A MODIFICATION OF CHENG'S METHOD: AN ALTERNATIVE FACTOR SCREENING METHOD FOR STOCHASTIC SIMULATION MODELS

Reza Yaesoubi

Harvard Medical School
641 Huntington Ave.
Boston, MA 02115, U.S.A.

Stephen D. Roberts

Edward P. Fitts Department of
Industrial and Systems Engineering
North Carolina State University
Campus Box 7906
Raleigh, NC 27695, U.S.A.

Robert W. Klein

Medical Decision Modeling Inc.
3600 Woodview Trace, Suite 317
Indianapolis, IN 46268, U.S.A.

ABSTRACT

Factor Screening experiments identify those factors with significant effect on a selected output. We propose a modification of Cheng's method as a new factor screening alternative for simulation models whose output has homogeneous variance and can be described by a second-order polynomial function. The performance of the proposed model is compared with several other factor screening alternatives through an empirical evaluation. The results show that the proposed method sustains its efficiency and accuracy as the number of factors or the homogeneous variance increases. However, its accuracy degrades as variance heterogeneity increases.

1 INTRODUCTION

Simulation models can consist of a large number of *inputs*, which are usually referred to as *factors* in Design of Experiments (DOE). Exploring the response (output) of large-scale and complex simulation models can become prohibitively costly and requires many time-consuming simulation runs. Factor Screening experiments, which assume that only a few factors are really important (*parsimony principle*), systematically examine the impact of input factors on a selected simulation response to identify those factors with significant effects.

Factor screening (Bettonvil and Kleijnen 1996) is generally employed in the *pilot phase* of complex simulation studies. The factors identified as being important can be further explored in later phases; e.g., the important factors might be cast as a metamodel and used in optimization. Moreover, the results of factor screening can be used not only for confirming prior expectations (which is an important step in validating the simulation models), but they are also informative when the simulation provides insights that do not match expectations. Many real-world models are characterized by high degree of uncertainty, and hence demand comprehensive datasets to accurately estimate the input parameters. Such data are often unavailable in practice. For instance, medical decision-making models generally suffer from inaccuracy in their input parameter estimates, simply due to the lack of sufficient data for obtaining accurate estimates. Performing factor screening analysis on the models created in such environments can help identify the influential factors in the models, and consequently, prioritize the data collection efforts giving emphasis to those input factors with greater significance.

Most of the factor screening methods introduced in recent years are created on the base of classical experimental design (Montgomery 2000) and sequential bifurcation (Bettonvil and Kleijnen 1996). Among those, 2^k Fractional Factorial design (Montgomery 2000), Cheng's Method (Cheng 1997), and fold-over Controlled Sequential Bifurcation (CSB-X) (Wan, Ankenman, and Nelson 2006b) have received the most attention. The reader is referred to Kleijnen et al. (2005) for a comprehensive review on designing simulation experiments. In this paper, we present a new factor screening method, called MCh-X, which is a modification of Cheng's method. MCh-X relaxes some of the restrictive assumptions

of Cheng's method in order to work under conditions posed by more complex simulation models. We compare the performance of MCh-X along with 2^k Fractional Factorial designs of resolution IV and CSB-X based on various criteria of efficiency, accuracy, and cost-effectiveness (achieving the highest accuracy with the least number of observations).

The remainder of this paper is organized as follows. Section 2 presents a literature review on the commonly-used factor screening methods: 2^k Fractional Factorial Design of resolution IV, Cheng's method, and the fold-over Controlled Sequential Bifurcation. In §3, we introduce a modification of Cheng's method (MCh-X) as a new alternative factor screening method. Section 4 compares the performance of the factor screening methods studied here. Section 5 concludes the paper and discusses future research.

2 LITERATURE REVIEW

Among the most popular factor screening methods are 2^k Fractional Factorial designs of resolution IV (2^k FF-IV) (Montgomery 2000). In these designs, no main effects are confounded with any other main or second-order interaction effects but the second-order interaction effects can be confounded with each other.

These designs assume that certain higher-order interactions are negligible so they can be confounded with main or second-order interaction effects. These designs also assume that errors are approximately normally distributed with mean zero and a homogeneous variance σ^2. Factorial designs have several attractive properties. First, they are very easy to construct and already available in almost all statistical software. Second, since they examine more than one factor at a time, these designs can identify certain important interaction effects. Third, they are *orthogonal* designs, i.e. the pair-wise correlation between any two columns (factors) in the design matrix is zero. This simplifies the analysis of the fitted regression model, because the estimates of the factors' effects and their contributions to the explanatory power of the regression model do not depend on other regressors included in the model. Finally, although ordinary least squares regression assumes that the error term is identically distributed, the regression coefficients are still unbiased estimators even if the underlying variance is not constant.

Sequential Bifurcation (SB), introduced by Bettonvil and Kleijnen (1996), was mainly designed for deterministic simulation models, and based on group screening designs (Johnson and Gastwirth 1999, Kleijnen 1987). SB, like all other group screening designs, assumes a low-order polynomial metamodel for the response function of the simulation model and known signs for the main effects. Knowing the main effects' signs assures that individual effects do not compensate for the effects of each other within a group. The criterion that SB uses for declaring a factor as important is the absolute value of the factor's main effect. Cheng (1997) extended SB to handle simulations models with stochastic response. His method assumes a first-order polynomial for the response function and a homogeneous variance over the entire experimental region. Since Cheng's method assumes known signs of main effects (β_j), the factor's upper and lower bounds should be determined such that $\beta_j \geq 0$, for all i. Cheng's method requires two parameters from a user: the delta limit (δ), which is the value that the main effect of a factor should reach to be considered important, and alpha (α), which is the probability of declaring an important factor unimportant.

Cheng's method does not control the probability of declaring an unimportant factor important, or the probability of declaring an important factor important at any step. Controlled Sequential Bifurcation (CSB), proposed by Wan, Ankenman, and Nelson (2003), Wan, Ankenman, and Nelson (2006a), is a procedure that incorporates a two-stage hypothesis-testing approach into SB to control the probability of type I error (i.e., the probability that an effect is classified as important when it is not) and power (i.e., the probability that an important effect is correctly classified). CSB controls the power at each bifurcation step and the type I error for each factor under *heterogeneous* variance conditions. In CSB, the analyst must specify two thresholds. The lower threshold (Δ_0) indicates the level that the main effect of a factor must reach to be considered *important*. The higher threshold (Δ_1) is the level that is considered critical when exceeded by a factor's main effect. In other words, if $\beta_j \leq \Delta_0$, then factor j is classified as unimportant and if $\beta_j \geq \Delta_1$, factor j is classified as critical, and otherwise if $\Delta_0 < \beta_j < \Delta_1$, it is classified as important.

From an error perspective, for those factors with effects $\leq \Delta_0$, CSB controls the type I error by declaring them important with probability less than α; and for those factors with effects $\geq \Delta_1$, CSB detects them with power greater than γ. Those factors whose effects fall between Δ_0 and Δ_1 are considered important and the CSB procedure has reasonable, though not guaranteed, power to identify them (Wan, Ankenman, and Nelson 2003). Furthermore, in CSB, dependence of outputs

across different factor settings due to common random numbers is permitted. For the special case, where $\alpha = 1 - \gamma$ (type I error is equal to one minus power), Wan, Ankenman, and Nelson (2006a) implemented a fully sequential test in CSB that has the same error control as the two-stage testing procedure. Wan, Ankenman, and Nelson (2006b) improved the CSB procedure by incorporating a fold-over design in the hypothesis test to identify important main effects even when two-factor interactions and quadratic terms are present. The new procedure, called CSB-X, still has the same error control for screening main effects. However, CSB-X is unable to estimate interaction and quadratic effects.

Since CSB-X controls both type I error and power to remain in the acceptable range, its efficiency decreases noticeably when applied to a model with a large variance. An example of such a case is provided in §4.1. For models with a high variance, however, if the assumption of homogeneous variance is still satisfied, then Cheng's method can be augmented with a few modifications to achieve the same degree of accuracy as CSB-X and still be desirably efficient. Details of such modifications, which result in a new factor screening method called MCh-X, are explained in the following section.

Among factor screening methods not discussed in this paper are Latin Hypercube (LH) designs and supersaturated designs (McKay, Beckman, and Conover 1979, Hinkelmann and Kempthorne 2005). Generating orthogonal LH designs with good space-filling property can be challenging and imposing restriction on the maximum number of factors that can be included in the experiment (for further discussion, refer to Steinberg and Lin (2006), Cioppa and Lucas (2007), Bingham, Sitter, and Tang (2009)). Likewise, generating and analyzing supersaturated designs pose restrictions and complications, which are outside the scope of this paper (for more information, refer to Hinkelmann and Kempthorne (2005), Georgiou (2008), Lin (1995), Eskridge et al. (2004)).

In addition, we do not investigate the two most recent derivatives of CSB: Controlled Sequential Factorial Design (CSFD) (Shen and Wan 2005) which combines sequential hypothesis-testing procedures with the traditional factorial design to control the Type I Error and power for each factor under heterogeneous variances conditions; and hybrid CSB-CSFD (Shen and Wan 2006) which employs a two-stage hybrid approach to combine CSB and CSFD to achieve better overall performance. We leave further investigation of these methods for future research.

3 A MODIFICATION OF CHENG'S METHOD

In terms of eliminating insignificant factors, Cheng's method is lenient in detecting the unimportant factors; that is, even if a main effect of a factor (or a group-factor) is less than δ, it is possible that the method does not eliminate it. Although this property does not affect the correctness of the method, it is counter to the fundamental assumption of most factor screening methods that only a *few* factors are important among *many* potentially important factors; hence it generally leads to declaring some unimportant factors important.

The CSB-X is the only sequential method that works with high degree of accuracy when variance is heterogeneous. Yet, as discussed in §4, it loses its efficiency rapidly as the (homogeneous or heterogeneous) variance increases. Therefore, we propose a modification of Cheng's method (MCh-X) in pursuit of enhancing the performance of Cheng's method to work efficiently as homogeneous variance increases.

MCh-X differs from the original Cheng's method in the following aspects: (1) MCh-X is based on testing the hypothesis that each factor is unimportant unless there a strong evidence to the contrary. (2) Cheng's method assumes a linear function for the response, which is rarely the case for many simulation models. MCh-X assumes a second-order polynomial function and uses a fold-over design to cope with the circumstances where the second-order interactions are not negligible. MCh-X is however unable to estimate these second-order interaction effects.

To quantify importance, MCh-X uses three parameters: significance level α, and two importance parameters δ and τ. The parameter δ is the value that a factor's main effect (or a group's main effect) should reach to be considered important. At any iteration, if a non-singleton group of factors has a main effect statistically greater than δ at a significance level α, then the group is categorized as important; otherwise the group will be eliminated and marked as unimportant. To classify a single factor, the procedure declares the factor important (unimportant) if its main effect is statistically greater (less) than δ at a significance level α; if such a decision cannot be made with the current number of observations, the procedure acquires more observations until the estimated half-width of the factor's main effect becomes less than the parameter τ at a significance level α. Therefore, the parameter τ can be used to determine the precision of the estimated main effects. A low value for τ forces the procedure to estimate the main effects with higher precision, and thus requires more observations.

From an error perspective, MCh-X may misclassify an unimportant group factor (i.e., a group with main effect less than δ) with probability α. Furthermore, when classifying a single factor, it may miss an important factor with main effect greater than $\delta + \tau$ with probability at most α. Like the original Cheng's method, MCh-X also assumes that the error term has homogeneous variance and the directions of the main effects are known.

The MCh-X assumes the following response function (y) around a given point $\mathbf{x} = (x_1, \ldots, x_K)$:

$$y = \beta_0 + \sum_{j=1}^{K} \beta_j x_j + \sum_{j=1}^{K-1} \sum_{j'=j+1}^{K} \beta_{j,j'} x_j x_{j'} + \sum_{j=1}^{K} \beta_{j,j} x_j^2 + e(\mathbf{x}), \tag{1}$$

where

  K: total number of factors in the experiment;
  x_j: value of factor j, standardized to lie in $[-1, +1]$;
  β_j: first-order or main effect of factor $j = 1, \ldots, K$ (β_0 is the intercept);
  $\beta_{j,j'}$:   interaction effect of the factors j and j', $1 \le j < j' \le K$;
  $\beta_{j,j}$:   pure second-order or quadratic effects of factor $j = 1, \ldots, K$;
  $e(\mathbf{x})$:   error term, assumed to be normally distributed with mean zero and constant variance σ^2.

It is also assumed that for all j, $\beta_j \ge 0$. In order to satisfy this assumption, for factor j, we can define its upper level (H_j) and lower level (L_j) such that $\beta_j \ge 0$.

Let $y^{(j)}$ denote the response value when the factors $1, \ldots, j$ are set at their upper levels (H_j) and the remaining factors ($j+1, \ldots, K$) are set at their lower levels (L_j). Therefore, in the standardized metamodel (1), the values of the first j factors are $+1$ and the values of the remaining factors are -1. Thus polynomial (1) yields:

$$y^{-(j)} = \beta_0 + \beta_1 + \cdots + \beta_{j-1} + \beta_j - \beta_{j+1} - \cdots - \beta_K$$
$$+ \sum_{i < i' \le j} \sum \beta_{i,i'} + \sum_{j < i < i'} \sum \beta_{i,i'} - \sum_{i \le j < i'} \sum \beta_{i,i'} + e(\mathbf{x}), \; j = 0, 1, \ldots, K.$$

The symbol $y^{-(j)}$, called the mirror observation of $y^{(j)}$, denotes the response value when the factors $1, \ldots, j$ are set at their lower levels (L_j) and the remaining factors ($j+1, \cdots, K$) are set at their upper levels (H_j). Therefore, in the standardized metamodel (1), the values of the first j factors are -1 and the values of the remaining factors are $+1$. So the polynomial (1) yields:

$$y^{(j)} = \beta_0 - \beta_1 - \cdots - \beta_{j-1} - \beta_j + \beta_{j+1} + \cdots + \beta_K$$
$$+ \sum_{i < i' \le j} \sum \beta_{i,i'} + \sum_{j < i < i'} \sum \beta_{i,i'} - \sum_{i \le j < i'} \sum \beta_{i,i'} + e(\mathbf{x}), \; j = 0, 1, \ldots, K.$$

If $j < k$, then the scaled difference $D(j,k) = [(y^{(k)} - y^{-(k)}) - (y^{(j-1)} - y^{-(j-1)})]/4$ has the expectation $E[D(j,k)] = \sum_{i=j}^{k} \beta_i$, and therefore, $D(j,k)$ can be regarded as an estimator for the sum of the main effects for factors j to k. If $y^{(k)}, y^{-(k)}, y^{(j-1)}$ and $y^{-(j-1)}$ are independent, then $\text{Var}[D(j,k)] = \sigma^2/4$.

If we have $r^{(j-1)}$ observations at level $j-1$ and mirror level $-(j-1)$, i.e., $y_r^{(j-1)}$ and $y_r^{-(j-1)}$, for $r = 1, 2, \ldots, r^{(j-1)}$, and $r^{(k)}$ observations at level k and mirror level $-k$, i.e., $y_r^{(k)}$ and $y_r^{-(k)}$, for $r = 1, 2, \ldots, r^{(k)}$, then:

$$\bar{D}(j,k) = \left[\frac{1}{r^{(k)}} \sum_{r=1}^{r^{(k)}} (y_r^{(k)} - y_r^{-(k)}) - \frac{1}{r^{(j-1)}} \sum_{r=1}^{r^{(j-1)}} (y_r^{(j-1)} - y_r^{-(j-1)}) \right] / 4, \tag{2}$$

and

$$\text{Var}[\bar{D}(j,k)] = \frac{1}{8} \sigma^2 \left(\frac{1}{r^{(j-1)}} + \frac{1}{r^{(k)}} \right).$$

The MCh-X method proceeds as follows.

Step 0: To initiate the process we make $r^{(0)} = N_0 > 1$ runs at levels 0 and its mirror level -0, and make $r^{(K)} = N_0 > 1$ runs at level K and its mirror level $-K$. Thus at the start of Step 1 we have four sets of observations:

$$\text{Observations at level } 0 \quad : \quad \{y_r^{(0)} : r = 1, ..., r^{(0)}\}, \tag{3}$$

$$\text{Observations at mirror level } -0 \quad : \quad \{y_r^{-(0)} : r = 1, ..., r^{(0)}\}, \tag{4}$$

$$\text{Observations at level } K \quad : \quad \{y_r^{(K)} : r = 1, ..., r^{(K)}\}, \tag{5}$$

$$\text{Observations at mirror level } -K \quad : \quad \{y_r^{-(K)} : r = 1, ..., r^{(K)}\}. \tag{6}$$

We place all the coefficients in a single, unclassified set $G_1 = \{\beta_1, \beta_2, ..., \beta_K\}$. Thus initially the number of sets is $p_1 = 1$. We assign 1 to s and proceed to the next step.

Step s $(s > 0)$: At the beginning of Step s, factors are partitioned into p_s sets:

$$G_{i,s} = \{\beta_j : k_{(i-1),s} < j \leq k_{i,s}\}, \text{ for } i = 1, 2, ..., p_s. \tag{7}$$

For example, at Step 1 ($s = 1$ and $p_1 = 1$), there is only one set of factors and Expression (7) results in $G_{i,1} = \{\beta_j : k_{(i-1),1} < j \leq k_{i,1}\}$, $i = 1$, where $k_{0,1} = 1$ and $k_{1,1} = K$. Thus $G_{1,1} = \{\beta_1, \beta_2, ..., \beta_K\}$. At the beginning of Step s, we have $2 \times (p_s + 1)$ sets of observations:

$$\text{Observations at level } k_{i,s} : \{y_r^{(k_{i,s})} : r = 1, ..., r^{(k_{i,s})}\}, \text{ for } i = 1, ..., p_s, \tag{8}$$

$$\text{Observations at mirror level } -k_{i,s} : \{y_r^{-(k_{i,s})} : r = 1, ..., r^{(k_{i,s})}\}, \text{ for } i = 1, ..., p_s, \tag{9}$$

$$\text{Observations at level } k_{(i-1),s} : \{y_r^{(k_{(i-1),s})} : r = 1, ..., r^{(k_{(i-1),s})}\}, \text{ for } i = 1, ..., p_s, \tag{10}$$

$$\text{Observations at mirror level } -k_{(i-1),s} : \{y_r^{-(k_{(i-1),s})} : r = 1, ..., r^{(k_{(i-1),s})}\}, \text{ for } i = 1, ..., p_s. \tag{11}$$

At this point, some of the sets are already classified, some are unclassified. If all sets have been classified, then the algorithm ends. Otherwise, we select any unclassified set; call this $G_{i,s}$. Two cases may occur:

Case 1: $G_{i,s}$ is not a singleton set

If $G_{i,s}$ is not a singleton set (set of size one), we check if all the factors can be classified as unimportant. This can be done by considering the expected main effect of the set $G_{i,s}$, which is $\bar{D}(k_{(i-1),s}, k_{i,s})$. Since $e(\mathbf{x})$ is assumed to be normally distributed, $\bar{D}(k_{(i-1),s}, k_{i,s})$ is also normally distributed with mean

$$\mu(G_{i,s}) = \sum_{i=k_{(i-1),s}}^{k_{i,s}} \beta_i,$$

and variance

$$\upsilon(G_{i,s}) = \frac{1}{8} \sigma^2 \left(\frac{1}{r^{(k_{(i-1),s}-1)}} + \frac{1}{r^{(k_{i,s})}} \right).$$

Let $S_{k_{i,s}}^2$ denote the estimated variance of the observations at level $k_{i,s}$, $S_{-k_{i,s}}^2$ denote the estimated variance of the observations at mirror level $-k_{i,s}$, $S_{k_{(i-1),s}-1}^2$ denote the estimated variance of the observations at level $k_{(i-1),s} - 1$, and $S_{-(k_{(i-1),s}-1)}^2$ denote the estimated variance of the observations at mirror level $-(k_{(i-1),s} - 1)$. Pooling these estimates gives an overall estimator of σ^2 for set $G_{i,s}$ as:

$$S_{G_{i,s}}^2 = \frac{(r^{(k_{(i-1),s}-1)} - 1)(S_{k_{(i-1),s}-1}^2 + S_{-(k_{(i-1),s}-1)}^2) + (r^{(k_{i,s})} - 1)(S_{k_{i,s}}^2 + S_{-k_{i,s}}^2)}{2(r^{(k_{(i-1),s}-1)} + r^{(k_{i,s})} - 2)}.$$

Pooling the estimates of σ^2 over all stages gives an overall estimator of σ^2 at the start of step s as:

$$S_s^2 = \frac{\sum_{i=1}^{p_s} (r^{(k_{(i-1),s}-1)} + r^{(k_{i,s})} - 2)S_{G_{i,s}}^2}{\sum_{i=1}^{p_s} (r^{(k_{(i-1),s}-1)} + r^{(k_{i,s})} - 2)}. \tag{12}$$

Now, in order to check if the main effect of set $G_{i,s}$ is significant, we test the following one-sided hypothesis: H_0: $\mu(G_{i,s}) \le \delta$.

If H_0 is rejected, then the factor set $G_{i,s}$ will be classified as important. If H_0 is not rejected, then the implication is that the factor set $G_{i,s}$ does not have a significant effect, and hence will be eliminated from the experiment. Because rejecting H_0 is a strong conclusion, this formulation forces the factor to demonstrate that its main effect exceeds δ. In other words, this formulation assumes that the effect is unimportant unless there is strong evidence to the contrary.

If the null hypothesis H_0: $\mu(G_{i,s}) \le \delta$ is correct, then the quantity

$$t_0 = \frac{\frac{\bar{D}(j,k)-\delta}{\sqrt{\text{Var}[\bar{D}(j,k)]}}}{\sqrt{\frac{SS_s}{\sigma^2}\big/\omega}} = \frac{\frac{\bar{D}(j,k)-\delta}{\sqrt{\frac{1}{8}\sigma^2(1\big/r^{(j-1)}+1\big/r^{(k)})}}}{\sqrt{\frac{S_s^2}{\sigma^2}}} = \frac{\bar{D}(j,k)-\delta}{\frac{1}{2\sqrt{2}}S_s\sqrt{1\big/r^{(j-1)}+1\big/r^{(k)}}}$$

has a t-distribution with ω_s degrees of freedom, where

$$\omega_s = 2\sum_{i=1}^{s} (r^{(k_{(i-1),s}-1)} + r^{(k_{i,s})} - 2) \tag{13}$$

and $SS_s = \omega S_s^2$.

If $t_{\alpha,\omega}$ denotes the upper α critical point of the t-distribution with ω degrees of freedom, then if $t_0 > t_{\alpha,\omega}$, we reject the null hypothesis H_0 and conclude that the corresponding group-factor is important. Having $t_0 > t_{\alpha,\omega}$ results in

$$\bar{D}(k_{(i-1),s}, k_{i,s}) > \delta + \frac{1}{2\sqrt{2}}t_{\alpha,\omega_s}S_s\sqrt{1\big/r^{(k_{(i-1),s}-1)}+1\big/r^{(k_{i,s})}}. \tag{14}$$

Therefore, if Inequality (14) holds, with confidence $1-\alpha$, $\mu(G_{i,s}) > \delta$; or equivalently $\sum_{i=k_{(i-1),s}}^{k_{i,s}} \beta_i > \delta$. Thus, with confidence $1-\alpha$, we can consider the set $G_{i,s}$ as important.

Next, $G_{i,s}$ will be split into two groups at $k_s = \lceil (k_{(i-1),s} + k_{i,s})/2 \rceil$ and the two new sets $\{\beta_{k_{(i-1),s}}, \ldots, \beta_{k_s}\}$ and $\{\beta_{k_s+1}, \ldots, \beta_{k_{i,s}}\}$ will replace $G_{i,s}$. We then obtain the following sets of observations to add to the observation sets (8)-(11):

$$\text{Observations at level } k_s \quad : \quad \{y_r^{(k_s)} : r = 1, \ldots, r^{(k_s)} = N_0\},$$
$$\text{Observations at mirror level } -k_s \quad : \quad \{y_r^{-(k_s)} : r = 1, \ldots, r^{(k_s)} = N_0\},$$
$$\text{Observations at level } k_s + 1 \quad : \quad \{y_r^{(k_s+1)} : r = 1, \ldots, r^{(k_s+1)} = N_0\},$$
$$\text{Observations at mirror level } -(k_s+1) \quad : \quad \{y_r^{-(k_s+1)} : r = 1, \ldots, r^{(k_s+1)} = N_0\}.$$

We next increment s and go to Step $s+1$.

Case 2: $G_{i,s}$ is a singleton set

If $G_{i,s} = \{\beta_k\}$, then we proceed to fully classify β_k. This can be done by considering the expected main effect of the set $\{\beta_k\}$, $\bar{D}(k,k)$, which can be estimated by Equation 2. A two-sided $1-\alpha$ confidence interval for β_k with upper and lower limits is given by:

$$\beta_k^{\pm} = \bar{D}(k,k) \pm \frac{1}{2\sqrt{2}}t_{\alpha/2,\omega_s}S_s\sqrt{1\big/r^{(k_s-1)}+1\big/r^{(k_s)}}, \tag{15}$$

where S_s and ω_s are calculated as before according to Equations 12 and 13, respectively.

The decision rule for classifying β_k is as follows:

1. If $\beta_k^- > \delta$, then β_k is classified as important.
2. If $\beta_k^+ < \delta$, β_k is classified as unimportant.
3. Otherwise, we do not have enough information to classify β_k and more observations are needed. If

$$\frac{1}{2\sqrt{2}} t_{\alpha,\omega} S_s \sqrt{1 \Big/ r^{(k_{(i-1),s}-1)} + 1 \Big/ r^{(k_{i,s})}} < \tau \tag{16}$$

does not hold, we make additional runs at levels $k_{(i-1),s}$ and $k_{i,s}$. If $r^{(k_{(i-1),s}-1)}$ and $r^{(k_{i,s})}$ are not equal, we add runs at the level with the smaller number of runs until $r^{(k_{(i-1),s}-1)} = r^{(k_{i,s})}$ or the Inequality 16 is satisfied. If $r^{(k_{(i-1),s}-1)} = r^{(k_{i,s})}$ but Inequality 16 is still not satisfied, we increase $r^{(k_{(i-1),s}-1)}$ and $r^{(k_{i,s})}$ together until Inequality 16 becomes true. Now if $\delta < \beta_k^-$, we classify β_k as important, and otherwise we classify it as unimportant.

Once β_k is classified, we increment s and go to Step $s+1$.

To use MCh-X, we need to specify the number of observations to be obtained at each iteration, N_0. If we want to estimate the main effects within the half-width τ, then a proper value for N_0 can be calculated by Equation 15 as follows. Given an estimate for the standard error ($\hat{\sigma}$), if we assume $r^{(k_{(i-1),s}-1)} = r^{(k_{i,s})} = N_0$ and $t_{\alpha/2,\omega_s} \approx z_{1-\frac{\alpha}{2}}$, where $z_{(1-\frac{\alpha}{2})}$ is the upper $1 - \frac{\alpha}{2}$ critical value of the standard normal distribution, then any $N_0 \geq \left(\frac{2\tau}{\hat{\sigma} z_{(1-\frac{\alpha}{2})}}\right)^2$ results in the half-width being less than τ.

4 PERFORMANCE EVALUATIONS OF THE FACTOR SCREENING METHODS

In this section, the performance of the factor screening methods is discussed in terms of efficiency (cost), accuracy (effectiveness) and cost-effectiveness. This comparison can be helpful for experimenters in selecting factor screening alternatives that best suit their models' characteristics.

The number of observations that a factor screening method requires to determine the set of important factors is generally considered a measure of efficiency (Kleijnen et al. 2005). To quantify the accuracy of a method, when the set of important factors is known, we can use the percentage of the important factors declared important $(1 - \hat{\alpha})$ as a measure for accuracy .

In practice, obtaining observations from complex simulation models can be prohibitively costly. In those situations, in detecting the set of important factors, the experimenter may be willing to sacrifice a small portion of accuracy to achieve better efficiency (fewer runs). Therefore, it is also valuable to rank the factor screening methods in terms of *cost-effectiveness*, which we measure as the ratio of cost (total number of observation needed) to effectiveness (accuracy):

$$\text{Cost-effectiveness ratio} = \frac{\text{Total \# of required observations}}{1 - \hat{\alpha}}. \tag{17}$$

In situations where sacrificing a small portion of accuracy to achieve better efficiency is legitimate, the experimenter may select the factor screening method that yields the smallest cost-effectiveness ratio.

In §4.1 we study the performance of factor screening methods under different variance settings. In §4.2, we investigate how the number of factors affects the performance of these methods. The results can be used to gain a *rough* insight into the performance of each factor screening method under certain conditions; yet it should be pointed out that a method cannot be *adequately* evaluated simply by observing its performance on a number of test beds; especially since the performance of factor screening methods are highly sensitive to their parameter settings. The sole purpose of this section is to obtain better understanding about the performance of each method under certain scenario settings. Nonetheless, understanding how each method performs allows us to make recommendations about which factor screening method can be expected to perform best under various conditions.

4.1 Performance Comparison: Effect of Variance

To evaluate and compare the performance of the discussed methods on different variance scenarios, we created a test bed in Microsoft Excel 2007 with the following settings. One hundred factors

are assumed to be involved, 10 of which are randomly selected to be important. The second-order interaction effects between any factor and an unimportant factor are assumed to be unimportant. The 10 important factors add 45 second-order interaction effects, 10 of which are randomly selected to be important. The coefficient of any important effect, including main or second-order interaction effects, is uniformly distributed over the intervals $[-1, -0.7]$ and $[0.7, 1]$, whereas for the unimportant factors, the coefficients are uniformly distributed over the interval $[-0.5, 0.5]$. All other effects associated with higher order interactions are considered to be zero.

The error term for the response value y is assumed to be normally distributed with mean zero and variance $(\sigma + cy)^2$, where σ and c are constant values. We test the factor screening methods on two different variance scenarios:

Model 1: Homogeneous variance with settings: $\sigma \in \{0.25, 0.50, 0.75, 1.00\}$ and $c = 0$.
Model 2: Heterogeneous variance with settings: $\sigma \in \{0.05, 0.10, 0.15, 0.20\}$ and $c = 0.2(\sigma + 0.1)$.

For each scenario, we use 2^k FF-IV, CSB-X, and MCh-X to identify the important factors. The parameter settings for each method are as follows. For 2^k FF-IV, we set $N = 40\sigma$ for Model 1 and $N = 200\sigma$ for Model 2; for MCh-X, we set $\alpha = 0.05$, $\delta = 0.5$, $\tau = 0.6$, and $N_0 = 40\sigma$ for Model 1 and $N_0 = 200\sigma$ for Model 2; and for CSB-X, we set $\alpha = 0.1$, $\Delta_0 = 0.5$, $\Delta_1 = 0.6$, and $N_0 = 5$.

Figures 1 and 2 compare the factor screening methods in terms of the different performance measures for the scenarios with homogeneous and heterogeneous variance, respectively. In these figures, accuracy is measured by the percentage of important factors detected by each method, efficiency is measure by total number of observations required by a method for the factor screening experiment, and the cost-effectiveness is calculated according to Equation 17. All confidence intervals are calculated at 5% significance level. In each figure, a sufficient number of observations have been obtained to create non-overlapping confidence intervals. Also note that both MCh-X and CSB-X assume that the signs of main effects are known; therefore prior to applying them to each model, an inexpensive Plackett-Burnman (PB) design (Montgomery 2000) was used to determine the signs of main effects. Moreover, for the heterogeneous variance scenario, the response observations were transformed by the Cox-Box method (Montgomery 2000) before using 2^k FF-IV.

According to Figures 1 and 2, the 2^k FF-IV designs are generally efficient and accurate in determining the important factors. In the case of variance heterogeneity, the 2^k FF-IV designs can still yield accurate results if an appropriate variance-stabilizing transformation (Montgomery 2000) is applied (see Figure 2(b)).

As shown in Figures 1(b) and 2(b), CSB-X rarely misclassifies an important factor since it has control over both type I error and power at each iteration. The main advantage of CSB-X is its capability to function under variance heterogeneity. However, as the variance increases, the efficiency of CSB-X deteriorates rapidly (see Figures 1(a) and 2(a)). Hence, for simulation models with relatively high variance and costly observations, this method may become prohibitively demanding. Nonetheless, for cases where accuracy is far more important than efficiency, CSB-X is the preferred choice.

Figures 1(a) and 1(b) show that MCh-X generally yields accurate results under variance homogeneity while sustaining its efficiency as the variance increases. However, since it employs the assumption of variance homogeneity, the MCh-X tends to misclassify the factors as the variance heterogeneity becomes more pronounced (see Figure 2(b)). Moreover, MCh-X generally has a low cost-effectiveness ratio for models with homogeneous variance (see Figure 1(c)). This implies that when the simulation runs are costly, the experimenter can get fairly accurate results with a reasonable number of simulation runs by using the MCh-X method.

4.2 Performance Comparison: Effect of Number of Factors

To observe the effect of the number factors on the performance of the methods studied here, we consider 4 models with $n \in \{50, 100, 150, 200\}$ factors. In each model, 10% of the factors are randomly selected to be important. The second-order interaction effects between any factor and an unimportant factor are assumed to be unimportant. Among the second-order interaction effects, $0.1n$ effects are randomly selected to be important. The coefficients of factors are determined with the same settings as in §4.1.

The error term is assumed to be normally distributed with mean zero and variance σ^2. For each set of models, we test the factor screening methods on two different variance scenarios: $\sigma = 0.25$ is assumed for the low variance scenario, and $\sigma = 0.75$ is assumed for the high variance scenario.

For each scenario, we use 2^k FF-IV, CSB-X, and MCh-X. The parameter settings for each method are as follows. For 2^k FF-IV, we set $N = 40\sigma$; for MCh-X, we set $\alpha = 0.05$, $\delta = 0.5$, $\tau = 0.6$, and $N_0 = 40\sigma$; and for CSB-X, we set $\alpha = 0.1$, $\Delta_0 = 0.5$, $\Delta_1 = 0.6$, and $N_0 = 5$.

Figures 3 and 4 compare the methods in terms of the various performance measures for the scenarios with different number of factors. All confidence intervals are calculated at 5% significance level. In each figure, a sufficient number of observations have been obtained to create non-overlapping confidence intervals.

Figures 3(a) and 4(a) show that for both low and high variance scenarios, MCh-X required the least number of runs to classify the factors, which increases by the number of factors at a linear rate. Moreover, according to Figures 3(b) and 4(b), when variance is homogenous, the factor screening methods studied here sustain their accuracy as the number of factors increases. Finally, Figures 3(c) and 4(c) display that for both low and high variance scenarios, MCh-X has the lowest cost-effectiveness ratio for different numbers of involved factors.

5   CONCLUSIONS AND FUTURE RESEARCH

The main methodological contribution of this work is the improvement made in Cheng's method. The original Cheng's method may fail to identify the important factors when the response variance is relatively high or the second-order interaction effects are significant. The modification of Cheng's method overcomes some of the original method's limitations, as discussed in detail in §3.

Early sequential factor screening methods require restrictive assumptions about the variance. Recent work, to some extent, has relaxed those assumptions. The CSB-X is designed for simulation models with heterogeneous variance. It, however, loses its efficiency as the response (homogeneous or heterogeneous) variance increases. In this paper, we introduced the MCh-X as an alternative factor screening method that sustains is efficiency as the homogeneous variance increases. Among the current factor screening alternatives, only 2^k FF-IV does not lose its accuracy and efficiency when applied to models with high heterogeneous variance, given that the appropriate transformation is applied. Sequential factor screening methods are still in need of further revision to provide efficiency under high heterogeneous variance.

One of the main limitations of the MCh-X method is the variance homogeneity assumption. For simulation models with heterogeneous response variance, 2^k Fractional Factorial of Resolution IV and CSB-X are still the better factor screening candidates. All factor screening methods discussed in this paper assume a second-order polynomial function for the model response (except the Cheng's method which assumes a first-order polynomial function). This assumption might not be satisfied as the complexity of a simulation model increases and therefore it might lead to misclassification of the factors. A factor screening method that can relax such a restrictive assumption is highly desired.

REFERENCES

Bettonvil, B. W. M., and J. P. C. Kleijnen. 1996. Searching for important factors in simulation models with many factors: Sequential bifurcation. *European Journal of Operational Research* 96:180–194.

Bingham, D., R. R. Sitter, and B. Tang. 2009. Orthogonal and nearly orthogonal designs for computer experiments. *Biometrika* 96 (1): 51–65.

Cheng, R. C. H. 1997. Searching for important factors: Sequential bifurcation under uncertainty. In *Proceedings of the 1997 Winter Simulation Conference*, 275–280. Atlanta, GA.

Cioppa, T. M., and T. W. Lucas. 2007. Efficient nearly orthogonal and space-filling latin hypercubes. *Technometrics* 49 (1): 45–55.

Eskridge, K. M., S. G. Gilmour, R. Mead, N. A. Butler, and D. A. Travnicek. 2004. Large supersaturated designs. *Journal of Statistical Computation and Simulation* 74 (7): 525–542.

Georgiou, S. D. 2008. Modelling by supersaturated designs. *Computational Statistics and Data Analysis* 53:428–435.

Hinkelmann, K., and O. Kempthorne. 2005. *Design and analysis of experiments, advanced experimental design*, Volume 2. Hoboken, New Jersey: John Wiley & Sons, Inc.

Johnson, W. O., and J. L. Gastwirth. 1999. Dual group screening. *Journal of Statistical Planning and Inference* 83:449–473.

Kleijnen, J. P. C. 1987. Review of random and group-screening designs. *Communications in Statistics: Theory and Methods* 16 (10): 2885–2837.

Kleijnen, J. P. C., S. M. Sanchez, T. W. Lucas, and T. M. Cioppa. 2005. State-of-the-art review: A users guide to the brave new world of designing simulation experiments. *INFORMS Journal on Computing* 17 (3): 263289.

Lin, D. K. J. 1995. Generating systematic supersaturated designs. *Technometrics* 37 (2): 213–225.

McKay, M. D., R. Beckman, and W. J. Conover. 1979. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* 21 (2): 239–245.

Montgomery, D. C. 2000. *Design and analysis of experiments*. 5 ed. New York: John Wiley & Sons.

Shen, H., and H. Wan. 2005. Controlled sequential factorial design for simulation factor screening. In *Proceedings of the 2005 Winter Simulation Conference*, 467–474. Orlando, FL.

Shen, H., and H. Wan. 2006. A hybrid method for simulation factor screening. In *Proceedings of the 2006 Winter Simulation Conference*, 382–389. Monterey, CA.

Steinberg, D. M., and D. K. J. Lin. 2006. A construction method for orthogonal latin hypercube designs. *Biometrika* 93 (2): 279–288.

Wan, H., B. E. Ankenman, and B. L. Nelson. 2003. Controlled sequential bifurcation: A new factor-screening method for discrete-event simulation. In *Proceedings of the 2003 Winter Simulation Conference*, 565–573. New Orleans, LA.

Wan, H., B. E. Ankenman, and B. L. Nelson. 2006a. 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. 2006b. Simulation factor screening with controlled sequential bifurcation in the presence of interactions. Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL 60208-3119, U.S.A.

AUTHOR BIOGRAPHIES

REZA YAESOUBI is a post-doctoral research fellow at Harvard Medical School and Brigham and Women's Hospital. He got his Ph.D. in Industrial and Systems Engineering from North Carolina State University in 2009. His research interests include health economics, simulation, optimization, game theory, and mechanism design. His e-mail is <reza.yaesoubi@gmail.com>.

STEPHEN D. ROBERTS is a professor in the Edward P. Fitts Department of Industrial and Systems Engineering at North Carolina State University. His research interests include simulation modeling methodology and the simulation modeling of medical decisions. His e-mail is <roberts@ncsu.edu>.

ROBERT KLEIN has nearly 30 years experience focusing on the application of simulation and other decision analysis tools to pharmacoeconomic and healthcare problems. Since joining Medical Decision Modeling Inc. in 1999, he has designed and/or analyzed over 50 models emphasizing the benefits and costs of pharmaceuticals or medical treatment strategies. Previously, he was the Decision Analyst for the Bowen Research Center at Indiana University and the Regenstrief Institute for Health Care. His e-mail is <rwk@mdm-inc.com>.
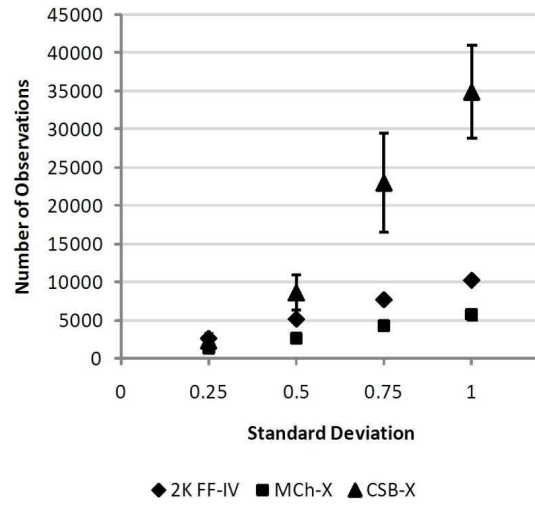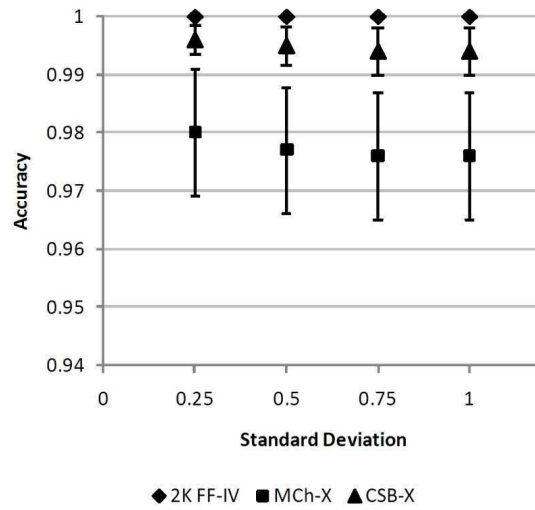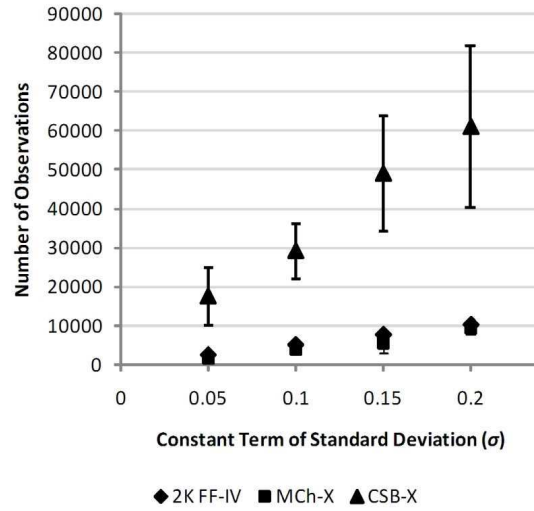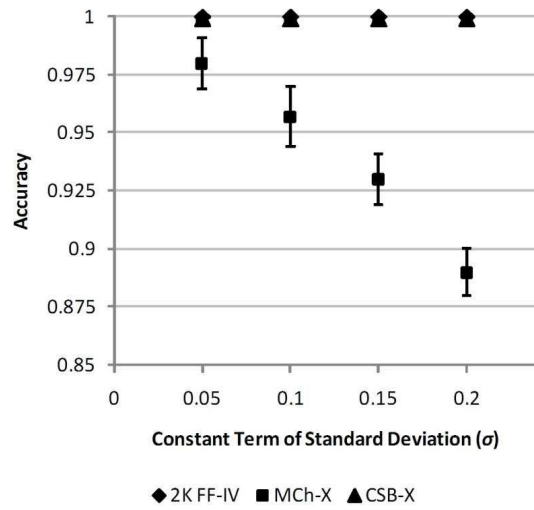
(a) Efficiency
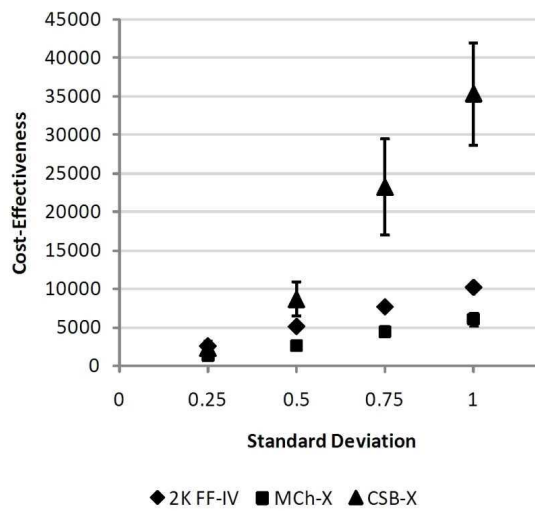
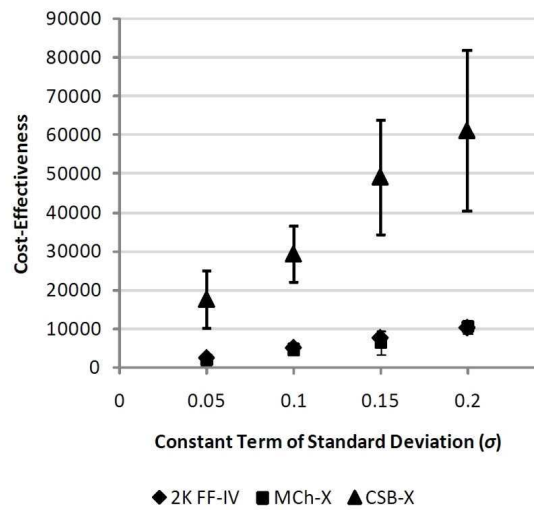(b) Accuracy

(c) Cost-Effectiveness

Figure 1: Performance evaluation for the homogeneous variance scenario
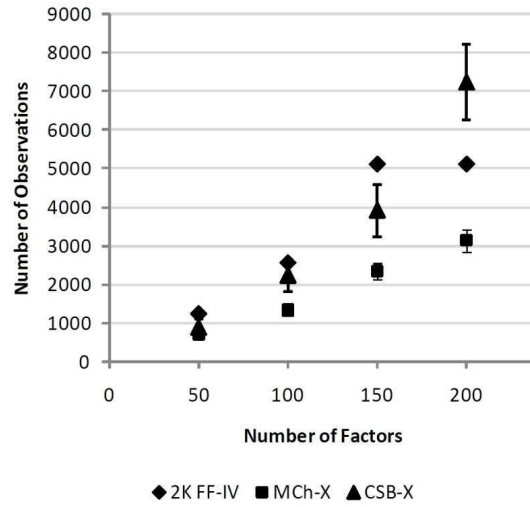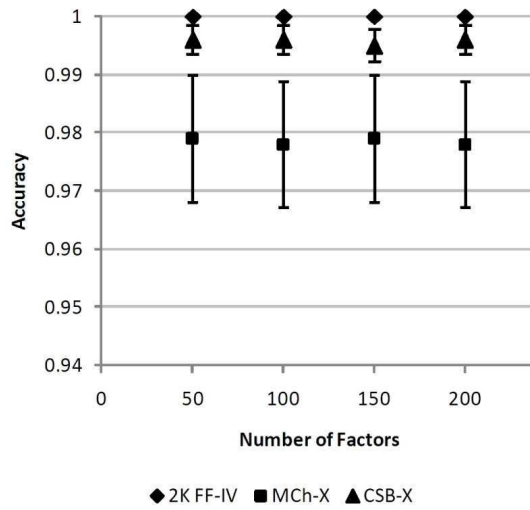
(a) Efficiency
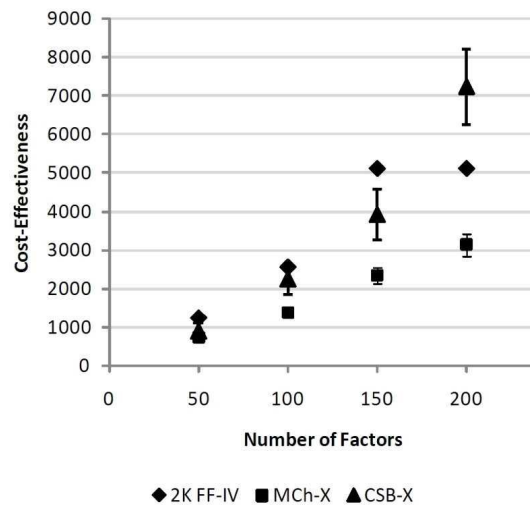

(b) Accuracy

(c) Cost-Effectiveness

Figure 2: Performance evaluation for the heterogeneous variance scenario
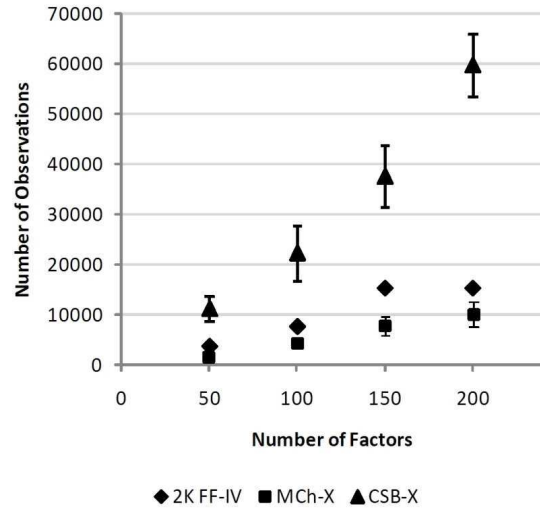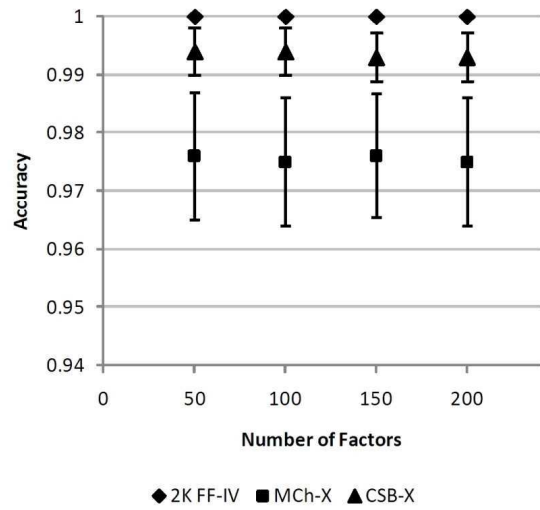
(a) Efficiency
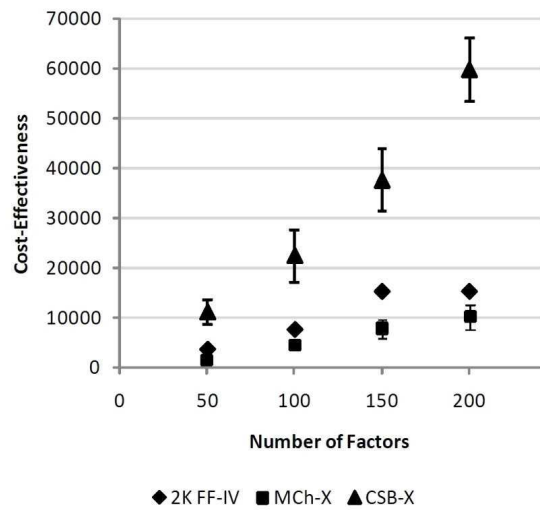

(b) Accuracy

(c) Cost-Effectiveness

Figure 3: Performance evaluation in terms of the number of factors involved for $\sigma = 0.25$

(a) Efficiency

(b) Accuracy

(c) Cost-Effectiveness

Figure 4: Performance evaluation in terms of the number of factors involved for $\sigma = 0.75$