Sequential bifurcation

Bettonvil, B.W.M.

Publication date:
1989

RSITY
UNIVERSITEIT
BRABANT

POSTBOX 90153
5000 LE TILBURG
THE NETHERLANDS

[redacted]

DEPARTMENT OF ECONOMICS

RESEARCH MEMORANDUM

[redacted]

SEQUENTIAL BIFURCATION: THE DESIGN OF A FACTOR SCREENING METHOD

Bert Bettonvil

FEW 381

march 1989

SEQUENTIAL BIFURCATION:

THE DESIGN OF A FACTOR SCREENING METHOD

Bert Bettonvil

Tilburg University/Eindhoven University of Technology

P O Box 90153

5000 LE Tilburg

Netherlands

We reconsider a method for factor screening called sequential bifurcation, which resembles binary search. Sequential bifurcation can be used in case a reponse can be represented by a model, additive in the input variables with known signs of the regression parameters. We introduce a new experimental design and compare our method to two older versions of sequential bifurcation and to other factor screening techniques. This comparison turns out to be in favour of our new design.

Keywords: Experimental Design, Screening, Aggregated Variables, Binary Search, Simulation.

1. INTRODUCTION

Suppose we are dealing with a problem in which a great many (100, 1000, 10000?) input variables play a role, but we think that just a few are really important. A straightforward screening method would use at least as many observations as there are input variables to be inspected. But an observation can be so time-consuming, that collecting so many data is prohibitive.

We assume a first order linear regression model:

$$y(x_1,x_2,\ldots,x_N) = \beta_0+\beta_1x_1+\beta_2x_2+\ldots+\beta_Nx_N, \tag{1.1}$$

in which we assume that we know the signs of $\beta_1,\ldots,\beta_N$.

The problem described often arises in simulation. Then (1.1) is an auxiliary or meta-model; see Kleijnen (1987). If, for example, x_1 represents number of servers, and y denotes expected waiting time, then $\beta_1\leq 0$.

Without loss of generality, we may take $\beta_\ell\geq 0$ ($\ell=1,\ldots,N$); this can always be achieved by reparametrization. To estimate a first-order model, a two-level experiment suffices, so we may take (reparametrization again) $x_\ell\in\{0,1\}$ for $\ell=1,\ldots,N$. In this paper we assume, for ease of survey, that we have negligible error, and that N is equal to some power of two: $N=2^m$ for some $m\in\mathbb{N}$. We want to find all "important" factors, calling a factor important, iff its regression parameter is "large". A regression parameter is called large, iff it is larger than some given number $\delta\geq 0$.

Our search routine should require relatively few observations, if the number of important regression parameters is small indeed, even if the total number of regression parameters is great. To this end we propose a modification of Jacoby and Harrison (1962)'s Sequential Bifurcation (see §4 of this paper). In two previous papers, Bettonvil (1988a,b), we used a different (and, as we will see, an inferior) design.

We will start with a global description and a small example in §2; in §3 we will give a formal description of the technique, which we will compare to our own former design in §4. The number of observations we need for our version of Sequential Bifurcation ("SB" in the sequel) and for related techniques is treated in §5. A discussion is given in §6.

2. GLOBAL DESCRIPTION

Our method exploits the assumption that, apart from β_0, all regression parameters are non-negative. This implies that the sequence β_0, $\beta_0+\beta_1$, $\beta_0+\beta_1+\beta_2$, ... , $\beta_0+\beta_1+\beta_2+...+\beta_{N-1}$, $\beta_0+\beta_1+\beta_2+...+\beta_{N-1}+\beta_N$ is non-decreasing. We introduce the following notation (using the symbol ":=" for "is defined as"):

$$\beta_i^+ := \Sigma_{j=0}^{i} \beta_j \quad (i=0,1,...,N), \tag{2.1}$$

so that the sequence $\beta_0^+,\beta_1^+,\beta_2^+,...,\beta_{N-1}^+,\beta_N^+$ is non-decreasing. We restrict ourselves to observations y_i, which have the first i input variables at their high levels, and the remaining N-i input variables at their low levels (i=0,1,2,...,N-1,N), so

$$y_i = \beta_i^+ \quad \text{for all observations } y_i. \tag{2.2}$$

We distinguish between y_i and β_i^+ to stress the difference between "potential" observations β_i^+ and "actual" observations y_i, because, except in pathological cases, we do not need all N+1 possible observations.

We start by observing y_0 and y_N (all inputs "off"; all inputs "on" respectively). Because of the assumption $\beta_\ell \geq 0$ (ℓ=1,2,...,N), we have $y_0 \leq y_N$. If $y_N - y_0$ $(=\beta_1+\beta_2+\ldots+\beta_N) \leq \delta$, then none of the effects is important: after only two observations the problem is solved! (We are aware that this situation will hardly occur in practice.) If $y_N - y_0 > \delta$, then one or more effects may be important (unless δ=0, a number of small parameters may add to a large sum). We then proceed with observation $y_{N/2}$ $(=y_{2^{m-1}}$; that is, we switch on the first half of the $N=2^m$ factors) and compute

$$y_{N/2} - y_0 = \beta_1+\beta_2+\ldots+\beta_{N/2} \tag{2.3}$$

and

$$y_N - y_{N/2} = \beta_{(N/2)+1}+\beta_{(N/2)+2}+\ldots+\beta_N. \tag{2.4}$$

If $y_{N/2} - y_0 \leq \delta$, then all parameters $\beta_1, \beta_2, \ldots, \beta_{N/2}$ are small $(x_1, x_2, \ldots, x_{N/2}$ are unimportant); if $y_{N/2} - y_0 > \delta$, then this set of parameters has to be investigated further (with the aid of observation $y_{N/4}$). In the same way, if $y_N - y_{N/2} \leq \delta$, then all parameters $\beta_{N/2+1}, \beta_{N/2+2}, \ldots, \beta_N$ are small; if $y_N - y_{N/2} > \delta$, we have to observe $y_{3N/4}$, and look at $y_{3N/4} - y_{N/2}$ and $y_N - y_{3N/4}$, and so on.

Example 2.1. For illustration purposes we take N=8 (so m=3) and δ=0. Suppose that $\beta_2>0$, $\beta_3>0$, and $\beta_1=\beta_4=\beta_5=\beta_6=\beta_7=\beta_8=0$. How will SB find this out?

Note that $\beta_0^+=\beta_1^+<\beta_2^+<\beta_3^+=\beta_4^+=\beta_5^+=\beta_6^+=\beta_7^+=\beta_8^+$. *We start by observing* y_0 *and* y_8. *We find that* $y_0<y_8$ *(as* $\beta_0^+<\beta_8^+$*), so at least one parameter must be positive, and we proceed by observing* y_4, *that is, the first half of the input variables is switched "on", the other half is switched "off". We find that* $y_0<y_4=y_8$. *Then* $\beta_5+\beta_6+\beta_7+\beta_8=y_8-y_4=0$, *so* $\beta_5=\beta_6=\beta_7=\beta_8=0$; *and* $\beta_1+\beta_2+\beta_3+\beta_4=y_4-y_0>0$, *so at least one of the latter four parameters is positive. Next, we observe* y_2. *We find that* $y_0<y_2<y_4$ *(as* $\beta_0^+<\beta_2^+<\beta_4^+$*). Then* $\beta_1+\beta_2=y_2-y_0>0$, *and* $\beta_3+\beta_4=y_4-y_2>0$. *To find out whether* β_1 *or* β_2 *or both are positive, we observe* y_1 *and find that* $y_0=y_1<y_2$, *so that* $\beta_1=0$ *and* $\beta_2>0$. *Finally, to get* β_3 *and* β_4, *we observe* y_3 *and find that* $y_2<y_3=y_4$, *implying that* $\beta_3>0$ *and* $\beta_4=0$. *Summarizing, for N=8 factors we need only 6 observations* (y_0,y_1,y_2,y_3,y_4 *and* y_8) *to identify which two individual factors are positive, and to measure their individual effects.*

3. FORMAL DESCRIPTION OF SB

The general procedure is such that we always start by observing y_0 and $y_N=y_{2^m}$. Then we know the sum $\beta_1+\beta_2+\ldots+\beta_N$. In general, if we know y_{i_1} and y_{i_2}, with $i_1<i_2$, then we know the sum $\beta_{i_1+1}+\beta_{i_1+2}+\ldots+\beta_{i_2}$ $(=y_{i_2}-y_{i_1})$. If this sum does not exceed δ, we do not investigate this group of parameters any further, since none of its components can exceed δ. On the other hand, if $y_{i_2}-y_{i_1}$ is larger than δ and $i_2-i_1=1$, then we have found that $\beta_{i_2}=y_{i_2}-y_{i_2-1}$ is large, so that factor i_2 is important. If $y_{i_2}-y_{i_1}$ exceeds δ and $i_2-i_1>1$, we split up the group: we observe y_{i_3}, with $i_3=(i_1+i_2)/2$ (since

$N=2^m$, i_1+i_2 is even) and we examine $y_{i_3}-y_{i_1} = \beta_{i_1+1}+\beta_{i_1+2}+\ldots+\beta_{i_3}$ and $y_{i_2}-y_{i_3} = \beta_{i_3+1}+\beta_{i_3+2}+\ldots+\beta_{i_2}$ in the same way as we did $y_{i_2}-y_{i_1}$.

So we start by considering all 2^m parameters simultaneously, and in the next steps (or stages) of the procedure, we split a group of parameters into two equally sized groups. First we have one single group of size $N=2^m$, next two groups of size 2^{m-1}, then four groups of size $2^{m-2}, \ldots, 2^j$ groups of size $2^{m-j}, \ldots,$ and finally $N=2^m$ groups of size $2^{m-m}=1$. If the sum of the parameters in any group is smaller than δ, then all parameters in this group are smaller than δ. For the sum of the parameters within the k^{th} group of 2^j groups (of size 2^{m-j}) we now introduce the symbol $\beta_{k|j}$, which we call the k^{th} parameter at stage j, or the accumulated effect of group k at stage j.

Example 3.1. Suppose m=3, so we have eight input variables. At stage 0 we have one group of size 8, and $\beta_{1|0}$ *is the sum of* β_1 *through* β_8*. At stage 1 we have two groups of size four each;* $\beta_{1|1}=\beta_1+\beta_2+\beta_3+\beta_4$; $\beta_{2|1}=\beta_5+\beta_6+\beta_7+\beta_8$*. In the same way we have* $\beta_{1|2}=\beta_1+\beta_2, \ldots, \beta_{4|2}=\beta_7+\beta_8$ *and* $\beta_{1|3}=\beta_1, \ldots,$ $\beta_{8|3}=\beta_8$.

At stage j we have 2^j groups of size 2^{m-j}. The first group runs from input variable 1 through 2^{m-j}, the second from $2^{m-j}+1$ through $2*2^{m-j}$, the k^{th} group from $(k-1)*2^{m-j}+1$ through $k*2^{m-j}$. Formally we have

DEFINITION 3.1:

$$\beta_{k|j} := \sum_{\ell=(k-1)*2^{m-j}+1}^{k*2^{m-j}} \beta_\ell \qquad (j=0,1,\ldots,m;\ k=1,2,..2^j).$$

The size of $\beta_{k|j}$ is found as the difference of two observations: $\beta_{k|j}=y_{i_2}-y_{i_1}$, where $i_2=k*2^{m-j}$ and $i_1=(k-1)*2^{m-j}$. As $y_i=\beta_i^+$ for all observations, we have

$$\beta_{k|j}=\beta^+_{k*2^{m-j}}-\beta^+_{(k-1)*2^{m-j}}. \tag{3.1}$$

Note that if j=m, we have $\beta_{k|m}=\beta_k=\beta_k^+-\beta_{k-1}^+$. By the introduction of $\beta_{k|j}(\geq 0)$ we can generalize the proporty that the sequence $\Sigma_{j=0}^{i}\beta_j$ (i=0,...,N) is non-decreasing, to all stages. The generalization of β_i^+ is then given by

DEFINITION 3.2:

$$\beta^+_{i|j} := \beta_0 + \Sigma_{k=1}^{i}\ \beta_{k|j} \qquad (j=0,\dots,m;\ i=0,\dots,2^j)$$

(in which we use the convention that empty sums ($\Sigma_{k=1}^{0}$) are zero). We find that for each j ($0\leq j\leq m$) the sequence $\beta^+_{0|j}$, $\beta^+_{1|j}$, ..., $\beta^+_{2^j|j}$ is non-decreasing. Now (3.1) reduces to

$$\beta_{k|j} = \beta^+_{k|j} - \beta^+_{k-1|j} \qquad (j=0,\dots,m;\ k=1,\dots,2^j). \tag{3.2}$$

We start the procedure at stage 0 by observing y_0 and y_N and computing $\beta_{1|0}$, the sum of all effect parameters (see definition 3.1). Not only do we call $\beta_{1|0}$ the parameter at stage 0; we also call y_0 and y_N the observations at stage 0. If $\beta_{1|0}>\delta$, then we proceed with observation $y_{N/2}$: the observation at stage 1. We compute $\beta_{1|1}=y_{N/2}-y_0$ and $\beta_{2|1}=y_N-y_{N/2}$: the parameters at stage 1. To denote the stage at which an observation is made, we use a notation analogous to definition 3.1:

DEFINITION 3.3:

$y_{i|j}$ is an observation at stage j; to obtain $y_{i|j}$ we take the first $i*2^{m-j}$ input variables at their high levels, and the remaining input variables at their low levels.

This definition implies that $y_{i|j}=y_{i*2^{m-j}}=\beta^+_{i*2^{m-j}}=\beta^+_{i|j}$. However, whereas $y_{i*2^{m-j}}=y_{2i*2^{m-j-1}}$ (as $i*2^{m-j} = 2i*2^{m-j-1}$), it is not true that $y_{i|j}=y_{2i|j+1}$, since in our design an observation at level j can not be an observation at level j+1. It is true that $\beta^+_{i|j}=\beta^+_{2i|j+1}(=\beta^+_{i*2^{m-j}})$.

The observations the procedure always starts with, have level 0 and are called $y_{1|0}(=y_N)$ and $y_{0|0}(=y_0)$. These observations yield $\beta_{1|0}$. If $\beta_{1|0}\le\delta$, then we stop. Otherwise (i.e., if $\beta_{1|0}>\delta$) we observe $y_{1|1}=y_{N/2}$. In general, if we have some $\beta_{i|j}\le\delta$, then we stop investigating this branch of the bifurcation tree; if $\beta_{i|m}>\delta$, we have found an important factor; if $\beta_{i|j}>\delta$ and $j<m$, then we proceed to an observation at level j+1. This observation should split $\beta_{i|j}$ into $\beta_{2i-1|j+1}$ and $\beta_{2i|j+1}$, so this observation must be $y_{2i-1|j+1}$! Note that the first index of y is odd. This makes the next lemma obvious.

LEMMA.

Observations y_0 and y_N can be written as $y_{0|0}$ and $y_{1|0}$ respectively; these are the observations at level 0. If observation y_ℓ with $0<\ell<N$ exists, then it can be uniquely written as $y_{i|j}$, where i and j are such that $\ell=i*2^{m-j}$ and i is odd.

If $j>0$, we cannot write $\beta_{i|j}=y_{i|j}-y_{i-1|j}$, since not both $y_{i|j}$ and $y_{i-1|j}$ can be observed in our method (either i or i-1 is even). However, it is true that $\beta_{i|j}=\beta^{+}_{i|j}-\beta^{+}_{i-1|j}$ (see definition 3.2), and there exist observations $y_{i_1|j_1}$ and $y_{i_2|j_2}$, such that $\beta_{i|j}=y_{i_1|j_1}-y_{i_2|j_2}$. To find these values of i_1, j_1, i_2 and j_2 for given i and j, we reformulate SB.

We start our search for the important factors by observing $y_{1|0}$ and $y_{0|0}$. As

$$y_{1|0} = \beta^{+}_{1|0} = \beta_0 + \beta_{1|0} = \beta_0+\beta_1+\beta_2+.....+\beta_N \tag{3.2}$$

and

$$y_{0|0} = \beta^{+}_{0|0} = \beta_0, \tag{3.3}$$

we have $y_{1|0}-y_{0|0}=\beta_{1|0}$, which is the sum of all factor effects. If $\beta_{1|0}\leq\delta$, then none of the parameters can be larger than δ, as all of them are non-negative. In this case the procedure terminates after only two observations. In case $\beta_{1|0}=\beta^{+}_{1|0}-\beta^{+}_{0|0}=y_{1|0}-y_{0|0}>\delta$, the sum of some small regression parameters may still yield a large value (small and large relative to δ): false signal. On the other hand, we may have one or more large β's. If $\beta_{1|0}>\delta$ and $\delta=0$, we surely have one or more large (now meaning positive) β's. Anyway, if $\beta_{1|0}>\delta$, we proceed to the next observation, namely

$$y_{1|1} = \beta^{+}_{1|1} = \beta_0 + \beta_{1|1}. \tag{3.4}$$

Now (3.3) and (3.4) give

$$\beta_{1|1} = \beta^{+}_{1|1}-\beta^{+}_{0|1} = \beta^{+}_{1|1}-\beta^{+}_{0|0} = y_{1|1} - y_{0|0} \tag{3.5}$$

and (3.2) and (3.4) yield

$$\beta_{2|1} = \beta^{+}_{2|1}-\beta^{+}_{1|1} = \beta^{+}_{1|0}-\beta^{+}_{1|1} = y_{1|0} - y_{1|1}. \tag{3.6}$$

Again, we analyze $\beta_{1|1}$ and $\beta_{2|1}$ just as we did $\beta_{1|0}$, that is, if $\beta_{1|1}\leq\delta$, then all its "component" parameters are small; and if $\beta_{2|1}\leq\delta$, all parameters within it are small. If $\beta_{1|1}>\delta$, we proceed to $y_{1|2}$; and if $\beta_{2|1}>\delta$, we proceed to $y_{3|2}$; so in the second stage (j=2) we have:

$$\beta_{1|2} = \beta^{+}_{1|2} - \beta^{+}_{0|2} = \beta^{+}_{1|2} - \beta^{+}_{0|0} = y_{1|2} - y_{0|0} \tag{3.7}$$
$$\beta_{2|2} = \beta^{+}_{2|2} - \beta^{+}_{1|2} = \beta^{+}_{1|1} - \beta^{+}_{1|2} = y_{1|1} - y_{1|2} \tag{3.8}$$
$$\beta_{3|2} = \beta^{+}_{3|2} - \beta^{+}_{2|2} = \beta^{+}_{3|2} - \beta^{+}_{1|1} = y_{3|2} - y_{1|1} \tag{3.9}$$
$$\beta_{4|2} = \beta^{+}_{4|2} - \beta^{+}_{3|2} = \beta^{+}_{1|0} - \beta^{+}_{3|2} = y_{1|0} - y_{3|2}. \tag{3.10}$$

Every $\beta_{i|2}$ (i=1,2,3,4) is first expressed as $\beta^{+}_{i|2}-\beta^{+}_{i-1|2}$, but can not be expressed as $y_{i|2}-y_{i-1|2}$, because we have only the observations $y_{1|0}$, $y_{0|0}$, $y_{1|1}$, $y_{1|2}$ and $y_{3|2}$. That is why we first use $\beta^{+}_{0|2}=\beta^{+}_{0|0}$, $\beta^{+}_{2|2}=\beta^{+}_{1|1}$ and $\beta^{+}_{4|2}=\beta^{+}_{1|0}$.

In general, if we have $\beta_{i|j}=\beta^{+}_{i|j}-\beta^{+}_{i-1|j}>\delta$ with $j<m$, then we observe $y_{2i-1|j+1}=\beta^{+}_{2i-1|j+1}$ and we can compute $\beta_{2i-1|j+1} = \beta^{+}_{2i-1|j+1}-\beta^{+}_{2i-2|j+1} = \beta^{+}_{2i-1|j+1}-\beta^{+}_{i-1|j}$ and $\beta_{2i|j+1} = \beta^{+}_{2i|j+1}-\beta^{+}_{2i-1|j+1} = \beta^{+}_{i|j}-\beta^{+}_{2i-1|j+1}$. If we have $\beta_{i|m}>\delta$, we have found an important regression parameter. If we have some $\beta_{i|j}\leq\delta$, we conclude that all regression parameters within $\beta_{i|j}$ are small.

Example 3.1. Our second example originates from Jacoby and Harrison (1962). We have $128=2^7$ variables, where only the variables numbered 68, 113 *and 120 have positive effects. We want to find these positive effects, so we take* $\delta=0$. *Our procedure starts by observing* $y_{1|0}$ *and* $y_{0|0}$, *resulting in* $\beta_{1|0}>0$. *Next we observe* $y_{1|1}$, *which gives* $\beta_{1|1}=\beta^{+}_{1|1}-\beta^{+}_{0|1}=\beta^{+}_{1|1}-\beta^{+}_{0|0}=y_{1|1}-y_{1|0}=0$, *so* $\beta_1=\beta_2=\ldots=\beta_{64}=0$; *and* $\beta_{2|1}=\beta^{+}_{2|1}-\beta^{+}_{1|1}=\beta^{+}_{1|0}-\beta^{+}_{1|1}=y_{1|0}-y_{1|1}>0$. *So next we have to observe* $y_{3|2}$.

And so on. See figure 1, where the positive parameters are indicated by $\underset{\sim}{\beta}_{i|j}$. *Figure 1 shows that we need only 16 observations to compute all 128 regression parameters.*

****

Figure 1 about here

****

4. COMPARISON TO AN OTHER DESIGN

In two previous papers (Bettonvil 1988a,b) we introduced SB using another design, which we demonstate with the aid of an example.

Example 4.1. In example 2.1 we had 8 input variables with $\beta_2>0$, $\beta_3>0$, *and* $\beta_1=\beta_4=\beta_5=\beta_6=\beta_7=\beta_8=0$. *SB used the observations* y_0, y_1, y_2, y_3, y_4 *and* y_8. *Bettonvil (1988a,b) would also use* $y_0=y(0,0,0,0,0,0,0,0)$ *and*

$y_8=y(1,1,1,1,1,1,1,1)$, but would then proceed with $y(0,0,0,0,1,1,1,1)$ to compute

$$\beta_{2|1}=y(0,0,0,0,1,1,1,1)-y(0,0,0,0,0,0,0,0)$$

and

$$\beta_{1|1}=\beta_{1|0}-\beta_{2|1}=y(1,1,1,1,1,1,1,1)-y(0,0,0,0,1,1,1,1).$$

Note that there is no essential difference between the use of $y_4=y(1,1,1,1,0,0,0,0)$ and $y(0,0,0,0,1,1,1,1)$; the difference arises in the next stage.

After finding out that $\beta_{1|1}>0$, $\beta_{2|1}=0$, we investigate $\beta_{1|2}$ and $\beta_{2|2}$. In the old design we would use $y(0,0,1,1,0,0,0,0)$ and compute

$$\beta_{2|2}=y(0,0,1,1,0,0,0,0)-y(0,0,0,0,0,0,0,0),$$

and $\beta_{1|2}$ as the difference of $\beta_{1|1}$ and $\beta_{2|2}$:

$$\beta_{1|2}=\beta_{1|1}-\beta_{2|2}= y(1,1,1,1,1,1,1,1)-y(0,0,0,0,1,1,1,1)$$
$$-y(0,0,1,1,0,0,0,0)+y(0,0,0,0,0,0,0.0).$$

We find out that $\beta_{2|2}>0$ and $\beta_{1|2}>0$. We investigate $\beta_{3|3}=\beta_3$ and $\beta_{4|3}=\beta_4$ with the aid of $y(0,0,0,1,0,0,0,0)$, and $\beta_{1|3}=\beta_1$ and $\beta_{2|3}=\beta_2$ with the aid of $y(0,1,0,0,0,0,0,0)$, the latter giving

$$\beta_{2|3}=y(0,1,0,0,0,0,0,0)-y(0,0,0,0,0,0,0,0),$$

and, because $\beta_{1|2}=\beta_{1|3}+\beta_{2|3}$, we have

$$\beta_{1|3}=\beta_{1|2}-\beta_{2|3}= y(1,1,1,1,1,1,1,1)-y(0,0,0,0,1,1,1,1)$$
$$-y(0,0,1,1,0,0,0,0)-y(0,1,0,0,0,0,0,0)+2y(0,0,0,0,0,0,0.0).$$

We see two things: the old design uses as many observations as the new design does (namely six); and the computation of β_1 is relatively complicated.

In general the old design works as follows: observe $y_0=y_{0|0}$ and $y_N=y_{1|0}$; compute $\beta_{1|0}=y_{1|0}-y_{0|0}$; suppose some $\beta_{i|j}>\delta$, then split the 2^m input variables into 2^{j+1} equally sized groups and observe the y for which the input variables in the $(2i)^{th}$ group are at their high levels, all other input variables are at their low levels. The difference between this observation and y_0 is $\beta_{2i|j+1}$; the difference between $\beta_{i|j}$ and $\beta_{2i|j+1}$ is $\beta_{2i-1|j+1}$. (Note. Jacoby and Harrison (1962) start the original Sequential Bifurcation by three observations: one with all input variables at their low levels, one with only the first half of the input varibles at their high levels, and one with only the second half at their high levels. If some $\beta_{i|j}>\delta$, they take two observations in the next stage: one with the input variables in group 2i-1 at their high levels, and one with those in group 2i at their high levels; for both observations all other variables are at their low levels. Trivially, this method uses about twice as much observations as our method.)

It is easely verified (e.g. by means of induction) that, in our old design, $\beta_{1|j}$ is computed as $y_N+(j-1)*y_0$-(the sum of j other observations). (Note. If at all stages we would take the variables in group 2i-1 at their high levels instead of those in group 2i, this computation would hold for $y_{2^j|j}$ instead of for $y_{1|j}$ because of symmetry.) We see that the computation of $\beta_1=\beta_{1|m}=y_N+(m-1)*y_0$-(the sum of m other observations) becomes awkward, especially if m is great; the case we are dealing with. A further drawback arises with the introduction of random error: assuming, instead of (1.1) that our model is

$$y(x_1,x_2,\dots,x_N,e) = \beta_0+\beta_1x_1+\beta_2x_2+\dots+\beta_Nx_N+e \tag{4.1}$$

where $e \sim N(0,\sigma^2)$ and the errors of all observations are independent, the estimator of $\beta_{1|m}$ would have a variance of $(1+(m-1)^2+m)\sigma^2 = (m^2-m+2)\sigma^2$, whereas in our new design all estimators are computed as the difference of two observations, yielding a variance of $2\sigma^2$!

Computational simplicity and, consequently, small variances in case of random error, give us enough reason to prefer the new design over the old one.

5. COMPARISON TO OTHER SCREENING TECHNIQUES

The computation of the required number of observations in case $k(\geq 0)$ out of 2^m input variables are important (with $\delta=0$, i.e. k out of 2^m effect parameters are positive) for SB and some other screening techniques has already been given in Bettonvil(1988a,b). (The previous section showed that the old and the new design of SB require the same number of observations.) Here we restrict ourselves to a résumé. We compute the worst-case number of observations.

For SB we always have two observations at level 0, and if $k=0$, then these two observations suffice. At level $j(>0)$ we can never have more than 2^{j-1} observations (the number of parameters at stage $j-1$ cannot exceed 2^{j-1}), nor can we have more than k observations (the number of positive parameters at each stage cannot exceed k). So the number of observations is limited to

$$2 + \Sigma_{j=1}^{m} \min\{k, 2^{j-1}\}. \tag{5.1}$$

Be ℓ such that $2^{\ell-1}<k\leq 2^{\ell}$. Then for $j\leq\ell$ we have $\min\{k,2^{j-1}\}=2^{j-1}$, and for $j>\ell$ we obtain $\min\{k,2^{j-1}\}=k$. So (5.1) can be replaced by

$$2 + \Sigma_{j=1}^{\ell} 2^{j-1} + \Sigma_{j=\ell+1}^{m} k = 1 + 2^{\ell} + k(m-\ell). \tag{5.2}$$

This worst case is achieved when the k important input variables have the least possible clustering.

Table 1 shows the worst-case number of observations needed to find k=0,1,2,...,8 important input variables out of 2^{10}=1024 candidate variables; not only for our version of SB, but also for three other factor screening methods, namely Two-Stage Group-Screening (see Mauro, 1984 or Mauro and Burns, 1984), Multi-Stage Group-Screening(see Patel, 1962 or Li, 1962) and Jacoby and Harrison (1962)'s version of Sequential Bifurcation. The worst case number of runs is given in table 1, in which G2 stands for Two-stage group-screening, GM for Multi-stage group screening, JH for the Jacoby and Harrison (1962) Sequential Bifurcation, and SB for our version of Sequential Bifurcation.

****

* Table 1 about here *

****

For G2, JH and SB we assumed the most unfavourable situation, namely minimal clustering. For G2 we further assumed that for the first stage we guessed the number of non-zero coefficients correctly. For GM we applied an approximating (favourable) formula and rounded up to the next integer.

It is clear that Sequential Bifurcation is far superior.

6. DISCUSSION

The modified version of Sequential Bifurcation is a screening method that requires relatively few observations, compared to related techniques, and besides, it is simple to handle: all (accumulated) factor effects are computed as the difference of two observations. In this paper we restricted ourselves numbers of factors, equal to 2^m ($m \in \mathbb{N}$), and to observations without random error. Extensions of Sequential Bifurcation to cope with general numbers of factors, and with observations with random error, are being prepared at the moment.

REFERENCES

Bettonvil, B. (1988a), "Sequential Bifurcation for Factor Screening", in *Operations Research Proceedings 1987*, eds. H.Schellhaas, P.van Beek, H.Iserman, R.Schmidt and M.Zijlstra, Springer Verlag, Heidelberg, 444-450.

Bettonvil, B. (1988b), "Factor Screening by Sequential Bifurcation", submitted for publication.

Jacoby, J.E. and Harrison, S. (1962), "Multi-Variable Experimentation and Simulation Models", *Naval Research Logistic Quarterly*, 9, 121-136.

Kleijnen, J.P.C. (1987), *Statistical Tools for Simulation Practitioners*, New York: Marcel Dekker.

Li, C.H. (1962), "A Sequential Method for Screening Experimental Variables", *American Statistical Association Journal*, 57, 455-477.

Mauro, C.A. (1984), "On the Performance of Two-Stage Group Screening Experiments", *Technometrics*, 26, 255-264.

Mauro, C.A. and Burns, K.C. (1984), "A Comparison of Random Balance and Two-Stage Group Screening Designs: a Case Study", *Communications in Statistics: Theory and Methods*, 13, 2625-2647.

Patel, M.S. (1962), "Group-Screening with More Than Two Stages", *Technometrics*, 4, 209-217.

Table 1. Maximum number of runs for given number of non-zero variables.

k	0	1	2	3	4	5	6	7	8
G2	2	68	96	116	136	148	164	172	188
GM	2	19	34	48	61	73	84	95	106
JH	3	21	39	55	71	85	99	113	127
SB	2	12	21	29	37	44	51	58	65

$$y_{0|0} \rightarrow \underset{\sim}{\beta}_{1|0} \leftarrow y_{1|0}$$

$$\downarrow$$

$$\beta_{1|1} \leftarrow y_{1|1} \rightarrow \underset{\sim}{\beta}_{2|1}$$

$$\downarrow$$

$$\underset{\sim}{\beta}_{3|2} \quad \leftarrow \quad y_{3|2} \quad \rightarrow \quad \underset{\sim}{\beta}_{4|2}$$

$$\downarrow \qquad\qquad\qquad\qquad \downarrow$$

$$\underset{\sim}{\beta}_{5|3} \leftarrow y_{5|3} \rightarrow \beta_{6|3} \qquad\qquad \beta_{7|3} \leftarrow y_{7|3} \rightarrow \underset{\sim}{\beta}_{8|3}$$

$$\downarrow \qquad\qquad\qquad\qquad \downarrow$$

$$\underset{\sim}{\beta}_{9|4} \leftarrow y_{9|4} \rightarrow \beta_{10|4} \qquad\qquad \underset{\sim}{\beta}_{15|4} \leftarrow y_{15|4} \rightarrow \beta_{16|4}$$

$$\downarrow \qquad\qquad\qquad\qquad \downarrow$$

$$\underset{\sim}{\beta}_{17|5} \leftarrow y_{17|5} \rightarrow \beta_{18|5} \qquad\qquad \underset{\sim}{\beta}_{29|5} \quad \leftarrow \quad y_{29|5} \quad \rightarrow \quad \underset{\sim}{\beta}_{30|5}$$

$$\downarrow \qquad\qquad\qquad\qquad \downarrow \qquad\qquad \downarrow$$

$$\beta_{33|6} \leftarrow y_{33|6} \rightarrow \underset{\sim}{\beta}_{34|6} \qquad \underset{\sim}{\beta}_{57|6} \leftarrow y_{57|6} \rightarrow \beta_{58|6} \quad \beta_{59|6} \leftarrow y_{59|6} \rightarrow \underset{\sim}{\beta}_{60|6}$$

$$\downarrow \qquad\qquad \downarrow \qquad\qquad\qquad\qquad \downarrow$$

$$\beta_{67|7} \leftarrow y_{67|7} \rightarrow \underset{\sim}{\beta}_{68|7} \quad \underset{\sim}{\beta}_{113|7} \leftarrow y_{113|7} \rightarrow \beta_{114|7} \qquad\qquad \beta_{119|7} \leftarrow y_{119|7} \rightarrow \underset{\sim}{\beta}_{120|7}$$

Figure 1. Example of error-free sequential bifurcation.

IN 1988 REEDS VERSCHENEN

297 Bert Bettonvil
Factor screening by sequential bifurcation

298 Robert P. Gilles
On perfect competition in an economy with a coalitional structure

299 Willem Selen, Ruud M. Heuts
Capacitated Lot-Size Production Planning in Process Industry

300 J. Kriens, J.Th. van Lieshout
Notes on the Markowitz portfolio selection method

301 Bert Bettonvil, Jack P.C. Kleijnen
Measurement scales and resolution IV designs: a note

302 Theo Nijman, Marno Verbeek
Estimation of time dependent parameters in lineair models using cross sections, panels or both

303 Raymond H.J.M. Gradus
A differential game between government and firms: a non-cooperative approach

304 Leo W.G. Strijbosch, Ronald J.M.M. Does
Comparison of bias-reducing methods for estimating the parameter in dilution series

305 Drs. W.J. Reijnders, Drs. W.F. Verstappen
Strategische bespiegelingen betreffende het Nederlandse kwaliteits-concept

306 J.P.C. Kleijnen, J. Kriens, H. Timmermans and H. Van den Wildenberg
Regression sampling in statistical auditing

307 Isolde Woittiez, Arie Kapteyn
A Model of Job Choice, Labour Supply and Wages

308 Jack P.C. Kleijnen
Simulation and optimization in production planning: A case study

309 Robert P. Gilles and Pieter H.M. Ruys
Relational constraints in coalition formation

310 Drs. H. Leo Theuns
Determinanten van de vraag naar vakantiereizen: een verkenning van materiële en immateriële factoren

311 Peter M. Kort
Dynamic Firm Behaviour within an Uncertain Environment

312 J.P.C. Blanc
A numerical approach to cyclic-service queueing models

313 Drs. N.J. de Beer, Drs. A.M. van Nunen, Drs. M.O. Nijkamp
Does Morkmon Matter?

314 Th. van de Klundert
Wage differentials and employment in a two-sector model with a dual labour market

315 Aart de Zeeuw, Fons Groot, Cees Withagen
On Credible Optimal Tax Rate Policies

316 Christian B. Mulder
Wage moderating effects of corporatism
Decentralized versus centralized wage setting in a union, firm, government context

317 Jörg Glombowski, Michael Krüger
A short-period Goodwin growth cycle

318 Theo Nijman, Marno Verbeek, Arthur van Soest
The optimal design of rotating panels in a simple analysis of variance model

319 Drs. S.V. Hannema, Drs. P.A.M. Versteijne
De toepassing en toekomst van public private partnership's bij de grote en middelgrote Nederlandse gemeenten

320 Th. van de Klundert
Wage Rigidity, Capital Accumulation and Unemployment in a Small Open Economy

321 M.H.C. Paardekooper
An upper and a lower bound for the distance of a manifold to a nearby point

322 Th. ten Raa, F. van der Ploeg
A statistical approach to the problem of negatives in input-output analysis

323 P. Kooreman
Household Labor Force Participation as a Cooperative Game; an Empirical Model

324 A.B.T.M. van Schaik
Persistent Unemployment and Long Run Growth

325 Dr. F.W.M. Boekema, Drs. L.A.G. Oerlemans
De lokale produktiestructuur doorgelicht.
Bedrijfstakverkenningen ten behoeve van regionaal-economisch onderzoek

326 J.P.C. Kleijnen, J. Kriens, M.C.H.M. Lafleur, J.H.F. Pardoel
Sampling for quality inspection and correction: AOQL performance criteria

327 Theo E. Nijman, Mark F.J. Steel
Exclusion restrictions in instrumental variables equations

328 B.B. van der Genugten
Estimation in linear regression under the presence of heteroskedasticity of a completely unknown form

329 Raymond H.J.M. Gradus
The employment policy of government: to create jobs or to let them create?

330 Hans Kremers, Dolf Talman
Solving the nonlinear complementarity problem with lower and upper bounds

331 Antoon van den Elzen
Interpretation and generalization of the Lemke-Howson algorithm

332 Jack P.C. Kleijnen
Analyzing simulation experiments with common random numbers, part II: Rao's approach

333 Jacek Osiewalski
Posterior and Predictive Densities for Nonlinear Regression.
A Partly Linear Model Case

334 A.H. van den Elzen, A.J.J. Talman
A procedure for finding Nash equilibria in bi-matrix games

335 Arthur van Soest
Minimum wage rates and unemployment in The Netherlands

336 Arthur van Soest, Peter Kooreman, Arie Kapteyn
Coherent specification of demand systems with corner solutions and endogenous regimes

337 Dr. F.W.M. Boekema, Drs. L.A.G. Oerlemans
De lokale produktiestruktuur doorgelicht II. Bedrijfstakverkenningen ten behoeve van regionaal-economisch onderzoek. De zeescheepsnieuwbouwindustrie

338 Gerard J. van den Berg
Search behaviour, transitions to nonparticipation and the duration of unemployment

339 W.J.H. Groenendaal and J.W.A. Vingerhoets
The new cocoa-agreement analysed

340 Drs. F.G. van den Heuvel, Drs. M.P.H. de Vor
Kwantificering van ombuigen en bezuinigen op collectieve uitgaven 1977-1990

341 Pieter J.F.G. Meulendijks
An exercise in welfare economics (III)

342 W.J. Selen and R.M. Heuts
A modified priority index for Günther's lot-sizing heuristic under capacitated single stage production

343 Linda J. Mittermaier, Willem J. Selen, Jeri B. Waggoner, Wallace R. Wood
Accounting estimates as cost inputs to logistics models

344 Remy L. de Jong, Rashid I. Al Layla, Willem J. Selen
Alternative water management scenarios for Saudi Arabia

345 W.J. Selen and R.M. Heuts
Capacitated Single Stage Production Planning with Storage Constraints and Sequence-Dependent Setup Times

346 Peter Kort
The Flexible Accelerator Mechanism in a Financial Adjustment Cost Model

347 W.J. Reijnders en W.F. Verstappen
De toenemende importantie van het verticale marketing systeem

348 P.C. van Batenburg en J. Kriens
E.O.Q.L. - A revised and improved version of A.O.Q.L.

349 Drs. W.P.C. van den Nieuwenhof
Multinationalisatie en coördinatie
De internationale strategie van Nederlandse ondernemingen nader beschouwd

350 K.A. Bubshait, W.J. Selen
Estimation of the relationship between project attributes and the implementation of engineering management tools

351 M.P. Tummers, I. Woittiez
A simultaneous wage and labour supply model with hours restrictions

352 Marco Versteijne
Measuring the effectiveness of advertising in a positioning context with multi dimensional scaling techniques

353 Dr. F. Boekema, Drs. L. Oerlemans
Innovatie en stedelijke economische ontwikkeling

354 J.M. Schumacher
Discrete events: perspectives from system theory

355 F.C. Bussemaker, W.H. Haemers, R. Mathon and H.A. Wilbrink
A (49,16,3,6) strongly regular graph does not exist

356 Drs. J.C. Caanen
Tien jaar inflatieneutrale belastingheffing door middel van vermogensaftrek en voorraadaftrek: een kwantitatieve benadering

357 R.M. Heuts, M. Bronckers
A modified coordinated reorder procedure under aggregate investment and service constraints using optimal policy surfaces

358 B.B. van der Genugten
Linear time-invariant filters of infinite order for non-stationary processes

359 J.C. Engwerda
LQ-problem: the discrete-time time-varying case

360 Shan-Hwei Nienhuys-Cheng
Constraints in binary semantical networks

361 A.B.T.M. van Schaik
Interregional Propagation of Inflationary Shocks

362 F.C. Drost
How to define UMVU

363 Rommert J. Casimir
Infogame users manual
Rev 1.2 December 1988

364 M.H.C. Paardekooper
A quadratically convergent parallel Jacobi-process for diagonal dominant matrices with nondistinct eigenvalues

365 Robert P. Gilles, Pieter H.M. Ruys
Characterization of Economic Agents in Arbitrary Communication Structures

366 Harry H. Tigelaar
Informative sampling in a multivariate linear system disturbed by moving average noise

367 Jörg Glombowski
Cyclical interactions of politics and economics in an abstract capitalist economy

IN 1989 REEDS VERSCHENEN

368 Ed Nijssen, Will Reijnders
"Macht als strategisch en tactisch marketinginstrument binnen de distributieketen"

369 Raymond Gradus
Optimal dynamic taxation with respect to firms

370 Theo Nijman
The optimal choice of controls and pre-experimental observations

371 Robert P. Gilles, Pieter H.M. Ruys
Relational constraints in coalition formation

372 F.A. van der Duyn Schouten, S.G. Vanneste
Analysis and computation of (n,N)-strategies for maintenance of a two-component system

373 Drs. R. Hamers, Drs. P. Verstappen
Het company ranking model: a means for evaluating the competition

374 Rommert J. Casimir
Infogame Final Report

375 Christian B. Mulder
Efficient and inefficient institutional arrangements between governments and trade unions; an explanation of high unemployment, corporatism and union bashing

376 Marno Verbeek
On the estimation of a fixed effects model with selective nonresponse

377 J. Engwerda
Admissible target paths in economic models

378 Jack P.C. Kleijnen and Nabil Adams
Pseudorandom number generation on supercomputers

379 J.P.C. Blanc
The power-series algorithm applied to the shortest-queue model

380 Prof. Dr. Robert Bannink
Management's information needs and the definition of costs, with special regard to the cost of interest