Some risks in the construction and analysis of supersaturated designs

B. Abraham **H. Chipman**
Department of Statistics and Actuarial Science
and Institute for Improvement in Quality and Productivity
University of Waterloo

and

K. Vijayan*
University of Western Australia
Nedlands, Western Australia

July 1998

ABSTRACT

A designed experiment in which the number of factors are at least as large as the number of runs is referred to as a supersaturated (SS) design. Recently these designs have received increased attention. Construction of such designs and analysis of data from these designs have been discussed by several authors. Our objective in this paper is to examine these designs and methods for their analysis. An important finding for practitioners is that the correlation structure inherent in SS designs can obscure real effects or promote non-real effects. While superior to stepwise regression, all subsets regression can still mislead. Hence one should be cautious with the use of SS designs.

Key Words: Hadamard matrices, Misspecification, Plackett and Burman Designs, Rubber data, Stepwise Regression, All Subsets Regression.

*Work was done while on leave at the University of Waterloo.

1 Introduction

In industrial experimentation it is natural to look at many factors simultaneously. Cost considerations will sometimes make it impractical to design experiments so that effects of all these factors are simultaneously estimated. Assuming that only a few factors are active (having large effects), there have been attempts to identify them with high probability. The supersaturated (SS) designs used for this purpose have fewer observations than the number of factors considered (see, for example, Booth and Cox (1962), Lin (1993)). The objective usually is to identify all or most of the active factors. The intention of this paper is to investigate these designs and their ability to identify active factors.

It is assumed that there is no interaction between the factors and that the effect is linear in the level of each factor. Hence the basic model for observations y_i is given by

$$y_i = \beta_0 + \sum_{j=1}^{m} \beta_j X_{ij} + \epsilon_i \ , \ i = 1, 2, \cdots, n \qquad n < m \tag{1}$$

where m is the number of factors and n is the number of observations (runs). Here X_{ij} is the level of factor X_j in run i and is restricted to two values ($+1$ or -1). The parameter β_j $(j = 1, \ldots, m)$ represents the effect of factor j and it is assumed that only a few of these parameters are different from 0. As the number of parameters is more than the number of observations, usual least squares methods cannot be used with (1) and all m factors. Some subset of factors must be selected. Lin (1993) suggests forward stepwise regression. All subsets regression, which evaluates all possible models, is a better alternative. In many SS experiments, different subsets of factors may explain the observed data well. All subsets regression is able to identify multiple models, while stepwise usually cannot.

There are several methods of constructing SS designs (choosing the "X" matrix). These include Wu (1993), Lin (1993), Nguyen (1996), Tang and Wu (1997), and Yamada and Lin (1997). Lin (1993) suggested using half fractions of Hadamard matrices. We will refer to this as the half replicate (HR) method. In Section 2 we will consider this procedure and discuss a set of data (the Rubber Data) which was previously considered by Williams (1968). Section 3 discusses the issue of misspecification and gives a simulation study to investigate the merits of the HR supersaturated design. In both sections 2 and 3, stepwise and all subsets methods are considered. Section 4 gives some concluding remarks.

2 Rubber Data

In order to illustrate the performance of the HR supersaturated design, Lin (1993) considered a set of data originally reported in Williams (1968) who, in the context of a rubber making process, used a Plackett-Burman (PB) design with 24 factors and 28 runs. The design and the data are given in Table 7 in the Appendix. As was pointed out by Box and Draper (1987) there was a typographical error in one of the entries of the original data (8th element from top in column 20) and this is corrected in our table. Also columns 13 and 16 are identical and hence we eliminate column 16; however we keep the labeling of the other factors the same.

The HR method can be summarised as follows: In a standard Hadamard matrix whose first column entries are all +1, take any other column as a branching column. Consider the matrix formed by the rows which have the same sign in the branching columns. Assign factors to the columns other than the first and the branching columns. The resulting matrix is recommended as an SS design.

The HR construction procedure implies that the design obtained is not unique. In the current example since 23 factors are being investigated we consider a Hadamard matrix of order 28 with +1 in the first column. The 23 factors given in Table 7 can be assigned to any 23 of the other 27 columns of the Hadamard matrix. The remaining 4 columns of the Hadamard matrix can be considered as branching columns and can be used to produce 8 designs (see Vijayan (1976)). The rows corresponding to these 8 designs are given in Table 8 in the Appendix. For the analysis, we assume that the data for each run in the different designs are the same as those obtained by Williams. Lin (1993) proposed design number 5 for this situation and used a stepwise regression procedure to identify the active factors. We adopted the same procedure for each of these 8 designs (the stepwise procedure in SAS - Statistical Analysis System). For inclusion and exclusion the default option ("significance" level .15) was used first. Then we considered the level .075 to get results for design 5 to be close to those of Lin (1993). Table 1 shows the factors which were picked as active by each design. For instance, in design 1 with significance level $(SL) = .075$, factor 17 was selected as active in the first step and factors 15, 4, 22 and 10 were selected in subsequent steps. As can be seen from the table, different factors are chosen in different designs.

In the original study Williams (1968) selected factors 15, 20, and 17 as having major effects and 4, 22, 14 and 8 as having moderate effects. Combining process knowledge with the experimental results Williams chose factors 15, 10, 20 and 4 for subsequent use. Design 5 leads to factors 15, 12, 20, 4 and 10 as active. This set is somewhat similar to the one Williams (1968) used. However, factors 17 and 22 were not chosen as active in design 5; instead 12 and 10 were selected.

For comparison, we adopted all subsets regression, which evaluates every possible subset of factors. The implementation we used (in S-Plus) can accommodate problems with more factors than runs. We restricted our search for the best subsets to models with 5 or less factors; note that there are $\sum_{i=1}^{5} \binom{23}{i} = 44,551$ subsets of size 1 to 5. By restricting attention to small subsets, the procedure runs quickly (under a second on a 200 Mhz pentium). The results are shown in Table 2. For example in design 1, factors 1, 15 and 20 provide the "best subset" of size 3; 1, 14, 15, and 20 provide the "best subset" of size 4; 1, 3, 14, 15, and 20 lead in subsets of size 5. Here best is in the sense of R^2. From Table 2 we note the following.

(i) Different designs lead to different choices with subsets of size $i = 1, 2, 3, 4, 5$. No single design identifies the same five factors as an analysis of the full 28 runs.

Table 1: Rubber data: Factors included after five steps in a stepwise procedure

Design	Steps	1	2	3	4	5	R^2
1	SL = .15	17	15	4	22	10	.90
	SL=.075	17	15	4	22	10	.90
2	SL=.15	15	24	18	13	8	.97
	SL=.075	15	24	18	13	8	.97
3	SL=.15	15	20	3	4	22	.92
	SL=.075	15	20	-	-	-	.73
4	SL=.15	2	13	8	3	20	.84
	SL=.075	2	13	8	-	-	.69
5*	SL=.15	15	12	20	4	10	.97
	SL=.075	15	12	20	4	10	.97
6	SL=.15	4	22	23	18	24	.88
	SL=.075	4	22	23	-	-	.74
7	SL=.15	14	12	11	23	-	.89
	SL=.075	14	12	11	23	-	.89
8	SL=.15	15	22	8	17	1	.94
	SL=.075	15	22	8	17	1	.94

Notes: * Design 5 is the same as that used in Lin (1993)
"-" indicates that the procedure stopped before

Table 2: Rubber Data: factors from 'all subsets selection' procedure. The row marked "all" refers to results using all 28 runs. Each row corresponds to a model identified by all subsets, with a ● indicating an active factor.

Design	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	17	18	19	20	21	22	23	24	R^2
all				●											●	●			●		●			0.6881
																●								0.314
				●											●									0.593
1	●														●				●					0.77
	●													●	●				●					0.874
	●			●										●	●				●					0.936
															●									0.583
															●								●	0.798
2															●		●						●	0.887
													●		●		●						●	0.935
								●					●		●		●						●	0.967
															●									0.556
															●				●					0.729
3					●			●							●									0.854
	●				●			●							●									0.957
	●				●			●							●					●				0.974
		●																						0.379
		●											●											0.548
4		●				●													●					0.694
		●				●										●			●					0.851
		●				●								●		●			●					0.903
															●									0.632
												●			●									0.74
5												●			●				●					0.871
				●								●			●				●					0.955
				●						●		●			●				●					0.973
				●																				0.362
		●																				●		0.632
6	●	●																				●		0.792
	●	●													●							●		0.895
	●	●													●			●				●		0.955
														●										0.297
												●		●										0.515
7											●	●		●										0.775
											●	●		●								●		0.893
								●			●	●		●								●		0.917
															●									0.483
															●						●			0.669
8								●							●						●			0.789
															●				●	●	●			0.873
	●							●							●	●					●			0.94

(ii) In designs 2 and 5 stepwise and all subsets regression identify the same factors (up to step 5). In design 7 the two procedures lead to the same factors up to step 4 and in design 8 only up to step 3. In other designs they lead to different factors, often with substantially inferior R^2 values. For example with design 4, the best 5 factor model identified with stepwise has $R^2 = .84$ while all subsets identifies a 5 factor model with $R^2 = .97$.

(iii) The R^2 criterion is overly optimistic. The best 5 factor model for the full 28 runs has $R^2 = 0.69$, while all eight designs with 14 runs yield models with $R^2 > 0.90$.

(iv) A half normal plot of effects for the original PB design identifies the five factors 4, 15, 17, 20, 22 as having the largest effects. The plot however indicates that factor 15 is the only important factor; it is difficult to make concrete statements about other factors.

Table 2 and (i)-(iv) above indicates sensitivity of results to the design used. Is looking only at the single best fitting model the reason important factors are not identified in some designs? To explore this possibility, we now consider the best five models of each size for each of the eight designs. This may be viewed as an extension of Table 2, in which each row is expanded to five rows, giving the five best models rather than the single best. In any situation with many predictors it is good practice to consider more than a single model, since often many models can fit the same data well.

Our findings were not entirely encouraging. An indication of this is factor 15, the only factor that is clearly active when the full 28 runs of data are used. In five of the eight designs, 15 is clearly identified in models of size 1 through 5. In the remaining three designs, 15 is not identified as important in all of the five different model sizes. The worst case is design 4, in which 15 is only identified among models with one active term. None of the models with 2-5 active terms for design 4 include factor 15, despite its clear importance among other factors when the full data are considered.

Whether a single model or many models are considered, the important point is that different designs lead to the identification of different factors as active. Even all subsets, which identifies more models than stepwise, can still mislead in some cases.

3 Misspecification

The discussion in the previous section points to two issues: (i) Choice of a design. (ii) Misspecification. We now discuss the second issue more fully.

When the number of factors is very large it is inevitable that some of the inactive factors would be selected with high probability. This is a serious problem but may be manageable as long as active factors are also selected with high probability. However, sometimes the procedures fail to select active factors and select inactive factors instead (see, for instance, Chipman et. al. (1997), Hamada and Wu (1992) and Wu (1993)). The nonorthogonality of the columns of the design matrix X is the root of the problem.

To highlight the misspecification issue we conducted a simulation study, with the two designs (14 run, 12 run) given in Lin (1993, 1995) and model (1). Although inferior to all subsets, we also consider stepwise since it is a commonly used tool.

3.1 14 run design with 23 factors used in Lin (1993)

Step 1: Data generation:
We generate $n = 14$ $N(0,1)$ random variates, $\epsilon_i, i = 1, \ldots, 14$ and then using the model (1) generate $y_1, \cdots, y_{14}$ for specific β values.
Simulated Models:

(1) Pure noise: $\beta_j = 0$ for all j.

(2) $$\beta_j = \begin{cases} .5, 1, 20 & j = 1 \\ 0 & j \neq 1 \end{cases}$$

(3) $\beta_2 = \beta_7 = 1,$ other β's zero
$\beta_2 = \beta_7 = 20,$ other β's zero

(4) $\beta_1 = \beta_2 = \beta_3 = 1,$ other β's zero
$\beta_2 = \beta_7 = \beta_{13} = .5,$ other β's zero
$\beta_2 = \beta_7 = \beta_{13} = 1,$ other β's zero
$\beta_2 = 5, \beta_7 = 10, \beta_{13} = 20,$ other β's zero
$\beta_2 = 14, \beta_7 = 20, \beta_{13} = 20,$ other β's zero
$\beta_2 = \beta_7 = \beta_{13} = 20,$ other β's zero

Step 2: Analysis for factor selection:
We use forward selection with a specific SL for including a factor, stop after step 5 and record the selected factors.

Step 3: Repeat Steps 1 and 2, $N = 200$ times and record the number of times each factor was selected.

In all cases the intercept $\beta_0 = 0$. For this particular design, correlations between columns are either ± 0.14 or ± 0.43. X_1 has a correlation of 0.43 with each of X_2, X_7, X_{13}, making case (4) a challenging problem.

In the pure noise case with SL = .05, 79% of the times at least one factor was selected as active in 5 steps. This was not unexpected. With 23 contrasts tested to zero individually and assuming them to be independent (this is not quite true in our case) there is a chance $1 - (1 - \alpha)^{23}$ of selecting at least one significant contrast where α is the SL for inclusion. When $\alpha = .05$ this probability $= .70$ which is in the vicinity of the simulated value. We cannot avoid this situation even if we have a large number of observations. On the other hand if the present experiment is considered as a device to reduce the number of factors which need to be carefully examined in a later experiment, then it is important to know whether the procedure leads to the active factors with high probability. For this purpose we consider cases 2-4 where one, two, or three β's are different from zero. The results from a stepwise procedure are presented in Table 3. We note the following from the table:

(i) When only one β is different from zero (or it is much larger than the rest), the corresponding factor gets selected in the first step. For instance, with a $SL = .1$, X_1 gets selected 100% of the cases if $\beta_1 = 20$, 87% of the cases when $\beta_1 = 1$ and only 36.5% when $\beta_1 = .5$. When SL=.05 and $\beta_1 = .5$, X_1 gets selected only in 27.5% of cases.

(ii) When two β's are important and very large ($\beta_2 = \beta_7 = 20$), the corresponding factors get selected about 78% of the time in 2 steps with SL=0.10. This drops to 55% when $\beta_2 = \beta_7 = 1$. In both cases the inactive factor X_1 is selected in the first forward step about 22% of the time.

(iii) When more than two β's are important, conclusions are different depending on which β's are non zero. For example when $\beta_1 = \beta_2 = \beta_3 = 1$ and other β's are zero (case 4), with $SL = .1$ the active factors X_1, X_2, X_3 get selected about 55% of the time in five steps. However, if $\beta_2 = \beta_7 = \beta_{13} = 1$ and other β's are zero (active factors are now assigned to columns 3, 8, and 14 but the design is the same), then with $SL = .1$ the combination (X_2, X_7, X_{13}) get selected only 4% of the time in five steps and .5% of the time in three steps; X_1 gets selected 91.5% of the time in the first step. The same value (=1) is used for both cases but the active factors are in different columns of the matrix and the non orthogonality of the columns creates the difficulty.

    Even when effects are very large, forward selection can mislead. When $\beta_2 = 14$, $\beta_7 = \beta_{13} = 20$ or $\beta_2 = \beta_7 = \beta_{13} = 20$, the combination (X_2, X_7, X_{13}) is never selected while X_1 is selected 100% of the time. On the other hand, if $\beta_2 = 5$, $\beta_7 = 10$, $\beta_{13} = 20$, the combination is selected in 100% of the cases. Factors selected as active depend on which columns of the design matrix are assigned to the real active factors and that is difficult to do in advance. This is a result of the correlation structure of the columns of the design matrix.

The behaviour of stepwise regression described in (iii) above can be better understood by some simple calculations. Since the true model is known, the expected value of the estimated regression coefficients $\hat{\beta}$ may be calculated for any specific model. If X_T is the design matrix corresponding to the true model, and X is the design for a model under consideration, then

$$\mathrm{E}(\hat{\beta}) = (X'X)^{-1}X'\mathrm{E}(Y) = (X'X)^{-1}X'X_T\beta \tag{2}$$

Calculation of (2) for the six simulated models in case 4 is given in Table 4. In each case, the calculation assumes only an intercept and a *single* predictor in the model. This corresponds to the choices that stepwise regression would be faced with at the first step. At this point in the algorithm, the standard error of the estimated coefficients will be $\mathrm{E}(s)/\sqrt{14} \approx 0.27$. The

Table 3: Simulation results: stepwise selection pattern of factors, Lin's 14 run design

Factors with Real Effects	Corresp-onding β values	Percent of times real effect factors selected in k steps $k = \#$ factors in col. 1		in 5 steps		Percent of time X_1 selected in the first step	
		SL=0.10	SL=0.05	SL=0.10	SL = 0.05	SL =0.10	SL=0.05
X_1	0.5	36.5	27.5	56.5	38		
	1	87	85.5	93	90.5		
	20	100	100	100	100		
X_2, X_7	1,1	55	56	67	62.5	24.5	22
	20, 20	78	74	78	74	22	26
X_1, X_2, X_3	1, 1, 1	40.5	38	55	45.5		
X_2, X_7, X_{13}	0.5, 0.5, 0.5	0	0	1	1	43.5	46.5
	1, 1, 1,	.5	.5	4	2	91.5	91
	5, 10, 20	100	100	100	100	0	0
	14, 20, 20	0	0	0	0	100	100
	20, 20, 20	0	0	0	0	100	100

Table 4: Expected value of first regression coefficient added under various true models. The expectations for coefficients other than 1, 2, 7, 13 are quite small and not presented.

Active terms in true model	First factor added: 1	2	7	13
$(\beta_2, \beta_7, \beta_{13}) = (0.5, 0.5, 0.5)$	0.64	0.36	0.36	0.36
$(\beta_2, \beta_7, \beta_{13}) = (1, 1, 1)$	1.29	0.71	0.71	0.71
$(\beta_2, \beta_7, \beta_{13}) = (5, 10, 20)$	15.00	0.71	6.43	17.86
$(\beta_2, \beta_7, \beta_{13}) = (14, 20, 20)$	23.14	8.29	15.14	15.14
$(\beta_2, \beta_7, \beta_{13}) = (20, 20, 20)$	25.71	14.29	14.29	14.29

Table 5: Simulation results with all subset procedure, Lin's 14 run design

Factors with Real Effects	β-values	Percent times correct factors selected in subsets of size k *		
		$k=3$	$k=4$	$k=5$
X_1, X_2, X_3	1,1,1	52 (77)	50 (77)	51 (70)
X_2, X_7, X_{13}	.5,.5,.5	1 (8)	4 (9)	5 (15)
	1,1,1	33 (57)	44 (65)	43 (67)
	5,10,20	100 (100)	100 (100)	100 (100)
	14,20,20	100 (100)	100 (100)	100 (100)
	20,20,20	100 (100)	100 (100)	100 (100)

* Note: The first number gives frequency that the highest ranked model contained all active factors. The number in parentheses gives the frequency with which one of the five highest ranked models contained all the active factors.

equality is not exact since the expected value of s may not be close to σ because the true model is not used. This value is quite small compared to coefficients as large as 20. In all but the third case, factor 1 has the largest expected value at the first step, and is consequently likely to be included. When the coefficients are large (the last two cases), stepwise always selects factor 1, even though it is inactive.

We also performed simulations for case (4) and used the all subsets regression procedure to see how the selection pattern would be affected. The results are given in Table 5. We note the following.

(i) All subsets performs notably better than stepwise in the more difficult problem involving X_2, X_7 and X_{13}. It finds the correct model more than a third of the time when all three coefficients are 1. In the cases of $(\beta_2, \beta_7, \beta_{13}) = (14, 20, 20)$ and $(20, 20, 20)$ where stepwise fails to find the right model, all subsets identifies the model 100% of the time.

(ii) Unlike stepwise, all subsets regression can identify different models of a given size with good fit. For this reason, Table 5 also gives the percentage of times that the correct model is among the five best of a given size. So for example, when $\beta_2 = \beta_7 = \beta_{13} = 1$, this model was ranked first among 3 term models by all subsets 33% of the time. It appeared in the top five models of size three 57% of the time.

Since SS designs are intended to screen variables, it is quite likely that practitioners will examine more than one model. The findings of this simulation study indicates that even when SS designs are used in this fashion, it is possible that the correct factors will not be identified.

Table 6: A Hadamard matrix of order 12

1	2	3	4	5	6	7	8	9	10	11	12
$\mathbf{1}$	$\mathbf{1}$	$-\mathbf{u}_3$	$\mathbf{1}$	$-\mathbf{u}_3$	$-\mathbf{u}_3$	$-\mathbf{u}_2$	$-\mathbf{u}_2$	$-\mathbf{u}_2$	$\mathbf{u}_1$	$\mathbf{u}_1$	$\mathbf{u}_1$
$\mathbf{1}$	$\mathbf{1}$	$\mathbf{u}_3$	$-\mathbf{1}$	$\mathbf{u}_1$	$\mathbf{u}_2$	$\mathbf{u}_2$	$\mathbf{u}_1$	$\mathbf{u}_3$	$-\mathbf{u}_3$	$-\mathbf{u}_2$	$-\mathbf{u}_1$
$\mathbf{1}$	$-\mathbf{1}$	$\mathbf{1}$	$\mathbf{u}_1$	$\mathbf{u}_3$	$\mathbf{u}_2$	$-\mathbf{u}_2$	$-\mathbf{u}_1$	$\mathbf{u}_2$	$-\mathbf{u}_3$	$\mathbf{u}_1$	$\mathbf{u}_3$
$\mathbf{1}$	$-\mathbf{1}$	$-\mathbf{1}$	$-\mathbf{u}_1$	$-\mathbf{u}_2$	$-\mathbf{u}_3$	$\mathbf{u}_1$	$\mathbf{u}_2$	$-\mathbf{u}_2$	$\mathbf{u}_3$	$-\mathbf{u}_3$	$-\mathbf{u}_1$

3.2 12 run design

Lin (1995) presented an algorithm to generate certain SS designs and designs for 12 runs. General construction methods are available to produce these designs with number of factors as high as 66. Basically the design is constructed from the columns of a standard Hadamard matrix. Using the notation $\mathbf{1}' = (1, 1, 1), \mathbf{u}_1' = (1, -1, -1), \mathbf{u}_2' = (-1, 1, -1)$, and $\mathbf{u}_3' = (-1, -1, 1)$ we can write down one such Hadamard matrix of order 12 as given in Table 6.

The different columns of the design matrix would be obtained by taking componentwise products of two columns of this basic matrix leading to a design matrix with 66 columns. We denote the componentwise product of the columns i and j by "$i \times j$". Lin (1995) demonstrated the usefulness of such SS designs by carrying out a simulation study using the above design. In the simulation study factors corresponding to columns $1 \times 4, 1 \times 8$, and 8×10 were given large effects (β values 17, 24 and 15 respectively), those corresponding to 1×11, and 8×11 were given moderate effects (β values 3), and others were taken as zero. The results of the simulation indicated that the 5 active factors were always selected in the first five steps of a forward selection procedure.

We wish to demonstrate that the results depend on which columns of the X matrix correspond to the large and moderate effect factors. In our simulation study we used the same design as in Lin (1995) and we assigned the β values $17, 24, 15$ (large effects in Lin's simulation) to columns $8 \times 10, 6 \times 12$, and 9×11 and the β value 3 to columns 3×5 and 4×7, and zero to the remaining columns. 12 observations were generated from model (1) using the β-values indicated and a forward selection procedure as in Lin (1995) was used to select the active factors. In 200 repetitions, none of the active factors corresponding to columns $6 \times 12, 9 \times 11, 3 \times 5$ and 4×7 were picked up in the first five steps while the factor corresponding to column 8×10 was selected only 29% of the time. The factor corresponding to column 1×2 (inactive) was always the choice in the first step.

4 Additional Discussion and Concluding Remarks

Data analysis in Section 2, and the simulations and discussion in Section 3 indicate that one should be very cautious with the use of SS designs. The simulations, in particular show that there is a high chance of missing the real active factors and selecting the inactive ones instead. The assignment of factors to columns of these designs is crucial because of the correlation structure among the columns of the design. If SS designs are to be used, the results of Section 3 indicate that all subsets variable selection should be used rather than stepwise regression. Even all subsets can be problematic and does not provide a means for deciding how many factors are active.

In Section 2 we commented that some additional criterion is necessary to select a design. One suggestion made in this regard is to consider a design in which the average value of S_{ij}^2, where $S_{ij} = X_i X_j'$ (say, $E(S^2)$), is made as small as possible (for example see Booth and Cox (1962), Wu (1993)). In the case of a 12-run design with 13 factors it can be shown that the minimum value of $E(S^2) = 48$ and is attained by a design given in Wu (1993). In the eight different 14-run designs considered in Section 2, E(S^2) is either 7.79 (for designs 1 and 3) or 7.92 (for other designs), leaving little theoretical grounds for distinguishing between the eight designs. The dramatically different conclusions found in the eight different subsets appears to be mainly due to random variation in the data.

Although our findings are for specific SS designs and specific simulations/data, we feel that they are indicative of problems that can occur for SS designs in general. We do not condemn the use of SS designs, but urge caution in their application and careful follow-up experimentation to validate any models suggested by data from SS designs. Meyer, Steinberg, and Box (1996) and associated discussion give a recent discussion of follow-up designs for model discrimination.

Acknowledgments

B. Abraham was supported by grants from the Natural Sciences and Engineering Research Council of Canada and the Manufacturing Research Corporation of Ontario. H. Chipman was supported by a grant from the Natural Sciences and Engineering Research Council of Canada. We also would like to thank an associate editor and two anonymous referee for some very helpful comments on an earlier version.

References:

Booth, K.H.V. and D.R. Cox (1962). "Some Systematic Supersaturated Experimental Designs", *Technometrics*, 4, 489-495.

Box, G.E.P. and N.R. Draper (1987). *Empirical Model-Building and Response Surfaces*, New York, John Wiley.

Chipman, H., M. Hamada and C.F.J. Wu (1997). "A Bayesian Variable Selection Approach for Analyzing Designed Experiments with Complex Aliasing", *Technometrics*, 39(4), 372–381.

Hamada, M. and C.F.J. Wu (1992). "Analysis of Designed Experiments with Complex Aliasing", *Journal of Quality Technology*, 24, 130–137

Lin, D.K.J. (1993). "A New Class of Supersaturated Designs", *Technometrics*, 35(1), 28–31.

Lin, D.K.J. (1995). "Generating Systematic Supersaturated Designs", *Technometrics*, 37(2), 213–225.

Meyer, R. D., Steinberg, D. M., and Box, G. E. P (1996) "Follow-up designs to resolve confounding in fractional factorials", *Technometrics*, 38, 303–313.

Nguyen, N. (1996). "A Note on the Construction of Near-Orthogonal Arrays With Mixed Levels and Economic Run Sizes", *Technometrics*, 279–283.

Tang, B. and Wu, C. F. J. (1997) "A Method for Constructing Super-saturated Designs and its Es^2 Optimality", *Canadian Journal of Statistics*, 25, 191–201.

Vijayan, K. (1976). "On Hadamard Matrices and Submatrices", *Journal of Australian Mathematical Society A*, 22, 469-75.

Williams, K.R. (1968). "Designed Experiments", *Rubber Age*, 100, 65-71.

Wu, C.F.J. (1993). "Construction of Supersaturated Designs Through Partially Aliased Interactions", *Biometrika*, 80, 661-669.

Yamada, S. and Lin, D. K. (1997) "Supersaturated Design Including an Orthogonal Base", *Canadian Journal of Statistics*, 25, 203–213.

Ye, J. (1998) "On Measuring and Correcting the Effects of Data Mining and Model Selection", *Journal of the American Statistical Association*, 93, 120-131.

Appendix

Table 7: Rubber data, from Williams (1968). Runs used in Lin's (1993) design are marked with a *. The level of run 8, factor 20 is here shown as +. In the source reference, the lower level appears: that seems to be a typographical error.

Run	Factor Number																								Response
Number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	y
1*	+	+	+	−	−	−	+	+	+	+	+	−	+	−	−	+	+	−	−	+	−	−	−	+	133
2	−	+	−	−	−	−	+	+	+	+	−	+	−	−	+	−	+	+	+	+	−	+	+	−	49
3*	+	−	−	−	−	−	+	+	+	−	−	−	+	+	+	+	−	+	−	−	+	+	−	−	62
4*	+	+	−	+	+	−	−	−	−	+	−	+	+	+	+	+	+	−	−	−	−	+	+	−	45
5	+	+	−	−	+	+	−	−	−	−	−	−	−	−	−	−	+	+	−	+	+	+	−	+	88
6*	+	+	−	+	−	+	−	−	−	+	+	−	+	−	+	+	−	+	+	+	−	−	−	−	52
7	−	−	+	+	+	+	+	+	−	+	−	+	+	−	−	+	−	+	−	−	−	−	−	−	300
8*	−	−	+	+	+	+	−	+	+	−	−	−	+	−	+	+	+	−	−	+	−	+	+	+	56
9*	−	−	+	+	+	+	+	−	+	+	+	−	−	+	+	−	+	+	+	+	+	+	−	−	47
10*	−	−	−	−	+	−	−	+	−	+	−	+	+	+	−	+	+	+	+	+	+	−	−	+	88
11	+	−	+	−	−	+	−	−	+	−	−	−	+	+	−	+	+	+	+	−	−	−	+	−	116
12	−	+	+	+	−	−	+	−	−	+	+	−	+	+	−	+	+	+	−	−	+	+	+	+	83
13*	−	+	+	−	−	+	−	+	−	+	−	−	−	−	−	−	−	−	+	−	+	+	+	−	193
14	−	−	−	+	−	−	−	−	+	+	−	−	−	−	+	−	−	−	−	−	+	−	−	+	230
15	+	−	+	−	+	−	+	−	−	+	−	−	−	+	+	−	−	−	+	+	−	−	+	+	51
16	−	+	+	−	+	−	−	−	+	−	+	+	+	−	+	+	+	−	+	−	+	−	−	−	82
17*	−	−	−	−	−	+	+	−	−	−	+	+	−	−	+	−	+	+	−	−	−	−	+	+	32
18	+	−	+	+	−	−	−	+	−	−	+	+	+	−	+	+	−	+	+	+	+	+	+	+	58
19	+	−	−	+	+	−	+	+	−	−	+	−	−	−	−	−	+	−	+	−	−	+	−	−	201
20	+	+	+	−	+	+	−	+	+	+	+	+	−	+	+	−	−	+	−	−	−	+	−	+	56
21	−	+	−	+	−	+	+	−	+	−	−	+	+	+	−	+	−	−	+	+	−	+	−	+	97
22*	+	+	+	+	−	+	+	+	−	−	−	+	−	+	+	−	+	−	+	−	+	−	−	+	53
23*	−	+	−	+	+	−	−	+	+	−	+	−	−	+	−	−	−	+	+	−	−	−	+	+	276
24*	+	−	−	−	+	+	+	−	+	+	+	+	+	−	−	+	−	−	+	−	+	+	+	+	145
25*	+	+	+	+	+	−	+	−	+	−	−	+	−	−	−	−	−	+	−	+	+	−	+	−	130
26	−	+	−	−	+	+	+	+	−	−	+	−	+	+	+	+	−	−	−	+	+	−	+	−	55
27	+	−	−	+	−	+	−	+	+	+	+	+	−	+	−	−	+	−	−	+	+	−	+	−	160
28*	−	−	+	−	−	−	−	−	−	−	+	+	−	+	−	−	−	−	−	+	−	+	−	−	127

Design Number	Rows from the PB design													
1 (2)	1	2	4	5	7	9	11	14	18	22	23	24	26	28
3 (4)	1	3	4	5	7	9	13	15	16	17	18	21	23	27
5 (6)	1	3	4	6	8	9	10	13	17	22	23	24	25	28
7 (8)	1	4	9	10	11	13	14	17	18	19	20	21	25	26

Table 8: 14-run supersaturated designs from a Hadamard matrix of order 28. Even numbered designs (in parentheses) are the complement of the rows listed.