SIMULATION SCREENING AND FALSE DISCOVERY RATE CONTROL FOR BOTH MAIN AND INTERACTION EFFECTS

Wen Shi

Jennifer Shang

School of Logistics and Engineering Management Hubei University of Economics 8 Yang Qiaohu Road Wuhan 430205, CHINA Katz Graduate School of Business University of Pittsburgh 4200 Fifth Avenue Pittsburgh, PA 15260, USA

Zhang Zhigang

School of Statistics Hubei University of Economics 8 Yang Qiaohu Road Wuhan 430205, CHINA

ABSTRACT

We propose a hybrid procedure that combines Morris' elementary effect, bootstrap-based hypothesis testing, and aggregated false discovery rate control to simultaneously identify the main and interaction effects in a statistical model. Numerical experiments demonstrate the efficiency and efficacy of our method.

1 INTRODUCTION

Many screening methods have been proposed to search for important factors with minimal number of simulation runs, including classic two-level designs, supersaturated designs, frequency domain experimentation, together with aggregated designs (Kleijnen 2015). Among them, sequential bifurcation (SB) and Morris method (MM) introduced by Bettonvil and Kleijnen (1997) and Morris (1991), respectively, are two known and accepted techniques. See Borgonovo and Plischke (2015, pp. 11-12) for recent review on sensitivity analysis based on simulation/computer experiments.

In this paper, we focus on MM, an alternative screening method in simulation experiments (Borgonovo and Plischke 2016). Unlike SB, MM is a *model-free* approach based on a series of ratios to find the elementary effect of a given factor; namely, no explicit statistical model such as low-order polynomial is used. Put another way, MM is not restricted to true input/output (I/O) function defined by the underlying simulation model. Through a sampling plan, a series of elementary effects (EEs) for a given factor can be derived. Morris (1991) maintains if the set of EEs for a given factor contains only zeros, then the specific factor does not significantly impact the response. If it exhibits a linear or additive effects on response, then it has constant effect. Otherwise, it has interaction effects. Campolongo and Braddock (1999) extend the traditional MM method by providing an estimates of two-factor interaction effects, while Cropp and Braddock (2002) further extend their results. To avoid cancellation effects and to ensure the robustness of MM method, Campolongo et al. (2007) propose a normalized approach. Boukouvalas et al. (2014) propose a sequential MM method, which uses a threshold value to separate the inputs with linear and nonlinear effects to save computational cost. Fédou and Rendas (2015) present designs based on MM method to enable the estimation of the graph of interactions between pairs of input factors of a function.

The above literature shows that few researchers have studied MM and it is a relatively new technique for factor screening. Although there have been several improvements and extensions, the MM method has not been used for simulation with *performance control*. In this paper, we propose a new procedure that provides error control for MM methods, and named it the *controlled MM* (CMM). The contribution of this research is that we propose a bootstrap method to test the hypotheses for main effects and interaction effects, and develop an aggregate FDR to control the *false discovery rate* (FDR) (Benjamini and Hochberg 1995). The FDR has been widely adopted for *large-scale* hypotheses. We conduct numerical experiments to validate our approach and to compare with the conventional factor screening methods.

2 MODEL REVIEW: MULTIPLE SEQUENTIAL BIFURCATION AND MORRIS METHOD

2.1 Multiple Sequential Bifurcation (MSB)

In this section, we review a specific SB, called controlled SB, which is a combination of basic SB (Bettonvil and Kleijnen 1997; Cheng 1997) and an error control procedure, initially developed for single response (Wan et al. 2006, 2010), and then for multiple responses (Shi et al. (2014) named MSB.

The controlled SB inherits the basic sequential iterative structure of the original SB, which consists of finitely many steps. The initial step starts with the aggregation of all factors into a whole group and tests that group's effect (i.e., main effect). If the group exhibits an important effect, implying that at least one factor in the group may bear importance, then SB splits the group into two subgroups (i.e., bifurcation); otherwise removes the group and SB stops. If the group is shown to be significant, SB proceeds in a similar way to test each subgroup's effects and each subgroup is categorized either as important for further testing or as insignificant and discarded. As the bifurcation procedure continues, the significant group becomes smaller and smaller; eventually each individual factor's effect is tested.

In addition to the basic sequential and group-screening abilities, the controlled SB controls both the Type-I error of each unimportant factor and the power of detecting the important effects. We add the superscript ^(SB) to denote the SB procedure. Let $\alpha^{(SB)}$, $\beta^{(SB)}$, $\Delta_0^{(SB)}$, $\Delta_1^{(SB)}$ be the users-specified parameters with $\Delta_0^{(SB)} \leq \Delta_1^{(SB)}$, where $\alpha^{(SB)}$ is the type-I error and $\beta^{(SB)}$ the type-II error, and $\Delta_0^{(SB)}$ is the threshold of declaring unimportance and $\Delta_1^{(SB)}$ the threshold of declaring importance. The objective of the controlled SB is to assign factors into one of the three classes: (a) unimportant factors, when main effect $\leq \Delta_0^{(SB)}$; (b) important factors, when main effect $\geq \Delta_1^{(SB)}$; (c) uncertain importance when the main effect is in the range of $[\Delta_0^{(SB)}, \Delta_0^{(SB)}]$. For type (a), we expect the probability of declaring them important to be less than $\alpha^{(SB)}$. For type (b), the power of discerning them to be important is $\geq 1 - \beta^{(SB)}$. For the last type, the controlled SB gives reasonable power to identify them as important; i.e., $[\alpha^{(SB)}, 1 - \beta^{(SB)}]$.

2.2 Morris Method (MM)

The MM method proposed by Morris (1991) focus on deterministic simulation. Suppose there are k factors in the simulation model and each factor are scaled to take values from [0,1], such that the overall screening experiments are implemented in a k-dimensional unit hypercube $\Omega = [0,1]^k$. Each factor is varied across p selected levels in [0,1]. Thus, the experimental domain Ω emerges as a k-dimensional and p-level grid. Morris (1991) categories the effect of a factor in the simulation as (a) negligible, (b) linear, and (c) nonlinear/involved interaction with other factors.

To determine where each factor belongs, they propose the so-called *elementary effects*. For a given set of factors $\mathbf{x} = (x_1, ..., x_k)$, the elementary effect of the *j*th factor is defined as:

$$d_{j}(\mathbf{x}) = \frac{y(\mathbf{x} + e_{j}\Delta) - y(\mathbf{x})}{\Delta} \quad (\mathbf{x} \in \Omega; \ j = 1, ..., k),$$
(1)

where $d_j(\mathbf{x})$ is actually the partial derivative with respect to x_j ; \mathbf{x} is randomly sampled from a subset of Ω such that the transformed factor combination $\mathbf{x} + e_j \Delta$ lies in Ω ; i.e., $(\mathbf{x} + e_j \Delta) \in \Omega$; $y(\mathbf{x})$ is the simulation output for factor combination \mathbf{x} . Also, Δ is a pre-specified multiple of 1/(p-1) such that x_j can take values from [0, 1/(p-1), 2/(p-1), ..., 1]; and e_j is a zero vector except for the *j*th element with a unit.

The finite distribution of $d_j(\mathbf{x})$ (i.e., the number of $d_j(\mathbf{x})$ is finite) associated with factor j is denoted by F_j such that $d_j(\mathbf{x}) \sim F_j$. The maximum number of elements of F_j is then $p^{k-1}[p - \Delta(p-1)]$, where p^{k-1} is the number of factor combinations formed by the remaining k-1 factors with p levels except for factor j. Also, $p - \Delta(p-1)$ is the possible levels that factor j can take to obtain elementary effects. For example, when $\Delta = 1/(p-1)$, factor j can only have p-1 levels, as the last level $x_j = 1$ is excluded. A recommended selection of p is even and $\Delta = p/[2(p-1)]$.

Morris suggests that a highly centralized distribution for F_j implies steady importance of factor j across the experimental region Ω ; and that a highly decentralized distribution corresponds to strong dependence of factor j on other factors (i.e., nonlinear or interaction effect). To characterize the distribution of F_j , two unbiased estimators, sample mean \overline{d}_j , and variance S_j^2 of the observed elementary effects are used:

$$\hat{\mu}_{j} = \frac{\sum_{i=1}^{N} d_{j;i}}{N} \quad (j = 1, ..., k),$$
(2)

$$\hat{\sigma}_{j}^{2} = \frac{\sum_{i=1}^{N} \left(d_{j;i} - \overline{d}_{j} \right)^{2}}{N - 1} \quad (j = 1, ..., k),$$
(3)

where $d_{j,i}$ denote the *i* th elementary effect of factor *j*.

In addition to (2) and (3), another measure of the individual elementary effect of the j th factor has been proposed by Campolongo et al. (2007) so that the possible cancellation of elementary effects in (2) can be eliminated due to non-monotonic factor effect, that is,

$$\hat{\mu}_{j} = \frac{\sum_{i=1}^{n} |d_{j;i}|}{N} \quad (j = 1, ..., k), \tag{4}$$

where $|\cdot|$ denotes the absolute value.

To obtain either (2) or (4) and (3), we need to sample N elementary effects with replacement from F_j , by designing an sampling matrix $[\mathbf{B}_1^+,...,\mathbf{B}_N^+]^{\mathrm{T}}$, where \mathbf{B}_i^+ is a random orientation form of B with i = 1,...,N, which is a (k+1)-by-k sampling matrix, with a 1-by-k null vector and a k-by-k lower triangular matrix, where all elements below the main diagonal are 1:

$$\mathbf{B} = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ 1 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 1 & 1 & \cdots & 1 \end{bmatrix}.$$
 (5)

Then ΔB could be used as a design matrix. Thus, B_i^+ or B would produce *only* one elementary effect for each factor, with the only distinction that B_i^+ is randomly selected. Note that, B can be converted to B_i^+ through a specially developed algorithm proposed in Morris (1991).

3 IMPROVING MM APPROACH THROUGH ERROR CONTROL

3.1 The Proposed CMM Framework

We now propose an efficient and effective hypothesis testing method to accurately classify factors into different categories (see Table 1). Since the distribution of F_j is unknown, we employ the bootstrap testing method (i.e. *simulation-based testing*) to generate an empirical distribution function (EDF) of F_j . Two bootstrap testing procedures for the main effect and interaction effect are proposed independently.

Due to the large number (2k) of hypotheses in our problem, it is necessary to control the *Type-I error* (i.e. *false positives*). The conventional approach is commonly used to control the Family-Wise Error Rate (FWER). However, the FWER procedure focuses on reducing Type-I error, which often results in an extremely low statistical power, especially for large-scale testing cases, thereby increasing the risk of finding no or few factors even in the presence of significant main or interaction effects. To circumvent the weakness of FWER in multiple hypotheses testing, in this research we propose to control the False Discovery Rate (FDR). FDR, introduced by Benjamini and Hochberg (1995), controls *the expected proportion* of falsely rejected hypotheses. The reasons why we make use of FDR is that it is more relevant and practicable than FWER in many real world applications (Sun et al. 2006). This explains why FDR is so pervasive since its introduction. Efron (2004, 2007) maintain that FDR is particularly appropriate for large-scales testing cases, as it can guarantee a desired level of power.

Table 1: Potential Classification of factor	assification of factor <i>j</i>	Potential	Table 1:
---	---------------------------------	-----------	----------

		Main effect					
		Unimportant	Important				
Internation offect	Unimportant	$\mu_{j}\!\leqslant\!\Delta_{\mathrm{EE}},\;\;\sigma_{j}\!\leqslant\!\Delta_{\mathrm{IE}}$	$\mu_{j}\!>\!\Delta_{ ext{EE}},\;\;\sigma_{j}\!\leqslant\!\Delta_{ ext{IE}}$				
Interaction effect	Important	$\mu_j \leqslant \Delta_{\rm EE}, \ \sigma_j > \Delta_{\rm IE}$	$\mu_j > \Delta_{\mathrm{EE}}, \ \sigma_j > \Delta_{\mathrm{IE}}$				

3.2 Adapting Bootstrap Method to Test the Main and Interaction Effects

Suppose that a set of elementary effects, $\mathbf{d}_j = (d_{j;1}, ..., d_{j;N})$, are obtained by random sampling from an unknown distribution F_j through (1). We want to test whether the mean \overline{d}_j , the unbiased estimator of μ_j , is less than Δ_{EE} , i.e.

$$H_0: \mu_j \leq \Delta_{\rm EE} \text{ vs. } H_1: \mu_j > \Delta_{\rm EE}, \tag{6}$$

where \leq implies that we use one-sided hypothesis and Δ_{EE} is used to define the threshold of insignificant elementary effects. An individual factor is declared unimportant if its elementary effect is $\leq \Delta_{\text{EE}}$.

To determine whether factor j interacts with other factors, we check if the sample variance S_j^2 , an unbiased estimator of σ_j^2 , is less than Δ_{IE}^2 . Similar to (6), we define the upper-tail test:

$$H_0: \sigma_j^2 \leq \Delta_{\rm IE}^2 \text{ vs. } H_1: \sigma_j^2 > \Delta_{\rm IE}^2, \tag{7}$$

where Δ_{IE} is similar to Δ_{EE} , indicating the threshold of a significant interaction effect. An individual factor is declared to have insignificant interaction effect if its standard deviation is larger than Δ_{IE} .

Since we test the main and the interaction effect for each factor, we use s=1 and s=2 to denote the main effect and the interaction effect, respectively. We use test statistic, $t_j^{(s)}$, to evaluate hypotheses (6) and (7). Due to little knowledge on the distribution of d_j , the population distribution F_j is unknown as well. Bootstrap enables us to approximate the null distribution \hat{F}_j , which is an empirical distribution function (EDF) of $F_j^{(s)}$. If the null hypothesis is true, we can then use the following test statistics to conduct hypotheses test for (6) and (7) respectively:

$$t_{j}^{(1)} = \frac{\overline{d}_{j} - \Delta_{\rm EE}}{S_{j}/\sqrt{N}}, \ j = 1, \dots, k,$$
(8)

and

$$t_{j}^{(2)} = \frac{(N-1)S_{j}^{2}}{\Delta_{\text{IE}}^{2}}, \ j = 1, \dots, k,$$
(9)

where $t_j^{(1)}$ and $t_j^{(2)}$ respectively follow the t and χ_{N-1}^2 distributions; and S_j^2 is the sample variance of factor j. We denote $t_{0;j}^{(s)}$ as the observed value corresponding to (8) and (9), derived from the original sample \mathbf{d}_j .

To apply (8) and (9), we should make certain the EDF \hat{F}_j formed by bootstrap samples is an appropriate estimate of F_j , and \hat{F}_j satisfies the null hypothesis. Specifically, both $t_j^{(1)}$ and $t_j^{(2)}$ must be asymptotic pivots (i.e., independent of parameters Δ_{EE} and Δ_{IE}), which according to MacKinnon (2009) are necessary to arrive at a more accurate test result under bootstrap. However, because the mean and the standard deviation of \hat{F}_j may differ from Δ_{EE} and Δ_{IE} , \hat{F}_j may not be a desired estimate for F_j . Thus, to obtain the desired mean and standard deviation for \hat{F}_j , we employ the following conversion process:

$$\mathbf{d}_{j}^{(1)} = \mathbf{d}_{j} - \overline{d}_{j} + \Delta_{\mathrm{EE}},\tag{10}$$

where $\mathbf{d}_{j}^{(1)}$ is the transformed vector based on the original sample \mathbf{d}_{j} . Similarly, we employ the following transformation process for (7),

$$\mathbf{d}_{j}^{(2)} = \frac{\left(\mathbf{d}_{j} - \bar{d}_{j}\right)\Delta_{\mathrm{IE}}}{S_{j}},\tag{11}$$

where the transformed sample $\mathbf{d}_{j}^{(2)}$ have variance Δ_{IE}^{2} .

From $\mathbf{d}_{j}^{(s)} = (d_{j;1}^{(s)}, ..., d_{j;N}^{(s)})$, we sample a set of bootstrapped data $\mathbf{d}_{j}^{(s)*} = (d_{j1}^{(s)*}, ..., d_{j;N}^{(s)*})$ using (10) and (11). For each bootstrap sample, we compute the test statistic for (6)

$$t_{j}^{(1)*} = \frac{\overline{d}_{j}^{(1)*} - \Delta_{\rm EE}}{S_{j}^{*}/\sqrt{N}} \ j = 1, \dots, k,$$
(12)

where $t_j^{(1)*} = t(\mathbf{d}_j^{(1)*})$, and $S_j^* = S(\mathbf{d}_j^{(1)*})$ is the standard deviation of a bootstrap sample. Similar to (12), we now derive a bootstrap test statistic for (7),

$$t_{j}^{(2)*} = \frac{(N-1)S_{j}^{*2}}{\Delta_{\rm IE}^{2}}.$$
(13)

We replicate this bootstrap approach *B* times through Monte Carlo experiment to obtain bootstrap estimates $t_{j;1}^{(s)*}, ..., t_{j;B}^{(s)*}$. In this research, we made B = 3,000 replications, to ensure our test power. Let subscript (*b*) be the *b* th ordered value with b = 1, ..., B. We sort these estimates in ascending order $(t_{j;1}^{(s)*} \leq \cdots \leq t_{j;(B)}^{(s)*})$, so that the EDF of $t_j^{(s)*}$ puts probability of 1/B on $t_{j;b}^{(s)*}$. The *p*-value of bootstrap test based on the original data is

$$\hat{p}_{j}^{(s)*} = 1 - \hat{F}_{j}^{(s)*} = \frac{\sum_{b=1}^{B} I(t_{j;b}^{(s)*} > t_{0;j}^{(s)})}{B}, \qquad (14)$$

where $\hat{p}_{j}^{(s)*} = \hat{p}(t_{0;j}^{(s)*})$ is the marginal significance level of $t_{j;0}^{(s)}$. $\hat{F}_{j}^{(s)*}$ denotes the EDF of $\mathbf{d}_{j}^{(s)*}$; while $I(\cdot)$ is an indicator function, which is equal to 1 if the argument in the bracket is true, and 0 otherwise. Hence, the bootstrap *p*-value is simply the proportion of the bootstrap test statistics $t_{j;b}^{(s)*}$ that are more extreme than the observed test statistic $t_{0;j}^{(s)}$.

3.3 Controlling the False Discovery Rate

To control the FDR at a specific level, say γ , we blend bootstrapped *p*-values together. This strategy is called aggregated/pooled FDR because we examine all tests together. We start with a pre-determined level $\gamma \in (0,1)$. Next, we mesh $\hat{p}_1^{(1)*}, ..., \hat{p}_k^{(1)*}$ with $\hat{p}_1^{(2)*}, ... \hat{p}_k^{(2)*}$, which are obtained from (14) with s = 1 and s = 2, respectively. For convenience, we drop superscripts ⁽¹⁾ and ⁽²⁾; then the original observed *p*-values become $\hat{p}_1^*, ..., \hat{p}_{2k}^*$. Let $\hat{p}_{(1)}^* \leq ... \leq \hat{p}_{(2k)}^*$ be the *p* values in ascending order. Following Benjamini and Hochberg (1995) (BH procedure), we are able to control the FDR by rejecting null hypotheses from the minimum $\hat{p}_{(1)}^*$ to $\hat{p}_{(j)}^*$, where

$$\hat{j} = \max\left\{j \in \{1, ..., 2k\}: \ \hat{p}_{(j)}^* \leq \frac{\gamma j}{k}\right\}.$$
 (15)

However, this procedure actually controls the FDR at level $\pi_0 \gamma$, where $\pi_0 \in [0,1]$, indicating the resulting FDR is more stringent as $\pi_0 \gamma \leq \gamma$. To achieve a desired FDR level, we make use of the *q*-values method (Storey 2002, 2003; Storey et al. 2004) to ensure FDR is at the γ level exactly. The estimates of the *q*-values are reached by adopting the recursive formula below:

$$\hat{q}_{(j)}^{*} = \min\left\{\frac{2\hat{\pi}_{0}k\hat{p}_{(j)}^{*}}{j}, \ \hat{q}_{(j+1)}^{*}\right\},\tag{16}$$

where $\hat{\pi}_0$ is an estimate of π_0 and $\hat{q}^*_{(j)} = \hat{\pi}_0 p_{(j)}$ when j = 2k. Loosely speaking, the *q*-value obtained from (16) is the minimum possible FDR for declaring the corresponding hypothesis significant. Thus, controlling FDR at level γ is equivalent to declare all tests with $\hat{q}^*_{(j)} \leq \gamma$ (j = 1, ..., 2k) significant.

4 NUMERICAL EVALUATION

4.1 Comparison of CMM and MM

To evaluate the proposed screening method, we first consider the low-order polynomial model presented in Morris (1991):

$$y = \beta_0 + \sum_{j=1}^{20} \beta_j w_j + \sum_{i(17)$$

where $w_j = 2(1.1x_j/(x_j + 0.1) - 0.5)$ are for factor j = 3, 5, 7; and $w_j = 2(x_j - 0.5)$ are for the remaining factors. x_j is normalized and uniformly distributed between [0,1]; as stated in Section 2.2. Coefficients of $\beta = [\beta_j, \beta_{i;j}, \beta_{l;i;j}, \beta_{s;l;i;j}]$ are specified at $\beta_j = 20$, for $1 \le j \le 10$; $\beta_{i;j} = -15$, for $1 \le i < j \le 6$; $\beta_{l;i;j} = -10$, for $1 \le l < i < j \le 5$; and $\beta_{s;l;i;j} = 5$, for $1 \le s < l < i < j \le 4$. The rest of the first- and second-order coefficients are randomly generated from a normal distribution with mean zero and unit standard deviation. The remainder of the third- and fourth-order coefficients are set to 0.

We examine the *efficacy* of CMM, under different Δ_{EE} and Δ_{IE} , both of which vary from 5 to 20 in increment of 5. To determine the significant effects, we use the aggregated FDR algorithm in equation (16). Table summarize the results of \hat{q}_{EE}^* and \hat{q}_{IE}^* of the 20 factors, where the q values less than the specified FDR level 0.05 are classified as significant.

4.2 Comparison of CMM and MSB

To compare CMM and MSB, we use the second numerical experiment in Shi et al. (2014), which is selected as it studied large-scale screening problems and compared SB and MSB. The metamodel used is

$$y^{(l)} = \beta_0^{(l)} + \sum_{j=1}^k \beta_j^{(l)} x_j + \sum_{i< j}^{k-1} \sum_{i=j+1}^k \beta_{i,j}^{(l)} x_i x_j + \sum_{j=1}^k \beta_{j;j}^{(l)} x_j^2 + \varepsilon^{(l)}, \ l = 1, ..., n,$$
(18)

where k = 100 factors and n = 2 responses.

From our experiments, we find except for (i) the noise of responses $\epsilon^{(i)}$, the efficiency of CMM is unaffected by (ii) the number of important factors, (ii) sparsity of effects, as well as (iii) the existence of Table 2: a-values under various Δm and Δm of CMM method

	$\Delta_{\rm E}$ and $\Delta_{\rm E}$ of civity method													
	$\Delta_{ ext{EE}}$:	=5	$\Delta_{ ext{EE}}$:	=5	Δ_{EE} =	=10	$\Delta_{ m EE}$ =	= 10	Δ_{EE} =	=15	$\Delta_{ ext{EE}}$:	= 15	Δ_{EE} =	= 20
Input	Δ_{IE} =	= 5	Δ_{IE} =	= 10	Δ_{IE} =	= 10	Δ_{IE} =	= 15	$\Delta_{\mathrm{IE}} =$	15	$\Delta_{ ext{IE}}$ =	=20	Δ_{IE} =	= 20
	${\hat q}_{ m EE}^*$	${\hat q}_{ m IE}^*$	${\hat q}_{ m EE}^{*}$	${\hat q}_{ m IE}^*$	${\hat q}_{ m EE}^{*}$	${\hat q}_{ m IE}^*$	${\hat q}_{ m EE}^{*}$	${\hat q}_{ m IE}^*$	$\hat{q}^*_{ ext{EE}}$	${\hat q}_{ m IE}^*$	${\hat q}_{ m EE}^*$	${\hat q}_{ m IE}^*$	$\hat{q}^*_{ ext{EE}}$	$\hat{q}_{ ext{IE}}^{*}$
#1 ^A	0.001	0	0.014	0	0.320	0	0.548	0	0.810	0	1	0	1	0
$\#2^{\text{A}}$	0	0	0	0	0.696	0	0.019	0	1	0	1	0	1	0
$\#3^{B}$	0.001	0	0	0	0.009	0	0	0	0	0	1	0	1	0
$#4^{\text{A}}$	0.003	0	0.014	0	0.021	0	0.150	0	0.605	0	1	0	1	0
$\#5^{\text{ C}}$	0	0	0	0	0	0	0	0	0	0	0.263	0	1	0
$\#6^{D}$	0	0	0	0	0.014	0	0.006	0	1	0	0.720	0	1	0
$\#7^{E}$	0	1	0	1	0	1	0	1	0	1	0.189	1	0.840	1
$\#8^{F}$	0	1	0	1	0	1	0	1	0	1	0	1	0	1
$\#9^{\mathrm{F}}$	0	1	0	1	0	1	0	1	0	1	0	1	0	1
$\#10^{\mathrm{F}}$	0	1	0	1	0	1	0	1	0	1	0	1	0	1
$\#11^{\text{ G}}$	1	1	1	1	1	1	1	1	1	1	1	1	1	1
$\#12^{\mathrm{G}}$	1	1	1	1	1	1	1	1	1	1	1	1	1	1
$\#13^{\rm \ G}$	1	1	1	1	1	1	1	1	1	1	1	1	1	1
$\#14^{\text{G}}$	1	1	1	1	1	1	1	1	1	1	1	1	1	1
$\#15^{\rm \;G}$	1	1	1	1	1	1	1	1	1	1	1	1	1	1
$\#16^{\rm \ G}$	1	1	1	1	1	1	1	1	1	1	1	1	1	1
$\#17^{\mathrm{G}}$	1	1	1	1	1	1	1	1	1	1	1	1	1	1
$\#18^{\rm G}$	1	1	1	1	1	1	1	1	1	1	1	1	1	1
$\#19^{\rm G}$	1	1	1	1	1	1	1	1	1	1	1	1	1	1
$\#20^{\text{G}}$	1	1	1	1	1	1	1	1	1	1	1	1	1	1

clustering of important factors, whereas both SB and MSB are strongly affected by these characteristics. We also find that CMM requires more simulation run time than SB and MSB, when the signs of main effects are known and heredity assumption holds.

In addition to efficiency, we also compare the efficacy between CMM and MSB. Like Shi et al. (2014) and Wan et al. (2010), the efficacy of CMM is also quantified by the fraction of trails, $\widehat{\Pr}$ {DI}. Due to space limitation, Table 2 only summarizes the results of $\widehat{\Pr}$ {DI} for combination 2, 7, 9 and 14, as similar results are observed in the remaining 12 combinations. We can see that CMM provides appealing screening results for both main effects and interaction effects.

The left column of Figure 1 demonstrates $\widehat{\Pr} \{DI\}$ for main effects, where the *x*-axis represents the number of factor ranging from 1 to 100 factors and the *y*-axis gives $\widehat{\Pr} \{DI\}$. We can see that $|\beta_j|$ has great bearing on the performance of $\widehat{\Pr} \{DI\}$; i.e. the larger $|\beta_j|$ is, the higher $\widehat{\Pr} \{DI\}$ is close to 1. Specifically, (i) when $|\beta_j| = 0 < \Delta_{EE} = 2$, $\widehat{\Pr} \{DI\}$ is considerably less than the type-I error $\alpha = 0.05$ and approximates 0; see the symbols " \bullet " in each plot. (ii) When $|\beta_j| = \Delta_{EE} = 2$, $\widehat{\Pr} \{DI\}$ approaches $\alpha = 0.05$; see the symbols " \bullet " in each plot. (iii) When $|\beta_j| > \Delta_{EE} = 2$, $\widehat{\Pr} \{DI\}$ increases as $|\beta_j|$ increases from 3 to 5. Contrasting to the locale of $\widehat{\Pr} \{DI\}$ for the same $|\beta_j|$ in Shi et al. (2014), we find the results of CMM

agrees with that of SB. Therefore, although CMM does not control type-II error directly, it provides comparably desired control error.



Figure 1: $\widehat{\Pr}$ (DI) of CMM for main and interaction effects in combination 2, 7, 9, 14

We also examine $\widehat{\Pr}$ {DI} for interaction effects. Because SB is incapable of giving screening results for interaction effects, we only show the results for CMM. Since all $\beta_{i:j}$ (i < j = 1, ..., 100) are generated from Normal(0,1), CMM should only reach a unique conclusion for all factors, either carrying significant interaction effect or no interactions. The right column of Figure 1 plots $\widehat{\Pr}$ {DI} of interaction effects for the 100 factors with Δ_{IE} =10. We can see that, regardless of combinations 2, 7, 9 or 14, all factors have nearly identical, i.e. $\widehat{\Pr}$ {DI} \approx 1. Even when Δ_{IE} takes a large value 10 relative to $\Delta_{EE} = 2$, CMM can still show significant interaction effects. The reason is that for factor j, although each individual $\beta_{i:j}$ is small,

their aggregated effect can reach or even exceed the threshold $\Delta_{\text{IE}} = 10$, causing CMM to identify them as important eventually.

5 CONCLUSION

CMM is an error control procedure that is capable of identifying both main effects and interaction effects simultaneously. Numerical evaluation shows that the performances of CMM are efficient and effective across various simulation settings. To further improve simulation efficiency, we will focus on optimal clustering matrix design in our future research. We will also develop stratified FDR control procedure to enhance statistical power.

ACKNOWLEDGMENTS

This work is partly supported by the National Natural Sciences Foundation of China under Grants No 71402048 and 71372134, and the China Postdoctoral Science Foundation Funded Project No 2015M582228.

REFERENCES

- Benjamini, Y., and Y. Hochberg. 1995. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society. Series B* (*Methodological*) 57 (1): 289-300.
- Bettonvil, B. W., and J. P. C. Kleijnen. 1997. "Searching for Important Factors in Simulation Models with Many Factors: Sequential Bifurcation." *European Journal of Operational Research* 96 (1): 180-194.
- Borgonovo, E., and E. Plischke. 2016. "Sensitivity Analysis: A Review of Recent Advances." *European Journal of Operational Research* 248 (3): 869-887.
- Boukouvalas, A., J. P. Gosling, and H. Maruri-Aguilar. 2014. "An Efficient Screening Method for Computer Experiments." *Technometrics* 56 (4): 422-431.
- Campolongo, F., and R. Braddock. 1999. "The Use of Graph Theory in the Sensitivity Analysis of the Model Output: A Second Order Screening Method." *Reliability Engineering & System Safety* 64 (1): 1-12.
- Campolongo, F., J. Cariboni, and A. Saltelli. 2007. "An Effective Screening Design for Sensitivity Analysis of Large Models." *Environmental Modelling & Software* 22 (10): 1509-1518.
- Russell C. H. Cheng. 1997. "Searching for important factors: sequential bifurcation under uncertainty". *In Proceedings of the 1997 Winter Simulation Conference*, edited by S. Andradóttir, K. J. Healy, D. H. Withers, and B. L. Nelson, 275-280. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Cropp, R. A., and R. D. Braddock. 2002. "The New Morris Method: An Efficient Second-Order Screening Method." *Reliability Engineering & System Safety* 78 (1): 77-83.
- Fédou, J.-M., and M. J. Rendas. 2015. "Extending Morris Method: Identification of the Interaction Graph Using Cycle-Equitable Designs." *Journal of Statistical Computation and Simulation* 85 (7): 1398-1419.
- Kleijnen, J. P. C. 2015. Design and Analysis of Simulation Experiments. 2ed. New York: Springer US
- MacKinnon, J. G. 2009. Bootstrap Hypothesis Testing, Handbook of Computational Econometrics.
- Morris, M. D. 1991. "Factorial Sampling Plans for Preliminary Computational Experiments." *Technometrics* 33 (2): 161-174.
- Shi, W., J. P. C. Kleijnen, and Z.-X. Liu. 2014. "Factor Screening for Simulation with Multiple Responses: Sequential Bifurcation." *European Journal of Operational Research* 237 (1): 136-147.

- Storey, J. D. 2002. "A Direct Approach to False Discovery Rates." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64 (3): 479-498.
- Storey, J. D. 2003. "The Positive False Discovery Rate: A Bayesian Interpretation and the Q-Value." Annals of statistics 31 (6): 2013-2035.
- Storey, J. D., J. E. Taylor, and D. Siegmund. 2004. "Strong Control, Conservative Point Estimation and Simultaneous Conservative Consistency of False Discovery Rates: A Unified Approach." *Journal* of the Royal Statistical Society: Series B (Statistical Methodology) 66 (1): 187-205.
- Sun, L., R. V. Craiu, A. D. Paterson, and S. B. Bull. 2006. "Stratified False Discovery Control for Large-Scale Hypothesis Testing with Application to Genome-Wide Association Studies." *Genetic Epidemiology* 30 (6): 519-530.
- Wan, H., B. E. Ankenman, and B. L. Nelson. 2006. "Controlled Sequential Bifurcation: A New Factor-Screening Method for Discrete-Event Simulation." *Operations Research* 54 (4): 743-755.
- Wan, H., B. E. Ankenman, and B. L. Nelson. 2010. "Improving the Efficiency and Efficacy of Controlled Sequential Bifurcation for Simulation Factor Screening." *INFORMS Journal on Computing* 22 (3): 482-492.

AUTHOR BIOGRAPHIES

WEN SHI is an Associate Professor in School of Logistics and Engineering Management at Hubei University of Economics. His research focuses on simulation modeling, DOE, supply chain management, and simulation optimization. His email address is shi3wen@163.com.

JENNIFER SHANG is a professor of business administration and arear director at Katz Graduate School of Business, University of Pittsburgh. She has won several outstanding research and teaching awards, and her main research interests include the design, planning, scheduling, and control of operational systems in manufacturing and service organizations. She conducts simulation for design and evaluation of integrated information/operational systems. Her email address is shang@katz.pitt.edu.

ZHANG ZHIGANG is an Associate Professor in the School of Statistics at Hubei University of Economics. His research interests lie in statistical modeling, time series analysis and statistical methods for big data. His email address is zzg@hbue.edu.cn.