Selection of the Best in the Presence of Subjective Stochastic Constraints

YUWEI ZHOU, Booth School of Business, University of Chicago, Chicago, United States
SIGRÚN ANDRADÓTTIR, H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, United States
SEONG-HEE KIM, H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, United States

We consider the problem of finding a system with the best primary performance measure among a finite number of simulated systems in the presence of subjective stochastic constraints on secondary performance measures. When no feasible system exists, the decision maker may be willing to relax some constraint thresholds. We take multiple threshold values for each constraint as a user's input and propose indifference-zone procedures that perform the phases of feasibility check and selection-of-the-best sequentially or simultaneously. Given that there is no change in the underlying simulated systems, our procedures recycle simulation observations to conduct feasibility checks across all potential thresholds. We prove that the proposed procedures yield the best system in the most desirable feasible region possible with at least a pre-specified probability. Our experimental results show that our procedures perform well with respect to the number of observations required to make a decision, as compared with straight-forward procedures that repeatedly solve the problem for each set of constraint thresholds, and that our simultaneously-running procedure provides the best overall performance.

CCS Concepts: • **Computing methodologies** → **Modeling methodologies**; **Simulation evaluation**; • **Theory of computation** → **Discrete optimization;**

Additional Key Words and Phrases: Ranking and selection, indifference-zone approach, fully sequential procedure, recycling observations, stochastic constraints, subjective constraints

ACM Reference Format:
Yuwei Zhou, Sigrún Andradóttir, and Seong-Hee Kim. 2024. Selection of the Best in the Presence of Subjective Stochastic Constraints. *ACM Trans. Model. Comput. Simul.* 34, 4, Article 22 (July 2024), 60 pages. https://doi.org/10.1145/3664814

1 INTRODUCTION

We consider the problem of selecting the best or near-best system with respect to a primary performance measure among a finite number of simulated systems while also satisfying stochastic

Authors' Contact Information: Yuwei Zhou, Booth School of Business, University of Chicago, Chicago, Illinois, United States; e-mail: yuwzhou@iu.edu; Sigrún Andradóttir, H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia, United States; e-mail: sa@gatech.edu; Seong-Hee Kim, H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia, United States; e-mail: skim@isye.gatech.edu.

ACM 1558-1195/2024/07-ART22
https://doi.org/10.1145/3664814

constraints on one or more secondary performance measures. When no feasible system exists with respect to a given set of threshold values, the decision maker may be willing to relax the threshold values of some constraints so that a feasible system can be found. In that sense, constraints with multiple thresholds can be considered as subjective constraints. The decision maker is often uncertain about the values of performance measures of simulated systems. Thus, the decision maker may prefer tight threshold values, but may worry that the desired thresholds will lead to infeasibility and settle for weaker thresholds. The decision maker could achieve this by starting with the desired thresholds and relaxing them until at least one feasible system is found. Alternatively, she could start with the most relaxed thresholds and tighten them until no feasible system exists. This iterative approach can be tedious and time-consuming. Our approach allows the decision maker to consider several sets of thresholds at the same time, with statistical validity, and hence removes the need for both trade-offs between feasible and desirable thresholds, and for iteratively considering different thresholds. We illustrate this problem with an example.

Suppose a decision maker wants to design an inventory policy such that the expected fill rate within each review period is maximized. She considers using an (s, S) inventory policy (namely ordering products to increase the inventory level up to S when the inventory level at a review period is below s and placing no order, otherwise). Two constraints exist, namely the probability that a shortage occurs between two successive review periods should be less than or equal to $q_1 = 1\%$ and the expected cost per review period should be less than or equal to q_2, where the value of q_2 is small. The decision maker thinks $q_2 = \$100$ is small but is willing to relax the threshold to \$105 or \$110 if no feasible system can be found with $q_2 = \$100$. If there is still no feasible systems with respect to $q_2 = \$110$, then the decision maker is willing to raise the threshold q_1 to 5%, still with three possible values for q_2. To solve this problem, a straightforward approach is to first rank the combinations of the thresholds of both constraints based on how the decision maker prioritizes the thresholds, for example, the threshold combinations (q_1, q_2) are preferred following the order of (1%, \$100), (1%, \$105), (1%, \$110), (5%, \$100), (5%, \$105), and (5%, \$110). Then, she can apply existing constrained optimization procedures repeatedly to identify the optimal system that is feasible with respect to each combination of the thresholds until an optimal feasible system is found. However, this iterative approach can be computational inefficient. Alternatively, since the decision maker has some flexibility with respect to the threshold on each constraint, the two constraints can be viewed as subjective constraints. Identifying the optimal system in the presence of two subjective constraints, where we consider all threshold combinations simultaneously, provides a new approach of solving the problem.

Ranking and selection (R&S) aims to identify a system with the best performance among finitely many systems whose performances are estimated by stochastic simulation. References [13] and [10] provide literature reviews on R&S. When the problem requires not only selecting the best system with respect to a primary performance measure but also determining the feasibility with respect to stochastic constraints on secondary performance measures, it becomes constrained R&S. There are three major approaches to solving constrained R&S, namely the **indifference-zone (IZ)** approach, the **optimal computing budget allocation (OCBA)** approach, and the Bayesian approach. References [11, 17], and [18] propose sampling frameworks that approximate the OCBA considering stochastic constraints. Reference [21] proposes a sequential policy from the Bayesian approach for allocating simulation effort to determine a set of systems with mean performance exceeding a threshold. For the IZ approach, the decision maker usually needs to specify an IZ parameter, which corresponds to the smallest significant difference of a performance measure that she values (see further discussion in Section 2.2). Reference [3] proposes a fully sequential procedure that finds a set of feasible systems given multiple constraints. Reference [1] proposes procedures that select the best with respect to the primary performance measure among a finite number of simulated systems in the presence of a single stochastic constraint on a secondary performance measure. Reference [9] applies the concept of dormancy to efficiently solve constrained R&S and [8] proposes procedures to select the best in the presence of multiple constraints.

For constrained R&S, if each constraint has one fixed threshold value, procedures due to [1] or [8] can be used. When the decision maker is willing to consider multiple threshold values, one may consider iteratively applying those procedures "from scratch" to each set of thresholds. However, this wastes all the information from the previous constrained R&S problems and becomes computationally inefficient. Given that there is no change in the simulation model of each system, a natural idea is to recycle simulation observations for constrained R&S with different thresholds. The idea of recycling simulation observations for computer experiments is proposed in [6]. However, they focus on estimation rather than comparison. Reference [22] proposes a procedure that performs feasibility determination when the decision maker wants to consider multiple threshold values on each constraint. They use the idea of recycling simulation observations and perform feasibility determination simultaneously with respect to all thresholds so that the overall required number of observations is reduced. However, their focus is on feasibility determination rather than on finding the best feasible system in the presence of subjective constraints.

In this article, we adopt the concept of recycling simulation observations in the context of constrained R&S when constraint thresholds vary. We provide fully sequential procedures that return the best feasible system with respect to the most preferred threshold values possible, where the preference order among thresholds is specified by the user. The threshold values for constraints are relaxed until there is at least one feasible solution. We prove that our procedures achieve a desired overall **probability of correct selection (PCS)** and also perform well in reducing the required number of observations until a decision is made compared with straight-forward repeating procedures, namely applying the procedures of [1] or [8] iteratively to each possible set of threshold values depending on whether the problem has a single constraint or multiple constraints.

It is worth mentioning that, besides the formulation of constrained R&S, there are two other approaches for dealing with multiple performance measures. A frequently used approach is to aggregate multiple objectives into a single objective by applying weights or a utility function, as discussed in [4]. However, determining the appropriate weights or utility function can be tricky, particularly when the units of the objectives differ (e.g., costs and probabilities). Furthermore, the optimal solution may vary as the weights or utility function changes. Another approach is to identify a Pareto set, which comprises non-dominated solutions for multiobjective optimization problems. A number of ranking and selection procedures have been developed to find Pareto sets for stochastic multi-objective problems, including [5, 7, 16], and [2]. While the approach of finding a Pareto set is in general applicable to the problem we discuss, our formulation and methods that utilize subjective constraints provide an alternative approach. Our proposed formulation provides two potential advantages regarding the problem discussed. First, the Pareto set may include several alternatives that excel in one performance measure while severely compromising other performance measures. Given that such extreme systems are unlikely to appeal to the decision maker, the computational effort spent to identify those systems may be avoided. Second, the Pareto set could consist of a large number of systems, leaving the decision maker with the challenge of identifying all non-dominated systems before eventually selecting one among the many systems present on the Pareto frontier for implementation. Our formulation overcomes this issue with the Pareto set formulation, as discussed in further detail in Sections 2.1 and 5 and through a case study in Section 7. However, there are other circumstances where identifying the entire Pareto set is desirable, such as when the decision maker does not aim to optimize a primary performance measure among all performance measures or when the decision maker wishes to understand the performance of all non-dominated systems (e.g., to study trade-offs post hoc). Due to the fact that our proposed procedures solve a different problem than MORS procedures (e.g., we optimize a primary performance measure subject to subjective constraints on secondary performance measures, rather than identifying the entire Pareto set), we do not directly compare the performance of our proposed procedures with that of the Pareto set approach.

The rest of the article is organized as follows: Section 2 provides the background for our problem. Sections 3 and 4 propose and analyze sequentially-running and simultaneously-running procedures, respectively, for the feasibility check and comparison phases. Section 5 discusses three major preference orders of the constraint thresholds. In Section 6, we present numerical results for the proposed procedures and compare their performances with the straight-forward procedures that apply existing constrained R&S procedures repeatedly to each set of thresholds. Section 7 further demonstrates the implementation and the performance of our proposed procedures through a case study based on an inventory policy example. Concluding remarks are provided in Section 8. Finally, the Appendices include the statement of one proposed procedure and two competing procedures, the detailed proof of the statistical validity of the proposed and competing procedures, the description of setting the required implementation parameters, and some additional experimental results.

2 BACKGROUND

In this section, we formulate our problem in Section 2.1 and discuss how we define the correct selection event in Section 2.2. The assumptions for the statistical validity of our proposed procedures are presented in Section 2.3.

2.1 Problem Formulation

We consider k systems whose primary performance measures, as well as s secondary performance measures, can be estimated through stochastic simulation. Let Γ denote the index set of all possible systems (i.e., $\Gamma = \{1, \ldots, k\}$). Let X_{in} be the observation associated with the primary performance measure of system i from replication n, and $Y_{i\ell n}$ be the observation associated with the ℓth stochastic constraint of system i from replication n, where $\ell = 1, \ldots, s$. We also define the expected values of the primary and secondary performance measures for each system $i \in \Gamma$ and constraint $\ell = 1, \ldots, s$ as $x_i = \mathrm{E}[X_{in}]$ and $y_{i\ell} = \mathrm{E}[Y_{i\ell n}]$, respectively. Constrained R&S is to select

$$\begin{array}{ll} \arg\max_{i\in\Gamma} & x_i \\ \text{s.t.} & y_{i\ell} \leq q_\ell \quad \text{for all } \ell = 1, \ldots, s, \end{array}$$

where q_ℓ denotes the constraint threshold for constraint ℓ.

For a given threshold vector $\mathbf{q} = (q_1, \ldots, q_s)$, procedures due to [1] can be used to find the best system if there is only one constraint. If there are multiple constraints, procedures due to [8] are suitable. In this article, we assume that the decision maker has a list of possible threshold values in consideration for each constraint and hopes to select the best system with respect to the most preferable thresholds possible. We further assume that $k \geq 2$ in this article. We let d_ℓ denote the number of distinct threshold values and $q_{\ell,m}$ denote the mth distinct threshold value on constraint ℓ, where $m = 1, \ldots, d_\ell$ and $\ell = 1, \ldots, s$. We assume $q_{\ell,1} < \cdots < q_{\ell,d_\ell}$, where $\ell = 1, \ldots, s$.

The threshold values for individual constraints are combined into an ordered list of vectors of threshold values $\{\mathbf{q}^{(1)}, \mathbf{q}^{(2)}, \ldots, \mathbf{q}^{(d)}\}$, where d denotes the total number of threshold vectors that the decision maker is interested to test. We assume that $\mathbf{q}^{(1)}$ is preferred to $\mathbf{q}^{(2)}$, $\mathbf{q}^{(2)}$ is preferred to $\mathbf{q}^{(3)}$, and so on. For the implementation of our procedures, a decision maker can input (i) the ordered list of threshold vectors, or (ii) an ordered list of threshold values for each constraint and a mechanism for constructing an ordered list of threshold vectors from the inputted threshold values (see Section 5). Note that the ordered list of threshold vectors should remain fixed throughout the implementation. We let $q_\ell^{(\theta)}$ be the threshold value on constraint ℓ in $\mathbf{q}^{(\theta)}$, where $\theta = 1, \ldots, d$ and $\ell = 1, \ldots, s$. Then we introduce the threshold index vector $\boldsymbol{I}^{(\theta)}$ to include the indices of the threshold values that form $\mathbf{q}^{(\theta)}$. Similar to the definition of $q_\ell^{(\theta)}$, $I_\ell^{(\theta)}$ represents the threshold index on constraint ℓ in $\mathbf{q}^{(\theta)}$.

Consider the example of selecting the best inventory control policy discussed in Section 1. Then $s = 2, d_1 = 2$ (i.e., two threshold values for the first constraint), $d_2 = 3$ (i.e., three threshold values for the second constraint), $q_{1,1} = 1, q_{1,2} = 5$, and $q_{2,1} = 100, q_{2,2} = 105$, and $q_{2,3} = 110$. Moreover, we consider the following $d = 6$ ordered threshold vectors

$$\mathbf{q}^{(1)} = \begin{bmatrix} 1 \\ 100 \end{bmatrix}, \quad \mathbf{q}^{(2)} = \begin{bmatrix} 1 \\ 105 \end{bmatrix}, \quad \mathbf{q}^{(3)} = \begin{bmatrix} 1 \\ 110 \end{bmatrix}, \quad \mathbf{q}^{(4)} = \begin{bmatrix} 5 \\ 100 \end{bmatrix}, \quad \mathbf{q}^{(5)} = \begin{bmatrix} 5 \\ 105 \end{bmatrix}, \text{ and } \mathbf{q}^{(6)} = \begin{bmatrix} 5 \\ 110 \end{bmatrix}.$$

Note that $q_1^{(1)} = q_1^{(2)} = q_1^{(3)} = 1, q_1^{(4)} = q_1^{(5)} = q_1^{(6)} = 5$, while $q_2^{(1)} = q_2^{(4)} = 100, q_2^{(2)} = q_2^{(5)} = 105$, and $q_2^{(3)} = q_2^{(6)} = 110$. The threshold index vectors are

$$\boldsymbol{I}^{(1)} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \boldsymbol{I}^{(2)} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad \boldsymbol{I}^{(3)} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}, \quad \boldsymbol{I}^{(4)} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \quad \boldsymbol{I}^{(5)} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}, \text{ and } \boldsymbol{I}^{(6)} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}.$$

Hence $I_1^{(1)} = I_1^{(2)} = I_1^{(3)} = 1, I_1^{(4)} = I_1^{(5)} = I_1^{(6)} = 2$, while $I_2^{(1)} = I_2^{(4)} = 1, I_2^{(2)} = I_2^{(5)} = 2$, and $I_2^{(3)} = I_2^{(6)} = 3$.

For $\theta \leq d$, we use A_θ to denote the region that is feasible under threshold vector $\mathbf{q}^{(\theta)}$ but not under threshold vectors $\mathbf{q}^{(1)}, \ldots, \mathbf{q}^{(\theta-1)}$ (if $\theta > 1$), and use A_{d+1} to denote the region that is infeasible to all $\mathbf{q}^{(1)}, \ldots, \mathbf{q}^{(d)}$. More specifically, we let

$$A_\theta = \begin{cases} \left\{(z_1, z_2, \ldots, z_s) : z_\ell \leq q_\ell^{(\theta)}, \ell = 1, 2, \ldots, s\right\}, & \text{if } \theta = 1; \\ \left\{(z_1, z_2, \ldots, z_s) : z_\ell \leq q_\ell^{(\theta)}, \ell = 1, 2, \ldots, s\right\} \setminus \cup_{\kappa=1}^{\theta-1} A_\kappa, & \text{if } \theta = 2, \ldots, d; \\ \mathbb{R}^s \setminus \cup_{\kappa=1}^{d} A_\kappa, & \text{if } \theta = d+1. \end{cases} \tag{1}$$

With this definition of A_θ, we can say that the decision maker wants to find the best among systems whose constraint mean configurations fall in A_1 but would consider systems in A_2 if no systems fall in A_1. She would further consider systems in A_3 if no systems fall in A_1 and A_2 and $d \geq 3$, and so on.

We assume that the ordered list of threshold vectors is such that when there is no trade-off, the decision maker always prefers "tighter" combinations of threshold values. Consider a case where there are two (non-negative) constraints, the first constraint has three thresholds, and the second constraint has two thresholds. Then it is not possible for the decision maker to prefer $(q_{1,3}, q_{2,1})$ to $(q_{1,2}, q_{2,1})$ in the preference order. Figure 1 shows $A_1, \ldots, A_5$ for an example with $d = 4$ combinations of threshold vectors. We see that $\mathbf{q}^{(1)} = (q_{1,2}, q_{2,1})$ does not correspond to the "tightest" combination of threshold values (i.e., $(q_{1,1}, q_{2,1})$), and similarly $\mathbf{q}^{(d)} = (q_{1,3}, q_{2,1})$ does not correspond to the "weakest" combination of threshold values (i.e., $(q_{1,3}, q_{2,2})$).

The following definition will facilitate the efficient implementation of our approaches.

Definition 2.1. Constraint ℓ has an increasing preference if $q_\ell^{(\theta)} \leq q_\ell^{(\theta')}$ for any $\theta, \theta' = 1, 2, \ldots, d$ with $\theta < \theta'$.

We consider the following two examples to further explain Definition 2.1. Figure 2 shows three preference orders of threshold vectors for two (non-negative) constraints with $d_1 = d_2 = 3$. Based on our definition of threshold vectors, Figure 2(a) formulates the threshold vectors as $\mathbf{q}^{(1)} = (q_{1,1}, q_{2,1}), \mathbf{q}^{(2)} = (q_{1,1}, q_{2,2}), \mathbf{q}^{(3)} = (q_{1,1}, q_{2,3}), \mathbf{q}^{(4)} = (q_{1,2}, q_{2,1})$, and so on. We see that constraint 1 has increasing preference whereas constraint 2 does not. On the other hand, we have

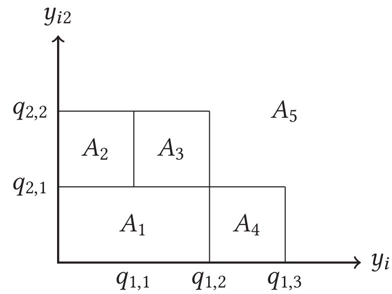


Fig. 1. A preference order where the "tightest" ("weakest") combination of thresholds is not "most" ("least") preferred threshold vector.

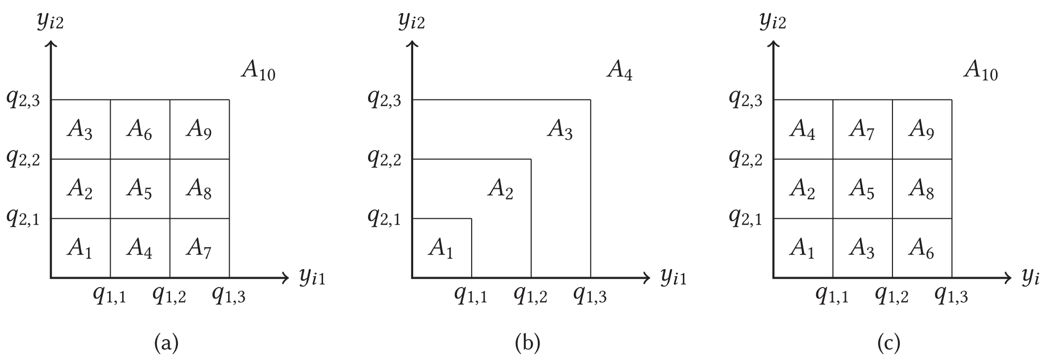


Fig. 2. Three preference orders.

$d = 3, \mathbf{q}^{(1)} = (q_{1,1}, q_{2,1}), \mathbf{q}^{(2)} = (q_{1,2}, q_{2,2})$, and $\mathbf{q}^{(3)} = (q_{1,3}, q_{2,3})$ in Figure 2(b), which satisfies Definition 2.1 for both constraints. Finally, in Figures 2(c) and 1, neither constraint has increasing preference.

2.2 Correct Selection

To solve the constrained R&S problem with subjective constraints described in Section 2.1, we consider two phases, namely Phase I to identify feasible systems and Phase II to select a system with the largest x_i based on a comparison among feasible systems. These phases are designed to correctly select the best feasible system with respect to the most preferred threshold vector possible, as described in this section.

For stochastic constraints, it is not always possible to guarantee a correct feasibility determination with respect to the stochastic constraints. Instead, [1] introduces a tolerance level, namely $\epsilon_\ell > 0$, for constraint ℓ, which is a positive real value predefined by the decision maker. This is often interpreted as the amount the decision maker is willing to be off from a given threshold value. Consider a threshold value $q_{\ell,m}$ for $m = 1, 2, \ldots, d_\ell$. Any systems with $y_{i\ell} \leq q_{\ell,m} - \epsilon_\ell$ are considered as desirable systems with respect to constraint ℓ and threshold value $q_{\ell,m}$. We let $D_\ell(q_{\ell,m})$ denote the set of desirable systems for constraint ℓ and $q_{\ell,m}$. Systems with $y_{i\ell} \geq q_{\ell,m} + \epsilon_\ell$ are considered as unacceptable systems for constraint ℓ and threshold $q_{\ell,m}$, and are placed in set $U_\ell(q_{\ell,m})$. Systems that fall within a tolerance level of $q_{\ell,m}$, which means $q_{\ell,m} - \epsilon_\ell < y_{i\ell} < q_{\ell,m} + \epsilon_\ell$, are considered as acceptable systems, placing them in the set $A_\ell(q_{\ell,m})$. More specifically,

$$D_\ell(q_{\ell,m}) = \{i \in \Gamma \mid y_{i\ell} \leq q_{\ell,m} - \epsilon_\ell\};$$

$$U_\ell(q_{\ell,m}) = \{i \in \Gamma \mid y_{i\ell} \geq q_{\ell,m} + \epsilon_\ell\}; \text{ and}$$
$$A_\ell(q_{\ell,m}) = \{i \in \Gamma \mid q_{\ell,m} - \epsilon_\ell < y_{i\ell} < q_{\ell,m} + \epsilon_\ell\}.$$

Remark 1. As discussed in [1], a feasible (infeasible) system i with $y_{i\ell} \in (q_{\ell,m} - \epsilon_\ell, q_{\ell,m})$ ($y_{i\ell} \in (q_{\ell,m}, q_{\ell,m} + \epsilon_\ell)$) that falls in the acceptable set with respect to constraint ℓ may be declared infeasible (feasible). This leads to potential errors in feasibility decisions, which are analogous to Type I and II errors of a hypothesis test. Therefore, $q_{\ell,m}$ and ϵ_ℓ should be chosen based on which error the decision maker views more important. For example, for the cost constraint of the inventory example in Section 1, if the decision maker wants to select systems whose expected cost is below 105 but eliminate all systems whose expected cost is above 110, she can set $q_{\ell,m} - \epsilon_\ell = 105$ and $q_{\ell,m} + \epsilon_\ell = 110$, which is equivalent to setting $q_{\ell,m} = 107.5$ and $\epsilon_\ell = 2.5$.

When feasibility check is performed to completion (until a decision is made), we let $\mathrm{CD}_{i\ell}(q_{\ell,m})$ denote the correct decision event of system i with respect to constraint ℓ and threshold $q_{\ell,m}$, which is defined as declaring system i as feasible if $i \in D_\ell(q_{\ell,m})$ and as infeasible if $i \in U_\ell(q_{\ell,m})$. Any feasibility decision is considered correct if $i \in A_\ell(q_{\ell,m})$. For any threshold vector $\mathbf{q}^{(\theta)}$, we say that system i is desirable with respect to $\mathbf{q}^{(\theta)}$ when it is desirable with respect to all the constraints, that is, $i \in D_\ell(q_\ell^{(\theta)})$ for all $\ell = 1, \ldots, s$. System i is unacceptable with respect $\mathbf{q}^{(\theta)}$ if it is unacceptable with respect to at least one constraint, that is, there exists ℓ such that $i \in U_\ell(q_\ell^{(\theta)})$. When system i is acceptable to some (or all) the constraints and desirable with respect to the other constraints, system i is called acceptable with respect to $\mathbf{q}^{(\theta)}$.

To select the best system with respect to the primary performance measure in Phase II, the decision maker needs to choose an indifference-zone parameter δ, which is the smallest absolute difference that the decision maker considers significant in terms of the primary performance measure. More specifically, any system whose primary performance measure is at least δ smaller (larger) than system i is considered as inferior (superior) to system i.

Let θ^* be the smallest θ such that $D_\ell(q_\ell^{(\theta)}) \neq \emptyset$ for all ℓ. If for each $\theta = 1, \ldots, d$, there exists at least one constraint ℓ_θ such that $D_{\ell_\theta}(q_{\ell_\theta}^{(\theta)}) = \emptyset$, that is, θ^* does not exist, then we set $\theta^* = d+1$. If $\theta^* \leq d$, then $\mathbf{q}^{(\theta^*)}$ is the most preferable threshold vector possible where at least one desirable system exists. Further, let B denote the set of desirable systems with respect to $\mathbf{q}^{(\theta^*)}$ (i.e., $B = \cap_{\ell=1}^s D_\ell(q_\ell^{(\theta^*)})$ and let $[b]$ be the index of the best system among the systems in B, so that $x_{[b]} \geq x_i$ for $i, [b] \in B$. Then if $\theta^* \leq d$, the correct selection event is to select a desirable or acceptable system with respect to $\mathbf{q}^{(\theta^*)}$ whose primary performance is not inferior to the best system, or an acceptable system with respect to a preferred threshold vector. More specifically,

$$\mathrm{CS} = \left\{ \text{select } i \text{ such that either } i \in \cap_{\ell=1}^s \left(D_\ell\left(q_\ell^{(\theta^*)}\right) \cup A_\ell\left(q_\ell^{(\theta^*)}\right)\right) \text{ and } x_i > x_{[b]} - \delta \right.$$
$$\left. \text{or } i \in \cup_{\theta<\theta^*} \cap_{\ell=1}^s \left(D_\ell\left(q_\ell^{(\theta)}\right) \cup A_\ell\left(q_\ell^{(\theta)}\right)\right) \right\}.$$

If $\theta^* = d+1$, CS is to either declare that no feasible systems exist or identify any acceptable system with respect to any of the threshold vectors $\mathbf{q}^{(1)}, \ldots, \mathbf{q}^{(d)}$.

Remark 2. If ϵ_ℓ is small enough that no acceptable systems exist, then a CS event corresponds to the selection of either system $[b]$ or an acceptable system i with respect to $\mathbf{q}^{(\theta^*)}$ where $x_i > x_{[b]} - \delta$. However, if there are acceptable systems with respect to $\mathbf{q}^{(\theta)}$ for $\theta < \theta^*$, then they may be declared feasible to $\mathbf{q}^{(\theta)}$. In this case, systems infeasible to $\mathbf{q}^{(\theta)}$ are eliminated including system $[b]$, and a CS event happens when selecting an acceptable system i (probably with the best primary performance

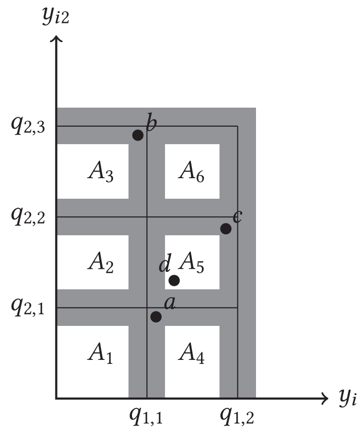


Fig. 3. Regions for two secondary performance measures and six threshold vectors.

measure but no guarantee whether $x_i > x_{[b]} - \delta$) from among those declared feasible with respect to $\mathbf{q}^{(\theta)}$.

To better illustrate the CS event, we consider a problem with two constraints where the first constraint has two thresholds and the second constraint has three thresholds. We consider all $d = 6$ possible threshold vectors $\mathbf{q}^{(1)}, \ldots, \mathbf{q}^{(6)}$. Figure 3 presents possible (non-negative) secondary performance means and thresholds where the shaded areas represent acceptable regions with respect to one or more threshold vectors, and $A_1, \ldots, A_6$ are defined as in Equation (1) and are separated by the solid lines. Assuming that there are four systems a, b, c, and d, we see that (i) $\theta^* = 5$; (ii) $a, b \in \cup_{\theta<\theta^*} \cap_{\ell=1}^{s} (D_\ell(q_\ell^{(\theta)}) \cup A_\ell(q_\ell^{(\theta)}))$; and (iii) $a, c, d \in \cap_{\ell=1}^{s}(D_\ell(q_\ell^{(5)}) \cup A_\ell(q_\ell^{(5)}))$. Then a CS event is to select system $i \in \{a, c, d\}$ such that $x_i > x_{[b]} - \delta$. Another possible CS event is to select a when a is declared feasible to $\mathbf{q}^{(1)}$ because systems $\{b, c, d\}$ are infeasible to $\mathbf{q}^{(1)}$. Similarly, if a is declared infeasible to $\mathbf{q}^{(1)}$ and $\mathbf{q}^{(3)}$ but b is declared feasible to $\mathbf{q}^{(3)}$, then the selection of b is a CS event. Finally, if a is declared infeasible to $\mathbf{q}^{(1)}$ but both a and b are declared feasible to $\mathbf{q}^{(3)}$, then $\{c, d\}$ are eliminated and the selection of a or b (with a better primary performance measure) becomes a CS event.

2.3 Notation and Assumptions

Throughout the article, we let $\mathbb{1}(\cdot)$ be the indicator function and $|S|$ be the cardinality of set S, and use the additional notation defined below:

$$
\begin{aligned}
n_0 &\equiv \text{initial sample size for each system } (n_0 \geq 2);\\
r_i &\equiv \text{number of observations so far for system } i \ (r_i \geq n_0);\\
\bar{X}_i(r_i) &\equiv \text{average value of } X_{i1}, \ldots, X_{ir_i} \text{ for system } i;\\
\bar{Y}_{i\ell}(r_i) &\equiv \text{average value of } Y_{i\ell 1}, \ldots, Y_{i\ell r_i} \text{ for system } i \text{ and constraint } \ell;\\
S^2_{X_{ij}}(n_0) &\equiv \text{sample variance of } X_{i1} - X_{j1}, \ldots, X_{in_0} - X_{jn_0} \text{ between system } i \text{ and } j;\\
S^2_{Y_{i\ell}}(n_0) &\equiv \text{sample variance of } Y_{i\ell 1}, \ldots, Y_{i\ell n_0} \text{ for system } i \text{ and constraint } \ell;\\
R(r_i; v, w, z) &\equiv \max\left\{0, \frac{(n_0-1)wz}{v} - \frac{v}{2c} r_i\right\} \text{ for } v, w, z \in \mathbb{R}^+ \text{ and } c \in \{1, 2, \ldots\};\\
g(\eta) &\equiv \sum_{j=1}^{c} (-1)^{j+1} \left(1 - \frac{1}{2}\mathbb{1}(j = c)\right) \times \left(1 + \frac{2\eta(2c-j)j}{c}\right)^{-(n_0-1)/2};
\end{aligned}
$$

$\alpha \equiv$ overall nominal error for a procedure under consideration, where $0 < \alpha < 1$.

Note that an integer parameter c is required for both $R(r_i; v, w, z)$ and $g(\eta)$. This is a user-defined parameter that impacts the shape of the continuation region defined by $(-R(r_i; v, w, z), R(r_i; v, w, z))$ (it becomes a longer triangle as c increases). The choice $c = 1$ is recommended as it guarantees a unique and easy solution when computing the implementation parameter η from $g(\eta)$. Reference [12] shows the derivation of $R(r_i; v, w, z)$ and also suggests that $c = 1$ is a good choice when the decision maker does not have information about the systems' mean configuration. The experimental results in the article are based on $c = 1$.

Our statistical analysis of our proposed procedures will rely on the following two assumptions.

Assumption 1. *For each system i, where $i = 1, \ldots, k$, we have*

$$\begin{bmatrix} X_{in} \\ Y_{i1n} \\ \vdots \\ Y_{isn} \end{bmatrix} \overset{iid}{\sim} N_{s+1}\left(\begin{bmatrix} x_i \\ y_{i1} \\ \vdots \\ y_{is} \end{bmatrix}, \Sigma_i \right), \qquad n = 1, 2, \ldots$$

where $\overset{iid}{\sim}$ denotes independent and identically distributed, N_{s+1} denotes $(s+1)$-dimensional multivariate normal, and Σ_i is the $(s+1) \times (s+1)$ positive definite covariance matrix of the vector $(X_{in}, Y_{i1n}, \ldots, Y_{isn})$. Furthermore, for the primary performance measure, we have

$$\begin{bmatrix} X_{1n} \\ \vdots \\ X_{kn} \end{bmatrix} \overset{iid}{\sim} N_k\left(\begin{bmatrix} x_1 \\ \vdots \\ x_k \end{bmatrix}, \Sigma' \right),$$

where Σ' is the $k \times k$ positive definite covariance matrix of the vector $(X_{1n}, \ldots, X_{kn})$.

Normally distributed data is a common assumption used in many R&S procedures due to the fact that it can be justified by the Central Limit Theorem when observations are either within-replication averages or batch means ([15]). Moreover, primary and secondary performance measures are usually correlated. When **common random numbers** (**CRN**) are introduced in simulating observations from each system, observations between systems are correlated. Our formulation allows correlations between both performance measures and systems. Note that $Y_{i\ell n}$ and $Y_{j\ell n}$ can be correlated for $i \neq j$ if CRNs are used. However, as feasibility determination involves comparisons between $Y_{i\ell n}$ and thresholds rather than $Y_{j\ell n}$, we do not require any assumptions about their covariance structure across systems.

Assumption 2. *If $\theta^* \leq d$, then for any system $i \in \cap_{\ell=1}^{s}(D_\ell(q_\ell^{(\theta^*)}) \cup A_\ell(q_\ell^{(\theta^*)}))$, where $i \neq [b]$, we assume $x_i \leq x_{[b]} - \delta$.*

Assumption 2 implies that there exists only one best system $[b]$ and any systems that are desirable or acceptable with respect to $q_\ell^{(\theta^*)}$ for all constraint $\ell = 1, \ldots, s$ are inferior to system $[b]$. In reality, one can choose a reasonably small δ to satisfy Assumption 2. This assumption is a standard assumption for proving the statistical validity of IZ approaches in the R&S literature.

3 SEQUENTIALLY-RUNNING PROCEDURES

In this section, we present two procedures, namely $\mathcal{ZAK}^{\mathcal{R}}$ and $\mathcal{ZAK}$, that implement Phases I and II sequentially.

References [1] and [8] also propose sequentially-running procedures to select the best system in the presence of multiple constraints. Our sequentially-running procedures use similar steps in

Phase II as [1] and [8], but the steps for Phase I are different because [1] and [8] consider one fixed set of thresholds while we consider multiple thresholds. Our approach for handling multiple threshold values builds on the work of [22] who developed $\mathcal{RF}$, an efficient fully-sequential procedure for checking the feasibility of all systems with respect to all constraints and all thresholds simultaneously. Reference [22] show that once a system i is declared feasible with respect to a threshold $q_{\ell,m}$ such that $q_{\ell,m} \geq y_{i\ell} + \epsilon_\ell$, this system will be declared feasible with respect to all thresholds $q_{\ell,m+1}, \ldots, q_{\ell,d_\ell}$ on constraint ℓ. Similarly, if a system i is declared infeasible with respect to a threshold $q_{\ell,m}$ such that $q_{\ell,m} \leq y_{i\ell} - \epsilon_\ell$, then this system will be declared infeasible with respect to all the thresholds $q_{\ell,1}, \ldots, q_{\ell,m-1}$. This fact is essential in our proposed procedures.

The $\mathcal{ZAK}^{\mathcal{R}}$ ("restart") procedure is statistically valid, while the $\mathcal{ZAK}$ procedure is heuristic. The two procedures are similar in the sense that both start by executing Phase I for all systems to identify the most preferred threshold vector possible, $\mathbf{q}^{(\theta^*)}$, as well as the feasible systems with respect to $\mathbf{q}^{(\theta^*)}$. The parameter θ keeps track of our current estimate of θ^* (initially $\theta = d$), M is a set of systems that are in consideration (initially M contains all the systems, that is, $M = \Gamma$), and F is a set of systems that are declared feasible with respect to threshold vector $\mathbf{q}^{(\theta)}$ (initially $F = \emptyset$). The procedures return $Z_{i,\ell,m} = 1$ ($Z_{i,\ell,m} = 0$) if system i is declared feasible (infeasible) with respect to constraint ℓ and threshold q_ℓ^m, and $Z_{i,\ell,m} = 2$ if no decision is made about the feasibility of system i with respect to threshold $q_{\ell,m}$ on constraint ℓ. Notice that once a system is declared feasible with respect to threshold vector $\mathbf{q}^{(\theta)}$ where $1 \leq \theta \leq d-1$, we do not need to check feasibility for any systems with respect to the less preferred threshold vectors $\mathbf{q}^{(\theta+1)}, \ldots, \mathbf{q}^{(d)}$.

The sequentially-running procedures, $\mathcal{ZAK}^{\mathcal{R}}$ and $\mathcal{ZAK}$, perform Phase II on the surviving systems from the completion of Phase I. More specifically, it selects the best system with respect to the primary performance measure among the subset of systems that are declared feasible with respect to the most preferred threshold vector possible identified in Phase I. The main difference between them lies in whether they collect observations on the primary performance measure during Phase I and recycle them in Phase II. In order to prove the statistical validity of $\mathcal{ZAK}^{\mathcal{R}}$ and avoid storing simulation results, the procedure avoids the correlation between the primary and secondary performance measures by not recycling any observations from Phase I and instead restarting "from scratch" when implementing comparisons in Phase II. Moreover, when CRN are used to compare systems in Phase II, we assume that the implementation of CRN is such that the simulation results for any surviving system in Phase II do not depend on the set of surviving systems F (e.g., the simulation results for any surviving system i would be the same as if $F = \Gamma$). $\mathcal{ZAK}^{\mathcal{R}}$ is described in Algorithm A.1 along with its statistical validity in Appendix A. A discussion about how to set the implementation parameters for $\mathcal{ZAK}^{\mathcal{R}}$ is given in Appendix B.1.

As $\mathcal{ZAK}^{\mathcal{R}}$ starts "from scratch" when performing the comparison, it discards all the information related to the primary performance measure obtained in Phase I, which can be quite inefficient in terms of the computation effort. One may consider collecting and storing all the observations of the primary performance measure in Phase I and then extracting information related to the primary performance measure when performing Phase II. However, as Phase I may require a lot of observations, this approach requires significant memory for storing the observations from Phase I. [19] proposes the Sequential Selection with Memory procedure ($\mathcal{SSM}$) that is specifically for use within an optimization-via-simulation algorithm when simulation is costly, and partial or complete information on solutions previously visited is maintained. When data storage is prohibitive, the procedure requires only summary statistics of the simulation output, which solves the memory space issue discussed above. We then present a sequentially-running procedure, namely $\mathcal{ZAK}$, that adopts the $\mathcal{SSM}$ procedure as its Phase II. The detailed description is shown in Algorithm 1.

ALGORITHM 1: Procedure $\mathcal{ZAK}$.

[**Setup:**] Select the overall nominal confidence level $1-\alpha$ and choose $0<\alpha_f, \alpha_c<1$ such that $\alpha_f+\alpha_c=\alpha$. Choose tolerance levels $\epsilon_1, \ldots, \epsilon_s$, indifference-zone parameter δ, threshold vectors $\{\mathbf{q}^{(1)}, \mathbf{q}^{(2)}, \ldots, \mathbf{q}^{(d)}\}$, and associated index vectors $\{\boldsymbol{I}^{(1)}, \boldsymbol{I}^{(2)}, \ldots, \boldsymbol{I}^{(d)}\}$. Set $M=\Gamma$ and $Z_{i,\ell,m}=2$ for all $i\in M$, $\ell=1,\ldots,s$, and $m=1,\ldots,d_\ell$. Set $F=\emptyset$ and $\theta=d$. Set η_f such that $g(\eta_f)=\alpha'_f$, where $0<\alpha'_f<1/s$ is set as a solution to

$$\left(1-\min\{s,d\}\alpha'_f\right)^{k-1}\times(1-s\alpha'_f)=1-\alpha_f,\ \text{ if systems are simulated independently;}$$

and set as

$$\alpha'_f=\alpha_f/[(k-1)\min\{s,d\}+s],\ \text{ if systems are simulated under CRN.}$$

Add any constraint ℓ, where $\ell=1,\ldots,s$, with increasing preference to set IP.

[**Initialization for Phase I:**]

for each system $i\in M$ **do**

– Obtain n_0 observations $Y_{i\ell 1}, Y_{i\ell 2}, \ldots, Y_{i\ell n_0}$ for $\ell=1,2,\ldots,s$. Also, obtain n_0 observations X_{in}, $n=1,\ldots,n_0$.

– Compute $\bar{Y}_{i\ell}(n_0)$ and $S^2_{Y_{i\ell}}(n_0)$.

– Compute $\bar{X}_i(n_0)$ and $S^2_{X_{ij}}(n_0)$ for all systems $j\neq i$.

– Set $r_i=n_0$, $\mathrm{ON}_i=\{1,2,\ldots,s\}$, and $\mathrm{ON}_{i\ell}=\{1,\ldots,d_\ell\}$ for $\ell=1,2,\ldots,s$.

end for

[**Feasibility Check:**]

for each system $i\in M$ **do**

  for $\ell\in\mathrm{ON}_i$ **do**

    for $m\in\mathrm{ON}_{i\ell}$ **do**,

      If $\bar{Y}_{i\ell}(r_i)+R(r_i;\epsilon_\ell,\eta_f,S^2_{Y_{i\ell}}(n_0))/r_i\le q_{\ell,m}$, set $Z_{i,\ell,m}=1$ and $\mathrm{ON}_{i\ell}=\mathrm{ON}_{i\ell}\setminus\{m\}$.

      If $\bar{Y}_{i\ell}(r_i)-R(r_i;\epsilon_\ell,\eta_f,S^2_{Y_{i\ell}}(n_0))/r_i\ge q_{\ell,m}$, set $Z_{i,\ell,m}=0$ and $\mathrm{ON}_{i\ell}=\mathrm{ON}_{i\ell}\setminus\{m\}$.

    end for

    If $\mathrm{ON}_{i\ell}=\emptyset$, set $\mathrm{ON}_i=\mathrm{ON}_i\setminus\{\ell\}$.

  end for

  If $\exists$ minimum $\kappa\le\theta$ s.t. $\prod_{\ell=1}^s Z_{i,\ell,I_\ell^{(\kappa)}}=1$, and either $\kappa<\theta$ or $i\notin F$, then

– If $\kappa<\theta$, then set $F=\emptyset$, $\theta=\kappa$, and for all $j\in M$ delete $q_{\ell,m}$ from $\mathrm{ON}_{j\ell}$ if $\ell\in\mathrm{IP}$ and $m>I_\ell^{(\theta)}$ (if $\ell\notin\mathrm{IP}$, then $q_{\ell,m}$ can be removed from $\mathrm{ON}_{j\ell}$ if $I_\ell^{(\theta')}\neq m$ for all $\theta'\le\kappa$), and set $\mathrm{ON}_j=\mathrm{ON}_j\setminus\{\ell\}$ if $\mathrm{ON}_{j\ell}=\emptyset$.

– Add system i to F.

  If $\prod_{\ell=1}^s Z_{i,\ell,I_\ell^{(\theta)}}=0$ or 1 and either $\theta=1$ or $\prod_{\ell=1}^s Z_{i,\ell,I_\ell^{(\kappa)}}=0$ for all $\kappa=1,\ldots,\theta-1$, then remove system i from M.

end for

[**Stopping Condition for Phase I:**]

If $M\neq\emptyset$, then for each system $i\in M$, set $r_i=r_i+1$, take one additional observation $Y_{i\ell r_i}$ and X_{i,r_i+1}, and update $\bar{Y}_{i\ell}(r_i)$ and $\bar{X}_i(r_i)$ for $\ell\in\mathrm{ON}_i$, then go to [**Feasibility Check**]. Else, check the following conditions.

– If $|F|=0$, stop and conclude no feasible systems;

– If $|F|=1$, stop and return the system in F as the best; or

– If $|F|>1$, go to [**Initialization for Phase II**].

[**Initialization for Phase II:**] Let η_c be a solution to $g(\eta_c)=\alpha'_c$, where

$$\alpha'_c=\begin{cases}1-(1-\alpha_c)^{1/(|F|-1)}, & \text{if systems are simulated independently;}\\ \alpha_c/(|F|-1), & \text{if systems are simulated under CRN.}\end{cases}$$

Let $M=F$ be the set of systems still in contention. Set $r=\min_{i\in F} r_i$ and go to [**Comparison**].

[**Comparison:**] For $i,j\in M$ s.t. $i\neq j$ and

$$r\bar{X}_i(r_i)>r\bar{X}_j(r_j)+R(r;\delta,\eta_c,S^2_{X_{ij}}(n_0)),$$

eliminate j from M.

[**Stopping Condition for Phase II:**] If $|M|=1$, then stop and return the system in M as the best. Otherwise, for each system $i\in M$ with $r_i\le r$, take one additional observation X_{i,r_i+1}, set $r_i=r_i+1$ and compute $\bar{X}_i(r_i)$. Then, set $r=r+1$ and go to [**Comparison**].

Similar to the discussion in [1], there are two difficulties in proving the statistical validity of $\mathcal{ZAK}$. First, as r_i, the number of observations X_{in} collected in Phase I, depends on $Y_{i\ell n}$ for system i, this dependency affects the comparison in Phase II. This dependency issue can be resolved by performing $\mathcal{ZAK}^{\mathcal{R}}$ instead as it restarts "from scratch" for the surviving systems of Phase I. Second, we use $g(\eta_c) = \alpha_c/(|F| - 1)$ instead of $g(\eta_c) = \alpha_c/(k-1)$ to compute the implementation parameter η_c for Phase II. Thus, we only allocate the nominal error for Phase II to the comparison between the best system $[b]$ and the surviving systems from Phase I, rather than all $k - 1$ other systems. As the comparison between $[b]$ and the other surviving systems is done with a larger nominal error, the resulting η_c is smaller, which helps improve the efficiency of our approach. However, the continuation region in Phase II now depends on the number of surviving systems from Phase I. We address the dependency between Phases I and II in $\mathcal{ZAK}$ by choosing the nominal errors α_f and α_c for Phases I and II as $\alpha_f + \alpha_c = \alpha$ to incorporate the correlation between the two phases. While $(1-\alpha_f)(1-\alpha_c)$ is always larger than $1 - (\alpha_f + \alpha_c)$, the difference is typically quite small. Although we have not proved the statistical validity of $\mathcal{ZAK}$, our experimental results (discussed in Section 6) do not show any violation of its validity.

The choices of α_f and α_c affect the performance of $\mathcal{ZAK}$. Similar to the discussion in Section B.1, the decision maker may choose $e_1 = \alpha_f/\alpha_c$ if she has knowledge on the relative difficulty of Phases I and II. The value of α_c can be found by solving $e_1 \times \alpha_c + \alpha_c = \alpha$, and the corresponding value of α_f can be found as $\alpha_f = e_1 \times \alpha_c$. If the decision maker does not have the information about the relative difficulty of Phases I and II, one possibility is to choose $\alpha_f = \alpha_c = \alpha/2$. Similar to $\mathcal{ZAK}^{\mathcal{R}}$, another possibility is to choose $e_2 = s\alpha'_f/\alpha'_c$ if $s \le d$ or to choose $e_2 = d\alpha'_f/\alpha'_c$ if $d < s$. Appendix B.2 provides a detailed discussion on how to set the implementation parameters α'_f, α'_c for Phase I.

4 SIMULTANEOUSLY-RUNNING PROCEDURE

In this section, we provide a procedure that implements Phases I and II simultaneously. This procedure aims to solve the problem from a different perspective. Specifically, by implementing Phase I and II simultaneously, the elimination of inferior and infeasible systems can happen simultaneously throughout the procedure. This procedure increases the opportunity to eliminate systems whose feasibility are still unknown but are clearly inferior to a certain system. As a result, the procedure is expected to be more efficient than the sequentially-running procedure. Section 4.1 describes the simultaneously-running procedure and Section 4.2 proves its statistical validity.

4.1 Procedure $\mathcal{ZAK}$+

In this section, we provide a procedure that runs Phases I and II simultaneously in Algorithm 2. Similar to the sequentially-running procedures $\mathcal{ZAK}^{\mathcal{R}}$ and $\mathcal{ZAK}$, we use the variable θ to keep track of the current most preferred threshold vector for which we are trying to determine feasibility. Initially, θ is set to d, which is the index of the least preferred threshold vector. We use sets M and F defined as in Section 3 and additionally define set SS_i as a set of systems found to be superior to system i in terms of the primary performance measure.

Rather than performing Phase II on the surviving systems from Phase I as $\mathcal{ZAK}^{\mathcal{R}}$ and $\mathcal{ZAK}$ do, we now perform both feasibility check and pairwise comparison for all systems that are still in consideration (i.e., $i \in M$) within each iteration. More specifically, for each system $i \in M$, we check whether there exists a minimum threshold vector that system i is feasible with respect to, use θ to keep track of this threshold index, and update set F if appropriate. When a feasible decision is made for system i, we perform an additional step in Phase I: eliminate system $j \in (M \cup F)$ if $i \in SS_j$ (system $i \in F$ is shown to be superior compared with system j) and system j is not feasible with respect to any of $\mathbf{q}^{(1)}, \ldots, \mathbf{q}^{(\theta-1)}$. In Phase II, once a system i is declared superior compared with system j in Phase II, we add system i to SS_j. Furthermore, if system $i \in (F \cap SS_j)$ and system j is infeasible with respect to all $\mathbf{q}^{(1)}, \ldots, \mathbf{q}^{(\theta-1)}$, then we eliminate system j from M and F.

Note that simultaneously-running procedures in [1] and [8] also use sets M, F, and SS_j, and their [**Comparison**] step is similar in the sense that pairwise comparison is performed among the systems whose superiority is not yet determined. However, the procedures in [1] and [8] are designed for a fixed set of thresholds, and thus there is no search for the most preferred threshold vector θ, and there is no resetting of set F. By contrast, $\mathcal{ZAK}$+ checks if a more preferred threshold vector is found at each iteration. Whenever a more preferred threshold vector is found, the index θ and F are reset, and systems feasible to the updated threshold vector θ are added to the reset set F.

A detailed description of the simultaneously-running procedure $\mathcal{ZAK}$+ is shown in Algorithm 2.

4.2 Statistical Validity of the Simultaneously Running Procedure

In this section, we present the proof of the statistical validity of the simultaneously-running procedure $\mathcal{ZAK}$+. Before presenting the main results, we need more definitions. Let θ^* be defined as in Section 2.2. We define the sets $S_a, S_u, S_{a'}$, and S_d as follows:

$$S_a = \text{ set of acceptable systems with respect to at least one of the threshold vectors } \mathbf{q}^{(1)}, \ldots, \mathbf{q}^{(\theta^*-1)};$$

$$S_u = \begin{cases} \text{set of unacceptable systems with respect to } \mathbf{q}^{(\theta^*)} \text{ among systems in } \Gamma \setminus S_a, & \text{if } \theta^* \leq d; \\ \Gamma \setminus S_a, & \text{if } \theta^* = d+1; \end{cases}$$

$$S_{a'} = \begin{cases} \text{set of acceptable systems with respect to } \mathbf{q}^{(\theta^*)} \text{ among systems in } \Gamma \setminus S_a, & \text{if } \theta^* \leq d; \\ \emptyset, & \text{if } \theta^* = d+1; \end{cases}$$

$$S_d = \begin{cases} \text{set of desirable systems with respect to } \mathbf{q}^{(\theta^*)} \text{ among systems in } \Gamma \setminus (S_a \cup \{[b]\}), & \text{if } \theta^* \leq d; \\ \emptyset, & \text{if } \theta^* = d+1. \end{cases}$$

We then let $j_a = |S_a|, j_{a'} = |S_{a'}|, j_d = |S_d|$, and $j_u = |S_u|$, and therefore $j_a + j_{a'} + j_d + j_u + \mathbb{1}(\theta^* \leq d) = k$. For correct selection, we must select a system in $S_a \cup \{[b]\}$ and eliminate the systems in $S_{a'} \cup S_d \cup S_u$ when $\theta^* \leq d$ (under Assumption 2); when $\theta^* = d+1$, CS involves eliminating all systems in S_u, and either declaring all systems infeasible or selecting a system in S_a.

To illustrate, recall the problem demonstrated in Figure 3, where $\theta^* = 5$. Figure 3 shows systems a and b as two examples of acceptable systems with respect to preferred threshold vectors (i.e., $a, b \in S_a$). Note that system a is acceptable with respect to $\mathbf{q}^{(1)}, \mathbf{q}^{(2)}, \mathbf{q}^{(3)}$, and $\mathbf{q}^{(4)}$ and desirable with respect to $\mathbf{q}^{(5)}$, while system b is acceptable with respect to $\mathbf{q}^{(3)}$ but unacceptable to $\mathbf{q}^{(1)}, \mathbf{q}^{(2)}, \mathbf{q}^{(4)}$, and $\mathbf{q}^{(5)}$. System c is acceptable with respect to $\mathbf{q}^{(5)}$ (i.e., $c \in S_{a'}$) and unacceptable with respect to $\mathbf{q}^{(1)}, \ldots, \mathbf{q}^{(4)}$.

We then introduce the following definitions for $i \in \Gamma$ and present two lemmas that are essential in proving the statistical validity of $\mathcal{ZAK}$+.

$$\mathcal{A}_1^*(i) = \left\{\text{system } i \text{ is declared infeasible to } \mathbf{q}^{(1)}, \ldots, \mathbf{q}^{(\min\{\theta^*, d\})}\right\};$$

$$\mathcal{A}_2^*(i) = \left\{\text{system } i \text{ is declared infeasible to } \mathbf{q}^{(1)}, \ldots, \mathbf{q}^{(\theta^*-1)} \text{ if } 1 < \theta^* \leq d\right\};$$

$$\mathcal{B}_1^* = \left\{\text{system } [b] \text{ is declared feasible to } \mathbf{q}^{(\theta^*)} \text{ if } \theta^* \leq d\right\}.$$

LEMMA 4.1. *Under Assumption 1, for a particular system i, the* [**Feasibility Check**] *steps in $\mathcal{ZAK}$+ ensure*

$$\Pr\left(\mathcal{A}_1^*(i)\right) \geq 1 - \min\{s, d\}\beta_f, \text{ if } i \in S_u;$$

ALGORITHM 2: $\mathcal{ZAK}$+.

[**Setup:**] Choose confidence level $1-\alpha$, tolerance levels $\epsilon_1,\ldots\epsilon_s$, indifference-zone parameter δ, threshold vectors $\{\mathbf{q}^{(1)},\mathbf{q}^{(2)},\ldots,\mathbf{q}^{(d)}\}$, and associated index vectors $\{\boldsymbol{I}^{(1)},\boldsymbol{I}^{(2)},\ldots,\boldsymbol{I}^{(d)}\}$. Set $M=\Gamma, SS_i=\emptyset$, and $Z_{i,\ell,m}=2$ for all $i\in M,\ell=1,\ldots,s$, and $m=1,\ldots,d_\ell$. Set $F=\emptyset$ and $\theta=d$. Choose $0<\beta_f<1/s, 0<\beta_c<1$ that satisfy

$$\min_{0\le j\le k-1}\left\{(1-\min\{s,d\}\beta_f)^j\times\left[(1-\min\{s,d-1\}\beta_f-\beta_c)^{k-j-1}-s\beta_f\right]\right\}$$
$$=1-\alpha \text{ and } 0<1-\min\{s,d-1\}\beta_f-\beta_c<1,$$
$$\text{if systems are simulated independently;}$$

$$\min_{0\le j\le k-1}\left\{1-[j\min\{s,d\}+(k-j-1)\min\{s,d-1\}+s]\,\beta_f-(k-j-1)\beta_c\right\}=1-\alpha,$$
$$\text{if systems are simulated under CRN.}$$

Set η_f and η_c such that $g(\eta_f)=\beta_f$ and $g(\eta_c)=\beta_c$. Add any constraint ℓ, where $\ell=1,\ldots,s$, with increasing preference to set IP.

[**Initialization:**]

for each system $i\in M$ **do**

— Obtain n_0 observations from system i.

— Compute $\bar{X}_i(n_0), \bar{Y}_{i\ell}(n_0), S^2_{X_{ij}}(n_0)$, and $S^2_{Y_{i\ell}}(n_0)$ for all $i,j\in M$, where $i\ne j$, and $\ell=1,\ldots,s$.

— Set $r=n_0$, $\mathrm{ON}_i=\{1,\ldots,s\}$, and $\mathrm{ON}_{i\ell}=\{1,\ldots,d_\ell\}$ for $\ell=1,\ldots,s$.

end for

[**Feasibility Check:**]

for $i\in M$ **do**

  for $\ell\in \mathrm{ON}_i$ **do**

    for $m\in\mathrm{ON}_{i\ell}$ **do**

      If $\bar{Y}_{i\ell}(r)+R(r;\epsilon_\ell,\eta_f,S^2_{Y_{i\ell}}(n_0))/r\le q_{\ell,m}$, set $Z_{i,\ell,m}=1$ and $\mathrm{ON}_{i\ell}=\mathrm{ON}_{i\ell}\setminus\{m\}$;

      If $\bar{Y}_{i\ell}(r)-R(r;\epsilon_\ell,\eta_f,S^2_{Y_{i\ell}}(n_0))/r\ge q_{\ell,m}$, set $Z_{i,\ell,m}=0$ and $\mathrm{ON}_{i\ell}=\mathrm{ON}_{i\ell}\setminus\{m\}$.

    end for

    If $\mathrm{ON}_{i\ell}=\emptyset$, set $\mathrm{ON}_i=\mathrm{ON}_i\setminus\{\ell\}$.

  end for

  If $\exists$ minimum $\kappa\le\theta$ s.t. $\prod_{\ell=1}^s Z_{i,\ell,I_\ell^{(\kappa)}}=1$, and either $\kappa<\theta$ or $i\notin F$, then

— If $\kappa<\theta$, then set $F=\emptyset,\theta=\kappa$, and for all $j\in M$ delete $q_{\ell,m}$ from $\mathrm{ON}_{j\ell}$ if $\ell\in$ IP and $m>I_\ell^{(\theta)}$ (if $\ell\notin$ IP, then $q_{\ell,m}$ can be removed from $\mathrm{ON}_{j\ell}$ if $I_\ell^{(\theta')}\ne m$ for all $\theta'\le\kappa$), and set $\mathrm{ON}_j=\mathrm{ON}_j\setminus\{\ell\}$ if $\mathrm{ON}_{j\ell}=\emptyset$.

— Add system i to F.

— For all $j\in M$, if $i\in SS_j$ and either $\theta=1$ or $\prod_{\ell=1}^s Z_{j,\ell,I_\ell^{(\kappa)}}=0$ for all $\kappa=1,\ldots,\theta-1$, then remove system j from M and F (if $j\in F$) and delete SS_j.

  If either $\prod_{\ell=1}^s Z_{i,\ell,I_\ell^{(\kappa)}}=0$ for all $1\le\kappa\le\theta$, or $\theta>1$, $\prod_{\ell=1}^s Z_{i,\ell,I_\ell^{(\kappa)}}=0$ for all $1\le\kappa\le\theta-1$, and there exists $j\in F\cap SS_i$, then remove i from M and delete SS_i.

end for

[**Comparison:**] For $i,j\in M$ s.t. $i\ne j, i\notin SS_j, j\notin SS_i$, and

$$r\bar{X}_i(r)>r\bar{X}_j(r)+R(r;\delta,\eta_c,S^2_{X_{ij}}(n_0)),$$

add system i to SS_j. If $i\in F$, then remove system j from M and F (if $j\in F$) if either $\theta=1$ or $\prod_{\ell=1}^s Z_{j,\ell,I_\ell^{(\kappa)}}=0$ for all $\kappa=1,\ldots,\theta-1$, and delete SS_j.

[**Stopping Condition:**] If $M=F$ and $|F|=1$, then stop and return the system in F as the best system. Else if $M=F$ and $|F|=0$, then stop and return no feasible systems exist. Otherwise, for all $i\in M$, set $r=r+1$, take one additional observation, update $\bar{X}_i(r)$ and $\bar{Y}_{i\ell}(r)$ for all $\ell\in\mathrm{ON}_i$, and go to [**Feasibility Check**].

$$\Pr\left(\mathcal{A}_2^*(i)\right) \geq 1 - \min\{s, d-1\}\beta_f,\ \textit{if } i \in S_d \cup S_{a'} \textit{ and } 1 < \theta^* \leq d;$$
$$\Pr\left(\mathcal{B}_1^*\right) \geq 1 - s\beta_f,\ \textit{if } \theta^* \leq d.$$

LEMMA 4.2. *Under Assumption 1, given i such that $x_i \leq x_{[b]} - \delta$, the* **[Comparison]** *steps in $\mathcal{ZAK}$+ run to completion ensure* $\Pr(\mathrm{CS}_i) \geq 1 - \beta_c$.

The proofs of Lemmas 4.1 and 4.2 are essentially the same as those of Lemmas A.2 and A.3 that are used to prove the statistical validity of $\mathcal{ZAK}^{\mathcal{R}}$. This is because both α'_f of $\mathcal{ZAK}^{\mathcal{R}}$ and β_f of $\mathcal{ZAK}$+ are the nominal error of feasibility check for one constraint of one system with a fixed threshold, and both α'_c of $\mathcal{ZAK}^{\mathcal{R}}$ and β_c of $\mathcal{ZAK}$+ are the nominal error of comparison between an inferior system and the best system $[b]$.

We are now ready to prove the statistical validity of $\mathcal{ZAK}$+.

THEOREM 4.3. *Under Assumptions 1 and 2, the $\mathcal{ZAK}$+ procedure guarantees* $\Pr\{\mathrm{CS}\} \geq 1 - \alpha$.

The proof of Theorem 4.3 is provided in Appendix C.

We now discuss how to choose implementation parameters β_f and β_c in simultaneous-running procedure $\mathcal{ZAK}$+. One approach is to first decide the choice of $e = s\beta_f/\beta_c$ when $s < d$ and $e = d\beta_f/\beta_c$ when $s \geq d$. Recall that this is the ratio of the error for a feasibility check of one system for all constraints and all thresholds to the error of a comparison between two systems. The ratio should be decided based on the decision maker's idea on whether she wants to allocate more error to feasibility check or comparison. A detailed discussion on how we compute β_f and β_c is included in Appendix D.

In reality, as the decision maker usually does not have detailed information on the mean performance measures, choosing the value of e is not straightforward. [8] consider a single threshold vector and choose $e = 1$ to balance the errors assigned to feasibility check and comparison. As it is reasonable to allocate more of the errors to feasibility check when multiple threshold vectors are under consideration, we use $e = 2$ for our experimental results to demonstrate the performance of our proposed procedure (based on the discussion in Section 6.2).

5 DIFFERENT PREFERENCE ORDERS OF INPUT THRESHOLD VECTORS

As discussed in Section 1, our procedures $\mathcal{ZAK}^{\mathcal{R}}$, $\mathcal{ZAK}$, and $\mathcal{ZAK}$+ require lists of threshold vectors $\{\mathbf{q}^{(1)}, \mathbf{q}^{(2)}, \ldots, \mathbf{q}^{(d)}\}$ and index vectors $\{\boldsymbol{I}^{(1)}, \boldsymbol{I}^{(2)}, \ldots, \boldsymbol{I}^{(d)}\}$. Having to manually enter preference order is tedious from both a problem formulation and implementation points of view. Techniques for facilitating this makes our approach more practical and useful. In this section, we discuss three preference orders for formulating the input threshold vectors, namely ranked constraints, equally important constraints, and total violation with ranked constraints. The experimental results for multiple constraints shown in Section 6 are based on these three preference orders.

Ranked constraints: The constraints are ranked with respect to their importance and the decision maker wants to relax the least important constraint first while keeping the rest of the constraints fixed at the current threshold values, and then move to the second least important constraint, and so on. Figure 2(a) shows A_θ for $\theta = 1, \ldots, 9$ when $s = 2$ and $d_1 = d_2 = 3$, the secondary performance measures are non-negative, and constraint 1 is more important than constraint 2. The inventory example discussed in Sections 1 and 2 also has ranked constraints with constraint 1 being more important than constraint 2.

Equally important constraints: All constraints are equally important and the decision maker wants to relax all constraints by one threshold at the same time. If the constraints do not all have the same number of thresholds, then constraints that have gone through all their thresholds keep the "loosest" threshold (i.e., q_{ℓ,d_ℓ} for constraint ℓ) while the other constraints relax. Figure 2(b) shows this case for two constraints and three thresholds on each constraint.

Total violation with ranked constraints: The decision maker wants to minimize the number of total violations on ranked constraints. For constraint ℓ with threshold $q_{\ell,m}$, its violation is defined as $m-1$ (relative to the tightest threshold $q_{\ell,1}$). Then the total violation is defined as the sum of the violations for all constraints. The decision maker always prefers threshold vectors that have fewer total violations, and among threshold vectors that have the same total violation, her preference order is based the priority of the constraints. In Figure 2(c), constraint 1 more important than constraint 2. Region A_1 is defined with respect to $(q_{1,1}, q_{2,1})$ and has total violation 0. Regions A_2 and A_3 are defined with respect to $(q_{1,1}, q_{2,2})$ and $(q_{1,2}, q_{2,1})$, respectively, and have total violation 1, with A_2 preferred to A_3 due to the ranking of constraints 1 and 2. In this preference order, we start with a threshold vector with total violation equal to 0 and then relax the total violation by relaxing the less important constraint first. The largest total violation is $\sum_{\ell=1}^{s}(d_\ell - 1)$.

The detailed algorithm statements of how to construct the three preference orders are included in Appendix E.

6 EXPERIMENTAL RESULTS

In this section, we present experimental results to demonstrate the performances of our proposed procedures $\mathcal{ZAK}^{\mathcal{R}}$, $\mathcal{ZAK}$, and $\mathcal{ZAK}+$. We compare them with alternative procedures that iteratively apply sequential or simultaneous procedures to threshold vectors $\mathbf{q}^{(1)}, \ldots, \mathbf{q}^{(d)}$. If a single constraint is considered, our alternative procedures use $\mathcal{AK}$ or $\mathcal{AK}+$ due to [1] for each threshold value. If multiple constraints are considered, our alternative procedures use $\mathcal{HAK}$ or $\mathcal{HAK}+$ due to [8] for each threshold vector. We name the procedures that iteratively implement $\mathcal{AK}$ and $\mathcal{AK}+$ as Restart$^{\mathcal{AK}}$ and Restart$^{\mathcal{AK}+}$, respectively. Similarly, we name the procedures that iteratively implement $\mathcal{HAK}$ and $\mathcal{HAK}+$ as Restart$^{\mathcal{HAK}}$ and Restart$^{\mathcal{HAK}+}$, respectively. Notice that Restart$^{\mathcal{AK}}$ (Restart$^{\mathcal{AK}+}$) is the special case of Restart$^{\mathcal{HAK}}$ (Restart$^{\mathcal{HAK}+}$) when the number of constraints is one and therefore does not need to be considered separately. We provide the algorithm statements and discussions of the statistical validity of procedures Restart$^{\mathcal{HAK}}$ and Restart$^{\mathcal{HAK}+}$ in Appendices F and G, respectively.

All the experimental results are based on 10,000 macro replications with $\alpha = 0.05$ and $n_0 = 20$ and we report average numbers of observations (OBS) and estimated PCS. We set $k = 100$ and $\delta = \epsilon_\ell = 1/\sqrt{n_0}$, where $\ell = 1, \ldots, s$. We discuss the experimental configurations in Section 6.1 and how we set the implementation parameters for our proposed procedures in Section 6.2. We then provide the experimental results to show that our proposed procedures are statistically valid and efficient in Sections 6.3 and 6.4, respectively. Experimental results for the inventory example discussed in Sections 1 and 2 are provided in Section 7. Appendix J discusses the impact of applying CRN in our proposed procedures.

6.1 Experimental Configurations

In this section, we discuss the mean and variance configurations for primary and secondary performance measures. We consider three mean configurations of systems, namely **difficult means** (**DM**), **monotone increasing means** (**MIM**), and **monotone decreasing means** (**MDM**). All the configurations depend on the number of systems b that are desirable with respect to threshold vector $\mathbf{q}^{(\theta^*)}$. As the existence of acceptable systems will not lower the PCS (because declaring acceptable systems feasible or infeasible with respect to a specific threshold value are both considered as correct feasibility decisions) and as [1] show by experiments that the presence of acceptable systems does not significantly affect the overall performance of procedures $\mathcal{AK}$ and $\mathcal{AK}+$, we do not include acceptable systems in our three configurations.

As the purpose of the DM configuration is to demonstrate the performance of the proposed procedures under a difficult case, we set the difference between any two consecutive thresholds on one constraint to the minimum possible value, so that the boundary of the unacceptable region of $q_{\ell,m}$ is the boundary of the desirable region of $q_{\ell,m+1}$. This is achieved by setting $q_{\ell,m+1} - q_{\ell,m}$ equal to $2\epsilon_\ell$ for all m and ℓ. When $\theta^* < d$, the means of all secondary performance measures are set to the boundary of the desirable region of $\mathbf{q}^{(\theta^*)}$ for b systems (i.e., the mean of secondary performance measure ℓ for b systems is $q_\ell^{(\theta^*)} - \epsilon_\ell$). For the other $(k-b)$ systems, to make the feasibility check difficult, the means of their secondary measures are set to the boundary of the desirable region of $\mathbf{q}^{(\theta^*+1)}$ (i.e., the means of secondary performance measure ℓ for $(k-b)$ systems is $q_\ell^{(\theta^*+1)} - \epsilon_\ell$). When $\theta^* = d$, the b systems that are feasible to $\mathbf{q}^{(\theta^*)}$ are set the same as when $\theta^* < d$. For the remaining $(k-b)$ systems, we set them at the boundary of the unacceptable region for the largest threshold of all constraints ℓ (i.e., $y_{i\ell} = q_{\ell,d_\ell} + \epsilon_\ell$ when $i = b+1, \ldots, k$). When $\theta^* = d+1$, as all systems are infeasible to all the threshold vectors considered (i.e., $b = 0$), the means of the secondary performance measures of all the systems are set as $y_{i\ell} = q_{\ell,d_\ell} + \epsilon_\ell$ for all i and ℓ.

Moreover, the DM configuration has one system whose mean performance of the primary performance is δ, the other systems that are feasible with respect to $\mathbf{q}^{(\theta^*)}$ have primary performances equal to 0, and all infeasible systems with respect to $\mathbf{q}^{(\theta^*)}$ have 2δ as their primary performance measures. This means that all the infeasible systems are superior compared with the best system while all other feasible systems are only δ inferior compared with the best system, which makes the comparison also difficult. More specifically, in the DM configuration,

$$x_i = \mathrm{E}\left[X_{in}\right] = \begin{cases} 0, & i = 1, 2, \ldots, b-1, \\ \delta, & i = b, \\ 2\delta, & i = b+1, \ldots, k. \end{cases}$$

For all constraints $\ell = 1, \ldots, s$, if $1 \le \theta^* \le d-1$,

$$y_{i\ell} = \mathrm{E}[Y_{i\ell n}] = \begin{cases} q_\ell^{(\theta^*)} - \epsilon_\ell, & i = 1, 2, \ldots, b, \\ q_\ell^{(\theta^*+1)} - \epsilon_\ell, & i = b+1, \ldots, k; \end{cases}$$

if $\theta^* = d$,

$$y_{i\ell} = \mathrm{E}[Y_{i\ell n}] = \begin{cases} q_\ell^{(\theta^*)} - \epsilon_\ell, & i = 1, 2, \ldots, b, \\ q_{\ell,d_\ell} + \epsilon_\ell, & i = b+1, \ldots, k; \end{cases}$$

and if $\theta^* = d+1$, $y_{i\ell} = \mathrm{E}[Y_{i\ell n}] = q_{\ell,d_\ell} + \epsilon_\ell$ for all i. We consider the case when the decision maker prefers threshold $q_{\ell,1} = 0$ for constraint ℓ, and relax the constraint threshold by $2\epsilon_\ell$ every time when she wants to consider a "looser" threshold value on that constraint. For example, we choose thresholds $\{0, 2\epsilon_\ell\}$ and $\{0, 2\epsilon_\ell, 4\epsilon_\ell, 6\epsilon_\ell\}$ on constraint ℓ when there are two or four thresholds in consideration, respectively.

On the other hand, as the purpose of the MIM and the MDM configurations is to show the efficiency of the proposed procedures in realistic settings, we set the differences between two consecutive thresholds larger than in the DM configuration to see how effectively the proposed procedures remove infeasible systems. In particular, we choose the smallest distance between two consecutive thresholds on constraint ℓ in the MIM and MDM configurations as $4\epsilon_\ell$. When $\theta^* \le d$, the means of all secondary performance measures are set to the interior of the desirable region of $\mathbf{q}^{(\theta^*)}$ for b systems and the other $(k-b)$ systems are evenly distributed over the interiors of

$A_{(\theta^*+1)},\ldots,A_{(d+1)}$ with respect to their secondary performance measures (i.e., for the systems in $A_{(\theta)}$, the mean of secondary performance measure ℓ is set within the desirable region of $\mathbf{q}^{(\theta)}$ as $q_\ell^{(\theta)}-2\epsilon_\ell$ where $\theta=\theta^*,\ldots,d$, and as $q_{\ell,d_\ell}+2\epsilon_\ell$ when $\theta=d+1$). When $\theta^*=d+1$, we set the means of the secondary performance measures to the interior of the unacceptable region for the largest thresholds of all constraints ℓ (i.e., $y_{i\ell}=q_{\ell,d_\ell}+2\epsilon_\ell$ for all constraint ℓ). We also let the means of the primary performance measure be monotonically increasing from 0 with an increment of δ for the MIM configuration, and let the primary performance measure be monotonically decreasing from $(k-1)\delta$ with a decrement of δ for the MDM configuration. This makes the comparison easier than in the DM configuration.

More specifically, we set $x_i=\mathrm{E}[X_{in}]=(i-1)\delta, i=1,\ldots,k$ for the MIM configuration and set $x_i=\mathrm{E}[X_{in}]=(k-i)\delta, i=1,\ldots,k$ for the MDM configuration. The means of the secondary performance measures of the MIM and the MDM configurations are the same. For all constraints $\ell=1,\ldots,s$, if $1\le\theta^*\le d$,

$$y_{i\ell}=\mathrm{E}[Y_{i\ell n}]=\begin{cases} q_\ell^{(\theta^*)}-2\epsilon_\ell, & i=1,2,\ldots,b,\\ q_\ell^{(\theta^*+1)}-2\epsilon_\ell, & i=b+1,\ldots,\lceil b+\frac{k-b}{d+1-\theta^*}\rceil,\\ q_\ell^{(\theta^*+2)}-2\epsilon_\ell, & i=\lceil b+\frac{k-b}{d+1-\theta^*}\rceil+1,\ldots,\lceil b+2\frac{k-b}{d+1-\theta^*}\rceil,\\ \cdots & \\ q_\ell^{(d)}-2\epsilon_\ell, & i=\lceil b+(d-\theta^*-1)\frac{k-b}{d+1-\theta^*}\rceil+1,\ldots,\lceil b+(d-\theta^*)\frac{k-b}{d+1-\theta^*}\rceil,\\ q_{\ell,d_\ell}+2\epsilon_\ell, & i=\lceil b+(d-\theta^*)\frac{k-b}{d+1-\theta^*}\rceil+1,\ldots,k; \end{cases}$$

and if $\theta^*=d+1$, $y_{i\ell}=\mathrm{E}[Y_{i\ell n}]=q_{\ell,d_\ell}+2\epsilon_\ell$ for all i. The decision maker prefers $q_{\ell,1}=0$, and we relax the constraint threshold by $4\epsilon_\ell$ when she wants to consider "looser" threshold values. For example, for the cases of two and four thresholds, we choose thresholds $\{0,4\epsilon_\ell\}$ and $\{0,4\epsilon_\ell,8\epsilon_\ell,12\epsilon_\ell\}$ on constraint ℓ, respectively.

We consider three variance configurations to test different levels of relative difficulty of the feasibility check and the comparison. We use $\sigma_{x_i}^2$ to denote the variance of the primary performance from system i, $\sigma_{y_{i\ell}}^2$ to denote the variance of the secondary performance ℓ from system i, and consider both low variance (L) and high variance (H). When the difficulty between feasibility checks and comparison are similar, we set $\sigma_{x_i}^2=1$ and $\sigma_{y_{i\ell}}^2=1$ (L/L); when the comparison is relatively more difficult than the feasibility checks, we set $\sigma_{x_i}^2=5$ and $\sigma_{y_{i\ell}}^2=1$ (H/L); and when the feasibility checks are relatively more difficult than comparison, we set $\sigma_{x_i}^2=1$ and $\sigma_{y_{i\ell}}^2=5$ (L/H).

Reference [1] shows that the correlation between the primary and secondary performance measures does not have a significant impact on the experimental results. References [8] and [22] also report the same tendency. Therefore, we assume the observations for the primary and secondary performance measures from each system are independent normal random variables through Sections 6.2–6.4. Section 7 illustrates how to apply our procedures in an inventory example where the observations are not necessarily normally distributed, the primary and secondary performance measures are correlated, and the secondary performance measures are also correlated.

With 10,000 macro replications, the first four digits of the OBS showed in the tables are meaningful, and the estimated PCS values are meaningful up to the 0.001th digit.

6.2 Implementation Parameters

In this section, we discuss how we set the implementation parameters e_1, e_2, and e for the proposed procedures $\mathcal{ZAK}^{\mathcal{R}}$, $\mathcal{ZAK}$, and $\mathcal{ZAK}+$.

As discussed in Appendix B, we introduce two approaches of setting the implementation parameters for procedures $\mathcal{ZAK}^R$ and $\mathcal{ZAK}$, namely setting $e_1 = \alpha_f/\alpha_c$ and setting $e_2 = s\alpha'_f/\alpha'_c$. We let $\mathcal{ZAK}_1^{\mathcal{R}}$ ($\mathcal{ZAK}_1$) denote the version of procedure $\mathcal{ZAK}^{\mathcal{R}}$ ($\mathcal{ZAK}$) that sets the implementation parameter as $e_1 = \alpha_f/\alpha_c$. Similarly, we let $\mathcal{ZAK}_2^{\mathcal{R}}$ ($\mathcal{ZAK}_2$) be the corresponding procedure that uses $e_2 = s\alpha'_f/\alpha'_c$. Note that $\mathcal{ZAK}$+ only has one setting of its implementation parameters, namely $e = s\beta_f/\beta_c$, as discussed in Sections 4.2 and Appendix D.

For brevity, experimental settings and results are given in Appendix H. As discussed in Appendices B and D, the optimal values of e_1, e_2, or e (that result in the smallest OBS) depend on the mean and variance configurations of the primary and secondary performance measures of the systems. In the experimental results we test, the difficulty of feasibility check for one threshold of one constraint is similar as for comparing one system with the best system $[b]$. This suggests that $e_1 = e_2 = e = s$ might be a good choice. In fact, the OBS achieves its minimum value for different choices of e_1, e_2, or e ranging from 1 to 7. In addition, we notice that the OBS is quite flat around the e_1, e_2, or e with the smallest OBS for each proposed procedure. We also notice that the OBS is similar between the two settings of the implementation parameters (e_1 and e_2) of $\mathcal{ZAK}^{\mathcal{R}}$ and $\mathcal{ZAK}$, respectively. For these reasons, in the remaining sections we only consider $\mathcal{ZAK}_2^{\mathcal{R}}$ and $\mathcal{ZAK}_2$ with $e_2 = 2 = s\alpha'_f/\alpha'_c$ and $\mathcal{ZAK}$+ with $e = 2 = s\beta_f/\beta_c$ (see also the discussion in Appendices B and D). In all cases, the minimum OBS is no more than 2.36% from the OBS when e_2 or e equals 2.

6.3 Statistical Validity

In this section, we present experimental results that document the statistical validity of our proposed procedures. The experimental results shown in this section are all under the DM mean configuration since correct selection is more difficult in the DM mean configuration than in the MIM or MDM mean configurations.

We first consider the case of a single constraint with four thresholds $\{0, 2\epsilon_1, 4\epsilon_1, 6\epsilon_1\}$. Table 1 shows the estimated PCS under our three variance configurations and all possible θ^* when $b \in \{25, 50, 75\}$ (except that $b = 0$ when $\theta^* = 5$ because all systems are infeasible). We see that the estimated PCS values of all proposed procedures are above the nominal level 0.95 under all variance configurations, all possible values of θ^*, and all values of b considered. One may also notice that $\theta^* = 5$ and $\theta^* = 1$ (to a lesser extent) achieve higher estimated PCS compared with other values of θ^*. During Phase I, one needs to ensure that three events happen, namely declaring systems in S_u infeasible to threshold vectors $\mathbf{q}^{(1)}, \ldots, \mathbf{q}^{(\theta^*)}$, declaring the best system $[b]$ feasible to $\mathbf{q}^{(\theta^*)}$, and declaring systems in $S_{a'} \cup S_d$ infeasible to threshold vectors $\mathbf{q}^{(1)}, \ldots, \mathbf{q}^{(\theta^*-1)}$ (see the detailed analysis in Sections A and 4.2). Moreover, when $\theta^* = d + 1$, the best system does not exist and therefore we do not need to perform Phase II to achieve CS. As a more preferred threshold vector does not exist when $\theta^* = 1$ and the best system does not exist when $\theta^* = 5$, we have fewer sources of error and therefore achieve a higher estimated PCS under those two cases.

Table 1 also indicates that for $1 < \theta^* \leq d$, the estimated PCS decreases in general when b increases. As the three events required by Phase I involve essentially making one difficult feasibility decision correctly for each system (i.e., declaring systems in S_u infeasible to $\mathbf{q}^{(\theta^*)}$, declaring system $[b]$ feasible to $\mathbf{q}^{(\theta^*)}$, and declaring the remaining $b-1$ systems infeasible to $\mathbf{q}^{(\theta^*-1)}$), different values of b do not affect the difficulty of Phase I much. However, increasing b requires more correct comparison decisions in order to eliminate the inferior systems (compared to $[b]$) that are feasible to $\mathbf{q}^{(\theta^*)}$ in Phase II. Combining Phases I and II, the estimated PCS is expected to decrease as b increases. On the other hand, when $\theta^* = 1$, as there does not exist threshold vector $\mathbf{q}^{(\theta^*-1)}$, there is

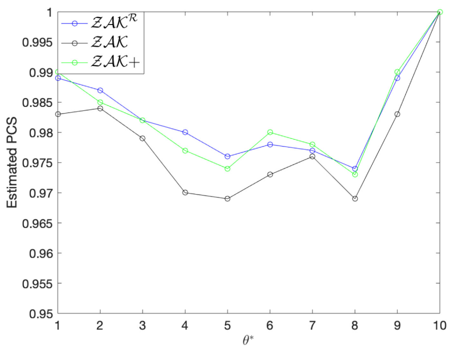
Table 1. Estimated PCS of $\mathcal{ZAK}^{\mathcal{R}}$, $\mathcal{ZAK}$, and $\mathcal{ZAK}$+ for $k = 100$ Systems and $s = 1$ Constraint with four Thresholds under the DM Configuration

		$\mathcal{ZAK}^{\mathcal{R}}$			$\mathcal{ZAK}$			$\mathcal{ZAK}$+		
	θ^*	$b = 25$	$b = 50$	$b = 75$	$b = 25$	$b = 50$	$b = 75$	$b = 25$	$b = 50$	$b = 75$
L/L	1	0.985	0.986	0.985	0.979	0.981	0.987	0.986	0.986	0.987
	2	0.977	0.971	0.964	0.971	0.971	0.963	0.977	0.972	0.967
	3	0.976	0.971	0.961	0.973	0.968	0.967	0.977	0.973	0.967
	4	0.981	0.969	0.967	0.974	0.969	0.965	0.978	0.973	0.962
	5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
L/H	1	0.984	0.986	0.987	0.986	0.991	0.995	0.985	0.987	0.988
	2	0.976	0.967	0.962	0.980	0.978	0.973	0.978	0.970	0.969
	3	0.977	0.967	0.966	0.980	0.973	0.972	0.978	0.973	0.964
	4	0.977	0.971	0.963	0.977	0.977	0.973	0.980	0.968	0.968
	5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
H/L	1	0.985	0.986	0.986	0.978	0.977	0.983	0.984	0.988	0.986
	2	0.978	0.970	0.965	0.969	0.965	0.964	0.977	0.973	0.964
	3	0.979	0.970	0.963	0.970	0.964	0.962	0.977	0.972	0.964
	4	0.979	0.973	0.967	0.969	0.964	0.961	0.979	0.970	0.968
	5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

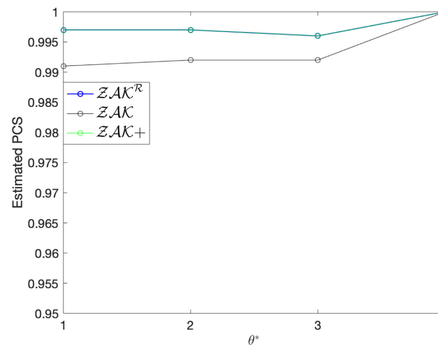
one less source of concluding incorrect decisions in Phase I (i.e., declaring $b - 1$ systems infeasible to $\mathbf{q}^{(\theta^*-1)}$). Thus increasing b makes Phase I less difficult. Combining Phases I and II, depending on the value of b and the error allocated to feasibility checks or comparison, the estimated PCS may behave differently. When $\theta^* = d + 1$, all systems are infeasible, which means that b remains 0. For simplicity, we fixed $b = 25$ when $\theta^* \leq d$ for the remainder of this section. Note that the estimated PCS values do not differ much for different variance configurations, thus, we also fix the L/L variance configuration in the rest of this section.

We then consider a case when two constraints are present. Each constraint contains three thresholds $\{0, 2\epsilon_\ell, 4\epsilon_\ell\}$ for $\ell = 1, 2$. Figure 4 shows the estimated PCS of the proposed procedures $\mathcal{ZAK}^{\mathcal{R}}$, $\mathcal{ZAK}$, and $\mathcal{ZAK}$+ with respect to all possible values of θ^* under our three preference orders. Thus, $d = 9$ for the ranked constraints and the total violation with ranked constraints formulations and $d = 3$ for the equally important constraints formulation.

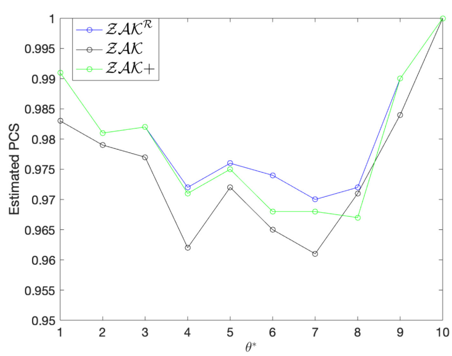
Figure 4 indicates that all three proposed procedures are statistically valid under our three preference orders. Note that the PCS is quite flat for all θ^* under the equally important constraints formulation. As the equally important constraints formulation relaxes all constraints by one threshold (if the constraint has at least one "looser" threshold) every time when one considers a less preferred threshold vector, declaring systems in S_u is easier than in the other two preference orders. Therefore, the estimated PCS for different θ^* under equally important constraints is relatively high in general. For the ranked constraints and the total violation with ranked constraints formulations, due to a similar reason as in the single constraint case, $\theta^* = 1$ and $\theta^* = d + 1$ achieve higher estimated PCS compared with the other θ^*. One may notice that $\theta^* = d$ also achieves a relatively high estimated PCS under these two preference orders. This is due to the mean configuration of the secondary performances we use for the systems that are infeasible to $\mathbf{q}^{(d)}$. In the DM configuration, we allocate b systems in A_{θ^*} and $(k - b)$ systems to A_{θ^*+1}. When $\theta^* = d$, we set all secondary performance measures of the $(k - b)$ systems that are infeasible to $\mathbf{q}^{(d)}$ equal to $y_{i\ell} = 5\epsilon_\ell$ (see the discussion in Section 6.1), which makes the detection of infeasibility of those systems easy (as the systems are infeasible to both constraints).



(a) Estimated PCS for ranked constraints



(b) Estimated PCS for equally important constraints



(c) Estimated PCS for total violation with ranked constraints

Fig. 4. Estimated PCS when $s = 2$ under our three threshold formulations as a function of θ^*.

6.4 Efficiency

In this section, we address the efficiency of our proposed procedures compared with the alternative procedures Restart$^{\mathcal{HAK}}$ and Restart$^{\mathcal{HAK}+}$ under the DM, MIM, and MDM configurations.

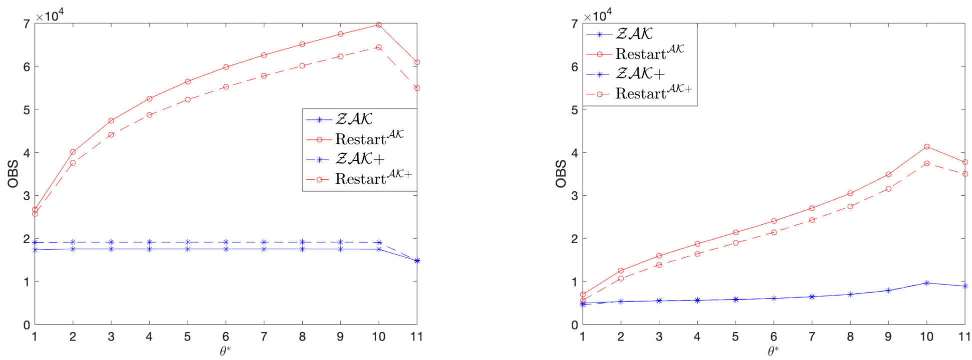
Table 2 shows OBS for the single constraint case under the DM configuration with four thresholds (the same experimental setting as in Table 1). We see that $\mathcal{ZAK}$ requires fewer OBS compared with $\mathcal{ZAK}^{\mathcal{R}}$ when $1 \leq \theta^* \leq 4$. This is expected as $\mathcal{ZAK}$ sets the implementation parameter for Phase II more efficiently than $\mathcal{ZAK}^{\mathcal{R}}$ (see the discussion in Section 3). When $\theta^* = 5$, $\mathcal{ZAK}^{\mathcal{R}}$ and $\mathcal{ZAK}$ have similar performance as all systems are infeasible to $\mathbf{q}^{(d)}$ and Phase II is not needed to achieve CS. Therefore, we omit the results for $\mathcal{ZAK}^{\mathcal{R}}$ from now on. We also see that the OBS increases with b for all three procedures. This is due to the fact that having more inferior systems that are feasible to $\mathbf{q}^{(\theta^*)}$ requires more correct feasibility and comparison decisions to achieve the final CS (on top of the feasibility decisions). One may also notice that all three proposed procedures require much fewer observations when $\theta^* = 5$ compared with other values of θ^*. This is because all systems are infeasible when $\theta^* = 5$ and thus do not require observations for Phase II to achieve correct selection. In terms of the comparison between $\mathcal{ZAK}$ and $\mathcal{ZAK}+$, we see that $\mathcal{ZAK}$ is more efficient than $\mathcal{ZAK}+$ in general under the L/L and H/L variance configurations while $\mathcal{ZAK}+$ is more efficient in general under the L/H variance configuration. This is because $\mathcal{ZAK}+$ performs the feasibility checks and comparison simultaneously. Hence inferior feasible

Table 2. Average Number of Observations of $\mathcal{ZAK}^{\mathcal{R}}$, $\mathcal{ZAK}$, and $\mathcal{ZAK}+$ for $k = 100$ Systems and $s = 1$ Constraint with four Thresholds under the DM Configuration

		$\mathcal{ZAK}^{\mathcal{R}}$			$\mathcal{ZAK}$			$\mathcal{ZAK}+$		
	θ^*	$b = 25$	$b = 50$	$b = 75$	$b = 25$	$b = 50$	$b = 75$	$b = 25$	$b = 50$	$b = 75$
L/L	1	22,659	29,344	35,537	17,350	20,628	24,208	19,037	22,218	24,885
	2	23,261	30,454	37,087	17,559	20,906	24,555	19,112	22,348	25,231
	3	23,241	30,416	37,008	17,531	20,891	24,580	19,119	22,350	25,231
	4	23,225	30,396	37,055	17,506	20,876	24,543	19,077	22,377	25,238
	5	8,904	8,904	8,904	8,924	8,924	8,924	8,893	8,893	8,893
L/H	1	81,402	87,996	94,254	73,911	73,924	74,139	65,610	59,200	52,238
	2	84,708	94,334	103,777	77,111	80,405	83,861	71,847	73,425	74,957
	3	84,711	94,421	103,925	77,160	80,383	83,867	71,764	73,383	75,001
	4	84,539	94,381	103,789	76,941	80,383	83,846	71,692	73,345	74,945
	5	44,119	44,119	44,119	44,215	44,215	44,215	44,065	44,065	44,065
H/L	1	53,562	86,764	117,509	39,708	69,173	100,260	50,006	79,456	106,285
	2	54,008	86,959	117,396	39,392	68,505	98,487	49,681	78,348	104,534
	3	53,975	87,151	117,446	39,476	68,365	98,184	49,537	78,446	104,392
	4	53,957	87,024	117,576	39,440	68,321	98,170	49,672	78,297	104,480
	5	8,904	8,904	8,904	8,924	8,924	8,924	8,893	8,893	8,893

systems with respect to $\mathbf{q}^{(\theta^*)}$ can be eliminated before knowing their feasibility with respect to $\mathbf{q}^{(\theta^*)}$, and this benefit is more obvious when the comparison is easier than the feasibility checks (i.e., L/H variance configuration). Also, we observe that the L/L variance configuration requires the smallest number of OBS. This is expected because lower variance results in an easier problem. However, H/L requires fewer OBS compared with L/H when b is relatively small (e.g., $b = 25$) whereas L/H is better when b is relatively large (e.g., $b = 75$). This is reasonable because the b inferior but feasible systems are often eliminated by comparison. Hence, the H/L variance configuration performs better when b is small. For simplicity, we fixed $b = 25$ and the L/L variance configuration in the rest of this section.

We then consider the single constraint case with ten thresholds under the L/L variance configuration. Figure 5 shows the results for OBS of the proposed procedures $\mathcal{ZAK}$ and $\mathcal{ZAK}+$ and their competing procedures Restart$^{\mathcal{AK}}$ and Restart$^{\mathcal{AK}+}$ under the DM and MIM configuration (the corresponding results for the MDM configuration are provided in Figure A.3). We see that $\mathcal{ZAK}$ and $\mathcal{ZAK}+$ outperform Restart$^{\mathcal{AK}}$ and Restart$^{\mathcal{AK}+}$, respectively. This is expected as Restart$^{\mathcal{AK}}$ and Restart$^{\mathcal{AK}+}$ allocate the nominal error for the ten thresholds and thus the resulting continuation regions used for feasibility check and for comparison are larger than those of $\mathcal{ZAK}$ and $\mathcal{ZAK}+$. We also see that the required observations increase dramatically for Restart$^{\mathcal{AK}}$ and Restart$^{\mathcal{AK}+}$ when θ^* increases, while the required observations for $\mathcal{ZAK}$ and $\mathcal{ZAK}+$ remain steady for all possible θ^*. This is because Restart$^{\mathcal{AK}}$ and Restart$^{\mathcal{AK}+}$ need to implement $\mathcal{AK}$ or $\mathcal{AK}+$ multiple times when θ^* gets larger and thus become very conservative, while $\mathcal{ZAK}$ and $\mathcal{ZAK}+$ are designed for one critical threshold per constraint regardless of the number of threshold values on that constraint. Note that $\mathcal{ZAK}$ and $\mathcal{ZAK}+$ require much fewer OBS when $\theta^* = 11$ compared with other values of θ^* (except for $\mathcal{ZAK}+$ under the MDM configuration). This is due to a similar reason as in the four thresholds case as all systems are eliminated by their infeasibility when $\theta^* = 11$ and thus we do not need to wait for comparison among feasible systems to be completed. (The different behavior of $\mathcal{ZAK}+$ under the MDM configuration is because under MDM the system with the highest mean falls in the most preferred region, and hence when $\theta^* \leq 10$, the



(a) OBS under DM

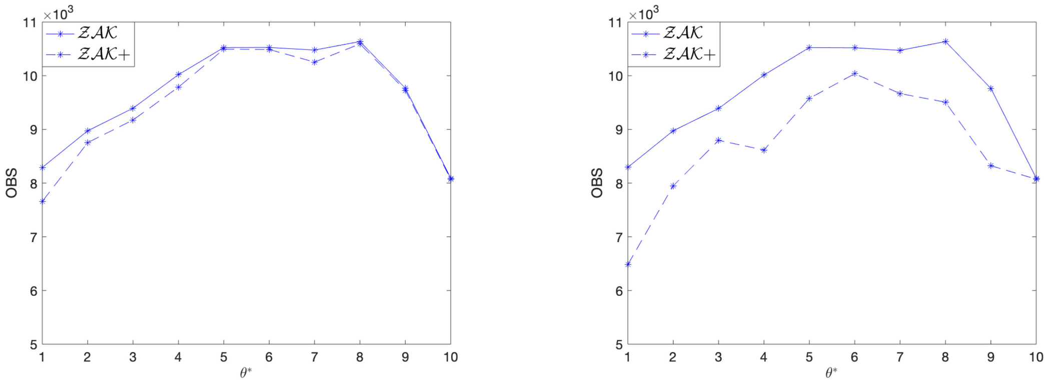
(b) OBS under MIM

Fig. 5. Average number of observations of $\mathcal{ZAK}$, Restart$^{\mathcal{AK}}$, $\mathcal{ZAK}$+ and Restart$^{\mathcal{AK}+}$ as functions of θ^* for $k = 100$ systems and $s = 1$ constraint with ten thresholds under the DM and the MIM configurations.

infeasible systems can be eliminated by both feasibility check and comparison while the infeasible systems under MIM can only be eliminated by the feasibility check.) We see that Restart$^{\mathcal{AK}}$ and Restart$^{\mathcal{AK}+}$ also show a sharp decrease in OBS when $\theta^* = 11$ (except for Restart$^{\mathcal{AK}+}$ under the MDM configuration), whereas OBS keeps increasing from $\theta^* = 1$ to 10. This is due to similar reasons as for $\mathcal{ZAK}$ and $\mathcal{ZAK}$+. However, as Restart$^{\mathcal{AK}}$ and Restart$^{\mathcal{AK}+}$ perform $\mathcal{AK}$ and $\mathcal{AK}$+ eleven times until its termination, the OBS is still relatively high when $\theta^* = 11$. As the performance of $\mathcal{ZAK}$ and $\mathcal{ZAK}$+ is expected to be significantly better than Restart$^{\mathcal{AK}}$ and Restart$^{\mathcal{AK}+}$, we omit the results for Restart$^{\mathcal{AK}}$ and Restart$^{\mathcal{AK}+}$ (and Restart$^{\mathcal{HAK}}$ and Restart$^{\mathcal{HAK}+}$ when multiple constraints are considered) and focus on comparing the performance of $\mathcal{ZAK}$ and $\mathcal{ZAK}$+ in the remainder of this section. Our results comparing all four procedures in the multiple constraints case are included in Appendix I. We see that $\mathcal{ZAK}$+ performs better or similar to $\mathcal{ZAK}$ and Restart$^{\mathcal{HAK}}$ performs better than Restart$^{\mathcal{HAK}+}$ under all cases we consider.

We now consider the two constraints case where each constraint contains three thresholds under the ranked constraints formulation and the MIM and MDM configurations (same experimental setting as when $s = 2$ under the ranked constraints formulation in Section 6.3 except for the mean configuration). Figure 6 shows the results of OBS for procedures $\mathcal{ZAK}$ and $\mathcal{ZAK}$+. We see that $\mathcal{ZAK}$+ performs significantly better than $\mathcal{ZAK}$ under the MDM configuration, while their performance is similar under the MIM configuration. This is because under the MDM configuration, the best system $[b]$ is feasible to the most preferred threshold vector $\mathbf{q}^{(1)}$. As $\mathcal{ZAK}$+ does not require both the comparison and feasibility decisions to be concluded to eliminate inferior systems or infeasible systems with respect to $\mathbf{q}^{(\theta^*)}$ (while $\mathcal{ZAK}$ needs to complete the feasibility check phase to eliminate infeasible systems with respect to $\mathbf{q}^{(\theta^*)}$), when the best system $[b]$ is feasible to $\mathbf{q}^{(1)}$, it can eliminate inferior systems once their feasibility is known to be no better than that of $[b]$ (this does not require concluding feasibility decisions for all the possible threshold vectors). On the other hand, the MIM configuration sets the infeasible systems with respect to $\mathbf{q}^{(\theta^*)}$ as superior systems compared with $[b]$, and hence those systems can only be eliminated once we make sure that they are not feasible to an improved threshold vector.

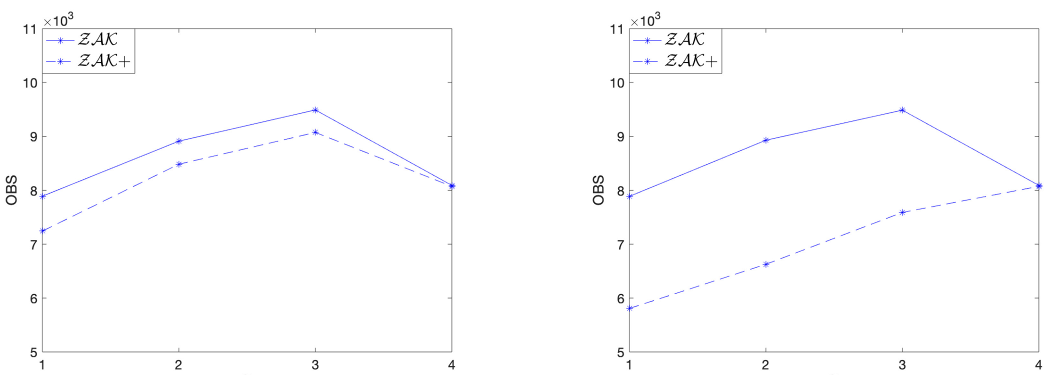
Figure 7 also shows the experimental results for two constraints with three thresholds on each constraint for the equally important constraints formulation and the MIM and the MDM configurations (same setting as in Figure 6 except for the preference order). The result shows a similar pattern as under the ranked constraints formulation. The dominance of $\mathcal{ZAK}$+ is more obvious



(a) OBS under MIM

(b) OBS under MDM

Fig. 6. Average number of observations of $\mathcal{ZAK}$ and $\mathcal{ZAK}$+ as functions of θ^* for $k = 100$ systems and $s = 2$ constraints under the MIM and MDM configurations for the ranked constraints formulation.



(a) OBS under MIM

(b) OBS under MDM

Fig. 7. Average number of observations of $\mathcal{ZAK}$ and $\mathcal{ZAK}$+ as functions of θ^* for $k = 100$ systems and $s = 2$ constraints under the MIM and MDM configurations for the equally important constraints formulation.

under the MDM configuration than under the MIM configuration. As the results for the total violation with ranked constraints formulation also show a similar pattern, we omit them here for the sake of space and include them in Appendix I.

As the MIM and MDM configurations aim to show the performance of the proposed procedures in realistic settings, we focus on the comparison between $\mathcal{ZAK}$ and $\mathcal{ZAK}$+ under those two configurations. Based on the results shown in this section and Appendix I, we see that $\mathcal{ZAK}$+ shows a significant improvement over $\mathcal{ZAK}$ under the MDM configuration while also outperforming $\mathcal{ZAK}$ in most cases under the MIM configuration. Therefore, since the decision maker usually does not have much information about the means of the systems in practice, we recommend $\mathcal{ZAK}$+ as it provides the best overall performance.

7 INVENTORY POLICY EXAMPLE

In this section, we study the implementation and performance of $\mathcal{ZAK}$ and $\mathcal{ZAK}$+, as well as their competing procedures Restart$^{\mathcal{HAK}}$ and Restart$^{\mathcal{HAK}+}$, on an (s, S) inventory policy example based on a similar setting as in [14]. Note that this example is similar to the problem we discussed in Sections 1 and 2 but with additional thresholds.

A decision maker controls inventory using an (s, S) policy, and the costs are given as (i) ordering cost at 3 per item; (ii) fixed ordering cost at 32 per order; (iii) holding cost at 1 per item per review period; and (iv) penalty cost at 5 per item of unsatisfied demand. Systems in consideration are given as

$$\Gamma = \{(s, S) \mid s = 20 + 2m', S = 40 + 10n', \text{ where } m' = 0, 1, 2, \ldots, 10, \text{ and } n' = 0, 1, 2, \ldots, 6\},$$

which contains 77 systems in total. Demand during each review period is assumed independent for different review periods and follows a Poisson distribution with mean 25. The decision maker particularly cares about three performance measures: (1) the fill rate per review period, which is the percentage of orders that can be fulfilled without running out of inventory, (2) the failure probability, which is the probability that a shortage occurs between two successive review periods; and (3) the expected cost per review period, which is the average total cost for each review period. In practice, performance measures often conflict with each other, such as fill rate and cost. Therefore, it is rare for one system to perform the best with respect to all performance measures under consideration. Instead, the decision maker may prioritize one performance measure and set it as the primary performance measure while treating the others as constraints to ensure that their values are within acceptable ranges. If she chooses her most preferred (fixed) threshold on each constraint, this may lead to no feasible systems when the chosen thresholds are too strict. Alternatively, she could consider multiple thresholds on each constraint so that she can identify both her most preferred combination of threshold values possible (that lead to some feasible systems) as well as the best system among the feasible systems with respect to the most preferred thresholds possible. In other words, she can formulate subjective stochastic constraints on the secondary performance measures. With a set of thresholds specified for each constraint, the decision maker has the flexibility to restrict the secondary performance measures to tighter ranges than would be possible using her weakest acceptable threshold values on all constraints. In this inventory problem, we assume that the decision maker treats the fill rate measure as the most important measure, meaning that she aims to maximize the fill rate as much as she can subject to maintaining reasonable values of the measures of failure probability and expected cost per review period. Thus, the problem can be formulated as described in Section 2 where the primary performance measure is set as the fill rate and the two secondary performance measures are set as the failure probability ($\ell = 1$) and the expected cost per review period ($\ell = 2$).

In our experiments, we set the run-length for each replication to 100 review periods and obtain one observation for the fill rate, failure probability, and average cost per review period from each replication, respectively, to estimate the primary and secondary performance measures. We also estimate the correlation between the primary performance measure and each constraint, as well as the correlation between the two constraints, based on 1,000 observations. The range of the correlation between the primary performance measure and the failure probability constraint (expected cost constraint) ranges from -1 to -0.7781 (from -0.7355 to 0.0731). The correlation between the two constraints ranges from -0.2334 to 0.5489.

We now address the selection of the implementation parameters. The choice of $\alpha \in (0, 1)$ depends on the desired nominal error and typically satisfies $\alpha \in \{0.01, 0.05, 0.1\}$. The value of n_0 determines the number of initial observations, which are used for the variance estimation, and one should choose it neither too large (which can result in collecting unnecessary observations) nor too small (which can result in a poor estimation of the variances). In the experiments, we set $\alpha = 0.05$ and $n_0 = 20$, which are common choices in the literature. Before implementing the proposed procedures $\mathcal{ZAK}$ and $\mathcal{ZAK}+$ to identify the most preferred system possible, the decision maker needs to specify additional implementation parameters, including the thresholds on the constraints, the associated tolerance levels of each constraint, the IZ parameter, and the choice of the preference order for the constraints. We now discuss how to choose those parameters separately.

Thresholds on the constraints. The choice of the thresholds on a particular constraint depends on the nature of the constraint and the expectation of the decision maker. One recommendation is to first choose (rough) lower and upper bounds for the constraint, which should be possible based on past or similar experiences or industry data, and then consider a relaxation level for each constraint and choose the thresholds based on an increment of the relaxation level. For example, the thresholds on the failure probability constraint should be limited between 0 and 1. Moreover, since identifying systems that are feasible to large thresholds, say 0.8 or 0.9, does not yield practical decisions (as decision maker is likely to expect the failure probability be lower than a relatively small value), it is natural to consider small thresholds. If the decision maker is willing to relax the constraints by roughly 5% every time, she may choose thresholds 0.01, 0.05, 0.1, 0.15, and so on. For the expected cost, thresholds from a wider (positive) range can be chosen. If she wants to relax the expected cost constraint by \$5 every time, she may consider thresholds such as \$100, \$105, \$110, and so on. As discussed in Sections 3 and 4, nominal errors such as α'_f, β_f, and β_c depend on $\min\{s, d\}$. In practice, s is typically smaller than d, which further implies that the nominal errors likely only depend on the number of constraints (rather than the number of threshold vectors). Thus, $\mathcal{ZAK}$ and $\mathcal{ZAK}$+ scale well with respect to the number of thresholds on the constraints and the decision maker can simply choose all possible thresholds that she is willing to consider if she does not have a clear idea on the performance of the constraints. In this case study, we particularly consider three thresholds on the first constraint ($q_1 \in \{0.01, 0.05, 0.1\}$) and eight thresholds on the second constraint ($q_2 \in \{100, 105, 110, 115, 120, 125, 130, 135\}$).

IZ parameter and Tolerance levels. The IZ parameter and the tolerance levels measure the absolute amount that the decision maker is indifferent to for the primary and secondary performance measures, respectively. Given that the primary performance measure, that is, the fill rate, has range in between 0 and 1, we assume that the decision maker has a small indifference level, say 0.1% of the range, and set $\delta = 0.001$. While the choices of the tolerance levels on the secondary performance measures depend on what differences the decision maker considers significant, the choices of the thresholds on the constraints also provide additional guidance. In general, when the thresholds are dense for a particular constraint, a smaller tolerance level is expected. This is because a dense set of thresholds usually indicates that the decision maker is more sensitive to the constraints and is less likely to accept a large amount off each threshold on that constraint. For the opposite reason, the decision maker may consider a larger tolerance level when the thresholds are sparse. In our experiments, with a relatively dense set of thresholds $\{0.01, 0.05, 0.1\}$ on the failure probability constraint, we set the tolerance level as $\epsilon_1 = 0.001$; and with a relatively sparse threshold set $\{100, 105, 110, \ldots, 130, 135\}$ on the expected cost constraint, we set $\epsilon_2 = 0.5$.

Preference order. With the chosen threshold constants on each constraint, the decision maker's preferred threshold combinations across all constraints depend on how much she prioritizes each constraint. Section 5 discusses three useful formulations. When the decision maker does not have a clear idea on how she prioritizes the constraints, choosing the equally important constraints formulation is natural and recommended. On the other hand, the decision maker should choose the ranked constraints formulation if she has a clear preference on the importance among the constraints. Finally, if the decision maker thinks all the constraints are important but she also

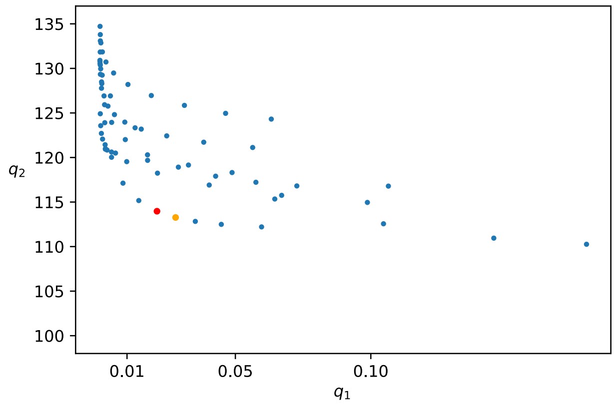


Fig. 8. The values of the secondary performance measures for all the systems. The best system $(s, S) = (28, 60)$ (highlighted in red) has a fill rate of 0.9981, a failure probability of 0.0211, and an expected cost per review period of 113.9701. The good system $(s, S) = (26, 60)$ (highlighted in yellow) has a fill rate of 0.9972, a failure probability of 0.0279, and an expected cost per review period of 113.2690.

has a preference among them, then she may want to relax the constraints one at a time following her preference. This can be modeled by choosing the total violation with ranked constraints formulation. In our experiments, we test all three preference order proposed in Section 5. For the ranked constraints and the total violation with ranked constraints formulations, consistent with the discussion in Section 1, we prioritize the first constraint over the second constraint (relax the second constraint first) and have 24 feasible regions (i.e., $d = 24$). For the equally important constraints formulation, we have 8 feasible regions (i.e., $d = 8$).

Figure 8 shows the analytical values of the secondary performance measures of all the systems based on a Markov chain model. The value of θ^* is 11, 4, and 11 for ranked constraints, equally important constraints, and total violation with ranked constraints, respectively, which corresponds to the threshold vector $(q_1, q_2) = (0.05, 115)$ for ranked constraints and total violation with ranked constraints and $(q_1, q_2) = (0.1, 115)$ for equally important constraints. The identity of the best system, $(s, S) = (28, 60)$, remains the same for all three preference orders considered. Note that the chosen IZ parameter $\delta = 0.001$ does not satisfy Assumption 2 (since the system $(s, S) = (26, 60)$ is feasible with respect to threshold vector $\mathbf{q}^{(\theta^*)}$ and has primary performance in a δ range of the best system under all three preference orders). Nevertheless, we do not see the statistical validity violated in the experiments.

We expect the comparison phase to be easier than the feasibility check phase because the variance of the difference in the fill rate is very small compared to the variance of cost per review period. Thus we do not employ CRN. The experimental results are based on 10,000 replications and are shown in Table 3. We see that under the ranked constraints and equally important constraints formulations, $\mathcal{ZAK}$ spends around 44% and 48% of the observations compared to those of Restart$^{\mathcal{HAK}}$ whereas $\mathcal{ZAK}$+ spends around 34% and 33% compared with Restart$^{\mathcal{HAK}+}$, respectively. When it comes to the total violation with ranked constraints formulation, the savings is more pronounced as $\mathcal{ZAK}$ and $\mathcal{ZAK}$+ spend around 29% and 23% of the observations compared to those of Restart$^{\mathcal{HAK}}$ and Restart$^{\mathcal{HAK}+}$, respectively. Both proposed procedures perform much better than their alternative procedures, while also remaining statistically valid (even though Assumption 2 is not satisfied). In terms of the comparison between $\mathcal{ZAK}$ and $\mathcal{ZAK}$+, we observe that $\mathcal{ZAK}$+ performs better under all three threshold formulations, while the advantage of

Table 3. Average Number of Observations and Estimated PCS (Reported in Parentheses) of $\mathcal{ZAK}$, Restart$^{\mathcal{HAK}}$, $\mathcal{ZAK}$+ and Restart$^{\mathcal{HAK}+}$ for the Inventory Policy Example

Preference Order	$\mathcal{ZAK}$	Restart$^{\mathcal{HAK}}$	$\mathcal{ZAK}$+	Restart$^{\mathcal{HAK}+}$
Ranked constraints	9,547 (1.000)	21,769 (1.000)	6,066 (1.000)	17,799 (1.000)
Equally important constraints	7,819 (1.000)	16,240 (1.000)	2,490 (1.000)	7,475 (1.000)
Total violation with ranked constraints	8,778 (1.000)	30,158 (1.000)	6,034 (1.000)	26,023 (1.000)

$\mathcal{ZAK}$+ is more obvious under the equally important constraints formulation. We also see that the comparison between Restart$^{\mathcal{HAK}}$ and Restart$^{\mathcal{HAK}}$ shows a similar pattern as Restart$^{\mathcal{HAK}+}$ performs better than Restart$^{\mathcal{HAK}}$ under all three threshold formulations and the equally important constraints formulation makes the dominance more clear. Note that this agrees with the results in Section 6.4.

8 CONCLUSION

We consider the selection-of-the-best problem when subjective stochastic constraints are present. When a decision maker has flexibility with thresholds, this allows her to start with tight threshold values for each constraint and then relax the thresholds until feasible systems are found and compared. We discuss how to combine thresholds on constraints into threshold vectors based on how a decision maker prioritizes each constraint. We propose two procedures that select the best system with respect to a primary performance measure while also satisfying constraints on secondary performance measures with respect to the most preferred thresholds possible. Our procedures differ in that one runs feasibility check and comparison sequentially while the other runs them simultaneously. We discuss how to set the implementation parameters for our procedures and prove their statistical validity. We also demonstrate through experiments that the required number of observations remains steady when the number of threshold vectors grows and address the impact of applying CRN when performing our procedures. Finally, our experimental results show that the proposed procedures perform well in reducing the average number of needed observations as compared with procedures that repeatedly solve the problem for each threshold vector. Overall, we recommend our simultaneously-running procedure as it provides the best performance in general.

APPENDICES

In Appendix A, we provide the detailed algorithm statement of Procedure $\mathcal{ZAK}^{\mathcal{R}}$ from Section 3 along with the discussion on its statistical validity. Appendix B describes how we set implementation parameters for the proposed sequentially-running procedures. We provide the proof of the statistical validity of Procedure $\mathcal{ZAK}$+ in Appendix C and include how to set its implementation parameters in Appendix D. Appendix E includes the algorithms that we use to generate the three example preference orders discussed in Section 5. In Appendices F and G, we describe procedures Restart$^{\mathcal{HAK}}$ and Restart$^{\mathcal{HAK}+}$ and discuss their statistical validity, respectively. Appendices H and I provide additional experimental results that are used to set the implementation parameters of our proposed procedures and to demonstrate the efficiency of our proposed procedures, respectively. Finally, Appendix J provides experimental results and a discussion on the impact of using CRNs.

A PROCEDURE $\mathcal{ZAK}^{\mathcal{R}}$

In this section, we provide the detailed description of the $\mathcal{ZAK}^{\mathcal{R}}$ procedure and prove its statistical validity.

Algorithm A.1 gives the full description of $\mathcal{ZAK}^{\mathcal{R}}$. Note that it is possible to use r rather than r_i in Phase I in $\mathcal{ZAK}^{\mathcal{R}}$. To prove the statistical validity of $\mathcal{ZAK}^{\mathcal{R}}$, we start with the following lemma.

Lemma A.1. *Under Assumption 1, for system i and constraint ℓ with specific threshold value $q_{\ell,m}$, the* **[Feasibility Check]** *steps in $\mathcal{ZAK}^{\mathcal{R}}$ that run to completion ensure* $\Pr(\mathrm{CD}_{i\ell}(q_{\ell,m})) \geq 1 - \alpha'_f$.

Proof. When system i and constraint ℓ with specific threshold $q_{\ell,m}$ are considered separately, the **[Feasibility Check]** steps in $\mathcal{ZAK}^{\mathcal{R}}$ either conclude a feasibility decision or eliminate threshold $q_{\ell,m}$ for further consideration (when system i is declared feasible with respect to a threshold vector and all preferred threshold vectors do not involve threshold value $q_{\ell,m}$ on constraint ℓ). We see that when a feasibility decision is concluded, the **[Feasibility Check]** steps in $\mathcal{ZAK}^{\mathcal{R}}$ are essentially the same as for the statistically-valid feasibility check procedure $\mathcal{F}$ in [1] for a single system and a single constraint with one threshold value with confidence level $1 - \alpha'_f$. The result now follows from the special case of Theorem 1 in [1] with $k = 1$. □

We use the same notation for $i \in \Gamma$ as in Section 4 as follows.

$$\mathcal{A}_1^*(i) = \left\{\text{system } i \text{ is declared infeasible to } \mathbf{q}^{(1)}, \ldots, \mathbf{q}^{(\min\{\theta^*,d\})}\right\};$$

$$\mathcal{A}_2^*(i) = \left\{\text{system } i \text{ is declared infeasible to } \mathbf{q}^{(1)}, \ldots, \mathbf{q}^{(\theta^*-1)} \text{ if } 1 < \theta^* \leq d\right\};$$

$$\mathcal{B}_1^* = \left\{\text{system } [b] \text{ is declared feasible to } \mathbf{q}^{(\theta^*)} \text{ if } \theta^* \leq d\right\}.$$

Lemma A.2. *Under Assumption 1, for a particular system i, the* **[Feasibility Check]** *steps in $\mathcal{ZAK}^{\mathcal{R}}$ ensure*

$$\Pr\left(\mathcal{A}_1^*(i)\right) \geq 1 - \min\{s,d\}\alpha'_f, \text{ if } i \in S_u;$$
$$\Pr\left(\mathcal{A}_2^*(i)\right) \geq 1 - \min\{s,d-1\}\alpha'_f, \text{ if } i \in S_d \cup S_{a'} \text{ and } 1 < \theta^* \leq d;$$
$$\Pr\left(\mathcal{B}_1^*\right) \geq 1 - s\alpha'_f, \text{ if } \theta^* \leq d.$$

Proof. First, consider $i \in S_u$. We discuss the following two cases depending on whether $\theta^* \leq d$ or $\theta^* = d + 1$.

When $\theta^* \leq d$, system i must be unacceptable to $\mathbf{q}^{(1)}, \ldots, \mathbf{q}^{(\theta^*)}$ because it is unacceptable to $\mathbf{q}^{(\theta^*)}$, not in S_a, and there are no desirable systems with respect to $\mathbf{q}^{(1)}, \ldots, \mathbf{q}^{(\theta^*-1)}$. As system i is unacceptable with respect to $\mathbf{q}^{(1)}, \ldots, \mathbf{q}^{(\theta^*)}$, then for each $\kappa = 1, \ldots, \theta^*$, there exist at least one constraint ℓ_κ such that $y_{i\ell_\kappa} \geq q_{\ell_\kappa}^{(\kappa)} + \epsilon_{\ell_\kappa}$. Then we have

$$\Pr\left(\mathcal{A}_1^*(i)\right) \geq \Pr\left(\cap_{\kappa=1}^{\theta^*}\mathrm{CD}_{i\ell_\kappa}(q_{\ell_\kappa}^{(\kappa)})\right) \geq 1 - \sum_{\kappa=1}^{\theta^*}\Pr\left(\mathrm{ICD}_{i\ell_\kappa}(q_{\ell_\kappa}^{(\kappa)})\right) \geq 1 - d\alpha'_f, \tag{2}$$

where we use $\mathrm{ICD}_{i\ell}(q_{\ell,m})$ to denote the event of incorrect decision of system i with respect to constraint ℓ and threshold $q_{\ell,m}$. The first inequality holds because declaring system i infeasible to constraint ℓ_κ is sufficient to declare system i infeasible to threshold vector $\mathbf{q}^{(\kappa)}$ and it is not possible to declare a system feasible with respect to a threshold vector without completing the comparison with all thresholds in that vector. The second inequality holds due to the Bonferroni inequality, and the last inequality holds due to Lemma A.1 and the fact of $\theta^* \leq d$.

ALGORITHM A.1: Procedure $\mathcal{ZAK}^{\mathcal{R}}$.

[**Setup:**] Select the overall nominal confidence level $1-\alpha$ and choose $0 < \alpha_f, \alpha_c < 1$ such that $(1-\alpha_f)(1-\alpha_c) = 1-\alpha$. Choose tolerance levels $\epsilon_1, \ldots, \epsilon_s$, indifference-zone parameter δ, threshold vectors $\{\mathbf{q}^{(1)}, \mathbf{q}^{(2)}, \ldots, \mathbf{q}^{(d)}\}$, and associated index vectors $\{\boldsymbol{I}^{(1)}, \boldsymbol{I}^{(2)}, \ldots, \boldsymbol{I}^{(d)}\}$. Set $M = \Gamma$ and $Z_{i,\ell,m} = 2$ for all $i \in M$, $\ell = 1, \ldots, s$, and $m = 1, \ldots, d_\ell$. Set $F = \emptyset$ and $\theta = d$. Set η_f such that $g(\eta_f) = \alpha'_f$, where $0 < \alpha'_f < 1/s$ is set as a solution to

$$\left(1 - \min\{s, d\}\alpha'_f\right)^{k-1} \times (1 - s\alpha'_f) = 1 - \alpha_f, \text{ if systems are simulated independently;}$$

and set as

$$\alpha'_f = \alpha_f / [(k-1)\min\{s, d\} + s], \text{ if systems are simulated under CRN.}$$

Add any constraint ℓ, where $\ell = 1, \ldots, s$, with increasing preference to set IP.
[**Initialization for Phase I:**]
for each system $i \in M$ **do**
— Obtain n_0 observations $Y_{i\ell 1}, Y_{i\ell 2}, \ldots, Y_{i\ell n_0}$ for $\ell = 1, 2, \ldots, s$.
— Compute $\bar{Y}_{i\ell}(n_0)$ and $S^2_{Y_{i\ell}}(n_0)$.
— Set $r_i = n_0$, $\mathrm{ON}_i = \{1, 2, \ldots, s\}$, and $\mathrm{ON}_{i\ell} = \{1, \ldots, d_\ell\}$ for $\ell = 1, 2, \ldots, s$.
end for
[**Feasibility Check:**]
for each system $i \in M$ **do**
  for $\ell \in \mathrm{ON}_i$ **do**
    for $m \in \mathrm{ON}_{i\ell}$ **do**,
      If $\bar{Y}_{i\ell}(r_i) + R(r_i; \epsilon_\ell, \eta_f, S^2_{Y_{i\ell}}(n_0))/r_i \le q_{\ell,m}$, set $Z_{i,\ell,m} = 1$ and $\mathrm{ON}_{i\ell} = \mathrm{ON}_{i\ell} \setminus \{m\}$.
      If $\bar{Y}_{i\ell}(r_i) - R(r_i; \epsilon_\ell, \eta_f, S^2_{Y_{i\ell}}(n_0))/r_i \ge q_{\ell,m}$, set $Z_{i,\ell,m} = 0$ and $\mathrm{ON}_{i\ell} = \mathrm{ON}_{i\ell} \setminus \{m\}$.
    end for
    If $\mathrm{ON}_{i\ell} = \emptyset$, set $\mathrm{ON}_i = \mathrm{ON}_i \setminus \{\ell\}$.
  end for
  If $\exists$ minimum $\kappa \le \theta$ s.t. $\prod_{\ell=1}^{s} Z_{i,\ell,I_\ell^{(\kappa)}} = 1$, and either $\kappa < \theta$ or $i \notin F$, then
— If $\kappa < \theta$, then set $F = \emptyset$, $\theta = \kappa$, and for all $j \in M$ delete $q_{\ell,m}$ from $\mathrm{ON}_{j\ell}$ if $\ell \in \mathrm{IP}$ and $m > I_\ell^{(\theta)}$ (if $\ell \notin \mathrm{IP}$, then $q_{\ell,m}$ can be removed from $\mathrm{ON}_{j\ell}$ if $I_\ell^{(\theta')} \ne m$ for all $\theta' \le \kappa$), and set $\mathrm{ON}_j = \mathrm{ON}_j \setminus \{\ell\}$ if $\mathrm{ON}_{j\ell} = \emptyset$.
— Add system i to F.
  If $\prod_{\ell=1}^{s} Z_{i,\ell,I_\ell^{(\theta)}} = 0$ or 1 and either $\theta = 1$ or $\prod_{\ell=1}^{s} Z_{i,\ell,I_\ell^{(\kappa)}} = 0$ for all $\kappa = 1, \ldots, \theta - 1$, then remove system i from M.
end for
[**Stopping Condition for Phase I:**]
If $M \ne \emptyset$, then for each system $i \in M$, set $r_i = r_i + 1$, take one additional observation $Y_{i\ell r_i}$, and update $\bar{Y}_{i\ell}(r_i)$ for $\ell \in \mathrm{ON}_i$, then go to [**Feasibility Check**]. Else, check the following conditions.
— If $|F| = 0$, stop and conclude no feasible systems;
— If $|F| = 1$, stop and return the system in F as the best; or
— If $|F| > 1$, go to [**Initialization for Phase II**].
[**Initialization for Phase II:**] Let η_c be a solution to $g(\eta_c) = \alpha'_c$, where

$$\alpha'_c = \begin{cases} 1 - (1-\alpha_c)^{1/(k-1)}, & \text{if systems are simulated independently;} \\ \alpha_c/(k-1), & \text{if systems are simulated under CRN.} \end{cases}$$

Let $M = F$ be the set of systems still in contention. For each system $i \in M$, perform an entirely new simulation and obtain $X_{i1}, \ldots, X_{in_0}$ independent of any $Y_{j\ell n}$ generated in Phase I. Compute $\bar{X}_i(n_0)$ and $S^2_{X_{ij}}(n_0)$ for $i, j \in M$ and $i \ne j$. Set $r = n_0$ and go to [**Comparison**].
[**Comparison:**] For $i, j \in M$ s.t. $i \ne j$ and

$$r\bar{X}_i(r) > r\bar{X}_j(r) + R(r; \delta, \eta_c, S^2_{X_{ij}}(n_0)),$$

eliminate j from M.
[**Stopping Condition for Phase II:**] If $|M| = 1$, then stop and select the system in M as the best. Otherwise, for each system $i \in M$, take one additional observation $X_{i,r+1}$ independent of any $Y_{j\ell n}$ generated in Phase I and compute $\bar{X}_i(r+1)$. Then, set $r = r + 1$ and go to [**Comparison**].

Observe that since there are only s constraints, the set $L = \{\ell_1, \ldots, \ell_{\theta^*}\}$ can have at most s distinct values. For $\ell \in L$, let $I_{i\ell}$ denote the largest threshold index on constraint ℓ that system i is unacceptable to, that is,

$$I_{i\ell} = \max_{1 \le m \le d_\ell} \left\{ m : y_{i\ell} \ge q_{\ell,m} + \epsilon_\ell \right\}.$$

Thus, we know that $q_{\ell,1} < q_{\ell,2} < \cdots < q_{\ell,I_{i\ell}} \le y_{i\ell} - \epsilon_\ell$ on constraint ℓ. Due to the discussion in [22], we know that $\mathrm{CD}_{i\ell}(q_{\ell,I_{i\ell}}) \subseteq \cdots \subseteq \mathrm{CD}_{i\ell}(q_{\ell,2}) \subseteq \mathrm{CD}_{i\ell}(q_{\ell,1})$. Then $\mathrm{CD}_{i\ell}(q_{\ell,I_{i\ell}}) \subseteq \mathrm{CD}_{i\ell}(q_\ell^{(\kappa)})$ for $\kappa = 1, \ldots, \theta^*$ with $\ell_\kappa = \ell$. Thus, we also have

$$\begin{aligned} \Pr\left(\mathcal{A}_1^*(i)\right) &\ge \Pr\left(\cap_{\kappa=1}^{\theta^*} \mathrm{CD}_{i\ell_\kappa}(q_{\ell_\kappa}^{(\kappa)})\right) \ge \Pr\left(\cap_{\ell \in L} \mathrm{CD}_{i\ell}(q_{\ell,I_{i\ell}})\right) \\ &\ge 1 - \sum_{\ell \in L} \Pr\left(\mathrm{ICD}_{i\ell}(q_{\ell,I_{i\ell}})\right) \ge 1 - |L|\alpha_f' \ge 1 - s\alpha_f', \end{aligned} \tag{3}$$

where the third inequality is due to the Bonferroni inequality and the forth inequality is due to Lemma A.1. By comparing Equations (2) and (3), we conclude that $\Pr\left(\mathcal{A}_1^*(i)\right) \ge 1 - \min\{s, d\}\alpha_f'$.

When $\theta^* = d + 1$, a similar argument yields

$$\Pr\left(\mathcal{A}_1^*(i)\right) \ge \Pr\left(\cap_{\kappa=1}^{d} \mathrm{CD}_{i\ell_\kappa}(q_{\ell_\kappa}^{(\kappa)})\right) \ge 1 - \sum_{\kappa=1}^{d} \Pr\left(\mathrm{ICD}_{i\ell_\kappa}(q_{\ell_\kappa}^{(\kappa)})\right) \ge 1 - d\alpha_f',$$

and, defining $L = \{\ell_1, \ldots, \ell_d\}$,

$$\begin{aligned} \Pr\left(\mathcal{A}_1^*(i)\right) &\ge \Pr\left(\cap_{\kappa=1}^{d} \mathrm{CD}_{i\ell_\kappa}(q_{\ell_\kappa}^{(\kappa)})\right) \ge \Pr\left(\cap_{\ell \in L} \mathrm{CD}_{i\ell}(q_{\ell,I_{i\ell}})\right) \\ &\ge 1 - \sum_{\ell \in L} \Pr\left(\mathrm{ICD}_{i\ell}(q_{\ell,I_{i\ell}})\right) \ge 1 - |L|\alpha_f' \ge 1 - s\alpha_f'. \end{aligned}$$

Therefore, $\Pr\left(\mathcal{A}_1^*(i)\right) \ge 1 - \min\{s, d\}\alpha_f'$.

Now, consider $i \in S_d \cup S_{a'}$ with $1 < \theta^* \le d$. As system i is not in S_a and there are no desirable systems with respect to $\mathbf{q}^{(1)}, \ldots, \mathbf{q}^{(\theta^*-1)}$, system i must be unacceptable with respect to $\mathbf{q}^1, \ldots, \mathbf{q}^{(\theta^*-1)}$. Then for each $\kappa = 1, \ldots, \theta^* - 1$, there exist at least one constraint ℓ_κ such that $y_{i\ell_\kappa} \ge q_{\ell_\kappa}^{(\kappa)} + \epsilon_{\ell_\kappa}$. Due to a similar argument as for $i \in S_u$, we have

$$\Pr\left(\mathcal{A}_2^*(i)\right) \ge \Pr\left(\cap_{\kappa=1}^{\theta^*-1} \mathrm{CD}_{i\ell_\kappa}(q_{\ell_\kappa}^{(\kappa)})\right) \ge 1 - \sum_{\kappa=1}^{\theta^*-1} \Pr\left(\mathrm{ICD}_{i\ell_\kappa}(q_{\ell_\kappa}^{(\kappa)})\right) \ge 1 - (d-1)\alpha_f'.$$

Based on a similar definition $L = \{\ell_1, \ldots, \ell_{\theta^*-1}\}$ and the discussion above, we have

$$\begin{aligned} \Pr\left(\mathcal{A}_2^*(i)\right) &\ge \Pr\left(\cap_{\kappa=1}^{\theta^*-1} \mathrm{CD}_{i\ell_\kappa}(q_{\ell_\kappa}^{(\kappa)})\right) \ge \Pr\left(\cap_{\ell \in L} \mathrm{CD}_{i\ell}(q_{\ell,I_{i\ell}})\right) \\ &\ge 1 - \sum_{\ell \in L} \Pr\left(\mathrm{ICD}_{i\ell}(q_{\ell,I_{i\ell}})\right) \ge 1 - |L|\alpha_f' \ge 1 - s\alpha_f'. \end{aligned}$$

Therefore, we have $\Pr\left(\mathcal{A}_2^*(i)\right) \ge 1 - \min\{s, d-1\}\alpha_f'$.

Finally, for [b], when $\theta^* \le d$, we have

$$\Pr\left(\mathcal{B}_1^*\right) = \Pr\left(\cap_{\ell=1}^{s} \mathrm{CD}_{i\ell}(q_\ell^{(\theta^*)})\right) \ge 1 - \sum_{\ell=1}^{s} \Pr\left(\mathrm{ICD}_{i\ell}(q_\ell^{(\theta^*)})\right) \ge 1 - s\alpha_f',$$

where the last inequality is due to Lemma A.1. □

For Lemma A.2, one may notice that $d > s$ holds in most cases, and therefore $\Pr\left(\mathcal{A}_1^*(1)\right) \geq 1 - s\alpha_f'$ and $\Pr\left(\mathcal{A}_2^*(1)\right) \geq 1 - s\alpha_f'$ hold in most cases. Note that when $d \geq s$ and the systems are simulated independently, the implementation parameter α_f' has a closed-form solution as

$$\alpha_f' = \frac{1}{s}\left[1 - (1-\alpha_f)^{1/k}\right].$$

When $d < s$, one may need to find α_f' by numerically solving $(1 - d\alpha_f')^{k-1} \times (1 - s\alpha_f') = 1 - \alpha_f$. As we always have $(1 - d \times 0)^{k-1} \times (1 - s \times 0) - (1 - \alpha_f) = \alpha_f > 0$ and $(1 - d \times \frac{1}{s})^{k-1} \times (1 - s \times \frac{1}{s}) - (1 - \alpha_f) = \alpha_f - 1 < 0$, there will always be a solution α_f' satisfying $0 < \alpha_f' < \frac{1}{s}$.

We then use CS_i to denote the correct selection between system $i \in S_{a'} \cup S_d$ and the best system $[b]$ and introduce the following lemma.

Lemma A.3. *Under Assumption 1, given i such that $x_i \leq x_{[b]} - \delta$, the* **[Comparison]** *steps for system i and $[b]$ in $\mathcal{ZAK}^{\mathcal{R}}$ that run to completion ensure*

$$\Pr(\mathrm{CS}_i) \geq 1 - \alpha_c'.$$

Proof. When only system i and $[b]$ are considered, the **[Comparison]** steps in $\mathcal{ZAK}^{\mathcal{R}}$ are the same as in the statistically-valid selection-of-the-best procedure provided in [12] when two systems are considered with confidence level $1 - \alpha_c'$. Therefore, the result follows from the special case of Theorem 1 of [12] with $k = 2$. □

We are now ready to give the main theorem about the statistical validity of $\mathcal{ZAK}^{\mathcal{R}}$ and provide the detailed proof of Theorem A.4.

Theorem A.4. *Under Assumptions 1 and 2, the $\mathcal{ZAK}^{\mathcal{R}}$ procedure guarantees*

$$\Pr\{\mathrm{CS}\} \geq 1 - \alpha.$$

Proof. We consider two cases, namely when $\theta^* \leq d$ and $\theta^* = d + 1$.
Case 1: $\theta^* \leq d$.
Note that any systems in $(S_{a'} \cup S_d)$ should not be declared feasible with respect to a more preferred threshold vector $\mathbf{q}^{(1)}, \ldots, \mathbf{q}^{(\theta^*-1)}$ as they could be selected as the best system otherwise. More specifically, we consider the following four events.

$$\mathcal{A}_1^* = \left\{\text{all systems in } S_u \text{ are eliminated by infeasibility} = \cap_{i \in S_u} \mathcal{A}_1^*(i)\right\};$$

$$\mathcal{A}_2^* = \left\{\text{all systems in } (S_{a'} \cup S_d) \text{ are declared infeasible to thresholds } \mathbf{q}^{(1)}, \ldots, \mathbf{q}^{(\theta^*-1)}\right\}$$
$$= \left\{\cap_{i \in S_{a'} \cup S_d} \mathcal{A}_2^*(i) \text{ when } \theta^* > 1\right\};$$

$$\mathcal{B}_2^* = \{\text{system } [b] \text{ would be selected as the best system among the systems in } S_{a'} \cup S_d\};$$

$$\mathcal{B}^* = \left\{\text{system } [b] \text{ is declared feasible with respect to } \mathbf{q}^{(\theta^*)} \text{ and is selected as the best system}\right.$$
$$\left.\text{among the surviving systems from Phase I}\right\}.$$

Notice that $\mathcal{B}_1^* \cap \mathcal{B}_2^* \subseteq \mathcal{B}^*$ and $\mathcal{A}_2^*$ is not defined when $\theta^* = 1$. This means

$$\Pr\{\mathrm{CS}\} \geq \begin{cases} \Pr(\mathcal{A}_1^* \cap \mathcal{B}^*), & \text{if } \theta^* = 1; \\ \Pr(\mathcal{A}_1^* \cap \mathcal{A}_2^* \cap \mathcal{B}^*), & \text{if } \theta^* > 1. \end{cases}$$

We see that $\Pr\{\mathrm{CS}\}$ achieves its lower bound when $\theta^* > 1$ (because the bounds on $\Pr(\mathcal{A}_1^*)$, $\Pr(\mathcal{B}_1^*)$, and $\Pr(\mathcal{B}_2^*)$ below do not depend on the value of θ^*), and thus we focus on the case when $\theta^* > 1$. We also see that $\mathcal{A}_1^*$, $\mathcal{A}_2^*$, and $\mathcal{B}_1^*$ are independent events when systems are simulated independently but are dependent events when systems are simulated under CRN. As we discard observations from Phase I and completely restart for Phase II, and as $\mathcal{B}_2^*$ involves making the correct selection from all systems in $S_{a'} \cup S_d$ (not only the ones surviving from Phase I), $\mathcal{B}_2^*$ is independent from $\mathcal{A}_1^*, \mathcal{A}_2^*$, and $\mathcal{B}_1^*$. We have

$$\begin{aligned}\Pr\{\text{CS}\} &\geq \Pr(\mathcal{A}_1^* \cap \mathcal{A}_2^* \cap \mathcal{B}^*) \geq \Pr(\mathcal{A}_1^* \cap \mathcal{A}_2^* \cap \mathcal{B}_1^* \cap \mathcal{B}_2^*)\\ &= \begin{cases}\Pr(\mathcal{A}_1^*) \times \Pr(\mathcal{A}_2^*) \times \Pr(\mathcal{B}_1^*) \times \Pr(\mathcal{B}_2^*), & \text{if systems are simulated independently;}\\ \left[\Pr(\mathcal{A}_1^*) + \Pr(\mathcal{A}_2^*) + \Pr(\mathcal{B}_1^*) - 2\right] \times \Pr(\mathcal{B}_2^*), & \text{if systems are simulated under CRN.}\end{cases}\end{aligned}$$

We discuss the cases depending on whether systems are simulated independently or under CRN. When systems are simulated independently, by Lemma A.2, we have

$$\begin{aligned}&\Pr(\mathcal{A}_1^*) \geq (1 - \min\{s, d\}\alpha_f')^{j_u};\\ &\Pr(\mathcal{A}_2^*) \geq (1 - \min\{s, d-1\}\alpha_f')^{j_{a'}+j_d} = (1 - \min\{s, d-1\}\alpha_f')^{k-j_a-j_u-1};\\ &\Pr(\mathcal{B}_1^*) \geq 1 - s\alpha_f'.\end{aligned}$$

Let N_{ij} denote the number of observations taken for system i before a comparison decision is made between systems i and j, and let N_i denote the maximum number of observations that system i takes within Phase II. That is

$$N_{ij} = \left\lceil \frac{2c\eta_c(n_0-1)S^2_{X_{ij}}(n_0)}{\delta^2} \right\rceil, \text{ and } N_i = \max_{j\neq i} N_{ij}.$$

Then we have

$$\begin{aligned}\Pr(\mathcal{B}_2^*) &\geq \Pr\left(\cap_{i\in S_{a'}\cup S_d}\text{CS}_i\right)\\ &= \text{E}\left[\Pr\left\{\cap_{i\in(S_d\cup S_{a'})}\text{CS}_i \middle| X_{[b]1}, \ldots, X_{[b],N_{[b]}}, S^2_{X_{i[b]}}(n_0)\right\}\right]\\ &= \text{E}\left[\prod_{i\in(S_d\cup S_{a'})} \Pr\left\{\text{CS}_i \middle| X_{[b]1}, \ldots, X_{[b],N_{[b]}}, S^2_{X_{i[b]}}(n_0)\right\}\right]\\ &\geq \prod_{i\in(S_d\cup S_{a'})} \text{E}\left[\Pr\left\{\text{CS}_i \middle| X_{[b]1}, \ldots, X_{[b],N_{[b]}}, S^2_{X_{i[b]}}(n_0)\right\}\right]\\ &= \prod_{i\in(S_d\cup S_{a'})} \Pr\{\text{CS}_i\} \geq \prod_{i\in(S_d\cup S_{a'})} (1-\alpha_c')\\ &= (1-\alpha_c')^{j_d+j_{a'}} \geq (1-\alpha_c')^{k-j_u-j_a-1},\end{aligned} \tag{4}$$

where the second inequality holds due to Lemma 2.4 in [20] and the third inequality follows from Lemma A.3.

Thus, we know that

$$\begin{aligned}\Pr\{\text{CS}\} &\geq (1 - \min\{s, d\}\alpha_f')^{j_u} \times (1 - \min\{s, d-1\}\alpha_f')^{k-j_a-j_u-1} \times (1 - s\alpha_f') \times (1-\alpha_c')^{k-j_u-j_a-1}\\ &\geq (1 - \min\{s, d\}\alpha_f')^{j_u} \times (1 - \min\{s, d\}\alpha_f')^{k-j_a-j_u-1} \times (1 - s\alpha_f') \times (1-\alpha_c')^{k-j_u-j_a-1}\\ &= (1 - \min\{s, d\}\alpha_f')^{k-j_a-1} \times (1 - s\alpha_f') \times (1-\alpha_c')^{k-j_u-j_a-1}\\ &\geq (1 - \min\{s, d\}\alpha_f')^{k-1} \times (1 - s\alpha_f') \times (1-\alpha_c')^{k-1}\\ &= (1-\alpha_f) \times \left[(1-\alpha_c)^{1/(k-1)}\right]^{k-1} = (1-\alpha_f)(1-\alpha_c) = 1-\alpha,\end{aligned}$$

where the third inequality holds since the lower bound of $(1-\min\{s,d\}\alpha_f')^{k-j_a-1}$ is achieved when $j_a = 0$ when $0 < \alpha_f' < 1/s$, and the lower bound of $(1-\alpha_c')^{k-j_u-j_a-1}$ is achieved when $j_a = j_u = 0$ for $0 \le 1-\alpha_c' < 1$.

When systems are simulated under CRN, by Lemmas A.2, A.3, and the Bonferroni inequality, we have

$$\begin{aligned}
\Pr(\mathcal{A}_1^*) &\ge 1 - j_u \min\{s,d\}\alpha_f';\\
\Pr(\mathcal{A}_2^*) &\ge 1 - (j_{a'} + j_d)\min\{s,d-1\}\alpha_f' = 1-(k-j_a-j_u-1)\min\{s,d-1\}\alpha_f';\\
\Pr(\mathcal{B}_1^*) &\ge 1 - s\alpha_f';\\
\Pr(\mathcal{B}_2^*) &\ge \Pr\left(\cap_{i\in S_{a'}\cup S_d}\mathrm{CS}_i\right) \ge 1 - \sum_{i\in(S_d\cup S_{a'})}\Pr(\mathrm{ICS}_i) \ge 1-(j_d+j_{a'})\alpha_c'\\
&= 1-(k-j_u-j_a-1)\alpha_c',
\end{aligned}$$

where ICS_i denotes the incorrect selection event between system $i \in S_d \cup S_{a'}$ and system $[b]$. Thus,

$$\begin{aligned}
\Pr\{\mathrm{CS}\} &\ge \left[1 - j_u\min\{s,d\}\alpha_f' + 1 - (k-j_a-j_u-1)\min\{s,d-1\}\alpha_f' + 1 - s\alpha_f' - 2\right]\\
&\qquad \times \left[1-(k-j_u-j_a-1)\alpha_c'\right]\\
&\ge \left[1 - j_u\min\{s,d\}\alpha_f' + 1 - (k-j_a-j_u-1)\min\{s,d\}\alpha_f' + 1 - s\alpha_f' - 2\right]\\
&\qquad \times \left[1-(k-j_u-j_a-1)\alpha_c'\right]\\
&= \left[1-(k-j_a-1)\min\{s,d\}\alpha_f' - s\alpha_f'\right] \times \left[1-(k-j_u-j_a-1)\alpha_c'\right]\\
&\ge \left[1-(k-1)\min\{s,d\}\alpha_f' - s\alpha_f'\right] \times [1-(k-1)\alpha_c'] = (1-\alpha_f)(1-\alpha_c) = 1-\alpha,
\end{aligned}$$

where the third inequality holds since $\alpha_f', \alpha_c' > 0$, and hence the lower bound of $(k-j_a-1)\min\{s,d\}\alpha_f'$ is achieved when $j_a = 0$, and the lower bound of $1-(k-j_u-j_a-1)\alpha_c'$ is achieved when $j_a = j_u = 0$.

Case 2: $\theta^* = d+1$**.**

If $\theta^* = d+1$, there are no desirable systems for any threshold vector. Based on the definition of CS, CS is to either declare all systems are infeasible or to select an acceptable system with respect to any of the threshold vectors $\mathbf{q}^{(1)}, \ldots, \mathbf{q}^{(d)}$. Therefore, CS is ensured by correctly concluding feasibility decisions for all system $i \in S_u$. Then $\Pr(\mathrm{CS}) \ge \Pr(\mathcal{A}_1^*)$ and Lemma A.2 and the Bonferroni inequality yield

$$\begin{aligned}
\Pr\{\mathrm{CS}\} &\ge \begin{cases} (1-\min\{s,d\}\alpha_f')^{j_u}, & \text{if systems are simulated independently,}\\ 1 - j_u\min\{s,d\}\alpha_f', & \text{if systems are simulated under CRN,}\end{cases}\\
&\ge \begin{cases} (1-\min\{s,d\}\alpha_f')^{k}, & \text{if systems are simulated independently,}\\ 1 - k\min\{s,d\}\alpha_f', & \text{if systems are simulated under CRN,}\end{cases}
\end{aligned}$$

where the last inequality is due to the fact that $1 \le j_u \le k$ and $0 < \min\{s,d\}\alpha_f' < 1$. When systems are simulated independently, we have

$$\begin{aligned}
\Pr\{\mathrm{CS}\} &\ge (1-\min\{s,d\}\alpha_f')^k \ge (1-\min\{s,d\}\alpha_f')^{k-1}\cdot(1-s\alpha_f')\\
&= 1-\alpha_f > 1-\alpha.
\end{aligned}$$

When systems are simulated under CRN, we have

$$\begin{aligned}\Pr\{\text{CS}\} &\geq 1 - k\min\{s,d\}\alpha'_f \geq 1-(k-1)\min\{s,d\}\alpha'_f - s\alpha'_f \\ &= 1-\alpha_f > 1-\alpha. \end{aligned}$$

□

B IMPLEMENTATION PARAMETERS FOR $\mathcal{ZAK}^{\mathcal{R}}$ AND $\mathcal{ZAK}$

In this section, we provide detailed discussion about how we set the implementation parameters for the two proposed sequentially-running procedures $\mathcal{ZAK}^{\mathcal{R}}$ and $\mathcal{ZAK}$ in Appendices B.1 and B.2, respectively.

B.1 Implementation Parameters for $\mathcal{ZAK}^{\mathcal{R}}$

The choices of α_f and α_c affect the performance of the $\mathcal{ZAK}^{\mathcal{R}}$ procedure. If Phase I is difficult (e.g., the secondary performance measures of many systems are close to some of the threshold values in threshold vectors $\mathbf{q}^{(1)},\dots,\mathbf{q}^{(\theta^*)}$), one may want to choose a larger value for α_f than α_c to improve the efficiency. On the other hand, if Phase I is relatively easy compared with Phase II, then it is more efficient to assign a larger value of α_c than α_f. If the decision maker has knowledge on the relative difficulty of the feasibility checks and the comparison, she may first decide the choice of $e_1 = \alpha_f/\alpha_c$, the ratio of the nominal error of Phase I to Phase II. Then we have

$$(1-e_1\times\alpha_c)(1-\alpha_c) = e_1\alpha_c^2 - (e_1+1)\alpha_c + 1 = 1-\alpha.$$

Since the left-hand side equals 1 when $\alpha_c = 0$ and 0 when $\alpha_c = \min\{1, 1/e_1\}$, there must be exactly one root α_c with $\alpha_c, e_1\times\alpha_c \in (0,1)$. We have $\alpha_c = \frac{e_1+1-\sqrt{(e_1+1)^2-4e_1\alpha}}{2e_1}$ (the other root does not satisfy $\alpha_c < \min\{1,1/e_1\}$) and $\alpha_f = e_1\times\alpha_c$. However, the decision maker usually does not have such information about the mean configurations of the primary and secondary performance measures of the systems. One possibility is to select $\alpha_f = \alpha_c = 1-(1-\alpha)^{1/2}$.

If $s \leq d$, the formulas for selecting α'_f and α'_c in Algorithm A.1 suggest one may first choose $e_2 = s\alpha'_f/\alpha'_c$ (the ratio of the nominal error for feasibility checks across all the constraints for one system and the nominal error for the comparison between best system $[b]$ and one inferior system) and further find α'_f and α'_c depending on the value of e_2. Similarly, one may consider $e_2 = d\alpha'_f/\alpha'_c$ if $d < s$.

We start with the case when $s \leq d$. When systems are simulated independently, we know that

$$1-\alpha = (1-\alpha_f)(1-\alpha_c) = (1-s\alpha'_f)^k \times (1-\alpha'_c)^{k-1} = (1-e_2\alpha'_c)^k(1-\alpha'_c)^{k-1},$$

where one can numerically solve for α'_c and $\alpha'_f = e_2\alpha'_c/s$. Since the right-hand side equals 1 when $\alpha'_c = 0$ and 0 when $\alpha'_c = \min\{1, 1/e_2\}$, there must be one exactly root α'_c with $\alpha'_c, e_2\times\alpha'_c \in (0,1)$ and it follows that $0 < \alpha'_f = e_2\alpha'_c/s < 1/s$ as desired. When systems are simulated under CRN, we know that

$$\begin{aligned}1-\alpha &= (1-\alpha_f)(1-\alpha_c) = (1-ks\alpha'_f)\times(1-(k-1)\alpha'_c) = (1-ke_2\alpha'_c)\times(1-(k-1)\alpha'_c) \\ &= e_2k(k-1)(\alpha'_c)^2 - (e_2k+k-1)\alpha'_c + 1.\end{aligned}$$

Since the right-hand side equals 1 when $\alpha'_c = 0$ and 0 when $\alpha'_c = \min\{\frac{1}{k-1}, \frac{1}{e_2k}\}$, there must be exactly one root α'_c with $(k-1)\alpha'_c, e_2k\alpha'_c \in (0,1)$. Thus, we have $\alpha'_c = \frac{e_2k+k-1-\sqrt{(e_2k+k-1)^2-4e_2k(k-1)\alpha}}{2e_2k(k-1)}$ (the other root does not satisfy $\alpha'_c < \min\{\frac{1}{k-1}, \frac{1}{e_2k}\}$).

We then discuss the case when $d < s$. We set $e_2 = d\alpha'_f/\alpha'_c$ and find α'_c by solving

$$\begin{cases}(1-e_2\alpha'_c)^{k-1}\times(1-e_2\frac{s}{d}\alpha'_c)\times(1-\alpha'_c)^{k-1}=1-\alpha, & \text{if systems are simulated independently;}\\ \left(1-e_2\left(k-1+\frac{s}{d}\right)\alpha'_c\right)\times\left[1-(k-1)\alpha'_c\right]=1-\alpha, & \text{if systems are simulated under CRN.}\end{cases}$$

The former can be solved numerically. As the left-hand side equals 1 when $\alpha'_c = 0$ and 0 when $\alpha'_c = \min\{\frac{d}{se_2}, 1\}$, there must be a root α'_c with $e_2\alpha'_c, \frac{se_2}{d}\alpha'_c, \alpha'_c \in (0,1)$ and it follows that $0 < \alpha'_f = e_2\alpha'_c/d < 1/s$ as desired. For the latter, since the left-hand side equals 1 when $\alpha'_c = 0$ and 0 when $\alpha'_c = \min\{\frac{1}{k-1}, \frac{1}{e_2(k-1+\frac{s}{d})}\}$, there must be one root α'_c with $(k-1)\alpha'_c, e_2(k-1+\frac{2}{d})\alpha'_c \in (0,1)$. Therefore, we have $\alpha'_c = \frac{e_2(k-1+\frac{s}{d})+k-1-\sqrt{[e_2(k-1+\frac{s}{d})+k-1]^2-4e_2(k-1+\frac{s}{d})(k-1)\alpha}}{2e_2(k-1+\frac{s}{d})(k-1)}$ as the other root does not satisfy $\alpha'_c < \min\{\frac{1}{k-1}, \frac{1}{e_2(k-1+\frac{2}{d})}\}$.

In reality, the decision maker usually does not have detailed information regarding the mean performance of each system. One recommendation is to balance the error between the feasibility checks and the comparison. For example, if one has a single threshold vector and wishes to allocate the same amount of error for feasibility checks for all constraints of one system as for the comparison of one system with the best system $[b]$, then $e_1 = 1$ and $e_2 = 1$ are appropriate choices. On the other hand, if one wants to allocate the same error for feasibility check for one constraint of one system as for comparison of one system with the best system $[b]$, then $e_1 = s$ and $e_2 = s$ are appropriate. Note that this agrees with the discussion from [8] who consider a single threshold vector under the MIM configuration and test the formulation using $e_1 = 1$. They recommend to set the ratio of the difficulty between feasibility checks and comparison to 1 on the grounds that this choice is robust to differing numbers of constraints, numbers of feasible systems, and variance configurations. When multiple threshold vectors are considered, we need to ensure more correct events during the feasibility checks (see the detailed analysis in the proof of statistical validity of $\mathcal{ZAK}^{\mathcal{R}}$ in this section and further analysis in Section 4.2). Therefore, larger values of e_1 and e_2 may be more appropriate than in the single threshold vector case. More specifically, most of our experimental results (Section 6) consider the e_2 formulation with $e_2 = 2$ (see the analysis in Section 6.2).

B.2 Implementation Parameters for $\mathcal{ZAK}$

To find the values of α'_f and α'_c, after choosing the value of e_2, one needs to solve

$$\begin{cases}\alpha = 1-(1-\min\{s,d\}\alpha'_f)^{k-1}\times(1-s\alpha'_f)+1-(1-\alpha'_c)^{|F|-1}, & \text{if systems are simulated independently;}\\ \alpha = [(k-1)\min\{s,d\}+s]\alpha'_f+(|F|-1)\alpha'_c, & \text{if systems are simulated under CRN.}\end{cases} \tag{5}$$

As the decision maker does not have the information on the number of surviving systems for Phase II (i.e., the value of $|F|$) prior to the execution of Algorithm 1, she may first find α'_f by assuming that the number of surviving systems for Phase II is k (i.e., by assuming $|F| = k$).

When $s \le d$, one may find α'_c by solving

$$\begin{cases}\alpha = 1-(1-e_2\alpha'_c)^k+1-(1-\alpha'_c)^{k-1}, & \text{if systems are simulated independently;}\\ \alpha = ke_2\alpha'_c+(k-1)\alpha'_c. & \text{if systems are simulated under CRN,}\end{cases}$$

When systems are simulated independently, the right-hand side equals 0 when $\alpha'_c = 0$. When $\alpha'_c = \min\{1, \frac{1}{e_2}\}$, one of the terms $1-(1-e_2\alpha'_c)^k, 1-(1-\alpha'_c)^{k-1}$ on the right-hand side equals 1 and the other is positive, and hence the right-hand side is greater than 1. Thus, there must be a root

α_c' with $\alpha_c', e_2\alpha_c' \in (0,1)$ and it follows that $0 < \alpha_f' = e_2\alpha_c/s < 1/s$ as desired. When systems are simulated under CRN, we find $\alpha_c' = \frac{\alpha}{ke_2+k-1}$. The corresponding α_f' can be found as $\alpha_f' = e_2\alpha_c'/s$.

When $d < s$, one may find α_c' by solving

$$\begin{cases} \alpha = 1 - (1 - e_2\alpha_c')^{k-1} \times (1 - \frac{s}{d}e_2\alpha_c') + 1 - (1 - \alpha_c')^{k-1}, & \text{if systems are simulated independently;} \\ \alpha = \frac{(k-1)d+s}{d} e_2\alpha_c' + (k-1)\alpha_c', & \text{if systems are simulated under CRN.} \end{cases}$$

When systems are simulated independently, the right-hand side equals 0 when $\alpha_c' = 0$. When $\alpha_c' = \min\{\frac{d}{se_2}, 1\}$, the right-hand side is greater than 1 (because one of the terms $(1 - e_2\alpha_c')^{k-1} \times (1 - \frac{s}{d}e_2\alpha_c')$, $1 - (1 - \alpha_c')^{k-1}$ equals 1 and the other one is positive). Thus, there must be a root α_c' with $e_2\alpha_c', \frac{se_2}{d}\alpha_c', \alpha_c' \in (0,1)$ and hence $0 < \alpha_f' = e_2\alpha_c'/d < 1/s$ as desired. When systems are simulated under CRN, we find $\alpha_c' = \frac{\alpha}{\frac{(k-1)d+s}{d}e_2+k-1}$. The corresponding α_f' can be found as $\alpha_f' = e_2\alpha_c'/d$.

After the completion of Phase I, with the information on the number of surviving systems $|F|$, we may solve for an updated value for α_c', namely α_c'', by solving Equation (5) where α_f' and α_c' are replaced by the value of α_f' we already computed (i.e., $\alpha_f' = e_2\alpha_c'/\min\{s,d\}$) and α_c'', respectively.

C STATISTICAL VALIDITY OF $\mathcal{ZAK}$+

In this section, we provide the proof of Theorem 4.3.

Proof. We consider two cases, namely when $\theta^* \le d$ and $\theta^* = d+1$.

Case 1: $\theta^* \le d$.

We consider the events $\mathcal{A}_1^*, \mathcal{A}_2^*, \mathcal{B}_1^*$, and $\mathcal{B}_2^*$ defined in Section A. Notice that $\mathcal{A}_2^* \cap \mathcal{B}_2^*$ is the event that all systems in $S_{a'} \cup S_d$ are declared infeasible to threshold vectors $\mathbf{q}^{(1)}, \dots, \mathbf{q}^{(\theta^*-1)}$ and are eliminated by comparison with system $[b]$, that is, $\mathcal{A}_2^* \cap \mathcal{B}_2^* = \cap_{i \in S_d \cup S_{a'}} \mathcal{A}_2^*(i) \cap \mathrm{CS}_i$. Similarly, $\mathcal{A}_1^* = \cap_{i \in S_u} \mathcal{A}_1^*(i)$.

We discuss the cases depending on whether systems are simulated independently or under CRN. When systems are simulated independently, as $\mathcal{ZAK}$+ performs Phases I and II simultaneously, events $\mathcal{A}_2^*, \mathcal{B}_1^*$, and $\mathcal{B}_2^*$ are dependent whereas $\mathcal{A}_1^*$ is independent of $\mathcal{A}_2^* \cap \mathcal{B}_1^* \cap \mathcal{B}_2^*$. We then have

$$\begin{aligned} \Pr\{\mathrm{CS}\} &\ge \Pr\left\{\mathcal{A}_1^* \cap \mathcal{A}_2^* \cap \mathcal{B}_1^* \cap \mathcal{B}_2^*\right\} \\ &= \Pr\left(\mathcal{A}_1^*\right) \times \Pr\left(\mathcal{A}_2^* \cap \mathcal{B}_1^* \cap \mathcal{B}_2^*\right) \\ &\ge \Pr\left(\mathcal{A}_1^*\right) \times \left[\Pr\left(\mathcal{A}_2^* \cap \mathcal{B}_2^*\right) + \Pr\left(\mathcal{B}_1^*\right) - 1\right]. \end{aligned}$$

By Lemma 4.1, we have

$$\begin{aligned} \Pr\left(\mathcal{A}_1^*\right) &\ge \left(1 - \min\{s,d\}\beta_f\right)^{j_u}; \\ \Pr\left(\mathcal{B}_1^*\right) &\ge 1 - s\beta_f. \end{aligned}$$

We use the same notation N_{ij} from the proof of Theorem A.4 and have

$$\begin{aligned} \Pr\left(\mathcal{A}_2^* \cap \mathcal{B}_2^*\right) &= \Pr\left(\cap_{i \in (S_d \cup S_{a'})}\left(\mathcal{A}_2^*(i) \cap \mathrm{CS}_i\right)\right) \\ &= \mathrm{E}\left[\Pr\left\{\cap_{i \in (S_d \cup S_{a'})}\left(\mathcal{A}_2^*(i) \cap \mathrm{CS}_i\right) \middle| X_{[b]1}, \dots, X_{[b],N_{[b]}}, S^2_{X_{i[b]}}(n_0)\right\}\right] \\ &= \mathrm{E}\left[\prod_{i \in (S_d \cup S_{a'})} \Pr\left\{\mathcal{A}_2^*(i) \cap \mathrm{CS}_i \middle| X_{[b]1}, \dots, X_{[b],N_{[b]}}, S^2_{X_{i[b]}}(n_0)\right\}\right] \\ &\ge \prod_{i \in (S_d \cup S_{a'})} \mathrm{E}\left[\Pr\left\{\mathcal{A}_2^*(i) \cap \mathrm{CS}_i \middle| X_{[b]1}, \dots, X_{[b],N_{[b]}}, S^2_{X_{i[b]}}(n_0)\right\}\right] \end{aligned}$$

$$
\begin{aligned}
&\geq \prod_{i\in(S_d\cup S_{a'})}\left[1-\mathrm{E}\left[\Pr\left\{\left(\mathcal{A}_2^*(i)\right)^c\middle|X_{[b]1},\ldots,X_{[b],N_{[b]}},S_{X^2_{i[b]}}(n_0)\right\}\right]\right.\\
&\qquad\left.-\,\mathrm{E}\left[\Pr\left\{\mathrm{ICS}_i\middle|X_{[b]1},\ \ldots,X_{[b],N_{[b]}},S_{X^2_{i[b]}}(n_0)\right\}\right]\right]\\
&= \prod_{i\in(S_d\cup S_{a'})}\left[1-\Pr\left\{\left(\mathcal{A}_2^*(i)\right)^c\right\}-\Pr\left\{\mathrm{ICS}_i\right\}\right]\\
&\geq \prod_{i\in(S_d\cup S_{a'})}\left(1-\min\{s,d-1\}\beta_f-\beta_c\right)=\left(1-\min\{s,d-1\}\beta_f-\beta_c\right)^{j_d+j_{a'}}\\
&= \left(1-\min\{s,d-1\}\beta_f-\beta_c\right)^{k-j_a-j_u-1},
\end{aligned}
$$

where we use A^c to denote the complement event of A. The first inequality is from Lemma 2.4 of [20], the second inequality holds due to the Bonferroni inequality, and the last inequality is from Lemmas 4.1 and 4.2.

Thus, we know that

$$
\begin{aligned}
\Pr\{\mathrm{CS}\} &\geq \left(1-\min\{s,d\}\beta_f\right)^{j_u}\times\left[\left(1-\min\{s,d-1\}\beta_f-\beta_c\right)^{k-j_a-j_u-1}+\left(1-s\beta_f\right)-1\right]\\
&\geq \left(1-\min\{s,d\}\beta_f\right)^{j_u}\times\left[\left(1-\min\{s,d-1\}\beta_f-\beta_c\right)^{k-j_u-1}-s\beta_f\right],
\end{aligned}
$$

where the second inequality holds since the lower bound of $(1-\min\{s,d-1\}\beta_f-\beta_c)^{k-j_a-j_u-1}$ is achieved when $j_a=0$ for $0<1-\min\{s,d-1\}\beta_f-\beta_c<1$. As $0\leq j_u\leq k-1$ (because $\theta^*\leq d$), we know that

$$
\Pr\{\mathrm{CS}\}\geq \min_{0\leq j\leq k-1}\left\{\left(1-\min\{s,d\}\beta_f\right)^j\times\left[\left(1-\min\{s,d-1\}\beta_f-\beta_c\right)^{k-j-1}-s\beta_f\right]\right\}=1-\alpha.
$$

When systems are simulated under CRN, events $\mathcal{A}_1^*,\mathcal{A}_2^*,\mathcal{B}_1^*$, and $\mathcal{B}_2^*$ are all dependent. Thus, we have

$$
\Pr\{\mathrm{CS}\}\geq\Pr\left\{\mathcal{A}_1^*\cap\mathcal{A}_2^*\cap\mathcal{B}_1^*\cap\mathcal{B}_2^*\right\}\geq\Pr\left(\mathcal{A}_1^*\right)+\Pr\left(\mathcal{A}_2^*\cap\mathcal{B}_2^*\right)+\Pr\left(\mathcal{B}_1^*\right)-2.
$$

By Lemmas 4.1 and 4.2, and the Bonferroni inequality, we have

$$
\begin{aligned}
\Pr\left(\mathcal{A}_1^*\right)&\geq 1-j_u\min\{s,d\}\beta_f;\\
\Pr\left(\mathcal{B}_1^*\right)&\geq 1-s\beta_f;\\
\Pr\left(\mathcal{A}_2^*\cap\mathcal{B}_2^*\right)&=\Pr\left(\cap_{i\in(S_d\cup S_{a'})}\left(\mathcal{A}_2^*(i)\cap\mathrm{CS}_i\right)\right)\\
&\geq 1-\sum_{i\in(S_d\cup S_{a'})}\left[\Pr\left(\mathcal{A}_2^*(i)\right)^c+\Pr(\mathrm{ICS}_i)\right]\\
&\geq 1-\sum_{i\in(S_d\cup S_{a'})}\left[\min\{s,d-1\}\beta_f+\beta_c\right]\\
&=1-(j_d+j_{a'})\left[\min\{s,d-1\}\beta_f+\beta_c\right]\\
&=1-(k-j_a-j_u-1)\left[\min\{s,d-1\}\beta_f+\beta_c\right],
\end{aligned}
$$

where the first inequality holds due to the Bonferroni inequality and the second inequality holds by Lemmas 4.1 and 4.2.

Thus, we know that

$$
\begin{aligned}
\Pr\{\mathrm{CS}\}&\geq 1-j_u\min\{s,d\}\beta_f+\left\{1-(k-j_a-j_u-1)\left[\min\{s,d-1\}\beta_f+\beta_c\right]\right\}+1-s\beta_f-2\\
&\geq 1-j_u\min\{s,d\}\beta_f-(k-j_u-1)\left[\min\{s,d-1\}\beta_f+\beta_c\right]-s\beta_f
\end{aligned}
$$

$$= 1 - [j_u \min\{s,d\} + (k - j_u - 1) \min\{s, d-1\} + s]\, \beta_f - (k - j_u - 1)\beta_c,$$

where the second inequality holds since the lower bound of $1-(k-j_a-j_u-1)\left[\min\{s,d-1\}\beta_f + \beta_c\right]$ is achieved when $j_a = 0$. As $0 \le j_u \le k-1$, we know that

$$\Pr\{\mathrm{CS}\} \ge \min_{0\le j\le k-1} \left\{1 - [j \min\{s,d\} + (k-j-1)\min\{s,d-1\} + s]\, \beta_f - (k-j-1)\beta_c\right\} = 1-\alpha.$$

Case 2: $\theta^* = d+1$.

If $\theta^* = d+1$, there are no desirable systems for any threshold vector. Similar to the discussion in the proof of Theorem A.4, CS is ensured by correctly concluding feasibility decisions for all systems $i \in S_u$. Then $\Pr\{\mathrm{CS}\} \ge \Pr(\mathcal{A}_1^*)$ and Lemma 4.1 and the Bonferroni inequality yield

$$\begin{aligned}\Pr\{\mathrm{CS}\} &\ge \begin{cases}(1-\min\{s,d\}\beta_f)^{j_u}, & \text{if systems are simulated independently,}\\ 1 - j_u \min\{s,d\}\beta_f, & \text{if systems are simulated under CRN.}\end{cases}\\ &\ge \begin{cases}(1-\min\{s,d\}\beta_f)^{k}, & \text{if systems are simulated independently,}\\ 1 - k \min\{s,d\}\beta_f, & \text{if systems are simulated under CRN,}\end{cases}\end{aligned}$$

where the last inequality is due to the fact that $1 \le j_u \le k$ and $0 < \min\{s,d\}\beta_f < 1$. When systems are simulated independently, we have

$$\begin{aligned}\Pr\{\mathrm{CS}\} &\ge \left(1-\min\{s,d\}\beta_f\right)^k \ge \left(1-\min\{s,d\}\beta_f\right)^{k-1}\left(1-s\beta_f\right)\\ &= \left(1-\min\{s,d\}\beta_f\right)^{k-1}\left[\left(1-\min\{s,d-1\}\beta_f-\beta_c\right)^{k-(k-1)-1} - s\beta_f\right]\\ &\ge \min_{0\le j\le k-1}\left\{\left(1-\min\{s,d\}\beta_f\right)^j\left[\left(1-\min\{s,d-1\}\beta_f-\beta_c\right)^{k-j-1} - s\beta_f\right]\right\}\\ &= 1-\alpha,\end{aligned}$$

where the second inequality holds since $\min\{s,d\} \le s$ and $0 < \min\{s,d\}\beta_f < 1$ and the first equality holds since $(1-\min\{s,d-1\}\beta_f - \beta_c)^0 = 1$.

When systems are simulated under CRN, we have

$$\begin{aligned}\Pr\{\mathrm{CS}\} &\ge 1 - k\min\{s,d\}\beta_f \ge 1 - [(k-1)\min\{s,d\} + s]\beta_f\\ &= 1 - [(k-1)\min\{s,d\} + (k-(k-1)-1)\min\{s,d-1\} + s]\,\beta_f - (k-(k-1)-1)\beta_c\\ &\ge \min_{0\le j\le k-1}\left[1 - [j\min\{s,d\} + (k-j-1)\min\{s,d-1\} + s]\,\beta_f - (k-j-1)\beta_c\right]\\ &= 1-\alpha.\end{aligned}$$

□

D IMPLEMENTATION PARAMETERS FOR $\mathcal{ZAK}$+

We start by considering the case when $s < d$, and the systems are simulated independently. In this case, we need to find β_f and β_c such that

$$\min_{0\le j\le k-1}\left\{(1-\min\{s,d\}\beta_f)^j \times \left[(1-\min\{s,d-1\}\beta_f-\beta_c)^{k-j-1} - s\beta_f\right]\right\} = 1-\alpha.$$

Let $\beta = s\beta_f = e\beta_c$. Then we have

$$\Pr\{\mathrm{CS}\} \ge \min_{0\le j\le k-1}\left\{(1-\beta)^j \times \left[(1-(1+1/e)\beta)^{k-j-1} - \beta\right]\right\}.$$

Let $f(j)$ be a function of j such that $f(j) = (1-\beta)^j \times \left[(1-(1+1/e)\beta)^{k-j-1} - \beta\right]$. We need to find the lower bound of $f(j)$ given that $0 \le j \le k-1$. Treating j as a continuous variable, the first derivative of $f(j)$ is

$$\begin{aligned}\frac{\partial}{\partial j}f(j) &= (1-\beta)^j\log(1-\beta)\left[(1-(1+1/e)\beta)^{k-j-1}-\beta\right]\\ &\quad -(1-\beta)^j(1-(1+1/e)\beta)^{k-j-1}\log(1-(1+1/e)\beta)\\ &=(1-\beta)^j\left\{[\log(1-\beta)-\log(1-(1+1/e)\beta)](1-(1+1/e)\beta)^{k-j-1}-\beta\log(1-\beta)\right\}>0,\end{aligned}$$

where the last inequality holds since $\log(1-\beta)>\log(1-(1+1/e)\beta)$ and $\log(1-\beta)<0$. Therefore, we know that $f(j)$ is increasing. Given that $0\le j\le k-1$, $f(j)$ achieves its minimum when $j=0$. Hence, to find β_f and β_c, we solve

$$(1-\beta)^0\times\left[(1-(1+1/e)\beta)^{k-0-1}-\beta\right]=(1-(1+1/e)\beta)^{k-1}-\beta=1-\alpha.$$

The resulting β is the common value of $e\beta_c$ and $s\beta_f$. We see that $(1-(1+1/e)\beta)^{k-1}-\beta$ equals 1 when $\beta=0$ and is negative when $\beta=\frac{e}{e+1}$. Thus, there exists a solution β with $0<\beta<\frac{e}{e+1}$ that solves $(1-(1+1/e)\beta)^{k-1}-\beta=1-\alpha$, which can be found numerically. It follows that $0<\beta_f=\beta/s<\frac{e}{e+1}\times\frac{1}{s}<\frac{1}{s}, 0<\beta_c=\frac{\beta}{e}<\frac{1}{e+1}<1$, and $0<1-(1+\frac{1}{e})\beta\le 1-\min\{s,d-1\}\frac{\beta}{s}-\frac{\beta}{e}=1-\min\{s,d-1\}\beta_f-\beta_c<1$ as desired.

We then consider the case when $s<d$ and the systems are simulated under CRN. We need to find β_f and β_c such that

$$\min_{0\le j\le k-1}\left\{1-[j\min\{s,d\}+(k-j-1)\min\{s,d-1\}+s]\beta_f-(k-j-1)\beta_c\right\}=1-\alpha.$$

By setting $\beta=s\beta_f=e\beta_c$, we have

$$\Pr\{\mathrm{CS}\}\ge\min_{0\le j\le k-1}\left\{1-\left(k+\frac{k-j-1}{e}\right)\beta\right\}=1-\left(k+\frac{k-1}{e}\right)\beta,$$

and the value of $s\beta_f$ and $e\beta_c$ can be found as $s\beta_f=e\beta_c=\alpha/[k+(k-1)/e]$.

When $s\ge d$, by setting $\beta=d\beta_f=e\beta_c$, we need to find β such that

$$\begin{cases}1-\alpha=\min_{0\le j\le k-1}\left\{(1-\beta)^j\times\left[\left(1-\frac{d-1}{d}\beta-\frac{1}{e}\beta\right)^{k-j-1}-\frac{s}{d}\beta\right]\right\}, & \text{if systems are simulated independently;}\\ 1-\alpha=\min_{0\le j\le k-1}\left\{1-\left[j+\frac{(d-1)(k-j-1)+s}{d}+\frac{k-j-1}{e}\right]\beta\right\}, & \text{if systems are simulated under CRN.}\end{cases}$$

When systems are simulated independently, for a fixed j such that $0\le j\le k-1$, $(1-\beta)^j\times[(1-\frac{d-1}{d}\beta-\frac{1}{e}\beta)^{k-j-1}-\frac{s}{d}\beta]$ equals 1 when $\beta=0$ and is non-positive when $\beta=\min\{\frac{1}{1-\frac{1}{d}+\frac{1}{e}},\frac{d}{s}\}$ (because $(1-\beta)^j\ge 0$ and $(1-\frac{d-1}{d}\beta-\frac{1}{e}\beta)^{k-j-1}=0$ when $\beta=\frac{1}{1-\frac{1}{d}+\frac{1}{e}}$ and $\frac{s}{d}\beta=1$ when $\beta=\frac{d}{s}$). Thus, there must be a solution β_j with $(1-\frac{1}{d}+\frac{1}{e})\beta_j, \frac{s}{d}\beta_j\in(0,1)$. We then let $f_j(\beta)$ be a function of β with a fixed j such that $f_j(\beta)=(1-\beta)^j\times[(1-\frac{d-1}{d}\beta-\frac{1}{e}\beta)^{k-j-1}-\frac{s}{d}\beta]$. The first derivative of $f_j(\beta)$ is

$$\begin{aligned}\frac{\partial}{\partial\beta}f_j(\beta) &= -j(1-\beta)^{j-1}\left[\left(1-\left(1-\frac{1}{d}+\frac{1}{e}\right)\beta\right)^{k-j-1}-\frac{s}{d}\beta\right]\\ &\quad -(1-\beta)^j\left[(k-j-1)\left(1-\frac{1}{d}+\frac{1}{e}\right)\left(1-\left(1-\frac{1}{d}+\frac{1}{e}\right)\beta\right)^{k-j-2}+\frac{s}{d}\right]<0,\end{aligned}$$

where the inequality holds for $0<\beta<\min\{\frac{1}{1-\frac{1}{d}+\frac{1}{e}},\frac{d}{s}\}$ such that $f_j(\beta)>0$. Given that $\frac{\partial}{\partial\beta}f_j(\beta)<0$ when $f_j(\beta)>0$, we know that the solution β_j is unique. We set $j_0\in\arg\min_{0\le j\le k-1}\beta_j$. As $\frac{\partial}{\partial\beta}f_j(\beta)<0$, which implies that $f_j(\beta)$ is a decreasing function in terms of β for a particular j, we know that $f_j(\beta_{j_0}) \geq 1-\alpha$ for all $1 \leq j \leq k-1$ and $f_{j_0}(\beta_{j_0}) = 1-\alpha$. We find β as $\beta = \beta_{j_0}$, which is the common value of $e\beta_c$ and $d\beta_f$. It follows that $0 < \beta_f = \frac{1}{d}\beta < \min\{\frac{1}{d+\frac{d}{e}-1}, \frac{1}{s}\} \leq \frac{1}{s}$, $0 < \beta_c = \frac{1}{e}\beta < \min\{\frac{1}{e-\frac{e}{d}+1}, \frac{1}{e}\} \leq \frac{1}{1+e(1-\frac{1}{d})} \leq 1$, and $0 < 1-(\frac{d-1}{d}+\frac{1}{e})\beta = 1-\min\{s,d-1\}\frac{\beta}{d} - \frac{\beta}{e} = 1-\min\{s,d-1\}\beta_f - \beta_c < 1$ as desired.

When systems are simulated under CRN, we find β such that

$$\begin{aligned} 1-\alpha &= \min_{0\leq j\leq k-1}\left\{1-\left[j+\frac{(d-1)(k-j-1)+s}{d}+\frac{k-j-1}{e}\right]\beta\right\} \\ &= \min_{0\leq j\leq k-1}\left\{1-\left[\left(\frac{1}{d}-\frac{1}{e}\right)j+\left(1-\frac{1}{d}+\frac{1}{e}\right)(k-1)+\frac{s}{d}\right]\beta\right\} \\ &= \begin{cases} 1-\left[\left(1-\frac{1}{d}+\frac{1}{e}\right)(k-1)+\frac{s}{d}\right]\beta, & \text{if } d \geq e, \\ 1-\left(k-1+\frac{s}{d}\right)\beta, & \text{if } d < e, \end{cases} \end{aligned}$$

and the value of $d\beta_f$ and $e\beta_c$ can be found as

$$\beta = \begin{cases} \alpha/\left[\left(1-\frac{1}{d}+\frac{1}{e}\right)(k-1)+\frac{s}{d}\right], & \text{if } d \geq e, \\ \alpha/\left(k-1+\frac{s}{d}\right), & \text{if } d < e. \end{cases}$$

We also see that $0 < \beta_f = \frac{1}{d}\beta \leq \frac{\alpha}{d(k-1+\frac{s}{d})} < \frac{1}{s}$ and $0 < \beta_c = \frac{1}{e}\beta \leq \beta < 1$ if $e \geq 1$ and $0 < \beta_c = \frac{1}{e}\beta \leq \frac{\alpha}{e(\frac{k-1}{e})} < 1$ if $e < 1 \leq d$, as desired.

E ALGORITHMS THAT CONSTRUCT THE THREE EXAMPLE PREFERENCE ORDERS

In this section, we include the algorithms used to generate the three example preference orders discussed in Section 5. More specifically, Algorithms A.2 – A.4 show the algorithm that generates ranked constraints, equally important constraints, and the total violation with ranked constraints formulation, respectively.

Note that the ranked constraints and the total violation with ranked constraints formulation require the rankings among constraints, without loss of generality, Algorithm A.2 and A.4 assume that the constraints are ranked from constraint 1 to constraint s.

ALGORITHM A.2: Constructing Threshold Vectors for Ranked Constraints.

Input $q_{\ell,m}$ for all $\ell = 1,\ldots,s$ and $m = 1,\ldots,d_\ell$. Let $\mathbf{Q}$ be an empty list of threshold vectors and let `threshold` be a vector of length s.
for $m_1 = 1,\ldots,d_1$ **do**
  for $m_2 = 1,\ldots,d_2$ **do**
    ...
    for $m_s = 1,\ldots,d_s$ **do**
      for $\ell = 1,\ldots,s$ **do**
        Set `threshold`$[\ell] = q_{\ell,m_\ell}$.
      end for
      Add `threshold` to $\mathbf{Q}$.
    end for
    ...
  end for
end for
return Q

ALGORITHM A.3: Constructing Threshold Vectors for Equally Important Constraints.

Input $q_{\ell,m}$ for all $\ell = 1, \dots, s$ and $m = 1, \dots, d_\ell$. Let $\mathbf{Q}$ be an empty list of threshold vectors and let `threshold` be a vector of length s. Set $L = \max_{\ell=1,\dots,s} d_\ell$.

```
for m = 1, ..., L do
    for ℓ = 1, ..., s do
        if m ≤ d_ℓ then
            Set threshold[ℓ] = q_{ℓ,m}.
        else
            Set threshold[ℓ] = q_{ℓ,d_ℓ}.
        end if
    end for
    Add threshold to Q.
end for
return Q
```

ALGORITHM A.4: Constructing Threshold Vectors for Total Violation with Ranked Constraints.

Input $q_{\ell,m}$ for all $\ell = 1, \dots, s$ and $m = 1, \dots, d_\ell$. Let $\mathbf{Q}$ be an empty list of threshold vectors and let `threshold` be a vector of length s.

```
for v = 0, ..., Σ_{ℓ=1}^{s} (d_ℓ − 1) do
    for v_1 = 0, ..., v do
        for v_2 = 0, ..., v − v_1 do
            for v_3 = 0, ..., v − (v_1 + v_2) do
                ...
                for v_s = v − Σ_{ℓ'=1}^{s−1} v_{ℓ'} do
                    for ℓ = 1, ..., s do
                        Set threshold[ℓ] = q_{ℓ, v_ℓ+1}.
                    end for
                end for
                ...
            end for
            Add threshold to Q.
        end for
    end for
end for
return Q
```

F PROCEDURES Restart$^{\mathcal{AK}}$ AND Restart$^{\mathcal{HAK}}$

In this section, we discuss the algorithms Restart$^{\mathcal{AK}}$ and Restart$^{\mathcal{HAK}}$ and their statistical validity. As Restart$^{\mathcal{AK}}$ is a special case of Restart$^{\mathcal{HAK}}$ when the number of constraints in consideration is one, we omit the discussion on the algorithm statement and the statistical validity of procedure Restart$^{\mathcal{AK}}$ for the sake of space.

Procedure Restart$^{\mathcal{HAK}}$ performs $\mathcal{HAK}$, due to [8], for threshold vectors $\mathbf{q}^{(1)}, \mathbf{q}^{(2)}, \dots, \mathbf{q}^{(\theta^*)}$ independently when $1 \leq \theta^* \leq d$, and for threshold vectors $\mathbf{q}^{(1)}, \mathbf{q}^{(2)}, \dots, \mathbf{q}^{(d)}$ independently when $\theta^* = d + 1$. As discussed in [8], $\mathcal{HAK}$ requires the user to choose a feasibility check procedure. In our experiments, we choose $\mathcal{F}_{\mathcal{B}}^{\mathcal{I}}$ in [8] as the feasibility check procedure. $\mathcal{HAK}$ also requires a user to input the ratio, denoted α_1/α_2, of the error for the feasibility checks and the comparison.

We set $\alpha_1/\alpha_2 = 1$ as recommended in [8] and the initial sample size when Restart$^{\mathcal{HAK}}$ applies $\mathcal{HAK}$ with respect to each threshold vector as $n_0 = 20$. Note that the results in this section can be easily generalized to a different α_1/α_2 ratio. A detailed description of Restart$^{\mathcal{HAK}}$ is shown in Algorithm A.5.

ALGORITHM A.5: Procedure Restart$^{\mathcal{HAK}}$.

[**Setup**:] Select the overall nominal confidence level $1-\alpha$. Choose tolerance levels $\epsilon_1,\ldots,\epsilon_s$, indifference-zone parameter δ, and threshold vectors $\{\mathbf{q}^{(1)},\mathbf{q}^{(2)},\ldots,\mathbf{q}^{(d)}\}$. Choose the procedure $\mathcal{F}_{\mathcal{B}}^{I}$ as the feasibility check procedure and set $\alpha' = 1-(1-\alpha)^{1/d}$.
for $\theta = 1,\ldots,d$ **do**
  [**Setup**] for $\mathcal{HAK}$: Same as in $\mathcal{HAK}$ except that α is replaced by α'. Set $\alpha_1 = \alpha_2 = \alpha'/2$.
  [**Initialization**], [**Feasibility Check**], [**Feasibility Stopping Rule**], [**Setup for Comparison**], [**Comparison**], and [**Comparison Stopping Rule**] are the same as in $\mathcal{HAK}$.
  [**Stopping Condition**]: If one system is found in [**Comparison Stopping Rule**], terminate the algorithm and select the system as the best. If no system is found in [**Feasibility Stopping Rule**] and $\theta = d$, declare no feasible system exists with respect to the given threshold vectors.
end for

As $\mathcal{HAK}$ is heuristic and Restart$^{\mathcal{HAK}}$ essentially applies $\mathcal{HAK}$ for threshold vectors $\mathbf{q}^{(1)},\mathbf{q}^{(2)},\ldots,\mathbf{q}^{(\min\{\theta^*,d\})}$, we do not prove the statistical validity of Restart$^{\mathcal{HAK}}$. However, if we consider a variation of $\mathcal{HAK}$, namely $\mathcal{HAK}^{\mathcal{R}}$ ("restart"), with a slight modification in the [**Setup**] for $\mathcal{HAK}$ (as Phases I and II are independent in $\mathcal{HAK}^{\mathcal{R}}$) and two changes in the [**Setup for Comparison**], we are able to prove the statistical validity of procedure Restart$^{\mathcal{HAK}^{\mathcal{R}}}$ that implements $\mathcal{HAK}^{\mathcal{R}}$ for threshold vectors $\mathbf{q}^{(1)},\mathbf{q}^{(2)},\ldots,\mathbf{q}^{(\min\{\theta^*,d\})}$ independently:

— In [**Setup**] for $\mathcal{HAK}$:
Set

$$\alpha_1 = \alpha_2 = \begin{cases} 1-(1-\alpha')^{k/(k+1)}, & \text{if systems are simulated independently;} \\ \frac{1}{2}\left(k+1-\sqrt{(k+1)^2-4k\alpha'}\right), & \text{if systems are simulated under CRN.} \end{cases}$$

Note that α_1 and α_2 are well-defined when systems are simulated under CRN since $(k+1)^2 - 4k\alpha' > 0$ always holds. This is because $0<\alpha'<1$ and thus $(k+1)^2-4k\alpha' > (k+1)^2-4k = (k-1)^2 \geq 0$.

— In [**Setup for Comparison**] in $\mathcal{HAK}$:
- Instead of using the observations of the primary performance measure $X_{i1},\ldots,X_{ir_i}$ collected from the [**Feasibility Check**] in $\mathcal{HAK}$, we perform a completely new simulation and collect $X_{i1},\ldots,X_{in_0}$ for system $i\in F$, and compute $\bar{X}_i(n_0)$ and $S^2_{X_{ij}}(n_0)$ for $i,j\in F$. Set $r_i = n_0$ for each system $i \in F$.
- Change $\beta_2 = \alpha_2/(|F|-1)$ to $\beta_2 = \begin{cases} 1-(1-\alpha_2)^{1/(k-1)}, & \text{if systems are simulated independently;} \\ \alpha_2/(k-1), & \text{if systems are simulated under CRN.} \end{cases}$

Note that [8] use F to denote the set of systems that are declared feasible with respect to $\mathbf{q}^{(\theta^*)}$ in Phase I.

To prove the statistical validity of Restart$^{\mathcal{HAK}^{\mathcal{R}}}$, we consider similar notation as in Section 2.2. Recall that we use $[b]$ to denote the index of the best system among the desirable systems with respect to $\mathbf{q}^{(\theta^*)}$. We further let $\mathrm{CS}^{(\theta)}$ be the correct selection event with respect to threshold vector

$\mathbf{q}^{(\theta)}$. Then if $\theta = 1, \ldots, \min\{\theta^*, d\}$,

$$\mathrm{CS}^{(\theta)} = \begin{cases} \left\{\text{declare no feasible system exists or select } i \text{ such that } i \in \cap_{\ell=1}^{s} \left(D_\ell\left(q_\ell^{(\theta)}\right) \cup A_\ell\left(q_\ell^{(\theta)}\right)\right)\right\}, & \text{if } \theta < \theta^*; \\ \left\{\text{select } i \text{ such that } i \in \cap_{\ell=1}^{s} \left(D_\ell\left(q_\ell^{(\theta)}\right) \cup A_\ell\left(q_\ell^{(\theta)}\right)\right) \text{ and } x_i > x_{[b]} - \delta\right\}, & \text{if } \theta = \theta^*. \end{cases}$$

We let $\mathrm{CS}^{\mathrm{Restart}}$ be the correct selection event of $\mathrm{Restart}^{\mathcal{HAK}^{\mathcal{R}}}$. As $\mathrm{Restart}^{\mathcal{HAK}^{\mathcal{R}}}$ iteratively applies $\mathcal{HAK}^{\mathcal{R}}$ for threshold vectors $\mathbf{q}^{(1)}, \mathbf{q}^{(2)}, \ldots, \mathbf{q}^{(\theta^*)}$ when $1 \le \theta^* \le d$ and for threshold vectors $\mathbf{q}^{(1)}, \mathbf{q}^{(2)}, \ldots, \mathbf{q}^{(d)}$ when $\theta^* = d + 1$, we have $\cap_{\theta=1}^{\min\{\theta^*, d\}} \mathrm{CS}^{(\theta)} \subset \mathrm{CS}^{\mathrm{Restart}}$.

Before we prove the statistical validity of $\mathrm{Restart}^{\mathcal{HAK}^{\mathcal{R}}}$, we first introduce the following notation:

$$S_a^{(\theta)} = \text{ set of acceptable systems with respect to threshold vector } \mathbf{q}^{(\theta)};$$

$$S_u^{(\theta)} = \text{ set of unacceptable systems with respect to threshold vector } \mathbf{q}^{(\theta)}.$$

Note that there do not exist desirable systems with respect to $\mathbf{q}^{(\theta)}$ when $\theta < \theta^*$. We then let

$$S_d^{(\theta^*)} = \begin{cases} \text{set of desirable systems with respect to } \mathbf{q}^{(\theta^*)} \text{ among systems in } \Gamma \setminus \{[b]\}, & \text{if } \theta^* \le d; \\ \emptyset, & \text{if } \theta^* = d+1, \end{cases}$$

and let $\mathrm{CS}_i^{(\theta^*)}$ be the correct selection event between system $i \in S_a^{(\theta^*)} \cup S_d^{(\theta^*)}$ and the best system $[b]$.

We then present two lemmas that we use to prove the statistical validity of $\mathrm{Restart}^{\mathcal{HAK}^{\mathcal{R}}}$.

Lemma F.1. *Under Assumption 1, for system i and constraint ℓ with threshold q_ℓ, the* **[Feasibility Check]** *steps in $\mathcal{HAK}^{\mathcal{R}}$ that run to completion ensure* $\Pr(\mathrm{CD}_{i\ell}(q_\ell)) \ge 1 - \beta_1$.

Lemma F.2. *Under Assumption 1, given i such that $x_i \le x_{[b]} - \delta$, the* **[Comparison]** *steps for system i and $[b]$ in $\mathcal{HAK}^{\mathcal{R}}$ that run to completion ensure*

$$\Pr\left(\mathrm{CS}_i^{(\theta^*)}\right) \ge 1 - \beta_2.$$

The proofs of Lemmas F.1 and F.2 are essentially same as those of Lemmas A.1 and A.3 when $c = 1$ (the case considered by [8]) because α'_f (α'_c) from $\mathcal{ZAK}^{\mathcal{R}}$ and β_1 (β_2) from $\mathcal{HAK}^{\mathcal{R}}$ both denote the nominal error of feasibility check for one constraint of one system with a fixed threshold (comparison between an inferior system and the best system $[b]$). We prove the statistical validity of $\mathrm{Restart}^{\mathcal{HAK}^{\mathcal{R}}}$ in the following theorem.

Theorem F.3. *Under Assumptions 1 and 2, the procedure* $\mathrm{Restart}^{\mathcal{HAK}^{\mathcal{R}}}$ *guarantees*

$$\Pr\{\mathrm{CS}^{\mathrm{Restart}}\} \ge 1 - \alpha.$$

Proof. We consider two cases, namely when $\theta^* \le d$ and $\theta^* = d + 1$.

Case 1: $\theta^* \le d$.

Recall from Section A that $\mathcal{B}_1^*$ denotes the event that system $[b]$ is declared feasible to $\mathbf{q}^{(\theta^*)}$. Similar to $\mathcal{B}_2^*$ and $\mathcal{A}_1^*$ in the proof of Theorem A.4, we define $\tilde{\mathcal{B}}_2^*$ as the event that selects the best system $[b]$ among the systems in $S_d^{(\theta^*)} \cup S_a^{(\theta^*)}$ and

$$\mathcal{A}^{(\theta)} = \left\{ \text{all systems in } S_u^{(\theta)} \text{ are declared infeasible with respect to } \mathbf{q}^{(\theta)}, \text{ where } \theta = 1, \ldots, d \right\}.$$

Note that when $\theta < \theta^*$, $\text{CS}^{(\theta)}$ can be ensured by only guaranteeing $\mathcal{A}^{(\theta)}$. When $\theta = \theta^*$, $\text{CS}^{(\theta)} \subseteq \mathcal{A}^{(\theta)} \cap \mathcal{B}_1^* \cap \tilde{\mathcal{B}}_2^*$. Thus,

$$\Pr\left(\text{CS}^{(\theta)}\right) \geq \begin{cases} \Pr\left(\mathcal{A}^{(\theta)} \cap \mathcal{B}_1^* \cap \tilde{\mathcal{B}}_2^*\right), & \text{if } \theta = \theta^*, \\ \Pr\left(\mathcal{A}^{(\theta)}\right), & \text{if } \theta < \theta^*. \end{cases}$$

As $\text{CS}^{(\theta)}$ achieves its lower bound when $\theta = \theta^*$ (because otherwise there is no need to make correct comparison decisions), we focus on this case. One may also notice that $\mathcal{A}^{(\theta^*)}$ and $\mathcal{B}_1^*$ are independent if systems are simulated independently and are dependent if systems are simulated under CRN. As we discard observations from Phase I and completely restart for Phase II in $\mathcal{HAK}^{\mathcal{R}}$, and $\tilde{\mathcal{B}}_2^*$ involves making the correct selection from all systems in $S_a^{(\theta^*)} \cup S_d^{(\theta^*)}$, $\tilde{\mathcal{B}}_2^*$ is independent from $\mathcal{A}^{(\theta^*)}$ and $\mathcal{B}_1^*$. Then, we have

$$\Pr\left(\text{CS}^{(\theta^*)}\right) \geq \begin{cases} \Pr\left(\mathcal{A}^{(\theta^*)}\right) \times \Pr\left(\mathcal{B}_1^*\right) \times \Pr\left(\tilde{\mathcal{B}}_2^*\right), & \text{if systems are simulated independently,} \\ \left[\Pr\left(\mathcal{A}^{(\theta^*)}\right) + \Pr\left(\mathcal{B}_1^*\right) - 1\right] \times \Pr\left(\tilde{\mathcal{B}}_2^*\right), & \text{if systems are simulated under CRN.} \end{cases}$$

We let $j_u^{(\theta)}$ denote the number of unacceptable systems with respect to $\mathbf{q}^{(\theta)}$, that is, $j_u^{(\theta)} = |S_u^{(\theta)}|$. We then discuss the cases depending on whether systems are simulated independently or under CRN.

When systems are simulated independently, by Lemma F.1 and the Bonferroni inequality, we have

$$\begin{aligned} \Pr\left(\mathcal{A}^{(\theta^*)}\right) &\geq \Pr\left(\cap_{i \in S_u^{(\theta^*)}} \cap_{\ell=1}^{s} \text{CD}_{i\ell}(q_\ell^{(\theta^*)})\right) = \prod_{i \in S_u^{(\theta^*)}} \Pr\left(\cap_{\ell=1}^{s} \text{CD}_{i\ell}(q_\ell^{(\theta^*)})\right) \\ &\geq \prod_{i \in S_u^{(\theta^*)}} \left[1 - \sum_{\ell=1}^{s} \Pr\left(\text{ICD}_{i\ell}(q_\ell^{(\theta^*)})\right)\right] \geq (1 - s\beta_1)^{j_u^{(\theta^*)}}; \end{aligned}$$

$$\Pr\left(\mathcal{B}_1^*\right) = \Pr\left(\cap_{\ell=1}^{s} \text{CD}_{[b]\ell}(q_\ell^{(\theta^*)})\right) \geq 1 - \sum_{\ell=1}^{s} \Pr\left(\text{ICD}_{[b]\ell}(q_\ell^{(\theta^*)})\right) \geq 1 - s\beta_1.$$

We use a similar approach as in Equation (4) from the proof of Theorem A.4 by replacing $S_{a'}$ and S_d with $S_a^{(\theta^*)}$ and $S_d^{(\theta^*)}$, respectively. We then have

$$\Pr(\tilde{\mathcal{B}}_2^*) \geq (1 - \beta_2)^{k - j_u^{(\theta^*)} - 1}.$$

Thus, we have

$$\Pr\left(\text{CS}^{(\theta^*)}\right) \geq (1 - s\beta_1)^{j_u^{(\theta^*)} + 1} \times (1 - \beta_2)^{k - j_u^{(\theta^*)} - 1}.$$

To find a lower bound of the above expression, we need to either maximize $j_u^{(\theta^*)}$ if $1 - s\beta_1 \leq 1 - \beta_2$ or minimize $j_u^{(\theta^*)}$ if $1 - s\beta_1 > 1 - \beta_2$. We also know that $0 \leq j_u^{(\theta^*)} \leq k - 1$. When $1 - s\beta_1 \leq 1 - \beta_2$, we have

$$\begin{aligned} (1 - s\beta_1)^{j_u^{(\theta^*)} + 1} \times (1 - \beta_2)^{k - j_u^{(\theta^*)} - 1} &\geq (1 - s\beta_1)^{(k-1)+1} \times (1 - \beta_2)^{k-(k-1)-1} \\ &= (1 - s\beta_1)^k = 1 - \alpha_1, \end{aligned}$$

where the last equality holds since procedure $\mathcal{HAK}$ sets $\beta_1 = (1 - (1 - \alpha_1)^{1/k})/s$ when systems are independent. When $1 - s\beta_1 > 1 - \beta_2$, we have

$$(1 - s\beta_1)^{j_u^{(\theta^*)} + 1} \times (1 - \beta_2)^{k - j_u^{(\theta^*)} - 1} \geq (1 - s\beta_1)^{0+1} \times (1 - \beta_2)^{k-0-1}$$

$$= (1 - s\beta_1) \times (1 - \beta_2)^{k-1}$$
$$= (1 - \alpha_1)^{1/k} \times (1 - \alpha_2)$$
$$= (1 - \alpha_1)^{(k+1)/k},$$

where the second equality holds as $\mathcal{HAK}$ sets $\beta_1 = (1 - (1 - \alpha_1)^{1/k})/s$ and $\mathcal{HAK}^{\mathcal{R}}$ sets $\beta_2 = 1 - (1 - \alpha_2)^{1/(k-1)}$ when systems are independent. Therefore, we have

$$\Pr\left(\mathrm{CS}^{(\theta^*)}\right) \geq \min\left[1 - \alpha_1, (1 - \alpha_1)^{(k+1)/k}\right] = (1 - \alpha_1)^{(k+1)/k}$$
$$= \left[1 - (1 - (1 - \alpha')^{k/(k+1)})\right]^{(k+1)/k} = 1 - \alpha'.$$

When systems are simulated under CRN, by Lemma F.1 and the Bonferroni inequality, we have

$$\Pr\left(\mathcal{A}^{(\theta^*)}\right) \geq \Pr\left(\cap_{i \in S_u^{(\theta^*)}} \cap_{\ell=1}^{s} \mathrm{CD}_{i\ell}(q_\ell^{(\theta^*)})\right) \geq 1 - \sum_{i \in S_u^{(\theta^*)}} \sum_{\ell=1}^{s} \mathrm{CD}_{i\ell}(q_\ell^{(\theta^*)}) \geq 1 - j_u^{(\theta^*)} s\beta_1;$$
$$\Pr\left(\mathcal{B}_1^*\right) \geq 1 - s\beta_1;$$
$$\Pr\left(\tilde{\mathcal{B}}_2^*\right) \geq \Pr\left(\cap_{i \in \left(S_a^{(\theta^*)} \cup S_d^{(\theta^*)}\right)} \mathrm{CS}_i^{(\theta^*)}\right) \geq 1 - \sum_{i \in \left(S_a^{(\theta^*)} \cup S_d^{(\theta^*)}\right)} \Pr\left(\mathrm{ICS}_i\right) \geq 1 - (k - j_u^{(\theta^*)} - 1)\beta_2.$$

Thus, we have

$$\Pr\left(\mathrm{CS}^{(\theta^*)}\right) \geq \left[1 - (j_u^{(\theta^*)} + 1)s\beta_1\right]\left[1 - (k - j_u^{(\theta^*)} - 1)\beta_2\right].$$

To find a lower bound of $\left[1 - (j_u^{(\theta^*)} + 1)s\beta_1\right]\left[1 - (k - j_u^{(\theta^*)} - 1)\beta_2\right]$, we see that

$$\left[1 - (j_u^{(\theta^*)} + 1)s\beta_1\right]\left[1 - (k - j_u^{(\theta^*)} - 1)\beta_2\right]$$
$$= -s\beta_1\beta_2 \times (j_u^{(\theta^*)})^2 + \left[(k-2)s\beta_1\beta_2 - s\beta_1 + \beta_2\right] \times j_u^{(\theta^*)} + (1 - s\beta_1)\left[1 - (k-1)\beta_2\right].$$

Given that $0 \leq j_u^{(\theta^*)} \leq k - 1$, we see that the above quadratic function achieves its minimum either when $j_u^{(\theta^*)} = 0$ or $j_u^{(\theta^*)} = k - 1$. When $j_u^{(\theta^*)} = 0$, we have

$$\left[1 - (j_u^{(\theta^*)} + 1)s\beta_1\right]\left[1 - (k - j_u^{(\theta^*)} - 1)\beta_2\right] = (1 - s\beta_1)(1 - (k-1)\beta_2)$$
$$= (1 - \alpha_1/k)(1 - \alpha_2)$$
$$= (1 - \alpha_1/k)(1 - \alpha_1),$$

where the second equality holds since procedure $\mathcal{HAK}$ sets $\beta_1 = \alpha_1/(ks)$ and $\mathcal{HAK}^{\mathcal{R}}$ sets $\beta_2 = \alpha_2/(k-1)$ when systems are correlated. When $j_u^{(\theta^*)} = k - 1$, we have

$$\left[1 - (j_u^{(\theta^*)} + 1)s\beta_1\right]\left[1 - (k - j_u^{(\theta^*)} - 1)\beta_2\right] = (1 - ks\beta_1) = 1 - \alpha_1,$$

where the second equality holds since $\mathcal{HAK}$ sets $\beta_1 = \alpha_1/(ks)$ when systems are correlated. Therefore, we have

$$\Pr\left(\mathrm{CS}^{(\theta^*)}\right) \geq \min\left[1 - \alpha_1, (1 - \alpha_1/k)(1 - \alpha_1)\right]$$
$$= (1 - \alpha_1/k)(1 - \alpha_1) = \frac{1}{k}\alpha_1^2 - \frac{k+1}{k}\alpha_1 + 1$$
$$= \frac{1}{k}\left[\frac{1}{2}\left(k + 1 - \sqrt{(k+1)^2 - 4k\alpha'}\right)\right]^2 - \frac{k+1}{2k}\left(k + 1 - \sqrt{(k+1)^2 - 4k\alpha'}\right) + 1$$

$$= 1 - \alpha'.$$

Note that although setting $\alpha_1 = \frac{1}{2}\left(k + 1 + \sqrt{(k+1)^2 - 4k\alpha'}\right)$ also yields $\Pr(\mathrm{CS}^{(\theta^*)}) \geq 1 - \alpha'$, it is not valid. This is because $\frac{1}{2}\left(k + 1 + \sqrt{(k+1)^2 - 4k\alpha'}\right) > \frac{1}{2}\left(k + 1 + \sqrt{(k+1)^2 - 4k}\right) = k \geq 1$ (as $0 < \alpha' < 1$) and hence selecting α_1 in this manner violates the fact that $0 < \alpha_1 < 1$.

Thus, we see that $\Pr\left(\mathrm{CS}^{(\theta)}\right) \geq \Pr\left(\mathrm{CS}^{(\theta^*)}\right) \geq 1 - \alpha'$ regardless whether systems are simulated independently or under CRN. Therefore, we have

$$\Pr\{\mathrm{CS}^{\text{Restart}}\} \geq \Pr\{\cap_{\theta=1}^{\theta^*} \mathrm{CS}^{(\theta)}\} \geq \Pr\{\cap_{\theta=1}^{d} \mathrm{CS}^{(\theta)}\} = \prod_{\theta=1}^{d} \Pr\left(\mathrm{CS}^{(\theta)}\right)$$
$$\geq (1 - \alpha')^d = (1 - (1 - (1 - \alpha)^{1/d}))^d = 1 - \alpha.$$

Case 2: $\theta^* = d + 1$.
If $\theta^* = d + 1$, there are no desirable systems for any threshold vector. Therefore, $\mathrm{CS}^{(\theta)}$ is ensured by correctly concluding feasibility decisions for all systems $i \in S_u^{(\theta)}$. Then $\Pr(\mathrm{CS}^{(\theta)}) \geq \Pr(\mathcal{A}^{(\theta)})$ and Lemma F.1 and the Bonferroni inequality yields

$$\Pr(\mathrm{CS}^{(\theta)}) \geq \begin{cases} (1 - s\beta_1)^{j_u^{(\theta)}}, & \text{if systems are simulated independently,} \\ 1 - j_u^{(\theta)} s\beta_1, & \text{if systems are simulated under CRN,} \end{cases}$$
$$\geq \begin{cases} (1 - s\beta_1)^{k}, & \text{if systems are simulated independently,} \\ 1 - ks\beta_1, & \text{if systems are simulated under CRN,} \end{cases}$$

where the last inequality is due to the fact that $0 \leq j_u^{(\theta)} \leq k$ for any $\theta = 1, \ldots, d$. When systems are simulated independently, we have

$$\Pr\left\{\mathrm{CS}^{(\theta)}\right\} \geq (1 - s\beta_1)^k = 1 - \alpha_1 > 1 - \alpha'.$$

When systems are simulated under CRN, we have

$$\Pr\left\{\mathrm{CS}^{(\theta)}\right\} \geq (1 - ks\beta_1) = 1 - \alpha_1 > 1 - \alpha'.$$

Thus, we have $\Pr\left(\mathrm{CS}^{(\theta)}\right) \geq 1 - \alpha'$ regardless whether systems are simulated independently or under CRN. Then it follows that

$$\Pr\left\{\mathrm{CS}^{\text{Restart}}\right\} \geq \Pr\{\cap_{\theta=1}^{d} \mathrm{CS}^{(\theta)}\} = \prod_{\theta=1}^{d} \Pr\left(\mathrm{CS}^{(\theta)}\right) \geq (1 - \alpha')^d = 1 - \alpha. \qquad \square$$

Remark 3. There are two potential improvement for Restart$^{\mathcal{HAK}^{\mathcal{R}}}$ in terms of setting the implement parameters:

(1) The proof of Theorem F.3 computes $\Pr(\mathcal{A}^{(\theta^*)}) \geq (1 - s\beta_1)^{j_u^{(\theta^*)}}$ when systems are simulated independently and $\Pr(\mathcal{A}^{(\theta^*)}) \geq 1 - j_u^{(\theta^*)} s\beta_1$ when systems are simulated under CRN, which is consistent with the choice of implementation parameters in Procedure $\mathcal{HAK}$ in Healey et al. [8]. However, these bounds can be improved using ideas in this article. In particular, similar to the argument in the proof of Lemma 2, for each system $i \in S_u^{(\theta^*)}$, let ℓ_i be a constraint such that system i is infeasible to threshold vector $q_{\ell_i}^{(\theta^*)}$. To declare system i infeasible to threshold vector $\mathbf{q}^{(\theta^*)}$, it is sufficient to make a correct feasibility decision for constraint ℓ_i

with respect to threshold $q_{\ell_i}^{(\theta^*)}$. Therefore, one may improve the efficiency of Restart$^{\mathcal{HAK}^{\mathcal{R}}}$ by computing $\Pr(\mathcal{A}^{(\theta^*)})$ as

$$\Pr\left(\mathcal{A}^{(\theta^*)}\right) \geq \Pr\left(\cap_{i\in S_u^{(\theta^*)}} \mathrm{CD}_{i\ell_i}(q_{\ell_i}^{(\theta^*)})\right) = \prod_{i\in S_u^{(\theta^*)}} \Pr\left(\mathrm{CD}_{i\ell_i}(q_{\ell_i}^{(\theta^*)})\right)$$
$$= \prod_{i\in S_u^{(\theta^*)}} \left[1 - \Pr\left(\mathrm{ICD}_{i\ell_i}(q_{\ell_i}^{(\theta^*)})\right)\right] \geq (1-\beta_1)^{j_u^{(\theta^*)}}.$$

when systems are simulated independently, and

$$\Pr\left(\mathcal{A}^{(\theta^*)}\right) \geq \Pr\left(\cap_{i\in S_u^{(\theta^*)}} \mathrm{CD}_{i\ell_i}(q_{\ell_i}^{(\theta^*)})\right) \geq 1 - \sum_{i\in S_u^{(\theta^*)}} \Pr\left(\mathrm{ICD}_{i\ell_i}(q_{\ell_i}^{(\theta^*)})\right) \geq 1 - j_u^{(\theta^*)}\beta_1.$$

when systems are simulated under CRN.

(2) The proof of Theorem F.3 allocates error to both Phases I and II in order to achieve $\mathrm{CS}^{(\theta)}$ for all $\theta = 1, \ldots, \theta^*$. One may improve the efficiency of Restart$^{\mathcal{HAK}^{\mathcal{R}}}$ by not allocating error to Phase II when $\theta < \theta^*$ (since there are no feasible systems exists with respect to $\mathbf{q}^{(\theta)}$ when $\theta < \theta^*$).

As the current approach is a natural and statistical valid way of restarting $\mathcal{HAK}$ for different threshold vectors, we do not consider an improved version of Restart$^{\mathcal{HAK}^{\mathcal{R}}}$ since this is not the main focus of the article.

As Restart$^{\mathcal{HAK}}$ reuses the observations from Phase I and assigns the error in Phase II more efficiently, it is expected to perform better than Restart$^{\mathcal{HAK}^{\mathcal{R}}}$. Although we do not prove the statistical validity of Restart$^{\mathcal{HAK}}$, we have not found any experiments that violate the statistical guarantee. We believe that Restart$^{\mathcal{HAK}^{\mathcal{R}}}$ and Restart$^{\mathcal{HAK}}$ are appropriate choices of sequentially-running approaches for comparison with $\mathcal{ZAK}^{\mathcal{R}}$ and $\mathcal{ZAK}$, respectively.

G PROCEDURES Restart$^{\mathcal{AK}+}$ AND Restart$^{\mathcal{HAK}+}$

In this section, we discuss the algorithms Restart$^{\mathcal{AK}+}$ and Restart$^{\mathcal{HAK}+}$ and their statistical validity. Similar to Appendix F, as Restart$^{\mathcal{AK}+}$ is a special case of Restart$^{\mathcal{HAK}+}$ when the number of constraints is one, we omit a separate discussion of Restart$^{\mathcal{AK}+}$.

Restart$^{\mathcal{HAK}+}$ performs procedure $\mathcal{HAK}+$ due to [8] independently for the threshold vectors $\mathbf{q}^{(1)}, \mathbf{q}^{(2)}, \ldots, \mathbf{q}^{(\theta^*)}$ when $1 \leq \theta^* \leq d$, and for threshold vectors $\mathbf{q}^{(1)}, \mathbf{q}^{(2)}, \ldots, \mathbf{q}^{(d)}$ independently when $\theta^* = d+1$. As discussed in [8], $\mathcal{HAK}+$ requires user to choose a feasibility check procedure. In our experiments, we choose $\mathcal{F}_{\mathcal{B}}^{\mathcal{I}}$ in [8] as the feasibility check procedure. $\mathcal{HAK}+$ also requires a user‘s input for the ratio, namely $e = s\beta_1/\beta_2$, of the error for the feasibility checks and the comparison. We set $e = 1$ as recommended in [8] and the initial sample size when Restart$^{\mathcal{HAK}+}$ applies $\mathcal{HAK}+$ with respect to each threshold vector is set as $n_0 = 20$. Note that the procedure and the proof discussed in this section can be easily generalized to a different value of e. A detailed algorithm description is shown in Algorithm A.6.

We utilize the same notation of $S_u^{(\theta)}, j_u^{(\theta)}$, $\mathrm{CS}^{(\theta)}$, and $\mathrm{CS}^{\mathrm{Restart}}$ as in Appendix F, and prove the statistical validity of Restart$^{\mathcal{HAK}+}$ in the following theorem.

Theorem G.1. *Under Assumptions 1 and 2, the procedure* Restart$^{\mathcal{HAK}+}$ *guarantees*

$$\Pr\{\mathrm{CS}^{\mathrm{Restart}}\} \geq 1-\alpha.$$

Proof. We consider two cases, namely $\theta^* \leq d$ and $\theta^* = d+1$.

ALGORITHM A.6: Procedure Restart$^{\mathcal{HAK}+}$.

[**Setup**:] Select the overall nominal confidence level $1-\alpha$. Choose tolerance levels $\epsilon_1,\ldots,\epsilon_s$, indifference-zone parameter δ, and threshold vectors $\{\mathbf{q}^{(1)},\mathbf{q}^{(2)},\ldots,\mathbf{q}^{(d)}\}$. Choose procedure $\mathcal{F}_{\mathcal{B}}^{\mathcal{I}}$ as the feasibility check procedure and set $\alpha' = 1-(1-\alpha)^{1/d}$.

for $\theta = 1,\ldots,d$ **do**

[**Setup**] for $\mathcal{HAK}$+: Same as in $\mathcal{HAK}$+ except that α is replaced by α'. Note that we set β_2 to the solution of $\beta_2 + 2[1-(1-\beta_2)^{(k-1)/2}] = \alpha'$ when systems are simulated independently and set $\beta_2 = \alpha'/k$ when systems are simulated under CRN. We also set $\beta_1 = \beta_2/s$.

[**Initialization**], [**Feasibility Check**], [**Comparison**], and [**Stopping Rule**] are the same as in $\mathcal{HAK}$+.

[**Stopping Condition**:] If one system is found in [**Stopping Rule**], terminate the algorithm and select the system as the best. If no system is found in [**Stopping Rule**] and $\theta = d$, declare no feasible system exists with respect to the given threshold vectors.

end for

Case 1: $\theta^* \le d$.

When systems are simulated independently and Assumptions 1 and 2 hold, due to Lemmas F.1 and F.2 and the arguments in the proof of Theorem F.3, the feasibility check and comparison procedures of $\mathcal{HAK}$+ satisfy Assumptions 3 and 5 of [8], respectively. Thus, we are able to apply Lemma 4.2 of [8]. That is, we have

$$\Pr\left\{\mathrm{CS}^{(\theta)}\right\} \ge (1-s\beta_1)^{j_u^{(\theta)}} + (1-s\beta_1) + (1-\beta_2)^{k-j_u^{(\theta)}-1} - 2, \tag{6}$$

when $j_u^{(\theta)} < k$ and $\Pr(\mathrm{CS}^{(\theta)}) \ge (1-s\beta_1)^k$ when $j_u^{(\theta)} = k$. Also, Remark 4.3 of [8] discusses that the smallest lower bond on $\Pr\left\{\mathrm{CS}^{(\theta)}\right\}$ is always achieved when $j_u^{(\theta)} < k$. As we set $\beta_2 = s\beta_1$ and β_2 as the solution to $\beta_2 + 2[1-(1-\beta_2)^{(k-1)/2}] = \alpha'$, we know that

$$\begin{aligned}(1-s\beta_1)^{j_u^{(\theta)}} + (1-s\beta_1) + (1-\beta_2)^{k-j_u^{(\theta)}-1} - 2 &= (1-\beta_2)^{j_u^{(\theta)}} + (1-\beta_2) + (1-\beta_2)^{k-j_u^{(\theta)}-1} - 2\\ &\ge (1-\beta_2)^{(k-1)/2} + (1-\beta_2) + (1-\beta_2)^{(k-1)/2} - 2\\ &= 1 - \left(\beta_2 + 2\left[1-(1-\beta_2)^{(k-1)/2}\right]\right)\\ &= 1-\alpha',\end{aligned}$$

where the inequality holds as the lower bound is achieved when $j_u^{(\theta)} = (k-1)/2$. By Theorem 4.4 of [8], we know that $\Pr(\mathrm{CS}^{(\theta)}) \ge 1-\alpha'$.

When systems are simulated under CRN and Assumptions 1 and 2 hold, due to Lemmas F.1 and F.2 and the arguments in the proof of Theorem F.3, the feasibility check procedure and the comparison procedure of $\mathcal{HAK}$+ satisfy Assumptions 4 and 6. With Assumption 1, we apply Lemma 4.6 of [8] and have

$$\Pr\left\{\mathrm{CS}^{(\theta)}\right\} \ge 1 - (j_u^{(\theta)}+1)s\beta_1 - (k-j_u^{(\theta)}-1)\beta_2, \tag{7}$$

when $j_u^{(\theta)} < k$ and $\Pr\{\mathrm{CS}^{(\theta)}\} \ge 1-ks\beta_1$ when $j_u^{(\theta)} = k$. As we set $\beta_2 = s\beta_1 = \alpha'/k$, we know that

$$1 - (j_u^{(\theta)}+1)s\beta_1 - (k-j_u^{(\theta)}-1)\beta_2 = 1-k\beta_2 = 1-\alpha'.$$

Then by Theorem 4.8 of [8], we know that $\Pr(\mathrm{CS}^{(\theta)}) \ge 1-\alpha'$.

As we have $\Pr(\mathrm{CS}^{(\theta)}) \geq 1-\alpha'$ regardless of whether the systems are simulated independently or under CRN, we have

$$\begin{aligned}\Pr\{\mathrm{CS}^{\mathrm{Restart}}\} &\geq \Pr\left\{\cap_{\theta=1}^{\theta^*}\mathrm{CS}^{(\theta)}\right\} \geq \Pr\left\{\cap_{\theta=1}^{d}\mathrm{CS}^{(\theta)}\right\} = \prod_{\theta=1}^{d}\Pr\left(\mathrm{CS}^{(\theta)}\right)\\ &\geq (1-\alpha')^d = (1-(1-(1-\alpha)^{1/d}))^d = 1-\alpha.\end{aligned}$$

Case 2: $\theta^* = d+1$**.**
If $\theta^* = d+1$, there are no desirable systems for any threshold vector. This means that we have $j_u^{(\theta)} = k$ for any $\theta = 1,\ldots,d$. Similar to the proof of Theorem F.3, $\mathrm{CS}^{(\theta)}$ is ensured by correctly concluding feasibility decisions for all systems $i \in S_u^{(\theta)}$. By Lemmas 4.2 and 4.6 from [8], we have

$$\Pr\left(\mathrm{CS}^{(\theta)}\right) \geq \begin{cases}(1-s\beta_1)^k, & \text{if systems are simulated independently,}\\ 1-ks\beta_1, & \text{if systems are simulated under CRN.}\end{cases}$$

When systems are simulated independently, by Remark 4.3 of [8], the lower bound of $(1-s\beta_1)^k$ is never smaller than the **Right-Hand Side** (**RHS**) of Equation (6) when $j_u^{(\theta^*)} = k-1$. Therefore, we have $(1-s\beta_1)^k \geq 1-\alpha'$.

When systems are simulated under CRN, by Remark 4.7 of [8], the lower bound of $1-ks\beta_1$ is equal to the RHS of Equation (7) when $j_u^{(\theta^*)} = k-1$. Therefore, we have $1-ks\beta_1 \geq 1-\alpha'$.

Thus, we have $\Pr(\mathrm{CS}^{(\theta)}) \geq 1-\alpha'$ both when the systems are simulated independently or under CRN. It then follows that

$$\Pr\left\{\mathrm{CS}^{\mathrm{Restart}}\right\} \geq \Pr\{\cap_{\theta=1}^{d}\mathrm{CS}^{(\theta)}\} = \prod_{\theta=1}^{d}\Pr\left(\mathrm{CS}^{(\theta)}\right) \geq (1-\alpha')^d = 1-\alpha. \qquad \square$$

Remark 4. Similar as in Appendix F, there are two potential improvement for $\mathrm{Restart}^{\mathcal{HAK}+}$ in terms of setting the implementation parameters:

(1) Due a similar reason as in Remark 1, the computation of $\Pr(\mathrm{CS}^{(\theta)})$ in the proof of Theorem G.1 can be improved. When systems are simulated independently, Equation (6) can be improved as

$$\Pr\left\{\mathrm{CS}^{(\theta)}\right\} \geq (1-\beta_1)^{j_u^{(\theta)}} + (1-s\beta_1) + (1-\beta_2)^{k-j_u^{(\theta)}-1} - 2.$$

When systems are simulated under CRN, Equation (7) can be improved as

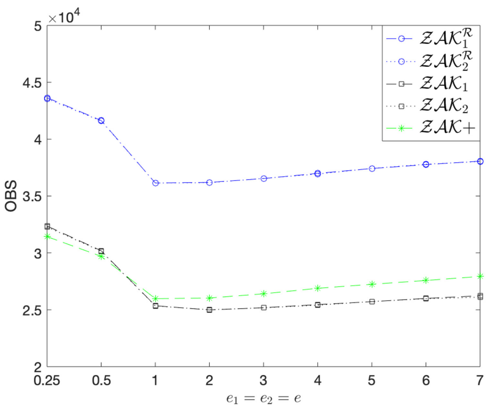
$$\Pr\left\{\mathrm{CS}^{(\theta)}\right\} \geq 1-(j_u^{(\theta)}+1)\beta_1 - (k-j_u^{(\theta)}-1)\beta_2.$$

(2) The proof of Theorem G.1 allocates error to both Phases I and II for all $\theta = 1,\ldots,\theta^*$. One may improve the efficiency of $\mathrm{Restart}^{\mathcal{HAK}+}$ by not allocating error to Phase II when $\theta < \theta^*$ (since there are no feasible systems exists with respect to $\mathbf{q}^{(\theta)}$ when $\theta < \theta^*$).

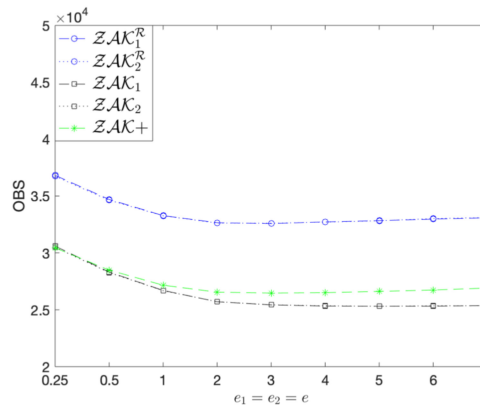
As the current setting is a natural and statistical valid way of restarting $\mathcal{HAK}+$ for different threshold vectors, we do not consider an improved version of $\mathrm{Restart}^{\mathcal{HAK}+}$ since this is not the main focus of the article.

Table A.1. Threshold Configuration for the four Constraints ($s = 4$) Case

Constraint	Threshold values of constraint ℓ
$\ell = 1$	$0, 2\epsilon_1, 4\epsilon_1, 6\epsilon_1$
$\ell = 2$	$0, 2\epsilon_2$
$\ell = 3$	$0, 2\epsilon_3, 4\epsilon_3$
$\ell = 4$	$0, 2\epsilon_4, 4\epsilon_4, 6\epsilon_4$



(a) OBS when $s = 2$ and $b = 50$



(b) OBS when $s = 4$ and $b = 25$

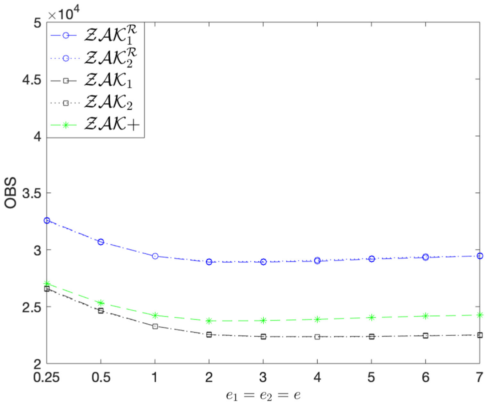
Fig. A.1. Average number of observations of procedures $\mathcal{ZAK}_1^{\mathcal{R}}, \mathcal{ZAK}_2^{\mathcal{R}}, \mathcal{ZAK}_1, \mathcal{ZAK}_2$, and $\mathcal{ZAK}+$ as functions of e_1, e_2, and e for $k = 100$ systems and $s = 2$ and 4 constraints.

H EXPERIMENTAL RESULTS FOR IMPLEMENTATION PARAMETERS

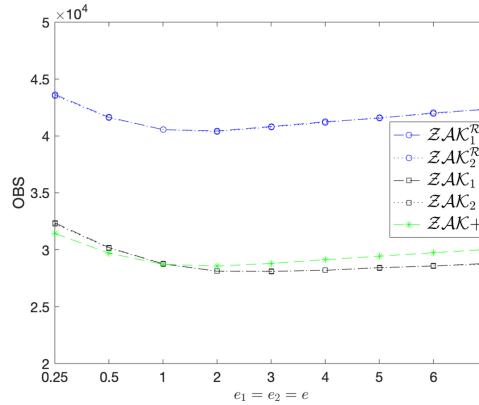
In this section, we present the experimental results that we use to choose the implementation parameters for the proposed procedures $\mathcal{ZAK}^{\mathcal{R}}$, $\mathcal{ZAK}$, and $\mathcal{ZAK}+$.

We test the performance of our proposed procedures in the DM mean configuration, the L/L variance configuration, and the ranked constraints preference order (where the constraints are ranked from constraint 1 to constraint s) when $k = 100$, $s = 2, 4, 6$, and $b = 25, 50$. When $s = 2$, both constraints have three thresholds $\{0, 2\epsilon_\ell, 4\epsilon_\ell\}$, for all $\ell = 1, 2$, and θ^* is set as $\theta^* = 5$. When $s = 4$, we consider the threshold values of each constraint shown in Table A.1 and $\theta^* = 50$. When $s = 6$, we let constraint ℓ have two thresholds $\{0, 2\epsilon_\ell\}$, where $\ell = 1, \ldots, 6$, and $\theta^* = 30$. The results of OBS when $s = 2$ and $b = 50$ and when $s = 4$ and $b = 25$ are shown in Figure A.1. Figures A.2(a) and A.2(b) show the experimental results for the case, where $s = 2$ and $b = 25$ and the case, where $s = 4$ and $b = 50$, respectively. The results for the six constraints case, where $b = 25$ and 50 are shown in Figures A.2(c) and A.2(d), respectively. Note that we fixed the ranges of $e_1, e_2, e \in \{0.25, 0.5, 1, 2, \ldots, 7\}$ and depict OBS on the scale $\{2, 2.5, \ldots, 5\} \times 10^4$ in all the figures to facilitate the comparison.

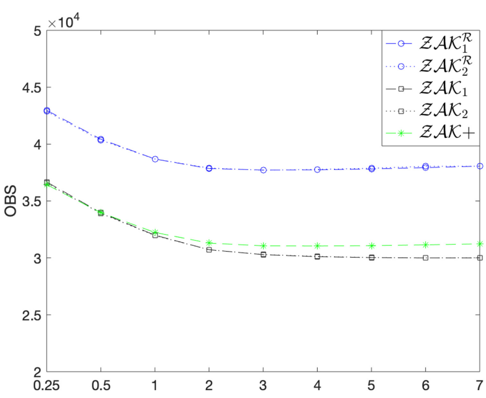
We see that for the four cases shown in Figure A.2, the values of e_1, e_2, and e where OBS achieves its minimum value ranges from 2 to 7 and the OBS is flat within this range. Note that the OBS is also similar between the two settings of the implementation parameters of $\mathcal{ZAK}^{\mathcal{R}}$ and $\mathcal{ZAK}$.



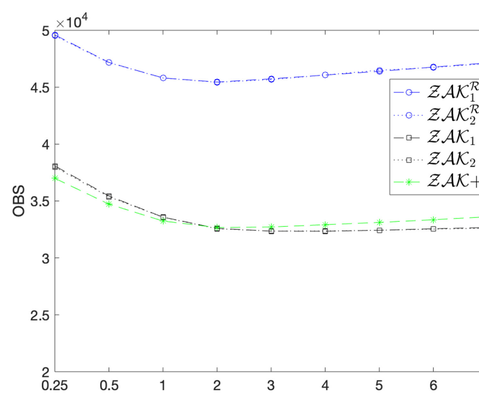
(a) OBS when $s = 2$ and $b = 25$



(b) OBS when $s = 4$ and $b = 50$



(c) OBS when $s = 6$ and $b = 25$



(d) OBS when $s = 6$ and $b = 50$

Fig. A.2. Average number of observations of procedures $\mathcal{ZAK}_1^{\mathcal{R}}$, $\mathcal{ZAK}_2^{\mathcal{R}}$, $\mathcal{ZAK}_1$, $\mathcal{ZAK}_2$, and $\mathcal{ZAK}+$ as functions of e_1, e_2, and e for $k = 100$ systems and $s = 2, 4, 6$ constraints.

I ADDITIONAL EXPERIMENTAL RESULTS FOR EFFICIENCY

In this section, we provide additional experimental results aimed at comparing the efficiency among all proposed procedures. Note that all the experimental results in this section are based on the L/L variance configuration.

Figure A.3 shows the OBS for a single constraint with ten thresholds under the MDM configuration (same experimental setting as in Figure 5 except for the mean configuration) for all four procedures $\mathcal{ZAK}$, Restart$^{\mathcal{AK}}$, $\mathcal{ZAK}+$, and Restart$^{\mathcal{AK}+}$. The pattern is similar when $1 \le \theta^* \le 10$ as in Figure 5(b) except that the benefit of $\mathcal{ZAK}+$ over $\mathcal{ZAK}$ is more substantial. When $\theta^* = 11$, $\mathcal{ZAK}+$ and Restart$^{\mathcal{ZAK}+}$ require more OBS than when $\theta^* = 10$. Since the problem is easier under the MDM configuration than with the MIM configuration for both $\mathcal{ZAK}+$ and Restart$^{\mathcal{ZAK}+}$ when $1 \le \theta^* \le 10$ and becomes the same when $\theta^* = 11$, this is expected. Both $\mathcal{ZAK}$ and $\mathcal{ZAK}+$ perform significantly better than the alternative procedures Restart$^{\mathcal{AK}}$ and Restart$^{\mathcal{AK}+}$.

Figures A.4, A.5, and A.6 show the OBS for two constraints with three thresholds on each constraint (same experimental setting as in Figures 6 and 7) for all four procedures

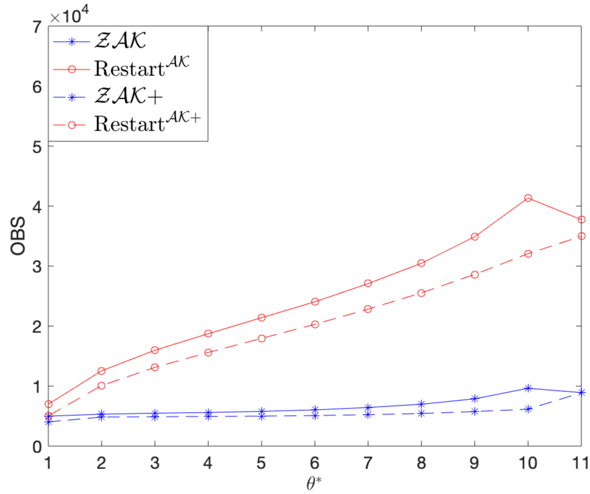


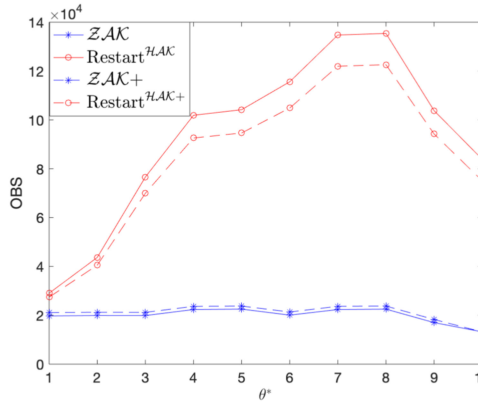
Fig. A.3. Average number of observations of $\mathcal{ZAK}$, Restart$^{\mathcal{AK}}$, $\mathcal{ZAK}$+, and Restart$^{\mathcal{AK}+}$ as functions of θ^* for $k = 100$ systems and $s = 1$ constraint with ten thresholds under the MDM configuration.

$\mathcal{ZAK}$, Restart$^{\mathcal{AK}}$, $\mathcal{ZAK}$+, and Restart$^{\mathcal{AK}+}$ under the ranked constraints, equally important constraints, and total violation with ranked constraints formulations, respectively. Each figure also contains the DM, MIM, and MDM configurations. As in the single constraint case, both $\mathcal{ZAK}$ and $\mathcal{ZAK}$+ show significant improvement compared with their competing procedures Restart$^{\mathcal{HAK}}$ and Restart$^{\mathcal{HAK}+}$ under all threshold formulations and all mean configurations. Note that the results of $\mathcal{ZAK}$ and $\mathcal{ZAK}$+ under the MIM and MDM configurations with the ranked constraints and equally important constraints formulations (Figures A.4(b), A.4(c), A.5(b), and A.5(c)) are the same as in Figures 6 and 7, but are shown on different scales because Restart$^{\mathcal{HAK}}$ and Restart$^{\mathcal{HAK}+}$ require much more observations than $\mathcal{ZAK}$ and $\mathcal{ZAK}$+.

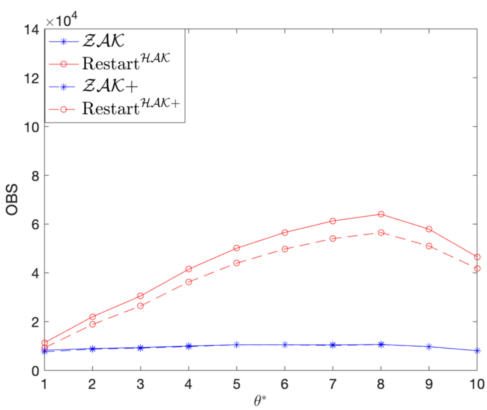
Finally, Figure A.7 shows the experimental results for two constraints with three thresholds on each constraint for procedures $\mathcal{ZAK}$ and $\mathcal{ZAK}$+ under the total violation with ranked constraints formulation and the MIM and MDM configurations (same setting as in Figures 6 and 7 except for the preference order). As discussed and explained in Section 6.4, the result shows a similar pattern as in Figure 6. We see that $\mathcal{ZAK}$+ performs slightly better or very similar to $\mathcal{ZAK}$ under the MIM configuration and performs significantly better than $\mathcal{ZAK}$ under the MDM configuration. Note that although the results for $\mathcal{ZAK}$ and $\mathcal{ZAK}$+ in Figures A.7(a) and A.7(b) are the same as in Figures A.6(b) and A.6(c), the scales of the plots are different due to the fact that Restart$^{\mathcal{HAK}}$ and Restart$^{\mathcal{HAK}+}$ require much more observations.

J EXPERIMENTAL RESULTS FOR THE IMPACT OF USING CRN

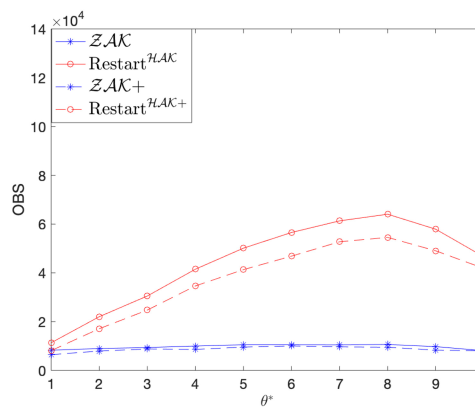
In this section, we discuss the impact of using CRN when applying the proposed procedures. To account for the dependency across systems induced by the use of CRN, the implementation parameters of both procedures take more conservative values than those with independent sampling. However, CRN often reduces the variance of the difference in the primary performance measures among systems. Thus, the feasibility check tends to require more observations while the comparison tends to require fewer observations. Whether CRN helps the overall performance of proposed procedures depends on how much savings we get in the comparison compared to the increment in observations in the feasibility check.



(a) OBS under DM



(b) OBS under MIM

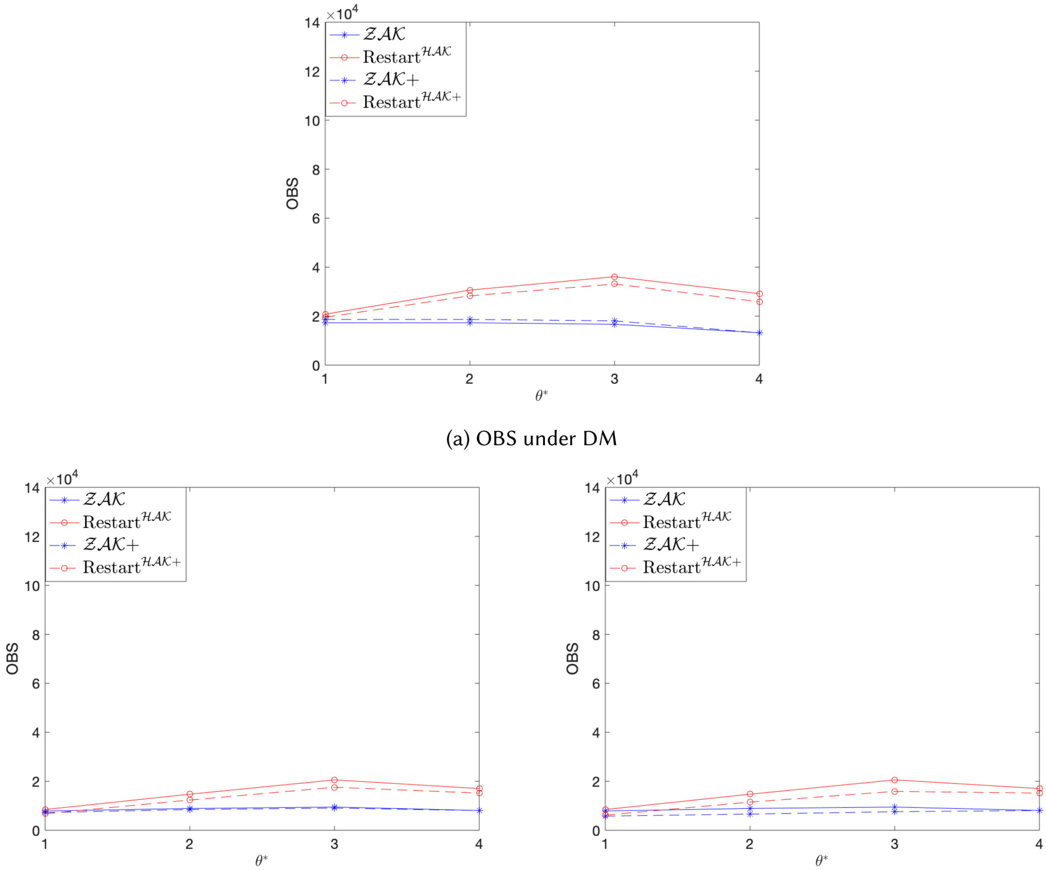


(c) OBS under MDM

Fig. A.4. Average number of observations of $\mathcal{ZAK}$, Restart$^{\mathcal{HAK}}$, $\mathcal{ZAK}$+, and Restart$^{\mathcal{HAK}+}$ as functions of θ^* for $k = 100$ systems and $s = 2$ constraints under the DM, MIM, and MDM configurations for the ranked constraints formulation.

We consider the case of a single constraint with two thresholds ($d = 2$) under the DM configuration and three different variance configurations (H/L, L/L, and L/H). Let ρ be the correlation between each pair of systems for the primary performance measure. Then the variance of the difference in the primary performance measure between two systems equals $2\sigma^2_{x_i}(1-\rho)$, while the variance of the secondary performance measure of each system is $\sigma^2_{y_{i\ell}}$. When systems are simulated independently (i.e., $\rho = 0$), the first two variance configurations (H/L and L/L) have more difficult comparison than feasibility check due to the larger value of $2\sigma^2_{x_i}$ than $\sigma^2_{y_{i\ell}}$. On the other hand, the L/H configuration has easier comparison than feasibility check. Thus, we expect the H/L and L/L variance configurations to show the benefit of CRN but not the L/H configuration. In our experiments, we consider $\rho \in \{0.25, 0.5, 0.75\}$ and all possible values of θ^* (i.e., $\theta^* \in \{1, 2, 3\}$), and fix $b = 25$. The results for the H/L, L/L, and L/H variance configurations are shown in Tables A.2, A.3, and A.4, respectively.

From Tables A.2 and A.3, we see that under the H/L and L/L variance configurations, $\mathcal{ZAK}$ and $\mathcal{ZAK}$+ both require fewer observations when CRN is applied with $\theta^* \in \{1, 2\}$ and $\rho \in$



(a) OBS under DM

(b) OBS under MIM

(c) OBS under MDM

Fig. A.5. Average number of observations of $\mathcal{ZAK}$, Restart$^{\mathcal{HAK}}$, $\mathcal{ZAK}$+, and Restart$^{\mathcal{HAK}+}$ as functions of θ^* for $k = 100$ systems and $s = 2$ constraints under the DM, MIM, and MDM configurations for the equally important constraints formulation.

$\{0.25, 0.5, 0.75\}$. As the variance of the pairwise comparison is reduced due to the CRN, the continuation region for comparison gets shorter and narrower and thus it takes fewer observations to complete the comparison among systems deemed feasible. Note that when $\theta^* = 3$, all systems are infeasible with respect to all threshold vectors considered, which means that the procedures are likely to be terminated by all systems deemed infeasible and there is no need to wait for the comparison decisions to be completed. Thus applying CRN does not help in this case. One may notice that the benefit of applying CRN is more obvious in Table A.2 than that in Table A.3. This is expected because the variance of the primary performance measure in the H/L configuration (Table A.2) is much larger than that in the L/L configuration (Table A.3). Therefore, reducing the variance of the pairwise comparison benefits the overall performance a lot more under the H/L configuration. We also see that for a fixed ρ, the performance of $\mathcal{ZAK}$ ($\mathcal{ZAK}$+) is similar under $\theta^* = 1$ or 2. This is expected as procedures $\mathcal{ZAK}$ and $\mathcal{ZAK}$+ are robust with respect to the values of θ^*. The OBS decreases when ρ increases for both $\mathcal{ZAK}$ and $\mathcal{ZAK}$+ when $\theta^* \in \{1, 2\}$. This is because higher correlation across systems

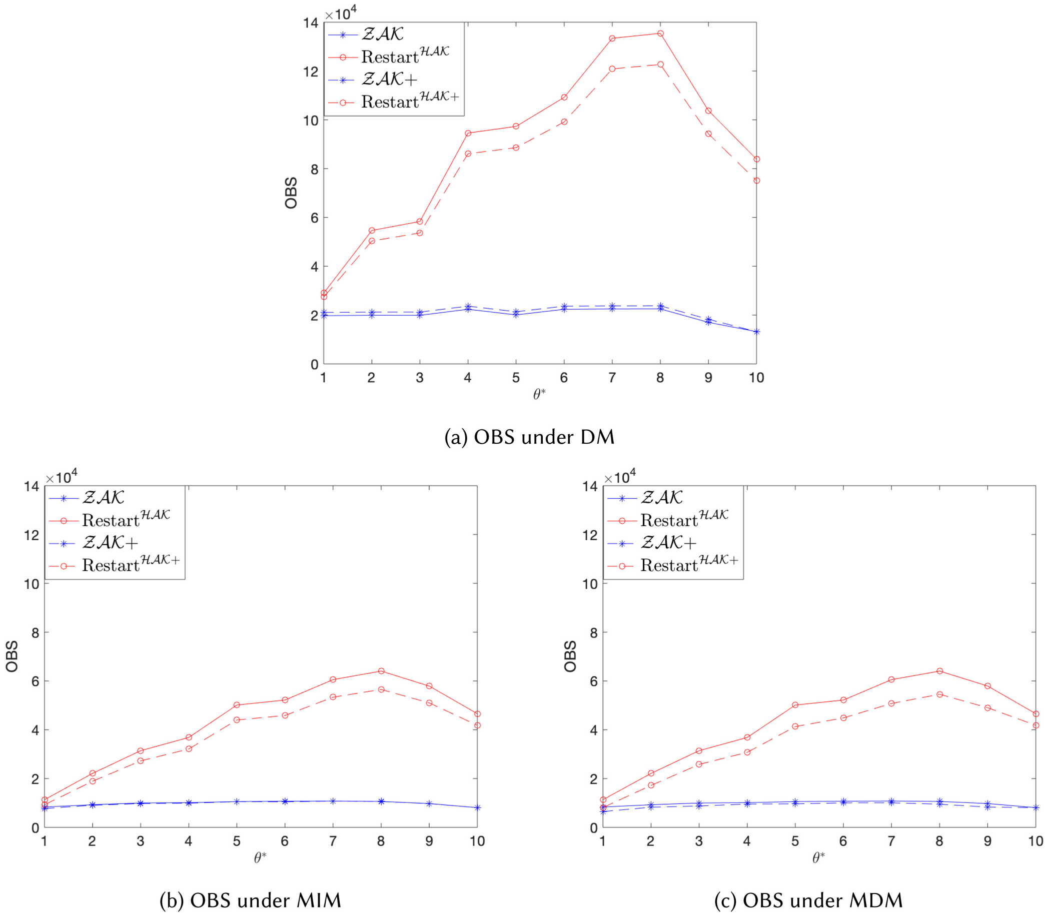
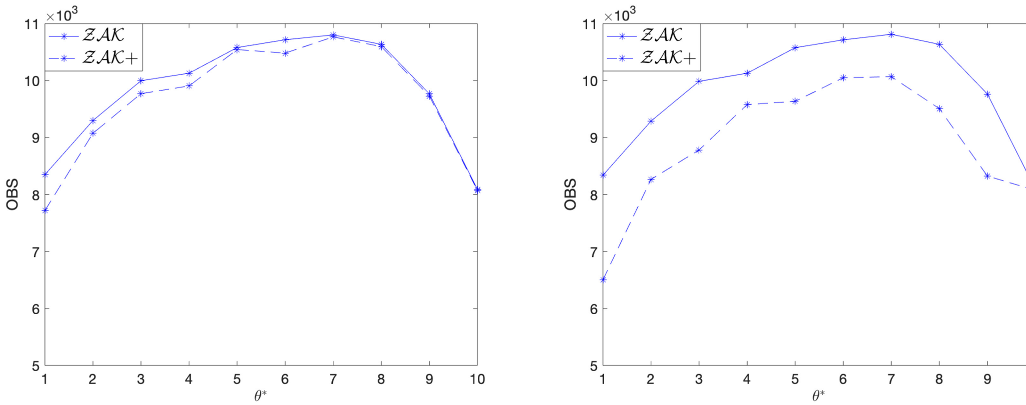


Fig. A.6. Average number of observations of $\mathcal{ZAK}$, Restart$^{\mathcal{HAK}}$, $\mathcal{ZAK}$+, and Restart$^{\mathcal{HAK}+}$ as functions of θ^* for $k = 100$ systems and $s = 2$ constraints under the DM, MIM, and MDM configurations for the total violation with ranked constraints formulation.

reduces the variance of the difference in the primary performance measures among systems, and thus both procedures become more efficient with larger ρ. When $\theta^* = 3$, however, as there are no feasible systems, reducing the variance of the difference in the primary performance measures among systems does not improve performance because no comparison is required to achieve CS.

Table A.4 shows the experimental results when the variance configuration is set to L/H. As the feasibility check is considered to be more difficult than the pairwise comparison, the benefit of CRN is expected to be smaller. Indeed, we do not see much savings in observations for both procedures. [8] discuss the required correlation to overcome the conservative Bonferroni bound required for the proof of the statistical validity of the proposed procedures under CRN. They show that the cross-correlation ρ needs to be sufficiently large to achieve a smaller number of observations under CRN than under independent sampling. When $\theta^* = 1$, our problem configuration becomes similar to that of [8] and we do see savings in observations for $\mathcal{ZAK}$+ (but not for $\mathcal{ZAK}$) when ρ is sufficiently large, which is consistent with the findings from [8]. When $\theta^* = 2, 3$, the benefit of



(a) OBS under MIM

(b) OBS under MDM

Fig. A.7. Average number of observations of $\mathcal{ZAK}$ and $\mathcal{ZAK}$+ as functions of θ^* for $k = 100$ systems and $s = 2$ constraints under the MIM, and MDM configurations for the total violation with ranked constraints formulation.

Table A.2. Average Number of Observations and Estimated PCS (Reported in Parentheses) of $\mathcal{ZAK}$ and $\mathcal{ZAK}$+ for $k = 100$ System and $s = 1$ Constraint with Two Thresholds Under the DM and H/L Configurations and $\rho \in \{0.25, 0.5, 0.75\}$

		With CRN		Without CRN	
	ρ	$\mathcal{ZAK}$	$\mathcal{ZAK}$+	$\mathcal{ZAK}$	$\mathcal{ZAK}$+
	0.25	37,610 (0.969)	47,509 (0.974)		
$\theta^* = 1$	0.5	32,316 (0.965)	40,273 (0.973)	39,429 (0.964)	49,674 (0.974)
	0.75	23,429 (0.953)	28,165 (0.972)		
	0.25	37,409 (0.960)	47,265 (0.968)		
$\theta^* = 2$	0.5	32,084 (0.955)	40,059 (0.967)	39,357 (0.960)	49,381 (0.965)
	0.75	23,351 (0.949)	28,041 (0.967)		
	0.25	15,015 (0.972)	14,896 (0.971)		
$\theta^* = 3$	0.5	15,020 (0.970)	14,888 (0.973)	14,986 (0.969)	14,814 (0.968)
	0.75	15,014 (0.972)	14,884 (0.973)		

CRN does not exist in this setting. When the feasibility check is more difficult than the pairwise comparison in the sense that it takes more observations to complete, it is possible that the use of CRN makes the overall performance worse than independent sampling. However, Table A.4 shows that the increment in observations does not seem significant.

Table A.3. Average Number of Observations and Estimated PCS (Reported in Parentheses) of $\mathcal{ZAK}$ and $\mathcal{ZAK}+$ for $k = 100$ System and $s = 1$ Constraint with Two Thresholds Under the DM and L/L Configurations and $\rho \in \{0.25, 0.5, 0.75\}$

		With CRN		Without CRN	
	ρ	$\mathcal{ZAK}$	$\mathcal{ZAK}+$	$\mathcal{ZAK}$	$\mathcal{ZAK}+$
$\theta^* = 1$	0.25	17,033 (0.966)	18,647 (0.975)	17,334 (0.967)	19,021 (0.974)
	0.5	16,202 (0.968)	17,305 (0.972)		
	0.75	15,221 (0.956)	15,306 (0.977)		
$\theta^* = 2$	0.25	17,212 (0.960)	18,675 (0.967)	17,462 (0.961)	19,043 (0.968)
	0.5	16,492 (0.958)	17,475 (0.967)		
	0.75	15,729 (0.956)	15,873 (0.968)		
$\theta^* = 3$	0.25	15,022 (0.973)	14,880 (0.971)	14,985 (0.970)	14,807 (0.971)
	0.5	15,023 (0.969)	14,885 (0.970)		
	0.75	15,014 (0.973)	14,875 (0.971)		

In summary, there is a trade-off between the required number of observations in the feasibility check and pairwise comparison when CRN is applied. CRN is unlikely to help when (i) the comparison is easier than the feasibility check or (ii) the induced correlation across systems for the primary performance measure is small. If the decision maker knows that the comparison is easier than the feasibility check or that the correlation is small, then it is better to use independent sampling. However, the decision maker may not have this information in practice. In that case, we recommend that the decision maker uses CRN because there is a possibility that CRN will reduce the number of observations significantly and, even when it does not, the number of observations with CRN appears to be similar to or only slightly larger than that with independent sampling.

Based on the results in Tables A.2, A.3, and A.4, we also observe that $\mathcal{ZAK}$ performs better than $\mathcal{ZAK}+$ when $\theta^* \in \{1, 2\}$ under the H/L and L/L configurations while $\mathcal{ZAK}+$ dominates $\mathcal{ZAK}$ when $\theta^* \in \{1, 2\}$ under the L/H configuration. Both $\mathcal{ZAK}$ and $\mathcal{ZAK}+$ perform similar when $\theta^* = 3$. This agrees with the finding from the single constraint with four thresholds case discussed in Section 6.4 (Table 2).

Table A.4. Average Number of Observations and Estimated PCS (Reported in Parentheses) of $\mathcal{ZAK}$ and $\mathcal{ZAK}$+ for $k = 100$ System and $s = 1$ Constraint with Two Thresholds Under the DM and L/H Configuration and $\rho \in \{0.25, 0.5, 0.75\}$

		With CRN		Without CRN	
	ρ	$\mathcal{ZAK}$	$\mathcal{ZAK}$+	$\mathcal{ZAK}$	$\mathcal{ZAK}$+
$\theta^* = 1$	0.25	74,008 (0.978)	69,321 (0.976)		
	0.5	73,930 (0.975)	68,283 (0.974)	73,842 (0.977)	69,288 (0.972)
	0.75	73,912 (0.979)	66,547 (0.975)		
$\theta^* = 2$	0.25	77,149 (0.969)	75,501 (0.967)		
	0.5	77,159 (0.969)	75,241 (0.967)	76,959 (0.967)	75,239 (0.966)
	0.75	77,176 (0.971)	74,959 (0.969)		
$\theta^* = 3$	0.25	74,484 (0.970)	73,521 (0.967)		
	0.5	74,506 (0.969)	73,528 (0.971)	74,339 (0.969)	73,266 (0.966)
	0.75	74,493 (0.970)	73,492 (0.968)		

ACKNOWLEDGMENTS

The second author was supported by NSF under grant CMMI-2127778.

REFERENCES

[1] Sigrún Andradóttir and Seong-Hee Kim. 2010. Fully sequential procedures for comparing constrained systems via simulation. *Naval Research Logistics* 57, 5 (2010), 403–421. https://onlinelibrary.wiley.com/doi/pdf/10.1002/nav.20408
[2] Sigrún Andradóttir and Judy S. Lee. 2021. Pareto set estimation with guaranteed probability of correct selection. *European Journal of Operational Research* 292, 1 (2021), 286–298.
[3] Demet Batur and Seong-Hee Kim. 2010. Finding feasible systems in the presence of constraints on multiple performance measures. *ACM Transactions on Modeling and Computer Simulation* 20, 3 (2010), Article 13 , 26 pages.
[4] John Butler, Douglas J. Morrice, and Peter W. Mullarkey. 2001. A multiple attribute utility theory approach to ranking and selection. *Management Science* 47, 6 (2001), 800–816.
[5] Guy Feldman, Susan R. Hunter, and Raghu Pasupathy. 2015. Multi-objective simulation optimization on finite sets: Optimal allocation via scalarization. In *Proceedings of the 2015 Winter Simulation Conference*. 3610–3621.
[6] Mingbin Feng and Jeremy Staum. 2017. Green simulation: Reusing the output of repeated experiments. *ACM Transactions on Modeling and Computer Simulation* 27, 4 (2017), Article 23 , 28 pages.
[7] Joshua Q. Hale, Helin Zhu, and Enlu Zhou. 2020. Domination measure: A new metric for solving multiobjective optimization. *INFORMS Journal on Computing* 32, 3 (2020), 565–581.
[8] Christopher Healey, Sigrún Andradóttir, and Seong-Hee Kim. 2014. Selection procedures for simulations with multiple constraints under independent and correlated sampling. *ACM Transactions on Modeling and Computer Simulation* 24, 3 (2014), Article 14 , 25 pages.
[9] Christopher M. Healey, Sigrún Andradóttir, and Seong-Hee Kim. 2013. Efficient comparison of constrained systems using dormancy. *European Journal of Operational Research* 224, 2 (2013), 340–352.
[10] L. Jeff Hong, Barry L. Nelson, and Jie Xu. 2015. Discrete optimization via simulation. In *Handbook of Simulation Optimization*, Michael C Fu (Ed.). Springer, New York, NY, 9–44.

[11] Susan R. Hunter and Raghu Pasupathy. 2013. Optimal sampling laws for stochastically constrained simulation optimization on finite sets. *INFORMS Journal on Computing* 25, 3 (2013), 527–542. https://doi.org/10.1287/ijoc.1120.0519
[12] Seong-Hee Kim and Barry L. Nelson. 2001. A fully sequential procedure for indifference-zone selection in simulation. *ACM Transactions on Modeling and Computer Simulation* 11, 3 (2001), 251–273.
[13] Seong-Hee Kim and Barry L. Nelson. 2006. Selecting the best system: Simulation. In *Handbooks in Operations Research and Management Science*, Shane G. Henderson and Barry L. Nelson (Eds.). Elsevier, 501–534.
[14] Lloyd W. Koenig and Averill M. Law. 1985. A procedure for selecting a subset of size m containing the l best of k independent normal populations, with applications to simulation. *Communications in Statistics - Simulation and Computation* 14, 3 (1985), 719–734. https://doi.org/10.1080/03610918508812467
[15] Averill M. Law and David M. Kelton. 2000. *Simulation Modeling and Analysis* (third ed.). Academic Press.
[16] Loo Hay Lee, Ek Peng Chew, Suyan Teng, and David Goldsman. 2010. Finding the non-dominated pareto set for multi-objective simulation models. *IIE Transactions* 42, 9 (2010), 656–674.
[17] Loo Hay Lee, Nugroho Artadi Pujowidianto, Ling-Wei Li, Chun-Hung Chen, and Chee Meng Yap. 2012. Approximate simulation budget allocation for selecting the best design in the presence of stochastic constraints. *IEEE Transactions on Automatic Control* 57, 11 (2012), 2940–2945.
[18] Raghu Pasupathy, Susan R. Hunter, Nugroho A. Pujowidianto, Loo Hay Lee, and Chun-Hung Chen. 2014. Stochastically constrained ranking and selection via SCORE. *ACM Transactions on Modeling and Computer Simulation* 25, 1(2014), Article 1 , 26 pages.
[19] Juta Pichitlamken, Barry L. Nelson, and L. Jeff Hong. 2006. A sequential procedure for neighborhood selection-of-the-best in optimization via simulation. *European Journal of Operational Research* 173, 1 (2006), 283–298.
[20] Ajit C. Tamhane. 1977. Multiple comparisons in model I one-way ANOVA with unequal variances. *Communications in Statistics - Theory and Methods* 6, 1 (1977), 15–32. https://doi.org/10.1080/03610927708827466
[21] Jing Xie and Peter I. Frazier. 2013. Sequential bayes-optimal policies for multiple comparisons with a known standard. *Operations Research* 61, 5 (2013), 1174–1189. https://doi.org/10.1287/opre.2013.1207
[22] Yuwei Zhou, Sigrún Andradóttir, Seong-Hee Kim, and Chuljin Park. 2022. Finding feasible systems for subjective constraints using recycled observations. *INFORMS Journal on Computing* 34, 6 (2022), 3080–3095. https://doi.org/10.1287/ijoc.2022.1227

Received 29 March 2023; revised 7 April 2024; accepted 30 April 2024