A New Class of Supersaturated Designs

Dennis K. J. Lin

Department of Statistics
The University of Tennessee
Knoxville, TN 37996

Supersaturated designs are useful in situations in which the number of active factors is very small compared to the total number of factors being considered. In this article, a new class of supersaturated designs is constructed using half fractions of Hadamard matrices. When a Hadamard matrix of order N is used, such a design can investigate up to $N - 2$ factors in $N/2$ runs. Results are given for $N \leq 60$. Extension to larger N is straightforward. These designs are superior to other existing supersaturated designs and are easy to construct. An example with real data is used to illustrate the ideas.

KEY WORDS: Hadamard matrices; Plackett and Burman designs; Random balance designs.

Many preliminary studies in industrial experimentation contain a large number of potentially relevant factors, but often only a few are believed to have actual effects. This is sometimes called *effect sparsity*. The basic problem here is how to identify these few active factors in an efficient way. Knowing every main effect can be wasteful because nonsignificant factors are not usually of interest. One approach is to use a so-called *supersaturated design*—namely, a factorial design with n observations and k factors, with $k > n - 1$. If the first-order model is assumed (as are all main-effect models) and if the number of significant factors is expected to be small, a supersaturated design can save considerable cost.

Satterthwaite (1959) suggested constructing such a design at random. Although the idea of *random balance designs* is interesting, the designs themselves are of questionable usefulness. (See Youden, Kempthorne, Tukey, Box, and Hunter 1959.) Booth and Cox (1962) were the first to examine this problem systematically. They provided seven supersaturated designs obtained via computer search. Apart from these computer-generated designs, the construction problem has not been addressed in the literature.

This article discusses a class of special supersaturated designs, which can be easily constructed via *half fractions of Hadamard matrices*. These designs can examine $k = N - 2$ factors with $n = N/2$ runs, where N is the order of Hadamard matrix used. The basic assumption here—that there are only a few active main effects—occurs frequently in industrial situations. Results are presented here for the cases $N \leq 60$ (i.e., supersaturated designs with $n = N/2 \leq 30$ runs); extension to higher order Hadamard matrices is straightforward. The Plackett and Burman (1946) designs, which can be viewed as a special class of Hadamard matrices, are used to illustrate the basic construction method.

The article is organized as follows. In Section 1, the construction method is introduced; comparisons with other supersaturated designs are also made. It is shown that the designs given here are superior to other existing supersaturated designs and are easy to construct. In Section 2, an example with real data is given, and a data-analysis method for supersaturated designs is demonstrated through the example.

1. CONSTRUCTION METHODS AND SOME RESULTS

Table 1 shows the original 12-run Plackett and Burman design. If we take the column **(11)** as the *branching column*, then the total 12 runs (rows) can be split into two groups, Group I with the sign of +1 in column **(11)** (rows 2, 3, 5, 6, 7, and 11), and Group II with the sign of −1 in column **(11)** (rows 1, 4, 8, 9, 10, and 12). Deleting column **(11)** from Group I causes the columns **1–10** then to form a supersaturated design to examine $N - 2 = 10$ factors in $N/2 = 6$ runs (runs 1–6, as indicated in Table 1). It can be shown that, if Group II is used, the resulting supersaturated design is an equivalent one.

In general, a Plackett and Burman (1946) design matrix can be split into two half fractions according to a specific branching column whose signs equal +1 or −1. Specifically, take only the rows that have +1 in the branching column. Then the $N - 2$ columns other than the branching column will form a supersaturated design for $N - 2$ factors in $N/2$ runs. Of course, the underlying model is the first-order (main-effect) model.

All possible choices of branching columns have been studied. For $N \leq 60$, the selection of a particular branching column makes no difference, except as

*Table 1. Supersaturated Design Derived From the Hadamard Matrix of Order 12 (using **11** as the branching column)*

Run	Row	*I*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*(11)*
	1	+	+	+	−	+	+	+	−	−	−	+	−
1	2	+	+	−	+	+	+	−	−	−	+	−	+
2	3	+	−	+	+	+	−	−	−	+	−	+	+
	4	+	+	+	+	−	−	−	+	−	+	+	−
3	5	+	+	+	−	−	−	+	−	+	+	−	+
4	6	+	+	−	−	−	+	−	+	+	−	+	+
5	7	+	−	−	−	+	−	+	+	−	+	+	+
	8	+	−	−	+	−	+	+	−	+	+	+	−
	9	+	−	+	−	+	+	−	+	+	+	−	−
	10	+	+	−	+	+	−	+	+	+	−	−	−
6	11	+	−	+	+	−	+	+	+	−	−	−	+
	12	+	−	−	−	−	−	−	−	−	−	−	−

noted later, for the resulting supersaturated designs. The only exception is the case $N = 52$, which is not a cyclic type. (For designs with cyclic structure, as are most Plackett and Burman designs, there exists a cross-balance among the columns.) Recall that the 52-run case was constructed (Plackett and Burman 1946, p. 323) via permuting five 10×10 blocks of signs, plus the first column whose signs are $(+, -, +, -, \ldots, +, -)$ in each block. Using the first column for branching results in 25 highly correlated pairs (five in each block, correlation being 22/26). If any other column is used, only the pair involving the first column produces correlation = 22/26. Thus deleting the first column and then using any one of others as the branching column gives a supersaturated design with 49 columns in 26 runs, and this is what I recommend.

Comparisons with designs given by Satterthwaite (1959) and Booth and Cox (1962) are made in Table 2. These are the only supersaturated designs available in the literature, apart from some ad hoc procedures. To provide a fair basis of comparison, Booth and Cox's criterion, $E(s^2) = \Sigma s_{ij}^2/\binom{k}{2}$, the average of s^2, where $s_{ij} = \Sigma x_i' x_j$ is the sum of cross-products of any two columns (x_i and x_j, say) from the k design columns, is adopted. $E(s^2)$ gives an intuitive measure of nonorthogonality—the smaller, the better.

Table 2. Comparison of the Expectations of s^2 for Selected Designs

n	*k*	*Random balance*	*Booth and Cox (1962)*	*HFHM**	*HFHM* largest* $\lvert s_{ij}\rvert$
12	22	13.09	—	6.85	.333
	16	13.09	7.06	6.27	
	18	13.09	9.68	6.59	
	24	13.09	10.26	—	
18	34	19.06	—	9.82	.333
	24	19.06	13.04	9.22	
	30	19.06	15.34	9.74	
	36	19.06	16.44	—	
24	46	25.04	—	12.80	.333
	30	25.04	12.06	11.59	
6	10	7.20	—	4.00	.333
10	18	11.11	—	5.88	.600
14	26	15.07	—	7.84	.429
22	42	23.05	—	11.80	.273
26	49	27.04	—	13.80	.385
30	58	31.03	—	15.79	.200

*HFHM = half fraction Hadamard matrices.

In the first portion of Table 2, seven designs given by Booth and Cox (1962) are compared with the designs obtained here. Note that the designs suggested in this article can examine up to $N - 2$ factors in $N/2$ runs and thus cannot be compared with the two designs given by Booth and Cox (1962, designs III and VI). The $E(s^2)$ value for random balance design is $n^2/(n - 1)$ (see Booth and Cox 1962, p. 494). In all cases, the $E(s^2)$ values of designs derived from Hadamard matrices are close to $(2n + 3)/4$. Judged by $E(s^2)$, the designs given here are superior to the others. When $k < N - 2$ factors are investigated, one can always remove certain columns from the complete design. Designs given in Table 2 select the k columns that have the best $E(s^2)$ values. In practice, any k columns can be used because their $E(s^2)$ values are pretty much the same. For example, when $(k, N - 2) = (30, 46)$, the ratio of the maximum and minimum $E(s^2)$ values among all the choices is 1.24:1. Such a ratio, of course, tends to 1 as k increases (and N fixed). When k is much smaller than $N - 2$, the $E(s^2)$ value can vary widely, but the supersaturated design is not recommended for such cases.

The second portion of Table 2 compares all other designs ($n = N/2 \leq 30$). The last column of Table 2 shows the maximum cross-products of any two columns in the design and measures the largest absolute correlation among the columns. We would like to keep this value as small as possible. Note that some Hadamard matrices are absent from Table 2 ($N = 2n = 16$, 32, 40, and 56). These designs are constructed via foldover. Half fractions of such folded designs result in two equivalent sets of columns and thus cannot be used as supersaturated designs.

Table 3. Half Fractions of Williams's (1968) Data

	Factor																								
Run	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	Response
1	+	+	+	−	−	−	+	+	+	+	+	−	+	−	−	+	+	−	−	+	−	−	−	+	133
2	+	−	−	−	−	−	+	+	+	−	−	−	+	+	+	+	−	+	−	−	+	+	−	−	62
3	+	+	−	+	+	−	−	−	−	+	−	+	+	+	+	+	+	−	−	−	−	+	+	−	45
4	+	+	−	+	−	+	−	−	−	+	+	−	+	−	+	+	−	+	+	+	−	−	−	−	52
5	−	−	+	+	+	+	−	+	+	−	−	−	+	−	+	+	+	−	−	+	−	+	+	+	56
6	−	−	+	+	+	+	+	−	+	+	+	−	−	+	+	−	+	+	+	+	+	+	−	−	47
7	−	−	−	−	+	−	−	+	−	+	−	+	+	+	−	+	+	+	+	+	+	−	−	+	88
8	−	+	+	−	−	+	−	+	−	+	−	−	−	−	−	−	−	−	+	−	+	+	+	−	193
9	−	−	−	−	−	+	+	−	−	−	+	+	−	−	+	−	+	+	−	−	−	−	+	+	32
10	+	+	+	+	−	+	+	+	−	−	−	+	−	+	+	−	+	−	+	−	+	−	−	+	53
11	−	+	−	+	+	−	−	+	+	−	+	−	−	+	−	−	−	+	+	−	−	−	+	+	276
12	+	−	−	−	+	+	+	−	+	+	+	+	+	−	−	+	−	−	+	−	+	+	+	+	145
13	+	+	+	+	+	−	+	−	+	−	−	+	−	−	−	−	−	+	−	+	+	−	+	−	130
14	−	−	+	−	−	−	−	−	−	−	+	+	−	+	−	−	−	−	−	+	−	+	−	−	127

2. EXAMPLE

I used the real data set given by Williams (1968) to illustrate the usefulness and data-analysis procedure for the supersaturated designs in Section 1. The original problem concerned the effects of 24 predictor variables, and 28 runs of a Plackett and Burman design were used. (The full data display was also given by Box and Draper [1987, p. 175].) From these 28-run results, judged by relative size of mean squared terms, the factors **15**, **20**, and **17** were identified as important; factors **4**, **22**, **14**, and **8** were moderate. After combining with other experimental results, however, the most important factors were identified as **15**, **10**, **20**, and **4** and were recommended for subsequent studies (Williams 1968, table III).

A half fraction of these 28 runs (i.e., supersaturated design with 14 runs) can be used to identify the important effects as follows: Making the unused orthogonal column, (+ − + + − + − + + + − − + − − − + − − − − + + + + − − +)′, as the branching column results in a supersaturated design with 14 runs consisting of rows 1, 3, 4, 6, 8, 9, 10, 13, 17, 22, 23, 24, 25, and 28. See Table 3 for the supersaturated design and corresponding observations.

There are many possible ways of analyzing the data to identify the important effects. One approach is to use the stepwise selection procedure (e.g., Draper and Smith 1981, p. 307). The results of the analysis are given in Table 4. The important factors were identified as **15**, **12**, **20**, **4**, and **10** with an R^2 = 97.3%. Table 4 also shows the estimated effects and their corresponding t ratios at each step. The examination of residuals showed no unusual patterns. The conclusion is quite similar to that of Williams (1968) but was obtained with only 14 observations.

3. CONCLUDING REMARKS

It is known that Hadamard matrices of orders up to 12 are unique, but this is not true for higher order N. The Plackett and Burman design is simply one of them. For example, Hall (1961, 1965) showed that there are precisely five nonequivalent classes of order 16 and three of order 20. I have also studied all of these nonequivalent Hadamard matrices, as well as

Table 4. Stepwise (forward) Selection for the Data in Table 3

	Entering variables						
Step	15	12	20	4	10	σ	R^2
1	−53.2 **(−4.54)**					43.9	63%
2	−56.4 **(−5.42)**	−22.3 **(−2.14)**				38.5	74%
3	60.5 **(−7.75)**	−26.4 **(−3.38)**	24.8 **(−3.17)**			28.5	87%
4	−70.5 **(−12.96)**	−25.3 **(−5.19)**	−29.2 **(−5.86)**	22.1 **(4.09)**		17.8	95%
5	−71.3 **(−15.96)**	−26.8 **(−6.63)**	−28.0 **(−6.80)**	20.7 **(4.64)**	−9.4 **(−2.33)**	14.5	97%

NOTE: The numbers given in the table are estimated effects and their t ratios. The constant term is 102.8 at all steps.

three $\mathbf{H}_{28}$'s and six $\mathbf{H}_{52}$'s given by Wallis, Street, and Wallis (1972, appendix K). By examining all of these nonequivalent Hadamard matrices, however, I found that the supersaturated designs given here are unique (subject to permutation of rows or columns and sign changes).

When a supersaturated design is employed, the experimenters must recognize the key assumption—the existence of only a few dominant effects. Otherwise, the results can be misleading (see the insightful comments by Herzberg and Cox [1969]).

ACKNOWLEDGMENT

I am grateful to the editor, the associate editor, and three referees for their constructive comments, which have led to major improvements in the presentation. My thanks also to Robert W. Mee, The University of Tennessee, for valuable comments. This research was partially supported by National Science Foundation Grant DMS-9204007 and the Professional Development Awards, The University of Tennessee.

[Received March 1991. Revised May 1992.]

REFERENCES

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

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

Draper, N. R., and Smith, H. (1981), *Applied Regression Analysis*, New York: John Wiley.

Hall, M. J. (1961), "Hadamard Matrix of Order 16," *Jet Propulsion Laboratory Research Summary*, 1, 21–26.

——— (1965), "Hadamard Matrix of Order 20," *Jet Propulsion Laboratory Technical Report*, 1, 32–76.

Herzberg, A. M., and Cox, D. R. (1969), "Recent Work on the Design of Experiments: A Bibliography and a Review," *Journal of the Royal Statistical Society*, Ser. A, 132, 29–39.

Plackett, R. L., and Burman, J. P. (1946), "The Design of Optimum Multifactorial Experiments," *Biometrika*, 33, 303–325.

Satterthwaite, F. (1959), "Random Balance Experimentation" (with discussion), *Technometrics*, 1, 111–137.

Wallis, W. D., Street, A. P., and Wallis, J. S. (1972), *Combinatorics: Room Squares, Sum-Free Sets, Hadamard Matrices*, Berlin: Springer-Verlag.

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

Youden, W. J., Kempthrone, O., Tukey, J. W., Box, G. E. P., and Hunter J. S. (1959), Discussion of "Random Balance Experimentation," by F. E. Satterthwaite, and "The Applications of Random Balance Designs," by T. A. Budne, *Technometrics*, 1, 157–184.