Another Look at Dorian Shainin's Variable Search Technique

TIRTHANKAR DASGUPTA

Department of Statistics, Harvard University, Cambridge, MA 02138

NAGESH ADIGA and C. F. JEFF WU

School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332

An in-depth analysis of Dorian Shainin's variable search (VS) method is carried out. Conditions under which the VS approach does and does not work well in comparison with traditional designs are identified. Explicit expressions for the expected number of runs and the probability of correct screening are derived under stated assumptions. The crucial roles of process knowledge and noise variation in successful application of the VS design are established through theoretical and simulation studies.

Key Words: Design of Experiment; Factorial Design; Projectivity; Quality Improvement; Variable Screening.

Introduction

Dorian Shainin, a well-known, but controversial quality consultant, developed a quality-improvement program popularly known as the *Shainin system*. The Shainin system has been reported to be useful to many industries. Among several new tools and techniques proposed by Shainin (Steiner et al. (2008)), variable search (VS) is one that has received quite a bit of attention from researchers and practitioners. The VS technique (Shainin (1986), Shainin and Shainin (1988)) can be described as a sequential screening process used to identify the key factors and their settings to optimize the response of a system by making use of available experimenter's knowledge. Verma et al. (2004) used some real-life numerical examples to demonstrate the superiority of VS over more traditional methods.

It is worth mentioning that another Shainin method called component search (CS) has been found to be quite popular among engineers. CS (Shainin and Shainin (1988)) is typically used when units can be disassembled and reassembled without damage or change to any of the components or subassemblies, with the objective of comparing families of variation defined by the assembly operation and individual components. CS has a stage called component swapping, from which VS was derived. See Steiner et al. (2008) for a detailed description of CS.

Ledolter and Swersey (1997) critically examined the VS method and argued, via an example involving seven factors, that a fractional factorial design is generally a better alternative compared with the VS method. In this article, we carry out a more in-depth analysis of Shainin's VS procedure by (a) summarizing the key properties of the VS design and investigating its suitability for factor screening, (b) identifying the statistical inference procedures associated with each step of the VS design, (c) combining the test procedures at individual steps to obtain a general expression for the probability of correct identification of active factors, and (d) examining the sensitivity of the VS method against correctness of engineering assumptions and varying levels of noise.

The remainder of this paper is organized as follows. In the next section, we provide a stage-by-stage description of the VS methodology. We then

Dr. Dasgupta is an Assistant Professor in the Department of Statistics. His email address is dasgupta@stat.harvard.edu.

Mr. Adiga is a Ph.D. Candidate in the School of Industrial and Systems Engineering. His email address is nagesh@gatech.edu.

Dr. Wu is Professor of Engineering Statistics and Coca-Cola Chair in the School of Industrial and Systems Engineering. His email address is jeffwu@isye.gatech.edu.

TABLE 1. Example of Variable Search Design

Run	1	2	3	4	5	6	7	y	Confidence interval	Remark
1	+	+	+	+	+	+	+	448		
2	+	+	+	+	+	+	+	453		
3	+	+	+	+	+	+	+	451		
4	−	−	−	−	−	−	−	63		
5	−	−	−	−	−	−	−	66		
6	−	−	−	−	−	−	−	70		
	Swapping for factor 1									
7	−	+	+	+	+	+	+	350	(441.2, 460.8)	1 is active
8	+	−	−	−	−	−	−	104	(56.2, 75.8)	1 is active
	Swapping for factor 2									
9	+	−	+	+	+	+	+	324	(441.2, 460.8)	2 is active
10	−	+	−	−	−	−	−	249	(56.2, 75.8)	2 is active
	Capping runs (factors 1 and 2)									
11	+	+	−	−	−	−	−	392	(441.2, 460.8)	Capping run unsuccessful
12	−	−	+	+	+	+	+	106	(56.2, 75.8)	Capping run unsuccessful
	Swapping for Factor 3									
13	+	+	−	+	+	+	+	403	(441.2, 460.8)	3 is active
14	−	−	+	−	−	−	−	96	(56.2, 75.8)	3 is active
	Capping runs (factors 1, 2, and 3)									
15	+	+	+	−	−	−	−	443	(441.2, 460.8)	Capping run successful
16	−	−	−	+	+	+	+	60	(56.2, 75.8)	Capping run successful

discuss the salient properties of the VS design and the statistical-inference procedures associated with individual stages of VS and their impact on the overall process. We then turn to the sensitivity of the VS method against correctness of engineering assumptions and varying levels of noise. Finally, concluding remarks are presented.

An Overview of Shainin's Variable Search

Here we explain the VS method with a hypothetical example with seven factors (and data) presented in Table 1. The method consists of four stages, and assumes that the following are known: (i) the order of importance of the factors under investigation and (ii) the best and worst settings for each of the factors.

In **stage 1**, the suspect factors are ranked in descending order of perceived importance. Two levels are assigned to each factor: "best" (+) level and "worst" (−) level. Assume that the objective is to maximize the response, i.e., larger values of response are preferred. The algorithm starts with two experimental runs: one with all factors at their best levels and the other with all factors at their worst levels. These two runs are then replicated thrice in random order (runs 1 to 6 in Table 1). These responses are used to test if there is a statistically significant difference between these two settings. Median and range for the three replications for the two experiments are computed. Denote these medians by M_{b} and M_{w}, and the ranges by R_{b} and R_{w}, respectively, where the suffixes b and w indicate best and worst settings, respectively. We then compute $R_m = (M_{\mathrm{b}} - M_{\mathrm{w}})/R_{\mathrm{avg}}$, where $R_{\mathrm{avg}} = (R_{\mathrm{b}} + R_{\mathrm{w}})/2$ denotes the average range. Note that R_{avg}/d_2 is an unbiased estimator of the underlying normal error standard deviation σ, where $d_2 = 1.693$ for a sample size of 3. A value of R_{m} greater than 1.07 (or, alternatively 1.25 as recommended by Bhote and Bhote (2000)) suggests the presence of at least one active factor and prompts the experimenter to move on to the next stage of the algorithm. In the example given in Table 1, $R_{\text{avg}} = (5+7)/2 = 6$ and $M_{\text{b}} - M_{\text{w}} = 451\text{-}66 = 385$, so that $R_{\text{m}} = 64.17$.

In **stage 2**, Shainin specifies confidence intervals for the mean response corresponding to the "best" and "worst" situations based on t-distribution with four degrees of freedom as $M_{\text{b}} \pm 2.776 R_{\text{avg}}/d_2$ and $M_{\text{w}} \pm 2.776 R_{\text{avg}}/d_2$, respectively, where $d_2 = 1.693$ because the sample size is three. These confidence intervals are used in the later stages to determine significance of factors or groups of factors. In our example, the confidence intervals corresponding to the "best" and "worst" settings are computed as (441.2, 460.8) and (56.2, 75.8), respectively.

Stage 3, also called *swapping*, is used to identify active factors by switching levels of each factor one at a time. Ideally, swapping should start from the most important factor and end with the least important factor. Assume, without loss of generality, that factor 1 is most important. Swapping of factor 1 is performed in the following way. The first run is conducted with factor 1 at "best" level and all the other factors at their "worst" levels. One more run is conducted with all these factor levels reversed. Factor 1 is declared *inert* (or insignificant) if both the response values are within the confidence intervals derived in stage 2, and *active* (or significant) otherwise. Similarly, swapping is performed with each other factor in order of perceived importance until two active factors are found. Once two active factors have been identified, we move to the next stage.

Stage 4, called *capping*, is used to check whether there are still more active factors to be identified (apart from the already identified ones). Two runs are conducted to confirm this. In the first run, all the factors identified active are set at their "best" levels and all the other factors at their "worst" levels. In the second run, all these levels are reversed. If the two responses from these two trials lie within the confidence intervals computed at stage 2, it is concluded that all the active factors have been identified successfully. Otherwise, one needs to go back to the swapping stage and search for some more active factors.

Swapping and capping runs are successively conducted till a capping run is "successful", which means there are no more active factors to be identified. In the example in Table 1, swapping of factor 1 (runs 7–8) and factor 2 (runs 9–10) declare these two factors as active. The follow-up capping run with these two factors (runs 11–12) is unsuccessful, which leads to the conclusion that there are possibly more active factors. Swapping of factor 3 (runs 13–14) declares it as active. Finally, capping of factors (1, 2 ,3) is successful and leads to termination of the VS process.

Properties of Variable Search Design

Ledolter and Swersey (1997) discussed some properties of the VS design and compared it with a fractional factorial design using the second order model

$$y = \beta_0 + x_1\beta_1 + x_2\beta_2 + x_3\beta_3 + (x_1x_2)\beta_{12} + (x_1x_3)\beta_{13} + (x_2x_3)\beta_{23} + \epsilon, \qquad (1)$$

where $\epsilon \sim N(0, \sigma^2)$. Assuming that out of seven factors $(1, 2, \ldots, 7)$ under investigation, three (1, 2, 3) are active through model (1), they argued that a 2^{7-3} fractional factorial design of resolution IV (Wu and Hamada (2000), Ch. 4) with defining relations $5 = 123$, $6 = 124$, and $7 = 234$ is superior to the VS design in terms of estimation efficiency. Note that the best possible VS design in this case with the correct conclusion is the one shown in Table 1.

However, it can be seen that, in this case, the VS design may have some advantages over the 2^{7-3} design in the context of early identification of active effects, provided the experimenter's knowledge regarding the relative importance of factors and their "best" and "worst" levels is perfect. Assume that the error variance σ^2 is sufficiently small to ensure that the statistical tests of hypotheses are powerful enough to guarantee the detection of significant effects both by the VS design and the fractional factorial design. Then, usual analysis of data obtained from the 16-run 2^{7-3} design will declare main effects $1, 2, 3$ and two factor interactions (2 fi's) $12, 13$ and 23 as significant. However, the 2 fi's 12, 13 and 23 are aliased with other 2 fi's, i.e., $12 = 35 = 46$, $13 = 25$ and $23 = 15 = 47$. Whereas the 2 fi's 46 and 47 can be ruled out using the effect heredity principle (Wu and Hamada (2009), Ch 5), the 2 fi's 35, 25 and 15 cannot, as one of the parent factors of each is significant. Thus, one would need additional orthogonal runs (at least 4) to disentangle these effects and identify the *true* active factors.

This problem will not arise for the VS design because the swapping and capping runs permit simultaneous testing of main effects and interactions associated with a single factor or a group of factors (this will be discussed in detail in the next section), thereby enabling the experimenter to completely rule out the presence of main effects or interactions of groups of effects. Therefore, in order to unambiguously identify the active factors, the fractional factorial may actually need more runs than the best possible VS design. This relative advantage of the VS design will be more predominant in this example if the model (1) contains a three-factor interaction (3 fi) term 123. In this case, regular analysis of the fractional factorial will identify the main effects 1, 2, 3 and 5 as significant, which will create more confusion regarding construction of orthogonal follow-up runs. The VS design will, however, remain the same if the sequence of runs is correct.

Now consider the case where the underlying true model is still the same as (1), but the number of factors under investigation is 20. Under the same assumption as before, the VS design will still be able to reach the correct conclusions in 16 runs, while a 16-run fractional factorial (2^{20-16}) clearly cannot be constructed. One should keep in mind, though, that the assumption of a very accurate level of process knowledge that leads to conducting swapping runs for the 3 active factors (out of 20) first is a very strong one.

Thus, whereas the VS design has some properties (e.g., estimation efficiency) that make it inferior to a comparable fractional factorial design, it also has certain properties that may give it an edge in *specific* situations. Apart from such plausible technical advantages, it should be noted that the VS method has certain *practical advantages.* The sequential nature of the experiments in the VS design permits the experimenter to obtain partial information at each stage of the experiment, whereas in fractional factorial or orthogonal array designs, one needs to wait for the experiments to be completed. The simplicity of the VS technique is appealing to experimenters. However, the VS method has some *practical drawbacks* as well. It will not be efficient if applied to a new process or to a process with no prior information. Also, the frequent change of settings will be expensive if the levels of factors are hard to change. By contrast, restricted randomization can be used in fractional factorial designs. Note that one critical disadvantage that may overshadow the practical benefits accrued from the sequential nature of experiments in the VS design (as stated earlier) is the presence of block effects in the form of uncontrollable factors drifting over time. However, because the experiment has variable block sizes, incorporating block effects into the analysis of VS designs appears to be quite nontrivial. Therefore, in the paper, we assume no block effect and leave the analysis with block effect as future research.

In the following subsections, we study some properties of the VS design in terms of run size and estimation efficiency. The proofs of all the results stated in this section are given in the Appendix.

Run Size of the VS Design

If we assume that each stage of the VS design would result in correct conclusion (the probability of which will be explored in the next section), then the number of runs will depend on the ordering of the factors according to perceived importance. Clearly, the best possible scenario would be one where all the active factors are explored first and the worst possible scenario would occur when the last factor to be explored is an active factor. The result in (2) is useful to compute the smallest possible, largest possible, and expected run length of a VS design. Suppose that the VS design identifies m active factors out of k factors under investigation, where $1 \leq m \leq k$. Then it is easy to see (Appendix) that the total number of runs N of the VS design satisfies

$$\begin{cases} N = 2(k+3), & \text{if } m = 1, \\ 4(m+1) \leq N \leq 2(k+m)+4, & \text{if } m > 1. \end{cases} \tag{2}$$

For example, if the VS design identifies 3 out of 7 factors under investigation as active (as in the example given in Table 1), the minimum and maximum number of runs of the VS design will be 16 and 24, respectively. The following result is helpful to compute the expected run size for the VS design under the assumption of a random ordering of factors chosen for swapping runs.

Result 1. Suppose p $(2 \leq p \leq k)$ out of k factors being investigated are actually active. Assume that

(i) Each swapping and capping run will lead to a correct conclusion.

(ii) There is a complete lack of experimenter's knowledge regarding relative importance of factors, which means any permutation of $(1, 2, \ldots, k)$ is equally likely to occur while investigating the k factors one by one.

Then the total number of runs in the VS design is a discrete random variable N that has the following probability function: $\Pr\{N = 4(p+1) + 2j\} = p\binom{k-p}{j}/(p+j)\binom{k}{p+j}$ and its expectation is given by $E(N) = 4(p+1) + 2(k-p)p/(p+1)$.

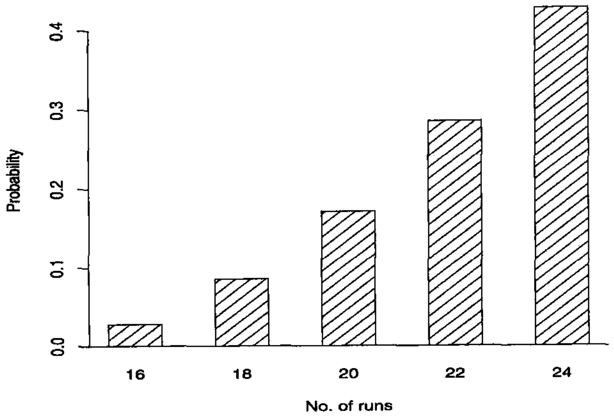


FIGURE 1. Probability Function of N for $k = 7$, $p = 3$.

Using Result 1, it is seen that, in the complete absence of knowledge about the relative importance of factors, the expected number of runs of a VS design in the example discussed earlier with $k = 7$ and $p = 3$ is 22. Its probability function is shown in Figure 1. A perfect knowledge of the relative importance will reduce the number of required runs to 16. An investigation with $k = 20$ factors will need, on average, about 42 runs if $p = 3$. The significant savings of runs that can be achieved by using VS design over comparable fractional factorial design is therefore quite evident.

Estimation of Main Effects from VS Design

We begin this discussion by noting that the VS design is neither a design with a fixed number of runs nor an orthogonal array. Assume that p out of k factors under investigation are active and the following model describes their relationship with the response y:

$$y = \beta_0 + \sum_{i=1}^{p} \beta_i x_i + \sum_{i<j} \sum_{j=1}^{p} \beta_{ij} x_i x_j + \ldots + \beta_{12\ldots p}\, x_1 x_2 \ldots x_p + \varepsilon, \tag{3}$$

where $x_i = -1$ or $+1$ according as the worst or best level of factor i is used, and $\epsilon \sim N(0, \sigma^2)$. Further, assume that the p active factors are correctly identified by the VS procedure.

Shainin recommended estimation of main effects of active factors from the $(2p + 2) \times p$ submatrix of the VS design that consists of two stage 1 runs and p pairs of swapping runs for the active factors.

TABLE 2. Model Matrix for Estimation from VS Design

Run	1	2	3	4	5	6	7	y
1	+	+	+	+	+	+	+	y_1
2	−	−	−	−	−	−	−	y_2
3	−	+	+	+	+	+	+	y_3
4	+	−	−	−	−	−	−	y_4
5	+	−	+	+	+	+	+	y_5
6	−	+	−	−	−	−	−	y_6
7	+	+	−	+	+	+	+	y_7
8	−	−	+	−	−	−	−	y_8
9	+	+	+	−	+	+	+	y_9
10	−	−	−	+	−	−	−	y_{10}
11	+	+	+	+	−	+	+	y_{11}
12	−	−	−	−	+	−	−	y_{12}
13	+	+	+	+	+	−	+	y_{13}
14	−	−	−	−	−	+	−	y_{14}
15	+	+	+	+	+	+	−	y_{15}
16	−	−	−	−	−	−	+	y_{16}

We shall denote this matrix by $\mathbf{X}_p$ in all subsequent discussions. Table 2 shows matrix $\mathbf{X}_7$. Shainin suggested estimating the main effect of factor i for $i = 1, 2, \ldots, p$ (defined as twice the regression coefficient β_i in model (3)) by comparing its swapping runs to the two corresponding stage 1 runs. For example, the regression coefficients β_1 and β_2 are estimated as

$$\hat{\beta}_1 = (y_1 - y_2 - y_3 + y_4)/4, \tag{4}$$
$$\hat{\beta}_2 = (y_1 - y_2 - y_5 + y_6)/4. \tag{5}$$

It is easy to see that the above estimators are unbiased, have a pairwise correlation of 0.5, and have the same standard error of 0.5σ. Ledolter and Swersey (1997) are particularly critical about this estimation procedure, owing to its large standard error (the standard error of $\hat{\beta}_i$ estimated from a 16-run factorial design would be 0.25σ) and correctly argue that a least-squares estimator would be a better choice, as it is statistically more efficient, although, still less efficient than a comparable fractional factorial design. For example, for $p = 7$, we find that the standard error of the least-squares estimator $\hat{\beta}_i$ is 0.33σ, and each pair of estimated effects $\hat{\beta}_i, \hat{\beta}_j$ has a correlation of approximately -0.14. We shall now see that the least-squares estimators obtained this way have some interesting properties that are summarized in result 2.

Result 2. Assume that $\mathbf{y}$, the $(2p+2) \times 1$ vector of responses, depends on the p active factors through model (3), and let $\boldsymbol{\beta}_{\text{main}} = (\beta_1, \ldots, \beta_p)'$.

(i) Then the least squares estimator of $\boldsymbol{\beta}_{\text{main}}$, given by

$$\hat{\boldsymbol{\beta}}_{\text{main}} = (\mathbf{X}_p'\mathbf{X}_p)^{-1}\mathbf{X}_p'\mathbf{y}, \qquad (6)$$

satisfies the following:

$$E(\hat{\boldsymbol{\beta}}_{\text{main}}) = \boldsymbol{\beta}_{\text{main}}$$

if all three and higher order interactions in model (3) are absent.

(ii) $$\sigma_p^2 = \frac{p^2 - 4p + 7}{8(p^2 - 3p + 4)}\sigma^2 \quad \text{for } i = 1, \ldots, p, \qquad (7)$$

$$\rho_p = -\frac{p-3}{p^2 - 4p + 7} \quad \text{for } i, j = 1, \ldots, p, \ i \neq j, \qquad (8)$$

where σ_p^2 and ρ_p denote the variance and pairwise correlation of estimated main effects when there are p active factors.

From Equations (7) and (8), we have that $\sigma_p^2 = 0.125\sigma^2$ for $p = 3$ and $\sigma_p^2 \to 0.125\sigma^2$ as $p \to \infty$. Also, $\rho_p = 0$ for $p = 3$ and $\rho_p \to 0$ as $p \to \infty$. Figure 2 shows plots of σ_p^2 (dotted curve in the left panel) and ρ_p (right panel) for $\sigma = 1$. From the above observations and Figure 2, the results in Equations (9)–(10) can easily be established. The variance σ_p^2 and the correlation coefficient ρ_p satisfy the following inequalities:

$$0.1071\sigma^2 \leq \sigma_p^2 \leq 0.1250\sigma^2, \qquad (9)$$

$$-0.167 \leq \rho_p \leq 0. \qquad (10)$$

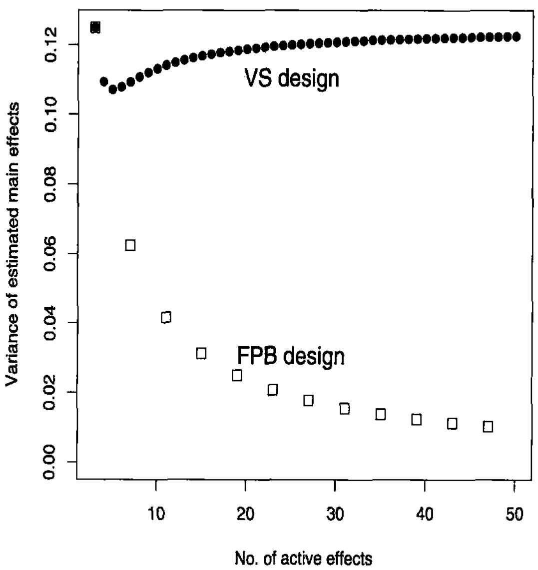
The lower and upper bounds in Equations (9) and (10) are attained for $p = 5$ and $p = 3$, respectively.

The findings from result 2 and Equations (9)–(10) can be summarized as follows:

1. Least-squares estimators of main effects of active factors obtained from stage 1 and swapping runs are *unbiased* and uncorrelated with estimators of 2 fi's under the assumption of negligible three-factor interactions.
2. The standard error of these estimators remains almost invariant (varying from $\sqrt{.1071}\sigma = 0.33\sigma$ to $\sqrt{.125}\sigma = 0.35\sigma$) with respect to the number of active factors.
3. The estimators are uncorrelated only if $p = 3$. For $p \geq 3$, they have a small negative correlation, which has the largest magnitude (-0.167) for $p = 5$.

Comparison of Efficiency with Folded-Over Plackett–Burman-Type Screening Designs

A large class of two-level orthogonal array designs with run size $N = 4n$ (where n is a positive integer) given by Plackett and Burman (1946) have been used as screening designs. Box and Hunter (1961) demonstrated that resolution IV designs can be obtained by folding over such designs. Miller and Sitter (2001) further studied these designs for $N = 12$ and proposed an analysis strategy.



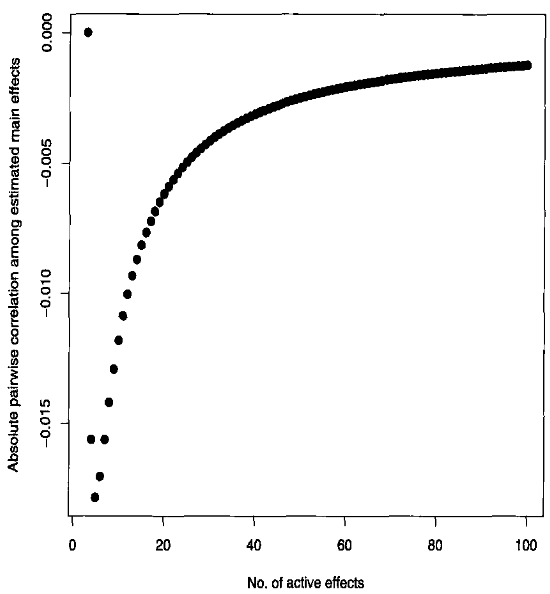


FIGURE 2. Plots of σ_p^2 (Left Panel) and ρ_p (Right Panel) for $\sigma = 1$.

Note that, when the number of factors p satisfies $p+1=4n$, a folded-over version of the Plackett–Burman (FPB) design is of order $(2p+2)\times p$ and is therefore comparable with the VS design matrix $\mathbf{X}_p$. The variance of $\hat{\beta}_i$ estimated from the FPB design with p factors is $\tilde{\sigma}_p^2=\sigma^2/(2p+2)$. Comparing this with Equation (7), the relative efficiency of the FPB design with respect to that of the VS design can be obtained as

$$e_p=\frac{\tilde{\sigma}_p^2}{\sigma_p^2}=\frac{1}{4}\left(p-\frac{p-7}{p^2-3p+4}\right). \qquad (11)$$

From Equation (11), we find that $e_p=1$ when $p=3$, i.e., both the designs are equally efficient with respect to estimation of main effects. However, for $p=7$, we have $e_p=7/4$, which means that the VS design is almost half as efficient as the FPB design. Further, $e_p\to\infty$ as $p\to\infty$. A comparison of variance of $\hat{\beta}_i$ obtained from the FPB and VS designs is shown in the left panel of Figure 2 for different values of p.

Although the above discussion clearly establishes the superiority of FPB designs over VS designs with respect to estimation efficiency, two important points should be kept in mind. First, the FPB design exists only when $p=4k-1$. For example, when $p=4,5$, or 6, an orthogonal design comparable with the VS design does not exist. Second, and more important, the FPB design would usually include all of the $k(>p)$ factors under investigation, whereas the VS design matrix $\mathbf{X}_p$ corresponds only to the p factors that are screened out as active. Thus, the above efficiency comparison may not always be meaningful. Further, when the number of factors under investigation is large, the VS method will have a clear-cut advantage in terms of run size as described earlier.

Estimation of Interaction Effects

So far we have discussed only the estimation of main effects. In the VS design, the 2 fi's are aliased with one another, and as observed by Ledolter and Swersey (1997), the swapping runs for factor i permit estimation of the sum of the 2 fi's associated with that factor (more discussion on this in the next section). We now discuss some properties of the VS design, summarized in the following two results (3 and 4), that helps us to devise a strategy to obtain unconfounded estimates of 2 fi's (and higher order interaction effects, if they exist) when the number of active factors does not exceed 4. It is well known that good screening designs should have projection properties, for which Box and Tyssedal (1996) gave the following definition.

Definition. An $N\times k$ design $\mathbf{D}$ with N runs and k factors each at 2 levels is said to be of projectivity P if every subset of P factors out of the possible k contains a complete 2^P full factorial design, possibly with some points replicated. The resulting design will then be called a (N,k,P) screen.

The above definition refers to designs that are orthogonal arrays. However, extending the definition of (N,k,P) screen to all $N\times k$ designs that are not necessarily orthogonal arrays, the following results can easily be established for a VS design.

Result 3. Consider an $N\times m$ submatrix $\mathbf{D}$ of a VS design matrix that consists of the columns corresponding to m (≥ 3) factors identified as active. Then

(i) submatrix $\mathbf{D}$ is a $(N,k,3)$ screen, i.e., has projectivity 3.

Further, when $m=4$ with four active factors A,B,C, and D identified by VS,

(ii) submatrix $\mathbf{D}$ contains a 2^{4-1}_{IV} fractional factorial design in these four factors with the defining relation $I=-ABCD$.

(iii) The above fractional factorial design is the *largest* orthogonal array that is contained in the submatrix $\mathbf{D}$.

(iv) It is possible to construct a 2^4 design in A,B,C, and D by addition of just four more runs, $(+--+-)$, $(+--+)$, $(-++-)$, and $(-+-+)$, to the VS design.

From result 3, it follows that, for $p=3$, the VS design permits unbiased and independent estimation of *all* factorial effects from the 2^3 design that it contains. Note that the design matrix is precisely the one discussed earlier in the context of least-squares estimation of main effects, which consists of two stage 1 and the swapping runs. Any regression coefficient in model (3) estimated in this way from the VS design will have a standard error of $\sigma/\sqrt{8}=0.35\sigma$. When $p=4$, the VS design still permits independent estimation of the four main effects and three aliased sets of 2 fi's from the 8-run fractional factorial design identified in result 3(ii) with a standard error of 0.35σ. The standard error of each main-effect estimate can be slightly reduced to 0.33σ by using least-squares estimation with the upper left 10×4 submatrix of the design matrix in Table 2; however, this will result in correlated estimates. A better strategy may be to conduct four additional runs as suggested in result 3(iv) and estimate each factorial effect of every possible order with a standard error of 0.125σ.

Statistical Inference at Different Stages of VS

The objective of this section is to understand the mechanism of hypothesis testing associated with the VS design, compute the probabilities of obtaining correct conclusions at different stages of the VS design (Equations (14), (17), and (22)), and eventually utilize these results to compute the probability of correct screening of active factors if the order of investigation of factors is fixed. Readers who wish to skip the technical details may skip the derivations in the first subsection and move on to the next subsection, which is of more practical importance.

Assume that VS is being performed to identify the active factors from a pool of k potential factors $x_1, \ldots, x_k$. The best and worst levels of each factor are known and represented by $+1$ and -1, respectively. The objective is to maximize the response y, which is related to the experimental factors through the following second-order model:

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i<j} \sum_{j=1}^{k} \beta_{ij} x_i x_j + \varepsilon, \qquad (12)$$

where $x_i = -1$ or $+1$ according as the worst or best level of factor i is used, and $\epsilon \sim N(0, \sigma^2)$. Note that correct knowledge of the best and worst levels of each factor implies that $\beta_i > 0$ for $i = 1, \ldots, k$ in model (12). Next, we introduce the following notation:

- y_i^+ (y_i^-): Observed response when factor x_i is at '+' ('−') level and the remaining $k-1$ factors are at '−' ('+') level.
- Let $\mathcal{F} \subset \{1, 2, \ldots, k\}$ be a set with cardinality q. Define $y_{\mathcal{F}}^+$ ($y_{\mathcal{F}}^-$) as the observed response when all the q factors $x_i, i \in \mathcal{F}$ are at '+' ('−') level and the remaining $k-q$ factors are at '−' ('+') level.
- y^+ (y^-): Observed response when all the k factors are at '+' ('−') level.

In the following subsections, we describe the statistical tests of hypothesis associated with different stages of the VS design. To keep our computations tractable, we shall (i) use the large-sample approximation for sample median, although the sample size in the VS design is only 3, and (ii) use sample standard deviation instead of sample range to estimate σ.

Power of Statistical Hypotheses Tested at Different Stages of VS and Probability of Correct Screening

Stage 1

As described before, the median M_{b} of three realizations of y^+ are compared with the median M_{w} of three realizations of y^-.

It is easy to verify (see Appendix for a detailed argument) the following:

- As observed by Ledolter and Swersey (1997), stage 1 of VS with respect to model (12) is equivalent to testing the hypothesis H_0 : $\sum_{i=1}^{k} \beta_i = 0$ against H_1 : $\beta_i \neq 0$ for at least one $i = 1, 2, \ldots, k$. An appropriate rejection rule will be the following: reject H_0 at level α if
$$\frac{|M_{\mathrm{b}} - M_{\mathrm{w}}|}{\hat{\sigma}\sqrt{\pi/3}} > t_{4,\alpha/2}. \qquad (13)$$
- The power (i.e., ability to detect presence of at least one active factor) of the test is given by
$$P_I = 1 - F_\delta(t_{4,\alpha/2}) + F_\delta(-t_{4,\alpha/2}), \qquad (14)$$
where $F_\delta(\cdot)$ denotes the cumulative distribution function of a noncentral t distribution with 4 degrees of freedom and noncentrality parameter $\delta = (2\sum_{i=1}^{k} \beta_i)/(\sigma\sqrt{\pi/3})$.

Clearly, P_I is a monotonically increasing function of $(\sum_{i=1}^{k} \beta_i)/\sigma$ through the parameter δ. Figure 3 shows a plot of the power function. It is seen that the power is approximately 95% if $(\sum_{i=1}^{k} \beta_i)/\sigma \approx 2.52$

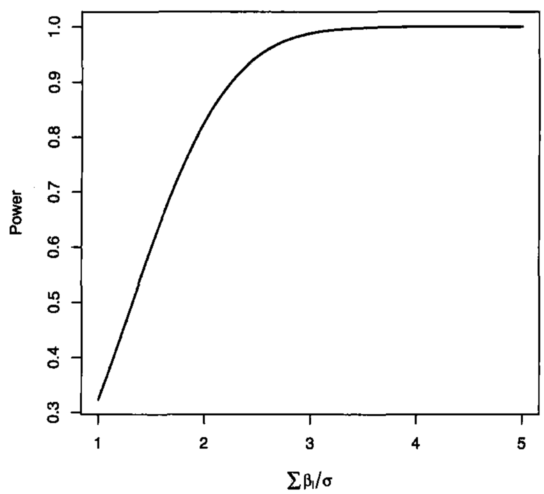


FIGURE 3. Power of Stage 1 Test as a Function of $\sum \beta_i/\sigma$.

and 99% if $(\sum_{i=1}^{k} \beta_i)/\sigma \approx 3.07$. In other words, the power of the stage 1 test is 95% and 99% if the sum of main effects $\sum_{i=1}^{k} \beta_i$ is, respectively, equal to 2.5 times and 3 times the error standard deviation σ.

Swapping

The objective of the swapping stage it to identify active factors. Here, with reference to the experiments in stage 1, the level of one particular factor is switched, and the difference in the response is observed. If this change is significant, then that factor is declared as active. Details of the hypothesis test associated with the swapping of factor x_i are described in the Appendix. The power of the test can be obtained as

$$P^{i}_{\text{swap}} = 1 - \{F_{\delta^+}(t_{4,\alpha/2}) - F_{\delta^+}(-t_{4,\alpha/2})\} \times \{F_{\delta^-}(t_{4,\alpha/2}) - F_{\delta^-}(-t_{4,\alpha/2})\}, \quad (15)$$

where $F_{\delta^+}(\cdot)$ and $F_{\delta^-}(\cdot)$ denote the cumulative distribution function of a noncentral t distribution with 4 degrees of freedom and noncentrality parameters $\delta^+ = (2\beta_i + 2\sum_{j\neq i}\beta_{ij})/(1.23\sigma)$ and $\delta^- = (2\beta_i - 2\sum_{j\neq i}\beta_{ij})/(1.23\sigma)$, respectively.

Thus, the power of the swapping phase is a function of β_i/σ and $\sum_{j\neq i}^{k}\beta_{ij}/\sigma$. Figure 4 shows a contour plot of the power against β_i/σ and $\sum_{j\neq i}^{k}\beta_{ij}/\sigma$. The darker regions represent low powers. As expected, the power of the test is small when both β_i/σ and $|\sum_{j\neq i}^{k}\beta_{ij}|/\sigma$ are small.

Remarks

1. The type 1 error of the test described above will be $1-(1-\alpha)^2$. To ensure that the test has a prespecified type 1 error α, the levels of significance associated with the events A^+ and A^- should be adjusted.
2. Note that each pair of swapping runs for an active factor x_i leads to estimation of the main effect β_i and sum of all interactions $\sum_{j\neq i}\beta_{ij}$ that involve x_i. Define $u_i = (s^+(x_i)+s^-(x_i))/4$ and $v_i = (s^+(x_i) - s^-(x_i))/4$. Then unbiased estimators of β_i and $\sum_{j\neq i}\beta_{ij}$ are given by
$$\hat{\beta}_i = u_i, \quad (16)$$
$$\sum_{j\neq i}\hat{\beta}_{ij} = v_i, \quad (17)$$
for $i = 1,\ldots,p$. The variance of $\hat{\beta}_i$ is given by $(1/8)(1+\pi/6)\sigma^2$ or $0.19\sigma^2$. Note that this estimation procedure is the same as the one recommended by Shainin and discussed earlier, except for the fact that six stage 1 runs are used instead of two. This reduces the variance of the estimates, but inflates the correlation between $\hat{\beta}_i$ and $\hat{\beta}_j$.
3. Thus, in a nutshell, swapping allows combined testing of all effects involving a single factor and hence is a useful tool to detect whether a single factor is "active" by conducting only two additional runs.

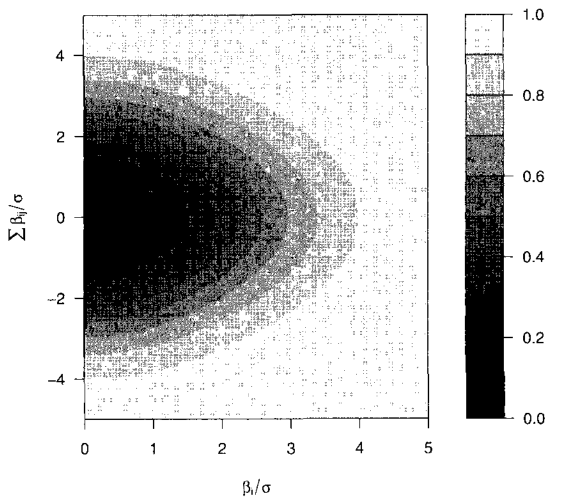


FIGURE 4. Power of Swapping to Detect Active Factor x_i as a Function of β_i/σ and $\sum_{j\neq i}\beta_{ij}/\sigma$.

Capping

The objective of the capping stage is to confirm whether all the active factors have been identified. Assume that q out of the k factors have been declared active in the swapping stage. Let $\mathcal{F}$ represent the set of indices of factors that have been declared active. Details of the hypothesis test associated with the capping of the factors in $\mathcal{F}$ is described in the Appendix. The power of the test can be obtained as

$$P^{\mathcal{F}}_{\text{cap}} = 1 - \{F_{\delta^+}(t_{4,\alpha/2}) - F_{\delta^+}(-t_{4,\alpha/2})\} \times \{F_{\delta^-}(t_{4,\alpha/2}) - F_{\delta^-}(-t_{4,\alpha/2})\}, \quad (18)$$

where $F_{\delta^+}(\cdot)$ and $F_{\delta^-}(\cdot)$ denote the cumulative distribution function of a noncentral t-distribution with 4 degrees of freedom and noncentrality parameters $\delta^+ = (2\sum_{i\notin\mathcal{F}}\beta_i + 2\sum_{i\notin\mathcal{F}}\sum_{j\in\mathcal{F}}\beta_{ij})/(1.23\sigma)$ and $\delta^- = (2\sum_{i\notin\mathcal{F}}\beta_i - 2\sum_{i\notin\mathcal{F}}\sum_{j\in\mathcal{F}}\beta_{ij})/(1.23\sigma)$, respectively. Similar to the swapping phase, the power of the capping test is small when both $\sum_{i\notin\mathcal{F}}\beta_i$ and $\sum_{i\notin\mathcal{F}}\sum_{j\in\mathcal{F}}\beta_{ij}$ are small.

Overall Probability of Correct Screening

We shall now use the power functions developed in the previous three subsections to obtain the probability of correct screening of p ($< k$) active factors under model (12) for a specific order in which the k factors are investigated, assuming that the best and worst levels of each factor are correctly identified. Without loss of generality, let $x_1, \dots, x_p$ represent the p active factors and $x_{p+1}, \dots, x_k$ the $k-p$ inactive factors. Consider the following two extreme situations: (i) the p active factors are investigated first (assume that VS investigates k factors in the sequence $x_1, \dots, x_p, x_{p+1}, \dots, x_k$) and (ii) the $k-p$ inert factors are investigated first (assume that VS investigates k factors in the sequence $x_k, \dots, x_{p+1}, x_p, \dots, x_1$). Then the probability of correctly identifying $\{x_1, \dots, x_p\}$ as the *only* active factors is given by

$$P_{CI} = \begin{cases} P_I(1-\alpha)(\prod_{i=1}^{p} P^i_{\text{swap}}) \\ \times(\prod_{i=2}^{p-1} P^{\mathcal{F}_i}_{\text{cap}}) \\ \times(1 - P^{\mathcal{F}_p}_{\text{cap}}), & \text{for situation (i)} \\ \\ P_I(1-\alpha)^{k-p} \\ \times(\prod_{i=1}^{p} P^i_{\text{swap}}) \\ \times(\prod_{i=2}^{p-1} P^{\mathcal{G}_i}_{\text{cap}}) \\ \times(1 - P^{\mathcal{G}_p}_{\text{cap}}), & \text{for situation (ii),} \end{cases} \tag{19}$$

where $\mathcal{F}_i = \{1, 2, \dots, i\}$, $\mathcal{G}_i = \{p, p-1, \dots, p-i+1\}$; and $P_I, P^i_{\text{swap}}, P^{\mathcal{F}}_{\text{cap}}$ are defined by Equations (14), (15), and (18), respectively.

Note that this probability will depend on the order in which the active factors are investigated if their impacts on the response are different. Consider an example where seven factors $\{x_1, \dots x_7\}$ are being investigated to identify three active factors $\{x_1, x_2, x_3\}$ that affect the response through the following model:

$$y = 0.8x_1 + 0.7x_2 + 0.8x_3 + 0.4x_1x_2 + 0.3x_2x_3 + 0.4x_1x_3 + \epsilon, \tag{20}$$

where $x_i = -1$ or $+1$ according as the worst or best level of factor i is used and $\epsilon \sim N(0, \sigma^2)$. Note that we do not include an intercept term in this model because the tests described in the previous subsection are independent of the intercept term β_0 in model (12). The powers associated with each individual hypothesis test at different stages are computed using Equations (14), (15), and (18) with $\sigma = 0.20$. Assuming a 5% level of significance for each test, the probability of detecting x_1, x_2, and x_3 as active and $x_4, \dots, x_7$ as inert can be computed for situations (i) and (ii) using Equation (19) as 0.8737 and 0.5818, respectively.

Inference in the Presence of Three-Factor Interactions

In the previous subsections, we have analyzed the VS technique where three-factor interaction effects were ignored. We now extend this to situations with three-factor interaction effects. Consider the following third-order model:

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i<j} \sum_{j=1}^{k} \beta_{ij} x_i x_j + \sum_{i<j,l} \sum_{j<l} \sum_{l=1}^{k} \beta_{ijl} x_i x_j x_l + \varepsilon. \tag{21}$$

Due to the presence of three-factor interactions, the test statistic for testing hypotheses at different stages of the VS approach have slightly different sampling distributions, as described in the Appendix. The test procedures, however, are identical to those derived earlier. It is worth noting that, under model (21), negative values of three-factor interactions (3 fi's) are likely to reduce the power of the tests at different stages. However, positive values of 3 fi's associated with a particular effect will increase the power of its detection as an active factor.

Robustness of VS with Respect to Noise Variation and Accuracy of Experimenter's Knowledge

In the earlier sections, we have seen that, for a given model, the performance of the VS design (with respect to the probability of correct screening) depends on the level of inherent noise variation σ^2 and the degree of correctness of the experimenter's knowledge. We shall now examine the effects of these variables (here we refrain from using the term "factors" to avoid mix-up with the original experimental factors to be screened by VS) and their interactions on the overall probability of correct screening using the results derived in the previous section. In this study, we consider seven factors $1, 2, \dots, 7$ to be investigated, out of which three factors $1, 2, 3$ are actually active. The response y depends on these three factors through the second-order model given in Equation (20). We consider the following three input variables that are known to affect the performance of VS:

1. A: Incorrect engineering assumption about setting of a particular factor (in this case we con-

TABLE 3. Design Matrix and Percentage of Correct Screening

Run	Setting of factor 3 (A)	Order of investigation (B)	σ (C)	100 $\times P_{CI}$
1	Correct	1-2-3-4-5-6-7	0.10	90.25
2	Correct	1-2-3-4-5-6-7	0.30	49.14
3	Correct	7-6-5-4-3-2-1	0.10	59.87
4	Correct	7-6-5-4-3-2-1	0.30	33.96
5	Incorrect	1-2-3-4-5-6-7	0.10	90.23
6	Incorrect	1-2-3-4-5-6-7	0.30	20.24
7	Incorrect	7-6-5-4-3-2-1	0.10	59.86
8	Incorrect	7-6-5-4-3-2-1	0.30	9.63

sider factor 3 without loss of generality). This variable has two levels: the + (−) level corresponds to a correct (incorrect) assumption, which means that the coefficients of x_3, x_{23} and x_{13} in model (20) are 0.8 (-0.8), 0.3(-0.3), and 0.4(-0.4), respectively.

2. B: Incorrect engineering assumption about relative importance of factors 1–7. Again we consider two levels of this variable: the − level corresponds to the correct order 1-2-3-4-5-6-7 and + level corresponds to the completely reverse (and incorrect) order 7-6-5-4-3-2-1.
3. C: The standard deviation σ of the error term ϵ in model (20). The two levels of this variable are chosen as $\sigma = 0.10$ and $\sigma = 0.30$.

A full factorial experiment was designed with these three input variables, and for each combination, the percentage of correct screening ($100 \times P_{CI}$) was computed using Equation (19). The results are summarized in Table 3.

The data from Table 3 are summarized in the form of significant main effects and interaction plots (see Figure 5). All the three variables A, B, and C are seen to affect the performance. As expected, the performance is poor with incorrect settings, wrong ordering, or high error variance. In particular, a three-times increase in error variance is seen to have a strong effect on the performance. Two interactions, $B \times C$ and $A \times C$, are also seen to affect P_{CI}. Clearly, higher noise level worsens the effect of lack of experimenter's knowledge on the performance. Also, the fact that a combination of incorrect setting, incorrect ordering, and high noise (run 8 in Table 3) leads to a very low (9.63) percentage of correct screening is indicative of the presence of a three-factor interaction $A \times B \times C$.

Concluding Remarks

We have investigated Shainin's variable search (VS) method with the objective of understanding it better and also identifying the type of settings under which it does and doesn't work well. The results in the previous three sections have established that VS is a useful method for screening of factors if (a) the engineering assumptions about the directions of the factor effects on the response and the relative order of importance are correct and (b) the error variance is not very high relative to the main effects and sum of 2 fi's involving each factor. The VS design permits unbiased estimation of the main effects under the assumption that interactions of order 3 and above are negligible. Further, it has some projection properties that permit independent estimation of main effects and 2 fi's for a maximum of four active factors.

Thus, the VS method is likely to be particularly useful for screening active factors if the number of factors under investigation is large, e.g., $k \geq 15$, where it can lead to a significant saving of runs in comparison with a comparable fractional factorial design, particularly if higher order interactions are actually present in the model or cannot be ruled out. In contrast, if the number of factors is not very large, e.g., $k \leq 10$, the experimenter's knowledge about the relative importance of factors is limited, and higher order interactions can be assumed away, fractional factorial designs or screening designs like Plackett–Burman designs will be a much better choice. Further, incorrect process knowledge and high error variance can result in poor performance of VS, both in terms of correctness of factor screening and run size.

Acknowledgments

The authors are thankful to the two referees and the editor for several constructive suggestions that helped in improving the content and presentation of this paper. The research of Adiga and Wu was supported by ARO grant W99INF-08-1-0368 and NSF grant DMS-0705261. The research of Dasgupta was supported by NSF grant CMMI-1000720.

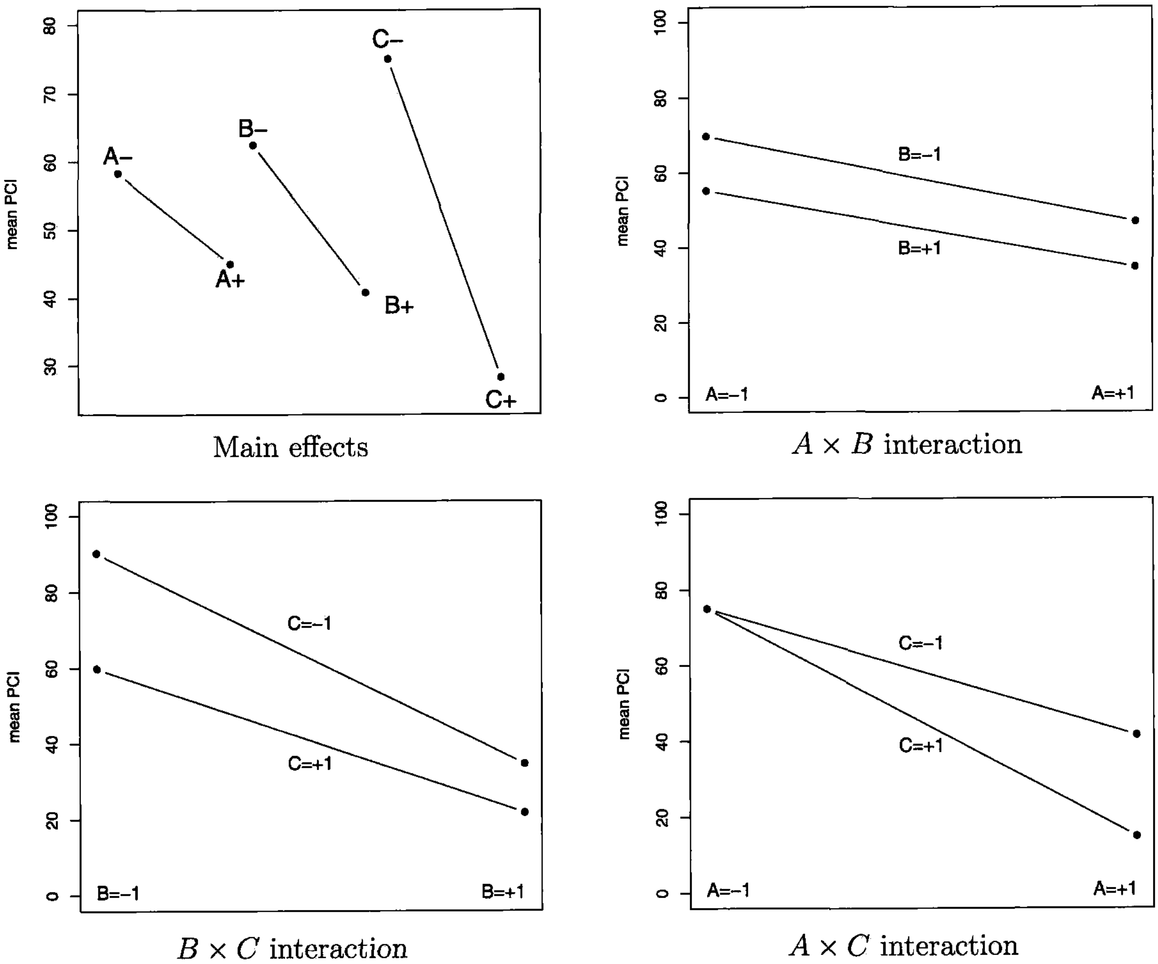


FIGURE 5. Plots of Factorial Effects from the Designed Experiment.

Appendix

Proof of Equation (2)

If VS identifies only one factor as active, it does not involve any capping run (a minimum of two active factors is necessary to conduct a capping run). However, each of the k factors undergoes swapping, which results in a total of $2k$ runs. Adding the 6 stage 1 runs to this, we have $N = 2(k+3)$. For $m > 2$, the total number of capping runs will be $2(m-1)$, and the total number of swapping terms will range from $2m$ to $2k$ depending on the order of swapping. Again, the result follows by adding the six stage 1 runs to the number of swapping and capping runs.

Proof of Result 1

From assumption (i), it follows that the total number of active factors identified by VS will be exactly p. By Equation (2), N can take integer values ranging from $4(p+1)$ to $2(k+p)+4$. Clearly, N varies *only* due to the number of extra swapping runs, which equals $2j$ if j inactive factors are examined, $j = 0, 1, \ldots, (k-p)$. Therefore, N can take only even integer values $4(p+1), 4(p+1)+2, \ldots, 4(p+1)+2(k-p)$.

The total number of mutually exclusive, exhaustive, and equally likely ways in which k factors can be examined is $k!$. The event $\mathcal{A}_j = \{N = 4(p+1)+2j\}$ would occur if the search contains j extra swaps, i.e., j inactive factors have to be explored before the pth active factor. This will happen if the $(p+j)$th factor examined is active and there are j inactive factors among the first $(p+j-1)$ factors examined. The total possible number of such arrangements is given by

$$n(\mathcal{A}_j) = \binom{p}{1}\binom{k-p}{j}\binom{p+j-1}{j} \times (p-1)!\, j!\, (k-p-j)!.$$

Because each of the above arrangements are mutually exclusive, exhaustive and equally likely, by classical definition of probability, it follows that

$$\Pr(\mathcal{A}_j) = \frac{n(\mathcal{A}_j)}{k!} = \frac{\binom{p}{1}\binom{k-p}{j}\binom{p+j-1}{j}(p-1)!j!(k-p-j)!}{k!},$$

which, after some algebraic manipulations, can be written in a more convenient form,

$$\Pr(\mathcal{A}_j) = p\binom{k-p}{j} \Big/ (p+j)\binom{k}{p+j}.$$

The expectation of N is given by

$$\begin{aligned} E(N) &= \sum_{j=0}^{k-p}(4(p+1)+2j)\Pr(\mathcal{A}_j) \\ &= 4(p+1) + 2\sum_{j=1}^{k-p} j\Pr(\mathcal{A}_j) \\ &\qquad \left(\text{because } \textstyle\sum_{j=0}^{k-p}\Pr(\mathcal{A}_j)=1\right) \\ &= 4(p+1) + 2\sum_{j=1}^{k-p} j\frac{p}{p+j}\frac{\binom{k-p}{j}}{\binom{k}{p+j}} \\ &= 4(p+1) + 2\frac{p}{p+1}(k-p) \end{aligned}$$

after some algebraic manipulations.

Proof of Result 2

(i) Let $\mathbf{Z}_p$ denote the $(2p+2)\times\binom{p}{2}$ matrix obtained by taking the pairwise product of the p columns of the $\mathbf{X}_p$ matrix. It is easy to see that each column of $\mathbf{X}_p$ is orthogonal to each column of $\mathbf{Z}_p$ and hence $\mathbf{X}_p'\mathbf{Z}_p = \mathbf{0}$. Also, let β_{INT} denote the $\binom{p}{2}\times 1$ vector of coefficients β_{ij}'s defined in model (3). Assuming absence of 3 fi's and higher order interactions, from model (3), we can write $E(\mathbf{y}) = \beta_0\mathbf{1}_{2p+2} + \mathbf{X}_p\beta_{\text{main}} + \mathbf{Z}_p\beta_{\text{INT}}$, where $\mathbf{1}_N$ denotes the $N\times 1$ vector of 1's. Now, by definition of $\hat{\beta}_{\text{main}}$,

$$\begin{aligned} E(\hat{\beta}_{\text{main}}) &= E\left[(\mathbf{X}_p'\mathbf{X}_p)^{-1}\mathbf{X}_p'\mathbf{y}\right] \\ &= E\left[(\mathbf{X}_p'\mathbf{X}_p)^{-1}(\mathbf{X}_p'\beta_0\mathbf{1}_{2p+2} + \mathbf{X}_p'\mathbf{X}_p\beta_{\text{main}} \right. \\ &\qquad \left. + \mathbf{X}_p'\mathbf{Z}_p\beta_{\text{INT}})\right] \\ &= \beta_{\text{main}} \end{aligned}$$

because $\mathbf{X}_p'\mathbf{1}_{2p+2} = \mathbf{0}$ and $\mathbf{X}_p'\mathbf{Z}_p = \mathbf{0}$.

(ii) The variance–covariance matrix of $\hat{\beta}_{\text{main}}$ is $\sigma^2\mathbf{D}_p^{-1}$, where $\mathbf{D}_p = (\mathbf{X}_p'\mathbf{X}_p)$. It is easy to verify that

$$\mathbf{D}_p = \begin{bmatrix} a_p & b_p & b_p & \cdots & b_p \\ b_p & a_p & b_p & \cdots & b_p \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ b_p & b_p & b_p & \cdots & a_p \end{bmatrix}, \tag{22}$$

where $a_p = 2p+2$ and $b_p = 2p-6$ are functions of p. The determinant of $\mathbf{D}_p$ can be obtained as

$$\det\mathbf{D}_p = \left[a_p + (p-1)b_p\right](a_p - b_p)^{p-1}. \tag{23}$$

Let $\mathbf{D}_p^r$ denote the $r\times r$ principal submatrix of $\mathbf{D}_p$ for $r = 1,\ldots,p$. Then

$$\det\mathbf{D}_p^r = \left[a_p + (r-1)b_p\right](a_p - b_p)^{r-1}. \tag{24}$$

Clearly, all the diagonal elements of $\mathbf{D}_p^{-1}$ will be the same and, when multiplied by σ^2, will represent $var(\hat{\beta}_i)$ for $i = 1,\ldots,p$. In particular, the $(1,1)$th element of $\mathbf{D}_p^{-1}$ will be equal to the adjugate of the $(1,1)$th element of $\mathbf{D}_p$ divided by $\det\mathbf{D}_p$. Because the adjugate of the $(1,1)$th element of $\mathbf{D}_p$ is $\det\mathbf{D}_p^{p-1}$, it follows that

$$\begin{aligned} \text{var}(\hat{\beta}_i) &= \sigma^2\frac{\det\mathbf{D}_p^{p-1}}{\det\mathbf{D}_p} \\ &= \sigma^2\frac{\left[a_p + (p-2)b_p\right]}{\left[a_p + (p-1)b_p\right](a_p - b_p)}, \\ &\qquad\qquad \text{by (23) and (24)} \\ &= \sigma^2\frac{p^2 - 4p + 7}{8(p^2 - 3p + 4)}. \end{aligned}$$

Proceeding in the same way, it can be shown that any off-diagonal element of $\mathbf{D}_p^{-1}$ can be written as $(\det\mathbf{C}_{p-1})/(\det\mathbf{D}_p)$, where

$$\mathbf{C}_p = \begin{bmatrix} b_p & b_p & b_p & \cdots & b_p \\ b_p & a_p & b_p & \cdots & b_p \\ b_p & b_p & a_p & \cdots & b_p \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ b_p & b_p & b_p & \cdots & a_p \end{bmatrix}$$

and a_p and b_p are defined as before. Because

$$\det\mathbf{C}_{p-1} = -b_p(a_p - b_p)^{p-2} = -8^{p-2}(2p-6),$$

we have

$$\text{cov}(\hat{\beta}_i, \hat{\beta}_j) = -\frac{p-3}{8(p^2 - 3p + 4)}\sigma^2,$$

by substituting the values of a_p and b_p.

Proof of Result 3

Part (i)

Consider any three columns of $\mathbf{D}$ that correspond to factors i, j, and k identified as active. Then the rows corresponding to stage 1 runs are $(+,+,+)$ and

$(-,-,-)$. It is straightforward to verify that the swapping runs for factors i, j, and k will generate the six remaining combinations of a 2^3 design.

Parts (ii)–(iv)

It is easy to see that the 8 swapping runs for any 4 factors A, B, C, and D generate a 2^{4-1} design with the defining relation $I = -ABCD$. Note that, if an additional column is introduced, six rows of that column will have the same symbol ($-$ or $+$) as in one of the four columns. Hence, the augmented matrix will not be an orthogonal array. Next, we show that $\mathbf{D}$ cannot contain an orthogonal array with more than 8 rows. Let $i_1, i_2, \ldots i_p$ denote the p active factors in the order in which they are identified as active. For each pair (i_k, i_{k+1}), the level combinations $(-,+)$ and $(+,-)$ can appear *twice and only twice* corresponding to the two swapping runs. Thus, the largest orthogonal submatrix of the $N \times 2$ matrix with columns i_k and i_{k+1} has 8 rows. To prove the last part, we note that the four runs along with two stage 1 runs and two capping runs $(+,+,-,-)$ and $(-,-,+,+)$ form a 2^{4-1} design with the defining relation $I = ABCD$. Therefore, the combination of the two half fractions with defining relations $I = -ABCD$ and $I = ABCD$ constitute a 2^4 full factorial design.

Hypothesis Test for Stage 1 and Derivation of /break Equations (13) and (14)

Because the variance of the median obtained from a random sample of size n drawn from a normal population can be approximated as $\pi/2n$ for large n, it is easy to see that the distribution of $M_\mathrm{b} - M_\mathrm{w}$ can be approximated as $N(2\sum_{i=1}^k \beta_i, (\pi\sigma^2)/3)$. Now let $\hat{\sigma_\mathrm{b}}^2$ and $\hat{\sigma_\mathrm{w}}^2$ denote the sample variances of the two sets of observations on y^+ and y^-, respectively, and $\hat{\sigma}^2 = (\hat{\sigma_\mathrm{b}}^2 + \hat{\sigma_\mathrm{w}}^2)/2$ denote the pooled variance. Then $(M_\mathrm{b} - M_\mathrm{w})/(\hat{\sigma}\sqrt{\pi/3})$ follows a noncentral t distribution with 4 degrees of freedom and noncentrality parameter $\delta = (2\sum_{i=1}^k \beta_i)/(\sigma\sqrt{\pi/3})$. Note that $\sum_{i=1}^k \beta_i > 0$ implies at least one of the factors is active.

Hypothesis Test for the Swapping Stage and Derivation of Equation (15)

Define the following statistics with respect to the swapping of factor x_i:

$$\begin{aligned} s^-(x_i) &= M_\mathrm{b} - y_i^-, \\ s^+(x_i) &= y_i^+ - M_\mathrm{w}. \end{aligned} \tag{26}$$

Because M_b is distributed approximately as $N(\beta_0 + \sum_j \beta_j + \sum_{j\neq k}\beta_{jk}, (\pi/6)\sigma^2)$ and $y_i^- \sim N(\beta_0 - \beta_i + \sum_{j\neq i}\beta_j - \sum_{j\neq i}\beta_{ij} + \sum_{j\neq i, k\neq i}\beta_{jk}, \sigma^2)$, the distribution of $s^-(x_i)$ can be approximated by $N(2(\beta_i + \sum_{j\neq i}\beta_{ij}), (1+\pi/6)\sigma^2)$. Similarly, the distribution of $s^+(x_i)$ can be approximated by $N(2(\beta_i - \sum_{j\neq i}\beta_{ij}), (1+\pi/6)\sigma^2)$. Note that factor x_i is active if and only if at least one of the following conditions holds good: (i) $\beta_i \neq 0$, (ii) $\sum_{j\neq i}\beta_{ij} \neq 0$. Thus, the following null hypothesis of interest at this stage is $H_0 : \beta_i = \sum_{j\neq i}\beta_{ij} = 0$. The hypothesis is rejected at level α if either or both of the following two events occur:

$$\begin{aligned} A^+ &: \frac{|s^+(x_i)|}{1.23\hat{\sigma}} > t_{4,\alpha/2}, \\ A^- &: \frac{|s^-(x_i)|}{1.23\hat{\sigma}} > t_{4,\alpha/2}, \end{aligned} \tag{27}$$

where $1.23 = (1+\pi/6)^{1/2}$. Because A^+ and A^- are independent events, Equation (15) follows immediately.

Hypothesis Test for the Capping Stage and Derivation of Equation (18)

Define the following statistics with respect to the capping of factors in $\mathcal{F}$:

$$\begin{aligned} C^+(\mathcal{F}) &= M_\mathrm{b} - y_\mathcal{F}^+, \\ C^-(\mathcal{F}) &= y_\mathcal{F}^- - M_\mathrm{w}. \end{aligned} \tag{28}$$

Proceeding as before, it can be easily seen that

$$C^+(\mathcal{F}) \sim N\left(2\sum_{i\notin\mathcal{F}}\beta_i + 2\sum_{i\notin\mathcal{F}}\sum_{j\in\mathcal{F}}\beta_{ij}, (1+\pi/6)\sigma^2\right),$$

$$C^-(\mathcal{F}) \sim N\left(2\sum_{i\notin\mathcal{F}}\beta_i - 2\sum_{i\notin\mathcal{F}}\sum_{j\in\mathcal{F}}\beta_{ij}, (1+\pi/6)\sigma^2\right).$$

Note that the capping runs will be successful (i.e., all factors $x_i, i \notin \mathcal{F}$ will be declared inert) if and only if both the following conditions hold good: (i) $\sum_{i\notin\mathcal{F}}\beta_i = 0$ and (ii) $\sum_{i\notin\mathcal{F}}\sum_{j\in\mathcal{F}}\beta_{ij} = 0$. Thus, the following null hypothesis of interest at this stage is $H_0 : \sum_{i\notin\mathcal{F}}\beta_i = \sum_{i\notin\mathcal{F}}\sum_{j\in\mathcal{F}}\beta_{ij} = 0$. The hypothesis is rejected (i.e., the capping run declared unsuccessful) at level α if either or both of the following two events occur:

$$\begin{aligned} B^+ &: \frac{|C^+(\mathcal{F})|}{1.23\hat{\sigma}} > t_{4,\alpha/2}, \\ B^- &: \frac{|C^-(\mathcal{F})|}{1.23\hat{\sigma}} > t_{4,\alpha/2}. \end{aligned} \tag{29}$$

Because B^+ and B^- are independent events, Equation (18) follows immediately.

Hypotheses Testing in the Presence of Three-Factor Interactions

Due to the presence of three-factor interactions, the expectations of the test statistics associated with stage 1, swapping and capping require the following modifications:

$$E(M_b - M_w) = 2\left(\sum_{i=1}^{k}\beta_i + \sum_{i<j,l}\sum_{j<l}\sum_{l=1}^{k}\beta_{ijl}\right),$$

$$E(s^+(x_i)) = 2\left(\beta_i + \sum_{j\neq i}\sum_{l\neq i,j}\beta_{ijl} + \sum_{j\neq i}\beta_{ij}\right),$$

$$E(s^-(x_i)) = 2\left(\beta_i + \sum_{j\neq i}\sum_{l\neq i,j}\beta_{ijl} - \sum_{j\neq i}\beta_{ij}\right),$$

$$E(C^+(\mathcal{F})) = 2\sum_{i\notin\mathcal{F}}\beta_i + 2\sum_{\psi(\mathcal{F})}\sum\sum\beta_{ijl} + 2\sum_{i\notin\mathcal{F}}\sum_{j\in\mathcal{F}}\beta_{ij},$$

$$E(C^-(\mathcal{F})) = 2\sum_{i\notin\mathcal{F}}\beta_i + 2\sum_{\psi(\mathcal{F})}\sum\sum\beta_{ijl} - 2\sum_{i\notin\mathcal{F}}\sum_{j\in\mathcal{F}}\beta_{ij},$$

where $\psi(\mathcal{F})$ is the set of all triplets (i, j, l), where one or all of the $i, j, l \notin \mathcal{F}$.

The variance expressions and the test procedures are exactly identical to those derived in the Appendices before.

References

BHOTE, K. R. and BHOTE, A. K. (2000). *World Class Quality.* New York, NY: Amacom.

BOX, G. E. P. and TYSSEDAL, J. (1996). "Projective Properties of Certain Orthogonal Arrays". *Biometrika* 83, pp. 950–955.

BOX, G. E. P. and HUNTER, J. (1961). "The 2^{k-p} Fractional Factorial Designs: Part I". *Technometrics* 3, pp. 311–351.

LEDOLTER, J. and SWERSEY, A. (1997). "Dorian Shainin's Variable Search Procedure: A Critical Assessment". *Journal of Quality Technology* 29, pp. 237–247.

MILLER, A. and SITTER, R. R. (2001). "Using the Folded-Over 12-Run Plackett–Burman Design to Consider Interactions". *Technomterics* 43, pp. 44–55.

PLACKETT, R. L. and BURMAN, J. P. (1946). "The Design of Optimum Multifactorial Experiments". *Biometrika* 33, pp. 305–325.

SHAININ, D. (1986). "Better than Taguchi Orthogonal Tables". *ASQC Quality Congress Transactions.* Anaheim, CA, pp. 446–481.

SHAININ, D. and SHAININ, P. (1988). "Better than Taguchi Orthogonal Tables". *Quality and Reliability Engineering International* 4, pp. 143–149.

STEINER, S. H.; MACKAY, R. J.; and RAMBERG, J. S. (2008). "An Overview of the Shainin System™ for Quality Improvement". *Quality Engineering* 20, pp. 6–19.

VERMA, A. K.; SRIVIDYA, A.; MANNIKAR, A. V.; PANKHAWALA, V. A; and RATHNARAJ, K. J. (2004). "SHAININ Method: Edge Over Other DOE Techniques". *IEEE International Engineering Management Conference, Proceedings* 20, pp. 6–19.

WU, C. F. J. and HAMADA, M. (2009). *Experiments: Planning, Analysis, and Optimization,* 2nd edition. New York, NY: Wiley.