noncentral F distribution, respectively, optimal settings for these parameters can also be considered. For a given fitted model, p is fixed so that the degrees of freedom would be influenced by the number of check points selected, k , and by N , the total number of observations. Finally, the experimental design and the number of check points also affect the power of the F test when $\hat{\sigma}_{\text{ext}}^2$ is used to estimate σ^2 . Thus the effect of the experimental design and the number of check points can also be investigated for the situation where the lack of fit test statistic is given by Eq. (3.3).

We now conclude our investigation of the check point approach to lack of fit testing and in the next chapter turn to an investigation of a near neighbor method for testing lack of fit.

CHAPTER FOUR
USE OF NEAR NEIGHBOR OBSERVATIONS
FOR TESTING LACK OF FIT

4.1 Introduction

In an experiment in which replicate response observations are available at one or more design points, lack of fit of a fitted model can be tested by a procedure which involves partitioning the residual sum of squares into two statistically independent portions. One portion is the sum of squares due to lack of fit (SS_{LOF}), and the second portion is the sum of squares due to pure error (SSE_{pure}) obtained from the replicates. As discussed in Section 2.2, this procedure was suggested by Draper and Smith (1981, p.120). Lack of fit is inferred if the calculated value of the ratio

$$F = \frac{MS_{LOF}}{MSE_{pure}} \quad (4.1)$$

exceeds the corresponding upper 100α percentage point of the central F distribution, where MS_{LOF} and MSE_{pure} are the mean square values found by dividing SS_{LOF} and SSE_{pure} by their respective degrees of freedom.

In order to test the fitted model for lack of fit when replicate observations are not available, Shillington (1979) suggested a procedure which uses observed response values

collected at points which are "near neighbors" in the factor space in place of replicates (see Section 2.3). Lack of fit is inferred when the calculated value of the ratio

$$F = \frac{MSE_B}{MSE_W} \quad (4.2)$$

exceeds the upper 100α percentage point of the central F distribution. The numerator, MSE_B , of the F ratio in Eq. (4.2) is a generalization of the numerator, MS_{LOF} , of the F ratio in Eq. (4.1). The form of MSE_B will be given in Eq. (4.5) of Section 4.3. The denominator, MSE_W , in Eq. (4.2) is a generalization of MSE_{pure} in Eq. (4.1), and the value of MSE_W is calculated using near neighbor observations in place of replicates (see Eq. (4.6) in Section 4.3).

Shillington's near neighbor method provides an alternative to the check points method when replicate observations are not available. Typically, near neighbors might appear either because an experiment was not designed to provide replicate observations or with a designed experiment consisting of a large number of design points in a bounded factor space which results in some points being near one another.

In this chapter we shall further study Shillington's (1979) near neighbor procedure for testing lack of fit. A question involving the correctness of the ordinary least squares technique suggested by Shillington for deriving the denominator, MSE_W , of the F ratio in Eq. (4.2) will

be raised. The question will be resolved by showing the equivalence of deriving MSE_W by ordinary least squares and of deriving MSE_W by a generalization of weighted least squares. We will verify that when the observable response values are assumed to have the normal distribution with homogeneous variance, σ^2 , the F ratio in Eq. (4.2) possesses a central F distribution when the fitted model is adequate, but possesses a doubly noncentral F distribution when the fitted model suffers from lack of fit. We shall also show that the F test for lack of fit which uses the statistic in Eq. (4.2) can have either an upper tailed or a lower tailed rejection region. Finally, the use of a clustering algorithm for defining groups of near neighbors will be proposed.

4.2 Notation

In this section we introduce the notation to be used in this chapter. Throughout our investigation of Shillington's near neighbor procedure for testing lack of fit, we shall assume the observed response values collected in an experiment can be grouped into g cells where the j th cell contains n_j observations, $j = 1, 2, \dots, g$. The observations in a cell are from points that are "near neighbors" in the sense that they are near one another in the factor (mixture) space. A model of the form

$$E(\underline{Y}) = X\underline{\beta}_1 \quad (4.3)$$

is fitted to the data using ordinary least squares, but the true model is assumed to have the form

$$E(\underline{Y}) = X\underline{\beta}_1 + X_2\underline{\beta}_2, \quad (4.4)$$

where $\underline{Y}$ is an $N \times 1$ vector of response values observable at the design points with $\text{var}(\underline{Y}) = \sigma^2 I_N$, X and X_2 are $N \times p$ and $N \times p_2$ matrices of known constants, respectively, and $\underline{\beta}_1$ and $\underline{\beta}_2$ are $p \times 1$ and $p_2 \times 1$ vectors of unknown regression coefficients, respectively. The vector $\underline{Y}$ is assumed to have the normal distribution.

Let us now define the following vector and matrix quantities to be used in developing the numerator, MSE_B , of the F ratio in Eq. (4.2):

$\bar{Y}_C =$ a $g \times 1$ vector with j th element equal to the average of the n_j observed response values in the j th cell of near neighbor observations, $j = 1, 2, \dots, g$.

$X_C =$ a $g \times p$ matrix whose j th row is the average of the n_j rows of X corresponding to the j th cell, $j = 1, 2, \dots, g$.

$X_{2C} =$ a $g \times p_2$ matrix whose j th row is the average of the n_j rows of X_2 corresponding to the j th cell, $j = 1, 2, \dots, g$.

$G_0 =$ a $g \times g$ diagonal matrix of the form

$$G_0 = \text{diag}[1/n_1, 1/n_2, \dots, 1/n_g].$$

To further illustrate the forms of $\bar{Y}_C$, X_C , X_{2C} , and G_0 , we present the following numerical example. Consider a data set consisting of response observations (Y) taken at $N = 8$ different combinations of the settings of the factors x_1 and x_2 , where the eight response observations are divided into $g = 4$ near neighbor cells. The vector of observed response values, $\underline{Y}$, and the matrix X corresponding to the first order model, $E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$, are

$$\underline{Y} = \begin{bmatrix} 10 \\ \hline 13 \\ 16 \\ \hline 15 \\ 18 \\ \hline 21 \\ \hline 27 \\ 30 \end{bmatrix} \quad X = \begin{bmatrix} 1 & 1 & 2 \\ \hline 1 & 2 & 5 \\ 1 & 2 & 4 \\ \hline 1 & 3 & 2 \\ 1 & 3 & 1 \\ \hline 1 & 4 & 2 \\ \hline 1 & 5 & 5 \\ 1 & 5 & 4 \end{bmatrix}.$$

The horizontal lines in $\underline{Y}$ and X delineate the four cells of near neighbors. In this example

$$\bar{Y}_C = \begin{bmatrix} 10 \\ (13 + 16)/2 \\ (15 + 18 + 21)/3 \\ (27 + 30)/2 \end{bmatrix} = \begin{bmatrix} 10 \\ 14.5 \\ 18 \\ 28.5 \end{bmatrix},$$

$$\begin{aligned}
 x_C &= \begin{bmatrix} 1 & 1 & 2 \\ (1+1)/2 & (2+2)/2 & (5+4)/2 \\ (1+1+1)/3 & (3+3+4)/3 & (2+1+2)/3 \\ (1+1)/2 & (5+5)/2 & (5+4)/2 \end{bmatrix} \\
 &= \begin{bmatrix} 1 & 1 & 2 \\ 1 & 2 & 4.5 \\ 1 & 3.3 & 1.7 \\ 1 & 5 & 4.5 \end{bmatrix} ,
 \end{aligned}$$

and

$$G_0 = \text{diag}(1, 1/2, 1/3, 1/2).$$

If the true model is second degree, $E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2$, then the x_2 and x_{2C} matrices have three columns corresponding to the terms $x_1 x_2$, x_1^2 , and x_2^2 , respectively. For this numerical example we have

$$x_2 = \begin{bmatrix} 2 & 1 & 4 \\ 10 & 4 & 25 \\ 8 & 4 & 16 \\ 6 & 9 & 4 \\ 3 & 9 & 1 \\ 8 & 16 & 4 \\ 25 & 25 & 25 \\ 20 & 25 & 16 \end{bmatrix}$$

and

$$x_{2C} = \begin{bmatrix} 2 & 1 & 4 \\ 9 & 4 & 20.5 \\ 5.7 & 11.3 & 3 \\ 22.5 & 25 & 20.5 \end{bmatrix} .$$

Next let us define the following quantities to be used in developing the denominator, MSE_W , of the F ratio in Eq. (4.2):

$\underline{W}$ = an $N \times 1$ vector of within cell deviations where the i th element, W_i , of $\underline{W}$ is equal to the difference between the i th element, Y_i , of $\underline{Y}$ and the average of the response values observed in the near neighbor cell containing Y_i , $i = 1, 2, \dots, N$.

X_W = an $N \times p$ matrix whose i th row is equal to the i th row of the X matrix minus the row of X_C corresponding to the cell containing the response value observed at the i th row of X .

r = $\text{rank}(X_W)$.

X_{2W} = an $N \times p_2$ matrix whose i th row is equal to the i th row of the X_2 matrix minus the row of X_{2C} corresponding to the cell containing the response value observed at the i th row of X .

Σ_0 = an $N \times N$ idempotent matrix of the form

$$\Sigma_0 = I_N - \text{diag}[(1/n_1)J_1, (1/n_2)J_2, \dots, (1/n_g)J_g]$$

where J_j is an $n_j \times n_j$ matrix of ones, and I_N is the identity matrix of order $N \times N$, with $N = \sum_{j=1}^g n_j$.

Let us illustrate the forms of W , X_W , X_{2W} , and Σ_0 by using the numerical example presented earlier in this section, where the eight response observations were distributed among four cells. For these data we have

$$\underline{W} = \begin{bmatrix} 10 & - & 10 \\ 13 & - & 29/2 \\ 16 & - & 29/2 \\ 15 & - & 18 \\ 18 & - & 18 \\ 21 & - & 18 \\ 27 & - & 57/2 \\ 30 & - & 57/2 \end{bmatrix} = \begin{bmatrix} 0 \\ -1.5 \\ 1.5 \\ -3 \\ 0 \\ 3 \\ -1.5 \\ 1.5 \end{bmatrix} ,$$

$$X_W = \begin{bmatrix} 1 & - & 1 & 1 & - & 1 & 2 & - & 2 \\ 1 & - & 1 & 2 & - & 2 & 5 & - & 9/2 \\ 1 & - & 1 & 2 & - & 2 & 4 & - & 9/2 \\ 1 & - & 1 & 3 & - & 10/3 & 2 & - & 5/3 \\ 1 & - & 1 & 3 & - & 10/3 & 1 & - & 5/3 \\ 1 & - & 1 & 4 & - & 10/3 & 2 & - & 5/3 \\ 1 & - & 1 & 5 & - & 5 & 5 & - & 9/2 \\ 1 & - & 1 & 5 & - & 5 & 4 & - & 9/2 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & .5 \\ 0 & 0 & -.5 \\ 0 & -.3 & .3 \\ 0 & -.3 & -.7 \\ 0 & .7 & .3 \\ 0 & 0 & .5 \\ 0 & 0 & -.5 \end{bmatrix} ,$$

$$X_{2W} = \begin{bmatrix} 2 & - & 2 & 1 & - & 1 & 4 & - & 4 \\ 10 & - & 9 & 4 & - & 4 & 25 & - & 41/2 \\ 8 & - & 9 & 4 & - & 4 & 16 & - & 41/2 \\ 6 & - & 17/3 & 9 & - & 34/3 & 4 & - & 3 \\ 3 & - & 17/3 & 9 & - & 34/3 & 1 & - & 3 \\ 8 & - & 17/3 & 16 & - & 34/3 & 4 & - & 3 \\ 25 & - & 45/2 & 25 & - & 25 & 25 & - & 41/2 \\ 20 & - & 45/2 & 25 & - & 25 & 16 & - & 41/2 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 4.5 \\ -1 & 0 & -4.5 \\ .3 & -2.3 & 1 \\ -2.7 & -2.3 & -2 \\ 2.3 & 4.7 & 1 \\ 2.5 & 0 & 4.5 \\ -2.5 & 0 & -4.5 \end{bmatrix} ,$$

and

$$\Sigma_0 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & -1/2 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1/2 & 1/2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2/3 & -1/3 & -1/3 & 0 & 0 \\ 0 & 0 & 0 & -1/3 & 2/3 & -1/3 & 0 & 0 \\ 0 & 0 & 0 & -1/3 & -1/3 & 2/3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1/2 & -1/2 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1/2 & 1/2 \end{bmatrix} .$$

4.3 Shillington's Procedure

We originally described Shillington's near neighbor procedure for testing lack of fit in Section 2.3 of Chapter Two. We now reintroduce the procedure by using the matrix

and vector quantities defined in Section 4.2. The test that Shillington proposed involves the use of an F ratio (see Eq. (4.2)) of two statistically independent mean square values, each of which is an unbiased estimate of σ^2 when the fitted model is the correct true model. The two independent mean squares become biased estimates of σ^2 when the fitted model suffers from lack of fit. Shillington's methodology detects lack of fit when the calculated value of the F ratio in Eq. (4.2) is large, thus his test is upper tailed. We shall see later in Section 4.7 that the test is not always upper tailed, and may be lower tailed. Shillington points out that the power of the test depends on the relative magnitudes of $E(MSE_B)$ and $E(MSE_W)$, that is, the power depends on the difference between the expected values of the numerator and of the denominator in the F ratio in Eq. (4.2). We shall be more specific than Shillington by discussing the power of the test in terms of parameters of the doubly noncentral F distribution.

We now turn to defining Shillington's test statistic in matrix notation. Shillington's F ratio takes the form (see Eq. (4.2))

$$F = \frac{SSE_B / (g - p)}{SSE_W / (N - g - r)}$$

$$= \frac{MSE_B}{MSE_W}$$

where SSE_B is the residual sum of squares from a weighted least squares regression analysis in which $\bar{y}_C$ is regressed on X_C , g is the number of cells of near neighbors, p is the number of terms in the fitted model, and r is the rank of X_W . The quantity SSE_B can be written as the quadratic form

$$SSE_B = \bar{y}_C' [G_0^{-1} - G_0^{-1} X_C (X_C' G_0^{-1} X_C)^{-1} X_C' G_0^{-1}] \bar{y}_C, \quad (4.5)$$

(see Graybill, 1976, p. 329; also see Draper and Smith, 1981, p. 109). The quantity SSE_W is defined as

$$SSE_W = \underline{W}' [I_N - X_W (X_W' X_W)^{-} X_W'] \underline{W}, \quad (4.6)$$

where $(X_W' X_W)^{-}$ is any generalized inverse of $(X_W' X_W)$. [A matrix A^{-} is defined as a generalized inverse of the matrix A if $AA^{-}A = A$.] The quadratic form SSE_W is the residual sum of squares from an ordinary least squares regression analysis in which $\underline{W}$ is regressed on X_W .

In the next two sections we shall discuss the development of the numerator and denominator, MSE_B and MSE_W , respectively, of the F ratio given in Eq. (4.2). We then suggest an alternative representation for MSE_W which relies on a generalization of weighted least squares. This alternative representation for MSE_W will be shown to be equivalent to Shillington's expression for MSE_W .

4.3.1 Development of MSE_B

The quantity $MSE_B = SSE_B/(g - p)$ is the numerator of the F ratio in Eq. (4.2). As mentioned in Section 4.3, the quantity SSE_B is the residual sum of squares from a weighted least squares regression analysis in which $\bar{Y}_C$ is regressed on X_C . The weighting is appropriate because $\text{var}(\bar{Y}_C) = \sigma^2 G_0$ not only when the fitted model is adequate (under model (4.3)), but also when the fitted model suffers from lack of fit (under model (4.4)). In order to further explain the $(\bar{Y}_C, X_C)$ system, we define the matrix M as

$$M = \text{diag}[(1/n_1)\underline{1}_1', \dots, (1/n_g)\underline{1}_g']$$

where $\underline{1}_j$ is an $n_j \times 1$ vector of ones, $j = 1, 2, \dots, g$. Then the $(\bar{Y}_C, X_C)$ system can be derived as a linear transformation of the (Y, X) system. That is, application of the transformation matrix M yields the following equalities

$$\bar{Y}_C = MY,$$

$$X_C = MX,$$

and

$$X_{2C} = MX_2.$$

From this it follows that $\text{var}(\bar{Y}_C) = M \text{var}(Y)M' = \sigma^2 MM' = \sigma^2 G_0$, since $MM' = G_0$. Under the hypothesized model of

Eq. (4.3), $E(\bar{Y}_C) = ME(\underline{Y}) = M\underline{X}\underline{\beta}_1 = \underline{X}_C\underline{\beta}_1$, whereas under the model of Eq. (4.4), $E(\bar{Y}_C) = M(\underline{X}\underline{\beta}_1 + \underline{X}_2\underline{\beta}_2) = \underline{X}_C\underline{\beta}_1 + \underline{X}_{2C}\underline{\beta}_2$.

We now consider the distribution of the random quantity SSE_B/σ^2 . It can be shown (Theorem 2, Searle, 1971, p. 57) that under the model of Eq. (4.3), SSE_B/σ^2 possesses a central chi-square distribution with $g - p$ degrees of freedom, but that under the model of Eq. (4.4), SSE_B/σ^2 possesses a noncentral chi-square distribution with $g - p$ degrees of freedom and noncentrality parameter Π_1 , where

$$\Pi_1 = (1/2\sigma^2)\underline{\beta}_2'\underline{X}_{2C}'[G_0^{-1} - G_0^{-1}\underline{X}_C(\underline{X}_C'G_0^{-1}\underline{X}_C)^{-1}\underline{X}_C'G_0^{-1}]\underline{X}_{2C}\underline{\beta}_2. \quad (4.7)$$

Here we point out that the noncentrality parameter for SSE_B/σ^2 given by Shillington (1979) is not correct and should be written as in Eq. (4.7).

Finally we note that SSE_B is equivalent to the usual lack of fit sum of squares, SS_{LOF} , where $SS_{LOF}/(g - p) = MS_{LOF}$ is the numerator of the F ratio in Eq. (4.1), when all observations in each cell are true replicates rather than near neighbors. Shillington pointed out this fact, but did not give a proof. We offer a proof in Appendix 7.

4.3.2 Development of MSE_W

The quantity $MSE_W = SSE_W/(N - g - r)$, where r denotes the rank of $\underline{X}_W$, is the denominator of the F ratio in Eq. (4.2). As mentioned in Section 4.3, the quantity SSE_W is the residual sum of squares from an ordinary least squares regression analysis in which $\underline{W}$ is regressed on $\underline{X}_W$. Using

Theorem 2 (Searle, 1971, p. 57) and noting that

$\underline{W} = \Sigma_0 \underline{Y}$ and $\Sigma_0' \Sigma_0 = \Sigma_0$, it can be shown that under the hypothesized model of Eq. (4.3), SSE_W/σ^2 possesses a central chi-square distribution with $N - g - r$ degrees of freedom, but under the model of Eq. (4.4), SSE_W/σ^2 possesses a noncentral chi-square distribution with $N - g - r$ degrees of freedom and noncentrality parameter Π_2 , where

$$\Pi_2 = (1/2\sigma^2) \beta_{-2}' X_{2W}' [I_N - X_W (X_W' X_W)^{-1} X_W'] X_{2W} \beta_{-2}. \quad (4.8)$$

Shillington (1979) points out that SSE_W reduces to the usual pure error sum of squares, SSE_{pure} , when all cells contain true replicates. This is easily seen by using the fact that $X_W = 0$ when all cells are composed entirely of true replicates so that $SSE_W = \underline{W}' \underline{W} = \underline{Y}' \Sigma_0' \Sigma_0 \underline{Y} = \underline{Y}' \Sigma_0 \underline{Y} = SSE_{\text{pure}}$.

We saw in Section 4.3.1 that the $(\bar{\underline{Y}}_C, X_C)$ system is derived as a linear transformation of the $(\underline{Y}, X)$ system. Similarly, the $(\underline{W}, X_W)$ system can be derived by applying the transformation matrix Σ_0 . Thus $\underline{W} = \Sigma_0 \underline{Y}$, $X_W = \Sigma_0 X$, and $X_{2W} = \Sigma_0 X_2$. It follows that $E(\underline{W}) = \Sigma_0 E(\underline{Y}) = \Sigma_0 X_{-1}^{\beta} = X_{W-1}^{\beta}$, under the model of Eq. (4.3), and $E(\underline{W}) = \Sigma_0 (X_{-1}^{\beta} + X_{2-2}^{\beta}) = X_{W-2}^{\beta} + X_{2W-2}^{\beta}$ under the model of Eq. (4.4). Furthermore, $\text{var}(\underline{W}) = \Sigma_0 \text{var}(\underline{Y}) \Sigma_0' = \sigma^2 \Sigma_0' \Sigma_0 = \sigma^2 \Sigma_0$, since Σ_0 is symmetric and idempotent.

Since the variance-covariance matrix of $\underline{W}$ is not equal to aI_N , for some positive constant a , SSE_W should have been derived as the residual sum of squares from a weighted least

squares regression analysis of $\underline{W}$ on X_W rather than from the ordinary least squares regression of $\underline{W}$ on X_W that Shillington (1979) suggested. We shall use the weighted least squares regression of $\underline{W}$ on X_W in an attempt to replace SSE_W in the F ratio in Eq. (4.2) by an expression we will call $SSE_W(\text{weighted})$. We later show that SSE_W and $SSE_W(\text{weighted})$ are equivalent.

The variance-covariance matrix of $\underline{W}$, which is $\sigma^2 \Sigma_0$, is of rank $N - g$ and is thus singular. Therefore the residual sum of squares from a weighted least squares regression analysis of $\underline{W}$ on X_W which is

$$\underline{W}' [\Sigma_0^{-1} - \Sigma_0^{-1} X_W (X_W' \Sigma_0^{-1} X_W)^{-1} X_W' \Sigma_0^{-1}] \underline{W}$$

cannot be used as an expression for $SSE_W(\text{weighted})$, since Σ_0^{-1} does not exist. The problem of performing a weighted least squares regression analysis when the variance-covariance matrix of $\underline{W}$ is singular is considered in the next section.

4.4 Development of $SSE_W(\text{weighted})$

If the variance-covariance matrix of $\underline{W}$, $\sigma^2 \Sigma_0$, is nonsingular then the weighted least squares formula for $SSE_W(\text{weighted})$ is

$$SSE_W(\text{weighted}) = (\underline{W} - X_W \hat{\beta}_1)' \Sigma_0^{-1} (\underline{W} - X_W \hat{\beta}_1)$$

where $\hat{\beta}_1$ is a solution to $X_W' \Sigma_0^{-1} X_W \hat{\beta}_1 = X_W' \Sigma_0^{-1} \underline{W}$ and can be written as $\hat{\beta}_1 = (X_W' \Sigma_0^{-1} X_W)^{-1} X_W' \Sigma_0^{-1} \underline{W}$. The quantity $SSE_W(\text{weighted})$ divided by the appropriate degrees of freedom provides an unbiased estimate of σ^2 under the model of Eq. (4.3).

However, since Σ_0 is singular, the weighted regression formula above cannot be used to calculate $SSE_W(\text{weighted})$. C. R. Rao (1971, 1972, and 1973) suggests an analog of weighted least squares for the case of a singular variance-covariance matrix. Rao suggests the existence of a matrix H such that $\hat{\beta}_1$ is a stationary vector value of $(\underline{W} - X_W \hat{\beta}_1)' H (\underline{W} - X_W \hat{\beta}_1)$ in which case σ^2 may be estimated using

$$\hat{\sigma}^2 = (\underline{W} - X_W \hat{\beta}_1)' H (\underline{W} - X_W \hat{\beta}_1) / (N - g - r)$$

where $(N - g - r) = \text{rank}(\Sigma_0 : X_W) - \text{rank}(X_W)$. The rank of the matrix $(\Sigma_0 : X_W)$ is equal to $N - g$ because $X_W = \Sigma_0 X_W$ so that the columns of X_W are spanned by the columns of Σ_0 , thus, $\text{rank}(\Sigma_0 : X_W) = \text{rank}(\Sigma_0) = N - g$.

One form of the matrix H is defined (Rao, 1971 and 1972) as

$$H = [\Sigma_0 + c^2 X_W X_W']^{-1} \quad (4.9)$$

where c is an arbitrary nonzero constant, so that with the model of Eq. (4.3), $\hat{\sigma}^2 = (\underline{W} - X_W \hat{\beta}_1)' H (\underline{W} - X_W \hat{\beta}_1) / (N - g - r)$

is an unbiased estimator for σ^2 . Thus a stationary vector value of $(\underline{W} - \underline{X}_W \underline{\beta}_1)' H (\underline{W} - \underline{X}_W \underline{\beta}_1)$ is given by $\hat{\underline{\beta}}_1 = (\underline{X}_W' H \underline{X}_W)^{-} \underline{X}_W' H \underline{W}$. Rao indicates (1972, p.3) that $\hat{\sigma}^2$ is invariant to the choices of the generalized inverses involved in $\hat{\sigma}^2$.

Rao's proofs for obtaining an unbiased estimator $\hat{\sigma}^2$ for σ^2 are not given in detail. Therefore we shall state and prove the following theorem which will be used to develop an expression for $SSE_W(\text{weighted})$. The notation A^- will be used to denote any generalized inverse of a matrix A , such that $AA^-A = A$.

Theorem 4.1. Let $\underline{Y} \sim (\underline{X}\underline{\beta}, \sigma^2 G)$, where G is singular, then $\hat{\sigma}^2 = f^{-1}(\underline{Y} - \underline{X}\hat{\underline{\beta}})' T^{-} (\underline{Y} - \underline{X}\hat{\underline{\beta}})$

- (i) is an unbiased estimate of σ^2 , and
- (ii) is unique with probability one, and
- (iii) is a scalar multiple of a central chi-square variable with f degrees of freedom of the form $(\sigma^2/f)\chi_f^2$ if $\underline{Y}$ has the multivariate normal distribution.

The vector $\underline{Y}$ is of order $N \times 1$, $\underline{\beta}$ is a $p \times 1$ vector of unknown regression coefficients, X is an $N \times p$ matrix of known constants, G is an $N \times N$ positive semi-definite matrix of known constants, $T = G + XX'$, $\hat{\underline{\beta}}$ is any solution to $X'T^{-}X\underline{\beta} = X'T^{-}\underline{Y}$, that is, $\hat{\underline{\beta}} = (X'T^{-}X)^{-}X'T^{-}\underline{Y}$, and $f = \text{rank}(G:X) - \text{rank}(X)$.

The proofs of parts (i), (ii) and (iii) of Theorem 4.1 are given in Appendices 9, 10, and 11, respectively. Lemma 4.1 which is used in the proof of Theorem 4.1 is stated and proved in Appendix 8.

The results of Theorem 4.1 can now be applied to our problem of finding an expression for $SSE_W(\text{weighted})$. We define $SSE_W(\text{weighted})$ as

$$SSE_W(\text{weighted}) = \underline{W}' [T_0^- - T_0^- X_W (X_W' T_0^- X_W)^{-1} X_W' T_0^-] \underline{W} \quad (4.10)$$

where $T_0 = \Sigma_0 + X_W X_W'$. Writing $SSE_W(\text{weighted})$ in Eq. (4.10) as $SSE_W(\text{weighted}) = (\underline{W} - X_W \hat{\beta}_1)' T_0^- (\underline{W} - X_W \hat{\beta}_1)$, from Theorem 4.1 we see that if the true model is of the form in Eq. (4.3) then $SSE_W(\text{weighted})/\sigma^2$ has a central chi-square distribution with $f = \text{rank}(\Sigma_0 : X_W) - \text{rank}(X_W) = N - g - r$ degrees of freedom. However, if the true model is of the form in Eq. (4.4), then $SSE_W(\text{weighted})/\sigma^2$ has a noncentral chi-square distribution with $N - g - r$ degrees of freedom and noncentrality parameter Π_2^* , where

$$\Pi_2^* = (1/2\sigma^2) \beta_2' X_{2W}' [T_0^- - T_0^- X_W (X_W' T_0^- X_W)^{-1} X_W' T_0^-] X_{2W} \beta_2.$$

The distribution of $SSE_W(\text{weighted})/\sigma^2$ under model (4.4) is verified by the following theorem.

Theorem 4.2. Let $\underline{Y} \sim N_N(X\underline{\beta} + X_2\underline{\beta}_2, \sigma^2 G)$, G singular, then $\hat{\sigma}^2/\sigma^2 \sim \chi_{f,\lambda}^2$ where $\hat{\sigma}^2 = f^{-1}(\underline{Y} - X\underline{\hat{\beta}})'T^-(\underline{Y} - X\underline{\hat{\beta}})$, $f = \text{rank}(G:X) - \text{rank}(X)$, $T = G + XX'$, and $\lambda = (1/2\sigma^2)\underline{\beta}_2'X_2'[T^- - T^-X(X'T^-X)^-X'T^-]X_2\underline{\beta}_2$.

The proof of Theorem 4.2 is given in Appendix 12.

4.5 Equivalence of SSE_W and $SSE_W(\text{weighted})$

In this section we shall show that our expression for $SSE_W(\text{weighted})$ in Eq. (4.10) is equal to Shillington's unweighted SSE_W in Eq. (4.6). Thus the complex calculations required for evaluating $SSE_W(\text{weighted})$ can be avoided by calculating the simpler form SSE_W .

Zyskind (1967) investigated conditions under which ordinary least squares estimators are BLUE (best linear unbiased estimators) even though $\underline{Y}$ in the model $\underline{Y} = X\underline{\beta} + \underline{\varepsilon}$, where $E(\underline{\varepsilon}) = \underline{0}$, does not have variance-covariance matrix equal to $\sigma^2 I$. Zyskind assumes that $\text{var}(\underline{Y}) = \sigma^2 V$, where V is non-negative and possibly singular, and then states and proves the following necessary and sufficient condition for ordinary least squares estimators to be BLUE.

Theorem 1 (Zyskind, 1967). A necessary and sufficient condition for all simple least squares linear estimators to be also best linear unbiased estimators of the

corresponding estimable parametric function $\lambda'\beta$ in the linear model

$$\underline{Y} = X\underline{\beta} + \underline{\varepsilon}, \quad E(\underline{\varepsilon}) = \underline{0}, \quad E(\underline{\varepsilon}\underline{\varepsilon}') = \sigma^2 V,$$

where V is a symmetric non-negative matrix and X is of rank r , is that there exist a subset of r orthogonal eigenvectors of V which forms a basis for the column space of the matrix X .

In a second theorem, Zyskind (1967) gives several other necessary and sufficient conditions for ordinary least squares estimators to be BLUE. These conditions are shown to be equivalent to the condition in Theorem 1 (Zyskind, 1967). The fifth of these conditions in Zyskind's second theorem is that $VP = PV$, where $P = X(X'X)^{-1}X'$.

Applying condition 5 of Theorem 2 (Zyskind, 1967) to our problem of regressing $\underline{W}$ on X_W we have

$$V = \Sigma_0$$

and

$$P = X_W(X_W'X_W)^{-1}X_W'$$

and therefore

$$\begin{aligned}
 VP &= \Sigma_0 X_W (X_W' X_W)^{-1} X_W' \\
 &= X_W (X_W' X_W)^{-1} X_W' ,
 \end{aligned}$$

since $\Sigma_0 X_W = \Sigma_0 \Sigma_0 X = \Sigma_0 X = X_W$. It follows that

$$\begin{aligned}
 VP &= X_W (X_W' X_W)^{-1} X_W' \Sigma_0 \\
 &= PV.
 \end{aligned}$$

Therefore by Theorem 2 (Zyskind, 1967) the ordinary least squares solutions from regressing $\underline{W}$ on X_W are BLUE estimators, and thus are equivalent to the solutions obtained from weighted least squares. We conclude therefore that $SSE_W = SSE_W(\text{weighted})$.

4.6 The Test Statistic

As stated in Section 4.1, Shillington (1979) proposed that a fitted model be tested for lack of fit by using the F ratio

$$F = \frac{MSB_B}{MSE_W}$$

given in Eq. (4.2). In this section we shall verify that Shillington's F ratio possesses a central F distribution when the true model is of the form in Eq. (4.3), and possesses a doubly noncentral F distribution when the true

model is of the form in Eq. (4.4). This information on the distribution of the F ratio will be needed in Section 4.7, where the power of the test is discussed. Additionally, we shall give the form of the expected values of both the numerator, MSE_B , and the denominator, MSE_W , of the F ratio in Eq. (4.2). These expected values will aid us in developing a procedure for calculating the power of the test, since they will be used to determine whether the test is upper tailed or lower tailed.

In developing the distribution of the F ratio in Eq. (4.2), we shall show that SSE_B/σ^2 and SSE_W/σ^2 are statistically independent. In this pursuit, let us use the expression for SSE_B in Eq. (4.5) and the fact that $\bar{Y}_C = M\bar{Y}$ to express SSE_B as

$$SSE_B = \bar{Y}'M'[G_0^{-1} - G_0^{-1}X_C(X_C'G_0^{-1}X_C)^{-1}X_C'G_0^{-1}]M\bar{Y}.$$

Also, using the expression for SSE_W in Eq. (4.6) (which is allowed because we showed in Section 4.5 that the correct form, $SSE_W(\text{weighted})$, is equal to SSE_W) and using the fact that $\underline{W} = \Sigma_0 \underline{Y}$, we can express SSE_W as

$$SSE_W = \bar{Y}'\Sigma_0[I_N - X_W(X_W'X_W)^{-1}X_W']\Sigma_0\bar{Y}.$$

By Theorem 4 (Searle, 1971, p.59), to show that SSE_B/σ^2 and SSE_W/σ^2 are statistically independent, it suffices to show that the matrix product

$$[M'\{G_0^{-1} - G_0^{-1}X_C(X_C'G_0^{-1}X_C)^{-1}X_C'G_0^{-1}\}M][\Sigma_0\{I_N - X_W(X_W'X_W)^{-1}X_W'\}\Sigma_0]$$

is equal to the zero matrix. This is seen to be true since $M\Sigma_0 = 0$, and therefore SSE_B/σ^2 and SSE_W/σ^2 are independent.

When the fitted model and the true model are both of the form in Eq. (4.3), then the F ratio in Eq. (4.2) possesses a central F distribution with $g - p$ and $N - g - r$ degrees of freedom in the numerator and denominator, respectively. Furthermore, the numerator, MSE_B , of the F ratio in Eq. (4.2) has expectation equal to

$$\begin{aligned} E(MSE_B) &= [\sigma^2/(g - p)]E\chi_{g-p}^2 \\ &= \sigma^2. \end{aligned}$$

Similarly, under model (4.3), the expected value of the denominator, MSE_W , of the F ratio has expectation equal to

$$\begin{aligned} E(MSE_W) &= [\sigma^2/(N - g - r)]E\chi_{N-g-r}^2 \\ &= \sigma^2. \end{aligned}$$

When the fitted model suffers from lack of fit and the true model is given by Eq. (4.4), the F ratio in Eq. (4.2) is a ratio of two statistically independent noncentral chi-square random variables, each divided by its respective

degrees of freedom. Thus the F ratio in Eq. (4.2) possesses a doubly noncentral F distribution with $g - p$ and $N - g - r$ degrees of freedom and noncentrality parameters Π_1 and Π_2 , where Π_1 and Π_2 were given in Eqs. (4.7) and (4.8), respectively. The expected value of the numerator, MSE_B , of the F ratio can be written as

$$E(MSE_B) = [\sigma^2/(g - p)] E\chi_{g-p, \Pi_1}^2$$

$$= \sigma^2 + \beta_2' C_1 \beta_2 / (g - p)$$

where

$$C_1 = X_2' M' [G_0^{-1} - G_0^{-1} X_C (X_C' G_0^{-1} X_C)^{-1} X_C' G_0^{-1}] M X_2. \quad (4.11)$$

Similarly under model (4.4), the expected value of the denominator, MSE_W , of the F ratio can be written as

$$E(MSE_W) = [\sigma^2/(N - g - r)] E\chi_{N-g-r, \Pi_2}^2$$

$$= \sigma^2 + \beta_2' C_2 \beta_2 / (N - g - r)$$

where

$$C_2 = X_2' \Sigma_0 [I_N - X_W (X_W' X_W)^{-1} X_W'] \Sigma_0 X_2. \quad (4.12)$$

4.7 The Testing Procedure and its Power

As discussed in Section 4.1, Shillington (1979) suggested that lack of fit of the fitted model be inferred when the value of the F statistic in Eq. (4.2) exceeds $F_{\alpha; g-p, N-g-r}$. The test, however, is not always upper tailed, and in fact can be lower tailed. The test is considered lower tailed when, because of lack of fit, the expected value of the numerator of the F ratio is less than the expected value of the denominator of the F ratio.

We suggest that lack of fit be tested with an upper tailed test using the F ratio $F = \text{MSE}_B / \text{MSE}_W$ when the matrix D, which is defined as

$$D = \frac{C_1}{g - p} - \frac{C_2}{N - g - r} \quad (4.13)$$

is found to be positive definite (which can only occur when C_1 is positive definite, by Theorem 3.1 in Appendix 4). The matrices C_1 and C_2 in Eq. (4.13) are defined in Eqs. (4.11) and (4.12), respectively. An upper tailed test is appropriate when the matrix D is positive definite because no matter what the value of β_2 is, the expected value of the numerator, MSE_B , of the F ratio will be greater than the expected value of the denominator, MSE_W , of the F ratio. However, there may be cases where D is negative definite (which can only occur when C_2 is positive definite), and in this case lack of fit should be tested with a lower tailed

rejection region. If D is indefinite, then the F test for lack of fit may be upper tailed, lower tailed, or lack of fit may not be testable depending upon the value of β_2 .

In those cases where D is indefinite, it is helpful to write the quantity $\beta_2' D \beta_2$, which represents the difference between the expected value of the numerator and the expected value of the denominator of the F ratio, $F = \text{MSE}_B / \text{MSE}_W$, as

$$\begin{aligned}\beta_2' D \beta_2 &= \beta_2' [U_1 : U_2 : U_3] \text{diag}[\Gamma_1, \Gamma_2 = 0, \Gamma_3] [U_1 : U_2 : U_3]' \beta_2 \\ &= \beta_2' U_1 \Gamma_1 U_1' \beta_2 + \beta_2' U_3 \Gamma_3 U_3' \beta_2,\end{aligned}$$

where U_1 , U_2 , and U_3 are matrices whose columns are orthonormal eigenvectors of D , and Γ_1 , Γ_2 , and Γ_3 are diagonal matrices whose elements are the positive, zero, and negative eigenvalues of D , respectively. Lack of fit is testable with an upper tailed test if β_2 is in the column space of $[U_1 : U_2]$, but not entirely in the column space of U_2 , since then $\beta_2' D \beta_2$ is positive. Similarly, lack of fit is testable with a lower tailed test if β_2 is in the column space of $[U_2 : U_3]$, but not entirely in the column space of U_2 , since then $\beta_2' D \beta_2$ is negative. If β_2 is in the column space of U_2 , then lack of fit cannot be tested, since $\beta_2' D \beta_2$ would equal zero.

We define $\underline{\beta}_2$ to be in the column space of U_2 if the matrix $L_1' L_1$ has a zero eigenvalue, where $L_1 = [\underline{\beta}_2 : U_2]$. Similarly, letting $L_2 = [\underline{\beta}_2 : U_1 : U_2]$ and $L_3 = [\underline{\beta}_2 : U_2 : U_3]$, $\underline{\beta}_2$ is in the column space of $[U_1 : U_2]$ if $L_2' L_2$ has a zero eigenvalue, and $\underline{\beta}_2$ is in the column space of $[U_2 : U_3]$ if $L_3' L_3$ has a zero eigenvalue.

When D is positive definite or D is indefinite but $\underline{\beta}_2$ is upper tailed testable, then the F test for lack of fit which makes use of the F ratio $F = \text{MSE}_B / \text{MSE}_W$ is a test of the hypotheses (see Theorem 3.2, Appendix 5)

$$H_0: \pi_1 = \pi_2 = 0$$

$$H_a: \pi_1 / (g - p) - \pi_2 / (N - g - r) > 0.$$

When D is negative definite or D is indefinite but $\underline{\beta}_2$ is lower tailed testable, then the F test tests

$$H_0: \pi_1 = \pi_2 = 0$$

$$H_a: \pi_1 / (g - p) - \pi_2 / (N - g - r) < 0.$$

In the case where D is indefinite and $\underline{\beta}_2$ is in the column space of U_2 , then no hypotheses concerning lack of fit of the fitted model can be tested.

When the test is upper tailed, the power of the F test for lack of fit is defined as

$$\text{Power} = P\{F''_{g-p, N-g-r; \Pi_1, \Pi_2} > F_{\alpha; g-p, N-g-r}\} \quad (4.14)$$

where $F_{\alpha; g-p, N-g-r}$ is the upper 100α percentage point of the central F distribution with $g - p$ and $N - g - r$ degrees of freedom. In the case of a lower tailed test, the power of the test is defined as

$$\text{Power} = P\{F''_{g-p, N-g-r; \Pi_1, \Pi_2} < F_{1-\alpha; g-p, N-g-r}\}. \quad (4.15)$$

4.8 Selection of Near Neighbor Groupings

In the preceding sections of this chapter, we have discussed a near neighbor procedure which uses the F ratio $F = \text{MSE}_B / \text{MSE}_W$ to test a fitted model for lack of fit. In this section we shall investigate the effect that different groupings of response observations into near neighbor cells have on the testing procedure and its power. From equations (4.14) and (4.15) in the previous section it is evident that the power of the F test for lack of fit, which makes use of the F ratio $F = \text{MSE}_B / \text{MSE}_W$, depends on the values of the numerator and denominator noncentrality parameters, Π_1 and Π_2 . Assuming the numerator and denominator degrees of freedom are fixed, and the test is upper tailed, then the power is an increasing function of increasing values of Π_1 and is a decreasing function of increasing values of Π_2 (see Appendix 1). When the test is lower tailed, the

power is an increasing function of Π_2 and is a decreasing function of Π_1 .

Since both the numerator and denominator noncentrality parameters, Π_1 and Π_2 , are functions of the groupings of the data points into near neighbor cells (as are the numerator and denominator degrees of freedom), we would like to investigate the effect of the number and composition of cells on the power of the F test. Intuitively, it would seem that homogeneous near neighbor cells would minimize the bias inherent in estimating σ^2 with MSE_W , and thus would minimize Π_2 and maximize the power of an upper tailed test. However, any grouping of the data points would also influence the numerator noncentrality parameter and the numerator and denominator degrees of freedom. Therefore while a grouping of data points into homogeneous cells might decrease Π_2 and thus apparently increase the power of an upper tailed test, the result of the grouping on the power also depends on how the degrees of freedom, $g - p$ and $N - g - r$, and the numerator noncentrality parameter, Π_1 , are affected.

We will attempt to find homogeneous cells of near neighbor points by using an iterative partitioning clustering algorithm. Two examples will be presented. The first example makes use of the stack loss data presented by Daniel and Wood (1971) and later analyzed by Shillington (1979). The second example involves data from a mixture experiment discussed by Piepel (1981).

Our objective is to investigate the effect of forming homogeneous cells of near neighbors on the F test for lack of fit which makes use of the test statistic $F = \text{MSE}_B / \text{MSE}_W$. It is hoped that such homogeneous groupings will increase the power of the F test (assuming that the rejection region is upper tailed) by decreasing Π_2 for a fixed number of near neighbor cells (and thus fixed values for the degrees of freedom). The effect that homogeneous grouping has on Π_1 is not clear, but is of interest, since the magnitude of Π_1 also affects the power of the test. Additionally, we will vary the number of cells of near neighbors in an attempt to determine how this affects the power of the test, since the number of cells affects both the noncentrality parameters and the degrees of freedom.

The algorithm used for grouping the data points into homogeneous near neighbor cells can be described as an iterative partitioning type of cluster analysis. The computations involved were accomplished using the RELOC procedure available in the CLUSTAN 1C computer package (Wishart, 1975). All computations were performed using data points whose coordinates were standardized by subtracting off sample means and dividing by sample standard deviations.

Initially, k clusters (near neighbor cells) of the N data points in the factor space are arbitrarily defined. Then the Euclidean distance between each point and the centroid (average vector value) of each of the k clusters is determined. If a point is found to be closer to the

centroid of one of the other $k - 1$ clusters than to the centroid of the cluster in which it is currently classified, then the point is reclassified into that nearest cluster (cell). The centroids of the k clusters are then recalculated, taking into account any reclassified point. The entire set of N points is scanned repeatedly in this manner until no reclassification occurs. This method of assigning points to clusters will be referred to as iterative relocation.

In the second stage of the algorithm, two of the k clusters arrived at by the iterative relocation procedure are fused, resulting in $k - 1$ clusters. The two clusters to be fused are selected as those which when fused produce the $k - 1$ clusters with minimum "error sum of squares." The error sum of squares is defined as the sum of squared Euclidean distances between every point and the centroid point of the cluster to which it belongs. After $k - 1$ clusters are determined using the error sum of squares criterion, iterative relocation is applied to the $k - 1$ groups in an effort to improve the clusterings. This alternation of fusion and iterative relocation continues for $k - 2$ clusters, $k - 3$ clusters, ..., 2 clusters, or until a specified minimum number of clusters is reached. The question of determining an "optimal" number of clusters is not addressed by this procedure.

4.8.1 Example 1--Stack Loss Data

The first example we investigate is the 21 observation stack loss data of Daniel and Wood (1971), which was analyzed by Shillington (1979). The data (see Table 8) consist of the values of three factors, x_1 (air flow), x_2 (cooling water inlet temperature), and x_3 (acid concentration) along with the values of a response variable, Y (stack loss). A first order regression equation of the form

$$E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

is fitted using 17 of the original 21 observations (Shillington discarded 4 of the original 21 observations as outliers). We assume the true model to have the form

$$E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{33} x_3^2$$

and thus contains $p_2 = 3$ terms in addition to the $p = 4$ terms in the fitted model. We wish to investigate the capability of the F test

$$F = \frac{MSE_B}{MSE_W}$$

in detecting lack of fit of the fitted model.

We first consider the use of the F statistic with the six cells of near neighbors used by Shillington (1979), which is the same near neighbor grouping suggested by Daniel

and Wood (1971). This 6 cell grouping (see Table 8 under the column heading "6**") is found to yield a matrix $D = [C_1/(g - p)] - [C_2/(N - g - r)]$ which is indefinite, since the eigenvalues of D have the values 12110, -7, and -1415 (see Table 9). Thus the test is not upper tailed, since D must be positive definite for an upper tailed test to exist for all values of $\underline{\beta}_2$, where in this example, $\underline{\beta}_2 = (\beta_{11}, \beta_{22}, \beta_{33})'$.

When the 6 cell grouping of near neighbors generated by the iterative partitioning clustering algorithm (see Table 8 under the column heading "6") is used, the values of the eigenvalues of D are 49090, 379, and -43, and the test is still not upper tailed for all values of $\underline{\beta}_2$.

We then use the iterative partitioning clustering algorithm to determine homogeneous cell groupings for 5, 7, 8, 9, 10, 11, and 12 cells. The matrix D is found to be indefinite for the groupings into 5, 7, or 8 cells, but D is positive definite for 9, 10, 11, or 12 cells of near neighbor groupings. Thus, no matter what the value of $\underline{\beta}_2$, lack of fit can be tested with an upper tailed test using the 9, 10, 11, or 12 cell groupings of near neighbors.

The value of $F = \text{MSE}_B/\text{MSE}_W$ was calculated using the matrix procedure from the 1979 version of SAS. None of the near neighbor groupings provided evidence of lack of fit, and thus we cannot conclude that there is lack of fit when the fitted model is $E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$

Table 8. Near Neighbor Cells for Stack Loss Data.

Observation Number	Air Flow x ₁	Water Temp. x ₂	Acid Concn. x ₃	Stack Loss y	Membership in Near Neighbor Cells for Five to Twelve Cells											
					5	6**	6	7	8	9	10	11	12			
1*	80	27	89	42	--	--	--	--	--	--	--	--	--	--	--	
2	80	27	88	37	1	1	1	1	1	1	1	1	1	1	1	
3*	75	25	90	37	--	--	--	--	--	--	--	--	--	--	--	
4*	62	24	87	28	--	--	--	--	--	--	--	--	--	--	--	
5	62	22	87	18	2	2	2	2	2	2	2	2	2	2	2	
6	62	23	87	18	2	2	2	2	2	2	2	2	2	2	2	
7	62	24	93	19	2	2	2	3	3	3	3	3	3	3	3	
8	62	24	93	20	2	2	2	3	3	3	3	3	3	3	3	
9	58	23	87	15	2	3	2	2	2	2	2	2	2	2	2	
10	58	18	80	14	3	4	3	4	4	4	4	4	4	4	4	
11	58	18	89	14	4	4	4	5	5	5	5	5	5	5	5	
12	58	17	88	13	4	4	4	5	5	5	5	5	5	5	5	
13	58	18	82	11	3	4	3	4	4	4	4	4	4	4	4	
14	58	19	93	12	4	4	4	5	5	6	6	6	6	6	6	
15	50	18	89	8	4	5	5	6	6	7	7	7	7	7	7	
16	50	18	86	7	4	5	5	6	6	7	7	7	7	7	7	
17	50	19	72	8	5	5	6	7	7	8	8	8	8	8	8	
18	50	19	79	8	5	5	6	7	7	8	8	8	8	8	8	
19	50	20	80	9	5	5	6	7	7	9	9	9	9	9	9	
20	56	20	82	15	3	6	3	4	4	4	4	4	4	4	4	
21*	70	20	91	15	--	--	--	--	--	--	--	--	--	--	--	

*Outliers, not included in analyses.

**This near neighbor cell grouping was proposed by Daniel and Wood (1971) and used by Shillington (1979). All other cell groupings were generated by an iterative partitioning cluster analysis using the CLUSTAN computer package. Numbers in the table refer to cell membership.

Table 9. Effect of Near Neighbor Cells on
F = MSE_B/MSE_W for Stack Loss Data.

Number of Cells (g)	5	6**	7	8	9	10	11	12
Numerator								
d.f. (g - p)	1	2	3	4	5	6	7	8
Denominator								
d.f. (N-g-r)	9	9	7	6	5	4	4	3
Eigenvalues								
of C ₁	65525	24408	98191	99915	102192	102859	103214	103445
	0	622	4082	9098	20054	23317	22394	22642
	0	0	0	17	17	85	106	107
Eigenvalues								
of C ₂	14900	16325	13539	8297	1017	318	318	104
	1478	121	112	50	48	24	24	10
	111	0	39	20	18	0	0	0
Eigenvalues								
of D	69297	12110	49090	33300	25540	17143	14745	12931
	-13	-7	379	1852	4847	3640	3120	2800
	-1592	-1415	-43	-4	-2	8	9	6
Calculated								
value of F	1.14NS	1.08NS	0.96NS	0.91NS	0.83NS	1.73NS	1.49NS	2.52NS
Π ₁	--	--	--	--	--	133	136	136
Π ₂	--	--	--	--	--	1	1	1
Approximate								
Power of Upper								
Tailed Test	--	--	--	--	--	0.98	0.96	0.79

Note: Fitted Model: $E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$

Assumed true model: $E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{33} x_3^2 + \beta_{22} x_2^2 + \beta_{33} x_3^2$

** Daniel and Wood (1971) grouping, other groupings generated by CLUSTAN.

NS Not greater than F_{.05;g-p,N-g-r}.

and the true model contains only pure quadratic terms in addition to the first degree terms.

For the groupings of near neighbors into cells which provide an upper tailed test (9, 10, 11, or 12 cells), the power of the upper tailed test can be approximated if Π_1 and Π_2 can be specified. This approximate power can be calculated using an approximation to the doubly noncentral F distribution given in Johnson and Kotz (1970, p.197), which is described in Appendix 6 of this dissertation. Thus we calculate an approximation for

$$P\{F''_{g-p, N-g-r; \Pi_1, \Pi_2} > F_{\alpha; g-p, N-g-r}\}.$$

In order to compare the power of the upper tailed F test for 9, 10, 11, and 12 cells, we will assume arbitrarily that the true value of the parameter vector $\underline{\beta}_2$ is $\underline{\beta}_2 = (\beta_{11}, \beta_{22}, \beta_{33})' = (.044, .329, -.033)'$ which is arrived at by taking $\underline{\beta}_2 = 10\hat{\underline{\beta}}_2$ where $\hat{\underline{\beta}}_2$ is the least squares estimate of $\underline{\beta}_2$ calculated from the data. Furthermore, taking $\sigma^2 = 1.6$ (since the residual mean square value from fitting the "true" second degree model is MSE = 1.6) we calculate the values for $\Pi_1 = \underline{\beta}_2' C_1 \underline{\beta}_2 / 2\sigma^2$ and $\Pi_2 = \underline{\beta}_2' C_2 \underline{\beta}_2 / 2\sigma^2$ for each of the 9, 10, 11, and 12 cell groupings. The calculated values of Π_1 and Π_2 , as well as the approximate power values for each of the four F tests (calculated using the approximation to F'' from Johnson and Kotz (1970, p.197)) are presented in Table 9. The power is

quite high ($> .96$) for 9, 10, or 11 cell groupings, but drops off to .79 for the 12 cell grouping. This drop in power seems to be due to the effect of having only 3 degrees of freedom in the denominator of the F ratio.

In summary, this example illustrates that the F test for lack of fit that makes use of the statistic $F = \text{MSE}_B / \text{MSE}_W$ is upper tailed only for certain groupings of the design points into near neighbor cells. For the near neighbor cell groupings that provide an upper tailed test, the power is generally high for the values of β_2 and σ^2 that we selected, but decreases slightly as we move from 9 to 10 to 11 cells and decreases more severely as we move from 11 to 12 cells. This more severe decrease in power is due to the decrease to only 3 denominator degrees of freedom.

4.8.2 Example 2--Glass Leaching Data

The second example we investigate is one in which the leachability, Y , of glass is assumed to be a function of the proportions of eleven chemicals of which the glass is composed (Piepel, 1981). A first order Scheffé polynomial model was fitted to the common logarithms of the leachability values, that is, the fitted model is of the form

$$E(\log_{10} Y) = \sum_{i=1}^{11} \beta_i x_i.$$

The experimental design coordinates and the values of the 44 data observations are presented in Table 10. For illustrative purposes, the true model is assumed to contain the $p_2 = 8$ second order cross product terms,

$\beta_{68}x_6x_8$, $\beta_{3,11}x_3x_{11}$, $\beta_{79}x_7x_9$, $\beta_{56}x_5x_6$, $\beta_{7,11}x_7x_{11}$, $\beta_{15}x_1x_5$, $\beta_{6,10}x_6x_{10}$, and $\beta_{59}x_5x_9$ in addition to the $p = 11$ first order terms in the fitted model. The 19-term model is the final fitted model proposed by Piepel and serves as our true model.

Piepel (1981) suggested that the four sets of observations (see Table 10)

- (a) 14 and 15
- (b) 18, 19, and 20
- (c) 25, 26, and 27
- (d) 39, 40, and 41

were intended to constitute four cells of replicate observations for use in estimating pure experimental error. However, the settings of the mixture components were not well controlled, so that each of the four cells contained near neighbors rather than replicates. By defining each of the remaining 33 data points as 33 cells containing one observation each, the 44 data observations are partitioned into 37 cells. If we choose to use the 37 cells to test the fitted model for lack of fit, with the test statistic $F = \text{MSE}_B / \text{MSE}_W$, we find that there are no degrees of freedom for MSE_W so that the F statistic cannot be calculated in this case. However, the F statistic for

Table 10. Glass Leaching Data.

Observation Number	x1 SiO ₂	x2 B ₂ O ₃	x3 Al ₂ O ₃	x4 CaO	x5 MgO	x6 Na ₂ O	x7 ZnO	x8 TiO ₂	x9 Cr ₂ O ₃	x10 Fe ₂ O ₃	x11 NiO	y* Leach Value
1	.448	.063	.143	.000	.000	.165	.059	.069	.000	.029	.025	8.10
2	.544	.063	.000	.138	.000	.154	.000	.071	.030	.000	.000	0.14
3	.565	.120	.000	.000	.000	.153	.061	.070	.031	.000	.000	0.01
4	.581	.061	.012	.000	.079	.108	.061	.070	.000	.028	.000	0.01
5	.432	.126	.000	.098	.079	.158	.000	.070	.000	.009	.028	7.40
6	.486	.064	.000	.140	.000	.159	.063	.000	.028	.031	.029	3.90
7	.429	.060	.145	.139	.013	.152	.062	.000	.000	.000	.000	91.40
8	.429	.124	.144	.138	.000	.105	.060	.000	.000	.000	.000	49.00
9	.443	.120	.000	.099	.080	.101	.061	.070	.000	.026	.000	0.11
10	.447	.114	.148	.012	.077	.150	.000	.000	.027	.025	.000	33.40
11	.427	.121	.145	.010	.078	.161	.000	.000	.029	.029	.000	39.50
12	.550	.060	.146	.000	.076	.111	.000	.030	.000	.000	.027	0.42
13a	.434	.058	.148	.127	.000	.109	.000	.069	.030	.025	.000	9.60
14a	.580	.122	.000	.135	.000	.110	.000	.000	.000	.025	.028	0.20
15	.575	.123	.000	.134	.000	.112	.000	.000	.000	.028	.028	0.24
16	.498	.066	.000	.134	.080	.105	.059	.000	.030	.000	.028	6.20
17b	.435	.118	.145	.000	.031	.102	.060	.051	.031	.000	.027	2.20
18b	.496	.092	.072	.067	.038	.125	.031	.035	.015	.015	.014	0.86
19b	.497	.091	.072	.066	.039	.129	.029	.034	.014	.015	.014	0.80
20	.497	.091	.072	.067	.039	.127	.030	.034	.014	.015	.014	0.86
21	.565	.125	.000	.010	.036	.163	.000	.070	.031	.000	.000	0.20
22	.428	.068	.073	.117	.084	.108	.000	.071	.000	.027	.024	1.47
23	.433	.064	.155	.047	.000	.159	.000	.073	.028	.022	.019	16.40
24c	.550	.141	.000	.000	.000	.156	.063	.079	.000	.011	.000	0.15
25c	.577	.127	.023	.000	.086	.094	.060	.000	.033	.000	.000	0.15
26c	.586	.126	.024	.000	.079	.093	.058	.000	.034	.000	.000	0.08
27	.585	.128	.023	.000	.082	.091	.059	.000	.032	.000	.000	0.15
28	.579	.068	.000	.000	.088	.146	.061	.000	.000	.031	.027	0.18
29	.541	.064	.153	.072	.000	.113	.000	.000	.031	.026	.000	2.04

Table 10--continued.

30	.417	.127	.152	.010	.080	.162	.000	.000	.031	.000	.021	51.80
31	.429	.133	.031	.130	.000	.144	.059	.074	.000	.000	.000	9.04
32	.454	.067	.142	.000	.000	.145	.060	.077	.033	.000	.022	2.09
33	.420	.130	.138	.000	.083	.155	.000	.074	.000	.000	.000	24.70
34	.574	.066	.000	.114	.000	.108	.000	.077	.034	.000	.027	0.02
35	.584	.067	.031	.128	.000	.165	.000	.000	.000	.025	.000	0.08
36	.464	.067	.142	.126	.083	.118	.000	.000	.000	.000	.000	37.20
37	.433	.134	.142	.038	.000	.093	.058	.072	.000	.030	.000	7.39
38 _d	.432	.134	.012	.129	.083	.096	.059	.000	.000	.030	.025	96.60
39 _d	.434	.068	.127	.131	.000	.153	.061	.000	.000	.000	.026	77.80
40 _d	.432	.067	.138	.131	.000	.146	.060	.000	.000	.000	.026	75.60
41	.431	.067	.139	.132	.000	.149	.061	.000	.000	.000	.021	76.40
42	.413	.127	.151	.011	.080	.165	.000	.000	.030	.000	.023	43.30
43	.430	.133	.144	.127	.000	.111	.000	.000	.031	.000	.024	25.40
44	.596	.068	.000	.011	.000	.136	.063	.077	.024	.000	.025	0.07

* Leach values are weight percent loss of glass. Analyses performed on log₁₀(Y).
a,b,c,d Four groups of intended replicates.

lack of fit can be calculated using from 12 to 33 cells of near neighbors, and thus we refer to the iterative partitioning clustering algorithm discussed earlier in this section to generate near neighbor groupings of 15, 20, 25, and 30 cells (see Table 11).

The clusterings of observations into 15, 20, or 25 cells each yields a D matrix that is indefinite (see Table 12), so that the test is not upper tailed. The 30 cell clustering produces a positive definite D matrix, so that the test is upper tailed for all nonzero values of $\underline{\beta}_2$, where $\underline{\beta}_2 = (\beta_{68}, \beta_{3,11}, \beta_{79}, \beta_{56}, \beta_{7,11}, \beta_{15}, \beta_{6,10}, \beta_{59})'$. Taking $\underline{\beta}_2 = (15.141, -112.429, -78.761, -78.275, 87.996, 13.356, -76.948, 34.721)'$, which is the least squares estimate of $\underline{\beta}_2$ from the data, taking $\sigma^2 = .008$ (which is MSE_{pure} with seven degrees of freedom from Piepel's analysis of the data), and using the approximation of Johnson and Kotz (1970, p.197) to approximate

$$P\{F_{19,4;\Pi_1,\Pi_2}'' > F_{.05;19,4}\}$$

where $\Pi_1 = 9.79$, $\Pi_2 = 0.08$ and $F_{.05;19,4} = 5.81$, we find that (using 30 near neighbor cells) a value for the power of the F test is .10. The power increases as the magnitudes of the elements of $\underline{\beta}_2$ are increased, so for example if all the elements of $\underline{\beta}_2$ above are doubled, then $\Pi_1 = 39.16$, $\Pi_2 = 0.32$, and the approximate power is .25. If the elements of $\underline{\beta}_2$ above are each multiplied by 5, then

Table 11. Near Neighbor Cells for Glass Leaching Data.

Observation Number	Membership in Near Neighbor cells*				Observation Number	Membership in Near Neighbor Cells			
	15	20	25	30		15	20	25	30
1	1	1	1	1	23	1	1	1	18
2	2	2	2	2	24	3	3	3	19
3	3	3	3	3	25	15	16	17	20
4	4	4	4	4	26	15	16	17	20
5	5	5	5	5	27	15	16	17	20
6	6	6	6	6	28	4	4	18	21
7	7	7	7	7	29	12	12	12	12
8	8	8	8	8	30	10	17	19	22
9	9	9	9	9	31	8	8	8	23
10	10	10	10	10	32	1	14	15	15
11	10	10	10	10	33	10	18	20	24
12	11	11	11	11	34	2	2	2	2
13	12	12	12	12	35	13	19	21	25
14	13	13	13	13	36	11	11	22	26
15	13	13	13	13	37	8	8	23	27
16	6	6	14	14	38	9	9	9	28
17	14	14	15	15	39	7	7	7	7
18	11	15	16	16	40	7	7	7	7
19	11	15	16	16	41	7	7	7	7
20	11	15	16	16	42	10	17	19	22
21	3	3	3	3	43	14	20	24	29
22	5	5	5	17	44	2	2	25	30

*Cell groupings generated by an iterative partitioning cluster analysis using the CLUSTAN computer package. Numbers in the table refer to cell membership.

Table 12. Effect of Near Neighbor Cells on
F = MSE_B/MSE_W for Glass Leaching Data.

Number of Cells (g)	15	20	25	30
Numerator d.f. (g - p)	4	9	14	19
Denominator d.f. (N-g-r)	19	14	9	4
Eigenvalues of C ₁				
	2.92×10 ⁻⁵	6.57×10 ⁻⁵	1.03×10 ⁻⁴	1.27×10 ⁻⁴
	1.92×10 ⁻⁵	2.56×10 ⁻⁵	2.84×10 ⁻⁵	3.24×10 ⁻⁵
	3.13×10 ⁻⁶	1.07×10 ⁻⁵	1.74×10 ⁻⁵	2.37×10 ⁻⁵
	6.60×10 ⁻⁷	4.34×10 ⁻⁶	6.45×10 ⁻⁶	7.38×10 ⁻⁶
	0	3.02×10 ⁻⁶	4.86×10 ⁻⁶	5.51×10 ⁻⁶
	0	1.07×10 ⁻⁶	1.23×10 ⁻⁶	1.88×10 ⁻⁶
	0	1.10×10 ⁻⁷	9.00×10 ⁻⁷	1.44×10 ⁻⁶
	0	4.76×10 ⁻⁸	3.33×10 ⁻⁷	7.97×10 ⁻⁷
Eigenvalues of C ₂				
	1.60×10 ⁻⁵	1.05×10 ⁻⁵	6.69×10 ⁻⁶	5.01×10 ⁻⁷
	1.28×10 ⁻⁵	5.03×10 ⁻⁶	5.06×10 ⁻⁷	5.89×10 ⁻⁸
	7.65×10 ⁻⁶	2.64×10 ⁻⁶	1.51×10 ⁻⁷	3.22×10 ⁻⁹
	1.84×10 ⁻⁶	8.10×10 ⁻⁷	9.50×10 ⁻⁸	1.21×10 ⁻⁹
	1.09×10 ⁻⁶	4.36×10 ⁻⁷	2.65×10 ⁻⁸	0
	7.43×10 ⁻⁷	7.31×10 ⁻⁸	9.49×10 ⁻⁹	0
	6.36×10 ⁻⁷	2.58×10 ⁻⁸	1.71×10 ⁻⁹	0
	2.30×10 ⁻⁸	2.72×10 ⁻⁹	0	0

Table 12--continued.

Eigenvalues of D					
	6.54×10^{-6}	7.05×10^{-6}	7.23×10^{-6}	6.36×10^{-6}	
	4.44×10^{-6}	2.38×10^{-6}	1.79×10^{-6}	1.67×10^{-6}	
	6.17×10^{-7}	9.39×10^{-7}	1.03×10^{-6}	1.24×10^{-6}	
	1.85×10^{-8}	4.15×10^{-7}	4.54×10^{-7}	3.87×10^{-7}	
	-4.43×10^{-8}	2.41×10^{-7}	3.42×10^{-7}	2.88×10^{-7}	
	-4.87×10^{-8}	6.86×10^{-8}	7.97×10^{-8}	9.30×10^{-8}	
	-1.38×10^{-7}	-2.93×10^{-8}	3.77×10^{-8}	7.18×10^{-8}	
	-4.80×10^{-7}	-1.52×10^{-7}	-1.97×10^{-7}	2.85×10^{-8}	
Calculated Value of F	0.42	1.01	5.69*	20.55*	
Π_1^{**}	--	--	--	9.79	
Π_2	--	--	--	0.08	
Approximate Power of Upper Tailed Test	--	--	--	0.10	

* Greater than F_{.05;g-p,N-g-r}.

** Values of Π_1 , Π_2 , and the power of the test were calculated by assuming $\beta_2 = (15.141, -122.429, -78.761, -78.275, 87.996, 13.456, -76.948, 34.721)$, and assuming $\sigma^2 = .008$.

$\Pi_1 = 244.75$, $\Pi_2 = 2.00$ and the approximate power is .83. From the entry in Table 12, we see that the calculated F value of 20.55 with the 30 cell clustering exceeds $F_{.05;19,4} = 5.81$, and we conclude that the fitted first degree model is inadequate.

4.9 Discussion

When a designed experiment includes replicated points, the adequacy of a fitted model can be tested by comparing the portion of the residual sum of squares due to lack of fit to a second portion due to pure error from the replicates. The test statistic is an F ratio of the mean square due to lack of fit to the mean square due to pure error, and lack of fit is inferred when the calculated value of this ratio is large (Draper and Smith, 1981, p.120).

When replicate points do not exist, lack of fit can be tested using near neighbor observations with the test statistic $F = \text{MSE}_B / \text{MSE}_W$. This F ratio has been shown to possess a central F distribution when the fitted model is adequate, and a doubly noncentral F distribution when the fitted model suffers from lack of fit.

When the fitted model is adequate, the expected values of both MSE_B and MSE_W are equal to σ^2 , so that the ratio $E[\text{MSE}_B] / E[\text{MSE}_W]$ equals unity. However, when lack of fit is present, both MSE_B and MSE_W are biased estimates of σ^2 , and we compare the magnitudes of the biases of these estimates (which are functions of the noncentrality parameters and degrees of freedom of the doubly noncentral F distribution)

in the F test. The test has an upper tailed rejection region if the bias corresponding to MSE_B exceeds the bias corresponding to MSE_W . The rejection region is lower tailed if the bias corresponding to MSE_W exceeds the bias corresponding to MSE_B . In other words, the test is upper tailed if the matrix D (see Eq. 4.13) is positive definite, and the test is lower tailed if D is negative definite. If D is indefinite then the test may be upper tailed, lower tailed or still yet lack of fit may not be testable depending upon the value of β_2 .

In two examples an iterative partitioning clustering algorithm is used to assign the data points to a preselected number of near neighbor cells. When the number of cells is low, the matrix D is found to be indefinite, so that the F test is not strictly upper tailed or lower tailed. However, by increasing the number of cells, it is possible in both examples to produce a positive definite matrix D, so that the test is upper tailed.

Increasing the number of cells not only produces an upper tailed test, but also affects the values of the parameters of the doubly noncentral F distribution. As the number of cells is increased (moving from left to right in Tables 9 and 12) we see that the smallest eigenvalue of C_1 increases and that the largest eigenvalue of C_2 decreases. Therefore a lower bound for π_1 ,

$$\frac{\beta_2' \beta_2 \psi_{\min}}{2\sigma^2} < \pi_1$$

increases as the number of cells increases (where $\psi_{\min}$ denotes the smallest eigenvalue of C_1). In addition, an upper bound for Π_2 ,

$$\Pi_2 \leq \frac{\frac{\beta_1' \beta_2 \rho_{\max}}{2-2}}{2\sigma^2}$$

decreases as the number of cells increases (where $\rho_{\max}$ denotes the largest eigenvalue of C_2). Finally, as the number of cells increases (moving from left to right in Tables 9 and 12), the numerator degrees of freedom, $g - p$, increase and the denominator degrees of freedom, $N - g - r$, decrease. Since the parameters of the doubly noncentral F distribution change as the number of cells changes, the power of the F test can be affected. For the stack loss data example, we see in Table 9 that the power of the upper tailed test decreases as we move from 9 to 10 to 11 to 12 cells.

An area for future study can be a further investigation of the effect of the number and composition of near neighbor cells on the power of the F test which makes use of $F = \text{MSE}_B / \text{MSE}_W$. This investigation would involve the effect of near neighbor cell selections on the parameters Π_1 , Π_2 , $g - p$, and $N - g - r$ of the doubly noncentral F distribution. It would be desirable to develop a method (perhaps an alternative to the iterative partitioning clustering algorithm) which could be used to select the number and composition of cells so as to maximize the power of the F test.

CHAPTER FIVE CONCLUSIONS AND RECOMMENDATIONS

Two general methods for testing a linear model fitted in a mixture space for lack of fit have been investigated in this dissertation. The first method makes use of response values observed at check points while the second method makes use of response values observed at design points which are near neighbors in the factor space.

In Chapter Two we discussed the work of several authors (Scheffé (1958), Gorman and Hinman (1962), Kurotori (1966), and Snee (1971)) for testing lack of fit which centered on measuring bias inherent in the fitted model when estimating the response at check points. Only the method suggested by Scheffé (1958) was an exact test. In Chapter Three, a method for selecting check points that maximizes the power of Scheffé's F test was devised. When replicate response observations were available, so that the experimental error variance could be estimated by $\hat{\sigma}_{\text{ext}}^2$ from the replicates, we saw that the power of this upper tailed F test was maximized by selecting check points that maximize (or approximately maximize) the noncentrality parameter λ_1 of the noncentral F distribution. When the matrix A_1 (where $\lambda_1 = \beta_2' A_1 \beta_2 / 2\sigma^2$) was found to be positive semi-definite it was determined that only a subset of possible values of the β_2 parameter vector could be detected as contributing to lack of fit.

When an estimate of the experimental error variance was not available from replicates, an extension of Scheffé's F test for lack of fit which replaced $\hat{\sigma}_{\text{ext}}^2$ with MSE (the residual mean square error from the fitted model) in the denominator was developed. We found that to maximize the power of the test it was necessary to select check points to maximize (or approximately maximize) the numerator noncentrality parameter, λ_1 , of the doubly noncentral F distribution when the test was upper tailed. When the test was lower tailed, we sought check point locations that minimized (or approximately minimized) λ_1 . A criterion was developed for determining whether the test was upper tailed or lower tailed by comparing the expected values of the numerator and denominator of the F ratio when the fitted model was inadequate. Finally, we discovered cases where, for some values of β_2 , lack of fit could not be tested.

An alternative to the check points method for testing lack of fit in a fitted model is a procedure that involves measuring the bias that is present in estimates of the response at the design points (the number of design points must exceed the number of terms in the model). When replicate observations are available, the well known procedure in which the test statistic is a ratio of the lack of fit mean square to the pure error mean square can be used to test for lack of fit (see Draper and Smith, 1981, p. 120). When replicate observations are not available, several techniques which make use of near neighbor

observations in place of replicates for testing lack of fit have been proposed in the literature (see Green (1971), Daniel and Wood (1971), and Shillington (1979)). Additionally, it has been suggested by Draper and Smith (1981, p. 42) that lack of fit can be tested by using near neighbor observations as substitutes for replicate observations in the usual lack of fit, pure error F ratio. However, the exact distributions of the test statistics proposed by Daniel and Wood (1971) and Draper and Smith (1981, p. 42) have not been defined, and Green's (1971) procedure requires an inordinately large number of observations. Thus because of these reasons we chose Shillington's (1979) procedure to study in greater detail in Chapter Four.

In Chapter Four the distributional properties of Shillington's test statistic were developed, and a method based on an iterative partitioning clustering algorithm for defining groups of near neighbor observations was proposed. It was shown that the power of Shillington's test depends on the parameters of the doubly noncentral F distribution, and that the manner in which observations are grouped as near neighbors can alter the values of the parameters of the doubly noncentral F distribution and thus affect the power of the test. We found that increasing the number of near neighbor cells so that individual cells become more compact produced an upper tailed F test in the

two examples studied, but that there are many other cases where the test will not be upper tailed.

Now that we have briefly summarized our findings from investigating the check point and near neighbor methods of testing lack of fit in a mixture model, a logical question is, "Which of the two methods is better?" It was not our original intent to address this question in this dissertation, but an interesting result that has been discovered in the latter stages of our investigations is as follows: Under certain circumstances, the check point method for testing lack of fit is equivalent to the usual method which partitions the residual sum of squares into sums of squares due to lack of fit and due to pure error (which was shown in Chapter Four to be a special case of the near neighbor method). Because we have not found a derivation of the equality of these methods in the literature, we shall show it here.

In Chapter Three, check points were used to test lack of fit in a fitted model of the form $E(\underline{Y}) = \underline{X}\underline{\beta}_1$. With k check points, the test statistic was of the form (see Eq. (3.3))

$$F = \frac{\underline{d}'\underline{V}_0^{-1}\underline{d}/k}{\hat{\sigma}_{\text{ext}}^2}$$

where $\hat{\sigma}_{\text{ext}}^2$ is an external estimate of σ^2 which can be

calculated from replicates, if they exist. The vector $\underline{d}$ in the F ratio was defined to be a vector of differences between observed and predicted response values at the k check points having the form

$$\underline{d} = \underline{y}^* - X^*(X'X)^{-1}X'\underline{y},$$

where $\underline{y}^*$ is the $k \times 1$ vector of observed response values at the k check points and X^* is the corresponding settings of the model terms at the check points. The matrix $\sigma^2 V_0$ was defined as the variance-covariance matrix of $\underline{d}$ where V_0 has the form

$$V_0 = I_k + X^*(X'X)^{-1}X^{*'}.$$

It can be shown (see Appendix 13) that if we define the vector $\underline{y}_A$ as

$$\underline{y}_A = \begin{bmatrix} \underline{y} \\ \underline{y}^* \end{bmatrix} = \begin{bmatrix} \text{observations at the original design points} \\ \text{observations at the check points} \end{bmatrix}$$

and similarly define the matrix X_A as

$$X_A = \begin{bmatrix} X \\ X^* \end{bmatrix} = \begin{bmatrix} \text{design point settings} \\ \text{check point settings} \end{bmatrix}$$

so that the original design points as well as the check points are all taken at once as design points in regressing $\underline{y}_A$ on X_A , then

$$SSE_A = \underline{Y}_A' [I_{(N+k)} - X_A (X_A' X_A)^{-1} X_A'] \underline{Y}_A = \underline{d}' V_0^{-1} \underline{d} + SSE. \quad (5.1)$$

Thus, the residual sum of squares, SSE_A , from the analysis of the fitted model when both the original design points and the check points are used to fit the model is equal to the sum of the quadratic form, $\underline{d}' V_0^{-1} \underline{d}$, used in the numerator of the check point F test and the residual sum of squares, SSE , from the analysis of the fitted model using data collected only from the original design points.

If we perform the usual partitioning of SSE_A into a lack of fit sum of squares, $SS_{LOF(A)}$, and a pure error sum of squares due to replicates, $SSE_{pure(A)}$, then from Eq. (5.1) we can write

$$SS_{LOF(A)} + SSE_{pure(A)} = \underline{d}' V_0^{-1} \underline{d} + SSE. \quad (5.2)$$

Thus from Eq. (5.2), when $SSE_{pure(A)}$ is equal to SSE , then $SS_{LOF(A)}$ becomes equal to $\underline{d}' V_0^{-1} \underline{d}$ so that the check point F ratio, $F = (\underline{d}' V_0^{-1} \underline{d} / k) / \hat{\sigma}_{ext}^2$, and the usual lack of fit F ratio, $F = MS_{LOF(A)} / MSE_{pure(A)}$, are equivalent. We now present an example to illustrate the result in Eq. (5.2).

Let us fit a second degree Scheffé polynomial model to the following hypothetical or artificial response observations collected at the six points of the $\{3,2\}$ simplex lattice design:

$$\underline{Y} = \begin{bmatrix} 2350 \\ 2370 \\ 2450 \\ 2430 \\ 2650 \\ 2670 \\ 2400 \\ 2420 \\ 2750 \\ 2730 \\ 2950 \\ 2970 \end{bmatrix} \quad \underline{X} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ .5 & .5 & 0 & .25 & 0 & 0 \\ .5 & .5 & 0 & .25 & 0 & 0 \\ .5 & 0 & .5 & 0 & .25 & 0 \\ .5 & 0 & .5 & 0 & .25 & 0 \\ 0 & .5 & .5 & 0 & 0 & .25 \\ 0 & .5 & .5 & 0 & 0 & .25 \end{bmatrix} .$$

The model is $\hat{Y} = 2360x_1 + 2440x_2 + 2660x_3 + 40x_1x_2 + 920x_1x_3 + 1640x_2x_3$. Since each of the six design points is replicated twice, there are six degrees of freedom available for estimating the experimental error variance. Let an estimate of the error variance be $\hat{\sigma}_{\text{ext}}^2 = \text{MSE}_{\text{pure}} = \text{SSE}_{\text{pure}}/6 = 1200/6 = 200$, and this value will be the denominator of the check point lack of fit F ratio.

Let us choose the three points $(2/3, 1/6, 1/6)$, $(1/6, 2/3, 1/6)$, and $(1/6, 1/6, 2/3)$ as check points and assume that we have observed the following values at these points

$$\underline{Y}^* = \begin{bmatrix} 2690 \\ 2770 \\ 2980 \end{bmatrix} \quad \underline{X}^* = \begin{bmatrix} 2/3 & 1/6 & 1/6 & 1/9 & 1/9 & 1/36 \\ 1/6 & 2/3 & 1/6 & 1/9 & 1/36 & 1/9 \\ 1/6 & 1/6 & 2/3 & 1/36 & 1/9 & 1/9 \end{bmatrix} .$$

The numerator of the check point lack of fit F ratio is then calculated to be $\underline{d}'\underline{V}_0^{-1}\underline{d} = 24546.7$, so that

$$F = (\underline{d}'\underline{V}_0^{-1}\underline{d}/k)/\hat{\sigma}_{\text{ext}}^2 = (24546.7/3)/200 = 40.91.$$

If we use all of the observed response values to fit the second order Scheffé polynomial, then the model is $\hat{Y} = 2360.7x_1 + 2437.7x_2 + 2661.7x_3 + 183.3x_1x_2 + 1071.3x_1x_3 + 1785.3x_2x_3$ and the residual sum of squares is $SSE_A = 25746.7$. This residual sum of squares can be partitioned into $SS_{LOF(A)} = 24546.7$ and $SSE_{pure(A)} = 1200$. The F ratio for testing lack of fit is calculated to be $F = MS_{LOF(A)} / MSE_{pure(A)} = [24546.7/3] / [1200/6] = 40.91$, which is identical to the previously calculated F value.

In the above example we note that $SSE_{pure(A)}$ is equal to SSE ($SSE = SSE_{pure(A)}$) so that $SS_{LOF(A)}$ is equal to $\underline{d}'V_0^{-1}\underline{d}$. Since both the check point F ratio and the usual lack of fit F ratio have produced the same value, $F = 40.91$, we conclude that the two methods for testing lack of fit in the fitted model are equivalent.

In order to put this dissertation in a better perspective, we now make some concluding remarks on the lack of fit testing procedures investigated, including possible drawbacks, extensions, and recommendations for future work.

An aspect of our investigations that may raise some questions is that our methods are dependent on the specification of the form of the true model believed to be responsible for lack of fit in the fitted model. Requiring the form of the true model to be specified was necessary in order to be able to investigate the power of the testing procedures. There are situations, however, where a complete or true model can reasonably be specified. One example

could be in fitting polynomial models, where the polynomial of one degree higher than the fitted model could be taken as the true model.

We now mention two ways in which our results can be applied to more general situations than may be readily apparent from our previous discussions. First, we point out that all examples in Chapters Three and Four dealt with polynomial models. This type of model was selected because of its popularity and wide applicability, however, our methods can be applied not only to polynomial models but to any models which are linear in their parameters. Secondly, it was our intent in this dissertation to discuss methods for testing lack of fit in a mixture model, but the methods discussed can certainly be used not only in mixture problems but also in general response surface problems in which a linear model is fitted. This generalization is illustrated for the near neighbor approach to lack of fit testing through the stack loss example in Chapter Four.

Topics for future research stemming from this dissertation were listed in the concluding paragraphs of Chapters Three and Four. One area suggested in Chapter Three was to investigate the effect of experimental design on the selection of check points and on the resulting power of the test. Perhaps a "minimum bias" design could be used for fitting the model, while lack of fit could be detected with "high bias" check points, but this is only speculation, and needs to be investigated.

The fact that the check points method and the standard method that partitions the residual sum of squares into lack of fit and pure error portions were found, under a certain condition, to be equivalent suggests that selecting check points to maximize the power of the check point F test may in general be equivalent to choosing points to augment the original design. The augmented points would be chosen to maximize the power of the F test that partitions the residual sum of squares into lack of fit and pure error sums of squares. An investigation of the selection of optimal check points versus the selection of optimal augmented design points would be of interest.

For the near neighbor test for lack of fit it was recommended in Chapter Four that other methods besides the iterative partitioning clustering algorithm might be considered for selecting groups of near neighbors. The effect of the number and composition of the groups selected on the power of the test through their effect on the parameters of the doubly noncentral F distribution could then be investigated.

In view of the equivalence of the check point method and the method that partitions the residual sum of squares when replicates exist (see Eq. (5.2)), it would be of interest to investigate whether there is also some equivalence between Shillington's near neighbor F ratio and the check point F ratio, $F = (\underline{d}'V_0^{-1}\underline{d}/k)/\text{MSE}$, to be used when an external estimate of σ^2 is not available. If the methods

are not equivalent, perhaps one could be shown to be preferable to the other as judged by comparing the power of the two procedures in testing for lack of fit.

Finally, the focus of this dissertation has been on testing lack of fit in linear models so that another area for future investigation can be the problem of testing lack of fit in models which are nonlinear in their parameters.

APPENDIX 1
INFLUENCE OF λ_1 ON $P\{F''_{v_1, v_2; \lambda_1, \lambda_2} > F_{\alpha; v_1, v_2}\}$

In this appendix we show that $P\{F''_{v_1, v_2; \lambda_1, \lambda_2} > F_{\alpha; v_1, v_2}\}$ is an increasing function of λ_1 .

Let $X_1, \dots, X_{v_1}, Y_1, \dots, Y_{v_2}$ be independent $N(0,1)$.

Then

$$F = (v_2/v_1) [(X_1 + \lambda_1^{1/2})^2 + \sum_{i=2}^{v_1} X_i^2] / [(Y_1 + \lambda_2^{1/2})^2 + \sum_{i=2}^{v_2} Y_i^2]$$

is distributed as $F''_{v_1, v_2; \lambda_1, \lambda_2}$ where v_1 and v_2 are the respective numerator and denominator degrees of freedom and λ_1 and λ_2 are the respective numerator and denominator noncentrality parameters (Scheffé, 1959, p. 412-413).

Fixing the values of v_1, v_2 , and λ_2 we wish to show $P\{F''_{v_1, v_2; \lambda_1, \lambda_2} > F_{\alpha; v_1, v_2}\}$ is a strictly increasing function of λ_1 , where $F_{\alpha; v_1, v_2}$ represents the upper 100 α percentage point of the central F distribution with v_1 and v_2 degrees of freedom. Let

$$\begin{aligned} f(\lambda_1^{1/2}) \\ = P\{ (v_2/v_1) [(X_1 + \lambda_1^{1/2})^2 + \sum_{i=2}^{v_1} X_i^2] / [(Y_1 + \lambda_2^{1/2})^2 + \sum_{i=2}^{v_2} Y_i^2] \\ > F_{\alpha; v_1, v_2} \}, \end{aligned}$$

then $f(\lambda_1^{1/2})$ may be rewritten as

$$f(\lambda_1^{1/2}) = P\{(X_1 + \lambda_1^{1/2})^2 > U\} = 1 - P\{(X_1 + \lambda_1^{1/2})^2 \leq U\}, \quad (A1.1)$$

$$\text{where } U = (v_1/v_2)[(Y_1 + \lambda_2^{1/2})^2 + \sum_{i=2}^{v_2} Y_i^2] F_{\alpha; v_1, v_2} - \sum_{i=2}^{v_1} X_i^2.$$

Note that the random variable U is independent of X_1 .

If $\lambda_{11}^{1/2}$ and $\lambda_{12}^{1/2}$ denote any two values of $\lambda_1^{1/2}$ such that $\lambda_{11}^{1/2} < \lambda_{12}^{1/2}$, then we shall prove that for $f(\lambda_1^{1/2})$

defined as in (A1.1), $f(\lambda_{11}^{1/2}) < f(\lambda_{12}^{1/2})$. Now

$$f(\lambda_1^{1/2}) = 1 - \int_0^{\infty} g_{\lambda_1^{1/2}}(u)p(u)du \text{ where } p(u) \text{ is the p.d.f. of } U,$$

and for any positive number, u' , $g_{\lambda_1^{1/2}}(u')$ denotes the conditional probability that $(X_1 + \lambda_1^{1/2})^2 < u'$, given

$U = u'$. However, this conditional probability must be the same as the unconditional probability, since X_1 and U are statistically independent.

Thus $g_{\lambda_1^{1/2}}(u')$ is the probability that the random variable X_1 falls in an interval of half length $u'^{1/2}$ centered at $-\lambda_1^{1/2}$. Since $X_1 \sim N(0,1)$, this is a decreasing function of $\lambda_1^{1/2}$. Therefore $g_{\lambda_{11}^{1/2}}(u') - g_{\lambda_{12}^{1/2}}(u') > 0$ for all $u' > 0$. Hence,

$$f(\lambda_{11}^{1/2}) - f(\lambda_{12}^{1/2}) = \int_0^{\infty} [-g_{\lambda_{11}^{1/2}}(u) + g_{\lambda_{12}^{1/2}}(u)]p(u)du < 0.$$

Thus $P\{F_{v_1, v_2; \lambda_1, \lambda_2}'' > F_{\alpha; v_1, v_2}\}$ is a strictly increasing function of λ_1 .

We note that this proof is a modification of the proof that $P\{F_{v_1, v_2; \lambda_1, \lambda_2}'' > F_{\alpha; v_1, v_2}\}$ is decreasing in λ_2 (Scheffé, 1959, p.136).

APPENDIX 2
A CONTROLLED RANDOM SEARCH PROCEDURE
FOR GLOBAL OPTIMIZATION

W. L. Price (1977) describes a conceptually simple random search procedure, called "a controlled random search procedure for global optimization," which is effective in searching for global minima of a function of n variables, with or without constraints. The procedure does not require the function to be differentiable or the variables to be continuous.

An initial search domain, V , is defined by specifying upper and lower bounds for each of the n variables, and a predetermined number, N , of trial points are chosen at random over V , consistent with any constraints. The function is evaluated at each of the N trial points and the position as well as the value of the function at each point are stored in an array, A' . At each iteration a new trial point, $\underline{p}$, is selected randomly from a set of possible trial points whose positions are related to the configuration of the N points currently in storage. If $\underline{p}$ satisfies the constraints, the function is evaluated at $\underline{p}$ and the function value, f_p , is compared with f_m , which is the greatest function value for the N points already in storage.

If $f_p < f_m$ then $\underline{M}$, the point in storage corresponding to f_m , is replaced, in the array A' , by $\underline{p}$. If $\underline{p}$ fails to satisfy

the constraints or if $f_p > f_m$ then the trial is discarded and a new point is chosen from the potential trial set.

As the algorithm proceeds, the set of N points in storage tend to cluster around minima. As Price states, "the probability that the points ultimately coverge onto the global minimum (minima) depends on the value of N , the complexity of the function, the nature of the constraints and the way in which the set of potential trial points is chosen."

Price notes that since the procedure is intended to find global minima, thoroughness of search is more important than speed of convergence, but if the procedure is to be more efficient than pure random search the probability of success ($f_p < f_m$) at each iteration must be sufficiently high. His procedure reaches a compromise between the requirements of search and convergence by defining the set of potential trial points in terms of the configuration of the N points already in storage. At each iteration $n + 1$ distinct points, $R_1, R_2, \dots, R_{n+1}$, are chosen at random from the N ($N > n$) currently in storage and these constitute a simplex of points in n -space. The point R_{n+1} is arbitrarily chosen as the vertex of the simplex, and the next trial point, P , is taken as the image of the vertex with respect to the centroid, G , of the remaining n points. Thus $P = 2G - R_{n+1}$. He notes that it is possible to speed up covergence by selecting the vertex as the point R_i , $i = 1, 2, \dots, n + 1$, which has the largest

function value of the points $\underline{R}_1, \underline{R}_2, \dots, \underline{R}_{n+1}$ but this would be detrimental to the thoroughness of the search.

The version of Price's procedure used in the work in this dissertation was programmed in the FORTRAN language by Michael Conlon of the Center for Instructional and Research Computing Activities, the University of Florida, Gainesville, Florida. This version of Price's procedure selects new trial points using the suggested criterion $\underline{P} = 2\underline{G} - \underline{R}_{n+1}$. The algorithm continues until an iteration limit is reached or a desired tolerance between the minimum and maximum function values in storage is achieved.

In our particular application, if $p_2 = 1$ so that A_1 is a scalar, we wish to maximize

$$A_1 = (X_2^* - X^*A)'V_0^{-1}(X_2^* - X^*A),$$

with respect to k check points, in order to maximize the power of an upper tailed test. For locating check points that maximize the power of a lower tailed test it is necessary to minimize A_1 . If $p_2 > 1$ so that A_1 is not a scalar, but is a $p_2 \times p_2$ matrix, then it will be necessary to maximize or minimize certain eigenvalues of A_1 .

All of these optimization problems can be handled by Price's procedure. Since the procedure finds minima, then to find maxima, we simply minimize the negative of the function under consideration. The restriction that the check points must be located within the experimental simplex

(or a subregion of the simplex) is taken care of by specifying constraints in the program.

To give a specific example, suppose we fit a second order canonical polynomial model in a three component mixture space, using a simplex centroid design. If we assume the true model is special cubic in the three components, then $p_2 = 1$, and

$$A_1 = (X_2^* - X^*A)'V_0^{-1}(X_2^* - X^*A)$$

is a scalar quantity. In order to locate a single check point that maximizes the power of an upper tailed test for lack of fit, we select the check point that maximizes A_1 . Since the experimental region we wish to search is the entire two dimensional simplex, we define the check point as $\underline{x}^* = (x_1, x_2, x_3)$, and in our program impose the constraints:

$$0 < x_1 < 1,$$

and

$$0 < x_2 < 1.$$

We then define x_3 as $x_3 = 1 - x_1 - x_2$, while requiring that $0 < x_3 < 1$. Price's random search procedure is used to search the two-dimensional simplex for the point (x_1, x_2) that maximizes A_1 .

Price suggests the use of $N = 50$ storage points for such a two-dimensional search, and we have generally found this to be adequate. For $k > 1$ check points to be located simultaneously in a three component system, the problem becomes one of searching in $2k$ dimensions. For the applications considered, $N = 50k$ appears to be adequate.

The only real problem encountered has been that of economics in that the procedure becomes costly in terms of computer time for these situations where the optimal value of the function is assumed by all points in a region. In these cases the algorithm searches in vain for points that will improve upon the functional values already in storage, which all lie in this optimum region. However, in other applications, the procedure converged quickly to an optimum (those that converged did so in 10,000 iterations or less, at a small cost in computational time).

APPENDIX 3
STATISTICAL INDEPENDENCE OF $\underline{d}'V_0^{-1}\underline{d}/\sigma^2$ AND SSE/σ^2

Let us write $\underline{d}'V_0^{-1}\underline{d}$ as

$$\begin{aligned}\underline{d}'V_0^{-1}\underline{d} &= (\underline{Y}^* - \hat{\underline{Y}}^*)'V_0^{-1}(\underline{Y}^* - \hat{\underline{Y}}^*) \\ &= \underline{Y}^{*'}V_0^{-1}\underline{Y}^* - \hat{\underline{Y}}^{*'}V_0^{-1}\underline{Y}^* - \underline{Y}^{*'}V_0^{-1}\hat{\underline{Y}}^* + \hat{\underline{Y}}^{*'}V_0^{-1}\hat{\underline{Y}}^* \\ &= \underline{Y}^{*'}V_0^{-1}\underline{Y}^* - 2 \underline{Y}^{*'}V_0^{-1}\hat{\underline{Y}}^* + \hat{\underline{Y}}^{*'}V_0^{-1}\hat{\underline{Y}}^*.\end{aligned}$$

Now let us write SSE as

$$SSE = \underline{Y}'(I_N - X(X'X)^{-1}X')\underline{Y}.$$

Since $\underline{Y}$ and $\underline{Y}^*$ are independent, SSE is independent of $\underline{Y}^{*'}V_0^{-1}\underline{Y}^*$. Rewriting $\underline{Y}^{*'}V_0^{-1}\hat{\underline{Y}}^*$ as

$$\underline{Y}^{*'}V_0^{-1}\hat{\underline{Y}}^* = \underline{Y}^{*'}V_0^{-1}X^* \underline{b}_1,$$

where $\underline{b}_1$ is the least squares estimator of $\underline{\beta}_1$, we have

$$\underline{Y}^{*'}V_0^{-1}\hat{\underline{Y}}^* = \underline{Y}^{*'}V_0^{-1}X^*(X'X)^{-1}X'\underline{Y}.$$

We now show that the second portion of $\underline{d}'\underline{V}_0^{-1}\underline{d}$ is independent of SSE if and only if

$$[\underline{V}_0^{-1}\underline{X}^*(\underline{X}'\underline{X})^{-1}\underline{X}'] [\underline{I}_N - \underline{X}(\underline{X}'\underline{X})^{-1}\underline{X}'] = 0.$$

Define

$$\begin{aligned} & \text{cov}[\underline{Y}^*, \underline{V}_0^{-1}\underline{X}^*(\underline{X}'\underline{X})^{-1}\underline{X}'\underline{Y}, \underline{Y}'(\underline{I}_N - \underline{X}(\underline{X}'\underline{X})^{-1}\underline{X}')\underline{Y}] \\ &= E[\underline{Y}^* \underline{V}_0^{-1}\underline{X}^*(\underline{X}'\underline{X})^{-1}\underline{X}'\underline{Y}\underline{Y}'(\underline{I}_N - \underline{X}(\underline{X}'\underline{X})^{-1}\underline{X}')\underline{Y}] \\ &\quad - E[\underline{Y}^* \underline{V}_0^{-1}\underline{X}^*(\underline{X}'\underline{X})^{-1}\underline{X}'\underline{Y}] E[\underline{Y}'(\underline{I}_N - \underline{X}(\underline{X}'\underline{X})^{-1}\underline{X}')\underline{Y}] \\ &= E(\underline{Y}^*) E[\underline{V}_0^{-1}\underline{X}^*(\underline{X}'\underline{X})^{-1}\underline{X}'\underline{Y}\underline{Y}'(\underline{I}_N - \underline{X}(\underline{X}'\underline{X})^{-1}\underline{X}')\underline{Y}] \\ &\quad - E(\underline{Y}^*) E[\underline{V}_0^{-1}\underline{X}^*(\underline{X}'\underline{X})^{-1}\underline{X}'\underline{Y}] E[\underline{Y}'(\underline{I}_N - \underline{X}(\underline{X}'\underline{X})^{-1}\underline{X}')\underline{Y}] \\ &= E(\underline{Y}^*) [\text{cov}(\underline{V}_0^{-1}\underline{X}^*(\underline{X}'\underline{X})^{-1}\underline{X}'\underline{Y}, \underline{Y}'(\underline{I}_N - \underline{X}(\underline{X}'\underline{X})^{-1}\underline{X}')\underline{Y})] \\ &= 0, \end{aligned}$$

if $\underline{V}_0^{-1}\underline{X}^*(\underline{X}'\underline{X})^{-1}\underline{X}'\underline{Y}$ is independent of $\underline{Y}'(\underline{I}_N - \underline{X}(\underline{X}'\underline{X})^{-1}\underline{X}')\underline{Y}$.

This occurs if and only if,

$$[\underline{V}_0^{-1}\underline{X}^*(\underline{X}'\underline{X})^{-1}\underline{X}'] [\underline{I}_N - \underline{X}(\underline{X}'\underline{X})^{-1}\underline{X}'] = 0,$$

see Searle (1971) p.59, Theorem 3. Now,

$$\begin{aligned}
& [V_0^{-1}X^*(X'X)^{-1}X'] [I_N - X(X'X)^{-1}X'] \\
&= V_0^{-1}X^*(X'X)^{-1}X' - V_0^{-1}X^*(X'X)^{-1}X'X(X'X)^{-1}X' \\
&= 0.
\end{aligned}$$

Therefore SSE is independent of the second portion of $\underline{d}'V_0^{-1}\underline{d}$. Now we must show that SSE is independent of the third portion of $\underline{d}'V_0^{-1}\underline{d}$. Write $\hat{\underline{y}}^*V_0^{-1}\hat{\underline{y}}^*$ as

$$\begin{aligned}
\hat{\underline{y}}^*V_0^{-1}\hat{\underline{y}}^* &= (X^* \underline{b}_1)' V_0^{-1} X^* \underline{b}_1 \\
&= \underline{y}' X(X'X)^{-1} X^* V_0^{-1} X^* (X'X)^{-1} X' \underline{y}.
\end{aligned}$$

Then SSE is independent of the third portion of $\underline{d}'V_0^{-1}\underline{d}$ if and only if

$$[X(X'X)^{-1}X^* V_0^{-1}X^* (X'X)^{-1}X'] [I_N - X(X'X)^{-1}X'] = 0,$$

see Searle (1971), p.59, Theorem 4. Continuing then,

$$\begin{aligned}
& [X(X'X)^{-1}X^* V_0^{-1}X^* (X'X)^{-1}X'] [I_N - X(X'X)^{-1}X'] \\
&= X(X'X)^{-1}X^* V_0^{-1}X^* (X'X)^{-1}X' - X(X'X)^{-1}X^* V_0^{-1}X^* (X'X)^{-1}X'X(X'X)^{-1}X' \\
&= X(X'X)^{-1}X^* V_0^{-1}X^* (X'X)^{-1}X' - X(X'X)^{-1}X^* V_0^{-1}X^* (X'X)^{-1}X' \\
&= 0.
\end{aligned}$$

Therefore SSE is independent of the third portion of $\underline{d}'V_0^{-1}\underline{d}$.

Finally, since SSE is independent of each of the three portions of $\underline{d}'V_0^{-1}\underline{d}$, we can conclude that SSE is independent of $\underline{d}'V_0^{-1}\underline{d}$ and therefore SSE/σ^2 is independent of $\underline{d}'V_0^{-1}\underline{d}/\sigma^2$.

APPENDIX 4
THEOREM 3.1

Theorem 3.1

Let A and B be $k \times k$ matrices. If $(A - B)$ is positive definite and B is positive semi-definite, then A is positive definite.

Proof

We assume that $(A - B)$ is positive definite. Then $\underline{z}'(A - B)\underline{z} > 0$, for all $\underline{z} \neq \underline{0}$. Thus $\underline{z}'A\underline{z} - \underline{z}'B\underline{z} > 0$, for all $\underline{z} \neq \underline{0}$, so that $\underline{z}'A\underline{z} > \underline{z}'B\underline{z} > 0$, for all $\underline{z} \neq \underline{0}$, since B is positive semi-definite. Therefore, $\underline{z}'A\underline{z} > 0$, all $\underline{z} \neq \underline{0}$.

Now if $\underline{z}'A\underline{z} = 0$, then $\underline{z} = \underline{0}$ for if $\underline{z} \neq \underline{0}$, then $\underline{z}'(A - B)\underline{z} > 0$ implies $\underline{z}'B\underline{z} < 0$. But this is a contradiction since by assumption $\underline{z}'B\underline{z} \geq 0$. Therefore $\underline{z}$ must be $\underline{0}$ and A must be positive definite.

APPENDIX 5
THEOREM 3.2

Theorem 3.2

Let A_1 and A_2 be $p_2 \times p_2$ positive semi-definite matrices. Let $\underline{\beta}_2$ be a p_2 -dimensional vector and define λ_1 and λ_2 as

$$\lambda_1 = \underline{\beta}_2' A_1 \underline{\beta}_2 / 2\sigma^2, \text{ and}$$

$$\lambda_2 = \underline{\beta}_2' A_2 \underline{\beta}_2 / 2\sigma^2,$$

where $\sigma^2 > 0$. Let $k > 0$, $N > 0$, $p > 0$, and $N > p$.

(a) If $[A_1/k - A_2/(N - p)]$ is positive definite then $[\lambda_1/k - \lambda_2/(N - p)] = 0$ if and only if $\lambda_1 = \lambda_2 = 0$.

(b) If $[A_1/k - A_2/(N - p)]$ is negative definite then $[\lambda_1/k - \lambda_2/(N - p)] = 0$ if and only if $\lambda_1 = \lambda_2 = 0$.

Proof of part (a).

Necessity. Let $[A_1/k - A_2/(N - p)]$ be positive definite and suppose that $[\lambda_1/k - \lambda_2/(N - p)] = 0$. We show that $\lambda_1 = \lambda_2 = 0$. The matrix $A_1/k - A_2/(N - p)$ being positive definite implies $\underline{\beta}_2' [A_1/k - A_2/(N - p)] \underline{\beta}_2 = 0$ iff $\underline{\beta}_2 = \underline{0}$, that is,

$$\beta_2' A_1 \beta_2 / k - \beta_2' A_2 \beta_2 / (N - p) = 0 \text{ iff } \beta_2 = \underline{0}. \text{ But,}$$

$$\beta_2' A_1 \beta_2 / 2k\sigma^2 - \beta_2' A_2 \beta_2 / 2(N - p)\sigma^2 = 0 \text{ iff } \beta_2 = \underline{0}. \text{ Hence}$$

$$\lambda_1 / k - \lambda_2 / (N - p) = 0 \text{ iff } \beta_2 = \underline{0}.$$

It follows that if $\lambda_1 / k - \lambda_2 / (N - p) = 0$, then $\lambda_1 = \lambda_2 = 0$.

Sufficiency. Obviously, if $\lambda_1 = \lambda_2 = 0$, then

$$\lambda_1 / k - \lambda_2 / (N - p) = 0 - 0 = 0.$$

Proof of part (b). This follows from part (a), since in this case $A_2 / (N - p) - A_1 / k$ is positive definite.

APPENDIX 6 AN APPROXIMATION TO THE DOUBLY NONCENTRAL F DISTRIBUTION

Johnson and Kotz (1970, p.197) indicate the following approximation for $P\{F''_{v_1, v_2; \lambda_1, \lambda_2} \leq F_{\alpha; v_1, v_2}\}$ where v_1 and v_2 are the numerator and denominator degrees of freedom, respectively, and λ_1 and λ_2 are the numerator and denominator noncentrality parameters, respectively:

$$\begin{aligned} P\{F''_{v_1, v_2; \lambda_1, \lambda_2} \leq F_{\alpha; v_1, v_2}\} &\doteq P\{cF_{v, v'} \leq F_{\alpha; v_1, v_2}\} \\ &= P\{F_{v, v'} \leq (1/c)F_{\alpha; v_1, v_2}\} \end{aligned}$$

where $F_{\alpha; v_1, v_2}$ is the upper 100α percentage point of the central F distribution with v_1 and v_2 degrees of freedom, and where $c = [1 + \lambda_1/v_1]/[1 + \lambda_2/v_2]$, $v = [v_1 + \lambda_1]^2/[v_1 + 2\lambda_1]$, $v' = [v_2 + \lambda_2]^2/[v_2 + 2\lambda_2]$, and $F_{v, v'}$ is a central F random variable with v and v' degrees of freedom.

APPENDIX 7
EQUIVALENCE OF SSE_B AND SS_{LOF} WHEN
REPLICATES REPLACE NEAR NEIGHBOR OBSERVATIONS

In this appendix we show that $SSE_B = SS_{LOF}$ when response observations are partitioned into g groups of true replicates rather than g groups of near neighbor observations.

From Chapter Two, Section 2.2, if each cell consists entirely of true replicates, then the sum of squares due to lack of fit can be expressed as

$$SS_{LOF} = SSE - SSE_{\text{pure}}$$

where SSE is the residual sum of squares from a least squares regression of $\underline{Y}$ on X and where SSE_{pure} is the sum of squares due to pure error, calculated from replicates.

Since $SSE_{\text{pure}} = \underline{Y}'\Sigma_0\underline{Y}$, where Σ_0 is defined as in Section 4.2, we have

$$\begin{aligned} SS_{LOF} &= \underline{Y}'(I_N - X(X'X)^{-1}X')\underline{Y} - \underline{Y}'\Sigma_0\underline{Y} \\ &= \underline{Y}'(I_N - \Sigma_0)\underline{Y} - \underline{Y}'X(X'X)^{-1}X'\underline{Y}. \end{aligned} \quad (A7.1)$$

We wish to show that when each cell is composed entirely of true replicates, SSE_B is equal to the expression in (A7.1).

Recalling from Section 4.3.1 that $\bar{\underline{Y}}_C = M\underline{Y}$, where $M = \text{diag}[(1/n_1)\underline{1}'_1, \dots, (1/n_g)\underline{1}'_g]$, we write

$$\begin{aligned} \text{SSE}_B &= \bar{\underline{Y}}'_C [G_0^{-1} - G_0^{-1}X_C(X'_C G_0^{-1}X_C)^{-1}X'_C G_0^{-1}] \bar{\underline{Y}}_C \\ &= \underline{Y}' M' [G_0^{-1} - G_0^{-1}X_C(X'_C G_0^{-1}X_C)^{-1}X'_C G_0^{-1}] M \underline{Y}, \end{aligned}$$

where from Section 4.2, $G_0 = \text{diag}[1/n_1, 1/n_2, \dots, 1/n_g]$. Recognizing that $G_0 = MM'$ and $X_C = MX$, we have

$$\begin{aligned} \text{SSE}_B &= \underline{Y}' [M'(MM')^{-1}M \\ &\quad - M'(MM')^{-1}MX\{X'M'(MM')^{-1}MX\}^{-1}X'M'(MM')^{-1}M] \underline{Y}. \end{aligned}$$

Since $M'(MM')^{-1}M = I_N - \Sigma_0$, we have

$$\begin{aligned} \text{SSE}_B &= \underline{Y}' (I_N - \Sigma_0) \underline{Y} \\ &\quad - \underline{Y}' (I_N - \Sigma_0) X \{X' (I_N - \Sigma_0) X\}^{-1} X' (I_N - \Sigma_0) \underline{Y} \end{aligned}$$

and since $\Sigma_0 X = 0$ when all cells are composed entirely of true replicates, we have

$$\begin{aligned} \text{SSE}_B &= \underline{Y}' (I_N - \Sigma_0) \underline{Y} - \underline{Y}' X (X' X)^{-1} X' \underline{Y} \\ &= \text{SS}_{\text{LOF}}, \end{aligned}$$

from (A7.1). Therefore, SSE_B is equal to the usual SS_{LOF} when cells are composed entirely of true replicates.

APPENDIX 8
LEMMA 4.1

Lemma 4.1

Let $\underline{Y} \sim (X\underline{\beta}, \sigma^2 G)$, G singular. Define $T = G + XX'$.

Define T^- such that $TT^-T = T$.

1. (i) $TT^-X = X$

(ii) $X'T^-T = X'$

2. $\text{rank}(X'T^-X) = \text{rank}(X)$

3. (i) $X(X'T^-X)^-(X'T^-X) = X$

(ii) $(X'T^-X)(X'T^-X)^-X' = X'$

4. $\underline{Y}$ is in the column space of T ($\underline{Y} \in C(T)$), with probability one, by which we mean that there exists a vector $\underline{a}$ such that letting $\underline{Y} = (y_1, y_2, \dots, y_N)'$ and $T\underline{a} = (x_1, x_2, \dots, x_N)'$, then
 $P\{|y_i - x_i| > \epsilon\} = 0$, for all $\epsilon > 0$, $i = 1, 2, \dots, N$.

Proof

1. (i) $T = XX' + G$

$= XX' + VV'$, where $G = VV'$

$= CC'$, where $C = [X:V]$.

Now, $CC'(CC')^-C = C$, from Pringle and Rayner (1971, p. 26), and therefore

$$TT^-[X:V] = [X:V]$$

from which it follows that $TT^-X = X$ (and $TT^-V = V$).

- (ii) The proof of (ii) follows directly from (i) by taking the transpose.
2. The proof of part 2 is given in Rao, 1973, p. 77, #30.
 3. The proof of part 3 is given in Rao and Mitra, 1971, p.22, Lemma 2.2.6(c).
 4. By definition, $\underline{Y} \sim (X\beta, \sigma^2 G)$ so that an equivalent representation for $\underline{Y}$ is $\underline{Y} = X\beta + \underline{\varepsilon}$, where $\underline{\varepsilon} \sim (0, \sigma^2 G)$. We wish to show that the random vector $\underline{Y}$ is in the column space of T , with probability one. It is sufficient to show that $TT^{-}\underline{Y} = \underline{Y}$, w.p.1. (see Pringle and Rayner, 1971, p.9). Rewriting $TT^{-}\underline{Y}$ we have

$$\begin{aligned} TT^{-}\underline{Y} &= TT^{-}(X\beta + \underline{\varepsilon}) \\ &= TT^{-}X\beta + TT^{-}\underline{\varepsilon}. \end{aligned}$$

By part 1 of Lemma 4.1, $TT^{-}X = X$, and therefore $X\beta \in C(T)$. The proof is complete if we show $TT^{-}\underline{\varepsilon} = \underline{\varepsilon}$, w.p.1. The difference $TT^{-}\underline{\varepsilon} - \underline{\varepsilon}$ can be written as $TT^{-}\underline{\varepsilon} - \underline{\varepsilon} = (TT^{-} - I_N)\underline{\varepsilon}$, therefore we must show (see explanation below) that

$$E[\underline{\varepsilon}'(TT^{-} - I_N)'(TT^{-} - I_N)\underline{\varepsilon}] = 0. \quad (A8.1)$$

The expectation in Eq. (A8.1) can be written as

$$\begin{aligned}
& E[\underline{\varepsilon}'(TT^{-} - I_N)'(TT^{-} - I_N)\underline{\varepsilon}] \\
&= \text{trace} [(TT^{-} - I_N)'(TT^{-} - I_N)\sigma^2 G] \\
&= \sigma^2 \text{trace} [(TT^{-} - I_N)'(TT^{-} - I_N)VV'] \\
&= 0
\end{aligned}$$

since $TT^{-}V = V$, by proof of part 1(i) of Lemma 4.1.

Therefore $\underline{Y} \in C(T)$, w.p.1.

We now show that proving the equality in (A8.1) is equivalent to proving that $TT^{-}\underline{\varepsilon} = \underline{\varepsilon}$, w.p.1. By the Markov Inequality

$$P\{ |u_i - v_i| > \varepsilon \} \leq [E(u_i - v_i)^2] / \varepsilon^2$$

and therefore if $E(u_i - v_i)^2 = 0$, we have $u_i = v_i$, w.p.1.

If $\underline{u}' = (u_1, u_2, \dots, u_N)$, $\underline{v}' = (v_1, v_2, \dots, v_N)$, and if $E(u_i - v_i)^2 = 0$, for $i = 1, 2, \dots, N$, then $u_i = v_i$, w.p.1, for $i = 1, 2, \dots, N$, which implies that $\underline{u} = \underline{v}$, w.p.1. But $E(u_i - v_i)^2 = 0$, for $i = 1, 2, \dots, N$ if and only if $\sum_{i=1}^N E(u_i - v_i)^2 = 0$, and since

$$\sum_{i=1}^N E(u_i - v_i)^2 = E(\underline{u} - \underline{v})'(\underline{u} - \underline{v})$$

we have $\underline{u} = \underline{v}$, w.p.1, if $E(\underline{u} - \underline{v})'(\underline{u} - \underline{v}) = 0$. In (A8.1) we take $\underline{u} = TT^{-}\underline{\varepsilon}$ and $\underline{v} = \underline{\varepsilon}$.

APPENDIX 9
PROOF OF THEOREM 4.1(i)

In this appendix we give the proof of Theorem 4.1(i). We show that $E(\hat{\sigma}^2) = \sigma^2$, where $\hat{\sigma}^2 = f^{-1}(\underline{Y} - X\hat{\underline{\beta}})'T^{-}(\underline{Y} - X\hat{\underline{\beta}})$. First we write $\hat{\sigma}^2$ as

$$\hat{\sigma}^2 = f^{-1}[\underline{Y}'T^{-}\underline{Y} - 2\hat{\underline{\beta}}'X'T^{-}\underline{Y} + \hat{\underline{\beta}}'X'T^{-}X\hat{\underline{\beta}}],$$

where $\hat{\underline{\beta}} = (X'T^{-}X)^{-}X'T^{-}\underline{Y}$. Now,

$$\begin{aligned}\hat{\underline{\beta}}'X'T^{-}X\hat{\underline{\beta}} &= \hat{\underline{\beta}}'X'T^{-}X(X'T^{-}X)^{-}X'T^{-}\underline{Y} \\ &= \hat{\underline{\beta}}'X'T^{-}\underline{Y},\end{aligned}$$

by Lemma 4.1, part 3(ii). Therefore

$$\begin{aligned}\hat{\sigma}^2 &= f^{-1}[\underline{Y}'T^{-}\underline{Y} - \hat{\underline{\beta}}'X'T^{-}\underline{Y}] \\ &= f^{-1}[\underline{Y}'T^{-}\underline{Y} - \{(X'T^{-}X)^{-}X'T^{-}\underline{Y}\}'X'T^{-}\underline{Y}] \\ &= f^{-1}[\underline{Y}'T^{-}\underline{Y} - \underline{Y}'(T^{-})'X(X'T^{-}X)^{-}X'T^{-}\underline{Y}] \\ &= f^{-1}\underline{Y}'A_0\underline{Y}\end{aligned}\tag{A9.1}$$

where $A_0 = T^{-} - (T^{-})'X(X'T^{-}X)^{-}X'T^{-}$. Using equation (A9.1),

and applying Theorem 1(i) (Searle, 1971, p.55), we can write the expected value of $\hat{\sigma}^2$ as

$$\begin{aligned}
 E(\hat{\sigma}^2) &= f^{-1} E[\underline{Y}' A_0 \underline{Y}] \\
 &= f^{-1} [\text{trace}\{A_0 \sigma^2 G\} + E(\underline{Y})' A_0 E(\underline{Y})] \\
 &= f^{-1} \text{trace}[A_0 \sigma^2 G], \tag{A9.2}
 \end{aligned}$$

since

$$\begin{aligned}
 E(\underline{Y})' A_0 E(\underline{Y}) &= \underline{\beta}' X' [T^- - (T^-)' X (X' T^- X)^- X' T^-] X \underline{\beta} \\
 &= \underline{\beta}' X' T^- X \underline{\beta} - \underline{\beta}' (X' T^- X) (X' T^- X)^- (X' T^- X) \underline{\beta} \\
 &= 0
 \end{aligned}$$

as $X' (T^-)' X = X' T^- X$, because T is symmetric and $X' T^- X$ is unique (see proof of Theorem 4.1(ii)). Thus

$$\begin{aligned}
 E(\hat{\sigma}^2) &= f^{-1} \text{trace}[A_0 \sigma^2 G] \\
 &= \sigma^2 f^{-1} \text{trace}[A_0 G] \\
 &= \sigma^2 f^{-1} \text{trace}[A_0 (T - XX')].
 \end{aligned}$$

By writing A_0 as in Eq. (A9.1), we get

$$\begin{aligned}
E(\hat{\sigma}^2) &= \sigma^2 f^{-1} \text{ trace } [\{T^- - (T^-)'X(X'T^-X)^{-}X'T^-\}\{T - XX'\}] \\
&= \sigma^2 f^{-1} \text{ trace } [T^-T - T^-XX' - (T^-)'X(X'T^-X)^{-}X'T^-T \\
&\quad + (T^-)'X(X'T^-X)^{-}X'T^-XX'] \\
&= \sigma^2 f^{-1} [\text{trace } T^-T - \text{trace } T^-XX' \\
&\quad - \text{trace } (T^-)'X(X'T^-X)^{-}X' + \text{trace } (T^-)'XX'],
\end{aligned}$$

by Lemma 4.1, parts 1 and 3, and so

$$E(\hat{\sigma}^2) = \sigma^2 f^{-1} [\text{trace } T^-T - \text{trace } (X'T^-X)^{-}(X'T^-X)],$$

since $X'(T^-)'X = X'T^-X$. Since T^-T and $(X'T^-X)^{-}(X'T^-X)$ are idempotent, and $\text{rank}(AA^-) = \text{rank}(A)$ for any matrix A , we see that

$$E(\hat{\sigma}^2) = \sigma^2 f^{-1} [\text{rank}(T) - \text{rank}(X'T^-X)]$$

and by Lemma 4.1 part 2 we have

$$E(\hat{\sigma}^2) = \sigma^2 f^{-1} [\text{rank}(T) - \text{rank}(X)].$$

Finally, since $f = \text{rank}(G:X) - \text{rank}(X)$, we can write

$$E(\hat{\sigma}^2) = \sigma^2 f^{-1} [\text{rank}(G:X) - \text{rank}(X)] = \sigma^2. \quad (\text{A9.3})$$

The proof of Theorem 4.1(i) is now completed by justifying the equality in (A9.3) by showing that $\text{rank}(T) = \text{rank}(G:X)$. First we write

$$\text{rank}(T) = \text{rank}(G + XX').$$

Replacing G by VV' , we have

$$\begin{aligned} \text{rank}(T) &= \text{rank}(VV' + XX') \\ &= \text{rank}(CC'), \text{ where } C = (V:X) \\ &= \text{rank}(C) \\ &= \text{rank}(V:X) \\ &= \text{rank}(G:X), \end{aligned}$$

since the column space of G is the same as the column space of V . The column space of G is the same as the column space of V if the columns of V belong to the column space of G , and vice versa, if the columns of G belong to the column space of V . Symbolically, this is written as

$V \subset C(G)$, and $G \subset C(V)$. To show that $V \subset C(G)$, it is sufficient to show that $GG^{-1}V = V$, but this is true because $GG^{-1}V = (VV')(VV')^{-1}V = V$. Now, $G \subset C(V)$ since by definition $VV' = G$.

APPENDIX 10
PROOF OF THEOREM 4.1(ii)

In Appendix 10 we prove part (ii) of Theorem 4.1, thus we show that $\hat{\sigma}^2 = f^{-1}(\underline{Y} - \underline{X}\hat{\beta})'T^{-}(\underline{Y} - \underline{X}\hat{\beta})$ is unique with probability one. The following theorem will be useful in our proof.

Theorem vi(c) (Rao, 1973, p.26).

Let B and D be non-null matrices. Then $BA^{-}D$ is invariant for any choice of A^{-} if and only if $C(B') \subset C(A')$ and $C(D) \subset C(A)$, where $C(\cdot)$ denotes column space.

The relationship $C(B') \subset C(A')$ holds if and only if $BA^{-}A = B$, and similarly $C(D) \subset C(A)$ holds if and only if $AA^{-}D = D$ (see Pringle and Rayner, 1971, p.9). Since the quantity $\hat{\sigma}^2$ is written as

$$\hat{\sigma}^2 = f^{-1}[\underline{Y}'T^{-}\underline{Y} - \underline{Y}'(T^{-})'X(X'T^{-}X)^{-}X'T^{-}\underline{Y}],$$

to show that $\hat{\sigma}^2$ is unique with probability one, it suffices to show that

$$\underline{Y}'T^{-}\underline{Y} - \underline{Y}'(T^{-})'X(X'T^{-}X)^{-}X'T^{-}\underline{Y} \quad (A10.1)$$

is invariant with probability one to the choice of the generalized inverses involved.

First we show that $\underline{Y}'T^-\underline{Y}$ is unique with probability one (w.p.1). From part 4 of Lemma 4.1, $\underline{Y} \in C(T)$, w.p.1.

Therefore $\underline{Y}' \in C(T')$, w.p.1, since T is symmetric, and then by Theorem vi(c) (Rao, 1973, p.26), $\underline{Y}'T^-\underline{Y}$ is unique, w.p.1.

Secondly we show that $\underline{Y}'(T^-)'X(X'T^-X)^-X'T^-\underline{Y}$ is unique with probability one in the following four part proof.

(1) Show $X'T^-X$ is unique.

From part 1(i) of Lemma 4.1, $TT^-X = X$ and thus $X \in C(T)$.

Since T is symmetric, we have $X \in C(T')$. By Theorem vi(c) (Rao, 1973, p.26), $X'T^-X$ is unique.

(2) Show $X'T^-\underline{Y}$ is unique, w.p.1.

By (1) above, $X \in C(T')$ and by part 4 of Lemma 4.1, $\underline{Y} \in C(T)$, w.p.1. Thus applying Theorem vi(c) (Rao, 1973, p.26), $X'T^-\underline{Y}$ is unique, w.p.1.

(3) Show $\underline{Y}'(T^-)'X$ is unique, w.p.1.

This follows from part (2), since $\underline{Y}'(T^-)'X$ is equal to the transpose of $X'T^-\underline{Y}$, which was shown in (2) to be unique, w.p.1.

(4) Using (1), (2), and (3) above, the second quantity in (A10.1) is unique, w.p.1, by Theorem vi(c) (Rao, 1973, p.26) if

(a) $[\underline{Y}'(T^-)'X]' \in C[(X'T^-X)']$, w.p.1, and

(b) $X'T^-\underline{Y} \in C(X'T^-X)$, w.p.1.

Part (a) is true not only with probability one but always because $\underline{Y}'(T^-)'X(X'T^-X)^-(X'T^-X) = \underline{Y}'(T^-)'X$, since by part

3(i) of Lemma 4.1, $X = X(X'T^{-1}X)^{-1} (X'T^{-1}X)$. Part (b) is true not only w.p.1 but always because

$(X'T^{-1}X)(X'T^{-1}X)^{-1} X'T^{-1}\underline{Y} = X'T^{-1}\underline{Y}$, by part 3(ii) of Lemma 4.1.

Therefore we have shown that both $\underline{Y}'T^{-1}\underline{Y}$ and $\underline{Y}'(T^{-1})'X(X'T^{-1}X)^{-1}X'T^{-1}\underline{Y}$ are unique with probability one, which allows us to conclude that $\hat{\sigma}^2$ is unique with probability one.

APPENDIX 11
PROOF OF THEOREM 4.1(iii)

In this appendix we prove Theorem 4.1(iii), that is we show that if $\underline{y}$ possesses an N-variate normal distribution then $\hat{f}_\sigma^2/\sigma^2 \sim \chi_f^2$, where $f = \text{rank}(G:X) - \text{rank}(X)$. Recall that $\hat{\sigma}^2 = f^{-1} \underline{y}' A_0 \underline{y}$ where $A_0 = T^- - (T^-)' X (X' T^- X)^- X' T^-$. Since we have shown in Theorem 4.1(ii) that $\hat{\sigma}^2$ is unique with probability one, the choice of the generalized inverses in the expression for $\hat{\sigma}^2$ may be made arbitrarily. Thus we choose each of the generalized inverses to be the unique Moore-Penrose inverse, and we denote the unique Moore-Penrose inverse of a matrix B by B^+ . The Moore-Penrose inverse has the following four properties (see Searle, 1971, p.16):

1. $BB^+B = B$
2. $B^+BB^+ = B^+$
3. $(BB^+)' = BB^+$
4. $(B^+B)' = B^+B$.

The quantity $\hat{f}_\sigma^2/\sigma^2$ can be expressed as

$$\hat{f}_\sigma^2/\sigma^2 = \underline{y}' A \underline{y}, \quad (\text{A11.1})$$

where $A = (1/\sigma^2) [T^+ - T^+ X (X' T^+ X)^+ X' T^+]$. We wish to show

that $\underline{y}'A\underline{y} \sim \chi^2_f$, which can be done by making use of the following corollary.

Corollary 2s.1 (Searle, 1971, p.69).

When $\underline{x}$ is $N(\underline{\mu}, V)$ whether V be singular or non-singular, $\underline{x}'A\underline{x} \sim \chi'^2$ with degrees of freedom equal to $\text{trace}(AV)$ and noncentrality parameter equal to $(1/2)\underline{\mu}'A\underline{\mu}$, where χ'^2 denotes a noncentral chi-square random variable, if and only if

- (i) $VAVAV = VAV$
- (ii) $\underline{\mu}'AV = \underline{\mu}'AVAV$, and
- (iii) $\underline{\mu}'A\underline{\mu} = \underline{\mu}'AVA\underline{\mu}$.

In our application, the matrices A and V in Corollary 2s.1 (Searle, 1971, p.69) are defined as

$$A = (1/\sigma^2)[I - T^+X(X'T^+X)^+X']T^+,$$

and

$$V = \sigma^2 G.$$

The proof of Theorem 4.1(iii) follows from Corollary 2s.1 (Searle, 1971, p.69) if we can show that $AVA = A$. To show that $AVA = A$, we first show that $AG = AT$, where as we recall, $AG = A(T - XX')$. Thus $AG = AT$ if $AXX' = 0$. Using the complete expression for A , we have

$$\begin{aligned}
AXX' &= (1/\sigma^2)[T^+ - T^+X(X'T^+X)^+X'T^+]XX' \\
&= (1/\sigma^2)[T^+XX' - T^+X(X'T^+X)^+(X'T^+X)X'],
\end{aligned}$$

and so by Lemma 4.1 part 3(i),

$$\begin{aligned}
AXX' &= (1/\sigma^2)[T^+XX' - T^+XX'] \\
&= 0.
\end{aligned}$$

Therefore, since $AG = AT$, we have $AVA = \sigma^2AGA = \sigma^2ATA$. We now show that $\sigma^2ATA = A$:

$$\begin{aligned}
\sigma^2ATA &= (1/\sigma^2)[I - T^+X(X'T^+X)^+X']T^+T[I - T^+X(X'T^+X)^+X']T^+ \\
&= (1/\sigma^2)[T^+T - T^+X(X'T^+X)^+X'T^+T][T^+ - T^+X(X'T^+X)^+X'T^+] \\
&= (1/\sigma^2)[T^+TT^+ - T^+X(X'T^+X)^+X'T^+TT^+ \\
&\quad - T^+TT^+X(X'T^+X)^+X'T^+ \\
&\quad + T^+X(X'T^+X)^+X'T^+TT^+X(X'T^+X)^+X'T^+] \\
&= (1/\sigma^2)[T^+ - T^+X(X'T^+X)^+X'T^+ - T^+X(X'T^+X)^+X'T^+ \\
&\quad + T^+X(X'T^+X)^+X'T^+X(X'T^+X)^+X'T^+]
\end{aligned}$$

since $T^+TT^+ = T^+$, by property 2 of the Moore-Penrose inverse. Therefore,

$$\begin{aligned}\sigma^2 ATA &= (1/\sigma^2)[T^+ - 2T^+X(X'T^+X)^+X'T^+ + T^+X(X'T^+X)^+X'T^+] \\ &= (1/\sigma^2)[I - T^+X(X'T^+X)^+X']T^+ \\ &= A.\end{aligned}$$

Since we have verified that $AVA = \sigma^2 ATA = A$ we can conclude that $\hat{f}\sigma^2/\sigma^2 = \underline{Y}'A\underline{Y} \sim \chi_f^2$, by Corollary 2s.1 (Searle, 1971, p.69). The quantity $\hat{f}\sigma^2/\sigma^2 \sim \chi_f^2$ and not $\chi_f'^2$, since the noncentrality parameter equals zero, which we now show.

The noncentrality parameter, from Corollary 2s.1 (Searle, 1971, p.69) is of the form $(1/2)\underline{\mu}'A\underline{\mu}$, where in our application, $\underline{\mu} = X\underline{\beta}$. Thus,

$$\begin{aligned}\underline{\mu}'A\underline{\mu} &= \underline{\beta}'X'AX\underline{\beta} \\ &= (1/\sigma^2)\underline{\beta}'X'[T^+ - T^+X(X'T^+X)^+X'T^+]X\underline{\beta} \\ &= (1/\sigma^2)[\underline{\beta}'X'T^+X\underline{\beta} - \underline{\beta}'X'T^+X(X'T^+X)^+(X'T^+X)\underline{\beta}]\end{aligned}$$

and so by Lemma 4.1 part 3(i),

$$\begin{aligned}\underline{\mu}'A\underline{\mu} &= (1/\sigma^2)[\underline{\beta}'X'T^+X\underline{\beta} - \underline{\beta}'X'T^+X\underline{\beta}] \\ &= 0.\end{aligned}$$

We now verify that the degrees of freedom are $f = \text{rank}(G:X) - \text{rank}(X)$. From Corollary 2s.1 (Searle, 1971, p.69) the degrees of freedom associated with $\underline{Y}'A\underline{Y}$ are equal to $f = \text{trace}(\sigma^2AG)$, and so

$$\begin{aligned}\text{trace}(\sigma^2AG) &= \text{trace}[I - T^+X(X'T^+X)^+X']T^+G \\ &= \text{trace}(T^+G) - \text{trace}[T^+X(X'T^+X)^+X'T^+G] \\ &= \text{trace}(T^+T - T^+XX') - \text{trace}[T^+X(X'T^+X)^+X'T^+T] \\ &\quad + \text{trace}[T^+X(X'T^+X)^+X'T^+XX'],\end{aligned}$$

since $G = T - XX'$. It follows that

$$\begin{aligned}\text{trace}(\sigma^2AG) &= \text{trace } T^+T - \text{trace } T^+XX' - \text{trace } X(X'T^+X)^+X'T^+ \\ &\quad + \text{trace } T^+XX',\end{aligned}$$

since $\text{trace } AB = \text{trace } BA$ for arbitrary matrices A and B , $T^+TT^+ = T^+$, and $X(X'T^+X)^+(X'T^+X) = X$ by Lemma 4.1 part 3(ii). Therefore

$$\begin{aligned}\text{trace}(\sigma^2AG) &= \text{trace } T^+T - \text{trace } (X'T^+X)(X'T^+X)^+ \\ &= \text{rank}(T) - \text{rank}(X'T^+X),\end{aligned}$$

since TT^+ and $(X'T^+X)(X'T^+X)^+$ are idempotent, and $\text{rank}(AA^+) = \text{rank } A$, for any matrix A . Finally, by Lemma 4.1 part 2 we have

$$\text{trace}(\sigma^2 AG) = \text{rank}(T) - \text{rank}(X)$$

and by the argument in the proof of Theorem 4.1(i),

$$\text{trace}(\sigma^2 AG) = \text{rank}(G:X) - \text{rank}(X).$$

APPENDIX 12
PROOF OF THEOREM 4.2

In this appendix we prove Theorem 4.2, thus we show that when $\underline{y} \sim N_N(\underline{X}\underline{\beta} + \underline{X}_{2-2}\underline{\beta}_{2-2}, \sigma^2 G)$ then $\hat{f}\sigma^2/\sigma^2 \sim \chi_{\hat{f},\lambda}^2$, where $\lambda = (1/2\sigma^2)\underline{\beta}'_2\underline{X}'_2[T^- - T^-X(X'T^-X)^-X'T^-]\underline{X}_{2-2}\underline{\beta}_{2-2}$.

From the proof of Theorem 4.1, we have $\hat{f}\sigma^2/\sigma^2 \sim \chi_{\hat{f},\lambda}^2$. By Corollary 2s.1 (Searle, 1971, p.69) the noncentrality parameter is

$$\lambda = (1/2)(\underline{X}\underline{\beta} + \underline{X}_{2-2}\underline{\beta}_{2-2})'A(\underline{X}\underline{\beta} + \underline{X}_{2-2}\underline{\beta}_{2-2}),$$

where $A = (1/\sigma^2)[T^- - T^-X(X'T^-X)^-X'T^-]$. Thus

$$\lambda = (1/2)[\underline{\beta}'_2\underline{X}'_2A\underline{\beta}_2 + \underline{\beta}'_2\underline{X}'_2A\underline{X}_{2-2}\underline{\beta}_{2-2} + \underline{\beta}'_2\underline{X}'_2A\underline{\beta}_2 + \underline{\beta}'_2\underline{X}'_2A\underline{X}_{2-2}\underline{\beta}_{2-2}].$$

From the proof of Theorem 4.1(iii), $\underline{\beta}'_2\underline{X}'_2A\underline{\beta}_2 = 0$. We now show that $\underline{\beta}'_2\underline{X}'_2A\underline{\beta}_2 = 0$:

$$\begin{aligned}\underline{\beta}'_2\underline{X}'_2A\underline{\beta}_2 &= \underline{\beta}'_2\underline{X}'_2[T^- - T^-X(X'T^-X)^-X'T^-]\underline{\beta}_2/\sigma^2 \\ &= \underline{\beta}'_2\underline{X}'_2[T^-\underline{\beta}_2 - T^-X(X'T^-X)^-(X'T^-X)\underline{\beta}_2]/\sigma^2,\end{aligned}$$

and so by Lemma 4.1 part 3(i),

$$\begin{aligned}\underline{\beta}'_2\underline{X}'_2A\underline{\beta}_2 &= \underline{\beta}'_2\underline{X}'_2[T^-\underline{\beta}_2 - T^-\underline{\beta}_2]/\sigma^2 \\ &= 0.\end{aligned}$$

Thus we conclude that

$$\lambda = (1/2)\beta_2' X_2' A X_2 \beta_2$$

$$= (1/2\sigma^2)\beta_2' X_2' [T^- - T^- X (X' T^- X)^{-1} X' T^-] X_2 \beta_2.$$

APPENDIX 13
PROOF OF THE EQUALITY $SSE_A = \underline{d}'V_0^{-1}\underline{d} + SSE$

In this appendix we show that the check point method for testing a fitted model for lack of fit and the method in which the residual sum of squares is partitioned into a lack of fit sum of squares and a pure error sum of squares are equivalent in the sense that $SSE_A = \underline{d}'V_0^{-1}\underline{d} + SSE$.

Let us define $\underline{y}_A$ and x_A as

$$\underline{y}_A = \begin{bmatrix} \underline{y} \\ \underline{y}^* \end{bmatrix} \quad (A13.1)$$

and

$$x_A = \begin{bmatrix} x \\ x^* \end{bmatrix} . \quad (A13.2)$$

Then the residual sum of squares from regressing $\underline{y}_A$ on x_A is

$$SSE_A = \underline{y}_A' [I - x_A (x_A' x_A)^{-1} x_A'] \underline{y}_A .$$

Using Eqs. (A13.1) and (A13.2) we can write SSE_A as

$$\begin{aligned}
 SSE_A &= \begin{bmatrix} \underline{Y} \\ \underline{\bar{Y}^*} \end{bmatrix}' \left[I - \begin{pmatrix} X \\ X^* \end{pmatrix} (X'X + X^{*'}X^*)^{-1} \begin{pmatrix} X \\ X^* \end{pmatrix}' \right] \begin{bmatrix} \underline{Y} \\ \underline{\bar{Y}^*} \end{bmatrix} \\
 &= \underline{Y}^{*'} [I - X^*(X'X + X^{*'}X^*)^{-1}X^{*'}] \underline{Y}^* \\
 &\quad - 2 \underline{Y}^{*'} X^* (X'X + X^{*'}X^*)^{-1} X' \underline{Y} \\
 &\quad + \underline{Y}' [I - X(X'X + X^{*'}X^*)^{-1}X'] \underline{Y} \\
 &= \underline{Y}^{*'} V_0^{-1} \underline{Y}^* - 2 \underline{Y}^{*'} X^* (X'X + X^{*'}X^*)^{-1} X' \underline{Y} \\
 &\quad + \underline{Y}' [I - X(X'X + X^{*'}X^*)^{-1}X'] \underline{Y}. \tag{A13.3}
 \end{aligned}$$

Eq. (A13.3) is true because from Eq. (8) (Morrison, 1976, p. 69) we can write V_0^{-1} as

$$\begin{aligned}
 V_0^{-1} &= [I + X^*(X'X)^{-1}X^{*'}]^{-1} \\
 &= I - X^*(X'X + X^{*'}X^*)^{-1}X^{*'}.
 \end{aligned}$$

We now write the quadratic form $\underline{d}'V_0^{-1}\underline{d}$ as

$$\begin{aligned}
 \underline{d}'V_0^{-1}\underline{d} &= (\underline{y}^* - \hat{\underline{y}}^*)'V_0^{-1}(\underline{y}^* - \hat{\underline{y}}^*) \\
 &= \underline{y}^{*'}V_0^{-1}\underline{y}^* - 2\underline{y}^{*'}V_0^{-1}\hat{\underline{y}}^* + \hat{\underline{y}}^{*'}V_0^{-1}\hat{\underline{y}}^* \\
 &= \underline{y}^{*'}V_0^{-1}\underline{y}^* - 2\underline{y}^{*'}V_0^{-1}X^*(X'X)^{-1}X'\underline{y} \\
 &\quad + \underline{y}'X(X'X)^{-1}X^{*'}V_0^{-1}X^*(X'X)^{-1}X'\underline{y}. \quad (A13.4)
 \end{aligned}$$

The first portion in Eq. (A13.3) is equal to the first portion in Eq. (A13.4). We now show that the second portions of Eqs. (A13.3) and (A13.4) are equal. It can be verified using Eq. (8) (Morrison, 1976, p. 69) that

$$(X'X + X^{*'}X^*)^{-1} = (X'X)^{-1} - (X'X)^{-1}X^{*'}V_0^{-1}X^*(X'X)^{-1}. \quad (A13.5)$$

Using Eq. (A13.5) the second portion of SSE_A in Eq. (A13.3) can be written as

$$\begin{aligned}
 &- 2\underline{y}^{*'}X^*(X'X + X^{*'}X^*)^{-1}X'\underline{y} \\
 &= -2\underline{y}^{*'}X^*[(X'X)^{-1} - (X'X)^{-1}X^{*'}V_0^{-1}X^*(X'X)^{-1}]X'\underline{y} \\
 &= -2\underline{y}^{*'}[I - X^*(X'X)^{-1}X^{*'}V_0^{-1}]X^*(X'X)^{-1}X'\underline{y}. \quad (A13.6)
 \end{aligned}$$

The second portion of SSE_A , given in Eq. (A13.6), is seen to equal the second portion of $\underline{d}'V_0^{-1}\underline{d}$ in Eq. (A13.4) using the fact that

$$\begin{aligned} I - X^*(X'X)^{-1}X^*V_0^{-1} &= I - (V_0 - I)V_0^{-1} \\ &= V_0^{-1}. \end{aligned}$$

We now show that the third portion of the expression for SSE_A in Eq. (A13.3) is equal to the sum of the third portion of $\underline{d}'V_0^{-1}\underline{d}$ in Eq. (A13.4) and SSE , where $SSE = \underline{y}'[I - X(X'X)^{-1}X']\underline{y}$. Using the result in Eq. (A13.5), the third portion of SSE_A in Eq. (A13.3) can be written as

$$\begin{aligned} &\underline{y}'[I - X(X'X + X^*X^*)^{-1}X']\underline{y} \\ &= \underline{y}'[I - X\{(X'X)^{-1} - (X'X)^{-1}X^*V_0^{-1}X^*(X'X)^{-1}\}X']\underline{y} \\ &= \underline{y}'[I - X(X'X)^{-1}X']\underline{y} + \underline{y}'X(X'X)^{-1}X^*V_0^{-1}X^*(X'X)^{-1}X'\underline{y}. \end{aligned}$$

Therefore, since the first two portions of SSE_A in Eq. (A13.3) are equal to the first two portions of $\underline{d}'V_0^{-1}\underline{d}$ in Eq. (A13.4), respectively, and the third portion of SSE_A in Eq. (A13.3) is equal to the sum of the third portion of $\underline{d}'V_0^{-1}\underline{d}$

in Eq. (A13.4) and $\underline{Y}'[I - X(X'X)^{-1}X']\underline{Y}$, we must have then

$$SSE_A = \underline{d}'V_0^{-1}\underline{d} + \underline{Y}'[I - X(X'X)^{-1}X']\underline{Y}$$

$$= \underline{d}'V_0^{-1}\underline{d} + SSE.$$

REFERENCES

- Atkinson, A. C. (1972). Planning experiments to detect inadequate regression models. Biometrika, Vol. 59, pp. 275-293.
- Barr, A. J., J. H. Goodnight, and J. P. Sall (1979). SAS Users Guide, 1979 Edition. SAS Institute, Inc., Raleigh, North Carolina.
- Becker, N.G. (1968). Models for the response of a mixture. Journal of the Royal Statistical Society, B, Vol. 30, pp. 349-358.
- Becker, N.G. (1978). Models and designs for experiments with mixtures. Australian Journal of Statistics, Vol. 20, pp. 195-208.
- Box, G. E. P., and N. R. Draper (1975). Robust designs. Biometrika, Vol. 62, pp. 347-352.
- Claringbold, P. J. (1955). Use of the simplex design in the study of the joint action of related hormones. Biometrics, Vol. 11, pp. 174-185.
- Cochran, W. G., and G. M. Cox (1957). Experimental Designs, 2nd Ed. John Wiley and Sons, New York.
- Cornell, J. A. (1973). Experiments with mixtures: A review. Technometrics, Vol. 15, pp. 437-455.
- Cornell, J. A. (1981). Experiments with Mixtures: Designs, Models, and the Analysis of Mixture Data. John Wiley and Sons, New York.
- Cornell, J. A., and I. J. Good (1970). The mixture problem for categorized components. Journal of the American Statistical Association, Vol. 65, pp. 339-355.
- Daniel, C., and F. S. Wood (1971). Fitting Equations to Data: Computer Analysis of Multifactor Data for Scientists and Engineers. John Wiley and Sons, New York.
- Draper N. R., and A. M. Herzberg (1971). On lack of fit. Technometrics, Vol. 13, pp. 231-241.

- Draper, N. R., and W. E. Lawrence (1965a). Mixture designs for three factors. Journal of the Royal Statistical Society, B, Vol. 27, pp. 450-465.
- Draper, N. R., and W. E. Lawrence (1965b). Mixture designs for four factors. Journal of the Royal Statistical Society, B, Vol. 27, pp. 473-478.
- Draper, N. R., and H. Smith (1981). Applied Regression Analysis, 2nd Ed. John Wiley and Sons, New York.
- Draper, N. R., and R. C. St. John (1977). A mixtures model with inverse terms. Technometrics, Vol. 19, pp. 37-46.
- Gorman, J. W., and J. E. Hinman (1962). Simplex-lattice designs for multicomponent systems. Technometrics, Vol. 4, pp. 463-487.
- Graybill, F. A. (1969). Introduction to Matrices with Applications in Statistics. Wadsworth Publishing Co., Inc., Belmont, California.
- Graybill, F. A. (1976). Theory and Application of the Linear Model. Duxbury Press, North Scituate, Massachusetts.
- Green, J. R. (1971). Testing departure from a regression, without using replicaton. Technometrics, Vol. 13, pp. 609-615.
- Johnson, N. L., and S. Kotz (1970). Distributions in Statistics: Continuous Univariate Distributions - 2. Houghton Mifflin, Boston.
- Kenworthy, O. O. (1963). Factorial experiments with mixtures using ratios. Industrial Quality Control, Vol. 19, pp. 24-26.
- Khuri, A. I., and J. A. Cornell (1981). Lack of fit revisited. Technical Report No. 167, Department of Statistics, University of Florida, Gainesville, Florida.
- Kurotori, I. S. (1966). Experiments with mixtures of components having lower bounds. Industrial Quality Control, Vol. 22, pp. 592-596.
- McLean, R. A., and V. L. Anderson (1966). Extreme vertices design of mixture experiments. Technometrics, Vol. 8, pp. 447-454.
- Morrison, D. F. (1976). Multivariate Statistical Methods, 2nd Ed. McGraw-Hill, New York.

- Mudholkar, G. S., Y. P. Chaubey, and C. Lin (1976). Approximations for the doubly noncentral-F distribution. Communications in Statistics, A, Vol. 5, pp. 49-63.
- Myers, R. H. (1971). Response Surface Methodology. Allyn and Bacon, Inc., Boston.
- Piepel, G. F. (1981). Measuring component effects in constrained mixture experiments. Unpublished manuscript presented at 1981 Joint Statistical Meetings of the American Statistical Association and the Biometric Societies, Detroit, Michigan, August 1981.
- Price, W. L. (1977). A controlled random search procedure for global optimisation. The Computer Journal, Vol. 20, pp. 367-370.
- Pringle, R. M., and A. A. Rayner (1971). Generalized Inverse Matrices with Applications to Statistics. Hafner Publishing Company, New York.
- Rao, C. R. (1971). Unified theory of linear estimation. Sankhyā, A, Vol. 33, pp. 371-394.
- Rao, C. R. (1972). Unified theory of least squares. Communications in Statistics, Vol. 1, pp. 1-8.
- Rao, C. R. (1973). Linear Statistical Inference and its Applications, 2nd Ed. John Wiley and Sons, New York.
- Rao, C. R., and S. K. Mitra (1971). Generalized Inverse of Matrices and its Applications. John Wiley and Sons, New York.
- Scheffé, H. (1958). Experiments with mixtures. Journal of the Royal Statistical Society, B, Vol. 20, pp. 344-360.
- Scheffé, H. (1959). The Analysis of Variance. John Wiley and Sons, New York.
- Scheffé, H. (1963). The simplex-centroid design for experiments with mixtures. Journal of the Royal Statistical Society, B, Vol. 25, pp. 235-263.
- Searle, S. R. (1971). Linear Models. John Wiley and Sons, New York.
- Shillington, E. R. (1979). Testing lack of fit in regression without replication. The Canadian Journal of Statistics, Vol. 7, pp. 137-146.

- Snee, R. D. (1971). Design and analysis of mixture experiments. Journal of Quality Technology, Vol. 3, pp. 159-169.
- Snee, R. D. (1973). Techniques for the analysis of mixture data. Technometrics, Vol. 15, pp. 517-528.
- Snee, R. D. (1977). Validation of regression models: methods and examples. Technometrics, Vol. 19, pp. 415-428.
- Thompson, W. O., and R. H. Myers (1968). Response surface designs for experiments with mixtures. Technometrics, Vol. 10, pp. 739-756.
- Wishart, D. (1975). Clustan 1C User Manual. David Wishart, University College, London.
- Zyskind, G. (1967). On canonical forms, non-negative covariance matrices and best and simple least squares linear estimators in linear models. Annals of Mathematical Statistics, Vol. 38, pp. 1092-1109.

BIOGRAPHICAL SKETCH

John Thomas Shelton was born on March 30, 1952, in Jacksonville, Florida, where he resided until graduating from Englewood High School in June, 1970. He then entered the University of Florida where he received a Bachelor of Science degree in mathematics in June, 1974.

John began graduate study at Virginia Polytechnic Institute and State University in Blacksburg, Virginia, in September, 1975, and there received a Master of Science degree in statistics in the summer of 1976. After two years as a Research Associate at Auburn University in Auburn, Alabama, he returned to the University of Florida in September, 1978, where he has since been pursuing a doctoral degree in statistics. While a graduate student at the University of Florida, John has worked as a Graduate Assistant performing statistical consulting duties in the School of Forest Resources and Conservation.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Andre' I. Khuri, Chairman
Assistant Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

John A. Cornell, Cochairman
Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Richard F. Fisher
Professor of Forest Resources
and Conservation

This dissertation was submitted to the Graduate Faculty of the Department of Statistics in the College of Liberal Arts and Sciences and to the Graduate Council, and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

May, 1982

Dean for Graduate Studies
and Research