Analysis of No-Confounding Designs using the Dantzig Selector

by

Archana Krishnamoorthy

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Approved October 2014 by the
Graduate Supervisory Committee:

Douglas Montgomery, Co-Chair
Connie Borror, Co-Chair
Rong Pan

ARIZONA STATE UNIVERSITY

December 2014

ABSTRACT

No-confounding designs (NC) in 16 runs for 6, 7, and 8 factors are non-regular fractional factorial designs that have been suggested as attractive alternatives to the regular minimum aberration resolution IV designs because they do not completely confound any two-factor interactions with each other. These designs allow for potential estimation of main effects and a few two-factor interactions without the need for follow-up experimentation. Analysis methods for non-regular designs is an area of ongoing research, because standard variable selection techniques such as stepwise regression may not always be the best approach. The current work investigates the use of the Dantzig selector for analyzing no-confounding designs. Through a series of examples it shows that this technique is very effective for identifying the set of active factors in no-confounding designs when there are three of four active main effects and up to two active two-factor interactions.

To evaluate the performance of Dantzig selector, a simulation study was conducted and the results based on the percentage of type II errors are analyzed. Also, another alternative for 6 factor NC design, called the Alternate No-confounding design in six factors is introduced in this study. The performance of this Alternate NC design in 6 factors is then evaluated by using Dantzig selector as an analysis method. Lastly, a section is dedicated to comparing the performance of NC-6 and Alternate NC-6 designs.

ACKNOWLEDGMENTS

I would like to express my gratitude to Dr. Douglas Montgomery for his guidance in every step and his generosity with time. He has always wowed me with his unfailing energy and love of learning. Being around him inspires me to work better. I am grateful to Dr. Connie Borror for her valuable feedback and time. I would like to thank Dr. Rong Pan for his guidance during different stages of my research.

I am also hugely thankful to Dr. Shilpa Shinde for her guidance during the initial stages of my research. My special thanks to Dr. Brian Stone, who took time away from his busy work to speak with me and answer my many questions.

My family and friends have provided the support for this research. Words cannot express my gratitude to my husband for his constant encouragement and support.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

Chapter 1

INTRODUCTION

1.1 Background

Screening experiments are widely used during the early stages of experimentation when many factors are considered and when the objective is to identify the active factors that have large effects. Fractional factorial designs, the most extensively used screening designs are classified into two broad types: Regular and Non-regular designs. The factor effects in regular designs are either unaliased or completely confounded. Non-regular designs exhibit a more complex aliasing but the effects are not completely confounded. Hence there is at least some chance that some information on the aliased effects may be available.

In the screening experiments involving 6-8 factors, resolution IV Fractional factorials designs (FFD) are widely used, as they are economical and provide clear estimates of main effects when three-factor and higher-order interactions are negligible. However, if the estimation of two-factor interaction is required, experimenters are frequently required to augment the original fraction with new runs to resolve ambiguities resulting from the factors being completely confounded.

1.2 Literature Survey

Non regular designs have received considerable attention ever since Plackett and Burman (PB) (1946) provided a series of two-level fractional factorial designs for examining (n - 1) factors in n runs, where n is a multiple of four and n <100. When only main effects are involved, PB designs can estimate all of them. Additionally, they have smaller run size requirements and are much more flexible in accommodating many factor levels. However, traditionally, non-regular factorials were not advocated because of their complex aliasing structure and difficulty in interpreting interactions. Hamada and Wu (1992) showed that for data from designs with complex aliasing, it is possible to detect interaction effects.

Consequently in recent years, PB designs enjoy wide applications and a number of analysis methods were put forth. Hamada and Wu (1992) proposed a traditional analysis approach to identify significant interactions while Chipman et al. (1997) proposed a Bayesian variable selection approach for analyzing experiments with complex aliasing. Tyssedal and Samset (1997) suggested using contrast plots and employ the aliasing structure of the non-regular designs to identify significant effects.

Supersaturated designs, where the run size is not enough for estimating all the active effects, have also been widely researched. Wu (1993) and Lin (1993) constructed supersaturated designs through partially aliased interactions and half fractions of Hadamard matrices, respectively. Wu and Hamada (2000) pointed out that the analysis of supersaturated designs is driven by effect sparsity and effect heredity.

Variable selection methods like all possible regression are not computationally feasible for a supersaturated design. Phoa et. all (2008) presented factor screening based on Dantzig selector. Marley et al (2009) compared four analysis strategies and confirmed that shrinkage methods perform better for supersaturated designs. Draguljic et al (2014) analyzed screening strategies in the presence of interactions and pointed out that Dantzig Selector outperforms other shrinkage methods. Mee (2013) discussed analysis of non-regular designs of different strengths using forward selection method.

Jones and Montgomery (2010) proposed a new set of non-regular orthogonal designs called the no-confounding designs (NC) that have no complete confounding of two factor interactions. For 6-8 factors, these designs are projections of the Hall designs (1961) created by selecting specific sets of columns. Since the two factor interactions are not completely aliased, these designs allow for the estimation of models containing both interactions and main effects. In this study, the NC designs will be analyzed using the Dantzig selector proposed by Candes and Tao (2007).

1.3 No-confounding designs:

This section presents the NC designs and highlights some of its properties. The recommended no-confounding design for 6 factors is given in Table 1.1. Figure 1.1 displays the correlation matrix for this design along with the correlation matrix for the regular fraction.

Table 1.1 Recommended 16-run six-factor no-confounding design

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

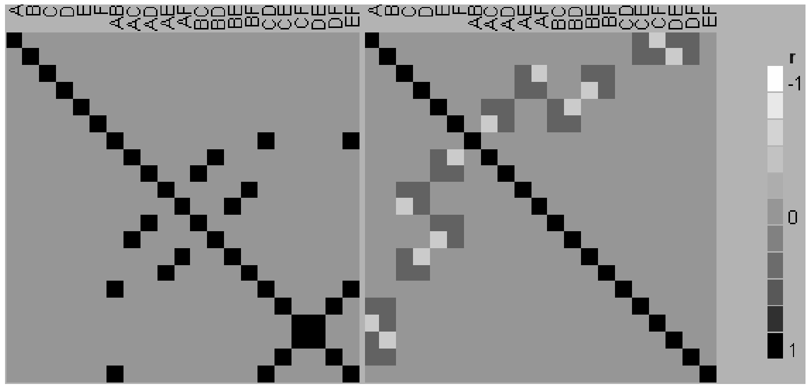

Figure 1.1 Correlation matrix (a) regular 2 $^{6-2}$ fractional factorial (b) the no-confounding design

Since all of the off-diagonal entries in the correlation matrix for the NC design are between –1 and +1, it is clear there are no two-factor interactions completely aliased with each other, unlike the regular 2^{6-2} fraction.

Tables 1.2 and 1.3 provide the no-confounding designs in 16 runs for seven and eight factors along with the correlation matrices.

Table 1.2 Recommended 16-run seven factor no-confounding design

Runs	A	B	C	D	E	F	G
1	1	1	1	1	1	1	1
2	1	1	1	−1	−1	−1	−1
3	1	1	−1	1	1	−1	−1
4	1	1	−1	−1	−1	1	1
5	1	−1	1	1	−1	1	−1
6	1	−1	1	−1	1	−1	1
7	1	−1	−1	1	−1	−1	1
8	1	−1	−1	−1	1	1	−1
9	−1	1	1	1	1	1	−1
10	−1	1	1	−1	−1	−1	1
11	−1	1	−1	1	−1	1	1
12	−1	1	−1	−1	1	−1	−1
13	−1	−1	1	1	−1	−1	−1
14	−1	−1	1	−1	1	1	1
15	−1	−1	−1	1	1	−1	1
16	−1	−1	−1	−1	−1	1	−1

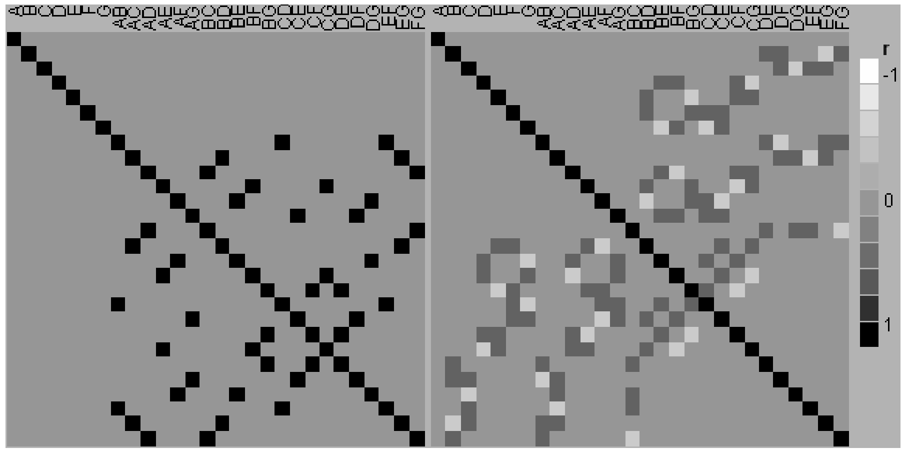

Figure 1.2 Correlation matrix (a) regular 2^{7-3} fractional factorial (b) the no-confounding design

Table 1.3 Recommended 16-run eight factor no-confounding design

Runs	A	B	C	D	E	F	G	H
1	1	1	1	1	1	1	1	1
2	1	1	1	1	−1	−1	−1	−1
3	1	1	−1	−1	1	1	−1	−1
4	1	1	−1	−1	−1	−1	1	1
5	1	−1	1	−1	1	−1	1	−1
6	1	−1	1	−1	−1	1	−1	1
7	1	−1	−1	1	1	−1	−1	1
8	1	−1	−1	1	−1	1	1	−1
9	−1	1	1	1	1	1	1	1
10	−1	1	1	−1	1	−1	−1	−1
11	−1	1	−1	1	−1	−1	1	−1
12	−1	1	−1	−1	−1	1	−1	1
13	−1	−1	1	1	−1	−1	−1	1
14	−1	−1	1	−1	−1	1	1	−1
15	−1	−1	−1	1	1	1	−1	−1
16	−1	−1	−1	−1	1	−1	1	1

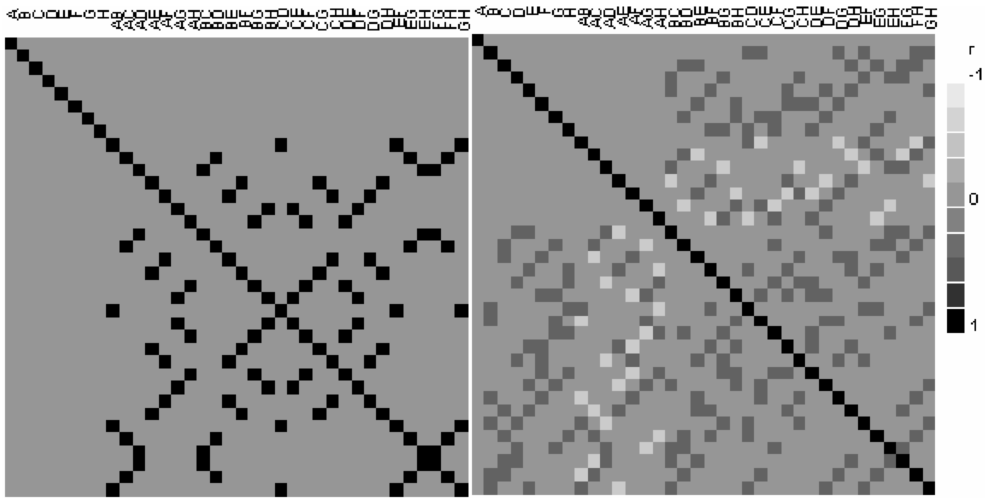

Figure 1.3 Correlation matrix (a) regular 2^{8-4} fractional factorial (b) the no-confounding design

To identify active factors, Jones and Montgomery (2010) illustrated using forward stepwise regression with all main effects and two-factor interactions as candidate effects. Stepwise regression is a popular method for model selection as it is widely available in commercial software. Shinde (2012) analyzed NC designs using stepwise regression and confirmed that the method does not perform well when the number of active effects exceeds four. In this study, an attempt is made to analyze the active factors in no-confounding designs using the Dantzig selector proposed by Candes and Tao (2007).

1.4 The Dantzig selector

The Dantzig selector (DS) is an effective variable selection technique which chooses the best subset of variables or active factors by solving a simple convex program. It is a shrinkage method that achieves variable selection by shrinking the estimated regression coefficients towards zero at a different rate.

For a linear regression model $\boldsymbol{y} = \boldsymbol{X\hat{\beta}}, +\boldsymbol{\varepsilon}$, where **y** is an $n \times 1$ vector of observations, **X** is an $n{\times}p$ model matrix, $\boldsymbol{\beta}$ is a $k{\times}1$ vector of unknown parameters to be estimated and ε is an n×1 vector of random errors, the estimate of $\boldsymbol{\beta}$ is chosen to satisfy,

$$\min_{\hat{\boldsymbol{\beta}} \in \boldsymbol{R}^k} \|\hat{\boldsymbol{\beta}}\|_{l1} \textit{subject to } \|\boldsymbol{X}^t \boldsymbol{r}\|_{l_\infty} \leq \boldsymbol{\delta}$$

where **r** is the residual vector $\mathbf{r}=\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}$ and $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I_n})$. Here $||\,\beta\,||_1 = |\beta_0| + \ldots + |\beta_k|$ is the l_1 norm $||\,\beta\,||_\infty = \max(|\beta_0|, \ldots, |\beta_k|)$ is the l_∞ norm, and δ is the tuning parameter. Here, we seek an estimator $\boldsymbol{\beta}$ with minimum complexity (as measured by the l_1-norm) among all objects that are consistent with the data.

A two-stage estimation approach called the Gauss-Dantzig selector was also proposed by Candes and Tao (2007). First the Dantzig selector is used to identify the active factors, and then the factors whose coefficient estimates are greater than γ are retained. Second, least squares estimates are found by regressing the response on the set of retained factors. The value of γ is the threshold between the signal and the noise and hence affects the selection of active factors.

In Chapter 2, we discuss how the Dantzig selector performs for selected examples. We propose an alternate NC-6 design in Chapter 3 and also present an extensive simulation study to analyze the performance of Dantzig selector, as a variable selection method for NC designs.

Chapter 2

EXAMPLES OF ANALYSING NO-CONFOUNDING DESIGNS USING DANTZIG SELECTOR

2.1. Background

The purpose of screening experiments is to identify the active factors correctly and economically. For 6, 7 and 8 factors, resolution IV regular fractional factorial designs in 16 runs are widely used as screening designs; however, because the two factor interactions are completely confounded, experimenters are frequently required to augment the original fraction with more runs to resolve the ambiguities.

The no-confounding designs introduced by Jones and Montgomery (2010) are non-regular designs like the Plackett-Burman designs. Since they do not completely confound any of the two factor interactions, they potentially allow for unambiguous estimation of models containing both main effects and a few two factor interactions. . In a six factor example Jones and Montgomery used forward stepwise regression with all main effects and two-factor interactions as candidate effects to successfully identify the active factors. However, stepwise regression methods may not always be the best approach. Analysis methods for non-regular designs is an area of on-going research. Mee (2013) provides an excellent discussion of several analysis methods for non-regular designs.

Candes and Tao (2007) propose the Dantzig selector as a technique for variable selection in problems where the number of factors exceeds the number of runs. Phao, Pan and Xu (2009) demonstrate the successful application of this technique to supersaturated designs. Because NC designs are essentially supersaturated designs for the main effects and all two-factor interactions, we suspect that the Dantzig selector would also perform well in this application. We evaluate how the Dantzig selector performs in analyzing the

Jones and Montgomery no-confounding designs involving 6, 7 and 8 factors. Each of the designs was tested with simulated models involving main effects only, models with main effects plus one hierarchical two-factor interaction, and models with main effects and two two-factor interactions as these scenarios are very typical of the situation frequently encountered in practice.

2.2. Analysis of the six factor NC design using Dantzig selector

The Dantzig selector is a shrinkage method where the estimates of the regression coefficients are chosen to satisfy,

$$\min_{\hat{\boldsymbol{\beta}} \in \boldsymbol{R}^k} \left\| \hat{\boldsymbol{\beta}} \right\|_{l1} \textit{subject to } \left\| \boldsymbol{X}^t \boldsymbol{r} \right\|_{l_\infty} \leq \boldsymbol{\delta}$$

where **r** is the residual vector $\mathbf{r}=\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}$, δ is a tuning parameter, Here $|| \boldsymbol{\beta} ||_1 = |\beta_0| + \ldots + |\beta_k|$ is the l_1 norm and $|| \boldsymbol{\beta} ||_\infty = \max(|\beta_0|, \ldots, |\beta_k|)$ is the l_∞ norm.

In applying a Gauss-Dantzig selector, Candes and Tao (2007) suggested choosing a set of active factors for a specific value of δ based on some model fitting criteria, and then using standard least squares to fit the reduced linear model and produce the final estimates of the model parameters. The choice of δ in the l_1-regularisation problem equation has a significant impact on the results. When δ is too large, it leads to the insignificance of all predictors on the change in response (type II errors) and when δ is too small, inactive factors with small magnitudes of coefficients are falsely included in the model (type I errors). In the examples below, the designs and models were tested for a range of values of δ. The analysis was performed using JMP. The penalized regression estimates for each value of δ is obtained and then the modified AIC criterion (mAIC) proposed by Phoa et al. (2009) was used for final model selection.

Screening designs should be effective in detecting real factor effects of moderately large size, say at least two standard deviations in magnitude. Smaller effects may be statistically significant, but unless they exert a relatively large effect on the mean response, they may not have much practical significance. To investigate how the six factor NC design performs with a simple model containing only main effects the first-order model with significant factors A, B, and F, $y=1.5+2.9A+4.5B-2.7F + \varepsilon$, was assumed and a response generated with standard normal random noise $\varepsilon \sim N(0, 1)$ constituted the data . Table 2.1 displays the NC-6 design with the simulated responses.

Table 2.1 NC-6 design with responses for the model $y=1.5+2.9A+4.5B-2.7F + \varepsilon$

Runs	**A**	**B**	**C**	**D**	**E**	**F**	**Y**
1	1	1	1	1	1	1	5.881004
2	1	1	-1	-1	-1	-1	10.51876
3	-1	-1	1	1	-1	-1	-5.86881
4	-1	-1	-1	-1	1	1	-7.17214
5	1	1	1	-1	1	-1	12.93893
6	1	1	-1	1	-1	1	7.439093
7	-1	-1	1	-1	-1	1	-9.89413
8	-1	-1	-1	1	1	-1	-3.90374
9	1	-1	1	1	1	-1	2.982714
10	1	-1	-1	-1	-1	1	-1.84172
11	-1	1	1	1	-1	1	-0.63985
12	-1	1	-1	-1	1	-1	5.742818
13	1	-1	1	-1	-1	-1	2.100142
14	1	-1	-1	1	1	1	-3.53534
15	-1	1	1	-1	1	1	0.013759
16	-1	1	-1	1	-1	-1	5.666989

Table 2.2 shows selected results when the above model was analyzed for various values of delta. Except for the columns shown, the estimated effects for all other variables were zero, and those columns were omitted from the table for clarity. A wide range of values for the tuning parameter δ (between 0.5 and 100) was tested. Intervals between consecutive δ may not be consistent as the δ between 0-10 was manually input with a 0.25 interval and the δ >10 were selected randomly. It is evident that as the δ increases, fewer effects are included in the model. It can be observed that the model that has the minimum value of mAIC correctly identifies the true model form, although the

parameter estimates depend on the value of δ. A least squares fit of the first-order model with factors A, B, and F produces the correct estimates of the model parameters.

Table 2.2 NC-6 results for the model $y=1.5+2.9A+4.5B-2.7F+\varepsilon$

Delta	A	B	C	D	E	F	mAIC
0.5	3.252418	4.637159	0	0	0	-2.4642	16.71075
0.75	3.236793	4.621534	0	0	0	-2.44857	16.71075
1	3.221168	4.605909	0	0	0	-2.43295	16.71075
1.25	3.205543	4.590284	0	0	0	-2.41732	16.71075
1.5	3.189918	4.574659	0	0	0	-2.4017	16.71075
1.75	3.174293	4.559034	0	0	0	-2.38607	16.71075
2	3.158668	4.543409	0	0	0	-2.37045	16.71075
2.25	3.143043	4.527784	0	0	0	-2.35482	16.71075
2.5	3.127418	4.512159	0	0	0	-2.3392	16.71075
2.75	3.111793	4.496534	0	0	0	-2.32357	16.71075
3	3.096168	4.480909	0	0	0	-2.30795	16.71075
3.25	3.080543	4.465284	0	0	0	-2.29232	16.71075
3.5	3.064918	4.449659	0	0	0	-2.2767	16.71075
3.75	3.049293	4.434034	0	0	0	-2.26107	16.71075
4	3.033668	4.418409	0	0	0	-2.24545	16.71075
4.25	3.018043	4.402784	0	0	0	-2.22982	16.71075
4.5	3.002418	4.387159	0	0	0	-2.2142	16.71075
4.75	2.986793	4.371534	0	0	0	-2.19857	16.71075
5	2.971168	4.355909	0	0	0	-2.18295	16.71075
5.25	2.955543	4.340284	0	0	0	-2.16732	16.71075
5.5	2.939918	4.324659	0	0	0	-2.1517	16.71075
5.75	2.924293	4.309034	0	0	0	-2.13607	16.71075
6	2.908668	4.293409	0	0	0	-2.12045	16.71075
13.44909	2.443099	3.827841	0	0	0	-1.65488	16.71075
25.89818	1.665031	3.049773	0	0	0	-0.87681	16.71075
38.34727	0.886963	2.271704	0	0	0	0	39.47342
50.79636	0	1.493636	0	0	0	0	48.18566
63.24545	0	0.715568	0	0	0	0	48.18566

To investigate the impact of a significant two-factor interaction effect another model, $y=1.5+2.2A+4.5B+1.9F+2.9AF + \varepsilon$ was created. Random normal noise, $\varepsilon \sim N(0, 1)$ was included to generate the responses. The design with responses and selected results for delta ranging between 0.5-15 are presented in Table 2.3 and 2.4, respectively.

Table 2.3 NC-6 design with responses for y=1.5+2.2A+4.5B+1.9F+2.9AF+ε

Runs	**A**	**B**	**C**	**D**	**E**	**F**	**Y**
1	1	1	1	1	1	1	14.34351
2	1	1	-1	-1	-1	-1	3.777049
3	-1	-1	1	1	-1	-1	-5.40884
4	-1	-1	-1	-1	1	1	-6.29689
5	1	1	1	-1	1	-1	2.84584
6	1	1	-1	1	-1	1	15.12935
7	-1	-1	1	-1	-1	1	-6.72336
8	-1	-1	-1	1	1	-1	-4.87035
9	1	-1	1	1	1	-1	-5.15461
10	1	-1	-1	-1	-1	1	4.342885
11	-1	1	1	1	-1	1	2.149369
12	-1	1	-1	-1	1	-1	5.012015
13	1	-1	1	-1	-1	-1	-6.33973
14	1	-1	-1	1	1	1	4.370357
15	-1	1	1	-1	1	1	1.766734
16	-1	1	-1	1	-1	-1	2.82219

Table 2.4 NC-6 results of the model for y=1.5+2.2A+4.5B+1.9F+2.9AF+ε

Delta	A	B	F	AF	mAIC
0.5	2.68431	4.500736	2.036185	2.976832	16.60565
1	2.65306	4.469486	2.027837	2.976832	16.60565
1.25	2.63743	4.453861	2.01742	2.976832	16.60565
1.5	2.62181	4.438236	2.007004	3.01138	16.60565
1.75	2.60618	4.422611	1.996587	2.98013	16.60565
2	2.59056	4.406986	1.98617	2.962856	16.60565
2.25	2.57493	4.391361	1.975754	2.947231	16.60565
2.5	2.55931	4.375736	1.965337	2.931606	16.60565
3	2.52806	4.344486	1.944504	2.901073	16.60565
3.25	2.51243	4.328861	1.934087	2.890656	16.60565
3.5	2.49681	4.313236	1.92367	2.880239	16.60565
3.75	2.48118	4.297611	1.913254	2.869823	16.60565
4	2.46556	4.281986	1.902837	2.857295	16.60565
4.520314	2.43304	4.249467	1.881157	2.824776	16.60565
8.040629	2.21302	4.029447	1.734477	2.604756	16.60565
11.56094	1.99300	3.809427	1.55234	2.384736	16.60565
15.08126	1.77298	3.589408	1.33232	2.164717	16.60565

It can be observed that the Dantzig selector method correctly identifies A, B, F, and AF as active factors without any type I or type II errors.

To test a model where two two factor interactions enter with strong heredity, another model **y=1.5+2.2A+4.5B+3.9C+2.3F-3.2AB+2.5AF+ ε,** assuming A, B, C, F, AB, AF as the active effects was generated and normal random noise **ε** ~ N (0, 1) was added to the mean response to generate the simulated data. The design table with the generated responses and the selected results are presented in Tables 2.5-2.6. Once again, models with minimum mAIC value correctly identify the active effects.

Table 2.5 NC-6 design with responses for y=1.5+2.2A+4.5B+3.9C+2.3F-3.2AB+2.5AF+ ε

Runs	A	B	C	D	E	F	y
1	1	1	1	1	1	1	13.75917
2	1	1	-1	-1	-1	-1	-3.21595
3	-1	-1	1	1	-1	-1	-4.13436
4	-1	-1	-1	-1	1	1	-12.4164
5	1	1	1	-1	1	-1	5.538028
6	1	1	-1	1	-1	1	5.265305
7	-1	-1	1	-1	-1	1	-5.45184
8	-1	-1	-1	1	1	-1	-12.0781
9	1	-1	1	1	1	-1	0.657267
10	1	-1	-1	-1	-1	1	3.601692
11	-1	1	1	1	-1	1	11.42639
12	-1	1	-1	-1	1	-1	2.152705
13	1	-1	1	-1	-1	-1	1.362273
14	1	-1	-1	1	1	1	2.742272
15	-1	1	1	-1	1	1	11.50986
16	-1	1	-1	1	-1	-1	2.98114

Table 2.6 NC-6 analysis for y=1.5+2.2A+4.5B+3.9C+2.3F-3.2AB+2.5AF+ ε

Delta	A	B	C	F	AB	AF	mAIC
0.5	2.088436	4.644177	3.85458	2.10700	-3.04173	2.2010	53.52532
0.75	2.088436	4.612927	3.82333	2.07575	-3.02611	2.1698	53.52532
1	2.0884364565	4.581677	3.79208	2.04450	3.01048	2.1385	53.52532
1.25	2.0877702955625	4.551094	3.76083	2.01325	2.99486	2.1073	53.52532
1.5	2.0721452955625	4.535469	3.72958	1.98200	2.97923	2.0760	53.52532
1.75	2.0565202955625	4.519844	3.69833	1.95075	-2.96361	2.0448	53.52532
2	2.0408952955625	4.504219	3.66708	1.91950	2.94798	2.0135	53.52532
2.25	2.0252702955625	4.488594	3.63583	1.88825	2.93236	1.9823	53.52532
2.5	2.0096452955625	4.472969	3.60458	1.88083	-2.91673	1.9510	53.52532
2.75	1.9940202955625	4.457344	3.57333	1.88083	-2.90111	1.9198	53.52532
3	1.9783952955625	4.441719	3.54208	1.88083	2.88548	1.8885	53.52532
3.25	1.9627702955625	4.426094	3.51083	1.88083	2.86986	1.8573	53.52532
3.5	1.9471452955625	4.410469	3.47958	1.88083	2.85423	1.8260	53.52532
3.75	1.9315202955625	4.394844	3.44833	1.88083	2.83861	1.7948	53.52532
4	1.9158952955625	4.379219	3.41708	1.88083	2.82298	1.7635	53.52532
4.25	1.9002702955625	4.363594	3.38583	1.88083	2.80736	1.7323	53.52532
4.5	1.8846452955625	4.347969	3.35458	1.88083	-2.79173	1.7010	53.52532
4.75	1.8690202955625	4.332344	3.32333	1.88044	-2.77611	1.6698	53.52532
5	1.8533952955625	4.316719	3.29208	1.87002	2.76048	1.6385	53.52532
5.25	1.8377702955625	4.301094	3.26083	1.85960	2.74486	1.6073	53.52532
5.5	1.8221452955625	4.285469	3.22958	1.84919	-2.72923	1.5760	53.52532
5.75	1.8065202955625	4.269844	3.19833	1.83877	-2.71361	1.5448	53.52532
6	1.7908952955625	4.254219	3.16708	1.82835	2.69798	1.5135	53.52532
8.15560	1.65617020313988	4.119494	2.89763	1.73854	2.56326	1.2441	53.52532
11.7334	1.43255765692857	3.895881	2.45040	1.58946	2.33964	0.7968	53.52532
15.3112	1.20894511071726	3.672268	2.00318	1.36639	-2.11603	0	73.37744
18.889	0.98533256450595	3.448656	1.64181	1.14278	-1.89242	0	73.37744
22.4668	0.76172001829464	3.225043	1.44795	0.91916	-1.66881	0	73.37744
26.0446	0.53810747208333	3.001431	1.22434	0.69555	-1.44519	0	73.37744
29.6224	0	2.777818	1.00073	0.47194	-1.22158	0	67.67144
33.2002	0	2.554206	0.77711	0	0.99797	0	60.99774
36.7780	0	2.374837	0.55350	0	0.77436	0	60.99774
40.3558	0	2.173625	0	0	0.55074	0	58.04743
43.9336	0	1.950013	0	0	0	0	57.58655
47.5114	0	1.7264	0	0	0	0	57.58655
51.0891	0	1.502788	0	0	0	0	57.58655
54.6671	0	1.279175	0	0	0	0	57.58655
58.2441	0	1.055563	0	0	0	0	57.58655
61.8226	0	0.83195	0	0	0	0	57.58655
65.4004	0	0.608338	0	0	0	0	57.58655

2.3. Analysis of the seven factor NC design using Dantzig selector

Three types of models, one with only main effects and two others with main effects and significant two-factor interactions were studied. To investigate a main effects only model using the NC-7 design, the following first-order model was generated: y=1.5+2.2A+4.5B+3.9C+1.9F + ε, where ε ~ N (0, 1). The design including the simulated responses and the results of the analysis for different delta values are presented in Tables 2.7 and 2.8. Once again, the minimum value of the mAIC criterion identifies the correct model

Table 2.7 NC-7 design table for y=1.5+2.2A+4.5B+3.9C+2.8F

Runs	A	B	C	D	E	F	G	Y
1	1	1	1	1	1	1	1	13.69104
2	1	1	1	-1	-1	-1	-1	10.15821
3	1	1	-1	1	1	-1	-1	1.387442
4	1	1	-1	-1	-1	1	1	7.538254
5	1	-1	1	1	-1	1	-1	4.892953
6	1	-1	1	-1	1	-1	1	0.786086
7	1	-1	-1	1	-1	-1	1	-7.74477
8	1	-1	-1	-1	1	1	-1	-2.97237
9	-1	1	1	1	1	1	-1	10.38366
10	-1	1	1	-1	-1	-1	1	3.046954
11	-1	1	-1	1	-1	1	1	1.987258
12	-1	1	-1	-1	1	-1	-1	-2.30385
13	-1	-1	1	1	-1	-1	-1	-4.85529
14	-1	-1	1	-1	1	1	1	0.315871
15	-1	-1	-1	1	1	-1	1	-9.50399
16	-1	-1	-1	-1	-1	1	-1	-8.22494

Table 2.8 NC-7 results for y=1.5+2.2A+4.5B+3.9C+2.8F+ ε

Delta	A	B	C	D	E	F	G	mAIC
0.5	2.274447	4.543464	3.609779	0	0	2.258809	0	28.99056
0.75	2.258822	4.527839	3.594154	0	0	2.243184	0	28.99056
1	2.243197	4.512214	3.578529	0	0	2.227559	0	28.99056
1.25	2.227572	4.496589	3.562904	0	0	2.211934	0	28.99056
1.5	2.211947	4.480964	3.547279	0	0	2.196309	0	28.99056
1.75	2.196322	4.465339	3.531654	0	0	2.180684	0	28.99056
2	2.180697	4.449714	3.516029	0	0	2.165059	0	28.99056
2.25	2.165072	4.434089	3.500404	0	0	2.149434	0	28.99056
2.5	2.149447	4.418464	3.484779	0	0	2.133809	0	28.99056
2.75	2.133822	4.402839	3.469154	0	0	2.118184	0	28.99056
3	2.118197	4.387214	3.453529	0	0	2.102559	0	28.99056
3.25	2.102572	4.371589	3.437904	0	0	2.086934	0	28.99056
3.5	2.086947	4.355964	3.422279	0	0	2.071309	0	28.99056
3.75	2.071322	4.340339	3.406654	0	0	2.055684	0	28.99056
4	2.055697	4.324714	3.391029	0	0	2.040059	0	28.99056
4.25	2.040072	4.309089	3.375404	0	0	2.024434	0	28.99056
4.5	2.024447	4.293464	3.359779	0	0	2.008809	0	28.99056
4.75	2.008822	4.277839	3.344154	0	0	1.993184	0	28.99056
5	1.993197	4.262214	3.328529	0	0	1.977559	0	28.99056
5.25	1.977572	4.246589	3.312904	0	0	1.961934	0	28.99056
5.5	1.961947	4.230964	3.297279	0	0	1.946309	0	28.99056
5.75	1.946322	4.215339	3.281654	0	0	1.930684	0	28.99056
6	1.930697	4.199714	3.266029	0	0	1.915059	0	28.99056
11.45649	1.589667	3.858684	2.924998	0	0	1.574028	0	28.99056
21.91298	0.936136	3.205153	2.271467	0	0	0.920497	0	28.99056
32.36947	0	2.551623	1.617937	0	0	0	0	46.92258
42.82596	0	1.898092	0.964406	0	0	0	0	46.92258
53.28245	0	1.244561	0	0	0	0	0	53.27399
63.73894	0	0.591031	0	0	0	0	0	53.27399

A two-factor interaction model y=1.5+2.5A+4.5C+3.2F+2.9AF+ ε with N (0, 1) errors was also investigated. The design and simulated responses is shown in Table 9. The results of the Dantzig selector analysis is shown in Table 2.10.

Table 2.9 NC-7 design table for y=1.5+2.5A+4.5C+3.2F+2.9AF

Runs	A	B	C	D	E	F	G	Y
1	1	1	1	1	1	1	1	16.02077
2	1	1	1	-1	-1	-1	-1	2.261747
3	1	1	-1	1	1	-1	-1	-7.59143
4	1	1	-1	-1	-1	1	1	5.699659
5	1	-1	1	1	-1	1	-1	11.77831
6	1	-1	1	-1	1	-1	1	2.090869
7	1	-1	-1	1	-1	-1	1	-6.07119
8	1	-1	-1	-1	1	1	-1	4.760928
9	-1	1	1	1	1	1	-1	4.087872
10	-1	1	1	-1	-1	-1	1	3.851497
11	-1	1	-1	1	-1	1	1	-3.73806
12	-1	1	-1	-1	1	-1	-1	-5.35274
13	-1	-1	1	1	-1	-1	-1	3.760795
14	-1	-1	1	-1	1	1	1	3.879211
15	-1	-1	-1	1	1	-1	1	-5.26657
16	-1	-1	-1	-1	-1	1	-1	-6.0011

Table 2.10 NC-7 results for y=1.5+2.5A+4.5C+3.2F+2.9AF

Delta	A	C	F	AF	EF	mAIC
0.5	2.076797599	4.228979206	3.019039	2.86467	0	29.05886
0.75	2.061172599125	4.166479206	3.003414	2.849045	0.428178	43.98173
1	2.045547599125	4.103979206	2.987789	2.83342	0.443803	43.98173
1.25	2.029922599125	4.041479206	2.961992	2.807623	0.459428	43.98173
1.5	2.014297599125	4.029909093	2.930742	2.776373	0.449588	43.98173
1.75	1.998672599125	4.029909093	2.908078	2.75371	0.433963	43.98173
2	**1.983047599125**	**4.029909093**	**2.892453**	**2.738085**	**0**	**29.05886**
2.25	**1.967422599125**	**4.029909093**	**2.876828**	**2.72246**	**0**	**29.05886**
2.5	**1.951797599125**	**4.029909093**	**2.861203**	**2.706835**	**0**	**29.05886**
2.75	**1.936172599125**	**4.029909093**	**2.845578**	**2.69121**	**0**	**29.05886**
3	**1.920547599125**	**4.029909093**	**2.829953**	**2.675585**	**0**	**29.05886**
3.25	**1.904922599125**	**4.029909093**	**2.814328**	**2.65996**	**0**	**29.05886**
3.5	**1.889297599125**	**4.029909093**	**2.798703**	**2.644335**	**0**	**29.05886**
3.75	**1.873672599125**	**4.029909093**	**2.783078**	**2.62871**	**0**	**29.05886**
4	**1.858047599125**	**4.029909093**	**2.767453**	**2.613085**	**0**	**29.05886**
4.25	**1.842422599125**	**4.029909093**	**2.751828**	**2.59746**	**0**	**29.05886**
4.5	**1.826797599125**	**4.029909093**	**2.736203**	**2.581835**	**0**	**29.05886**
4.75	**1.811172599125**	**4.029909093**	**2.720578**	**2.56621**	**0**	**29.05886**
5	**1.795547599125**	**4.027102365**	**2.704953**	**2.550585**	**0**	**29.05886**
5.25	**1.779922599125**	**4.016685699**	**2.689328**	**2.53496**	**0**	**29.05886**
5.5	**1.764297599125**	**4.006269032**	**2.673703**	**2.519335**	**0**	**29.05886**
5.75	**1.748672599125**	**3.995852365**	**2.658078**	**2.50371**	**0**	**29.05886**
6	**1.733047599125**	**3.985435699**	**2.642453**	**2.488085**	**0**	**29.05886**
8.638383	1.5681486878125	3.875503091	2.034929	1.88056	0	47.01734
11.18451	1.409015717375	3.756691291	1.568759	1.414391	0	47.01734
13.73064	1.2498827469375	3.59755832	1.09136	0.936992	0	68.24936
16.27677	**1.0907497765**	**3.43842535**	**2.000155**	**1.845787**	**0**	**29.05886**
18.82289	**0.9316168060625**	**3.279292379**	**1.841023**	**1.686654**	**0**	**29.05886**
21.36902	**0.772483835625**	**3.120159409**	**1.68189**	**1.527521**	**0**	**29.05886**
23.91515	**0.6133508651875**	**2.961026438**	**1.522757**	**1.368388**	**0**	**29.05886**
26.46128	**0.45421789475**	**2.801893468**	**1.363624**	**1.209255**	**0**	**29.05886**
29.0074	0	2.642760497	1.204491	1.050122	0	44.61054
31.55353	0	2.483627527	1.045358	0.890989	0	44.61054

For delta values less than 2, an inactive factor EF is also included as being active. However, this is not the minimum mAIC model. As with the earlier cases, the models with the minimum mAIC correctly identify the active factors.

To investigate a model with two two-factor interactions, the model y=1.5+2.9A+3.2C+2.5AC+2.2AB + ε with N (0, 1) random errors was constructed. Notice that this model contains one hierarchical interaction (AC) and one interaction that has one weak heredity (AB). The design with the simulated response values and the results of the Dantzig selector analysis are shown in tables 2.11 and 2.12. The minimum value of mAIC (41.4161) identifies the correct model without any type I or type II errors.

Table 2.11 NC-7 design for the model y=1.5+2.9A+3.2C+2.5AC+2.2AB + ε

Runs	A	B	C	D	E	F	G	Y
1	1	1	1	1	1	1	1	13.06209
2	1	1	1	-1	-1	-1	-1	11.60521
3	1	1	-1	1	1	-1	-1	-0.01448
4	1	1	-1	-1	-1	1	1	0.709439
5	1	-1	1	1	-1	1	-1	6.7633
6	1	-1	1	-1	1	-1	1	7.919788
7	1	-1	-1	1	-1	-1	1	-4.34065
8	1	-1	-1	-1	1	1	-1	-2.90166
9	-1	1	1	1	1	1	-1	-3.80524
10	-1	1	1	-1	-1	-1	1	-0.88725
11	-1	1	-1	1	-1	1	1	-3.69191
12	-1	1	-1	-1	1	-1	-1	-3.30848
13	-1	-1	1	1	-1	-1	-1	2.997959
14	-1	-1	1	-1	1	1	1	1.524573
15	-1	-1	-1	1	1	-1	1	0.761699
16	-1	-1	-1	-1	-1	1	-1	0.160675

Table 2.12 NC-7 results for y=1.5+2.9A+3.2C+2.5AC+2.2AB + ε

Delta	A	B	C	AB	AC	AE	AF	DE	DF	EG	mAIC
0.5	2.4094	0.0000	2.2322	0.0000	1.3003	0.3756	0.4587	1.0081	0.0000	2.1422	98.1185
0.75	2.3938	0.0000	2.2009	0.0000	1.3003	0.2600	0.4431	0.9925	0.0000	2.1422	98.1185
1	2.3782	0.0000	3.1465	1.9537	2.2771	0.3370	0.4274	0.0000	0.0000	0.0000	43.6431
1.25	2.3626	0.0000	3.0996	1.9224	2.2615	0.3214	0.4118	0.0000	0.0000	0.0000	43.6431
1.5	2.3469	0.0000	3.0528	1.8912	2.2459	0.3100	0.3962	0.0000	0.0000	0.2511	69.4156
1.75	2.3313	0.0000	3.0059	1.8599	2.2303	0.2986	0.3806	0.0000	0.0000	0.2823	69.4156
2	2.3157	0.0000	2.9590	1.8287	2.2146	0.2881	0.3649	0.0000	0.0000	0.3136	69.4156
2.25	2.3001	0.0000	2.9248	1.7974	2.1863	0.2589	0.3493	0.0000	0.0000	0.3448	69.4156
2.5	2.2844	0.0000	2.8936	1.7662	2.1551	0.2433	0.3337	0.0000	0.0000	0.3761	69.4156
3	2.2532	0.0000	2.8311	1.7037	2.0926	0.0000	0.2970	0.0000	0.0000	0.4386	54.9610
3.25	2.2376	0.0000	2.7998	1.6724	2.0613	0.0000	0.2657	0.0000	0.0000	0.4698	54.9610
3.5	**2.2219**	**0.000**	**2.768**	**1.6412**	**2.0301**	**0.000**	**0.000**	**0.000**	**0.000**	***0.5011***	**41.4161**
4	**2.1907**	**0.000**	**2.706**	**1.5787**	**1.9676**	**0.000**	**0.000**	**0.000**	**0.000**	***0.563***	**41.4161**
4.25	**2.1751**	**0.000**	**2.674**	**1.5474**	**1.9363**	**0.0000**	**0.000**	**0.000**	**0.000**	***0.594***	**41.4161**
4.5	**2.159**	**0.000**	**2.643**	**1.5162**	**1.9051**	**0.000**	**0.000**	**0.000**	**0.000**	***0.6261***	**41.4161**
5	**2.128**	**0.000**	**2.5811**	**1.4537**	**1.842**	**0.000**	**0.000**	**0.000**	**0.000**	***0.688***	**41.4161**
5.25	**2.112**	**0.000**	**2.549**	**1.4224**	**1.8113**	**0.000**	**0.000**	**0.000**	**0.000**	***0.7198***	**41.4161**
5.5	**2.096**	**0.000**	**2.518**	**1.3912**	**1.7801**	**0.000**	**0.000**	**0.000**	**0.000**	***0.7511***	**41.4161**
6	**2.065**	**0.000**	**2.456**	**1.3287**	**1.7176**	**0.000**	**0.000**	**0.000**	**0.000**	***0.8136***	**41.4161**
7.7538	**1.956**	**0.000**	**2.236**	**1.1094**	**1.498**	**0.000**	**0.000**	**0.000**	**0.000**	***1.0328***	**41.4161**
10.005	**1.815**	**0.828**	**0.738**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.3889**	***3.3592***	**41.4161**
12.256	**1.674**	**0.546**	**0.738**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.3889**	***3.0777***	**41.4161**
14.5076	1.5340	0.0000	0.7385	0.0000	0.0000	0.0000	0.0000	0.0000	0.3889	2.7963	50.0715
16.7589	1.3933	0.0000	0.7385	0.0000	0.0000	0.0000	0.0000	0.0000	0.3808	2.5230	50.0715
19.0101	1.2526	0.0000	0.7385	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	2.3823	41.9006
23.5127	0.9711	0.0000	0.7110	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	2.1147	41.9006
30.2665	0.5490	0.0000	0.4296	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.8333	41.9006
34.7690	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.6457	42.4815
41.5228	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.3445	42.4815
46.0254	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0631	42.4815

For investigating a model where two two-factor interactions enter with strong heredity, the model, **y=1.5+2.4A+2.8B-3.1D+2.5AB+2.2AD + ε** with N (0, 1) errors was constructed. The design with the simulated responses and the result from the Dantzig selector analysis for different delta values are presented in tables 2.13 and 2.14. As with the earlier cases, the minimum value of mAIC (22.334) identifies the active factors without any errors.

Table 2.13. NC-7 Design for the Model y= y=1.5+2.4A+2.8B-3.1D+2.5AB+2.2AD+ ε

Runs	A	B	C	D	E	F	G	Y
1	1	1	1	1	1	1	1	7.4192
2	1	1	1	-1	-1	-1	-1	9.9557
3	1	1	-1	1	1	-1	-1	8.6761
4	1	1	-1	-1	-1	1	1	10.11
5	1	-1	1	1	-1	1	-1	-2.058
6	1	-1	1	-1	1	-1	1	-1.528
7	1	-1	-1	1	-1	-1	1	-1.213
8	1	-1	-1	-1	1	1	-1	-0.623
9	-1	1	1	1	1	1	-1	-5.617
10	-1	1	1	-1	-1	-1	1	4.9566
11	-1	1	-1	1	-1	1	1	-5.699
12	-1	1	-1	-1	1	-1	-1	4.4189
13	-1	-1	1	1	-1	-1	-1	-6.624
14	-1	-1	1	-1	1	1	1	3.7118
15	-1	-1	-1	1	1	-1	1	-6.209
16	-1	-1	-1	-1	-1	1	-1	3.8419

Table 2.14. NC-7 Results for y=1.5+2.4A+2.8B-3.1D+2.5AB+2.2AD + ε

Delta	A	B	D	AB	AD	BF	CG	mAIC
0.5	2.34	2.7	-2.76	2.29	2.12	0	0	22.335
0.75	2.33	2.69	-2.77	2.27	2.13	0	0	22.335
1	2.31	2.67	-2.71	2.26	2.07	0	0	22.335
1.25	2.29	2.66	-2.69	2.24	2.05	0	0	22.335
1.5	2.28	2.64	-2.66	2.22	2.03	0	0	22.335
1.75	2.26	2.63	-2.64	2.21	2	0	0	22.335
2	2.25	2.61	-2.61	2.19	1.97	-0.3	0	43.354
2.25	2.23	2.59	-2.58	2.18	1.94	-0.3	0	43.354
2.5	2.22	2.58	-2.55	2.16	1.91	-0.4	0	43.354
2.75	2.2	2.56	-2.64	2.15	2	0	0	22.335
3	2.18	2.55	-2.63	2.13	1.99	0	0	22.335
3.25	2.17	2.53	-2.45	2.11	1.81	-0.5	0	43.354
3.5	2.15	2.52	-2.4	2.1	1.76	-0.5	0	43.354
3.75	2.14	2.5	-2.35	2.08	1.72	-0.5	0	43.354
4	2.12	2.48	-2.31	2.07	1.67	-0.6	0	43.354
4.25	2.11	2.47	-2.26	2.05	1.62	-0.6	0	43.354
4.5	2.09	2.45	-2.21	2.04	1.58	-0.6	0	43.354
5	2.06	2.42	-2.12	2.01	1.48	-0.7	0	43.354
5.25	2.04	2.41	-2.07	1.99	1.44	-0.7	0	43.354
5.5	2.03	2.39	-2.03	1.97	1.39	-0.7	0.3	61.69
6	2	2.36	-1.93	1.94	1.3	-0.8	0.4	61.69
7.595	1.9	2.26	-2.06	1.84	1.42	0	0.6	37.244
10.89	1.69	2.05	-1.01	1.64	0.38	-1.4	1	61.69
12.54	1.59	1.95	-0.7	1.53	0	-1.6	1.2	88.903
15.84	1.38	1.74	-1.82	1.33	1.19	0	0	22.335
17.49	1.28	1.64	-1.72	1.22	1.08	0	0	22.335
20.79	1.07	1.44	-1.51	1.02	0.88	0	0	22.335
24.08	0.87	1.23	-1.31	0.81	0.67	0	0	22.335
27.38	0.66	1.02	-1.1	0.61	0.47	0	0	22.335
30.68	0.45	0.82	-0.9	0.4	0	0	0	58.491
32.33	0.35	0.71	-0.79	0.3	0	0	0	58.491
35.63	0	0.51	-0.6	0	0	0	0	52.93
38.92	0	0.34	-0.45	0	0	0	0	52.93
40.57	0	0.27	-0.35	0	0	0	0	52.93

2.4. Analysis of the eight factor NC design using Dantzig selector

For investigating the utility of the Dantzig selector with the NC-8 design, three random models were created and standard normal error was added to the mean at each treatment combination to produce simulated responses. The responses and the results of the model with four main effects are presented in Tables 2.15 and 2.16. As shown in Table 14, the minimum value of mAIC correctly identifies the four active main effects.

Table 2.15 NC-8 design table for y=1.5+2.2A+4.5B+3.9C+2.8F + ε

Runs	A	B	C	D	E	F	G	H	Y
1	1	1	1	1	1	1	1	1	15.44734
2	1	1	1	1	-1	-1	-1	-1	10.92649
3	1	1	-1	-1	1	1	-1	-1	7.261216
4	1	1	-1	-1	-1	-1	1	1	0.514923
5	1	-1	1	-1	1	-1	1	-1	-0.29637
6	1	-1	1	-1	-1	1	-1	1	5.78224
7	1	-1	-1	1	1	-1	-1	1	-7.02248
8	1	-1	-1	1	-1	1	1	-1	-0.38851
9	-1	1	1	1	1	1	1	1	11.25737
10	-1	1	1	-1	1	-1	-1	-1	4.477287
11	-1	1	-1	1	-1	-1	1	-1	-3.40267
12	-1	1	-1	-1	-1	1	-1	1	1.316205
13	-1	-1	1	1	-1	-1	-1	1	-3.73184
14	-1	-1	1	-1	-1	1	1	-1	3.06383
15	-1	-1	-1	1	1	1	-1	-1	-7.81701
16	-1	-1	-1	-1	1	-1	1	1	-12.4118

Table 2.16 NC-8 results for y=1.5+2.2A+4.5B+3.9C+2.8F + ε

Delta	A	B	C	F	mAIC
0.5	2.4358386749165	4.382504	4.273527	2.898069	25.53061
0.75	2.4202136749165	4.366879	4.257902	2.882444	25.53061
1	2.4045886749165	4.351254	4.242277	2.866819	25.53061
1.25	2.3889636749165	4.335629	4.226652	2.851194	25.53061
1.5	2.3733386749165	4.320004	4.211027	2.835569	25.53061
1.75	2.3577136749165	4.304379	4.195402	2.819944	25.53061
2	2.3420886749165	4.288754	4.179777	2.804319	25.53061
2.25	2.3264636749165	4.273129	4.164152	2.788694	25.53061
2.5	2.3108386749165	4.257504	4.148527	2.773069	25.53061
2.75	2.2952136749165	4.241879	4.132902	2.757444	25.53061
3	2.2795886749165	4.226254	4.117277	2.741819	25.53061
3.25	2.2639636749165	4.210629	4.101652	2.726194	25.53061
3.5	2.2483386749165	4.195004	4.086027	2.710569	25.53061
3.75	2.2327136749165	4.179379	4.070402	2.694944	25.53061
4	2.2170886749165	4.163754	4.054777	2.679319	25.53061
4.25	2.2014636749165	4.148129	4.039152	2.663694	25.53061
4.5	2.1858386749165	4.132504	4.023527	2.648069	25.53061
4.75	2.1702136749165	4.116879	4.007902	2.632444	25.53061
5	2.1545886749165	4.101254	3.992277	2.616819	25.53061
5.25	2.1389636749165	4.085629	3.976652	2.601194	25.53061
5.5	2.1233386749165	4.070004	3.961027	2.585569	25.53061
5.75	2.1077136749165	4.054379	3.945402	2.569944	25.53061
6	2.0920886749165	4.038754	3.929777	2.554319	25.53061
9.827509	1.85286936829909	3.799535	3.690558	2.3151	25.53061
18.65502	1.30115006168169	3.247816	3.138838	1.763381	25.53061
27.48253	0.749430755064277	2.696097	2.587119	1.211661	25.53061
36.31004	0	2.144377	2.0354	0.659942	48.56203
45.13754	0	1.592658	1.48368	0	51.68206
53.96505	0	1.040939	0.931961	0	51.68206
62.79256	0	0.489219	0	0	58.35857

The second model contains three active factors, A, C and F, and a two-factor interaction with strong heredity. The design with the simulated responses is shown in Table 2.17. The results of using the Dantzig selector are in Table 2.18. These results show that the model with the minimum mAIC value of 36.14368 correctly identifies the active factors.

Table 2.17 NC-8 design table for y=1.5+2.5A+4.5C+3.2F+2.9AF + ε

Runs	A	B	C	D	E	F	G	H	Y
1	1	1	1	1	1	1	1	1	15.01554
2	1	1	1	1	-1	-1	-1	-1	2.124344
3	1	1	-1	-1	1	1	-1	-1	5.315324
4	1	1	-1	-1	-1	-1	1	1	-7.31853
5	1	-1	1	-1	1	-1	1	-1	4.025293
6	1	-1	1	-1	-1	1	-1	1	13.69961
7	1	-1	-1	1	1	-1	-1	1	-6.74345
8	1	-1	-1	1	-1	1	1	-1	6.286797
9	-1	1	1	1	1	1	1	1	3.974207
10	-1	1	1	-1	1	-1	-1	-1	4.43829
11	-1	1	-1	1	-1	-1	1	-1	-4.63541
12	-1	1	-1	-1	-1	1	-1	1	-3.37166
13	-1	-1	1	1	-1	-1	-1	1	2.992898
14	-1	-1	1	-1	-1	1	1	-1	2.256046
15	-1	-1	-1	1	1	1	-1	-1	-5.82062
16	-1	-1	-1	-1	1	-1	1	1	-5.80214

Table 2.18 NC-8 results for y=1.5+2.5A+4.5C+3.2F+2.9AF+ε

delta	A	C	F	AF	BE	EG	FH	mAIC
0.5	2.367081862375	3.101022	2.572449	2.566908	0.827	0.450707	1.051352	101.824
0.75	2.351456862375	3.163522	2.546382	2.592975	0.796	0.388207	1.020102	80.65336
1	2.335831862375	3.221213	2.570045	2.564503	0.769	0.330516	0.988852	101.824
1.25	2.320206862375	3.262829	2.559603	2.554061	0.759	0	0.957602	55.52304
1.5	2.304581862375	3.497921	2.591628	2.586087	0.663	0	0.926352	55.52304
1.75	2.288956862375	3.387829	2.520957	2.515415	0.774	0	0.864628	55.52304
2	2.273331862375	3.3738	2.498318	2.492776	0.788	0	0.847406	55.52304
2.25	2.257706862375	3.34255	2.467068	2.461526	0.819	0	0.847406	55.52304
2.5	2.242081862375	3.3113	2.435818	2.430276	0.85	0	0.847406	55.52304
2.75	2.226456862375	3.202747	2.365916	2.360374	0.959	0	0.92471	55.52304
3	2.210831862375	3.031693	2.264764	2.259222	1.13	0	0.95596	55.52304
3.25	2.195206862375	3.21755	2.342068	2.336526	0.944	0	0.847406	55.52304
3.5	2.179581862375	3.1863	2.310818	2.305276	0.975	0	0.847406	55.52304
3.75	2.163956862375	3.15505	2.279568	2.274026	1.006	0	0.847406	55.52304
4	2.148331862375	3.1238	2.248318	2.242776	1.038	0	0.847406	55.52304
4.25	**2.1327068623**	**3.939956**	**2.084666**	**2.079124**	**0**	**0**	**0**	**36.14368**
4.5	**2.11708186237**	**3.908706**	**2.037791**	**2.032249**	**0**	**0**	**0**	**36.14368**
4.75	**2.1014568623**	**3.877456**	**2.154568**	**2.149026**	**0**	**0**	**0**	**36.14368**
5	2.085831862375	3.846206	2.123318	2.117776	0.315	0	0	55.52304
5.25	2.070206862375	3.814956	2.092068	2.086526	0.346	0	0	55.52304
5.5	2.054581862375	3.783706	2.060818	2.055276	0.378	0	0	55.52304
5.75	2.038956862375	3.752456	2.029568	2.024026	0.409	0	0	55.52304
6	2.023331862375	3.721206	1.998318	1.992776	0.44	0	0	55.52304
7.793407	1.911243924052	3.280288	1.665771	1.660229	0.881	0	0	55.52304
10.05788	1.7697146112777	2.714171	1.241183	1.235641	1.447	0	0	79.27787
12.32235	1.628185298503	2.930913	1.208024	1.202483	1.23	0	0	55.52304

The same model with two two-factor interactions that was used with the 7-factor design was used to investigate the performance of the Dantzig selector for the 8-factor design. Tables 2.19 shows the design and responses and Table 2.20 provide the results of applying the Dantzig selector. As observed the earlier cases, minimum mAIC value correctly identifies the active effects.

Table 2.19 NC-8 design for model y=1.5+2.9A+3.2C+2.5AC+2.2AB+ε

Runs	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**Y**
1	1	1	1	1	1	1	1	1	11.31217
2	1	1	1	1	-1	-1	-1	-1	13.25942
3	1	1	-1	-1	1	1	-1	-1	-0.22
4	1	1	-1	-1	-1	-1	1	1	1.317641
5	1	-1	1	-1	1	-1	1	-1	5.540407
6	1	-1	1	-1	-1	1	-1	1	8.456294
7	1	-1	-1	1	1	-1	-1	1	-4.06881
8	1	-1	-1	1	-1	1	1	-1	-4.34169
9	-1	1	1	1	1	1	1	1	-3.18056
10	-1	1	1	-1	1	-1	-1	-1	-3.07519
11	-1	1	-1	1	-1	-1	1	-1	-4.17996
12	-1	1	-1	-1	-1	1	-1	1	-4.79309
13	-1	-1	1	1	-1	-1	-1	1	2.461678
14	-1	-1	1	-1	-1	1	1	-1	0.717867
15	-1	-1	-1	1	1	1	-1	-1	0.846125
16	-1	-1	-1	-1	1	-1	1	1	0.531216

Table 2.20 NC-8 results for y=1.5+2.9A+3.2C+2.5AC+2.2AB+ε

Delta	A	C	AB	AC	BD	CE	EF	EG	mAIC
0.5	2.589	3.092	2.448	2.527	0.000	0.000	0.000	0.000	35.042
0.75	**2.574**	**3.020**	**2.389**	**2.455**	**0.000**	**0.000**	**0.000**	**0.000**	**23.460**
1	**2.558**	**3.043**	**2.307**	**2.478**	**0.000**	**0.000**	**0.000**	**0.000**	**23.460**
1.5	**2.527**	**3.012**	**2.283**	**2.447**	**0.000**	**0.000**	**0.000**	**0.000**	**23.460**
1.75	**2.511**	**2.996**	**2.184**	**2.431**	**0.000**	**0.000**	**0.000**	**0.000**	**23.460**
2	2.496	2.981	2.069	2.416	0.000	-0.394	0.000	0.000	39.673
2.25	2.480	2.965	2.077	2.400	0.000	-0.362	0.000	0.000	39.673
2.5	2.464	2.949	2.085	2.385	0.000	-0.331	0.000	0.000	39.673
2.75	2.449	2.914	2.027	2.349	0.000	0.000	0.317	0.000	40.377
3	**2.433**	**2.918**	**2.190**	**2.353**	**0.000**	**0.000**	**0.000**	**0.000**	**23.460**
3.25	**2.417**	**2.903**	**2.174**	**2.338**	**0.000**	**0.000**	**0.000**	**0.000**	**23.460**
4	**2.371**	**2.856**	**2.129**	**2.291**	**0.000**	**0.000**	**0.000**	**0.000**	**23.460**
4.25	**2.355**	**2.829**	**2.123**	**2.264**	**0.000**	**0.000**	**0.000**	**0.000**	**23.460**
4.5	**2.339**	**2.824**	**2.116**	**2.260**	**0.000**	**0.000**	**0.000**	**0.000**	**23.460**
5	**2.308**	**2.793**	**2.107**	**2.228**	**0.000**	**0.000**	**0.000**	**0.000**	**23.460**
5.25	**2.292**	**2.778**	**2.107**	**2.213**	**0.000**	**0.000**	**0.000**	**0.000**	**23.460**
5.5	**2.277**	**2.762**	**2.107**	**2.197**	**0.000**	**0.000**	**0.000**	**0.000**	**23.460**
6.6	**2.208**	**2.693**	**2.073**	**2.128**	**0.000**	**0.000**	**0.000**	**0.000**	**23.460**
8.0	2.121	2.480	1.992	1.915	0.000	0.000	0.000	0.340	41.352
10.8	1.946	1.785	1.817	1.221	0.706	0.000	0.000	0.673	62.827
12.2	1.858	1.903	1.729	1.338	0.881	0.000	0.000	0.000	40.939
13.6	1.771	1.728	1.642	1.163	1.056	0.000	0.000	0.000	40.939
15	**1.683**	**2.168**	**1.554**	**1.603**	**0.000**	**0.000**	**0.000**	**0.000**	**23.460**
16.400	1.595	1.377	1.467	0.813	1.406	0.000	0.000	0.000	40.939
19.2	**1.420**	**1.906**	**1.292**	**1.341**	**0.000**	**0.000**	**0.000**	**0.000**	**23.460**
22	**1.245**	**1.731**	**1.117**	**1.166**	**0.000**	**0.000**	**0.000**	**0.000**	**23.460**
24.8	**1.070**	**1.556**	**0.942**	**0.991**	**0.000**	**0.000**	**0.000**	**0.000**	**23.460**
26.2	**0.983**	**1.468**	**0.854**	**0.903**	**0.000**	**0.000**	**0.000**	**0.000**	**23.460**
29	**0.808**	**1.293**	**0.679**	**0.728**	**0.000**	**0.000**	**0.000**	**0.000**	**23.460**
30.4	**0.720**	**1.206**	**0.592**	**0.641**	**0.000**	**0.000**	**0.000**	**0.000**	**23.460**
33.2	**0.545**	**1.031**	**0.417**	**0.466**	**0.000**	**0.000**	**0.000**	**0.000**	**23.460**
34.6	**0.458**	**0.943**	**0.329**	**0.378**	**0.000**	**0.000**	**0.000**	**0.000**	**23.460**
36.0	0.370	0.856	0.000	0.000	0.000	0.000	0.000	0.000	49.617
38.8	0.000	0.681	0.000	0.000	0.000	0.000	0.000	0.000	50.205
40.2	0.000	0.593	0.000	0.000	0.000	0.000	0.000	0.000	50.205

To test a model where two two-factor interactions enter with strong heredity, another model $y= 1.5+2.9A+3.2B+2.8C+2.5AB-2.2AC+\varepsilon$, assuming A, B, C, AB, and AC as the active effects was created and normal random noise $\varepsilon \sim N(0, 1)$ was added to the mean response to generate the simulated data. The design with the generated responses and the analysis results are presented in Tables 2.21 and 2.22. The models with the minimum value of the mAIC criterion correctly identify the active effects.

Table 2.21. NC-8 Design for $y=1.5+2.9A+3.2B+2.8C+2.5AB-2.2AC+\varepsilon$

Runs	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**Y**
1	1	1	1	1	1	1	1	1	10.9434
2	1	1	1	1	-1	-1	-1	-1	11.3763
3	1	1	-1	-1	1	1	-1	-1	9.82478
4	1	1	-1	-1	-1	-1	1	1	9.2585
5	1	-1	1	-1	1	-1	1	-1	-0.95296
6	1	-1	1	-1	-1	1	-1	1	-0.34907
7	1	-1	-1	1	1	-1	-1	1	-2.31745
8	1	-1	-1	1	-1	1	1	-1	-0.71584
9	-1	1	1	1	1	1	1	1	3.44992
10	-1	1	1	-1	1	-1	-1	-1	5.46637
11	-1	1	-1	1	-1	-1	1	-1	-5.41717
12	-1	1	-1	-1	-1	1	-1	1	-5.30063
13	-1	-1	1	1	-1	-1	-1	1	2.73516
14	-1	-1	1	-1	-1	1	1	-1	2.20055
15	-1	-1	-1	1	1	1	-1	-1	-5.14266
16	-1	-1	-1	-1	1	-1	1	1	-7.20666

Table 2.22. NC-8 Results for the Model y=1.5+2.9A+3.2B+2.8C+2.5AB-2.2AC+ε

Delta	A	B	C	AB	AC	DH	EF	GH	mAIC
0.5	2.861	3.002	2.346	2.301	-1.725	0	0	0	36.784
0.75	2.846	2.987	2.331	2.285	-1.71	0	0	0	36.784
1	2.83	2.971	2.315	2.269	-1.694	0	0	0	36.784
1.25	2.815	2.877	2.3	2.176	-1.679	0	0	0.45	57.125
1.5	2.799	2.757	2.284	2.055	-1.663	0	0	0.472	57.125
1.75	2.783	2.718	2.268	2.016	-1.647	0	0	0.488	57.125
2	2.768	2.783	2.298	2.082	-1.586	0	0.308	0	58.722
2.25	2.752	2.893	2.267	2.191	-1.586	0	0	0	36.784
2.5	2.736	2.877	2.088	2.176	-1.467	0	0	0	57.14
2.75	2.721	2.862	2.057	2.16	-1.436	0	0	0	57.14
3	2.705	2.846	2.182	2.144	-1.561	0	0	0	36.784
3.25	2.69	2.83	2.175	2.129	-1.554	0	0	0	36.784
3.5	2.674	2.815	2.159	2.113	-1.538	0	0	0	36.784
4	2.643	2.534	2.128	1.832	-1.507	0	0.558	0	58.722
4.5	2.611	2.343	2.096	1.642	-1.475	0	0.51	0.66	83.056
5	2.58	2.281	2.065	1.579	-1.444	0	0.573	0.66	83.056
5.5	2.549	2.432	2.034	1.731	-1.413	0	0.573	0	58.722
6	2.518	2.658	1.972	1.957	-1.351	0	0	0	36.784
8.132	2.384	2.525	1.705	1.824	-1.084	0.329	0	0	57.508
10.985	2.206	2.347	1.348	1.645	-0.727	0.685	0	0	57.508
12.411	2.117	1.607	1.17	0.905	-0.549	0.864	1.36	0	83.441
15.264	1.939	2.079	1.424	1.378	-0.803	0	0	0	36.784
16.69	1.85	1.99	1.335	1.289	-0.714	0	0	0	36.784
18.117	1.76	1.901	1.245	1.2	-0.624	0	0	0	36.784
19.543	1.671	1.812	1.156	1.111	-0.535	0	0	0	36.784
20.97	1.582	1.723	1.067	1.021	-0.446	0	0	0	36.784
23.822	1.404	1.545	0.889	0.843	0	0	0	0	55.799
25.249	1.315	1.455	0.8	0.754	0	0	0	0	55.799
28.102	1.136	1.277	0.621	0.576	0	0	0	0	55.799
30.954	0.958	1.099	0.532	0.397	0	0	0	0	55.799
32.381	0.869	1.01	0.487	0.308	0	0	0	0	55.799
35.234	0.691	0.831	0.387	0	0	0	0	0	55.947
38.086	0.512	0.677	0	0	0	0	0	0	53.858
40.939	0.334	0.57	0	0	0	0	0	0	53.858
45.218	0	0.383	0	0	0	0	0	0	54.09

2.5. Conclusion

The examples presented in this chapter illustrates the effectiveness of the Dantzig selector in identifying the active factor effects for several different types of models containing both main effects and interactions for the 16-run no confounding designs in 6, 7, and 8 factors. We observe that the mAIC is a good criterion and that the minimum value of mAIC is effective in identifying the correct model. The choice of δ plays a major role as expected. Even though δ was varied between 0-100 in all the cases, smaller δ (~<10) seems to be adequate to identify the model that has the minimum value of mAIC. The true models tested in this study have about half of the main effects active and either one or two active two-factor interactions. All of the models studied have significant factors with relatively large effects, at least two standard deviations in magnitude. An obvious useful extension of this work would be to investigate how the Dantzig selector performs in situations where more effects are active and smaller factor effects are present. However, since the primary objective of screening designs is to detect factor with large effects, we think this work clearly illustrates the utility of the Dantzig selector in this situation.

Chapter 3

DANTZIG SELECTOR ANALYSIS USING SIMULATION STUDY

3.1 Background

Dantzig selector introduced by Candes and Tao (2007) has received considerable attention in the recent years for its property of identifying sparse signals from huge data. Consequently, studies were conducted to demonstrate the effectiveness of Dantzig selector for factor screening experiments. Phoa et al. (2009) analyzed supersaturated designs using Dantzig selector and pointed out that this method is powerful, while also being simple and fast to use.

In this chapter we evaluate the performance of Dantzig selector as a model selection method for analyzing no-confounding (NC) designs in 6,7 and 8 factors. NC designs are non-regular designs introduced by Jones and Montgomery (2010) and they have no complete confounding of two factor interactions. Hence they could allow for estimation of models containing main effects and interactions without having the need to run follow-up experiments like other regular designs.

A comprehensive study is carried out to study the effectiveness of using Dantzig selector for NC designs in 6,7 and 8 factors in this chapter. Also we introduce an alternate 6 factor NC design in 16 runs and present a comparative study on the performance of NC6 and the alternate NC6 designs.

3.2 Dantzig selector and choice of parameters

The Dantzig selector performs variable selection by shrinking the estimated regression coefficients towards zero. The estimated β's are the solution to the below regularization problem,

$$\min_{\hat{\beta}\in R^k} \|\hat{\beta}\|_{l1} \text{ subject to } \|X^t r\|_{l_\infty} \leq \delta$$

where **r** is the residual vector $\mathbf{r}=\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}$. $||\beta||_1 = |\beta_0| + \ldots + |\beta_k|$ is the l_1 norm and $||\beta||_\infty = \max(|\beta_0|, \ldots, |\beta_k|)$ is the l_∞ norm. Being a form of penalized regression, the method requires selection of tuning parameter δ.

Selection of an appropriate tuning parameter is crucial. The procedure recommended by Phoa et. al (2009) to select the δ automatically was used in the simulation study. As the first step, the Dantzig selector estimates of **β** are obtained for a wide range of δ. Secondly, a list of models is obtained for a fixed value of $\gamma \geq 0$. The models obtained are then compared based on a criterion and then the value of δ that yields the best model is selected.

The value of γ is the threshold between the signal and the noise and hence affects the selection of active factors. Since no prior knowledge of the magnitude of coefficients is available, for the simulations, the value for γ was set to be 0.1 of the largest |β| in the model when δ=0, as recommended by Phoa et al (2009).

The modified AIC criterion was used as the model selection criterion. Phoa et al (2009) demonstrated that since effect sparsity is an important assumption while using Dantzig selector, imposing a heavy penalty on the model complexity worked well in their simulation. However, our preliminary studies showed that the model chosen based on minimum mAIC value selected too few active factors. Hence, we decided to pick two models from each simulation. One based on the minimum value of mAIC and another one based on the second minimum value of mAIC. The following sections show that this approach reduced the type II error rate considerably.

3.3 Simulation set-up

Since the Dantzig selector is not incorporated into any statistical package yet, JMP was used to code and perform the simulations. The true model is unknown in general; hence each of the designs was tested for main effects only, main effects + 1 hierarchical interaction and main effects + 2 hierarchical interaction cases. Additionally, three different coefficient / noise ratios were tested for each of the above models. The true models for each of the cases were kept constant across the NC 6, 7 and 8 factor designs. The different settings of the simulation parameters are listed in Table 3.1.

Table 3.5 Simulation settings

True Model	Main effects only, Main effects+1 hierarchical interaction, Main effects+2 hierarchical interaction	
No. of variables	6, 7, 8	
Coefficient/Noise ratio	1, 2, 3	
Number of active factors	*Main effects only models*	2, 3, 4
	Main effects + 1 hierarchical interaction	2+1, 3+1, 4+1
	Main effects + 2 hierarchical interaction	3+2, 4+2

For every combination of factor types and coefficient/noise ratios, 5000 iterations were performed for each run. For every iteration, the true model was kept constant and random normal error (N (0, 1)) was added to the responses. Two models are chosen for each iteration based on the values of minimum mAIC and second minimum mAIC value. The results of each simulation run were evaluated by calculating the following parameters separately for the two models chosen:

π_1: Percentage of runs where the true model was identified

π_2: Percentage of runs where true model + some inactive factors were identified (Type I error)

π_3 : Percentage of runs where some or all of the active factors were missed (Type II error)

Ideally it is desired to have π_1 close to 100% and the others close to zero, since we are looking at factor screening experiments. Also, in general, experimenters tend to have less tolerance towards type II errors as once active factors are excluded from the initial stages, it is unlikely to be reconsidered in the next stage. Apart from calculating the individual type II error percentage for the models chosen from each iteration, the combined type II error percentage - where either of the two models chosen missed one or more active factors was also calculated.

3.4 Six factor designs

Figures 3.1-3.3 display the graphical summaries of the type II error rate for different coefficient sizes based on the noise ratio. The range of delta value was varied between 1 and 100. When there are only large or medium sized main effects in the model, the error rate is zero for true models containing only main effects. But when the effect sizes are small, the error rate is high in all the cases tested.

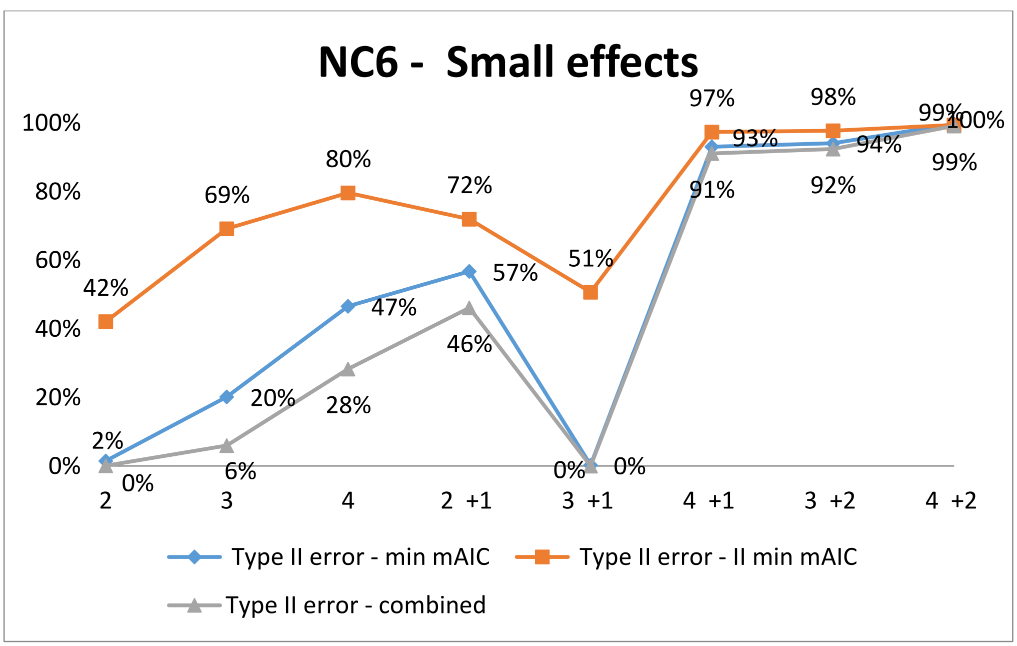

Figure 3.1 NC-6 factor design with Small effects

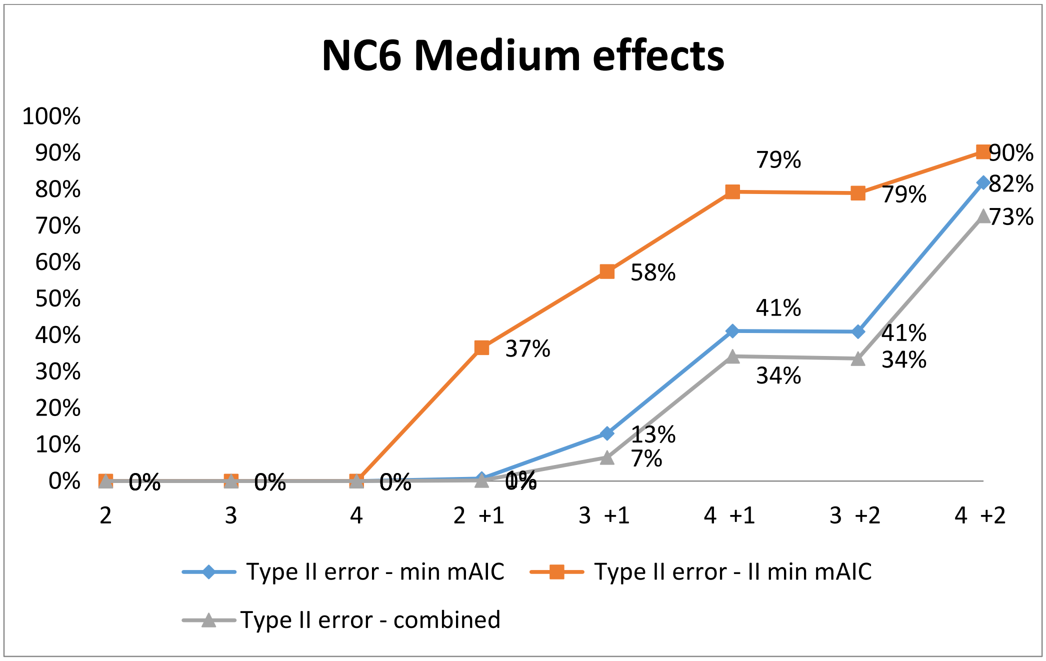

Figure 3.2 NC-6 factor design with Medium effects

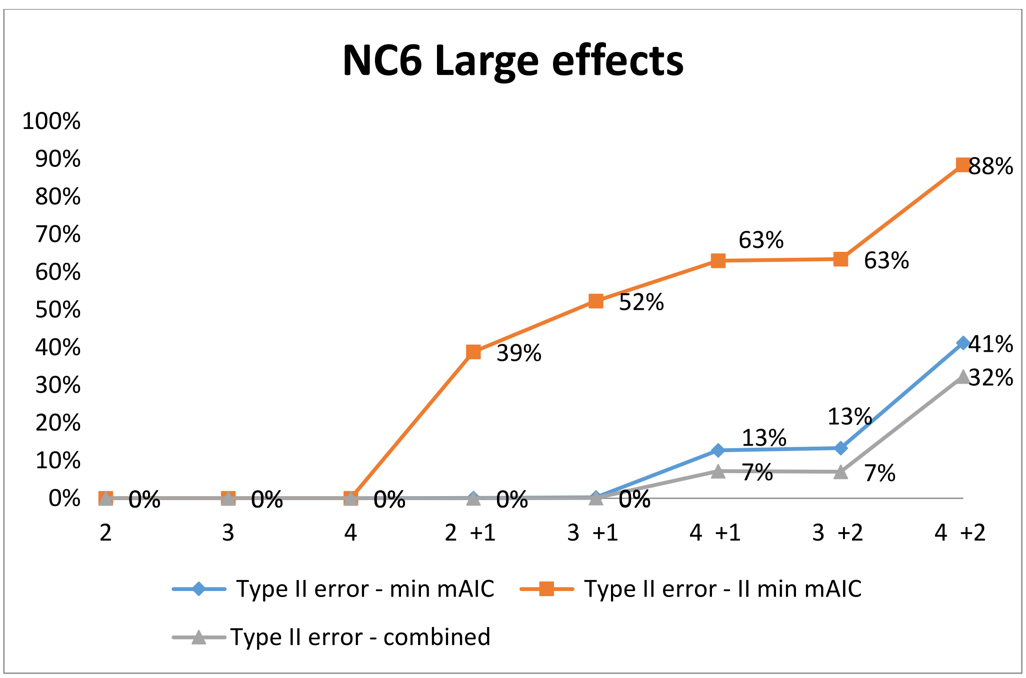

Figure 3.3 NC-6 factor designs with large effects

It was also observed that the models selected based on the second minimum value of mAIC have higher error rates. However, the combined error rate, which denotes the percentage of runs where both models fail to identify the active factors, is considerably lower. As the number of factors increases, the error rates in all the three cases go beyond 80%. This could be due to wide range of delta values, which are much higher in magnitude than the coefficient sizes. Hence another set of simulations was run by restricting the range of delta.

Figures 3.4-3.5 show the summary of error rate when the range of delta is varied between 0-10. In the case of medium and large effects, the error rate is zero for models with only main effects as well as for models with interaction. When the effect sizes are small, the combined error rate in the all cases is <10%. Hence it is evident that when the delta value is chosen reasonably, the Dantzig selector performs well in identifying the active factors. For the models selected based on the second minimum mAIC, high type I error rates are observed when the true model contains 1 or 2 interactions. Tables 3.2-3.3 summarize the results obtained for the two models selected during each run.

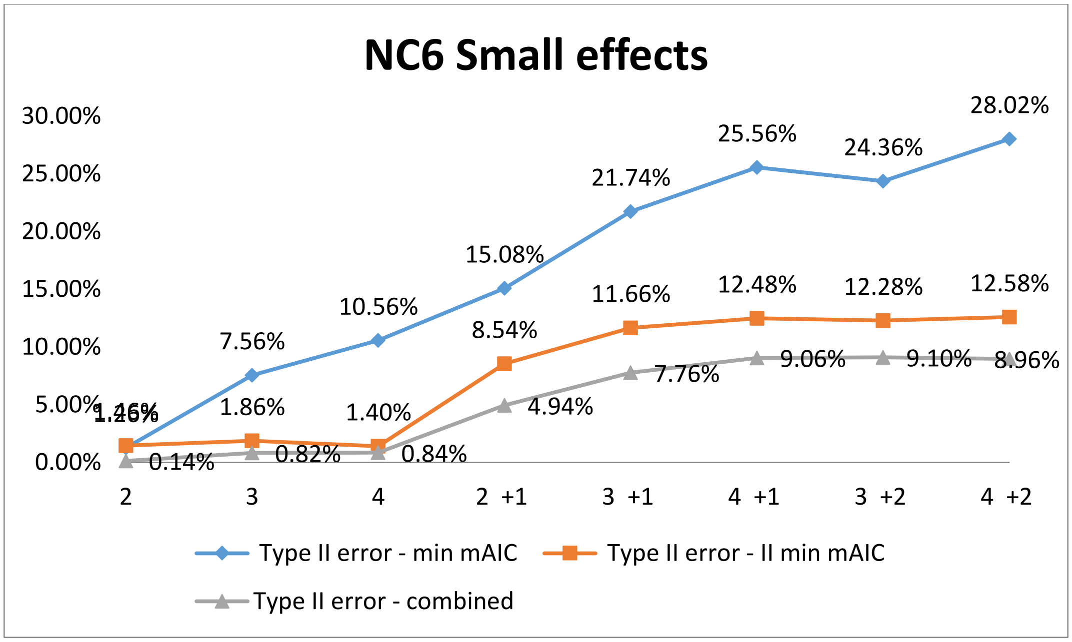

Figure 3.4 NC-6 small effects – δ < 10

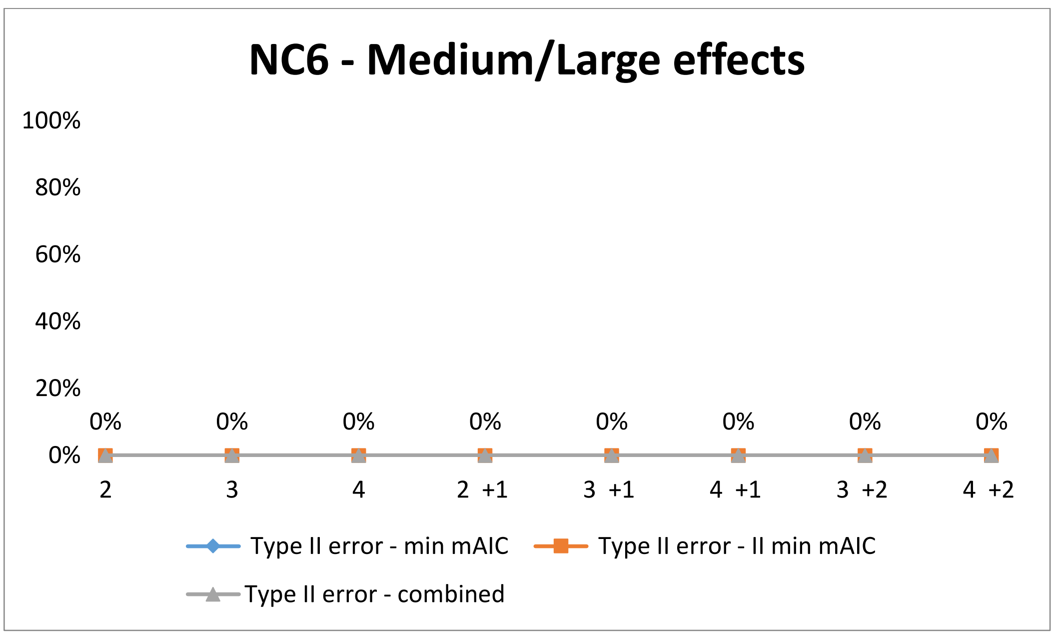

Figure 3.5 NC-6 medium/large effects - δ < 10

Table 3.2 NC6 Results summary - models selected based on minimum mAIC

Coefficient/ Noise ratio	Active factors in true model	All AF's identified+ No inactive	Type I error percent	Type II error percent
3 SD	2	96.96%	3.04%	0.00%
	3	99.32%	0.68%	0.00%
	4	99.68%	0.32%	0.00%
	2 +1	66.18%	33.82%	0.00%
	3 +1	68.04%	31.96%	0.00%
	4 +1	68.72%	31.28%	0.00%
	3 +2	67.82%	32.18%	0.00%
	4 +2	70.16%	29.84%	0.00%
2 SD	2	97.00%	3.00%	0.00%
	3	99.28%	0.72%	0.00%
	4	99.66%	0.34%	0.00%
	2 +1	54.92%	45.08%	0.02%
	3 +1	52.64%	47.36%	0.00%
	4 +1	51.52%	48.48%	0.00%
	3 +2	52.66%	47.34%	0.00%
	4 +2	55.48%	44.52%	0.00%
1 SD	2	95.06%	3.72%	1.26%
	3	90.66%	1.92%	7.56%
	4	88.50%	0.98%	10.56%
	2 +1	30.38%	68.40%	15.08%
	3 +1	25.94%	70.74%	21.74%
	4 +1	26.12%	69.16%	25.56%
	3 +2	25.18%	71.08%	24.36%
	4 +2	30.08%	63.52%	28.02%

Table 3.3 NC6 Results summary - models selected based on second minimum mAIC

Coefficient/ Noise ratio	Active factors in true model	All AF's identified+ No inactive	Type I error percent	Type II error percent
3 SD	2	65.10%	34.90%	0.00%
	3	73.46%	26.54%	0.00%
	4	81.56%	18.44%	0.00%
	2 +1	22.16%	77.84%	0.00%
	3 +1	19.82%	80.18%	0.00%
	4 +1	20.64%	79.36%	0.00%
	3 +2	19.24%	80.76%	0.00%
	4 +2	23.66%	76.34%	0.00%
2 SD	2	43.22%	56.78%	0.00%
	3	48.80%	51.20%	0.00%
	4	60.94%	39.06%	0.00%
	2 +1	7.68%	92.32%	0.00%
	3 +1	4.90%	95.10%	0.00%
	4 +1	5.14%	94.86%	0.00%
	3 +2	4.66%	95.34%	0.00%
	4 +2	7.38%	92.62%	0.00%
1 SD	2	13.32%	85.32%	1.46%
	3	25.58%	73.00%	1.86%
	4	42.90%	56.06%	1.40%
	2 +1	4.52%	95.02%	8.54%
	3 +1	2.72%	96.78%	11.66%
	4 +1	3.26%	96.14%	12.48%
	3 +2	2.94%	96.68%	12.28%
	4 +2	5.50%	93.98%	12.58%

3.5 Seven factor designs

As in the case of six factor designs, when only medium or large main effects are present, the combined error rate is zero. But when the true models have interaction, the error rate increases with the number of factors. Figures 3.6-3.8 show the graphical summary of the type II error rate when the range of delta was between 1- 100.

When the value of delta was restricted to be within 0-10, the error rate dropped considerably. For the medium and large coefficients, the combined type II error was close to zero for true models with a total of 5 active effects including interactions. However, the design breaks down when the true model has 4 main effects + 2 interactions. The error rates for smaller coefficients are much higher. It is also interesting to note that, the model selected using the second minimum value of mAIC has lower error rate than the models selected using minimum mAIC when the size of the coefficient is medium or large. Figures 3.9-3.11 present the results for cases where the delta value is less than 10.

Tables 3.4-3.5 present the individual results summary for the two models selected for each iteration. From the results, it is evident that the design breaks down when the number of active factors reach 6. Also the type I error is higher for the models selected based on the second minimum value of mAIC.

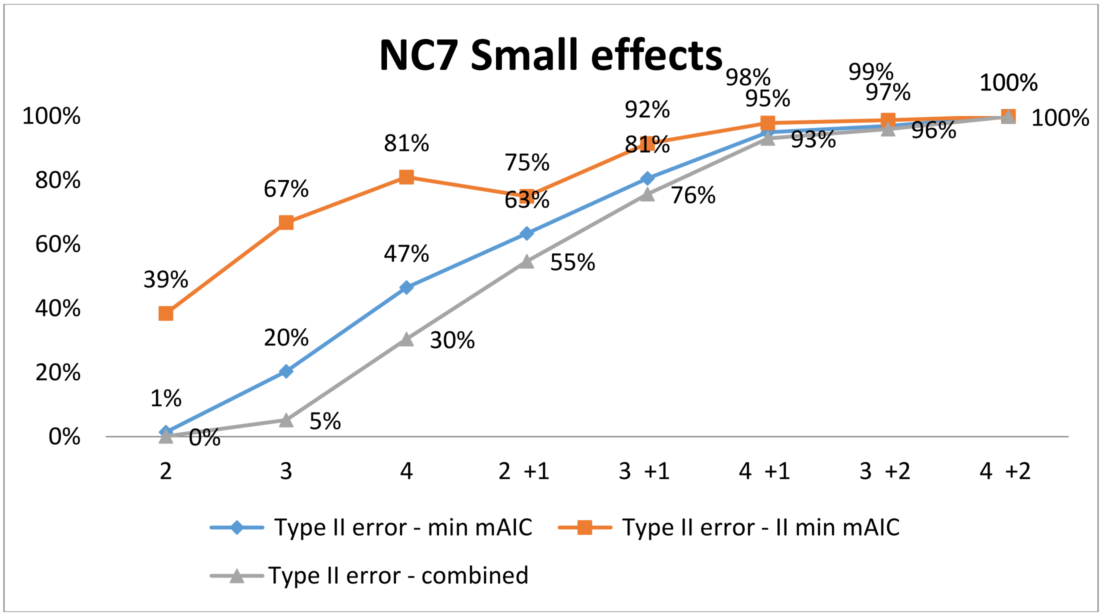

Figure 3.6 NC7 factor design with small coefficients

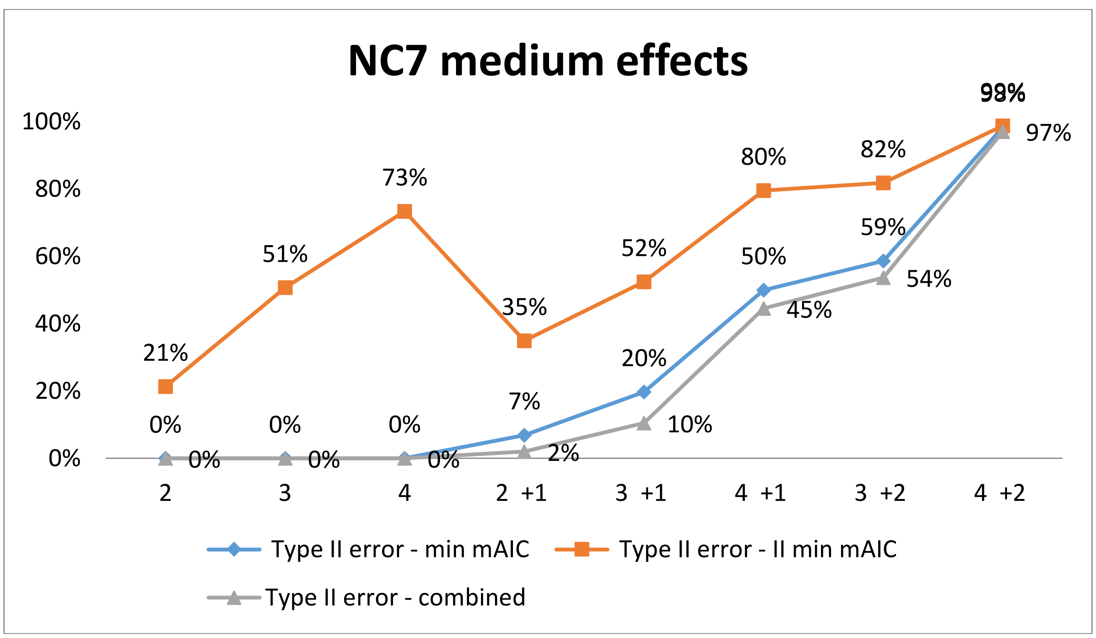

Figure 3.7 NC7 factor design with medium coefficients

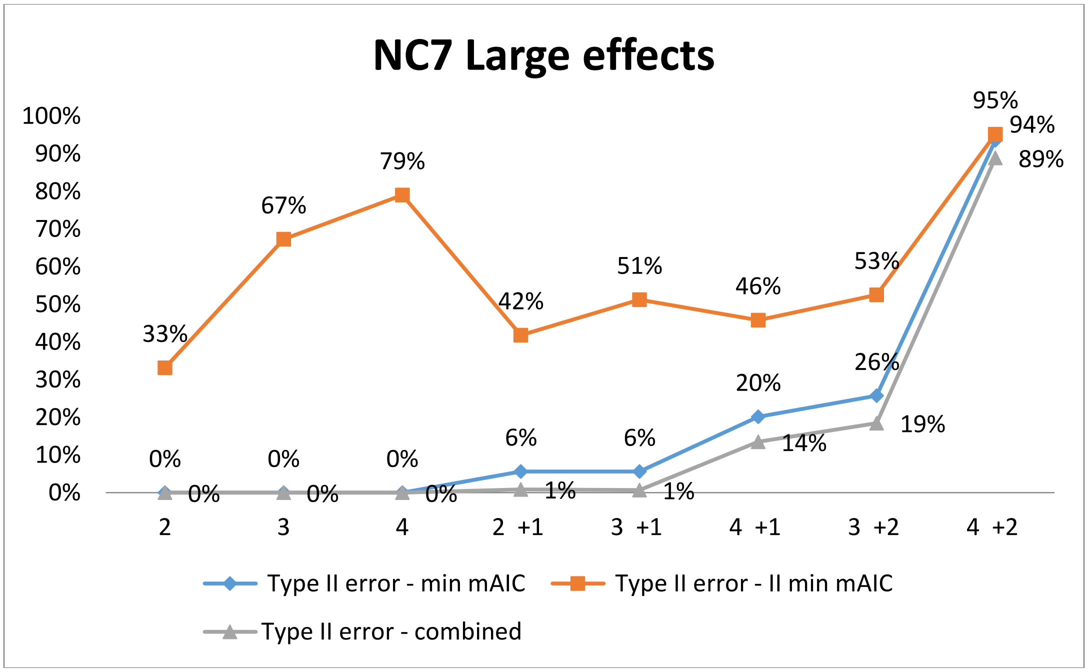

Figure 3.8 NC7 factor design with large coefficients

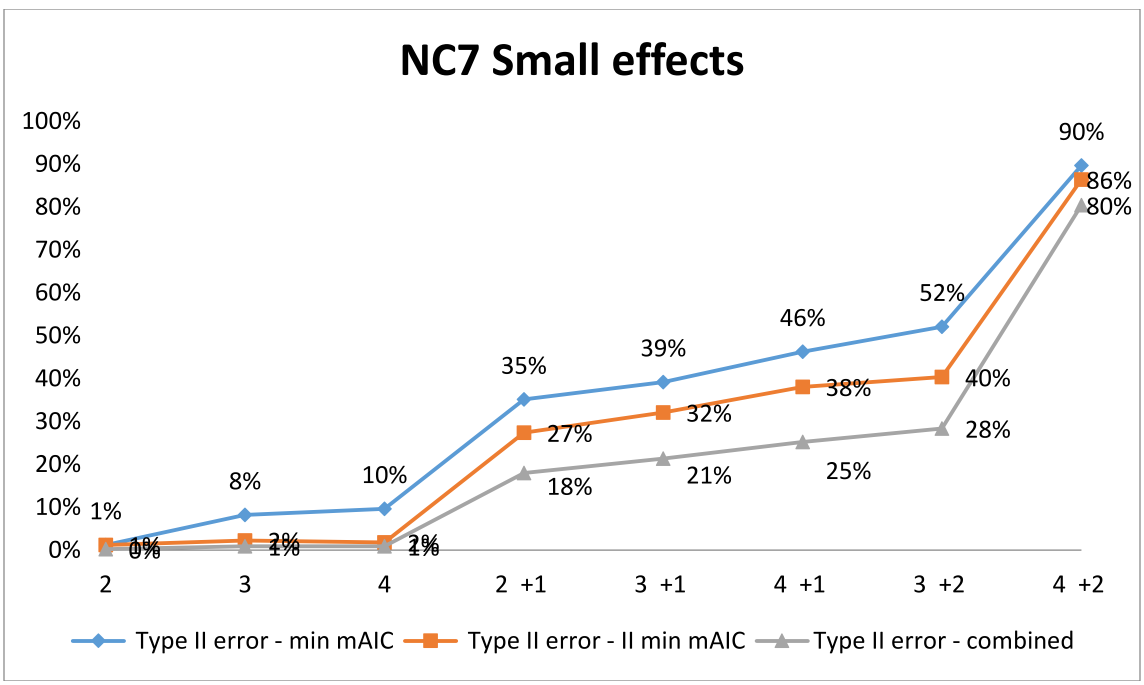

Figure 3.9 NC-7 small effects - $\delta < 10$

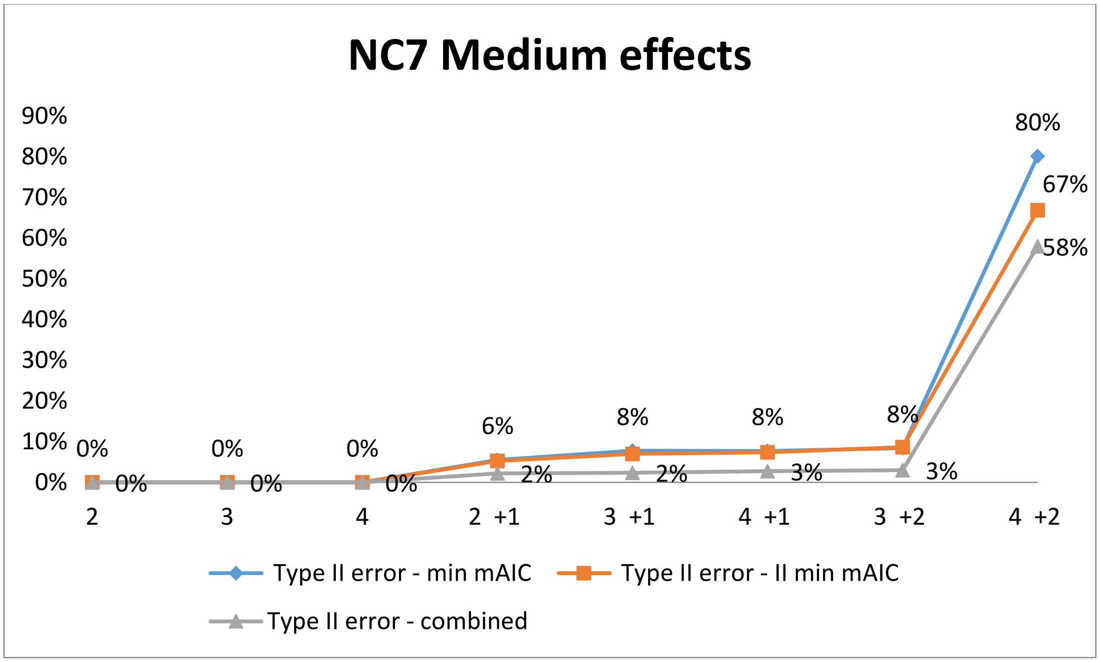

Figure 3.10 NC-7 medium effects - δ < 10

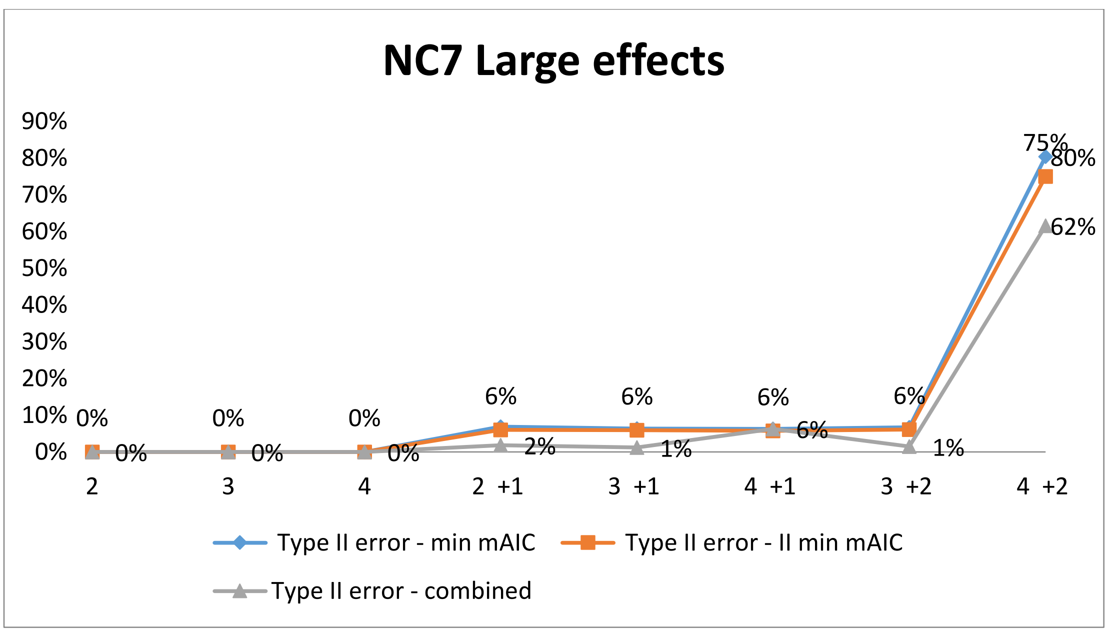

Figure 3.11 NC-7 large effects - δ < 10

Table 3.4 NC7 Results summary - models selected based on minimum mAIC

Coefficient /Noise ratio	Active factors in true model	All AF's identified+ No inactive	Type I error percent	Type II error percent
3 SD	2	96.80%	3.20%	0.00%
	3	99.02%	0.98%	0.00%
	4	99.74%	0.26%	0.00%
	2 +1	65.76%	34.24%	6.86%
	3 +1	67.82%	32.18%	6.38%
	4 +1	60.56%	39.44%	6.28%
	3 +2	53.72%	46.28%	6.74%
	4 +2	9.80%	90.20%	80.38%
2 SD	2	95.84%	4.16%	0.00%
	3	98.92%	1.08%	0.00%
	4	99.52%	0.48%	0.00%
	2 +1	54.04%	45.96%	5.54%
	3 +1	50.74%	49.26%	7.76%
	4 +1	43.20%	56.80%	7.74%
	3 +2	34.16%	65.84%	8.40%
	4 +2	6.52%	93.48%	80.14%
1 SD	2	94.02%	4.86%	1.16%
	3	89.94%	2.00%	8.16%
	4	89.56%	0.94%	9.60%
	2 +1	27.94%	70.40%	35.08%
	3 +1	25.18%	71.64%	39.16%
	4 +1	19.06%	77.44%	46.24%
	3 +2	12.92%	84.84%	52.08%
	4 +2	2.50%	97.42%	89.64%

Table 3.5 NC7 Results summary - models selected based on second minimum mAIC

Coefficient /Noise ratio	Active factors in true model	All AF's identified+ No inactive	Type I error percent	Type II error percent
3 SD	2	60.42%	39.58%	0.00%
	3	66.90%	33.10%	0.00%
	4	73.98%	26.02%	0.00%
	2 +1	24.44%	75.56%	5.98%
	3 +1	21.94%	78.06%	5.90%
	4 +1	6.60%	93.40%	5.70%
	3 +2	2.44%	97.56%	6.06%
	4 +2	8.78%	91.22%	74.98%
2 SD	2	39.12%	60.88%	0.00%
	3	39.36%	60.64%	0.00%
	4	49.36%	50.64%	0.00%
	2 +1	9.94%	90.06%	5.28%
	3 +1	6.18%	93.82%	7.00%
	4 +1	1.88%	98.12%	7.38%
	3 +2	1.48%	98.52%	8.66%
	4 +2	7.38%	92.62%	66.84%
1 SD	2	31.08%	67.84%	1.18%
	3	19.78%	78.56%	2.22%
	4	29.50%	69.18%	1.78%
	2 +1	5.36%	94.18%	27.36%
	3 +1	3.14%	96.16%	32.08%
	4 +1	3.34%	96.22%	38.06%
	3 +2	2.02%	97.68%	40.34%
	4 +2	1.44%	98.54%	86.40%

3.6 NC8 factor design

When the true model contains only the main effects, NC8 designs have zero error rate when the size of the coefficients are medium or large. When the effects are large, NC8 is even able to detect true models which contain 3 main effects and 1 hierarchical interaction with close to a zero error rate. Figures 3.12-3.14 present graphical summary of results when the delta is varied between 1 and 100.

For the cases where the delta is restricted between 1-10, as observed with the NC6 and NC7 designs, there is a considerable drop in the combined type II error rate. Additionally the error rates are close to zero for true models with up to 3 main effects and 2 hierarchical interactions when the coefficient sizes are medium. Also when the effect sizes are small, the type II error rate exceeds 30% when interactions are present in the true model. When the effects are large, true models with 3 or 4 main effects and 2 hierarchical interactions have lower error rates than the models with 1 hierarchical interaction. Individually, the models selected based on the second minimum mAIC have lower type II error rates than the models selected based on minimum mAIC. Figures 3.15-3.17 present graphical summary of results when the delta is varied between 1 and 10.

Tables 3.6-3.7 show the results summary individually for the two models selected after each iteration. From the results, it is evident that the models selected based on second minimum mAIC consistently performs better than the model selected by minimum mAIC when it comes to type II error rate. But the type I error rates are lower for the models selected through minimum mAIC value.

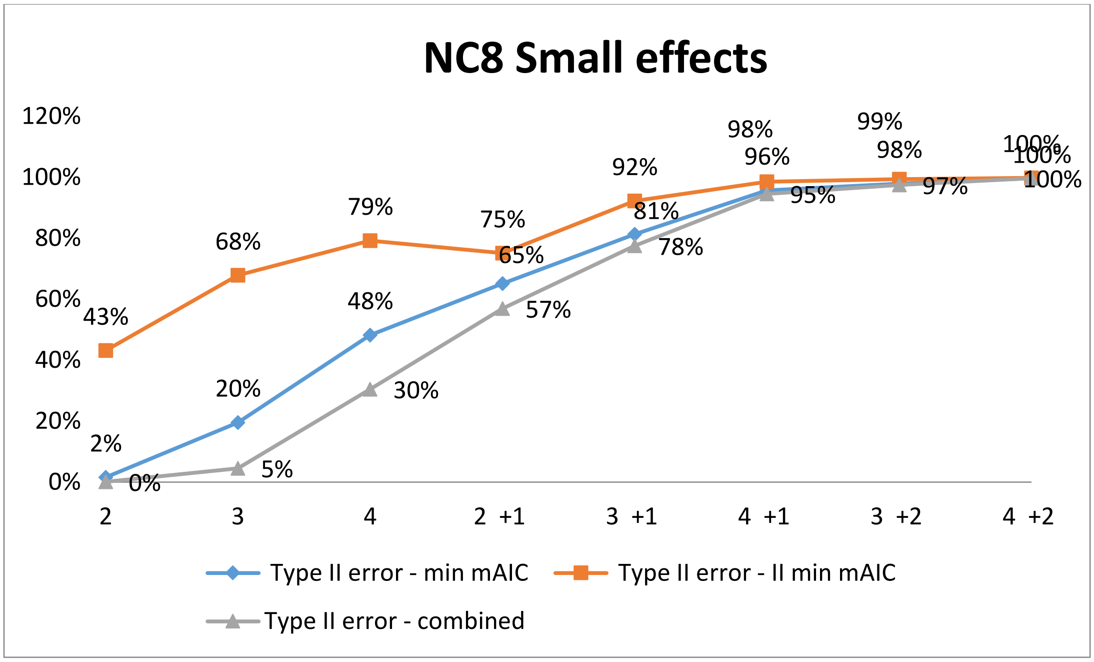

Figure 3.32 NC8 factor design with small coefficients

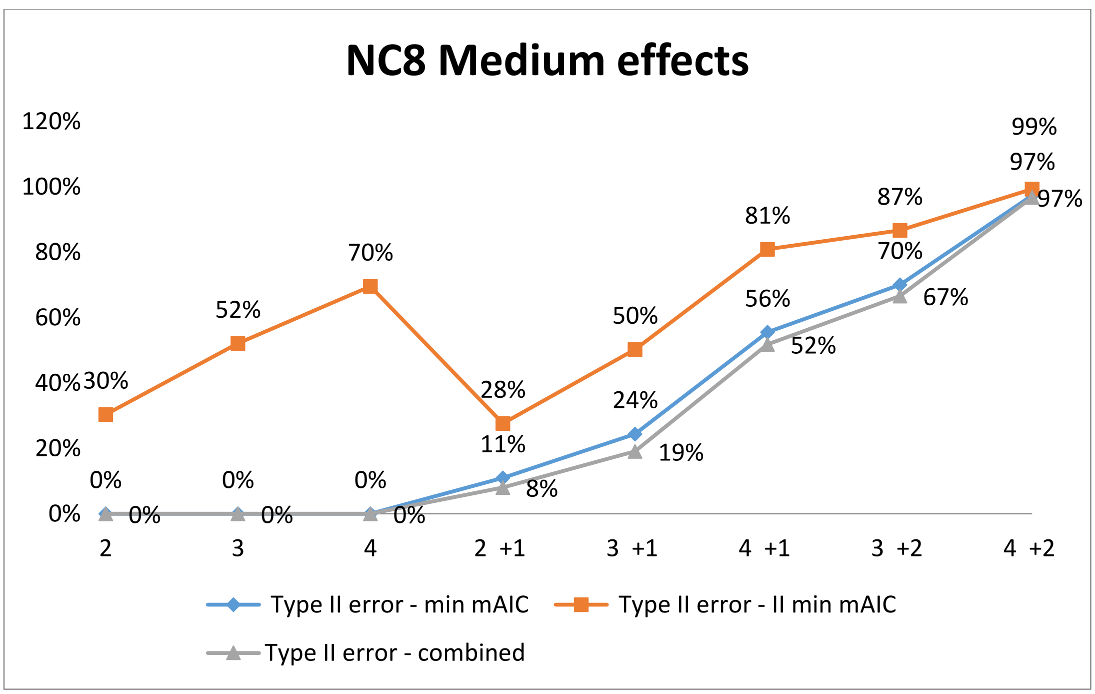

Figure 3.13 NC8 factor design with medium coefficients

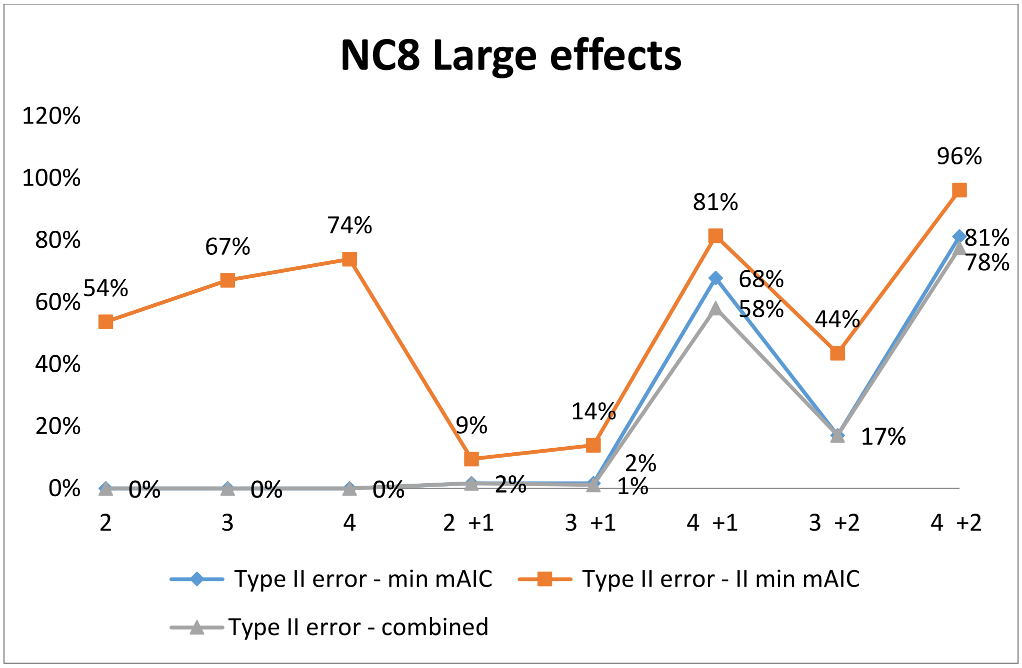

Figure 3.14 NC8 factor design with large coefficients

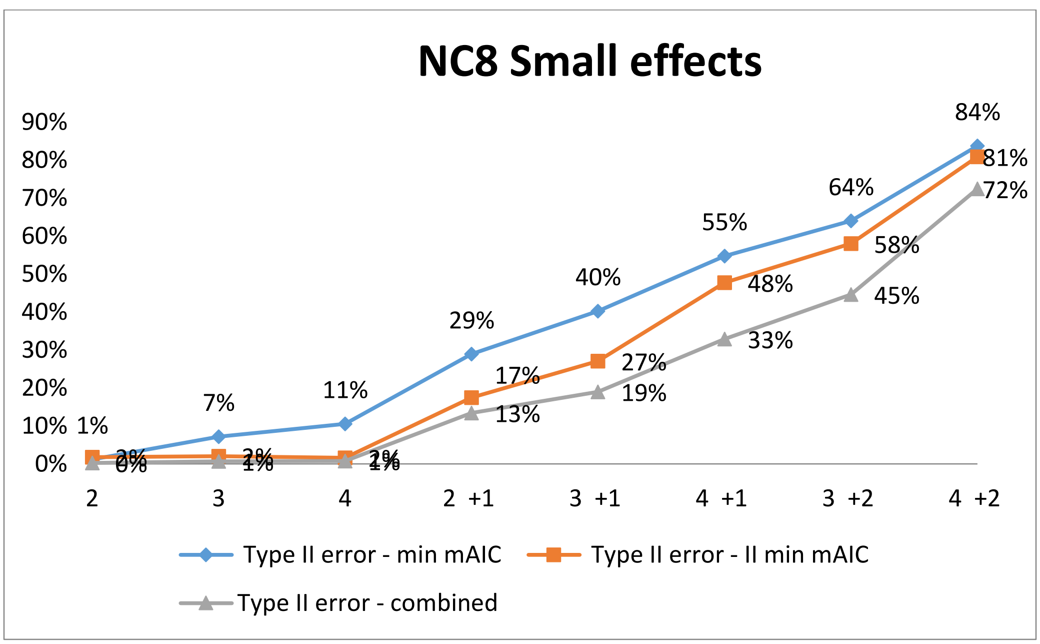

Figure 3.15 NC-8 small effects - δ < 10

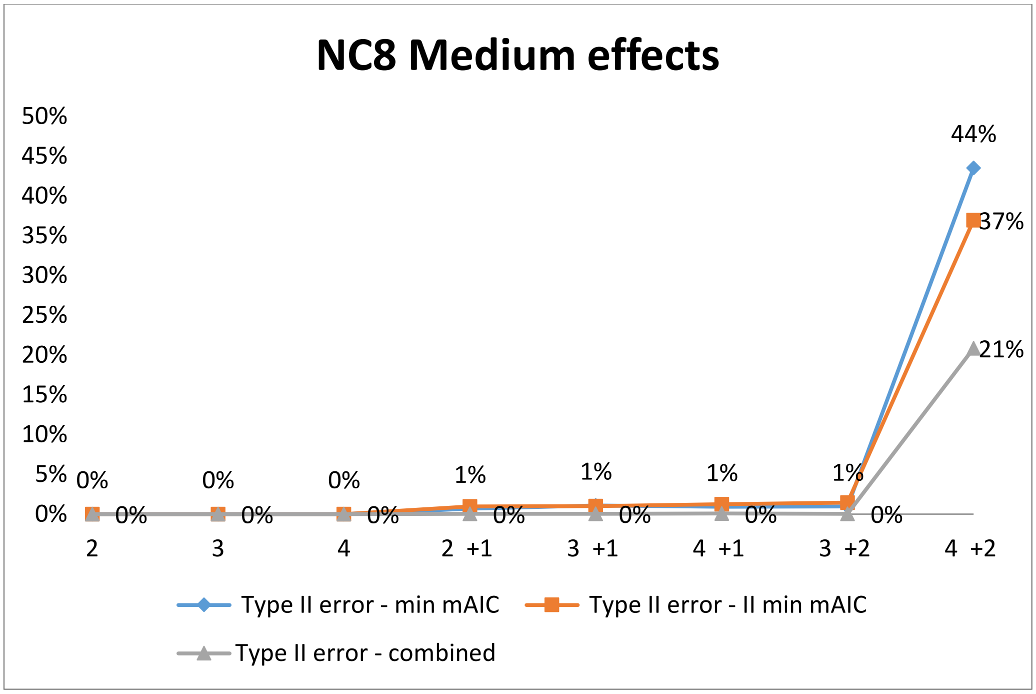

Figure 3.16 NC-8 medium effects - δ < 10

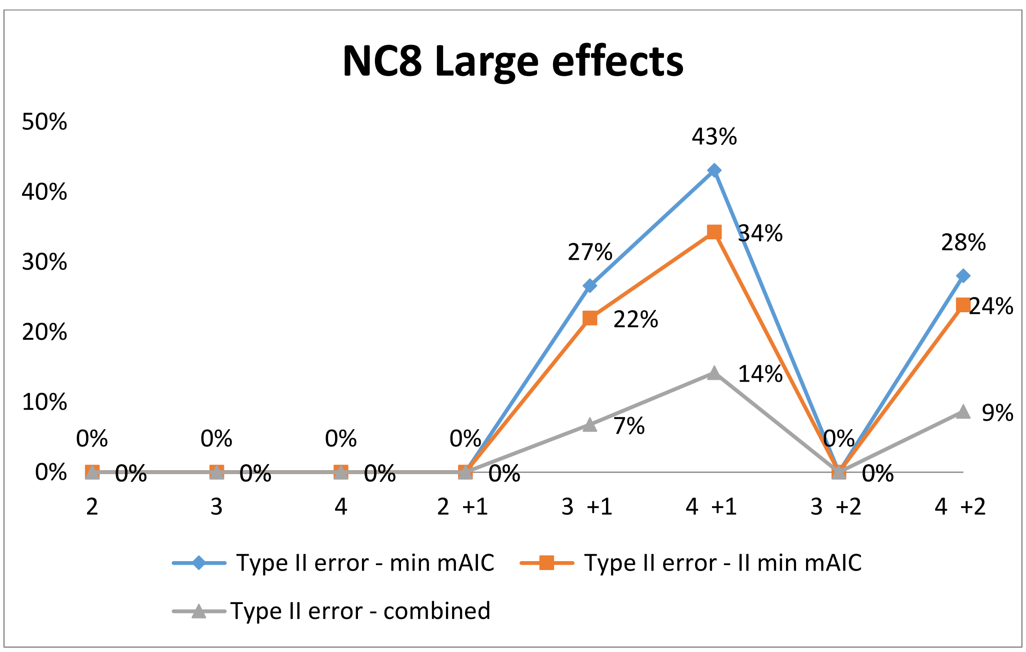

Figure 3.17 NC-8 large effects - δ < 10

Table 3.6 NC8 Results summary - models selected based on minimum mAIC

Coefficient/ Noise ratio	Active factors in true model	All AF's identified + No inactive	Type I error percent	Type II error percent
3 SD	2	96.00%	4.00%	0.00%
	3	99.02%	0.98%	0.00%
	4	99.52%	0.48%	0.00%
	2 +1	61.76%	38.24%	0.00%
	3 +1	33.88%	66.12%	26.62%
	4 +1	31.66%	68.34%	43.10%
	3 +2	65.10%	34.90%	0.00%
	4 +2	34.80%	65.20%	28.00%
2 SD	2	95.64%	4.36%	0.00%
	3	98.84%	1.16%	0.00%
	4	99.14%	0.86%	0.00%
	2 +1	41.84%	58.16%	0.68%
	3 +1	39.54%	60.46%	1.08%
	4 +1	33.48%	66.52%	0.90%
	3 +2	21.72%	78.28%	0.94%
	4 +2	7.74%	92.26%	43.52%
1 SD	2	93.16%	5.72%	1.18%
	3	89.82%	3.08%	7.20%
	4	88.02%	1.48%	10.54%
	2 +1	26.26%	72.48%	28.88%
	3 +1	22.52%	74.48%	40.26%
	4 +1	12.46%	86.08%	54.76%
	3 +2	7.54%	91.28%	63.98%
	4 +2	2.24%	97.70%	83.74%

Table 3.7 NC8 Results summary - models selected based on second minimum mAIC

Coefficient/ Noise ratio	Active factors in true model	All AF's identified + No inactive	Type I error percent	Type II error percent
3 SD	2	51.84%	48.16%	0.00%
	3	58.88%	41.12%	0.00%
	4	66.68%	33.32%	0.00%
	2 +1	10.40%	89.60%	0.00%
	3 +1	5.62%	94.38%	22.00%
	4 +1	7.22%	92.78%	34.28%
	3 +2	0.76%	99.24%	0.00%
	4 +2	4.94%	95.06%	23.88%
2 SD	2	19.54%	80.46%	0.00%
	3	31.30%	68.70%	0.00%
	4	40.72%	59.28%	0.00%
	2 +1	5.26%	94.74%	0.94%
	3 +1	0.78%	99.22%	1.00%
	4 +1	0.28%	99.72%	1.24%
	3 +2	0.18%	99.82%	1.44%
	4 +2	1.06%	98.94%	36.94%
1 SD	2	6.64%	91.88%	1.78%
	3	12.12%	86.38%	2.04%
	4	22.16%	76.80%	1.62%
	2 +1	4.02%	95.48%	17.46%
	3 +1	2.56%	96.96%	27.04%
	4 +1	0.84%	99.00%	47.74%
	3 +2	0.74%	99.16%	57.98%
	4 +2	0.36%	99.64%	80.88%

3.7 Alternate no-confounding designs for 6 factors

An alternative six factor no-confounding design and its performance based on analysis using a Dantzig selector are presented in this section. A comparative study for the two NC-6 designs is also presented.

We recommend the following alternate no-confounding design in six factors (alternate NC-6) in 16 runs, as shown in Table 3.8. The design is orthogonal and has no confounding of two factor interactions like the NC-6 design. The correlation matrix for the alternate NC-6 design is presented in Figure 3.18.

Table 3.8 Recommended Alternate NC-6 design in 16 runs

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

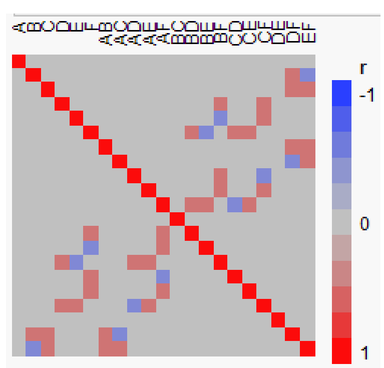

Figure 3.18 Correlation matrix for Alternate NC-6 design

3.8 Analysis of alternate NC-6 design using Dantzig selector

Figures 3.19-3.21 present the graphical summaries of type II error for the three different coefficient sizes tested when the delta value is varied between 1 and 100. As observed in the case of NC-6 designs, when the true model contains only main effects, the type II error rate is zero for medium and large sized coefficients. When the true model includes 1 or 2 interactions, the maximum error rate observed was 64% when the coefficient size was medium. True models with small coefficient sizes had higher error rates.

Figures 3.22-3.25 present the summaries of error rate when the delta value is varied between 1 and 10. When the delta value is small, the alternate NC-6 designs perform well for all ranges of coefficients. The error rate was zero for both medium and large coefficients. For the small coefficients, the maximum error rate observed was 9%.

Tables 3.9-3.10 present the results summary individually for the two models chosen in each simulation run. When it comes to type I error, the models chosen based on minimum mAIC perform better. But when type II error is considered, models based on second minimum mAIC have smaller error rates.

A graphical comparison of success rate between NC-6 and the alternate NC-6 design, where the method identifies the true model without any error is shown in Figure 3.26. When the coefficients are medium or large sized, the type II error rates are zero for both the designs. But when it comes to smaller coefficients, the alternate NC-6 design has a higher success rate. It is also evident that the alternate NC-6 design performs much better even when the true model contains as many as 6 active factors including 2 hierarchical interactions.

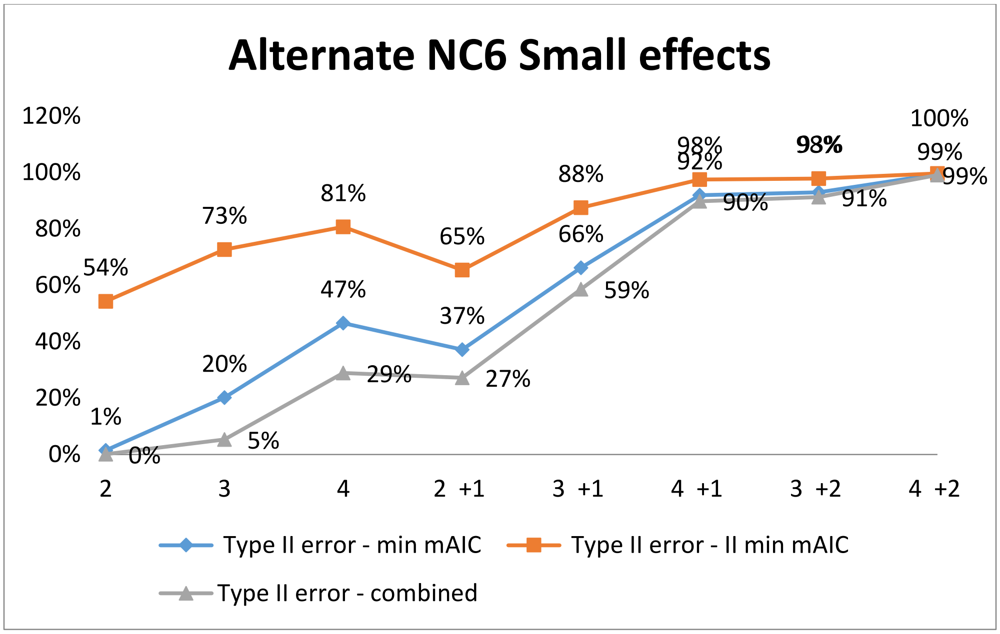

Figure 3.19 Alternate NC-6 factor design with small coefficients

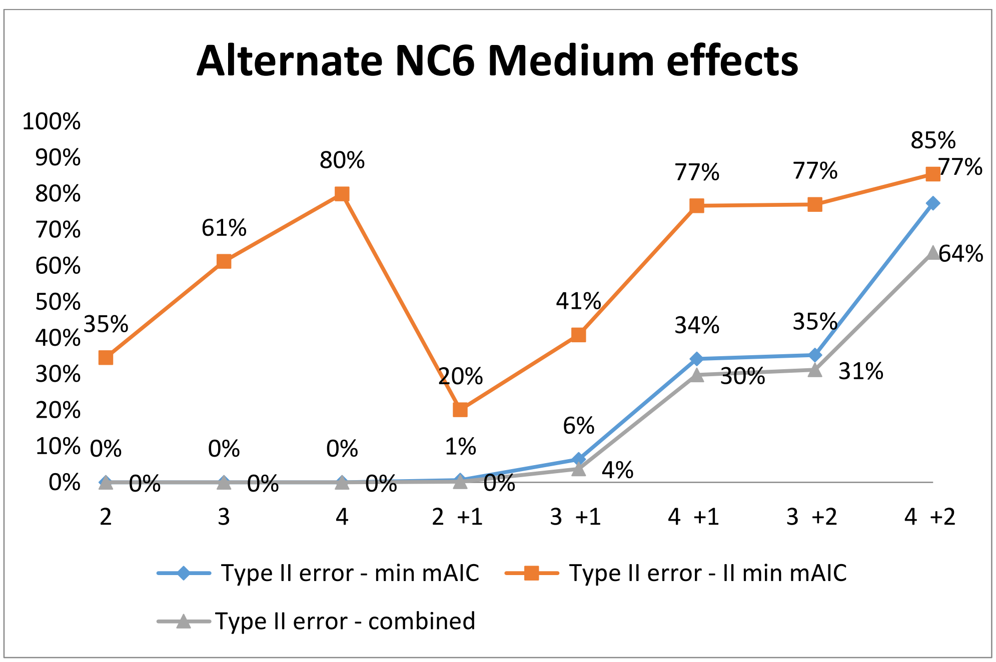

Figure 3.20 Alternate NC-6 factor design with medium coefficients

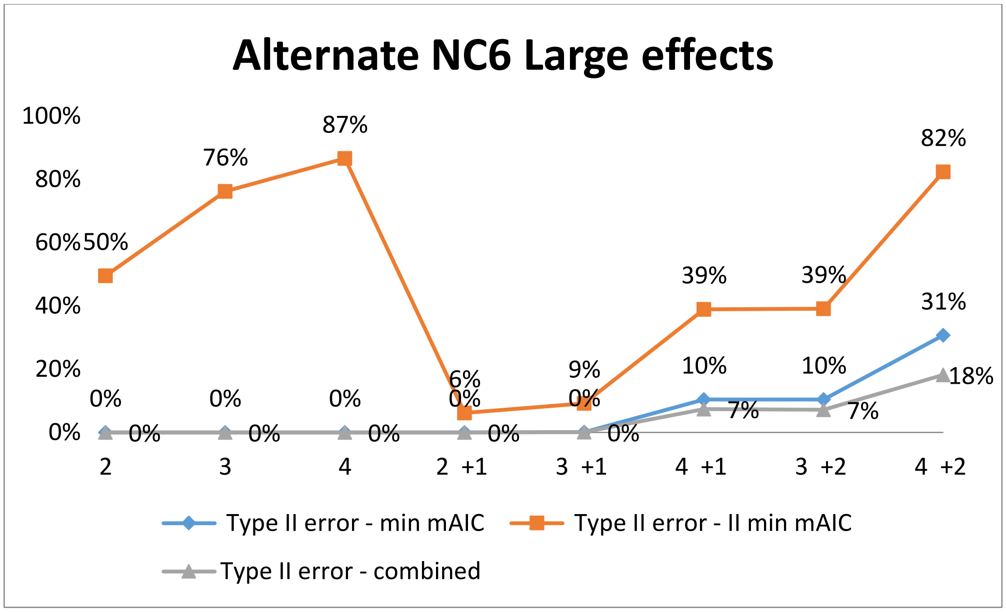

Figure 3.21 Alternate NC-6 factor design with large coefficients

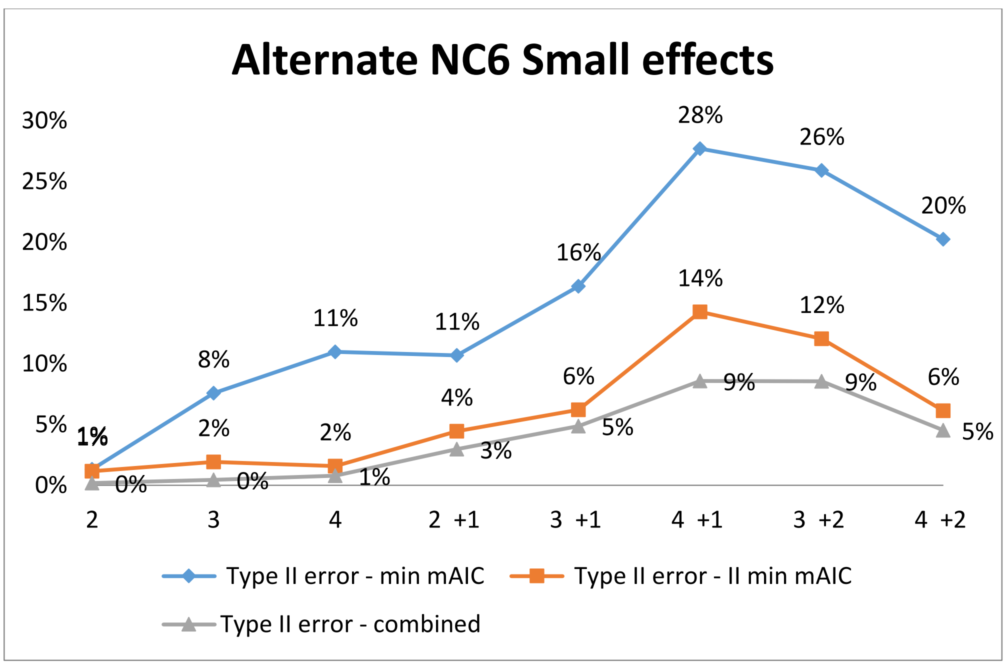

Figure 3.22 Alternate NC-6 small effects - δ < 10

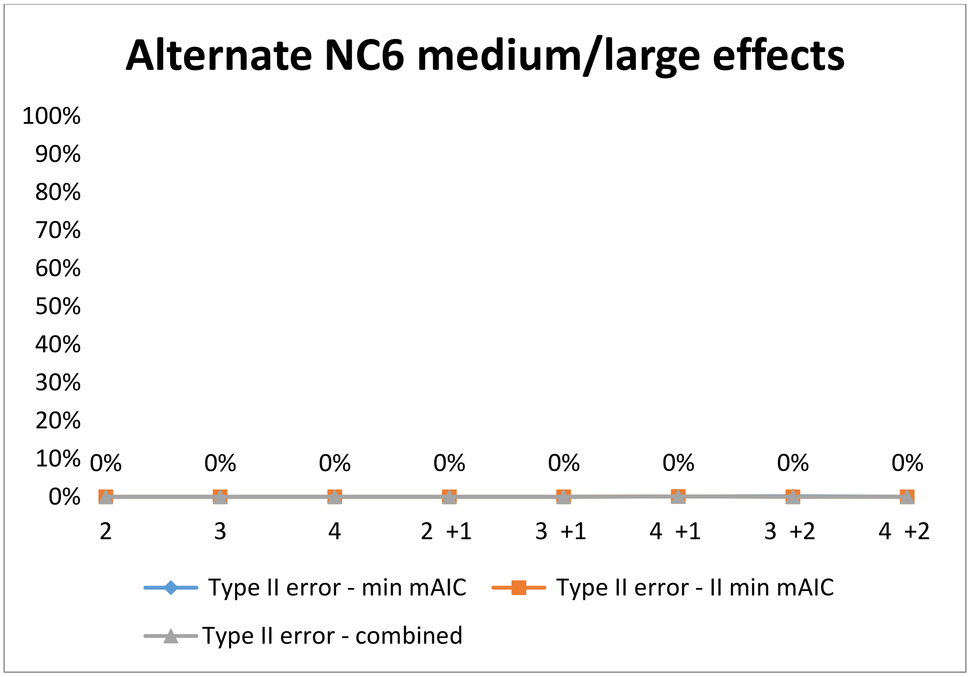

Figure 3.23 NC-6 medium/large effects - δ < 10

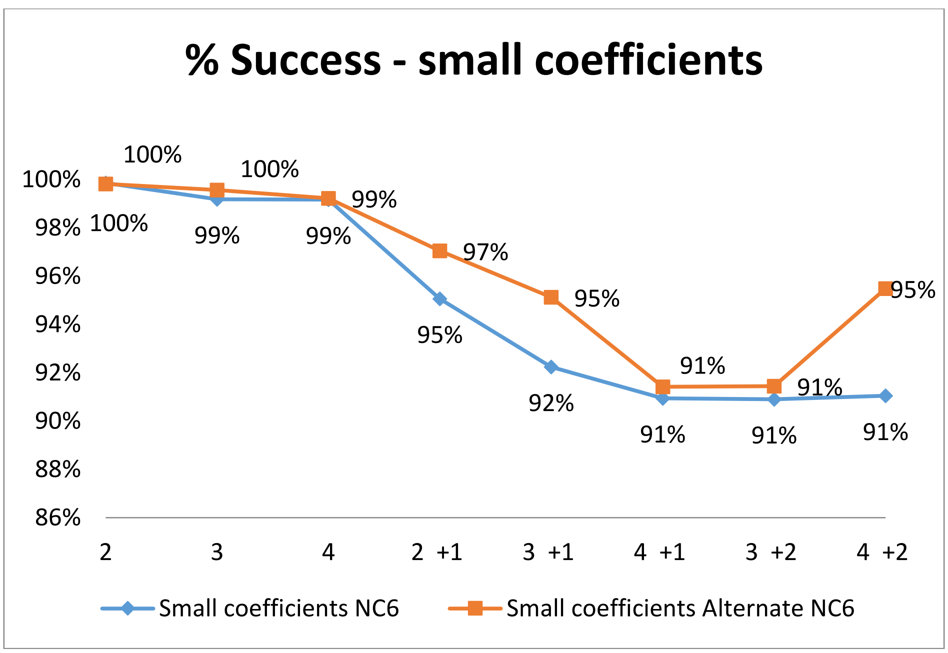

Figure 3.24 Comparison - NC6 and Alternate NC6 success rate

Table 3.9 Alternate NC-6 results summary - models selected based on minimum mAIC

Coefficient /Noise ratio	Active factors in true model	All AF's identified+ No inactive	Type I error percent	Type II error percent
3 SD	2	97.06%	2.94%	0.00%
	3	99.48%	0.52%	0.00%
	4	99.82%	0.18%	0.00%
	2 +1	81.46%	18.54%	0.00%
	3 +1	81.88%	18.12%	0.00%
	4 +1	73.84%	26.16%	0.00%
	3 +2	73.26%	26.74%	0.00%
	4 +2	85.96%	14.04%	0.00%
2 SD	2	97.06%	2.94%	0.00%
	3	99.16%	0.84%	0.00%
	4	99.72%	0.28%	0.00%
	2 +1	72.66%	27.34%	0.00%
	3 +1	70.82%	29.18%	0.04%
	4 +1	56.12%	43.88%	0.14%
	3 +2	53.18%	46.82%	0.16%
	4 +2	76.84%	23.16%	0.00%
1 SD	2	94.82%	4.00%	1.30%
	3	91.02%	1.48%	7.58%
	4	88.42%	0.64%	10.98%
	2 +1	59.26%	38.58%	10.68%
	3 +1	54.32%	40.20%	16.38%
	4 +1	38.24%	55.52%	27.70%
	3 +2	37.94%	56.46%	25.90%
	4 +2	57.30%	32.86%	20.24%

Table 3.10 Alternate NC-6 results summary - models selected based on second minimum mAIC

Coefficient /Noise ratio	Active factors in true model	All AF's identified+ No inactive	Type I error percent	Type II error percent
3 SD	2	62.42%	37.58%	0.00%
	3	72.64%	27.36%	0.00%
	4	81.48%	18.52%	0.00%
	2 +1	11.86%	88.14%	0.00%
	3 +1	6.82%	93.18%	0.00%
	4 +1	3.68%	96.32%	0.00%
	3 +2	3.82%	96.18%	0.00%
	4 +2	7.82%	92.18%	0.00%
2 SD	2	43.24%	56.76%	0.00%
	3	48.88%	51.12%	0.00%
	4	63.68%	36.32%	0.00%
	2 +1	7.28%	92.72%	0.02%
	3 +1	3.14%	96.86%	0.04%
	4 +1	0.42%	99.58%	0.08%
	3 +2	0.38%	99.62%	0.06%
	4 +2	4.80%	95.20%	0.04%
1 SD	2	32.72%	66.18%	1.16%
	3	23.52%	74.92%	1.92%
	4	43.00%	55.90%	1.58%
	2 +1	3.40%	95.94%	4.44%
	3 +1	4.44%	94.84%	6.20%
	4 +1	4.62%	94.72%	14.26%
	3 +2	4.04%	95.66%	12.06%
	4 +2	8.42%	90.90%	6.14%

3.9 Conclusion

This chapter studies the effectiveness of the Dantzig selector as a variable selection method for analyzing NC designs. Stepwise regression is the most commonly used method for variable selection. However, an extensive study on their performance by Shinde (2012) confirmed that stepwise regression does not work well once the total number of active factors exceeds four. Based on the simulations conducted in this study, it is evident that Dantzig selector performs better even for true models that contain up to six active factors including two interactions. Additionally, the results also indicate that when the delta values chosen are within a reasonable range, Dantzig selector performs exceptionally well for medium and large sized coefficients.

An alternate NC-6 design was also introduced in this chapter. A comparative study suggests that the alternate NC-6 performs better even when the true models contain small coefficients.

While using the modified AIC criterion for model selection, selecting two models from each experiment based on the minimum and second minimum value worked well by reducing the error rate considerably.

We believe that NC designs are good alternatives to the fractional factorial designs when both main effects and interactions are to be identified, especially in cases where running follow-up experiments is not an option.

Chapter 4

CONCLUSIONS AND FUTURE WORK

4.1 Conclusion and future work

With increasing computational efficiency, supersaturated designs are continuing to receive more attention for their ability to handle a large number of variables in fewer runs. When developments and discovery involve complex systems with many factors, running regular fractional factorial designs could become expensive and challenging, as they would require follow up experiments to identify main effects and interactions. From the simulation study, it is evident that NC designs provide a good alternative to regular designs and can allow for the estimation of main effects and some two-factor interactions without the need of follow up experiments.

The simulation study also confirmed that Dantzig selector performs well in identifying the active effects when true models contain few interactions. Shinde (2012) observed that stepwise regression breaks down once the number of active factors exceeds four. However, from this study, it can be observed that Dantzig selector was able to identify even 6 active factors with good accuracy. Additionally, the results also indicate that when the delta values chosen are within a reasonable range, Dantzig selector performs exceptionally well for medium and large sized coefficients.

The proposed alternate NC-6 designs work better even when the coefficient to noise ratio is one. While using the mAIC criterion for model selection, choosing two models based on minimum and second minimum value of mAIC works better than choosing a single model based on the minimum value. The fact that Dantzig selector is a linear program makes it fast and easy to use, since many software and packages contain algorithms for linear programming.

We believe that NC designs are a good alternate to FF designs. When used with effective variable selection methods like the Dantzig selector, they can eliminate the need for running follow up experiments, thereby saving cost, time and resources. With the initial 16 runs, NC designs can estimate both main effects and interactions when effect sparsity holds true.

The threshold value (γ) is the minimum value for an effect to be considered active. In this study, a common threshold value was applied to each of the estimates calculated, irrespective of whether it is a main effect or an interaction. The present work could also be extended by defining different threshold for different types of effects. NC designs for 9 through 14 factors in 16 runs were introduced by Montgomery (2012). Running a simulation study to evaluate the effectiveness of Dantzig selector for these designs could be a possible extension to the present work.

As part of this study, I examined at true models with a maximum of 6 active factors. It would also be interesting to explore how the Dantzig selector performs for more complex models with higher number of factors. Also, exploring the possibilities of combining the Dantzig selector with other existing variable selection methods for improving variable selection could be another extension to the present work.

REFERENCES

Candes, E., & Tao, T. (2007). The Dantzig selector: Statistical estimation when p is much larger than n. The Annals of Statistics, 35(6), 2313-2351.

Chipman, H., Hamada, M., & Wu, C. F. J. (1997). A Bayesian variable-selection approach for analyzing designed experiments with complex aliasing. Technometrics, 39(4), 372-381.

Draguljic, D., D. C. Woods, A. M. Dean, S. M. Lewis, and A.-J. E. Vine (2014). Screening strategies in the presence of interactions. Technometrics 56(1), 1–16.

Hall, M. Jr. (1961). Hadamard matrix of order 16, Jet Propulsion Laboratory Research Summary, Vol. 1, pp.21–26.

Hamada, M., & Wu, C. J. (1992). Analysis of designed experiments with complex aliasing. Journal of Quality Technology, 24, 130-137.

Jones, B., & Montgomery, D. C. (2010). Alternatives to resolution IV screening designs in 16 runs. International Journal of Experimental Design and Process Optimisation, 1(4), 285-295

Lin, D. K. J. (1993). A new class of supersaturated designs. Technometrics, 35, 28–31.

Marley, C. J. and Woods, D. C. (2010), A comparison of design and model selection methods for supersaturated experiments," Computational Statistics and Data Analysis, 54, 3158-3167

Mee, R. (2013), "Tips for Analyzing Nonregular Fractional Factorial Experiments," Journal of Quality Technology, 45, 330-349.

Montgomery, D. C. (2012), Design and Analysis of Experiments, New York: Wiley, 8th ed.

Phoa, F. K. H., Pan, Y.-H., and Xu, H. (2009a), Analysis of supersaturated designs via the Dantzig selector," Journal of Statistical Planning and Inference, 139, 2362-2372.

Plackett, R. L., & Burman, J. P. (1946). The design of optimum multifactorial experiments. Biometrika, 305-325.

Shinde, Shilpa Madhavan (2012) Projection Properties and Analysis Methods for Six to Fourteen Factor No Confounding Designs in 16 Runs, PhD Dissertation – Arizona State University

Tyssedal, J., & Samset, O. (1997). Analysis of the 12 run Plackett-Burman design. Preprint, Statistics, (8).

Wu, C. F. J. (1993). Construction of supersaturated designs through partially aliased interactions. Biometrika, 80, 661–669.

Wu, C. F. J. and Hamada, M. (2000). Experiments: Planning, Analysis and Parameter Design Optimization. New York: Wiley.