BIASED ESTIMATION USING RIDGE REGRESSION TECHNIQUES

BY

PHILEMON O. WACHARA

A dissertation submitted in partial fulfilment for the degree of Master of Science in Mathematical Statistics in the Department of Mathematics

UNIVERSITY OF NAIROBI

JULY, 1994

DECLARATION

This dissertation is my own work and has not been presented for a degree in any other university

signature.....

WACHARA P . O

This dissertation has been submitted for examination with my approval as the university supervisor

signature.....

DR. F NJUI

Department of Mathematics

University of Nairobi

P.O.BOX 30197

Nairobi, Kenya

To Lynette .O. Adhiambo

you are the wind beneath my wings

SUMMARY OF CONTENTS

In this dissertation emphasis is placed on the practicability of a non-iterative procedure for generalized ridge regression as given by Hemmerle (1975) and its application on third order rotatable designs.

In chapter I, we outline basic concepts on the theory of regression analysis and give an introduction on the principles of biased estimation and ridge regression techniques. A statement of the problem of study is then given at the end of the chapter.

Chapter II examines the minimum mean square error estimator (MIMSEE) in ridge regression in its general form and describes a non-iterative solution for generalized ridge regression as given by Hemmerle (1975).

Chapter III gives a general review of rotatable designs of up to third order, giving particular consideration to the estimation of the coefficients of a third order rotatable design.

The application of ridge regression on third order rotatable designs is considered in chapter IV, and an illustration of its practicability given.

ACKNOWLEDGEMENT

I am grateful to the many people who, in one way or another, have contributed to make this study a success.

Special thanks go to my supervisor, Dr. F Njui for sacrificing a lot of his time to read through my work and for his valuable suggestions and advice.

I would also like to express my appreciation to my lecturers for effectively disbursing the knowledge which triggered my interest in the field of statistics.

I am indebted to my colleagues; Kipchirchir, Nyapola, and Obonyo for their encouragement and company.

Finally, I wish to express my heartfelt thanks to my parents, Ezra Obuya and Catherine Awiti for sacrificing so much, that I may have tomorrow what they could not have yesterday.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

1.1 Introduction

In order to make a decision it is often necessary to prepare a forecast. In order to prepare a budget it is necessary to predict revenues. These and other decisions can be made easier if a relationship can be established between the variable to be predicted and some other variables that are either known or are significantly easier to anticipate. For example, in an industrial situation it may be known that the tar content in the outlet stream in a chemical process is related to the inlet temperature and it may be of interest to develop a procedure for estimating the tar content for various levels of the inlet temperature from experimental information. The statistical aspect of the problem then becomes one of arriving at the best estimate of the relationship between the variables. This leads us therefore to regression analysis. The relationship is expressed in form of an equation connecting the response or dependent variable y, and one or more independent variables, $x_1, x_2, \ldots, x_p$, which are measured with negligible error and are often controlled in the experiment. The relationship, fitted to a set of experimental data, is characterized by a prediction equation called a regression equation.

31.2 The general linear model and least squares estimation

Suppose the experimental data consists of n observations on the response variable y and independent variables $x_1, \ldots, x_p$. Then the relationship between y and $x_1, \ldots, x_p$ can be formulated as a linear model

$$y_j = \beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j} + \ldots + \beta_p x_{pj} + \varepsilon_j \qquad j=1, 2, \ldots, n \quad (1.2.1)$$

where $\beta_0, \beta_1, \ldots, \beta_p$ are constants called the model regression parameters and ε is a random disturbance for which it is assumed that for every fixed value of $x_1, x_2, \ldots, x_p$, the ε's are independently distributed with mean zero and a common variance σ^2.

More compactly, (1.2.1) can be represented in matrix notation as;

$$y = X\beta + \varepsilon$$

where

$$y = [y_1, y_2, \ldots, y_n]',$$

$$\beta = [\beta_0, \beta_1, \ldots, \beta_p]' \quad , \quad \varepsilon = [\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n]'$$

and

$$X = \begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{p1} \\ 1 & x_{12} & x_{22} & \cdots & x_{p2} \\ \vdots & \vdots & \vdots & \vdots\vdots\vdots & \vdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{pn} \end{bmatrix}$$

(1.2.2)

The nx(p+1) matrix is called the regression matrix and is assumed to be of full rank (p+1). Further assumptions are

$$E(\varepsilon)=0 \ , \text{ and } \ E(\varepsilon\varepsilon')= \sigma^2 I_n \ . \qquad (1.2.3)$$

We use the method of least squares to obtain an estimate of β. This method entails minimizing $\Sigma\varepsilon^2_i$, that is the sum of squares due to the random disturbances, or the residual sum of squares (R.S.S) with respect to β. From (1.2.2) we can write

$$\begin{aligned} \Sigma\varepsilon^2_i &= \varepsilon'\varepsilon \\ &= (y - X\beta)'(y - X\beta) \\ &= y'y - \beta'X'y - y'X\beta + \beta'X'X\beta \\ &= y'y - 2\beta'X'y + \beta'X'X\beta \end{aligned} \qquad (1.2.4)$$

The least squares estimate of β is then the value b which minimizes $\varepsilon'\varepsilon$, is determined by differentiating equation (1.2.4) with respect to β and setting the resultant matrix equation equal to zero. This gives,

$$-2X'y + 2X'X\beta = 0$$

or

$$X'X\beta = X'y \ , \qquad (1.2.5)$$

which gives us n equations, called the normal equations. Since X is of full rank, X'X is nonsingular and hence its inverse exists. In this case the solution of the normal equations can be written as

$$b = (X'X)^{-1}X'y. \qquad (1.2.6)$$

We note that

$$E(b) = E[(X'X)^{-1}X'y]$$

$$= E[(X'X)^{-1}X'(X\beta + \varepsilon)]$$

$$= \beta + (X'X)^{-1}XE(\varepsilon),$$

that is

$E(b) = \beta$, using the assumption in (1.2.3). Hence the least squares estimators are unbiased and have variance given by,

$$Var(b) = E[(b - \beta)(b - \beta)']$$

$$= E[(X'X)^{-1}X'\varepsilon\varepsilon'X(X'X)^{-1}]$$

$$= (X'X)^{-1}X'\sigma^2 I_n X(X'X)^{-1} \ , \quad \text{using (1.2.3)}$$

$$= (X'X)^{-1}\sigma^2 \qquad (1.2.7)$$

1.3 Biased estimation

Although it can be shown that the estimator **b** in (1.2.6) has minimum variance in the class of linear unbiased estimators(Gauss-Markov theorem), there is no guarantee that its variance will be small. If there is an excessive amount of multicollinearity among the independent variables, the $X'X$ matrix approaches a near singular condition, resulting in extremely large values along the diagonal of $(X'X)^{-1}$. This can readily be established by noting that multicollinearity is synonymous with small eigenvalues of the $X'X$ matrix. Now from (1.2.7) the total variance of the estimator can be expressed as,

$$\text{Total variance} = \text{trace}[\text{var}(\mathbf{b})]$$

$$= \sigma^2 \, \text{trace}(X'X)^{-1}$$

$$= \sigma^2 \sum_{i=0}^{i=p} \lambda_i^{-1} \quad , \tag{1.3.1}$$

where $\lambda_0 \geq \lambda_1 \geq \ldots \geq \lambda_p > 0$ are the eigenvalues of $X'X$. It is evident that this value may be too large for practical purposes given a large amount of multicollinearity among the independent variables.

Since the correlation among the independent variables is often a natural phenomena, one cannot always alleviate the difficulty brought about by the multicollinearity simply

by changing the experimental design. In such cases one is forced to drop factors to destroy the correlation bonds among the independent variables. This proves to be unsatisfactory if the initial desire was to use the estimated independent variables for control and optimization. An alternative is to abandon the usual least squares procedure and resort to biased estimation techniques. In using a biased estimation procedure, one is essentially willing to allow for a certain amount of bias in the estimates in order to reduce the variances of the estimators.

1.4 Ridge regression

Hoerl and Kennard [12] have established that on the average, the distance from $\mathbf{b}$ to β will tend to be large if there is a small eigenvalue of $\mathbf{X'X}$. In particular, the worse the conditioning of $\mathbf{X'X}$, the more $\mathbf{b}$ can be expected to be too long and the further one can move from it without an appreciable increase in the residual sum of squares. Thus if $\mathbf{B}$ is an estimate of the vector β, it is only reasonable to fix its squared length at ρ, where ρ is a finite positive constant, and then try to locate the value $\mathbf{b}^*$ of $\mathbf{B}$ that gives a minimum sum of squares for the model (1.2.1). Hence, we need to minimize the residual sum of squares subject to the single constraint that $\mathbf{B'B} = \rho$. The method of Lagrange multipliers requires the differentiation of

$$F = (y - XB)'(y - XB) + k(B'B - \rho) \qquad (1.4.1)$$

with respect to B and equating to zero, where $k \geq 0$ is the Lagrange multiplier.

Now (1.4.1) can be expanded as

$$F = (y - XB)'(y - XB) + k(B'B - \rho)$$

$$= y'y - 2B'X'y + B'X'XB + k(B'B - \rho) \qquad (1.4.2)$$

Then

$$\frac{\partial F}{\partial B} = -2X'y + 2X'XB + 2kB = 0 \quad , \qquad (1.4.3)$$

which can be simplified to,

$$[X'X + kI]B = X'y$$

or

$$b^* = B = (X'X + kI)^{-1}X'y \qquad (1.4.4)$$

where $b^* = [b_0^*, b_1^*, \ldots, b_p^*]$ is the new estimate, called the ridge regression estimate of β.

Now from (1.4.4) we can write,

$$E(b^*) = (X'X + kI)^{-1}X'E(X\beta + \varepsilon)$$

$$= (X'X + kI)^{-1}X'X\beta$$

$$= [I + k(X'X)^{-1}]^{-1}\beta$$

$$= A\beta, \qquad (1.4.5)$$

where $A = [I + k(X'X)^{-1}]^{-1}$. Thus b^* is a biased estimate of β and has variance given by

$$Var(b^*) = Var\{(X'X + kI)^{-1}X'y\}$$

$$= (X'X + kI)^{-1}X'X(X'X + kI)^{-1}\sigma^2 \qquad (1.4.6)$$

From this, we get,

$$Total\ Variance = trace\ Var(b^*)$$

$$= \sigma^2 \sum_{i=0}^{i=p} \frac{\lambda_i}{(\lambda_i + k)^2} \qquad (1.4.7)$$

where $\lambda_0 \geq \lambda_1 \geq \ldots \geq \lambda_p > 0$ are the eigenvalues of $X'X$.

The total mean square error is

$$E[(b^* - \beta)'(b^* - \beta)] = trace[E(b^* - \beta)(b^* - \beta)']$$

$$= trace[E(b^* - A\beta)(b^* - A\beta)' + E(A\beta - \beta)(A\beta - \beta)']$$

$$= \sigma^2 \text{ trace}[(X'X + kI)^{-1}X'X(X'X + kI)^{-1}] +$$

$$\text{trace}(A\beta - \beta)(A\beta - \beta)' \qquad (1.4.8)$$

which in terms of the eigenvalues of $X'X$ can be expressed as

$$E[(b^* - \beta)'(b^* - \beta)] = \sigma^2 \sum_{i=0}^{i-p} \frac{\lambda_i}{(\lambda_i + k)^2} + \sum_{i=0}^{i-p} \frac{k^2\alpha_i^2}{(\lambda_i + k)^2} ,$$

(1.4.9)

where $\alpha = P\beta = (\alpha_0, \alpha_1, \ldots, \alpha_p)'$ and P is an orthogonal matrix satisfying

$$P'(X'X)P = \Lambda = \text{Diag}(\lambda_0, \lambda_1, \ldots, \lambda_p)$$

and $P'P = PP' = I$ (1.4.10)

Having obtained the above equations it can easily be shown (see Kibua[14]) that ridge regression estimators are an improvement on ordinary least squares estimators by comparing the mean square errors in (1.3.1) and (1.4.9).

1.5 Method of the ridge trace

The essential parameter that distinguishes ridge regression from ordinary least squares is k, which may be referred to as the bias parameter. As k increases from zero, the bias of the estimates also increase and the regression estimates all tend toward zero. Hoerl and

Kennard [12] have proved that there is always a positive value of $k < \sigma^2 / \alpha^2_{max}$ for which the ridge estimates will be stable with respect to small changes in the estimation data. In practice, a value of k is chosen by computing b_0^*, b_1^*, ..., b_p^* for a range of values of k between 0 and 1 and plotting the result against k. The resulting graph is known as the ridge trace and is used to select an appropriate value for k. It should be noted, however, that although in most practical cases the value of k falls between 0 and 1, there do exist instances for which the value is actually greater than 1, as has been demonstrated by Brown and Payne [4].

Example 1.

Consider the data in table 1.1 in which measurements were taken on 9 infants. The aim of the experiment was to arrive at a suitable estimating equation relating the length of an infant to all the independent variables, namely age, length at birth, weight at birth and the chest size at birth.

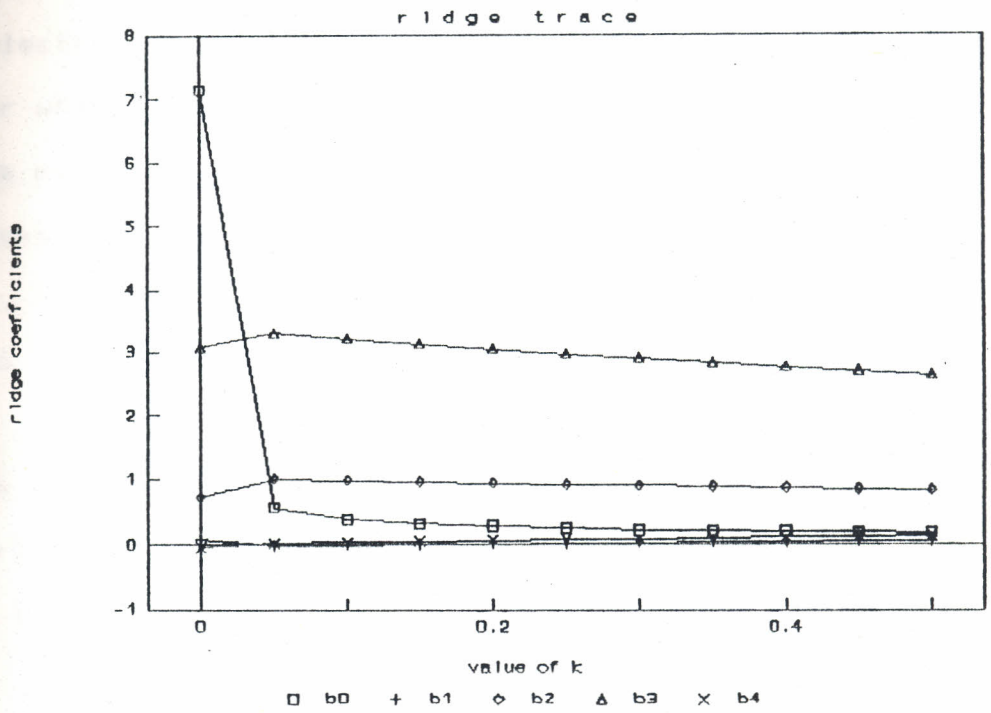
Table 1.1 data relating to infant length.

Infant length, y(cm)	Age, x_1(days)	Length at birth, x_2(cm)	Weight at birth, x_3(Kg)	Chest size at birth, x_4(cm)
57.5	78	48.2	2.75	29.5
52.8	69	45.5	2.15	26.3
61.3	77	46.3	4.41	32.2
67.0	88	49.0	5.52	36.5
53.5	67	43.0	3.21	27.2
62.7	80	48.0	4.32	27.7
56.2	74	48.0	2.31	28.3
68.5	94	53.0	4.30	30.3
69.2	102	58.0	3.71	28.7

Source: Walpole, R.E and Myers, R.H 1978[16]

The computations for $b_0^*, \ldots, b_4^*$ for $0 \leq k \leq 0.5$ were carried out on a computer and plotted against k. The resulting ridge trace is shown in Figure 1.1

FIGURE 1.1



It is evident that some of the coefficients estimated by ordinary least squares method are overestimated and are not collectively stable. In particular, moving a short distance from the least squares point k=0, shows a rapid decrease in the absolute value of b_o.

The selection of k

Since for k>0, **b*** is biased and its bias increases with k while the total variance is a decreasing function of k, the idea of ridge regression is to pick a value of k for

which the reduction in total variance is not exceeded by the increase in bias. The behaviour of **b*** as a function of k is easily observed from the ridge trace. The value of k selected is the smallest value for which **b*** is stable and for which the residual sum of squares should remain close to its minimum value while the variance-covariance matrix of **b*** gives the appearance of an orthogonal system.

1.6 A brief litterature review on rotatable designs

The concept of response surface and designs for the exploration begun in the chemical industry. Much of the early work was done by statisticians and chemical engineers in the Imperial chemical industries in Great Britain.

Box and Hunter [3] suggested the property of rotatability as a desirable quality in an experimental design. This property requires that one should be able to rotate the design through any angle around its centre and yet leave the variances of the response unchanged. Such designs permit a response surface to be fitted easily and provide spherical information contours. Box and Hunter[3] went further to derive the necessary and sufficient conditions for a design of order two to be rotatable. They also constructed designs through geometrical configurations and obtained several rotatable designs of order two. Draper[8] obtained some rotatable designs of order three in three factors and also some specific rotatable designs of

the same order but which were sequential. Kibua [14] applied ordinary ridge regression in the analysis of second order rotatable designs and demonstrated its superiority in the case where there was 'ill-conditioning' in the design matrix. However, the ordinary ridge regression method is subject to further improvement by using the minimum mean square error estimation technique, described in the next chapter.

1.7 Aim and scope of the study

Ridge regression is a relatively new technique and, as such, its full impact as a practical tool has yet to be adequately investigated. However, there are a few researches that have begun to provide the type of insight that a practitioner needs. It was in this respect then, that this work was conducted to explore a practical procedure developed by Hemmerle [11] for obtaining the minimum mean square error estimator for generalized ridge regression methods and to illustrate its usefulness in the analysis of third order rotatable designs.

CHAPTER II

MINIMUM MEAN SQUARE ERROR ESTIMATION

2.1 Introduction

In this chapter we examine the minimum mean square error linear estimator (MIMSEE) in ridge regression in its general form. This is an estimator for which the value k has been chosen so as to minimize the total mean square error. Thus instead of taking a single value for k, we consider several different values, say k_0, k_1, ..., k_p, that is , we consider separate ridge parameters for each of the regression coefficients. Now since $X'X$ is a real p-square symmetric matrix, the conditions stipulated in (1.4.10) are applicable to the model (1.2.2) and the multiple linear regression model can be written in canonical form as

$$y = X^*\alpha + \varepsilon$$

Where

$$X^* = XP' \quad \text{and} \quad \alpha = P\beta \qquad (2.1.1)$$

2.2 The generalized ridge regression estimates and MIMSEE

The generalized ridge regression estimates are generated by a constrained least squares approach in which the error sum of squares is minimized subject to the constraint that $B_i = \rho_i$, i=0, 1, ..., p, where the ρ_i's are finite positive constants. As a result of the constraint applied to the B_i's the relationship

$$\hat{\alpha}^* = Pb^*$$

necessarily restricts the magnitudes of the $\hat{\alpha}^*_i$'s from becoming too large. To minimize the error sum of squares for the canonical model (2.1.1) subject to the above constraints thus requires the use of p Lagrange multipliers, $k_0, k_1, \ldots, k_p$. Taking derivatives of

$$F = (y - XB)(y - XB)' + K(B'B - \rho) \qquad (2.2.1)$$

with respect to the unknown parameters and equating to zero, we thus obtain the system of equations

$$\hat{\alpha}^* = (X^{*\prime}X^* + K)^{-1} X^{*\prime}y \qquad (2.2.2)$$

where

$$K = \mathrm{Diag}(k_0, k_1, \ldots, k_p)$$

with $k_i > 0$ for $i = 0, 1, \ldots, p$

By restricting the magnitudes of the coefficients in the minimization procedure, we have, in effect, added constants to the diagonal elements of $X^{*\prime}X^*$ and consequently introduced some bias into the estimates. However, the addition causes the matrix $X^{*\prime}X^*$ to behave as if the variables are orthogonal to each other, thus increasing stability in the estimates of the coefficients.

values for the k_i's

Optimal values for the k_i's in (2.2.2) are those that minimize the total mean square error. From (1.4.9) it follows that

$$Q = E[(\hat{\alpha}^* - \alpha)'(\hat{\alpha}^* - \alpha)]$$

$$= \sum_{i=0}^{i=p} \frac{\sigma^2\lambda_i + \alpha_i^2 k_i^2}{(\lambda_i + k_i)^2}$$

(2.2.3)

The differentiation of (2.2.3) with respect to the k_i's yields the minimization equations

$$\frac{\partial Q}{\partial k_i} = \frac{2\lambda_i(\lambda_i + k_i)(k_i\alpha_i^2 - \sigma^2)}{(\lambda_i + k_i)^4}$$

$$= 0 , \qquad i = 0, 1, \ldots, p.$$

(2.2.4)

From the full rank assumption of $X'X$ we have that $\lambda_i > 0 \ \forall i$ so that restricting the k_i's to be non-negative yields the solutions

$$k_i = \frac{\sigma^2}{\alpha_i^2} \qquad , i = 0, 1, \ldots, p.$$

(2.2.5)

unfortunately, σ^2 and α_i's are unknown and hence are not operational. Thus we are forced to use their estimates. In practice one estimates σ^2 by s^2, where

$$s^2 = \frac{(y-X^*\hat{\alpha})'(y-X^*\hat{\alpha})}{n-p} \qquad (2.2.6)$$

using the ordinary least squares procedures. As for the estimation of α_i and consequently k_i, i=0, 1, ..., p, suggestion is given on the use of the following iterative procedure:

Step 1. Using ordinary least squares procedures on the canonical model, estimate the α_i's by computing

$$\hat{\alpha} = [X^{*\prime}X^*]^{-1}X^{*\prime}y$$

and estimate σ^2 by s^2.

Step 2. Use the value of s^2 and the α_i's from step 1 to compute

$$k_i = \frac{s^2}{\hat{\alpha}_i^2} \quad , \; i = 0,1,\ldots,p$$

Step 3. Use the k_i's to solve the expression

$$\hat{\alpha}^* = [(X^*)'(X^*) + K]^{-1}(X^*)'y$$

and thus obtain initial estimates of the $\hat{\alpha}^*_i$'s.

Next compute

$$\hat{\alpha}^{*\prime}\hat{\alpha}^* = \sum_{i=0}^{i=p} \hat{\alpha}_i^{*2}$$

Step 4. Repeat steps 2 and 3 using the α^*_i's from step 3

and again compute $\hat{\alpha}^{*\prime}\hat{\alpha}^{*}$.

Step 5. Continue the iterative procedure and terminate only when stability is achieved in $\hat{\alpha}^{*\prime}\hat{\alpha}^{*}$.

Step 6. The generalized ridge regression coefficients are then computed from the formula $\mathbf{b}^{*} = \mathbf{P}'\hat{\alpha}^{*}$, obtained from (2.1.1)

Example 2

Using the data provided in table 1.1, and with the help of the computer program given in the appendix, the results of the above iterative procedure are given below.

The orthogonal matrix P is given by

$$\boldsymbol{P} = \begin{bmatrix} 0.997400 & 0.019620 & -0.046130 & -0.050710 & -0.005130 \\ 0.009957 & 0.818800 & 0.490100 & 0.036790 & 0.296400 \\ -0.028000 & 0.380100 & -0.823800 & 0.318000 & 0.273700 \\ 0.060850 & -0.126100 & 0.245200 & 0.943000 & -0.176100 \\ 0.0228900 & -0.410800 & 0.137200 & 0.075580 & 0.897900 \end{bmatrix} \qquad (2.2.7)$$

and

$$\boldsymbol{X}^{*} = \begin{bmatrix} 0.013388 & 96.34414 & -1.13871 & -0.55801 & 1.289425 \\ 0.008214 & 86.68112 & -3.40199 & -0.08743 & 1.697557 \\ -0.01662 & 95.45551 & 1.31328 & 0.19212 & 3.989337 \\ -0.00372 & 107.1009 & 4.80001 & -0.24344 & 3.785841 \\ 0.025926 & 84.12403 & -1.51928 & 0.39286 & 3.064381 \\ -0.00851 & 97.40796 & -0.20715 & 0.93824 & -1.05717 \\ -0.02739 & 92.59906 & -2.96271 & -0.30625 & 1.794449 \\ 0.023176 & 112.0915 & 1.70051 & -0.07788 & -3.78934 \\ -0.01237 & 120.5965 & -0.00323 & -0.13529 & -7.87097 \end{bmatrix} \qquad (2.2.8)$$

so that

$$\mathbf{X}^{*\prime}\mathbf{X}^{*} = \begin{bmatrix} 0.002721 & 0 & 0 & 0 & 0 \\ 0 & 89630 & 0 & 0 & 0 \\ 0 & 0 & 51.66 & 0 & 0 \\ 0 & 0 & 0 & 1.568 & 0 \\ 0 & 0 & 0 & 0 & 124.8 \end{bmatrix}$$

(2.2.9)

The ordinary least squares estimate of the coefficients are then

$\mathbf{b}=(7.2177, -0.27722, 1.557625, 4.544891, -0.06157)'$

with the mean square error calculated to be 752.922

(2.2.10)

On the other hand the ridge method gives

$\mathbf{b}^{*}=(0.26998, -0.08811, 1.13259, 3.84959, -0.04645)'$

as the estimate of the regression coefficients with a mean square error of 11.84962. (2.2.11)

Thus using the mean square error as our basis for comparison, then as expected, the ridge method is preferred to the ordinary least squares procedure, due to its smaller value.

2.3 A non-iterative solution for generalized ridge regression

We assume the model has already been reduced to the canonical form expressed in (2.1.1). The general ridge estimation procedure is then defined as

$$\hat{\alpha}^* = [(X^*)'(X^*) + K]^{-1}(X^*)'y \ , \qquad (2.3.1)$$

where K is a diagonal matrix with non-negative diagonal elements $k_0, k_1, \ldots, k_p$.

The iterative procedure to estimate k_i is described by the formula

$$k_{i(j)} = \frac{s^2}{(\hat{\alpha}^*_{i(j)})^2} \ , \qquad i=0,1,\ldots,p \qquad (2.3.2)$$

where the bracketed j subscript is used to denote the j^{th} iterate, and take $\hat{\alpha}^*_{i(0)} = \hat{\alpha}_i$, i=0, 1, ..., p as initial values, where $\hat{\alpha}_i$ is the ordinary least squares estimate of α_i.

We first give an explicit formulation of the above procedure, in terms of matrices. To this end we represent the (p+1) vectors of $(X^*)'y$ and $\hat{\alpha}^*_{(j)}$ as diagonal matrices. Let

$$B = \begin{bmatrix} [(X^*)'y]_0 & 0 & \cdots & 0 \\ 0 & [(X^*)'y]_1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & [(X^*)'y]_p \end{bmatrix}$$

and

$$A_j = \begin{bmatrix} \hat{\alpha}^*_{0(j)} & 0 & \cdots & 0 \\ 0 & \hat{\alpha}^*_{1(j)} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \hat{\alpha}^*_{p(j)} \end{bmatrix} \tag{2.3.3}$$

As a consequence we have that

$$A_0 = \Lambda^{-1} B \tag{2.3.4}$$

where Λ is the diagonal matrix of characteristic roots of $X'X$. The iterative procedure may now be described by the matrix formula

$$A_{j+1} = (\Lambda + s^2 A^{-2}_j)^{-1} B \tag{2.3.5}$$

or

$$A_{j+1} = (\Lambda + s^2 A^{-2}_j)^{-1} \Lambda A_0 \tag{2.3.6}$$

which can be reduced to

$$A_{j+1} = (I + s^2 \Lambda^{-1} A^{-2}_j)^{-1} A_0 \tag{2.3.7}$$

Next, we let

$$D = s^{-2}\,\Lambda \qquad \text{in (2.3.7) and obtain}$$

$$A_{j+1} = (I + D^{-1}A^{-2j})^{-1}\,A_0 \tag{2.3.8}$$

An expression for A^{-2j+1} is then given by

$$A^{-2j+1} = A^{-1}{}_0(I + D^{-1}A^{-2j})A^{-1}{}_0(I + D^{-1}A^{-2j}) \tag{2.3.9}$$

However, the matrices in (2.3.9) are diagonal and commute. Thus we can write

$$A^{-2j+1} = A^{-2}{}_0(I + D^{-1}A^{-2j})^2 \tag{2.3.10}$$

so that

$$D^{-1}A^{-2j+1} = D^{-1}A^{-2}{}_0(I+D^{-1}A^{-2j})^2 \tag{2.3.11}$$

and if we let

$$E_j = D^{-1}A^{-2j} \tag{2.3.12}$$

the iterative procedure is reduced to the simple formula

$$E_{j+1} = E_0(I+E_j)^2 \tag{2.3.13}$$

Now assuming that $a_i \neq 0 \;\forall$ i and that the iterative procedure is convergent, that is

$$\lim_{j\to\infty} E_j = E^* \qquad (2.3.14)$$

then from (2.3.13) and (2.3.14), we must have the relationship

$$E^* = E_0(I + E^*)^2 \qquad (2.3.15)$$

or, if expanded,

$$(E^*)^2 + (2I - E^{-1}_0)E^* + I = 0 \qquad (2.3.16)$$

Now (2.3.16) consists of (p+1) equations of the form

$$(e^*)^2 + (2 - 1/e_0)e^* + 1 = 0 \qquad (2.3.17)$$

where e_0 and e^* are scalars. Solving for e^* we obtain

$$e^* = \frac{(1-2e_0) \pm \sqrt{(1-4e_0)}}{2e_0} \qquad (2.3.18)$$

Convergence conditions

The matrix formula (2.3.13) also consists of (p+1) separate iterative expressions of the form

$$e_{j+1} = e_0(1 + e_j)^2, \qquad (2.3.19)$$

where e_0, e_j, e_{j+1} , are scalars and the subscript j is used to denote the j^{th} iterate. It is of much interest to know for what range of values for e_0 the expression (2.3.19) converges or diverges. But first we need the following lemma.

Lemma

The iterative procedure defined by (2.3.19) converges for $e_0 = 1/4$.

proof

We first observe that for $e_0>0$ we have

$$e_1 = e_0(1 + e_0)^2 > e_0$$

$$e_2 = e_0(1 + e_1)^2 > e_0(1 + e_0)^2 = e_1$$

and proceeding inductively, we obtain

$$e_{j+1} > e_j \text{ , for } e_0 \qquad (2.3.20)$$

Now for $e_0 = 1/4$ let

$$\sqrt{e_1} = \sqrt{e_0}(1 + e_0) = g_1 = 1 - 3/2^3$$

$$\sqrt{e_2} = \sqrt{e_0}(1 + e_1) \leq \sqrt{e_0}(1 + \sqrt{e_1})$$

$$= \sqrt{e_0}(1 + g_1) = g_2 = 1 - 3/2^4$$

In order to argue inductively, assume that

$$\sqrt{e_j} \leq g_j = 1 - 3/2^{j+2}$$

Then

$$\sqrt{e_{j+1}} = \sqrt{e_0}(1 + e_j)$$

$$\leq \sqrt{e_0}(1 + \sqrt{e_j}) \leq \sqrt{e_0}(1 + g_j)$$

and

$$\sqrt{e_0}(1 + g_j) = 1 - 3/2^{j+3} = g_{j+1}$$

Consequently for all j we have

$$\sqrt{e_j} < g_j < 1 , \quad \text{for } e_0 = 1/4 \qquad (2.3.21)$$

Combining this result with (2.3.20) yields

$$0<e_0<e_1<\ldots<e_j<\ldots<1 , \quad \text{for } e_0 = 1/4 \qquad (2.3.22)$$

which is a monotonically increasing sequence of real numbers bounded from above, so the iterative procedure converges. ▪

Observe from (2.3.18) that we must have

$$e^* = \lim_{j\to\infty} e_j = 1$$

for $e_0 = 1/4$. (2.3.23)

Now suppose that the iterative procedure defined by (2.3.19) converges for e_0 and that we have $0 < e'_0 \leq e_0$. Clearly, for the primed sequence $e'_0, e'_1, \ldots, e'_j, \ldots$, we must have $e'_j \leq e_j$ so that the primed sequence converges to a limit $(e^*)' \leq e^*$. This fact suggests that the iterative procedure defined by (2.3.19) converges whenever $0 < e_0 \leq \frac{1}{4}$.

Now consider equation (2.3.18), and take $0 < e_0 < \frac{1}{4}$. Taking the positive sign and squaring the right hand side gives

$$e^{*2} = \left[\frac{(1 - 2e_0) + \sqrt{(1 - 4e_0)}}{2e_0}\right]^2$$

$$= \frac{(1 - 2e_0)^2 + 2(1 - 2e_0)\sqrt{(1 - 4e_0)} + (1 - 4e_0)}{4e_0^2}$$

$$> \frac{(1 - 2e_0)^2 + 2(1 - 4e_0) + (1 - 4e_0)}{4e_0^2}$$

$$= 1 + \frac{(1 - 4e_0)}{e_0^2}$$

$$> 1,$$

which contradicts the fact that we must have $e^* \leq 1$. Consequently, an explicit solution for (2.3.19), valid whenever $0 < e_0 \leq \frac{1}{4}$ is given by

$$e^* = \frac{(1 - 2e_0) - \sqrt{(1 - 4e_0)}}{2e_0} \quad , \qquad 0 < e_0 \leq \frac{1}{4} \; .$$

(2.3.24)

For $e_0 > 1/4$, $(1 - 4e_0)$ is negative and the iterative procedure diverges and thus

$$\lim_{j \to \infty} e_j = \infty \; , \; for \; e_0 > \frac{1}{4}$$

(2.3.25)

Explicit solution for optimal $\hat{a}^*$

Let

$$e_i^* = \lim_{j \to \infty} e_{i(j)}$$

$$\hat{a}_i^* = \lim_{j \to \infty} \hat{a}_{i(j)}^*$$

(2.3.26)

Then since

$$e_{i(j)} = \frac{s^2}{\lambda_i (\hat{a}_{i(j)}^*)^2} \quad ,$$

(2.3.27)

we have that

$$\lim_{j \to \infty} \hat{a}_{i(j)}^* = 0 \quad for \quad e_{i(0)} > \frac{1}{4}$$

(2.3.28)

whenever the procedure defined by (2.3.19) diverges for the

i^{th} equation. Thus we let

$$\hat{\alpha}^*_i = 0 \quad , \quad \text{for} \quad e_{i(o)} > \tfrac{1}{4} \tag{2.3.29}$$

When the procedure converges for the i^{th} equation we have that

$$\hat{\alpha}^*_i = \frac{[(X^*)'y]_i}{\lambda_i + \lambda_i e^*_i} = \frac{\hat{\alpha}_i}{(1 + e^*_i)} \quad , \tag{2.3.30}$$

for $0 < e_{i(o)} \leq \tfrac{1}{4}$, where e^*_i is evaluated using the formula (2.3.24).

By applying (2.3.29) and (2.3.30), we obtain an explicit solution for all of the optimum generalized ridge regression estimators $\hat{\alpha}^*_i$, i=0, 1, ..., p.

Example 3

Still making use of the data in table 1.1 together with some of the results obtained in example 2 of section 2.1, we can calculate the following;

$s^2=2.045$

$\hat{\alpha}=(7.50697,\ 0.613404,\ 0.241795,\ 4.274334,\ 0.450613)'$ (2.3.31)

so that from (2.3.12),

$$e_{1(0)} = \frac{s^2}{\lambda_1(\hat{R}_{1(0)})^2}$$

(2.3.32)

and hence

$e_{0(0)}$=13.336 . This is greater than 0.25 and so we assign the value zero to $\hat{a}^*_0$.

Also

$e_{1(0)}$=0.00006 . This is less than 0.25 so that

e^*_1=0.000058, giving $\hat{a}^*_1$=0.6134

Similarly, for the remaining ones we get,

$e_{2(0)}$=0.677, $\hat{a}^*_2$=0

$e_{3(0)}$=0.071, $\hat{a}^*_3$=3.9436 and

$e_{4(0)}$=0.081, $\hat{a}^*_4$=0.4107 (2.3.33)

Hence

$$\hat{a}^* = (0,\ 0.6134,\ 0,\ 3.9436,\ 0.4107)'$$

and the resulting solution for **b*** is given by

$$\mathbf{b}^* = (0.2555, -0.1638, 1.3239, 3.772, -0.1439)'$$

from which the mean square error is equal to 5.6945 which, as expected, is less than that of the previous example.

CHAPTER III

ROTATABLE DESIGNS OF UPTO THIRD ORDER

3.1 Response surfaces

Suppose we have a system which involves a response variable, y, which depends on the level of some random variables $\zeta_1, \zeta_2, \ldots, \zeta_p$. We assume that the levels of the ζ_i's can be controlled by the experimenter with negligible error. The experiment is conducted with design variables $x_1, x_2, \ldots, x_p$, which are usually simple transformations of ζ. Each treatment can then be represented by a point with coordinates $(x_{1i}, x_{2i}, \ldots, x_{pi})$ in a p-dimensional factor space at the point y_i , i=1, 2, ...,n. Since the freedom to choose the levels of the ζ's is ours, it is normally convenient to have them equally spaced. The relationship between ζ and x can then be expressed as

$$x_{ji} = \frac{\zeta_{ji} - \bar{\zeta}_j}{\zeta_{ji} - \zeta_{j(i-1)}} ,$$

where $$\bar{\zeta}_j = \sum_{i=1}^{n} \frac{\zeta_{ji}}{n} .$$

(3.1.1)

In general, the response is a function of the random variables $y=f(\zeta_1, \zeta_2, \ldots, \zeta_p)$ and is often unknown. The

basic response procedure is thus to approximate f(.) with a low order polynomial and to use sample data to fit least squares estimates of the coefficients of the polynomial.

Occasionally, in factorial experiments, the relationship between the response and factor levels is studied. In order to get the relationship between the response and the factor levels, a set of treatment combinations are suitably chosen. Such a set of treatment combinations is called a response surface design, and the prior mentioned relationship, a response surface. These surfaces can be linear, second degree or higher degree polynomials. The main considerations connected with the exploration of the response surface are;

i) performing statistically designed experiments
ii) estimating the coefficients in the response surface equation
iii) checking on the adequacy of the equation and
iv) studying the response surface in the region of interest.

In this work, particular consideration is given to the estimation of the coefficients in the response surface equation of order three using ridge regression techniques.

3.2 Rotatable designs

Let p factors $F_1, F_2, \ldots, F_p$ affect the yield in a particular character and let the expected yield satisfy the functinal relation $y=f(x_1, x_2, \ldots, x_p)$ where $x_1, x_2, \ldots, x_p$ are the levels of the factors $F_1, F_2, \ldots, F_p$, respectively, used for getting that response. We assume that f(.) can be represented by a polynomial of degree `d'.

Definition

A p-dimensional design of order d is said to be a rotatable design if the variance of the estimated response at the point $(x_1, x_2, \ldots, x_p)$ is a function of a constant, ρ, given by

$$\rho^2 = \sum_{i=1}^{p} x_i^2 \quad .$$

(3.2.1)

This property is a reasonable one to adopt for exploratory work, in which the experimenter does not know in advance how the response surface will orient itself with respect to the x-axis. Consequently, he has no rational basis for specifying that the variance of the estimated response should be smaller in some directions than in others.

3.3 First order designs

The simplest kind of surface which could possibly be used to describe the response y, as a function of the p factors is a first order surface. Its equation is of the form

$$y_u = \beta_0 x_{0u} + \Sigma \beta_i x_{iu} + \varepsilon_u , \quad u=1, 2, \ldots, N \qquad (3.3.1)$$

where β's are unknown constants and $x_{0u} = 1$, $\forall u$

$E(\varepsilon_u)=0$, $Var(\varepsilon_u)=\sigma^2$ (unknown)

and $Cov(\varepsilon_u \varepsilon_{u'})=0$, $u \neq u'$.

The summation is taken over i=1, 2, ..., p

The experimental design to estimate β's can then be written as $E(\mathbf{y})=\mathbf{X}\boldsymbol{\beta}$ where $\mathbf{X}$ is of the form (1.2.2)

We can choose $\mathbf{X}$ such that the estimates of $\boldsymbol{\beta}$ are orthogonal and the experimental design is rotatable, that is, $Var(y_u)$ is fixed for all $(x_{1u}, x_{2u}, \ldots, x_{pu})$, which satisfy

$$\sum_{i=1}^{p} x_{iu}^2 = \rho^2 (\textit{some constant}), u=1,2,\ldots,n. \qquad (3.3.2)$$

3.4 Second order designs

A second order response surface is represented by the polynomial equation

$$y = \beta_0 x_0 + \sum_{i=1}^{p} \beta_i x_i + \sum_{\substack{i=1 \\ i \le j}}^{p} \sum_{j=1}^{p} \beta_{ij} x_i x_j \quad , \tag{3.4.1}$$

where $\beta_0, \beta_1, \ldots, \beta_p, \beta_{11}, \ldots, \beta_{pp}, \beta_{12}, \ldots, \beta_{(p-1)p}$ are constants.

The design used to estimate the unknown constants is known as second order response surface.

Box and Hunter [3] showed that the moment conditions necessary for the n points $(x_{1u}, x_{2u}, \ldots, x_{pu})$ u=1, 2, ..., n to form a second order rotatable design are

$$\sum_{u=1}^{n} x_{iu} = 0 \quad , \qquad \sum_{u=1}^{n} x_{iu} x_{ju} = 0 \quad , \qquad \sum_{u=1}^{n} x_{iu} x_{ju}^2 = 0$$

$$\sum_{u=1}^{n} x_{iu}^3 = 0 \quad , \qquad \sum_{u=1}^{n} x_{iu} x_{ju}^3 = 0 \quad , \qquad \sum_{u=1}^{n} x_{iu} x_{ju} x_{ku} = 0$$

$$\sum_{u=1}^{n} x_{iu} x_{ju} x_{ku}^2 = 0 \quad \text{and} \quad \sum_{u=1}^{n} x_{iu} x_{ju} x_{ku} x_{lu} = 0 \quad , \text{ for } i \neq j \neq k \neq l \, , \tag{3.4.2}$$

that is, each of the sum of powers or sum of the product of powers of the x_{iu}'s with at least one power odd is zero. Further,

$$\sum_{u=1}^{n} x_{iu}^2 = n\mu_2 \quad , \quad \sum_{u=1}^{n} x_{iu}^4 = 3n\mu_4$$

$$\sum_{u=1}^{n} x_{iu}^2 x_{ju}^2 = n\mu_4 \qquad for \quad i \neq j \ ,$$

(3.4.3)

where μ's are constants.

They also showed that in order that $\mathbf{X'X}$ be non-singular, μ_2 and μ_4 must satisfy

$$\frac{\mu_4}{\mu_2^2} > \frac{p}{p+2} \ .$$

(3.4.4)

Also, if the design matrix $\mathbf{X}$ of n runs does not satisfy (3.4.4) then an addition of one or more central points (0, 0, ...,0) could satisfy the condition (3.4.4) and the final design matrix will be a rotatable design of order two.

3.5 Third order designs

If y_u is the response at the u^{th} experimental point then the polynomial equation of third order may be expressed as

$$y_u = \beta_0 x_{0u} + \sum_{i=1}^{p} \beta_i x_{iu} + \sum_{i=1}^{p} \sum_{j \geq i}^{p} \beta_{ij} x_{iu} x_{ju}$$

$$+ \sum_{i=1}^{p} \sum_{j \geq i}^{p} \sum_{k \geq j}^{p} \beta_{ijk} x_{iu} x_{ju} x_{ku} + \epsilon_u \quad . \qquad (3.5.1)$$

Or, in vector notation, as

$$y_u = X_u \beta + \varepsilon_u \qquad \forall\, u,$$

where

$X_u = (x_{0u}, x_{1u}, \ldots, x_{pu}, x^2_{1u}, \ldots, x^2_{pu}, x_{1u}x_{2u}, \ldots,$
$x_{(p-1)u}x_{pu}, x_{1u}x_{2u}x_{3u}, \ldots, x_{(p-2)u}x_{(p-1)u}x_{pu})$, and
$x_{0u} = 1 \quad \forall\, u.$

also,

$\beta = (\beta_0, \beta_1, \ldots, \beta_p, \beta_{11}, \ldots, \beta_{pp}, \beta_{12}, \ldots, \beta_{(p-1)p},$
$\beta_{111}, \ldots, \beta_{ppp}, \ldots, \beta_{(p-2)(p-1)p}),$

$E(\varepsilon_u) = 0$, $Var(\varepsilon_u) = \sigma^2$

and $Cov(\varepsilon_u \varepsilon_{u'}) = 0$ for $u \neq u'$, $u = 1, 2, \ldots, n$.

For third order rotatable designs, Box and Hunter[3] showed that the moment conditions for a set of points to form a rotatable arrangement are:

a) the sum of powers or the sum of the product of powers of the x_{iu}'s with at least one power odd is zero,

b) the sum of powers or the sum of the product of powers of the x_{iu}'s with even powers is constant, that is,

$$\sum_{u=1}^{n} x_{iu}^2 = n\mu_2 \quad , \qquad \sum_{u=1}^{n} x_{iu}^4 = 3n\mu_4 \quad ,$$

$$\sum_{u=1}^{n} x_{iu}^2 x_{ju}^2 = n\mu_4 \quad , \qquad \sum_{u=1}^{n} x_{iu}^6 = 15n\mu_6 \quad ,$$

$$\sum_{u=1}^{n} x_{iu}^2 x_{ju}^4 = 3n\mu_6 \quad , \qquad \text{for } i \neq j \text{ and,}$$

$$\sum_{u=1}^{n} x_{iu}^2 x_{ju}^2 x_{ku}^2 = n\mu_6 \quad , \qquad \text{for } i \neq j \neq k \ .$$

(3.5.2)

Further the design matrix must satisfy the additional conditions

$$\frac{\mu_4}{\mu_2^2} > \frac{p}{p+2} \quad ,$$

(3.5.3)

and

$$\frac{\mu_2\mu_6}{\mu_4^2} > \frac{p+2}{p+4} \quad , \tag{3.5.4}$$

in order that $X'X$ be non-singular.

However, if the design matrix X does not satisfy the conditions of non-singularity due to equality in (3.5.3), then the addition of no central points (0, 0, ...,0) could satisfy (3.4.7). The corresponding values of μ_2, μ_4, μ_6 obtained by the addition of these points would then be,

$$\mu_2^* = \frac{n\mu_2}{n+n_0} \quad , \quad \mu_4^* = \frac{n\mu_4}{n+n_0} \quad \text{and} \quad \mu_6^* = n\mu_6^* = \frac{n\mu_6}{n+n_0} \ .$$

$$\text{Thus} \quad \frac{\mu_4^*}{\mu_2^{*2}} = \left(1+\frac{n_0}{n}\right)\frac{\mu_4}{\mu_2^2} > \frac{p}{p+2} \quad , \tag{3.5.5}$$

which implies that (3.5.3) is satisfied. However in case of equality in (3.5.4) the addition of no central points implies that

$$\frac{\mu_2^*\mu_6^*}{\mu_4^*} = \frac{\mu_2\mu_6}{\mu_4^2} = \frac{p+2}{p+4} \ . \tag{3.5.6}$$

Thus (3.5.4) is not satisfied. Hence the mere addition of

central points will not enable the design to satisfy condition (3.5.4). Draper [7] proved that in order to obtain third order rotatable designs in this case, one must combine at least two spherical sets of points with different positive radii.

3.6 The moment matrix X'X of third order rotatable designs

It is noted that (3.5.1) has its right hand side containing all possible terms up to and including third order. However, the terms are grouped in a particular form so that the sub-matrices in **X'X** are easy to handle, as they appear in the diagonal of the moment matrix. The model in expanded form is

$$
\begin{aligned}
E(y) = {} & \beta_0 + \beta_{11}x_1^2 + \beta_{22}x^2_2 + \ldots + \beta_{pp}x^2_p \\
& + \beta_{12}x_1x_2 + \beta_{13}x_1x_3 + \ldots + \beta_{(p-1)p}x_{p-1}x_p \\
& + \beta_1x_1 + \beta_{111}x^3_1 + \ldots + \beta_{1pp}x_1x^2_p \\
& + \beta_2x_2 + \beta_{222}x^3_2 + \ldots + \beta_{2pp}x_2x^2_p \\
& \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
& + \beta_{123}x_1x_2x_3 + \beta_{124}x_1x_2x_4 + \ldots \\
& + \beta_{(p-2)(p-1)p}x_{p-2}x_{p-1}x_p \qquad (3.6.1)
\end{aligned}
$$

If we write

$$\mathbf{y} = (y_1, y_2, \ldots, y_n)'$$

$$\boldsymbol{\beta} = (\beta_0, \beta_{11}, \beta_{22}, \ldots, \beta_{12}, \beta_{13}, \ldots, \beta_{(p-2)(p-1)p})'$$

$$\mathbf{x}_u = (1, x^2_{1u}, x^2_{2u}, \ldots, x_{1u}x_{2u}, \ldots, x_{(p-2)u}x_{(p-1)u}x_{pu})' \qquad u = 1, 2, \ldots, n$$

and define $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)'$, then the model can be written as

$$E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta} \qquad (3.6.2)$$

Here $\mathbf{y}$ is an $n \times 1$ vector, $\mathbf{X}$ is an $n \times (p+3)(p+2)(p+1)/6$ matrix and $\boldsymbol{\beta}$ is a $(p+3)(p+2)(p+1)/6 \times 1$ vector. Hence the moment matrix $\mathbf{X}'\mathbf{X}$ is a square matrix of dimension $(p+3)(p+2)(p+1)/6$.

When a third order rotatable design is used to estimate the coefficient of the cubic model (3.6.1), the moment matrix $\mathbf{X}'\mathbf{X}$ takes a particularly simple form due to the fact that the moments of the design obey the conditions given in (3.5.2).

Thus we have

$$X'X = n\begin{bmatrix} G & O & O & O & O & \cdots & O & O \\ & \mu_4 I_1 & O & O & O & \cdots & O & O \\ & & K_1 & O & O & \cdots & O & O \\ & & & K_2 & O & \cdots & O & O \\ & (symm.) & & & K_3 & \cdots & O & O \\ & & & & & & \vdots & \vdots \\ & & & & & & K_p & O \\ & & & & & & & \mu_6 I_2 \end{bmatrix}$$

(3.6.3)

Where **O** denotes a matrix consisting entirely of zeros and of dimensions appropriate to its position in **X'X** and where the other sub-matrices are defined as follows:

$$G = \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix}$$

(3.6.4)

is of dimension (p+1) × (p+1), where further

$$G_{11} = 1 \quad , \quad G_{12} = G'_{21} = \mu_2 1_{1\times p}$$

and

$$G_{22} = \mu_4 \begin{bmatrix} 3 & 1 & 1 & \cdots & 1 \\ 1 & 3 & 1 & \cdots & 1 \\ 1 & 1 & 3 & \cdots & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 1 & \cdots & 3 \end{bmatrix}_{p\times p}$$

(3.6.5)

Also

$$K_1 = \begin{bmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{bmatrix} \tag{3.6.6}$$

where

$$K_{11} = \begin{bmatrix} \mu_2 & 3\mu_4 \\ 3\mu_4 & 15\mu_6 \end{bmatrix}_{2\times 2} \tag{3.6.7}$$

$$K_{12} = K'_{21} = \begin{bmatrix} \mu_4 & \mu_4 & \cdots & \mu_4 \\ 3\mu_6 & 3\mu_6 & \cdots & 3\mu_6 \end{bmatrix}_{2\times(p-1)} \tag{3.6.8}$$

and

$$K_{22} = \mu_6 \begin{bmatrix} 3 & 1 & 1 & \cdots & 1 \\ 1 & 3 & 1 & \cdots & 1 \\ 1 & 1 & 3 & \cdots & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 1 & \cdots & 3 \end{bmatrix}_{(p-1)\times(p-1)} \tag{3.6.9}$$

K_2, K_3, ..., K_p are similarly defined and are all of dimension $(p+1) \times (p+1)$.

I_1 and I_2 are identity matrices of dimensions

$\frac{1}{2}\, p(p-1) \times \frac{1}{2}\, p(p-1)$ and $p(p-1)(p-2)/6 \times p(p-1)(p-2)/6$ respectively. (3.6.10)

The inversion of the moment matrix $X'X$ now reduces to the inversion of the submatrices. The inverses of the

submatrices are worked out below.

$$G^{-1} = \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix} \tag{3.6.11}$$

where

$$H_{11} = (G_{11} - G_{12}\, G^{-1}{}_{22}\, G_{21})^{-1}\ , \tag{3.6.12}$$

$$H_{12} = H'_{21} = -\,(G_{11} - G_{12}\, G^{-1}{}_{22}\, G_{21})^{-1}\, G_{12}\, G^{-1}{}_{22} \tag{3.6.13}$$

and

$$H_{22} = G^{-1}{}_{22} + G^{-1}{}_{22}\, G_{21}\, (G_{11} - G_{12}\, G^{-1}{}_{22}\, G_{21})^{-1}\, G_{12}\, G^{-1}{}_{22} \tag{3.6.14}$$

Now

$$G_{22}^{-1} = \frac{1}{2\mu_4 (p+2)} \begin{bmatrix} p+1 & -1 & -1 & \cdots & -1 \\ -1 & p+1 & -1 & \cdots & -1 \\ -1 & -1 & p+1 & \cdots & -1 \\ \vdots & \vdots & \vdots & \vdots\ \vdots\ \vdots & \vdots \\ -1 & -1 & -1 & \cdots & p+1 \end{bmatrix}_{p \times p} \tag{3.6.15}$$

$$G_{12} G_{22}^{-1} = \frac{\mu_2}{\mu_4 (p+2)} 1'_{1 \times p}\ , \tag{3.6.16}$$

and

$$G_{12}\, G_{22}^{-1}\, G_{21} = \frac{\mu_2^2 p}{\mu_4 (p+2)} \quad ,$$

(3.6.17)

so that

$$H_{11} = \frac{\mu_4 (p+2)}{\mu_4 (p+2) - \mu_2^2 p} \quad ,$$

$$H_{12} = H'_{21} = \frac{\mu_2}{\mu_4 (p+2) - \mu_2^2 p} 1'_{1\times p} \quad ,$$

and

$$H_{22} = \begin{bmatrix} c & d & d & \cdots & d \\ d & c & d & \cdots & d \\ d & d & c & \cdots & d \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ d & d & d & \cdots & c \end{bmatrix}_{p\times p}$$

where

$$c = \frac{\mu_4 (p+1)\,(p+2) - \mu_2^2\,[p(p+1) - 2]}{2\mu_4 (p+2)\,[\mu_4 (p+2) - \mu_2^2 p]} \quad ,$$

and

$$d = \frac{\mu_2^2 - \mu_4}{2\mu_4\,[\mu_4 (p+2) - \mu_2^2 p]} \quad .$$

(3.6.18)

Similarly,

$$K^{-1} = \begin{bmatrix} L_{11} & L_{12} \\ L_{21} & L_{22} \end{bmatrix}$$

(3.6.19)

where the entries of K^{-1} are defined similar to expressions (3.6.12), (3.6.13), (3.6.14). Then

$$K_{22}^{-1} = \frac{1}{2\mu_6(p+1)} \begin{bmatrix} p & -1 & -1 & \cdots & -1 \\ -1 & p & -1 & \cdots & -1 \\ -1 & -1 & p & \cdots & -1 \\ \vdots & \vdots & \vdots & \vdots\ \vdots\ \vdots & \vdots \\ -1 & -1 & -1 & \cdots & p \end{bmatrix}_{(p-1)\times(p-1)}$$

(3.6.20)

$$K_{12}\, K_{22}^{-1} = \frac{1}{\mu_6(p+1)} \begin{bmatrix} \mu_4 & \mu_4 & \cdots & \mu_4 \\ 3\mu_6 & 3\mu_6 & \cdots & 3\mu_6 \end{bmatrix}_{2\times(p-1)}$$

(3.6.21)

and

$$K_{12}\, K_{22}^{-1}\, K_{21} = \frac{(p-1)}{\mu_6(p+1)} \begin{bmatrix} \mu_4^2 & 3\mu_4\mu_6 \\ 3\mu_4\mu_6 & 9\mu_6^2 \end{bmatrix}$$

(3.6.22)

so that

$$L_{11} = \begin{bmatrix} e & f \\ f & g \end{bmatrix}$$

(3.6.23)

Where

$$e = \frac{\mu_6(p+4)}{\mu_2\mu_6(p+4) - \mu_4^2(p+2)} \quad ,$$

$$f = \frac{-\mu_4}{\mu_2\mu_6(p+4) - \mu_4^2(p+2)} \quad ,$$

$$g = \frac{\mu_2\mu_6(p+1) - \mu_4^2(p-1)}{6\mu_6[\mu_2\mu_6(p+4) - \mu_4^2(p+2)]} \quad .$$

(3.6.24)

Also

$$L_{12} = L'_{21} = \begin{bmatrix} f & f & \cdots & f \\ h & h & \cdots & h \end{bmatrix}_{2\times(p-1)}$$

(3.6.25)

where f is as defined in (3.6.24) and

$$h = \frac{\mu_4^2 - \mu_2\mu_6}{2\mu_6[\mu_2\mu_6(p+4) - \mu_4^2(p+2)]} \quad .$$

(3.6.26)

Then

$$L_{22} = \begin{bmatrix} 1 & h & h & \cdots & h \\ h & 1 & h & \cdots & h \\ h & h & 1 & \cdots & h \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ h & h & h & \cdots & 1 \end{bmatrix}_{(p-1)\times(p-1)}$$

(3.6.27)

where **h** is as in (3.6.26) and

$$l = \frac{\mu_2\mu_6(p+3) - \mu_4^2(p+1)}{2\mu_6[\mu_2\mu_6(p+4) - \mu_4^2(p+2)]} \quad .$$

(3.6.28)

Finally,

$$(\mu_4 I_1)^{-1} = \frac{1}{\mu_4} I_1 \quad ,$$

and

$$(\mu_6 I_2)^{-1} = \frac{1}{\mu_6} I_2 \quad ,$$

(3.6.29)

thus

$$(X'X)^{-1} = \frac{1}{n}\begin{bmatrix} G^{-1} & O & O & \cdots & O & O \\ & \frac{1}{\mu_4} I_1 & O & \cdots & O & O \\ & & K^{-1} & \cdots & O & O \\ & & & & \vdots & \vdots \\ (symm.) & & & & K^{-1} & O \\ & & & & & \frac{1}{\mu_6} I_2 \end{bmatrix}$$

(3.6.30)

The least squares estimates of the coefficients may then be obtained from (1.2.6), and the corresponding estimates of the coefficients using ridge regression techniques, from (2.1.1) and (2.2.1). However, with the latter, we need not find an expressions similar to (3.6.30) for $(X'X + K)^{-1}$, since the non-iterative solution given in section 2.3 gives the value of $\hat{\alpha}^*$ in (2.3.1) explicitly without first obtaining the matrix K. This is illustrated in the next chapter.

CHAPTER IV

APPLICATION OF RIDGE REGRESSION ON THIRD ORDER ROTATABLE DESIGNS IN THREE FACTORS

4.1 Introduction

Suppose that an experimenter is interested in three experimental factors and that these factors have been coded so that an experimental design for examining them can be described by referring to a set of points (x_1, x_2, x_3) in three dimensional cartesian space, rather than to the actual factor levels. Suppose, further, that a dependent variable y is observed as resulting from the experiment and that it is desired to estimate by least squares, the coefficients of a selected polynomial function of x_1, x_2 and x_3 which, it is thought, will give a satisfactory representation of y. For the estimation of a cubic function of the form given in (3.6.1), the model under consideration becomes,

$$\begin{aligned} E(y) = {} & \beta_0 + \beta_{11}x_1^2 + \beta_{22}x^2_2 + \beta_{33}x^2_3 \\ & + \beta_{12}x_1x_2 + \beta_{13}x_1x_3 + \beta_{23}x_2x_3 \\ & + \beta_1x_1 + \beta_{111}x^3_1 + \beta_{122}x_1x^2_2 + \beta_{133}x_1x^2_3 \\ & + \beta_2x_2 + \beta_{222}x^3_2 + \beta_{112}x^2_1x_2 + \beta_{233}x_2x^2_3 \\ & + \beta_3x_3 + \beta_{333}x^3_3 + \beta_{113}x^2_1x_3 + \beta_{223}x^2_2x_3 \\ & + \beta_{123}x_1x_2x_3. \end{aligned} \qquad (4.1.1)$$

Certain third order rotatable designs have been proposed by Draper[7], suitable for estimating a third order or cubic polynomial. These are such that the designs consist of two parts, each of which is a second order rotatable design. Thus an initial quadratic fitting can be attempted using one part of the design only. If the model is found to be inadequate, the second part, which completes the third order rotatable design, is then added. For our pruposes, we shall assume the latter to be the case.

The model (4.1.1) is of the same form as the one in (2.1.1). Thus we will use the generalized ridge regression method described in chapter II in the estimation of the unknown coefficients as illustrated below.

4.2 Illustration

As noted in the previous section, Draper[7] established that certain pairs of second order rotatable design classes could be combined in such a way that infinite classes of third order rotatable designs for three factors were formed. Two such classes of second order rotatable designs are shown in table 4.1. For each class is shown a reference designation D_1, the points which form the class, the number of points in the class and a restriction on the parameters which must hold in order that the fourth order relationships of equations (3.4.3) can be satisfied. The third order rotatable design given here is derived from

these two classes of second order rotatable designs by combining them as demonstrated by Draper[7].

Table 4.1 two classes of second order rotatable designs

reference	D_1	D_2
points in class	$(\pm a, \pm a, \pm a)$	$(\pm f, \pm f, 0)$
	$(\pm c_1, 0, 0)$	$(\pm f, 0, \pm f)$
	$(0, \pm c_1, 0)$	$(0, \pm f, \pm f)$
	$(0, 0, \pm c_1)$	$(\pm a, \pm a, \pm a)$
	$(\pm c_2, 0, 0)$	$(\pm c, 0, 0)$
	$(0, \pm c_2, 0)$	$(0, \pm c, 0)$
	$(0, 0, \pm c_2)$	$(0, 0, \pm c)$
No. of points	20	26
restriction	$c^4_1 + c^4_2 = 8a^4$	$c^4 = 2f^4 + 8a^4$

Source: Draper, N.R (1962) [8]

We take D_1 with $a = 1$, $c_1 = c_2 = \sqrt{2}$

and

D_2 with $f = \sqrt{2}$, $a = 1$, $c = 2$

Then for the third order rotatable design,

$n = 50, \quad \mu_2 n = 48, \quad \mu_4 n = 32, \quad$ and $\quad \mu_6 n = 16$

so that

$$\frac{\mu_4}{\mu_2^2} = \frac{50}{72} \qquad \textit{and} \qquad \frac{\mu_6 \mu_2}{\mu_4^2} = \frac{3}{4} \tag{4.2.1}$$

There are two stages in this example. In the first stage D_1 is used with four centre points and a response surface of second order is fitted to the 24 experimental results. Assuming that the model provides an inadequate representation of the true response relationship, it is decided, then, to complete the design by running the 26 point D_2 portion with no additional centre points, and to analyze the total set of 50 observations by fitting a third order response surface. For the analysis, we use the generalized ridge regression techniques of section 2.3, which, as has already been noted, is possible due to the nature of the model (4.1.1). The 24 observations from the first part of the design are listed in Table 4.2 and the 26 observations from the second part appear in Table 4.3.

Table 4.2

The design points and the observations, first set.

Number	x_1	x_2	x_3	y
1	-1	-1	-1	34.727
2	1	-1	-1	38.917
3	-1	1	-1	44.907
4	1	1	-1	24.641
5	-1	-1	1	24.658
6	1	-1	1	45.636
7	-1	1	1	33.702
8	1	1	1	5.374
9	1.4142	0	0	33.414
10	-1.4142	0	0	38.540
11	0	1.4142	0	40.393
12	0	-1.4142	0	40.687
13	0	0	1.4142	23.869

Table 4.2(cont.)

Number	x_1	x_2	x_3	y
14	0	0	-1.4142	33.727
15	1.4142	0	0	34.453
16	-1.4142	0	0	39.201
17	0	1.4142	0	38.335
18	0	-1.4142	0	40.092
19	0	0	1.4142	25.823
20	0	0	-1.4142	33.068
21	0	0	0	44.562
22	0	0	0	41.187
23	0	0	0	43.832
24	0	0	0	42.165

Table 4.3

The design points and observations, second set.

Number	x_1	x_2	x_3	y
25	-1.4142	-1.4142	0	17.252
26	1.4142	-1.4142	0	56.277
27	-1.4142	2	0	55.478
28	1.4142	2	0	5.946
29	-1.4142	0	-1.4142	40.856
30	1.4142	0	-1.4142	22.157
31	-1.4142	0	1.4142	19.753
32	1.4142	0	1.4142	7.718
33	0	-1.4142	-1.4142	39.282

Table 4.3 (cont.)

Number	x_1	x_2	x_3	y
34	0	2	-1.4142	33.651
35	0	-1.4142	1.4142	33.942
36	0	2	1.4142	0.059
37	-1	-1	-1	32.802
38	1	-1	-1	39.394
39	-1	1	-1	47.553
40	1	1	-1	24.402
41	-1	-1	1	22.005
42	1	-1	1	46.184
43	-1	1	1	35.271
44	1	1	1	5.314
45	2	0	0	31.472
46	-2	0	0	35.209
47	0	2	0	33.338
48	0	-2	0	39.156
49	0	0	2	1.883
50	0	0	-2	27.891

Source: Draper, N.R (1962)[8]

For the third order rotatable design, with parameters given by (4.2.1), the sub-matrices of the moment matrix $\mathbf{X'X}$ in (3.6.3) are,

$$G = \begin{bmatrix} 1.00 & 0.96 & 0.96 & 0.96 \\ & 1.92 & 0.64 & 0.64 \\ (symm.) & & 1.92 & 0.64 \\ & & & 1.92 \end{bmatrix}$$

(4.2.2)

$$K = \begin{bmatrix} 0.96 & 1.92 & 0.64 & 0.64 \\ & 4.80 & 0.96 & 0.96 \\ (symm.) & & 0.96 & 0.32 \\ & & & 0.96 \end{bmatrix}$$

(4.2.3)

$$\mu_4 \; I_1 = 0.64 I_1 \qquad \text{and} \qquad \mu_6 \; I_2 = 0.32 \text{ (a scalar)}$$

(4.2.4)

The orthogonal matrix is then

$$P = \begin{bmatrix} R & O & O & O \\ & S & O & O \\ & & T & O \\ (symm.) & & & U \end{bmatrix}$$

(4.2.5)

Where

$$R = \begin{bmatrix} 0.886 & -0.371 & -0.273 & -0.048 \\ -0.385 & -0.497 & -0.463 & -0.625 \\ -0.016 & -0.605 & 0.790 & -0.095 \\ 0.258 & 0.499 & 0.295 & -0.773 \end{bmatrix}$$

(4.2.6)

$$S = \begin{bmatrix} 0.707 & 0.707 & 0 \\ 0.707 & -0.707 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

(4.2.7)

$$T = \begin{bmatrix} 0.64 & -0.19 & -0.21 & -0.13 & 0.64 & -0.19 & -0.21 & -0.13 \\ -0.27 & -0.62 & 0.16 & -0.16 & -0.27 & -0.62 & -0.16 & -0.16 \\ 0.15 & -0.30 & 0.46 & 0.42 & 0.15 & -0.30 & 0.46 & 0.42 \\ 0.04 & 0.003 & 0.47 & -0.53 & 0.04 & 0.003 & 0.47 & -0.53 \\ 0.64 & -0.19 & -0.21 & -0.13 & -0.64 & 0.19 & 0.21 & 0.13 \\ -0.27 & -0.62 & -0.16 & -0.16 & 0.27 & 0.62 & 0.16 & 0.16 \\ 0.15 & -0.30 & 0.46 & 0.42 & -1.51 & 0.30 & -0.46 & -0.42 \\ 0.04 & 0.003 & 0.47 & -0.53 & -0.04 & -0.003 & -0.47 & 0.53 \end{bmatrix}$$

(4.2.8)

and

$$U = \begin{bmatrix} 0.900 & -0.262 & -0.291 & -0.189 & 0 \\ -0.374 & -0.870 & -0.222 & -0.232 & 0 \\ 0.213 & -0.417 & 0.652 & 0.596 & 0 \\ 0.060 & 0.004 & 0.664 & -0.745 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

(4.2.9)

The vector of eigenvalues is

λ=(6.371, 144.719, 62.629, 50.306, 32.000, 32.000, 32.000,
1.435, 292.150, 33.521, 31.654, 1.435, 292.150, 33.521,
31.654, 1.435, 292.150, 33.521, 31.654, 16.000)'

Using the notation of chapter II, the values of $\hat{\alpha}_{1(0)}$, $e_{1(0)}$, e^*_1, $\hat{\alpha}^*_1$ and b^{*1} have been tabulated in table 4.4 below.

Table 4.4

Summary of calculations

i j k	$\hat{a}_{ijk(o)}$	$e_{ijk(o)}$	e^*_{ijk}	$\hat{a}^*_{ijk}$	b^*_{ijk}
0	56.4716	0.0003	0.0003	56.455	56.004
11	-19.976	0.0025	0.0025	-19.926	-21.424
22	3.245	0.095	0.119	2.900	12.221
33	13.313	0.0056	0.0057	13.237	-4.772
12	-6.672	0.0225	0.024	-6.518	-10.264
13	-8.123	0.0152	0.0156	-7.998	1.046
23	-3.468	0.0831	0.101	-3.151	-3.151
1	-6.855	0.0213	0.0222	-6.706	-4.827
111	1.449	0.476	∞	0	0.526
122	-3.246	0.095	0.119	-2.902	0.0626
133	2.145	0.2173	0.469	1.4601	2.175
2	-1.237	0.6534	∞	0	-2.909
222	-0.1452	47.418	∞	0	2.542
112	6.689	0.0223	0.0234	6.536	-3.531
233	-3.231	0.0958	0.1202	-2.884	-0.865

KEY: **i j k** = 0 means that there is no entry.

Table 4.4(cont.)

i j k	$\hat{\alpha}_{ijk(0)}$	$e_{ijk(0)}$	e^*_{ijk}	$\hat{\alpha}^*_{ijk}$	b^*_{ijk}
3	-1.2	0.694	∞	0	-0.4047
333	2.1895	0.2086	0.4215	1.54	-1.34
133	-1.41	0.5055	∞	0	0.643
223	0.1212	68.108	∞	0	0.00595
123	-3.078	0.0156	0.1363	-2.708	-2.708

KEY:

i j k = 0 means that there is no entry.

The above illustration clearly demonstrates the simplicity with which the concept of minimum mean square error and ridge regression can be combined to give the 'best' estimates of the coefficients of a third order rotatable design in three factors. It should be noted that the same procedure can be applied to any number of factors.

CONCLUDING REMARKS

Statisticians have begun to realize that certain deliberately induced biases can dramatically improve estimation properties when there are several parameters to be estimated. This represents a radical departure from the tradition of unbiased estimation which has dominated statistical thinking since Gauss' development of least squares method.

Unbiased estimators have been used on literally millions of real problems, with generally satisfactory results. Biased estimators have not. As such, their theoretical superiority has yet to be tested in the rigors of wide spread application. Primarily, the ridge regression procedure is intended to overcome 'ill-conditioned' situations where correlations between the various predictor variables in the model cause the $\mathbf{X'X}$ matrix to be close to singular, and in the process giving rise to unstable parameter estimates. The estimates may, for example, have the wrong sign or be much larger than physical or practical considerations would deem reasonable. This work brings to notice one of the many statistical domains, namely, response surface designs, in which biased estimation may be a useful tool.

APPENDIX

```
      PROGRAM ITMETHOD
C     THIS PROGRAM READS  THE REGRESSION MATRIX  X(M,N),
      THE OBSERVED VALUES Y  AND FORMS THE DESIGN MATRIX
      FROM  WHICH IT  REDUCES  THE  DESIGN TO  CANONICAL
      FORM.       IT  THEN   CALCULATES   THE  REGRESSION
      COEFICIENTS AND THE MEAN SQUARE  ERRORS USING BOTH
      THE USUAL  REGRESSION METHOD AND  RIDGE REGRESSION
      METHOD BY ITERATIONS.
C
      DIMENSION  A(20,20),  X(20,20), Y(20),  BETA1(20),
      BETA2(20), G(20), XSTAR(20,20)
      REAL MSE1, MSE2, LAMBDA(20), P(20)
C
      OPEN(UNIT=5, FILE='INPUT', STATUS='OLD')
      OPEN(UNIT=6, FILE='OUTPUT', STATUS='NEW')
      M=9
      N=5
      READ(5, 100)((X(I,J),I=1,M),J=1,N)
      READ(5, 101)(Y(I),I=1,M)
      DO 5 I=1,N
      BETA1(I)=0.
      BETA2(I)=0.
      G(I)=0.
5     CONTINUE
```

```
      DO 10 I=1,N
      DO 10 J=1,N
      A(I,J)=0.
      B(I,J)=0.
10    CONTINUE
      DO 20 I=1,N
      DO 20 J=1,N
      DO 20 K=1,M
      B(I,J)=X(K,I)*X(K,J)+B(I,J)
20    CONTINUE
      CALL EIGEN(B,N,P,LAMBDA)
      DO 25 I=1,M
      DO 25 J=1,N
      XSTAR(I,J)=0
25    CONTINUE
      DO 30 I=1,M
      DO 30 J=1,N
      DO 30 K=1,N
      XSTAR(I,J)=X(I,K)*P(K,J)+XSTAR(I,J)
30    CONTINUE
      DO 40 I=1,N
      DO 40 J=1,N
      DO 40 K=1,M
      A(I,J)=XSTAR(K,I)*XSTAR(K,J)+A(I,J)
40    CONTINUE
      DO 50 I=1,N
```

```
      DO 50 J=1,M
      G(I)=XSTAR(J,I)*Y(J)+G(I)
50    CONTINUE
      YY=0
      DO 60 I=1,9
      YY=YY+Y(I)**2
60    CONTINUE
      CALL SOLVE(A,G,M,N,YY,K,ALFA1,ALFA2,SIG)
      MSE1=0
      MSE2=0
      DO 70 I=1,5
      MSE1=MSE1+(SIG*LAMBDA(I)+(ALFA1(I)*K(I,I))**2)/
      (LAMBDA(I)+K(I,I))**2
      MSE2=MSE2+(SIG*LAMBDA(I)+(ALFA2(I)*K(I,I))**2)/
      (LAMBDA(I)+K(I,I))**2
70    CONTINUE
      DO 80 I=1,N
      DO 80 J=1,N
      BETA1(I)=P(J,I)*ALFA1(J)+BETA1(I)
      BETA2(I)=P(J,I)*ALFA2(J)+BETA1(I)
80    CONTINUE
100   FORMAT(5F5.2/)
101   FORMAT(9F3.1/)
200   FORMAT(5F7.4/)
201   FORMAT(5F7.4/)
202   FORMAT(5F8.4/)
```

```
203   FORMAT('MSE BY OLSE METHOD IS ',F10.4/)
204   FORMAT(5F8.4/)
205   FORMAT('MSE BY RIDGE METHOD IS ',F10.4/)
300   FORMAT('INITIAL DESIGN MATRIX IS '/)
301   FORMAT('OBSERVED VALUES ARE'/)
400   FORMAT('THE ORTHOGONAL MATRIX IS'/)
401   FORMAT('VECTOR OF EIGEN VALUES IS'/)
402   FORMAT('EST. COEFF. BY OLSE IS'/)
404   FORMAT('EST. COEFF. BY RIDGE IS'/)
      WRITE(6,300)
      WRITE(6,100)((X(I,J),I=1,M),J=1,N)
      WRITE(6,301)
      WRITE(6,101)(Y(I),I=1,M)
      WRITE(6,400)
      WRITE(6,200)((P(I,J),I=1,N),J=1,N)
      WRITE(6,401)
      WRITE(6,201)(LAMBDA(I),I=1,N)
      WRITE(6,402)
      WRITE(6,202)(BETA1(I),I=1,N)
      WRITE(6,404)
      WRITE(6,204)(BETA2(I),I=1,N)
      WRITE(6,203)MSE1
      WRITE(6,205)MSE2
      STOP
      END
```

****

```
      SUBROUTINE EIGEN(A,N,U,LAMDA)
      REAL
     A(20,20),U(20,20),LAMBDA(20),B(20,20),C(20,20),D(20,20),
     IDENT(20,20),V(20),VZERO(20),Y(20),L,LZERO
                MMAX=10000
                MFREQ=500
                EPS=1.0E-9
                DO 1 I=1,N
           1    VZERO(I)=1
                DO 2 I=1,N
                DO 2 J=1,N
           2    IDENT(I,I)=1
                CALL MATEQ(IDENT,B,N,N)
                DO 11 I=1,N
                CALL MATVEC(B,VZERO,V,N,N)
                CALL VECLEN(V,LZERO,N)
                DO 5 M=1,MMAX
                IF ((M/MFREQ)*MFREQ.NE.M)GOTO 4
                CALL MATVEC(B,V,Y,N,N)
                CALL SCAVEC(1/L,Y,V,N)
           4    CALL MATVEC(A,V,Y,N,N)
                CALL VECLEN(Y,L,N)
                CALL SCAVEC(1/L,Y,V,N)
                IF(ABS(L-LZERO)/LZERO).LT.EPS)GOTO 7
           5    LZERO=L
                IM1=I-1
```

```
      FLAG=0
7     CALL MATVEC(A,V,Y,N,N)
      DO 8 K=1,N
      IF (ABS(V(K).LT.1E-3)GOTO 8
      IF (V(K)*Y(K).LT.0)L=-L
      GOTO 9
8     CONTINUE
9     LAMBDA(I)=L
      DO 10 K=1,N
10    V(K,I)=V(K)
      IF(I.GE.N)GOTO 11
      CALL SCAMAT(L,IDENT,C,N,N)
      CALL MATSUB(A,C,D,N,N)
      CALL MATMLT(D,B,C,N,N,N)
      CALL MATEQ(C,B,N,N)
11    CONTINUE
      RETURN
      END
++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
SUBROUTINE SOLVE(A,G,M,N,YY,K,ALFA1,ALFA2,SIG)
REAL A(20,20),G(20),K(20,20),B(20,20),BINV(20,20),ALFA1(20),
ALFA2(20),C(20),ALFAMD,ALFA,D,TOL
      TOL=1E-4
      ALFAMD=0
      D=0
      DO 1 I=1,N
```

```
      DO 1 J=1,N
      K(I,J)=0
      B(I,J)=0
      BINV(I,J)=0
1     CONTINUE
      ITER=1
2     DO 3 I=1,N
      B(I,I)=A(I,I)+K(I,I)
      BINV(I,I)=1/B(I,I)
3     CONTINUE
      DO 4 I=1,N
      ALFA(I)=0
      C(I)=0
4     CONTINUE
      DO 6 I=1,N
      C(I)=C(I)+BINV(I,I)*G(I)
      IF(ITER.EQ.1)GOTO 5
      ALFA2(I)=C(I)
      ALFA(I)=C(I)
      ALFAMD=ALFAMD+ALFA(I)**2
      GOTO 6
5     ALFA1(I)=C(I)
      ALFA(I)=C(I)
      D=D+ALFA1(I)*G(I)
      ALFAMD=ALFAMD+ALFA(I)**2
6     CONTINUE
```

REFERENCES

[1] ArapKoske, J.K : Fourth order rotatable designs. *Unpublished Thesis submitted to University of Nairobi, March 1984.*

[2] Box, G.E.P(1954) : The exploration and exploitation of response surfaces: some general considerations and examples. *Biometrics, Vol. 10, 16-60.*

[3] Box, G.E.P and J.S, Hunter(1957) : Multi-factor experimental designs for exploring response surfaces. *Annals of Mathematical Statistics, Vol. 28, 195-241.*

[4] Brown, P.J and C, Payne(1975) :Election night forecasting. *Journal of the Royal Statistical Society, series A, Vol. 138, 463-483.*

[5] Chatterjee, S and B, Price(1977) : Regression analysis by example. *John Wiley & sons, Inc., 175-192.*

[6] Cochran, W.G and G.M, Cox(1968) : Experimental designs. *John Wiley & sons, Inc., 335-353.*

[7] Draper, N.R(1960) : Third order rotatable designs in three dimensions. *Annals of Mathematical Statistics, Vol. 31, 865-874.*

[8] Draper, N.R(1960) : Third order rotatable designs in three dimensions. *Annals of Mathematical*

Statistics, Vol. 31, 865-874.

[9] Draper, N.R(1962) : Third order rotatable designs in three factors: Analysis. *Technometrics, Vol. 4, 219-234*

[10] Farebrother, R.W(1975) : The Minimum mean square error linear estimator and Ridge regression. *Technometrics, Vol. 17, 127-128.*

[11] Hemmerle, W.J(1975) : An explicit solution for generalized Ridge regression. *Technometrics, Vol. 17, 309-314.*

[12] Hoerl, A.E and R.W, Kennard(1970) : Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics, Vol. 12, 55-67.*

[13] Hoerl, A.E and R.W, Kennard(1970) : Ridge regression: Applications to nonorthogonal problems. *Technometrics, Vol. 12, 69-82.*

[14] Kibua, T.K : Improved estimation on regression analysis. *Unpublished dissertation submitted to University of Nairobi, June 1985.*

[15] Marquardt, D.W and R.D, Snee(1975) : Ridge regression in practice. *The American Statistician, Vol. 29, 3-19.*

[16] Walpole,R.E and R.H, Myers(1978) : Probability and statistics for engineers and scientists. *Macmillan Publishing Co., Inc., 349-358.*