The choice of screening design

John Tyssedal and Muhammad Azam Chaudhry
Department of Mathematical Sciences,
Norwegian University of Science and Technology (NTNU),
7491 Trondheim, Norway
Email: tyssedal@stat.ntnu.no
azam@math.ntnu.no

Abstract

A screening design is an experimental plan used for identifying the expectedly few active factors from potentially many. In this paper, we compare the performances of three experimental plans, a Plackett-Burman design, a minimum run resolution IV design and a definitive screening design, all with 12 and 13 runs, when they are used for screening and three out of six factors are active. The functional relationship between the response and the factors was allowed to be of two types, a second order model and a model with all main effects and interactions included. D-efficiencies for the designs ability to estimate parameters in such models were computed, but it turned out that these are not very informative for comparing the screening performances of the two-level designs to the definitive screening design. The overall screening performance of the two-level designs was quite good, but there exist situations where the definitive screening design, allowing both screening and estimation of second order models in the same operation, has a reasonable high probability of being successful.

Key words: factor screening, definitive screening design, minimum run resolution IV design, Plackett-Burman design

1. Introduction

At the initial stage of an experimental investigation, there may be a large number of factors that potentially affect a response. In reality, though, it is most often found that only a small subset of these, typically two or three, maybe four are really important for describing the variation in the response values. The extraction of this subset of active factors is known as factor screening.

Two-level designs are normally the preferred choice for screening purposes. These can be described as belonging to one of two classes, regular or nonregular. Regular designs are the factorial 2^k designs and their regular fractions denoted as $2^{k\text{-}p}$ designs and are widely used in industrial experimentation. Their properties and usefulness are nicely explained in several textbooks like Box et al. [1] and Montgomery [2].

The nonregular designs are fractions of full 2^k designs that are not regular. Two-level Plackett-Burman (PB) designs [3], with the number of runs $n \neq 2^k$, are the most well-known two-level nonregular designs. As for the regular designs, contrast columns are orthogonal, and they can accommodate *n*-1 factors in *n* runs. Another class of nonregular two-level designs is the minimum run resolution IV (MinResIV) designs (Webb [4]) later discussed by Anderson and Whitcomb [5]. These designs can accommodate *k* factors in 2*k* runs, but factor columns are not in general orthogonal.

Recently, Jones and Nachtsheim [6] introduced the definitive screening designs (DSDs). For *k* factors, these designs also require 2*k* runs or preferably 2*k*+1. They are three-level designs and as such, they allow the estimation of quadratic effects which is clearly impossible using two-level designs.

With several options for screening a set of factors, it may be unclear which design to choose. For instance, if some of the factors have quadratic effects, how will these affect the screening procedure if a two-level design is chosen? On the other hand, suppose that no factor has a quadratic effect. Will there then be any dramatic loss in the screening efficiency if a DSD is chosen? The motivation for this paper is to give some answers to these questions.

This paper is organized as follows. In Section 2 we will present some designs belonging to two of the respective classes mentioned and discuss their properties. In Section 3 we will give the background and motivation for our simulation study. The overall screening performances of the designs are given in Section 4 followed by discussion and concluding remarks in Section 5.

2. Three designs and their properties

To find some answers to our objective, we have chosen three designs where each of them allows screening of six factors in 12 runs. The designs are 6 columns from a PB design, a MinResIV design and a DSD all with 12 runs. Hence we are comparing two nonregular two-level designs, one with orthogonal design columns and one with not, with a three-level design.

Nonregular orthogonal two-level designs have several advantages compared to the regular ones. One is their flexible run sizes, they apparently exist for all number of runs, $n = 4m$, $m \geq 3$. Second, they have far better projection properties than the regular ones. Box and Tyssedal [7] defined projectivity of two-level designs as follows: *A* $n \times k$ *design with* n *runs and* k *factors each at two levels is said to be of projectivity* P *if the design contains a complete* 2^P *factorial in every possible subset of* P *out of the* k *factors, possibly with some points replicated.* Projection properties thus concern the properties of a design when restricted to a subset of P factors which fit well into the intention of screening. While most orthogonal nonregular two-level designs are of projectivity $P = 3$ in n-1 factors, regular two-level designs can only accommodate $\frac{n}{2}$ factors in order to bear this property. A third advantage is that for most of the nonregular orthogonal two-level designs, effects are only partial aliased and hence can be estimated from the data. It must be noted, however, that in order to take advantage of the partial aliasing, the number of interactions should be small. Their alias pattern may be rather complex, and as a result, their analysis is considered difficult by many practitioners. Several methods for analyzing orthogonal nonregular two-level designs exist, (Hamada and Wu [8], Box and Meyer [9], Tyssedal and Samset [10], Chipman et al. [11], Tyssedal and Niemi [12] to mention a few). Still, ways to analyze these designs need more research, especially in cases when the number of active factors exceeds the projectivity of the designs (Tyssedal and Hussain [13]).

An example of an orthogonal nonregular two-level design, the twelve run PB (PB_{12}) design, is given in Table 1.

Table 1 about here

The PB_{12} design has two nonisomorphic projections onto six factors (Lin and Draper [14]). One of the designs, design PB_{12}(6.1), have no mirror image run (columns A to F in Table 1**)** while the other design, PB_{12}(6.2), has two mirror image runs (for instance run 7 and 10 in columns A to E and G). According to Wang and Wu [15], design PB_{12}(6.1) has higher efficiency than design PB_{12}(6.2), and is therefore normally preferred when 6 columns are to be selected from a PB_{12} design.

The MinResIV designs can accommodate *k* factors in 2*k* runs. Thereby they have very flexible run sizes but at the expense of the number of experimental factors allowed. Their runs consist of *k* mirror image pairs, and as a result, main effects and two-factor interactions are not aliased and thereby their name. This is one of the reasons for their attractiveness. However, not all main effect columns are orthogonal and they may thus be aliased with each other. The same applies to two-factor interaction columns. A six-factor MinResIV (MinResIV_{12}) design is given in Table 2. The design is the one given in Design Expert.

Table 2 about here

Both the $\text{PB}_{12}(6.1)$ and the MinResIV_{12} designs are projectivity *P*=3 designs. While for the PB_{12} design all projections onto three dimensions consist of a full 2^3 design + a half fraction, the MinResIV_{12} design has two types of projections onto three dimensions. For more on projection properties of MinResIV designs we refer to Hussain and Tyssedal [16].

As already mentioned the 12 run DSD $\left(\text{DSD}_{12}\right)$ is a three level design. Two-factor settings of zero is added to each column. With an additional center run added for all factors, the design given in Table 3, a DSD_{12}, projects onto a full 3^2 in every two dimensions. DSDs have the same flexibility in run sizes and the same restriction on the number of experimental factors as the MinResIV designs, although one extra run is recommended. Main effect columns are not aliased with other main effect columns, with two-factor interaction columns or quadratic effect columns. The two-factor interaction columns, however, are aliased with each other and so are the quadratic effect columns. Two-factor interaction columns and quadratic effect columns are also aliased.

Table 3 about here

3. Choice of models and motivational examples

If the number of experimental factors exceeds half the number of runs, it is natural to use a nonregular orthogonal two-level design for screening. If not, the choice may not be that obvious. Let us consider a screening situation with six experimental factors in 12 or 13 runs. Also assume that not more than three factors are active, a natural assumption considering the number of runs and factors.

For some comparison between these designs we will now consider two types of models. One model, type 1, which is a full second order model in three factors.

$$Y = \beta_0 + \sum_{i=1}^{3} \beta_i x_i + \sum_{i<j}^{3} \beta_{ij} x_i x_j + \sum_{i=1}^{3} \beta_{ii} x_i^2 + \varepsilon \tag{1}$$

and another model, type 2, consisting of main effects and interactions up to third order.

$$Y = \beta_0 + \sum_{i=1}^{3} \beta_i x_i + \sum_{i<j}^{3} \beta_{ij} x_i x_j + \beta_{123} x_1 x_2 x_3 + \varepsilon \tag{2}$$

β_0 is the intercept, β_i, β_{ij} and β_{ijk} represent half the main effects, two-factor interactions, and three-factor interactions respectively while β_{ii} represent the sizes of the quadratic terms. The ε is an error term assumed normally distributed with mean 0 and variance σ^2.

A measure of the efficiency by which a design is able to estimate the parameters in the model, is the D-efficiency of the design matrix, $\mathbf{X}$, defined as $\mathrm{D}_{eff} = \frac{\left|\mathbf{X}^{t}\mathbf{X}\right|^{\frac{1}{p}}}{n}$, where p is the number of parameters, intercept included, and n is the number of runs. Table 4 gives us the D_{eff} that can be obtained for estimating the parameters in various models with three active factors using a PB_{12} (6.1) design, a $\mathrm{MinResIV}_{12}$ design or a DSD_{12} without and with an extra center run.

Table 4 about here.

Obviously, without adding a center run the DSD_{12} is the only design that can estimate quadratic effects, but if three quadratic terms are present there will be nearly linear dependency among the effect columns, as shown by the low value for D_{eff}, and MATLAB for instance, will leave out the column for the last quadratic term. For models without quadratic terms, the PB_{12} (6.1) design has the highest D-efficiency, and we notice that by adding a center run, the D_{eff} is increased for models with quadratic terms and decreased for those without. The DSD_{12} is not a P=3 design, but we notice that it has the possibility to allow the estimation of three main effects and all their interactions if no quadratic term is in the model.

As a motivational example, we performed a simulation study where these three designs with and without a center run were tested out on how well they were able to identify the correct subset of three active factors from the following two models, one of type 1 and the other of type 2. These two models thus represent two phenomena to be investigated with the three designs.

$$Y_1 = 2 + x_1 + 0.5\,x_2 + 0.5\,x_3 + x_1\,x_2 + 1.5\,x_1\,x_3 + 1.5\,x_2\,x_3 + x_1^2 + 2x_2^2 + 3x_3^2 + \varepsilon \qquad (3)$$

$$Y_2 = 2 + x_1 + 0.5x_2 + 0.5x_3 + x_1x_2 + 1.5x_1x_3 + 1.5x_2x_3 + x_1x_2x_3 + \varepsilon \qquad (4)$$

The procedure was as follows

1. For all given design matrices, **X**, all projections onto 3 factors were found.

2. Starting with the model in (3), a vector of response values was generated with each of the 6 possible designs, 12 response values for designs without a center run and 13 response values for designs with a center run. The errors were generated as independent and identically normally distributed with mean 0 and variance σ^2 to be varied.

3. For each projection from the DSD_{12}, a model of type 1 in the respective three factors was fitted to its generated response and similar for the DSD_{12} with a center run $\left(\mathrm{DSD}_{12}\mathrm{cr}\right)$. For each projection from the PB_{12} (6.1) and $\mathrm{MinResIV}_{12}$ designs, a model of type 2 in the respective three factors was fitted to their respective responses and similar when a center run was added.

4. The mean square error, MSE = $\frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{n-p-1}$, was calculated for each projection where *p* is the number of terms and $\hat{y}_i$ is the fitted response.

5. The procedure was repeated 1000 times for each value of σ^2 varying from 0.1 to 1 in steps of 0.1, and the number of times the model with the correct subset of active factors had the smallest MSE for each design was recorded. We will call this the *success frequency* (SF).

6. Steps 2-5 were then repeated using model (4).

The results from our simulation study are given in Table 5. For each of the three designs, the screening is performed with and without a center run added.

<u>Table 5 about here</u>

In order to evaluate if there is a difference in performances, we look at the number of successes in 1000 simulations and use a binomial distribution. For various success probabilities, we then have the following standard deviations as given in Table 6.

Table 6 about here

Surprisingly, even though the chosen type 1 model, model (3), had rather small first order effects compared to quadratic effects, the two level designs without a center run outperformed the DSD_{12} for variances greater than 0.4. The PB_{12} (6.1) design performed clearly the best and seemed to benefit from the extra center run as the variance increased while the $\mathrm{MinResIV}_{12}$ did not. Also for the chosen model of type 2, model (4), the PB_{12} (6.1) design performed the best. There was little difference between designs with and without a center run, and the two DSDs were clearly inferior to the others.

Now, consider a model given by

$$\mathrm{E}(Y)=f(v_1,v_2,v_3)=\beta_0+\sum_{i=1}^{3}\beta_i\left(\frac{v_i-\bar{v}_i}{l_i}\right)+\sum_{i<j}^{3}\beta_{ij}\left(\frac{v_i-\bar{v}_i}{l_i}\right)\left(\frac{v_j-\bar{v}_j}{l_j}\right)+\sum_{i=1}^{3}\beta_{ii}\left(\frac{v_i-\bar{v}_i}{l_i}\right)^2$$

Each v_i, $i=1,2,3$ is assumed to take values on an interval $[\bar{v}_i-l_i,\bar{v}_i+l_i]$. Hence $\bar{v}_i$ is the midpoint of the interval and $2l_i$ is the length. Further let us assume that we are interested in investigating the function $f(v_1,v_2,v_3)$ and choose the three levels $\bar{v}_i-l_i$, $\bar{v}_i$ and $\bar{v}_i+l_i$ for each of the v_i, $i=1,2,3$. With the transformation $x_i=\frac{v_i-\bar{v}_i}{l_i}$, each x_i, $i=1,2,3$, takes the values -1, 0 and 1. Suppose instead that it was decided to investigate $f(v_1,v_2,v_3)$ in the region determined by $[\bar{v}_i-\frac{l_i}{2},\bar{v}_i+\frac{3l_i}{2}]$,

$i = 1,2,3$. The midpoints are now $z_i = \overline{v}_{\mathrm{i}} + \frac{l_{\mathrm{i}}}{2}$. Using the standard transformations $z_i = \frac{v_{\mathrm{i}} - \left(\overline{v}_{\mathrm{i}} + \frac{l_{\mathrm{i}}}{2} \right)}{l_{\mathrm{i}}}$, $i = 1,2,3$, gives $v_i = l_i z_i + \left(\overline{v}_i + \frac{l_i}{2} \right)$, and we get for $z_i \in [-1,1]$, $i = 1,2,3$, that

$$f(z_1, z_2, z_3) = \beta_0 + \sum_{i=1}^{3} \beta_{\mathrm{i}} \left(z_i + \frac{1}{2} \right) + \sum_{i<j}^{3} \beta_{ij} \left(z_i + \frac{1}{2} \right)\left(z_j + \frac{1}{2} \right) + \sum_{i=1}^{3} \beta_{ii} \left(z_i + \frac{1}{2} \right)^2$$

or:

$$f(z_1, z_2, z_3) = \beta_0 + \sum_{i=1}^{3} \left(\frac{\beta_{\mathrm{i}}}{2} + \frac{\beta_{ii}}{4} \right) + \sum_{\mathrm{i<j}}^{3} \frac{\beta_{ij}}{4} + \sum_{i=1}^{3} (\beta_{\mathrm{i}} + \beta_{ii}) z_i + \sum_{i<j}^{3} \frac{\beta_{ij}}{2} (z_i + z_j) + \sum_{i<j}^{3} \beta_{ij}\, z_i\, z_j + \sum_{i=1}^{3} \beta_{ii}\, z_{\mathrm{i}}^2$$

Similarly, we may arrive at a model of the form:

$$f(z_1, z_2, z_3) = \beta_0 + \sum_{i=1}^{3} \frac{\beta_{\mathrm{i}}}{2} + \sum_{i<j}^{3} \frac{\beta_{ij}}{4} + \frac{\beta_{123}}{8} + \sum_{i=1}^{3} \left(\beta_i + \frac{\beta_{123}}{4} \right) z_i + \sum_{i<j}^{3} \frac{\beta_{ij}}{2} (z_i + z_j) + \sum_{i<j}^{3} \left(\beta_{ij} + \frac{\beta_{123}}{2} \right) z_i\, z_j + \beta_{123}\, z_1\, z_2\, z_3$$

for a model of type 2.

Hence for a model of type 1 we observe that changing the experimental region, keeping the variation width of each factor constant, only affects the intercept and the main effects. For a model of type 2, both intercept, main effects and two-factor interactions are affected. If also the variation width is changed, all terms may be affected.

Now suppose we changed the region such that each variable is moved a quarter of the interval length to the right. Then, accordingly to the development above, in variables taking the values -1, 0 and 1, we get the following models for the responses Y_1 and Y_2.

$$Y_1^* = 5.5 + 3.25x_1 + 3.75\, x_2 + 5\, x_3 + x_1\, x_2 + 1.5\, x_1\, x_3 + 1.5\, x_2\, x_3 + x_1^2 + 2x_2^2 + 3x_3^2 + \varepsilon \quad (5)$$

$$Y_2^* = 4.125 + 2.5x_1 + 2x_2 + 2.25x_3 + 1.5x_1x_2 + 2x_1x_3 + 2x_2x_3 + x_1x_2x_3 + \varepsilon \quad (6)$$

Repeating our simulation study we get the results in Table 7.

<u>Table 7. about here</u>

The role is now changed. The DSD_{12} outperforms the two-level designs, even though it is the main effects and not the quadratic terms that have been increased in (5) and no quadratic effect is included in (6).

Now in such an investigation the results will depend on how well the correct model separates from the others. The cause of the bad performance with model 4 using DSD_{12}, is that without noise there are three other models with three active factors having a smaller MSE than the correct one. For model 5 and 6, it is the DSD_{12} that is superior in separating the correct model from the others. This example illustrates that the best screening design to be used in a given situation is not the one giving the largest D_{eff} of the projection onto the right factors, even a design with D_{eff} =0 for such a projection may work fine. Rather it is the chosen experimental region and the variation width of each factor that determines the success.

4. Test of overall performances of the designs

As we have seen, for models of type 1 and type 2 the regression coefficients, after the factors levels have been transformed to take the values -1, 0 and 1, will depend on our experimental region and the variation width of the factors and this will again affect the screening efficiency of the designs. To get a better impression of the overall performance of each of the designs, we performed a simulation study again with 1000 simulations, but for each simulation we now randomly draw each coefficient uniformly from an interval. The constant term was held fixed at 2. The interval [-1, 1] was chosen to represent small values, and the interval [-3, 3] to represent large values. For a model of type 1 the four test cases to be investigated are:

1. All coefficients are drawn uniformly from the interval [-1, 1].

2. Coefficients in front of linear and product terms are drawn uniformly from the interval [-3, 3] and coefficients in front of quadratic terms from the interval [-1, 1].
3. Coefficients in front of linear and product terms are drawn uniformly from the interval [-1, 1] and coefficients in front of quadratic terms from the interval [-3, 3].
4. All coefficients are drawn uniformly from the interval [-3, 3].

The results are summarized in Tables 8 to 11

Table 8 about here

Table 9 about here

Table 10 about here

Table 11 about here

First of all, we notice that adding a center run did not seem to affect the procedures much except for the case with small coefficients in front of linear and product terms and large coefficients in front of quadratic terms, where the performance of the two-level designs is clearly worse when a center run is added. Now since the coefficients in front of the terms are chosen in a low/high manner, one way to summarize the results is to define two factors F_1 and F_2 where F_1 represents the size of the coefficients in front of linear and product terms and F_2 represents the size of the terms in front of the quadratic terms. If we, as a response, take the amount of noise for which the SF exceeds 900 and only use designs without center runs, we get the results in Table 12.

Table 12 about here

From Table 12 it seems like it is the size of coefficients in front of the main effects and product terms that determines how efficient the screening is. The size of the terms in front of the quadratic terms has little effect. Also from the averages, the overall performance of the two-level designs is clearly better than for the DSD_{12} design and the $\mathrm{PB}_{12}(6.1)$ design comes out the best.

The same simulation study was also performed for the model of type 2. The four test cases were:

1. All coefficients are drawn uniformly from the interval [-1, 1].
2. Coefficients in front of linear terms are drawn uniformly from the interval [-3, 3] and coefficients in front of product terms from the interval [-1, 1].
3. Coefficients in front of linear terms are drawn uniformly from the interval [-1, 1] and coefficients in front of product terms from the interval [-3, 3].
4. All coefficients are drawn uniformly from the interval [-3, 3].

Tables 13-16 summarize the results.

<u>Table 13 about here</u>

<u>Table 14 about here</u>

<u>Table 15 about here</u>

<u>Table 16 about here</u>

The difference between the performance of the designs with and without center run is now almost negligible in all cases. The PB_{12} (6.1) design also now came out the best, but the differences between the performances of the two two-level designs are rather small. The performance of the DSD_{12} is now considerably worse than for the two-level designs, and the sizes of the terms obviously matters. It is beneficial with large terms in front of the linear terms while large terms in front of the product terms appear to have a negative effect on the SF for the DSD_{12}.

One of the referee suggested to test more situations, and mentioned especially cases with a certain mix of positive and negative coefficients like negative linear terms and positive quadratic terms for instance. We therefore performed another simulation study where the responses where generated from a model of type 1 with the following four test cases.

1. All coefficients are drawn uniformly from the interval [-3, 0].
2. Coefficients in front of linear and product terms are drawn uniformly from the interval [-3, 0] and coefficients in front of quadratic terms from the interval [0, 3].
3 Coefficients in front of linear and product terms are drawn uniformly from the interval [0, 3] and coefficients in front of quadratic terms from the interval [-3, 0]
4. All coefficients are drawn uniformly from the interval [0, 3].

The simulation study is closely related to the one reported in Table 11. The difference is that in Table 11 all coefficients were uniformly drawn from [-3, 3]. As expected, case 1 and case 4 gave almost identical SFs and so did case 2 and case 3. For the DSD_{12} we found only small differences between the four cases and all SFs were well within the uncertainty range when compared with the SFs in Table 11. The same conclusion also applies to the two level designs except that adding a center run had some small negative impact on the SFs for the PB_{12} (6.1) design in the cases 2 and 3.

We then changed case 2 and case 3 such that product terms had the same signs as the quadratic terms. Again the conclusions were exactly the same.

We also tried out how a mix of coefficients affected the SFs when the response values were generated from a model of type 2. The four tested cases were

1. All coefficients are drawn uniformly from the interval [-3, 0].

2. Coefficients in front of linear terms are drawn uniformly from the interval [-3, 0] and coefficients in front of product terms from the interval [0, 3].
3. Coefficients in front of linear terms are drawn uniformly from from the interval [0, 3] and coefficients in front of product terms from the interval [-3, 0].
4. All coefficients are drawn uniformly from the interval [0, 3].

Also in this situation cases 1 and 4 gave almost identical SFs and so did the two other cases. Except for a few cases the obtained SFs were well within the uncertainty ranges for the numbers in Table 16. Identical signs seem to have a small positive effect on the overall performance of the DSDs, while for the two-level designs opposite signs have a small positive effect.

4. Discussion and Concluding remarks

Although our investigation is limited, it illustrates several concerns to be aware of when comes to screening. Screening is about finding the subset of factors that really explains most of the variation in the data. A design's ability to extract out this subset of factors from possibly many should not be confused with its ability to estimate the possible effects of the factors in this subset. In our simulations, the two-level designs had an overall better screening performance than the DSD_{12} design even if the true response was a second order function for which their $D_{eff} = 0$ for estimating all the model coefficients. Therefore, in choosing a screening design one should carefully consider what the goal is. Is it just to extract the subset of active factors and then follow up with a closer investigation of the relationship between the response and the factors afterwards, or is the purpose to do both things at once? In the first case, the two-level designs seem to perform the best. However, if the response is a second order model, our simulations show that doing both things at once has a reasonable probability of success if the variance of the response values is low. If that is not the case, and especially for a type 2 model, one may easily end up with the wrong subset of active factors, and the existing possibility to estimate quadratic effects may be of little value. We also noticed that adding a center run did not improve the overall screening performance of the three types of design. In fact for a model with quadratic terms the SFs were lowered in some cases for the two-level designs. However, it improves the estimation efficiency of the DSD_{12} if the response has quadratic terms, and should be added for the DSD_{12} if both screening and estimation is considered simultaneously.

The PB_{12} (6.1) design came out best in our simulation study. It performed a little better than the $\mathrm{MinResIV}_{12}$ design and in the comparison between the two-level designs, the screening performances were reasonably consistent with the D_{eff} of the projections. It may be argued that in our way of analysis, we did not take into account some of the attractive properties of the MinResIV designs and the DSDs, especially that main effects can be estimated unbiased from two-factor interactions and for the DSDs also unbiased from quadratic effects. Also, it is well demonstrated, for instance Wolters and Bingham [17], that the "correct model" is not necessarily the one with the lowest MSE, but that it is normally among the ones with the lowest MSE. Hence our reported success frequencies may, and especially for the $\mathrm{MinResIV}_{12}$ and the DSD_{12} designs, be a little pessimistic. Nevertheless, whatever method of analysis that is used and set of assumed active factors that is obtained, if there exist a different subset of factors that produces a lower MSE, it will normally cause some ambiguity which often has to be solved by follow-up experiments.

Finally, as showed in Section 3, the choice of the best screening design in a given situations also depend on where in the experimental region the experiment is performed and what levels that are chosen for the experimental factors.

Acknowledgement

We are thankful to the AE and two anonymous referees for their helpful comments and suggestions that led to improvements to this article.

References

1) Box, G., Hunter, W., and Hunter, J. (1978, 2005). *Statistics for experimenters* (1st and 2nd edition). John Wiley and Sons, NewYork.

2) Montgomery, D. (2012). *Design and Analysis of experiment* (8th edition). John Wiley and Sons, New york.

3) Plackett, R., and Burman, J. (1946). The design of optimum multifactorial experiments. *Biometrika*, 33(4), pp. 305-325.

4) Webb, S. (1968). Non-orthogonal designs of even resolution. *Technometrics*, 10(2): pp. 291-299.

5) Anderson, M., and Whitcomb, P. (2004). Screening process factors in the presence of interactions. *AQC Toronto Proceedings*.

6) Jones, B., and Nachtsheim, C. (2011). A Class of Three-Level Designs for Definitive Screening in the Presence of Second-Order Effects. *Journal of Quality Technology,* 43, pp. 1-15.

7) Box, G., and Tyssedal, J. (1996). Projective properties of certain orthogonal arrays. *Biometrika,* 83(4), pp. 950-955.

8) Hamada, M., and Wu, C. F. (1992). Analysis of Designed Experiments with Complex Aliasing. *Journal of Quality Technology,* 24(3), pp. 130-137.

9) Box, G., and Meyer, R. (1993). Finding the active factors in the fractionated screening experiments. *Journal of Quality Technology,* 25(2), pp. 94-105.

10) Tyssedal, J., and Samset, O. (1997). Analysis of the 12 run Plackett and Burman design. Technical Report no 8, The Norwegian University of Science and Technology, Department of Mathematical Sciences.

11) Chipman, H., Hamada, M., and Wu, C. (1997). A bayesian-variable selection approach for analysing designed experiment with complex aliasing. *Technometrics* 39(4), pp. 372 - 381.

12) Tyssedal, J. S., and Niemi, R. (2014). Graphical Aids for the Analysis of Two-Level Nonregular Designs. *Journal of Computational and Graphical Statistics*, 23(3).

13) Tyssedal, J., and Hussain, S. (2016). Factor screening in nonregular two-level designs based on projection-based variable selection. *Journal of Applied Statistics*, 43(3), pp 490-508.

14) Lin, D., and Draper, N. (1992). Projection Properties of Plackett and Burman. *Technometrics,* 34(4), pp. 423-428.

15) Wang, J. C., and Wu, C. F. (1995). A Hidden Projection Property of Plackett-Burman and

Related Designs. *Statistica Sinica*, 5, pp. 235-250.

16) Hussain, S. and Tyssedal, J. (2016). Projection Properties of Blocked Non-regular Two-level Designs. *Quality and Reliability Engineering International*, 32 (8), pp 3011-3021.

17) Wolters, M.A. and Bingham, D (2011). Simulated Annealing Model Search for Subset Selection in Screening Experiments. *Technometrics* 53(3), 225-237.

Tables

Table 1: The twelve run PB (PB_{12}) design

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

Table 2: The $MinResIV_{12}$ design for 6 factors

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

Table 3: The 12 run definitive screening design (DSD_{12}) for six factors

Runs	A	B	C	D	E	F
1	0	1	-1	-1	-1	-1
2	0	-1	1	1	1	1
3	1	0	-1	1	1	-1
4	-1	0	1	-1	-1	1
5	-1	-1	0	1	-1	-1
6	1	1	0	-1	1	1
7	-1	1	1	0	1	-1
8	1	-1	-1	0	-1	1
9	1	-1	1	-1	0	-1
10	-1	1	-1	1	0	1
11	1	1	1	1	-1	0
12	-1	-1	-1	-1	1	0

Table 4: D-efficiencies of various models with 3 active factors using a PB_{12}(6.1)design, a $\mathrm{MinResIV}_{12}$ design and a DSD_{12} without and with (cr) an extra center run

	D-efficiencies					
Effects / Designs	PB_{12}(6.1)	PB_{12}(6.1) cr	$\mathrm{MinResIV}_{12}$	$\mathrm{MinResIV}_{12}$ cr	DSD_{12}	DSD_{12} cr
x_1, x_2, x_3	100	94.17	92.77-97.04	87.36-91.44	87.21	82.14
x_1, x_2, x_3, x_{12}	97.67	91.61	91.98-97.67	86.44-91.61	82.66	77.53
$x_1, x_2, x_3, x_{12}, x_{13}$	96.15	89.94	90.48-96.15	84.85-90.08	78.89	73.82
$x_1, x_2, x_3, x_{12}, x_{13}, x_{23}$	95.07	88.77	88.90-95.07	83.23-88.89	75.04	70.06
$x_1, x_2, x_3, x_{12}, x_{13}, x_{23}, x_{123}$	94.28	88.00	87.73-94.28	81.99-88.00	69.36	64.67
$x_1, x_2, x_3, x_{12}, x_{13}, x_{23}, x_1^2$	0	64.73	0	61.04-64.73	57.47	56.83
$x_1, x_2, x_3, x_{12}, x_{13}, x_{23}, x_1^2, x_2^2$	0	0	0	0	45.24	48.27
$x_1, x_2, x_3, x_{12}, x_{13}, x_{23}, x_1^2, x_2^2, x_3^2$	0	0	0	0	1.54	42.35
x_1, x_2, x_3, x_1^2	0	56.1567	0	52.88-54.84	60.39	60.46
$x_1, x_2, x_3, x_1^2, x_2^2$	0	0	0	0	46.95	49.14
$x_1, x_2, x_3, x_1^2, x_2^2, x_3^2$	0	0	0	0	38.86	42.29

Table 5: The SFs of the three types of designs without and with a center run (cr), varying the noise σ^2 from 0.1 to 1. Responses are generated with model (3) and model (4)

SFs of the designs when the responses are generated with model (3)										
σ^2	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
DSD_{12}	1000	1000	999	976	874	781	656	571	487	376
$\mathrm{DSD}_{12}\mathrm{cr}$	1000	1000	998	974	903	770	658	551	480	400
$\mathrm{PB}_{12}(6.1)$	1000	1000	1000	1000	992	976	944	871	832	754
$\mathrm{PB}_{12}(6.1)\mathrm{cr}$	1000	1000	1000	1000	998	991	958	918	885	822
$\mathrm{MinResIV}_{12}$	1000	1000	998	978	920	868	794	711	646	627
$\mathrm{MinResIV}_{12}\mathrm{cr}$	995	872	814	702	674	616	561	547	482	444
SFs of the designs when the responses are generated with model (4)										
σ^2	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
DSD_{12}	0	28	63	91	114	126	130	130	152	139
$\mathrm{DSD}_{12}\mathrm{cr}$	0	27	62	84	109	116	140	117	120	140
$\mathrm{PB}_{12}(6.1)$	1000	1000	997	983	935	852	735	634	564	513
$\mathrm{PB}_{12}(6.1)\mathrm{cr}$	1000	1000	999	983	931	833	745	644	572	495
$\mathrm{MinResIV}_{12}$	1000	999	976	883	728	664	498	447	383	326
$\mathrm{MinResIV}_{12}\mathrm{cr}$	1000	1000	973	850	743	618	509	476	413	356

Table 6: Standard deviations for various counted numbers

Probability	0.99	0.95	0.85	0.75	0.65	0.55	0.50
Standard deviation	3.1	6.9	11.3	13.7	15.1	15.7	15.8

Table 7: The SFs of the three types of designs without and with a center run (cr), varying the noise σ^2 from 0 to 2.4 in steps of 0.2 when the region is changed. Responses are generated with model (5) and model (6)

SFs of the designs when the responses are generated with model (5)													
σ^2	0	0.2	0.4	0.6	0.8	1	1.2	1.4	1.6	1.8	2	2.2	2.4
DSD_{12}	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	997	995	988
$DSD_{12}cr$	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	995	993
$PB_{12}(6.1)$	1000	1000	1000	1000	1000	1000	1000	996	990	974	965	923	888
$PB_{12}(6.1)cr$	1000	1000	1000	1000	1000	1000	999	999	994	988	954	927	886
$MinResIV_{12}$	1000	1000	1000	1000	1000	1000	996	991	965	927	886	850	790
$MinResIV_{12}cr$	1000	1000	1000	1000	1000	999	996	984	957	920	862	829	782

SFs of the designs when the responses are generated with model (6)													
σ^2	0	0.2	0.4	0.6	0.8	1	1.2	1.4	1.6	1.8	2	2.2	2.4
DSD_{12}	1000	1000	1000	1000	1000	1000	996	982	949	921	893	813	722
$DSD_{12}cr$	1000	1000	1000	1000	1000	999	998	980	950	917	858	800	735
$PB_{12}(6.1)$	1000	1000	997	952	804	664	521	462	350	313	291	269	264
$PB_{12}(6.1)cr$	1000	1000	999	948	830	670	544	432	403	345	287	263	279
$MinResIV_{12}$	1000	1000	1000	973	911	818	719	628	547	493	452	418	355
$MinResIV_{12}cr$	1000	1000	1000	976	916	812	691	645	550	491	419	385	340

Table 8: The SFs for each of the three types of designs without and with a center run (cr), varying the noise σ^2 from 0.1 to 1. All coefficients are drawn uniformly from the interval [-1, 1]

Overall performance of all six designs when responses are generated with a model of type 1. Case 1										
σ^2	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
DSD_{12}	870	771	677	566	501	465	378	330	278	256
$DSD_{12}cr$	876	771	669	577	504	446	356	330	293	266
$PB_{12}(6.1)$	997	984	927	854	757	639	528	446	406	336
$PB_{12}(6.1)cr$	991	973	930	838	731	662	520	475	375	318
$MinResIV_{12}$	991	947	888	791	720	627	519	496	364	313
$MinResIV_{12}cr$	981	928	862	801	684	614	524	445	379	313

Table 9: The SFs of the three types of designs without and with center run (cr), varying the noise σ^2 from 0.1 to 1. Coefficients in front of linear and product terms are drawn uniformly from the interval [-3, 3] and coefficients in front of quadratic terms from the interval [-1, 1]

Overall performance of all six designs when responses are generated with a model of type 1. Case 2										
σ^2	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
DSD_{12}	960	936	887	835	803	784	742	710	695	676
$DSD_{12}cr$	956	938	880	824	816	794	740	720	665	628
$PB_{12}(6.1)$	1000	1000	997	992	988	983	960	956	949	915
$PB_{12}(6.1)cr$	998	999	996	992	987	977	960	967	944	905
$MinResIV_{12}$	1000	998	993	975	961	953	942	903	900	864
$MinResIV_{12}cr$	994	994	992	970	962	953	928	912	894	848

Table 10: The SFs of the three types of designs without and with a center run (cr), varying the noise σ^2 from 0.1 to 1.0. Coefficients in front of linear and product terms are drawn uniformly from the interval [-1, 1] and coefficients in front of quadratic terms from the interval [-3, 3]

Overall performance of all six designs when responses are generated with a model of type 1. Case 3										
σ^2	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
DSD_{12}	866	781	667	573	522	430	400	336	300	242
$DSD_{12}cr$	875	789	654	585	504	432	401	328	307	260
$PB_{12}(6.1)$	995	977	931	843	736	645	530	472	395	302
$PB_{12}(6.1)cr$	923	885	837	740	641	573	495	409	327	285
$MinResIV_{12}$	987	949	897	798	724	616	535	448	399	363
$MinResIV_{12}cr$	889	840	792	707	631	563	465	426	352	349

Table 11: The SFs of the three types of designs without and with a center run (cr), varying the noise σ^2 from 0.1 to 1. All coefficients are drawn uniformly from the interval [-3,3]

Overall performance of all six designs when responses are generated with a model of type 1. Case 4										
σ^2	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
DSD_{12}	956	919	867	842	788	778	713	693	659	657
$DSD_{12}cr$	956	920	878	844	805	779	721	710	654	647
$PB_{12}(6.1)$	1000	999	998	996	996	978	974	950	930	896
$PB_{12}(6.1)cr$	994	994	997	986	984	977	962	936	922	901
$MinResIV_{12}$	998	993	987	970	955	951	934	903	890	876
$MinResIV_{12}cr$	988	979	967	957	956	937	909	871	845	843

Table 12: A 2^2 design showing the effect of four combinations of effect sizes. F_1 represents the size of the coefficients in front of linear and product terms and F_2 represents the size of the terms in front of the quadratic terms

F_1	F_2	DSD_{12}	$PB_{12}(6.1)$	$MinResIV_{12}$
-	-	0	0.3	0.2
+	-	0.2	1.0	0.9
-	+	0	0.3	0.2
+	+	0.2	0.9	0.7
		Ave.: 0.1	Ave.: 0.625	Ave.: 0.50

Table 13: The SFs of the three types of designs without and with a center run (cr), varying the noise σ^2 from 0.1 to 1.0. All coefficients are drawn uniformly from the interval [-1, 1]

Overall performance of all six designs when responses are generated with a model of type 2. Case 1										
σ^2	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
DSD_{12}	491	442	439	357	341	290	259	237	223	173
DSD_{12}cr	495	451	436	346	303	303	249	220	227	172
$PB_{12}(6.1)$	996	980	931	860	757	693	572	493	426	372
$PB_{12}(6.1)$cr	993	983	928	880	790	673	591	500	436	407
$MinResIV_{12}$	997	969	911	835	746	647	597	509	448	368
$MinResIV_{12}$cr	992	950	897	843	731	684	601	519	427	407

Table 14: The SFs of the three types of designs without and with a center run (cr), varying the noise σ^2 from 0.1 to 1.0. Coefficients in front of linear terms are drawn uniformly from [-3, 3] and coefficients in front of product terms from the interval [-1, 1]

Overall performance of all six designs when responses are generated with a model of type 2. Case 2										
σ^2	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
DSD_{12}	785	776	746	717	680	670	646	633	616	561
DSD_{12}cr	779	780	762	732	646	678	662	645	600	571
$PB_{12}(6.1)$	1000	995	976	960	910	871	822	788	730	642
$PB_{12}(6.1)$cr	999	995	982	955	927	884	832	776	724	654
$MinResIV_{12}$	996	985	974	938	900	851	833	781	753	672
$MinResIV_{12}$cr	997	987	976	935	901	856	812	787	739	673

Table 15: The SFs of the three types of designs without and with a center run (cr), varying the noise σ^2 from 0.1 to 1. Coefficients in front of linear terms are drawn uniformly from [-1, 1] and coefficients in front of product terms from the interval [-3, 3]

Overall performance of all six designs when responses are generated with a model of type 2. Case 3										
σ^2	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
DSD_{12}	185	182	133	144	119	97	93	83	90	86
DSD_{12}cr	183	176	143	140	125	97	85	79	78	84
$\text{PB}_{12}(6.1)$	1000	1000	993	994	989	970	970	947	917	886
$\text{PB}_{12}(6.1)\text{cr}$	1000	1000	995	993	990	975	971	947	905	900
MinResIV_{12}	999	993	991	974	970	944	921	902	853	827
$\text{MinResIV}_{12}\text{cr}$	1000	997	992	984	967	945	944	916	851	847

Table 16: The SFs of the three types of designs without and with a center run (cr), varying the noise σ^2 from 0.1 to 1.0. All coefficients are drawn uniformly from the interval [-3,3]

Overall performance of all six designs when responses are generated with a model of type 2. Case 4										
σ^2	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
DSD_{12}	542	528	486	492	471	436	404	400	396	404
DSD_{12}cr	536	522	483	489	466	471	421	420	401	390
$\text{PB}_{12}(6.1)$	1000	1000	998	996	986	976	967	943	920	913
$\text{PB}_{12}(6.1)\text{cr}$	1000	1000	997	996	985	974	965	943	939	922
MinResIV_{12}	999	998	990	983	960	946	926	908	911	878
$\text{MinResIV}_{12}\text{cr}$	999	995	990	982	969	958	952	910	906	887