

SPOTLIGHT INTERACTION EFFECTS IN MAIN-EFFECT PLANS: A SUPERSATURATED DESIGN APPROACH

Dennis K. J. Lin

Pennsylvania State University
University Park, Pennsylvania 16802-1913

Key Words

Effect sparsity; Normal plot; Plackett and Burman designs; Screening; Stepwise regression.

Introduction

Screening designs are typically used in the initial stage of an experimental investigation where many possible factors are suggested but only a small subset are anticipated to be “real” or to exhibit the so-called *effect sparsity* phenomenon. Conventionally, a first-order model is assumed, and interaction effects are tentatively ignored. A design suitable for such a first-order model is called a *main-effect design*.

The construction of first-order main-effect designs that are optimal in some sense has received a great deal of attention in the literature. Because of their relative simplicity of use, Plackett and Burman-type designs (including 2^{k-p} designs) are very popular in practice. When only main effects exist, these designs allow unbiased estimation of all of them. They are extremely important in screening situations.

Draper (1) comments that Plackett and Burman designs can be confusing unless (i) the interactions are small or negligible or (ii) there are relatively few “important” factors. Indeed, if interactions exist, the estimation is blurred (see Ref. 2), leading to incorrect conclusions. Box and Meyer (3) comment that “. . . in actual experiments conducted using Plackett and Burman nongeometric designs, potentially important effects involving interactions have probably been

missed.” It is shown here that important interaction effects can, in fact, be found easily in such nongeometric designs.

Certainly, a main-effect design is not intended for identifying interaction effects. Statistical designs for detecting interaction effects have been developed by many researchers, such as the resolution V design in Ref. 4 and the search design in Ref. 5. In this article, we study the case in which an experiment based on a main-effect design has been conducted and how we can spotlight interaction effects (which in the conventional wisdom is not believed to be possible). The proposed method, building on the recent work in Refs. 2 and 6–8, is introduced through an example in the next section. Some comparisons with existing methods are also made there. Theoretical formulation and assumptions are then discussed, followed by a more complicated example. Discussion and future research directions are given at the end.

Example

Consider the experiment in Ref. 9. A 12-run Plackett and Burman design was used to study the effects of seven factors (designated here as **A**, **B**, . . . , **G**) on the fatigue life of weld-repaired castings. The design and responses are given in Table 1. For the details of factors and level values, see Ref. 9. See also Refs. 10 and 11, for analyses of these data.

Plackett and Burman designs are traditionally known as *main-effect designs* because, if all interactions can tentatively be ignored, they can be used to estimate all main

Table 1. The Cast Fatigue Experiment

RUN	A	B	C	D	E	F	G	8	9	10	11	RESPONSE
1	1	1	-1	1	1	1	-1	-1	-1	1	-1	6.058
2	1	-1	1	1	1	-1	-1	-1	1	-1	1	4.733
3	-1	1	1	1	-1	-1	-1	1	-1	1	1	4.625
4	1	1	1	-1	-1	-1	1	-1	1	1	-1	5.899
5	1	1	-1	-1	-1	1	-1	1	1	-1	1	7.000
6	1	-1	-1	-1	1	-1	1	1	-1	1	1	5.752
7	-1	-1	-1	1	-1	1	1	-1	1	1	1	5.682
8	-1	-1	1	-1	1	1	-1	1	1	1	-1	6.607
9	-1	1	-1	1	1	-1	1	1	1	-1	-1	5.818
10	1	-1	1	1	-1	1	1	1	-1	-1	-1	5.917
11	-1	1	1	-1	1	1	1	-1	-1	-1	1	5.863
12	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	4.809

effects. There are many ways to analyze such a main-effect design. One popular way is the normal plot (see Ref. 10, Fig. 1). It appears that factor **F** is the only significant main effect. Consequently, a main-effect model is fitted as follows:

$$\hat{y} = 5.73 + 0.458\mathbf{F} \quad (R^2 = 44.5\%).$$

Note that the low R^2 is not so impressive. Can we safely ignore the interaction effects? Hunter et al. (9) claim that the design did not generate enough information to identify specific conjectured interaction effects. If this is not the case here, can we detect significant interaction effects?

Hamada and Wu (10) introduced the concept of *effect heredity*. After main effects were identified, they used forward selection regression to identify significant effects in a group, which consists of (1) the effects already identified and (2) the two-factor interactions having at least one component factor appearing among the main effects in item 1. In this particular example, a model for factor **F** and interaction **FG** was chosen and is given as follows:

$$\hat{y} = 5.7 + 0.458\mathbf{F} - 0.459\mathbf{FG} \quad (R^2 = 89\%). \quad (1)$$

The assumption of effect heredity is debatable in practice. Much evidence indicates that strong interactions do not necessarily contain factors associated with significant main effects (see, for example, numerous examples with real data sets in Ref. 12). Rather, the use of interaction plots in Ref. 13 to search two-factor interactions is considered to be an improved method. Graphical methods are always subjective. In this case, there are $\binom{7}{2} = 21$ interaction plots to be compared, meaning that some guessing will be necessary. There is no basic guideline to determine which interaction should be included in the final selection.

Now, from Table 1, if we generate all interaction columns, **AB**, **AC**, . . . , **FG**, together with all main-effect columns, **A**, **B**, . . . , **G**, we have $7 + 21 = 28$ columns. Treat all of those 28 columns in 12 runs as a supersaturated design (Ref. 8) as shown in Table 2. The largest correlation between any pairs of the 28 design columns is $\pm 1/3$.

Table 3 shows the results from a regular stepwise regression analysis (with $\alpha = 5\%$ for entering variables). The model

$$\hat{y} = 5.73 + 0.394\mathbf{F} - 0.395\mathbf{FG} - 0.191\mathbf{AE} \quad (R^2 = 95\%) \quad (2)$$

is a significantly better fit to the data than is Eq. (1). Note that the **AE** interaction, in general, would never be chosen under the effect heredity assumption. Practitioners may consider adding main effects **A**, **E**, and **G** to the final model because of the significance of interactions **FG** and **AE**.

In general, for most main-effect designs, such as Plackett and Burman-type designs (except for 2^{k-p} fractional factorials), one can apply the following procedure (but see the limitation in a later section):

Step 1. Generate all interaction columns, and combine them with the main-effect columns. We have now $k + \binom{k}{2} = k(k+1)/2$ design columns.

Step 2. Analyze these $k(k+1)/2$ columns with n experimental runs as a supersaturated design. Data analysis methods for such a supersaturated design are available; see, for example, Refs. 7, 8, 14, and 15.

Note that if the interactions are indeed inert, the procedure will work well; and if the effect heredity assumption is indeed true, the procedure will end up with the same

Table 3. Stepwise Selection for the Cast Fatigue Experiment

STEP	ENTERING VARIABLES						$\hat{\sigma}^2$	R^2
	FG	F	AE	EF	D	E		
1	-0.459 (-2.85)						0.558	44.74%
2	-0.459 (-6.12)	0.458 (6.11)					0.260	89.25%
3	-0.395 (-7.01)	0.394 (6.99)	-0.191 (-3.19)				0.183	95.26%
4	-0.395 (-8.05)	0.394 (8.03)	-0.191 (-3.66)	-0.087 (-1.89)			0.159	96.86%
5	-0.365 (-14.82)	0.406 (16.98)	-0.154 (-5.85)	-0.130 (-5.41)	-0.128 (-4.87)		0.077	99.4%
6	-0.387 (-33.97)	0.408 (39.66)	-0.148 (-13.01)	-0.128 (-12.34)	-0.122 (-10.76)	-0.054 (-5.24)	0.033	99.9%

conclusion as that of Ref. 10. The proposed procedure will always results in a better (or equal) performance than using Hadama and Wu's procedure. Next, we discuss the theoretical formulation, assumptions, and limitations of the proposed method.

Theoretical Formulation

Consider a supersaturated design in n experimental runs to investigate k ($>n-1$) factors. If \mathbf{X} denotes the $n \times k$ "design matrix" (namely original design matrix with all interaction columns, but without intercept column), our model is

$$\mathbf{Y} = \mu \mathbf{1} + \mathbf{X}\beta + \varepsilon,$$

where \mathbf{Y} is the $n \times 1$ observable data vector, μ is the intercept term, and $\mathbf{1}$ is an n vector of 1's; β is a $k \times 1$ fixed-parameter vector for the unknown factor effects, and ε is a vector assumed to be distributed as $N(0, \sigma^2 \mathbf{I}_n)$. Because k is larger than $n-1$, it is clear that the \mathbf{X} matrix cannot be full rank, and orthogonality is only possible for certain pairs of design columns.

Let $\mathbf{A} = \{i_1, i_2, \dots, i_p\}$ and $\mathbf{N} = \{i_{p+1}, i_{p+2}, \dots, i_k\}$ denote indexes of active and inert factors, respectively, so that $\mathbf{N} \cup \mathbf{A} = \{1, 2, \dots, k\} = \mathbf{S}$. We have null and alternative pairs $\mathbf{H}_j: \beta_j = 0$ and $\mathbf{H}_j^c: \beta_j \neq 0$, with \mathbf{H}_j true for $j \in \mathbf{N}$ and \mathbf{H}_j^c true for $j \in \mathbf{A}$. Under an effect sparsity assumption, we suppose that p is small relative to k . Forward selection proceeds by identifying the maximum F -statistics at successive stages. Let $F_j^{(s)}$ denote the F -statistics for testing \mathbf{H}_j at stage s , $s = 1, 2, \dots$. Sequentially, define the following:

$$j_1 = \arg \max_{j \in \mathbf{S}} F_j^{(1)},$$

$$j_2 = \arg \max_{j \in \mathbf{S} - \{j_1\}} F_j^{(2)},$$

$$j_3 = \arg \max_{j \in \mathbf{S} - \{j_1, j_2\}} F_j^{(3)},$$

and so forth, where

$$F_j^{(s)} = \frac{\text{MSE}(j | j_1, \dots, j_{s-1})}{\text{MSE}(j, j_1, \dots, j_{s-1})}.$$

Letting $F^{(s)} = \max F_j^{(s)}$, the forward selection procedure is defined by selecting variables j_1, \dots, j_f , where $F^{(f)} \leq \alpha$ and $F^{(f+1)} > \alpha$. If $F^{(1)} > \alpha$, then no variables are selected.

The basic assumption made here is the presence of both (1) *effect sparsity* (as will be discussed below) and (2) *effect prominence*—the active factors have (main or interaction) effects large enough to stand out from the experimental error or from the combined effects of unimportant factors—an assumption that is shared by all screening designs. In general, if an effect (main or interaction) is three times (or more) larger than the overall pooled standard deviation, such an effect can always be identified (Ref. 14).

The success of a supersaturated design, like most main-effect plans used in the screening process, depends heavily on the assumption of effect sparsity, a tacit assumption in many popular analysis tools, such as Normal plots. Box and Meyer (3) suggest a "20% rule" as a standard. Whereas Srivastava (5) shows that the number of active factors, p , must not exceed half of the experimental runs, n , for identifiability. This can be viewed as a guideline for the effect

Table 4. A 28-Run Plackett and Burman Design

RUN	FACTORS																												STRIP ADHESION y
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24					
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		

Source: Ref. 15.

sparsity assumption. In our specific case, the proposed method can only pick up, at most, $n/2$ relatively large (main or interaction) effects, depending on the particular design.

A Second Example

To investigate the epoxide adhesive system, Ref. 15 concerned 24 predictor variables. A 28-run Plackett and Burman design was conducted as given in Table 4. Note that columns 13 and 16 are identical. These are the temperatures for Oven 1 and Oven 2, respectively. For the simplicity of illustrating the proposed method, we simply delete column 16 in the following analysis.

A typical main-effect analysis (such as the Normal plot or regression method) indicates that factors **15**, **20**, **17**, **4** (and perhaps **22**, **14**, **8** and **1**) are significantly important with $R^2 = 64\%$ (or $R^2 = 81\%$ using all eight predicted variables; see Ref. 12, pp. 545–546). This conclusion, as noted in Ref. 15 (Table III), is very different from other experiments. Such an inconsistency clearly indicates the possibility of significant interaction effects. Of course, in the “standard” main-effect analysis, all interaction effects have been ignored. Can we alleviate such an assumption here?

If two-factor interactions are also under consideration, can we identify certain potential two-factor interaction effects using the original experimental results in Table 4. Following the procedure previously mentioned, we shall (1) generate all two-factor interaction columns and (2) perform a stepwise regression analysis with all main and two-factor interaction ($23 + \binom{23}{2} = 276$) effects columns as regressors. It is found that factor **15** and interactions **5** × **21**, **18** × **21**, **11** × **19**, **10** × **11**, **15** × **21**, and **3** × **5** are significant (with $R^2 = 92\%$). This indicates that (1) factor **15** is indeed active and (2) apart from factor **15**, the response can be explained either by main effects **20**, **17**, and **4**, or by interactions **5** × **21**, **18** × **21**, **11** × **19**, **10** × **11**, **15** × **21**, and **3** × **5**.

A close look at the correlation structure shows that factor **20** is somewhat correlated (correlation 3/7) with **5** × **21** and **18** × **21**; Factor **4** is correlated with **5** × **21** and **11** × **19**; all other effects are either orthogonal or with correlation $\pm 1/7$. This may well explain why factors **20** and **4** are significant in the main-effect fitting. A follow-up experiment thus needs to include factors **15**, **21**, **11**, **10**, **19**, **18**, and **5**. A 12-run Plackett and Burman design (or a resolution V design of 16 runs) is recommended. Note that a regular approach to estimate all two-factor interactions will result in a resolution V design which requires 512 runs for examining 23 factors (see Ref. 6).

Discussion

Conventional wisdom for analyzing a main-effect design is based on certain hierarchical assumptions, such as “all interactions are null.” In this article, I have introduced a technique from a supersaturated design perspective to spotlight two-factor interactions from a main-effect plan, a more flexible analysis to alleviate these assumptions. The major risk taken here is a *false positive effect* (i.e., to misclassify a null factor). Incorporating some hierarchical assumptions may reduce such an error. Given the large number of effects, a smaller α -level (such as 1%) may be required. In addition, as pointed out in Ref. 14, a screening experiment is typically conducted in the early stage of a study. In this case, a *false negative effect* (i.e., missing a real active factor) is much more serious than a false positive one. The latter can generally be resolved by consequent confirmatory runs. In fact, a data analysis method to control false positive errors can be found in Ref. 16.

The proposed method can be used for almost any main-effect plan, provided that the interaction columns are not *fully* confounded with the main-effect or other interaction columns. Certainly, as a general rule, the weaker the confounding, the better. The 12-run Plackett and Burman design is a particularly good example for the application of such a technique. Other examples given in Ref. 10, some with more than two levels, can also be analyzed in a similar manner. For two-level Plackett and Burman-type designs, the confounding patterns have been studied in Ref. 2. Theoretical justifications for those supersaturated designs have been recently developed (see Ref. 17). Note that the main-effect (resolution III) 2^{k-p} fractional factorial designs do not form a good basis for the analysis described here.

Higher-order (higher than second-order) interactions can also be considered, if necessary, provided they are not fully confounded with other design columns. Of course, as more columns enter the selection procedure, the probability for a false positive error also increases. In general, it is wise to combine several data analysis methods to ensure the results. Because of its flexibility, it is recommended to include the proposed procedure in the regular toolbox for analyzing “main-effect” designs.

Acknowledgments

The author is most grateful to Max Morris (Oak Ridge National Laboratory) and Robert Mee (University of Tennessee) for many constructive remarks which led to major improvements in the presentation. This research was par-

tially supported by the National Science Foundation, via Grant DMS-9704711.

References

1. Draper, N. R., Plackett and Burman Designs, in *Encyclopedia of Statistical Sciences*, John Wiley & Sons, New York, 1985, pp. 754–758.
2. Lin, D. K. J. and Draper, N. R., Generating Alias Relationships for Two-Level Plackett and Burman Designs, *Comput. Statist. Data Anal.*, 15, 147–157 (1993).
3. Box, G. E. P. and Meyer, R. D., An Analysis for Unreplicated Fractional Factorials, *Technometrics*, 28, 11–18 (1986).
4. Box, G. E. P. and Hunter, J. S., The 2^{k-p} Fractional Factorial Designs (Parts 1 and 2), *Technometrics*, 3, 311–351, 449–458 (1961).
5. Srivastava, J. N., Designs for Searching Nonnegligible Effects, in *A Survey of Statistical Design and Linear Models*, edited by J. N. Srivastava, North-Holland, Amsterdam, 1975, pp. 507–519.
6. Draper, N. R. and Lin, D. K. J., Capacity Consideration for Two-Level Fractional Factorial Designs, *J. Statist. Planning Inference*, 24, 25–35 (1990).
7. Lin, D. K. J., Systematic Supersaturated Designs, Working Papers No. 264, College of Business Administration, University of Tennessee (1991).
8. Lin, D. K. J., A New Class of Supersaturated Designs, *Technometrics*, 35, 28–31 (1993).
9. Hunter, G. B., Hodi, F. S., and Eager, T. W., High-Cycle Fatigue of Weld Repaired Cast Ti-6Al-4V, *Metall. Trans.*, 13A, 1589–1594 (1982).
10. Hamada, M. and Wu, C. F. J., Analysis of Designed Experiments with Complex Aliasing, *J. Qual. Technol.*, 24, 130–137 (1992).
11. Box, G. E. P. and Meyer, R. D., Finding the Active Factors in Fractionated Screening, *J. Qual. Technol.*, 25, 94–105 (1993).
12. Box, G. E. P. and Draper, N. R., *Empirical Model-Building and Response Surface*, John Wiley & Sons, New York, 1987.
13. Jan, H. W. and Wang, P. C., To Analyze Data from $L_{12}(2^{11})$ and $L_{18}(2 \times 3^7)$ Orthogonal Arrays, *J. Chin. Statist. Assoc.*, 32, 21–35 (1993) (in Chinese).
14. Lin, D. K. J., Generating Systematic Supersaturated Designs, *Technometrics*, 37, 213–225 (1995).
15. Williams, K. R., Designed Experiment, *Rubber Age*, 100, 65–71 (1968).
16. Westfall, P. H., Young, S. S., and Lin, D. K. J., Forward Selection Error Control in the Analysis of Supersaturated Designs, *Statist. Sinica*, 8, 101–117 (1998).
17. Deng, L. Y., Lin, D. K. J., and Wang, J. N., Supersaturated Designs Using Hadama Matrices, IBM Research Report No. 19470, IBM T. J. Watson Research Center (1994).

About the Author: Dennis K. J. Lin is a professor of management science and information systems and of statistics at The Pennsylvania State University. Dr. Lin received his Ph.D. in statistics from the University of Wisconsin–Madison in 1988. His current research interests are response surface methodology, quality engineering, computer experiments, and reliability. He has published more than 40 articles in such applied statistical journals as *Technometrics*, *Journal of Quality Technology*, *Applied Statistics (JRSSC)*, *Statistica Sinica*, and others. He is currently a member of the Editorial Review Board for *Journal of Quality Technology*, an associate editor for *The American Statistician*, and a managing editor for *Statistica Sinica*.