Improving the Efficiency and Efficacy of Controlled Sequential Bifurcation for Simulation Factor Screening

Hong Wan

School of Industrial Engineering, Purdue University, West Lafayette, Indiana 47907,
hwan@purdue.edu

Bruce E. Ankenman, Barry L. Nelson

Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, Illinois 60208
{ankenman@northwestern.edu, nelsonb@northwestern.edu}

Controlled sequential bifurcation (CSB) is a factor-screening method for discrete-event simulations. It combines a multistage hypothesis testing procedure with the original sequential bifurcation procedure to control both the power for detecting important effects at each bifurcation step and the Type I error for each unimportant factor under heterogeneous variance conditions when a main-effects model applies. This paper improves the CSB procedure in two aspects. First, a new fully sequential hypothesis-testing procedure is introduced that greatly improves the efficiency of CSB. Moreover, this paper proposes CSB-X, a more general CSB procedure that has the same error control for screening main effects that CSB does, even when two-factor interactions are present. The performance of the new method is proven and compared with the original CSB procedure.

1. Introduction

Screening experiments are designed to investigate the controllable factors in an experiment aimed at eliminating the unimportant ones. Controlled sequential bifurcation (CSB) has been proposed as an effective and efficient factor-screening method for computer simulation experiments (Wan et al. 2006). CSB emphasizes situations that often occur in simulation experiments (but occur less frequently in physical experiments). These situations include having a large number of factors including, easy switching between factor-level combinations, and the use of common random numbers (CRN) to reduce the variances of estimated effects.

CSB extends the basic sequential bifurcation (SB) procedure (Bettonvil 1990, 1995; Bettonvil and Kleijnen 1997; Cheng 1997) to provide error control for random responses. By incorporating a multistage hypothesis-testing approach into SB, CSB is the first screening strategy to simultaneously control the Type I error for each factor and power for each bifurcation step under heterogeneous variance conditions. The methodology is easy to implement and is more efficient than traditional designs in many situations, making it attractive for a wide variety of simulation applications.

The error control property of CSB depends on the hypothesis testing procedure used at each step. Wan et al. (2006) proposes two hypothesis testing procedures that have the desired error control properties: the two-stage testing procedure takes N_0 observations at the first stage; if no conclusion can be reached, another $N - N_0$ observations will be taken. The fully sequential testing procedure takes one observation at a time after the first N_0 observations and stops as soon as a conclusion can be reached. Numerical evaluation shows that the fully sequential testing procedure is much more efficient, but it is only applicable for certain levels of Type I error and power. In this paper, we improve the efficiency of CSB by substantially generalizing the fully sequential testing procedure introduced in Wan et al. (2006). The original fully sequential testing procedure applies to only a very specific setting—when the required Type I error equals one minus the power, which is rarely encountered in practice; the new test presented here is general. When variances are large, the two-stage hypothesis testing procedure in Wan et al. (2006) may require many observations to achieve the desired error control. The extended fully sequential procedure introduced here can save as much as two-thirds of the

computational effort relative to the two-stage hypothesis testing procedure.

Wan et al. (2006) assume that the simulation output can be represented by a main-effects model

$$Z = \beta_0 + \sum_{i=1}^{K} \beta_i x_i + \varepsilon \tag{1}$$

with all $\beta_i \geq 0$, $0 < i \leq K$. The error term, ε, is assumed to be a $\mathrm{Nor}(0, \sigma^2(\mathbf{x}))$ random variable where $\sigma^2(\mathbf{x})$ is unknown and may depend on $\mathbf{x} = (x_1, x_2, \ldots, x_K)$, and Z is the simulation output response of interest. Unfortunately, it is not always realistic to assume that there are no interactions (and it is very unusual to know that there are none, as shown in §6). Furthermore, even when subject matter experts can specify the directions of the main effects, the signs of the interactions may still be unknown. As we will demonstrate in §§5 and 6, if interactions are present and we ignore them, the results given by the original CSB can be misleading even when the interactions themselves are not important. Kleijnen et al. (2006) proposed a fold-over design for SB in which a more general model that includes all two-factor interactions and quadratic terms is assumed:

$$Z = \beta_0 + \sum_{i=1}^{K} \beta_i x_i + \sum_{i=1}^{K} \sum_{j=i}^{K} \beta_{ij} x_i x_j + \varepsilon, \tag{2}$$

where $\beta_i \geq 0$, $0 < i \leq K$ is the main effect of factor i, and β_{ij} is the interaction between factors i and j. When $i = j$, β_{ii} represents the quadratic effect of factor i. The error term ε has the same properties as in the main-effects model. This paper shows the implementation of the fold-over design in CSB, termed CSB-X. The resulting CSB-X method retains the error control property of CSB in the presence of two-factor interactions. Even though the number of design points is doubled, the number of observations required is about the same, if not less, than CSB in many situations.

The goal of screening with CSB-X is still to identify the factors with important main effects, so the strategy is to include new design points that allow us to eliminate the bias in the main-effects estimators as a result of the second-order terms. The signs of the interactions and quadratic terms are irrelevant and do not influence the final result, so no prior knowledge about the interactions is needed. Because the method does not estimate the interactions themselves, it will miss those factors with small main effects but big interactions with other factors. For the purpose of screening, however, it is often reasonable to assume that only the main effects are of interest. Specifically, if one accepts the heredity property (Wu and Hamada 2000), then for an interaction to be important at least

one of its parent factors should be important; therefore, the pursuit of important interactions can focus on interactions that involve at least one of the important main effects after the screening experiment.

The rest of this paper is organized as follows: In §2, we give a thorough review of the CSB method and its performance guarantees. The extended fully sequential testing procedure is discussed in §3, and the potential savings of computational effort is demonstrated numerically. In §4, we discuss the new CSB-X procedure based on fold-over designs. The performance is proven, and the screening efficacy and effectiveness of the original CSB and CSB-X are compared. Further evaluation of CSB and CSB-X both employing the newer, more general, fully sequential testing procedure is provided in §5. A realistic example, first presented in Wan et al. (2006), is reevaluated with CSB-X in §6, and §7 concludes the paper.

2. Review of CSB

In this section, we review the CSB method proposed by Wan et al. (2006). Let α, γ, Δ_0, and Δ_1 be user-specified parameters with $0 < \alpha < 0.5$, $0.5 < \gamma < 1$, and $\Delta_1 > \Delta_0 > 0$. The primary objective of CSB is to divide the factors into two groups: those that are unimportant, which means $\beta_i \leq \Delta_0$, and those that are important, meaning $\beta_i \geq \Delta_0$. For those factors with effects $\leq \Delta_0$, we want to control the Type I error of declaring them important to be $\leq \alpha$; for effects that are $\geq \Delta_1$, we want to provide power for identifying them as important to be $\geq \gamma$. For those factors with effects between Δ_0 and Δ_1, the procedure should have reasonable power to identify them. Wan et al. (2006) propose a cost model that connects both the thresholds of importance (Δ_0 and Δ_1) and factor levels with the cost to change them to guarantee a fair comparison. More generally, these parameters should be determined by the goal and properties of the specific problem. Both the CSB procedure and the CSB-X procedure described in this paper are independent of how the factor levels and thresholds of importance are determined.

A formal description of CSB appears in Figure 1. In each step of this sequential procedure, the cumulative effect of a group of factors is tested for importance. The first step begins with all K factors of interest in a single group and tests that group's effect. If the group's effect is important (indicating that at least one factor in the group may have an important effect), then the group is split into two subgroups ("bifurcation"). The cumulative effects of these two subgroups are then tested in subsequent steps, and each subgroup is either classified as unimportant or split into two subgroups for further testing. As the experiment proceeds, the groups become smaller until eventually

Initialization: Create an empty LIFO queue for groups. Add the group $\{1, 2, \ldots, K\}$ to the LIFO queue.
While queue is not empty, do
  Remove: Remove a group from the queue.
  Test:
    Unimportant: If the group is unimportant, then classify all factors in the group as unimportant.
    Important (size = 1): If the group is important and of size 1, then classify the factor as important.
    Important (size > 1): If the group is important and size is greater than 1, then split it into two subgroups such that all factors in the first subgroup have smaller index than those in the second subgroup. Add each subgroup to the LIFO queue.
  End Test
End While

Figure 1   Structure of CSB

all factors that have not been classified as unimportant are tested individually. The method assumes a main-effects model with normal errors and all factor effects $\beta_i \geq 0$.

More specifically, suppose the group to be tested contains factors $\{k_1 + 1, \ k_1 + 2, \ldots, k_2\}$ with $k_1 < k_2$. The Test step in Figure 1 tests the following hypothesis to determine if the group might contain important factors:

$$H_0: \sum_{i=k_1+1}^{k_2} \beta_i \leq \Delta_0 \quad \text{versus} \quad H_1: \sum_{i=k_1+1}^{k_2} \beta_i > \Delta_0.$$

Let x_i represent the setting of factor i. *An experiment at level k* is defined as the following factor settings:

$$x_i(k) = \begin{cases} 1, & i = 1, 2, \ldots, k; \\ 0, & i = k+1, k+2, \ldots, K. \end{cases} \tag{3}$$

Thus, "level k" indicates an experiment at which factors $1, 2, \ldots, k$ are at their high settings and factors $k + 1, \ k + 2, \ldots, K$ are at their low settings. We use $Z(k)$ to represent the response at level k so that $Z(k) = \beta_0 + \sum_{i=1}^k \beta_i + \varepsilon(k)$, given model (1). Let $Z_l(k)$ denote the lth simulation replication of an experiment at level k, $0 \leq k \leq K$. Therefore,

$$D_l(k_1, k_2) = Z_l(k_2) - Z_l(k_1) \tag{4}$$

is an unbiased estimator of the group effect $\sum_{i=k_1+1}^{k_2} \beta_i$.

Define a *qualified* hypothesis test as a test that guarantees

• $\Pr\{$Declare a group effect $\sum_{i=k_1+1}^{k_2} \beta_i$ important $\big|$ $\sum_{i=k_1+1}^{k_2} \beta_i \leq \Delta_0\} \leq \alpha$, and

• $\Pr\{$Declare a group effect $\sum_{i=k_1+1}^{k_2} \beta_i$ important $\big|$ $\sum_{i=k_1+1}^{k_2} \beta_i \geq \Delta_1\} \geq \gamma$.

Given a qualified hypothesis testing procedure and assuming model (1) holds, we can prove the following two theorems for CSB (Wan et al. 2006).

Theorem 1. *CSB with a qualified testing procedure guarantees that*

$$\Pr\{Declare\ factor\ i\ important \mid \beta_i \leq \Delta_0\} \leq \alpha$$

for each factor i individually.

Theorem 2. *CSB with a qualified testing procedure guarantees that*

$$\Pr\Bigg\{Declare\ a\ group\ effect \\ \sum_{i=k_1+1}^{k_2} \beta_i\ important\ \Bigg|\ \sum_{i=k_1+1}^{k_2} \beta_i \geq \Delta_1\Bigg\} \geq \gamma$$

for each group tested.

The Type I error control, described in Theorem 1, holds for each factor individually. We will briefly discuss the experiment-wide Type I error control of CSB by evaluating the expected number of factors that are falsely classified as important, denoted $E[F_K]$, for two extreme cases. To simplify the analysis, we assume that there are $K = 2^L$ factors, where L is an integer, and all tests are independent. Then we have the following two theorems (Wan et al. 2006):

Theorem 3. *If model (1) holds and $\sum_{i=1}^K \beta_i \leq \Delta_0$, then CSB with a qualified testing procedure guarantees that $E[F_K] \leq \alpha$.*

Theorem 4. *If model (1) holds and $\beta_i \leq \Delta_0, i = 1, 2, \ldots, K$, but $\beta_i + \beta_j \geq \Delta_1$ for all $i \neq j$, then CSB with a qualified testing procedure guarantees that $E[F_K] \leq K\alpha$.*

Realistic problems should be between these two extreme cases, but closer to Theorem 3. Therefore, CSB provides strong control of the "false-positive" rate, regardless of the number of factors.

Wan et al. (2006) proposed two qualified tests: a two-stage testing procedure and a fully sequential testing procedure for a very special case ($\alpha = 1 - \gamma$). The two-stage testing procedure takes a small number of observations in the first stage. If no conclusion can be reached, more samples are collected in the second stage. The sample size depends on the estimated variances in the first stage. The fully sequential testing procedure also has a first stage with a small number of observations, but then it takes one observation at a time until a conclusion can be reached. The fully sequential testing procedure is usually more efficient than the two-stage test when applicable.

In summary, CSB is a sequential screening method for stochastic simulation experiments. Given an appropriate hypothesis testing procedure, the method controls the Type I error for each factor and power for each bifurcation step. The sequential property of the method makes it well suited for simulation experiments. Examples show that CSB is highly efficient

for large-scale problems when there are only a few important factors and they are clustered, because CSB can eliminate unimportant factors in groups (assuming $\beta_i \geq 0$ for all i).

3. Extended Fully Sequential Testing Procedure

A key feature of CSB is that it provides strong statistical inference (Theorems 1–4) relative to other factor-screening methods. There is, however, a price to be paid for such strong inference (particularly the power guarantee) in terms of the simulation effort required to attain it. Although the structure of CSB makes it efficient when there are a small number of important factors, the efficiency of the testing procedure used in each step is just as critical to the overall computational cost to run CSB.

Wan et al. (2006) described a fully sequential testing procedure for the special case when the experimenter is willing to set $\alpha = 1 - \gamma$, where α is the required probability of Type I error and γ is the required power. The procedure is based on a ranking-and-selection procedure described by Kim (2005); it adds one replication at a time and terminates as soon as a conclusion can be reached, and in most cases it is much more efficient than their two-stage testing procedure that does not link Type I error and power. However, the requirement that $\alpha = 1 - \gamma$ limits the fully sequential test's usefulness. For example, the experimenter may require high power (have low probability of missing important factors)—say, 0.95—and $\alpha = 0.1$ may be a sufficient level of Type I error. In this section, we generalize the fully sequential testing procedure so that it can be used when $\alpha \neq 1 - \gamma$. This is a nontrivial extension that requires a significantly different development. The extended fully sequential testing procedure is a qualified test for CSB and is invoked each time the Test step is encountered in Figure 1.

3.1. Procedure

Unlike the two-stage testing procedure for which the second-stage sample size is based on the worst-case scenario (Wan et al. 2006), the extended fully sequential testing procedure adds one replication at a time to both the upper and lower levels of the group being tested. It will reach a conclusion as soon as the information is sufficient. We use r to represent the current number of replications at levels k_1 and k_2 (the replications are always paired for the upper and lower levels of the testing group). The initial experiment at any level consists of N_0 replications, but more generally, n_k denotes the number of replications that have been taken at level k. For a specific selected group with factors $\{k_1 + 1, k_1 + 2, \ldots, k_2\}$, it is easy to see that given
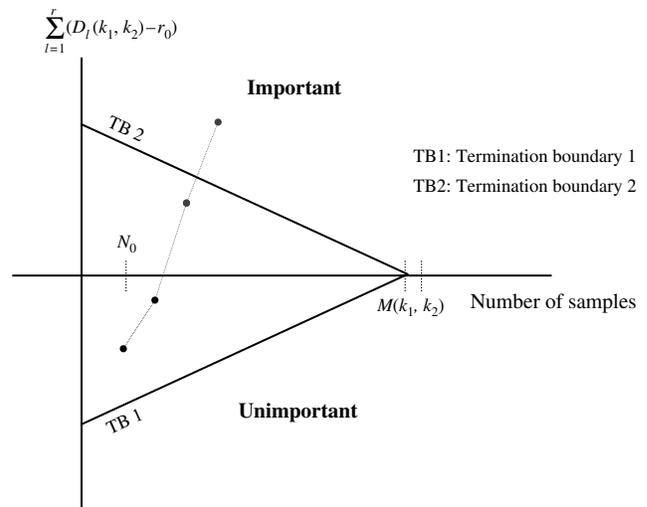
Figure 2 **Fully Sequential Test**

model (1), $\mathrm{E}[D_l(k_1, k_2)] = \sum_{i=k_1+1}^{k_2} \beta_i$, which is the sum of the main effects of the group. The test also uses

$$S^2(k_1, k_2) = \frac{1}{N_0 - 1} \sum_{l=1}^{N_0} (D_l(k_1, k_2) - \bar{D}(k_1, k_2))^2,$$

the first-stage (initial N_0 observations) sample variance of the paired differences (see Equation 7 in §4). The fully sequential test takes paired observations from each level, one pair at a time, and checks whether $\sum_{l=1}^{r} (D_l(k_1, k_2) - r_0)$ crosses one of two termination boundaries, where r_0 is a drift parameter (see Figure 2). If Termination boundary 1 is crossed, the group is declared unimportant. If Termination boundary 2 is crossed, the group is declared important. The maximum number of paired observations that will be taken is one more than $M(k_1, k_2) = \lfloor a(k_1, k_2)/\lambda \rfloor$, where $a(k_1, k_2)$ and λ are parameters described below.

The procedure is illustrated in Figure 2, where the dots represent the observed value of the test statistic as a function of the number of paired observations. The algorithm is presented in Figure 3. After the initial N_0 replications, the test adds one replication at a time to both the upper and lower levels of the group being tested until a decision is made. The critical region for the extended fully sequential test is a function of the following quantities:

- $\pm a(k_1, k_2) = \pm a_0 S^2(k_1, k_2)$, the intercepts of the triangular region in Figure 2; and
- $\pm \lambda = \pm (\Delta_1 - \Delta_0)/4$, the slopes of the sides of the triangular region.

The constants a_0 and r_0 are the solutions of the following equations:

$$\int_0^\infty \int_{-\infty}^\infty \theta(x) \frac{\nu(y)}{1 - \nu(y)} \phi\left(\frac{y\theta(x) - (r_0 - \Delta_0)}{\theta(x)}\right)$$
$$\cdot f(x, N_0 - 1) \, dx \, dy = \alpha \qquad (5)$$

0. Set $r = n_{k_1} = n_{k_2}$.
1. If $r > M(k_1, k_2)$ then
   (a) If $\sum_{l=1}^{r} (D_l(k_1, k_2) - r_0) \leq 0$, then stop and classify the group as unimportant.
   (b) Else stop and classify the group as important.
2. Else (i.e., $r \leq M(k_1, k_2)$)
   (a) If $\sum_{l=1}^{r} (D_l(k_1, k_2) - r_0) \leq -a(k_1, k_2) + \lambda r$ (Termination boundary 1), then classify the group as unimportant.
   (b) Else if $\sum_{l=1}^{r} (D_l(k_1, k_2) - r_0) \geq a(k_1, k_2) - \lambda r$ (Termination boundary 2), then classify the group as important.
   (c) Else take one more replication at both levels k_1 and k_2, set $r = r + 1$, and go to Step 1.

Figure 3 **Extended Fully Sequential Test**

and

$$\int_0^\infty \int_{-\infty}^\infty \theta(x) \frac{\nu(y)}{1 - \nu(y)} \phi\left(\frac{y\theta(x) - (\Delta_1 - r_0)}{\theta(x)}\right)$$
$$f(x, N_0 - 1)\, dx\, dy = 1 - \gamma, \qquad (6)$$

where

$$\nu(y) = e^{-2\lambda y}, \quad \phi(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}, \quad \theta(x) = \frac{\sqrt{N_0 - 1}}{\sqrt{x a_0/\lambda}},$$

and

$$f(x, n) = \begin{cases} \dfrac{1}{2^{n/2}\Gamma(n/2)} x^{n/2-1} e^{-x/2}, & x \geq 0, \\ 0, & x < 0, \end{cases}$$

which is the χ^2 distribution with n degrees of freedom, and $\Gamma(n/2) = \int_0^\infty x^{n/2-1} e^{-x}\, dx$.

If $\alpha = 1 - \gamma$ (the symmetric case), then $a_0 = 2\eta(N_0 - 1)/(\Delta_1 - \Delta_0)$ and $r_0 = (\Delta_1 + \Delta_0)/2$, where $\eta = (e^{-2\ln(2\alpha)/(N_0-1)} - 1)/2$ (Hartmann 1991, Kim 2005). Our contribution is to extend the procedure to the asymmetric case ($\alpha \neq 1 - \gamma$). Because achieving both high power and small Type I error often requires a very large number of replications, our effort to decouple the choice of α and γ improves the efficiency of the fully sequential testing procedure in practical applications.

As in Wan et al. (2006), replications are obtained whenever new groups are formed: when forming a new group containing the factors $\{k_1 + 1, k_1 + 2, \ldots, k_2\}$ with $k_1 < k_2$, the number of observations at levels k_1 and k_2 are equalized in the following way before beginning the test:

If $n_{k_1} = 0$, then collect N_0 observations at level k_1 and set $n_{k_1} = N_0$;

If $n_{k_2} = 0$, then collect N_0 observations at level k_2 and set $n_{k_2} = N_0$;

If $n_{k_1} < n_{k_2}$, then make $n_{k_2} - n_{k_1}$ additional replications at level k_1 and set $n_{k_1} = n_{k_2}$; and

If $n_{k_2} < n_{k_1}$, then make $n_{k_1} - n_{k_2}$ additional replications at level k_2 and set $n_{k_2} = n_{k_1}$.

When the group effect is significantly larger than Δ_1 or smaller than Δ_0, the test can be much more efficient than a two-stage testing procedure because of possible termination. The procedure proposed here is applicable to both symmetric (i.e., $\alpha = 1 - \gamma$) and asymmetric cases (i.e., $\alpha \neq 1 - \gamma$). When $\alpha = 1 - \gamma$, the parameters a_0 and r_0 can be calculated directly; otherwise, the closed forms are not available and numerical methods must be used to solve Equations (5) and (6); see §3.2.

3.2. Determining Critical Values
The challenge of implementing the extended fully sequential testing procedure lies in solving Equations (5) and (6) to get a_0 and r_0. Numerical methods are used to evaluate the double integrals, and a two-dimensional line search is then required to find a solution. A Matlab (The Mathworks, Inc.) algorithm is presented in Appendix C in the Online Supplement (available at http://joc.pubs. informs.org/ecompanion.html). The following lemmas make the search relatively easy.

LEMMA 1. *For a_0 fixed, both the power and the Type I error are decreasing in r_0.*

LEMMA 2. *If $\alpha \leq 1/2$ and $\gamma \geq 1/2$, then $\Delta_0 \leq r_0 \leq \Delta_1$.*

The proofs of the lemmas are given in Appendix A in the Online Supplement. Lemma 2 gives initial upper and lower bounds for r_0; based on the monotonicity property of Lemma 1, the bounds can be updated during the search process to close in on the solution.

3.3. Performance of the Extended Fully Sequential Testing Procedure
The extended fully sequential testing procedure controls the Type I error and power for each bifurcation step; i.e., it is a qualified testing procedure. More specifically, we can prove the following theorem for every single bifurcation step. The proof of the performance of the entire CSB and CSB-X procedure then follows as in Wan et al. (2006); see Theorems 1–4.

THEOREM 5. *Suppose the responses $D_l(k_1, k_2)$ (Equation 4) are independent and identically distributed (i.i.d.) normal random variables.*
- *If $\mathrm{E}[D_l(k_1, k_2)] \leq \Delta_0$, then $\Pr\{$Declare the group important$\} \leq \alpha$.*
- *If $\mathrm{E}[D_l(k_1, k_2)] \geq \Delta_1$, then $\Pr\{$Declare the group important$\} \geq \gamma$.*

The proof is given in Appendix A in the Online Supplement.

Figure 4 illustrates the potential for increased efficiency by decoupling α and $1 - \gamma$. The outer triangle is the continuation region for a case where
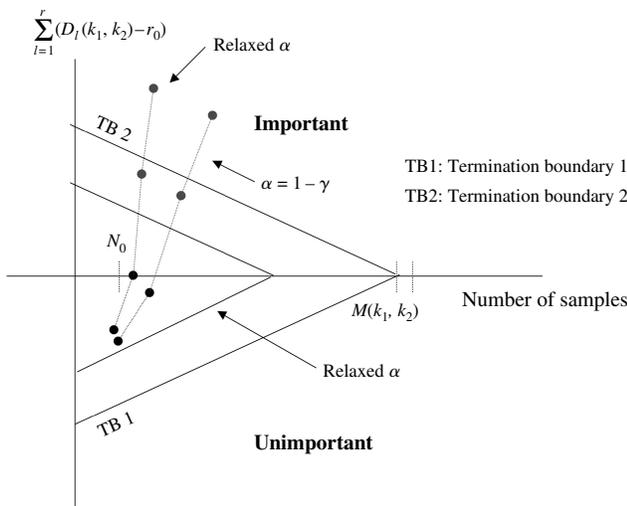
Figure 4 Illustration of Efficiency Improvement with Reduced Type I Error Requirements (the Smaller Triangular Region and the "Relaxed α" Sample Path)

Table 1 Triangle Region Width Comparison with Different α and γ;
Ratio 1 $= T(\alpha, \gamma, \Delta_0)/(T(0.05, 0.95, \Delta_0))$,
Ratio 2 $= T(\alpha, \gamma, \Delta_1)/(T(0.05, 0.95, \Delta_1))$

		$N_0 = 25$		$N_0 = 10$	
α	γ	Ratio 1	Ratio 2	Ratio 1	Ratio 2
0.05	0.95	1	1	1	1
0.05	0.90	0.76	0.92	0.74	0.91
0.05	0.85	0.63	0.88	0.60	0.88
0.05	0.80	0.53	0.87	0.51	0.87
0.05	0.75	0.47	0.86	0.45	0.87
0.05	0.70	0.41	0.88	0.39	0.89
0.10	0.95	0.92	0.76	0.92	0.74
0.15	0.95	0.88	0.63	0.88	0.60
0.20	0.95	0.87	0.54	0.87	0.51
0.25	0.95	0.86	0.47	0.88	0.45
0.30	0.95	0.87	0.41	0.89	0.39

$\alpha = 1 - \gamma$. The inner triangle is obtained when either Type I error is relaxed (larger α) or power is decreased (smaller γ). A smaller continuation region leads to earlier termination. Further enhancing this effect, the drift parameter r_0 (which is subtracted in the partial sum) decreases as α increases, tending to make a sample path that drifts up exit the region more quickly, as illustrated in Figure 4. An analogous effect occurs when γ decreases.

Another way to evaluate the potential improvement is to compare the number of observations until the tests terminate, assuming the partial sum follows the expected value of its sample path:

$$E\left[\sum_{l=1}^{r}(D_l(k_1, k_2) - r_0)\right], \quad r = 1, 2, \ldots .$$

When $E[D_l(k_1, k_2)] = \Delta_0$ or Δ_1, it is easy to show that the test terminates after

$$T(\alpha, \gamma, \Delta_0) = \left\lceil \frac{-a_0 S^2(k_1, k_2)}{\Delta_0 - r_0 - \lambda} \right\rceil$$

or

$$T(\alpha, \gamma, \Delta_1) = \left\lceil \frac{a_0 S^2(k_1, k_2)}{\Delta_1 - r_0 + \lambda} \right\rceil$$

observations, respectively. In Table 1, we examine the ratio

$$\frac{T(\alpha, \gamma, \Delta)}{T(\alpha_0, 1 - \alpha_0, \Delta)}$$

for $\Delta_0 = 2$, $\Delta_1 = 4$ (so $\lambda = (\Delta_1 - \Delta_0)/4 = 0.05$), $\alpha_0 = 0.05$ and $N_0 = 10$ and 25, ignoring rounding (notice that the ratio is free of $S^2(k_1, k_2)$). We see that if we relax either the Type I error or power, this ratio can be substantially less than one.

4. CSB with Fold-Over Designs (CSB-X)

In this section, we introduce an improved CSB methodology called CSB-X. The new procedure can handle two-factor interactions and gives the same error control for main-effect screening as the original CSB. This work exploits the fold-over design for SB of Kleijnen et al. (2006) (also see Bettonvil 1990, Bettonvil and Kleijnen 1997). Notice that the fold-over design presented here requires three levels of each factor to eliminate both quadratic effects and two-factor interactions. The fold-over design can be conducted with two levels if we code the low levels at "-1" and high levels as "$+1$". The number of design points for CSB-X in that case will be one less (the center point), and the design will eliminate the two-factor interactions but not quadratic effects. The difference is small. We use the three-level setting to be consistent with the coding of the CSB procedure in Wan et al. (2006).

When quadratic effects and two-factor interactions exist (model 2), then

$$Z(k) = \beta_0 + \sum_{i=1}^{k}\beta_i + \sum_{i=1}^{k}\sum_{j=i}^{k}\beta_{ij} + \varepsilon(k) \quad \text{and}$$

$$E[D_l(k_1, k_2)] = \sum_{i=k_1}^{k_2}\beta_i + \sum_{i=1}^{k_1}\sum_{j=k_1+1}^{k_2}\beta_{ij} + \sum_{i=k_1+1}^{k_2}\sum_{j=i}^{k_2}\beta_{ij}.$$

Therefore, the group main effect is biased by the two-factor interactions. Specifically, when large negative interactions exist, important main effects may get canceled and the screening result is misleading. To eliminate the influence of the quadratic effects and interactions, new design points called *mirror levels* will be run. For each level $k \geq 0$, the mirror level of the experiment at level k, denoted by level $-k$, is defined by the following factor settings (also see Equation 3):

$$x_i(-k) = \begin{cases} -1, & i = 1, 2, \ldots, k; \\ 0, & i = k+1, \ k+2, \ldots, K. \end{cases}$$

Therefore, $Z(-k) = \beta_0 - \sum_{i=1}^{k} \beta_i + \sum_{i=1}^{k} \sum_{j=1}^{i} \beta_{ij} + \varepsilon(-k)$.

The initial experiment at any level consists of N_0 replications, but more generally, n_k denotes the number of replications that have been taken at level k. If we define $Y_l(k) = (Z_l(k) - Z_l(-k))/2, 0 \leq k \leq K$, and redefine the difference $D_l(k_1, k_2)$ as

$$D_l(k_1, k_2) = Y_l(k_2) - Y_l(k_1),$$
$$l = 1, 2, \ldots, \quad \min\{n_{k_1}, n_{k_2}\}, \quad (7)$$

then it is easy to see that given model (2) $E[Y_l(k)] = \sum_{i=1}^{k} \beta_i$ and $E[D_l(k_1, k_2)]$ is $\sum_{i=k_1+1}^{k_2} \beta_i$, which is the sum of the main effects of the group. The effects of interactions and quadratic terms are eliminated. This is the main improvement of CSB-X over CSB and it is accomplished through a simple data transformation from the Zs to the Ys. With this data transformation, CSB-X is exactly the same as CSB. In CSB-X, whenever data are collected at any level k, the same amount of data is also collected at level $-k$. All of the definitions and notations from Wan et al. (2006) carry over to the new $Y_l(k)$. Given model (2), both $Y_l(k)$ in CSB and $Y_l(k)$ in CSB-X are normally distributed estimators of $\sum_{l=1}^{k} \beta_i$; hence, the data transformation introduced in CSB-X does not influence the validity of the statistical test, so both the two-stage testing procedure introduced in Wan et al. (2006) and the fully sequential testing procedure described above are qualified procedures and can be implemented directly. The error control of CSB-X is the same as in CSB (Theorems 1–4). Specifically, CSB-X with a qualified hypothesis testing procedure controls the Type I error for each factor individually, and guarantees the power for each step. The screening procedure does not require an equal variance assumption and is valid with or without CRN when using our tests (Wan et al. 2006).

It should be noted that although CSB-X doubles the number of design points, the variance of the test statistics is reduced compared with the test statistics of CSB. It can be shown that in homogeneous variance cases, the expected number of runs required for CSB-X and CSB are the same (see Appendix B in the Online Supplement). Therefore CSB-X can achieve the same level of efficiency as CSB while avoiding mistakes in the presence of interactions.

5. Numerical Evaluation of Extended CSB and CSB-X in the Presence of Interactions

In this section, we present numerical evaluations of CSB-X in comparison with original CSB and fractional factorial designs.

5.1. Comparison of CSB-X and CSB in the Presence of Interactions

To illustrate the influence of interactions on the screening results of CSB and CSB-X, three cases with $K = 10$ were explored with $\Delta_0 = 2$ and $\Delta_1 = 4$. The response follows model (2):

$$Y = \beta_0 + \sum_{i=1}^{K} \beta_i x_i + \sum_{i=1}^{K} \sum_{j=i}^{K} \beta_{ij} x_i x_j + \varepsilon,$$

where the error ε is normally distributed with mean zero and a standard deviation equal to $1 +$ |expected response|. We also set $\alpha = 0.05$ and $\gamma = 0.90$. The main effects β_i $(i = 1, 2, \ldots, K)$ are fixed, and interactions β_{ij} $(0 < i, j \leq K)$ are randomly generated from a normal distribution with mean zero and variance four. This allows for a substantial number of important interactions thus stressing the method's ability to estimate the main effects. For each case considered, CSB and CSB-X are applied 1,000 times (macro replications), using the fully sequential testing procedure described in §3. So as not to overuse a particularly favorable interaction set, a new interaction set (β_{ij}) is generated for each trial. The fraction of trials in which each factor is declared important, $\widehat{\Pr}\{DI\}$, is recorded, which is an unbiased estimator of $\Pr\{\text{factor } i \text{ is declared important}\}$.

In Case 1, we set $(\beta_1, \beta_2, \ldots, \beta_{10}) = (0, 0, 0, 0, 0, 0, 0, 0, 0, 0)$; the observed frequency that each factor is declared important should be close to zero; in Case 2, $(\beta_1, \beta_2, \ldots, \beta_{10}) = (2, 2, 2, 2, 2, 2, 2, 2, 2, 2)$; the observed frequency that each factor is declared important should be no larger than 0.05; in Case 3, $(\beta_1, \beta_2, \ldots, \beta_{10}) = (2, 2.44, 2.88, 3.32, 3.76, 4.2, 4.64, 5.08, 5.52, 6)$; the observed frequency that β_1 is declared important should be no larger than 0.05, but for $\beta_6, \beta_7, \ldots, \beta_{10}$ it should be near 0.90. Case 3 is clearly not a typical screening situation because all factors are important. However, it is included to demonstrate the error and power control of CSB-X. The results of the experiments are given in Table 2.

We can see that CSB-X gives the desired screening results with appropriate Type I error and power control. CSB, on the other hand, loses the control of both Type I error and power when the interactions are present. For example, in Case 3 the observed $\widehat{\Pr}\{DI\}$ of β_1 is 0.17 in CSB, much larger than the desired Type I error 0.05. In CSB-X, the $\widehat{\Pr}\{DI\}$ is 0. For a large factor effect such as $\beta_{10} = \Delta_0 + \Delta_1 = 6$, the power of identifying the factor as important in CSB is only 0.46, which means that more than half of the time the factor is misclassified. On the other hand, factor 10 is always identified as important by CSB-X.

Because CSB gives misleading results in these cases, it is meaningless to compare the number of replications required by the two methods. The absolute number of replications required by CSB-X can be large,

Table 2   Screening Results ($\widehat{\Pr}\{DI\}$) for CSB and CSB-X with Interactions Present

Factor no.	Effect	$\widehat{\Pr}\{DI\}$ CSB	CSB-X	Effect	$\widehat{\Pr}\{DI\}$ CSB	CSB-X	Effect	$\widehat{\Pr}\{DI\}$ CSB	CSB-X
1	0	0.01	0	2	0.14	0.00	2.00	0.17	0.00
2	0	0.07	0	2	0.25	0.03	2.44	0.33	0.10
3	0	0.09	0	2	0.30	0.04	2.88	0.39	0.34
4	0	0.12	0	2	0.28	0.03	3.32	0.40	0.62
5	0	0.15	0	2	0.32	0.04	3.76	0.44	0.84
6	0	0.13	0	2	0.27	0.03	4.20	0.40	0.95
7	0	0.16	0	2	0.27	0.05	4.64	0.43	0.98
8	0	0.17	0	2	0.28	0.04	5.08	0.50	1.00
9	0	0.15	0	2	0.28	0.04	5.52	0.46	1.00
10	0	0.17	0	2	0.32	0.04	6.00	0.46	1.00
No. of replications		1,781	971		6,106	21,408		10,464	19,793

especially when response variances are large, but this is necessary to provide the power guarantee. Table 3, which revisits Case 3 from Table 2, shows that the computational effort can be reduced by softening the Type I error and power requirements. In the special case when only Type I error control is required, a one-sided t test suffices and the number of replications is bounded above by $2N_0(K+1)$ for CSB-X.

5.2. Comparison of CSB-X and Fractional Factorial Designs for Large-Scale Problems

In this section we study screening problems with 200 factors and 500 factors, respectively, and compare CSB-X to a standard unreplicated fractional factorial design (i.e., an orthogonal array). For each case, only 2% of the factors are important. The important factors have effects equal to five and the unimportant factors have effects equal to zero. Normal errors are assumed with mean zero and standard deviation one (equal variances across different levels). The interactions β_{ij} are generated as in previous section from a normal distribution with mean zero and variance four. The threshold of importance, Δ_0, is set to two; and the critical threshold, Δ_1, is set to four. The initial number

Table 3   Screening Results for Case 3 with Different α and γ

Factor effect	$\alpha = 0.05$ $\gamma = 0.80$	$\alpha = 0.05$ $\gamma = 0.70$	$\alpha = 0.15$ $\gamma = 0.90$	$\alpha = 0.30$ $\gamma = 0.90$	$\alpha = 0.30$ $\gamma = 0.70$
2.00	0.00	0.00	0.01	0.10	0.04
2.44	0.07	0.05	0.26	0.45	0.28
2.88	0.27	0.20	0.51	0.63	0.47
3.32	0.46	0.41	0.72	0.79	0.56
3.76	0.71	0.60	0.86	0.88	0.65
4.20	0.86	0.78	0.93	0.93	0.77
4.64	0.95	0.91	0.97	0.97	0.84
5.08	0.98	0.97	0.99	0.99	0.88
5.52	0.99	0.98	1.00	0.99	0.90
6.00	1.00	0.99	1.00	1.00	0.94
No. of runs	17,307	15,216	12,757	8,458	4,238

of runs at each level, N_0, is equal to five. The Type I error is set to be $\alpha = 0.05$ and the power requirement is $\gamma = 0.90$.

For each case, there are two scenarios. The first scenario has all important factors clustered together with the smallest indices so that the number of important groups is as small as possible at each step. The second scenario has the important factors evenly spread so there are the maximum number of important groups remaining at each step. Obviously, CSB-X is more efficient with the first scenario than with the second scenario.

For each case and scenario considered, CSB-X with the fully sequential testing procedure is applied 1,000 times, and the average number of replications required for screening is recorded. The lower bound of the number of replications required for the fractional factorial design (FFD) is the smallest number of design points required to estimate 200 or 500 main effects with a Resolution IV design (i.e., main effects are not confounded with two-factor interactions but two-factor interactions are confounded with each other). This lower bound equals $2K + 1$ for K factors (Wu and Hamada 2000) and is not influenced by the scenario. The comparison is demonstrated in Table 4.

We can see that for both cases of K, CSB-X takes only approximately 28% and 19% of the minimum number of replications required by fractional factorial designs, respectively, in the clustered scenario. In the other scenario, CSB-X requires 77% and 75% of the lower bounds, respectively. In addition, if the conditions of Theorem 3 or Theorem 4 are satisfied, the expected number of factors that will be falsely classified as important ($E[F_K]$) for the fractional factorial design always equals αK, which is greater than or equal to that of CSB-X, especially for the conditions of Theorem 3. Furthermore, the fractional factorial design does not have power control. Therefore, from the error control point of view, CSB-X is superior. Moreover, the typical test implemented in a fractional factorial design assumes equal variance across design points, which CSB does not require. However, the fractional factorial design does not

Table 4   Comparison of CSB-X and Fractional Factorial Design

Scenarios	No. of runs required CSB-X	FFD (lower bound)
200 factors, {1, 2, 3, 4} important	111	401
200 factors, {1, 51, 101, 151} important	310	401
500 factors, {1, 2, 3, 4, 5, 6, 7, 8, 9, 10} important	186	1,001
500 factors, {1, 51, 101, 151, 201, 251, 301, 351, 401, 451} important	754	1,001

require $\beta_i \geq 0$ $(i > 0)$ and can directly estimate group of interactions (Resolution IV designs) or individual interactions (Resolution V design). CSB-X can only eliminate quadratic effects and two-factor interactions but not estimate them.

6. Semiconductor Case Study Revisited

Wan et al. (2006) implemented CSB to study a simplified semiconductor manufacturing system to identify the important machines and transporters that are worthy of investment. In this section we use CSB-X with the extended fully sequential test to study the case again. The results from CSB-X and CSB are compared.

The detailed description of the semiconductor manufacturing system will not be repeated here. In summary, there are two main processes, diffusion and lithography, each of which contains substeps. The raw material will be released at the rate of one cassette per hour and processed in single-cassette loads, beginning with the CLEAN step of diffusion, and after diffusion, proceeding to the lithography process. The diffusion and lithography then alternate until the product completes processing (see Figure 5). The numbers of passes required are different among the products. The time for the movement between each process is negligible and the movement between diffusion and lithography will be handled by a transporter. The factors we are interested in are the number of fast and slow machines at each substep and the numbers of each kind of transporter, which are listed in Table 5.

In addition to the high and low settings in CSB, CSB-X (and the factorial design discussed later) requires a mirror setting for each factor. To accommodate this requirement, the original low setting for each factor is used as the mirror setting; the original high setting for each factor is used as the low setting; and a new high setting is determined by the cost model in Wan et al. (2006). The new experimental design is given in Table 6. The CSB procedure uses the mirror settings and high settings as the low and high settings to guarantee that the measured effects of CSB and CSB-X are comparable.

The simulation programming of the manufacturing system was done in simlib, a collection of ANSI-standard C support functions for simulation (Law and Kelton 2000). CRN was implemented by assigning each station a separate stream of random numbers. CSB and CSB-X were implemented in C++ with the extended fully sequential testing procedure. For each replication, 365 days of operation were simulated with a 300 hour warm-up period. The performance measure was the long-run average cycle time (hours) weighted by percentage of different products.
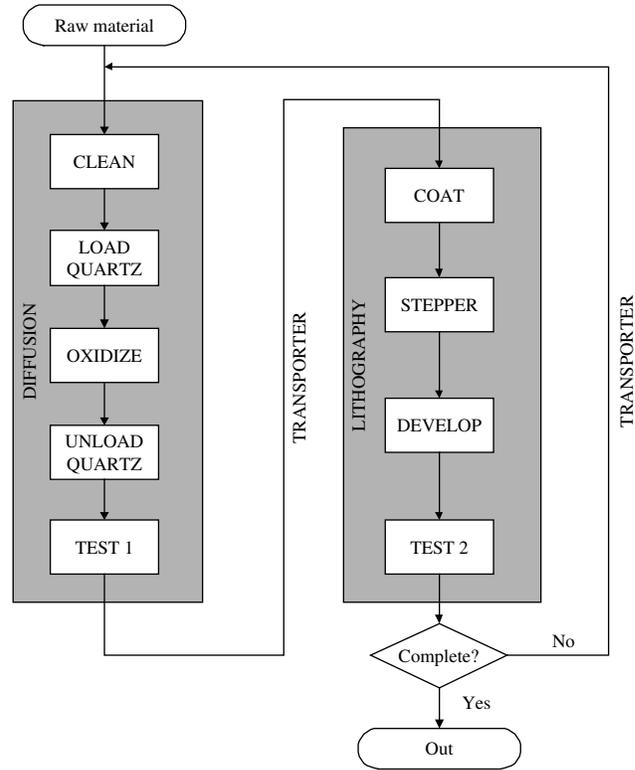
Figure 5 Production Process of the Semiconductor Manufacturing System

Thus, for this example it was more natural to define all effects as negative instead of positive. Here, Δ_0 was the minimum acceptable decrease in long-run cycle time, which we set to be one hour in this test; Δ_1 was the decrease in long-run cycle time that we do not want to miss, which we set to be two hours in this test.

Both CSB and CSB-X are implemented with $\alpha = 0.05$, $\gamma = 0.9$, and $N_0 = 25$. CSB consumed 500 replications identifying eight factors as important and CSB-X consumed 950 replications identifying eight factors as important (Table 7). Both CSB and CSB-X identify factors 3, 5, 6, 16, and 20 as important. CSB identifies

Table 5 Mean Processing Time per Cassette for Each Step (Hours) and Cost of Machines (Millions of Dollars)

Stations	Fast machine	Cost per unit	Slow machine	Cost per unit
CLEAN	1.5	1.38	2.5	0.83
LOAD QUARTZ	0.19	0.63	0.31	0.38
OXIDIZE	3.5	3.25	5.4	1.95
UNLOAD QUARTZ	0.19	0.63	0.31	0.38
TEST 1	0.5	1.25	1.25	0.75
COAT	0.75	1.13	1.50	0.68
STEPPER	0.85	2.25	1.8	1.35
DEVELOP	0.38	0.25	0.63	0.15
TEST 2	0.5	1.25	1.25	0.75
AGV	0.028	1.05	NA	NA
CONVEYOR	NA	NA	0.19	0.635

Table 6   Factor Description and Settings (Unit Number)

Factor ID	Factor description	Mirror	Low	High
1	Number of slow machines in OXIDIZE	92	93	94
2	Number of fast machines in STEPPER	0	1	2
3	Number of fast machines in COAT	0	2	4
4	Number of slow machines in CLEAN	42	45	48
5	Number of fast machines in TEST 1	0	2	4
6	Number of fast machines in TEST 2	0	2	4
7	Number of slow machines in STEPPER	30	32	34
8	Number of slow machines in COAT	25	29	33
9	Number of slow machines in CLEAN	0	2	4
10	Number of slow machines in TEST 1	21	25	29
11	Number of slow machines in TEST 2	21	25	29
12	Number of slow machines in LOAD QUARTZ	5	13	21
13	Number of slow machines in UNLOAD QUARTZ	5	13	21
14	Number of fast machines in LOAD QUARTZ	0	5	10
15	Number of fast machines in UNLOAD QUARTZ	0	5	10
16	Number of AGVs	0	5	10
17	Number of slow machines in DEVELOP	10	31	52
18	Number of CONVEYORS	6	9	12
19	Number of fast machines in OXIDIZE	0	1	2
20	Number of fast machines in DEVELOP	0	13	26

Table 7   CSB, CSB-X, and Factorial Design Screening Results

Factor ID	$\hat{\beta}^{\text{CSB}}$	$\hat{\beta}^{\text{CSB-X}}$	Confidence interval (95%) from the factorial design
1	0.08	−0.00527	(−0.0057, 0.0467)
2	−1.39	−1.269	(**−1.333**, **−1.205**)
3	**−2.20**	**−2.006**	(**−2.070**, **−1.943**)
4	−0.39	−0.162	(−0.206, −0.118)
5	**−3.03**	**−2.772**	(**−2.812**, **−2.732**)
6	**−3.27**	**−2.784**	(**−2.815**, **−2.753**)
7	−0.14	−0.239	(−0.283, −0.195)
8	−0.02	−0.295	(−0.344, −0.247)
9	−1.34	−1.404	(**−1.471**, **−1.338**)
10	0.19	−0.291	(−0.335, −0.247)
11	−0.14	−0.281	(−0.325, −0.238)
12	**−1.62**	−0.846	(−0.888, −0.804)
13	**−1.56**	−0.838	(−0.895, −0.781)
14	−0.83	**−1.676**	(**−1.709**, **−1.644**)
15	−0.74	**−1.638**	(**−1.674**, **−1.603**)
16	**−4.14**	**−3.169**	(**−3.211**, **−3.126**)
17	**−1.78**	−0.840	(−0.858, −0.822)
18	0.08	**−1.00**	(−1.053, −0.942)
19	−0.55	−0.569	(−0.607, −0.532)
20	**−1.77**	**−2.476**	(**−2.566**, **−2.385**)

Note. Identified important effects are in bold.

Table 8   Estimate of Large Interactions from the Factorial Experiments

Interactions	Estimate	Confidence interval (95%)
(5, 5)	0.56	(0.17, 0.95)
(13, 15)	0.83	(0.79, 0.88)
(14, 14)	0.91	(0.58, 1.24)
(15, 15)	0.64	(0.37, 0.91)
(16, 16)	2.62	(2.38, 2.87)
(16, 18)	1.00	(0.94, 1.04)
(17, 20)	0.81	(0.75, 0.87)
(20, 20)	1.81	(1.48, 2.15)

factors 12, 13, and 17 as important, whereas CSB-X does not; CSB-X identifies factors 14, 15, and 18 as important whereas CSB does not.

To explain the different results between the two methods, we performed an additional experiment using a Resolution V central composite design with 553 design points (the mirror level is coded as "−1" low level as "0" and high level as "+1") and took 10 independent replications at each design point (Sanchez and Sanchez 2005). Therefore, we can get 10 independent estimates of each β_i and β_{ij}. Regression analysis shows that the model with two-factor interactions and quadratic effects has $R^2 > 97\%$ and $R^2(adj) > 96\%$, whereas the model with only main effects has R^2 around 0.79 and $R^2(adj)$ around 78%. This implies that two-factor interactions and quadratic effects are statistically significant (nonzero) and ignoring them may lead to biased results, and that the second-order model can predict the responses well. There are both positive and negative interactions and quadratic effects and most of them are small. In Table 8 the relatively large interactions and quadratic effects (greater than $\Delta_0/2$) are listed. We can see that the important ones are sparse. With simple algebra, we can use the estimated regression coefficients $\hat{\beta}_i^F$ and $\hat{\beta}_{ij}^F$ (F represents factorial design) to approximate the "main-effect coefficient" if we assume a main-effects model even though there are interactions, as CSB does (β_m^{CSB}), and the main-effect coefficients if we eliminate the bias in the main-effects estimators because of the two-factor interactions and quadratic terms, as CSB-X does ($\beta_m^{\text{CSB-X}}$). Specifically, $\beta_m^{\text{CSB}} = \text{E}[(Z(m) - Z(m-1))/2] \approx \hat{\beta}_m^F + \sum_{i=1}^{m-1} \hat{\beta}_{im}^F - \sum_{i=m+1}^{K} \hat{\beta}_{im}^F$, and $\beta_m^{\text{CSB-X}} = \text{E}[(Z(m) - Z(-m) - Z(m-1) + Z(-m+1))/4] \approx \hat{\beta}_m^F$.

Table 7 shows the 95% confidence intervals for the estimated main effects from the factorial design and the implied values of β_m^{CSB}s and $\beta_m^{\text{CSB-X}}$s (see above), as well as the screening results for CSB, CSB-X, and the factorial design (through confidence intervals). It is clear that factors 12, 13, and 17 are actually not important, and factors 14 and 15 are important. Factor 18 is on the boundary. This case study demonstrates that even though most of the interactions are not important, they can still bias the results of CSB. In this sense, CSB-X is preferred. In addition, to achieve confidence intervals for the estimated βs without an equal variance assumption, we felt that a Resolution V central composite design would need at least three replications per design point, a total of 1,659 observations (the results presented above are based on 10 replications per design point, a total of 5,530 observations). CSB-X only takes a fraction of this, and the results are satisfactory.

7. Conclusion

In this paper, we extended the fully sequential testing procedure proposed in Wan et al. (2006) to general cases and applied it in factor screening. The extended test does not enforce $\alpha = 1 - \gamma$ and can be used as a stand-alone hypothesis testing procedure with controlled Type I error rate and power in applications other than factor screening. Numerical evaluation shows that decoupling of Type I error and power can significantly improve the efficiency of the procedure.

In addition, we incorporated a fold-over design into CSB to identify important main effects even when two-factor interactions and quadratic terms are present. The new CSB-X procedure doubles the number of design points compared to CSB; however, the decrease of the variance of the test statistic typically compensates for this when a main-effects model applies. When interactions exist, CSB is not appropriate so sample size comparisons are irrelevant. Thus, CSB-X provides all of the benefits of CSB with more generality and without a loss of efficiency.

References

Bettonvil, B. 1990. *Detection of Important Factors by Sequential Bifurcation*. Tilburg University Press, Tilburg, The Netherlands.

Bettonvil, B. 1995. Factor screening by sequential bifurcation. *Comm. Statist.: Simulation Comput.* **24**(1) 165–185.

Bettonvil, B., J. P. C. Kleijnen. 1997. Searching for important factors in simulation models with many factors: Sequential bifurcation. *Eur. J. Oper. Res.* **96**(1) 180–194.

Cheng, R. C. H. 1997. Searching for important factors: Sequential bifurcation under uncertainty. D. H. Withers, B. L. Nelson, S. Andradóttir, K. J. Healy eds. *Proc. 1997 Winter Simulation Conf.*, Institute of Electrical and Electronics Engineers, Piscataway, NJ, 275–280.

Hartmann, M. 1991. An improvement of Paulson's procedure for selecting the population with the largest mean from k normal populations with a common unknown variance. *Sequential Anal.* **10**(1–2) 1–16.

Kim, S.-H. 2005. Comparison with a standard via fully sequential procedures. *ACM Trans. Modeling Comput. Simulation* **15**(2) 155–174.

Kleijnen, J. P. C., B. Bettonvil, F. Persson. 2006. Finding the important factors in large discrete-event simulation: Sequential bifurcation and its applications. A. Dean, S. Lewis, eds. *Screening: Methods for Experimentation in Industry, Drug Discovery, and Genetics*. Springer, New York, 287–307.

Law, A. M., W. D. Kelton. 2000. *Simulation Modeling and Analysis*, 3rd ed. McGraw Hill, New York.

Sanchez, S. M., P. J. Sanchez. 2005. Very large fractional factorial and central composite designs. *ACM Trans. Modeling Comput. Simulation* **15**(4) 362–377.

Wan, H., B. E. Ankenman, B. L. Nelson. 2006. Controlled sequential bifurcation: A new factor-screening method for discrete-event simulation. *Oper. Res.* **54**(4) 743–755.

Wu, J. C. F., M. Hamada. 2000. *Experiments: Planning, Analysis, and Parameter Design Optimization*. John Wiley & Sons, New York.