MAJOR RESEARCH PROJECT

ON

REDUCTION OF DIMENSIONALITY OF RESPONSE SURFACE DESIGN MODEL - BAYESIAN APPROACH

(UGC-MRP File No. 42-37/2013 Dated 12-03-2013)

BY

Dr. N. CH. BHATRA CHARYULU

PRINCIPAL INVESTIGATOR
DEPARTMENT OF STATISTICS
UNIVERSITY COLLEGE OF SCIENCE
OSMANIA UNIVERSITY, HYDERBAD-7

REPORT SUBMITTED TO

UNIVERSITY GRANTS COMMISSION, NEW DELHI

(Project Duration: 01-04-2013 to 31-03-2017)

Dr. N.Ch. Bhatra Charyulu
M.Sc., M.Phil., Ph.D.
Principal Investigator
UGC- MRP

Department of Statistics
University College of Science,
Osmania University,
HYDERABAD- 500007.

Date:

FOREWARD

This is the final report of UGC sanctioned Major Research Project entitled **"Reduction of Dimensionality of Response Surface Design Model – Bayesian Approach"** bearing F.No. 42-37/2013 (SR) dated 12-03-2013, has been carried out by me as Principal Investigator, during the period 1st April, 2013 to 31st March, 2017, in the Department of Statistics, University College of Science, Osmania University, Hyderabad, Telangana State, India, 500 007. This report contains the detailed procedure for objectives / results achieved and problems attempted related to the objectives mentioned in the above entitled project submitted to UGC for sanctioning the MRP of problems attempted and the solutions or results obtained so far under the project. This project was assisted by **Dr. Sk. Ameen Saheb**, **Smt. T. Deepthi**, **Mr. K. Srinivas** and their cooperation is gratefully acknowledged.

(**N.Ch. Bhatra Charyulu**)

ACKNOWLEDGMENTS

I would like to express my sincere thanks to the **University Grants Commission**, **New Delhi** for providing financial assistance to carry out this project under UGC-MRP. I am also grateful to **Registrar**, Osmania University, **Dean**, Development & UGC-Affairs, Osmania University, and **Principal,** UCS, Osmania University for providing necessary facilities to carry out this research project smoothly.

I am thankful to **Prof. M. Vittal, Prof. M. Gopal Rao, Prof. P. Lakshmi Manga, Dr. K. Vani, Dr. C. Jayalakshmi, Dr. G. Jayasree**, for their help and encouragement throughout the period of this project. I am also thankful to all my well wishers who helped directly or indirectly in completing this project.

I would not have been able to complete this project work without the support from my family members and understanding of my goals and aspirations towards research.

- **Dr. N. Ch. Bhatra Charyulu**

CONTENTS

1. INTRODUCTION

1.1 INTRODUCTION

The Design of Experiments plays a key role in any scientific investigation. In early development stages, in the theory of experimental designs, only few designs were needed primarily for their applications in agricultural experiments. Subsequently their utility was found in various fields like physical, biological, sociological and engineering studies. As a result, many of the designs were developed with the increase of applications of experimental designs in Science & Technology, Medicine, Social Sciences, and Industry etc.

Response Surface Methodology is a collection of mathematical and statistical techniques developed by Box and Wilson in 1951 to find optimal settings of input factors or design variables that maximize or minimize or target measured response variables. The study to investigate the functional relationship between the response and the factor combination is known as response surface study. The response variable is a measured quantity whose value is assumed to be affected with changing the levels of the factors. Mathematically the response function is denoted as $\boldsymbol{y} = f\,(x_1, x_2, \ldots x_v) + \varepsilon$, where $f\,(x_1, x_2, \ldots x_v)$ is a polynomial function of degree k with v factors. The purpose of response surface is to determine and quantify the relationship between the values of one or more measurable response variable(s) and settings of a group of experimental factors presumed to affect the response(s) and also to find the settings of the experimental factors that produce the best value or best set of values of the response(s).

When the degree of polynomial increases, the size of the response surface design model also increases, then the complexity of analysis increases and the model is relatively inaccurate in higher dimensions, which leads to problem of dimensionality. The Dimensionality Reduction of model is minimizing the number of factors in the model with minimum loss of information in the data which is useful for concentrating on the detailed analysis about the significant factors in the model.

1.2 SELECTION OF BEST REGRESSION PROCEDURES

1.2.1 All Possible Regression: Let Y be the response variable and $\underline{X} = (X_1, X_2, \ldots X_p)$ be the vector of predictors. Assume Y and X are linearly related. The number of all possible combinations of linear functions existing are 2^p between the response and predictor variable(s) (each model includes the constant term). Examine R^2, S_e^2 and C_p statistic for each of the fitted model and select the best model among them.

1.2.2 Forward Selection: This Procedure begins with no predictors in the model except an intercept. Variables are checked one at a time and the most significant variable is added to the model at each stage. Select the maximum correlated variable with the response from the set of predictor variables and fit the model and test its significance. Compute the partial F-test for each of the independent variable which are not in the model and select the largest partial F value and compare this with a preselected value (F_1) for some α. If this value is greater than the preselected value then add this variable to the equation. The procedure is terminated when no independent variables in the equation have no significant contribution to the response variable.

1.2.3 Backward Elimination: Backward elimination procedure works in opposite direction of a Forward selection procedure. In this procedure, we begin with a model that contains all the predictors. Computing the partial F-test for each of the independent variables and select the lowest partial F value and compare this with a preselected value (F_1) for some α. If this value is less than the preselected value then remove the variable from the equation. Repeat the procedure until all the independent variables in the equation provides significant contribution to the response variable.

1.2.4 Step-wise Regression: This procedure is a combination of Forward and Backward procedures, in which at each step the predictor is selected with highest partial correlation with the response and examine its significance. The procedure terminates when the smallest partial F-value is greater than a preselected value and the smallest partial F-value of the next best predictor less than a preselected value.

1.3 SUMMARY OF THE REPORT

This report is devoted to study the Reduction of Dimensionality of Response Surface Design Model and its reduction in Bayesian Approach. The presentation of research work done is organized in six chapters. The present chapter of the thesis provides a brief introduction of Response Surface Design with a bird's eye view of some of the useful methods and chapter-wise summary of the report.

In Chapter – 2, an introduction to response surface design models and its Analysis and complete literature available on the Reduction of Dimensionality of Regression model and Response Surface Design Model for first and second order are presented briefly. Few major methods used for the reduction of size of the model (regression model and response surface model) with suitable examples is presented. An attempt is made to compare the various methods.

In Chapter – 3, the basic concepts for Bayesian theory, the Bayesian approach for the estimation of parameters of a simple and multiple regression model, Bayesian simulation algorithms like Gibb Sampler, Metropolis Hastings algorithm etc and the role of Bayesian theory for selecting the variables are presented. A brief review on Bayesian variable selection proposed by different authors are listed and also three Bayesian variable selection methods are described for selecting the significant factors of a regression model with suitable illustrations. An introduction to simulation software and a comparison of various Bayesian variable selection procedures is presented.

In Chapter – 4, variance of estimated response for first and second order response surface design models with the restrictions on the moment matrix is presented. Variance component indices for first order and second order response surface design models with restrictions on the moment matrix of the design are derived. Proportion of variance indices for first and second order response surface design models with and without restrictions on the moment matrix are illustrated with suitable examples.

In Chapter - 5, a brief introduction for response surface design and the necessity for reduction of dimensionality and the concept of multi-collinearity is discussed. In section 5.2 and section 5.3 an attempt is made to reduce the size of first and second order response surface design model with detailed step by step procedure of finding the best choice model with significant factors is illustrated with suitable examples under imposing and not imposing restrictions on moment matrix in nested approach is presented in theoretical, algorithmic approach. To compare estimated parameters with various existed regression methods is presented in the last section 5.5.

In chapter - 6, a brief introduction of Bayesian procedure and Bayes theorem is presented. In section 6.2 an attempt is made to propose the reduction size of the first and second order response surface design model by estimating its parameters in Bayesian approach using simulation procedures and the method is illustrated with suitable examples with and without restrictions imposed on the moment matrix. An attempt is made to compare the various existing methods are presented in the last section 6.5.

In chapter - 7, conclusions on the reduction in Dimensionality of Response Surface Design Model and its future scope using Machine learning are presented. Programs used for estimating the parameters using Bayesian approach and evaluation of posterior probability are presented in Appendix. An up to date literature used is presented in the bibliography.

2. REVIEW ON REDUCTION OF DIMENSIONALITY OF RESPONSE SURFACE DESIGN MODEL

2.1 INTRODUCTION

Let $x_1, x_2, \dots, x_v$ are 'v' factors each has 's' levels for experimentation and D denote the design matrix of the combination of the factor levels, given by

$$D = ((x_{u1}, x_{u2}, \dots , x_{uv})) \qquad (2.1.1)$$

where x_{ui} be the level of the i^{th} factor in the u^{th} factors combination (i=1, 2, ... v; u =1, 2 … N). Let Y_u denote the response at the u^{th} combination. The factor-response relationship is given by

$$E(Y_u) = f(x_{u1}, x_{u2}, \dots , x_{uv}) \qquad (2.1.2)$$

is called the response surface. Design used for fitting the response surface models are termed as 'Response Surface Design'. The functional form of the response surface may be first order, second order, etc. The general Response Surface Design Model for the given responses is

$$Y = X\underline{\beta} + \underline{\varepsilon} \qquad (2.1.3)$$

where,

Y_{Nx1}= [Y_1 Y_2 … Y_N] ' is the vector of responses,

$X_{N\times k}$ is the design matrix,

$\underline{\beta}_{k\times 1}$ is the vector of parameters and

$\underline{\boldsymbol{\varepsilon}}_{N\times 1}$ = [$\varepsilon_1, \varepsilon_2, \dots , \varepsilon_N$] ′ is the vector of random errors and follows $N(0, \sigma^2 I)$.

2.1.1 First Order Response Surface Design Model: The first order response surface design model at the u^{th} design point is

$$Y_u = \beta_0 + \beta_1 x_{u1} + \beta_2 x_{u2} + \ldots \ldots + \beta_v x_{uv} + \boldsymbol{\varepsilon}_{\mathbf{u}} \tag{2.1.4}$$

where,

Y_u is the response at u^{th} design point X_u, ($u = 1, 2, \ldots, N$)

$X = \begin{bmatrix} 1 & x_{11} & x_{12} & \ldots & x_{1v} \\ 1 & x_{21} & x_{22} & \ldots & x_{2v} \\ 1 & x_{31} & x_{32} & \ldots & x_{3v} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & x_{N1} & x_{N2} & \ldots & x_{Nv} \end{bmatrix}$ be the design matrix and

$\underline{\beta} = (\beta_0, \beta_1, \beta_2, \ldots \beta_v)$ is the vector of parameters

$\boldsymbol{\varepsilon}_{\mathbf{u}}$ is the random error corresponding to Y_u at u^{th} design point.

Then the estimated response at u^{th} design point is $\hat{Y}_u = X(X'X)^{-1}X'Y$ and variance covariance matrix of $\hat{Y}_u$ is $V(\hat{Y}_u) = (X'X)^{-1}\sigma^2$. The moment matrix $(X'X)$ is

$$X'X = \left[\begin{array}{c|ccc} N & \sum_{u=1}^{N} x_{u1} & \ldots & \sum_{u=1}^{N} x_{uv} \\ \hline \sum_{u=1}^{N} x_{u1} & \sum_{u=1}^{N} x_{u1}^2 & \ldots & \sum_{u=1}^{N} x_{u1}x_{uv} \\ \ldots & \ldots & \ldots & \ldots \\ \sum_{u=1}^{N} x_{uv} & \sum_{u=1}^{N} x_{u1}x_{uv} & \ldots & \sum_{u=1}^{N} x_{uv}^2 \end{array}\right] \tag{2.1.5}$$

The variance of the predicted response at the u^{th} design point is:

$$V(\hat{Y}_u) = V(\hat{\beta}_0) + \sum_{i=1}^{v} x_{ui}^2 V(\hat{\beta}_i) + \sum_{i=1}^{v}\sum_{j=1}^{v} x_{ui}x_{uj} \operatorname{Cov}(\hat{\beta}_i, \hat{\beta}_j) \tag{2.1.6}$$

Suppose the restrictions are imposed on X'X given in (2.1.5) towards reaching to orthogonality, i.e. as $\boldsymbol{\Sigma}\, x_{ui} = \boldsymbol{\Sigma}\, x_{ui}x_{uj} = 0$ and let $\boldsymbol{\Sigma}\, x^2_{ui} = N\lambda_2$ (the summation over $u = 1, 2, \ldots N$ and $i \neq j = 1,2\ldots v$). Then, the variance – covariance matrix X'X will be in the form

$$X'X = \left[\begin{array}{c|c} N & O \\ \hline O & N\lambda_2 I \end{array}\right]$$

Then the variance of the estimated response at the u^{th} design point is

$$V(\hat{Y}_u) = V(\hat{\beta}_0) + V(\hat{\beta}_i) \sum_{i=1}^{v} x_{ui}^2 \tag{2.1.7}$$

where, $V(\hat{\beta}_0) = \sigma^2/N$, $V(\hat{\beta}_i) = \sigma^2/N\lambda_2$ for i = 1, 2, . . . v.

2.1.2 Second Order Response Surface Design Model: The second order response surface design model at the u^{th} design point is

$$Y_u = \beta_0 + \sum_{i=1}^{v} \beta_i x_{ui} + \sum_{i=1}^{v} \beta_{ii} x_{ui}^2 + \sum_{\mathbf{i<j}}^{\mathbf{v}} \beta_{ij} x_{ui} x_{uj} + \boldsymbol{\varepsilon} \tag{2.1.8}$$

where,

Y_u is the response at the u^{th} design point,

$$X=\left[\begin{array}{c|cccc|cccc|cccc} 1 & x_{11} & x_{12} & \dots & x_{1v} & x_{11}^2 & x_{12}^2 & \dots & x_{1v}^2 & x_1x_2 & x_1x_3 & \dots & x_1x_v \\ 1 & x_{21} & x_{22} & \dots & x_{2v} & x_{21}^2 & x_{22}^2 & \dots & x_{2v}^2 & x_2x_3 & x_2x_4 & \dots & x_2x_v \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & x_{N1} & x_{N2} & \dots & x_{Nv} & x_{N1}^2 & x_{N2}^2 & \dots & x_{Nv}^2 & x_Nx_1 & x_Nx_2 & \dots & x_Nx_v \end{array}\right]$$ be the design matrix

$x_u = (1, x_{u1}, x_{u2}, \dots x_{uv}, x^2_{u1}, x^2_{u2}, \dots x^2_{uv}, x_{ul}x_{u2}, \dots x_{uv-1}x_{uv})$ be the u^{th} row of X

$\underline{\beta} = (\beta_0, \beta_1, \beta_2, \dots \beta_v, \beta_{11}, \beta_{22}, \dots \beta_{vv}, \beta_{l2}, \dots \beta_{v-1v})'$ be the vector of parameters

$\boldsymbol{\varepsilon_u}$ is the random error corresponding to Y_u.

Then the estimated response at u^{th} design point is $\hat{Y}_u = X(X'X)^{-1}X'Y$ and variance covariance matrix of $\hat{Y}_u$ is $V(\hat{Y}_u) = (X'X)^{-1}\sigma^2$. The moment matrix $(X'X)$ is

$$
\left[\begin{array}{c|ccc|ccc|ccc}
N & \sum_{i=1}^{N} x_{il} & \cdots & \sum_{i=1}^{N} x_{iv} & \sum_{i=1}^{N} x_{i1}^{2} & \cdots & \sum_{i=1}^{N} x_{iv}^{2} & \sum_{i=1}^{N} x_{i1}x_{i2} & \cdots & \sum_{i=1}^{N} x_{iv-1}x_{iv} \\
\hline
\sum_{i=1}^{N} x_{il} & \sum_{i=1}^{N} x_{i1}^{2} & \cdots & \sum_{i=1}^{N} x_{i1}x_{iv} & \sum_{i=1}^{N} x_{i1}^{3} & \cdots & \sum_{i=1}^{N} x_{i1}x_{iv}^{2} & \sum_{i=1}^{N} x_{i1}^{2}x_{i2} & \cdots & \sum_{i=1}^{N} x_{i1}x_{iv-1}x_{iv} \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\sum_{i=1}^{N} x_{iv} & \sum_{i=1}^{N} x_{i1}x_{iv} & \cdots & \sum_{i=1}^{N} x_{iv}^{2} & \sum_{i=1}^{N} x_{i1}^{2}x_{iv} & \cdots & \sum_{i=1}^{N} x_{iv}^{3} & \sum_{i=1}^{N} x_{i1}x_{i2}x_{iv} & \cdots & \sum_{i=1}^{N} x_{iv-1}x_{iv}^{2} \\
\hline
\sum_{i=1}^{N} x_{i1}^{2} & \sum_{i=1}^{N} x_{i1}^{3} & \cdots & \sum_{i=1}^{N} x_{i1}^{2}x_{iv} & \sum_{i=1}^{N} x_{i1}^{4} & \cdots & \sum_{i=1}^{N} x_{i1}^{2}x_{iv}^{2} & \sum_{i=1}^{N} x_{i1}^{3}x_{i2} & \cdots & \sum_{i=1}^{N} x_{i1}^{2}x_{iv-1}x_{iv} \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\sum_{i=1}^{N} x_{iv}^{2} & \sum_{i=1}^{N} x_{i1}x_{iv}^{2} & \cdots & \sum_{i=1}^{N} x_{iv}^{3} & \sum_{i=1}^{N} x_{i1}^{2}x_{iv}^{2} & \cdots & \sum_{i=1}^{N} x_{iv}^{4} & \sum_{i=1}^{N} x_{i1}x_{i2}x_{iv}^{2} & \cdots & \sum_{i=1}^{N} x_{iv-1}x_{iv}^{3} \\
\hline
\sum_{i=1}^{N} x_{i1}x_{i2} & \sum_{i=1}^{N} x_{i1}^{2}x_{i2} & \cdots & \sum_{i=1}^{N} x_{i1}x_{i2}x_{iv} & \sum_{i=1}^{N} x_{i1}^{3}x_{i2} & \cdots & \sum_{i=1}^{N} x_{i1}x_{i2}x_{iv}^{2} & \sum_{i=1}^{N} x_{i1}^{2}x_{i2}^{2} & \cdots & \sum_{i=1}^{N} x_{i1}x_{i2}x_{iv-1}x_{iv} \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\sum_{i=1}^{N} x_{iv-1}x_{iv} & \sum_{i=1}^{N} x_{i1}x_{iv-1}x_{iv} & \cdots & \sum_{i=1}^{N} x_{iv-1}x_{iv}^{2} & \sum_{i=1}^{N} x_{i1}^{2}x_{iv-1}x_{iv} & \cdots & \sum_{i=1}^{N} x_{iv-1}x_{iv}^{3} & \sum_{i=1}^{N} x_{i1}x_{i2}x_{iv-1}x_{iv} & \cdots & \sum_{i=1}^{N} x_{iv-1}^{2}x_{iv}^{2}
\end{array}\right]
$$

The $V(\hat{Y}_u)$ is not in a simplified form as the elements in moment matrix are in higher order, hence imposing the restrictions on the moment matrix that all odd power summations are zero, towards reaching to near orthogonality, i.e. $\sum x_{ui}^{\delta 1} x_{uj}^{\delta 2} x_{uk}^{\delta 3} x_{ul}^{\delta 4} = 0$. Let $\Sigma\, x^2_{ui} = N\,\lambda_2$; $\Sigma\, x^4_{ui} = CN\,\lambda_4$; $\Sigma x^2_{ui}\, x^2_{uj} = N\lambda_4$ then the moment matrix will be in the form

$$
N^{-1}X'X = \left[\begin{array}{c|ccc}
1 & 0 & \lambda_2 J & 0 \\
\hline
0 & \lambda_2 J & 0 & 0 \\
\lambda_2 J & 0 & [(c-1)I + J]\lambda_4 & 0 \\
0 & 0 & 0 & \lambda_4 I
\end{array}\right]
\text{ and } N(X'X)^{-1} = \left[\begin{array}{c|ccc}
\lambda_4(c+v+1)\Delta^{-1} & 0 & (c+v-1)(c-1)J\Delta^{-1} & 0 \\
\hline
0 & \lambda_2^{-1} I & 0 & 0 \\
-2\lambda_2 J\Delta^{-1} & 0 & Z_{VxV} & 0 \\
0 & 0 & 0 & \lambda_2^{-1}
\end{array}\right]
$$

where, $\Delta = [\,\lambda_4\,(c+v-1) - v\lambda_2^2]\; > 0$; $Z_{v\times v} = \dfrac{[(c+v-1)I_v - J_{v,v}]}{\lambda_4(c-1)(c+v-1)} + \dfrac{[\lambda_2^2(c+v-1)(c-1)^2]}{[\lambda_4(c+v-1) - k\lambda_2^2]} J_{k,k}$

Let $\sum_{i=1}^{v} x_{ui}^2 = \rho^2$; then $\sum_{i=1}^{v} x_{ui}^4 = \rho^4 - 2\sum_{i<j}^{v} x_{ui}^2 x_{uj}^2$.

The variance of the estimated response at the u^{th} design point is

$$V(\hat{Y}_u) = V(\hat{\beta}_0) + \rho^2 V(\hat{\beta}_i) + [\rho^4 - 2\sum_{i<j}^{v} x_{ui}^2 x_{uj}^2]\, V(\hat{\beta}_{ii}) + \sum_{i<j}^{v} x_{ui}^2 x_{uj}^2\, V(\hat{\beta}_{ij}) + 2\,Cov(\hat{\beta}_0, \hat{\beta}_{ii})\rho^2$$

$$+\ 2\ Cov(\hat{\beta}_{ii}, \hat{\beta}_{jj}) \sum_{i<j}^{v} x_{ui}^2 x_{uj}^2 \qquad (2.1.9)$$

where,

$$V(\hat{\beta}_0) = [\lambda_4 (c+k-1)/ N\Delta]\, \sigma^2$$

$$V(\hat{\beta}_i) = (1/ N\lambda_2)\, \sigma^2$$

$$V(\hat{\beta}_{ij}) = (1/ N\lambda_4)\, \sigma^2$$

$$Cov(\hat{\beta}_0, \hat{\beta}_{ii}) = [-\lambda_2 / N\Delta]\, \sigma^2$$

$$V(\hat{\beta}_{ii}) = [\{\lambda_4(c+k-2)-(k-1)\lambda_2^2\}/\{N\lambda_4 (c-1)\, \Delta\}]\, \sigma^2$$

$$Cov(\hat{\beta}_{ii}, \hat{\beta}_{jj}) = [(\lambda_2^2 - \lambda_4) / \{(c-1)\, N\lambda_4\, \Delta\}]\, \sigma^2$$

$$V(\hat{Y}_u)=V(\hat{\beta}_0)+[V(\hat{\beta}_i)+2Cov(\hat{\beta}_0, \hat{\beta}_{ii})]\rho^2+V(\hat{\beta}_{ii})\rho^4+[V(\hat{\beta}_{ij})-2V(\hat{\beta}_{ii})+2Cov(\hat{\beta}_{ii}, \hat{\beta}_{jj})] \sum_{i<j}^{v} x_{ui}^2 x_{uj}^2$$

$$\Rightarrow V(\hat{Y}_u) = \frac{\sigma^2}{N\Delta}\left[\left\{\frac{\Delta - 2\lambda_2^2}{\lambda_2}\right\}\rho^2 + \left\{\frac{\Delta - \lambda_2^2}{\lambda_4 (c-1)}\right\}\rho^4 - \left\{\Delta + v\lambda_2^2\right\}\right] + \left[\frac{(c-3)}{(c-1)N\lambda_4}\sigma^2\right]\sum_{i<j}^{v} x_{ui}^2 x_{uj}^2$$

(2.1.10)

If the variance of the estimated response at any design point is a function of ρ^2, i.e. the distance from design point to the origin. Then the design is called second order rotatable design. In other words, it follows that $V(\hat{Y}_u)$ at points equidistant from the centre are same. If c=3, the variance of estimated response can be expressed in the form of a function of ρ^2 as

$$V(\hat{Y}_u) = A\rho^4 + B\rho^2 + C \qquad (2.1.11)$$

where,

$$A = \frac{\sigma^2}{N\Delta}\left[\frac{\Delta - \lambda_2^2}{\lambda_4 (c-1)}\right]; \qquad B = \frac{\sigma^2}{N\Delta}\left[\frac{\Delta - 2\lambda_2^2}{\lambda_2}\right]; \qquad C = \frac{\sigma^2}{N\Delta}\left[\Delta + v\lambda_2^2\right].$$

2.2 REVIEW ON REDUCTION IN DIMENSIONALITY OF MODEL

When the number of factors and order of model are increasing, the difficulty level of the analysis of the response surface design model increases. It leads to the problem of dimensionality of the model. If the experimenter is interest to study only a few factors (or its combinations) and it is possible to eliminate the insignificant factors or its combinations from the model, which are not affecting much the response, then the size of the model can be reduced. It minimizes the time, cost, effort and data complexity with little loss of information.

Some of the well known dimensionality reduction techniques are Principal Component Analysis, Factor Analysis, Multidimensional Scaling, Discriminant Analysis, projection pursuit method etc. Most of the literature is found on the reduction of the dimensionality of Regression Models. No theoretical work has been carried out on the Response Surface Design Model reduction.

2.2.1 Reduction in size of Regression Model: Several researchers made attempts on the reduction of dimensionality of regression models. Some of them are: Friedman and Tukey (1974), Hastie and Tibshirani (1984), Breiman, et. al (1984), Diaconis and Friedman (1984), Breiman and Friedman (1985), Stone (1985, 1986), Engle, et. al (1986), Chen (1988), Loh and Vanichsetakul (1988), Haste and Stuetzle (1989), Saund (1989), Kramer (1991), Lin (1991), Foster and George (1994), Tibshirani (1996), Jin and Shaoping (2000), Li et. al (2000), Fan and Li (2001), Cheng and Wu (2001), Fan and Li (2002), Broman and Speed (2002), Efron, et. al (2004), Janathi, et. al (2004), Zou and Hastie (2005), Yuan and Lin (2006), Steel and Uys (2007), Wu and Liu (2009), Xu and Ying (2010), Radchenko and James (2011), Tengfei Long and Weili Jiao (2012), Luan Jaupi (2014), Parpoula Christina, et. al (2014), Tengfei Long, et. al (2014).

Friedman and Tukey (1974) proposed a projection pursuit method for the reduction of dimensionality. This method searches for linear projection onto the lower dimensional space that robustly reveals structures in the data. The fundamental idea behind projection pursuit is to search linear projection of the data onto a lower dimensional space their distribution is “interesting”; interesting is defined as being “far from the normal distribution", i.e. the normal distribution is assumed to be most uninteresting. The degree of "far from the normal distribution" is defined as being a projection index.

Hastie and Tibshirani (1984) proposed a flexible method to identify and characterize the effect of potential prognostic factors on an outcome variable in clinical trials. These methods are called generalized additive models. In Logistic Regression models, the effects of prognostic factors x_j are expressed in terms of a linear predictor of the form $\sum \beta_j x_j$. The additive models replaces $\sum \beta_j x_j$ with $\sum f_j(x_j)$ where f_j is an unspecified non-parametric function. This function is estimated in a flexible manner using a scatter plot smoother. The estimated function of $f_j(x_j)$ can reveal possible nonlinearities in the effect of x_j. Breiman et.al. (1984) proposed a tree-structured approach in which a regression model can be reduced by constructing tree by splitting the data with respect to the characterization at each step, with a partition of the entire data into several homogeneous groups. Since one can split the data in many possible ways, this leaves a great deal of flexibility criterion seems to project which are close to Gaussian. Diaconis and Friedman (1984) proposed a method of distribution of Projections. Most of the results are stated for one-dimensional projections. The data is set by projections and categorized into Gaussian (or nearly Gaussian) and non Gaussian projections.

Breiman and Friedman (1985) proposed the Alternating Conditional Expectation algorithm for estimating the transformations of a response and a set of predictor variables in multiple regression which produce the maximum linear effect between the transformed independent variables and response variable which gives the data analyst insight into the relationships between the variables, so that the relationships between them will be best described and nonlinear relationships can be uncovered. Stone (1985, 1986) developed the additive regression model for easier interpretation of the contribution of each explanatory variable and may be preferable to a fully nonparametric regression model for a moderate sample size. Engle et.al. (1986) proposed a class of partial spline curve models for smoothing data with reduced dimensions.

Chen (1988), Loh and Vanichsetakul (1988), constructed trees based on the information provided from the learning sample of objects with known class for each of the components of classifier construction using factor analysis and principal component analysis methods.

Hastie and Stutzle (1989) proposed the concept of a principal curve and developed a concrete algorithm to find the principal curve, which is represented by a parametric curve. The principal curves of a given distribution are not always unique. Saund (1989) used a three layer neural network with a single hidden layer which he called a “connectionist network”. While training the network, he noticed that in order to obtain a good dimensionality reduction, we need to identify the most effective number of units in the middle bottleneck layer but, he does not provide us with a way of directly identifying it.

Kramer (1991) proposed a non linear principal component analysis approach for training the feed forward network to obtain an identity mapping using sequential networks in cascade. Each network has single bottleneck layer and the output of one is fed into the second and the whole network is trained. Lin (1991) proposed sliced inverse regression for selecting the choice of variables in the regression model.

Foster and George (1994) developed a new criterion called the risk inflation, is the maximum possible increase in risk of the consequent selection or estimation procedure for selecting correct predictors which is used for the evaluation of variable selection procedures in multiple regression. Tibshirani (1996) proposed a new technique called Lasso - “least absolute shrinkage and selection operator” for estimating the parameters in the model so that it shrinks some of the coefficients to zero and hence try to retain the good features of both subset selection and ridge regression. In general, the idea of Lasso is extended to generalized regression models and tree-based models.

Jin and Shaoping (2000) used neural network approach for reduction of dimensionality for Chinese character recognition. Li et. al (2000) proposed iterative Tree-Structured regression for finding a direction, along which the regression surface bends. The direction is used for splitting the data into two regions, within each region another direction is found and the partition is done in the same manner. The process continues recursively until the entire regressor domain is decomposed into regions, where, the surface no longer bends significantly and linear regression fit becomes appropriate. For implementing the direction search, the Principal Hessian directions are used.

Fan and Lin (2001) proposed a unified non-concave penalized likelihood least squares regression approach called SCAD which performs variable selection and regression coefficient estimation simultaneously. Cheng and Wu (2001) made an attempt on factor screening and response surface exploration sequentially. This method is based on two stage analysis, even though it is based on two stage, it consists of three parts: screening analysis in stage 1, projection that links between stages 1 and 2 i.e. screening a larger number of factors and the more intensive study of the response surface over a smaller number of factors, and response surface exploration in stage 2.

Fan and Lin (2002) made an attempt on two new variable selection methods based on Fan and Lin (2001) paper and this non-concave penalized likelihood approach is extended to the Cox proportional hazards model and Cox proportional hazards frailty model which are used for survival analysis. And finally conclude that this new methods have better theoretic properties and finite sample performance. Broman and Speed (2002) focused on back-cross designs by considering the problem of identifying the genetic loci (known as quantitative trait loci QTLs) by reducing multi dimensional data to one dimensional data.

Efron, et. al (2004) developed a Least Angle Regression (LARS) algorithm for linear regression model which relates to the classic model selection method. And had given some modifications that turn LARS into Lasso or Stage-wise. Janathi, et. al (2004) described a method for data dimensionality reduction in non-linear projection of multidimensional data using a multi-layer neural network in auto-associative mode with a fast updating rule, based on a conjugate gradient algorithm. This connectionist approach gives an effective result in data dimensionality reduction when compared to principal component analysis method. And it has been demonstrated with an application to remotely sensed imagery (Landsat TM image of Ke´nitra region Morocco).

Zou and Hastie (2005) proposed a new technique called elastic net for the regression model which performs both automatic variable selection and continuous shrinkage. It selects the group of correlated variables and is like a stretchable fishing net that retains 'all the big fish'.

Yuan and Lin (2006) focused on the accuracy of estimation and extended the Lasso, the LARS algorithm and non-negative garrote for selecting the grouped variables (factors) in regression. And also proposed the efficient algorithms for the extensions of these methods for factor selection and showed that these extensions give superior performance to the traditional methods in regression. And also studied in detail about the similarities and differences between these methods with suitable illustrations.

Steel and Uys (2007) studied the influence measures based on Mallows' C_p statistic and Akaike's information criteria (AIC) for selecting the influence of an individual data in multiple linear regression. Wu and Liu (2009) developed variable selection in penalized quantile regression and extend the oracle properties of the SCAD and Adaptive – LASSO penalties to this quantile regression. Xu and Ying (2010) considered the median regression with a lasso type penalty term for variable selection and proposed two stage methods with fixed number of variables for simultaneous estimation and variable selection.

Radchenko and James (2011) proposed two variable selection methods, one is on Lasso, which computes highly shrunk regression coefficients and the other is based on Forward Selection, which uses no shrinkage on regression coefficients. From these two, **F**orward-**L**asso **A**daptive **SH**rinkage (FLASH) which includes both Lasso and Forward selection as special cases.

Tengfei Long and Weili Jiao (2012) solved the problem of Rational Functional Model consisting of 78 Rational Polynomial Coefficients (RPCs). They made an attempt to reduce the size of the rational functional model by converting the given polynomial into multiple linear regression model. The significant RPCs are selected one by one according to the criteria of goodness of fit.

Luan Jaupi (2014) made an attempt on the variable selection for Multivariate Statistical Process Control in two different approaches. One is on variable selection with pre-assigned role by assuming a two class system to classify the variables as primary and secondary based on different criteria. Then a double reduction of dimensionality is applied to select relevant primary variables that represent the whole set of variables. And the second one is on variable selection with Cost-utility Analysis, which is used to compare different variable subsets that might be used for process monitoring. The subset of relevant variables is selected in a manner such that the structure and information carried by the full set of original variables.

Parpoula Christina et. al. (2014) focused and discussed on the problem of selecting the significant variables in regression using SCAD, LASSO and Hard methods by using Supersaturated Designs (SSDs). These methods are much faster and provide more effective than penalized likelihood methods. Tengfei Long, et. al (2014) proposed an automatic variable selection of Rational Polynomial Coefficients (RPCs) based on nested regression and goodness of fit is used to evaluate the coefficients and applied the method on remote sensing images which includes (Quick Bird, SPOT5, Landsat-5, and ALOS).

2.2.2 Reduction in size of Response Surface Design Model: Only few authors made attempts on the reduction of size of the response surface design models. Kaufman, et. al (1996), Homma and Saltelli (1996), Venter, et. al (1998), Vignaux and Scott (1999), Lacey and Steele (2006) etc have studied on the reduction of the number of factors (or factor combinations) in the fitted response surface models for some experimental data. But, no much theoretical work is done on reduction of dimensionality in response surface designs and on its analysis.

Kaufman et. al (1996) proposed the variable complexity modelling approach is adapted for use with response surface approximation techniques. These methods are applied to multidisciplinary design of a High Speed Civil Transport (HSCT). Homma and Saltelli (1996) deal with new method of global sensitivity analysis of nonlinear models. In particular, global sensitivity analysis techniques are explored. By using FAST and Sobol series developments, the total variance is divided into sum of terms of increasing dimensionality. Computations of indices are done by Monte Carlo.

Venter, et. al. (1998) considered the data on the modelling of a plate with an abrupt change in thickness for a mechanical problem and fitted the data for linear response surface model and tried to reduce the number of variables using finite element analysis approach by solving the equations obtained through the boundary conditions. They reduced the number of variables from nine to seven. Vignaux and Scott (1999) proposed a method using statistical data from a survey.

Lacey and Steele (2006) applied the method of several engineering case studies including FE based example Show that better accuracy of the Response surface Designs can be obtained by using non dimensional variables.

2.3 METHODS FOR REDUCTION IN DIMENSIONALITY OF MODEL

In this section, popular methods proposed by the researchers used for the reduction of the size of the regression / response surface model are reviewed in detail with suitable examples.

METHOD 2.3.1: Saltelli and Chan (2000) proposed a method to reduce the size of the multiple regression models by eliminating some insignificant variables. A multiple regression model $Y = f(X) + \varepsilon$, with ‘p’ explanatory variables (X) and response variable (Y) is considered with $E(Y) = f(X)$. Standardize each input variable such that the inputs are normally distributed with mean zero and variance unity. The estimated coefficients of variables in the model will provide an insight into the relative significance of the inputs. For a standardized model, the square of the regression coefficients $\left(\beta_i^2\right)$ provides an estimate of a sensitivity index (S_i), which helps for ranking the input factors. Compute R^2 values and test the parameters for their significance to the reduction of size of the model. The method is illustrated in the example 2.3.1.

EXAMPLE 2.3.1: Let Y be the response variable measured along with its dependent variables X_1, X_2. X_3, X_4, X_5, X_6 and X_7 (Draper and Smith (1998)). The response and its associated independent variables observed for a sample of size 15 is presented below.

Table 2.3.1

S.No	**Y**	X_1	X_2	X_3	X_4	X_5	X_6	X_7
1	**201**	4.46	4.42	4.23	4.1	4.56	4.37	4.11
2	**224**	4.11	3.82	3.29	3.6	3.99	3.82	3.38
3	**301**	3.58	3.31	3.24	3.76	4.39	3.75	3.17
4	**301**	4.42	4.37	4.34	4.4	3.63	4.27	4.39
5	**301**	4.62	4.47	4.53	4.67	4.63	4.57	4.69
6	**309**	3.18	3.82	3.92	3.62	3.5	4.14	3.25
7	**311**	2.47	2.79	3.58	3.5	2.84	3.84	2.84
8	**311**	4.29	3.92	4	3.76	2.76	4.11	3.95
9	**312**	4.41	4.36	4.27	4.75	4.59	4.41	4.18
10	**312**	4.59	4.34	4.24	4.39	2.64	4.38	4.44
11	**333**	4.55	4.45	4.43	4.57	4.45	4.4	4.47
12	**351**	3.71	3.41	3.39	4.18	4.06	4.06	3.17
13	**411**	4.28	4.45	4.1	4.07	3.76	4.43	4.15
14	**424**	4.24	4.38	4.35	4.48	4.15	4.5	4.33
15	**424**	4.67	4.64	4.52	4.39	3.48	4.21	4.61

The sample correlation matrix is

	Y	X_1	X_2	X_3	X_4	X_5	X_6	X_7
Y	1.000	0.092	0.222	0.316	0.322	-0.155	0.302	0.276
X_1	0.092	1.000	0.901	0.673	0.736	0.291	0.677	0.897
X_2	0.222	0.901	1.000	0.852	0.740	0.278	0.835	0.936
X_3	0.316	0.673	0.852	1.000	0.757	0.093	0.882	0.911
X_4	0.322	0.736	0.740	0.757	1.000	0.437	0.812	0.822
X_5	-0.155	0.291	0.278	0.093	0.437	1.000	0.277	0.178
X_6	0.302	0.677	0.835	0.882	0.812	0.277	1.000	0.843
X_7	0.276	0.897	0.936	0.911	0.822	0.178	0.843	1.000

When a linear regression model is fitted to the data, estimates of the standardized regression coefficients are given as $\hat{\beta}_1$ = -2.256; $\hat{\beta}_2$ = 1.466; $\hat{\beta}_3$ = -1.690; $\hat{\beta}_4$ = 0.910; $\hat{\beta}_5$ = - 0.450; $\hat{\beta}_6$= -0.125; $\hat{\beta}_7$ = 1.907. The squares of the respective regression coefficients are: 5.089536, 2.149156, 2.8561, 0.8281, 0.2025, 0.015625, 3.636649. The parameters are tested for their significance based on the R^2 values that show percentage of output variation corresponding to each parameter in the model. As the parameter corresponding to the variable X_6 is insignificant, it is eliminated from the model to reduce its size.

METHOD 2.3.2: Mc Kay (1999) proposed a method using the correlations. Consider a multiple linear regression model $Y = f(X) + \varepsilon$, with 'p' explanatory variables (X) and response variable (Y). Assume the relationship between Y and X as $E(Y) = f(X)$. Let $r_{y,xi}$ be the correlation coefficient between the response variable (Y) and its associated variable (X_i). Evaluate the contribution of variance of each factor (X_i) with the output variance. Reduce the the size of the model by eliminating the factors from the model based on the strength and significance of the correlation between variables Y and X_i and contribution of the variance proportioned variance. The method is illustrated in the example 2.3.2.

EXAMPLE 2.3.2: Consider the data presented in Example 2.3.1, The correlation between the response and independent variables is {$r_{(X1,Y)}$, $r_{(X2,Y)}$, $r_{(X3, Y)}$, $r_{(X4 ,Y)}$, $r_{(X5,Y)}$, $r_{(X6,Y)}$, $r_{(X7,Y)}$}={ 0.092 0.222, 0.316, 0.322, - 0.155, 0.302, 0.276 }. The contribution of variances of the variables are examined which reduces the size of the model by eliminating X_6.

METHOD 2.3.3: Lin (1991) proposed a method called linear projection method. The detailed step by step procedure is presented below.

Step 1: Standardize the input vector of variables x as $z = (\hat{\Sigma}_{xx})^{-½} (x - \bar{x})$, where $\hat{\Sigma}_{xx}$ is the sample covariance matrix and $\bar{x}$ is the sample mean.

Step 2: Partition the range of response (y) into k slices I_1, $I_2 \ldots I_k$. Obtain the projection pursuits in p-dimensions for each slice as $\alpha = (\alpha_1^i, \alpha_2^i, \ldots, \alpha_p^i)$ for i= 1, 2, … , k

Step3: Obtain the variance-covariance matrix for the projection matrix **α.** Then, obtain the Eigen values and select first few largest eigen values and obtain their normalized Eigen vectors $\hat{e}_i$, i= 1, 2, …, m.

Step 4: Obtain $\hat{\beta}_i = \hat{e}_i (\hat{\Sigma}_{xx})^{-½}$ as direction estimates of the effective dimension reduction.

Remark: It is a simulated method so it has more complexity.

METHOD 2.3.4: This is a sensitivity method developed by Sobol (1993) to reduce the size of the model. The detailed step by step procedure for the evaluation of sensitivity indices is presented below.

Step-1: Consider a regression model Y= f (X) with X = (X_1, X_2, … , X_p) as the input vector of variables and Y the response variable. Let y = (y_1, y_2, … , y_n) be the vector of responses observed on the vector of variables X. Let the model be in the form

$$f(x) = \beta_0 + \sum \beta_i x_i + \sum\sum \beta_{ij} x_i x_j + \ldots \ldots \ldots + \beta_{12\ldots k} (x_1 x_2 \ldots x_p)$$

Step2: Decompose the variance D into partial variances associated with the each random input component as per the model.

$$\text{i.e., } D = \Sigma D_i + \Sigma D_{ij} + \ldots + D_{12\ldots p},$$

Step3: Compute the sobol sensitivity indices as proportion of the variances of parameters to total variance. Based on these indices, eliminate the component from the model, if it is negligible.

EXAMPLE 2.3.4: Consider the problem related to the two stage chemical process with five factors given by Box and Draper (1986), the variables under study are temperatures at two stages: X_1 and X_2, times of reactions at the two stages: X_3 and X_4, and the concentration of one of the treatments at the first stage: X_5 .

Table 2.3.2

Design Points	X_1	X_2	X_3	X_4	X_5	Response (Y)
1	-1	-1	-1	-1	1	49.2
2	1	-1	-1	-1	-1	51.2
3	-1	1	-1	-1	-1	50.4
4	1	1	-1	-1	1	52.4
5	-1	-1	1	-1	-1	49.2
6	1	-1	1	-1	1	67.1
7	-1	1	1	-1	1	59.6
8	1	1	1	-1	-1	67.9
9	-1	-1	-1	1	-1	59.3
10	1	-1	-1	1	1	70.4
11	-1	1	-1	1	1	69.6
12	1	1	-1	1	-1	64.0
13	-1	-1	1	1	1	53.1
14	1	-1	1	1	-1	63.2
15	-1	1	1	1	-1	58.4
16	1	1	1	1	1	64.3
17	3	-1	-1	1	1	63.0
18	1	-3	-1	1	1	63.8

Table 2.3.2

Design Points	X_1	X_2	X_3	X_4	X_5	Response (Y)
19	1	-1	-3	1	1	53.5
20	1	-1	-1	3	1	66.8
21	1	-1	-1	1	3	67.4
22	1.23	-0.56	-0.03	0.69	0.7	72.3
23	0.77	-0.82	1.48	1.88	0.77	57.1
24	1.69	-0.3	-1.55	-0.5	0.62	53.4
25	2.53	0.64	-0.1	1.51	1.12	62.3
26	-0.08	-1.75	0.04	-0.13	0.27	61.3
27	0.78	-0.06	0.47	-0.12	2.32	64.8
28	1.68	-1.06	-0.54	1.5	-0.93	63.4
29	2.08	-2.05	-0.32	1	1.63	72.5
30	0.38	0.93	0.25	0.38	-0.24	72.0
31	0.15	-0.38	-1.2	1.76	1.24	70.4
32	2.3	-0.74	1.13	-0.38	0.15	71.8

Consider the first order response surface model to be fitted as

$$Y = \beta_o + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5$$

The least squares estimates of the un-standardized parameters are $\hat{\beta}_0 = 59.687$; $\hat{\beta}_1 = -2.188$; $\hat{\beta}_2 = 0.792$; $\hat{\beta}_3 = 1.773$; $\hat{\beta}_4 = 2.411$; $\hat{\beta}_5 = 1.267$; and the standard errors of the estimated parameters are 1.409, 1.078, 1.150, 1.121, 1.146, 1.146. The proportion of variance indices corresponding to the parameters are $s_0 = 1.158$; $s_1 = 0.677$; $s_2 = 0.776$; $s_3 = 0.854$; $s_4 = 0.766$; $s_5 = 0.766$, which are used for ranking the parameters β_i's to reduce the model.

METHOD 2.3.5: Homma and Saltelli (1996) made an attempt to reduce the size of the response surface design model using sobol sensitivity indices. The detailed step by step procedure is presented below.

Step 1: Let $Y = f(X)$ be the response surface design model, in the form of $Y = X\beta + \varepsilon$, where $\underline{\mathbf{Y}}_{N\times1} = (Y_1, Y_2, \ldots, Y_N)'$ be the vector of observations corresponding to the N treatment combination of 'p' factors $(X_1, X_2, \ldots, X_p)$ and X be the design matrix. Assume the model contains k components.

Step-2: Estimate the vector of parameters in the model and evaluate the variances of each component C_i in the model and the variance of estimated response $V(C_i)$.

Step-3: Compute the ratio's of individual component variance to total variance, i.e. $S_i = V(C_i) / V(\hat{Y})$ for i= 1, 2, … k

Step-4: Eliminate the components whose ratio indices S_i 's are insignificant.

EXAMPLE 2.3.5: A chemical engineer is investigating the yield (Y) of a process, processed through each of the process variable Temperature (^{0}c), Pressure (Psig), and Catalyst Concentration (g/l) was run at a low and a high level, and the engineer decides to run a 2^3 design with four center points. The process variables coded in terms of levels ±1 for each of the design factor. The design and the resulting yields are presented in Table 2.3.3.

Table 2.3.3

S.No.	Natural Variables			Coded Variables			Yield
	Temperature (^{0}c)	Pressure (Psig)	Catalyst Concentration (g/l)	X_1	X_2	X_3	Y
1	125	41	14	-0.75	-0.95	-1.133	32
2	158	40	15	0.9	-1	-1	46
3	121	82	15	-0.95	1.1	-1	57
4	160	80	15	1	1	-1	65
5	118	39	33	-1.1	-1.05	1.14	36
6	163	40	30	1.15	-1	1	48
7	122	80	30	-0.9	1	1	57
8	165	83	30	1.25	1.15	1	68
9	140	60	22.5	0	0	0	50
10	140	60	22.5	0	0	0	44

11	140	60	22.5	0	0	0	53
12	140	60	22.5	0	0	0	56

Assume the experimental data satisfies first order model, $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$. The $X'X$ and $(X'X)^{-1}$ are

$$X'X = \begin{bmatrix} 12 & 0.6 & 0.25 & 0.007 \\ 0.6 & 8.18 & 0.31 & 0.14575 \\ 0.25 & 0.31 & 8.5375 & -0.07065 \\ 0.007 & 0.14575 & -0.07065 & 8.583289 \end{bmatrix};(X'X)^{-1} = \begin{bmatrix} 0.08368 & -0.00605 & -0.00223 & 0.00002 \\ -0.00605 & 0.12289 & -0.00430 & -0.00212 \\ -0.00223 & -0.00430 & 0.11736 & 0.00104 \\ 0.00002 & -0.00212 & 0.00104 & 0.11655 \end{bmatrix}$$

Estimated values for parameters are: $\hat{\beta}_0 = 50.520$; $\hat{\beta}_1 = 5.372$; $\hat{\beta}_2 = 10.130$; $\hat{\beta}_3 = 1.091$ and the estimated variance is 104.8626. The sobol indices for each component in the model at a particular design point are: $S_0 = 0.16345$; $S_1 = 0.37505$; $S_2 = 0.30315$; $S_3 = 0.22764$. Reduce the size of the model by ranking the indices.

Note: It can be noted that there is ambiguity in the method. The elimination of components is changing from one design point to another design point. The ratio indices at each design point are evaluated and presented in Table 2.3.4.

Table 2.3.4

S.No.	Design Variables			Indices corresponding to each component in the model		
	X_1	X_2	X_3	X_1	X_2	X_3
1	-0.75	-0.95	-1.133	0.16692	0.255769	0.361287
2	0.9	-1	-1	0.234601	0.276598	0.274688
3	-0.95	1.1	-1	0.239871	0.307128	0.252073
4	1	1	-1	0.294388	0.281141	0.2792
5	-1.1	-1.05	1.14	0.283681	0.246847	0.288969
6	1.15	-1	1	0.343139	0.247787	0.246077
7	-0.9	1	1	0.227658	0.268412	0.266559
8	1.25	1.15	1	0.375051	0.303158	0.227649

9	0	0	0	0	0	0
10	0	0	0	0	0	0
11	0	0	0	0	0	0
12	0	0	0	0	0	0

2.4 REMARKS ON REDUCTION IN DIMENSIONALITY OF MODELS

High-dimensional data sets or models make challenges as well as some opportunities bound to give rise to new theoretical developments. This can be studied in two aspects:

(i) Minimizing the number of factors or factor combinations in the model with minimum loss of information in the data and

(ii) Constructing the designs with minimum number of design points keeping in view the factors that are more active.

Though, several methods on the reduction of dimensionality of the regression model are available in the literature, it appears that no significant work has been done to give a proper criterion for the selection of variables in the regression model. No specific exclusive method is used for model reduction. The statistical techniques used for model reduction by the researchers are Principal Component Analysis, Factor Analysis, Multidimensional Scaling, Stepwise regression, Correlation approach, Regression approach, Sobol sensitivity approach, Fourier amplitude and Projection pursuit, Finite element methods etc. All these techniques are well known statistical / mathematical methods used to fit experimental data model and reduce the model by eliminating insignificant variables based on the parameter in the response surface model.

The feasibility of the Principal Component Analysis, Factor Analysis, Multidimensional Scaling, Stepwise regression, Correlation approach, Regression approach are studied for the response surface design models and observed that these methods have given fruitful results. Principal Component Analysis searches lower dimensional space that computes majority of the variation within the data and discovers linear structure in the data. However, this is ineffective in analyzing nonlinear structures, i.e. curves, surfaces or clusters. Hence it is not applicable for response surface model of second and higher orders. Factor Analysis is not applicable for response surface models due to the study of the interaction effects in the model and estimation of main effects of individual factors. Correlation approach, Regression approach, Sobol sensitivity approach, Fourier amplitude test and Projection pursuit can be used for response surface model reduction but Fourier amplitude test and Projection pursuit methods are very time consuming and their time complexity is more.

3. SELECTION OF VARIABLES IN BAYESIAN APPROACH

3.1 INTRODUCTION

Let A_1, A_2, … A_n are ‘n’ mutually exclusive and exhaustive events occurring in the sample space S (i.e. $S = \cup A_i$ and $A_i \cap A_j = \phi$, $i \neq j = 1, 2, \ldots n$) and B is any other event occurring in the same sample space(i.e $B \subseteq S$), such that P(B) > 0 then the probability of A_i given B, denoted by $P[A_i/B]$ is

$$P[A_i / B] = \frac{P[A_i]P[B/A_i]}{\sum_{i=1}^{n} P[A_i]P[B/A_i]} \quad \forall\ i = 1, 2, \ldots n \tag{3.1.1}$$

It states that if A and B are events, B represents our observation and P(B) > 0. The event A represents a model that might be true event. The P(A), is our initial belief about the probability of A being true, prior estimate of probability,. The P(B/A) is the probability of event B occurring if A is true, called likelihood factor. The P(A/B) is the probability of A being true given that B has been observed, posterior estimate. That is finally, consider the range of possible A’s in B to be estimated based on the observed B, Since B did happen, the total probability of B happening for any A, divide with P(B) to normalize the answer.

$$P(A/B) = P(A)\, P(B/A) / \ P(B)$$

Let $Y = [\ Y_1\ Y_2 \ldots Y_n\ ]'$ be the observed sample drawn from a population whose density is $f\,(Y, \theta)$ where the parameter θ is unknown and follows a probability distribution $f(\theta)$. The joint density function of observed sample $\underline{Y}$ for the given parameter θ, called the ‘likelihood function’ of $\underline{Y}$ denoted by $f\,(Y/\theta)$ and the probability distribution for the parameter θ encapsulates the prior beliefs held about their most likely values called ‘Prior distribution’ of θ denoted by $f\,(\theta)$. Then an updated measure of our beliefs for each of the parameter values θ based on our prior beliefs and given knowledge of the data $\underline{Y}$ called the 'Posterior distribution' of θ given Y denoted by $f(\theta/Y)$ is

$$f(\theta/Y) = \frac{f(\theta)\ f(Y/\theta)}{\int f(\theta)\, f(Y/\theta)\ d\theta} \tag{3.1.2}$$

It can be expressed as

P(Parameter / data) = [P(data / parameter) * P(parameter)] / P(data)

i.e. Posterior distribution is equal to Likelihood * Prior / Constant

Bayesian probability of any event is the ratio between the value at which an expectation depending on the happening of the event ought to be computed and the chance of the thing expected upon it's happening It demonstrates the estimation of future occurrences of an event, given information of the history of the event. It also allows us to obtain useful information even from a single piece of evidence or all evidence to be taken into account in an explicit way and combined in the overall probability model or included via the prior belief. The prior must be defined for every parameter to strengthen in this approach.

The Bayesian approach allows us to examine the probability that a possible model is true and also to compare the different possible models by assessing their relative probabilities of being true based on the given data. Obtaining appropriate statistical inferential statements from the posterior distribution is the process of fitting a Probabilistic model to an observed data and summarizing the result by a Probability distribution on the parameters of the model and on observed quantities such as predictions for future unobserved data.

If the Prior probability distribution $f(\theta)$ and the Posterior distribution $f(\theta/Y)$ belong to same family then the distributions are said to be 'Conjugate Pair of distributions'. Few of the conjugate family of distributions are presented in Table 3.1.1.

Table: 3.1.1 Conjugate Family of Distributions

S.No.	Likelihood Distribution	Prior Distribution	Posterior Distribution
1	Bernoulli (p)	Beta (a, b)	Beta First Kind ($a + \Sigma Y_i$, $b+n-\Sigma Y_i$)
2	Binomial (k, p)	Beta First kind (a, b)	Beta First Kind ($a + \Sigma Y_i$, $b+nk-\Sigma Y_i$)
3	Binomial (k, p)	Uniform (0,1)	Beta First Kind ($a + \Sigma Y_i$, $b+nk-\Sigma Y_i$)
4	Poisson (λ)	Gamma (a, b)	Gamma ($a + \Sigma x_i$, b+n)
5	Negative Binomial (p, r)	Beta (α , β)	Beta First Kind ($a - \Sigma x_i$, b+ r n)
6	Geometric (p)	Beta (α, β)	Beta ($\alpha' = \alpha+n$; $\beta' = \beta + \Sigma x_i$)
7	Hyper-geometric (N,M,n)	Beta-binomial n=N,$\alpha\beta$	Beta ($a + \Sigma x_i$, $b+\Sigma n-\Sigma x_i$)
8	Multinomial (p_1, .., .p_k)	Dirichlet (a)	Dirchlet(α') where $\alpha' = a + \Sigma xi$
9	Normal (μ , σ_0^2)	Normal (μ , σ^2)	Normal (μ' , σ'^2)
10	Normal (μ_0 , σ^2)	Gamma (a, b)	Gamma (a', b')

A Bayesian analysis provides direct statements about the quantities of interest, providing more intuitive results and feeding naturally into a decision making process. Bayesian statistics allows us to obtain some useful information even from a single piece of evidence, whereas many more samples would be required for a frequentist statistical result. The Bayesian approach allows all evidence to be taken into account in an explicit way. Different forms of evidence can be combined in the overall probability model or included via the prior belief. Analyzing the data using different priors allows the data to be interpreted from different points of view, such as regulatory or business.

The Bayesian framework more naturally allows for the modeling of biases or systematic error, and for modeling any hierarchical structure in the data or in the problem. Bayesian statistics provides greater flexibility. It offers a natural way to adapt an experiment in progress in the light of results collected so far, or to halt the experiment early if the result is clearer than expected and no more data is needed to achieve the desired degree of certainty.

3.2 BAYESIAN ESTIMATION OF PARAMETERS

Let $Y = (Y_1, Y_2, \ldots, Y_n)$ be the observed random sample drawn from a population with density function $f(x,\theta)$ or $P(Y, \theta)$ where the parameter θ is unknown and the distribution of the parameter θ is $f(\theta)$ or $P(\theta)$. Evaluate the likelihood function $P(Y / \theta)$, the joint density of Y and θ is defined as $P(Y, \theta) = P(\theta)\, P(Y/\theta)$. Estimate the value of the parameter θ, based on the observed sample, then the distribution of parameter θ is $f(\theta)$ or $P(\theta)$ (Prior probability). Evaluate the likelihood function with the estimated parameter $f(Y/\theta)$. Evaluate the conditional distribution of the parameter based on the sample using Bayesian rule as $f(\theta/Y) = [f(\theta).\, f(Y/\theta)] / [f(Y)]$, called Posterior distribution of parameter, with the evaluation of Normalized constant $f(y) = \int f(\theta).\, f(Y/\theta)\, d\theta$. or $P(Y) = \int P(\theta)\, P(Y/\theta)\, d\theta$.

Identify the full probability model of all observable and unobservable quantities, which involves defining the likelihood and prior density functions to be used in estimation. Then evaluate the conditional density of the variables to be estimated given the observed data (posterior) empirically. Evaluate the implication of the posterior and check the accuracy of the estimated quantities. The diagrammatically representation of the process is presented in Fig 3.2.1.

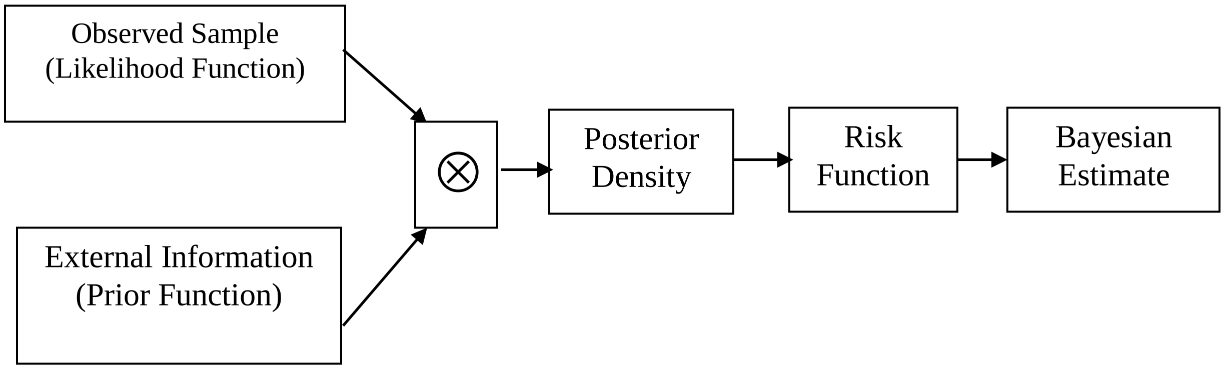

Fig 3.2.1

3.2.1 Metropolis Hastings Algorithm: Bayesian Statistics does not depend on Markov chains. The Metropolis Hasting algorithm is a Markov chain Monte Carlo method used to generate a sequence of random samples from a probability distribution for which direct sampling is difficult. This sequence can be used to approximate the distribution or to compute an expected value (integral). It generates a sequence of sample values in such a way that, as more and more sample values are produced, the distribution of values more closely approximates the desired distribution. These sample values are produced iteratively, with the distribution of the next sample being dependent only on the current sample value. At each iteration the algorithm picks a candidate for the next sample value based on the current sample value. Then with some probability, the candidate is either accepted or rejected the probability of acceptance is determined by comparing the likelihoods of the current or candidate sample values with respect to the desired distribution. It is making the computations in Bayesian Statistics easier. It is used to generate data from the posterior without needing to know some of the parts that are harder to find.

Generate a time reversible Markov chain whose stationary probabilities are $P_j = b_j /B$; for ($j = l, 2,. . .$), where b_j ($j = 1, 2, . . .$) be positive numbers and B is sum of the positive numbers chosen, which is finite. Define a Markov chain $\{X_n, n \geq 0\}$ as, when $X_n = i$, generate a random variable Y such that $P[Y = j] = q_{ij}$, ($j = 1, 2, ...$). Let $Q = ((q_{ij}))$ be any specified irreducible Markov transition probability matrix on the integers. If Y = j, then set $X_{n+1} = j$ with probability α_{ij}, and Y = i with probability $1- \alpha_{ij}$. Under these conditions, the sequence of states constitutes a Markov chain with transition probabilities P_{ij} are given by $P_{ij} = q_{ij}\alpha_{ij}$, if $j \neq i$ and

$P_{ij} = q_{ii} + q_{ik} (l - \alpha_{ik})$ if $j=i$. This Markov chain will be time reversible and have stationary probabilities P_j $P_j = b_j/B$, if $P_i P_{ij} = P_j P_{ji}$; for $j \neq i$ which is equivalent to $P_i q_{ij} \alpha_{ij} = P_j q_{ji} \alpha_{ji}$; If $\alpha_{ji} = 1$ then, $\alpha_{ij} = q_{ji} P_j / q_{ij} P_i$ and if $\alpha_{ij} = 1$ then $\alpha_{ji} = q_{ij} P_i / q_{ji} P_j$. From the above two we have $\alpha_{ij} = \min\{ q_{ji}P_j / q_{ij}P_i , 1\}$ and $\alpha_{ij} = \min\{ b_jq_{ji} / b_iq_{ij}, 1 \}$ which shows that the value of B is not needed to define the Markov chain, because the values b_j suffice. The implementation of the algorithm is presented below.

Step-1: Specify the target distribution from which the samples to be generated $P(\theta)$ (Prior distribution) which is the stationary distribution of Markov chain.

Step-2: Select an initial point θ^* at random from the proposed distribution $q(\theta^*/\theta_{t-1})$

Step-3: Calculate $\alpha = \text{Min} \{1, [P(\theta^*)/P(\theta)] . [q(\theta_{t-1}/\theta^*) / q(\theta^*/\theta_{t-1})] \}$

Step-4: Generate a uniform random number $u \sim U[0,1]$.

Step-5: Examine whether $u \leq \alpha$. If $u \leq \alpha$ then define $\theta_t = \theta^*$

Step-6: Repeat steps 2 to 5 until it generates a sequence $\{\theta_1, \theta_2, ... \theta_n\}$

3.2.2 Gibbs Sampling: Gibbs sampling is a particular form of Markov chain Monte Carlo (MCMC) algorithm for approximating the joint and marginal distribution by sampling from conditional distributions. If the joint distribution is not known explicitly or is difficult to sample from directly, but the conditional distribution is known or easy to sample from. Even if the joint distribution is known, the computational burden needed to calculate it may be huge. Gibbs Sampling algorithm could generate a sequence of samples from conditional individual distributions, which constitutes a Markov chain, to approximate the joint distribution. It can sample from conditional distribution while other parameters are fixed.

Let us assume that the posterior distribution is $P(\underline{\beta})$ where $\underline{\beta} = (\beta_1, \beta_2, \dots \beta_p)'$ vector of parameters. The full conditional distributions $P_i(\beta_i) = P(\beta_i / \beta_{-i})$; $i = 1, 2, \dots, p$ are available i.e., these are completely known and can be sampled from $P(\underline{\beta})$, where $\beta_{-i} = (\beta_1, \dots \beta_{i-1}, \beta_{i+1}, \dots \beta_p)'$.

Step 1: Set the initial values for the vector of parameters $\underline{\beta}$ as $\beta^{(0)} = (\beta_1^{(0)}, \beta_2^{(0)}, \dots, \beta_p^{(0)})'$

Step 2: Obtain the updated estimated values for the vector of parameters $\underline{\beta}$ as $\underline{\beta}^{(j)}$ where $\underline{\beta}^{(j)} = (\beta_1^{(j)}, \beta_2^{(j)}, \dots, \beta_p^{(j)})$ based on the distributions

$$\beta_1^{(j)} \sim P(\beta_1 / \beta_2^{(j-1)}, \beta_3^{(j-1)}, \dots, \beta_p^{(j-1)});$$

$$\beta_2^{(j)} \sim P(\beta_2 / \beta_1^{(j-1)}, \beta_3^{(j-1)}, \dots, \beta_p^{(j-1)}),$$

$$\dots \quad \dots \quad \dots \quad \dots$$

$$\beta_p^{(j)} \sim P(\beta_d / \beta_1^{(j)}, \beta_2^{(j)}, \dots, \beta_{p-1}^{(j)})$$

Step 3: Change the state j to j+1 and return to step 2. The process is continues until convergence is obtained. When convergence is obtained the values $\beta^{(j)}$ correspond to

$P(\beta_1, \beta_2, \dots \beta_p / \mathbf{y})$.

Note: To evaluate the posterior distribution, some of the software's can be used are Win BUG, Open BUG, R-BRIGGS, etc. or can also be evaluated using R-programming based on the prior and likelihoods.

3.3 BAYESIAN SELECTION OF VARIABLES

Several authors made attempts on the optimal choice of the model using Bayesian approaches. Some of them are: Mitchell and Beauchamp (1988), George and McCulloch (1993), Carlin and Siddhartha Chib (1995), Chipman (1996), Dellaportas et. al. (1997), George and McCulloch (1997), Chipman, et. al (1997), Kuo and Mallick (1998), George and Foster (2000), Wang and George (2004), Park and Casella (2008), O'Hara and Sillanppa (2009), Lindsey Charles and Sheather Simon (2010), Chen and Wang (2010), Weinwurm Stephan, et. al (2013), Otava Martin, et. al (2014), Elangovan and Lokeshmaran (2014), and Xu and Ghosh (2015).

Mitchell and Beauchamp (1988) proposed a Bayesian variable selection method for selecting the best subset of predictor variables in a linear regression model for the prediction of a dependent variable by introducing "spike and slab" prior distributions for the regression coefficients in linear regression models.

George and McCulloch (1993) developed the Stochastic Search Variable Selection (SSVS) procedure via Gibbs sampler for selecting or identifying the most promising subsets of predictor variables for the entire regression setup in a hierarchical Bayes normal mixture model with real data examples.

Carlin and Siddhartha Chib (1995) adopted a framework for Bayesian model choice, along with an MCMC algorithm called Gibbs sampling and applied this algorithm to two data examples.

Chipman (1996) developed the standard independent priors which are used in any variable selection procedures. These priors are incorporated in Bayesian variable selection algorithm for any type of linear model. This application is illustrated through Stochastic Search Variable Selection algorithm.

Dellaportas et. al. (1997) developed Gibbs Variable selection for selecting the independent variables for the linear regression model. George and Mc Culloch (1997) described a variety of approaches to Bayesian Variable Selection for hierarchical mixture models. The priors in these models describe the uncertainty present in variable selection. Chipman, Hamada and WU (1997) made an attempt on the selection of variables in complex aliasing using Bayesian approach and illustrated with four examples, three of which come from actual industrial experiments.

Kuo and Mallick (1998) made an attempt for selecting the suitable predictors in multiple regression using Bayesian approach. George and Foster (2000) proposed and developed two empirical Bayes selection criteria Maximum Marginal Likelihood (MML) and Conditional Marginal Likelihood (CML) and showed that these criteria select the maximum posterior models under implicit hyper-parameter choices for a particular hierarchical Bayes formulation when compared with AIC and BIC.

Wang and George (2004) developed and evaluated a new selection criteria based on Empirical Bayes and fully Bayes for Generalized Linear Models. They also introduced a general hierarchical mixture Bayesian setup for the variable selection problem.

Park and Casella (2008) developed a fully Bayesian hierarchical model and an efficient Gibbs sampler for the lasso problem.

O'Hara and Sillanppa (2009) had given review on different Bayesian variable selection methods like Kuo and Mallick, Gibbs Variable Selection (GVS), Stochastic Search Variable Selection (SSVS), Adaptive Shrikage with Jeffrey's prior or Laplacian prior and Reversible Jump MCMC with real and simulated data examples using BUGS package.

Lindsey Charles and Sheather Simon (2010) presented a new Stata program called v-select that performs variable selection after fitting a linear regression. Also they demonstrated the use of each method of variable selection: Forward Selection and Backward Elimination and best subset selection with variety of datasets. Chen and Wang (2010) proposed an application of MCMC method to the Bayesian variable selection problem for Gaussian process regression model and is applied to the chemometric calibration of near infrared (NIR) spectroscopic data.

Weinwurm Stephan et. al. (2013) conducted a detailed comparison on Bayesian Penalized Regression methods: Bayesian Lasso (BLA), Bayesian Ridge Regression (BRR) with Stochastic Search Variable Selection (SSVS) and hybrid Correlation Based Search (hCBS) on three simulated datasets. Each method has capability to predict phenotypes based on the selected Single-Nucleotide Polymorphisms (SNPs) and their computational demands are studied.

Otava Martin et. al. (2014) focused on Bayesian Variable Selection (BVS) models for order-restricted one-way ANOVA models for dose-response data that offers a framework to establish a inference and estimation simultaneously. BVS is similar to Gibbs variable selection. Elangovan and Lokeshmaran (2014) discussed the Bayesian variable selection for Cox's regression model and showed that this method works effectively by taking a real data example.

Xu and Ghosh (2015) proposed a Bayesian group lasso model with spike and slab priors by selecting the variables both at the group level and within a group also.

3.4 METHODS FOR THE SELECTION OF VARIABLES IN BAYESIAN APPROACH

In this section, an attempt is made to present some methods used to reduce the size of the model by selecting the significant variables for a multiple linear regression model using Bayesian approach.

3.4.1 Stochastic Search Variable Selection Method: It was proposed by George and McCulloch (1993), to reduce the size of the model by selecting the significant variables in the model by introducing a latent indicator vector that indicates predictors are included in and excluded from the current model.

Let us consider $\Theta_j = \gamma_j \theta_j$; where γ_j represents the indicator variable, that implies $P(\gamma_j,\theta_j) = P(\theta_j / \gamma_j)\, P(\gamma_j)$. Here Mixture prior for θ is used: $P(\theta_j / \gamma_j) = (1-\gamma_j)\, N(0, \sigma^2) + \gamma_j\, N(0, c\sigma^2)$. It proceeds by using Gibbs sampling to sample from the set of possible subset choices. From these subsets with higher probability can be identified by their more frequent appearance in the Gibbs sample.

Note: Unlike in Gibbs Variable Selection values of the prior parameters when $\gamma_j=0$ influence the posterior. In this, $P(\theta_j / \gamma_j=0)$ needs to be very small but at the same time not too restricted around zero. The advantage of Stochastic Search Variable Selection is that it avoids the overwhelming problem of calculating the posterior probabilities of all 2^M subsets.

EXAMPLE 3.4.1: Consider the experimental data presented in the example 2.3.1. Consider the model to be fitted is $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$. The estimated values for the parameters are: $\hat{\beta}_0 = 50.520$; $\hat{\beta}_1 = 5.372$; $\hat{\beta}_2 = 10.130$; $\hat{\beta}_3 = 1.091$. The variance-covariance matrix $(X'X)$ and its inverse $(X'X)^{-1}$) are:

$$\begin{bmatrix} 12 & 0.6 & 0.25 & 0.007 \\ 0.6 & 8.18 & 0.31 & 0.14575 \\ 0.25 & 0.31 & 8.5375 & -0.07065 \\ 0.007 & 0.14575 & -0.07065 & 8.583289 \end{bmatrix}; \begin{bmatrix} 0.08368 & -0.00605 & -0.00223 & 0.00002 \\ -0.00605 & 0.12289 & -0.00430 & -0.00212 \\ -0.00223 & -0.00430 & 0.11736 & 0.00104 \\ 0.00002 & -0.00212 & 0.00104 & 0.11655 \end{bmatrix}$$

Corresponding to each regression parameter the posterior Mean, Posterior Standard deviation and its probabilities (%) are presented in the Table 3.4.1.

Table 3.4.1

S.No.	Regression Parameters	Posterior Mean	Posterior S.D	Posterior Probabilities (%)
1	β_0	50.5197	1.0283	100
2	β_1	5.3846	1.2460	100
3	β_2	10.1232	1.2177	100
4	β_3	0.3685	0.8782	33.8

Hence the reduced model with two variables is: $Y = 50.520 + 5.391 X_1 + 10.12 X_2$ with R^2 value 0.910 selected with highest posterior probability value 0.662.

3.4.2 GIBBS Variable Selection Method: It was proposed by Dellaportas et. al. (1997). Consider $\Theta_j = \gamma_j \theta_j$; where γ_j represents the indicator variable, that implies $P(\gamma_j, \theta_j) = P(\theta_j / \gamma_j) P(\gamma_j)$. Since prior distributions of Indicator and effects are assumed to be dependent on each other. In GVS, we consider pseudo – prior for $\gamma_j=0$ otherwise it will be same as Kuo and

Mallick. Here Mixture prior is assumed for θ_j is $P(\theta_j / \gamma_j) = (1-\gamma_j)\, N(\hat{\mu}, \sigma^2) + \gamma_j\, N(0, \tau^2)$ When $\gamma_j = 0$ if $P(\theta_j / \gamma_j = 1)$ is large then the corresponding variable of θ_j is included in the model.

EXAMPLE 3.4.2: Consider the experimental data presented in the example 2.3.1. The estimated parameters are: $\hat{\beta}_0 = 50.520$; $\hat{\beta}_1 = 5.372$; $\hat{\beta}_2 = 10.130$; $\hat{\beta}_3 = 1.091$. Corresponding to X_1 and X_2 variables the model attains high posterior probability by considering the pseudo prior for indicator variable. We observe that this method also gives same as method 3.4.1.

3.4.3 Kuo and Mallick Variable Selection Method: Kuo and Mallick (1998) made an attempt to reduce the size of the regression model using indicator variables and conclude that if the estimated parameter has high posterior probability then the corresponding variable is included in the model. Let $\Theta_j = \gamma_j \theta_j$; where γ_j represents the indicator variable, that implies $P(\gamma_j, \theta_j) = P(\gamma_j)\, P(\theta_j)$. Since effects and Indicator variables are independent. In particular, If $\gamma_j = 0$ then θ_j is sampled from the full conditional distribution which is taken as its prior. By taking this prior distribution, if θ_j has high posterior distribution then the corresponding variable of θ_j is included in the model.

EXAMPLE 3.4.3: Consider the experimental data presented in the example 2.3.1. The estimated parameters are: $\hat{\beta}_0 = 50.520$; $\hat{\beta}_1 = 5.372$; $\hat{\beta}_2 = 10.130$; $\hat{\beta}_3 = 1.091$. In this method, we define the full conditional distribution which is taken as prior distribution. This method can also be observed that resulting to same model with high posterior probability value 0.662 containing the variables X_1 and X_2 variables with same R^2 value.

3.5 WINBUG SOFTWARE USED FOR POSTERIOR EVALUATION

Win BUG and Open BUG are general-purpose software used for analyzing complex statistical Bayesian models using Markov Chain Monte Carlo (MCMC) methods. It uses the Gibbs sampling algorithm to construct the transition kernel for its Markov chain samplers. Each iteration of Gibbs sampler involves drawing a new value for each parameter from its full conditional distribution. This software is used to generate a posterior sample and to estimate the parameters of posterior distributions by evaluating the Normalized constant. To evaluate the same programming code is written by specifying observed sample, the distribution of observed sample and its estimated parameters based on the sample, prior distribution and its estimated values from the observed sample.

4. REDUCTION IN DIMENSIONALITY OF RESPONSE SURFACE DESIGN MODEL USING VARIANCE COMPONENT INDICES

4.1 INTRODUCTION

Variance component indices are the global sensitivity indices. These indices can be evaluated by decomposing variance based on the input-output relationship between the response and input variables. From the model, this index measures the contribution of individual components variance to the variance of the output variable. In this chapter an attempt is made to derive the variance component indices and the proportion of the variances of estimated parameters and the response in case of first order and second order response surface design model are presented with suitable examples.

4.2 VARIANCE COMPONENT INDICES FOR FIRST ORDER RSD MODEL

The variance component indices for first order response surface design model under without imposing restrictions on $X'X$ and imposing restrictions on $X'X$ towards orthogonality are presented in sections 4.2.1 and 4.2.2 respectively

4.2.1 Without imposing restrictions on Moment matrix: Consider the following first order response surface design model in v factors

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots\ldots + \beta_v X_v + \boldsymbol{\varepsilon} \tag{4.2.1}$$

with $x_u = (1, x_{u1}, x_{u2} \ldots x_{uv})$ be the u^{th} row of X and $\beta = (\beta_0, \beta_1, \beta_2 \ldots \beta_v)$ be the vector of parameters. The estimated response at u^{th} design point is

$$\hat{Y}_u = \hat{\beta}_0 + \hat{\beta}_1 x_{u1} + \hat{\beta}_2 x_{u2} + \ldots\ldots + \hat{\beta}_v x_{uv} \tag{4.2.2}$$

The variance-covariance matrix of $\hat{\beta}$ is

$$V(\hat{\beta}) = (X'X)^{-1}\sigma^2 \tag{4.2.3}$$

where, the X'X matrix is

$$X'X = \left[\begin{array}{c|cccc} N & \sum_{u=1}^{N} x_{u1} & \cdots & \sum_{u=1}^{N} x_{uv} \\ \hline \sum_{u=1}^{N} x_{u1} & \sum_{u=1}^{N} x_{u1}^2 & \cdots & \sum_{u=1}^{N} x_{u1}x_{uv} \\ \cdots & \cdots & \cdots & \cdots \\ \sum_{u=1}^{N} x_{uv} & \sum_{u=1}^{N} x_{u1}x_{uv} & \cdots & \sum_{u=1}^{N} x_{uv}^2 \end{array}\right] \tag{4.2.4}$$

The variance of the predicted response at the u^{th} point is

$$V(\hat{Y}_u) = V(\hat{\beta}_0) + \sum_{i=1}^{v} x_{ui}^2 V(\hat{\beta}_i) + \sum_{i=1}^{v}\sum_{j=1}^{v} x_{ui}x_{uj} \operatorname{Cov}(\hat{\beta}_i, \hat{\beta}_j) \tag{4.2.5}$$

Then the variance component indices S_i' s corresponding to each of the component are

$$S_0 = V(\hat{\beta}_0) / V(\hat{Y}_u)$$

$$S_i = [\, x_{ui}^2 . V(\hat{\beta}_i) \,] / V(\hat{Y}_u) \tag{4.2.6}$$

$$S_{ij} = [\, x_{ui}\, x_{uj} \operatorname{Cov}(\hat{\beta}_i, \hat{\beta}_j) \,] / V(\hat{Y}_u)$$

Note:

1. The variance component indices can be expressed as product of function of the level of the factor and ratio index, where the ratio index is the variance ratio of the parameter to the estimated response.
2. The indices are depends on the design point chosen and its factor level which is different from one design point to another. So the insignificance of component is depends on the design point chosen.

3. **3.** Even though the variance component indices in (4.2.6) satisfy $S_0 + \sum_{i=1}^{v} S_i + \sum_{i=1}^{v}\sum_{j=1}^{v} S_{ij} = 1$ the indices are not theoretically in compressed form due to $\sum_{u=1}^{N} x_{ui}$, $\sum_{u=1}^{N} x_{ui}x_{uj}$ are theoretically not in compressed form.
4. **4.** Even though not imposing restrictions on X'X, But it is possible to evaluate the variance component indices for numerical experimental data.
5. **5.** If the variance ratio index of parameter and estimated response at u^{th} point is insignificant and also insignificant when it is multiplied with function of the factor level(s) then the dimensionality of the model can be reduced at those design points only.

   It can be illustrated through the example 4.2.1.

EXAMPLE 4.2.1: Let 120, 136, 137, 140, 141, 143, 161, 170, 213, 215, 269, 271, 313, 345, 381, 394, 398, 402 are the responses obtained through a first order response surface design model with four factors at the respective design points (0,1,-1,0), (1,-1,-1,1), (1,1,-1,1), (-1,1,-1,-1), (-1,-1,1,1,), (-1,-1,-1,1), (0,1,0,-1), (1,-1,-1,-1), (-1,1,-1,1), (-1,1,1,-1), (-1,-1,1,-1), (-1,-1,-1,-1), (1,1,0,0), (1,1,1,-1),(-1,1,-1,-1),(1,-1,0,-1),(-1,0,-1,-1), (-1,1,-1,1). There are no restrictions that are imposed the design matrix. Then the estimated values for parameters are $\hat{\beta}_0$ =328.3; $\hat{\beta}_1$ = -7.365; $\hat{\beta}_2$ =2.97; $\hat{\beta}_3$ =2.57; and $\hat{\beta}_4$ = -51.07 and the variance of estimated response is 11202.61. The variances of the parameters, variance of the estimated response at a design point (x_{u1}, x_{u2}, …, x_{uv}) and corresponding variance component indices in the model (3.2.1) at design point (x_{u1}, x_{u2}, …, x_{uv}) can be evaluated using (4.2.5) and (4.2.6), which are depends on design point chosen.

4.2.2 With imposing restrictions on Moment matrix: Consider the model (5.2.1) to be fitted is first order response surface design model in v factors. Suppose the restrictions are imposed on X'X given in (3.4) towards reaching to orthogonality, as $\Sigma\, x_{ui} = \Sigma\, x_{ui}x_{uj} = 0$ and let $\Sigma\, x^2_{ui} = N\lambda_2$ (the summation over u=1, …N and $i \neq j = 1,2\ldots v$). Then, the variance – covariance matrix X'X can be obtained as

$$X'X = \begin{bmatrix} N & 0 \\ 0 & N\lambda_2 I \end{bmatrix} \quad (4.2.7)$$

The variance of the estimated response at the u^{th} design point is

$$V(\hat{Y}_u) = V(\hat{\beta}_0) + \sum_{i=1}^{v} x_{ui}^2 V(\hat{\beta}_i) \quad (4.2.8)$$

Where, $V(\hat{\beta}_0) = \sigma^2/N$; $V(\hat{\beta}_i) = \sigma^2/N\lambda_2$ for i=1,2, ... v.

Then, the variance component indices using the model (4.2.8) can be obtained using the equations

$S_0 = \dfrac{V(\hat{\beta}_0)}{V(\hat{Y}_u)}$ and $S_i = \dfrac{V(\hat{\beta}_i)}{V(\hat{Y}_u)} x_{ui}^2$ as

$$S_0 = N^{-1} \quad \text{and} \quad S_i = (N\lambda_2)^{-1} x_{ui}^2 \text{ for } i = 1, 2, \ldots, v. \quad (4.2.9)$$

Note:

1. The variance component indices are in compressed form and satisfying $S_0 + \sum_{i=1}^{v} S_i = 1$.
2. The variance ratio index of parameter, estimated response at u^{th} point are same for all the factors.
3. Even though the indices are in compressed form, depends on the factor level(s) which is different from one design point to another. So the insignificance of component is depends on the design point chosen.

   It is illustrated in the example 4.2.2.

EXAMPLE 4.2.2: Let 27.6, 16.6, 15.4, 17.4, 17.0, 19.0, 17.4, 12.6, 18.6, 22.4, 21.4, 14.0, 24.0, 15.6, 13.0, 14.4 are the responses obtained through a first order response surface design model with four factors at the respective design points (1, 1, 1, 1), (-1, -1, 1, 1), (1, -1, -1, 1), (-1, 1, -1, 1), (1,-1,1,-1), (-1,1,1,-1), (1,1,-1,-1), (-1,-1,-1,-1), (1,-1,1,1), (-1,1,1,1), (1,1,-1,1), (-1, -1, -1, 1), (1, 1, 1, -1), (-1, -1, 1, -1), (1, -1, -1, -1), (-1, 1, -1, -1). In this design, $\Sigma x_{ui} = \Sigma x_{ui}x_{uj}=0$ and let $\Sigma x^2_{ui} = N\lambda_2$. The estimated values of parameters are $\hat{\beta}_0 = 17.9$, $\hat{\beta}_1 = 64$, $\hat{\beta}_2 = 2.55$, $\hat{\beta}_3 = 2.2$ and $\hat{\beta}_4 = 1.275$. The variance of the response is

17.30. The variances of the parameters, variance of the estimated response at a design point (x_{u1}, x_{u2}, …, x_{uv}) and corresponding variance component indices in the model (4.2.2) at design point (x_{u1}, x_{u2}, …, x_{uv}) can be evaluated using (4.2.5) and (4.2.6), which are depends on design point chosen.

EXAMPLE 4.2.3: Consider the experimental data presented in the example 2.3.5. The estimated parameters are: $\hat{\beta}_0$ = 50.520; $\hat{\beta}_1$ = 5.372; $\hat{\beta}_2$ = 10.130; $\hat{\beta}_3$ = 1.091. Then the sobol indices are: S_0 = 0.16345; S_1 = 0.37505; S_2 = 0.30315; S_3 = 0.22764. We observe that each index gives same as method 2.3.5.

Note:

1. The reduction of size of the model depends on the design points.
2. It is difficult to reduce the size of the model if the design is orthogonal in case of first order model and rotatable in case of second order model.

4.3 VARIANCE COMPONENT INDICES FOR SECOND ORDER RSD MODEL

The variance component indices for second order response surface design model under without imposing restrictions on $X'X$ and imposing restrictions on $X'X$ towards orthogonality are presented in sections 4.3.1 and 4.3.2 respectively.

4.3.1 Without imposing restrictions on Moment matrix: Consider the second-order response surface design model in v factors at the uth design point as

$$Y_u = \beta_0 + \sum_{i=1}^{v} \beta_i x_{ui} + \sum_{i=1}^{v} \beta_{ii} x_{ui}^2 + \sum_{i<j}^{v} \beta_{ij} x_{ui} x_{uj} + \boldsymbol{\varepsilon} \qquad (4.3.1)$$

Let $X_u = (1, x_{u1}, x_{u2} \dots x_{uv}, x^2_{u1}, x^2_{u2} \dots x^2_{uv}, x_{ul}x_{u2} \dots x_{uv-1}x_{uv})$ is the uth row of X,

$\underline{\beta} = (\beta_0, \beta_1, \beta_2 \dots \beta_v, \beta_{11}, \beta_{22} \dots \beta_{vv}, \beta_{l2} \dots \beta_{v-1v})'$ is the vector of parameters.

Then estimated response at the uth design point is

$$\hat{Y}_u = \hat{\beta}_0 + \sum_{i=1}^{v} \hat{\beta}_i x_{ui} + \sum_{i=1}^{v} \hat{\beta}_{ii} x_{ui}^2 + \sum_{i<j}^{v} \hat{\beta}_{ij} x_{ui} x_{uj} \qquad (4.3.2)$$

The variance-covariance matrix of $\hat{\beta}$ is $V(\hat{\beta}) = (X'X)^{-1}\sigma^2$, where $X'X$ is

$$\left[\begin{array}{c|ccc|ccc|ccc}
N & \sum_{i=1}^{N} x_{il} & \cdots & \sum_{i=1}^{N} x_{iv} & \sum_{i=1}^{N} x_{i1}^2 & \cdots & \sum_{i=1}^{N} x_{iv}^2 & \sum_{i=1}^{N} x_{i1}x_{i2} & \cdots & \sum_{i=1}^{N} x_{iv-1}x_{iv} \\
\hline
\sum_{i=1}^{N} x_{il} & \sum_{i=1}^{N} x_{i1}^2 & \cdots & \sum_{i=1}^{N} x_{i1}x_{iv} & \sum_{i=1}^{N} x_{i1}^3 & \cdots & \sum_{i=1}^{N} x_{i1}x_{iv}^2 & \sum_{i=1}^{N} x_{i1}^2x_{i2} & \cdots & \sum_{i=1}^{N} x_{i1}x_{iv-1}x_{iv} \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\sum_{i=1}^{N} x_{iv} & \sum_{i=1}^{N} x_{i1}x_{iv} & \cdots & \sum_{i=1}^{N} x_{iv}^2 & \sum_{i=1}^{N} x_{i1}^2x_{iv} & \cdots & \sum_{i=1}^{N} x_{iv}^3 & \sum_{i=1}^{N} x_{i1}x_{i2}x_{iv} & \cdots & \sum_{i=1}^{N} x_{iv-1}x_{iv}^2 \\
\hline
\sum_{i=1}^{N} x_{i1}^2 & \sum_{i=1}^{N} x_{i1}^3 & \cdots & \sum_{i=1}^{N} x_{i1}^2x_{iv} & \sum_{i=1}^{N} x_{i1}^4 & \cdots & \sum_{i=1}^{N} x_{i1}^2x_{iv}^2 & \sum_{i=1}^{N} x_{i1}^3x_{i2} & \cdots & \sum_{i=1}^{N} x_{i1}^2x_{iv-1}x_{iv} \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\sum_{i=1}^{N} x_{iv}^2 & \sum_{i=1}^{N} x_{i1}x_{iv}^2 & \cdots & \sum_{i=1}^{N} x_{iv}^3 & \sum_{i=1}^{N} x_{i1}^2x_{iv}^2 & \cdots & \sum_{i=1}^{N} x_{iv}^4 & \sum_{i=1}^{N} x_{i1}x_{i2}x_{iv}^2 & \cdots & \sum_{i=1}^{N} x_{iv-1}x_{iv}^3 \\
\hline
\sum_{i=1}^{N} x_{i1}x_{i2} & \sum_{i=1}^{N} x_{i1}^2x_{i2} & \cdots & \sum_{i=1}^{N} x_{i1}x_{i2}x_{iv} & \sum_{i=1}^{N} x_{i1}^3x_{i2} & \cdots & \sum_{i=1}^{N} x_{i1}x_{i2}x_{iv}^2 & \sum_{i=1}^{N} x_{i1}^2x_{i2}^2 & \cdots & \sum_{i=1}^{N} x_{i1}x_{i2}x_{iv-1}x_{iv} \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\sum_{i=1}^{N} x_{iv-1}x_{iv} & \sum_{i=1}^{N} x_{i1}x_{iv-1}x_{iv} & \cdots & \sum_{i=1}^{N} x_{iv-1}x_{iv}^2 & \sum_{i=1}^{N} x_{i1}^2x_{iv-1}x_{iv} & \cdots & \sum_{i=1}^{N} x_{iv-1}x_{iv}^3 & \sum_{i=1}^{N} x_{i1}x_{i2}x_{iv-1}x_{iv} & \cdots & \sum_{i=1}^{N} x_{iv-1}^2x_{iv}^2
\end{array}\right]$$

(4.3.3)

The $V(\hat{Y}_u)$ is not in a simplified form as the elements in moment matrix are in higher order. Then the estimated response at uth design point is $\hat{Y}_u = X(X'X)^{-1}X'Y$ and variance covariance matrix of $\hat{Y}_u$ is $V(\hat{Y}_u) = (X'X)^{-1}\sigma^2$ not in compressed form. Let S_0 S_i, S_{ii} and S_{ij} are the variance-component indices of the model (4.3.2)

$$\left.\begin{array}{l} S_0 = V(\hat{\beta}_0) / V(\hat{Y}_u); \\ S_i = [x_{ui}^2 . V(\hat{\beta}_i)] / V(\hat{Y}_u) \end{array}\right\} \qquad (4.3.4)$$

$S_{ii} = [x_{ui}^4 . V(\hat{\beta}_{ii})] \ / \ V(\hat{Y}_u)$

$S_{ij} = [x_{ui}^2 \, x_{uj}^2 \, Cov(\hat{\beta}_i, \hat{\beta}_j)] \ / \ V(\hat{Y}_u)$

Note:

1. The variance components indices can be expressed in (4.3.4) are product of function of the level of the factor and ratio index, where the ratio index is the variance ratio of the parameter to the estimated response.
2. The variance component indices are depends on the design point chosen and its factor level which is different from one design point to another. So the insignificance of component is depends on the design point chosen.
3. Even though the variance component indices in (4.3.2) satisfy $S_0 + \sum_{i=1}^{v} S_i + \sum_{i=1}^{v} S_{ii} + \sum_{i}^{v} \sum_{j=1}^{v} S_{ij}$ = 1 the indices are not theoretically in compressed form due to X′X is not in compressed form depends on design points.
4. Even though not imposing restrictions on X'X, But it is possible to evaluate the variance component indices for numerical experimental data.
5. If the variance ratio index of parameter and estimated response at u^{th} point is insignificant and also insignificant when it is multiplied with function of the factor level(s) then the dimensionality of the model can be reduced at those design points only.

   It can be illustrated through the example 4.3.1.

EXAMPLE 4.3.1: Let 2.83, 3.25, 3.56, 2.53, 3.01, 3.19, 2.23, 2.65, 3.06, 2.57, 3.08, 3.50, 2.42, 2.79, 3.03, 2.07, 2.85, 3.12 are the responses obtained through a second order response surface model with three factors at the design points (-1, -1, -1), (0, -1, -1), (1,-1, -1), (-1, 0, -1), (0, 0, 1), (1, 0, 1), (-1, 1, 1), (0, 1, 1), (1, -1, 0), (-1, -1, 0), (0, -1, -1), (1, 0, -1), (-1, 0, 1), (0, 0, -1), (1, 1, 0), (-1, 1, 1), (-1, 1, 1), (-1, 1, -1). Then estimated values of the parameters are $\hat{\beta}_0$ =51.74, $\hat{\beta}_1$ =0.44, $\hat{\beta}_2$ = -0.209, $\hat{\beta}_3$ = -0.038;

$\hat{\beta}_{11} = 0.023$, $\hat{\beta}_{22} = 0.539$, $\hat{\beta}_{33} = 0.840$, $\hat{\beta}_{12} = 0.170$, $\hat{\beta}_{13} = -0.066$ and $\hat{\beta}_{23} = -0.513$. The variance of the estimated response is 0.165. The variances of the parameters, variance of the estimated response at a design point (x_{u1}, x_{u2}, …, x_{uv}) and corresponding variance component indices in the model (4.3.2) at design point (x_{u1}, x_{u2}, …, x_{uv}) can be evaluated using (4.3.3) and (4.3.4), which are depends on design point chosen. These are used for ranking the parameters β_i s, to reduce the model.

4.3.2. With imposing restrictions on Moment matrix: Consider the second-order response surface design model in v factors at the u^{th} design point given in (4.3.1). Suppose the restrictions are imposed on the moment matrix X'X towards reaching to orthogonality for second order model, as $\Sigma x_{ui}^{\delta i} x_{uj}^{\delta j} x_{uk}^{\delta k} x_{ul}^{\delta l} = 0$; $i \neq j \neq k \neq l = 1, 2, \dots v$ for any δvalue is odd and $\Sigma x^2_{ui} = N\lambda_2$; $\Sigma x^4_{ui} = CN\lambda_4$; $\Sigma x^2_{ui} x^2_{uj} = N\lambda_4$ where the summation is over all the design points, and let $\Delta = \lambda_4(C+v-2)-v\lambda_2^2 > 0$. Then the moment matrix X'X and $(X'X)^{-1}$ can be obtained as

$$X'X = \left[\begin{array}{c|ccc} 1 & 0 & \lambda_2 J & 0 \\ \hline 0 & \lambda_2 I & 0 & 0 \\ \lambda_2 J & 0 & [(c-1)I+J]\lambda_4 & 0 \\ 0 & 0 & 0 & \lambda_4 I \end{array}\right]; \quad (X'X)^{-1} = \left[\begin{array}{c|ccc} \lambda_4(c+k+1)\Delta^{-1} & 0 & (c+k-1)(c-1)J\Delta^{-1} & 0 \\ \hline 0 & \lambda_2^{-1} I & 0 & 0 \\ -2\lambda_2 J\Delta^{-1} & 0 & Z & 0 \\ 0 & 0 & 0 & \lambda_2^{-1} \end{array}\right]$$

where $\Delta = \lambda_4 (c+k-1) - k\lambda_2^2 > 0$ and $Z_{kxk} = \dfrac{[(c+v-1)I_v - J_{v,v}]}{\lambda_4(c-1)(c+v-1)} + \dfrac{[\lambda_2^2(c+v-1)(c-1)^2]}{[\lambda_4(c+v-1)-k\lambda_2^2]} J_{k,k}$

variances of the estimated parameters can be obtained as,

$$V(\hat{\beta}_0) = [\lambda_4 (c+k-1)/ N\Delta]\sigma^2$$

$$V(\hat{\beta}_i) = (1/ N\lambda_2)\sigma^2 ;$$

$$V(\hat{\beta}_{ij}) = (1/ N\lambda_4)\ \sigma^2$$

$$Cov(\hat{\beta}_0, \hat{\beta}_{jj}) = [-\lambda_2 / N\Delta]\ \sigma^2$$

$$V(\hat{\beta}_{ii}) = [\,\{\lambda_4(c+k-2)-(k-1)\lambda_2^2\} / \{N\lambda_4 (c-1)\,\Delta\}\,]\,\sigma^2$$

$$Cov(\hat{\beta}_{ii}, \hat{\beta}_{jj}) = [\,(\lambda_2^2 - \lambda_4) / \{(c-1)\,N\lambda_4\,\Delta\}\,]\,\sigma^2$$

and other covariance's are vanishes.

Let $\sum_{i=1}^{v} x_{ui}^2=\rho^2$ then $\sum_{i=1}^{v} x_{ui}^4=\rho^4-2\sum_{i<j}^{v} x_{ui}^2 x_{ui}^2$. Then variance of estimated response is

$$V(\hat{Y}_u) = V(\hat{\beta}_0) + [\,V(\hat{\beta}_i) + 2\,Cov(\hat{\beta}_0, \hat{\beta}_{ii})]\,\rho^2 + V(\hat{\beta}_{ii})\,\rho^4 +$$

$$[\,V(\hat{\beta}_{ij}) - 2V(\hat{\beta}_{ii}) + 2\,Cov(\hat{\beta}_{ii}, \hat{\beta}_{jj})]\sum_{i<j}^{v} x_{ui}^2 x_{uj}^2$$

The resulting variance of estimated response from the above equations is

$$V(\hat{Y}_u) = \frac{\sigma^2}{N\Delta}\left[\left\{\frac{\Delta-2\lambda_2^2}{\lambda_2}\right\}\rho^2 + \left\{\frac{\Delta-\lambda_2^2}{\lambda_4(c-1)}\right\}\rho^4 - \{\Delta+v\lambda_2^2\}\right] + \left[\frac{(c-3)}{(c-1)N\lambda_4}\sigma^2\right]\sum_{i}^{v} x_i^2 x_j^2$$

The variance-component indices consider thus interaction among the input variables.

$$S_0 = \left[\frac{\lambda_4(c+k-1)/\Delta}{V(\hat{Y}_u)}\right] = \left[\frac{\lambda_4(c+k-1)/\Delta}{\frac{1}{\Delta}\left[\frac{\Delta-\lambda_2^2}{\lambda_4(c-1)}\rho^4 + \frac{\Delta-2\lambda_2^2}{\lambda_2}\rho^2 + (\Delta+k\lambda_2^2) + \frac{(c-3)}{(c-1)}\frac{\Delta}{\lambda_4}\sum_{i=1}^{k} x_i^2 x_j^2\right]}\right]$$

$$S_i = \left[\frac{1/\lambda_2}{V(\hat{Y}_u)}\right] = \left[\frac{1/\lambda_2}{\frac{1}{\Delta}\left[\frac{\Delta-\lambda_2^2}{\lambda_4(c-1)}\rho^4 + \frac{\Delta-2\lambda_2^2}{\lambda_2}\rho^2 + (\Delta+k\lambda_2^2) + \frac{(c-3)}{(c-1)}\frac{\Delta}{\lambda_4}\sum_{i=1}^{k} x_i^2 x_j^2\right]}\right]$$

$$S_{ij} = \left[\frac{1/\lambda_4}{V(\hat{Y}_u)}\right] = \left[\frac{1/\lambda_4}{\frac{1}{\Delta}\left[\frac{\Delta-\lambda_2^2}{\lambda_4(c-1)}\rho^4 + \frac{\Delta-2\lambda_2^2}{\lambda_2}\rho^2 + (\Delta+k\lambda_2^2) + \frac{(c-3)}{(c-1)}\frac{\Delta}{\lambda_4}\sum_{i=1}^{k} x_i^2 x_j^2\right]}\right]$$

$$S_{ii} = \left[\frac{\dfrac{\lambda_4(c+k-2)-(k-1)\lambda_2^2}{\lambda_4(c-1)[\lambda_4(c+k-1)-k\lambda_2^2]}}{V(\hat{Y}_u)} \right] = \left[\frac{\dfrac{\lambda_4(c+k-2)-(k-1)\lambda_2^2}{\lambda_4(c-1)[\lambda_4(c+k-1)-k\lambda_2^2]}}{\dfrac{1}{\Delta}\left[\dfrac{\Delta-\lambda_2^2}{\lambda_4(c-1)}\rho^4 + \dfrac{\Delta-2\lambda_2^2}{\lambda_2}\rho^2 + (\Delta+k\lambda_2^2) + \dfrac{(c-3)}{(c-1)}\dfrac{\Delta}{\lambda_4}\sum_{i=1}^{k} x_i^2 x_j^2 \right]} \right]$$

Note:

1. The variance component indices satisfying $S_0 + \sum_{i=1}^{v} S_i + \sum_{i=1}^{v} S_{ii} + \sum_{i}^{v} \sum_{j=1}^{v} S_{ij} = 1$.
2. The variance component indices are depends on the distance between the design point to the origin.
3. Even though the variance-component indices are in compressed form, depends on the factor level(s) which is different from one design point to another (design points are hided in the indices). So the insignificance of component is depends on the design point chosen.
4. The variance component indices are in the ratio

$$S_0 : S_i : S_{ii} : S_{ij} = \frac{\lambda_4(c+v-1)}{[\lambda_4(c+v-1)-v\lambda_2^2]} : \frac{1}{\lambda_2} : \frac{1}{\lambda_4} : \frac{\lambda_4(c+v-2)-(v-1)\lambda_2^2}{\lambda_4(c-1)[\lambda_4(c+v-1)-v\lambda_2^2]}$$

5. When S_i and S_{ii} are compared, they are in the ratio $\dfrac{1}{\sum_{u=1}^{N} x_{ui}^2} : \dfrac{c}{\sum_{u=1}^{N} x_{ui}^4}$. In general, for any design matrix, $\sum x_{ui}^4 \geq \sum x_{ui}^2$ therefore if $\dfrac{1}{c}\sum_{u=1}^{N} x_{ui}^4 \geq \sum_{u=1}^{N} x_{ui}^2$ then, $S_i > S_{ii}$ and if $\dfrac{1}{c}\sum_{u=1}^{N} x_{ui}^4 = \sum_{u=1}^{N} x_{ui}^2$ then, $S_i > S_{ii}$ and $\sum_{i=1}^{v} s_i = \sum_{i=1}^{v} s_{ii}$
6. The indices S_i and S_{ij} are in the ratio, $S_i : S_{ij} = \dfrac{1}{\lambda_2} : \dfrac{1}{\lambda_4} \times \dfrac{1}{(c-1)} \times (1-\delta)$ where $\delta = \dfrac{(\lambda_4-\lambda_2^2)}{[\lambda_4(c+v-1)-v\lambda_2^2]}$. It can be noted that, $S_i > S_{ij}$ for all $c>1$ and δ is very small.

7. The indices S_{ii} and S_{ij} are in the ratio $\frac{1}{(c-1)}\times(1-\delta)$: 1. When c > 1, $S_{ii} < S_{ij}$

8. The indices S_0 and S_i are in the ratio $\frac{\lambda_4(c+v-1)}{\Delta}:\frac{1}{\lambda_2}$. The comparison exists for non-standardized model and valid only if $\lambda_4(c+v-1) > v\lambda_2^2$. If $\lambda_2^2 \le \lambda_4$ and $c \ge 1$ then, $S_0 < S_i$ for all $\lambda_2 \ge 1$. If $\lambda_2^2 \le \lambda_4$ and c = 1 then, $S_0 > S_i$ for all $\lambda_2 \ge 1$. If $\lambda_2^2 = \lambda_4$ and c = 1 then, S_0 is highly significant for all $\lambda_2 \ge 1$

9. The indices S_0 and S_{ii} are in the ratio $\frac{\lambda_4(c+v-1)}{\Delta}:\frac{\lambda_4(c+v-2)-(v-1)\lambda_2^2}{\Delta\cdot\lambda_4(c-1)}$

10. The indices S_0 and S_{ij} are in the ratio $\frac{\lambda_4(c+v-1)}{\Delta}:\frac{1}{\lambda_4}=\frac{\lambda_4(c+v-1)}{\lambda_4(c+v-1)-v\lambda_2^2}:\frac{1}{\lambda_4}$. It is noted that $\frac{\lambda_4(c+v-1)}{\lambda_4(c+v-1)-v\lambda_2^2} > 1$ for $\forall\ \lambda_4 > 1$. Therefore $S_0 > S_{ij}$ $\forall\ \lambda_4 > 1$.

EXAMPLE 4.3.2: Let 63.03, 62.19, 64.01, 61.60, 58.95, 78.34, 45.75, 72.66, 46.36, 68.62, 35.16, 59.24, 71.62, 84.01, 61.18, 77.78, 61.15, 74.83, 52.45, 65.72 are the responses obtained through a second order response surface design model with four factors at the respective design points :(0, 0, 0,0), (0, 0, 0,0) (0, 0, 0,0) (0, 0, 0,0), (-1,-1,-1,-1), (1,-1,-1,-1), (-1, 1,-1,-1), (1,1,-1,-1), (-1,-1,1,-1), (1,-1,1,-1), (-1, 1, 1, -1), (1,1,1,-1), (-1,-1,-1,1), (1,-1,-1,1), (-1,1,-1,1), ((1,1,-1,1), ((-1,-1,1,1), (1,-1,1,1), (-1,1,1,1), (1,1,1,1).

The estimated values of the parameters are $\hat{\beta}_0$= 62.70, $\hat{\beta}_1$= 9.28, $\hat{\beta}_2$= -4.62, $\hat{\beta}_3$= -5.22; $\hat{\beta}_4$= -5.22, $\hat{\beta}_{22}$= -94.13, $\hat{\beta}_{33}$= -94.13, $\hat{\beta}_{44}$= 27.6, $\hat{\beta}_{12}$ =0.821, $\hat{\beta}_{13}$= -0.12, $\hat{\beta}_{14}$= -2.294, $\hat{\beta}_{11}$= 161.22, $\hat{\beta}_{23}$= -0.17, $\hat{\beta}_{24}$= 0.311, and $\hat{\beta}_{34}$= 0.367. The variance of the estimated response is 144.33. The variances of the parameters, variance of the estimated response at a design point (x_{u1}, x_{u2}, …, x_{uv}) and corresponding variance component indices in the model (4.3.1) at design point (x_{u1}, x_{u2}, …, x_{uv}) can be evaluated, which are depends on design point chosen. These are used for ranking the parameters β_i's to reduce the model.

4.4 COMPARISON IN DIMENSIONALITY OF REDUCED MODELS

Consider the first order response surface design model with the experimental data given by a chemical engineer to investigate the yield of a process on temperature, pressure and catalyst concentration and made an attempt to reduce the size of the model with different approaches proposed by Homma and Saltelli (1996) and in section 4.2. A comparison among the methods with respect to Coefficient of Regression parameters, Confidence Interval, Significance value, R^2 value and Sum of squares due to residual (S_e^2) for the data presented in example 2.3.5 is presented in Table 2.4.1 and found that all the methods gives same reduced models when comparing with the full model.

Table 2.4.1

Parameters	Confidence Interval		Significance Value	Full Model	Homma and Saltelli (1996)	Proposed in Section 4.2
	Lower	Upper				
β_0	48.130	52.908	48.761	50.52	50.52	50.52
β_1	2.476	8.267	4.278	5.372	5.372	5.372
β_2	7.300	12.959	8.256	10.13	10.13	10.13
β_3	-1.729	3.910	0.892	1.091	-	-
SSE				102.621	102.621	102.621
R^2				0.918	0.918	0.918

5. REDUCTION IN DIMENSIONALITY OF RESPONSE SURFACE DESIGN MODEL IN NESTED APPROACH

5.1 INTRODUCTION

Let $\mathbf{Y}_{NX1}$ be the vector of response corresponding to a design matrix X = $((x_{u1}, x_{u2}, \ldots, x_{uv}))$, where x_{ui} be the level of the i^{th} factor in the u^{th} treatment combination, u = 1, 2, … N, i = 1, 2, … p. The functional form of the response surface design model can be expressed as

$$Y = X\underline{\beta} + \underline{\varepsilon} \tag{5.1.1}$$

where $Y_{N\times 1} = (Y_1, Y_2, \ldots, Y_N)'$ is the vector of observations, $X_{N\times p}$ be the Design matrix, $\underline{\beta}_{p\times 1}$ be the vector of parameters and $\underline{\boldsymbol{\varepsilon}}_{N\times 1} = (\varepsilon_1, \varepsilon_2, \ldots \varepsilon_N)'$ be the vector of random errors and assume that $\underline{\boldsymbol{\varepsilon}} \sim N(0,\sigma^2 I)$. The factor-response relationship is given by $E(Y) = f(x_1, x_2, \ldots, x_v)$ is called the 'Response Surface'. Design used for fitting the response surface models are termed as 'Response Surface Design'. The least square estimate of the parameter is $\hat{\beta} = (X'X)^{-1}X'Y$ and its variance–covariance is $V(\hat{\beta}) = (X'X)^{-1}\sigma^2$.

Multi-collinearity is a phenomenon in which two or more predictor variables in a multiple regression model are highly correlated, meaning that one can be linearly predicted from the others with a non-trivial degree of accuracy. It refers to a situation in which two or more explanatory variables in a multiple regression model are highly linearly related.

5.2. REDUCTION OF FIRST ORDER RSD MODEL IN NESTED APPROACH

Consider the linear functional relationship between the response and 'v' factors.

$$Y_u = \beta_0 + \beta_1 x_{u1} + \beta_2 x_{u2} + \quad \dots \quad \dots \quad \dots + \beta_v x_{uv} + \varepsilon_u \tag{5.2.1}$$

where,

Y_u be the u^{th} response at the design point X_u,

$x_u = (1, x_{u1}, x_{u2}, \dots x_{uv})$ be the u^{th} treatment combination of 'v' factors,

$\underline{\beta} = [\beta_0, \beta_1, \beta_2, \dots \beta_v]'$ be the vector of parameters and

ε_u be the random error corresponding to u^{th} response Y_u. Assume $\underline{\boldsymbol{\varepsilon}} \sim N(0, \sigma^2 I)$.

When the number of predictor variables is more, each variable affects the estimates and useful to predict the response in the model. To avoid or alleviate the problem of multi-collinearity, fitting process will play a key role. From the choice of k independent variables, one variable that best fits the objective vector is selected; select the variables sequentially one by one from the original set which have most significant correlation with the estimated error and how much contribution of variable is in the estimate of response and stops the procedure until remained variables in the variable set are not significant for the objective vector. An attempt is made to fit a response surface design model in nested approach.

The detailed step by step procedure is presented below.

Step 1: Let $(Y_1, Y_2, \dots, Y_N)'$ be the vector of N observations, and X_{Nxv} be the design matrix, $F_1, F_2, \dots, F_v$ are v factors. Assume initially $Y = \varepsilon_1$.

Step 2: Choose the maximum correlation coefficient factor as F_i ($= X_1$) with Y and assume the nested model as $Y = \beta_{01} + \beta_1 X_1 + \varepsilon_2$.

Step 3: Evaluate the estimated responses and residuals ($\varepsilon_{i+1} = \varepsilon_i - \hat{Y}$ for i = 1, 2, .. .v-1).

Step 4: Choose the maximum correlated factor as F_j ($=X_{i+1}$) in the i^{th} step, between the residual and unselected factors, and assume the nested model as $\varepsilon_{i+1} = \beta_{0i+1} + \beta_{i+1}X_{i+1} + \varepsilon_{i+2}$.

Step 5: Estimate the nested model as $\hat{\varepsilon}_{i+1} = \hat{\beta}_{0i+1} + \hat{\beta}_{i+1} X_{i+1}$. Test the significance of the model.

Step 6: Repeat the steps 3-5, if the model is significant and the resulting model is the best model with the selected factors can be expressed in the form $\hat{Y} = \hat{\beta}_0 + \sum_{m=1}^{k} \hat{\beta}_m X_m$ where $\hat{\beta}_0 = \hat{\beta}_{01} + \hat{\beta}_{02} + ... + \hat{\beta}_{0k}$ and $k \leq v$.

The flow chart of the method is presented in fig 5.2.1.

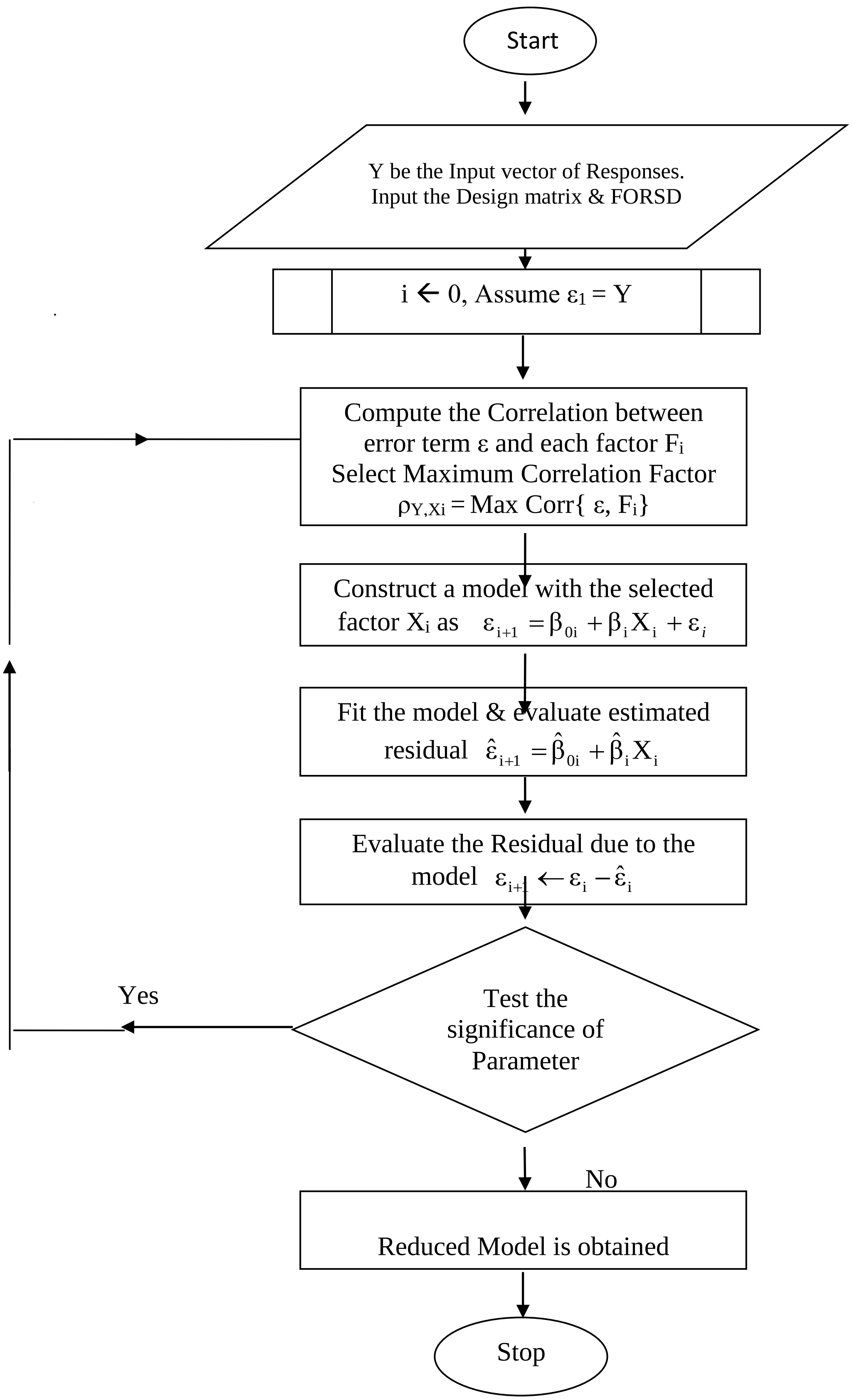

Fig 5.2.1: Nested procedure for First Order RSD Model

The method for reduction of first order response surface design model is illustrated in case of orthogonal design and non orthogonal design in the examples 5.2.1 and 5.2.2.

EXAMPLE 5.2.1: Five factors in a manufacturing process for an integrated circuit were investigated in a 2^{5-1} design with the objective of improving the process yield (Y). The five factors were A: Aperture Setting (Small, Large); B: Exposure time (20 percent below nominal, 20 percent above nominal); C: Develop Time (30 & 45 Sec); D: Mask Dimension (Small, Large) and E: Etch Time (14.5 & 15.5 Min). The construction of the 2^{5-1} design is shown below with sixteen design runs. Therefore five factors with coded variables and corresponding response values are given below.

Let Y = [8 9 34 52 16 22 45 60 6 10 30 50 15 21 44 63]′ be the vector of responses obtained at sixteen design runs [(-1 -1 -1 1 1), (1 -1 -1 1 -1), (-1 1 -1 1 -1), (1 1 -1 1 1), (-1 -1 1 1 -1), (1 -1 1 1 1), (-1 1 1 1 1), (1 1 1 1 -1), (-1 -1 -1 -1 -1), (1 -1 -1 -1 1), (-1 1 -1 -1 1), (1 1 -1 -1 -1), (-1 -1 1 -1 1), (1 -1 1 -1 -1), (-1 1 1 -1 -1), (1 1 1 -1 1)]′ respectively.

The correlation between response variable and the factors are: r (Y, X_1) = 0.293, r (Y, X_2) = 0.891, r (Y, X_3) = 0.286, r (Y, X_4) = 0.023, r (Y, X_5) = 0.016. The estimated model with maximum correlated factor is $\varepsilon_1 = 30.313 + 16.937\ X_2 + \varepsilon_2$. Evaluate ε_2 corresponding to each X_2. Mean square due to model (MSR) is 4590.063 and Mean square Error (MSE) is 84.670 and $R^2 = 0.795$; the factor X_2 is Significant.

The correlations between residual term and factors are: r (ε_2, X_1) = 0.646, r (ε_2, X_3) = 0.632, r (ε_2, X_4) = 0.051, r (ε_2, X_5) = 0.036. X_1 is selected with maximum correlation and the new model is $\varepsilon_2 = 0.000 + 5.562\ X_1 + \varepsilon_3$. MSR=495.063, MSE=49.308; R^2 =0.418; the factor X_1 is Significant.

The correlation between residual variable and unselected variables are: r (ε_3, X_3) = 0.828, r (ε_3, X_4) = 0.067, r (ε_3, X_5) = 0.048. X_3 is selected with maximum correlation and the new model is ε_3= - 0.001+ 5.437 X_3+ ε_4. MSR=473.062, MSE =15.518; R^2 =0.685; the factor X_3 is Significant. Repeat the steps until no significant factor is selected. The correlation between residual variable and unselected variables are: r (ε_4, X_4) = 0.119, r (ε_4, X_5) = 0.085; where $\varepsilon_4 = \varepsilon_3 - \hat{Y}$ ε_4= 0.00 + 0.437 X_4 + ε_5. MSR=3.062, MSE =15.299; R^2 =0.014; X_4 is Insignificant.

The nested reduced model is Y = 30.312 + 16.937 X_2 + 5.562 X_1 + 5.437 X_3, with error sum of squares 217.250 with 12 degrees of freedom and with an R^2 value is 0.962.

EXAMPLE 5.2.2: Taguchi (1987) studied PVC insulation of electric wire experiment conducted in 27 runs with nine factors. Among the nine factors, two are about Plasticizer: DOA (X_1) and DOP (X_2); two about stabilizer: Tribase (X_3) and Dyphos (X_4); three about Filler: Clay (X_5), Titanium white (X_6), and Carbon (X_7); the remaining two about process condition: number of revolutions of screw (X_8) and cylinder temperature (X_9). All nine factors are continuous and their levels are chosen to be equally spaced. The measure is the embrittlement temperature (Y). The response vector of 27 runs are:

Let Y = [5 2 8 -15 -6 -10 -28 -19 -23 -13 -17 -7 -23 -31 -23 -34 -37 -29 -27 -27 -30 -35 -35 -38 -39 -40 -41]′ corresponding at the design points [(0 0 0 0 0 0 0 0 0), (0 0 0 0 1 1 1 1 1), (0 0 0 0 2 2 2 2 2), (0 1 1 1 0 0 0 2 2), (0 1 1 1 1 1 1 0 0), (0 1 1 1 2 2 2 1 1), (0 2 2 2 0 0 0 1 1), (0 2 2 2 1 1 1 2 2), (0 2 2 2 2 2 2 0 0), (1 0 1 2 0 1 2 0 1), (1 0 1 2 1 2 0 1 2), (1 0 1 2 2 0 1 2 0), (1 1 2 0 0 1 2 2 0), (1 1 2 0 1 1 2 0 0 1), (1 1 1 2 0 2 0 1 1 2), (1 2 0 1 0 1 2 1 2), (1 2 0 1 1 2 0 2 0), (1 2 0 1 2 0 1 0 1), (2 0 2 1 0 2 1 0 2), (2 0 2 1 1 0 2 1 0), (2 0 2 1 2 1 0 2 1), (2 1 0 2

0 2 1 2 1), (2 1 0 2 1 0 2 0 2), (2 1 0 2 2 1 0 1 0), (2 2 1 0 0 2 1 1 0), (2 2 1 0 1 0 2 2 1), (2 2 1 0 2 1 0 0 2)]′ respectively.

The correlation between response variable and the factors are: $r(Y, X_1) = -0.762$, $r(Y, X_2) = -0.601$, $r(Y, X_3) = -0.124$, $r(Y, X_4) = -0.108$, $r(Y, X_5) = 0.052$, $r(Y, X_6) = -0.039$, $r(Y, X_7) = 0.114$, $r(Y, X_8) = 0.007$, $r(Y, X_9) = 0.026$. The estimated model with maximum correlated factor is $\varepsilon_1 = - 8.830 + -13.343X_1 + \varepsilon_2$. Evaluate ε_2 corresponding to each X_1. Mean square due to model is (MSR) is 3019.919 and Mean Square Error is 87.283 and $R^2 = 0.581$; the factor X_1 is Significant.

The correlation between residual term and factors are: $r(\varepsilon_2, X_2) = -0.861$, $r(\varepsilon_2, X_3) = -0.124$, $r(\varepsilon_2, X_4) = -0.099$, $r(\varepsilon_2, X_5) = 0.148$, $r(\varepsilon_2, X_6) = 0.007$, $r(\varepsilon_2, X_7) = 0.244$, $r(\varepsilon_2, X_8) = -0.057$, $r(\varepsilon_2, X_9) = -0.108$. X_2 is selected with maximum correlation and the new model is $\varepsilon_2 = 9.481 - 9.481X_2 + \varepsilon_3$. MSR = 1617.990, MSE = 22.564; $R^2 = 0.741$; the factor X_2 is Significant.

The correlation between residual variable and unselected variables are: $r(\varepsilon_3, X_3) = -0.245$, $r(\varepsilon_3, X_4) = -0.195$, $r(\varepsilon_3, X_5) = 0.291$, $r(\varepsilon_3, X_6) = 0.013$, $r(\varepsilon_3, X_7) = 0.480$, $r(\varepsilon_3, X_8) = -0.113$, $r(\varepsilon_3, X_9) = -0.212$. X_7 is selected with maximum correlation and the new model is $\varepsilon_3 = - 2.685 + 2.686X_7 + \varepsilon_4$. MSR = 129.836, MSE = 17.370; $R^2 = 0.230$; the factor X_7 is Significant. Repeat the steps until no significant factor is selected.

The correlation between residual variable and unselected variables are: $r(\varepsilon_4, X_3) = -0.279$, $r(\varepsilon_4, X_4) = -0.222$, $r(\varepsilon_4, X_5) = 0.332$, $r(\varepsilon_4, X_6) = 0.015$, $r(\varepsilon_4, X_8) = - 0.128$, $r(\varepsilon_4, X_9) = - 0.241$,where $\varepsilon_4 = \varepsilon_3 - \hat{Y}$; $\varepsilon_4 = - 1.631 + 1.630 X_5 + \varepsilon_5$. MSR = 47.834, MSE = 15.457; $R^2 = 0.110$; X_5 is insignificant.

The resulting nested reduced model is $Y = -2.034 - 13.343 X_1 - 9.481 X_2 + 2.686 X_7$, with error sum of squares 431.513 with 23 degrees of freedom and with an R^2 value is 0.917.

5.3 REDUCTION OF SECOND ORDER RSD MODEL IN NESTED APPROACH

Consider the functional relationship between the response and factors be defined as

$$Y_u = \beta_0 + \sum_{i=1}^{v} \beta_i x_{ui} + \sum_{i=1}^{v} \beta_{ii} x_{ui}^2 + \sum_{i<j}^{v} \beta_{ij} x_{ui} x_{uj} + \boldsymbol{\varepsilon}_u \qquad (5.3.1)$$

Where Y_u be the u^{th} response at the design point X_u for u=1,2,..N,

$X_u = (1, x_{u1}, x_{u2} \ldots x_{uv}, x^2_{u1}, x^2_{u2} \ldots x^2_{uv}, x_{ul}x_{u2}, \ldots, x_{uv-1}x_{uv})$ be the u^{th} treatment combination,

$\underline{\beta} = (\beta_0, \beta_1, \beta_2 \ldots \beta_v, \beta_{11}, \beta_{22} \ldots \beta_{vv}, \beta_{l2} \ldots \beta_{v-1v})'$ be the vector of parameters

ε_u be the random error corresponding to the response Y_u . Assume $\underline{\varepsilon} \sim N(0, \sigma^2 I)$.

METHOD: Let $Y = \beta_0 + \sum_{i=1}^{v} \beta_i x_{ui} + \sum_{i=1}^{v} \beta_{ii} x_{ui}^2 + \sum_{i<j}^{v} \beta_{ij} x_{ui} x_{uj} + \varepsilon$ be the full model with 'v' factors. Evaluate the correlation between the residual variables with factors and arrange them in decreasing order. Assume $r_1 < r_2 < \ldots < r_{vv-1}$ where r_i= Corr (ε_i, X_i), r_{ii}= Corr (ε_i, X_i^2), r_{ij}= Corr $(\varepsilon_{ij}, X_iX_j)$ for i, j=1, 2, 3, …, v., i≠j. Let $Y = \varepsilon_1$ and $\varepsilon_1 = \beta_{01} + \beta_1 X_1 + \varepsilon_2$, $\varepsilon_2 = \beta_{02} + \beta_2 X_2 + \varepsilon_3$, $\varepsilon_3 = \beta_{03} + \beta_3 X_3 + \varepsilon_4$, … $\varepsilon_v = \beta_{0v} + \beta_v X_v + \varepsilon_{v+1}$, … , $\varepsilon_{vv-1} = \beta_{0vv-1} + \beta_{vv-1} X_{v-1} X_v + \varepsilon_{vv}$. Select the maximum correlated factor as X_1 with response variable and estimate the parameters as $\hat{Y} = \hat{\beta}_{01} + \hat{\beta}_1 X_1$. Then evaluate the residuals and select the maximum correlated variable with the residual term and estimate the nested model as $\hat{\varepsilon}_r = \hat{\beta}_{0r} + \hat{\beta}_r X_r$ .for all possible values of r = 2, 3, …v, 11, 22, … vv, 12, 13, .. v(v-1). (denote $X_{ij} = X_i X_j$). Examine the significance of the model to include in the nested model. Repeat the procedure until no more factors will be included in the model. The resulting model will be the best choice of selected factors in the model $\hat{Y}$ =

$\hat{\beta}_0 + \sum_{i=1}^{k1} \hat{\beta}_i X_i + \sum_{i=1}^{k2} \hat{\beta}_{ii} X_i^2 + \sum_{i=1}^{k3} \sum_{j=1}^{k4} \hat{\beta}_{ij} X_i X_j$ where $\hat{\beta}_0 = \hat{\beta}_{01} + \hat{\beta}_{02} + \ldots + \hat{\beta}_{0k}$ and $k \le 2v + {}^{v}C_2$.

The Nested Approach is presented diagrammatically in fig 5.3.2.

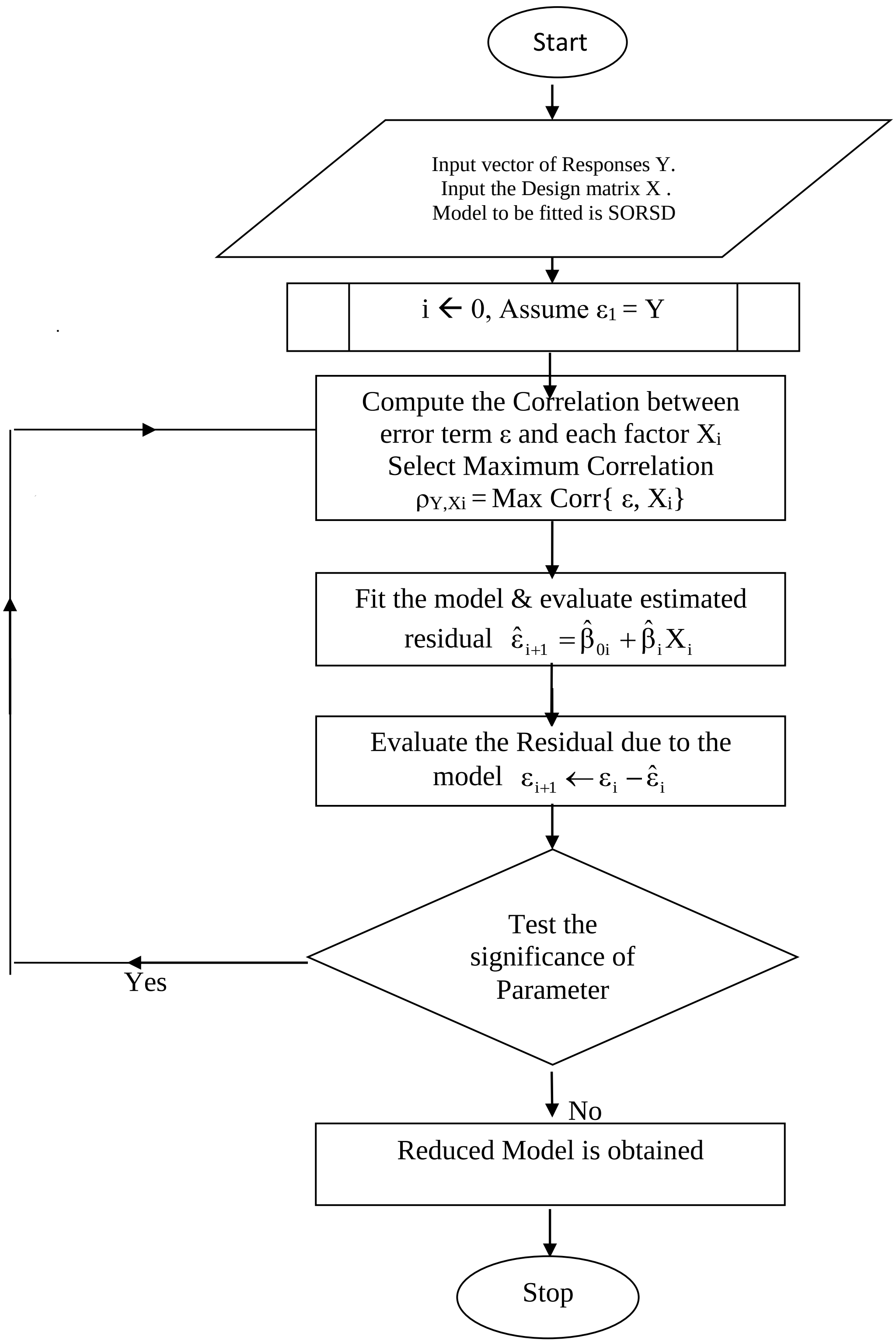

Fig 5.3.2: Nested procedure for Second Order RSD Model

The method for reduction of second order response surface design model in case of without restrictions and restrictions towards rotatability on moment matrix are illustrated in the examples 5.3.1 and 5.3.2 are presented below.

EXAMPLE 5.3.1: Khuri and Cornel (1996) studied the 2^4 factorial experimental design of the hydrogenolysis of Canadian lignite using carbon-monoxide and hydrogen mixtures as reducing agents, the input variables studied were X_1 = Temperature; X_2 = CO (H_2 ratio); X_3 = Pressure; and X_4 = Contact time. One of the response variables under investigation was Y = Percentage lignite conversion. The levels of four factors are with, X_1: Reaction temperature: 380^{o}C, 460^{o}C; X_2: Initial CO / H_2 ratio (molar ratio) is ¼: ¾ ; X_3: Initial Pressure (MPa) 7.10, 11.10; X_4: Contact time at reaction temperature (min) 10, 50.

Let Y = [53.3 78 62.4 78.9 75.9 75.4 71.3 84.4 64.5 67.5 72.8 85.3 71.4 83.3 82.9 81.7]′ be the vector of responses obtained at the sixteen design points [(-1 -1 -1 -1), (1 -1 -1 -1), (-1 1 -1 -1), (1 1 -1 -1), (-1 -1 1 -1), (1 -1 1 -1), (-1 1 1 -1), (1 1 1 -1), (-1 -1 -1 1), (1 -1 -1 1), (-1 1 -1 1), (1 1 -1 1), (-1 -1 1 1), (1 -1 1 1), (-1 1 1 1), (1 1 1 1)]′ respectively.

The correlation between response variable and the factors are: r (Y,X_1) = 0.574, r (Y,X_2) = 0.361, r (Y,X_3) = 0.456, r (Y,X_4) = 0.214, r (Y,X_1X_2) = 0.013, r (Y,X_1X_3) = - 0.239, r (Y, X_1X_4) = -0.198, r (Y, X_2X_3) = -0.156, r (Y, X_2X_4) = 0.155, r (Y, X_3X_4) = - 0.037. The estimated model with maximum correlated factor is $\varepsilon_1 = 74.313 + 5.0\ X_1 + \varepsilon_2$. Evaluate ε_2 corresponding to each X_1. Mean square due to model (MSR) is 400 and Mean square Error (MSE) is 58.278 and R^2=0.329; the factor X_1 is Significant.

The correlation between residual term and factors are: $r(\varepsilon_2, X_2) = 0.441$, $r(\varepsilon_2, X_3) = 0.557$, $r(\varepsilon_2, X_4) = 0.261$, $r(\varepsilon_2, X_1X_2) = 0.016$, $r(\varepsilon_2, X_1X_3) = -0.292$, $r(\varepsilon_2, X_1X_4) = -0.242$, $r(\varepsilon_2, X_2X_3) = -0.191$, $r(\varepsilon_2, X_2X_4) = 0.189$, $r(\varepsilon_2, X_3X_4) = -0.046$. X_3 is selected with maximum correlation and the new model is $\varepsilon_2 = -0.001 + 3.975 X_3 + \varepsilon_3$. MSR = 252.810, MSE=40.221; R^2 =0.310; the factor X_3 is Significant.

The correlation between residual variables and unselected variable are: $r(\varepsilon_3, X_2) = 0.531$, $r(\varepsilon_3, X_4) = 0.314$, $r(\varepsilon_3, X_1X_2) = 0.019$, $r(\varepsilon_3, X_1X_3) = -0.352$, $r(\varepsilon_3, X_1X_4) = -0.291$, $r(\varepsilon_3, X_2X_3) = -0.230$, $r(\varepsilon_3, X_2X_4) = 0.228$, $r(\varepsilon_3, X_3X_4) = -0.055$. X_2 is selected with maximum correlation and the new model is $\varepsilon_3 = 0.0000 + 3.15 X_2 + \varepsilon_4$. MSR = 158.760, MSE = 28.881; $R^2 = 0.282$; the factor X_2 is Significant. Repeat the steps until no significant factor is selected.

The correlation between residual variables and unselected variable are: $r(\varepsilon_4, X_4) = 0.371$, $r(\varepsilon_4, X_1X_2) = 0.022$, $r(\varepsilon_4, X_1X_3) = -0.415$, $r(\varepsilon_4, X_1X_4) = -0.343$, $r(\varepsilon_4, X_2X_3) = -0.271$, $r(\varepsilon_4, X_2X_4) = 0.269$, $r(\varepsilon_4, X_3X_4) = -0.065$; where $\varepsilon_4 = \varepsilon_3 - \hat{Y}$ and $\varepsilon_4 = 0.001 - 2.087 X_1X_3 + \varepsilon_5$. MSR = 69.722, MSE = 23.9; R^2 =0.172; X_1X_3 is Insignificant.

The nested reduced model is $Y = 74.312 + 5 X_1 + 3.975 X_3 + 3.15 X_2$, with error sum of squares 404.328 with 12 degrees of freedom and with an R^2 value is 0.667.

EXAMPLE 5.3.2: A chemical Engineer is investigating the yield (Y) of a process. Three process variables are of interest: Temperature (A), Pressure (B), and Catalyst Concentration (C). Each variable can be run at a low and a high level, and the engineer decides to run a 2^3 design with four center points. The design and the resulting yields are shown below.

Let $Y = [32 \quad 46 \quad 57 \quad 65 \quad 36 \quad 48 \quad 57 \quad 50 \quad 44 \quad 53 \quad 56]'$ be the yield of a process corresponding to coded variables X_1, X_2 & X_3 are [(-1 -1 -1), (1 -1 -1), (-1 1 -1), (1 1 -1), (-1 -1 1), (1 -1 1), (-1 1 1), (0 0 0), (0 0 0), (0 0 0), (0 0 0)]′ respectively.

The correlation between response variable and the factors are: r $(Y, X_1) = 0.328$, $r(Y, X_2) = 0.824$, $r(Y, X_3) = -0.118$, $r(Y, X_1^2) = -0.106$, $r(Y, X_2^2) = -0.106$, $r(Y, X_3^2) = -0.106$, r $(Y, X_1X_2) = -0.317$, $r(Y, X_1X_3) = -0.218$, $r(Y, X_2X_3) = -0.267$. The estimated model with maximum correlated factor is $\varepsilon_1 = 50.329 + 9.618\ X_2 + \varepsilon_2$. Evaluate ε_2 corresponding to each X_2. Mean square due to model (MSR) is 639.188 and Mean square Error (MSE) is 33.504 and $R^2 = 0.679$; the factor X_2 is Significant.

The correlation between residual term and factors are: r $(\varepsilon_2, X_1) = 0.809$, r $(\varepsilon_2, X_3) = 0.021$, r $(\varepsilon_2, X_1^2) = -0.061$, $r(\varepsilon_2, X_2^2) = -0.061$, r $(\varepsilon_2, X_3^2) = -0.061$, r $(\varepsilon_2, X_1X_2) = -0.330$, r $(\varepsilon_2, X_1X_3) = -0.155$, r $(\varepsilon_2, X_2X_3) = -0.242$. X_1 is selected with maximum correlation and the new model is $\varepsilon_2 = 0.486 + 5.348\ X_1 + \varepsilon_3$. MSR = 197.577, MSE = 11.551; $R^2 = 0.655$; the factor X_1 is Significant. Evaluate the correlation between residual variables and unselected variables are: r $(\varepsilon_3, X_3) = 0.253$, r $(\varepsilon_3, X_1^2) = 0.016$, r $(\varepsilon_3, X_2^2) = 0.016$, r $(\varepsilon_3, X_3^2) = 0.016$, r $(\varepsilon_3, X_1X_2) = -0.344$, r $(\varepsilon_3, X_1X_3) = -0.045$, r $(\varepsilon_3, X_2X_3) = -0.195$; where $\varepsilon_3 = \varepsilon_2 - \hat{Y}$. $\varepsilon_3 = -0.121 – 1.334\ X_1X_2 + \varepsilon_4$. MSR=12.301, MSE =10.185; R^2 =0.118; X_1X_2 is insignificant.

The nested reduced model is $Y = 50.815 + 9.618\ X_2 + 5.348\ X_1$, with error sum of squares 98.906 with 8 degrees of freedom and with an R^2 value is 0.895.

EXAMPLE 4.3.3: Sosada (1993) studied the effects of extraction time (X_1), solvent volume (X_2), ethanol concentration (X_3), and temperature (X_4) on the yield and phosphatidylcholine enrichment (PCE) (Y) of de-oiled rapeseed lecithin when fractionated with ethanol.

Let Y = [27.6 16.6 15.4 17.4 17.0 19.0 17.4 12.6 18.6 22.4 21.4 14.0 24.0 15.6 13.0 14.4 22.6 23.4 20.6 22.6 13.4 20.6 15.6 21.0 17.6]′ be the vector of responses obtained at 25 design points [(1 1 1 1), (-1 -1 1 1), (1 -1 -1 1), (-1 1 -1 1), (1 -1 1 -1), (-1 1 1 -1), (1 1 -1 -1), (-1 -1 -1 -1), (1 -1 1 1), (-1 1 1 1), (1 1 -1 1), (-1 -1 -1 1), (1 1 1 -1), (-1 -1 1 -1), (1 -1 -1 -1), (-1 1 -1 -1), (0 0 0 0), (1.414 0 0 0), (-1.414 0 0 0), (0 1.414 0 0), (0 -1.414 0 0), (0 0 1.414 0), (0 0 -1.414 0), (0 0 0 1.414), (0 0 0 -1.414)]′ respectively.

The correlation between response variable and the factors are: r $(Y, X_1) = 0.305$, r $(Y, X_2) = 0.623$, r $(Y, X_3) = 0.489$, r $(Y, X_4) = 0.292$, r $(Y, X_1^2) = 0.061$, r $(Y, X_2^2) = -0.231$, r $(Y, X_3^2) = -0.224$, r $(Y, X_4^2) = -0.136$, r $(Y, X_1X_2) = 0.155$, r $(Y, X_1X_3) = 0.062$, r $(Y, X_1X_4) = 0.036$, r $(Y, X_2X_3) = 0.124$, r $(Y, X_2X_4) = 0.098$, r $(Y, X_3X_4) = -0.016$. The estimated model with maximum correlated factor is $\varepsilon_1 = 18.552 + 2.691X_2 + \varepsilon_2$. Evaluate ε_2 corresponding to each X_2. Mean square due to the model (MSR) is 144.778 and Mean square Error (MSE) is 9.932 and $R^2 = 0.388$; the factor X_2 is Significant.

The correlation between residual term and factors are: r $(\varepsilon_2, X_1) = 0.390$, r $(\varepsilon_2, X_3) = 0.625$, r $(\varepsilon_2, X_4) = 0.373$, r $(\varepsilon_2, X_1^2) = 0.079$, r $(\varepsilon_2, X_2^2) = -0.296$, r $(\varepsilon_2, X_3^2) = -0.286$, r $(\varepsilon_2, X_4^2) = -0.174$, r $(\varepsilon_2, X_1X_2) = 0.198$, r $(\varepsilon_2, X_1X_3) = 0.079$, r $(\varepsilon_2, X_1X_4) = 0.046$, r $(\varepsilon_2, X_2X_3) = 0.159$, r $(\varepsilon_2, X_2X_4) = 0.126$, r $(\varepsilon_2, X_3X_4) = -0.020$. X_3 is selected with maximum correlation and the new model is $\varepsilon_2 = -3.553\text{E-}017 + 2.114 X_3 + \varepsilon_3$. MSR = 89.343, MSE = 6.048, $R^2 = 0.391$; the factor X_3 is Significant. The correlation between residual variables and unselected variables are: r $(\varepsilon_3, X_1) = 0.500$, r $(\varepsilon_3, X_4) = 0.478$, r $(\varepsilon_3, X_1^2) = 0.101$, r $(\varepsilon_3, X_2^2) = -0.379$, r $(\varepsilon_3, X_3^2) = -0.367$, r $(\varepsilon_3, X_4^2) = -0.223$, r $(\varepsilon_3, X_1X_2) = 0.254$, r $(\varepsilon_3, X_1X_3) = 0.102$, r $(\varepsilon_3, X_1X_4) = 0.059$, r $(\varepsilon_3, X_2X_3) = 0.203$, r $(\varepsilon_3, X_2X_4) = 0.161$, r $(\varepsilon_3, X_3X_4) = -0.025$. X_1 is selected with maximum correlation and the new model is $\varepsilon_3 = 0.000 + 1.318 X_1 + \varepsilon_4$. MSR = 34.742, MSE = 4.537; $R^2 = 0.250$; the factor X_1 is Significant. Repeat the steps until no significant factor is selected.

The correlation between residual variable and unselected variables are: r $(\varepsilon_4, X_4) = 0.552$, r $(\varepsilon_4, X_1^2) = 0.116$, r $(\varepsilon_4, X_2^2) = -0.438$, r $(\varepsilon_4, X_3^2) = -0.424$, r $(\varepsilon_4, X_4^2) = -0.258$, r $(\varepsilon_4, X_1X_2) =$

0.294, r (ε_4, X_1X_3) = 0.117, r (ε_4, X_1X_4) = 0.069, r (ε_4, X_2X_3) = 0.235, r (ε_4, X_2X_4) = 0.186, r (ε_4, X_3X_4) = - 0.029. ε_4= 1.776E-017 + 1.260 X_4+ ε_5. MSR = 31.773, MSE = 3.156; R^2 = 0.304; the factor X_4 is Significant. The correlation between residual variable and unselected variables are: r (ε_5, X_1^2) = 0.139, r (ε_5, X_2^2) = - 0.525, r (ε_5, X_3^2) = - 0.508, r (ε_5, X_4^2) = - 0.309, r (ε_5, X_1X_2) = 0.352, r (ε_5, X_1X_3) = 0.141, r (ε_5, X_1X_4) = 0.082, r (ε_5, X_2X_3) = 0.282, r (ε_5, X_2X_4) = 0.223, r (ε_5, X_3X_4) = -0.035. ε_5= 1.264 - 1.580 X_2^2 + ε_6. MSR = 19.976, MSE = 2.287; R^2 = 0.275; the factor X_2^2 is Significant.

The correlation between residual variable and unselected variables are: r (ε_6, X_1^2) = 0.164, r (ε_6, X_3^2) = - 0.597, r (ε_6, X_4^2) = - 0.363, r (ε_6, X_1X_2) = 0.414, r (ε_6, X_1X_3) = 0.165, r (ε_6, X_1X_4) = 0.097, r (ε_6, X_2X_3) = 0.331, r (ε_6, X_2X_4) = 0.262, r (ε_6, X_3X_4) = - 0.041, ε_6 = 1.224 - 1.530 X_3^2 + ε_7. MSR = 18.723, MSE = 1.473; R^2 = 0.356; X_3^2 is Significant. The correlation between residual variable and unselected variables are: r (ε_7, X_1^2) = 0.204, r (ε_7, X_4^2) = - 0.452, r (ε_7, X_1X_2) = 0.515, r (ε_7, X_1X_3) = 0.206, r (ε_7, X_1X_4) = 0.120, r(ε_7, X_2X_3) = 0.412, r (ε_7, X_2X_4) = 0.326, r (ε_7, X_3X_4) = - 0.052. ε_7 = 0.000 + 0.750 X_1X_2 + ε_8. MSR = 9.000, MSE = 1.082; R^2 = 0.266; X_1X_2 is Significant. The correlation between residual variable and unselected variables are: r (ε_8, X_1^2) = 0.238, r(ε_8, X_4^2) = - 0.527, r (ε_8, X_1X_3) = 0.241, r (ε_8, X_1X_4) = 0.140, r (ε_8, X_2X_3) = 0.481, r (ε_8, X_2X_4) = 0.381, r (ε_8, X_3X_4) = - 0.060. ε_8 = 0.744 - 0.930 X_4^2 + ε_9. MSR = 6.912, MSE = 0.781; R^2 = 0.278; X_4^2 is Significant.

The correlation between residual variable and unselected variables are: r (ε_9, X_1^2) = 0.281, r (ε_9, X_1X_3) = 0.283, r (ε_9, X_1X_4) = 0.165, r (ε_9, X_2X_3) = 0.566, r (ε_9, X_2X_4) = 0.488, r (ε_9, X_3X_4) = -0.071. ε_9= 0.000 + 0.600 X_2X_3 + ε_{10}. MSR = 5.760, MSE = 0.531; R^2 = 0.320; X_2X_3 is Significant. The correlation between residual variable and unselected variables are: r (ε_{10}, X_1^2) = 0.320, r (ε_{10}, X_1X_3) = 0.343, r (ε_{10}, X_1X_4) = 0.200, r (ε_{10}, X_2X_4) = 0.544, r (ε_{10},

X_3X_4) = -0.086. ε_{11}= 0.000 + 0.475 X_2X_4 + ε_{12}. MSR = 3.610, MSE = 0.374; R^2 = 0.296; X_2X_4 is Significant.

The correlation between residual variable and unselected variables are: r (ε_{11}, X_1^2) = 0.405, r (ε_{11}, X_1X_3) = 0.409, r (ε_{11}, X_1X_4) = 0.239, r (ε_{11}, X_3X_4) = - 0.102. ε_{12}= 0.000 + 0.3 X_1X_3 + ε_{12}. MSR = 1.440, MSE = 0.311; R^2 = 0.167; X_1X_3 is Significant. The correlation between residual variable and unselected variables are: r (ε_{12}, X_1^2) = 0.444, r (ε_{12}, X_1X_4) = 0.262, r (ε_{12}, X_3X_4) = - 0.112. ε_{13}= - 0.337 + 0.421 X_1^2 + ε_{14}. MSR = 1.415, MSE = 0.250; R^2 = 0.197; X_1^2 is Significant.

The correlation between residual variable and unselected variables are: r (ε_{13}, X_1X_4) = 0.292, r (ε_{13}, X_3X_4) = -0.125. ε_{14}= 2.513e-005 + 0.175 X_1X_4+ ε_{15}. MSR = 0.490, MSE = 0.229; R^2 = 0.085; X_1X_4 is insignificant.

The resulting nested reduced model is Y= 21.447 + 2.691 X_2 + 2.114 X_3 + 1.318 X_1 + 1.260 X_4 – 1.580 X_2^2 – 1.530 X_3^2 + 0.750 X_1X_2 – 0.930 X_4^2 + 0.6 X_2X_3 + 0.475 X_2X_4 + 0.3 X_1X_3 + 0.421 X_1^2 with error sum of squares 5.479 with 12 degrees of freedom and with an R^2 value 0.985.

5.4 COMPARISON OF NESTED MODEL WITH CLASSICAL MODELS AND THEIR ANALYSIS

An attempt is made to analyze the reduced First order (examples 5.2.1, 5.2.2) and Second Order (examples 5.3.1, 5.3.2 and 5.3.3) Response Surface Design Models in Nested approach are presented. The comparison in respect of (a) the estimated values for the parameters and the selection of best models using Step-wise approach, Forward approach and Backward elimination approach (b) the estimated values for the parameters and confidence interval at 95% level for the parameters and (c) the sum of squares due to reduced model and its R^2 are evaluated.

5.4.1 FIRST ORDER RSD MODEL

a) The estimated values for parameters in case of full model and reduced models in different approaches is presented in Table 5.4.1

Table 5.4.1: Example 5.2.1

S.No.	Parameters	Full Model	Stepwise	Forward	Backward	Nested
1	β_0	30.313	30.313	30.313	30.313	30.313
2	β_1	5.563	5.563	5.563	5.563	5.563
3	β_2	16.937	16.937	16.937	16.937	16.937
4	β_3	5.437	5.437	5.437	5.437	5.437
5	β_4	0.438	-	-	-	-
6	β_5	0.312	-	-	-	-

b) The estimated values for the parameters and their Confidence Intervals at 95% level corresponding to the data presented in example 5.2.1 (orthogonal design) of First Order Response Surface Model are presented in the table 5.4.2.

Table 5.4.2: Example 5.2.1

S.No.	Parameters	Full model	95% Credible Intervals		Significance Value
			Lower	Upper	
1	β_0	30.313	27.744	32.881	26.295
2	β_1	5.563	2.994	8.131	4.825
3	β_2	16.937	14.369	19.506	14.693
4	β_3	5.437	2.869	8.006	4.717
5	β_4	0.438	-2.131	3.006	0.380
6	β_5	0.312	-2.256	2.881	0.271

c) Comparison of S.S & R^2 of full model with reduced models is presented in table 5.4.3

Table 5.4.3: Example 5.2.1

	Full model	Stepwise	Forward	Backward	Nested
SSR	5562.813	5558.19	5558.19	5558.19	5558.188
ESS	212.625	217.25	217.25	217.25	217.25
MSE	21.262	18.104	18.104	18.104	18.104
R^2 value	0.963	0.962	0.962	0.962	0.962

Note: It can be noted that the values of parameters β_4, β_5 are insignificant in nested and other regression approaches. So in all the approaches the reduced model is same and the mean square due to residual and R^2 values are same.

5.4.2 FIRST ORDER RSD MODEL IN EXAMPLE 5.2.2

a) The estimated values for parameters in case of full model and reduced models in different approaches are presented in Table 5.4.4.

Table 5.4.4: Table 4.4.1: Example 5.2.2

S.No.	Parameters	Full Model	Stepwise	Forward	Backward	Nested
1	β_0	0.256	-2.408	-2.408	-2.408	-2.034
2	β_1	-12.996	-12.94	-12.94	-12.94	-13.343
3	β_2	-9.5	-9.503	-9.503	-9.503	-9.481
4	β_3	-1.389	-	-	-	-
5	β_4	-1.111	-	-	-	-
6	β_5	1.611	-	-	-	-
7	β_6	0.055	-	-	-	-
8	β_7	2.666	2.663	2.663	2.663	2.686
9	β_8	-0.611	-	-	-	-
10	β_9	-1.166	-	-	-	-

b) The estimated values for the parameters and their Confidence Intervals at 95% level corresponding to the data presented in example 5.2.2 (non-orthogonal design) of First Order Response Surface Model are presented in the Table 5.4.5.

Table 5.4.5: Table 5.4.1: Example 5.2.2

S.No.	Parameters	Full model	95% Credible Intervals		Significance Value
			Lower	Upper	
1	β_0	0.256	-6.110	6.621	0.085
2	β_1	-12.996	-15.164	-10.827	-12.644
3	β_2	-9.5	-11.581	-7.419	-9.633
4	β_3	-1.389	-3.470	0.692	-1.409
5	β_4	-1.111	-3.192	0.969	-1.127
6	β_5	1.611	-0.470	3.691	1.633
7	β_6	0.055	-2.024	2.136	0.056
8	β_7	2.666	0.586	4.747	2.704
9	β_8	-0.611	-2.692	1.470	-0.619
10	β_9	-1.166	-3.247	0.914	-1.183

c) Comparison of S.S & R^2 of full model with reduced models is presented in table 5.4.6.

Table 5.4.6: Example 5.2.2

	Full model	Stepwise	Forward	Backward	Nested
SSR	4905.37	4770.49	4770.49	4770.49	4770.49
ESS	296.629	431.513	431.513	431.513	431.513
MSE	17.449	18.761	18.761	18.761	18.761
R^2 value	0.943	0.917	0.917	0.917	0.917

Note: It can be noted that the values of parameters β_3, β_4, β_5, β_6, β_8, β_9 are insignificant in nested and other regression approaches. So, in all the approaches the reduced model is same and the mean square due to residual and R^2 values are same.

5.4.3 SECOND ORDER RSD MODEL IN EXAMPLE 5.3.1

a) The estimated values for parameters in case of full model and reduced models in different approaches is presented in Table 5.4.7

Table 5.4.7: Example 5.3.1

S.No.	Parameters	Full Model	Stepwise	Forward	Backward	Nested
1	β_0	74.313	74.313	74.313	74.313	74.313
2	β_1	5.000	5	5	5	5
3	β_2	3.15	-	-	3.15	3.15
4	β_3	3.975	3.975	3.975	3.975	3.975
5	β_4	1.862	-	-	-	-
6	β_{12}	0.113	-	-	-	-
7	β_{13}	-2.088	-	-	-	-
8	β_{14}	-1.725	-	-	-	-
9	β_{23}	-1.363	-	-	-	-
10	β_{24}	1.35	-	-	-	-
11	β_{34}	-0.325	-	-	-	-

b) The estimated values for the parameters and their Confidence Intervals at 95% level are presented in example 5.3.1 (orthogonal design) of Second Order Response Surface Design Model are presented in the table 5.4.8.

Table 5.4.8: Example 5.3.1

S.No.	Parameters	Full model	95% Credible Intervals		Significance Value
			Lower	Upper	
1	β_0	74.313	70.557	78.068	50.868
2	β_1	5.000	1.245	8.755	3.423

Table 5.4.8: Example 5.3.1

S.No.	Parameters	Full model	95% Credible Intervals		Significance Value
			Lower	Upper	
3	β_2	3.15	-0.605	6.905	2.156
4	β_3	3.975	0.220	7.730	2.721
5	β_4	1.862	-1.893	5.618	1.275
6	β_{12}	0.113	-3.643	3.868	0.077
7	β_{13}	-2.088	-5.83	1.668	-1.429
8	β_{14}	-1.725	-5.480	2.030	-1.81
9	β_{23}	-1.363	-5.118	2.393	-0.933
10	β_{24}	1.35	-2.405	5.105	0.924
11	β_{34}	-0.325	-4.080	3.430	-0.222

c) Comparison of S.S & R^2 of full model with reduced models is presented in table 5.4.9

Table 5.4.9: Example 5.3.1

	Full model	Stepwise	Forward	Backward	Nested
SSR	1045.16	652.81	652.81	811.57	811.57
ESS	170.738	563.088	563.088	404.328	404.328
MSE	34.148	43.314	43.314	33.694	33.694
R^2 value	0.860	0.537	0.537	0.667	0.667

Note: It can be noted that the values of parameters β_4, β_{11}, β_{22}, β_{33}, β_{44}, β_{12}, β_{13}, β_{14}, β_{23}, β_{24}, β_{34} are insignificant in nested and backward approaches whereas in stepwise and forward β_2, β_4, β_{11}, β_{22}, β_{33}, β_{44}, β_{12}, β_{13}, β_{14}, β_{23}, β_{24}, β_{34} are insignificant. Hence the corresponding reduced model, Mean square error due to residual and R^2 values are same.

5.4.4 SECOND ORDER RSD MODEL IN EXAMPLE 5.3.2:

a) The estimated values for parameters in case of full model and reduced models in different approaches is presented in table 5.4.10

Table 5.4.10: Example 5.3.2

S.No.	Parameters	Full Model	Stepwise	Forward	Backward	Nested
1	β_0	50.75	50.906	50.906	50.906	50.815
2	β_1	5	5.484	5.484	5.484	5.348
3	β_2	10	10.484	10.484	10.484	9.618
4	β_3	0.5	-	-	-	-
5	β_{33}	-0.25	-	-	-	-
6	β_{12}	-1.5	-	-	-	-
7	β_{13}	-0.5	-	-	-	-
8	β_{23}	-1	-	-	-	-

b) The estimated values for the parameters and their Confidence Intervals at 95% level corresponding to the data presented in example 5.3.2 (non-orthogonal design) of Second Order Response Surface Model are presented in the table 5.4.11

Table 5.4.11: Example 5.3.2

S.No.	Parameters	Full model	95% Credible Intervals		Significance Value
			Lower	Upper	
1	β_0	50.75	42.597	58.902	19.811
2	β_1	5	-3.152	13.152	1.952
3	β_2	10	1.847	18.152	3.904
4	β_3	0.5	-7.652	8.652	0.195
5	β_{33}	-0.25	-11.779	11.279	-0.069
6	β_{12}	-1.5	-9.652	6.652	-0.586
7	β_{13}	-0.5	-8.652	7.652	-0.195

8	β_{23}	-1	-9.152	7.152	-0.390

c) Comparison of S.S & R^2 of full model with reduced models is presented in table 5.4.12

Table 5.4.12: Example 5.3.2

	Full model	Stepwise	Forward	Backward	Nested
SSR	861.977	841.821	841.821	841.821	841.821
ESS	78.75	98.906	98.906	98.906	98.906
MSE	26.25	12.363	12.363	12.363	12.363
R^2 value	0.916	0.895	0.895	0.895	0.895

Note: It can be noted that the values of parameters β_3, β_{11}, β_{22}, β_{33}, β_{12}, β_{13}, β_{23} are insignificant in nested and other regression approaches. So in all the approaches the reduced model is same and the mean square error due to residual and R^2 values are same.

5.4.5 SECOND ORDER RSD MODEL IN EXAMPLE 5.3.3

a) The estimated values for parameters in case of full model and reduced models in different approaches is presented in table 5.4.13

Table 5.4.13: Example 5.3.3

S.No.	Parameters	Full Model	Stepwise	Forward	Backward	Nested
1	β_0	21.447	21.783	21.783	21.783	21.447
2	β_1	1.318	1.318	1.318	1.318	1.318
3	β_2	2.691	2.691	2.691	2.691	2.691
4	β_3	2.114	2.114	2.114	2.114	2.114
5	β_4	1.26	1.26	1.26	1.26	1.26
6	β_{11}	0.421	-	-	-	0.421
7	β_{22}	-1.58	-1.58	-1.58	-1.58	-1.58
8	β_{33}	-1.53	-1.53	-1.53	-1.53	-1.53

9	β_{44}	-0.93	-0.93	-0.93	-0.93	-0.93
10	β_{12}	0.75	0.75	0.75	0.75	0.75
11	β_{13}	0.3	-	-	-	0.3
12	β_{14}	0.175	-	-	-	-
13	β_{23}	0.6	0.6	0.6	0.6	0.6
14	β_{24}	0.475	0.475	0.475	0.475	0.475
15	β_{34}	-0.075	-	-	-	-

b) The estimated values for the parameters and their Confidence Intervals at 95% level corresponding to the data presented in the example 5.3.3 (central composite design) of Second Order Response Surface Model are presented in the table 5.4.14.

Table 5.4.14: Example 5.3.3

S.No.	Parameters	Full model	95% Credible Intervals		Significance Value
			Lower	Upper	
1	β_0	21.447	20.486	22.408	49.725
2	β_1	1.318	0.960	1.676	8.197
3	β_2	2.691	2.332	3.049	16.732
4	β_3	2.114	1.755	2.472	13.144
5	β_4	1.26	0.902	1.619	7.839
6	β_{11}	0.421	-0.146	0.988	1.655
7	β_{22}	-1.58	-2.147	-1.013	-6.213
8	β_{33}	-1.53	-2.097	-0.963	-6.016
9	β_{44}	-0.93	-1.496	-0.363	-3.656
10	β_{12}	0.75	0.349	1.151	4.172
11	β_{13}	0.3	-0.101	0.701	1.669
12	β_{14}	0.175	-0.226	0.576	0.973
13	β_{23}	0.6	0.199	1.001	3.337
14	β_{24}	0.475	0.074	0.876	2.642
15	β_{34}	-0.075	-0.476	0.326	-0.417

c) Comparison of S.S & R^2 of full model with reduced models is presented in table 5.4.15

Table 5.4.15: Example 5.3.3

	Full model	Stepwise	Forward	Backward	Nested
SSR	368.053	364.618	364.618	364.618	367.473
ESS	5.169	8.604	8.604	8.604	5.749
MSE	0.517	0.615	0.615	0.615	0.479
R^2 value	0.986	0.977	0.977	0.977	0.985

Note: It can be noted that the values of parameters β_{14}, β_{34} are insignificant in nested approach. Whereas β_{11}, β_{13}, β_{14}, β_{34} are insignificant in all other approaches.

5.5 REMARKS ON NESTED SELECTION OF MODEL

1. The nested approach avoids the problem of multi-collinearity by selecting single variable at each step.
2. It reduces the size of the original model by selecting a subset of variables from the original set of variables iteratively in a forward approach.
3. If the number of factors are more one can be linearly predicted from the others with a non-trivial degree of accuracy.
4. The time complexity of Nested selection of model is same as that of classical standard procedures in estimation and testing the parameters significance in each step of iteration.
5. The selection of variables in the model is almost same when compared with all classical approaches in addition to that it selects some more variables in some cases which was observed when testing number of models.
6. The maximum number of iterations depends on the number of components in the model.

6. REDUCTION IN DIMENSIONALITY OF RESPONSE SURFACE DESIGN MODEL USING BAYESIANAPPROACH

6.1 INTRODUCTION

Let $\underline{Y} = [Y_1\ Y_2\ \ldots\ Y_n]'$ be the observed sample drawn from a population with parameter θ. The joint density function of observed sample $\underline{Y}$ for the given parameter θ, called the likelihood function of $\underline{Y}$ denoted by P(Y/θ). Probability distribution for each parameter encapsulates the prior beliefs held about their most likely values called Prior distribution. P(θ) explains the information about the parameter θ called prior distribution of θ. An updated measure for each of the parameter values θ based on prior and given knowledge on Y called the posterior distribution of θ given Y. Bayesian approach is a powerful technique can be used to estimate the uncertainties of parameters based on the posterior probabilities. Posteriors can be computed using the likelihood and prior distributions, since the marginal likelihood is simply a normalized constant, which need not be explicitly calculated. The prior must be defined for every parameter to strengthen the Bayesian approach.

$$f\,(\theta/Y) \quad = \quad \frac{f(\theta)\ f(Y/\theta)}{\int f(\theta)\, f(Y/\theta)\ d\theta} \tag{6.1.1}$$

$\Rightarrow$ Posterior distribution ∝ Prior distribution * Likelihood function

The posterior probability of θ given Y can be evaluated by generating a sequence of sample values in such a way that, as more and more sample values as possible, such that the distribution of sample values more closely approximates the desired distribution and is used to evaluate the normalized constant $f\,(\underline{Y}) = \int f\,(\theta).f\,(\underline{Y}/\theta)\ d\theta$. The computation of normalized constant manually is more difficult due to not knowing about all possible samples of population of parameter. To evaluate the normalized constant, we can also use Gibbs sampling technique or Metropolis Hasting Algorithm etc.

6.2 POSTERIOR DISTRIBUTION OF REGRESSION PARAMETER

Consider a simple linear regression model, $Y = \beta_0 + \beta_1 X + \varepsilon$ expressed in the form $Y = X\beta + \varepsilon$ to be fitted for the data based on the observed sample. Let $(x_1, y_1), (x_2, y_2), \dots , (x_n, y_n)$ be a random sample of size 'n' drawn from a bi-variate normal population with means μ_x and μ_y and variance σ_x^2 and σ_y^2. Assume $Y \sim N(X\beta, \sigma^2)$ and $\varepsilon \sim N(\mathbf{\underline{0}}, \sigma^2 I)$.

$$\begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \dots & \dots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_n \end{bmatrix} \qquad (6.2.1)$$

The likelihood function of observed sample $(y_1, y_2, \dots , y_n)$ is

$$L(y) = (2\pi\sigma^2)^{-n/2} \exp\{-(2\sigma^2)^{-1}(Y - X\beta)'(Y - X\beta)\} \qquad (6.2.2)$$

$$= (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}(Y - \hat{\beta}X + \hat{\beta}X - \beta X)^2\right\}$$

$$= (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\left[(Y - X\hat{\beta})^2 + \Sigma X_i^2(\beta - \hat{\beta})^2\right]\right\}$$

$$= (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\left[\hat{\sigma}^2(n-1) + \Sigma X_i^2(\beta - \hat{\beta})^2\right]\right\}$$

$$= (2\pi\sigma^2)^{-n/2}.\exp\left\{\frac{-\hat{\sigma}^2(n-1)}{2\sigma^2}\right\}.\exp\left\{-\frac{1}{2\sigma^2/\Sigma X_i^2}(\beta - \hat{\beta})^2\right\}$$

$$= (\sqrt{2\pi})\sigma^{-1} \exp\left\{-\frac{1}{2\sigma^2/\Sigma X_i^2}(\beta - \hat{\beta})^2\right\}.(\sigma^2)^{-(n-1)/2}.\exp\left\{\frac{-\hat{\sigma}^2(n-1)}{2\sigma^2}\right\}$$

$$\therefore \ \beta \sim N\left(\hat{\beta}, \frac{\sigma^2}{\Sigma X_i^2}\right) \quad \text{and} \quad \sigma^2 \sim IG\left(\frac{n-1}{2}, \frac{(n-1)\hat{\sigma}^2}{2}\right) \qquad (6.2.3)$$

The posterior distribution of parameters can be evaluated using Bayesian rule as

$$P(\beta, \sigma^2 / y) = P(\sigma^2) \,.\, P(y/ \beta, \sigma^2) = \prod_{i=1}^{n} (\sigma^2)^{-1/2} \exp\{ (-1/ 2\sigma^2) [y_i - f_i(\boldsymbol{x}, \beta)]^2 \qquad (6.2.4)$$

Integrating with respect to σ^2 we can obtain

$$P(\beta) \quad \propto \quad \prod ([y_i - f_i(\boldsymbol{x}, \beta)]^2)^{-n/2} \, d\beta \qquad (6.2.5)$$

The parameter β can be estimated using the observed sample vector $\underline{\mathbf{y}}$ and based on the prior distribution P(β), which follows $N(X\beta, \sigma^2/\sum X_i^2)$.

6.3 BAYESIAN SELECTION OF REDUCED FIRST ORDER RESPONSE SURFACE DESIGN MODEL

Let there be 'v' factors each at 's' levels for experimentation. Consider the N treatment combinations of 'v' factors with different levels to form a design $X_{N\times v} = ((x_{ui}))$ u=1, …N, i=1,2,…v, where x_{ui} be the level of the i^{th} factor in the u^{th} treatment combination. Let us assume that there are k independent variables $X_1, X_2, \ldots X_v$ and one dependent variable y. If the model is of first order, then

$$Y_u = \beta_0 + \beta_1 x_{u1} + \beta_2 x_{u2} + \ldots\ldots. + \beta_k x_{uv} + \varepsilon_u \qquad (6.3.1)$$

It can be expressed in matrix form as

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \cdots \\ \cdots \\ \cdots \\ Y_N \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \ldots & \ldots & x_{1v} \\ 1 & x_{21} & x_{22} & \ldots & \ldots & x_{2v} \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & x_{u1} & x_{u2} & \ldots & \ldots & x_{uv} \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & x_{N1} & x_{N2} & \ldots & .. & x_{Nv} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ . \\ . \\ . \\ \beta_v \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdots \\ \cdots \\ \cdots \\ \varepsilon_N \end{bmatrix}$$

$$\underline{Y} = X \underline{\boldsymbol{\beta}} + \underline{\boldsymbol{\varepsilon}} \qquad (6.3.2)$$

Where $Y = [Y_1, Y_2, \ldots, Y_N]'$ be the observed sample on the random variable $Y \sim N(X\beta, \sigma^2)$.

X = be the data matrix of size N with 'v' predictor variables. Assume X ~ N(μ_x ,σ_x^2).

β = [β_0 β_1 …, β_v]′ be the vector of parameters.

ε = [ε_1, ε_2, … , ε_N]′ be the vector of random errors. $\underline{\boldsymbol{\varepsilon}}$ ~ MVN ($\underline{\mathbf{0}}$, $\sigma^2\mathbf{I}_N$)).

The likelihood function of an observed sample Y with the given parameters β and σ^2 is

$$L(Y/\beta,\sigma^2)=(2\pi\sigma^2)^{-\frac{N}{2}}\exp\left\{-(2\sigma^2)^{-1}\sum_{i=1}^{N}(Y_i-X_i\beta)'(Y_i-X_i\beta)\right\} \quad (6.3.3)$$

$$L(Y,\beta,\sigma^2)=\frac{1}{(2\pi\sigma^2)^{N/2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(Y_i-X_i\beta)'(Y_i-X_i\beta)\right\}$$

$$\Rightarrow L(Y,\beta,\sigma^2)=\frac{1}{(2\pi\sigma^2)^{N/2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{N}[(Y_i-X_i\hat{\beta}+X_i\hat{\beta}-X_i\beta)'(Y_i-X_i\hat{\beta}+X_i\hat{\beta}-X_i\beta)]\right\}$$

$$\Rightarrow L(Y/\beta,\sigma^2)=\frac{1}{(2\pi\sigma^2)^{N/2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{N}[(Y_i-X_i\hat{\beta})'(Y_i-X_i\hat{\beta})+(\beta-\hat{\beta})'X'X(\beta-\hat{\beta})]\right\}$$

$$\Rightarrow L(Y/\beta,\sigma^2)=(2\pi\sigma^2)^{-\frac{N}{2}}\exp\left\{-(2\sigma^2)^{-1}\sum_{i=1}^{N}[(Y_i-X_i\hat{\beta})'(Y_i-X_i\hat{\beta})+(\beta-\hat{\beta})'X'X(\beta-\hat{\beta})]\right\}$$

(6.3.4)

$$\frac{\partial L}{\partial\beta}\Rightarrow\hat{\beta}=(X'X)^{-1}X'Y \quad \text{and} \quad \frac{\partial L}{\partial\sigma^2}\Rightarrow\hat{\sigma}^2=\frac{(Y-X\beta)'(Y-X\beta)}{N-v} \quad (6.3.5)$$

Then, the distribution of the parameters is:

$$\Rightarrow L(Y/\beta,\sigma^2)=(2\pi\sigma^2)^{-\frac{N}{2}}\exp\left\{-(2\sigma^2)^{-1}\sum_{i=1}^{N}[(Y_i-X_i\hat{\beta})'(Y_i-X_i\hat{\beta})+(\beta-\hat{\beta})'X'X(\beta-\hat{\beta})]\right\}$$

$$=(2\pi\sigma^2)^{-N/2}\exp\left\{-\frac{1}{2\sigma^2}\left[(Y-X\hat{\beta})^2+\sum_{i=1}^{N}X_i^2(\beta-\hat{\beta})^2\right]\right\}$$

$$=(2\pi\sigma^2)^{-N/2}\exp\left\{-\frac{1}{2\sigma^2}\left[\hat{\sigma}^2(N-v)+\sum_{i=1}^{N}X_i^2(\beta-\hat{\beta})^2\right]\right\}$$

$$= \quad (2\pi\sigma^2)^{-n/2}.\exp\left\{\frac{-\hat{\sigma}^2(N-v)}{2\sigma^2}\right\}.\exp\left\{-\frac{1}{2\sigma^2/\Sigma X_i^{\ 2}}(\beta-\hat{\beta})^2\right\}$$

$$= \quad (\sqrt{2\pi})\sigma^{-1}\ \exp\left\{-\frac{1}{2\sigma^2/\Sigma X_i^{\ 2}}(\beta-\hat{\beta})^2\right\}.(\sigma^2)^{-(N-1)/2}.\exp\left\{\frac{-\hat{\sigma}^2(N-v)}{2\sigma^2}\right\} \quad (6.3.6)$$

Then the posterior distribution of the parameter β is

$$P(\beta / X,Y) \ \propto\ \sigma^{-N-1}\ \exp\left\{-\frac{1}{2\sigma^2}\left[(Y - X\hat{\beta})'(Y - X\hat{\beta}) + (\beta - \hat{\beta})'\,X'X\,(\beta - \hat{\beta})\right]\right\}$$

$$\propto \quad \sigma^{-N-1}\exp\ \{-\frac{1}{2\sigma^2}[\,(N-v)\hat{\sigma}^2 + (\beta-\hat{\beta})'\,X'X\,(\beta-\hat{\beta})\,]\,\}$$

$$\text{Since}\,(N-v)\,\hat{\sigma}^2 = (Y - X\hat{\beta})'\,(Y - X\hat{\beta})\,)$$

$$\propto \ \sigma^{-N-1}\exp\,(-\frac{1}{2\sigma^2}(\beta-\hat{\beta})'X'X(\beta-\hat{\beta})\})$$

$$\sim \text{MVN}\,(\hat{\beta},(X'X)^{-1}\sigma^2) \qquad (6.3.7)$$

And, the posterior distribution of the parameter σ^2 is

$$P(\sigma^2 / X,Y) \propto \sigma^{-N-1}\exp(-\frac{1}{2\sigma^2}\{(N-v)\hat{\sigma}^2 + (\beta-\hat{\beta})'X'X(\beta-\hat{\beta})\})$$

$$\propto \sigma^{-(N-v+1)}\exp(-\frac{1}{2\sigma^2}(N-v)\,\hat{\sigma}^2)$$

$$\sim \text{IG}\,[\,(N-v)/2,\,(N-v)\hat{\sigma}^2/2\,] \qquad (6.3.8)$$

Let Y be the response variable follow Normal distribution (say) and let X_1, X_2, …, X_v be the predictor variables for the selection of best model for the response Y. Let M_1, M_2, …, M_m be the possible models ($m = 2^v$) each containing subsets of X_1, X_2, … X_v. For each j, Model M_j (j=1,2, … k) is defined by a family of distributions $P_{\theta j}$ where $P_{\theta j}$ has some prior distribution. The dimension of $P_{\theta j}$ for each j need not be same. The posterior probability for all possible models M_1, M_2, … M_m can be evaluated based on their priors, likelihood and Normalized constant (unknown). Then, select the best model M* with the highest posterior probability from M_1, M_2, ... M_m. The method is illustrated in case of orthogonal design and non orthogonal designs in the examples 6.3.1 and 6.3.2.

EXAMPLE 6.3.1: Consider the experimental data considered in the example 5.2.1 with five factors in 16 design points. Using R the posteriors for all the 32 possible linear combination models are evaluated and the highest posterior probability is presented in Table 6.3.1

Table 6.3.1

S.No.	Parameters	Posterior Mean	Posterior S.D.	P≠0
1	β_0	30.313	1.082	1.00
2	β_1	5.562	1.082	1.00
3	β_2	16.937	1.082	1.00
4	β_3	5.437	1.082	1.00
5	β_4	0.095	0.551	0.21
6	β_5	0.065	0.526	0.20

.

Bayesian estimated parameters values are: $\hat{\beta}_0 = 30.31$; $\hat{\beta}_1 = 5.562$; $\hat{\beta}_2 = 16.94$; $\hat{\beta}_3 = 5.436$; $\hat{\beta}_4 = 0.4381$; $\hat{\beta}_5 = 0.3126$. The resulting reduced Bayesian model is Y = 30.313 + 5.562 X_1 + 16.937 X_2 + 5.437 X_3, with error sum of squares 217.250 with 12 degrees of freedom and with an R^2 value is 0.962 selected with highest probability value 0.618 with three variables.

EXAMPLE 6.3.2: Consider the experimental data considered in the example 5.2.2 with nine factors in 27 design points. Using R the posteriors for all the 512 ($m = 2^9$) possible linear combination models are evaluated and the highest posterior probability is presented in Table

6.3.2. The Bayesian estimated parameters values are: $\hat{\beta}_0$ = 0.417; $\hat{\beta}_1$ = - 12.78; $\hat{\beta}_2$ = - 9.329; $\hat{\beta}_3$ = - 1.219; $\hat{\beta}_4$ = - 0.942; $\hat{\beta}_5$ = - 0.323; $\hat{\beta}_6$ = 0.225; $\hat{\beta}_7$ = 2.835; $\hat{\beta}_8$ = - 0.433; $\hat{\beta}_9$ = - 0.984.

Table 6.3.2

S.No.	Parameters	Posterior Mean	Posterior S.D	P≠0
1	β_0	-1.795	2.610	1.00
2	β_1	-12.955	1.042	1.00
3	β_2	-9.503	1.004	1.00
4	β_3	-0.544	0.917	0.39
5	β_4	-0.301	0.716	0.27
6	β_5	0.796	1.060	0.49
7	β_6	0.006	0.344	0.11
8	β_7	2.542	1.123	0.95
9	β_8	-0.093	0.449	0.15
10	β_9	-0.340	0.754	0.29

The resulting reduced Bayesian model is $Y = - 2.408 – 12.940 X_1 – 9.503 X_2 + 2.663 X_7$, with error sum of squares 431.513 with 23 degrees of freedom and with an R^2 value is 0.917 selected with highest probability value 0.105 with three variables.

6.4 BAYESIAN SELECTION OF REDUCED SECOND ORDER RESPONSE SURFACE DESIGN MODEL

Let there be ‘v’ factors each at‘s’ levels. Consider the N treatment combinations of ‘v’ factors with different levels to form a basic design $X_{N\times v} = ((x_{ui}))$ u = 1,…N, i=1,2,…v, where x_{ui} be the level of the i^{th} factor in the u^{th} treatment combination. Let us assume that there are v independent variables X_1, X_2, …. X_v and one dependent variable Y. If model is of second order, then

$$Y_u = \beta_0 + \sum_{i=1}^{v} \beta_i x_{ui} + \sum_{i=1}^{v} \beta_{ii} x_{ui}^2 + \sum_{i<j}^{v} \beta_{ij} x_{ui} x_{uj} + \boldsymbol{\varepsilon} \qquad (6.4.1)$$

It can be expressed in matrix form as

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_u \\ \vdots \\ Y_N \end{bmatrix} = \left[\begin{array}{c|cccc|cccc|ccc} 1 & x_{11} & x_{12} & \cdots & x_{1v} & x_{11}^2 & x_{12}^2 & \cdots & x_{1v}^2 & x_{11}x_{12} & \cdots & x_{1v-1}x_{1v} \\ 1 & x_{21} & x_{22} & \cdots & x_{2v} & x_{21}^2 & x_{22}^2 & \cdots & x_{2v}^2 & x_{21}x_{22} & \cdots & x_{2v-1}x_{2v} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{u1} & x_{u2} & \cdots & x_{uv} & x_{u1}^2 & x_{u2}^2 & \cdots & x_{uv}^2 & x_{u1}x_{u2} & \cdots & x_{uv-1}x_{uv} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{N1} & x_{N2} & \cdots & x_{Nv} & x_{N1}^2 & x_{N2}^2 & \cdots & x_{Nv}^2 & x_{N1}x_{N2} & \cdots & x_{Nv-1}x_{Nv} \end{array}\right] \left[\begin{array}{c} \beta_0 \\ \hline \beta_1 \\ \beta_2 \\ \vdots \\ \beta_v \\ \hline \beta_{11} \\ \beta_{22} \\ \vdots \\ \beta_{vv} \\ \hline \beta_{12} \\ \beta_{13} \\ \vdots \\ \beta_{v-1v} \end{array}\right] + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \vdots \\ \vdots \\ \varepsilon_N \end{bmatrix}$$

$$\underline{Y} = X\, \underline{\boldsymbol{\beta}} + \underline{\boldsymbol{\varepsilon}} \qquad (6.4.2)$$

Where,

$\underline{\mathbf{Y}} = (Y_1, Y_2, \dots, Y_N)'$ is the vector of observations,

$\underline{\boldsymbol{\varepsilon}} = (\varepsilon_1, \varepsilon_2, \dots \varepsilon_N)'$ is the vector of random errors and assume that $\underline{\boldsymbol{\varepsilon}} \sim N(0, \sigma^2 I)$.

$X_u = (1, x_{u1}, x_{u2} \dots x_{uv}, x^2_{u1}, x^2_{u2} \dots x^2_{uv}, x_{ul}x_{u2} \dots x_{uv-1}x_{uv})$ is the u^{th} row of X

$\underline{\beta} = (\beta_0, \beta_1, \beta_2 \dots \beta_v, \beta_{11}, \beta_{22} \dots \beta_{vv}, \beta_{l2} \dots \beta_{v-1v})'$ is the vector of parameters.

Let the total number of terms in the model be k $(= 2v + {}^vC_2)$.

The likelihood function of an observed sample Y with the given parameters β and σ^2 is

$$L(Y/\beta, \sigma^2) = (2\pi\sigma^2)^{-\frac{N}{2}} \exp\left\{ -(2\sigma^2)^{-1} \sum_{i=1}^{N} (Y_i - X_i\beta)'(Y_i - X_i\beta) \right\} \qquad (6.4.3)$$

$$L(Y, \beta, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{N} (Y_i - X_i\beta)'(Y_i - X_i\beta) \right\}$$

$$\Rightarrow L(Y,\beta,\sigma^2)=\frac{1}{(2\pi\sigma^2)^{N/2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{N}[(Y_i-X_i\hat{\beta}+X_i\hat{\beta}-X_i\beta)'(Y_i-X_i\hat{\beta}+X_i\hat{\beta}-X_i\beta)]\right\}$$

$$\Rightarrow L(Y/\beta,\sigma^2)=\frac{1}{(2\pi\sigma^2)^{N/2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{N}[(Y_i-X_i\hat{\beta})'(Y_i-X_i\hat{\beta})+(\beta-\hat{\beta})'X'X(\beta-\hat{\beta})]\right\}$$

$$\Rightarrow L(Y/\beta,\sigma^2)=(2\pi\sigma^2)^{-\frac{N}{2}}\exp\left\{-(2\sigma^2)^{-1}\sum_{i=1}^{N}[(Y_i-X_i\hat{\beta})'(Y_i-X_i\hat{\beta})+(\beta-\hat{\beta})'X'X(\beta-\hat{\beta})]\right\}$$

(6.4.4)

$\frac{\partial L}{\partial\beta}\Rightarrow\hat{\beta}=(X'X)^{-1}X'Y$ and $\frac{\partial L}{\partial\sigma^2}\Rightarrow\hat{\sigma}^2=\frac{(Y-X\beta)'(Y-X\beta)}{N-k}$ where k = 2v + $^vC_2$ (6.4.5)

Then, the distribution of the parameters is:

$$\Rightarrow L(Y/\beta,\sigma^2)=(2\pi\sigma^2)^{-\frac{N}{2}}\exp\left\{-(2\sigma^2)^{-1}\sum_{i=1}^{N}[(Y_i-X_i\hat{\beta})'(Y_i-X_i\hat{\beta})+(\beta-\hat{\beta})'X'X(\beta-\hat{\beta})]\right\}$$

$$=(2\pi\sigma^2)^{-N/2}\exp\left\{-\frac{1}{2\sigma^2}\left[(Y-X\hat{\beta})^2+\sum_{i=1}^{N}X_i^2(\beta-\hat{\beta})^2\right]\right\}$$

$$=(2\pi\sigma^2)^{-N/2}\exp\left\{-\frac{1}{2\sigma^2}\left[\hat{\sigma}^2(N-k)+\sum_{i=1}^{N}X_i^2(\beta-\hat{\beta})^2\right]\right\}$$

$$=(2\pi\sigma^2)^{-N/2}.\exp\left\{\frac{-\hat{\sigma}^2(N-k)}{2\sigma^2}\right\}.\exp\left\{-\frac{1}{2\sigma^2/\Sigma X_i^{\,2}}(\beta-\hat{\beta})^2\right\}$$

$$=(\sqrt{2\pi})\sigma^{-1}\exp\left\{-\frac{1}{2\sigma^2/\Sigma X_i^{\,2}}(\beta-\hat{\beta})^2\right\}.(\sigma^2)^{-(N-1)/2}.\exp\left\{\frac{-\hat{\sigma}^2(N-k)}{2\sigma^2}\right\}$$

Then the posterior distribution of the parameter β is

$$P(\beta/X,Y)\propto\sigma^{-N-1}\exp\left\{-\frac{1}{2\sigma^2}\left[(Y-X\hat{\beta})'(Y-X\hat{\beta})+(\beta-\hat{\beta})'X'X(\beta-\hat{\beta})\right]\right\}$$

$$\propto \quad \sigma^{-N-1}\exp\ \{-\frac{1}{2\sigma^2}[\,(N-k)\hat{\sigma}^2 + (\beta-\hat{\beta})'\,X'X\,(\,\beta-\hat{\beta})\,]\}$$

since $(N-k)\,\hat{\sigma}^2 = (Y-X\hat{\beta})'\,(Y-X\hat{\beta})\,)$

$$\propto \quad \sigma^{-N-1}\exp\{-\frac{1}{2\sigma^2}(\beta-\hat{\beta})'\,X'X\,(\beta-\hat{\beta})\ \}$$

$$\sim \mathrm{MVN}\,(\hat{\beta},(X'X)^{-1}\sigma^2) \qquad (6.4.5)$$

And, the posterior distribution of the parameter σ^2 is

$$P\,(\sigma^2 / X,Y) \quad \propto \sigma^{-n-1}\exp(-\frac{1}{2\sigma^2}\{(N-k)\hat{\sigma}^2 + (\beta-\hat{\beta})'X'X(\beta-\hat{\beta})\})$$

$$\propto \sigma^{-(n-k+1)}\exp(-\frac{1}{2\sigma^2}(N-k)\hat{\sigma}^2)$$

$$\sim IG\,(\frac{(N-k)}{2},\frac{(N-k)\hat{\sigma}^2}{2})\quad \text{where } k = 2v + {}^{v}C_2. \qquad (6.4.6)$$

The method for reduction of second order response surface design model in case of without restrictions and restrictions towards rotatability on moment matrix are illustrated in the examples 6.4.1 and 6.4.2 are presented below.

EXAMPLE 6.4.1: Consider the experimental data considered in the example 5.3.1 with four factors in 16 design points. Using R the posteriors for all the 16,384 possible linear combination models are evaluated and the highest posterior probability is presented in Table 5.4.1. The Bayesian estimated parameters values are: $\hat{\beta}_0$ = 74.31; $\hat{\beta}_1$ = 5; $\hat{\beta}_2$ = 3.149; $\hat{\beta}_3$ = 3.974; $\hat{\beta}_4$ = 1.86; $\hat{\beta}_{12}$ = 0.112; $\hat{\beta}_{13}$ = -2.089; $\hat{\beta}_{14}$ = -1.725; $\hat{\beta}_{23}$ = - 1.366; $\hat{\beta}_{24}$ = 1.349; $\hat{\beta}_{34}$ = 0.326;

Table 6.4.1

S.No.	Parameters	Posterior Mean	Posterior S.D	P≠0
1	β_0	74.313	1.379	1.00
2	β_1	5	1.379	1.00

3	β_2	3.033	1.469	0.96
4	β_3	3.969	1.386	0.99
5	β_4	1.030	1.360	0.55
6	β_{11}	-0.181	0.746	0.11
7	β_{22}	-0.175	0.738	0.11
8	β_{33}	-0.169	0.729	0.11
9	β_{44}	-0.162	0.722	0.11
10	β_{12}	-0.025	0.550	0.11
11	β_{13}	-1.042	1.505	0.66
12	β_{14}	-0.531	1.283	0.48
13	β_{23}	-0.229	1.007	0.33
14	β_{24}	0.273	0.944	0.32
15	β_{34}	-0.032	0.504	0.12

The resulting reduced Bayesian model is $Y = 74.312 + 5\ X_1 + 3.15\ X_2 + 3.975\ X_3 + 1.862\ X_4 - 2.088\ X_{13} - 1.725\ X_{14}$, with error sum of squares 231.493 with 9 degrees of freedom and with an R^2 value is 0.810 selected with highest probability value 0.025 with six variables.

EXAMPLE 6.4.2: Consider the experimental data considered in the example 5.3.2 with three factors in 11 design points. Using R the posteriors for all the 512 possible linear combination models are evaluated and the highest posterior probability is presented in Table 5.4.2. The Bayesian estimated parameters values are: $\hat{\beta}_0 = 50.71$; $\hat{\beta}_1 = 5.001$; $\hat{\beta}_2 = 10$; $\hat{\beta}_3 = 0.499$; $\hat{\beta}_{33} = -0.249$; $\hat{\beta}_{12} = -1.50$; $\hat{\beta}_{13} = -0.501$; $\hat{\beta}_{23} = -0.999$;

Table 6.4.2

S.No.	Parameters	Posterior Mean	Posterior S.D	$P \neq 0$
1	β_0	50.813	1.589	1.00
2	β_1	5.446	1.519	1.00
3	β_2	10.446	1.519	1.00

4	β_3	0.346	1.006	0.30
5	β_{11}	0.037	1.032	0.16
6	β_{22}	0.034	1.019	0.16
7	β_{33}	0.031	1.006	0.16
8	β_{12}	-0.413	1.069	0.32
9	β_{13}	-0.011	0.725	0.19
10	β_{23}	-0.159	0.821	0.22

The resulting reduced Bayesian model is $Y = 50.906 + 5.484\, X_1 + 10.484\, X_2$, with error sum of squares 98.906 with 8 degrees of freedom and with an R^2 value is 0.895 selected with highest probability value 0.128 with two variables.

EXAMPLE 6.4.3: Consider the experimental data considered in the example 5.3.3 with four factors in 25 design points (CCD). Using R the posteriors for all the 16,384 possible linear combination models are evaluated and the highest posterior probability is presented in Table 5.4.3. The Bayesian estimated parameters values are: $\hat{\beta}_0 = 20.87$; $\hat{\beta}_1 = 1.350$; $\hat{\beta}_2 = 2.435$, $\hat{\beta}_3 = 2.215$, $\hat{\beta}_4 = 1.317$, $\hat{\beta}_{11} = 0.228$, $\hat{\beta}_{22} = -1.64$, $\hat{\beta}_{33} = -0.893$, $\hat{\beta}_{44} = -0.483$, $\hat{\beta}_{12} = 0.751$, $\hat{\beta}_{13} = 0.299$, $\hat{\beta}_{14} = 0.174$, $\hat{\beta}_{23} = 0.599$, $\hat{\beta}_{24} = 0.475$, $\hat{\beta}_{34} = -0.076$.

Table 6.4.3

S.No.	Parameters	Posterior Mean	Posterior S.D	$P \neq 0$
1	β_0	21.525	0.443	1.00
2	β_1	0.689	0.668	0.52
3	β_2	2.691	0.161	1.00
4	β_3	2.114	0.161	1.00
5	β_4	1.260	0.161	1.00

6	β_{11}	0.323	0.282	0.76
7	β_{22}	- 1.579	0.255	1.00
8	β_{33}	-1.529	0.255	1.00
9	β_{44}	-0.929	0.255	1.00
10	β_{12}	0.750	0.180	1.00
11	β_{13}	0.232	0.200	0.77
12	β_{14}	0.057	0.131	0.32
13	β_{23}	0.600	0.180	1.00
14	β_{24}	0.475	0.180	1.00
15	β_{34}	-0.012	0.077	0.15

The resulting reduced Bayesian model is $Y = 21.446 + 1.318 X_1 + 2.691 X_2 + 2.114 X_3 + 1.261 X_4 + 0.421 X_{11} - 1.58 X_{22} - 1.53 X_{33} - 0.929 X_{44} + 0.75 X_{12} + 0.30 X_{13} + 0.60 X_{23} + 0.475 X_{24}$, with error sum of squares 5.749 with 12 degrees of freedom and with an R^2 value is 0.985 selected with highest probability value 0.144 with twelve variables.

6.5 COMPARISON OF BAYESIAN SELECTION MODEL WITH CLASSICAL SELECTION MODELS & THEIR ANALYSIS:

In this section an attempt is made to compare the reduced First order and Second Order Response Surface Design Models in Bayesian approach with Step-wise approach, Forward approach, Backward elimination approach and Nested approach and confidence intervals for Estimated parameters and the R^2 are presented.

6.5.1 FIRST ORDER RSD MODEL IN EXAMPLE 6.2.1

a) The data presented in the example 6.2.1 is pertaining to orthogonal design of First Order Response Surface Model. The estimated values for the parameters and their Confidence Intervals at 95% level are presented in the Table 6.5.1.

Table 6.5.1

S.No.	Parameters	Full model	95% Confidence		Significance Value
			Lower	Upper	
1	β_0	30.313	27.744	32.881	26.295
2	β_1	5.563	2.994	8.131	4.825
3	β_2	16.937	14.369	19.506	14.693
4	β_3	5.437	2.869	8.006	4.717
5	β_4	0.438	-2.131	3.006	0.380
6	β_5	0.312	-2.256	2.881	0.271

b) The estimated values for parameters in case of full model and reduced models in different approaches is presented in Table 6.5.2

Table 6.5.2

S.No.	Parameters	Full Model	Stepwise	Forward	Backward	Nested	Bayesian
1	β_0	30.313	30.313	30.313	30.313	30.313	30.313
2	β_1	5.563	5.563	5.563	5.563	5.563	5.563
3	β_2	16.937	16.937	16.937	16.937	16.937	16.937
4	β_3	5.437	5.437	5.437	5.437	5.437	5.437
5	β_4	0.438	-	-	-	-	-
6	β_5	0.312	-	-	-	-	-

c) An analysis in case of full and reduced models is presented in Table 6.5.3.

Table 6.5.3

	Full model	Stepwise	Forward	Backward	Nested	Bayesian
SSR	5562.813	5558.19	5558.19	5558.19	5558.188	5558.188
ESS	212.625	217.25	217.25	217.25	217.25	217.25
MSE	21.262	18.104	18.104	18.104	18.104	18.104
R^2 value	0.963	0.962	0.962	0.962	0.962	0.962

Note: The reduced model in all the approaches is same and the mean square due to residual and R^2 values are same. Cumulative Posterior Probability value for best four models is 1

6.5.2 FIRST ORDER RSD MODEL IN EXAMPLE 6.2.2

a) The data presented in the example 6.2.2 is pertaining to non-orthogonal design of First Order Response Surface Model. The estimated values for the parameters and their Confidence Intervals at 95% level are presented in the Table 6.5.4

Table 6.5.4

S.No.	Parameters	Full model	95% Confidence		Significance Value
			Lower	Upper	
1	β_0	0.256	-6.110	6.621	0.085
2	β_1	-12.996	-15.164	-10.827	-12.644
3	β_2	-9.5	-11.581	-7.419	-9.633
4	β_3	-1.389	-3.470	0.692	-1.409
5	β_4	-1.111	-3.192	0.969	-1.127
6	β_5	1.611	-0.470	3.691	1.633
7	β_6	0.055	-2.024	2.136	0.056
8	β_7	2.666	0.586	4.747	2.704
9	β_8	-0.611	-2.692	1.470	-0.619
10	β_9	-1.166	-3.247	0.914	-1.183

b) The estimated values for parameters in case of full model and reduced models in different approaches is presented in Table 6.5.5

Table 6.5.5

S.No.	Parameters	Full Model	Stepwise	Forward	Backward	Nested	Bayesian
1	β_0	0.256	-2.408	-2.408	-2.408	-2.034	-2.408
2	β_1	-12.996	-12.94	-12.94	-12.94	-13.343	-12.94
3	β_2	-9.5	-9.503	-9.503	-9.503	-9.481	-9.503
4	β_3	-1.389	-	-	-	-	-
5	β_4	-1.111	-	-	-	-	-
6	β_5	1.611	-	-	-	-	-
7	β_6	0.055	-	-	-	-	-
8	β_7	2.666	2.663	2.663	2.663	2.686	2.663
9	β_8	-0.611	-	-	-	-	-
10	β_9	-1.166	-	-	-	-	-

c) An analysis in case of full and reduced models is presented in Table 6.5.6

Table 6.5.6

	Full model	Stepwise	Forward	Backward	Nested	Bayesian
SSR	4905.37	4770.49	4770.49	4770.49	4770.49	4770.49
ESS	296.629	431.513	431.513	431.513	431.513	431.513
MSE	17.449	18.761	18.761	18.761	18.761	18.761
R^2 value	0.943	0.917	0.917	0.917	0.917	0.917

Note: The reduced model in all the approaches is same and the mean square due to residual and R^2 values are same. Cumulative Posterior Probability value for best five models is 0.3733

6.5.3 SECOND ORDER RSD MODEL IN EXAMPLE 6.4.1

a) The data presented in the example 6.3.1 is pertaining to second order Response Surface Model. The estimated values for the parameters and their Confidence Intervals at 95% level are presented in the Table 6.5.7

Table 6.5.7

S.No.	Parameters	Full model	95% Confidence		Significance Value
			Lower	Upper	
1	β_0	74.313	70.557	78.068	50.868
2	β_1	5.000	1.245	8.755	3.423
3	β_2	3.15	-0.605	6.905	2.156
4	β_3	3.975	0.220	7.730	2.721
5	β_4	1.862	-1.893	5.618	1.275
6	β_{12}	0.113	-3.643	3.868	0.077
7	β_{13}	-2.088	-5.83	1.668	-1.429
8	β_{14}	-1.725	-5.480	2.030	-1.81
9	β_{23}	-1.363	-5.118	2.393	-0.933
10	β_{24}	1.35	-2.405	5.105	0.924
11	β_{34}	-0.325	-4.080	3.430	-0.222

b) The estimated values for parameters in case of full model and reduced models in different approaches is presented in Table 6.5.8

Table 6.5.8

S.No.	Parameters	Full Model	Stepwise	Forward	Backward	Nested	Bayesian
1	β_0	74.313	74.313	74.313	74.313	74.313	74.313
2	β_1	5.000	5	5	5	5	5
3	β_2	3.15	-	-	3.15	3.15	3.15
4	β_3	3.975	3.975	3.975	3.975	3.975	3.975
5	β_4	1.862	-	-	-	-	1.862
6	β_{12}	0.113	-	-	-	-	-
7	β_{13}	-2.088	-	-	-	-	-2.088
8	β_{14}	-1.725	-	-	-	-	-1.725
9	β_{23}	-1.363	-	-	-	-	-
10	β_{24}	1.35	-	-	-	-	-
11	β_{34}	-0.325	-	-	-	-	-

c) An analysis in case of full and reduced models is presented in Table 6.5.9

Table 6.5.9

	Full model	Stepwise	Forward	Backward	Nested	Bayesian
SSR	1045.16	652.81	652.81	811.57	811.57	984.405
ESS	170.738	563.088	563.088	404.328	404.328	231.493
MSE	34.148	43.314	43.314	33.694	33.694	25.721
R^2 value	0.860	0.537	0.537	0.667	0.667	0.810

Note: The reduced model in all the approaches is not same and the mean square error values for full and Bayesian reduced models are differ. It can be noted that the values of parameters β_4, β_{11}, β_{22}, β_{33}, β_{44}, β_{12}, β_{13}, β_{14}, β_{23}, β_{24}, β_{34} are insignificant in nested and backward approaches whereas in Bayesian β_{11}, β_{22}, β_{33}, β_{44}, β_{12}, β_{23}, β_{24}, β_{34} are insignificant and in stepwise and forward β_2, β_4, β_{11}, β_{22}, β_{33}, β_{44}, β_{12}, β_{13}, β_{14}, β_{23}, β_{24}, β_{34} are insignificant. Cumulative Posterior Probability value for best five models is 0.1051.

6.5.3 SECOND ORDER RSD MODEL IN EXAMPLE 6.4.2

a) The data presented in the example 6.3.2 is pertaining to second order Response Surface Model. The estimated values for the parameters and their Confidence Intervals at 95% level are presented in the Table 6.5.10.

Table 6.5.10

S.No.	Parameters	Full model	95% Confidence		Significance Value
			Lower	Upper	
1	β_0	50.75	42.597	58.902	19.811
2	β_1	5	-3.152	13.152	1.952
3	β_2	10	1.847	18.152	3.904
4	β_3	0.5	-7.652	8.652	0.195

5	β_{33}	-0.25	-11.779	11.279	-0.069
6	β_{12}	-1.5	-9.652	6.652	-0.586
7	β_{13}	-0.5	-8.652	7.652	-0.195
8	β_{23}	-1	-9.152	7.152	-0.390

b) The estimated values for parameters in case of full model and reduced models in different approaches is presented in Table 6.5.11

Table 6.5.11

S.No.	Parameters	Full Model	Stepwise	Forward	Backward	Nested	Bayesian
1	β_0	50.75	50.906	50.906	50.906	50.815	50.906
2	β_1	5	5.484	5.484	5.484	5.348	5.484
3	β_2	10	10.484	10.484	10.484	9.618	10.484
4	β_3	0.5	-	-	-	-	-
5	β_{33}	-0.25	-	-	-	-	-
6	β_{12}	-1.5	-	-	-	-	-
7	β_{13}	-0.5	-	-	-	-	-
8	β_{23}	-1	-	-	-	-	-

c) An analysis in case of full and reduced models is presented in table 6.5.12

Table 6.5.12

	Full model	Stepwise	Forward	Backward	Nested	Bayesian
SSR	861.977	841.821	841.821	841.821	841.821	841.821

ESS	78.75	98.906	98.906	98.906	98.906	98.906
MSE	26.25	12.363	12.363	12.363	12.363	12.363
R^2 value	0.916	0.895	0.895	0.895	0.895	0.895

Note: The reduced models in all the approached is same and the mean square due to residual and R^2 values are same. Cumulative Posterior Probability value for best five models is 0.3502.

6.5.4 SECOND ORDER RSD MODEL IN EXAMPLE 6.4.3

a) The data presented in the example 6.4.3 is pertaining to second order Central Composite Response Surface Design Model. The estimated values for the parameters and their Confidence Intervals at 95% level are presented in the Table 6.5.13

Table 6.5.13

S.No.	Parameters	Full model	95% Confidence		Significance Value
			Lower	Upper	
1	β_0	21.447	20.486	22.408	49.725
2	β_1	1.318	0.960	1.676	8.197
3	β_2	2.691	2.332	3.049	16.732
4	β_3	2.114	1.755	2.472	13.144
5	β_4	1.26	0.902	1.619	7.839
6	β_{11}	0.421	-0.146	0.988	1.655
7	β_{22}	-1.58	-2.147	-1.013	-6.213
8	β_{33}	-1.53	-2.097	-0.963	-6.016
9	β_{44}	-0.93	-1.496	-0.363	-3.656
10	β_{12}	0.75	0.349	1.151	4.172
11	β_{13}	0.3	-0.101	0.701	1.669
12	β_{14}	0.175	-0.226	0.576	0.973
13	β_{23}	0.6	0.199	1.001	3.337
14	β_{24}	0.475	0.074	0.876	2.642
15	β_{34}	-0.075	-0.476	0.326	-0.417

b) The estimated values for parameters in case of full model and reduced models in different approaches is presented in Table 6.5.14

Table 6.5.14

S.No.	Parameters	Full Model	Stepwise	Forward	Backward	Nested	Bayesian
1	β_0	21.447	21.783	21.783	21.783	21.447	21.447
2	β_1	1.318	1.318	1.318	1.318	1.318	1.318
3	β_2	2.691	2.691	2.691	2.691	2.691	2.691
4	β_3	2.114	2.114	2.114	2.114	2.114	2.114
5	β_4	1.26	1.26	1.26	1.26	1.26	1.26
6	β_{11}	0.421	-	-	-	0.421	0.421
7	β_{22}	-1.58	-1.58	-1.58	-1.58	-1.58	-1.58
8	β_{33}	-1.53	-1.53	-1.53	-1.53	-1.53	-1.53
9	β_{44}	-0.93	-0.93	-0.93	-0.93	-0.93	-0.93
10	β_{12}	0.75	0.75	0.75	0.75	0.75	0.75
11	β_{13}	0.3	-	-	-	0.3	0.3
12	β_{14}	0.175	-	-	-	-	-
13	β_{23}	0.6	0.6	0.6	0.6	0.6	0.6
14	β_{24}	0.475	0.475	0.475	0.475	0.475	0.475
15	β_{34}	-0.075	-	-	-	-	-

c) An analysis in case of full and reduced models is presented in table 6.5.15

Table 6.5.15

	Full model	Stepwise	Forward	Backward	Nested	Bayesian
SSR	368.053	364.618	364.618	364.618	367.473	367.473
ESS	5.169	8.604	8.604	8.604	5.749	5.749
MSE	0.517	0.615	0.615	0.615	0.479	0.479
R^2 value	0.986	0.977	0.977	0.977	0.985	0.985

Note: The value of the parameters β_{11}, β_{13}, β_{14}, β_{23} are insignificant in stepwise, Forward and Backward approaches whereas in Nested and Bayesian β_{11}, β_{14}, β_{23} are insignificant. The reduced model in Nested and Bayesian approaches are same. The mean square error values and R^2 values are also same in proposed methods. Cumulative Posterior Probability value for best five models is 0.5096.

6.6 PROPERTIES OF BAYESIAN ESTIMATED PARAMETER

An attempt is made to study the properties of Bayes estimated parameter like Mean, variance, consistency and sufficiency of estimator are derived for a design model and presented below.

THEOREM 6.6.1: In a linear design model, Bayes estimator of a parameter $\hat{\beta}_i$ is an unbiased estimator i.e. $E(\hat{\beta}) = \beta$

Proof: The vector of estimated parameter follows multivariate normal distribution i.e. $\hat{\beta} \sim MVN(\beta, (X'_{N\text{-}m} X_{N\text{-}m})^{-1}\hat{\sigma}^2)$. Then β_i is

$$f(\hat{\beta}_i) = \left(\sqrt{\frac{1}{2\pi\sigma^2 / \Sigma X_i^2}}\right) \exp\left\{-\frac{1}{2\sigma^2/\Sigma X_i^{\,2}}(\hat{\beta}_i - \beta_i)^2\right\}$$

$$E(\hat{\beta}_i) = \left(\frac{1}{\sqrt{2\pi\sigma^2/\Sigma X_i^2}}\right) \int_{-\infty}^{\infty} \hat{\beta}_i \exp\left\{-\frac{1}{2\sigma^2/\Sigma X_i^{\,2}}(\hat{\beta}_i - \beta_i)^2\right\} d\hat{\beta}_i$$

Let $\sigma z = (\hat{\beta}_i - \beta_i)\sqrt{\Sigma X_i^2}$ then

$$E(\hat{\beta}_i) = \left(\frac{1}{\sqrt{2\pi\sigma^2/\Sigma\Sigma_i^{\,2}}}\right) \int_{-\infty}^{\infty} \left(\frac{\sigma z}{\sqrt{\Sigma X_i^2}} + \beta_i\right) \exp\left\{-\frac{z^2}{2}\right\} \frac{\sigma\, dz}{\sqrt{\Sigma X_i^2}}$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right) \int_{-\infty}^{\infty} \left(\frac{z}{\sqrt{\Sigma X_i^2}}\right) \exp\left\{-\frac{z^2}{2}\right\} dz + \left(\frac{1}{\sqrt{2\pi}}\right) \int_{-\infty}^{\infty} \beta_i \exp\left\{-\frac{z^2}{2}\right\} dz$$

$$= \beta_i \tag{6.6.1}$$

THEOREM 6.6.2: The Variance of Bayes estimator of parameter $\hat{\beta}_i$ is Var $(\hat{\beta}_i) = \sigma^2 / \Sigma X_i^2$

Proof: The second non-central moment of β is

$$E(\hat{\beta}_i^2) = \left(\frac{1}{\sqrt{2\pi\sigma^2/\Sigma X_i^2}}\right) \int_{-\infty}^{\infty} \hat{\beta}_i^2 \exp\left\{-\frac{1}{2\sigma^2/\Sigma X_i^2}(\hat{\beta}_i - \beta_i)^2\right\} d\hat{\beta}_i$$

Let $\sigma z = (\hat{\beta}_i - \beta_i)\sqrt{\Sigma X_i^2}$ then

$$E(\hat{\beta}_i^2) = \left(\frac{1}{\sqrt{2\pi\sigma^2/\Sigma X_i^2}}\right) \int_{-\infty}^{\infty} \left(\frac{\sigma z}{\sqrt{\Sigma X_i^2}} + \beta_i\right)^2 \exp\left\{-\frac{z^2}{2}\right\} \frac{\sigma\, dz}{\sqrt{\Sigma X_i^2}}$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right) \int_{-\infty}^{\infty} \frac{\sigma z}{\sqrt{\Sigma X_i^2}} \exp\left\{-\frac{z^2}{2}\right\} dz + \left(\frac{1}{\sqrt{2\pi}}\right) \int_{-\infty}^{\infty} \beta_i^2 \exp\left\{-\frac{z^2}{2}\right\} dz + \int_{-\infty}^{\infty} \frac{2}{\sqrt{2\pi}} \frac{\sigma z \beta_i}{\sqrt{\Sigma X_i^2}} \exp\left\{-\frac{z^2}{2}\right\} dz$$

$$= \frac{\sigma^2}{\Sigma X_i^2} + \beta_i^2$$

Hence the variance of the estimator is Var $(\hat{\beta}_i) = \sigma^2 / \Sigma X_i^2$ (6.6.2)

THEOREM 6.6.3: Bayes estimator $\hat{\beta}_i$ is a consistent estimator.

Proof: Using Chebychev's inequality

$$P\left[|\hat{\beta}_i - \beta| \geq \varepsilon\right] < \frac{\mathrm{Var}(\hat{\beta}_i)}{\varepsilon^2} \text{ where Var } (\hat{\beta}_i) = \sigma^2/\Sigma X_i^2$$

$$\Rightarrow P\left[|\hat{\beta}_i - \beta| \geq \varepsilon\right] < \frac{\sigma^2/\Sigma X_i^2}{\varepsilon^2} = \frac{2\sigma^4}{\varepsilon^2 n} \text{ Because } V(\beta) = \frac{(\sigma^2)^2}{n^2} V(\chi^2_{(n)})$$

$$\Rightarrow P\left[|\hat{\beta}_i - \beta| \geq \varepsilon\right] \to 0 \text{ as } n \to \infty$$

Hence β is a consistent estimator for σ^2. (6.6.3)

THEOREM 6.6.4: Sufficient statistic of a Bayesian estimator β is $\sum X_i^2$.

Proof: $$f(\beta) = (\sqrt{2\pi})\sigma^{-1}\exp\left\{-\frac{1}{2\sigma^2/\Sigma X_i^{\,2}}\left(\hat{\beta}-\beta\right)^2\right\}$$

$$L(\beta,\sigma^2) = (\sqrt{2\pi})^{-(N-m)}\,\sigma^{-(N-m)}\exp\left\{-\frac{1}{2\sigma^2/\Sigma\ X_i^{\,2}}\left(\hat{\beta}-\beta\right)^2\right\}$$

$$\propto \quad \sigma^{-(N-m)}\exp\left\{-\frac{1}{2\sigma^2/\Sigma\ X_i^{\,2}}\left(\hat{\beta}-\beta\right)^2\right\}$$

$$\propto \quad \sigma^{-(N-m)}\exp\left\{-\frac{1}{2\sigma^2}\left[\Sigma X_i^{\,2}(\hat{\beta}-\beta)^2\right]\right\}$$

Hence the sufficient estimator for $\hat{\beta}$ is ΣX_i^2. (6.6.4)

6.7 REMARKS ON BAYESIAN SELECTION OF MODEL

1. In the maximum likelihood estimator, the analytical likelihood function and behavior of likelihood function are crucial.
2. The Maximum Likelihood estimator and Bayesian estimator will result same.
3. The goal of the Bayesian estimation is to characterize the posterior distribution of parameters, which is defined as the distribution of parameters conditional on data.
4. Bayesian methods don't require an analytical likelihood function.
5. When the prior is uniform, the posterior is the same as the likelihood (as a function of parameters). The tighter the prior, the less the posterior reflects information from the data.
6. The Bayesian estimator draws a large sample, which can be used to compute the parameters of interest using Gibbs sampling algorithm.

7. Bayesian approach does not provide any formulae to estimate missing values even in case of single missing observation.

8. It is difficult to estimate the missing values and parameters using Bayesian approach manually.

9. Bayesian approach depends on the priors and posteriors of the parameters and observations in the design model.

10. Bayes method is complicated when compared to least squares method because Bayes procedure depends on distribution for generation of samples.

11. The confidence intervals for the vector of parameters β with (1 − α) 100% confidence level can be evaluated as $\hat{\beta}_i \pm t_{\alpha/2, k}\sqrt{C_{ii}}$. Where C_{ii} can be obtained from $(X'X)^{-1}$ matrix.

12. The bias due to the estimated missing value can be obtained from the difference of two error sum of squares which are evaluated with and without imposing the condition H_0 $\alpha_i=0$.

13. Bayesian approach reduces the size of the original model by selecting the significant factors from the original set of variables based on the posterior probability values.

14. Even though the manual computation is very difficult and time consuming Bayesian approach gives same results when compared with other traditional approaches like Stepwise, Forward, Backward elimination and Nested.

15. In Bayesian approach, selecting the variables are more when compared to other methods as it includes the variables based on their probability value.

16. At initial point the estimation value of the parameter uses least square method.

7. SUMMARY OF FINDINGS AND FUTURE SCOPE

7.1 OBJECTIVES OF THE PROJECT

The objectives mentioned in the major research project proposal submitted to UGC entitled on “Reduction of Dimensionality of Response Surface Design Model - Bayesian Approach” are

1. Addressing the problems involved in the reduction of dimensionality in Response Surface Design model.
2. Deriving optimal methods to reduce the dimensionality of the Response Surface Design model with minimum loss in the information.
3. Construction of optimal response surface design model (first and second orders) and carrying out its analysis.
4. Examining the reduced response surface model for some specific application / experimental data.

7.2 SUMMARY OF FINDINGS

All the objectives mentioned in the proposal submitted to UGC entitled Reduction of Dimensionality of Response Surface Design Model-Bayesian Approach are achieved. The summarized conclusions on the research work presented in this thesis are,

1. A complete literature survey is conducted on the reduction of dimensionality of linear regression model and complete literature survey on the reduction of Response Surface

Design Model. All the methods are studied and tested with suitable examples and evaluated their analysis.

2. It can be noted that even though there are several methods available for selecting the best model, most of the research is on selecting the best regression model but no specific method was developed for the reduction of dimensionality of response surface design model.

3. This Project report provides three new approaches to reduce the dimensionality of response surface design model, one is using variance component indices, second one is using nested approach and the third one is in Bayesian approach.

4. The variance component indices are derived for first and second order models which are used for ranking to reduce the size of the response surface design models under with and without restrictions on the moment matrix to reduce the dimensionality of the model.

5. Nested Approach provides best selection model when compared with Step-wise, Forward and Back ward approaches. Nested Approach is avoids the problem of Multi-collinearity.

6. Derived the probability distributions for priors and posterior of parameters to estimate the parameters and evaluated posterior probabilities for all possible models and selected best model based on maximum posterior probability in case of first and second order response surface design models to reduce the dimensionality of the model.

7. Bayesian approach reduces the size of the original model by selecting the significant factors from the original set of variables based on the posterior probability values. It provides more efficient result for selecting the optimal model when compared with all the

classical methods like, All possible, Forward, Backward and Step-wise procedures and Nested approach.

8. The properties for the Bayesian estimated parameters like variance, unbiasedness, consistency, sufficiency are derived. An analysis of Bayesian selected model and confidence interval for the parameters are also presented. Even though the manual computation is difficult and time consuming Bayesian approach provides same / better results when compared with other traditional approaches.

9. In Bayesian approach, is a simulated process and depends on distribution for generation of samples. Selecting the variables are more when compared to other methods as it includes the variables based on their probability value and it depends on the priors and posteriors of the parameters and observations in the design model.

10. The approaches compared with respect to estimated values for parameters, their confidence intervals, reduced models, Mean square error and R^2 values are presented.

7.3 FUTURE SCOPE

Model selection plays a major role in the multi-factor experiments. Dimension Reduction refers to the process of converting a set of data having vast dimensions into data with lesser dimensions ensuring that it conveys similar information concisely. These techniques are typically used while solving machine learning problems to obtain better features for a classification or regression task. It can reduce ‘n’ dimensions of data set to ‘k’ dimensions ($k < n$). These k dimensions can be directly identified (filtered) or can be a combination of dimensions (weighted averages of dimensions) or new dimension(s) that represent existing multiple dimensions well.

APPENDIX

1 R code to find the initial parameters corresponding to the data presented

```
y=matrix(c(8,9,34,52,16,22,45,60,6,10,30,50,15,21,44,63),16,1);
x=matrix(c(-1,1,-1,1,-1,1,-1,1,-1,1,-1,1,-1,1,-1,1,
           -1,-1,1,1,-1,-1,1,1,-1,-1,1,1,-1,-1,1,1,
           -1,-1,-1,-1,1,1,1,1,-1,-1,-1,-1,1,1,1,1,
            1,1,1,1,1,1,1,1,-1,-1,-1,-1,-1,-1,-1,-1,
            1,-1,-1,1,-1,1,1,-1,-1,1,1,-1,1,-1,-1,1),16,5);
fit1<-lm(y~x);
summary(fit1);
confint(fit1,level=0.95);
                   #Load BMA package to find posterior values.
bicreg(x,y);
summary(bicreg(x,y))
```

2 R code to find the initial parameters corresponding to the data presented

```
y=matrix(c(5,2,8,-15,-6,-10,-28,-19,-23,-13,-17,-7,-23,-31,-23,-34,-37,-29,-27,-27,-30,-35,-
  35,-38,-39,-40,-41),27,1);
X=matrix(c(0,0,0,0,0,0,0,0,0,1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,2,2,2,
           0,0,0,1,1,1,2,2,2,0,0,0,1,1,1,2,2,2,0,0,0,1,1,1,2,2,2,
           0,0,0,1,1,1,2,2,2,1,1,1,2,2,2,0,0,0,2,2,2,0,0,0,1,1,1,
           0,0,0,1,1,1,2,2,2,2,2,2,0,0,0,1,1,1,1,1,1,2,2,2,0,0,0,
           0,1,2,0,1,2,0,1,2,0,1,2,0,1,2,0,1,2,0,1,2,0,1,2,0,1,2,
           0,1,2,0,1,2,0,1,2,1,2,0,1,2,0,1,2,0,2,0,1,2,0,1,2,0,1,
           0,1,2,0,1,2,0,1,2,2,0,1,2,0,1,2,0,1,1,2,0,1,2,0,1,2,0,
           0,1,2,2,0,1,1,2,0,0,1,2,2,0,1,1,2,0,0,1,2,2,0,1,1,2,0,
           0,1,2,2,0,1,1,2,0,1,2,0,0,1,2,2,0,1,2,0,1,1,2,0,0,1,2),27,9);
fit2<-lm(y~X);
summary(fit2);
confint(fit2,level=0.95);
bicreg(X,y);
summary(bicreg(X,y))
```

3 R code to find the initial parameters corresponding to the data presented

```
y=matrix(c(53.3,78,62.4,78.9,75.9,75.4,71.3,84.4,64.5,67.5,72.8,85.3,71.4,83.3,82.9,81.7),16,1);
X=matrix(c(-1,1,-1,1,-1,1,-1,1,-1,1,-1,1,-1,1,-1,1,
           -1,-1,1,1,-1,-1,1,1,-1,-1,1,1,-1,-1,1,1,
           -1,-1,-1,-1,1,1,1,1,-1,-1,-1,-1,1,1,1,1,
           -1,-1,-1,-1,-1,-1,-1,-1,1,1,1,1,1,1,1,1,
            1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,
            1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,
            1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,
            1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,
            1,-1,-1,1,1,-1,-1,1,1,-1,-1,1,1,-1,-1,1,
            1,-1,1,-1,-1,1,-1,1,1,-1,1,-1,-1,1,-1,1,
```

```
                  1,-1,1,-1,1,-1,1,-1,-1,1,-1,1,-1,1,-1,1,
                  1,1,-1,-1,-1,-1,1,1,1,1,-1,-1,-1,-1,1,1,
                  1,1,-1,-1,1,1,-1,-1,-1,-1,1,1,-1,-1,1,1,
                  1,1,1,1,-1,-1,-1,-1,-1,-1,-1,-1,1,1,1,1),16,14);
fit3<-lm(y~X);
summary(fit3);
confint(fit3,level=0.95);
bicreg(X,y);
summary(bicreg(X,y))
```

4 R code to find the initial parameters corresponding to the data presented

```
y=matrix(c(32,46,57,65,36,48,57,50,44,53,56),11,1);
X=matrix(c(-1,1,-1,1,-1,1,-1,0,0,0,0,
           -1,-1,1,1,-1,-1,1,0,0,0,0,
           -1,-1,-1,-1,1,1,1,0,0,0,0,
            1,1,1,1,1,1,1,0,0,0,0,
            1,1,1,1,1,1,1,0,0,0,0,
            1,1,1,1,1,1,1,0,0,0,0,
            1,-1,-1,1,1,-1,-1,0,0,0,0,
            1,-1,1,-1,-1,1,-1,0,0,0,0,
            1,1,-1,-1,-1,-1,1,0,0,0,0),11,9);
fit4<-lm(y~X);
summary(fit4);
confint(fit4,level=0.95);
```

5 R code to find the initial parameters corresponding to the data presented

```
y=matrix(c(27.6,16.6,15.4,17.4,17,19,17.4,12.6,18.6,22.4,21.4,14,24,15.6,13,14.4,22.6,23.4,
           20.6,22.6,13.4,20.6,15.6,21,17.6),25,1);
X=matrix(c(1,-1,1,-1,1,-1,1,-1,1,-1,1,-1,1,-1,1,-1,0,1.414,-1.414,0,0,0,0,0,0,
           1,-1,-1,1,-1,1,1,-1,-1,1,1,-1,1,-1,-1,1,0,0,0,1.414,-1.414,0,0,0,0,
           1,1,-1,-1,1,1,-1,-1,1,1,-1,-1,1,1,-1,-1,0,0,0,0,0,1.414,-1.414,0,0,
           1,1,1,1,-1,-1,-1,-1,1,1,1,1,-1,-1,-1,-1,0,0,0,0,0,0,0,1.414,-1.414,
           1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,0,1.999,1.999,0,0,0,0,0,0,
           1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,0,0,0,1.999,1.999,0,0,0,0,
           1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,0,0,0,0,0,1.999,1.999,0,0,
           1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,0,0,0,0,0,0,0,1.999,1.999,
           1,1,-1,-1,-1,-1,1,1,-1,-1,1,1,1,1,-1,-1,0,0,0,0,0,0,0,0,0,
           1,-1,-1,1,1,-1,-1,1,1,-1,-1,1,1,-1,-1,1,0,0,0,0,0,0,0,0,0,
           1,-1,1,-1,-1,1,-1,1,1,-1,1,-1,-1,1,-1,1,0,0,0,0,0,0,0,0,0,
           1,-1,1,-1,-1,1,-1,1,-1,1,-1,1,1,-1,1,-1,0,0,0,0,0,0,0,0,0,
           1,-1,-1,1,1,-1,-1,1,-1,1,1,-1,-1,1,1,-1,0,0,0,0,0,0,0,0,0,
           1,1,-1,-1,-1,-1,1,1,1,1,-1,-1,-1,-1,1,1,0,0,0,0,0,0,0,0,0),25,15);
fit5<-lm(y~X);
summary(fit5);
confint(fit5,level=0.95);
bicreg(X,y);
```

```
summary(bicreg(X,y))
```

6 Win-BUG Program to find the posterior estimates of parameters corresponding to the data presented

```
model
        {
          for(i in 1:16)
          {
            mu.y[i]<- beta0+beta1*x1[i]+beta2*x2[i]+beta3*x3[i]+beta4*x4[i]+beta5*x5[i]
            y[i] ~ dnorm(mu.y[i],prec)
            }
            mu ~ dnorm(30.3125,24.064325)
            beta0 ~ dnorm(30.313,24.064325)
            beta1 ~ dnorm(5.563, 24.064325)
            beta2 ~ dnorm(16.937,      24.064325)
            beta3 ~ dnorm(5.437, 24.064325)
            beta4 ~ dnorm(0.438, 24.064325)
            beta5 ~ dnorm(0.312, 24.064325)
            prec ~ dgamma(7.5,0.00036)
            s2<-1/prec
            }
            list(y=c(8,9,34,52,16,22,45,60,6,10,30,50,15,21,44,63),
            x1=c(-1,1,-1,1,-1,1,-1,1,-1,1,-1,1,-1,1,-1,1),
            x2=c(-1,-1,1,1,-1,-1,1,1,-1,-1,1,1,-1,-1,1,1),
            x3=c(-1,-1,-1,-1,1,1,1,1,-1,-1,-1,-1,1,1,1,1),
            x4=c(1,1,1,1,1,1,1,1,-1,-1,-1,-1,-1,-1,-1,-1),
            x5=c(1,-1,-1,1,-1,1,1,-1,-1,1,1,-1,1,-1,-1,1))
            list(beta0 = 30.313, beta1 = 5.563, beta2 = 16.937, beta3 = 5.437, beta4 = 0.438,
            beta5 = 0.312, prec = 0.00260, mu =30.3125)
```

Output:

Node	Posterior Mean	Posterior S.d	Posterior Median	Start	Sample
beta0	30.31	0.1951	30.31	1	100000
beta1	5.562	0.1944	5.563	1	100000
beta2	16.94	0.1949	16.94	1	100000
beta3	5.436	0.1942	5.437	1	100000
beta4	0.4381	0.1948	0.4388	1	100000
beta5	0.3126	0.1944	0.3123	1	100000
Mu	30.31	0.2036	30.31	1	100000
Prec	0.1432	0.0365	0.14	1	100000

7 Win-BUG Program to find the posterior estimates of parameters corresponding to the data presented

```
model
```

```
{
 for(i in 1:27)
 {
 mu.y[i]<-
  beta0+beta1*x1[i]+beta2*x2[i]+beta3*x3[i]+beta4*x4[i]+beta5*x5[i]+beta6*x6[i]+be
t
  a7*x7[i]+beta8*x8[i]+beta9*x9[i]
  y[i] ~ dnorm(mu.y[i],prec)
 }
mu ~ dnorm(-22.7,7.4074)
beta0 ~ dnorm(0.256,7.4074)
beta1 ~ dnorm(-12.996,4.3478)
beta2 ~ dnorm(-9.5,4.4444)
beta3 ~ dnorm(-1.389,4.4444)
beta4 ~ dnorm(-1.111,4.4444)
beta5 ~ dnorm(-1.611,4.4444)
beta6 ~ dnorm(0.055,4.4444)
beta7 ~ dnorm(2.666,4.4444)
beta8 ~ dnorm(-0.611,4.4444)
beta9 ~ dnorm(-1.166,4.4444)
prec ~ dgamma(13,0.00041)
s2<-1/prec
}
list(y=c(5,2,8,-15,-6,-10,-28,-19,-23,-13,-17,-7,-23,-31,-23,-34,-37,-29,-27,-27,-30,-35,-
35,-38,-39,-40,-41),
x1=c(0,0,0,0,0,0,0,0,0,1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,2,2,2),
x2=c(0,0,0,1,1,1,2,2,2,0,0,0,1,1,1,2,2,2,0,0,0,1,1,1,2,2,2),
x3=c(0,0,0,1,1,1,2,2,2,1,1,1,2,2,2,0,0,0,2,2,2,0,0,0,1,1,1),
x4=c(0,0,0,1,1,1,2,2,2,2,2,2,0,0,0,1,1,1,1,1,1,2,2,2,0,0,0),
x5=c(0,1,2,0,1,2,0,1,2,0,1,2,0,1,2,0,1,2,0,1,2,0,1,2,0,1,2),
x6=c(0,1,2,0,1,2,0,1,2,1,2,0,1,2,0,1,2,0,2,0,1,2,0,1,2,0,1),
x7=c(0,1,2,0,1,2,0,1,2,2,0,1,2,0,1,2,0,1,1,2,0,1,2,0,1,2,0),
x8=c(0,1,2,2,0,1,1,2,0,0,1,2,2,0,1,1,2,0,0,1,2,2,0,1,1,2,0),
x9=c(0,1,2,2,0,1,1,2,0,1,2,0,0,1,2,2,0,1,2,0,1,1,2,0,0,1,2))
list(beta0 = 0.256,beta1 = - 12.996,beta2 = - 9.5,beta3 = - 1.389, beta4 = -1.111, beta5
= 1.611, beta6 = 0.055, beta7 = 2.666, beta8 = - 0.611, beta9 = - 1.166, prec = 0.005,mu = - 22.7)
```

Output:

Node	Posterior Mean	Posterior S.d	Posterior Median	Start	Sample
beta0	0.4174	0.3517	0.4158	1	100000
beta1	-12.78	0.3742	-12.78	1	100000
beta2	-9.329	0.3682	-9.329	1	100000
beta3	-1.219	0.368	-1.219	1	100000
beta4	-0.9415	0.3663	-0.9412	1	100000
beta5	-0.3229	0.3954	-0.3183	1	100000

beta6	0.2252	0.3673	0.226	1	100000
beta7	2.835	0.3667	2.834	1	100000
beta8	-0.4328	0.365	-0.4329	1	100000
beta9	-0.984	0.366	-0.9865	1	100000
Mu	-22.7	0.3673	-22.7	1	100000
Prec	0.1337	0.0284	0.1315	1	100000

8. Win-BUG Program to find the posterior estimates of parameters corresponding to the data presented

```
model
{
for(i in 1:16)
{
mu.y[i]<-
beta0+beta1*x1[i]+beta2*x2[i]+beta3*x3[i]+beta4*x4[i]+beta5*x1x2[i]+beta6*x
1x3[i]+beta7*x1x4[i]+beta8*x2x3[i]+beta9*x2x4[i]+beta10*x3x4[i]
y[i] ~ dnorm(mu.y[i],prec)
}
mu ~ dnorm(74.3125,5.06624)
beta0 ~ dnorm(74.313,5.06624)
beta1 ~ dnorm(5,5.06624)
beta2 ~ dnorm(3.15,5.06624)
beta3 ~ dnorm(3.975,5.06624)
beta4 ~ dnorm(1.862,5.06624)
beta5 ~ dnorm(0.113,5.06624)
beta6 ~ dnorm(-2.088,5.06624)
beta7 ~ dnorm(-1.725,5.06624)
beta8 ~ dnorm(-1.363,5.06624)
beta9 ~ dnorm(1.35,5.06624)
beta10 ~ dnorm(-0.325,5.06624)
prec ~ dgamma(7.5,0.00191)
s2<-1/prec
}
list(y=c(53.3,78,62.4,78.9,75.9,75.4,71.3,84.4,64.5,67.5,72.8,85.3,71.4,83.3,82.9,81.7),
x1=c(-1,1,-1,1,-1,1,-1,1,-1,1,-1,1,-1,1,-1,1),
x2=c(-1,-1,1,1,-1,-1,1,1,-1,-1,1,1,-1,-1,1,1),
x3=c(-1,-1,-1,-1,1,1,1,1,-1,-1,-1,-1,1,1,1,1),
x4=c(-1,-1,-1,-1,-1,-1,-1,-1,1,1,1,1,1,1,1,1),
x1x2=c(1,-1,-1,1,1,-1,-1,1,1,-1,-1,1,1,-1,-1,1),
x1x3=c(1,-1,1,-1,-1,1,-1,1,1,-1,1,-1,-1,1,-1,1),
x1x4=c(1,-1,1,-1,1,-1,1,-1,-1,1,-1,1,-1,1,-1,1),
x2x3=c(1,1,-1,-1,-1,-1,1,1,1,1,-1,-1,-1,-1,1,1),
x2x4=c(1,1,-1,-1,1,1,-1,-1,-1,-1,1,1,-1,-1,1,1),
x3x4=c(1,1,1,1,-1,-1,-1,-1,-1,-1,-1,-1,1,1,1,1))
list(beta0 = 74.313, beta1 =5, beta2 =3.15, beta3 = 3.975, beta4 = 1.862, beta5 =
0.113, beta6 = -2.088, beta7 = -1.725, beta8 = - 1.363, beta9 = 1.35, beta10 =
```

0.325, prec = 0.01234,mu = 74.3125)

Output:

Node	Posterior Mean	Posterior S.d	Posterior Median	Start	Sample
beta0	74.31	0.3625	74.31	1	100000
beta1	5.00	0.3622	4.999	1	100000
beta2	3.149	0.364	3.15	1	100000
beta3	3.974	0.3637	3.975	1	100000
beta4	1.86	0.3626	1.86	1	100000
beta5	0.1124	0.3631	0.1139	1	100000
beta6	-2.089	0.3634	-2.09	1	100000
beta7	-1.725	0.3618	-1.725	1	100000
beta8	-1.366	0.3641	-1.365	1	100000
beta9	1.349	0.3637	1.349	1	100000
beta10	-0.3263	0.3633	-0.3249	1	100000
Mu	74.31	0.444	74.31	1	100000
Prec	0.1603	0.0416	0.1565	1	100000

9 Win-BUG Program to find the posterior estimates of parameters corresponding to the data presented

```
model
{
for(i in 1:11)
{
mu.y[i]<-
beta0+beta1*x1[i]+beta2*x2[i]+beta3*x3[i]+beta4*x3x3[i]+beta5*x1x2[i]+beta6*
x1x3[i]+beta7*x2x3[i]
y[i] ~ dnorm(mu.y[i],prec)
}
mu ~ dnorm(49.45455,8.55207)
beta0 ~ dnorm(50.750,8.55207)
beta1 ~ dnorm(5,23.51818)
beta2 ~ dnorm(10,23.51818)
beta3 ~ dnorm(0.5,23.51818)
beta4 ~ dnorm(-0.250,23.51818)
beta5 ~ dnorm(-1.5,23.51818)
beta6 ~ dnorm(-0.5,23.51818)
beta7 ~ dnorm(-1,23.51818)
prec ~ dgamma(5,0.00232)
s2<-1/prec
}
list(y=c(32,46,57,65,36,48,57,50,44,53,56),
x1=c(-1,1,-1,1,-1,1,-1,0,0,0,0),
x2=c(-1,-1,1,1,-1,-1,1,0,0,0,0),
```

```
x3=c(-1,-1,-1,-1,1,1,1,0,0,0,0),
x3x3=c(1,1,1,1,1,1,1,0,0,0,0),
x1x2=c(1,-1,-1,1,1,-1,-1,0,0,0,0),
x1x3=c(1,-1,1,-1,-1,1,-1,0,0,0,0),
x2x3=c(1,1,-1,-1,-1,-1,1,0,0,0,0))
list(beta0 = 50.75, beta1 =5, beta2 =10, beta3 = 0.5, beta4 = - 0.250, beta5 = - 1.5,
beta6 = - 0.5, beta7 = -1, prec = 0.01063,mu = 49.45455)
```

Output:

Node	Posterior Mean	Posterior S.d	Posterior Median	Start	Sample
beta0	50.75	0.2986	50.75	1	100000
beta1	5.001	0.1986	5.001	1	100000
beta2	10.00	0.1991	10.00	1	100000
beta3	0.4992	0.1987	0.4992	1	100000
beta4	-0.2497	0.1999	-0.2497	1	100000
beta5	-1.500	0.1986	-1.500	1	100000
beta6	-0.5008	0.199	-0.5007	1	100000
beta7	-0.9996	0.1983	-0.9992	1	100000
Mu	49.46	0.3423	49.46	1	100000
Prec	0.2571	0.0795	0.2486	1	100000

10 Win-BUG Program to find the posterior estimates of parameters corresponding to the data presented

```
model
{
for(i in 1:25)
{
mu.y[i]<-
beta0+beta1*x1[i]+beta2*x2[i]+beta3*x3[i]+beta4*x4[i]+beta5*x1x1[i]+
beta6*x2x2[i]+beta7*x3x3[i]+beta8*x4x4[i]+beta9*x1x2[i]+beta10*x1x3[i]+bet
a11*x1x4[i]+beta12*x2x3[i]+ beta13*x2x4[i]+ beta14*x3x4[i]
y[i] ~ dnorm(mu.y[i],prec)
}
mu ~ dnorm(18.552,0.62204)
beta0 ~ dnorm(21.447,0.62204)
beta1 ~ dnorm(1.318,0.77759)
beta2 ~ dnorm(2.691,0.77759)
beta3 ~ dnorm(2.114,0.77759)
beta4 ~ dnorm(1.260,0.77759)
beta5 ~ dnorm(0.421,1.94457)
beta6 ~ dnorm(-1.580,1.94457)
beta7 ~ dnorm(-1.530,1.94457)
beta8 ~ dnorm(-0.930,1.94457)
beta9 ~ dnorm(0.750,0.97193)
beta10 ~ dnorm(0.3,0.97193)
```

```
beta11 ~ dnorm(0.175,0.97193)
beta12 ~ dnorm(0.6,0.97193)
beta13 ~ dnorm(0.475,0.97193)
beta14 ~ dnorm(-0.075,0.97193)
prec ~ dgamma(12,0.00543)
s2<-1/prec
}
list(y=c(27.6,16.6,15.4,17.4,17,19,17.4,12.6,18.6,22.4,21.4,14,24,15.6,13,14.4,22.
6,23.4,20.6,22.6,13.4,20.6,15.6,21,17.6),
x1=c(1,-1,1,-1,1,-1,1,-1,1,-1,1,-1,1,-1,1,-1,0,1.414,-1.414,0,0,0,0,0,0),
x2=c(1,-1,-1,1,-1,1,1,-1,-1,1,1,-1,1,-1,-1,1,0,0,0,1.414,-1.414,0,0,0,0),
x3=c(1,1,-1,-1,1,1,-1,-1,1,1,-1,-1,1,1,-1,-1,0,0,0,0,0,1.414,-1.414,0,0),
x4=c(1,1,1,1,-1,-1,-1,-1,1,1,1,1,-1,-1,-1,-1,0,0,0,0,0,0,0,1.414,-1.414),
x1x1=c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,0,1.999,1.999,0,0,0,0,0,0),
x2x2=c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,0,0,0, 1.999,1.999,0,0,0,0),
x3x3=c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,0,0,0,0,0, 1.999,1.999,0,0),
x4x4=c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,0,0,0,0,0,0,0,1.999,1.999),
x1x2=c(1,1,-1,-1,-1,-1,1,1,-1,-1,1,1,1,1,-1,-1,0,0,0,0,0,0,0,0,0),
x1x3=c(1,-1,-1,1,1,-1,-1,1,1,-1,-1,1,1,-1,-1,1,0,0,0,0,0,0,0,0,0),
x1x4=c(1,-1,1,-1,-1,1,-1,1,1,-1,1,-1,-1,1,-1,1,0,0,0,0,0,0,0,0,0),
x2x3=c(1,-1,1,-1,-1,1,-1,1,-1,1,-1,1,1,-1,1,-1,0,0,0,0,0,0,0,0,0),
x2x4=c(1,-1,-1,1,1,-1,-1,1,-1,1,1,-1,-1,1,1,-1,0,0,0,0,0,0,0,0,0),
x3x4=c(1,1,-1,-1,-1,-1,1,1,1,1,-1,-1,-1,-1,1,1,0,0,0,0,0,0,0,0,0))
list(beta0 =21.447, beta1 =1.318, beta2 =2.691, beta3 = 2.114, beta4 =1.260, beta5
= 0.421, beta6 = - 1.580, beta7 = - 1.530, beta8 = - 0.930, beta9=0.750,
beta10=0.3, beta11=0.175, beta12=0.6, beta13=0.475, beta14=- 0.075, prec =
0.06430,mu = 18.552)
```

Output:

Node	Posterior Mean	Posterior S.d	Posterior Median	Start	Sample
beta0	20.87	0.4989	20.86	1	100000
beta1	1.350	0.2094	1.350	1	100000
beta2	2.435	0.2143	2.435	1	100000
beta3	2.215	0.2137	2.215	1	100000
beta4	1.317	0.2158	1.317	1	100000
beta5	0.2281	0.3194	0.227	1	100000
beta6	-1.64	0.321	0.00104	1	100000
beta7	-0.893	0.3244	-0.8885	1	100000
beta8	-0.483	0.3413	-0.4793	1	100000
beta9	0.7509	0.2257	0.7517	1	100000
beta10	0.2993	0.2267	0.2991	1	100000
beta11	0.1744	0.2262	0.1746	1	100000
beta12	0.5989	0.2268	0.5993	1	100000
beta13	0.4755	0.2263	0.4757	1	100000
beta14	-0.0758	0.2263	-0.0755	1	100000
Mu	18.55	1.268	18.55	1	100000

Prec	1.222	0.2889	1.199	1	100000

BIBLIOGRAPHY

1. Bansal A.K. (2007): “Bayesian Parametric Inference”, Narosa Publishing House, New Delhi.
2. Berger J.O. (1985): “Statistical Decision Theory and Bayesian Analysis”, Second Edition, Springer-Verlog, New York.
3. Bernardo J.M. and Smith A.F. (1994): “Bayesian Theory”, Wiley & Sons, New York
4. Bhatra charyulu N.Ch. (1994): “Some studies on the second order response surface designs” unpublished Ph.D. thesis submitted to Osmania University, Hyderabad.
7. Box G.E.P and Wilson K.B. (1951): “On the experimental attainment of optimum conditions”, Journal of Royal Statistical Society, Series B, Vol. 13, pp 1 – 45.
8. Box G.E.P. and Draper N.R. (1987): “Empirical Model-Building and response surfaces”. John Wiley, New York.
9. Box G.E.P. and Hunter J.S.(1957): “Multifactor experimental designs for exploring response surfaces”, Annals of Mathematical Statistics, Vol. 28, pp 195 – 241.
10. Box G.E.P. and Meyer R.D.(1993): “Finding the active factors in fractionated screening experiments”, Journal of Quality Technology, Vol. 25, pp 94 – 105.
11. Box G.E.P. and Tiao G.C. (1973): “Bayesian Inference in Statistical Analysis” Wiley and Sons.
12. Breiman L. (1995): “Better subset regression using the nonnegative garrotte”, Technometrics, Vol. 37, pp 373 – 384.
13. Breiman L., and Friedman J. (1985): “Estimating Optimal Transformations for Multiple Regression and Correlation”, Journal of the American Statistical Association, Vol. 80, pp 580 – 597.

14. Breiman L., Friedman J., Olshen R., and Stone C. (1984): "Classification and Regression Trees". Belmont, CA: Wadsworth.

15. Broman Karl W and Speed Terence P. (2002): "A model selection approach for the identification of quantitative trait loci in experimental crosses", Journal of Royal Statistical Society, Series B, Vol. 64, part 4, pp 641 – 656.

16. Carlin B.P. and Siddhartha Chib (1995):"Bayesian Model Choice Via Markov Chain Monte Carlo Methods", Journal of the Royal Statistical Society, Series B, Vol. 57, pp 473 – 484.

17. Chen Tao and Wang Bo (2010): "Bayesian variable selection for Gaussian process regression: application to chemometric calibration of spectrometers", Neuro Computing, Vol. 73, pp 13 – 15.

18. Chen, H. (1988): "Convergence Rates for Parametric Components in a Partly Linear Model", Annals of Statistics, Vol.16, pp 136 – 146.

19. Cheng and Wu (2001): "Factor Screening and Response Surface Exploration", Statistica Sinica Vol. 11, pp 553 – 604.

20. Chipman H. (1996): "Bayesian variable selection with related predictors", The Canadian Journal of Statistics, Vol. 24, pp 17 – 36.

21. Chipman H., Hamada M. and WU C.F.T (1997): "A Bayesian variable selection approach for analyzing designed experiments with complex aliasing", Technometrics, Vol. 39, pp 372 – 381.

22. Christian Gogu, Haftka R. T., Satish K .B. and Bhavani (2009): "Dimensionality Reduction Approach for Response Surface Approximation: Application to Thermal Design", AIAA Journal, Vol. 47, pp 1700 – 1708.

23. Christine M. Anderson-Cook, ConnieM. Borror, Douglas C. Montgomery (2009): “Response Surface Design evaluation and comparison”, Journal of Statistical Planning and Inference, Vol. 139, pp 629 – 641.

24. Das M.N. and Giri N.C (1986):”Design and Analysis of Experiments”; New Age International Publishers, New Delhi, 2nd edition.

25. Degroot M.H. (1970):”Optimal Statistical Decisions”, Mcgraw-Hill, New York.

26. Dellaportas P, Forster J.J. and Ntzoufras I. (2000): “Bayesian Variable Selection using the Gibbs Sampling”, Generalized linear models: a Bayesian perspective, pp 273 – 286. Marcel Dekker, Inc., New York.

27. Dellaportas P, Forster J.J. and Ntzoufras I. (2002): “On Bayesian model and Variable Selection using MCMC”, Statistics and Computing, Vol. 12, pp 27 – 36.

28. Dey Tanujit, (2013): “Model Sampler: An R Tool for Variable Selection and Model Exploration in Linear Regression”, Journal of Data Science, Vol. 11, pp 343 – 370.

29. Diaconis P., and Freedman D. (1984): “Asymptotics of Graphical Projection Pursuit”, Annals of Statistics, Vol. 12, pp 793 – 815.

30. Draper N.R. Smith H (1998):”Applied Regression Analysis”, John Wiley and Sons, USA, 3rd edition.

31. Elangovan R and Lokeshmaran A (2014): “Bayesian Variable Selection for Cox’s Regression Model”, Asia Pacific Journal of Research, Vol. I, Issue XIII, pp 11 – 23.

32. Engle R. F., Granger C.W. I., Rice J., and Weiss A. (1986): “Semi-parametric Estimates of the relation between Weather and Electricity sales”, Journal of the American Statistical Association, Vol. 81, pp 310 – 320.

33. Fan J and Lin (2002): “Variable Selection for Cox’s Proportional Hazards Model and Frailty Model”, Annals of Statistics, Vol. 30, No. 1, pp 74 – 99.

34. Fan J. and Lin R. (2001): “Variable selection via non-concave penalized likelihood and its oracle properties”, Journal of the American Statistical Association, Vol. 96, pp 1348 – 1360.

35. Feng-Jenq Lin (2008): “Solving Multi-collinearity in the process of Fitting Regression Model using the Nested Estimate Procedure”, Quality and Quantity, Vol. 42, pp 417 – 426.

36. Foster and George (1994): “The Risk Inflation Criterion for Multiple Regression”, Annals of Statistics, Vol. 22, No. 4, pp 1947 – 1975.

37. Friedman and Tukey (1974): “A Projection Pursuit Algorithm for Exploratory Data Analysis”, IEEE Transactions on Computers, Vol. 23, pp 881 – 890.

38. Gelman A., Carlin J.B., Stern H.S. and Rubin D.B. (1995): “Bayesian Data Analysis”, chapman & Hall London.

39. George (2000): “The Variable Selection Problem”, Journal of the American Statistical Association, Vol. 95, No. 452, pp 1304 – 1308.

40. George and Mc Culloch (1997): “Approaches for Bayesian Variable Selection”, Statistica Sinica 7, pp 339 – 373.

41. George E. I. and Foster D. P. (2000): “Calibration and empirical Bayes variable selection”, Biometrica, Vol. 87, pp 731 – 747.

42. George E.I. and Mc Culloch R.E. (Sep 1993): “Variable Selection via Gibbs Sampling”; Journal of the American Statistical Association, Vol. 88, No. 423, pp 881 – 889.

43. Gewekke J (1995): “Variable selection and model comparison in regression” Proceedings of the fifth valenceia international meeting on Bayesian Statistics, Oxford University Press, Oxford, U.K.

44. Ghosh JK Delampady M and Samanta J (2006): “An introduction to Bayesian Analysis: Theory and Methods”, Springer Texts in Statistics, New York.

45. Giunta Anthony A, Balabanov Vladimir, Kaufman Matthew, Burgee Susan, Grossman Bernard, Haftka Raphael T, Mason William H and Watson Layne T (1996): “Variable-Complexity Response Surface Design of an HSCT Configuration”. https://www.researchgate.net/publication/2774198.

46. Gupta S.C. and Kapoor V.K. (1976): “Fundamentals of Applied Statistics”, Sultan Chand and sons, New Delhi, 1st edition.

47. Hastie and Tibshirani (1986): “Generalized Additive Models”, Statistical Science, Vol. 1, No. 3, pp 297 – 318.

48. Hastie, T and Stuetzle. W. (1989): “Principal curves”, Journal of the American Statistical Association, Vol. 84, No. 406, pp 502 - 516.

49. Homma Toshimitsu and Saltelli Andrea (1996): “Importance measures in global sensitivity analysis of nonlinear models”, Reliability Engineering and System Safety, Vol. 52, pp 1 – 17.

50. Huber, P. (1985): “Projection Pursuit: with discussion”, Annals of Statistics, Vol. 13, pp 435 - 526.

51. Janathi Idrissi, M., Sbihi, A., and Touahni, R. (2004): “An improved neural network technique for data dimensionality reduction in remotely sensed imagery”, International Journal of Remote Sensing, Vol. 25, pp 1981 - 1986.

52. Jin. Y., and Shaoping. M. (2000): “A neural- network dimension reduction method for the large- set pattern classification”, Lecture Notes in Computer Science, Springer, Vol. 1948, pp 426 - 433.

53. Johnson R.A. and Wichern D.W. (2013): “Applied Multivariate Statistical Analysis”, PHI Learning Pvt. Ltd. New Delhi, 6th edition.

54. Joyee Ghosh and Andrew E. Ghattas (2015): “Bayesian Variable Selection under Collinearity.” The American Statistician, Vol. 69, pp 165 – 173.

55. Karenchan, Saltelli., A. and Tarantola S. (1997): “Sensitivity analysis of model output: variance-based methods make the difference”, Proceedings of the 1997 winter simulation conference.

56. Kaufman M., Balabano V. V., Grossman B., Manson W. H., Watson, L. J. and Haftaoka, R. J. (1996): “Multidisciplinary optimization via Response surface techniques”, Proceedings of the 36th Israel Conference on Aerospace Sciences, A57 - A67.

57. Khuri, A.I and Cornell, J.A. (1996): “Response surface. Design and analysis”, 2nd edition Marcel Dekker, New York.

58. Kuo L and Mallick B. (April 1998): ”Variable Selection for Regression Models”, Sankhya, The Indian Journal of Statistics, Series B, Vol. 60, No1, pp 65 – 81.

59. Lacey and Steele (2006): “The use of Dimensional Analysis to Augment Design of Experiments for Optimization and Robistification”, Journal of Engineering Design, Vol. 17, No. 1, pp 55 – 73.

60. Li ker-Chau, Lue Heng-Hui and Chen Chun-Houh (2000): “Interactive Tree Structured Regression via Principal Hessian directions”, Journal of the American Statistical Association, Vol. 95, pp 547 – 560.

61. Lin K.C. (1991): "Sliced inverse regression for dimension reduction", Journal of American Statistics, Vol. 86, pp 316 – 342.

62. Lindley D.V. (2000): "A Bayes Factor with Reasonable Model Selection Consistency for ANOVA Model", Arxiv preprint arXiv: 0906.4329.

63. Lindsey Charles and Sheather Simon (2010): "Variable Selection in Linear Regression", The Stata Journal, Vol. 10, No. 4, pp 650 – 669.

64. Loh, W. Y., and Vanichsetakul, N. (1988): "Tree-Structured Classification via Generalized Discriminant Analysis", Journal of the American Statistical Association, Vol. 83, pp 715 – 728.

65. Luan Jaupi (2014): "Variable Selection Methods for Multivariate Process Monitoring", Proceedings of the World Congress on Engineering 2014 Vol. II, July 2 – 4, London, U.K.

66. McKay M.D., R.J. Beckman and W.J. Conover (1979): "A comparison of three methods for selecting Values of input variables in the analysis of output from a computer code", Technometrics Vol. 42, No. 1, pp 55 – 61.

67. Medhi J (1982): "Stochastic Process", Wiley Eastern Limited, New Delhi.

68. Mitchell T.J, and Beauchamp J.J. (1988): "Bayesian Variable selection the Linear Regression". Journal of the American Statistics Association, Vol. 83, pp 1023 – 1036.

69. Montgomery D.C. (2001): "Design and Analysis of Experiments", John Wiley and Sons, USA, 5th edition.

70. Myers R.H. and Montgomery D.C. (1995): "Response surface methodology: Process and Product optimization using designed experiments". John Wiley, New York.

71. Myers R.H., Kim Y and Griffins S.K. (1997): "Recent Developments in response surface methodology and its applications in industry in: statistical process monitoring and

optimization“, sung H. Park and G.G. Vining, Eds. New York: Marcel Dekker, pp 457 – 481.

72. Myers R.H., Montgomery D.C., Vining G.G., Borror C.M., and Kowalski S.M. (2004): “Response surface methodology: A retrospective and literature survey”. Journal of Quality Technology, Vol. 36, pp 53 - 77.

73. Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller (1953): ”Equation of State Calculations by Fast Computing Machines”; The journal of Chemical Physics, Vol. 21, No. 6, pp 1087 – 1092.

74. O’Hagan A. and Forster J. (2004): “Kendall’s Advanced Theory of Statistics: Vol. 2B: Bayesian Inference”, Second Edition, John Wiley and Sons Ltd, London, U.K.

75. O’Hara R.B. and Sillanpaa M.J. (2009): “A Review of Bayesian Variable Selection Methods: What, How and Which”, Bayesian Analysis: Vol 4, No. 1, pp 85 – 118.

76. Otava Martin, Shkedy Ziv, Lin Dan, Göhlmann Hinrich W. H., Bijnens Luc, Talloen Willem & Kasim Adetayo (2014): “Dose–Response Modeling Under Simple Order Restrictions Using Bayesian Variable Selection Methods”, Statistics in Biopharmaceutical Research, Vol. 6, No. 3, pp 252 – 262.

77. Park, T. and Casella, G. (2008): “The Bayesian lasso.” Journal of the American Statistical Association, Vol. 103, pp 681 – 686.

78. Parpoula Christina, Drosou Krystallenia, Koukouvinos Christos and Mylona Kalliopi (2014): “A New Variable Selection approach Inspired by Supersaturated Designs Given a Large-Dimensional Dataset”, Journal of Data Science, Vol. 12, pp 35 – 52.

79. Peter M. Lee (2012): “Bayesian Statistics- An introduction”, John Wiley & Sons Ltd, Publications, 4th Edition.

80. Plackett R.L and Burman J.P (1946): "The Design of optimum multifactorial Experiments", Biometrika, Vol. 33, pp 305 – 325.

81. Radchenko P and James G.M. (2011): "Improved Variable Selection with forward-lasso adaptive shrinkage", Annals of Applied Statistics, Vol. 5, No. 1, pp 427 – 448.

82. Raghavarao D. (1970): "Constructions and Combinatorial problems in Design of Experiments", John Wiley and Sons.

83. Rudd Wetzels, Raoul P.P.P. Grasman & Eric-Jan Wagenmatkes (2012): "A Default Bayesian Hypothesis Test for ANOVA Designs." The American Statistician, Vol. 66, pp 104 – 111.

84. Saltelli A. (2002): "Sensitivity analysis for importance assessment", Risk analysis, Vol. 22, No. 3, pp 1 – 12.

85. Saltelli A.S. Tarantola and Chan K (1999): "A quantitative model independent method for global sensitivity analysis of model output". Technometrics, Vol. 41, No. 1, pp 39 – 56.

86. Saund (1989): "Dimensionality-reduction using connectionist networks", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 11, pp 304 – 314.

87. Sheldon M. Ross (2013):"Simulation", Academic Press, Uk, 5th edition.

88. Siddhartha Chib and Edward Greenberg (Nov. 1995): "Understanding the Metropolis – Hastings Algorithm", The American Statistician, Vol. 49, No. 4 pp 327 – 335.

89. Sinha S.K. (1998): "Bayesian Estimation", New Age international Pvt. Ltd, New Delhi.

90. Sobol I.M (1993): "Sensitivity estimates for nonlinear mathematical models, mathematical modeling and computation", Vol. 1, No. 4, pp 407 – 414.

91. Sosada M (1993): "Optimal Conditions for Fractionation of rapeseed lecithin with alcohols", Journal of the American Oil Chemists, Society, Vol. 70, pp 405 – 410.

92. Steel S.J and Uys D.W. (2007): "Variable selection in Multiple Linear Regression: The influence of individual cases", Orion Journals, Vol. 23, No. 2, pp 123 – 136.

93. Stone (1986): "The dimensionality reduction principal for generalized additive models", Annals of Statistics, Vol. 14, pp 592 – 606.

94. Stone C.J. (1985): "Additive Regression and other nonparametric models", Annals of Statistics, Vol. 13, pp 685 – 705.

95. Sudret B (2009): "Global sensitivity analysis using polynomial chaos expansion", Reliability Engineering & amp, System Safety, Vol. 93 (Issue 7), pp 964 – 979.

96. Taguchi (1987): "System of Experimental Design", Unipub / Kraus, International Publication.

97. Tengfei Long and Weili Jiao (2012): "An Automatic Selection and Solving Method for Rational Polynomial Coefficients based on Nested Regression", proceedings of the 33rd Asian Conference on Remote Sensing held on November 26-30, 2012. Ambassador City, Pattaya, Thailand.

98. Tengfei Long, Weili Jiao and Guojin He (2014): "Nested Regression Based Optimal Selection (NRBOS) of Rational Polynomial Coefficients", American Society for Photogrammetric Engineering & Remote Sensing, Vol. 80, No. 3, pp 261 – 269.

99. Tibshirani Robert (1996): "Regression shrinkage and Selection via the Lasso", Journal of Royal Statistical society, Series B, Vol. 58, Issue I, pp 267 – 288.

100. Venter G., Haftka R. T. and Stammer Jr. J. H. (1998): "Construction of response surface approximations for design optimization", Journal of AIAA, Vol. 36, No. 12, pp 2242 – 2249.

101. Vignaux V. A. and Scott J. L. (1999): “Simplifying regression models using dimensional analysis”, Australian & Newzeland Journal of Statistics, Vol. 41, pp 31 – 41.

102. W.K. Hastings (April 1970):” Monte Carlo Sampling Methods Using Markov Chains and their Applications”; Biometrika, Vol. 57, No. 1, pp 97 – 109.

103. Wang Xinlei and George Edward I. (2004): “A Hierarchical Bayes Approach to Variable Selection for Generalized Liner Models”.

104. Weinwurm Stephan, Sölkner Johann and Waldmann Patrik (2013): “The Effect of Linkage Disequilibrium on Bayesian Genome-wide Association Methods”, Journal of Biometrics and Biostatistics, Vol. 4, Issue 5, pp 1 – 11.

105. Wu Yichao and Liu Yufeng (2009): ”Variable Selection in Quantile Regression”, Statistica, Sinica, Vol. 19, pp 801 – 817.

106. Xu J. and Ying, Z. (2010): “Simultaneous estimation and variable selection in median regression using Lasso – type penalty”, Annals of the Institute of Statistical Mathematics, Vol. 62, pp 487 – 514.

107. Xu Xiaofan and Ghosh Malay (2015): “Bayesian Variable Selection and Estimation for Group Lasso”, Bayesian Analysis, Vol. 10, No. 4, pp 909 – 936.

108. Yamada, S. (2004): “Selection of active factors by stepwise regression in the data analysis of Supersaturated Design”, Journal of Quality Engineering, Vol. 16, pp 501 – 513.

109. Yuan M. and Lin Y. (2005): “Efficient empirical Bayes variable selection and estimation in linear models”, Journal of the American Statistical Association, Vol. 100, No. 472, pp 1215 – 1225.

110. Yuan M. and Lin Y. (2006): “Model selection and estimation in regression with grouped variables”, Journal of Royal statistical Society, Series B, Vol. 68, pp 49 – 67.

111. Yuan M. Joseph R.V. and Lin Y (2007): “An efficient variable selection approach for analyzing designed experiments”, Technometrics, Vol. 49, pp 430 – 439.

112. Zou Hui and Hastie Trevor (2005): “Regularization and Variable Selection Via the elastic net”, Journal of Royal Statistical Society B, Vol. 67, Part 2, pp 301 – 320.