

Tilburg University

Full Screening Another View at One-Factor-at-a-Time Designs

Bettonvil, B.W.M.

Publication date:
2010

[Link to publication in Tilburg University Research Portal](#)

Citation for published version (APA):

Bettonvil, B. W. M. (2010). *Full Screening Another View at One-Factor-at-a-Time Designs*. (CentER Discussion Paper; Vol. 2010-139). Information Management.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

No. 2010-139

**FULL SCREENING
ANOTHER VIEW AT ONE-FACTOR-AT-A-TIME DESIGNS**

By Bert Bettonvil

December 2010

ISSN 0924-7815

FULL SCREENING ANOTHER VIEW AT ONE-FACTOR-AT-A-TIME DESIGNS

Bert Bettonvil

Tilburg University

Key Words: simulation, design of experiments, optimization, screening.

Jel code: C0, C1, C9.

Abstract

Textbooks on Design Of Experiments invariably start by explaining why one-factor-at-a-time (OAT) is an inferior method. Here we will show that in a model with all interactions a variant of OAT is extremely efficient, provided that we only have non-negative parameters and that there are only a few large parameters. In the extreme, this means that there is one positive parameter. In that case, for m variables, that is, for $2^m - 1$ parameters, the procedure only needs $m + 2$ observations to find. In other examined cases, the proposed method is also very fast.

1. Introduction

Screening techniques are used to reduce the number of potentially important factors influencing a system's performance to the few really important factors, and to do so in an economical way. The latter is important because either we are dealing with a real system, for which it is difficult or expensive to change factor values for a new run, or we are dealing with a complicated simulation for which a single simulation run may take so long (hours, days) that performing many runs is unacceptable.

The first publications on (group) screening appear in the 1960's: see Watson (1961), Patel (1962), or Li (1962); an overview of the various methods of screening is given in Morris (2006) and Webb (2010), who both discuss among others Bettonvil's (1995) Sequential Bifurcation. The latter technique has the desirable property that if exactly one factor out of 2^m is nonzero, it will be found in $m + 2$ observations. In the sequel we will see a corresponding logarithmic relation between the number of parameters in the problem at hand and the number of observations.

In this paper we assume we are dealing with (a simulation of) a system depending on m factors, each of which can be in two positions: low and high. Consequently, the system can be in any of 2^m states. We assume that the real system is, to start with, in the state with all factors "low"; that going from one state to another by increasing one (or more) factors does not decrease the output, or yield, of the system (so the yield either increases or remains the same), but on the other hand, there is are known (positive) costs involved for the increment of each factor. We do not want to maximize the yield of the system; that would be trivial (the yield is maximal in the state with all factors "high"), but instead we want to maximize the profit reached by increasing some factors, that is,

maximize the difference between yield and costs, or the quotient of yield and costs.

Our model is

$$y(x_1, \dots, x_m) = \beta_0 + \beta_1 x_1 + \dots + \beta_m x_m + \beta_{1,2} x_1 x_2 + \dots + \beta_{m-1,m} x_{m-1} x_m + \beta_{1,2,3} x_1 x_2 x_3 + \dots + \beta_{m-2,m-1,m} x_{m-2} x_{m-1} x_m + \dots + \beta_{1,2,\dots,m} x_1 x_2 \dots x_m. \quad (1)$$

where, without loss of generality, we assume that all $x_i \in \{0, 1\}$. Further, we assume that (apart from β_0) all β 's are non-negative, with the majority of them zero or small, and with only a few large ones, where we will discuss "small" and "large" later. Model (1) represents a system with m inputs and one output; we think of systems like production lines, where changing input variables like conveyor-belt speed, buffer sizes, and the like, can increase the production volume. We assume that, starting from the "present situation" with all input variables at their low levels (all $x_i = 0$), changing one or more input variables to their high levels will increase (at least: not decrease) the yield y , and that there are only a few *combinations* that do increase the output, with the ideal case: there is exactly one combination of input variables that plays a decisive role in determining the output.

Most screening methods are restricted to main effect models. The exceptions are Morris (1991), who detects (but does not measure) quadratic effects and interactions, and Bettonvil (1990), who finds important main effects in models with first order interactions, using the Box and Wilson's (1951) foldover principle. As (1) shows, we assume a model with all interactions, but without quadratic effects.

In the following, we deviate from standard notation to improve readability. Instead of writing $y(x_1, \dots, x_m)$, where (x_1, \dots, x_m) is a series of 0's and 1's, we write y_S , where S is the set of inputs at their high levels; so y_\emptyset is the present yield: all inputs at their low levels; and y_U is the yield with all inputs at their high levels. If the inputs with indices in set S are at their high levels, we have costs $C_S = \sum_{s \in S} c_s$ with $c_s > 0$ ($s = 1, \dots, m$), and yield y_S . We can either maximize the absolute profit $y_S - C_S$, or the relative profit $(y_S - y_\emptyset)/C_S$. In the sequel we will mainly deal with the relative profit, where for simplicity we assume all c 's equal to one, so we concentrate on $(y_S - y_\emptyset)/\#S$, but this not a fundamental restriction. Also see Appendix A, remark 1.1.

In section 2 we consider three versions (including a trivial one) of, what we call, "perfect" cases, namely the case where all parameters are zero, the cases where exactly one parameter (be it a main effect or an interaction of any order) is positive, and the case with only main effects. In section 3 we switch to the general case, and we show how we arrive at the optimal vertex. In section 4 we give some examples, and in section 5 we discuss our findings.

2. "Perfect" cases

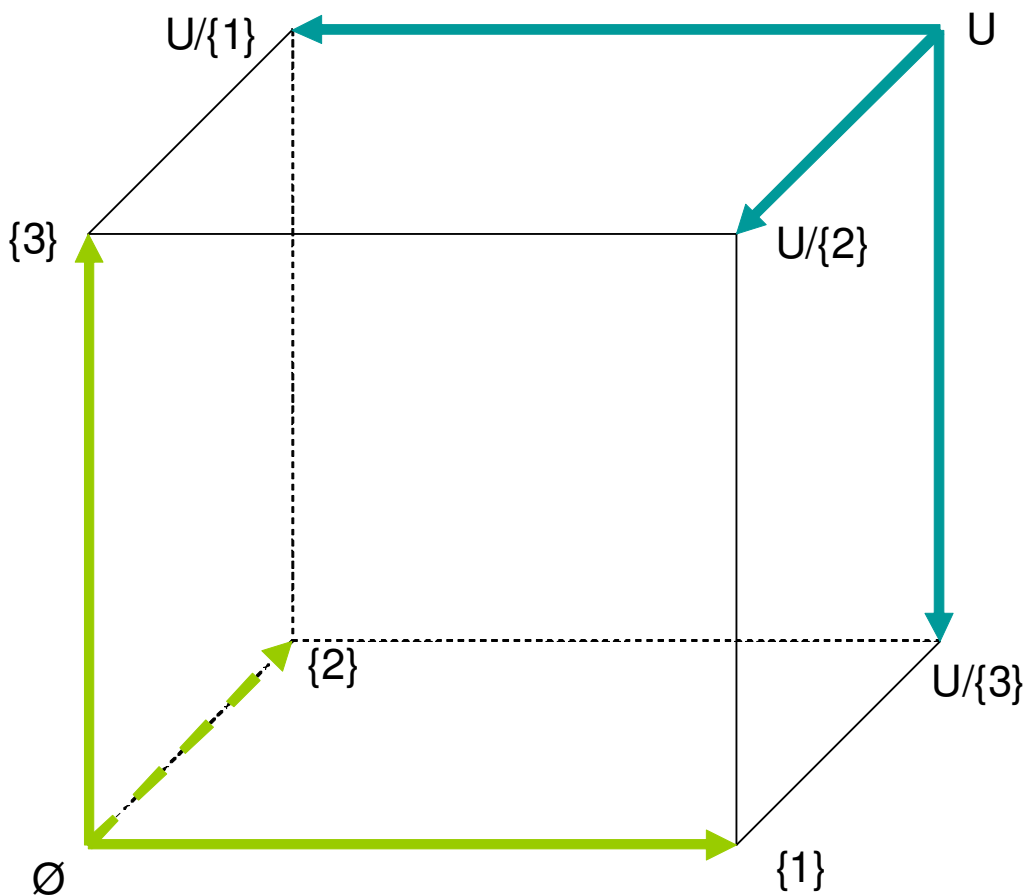
2.0. All parameters zero

Just for the sake of completeness we consider the case where all parameters are zero. In that case the yield is the same, whatever measures are taken to increase it, so the best policy is: do not invest any money in taking such measures: y_\emptyset is optimal. We can see that all parameters are zero by observing y_\emptyset and y_U . All parameters are zero if and only if (in the sequel: iff) $y_U = y_\emptyset$.

2.1. Exactly one positive parameter

If exactly one parameter is positive, then $y_U > y_\emptyset$. If we observe this, we turn to OAT. In general, OAT is used starting at the origin, so apart from y_\emptyset we would observe $y_{\{1\}}, y_{\{2\}}, \dots, y_{\{m\}}$. But only if the one positive effect is a main effect, we get to see it. If we observe $y_\emptyset, y_{\{1\}}, y_{\{2\}}, \dots, y_{\{m\}}$, and y_U , and we find that there one index i with $y_{\{i\}} = y_U$, and all other $y_{\{j\}} = y_\emptyset (< y_U)$, we conclude that there is only one effect, and this is a main effect. Applied in this way, OAT has very limited use.

Figure 2.1.1. Two ways of OAT



But, there is an alternative. Suppose, apart from y_\emptyset and y_U , we observe $y_{U \setminus \{1\}}, y_{U \setminus \{2\}}, \dots, y_{U \setminus \{m\}}$; that is, OAT starting from U : the point where all inputs are at their high levels. Notice the correspondence between $y_U - y_\emptyset$ and $y_{U \setminus \{i\}} - y_\emptyset$: while the first is the sum of all regression parameters in a model with m variables, the latter is its equivalent in a model with $m - 1$ variables, where variable i is left out, so

PROPERTY

(P1) $y_{U \setminus \{i\}} - y_\emptyset$ is the sum of all regression parameters that do not have i as an index.

(P2) $y_U - y_{U \setminus \{i\}}$ is the sum of all regression parameters that have i as an index.

(P2) is an immediate consequence of (P1). This partitioning of parameters is similar to Homma and Saltelli's (1996).

If some $\beta_S > 0$ and this is the only positive parameter, then, as a consequence of (P1), all observations $y_{U \setminus \{i\}}$ with $i \in S$ will be equal to y_\emptyset , and, as a consequence of (P2), all observations $y_{U \setminus \{i\}}$ with $i \notin S$ will be equal to y_U .

And the reverse also holds. Suppose we observe y_\emptyset, y_U , and $y_{U \setminus \{1\}}, y_{U \setminus \{2\}}, \dots, y_{U \setminus \{m\}}$, and we find that $y_\emptyset < y_U$, and there is a set $S \subset \{1, 2, \dots, m\}$ such that for $i \in S$, $y_{U \setminus \{i\}} = y_\emptyset$, and that for $i \notin S$, $y_{U \setminus \{i\}} = y_U$. Then β_S is the only positive parameter. See proof 2.1.1 in Appendix B.

This is an important property of the systems under study: in case there is a single non-zero parameter (in a collection of $2^m - 1$ parameters), it takes $m + 2$ observations to find it.

See appendix A, remarks 2.1.1 through 2.1.3 for some further aspects of the one positive parameter case.

2.2. Only main effects

Suppose that only the main effects are positive. For ease of notation we take $y_\emptyset = 0$. It is plain to see that

$$y_U = \sum_j \beta_j \tag{2a}$$

$$y_{U \setminus \{i\}} = \sum_j \beta_j - \beta_i \tag{2b}$$

so that $\sum_i (y_U - y_{U \setminus \{i\}}) = \sum_i \beta_i = y_U$.

If, on the other hand, we observe that $\sum_i (y_U - y_{U \setminus \{i\}}) = y_U$, can we conclude then that we are dealing with main effects only? Because $y_U - y_{U \setminus \{i\}}$ is the sum of all parameters with an index i , we have

$$\sum_i (y_U - y_{U \setminus \{i\}}) = \sum_i \beta_i + 2\sum_{i,j} \beta_{i,j} + 3\sum_{i,j,k} \beta_{i,j,k} + \dots + m\beta_{1,2,\dots,m}. \tag{3}$$

Any $\beta_{i,j}$ appears twice in this summation, one time coming from i , one time coming from j ; any $\beta_{i,j,k}$ appears three times, caused by i, j , and k ; etcetera. But also

$$y_U = \sum_i \beta_i + \sum_{i,j} \beta_{i,j} + \sum_{i,j,k} \beta_{i,j,k} + \dots + \beta_{1,2,\dots,m} \tag{4}$$

trivially. Equality of (3) and (4) is only possible if all terms with coefficient greater than one are zero, due to the fact that all regression coefficients are non-negative.

3. The general case

3.1. Admissible observations

Given the observations $y_\emptyset < y_U$, and $y_{U \setminus \{i\}}$ with $y_\emptyset \leq y_{U \setminus \{i\}} \leq y_U$ ($i = 1, \dots, m$), we have to determine whether these observations satisfy the model constraints, that is, model (1) with all non-negative regression coefficients. We already saw, see remark 1, that not all data with $y_\emptyset \leq y_{U \setminus \{i\}} \leq y_U$ ($i = 1, \dots, m$) satisfy our model demands: for $i = 1, \dots, m$ $y_{U \setminus \{i\}}$ may not be too large, or, equivalently, $y_U - y_{U \setminus \{i\}}$ may not be too small. The derivation in section 2.2 leads to the idea that $\sum_{i=1}^m (y_U - y_{U \setminus \{i\}}) \geq y_U - y_\emptyset$.

PROPOSITION

Model (1) holds with only non-negative regression parameters if and only if $\sum_{i=1}^m (y_U - y_{U \setminus \{i\}}) \geq y_U - y_\emptyset$ (provided that $y_\emptyset \leq y_{U \setminus \{i\}} \leq y_U$ ($i = 1, \dots, m$)).

The proof is given in Appendix B, proof 3.1.1. The outlines of the proof and the following example are the same.

EXAMPLE

We have 6 variables; $y_\emptyset = 0$, $y_U = 6$ and $y_U - y_{U \setminus \{i\}}$ ($i = 1, \dots, 6$) are given in Table 1.1 below. To simplify matters, we ordered the observations in ascending order. Now we will show how to arrive at seven positive parameters that are in accordance with these observations, and with model (1). We start with the tentative parameters $\tilde{\beta}_1, \tilde{\beta}_2, \dots, \tilde{\beta}_6$ and $\tilde{\beta}_{1,2,3,4,5,6}$ as given in Table 1.2, where $x \in [0, 0.228]$, that is, we start with giving each $\tilde{\beta}_i$ the value $y_U - y_{U \setminus \{i\}}$ ($i = 1, \dots, 6$) and next we let them all decrease, and at the same time we let $\tilde{\beta}_{1,2,3,4,5,6}$ increase in such a way that the sum of the parameters containing an index i remains $y_U - y_{U \setminus \{i\}}$ ($i = 1, \dots, 6$). The total sum of these parameters is far too large: it goes from 18.092 to 16.952 as x goes from 0 to 0.228. Now we take a next step: after "exhausting" $\tilde{\beta}_6$ and maximizing $\tilde{\beta}_{1,2,3,4,5,6}$, that is, after maximizing x , we will gradually replace $\tilde{\beta}_5$ by $\tilde{\beta}_{1,2,3,4,5}$. We switch to Table 1.3, where $x \in [0, 0.313]$ (Table 1.2 with $x = 0.228$ is identical to Table 1.3 with $x = 0$). Still, for all $i = 1, \dots, 6$ the sums of the parameters satisfy the equality to $y_U - y_{U \setminus \{i\}}$, as is easily verified, and as x increases from 0 to 0.313, the sum of all parameters decreases from 16.952 to 15.700. In the next step, we replace $\tilde{\beta}_4$ by $\tilde{\beta}_{1,2,3,4}$ and adapt the other parameters to keep the sums OK: see Table 1.4. In this table, as x increases from 0 to 0.290, the total decreases from 15.700 to 14.830, which is still too large. Next, we replace $\tilde{\beta}_3$ by $\tilde{\beta}_{1,2,3}$ and again adapt the parameters, see Table 1.5. Here, as x increases from 0 to 4.598, the total sum decreases from 14.830 to 5.634, which is smaller than the "goal value" 6.000. However, for $x=4.415$ we have a solution which satisfies our demands; see Table 1.6.

$y_U - y_{U \setminus \{1\}}$	5.560
$y_U - y_{U \setminus \{2\}}$	5.503
$y_U - y_{U \setminus \{3\}}$	5.429
$y_U - y_{U \setminus \{4\}}$	0.831
$y_U - y_{U \setminus \{5\}}$	0.541
$y_U - y_{U \setminus \{6\}}$	0.229
y_U	6

Table 1.1

$\tilde{\beta}_1$	$5.560 - x$
$\tilde{\beta}_2$	$5.503 - x$
$\tilde{\beta}_3$	$5.429 - x$
$\tilde{\beta}_4$	$0.831 - x$
$\tilde{\beta}_5$	$0.541 - x$
$\tilde{\beta}_6$	$0.228 - x$
$\tilde{\beta}_{1,2,3,4,5,6}$	x

Table 1.2

$\tilde{\beta}_1$	$5.332 - x$
$\tilde{\beta}_2$	$5.275 - x$
$\tilde{\beta}_3$	$5.201 - x$
$\tilde{\beta}_4$	$0.603 - x$
$\tilde{\beta}_5$	$0.313 - x$
$\tilde{\beta}_{1,2,3,4,5}$	x
$\tilde{\beta}_{1,2,3,4,5,6}$	0.228

Table 1.3

$\tilde{\beta}_1$	$5.019 - x$
$\tilde{\beta}_2$	$4.962 - x$
$\tilde{\beta}_3$	$4.888 - x$
$\tilde{\beta}_4$	$0.290 - x$
$\tilde{\beta}_{1,2,3,4}$	x
$\tilde{\beta}_{1,2,3,4,5}$	0.313
$\tilde{\beta}_{1,2,3,4,5,6}$	0.228

Table 1.4

$\tilde{\beta}_1$	$4.729 - x$
$\tilde{\beta}_2$	$4.672 - x$
$\tilde{\beta}_3$	$4.598 - x$
$\tilde{\beta}_{1,2,3}$	x
$\tilde{\beta}_{1,2,3,4}$	0.290
$\tilde{\beta}_{1,2,3,4,5}$	0.313
$\tilde{\beta}_{1,2,3,4,5,6}$	0.228

Table 1.5

$\tilde{\beta}_1$	0.314
$\tilde{\beta}_2$	0.257
$\tilde{\beta}_3$	0.183
$\tilde{\beta}_{1,2,3}$	4.415
$\tilde{\beta}_{1,2,3,4}$	0.290
$\tilde{\beta}_{1,2,3,4,5}$	0.313
$\tilde{\beta}_{1,2,3,4,5,6}$	0.228

Table 1.6

It follows that we can write the $m + 2$ observations y_\emptyset , y_U , and $y_{U \setminus \{i\}}$ ($i = 1, \dots, m$) in terms of β_0 and $m + 1$ positive (to cover exceptional cases: non-negative) regression parameters, and $2^m - m - 2$ regression parameters equal to zero. This is not only true for these particular observations, but for any set of observations generated by model (1), which includes y_\emptyset .

LEMMA.

If the $k + 1$ observations y_\emptyset, y_{S_j} ($j = 1, \dots, k$) are generated by model (1), then k

non-negative parameters suffice to explain $y_{S_j} - y_\emptyset$. That is, let
$$\begin{bmatrix} y_{S_1} - y_\emptyset \\ y_{S_2} - y_\emptyset \\ \vdots \\ y_{S_k} - y_\emptyset \end{bmatrix} = X\beta$$
 with X

a $k \times \ell$ matrix with $k < \ell$, the elements of X and β follow from model (1), which implies that X contains only 0's and 1's, and that β contains only non-negative elements. Then we can reduce the number of elements of β from ℓ to k ; that is, we can select k columns of X giving us a $k \times k$ matrix \tilde{X} , and find a k -vector $\tilde{\beta}$ with only non-negative elements such that

$$\begin{bmatrix} y_{S_1} - y_\emptyset \\ y_{S_2} - y_\emptyset \\ \vdots \\ y_{S_k} - y_\emptyset \end{bmatrix} = X\beta = \tilde{X}\tilde{\beta}.$$

See proof 3.1.2 in Appendix B.

3.2. Candidates

Before describing the way in which the optimal vertex is found, that is, where is the profit maximal, we must consider the matter of candidates. It is quite possible that not every vertex is acceptable as a solution. It may be that the company fixed an upper limit of the costs; maybe taking some measure prohibits taking some other measure, or prescribes some other measure. Anyway, it is possible, and even probable, that not all vertices are acceptable as a solution. The vertices that are acceptable as a solution, are called candidates, and we indicate them as CAND(0). In the sequel we will see how to arrive at just one candidate: the optimal solution.

3.3. Determination of minima and maxima

In this section, we assume that $\beta_0 = y_0 = 0$. Assume we have performed k observations, collected in vector y_S . There is a k -vector $\tilde{\beta}$ (a basic solution) such that $y_S = \tilde{X}\tilde{\beta}$, where the elements of the $k \times k$ matrix \tilde{X} are determined by the indices of the observations and those of the regression coefficients. For any \tilde{y} , not yet observed, we can compute the value, based on the values of the elements of $\tilde{\beta}$: $\tilde{y} = x'_n \tilde{\beta}$, where x_n is a k -vector, the elements of which are determined by the index of \tilde{y} and the indices of $\tilde{\beta}$. This does not give us the value of \tilde{y} if we actually performed this observation, but it is a feasible value; that is, there is an interval, although unknown at the moment, in which the value of \tilde{y} lies, and also $x'_n \tilde{\beta}$ lies within this interval. So we have

$$\begin{bmatrix} y_S \\ \tilde{y} \end{bmatrix} = \begin{bmatrix} \tilde{X} \\ x'_n \end{bmatrix} \tilde{\beta}.$$

Now we add one extra elements β_e to this system, and we give it the value 0, so it leaves everything unchanged. The system changes into

$$\begin{bmatrix} y_S \\ \tilde{y} \end{bmatrix} = \begin{bmatrix} \tilde{X} & x_e \\ x'_n & x_{ne} \end{bmatrix} \begin{bmatrix} \tilde{\beta} \\ \beta_e \end{bmatrix} \quad (5)$$

where the elements of x_e and the value of x_{ne} are determined by the y - and β -indices. What is the effect if we let β_e increase, taking care that y_S does not change? We see that $y_S = \tilde{X}\tilde{\beta} + x_e\beta_e$, implying that $\tilde{\beta} = \tilde{X}^{-1}y_S - \tilde{X}^{-1}x_e\beta_e$, provided that \tilde{X} is not singular. For this issue, see remark 3.3.1. We also see that $\tilde{y} = x'_n\tilde{\beta} + x_{ne}\beta_e$. In this expression we substitute the value of $\tilde{\beta}$ just found, giving us $\tilde{y} = x'_n\tilde{X}^{-1}y_S + (x_{ne} - x'_n\tilde{X}^{-1}x_e)\beta_e$.

Depending on the choice of e (that is, which extra β do we add to the system) we can let \tilde{y} increase or decrease. Assume that we want to let \tilde{y} increase (decreasing goes in a similar way). We find an index e such that $x_{ne} - x'_n\tilde{X}^{-1}x_e > 0$ and let β_e increase while keeping y_S constant, and keeping all elements of $\tilde{\beta}$ non-negative. The latter demand gives the limit of how far we can let β_e increase. At the point where one element of $\tilde{\beta}$ becomes 0, we stop increasing β_e and, at the same point, we can remove this zero-element from $\tilde{\beta}$ and replace it by β_e .

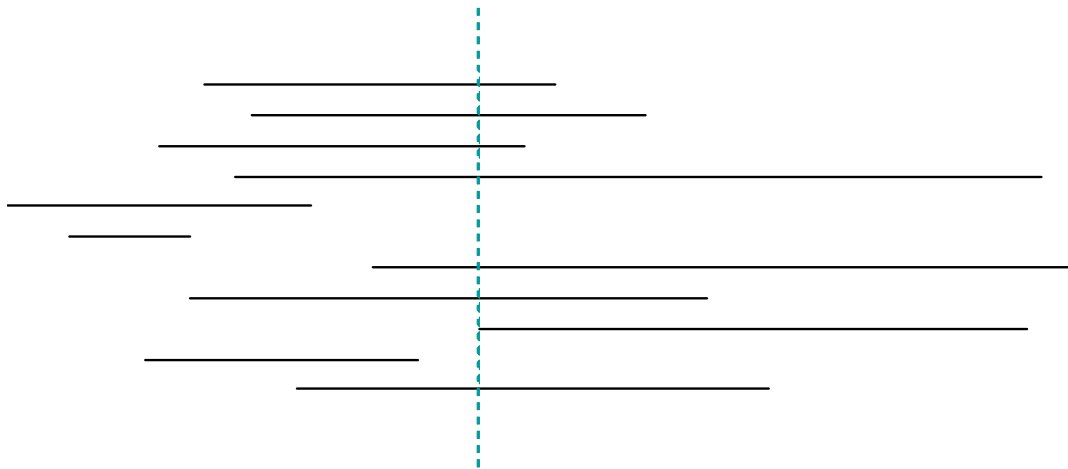
We then can look for another element to increase \tilde{y} further, and repeat this process until \tilde{y} can not be increased anymore.

3.4. First application of minima and maxima

For all candidates we can now compute the minimum and maximal value attainable. Consider the maximin of these: what is the maximum of the minimum values. All candidates with a maximum attainable value smaller than this maximum can be

discarded, and in this way the set of candidates is reduced from CAND(0) to CAND(1).

Figure 3.4.1. Reducing the number of candidates.



3.5. Second application of minima and maxima

As a next step, we compute the minimum and the maximum values attainable for all vertices not yet observed. This is done in the same way as the computation of these quantities for the candidates. We choose a value a bit larger than the minimum and a value a bit smaller than the maximum (see remark 3.5.1 in Appendix A), and for each of these we continue as if we had observed this value.

We then can compute what would happen to the candidates, had that particular value been observed ("first application of minima and maxima"). That gives us for all vertices (not yet observed) a reduction of the number of candidates for a small value in the vertex as well as such a reduction for a large value. The optimal new observation is the one with the largest reduction of candidates.

To be more specific:

We compute the reduction for a small value and for a large value in a vertex (say) x . Call the minimum of these two \min_x and the maximum \max_x . We do so for all vertices (not yet observed). Then we select

(a) the vertex where \min_x is maximal: $\max(\min_x | x \text{ not observed})$. If this maximum is attained in a unique point, this point is selected. If not

(b) among the points with maximal minimum, we select the point or points with the largest maximum: $\max(\max_x | x \text{ not observed})$. Again, if this maximum is attained in a unique point, this point is selected. If not,

(c) among the points, found in (b), the point with the smallest difference in attainable large value and small value is selected (see remark 3.5.2 in Appendix A). If this still does not give a unique point, a point is selected randomly from the group now left over.

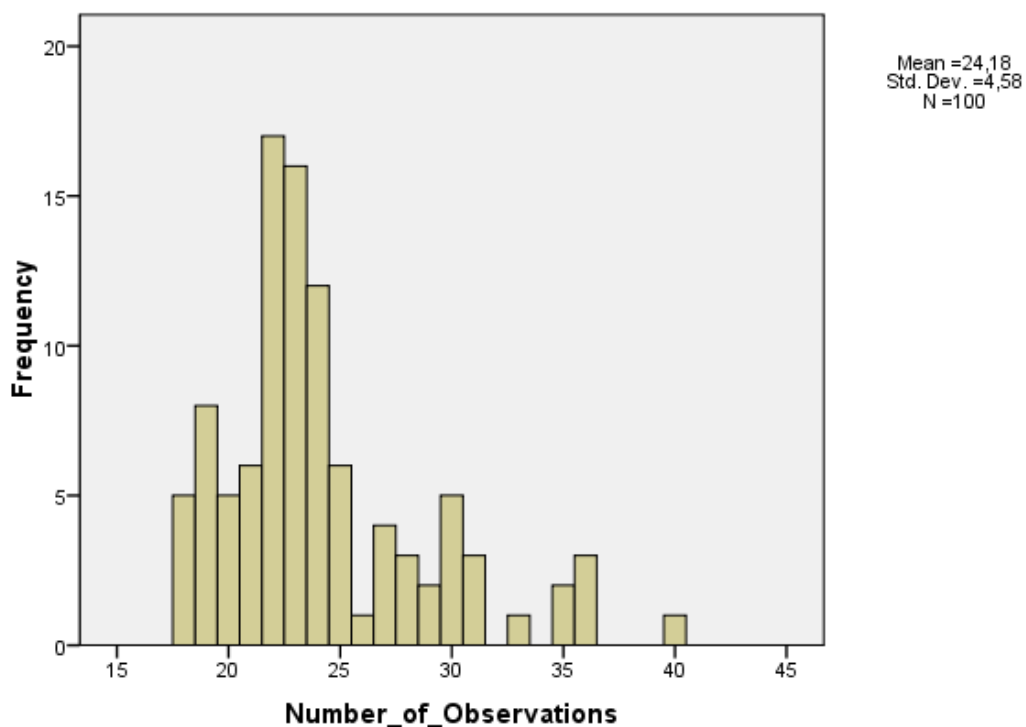
In this way, a vertex is selected and in this vertex an observation is performed. First it is checked whether this observation falls in the interval of attainable values (computed earlier). If it is not, the model restrictions are violated, so we must stop. However, if the observed value is admissible, then a new basic solution is computed (see remark 3.5.3 in Appendix A), and the set of candidates is updated.

If necessary, this procedure is repeated until we have found an optimum.

4. Some synthetic examples

To see how the screening procedure works out, we generated data as follows: we drew a random sample of size 255 from the negative exponential distribution (with parameter 1) and assigned these values to the parameters in model (1) (we gave β_0 value zero). Using the vertex values based on these parameters we ran the procedure. This was replicated 100 times. The number of observations needed is given in figure 4.1. We see a minimum of 18, a maximum of 40, and a mean of 24 observations. Note that without constraints 256 observations would be needed.

Figure 4.1



But our data do not really satisfy the demand of "a few large, the majority small". For each replication we computed the number of parameters that are larger than the mean of these parameters. This measure of skewness was inspired by the intuitively appealing Cyhelský's a , see Appendix A, remark 4.1. For our 100 replicates the number of large parameters varied between 81 and 104, with a mean of 93. This is by no means "a few". So for far-from-ideal data the method works efficiently.

To improve the skewness, we simply squared all parameters, and we did so again, and again, and again. So we have our original parameters (from the negative exponential distribution), and these values squared, and raised to the powers 4, 8, and 16. The results are in table 4.1. This table shows that with a decreasing number of large

parameters, the number of observations needed decreases, as is to be expected.

Power	MinObs	MaxObs	MeanObs	MinLarge	MaxLarge	MeanLarge
1	18	40	24	81	104	93
2	15	32	23	48	76	62
4	11	31	18	15	46	31
8	10	25	14	2	23	11
16	10	12	10	1	12	4

Table 4.1. Simulation Results

5. Discussion

Finding the optimal vertex in $\{0, 1\}^m$, given a model with only non-negative regression parameters, can be done in an efficient way, especially for the cases (a) there is exactly one positive parameter, and (b) only the main effect parameters are positive. These cases may only be of theoretical interest, the method described in this paper also proves its value in more general cases, as is shown in the examples.

Some questions are still open. First, how robust is the method against "minor" assumption violations (small negative parameters)? Second, how can the method be adapted for models with random error? Third, can the method be accelerated by using more sophisticated techniques than LP? Finally, we do not yet have a real-world example.

Appendix A. Remarks

REMARK 1.1.

The cost structure should be such that going from one vertex in the hypercube $\{0, 1\}^m$ to a neighbor by replacing one 0 into a 1, the cost always increases ("nothing comes for free"). This can of course be achieved in many ways, but a rather general formula is $C = x^T A x$ where A is a (symmetric) positive definite matrix. Further, the profit is a function of yield and cost, in such a way that if $y_1 < y_2$, then $P(y_1, C) < P(y_2, C)$ and if $C_1 < C_2$, then $P(y, C_1) > P(y, C_2)$. The exact shape of the cost structure and the profit structure is irrelevant, but both must be known functions.

REMARK 2.1.1.

There are 2^m ways in which $y_{U\setminus\{1\}}, y_{U\setminus\{2\}}, \dots, y_{U\setminus\{m\}}$ can assume the values y_U and y_\emptyset (with $y_U > y_\emptyset$). Each of these combinations corresponds to one non-zero parameter, except the combination $y_{U\setminus\{1\}} = y_{U\setminus\{2\}} = \dots = y_{U\setminus\{m\}} = y_U$. The latter would mean that no parameter is important, and that is impossible when $y_U > y_\emptyset$.

REMARK 2.1.2.

In order that the two statements

(i) only $\beta_S > 0$

(ii) for $i \in S$, $y_{U\setminus\{i\}} = y_\emptyset$, and for $i \notin S$, $y_{U\setminus\{i\}} = y_U$

are equivalent, it is not necessary to assume that all parameters are non-negative. It suffices to assume that

(a) increasing a factor does not decrease the output, for any combination of values of the remaining factors: $\forall_i \forall_{T \subset U} : y_{T \cup \{i\}} \geq y_T$

(b) consider two points in $\{0, 1\}^m$, both with index i at value 0, where one of the points has all its indices at least as high as the indices of the other point. Then the effect of increasing factor i at the first point is at least as big as the effect of increasing factor i at the second point: $\forall_i \forall_{T \subset V \subset U, i \notin V} : y_{V \cup \{i\}} - y_V \geq y_{T \cup \{i\}} - y_T$.

This is proven in Appendix B, proof 2.1.1. Nevertheless, we maintain our demand of non-negative parameters to be able to find the optimal point in "imperfect" cases.

REMARK 2.1.3.

If there is exactly one non-negative parameter (say) β_S , then only the adaptation of the system from the "present situation", indicated by \emptyset , which is "no change", to S MAY make sense. We know the cost of going from \emptyset to S is c_S and the extra yield is β_S and these must be weighted one against the other.

REMARK 2.2.1.

"Main effects only" is the *only* situation in which OAT from the origin (the "classic" or "naive" approach) is useful; in order to verify whether all interaction effects are zero, it suffices to perform the observations y_\emptyset , $y_{\{i\}}$, and y_U . If indeed $y_U - y_\emptyset = \sum_i (y_{\{i\}} - y_\emptyset)$, then there are only main effects. Notice that in addition to the standard observations, y_U has to be observed. So we can give the advice: if for some reason you want to apply the classic OAT design, you can check the quality of the underlying main-effects-only model by adding observation y_U .

REMARK 2.2.2.

If only main effects play a role in $y(x)$, then it depends on the relationship between $p(x)$, $y(x)$, and $C(x)$, which is the optimal solution:

- if $p(x) = y(x) - C(x)$, then the optimal solution includes all variables i for which $\beta_i > c_i$;
- if $p(x) = (y(x) - y_\emptyset) / C(x)$, then choose variable i for which β_i / c_i is maximal.

REMARK 3.3.1.

Matrix \tilde{X} in (5) may be singular, implying that only $k - 1$ parameters are needed to explain k observations. In that exceptional case, we can remove a row in \tilde{X} , and an element of $\tilde{\beta}$, in the way described in proof 3.1.2 in Appendix B. Next we add an row that does not depend on the rows already in \tilde{X} (which is always possible, given the structure of model (1)), and the corresponding element (say) $\tilde{\beta}_j$, which is given value 0. Only minor adaptations are needed to the search procedure of the main text.

REMARK 3.5.1.

If we would choose the maximum and minimum value attainable in a not yet observed vertex, then we would certainly arrive in a situation where at least one regression parameter is zero. Take a look at equation (5), and let us assume that we are maximizing \tilde{y} . Then there is some final step in the search procedure. Let that be β_e . This parameter is increased until one of the elements in $\tilde{\beta}$ becomes zero. According to our final step assumption we have reached the maximum, and one of the elements is zero.

REMARK 3.5.2.

After steps (a) and (b) we are left with some vertices for which the reductions of the number of candidates are the same. We tested the following situation, which is not at all like having one, or a few, large parameters, namely we took a model with all parameters in (1) equal to one, and we compared selecting the vertex with the largest and the smallest difference in attainable large value and small value. It turned out that selecting the smallest difference results in fewer observations.

REMARK 3.5.3.

Starting from the "old" basic solution, see whether the observed value in the selected vertex is smaller or larger than the value based on the actual $\beta'_s : x'_n \tilde{\beta}$ in terms of model (5). Then increase or decrease the value of \tilde{y} until the observed value is reached. This is the new basic solution.

REMARK 4.1.

In the ideal case, there is only one positive parameter. The situation where all parameters are equal to each other, is very unfavorable for our technique. The ideal case is very skew, the second case mentioned is symmetric ("unskew"). We want a measure that is informative on the skewness after ordering the parameters (say) for small to large. We could, of course, use the skewness measure $\gamma_1 = \mu_3/\sigma^3$, but this is not intuitively appealing. Cyhelský's skewness coefficient

$a = (\text{number of observations below the mean} - \text{number of observations above the mean})/\text{total number of observations}$, however, is intuitively appealing. But even more, just the number of observations above the mean is a good measure of "large parameters". We encountered Cyhelský's skewness coefficient on <http://en.wikipedia.org/wiki/Skewness>.

Appendix B. Proofs

PROOF 2.1.1.

First, for all $i \notin S$, all parameters with an index i in it, are zero, because $y_U - y_{U \setminus \{i\}}$ is the sum of all parameters containing index i , $y_U - y_{U \setminus \{i\}}$ is zero, and all parameters are non-negative, so all parameters with an index i in it, are zero.

Second, for all $i \in S$, the sum of all parameters containing index i , equals $y_U - y_\emptyset$, but this can only hold if only $\beta_S = y_U - y_\emptyset$.

PROOF 3.1.1.

(a) Suppose model (1) holds with only non-negative regression parameters. Then

$$\begin{aligned} \sum_{i=1}^m (y_U - y_{U \setminus \{i\}}) &= (y_U - y_{U \setminus \{1\}}) + (y_U - y_{U \setminus \{2\}}) + \dots + (y_U - y_{U \setminus \{m\}}) \geq \\ &\geq (y_U - y_{U \setminus \{1\}}) + (y_{U \setminus \{1\}} - y_{U \setminus \{1,2\}}) + \dots + (y_{\{m\}} - y_\emptyset) = y_U - y_\emptyset. \end{aligned}$$

This is because any $y_U - y_{U \setminus S} \geq y_{U \setminus T} - y_{U \setminus (S \cup T)}$. For the first term simply contains more (non-negative) regression parameters than the second term.

(b) Suppose $\sum_{i=1}^m (y_U - y_{U \setminus \{i\}}) \geq y_U - y_\emptyset$. Without loss of generality we may assume that $y_{U \setminus \{1\}} \leq y_{U \setminus \{2\}} \leq \dots \leq y_{U \setminus \{m\}}$, which is equivalent to $y_U - y_{U \setminus \{1\}} \geq y_U - y_{U \setminus \{2\}} \geq \dots \geq y_U - y_{U \setminus \{m\}}$. We consider $m - 1$ β -structures. In the first structure we take $\tilde{\beta}_i = y_U - y_{U \setminus \{i\}} - x$ ($i = 1, \dots, m$) and $\tilde{\beta}_{1,2,\dots,m} = x$ with $x \in [0, y_U - y_{U \setminus \{m\}}]$.

The sum of these regression parameters is $\sum_{i=1}^m (y_U - y_{U\setminus\{i\}}) - (m-1)x$. In the second structure, we take $\tilde{\beta}_i = y_{U\setminus\{m\}} - y_{U\setminus\{i\}} - x$ ($i = 1, \dots, m-1$), $\tilde{\beta}_{1, \dots, m-1} = x$ with $x \in [0, y_{U\setminus\{m\}} - y_{U\setminus\{m-1\}}]$, and $\tilde{\beta}_{1, 2, \dots, m} = y_U - y_{U\setminus\{m\}}$. And so on.

In the k^{th} structure ($k = 3, \dots, m-1$) we take $\tilde{\beta}_i = y_{U\setminus\{m-k+2\}} - y_{U\setminus\{i\}} - x$ ($i = 1, \dots, m-k+1$), $\tilde{\beta}_{1, \dots, m-k+1} = x$ with $x \in [0, y_{U\setminus\{m-k+2\}} - y_{U\setminus\{m-k+1\}}]$, $\tilde{\beta}_{1, \dots, m-k+2} = y_{U\setminus\{m-k+3\}} - y_{U\setminus\{m-k+2\}}$, \dots , $\tilde{\beta}_{1, 2, \dots, m} = y_U - y_{U\setminus\{m\}}$. In this way, in all structures and for all x , all $y_U - y_{U\setminus\{i\}}$ are the sum of the $\tilde{\beta}$ parameters containing index i ($i = 1, \dots, m$), and the sum of all parameters starts at $\sum_{i=1}^m (y_U - y_{U\setminus\{i\}})$ in the first structure, $x = 0$; it is a continuous function, and with the value $y_U - y_{U\setminus\{1\}}$ in structure $m-1$, $x = y_{U\setminus\{m-k+2\}} - y_{U\setminus\{m-k+1\}}$. Because of the fact that $\sum_{i=1}^m (y_U - y_{U\setminus\{i\}}) \geq y_U - y_{\emptyset}$ and $y_U - y_{U\setminus\{1\}} \leq y_U - y_{\emptyset}$, model (1) holds with only non-negative regression parameters.

PROOF 3.1.2.

Since $\ell > k$, there exists a non-trivial ℓ -vector t such that $Xt = 0$. Since X contains only non-negative elements (0's and 1's), t must contain positive as well as negative elements. Then there exists a λ such that $\beta + \lambda t$ contains only non-negative elements and (at least) one element equal to zero. This element can be removed from β , and the corresponding column of X can be deleted. This can be repeated until we have k columns left.

Any k -vector β that suffices to explain the k -vector $y_S - y_{\emptyset}$ is called a basic solution of order k . It is characterized by its indices and its values.

References

- BETTONVIL, B. (1990), *Detection of important factors by sequential bifurcation*. Tilburg University Press.
- BETTONVIL, B. (1995), "Factor screening by sequential bifurcation", *Communications in Statistics - Simulation and Computation* **24**(1) 165-185.
- BOX, G.E.P., and K.B. WILSON (1951), "On the experimental attainment of optimum conditions", *Journal of the Royal Statistical Society, Series B* **13** 1-38.
- HOMMA, T., and A. SALTELLI (1996), "Importance measures in global sensitivity analysis of non linear models", *Reliability Engineering and System Safety* **52** 1-17.
- LI, C.H. (1962), "A sequential method for screening experimental variables", *American Statistical Association Journal* **57** 455-477.
- MORRIS, M.D. (1991), "Factorial sampling plans for preliminary computational experiments", *Technometrics* **33**(2) 161-174.
- MORRIS, M.D. (2006), "An overview of group factor screening", in: A. Dean and S. Lewis. *Screening methods for experimentation in industry, drug discovery, and genetics*. Springer. 191-206.
- PATEL, M.S. (1962), "Group-screening in more than two stages", *Technometrics* **4** 209-217.
- WEBB, S. (2010), "Screening experiments with many factors", Submitted for publication in: *Communications in Statistics - Theory and Methods*.
- WATSON, G.S. (1961), A study of the group screening method. *Technometrics* **3** 371-388.