



**NTNU – Trondheim**  
Norwegian University of  
Science and Technology

# Methods for Analyzing the 12 Run Plackett-Burman Design

**Helene Eide Wiik**

Master of Science in Physics and Mathematics

Submission date: June 2014

Supervisor: John Sølve Tyssedal, MATH

Norwegian University of Science and Technology  
Department of Mathematical Sciences



## Preface

The objective of this master thesis is to discuss methods for identifying active factors in experiments using a 12 run Plackett-Burman design.

This thesis is the result of the course TMA4905 - Statistics, where the final result counts for 30.0 credits. The work was carried out during the spring semester of 2014 under the supervision of John Tyssedal. The completion of this master thesis concludes my Master of Science in Physics and Mathematics at the Norwegian University of Science and Technology (NTNU).

My sincere appreciation goes to my supervisor John Tyssedal for his contributions and guidance during the work of this master thesis. He has provided me with valuable feedback and constructive discussions which has improved this thesis as well as my understanding. His interest in the topic of study has been a great source of motivation, which has ensured a steady progress during the semester.





## Abstract

In the early stages of an experimental analysis it is often of interest to determine the most important factors. The main goal of this master thesis is to study and compare different methods for determining the active factors in a 12 run Plackett-Burman experiment. The methods discussed in this thesis are the Dantzig selector, a graphical Dantzig selector, a projection based Dantzig selector, the Lasso, a projection based method, the partial F-test, a graphical method using orthogonalization and the residual variance of fitted models. The methods were implemented in R and applied to data generated from a total of 15 different models. The form of these models vary in terms of the number of active factors and the presence of interaction effects. This diversity of models facilitate exposure of strengths, weaknesses and limitations of the methods. From the study conducted in this thesis it was confirmed that methods using the Dantzig selector perform significantly better for designs following the uniform uncertainty principle. The results produced by the Lasso were found to be relatively close to what was obtained by the Dantzig selector. Further, it was found that a combination of the projection based method, residual variance of fitted models and the graphical method using orthogonalization together form a highly valuable procedure for determining active factors.



## Sammendrag

I en tidlig fase av en eksperimentell analyse er det ofte av interesse å identifisere faktorene som har størst betydning. Hovedmålet med denne masteroppgaven er å undersøke og sammenligne ulike metoder for å finne de aktive faktorene i et 12 run Plackett-Burman forsøk. Metodene som er diskutert i denne oppgaven er Dantzig selectoren, en grafisk Dantzig selector, en projeksjonsbasert Dantzig selector, Lasso, en projeksjonsbasert metode, den partielle F-testen, en grafisk metode som benytter seg av ortogonalisering og residual variansen av tilpassede modeller. Metodene ble implementert i R og koden ble testet på data som ble generert fra 15 forskjellige modeller. Formen på modellene varierer i forhold til antall aktive faktorer og om det er interaksjonsledd tilstede i modellene. Dette gjør at man kan avdekke mulige styrker, svakheter og begrensninger ved metodene. Fra undersøkelsene som er gjennomført i denne oppgaven ble det konkludert med at metodene som inkluderer Dantzig selectoren gir signifikant bedre resultater i forsøk som følger det uniforme usikkerhetsprinsippet. Lasso estimatoren ga resultater som var relativt like til de som ble funnet ved å bruke Dantzig selectoren. Videre ble det funnet at kombinasjonen av den projeksjonsbaserte metoden, residual variansen av tilpassede modeller og den grafiske metoden som bruker ortogonalisering tilsammen utgjør en svært god metode for å identifisere aktive faktorer.



# Contents

<b>Preface</b>	<b>i</b>
<b>Abstract</b>	<b>iii</b>
<b>Sammendrag</b>	<b>v</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Theory</b>	<b>3</b>
2.1 Experimental designs . . . . .	3
2.1.1 Factorial designs at two levels . . . . .	3
2.1.2 Regular and non-regular designs . . . . .	5
2.1.3 Projectivity . . . . .	6
2.1.4 Plackett-Burman designs . . . . .	8
2.2 Variable selection . . . . .	9
2.2.1 The Dantzig selector . . . . .	11
2.2.2 The Lasso . . . . .	13
2.2.3 Orthogonalization of augmented regression models . . . . .	14
<b>3 Methods</b>	<b>16</b>
3.1 The Dantzig selector . . . . .	16
3.2 A graphical Dantzig selector . . . . .	17
3.3 A projection based Dantzig selector . . . . .	19
3.3.1 An expansion of the method . . . . .	20
3.4 The Lasso . . . . .	21
3.5 A projection based method . . . . .	22
3.6 The partial F-test . . . . .	23
3.7 A graphical method using orthogonalization . . . . .	23
3.8 R software . . . . .	26
3.8.1 Background . . . . .	26
3.8.2 Implementation of code . . . . .	26
<b>4 Examples</b>	<b>28</b>
4.1 Example 1 . . . . .	30
4.1.1 Effect identification using the Dantzig selector . . . . .	30
4.1.2 Variable selection using a graphical Dantzig selector . . . . .	32

4.1.3	Selection of effects using a projection based Dantzig selector . . . . .	34
4.1.4	Identifying active effects using the Lasso . . . . .	36
4.1.5	Factor screening using a projection based method . . . . .	41
4.2	Example 2 . . . . .	44
4.2.1	Effect identification using the Dantzig selector . . . . .	44
4.2.2	Variable selection using a graphical Dantzig selector . . . . .	46
4.2.3	Selection of effects using a projection based Dantzig selector . . . . .	48
4.2.4	Identifying active effects using the Lasso . . . . .	50
4.2.5	Factor screening using a projection based method . . . . .	55
4.2.6	Focus on using methods that can confirm each other . . . . .	57
<b>5</b>	<b>Discussion</b>	<b>70</b>
<b>6</b>	<b>Conclusion</b>	<b>73</b>
	<b>Bibliography</b>	<b>76</b>
	<b>Appendix A</b>	<b>77</b>
	<b>Appendix B</b>	<b>79</b>
	<b>Appendix C</b>	<b>101</b>

# Chapter 1

## Introduction

Statistical experimental designs have been a key reason for progress in several scientific fields, and is still highly valuable. For leading companies in today's society it is crucial to focus on developing new products and optimizing existing solutions. Experiments are often conducted as a basis for acquiring the knowledge needed for this development. In the industry heavy machinery is trusted to perform critical and difficult tasks, creating a demand for highly reliable systems. In that regard, experiments are performed to ensure safety for personnel, equipment and environment. When time and resources are limited it is important to identify the factors that have the greatest effect on the production rate and the quality of the product. When an experiment is conducted the factors with a significant amount of impact on the response are called active. Factors that have negligible impact on the response, are called inert.

As science and technology advance at an increasing rate it is necessary to be able to investigate complex processes. It is often the case that these processes include many factors that might have a significant impact on the response. As each run in such experiments may be expensive and time consuming it is often desirable to choose an experimental design that is capable of investigating a large amount of factors in a small number of runs. A group of saturated experimental designs that are known to hold this property is the two-level Plackett-Burman designs, hereinafter referred to as the two-level PB designs. A design with  $s$  levels and  $k$  factors are called saturated if the number of runs are equal to  $k(s - 1) + 1$ . For a two-level saturated design, this means that  $k$  factors can be investigated in  $k + 1$  runs. Supersaturated designs are designs with the number factors greater or equal to the number of experimental runs. Such designs are also often used for screening, i.e. to identify a set of active factors from a larger set of factors.

The objective of this master thesis is to discuss methods for determining active factors in a 12 run PB experiment. Three of the methods studied in this thesis involve the use of the Dantzig selector, an estimator that has received a lot of attention lately. This estimator was initially indented for signal processing in supersaturated designs, but it may also be used on designs with less factors than experimental runs. The main contribution of this master thesis is to study the performance of

the Dantzig selector when applied to data generated from generalized projection models. A generalized projection model is a model with the ability to estimate all main effects and interaction effects up to a given order. Another contribution of this thesis is to show how a combination of methods can be used to obtain a more reliable analysis. The results obtained by the methods using the Dantzig selector are compared to results obtained by the Lasso estimator. Thereafter, the projection based method used in Wiik (2013), see also Tyssedal and Samset (1997), is tested in combination with the partial F-test, a graphical method using orthogonalization and the residual variance of fitted models. All of the methods studied in this thesis are based on the principle of factor sparsity. The principle of factor sparsity states that the dimension of the space containing the active factors is likely to be low. This means that the variation in the data is likely to be explained by only a few factors. The methods are tested on models with up to four factors, though it might be questioned whether four active factors are sparse enough in a 12 run PB experiment.

In the following chapter an introduction to theory considered relevant for the methods studied in this thesis is given. In the first section of Chapter 2 the main focus is directed towards experimental designs. A general introduction to factorial designs is given, followed by a discussion of regular and non-regular designs. An introduction to projectivity and PB designs are found in this section as well. In Section 2.2 the problem of variable selection is approached by a general introduction to the subject, followed by a discussion of the Dantzig selector, the Lasso and the method of orthogonalization. The methods studied in this thesis are discussed in details in Chapter 3, and a short introduction to the R code implemented for these methods is given at the end. In Chapter 4 the methods are tested on data generated from a total of 15 different models. The models are divided into two examples, with the models that are considered to be relatively easy to analysis in the first example and the more challenging ones in the second. The data generated are given in Appendix A, and the entire R code implemented for the methods is found in Appendix B. Some of the plots produced by the methods are given in the text and some are displayed in Appendix C. Finally, a formal discussion of the findings of this thesis is given, followed by some concluding remarks in the last chapter.



# Chapter 2

## Theory

### 2.1 Experimental designs

When conducting an experiment, the experimenter chooses a set of factors that are assumed to influence the response. For each factor a predefined number of levels is defined. In the early stages of an experimentation, the experiment is often simplified by only including two levels, where one level is defined as high and the other as low. The different combinations of high and low level for the factors are presented in a design matrix. The experiments are conducted for different combinations of levels for the given factors, and the response is measured after each run. The analysis can then be performed based on the measured variation in the response.

Proper planning should be conducted prior to performing the experiment. The planning should ensure a time efficient and economic analysis, without compromising the information being searched for. As an initial step, the goal of the analysis should be clearly defined. The goal may for example be to optimize the system in terms of quality, capacity or efficiency. With the given goal in mind, the experimental design should be chosen carefully. The design should provide a proper balance between the need for accurate information and limitations in terms of time, cost and complexity in the analysis.

#### 2.1.1 Factorial designs at two levels

There are several types of designs to choose from when conducting an experiment. One widely known class of designs is the full two-level factorial designs. In these designs all possible combinations of levels for all the factors are included. Hence, for a full factorial experiment comprising  $k$  factors with two levels, the number of runs is equal to  $2^k$ . The advantage of this design is that all the factor effects can be estimated independently. The disadvantage of the full factorial designs is the large number of runs to be performed for experiments with many factors.

If it is reasonable to assume that some of the factors are inert it may be sufficient to perform only a fraction of the experimental runs. Such designs are called

fractional factorial design. The number of runs for these designs is equal to  $2^{k-p}$ , when each of the  $k$  factors have two levels, and  $p$  is the fraction of the design. When  $p$  is equal to one, the design is referred to as a half fraction of the full factorial design. For fractional factorial designs the effects are either fully aliased or it is possible to estimate them independent of the other effects. The  $2^{k-p}$  designs are often preferred over the full factorial designs because they require less runs. For comparison, let us consider a two-level design with four factors. The full factorial design would demand  $2^4 = 16$  runs, while the half fraction,  $2^{4-1}$ , would require 8 runs. A full factorial  $2^4$  design is shown in Table 2.1 and a  $2^{4-1}$  fractional factorial design in Table 2.2.

Run	A	B	C	D
1	-	-	-	-
2	+	-	-	-
3	-	+	-	-
4	+	+	-	-
5	-	-	+	-
6	+	-	+	-
7	-	+	+	-
8	+	+	+	-
9	-	-	-	+
10	+	-	-	+
11	-	+	-	+
12	+	+	-	+
13	-	-	+	+
14	+	-	+	+
15	-	+	+	+
16	+	+	+	+

Table 2.1: A full factorial  $2^4$  design.

In order to construct a  $2^{k-p}$  fractional factorial design,  $p$  generators are used. A generator is the relation where the aliased effects are set equal to each other. Aliased effects are effects where the signs in the design columns are equal. For example, if the signs in the column for A are equal to the signs for BC, then A and BC are aliased, and  $A=BC$  is the generator of the design. By multiplying both sides of this relation by A the defining relation,  $I=ABC$ , is obtained. The resolution is another property of the factorial designs. The resolution is equal to  $R$  if all  $p$ -factor effects are aliased with effects comprising  $R - p$  factors or more. In other words, the resolution is equal to the number of letters in the shortest word in the defining relation.

For the design given in Table 2.2 the generator is the relation  $D=ABC$ . This can be seen by comparing the signs for ABC and the signs for D. By multiplying both sides of the generator by D, the defining relation is found to be  $I=ABCD$ . As the defining relation contains four letters, the resolution is equal to four.

Run	A	B	C	D
1	-	-	-	-
2	+	-	-	+
3	-	+	-	+
4	+	+	-	-
5	-	-	+	+
6	+	-	+	-
7	-	+	+	-
8	+	+	+	+

Table 2.2: A  $2^{4-1}$  fractional factorial design.

When performing an analysis with a two-level factorial design, the estimates of main effects and interaction effects are often of interest. For main effects the definition given by Tyssedal (2011) will be used, which states that "For two-level designs we define the main effect of a factor as: Expected average response when the factor is on the high level - the expected average response when the factor is at the low level". The main effect of a factor,  $A$ , is estimated by

$$\hat{A} = \bar{y}_H - \bar{y}_L \quad (2.1)$$

where  $\bar{y}_H$  is the average response when the factor is on its high level, and  $\bar{y}_L$  the average response when the factor is on its low level. Interactions are present when the effect of one factor is dependent on the level of one or more of the other effects. Interaction effects are estimated by similar calculations as those for main effects. Thus, an interaction effect is estimated by the average response when the interaction is on its high level, minus the average response when the interaction is on its low level.

## 2.1.2 Regular and non-regular designs

When performing an experiment with a factorial design the experimenter has the option to choose between a regular design and a non-regular design. The type of design preferred is dependent on the conditions of the experiment at hand and on the goal of the analysis.

### Regular designs

The class of  $2^k$  designs and  $2^{k-p}$  designs are called regular designs. Analysis with regular designs are relatively easy to perform, and for this reason these designs are often used in the industry. The drawback of regular designs is that they only exist for the number of runs equal to a power of two. As a consequence, these designs are less economic and time efficient compared to some of the members of the non-regular designs.

To identify active effects in regular designs without replicated rows, a normal plot and Lenth's method are commonly used. The basis for the normal plot is that inert effects should be normally distributed with zero means and equal variances. As a consequence, their estimates should lie on a straight line in a normal plot. All deviating points are indications of significant effects. A description of Lenth's method is given in Hamada and Wu (2011).

### Non-regular designs

All two-level, orthogonal designs that are not a  $2^{k-p}$  design, are said to be non-regular. The non-regular designs provide flexibility in terms of run sizes compared to the regular designs. This advantage is one of the main reasons why non-regular designs are commonly used for screening experiments. Another advantage is the properties the non-regular designs holds when a subset of the factors is considered. These properties are referred to as projection properties, and these will be discussed further in Subsection 2.1.3.

One difference between the regular and the non-regular designs, is that regular designs have a defining relation, while most non-regular designs do not. In contrast to the  $2^{k-p}$  regular designs, the effects for most of the non-regular designs are not fully aliased, but only partially aliased. Their alias strings may be long and they have been criticized for having a rather complex alias pattern. On the other hand, the non-regular designs have a higher ability than designs with fully confounded effects, to select one single model instead of a set of possible models.

### 2.1.3 Projectivity

If only a small subset of the factors are active it is of importance to know how well a design projects onto such a small subset. The projectivity  $P$  for a two-level design with  $n$  runs and  $k$  factors was defined by Box and Tyssedal (1996) as "(...) the design contains a complete  $2^P$  factorial in every possible subset of  $P$  out of the  $k$  factors, possibly with some points replicated." An important property of designs with projectivity  $P$  is that unbiased estimates are available for main effects and interactions up to  $P$  factors, given that all other factors are inert. Also, if a subset of factors contains the only active factors, the difference between expected values in replicated runs is equal to zero. Hence, estimates of the variance within each group of replicated runs should be low. This property is of great value for the projection based method that will be discussed in Section 3.5.

There exist some general rules for projections in regular designs. Any subset of factors in a regular design will consist of a full factorial design or a fractional factorial design. Also, if the resolution of a two-level regular design is equal to  $R$ , then the projectivity is equal to  $P = R - 1$ . Another useful rule is that the projectivity of regular, saturated fractional factorial designs is always equal to 2.

Lets consider the  $2^{4-1}$  fractional factorial design given in Table 2.2. If one of the four factors is found to be inert, this column in the design matrix can be removed. The remaining design consists of three factors with eight runs, and hence, the reduced design is a  $2^3$  factorial design. By using the definition of projectivity given by Box and Tyssedal (1996), it is clear that the  $2^{4-1}$  design has projectivity 3. In the case where only two factors are found to be active, the remaining design consists of two  $2^2$  factorial designs. Knowledge about these projection properties can be used as a tool when analysis is performed.

There has been an increasing interest for knowledge regarding projections in non-regular designs. One of these is the hidden projection property. Hidden projection property is described by Wang and Wu (1995) as "A design is said to possess a hidden projection property if it allows some (or all) interactions to be estimated even when the projected design does not have the right resolution or other combinatorial design property for the same interactions to be estimated." Designs with hidden projection properties are useful for screening, especially when the number of runs in the design is low. For such designs, the run size does not prevent estimation of interactions.

It is of special importance to consider the projection properties of screening designs, as they are constructed to identify a set of active factors from a bigger set of factors. The PB designs are known to have outstanding projection properties that can be taken advantage of. The projection properties of the PB designs will be discussed in Subsection 2.1.4.

The method used to construct an orthogonal two-level design can in some cases be used to determine the projectivity of the design. For the method called cyclic generation, two-level designs are constructed from a single sequence of plus and minuses. This initial sequence is used as the top row of the design matrix. For each new row the signs are shifting one place to the right, and the last sign is placed in the front. A second method used to construct orthogonal two-level designs is called doubling. This method uses a specific  $n \times n$  matrix called a Hadamard matrix, denoted  $H_n$ . A Hadamard matrix is a matrix consisting of the values -1 and +1, where  $H_n^T H_n = H_n H_n^T = nI$  and where  $I$  is the identity matrix. This matrix can be obtained by adding a column of plus signs in front of the cyclic generated design. The technique of doubling can be explained as follows. Let  $H_n$  be a  $n \times n$  Hadamard matrix with all entries in the first column equal to +1. The matrix

$$\begin{bmatrix} H_n & H_n \\ H_n & -H_n \end{bmatrix}$$

is then a  $2n \times 2n$  Hadamard matrix whose last  $2n - 1$  columns constitute the orthogonal two-level array.

Box and Tyssedal (1996) proved three results where projectivity is determined based on how the design matrix was obtained. These results were given as follows.

- I *A saturated design obtained from a doubled  $n \times n$  orthogonal array is always of projectivity  $P=2$  and only 2.*
- II *A saturated design obtained from a cyclic orthogonal array is either a factorial orthogonal array with  $P=2$  and only 2, or else has projectivity at least  $P=3$ .*
- III *Any saturated two-level design obtained from an orthogonal array containing  $n=4m$  runs, with  $m$  odd, is of projectivity at least  $P=3$ .*

These results are particularly useful for determining the projectivity of PB designs.

### 2.1.4 Plackett-Burman designs

The most well known non-regular designs are the PB designs introduced by Plackett and Burman (1946). They have become known for their ability to investigate a large number of factors in a relatively low number of experimental runs. Due to this property the PB designs are often used for screening. The number of runs  $n$  in a PB design is equal to a multiple of four. Plackett and Burman only included the designs with  $n \leq 100$ , and they also omitted the design where  $n = 92$ . For PB designs where the number of runs is equal to a power of two the designs coincide with the regular ones, and the rest of the PB designs are non-regular. The 12 run PB design matrix is shown in Table 2.3.

Run	A	B	C	D	E	F	G	H	I	J	K
1	+	+	-	+	+	+	-	-	-	+	-
2	-	+	+	-	+	+	+	-	-	-	+
3	+	-	+	+	-	+	+	+	-	-	-
4	-	+	-	+	+	-	+	+	+	-	-
5	-	-	+	-	+	+	-	+	+	+	-
6	-	-	-	+	-	+	+	-	+	+	+
7	+	-	-	-	+	-	+	+	-	+	+
8	+	+	-	-	-	+	-	+	+	-	+
9	+	+	+	-	-	-	+	-	+	+	-
10	-	+	+	+	-	-	-	+	-	+	+
11	+	-	+	+	+	-	-	-	+	-	+
12	-	-	-	-	-	-	-	-	-	-	-

Table 2.3: A 12 run PB design matrix.

Earlier, the PB designs were often rejected in favor of other designs due to their complex alias structure. For PB designs, each of the  $k$  factors are partially confounded with all two-factor interactions not involved with the given factor. The analysis of these designs was considered to be complicated, and was expected to give unclear

results. Thus, these designs were considered to only be capable of identifying main effects. An analysis strategy proposed by Hamada and Wu (1991) showed that when only a few factors are active, their partial aliasing can be considered as an advantage, rather than a barrier for analysis. As the knowledge and understanding of projection properties were developed, an increasing interest was directed towards the PB designs. A thorough discussion of these properties is given by Box and Tyssedal (1996).

The PB designs with the number of runs  $n = 40, 56, 88$  and  $96$  are constructed by doubling, and hence, they are of projectivity  $P = 2$ . Box and Tyssedal (1996) were able to show that all the non-regular PB designs, except for the four with  $P = 2$ , are of projectivity at least  $P = 3$ . Computer search has been a key tool for investigating designs of projectivity  $P = 4$  and designs with hidden projection properties. By this method it was found that the PB designs with  $n = 68, 72, 80$  and  $84$  are of projectivity  $P = 4$ , as seen in Tyssedal og Samset (1999).

In this thesis a 12 run PB design with 11 factors is considered. For this design the projectivity is equal to three. The derivation of this result is found in Lin and Draper (1992) and it also follows from III in the list given in Subsection 2.1.3. When considering a subset of any combination of three factors, the 12 run PB design is reduced to a design corresponding to a full factorial  $2^3$  design plus the  $2^{3-1}$  half fraction of the  $2^3$  design.

Although the 12 run PB design has too few runs to be of projectivity four, some useful properties have been found when projecting onto a subset of four factors. Lin and Draper (1992) found that if it is appropriate to assume that all three- and four-factor interactions are negligible, then in a projection onto any four factors it is possible to estimate all main effects and all two-factor interactions. By comparing these findings to the definition given by Wang and Wu (1995), one will recognize that the 12 run PB design has hidden projection properties.

## 2.2 Variable selection

Consider the regression model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad (2.2)$$

where  $\mathbf{y}$  contains  $n$  observed values,  $\boldsymbol{\beta}$  is a vector of  $k$  parameter values,  $\mathbf{X}$  is a  $n \times k$  design matrix, and  $\boldsymbol{\epsilon}$  is the noise. In this thesis the noise is assumed to be independent and identically normally distributed with zero mean and variance equal to  $\sigma^2$ . Let the notation used for Equation 2.2 be valid throughout this thesis.

The goal of variable selection is to identify the values of  $\boldsymbol{\beta}$  that are significantly different from zero. Under the assumption of factor sparsity, variable selection can contribute to an improved accuracy of estimation and reduce the cost of computation for further analysis. The variable selection methods should facilitate estimation of

parsimonious models with great explanatory power. Parsimonious models are usually desired as they have more degrees of freedom, are easier to interpret and do not add additional noise to predictions.

The procedures used for variable selection may either be factor based or effect based. The methods with a factor based approach searches through a set of factors that are considered to be of interest, with the goal of identifying the active factors. Except for factor sparsity, few assumptions are usually needed for these procedures. Factor based procedures are appropriate for designs with good projection properties. For the effect based procedures the focus is directed towards selection of active effects. For these methods effect sparsity is assumed and the principle of effect heredity is often considered in the search for active effects. The principle of effect heredity states that either one or both of the factors involved with an interaction should also be included in the model as main effects.

Several of the methods studied in this thesis result in a list, where the effects are ordered according to how likely they are considered to be active. Hence, an additional challenge that the analyst encounters when these methods are performed, is to determine the number of effects to include in the model. Procedures such as forward selection, backward elimination and stepwise procedures can be used for this purpose. By the method referred to as forward selection, a model with a single effect is fitted to the data, and then more and more effects are added as long as all of them are considered to be significant. The effects are added in the order suggested by the given method. Another possibility is to perform backward elimination. Then, the first step is to fit a model with a preselected number of effects. In this thesis, it was chosen to fit the six first effects to a linear model. Afterwards, effects are removed based on their p-value as long as all of the remaining effects are considered to be significant. The stepwise procedure is a combination of these two methods, where effects are added and removed depending on their p-value or according to a criterion. It may be discussed which p-value that is optimal, but in this thesis it was chosen to use 0.10 for forward selection and 0.05 for backward elimination.

There are some challenges related to variable selection, especially when considering higher dimensional cases. Non orthogonal effect columns may cause problems as they can lead to selection of incorrect models and result in over-fitted models. In general, one variable selection procedure does not provide any guarantee that it will find the true active factors. Hence, it is wise to perform more than one variable selection method, and interpret the results in relation with the opinion of professionals who have experience within the given field. Another challenge with variable selection is that the methods usually are restricted by assumptions. In practice, it is often the case that these conditions are not met, or that the assumptions are hard to verify. These challenges are important to keep in mind when performing variable selection.



## 2.2.1 The Dantzig selector

The Dantzig selector is an estimator proposed by Candes and Tao (2007) designed to identify active factors and estimate their corresponding parameter values in linear regression models that might have more parameters than observations. The estimate of parameters using the Dantzig selector is the solution to the convex optimization problem

$$\min_{\hat{\boldsymbol{\beta}} \in \mathbb{R}^k} \|\hat{\boldsymbol{\beta}}\|_{l_1} \text{ subject to } \|\mathbf{X}^T \mathbf{r}\|_{l_\infty} \leq \delta \quad (2.3)$$

where  $\delta$  is a tuning parameter and  $\mathbf{r} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$  is a vector of residuals. The norms in this problem are given by  $\|\hat{\boldsymbol{\beta}}\|_{l_1} = \sum_{i=1}^k |\hat{\beta}_i|$  and  $\|\mathbf{X}^T \mathbf{r}\|_{l_\infty} = \max_{1 \leq i \leq k} |\mathbf{X}_i^T (\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}})|$  where  $\mathbf{X}_i$  is the  $i$ th column of the design matrix under consideration. Among the models that are consistent with the data, the estimator aims to select the one with the least complexity. This implies selecting the most sparse vector of parameters. A well known property of the Dantzig selector is that the expression in Equation (2.3) can be converted into a linear program. The linear program is written

$$\min \mathbf{c}^T \mathbf{x} \text{ subject to } \mathbf{A}\mathbf{x} \geq \mathbf{b} \text{ and } \mathbf{x} \geq \mathbf{0} \quad (2.4)$$

where

$$\mathbf{c} = \begin{bmatrix} \mathbf{1}_k \\ \mathbf{0}_k \end{bmatrix}, \mathbf{A} = \begin{bmatrix} \mathbf{X}^T \mathbf{X} & -\mathbf{X}^T \mathbf{X} \\ -\mathbf{X}^T \mathbf{X} & \mathbf{X}^T \mathbf{X} \\ 2\mathbf{I}_k & -\mathbf{I}_k \end{bmatrix}, \mathbf{b} = \begin{bmatrix} -\mathbf{X}^T \mathbf{y} - \delta \mathbf{1}_k \\ \mathbf{X}^T \mathbf{y} - \delta \mathbf{1}_k \\ \mathbf{0}_k \end{bmatrix} \text{ and } \mathbf{x} = \begin{bmatrix} \mathbf{u} \\ \mathbf{u} + \boldsymbol{\beta} \end{bmatrix}.$$

For supersaturated designs the linear system contains more unknowns than available equations. Hence, it is necessary to assume factor sparsity when applying the Dantzig selector. In this case only some of the entries in  $\hat{\boldsymbol{\beta}}$  are different from zero, making the system solvable. The fact that the Dantzig selector can be rewritten into a linear programming problem is one of the main advantages of the method. Algorithms for linear programming problems are in general considered to be relatively easy to use and fast to perform. A large range of software provide built in functions that are designed to solve linear programming problems.

One of the challenges related to the Dantzig selector is the estimation of the optimal value of the tuning parameter,  $\delta$ . This is an important issue as the consistency of the Dantzig selector for estimation and model selection is highly dependent on this value. From the expression in Equation (2.3) it is seen that the tuning parameter controls the correlation between the design matrix and the residuals. A low value of  $\delta$  would cause the method to have small residuals. This would lead to over-fitting of the model because some inert factors are included in the model in order to explain the variation in the response. In the opposite case where the value of  $\delta$  is chosen too high, large residuals are allowed in the estimation. These residuals may explain so much of the variation in the response that too few of the factors are identified as active. By Candes and Tao (2007) it was suggested to choose

$$\delta = \sigma \sqrt{2 \log(k)}, \quad (2.5)$$

but they emphasize that this is not the optimal value in all cases. This value of  $\delta$  entails estimation of  $\sigma$ , as this value rarely is known in real experiments. Whenever time and resources are available, Candes and Tao recommend to use Monte Carlo simulations instead of the expression in Equation (2.5), in order to find the optimal  $\delta$ .

There are some challenges and restrictions related to the Dantzig selector. It is known that the Dantzig selector is a biased estimator, producing parameter values that are biased downwards. In order to obtain unbiased estimates and improve the quality of the analysis, Candes and Tao (2007) introduced a version of the method called the Gauss-Dantzig selector. They recommended to use the Dantzig selector to determine which terms to include in the model, and then use the method of least squares to estimate the parameters in the model. A lot of effort has been spent in the attempt to further improve the Dantzig selector. Some of the other well known methods are named generalized Dantzig, double Dantzig, ridge Dantzig and interpolated Dantzig selector. These methods are introduced and discussed in James and Radchenko (2009).

For some experimental designs the Dantzig selector is inconsistent for model selection. The Dantzig selector is intended for sparse cases where the design matrix follows the uniform uncertainty principle, hereinafter referred to as the UUP. The UUP was discussed in Candes and Tao (2007), and it can be explained as follows. Let  $X_T$  denote a matrix containing the columns of the design matrix corresponding to the values in the subset  $T \subset \{1, \dots, k\}$ . The smallest value of  $\delta_S$  such that

$$(1 - \delta_S) \|c\|_{l_2}^2 \leq \|X_T c\|_{l_2}^2 \leq (1 + \delta_S) \|c\|_{l_2}^2 \quad (2.6)$$

for all subsets  $T$  with the number of columns less than or equal to  $S$ , and with coefficient sequence  $(c_j)_{j \in T}$ , is referred to as the  $S$ -restricted isometry constant. If the inequality

$$\delta_S + \delta_{2S} + \delta_{3S} < 1 \quad (2.7)$$

is true, then the Dantzig selector will be able to estimate the values of  $\beta$  with a probability close to one, given that sufficiently few factors are active. For this requirement to hold all subsets comprising  $S$  or less columns have to be uniformly close to orthonormal. Under the UUP the Dantzig selector holds the oracle property, i.e. it has the ability to select active effects with a probability converging to one, and provide estimates of the non-zero coefficients that are asymptotically normally distributed, with means and covariances identical to what they would have been if the zero coefficients had been known in advance. Given that the requirements of the factor sparsity and the UUP are met, Candes and Tao (2007) found that with a probability close to one

$$\|\hat{\beta} - \beta\|_{l_2}^2 \leq C^2 \cdot 2 \log k \cdot \left( \sigma^2 + \sum_i \min(\beta_i^2, \sigma^2) \right) \quad (2.8)$$

where  $C$  is a constant. From Equation (2.8) we see that the error made is relatively small for designs with a low number of columns, but that the Dantzig selector may be inconsistent for variable selection whenever  $k$  is large.

In Phoa, Pan and Xu (2009) it was mentioned that the given requirements are too strong to be valid for most designs, especially in the high dimensional case. This is mainly due to columns not being orthogonal in the design matrix. For experiments with large designs and many active factors, the computations needed in order to check these requirements are highly expensive. If the design matrix does not meet the requirements of the UUP it may be problematic to estimate all the parameters. This is because subsets of factor columns are correlated, in the sense that they are not orthogonal to each other. However, it may be the case that only some of the active factor columns violate the UUP. In this case most of the parameters can be estimated, even though the design matrix does not follow the UUP. It might be tempting to use the Dantzig selector for cases where the inequalities in (2.6) and (2.7) are not true, but for this case, there is no guarantee for the performance of the method.

### 2.2.2 The Lasso

Another well known method for linear model estimation in high dimensions is the Lasso. An introduction and discussion of this method is given in Tibshirani (1996). The letters in Lasso are the initials for Least Absolute Shrinkage and Selection Operator. The estimated parameters are the solution to the optimization problem

$$\min_{\hat{\boldsymbol{\beta}} \in \mathbb{R}^k} \|\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|_{l_2} \text{ subject to } \|\hat{\boldsymbol{\beta}}\|_{l_1} \leq t \quad (2.9)$$

where  $\|\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|_{l_2}^2 = \sum_{i=1}^n (y_i - x_i^T \hat{\boldsymbol{\beta}})^2$  is the squared  $l_2$ -norm of the residuals. The squared error loss is restricted by the positive parameter  $t$ . The Lasso can also be written on the Lagrangian form

$$\min_{\hat{\boldsymbol{\beta}} \in \mathbb{R}^k} \frac{1}{2} \|\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|_{l_2}^2 + \lambda \|\hat{\boldsymbol{\beta}}\|_{l_1} \quad (2.10)$$

where  $\lambda$  is a positive regularization parameter. A number of well known statistical softwares provide built in algorithms with solution to the quadratic programming problem in the Lasso.

For design matrices with rank equal to the number of factors, the problem in Equation (2.9) is strictly convex, which ensures a unique solution to the problem. This is the case for the 12 run PB design considered in this thesis, but not for the expanded matrix which also includes interaction columns. The required conditions for Lasso to perform well are presented and discussed in van de Geer and Bühlmann (2009). Given that the conditions of the Lasso holds and the vector of parameters is sufficiently sparse

$$\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_{l_2}^2 \leq C \cdot n\sigma^2 \quad (2.11)$$

with a probability close to one, for a constant  $C$ .

The solution produced by the Lasso is quite close to the results obtained by the Dantzig selector. In Meinshausen, Rocha and Yu (2007) it is stated that under the condition

$$(\mathbf{X}^T \mathbf{X})_{jj}^{-1} > \sum_{i \neq j} |(\mathbf{X}^T \mathbf{X})_{ij}| \quad \text{for all } j = 1, \dots, k \quad (2.12)$$

the two methods produce identical solutions for all values  $\delta > 0$ . Even though the results are quite similar for cases where the inequality in (2.12) does not hold, there are small differences in the estimates found by the methods. There are still disagreements among statisticians regarding which of the two methods that are generally preferred.

### 2.2.3 Orthogonalization of augmented regression models

Consider now the model in Equation (2.2) rewritten as

$$\mathbf{y} = \mathbf{X}_1 \boldsymbol{\beta}_1 + \mathbf{X}_2 \boldsymbol{\beta}_2 + \boldsymbol{\epsilon}. \quad (2.13)$$

In this expression,  $\mathbf{X}_1$  could for instance consist of main effect columns and  $\mathbf{X}_2$  of two-factor interaction columns. In this case, there might not be orthogonality between  $\mathbf{X}_1$  and  $\mathbf{X}_2$ . By orthogonalization, which is the process of making vectors orthogonal to each other, it is possible to obtain unbiased estimates of effects. The goal of the orthogonalization is to find an alternative form of Equation (2.13) which is composed of two orthogonal parts. A thorough derivation of this expression is given in Box and Draper (2007), and the main steps are presented below. Consider first the model in Equation (2.2). The derivation is based on the fact that the residuals  $\mathbf{y} - \hat{\mathbf{y}}$  obtained by least squares estimation are orthogonal to the columns of the design matrix  $\mathbf{X}$ . Let  $\mathbf{y}_{\cdot 1} = \mathbf{y} - \hat{\mathbf{y}}$  denote the component of the response that is orthogonal to the columns of  $\mathbf{X}_1$ . Then the expression

$$\mathbf{y}_{\cdot 1} = (\mathbf{I} - \mathbf{R}_1) \mathbf{y} \quad (2.14)$$

where  $\mathbf{R}_1 = \mathbf{X}_1 (\mathbf{X}_1^T \mathbf{X}_1)^{-1} \mathbf{X}_1^T$  is found by writing out the expression for  $\hat{\mathbf{y}}$  and inserting the least squares estimates of the parameters,  $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$ . Further, let  $\mathbf{X}_{2 \cdot 1} = \mathbf{X}_2 - \hat{\mathbf{X}}_2$  denote a matrix where all of the columns are orthogonal to the columns of  $\mathbf{X}_1$ . With the same reasoning as above, this expression is given by

$$\mathbf{X}_{2 \cdot 1} = \mathbf{X}_2 - \mathbf{X}_1 \mathbf{A}, \quad (2.15)$$

where

$$\mathbf{A} = (\mathbf{X}_1^T \mathbf{X}_1)^{-1} \mathbf{X}_1^T \mathbf{X}_2 \quad (2.16)$$

is called the alias or bias matrix. The alternative form of the model in Equation (2.13) is then written

$$\begin{aligned}\mathbf{y} &= \mathbf{X}_1(\boldsymbol{\beta}_1 + \mathbf{A}\boldsymbol{\beta}_2) + (\mathbf{X}_2 - \mathbf{X}_1\mathbf{A})\boldsymbol{\beta}_2 + \boldsymbol{\epsilon} \\ &= \mathbf{X}_1\boldsymbol{\beta} + \mathbf{X}_{2.1}\boldsymbol{\beta}_2 + \boldsymbol{\epsilon}.\end{aligned}\tag{2.17}$$

As  $\mathbf{X}_1$  and  $\mathbf{X}_{2.1}$  are orthogonal, unbiased estimates are obtained by performing the method of least squares on each of the two parts of Equation (2.17) separately. The corresponding expressions are given by

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}_1^T \mathbf{X}_1)^{-1} \mathbf{X}_1^T \mathbf{y}\tag{2.18}$$

and

$$\hat{\boldsymbol{\beta}}_2 = (\mathbf{X}_{2.1}^T \mathbf{X}_{2.1})^{-1} \mathbf{X}_{2.1}^T \mathbf{y}.\tag{2.19}$$

When performing orthogonalization as part of a variable selection procedure,  $\mathbf{X}_1$  and  $\mathbf{X}_2$  are usually chosen by the analyst. Hence, the values in Equation (2.18) can be calculated directly, and  $\hat{\boldsymbol{\beta}}_2$  can be found after calculating  $\mathbf{X}_{2.1} = \mathbf{X}_2 - \mathbf{X}_1\mathbf{A}$ . The estimated values for  $\hat{\boldsymbol{\beta}}_1$  are either found by  $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_1 + \mathbf{A}\hat{\boldsymbol{\beta}}_2$ , or by swapping the matrices  $\mathbf{X}_1$  and  $\mathbf{X}_2$  in the calculation of the bias matrix and  $\hat{\boldsymbol{\beta}}_{2.1}$ . Given that the columns of the correct active effects are included in  $\mathbf{X}_1$  and  $\mathbf{X}_2$ , the estimate  $\hat{\boldsymbol{\beta}}_2$  is unbiased. The method requires that the total number of columns does not exceed  $n - 1$ . Due to this restriction it is often necessary to divide  $\mathbf{X}_2$  into parts and perform orthogonalization on each of them.

# Chapter 3

## Methods

An introduction to the methods studied in this thesis is given in this chapter. The methods are introduced in the order of the Dantzig selector, the graphical Dantzig selector, the projection based Dantzig selector, the Lasso, the projection based method, the partial F-test and the graphical method using orthogonalization. The same ordering is used in Chapter 4 where the methods are tested on some predefined models. At the end of the chapter a short introduction to the R code is given.

### 3.1 The Dantzig selector

The Dantzig selector is performed as described in Subsection 2.2.1 with the tuning parameter  $\delta = \sigma\sqrt{2\log(k)}$ , as suggested by Candès and Tao (2007). After solving the linear problem in Equation (2.4) a vector with the values of  $\hat{\beta}$  is returned. The effects are ordered from high to low absolute value of the estimated parameters. In this thesis it was chosen to display the six first values in a table, but this may not be the optimal number of effects for all cases.

When the design matrix is used as  $\mathbf{X}$  in Equation (2.3) only parameters for main effects are estimated. When interaction effects are to be investigated, the design matrix is expanded with the corresponding interaction columns. For the examples in this thesis, the six first effects obtained for the main effects only, and the six first for the main effects and two-factor interactions together are presented in the same table. Whenever the very same effect is listed as the most likely effect to be active in both cases, the confidence in that effect is increased. To determine how many effects to include in the model, the estimated parameter values may be considered. It is also possible to use forward selection, backward elimination or a stepwise procedure as discussed in Section 2.2.

Ever since Candès and Tao published their article about the Dantzig selector it has been searched for ways to expand the method in order to avoid the selection of the optimal  $\delta$ . Among these is the graphical method described in the next section.

## 3.2 A graphical Dantzig selector

A graphical method using the Dantzig selector was proposed by Phoa, Pan and Xu (2009) for factor screening in supersaturated designs. This graphical method avoids the challenge of selecting an optimal value for  $\delta$  by solving the Dantzig selector for an entire set of values for  $\delta$ . The procedure suggested by Phoa, Pan and Xu (2009) results in a plot of the estimated parameters  $\hat{\beta}$  against  $\delta$  by performing the following four steps.

1. Standardize data so that  $\mathbf{y}$  has mean 0 and the columns of  $\mathbf{X}$  have equal lengths. Compute  $\delta_0 = \max |\mathbf{x}_i^t \mathbf{y}|$ , where  $\mathbf{x}_i$  is the  $i$ th column of  $\mathbf{X}$ .
2. Solve the linear program in Equation (2.4) to obtain the Dantzig selector  $\hat{\beta}$  for values of  $\delta$  ranging from 0 to  $\delta_0$ .
3. Make a profile plot of the estimates by plotting  $\hat{\beta}$  against  $\delta$ .
4. Identify important effects by inspection of the profile plot.

The vector containing the values of  $\delta$  should consist of hundreds of points evenly spaced from 0 to  $\delta_0$ . An effect that is large relative to the other effects and also reduces slowly to zero is regarded as an important effect. Similarly to the Dantzig selector, this method also allows for investigation of interaction effects. This is done by expanding the design matrix with the corresponding interaction columns, and performing the same procedure as given above. To determine how many terms to include in the model, forward selection can be performed where the terms are added in the order indicated by the plot. An alternative is to perform backward elimination with the most likely active effects found by the graphical Dantzig selector. In this thesis it was determined to perform backward elimination on the six first effects, but this may not always be the optimal number of effects to start with.

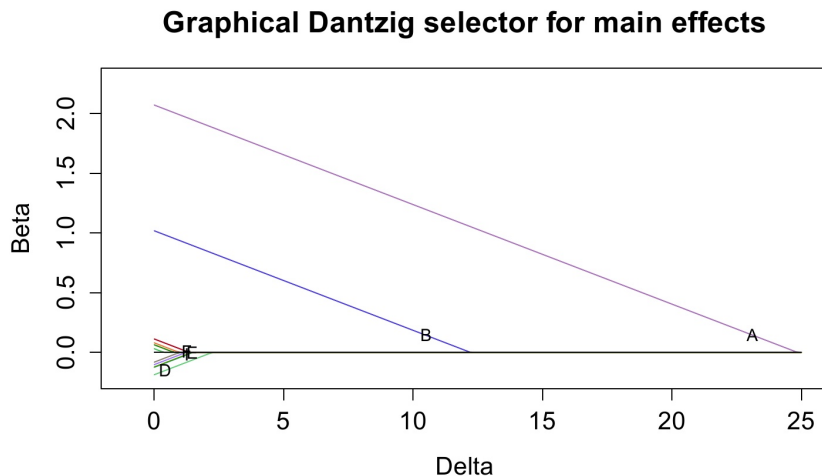


Figure 3.1: Example of a plot obtained by the graphical Dantzig selector for data generated from  $y = 2 + 2A + B + \epsilon$  with  $\epsilon \sim N(0, \sigma^2)$  where  $\sigma^2 = 0.1$ .

Lets consider an example of the procedure described above using a 12 run PB design. The data used in this example was generated from the model  $y = 2 + 2A + B + \epsilon$  where the noise is given by  $\epsilon \sim N(0, \sigma^2)$  with  $\sigma^2 = 0.1$ . Figure 3.1 displays the plot obtained when only considering main effects. The plot indicates that A and B are active, because these main effects are relatively large and long lasting. To determine whether any of the other factors should be included in the model, a model may be fitted with only the main effect A. Then the effects  $B, D, F$  and so on can be added to the model as long as all the effects are found to be significant.

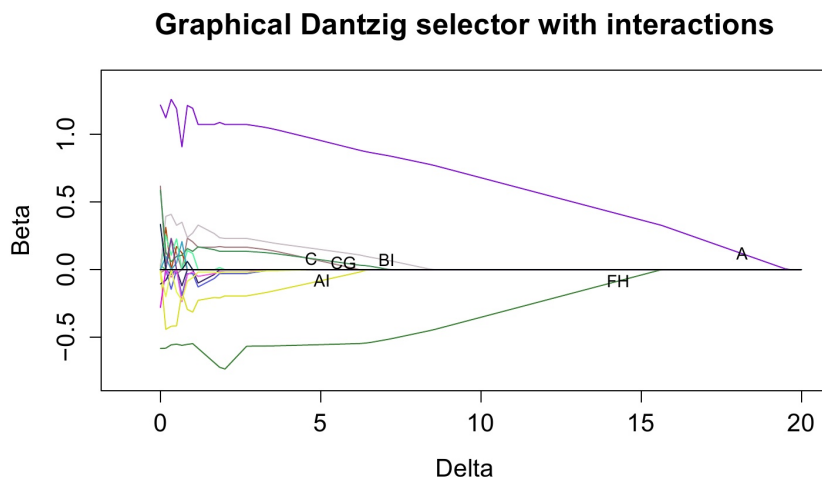


Figure 3.2: Example of a plot obtained by the graphical Dantzig selector for data generated from  $y = 2 + 2A + C + BD + \epsilon$  with noise  $\epsilon \sim N(0, \sigma^2)$  where  $\sigma^2 = 0.1$ .

As an additional example the procedure is performed using a 12 run PB design expanded with all two-factor interaction columns. The plot obtained when the graphical Dantzig selector is applied to data generated from the model  $y = 2 + 2A + C + BD + \epsilon$  with  $\epsilon \sim N(0, 0.1)$  is given in Figure 3.2. From this plot one may erroneously conclude that A and FH are active. The design matrix including the interaction columns did not meet the requirements of the UUP, and hence, there was no guarantee for the performance of the method. Also, as there are four active factors in the model it may be questioned whether factor sparsity is fulfilled in this example.

There are many cases where the methods using the Dantzig selector are not appropriate because the UUP does not hold. In the paper written by Phoa, Pan and Xu (2009) it is recommended to perform other analysis in addition to this method in order to verify the results. The Dantzig selector involves solving a linear programming problem, which is in general regarded as fast and easy to perform. For this procedure the linear program has to be solved for over a hundred values of  $\delta$ , making the code more computationally expensive.



### 3.3 A projection based Dantzig selector

In the search for an improved algorithm an expanded and slightly modified version of the graphical procedure in Section 3.2 is suggested. For this method it is assumed at most six active effects and at most four active factors. The active effects are identified by performing the following six steps.

1. For all combinations of  $n \leq 4$  factors, make a design matrix with the  $n$  main effects together with their two-factor interactions.
2. Run each of the design matrices through the code described in Section 3.2.
3. For each outcome, sort the effects according to how slowly the graph reduces to zero, from longest lasting to shortest lasting.
4. For each combination of the  $n$  factors, fit a model with the six longest lasting effects.
5. Calculate and compare the residual variances of the fitted models.
6. Identify the correct model by selecting the model with the lowest residual variance.

There is a challenge associated with the selection of factors in Step 3. By Phoa, Pan and Xu (2008) it was suggested to select the largest and longest lasting effects. As the code in Section 3.2 has to be run a large number of times the selection has to be done without visually evaluating each of the plots. Hence, it was decided that the selection is to be done by only considering how long the effects are lasting. For the 12 run PB design, this simplification will not influence the selection of factors when only considering main effects, as these graphs are parallel. However, when including interactions it may cause the algorithm to make some incorrect selections of effects. For most cases this simplification is not expected to have a great impact on the results, as the active effects should possess both of these properties.

Due to hidden projection properties, it is possible to perform the procedure described above with  $n = 4$  when using the 12 run PB design. When projecting onto any four factors it is possible to estimate all main effects and two-factor interactions when using the 12 run PB design, given that it is appropriate to assume that all three- and four-factor interactions are negligible. It is usually preferred to use  $n = 4$  whenever the experimental design allows for it. This is because the use of  $n \leq 3$  will eliminate the possibility of investigating four active factors. An advantage of this method is that the procedure has the ability to select a model with less than four active factors even though the method was performed with  $n = 4$ . Due to these advantages,  $n = 4$  was chosen for the examples in this thesis.

The procedure presented in Section 3.2 has to be performed for all possible combinations of four columns. For the 12 run PB design with 11 factors this means that the code has to be run  $\binom{11}{4} = 330$  times. For a  $\delta$  comprising 100 values, the

Dantzig selector has to be solved 33 000 times. Hence, this method is even more computationally expensive than the previously described method.

It might not always be the case that the true model has the lowest variance. A model that is not sparse may explain the variation in the response equally well as the true model. For cases with large noise there may exist models that provide a better fit for the data compared to the model with the correct effects. As a start it may be of interest to rank the models according to the estimated residual variance and find out if the true model belongs to a subset of the  $s$  models with the smallest residual variances. In the code made for this thesis  $s$  was set equal to 10, but this number can easily be changed in the R code. It should be noted that there exists several models containing all of the correct effects whenever there are less than six terms in the true model. The fewer active factors, the more models comprising the correct effects. This increases the probability of finding a model with all the correct terms ranked first on the list. Further analysis can be performed in order to be able to select one single model. One suggestion is the method described in Subsection 3.3.1.

### 3.3.1 An expansion of the method

It is suggested to expand the projection based Dantzig selector by considering the ten models with the smallest residual variances found by the procedure described above. Each of these ten models are fitted to a linear model. In the attempt to avoid over-fitted models, backward elimination is used to reduce the models. The reduced model with the lowest residual variance is chosen as the final model. A second alternative is to select the model according to the Akaike information criterion (AIC). For a model with  $k$  parameters the expression for the AIC is give by

$$AIC = 2k - 2 \ln(L) \tag{3.1}$$

where  $\ln L$  is the maximal value of the log-likelihood function. A model with a low AIC value is to be preferred over a model with a high value of AIC. This criterion is appropriate in the case of over-fitting as it adds a penalty to complex models. A third approach is the Bayesian information criterion (BIC). This criterion is quite similar to the AIC, but it adds an even stronger penalty to complex models. The expression for BIC is given by

$$BIC = k \ln(n) - 2 \ln(L) \tag{3.2}$$

where  $n$  is the number of observations. A model with a low BIC is preferred over a model with a high BIC. During the study of this method, all of these three criteria were tested, and the results were compared to each other.

### 3.4 The Lasso

With the purpose of comparison to the Dantzig selector, the Lasso was studied. It should be noted that Lasso initially was intended for prediction, so variable selection is outside its original area of use. When Lasso is applied to the data, the standardized coefficients are plotted against the standardized  $l_1$ -norm of the vector of coefficients. In this plot the solution path for each effect is piecewise linear. A large absolute value of the standardized coefficient which slowly decays as the  $l_1$ -norm goes to zero, is an indication of an active effect.

Consider the plot given in Figure 3.3 where the Lasso has been applied to data generated from the model  $y = 2 + A + B + \epsilon$  with  $\epsilon \sim N(0, \sigma^2)$  where  $\sigma^2 = 0.1$ . The numbers to the right in the plot indicate which effects are plotted, as the solution paths are numbered in the order given in the design matrix. Hence, for a design matrix with headers given by the alphabet number 1 corresponds to main effect A, 2 corresponds to main effect B, and so on.

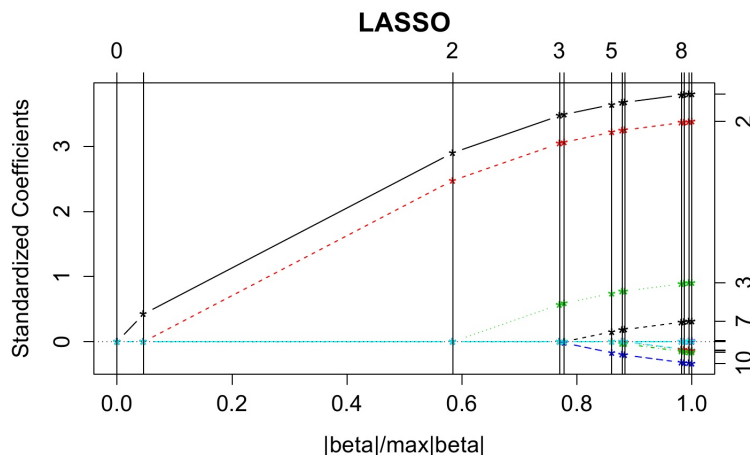


Figure 3.3: Example of plot obtained by the Lasso for data generated from the model  $y = 2 + A + B + \epsilon$  with  $\epsilon \sim N(0, \sigma^2)$  and  $\sigma^2 = 0.1$ .

In Figure 3.3 it is seen that the built-in function for the Lasso in the programming language R does not label all of the solution paths. However, it is possible to get a printout of the effects given in the order of importance found by the Lasso. The printout in R revealed that the longest lasting solution path in Figure 3.3 belongs to A. Hence, the plot in Figure 3.3 indicates that A and B are active. To test whether any of the other effects should be included in the model, it is possible to use the methods discussed in Section 2.2.

The optimal value of  $\lambda$  in Equation (2.10) used for parameter estimation is chosen by cross-validation in R. The estimated parameters are ordered according to decreasing absolute value.

### 3.5 A projection based method

In Wiik (2013) a factor based approach for variable selection in non-regular designs was studied. This method was also discussed in Tyssedal og Samset (1997). This method utilizes the projection properties of the design, and hence, it is referred to as the projection based method in this thesis. For the models tested in Wiik (2013) this method proved to be valuable for identifying active factors in experiments using the 12-run PB design. When considering projections containing up to three factors it can be seen that there are repeated rows in the subset of the design matrix. If a given subset contains the only active factors, the expected values within each group of replicated rows are equal and the variance within each group should be low. The variance within each groups of replicated rows can thereby be used to identify the active factors by selecting the subset of factors with the lowest estimated variance. For subsets of four factors there is only one replicated row. Hence, for models with more than three active factors it is necessary to base the analysis on other methods.

In the 12 run PB design with 11 factors there are 11 main effects,  $\binom{11}{2} = 55$  two-factor interactions and  $\binom{11}{3} = 165$  three-factor interactions. In the search for active factors the variance within each group of repeated rows should be estimated for all of these possible combinations of columns in the design matrix. The formula

$$\hat{s}^2 = \frac{1}{\sum_{i=1}^g (r_i - 1)} \sum_{i:\text{replicated}} \sum_{j=1}^{r_i} (y_{ij} - \bar{y}_i)^2 \quad (3.3)$$

is used, where  $\mathbf{y}$  is the response vector,  $g$  the number of groups and  $r_i$  the number of repeated rows within each group. This formula is also given in Tyssedal, Grinde and Røstad (2006). The number of groups and repeated rows depends on the number of active factors, as presented in the list below.

**One active** Two groups containing six repeated rows

**Two active** Four groups with three repeated rows

**Three active** Four pairs of repeated rows, and four non-repeated rows.

It should be kept in mind that the method proposed in this chapter does not assume any particular form of the model. Hence, if the results for three active factors shows that ABC provides a lower variance than any other combination of factors, then this does not mean that the interaction ABC is present in the true model. It only suggests that the factors A, B and C are active.

An advantage of this method is that few assumptions are needed. A problem that may occur when using the method presented in this chapter is the risk of over-fitting. Over-fitting may be detected by observing that many of the estimated variances are small. The method have some limitations to be aware of. As there are only one replicated row for a subset of four factors, the analysis is restricted to at most three active factors. If there are more than three active factors, the method may produce erroneous results.

### 3.6 The partial F-test

The partial F-test is used to determine whether it is appropriate to include additional terms in a model. Hence, it can be used in combination with some of the previously discussed methods in order to verify that all terms have been found. When the partial F-test is performed a set of new models are fitted by expanding the original model with effects one wants to find out whether should be included in the model. The original model is often referred to as the reduced model, and the expanded as the full model. Let  $n$  denote the number of observations,  $r$  the number of effects in the reduced model and  $f$  the number of effects in the full model. The test statistic for the partial F-test is then given by

$$F = \frac{(SSR_{full\ model} - SSR_{reduced\ model})/(f - r)}{SSE_{full\ model}/(n - f - 1)} \quad (3.4)$$

where  $SSR$  is the regression sum of squares and  $SSE$  the error sum of squares. The effects added to the reduced model are found to be significant if  $F > F_{\alpha, (f-r), (n-f-1)}$ . It is often desirable to investigate whether an additional factor should be added to the model. In this case a set of new models are fitted by expanded the model with each of the remaining factors, together with their interactions with the effects already in the model. For example, if the factor A has been found to be active, the model is expanded by adding  $X + AX$  for  $X = B, \dots, K$ . The partial F-test is applied to all of these ten models in order to test for significance of the added effects. For a model with the factors A and B the model is expanded with  $X + AX + BX$  for  $X = C, \dots, K$ . For a model with the three factors A, B and C the model is expanded with  $X + AX + BX + CX$  for  $X = D, \dots, K$ . Three-factor interactions could also be added to these models.

### 3.7 A graphical method using orthogonalization

Scatter plot of effects plotted against the absolute value of their estimated contrasts can provide useful information regarding active effects. These plots are often referred to as contrast plots. For the 12 run PB design the estimated contrasts are equal to  $\frac{1}{6}\mathbf{X}^T\mathbf{y}$ , which is twice their parameter value obtained by least squares. When considering contrasts of main effects,  $\mathbf{X}$  is equal to the design matrix. Similarly, when considering contrasts of interaction effects,  $\mathbf{X}$  is equal to the matrix comprising the columns of the interactions in question. In this section the graphical method proposed by Tyssedal and Niemi (2011) is discussed where the analysis is based on interpretation of contrast plots. In order to ease the interpretation of these plots, orthogonalization is performed as discussed in Subsection 2.2.3.

For the 12 run PB design all two-factor interactions are partially aliased with all main effects, except for the two factors that it is involved with. Any two-factor interaction is orthogonal to the columns of the main effects that are included in the given interaction. The aliasing coefficient is zero in the orthogonal case, and either  $-\frac{1}{3}$  or  $\frac{1}{3}$  otherwise. Hence, the estimated contrasts of the factors that are not

included in the true model, neither as a main effect or as part of an interaction, have a certain structure. For a model with one interaction term, the main effect contrasts of the inert factors have approximately the same value in a contrast plot. Contrasts deviating from this pattern is an indication of active effects. For models with two interaction terms the main effect contrasts of the inert effects form approximately two horizontal lines in the plot. For the case with more than two interaction terms, the contrast plot may be difficult to interpret. After the contrast deviating from the pattern in the main effects contrast plot have been detected, the analysis is proceeded by plotting interaction contrast plots. The principle of effect heredity can be used as a guidance to which interaction contrasts to plot. Due to this principle the contrast plot of the interactions involved with factors that seems to be active based on the main effects contrast plot should be performed.

Consider the main effects contrast plot in Figure 3.4a. This plot was obtained for data generated from the model

$$y = 2 + 4A + 2C + BC + 2CD + \epsilon \quad (3.5)$$

with noise  $\epsilon \sim N(0, \sigma^2)$  and  $\sigma^2 = 0.1$ , using a 12 run PB design. It is observed that the absolute value of the contrasts for B and D-K approximately cluster in two lines in this plot, one with values close to 0.5 and the others with values close to 2. Further, it is observed that A and C deviates form this pattern. Based on this observation it is concluded that there are two interaction terms, and that A and C are important for the model in question. For further analysis it is natural to proceed by plotting the contrast plots for interactions with A and C, as shown in Figure 3.4b and Figure 3.4c. The plot in Figure 3.4b is not easy to interpret as no effects stands out as clear candidates for being active. In the interaction contrast plot in Figure 3.4c it is seen that AC is close to zero and that BC and CD are seemingly on their own level. The rest of the interactions involved with C lie approximately on one line, indicating that they may be inert. This indicates that A or C or both is included in the model as main effects, and that BC and CD are the two interaction effects.

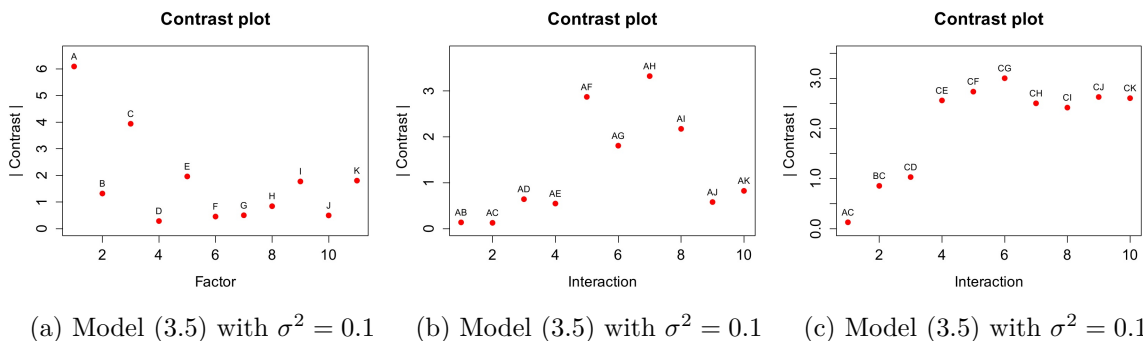


Figure 3.4: Example of a main effects contrast plot where data was generated from Model 3.5 with noise  $\epsilon \sim N(0, \sigma^2)$  and  $\sigma^2 = 0.1$ .

In order to support the interpretation of contrast plots, orthogonalization can be performed as described in Subsection 2.2.3. Tyssedal and Niemi (2011) proposed a graphical method where interpretation of contrast plots were combined with a orthogonalization procedure. This method was given as follows

1. Perform a main effects contrast plot. Consider contrasts that cluster in groups likely to represent inert factors and those who don't likely to represent the active ones.
2. Perform two-factor interaction plots and interpret them the same way as for the main effects contrast plot. Normally these plots are made for interactions involved with main effects that potentially have been judged active from step 1. For step 1 and 2, be aware of that if a contrast who don't cluster in groups came out with a value close to zero a possible explanation may be that the contrast is orthogonal to some active effects. Hence, small values of contrasts may point to potentially active effects.
3. Whenever interpretation of main effects and two-factor interactions contrast plots becomes difficult due to masking of effects, large noise or many active two-factor interactions, perform orthogonalization possibly in several steps. This is a very powerful procedure that in a few operations may reveal the complexity of the problem at hand.

The orthogonalization procedure was here referred to as alias reduction. When orthogonalization is performed there are many possibilities in terms of the choice of  $\mathbf{X}_1$  and  $\mathbf{X}_2$  in Equation (2.17), but it is wise to base the choice on interpretation of the contrast plots. In this thesis  $\mathbf{X}_1$  is chosen as the main effects and interactions that are considered to be active based on the contrast plots. Effect columns that are to be investigated are included in  $\mathbf{X}_2$ . Recall that the number of columns in the two matrices cannot exceed  $n - 1$ , such that it often will be necessary to divide  $\mathbf{X}_2$  into parts and perform orthogonalization on each of them. The estimates of  $\hat{\beta}_1$  and  $\hat{\beta}_2$  are found as described in Subsection 2.2.3. After orthogonalization is performed the matrices may be updated and the procedure repeated. In order to be able to directly compare the contrast plots before and after orthogonalization, it is important to remember to multiply the estimated parameters  $\hat{\beta}_1$  and  $\hat{\beta}_2$  by two.

After orthogonalization the contrast plots in Figure 3.5 were obtained for Model (3.5). Figure 3.5b was obtained after orthogonalization with  $\mathbf{X}_1 = [BC, CD]$  and  $\mathbf{X}_2 = [A, B, C, D, E, F]$ , and thereafter with  $\mathbf{X}_1 = [A, C, BC, CD]$  and  $\mathbf{X}_2 = [G, H, I, J, K]$ . The orthogonalization conducted in order to obtain the plot in Figure 3.5b, was first performed with  $\mathbf{X}_1 = [A, C]$  and  $\mathbf{X}_2 = [AC, BC, CD, CE, CF, CG]$ , and then with  $\mathbf{X}_1 = [A, C, BC, CD]$  and  $\mathbf{X}_2 = [CH, CI, CJ, CK]$ . From these plots it seems reasonable to conclude that the model consists of the effects A, C, BC and CD.

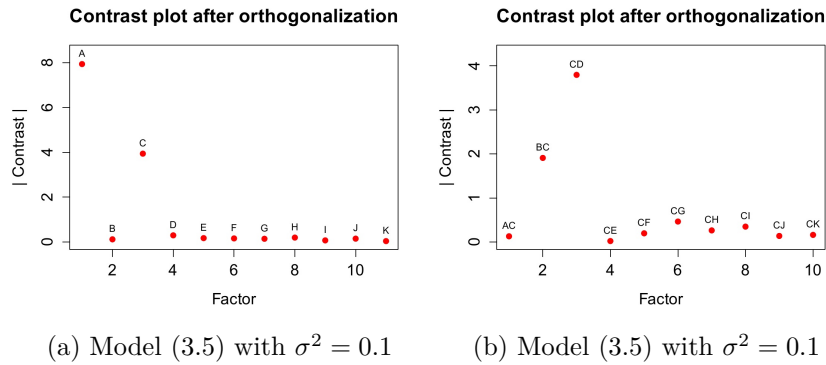


Figure 3.5: Example of a main effects contrast plot after orthogonalization, where data was generated from Model 3.5 with noise  $\epsilon \sim N(0, \sigma^2)$  and  $\sigma^2 = 0.1$ .

The method presented in this section has the advantages that a large amount of information usually can be drawn from the contrast plots with very little computational effort. The procedure might provide an indication of active effects, or it might give a clear impression of the complexity of the problem at hand. Either way, it forces the analyst to consider different aspects of the problem. This graphical method may be a valuable tool in combination with other methods, as it often provides information regarding the form of the model.

## 3.8 R software

### 3.8.1 Background

The programming language R was used when implementing the methods presented in this project. This language is mainly used for statistical computing and as a graphical tool. The R software is free for anyone to use, and this is one of the reasons why it is extensively used by statisticians. It provides the user with the ability to develop own methods, and to perform analysis based on embedded functions. For an experienced user of R the embedded functions provide a lot of information compared to the time and effort spent. Often, a few lines of code is enough to gather the desired information. For inexperienced users the embedded function can be challenging to use as limited information is available on the internet. When using R one should be aware of the fact that R does not provide any warranty. It is important to have knowledge about the analysis performed in R in order to detect any failures in the results produced.

### 3.8.2 Implementation of code

Most of the methods in this project were implemented using own coding, but some embedded functions were used as well. The function  $lp()$  was used to solve the linear problem in Equation 2.4 as part of the Dantzig selector. The function  $lm()$  was used to fit data to a linear model. The Lasso method was implemented using the built in function  $lars()$  in R. Lars is a least angle regression algorithm, where a slight



modification of the algorithm provides the solution path of the Lasso. The optimal value for parameter estimation was found using *mylars()*. It should be noted that *mylars()* makes use of an estimation procedure that provides slightly different estimates each time it is run. However, the variation in the estimated values are so small that it is expected to have a small influence on the selection of factors.

The code implemented for the methods studied in this thesis are given in Appendix B. The code was developed with the intention for it to be time efficient and provide accurate results with as few lines of code as possible. During implementation of the code it was focused on developing a flexible solution such that it would work on all of the data sets with either non or just a slight amount of modification. This ensured that the number of code lines was reduced to a minimum. To make sure that the code was implemented correctly the results obtained were compared to analysis performed by others and to functions embedded in R whenever available. The R code implemented for the procedure in Section 3.2 was tested by performing the same experiments as in Phoa, Pan and Xu (2009). The plot obtained was exactly equivalent to the plot displayed in their report.

The R code has some restrictions to be aware of. Most of the implemented functions are intended for the 12 run PB design matrix, but only small adjustments have to be made to the code in order for it to work on all design matrices. The code implemented for the graphical method using orthogonalization has to be modified for each model. This is because the form of the matrices in Equation (2.17) are dependent on visual interpretation of the contrast plots. In order to save space in the appendix, the code used for each single model was omitted. Only an example code is presented in the appendix. In order to perform orthogonalization on the other models, the code has to be adapted to the problem at hand.

Standard procedure in most programming languages, including R, is to use local variables whenever available, unless specifically stated otherwise in the code. During the implementation of the method described in Section 3.3 a deviation from this procedure was discovered. A local variable is a variable that is only available for a certain part of the code. For a function in R the local variables are either given to the function as input or they are declared within the function itself. The unexpected event occurred when using the embedded function *paste()* in order to define the form of a model. When this model was fitted by *lm()*, the embedded function caused the code to ignore the response vector given as input to the function, and instead fit a model with a different response vector which was defined outside of the function. This is very unexpected and it is something that users of R should be aware of. The solution to this problem was to assign the vector under consideration with the same name both inside and outside of the function.

# Chapter 4

## Examples

In order to investigate the methods described in Chapter 3 the R code was tested on data generated from a selection of predefined models. The data was generated with noise that was assumed to be normally distributed with zero mean and equal variance. The design matrix used was the slightly rewritten 12 run PB design matrix given in Table 4.1.

Run	A	B	C	D	E	F	G	H	I	J	K
1	-	-	-	-	-	-	-	-	-	-	-
2	-	-	-	-	-	+	+	+	+	+	+
3	-	-	+	+	+	-	-	-	+	+	+
4	-	+	-	+	+	-	+	+	-	-	+
5	-	+	+	-	+	+	-	+	-	+	-
6	-	+	+	+	-	+	+	-	+	-	-
7	+	-	+	+	-	-	+	+	-	+	-
8	+	-	+	-	+	+	+	-	-	-	+
9	+	-	-	+	+	+	-	+	+	-	-
10	+	+	+	-	-	-	-	+	+	-	+
11	+	+	-	+	-	+	-	-	-	+	+
12	+	+	-	-	+	-	+	-	+	+	-

Table 4.1: The 12 run PB design matrix slightly rewritten.

Data from a total of 15 different models was investigated, with six in the first example and nine in the second. This large amount of models was chosen in order to ensure a diversity of models that would facilitate exposure of strengths, weaknesses and limitations of the methods. The number of active factors range from one to four, where some include interaction effects and some only are composed of main effects. Data generated from the models in the first example were considered to be relatively easy to analyze because they have few active factors. The data from the models in the second example was expected to be more challenging to analyze.

To investigate different aspects of the methods, generation of data was performed two times for each model. For the first generation of data, the variance of the noise

was chosen to be  $\sigma^2 = 0.1$ , and for the second generation it was set equal to  $\sigma^2 = 0.5$ . The value

$$\sigma \leq \frac{\textit{smallest coefficient}}{2}$$

is often used as an upper noise limit for which the methods are expected to perform well. As  $\sigma = \sqrt{0.1} \approx 0.32$  is below this limit the methods were expected to provide reasonably good results when applied to the first data sets. It was expected that the methods would encounter some problems when using  $\sigma = \sqrt{0.5} \approx 0.71$ , as this exceeds the given limit.

In order to be able to directly compare the methods, the very same data sets were examined by all the methods. The data was generated two times for each model, one for each level of noise. The 30 data sets obtained were saved to a file, and loaded during the testing of the methods. Unless specifically stated otherwise in the text, these and only these data sets were used during the study of the methods. The generated values are presented in Appendix A.

Before the methods were applied to the data, some thoughts were made regarding their performance. The Dantzig selector was originally intended for sparse parameter vectors with noiseless data obtained by using a design matrix that meets the requirements of the UUP. For the 12 run PB design used in these examples the UUP holds when only main effects are considered. The design matrix that is expanded to also include two-factor interactions does not meet the requirements of the UUP. When including the columns of the three-factor interactions the design is even further away from fulfilling the requirements. Hence, it was expected that the methods that utilizes the Dantzig selector would perform significantly better on models with only main effects. The results found by the Lasso were expected to be quite similar to the ones obtained by the Dantzig selector, as discussed in James, Radchenko and Lv (2009). Further, it was anticipated that the projection based method would perform well for models with less than four active factors. For the graphical method using orthogonalization it was believed that the method either would be able to identify the active effects in the model, or that it otherwise would provide information that could be valuable to other analysis procedures. The results could for instance point to some of the active effects, indicate the number of interaction effects or emphasize the complexity of the problem. Because some of the data sets were generated with a high level of noise, and as some of the models pushed the limit in terms of factor sparsity, it was expected that a combination of methods would be necessary in order to identify the model effects.

The results produced by the methods are gathered in tables, and for some of them the results are displayed in the form of a plot. Most of the tables in the examples are organized with a reference to the true model to the left, the results for main effects in the center, and the results obtained when also including two-factor interactions to the right. Some of the plots obtained for  $\sigma^2 = 0.1$  are given in the text, but the rest are given in Appendix C.

## 4.1 Example 1

In this example all the methods in Chapter 3 were tested on the models

$$y = k + A + \epsilon \tag{4.1}$$

$$y = k + 2A + B + \epsilon \tag{4.2}$$

$$y = k + A + B + \epsilon \tag{4.3}$$

$$y = k + 3A + 2B + AB + \epsilon \tag{4.4}$$

$$y = k + 2A + 2B + AB + \epsilon \tag{4.5}$$

$$y = k + 2A + B + AB + \epsilon \tag{4.6}$$

with the constant  $k$  equal to 2. The data was generated with noise  $\epsilon \sim N(0, \sigma^2)$  using both  $\sigma^2 = 0.1$  and  $\sigma^2 = 0.5$ . Note that the first three models consist of main effects only, while the last three models include a two-factor interaction term. Due to the fact that the interaction term  $AB$  is orthogonal to both  $A$  and  $B$  it was expected that none of these models would be too difficult to analyze.

### 4.1.1 Effect identification using the Dantzig selector

As an initial attempt to identify the active factors in the models (4.1) - (4.6), the Dantzig selector was applied to the data given in Appendix A. The analysis was conducted as described in Section 3.1 with  $\delta = \sigma\sqrt{2\log(k)}$ . The estimated parameters were ordered according to decreasing absolute value. The six first effects found for the data generated with  $\sigma^2 = 0.1$  are shown in Table 4.2. The table displays the results for all the models in the panel, as indicated in the column to the left.

Model	Main effects						With two-factor interactions					
(4.1)	A	G	K	B	C	H	A	IJ	G	BD	DF	BJ
	0.91	-0.20	-0.05	0.02	0.01	0.00	0.82	-0.12	-0.12	-0.06	0.01	0.00
(4.2)	A	B	J	F	E	K	A	B	DF	CD	IJ	FH
	1.99	0.75	-0.06	0.03	-0.01	0.00	1.90	0.66	0.09	-0.06	-0.05	-0.01
(4.3)	A	B	I	J	D	G	A	B	GK	IK	AC	HJ
	0.96	0.82	0.04	0.03	0.02	0.00	0.89	0.76	-0.10	-0.02	-0.01	-0.01
(4.4)	A	B	F	K	D	J	A	B	AB	FI	DI	EG
	3.04	1.90	-0.36	0.36	-0.35	0.32	2.78	1.65	0.71	-0.21	-0.16	0.11
(4.5)	B	A	K	D	E	G	B	A	HJ	CJ	CH	EK
	2.00	1.93	0.47	-0.37	-0.37	-0.34	1.02	0.95	-0.72	-0.68	-0.63	-0.28
(4.6)	A	B	F	H	C	E	A	B	AB	EG	GK	HJ
	1.99	1.04	-0.47	-0.34	-0.34	-0.28	1.63	0.67	0.57	0.31	-0.23	-0.15

Table 4.2: The effects with the largest parameters in absolute value, estimated by the Dantzig selector for the data generated with  $\sigma^2 = 0.1$  from the models indicated in the column to the left. The results were found using  $\delta = 0.69$  when considering main effects, and  $\delta = 0.92$  when also including two-factor interactions.

From the results in the center of the table it is seen that the Dantzig selector performed well for all of the models when only considering main effects. When also including two-factor interactions in the design matrix, the method succeeded to identify the correct effects for all of the models, except for Model (4.5). Even though the estimation is not perfectly accurate, one gets an indication of the form of the model. As the goal of the analysis is only to study the performance of the methods for variable selection, the estimated values were only included as a guide for the number of terms to include in the model. The procedures discussed in Section 2.2 can also be used in order to remove the inert effects from the ones suggested in Table 4.2.

To test the method further, the Dantzig selector was applied to the data generated with  $\sigma^2 = 0.5$  given in Appendix A. The estimated parameter values obtained are given in Table 4.3.

Model	Main effects						With two-factor interactions					
(4.1)	A	H	E	K	J	G	A	FK	EI	DE	AJ	BH
	0.73	-0.13	-0.05	-0.05	0.04	0.00	0.57	0.15	0.09	-0.06	0.05	-0.03
(4.2)	A	B	E	K	H	G	A	B	IJ	CH	BK	DF
	1.90	0.78	-0.20	-0.17	0.14	-0.06	1.62	0.54	-0.23	0.21	0.14	0.09
(4.3)	B	A	D	H	K	C	B	A	HJ	FG	CD	FH
	0.88	0.78	0.29	0.24	0.12	-0.04	0.43	0.38	-0.26	-0.22	-0.20	-0.18
(4.4)	A	B	C	F	J	H	A	B	CH	AB	HJ	DI
	2.99	1.65	-0.56	-0.48	0.45	-0.35	2.43	1.09	0.49	0.41	-0.40	-0.28
(4.5)	B	A	H	E	F	J	B	A	GK	CH	EG	CJ
	1.81	1.64	-0.46	-0.43	-0.40	0.33	1.12	0.95	-0.75	0.51	0.33	-0.27
(4.6)	A	B	D	E	F	J	A	B	AB	CH	DI	CJ
	1.78	1.11	-0.54	-0.34	-0.27	0.25	1.28	0.61	0.42	0.36	-0.30	-0.25

Table 4.3: The results produced by the Dantzig selector for the data generated from the models (4.1) - (4.6) with  $\sigma^2 = 0.5$ . The estimated values were obtained by using  $\delta = 1.55$  when only considering main effects, and by using  $\delta = 2.05$  when also including two-factor interactions.

From Table 4.3 it is seen that the method provided fairly good results for the data generated with  $\sigma^2 = 0.5$ . However, by comparing Table 4.3 to the results in Table 4.2 it is observed that the performance of the Dantzig selector was reduced as the level of noise was increased. For all of the models, except for Model (4.5), the correct terms are on the list, but the estimation of parameters are less accurate. For Model (4.4) the term CH was ranked before AB, but the correct model was found when backward selection was applied to a model fitted with the six terms presented in Table 4.3. For Model (4.5) the interaction term AB was not ranked as one of the six most likely terms, which means that this interaction effect was lost by the Dantzig selector.

In order to get a better understanding of the performance of the Dantzig selector, the experiments were conducted ten times for each model and for each level of noise. Table 4.4 displays the fraction of times the correct effects were found by the

Dantzig selector. The results were counted as success given that all the correct terms were on the list of the six effects with the largest estimated parameter values, when also considering two-factor interactions. The thought was that backward elimination or other similar methods in many cases would be able to select the active effects as long as they were on the list, even though the Dantzig selector had estimated a higher parameter value for some of the inert effects.

Model	$\sigma^2 = 0.1$	$\sigma^2 = 0.5$
(4.1)	10/10	10/10
(4.2)	10/10	10/10
(4.3)	10/10	10/10
(4.4)	9/10	5/10
(4.5)	9/10	3/10
(4.6)	9/10	5/10

Table 4.4: The fraction of times the Dantzig selector gave all the correct effects estimated parameter values corresponding to one of the six most likely active effects.

Based on the fraction of successes displayed in Table 4.4 it seems like the Dantzig selector is highly reliable for models with only main effects. This finding is consistent with the discussion in Candès and Tao (2007), as the design used in this case meets the requirements of the UUP. It is seen that the performance was rather good for all of these models as long as the variance of the noise was low. It is observed that ability of the Dantzig selector to select active effects was significantly reduced for the models (4.4) - (4.6) when the level of noise was increased. It should be noted that the results might have been improved given that Monte Carlo simulations had been used to find the optimal  $\delta$ .

### 4.1.2 Variable selection using a graphical Dantzig selector

As a second approach for identifying the active effects, the graphical Dantzig selector was conducted as described in Section 3.2. When this method is performed, no estimate for  $\beta$  is found, the method is only used to identify the active effects. The method results in a plot which is interpreted as described in Section 3.2. The results found for the data generated with  $\sigma^2 = 0.1$  are shown in Table 4.5. This table is structured in the same way as the table for the previous method, with the effects ranked according to how long each solution path lasted. This criterion was chosen in order to obtain a R code that could provide results without visually analyzing the plots. Hence, when performing the graphical Dantzig selector the plots should be considered as well before making any conclusions. The resulting plots for main effects are given in Figure C.1 in Appendix C. The plots obtained when also including two-factor interactions are given in Figure 4.1. For the purpose of comparison, these plots are given in Subsection 4.1.4, where they are presented in the same figure as the corresponding plots obtained by the Lasso.

Model	Main effects						With two-factor interactions					
(4.1)	A	G	K	B	C	H	A	IJ	G	BD	DF	CH
(4.2)	A	B	J	F	E	G	A	B	DF	IJ	CD	FH
(4.3)	A	B	I	J	D	G	A	B	GK	IK	HJ	AC
(4.4)	A	B	D	F	K	J	A	B	FI	DI	EG	HJ
(4.5)	B	A	K	D	E	G	B	A	EK	DI	CJ	EG
(4.6)	A	B	F	C	H	E	A	EG	GK	HJ	DI	DF

Table 4.5: Ordering of effects found by the graphical Dantzig selector for the data generated with  $\sigma^2 = 0.1$ .

From Table 4.5 it is seen that the correct factors were identified for the first three models. The same observation was made when examining the plots in Figure 4.1. With the ordering of effects given in Table 4.5 the interaction term AB was not found for Model (4.4), and neither B nor AB was found for Model (4.6). It was difficult to add further information by considering the plots in Figure 4.1. Hence, the Dantzig selector performed better than the graphical Dantzig selector for the data generated from these two models. For Model (4.5) the interaction effect AB was not found. However, by considering the plot in Figure 4.1i the AB term may be considered to be of interest, as this effect has high estimated values for low values of  $\delta$ . It is observed that AB was not included in Table 4.5 for Model (4.5) as a consequence of the steep solution path obtained for this interaction effect.

It was of interest to find out whether the graphical Dantzig selector would be capable of identifying the active factors for the data generated with  $\sigma^2 = 0.5$ . The results are given in Table 4.6, and the corresponding plots are shown in Figure C.2.

Model	Main effects						With two-factor interactions					
(4.1)	A	H	E	K	J	G	A	FK	EI	BH	DE	AJ
(4.2)	A	B	E	K	H	G	A	B	IJ	CH	DF	BK
(4.3)	B	A	D	H	K	C	B	HJ	A	CD	FG	CE
(4.4)	A	B	C	F	J	H	A	B	EG	CH	HJ	DI
(4.5)	B	A	H	E	F	J	B	GK	A	CH	EG	CJ
(4.6)	A	B	D	E	F	J	A	DI	CH	EK	CJ	B

Table 4.6: Results found by the graphical Dantzig selector for the data with  $\sigma^2 = 0.5$ .

When considering the results for main effects only, the method correctly indicated that A is an active factor in Model (4.1), and that A and B are important main effects for the rest of the models. When also considering two-factor interactions, it was found that all of the active factors were identified for the first three models. For the models in (4.4) - (4.6) on the other hand, it was observed that the performance of the method was not satisfactory, as AB was lost from the analysis. The plots in Figure C.2 offered limited information of value, beyond the ordering of effects that were already given in Table 4.6. The methods presented in Section 2.2 can be used to determine how many effects to include in the model. For example, backward elimination with  $\alpha = 0.05$  could be applied to fitted models with the six terms given

in Table 4.6. When this was done for the data generated from Model (4.3) with  $\sigma^2 = 0.5$  it became clear that A and B were the only active factors in the model, even though HJ was ranked before factor A when considering two-factor interactions.

In order to get a better understanding of the performance of the graphical Dantzig selector, ten different data sets were generated from the models in (4.1) - (4.6) for each level of noise. The fraction of times all the correct effects were ranked as one of the six most likely active effects, are presented in Table 4.7.

Model	$\sigma^2 = 0.1$	$\sigma^2 = 0.5$
(4.1)	10/10	10/10
(4.2)	10/10	10/10
(4.3)	10/10	10/10
(4.4)	0/10	0/10
(4.5)	0/10	0/10
(4.6)	0/10	0/10

Table 4.7: The fraction of times all the correct terms were ranked as one of the six most likely active effects, determined by the graphical Dantzig selector.

It is seen that the graphical Dantzig selector was highly reliable for the models with main effects only, but that none of the models were fully identified for the last three models. The ordering of effects were only based on the longest lasting solution paths. The fraction of successes in Table 4.7 may have been improved given that the plots had been considered as well. The reduction in performance compared to the Dantzig selector for the models (4.4) - (4.6), may be explained by discussing the solution paths found by the graphical method. When only considering main effects, the solution paths plotted by the graphical Dantzig selector are parallel. Hence, the very same ordering of effects are found by the two methods. However, when also considering two-factor interactions the solution paths may cross each other. Hence, an active factor may have a large estimated parameter value for the optimal  $\delta$ , but a steep slope of the graph could cause the graphical Dantzig selector to disregard this factor in favor of inert factors.

### 4.1.3 Selection of effects using a projection based Dantzig selector

In the search for an improved procedure using the Dantzig selector, the method in Section 3.3 was tested. The columns of four and four factors together with their interactions were run through the code made for the method in Section 3.2. Models were fitted with the six effects with the largest estimated parameter values. The resulting models were ordered according to the values of the residual variances, from lowest to highest value. As six effects were included for each model, there were some effects in addition to the active ones. Hence, for the models in this example there were several models that contained all the correct effects. The one of these providing the lowest residual variance was of interest. The results obtained from ten individual data sets generated with  $\sigma^2 = 0.1$  and  $\sigma^2 = 0.5$  are displayed in Table 4.8. The



numbers in the table indicate the ranking of the fitted model that contained all the correct terms in the model. References to the models that are investigated are given in the column to the left. The first number given for each model corresponds to the results obtained for the data sets in Appendix A.

Model	$\sigma^2 = 0.1$					$\sigma^2 = 0.5$				
(4.1)	1	1	1	1	1	1	1	1	1	1
	1	1	1	1	1	1	1	1	1	1
(4.2)	1	1	1	1	1	2	1	1	1	1
	1	1	1	1	1	1	1	1	1	2
(4.3)	1	1	1	1	1	1	1	1	1	1
	1	1	1	1	1	1	2	2	1	1
(4.4)	1	1	1	1	1	1	1	1	1	1
	1	1	1	1	1	1	1	1	1	1
(4.5)	1	1	1	1	1	1	1	1	1	1
	1	1	1	1	1	1	1	1	1	1
(4.6)	1	1	1	1	1	2	1	1	1	1
	1	1	1	1	1	1	1	3	1	2

Table 4.8: The ranking of the model containing all the correct terms, in the list where fitted models were ranked according to the value of the residual variances.

This method performed relatively good for all of the models, even for the data sets generated with the high level of noise. Compared to the two previously studied methods, the projection based Dantzig selector seemed slightly more reliable for the models (4.4) - (4.6). In Section 3.3 a discussion was given on how to further advance on this method. This expanded method was studied, and the results are given below.

### Residual variance, AIC and BIC of reduced models

With this expanded method backward elimination with F-test and  $\alpha = 0.05$  was applied to the ten first models, ranked by the projection based Dantzig selector. The final model was selected based on either the residual variance, the AIC or the BIC of the reduced models. These three criteria were tested on the data in Appendix A, and the results are given in Table 4.9.

Model	Variance/ AIC/ BIC
(4.1)	A,G,K,FK,AF
(4.2)	A,B,J,AD
(4.3)	A,B,GK,AK,BK
(4.4)	A,B,AB,JK
(4.5)	B,A,AB,BG,AG,AI
(4.6)	A,B,AB,F,AJ,AF

Table 4.9: The effects found by the expanded projection based Dantzig selector as discussed in Subsection 3.3.1, when applied to the data generated with  $\sigma^2 = 0.1$ .

As indicated by Table 4.9, the very same effects were selected when using the criteria of residual variance, the AIC and the BIC. It is seen that the correct terms were found for all of them, but that some inert effects were added as well. It was tested whether this over-fitting could be detected by performing forward selection or a stepwise procedure based on p-values. The results in Table 4.9 were not improved by either of these methods.

The same procedure was performed on the data generated with  $\sigma^2 = 0.5$ . The results obtained by this expanded procedure are shown in Table 4.10.

Model	Variance/ AIC/ BIC
(4.1)	A,FK,GK,K,AF
(4.2)	A,B,AD,DG
(4.3)	B,A,BI,BJ,J
(4.4)	A,B,AB,CH,BC,C
(4.5)	B,A,AB,AJ,C
(4.6)	A,HJ,D,DH,AJ,H

Table 4.10: The effects determined by the method in Subsection 3.3.1 when applied to the data generated with  $\sigma^2 = 0.5$ .

Also for the data generated with  $\sigma^2 = 0.5$ , the very same effects were identified when the final model was selected based on the variance, the AIC and the BIC. From Table 4.10 it is seen that the method was able to identify the correct terms for the first five models, but that some additional effects were included. For Model (4.6) both B and AB were lost, and some inert effects were added. By using a stepwise procedure to prevent over-fitting, the results were not changed. When forward selection was performed the results were not improved for the first three models, but some inert factors were removed for the last three. However, all of these models were still considerably over-fitted. Hence, it was concluded that other types of criteria should be considered in order to remove the inert effects from the models.

#### 4.1.4 Identifying active effects using the Lasso

The Lasso was applied to the data in Appendix A, and the results were compared to the ones obtained by the Dantzig selector. The results from the Lasso were obtained by using a built-in function in R, where the optimal value for variable selection was chosen by cross-validation. The six effects with the largest estimated parameters in absolute value for the data generated with  $\sigma^2 = 0.1$ , are given in Table 4.11. The table is structured in the same way as for the results of the Dantzig selector, with effects ordered according to decreasing absolute value of the estimated parameters. The plots obtained when only considering main effects are given in Figure C.3 in Appendix C. The resulting plots found by the Lasso when also considering two-factor interactions are given in Figure 4.1, together with the corresponding plots obtained by the Dantzig selector.

Model	Main effects						With two-factor interactions					
(4.1)	A	G	K	B	C	H	A	G	IJ	BD	DF	CH
	0.95	-0.24	-0.10	0.06	0.05	0.04	0.86	-0.15	-0.14	-0.09	0.04	0.02
(4.2)	A	B	J	F	E	K	A	B	AG	AK	CD	DF
	2.03	0.79	-0.10	0.07	-0.06	-0.03	1.98	0.70	-0.07	0.06	-0.05	0.05
(4.3)	A	B	I	J	D	G	A	B	GK	AK	AC	IK
	1.01	0.88	0.09	0.08	0.08	0.04	0.96	0.82	-0.11	-0.05	-0.04	-0.03
(4.4)	A	B	F	K	D	J	A	B	FI	AB	DI	EG
	2.09	1.78	-0.23	0.23	-0.22	0.20	2.24	1.11	-0.21	0.17	-0.16	0.11
(4.5)	B	A	K	D	E	G	B	A	EK	EG	GK	DI
	1.86	1.79	0.33	-0.24	-0.24	-0.20	0.96	0.90	-0.48	0.38	-0.26	-0.19
(4.6)	A	B	F	H	C	E	A	B	AB	EG	GK	HJ
	1.89	0.93	-0.37	-0.24	-0.24	-0.18	1.77	0.80	0.71	0.26	-0.18	-0.13

Table 4.11: The effects with the largest parameters in absolute values, estimated by applying the Lasso to the data generated from the models (4.1) - (4.6) with  $\sigma^2 = 0.1$ .

From Table 4.11 it is seen that all of the active effects were identified by the Lasso, except for Model (4.5). By comparison, it was found that the results produced are rather similar to the results produced by the Dantzig selector. When only considering main effects the ordering of effects are the same for both methods, but there are some small differences in the results found when also considering two-factor interactions. Similarities are also observed when comparing the plots in Figure 4.1.

The results produced by the Lasso for the data generated with  $\sigma^2 = 0.5$  are presented in Table 4.12. The corresponding plots are displayed in Figure C.4 in Appendix C.

Model	Main effects						With two-factor interactions					
(4.1)	A	H	E	K	J	G	A	FK	EI	DE	AJ	GK
	0.77	-0.18	-0.10	-0.10	0.09	0.04	0.61	0.20	0.16	-0.11	0.09	-0.08
(4.2)	A	B	E	K	H	G	A	B	IJ	CH	DF	BK
	1.95	0.83	-0.25	-0.22	0.18	-0.10	1.56	0.47	-0.21	0.20	0.08	0.07
(4.3)	B	A	D	H	K	C	B	A	HJ	CD	FG	FH
	1.00	0.91	0.42	0.37	0.24	-0.17	0.40	0.32	-0.23	-0.18	-0.15	-0.12
(4.4)	A	B	C	F	J	H	A	B	CH	HJ	EG	DI
	2.91	1.57	-0.48	-0.41	0.37	-0.28	2.09	0.75	0.44	-0.40	0.27	-0.18
(4.5)	B	A	H	E	F	J	B	A	GK	CH	EG	CJ
	1.91	1.73	-0.55	-0.53	-0.50	0.43	1.25	1.08	-0.75	0.51	0.33	-0.27
(4.6)	A	B	D	E	F	J	A	B	AB	DI	CH	D
	1.73	1.06	-0.49	-0.29	-0.22	0.20	1.51	0.84	0.60	-0.30	0.24	-0.17

Table 4.12: The ordering of effects based on the absolute value of their estimated parameters, found by the Lasso for the data generated with  $\sigma^2 = 0.5$ , from the models indicated in the column to the left.

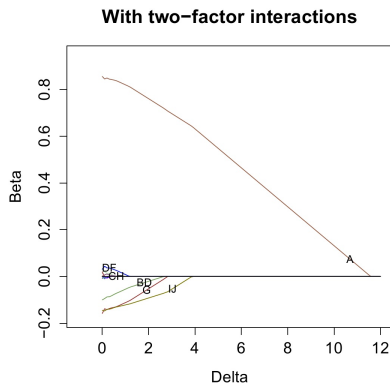
From Table 4.12 it is seen that the results found by the Lasso are quite similar to those found by the Dantzig selector. The Lasso was able to identify the main effects in all of the models, but the interaction effect AB was lost for Model (4.4) and Model (4.5). Both the estimated values and the plots produced by the Lasso can be considered in order to determine how many effects to include in the model. The methods mentioned in Section 2.2 can also be used for this purpose.

The Lasso was then applied to ten different data sets for each model and for each level of noise. The fraction of successes are noted in Table 4.13. The outcome was considered as a success given that all of the correct terms were ranked as one of the six effects with the highest estimated parameter values. It should be noted that these ten data sets were not the same as the ten data sets investigated by the Dantzig selector. Hence, the generated noise may have had an influence on the fraction of successes.

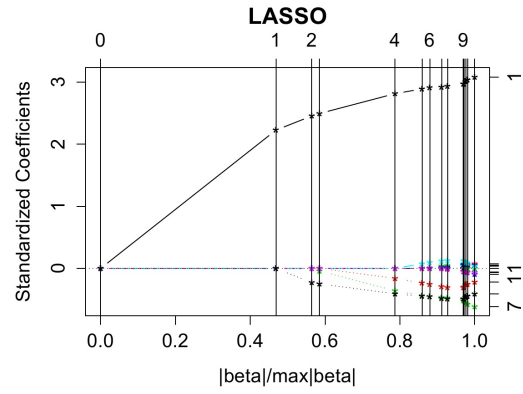
Model	$\sigma^2 = 0.1$	$\sigma^2 = 0.5$
(4.1)	10/10	10/10
(4.2)	10/10	10/10
(4.3)	10/10	9/10
(4.4)	8/10	2/10
(4.5)	7/10	4/10
(4.6)	8/10	3/10

Table 4.13: The fraction of times all the correct terms were ranked as one of the six most likely active effects by the Lasso.

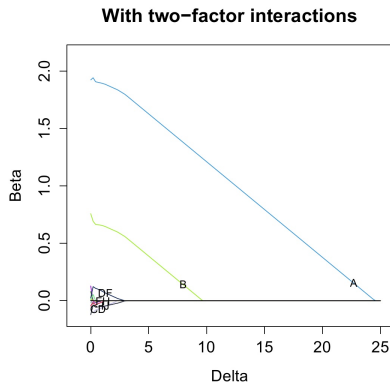
It is seen that the Lasso was quite reliable for the models consisting of only main effects. It is observed that the performance is reduced when interaction terms are added and when the level of noise is increased. For the ten data sets generated from these models, it was found that the Dantzig selector produced slightly better results, but the level of performance was rather close. The Lasso and the Dantzig selector should be applied to a greater amount of data sets in order to make any general conclusions.



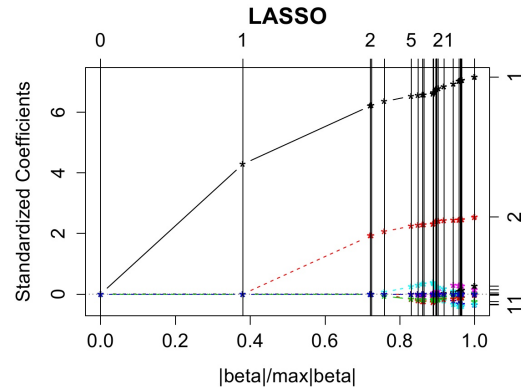
(a) Graphical Dantzig selector for Model (4.1) with  $\sigma^2 = 0.1$ , including two-factor interactions.



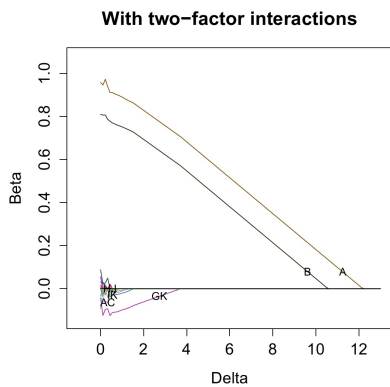
(b) Lasso for Model (4.1) with  $\sigma^2 = 0.1$ , including two-factor interactions.



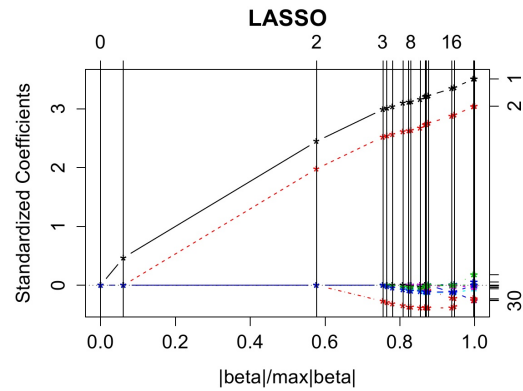
(c) Graphical Dantzig selector for Model (4.2) with  $\sigma^2 = 0.1$ , including two-factor interactions.



(d) Lasso for Model (4.2) with  $\sigma^2 = 0.1$ , including two-factor interactions.

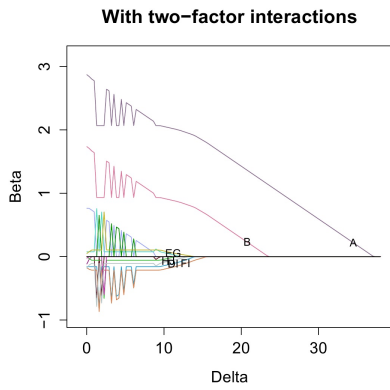


(e) Graphical Dantzig selector for Model (4.3) with  $\sigma^2 = 0.1$ , including two-factor interactions.

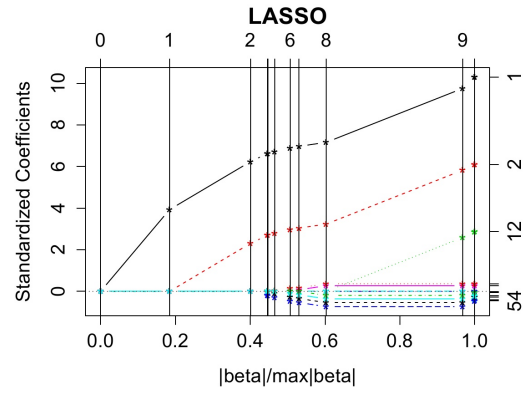


(f) Lasso for Model (4.3) with  $\sigma^2 = 0.1$ , including two-factor interactions.

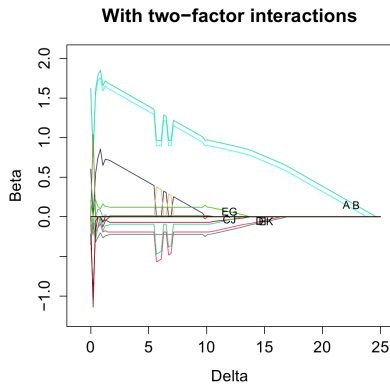
Figure 4.1: Plots obtained by the graphical Dantzig selector and the Lasso for the data generated from the models (4.1) - (4.6) with  $\sigma^2 = 0.1$ , where two-factor interactions had been included in the analysis.



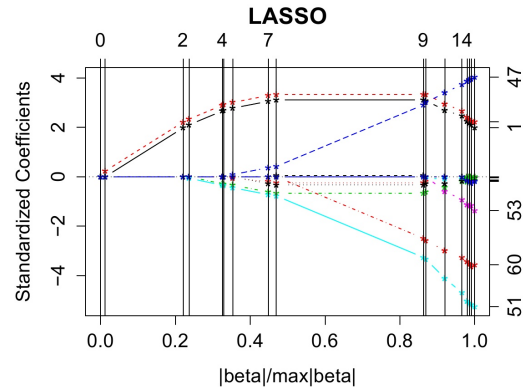
(g) Graphical Dantzig selector for Model (4.4) with  $\sigma^2 = 0.1$ , including two-factor interactions.



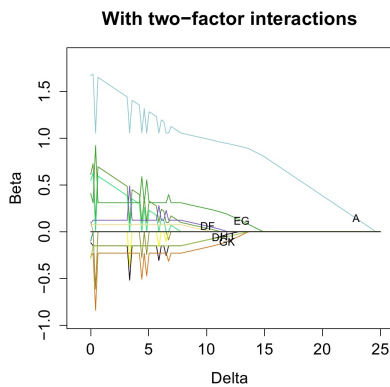
(h) Lasso for Model (4.4) with  $\sigma^2 = 0.1$ , including two-factor interactions.



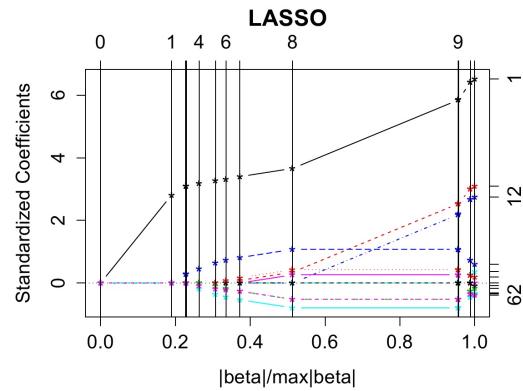
(i) Graphical Dantzig selector for Model (4.5) with  $\sigma^2 = 0.1$ , including two-factor interactions.



(j) Lasso for Model (4.5) with  $\sigma^2 = 0.1$ , including two-factor interactions.



(k) Graphical Dantzig selector for Model (4.6) with  $\sigma^2 = 0.1$ , including two-factor interactions.



(l) Lasso for Model (4.6) with  $\sigma^2 = 0.1$ , including two-factor interactions.

Figure 4.1: Plots obtained by the graphical Dantzig selector and the Lasso for the data generated from the models (4.1) - (4.6) with  $\sigma^2 = 0.1$ , where two-factor interactions had been included in the analysis.

### 4.1.5 Factor screening using a projection based method

The projection based method in Section 3.5 was applied to the panel of models in (4.1) - (4.6). The results obtained were ordered from low to high value of estimated variance,  $\hat{\sigma}^2$ . The four factors with the lowest values of  $\hat{\sigma}^2$  when considering one, two and three active factors were presented in a table. The estimated variances for the data sets in Appendix A generated with  $\sigma = 0.1$  are given in Table 4.14.

Model	One active factor				Two active factors				Three active factors			
(4.1)	A	G	B	K	AG	AC	AH	AK	AGH	AFG	AGK	ACH
	0.11	1.15	1.22	1.22	0.04	0.10	0.11	0.11	0.01	0.03	0.03	0.04
(4.2)	A	B	J	C	AB	AG	AK	AI	ABE	ABC	ABI	ABK
	0.82	5.06	5.82	5.83	0.05	0.82	0.83	0.88	0.01	0.02	0.04	0.04
(4.3)	A	B	D	I	AB	AC	AD	AF	ABE	ABK	ABI	ABJ
	0.96	1.27	2.19	2.19	0.04	0.98	1.01	1.04	0.02	0.02	0.03	0.03
(4.4)	A	B	F	K	AB	AD	AH	AE	ABI	ABE	ABH	ABC
	5.92	12.80	17.20	17.20	0.06	6.42	6.46	6.49	0.01	0.02	0.02	0.07
(4.5)	B	A	K	D	AB	BG	BK	BF	EGK	ABG	FGH	ABH
	6.25	6.58	10.98	11.08	0.12	6.45	6.62	6.85	0.03	0.04	0.04	0.06
(4.6)	A	B	F	C	AB	AJ	AF	AD	ABF	ABG	ABJ	ABK
	2.72	6.33	7.43	7.58	0.12	2.85	2.88	3.00	0.05	0.05	0.07	0.08

Table 4.14: The variance estimated by the projection based method for data generated with  $\sigma^2 = 0.1$ . The variances are given in the order of increasing value.

For Model (4.1) factor A has a considerably lower variance compared to the other main effects. For two active factors the estimated variances are all low and also rather similar to the variance found for factor A. This indicates that including two active factors would be an over-fitting for the model. Thus, the results in Table 4.14 indicate that A is the only active factor. For all of the models in (4.2) - (4.6) the factors A and B have the lowest values for  $\hat{\sigma}^2$  for one active factor. When considering two active factors, AB has considerably lower  $\hat{\sigma}^2$  than the rest. Also, AB has lower  $\hat{\sigma}^2$  than the variances estimated for one active factor. For three active factors the estimated variances are all low and also fairly similar, and they do not deviate much from the variance estimated for AB. These values indicate that three active factors would be an over-fitting. These results indicate that A and B are the most important factors in the models (4.2) - (4.6). It was seen that the method performed perfectly for the data generated with  $\sigma = 0.1$ , given that the number of active factors had been known. However, it is wise to perform additional analysis in order to confirm the number of active factors. This was done using the partial F-test, as presented in the next subsection.

It is of interest to study the ability of the projection based method to identify the active factors for the data generated with  $\sigma^2 = 0.5$ . When the method was applied to the data in Appendix A with the high level of noise, the results in Table 4.15 were found.

Model	One active factor				Two active factors				Three active factors			
(4.1)	A	H	E	K	AJ	AH	AE	AG	AEJ	AHJ	AJK	ABJ
	0.25	1.04	1.08	1.08	0.09	0.21	0.24	0.25	0.02	0.03	0.05	0.08
(4.2)	A	B	E	K	AB	AI	AK	AG	ABD	ABK	ABE	ABC
	1.46	5.41	6.28	6.30	0.57	1.33	1.33	1.50	0.21	0.28	0.39	0.40
(4.3)	B	A	D	H	AB	BK	AG	BD	ABK	CDE	CHJ	ABC
	1.51	1.73	2.51	2.56	0.58	1.23	1.38	1.49	0.06	0.24	0.25	0.34
(4.4)	A	B	C	F	AB	AJ	AH	AC	ABH	ABC	ABJ	ABE
	5.91	13.78	17.00	17.12	0.51	5.71	6.07	6.08	0.11	0.33	0.42	0.47
(4.5)	B	A	H	E	AB	BI	BH	BC	CDE	EGK	ABF	ABI
	5.28	6.07	9.40	9.43	0.62	5.12	5.54	5.66	0.30	0.31	0.33	0.36
(4.6)	A	B	D	E	AB	AF	AD	AE	ABC	ABD	ABH	ABF
	3.45	5.96	7.29	7.55	0.34	3.29	3.48	3.67	0.23	0.24	0.24	0.28

Table 4.15: The variance estimated by the projection based method for data generated from the models (4.1) - (4.6) with  $\sigma^2 = 0.5$ . The factors are given in the order of increasing value of the estimated variances.

From Table 4.15 it is seen that the estimated variance for A is significantly lower than for the rest of the main effects, when considering Model (4.1). For two active factors the estimated variance for AJ is lower than  $\hat{s}^2$  for A. For three active factors the estimated variances are rather similar and not much lower than for two active factor. For this model, it is not clear whether A is sufficient for the model, or if factor J should be included as well. When comparing  $\hat{s}^2$  for the models (4.2) - (4.6) it is seen that the estimated variances get smaller as the number of active factors is increased. For the models (4.2) - (4.6) the difference in  $\hat{s}^2$  for two and three active factors is of a size where one is not certain of how many factors that are appropriate for the model. As was expected beforehand, the method gave a more clear indication of the number of active factors for the data generated with  $\sigma^2 = 0.1$  compared to the data generated with  $\sigma^2 = 0.5$ .

As an additional analysis, ten models were generated for each model and for each level of noise. The fraction of times the correct combination of factors was given the lowest estimated variance is given in Table 4.16. Here it was assumed that the number of active factors was known in advance.

Model	$\sigma^2 = 0.1$	$\sigma^2 = 0.5$
(4.1)	10/10	10/10
(4.2)	10/10	10/10
(4.3)	10/10	10/10
(4.4)	10/10	10/10
(4.5)	10/10	10/10
(4.6)	10/10	10/10

Table 4.16: The fraction of times the active factors were ranked first in the list produced by the projection based method.



It is seen that the projection based method was able to identify all of the active factors for these models. The main challenge of this method was not to identify the active factors, but to determine how many to include. It is seen that this method is superior compared to the previously studied methods for the models in this example.

### Partial F-test

For some of the models the projection based method indicated how many factors to include, but for some models it was less evident. In order to verify the number of active factors the partial F-test was conducted. Models were fitted with the main effects and two-factor interactions as indicated by the projection based method. These models were tested against the models obtain by adding the factor given in Table 4.14 and Table 4.15. For example, from the estimated variances in Table 4.14 it is evident that at least two factors are present in Model (4.4). To verify that two factors are enough a model with A+B+AB was tested against A+B+E+AB+AE+BE. The factors found to be active by the partial F-test are given in Table 4.17.

Model	$\sigma^2 = 0.1$	$\sigma^2 = 0.5$
(4.1)	A,G	A,J
(4.2)	A,B	A,B
(4.3)	A,B	A,B,K
(4.4)	A,B	A,B
(4.5)	-	-
(4.6)	A,B	A,B

Table 4.17: The factors found to be active based on the results from the partial F-test, when evaluating the models suggested by the projection based method.

It is seen that an additional factor erroneously was added to Model (4.1), and also to Model (4.3) for  $\sigma^2 = 0.5$ . For the rest of the models the partial F-test correctly indicated that A and B are the active factors. For the data generated with  $\sigma^2 = 0.1$  from Model (4.5), the projection based method suggested the combination AB when considering two active factors and the factors EGK for three active factors. As AB is not included in EGK, the partial F-test is not suitable for this case. Hence, other types of argumentation had to be used. As  $\hat{\sigma}^2$  for three active factors are all low, relatively similar to each other and also quite similar to the estimated variance for AB, the confidence was high that two active factors were enough for this model. As an additional check the partial F-test was performed on ABG for  $\sigma^2 = 0.1$ , which was the second combination of factors found by the projection based method. The test indicated that the model with A and B provided a better fit for the data compared to the model with three active factors. By also comparing the results of the projection based method to the Dantzig selector it seemed reasonable to conclude that A and B were the active factors in the model. Similar argumentation can be used for Model (4.5) with  $\sigma^2 = 0.5$ . For the rest of the models the partial F-test proved to be a valuable tool to be use in cooperation with the projection based method, for cases where there are one or two active factors. For these models the results from the combination of the projection based method and the partial F-test were considered to be sufficiently evident.

## 4.2 Example 2

To test the methods further, data was generated from the models

$$y = k + A + 2AB + 2AC + \epsilon \quad (4.7)$$

$$y = k + A + AB + 2AC + \epsilon \quad (4.8)$$

$$y = k + A + 1.5B + 2C + AB + 1.5AC + \epsilon \quad (4.9)$$

$$y = k + A + 1.5B + 2C + 1.5ABC + \epsilon \quad (4.10)$$

$$y = k + 2A + BC + \epsilon \quad (4.11)$$

$$y = k + A + C + BC + CD + \epsilon \quad (4.12)$$

$$y = k + 2A + 3B + 2C + D + 3CD + \epsilon \quad (4.13)$$

$$y = k + 4A + B + C + D + 2AD + \epsilon \quad (4.14)$$

$$y = k + 2A + 4C + 2BC + 2CD + \epsilon \quad (4.15)$$

with  $k = 2$  and noise  $\epsilon \sim N(0, \sigma^2)$ . The data was generated with both  $\sigma^2 = 0.1$  and  $\sigma^2 = 0.5$  as shown in Appendix A. Note that the first five models have three active factors, while the remaining have four active factors. Factor sparsity is an important requirement for all the methods in Chapter 3. During the study of the methods it was of interest to find out whether four active factors are sparse enough for the 12 run PB design. Due to the number of active factors it was expected that the methods would perform better for the models (4.7)-(4.11) than for (4.12)-(4.15).

For Model (4.10) a three-factor interaction term is present. Adjustments were made to the functions in R, in order for them to take this three-factor interaction effect into consideration. It was expected that the methods using the Dantzig selector and the Lasso would have greater difficulties with identifying the active factors for this model because of the additional columns that have to be added to the design matrix. As the projection based method does not assume any particular form of the model, these results are not influenced by this three-factor interaction term. It was also anticipated that the method using contrast plots and orthogonalization would perform well for Model (4.10).

### 4.2.1 Effect identification using the Dantzig selector

The first method tested on the models in the panel above was the Dantzig selector, as described in Section 3.1. Similarly to the analysis conducted in Example 1, the estimated values were found using  $\delta = \sigma\sqrt{2\log(k)}$ . The results were ordered according to decreasing absolute value of the estimated parameters. The first six effects were presented in a table. The results found for the data in Appendix A generated with  $\sigma^2 = 0.1$ , are displayed in Table 4.18. The results found when also considering three-factor interactions are not included in this table. For Model (4.10) the result obtained are discussed in the text instead.

Model	Main effects						Two-factor interactions					
(4.7)	E	F	D	K	A	C	AC	AB	DI	GJ	HI	DJ
	-1.42	-1.33	-1.26	1.21	1.00	-0.70	1.11	0.87	-0.69	0.55	0.47	0.38
(4.8)	F	K	D	E	A	B	AC	EH	DI	GJ	EG	DJ
	-0.99	0.96	-0.95	-0.93	0.92	-0.55	1.03	-0.53	-0.50	0.44	0.42	0.41
(4.9)	C	B	A	D	F	K	CJ	JK	C	CI	CH	DJ
	1.60	0.93	0.91	-0.85	-0.73	0.71	-1.09	-1.01	0.98	0.82	0.61	0.53
(4.10)	C	B	A	J	G	E	C	BD	AD	CD	CE	AE
	2.00	1.42	1.06	-0.59	-0.56	0.49	0.96	-0.75	-0.72	-0.71	-0.69	-0.63
(4.11)	A	K	E	I	D	F	A	DF	DK	EF	EK	GJ
	1.91	-0.45	-0.34	0.33	-0.31	0.28	1.25	0.41	-0.38	0.38	-0.27	0.21
(4.12)	C	I	K	E	B	D	C	BI	I	AD	AF	B
	1.09	0.66	-0.59	-0.55	-0.35	-0.26	0.85	0.69	0.63	0.50	-0.25	-0.16
(4.13)	B	C	F	K	J	D	GK	HI	AF	DE	CE	C
	2.08	2.02	-1.13	-1.04	1.03	1.03	-2.60	-0.69	-0.54	-0.53	-0.52	0.42
(4.14)	A	D	H	E	F	K	A	IJ	BF	FG	AD	EJ
	4.03	0.86	0.70	-0.69	0.68	-0.62	3.35	-1.25	0.48	-0.47	0.36	-0.36
(4.15)	C	E	I	K	B	D	C	AF	DK	AH	BJ	FG
	3.98	-1.35	1.30	-1.26	-0.75	-0.73	2.16	-1.46	-1.33	1.18	0.96	0.65

Table 4.18: The effects with the largest parameters in absolute value, estimated by the Dantzig selector for data generated from the models (4.7) - (4.15) using  $\sigma^2 = 0.1$ . The estimates were obtained using  $\delta = 0.69$  when considering only main effects, and  $\delta = 0.92$  when expanding the design matrix to also include two-factor interactions.

From Table 4.18 it is observed that the Dantzig selector was not able to identify all of the active effects for any of the models. On the other hand, it is seen that at least one of the correct terms is on the list of the six effects with the largest estimated parameter values. For Model (4.7) the correct interaction terms were ranked first when including two-factor interactions, but it is seen that A is lost by this method. For the second model the interaction AC was found while A and AB were lost by the Dantzig selector. Similar observations are made for the rest of the models, where some effects are found and some are lost by the method. For the models (4.10) - (4.12) and (4.14) - (4.15) the same effect is ranked first in the results for main effects and in the results where two-factor interactions are included as well. This observation increases the confidence in the given main effect. Due to the ABC term in Model (4.10) the code was run with a design matrix with columns corresponding to all the main effect, two-factor and three-factor interactions. From the results produced it was found that only factor C was included in the six first terms. From the results given to the right in Table 4.18 it is noticed that the estimated values found for the active effects are far from accurate for all of the models.

The six largest parameters in absolute value estimated by the Dantzig selector for the data generated with  $\sigma^2 = 0.5$  are given in Table 4.19.

Model	Main effects						Two-factor interactions					
(4.7)	D	F	K	A	E	B	AC	AB	EH	EG	DJ	GJ
	-1.18	-1.10	1.07	0.71	-0.54	-0.37	0.79	0.68	-0.68	0.57	0.23	0.18
(4.8)	F	A	K	D	E	I	AC	EG	DI	EH	DJ	GJ
	-0.95	0.83	0.82	-0.82	-0.56	-0.34	0.82	0.58	-0.48	-0.37	0.29	0.27
(4.9)	C	A	D	F	K	E	FJ	EH	C	BI	A	FH
	1.59	1.11	-1.06	-0.85	0.75	-0.73	1.09	-1.08	0.88	0.84	0.41	-0.32
(4.10)	C	B	A	F	I	E	CE	FI	C	AE	FK	GI
	1.84	1.22	0.95	-0.65	0.49	0.48	-1.36	-1.17	0.82	-0.62	-0.51	-0.30
(4.11)	A	J	K	H	B	F	A	EF	DF	CE	EK	IJ
	1.52	-0.61	-0.52	0.40	0.22	0.17	0.79	0.48	0.40	-0.31	-0.31	-0.20
(4.12)	K	C	B	I	F	E	BI	BD	CD	K	IK	B
	-0.89	0.68	-0.54	0.40	-0.37	-0.28	0.60	-0.54	0.35	-0.31	0.24	-0.21
(4.13)	C	B	K	I	G	H	GK	HI	AF	BI	DE	GH
	1.55	1.50	-1.20	0.98	0.97	-0.80	-1.99	-0.85	-0.59	0.55	-0.43	-0.31
(4.14)	A	J	D	F	I	E	A	IJ	FG	J	GH	HK
	4.23	0.96	0.91	0.57	-0.54	-0.54	3.39	-1.67	-0.48	0.42	0.36	-0.24
(4.15)	C	I	K	E	A	B	AH	C	AF	CE	GH	DK
	3.64	1.47	-1.46	-1.02	0.72	-0.51	1.78	1.74	-1.73	0.64	-0.60	-0.57

Table 4.19: The largest parameters in absolute value, estimated by applying the Dantzig selector to the data generated with  $\sigma^2 = 0.5$ , from the models indicated in the column to the left. The estimated values were found by using  $\delta = 1.55$  for the value in the center of the table, and  $\delta = 2.05$  for the ones to the right.

It is seen that the performance of the Dantzig selector was reduced when the level of noise was increased. The estimation of parameters was less accurate and for Model (4.13) non of the terms in the model were ranked as one of the six most likely effects. For Model (4.10) non of the active factors were identified by the Dantzig selector when including three-factor interactions in the analysis.

In Example 1 there was given a table containing the fraction of times all of the correct effects were ranked as one of the effects with the six largest parameter values. Due to the relatively low level of performance of the Dantzig selector for the models (4.7) - (4.15), it was determine not to include such a table in this example. For these models additional analysis should be performed in order to determine the active effects.

## 4.2.2 Variable selection using a graphical Dantzig selector

As a second attempt to find the active effects in the models (4.7) - (4.15), the graphical method in Section 3.2 was conducted. The six effects with the longest lasting solution paths for the data with  $\sigma^2 = 0.1$  are shown in Table 4.20. The plots found when considering main effects are given in Figure C.5 in Appendix C. The plots obtained when also including two-factor interactions are found in the same figure as the corresponding plots from the Lasso in Figure 4.2.

Model	Main effects						Two-factor interactions					
(4.7)	E	F	D	K	A	C	AC	AB	DI	GJ	HI	EG
(4.8)	F	K	F	E	A	B	AC	EH	DI	GJ	AB	EG
(4.9)	C	B	A	D	F	E	C	JK	FH	BI	FI	BG
(4.10)	C	B	A	J	G	E	C	FK	AD	CD	FI	B
(4.11)	A	K	E	I	D	F	A	GH	GJ	DK	IJ	DF
(4.12)	C	I	K	E	B	D	C	BI	AH	AF	DK	CD
(4.13)	B	C	F	K	D	J	GK	HI	AF	DE	CE	B
(4.14)	A	D	E	F	H	K	A	IJ	FG	AD	HK	EJ
(4.15)	C	E	I	K	B	D	C	AF	DK	AH	BI	IK

Table 4.20: The six longest lasting effects in the plot obtained by the graphical Dantzig selector for the data in Appendix A generated with  $\sigma^2 = 0.1$ .

From the results in Table 4.20 it is observed that the quality of the outcome produced by the graphical Dantzig selector was considerably lower in this example compared to Example 1. However, it is seen that the method found at least one of the correct terms for each model. None of the active effects were found for Model (4.10) when also considering three-factor interactions. The findings in Table 4.20 are supported by the plots in Figure 4.2, but limited additional information was offered by these plots. In some of the plots, such as the one in Figure 4.2e it is observed that non of the effects clearly stands out from the others in terms of a high value and a long lasting solution path. This might give an indication that the methods using the Dantzig selector are not appropriate for the model under consideration.

With  $\sigma^2 = 0.5$  the graphical Dantzig selector provided the results in Table 4.21. The corresponding plots are given in Figure C.6 in Appendix C.

Model	Main effects						Two-factor interactions					
(4.7)	D	F	K	A	E	B	AC	AB	EH	EG	GJ	DJ
(4.8)	F	A	D	K	E	I	AC	DI	EG	EH	DJ	GJ
(4.9)	C	A	D	F	K	E	EH	C	BI	BG	DJ	JK
(4.10)	C	B	A	F	E	H	C	FK	AE	BD	AD	GI
(4.11)	A	J	K	H	B	E	A	GH	IJ	DF	EF	CE
(4.12)	K	C	B	I	F	E	AH	BI	K	BD	CD	GJ
(4.13)	C	B	K	G	I	H	GK	AF	HI	DE	BI	CE
(4.14)	A	J	D	F	I	B	A	IJ	FG	HK	J	BG
(4.15)	C	I	K	E	A	B	C	AF	AH	DK	BI	FJ

Table 4.21: The results found by the graphical Dantzig selector for data generated from the models (4.7) - (4.15) with  $\sigma^2 = 0.5$ .

Except for Model (4.13) the graphical Dantzig selector was able to identify at least one of the active effects. Also in this case, non of the active effects were found for Model (4.10) when also considering three-factor interactions. It was hard to find more useful information regarding the ordering of effects from the plots in Figure C.6, other than what was already given in Table 4.21. However, these plots can be consulted when the number of active effects are to be determined.

### 4.2.3 Selection of effects using a projection based Dantzig selector

The projection based Dantzig selector was conducted as described in Section 3.3. Recall that the six effects with the largest parameters in absolute value were fitted to a linear model. The code was applied to ten different data sets for each model and the ranking of the model containing all the correct effects was noted. For all the models in this example there were several models that contained all the correct effects, but only the one with the lowest ranking was noted in the table. As there is a three-factor interaction term in Model (4.10) the R function was designed to account this type of models. For this model the Dantzig selector was applied to all subsets of four factors, together with their two-factor and three-factor interactions. The method was then proceeded the same way as for the other models. The results found for the data generated with  $\sigma^2 = 0.1$  and  $\sigma^2 = 0.5$  are shown in Table 4.22. The first number noted for each model corresponds to the results obtained for the data sets in Appendix A.

Model	$\sigma^2 = 0.1$					$\sigma^2 = 0.5$				
(4.7)	1	1	2	1	1	1	2	1	1	1
	1	1	1	1	1	1	1	1	2	1
(4.8)	3	1	1	1	1	2	1	1	1	2
	1	1	1	1	1	1	1	1	1	1
(4.9)	1	1	1	1	1	1	1	2	2	3
	1	1	1	1	1	11+	1	2	2	9
(4.10)	3	4	9	5	6	1	9	11+	7	11+
	1	2	1	2	2	11+	11+	8	11+	10
(4.11)	1	2	1	1	1	1	4	3	1	1
	1	1	1	1	1	3	1	2	1	2
(4.12)	1	5	1	6	1	11+	11+	11+	9	11+
	7	5	1	1	2	11+	11+	4	4	4
(4.13)	1	1	2	2	3	11+	11+	11+	11+	11+
	1	11+	3	4	2	5	4	11+	1	8
(4.14)	1	1	11+	3	3	6	11+	1	11+	11+
	5	11+	11+	1	1	11+	11+	2	4	4
(4.15)	4	1	1	4	2	2	2	3	1	1
	2	1	1	1	3	1	2	2	2	1

Table 4.22: The ranking of the model comprising all the correct effects, where models have been ranked from lowest to highest residual variance.

From Table 4.22 it is seen that the method provides reasonably good results for the first three models in the panel. The performance of the methods was somewhat reduced for Model (4.10) which contains a three-factor interaction term. This was expected as the 12 run PB design only allows for estimation of main effects and two-factor interactions when projected onto four factors. For the models with four active factors the true model was in general ranked lower on the list compared to the ranking of the models with three active factors. It was expected that the method

would perform better on the data with the low level of noise compared to the data with the high level. From Table 4.22 it is seen that the prior beliefs are confirmed.

### Residual variance, AIC and BIC of reduced models

The expanded method that selects model based on the residual variance, AIC or BIC of the reduced models was also applied to the data in Appendix A. The results for the data with  $\sigma^2 = 0.1$  are displayed in Table 4.23.

Model	Variance/ AIC/ BIC
(4.7)	AB,AC,A,CE
(4.8)	EG,BG,E,B,CG,CE
(4.9)	C,AC,B,A,AB
(4.10)	AHI,C,A,ACH,AI
(4.11)	A,BC,CD
(4.12)	C,CD,BC,A
(4.13)	B,C,CD,BD,A,D
(4.14)	A,AD,D,B,C
(4.15)	C,GJ,K,CJ,JK

Table 4.23: Effects selected by using the residual variance, AIC and BIC from the models indicated in the column to the left with variance of the noise  $\sigma^2 = 0.1$ .

As indicated by Table 4.23 the exact same terms were selected regardless of the choice of criteria. For the models (4.7), (4.9) and (4.11) - (4.14) all the correct terms were identified, even though some additional terms were suggested for some of them. For (4.10) and (4.15) only parts of the models were found, and for Model (4.8) non of the active effects were identified. In the attempt to remove the additional inert factors in Table 4.23 both forward selection and a stepwise procedure were tested. The stepwise procedure removed the inert effect added to Model (4.13). For the rest of the models the results were not improved. When forward selection was performed both some inert and active effects were removed from some of the models. Hence, forward selection is not preferred in cooperation with this method.

Model	Variance/ AIC/ BIC
(4.7)	AC,AB,A,E,C
(4.8)	EG,BG,F,BF,E,B
(4.9)	C,A,AC,AB,B
(4.10)	AHI,ABI,A,AH,B
(4.11)	A,BC,B,AB
(4.12)	AH,K,EH,AK
(4.13)	GK,K,G,EI,E
(4.14)	A,IJ,J,I,AJ,BI
(4.15)	C,I,K,E,CI,CK

Table 4.24: Effects selected using the residual variance, AIC and BIC of the reduced models from the data with variance of the noise  $\sigma^2 = 0.5$ .

The results for  $\sigma^2 = 0.5$  are shown in Table 4.24. From this table it is seen that the performance was reduced compared to the data sets generated with  $\sigma^2 = 0.1$ . All the terms were found for the models (4.7), (4.9) and (4.11), even though some additional terms were suggested as well. Non of the correct terms were identified for (4.8), (4.12) and (4.13). In the remaining three models only some of the active factors were found. It is also observed that the model with the correct effects for Model (4.10) was ranked first in Table 4.22. However, when the ten first models were reduced, the model with the effects AHI, ABI, A, AH and B was chosen in favor of the model containing the active ones.

#### 4.2.4 Identifying active effects using the Lasso

The Lasso was applied to the data generated from the models (4.7) - (4.15). The results found for the data with  $\sigma^2 = 0.1$  are found in Table 4.25. The corresponding plots for main effects are shown in Figure C.7 in Appendix C, and the plots found when also including two-factor interactions are given in Figure 4.2.

Model	Main effects						With two-factor interactions					
(4.7)	E	F	D	K	A	C	AB	AC	DI	A	GJ	HI
	-1.46	-1.37	-1.30	1.26	1.04	-0.74	1.31	1.10	-0.27	0.23	0.23	0.06
(4.8)	F	K	D	E	A	B	AC	AB	A	EH	DI	DJ
	-0.74	0.72	-0.71	-0.68	0.68	-0.31	1.55	0.53	0.52	-0.13	-0.09	0.04
(4.9)	C	B	A	D	F	K	FH	C	FI	CH	BI	JK
	1.22	0.56	0.53	-0.48	-0.35	0.33	-0.81	0.53	-0.39	0.31	0.27	-0.23
(4.10)	C	B	A	J	G	E	C	FK	B	A	GI	FI
	1.72	1.14	0.78	-0.31	-0.28	0.21	0.93	-0.92	0.65	0.63	-0.44	-0.35
(4.11)	A	K	E	I	D	F	A	BC	GJ	GH	DF	IJ
	1.78	-0.32	-0.21	0.20	-0.19	0.15	1.61	0.37	0.19	-0.15	0.13	-0.12
(4.12)	C	I	K	E	B	D	AH	AF	C	BI	DK	CD
	1.14	0.72	-0.64	-0.61	-0.41	-0.31	0.42	-0.39	0.38	0.21	-0.18	0.08
(4.13)	B	C	F	K	J	D	GK	HI	AF	DE	CE	C
	1.52	1.46	-0.57	-0.48	0.47	0.47	-2.50	-0.63	-0.48	-0.40	-0.32	0.13
(4.14)	A	D	H	E	F	K	A	IJ	AD	CK	FG	HK
	3.73	0.56	0.41	-0.40	0.38	-0.33	3.55	-1.15	0.59	-0.34	-0.21	-0.19
(4.15)	C	E	I	K	B	D	C	AF	AH	DK	BI	IK
	4.01	-1.38	1.33	-1.29	-0.78	-0.76	2.42	-0.94	0.89	-0.55	0.51	0.17

Table 4.25: The effects with the largest parameters in absolute value, estimated by applying the Lasso to the data generated with  $\sigma^2 = 0.1$ .

From Table 4.25 it is seen that the ability of the Lasso to identify the active factors are approximately on the same level as for the Dantzig selector. However, for some of the models the Lasso performed slightly better. For Model (4.7) and (4.8) it is seen that all of the active effects are ranked as one of the six effects with the largest estimated parameter values. Further, it is observed that at least one active factor is identified for the rest of the models. When the Lasso was applied to Model (4.10)



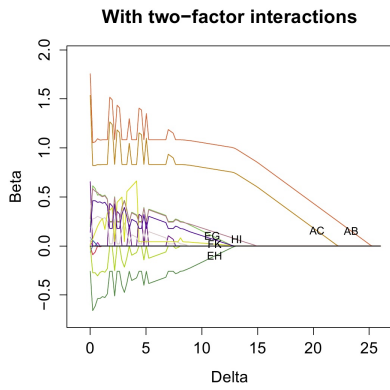
with columns corresponding to the main effects, two-factor and three-factor interactions included in the design matrix, only the main effect C was found. The same result was found by the Dantzig selector. By considering Figure 4.2 it is seen that there are great similarities in the shape of the plots obtained by the Lasso and the graphical Dantzig selector.

The results found for the data in Appendix A generated with  $\sigma^2 = 0.5$  are given in Table 4.26, and the corresponding plots are given in Figure C.8 in Appendix C.

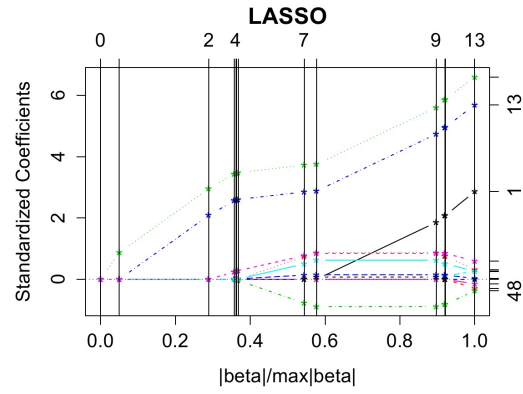
Model	Main effects						With two-factor interactions					
(4.7)	D	F	K	A	E	B	AC	AB	EH	EG	GJ	DJ
	-1.23	-1.15	1.12	0.76	-0.59	-0.42	0.82	0.71	-0.57	0.46	0.04	0.02
(4.8)	F	A	K	D	E	I	AC	EG	DI	EH	GJ	DJ
	-0.76	0.65	0.64	-0.64	-0.38	-0.15	0.79	0.70	-0.49	-0.42	0.39	0.33
(4.9)	C	A	D	F	K	E	BI	EH	BG	DJ	JK	C
	1.62	1.14	-1.10	-0.89	0.79	-0.76	0.61	-0.55	-0.50	0.37	-0.36	0.25
(4.10)	C	B	A	F	I	E	B	GI	C	BD	GK	FK
	1.61	1.00	0.72	-0.43	0.26	0.26	0.98	-0.97	0.65	-0.57	-0.23	-0.21
(4.11)	A	J	K	H	B	F	A	EF	DF	CE	GH	EK
	1.46	-0.56	-0.46	0.34	0.16	0.11	0.78	0.44	0.40	-0.27	-0.21	-0.20
(4.12)	K	C	B	I	F	E	AH	AG	AF	BI	GJ	BD
	-0.72	0.51	-0.37	0.24	-0.20	-0.11	0.54	0.39	-0.30	0.23	0.22	-0.12
(4.13)	C	B	K	I	G	H	GK	HI	AF	BI	DE	GH
	1.19	1.14	-0.84	0.63	0.62	-0.44	-1.99	-0.90	-0.57	0.56	-0.41	-0.32
(4.14)	A	J	D	F	E	I	A	IJ	FG	J	HK	BG
	3.99	0.72	0.67	0.33	-0.30	-0.29	3.23	-1.48	-0.50	0.41	-0.23	-0.15
(4.15)	C	I	K	E	A	B	C	AF	AH	DK	BI	FJ
	3.75	1.59	-1.57	-1.14	0.83	-0.62	2.02	-1.09	1.06	-0.50	0.32	-0.22

Table 4.26: The six effects with the largest parameters in absolute values, estimated by the Lasso for the data generated with  $\sigma^2 = 0.5$ .

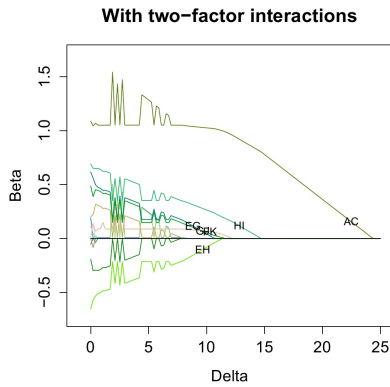
From Table 4.26 it is found that the performance of the Lasso was slightly reduced as the level of noise was increased. For Model (4.12) and Model (4.13) non of the correct effects got a large enough estimated parameter value to be ranked as one of the six most likely terms in the model. When also including three-factor interactions in the design matrix, non of the active factors were found for Model (4.10). By comparison of results, it is found that Dantzig selector performed slightly better for Model (4.9) and Model (4.12). Except for this, the results are rather similar.



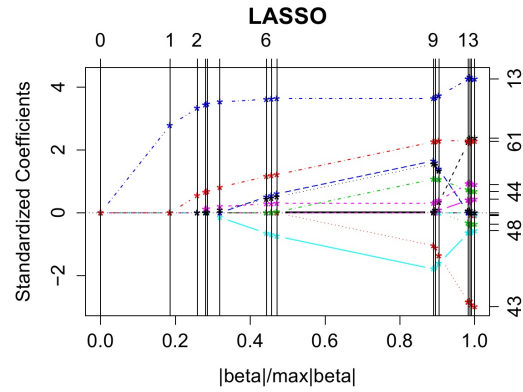
(a) Graphical Dantzig selector for Model (4.7) with  $\sigma^2 = 0.1$ , including two-factor interactions.



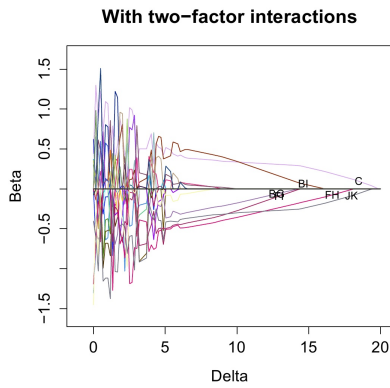
(b) Lasso for Model (4.7) with  $\sigma^2 = 0.1$ , including two-factor interactions.



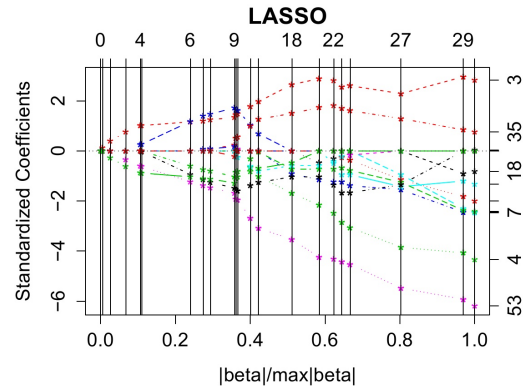
(c) Graphical Dantzig selector for Model (4.8) with  $\sigma^2 = 0.1$ , including two-factor interactions.



(d) Lasso for Model (4.8) with  $\sigma^2 = 0.1$ , including two-factor interactions.

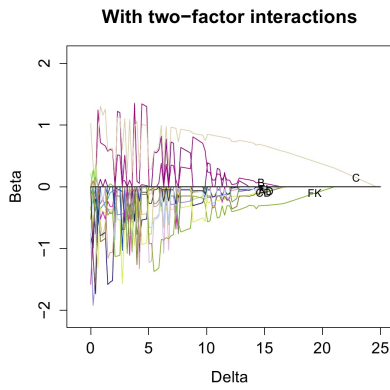


(e) Graphical Dantzig selector for Model (4.9) with  $\sigma^2 = 0.1$ , including two-factor interactions.

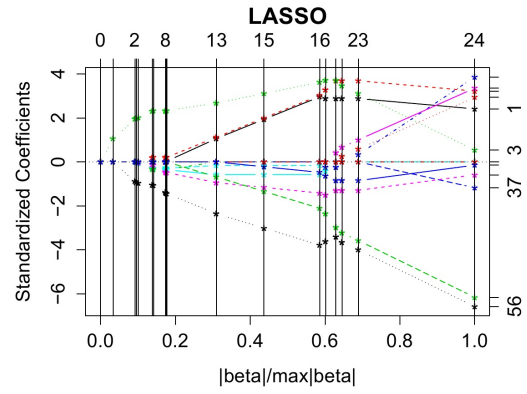


(f) Lasso for Model (4.9) with  $\sigma^2 = 0.1$ , including two-factor interactions.

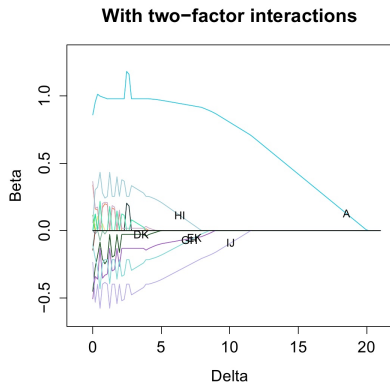
Figure 4.2: Plots obtained by the graphical Dantzig selector and the Lasso for the data generated from models (4.7) - (4.15) with  $\sigma^2 = 0.1$ . Two-factor interactions have been included in the analysis.



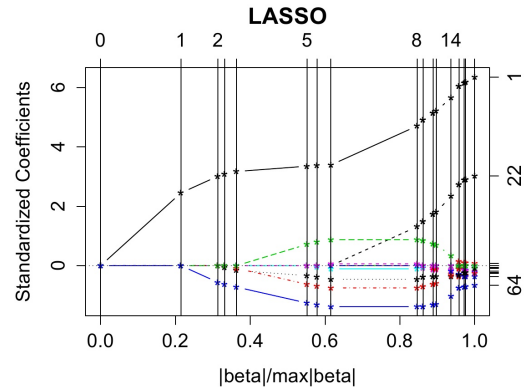
(g) Graphical Dantzig selector for Model (4.10) with  $\sigma^2 = 0.1$ , including two-factor interactions.



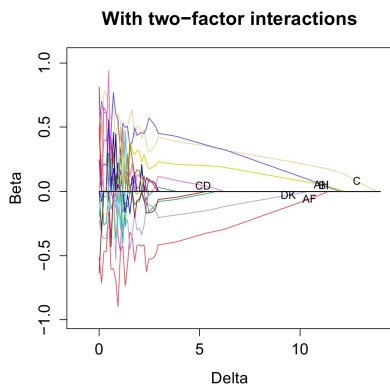
(h) Lasso for Model (4.10) with  $\sigma^2 = 0.1$ , including two-factor interactions.



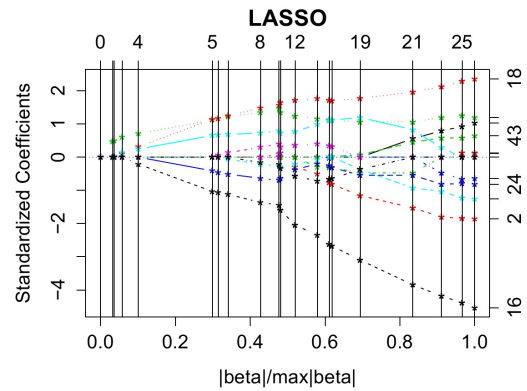
(i) Graphical Dantzig selector for Model (4.11) with  $\sigma^2 = 0.1$ , including two-factor interactions.



(j) Lasso for Model (4.11) with  $\sigma^2 = 0.1$ , including two-factor interactions.

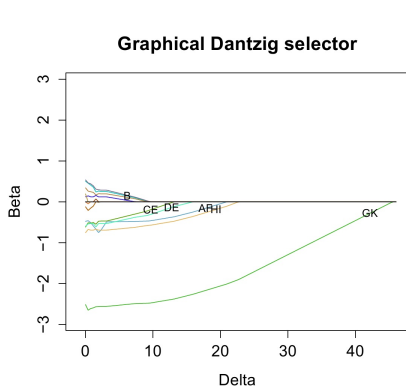


(k) Graphical Dantzig selector for Model (4.12) with  $\sigma^2 = 0.1$ , including two-factor interactions.

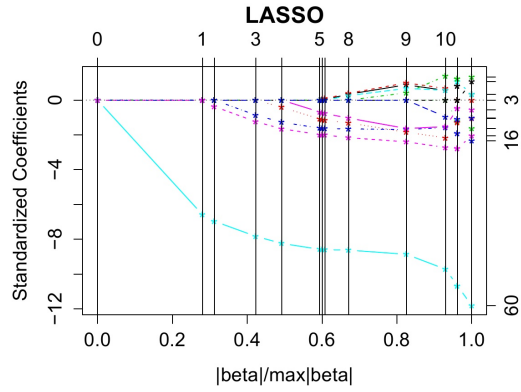


(l) Lasso for Model (4.12) with  $\sigma^2 = 0.1$ , including two-factor interactions.

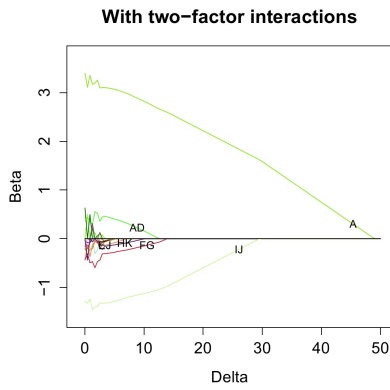
Figure 4.2: Plots obtained by the graphical Dantzig selector and the Lasso for the data generated from models (4.7) - (4.15) with  $\sigma^2 = 0.1$ . Two-factor interactions have been included in the analysis.



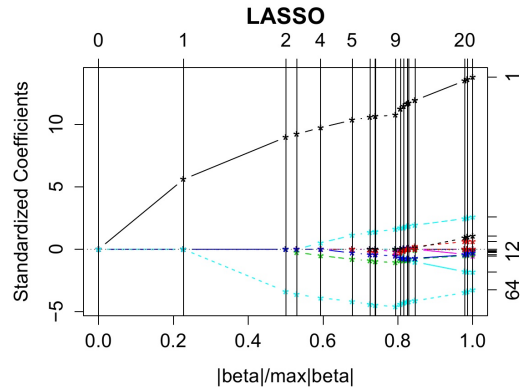
(m) Graphical Dantzig selector for Model (4.13) with  $\sigma^2 = 0.1$ , including two-factor interactions.



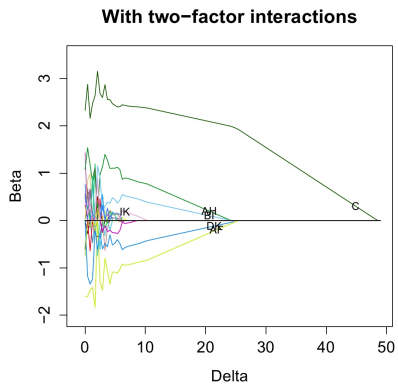
(n) Lasso for Model (4.13) with  $\sigma^2 = 0.1$ , including two-factor interactions.



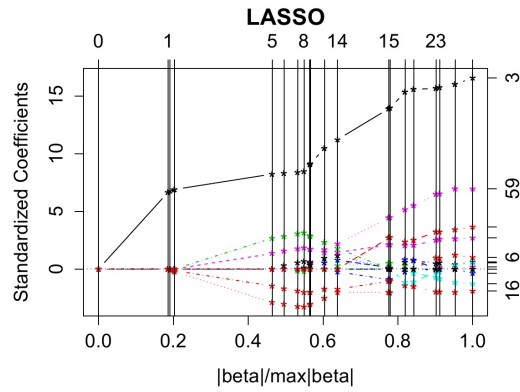
(o) Graphical Dantzig selector for Model (4.14) with  $\sigma^2 = 0.1$ , including two-factor interactions.



(p) Lasso for Model (4.14) with  $\sigma^2 = 0.1$ , including two-factor interactions.



(q) Graphical Dantzig selector for Model (4.15) with  $\sigma^2 = 0.1$ , including two-factor interactions.



(r) Lasso for Model (4.15) with  $\sigma^2 = 0.1$ , including two-factor interactions.

Figure 4.2: Plots obtained by the graphical Dantzig selector and the Lasso for the data generated from models (4.7) - (4.15) with  $\sigma^2 = 0.1$ . Two-factor interactions have been included in the analysis.

## 4.2.5 Factor screening using a projection based method

The projection based method was then applied to the models (4.7) - (4.15). The variances estimated for the data generated with  $\sigma^2 = 0.1$  are shown in Table 4.27.

Model	One active factor				Two active factors				Three active factors			
(4.7)	E	F	D	K	FK	AB	AC	DI	ABC	AFK	ADE	EFK
	8.72	9.05	9.26	9.41	4.89	5.32	6.07	6.34	0.14	0.16	0.22	1.59
(4.8)	F	K	D	E	AC	FK	EH	DI	ABC	ACD	DFG	DHI
	5.85	5.92	5.94	6.00	1.39	3.46	4.29	4.46	0.04	1.25	1.26	1.34
(4.9)	C	B	A	D	AC	CH	CJ	CD	ABC	AFK	CEG	ADE
	5.57	7.68	7.74	7.87	4.24	4.88	4.96	5.26	0.10	0.96	0.97	0.99
(4.10)	C	B	A	J	CD	BC	CE	AC	ABC	CHJ	CDE	AFK
	6.68	9.13	10.24	11.25	4.75	4.94	5.13	6.07	0.13	0.24	0.28	0.51
(4.11)	A	K	E	I	AK	AE	AF	AH	AHJ	AFK	AGI	ABC
	1.32	5.56	5.77	5.77	1.07	1.13	1.16	1.23	0.03	0.04	0.04	0.06
(4.12)	C	I	K	E	CD	CI	CK	BC	ACE	CFH	CGJ	CEI
	2.04	3.00	3.12	3.17	1.47	1.59	1.69	1.81	0.08	0.46	0.50	0.59
(4.13)	B	C	F	K	GK	CE	BC	BI	GJK	BHI	BGK	ABE
	16.41	16.71	20.18	20.44	2.86	12.45	12.60	13.27	0.18	1.26	1.55	1.63
(4.14)	A	D	H	E	AD	AE	AF	AH	ADE	AGH	AIJ	ABD
	5.14	24.14	24.46	24.47	2.15	4.41	5.36	5.45	0.07	0.63	0.71	1.48
(4.15)	C	E	I	K	CD	BC	CI	CE	ACE	CGJ	CFH	CEI
	8.75	25.97	26.16	26.29	6.89	7.17	7.13	7.42	0.02	1.70	1.72	2.45

Table 4.27: The variance estimated by the projection based method for the data generated with  $\sigma^2 = 0.1$ . The variances are given in the order of increasing value.

For the models (4.7) - (4.10) it is observed that the estimated variances,  $\hat{s}^2$ , are reduced when going from one to two active factors, and further reduced when considering three active factors. From these results one can conclude that there at least should be three active factors in the model, but the method does not provide any information whether more factors should be included. It is seen that ABC provided a lower variance than the rest of the combination of three factors. Given that other analysis methods had indicated that three active factors were appropriate for the models, it would be reasonable to conclude that A, B and C are the most important factors in the models (4.7) - (4.10). However, for Model (4.7) the variances for ABC and AFK are rather similar. This was expected as the projection based method allows for three-factor interactions to be present in the model, and with these included, there are three models that equally well can explain the variation in the data. For Model (4.11) there are five models that will produce the exact same data whenever noise is omitted. However, if the sparsity principle is applied the factors A, B and C are chosen in both cases. Based on  $\hat{s}^2$  one may conclude that A, C and E are active for Model (4.12), that G, J and K are the most important in Model (4.13), that A, D and E are most important for (4.13) and A, C and E the active in (4.15). For all of these four models the true active factors are A, B, C and D. Hence, it is seen that erroneous conclusions could be made for these models. This was expected

as the projection based method is restricted to at most three active factors. These results illustrates a weakness of the method. From the results it is not evident that three factors are not enough for the last four models, and it is not possible to check by this method alone. Hence, additional analysis is required to support this method.

The results found for the data generated with  $\sigma^2 = 0.5$  are given in Table 4.28.

Model	One active factor				Two active factors				Three active factors			
(4.7)	D	F	K	A	FK	AC	DJ	DI	ADE	DFK	ABC	DIJ
	5.40	5.64	5.74	6.61	2.42	3.91	3.96	4.08	0.04	0.62	0.65	0.66
(4.8)	F	A	K	D	AC	FK	DI	EG	ABC	EGH	ACJ	DIJ
	4.68	4.96	4.97	4.98	1.44	2.25	2.87	3.84	0.25	0.33	0.74	0.84
(4.9)	C	A	D	F	AC	CH	CD	CJ	ABC	FIJ	ACG	CDI
	6.77	8.46	8.60	9.14	3.83	6.19	6.25	6.27	0.21	1.79	2.26	2.26
(4.10)	C	B	A	F	CE	BC	CD	AC	ABC	CHJ	ADE	AFK
	6.97	9.40	10.20	10.87	5.10	5.68	5.91	6.66	0.16	0.20	0.21	0.61
(4.11)	A	J	K	H	AK	AI	AJ	AH	ABC	AHJ	AGI	AFK
	2.08	4.68	4.84	5.00	1.65	1.71	1.76	1.91	0.25	0.35	0.47	0.67
(4.12)	K	C	B	I	CK	IK	BI	CD	ABF	FGK	BDI	BCK
	2.51	2.97	3.22	2.41	1.54	1.64	1.75	1.96	0.18	0.43	0.47	0.49
(4.13)	C	B	K	I	GK	BI	CE	HI	GJK	CFI	CGK	BHI
	13.00	13.20	14.26	14.89	1.68	8.83	9.95	10.03	0.33	0.81	1.03	1.10
(4.14)	A	J	D	F	AD	AE	AJ	AC	AIJ	ADE	AGH	ABD
	6.48	27.90	28.02	28.73	3.87	5.01	6.07	6.62	0.34	0.35	0.54	2.08
(4.15)	C	I	K	E	CI	CK	BC	CE	ACE	CEI	CEK	CFH
	9.44	23.39	23.43	24.87	6.79	7.10	8.29	8.43	0.04	1.83	2.16	3.08

Table 4.28: The variance for the models (4.7) - (4.15), estimated by the projection based method for the data with  $\sigma^2 = 0.5$ , given in the order of increasing value.

From Table 4.28 it is seen that  $\hat{s}^2$  is reduced as the number of factors is increased, for all of the models. The method had some difficulties with identifying the active factors in the first model. One may erroneously conclude that A, D and E are the active factors in the first model. With these three factors it is possible to form a model that will provide the same data as Model (4.7) whenever noise is omitted. On the contrary, this is not the case for the factors D, F and K. The fact that these factors were ranked before ABC may be explained by the high level of noise. The values given in Table 4.28 indicate that A, B and C are the active factors in (4.8) - (4.11). However, additional analysis should be conducted in order to confirm that three active factors are sufficient for the models. For the last four models there are four active factors in the true model. From the values in Table 4.28 one could easily come to erroneous conclusions regarding the active factors. From the results in Table 4.27 and Table 4.28 it is found that the projection based method provided fairly good results for the models with less than four active factors. From the given study it is also clear that the method should be performed in cooperation with other methods in order to be confirm the number of active factors.

As the projection based method performed relatively well for the models (4.7) - (4.11), the method was tested on ten data sets for each level of noise. As the projection based method is not appropriate for models with more than three active factors, the results for the last four models were omitted. The fraction of times the correct combination of factors were given the lowest estimated variance when considering three active factors are given in Table 4.29.

Model	$\sigma^2 = 0.1$	$\sigma^2 = 0.5$
(4.7)	4/10	3/10
(4.8)	10/10	7/10
(4.9)	10/10	8/10
(4.10)	10/10	5/10
(4.11)	2/10	3/10

Table 4.29: The fraction of times the active factors were ranked first in the list produced by the projection based method, for the models (4.7) - (4.11).

It is seen that the projection based method was able to identify the correct terms for the data generated from the models (4.8) - (4.10) with  $\sigma^2 = 0.1$ . For Model (4.7) and Model (4.11) there are three and five models, respectively, that would generate the very same data whenever noise is omitted. Hence, the relatively low fraction of successes for these two models was not unexpected. The rate of successes was decreased for the first four models when the level of noise was increased.

#### 4.2.6 Focus on using methods that can confirm each other

In this subsection three different methods are used together in order to obtain a more reliable analysis. It is not possible to know how many factors that are active before performing any analysis, but it is assumed that there are at most four. Contrast plots are used as a basis for identifying active factors. Whenever interpretation of the contrast plots are unclear additional analysis is performed using either

1. Orthogonalization
2. The projection based method
3. Residual variance of fitted models of a given form with certain effects included.

The methods are tested in the order given in this list. The idea was that the methods together with the contrast plots would provide adequate information about the model under consideration. A large amount of plots are needed for the first procedure, so in order to save space, only the most informative contrast plots are presented in this thesis. For the data generated with  $\sigma^2 = 0.1$  most of the plots are presented in this subsection, but some of the less important plots are given in Appendix C. In order to limit the number of pages even more the most important plots obtained for the data generated with  $\sigma^2 = 0.5$  are all presented in Appendix C. Code on how to obtain these plots is given in Appendix B.

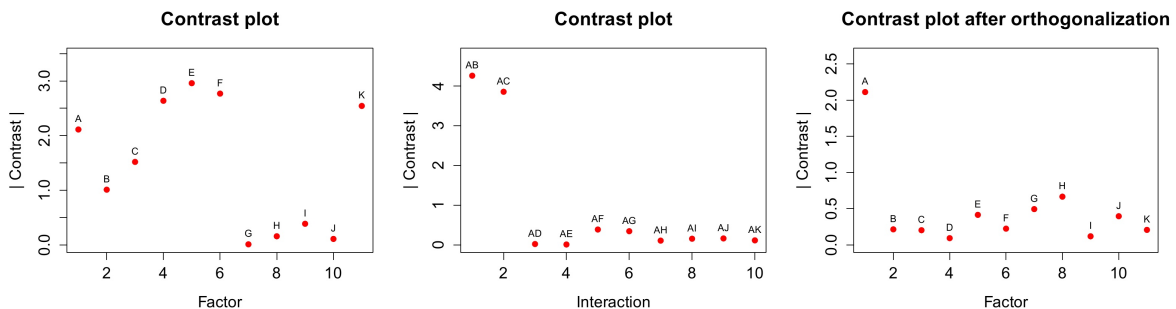
When using the third method, a set of models are fitted with the factors and the number of interaction terms as indicated by the main effects contrast plot. Consider an example where the analysis performed so far pointed to a model with four active factors and one interaction term. When the analysis also indicated that the factors B and C are important for the model, one can proceed the analysis by fitting

$$E(Y) = B + C + X_1 + X_2 + X_{interaction} \quad (4.16)$$

where  $X_1 + X_2$  represents all combinations of factors, excluding B and C, and  $X_{interaction}$  is a two-factor interaction term comprising the factors in the given model. For this specific case there are  $6 \cdot \binom{9}{2} = 216$  models to be fitted and evaluated.

### Considering the data in Appendix A generated with $\sigma^2 = 0.1$ .

As a starting point for the analysis the main effects contrast plot was made for Model (4.7), as shown in Figure 4.3a. From this plot it is observed that most of the points cluster in two lines, and that B, C and possibly A breaks the pattern of these lines. This indicates that there are two interaction terms in the model and that A, B and C may be the active factors. Contrast plots for interactions with these three factors were plotted. From the interactions contrast plot for factor A in Figure 4.3b it seems reasonable to conclude that AB and AC are the interaction terms in the model. The same conclusion was made from the other two interaction plots, as seen in Figure C.9a and Figure C.9b in Appendix C. By first performing orthogonalization with  $\mathbf{X}_1 = [AB, AC]$  and  $\mathbf{X}_2 = [A, B, C, D, E, F]$ , and then with  $\mathbf{X}_1 = [A, AB, AC]$  and  $\mathbf{X}_2 = [G, H, I, J, K]$ , the main effects contrast plot in Figure 4.3c was obtained. From these contrast plots it seems reasonable to conclude that A, AB and AC are the active effects in the model. From the results of the projection based method it is evident that the model consist of more than two active factors, but the variances are rather similar for the combinations ABC, AFK and ADE. The graphical method points to the most parsimonious of these models. This is also the model that normally will be considered as the most plausible explanation, but the projection based method also reveals the closest alternatives.



(a) Model (4.7) with  $\sigma^2 = 0.1$     (b) Model (4.7) with  $\sigma^2 = 0.1$     (c) Model (4.7) with  $\sigma^2 = 0.1$

Figure 4.3: Contrast plots for the data generated from Model (4.7) with  $\sigma^2 = 0.1$ .



The main effects contrast plot for Model (4.8) is shown in Figure 4.4a. From this plot it can be concluded that there are two interaction terms in the model because most of the points are distributed along two horizontal lines. As B deviates from these two lines it is of interest to look at the interactions contrast plot for this factor. From Figure 4.4b it is seen that the interaction contrasts lie on two lines, which indicates that there are two active effects in addition to factor B. Because it was difficult to find more useful information from these plots, the results from the projection based method were considered. Because the results obtained from this method indicated that A, B and C are the active factors, the interactions contrast plots for A and C were plotted as well. From Figure 4.4c it is evident that the interactions AB and AC are part of the model. The contrast plot for interactions with C in Figure C.9c support this belief as AC has a higher value than the other interaction contrasts in this plot. After orthogonalization with  $\mathbf{X}_1 = [AB, AC]$  it is found that the main effect of A is active, as indicated by the plot in Figure 4.4d. Consequently, it seems reasonable to conclude that A, AB and AC are the terms to be included in the model.

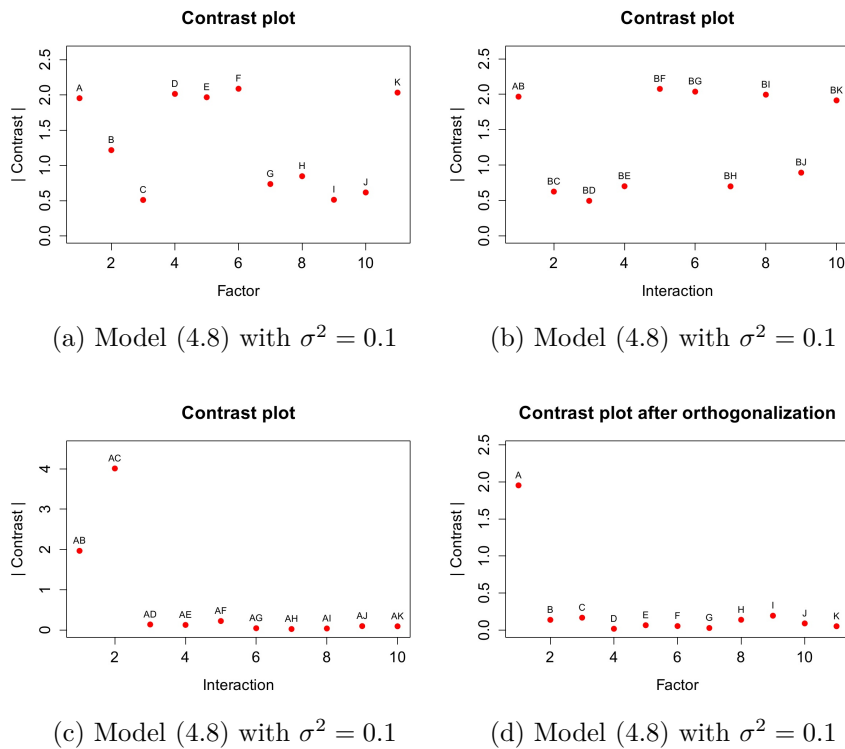


Figure 4.4: Contrast plots for the data generated from Model (4.8) with  $\sigma^2 = 0.1$ .

For Model (4.9) the main effects contrast plot displayed in Figure 4.5a indicates that factor C should be included in the model, and furthermore that two interaction effects are present. Based on these observations it is natural to plot the contrasts for interactions with C. As seen in Figure 4.5b this plot is not easy to interpret. As limited information was available from these plots, the results from the projection based method was included in the analysis. The projection based method suggested

that A, B and C are important for the model. Thus, the interaction plots for A and B were plotted as well, as shown in Figure C.9d and Figure C.9e in Appendix C. It is not evident how to interpret these plots either, but the interactions contrast plot with B gives a slight indication that BG and BI should be included in the model. This is not in correspondence with the results from the projection based method, and it also seems rather implausible from the plot in Figure 4.5a as both G and I are on the same level as the factors H and J. As no clear indication of active effects was found by these two methods, additional analysis had to be performed. From the analysis performed so far there was no reason to believe that there were more than three active factors in the model. Hence, a set of models with three active factors were fitted. The factor C was included in all of them, while the two remaining factors were found by iterating through all combinations of two factors. The models also included two interaction terms, as this was indicated by Figure 4.5a. The terms of the five models with the smallest residual variances are presented in Table 4.30. It is seen that the model with the smallest residual variance is the one with the effects A, B, C, AB and AC. As this is in agreement with the result from the projection based method, and does not contradict the main effects contrast plot the confidence in this model was increased.

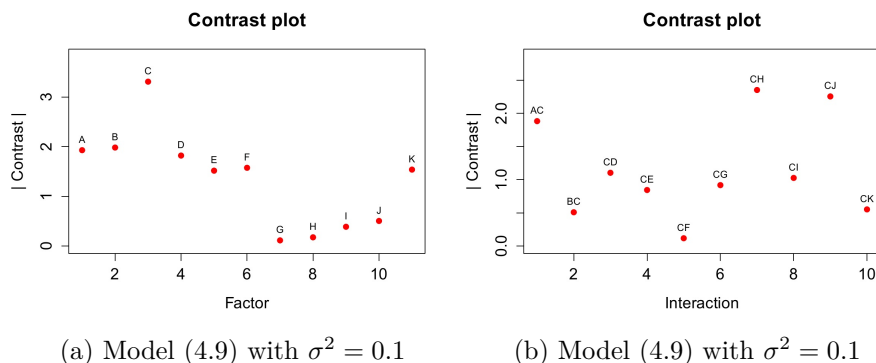


Figure 4.5: Contrast plots for the data generated from Model (4.9) with  $\sigma^2 = 0.1$ .

Residual variance	Terms in the fitted model
0.28	A, B, C, AB, AC
1.28	C, E, G, CG, EG
1.31	C, D, I, CI, DI
1.33	C, F, J, CJ, FJ
1.34	C, H, K, CH, HK

Table 4.30: The effects in the five models with the smallest residual variances, fitted to the data from Model (4.9) with  $\sigma^2 = 0.1$ . All of the models were fitted with the main effect C, a total of three active factors and two interaction terms.

From the contrast plot in Figure 4.6a it seems like A, B and C are the active factors and that there are one interaction term in Model (4.10). The interactions contrast plots with these three factors were plotted, and they all suggested interactions with

D and E, as seen in the figures C.9f, C.9g and C.9h. This points to a model with five active factors with more interaction effects than what was indicated by the main effect plot. Also, the contrasts of D and E are on the same level as the factors F-K in this plot. Although it is possible that active factors are located on the same level as inert factors there seems to be a contradiction here. In cases where there are contradictions between the plots one should confirm the results by performing other methods. By the projection based method it was suggested that A, B and C are the active factors in the model. As the result from the main effects contrast plot is in agreement with the projection based method the confidence in these factors is increased. One explanation is that the interaction term in the model is comprised of three factors. Thus, three-factor interactions contrast plots were produced. The plot in Figure 4.6b indicates that ABC should be included in the model as this interaction breaks the pattern of the two horizontal lines. The main effects contrast plot made after orthogonalization is given in Figure 4.6c. In contrast to the columns of two-factor interactions, the columns of three-factor interactions are not orthogonal to the column of the constant term. Hence, when performing orthogonalization the constant terms has to be included. From Figure 4.6c and Figure 4.6d, together with the discussion given above, it seems reasonable to conclude that A, B, C and ABC are the true terms the model.

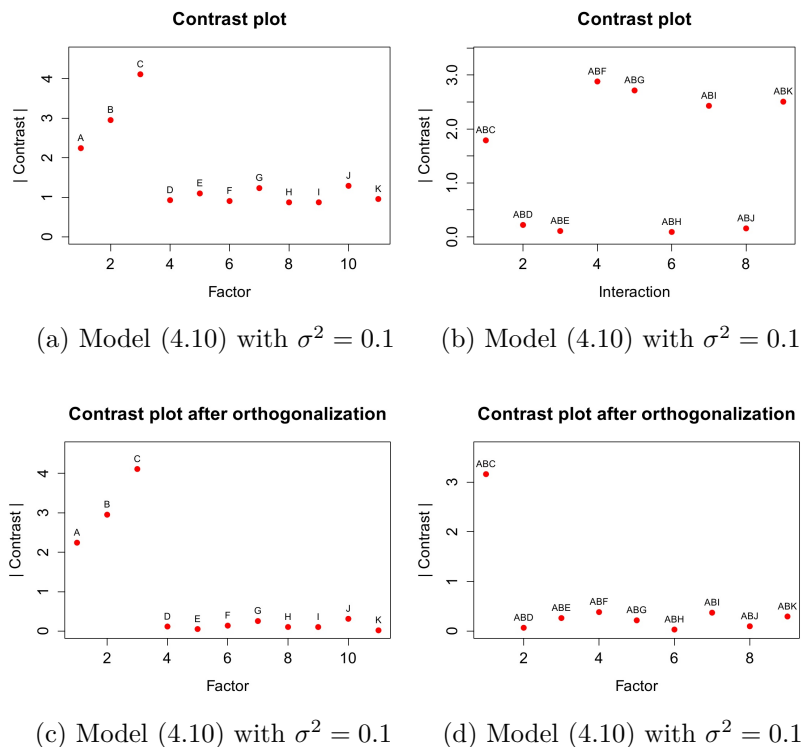
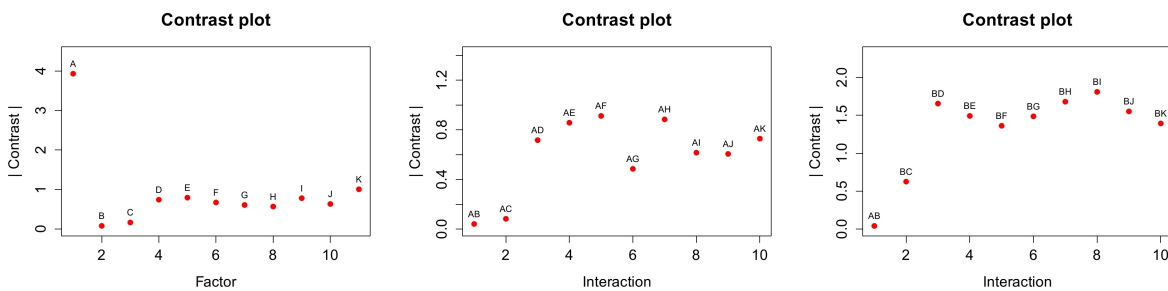


Figure 4.6: Contrast plots for the data generated from Model (4.10) with  $\sigma^2 = 0.1$ .

The contrast plot for the main effects obtained for Model (4.11) is displayed in Figure 4.7a. This plot strongly indicates that the main effect of A should be included in the model, as well as one interaction term. From this plot it is observed that the contrasts for the factors D-K lie on a line, and that the contrasts for B and

C are close to zero. This indicates that B and C are involved in an interaction term, because their corresponding columns in the design matrix are orthogonal to columns of interactions consisting of these two factors. Thus, these factors are not affected to the same extent as the other factors. In the contrast plot found for the interactions involved with A in Figure 4.7b, it is observed that the contrasts of AB and AC are approximately zero while the others are on a higher level. The interactions contrast plot in Figure 4.7c suggests that BC is the interaction term to include in the model, because this contrast is on its own level. The contrast AB is close to zero in this plot as this effect is orthogonal to A. The other contrasts are on approximately the same level, indicating that there are one other term besides BC. Similar observations are made from the contrast plot for the interactions with C in Figure C.9i. Interpretation of these plots are relatively clear, but orthogonalization was performed in order to confirm that the model consists of the terms A and BC. These plots are found in Figure C.9j and Figure C.9k in Appendix C. The contrast plots after orthogonalization was in agreement with this conclusion. Note again that the graphical method points to the most parsimonious model for explanation of the variation in the data.



(a) Model (4.11) with  $\sigma^2 = 0.1$     (b) Model (4.11) with  $\sigma^2 = 0.1$     (c) Model (4.11) with  $\sigma^2 = 0.1$

Figure 4.7: Contrast plots for the data generated from Model (4.11) with  $\sigma^2 = 0.1$ .

Consider now Model (4.12). From the main effects contrast plot in Figure 4.8a it is found that C should be included in the model. Due to this observation the corresponding interactions contrast plot was made. The plot in Figure 4.8b strongly suggests that BC and CD should be included in the model, as these breaks the pattern of the other effects. The fact that AC is close to zero indicates that either A or C or both are active. By performing orthogonalization with  $\mathbf{X}_1 = [BC, CD]$  the plot in Figure 4.8c was obtained. From this plot together with Figure 4.8b it seems reasonable to conclude that the model consists of A, C, BC and CD. The interactions contrast plot with C after orthogonalization in Figure 4.8d, supports this selection of the effects. It is observed that this result is not in correspondence with the combination of factors ACE obtained by the projection based method. This is explain by the fact that this method is restricted to at most three active factors. The projection based method provided no signs or warnings that three active factors were not enough. This emphasizes the importance of using several methods when performing variable selection.

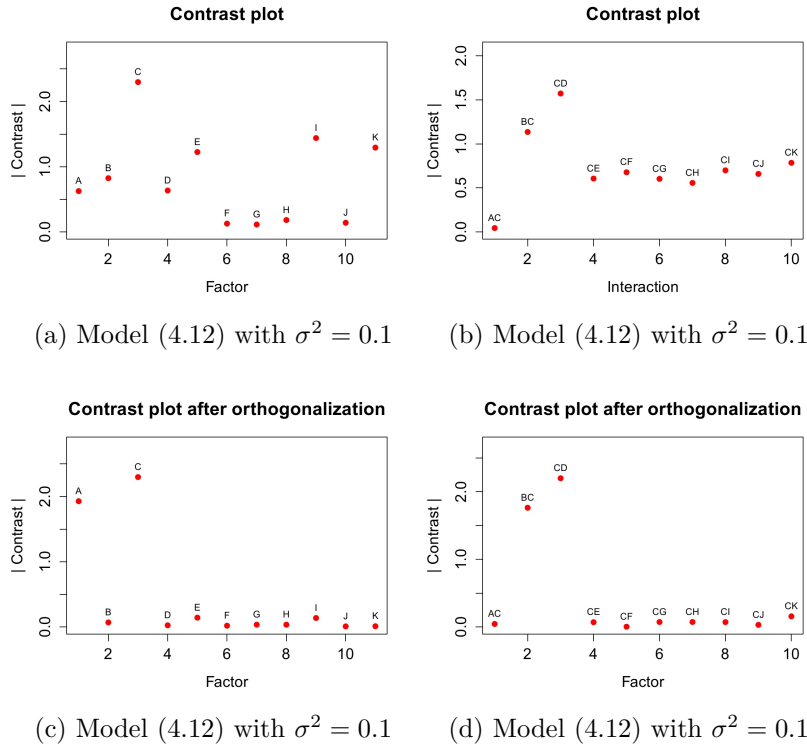
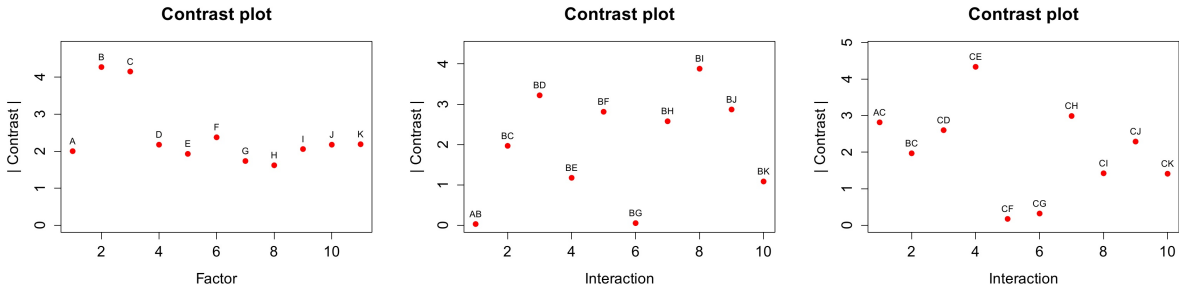


Figure 4.8: Contrast plots for the data generated from Model (4.12) with  $\sigma^2 = 0.1$ .

For Model (4.13) the main effects contrast plot in Figure 4.9a was produced. This plot indicates that B and C are important for the model, and further that one interaction term is present. The contrast plots for the interactions with B and C are shown in Figure 4.9b and Figure 4.9c respectively. It is seen that these are both hard to interpret, but that they vaguely indicate the presence of the terms BI and CE. This seems rather implausible since the main effects contrasts plot only indicated one interaction effect. By this method alone it is difficult to find a path for further analysis. Hence, the results of the projection based method was considered. This method suggested the combination of factors GJK, but it is noticed that this result is not in agreement with the main effects contrast plot. A possible solution is that there might be four active factors in the model. As the projection based method is restricted to at most three active factors, a third approach is required in order to find the active effects in the model. A set of models were fitted with B and C included in all of them, together with all combinations of two additional factors and one interaction term. The fitted models were evaluated according to their residual variance, and the terms in the models with the smallest variances are given in Table 4.31. From this table it is observed that the first three models provided a much lower residual variance than the last two. It is also seen that the model comprising the effects B, C, F, H and BH, and the model with the terms B, C, G, K, and GK were ranked before Model (4.13). Some investigation revealed that the models  $y \sim k + 2B + 3C - 2F - H - 3BH$  and  $y \sim k + B + C + G - K - 3GK$  will generate the exact same data as Model (4.13), whenever noise is omitted.



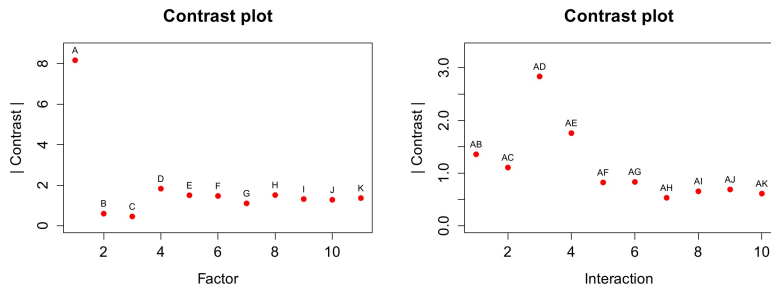
(a) Model (4.13) with  $\sigma^2 = 0.1$     (b) Model (4.13) with  $\sigma^2 = 0.1$     (c) Model (4.13) with  $\sigma^2 = 0.1$

Figure 4.9: Contrast plots for the data generated from Model (4.13) with  $\sigma^2 = 0.1$ .

Residual variance	Terms in the fitted model
0.29	B, C, F, H, BH
0.41	B, C, G, K, GK
0.46	A, B, C, D, CD
2.45	B, C, E, J, CE
2.54	B, C, E, I, EI

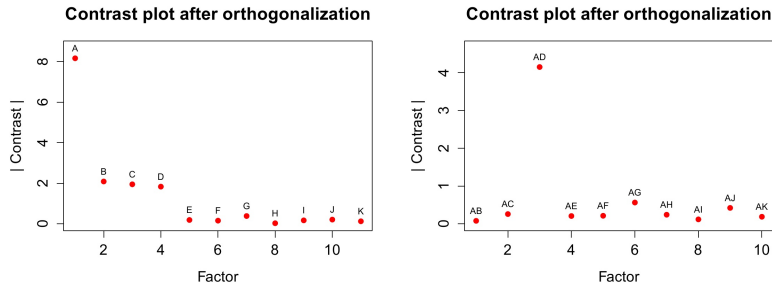
Table 4.31: The models with the smallest residual variances, for Model (4.13) with  $\sigma^2 = 0.1$ . Models were fitted with B and C, four active factors and one interaction.

The main effects contrast plot in Figure 4.10a was obtained for Model (4.14). From this plot it is evident that A should be included in the model, and there seems to be one interaction term. It is also observed that B and C are close to zero. Due to these findings the interactions contrast plots for A, B and C were plotted. It is seen that the interactions contrast plots in Figure C.9l and Figure C.9m are not easy to interpret. From the plot in Figure 4.10b it seems like AD should be included in the model. When performing orthogonalization with  $\mathbf{X}_1 = [AD]$ , the plot in Figure 4.10c was obtained. The interactions contrast plot with A after orthogonalization is given in Figure 4.10d. Based on these plots it seems reasonable to conclude that the model consists of the terms A, B, C, D and AD. Note that this result is not in correspondence with the results of the projection based method, but this was expected as the model has four active factors.



(a) Model (4.14) with  $\sigma^2 = 0.1$     (b) Model (4.14) with  $\sigma^2 = 0.1$

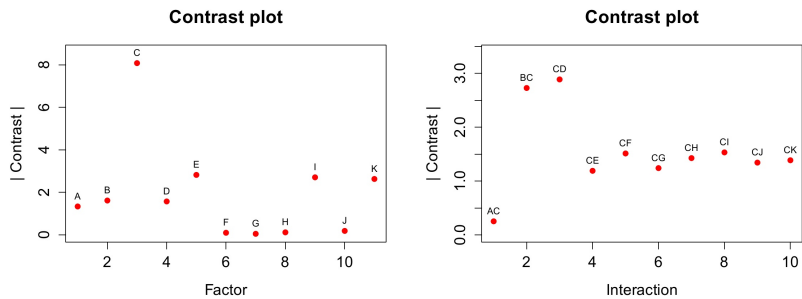
Figure 4.10: Contrast plots for the data generated from Model (4.14) with  $\sigma^2 = 0.1$ .



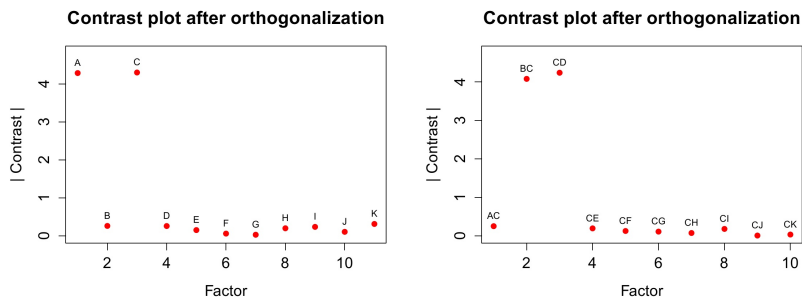
(c) Model (4.14) with  $\sigma^2 = 0.1$       (d) Model (4.14) with  $\sigma^2 = 0.1$

Figure 4.10: Contrast plots for the data generated from Model (4.14) with  $\sigma^2 = 0.1$ .

For Model (4.15) the plot in Figure 4.11a was found. From this plot it is concluded that C is involved with the model. Consequently, the interaction plot with C was plotted, as displayed in Figure 4.11b. This plot strongly indicates that BC and CD should be included in the model, and that there are one other factor in the model, which is not included in these interaction terms. Also, as AC tends to zero, it is indicated that this term is orthogonal to one or both of the factors in the model. After orthogonalization was performed the plots in Figure 4.11c and Figure 4.11d were obtained. From these plots it seems reasonable to conclude that A, C, BC and CD are the true terms in the model. As for the previous three models, this result differs from the findings of the projection based methods, as there are four active factors in the model.



(a) Model (4.15) with  $\sigma^2 = 0.1$       (b) Model (4.15) with  $\sigma^2 = 0.1$



(c) Model (4.15) with  $\sigma^2 = 0.1$       (d) Model (4.15) with  $\sigma^2 = 0.1$

Figure 4.11: Contrast plots for the data generated from Model (4.15) with  $\sigma^2 = 0.1$ .

**Consider the data in Appendix A generated with  $\sigma^2 = 0.5$ .**

From the data generated from Model (4.7) with  $\sigma^2 = 0.5$  the main effects contrast plot in Figure C.10a in Appendix C was obtained. From this plot it is not clear which factors to include in the model, but it may appear that the contrasts cluster in two lines and that A, B, C and E breaks this pattern. The interactions contrast plot with these four factors were plotted. The first three plots strongly suggested that AB and AC should be included in the model. This is especially clear in the plot given in Figure C.10b. From the contrast plot for the interactions with E, the terms EG and EH were found to be important for the model. If all of these terms are included, there are six active factors in the model. As the true model should hold the property of factor sparsity, one should consult with other analysis methods. For Model (4.7) the projection based method indicated that ADE are the most important factors, but this is not in correspondence with the plot in Figure C.10a. One may try to explain the deviation in results by a high level of noise, or one may erroneously conclude that there are more than three active factors in the model. However, orthogonalization with  $\mathbf{X}_1 = [AB, AC]$  was performed as a starting point for the analysis, which resulted in the plot in Figure C.10c. It is seen that the plots indicate that the terms A, E, AB and AC should be included in the model, which would lead to an erroneous conclusion regarding the factor E, due to the high level of noise.

From the main effects contrast plot produced for Model (4.8) it is found that E and J breaks the pattern of the two horizontal lines in Figure C.10d. When considering the contrast plots for interactions with E and J, it was indicated that EG and EH are active effects. This does not seem plausible from the plot in Figure C.10d as the contrasts of G and H are on the same line as three other factors. In this case it is wise to perform additional analysis. When the results of the projection based method were taken into account, it was realized that interactions with A, B and C should be investigated, as ABC provided a lower estimated variance than any other combination of factors. The interaction plot for A is given in Figure C.10e. This plot indicates that the effects AB and AC are part of the model. After orthogonalization was performed, the main effects contrast plot in Figure C.10f was obtained. Based on these plots the most plausible effects in the model are A, AB and AC.

The main effects contrast plot in Figure C.10g was produced for Model (4.9). Based on interpretation of this plot it was found that C and possibly B should be a part of the model. Further, as the rest of the effects cluster in approximately two lines, there should be two interaction terms in the model. The corresponding interactions contrast plot with B in Figure C.10g suggested that BG and BI are appropriate for the model. This seems unlikely from the plot in Figure C.10h. The interaction plot with C did not provide any clear indication of active effects. To proceed the analysis, the result from the projection based method was considered. This method pointed to A, B and C as the active factors. Due to this finding, the contrast plot for interactions with A was made as well, but no obvious interpretation was available from this plot. As there was no clear path for further analysis the third method was conducted. From Figure C.10g it is only evident that C is important for the model.



Because of this, all possible models with three active factors and two interaction effects were fitted, all including the factor C. The effects in the five models with the smallest residual variances are presented in Table 4.32. It is seen that the model with the smallest residual variance consists of the terms A, B, C, AB and AC.

Residual variance	Terms in the fitted model
0.39	A, B, C, AB, AC
1.56	C, D, I, CI, DI
1.71	C, F, J, CJ, FJ
1.73	C, E, G, CG, EG
1.75	C, D, I, CD, DI

Table 4.32: The terms in the five models with the smallest residual variances found for models fitted to the data generated from Model (4.9) with  $\sigma^2 = 0.5$ . The models were all fitted with the main effect C, three active factors and two interaction terms.

From Figure C.10i it seems like A, B and C should be included in Model (4.10), as these points clearly breaks the pattern formed by the other contrasts. It also seems like there are one interaction term in the model. The interactions contrast plot with these factors indicated that interactions with D and E are appropriate for the model. As this is unlikely from the interpretation of Figure C.10i the results from the projection based method was considered. In agreement with the main effects contrast plot this method also indicated that A, B and C are active. Hence, it was determined to plot the three-factor interaction contrasts. From Figure C.10j it might look like ABC is important for the model, but the results are not clear. By performing orthogonalization, the plots in Figure C.10k was obtained. Again, the constant terms has to be included in the orthogonalization as this terms is not orthogonal to the three-factor interaction effect. From these plots it seems reasonable to conclude that the model is comprised of the effects A, B, C and ABC. To confirm, the main effects contrast plot after orthogonalization was performed, and this plot supported this selection of effects.

Consider now Figure C.10l where the main effects contrasts for Model (4.11) are plotted. This plot strongly suggests that A is important for the model. Furthermore, it is seen that the points gather in one single line and that the absolute value of the contrast for factor C is close to zero. Based on these observations it was determined to plot the interactions with A and C. The contrast plot for interactions with A was found to be difficult to analyze. From the contrast plot in Figure C.10m it is seen that AC is close to zero, which gives an indication that either A or C or both should be included in the model. As C is close to zero in Figure C.10l it is concluded that A is likely to be a factor in the model. Based on this line of thought the analysis was proceeded by first orthogonalization with  $\mathbf{X}_1 = [A]$  and  $\mathbf{X}_2 = [AC, BC, CD, CE, CF, CG]$ , and thereafter with  $\mathbf{X}_1 = [A, BC]$  and  $\mathbf{X}_2 = [CH, CI, CJ, CK]$ . The interactions contrast plot obtained after orthogonalization is given in Figure C.10n. From these plots it seems reasonable to conclude that the terms in the model are A and BC. To confirm, orthogonalization was per-

formed with  $\mathbf{X}_1 = [BC]$  and with  $\mathbf{X}_2$  containing the main effects. The plot obtained was in agreement with the original conclusion. The projection based method indicated that A, B and C were active, but it also presented other possible solutions. Again, the graphical method points to the most parsimonious solution.

The next model to be evaluated was Model (4.12). It is seen that it is not clear how to interpret the main effects contrast plot in Figure C.10o. The plot gives a slight indication that there are two interaction effects and that K and possibly B and C should be part of the model. As G is close to zero one may think that this factor is involved in one of the interactions. Contrast plots for interactions with B, C, G and K were produced. The contrast plots for interactions with B and C are given in Figure C.10p and Figure C.10q, and they indicate the presence of the effects BD, BI and CD. The contrast plots for interactions with G and K indicated that AG, GJ and IK are appropriate for the model. As this would lead to a model with a lot of active factors, it was desirable to perform other methods to support interpretation of the contrast plots. Hence, the result ABF from the projection based method was taken into consideration. As there is lack of correspondence between the results from the projection base method and the main effects contrast plot, an additional analysis was needed. Because non of the factors stood out as clear candidates for being active, the third method was not a good alternative either. Due to lack of correspondence between the methods and because there was no clear path for further analysis, the safest thing to do would be to perform additional experiments in order to collect more data. One may explain the lack of correspondence between the methods by a high level of noise and the presence of more than three active factors.

For Model (4.13) the contrast plot for main effects in Figure C.10r was produced. From this plot it seems reasonable to concluded that B and C are important for the model and that there exists one interaction term. The interactions contrast plot with B and C were both hard to interpret. The projection based method was consulted in order to ease the interpretation. This method pointed to GJK as the active factors, but this is not in agreement with the plot in Figure C.10r. As there was no clear path for further analysis by these two methods, the third approach was applied. A set of models were fitted, where all contained the factors B and C, together with all combination of two additional factors and one interaction term. The effects in the models with the smallest residual variances are given in Table 4.33. It is seen that the model with the lowest residual variance is the one with the terms B, C, G, K and GK. This was not unexpected as there exist a model with these effects which would generate the exact same data as Model (4.13) whenever noise is omitted. It is also observed that there are three models with a low residual variance compared to the others. With this analysis the number of plausible models have been reduced from 330 to three, but follow-up experiments may be necessary in order to perform the final selection of variables.

By considering Figure C.10s it is found that A most certainly should be included in Model (4.14), and that there is one interaction term. It was natural to proceed the analysis by plotting the interactions contrast plot with the factor A. From Figure

Residual variance	Terms in the fitted model
0.50	B, C, G, K, GK
0.72	A, B, C, D, CD
0.84	B, C, H, F, BH
2.13	B, C, I, K, BC
2.26	B, C, I, J, BI

Table 4.33: The terms in the five models with the smallest residual variances found for models fitted to the data generated from Model (4.13) with  $\sigma^2 = 0.5$ . The models were all fitted with the main effects B and C, four active factors and one interaction.

C.10t it is seen that interpretation of this plot is not clear. To proceed the analysis, orthogonalization was performed with AD, and also with AE and AC. The resulting plots revealed which of these interaction effects that were appropriate for the model. The plot in Figure C.10u was obtained after orthogonalization with AD. Based on this plot it seems reasonable to conclude that the true effects in the model are A, B, C, D and AD. The main effects contrast plots after orthogonalization with AE and AC confirmed that they are not active effects in the model.

Finally, the main effects contrast plot in Figure C.10v was made for Model (4.15). This plot strongly indicates that C is important, and it also looks like there are two interaction effects in the model. From the interactions contrast plot with C it was found that BC and CD are likely active factors, and that AC is orthogonal to one or two of the factors in the model. By orthogonalization the two plots in Figure C.10w and Figure C.10x were obtained. From these plots it seems reasonable to conclude that A, C, BC and CD are the active effects in the model. It is noted that this result is not in correspondence with the finding of the projection based method. Because the contrast plots together with orthogonalization suggested four active factors, these results seems more plausible than the results obtained by the projection based method.

# Chapter 5

## Discussion

To get a better understanding of the methods it is important to discuss the results produced by each of them. It is of interest to find out if our prior beliefs coincide with the results from the examples, and try to explain any deviations from our expectations. The results produced by the different methods are compared to each other, as well as to the results obtained for data generated with the different level of noise.

The Dantzig selector performed significantly better when applied to data from models with only main effects. This was expected as the 12 run PB design do not meet the requirements of the UUP when two-factor interaction columns are included in the regression matrix. For the models in Example 1 containing interaction effects, the projection based Dantzig selector was found to be more reliable than the Dantzig selector and the graphical Dantzig selector. Furthermore, it was found that the Dantzig selector was preferred over the graphical Dantzig selector. As was discussed during the study of the methods, the difference in results between the Dantzig selector and the graphical Dantzig selector, can be explained by the fact that the solution paths for the effects may cross each other. Hence, an active factor may have a large estimated parameter value for the optimal  $\delta$ , but if the slope of the solution path is steep, it may erroneously be considered as an inert effect by the graphical Dantzig selector. A challenge of these methods is that the inert effects have to be removed from the six terms that are initially suggested. The expanded projection based Dantzig selector discussed in Subsection 3.3.1 aims at removing all the inert effects, before the model is selected. The criteria used for the final selection of model effects were either the residual variance, the AIC or the BIC. It was found that this method is not preferred as it has a high risk of selecting over-fitted models. From the results in Example 2 it was observed that non of the methods using the Dantzig selector performed satisfactory for variable selection. As the methods were tested on designs with more complex properties than what they were intended for, the outcome of these methods were highly unreliable. For such cases, the analysis should be performed with a critical mind and the results used with care. However, even for the most complex models the methods using the Dantzig selector were in most cases able to identify the most important effect in the model. Thus, these methods can be used as support for other analysis, even when the requirements of

the UUP are not met. It is possible that the Dantzig selector would have performed better if it had been restricted by certain assumptions, such as effect heredity.

As was expected beforehand, the results found by the Lasso were relatively close to the results found by the Dantzig selector. It was observed that the Lasso performed slightly better than the Dantzig selector for some of the models, but there were also examples of the opposite. For the models in Example 1 the Lasso was able to identify all of the main effects, but the interaction term AB was lost for some of the models. For most of the models in Example 2 the Lasso was unable to identify all of the active factors. However, for most of them the Lasso was also able to point to the most important effect in the model.

As has been discussed in this thesis, the 12 run PB design have several properties that makes it well suited for experimental analysis. When it is appropriate to assume at most four active factors, the 12 run PB design provides a lot of information in relatively few runs. The projection properties of the 12 run PB design are utilized in the projection based method described in Section 3.5. For the models in Example 1 it was observed that the projection based method was able to identify all of the correct factors for both the low and the high level of noise. The main challenge of this method was to determine how many factors to include. The partial F-test was found to be useful for this task when applied to the models with one and two active factors, even though one inert effect was added to two of the models. This additional factor may be explained by correlation between the noise column and the given effect column. For the models with three active factors in Example 2 the method performed well, even for the data with the high level of noise. Because the method is restricted to at most three active factors, the method failed to provide reasonable results for the models with four active factors. The main drawback of this method is that it may provide no sign of error whenever three active factors are not enough. Hence, whenever the method indicate that two active factors are insufficient, this method cannot be trusted without support from other methods. Because the projection based method does not assume any particular form of the model except sparsity, it can point to models that are disregarded by other methods due to heredity assumptions. Even though the most sparse model is usually preferred it might be useful to consider the closest alternatives.

The partial aliasing of the 12 run PB design is utilized in the graphical method using contrast plots and orthogonalization, as described in Section 3.7. The performance of this method is dependent on the parameter values, the level of heredity, the number of interaction effects and the number of active factors. The principle of effect heredity was used during the performance of the method, but it was found that it was capable of identifying the active effects also for the model without heredity. For some of the models the procedure was able to identify all of the active factors, but it was found that it in some cases needed support from other analysis. Hence, the focus was directed towards methods that could complement each other. The combination of contrast plots, orthogonalization, the projection based method and fitted models of a given form with certain effects included, proved to form a highly

valuable procedure for variable selection. These methods were able to identify all of the active factors for most of the models, even with high level of noise.

For some models there were lack of correspondence in the results produced by the methods in terms of which effects they found to be active. The deviation in results could for some cases be explained by a high level of noise, or by the fact that the requirements for some of the methods was not met. Another explanation was that there for some models existed other models that equally well could explain the variation in the data. This is particularly true for the projection based method, as this method does not assume any particular form of the model. This opens up for even more models that produce the exact same data whenever noise is omitted. In these cases, methods such as the Dantzig selector would choose the most sparse model, while other methods base their selection upon other types of criteria. Hence, it is expected that these methods occasionally suggest different effects to include in the model. In this thesis the parameters were all integers, but this is rather rare in real experiments. Thus, in real experiments there are less cases where more than one model have the ability to explain the data equally well.

For some of the methods, a list of possible models were suggested instead of a single one. In some cases, considering principles such as effect heredity and factor sparsity can provide additional guidance for which of these models to select as the final one. It is also possible to consult with other methods, and investigate any lack of correspondence between the results. A challenge shared by the Dantzig selector, the graphical Dantzig selector, the projection based Dantzig selector, the Lasso and the projection based method is that the inert effects should be removed from the final model. Forward selection, backward elimination or a stepwise procedures can be used, but they often select models that are over-fitted, at least when they are based on the criteria of AIC and BIC, and the data is generated from the 12 run PB design. For some of the models in this thesis the results were improved by these methods, but many of them were still over-fitted. There were also cases where some active factors were removed as well.

From the results in Chapter 4 it is seen that the methods generally produce less accurate results as the number of active factors is increased. For the models with four active factors the contrast plots and orthogonalization performed significantly better than the other methods. For the data sets generated from the more complex models with the high level of noise, most of the methods encountered some problems. The methods presented in this thesis have different strengths and weaknesses that makes them suitable for different types of models. Hence, several analysis methods should be performed in order to ensure a reliable result.

# Chapter 6

## Conclusion

The methods tested in this thesis were found to have a varying level of performance for identifying active factors in a 12 run PB experiment. Based on the results in Chapter 4 it is concluded that the Dantzig selector, the graphical Dantzig selector, and the projection based Dantzig selector are valuable methods for variable selection for sparse cases where the requirements of the UUP are fulfilled. Whenever these requirements are exceeded the results should be used with care, and they should be compared with the results obtained by other methods. As expected, it was found that the difference in the results produced by the Dantzig selector and the Lasso was relatively small. There was no clear sign of which of the two methods to generally prefer. Based on the study conducted in this thesis it is concluded that the projection based method performs reasonably well for models with at most three active factors. However, it is not appropriate to use this method alone whenever the results indicate that two active factors are insufficient. This is due to the fact that it may provide wrong results with no sign of error in cases where more than three factors are active. It was found that the method using contrast plots and orthogonalization performed reasonably well for many of the models tested in this thesis. The principle of effect heredity was used during the analysis, but the method proved to perform well even for the model where heredity was not fulfilled. Interpretation of contrast plots after orthogonalization was used in cooperation with the projection based method, and a method where selection of effects were based the residual variances of fitted models of a certain form with the effects indicated by the contrast plots. A main conclusion found during the study of these method is that a combination of these three methods together form a highly powerful variable selection procedure.

The results found in Chapter 4 together with the discussion in Chapter 5 emphasizes the importance of having proper theoretical knowledge about the method and the design at hand. None of the methods studied in this thesis were able to perform an accurate selection of active factors for all of the models on their own. To avoid erroneous conclusions it is highly important to perform several different methods when performing variable selection. Whenever there are lack of correspondence between the methods, one should be extra critical, consult an expert on the given theoretical

field or try to explain why some of the methods provided wrong results. Follow-up experiments may be needed for complex models with a high level of noise.



# Bibliography

- Box, G. and N. R. Draper, 2007: *Response Surfaces, Mixtures, and Ridge Analyses*. Wiley Series in Probability and Statistics, Wiley.
- Box, G. and J. Tyssedal, (1996): Projective properties of certain orthogonal arrays. *Biometrika*, **83** (4), 950–955.
- Candes, E. and T. Tao, (2007): The Dantzig selector: Statistical estimation when  $p$  is much larger than  $n$ . *The Annals of Statistics*, **35** (6), 2313–2351.
- F.K.H. Phoa, Y.-H. P. and H. Xu, (2009): Analysis of Supersaturated Designs via the Dantzig Selector. *Journal of Statistical Planning and Inference*, **139** (7), 2362–2372.
- James, G. and P. Radchenko, (2009): A generalized Dantzig selector with shrinkage tuning. *Biometrika*, **96** (2), 323–337.
- James, G. M., P. Radchenko, and J. Lv, (2009): Dasso: connections between the Dantzig selector and lasso. *Journal of the Royal Statistical Society B*, **71** (1), 127–142.
- Lin, D. and N. Draper, (1992): Projection Properties of Plackett and Burman Designs. *Technometrics*, **34** (4), 423–428.
- Meinshausen, N., G. Rocha, and B. Yu, (2007): Discussion: A tale of three cousins: Lasso, L2Boosting and Dantzig. *The Annals of Statistics*, **35**, 2373–2384.
- Tibshirani, R., (1996): Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society (Series B)*, **58**, 267–288.
- Tyssedal, J., (2011): Two-level factorial designs and how to block them.
- Tyssedal, J., H. Grinde, and C. Røstad, (2006): The use of a 12-run Plackett–Burman Design in the Injection Moulding of a Technical Plastic Component. *Qual. Reliab. Engng. Int.*, **22**, 651–657.
- Tyssedal, J. and R. Niemi, (2011): Graphical Aids for the Analysis of Two-level Non-regular Designs. *Journal of Computational and Graphical Statistics - to be published*.
- Tyssedal, J. and O. Samset, (1997): Analysis of the 12 run Plackett and Burman design. *Preprint Statistics (8/1997)*.

- Tyssedal, J. and O. Samset, (1999): Two-Level Designs with Good Projection Properties. *Preprint Statistics (12/1999)*.
- van de Geer, S. A. and P. Bühlmann, (2009): On the conditions used to prove oracle results for the Lasso. *Electronic Journal of Statistics*, **3**, 1360–1392.
- Wang, J. and C. Wu, (1995): A hidden projection property of Plackett-Burman and related designs. *Statistica Sinica*, **5**, 235–250.
- Wiik, H., (2013): Non-regular designs with censored data.
- Wu, C. and M. Hamada, (2011): *Experiments: Planning, Analysis, and Optimization*. Wiley Series in Probability and Statistics, Wiley.
- Wu, C. F. J. and M. Hamada, (1991): Analysis of Censored Data From Highly Fractionated Experiments. *Technometrics*, **33**, 25–38.

# Appendix A

## A.1 Data used for the study in Example 1

The data generated with  $\sigma^2 = 0.1$ :

$$\begin{aligned}y_{(4.1)} &= (1.21, 0.73, 1.27, 0.77, 1.64, 1.18, 2.98, 2.54, 3.39, 3.38, 3.23, 2.85) \\y_{(4.2)} &= (-0.68, -0.86, -1.27, 0.69, 0.75, 0.92, 3.10, 3.21, 3.48, 4.92, 4.85, 4.522) \\y_{(4.3)} &= (-0.27, 0.09, 0.38, 1.73, 1.68, 1.98, 2.16, 1.95, 2.09, 3.70, 3.84, 4.04) \\y_{(4.4)} &= (-2.14, -2.08, -1.88, 0.09, 0.05, -0.64, 2.10, 2.26, 1.95, 8.15, 7.92, 8.15) \\y_{(4.5)} &= (-1.32, -0.62, -1.40, 0.96, 1, 0.48, 0.61, 0.77, 0.65, 7.32, 7.12, 6.44) \\y_{(4.6)} &= (0.09, -0.62, -0.16, 0.50, -0.57, -0.02, 2.10, 1.81, 1.74, 5.99, 6, 6.2)\end{aligned}$$

The data generated with  $\sigma^2 = 0.5$ :

$$\begin{aligned}y_{(4.1)} &= (1.63, 1.06, 1.02, 0.69, 0.73, 1.46, 3.38, 2.59, 2.32, 2.00, 3.28, 3.29) \\y_{(4.2)} &= (-0.12, -1.04, -2.43, 0.42, 1.16, 0.64, 3.39, 2.13, 3.42, 5.05, 4.87, 4.15) \\y_{(4.3)} &= (-1.01, 0.05, 0.03, 3.35, 1.41, 1.50, 1.93, 1.19, 2.58, 3.70, 3.95, 2.93) \\y_{(4.4)} &= (-1.21, -2.26, -1.66, 0.23, -0.40, -1.48, 2.45, 1.92, 2.01, 6.98, 8.47, 8.78) \\y_{(4.5)} &= (-0.43, -1.53, -0.44, 0.62, 1.11, 1.96, 2.12, 0.93, -0.43, 6.48, 6.65, 6.73) \\y_{(4.6)} &= (-0.02, 0.64, -0.71, 0.26, 0.89, -0.06, 2.09, 1.72, 1.27, 7.02, 5.61, 6.19)\end{aligned}$$

## A.2 Data used for the study in Example 2

The data generated with  $\sigma^2 = 0.1$ :

$$\begin{aligned}y_{(4.7)} &= (4.79, 5.21, 0.71, 0.87, -3.2, -2.77, 3, 2.18, -0.92, 7.46, 3.14, 3.42) \\y_{(4.8)} &= (3.69, 4.09, 0.11, 2.11, -1.86, -1.91, 4.09, 3.85, -0.09, 6.29, 1.75, 2.07) \\y_{(4.9)} &= (0.50, -0.14, 1.11, 0.97, 2.37, 2.12, 3.70, 4.30, -2.7, 9.05, 2.05, 2.11) \\y_{(4.10)} &= (-3.91, -4.63, 2.77, 1.67, 2.93, 2.82, 1.72, 2.36, 1.20, 8.10, 0.98, 0.74) \\y_{(4.11)} &= (0.84, 0.62, -1.39, -1.82, 0.74, 1.03, 3.52, 3.20, 5.26, 5.20, 2.98, 3.44) \\y_{(4.12)} &= (1.88, 1.97, 2.61, -2.16, 1.96, 4.15, 4.24, 2.27, 1.78, 3.95, -0.23, 2.15) \\y_{(4.13)} &= (-3.10, -3.31, 3.34, -0.65, 1.57, 8.91, 7.52, -1.07, -2.79, 5.57, 2.96, 7.21) \\y_{(4.14)} &= (-3.08, -2.60, -3.70, -3.11, 0.74, -1.15, 9.07, 3.04, 7.08, 4.97, 8.90, 3.04) \\y_{(4.15)} &= (0.16, 0.49, 4.11, -8.32, 4.01, 7.71, 8.36, 4.07, -0.18, 8.16, -4.24, 0.00)\end{aligned}$$

**The data generated with  $\sigma^2 = 0.5$ :**

$$y_{(4.7)} = (3.84, 3.97, 1.56, 1.45, -1.38, -3.07, 2.55, 3.46, -0.98, 6.09, 2.03, 3.26)$$

$$y_{(4.8)} = (3.27, 3.29, 0.12, 2.57, -1.11, -2.39, 3.81, 3.59, -0.48, 5.48, 2.18, 2.70)$$

$$y_{(4.9)} = (0.63, 0.79, 1.64, -0.36, 1.41, 0.82, 4.53, 4.85, -2.75, 9.38, 1.33, 2.41)$$

$$y_{(4.10)} = (-3.98, -4.01, 3.70, 2.14, 2.70, 2.24, 2.38, 1.69, 1.35, 8.34, 0.6, 1.36)$$

$$y_{(4.11)} = (1.24, 0.09, -2.16, 0.09, 1.48, 1.98, 3.05, 2.69, 5.61, 5.45, 2.56, 3.12)$$

$$y_{(4.12)} = (2.90, 1.12, 2.34, -2.22, 1.58, 2.47, 4.61, 1.33, 1.91, 2.72, -1.18, 2.85)$$

$$y_{(4.13)} = (-2.18, -2.31, 2.37, -1.3, 1.08, 8.49, 6.19, -0.84, -2.01, 3.75, 1.35, 7.37)$$

$$y_{(4.14)} = (-4.58, -3.50, -3.36, -3.72, 1.06, -1.59, 9.28, 2.60, 6.25, 4.99, 9.88, 3.67)$$

$$y_{(4.15)} = (-0.21, 0.13, 3.95, -7.9, 4.07, 8.17, 8.13, 3.60, 1.51, 7.97, -3.99, 1.15)$$

# Appendix B

## B.1 Introduction

---

### Notation used in the R code

---

a: Choose model based on a=1: variance, a=2: AIC or a=3: BIC  
af: Vector of factors that are known to be active  
ff: Vector of terms in the model  
inter: Equal to 2 or 3, indicate if only 2-factor inter. are added, or also 3-factor inter.  
model: Reduced model in the partial F-test  
n.i: Number of interaction terms in the model  
numb: Number of factors in the model  
full: The full model in the partial F-test  
sigma: Standard deviation of the noise of the generated data  
x: The design matrix, or the design matrix together with interaction columns  
y: Response vector

Recall that `y` has to be defined with the correct values outside of the function, before it is given as input. This is important as R otherwise would produce erroneous results, as discussed in Section 3.8.

---

### Overview of functions and R code

---

```
design.matrix()
add.inter(x,inter)
make.headers(x,ff,inter)
Generation of data
DS(x,y,sigma)
DS.plot(x,y)
DS.sub(x,y,ff,inter)
DS.projection(x,y,inter)
reduce(x,y,ff,inter)
DS.var.aic.bic(x,y,a,inter)
l.lasso <- function(x,y)
equal.rows1(x)
```

```
equal.rows2(x)
equal.rows3(x)
proj.based1(x,y)
proj.based2(x,y)
proj.based3(x,y)
part.f.test(x,y,model,full)
Orthogonalization
min.var(af,x,y,numb,n.i)
```

---

## Libraries

---

```
library(lpSolve)
library(lars)
library(parcor)
library(MASS)
library(flare)
library(stringr)
```

## B.2 Design matrix

---

### Generate 12 run PB matrix

---

```
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)  
}
```

---

### Add interaction columns to the design matrix

---

```
add.inter <- function(x,inter){  
  
  n = dim(x)[1]  
  k = dim(x)[2]  
  s2 = 0; s3 = 0  
  
  numb.of.col = k+k*(k-1)/2+as.numeric(I(inter==3))  
  *k*(k-1)*(k-2)/6 # Number of columns  
  header = rep(0,numb.of.col) # Headers  
  ex.mat = matrix(0,n,numb.of.col) # Extended design matrix  
  
  for (i in 1:k){  
    ex.mat[,1:k] = x # Design matrix  
    header[i]= paste(letters[i]) # Make header  
    for (j in (i+1):k){  
      if (i != j & i != k){  
        s2 = s2 + 1  
        ex.mat[,k+s2] = x[,i]*x[,j] # Add 2-factor inter. columns  
        header[k+s2] = paste(letters[i],letters[j],sep="") # Make header  
      }  
      for (l in (j+1):k){  
        if (i!=j & i!=l & j!=l & i!=k & i!=(k-1) &  
            j!=k & inter==3){ # If inter == 3  
          s3 = s3 + 1  
          ex.mat[,k+k*(k-1)/2+s3] = x[,i]*x[,j]*x[,l] # Add 3-factor inter.
```

```

        header[k+k*(k-1)/2+s3] = paste(letters[i],
                                       letters[j],letters[l],sep="")}}}} # Add header

colnames(ex.mat) = toupper(header)
return(ex.mat)
}

```

---

Make header for the design matrix

---

```

make.headers <- function(ff,inter){

  x = design.matrix()
  n.x = colnames(x) # Design matrix headers

  h = vector()
  t = paste(ff,collapse="") # All factors
  u = unlist(strsplit(t,"")) # Split interactions into factors
  for (i in 1:11){
    if (length(which(n.x[i]==u))>0){h = c(h,i)} # Identify factors
  }

  nc = length(h)
  header = 0; s2 = 0; s3 = 0
  for (i in 1:nc){
    header[i]= paste(letters[h[i]]) # Make header
    for (j in (i+1):nc){
      if (i != j & i != nc){
        s2 = s2 + 1
        header[nc+s2] = paste(letters[h[i]],
                              letters[h[j]],sep="") # Add header 2-factor inter.
      }
      for (l in (j+1):nc){
        if (i!=j & i!=l & j!=l & i!=nc &
            i!=(nc-1) & j!=nc & inter==3){ # If inter == 3
          s3 = s3 + 1
          header[nc+nc*(nc-1)/2+s3] =
            paste(letters[h[i]],
                  letters[h[j]],letters[h[l]],sep="")}}}} # Add header 3-factor inter.
        header = toupper(header) # Uppercase letters
      }
    }
  }
  return(header)
}

```



## B.3 Generation of data

---

### Models used in the examples

---

```
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)
```

```
s1 = sqrt(0.1)
s5 = sqrt(0.5)
```

```
for (i in 1:12){y1.01[i] = 2 + A[i] + rnorm(1,0,s1)}
for (i in 1:12){y2.01[i] = 2 + 2*A[i] + B[i] + rnorm(1,0,s1)}
for (i in 1:12){y3.01[i] = 2 + A[i] + B[i] + rnorm(1,0,s1)}
for (i in 1:12){y4.01[i] = 2 + 3*A[i] + 2*B[i] + A[i]*B[i] + rnorm(1,0,s1)}
for (i in 1:12){y5.01[i] = 2 + 2*A[i] + 2*B[i] + A[i]*B[i] + rnorm(1,0,s1)}
for (i in 1:12){y6.01[i] = 2 + 2*A[i] + B[i] + A[i]*B[i] + rnorm(1,0,s1)}
```

```
for (i in 1:12){y7.01[i] = 2 + A[i] + 2*A[i]*B[i] + 2*A[i]*C[i] + rnorm(1,0,s1)}
for (i in 1:12){y8.01[i] = 2 + A[i] + A[i]*B[i]+2*A[i]*C[i] + rnorm(1,0,s1)}
for (i in 1:12){y9.01[i] = 2 + A[i] + 1.5*B[i] + 2*C[i] + A[i]*B[i]+1.5*A[i]*C[i] +
  rnorm(1,0,s1)}
for (i in 1:12){y10.01[i] = 2 + A[i] + 1.5*B[i] + 2*C[i] + 1.5*A[i]*B[i]*C[i] +
  rnorm(1,0,s1)}
for (i in 1:12){y11.01[i] = 2 + 2*A[i] + B[i]*C[i] + rnorm(1,0,s1)}
for (i in 1:12){y12.01[i] = 2 + A[i] + C[i] + B[i]*C[i]+C[i]*D[i] + rnorm(1,0,s1)}
for (i in 1:12){y13.01[i] = 2 + 2*A[i] + 3*B[i] + 2*C[i] + D[i] + 3*C[i]*D[i] +
  rnorm(1,0,s1)}
for (i in 1:12){y14.01[i] = 2 + 4*A[i] + B[i] + C[i] + D[i] + 2*A[i]*D[i] +
  rnorm(1,0,s1)}
for (i in 1:12){y15.01[i] = 2 + 2*A[i] + 4*C[i] + 2*B[i]*C[i] + 2*C[i]*D[i] +
  rnorm(1,0,s1)}
```

```
for (i in 1:12){y1.05[i] = 2 + A[i] + rnorm(1,0,s5)}
for (i in 1:12){y2.05[i] = 2 + 2*A[i] + B[i] + rnorm(1,0,s5)}
for (i in 1:12){y3.05[i] = 2 + A[i] + B[i] + rnorm(1,0,s5)}
for (i in 1:12){y4.05[i] = 2 + 3*A[i] + 2*B[i] + A[i]*B[i] + rnorm(1,0,s5)}
for (i in 1:12){y5.05[i] = 2 + 2*A[i] + 2*B[i] + A[i]*B[i] + rnorm(1,0,s5)}
for (i in 1:12){y6.05[i] = 2 + 2*A[i] + B[i] + A[i]*B[i] + rnorm(1,0,s5)}
```

```
for (i in 1:12){y7.05[i] = 2 + A[i] + 2*A[i]*B[i] + 2*A[i]*C[i] + rnorm(1,0,s5)}
for (i in 1:12){y8.05[i] = 2 + A[i] + A[i]*B[i]+2*A[i]*C[i] + rnorm(1,0,s5)}
for (i in 1:12){y9.05[i] = 2 + A[i] + 1.5*B[i] + 2*C[i] + A[i]*B[i]+1.5*A[i]*C[i] +
  rnorm(1,0,s5)}
```

```

for (i in 1:12){y10.05[i] = 2 + A[i] + 1.5*B[i] + 2*C[i] + 1.5*A[i]*B[i]*C[i] +
                    rnorm(1,0,s5)}
for (i in 1:12){y11.05[i] = 2 + 2*A[i] + B[i]*C[i] + rnorm(1,0,s5)}
for (i in 1:12){y12.05[i] = 2 + A[i] + C[i] + B[i]*C[i]+C[i]*D[i] + rnorm(1,0,s5)}
for (i in 1:12){y13.05[i] = 2 + 2*A[i] + 3*B[i] + 2*C[i] + D[i] + 3*C[i]*D[i] +
                    rnorm(1,0,s5)}
for (i in 1:12){y14.05[i] = 2 + 4*A[i] + B[i] + C[i] + D[i] + 2*A[i]*D[i] +
                    rnorm(1,0,s5)}
for (i in 1:12){y15.05[i] = 2 + 2*A[i] + 4*C[i] + 2*B[i]*C[i] + 2*C[i]*D[i] +
                    rnorm(1,0,s5)}

```

```

ymat = cbind(y1.01,y2.01,y3.01,y4.01,y5.01,y6.01,y7.01,y8.01,y9.01,y10.01,
y11.01,y12.01,y13.01,y14.01,y15.01,y1.05,y2.05,y3.05,y4.05,y5.05,y6.05,y7.05,
y8.05,y9.05,y10.05,y11.05,y12.05,y13.05,y14.05,y15.05)

```

```

save(y.mat,file="C:/R filer/SaveRespons.RData")
load(file="C:/R filer/SaveRespons.RData")

```

## B.4 Methods using the Dantzig selector

---

### Dantzig selector

---

```
DS <-function(x,y,sigma){

  n = dim(x)[1]           # Rows in design matrix
  k = dim(x)[2]           # Columns in design matrix
  z = (y-mean(y))/sqrt(var(y)) # Standardize data

  delta = sigma*sqrt(2*log(k))

  c = c(rep(1,k), rep(0,k)) # Vector c
  A = cbind(rbind(t(x)%*%x,-t(x)%*%x,2*diag(1,k)), # Matrix A
            rbind(-t(x)%*%x,t(x)%*%x,-diag(1,k)))

  beta = rep(0,k)         # Make vector for the beta values
  b = c(-t(x)%*%y-delta*rep(1,k),
        t(x)%*%y-delta*rep(1,k), rep(0,k)) # Make vector b
  x.ub = lp("min", c(rep(1,k),rep(0,k)),
            A,rep(">=",length(b)),b)$solution

  for (j in 1:k){beta[j] = x.ub[k+j]-x.ub[j]} # Calculate estimated beta
  names(beta) = colnames(x)
  beta = beta[beta != 0] # All non zero beta values
  part = round(beta[order(abs(beta), # Sort, round and select six first
decreasing=TRUE)][1:6],digits=2)
  return(part)
}
```

---

### Graphical Dantzig selector

---

```
DS.plot <- function(x,y){

  n = dim(x)[1]           # Rows in design matrix
  k = dim(x)[2]           # Columns in design matrix
  z = (y-mean(y))/sqrt(var(y)) # Standardize data
  #x = x/sqrt(sum(abs(x[,1])))

  delta_0 = t(x[,1])%*%y # Make vector with  $X^t y$ 
  for (i in 2:k){delta_0 = c(delta_0,t(x[,i])%*%y)}
  delta_0 = max(abs(delta_0)) #  $\Delta_0 = \max |x_i^t y|$ 
  delta = seq(0,ceiling(delta_0),length.out=120) # Make delta values
  l = length(delta)       # Number of delta values

  c = c(rep(1,k), rep(0,k)) # Vector c
  A = cbind(rbind(t(x)%*%x,-t(x)%*%x,2*diag(1,k)), # Matrix A
            rbind(-t(x)%*%x,t(x)%*%x,-diag(1,k)))
```

```

sol.matrix = matrix(0,length(delta),k)           # Make matrix to store the solutions
beta = rep(0,k)                                 # Make vector for the beta values

for (i in 1:l){
  b = c(-t(x)%*%y-delta[i]*rep(1,k),
        t(x)%*%y-delta[i]*rep(1,k), rep(0,k)) # Make vector b
  x.ub = lp("min", c(rep(1,k),rep(0,k)),A,
           rep(">=",length(b)),b)$solution
  for (j in 1:k){
    beta[j] = x.ub[k+j]-x.ub[j]                # Calculate estimated beta
  }
  sol.matrix[i,] = beta                        # Store the solution in sol.matrix
}

ymax = max(abs(sol.matrix))*1.1
ymin = min(sol.matrix)*1.1
col<-rgb(runif(k),runif(k),runif(k))
plot(delta,sol.matrix[,1],type="l",col=col[1],
      xlab="Delta", ylab="Beta",
      xlim=c(-1,ceiling(delta_0)),ylim=c(ymin,ymax)) # Plot solution for factor A
for (i in 2:k){lines(delta,sol.matrix[,i],
                    col=col[i])}                # Plot the rest of the factors
lines(c(0,ceiling(delta_0)),c(0,0))

longest = rep(0,k)
for (i in 1:k){
  if (max(abs(sol.matrix[,i]))==0){longest[i] = 0} # Maximal beta value
  if (max(abs(sol.matrix[,i]))>0){longest[i] =
    max(which(sol.matrix[,i] != 0))}             # Maximal beta value
ltr = order(longest,decreasing=T)[1:6]          # Sort factors
ltr = ltr[which(longest[ltr] != 0)]
ltr = ltr[!is.na(ltr)]

header = colnames(x)                            # Make headers
t = length(ltr); lab = rep(0,t)                 # Initial values

for (i in 1:t){
  pos.max = longest[ltr[i]]
  lab[i] = header[ltr[i]]
  if (pos.max > 8){pos.max=pos.max-8}
  text(delta[pos.max],sol.matrix[pos.max,ltr[i]],
        labels=lab[i],cex=0.7)                 # Write names on plot
}

colnames(sol.matrix) = colnames(x)
return(lab)
}

```

Support function for the projection based Dantzig selector

```

DS.sub <- function(x,y,ff,inter){

  nc = length(ff)
  n = dim(x)[1] # Rows in design matrix
  k = dim(x)[2] # Columns in design matrix
  z = (y-mean(y))/sqrt(var(y)) # Standardize data

  delta_0 = t(x[,1])%*%y # Make vector with X^t*y
  for (i in 2:k){delta_0 = c(delta_0,t(x[,i])%*%y)}
  delta_0 = max(abs(delta_0)) # Delta_0 = max|x_i^t*y|
  delta = seq(0,ceiling(delta_0),length.out=120) # Make delta values
  l = length(delta) # Number of delta values

  c = c(rep(1,k), rep(0,k)) # Vector c
  A = cbind(rbind(t(x)%*%x,-t(x)%*%x,2*diag(1,k)), # Matrix A
            rbind(-t(x)%*%x,t(x)%*%x,-diag(1,k)))

  sol.matrix = matrix(0,length(delta),k) # Make matrix to store the solutions
  beta = rep(0,k) # Make vector for the beta values

  for (i in 1:l){
    b = c(-t(x)%*%y-delta[i]*rep(1,k), # Make vector b
          t(x)%*%y-delta[i]*rep(1,k), rep(0,k))
    x.ub = lp("min", c(rep(1,k),rep(0,k)), # Solve linear problem
             A,rep(">=",length(b)),b)$solution
    for (j in 1:k){
      beta[j] = x.ub[k+j]-x.ub[j]} # Calculate estimated beta
    sol.matrix[i,] = beta} # Store the solution in sol.matrix

  longest = rep(0,k)
  for (i in 1:k){
    if (max(abs(sol.matrix[,i]))==0){longest[i]=0} # Find maximal value
    if (max(abs(sol.matrix[,i]))>0){
      longest[i]=max(which(sol.matrix[,i] != 0))}}
  ltr = order(longest,decreasing=T)[1:6] # Sort decreasing value
  ltr = ltr[which(longest[ltr] != 0)]
  ltr = ltr[!is.na(ltr)] # Remove na

  header = make.headers(ff,inter) # Make headers
  t = length(ltr); lab = rep(0,t)

  for (i in 1:t){lab[i] = header[ltr[i]]} # Terms in final model
  return(lab)
}

```

---

## Projection based Dantzig selector

---

```

DS.projection <- function(x,y,inter){

  k = dim(x)[2]
  rows = factorial(k)/(factorial(k-4)*factorial(4))           # Number of rows
  var.mat = cbind(seq(1,rows),rep(0,rows))                    # Initial values
  fac.mat = matrix(0,rows,6)                                  # Initial values
  t = 0; v.min = Inf; res = rep(0,1)                          # Initial values

  for (i in 1:(k-3)){
    for (j in i:(k-2)){
      for (l in j:(k-1)){
        for (h in l:k){
          if (i!=j & i!=l & i!=h & j!=l & j!=h & l!=h){
            t = t + 1
            hn = c(i,j,l,h)                                   # Four columns
            fh = c(paste(letters[i]),paste(letters[j]),
                  paste(letters[l]),paste(letters[h]))
            fh = toupper(fh)

            xdm = cbind(x[,i],x[,j],x[,l],x[,h])
            xdm2 = add.inter(xdm,inter)                         # Add interactions
            ds = DS.sub(xdm2,y,fh,inter)                        # Solve using DS.sub

            header = make.headers(fh,inter)                    # Make header
            colnames(xdm2) = header                            # Add header to xdm2
            lh = length(header)
            for (r in 1:lh){assign(header[r],xdm2[,r])}        # Assign headers to columns

            temp = paste(ds[1:length(ds)], collapse="+")      # Terms in the model
            form = paste("y~", temp)
            sig = (summary(lm(formula=
                              form,data.frame(xdm2)))$sigma)**2 # Sigma of fitted model
            var.mat[t,2] = round(sig,digits=3)                 # Save std. dev.
            fac.mat[t,] = ds                                    # Save factors
          }}}}}

  ink = var.mat[sort.list(var.mat[,2]),]                       # Sort matrix
  ink10 = ink[1:10,]                                          # Pick out ten first
  factors = fac.mat[ink10[,1],]                               # Save ten first factors
  return(cbind(ink10[,2],fac.mat[ink10[,1],]))
}

```

---

## Reduce model according to p-value

---

```

reduce <- function(x,y,ff,inter){

  x.2 = add.inter(x,inter)                                    # Expand matrix
  l = 0; p.val = 0.05

```

```

temp = paste(ff[1:length(ff)], collapse="+")
form = paste("y~", temp) # Fit model with terms in ff
x2m.f = x.2[,ff[1]]
for (i in 2:length(ff)){
  x2m.f = cbind(x2m.f,x.2[,ff[i]]) # Make matrix with columns of ff
colnames(x2m.f) = ff

mod = lm(formula= form,data.frame(x2m.f),y=TRUE) # Fit model
p = coef(summary(mod))[,4] # Estimated parameter values
while((max(p[2:length(p)])>=p.val) &length(p)>1){
  rem = names(which(p == max(p[2:length(p)])))
  mod = update(mod,as.formula(paste("~-.",
    rem,sep=""))) # Remove term with highest p-value
  p = coef(summary(mod))[,4]
}

name = names(coef(summary(mod))[,4])
res = name[2:length(name)] # Terms in reduced model
return(res)
}

```

---

### Expansion of the projection based Dantzig selector

---

```

DS.var.aic.bic <- function(x,y,a,inter){

k = dim(x)[2]
x.2 = add.inter(x,inter)
rows = factorial(k)/(factorial(k-4)*factorial(4)) # Number of rows
var.mat = cbind(seq(1,rows),rep(0,rows)) # Initial values
fac.mat = matrix(0,rows,6) # Initial values
t = 0; v.min = Inf; res = rep(0,1) # Initial values

for (i in 1:(k-3)){
  for (j in i:(k-2)){
    for (l in j:(k-1)){
      for (h in l:k){
        if (i!=j & i!=l & i!=h & j!=l & j!=h & l!=h){

t = t + 1
hn = c(i,j,l,h) # Four columns
fh = c(paste(letters[i]),paste(letters[j]),
  paste(letters[l]),paste(letters[h]))
fh = toupper(fh)

x2m = cbind(x[,i],x[,j],x[,l],x[,h])
x2m2 = add.inter(x2m,inter) # Add interactions
ds = DS.sub(x2m2,y,fh,inter) # Solve using DS.sub

header = make.headers(fh,inter)
colnames(x2m2) = header # Add header to x2m2
lh = length(header)
for (r in 1:lh){assign(header[r],x2m2[,r])} # Assign headers to columns

```

```

    temp = paste(ds[1:length(ds)], collapse="+") # Terms in the model
    form = paste("y~", temp)
    sig = (summary(lm(formula=
        form,data.frame(xdm2)))$sigma)**2 # Sigma of fitted model
    var.mat[t,2] = round(sig,digits=3) # Save std. dev.
    fac.mat[t,] = ds # Save factors
    }}}}

ink = var.mat[sort.list(var.mat[,2]),] # Sort matrix
ink10 = ink[1:10,] # Pick out ten first
factors = fac.mat[ink10[,1],] # Save ten first factors

for (f in 1:10){
    ff = factors[f,] # Select one model
    d = reduce(x,y,ff,inter) # Reduce model
    temp = paste(d[1:length(d)], collapse="+") # Terms in reduced model
    form = paste("y~", temp)
    xdm.f = cbind(x.2[,ff[1]],x.2[,ff[2]],x.2[,ff[3]],
        x.2[,ff[4]],x.2[,ff[5]],x.2[,ff[6]]) # Matrix of ff
    colnames(xdm.f) = ff
    mod = lm(formula= form,data.frame(xdm.f),y=TRUE) # Fit model

    if (a == 1){v.temp = summary(mod)$sigma} # Variance
    if (a == 2){v.temp = AIC(mod)} # AIC
    if (a == 3){v.temp = BIC(mod)} # BIC

    if (v.temp < v.min){ # If this value is the smallest
        v.min = v.temp
        name = names(coef(summary(mod))[,4]) # Save terms
        res = name[2:length(name)]
    }
}
return(res)
}

```



## B.5 The Lasso

---

### The Lasso

---

```
l.lasso <- function(x,y){

  lassofit = lars(x,y,type="lasso")           # Perform Lasso for main effects
  lambda = mylars(x,y)$lambda.opt           # Optimal variable selection value
  l.b = coef(lassofit,s=lambda,mode="lambda") # Estimates
  print(round(l.b[order(abs(l.b),
    decreasing=TRUE)][1:6],digits=2))       # Sort parameters
  plot.lars(lassofit)                       # Plot

  x.2 = add.inter(x,2)
  lassofit = lars(x.2,y,type="lasso")       # Perform Lasso with two-fac. inter.
  lambda = mylars(x.2,y)$lambda.opt         # Optimal variable selection value
  l.b = coef(lassofit,s=lambda,mode="lambda") # Estimates
  print(round(l.b[order(abs(l.b),
    decreasing=TRUE)][1:6],digits=2))       # Sort parameters
  plot.lars(lassofit)                       # Plot

  x.3 = add.inter(x,3)
  lassofit = lars(x.3,y,type="lasso")       # Perform Lasso with two-fac. inter.
  lambda = mylars(x.3,y)$lambda.opt         # Optimal variable selection value
  l.b = coef(lassofit,s=lambda,mode="lambda") # Estimates
  print(round(l.b[order(abs(l.b),
    decreasing=TRUE)][1:6],digits=2))       # Sort parameters
  plot.lars(lassofit)                       # Plot
}
```

## B.6 Projection based method

---

### Find equal rows for one active factor

---

```
equal.rows1 <- function(x){
  eq.mat = matrix(0,11,12)           # Matrix for saving equal rows
  c1 = 0

  for (i in 1:11){
    fac = x[,i]                       # Columns with row numbers
    c1 = c1+1
    eq.mat[c1,1:6] = which(fac==1)
    eq.mat[c1,7:12] = which(fac==-1)}
  return(eq.mat)
}
```

---

### Find equal rows two active factors

---

```
equal.rows2 <- function(x){
  eq.mat = matrix(0,55,12)           # Matrix for saving equal rows
  c1 = 0

  for (i in 1:10){
    for (j in (i+1):11){
      if (i != j){

        fac = cbind(x[,c(i,j)],seq(1:12)) # Columns with row numbers
        c1 = c1+1; c2 = 0

        while (eq.mat[c1,12] == 0){
          eq = which(rowSums(abs(t(
            apply(-fac[,1:2],1,'+',fac[1,1:2]))) == 0) # Search for equal row
          eq.mat[c1,(1+3*c2):(3+3*c2)] = fac[eq,3] # Insert equal rows into eq.mat
          c2 = c2+1
          fac = fac[-eq,]}}}}
  return(eq.mat)
}
```

---

### Find equal rows for three active factors

---

```
equal.rows3 <- function(x){
  eq.mat = matrix(0,165,8)           # Matrix for equal rows
  c1 = 0

  for (i in 1:9){
    for (j in (i+1):10){
      for (k in (j+1):11){
        if (i != j & i != k & j != k){
```

```

fac = cbind(x[,c(i,j,k)],seq(1:12))          # Columns with row numbers
c1 = c1+1; c2 = 0

while (eq.mat[c1,8] == 0){
  eq = which(rowSums(abs(t(
  apply(-fac[,1:3],1,'+',fac[1,1:3]))) == 0) # Search for equal row
  if (length(eq) == 2){
    eq.mat[c1,(1+2*c2):(2+2*c2)] = fac[eq,4] # Insert pairs into eq.mat
    c2 = c2+1
  }
  fac = fac[-eq,]}}}} # Remove rows
return(eq.mat)
}

```

---

### Projection based method one active

---

```

proj.based1 <- function(x,y){
  var = rep(0,11)
  header = rep(0,11)

  for(i in 1:11){

    fac = x[,i]          # One column of the design matrix
    p = which(fac[]==1)  # High level
    m = which(fac[]== -1) # Low level

    var[i] = (sum((y[p]-mean(y[p]))^2)+ # Estimate variance:
              sum((y[m]-mean(y[m]))^2))/10

    header[i] = paste(letters[i]) # Letters for headers
    names(var) = header          # Insert headers
    part = sort(round(var,digits=2)) # Sort
    return(part)
  }
}

```

---

### Projection based method two active

---

```

proj.based2 <- function(x,y){
  var = rep(0,55)
  header = rep(0,55)
  k = 1
  erm = equal.rows2(x) # Find equal rows in design matrix

  for (i in 1:10){
    for (j in (i+1):11){
      if (i != j){

        fac = x[,c(i,j)] # Two columns of the design matrix
        p = matrix(erm[k,],4,3,byrow=TRUE) # All pairs in fac

```

```

    sum.mi = 0
    for (l in 1:4){sum.mi = sum.mi +
    sum((y[p[l,]]-mean(y[p[l,]]))^2)}      # Sum variances
    var[k] = sum.mi/8                      # Estimated variance

    header[k]=paste(letters[i],letters[j],sep="")
    k = k+1}}}
names(var) = header                      # Insert headers
part = sort(round(var,digits=2))          # Sort
return(part)
}

```

---

### Projection based method three active

---

```

proj.based3 <- function(x,y){
  var = rep(0,165)
  header = rep(0,165)
  l = 1
  erm = equal.rows3(x)                   # Equal rows in the design matrix

  for (i in 1:9){
    for (j in (i+1):10){
      for (k in (j+1):11){
        if (i != j & i != k & j != k){

          fac = x[,c(i,j,k)]              # Three columns of the design matrix
          p = matrix(erm[l,],4,2,byrow=TRUE) # All pairs in fac

          sum.mi = 0
          for (m in 1:4){sum.mi = sum.mi +
          sum((y[p[m,]]-mean(y[p[m,]]))^2)} # Sum variances
          var[l] = sum.mi/4

          header[l] = paste(letters[i],letters[j],letters[k],sep="")
          l = l+1}}}}
names(var) = header                      # Insert headers
part = sort(round(var,digits=2))          # Sort
return(part)
}

```

## B.7 Partial F-test

---

### Partial F-test

---

```
part.f.test <- function(x,y,model,full){  
  
  x2 = add.inter(x,2) # Expand matrix  
  
  header = colnames(x2)  
  lh = length(header)  
  for (r in 1:lh){assign(header[r],x2[,r])}  
  
  temp.m = paste(model[1:length(model)], collapse="+") # Terms in reduced model  
  form.m = paste("y~", temp.m) # Make form of the model  
  fit.mod = lm(formula = form.m,data.frame(x2)) # Fit reduced model  
  
  temp.f = paste(full[1:length(full)], collapse="+") # Terms in full model  
  form.f = paste("y~", temp.f) # Make form of the model  
  fit.full = lm(formula = form.f,data.frame(x2)) # Fitt full model  
  
  p.val = anova(fit.mod, fit.full,test="F")$P[2] # Partial F-test  
  if (identical(model,full)){p.val = 0}  
  if (p.val < 0.01){ # Note the result of the test  
    model = full  
  }  
  return(model)  
}
```

## B.8 Orthogonalization

---

### Orthogonalization

---

```
----- Preparation -----

x1 = design.matrix()           # Design matrix
x2 = add.inter(x1,2)           # With 2.fac-inter
x3 = add.inter(x1,3)           # With 3.fac-inter
header = colnames(x3)         # Header
lh = length(header)
for (r in 1:lh){assign(header[r],x3[,r])} # Assign values to header

label.m = colnames(x1)
label.i = colnames(x2)
label.t = colnames(x3)

----- Before orthogonalization -----

y = y.mat[,7]                  # Response
eff = abs((1/6)*t(x1)%*%y)     # Main contrasts
eff.int = abs((1/6)*t(x2)%*%y) # Inter. contrasts
eff.int.3 = abs((1/6)*t(x3)%*%y) # 3-fac. inter. contrasts

plot(eff, main="Contrast plot", xlab="Factor", # Plot contrasts
     ylab="| Contrast |", col="red", pch = 19,
     ylim=c(0,max(eff) + 0.5))
text(eff,label.m,cex=0.6,pos=3) # Names on plot

----- Two-factor interactions, names and plot -----

interA = seq(12,21)           # Interactions
interB = c(12,seq(22,30))
interC = c(13,22,seq(31,38))
interD = c(14,23,31,seq(39,45))
interE = c(15,24,32,39,seq(46,51))
interF = c(16,25,33,40,46,seq(52,56))
interG = c(17,26,34,41,47,52,seq(57,60))
interH = c(18,27,35,42,48,53,57,seq(61,63))
interI = c(19,28,36,43,49,54,58,61,seq(64,65))
interJ = c(20,29,37,44,50,55,59,62,64,66)
interK = c(21,30,38,45,51,56,60,63,65,66)

interX = interA                # Inter. of interest
plot(eff.int[interX],main="Contrast plot",xlab="Interaction", # Plot
     ylab="| Contrast |",col= "red", pch = 19,
```

```

        ylim=c(0,max(eff.int[interX])+0.5))
text(eff.int[interX],label.i[interX],cex=0.6,pos=3)           # Names on plot

----- Three-factor interactions, names and plot -----

interAB = seq(67,75)
interAC = c(67,seq(76,83))
interBC = c(67,seq(112,119))

interX = interAB                                             # Inter. of interest
plot(eff.int.3[interX],main="Contrast plot",xlab="Interaction", # Plot
      ylab="| Contrast |",col= "red", pch = 19,
      ylim=c(0,max(eff.int.3[interX])+0.5))
text(eff.int.3[interX],label.t[interX],cex=0.6,pos=3)       # Names on plot

----- Orthogonalization main effects -----

x1.1 = cbind(AB,AC)                                         # X_1
x2.1 = cbind(A,B,C,D,E,F)                                   # X_2
a.mat.1 = solve((t(x1.1)%*%x1.1))%*%(t(x1.1)%*%x2.1)       # A matrix
x2.o1 = x2.1-x1.1%*%a.mat.1                                 # X_2.1
eff.o1 = solve((t(x2.o1)%*%x2.o1))%*%(t(x2.o1)%*%y)        # Contrasts

x1.2 = cbind(A,AB,AC)                                       # X_1
x2.2 = cbind(G,H,I,J,K)                                     # X_2
a.mat.2 = solve((t(x1.2)%*%x1.2))%*%(t(x1.2)%*%x2.2)       # A matrix
x2.o2 = x2.2-x1.2%*%a.mat.2                                 # X_2.1
eff.o2 = solve((t(x2.o2)%*%x2.o2))%*%(t(x2.o2)%*%y)        # Contrasts

eff.x1 = solve((t(x1.2)%*%x1.2))%*%(t(x1.2)%*%y)-
  a.mat.2%*%eff.o2                                         # Contrasts in X_1

effects = abs(c(eff.x1[1],eff.o1[2:5],eff.o2))              # Gather contrasts
label = label.m
plot(effects,main="Contrast plot after orthogonalization",   # Plot
      xlab="Factor",ylab="| Contrast |",col= "red",
      pch = 19,ylim=c(0,max(effects)+0.5))
text(effects,label,cex=0.6,pos=3)                          # Names on plot

----- Orthogonalization interactions -----

x1.1 = cbind(AB,AC)                                         # X_1
x2.1 = cbind(AD,AE,AF,AG)                                   # X_2
a.mat.1 = solve((t(x1.1)%*%x1.1))%*%(t(x1.1)%*%x2.1)       # A matrix
x2.o1 = x2.1-x1.1%*%a.mat.1                                 # X_2.1
eff.o1 = solve((t(x2.o1)%*%x2.o1))%*%(t(x2.o1)%*%y)        # Contrasts

```

```

x1.2 = cbind(AB,AC) # X_1
x2.2 = cbind(AH,AI,AJ,AK) # X_2
a.mat.2 = solve((t(x1.2)%*%x1.2))%*%(t(x1.2)%*%x2.2) # A matrix
x2.o2 = x2.2-x1.2%*%a.mat.2 # X_2.1
eff.o2 = solve((t(x2.o2)%*%x2.o2))%*%(t(x2.o2)%*%y) # Contrasts

eff.x1 = solve((t(x1.2)%*%x1.2))%*%(t(x1.2)%*%y)-
  a.mat.2%*%eff.o2 # Contrasts in X_1

inter.o2 = abs(c(eff.x1[1],eff.x1[2],eff.o1,eff.o2)) # Gather contrasts
label = label.i[interA] # Names
plot(inter.o2,main="Contrast plot after orthogonalization", # Plot
      xlab="Factor",ylab="| Contrast |",col= "red",pch = 19,
      ylim=c(0,max(inter.o2)+0.5))
text(inter.o2,label,cex=0.6,pos=3) # Names on plot

```



## B.9 Fit models with some factors known

---

### Fit models with some factors known

---

```
min.var <- function(af,x,y,numb,n.i){

  n.x = colnames(x)                # Names of factors
  x2 = add.inter(x,2)              # Expand design matrix

  sig = rep(Inf,5)
  h = vector()                    # Initial values

  for (r in 1:11){
    if (length(which(n.x[r]==af))>0){h = c(h,r)}} # Identify factors

  if (length(af) == 1 & numb == 4 & n.i == 1){
    res = matrix(Inf,5,5)
    for (i in 1:8){
      for (j in i:9){
        for (l in j:10){
          if (i!=j & i!=1 & j!=1){
            x.min = x1[,-h]        # DM without af factor
            label = colnames(x.min)
            ff = c(af,label[i],label[j],label[l]) # Combination of factors

            m = make.headers(ff,2)[5:10] # Find interactions
            for (n in 1:6){
              ffi = c(ff,m[n])      # Add interaction
              temp = paste(ffi[1:length(ffi)], # Make form of the model
                collapse="+")
              form = paste("y~", temp)
              mod = lm(formula= form,    # Fit model
                data.frame(x2[,ffi]),y=TRUE)
              v.temp = summary(mod)$sigma # Find variance

              if (v.temp < sig[5]){ # If small variance
                sig[5] = round(v.temp,digits=3) # Update variance
                res[5,] = ffi # Insert effects
                place = order(sig,decreasing=FALSE) # Smallest variances
                sig = sort(sig) # Sort sig
                res = res[place,] # Update res
              }}}}}}

  if (length(af) == 2 & numb == 4 & n.i == 1){
    res = matrix(Inf,5,5)
    for (i in 1:8){
      for (j in i:9){
        if (i!=j){
          x.min = x1[,-h] # DM without factors
          label = colnames(x.min)
```

```

ff = c(af,label[i],label[j])           # Combination of factors

m = make.headers(ff,2)[5:10]           # Find interactions
for (l in 1:6){
  ffi = c(ff,m[l])                     # Add interaction
  temp = paste(ffi[1:length(ffi)],     # Make form of the model
              collapse="+")
  form = paste("y~", temp)
  mod = lm(formula= form,               # Fit model
            data.frame(x2[,ffi]),y=TRUE)
  v.temp = summary(mod)$sigma          # Find variance

  if (v.temp < sig[5]){                # If small variance
    sig[5] = round(v.temp,digits=3)    # Update variance
    res[5,] = ffi                      # Insert effects
    place = order(sig,decreasing=FALSE) # Smallest variances
    sig = sort(sig)                    # Sort sig
    res = res[place,]                  # Update res
  }}}}}

if (length(af) == 1 & numb == 3 & n.i == 2){
  res = matrix(Inf,5,5)
  for (i in 1:9){
    for (j in i:10){
      if (i!=j){
        x.min = x1[,-h]                 # DM without factors
        label = colnames(x.min)
        ff = c(af,label[i],label[j])   # Combination of factors

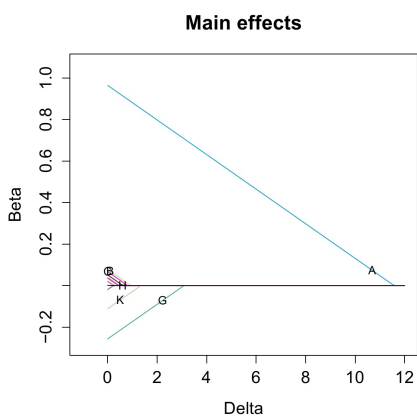
        m = make.headers(ff,2)[4:6]    # Find interactions
        m.d = c(m[1],m[2],m[1],m[3],m[2],m[3]) # Interactions
        for (n in 1:3){
          ffi = c(ff,m.d[2*n-1],m.d[2*n]) # Add interactions
          temp = paste(ffi[1:length(ffi)], # Make form of the model
                      collapse="+")
          form = paste("y~", temp)
          mod = lm(formula= form,         # Fit model
                  data.frame(x2[,ffi]),y=TRUE)
          v.temp = summary(mod)$sigma    # Find variance

          if (v.temp < sig[5]){          # If small variance
            sig[5] = round(v.temp,digits=3) # Update variance
            res[5,] = ffi                # Insert effects
            place = order(sig,decreasing=FALSE) # Smallest variances
            sig = sort(sig)              # Sort sig
            res = res[place,]           # Update res
          }}}}}
  return(cbind(sig,res))
}

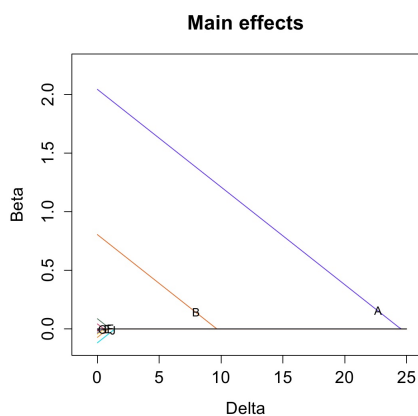
```

# Appendix C

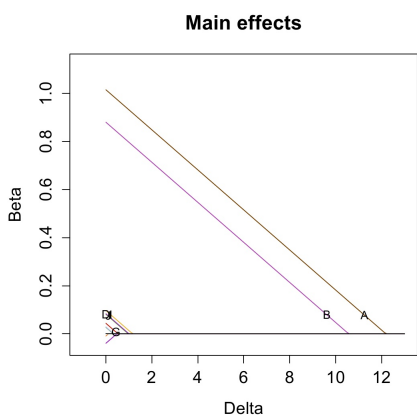
## C.1 Graphical Dantzig selector for data from the models (4.1) - (4.6) generated with $\sigma^2 = 0.1$ when considering main effects.



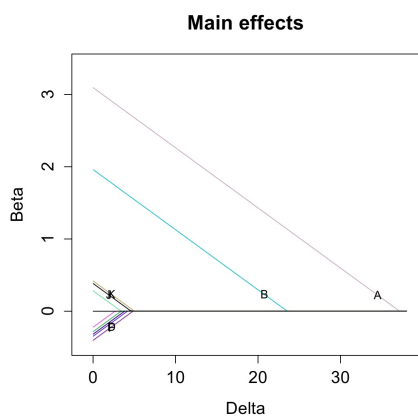
(a) Model (4.1) with  $\sigma^2 = 0.1$



(b) Model (4.2) with  $\sigma^2 = 0.1$

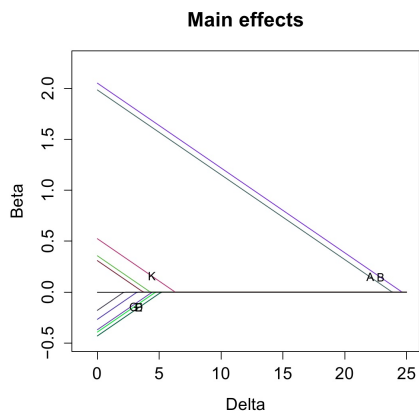


(c) Model (4.3) with  $\sigma^2 = 0.1$

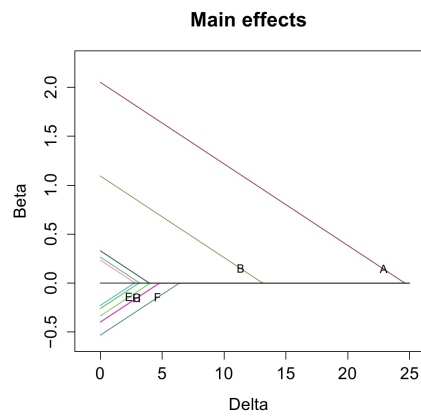


(d) Model (4.4) with  $\sigma^2 = 0.1$

Figure C.1: Plots obtained by the graphical Dantzig selector as introduced in Section 3.2, when considering main effects for data generated from the models (4.1) - (4.6) with  $\sigma^2 = 0.1$ .



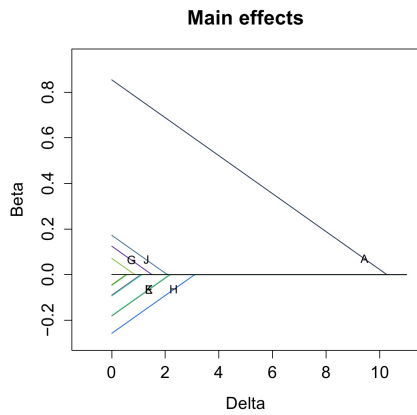
(e) Model (4.5) with  $\sigma^2 = 0.1$



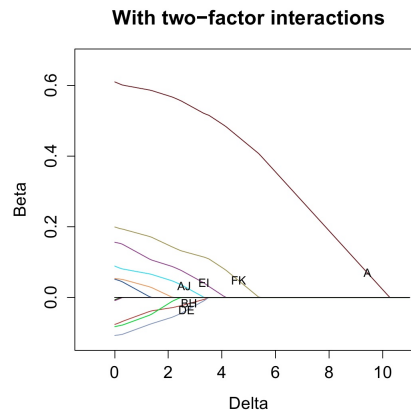
(f) Model (4.6) with  $\sigma^2 = 0.1$

Figure C.1: Plots obtained by the graphical Dantzig selector as introduced in Section 3.2, when considering main effects for data generated from the models (4.1) - (4.6) with  $\sigma^2 = 0.1$ .

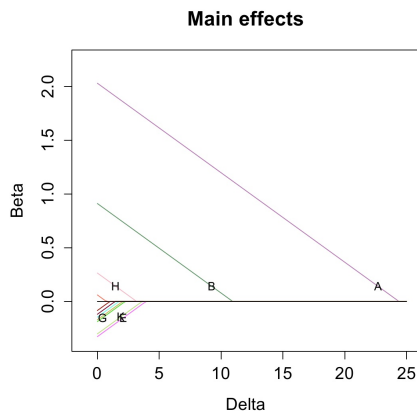
## C.2 Graphical Dantzig selector for data from the models (4.1) - (4.6) generated with $\sigma^2 = 0.5$ .



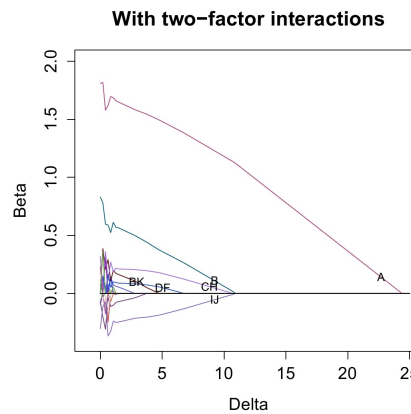
(a) Model (4.1) with  $\sigma^2 = 0.5$



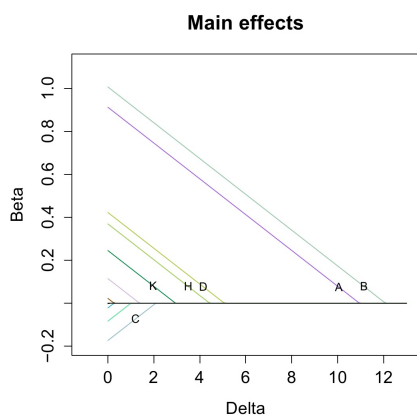
(b) Model (4.1) with  $\sigma^2 = 0.5$



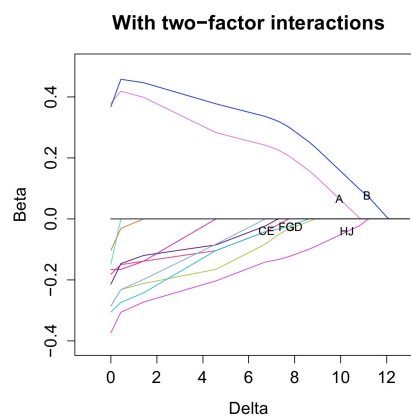
(c) Model (4.2) with  $\sigma^2 = 0.5$



(d) Model (4.2) with  $\sigma^2 = 0.5$

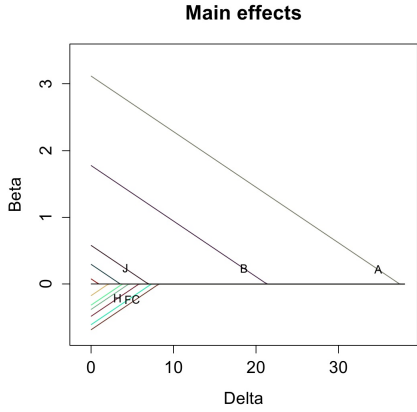


(e) Model (4.3) with  $\sigma^2 = 0.5$

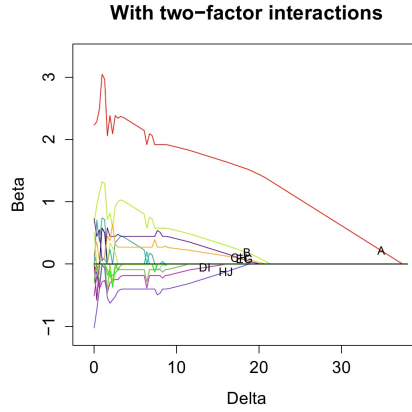


(f) Model (4.3) with  $\sigma^2 = 0.5$

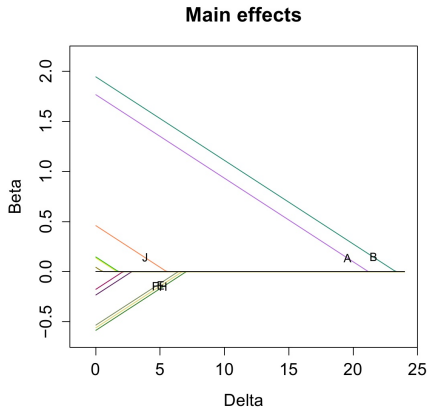
Figure C.2: Plots obtained by the graphical Dantzig selector as introduced in Section 3.2, for data generated from the models (4.1) - (4.6) with  $\sigma^2 = 0.5$ . Plots when considering main effects are given to the left, and when also considering two-factor interaction are given to the right.



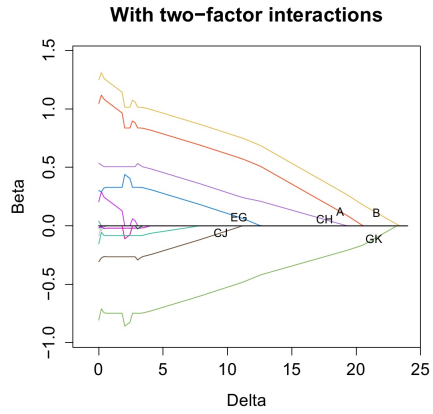
(g) Model (4.4) with  $\sigma^2 = 0.5$



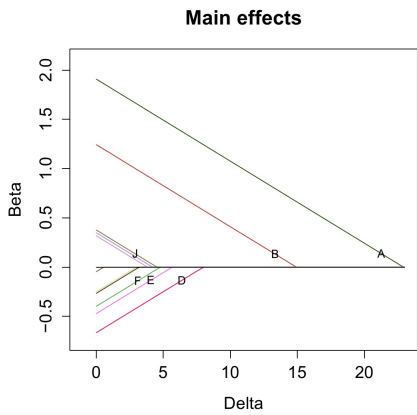
(h) Model (4.4) with  $\sigma^2 = 0.5$



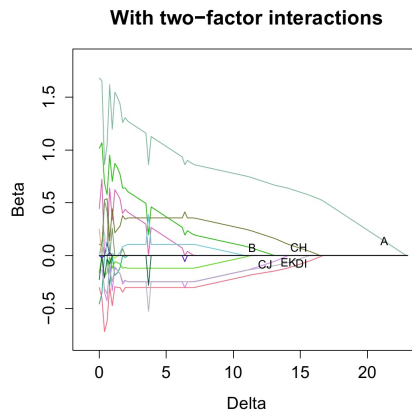
(i) Model (4.5) with  $\sigma^2 = 0.5$



(j) Model (4.5) with  $\sigma^2 = 0.5$



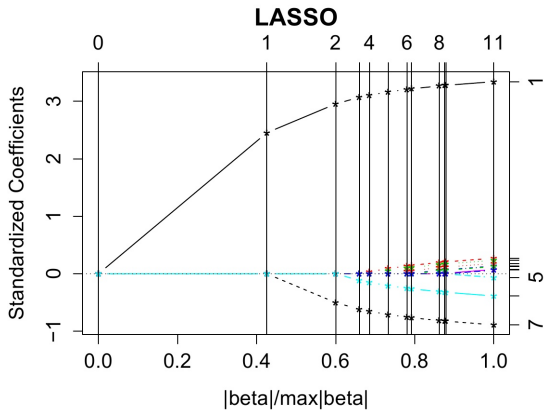
(k) Model (4.6) with  $\sigma^2 = 0.5$



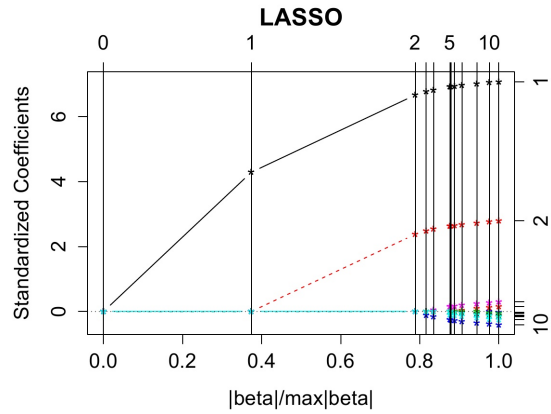
(l) Model (4.6) with  $\sigma^2 = 0.5$

Figure C.2: Plots obtained by the graphical Dantzig selector as introduced in Section 3.2, for data generated from the models (4.1) - (4.6) with  $\sigma^2 = 0.5$ . Plots when considering main effects are given to the left, and when also considering two-factor interaction are given to the right.

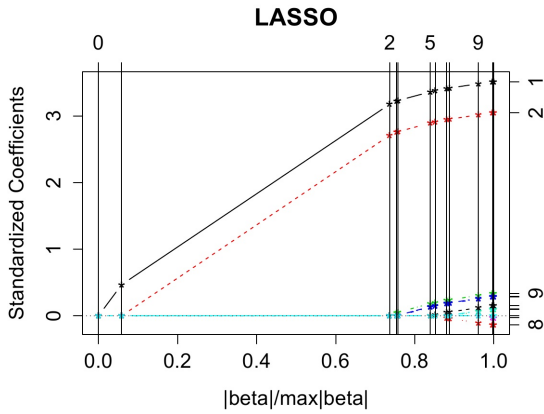
### C.3 Lasso for data generated from the models (4.1) - (4.6) with $\sigma^2 = 0.1$ when considering main effects.



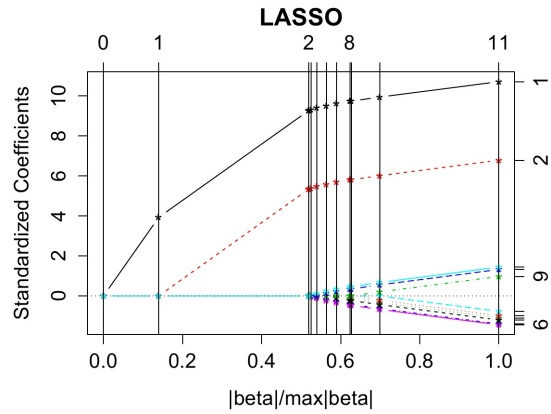
(a) Main effects for Model (4.1) with  $\sigma^2 = 0.1$



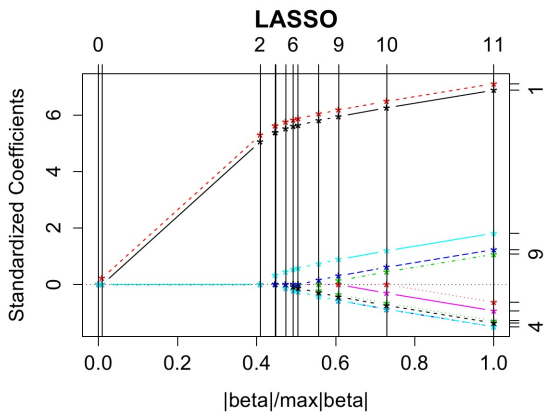
(b) Main effects for Model (4.2) with  $\sigma^2 = 0.1$



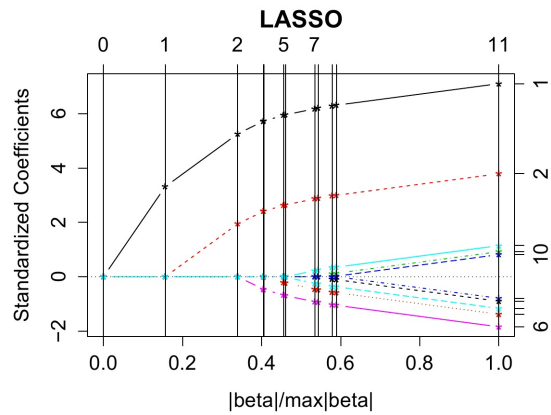
(c) Main effects for Model (4.3) with  $\sigma^2 = 0.1$



(d) Main effects for Model (4.4) with  $\sigma^2 = 0.1$



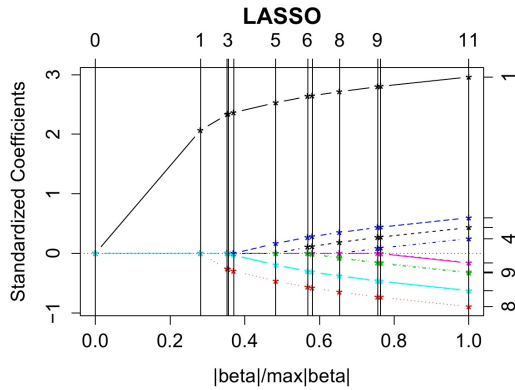
(e) Main effects for Model (4.5) with  $\sigma^2 = 0.1$



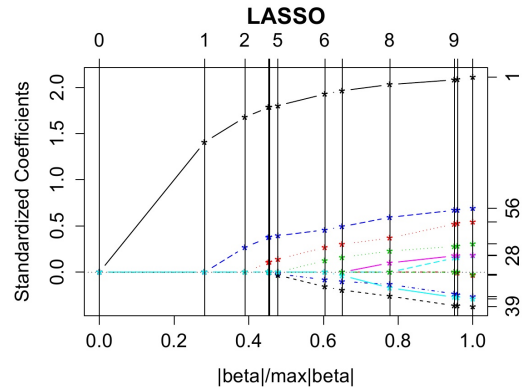
(f) Main effects for Model (4.6) with  $\sigma^2 = 0.1$

Figure C.3: Plots obtained by the Lasso as introduced in Section 3.4, when considering main effects for data generated from the models (4.1) - (4.6) with  $\sigma^2 = 0.1$ .

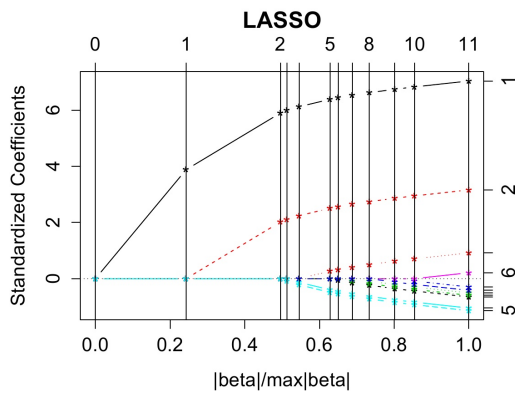
## C.4 Lasso for data from the models (4.1) - (4.6) generated with $\sigma^2 = 0.5$ .



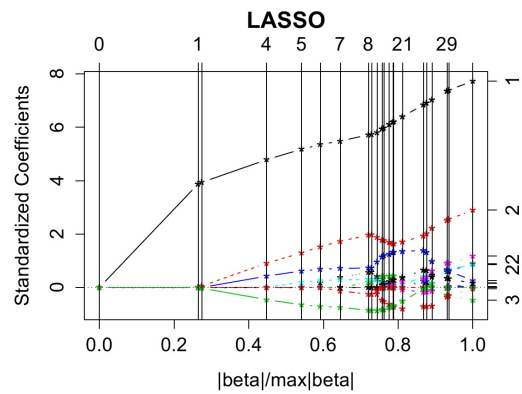
(a) Main effects for Model (4.1) with  $\sigma^2 = 0.1$



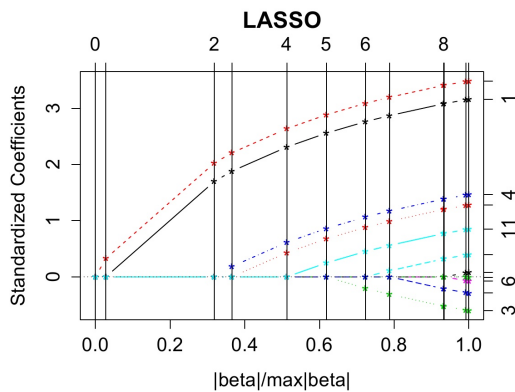
(b) With two-factor interactions for Model (4.1) with  $\sigma^2 = 0.1$



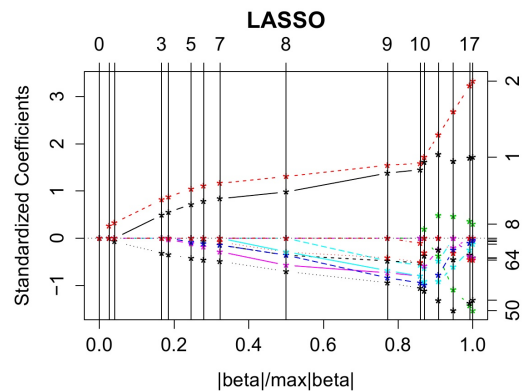
(c) Main effects for Model (4.2) with  $\sigma^2 = 0.1$



(d) With two-factor interactions for Model (4.2) with  $\sigma^2 = 0.1$



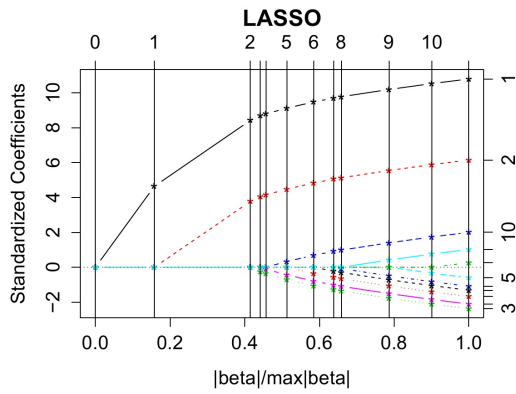
(e) Main effects for Model (4.3) with  $\sigma^2 = 0.1$



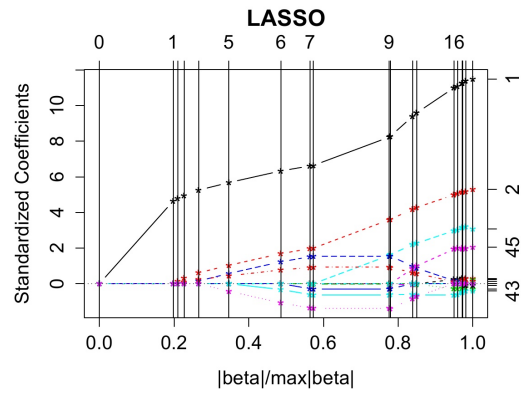
(f) With two-factor interactions for Model (4.3) with  $\sigma^2 = 0.1$

Figure C.4: Plots obtained by the Lasso as introduced in Section 3.4, for data generated from the models (4.1) - (4.6) with  $\sigma^2 = 0.5$ . Plots when considering main effects are given to the left, and when also considering two-factor interaction to the right.

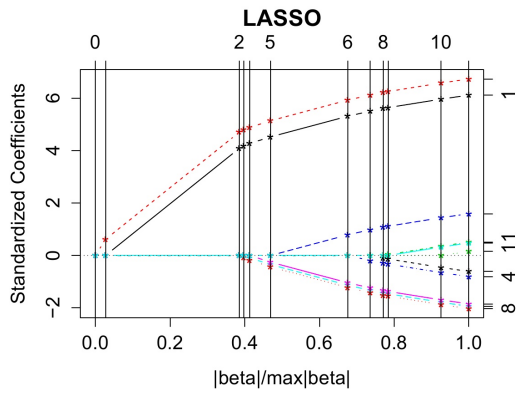




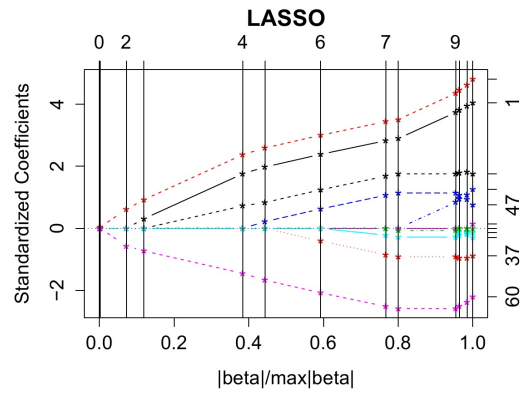
(g) Main effects for Model (4.4) with  $\sigma^2 = 0.1$



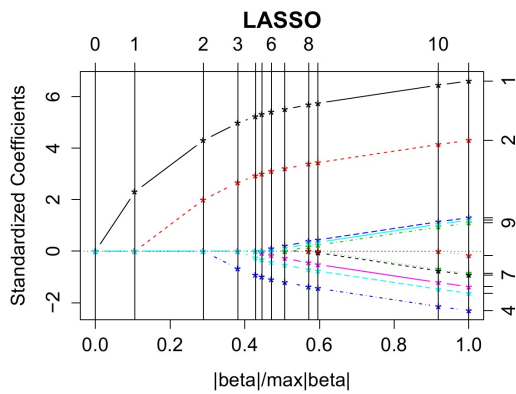
(h) With two-factor interactions for Model (4.4) with  $\sigma^2 = 0.1$



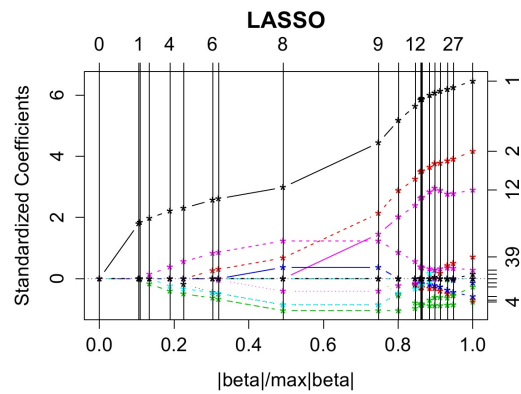
(i) Main effects for Model (4.5) with  $\sigma^2 = 0.1$



(j) With two-factor interactions for Model (4.5) with  $\sigma^2 = 0.1$



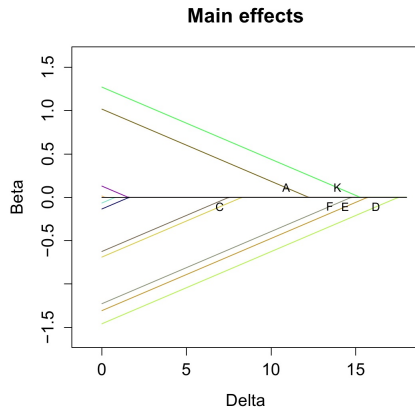
(k) Main effects for Model (4.6) with  $\sigma^2 = 0.1$



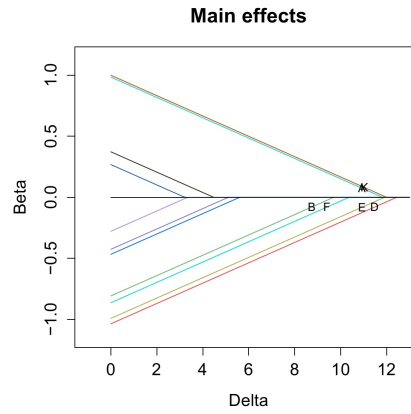
(l) With two-factor interactions for Model (4.6) with  $\sigma^2 = 0.1$

Figure C.4: Plots obtained by the Lasso as introduced in Section 3.4, for data generated from the models (4.1) - (4.6) with  $\sigma^2 = 0.5$ . Plots when considering main effects are given to the left, and when also considering two-factor interaction to the right.

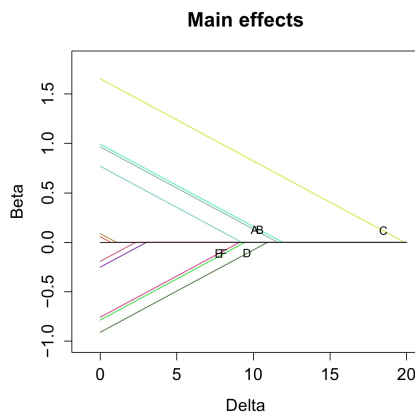
## C.5 Graphical Dantzig selector for data from the models (4.7) - (4.15) generated with $\sigma^2 = 0.1$ when considering main effects.



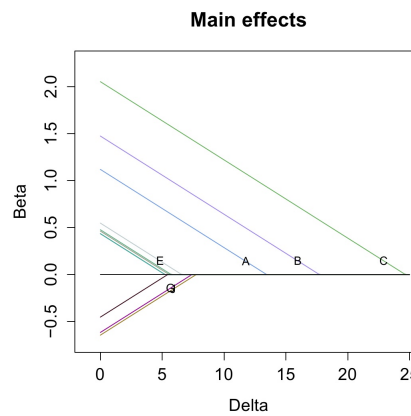
(a) Model (4.7) with  $\sigma^2 = 0.1$



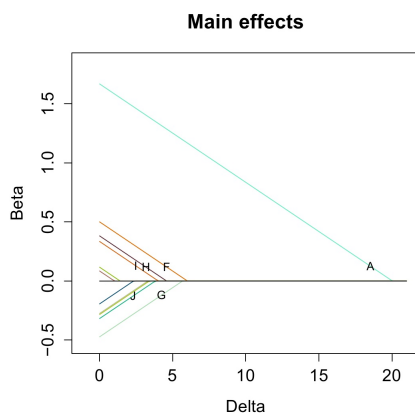
(b) Model (4.8) with  $\sigma^2 = 0.1$



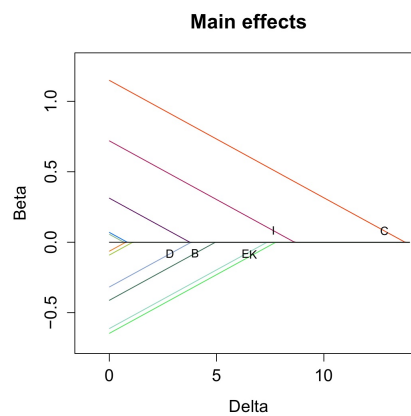
(c) Model (4.9) with  $\sigma^2 = 0.1$



(d) Model (4.10) with  $\sigma^2 = 0.1$

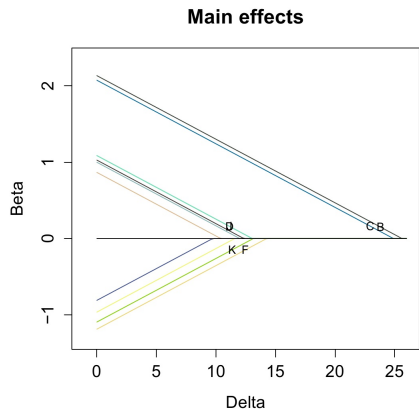


(e) Model (4.11) with  $\sigma^2 = 0.1$

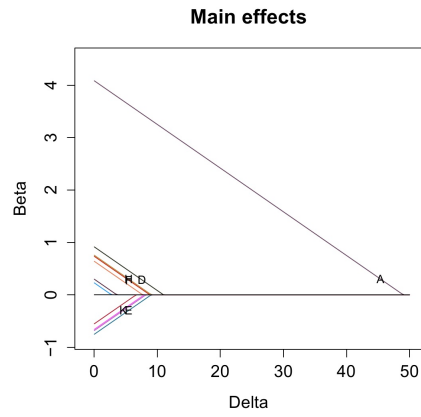


(f) Model (4.12) with  $\sigma^2 = 0.1$

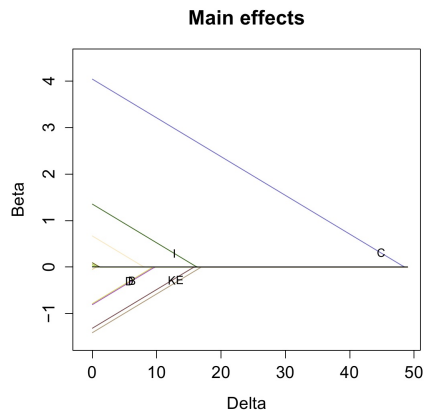
Figure C.5: Plots obtained by the graphical Dantzig selector as introduced in Section 3.2, when considering main effects for data generated from the models (4.7) - (4.15) with  $\sigma^2 = 0.1$ .



(g) Model (4.13) with  $\sigma^2 = 0.1$



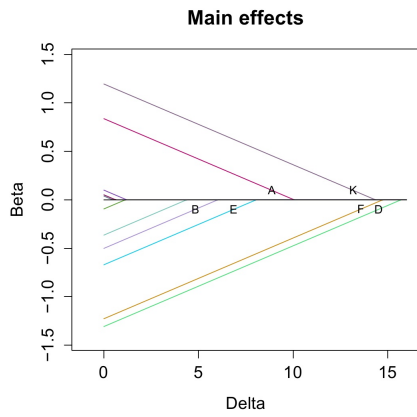
(h) Model (4.14) with  $\sigma^2 = 0.1$



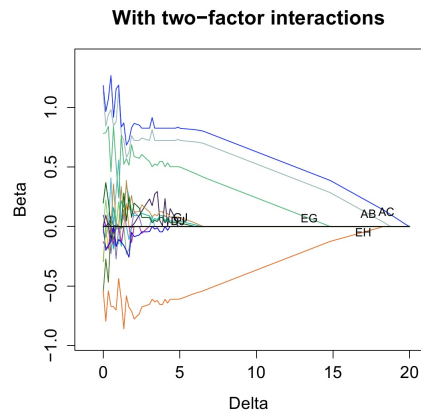
(i) Model (4.15) with  $\sigma^2 = 0.1$

Figure C.5: Plots obtained by the graphical Dantzig selector as introduced in Section 3.2, when considering main effects for data generated from the models (4.7) - (4.15) with  $\sigma^2 = 0.1$ .

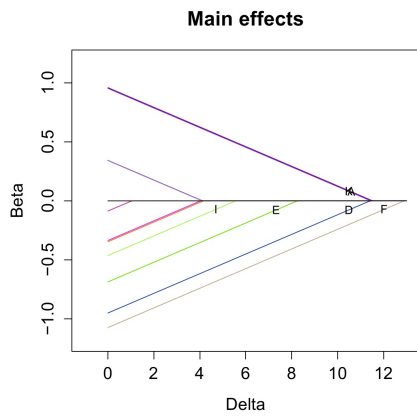
## C.6 Graphical Dantzig selector for data from the models (4.7) - (4.15) generated with $\sigma^2 = 0.5$ .



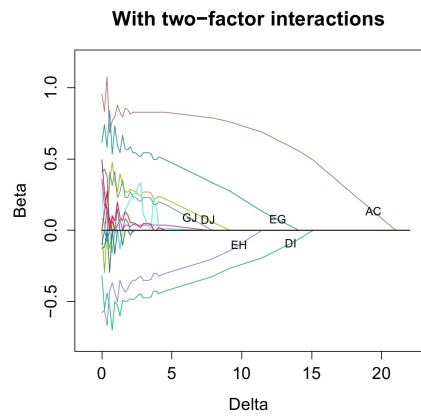
(a) Model (4.7) with  $\sigma^2 = 0.5$



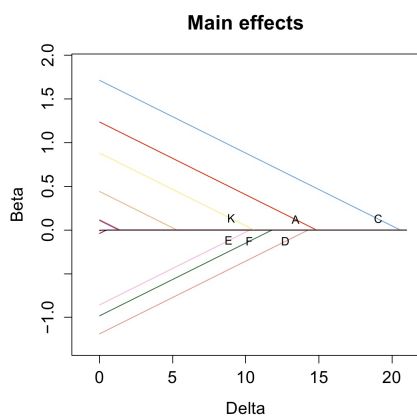
(b) Model (4.7) with  $\sigma^2 = 0.5$



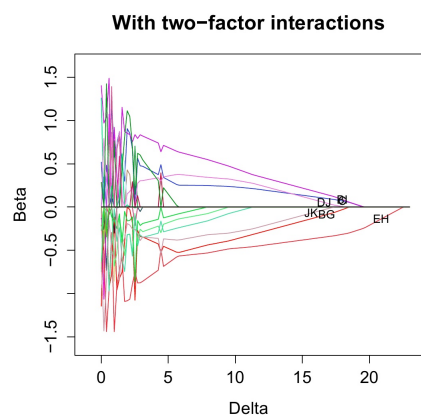
(c) Model (4.8) with  $\sigma^2 = 0.5$



(d) Model (4.8) with  $\sigma^2 = 0.5$

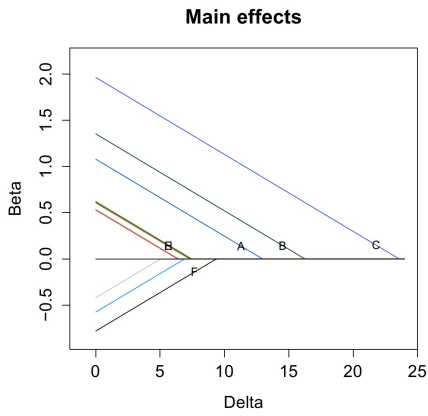


(e) Model (4.9) with  $\sigma^2 = 0.5$

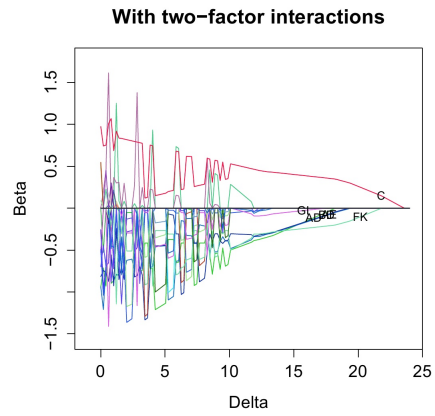


(f) Model (4.9) with  $\sigma^2 = 0.5$

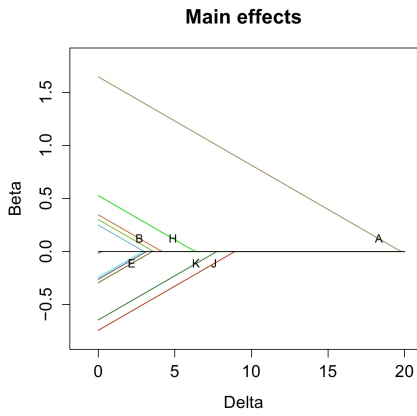
Figure C.6: Plots obtained by the graphical Dantzig selector as introduced in Section 3.2, for data generated from the models (4.7) - (4.15) with  $\sigma^2 = 0.5$ . Plots when considering main effects are given to the left, and when also considering two-factor interaction to the right.



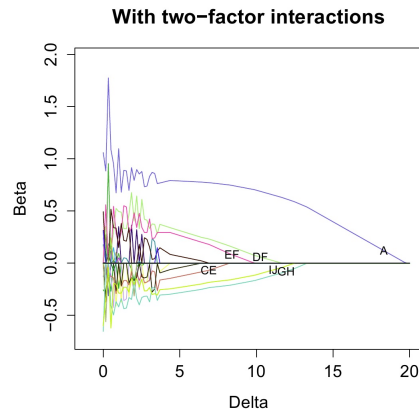
(g) Model (4.10) with  $\sigma^2 = 0.5$



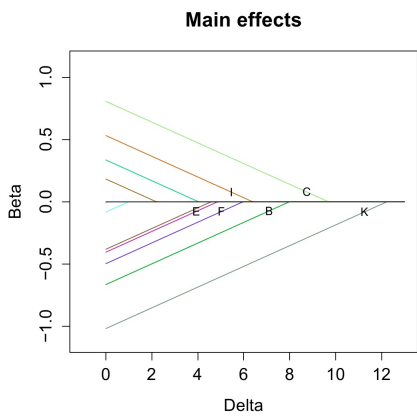
(h) Model (4.10) with  $\sigma^2 = 0.5$



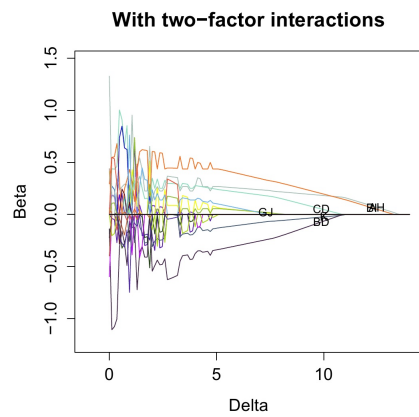
(i) Model (4.11) with  $\sigma^2 = 0.5$



(j) Model (4.11) with  $\sigma^2 = 0.5$

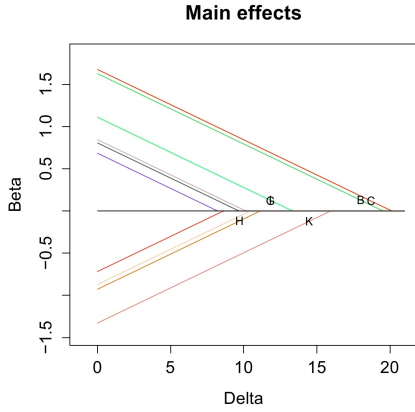


(k) Model (4.12) with  $\sigma^2 = 0.5$

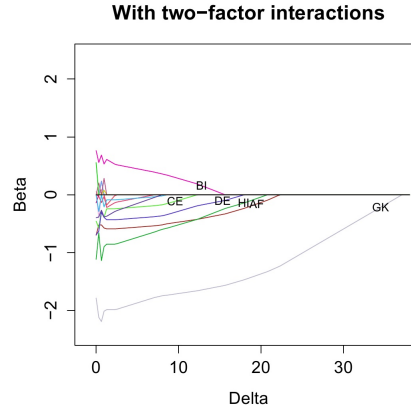


(l) Model (4.12) with  $\sigma^2 = 0.5$

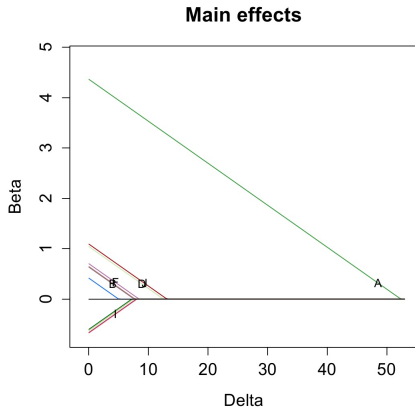
Figure C.6: Plots obtained by the graphical Dantzig selector as introduced in Section 3.2, for data generated from the models (4.7) - (4.15) with  $\sigma^2 = 0.5$ . Plots when considering main effects are given to the left, and when also considering two-factor interaction to the right.



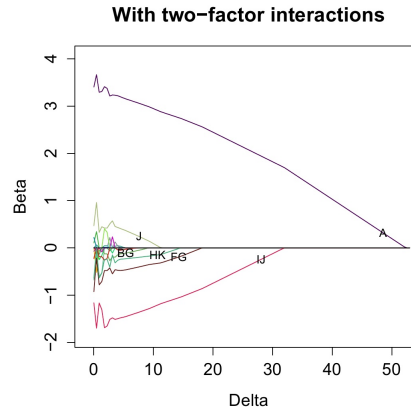
(m) Model (4.13) with  $\sigma^2 = 0.5$



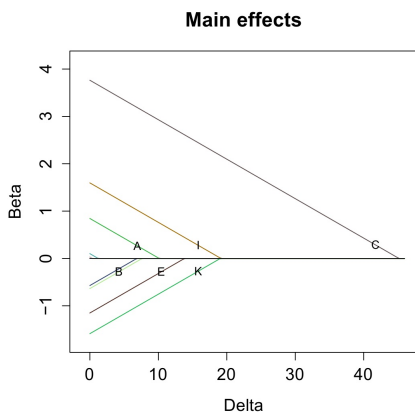
(n) Model (4.13) with  $\sigma^2 = 0.5$



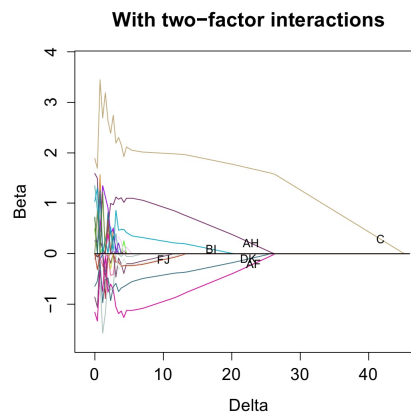
(o) Model (4.14) with  $\sigma^2 = 0.5$



(p) Model (4.14) with  $\sigma^2 = 0.5$



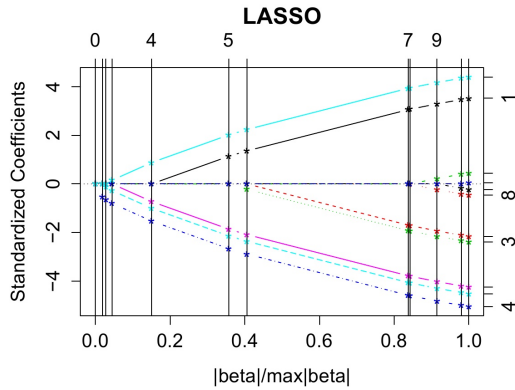
(q) Model (4.15) with  $\sigma^2 = 0.5$



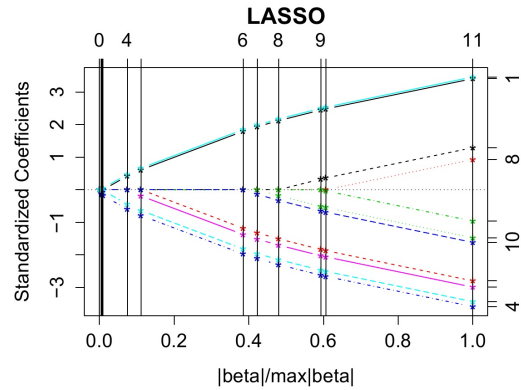
(r) Model (4.15) with  $\sigma^2 = 0.5$

Figure C.6: Plots obtained by the graphical Dantzig selector as introduced in Section 3.2, for data generated from the models (4.7) - (4.15) with  $\sigma^2 = 0.5$ . Plots when considering main effects are given to the left, and when also considering two-factor interaction to the right.

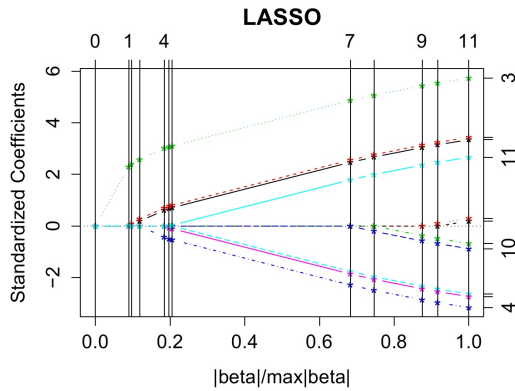
## C.7 Lasso for data generated from the models (4.7) - (4.15) with $\sigma^2 = 0.1$ when considering main effects.



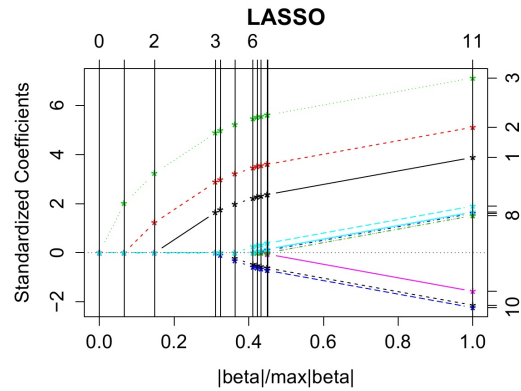
(a) Main effects for Model (4.7) with  $\sigma^2 = 0.1$



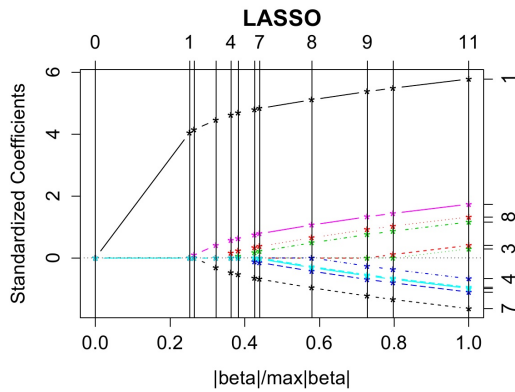
(b) Main effects for Model (4.8) with  $\sigma^2 = 0.1$



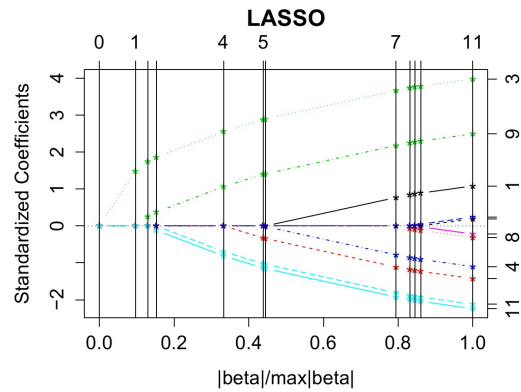
(c) Main effects for Model (4.9) with  $\sigma^2 = 0.1$



(d) Main effects for Model (4.10) with  $\sigma^2 = 0.1$

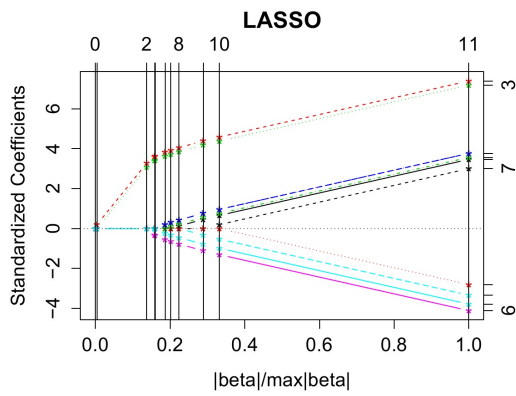


(e) Main effects for Model (4.11) with  $\sigma^2 = 0.1$

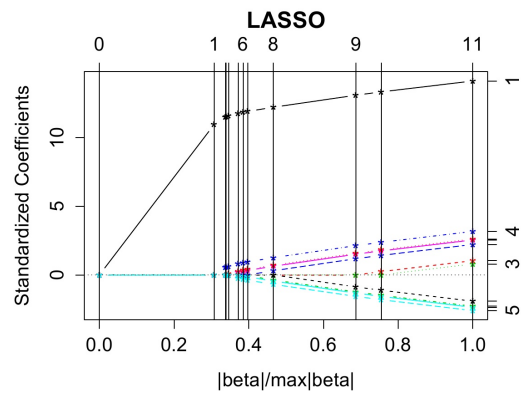


(f) Main effects for Model (4.12) with  $\sigma^2 = 0.1$

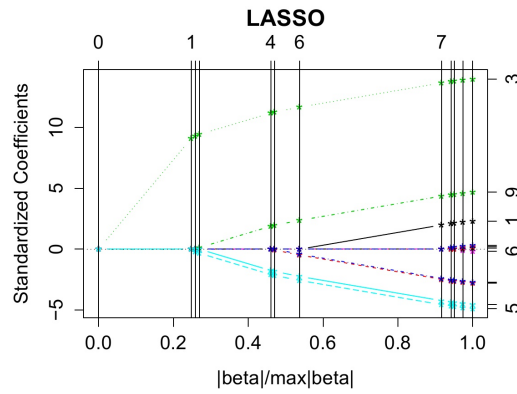
Figure C.7: Plots obtained by the Lasso as introduced in Section 3.4, when considering main effects for data generated from the models (4.7) - (4.15) with  $\sigma^2 = 0.1$ .



(g) Main effects for Model (4.13) with  $\sigma^2 = 0.1$



(h) Main effects for Model (4.14) with  $\sigma^2 = 0.1$

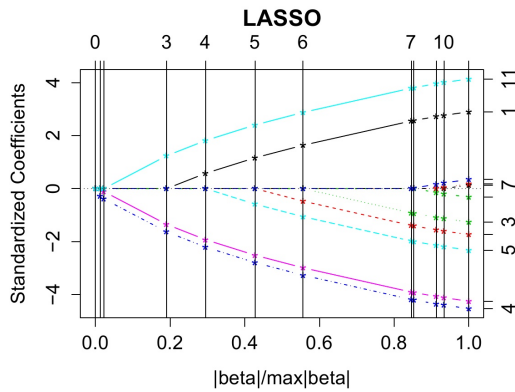


(i) Main effects for Model (4.15) with  $\sigma^2 = 0.1$

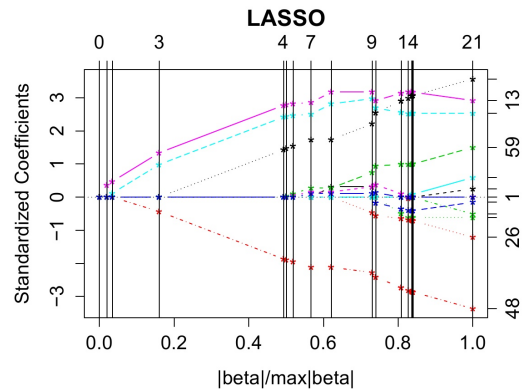
Figure C.7: Plots obtained by the Lasso as introduced in Section 3.4, when considering main effects for data generated from the models (4.7) - (4.15) with  $\sigma^2 = 0.1$ .



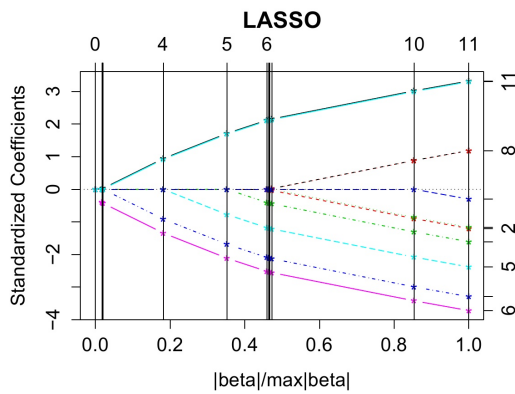
## C.8 Lasso for data from the models (4.7) - (4.15) generated with $\sigma^2 = 0.5$ .



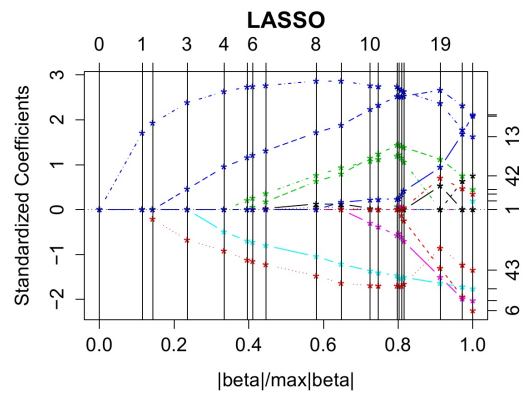
(a) Main effects for Model (4.7) with  $\sigma^2 = 0.5$



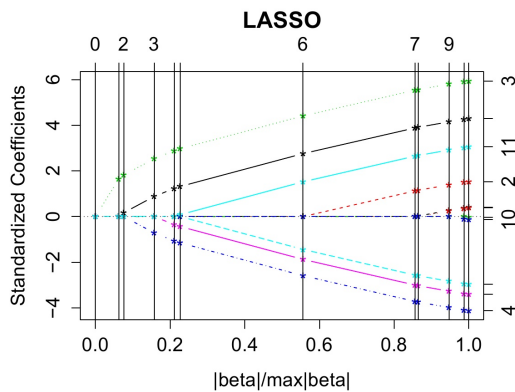
(b) With two-factor interactions for Model (4.7) with  $\sigma^2 = 0.5$



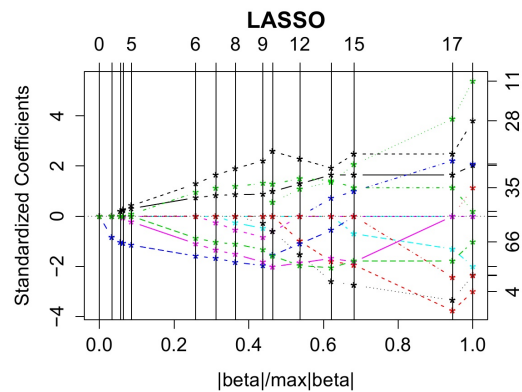
(c) Main effects for Model (4.8) with  $\sigma^2 = 0.5$



(d) With two-factor interactions for Model (4.8) with  $\sigma^2 = 0.5$

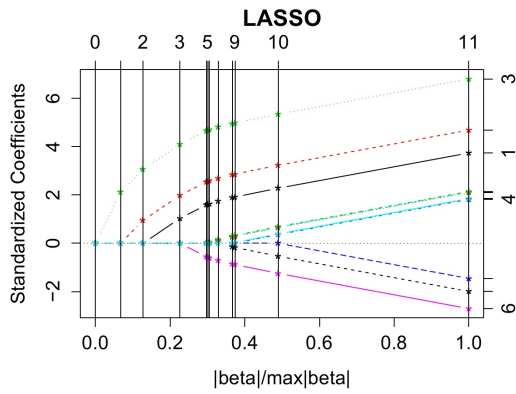


(e) Main effects for Model (4.9) with  $\sigma^2 = 0.5$

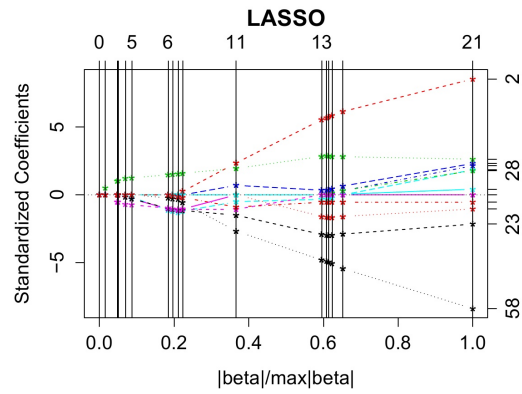


(f) With two-factor interactions for Model (4.9) with  $\sigma^2 = 0.5$

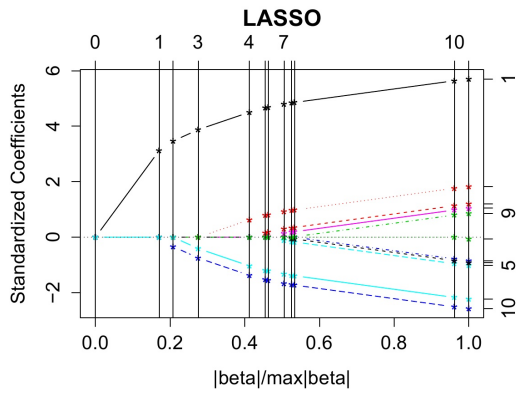
Figure C.8: Plots obtained by the Lasso as introduced in Section 3.4, for data generated from the models (4.7) - (4.15) with  $\sigma^2 = 0.5$ . Plots when considering main effects are given to the left, and when also considering two-factor interaction to the right.



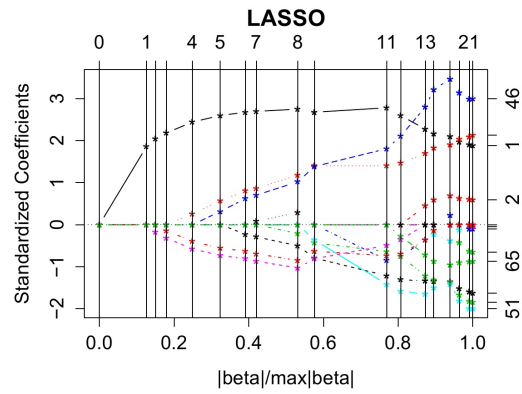
(g) Main effects for Model (4.10) with  $\sigma^2 = 0.5$



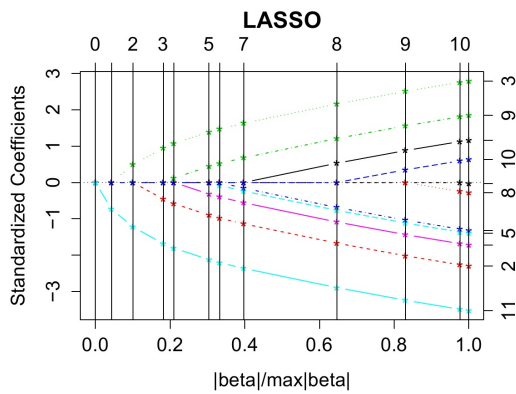
(h) With two-factor interactions for Model (4.10) with  $\sigma^2 = 0.5$



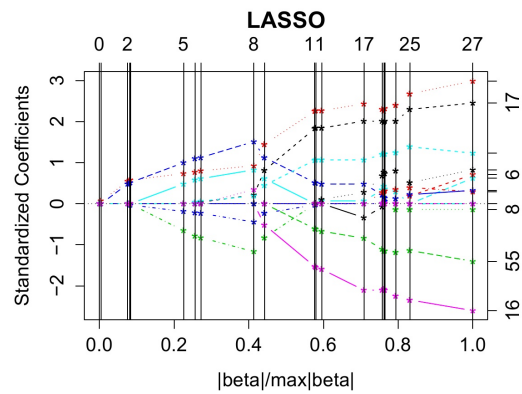
(i) Main effects for Model (4.11) with  $\sigma^2 = 0.5$



(j) With two-factor interactions for Model (4.11) with  $\sigma^2 = 0.5$

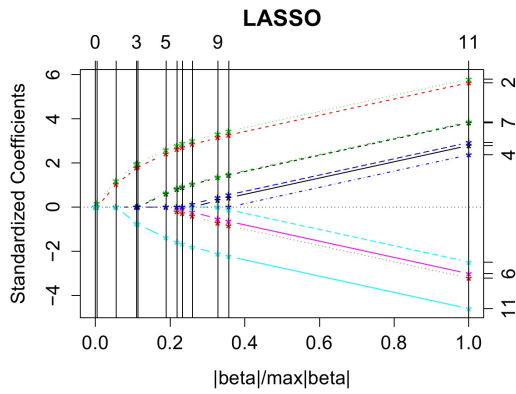


(k) Main effects for Model (4.12) with  $\sigma^2 = 0.5$

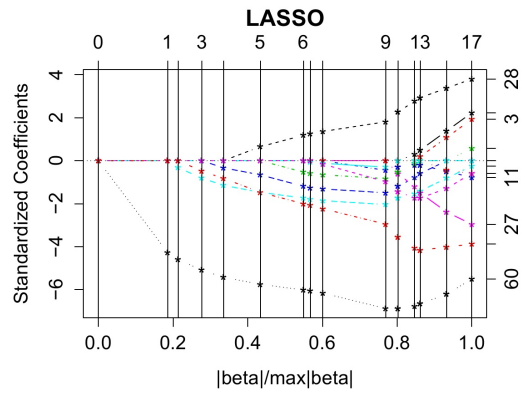


(l) With two-factor interactions for Model (4.12) with  $\sigma^2 = 0.5$

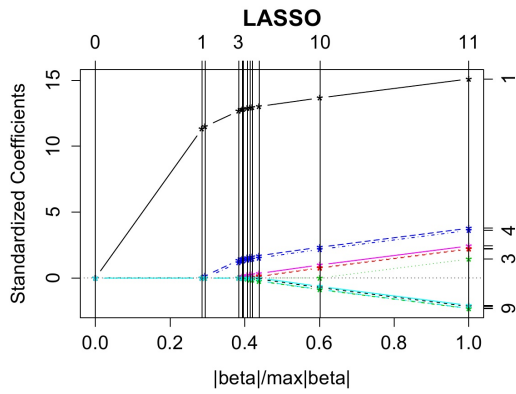
Figure C.8: Plots obtained by the Lasso as introduced in Section 3.4, for data generated from the models (4.7) - (4.15) with  $\sigma^2 = 0.5$ . Plots when considering main effects are given to the left, and when also considering two-factor interaction to the right.



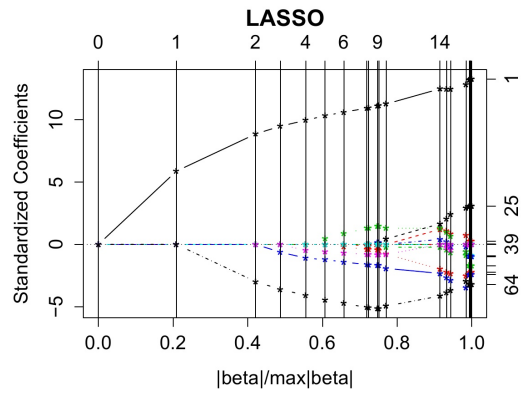
(m) Main effects for Model (4.13) with  $\sigma^2 = 0.5$



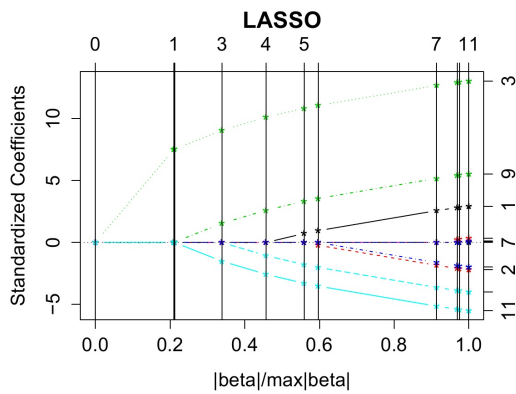
(n) With two-factor interactions for Model (4.13) with  $\sigma^2 = 0.5$



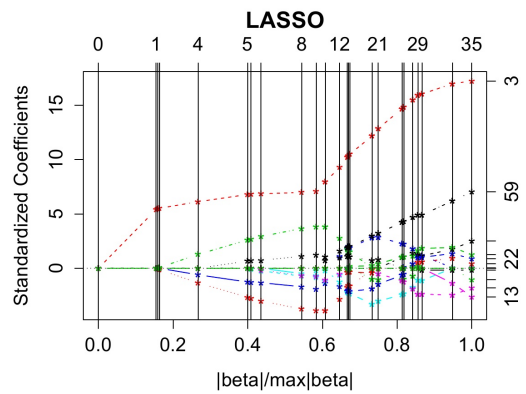
(o) Main effects for Model (4.14) with  $\sigma^2 = 0.5$



(p) With two-factor interactions for Model (4.14) with  $\sigma^2 = 0.5$



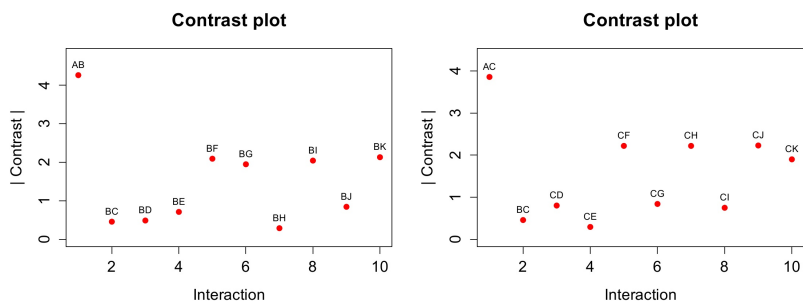
(q) Main effects for Model (4.15) with  $\sigma^2 = 0.5$



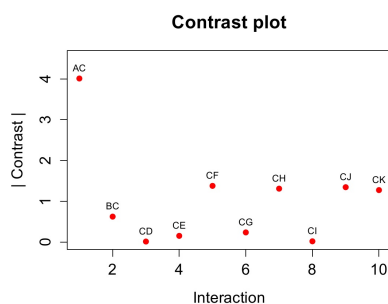
(r) With two-factor interactions for Model (4.15) with  $\sigma^2 = 0.5$

Figure C.8: Plots obtained by the Lasso as introduced in Section 3.4, for data generated from the models (4.7) - (4.15) with  $\sigma^2 = 0.5$ . Plots when considering main effects are given to the left, and when also considering two-factor interaction to the right.

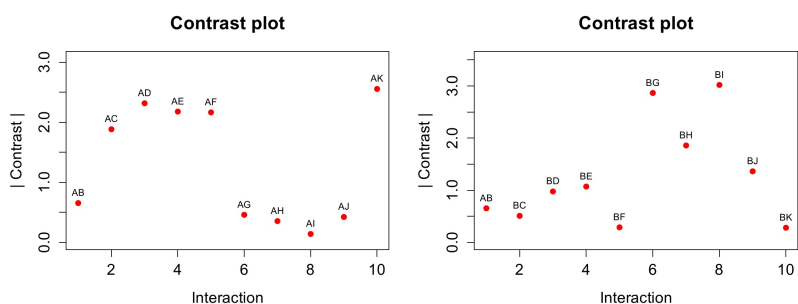
## C.9 Additional contrast plots for data generated from the models (4.7) - (4.15) with $\sigma^2 = 0.1$ .



(a) Model (4.7) with  $\sigma^2 = 0.1$       (b) Model (4.7) with  $\sigma^2 = 0.1$

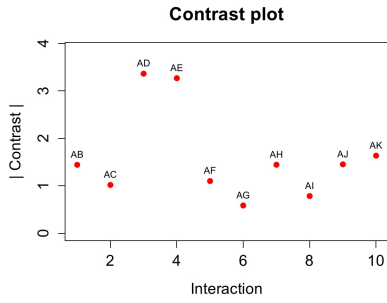


(c) Model (4.8) with  $\sigma^2 = 0.1$

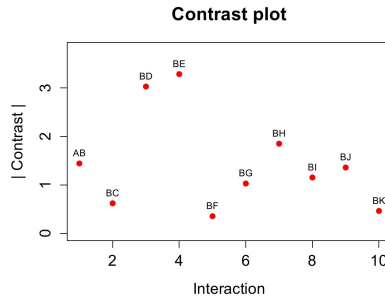


(d) Model (4.9) with  $\sigma^2 = 0.1$       (e) Model (4.9) with  $\sigma^2 = 0.1$

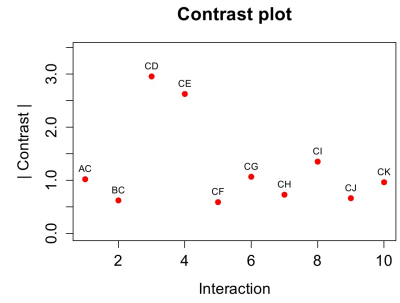
Figure C.9: Contrast plots used as support for the analysis performed in Subsection 4.2.6 for the data generated from the models (4.7) - (4.15) with  $\sigma^2 = 0.1$ .



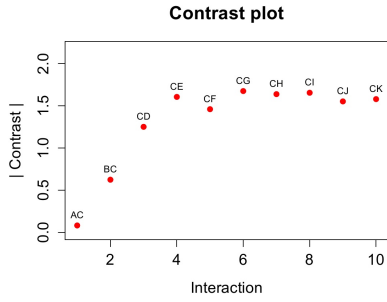
(f) Model (4.10) with  $\sigma^2 = 0.1$



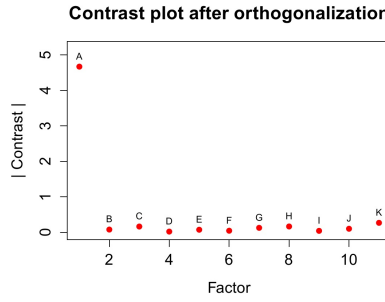
(g) Model (4.10) with  $\sigma^2 = 0.1$



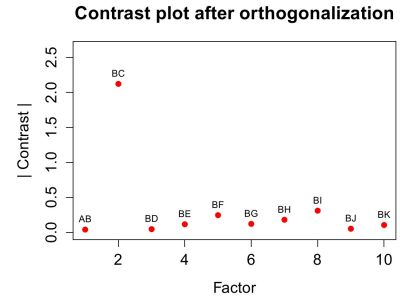
(h) Model (4.10) with  $\sigma^2 = 0.1$



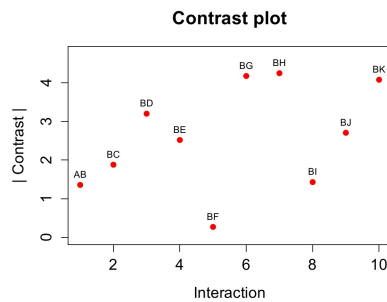
(i) Model (4.11) with  $\sigma^2 = 0.1$



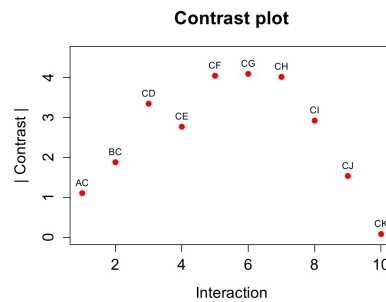
(j) Model (4.11) with  $\sigma^2 = 0.1$



(k) Model (4.11) with  $\sigma^2 = 0.1$



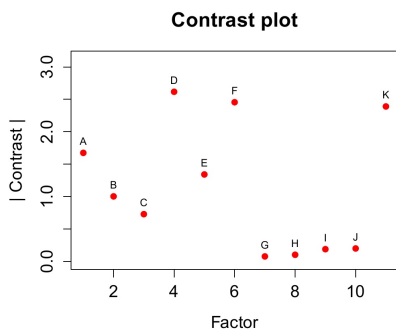
(l) Model (4.14) with  $\sigma^2 = 0.1$



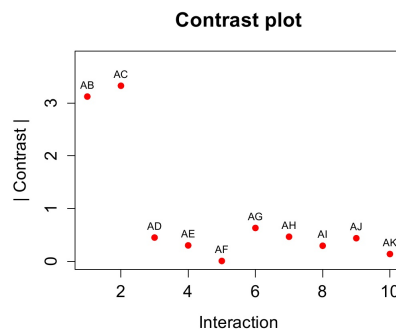
(m) Model (4.14) with  $\sigma^2 = 0.1$

Figure C.9: Contrast plots used as support for the analysis performed in Subsection 4.2.6 for the data generated from the models (4.7) - (4.15) with  $\sigma^2 = 0.1$ .

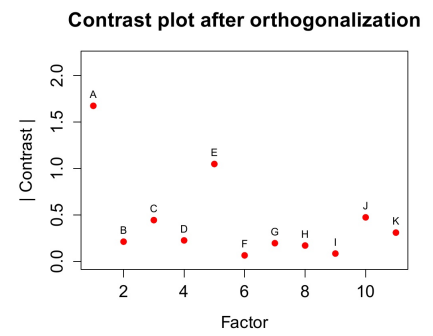
## C.10 Contrast plots for data generated from the models (4.7) - (4.15) with $\sigma^2 = 0.5$ .



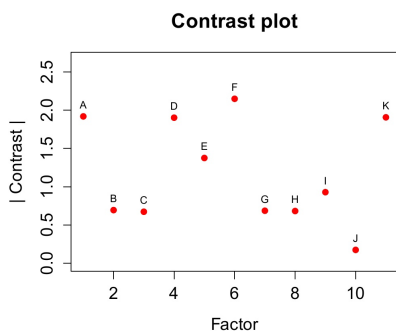
(a) Model (4.7) with  $\sigma^2 = 0.5$



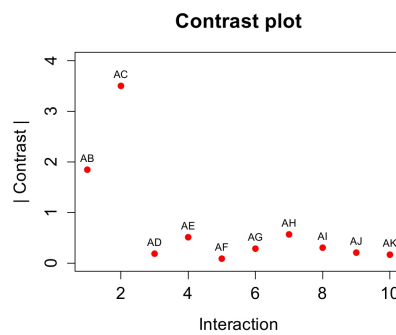
(b) Model (4.7) with  $\sigma^2 = 0.5$



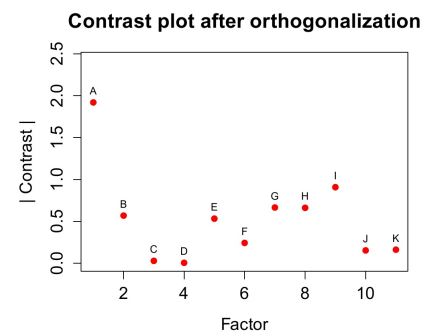
(c) Model (4.7) with  $\sigma^2 = 0.5$



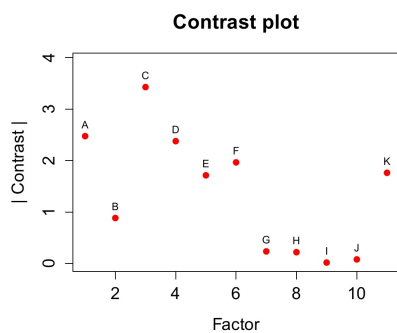
(d) Model (4.8) with  $\sigma^2 = 0.5$



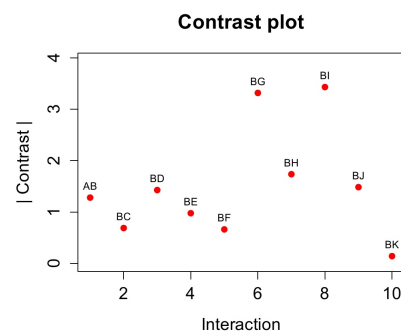
(e) Model (4.8) with  $\sigma^2 = 0.5$



(f) Model (4.8) with  $\sigma^2 = 0.5$

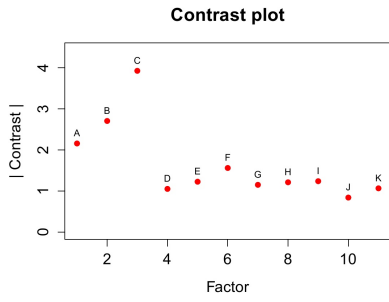


(g) Model (4.9) with  $\sigma^2 = 0.5$

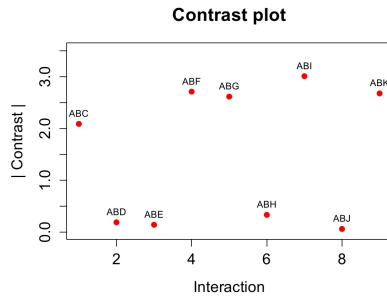


(h) Model (4.9) with  $\sigma^2 = 0.5$

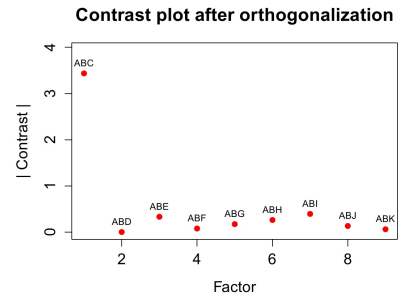
Figure C.10: Contrast plots used in the analysis performed in Subsection 4.2.6 for the data generated from the models (4.7) - (4.15) with  $\sigma^2 = 0.5$ .



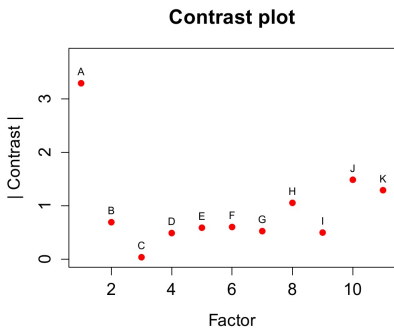
(i) Model (4.10) with  $\sigma^2 = 0.5$



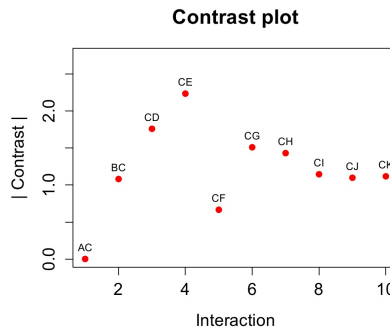
(j) Model (4.10) with  $\sigma^2 = 0.5$



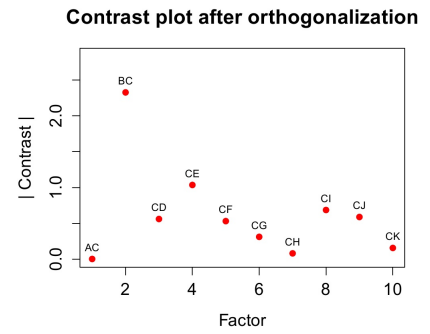
(k) Model (4.10) with  $\sigma^2 = 0.5$



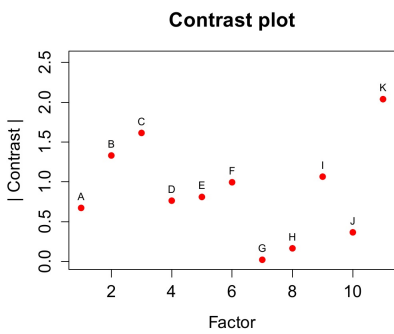
(l) Model (4.11) with  $\sigma^2 = 0.5$



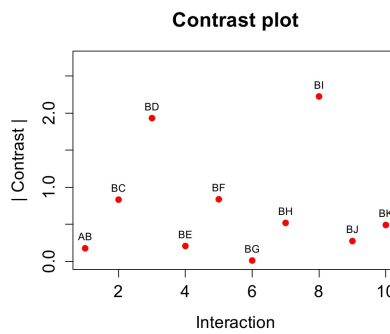
(m) Model (4.11) with  $\sigma^2 = 0.5$



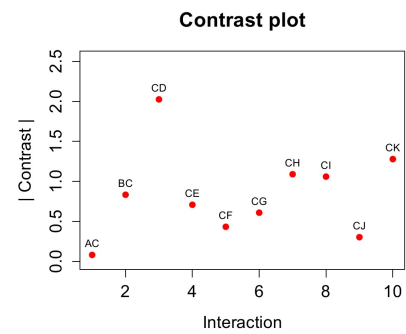
(n) Model (4.11) with  $\sigma^2 = 0.5$



(o) Model (4.12) with  $\sigma^2 = 0.5$

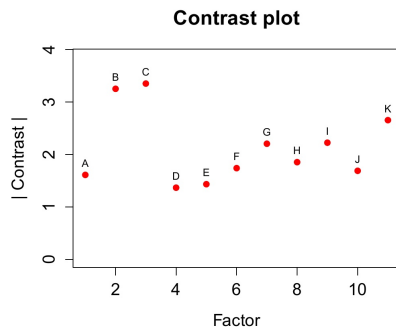


(p) Model (4.12) with  $\sigma^2 = 0.5$

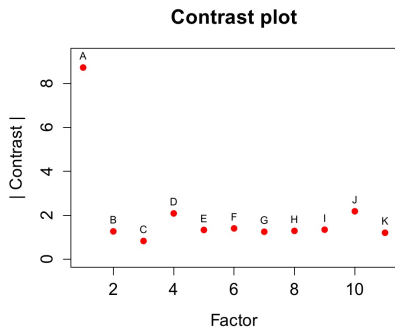


(q) Model (4.12) with  $\sigma^2 = 0.5$

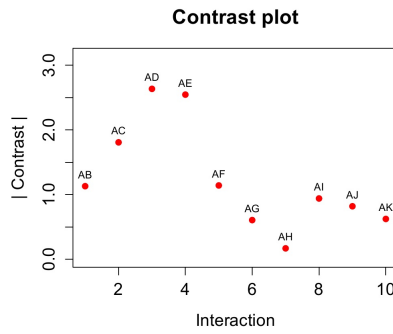
Figure C.10: Contrast plots used in the analysis performed in Subsection 4.2.6 for the data generated from the models (4.7) - (4.15) with  $\sigma^2 = 0.5$ .



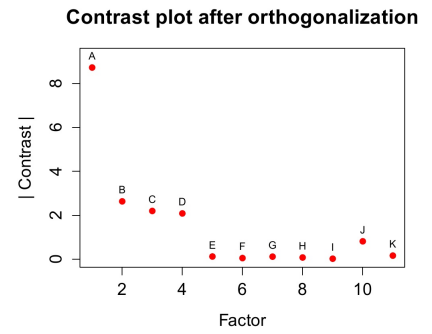
(r) Model (4.13) with  $\sigma^2 = 0.5$



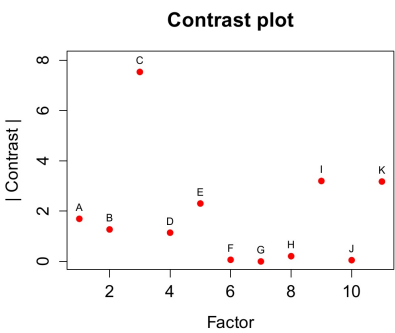
(s) Model (4.14) with  $\sigma^2 = 0.5$



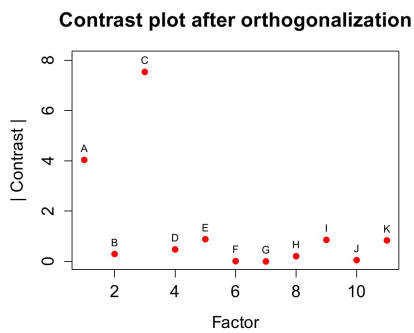
(t) Model (4.14) with  $\sigma^2 = 0.5$



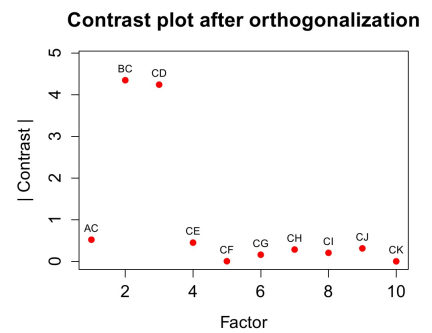
(u) Model (4.14) with  $\sigma^2 = 0.5$



(v) Model (4.15) with  $\sigma^2 = 0.5$



(w) Model (4.15) with  $\sigma^2 = 0.5$



(x) Model (4.15) with  $\sigma^2 = 0.5$

Figure C.10: Contrast plots used in the analysis performed in Subsection 4.2.6 for the data generated from the models (4.7) - (4.15) with  $\sigma^2 = 0.5$ .