

Factor screening in a 12 Run Plackett-Burman
design assuming four active factors

Yingzi Jin

MASTER THESIS

Department of Mathematical Sciences

November 2016

Supervisor: Professor John soelve Tyssedal

Preface

This Master's thesis is the final project for receiving the Master of Science degree in Mathematical Sciences at the Norwegian University of Science and Technology(NTNU).

I express my sincere gratitude to my supervisor Professor John Soelve Tyssedal for his advice, constant support and encouragement. He also spent much time to improve my writing and provide me relevant advice.

Finally, I would like to thank my parents for their support.

Yingzi Jin

Trondheim, November, 2016

Abstract

In this thesis we perform factor screening in a non-regular two-level design by reducing the number of possible sets of active factors to a certain number. The 12 Run Plackett-Burman(PB) design with four active factors is mainly concerned. Our proposed method works through picking up the 6 effects with the highest absolute value out of 10 in each projection model. To evaluate this method, we used the same example as was used in Tyssedal and Shahrukh[15] where variable selection methods such as AIC , F test and ΔR^2 -method used on projection models. A real example is included at the end to show how our proposed factor screening method can be done in practice.

Keywords: Non regular design, 12 Run PB design, Projection model, Factor screening

Contents

Preface	i
Abstract	ii
1 Introduction	1
2 Theory	4
2.1 Factor screening design	4
2.2 Regression Analysis	4
2.3 Two-level factorial designs	5
2.3.1 The 2^2 design	6
2.3.2 The 2^3 design	7
2.3.3 Special Feature	9
2.4 Fractional factorial designs	11
2.4.1 Fractions of 2^p experiments	12
2.4.2 Generation and defining relation	12
2.4.3 Resolution	13
2.5 Plackett-Burman Design(nonregular two-level designs)	15
2.5.1 Construction of Plackett-Burman design	15
2.5.2 Design Projectivity	17
2.5.3 Effect heredity	18
2.5.4 Confounding Pattern	18
3 Some variable selection techniques	19
3.1 The AIC criteria in forward selection	21
3.2 F test in forward selection	22
3.3 R square method in forward selection	22
3.4 An example	24
4 A new way for screening in a 12 run PB design with four active factors	28
4.1 Suggested variable selection in certain situation	29

4.2 Comparing the results	52
5 A real example	54
6 Concluding Remarks	59
7 Appendix A	63
8 Appendix B	69

1 Introduction

At the early stage of an experimental search, the important goal is to identify the subspace of active factors among many, which is known as factor screening. Two-level nonregular designs are important for factor screening and include all orthogonal two-level designs that does not belong to the 2^{k-p} family, that is, regular $1/2^p$, $p = 0, 1, \dots, k - 1$, fraction of 2^k factorials also known as regular designs or two-level fractional factorial designs. Fractional factorial designs are often used for the situations where the experiment wants to investigate more than k factors in 2^k experiments runs. Non-regular designs are important for two main reasons. One is their projective properties, that is, they project far better into lower dimensions than the ones for regular designs, see [1][3] and [10]. The other is that they apparently exist for every n as a multiple of four and $n \geq 12$, and therefore provide us with a lot more design alternatives that can be obtained from two-level fractional factorial designs.

The drawback of non-regular designs is their complicated alias pattern among main effects and interactions. Plackett-Burman designs were proposed more than six decades ago[12] and is the most well-known non-regular designs. For example, consider the widely used 12 Run PB design with 11 factors. There are 55 two-factor interactions and the main effect of each factor X is partially aliased with the 45 two-factor interactions not involving X , thereby making it difficult to interpret the significance of interactions. Therefore, methods such as normal and half-normal plots or more quantitative methods such as Lenth's method, often used for analyzing replicated two-level designs maybe of little value since these methods are based on being able to separate active contrasts from contrasts estimating only noise.

However, there are still several methods available for analyzing replicated nonregular designs and there can be mainly classified as factor-based or effect-based search procedures. The aim of factor-based search is to identify the subspace of active factors of normally low dimension, typically 2, 3 or 4, which is known as factor sparsity principle. These active factors are responsible for most of the variation in the data. The performance of such a procedure depends heavily on the projective properties of the design used. Examples of such methods

are given in [2],[7]and[14]. The aim of effect-based search is to identify the significant effects which are most likely active main effects and two-factor interactions. Consider a linear regression model consisting of main effects and interactions and assumed to explain the response well. Such a model can be defined as $Y = X\beta + \varepsilon$ where the $X = \begin{bmatrix} 1 & x_1 & x_2 & \cdots & x_k \end{bmatrix}$ is called the design matrix, Y is the response and ε is the noise and assumed to follow a normal distribution with mean equal to 0 and variance equal to σ^2 . For effect based methods, the principle of heredity is included. Some examples of such methods are the stepwise regression procedure[6] and the Bayesian stochastic search variable selection[4].

When performing a screening, not all methods work equally well on specific examples, and their performance when applied to a variety of models and noise has not been tested out. In order to reducing the number of candidate sets and identifying the correct model, [15] proposed a factor based procedure based on projection properties of the designs. Forward selection was performed on projection models using three different criteria, the *AIC*, the *F* test and the ΔR^2 -method.

The purpose of this thesis is to give a different method to do the factor screening. It reduces the number of terms in the projection models to simplify the complicated alias relationship. For instance, in a 12 Run PB design, there are 330 possible candidate sets of active factors assuming four factors are active. For each candidate set of four factors, we assumed a projection model consisting of a constant, four main effects and six two-factor interactions, altogether 11 terms. Thus, 11 terms are reduced to 6 by picking out the 6 effects with highest absolute value. We reduced 330 candidate sets to 10 with the smallest MSE which is known as the mean square error. In order to compare the results from our method to the other methods used in [15], we consider the same examples.

This paper begins with the introduction of regular and non-regular designs. In section 3, we explain how some variable selection techniques like *AIC*, *F* test and ΔR^2 -method work and their limitations. Our suggested method and its algorithm are given in section 4 together with the analysis of four examples taken from [15]. The results from different methods and

the discussion are still provided in section 4 followed by an analysis of a real example in section 5. Concluding Remarks are given in section 6.

2 Theory

2.1 Factor screening design

At the early stage of an experimental search, it is important to identify the subspace of active factors among many, in other words, many factors are considered and the objective is to identify those factors that have large effects. This is known as factor screening.

In general, the dimension of the subspace will be lower than the one of tested factors, i.e. typically 2, 3 or 4. This principle is also known as factor sparsity which actually refers to the idea that a few factors are responsible for most of the variation in the data. All the designs talked about in this article are factor screening design.

2.2 Regression Analysis

Regression analysis is a statistical methodology for finding a fractional relationship between response variables and a set of explanatory variables. It can also be used for assessing the effects of factors on responses, which is the goal in design of experiments.

Generally, the multiple linear regression model with a single response is specified to take the form

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \varepsilon \quad (1)$$

where x_1, x_2, \dots, x_n are n explanatory variables thought to be related to the response variable Y . $\beta_0, \beta_1, \dots, \beta_n$ are called regression coefficients, ε is called the noise.

With all observations taken into account, (1) can be written on matrix form as follows.

$$Y = X\beta + \varepsilon \quad (2)$$

where the $X = \begin{bmatrix} 1 & x_1 & x_2 & \cdots & x_k \end{bmatrix}$ is called the design matrix, then the estimators for the coefficients is given by:

$$\hat{\beta} = (X^T X)^{-1} X^T Y. \quad (3)$$

Note that the design matrix plays an important role on how easy it is to find a good model in the end.

2.3 Two-level factorial designs

The factorial experiments, where all combination of the levels of the factors are run, are usually referred to as a full factorial experiment. In general, a full factorial two level design with k factors requires 2^k runs for a single replicate. For instance, a two level experiment with two factors will require $2^2 = 4$ runs. As for the "factors" used in the two level experiments, some of them naturally have two levels such as "gender", obviously the male and female are its two levels. And another example, if concentration is a factor that varies from $20kg/m^3$ to $40kg/m^3$, then the two levels for this factor used in the 2^k designs would be $20kg/m^3$ and $40kg/m^3$.

The two levels of the factor are normally marked by 1 or -1, which makes the factor columns orthogonal. Therefore, the coefficients can be computed in a simple way to be mentioned.

When doing the factorial design there are two classes of effects that we are interested in: Main Effects and Interactions.

For two-level designs we define the main effect of a factor as: Excepted average response when the factor is on the high level minus excepted average response when the factor is at the low level.

The interaction between two factors is defined as: Half the main effect of a factor when the other is on the high level minus half the main effect of a factor when the other factor is at its low level.

Note that two-level factorial designs are normally called the regular designs.

2.3.1 The 2^2 design

The simplest of the two level factorial experiments is the 2^2 design where two factors named A and B are investigated at two levels. Four runs will be required for a single replicate of this design. In this situation, there are two main effects A , B and an interaction effect AB . Main effect of factor A means the overall difference of the expected average response when factor A is in high level and low level. Similarly, main effect of factor B refers to overall difference of the expected average response when factor B is in high level and low level. As for the interaction of AB , it means the overall difference of half the main effect of factor A when factor B is on the high level and on its low level.

The treatment for this design is shown in Table 1.

Table 1: The 2^2 Factorial Design

Run	Treatment Combination	I	A	B	AB
1	1	1	-1	-1	1
2	a	1	1	-1	-1
3	b	1	-1	1	-1
4	ab	1	1	1	1

The notation for "Treatment Combination" shown in Table 1 comes from the letter for the main factor which is the only one marked with the high level, while (1) represents the treatment combination where all main factors involved are at low level.

A linear regression model consisting of main effects and interactions is assumed to give an adequate approximation of the response. Such a model is given in Section 2.2 equation

(2), in this situation, the design matrix becomes $X =$

$$\begin{matrix} & I & A & B & AB \\ \begin{pmatrix} +1 & -1 & -1 & +1 \\ +1 & +1 & -1 & -1 \\ +1 & -1 & +1 & -1 \\ +1 & +1 & +1 & +1 \end{pmatrix} \end{matrix}$$

Section 2.2 also mentions the estimator of the coefficients which is given in the equation

(3), in order to obtain the estimated main effects and interactions, we have to multiply the corresponding coefficients with 2.

As noted earlier that the design matrix has orthogonal columns, which means that the sum of the terms resulting from the dot product of any two columns of the design matrix is zero. In fact, all 2^k designs are orthogonal designs.

The 2^2 design can also be presented in a geometrical way, using a square with the four treatment combinations lying at the four corners.

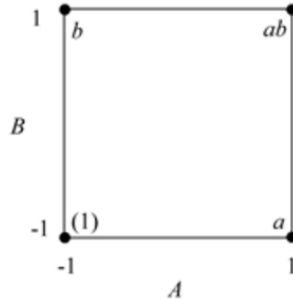


Figure 1: 2^2 design in a geometrical way

2.3.2 The 2^3 design

Similarly, suppose three factors named A, B, C are considered for a two-level design, then in this case, there are three main effects A, B and C , four interaction effects including three two-factor interaction effects AB, AC, BC and one three-factor interaction effect ABC . Therefore, we would have $2^3 = 8$ treatment combinations totally, which is shown in Table 2.

In this situation, I, A, B, \dots, ABC are called factorial effects.

Table 2: The 2^3 Factorial Design

Run	Treatment Combination	I	A	B	C	AB	AC	BC	ABC
1	a	1	1	-1	-1	-1	-1	1	1
2	b	1	-1	1	-1	-1	1	-1	1
3	c	1	-1	-1	1	1	-1	-1	1
4	abc	1	1	1	1	1	1	1	1
5	ab	1	1	1	-1	1	-1	-1	-1
6	ac	1	1	-1	1	-1	1	-1	-1
7	bc	1	-1	1	1	-1	-1	1	-1
8	(1)	1	-1	-1	-1	1	1	1	-1

In this case, the design matrix is $X=$

$$\begin{pmatrix} I & A & B & C & AB & AC & BC & ABC \\ +1 & +1 & -1 & -1 & -1 & -1 & +1 & +1 \\ +1 & -1 & +1 & -1 & -1 & +1 & -1 & +1 \\ +1 & -1 & -1 & +1 & +1 & -1 & -1 & +1 \\ +1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 \\ +1 & +1 & +1 & -1 & +1 & -1 & -1 & -1 \\ +1 & +1 & -1 & +1 & -1 & +1 & -1 & -1 \\ +1 & -1 & +1 & +1 & -1 & -1 & +1 & -1 \\ +1 & -1 & -1 & -1 & +1 & +1 & +1 & -1 \end{pmatrix}$$

The estimated coefficients can be computed by $\hat{\beta} = (X^T X)^{-1} X^T Y$ as mentioned in section 2.2. The estimated main effects and interaction are given by $2\hat{\beta}$.

The 2^3 design can also be presented geometrically using a cube with the eight treatment combinations lying at the eight corners.

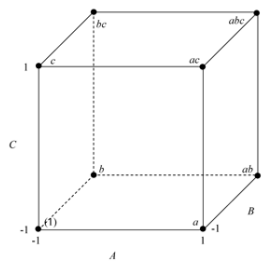


Figure 2: 2^3 design in a geometrical way

2.3.3 Special Feature

One special feature of the two-level factorial design is the term $(X^T X)^{-1}$, where X is the design matrix. Let's take the 2^3 design as an example. Since the design matrix of 2^3 design has been shown above, we can directly get the $(X^T X)^{-1}$ matrix.

$$(X^T X)^{-1} = \begin{bmatrix} 0.125 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.125 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.125 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.125 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.125 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.125 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.125 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.125 \end{bmatrix}.$$

The variance-covariance matrix of the coefficients are given by:

$$\text{Var}(\hat{\beta}) = (X^T X)^{-1} \sigma^2 \quad (4)$$

We notice that $(X^T X)^{-1}$ is diagonal due to the orthogonal design matrix X and it can be simplified as

$$(X^T X)^{-1} = 0.125 \cdot I = \frac{1}{2^3} \cdot I \quad (5)$$

Where I is the identity matrix which has the same order as the design matrix X . Thus, the $(X^T X)^{-1}$ matrix for any 2^k design with a single replicate can be written as [8]:

$$(X^T X)^{-1} = \frac{1}{2^k} \cdot I \quad (6)$$

Then the variance-covariance matrix for the coefficients is:

$$C = \sigma^2 \cdot \frac{I}{2^k} \quad (7)$$

which shows that the estimated effect coefficients are uncorrelated and have a constant variance σ^2 . In other words, the main factors and the interactions in the 2^k design are independent of each other. Note that the parameter $\hat{\beta}$ of the equation (4) is the same as the one of the equation (3).

2.4 Fractional factorial designs

Fractional factorial designs consist of a subset of runs from a full factorial designs. It is used to experiment with more than K factors while only 2^k experiments are allowed to be done. Thus, in order to find out the active effects, we should be careful with choosing the subset of the experimental runs of a full factorial design.

For example, we consider the same full factorial design as in the last section, three factors, 8 experiments in total. Suppose the first 4 experiments which are the Run 1 to 4 are carried out in Table 2, i.e

Run	Treatment Combination	I	A	B	C	AB	AC	BC	ABC
1	a	1	1	-1	-1	-1	-1	1	1
2	b	1	-1	1	-1	-1	1	-1	1
3	c	1	-1	-1	1	1	-1	-1	1
4	abc	1	1	1	1	1	1	1	1

It is clear to see from above chose experiments that I and ABC have the same signs, as well as the A and BC, B and AC, C and AB. Hence, we get

$$\begin{aligned} I &= ABC \\ A &= BC \\ B &= AC \\ C &= AB \end{aligned}$$

For this example, the $I = ABC$ is called the defining relation, which is an important characteristic of fractional factorial designs. Instead, if we only carry out the Run 5 to 8 in Table 2, the defining relation would be $I = -ABC$. And $A = BC$ is called the generator for

Run	Treatment Combination	I	A	B	C	AB	AC	BC	ABC
5	ab	1	1	1	-1	1	-1	-1	-1
6	ac	1	1	-1	1	-1	1	-1	-1
7	bc	1	-1	1	1	-1	-1	1	-1
8	(1)	1	-1	-1	-1	1	1	1	-1

the design. Besides, this designs are known as a half fractions of a 2^3 design, also called 2^{3-1} designs, since we experiment with 3 factors in 2^{3-1} runs. The advantage of such designs is to save effort and costs.

2.4.1 Fractions of 2^p experiments

Half-fraction of the 2^p design involves running only half of the treatments of the full factorial design, and is denoted as a 2^{p-1} design.

If the situation considers a large number of factors, which means there are still a lot of experiments to do even if we do the half-fraction factorial design, a quarter-fraction or smaller fractions will be considered.

2.4.2 Generation and defining relation

The generators in a fractional factorial design determine how the fraction (or subset of runs) is selected from the full set of runs.

Suppose we wish to construct a 1/2 fraction, 4-factor design using the design generator $D = ABC$. As we know, this design would require $2^{4-1} = 8$ runs.

Firstly, construct the full 3-factor design where 1 and +1 represent the low and the high factor levels, respectively.

A	B	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

Secondly, generate the runs for factor D by multiplying the settings for factors A, B, and C together. For example, the setting for factor D for the first run is $-1 * -1 * -1 = -1$.

In this case, the design generator is $D = ABC$, meaning factor D is confounded with the ABC interaction. Because effects that are confounded cannot be estimated separately from each other, design generators should be carefully chosen.

A	B	C	D=ABC
-1	-1	-1	-1
+1	-1	-1	+1
-1	+1	-1	+1
+1	+1	-1	-1
-1	-1	+1	+1
+1	-1	+1	-1
-1	+1	+1	-1
+1	+1	+1	+1

As mentioned earlier, defining relation is a vital characteristic of the factorial design, to be specific, the collection of generating relation for a factorial design is called a defining relation. For the previous example, if we multiply the factor D on the both sides of the design generator, we get the defining relation, $I = ABCD$.

$$D = ABC \iff D^2 = I = ABCD$$

2.4.3 Resolution

Resolution is another vital property of fractional design. The length of the shortest word in the defining relation is called the resolution of the design, which describes the degree to which estimated main effects are aliased or confounded with estimated two-factor interactions, three-factor interactions, etc.

Resolution III

Main effects are aliased with two-factor interactions. The 2^{3-1} given in section 2.4 is of resolution III with $I = ABC$.

Resolution IV

Similarly, main effects are confounded with at worst three-factor interactions. Or two-factor interactions are confounded with certain other two-factor interactions. Example 2^{4-1} , $I = ABCD$.

Resolution V

Main effects are aliased with four-factor interactions. Or two-factor interactions are aliased with certain three-factor interactions. Example 2^{5-1} , $I = ABCDE$.

Note: Assume that we have resolution II, then a possible generator relation would be $A = B$, which makes that main-effects are fully aliased. Thus, resolution II is an undesirable situation. Moreover, a Resolution IV design is better than a resolution III design because we have a less severe confounding pattern since higher order interactions are usually assumed to be much important than low-order interactions. In general, a higher resolution design for the same number of factors will, however, require more runs and it is worse than a lower-order design in that sense.

2.5 Plackett-Burman Design(nonregular two-level designs)

Plackett-Burman(PB) design[12] are the most well-known non-regular designs and was introduced in 1946 by Robin L.Plackett and J.P.Burman.

For fractional factorial two-level design, the number of the experiments n are 2^p , however, for the PB design, $n = 4k$. If $2^p = 4k$, the PB designs become the fractional factorial designs.

The aim of PB designs was to find experimental designs for investigating the independence of some measured quality on a number of factors which are independent, each taking L levels, in such a way as to minimize the variance of the estimates of these dependencies using a limited number of experiments.

For the case of two levels designs($L=2$), Plackett and Burman used the method for generating orthogonal matrices whose arrays are all either 1 or -1, which is the same idea as for factorial designs. As noted earlier, the number of PB designs $n = 4k$, but n has some limitations which are $n \leq 100$ and $n \neq 92$. For instance, if $n = 12$, we call it a 12 Run PB design, which can be used to screen $12 - 1 = 11$ factors.

2.5.1 Construction of Plackett-Burman design

The 12 and 20 Run PB design are given in Table 3 and 4.

2.5.2 Design Projectivity

Compared to regular designs, two-level nonregular designs such as PB designs have very favorable properties when projected onto a subset of factor columns. Box and Tyssedal(1996)[1] introduced and defined the projectivity of two-level designs which is stated as follows:

A $n \times k$ design with n runs and k factors each at two levels is said to be of projectivity P if the design contains a complete 2^P factorial in every possible subset of P out of the k factors, possibly with some points required. It is convenient to describe such design as (n, k, p) screens.

Box and Tyssedal[1] proved three propositions for orthogonal two-level arrays that can be used for classifying PB designs and later Samset and Tyssedal[13] discovered that PB designs with $n=68, 72, 80$ and 84 are also $(n, n - 1, 4)$ screens. However, most of the nonregular designs are $(n, n - 1, 3)$ screens. A complete listing of the projectivity of PB designs are shown in Table 5. Clearly, 12 Run PB design is a $P=3$ design in 11 factors and can be described as

Table 5: The projectivity of Plackett Burman designs

Screens	Number of runs,n
P=2	4,8,16,32,40,56,64,88,96
P=3	12,20,24,28,36,44,48,52,60,76,100
P=4	68,72,80,84

$(12, 11, 3)$ screens.

Samset and Tyssedal also discovered that all PB designs with projectivity $P = 3$ have the hidden projection properties(Lin and Draper[10] and Wang and Wu[16])that all main-effects and two-factor interactions for any four factors are estimable when the higher-order interactions are negligible. This is in line with the concept of generalized projectivity introduced by Evangelaras and Koukouvinos(2004)[5] A $n \times k$ design with n runs and k factors each at two levels is said to be of generalized projectivity P_a , if for any selection of P columns of the design all factorial effects including up to α -factor interactions are estimable.

For example, projecting the 12 Run PB design onto four factors, any of the $\binom{11}{4} = 330$

projections allow all main effects and two-factor interactions to be estimated. All projections onto k factors, $k \leq 4$, are identical, thereby these factors have the equal chance to be identified regardless of which factors they are. Lin and Draper(1991,1992)[10] further considered projections of 16, 20 and 24 run PB designs onto four factors, they look messier. Therefore, the 12 Run PB design has very attractive screening properties and deserved to be explored for such purposes.

2.5.3 Effect heredity

For a two-factor interaction to be included in a model, at least one(weak heredity) or both(strong heredity) of the corresponding main effects need to be significant.

2.5.4 Confounding Pattern

Confounding Pattern is about the alias relationships which is linked the main-effects and the two-factor interactions if we ignore the higher orders. According to [9], the alias relationships for 12-run PB design is complicated. Because for this situation, 11 main effects and 45 two-factor interactions are involved.

3 Some variable selection techniques

Consider a multiple linear regression model which has defined in the beginning, equation (1).

Again the estimator for the coefficients is given by

$$\hat{\beta} = (X^T X)^{-1} X^T Y. \quad (8)$$

Since $E(\varepsilon) = 0$, we have $E(\hat{\beta}) = \beta$. Therefore, the estimator is unbiased.

A measure of the total variance in the response is given by $\sum_{i=1}^n (y_i - \bar{y})^2$, where \bar{y} is the average response. We can split up this as follows:

Firstly we write

$$y_i - \bar{y} = y_i - \hat{y}_i + \hat{y}_i - \bar{y} \quad (9)$$

And then

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 + 2 \sum_{i=1}^n (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) + \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \quad (10)$$

$$= \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \quad (11)$$

Note that

$$\sum_{i=1}^n (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) = \sum_{i=1}^n (y_i - \hat{y}_i)\hat{y}_i - \sum_{i=1}^n (y_i - \hat{y}_i)\bar{y} = 0 \quad (12)$$

Since both the two last expressions are zero as the normal equations have to be fulfilled.

Equation (11) above can also be written as

$$SST = SSE + SSR \quad (13)$$

Where SST is known as the total sum of squares, SSE represents the error sum of squares, SSR represents the regression sum of squares.

As noted in the previous section, the hidden projection property is true for 12 Run PB

design which is considered to be the most important non-regular two-level design for experimentation. Therefore, when doing the factor screening, how well a candidate set's ability to explain the variation in the response depends on how well a model with main factors and all interactions included explains this variation. Such a model is known as a projection model. Therefore, for a projectivity $P = 3$ design one for three factors, all candidate sets are consisted of three main effects, three two-factor interactions and a three-factor interaction. This method is called the full projection method.

The 12 run PB design are of projectivity $P = 3$ as shown in Table 5, hence when we consider the case: identifying the 4 active main factors for the 12 Run Plackett-Burman design, the full projection method is not a good choice. As mentioned earlier, for the 12 run PB design, such models have too many terms. Hence, it's almost impossible to discriminate between the candidate sets ability to explain the variation in the data. The forward selection is suggested to perform on each projection model in [15].

Forward selection is one of the most common variable selection methods and the specific algorithm is presented as follows.

1. Start with only constant term, β_0 , in the equation (1).

2. Find $\max_j R(\beta_j) = \max_j SSR(\beta_j | \beta_0) = \max_j \{SSR(\beta_0, \beta_j) - SSR(\beta_0)\}$

3. If $\max_j \frac{R(\beta_j)}{\frac{SSE}{n-2}} = \frac{R(\beta_m)}{\frac{SSE}{n-2}} < f_{\alpha,1,n-2}$ stop, no variable is entered into the model.

4. If $\frac{R(\beta_j)}{\frac{SSE}{n-2}} \geq f_{\alpha,1,n-2}$, then add x_m to the equation (1).

Find $\max_{j \neq m} R(\beta_j | \beta_m) = \max_{j \neq m} \{SSR(\beta_0, \beta_m, \beta_j) - SSR(\beta_0, \beta_m)\}$ and go to step 3, if $\max_{j \neq m} \frac{R(\beta_j)}{\frac{SSE}{n-3}} < f_{\alpha,1,n-3}$ no more variables are entered into the model. Otherwise proceed in

the same way. Note that the degrees of freedom in the partial F-test is reduced by one for each variable that is entered into the model.

Its advantage is that there is no limitations on the number of active factors and the disadvantage is once a main factor is retained, it cannot be discarded, which may lead to the nesting problem.

There are some common criteria, the AIC criterion, the F-Method and the R Square Method, which also can be used for entering or removing terms in a forward selection procedure.

3.1 The AIC criteria in forward selection

The AIC-criteria is a default criterion and given by

$$AIC = 2p + n \ln\left(\frac{SSE}{n}\right) \quad (14)$$

Where the SSE is the residual sum of squares.

$$SSE = \sum_{i=1}^n (y_i - \hat{y}_i^{(k)})^2 \quad (15)$$

and $\hat{y}_i^{(k)}$ is the estimated response value for a model with k regression variables and $p = k + 1$. If all the models have the same k , the minimum AIC comes from the minimum SSE which is the common objective of least squares fitting. We expect the smaller AIC when adding the terms, in other words, for each step we add the term which decreases the AIC most and the procedure stops if the AIC increases.

3.2 F test in forward selection

When adding one more term in the model, the F-statistic which reflects the variable's contribute to the model if it is included can be written as

$$F = \frac{(R_{k+1}^2 - R_k^2)(n - k - 2)}{1 - R_{k+1}^2} \quad (16)$$

Where R_k^2 is the model with k regression variables. The decision on whether to enter a new term or not will be based on comparing the maximized F-statistic value with $f_\alpha(1, n - k - 2)$. To be specific, $f_\alpha(1, n - k - 2)$ is $1 - \alpha$ percentile in the F-distribution with 1 and $n - k - 2$ degrees of freedom.

Note that there is no way of removing variable once entered when doing forward selection. It may be difficult at the beginning for variable to be judged significant when using the F-test since the variation in the response caused by the rest of the regression variables is included in the error variance. Hence, the procedure may stop before all the active effects are entered. To reduce this problem, Tyssedal and Hussain[15] introduced the ΔR^2 -method.

3.3 R square method in forward selection

As we know, the coefficient of determination is defined as R^2 , where R is the (sample)correlation coefficient of the data and the fitted values that is of the pairs (y_i, \hat{y}_i) :

$$R := \frac{\sum(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum(y_i - \bar{y})^2 \sum(\hat{y}_i - \bar{\hat{y}})^2}} \quad (17)$$

Clearly, $-1 \leq R \leq 1$, $-1 \leq R^2 \leq 1$, and R^2 is a measure of the goodness of fit of the fitted values of the data, which can be written as

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (18)$$

$$= \frac{SSR}{SST} \quad (19)$$

$$= \frac{SST - SSE}{SST} \quad (20)$$

$$= 1 - \frac{SSE}{SST} \quad (21)$$

For example $R^2 = 0.75$ implies that 75% of the variation in the data can be explained by the model. Note that $R^2 = 1$ if and only if $SSE = 0$, in other words, all the residuals are 0 and the fitted value are the true values. We can also see from equation (21) that the smaller SSE, the bigger R^2 , which means the better fit of our regression model to the data.

The ΔR^2 method as introduced in [15] is

$$\Delta R^2(k+1) = R_{k+1}^2 - R_k^2 \quad (22)$$

We can see that this is just the change in the coefficient of determination increasing the number of regression variables from k to $k+1$. For a model with k regression variables, let

$$SSR(k) = \sum_{i=1}^n (\hat{y}_i^{(k)} - \bar{y})^2 \quad (23)$$

Then (20) becomes

$$\Delta R^2(k+1) = \frac{SSR(k+1) - SSR(k)}{SST} \quad (24)$$

$$= \frac{SSR(k+1) - SSR(k)}{\sigma^2} \cdot \frac{\sigma^2}{SST} \quad (25)$$

$\frac{SSR(k+1) - SSR(k)}{\sigma^2}$ follows a χ^2 -distribution with 1 degree of freedom under the assumption that the model with k regression variables is the correct one. Then, according to (25), $\Delta R^2(k+1) \cdot \frac{SST}{\sigma^2}$ follows a $\chi^2(1)$ distribution. σ^2 in this case is the variance of the observations and normally unknown. For the correct model $E(\frac{SSE}{n-k-1}) = \sigma^2$, we set $SSE = t \cdot SST$, $0 < t < 1$, then σ^2 depends on $\frac{t \cdot SST}{n-k-1}$, which is similar to the F-method. A new term can be entered under the assumption that the maximum value of $\Delta R^2(k+1) \geq \chi_{1,\alpha}^2 \cdot \frac{t}{n-k-1}$, or the procedure stops. Note that reasonable values for α and t need to be decided.

The ΔR^2 -method is based on using $\Delta R^2(k+1)$ in forward regression.

3.4 An example

We look at one of the examples used in [15] to illustrate how these criteria work.

In this example, a 12 run PB Design has been considered. The model is given by

$$Y = x_1 + x_3 + x_2x_3 + x_3x_4 + \varepsilon \quad (26)$$

where $\varepsilon \sim N(0, 0.7)$.

The design and the response simulated from the model are given in the Table 6.

In this case, 4 active factors are involved, which are x_1 , x_2 , x_3 and x_4 and the active effects

Table 6: The 12 Run Plackett-Burman Design and response values used in the example one.

Run	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	y
1	+	-	+	-	-	-	+	+	+	-	+	0.5456
2	+	+	-	+	-	-	-	+	+	+	-	-1.7307
3	-	+	+	-	+	-	-	-	+	+	+	0.4875
4	+	-	+	+	-	+	-	-	-	+	+	1.9211
5	+	+	-	+	+	-	+	-	-	-	+	-2.0272
6	+	+	+	-	+	+	-	+	-	-	-	2.0617
7	-	+	+	+	-	+	+	-	+	-	-	1.4472
8	-	-	+	+	+	-	+	+	-	+	-	0.9961
9	-	-	-	+	+	+	-	+	+	-	+	-1.9956
10	+	-	-	-	+	+	+	-	+	+	-	2.4805
11	-	+	-	-	-	+	+	+	-	+	+	-2.5985
12	-	-	-	-	-	-	-	-	-	-	-	-0.7527

are x_1 , x_3 , x_2x_3 and x_3x_4 correspondingly. The projection model for this situation involves x_1 , x_2 , x_3 , x_4 and their six two-factor interactions which are x_1x_2 , x_1x_3 , x_1x_4 , x_2x_3 , x_3x_4 .

Table 7 shows the values of the three criteria for entering terms in this particular projection model.

Table 7: The values of the $\Delta R^2(k+1)$, F -statistic and the AIC -criterion for entering terms in a projection model where $Y = x_1 + x_3 + x_2x_3 + x_3x_4 + \varepsilon$, $\varepsilon \sim N(0, 0.7)$. Data are obtained from a 12 Run PB design.

Terms entered	ΔR^2	F	AIC
x_3	0.458302	9.30652	7.84795
x_2x_3	0.101334	2.30115	7.36265
x_1	0.161188	5.19634	3.89354
x_3x_4	0.218949	29.0834	-12.5112
x_1x_3	0.014117	2.14321	-13.7164
x_2	0.011568	7.5466	-21.4889
x_2x_4	0.008648	3.67284	-26.098
x_4	0.006890	5.64417	-34.6587
x_1x_4	0.003385	6.77945	-46.8388
x_1x_2	0.0002794	0.45856	-47.3159

The three methods will enter terms in the same order and we expect to find the stopping value by looking for the maximum ΔR^2 , maximum F statistic and the minimum AIC . However, for the ΔR^2 method and F test, the procedure might stop too early in this case, but it seems more likely to happen for the F -test. It is clearly shown in Table 7 that the value of ΔR^2 and F decrease just after entering the term x_2x_3 . While the value for AIC decreases all the way until entering all the main factors and 2-factor interactions, since we look for the minimum AIC value, the procedure stops after entering all the terms.

Tyssedal and Hussain [15] explore further for the ΔR^2 method and F test and try to find the same stopping value and the best stopping value by choosing a panel of eight models reasonably. These models take into different situations account such as the number of active factors, the number of active effects, the number of interactions, type of interactions, the size of the active effects and whether the model obeys the heredity principles or not could effect the stopping value.

The eight models are as follows:

Modell:

$$Y_1 = x_1 + 2x_1x_2 + 2x_1x_3 + \varepsilon, \varepsilon \sim N(0, \sigma^2)$$

Model2:

$$Y_2 = x_1 + 1.5x_2 + 2x_3 + x_1x_2 + 1.5x_1x_3 + \varepsilon, \varepsilon \sim N(0, \sigma^2)$$

Model3:

$$Y_3 = x_1 + 1.5x_2 + 2x_3 + 1.5x_1x_2x_3 + \varepsilon, \varepsilon \sim N(0, \sigma^2)$$

Model4:

$$Y_4 = 2x_1 + x_2x_3 + \varepsilon, \varepsilon \sim N(0, \sigma^2)$$

Model5:

$$Y_5 = x_1 + x_3 + x_2x_3 + x_3x_4 + \varepsilon, \varepsilon \sim N(0, \sigma^2)$$

Model6:

$$Y_6 = 2x_1 + 3x_2 + 2x_3 + x_4 + 3x_3x_4 + \varepsilon, \varepsilon \sim N(0, \sigma^2)$$

Model7:

$$Y_7 = 4x_1 + x_2 + x_3 + x_4 + 2x_1x_4 + \varepsilon, \varepsilon \sim N(0, \sigma^2)$$

Model8:

$$Y_8 = 2x_1 + 4x_3 + 2x_2x_3 + 2x_3x_4 + \varepsilon, \varepsilon \sim N(0, \sigma^2)$$

The conclusion in [15] is that the number of active factors and the number of experimental factors both affect the performance of the procedures in negative way. And the type of model is important cause it is more difficult to do forward selection for the models with many terms of same size. However, it seems from the results of [15] that the type and size of interactions are not that important. As for the heredity, it may have an effect. Non-heredity models tend to have many factors which may open up for many possible explanations for the variation in the data. Furthermore, the variance of the observations, σ^2 also affect the performance of procedures.

When it comes to the three different criteria, considering the three active factors, F-Method performs well for the model1, 3 and 4, but not for model2. While ΔR^2 method and *AIC* method seem to be very reliable. As for the four active factors, the performance of the *AIC* method is very poor and worse than the *F* method, while ΔR^2 method performs quite well. Besides, model 5 is the worst among all the four active factors models due to the smallest capture probability. Overall, the ΔR^2 method is equal to or better than the *AIC* and the *F* method.

4 A new way for screening in a 12 run PB design with four active factors

We continue to consider the example in the last section, which is a 12 Run PB design. And the response is simulated from the model, $Y = x_1 + x_3 + x_2x_3 + x_3x_4 + \varepsilon, \varepsilon \sim N(0, 0.7)$. As we can see, four active factors are involved for this case. We look at all the $\binom{11}{4} = 330$ projections onto four factors by and try to do more analysis on the 12 Run PB design. The programming is done in R and shown in Appendix A.

In the example, σ^2 is given by 0.7, which seems large, we change the noise in the further investigation and then run 1000 simulations and it was counted the number of times the correct subset of active factors was contained in the candidate set with the 10 smallest MSE for each simulation. The number is defined as the capture frequency in [15]. Note that there is no need to restrict the number in the reduced set exactly to 10, but 10 is a reasonable compromise in order to not losing the right set. As for the MSE, it is known as the mean square error and measures the average of the squares of the error. In other words, it is the difference between the estimator and what is estimated, which is a measure of how well a candidate set is able to explain the variation in the response. In the regression model, the MSE can be written as:

$$MSE = \frac{SSE}{n - k - 1} \quad (27)$$

Where the SSE is known as the residual sum of squares.

$$SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (28)$$

Note that the response variable Y is a vector of n response values.

4.1 Suggested variable selection in certain situation

The specific algorithm we propose for variable selection has several steps.

1. Assume n_a active factors among all factors we have.
2. For each subset of n_a factors, make a design matrix with the constant column(1's) and all feasible main effects and two-factor interactions. Estimate this projection model to the response y .
3. Look at the absolute value of the coefficients, pick out the 6 ones with the highest absolute value.
4. Fit a model with the estimated constant coefficient and the 6 terms that correspond to the largest absolute value of the coefficients.
5. Compute the MSE, it is the difference between the estimated response in step 4 and the response y in step 2, definition is given by equation (27).
6. Repeat the step 1 to 5 for all the possible model selections of the n active factors.
7. Rank the MSE and make a list of the 10 smallest ones.

We will analyze the models with four active factors one by one.

Starting with the most difficult model from [15]

Model 5

$$Y = A + C + BC + CD + \varepsilon \quad (29)$$

where $\varepsilon \sim N(0, \sigma^2)$.

Note: We use A, B, C, D instead of the x_1, x_2, x_3, x_4 in [15], analogously, E, F, ..., K represent x_5, x_6, \dots, x_{11} .

First, we fix the response Y by using the value from Table 6 as taken from [15] where $\sigma^2 = 0.7$.

We get 10 candidate sets with the smallest MSE by our proposed method, the values for them are as follows:

Table 8: 10 candidate sets and their MSE values for the Model 5 with $\sigma^2 = 0.7$

Combinations	1	2	3	4	5	6	7	8	9	10
	B	D	B	A	A	C	A	D	A	G
	G	G	C	B	F	G	B	E	B	H
	I	H	J	H	H	H	C	J	H	I
	K	I	K	I	K	I	D	K	J	K
10 smallest MSE	0.073	0.132	0.166	0.194	0.235	0.248	0.257	0.277	0.288	0.314

Note again our proposed method in this thesis works with the 6 terms. We can see from Table 8 the combination A, B, C, D is among them which is compatible with the fact.

Change the value of σ^2 to find the "breakdown" value. The breakdown value means the capture frequency for $ABCD$ is less than 950, which means the corresponding capture probability is less than 0.95. Because it is meaningless to analyze the data when the capture frequency for the correct set is too low.

Table 9: 10 candidate sets and their capture frequency for the Model 5 with $\sigma^2 = 0.1$

Combinations	1	2	3	4	5	6	7	8	9	10
	G	C	A	C	B	D	F	F	A	A
	H	G	B	G	G	G	G	G	C	F
	I	H	C	I	I	H	I	H	G	G
	K	I	D	K	K	I	K	I	I	I
capture frequency	986	955	945	933	920	917	692	674	625	623

Table 10: 10 candidate sets and their capture frequency for the Model 5 with $\sigma^2 = 0.2$

Combinations	1	2	3	4	5	6	7	8	9	10
	G	C	A	C	B	D	F	F	A	A
	H	G	B	G	G	G	G	G	C	F
	I	H	C	H	I	H	H	I	G	G
	K	I	D	I	K	I	I	K	I	I
capture frequency	982	942	937	936	922	919	710	692	618	611

Table 11: 10 candidate sets and their capture frequency for the Model 5 with $\sigma^2 = 0.3$

Combinations	1	2	3	4	5	6	7	8	9	10
	G	A	C	C	B	D	F	F	C	A
	H	B	G	G	G	G	G	G	F	C
	I	C	H	I	I	H	I	H	G	G
	K	D	I	K	K	I	K	I	I	I
capture frequency	958	935	907	900	895	888	661	624	582	575

Table 12: 10 candidate sets and their capture frequency for the Model 5 with $\sigma^2 = 0.4$

Combinations	1	2	3	4	5	6	7	8	9	10
	G	A	C	C	D	B	F	F	A	A
	H	B	G	G	G	G	G	G	F	C
	I	C	I	H	H	I	H	I	G	G
	K	D	K	I	I	K	I	K	I	I
capture frequency	854	843	770	769	749	734	506	499	463	462

From the above Table 9-12, we observe that when σ^2 increases, the ratio of the success will go down, and the breakdown value 0.2 is suggested. Although the capture frequency for $ABCD$ is between 935 to 945 when σ^2 changes from 0.1 to 0.3. Considering that Model 5 is the most difficult model among all the eight models in the panel. In this case, we have lowered the criterion a little bit. Besides, no matter how the σ^2 has been changed, the value of

capture frequency for the combination "GHIK" is always bigger than the value for "ABCD", which implies that on some level "GHIK" explains the variation in the response better than "ABCD".

In order to do further analysis, firstly we have a look at the 10 combinations with smallest MSE when the response Y is simulated from model 5 with $\sigma^2 = 0$.

Table 13: 10 candidate sets and their MSE values for the Model 5 with $\sigma^2 = 0$

	Combinations	MSE
1	C F H K	$3.7717 \cdot 10^{-31}$
2	G H I K	$3.8923 \cdot 10^{-31}$
3	A C G I	$8.7514 \cdot 10^{-31}$
4	A F G I	$8.8007 \cdot 10^{-31}$
5	F G I K	$9.5280 \cdot 10^{-31}$
6	C G I K	$1.0769 \cdot 10^{-30}$
7	C G H I	$1.1070 \cdot 10^{-30}$
8	A B C D	$1.9151 \cdot 10^{-30}$
9	F G H I	$4.2907 \cdot 10^{-30}$
10	D F I K	$6.0000 \cdot 10^{-1}$

It is obvious to see from Table 13 that the first 9 candidate sets all have a MSE which almost equals 0, and "GHIK" is among them as we expect. Thus, we compute the effects and corresponding coefficients for each combination.

Table 14: The effects and the corresponding coefficients of the models which are potential equivalent to the Model 5

	Combinations	Effects	Corresponding Coefficients
1	C F H K	F C H K CH CK	1 1 -0.5 -0.5 0.5 0.5
2	G H I K	GI GK HI HK G IK	1.5 -0.5 -0.5 -0.5 0 0
3	A C G I	A C AC CG CI GI	-0.5 0.5 -0.5 -0.5 -0.5 1.5
4	A F G I	G I AF AG AI GI	0.5 0.5 1 -0.5 -0.5 1
5	F G I K	K I FI FK GI GK	-0.5 -0.5 -0.5 -0.5 1 -1
6	C G I K	C K CG GI GK I	0.5 -0.5 -0.5 1 -0.5 0
7	C G H I	C H CI HI GI CH	0.5 -0.5 -0.5 -0.5 1 0
8	A B C D	A C BC CD AB B	1 1 1 1 1 0 0
9	F G H I	H G FG FH GI HI	-0.5 -0.5 -0.5 -0.5 1 -1

Table 14 shows that there are some coefficients which equals 0 among the 2nd, 6th, 7th and 8th combinations. The final significant effects for each model are given in Table 15.

Table 15: The significant effects of the models which are potential equivalent to the Model 5

	Combinations	Effects
1	C F H K	F C H K CH CK
2	G H I K	GI GK HI HK
3	A C G I	A C AC CG CI GI
4	A F G I	G I AF AG AI GI
5	F G I K	K I FI FK GI GK
6	C G I K	C K CG GI GK
7	C G H I	C H CI HI GI
8	A B C D	A C BC CD
9	F G H I	H G FG FH GI HI

Then the potential alternative models are suggested as follows:

Model5.1:

$$Y_{5.1} = C + F - 0.5H - 0.5K + 0.5CH + 0.5CK + \varepsilon, \varepsilon \sim N(0, \sigma^2)$$

Model5.2:

$$Y_{5.2} = 1.5GI - 0.5GK - 0.5HI - 0.5HK + \varepsilon, \varepsilon \sim N(0, \sigma^2)$$

Model5.3:

$$Y_{5.3} = -0.5A + 0.5C - 0.5AC - 0.5CG - 0.5CI + 1.5GI + \varepsilon, \varepsilon \sim N(0, \sigma^2)$$

Model5.4:

$$Y_{5.4} = 0.5G + 0.5I + AF - 0.5AG - 0.5AI + GI + \varepsilon, \varepsilon \sim N(0, \sigma^2)$$

Model5.5:

$$Y_{5.5} = -0.5K - 0.5I - 0.5FI - 0.5FK + GI - GK + \varepsilon, \varepsilon \sim N(0, \sigma^2)$$

Model5.6:

$$Y_{5.6} = 0.5C - 0.5K - 0.5CG + GI - 0.5GK + \varepsilon, \varepsilon \sim N(0, \sigma^2)$$

Model5.7:

$$Y_{5.7} = 0.5C - 0.5H - 0.5CI - 0.5HI + GI + \varepsilon, \varepsilon \sim N(0, \sigma^2)$$

Model5.8:

$$Y_{5.8} = -0.5H - 0.5G - 0.5FH - 0.5FG - HI + GI + \varepsilon, \varepsilon \sim N(0, \sigma^2)$$

We can see from Table 15 that except the first combination "CFHK", the number of two-factor interactions is at least the half of the significant effects for other models.

Since the breakdown noise is after 0.3, let's have a look at the table from 9 to 11 whose noise is between 0.1 to 0.3, the capture frequency is broken down from the 7th combination for all of them, which seems unreasonable because there are 9 alternative models as noted earlier.

This example seems a bit complicated since we find eight equivalent models as the original model and also some problems stated above exist. It is also reasonable because this model is considered as the worst one in [15] when the other three criterias are performed.

The second example is Model 6, which is given by

$$Y = 2A + 3B + 2C + D + 3CD + \varepsilon \tag{30}$$

where $\varepsilon \sim N(0, \sigma^2)$

Note: We use A, B, C, D instead of the x_1, x_2, x_3, x_4 in [15], analogously, E, F, ..., K represent x_5, x_6, \dots, x_{11} .

Firstly, we change the value of the noise to find the breakdown value.

Table 16: 10 candidate sets and their capture frequency for the Model 6 with $\sigma^2=0.1$

Combinations	1	2	3	4	5	6	7	8	9	10
	B	B	A	B	A	C	B	A	B	C
	C	C	B	E	C	I	D	B	E	D
	I	E	C	F	I	J	E	D	G	I
	K	J	D	J	K	K	J	G	J	K
capture frequency	1000	1000	1000	730	730	713	707	534	533	517

Table 17: 10 candidate sets and their capture frequency for the Model 6 with $\sigma^2=0.2$

Combinations	1	2	3	4	5	6	7	8	9	10
	B	B	A	A	B	C	C	A	B	C
	C	C	B	B	E	E	D	B	D	I
	I	E	C	E	G	I	I	D	E	J
	K	J	D	J	J	K	K	J	J	K
capture frequency	1000	1000	1000	606	599	570	568	551	535	533

Table 18: 10 candidate sets and their capture frequency for the Model 6 with $\sigma^2=0.3$

Combinations	1	2	3	4	5	6	7	8	9	10
	B	B	A	C	A	B	A	C	B	B
	C	C	B	E	B	E	B	D	E	D
	I	E	C	I	D	G	E	I	F	E
	K	J	D	K	G	J	J	K	J	J
capture frequency	1000	1000	1000	574	566	558	554	552	455	446

Table 19: 10 candidate sets and their capture frequency for the Model 6 with $\sigma^2=0.4$

Combinations	1	2	3	4	5	6	7	8	9	10
	B	B	A	B	C	A	C	A	C	B
	C	C	B	E	D	B	E	B	I	D
	I	E	C	G	I	E	I	D	J	E
	K	J	D	J	K	J	K	G	K	J
capture frequency	1000	1000	1000	554	540	535	524	516	403	399

Table 20: 10 candidate sets and their capture frequency for the Model 6 with $\sigma^2=0.5$

Combinations	1	2	3	4	5	6	7	8	9	10
	B	B	A	A	C	A	B	C	B	C
	C	C	B	B	E	B	E	D	D	I
	E	I	C	E	I	D	G	I	E	J
	J	K	D	J	K	G	J	K	J	K
capture frequency	1000	998	998	520	507	504	493	487	362	350

Table 21: 10 candidate sets and their capture frequency for the Model 6 with $\sigma^2=0.6$

Combinations	1	2	3	4	5	6	7	8	9	10
	B	A	B	A	C	B	A	C	C	B
	C	B	C	B	E	E	B	D	I	D
	E	C	I	E	I	G	D	I	J	E
	J	D	K	J	K	J	G	K	K	J
capture frequency	988	986	984	474	468	444	444	432	337	304

Table 22: 10 candidate sets and their capture frequency for the Model 6 with $\sigma^2=0.7$

Combinations	1	2	3	4	5	6	7	8	9	10
	B	A	B	A	C	A	C	B	C	B
	C	B	C	B	E	B	D	E	I	D
	E	C	I	E	I	D	I	G	J	E
	J	D	K	J	K	G	K	J	K	J
capture frequency	955	951	946	460	450	442	441	418	316	305

Table 23: 10 candidate sets and their capture frequency for the Model 6 with $\sigma^2=0.8$

Combinations	1	2	3	4	5	6	7	8	9	10
	B	B	A	C	A	A	B	C	C	B
	C	C	B	E	B	B	E	D	I	D
	E	I	C	I	E	D	G	I	J	E
	J	K	D	K	J	G	J	K	K	J
capture frequency	889	887	862	412	400	398	385	372	312	278

Table 22 shows that capture frequency for "ABCD" is 951 which is still high when the $\sigma^2 = 0.7$, it is reduced to 862 when σ^2 comes to 0.8 in Table 23. Therefore, in this case the breakdown value is 0.8. And we also notice the combination "BCEJ" and "BCIK" both have the high capture frequency whenever the value of σ^2 changes from 0 to 0.7.

Similarly, we computer the 10 combinations with the smallest MSE when the response Y is simulated from the model 6 with $\sigma^2 = 0$.

Table 24: 10 candidate sets and their MSE values for the Model 6 with $\sigma^2 = 0$

	Combinations	MSE
1	A B C D	$6.2862 \cdot 10^{-30}$
2	B C E J	$6.7743 \cdot 10^{-30}$
3	B C I K	$8.5370 \cdot 10^{-30}$
4	B D E J	$1.0000 \cdot 10^{+00}$
5	C I J K	$1.0000 \cdot 10^{+00}$
6	C E I K	$1.0000 \cdot 10^{+00}$
7	B E F J	$1.0000 \cdot 10^{+00}$
8	C E F J	$1.1000 \cdot 10^{+00}$
9	B E G J	$1.1000 \cdot 10^{+00}$
10	C D I K	$1.1000 \cdot 10^{+00}$

Table 24 shows the values of MSE for first three models are close to 0. The effects and corresponding Coefficients are computed as follows:

Table 25: The effects and the corresponding coefficients of the models which are potential equivalent to the Model 6

	Combinations	Effects	Corresponding Coefficients
1	A B C D	A B C D CD AB	2 3 2 1 3 0
2	B C E J	B C E J BJ CJ	2 1 2 1 3 0
3	B C I K	B C I K IK BC	1 3 1 1 3 0

Ignoring the insignificant effects, the significant effects for each combination are shown below:

Table 26: The significant effects of the models which are potential equivalent to the Model 6

	Combinations	Effects
1	A B C D	A B C D CD
2	B C E J	B C E J BJ
3	B C I K	B C I K IK

Therefore, the equivalent models as the Model 6 are given by

Model6.1:

$$Y_{6.1} = 2B + C + 2E + J + 3BJ + \varepsilon, \varepsilon \sim N(0, \sigma^2)$$

Model6.2:

$$Y_{6.2} = B + 3C + I + K + 3IK + \varepsilon, \varepsilon \sim N(0, \sigma^2)$$

The result for this situation is much clearer than the previous example, no other combinations come up, only the combination "BCEJ" and "BCIK" among all the 330 candidate sets can also explain the variation in the response simulated from the Model 6. We also observe that Model 6.1 and Model 6.2 look quite similar as Model 6 in terms of the number and type of the active factors and interactions. In other words, four main effects and an interaction are involved in all these models.

The third example is Model 7 from [15], which is

$$Y = 4A + B + C + D + 2AD + \varepsilon \quad (31)$$

where $\varepsilon \sim N(0, \sigma^2)$.

Note: We use A, B, C, D instead of the x_1, x_2, x_3, x_4 in [15], analogously, E, F, ..., K represent x_5, x_6, \dots, x_{11} .

Similarly, change the value of noise to find the breakdown value.

Table 27: 10 candidate sets and their capture frequency for the Model 7 with $\sigma^2 = 0.1$

Combinations	1	2	3	4	5	6	7	8	9	10
	A	A	A	A	A	D	A	A	A	A
	H	D	D	B	B	E	D	D	C	E
	I	I	E	I	C	I	J	E	D	H
	K	K	J	K	D	J	K	I	E	K
capture frequency	1000	1000	1000	1000	1000	997	612	516	279	269

Table 28: 10 candidate sets and their capture frequency for the Model 7 with $\sigma^2 = 0.2$

Combinations	1	2	3	4	5	6	7	8	9	10
	A	A	A	A	A	D	A	A	A	A
	B	D	B	D	H	E	D	D	E	B
	C	I	I	E	I	I	E	J	H	D
	D	K	K	J	K	J	I	K	K	E
capture frequency	1000	997	997	994	979	874	447	416	250	236

Table 29: 10 candidate sets and their capture frequency for the Model 7 with $\sigma^2 = 0.3$

Combinations	1	2	3	4	5	6	7	8	9	10
	A	A	A	A	A	D	A	A	A	A
	B	D	D	B	H	E	D	D	C	C
	C	I	E	I	I	I	J	E	D	D
	D	K	J	K	K	J	K	I	J	K
capture frequency	1000	966	958	936	886	723	395	389	289	276

Table 30: 10 candidate sets and their capture frequency for the Model 7 with $\sigma^2 = 0.4$

Combinations	1	2	3	4	5	6	7	8	9	10
	A	A	A	A	A	D	A	A	A	A
	B	D	B	D	H	E	D	B	C	B
	C	I	I	E	I	I	E	D	D	D
	D	K	K	J	K	J	I	I	K	E
capture frequency	999	891	873	827	783	544	395	309	308	305

Table 31: 10 candidate sets and their capture frequency for the Model 7 with $\sigma^2 = 0.5$

Combinations	1	2	3	4	5	6	7	8	9	10
	A	A	A	A	A	D	A	A	A	A
	B	D	B	H	D	E	D	C	C	B
	C	I	I	I	E	I	E	D	D	D
	D	K	K	K	J	J	I	J	K	E
capture frequency	993	826	804	700	685	431	373	313	298	298

Table 32: 10 candidate sets and their capture frequency for the Model 7 with $\sigma^2 = 0.6$

Combinations	1	2	3	4	5	6	7	8	9	10
	A	A	A	A	A	D	A	A	A	A
	B	D	B	D	H	E	D	C	B	C
	C	I	I	E	I	I	E	D	D	D
	D	K	K	J	K	J	I	K	I	J
capture frequency	979	740	722	636	624	383	363	338	338	291

Table 33: 10 candidate sets and their capture frequency for the Model 7 with $\sigma^2 = 0.7$

Combinations	1	2	3	4	5	6	7	8	9	10
	A	A	A	A	A	A	A	D	A	A
	B	D	B	H	D	D	C	E	C	B
	C	I	I	I	E	E	D	I	D	D
	D	K	K	K	J	I	K	J	J	I
capture frequency	904	684	662	571	496	344	336	311	295	289

The breakdown value 0.7 for $ABCD$ is suggested. We observe the candidate sets $ADIK$, $ABIK$ and $ADEJ$ have the high capture frequency when σ^2 ranges from 0.1 to 0.3, while their capture frequencies go down quickly after $\sigma^2 = 0.4$.

We compute 10 combinations with the smallest MSEs for this situation to see if these three combinations are able to be the equivalent models as Model 7. Note the response Y are simulated from the model 7 with $\sigma^2 = 0$.

Table 34: 10 candidate sets and their MSE values for the Model 7 with $\sigma^2 = 0$

	Combinations	MSE
1	A B C D	$6.3922 \cdot 10^{-30}$
2	A B I K	$6.0000 \cdot 10^{-01}$
3	A D I K	$6.0000 \cdot 10^{-01}$
4	D E I J	$6.0000 \cdot 10^{-01}$
5	A D E J	$6.0000 \cdot 10^{-01}$
6	A H I K	$6.0000 \cdot 10^{-01}$
7	A D J K	$9.0000 \cdot 10^{-01}$
8	A B D K	$1.0000 \cdot 10^{+00}$
9	A E H K	$1.0000 \cdot 10^{+00}$
10	A B D J	$1.0000 \cdot 10^{+00}$

We observe from Table 34 that there is no model equivalent to the original model since the closest MSE value is 0.6. Thus, this example is quite simple.

The last example is Model 8, which is

$$Y = 2A + 4C + 2BC + 2CD + \varepsilon \tag{32}$$

where $\varepsilon \sim N(0, \sigma^2)$.

Note: We use A, B, C, D instead of the x_1, x_2, x_3, x_4 in [15], analogously, E, F, ..., K represent x_5, x_6, \dots, x_{11} .

Firstly, change the noise to see the breakdown value.

Table 35: 10 candidate sets and their capture frequency for the Model 8 with $\sigma^2=0.1$

Combinations	1	2	3	4	5	6	7	8	9	10
	C	C	C	C	A	A	C	B	C	B
	G	G	F	F	C	B	D	C	D	C
	I	H	H	G	G	C	E	H	F	G
	K	I	K	I	I	D	K	J	K	H
capture frequency	1000	1000	1000	1000	1000	1000	269	251	242	239

Table 36: 10 candidate sets and their capture frequency for the Model 8 with $\sigma^2=0.2$

Combinations	1	2	3	4	5	6	7	8	9	10
	C	C	C	C	A	A	C	B	C	B
	G	G	F	F	C	B	D	C	D	C
	I	H	H	G	G	C	E	H	F	G
	K	I	K	I	I	D	K	J	K	H
capture frequency	1000	1000	1000	1000	1000	1000	269	251	242	239

Table 37: 10 candidate sets and their capture frequency for the Model 8 with $\sigma^2=0.3$

Combinations	1	2	3	4	5	6	7	8	9	10
	C	C	C	C	A	A	B	C	B	C
	G	G	F	F	C	B	C	D	C	D
	I	H	H	G	G	C	H	E	F	F
	K	I	K	I	I	D	J	K	H	K
capture frequency	1000	1000	1000	1000	1000	1000	265	258	243	240

Table 38: 10 candidate sets and their capture frequency for the Model 8 with $\sigma^2=0.4$

Combinations	1	2	3	4	5	6	7	8	9	10
	C	C	C	C	A	A	C	C	B	C
	G	G	F	F	B	C	D	D	C	D
	I	H	H	G	C	G	E	F	H	E
	K	I	K	I	D	I	K	K	J	H
capture frequency	1000	1000	1000	1000	1000	998	283	256	249	235

Table 39: 10 candidate sets and their capture frequency for the Model 8 with $\sigma^2=0.5$

Combinations	1	2	3	4	5	6	7	8	9	10
	C	C	A	A	C	C	C	B	B	C
	G	G	B	C	F	F	D	C	C	D
	I	H	C	G	H	G	E	H	F	F
	K	I	D	I	K	I	K	J	H	K
capture frequency	1000	1000	1000	996	992	992	256	255	246	235

Table 40: 10 candidate sets and their capture frequency for the Model 8 with $\sigma^2=0.6$

Combinations	1	2	3	4	5	6	7	8	9	10
	A	C	C	C	C	A	B	C	C	B
	B	G	G	F	F	C	C	D	D	C
	C	H	I	G	H	G	H	E	F	F
	D	I	K	I	K	I	J	K	K	H
capture frequency	997	994	989	970	960	959	249	235	221	217

Table 41: 10 candidate sets and their capture frequency for the Model 8 with $\sigma^2=0.7$

Combinations	1	2	3	4	5	6	7	8	9	10
	A	C	C	A	C	C	C	B	B	C
	B	G	G	C	F	F	D	C	C	D
	C	I	H	G	G	H	E	H	G	I
	D	K	I	I	I	K	K	J	H	K
capture frequency	988	970	968	904	893	891	247	246	239	223

Table 42: 10 candidate sets and their capture frequency for the Model 8 with $\sigma^2=0.8$

Combinations	1	2	3	4	5	6	7	8	9	10
	A	C	C	C	C	A	C	B	C	C
	B	G	G	F	F	C	D	C	D	D
	C	I	H	G	H	G	E	H	I	E
	D	K	I	I	K	I	K	J	K	H
capture frequency	955	920	912	795	791	788	259	241	241	214

Table 43: 10 candidate sets and their capture frequency for the Model 8 with $\sigma^2=0.9$

Combinations	1	2	3	4	5	6	7	8	9	10
	A	C	C	A	C	C	B	C	C	B
	B	G	G	C	F	F	C	D	D	C
	C	I	H	G	G	H	H	F	E	G
	D	K	I	I	I	K	J	K	K	H
capture frequency	915	859	844	698	684	665	236	228	221	220

Table 35-43 above suggest that the breakdown value for the noise is 0.9 since the capture frequency for "ABCD" goes down to 915 from 955 when the noise changes from 0.9 to 0.8.

Similarly, the 10 combinations with the smallest MSEs are computed in Table 44 for this situation where the response Y is simulated from the Model 8 with $\sigma^2 = 0$

Table 44: 10 candidate sets and their MSE values for the Model 8 with $\sigma^2 = 0$

	Combinations	MSE
1	C G H I	$8.2436 \cdot 10^{-30}$
2	A B C D	$9.8805 \cdot 10^{-30}$
3	C G I K	$1.0788 \cdot 10^{-29}$
4	C F H K	$1.0886 \cdot 10^{-29}$
5	C F G I	$1.5432 \cdot 10^{-29}$
6	A C G I	$1.6043 \cdot 10^{-29}$
7	C F G H	$2.4000 \cdot 10^{+00}$
8	B C G H	$2.4000 \cdot 10^{+00}$
9	B C F H	$2.4000 \cdot 10^{+00}$
10	C D F K	$2.4000 \cdot 10^{+00}$

Considering the first 6 models whose values of MSE are close to 0. The effects and corresponding coefficients are shown in below table.

Table 45: The effects and the corresponding coefficients of the models which are potential equivalent to the Model 8

	Combinations	Effects	Corresponding Coefficients
1	C G H I	C H CI GI HI CH	3 1 1 2 1 0
2	A B C D	A C BC CD D BD	2 4 2 2 0 0
3	C G I K	C K CG GI GK CI	3 1 1 2 1 0
4	C F H K	C F H K CH CK	4 2 1 1 1 1
5	C F G I	C CG CI FG FI GI	4 1 1 1 1 2
6	A C G I	A C AC CG CI GI	1 3 1 1 1 3

Ignoring the insignificant effects whose coefficients are quite small and very close to 0.

Table 46: The significant effects of the models which are potential equivalent to the Model 8

	Combinations	Effects
1	C G H I	C H CI GI HI
2	A B C D	A C BC CD
3	C G I K	C K CG GI GK
4	C F H K	C F H K CH CK
5	C F G I	C CG CI FG FI GI
6	A C G I	A C AC CG CI GI

Thus, the equivalent models as the Model 8 are as follows:

Model8.1:

$$Y_{8,1} = 3C + H + CI + 2GI + HI + \varepsilon, \varepsilon \sim N(0, \sigma^2)$$

Model8.2:

$$Y_{8,2} = 3C + K + CG + 2GI + GK + \varepsilon, \varepsilon \sim N(0, \sigma^2)$$

Model8.3:

$$Y_{8,3} = 4C + 2F + H + K + CH + CK + \varepsilon, \varepsilon \sim N(0, \sigma^2)$$

Model 8.4:

$$Y_{8.4} = 4C + CG + CI + FG + FI + 2GI + \varepsilon, \varepsilon \sim N(0, \sigma^2)$$

Model 8.5:

$$Y_{8.5} = A + 3C + AC + CG + CI + 3GI + \varepsilon, \varepsilon \sim N(0, \sigma^2)$$

We notice the capture frequency for the first 6 combinations in Table 35 to 40 has a big difference to the last 4 ones and these 6 selections are perfectly consistent with the first 6 ones Table 44 without considering the order.

Also, we find that the capture frequency of combination "ABCD" is going down slowly as the noise increases. After the noise value larger than 0.6, its capture frequency becomes the highest. However, "C F G I", "C F H K" and "A C G I", the capture frequency of them started to go down after the noise value is 0.4. We compute the variance-covariance matrix between the "Effects" of each model which is defined as $(X^T X)^{-1}$ to see if they are less efficient. For experimental designs, efficiency is a measure of quality of a estimator.

Model 8.1: combination "C G H I"

Table 47: variance-covariance matrix for "CGHI"

	1	C	GI	H	HI	CI	CH
1	0.083333	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000
C	0.000000	0.11209239	-0.04891304	-0.01834239	0.037364130	-0.0061141	0.0163043
GI	0.000000	-0.04891304	0.13043478	0.04891304	-0.016304348	0.0163043	-0.0434783
H	0.000000	-0.01834239	0.04891304	0.11209239	-0.006114130	0.0373641	-0.0163043
HI	0.000000	0.03736413	-0.01630435	-0.00611413	0.095788043	-0.0020380	0.0054348
CI	0.000000	-0.00611413	0.01630435	0.03736413	-0.002038043	0.0957880	-0.0054348
CH	0.000000	0.01630435	-0.04347826	-0.01630435	0.005434783	-0.0054348	0.0978261

Model 8.2: combination "A B C D"

Table 48: variance-covariance matrix for "ABCD"

	1	C	A	BC	CD	D	BD
1	0.0833333	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000
C	0.0000000	0.095788043	-0.01630435	-0.00611413	-0.0054348	-0.0020380	0.0373641
A	0.0000000	-0.016304348	0.13043478	0.04891304	0.0434783	0.0163043	-0.0489130
BC	0.0000000	-0.006114130	0.04891304	0.11209239	0.0163043	0.0373641	-0.0183424
CD	0.0000000	-0.005434783	0.04347826	0.01630435	0.0978261	0.0054348	-0.0163044
D	0.0000000	-0.002038043	0.01630435	0.03736413	0.0054348	0.0957880	-0.0061141
BD	0.0000000	0.037364130	-0.04891304	-0.01834239	-0.0163043	-0.0061141	0.1120924

Model 8.3:combination "C G I K"

Table 49: variance-covariance matrix for "CGIK"

	1	C	GI	GK	CG	K	CI
1	0.083333333	0.00000000	0.0000000	0.00000000	0.00000000	0.00000000	0.0000000
C	0.000000000	0.11015625	-0.040625	0.03984375	-0.00390625	-0.01171875	-0.009375
GI	0.000000000	-0.04062500	0.112500	-0.00937500	0.01562500	0.04687500	-0.012500
GK	0.000000000	0.03984375	-0.009375	0.11015625	0.00390625	0.01171875	-0.040625
CG	0.000000000	-0.00390625	0.015625	0.00390625	0.09765625	0.04296875	-0.015625
K	0.000000000	-0.01171875	0.046875	0.01171875	0.04296875	0.12890625	-0.046875
CI	0.000000000	-0.00937500	-0.012500	-0.04062500	-0.01562500	-0.04687500	0.112500

Model 8.4:combination "C F H K"

Table 50: variance-covariance matrix for "CFHK"

	1	C	F	H	CH	CK	K
1	0.083333333	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000
C	0.000000000	0.083333333	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000
F	0.000000000	0.00000000	0.11111111	0.01388889	0.041666667	0.0416667	0.0138889
H	0.000000000	0.00000000	0.01388889	0.095486111	0.005208333	0.0364583	0.0017361
CH	0.000000000	0.00000000	0.041666667	0.005208333	0.109375000	0.0156250	0.0364583
CK	0.000000000	0.00000000	0.041666667	0.036458333	0.015625000	0.1093750	0.0052083
K	0.000000000	0.00000000	0.01388889	0.001736111	0.036458333	0.0052083	0.0954861

Model 8.5: combination "C F G I"

Table 51: variance-covariance matrix for "CFG I"

	1	C	GI	FI	CG	FG	CI
1	0.0833333	0.00000000	0.000000000	0.00000000	0.00000000	0.00000000	0.00000000
C	0.0000000	0.13043478	-0.043478261	0.04891304	-0.0163043	0.0489130	-0.0163043
GI	0.0000000	-0.04347826	0.097826087	-0.01630435	0.0054348	-0.0163044	0.0054348
FI	0.0000000	0.04891304	-0.016304348	0.11209239	-0.0373641	0.0183424	-0.0061141
CG	0.0000000	-0.01630435	0.005434783	-0.03736413	0.0957880	-0.0061141	0.0020380
FG	0.0000000	0.04891304	-0.016304348	0.01834239	-0.0061141	0.1120924	-0.0373641
CI	0.0000000	-0.01630435	0.005434783	-0.00611413	0.0020380	-0.0373641	0.0957880

Model 8.6: combination "A C G I"

Table 52: variance-covariance matrix for "ACGI"

	1	GI	C	AC	CG	CI	A
1	0.083333333	0.000000	0.000000000	0.000000000	0.000000000	0.000000000	0.000000
GI	0.000000000	0.13125	-0.04375000	-0.04375000	-0.01875000	-0.01875000	-0.05625
C	0.000000000	-0.04375	0.09791667	0.01458333	0.00625000	0.00625000	0.01875
AC	0.000000000	-0.04375	0.01458333	0.09791667	0.00625000	0.00625000	0.01875
CG	0.000000000	-0.01875	0.00625000	0.00625000	0.09791667	0.01458333	0.04375
CI	0.000000000	-0.01875	0.00625000	0.00625000	0.01458333	0.09791667	0.04375
A	0.000000000	-0.05625	0.01875000	0.01875000	0.04375000	0.04375000	0.13125

It's hard to explain from the above tables why the last three models which are "C F H K", "C F G I" and "A C G I" are less efficiency than the first three ones.

Since the capture frequencies for all models we got above is random, there will be slightly difference every time we repeat it.

Let's have a look at the variance of capture probability p which is the corresponding proportion of the capture frequency. In general, for an indicator function I , we have

$$I = \begin{cases} 0 & \text{with probability } 1-p \\ 1 & \text{with probability } p \end{cases}$$

$$\sigma_{\hat{p}} = \sqrt{Var(\hat{p})} = \sqrt{\frac{p(1-p)}{n}} \quad (33)$$

A capture probability in the range 0.8-0.99 gives a standard deviation in the range 1.26%-0.31%. The largest standard deviation corresponds to a capture probability equal to 1/2 and is 1.58%.

We have a look at the D criteria which is one of popular efficiency criteria to be used to check the design goodness. In this case, the parameter-estimator β is a vector and its variance is a matrix. The inverse matrix of the variance-matrix $(X^T X)^{-1}$ is called the information matrix. D-optimality seeks to minimize $(X^T X)^{-1}$, or equivalently maximize the determinant of the information matrix of the design. Thus, it can be defined as

$$D = |X^T X| \quad (34)$$

We compute the D value for each model.

Model1(CGHI):D=18087936

Model2(ABCD):D=18087936

Model3(CGIK):D=15728640

Model4(CFHK):D=21233664

Model5(CFGI):D=18087936

Model6(ACGI):D=17694720

It is obvious to show that Model 4, combination *CFHK* has the maximum D value, which suggests it has the highest efficiency. While this combination is one of the three whose capture frequency started to go down quickly after the noise value 0.4, which seems less efficient. Thus, the D criteria cannot explain this situation well.

We continue to look at A criteria which seeks to minimize the trace of the inverse of the information matrix. This criterion results in minimizing the average variance of the estimates of the regression coefficients. It is defined as

$$A = tr(X^T X)^{-1} \quad (35)$$

Model1(CGHI):A=0.7273551

Model2(ABCD):A=0.7273551

Model3(CGIK):A=0.7552083

Model4(CFHK):A=0.6875

Model5(CFGI):A=0.7273551

Model6(ACGI):A=0.7375

Still, Model 4, the combination *CFHK* has the minimum A value, which suggests it has the highest efficiency.

Therefore, in this paper, we cannot find the obvious characteristic for those equivalent models as the Model 8, perhaps it happens just because of the randomness. After all, there

are totally $\binom{11}{4} = 330$ projections onto four dimensions for the 12 run PB design. We reduced it to five models in this case. Of course the correct set $ABCD$ is among them and it always has the equal or higher capture probability than the other nine candidate sets no matter how the noise varies between 0.1 and 1.0.

4.2 Comparing the results

Although we still don't get very high capture probability for Model 5 with our suggested method, Table 53 shows the capture frequency for Model 5 is between 935 to 945 when σ^2 varies from 0.1 to 0.3, which is still a satisfied result.

As for Model 6, Model 7 and Model 8, Table 54, 55 suggest our suggested method performs better than ΔR^2 -method, F method and AIC method since higher values of σ^2 are obtained.

Note that the result about ΔR^2 -method, F method and AIC method are taken from [15].

Table 53: The capture frequencies for Model 5-8 with different values of σ^2

$\sigma^2 \backslash$ Models	Model5	Model6	Model7	Model8
0.1	945	1000	1000	1000
0.2	937	1000	1000	1000
0.3	935	1000	1000	1000
0.4	843	1000	999	1000
0.5	674	998	993	1000
0.6	530	986	979	989
0.7	354	951	904	988
0.8	273	862	834	955
0.9	188	789	716	915
1.0	138	677	627	850

Table 54: Values of σ^2 for the ΔR^2 -Method, F -Method, AIC -Method and our new method for which we have capture probability of 1, first line, and ≥ 0.95 , second line, for Model 5 and 6.

Methods Capture Probabilities		Model5				Model6			
		New Method	ΔR^2	F	AIC	New Method	ΔR^2	F	AIC
1		-	-	-	-	0.4	0.3	-	-
≥ 0.95		0.1	-	-	-	0.7	0.6	0.2	-

Table 55: Values of σ^2 for the ΔR^2 -Method, F -Method, AIC -Method and our new method for which we have capture probability of 1, first line, and ≥ 0.95 , second line, for Model 7 and 8.

Methods Capture Probabilities		Model7				Model8			
		New Method	ΔR^2	F	AIC	New Method	ΔR^2	F	AIC
1		0.3	0.3	0.3	-	0.5	-	-	-
≥ 0.95		0.6	0.5	0.5	-	0.8	0.2	0.2	-

5 A real example

Xu, Phoa and Wong [11] reanalysed data from three chemical experiments and their purpose was to demonstrate the shortcomings of the traditional analysis for 12 run PB design where the interactions are ignored. In each experiment, they have discovered important but missed interactions in the traditional analysis and the result shows all the new models they suggested have improved fit to the data. We will now use the method suggested in this paper to analyse the third of these examples where a two-step analytical methodology was developed to allow the chemical characterization of white grapes by simultaneously determining their most important phenolic compound and organic acids. In this experiment, a 12 Run PB design was used to select 8 factors. Factors and levels are given in Table 56 as taken from [11].

Table 56: Factors and levels for the Compound extraction experiment

Symbol	Factor	Unit	Low(-)	high(+)
A	Extraction solvent		Acid Water	MeOH
B	Extraction volume	mL	50	250
C	Extraction time	min	5	20
D	Temperature	°C	40	50
E	Extraction type		Ultrasonic	Stirring
F	Sorbent type		EC	NEC
G	Elution solvent		EtOH	MeOH
H	Elution volume	mL	20	150

We notice the design matrix and responses in their example are allocated to columns in the same way as the one we use in this paper given in Table 3. Design matrix and the corresponding response value are given in Table 57 as taken from [11].

Table 57: Design matrix and response for the Compound extraction experiment

Run	A	B	C	D	E	F	G	H	Y
1	+	-	+	-	-	-	+	+	6.98
2	+	+	-	+	-	-	-	+	5.31
3	-	+	+	-	+	-	-	-	9.67
4	+	-	+	+	-	+	-	-	6.45
5	+	+	-	+	+	-	+	-	5.23
6	+	+	+	-	+	+	-	+	5.34
7	-	+	+	+	-	+	+	-	4.03
8	-	-	+	+	+	-	+	+	3.76
9	-	-	-	+	+	+	-	+	2.10
10	+	-	-	-	+	+	+	-	2.65
11	-	+	-	-	-	+	+	+	7.40
12	-	-	-	-	-	-	-	-	7.14

Assuming 4 active factors among 8 factors, then there are totally $\binom{8}{4} = 70$ projections onto four dimensions. Table 58 shows 10 candidate sets with the smallest MSE obtained by our suggested method. Clearly, first set $ACDF$ has a much smaller MSE comparing with others. We will explore it further.

Besides, we observe from Table 58 that factor A, C, D are contained in five of the candidate sets with the smallest MSE and will be considered for further exploration.

We compute their corresponding models with 6 terms separately shown in Table 59.

Table 58: 10 candidate sets and their MSE values for the Compound extraction experiment with the response given in Table 57

	Combinations	MSE
1	A C D F	0.0721
2	A B G H	0.1738
3	A C D G	0.2117
4	D E G H	0.2217
5	A C D E	0.2681
6	B E G H	0.3452
7	A C D H	0.3873
8	A B C D	0.3955
9	A C F G	0.5030
10	A D E H	0.6231

Table 59: The effects and the corresponding coefficients of 10 candidate sets for compound extraction experiment

	Combinations	Effects	Corresponding Coefficients
1	A C D F	AD C D AF F A	1.69 1.22 -1.11 -0.42 -0.28 -0.17
2	A C D G	AD C D CD G CG	2.01 1.24 -1.09 -0.47 0.37 0.33
3	A C D E	AD C D CD E DE	1.65 1.17 -1.14 -0.47 -0.39 0.31
4	A C D H	AD C D A CD DH	1.77 1.01 -0.97 -0.36 -0.29 -0.20
5	A B C D	AD C D A CD BD	1.63 1.15 -1.04 -0.58 -0.50 0.35

We obtain the following models with 6 effects with the highest absolute value by our proposed method. Note that the intercept equals $5.505 \approx 5.51$.

Model CE 1(A C D F):

$$\hat{Y}_{CE1} = 5.51 + 1.22C - 1.11D + 1.69AD - 0.28F - 0.42AF - 0.17A$$

Model CE 2(A C D G):

$$\hat{Y}_{CE2} = 5.51 + 1.24C - 1.09D + 2.01AD - 0.47CD + 0.37G + 0.33CG$$

Model CE 3(A C D E):

$$\hat{Y}_{CE3} = 5.51 + 1.17C - 1.14D + 1.65AD - 0.47CD - 0.39E + 0.31DE$$

Model CE 4(A C D H):

$$\hat{Y}_{CE4} = 5.51 + 1.01C - 0.97D + 1.77AD - 0.36A - 0.29CD - 0.20DH$$

Model CE 5(A B C D):

$$\hat{Y}_{CE5} = 5.51 + 1.15C - 1.04D + 1.63AD - 0.58A - 0.50CD + 0.35BD$$

We perform ordinary regression on all the models. For Model CE 1, $R^2 = 0.995$, the residual standard error $\hat{\sigma} = 0.234$, the effect C, D, AD, AF are significant at a 1% level and we observe AD and AF in this model do not obey the heredity principle. In the four other models, we notice that effects C, D, AD are the only significant effects at a 1% level and R^2 is between 0.970 to 0.993 as given in Table 66. Therefore, we arrive at the same active effects, C (Extraction time), D (Temperature) and interaction AD as in [11], but allowing four to be active. We also found that AF was significant at a 1% level.

We observe from Table 58 that there are some candidate sets such as $ABGH, DEGH$ that can explain the variation well for this response.

This example was to demonstrate how 70 possible candidate sets could efficiently be reduced to five by our proposed method and then find the active factors by doing the ordinary regression on them.

Let us now change the observed response to $Y + 1.5F$, where Y is the response from Table 57 and F is an artificial main effect.

Table 60: 10 candidate sets and their MSE values for the Compound extraction experiment with the response simulated from $Y + 1.5F$, Y is given by Table 57.

	Combinations	MSE
1	A C D F	0.0721
2	C E G H	0.2397
3	C D G H	0.2877
4	A C G H	0.2913
5	A C E G	0.3157
6	B D E H	0.4882
7	A C D H	0.5017
8	A C F G	0.5030
9	A B C F	0.5058
10	C D E G	0.5251

We observe from Table 60 that the combination "ACDF" is still the one with the smallest

MSE, the corresponding model with 6 terms is given by Model NEW:

$$\hat{Y}_{NEW} = 5.51 + 1.69AD + 1.22F + 1.22C - 1.11D - 0.42AF - 0.17A$$

Since the new response is the old response plus $1.5F$, the only difference between Model CE 1 and Model NEW is the coefficient of the effect "F". Doing ordinary regression with the Model NEW, this model has $R^2 = 0.994$ and the residual standard error $\hat{\sigma} = 0.234$. All effects except "A" in the Model NEW are significant at a 1% level. Thus, when the response in this experiment has an added $1.5F$ where F is one of the factors, the significant effects become C , D , AD , F and AF .

Hence, our algorithm was able to detect a model with four active factors and five active effects in this situation.

6 Concluding Remarks

In this paper, we have tested the ability of a projection based methods on factor screening. We have only considered the situation where the number of candidate sets of active factors is 11. The testing has been performed on the 12 run PB designs using four different models with four active factors taken from [15]. The procedure is based on using projection models with main effects and two-factor interactions included. Thus, every projection model has 11 terms including the constant. The number of terms in the projection models is reduced to 6 by picking up the 6 highest absolute value of coefficients. The aim of reducing the terms is to simplify the alias relationship due to the complexity of confounding pattern of 12 Run PB design.

Measured in capture frequency and corresponding capture probability, it turns out our suggested method have capture probabilities equal to one or close to one for quite noisy data. In other words, the breakdown value of the noise is quite high except for model 5, see Table 65 in Appendix B. Higher values of σ^2 are observed by the suggested method than the values obtained by ΔR^2 -method, F-method and AIC method for model 6, 7 and 8 when the capture is 1 or ≥ 0.95 . As for the model 5, although no values of σ^2 is obtained to reach the value of at least 0.95 for capture probability by all the methods mentioned above, our suggested method got it between 0.935 and 0.945 when σ^2 varies from 0.1 to 0.3, which are quite close to 0.95. Overall, the suggested method performs better than those three criteria for screening four active factors for 12 Run PB design.

Furthermore, we found some equivalent models for each tested model among 330 projection ones especially the model 5 and model 8 which have 8 and 5 equivalent models separately. We tried to find out the reason by looking at the information matrix, the results didn't show any valuable information. However, an assumption came out that the number of equivalent models depends on the type of model, more two-factor interactions are more complicated or the ratio of the number of main effects to the number of interactions in tested models bigger, more equivalent models. Or the appearance of equivalent models is just due to the

randomness. After all, 5 or 8 among 330 projection models is just a small number. The topic is still open for more research.

References

- [1] George Box and John Tyssedal. “Projective properties of certain orthogonal arrays”. In: *Biometrika* 83.4 (1996), pp. 950–955.
- [2] GEP Box and R Daniel Meyer. “Finding the active factors in fractionated screening experiments”. In: *Journal of Quality Technology* 25 (1993), pp. 94–94.
- [3] Ching-Shui Cheng. “Some projection properties of orthogonal arrays”. In: *The Annals of Statistics* (1995), pp. 1223–1233.
- [4] Hugh Chipman, Michael Hamada, and CFJ Wu. “A Bayesian variable-selection approach for analyzing designed experiments with complex aliasing”. In: *Technometrics* 39.4 (1997), pp. 372–381.
- [5] H Evangelaras and C Koukouvinos. “On generalized projectivity of two-level screening designs”. In: *Statistics & probability letters* 68.4 (2004), pp. 429–434.
- [6] Michael Hamada and CF Jeff Wu. “Analysis of designed experiments with complex aliasing”. In: *Journal of Quality Technology;(United States)* 24.3 (1992).
- [7] Murat Kulahci and George EP Box. “Catalysis Of Discovery And Development In Engineering And Industry: QUALITY QUANDARIES”. In: *Quality Engineering* 15.3 (2003), pp. 513–517.
- [8] Russell V Lenth. “Quick and easy analysis of unreplicated factorials”. In: *Technometrics* 31.4 (1989), pp. 469–473.
- [9] Dennis KJ Lin and Norman R Draper. “Generating alias relationships for two-level Plackett and Burman designs”. In: *Computational statistics & data analysis* 15.2 (1993), pp. 147–157.
- [10] Dennis KJ Lin and Norman R Draper. “Projection Properties of Plackett and Burrnan Designs”. In: *Technometrics* 34.4 (1992), pp. 423–428.
- [11] Frederick KH Phoa, Weng Kee Wong, and Hongquan Xu. “The need of considering the interactions in the analysis of screening designs”. In: *Journal of Chemometrics* 23.10 (2009), pp. 545–553.

- [12] Robin L Plackett and J Peter Burman. “The design of optimum multifactorial experiments”. In: *Biometrika* 33.4 (1946), pp. 305–325.
- [13] Oddgeir Samset and John Tyssedal. “Two-level designs with good projection properties”. In: *Department of Mathematical Sciences, The Norwegian University of Science and Technology, Norway* (1999).
- [14] John Tyssedal, Hallgeir Grinde, and Carl C Røstad. “The Use of a 12-run Plackett–Burman Design in the Injection Moulding of a Technical Plastic Component”. In: *Quality and Reliability Engineering International* 22.6 (2006), pp. 651–657.
- [15] John Tyssedal and Shahrukh Hussain. “Factor screening in nonregular two-level designs based on projection-based variable selection”. In: *Journal of Applied Statistics* 43.3 (2016), pp. 490–508.
- [16] JC Wang and CF Jeff Wu. “A hidden projection property of Plackett-Burman and related designs”. In: *Statistica Sinica* (1995), pp. 235–250.

7 Appendix A

Codes below are to find the combination of 4 active factors among the 11 factors which has the smallest MSE.

```
design.matrix <- function(){
  A=c(1,1,-1,1,1,1,-1,-1,-1,1,-1,-1)
  B=c(-1,1,1,-1,1,1,1,-1,-1,-1,1,-1)
  C=c(1,-1,1,1,-1,1,1,1,-1,-1,-1,-1)
  D=c(-1,1,-1,1,1,-1,1,1,1,-1,-1,-1)
  E=c(-1,-1,1,-1,1,1,-1,1,1,1,-1,-1)
  F=c(-1,-1,-1,1,-1,1,1,-1,1,1,1,-1)
  G=c(1,-1,-1,-1,1,-1,1,1,-1,1,1,-1)
  H=c(1,1,-1,-1,-1,1,-1,1,1,-1,1,-1)
  I=c(1,1,1,-1,-1,-1,1,-1,1,1,-1,-1)
  J=c(-1,1,1,1,-1,-1,-1,1,-1,1,1,-1)
  K=c(1,-1,1,1,1,-1,-1,-1,1,-1,1,-1)
  x= cbind(A,B,C,D,E,F,G,H,I,J,K)
  return(x)
} #12 run

x
h=colnames(x)
h1=as.matrix(h)

#####
a=c("A","B","C","D","E","F","G","H","I","J","K")
m=combn(a,4,FUN=NULL,simplify=TRUE)
V=as.matrix(m) #matrix(Row 4 Column 330)

#####

n = dim(x)[1] # 12 Run
```

```

T=0
numb.of.col =dim(V)[2]*dim(V)[1]# number of columns equals 330*4
header = rep(0,numb.of.col)
mat = matrix(0,n,numb.of.col) # design matrix 12*(330*4)

for (i in 1:330){
  for (j in 1:4){
    for(k in 1:11){
      if (V[j,i]==h1[k,]){
        T = T + 1
        mat[,T]=x[,h1[k,]]
        header[T]=h1[k,]
      }}}
colnames(mat)=header
mat #design matrix(Row 12 Column 330*4) for 330 candidate sets
#each with 4 terms(4 main factors).
#####
n1 = dim(x)[1] # 12 Run
k=4 #the number of factors equals to 4
T1= 0
numb.of.col1 =330*( k+k*(k-1)/2)# number of columns equals 330*10
header1 = rep(0,numb.of.col1) # headers
ex.mat = matrix(0,n1,numb.of.col1) # extend design matrix 12*(330*10)
for (l in 0:329){
  for (i in 1:k){
    ex.mat[, (10*l+1):(10*l+k)] = mat[, (4*l+1):(4*l+k)] # Design matrix
    header1[(10*l+1):(10*l+k)]= header[(4*l+1):(4*l+k)] # Make header
  }}
colnames(ex.mat) = header1
ex.mat

```

```
#####
for (l in 0:329){
  for (i in 1:k){
    for (j in (i+1):k){
      if (i != j & i != k){
        T1 = T1 + 1
        ex.mat[,4*(l+1)+T1] = mat[,4*l+i]*mat[,4*l+j]# Add 2-factor inter. columns
        header1[4*(l+1)+T1] = paste(header[4*l+i],header[4*l+j],sep="")
      }
    }
  }

colnames(ex.mat) = header1
ex.mat #design matrix(Row 12 Column 330*10) for 330 candidate sets
#each with 10 terms(4 main factors and 6 two-factor interactions).

#####
#general case for choosing the model
sim_10000_matrix=matrix(0,4,1000*10)

for (b in 1:1000){
  #y<-A+C+B*C+C*D+rnorm(12,mean=0,sd=0.7)
  #y<-G+I+H*I+I*K+rnorm(12,mean=0,sd=0)
  #y<-2*A+3*B+2*C+D+3*C*D+rnorm(12,mean=0,sd=0.7)
  #y<-4*A+B+C+D+2*A*D+rnorm(12,mean=0,sd=0)
  y<-2*A+4*C+2*B*C+2*C*D+rnorm(12,mean=0,sd=0)
  #y<-A+C+B*C+C*D+rnorm(12,mean=0,sd=0) #the response is fixed
  #y<-c(0.54,-1.73,0.48,1.92,-2.02,2.06,1.44,0.99,-1.99,2.48,
    #-2.59,-0.75)#the variance of the noise equals to 0.7

P= matrix(0,10,330)
```

```

Q=matrix(0,6,330)

Effects=matrix(0,6,330)

for (l in 0:329) {
  M=data.frame(ex.mat[, (10*l+1):(10*l+10)])
  P[,l+1]=cbind((lm(y~1+M[,1]+M[,2]+M[,3]+M[,4]+M[,5]+M[,6]+
    #M[,7]+M[,8]+M[,9]+M[,10])$coefficients)[-1])

  Q[,l+1]=sort(abs(P[,l+1]),decreasing=TRUE)[1:6]
  Effects[,l+1]=cbind(names(sort(abs((lm(y~1+M[,1]+M[,2]+M[,3]+M[,4]+
    #M[,5]+M[,6]+M[,7]+M[,8]+M[,9]+M[,10])
    #$coefficients)[-1]),decreasing=TRUE)[1:6])))}

Effects

for (l in 0:329){
  for (j in 1:6){
    for (i in 1:10){
      if (abs(P)[i,l+1]==Q[j,l+1]){
        Q[j,l+1]=P[i,l+1]
      }}}
  Q #the coefficients for 6 effects which have the
    #highest absolute value for all 330 candidate sets

#####
for (k in 1:dim(Effects)[2]){
  for (j in 1:10){
    for (i in 1:6){
      if (Effects[i,k]==paste("M[, ",toString(j),"]", sep="")){

```

```

      Effects[i,k]=j
    }}}}

```

Effects #6 effects with the highest absolute value for all 330 candidate sets

```
#####
```

```

New_design.mat1=matrix(0,12,6*dim(Effects)[2])
for (l in 0:329){
  M=data.frame(ex.mat[, (10*l+1):(10*l+10)])
  New_design.mat1[, (6*l+1):(6*l+6)]=cbind(M[,as.numeric(Effects[1,l+1])],
    #M[,as.numeric(Effects[2,l+1])],M[,as.numeric(Effects[3,l+1])],
    #M[,as.numeric(Effects[4,l+1])],M[,as.numeric(Effects[5,l+1])],
    #M[,as.numeric(Effects[6,l+1])])
}

```

New_design.mat1 #the design matrix with 6 terms for all 330 candidate sets

```
#####
```

```

yHat=matrix(0,12,330)
MSE=matrix(0,330,1)
intercept=(lm(y~1+M[,1]+M[,2]+M[,3]+M[,4]+M[,5]+
  #M[,6]+M[,7]+M[,8]+M[,9]+M[,10])$coefficients)[1]

for (l in 0:329){
  yHat[,l+1]=New_design.mat1[, (6*l+1):(6*l+6)]%*%Q[,l+1]+intercept
  MSE[l+1,]=sum((y-yHat[,l+1])^2)/5 #the formula for MSE
}

```

MSE #the mean square error

```
#####
```



```

MSE=data.frame(MSE)
MSE_smallest=order(MSE)[1:10] #the column number of the 330 combinations

Eff=matrix(0,4,10)
for (i in 1:10){
  Eff[,i]=V[,MSE_smallest[i]]
}
Eff #10 smallest MSE candidate sets when running the procedure once
sim_10000_matrix[, (10*b-9):(10*b)]=Eff
}

#####
num<-matrix(0,1,330)

for (j in 1:330){
  for (i in 1:10000){
    if (paste(toString (sim_10000_matrix[, i]),sep="")==
        #paste(toString (V[,j]),sep="")){
      num[j]=num[j]+1
    }
  }
}

total <- t(rbind(V, num))
total<-data.frame(total)
colnames(total)[5]="num"
high_prob_rownames=order(num)[330:321]

sim_highprob_matrix=rbind(V[,high_prob_rownames],num[,high_prob_rownames])
sim_highprob_matrix #10 smallest MSE candidate sets with capture frequency
#The End

```

8 Appendix B

Table 61: 10 candidate sets and their capture frequency for the Model 5 with σ^2 equals to 0.5,0.6,...1.0

Combinations	1	2	3	4	5	6	7	8	9	10
	G	A	D	C	B	C	F	F	A	C
	H	B	G	G	G	G	G	G	C	F
	I	C	H	H	I	I	H	I	G	H
	K	D	I	I	K	K	I	K	I	K
capture frequency	721	674	575	574	574	562	375	341	329	308
Combinations	1	2	3	4	5	6	7	8	9	10
	G	A	C	C	D	B	A	F	C	F
	H	B	G	G	G	G	C	G	F	G
	I	C	H	I	H	I	G	H	H	I
	K	D	I	K	I	K	I	I	K	K
capture frequency	549	530	451	419	402	399	259	233	229	228
Combinations	1	2	3	4	5	6	7	8	9	10
	G	A	C	C	D	B	A	A	C	F
	H	B	G	G	G	G	F	C	F	G
	I	C	H	I	H	I	G	G	G	H
	K	D	I	K	I	K	I	I	I	I
capture frequency	424	354	336	327	319	311	191	191	169	164
Combinations	1	2	3	4	5	6	7	8	9	10
	G	A	D	C	C	B	A	F	A	A
	H	B	G	G	G	G	C	G	F	F
	I	C	H	I	H	I	G	I	H	G
	K	D	I	K	I	K	I	K	K	I
capture frequency	326	273	245	243	243	228	167	152	144	143
Combinations	1	2	3	4	5	6	7	8	9	10
	G	C	C	B	A	D	A	A	F	A
	H	G	G	G	B	G	F	F	G	C
	I	H	I	I	C	H	G	H	H	G
	K	I	K	K	D	I	I	K	I	I
capture frequency	269	215	209	194	188	181	148	135	127	121
Combinations	1	2	3	4	5	6	7	8	9	10
	G	C	C	A	D	B	A	F	C	C
	H	G	G	B	G	G	C	G	F	F
	I	H	I	C	H	I	G	H	G	H
	K	I	K	D	I	K	I	I	I	K
capture frequency	214	165	162	138	136	132	131	112	111	107

Table 62: 10 candidate sets and their capture frequency for the Model 6 with σ^2 equals to 0.9 and 1.0

Combinations	1	2	3	4	5	6	7	8	9	10
	B	B	A	C	A	C	A	B	C	C
	C	C	B	E	B	D	B	E	I	G
	I	E	C	I	E	I	D	G	J	I
	K	J	D	K	J	K	G	J	K	K
capture frequency	827	799	789	432	392	377	361	332	314	306
Combinations	1	2	3	4	5	6	7	8	9	10
	B	B	A	C	A	C	A	B	C	C
	C	C	B	E	B	D	B	E	G	I
	I	E	C	I	E	I	D	G	I	J
	K	J	D	K	J	K	G	J	K	K
capture frequency	716	694	677	405	377	352	340	321	310	296

Table 63: 10 candidate sets and their capture frequency for the Model 7 with σ^2 equals to 0.8,0.9 and 1.0

Combinations	1	2	3	4	5	6	7	8	9	10
	A	A	A	A	A	A	A	A	A	D
	B	D	B	H	D	C	B	D	D	E
	C	I	I	I	E	D	D	E	G	I
	D	K	K	K	J	K	I	I	J	J
capture frequency	834	616	583	489	453	315	312	310	298	293
Combinations	1	2	3	4	5	6	7	8	9	10
	A	A	A	A	A	A	A	A	A	A
	B	B	D	H	D	D	D	B	C	C
	C	I	I	I	E	E	G	D	D	D
	D	K	K	K	J	I	J	I	K	J
capture frequency	716	519	512	430	377	309	298	295	292	262
Combinations	1	2	3	4	5	6	7	8	9	10
	A	A	A	A	A	A	A	A	A	A
	B	B	D	H	D	D	D	B	C	C
	C	I	I	I	E	E	G	D	D	D
	D	K	K	K	J	I	J	I	K	J
capture frequency	627	464	440	385	330	301	297	280	274	250

Table 64: 10 candidate sets and their capture frequency for the Model 8 with $\sigma^2=1.0$

Combinations	1	2	3	4	5	6	7	8	9	10
	A	C	C	A	C	C	C	C	C	B
	B	G	G	C	F	F	D	D	D	C
	C	I	H	G	H	G	F	E	E	H
	D	K	I	I	K	I	K	K	H	J
capture frequency	850	776	763	586	582	578	245	242	238	227

Table 65: The breakdown values for each tested model

	Model 5	Model 6	Model 7	Model 8
Breakdown value	0.2	0.8	0.7	0.9

Table 66: Ordinary regression analysis on the Model CE 1-5 in the Compound extraction experiment

	R^2	$\hat{\sigma}$	Significant effects at a 1% level
Model CE 1	0.995	0.234	C,D,AD,AF
Model CE 2	0.993	0.295	C,D,AD
Model CE 3	0.992	0.321	C,D,AD
Model CE 4	0.970	0.627	C,D,AD
Model CE 5	0.975	0.575	C,D,AD