ADVANCES IN RANKING AND SELECTION: VARIANCE ESTIMATION AND CONSTRAINTS

A Thesis
Presented to
The Academic Faculty

by

Christopher M. Healey

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
The H. Milton Stewart School of Industrial and Systems Engineering

Georgia Institute of Technology
August 2010

ADVANCES IN RANKING AND SELECTION: VARIANCE ESTIMATION AND CONSTRAINTS

Approved by:

Professor Seong-Hee Kim, Advisor
The H. Milton Stewart School of
Industrial and Systems Engineering
Georgia Institute of Technology

Professor Sigrún Andradóttir, Advisor
The H. Milton Stewart School of
Industrial and Systems Engineering
Georgia Institute of Technology

Professor David Goldsman, Advisor
The H. Milton Stewart School of
Industrial and Systems Engineering
Georgia Institute of Technology

Professor Shabbir Ahmed
The H. Milton Stewart School of
Industrial and Systems Engineering
Georgia Institute of Technology

Professor Brani Vidakovic
The Wallace H. Coulter Department
of Biomedical Engineering
Georgia Institute of Technology

Date Approved: 6 July 2010

DEDICATION

To my parents,

Dawn and Kevin.

ACKNOWLEDGEMENTS

I wish to thank my advisors, Dr. Seong-Hee Kim, Dr. Sigrun Andradóttir, and Dr. David Goldsman, for their insight, direction, and patience as this thesis was born and matured. Their guidance and consistent contact was invaluable and unforgettable.

I would also like to thank the members of my committee, Dr. Shabbir Ahmed and Dr. Brani Vidakovic. Their comments and criticism helped make this work strong and cohesive.

There are many others who served as an inspiration for this thesis and provided support for me during my time at Georgia Tech, including my close friends in the doctoral program, the faculty and staff of the H. Milton Stewart School of Industrial and Systems Engineering, and my friends and fellow competitors in the Georgia Tech running club. I appreciate everything they have done to help me,.

Finally, I cannot express enough gratitude for the support of my brother, parents, and grandparents, who made all of this possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

In this thesis, we first show that the performance of ranking and selection (R&S) procedures depends highly on the quality of the variance estimates that are used. We study the performance of R&S procedures using three variance estimators — overlapping area, overlapping Cramér–von Mises, and overlapping modified jackknifed Durbin–Watson estimators — that show better long-run performance than other estimators previously used in conjunction with R&S procedures for steady-state simulations. We devote additional study to the development of the new overlapping modified jackknifed Durbin–Watson estimator and demonstrate some of its useful properties.

Next, we consider the problem of finding the best simulated system under a primary performance measure, while also satisfying stochastic constraints on secondary performance measures, known as constrained ranking and selection. We first present a new framework that allows certain systems to become dormant, halting sampling for those systems as the procedure continues. Secondly, we develop general procedures for constrained R&S that guarantee a nominal probability of correct selection, under any number of constraints and correlation across systems. In addition, we address topics critical to efficiency of the these procedures, namely the allocation of error between feasibility check and selection, the use of common random numbers (CRN), and the setup costs incurred when switching between systems. The use of CRN within our procedures can result in degradation of the probability of correct selection, so we also provide several new variance estimates to address this issue.

CHAPTER I

INTRODUCTION

In ranking and selection (R&S), we are concerned with the selection of the best system out of a number, say k, of alternatives. In this context, the best system is commonly one that has either the largest or smallest expected value of a specific performance measure. We also require a certain probability of correct selection (PCS) to be achieved by our procedures.

In this thesis, we contribute to two fields within simulation analysis methodology, namely steady-state output analysis and ranking and selection. Since these fields are already well-established, we will only review literature closely related to the research of this thesis. Readers interested in a comprehensive backgrounds in steady-state output analysis and ranking and selection may consult the chapters of Alexopoulos and Seila [4] and Kim and Nelson [32], respectively.

Many R&S procedures have been developed assuming that the basic observations are independent and identically distributed (i.i.d.) normal random variates. Algorithms have been designed to determine the best system in simulations, for example, the indifference-zone (IZ) methods of Dudewicz and Dalal [19], Rinott [43], Kim and Nelson [30, 32], and Hong and Nelson [28], the optimal computing budget allocation (OCBA) approaches of Chen [14] and Chen et al. [15], and the Bayesian methods of Chick and Inoue [17, 18], and Chick [16].

Those R&S procedures can be used for steady-state simulation if the experimenter is willing to use as basic observations the within-replication averages from multiple replications (after deletion of initial, potentially biased, data) or the batch means from a single replication. However, Goldsman et al. [24] and Kim and Nelson [31] found that both approaches could diminish the efficiency of fully sequential R&S procedures in terms of the overall sample size requirements, and they proposed two procedures that take individual observations (such as consecutive wait times) as the basic observations from a single replication.

Given observations $\{X_i : i = 1, \ldots, n\}$ of a stationary stochastic output process, we can estimate the long-run average mean μ with the sample mean $\bar{X}_n \equiv \frac{1}{n}\sum_{i=1}^{n} X_i$, and we often compute an estimator of the quantity $\sigma^2 \equiv \lim_{n\to\infty} n\mathrm{Var}(\bar{X}_n)$, the variance parameter. Unfortunately, it is well known that the sample variance is inappropriate for use as an estimator in the current context, because the sample variance of stationary data can be severely biased for σ^2 when correlations exist among observations.

There have been several alternative estimators for σ^2 suggested in the literature, some employing methods such as nonoverlapping batch means, overlapping batch means, and standardized time series (STS) (Law and Kelton [33]). We are particularly interested in STS estimators—specifically, the area estimator (Goldsman et al. [26]), the Cramér von–Mises (CvM) estimator (Goldsman et al. [23]), and a combination of the two estimators called the modified jackknifed Durbin–Watson (MJDW) estimator (Batur et al. [6]).

Most selection procedures require estimates for the so-called variance parameters of the competitors, which are unknown in many simulation applications. For instance, the procedures due to Goldsman et al. [24] and Kim and Nelson [31] developed for steady–state simulation — called $\mathcal{R}+$, $\mathcal{KN}+$, and $\mathcal{KN}++$ — use well-known variance parameter estimators that happen to be asymptotically chi-squared distributed.

In this thesis, we investigate the use of overlapping STS variance estimates within steady-state R&S procedures and introduce our own variance estimator, called overlapping modified jackknifed Durbin–Watson (OM). Our study shows that the overlapping estimators, namely overlapping area and overlapping CvM estimators of Alexopolous et al. [3] and our new OM estimator, can provide considerable savings. We also show experimentally some useful properties of the OM estimator.

We next turn to the topic of constrained ranking and selection. There are many R&S procedures available to determine the best system out of a number of simulated alternatives, as stated earlier, but there are only a few that consider the added difficulty of satisfying one or more stochastic constraints. These constraints can be placed on any secondary performance measure, but as the performance measures must be estimated by outputs from a simulation, we cannot be certain whether a system satisfies them or not. The more complicated task of finding the best feasible system, which we call constrained R&S, will require more computational overhead, additional analysis, and possibly more observations than selecting a system according to just one performance measure.

There has been recent interest in multiple objective R&S and constrained R&S. A special case of this problem was introduced by Santner and Tamhane [44], namely, to find the best system under a constraint on the system's variance. Lee et al. [34, 35] and Chen and Lee [12] consider the multi-objective problem, namely the sequential selection of a Pareto set of systems that are non-dominated in terms of all performance measures. Another multi-objective selection approach by Butler, Morrice, and Mullarkey [11] uses utility and weighting functions to construct a two-stage procedure to find the best system when tradeoffs between performance measures are known. Morrice and Butler [37] utilized multiple attribute utility theory to develop a two-stage procedure to select the best system with constraints. Pujowidianto et al. [42] develop a procedure for constrained R&S under multiple constraints within the OCBA approach, and Kabirian and Ólafsson [29] suggest an indifference-zone approach for the selection of the best system while considering the probability that several stochastic constraints are feasible. Andradóttir and Kim [5] propose several fully sequential constrained R&S procedures for independent systems under one constraint.

We embrace the fully sequential IZ approach to R&S, as fully-sequential procedures have been shown to reduce the number of necessary observations to reach a decision while guaranteeing a nominal PCS, see Paulson [40], Hartmann [27], and Kim and Nelson [30]. The approach utilizes an IZ parameter, which indicates the smallest difference between systems worth detecting. Thus, we can be satisfied in choosing any system with a mean inside an indifference zone of the best system's mean. Fully sequential procedures attempt to limit the number of necessary observations by determining which systems require additional observations after each stage of sampling. Stages can consist of as little as one data point for each system in contention, so decisions are made efficiently without compromising the desired PCS.

The constrained R&S part of the thesis is closely related to the work of Andradóttir and Kim (2010). Andradóttir and Kim (2010) introduce a fully sequential, indifference-zone framework for constrained R&S consisting of two phases, i.e., feasibility check and selection of the best (comparison). These phases may be addressed either *sequentially* (the feasibility of each system is determined before comparison begins) or *simultaneously* (the feasibility check and comparison screening occur simultaneously after each additional sample). Andradóttir and Kim [5] provide the $\mathcal{AK}$ procedure as an example of a sequentially running procedure and the $\mathcal{AK}+$ procedure as a simultaneously running procedure.

Simultaneously running procedures are particularly interesting, as they are statistically valid and efficient in many mean configurations, due to the feasibility check and comparison screening after each stage of sampling. Simultaneous procedures keep systems in contention only while they have been found neither infeasible nor inferior to a feasible system.

In this thesis, we present a new framework for fully sequential constrained R&S based on the concept of dormancy, extending fully sequential constrained R&S procedures to incorporate both any number of constraints and any correlation across systems (allowing for the use of CRN), and present a new procedure that minimizes the number of switches (setup cost of starting and stopping simulations) between the simulated alternatives.

The thesis is organized as follows: Chapter 2 introduces the new overlapping modified jackknifed Durbin–Watson variance estimator. Chapter 3 features our study of overlapping variance estimators in steady-state R&S procedures. Chapters 4 through 6 focus on constrained R&S. In Chapter 4, we introduce the new dormancy framework for comparison of constrained systems. Chapter 5 provides general procedures for multiple constraints and any correlation across systems. We present our minimal switching procedure in Chapter 6 and conclude with a summary of the contributions of this thesis in Chapter 7.

CHAPTER II

AN OVERLAPPING DURBIN–WATSON VARIANCE ESTIMATOR FOR SIMULATIONS

The modified jackknifed Durbin–Watson (MJDW) estimator of Batur, Goldsman, and Kim [6] has characteristics that we desire in a preferable estimator for σ^2, e.g., comparatively low bias and low variance. In particular, Batur, Goldsman, and Kim [6] show analytically and empirically that the MJDW estimator outperforms the area and Cramér von–Mises (CvM) estimators in terms of variance, while maintaining a similar bias. Meanwhile, Alexopoulos et al. [3] show that *overlapping* batched versions of the area and CvM estimators have significantly lower variance than the analogous estimators incorporating nonoverlapping batches, again without increasing bias. The current chapter combines the overlapping and MJDW methodologies, with the hope that the resulting overlapping MJDW estimator will be superior to its nonoverlapping counterpart. We would also like to see if the overlapping MJDW estimator has better properties than do the overlapping area and overlapping CvM estimators.

This chapter is organized as follows: In Section 2.1, we give background, assumptions, and definitions needed for the discussion of our new estimator. In Section 2.2, we introduce the overlapping MJDW variance estimator. Section 2.3 presents some experimental results and points out a useful property of the overlapping MJDW estimator. We conclude the chapter in Section 2.4.

2.1 Background

In this section, we provide background material that will be needed to define our new overlapping MJDW estimator in Section 2.2.

2.1.1 Standardized Time Series

We assume that our sample of output data $\{X_i : i = 1, \ldots, n\}$ is from a stationary stochastic process and that it satisfies a Functional Central Limit Theorem (FCLT) (see, e.g., Glynn and Iglehart [21]):

Assumption 1. For the stationary process $\{X_i;\ i = 1, \ldots, n\}$, there exist constants μ and $\sigma > 0$ such that

$$X_n(t) \equiv \frac{\lfloor nt \rfloor \left(\bar{X}_{\lfloor nt \rfloor} - \mu\right)}{\sqrt{n}} \Rightarrow \sigma\mathcal{W}(t) \text{ for } t \in [0,1],$$

where $\lfloor\cdot\rfloor$ is the floor function; $\bar{X}_\ell \equiv \frac{1}{\ell}\sum_{i=1}^{\ell} X_i$, $\ell = 1, \ldots, n$; $\Rightarrow$ denotes weak convergence (as $n \to \infty$) in the Skorohod space $D[0,1]$ of real-valued functions on $[0,1]$ that are right-continuous with left-hand limits; and $\mathcal{W}(\cdot)$ is a standard Brownian motion process.

From here on, we divide the output into batches—either nonoverlapping or overlapping—of size m, and we define the ratio $b \equiv n/m$. Thus, nonoverlapping batch j consists of the observations $\{X_{(j-1)m+i} : i = 1, \ldots, m\}$, for $j = 1, \ldots, b$. The standardized time series (STS) for nonoverlapping batch j is (Schruben [46])

$$T_{j,m}(t) \equiv \frac{\lfloor mt \rfloor(\bar{X}_{j,m} - \bar{X}_{j,\lfloor mt \rfloor})}{\sigma\sqrt{m}} \text{ for } t \in [0,1] \text{ and } j = 1, \ldots, b,$$

where the ℓth cumulative mean of the jth nonoverlapping batch is $\bar{X}_{j,\ell} \equiv \frac{1}{\ell}\sum_{p=1}^{\ell} X_{(j-1)m+p}$ for $j = 1, \ldots, b$ and $\ell = 1, \ldots, m$.

Similarly, overlapping batch j consists of the observations $\{X_{j+i} : i = 0, \ldots, m-1\}$ for $j = 1, \ldots, n-m+1$. The STS for the jth overlapping batch is

$$T^O_{j,m}(t) \equiv \frac{\lfloor mt \rfloor(\bar{X}^O_{j,m} - \bar{X}^O_{j,\lfloor mt \rfloor})}{\sigma\sqrt{m}} \text{ for } t \in [0,1] \text{ and } j = 1, \ldots, n-m+1,$$

where the ℓth cumulative mean of the jth overlapping batch is $\bar{X}^O_{j,\ell} \equiv \frac{1}{\ell}\sum_{p=0}^{\ell-1} X_{j+p}$ for $j = 1, \ldots, n-m+1$ and $\ell = 1, \ldots, m$.

2.1.2 Area Estimators for σ^2

The area estimator from a particular batch of observations is computed as the square of weighted area of the corresponding STS. One then averages the area estimators from various individual batches to obtain an "overall" nonoverlapping or overlapping batched area estimator.

2.1.2.1 Nonoverlapping Batched Area Estimator

We first consider the nonoverlapping batched version of the area estimator. The area estimator from the jth nonoverlapping batch is defined by

$$A_j(f;m) \equiv \left[\frac{1}{m}\sum_{\ell=1}^{m} f(\tfrac{\ell}{m})\sigma T_{j,m}(\tfrac{\ell}{m})\right]^2 \text{ for } j = 1,\ldots,b,$$

where $f(\cdot)$ is defined as a continuous weighting function on the interval $[0,1]$ and normalized so that $\int_0^1\int_0^1 f(s)f(t)(\min(s,t)-st)\,\mathrm{d}s\,\mathrm{d}t = 1$.

The nonoverlapping batched area estimator for σ^2 is

$$A(f;b,m) \equiv \frac{1}{b}\sum_{j=1}^{b} A_j(f;m).$$

It can be shown (Schruben [46]) that $A(f;b,m) \Rightarrow \sigma^2\chi_b^2/b$, as $m\to\infty$, where χ_b^2 denotes a χ^2 random variable with b degrees of freedom (d.f.). If we use the weighting function $f_2(t) \equiv \sqrt{840}(3t^2-3t+1/2)$—which is known to elicit good performance properties—we have (Aktaran-Kalaycı et al. [1])

$$\mathrm{E}[A(f_2;b,m)] = \sigma^2 + \frac{7(\sigma^2-6\gamma_2)}{2m^2} + O(\frac{1}{m^3}) \tag{1}$$

and

$$\lim_{m\to\infty}\mathrm{Var}[A(f_2;b,m)] = \frac{2\sigma^4}{b},$$

where $\gamma_j \equiv \sum_{i=1}^{\infty} i^j R_i$ for $j=1,2,\ldots$, and $R_i \equiv \mathrm{Cov}(X_1,X_{1+i})$ for $i=0,1,\ldots$.

2.1.2.2 Overlapping Batched Area Estimator

Now we consider the overlapping version of the area estimator. Alexopoulos et al. [2] define the overlapping area estimator from the jth overlapping batch as

$$A_j^O(f;m) \equiv \left[\frac{1}{m}\sum_{\ell=1}^{m} f(\tfrac{\ell}{m})\sigma T_{j,m}^O(\tfrac{\ell}{m})\right]^2 \text{ for } j = 1,\ldots,n-m+1$$

and the overlapping batched area (OA) estimator for σ^2 as

$$A^O(f;b,m) \equiv \frac{1}{n-m+1}\sum_{j=1}^{n-m+1} A_j^O(f;m),$$

where $b \equiv n/m$. Moreover, Alexopoulos et al. [2] show that for $b \geq 2$,

$$\mathrm{E}[A^O(f_2;b,m)] = \mathrm{E}[A(f_2;b,m)], \quad \lim_{m\to\infty} \mathrm{Var}[A^O(f_2;b,m)] = \frac{3514b-4359}{4290(b-1)^2}\sigma^4, \quad (2)$$

and

$$A^O(f_2;b,m) \approx \frac{\sigma^2\chi^2_{\nu_{\text{eff}}}}{\nu_{\text{eff}}} \text{ for large } m \text{ and } b,$$

where ν_{eff} is the effective d.f. calculated by the method of Satterthwaite [45],

$$\nu_{\text{eff}} = \left[\!\left[\frac{2\mathrm{E}^2[A^O(f_2;b,m)]}{\mathrm{Var}[A^O(f_2;b,m)]}\right]\!\right] = \left[\!\left[\frac{8580(b-1)^2}{3514b-4359}\right]\!\right], \quad (3)$$

and $[\![\cdot]\!]$ rounds to the nearest integer.

2.1.3 Cramér–von Mises Estimators for σ^2

The CvM estimator from a particular batch is the weighted area of the square of the corresponding STS. As in Section 2.1.2, we can produce nonoverlapping and overlapping versions of the "overall" batched CvM estimators.

2.1.3.1 Nonoverlapping Batched Cramér–von Mises Estimator

The CvM estimator for σ^2 from the jth nonoverlapping batch of data is

$$C_j(g;m) \equiv \frac{1}{m}\sum_{\ell=1}^{m} g(\tfrac{\ell}{m})\left[\sigma T_{j,m}(\tfrac{\ell}{m})\right]^2 \text{ for } j = 1,\ldots,b,$$

where $g(\cdot)$ is a continuous weighting function on the interval $[0,1]$, and is normalized so that $\int_0^1 g(t)t(1-t)\,\mathrm{d}t = 1$.

The nonoverlapping batched CvM estimator for σ^2 is

$$C(g;b,m) \equiv \frac{1}{b}\sum_{j=1}^{b} C_j(g;m).$$

Using $g_2(t) \equiv -24 + 150t - 150t^2$ as the weighting function, we have (Goldsman, Kang, and Seila [23] and Aktaran-Kalaycı et al. [1])

$$\mathrm{E}[C(g_2;b,m)] = \sigma^2 + \frac{4(\sigma^2 - 6\gamma_2)}{m^2} + O(\frac{1}{m^3}) \tag{4}$$

and

$$\lim_{m\to\infty} \mathrm{Var}[C(g_2;b,m)] \doteq \frac{1.729}{b}\sigma^4.$$

2.1.3.2 Overlapping Batched Cramér–von Mises Estimator

Alexopoulos et al. [3] define the CvM estimator for σ^2 from the jth overlapping batch,

$$C_j^O(g;m) \equiv \frac{1}{m}\sum_{\ell=1}^{m} g(\tfrac{\ell}{m})\left[\sigma T_{j,m}^O(\tfrac{\ell}{m})\right]^2 \text{ for } j = 1,\ldots,n-m+1,$$

along with the overlapping batched CvM (OC) estimator for σ^2,

$$C^O(g;b,m) \equiv \frac{1}{n-m+1}\sum_{j=1}^{n-m+1} C_j^O(g;m).$$

For $b \geq 2$, one can obtain

$$\mathrm{E}[C^O(g_2;b,m)] = \mathrm{E}[C(g_2;b,m)] \quad \text{and} \quad \lim_{m\to\infty} \mathrm{Var}[C^O(g_2;b,m)] = \frac{10768b - 13605}{13860(b-1)^2}\sigma^4. \tag{5}$$

Once again using Satterthwaite [45], we have

$$C^O(g_2;b,m) \approx \frac{\sigma^2\chi^2_{\nu_{\text{eff}}}}{\nu_{\text{eff}}} \text{ for large } m \text{ and } b,$$

where

$$\nu_{\text{eff}} = \left[\!\left[\frac{27720(b-1)^2}{10768b - 13605} \right]\!\right]. \tag{6}$$

2.2 Durbin–Watson Estimators for σ^2

The nonoverlapping batched Durbin–Watson estimator, first studied in Goldsman et al. [22] and enhanced in Batur, Goldsman, and Kim [6], combines certain area and CvM estimators. We review this estimator in Section 2.2.1 and then finally introduce our new overlapping version in Section 2.2.2.

Before proceeding, we define the following quantity from the jth overlapping batch of size m,

$$D_j^O(m) \equiv 2C_j^O(g_0; m) - A_j^O(f_0; m) \text{ for } j = 1, \ldots, n - m + 1,$$

where $f_0(t) \equiv \sqrt{12}$ and $g_0(t) \equiv 6$ for $t \in [0, 1]$.

2.2.1 Nonoverlapping Batched MJDW Estimator

The MJDW estimator for σ^2 from the jth nonoverlapping batch of the data is (cf. Batur, Goldsman, and Kim [6], who give an equivalent definition with different notation)

$$\widetilde{D}_{\mathrm{J},j}(m) \equiv 2D^O_{(j-1)m+1}(m) - \frac{1}{2}D^O_{(j-1)m+1}(\tfrac{m}{2}) - \frac{1}{2}D^O_{(j-1)m+\frac{m}{2}+1}(\tfrac{m}{2}) \text{ for } j = 1, \ldots, b.$$

This allows us to construct the nonoverlapping batched MJDW estimator,

$$\widetilde{D}_{\mathrm{J}}(b, m) \equiv \frac{1}{b}\sum_{j=1}^{b} \widetilde{D}_{\mathrm{J},j}(m).$$

Batur, Goldsman, and Kim [6] demonstrate that

$$\mathrm{E}[\widetilde{D}_{\mathrm{J}}(b, m)] = \sigma^2 + \frac{2(\sigma^2 - 12\gamma_2)}{m^2} + O(\frac{1}{m^3}) \tag{7}$$

and

$$\mathrm{Var}[\widetilde{D}_{\mathrm{J}}(b, m)] = \frac{1.2}{b}\sigma^4.$$

The nonoverlapping batched MJDW estimator performs well, with low bias and lower variance than the nonoverlapping batched area and CvM estimators.

2.2.2 Overlapping Batched MJDW Estimator

The overlapping modified jackknifed Durbin–Watson (OM) estimator is constructed from MJDW estimators corresponding to all $n-m+1$ overlapping batches of size m. The MJDW estimator from the jth overlapping batch is

$$\widetilde{D}^O_{\mathrm{J},j}(m) \equiv 2D^O_j(m) - \frac{1}{2}D^O_j(\tfrac{m}{2}) - \frac{1}{2}D^O_{j+\frac{m}{2}}(\tfrac{m}{2}) \text{ for } j = 1,\ldots,n-m+1.$$

The OM estimator for σ^2 is then

$$\widetilde{D}^O_{\mathrm{J}}(b,m) \equiv \frac{1}{n-m+1}\sum_{j=1}^{n-m+1} \widetilde{D}^O_{\mathrm{J},j}(m).$$

Since we know the expected value of the MJDW estimator, it follows that

$$\mathrm{E}[\widetilde{D}^O_{\mathrm{J}}(b,m)] = \mathrm{E}[\widetilde{D}_{\mathrm{J}}(b,m)]. \tag{8}$$

The asymptotic variance of the OM estimator is much harder to compute analytically. In the absence of theoretical computations of asymptotic variance, we have included Monte Carlo estimates of the OM estimator's variance for specific values of $b = n/m$ in the next section.

2.3 *Experimental Results*

To analyze the distribution and performance of the OM estimator, we present the results of our Monte Carlo experiments.

2.3.1 Configurations and Experimental Design

We compare our variance estimators by testing them on a first-order autoregressive (AR(1)) process. This process is defined by $X_i = \mu_i + \phi(X_{i-1} - \mu_i) + \epsilon_i$, $i \geq 1$, where X_0 is a Nor(0,1) random variable and the ϵ_i's are i.i.d $\mathrm{Nor}(0, 1-\phi^2)$. Since the AR(1) has a simple covariance structure characterized by the lag-k covariance $R_k = \phi^{|k|}$, the asymptotic variance parameter for this process can easily be computed to be $\sigma^2 = (1+\phi)/(1-\phi)$. We concentrate our experiments on the case where $\phi = 0.9$ (so that $\sigma^2 = 19$), performing 100,000 macro-replications.

2.3.2 Comparison of OM with Other Estimators

Table 1 gives the estimated expected values (E) and variances (V) of several variance estimators — including OA, OC, and OM — as we set the number of observations $n = 4096$, but change the value of $b = n/m$. We see that for any fixed (b, m), the estimated expected values for the variance estimators are roughly the same. In particular, for large batch size m, the estimated expected values of all of the estimators are nearly equal to σ^2; but as the batch size m decreases, all of the estimators become more biased.

We can gain additional insight into the bias of the various estimators via a closer examination of Equations (1), (2), (4), (5), (7), and (8). If we assume that $\sigma^2 \ll \gamma_2$ (which Aktaran-Kalaycı et al. [1] show to be the case for the AR(1) with $\phi = 0.9$), then the $O(m^{-2})$ bias term of the OM estimator has slightly higher magnitude than do those of the OA and OC estimators. But in any case, this bias term is evidently quite small since m is itself fairly large.

We also find from Table 1 that the estimated variance of the OM estimator is only about two-thirds that of its nonoverlapping counterpart, MJDW (which in turn has lower variance than the nonoverlapping area and CvM estimators under study here); and Figure 1 shows that the empirical probability distribution function (p.d.f.) for OM is clearly less variable than that of MJDW. On the other hand, all of the overlapping estimators have approximately the same variance. In fact, Figure 2 plots the empirical p.d.f.'s of the OA, OC, and OM estimators based on the same 100,000 replications of the AR(1) with $\phi = 0.9$, $m = 1000$, and $b = 20$; and the three p.d.f.'s are remarkably similar. Yet we see that the p.d.f. of OM seems to fall between those of OA and OC, indicating that the variance of OM lies between the variances of OA and OC. We will have more to say on this point in Section 2.3.3.

Table 1: Estimated means and variances of the nonoverlapping and overlapping batched area, CvM, and MJDW estimators for the variance parameter of an AR(1) process with $\phi = 0.9$ and $n = 4096$ ($\sigma^2 = 19$).

Estimator	$b=4$ $m=1024$		$b=8$ $m=512$		$b=16$ $m=256$	
	E	V	E	V	E	V
$A(f_2; b, m)$	18.84	176	18.76	89	18.13	41
$A^O(f_2; b, m)$	18.98	90	18.74	40	18.06	17
$C(g_2; b, m)$	18.86	150	18.73	75	18.06	34
$C^O(g_2; b, m)$	18.97	85	18.72	38	18.00	17
$\widetilde{D}_{\mathrm{J}}(b, m)$	18.88	104	18.70	51	17.96	23
$\widetilde{D}_{\mathrm{J}}^O(b, m)$	18.97	89	18.74	38	17.94	17

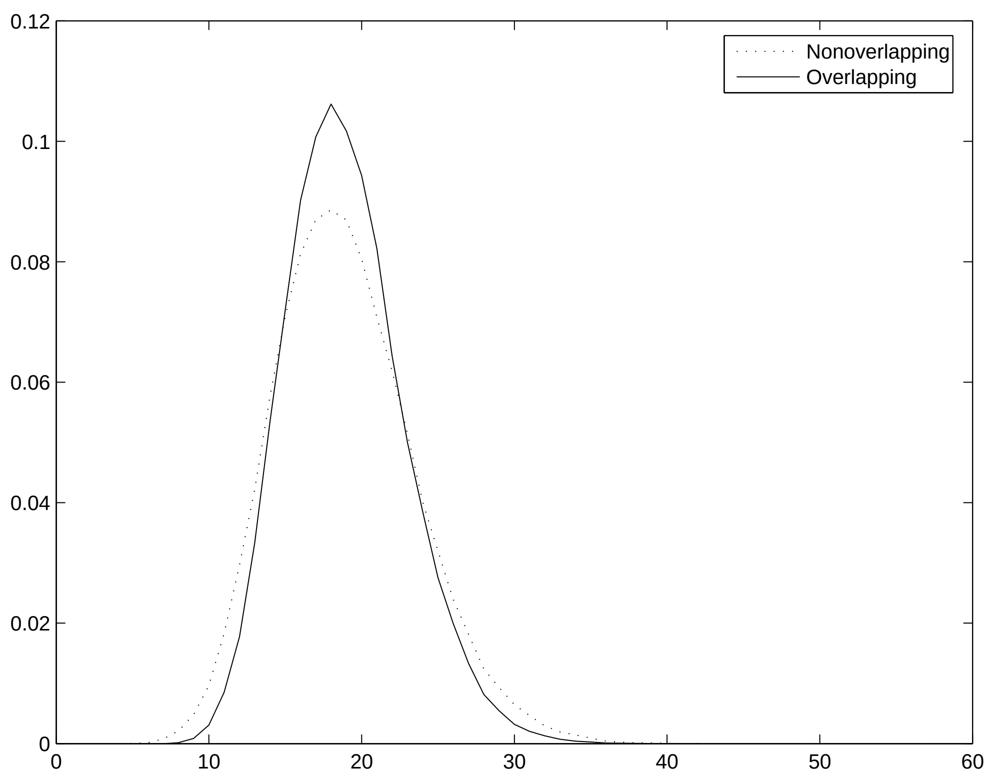

Figure 1: Empirical p.d.f.'s for nonoverlapping MJDW and overlapping MJDW based on 100,000 replications of an AR(1) process with $\phi = 0.9$, $\sigma^2 = 19$, $m = 1000$, and $b = 20$.

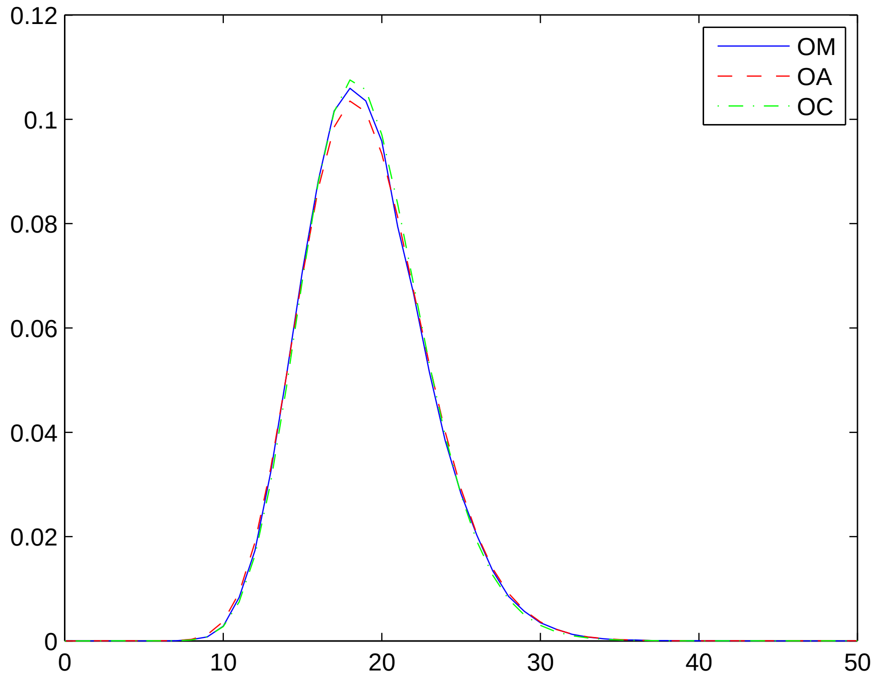

Figure 2: Empirical p.d.f.'s for OA, OC, and OM based on 100,000 replications of an AR(1) process with $\phi = 0.9$, $\sigma^2 = 19$, $m = 1000$, and $b = 20$.

2.3.3 Approximate χ^2 Distribution of OM

Since OA and OC are both approximately χ^2 distributed, we conjecture that the OM estimator is as well. Before presenting an example to test our conjecture, we will conduct a preliminary Monte Carlo study to obtain the effective degrees of freedom (d.f.) for the OM estimator for a variety of b values. By Satterthwaite [45] and Equation (8), we have

$$\nu_{\text{eff}} = \left[\!\!\left[\frac{2\mathrm{E}^2[\widetilde{D}_{\mathrm{J}}^{O}(b,m)]}{\mathrm{Var}[\widetilde{D}_{\mathrm{J}}^{O}(b,m)]} \right]\!\!\right] \approx \left[\!\!\left[\frac{2\sigma^4}{\mathrm{Var}[\widetilde{D}_{\mathrm{J}}^{O}(b,m)]} \right]\!\!\right]. \tag{9}$$

Equation (9) holds for any stochastic process satisfying Assumption 1; so without loss of generality, we can estimate $\mathrm{Var}[\widetilde{D}_{\mathrm{J}}^{O}(b,m)]$, and hence ν_{eff}, using a simple i.i.d. Nor(0,1) process (with $\sigma^2 = 1$). In particular, we ran 100,000 independent replications of the OM estimators with batches of "asymptotic" size $m = 128$ and various choices of b to obtain estimates of $\mathrm{Var}[\widetilde{D}^O_{\mathrm{J}}(b,m)]$ and ν_{eff}. Table 2 gives the resulting estimated d.f. for the OA, OC, and OM estimators from, respectively, Equations (3) and (6) and the Monte Carlo study carried out based on Equation (9). Generally speaking, for fixed b, the effective d.f. for the OM estimator falls between those for OA and OC.

Table 2: Estimated d.f. for OA, OC, and OM estimators.

b	OA	OC	OM	b	OA	OC	OM
2	3	3	3	27	64	68	66
3	6	6	6	28	67	70	69
4	8	8	8	29	69	73	71
5	10	11	11	30	71	75	74
6	13	14	13	31	74	78	76
7	15	16	16	32	76	80	78
8	18	19	18	33	79	83	82
9	20	21	21	34	81	86	84
10	23	24	23	35	84	88	87
11	25	26	26	36	86	91	89
12	27	29	28	37	88	93	91
13	30	32	31	38	91	96	94
14	32	34	33	39	93	99	96
15	35	37	36	40	96	101	99
16	37	39	39	41	98	104	101
17	40	42	41	42	101	106	104
18	42	44	43	43	103	109	107
19	45	47	46	44	106	111	106
20	47	50	49	45	108	114	112
21	49	52	51	46	110	117	113
22	52	55	53	47	113	119	117
23	54	57	56	48	115	122	118
24	57	60	58	49	118	124	121
25	59	62	61	50	120	127	125
26	62	65	64	51	123	129	126

With the estimated d.f. from Table 2 in hand, we finally conduct a Monte Carlo experiment to check if the OM estimator is approximately χ^2. Figure 3 plots the empirical and fitted p.d.f.'s for the OM estimator based on 100,000 replications of an AR(1) process with $\phi = 0.9$ ($\sigma^2 = 19$), $m = 2000$, and $b = 20$ ($\nu_{\mathrm{eff}} = 49$). The fit seems to be excellent, so that OM is indeed approximately χ^2 distributed.

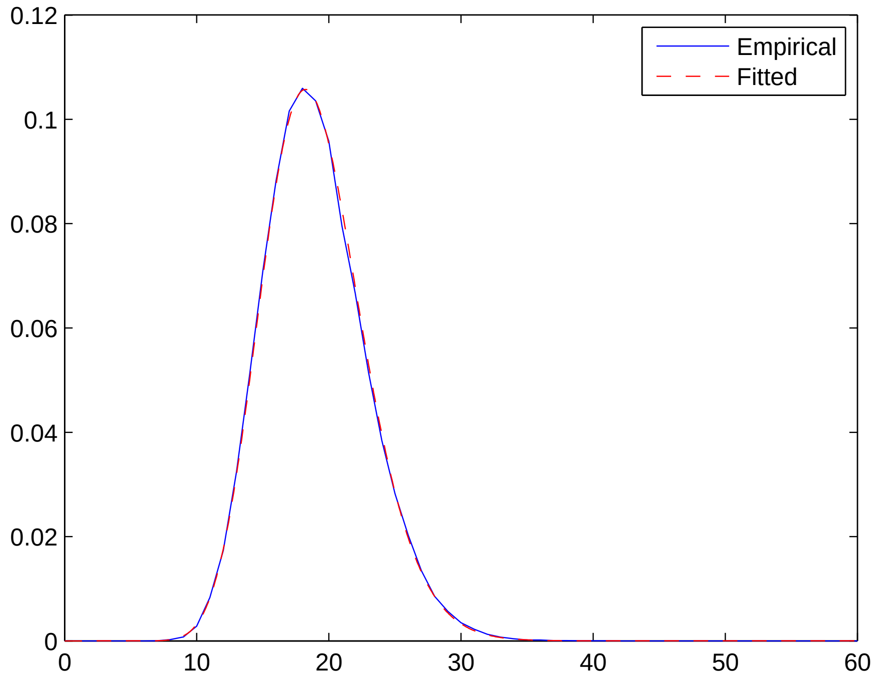

Figure 3: Empirical and fitted p.d.f.'s for OM based on 100,000 replications of an AR(1) process with $\phi = 0.9$, $\sigma^2 = 19$, $m = 2000$, and $b = 20$.

2.4 Conclusion

In this chapter, we introduced a new estimator for the variance parameter of output data from steady-state simulations. The OM estimator improves on the MJDW estimator of Batur, Goldsman, and Kim [6] by incorporating MJDW into overlapping batches.

The OM estimator exhibits characteristics of a good estimator — low bias and low variance. In terms of variance, the OM estimator outperforms all of the nonoverlapping estimators under consideration, and does about as well as the OA and OC overlapping estimators.

We also showed that OM has an approximate χ^2 distribution with about the same d.f. as OA and OC. This is a useful property and has been exploited in the next chapter, which studies an application in ranking and selection.

CHAPTER III

RANKING AND SELECTION TECHNIQUES WITH OVERLAPPING VARIANCE ESTIMATORS FOR SIMULATIONS

A number of new variance parameter estimators have recently been developed in the literature. For example, Alexopoulos et al. [2] propose various overlapping standardized time series (STS) estimators. These overlapping STS estimators have the same bias as, but smaller asymptotic variance than, their nonoverlapping counterparts. Thus, as better variance estimators are introduced, one might become interested in determining whether these new variance estimators can be incorporated into R&S procedures with beneficial results in terms of the required number of observations and the attained PCS. In the current chapter, we investigate such issues. We also discuss the choice of a batch size for overlapping STS estimators in R&S procedures.

This chapter is organized as follows: Section 3.1 defines notation and introduces the variance estimators considered herein. Section 3.2 gives an overview of three R&S procedures specifically designed for steady-state simulation. In Sections 3.3 and 3.4, we discuss our experimental setup and results, showing that the new overlapping variance estimators do indeed yield improved R&S procedure performance. We give conclusions in Section 3.5.

3.1 Variance Estimators

This section describes the notation used throughout the chapter and introduces the variance estimators that we will implement in the selection procedures.

3.1.1 Notation

Let $\mathbf{X}_i \equiv \{X_{i,j} : j = 1, \ldots, n\}$ be a realization from a single run of a simulation of system i, where, throughout the chapter, we assume that $i = 1, \ldots, k$. For example, $X_{i,j}$ could be the jth individual waiting time in the ith queueing system under consideration. After deleting some initial data during a carefully chosen warm-up period, this process is believed to be stationary.

Throughout the chapter, we assume that $\mathbf{X}_i$ satisfies a Functional Central Limit Theorem (FCLT):

Assumption 2.. *For the process $\mathbf{X}_i$, there exist constants μ_i and $\sigma_i > 0$ such that*

$$X_{i,n}(t) \equiv \frac{\lfloor nt \rfloor \left(\bar{X}_{i,\lfloor nt \rfloor} - \mu_i\right)}{\sqrt{n}} \Rightarrow \sigma_i \mathcal{W}_i(t) \quad \text{for } t \in [0,1]$$

where $\lfloor \cdot \rfloor$ is the floor function; $\bar{X}_{i,\ell} \equiv \sum_{j=1}^{\ell} X_{i,j}/\ell$, $\ell = 1, \ldots, n$; $\Rightarrow$ denotes convergence in distribution as $n \to \infty$; and $\mathcal{W}_i$ stands for a standard Brownian motion process.

For a stationary process, the FCLT holds using the steady-state mean μ_i and the variance parameter $\sigma_i^2 \equiv \lim_{n\to\infty} n\mathrm{Var}[\bar{X}_{i,n}]$ (see, for example, Glynn and Iglehart [21]). Further, we make the following assumption concerning competing alternatives:

Assumption 3.. *$\mathbf{X}_i$ and $\mathbf{X}_\ell$ for $i \neq \ell$ are independent.*

This assumption precludes use of certain simulation variance reduction techniques, such as common random numbers.

As σ_i^2 is unknown, it needs to be estimated from the data $\mathbf{X}_i$. We consider four STS variance estimators from the literature: the batched area, overlapping area, overlapping Cramér–von Mises, and overlapping Durbin–Watson estimators.

3.1.2 Batched Area Estimator

To calculate a batched area estimator from a set of n observations $\mathbf{X}_i$, we first split the n data points into b adjacent batches of size m (where $n = bm$). The STS for nonoverlapping batch j of system i is (Schruben [46])

$$T_{i,j,m}(t) \equiv \frac{\lfloor mt \rfloor (\bar{X}_{i,j,m} - \bar{X}_{i,j,\lfloor mt \rfloor})}{\sigma_i \sqrt{m}} \quad \text{for } t \in [0,1] \text{ and } j = 1, \ldots, b,$$

where

$$\bar{X}_{i,j,\ell} \equiv \frac{1}{\ell} \sum_{p=1}^{\ell} X_{i,(j-1)m+p} \quad \text{for } j = 1, \ldots, b \text{ and } \ell = 1, \ldots, m.$$

The area estimator from the jth batch from system i is

$$A_{i,j}(f;m) \equiv \left[\frac{1}{m} \sum_{\ell=1}^{m} f(\tfrac{\ell}{m}) \sigma_i T_{i,j,m}(\tfrac{\ell}{m}) \right]^2 \quad \text{for } j = 1, \ldots, b,$$

where $f(\cdot)$ is a continuous weighting function on the interval $[0,1]$ and normalized so that $\int_0^1 \int_0^1 f(s)f(t)(\min(s,t) - st)\, ds\, dt = 1$. Finally, the batched area (A) estimator for σ_i^2 is defined as

$$A_i(f;b,m) \equiv \frac{1}{b} \sum_{j=1}^{b} A_{i,j}(f;m).$$

3.1.3 Overlapping Area Estimator

The overlapping area estimator is similar to the batched area estimator, but differs in that we now incorporate estimators from $n - m + 1$ overlapping batches of size m, where the jth overlapping batch from system i consists of the observations $X_{i,j}, X_{i,j+1}, \ldots, X_{i,j+m-1}$. The STS for the jth overlapping batch from system i is

$$T^O_{i,j,m}(t) \equiv \frac{\lfloor mt \rfloor (\bar{X}^O_{i,j,m} - \bar{X}^O_{i,j,\lfloor mt \rfloor})}{\sigma_i \sqrt{m}} \quad \text{for } t \in [0,1] \text{ and } j = 1, \ldots, n-m+1,$$

where

$$\bar{X}^O_{i,j,\ell} \equiv \frac{1}{\ell} \sum_{p=0}^{\ell-1} X_{i,j+p} \quad \text{for } j = 1, \ldots, n-m+1 \text{ and } \ell = 1, \ldots, m.$$

Alexopoulos et al. [2] define the overlapping area estimator for σ_i^2 from the jth overlapping batch as

$$A^O_{i,j}(f;m) \equiv \left[\frac{1}{m} \sum_{\ell=1}^{m} f(\tfrac{\ell}{m}) \sigma_i T^O_{i,j,m}(\tfrac{\ell}{m}) \right]^2 \quad \text{for } j = 1, \ldots, n-m+1,$$

and the (overall) overlapping area (OA) estimator for σ_i^2 as

$$A_i^O(f;b,m) \equiv \frac{1}{n-m+1} \sum_{j=1}^{n-m+1} A_{i,j}^O(f;m),$$

where $b = n/m$ (though b can no longer be interpreted as "the number of batches").

3.1.4 Overlapping Cramér–von Mises Estimator

The Cramér–von Mises (CvM) estimator for σ_i^2, obtained from the jth overlapping batch, is

$$C_{i,j}^O(g;m) \equiv \frac{1}{m} \sum_{\ell=1}^{m} g(\tfrac{\ell}{m}) \left[\sigma_i T_{i,j,m}^O(\tfrac{\ell}{m})\right]^2 \quad \text{for } j = 1, \ldots, n-m+1,$$

where $g(\cdot)$ is a normalized weighting function on the interval $[0,1]$ such that $\int_0^1 g(t)t(1-t)\,dt = 1$. Alexopoulos et al. [2] define the (overall) overlapping CvM (OC) estimator for σ_i^2 as

$$C_i^O(g;b,m) \equiv \frac{1}{n-m+1} \sum_{j=1}^{n-m+1} C_{i,j}^O(g;m).$$

3.1.5 Overlapping Modified Jackknifed Durbin–Watson Estimator

The Durbin–Watson (DW) estimator for σ_i^2, obtained from the jth overlapping batch, is

$$D_{i,j}^O(m) \equiv 2C_{i,j}^O(g_0;m) - A_{i,j}^O(f_0;m) \quad \text{for } j = 1, \ldots, n-m+1,$$

where $g_0(t) \equiv 6$ and $f_0(t) \equiv \sqrt{12}$ for $t \in [0,1]$. It can be shown that the DW estimator has relatively low variance but suffers from high small-sample bias (Goldsman et al. [22]). To overcome this bias problem at only a modest cost in variance, Batur, Goldsman, and Kim [6] define the modified jackknifed DW estimator from the jth overlapping batch,

$$\widetilde{D}_{\mathrm{J},i,j}^O(m) \equiv 2D_{i,j}^O(m) - \tfrac{1}{2}D_{i,j}^O(\tfrac{m}{2}) - \tfrac{1}{2}D_{i,j+\frac{m}{2}}^O(\tfrac{m}{2}) \quad \text{for } j = 1, \ldots, n-m+1,$$

where we assume that $m/2$ is an integer.

Chapter 2 defines the (overall) overlapping modified jackknifed Durbin–Watson (OM) estimator for σ_i^2 as

$$\widetilde{D}^O_{\mathrm{J},i}(b,m) \equiv \frac{1}{n-m+1}\sum_{\ell=1}^{n-m+1}\widetilde{D}^O_{\mathrm{J},i,\ell}(m).$$

Chapter 2 explained how to determine the degrees of freedom for each variance estimator.

3.2 Selection Procedures

In this section, we elaborate on the details of three selection procedures, each of which we will implement with the A, OA, OC, and OM estimators. Henceforth, let $\widehat{\sigma}_i^2(b,m)$ denote a generic estimator for σ_i^2 using batch size m and sample-size-to-batch-size ratio $b = n/m$.

3.2.1 Extended Rinott Procedure ($\mathcal{R}+$)

The following two-stage "indifference-zone" procedure is an extension for use in steady-state simulation of Rinott's [43] classic procedure, and was studied in Goldsman and Marshall [25] and Goldsman et al. [24]. For more details on Rinott's procedure, see Mukhopadhyay [38].

1. **Setup:** Select a confidence level (nominal PCS) $1/k < 1-\alpha < 1$, indifference-zone parameter $\delta > 0$, first-stage sample size $n_0 \geq 2$, and batch size $m_0 < n_0$. The indifference-zone parameter δ is chosen as the smallest difference between systems that the experimenter deems as "worth detecting" and is explained more fully in, for example, Bechhofer, Santner, and Goldsman [8].

2. **Initialization:** Obtain Rinott's constant $h = h(\nu, k, 1-\alpha)$ from, say, Bechhofer, Santner, and Goldsman [8], where ν is the degrees of freedom for the associated variance estimator. For each system $i = 1,\ldots,k$, take n_0 observations $X_{i,j}$, $j = 1,\ldots,n_0$, and compute the estimator for σ_i^2, i.e., $\widehat{\sigma}_i^2(b_0,m_0)$ with $b_0 = n_0/m_0$. Let

$$N_i = \max\left\{n_0, \left\lceil \frac{h^2\widehat{\sigma}_i^2(b_0,m_0)}{\delta^2}\right\rceil\right\},$$

for $i = 1, \ldots, k$, where $\lceil \cdot \rceil$ is the ceiling function.

3. **Stopping Rule:** If $n_0 \geq \max_i N_i$, then stop and select the system with the largest first-stage sample mean $\bar{X}_{i,n_0}$ as the best. Otherwise, take $N_i - n_0$ additional observations $X_{i,n_0+1}, X_{i,n_0+2}, \ldots, X_{i,N_i}$ from each system i for which $N_i > n_0$. Select the system with the largest overall sample mean $\bar{X}_{i,N_i}$ as the best.

3.2.2 Extended Kim and Nelson Procedure ($\mathcal{KN}+$)

The next procedure, due to Kim and Nelson [31], is a sequential indifference-zone procedure and is more efficient with observations than Rinott's method. This savings of observations is gained by screening out clearly inferior systems. Here we require an estimator for the variance parameter of the difference between systems i and ℓ, denoted $\sigma^2_{i,\ell}$, which is equal to $\sigma^2_i + \sigma^2_\ell$ under Assumptions 2 and 3. Given the initial sample size n_0, batch size m_0, and $b_0 = n_0/m_0$, we denote the estimator of $\sigma^2_{i,\ell}$ as $\widehat{\sigma}^2_{i,\ell}(b_0, m_0)$, which we calculate using the estimators in Section 3.1 with the data points of the difference $Z_{i,\ell,j} \equiv X_{i,j} - X_{\ell,j}$ for $j = 1, \ldots, n_0$.

1. **Setup:** Select a confidence level $1/k < 1 - \alpha < 1$, indifference-zone parameter $\delta > 0$, first-stage sample size $n_0 \geq 2$, and batch size $m_0 < n_0$. Calculate the constant

$$h^2 \equiv \nu \left([2(1 - (1-\alpha)^{1/(k-1)})]^{-2/\nu} - 1 \right),$$

where the degrees of freedom ν is determined by which variance estimator is used.

2. **Initialization:** Let $I = \{1, \ldots, k\}$ be the set of systems in contention. For each system $i = 1, \ldots, k$, obtain n_0 observations $X_{i,j}$, $j = 1, \ldots, n_0$, and compute the first-stage sample mean $\bar{X}_{i,n_0}$. In addition, for all $i \neq \ell$, use the first n_0 observations to compute the estimator $\widehat{\sigma}^2_{i,\ell}(b_0, m_0)$ for $\sigma^2_{i,\ell}$. Set the observation counter $r = n_0$ and go to **Screening**.

3. **Screening:** Set $I^{\text{old}} = I$. Let

$$I \equiv \left\{ i : i \in I^{\text{old}} \text{ and } \bar{X}_{i,r} \geq \bar{X}_{\ell,r} - W_{i,\ell}(r), \forall \ell \in I^{\text{old}}, \ell \neq i \right\},$$

where

$$W_{i,\ell}(r) \equiv \max\left\{0, \frac{\delta}{2r}\left(\frac{h^2\widehat{\sigma}^2_{i,\ell}(b_0,m_0)}{\delta^2} - r\right)\right\}.$$

4. **Stopping Rule:** If the cardinality $|I| = 1$, then stop and select the system whose index is in I as the best. Otherwise, take one additional observation $X_{i,r+1}$ from each system $i \in I$, set $r = r+1$, and go to **Screening**.

3.2.3 Extended Kim and Nelson Procedure with Updates ($\mathcal{KN}$++)

Goldsman et al. [24] and Kim and Nelson [31] present another selection procedure similar to $\mathcal{KN}+$, but one that updates the variance estimator according to a batching sequence (b_r, m_r), where m_r and b_r denote the batch size and ratio r/m_r, respectively. Both m_r and b_r are non-decreasing functions of the number of observations, r. Goldsman et al. [24] present three batching sequences; we consider here only the sequence that takes $m_r = b_r = \lfloor\sqrt{r}\rfloor$, but with more-frequent updates of b_r when r is small.

1. **Setup:** Same setup as $\mathcal{KN}+$.

2. **Initialization:** Let $I = \{1, \ldots, k\}$ be the set of systems in contention. Obtain n_0 observations $X_{i,j}$, $j = 1, \ldots, n_0$, from each system $i = 1, \ldots, k$. Set the observation counter $r = n_0$ and $m_r = m_0$.

3. **Update:** If m_r has changed since the last update, then for all $i \neq \ell$, $i, \ell \in I$, recalculate the estimator $\widehat{\sigma}^2_{i,\ell}(b_r, m_r)$, ν, and h^2.

4. **Screening:** Set $I^{\text{old}} = I$. Let I be updated as in procedure $\mathcal{KN}+$, where we now use

$$W_{i,\ell}(r) \equiv \max\left\{0, \frac{\delta}{2r}\left(\frac{h^2\widehat{\sigma}^2_{i,\ell}(b_r,m_r)}{\delta^2} - r\right)\right\}.$$

5. **Stopping Rule:** If $|I| = 1$, then stop and select the system whose index is in I as the best. Otherwise, take one additional observation $X_{i,r+1}$ from each system $i \in I$, set $r = r+1$, and go to **Update**.

3.3 Experimental Setup

At this point, we are interested in the performance of the R&S procedures when they incorporate the new variance estimators. We follow the same experimental setup that Goldsman et al. [24] used. In particular, we take system 1 as the best system, i.e., the system with the largest mean. For all of the experiments, we set the nominal PCS to 0.95. For purposes of conducting our experiments, we set the indifference-zone parameter $\delta = \sigma_1/\sqrt{n_0}$, where σ_1 is the square root of the variance parameter of the best system.

We tested two different configurations for the mean performance measure: the slippage configuration (SC) and the monotone decreasing means (MDM) configuration. For the SC, all inferior systems are separated from the best system by a distance of δ. For example, $\mu_1 = \delta$, while $\mu_2 = \cdots = \mu_k = 0$. For the MDM configuration, we have $\mu_i = \mu_1 - (i-1)\delta$, $i = 2, \ldots, k$. The MDM configuration tests a procedure's ability to discard clearly inferior systems quickly, while the SC configuration is a "difficult" scenario where the means of all inferior systems are equal and very close to that of the best system (and is often used to test the statistical validity of the procedure).

For our analysis, we concentrate on two key measures: the observed PCS and the sample average number of total raw observations. All experimental results are based on 1000 independent replications.

In our testing, we rank on the mean values of two common processes: the means of AR(1) processes and the mean waiting times for customers in a steady-state M/M/1 queue.

AR(1) processes $\mathbf{X}_i$ for $i = 1, \ldots, k$, are defined as

$$X_{i,j} = \mu_i + \phi(X_{i,j-1} - \mu_i) + \epsilon_{i,j},$$

where $\epsilon_{i,j}$, $j = 1, 2, \ldots$, are i.i.d. Norm(0,$1-\phi^2$) random variables with $\phi \in (-1, 1)$. In this chapter, we chose a fairly high value for the serial correlation coefficient, $\phi = 0.9$. For AR(1) processes, larger μ_i is better.

The waiting times, $\mathbf{X}_i$, for customers of alternative i of an M/M/1 queuing system are given by

$$X_{i,j} = \max\{0, X_{i,j-1} + S_{i,j-1} - T_{i,j}\},$$

where the service times $S_{i,j-1}$ are i.i.d. exp(μ_i) and the interarrival times $T_{i,j}$ are i.i.d. exp(λ). For each system, the utilization $\rho_i \equiv \lambda/\mu_i$, so that the true expected waiting time $w_i = \rho_i^2/\lambda(1-\rho_i)$. For M/M/1 queueing systems, *smaller* w_i *is better.* We wanted the waiting times to have significant correlation, so we used a high starting value for the utilization, $\rho_1 = 0.9$. It should be noted that as the expected waiting time increases, so does the variance of the expected waiting time. This makes the M/M/1 case somewhat more interesting than the AR(1) process.

3.4 Results

Goldsman et al. [24] tested the performance of R&S procedures when nonoverlapping batch means, overlapping batch means, and A estimators were considered. Their experimental results show that the R&S procedures generally achieve at least the nominal PCS when a large enough batch size is used for a particular variance estimator. In addition, they found that the A estimator often produces the best performance in terms of the number of total observations when compared to implementations of batch means and overlapping batch means variance estimators. So, we are interested here in comparing the performance of the R&S procedures incorporating the OA, OC, and OM estimators with that of the A estimator.

Our experiments show that overlapping variance estimators provide a substantial improvement in observations required, without sacrificing correct selections. The savings in observations garnered with the use of the OA, OC, or OM estimators (compared to the A estimator) depend on the choice of batch size and selection procedure, but typically range from 10% to 50%.

We illustrate our results in pairs of tables, which show the sample average of the total number of raw observations and the estimated PCS, over the 1000 replications, for various choices of the initial batch size m_0 with a fixed n_0. Tables 3 and 4 display results when AR(1) processes are tested with $k = 2$ under the SC configuration, while Tables 5 and 6 are devoted to AR(1) processes with $k = 10$ under the SC configuration. Tables 7 and 8 give results for M/M/1 queue-waiting-time processes with $k = 2$ under the SC configuration. Results under the MDM configuration are provided in Tables 9 and 10 for AR(1) processes with $k = 10$ and in Tables 11 and 12 for M/M/1 queue-waiting-time processes with $k = 5$.

3.4.1 Slippage Configuration

Experiments under the SC configuration are usually performed to test a procedure's ability to handle difficult scenarios. Kim and Nelson [31] point out that the observed PCS does not always meet the nominal PCS for the A estimator, and there is some degradation in the observed PCS from the nominal level at small initial batch sizes m_0. However, they show that (i) such degradation is not significant, (ii) a large m_0 helps satisfy the PCS requirement, and (iii) the coverage problem goes away either with large k or under the MDM configuration.

We observe precisely the same tendencies when the OA, OC, and OM estimators are used. In particular, in most cases, the actual PCS with the OA, OC, and OM estimators is at least that of the A estimator for all three R&S procedures; see Tables 3, 5, and 7.

The non-normality of observations from the M/M/1 queue-waiting-time processes affects PCS adversely — but not too significantly as long as m_0 is large — as shown in Table 7.

We notice that a large m_0 helps achieve the nominal level of PCS, but at the cost of more observations. The good news is that the new overlapping variance estimators dramatically decrease the number of observations needed to reach a decision — especially for the large-m_0 case — as shown in Tables 4, 6, and 8. For example, Table 6 reveals a 65% savings in the number of observations from 1,457,600 for the A estimator to 516,500 for the OA estimator when the $\mathcal{R}+$ procedure is implemented on AR(1) processes with $k = 10$ and $m_0 = 500$ under the SC configuration.

3.4.2 MDM configuration

As one would expect, the estimated PCS values under the MDM configuration tend to be higher than those under the SC. For instance, for $k = 10$ AR(1) processes, Table 5 for the SC case shows that a number of estimated PCS values are lower than the nominal 0.95, while Table 9 for the MDM configuration shows that all PCS values are substantially larger than the nominal level. As the coverage problem is less problematic with the MDM configuration, we focus on discussing the efficiency of the R&S procedures in terms of sample size.

The advantages of implementing the overlapping estimators are most clearly seen with respect to the $\mathcal{R}+$ and $\mathcal{KN}+$ procedures. For instance, we notice in Table 4 that, for $k = 2$ AR(1) processes, procedures $\mathcal{R}+$ and $\mathcal{KN}+$ using any of our new overlapping variance estimators record savings of roughly up to 40% over the A estimator, especially when the procedures use relatively large m_0 sizes. Table 10 shows that we obtain even more savings by using the overlapping estimators for $\mathcal{R}+$ and $\mathcal{KN}+$ when $k = 10$ — up to 65%. This demonstrates that variance estimates with good statistical properties (low bias and low variance) can improve the efficiency of R&S procedures significantly.

Table 3: Estimated PCS when AR(1) processes are tested with the SC configuration, $k = 2$, $\phi = 0.9$, $n_0 = 1000$, $1 - \alpha = 0.95$, and $\delta = \sigma_1/\sqrt{n_0}$.

	$\mathcal{R}+$				$\mathcal{KN}+$				$\mathcal{KN}++$			
m_0	A	OA	OC	OM	A	OA	OC	OM	A	OA	OC	OM
500	0.966	0.970	0.969	0.966	0.964	0.953	0.953	0.957	0.935	0.934	0.936	0.942
250	0.943	0.939	0.945	0.944	0.942	0.936	0.936	0.943	0.925	0.927	0.931	0.925
200	0.951	0.940	0.943	0.945	0.949	0.949	0.943	0.943	0.942	0.941	0.942	0.940
125	0.939	0.934	0.938	0.943	0.946	0.924	0.923	0.925	0.929	0.931	0.931	0.925
100	0.938	0.937	0.932	0.930	0.919	0.914	0.912	0.913	0.926	0.931	0.932	0.923

Table 4: Sample average of total number of raw observations when AR(1) processes are tested with the SC configuration, $k = 2$, $\phi = 0.9$, $n_0 = 1000$, $1 - \alpha = 0.95$, and $\delta = \sigma_1/\sqrt{n_0}$. Entries are shown in units of 10^4.

	$\mathcal{R}+$				$\mathcal{KN}+$				$\mathcal{KN}++$			
m_0	A	OA	OC	OM	A	OA	OC	OM	A	OA	OC	OM
500	4.08	2.35	2.35	2.32	2.54	1.48	1.47	1.45	0.74	0.67	0.67	0.67
250	1.74	1.24	1.22	1.23	1.10	0.77	0.77	0.77	0.64	0.60	0.60	0.60
200	1.51	1.14	1.11	1.09	0.97	0.72	0.69	0.69	0.63	0.57	0.57	0.57
125	1.09	0.92	0.91	0.90	0.72	0.58	0.58	0.56	0.55	0.51	0.51	0.50
100	0.94	0.83	0.82	0.79	0.59	0.52	0.51	0.49	0.48	0.46	0.46	0.44

Table 5: Estimated PCS when AR(1) processes are tested with the SC configuration, $k = 10$, $\phi = 0.9$, $n_0 = 1000$, $1 - \alpha = 0.95$, and $\delta = \sigma_1/\sqrt{n_0}$.

	$\mathcal{R}+$				$\mathcal{KN}+$				$\mathcal{KN}++$			
m_0	A	OA	OC	OM	A	OA	OC	OM	A	OA	OC	OM
500	0.964	0.988	0.992	0.987	0.972	0.987	0.990	0.981	0.901	0.943	0.947	0.942
250	0.956	0.963	0.968	0.967	0.957	0.951	0.958	0.950	0.923	0.943	0.947	0.942
200	0.963	0.964	0.953	0.959	0.945	0.948	0.947	0.951	0.929	0.954	0.953	0.951
125	0.957	0.964	0.941	0.940	0.949	0.933	0.931	0.928	0.909	0.922	0.920	0.916
100	0.932	0.947	0.923	0.920	0.919	0.917	0.917	0.908	0.896	0.902	0.900	0.892

Table 6: Sample average of total number of raw observations when AR(1) processes are tested with the SC configuration, $k = 10$, $\phi = 0.9$, $n_0 = 1000$, $1 - \alpha = 0.95$, and $\delta = \sigma_1/\sqrt{n_0}$. Entries are shown in units of 10^4.

	$\mathcal{R}+$				$\mathcal{KN}+$				$\mathcal{KN}++$			
m_0	A	OA	OC	OM	A	OA	OC	OM	A	OA	OC	OM
500	145.76	51.65	51.54	51.27	90.83	32.45	33.06	32.92	7.20	6.52	6.56	6.48
250	31.03	17.92	17.86	17.76	18.51	9.78	9.78	9.73	6.34	5.95	5.96	5.92
200	24.04	15.83	15.30	15.15	13.95	8.48	8.05	7.98	6.07	5.79	5.75	5.69
125	15.98	12.26	12.08	11.91	8.46	6.22	6.12	6.05	5.32	5.12	5.09	4.96
100	13.34	10.94	10.86	10.42	7.02	5.45	5.39	5.20	4.85	4.74	4.71	4.52

Table 7: Estimated PCS when M/M/1 processes are tested with the SC configuration, $k = 2$, $\rho = 0.9$, $n_0 = 24000$, $1 - \alpha = 0.95$, and $\delta = \sigma_1/\sqrt{n_0}$.

	$\mathcal{R}+$				$\mathcal{KN}+$				$\mathcal{KN}++$			
m_0	A	OA	OC	OM	A	OA	OC	OM	A	OA	OC	OM
12000	0.938	0.930	0.937	0.942	0.952	0.949	0.952	0.950	0.924	0.928	0.922	0.935
8000	0.926	0.923	0.912	0.919	0.930	0.923	0.931	0.926	0.923	0.915	0.915	0.917
6000	0.900	0.910	0.914	0.906	0.938	0.922	0.923	0.921	0.920	0.912	0.914	0.919
4800	0.909	0.908	0.903	0.902	0.932	0.914	0.912	0.919	0.903	0.902	0.901	0.898
4000	0.911	0.893	0.900	0.905	0.914	0.912	0.910	0.908	0.915	0.912	0.906	0.906
3000	0.904	0.899	0.897	0.897	0.901	0.899	0.900	0.905	0.909	0.911	0.910	0.905
2400	0.900	0.898	0.902	0.891	0.911	0.902	0.904	0.902	0.891	0.893	0.892	0.889

Table 8: Sample average of total number of raw observations when M/M/1 processes are tested with the SC configuration, $k = 2$, $\rho = 0.9$, $n_0 = 24000$, $1 - \alpha = 0.95$, and $\delta = \sigma_1/\sqrt{n_0}$. Entries are shown in units of 10^5.

	$\mathcal{R}+$				$\mathcal{KN}+$				$\mathcal{KN}++$			
m_0	A	OA	OC	OM	A	OA	OC	OM	A	OA	OC	OM
12000	14.01	8.23	8.20	8.21	7.26	4.25	4.23	4.25	2.01	1.79	1.78	1.80
8000	7.81	5.09	5.08	5.02	4.43	2.67	2.66	2.67	1.81	1.60	1.59	1.60
6000	5.99	4.42	4.41	4.36	3.28	2.36	2.34	2.31	1.65	1.48	1.48	1.47
4800	5.00	4.04	3.94	3.86	2.89	2.12	2.07	2.04	1.63	1.51	1.50	1.49
4000	4.44	3.68	3.63	3.56	2.39	1.93	1.91	1.88	1.51	1.42	1.40	1.40
3000	3.62	3.23	3.18	3.09	1.95	1.69	1.67	1.64	1.39	1.30	1.30	1.26
2400	3.24	2.86	2.85	2.70	1.73	1.54	1.53	1.47	1.30	1.24	1.23	1.18

We observe similar trends in our M/M/1 waiting-time experiments, with savings for the implementations with overlapping variance estimators of up to 40% as in Table 8 for $k = 2$ and 60% as in Table 12 for $k = 5$.

The relative savings are much more modest when using OA, OC, or OM in procedure $\mathcal{KN}$++, mainly because the procedure is already extremely efficient. Specifically, the updating procedure of $\mathcal{KN}$++ quickly recovers from a poor variance estimate by allowing us to recalculate variance estimates as the procedure progresses from much larger collections of data than the initial sample. So the OA, OC, OM, and A estimators can all eventually produce good variance estimates. Nevertheless, we still can save up to 10% by implementing the OA, OC, or OM estimators over the A estimator.

3.4.3 Batch Size

As we decrease the initial batch size m_0, fewer observations are needed until a decision is made, and the percentage savings of observations required by the overlapping estimators compared to the A estimator tends to decrease for all three R&S procedures. This is because as the batch size decreases for a given n_0, the number of batches increases and the χ^2-like empirical distributions of the various estimators seem to approach each other. This in turn implies similar statistical properties (including the mean and variance) of the four estimators. A side effect of a small batch size is that the procedures often require a smaller-than-necessary number of observations until a decision; and this may result in PCS falling below the nominal level. For example, we see this effect in the observed PCS of Table 3 with $m_0 \leq 250$.

We recommend an initial batch size that is roughly one-quarter of the initial sample. This guarantees that the degrees of freedom of any variance estimator is not too small, ensuring estimated PCS close to (or above) the nominal level and significant savings in observations compared to the A estimator.

Table 9: Estimated PCS when AR(1) processes are tested with the MDM configuration, $k = 10$, $\phi = 0.9$, $n_0 = 1000$, $1 - \alpha = 0.95$, and $\delta = \sigma_1/\sqrt{n_0}$.

	$\mathcal{R}+$				$\mathcal{KN}+$				$\mathcal{KN}{++}$			
m_0	A	OA	OC	OM	A	OA	OC	OM	A	OA	OC	OM
500	0.994	0.995	0.998	0.997	0.995	0.999	0.997	1.000	0.995	0.992	0.993	0.993
250	0.995	0.993	0.994	0.992	0.995	0.992	0.993	0.991	0.987	0.990	0.990	0.987
200	0.995	0.993	0.992	0.992	0.996	0.992	0.992	0.992	0.990	0.994	0.994	0.992
125	0.989	0.985	0.987	0.988	0.986	0.987	0.990	0.986	0.982	0.985	0.984	0.982
100	0.983	0.983	0.985	0.988	0.987	0.984	0.984	0.984	0.990	0.986	0.986	0.980

Table 10: Sample average of total number of raw observations when AR(1) processes are tested with the MDM configuration, $k = 10$, $\phi = 0.9$, $n_0 = 1000$, $1 - \alpha = 0.95$, and $\delta = \sigma_1/\sqrt{n_0}$. Entries are shown in units of 10^4.

	$\mathcal{R}+$				$\mathcal{KN}+$				$\mathcal{KN}{++}$			
m_0	A	OA	OC	OM	A	OA	OC	OM	A	OA	OC	OM
500	143.49	51.33	51.26	51.21	40.10	14.25	14.28	14.30	4.11	3.47	3.51	3.47
250	31.35	18.02	17.96	17.83	8.19	4.32	4.31	4.28	3.28	2.88	2.88	2.86
200	24.00	15.88	15.35	15.19	6.02	3.74	3.58	3.53	3.08	2.74	2.61	2.70
125	15.90	12.28	12.09	11.90	3.75	2.77	2.75	2.72	2.66	2.44	2.43	2.37
100	13.35	10.92	10.86	10.39	3.16	2.49	2.46	2.39	2.43	2.25	2.24	2.17

Table 11: Estimated PCS when M/M/1 processes are tested with the MDM configuration, $k = 5$, $\rho = 0.9$, $n_0 = 24000$, $1 - \alpha = 0.95$, and $\delta = \sigma_1/\sqrt{n_0}$.

	$\mathcal{R}+$				$\mathcal{KN}+$				$\mathcal{KN}{++}$			
m_0	A	OA	OC	OM	A	OA	OC	OM	A	OA	OC	OM
12000	0.958	0.994	0.995	0.990	0.983	0.994	0.994	0.993	0.960	0.960	0.961	0.961
8000	0.972	0.979	0.987	0.982	0.989	0.982	0.979	0.982	0.968	0.968	0.968	0.968
6000	0.963	0.984	0.985	0.986	0.978	0.974	0.975	0.969	0.968	0.968	0.967	0.967
4800	0.964	0.987	0.987	0.987	0.975	0.963	0.966	0.960	0.962	0.962	0.960	0.960
4000	0.966	0.988	0.988	0.988	0.967	0.958	0.963	0.962	0.961	0.961	0.960	0.960
3000	0.976	0.990	0.989	0.987	0.959	0.954	0.954	0.955	0.955	0.955	0.957	0.957
2400	0.976	0.988	0.987	0.984	0.956	0.948	0.949	0.945	0.961	0.961	0.958	0.958

Table 12: Sample average of total number of raw observations when M/M/1 processes are tested with the MDM configuration, $k = 5$, $\rho = 0.9$, $n_0 = 24000$, $1 - \alpha = 0.95$, and $\delta = \sigma_1/\sqrt{n_0}$. Entries are shown in units of 10^5.

	$\mathcal{R}+$				$\mathcal{KN}+$				$\mathcal{KN}{++}$			
m_0	A	OA	OC	OM	A	OA	OC	OM	A	OA	OC	OM
12000	215.55	97.14	96.81	95.31	75.07	30.74	30.67	30.36	7.46	6.51	6.51	6.50
8000	89.69	46.14	46.00	45.60	29.30	13.31	13.30	13.14	6.51	5.63	5.46	5.45
6000	61.69	38.24	38.12	37.47	18.56	10.60	10.56	10.39	6.08	5.39	5.38	5.37
4800	47.78	33.49	32.51	31.57	14.28	9.08	8.70	8.48	5.62	4.99	4.91	4.88
4000	39.82	29.62	29.04	28.33	11.49	7.82	7.62	7.48	5.33	4.83	4.81	4.79
3000	30.32	24.62	24.52	23.09	8.46	6.35	6.29	5.97	4.84	4.45	4.34	4.34
2400	25.01	21.11	21.10	19.46	6.46	5.38	5.37	4.98	4.30	4.07	3.98	3.96

3.4.4 Variance Estimators

The newly implemented variance estimators, OA, OC, and OM, performed similarly under all configurations and conditions. All three perform substantially better than the benchmark A estimator, by saving observations without degrading PCS. We attribute the savings over A to the additional degrees of freedom possessed by the three overlapping variance estimators. However, little separates the degrees of freedom of OA, OC, and OM, and in fact we have seen that their performance results do not differ much. Among the three new estimators, there is perhaps a slight advantage in using the OM estimator, as our Monte Carlo results seem to indicate that it is often a bit more parsimonious in terms of observations.

3.5 Conclusion

We have shown through our experiments that implementing any of the three new overlapping variance estimators can provide a significant savings over the vanilla area estimator in terms of observations needed until a decision is made. This savings is gained without sacrificing the probability of finding the correct system. Our results show the overlapping variance estimators, OA, OC, and OM, should be preferred over the area estimator and thus over other previously studied estimators, including nonoverlapping batch means and overlapping batch means. For the best performance in terms of observations needed, we give a slight nod to use of the OM estimator in the $\mathcal{KN}$++ algorithm with one-quarter of the initial sample as the initial batch size.

CHAPTER IV

A DORMANCY FRAMEWORK FOR EFFICIENT COMPARISON OF CONSTRAINED SYSTEMS

We present a framework that improves on simultaneously-running procedures for constrained ranking and selection (R&S) in that it additionally allows the procedure to pause a system's sampling when it is found inferior to *any* system still in contention. We cannot eliminate such systems until their superior system's feasibility is verified, but rather we keep them "dormant." If the superior system is indeed found to be feasible, then we have achieved some savings over the original procedure. Otherwise, we allow the dormant system to return to contention. If the feasibility check phase requires a good deal more observations than the selection phase, an algorithm with dormancy can be much more efficient than the original procedure. A case displayed in our experimental study shows that dormancy can save many samples, presumably by saving almost every system from completing a feasibility check.

The implementation of the dormancy framework does face a challenge, namely comparison between unevenly sampled systems. If a dormant system returns to the set of contending systems, that system will most likely have fewer observations than the other contending systems. This complicates the selection phase of the algorithm. The three proposed dormancy approaches in this chapter use different rules to handle comparison of systems under uneven samples sizes. We will elaborate on these differences in the following sections, providing results to demonstrate the efficiency of each of our approaches.

To summarize, in this chapter we introduce our framework for the general constrained R&S problem, provide three approaches for implementing this framework, elaborate on how to apply the dormancy framework to improve upon a specific algorithm, $\mathcal{AK}+$, for one constrained performance measure, and compare the experimental results of $\mathcal{AK}+$, with or without dormancy, with those of the sequentially-running $\mathcal{AK}$. However, the dormancy framework is not limited to this case, as it can be applied to simultaneous procedures considering any number of constrained performance measures. For additional information on procedures for multiple constraints, see Chapter 5.

The chapter progresses as follows: Section 4.1 provides notation, assumptions, and modeling formulations needed to properly present and analyze the framework. Section 4.2 outlines the general dormancy framework and provides three implementation approaches that fall within the framework, including one provably valid approach. In Section 4.3, we combine the framework with $\mathcal{AK}+$ to generate three new procedures, and suggest some additional heuristic modifications. Then we compare the performance of our new procedures with that of $\mathcal{AK}$ and $\mathcal{AK}+$ through experiments in Section 4.4. We conclude the chapter in Section 4.5.

4.1 Background

In this section, we define our problem and present assumptions that govern our framework. We first describe the nature of constrained R&S in Section 4.1.1. We then turn to notation and assumptions needed to achieve a statistically valid selection in Section 4.1.2.

4.1.1 Problem Formulation

We are concerned with the selection of the best system with respect to the mean of a primary performance measure in the presence of constraints on s secondary performance measures. Let X_{in} be the nth observation of the ith system for the primary performance measure. Similarly, let $Y_{i\ell n}$ be the nth observation of the ℓth secondary performance measure of the ith system. We consider k systems or configurations, so the set S of all possible systems ranges from 1 to k.

We let $x_i = \mathrm{E}[X_{in}]$ and $y_{i\ell} = \mathrm{E}[Y_{i\ell n}]$ be the mean values of the primary and secondary performance measures for each system $i \in S$. Therefore, the objective is to determine which system has the largest primary performance measure, while also having mean secondary performance measure ℓ less than q_ℓ for $\ell = 1, 2, \ldots, s$:

$$\begin{aligned} &\arg\max_{i \in S} x_i \\ &\text{s.t. } y_{i\ell} \leq q_\ell \text{ for all } \ell = 1, 2, \ldots, s. \end{aligned} \tag{10}$$

This objective is accomplished through simulation and use of the IZ approach. The IZ approach applies to both types of performance measures, as statistically validity is required for both the comparison and feasibility check phases.

We let the IZ parameter, δ, be the smallest distance that we consider significant for the primary performance measure. We are essentially indifferent among the feasible systems whose primary performance measures are within δ of each other. For the remainder of the chapter, if x_i is found to be greater than x_j, then we use the terminology that system i is superior to system j (or equivalently system j is inferior to system i).

We employ a similar approach for the secondary performance measures. We designate ϵ_ℓ as the tolerance level associated with constraint ℓ. Any system with $y_{i\ell} \leq q_\ell - \epsilon_\ell$ for all ℓ is considered desirable (and clearly feasible). The set of desirable systems is denoted S_D. Systems that satisfy $y_{i\ell} < q_\ell + \epsilon_\ell$ for all ℓ, but are not in S_D, fall within the tolerance level of the constraints. These systems are declared acceptable and are placed in the set S_A. The other systems have at least one ℓ with $y_{i\ell} \geq q_\ell + \epsilon_\ell$ and are unacceptable and infeasible, placing them in the set S_U.

Let $[b]$ be the index of the best desirable system. We let CS denote the event that we make a correct selection of the best feasible (desirable or acceptable) system whose mean is greater than $x_{[b]} - \delta$ (i.e., that a system in the set S_{CS} is selected, where $S_{CS} = \{i : i \in S_D \cup S_A \text{ with } x_i > x_{[b]} - \delta\}$). Given the stochastic nature of the problem in (10), we cannot always choose the best feasible system. Hence, we seek procedures that choose the best system with a nominal probability $P(CS) \geq 1 - \alpha$.

4.1.2 Assumptions

To analyze our dormancy framework and specific implementations thereof, we need the following assumptions:

Assumption 4. *The original simultaneous procedure guarantees $P(CS) \geq 1 - \alpha$ by ensuring that with probability no smaller than $1 - \alpha$, a particular system in S_{CS} is declared feasible and all other systems in S would eventually be either declared infeasible or eliminated by that particular system (if they are not eliminated by another system first).*

Assumption 5. *If a feasibility decision is made for the dormancy procedure, then the same decision would have been made at the same sample size for the original procedure (if the system in question were not eliminated by another system first). Similarly, if a comparison decision is made for the dormancy procedure, then the same decision would have been made at the same sample size for the original procedure (if both systems were still in contention).*

Assumption 6. *Observation n of system i (i.e., X_{in} and $Y_{i\ell n}$ for $\ell = 1, \ldots, s$) should not depend on the order the systems are sampled.*

Assumption 7. *The parameters for feasibility check and comparison in the original procedure (namely the IZ parameters, variance estimates, and other parameters necessary for validity) depend only on first-stage samples for each system and do not change as a function of the systems remaining in contention.*

Assumption 4 concerns the validity of the original simultaneous procedure and how that validity is established. Assumption 5 ensures the validity of the feasibility and comparison phases of the new procedure we will create by applying the dormancy framework. It makes sure that decisions are made in an identical, valid manner in both procedures. Finally, Assumptions 6 and 7 are used to verify Assumption 5.

We note that the $\mathcal{AK}+$ procedure of Andradóttir and Kim [5] satisfies Assumptions 4 and 7. Moreover, Assumption 6 is satisfied if seeds for each system are kept separately.

While Assumption 5 maintains identical decision criteria for the two procedures, the procedures with and without dormancy may eliminate systems at different times and in different orders. We will show in the following section that this does not affect the validity of procedures utilizing dormancy.

4.2 Dormancy Framework

We introduce the dormancy framework and three specific approaches for implementing it. The framework utilizes the feasibility check and comparison steps of simultaneously-running procedures for constrained selection.

Simultaneously-running procedures as described in Andradóttir and Kim [5] keep F, the set of systems found feasible, M, the set of systems whose feasibility is yet to be determined, and perform two steps after each stage of sampling. First, feasibility screening is performed for undetermined systems in M, eliminating systems that are infeasible. Second, the procedure compares systems in contention. If a system i is found inferior to a feasible system, the inferior system i is eliminated. If a system is found inferior to a system in M, the procedure cannot eliminate the inferior system. Sampling from the inferior system continues until the inferior system is declared infeasible, the superior system is declared feasible, or the inferior system is either eliminated by another feasible system or selected as the best. Thus, we always obtain additional samples for all surviving systems. In general, systems are sampled only while they are not infeasible and not found inferior to another feasible system.

With dormancy, we seek to make simultaneous procedures more efficient, by introducing D, the set of dormant systems, in addition to F and M. We now specify our dormancy framework, which maintains the set D and directs sampling non-dormant systems. In particular, the original simultaneously-running procedure should be modified in the following ways:

General Dormancy Framework

Entering Dormancy: If system $j \in M$ is found superior to system i, make i dormant and add it to D.

Exiting Dormancy: If superior system j is found infeasible, remove system i (and other systems inferior to j) from D. If superior system j is eliminated by a superior feasible system, remove system i (and other systems inferior to j) from D.

Elimination: If superior system j is found feasible, remove i from both D and $M \cup F$.

Sampling: Do not obtain additional samples from system i, while $i \in D$. If system i returns from dormancy, it may have fewer observations than other contending systems. Take observations from systems with the lowest number of samples first.

This dormancy framework operates under the assumption that a system can be inferior to only one system at a time. Variations of the general framework can be constructed to consider multiple superior systems, so that systems return from dormancy if all of their superior systems are eliminated.

The use of dormancy creates a definite shift in the order that systems are simulated. By halting sampling of inferior systems, we systematically collect observations from the systems still in contention with the highest probable primary performance measure first. This framework will be more aggressive than simultaneous procedures that do not highlight superior systems until their feasibility is confirmed. We expect our procedures with dormancy to perform no worse than the original simultaneous procedures, which we confirm in our experimental results in Section 4.4, and to provide a good PCS, which will be addressed next.

Proposition 1. *If the general dormancy framework is combined with a valid simultaneously-running procedure that satisfies Assumption 4 and the framework is applied in a way that satisfies Assumption 5, then the resulting dormancy procedure guarantees $P(CS) \geq 1 - \alpha$.*

Proof: Let E be the event, in the original procedure, that a particular system in S_{CS} is declared feasible and all other systems in S would eventually be either declared infeasible or eliminated by that particular system. Then $P(E) \geq 1-\alpha$ by Assumption 4. Let $\omega \in E$ and let $j^*_\omega \in S_{CS}$ be the system returned as best by the original procedure under the sample path ω. We will show that the procedure with dormancy will also return j^*_ω as best.

Note that Assumption 4 implies that the procedure with dormancy does not eliminate system j^*_ω due to infeasibility (because system j^*_ω is declared feasible and is not eliminated by another system in the original algorithm). Suppose now that the algorithm with dormancy selects a system in $S \setminus \{j^*_\omega\}$ as best. But then there must exist a system $i \in S \setminus \{j^*_\omega\}$ that is declared feasible and eliminates j^*_ω in the procedure with dormancy. Assumption 5 implies that system i would eventually be declared feasible and superior to j^*_ω in the original procedure (if it were not eliminated first by another system). This contradicts the definition of ω, and concludes the proof. $\square$

The general framework does not specify how we plan to ensure that Assumption 2 holds under the application of dormancy. Dormancy does not affect feasibility check, so half of Assumption 5 is easy to verify under Assumptions 6 and 7. Moreover, when all systems are active with equal sample sizes and Assumptions 6 and 7 hold, a procedure with dormancy runs exactly the same comparison as the original simultaneous procedure. However, uneven sample sizes often arise. We will focus the following subsections on how comparison of systems with uneven sample sizes can be handled. In Section 4.2.1, we provide a statistically valid procedure that fits within our framework, namely the dormancy approach with recall, with proof. We also present two heuristic approaches in Section 4.2.2, which utilize the dormancy framework with different strategies to handle sample size discrepancies.

4.2.1 Dormancy with Recall Approach

In our first approach to handle dormancy and the differing sample sizes it causes, we keep track of the number of observations, r_i, for each system i, and when there are dormant systems, we store some past observations for the primary performance measure of all active systems. Thus, the algorithm can "recall" sums from previous sampling stages. The storage of primary performance measure samples enables the comparison of systems at equal sample sizes throughout the entire procedure.

Dormancy with Recall Approach

Utilize the general dormancy framework and handle comparison of systems as follows:

Comparison: When comparing contending systems i and j with r_i and r_j samples, compute the statistic for comparison of both systems using samples up to time $r = \min(r_i, r_j)$ only, even in the presence of additional samples for one of the systems.

Theorem 1. *If a simultaneously-running procedure satisfies Assumptions 4, 6, and 7, the dormancy with recall approach applied to the simultaneous procedure guarantees* $P(CS) \geq 1 - \alpha$.

Proof: As Assumption 4 is assumed to hold, we seek to show that Assumption 5 is satisfied, and thus Proposition 1 applies. The key point to observe is that due to the recall of data and Assumptions 6 and 7, the feasibility check and comparison decisions for each pair of systems are based on the exact same data and criteria for the procedure with and without dormancy. Thus, we will reach the same decisions at the same sample sizes. The result now follows from Proposition 1. □

4.2.2 Heuristic Dormancy Approaches

We also present two heuristic approaches with dormancy, namely dormancy with catch-up and dormancy with averages. These two approaches attempt to capture the efficiency of dormancy without the required storage for recall and store only the summary statistics of the observations of the primary performance measure for each system. When a dormant system returns to the set of contending systems, the two approaches will handle comparison differently.

The dormancy with catch-up approach compares systems with equal sample sizes only. To remove the need for the selection procedure to handle uneven sample sizes, we gather additional observations from the lagging system until it catches up to the other contending systems in terms of number of observations. During this catch-up process, we will test the system's feasibility (if needed). However, comparison will only resume once all contending systems have the same number of observations. The dormancy with catch-up approach does not require storage of past observations, but can be conservative compared to dormancy with recall, because comparison decisions can be delayed.

Dormancy with Catch-up Approach

Utilize the general dormancy framework and handle comparison of systems as follows:

Comparison: When two contending systems i and j have equal sample sizes $r_i = r_j$, compare the systems. Otherwise, wait until sample sizes become equal to compare the systems.

In our third approach, dormancy with averages, comparison among newly returned systems and other contending systems is performed by weighing summary statistics (i.e., partial sums) as in Pichitlamken et al. [41]. In particular, for two systems with different sample sizes, we compare summary statistics scaled by the number of samples available (like an average). Similar to dormancy with catch-up, this approach does not require storage of individual samples. Dormancy with averages should require fewer observations than dormancy with catch-up, though, as comparison decisions can be made at uneven sample sizes.

Dormancy with Averages Approach

Utilize the general dormancy framework and handle comparison of systems as follows:

Comparison: When comparing two contending systems i and j with r_i and r_j samples, let $r = \min(r_i, r_j)$ and compute summary statistics for both systems considering *all* samples. Then weigh the statistics by r/r_i and r/r_j for systems i and j, respectively, and compare the systems.

When considering validity, these two approaches do not meet the requirements of Assumption 2. The use of catch-up or averages changes the way the procedure compares systems. While comparison in these approaches may be valid in some cases, the difficulty lies in comparing two systems when the sample sizes may be pegged at times specified by the completion of a feasibility check or a comparison. The completion of the feasibility check may be a poor time to observe the primary performance measure due to correlation between primary and secondary performance measure samples. This correlation can induce bias, forcing the summary statistic well above or below its true mean value. Similarly, the primary performance measure may be biased at the time when a system returns from dormancy, as the system inevitably had previously been deemed inferior. Dormancy with catch-up and averages commonly compare systems at random times determined by the end of the feasibility check and comparison steps, inviting bias to occur. This bias violates the validity assumptions for many comparison procedures.

To illustrate the bias in the primary performance measure at the time of completion of the feasibility check, we consider a system with one primary and one secondary performance measure (so that $s = 1$), where both measures are normally distributed with means $x_1 = 0$ and $y_{11} = -\epsilon = -1/\sqrt{20}$ and equal variances $\sigma^2_{x_1} = \sigma^2_{y_{11}} = 1$. The first stage sample size equals 20. We test for the feasibility of the secondary performance measure, $y_{11} \leq q_1 = 0$. In Table 13, we observe the sum of samples X_{1n} under different levels of correlation $\rho = \mathrm{Cov}(X_{1n}, Y_{11n})/\sqrt{\sigma^2_{x_1}\sigma^2_{y_{11}}}$ when the feasibility check, Algorithm I of Andradóttir and Kim [5], is completed at time $T_f = 70$. We choose $T_f = 70$ to show how conditioning on the completion of the feasibility check results in $\mathrm{E}\left[\frac{1}{T_f}\sum_{n=1}^{T_f} X_{1n} | T_f = T\right] \neq x_1 = 0$ for some values of T, where the average is taken over all sample paths regardless of whether the system is found feasible or infeasible. Table 13 shows that the sample mean of the primary performance measure can be considerably biased at the time when feasibility is determined. This issue affects not only the dormancy with catch-up and dormancy with averages approaches, but also sequentially-running procedures such as $\mathcal{AK}$.

Table 13: Estimated expected value of X_{1n} when $T_f = 70$ after 5,000 replications for each level of correlation.

ρ	-0.9	-0.5	0.0	0.5	0.9
Estimated Expected Value of X_{1n}	0.53	0.34	0.00	-0.32	-0.55
Standard Error	0.0023	0.0025	0.0024	0.0025	0.0020

We also present an example of how systems can be biased at the end of comparison. In this case, we consider two systems with $x_1 = 0$, $x_2 = \delta = 1/\sqrt{20} \approx 0.2236$, equal variances $\sigma^2_{x_1} = \sigma^2_{x_2} = 1$, and first stage sample size $n_0 = 20$. We take samples X_{1n} and X_{2n} from systems 1 and 2, respectively, and compare the two systems with the fully-sequential $\mathcal{KN}$ procedure of Kim and Nelson [30], the basis for even-sample comparison in the procedures detailed in Section 4.3. Table 14 shows the observed expected value of primary performance measure averages for systems 1 and 2 at the completion time of comparison, T_c, regardless of the comparison decision.

Table 14: Estimated expected value of X_{1n} and X_{2n} for comparisons ending at time T_c after 5,000 replications for each completion time.

	$T_c = 40$	$T_c = 50$	$T_c = 60$	$T_c = 80$	$T_c = 120$	$T_c = 160$
Estimated Expected Value of X_{1n}	-0.1206	-0.0906	-0.0602	-0.0240	0.0167	0.0425
Standard Error	0.0021	0.0018	0.0016	0.0014	0.0011	0.0009
Estimated Expected Value of X_{2n}	0.3014	0.3014	0.2816	0.2528	0.2095	0.1822
Standard Error	0.0041	0.0041	0.0034	0.0030	0.0024	0.0012

Table 14 shows statistically significant bias for both systems in this case. In particular, for the inferior system that should go dormant when comparison is completed, the bias ranges from a strong negative bias at low sample sizes to a positive bias at large sample sizes. Thus, for small sample size eliminations, the inferior system is more likely to be undervalued. At large sample size eliminations, the inferior system average must be reasonably close to the superior system average, or it would have been eliminated earlier. The positive bias at large completion times suggests that the procedure may be more likely to select the inferior system as the best at such times. We will now investigate this issue further.

Figure 4 displays the percentage of completed comparisons occurring at a given sample size, along with the estimated PCS of the completed comparison, for our two system case with the $\mathcal{KN}$ procedure and nominal PCS $= 0.95$. It is interesting to note that $\mathcal{KN}$ does not guarantee constant PCS for all completion times, T_c. In particular, the PCS first increases and then decreases in T_c, and is smaller than the nominal PCS for both small and large T_c.

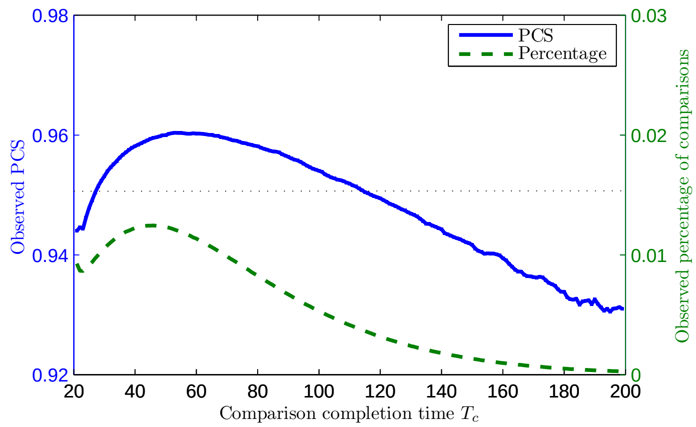

Figure 4: Empirical plot of PCS and percentage of comparisons as a function of the comparison completion time T_c after 10^8 total replications.

The observation that the PCS is smaller for large T_c than for moderate T_c may be counterintuitive in light of the Law of Large Numbers, but may be explained by the fact that the stopping time T_c depends on the sample paths of the two systems being compared. When the inferior system survives for a long time, its sample path may be biased high. Moreover, the $\mathcal{KN}$ comparison approach uses a triangular continuation region and hence smaller differences in sample means are sufficient to complete comparison for larger values of T_c. Indeed, the impact of the triangular shape of the continuation region is confirmed by showing that when the continuation region is specified by two parallel lines and the variances $\sigma_{x_1}^2$ and $\sigma_{x_2}^2$ are known, the PCS does not depend on T_c (see equations (2.3.0.2) and (2.3.0.6(b)) of Borodin and Salminen, [10]). By contrast, numerical results not included here for reasons of brevity show that the version of $\mathcal{KN}$ with known variances $\sigma_{x_1}^2$ and $\sigma_{x_2}^2$ has PCS that decreases with T_c. An implication of this latter result is that it is the use of estimated variances in $\mathcal{KN}$ that explains why the PCS increases with T_c for small values in Figure 4. This is reasonable because when such variance estimates are too small, comparison decisions may be made prematurely, resulting in incorrect selection.

We have documented bias in the primary performance measure at times when feasibility check or comparison is completed. This bias implies that the validity of comparison between systems anchored at these points will be difficult to prove. However, we will show empirically in Section 4.4 that procedures implemented with these approaches can still produce good PCS results.

4.3 Example Procedures for One Constraint

As mentioned earlier, the dormancy framework is flexible enough to be applied to many types of simultaneous procedures with any number of constraints. For this chapter, we evaluate the performance of our dormancy approaches when applied to the simultaneously-running $\mathcal{AK}+$ procedure of Andradóttir and Kim [5] for a single constraint (we only consider the case where the parameter c of $\mathcal{AK}+$ equals one). This will give us a chance to provide detailed implementations in a specific setting and allow us to compare the new dormancy framework with established procedures. Consideration of multiple constraints falls outside the scope of this chapter, but the reader is referred to Chapter 5 for an in-depth discussion of efficient implementation of procedures for multiple constraints.

In Section 4.3.1, we discuss necessary notation and assumptions to ensure validity and proper implementation of the new procedures. Section 4.3.2 outlines the Dormant with Recall, $\mathcal{D}_R$, procedure and includes a proof of validity. Sections 4.3.3 and 4.3.4 contain the Dormant with Catch-up, $\mathcal{D}_C$, and Dormant with Averages, $\mathcal{D}_A$, algorithms, respectively. We finish with some useful heuristic modifications in Section 4.3.5.

4.3.1 Additional Notation and Assumptions

Before presenting the algorithms, we provide some additional notation. Note that because we have only one constraint, we now require fewer subscripts for the samples of the secondary performance measure:

$n_0 \equiv$ the initial sample size;

$S^2_{Y_i} \equiv$ the sample variance of $\{Y_{i1}, \ldots, Y_{in_0}\}$;

$S^2_{X_{ij}} \equiv$ the sample variance of the difference of $\{X_{i1}, \ldots, X_{in_0}\}$ and $\{X_{j1}, \ldots, X_{jn_0}\}$;

$SS_i \equiv$ the set of systems superior to system i in terms of x_i;

$R(r; a, b, c) \equiv \max\{0, \frac{bc}{2a} - \frac{a}{2}r\}$, for $a, b, c \in \mathbb{R}^+$ and $a \neq 0$;

$g(\eta, d) \equiv \frac{1}{2}(1 + 2\eta)^{-(d-1)/2}$.

We also need additional assumptions for the validity of $\mathcal{AK}+$.

Assumption 8. *For each $i = 1, 2, \ldots, k$,*

$$\begin{bmatrix} X_{in} \\ Y_{in} \end{bmatrix} \overset{iid}{\sim} BN\left(\begin{bmatrix} x_i \\ y_i \end{bmatrix}, \Sigma_i \right) n = 1, 2, \ldots$$

where $\overset{iid}{\sim}$ denotes independent and identically distributed, BN denotes bivariate normal, and Σ_i is the 2×2 positive definite covariance matrix of the vector (X_{in}, Y_{in}). Also, (X_{in}, Y_{in}) is independent of $(X_{jn'}, Y_{jn'})$ for $(i, n) \neq (j, n')$, which rules out the use of common random numbers.

Assumption 9. *For any $i \in S_D \cup S_A$ with $i \notin [b]$, $x_i \leq x_{[b]} - \delta$.*

The assumption of normality in the observations can be satisfied through use of within-replication averages or batch means (see, e.g., Law and Kelton [33]). The second assumption allows only one possible best feasible system (i.e., $S_{CS} = \{[b]\}$).

4.3.2 The $\mathcal{D}_R$ Procedure

We present the $\mathcal{D}_R$ procedure created by combining the dormancy with recall approach with $\mathcal{AK}+$.

Procedure [Dormant with Recall $\mathcal{D}_R$]

Setup: Select the overall confidence level $1/k \leq 1-\alpha < 1$. Choose ϵ, q, δ, and $n_0 \geq 2$. Find η, a solution to the equation $g(\eta, n_0) = \beta$, where β is the solution to the equation $\beta + 2[1 - (1-\beta)^{(k-1)/2}] = \alpha$.

Initialization: Let $M = S$ be the set of undetermined systems, $F = \emptyset$ be the set of feasible systems, and $D = \emptyset$ be the set of dormant systems. Also, let $SS_i = \emptyset$ be the set of systems superior to system i in terms of x_i. Let $h^2 = 2\eta(n_0 - 1)$.

Obtain n_0 observations X_{in} and Y_{in} from each system $i = 1, 2, \ldots, k$.

For all i and $j \neq i$, compute the estimators $S^2_{Y_i}$ and $S^2_{X_{ij}}$.

Set the observation counters $r_i = n_0$ for all i and $r = n_0$.

Feasibility Check: For $i \in M \setminus D$ and $r_i = r$, if

$$\sum_{n=1}^{r_i}(Y_{in} - q) \leq -R(r_i; \epsilon, h^2, S^2_{Y_i}),$$

move i from M to F. For all $j \in M \cup F$ with $i \in SS_j$, eliminate j from M or F, delete SS_j, and remove j from D, if applicable. Else if

$$\sum_{n=1}^{r_i}(Y_{in} - q) \geq +R(r_i; \epsilon, h^2, S^2_{Y_i}),$$

eliminate i from M and any existing SS_j and delete SS_i. If $i \in SS_j$ and $j \in D$, remove j from D and let $r = \min\{r, r_j\}$.

Comparison: For each $i, j \in (M \cup F) \setminus D$ such that $i \neq j$, r_i or r_j is equal to r, and

$$\sum_{n=1}^{r} X_{in} \leq \sum_{n=1}^{r} X_{jn} - R(r; \delta, h^2, S^2_{X_{ij}}),$$

if $j \in F$, then eliminate i from M or F, delete SS_i, and for all $j' \in D \setminus \{i, j\}$ with $i \in SS_{j'}$, eliminate i from $SS_{j'}$, remove j' from D, and let $r = \min\{r, r_{j'}\}$; otherwise if $j \notin F$, then add index j to SS_i and i to D.

Stopping Rule: If $|M| = 0$ and $|F| = 1$, then stop and select the system whose index is in F as the best. If $|M| = 0$ and $|F| = 0$, then stop and report that there is no feasible system. Otherwise, for all systems $i \in (M \cup F) \setminus D$ such that

$r_i = r$, take one additional observation X_{i,r_i+1} and Y_{i,r_i+1} and set $r_i = r_i + 1$. Set $r = r + 1$. Then go to **Feasibility Check**.

We comment that it may be more efficient to let $r = \min_{j\in(M\cup F)\setminus D} r_j$ when i is eliminated in **Feasibility Check** or **Comparison** (e.g., if $r_i \ll r_j, \forall j \neq i$). This would prevent a situation where $r_i \neq r, \forall i \in (M \cup F) \setminus D$. We next prove the validity of the $\mathcal{D}_R$.

Theorem 2. *Under Assumptions 6, 8, and 9, the $\mathcal{D}_R$ procedure guarantees $P(CS) \geq 1 - \alpha$.*

Proof: Under Assumptions 8 and 9, Theorem 4 of Andradóttir and Kim [5] and its proof show that $\mathcal{AK}+$ guarantees $P(CS) \geq 1-\alpha$, in a manner satisfying Assumption 4. Moreover, $\mathcal{AK}+$ clearly satisfies Assumption 7. The result now follows from Theorem 1. □

4.3.3 The $\mathcal{D}_C$ Procedure

The $\mathcal{D}_C$ procedure is formed by applying the dormancy with catch-up framework to $\mathcal{AK}+$. This procedure is heuristic.

Procedure [Dormant with Catch-up $\mathcal{D}_C$]

Setup: Same as in $\mathcal{D}_R$.

Initialization: Same as in $\mathcal{D}_R$.

Feasibility Check: Same as in $\mathcal{D}_R$.

Comparison: For each $i, j \in (M \cup F) \setminus D$ such that $i \neq j$, $r_i = r_j = r$, and

$$\sum_{n=1}^{r} X_{in} \leq \sum_{n=1}^{r} X_{jn} - R(r; \delta, h^2, S^2_{X_{ij}}),$$

if $j \in F$, then eliminate i from M or F, delete SS_i, and for all $j' \in D \setminus \{i, j\}$ with $i \in SS_{j'}$, eliminate i from $SS_{j'}$, remove j' from D, and let $r = \min\{r, r_{j'}\}$; otherwise if $j \notin F$, then add index j to SS_i and i to D.

Stopping Rule: Same as in $\mathcal{D}_R$.

4.3.4 The $\mathcal{D}_A$ Procedure

The $\mathcal{D}_A$ procedure is formed by applying the dormancy with averages framework to $\mathcal{AK}+$. This procedure is also heuristic.

Procedure [Dormant with Averages $\mathcal{D}_A$]

Setup: Same as in $\mathcal{D}_R$.

Initialization: Same as in $\mathcal{D}_R$.

Feasibility Check: Same as in $\mathcal{D}_R$.

Comparison: For each $i, j \in (M \cup F) \setminus D$ such that $i \neq j$, r_i or r_j is equal to r, and

$$\frac{r}{r_i} \sum_{n=1}^{r_i} X_{in} \leq \frac{r}{r_j} \sum_{n=1}^{r_j} X_{jn} - R(r; \delta, h^2, S^2_{X_{ij}}),$$

if $j \in F$, then eliminate i from M or F, delete SS_i, and for all $j' \in D \setminus \{i, j\}$ with $i \in SS_{j'}$, eliminate i from $SS_{j'}$, remove j' from D, and let $r = \min\{r, r_{j'}\}$; otherwise if $j \notin F$, then add index j to SS_i and i to D.

Stopping Rule: Same as in $\mathcal{D}_R$.

4.3.5 Heuristic Modifications

We also introduce four types of heuristic modifications which can use any of the dormant procedures, $\mathcal{D}_R$, $\mathcal{D}_C$, or $\mathcal{D}_A$, as a basis.

$\mathcal{D}^T$ hopes to improve the efficiency of dormant algorithms by expanding eliminations past simple pairwise comparisons. In this algorithm, if system i is eliminated by a feasible system, we also eliminate all systems inferior to system i. Thus, we achieve a transitive effect. As most selection procedures are based on pairwise comparisons only, this is a heuristic step. Even under Assumption 5, the best system might be found inferior to an infeasible system i and then be eliminated with system i.

$\mathcal{D}^I$ modifies the continuation regions of the feasibility check and comparison steps to adjust for differences in means larger than the tolerance level or indifference-zone parameters. In particular, we use an adjusted tolerance level,

$\epsilon_i \equiv \max(\epsilon, |\frac{1}{n_0}\sum_{n=1}^{n_0}(Y_{in} - q)|)$ for each system i. We also adjust the indifference-zone parameter $\delta_{ij} \equiv \max(\delta, |\frac{1}{n_0}\sum_{n=1}^{n_0}(X_{in} - X_{jn})|)$. These modifications let us utilize first stage sample means to aid in the decision making process, an idea highlighted in Chen and Kelton [13]. These new tolerance level or indifference-zone parameters will allow the procedure to make decisions quicker, at some expense of observed PCS.

$\mathcal{D}^{I'}$ is a more conservative variant of $\mathcal{D}^I$. The tolerance levels and indifference-zone parameters are adjusted slightly, so that $\epsilon_i \equiv \max(\epsilon, |\frac{1}{n_0}\sum_{n=1}^{n_0}(Y_{in} - q)| - 2\sqrt{S_{Y_i}^2/n_0})$ and $\delta_{ij} \equiv \max(\delta, |\frac{1}{n_0}\sum_{n=1}^{n_0}(X_{in} - X_{jn})| - 2\sqrt{S_{X_{ij}}^2/n_0})$. By including the standard error in the indifference-zone computations, we hope to gain savings while preserving PCS.

The final modification, $\mathcal{D}^+$, utilizes the variance updating strategy of Kim and Nelson [31]. At fixed intervals, the procedure recomputes all variance estimates, $S_{Y_i}^2$ for feasibility check and $S_{X_{ij}}^2$ for comparison, utilizing all available samples. This recalculation also requires the modification of procedural parameters η and h^2 to account for the additional samples, due to more degrees of freedom.

We can also consider combinations of $\mathcal{D}^T$, $\mathcal{D}^I$ or $\mathcal{D}^{I'}$, and $\mathcal{D}^+$. The resulting modifications will feature the transitive property, sample-mean adjusted tolerance levels and indifference zones, and/or updates (for variance and possibly tolerance level and indifference zone means).

For our experimental study, any procedure featuring one or more modifications will be noted through the application of superscripts. For example, the dormant with averages procedure $\mathcal{D}_A$ combined with the conservative indifference-zone modification, $\mathcal{D}^{I'}$, and featuring updating of both mean and variance estimates, $\mathcal{D}^+$, will be expressed as $\mathcal{D}_A + \mathcal{D}^{I'} + \mathcal{D}^+ = \mathcal{D}_A^{I'+}$.

4.4 Experiments and Results

In this section, we illustrate the performance of our dormancy framework under various configurations. We describe our experimental setup in Section 4.4.1, followed by an exposition and analysis of our experimental results in Section 4.4.2.

The results were obtained based on 10,000 replications, while seeking a PCS of $1 - \alpha = 0.95$. For each setup, we consider k different systems. Of the k possible systems, we let b of them be desirable (clearly feasible), while a systems are acceptable. We also set $\delta = 1/\sqrt{n_0}$ and $\epsilon = 1/\sqrt{n_0}$, so that both the indifference-zone parameter and tolerance level will be equivalent to the first-stage standard deviation of a system with variance equal to 1 for both primary and secondary performance measures. We let $n_0 = 20$.

In addition to providing numerical results for $\mathcal{AK}+$ with or without dormancy, we include experimental results for the $\mathcal{AK}$ procedure of Andradóttir and Kim [5], which considers feasibility check and comparison in sequence. First, the algorithm determines each systems feasibility. Then, the algorithm selects the best among the feasible systems. This approach works well if the feasibility check is not difficult compared to the selection of the best system, but is heuristic. Still, $\mathcal{AK}$ achieves the nominal PCS in experiments and at times is more efficient than $\mathcal{AK}+$, so we include it in our analysis.

4.4.1 System Mean and Variance Configurations

We consider two configurations of means and three configurations of variances. These configurations are described in Sections 4.4.1.1 and 4.4.1.2, respectively.

4.4.1.1 Means Configurations

We use the difficult means (DM) configuration to test the validity of the algorithms. In the DM configuration, we make both selection and feasibility determination hard, by creating some slightly infeasible, but far superior systems. In addition, the inferior feasible systems will all be less favorable by only a slight amount. The DM configuration was considered previously by Andradóttir and Kim [5], and involves structuring the means as follows:

$$x_i = E[X_{in}] = \begin{cases} 0, & i = 1, 2, \ldots, b-1, \\ \delta, & i = b, \\ 0 & i = b+1, \ldots, b+a, \\ (i-1)\delta, & i = b+a+1, \ldots, k, \end{cases}$$

and

$$y_i = E[Y_{in}] = \begin{cases} -\epsilon, & i = 1, 2, \ldots, b, \\ 0, & i = b+1, \ldots, b+a, \\ \epsilon & i = b+a+1, \ldots, k, \end{cases}$$

where again δ is the indifference-zone parameter and ϵ is the tolerance level. We set the constraint level, q, to zero.

The monotone increasing means (MIM) configuration tests an algorithm's ability to quickly distinguish clearly inferior and/or infeasible systems. Since many of the system means are located a good distance away from the indifference zone, the algorithm should be able to make a decision more quickly. The following MIM configuration is also used by Andradóttir and Kim [5] with $q = 0$:

$$x_i = E[X_{in}] = \begin{cases} (i-1)\delta, & i = 1, 2, \ldots, b, \\ (b-2)\delta, & i = b+1, \ldots, b+a, \\ (i-1)\delta, & i = b+a+1, \ldots, k, \end{cases}$$

and

$$y_i = E[Y_{in}] = \begin{cases} -(b-i+1)\epsilon, & i = 1, 2, \ldots, b, \\ 0, & i = b+1, \ldots, b+a, \\ (i-b)\epsilon & i = b+a+1, \ldots, k, \end{cases}$$

Once again, we have a setup where infeasible systems have attractive primary performance measures. However, in this case, infeasible systems are not necessarily close to the constraint.

For all of our experiments (except Table 15), we set $a = 0$ and $b = \lceil \frac{k+1}{2} \rceil$, where $\lceil \cdot \rceil$ is the ceiling function. Andradóttir and Kim [5] show that these choices of a and b result in the smallest possible PCS for $\mathcal{AK}+$.

4.4.1.2 Variance Configurations and Correlation

To illustrate trends, we also consider several variance configurations for both the primary performance measure and the secondary constrained performance measure, denoted by $\sigma_{x_i}^2$ and $\sigma_{y_i}^2$, respectively. We generalize the variance setups of Andradóttir and Kim [5] and Kim and Nelson [30].

We let the factor f be a measure of the relative difficulties of feasibility check and comparison. The difficulties will be controlled through the values of $\sigma_{x_i}^2$ and $\sigma_{y_i}^2$, as high values $\sigma_{x_i}^2$ will indicate a hard comparison, while similarly high values for $\sigma_{y_i}^2$ signal a hard feasibility check. We utilize the factor f, so that when $f > 1$ feasibility check is generally harder than comparison and when $f < 1$ comparison is more difficult.

In particular, for the primary performance measure, a configuration with constant (CONST) variance has $\sigma_{x_i}^2 = \frac{1}{f}$ for all i. For increasing (INC) variance, $\sigma_{x_i}^2 = \frac{1}{f}(1 + (i-1)\delta)$ for all i. And in the decreasing (DEC) variance setup, $\sigma_{x_i}^2 = \frac{1}{f}/(1+(i-1)\delta)$ for all i. A similar pattern is used for the variances of the secondary performance measure, except δ and $\frac{1}{f}$ are replaced by ϵ and f. For each mean configuration, we consider as many as five variance configurations, namely CONST $\sigma_{x_i}^2$/CONST $\sigma_{y_i}^2$, INC $\sigma_{x_i}^2$/INC $\sigma_{y_i}^2$, INC $\sigma_{x_i}^2$/DEC $\sigma_{y_i}^2$, DEC $\sigma_{x_i}^2$/INC $\sigma_{y_i}^2$, and DEC $\sigma_{x_i}^2$/DEC $\sigma_{y_i}^2$.

In practice, there may be some correlation (positive or negative) between the primary and secondary performance measures. We induce several different values of correlation, denoted ρ, with $\rho \in \{-0.9, -0.6, -0.3, 0, 0.3, 0.6, 0.9\}$.

4.4.2 Results

We now present selected results of our experiments. In Section 4.4.2.1, we identify relatively favorable and unfavorable configurations for the performance of our dormancy approaches, showing how dormancy can provide substantial savings when feasibility check is difficult. We include an analysis of different correlation structures in Section 4.4.2.2, display our algorithms' observed PCS and efficiency under favorable and unfavorable configurations in Sections 4.4.2.3 and 4.4.2.4, respectively, and discuss the usefulness of our heuristic modifications in Section 4.4.2.5.

4.4.2.1 Difficult Feasibility Check or Comparison

Depending on the number of observations needed for feasibility check or comparison, the effectiveness of our new framework can vary considerably. To illustrate this characteristic, we include two figures that compare the performance of all five procedures $\mathcal{AK}$, $\mathcal{AK}+$, $\mathcal{D}_R$, $\mathcal{D}_C$, and $\mathcal{D}_A$ as the measure f of the difficulty of the feasibility check varies while $\rho = 0$. Figure 5 displays the number of required observations for the procedures under the DM configuration with CONST/CONST variance. Figure 6 shows the the number of required observations for all procedures under the MIM configuration with the same variance structure.

The figures show that when the variance of the secondary performance measure is much higher than the variance of the primary performance measure, the savings from dormancy are substantial. For example, when $f = 10$, we see up to 30% savings over $\mathcal{AK}+$, and the savings over $\mathcal{AK}$ are greater. Thus, the dormancy framework is a promising approach to handle hard feasibility check configurations.

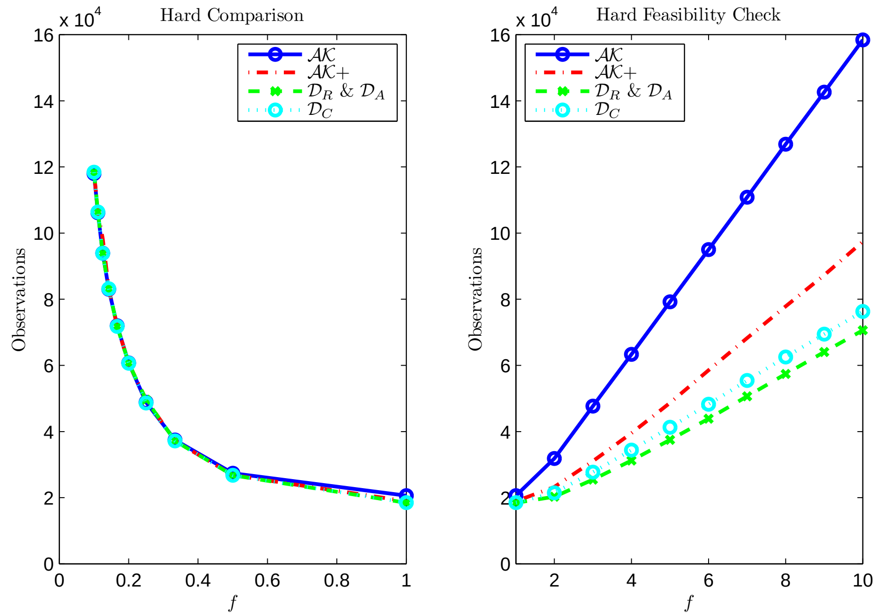

Figure 5: Number of needed observations as a function of f in a DM configuration with CONST $\sigma_{x_i}^2$/CONST $\sigma_{y_i}^2$, $k = 101$, $b = 51$, and $a = 0$.

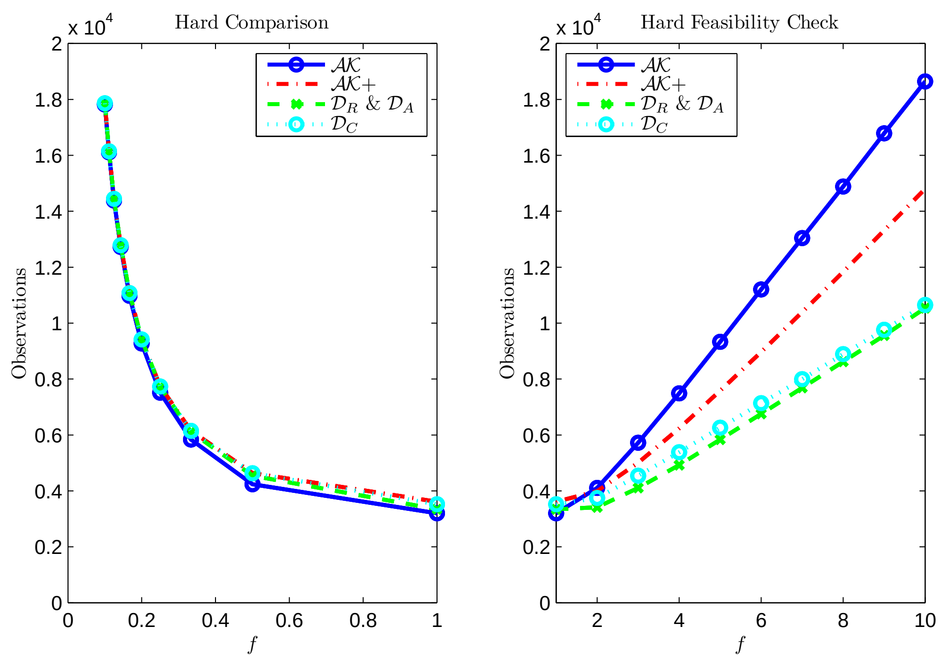

Figure 6: Number of needed observations as a function of f in a MIM configuration with CONST $\sigma_{x_i}^2$/CONST $\sigma_{y_i}^2$, $k = 101$, $b = 51$, and $a = 0$.

When comparison is relatively harder than feasibility check, as when $f \leq 1$, we see that all algorithms exhibit similar performance and differ by no more than 1%, so that they are indistinguishable in our figures. We note that while dormancy does not improve on $\mathcal{AK}$ under hard comparison, the small difference shows that dormancy is generally capable of good performance for a wide range of scenarios. Dormancy outperforms $\mathcal{AK}+$ in all cases.

From now on, we will let $f = 1$ and $f = 5$ be indicative of difficult comparison and difficult feasibility check, respectively. For difficult comparison, we choose $f = 1$ since all values of f less than one produce similar results among the procedures and $f = 1$ requires fewer overall observations.

In Table 15, we present the required number of samples and observed PCS for $k = 101$ systems where all systems are feasible ($b = k$) under DM and differing variance configurations with $f = 5$ and $\rho = 0$. Table 15 shows how the savings over $\mathcal{AK}$ and $\mathcal{AK}+$ from implementing dormancy can be almost limitless, if we evaluate a setup where many inferior systems are feasible, but feasibility check is hard. In particular, in the DEC/INC variance setup, we find 98% savings over $\mathcal{AK}$ and 80% savings over $\mathcal{AK}+$. Results for the MIM configuration are not included here, but similarly we can find large savings (up to 75%) with an implementation of dormancy over $\mathcal{AK}$ and $\mathcal{AK}+$.

Table 15: Average number of needed observations and observed PCS under the DM configuration with $k = 101$, $b = 101$, $a = 0$, and $f = 5$.

	$\sigma^2_{x_i}$ CONST	$\sigma^2_{y_i}$ CONST	$\sigma^2_{x_i}$ INC	$\sigma^2_{y_i}$ INC	$\sigma^2_{x_i}$ INC	$\sigma^2_{y_i}$ DEC	$\sigma^2_{x_i}$ DEC	$\sigma^2_{y_i}$ INC	$\sigma^2_{x_i}$ DEC	$\sigma^2_{y_i}$ DEC
	REP	PCS	REP	PCS	REP	PCS	REP	PCS	REP	PCS
$\mathcal{AK}$	79357	1.000	968506	1.000	72179	0.986	967779	1.000	11675	1.000
$\mathcal{AK}+$	24800	0.987	215966	0.989	59544	0.986	93448	0.999	2258	0.996
$\mathcal{D}_R$	6487	0.988	113926	0.993	59226	0.980	18120	0.999	2067	0.998
$\mathcal{D}_C$	14509	0.994	155666	0.991	59518	0.980	18394	0.999	2069	0.997
$\mathcal{D}_A$	6447	0.991	108205	0.992	59302	0.980	18137	0.999	2067	0.996

4.4.2.2 Performance under Correlation

We have seen that the difficulty of comparison and feasibility check plays a large part in the efficiency of applying dormancy, but this is not the case with correlation between performance measures. Tables 16 and 17 display the average number of needed observations to select a best system under varying correlation for the MIM configuration with constant variances and $f = 1$ or $f = 5$, respectively. These results are consistent with the results displayed in Andradóttir and Kim [5] for the underlying simultaneous procedure, $\mathcal{AK}$+. In particular, the algorithms perform equally well under all levels of correlation, and the percentage savings we gain with the dormancy remains roughly the same. The PCS values were also similar, satisfying the nominal PCS under all correlations. Therefore, the remainder of our experiments will feature $\rho = 0$.

Table 16: Average number of needed observations under the MIM configuration with CONST $\sigma_{x_i}^2$, CONST $\sigma_{y_i}^2$, $k = 101$, $b = 51$, $a = 0$, $f = 1$, and varying correlation, ρ.

	$\rho = -0.9$	$\rho = -0.6$	$\rho = -0.3$	$\rho = 0$	$\rho = 0.3$	$\rho = 0.6$	$\rho = 0.9$
$\mathcal{AK}$	20417	20706	20638	20625	20592	20396	20038
$\mathcal{AK}$+	18360	18761	18731	18940	19074	19075	18891
$\mathcal{D}_R$	18185	18496	18308	18451	18483	18494	18245
$\mathcal{D}_C$	18197	18519	18351	18499	18542	18546	18273
$\mathcal{D}_A$	18186	18498	18312	18455	18490	18498	18246

Table 17: Average number of needed observations under the MIM configuration with CONST $\sigma_{x_i}^2$, CONST $\sigma_{y_i}^2$, $k = 101$, $b = 51$, $a = 0$, $f = 5$, and varying correlation, ρ.

	$\rho = -0.9$	$\rho = -0.6$	$\rho = -0.3$	$\rho = 0$	$\rho = 0.3$	$\rho = 0.6$	$\rho = 0.9$
$\mathcal{AK}$	79355	79277	79496	79201	79142	79625	79031
$\mathcal{AK}$+	49067	50380	51047	51089	51132	51063	49767
$\mathcal{D}_R$	37481	37486	37512	37407	37440	37610	37224
$\mathcal{D}_C$	40751	41227	41542	41624	41458	41388	40539
$\mathcal{D}_A$	37455	37486	37508	37383	37434	37582	37204

4.4.2.3 Probability of Correct Selection

Tables 18 and 19 show the observed PCS of our new procedures under the DM configuration in the unfavorable setting $f = 1$ and the favorable setting $f = 5$, respectively. As k increases, the PCS of the procedures generally increases, probably due to the conservative nature of the bounds that ensure validity, so we discuss PCS for a relatively small number of systems, $k = 5$. These tables show that our algorithms display almost identical PCS results to $\mathcal{AK}+$. This is not surprising, as the algorithms feature similar elimination decisions. While we know $\mathcal{AK}+$ and $\mathcal{D}_R$ are valid, $\mathcal{D}_C$ and $\mathcal{D}_A$ perform equally well, always meeting the nominal PCS of 0.95.

Table 18: Average number of needed observations and observed PCS under the DM configuration with $k = 5$, $b = 3$, $a = 0$, and $f = 1$.

	$\sigma^2_{x_i}$ CONST	$\sigma^2_{y_i}$ CONST	$\sigma^2_{x_i}$ INC	$\sigma^2_{y_i}$ INC	$\sigma^2_{x_i}$ INC	$\sigma^2_{y_i}$ DEC	$\sigma^2_{x_i}$ DEC	$\sigma^2_{y_i}$ INC	$\sigma^2_{x_i}$ DEC	$\sigma^2_{y_i}$ DEC
	REP	PCS	REP	PCS	REP	PCS	REP	PCS	REP	PCS
$\mathcal{AK}$	576	0.969	805	0.970	586	0.968	670	0.971	433	0.971
$\mathcal{AK}+$	555	0.956	778	0.960	593	0.961	616	0.956	417	0.960
$\mathcal{D}_R$	545	0.957	760	0.961	591	0.961	580	0.957	412	0.960
$\mathcal{D}_C$	548	0.956	770	0.960	592	0.961	586	0.957	413	0.960
$\mathcal{D}_A$	545	0.956	762	0.960	591	0.961	581	0.957	412	0.960

Table 19: Average number of needed observations and observed PCS under the DM configuration with $k = 5$, $b = 3$, $a = 0$, and $f = 5$.

	$\sigma^2_{x_i}$ CONST	$\sigma^2_{y_i}$ CONST	$\sigma^2_{x_i}$ INC	$\sigma^2_{y_i}$ INC	$\sigma^2_{x_i}$ INC	$\sigma^2_{y_i}$ DEC	$\sigma^2_{x_i}$ DEC	$\sigma^2_{y_i}$ INC	$\sigma^2_{x_i}$ DEC	$\sigma^2_{y_i}$ DEC
	REP	PCS	REP	PCS	REP	PCS	REP	PCS	REP	PCS
$\mathcal{AK}$	2001	0.987	2887	0.986	1450	0.986	2879	0.987	1444	0.985
$\mathcal{AK}+$	1421	0.963	2170	0.959	942	0.957	2157	0.964	938	0.962
$\mathcal{D}_R$	1032	0.963	1695	0.959	661	0.958	1662	0.964	639	0.962
$\mathcal{D}_C$	1058	0.963	1780	0.957	703	0.958	1676	0.958	646	0.962
$\mathcal{D}_A$	1032	0.964	1695	0.959	661	0.959	1661	0.963	639	0.962

We also present Tables 20 and 21, which display the PCS and number of required observations to find the best system under 101 systems with 51 feasible for unfavorable and favorable difficulty ratios f under the MIM configuration. In this configuration, we expect PCS will be much higher due to the use of the IZ approach. Our results show that the three dormant algorithms provide an observed PCS much higher than 0.95 under MIM, but these results are consistent with previous works and similar to the observed PCS of $\mathcal{AK}$ and $\mathcal{AK}+$.

Table 20: Average number of needed observations and observed PCS under the MIM configuration with $k = 101$, $b = 51$, $a = 0$, and $f = 1$.

	$\sigma^2_{x_i}$ CONST	$\sigma^2_{y_i}$ CONST	$\sigma^2_{x_i}$ INC	$\sigma^2_{y_i}$ INC	$\sigma^2_{x_i}$ INC	$\sigma^2_{y_i}$ DEC	$\sigma^2_{x_i}$ DEC	$\sigma^2_{y_i}$ INC	$\sigma^2_{x_i}$ DEC	$\sigma^2_{y_i}$ DEC
	REP	PCS	REP	PCS	REP	PCS	REP	PCS	REP	PCS
$\mathcal{AK}$	3328	1.000	31484	0.999	18703	1.000	22491	0.999	2031	1.000
$\mathcal{AK}+$	3591	0.999	33947	0.999	18695	1.000	18686	0.999	2032	1.000
$\mathcal{D}_R$	3322	0.999	29686	0.999	18696	1.000	14527	0.999	2032	1.000
$\mathcal{D}_C$	3520	0.999	33500	0.999	18695	1.000	14619	0.999	2032	1.000
$\mathcal{D}_A$	3327	0.999	29851	0.999	18695	1.000	14527	0.999	2032	1.000

Table 21: Average number of needed observations and observed PCS under the MIM configuration with $k = 101$, $b = 51$, $a = 0$, and $f = 5$.

	$\sigma^2_{x_i}$ CONST	$\sigma^2_{y_i}$ CONST	$\sigma^2_{x_i}$ INC	$\sigma^2_{y_i}$ INC	$\sigma^2_{x_i}$ INC	$\sigma^2_{y_i}$ DEC	$\sigma^2_{x_i}$ DEC	$\sigma^2_{y_i}$ INC	$\sigma^2_{x_i}$ DEC	$\sigma^2_{y_i}$ DEC
	REP	PCS	REP	PCS	REP	PCS	REP	PCS	REP	PCS
$\mathcal{AK}$	9332	0.999	116274	0.999	4467	0.994	116585	0.998	2189	0.999
$\mathcal{AK}+$	7587	0.999	98107	0.998	5034	0.999	97106	0.997	2134	0.999
$\mathcal{D}_R$	5834	0.999	75121	0.998	4982	0.999	70831	0.997	2115	0.999
$\mathcal{D}_C$	6261	0.999	86147	0.998	5020	0.999	71544	0.998	2115	0.999
$\mathcal{D}_A$	5849	0.999	75161	0.998	4983	0.999	70931	0.997	2115	0.999

4.4.2.4 Number of Required Observations

In this section, we discuss the performance of the dormancy approaches in both favorable and unfavorable settings in terms of required samples. As shown in Tables 15 through 21, the three dormancy algorithms show at least a small amount of savings over $\mathcal{AK}+$ in most cases. The only exceptions are two configurations in Table 20 with DEC $\sigma^2_{y_i}$ where the performance of $\mathcal{AK}+$, $\mathcal{D}_R$, $\mathcal{D}_C$, and $\mathcal{D}_A$ is virtually identical. The size of the savings is usually much larger when $f = 5$ than when $f = 1$, and also when $\sigma^2_{y_i}$ is CONST or INC, but the performance of dormant approaches tends to vary for different mean and variance configurations and number of systems considered.

$\mathcal{D}_C$ often requires more observations than $\mathcal{D}_R$ or $\mathcal{D}_A$, since comparison cannot occur until the lagging system reaches the same number of observations as every other contending system. Thus, $\mathcal{D}_R$ and $\mathcal{D}_A$ are preferable to $\mathcal{D}_C$ in most situations. Due to the ability to compare systems at uneven sample sizes, $\mathcal{D}_R$ and $\mathcal{D}_A$ perform similarly.

While the dormant algorithms outperform $\mathcal{AK}+$ in most cases, we must also compare the performance of our new algorithms against the performance of $\mathcal{AK}$. When $f = 5$, we see savings in all configurations except MIM with INC $\sigma^2_{x_i}$/DEC $\sigma^2_{y_i}$. In the unfavorable case, $f = 1$, $\mathcal{D}_R$ and $\mathcal{D}_A$ improve on $\mathcal{AK}$ in all configurations except DM with INC $\sigma^2_{x_i}$/DEC $\sigma^2_{y_i}$ and MIM with DEC $\sigma^2_{x_i}$/DEC $\sigma^2_{y_i}$. These cases require just a small number observations for the feasibility check. For the MIM and DEC $\sigma^2_{x_i}$/DEC $\sigma^2_{y_i}$ configuration, the procedures require barely more than the first-stage samples to make decisions, allowing little room for improvement. The configurations with DEC $\sigma^2_{y_i}$ are where $\mathcal{AK}$ performs best, quickly removing infeasible systems. In a heuristic step, $\mathcal{AK}$ recalculates all parameters after the feasibility check, so reducing the number of contending systems after a fast feasibility check allows it to complete comparison efficiently. However, in these variance configurations, the extra required observations for procedures with dormancy is small, while ensuring validity in the case of $\mathcal{D}_R$.

4.4.2.5 Performance of Additional Heuristic Modifications

In this subsection, we demonstrate the performance of six heuristic treatments within the $\mathcal{D}_A$ algorithm. $\mathcal{D}_A$ is an appealing choice of a heuristic, combining good performance with limited storage requirements. We see similar results when the other procedures are applied with the heuristic modifications. The performance of the heuristics for the DM and MIM configurations can be seen in Tables 22 and 23, respectively, along with results for $\mathcal{D}_A$, as we seek to improve performance under the unfavorable configuration of $f = 1$. Not surprisingly, our experiments under $f = 5$, omitted here to conserve space, showed a larger benefit from the use of dormancy, but reached similar savings and PCS conclusions. We prefer including the results for $f = 1$ to document the ability of the heuristics to improve worst-case performance.

Table 22: Average number of needed observations and observed PCS under the DM configuration with $k = 25$, $b = 13$, $a = 0$, and $f = 1$.

	$\sigma^2_{x_i}$ CONST	$\sigma^2_{y_i}$ CONST	$\sigma^2_{x_i}$ INC	$\sigma^2_{y_i}$ INC	$\sigma^2_{x_i}$ INC	$\sigma^2_{y_i}$ DEC	$\sigma^2_{x_i}$ DEC	$\sigma^2_{y_i}$ INC	$\sigma^2_{x_i}$ DEC	$\sigma^2_{y_i}$ DEC
	REP	PCS	REP	PCS	REP	PCS	REP	PCS	REP	PCS
$\mathcal{D}_A$	3671	0.960	12650	0.960	6827	0.969	7436	0.964	1378	0.968
$\mathcal{D}_A^T$	3666	0.960	12639	0.960	6827	0.969	7439	0.961	1377	0.968
$\mathcal{D}_A^I$	2494	0.833	4991	0.403	3355	0.737	3383	0.507	1142	0.948
$\mathcal{D}_A^{I'}$	3631	0.958	12339	0.944	6659	0.961	7324	0.955	1373	0.968
$\mathcal{D}_A^+$	2988	0.958	10244	0.967	5736	0.968	5830	0.961	1156	0.962
$\mathcal{D}_A^{I+}$	1925	0.848	4592	0.564	2907	0.806	3067	0.655	947	0.943
$\mathcal{D}_A^{I'+}$	2810	0.948	9364	0.946	5264	0.959	5435	0.947	1124	0.959

Table 23: Average number of needed observations and observed PCS under the MIM configuration with $k = 25$, $b = 13$, $a = 0$, and $f = 1$.

	$\sigma^2_{x_i}$ CONST	$\sigma^2_{y_i}$ CONST	$\sigma^2_{x_i}$ INC	$\sigma^2_{y_i}$ INC	$\sigma^2_{x_i}$ INC	$\sigma^2_{y_i}$ DEC	$\sigma^2_{x_i}$ DEC	$\sigma^2_{y_i}$ INC	$\sigma^2_{x_i}$ DEC	$\sigma^2_{y_i}$ DEC
	REP	PCS	REP	PCS	REP	PCS	REP	PCS	REP	PCS
$\mathcal{D}_A$	1347	0.994	4746	0.993	3173	0.995	2446	0.996	614	0.994
$\mathcal{D}_A^T$	1346	0.994	4738	0.993	3173	0.995	2445	0.996	614	0.994
$\mathcal{D}_A^I$	902	0.970	2015	0.866	1490	0.919	1180	0.917	573	0.993
$\mathcal{D}_A^{I'}$	1201	0.994	3951	0.991	2805	0.994	2010	0.995	606	0.994
$\mathcal{D}_A^+$	1143	0.994	3836	0.995	2647	0.995	1980	0.994	583	0.993
$\mathcal{D}_A^{I+}$	777	0.976	1610	0.921	1232	0.952	948	0.944	550	0.992
$\mathcal{D}_A^{I'+}$	973	0.993	2666	0.993	1990	0.993	1393	0.991	573	0.992

The first heuristic, $\mathcal{D}_A^T$, does not provide much of an advantage in either the DM or MIM configurations, so therefore we do not consider combining it with any other modifications. The other heuristics show more promise. The two indifference-zone treatments, featured in $\mathcal{D}_A^I$ and $\mathcal{D}_A^{I'}$, show good improvement over $\mathcal{D}_A$. The aggressive $\mathcal{D}_A^I$ displays up to 60% savings in observations under DM and MIM over $\mathcal{D}_A$, but also exhibits a severe decay of PCS in most DM configurations. The more conservative $\mathcal{D}_A^{I'}$ posts modest improvements in DM configurations and 1% to 18% gains in MIM configurations over $\mathcal{D}_A$, while retaining a PCS close to or better than nominal.

The variance updating modification provides additional efficiency. Our implementation performs updating after every stage of sampling. A procedure utilizing only dormancy and variance-updating, $\mathcal{D}_A^+$, features consistent savings of about 20% over $\mathcal{D}_A$ (except for MIM with DEC $\sigma_{x_i}^2$/DEC $\sigma_{y_i}^2$), while experiencing better than nominal PCS for all configurations.

The combination of indifference-zone and variance updating modifications leads to the attractive heuristics, $\mathcal{D}_A^{I+}$ and $\mathcal{D}_A^{I'+}$. Tables 22 and 23 show that variance updating provides at least 10% savings when combined with other heuristics. In addition to the savings, variance updating achieves similar PCS results when the original approach had good PCS and significantly better PCS when the original procedure did not have good PCS. $\mathcal{D}_A^{I+}$ is the most efficient heuristic, but its PCS remains poor in the DM configuration. $\mathcal{D}_A^{I'+}$ improves on its non-updating counterpart, without significant PCS degradation. We note two trends in the relative performance of the heuristics across different variance configurations. Under DEC $\sigma_{x_i}^2$/DEC $\sigma_{y_i}^2$, the savings is limited, as all decisions are made almost immediately after the first stage and not much improvement is possible. Under a configuration with INC $\sigma_{x_i}^2$ or INC $\sigma_{y_i}^2$ (or both), the procedures featuring the indifference-zone modifications perform relatively better.

As heuristics, both $\mathcal{D}_A^{I+}$ and $\mathcal{D}_A^{I'+}$ are the promising options. The choice between the two falls to the user, as $\mathcal{D}_A^{I'+}$ provides overall efficiency and good PCS in DM, while $\mathcal{D}_A^{I+}$ trades PCS for quick decisions and efficiency in MIM.

4.5 Conclusion

We have introduced a new framework for simultaneous procedures in constrained R&S that select the best simulated system according to a primary performance measure while secondary performance measures satisfy constraints. Our dormancy framework involves the pausing of sampling for systems dominated (in terms of the primary performance measure) by any other system whose feasibility is undetermined. This modification saves simultaneously-running procedures from taking unnecessary observations, including observations used to determine the feasibility of inferior systems.

We present three approaches for implementing the dormancy framework, namely one that was proved statistically valid (dormant with recall) and two heuristics (dormant with catch-up and dormant with averages). These three approaches differ in the way they compare systems with uneven sample sizes, a situation that occurs when systems return from dormancy and must be compared again to contending systems. These dormancy approaches are combined with a known procedure for selection with one constraint to test the validity of the new framework and compare our new procedures to previously studied algorithms.

Our numerical results show that the percentage of observations saved by using dormancy can be very large when the feasibility determination is difficult. The three dormant procedures almost always outperform previously studied algorithms in the number of required observations, while displaying similar observed PCS. We recommend the use of dormancy with recall in almost all situations, while dormancy with averages is an apt alternative if storage is an issue. Dormancy may also be implemented with heuristic treatments that improve the efficiency of the overall procedure, with some possible loss of nominal PCS.

CHAPTER V

FULLY-SEQUENTIAL SELECTION PROCEDURES IN SIMULATIONS WITH MULTIPLE CONSTRAINTS

This chapter is most closely related to the work of Andradóttir and Kim [5] and Chapter 4. Andradóttir and Kim [5] introduced a fully-sequential, indifference-zone framework for constrained R&S consisting of two phases, i.e., feasibility check and selection of the best (comparison). These phases may be addressed either *sequentially* (the feasibility of each system is determined before comparison begins) or *simultaneously* (the feasibility check and comparison screening occur simultaneously after each additional sample). Andradóttir and Kim [5] and Chapter 4 proposed and analyzed several fully-sequential indifference-zone R&S procedures within this framework for independent systems with one constraint.

In this chapter, we elaborate on the framework of Andradóttir and Kim [5] and extend fully-sequential procedures to select the best system under any number of constraints and correlation across systems. This is a substantial extension of previous research that has only provided valid and heuristic procedures for independent systems and one constraint. Our procedures are combinations of valid feasibility check techniques for multiple constraints (e.g., Batur and Kim [7]) and valid comparison techniques. We show how to bring such techniques together to achieve statistically valid R&S procedures for multiple constraints.

R&S procedures should not allow the handling of multiple constraints to shift emphasis unduly towards feasibility verification. Thus, we consider how error should be allocated between the feasibility check and the comparison phases of the procedures. With the support of experimental results, we devise general, robust, and efficient error allocation rules as functions of the number of constraints for both simultaneously-running and sequential-running constrained selection procedures.

One topic of interest is the impact of multiple constraints on computational efficiency. Valid procedures for constrained R&S may require more observations to select the best feasible system than standard R&S due to a lengthy feasibility verification and the splitting of error between feasibility check and comparison. But within constrained R&S, there has been no study that we know of concerning the difficulty of satisfying multiple constraints. For example, what is the difference in the number of samples needed to find the best feasible system under one constraint or five constraints? We conduct an experimental study and show how many more (or less, somewhat surprisingly) observations a constrained R&S procedure can require when considering multiple constraints, while still guaranteeing a nominal PCS.

Our extension to allow correlation across systems is also significant, because it allows for the use of common random numbers (CRN). CRN have been shown to reduce the number of required samples in R&S procedures, see for example, Nelson and Matejcik [36], Chick and Inoue [18], and Kim and Nelson [30], and we seek to analyze the implementation and performance of CRN in constrained R&S. We will investigate when and how CRN should be used within constrained R&S procedures to reduce the observations necessary to make valid selection of the best feasible system, due to a more efficient comparison phase.

The chapter is organized as follows. Section 5.1 provides necessary background material, namely the formulation, notation, assumptions, and feasibility check approaches vital to our procedures. In Section 5.2, we present our procedures for multiple constraints and prove their validity in Section 5.3. In Section 5.4, we discuss issues associated with efficient implementation and provide analysis for the design of the procedures, specifically appropriate error allocation and the use of CRN. We discuss experiment setup and analyze experimental results in Section 5.5, and finally conclude the chapter in Section 5.6.

5.1 Background

This section details the background needed to formulate and analyze the general constrained R&S problem and procedures for solving it. In Section 5.1.1, we describe the problem formulation and indifference-zone approach to finding the best feasible system. Sections 5.1.2 and 5.1.3 detail notation and assumptions necessary for the feasibility check and comparison phases of our constrained R&S procedures and their validity. We also include two feasibility check procedures for multiple constraints in Section 5.1.4 that will be implemented in our general R&S procedures.

5.1.1 Formulation

Constrained R&S attempts to select the best system with respect to the mean of a primary performance measure in the presence of constraints on one or more secondary performance measures. Let $(X_{in}, Y_{i1n}, \ldots, Y_{isn})$ be the nth observation of the ith system for the primary performance measure and s secondary performance measures. The set of all possible systems is denoted $S = \{1, \ldots, k\}$.

We let $x_i = \mathrm{E}[X_{in}]$ and $y_{i\ell} = \mathrm{E}[Y_{i\ell n}]$ be the expected values of the primary and secondary constrained performance measures for each system $i \in S$ and constraints $\ell = 1, \ldots, s$. Our objective is to select the system with the best primary performance measure that also satisfies all of the constraints:

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

This objective is accomplished through the indifference-zone (IZ) approach. The IZ approach is extended to include both the comparison of primary performance measures and feasibility check of multiple secondary performance measures.

For the primary performance measure, we denote δ, the IZ parameter, to be the smallest distance that we consider significant. We are essentially indifferent among the feasible systems whose primary performance measures are within δ of each other. If x_i is found to be greater than x_j, then we say that system i is superior to system j (or equivalently system j is inferior to system i).

We also employ the IZ approach for each of the secondary performance measures, but in this case, the smallest significant distance is ϵ_ℓ, the tolerance level associated with the constraint ℓ. Any system with $y_{i\ell} \leq q_\ell - \epsilon_\ell$ for all $\ell = 1, \ldots, s$ is considered desirable. The set of all desirable systems is denoted S_D. Systems that have at least one mean secondary performance measure greater than q_ℓ (i.e., $y_{i\ell} \geq q_\ell + \epsilon_\ell$ for some ℓ) are unacceptable and infeasible, placing them in the set S_U. Systems that fall within the tolerance level of q_ℓ for some ℓ, so that $q_\ell - \epsilon_\ell < y_{i\ell} < q_\ell + \epsilon_\ell$, and below the tolerance level for the remaining constraints are acceptable and are placed in the set S_A. The goal is to identify a desirable or acceptable system whose primary performance measure value is no worse than an indifference zone away from that of the best desirable system.

5.1.2 Notation

To accurately ensure validity of the overall procedures, some notation must be described before we advance:

n_0 = the first stage sample size;

$S^2_{X_{ij}}$ = the sample variance of $\{X_{i1} - X_{j1}, \ldots, X_{in_0} - X_{jn_0}\}$;

$S^2_{Y_{i\ell}}$ = the sample variance of $\{Y_{i\ell 1}, \ldots, Y_{i\ell n_0}\}$ (the ℓth constraint of system i);

$\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, \ldots, \epsilon_s)^T$, $\epsilon_\ell \in \mathbb{R}^+$;

$\boldsymbol{q} = (q_1, q_2, \ldots, q_s)^T$, $q_\ell \in \mathbb{R}$;

$\boldsymbol{a} = (a_1, a_2, \ldots, a_s)^T$, $a_\ell \in \mathbb{R}^+$;

$\boldsymbol{Y}_{in} = (Y_{i1n}, Y_{i2n}, \ldots, Y_{isn})^T$;

$q^a = \boldsymbol{a}^T \boldsymbol{q}$;

$\epsilon^a = \boldsymbol{a}^T \boldsymbol{\epsilon}$;

$Y_{in}^a = \boldsymbol{a}^T \boldsymbol{Y}_{in}$;

$S_{Y_i^a}^2$ = the sample variance of $\{Y_{i1}^a, \ldots, Y_{in_0}^a\}$;

$R(r; b, c, d) = \max\{0, \frac{cd}{2b} - \frac{b}{2}r\}$, for $b, c, d \in \mathbb{R}^+$ and $b \neq 0$;

CS = the event that correct selection is made of the best feasible system, $x_{[b]}$, if a feasible system exists, given $x_{[b]} \geq x_i + \delta$ for all $i \in S_D \cup S_A$; if no feasible systems exist, all systems should be eliminated;

CS_i = the event that a good selection is made in comparison between inferior system i and the best feasible system, given $x_{[b]} \geq x_i + \delta$ for all $i \in S_D \cup S_A$;

CD_i = the event that a correct feasibility decision is made on system $i \in S$ (when $i \in S_A$ a feasible or infeasible decision are both correct);

β_1 = the nominal error of an individual feasibility check for one performance measure of one system;

β_2 = the nominal error of an individual comparison between two systems.

5.1.3 Assumptions for Validity

We need some assumptions about the data, the systems, and the feasibility check and comparison procedures.

Assumption 10. *For each $i = 1, 2, \ldots, k$,*

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

where $\stackrel{iid}{\sim}$ denotes independent and identically distributed, MN denotes multivariate normal, and Σ_i is the $(s+1) \times (s+1)$ covariance matrix of the vector $(X_{in}, Y_{i1n}, \ldots, Y_{isn})$.

Normally-distributed data is a common, not particularly restrictive, assumption. Law and Kelton [33] explain how normality can be achieved through within-replications averages or batch means. Commonly, primary and secondary performance measures will be correlated. Moreover, if CRN are used to simulate different systems, $(X_{in}, Y_{i1n}, \ldots, Y_{isn})$ and $(X_{jn}, Y_{j1n}, \ldots, Y_{jsn})$ will typically be correlated. Therefore, we allow correlation across systems and across performance measures.

Assumption 11. *For any $i \in S_D \cup S_A$ with $i \notin [b]$, $x_i \leq x_{[b]} - \delta$.*

This assumption allows only one possible best feasible system, as all systems that could be deemed feasible are inferior to $[b]$.

Assumption 12. *If the systems are simulated independently, the feasibility check phase guarantees $\mathsf{Pr}\{\cap_{i \in S'} CD_i\} \geq (1 - s\beta_1)^t$ for any $1 \leq t \leq k$ and any subset $S' \subseteq S$ with cardinality t.*

Assumption 13. *If the systems are simulated under CRN, the feasibility check phase guarantees $\mathsf{Pr}\{\cap_{i \in S'} CD_i\} \geq (1 - ts\beta_1)$ for any $1 \leq t \leq k$ and any subset $S' \subseteq S$ with cardinality t.*

We assume that the feasibility check procedure can correctly determine the feasibility of any number of systems with s constraints with a certain probability. Systems simulated under CRN require different bounds than independently-simulated systems.

Assumption 14. *If the systems are simulated independently, the comparison phase guarantees $\mathsf{Pr}\{\cap_{i \in S'} CS_i\} \geq (1 - \beta_2)^t$ for any $1 \leq t \leq k - 1$ and any subset S' of $\{i \in \{1, \ldots, k\} : x_i \leq x_{[b]} - \delta\}$ with cardinality t.*

Assumption 15. *If the systems are simulated under CRN, the comparison phase guarantees $\mathsf{Pr}\{\cap_{i \in S'} CS_i\} \geq (1 - t\beta_2)$ for any $1 \leq t \leq k - 1$ and any subset S' of $\{i \in \{1, \ldots, k\} : x_i \leq x_{[b]} - \delta\}$ with cardinality t.*

Given that we start with a set of systems inferior to system $[b]$, we require that pairwise comparison of this set with $[b]$ concludes with a selection of $[b]$ as the best with a certain probability. Again, the use of CRN requires different bounds than when considering independent systems. Several IZ-based comparison procedures, such $\mathcal{KN}$ of Kim and Nelson [30], satisfy Assumptions 14 and 15, but not all procedures are valid under CRN.

Assumption 16. *Observation n of system i (i.e., X_{in} and $Y_{i\ell n}$ for $\ell = 1, \ldots, s$) should not depend on the order the systems are sampled.*

This assumption is critical to the proof of any procedure that implements the dormancy framework (Chapter 4). This makes sure procedures with and without dormancy produce identical results.

5.1.4 Feasibility Check Procedures for Multiple Constraints

For the feasibility check phase under multiple constraints, we feature the fully-sequential procedures, $\mathcal{F}_{\mathcal{B}}^{\mathcal{I}}$ and $\mathcal{F}_{\mathcal{A}}^{\mathcal{I}}$, of Batur and Kim [7]. $\mathcal{F}_{\mathcal{B}}^{\mathcal{I}}$ is a fully–sequential feasibility check procedure for one or more constraints whose validity is established through the use of Bonferroni bounds. The $\mathcal{F}_{\mathcal{A}}^{\mathcal{I}}$ procedure features an artificial constraint, obtained by aggregation (or linear combination) of all secondary performance measures and their constrained levels. These procedures share a common setup, with additional steps to accommodate the aggregation in $\mathcal{F}_{\mathcal{A}}^{\mathcal{I}}$. To account for every system's status during the feasibility check, we utilize a set M of systems with undetermined feasibility, a set F of feasible systems, a set K_i that tracks the individual performance measures that have been deemed feasible for system i, for all $i \in S$, and a set A containing all systems whose feasibility according to the aggregate constraint has not been determined. We also denote the cardinality of a set as $|\cdot|$.

Section 5.1.4.1 provides a detailed implementation of $\mathcal{F}_{\mathcal{B}}^{\mathcal{I}}$. Section 5.1.4.2 features a similar description of $\mathcal{F}_{\mathcal{A}}^{\mathcal{I}}$ and a proof that the procedure satisfies Assumption 12.

5.1.4.1 Basic Feasibility Check for Multiple Constraints – $\mathcal{F}_{\mathcal{B}}^{\mathcal{I}}$

This approach involves sequential screening on every constrained performance measure. If a constraint is found to be violated, the system is declared infeasible. A system is declared feasible only if all constraints have been deemed feasible. Batur and Kim [7] proved that with $\beta_1 = \alpha/(ks)$ for correlated systems and $\beta_1 = (1-(1-\alpha)^{1/k})/s$ for independent systems, $\mathcal{F}_{\mathcal{B}}^{\mathcal{I}}$ guarantees the event that $S_D \subset F \subset S_D \cup S_U$ occurs with probability at least $1-\alpha$ when Assumption 10 holds. It also satisfies Assumptions 12 and 13 in this situation, a result of the proofs of Lemma 1 and Corollary 1 of Batur and Kim [7]. We present an instance of $\mathcal{F}_{\mathcal{B}}^{\mathcal{I}}$ when the continuation region parameter is set to $c = 1$.

Procedure [$\mathcal{F}_{\mathcal{B}}^{\mathcal{I}}$]

Setup: Select a first–stage sample size, $n_0 \geq 2$. Choose ϵ_ℓ and q_ℓ for $\ell = 1, 2, \ldots, s$. Let $\eta_1 = \frac{1}{2}((2\beta_1)^{-2/(n_0-1)} - 1)$ and $h_1^2 = 2\eta_1(n_0 - 1)$.

Initialization: Obtain n_0 observations from each constrained performance measure $\ell = 1, 2, \ldots, s$ from every system $i = 1, 2, \ldots, k$. For all i and ℓ, compute the estimators $S^2_{Y_{i\ell}}$. Set the observation counter $r_i = n_0$ and $K_i = \emptyset$ for $i = 1, 2, \ldots, k$. Let M contain all systems and $F = \emptyset$.

Feasibility Check: For all $i \in M$ and any $\ell \notin K_i$, if

$$\sum_{n=1}^{r_i}(Y_{i\ell n} - q_\ell) \geq R(r_i; \epsilon_\ell, h_1^2, S^2_{Y_{i\ell}}),$$

then remove i from M. Else if

$$\sum_{n=1}^{r_i}(Y_{i\ell n} - q_\ell) \leq -R(r_i; \epsilon_\ell, h_1^2, S^2_{Y_{i\ell}}),$$

then add ℓ to K_i. If $|K_i| = s$, remove i from M and add i to F.

Stopping Rule: If $|M| = 0$, then stop and return the set F as feasible systems. Otherwise, for all systems $i \in M$, take one additional observation $\boldsymbol{Y}_{i,r_i+1}$ and set $r_i = r_i + 1$. Then go to **Feasibility Check**.

5.1.4.2 *Accelerated Feasibility Check for Multiple Constraints – $\mathcal{F}_{\mathcal{A}}^{\mathcal{I}}$*

If two or more constrained performance measures are involved in the feasibility check, then it is possible to accelerate the feasibility determination for systems that are infeasible for multiple constraints. In particular, Batur and Kim [7] introduce an artificial, aggregate constraint to the feasibility check. This aggregate constraint adds some complexity, but can quickly eliminate systems that violate multiple constraints. The constraint is a linear function of all secondary performance measure samples, with weights $a_1, a_2, \ldots, a_s$ for each constraint $1, 2, \ldots, s$, respectively, and can only be used to declare systems infeasible. Batur and Kim [7] suggest the values $a_\ell = \prod_{\nu=1,\nu\neq\ell}^{s} \epsilon_\nu$, for $\ell = 1, 2, \ldots, s$, to minimize the area where systems may be infeasible for all constraints, but still not be found infeasible due by the aggregate constraint.

Batur and Kim [7] show that when $\beta_1 = \alpha/(k(s+1))$, $\mathcal{F}_{\mathcal{A}}^{\mathcal{I}}$ for correlated systems guarantees that the event $S_D \subset F \subset S_D \cup S_U$ occurs with probability at least $1-\alpha$ when Assumption 10 holds. The proof of Lemma 2 of Batur and Kim [7] shows that $\mathcal{F}_{\mathcal{A}}^{\mathcal{I}}$ satisfies Assumption 13 in this situation. At the end of the section, we strengthen Corollary 2 of Batur and Kim [7] whose proof shows $\mathcal{F}_{\mathcal{A}}^{\mathcal{I}}$ satisfies Assumption 12 for independently-simulated systems. Note that Batur and Kim [7] recommended defining β_1 heuristically, in terms of s instead of $s+1$ constraints (so that $\beta_1 = \alpha/(ks)$), to ensure that $\mathcal{F}_{\mathcal{A}}^{\mathcal{I}}$ performs more efficiently than $\mathcal{F}_{\mathcal{B}}^{\mathcal{I}}$, while showing only a small, practically insignificant loss in PCS. Our experiments will feature this aggressive definition of β_1. We present an instance of $\mathcal{F}_{\mathcal{A}}^{\mathcal{I}}$ when the continuation region parameter is set to $c = 1$.

Procedure $[\mathcal{F}_{\mathcal{A}}^{\mathcal{I}}]$

Setup: Same as in $\mathcal{F}_{\mathcal{B}}^{\mathcal{I}}$.

Initialization: Same as in $\mathcal{F}_{\mathcal{B}}^{\mathcal{I}}$, except for the following addition: Compute the estimator $S^2_{Y_i^a}$ for all i and let $A = S$.

Feasibility Check: Same as in $\mathcal{F}_{\mathcal{B}}^{\mathcal{I}}$ except for the following addition: If $i \in M \cap A$ and

$$\sum_{n=1}^{r_i}(Y_{in}^a - q^a) \geq R(r_i; \epsilon^a, h_1^2, S_{Y_i^a}^2),$$

the remove i from M and A. For $i \in M \cap A$ with

$$\sum_{n=1}^{r_i}(Y_{in}^a - q^a) \leq R(r_i; \epsilon^a, h_1^2, S_{Y_i^a}^2),$$

remove i from A.

Stopping Rule: Same as in $\mathcal{F}_{\mathcal{B}}^{\mathcal{I}}$, except for the following addition: If taking an additional observation from system $i \in M \cap A$, calculate Y_{i,r_i+1}^a.

We conclude this section on $\mathcal{F}_{\mathcal{A}}^{\mathcal{I}}$ with a short proof that shows $\mathcal{F}_{\mathcal{A}}^{\mathcal{I}}$ satisfies Assumption 12. Let $CD_{i\ell}$ and $ICD_{i\ell}$ denote the events of a correct and an incorrect decision of the feasibility of constraint ℓ of system i, respectively. Similarly, let CD_{ia} and ICD_{ia} denote the events of a correct and an incorrect decision of the feasibility of the aggregate constraint of system i, respectively. Andradóttir and Kim [5] have shown that $\mathsf{Pr}\{CD_{i\ell}\} = 1 - \mathsf{Pr}\{ICD_{i\ell}\} \geq 1 - \beta_1$ and $\mathsf{Pr}\{CD_i^a\} = 1 - \mathsf{Pr}\{ICD_i^a\} \geq 1 - \beta_1$. Batur and Kim [7] show that if systems are simulated independently and β_1 satisfies $(1 - s\beta_1)^k + (1 - \beta_1)^k = 1 - \alpha$, then $\mathsf{Pr}\{\cap_{i \in S} CD_i\} \geq 1 - \alpha$. We now strengthen this result and show that if $\beta_1 = (1 - (1 - \alpha)^{1/k})/(s + 1)$, Assumption 12 is satisfied.

Theorem 3. *If the systems are simulated independently and* $0 < \beta_1 < \frac{1}{s+1}$ *is chosen such that* $(1 - (s + 1)\beta_1)^k = 1 - \alpha$, *then* $\mathcal{F}_{\mathcal{A}}^{\mathcal{I}}$ *satisfies* $\mathsf{Pr}\{\cap_{i \in S'} CD_i\} \geq 1 - \alpha$ *for any* $1 \leq t \leq k$ *and any subset* $S' \subseteq S$ *with cardinality* t.

Proof: We have

$$\begin{aligned}
\mathsf{Pr}\{\cap_{i\in S'}CD_i\} &\geq \mathsf{Pr}\{(\cap_{i=1}^{t}\cap_{\ell=1}^{s} CD_{i\ell}) \cap (\cap_{i=1}^{t} CD_i^a)\} \\
&= \mathsf{Pr}\{\cap_{i=1}^{t}(\cap_{\ell=1}^{s} CD_{i\ell} \cap CD_i^a)\} \\
&= \prod_{i=1}^{t} \mathsf{Pr}\{\cap_{\ell=1}^{s} CD_{i\ell} \cap CD_i^a\} \\
&\geq \prod_{i=1}^{t}(1 - \sum_{\ell=1}^{s} \mathsf{Pr}\{ICD_{i\ell}\} - \mathsf{Pr}\{ICD_i^a\}) \\
&\geq \prod_{i=1}^{t}(1-(s+1)\beta_1) \text{ since } 1-(s+1)\beta_1 \geq 0 \\
&= (1-(s+1)\beta_1)^t \\
&\geq 1-\alpha,
\end{aligned}$$

where the second equality is due to the systems being simulated independently, the second inequality is due to the Bonferroni inequality, and the third inequality is due to the definition of β_1. □

5.2 General Constrained R&S Procedures

In this section, we present three procedures for constrained R&S with multiple constraints. The procedures generalize approaches of Andradóttir and Kim [5] and Chapter 4 that were originally formulated to compare independent systems with a single constrained performance measure. Our generalized algorithms incorporate a fully-sequential feasibility check for any number of constraints, and two of them allow for the valid incorporation of CRN.

In Section 5.2.1, we describe a sequentially-running procedure. Sections 5.2.2 and 5.2.3 feature simultaneously-running procedures.

5.2.1 A Sequentially-running Procedure – $\mathcal{HAK}$

In this section, we extend the $\mathcal{AK}$ procedure of Andradóttir and Kim [5]. This procedure performs feasibility check and comparison in sequence, first completing the feasibility check for all systems and constraints, then proceeding to select the best out of the surviving feasible systems. This procedure can be very efficient if feasibility is quickly determined and several infeasible systems are eliminated. Since feasibility check may be completed at different sample sizes for each system, the $\mathcal{SSM}$ procedure of Pichitlamken et al. [41] is used to perform comparison.

While the $\mathcal{AK}$ procedure is heuristic, Andradóttir and Kim [5] show that any degradation in PCS is very limited and its performance can be competitive. Therefore, it is a useful algorithm to extend to multiple constraints. Andradóttir and Kim [5] present a similar, less efficient sequentially-running procedure that utilizes restarting to make a valid selection of the best feasible system. This procedure can also be extended to include multiple constraints for independent and correlated systems, but the details fall outside the scope of this chapter.

Our $\mathcal{HAK}$ procedure for multiple constraints is described next.

Procedure [$\mathcal{HAK}$]

Setup: Select the overall confidence level $1/k \leq 1-\alpha < 1$ and choose the confidence levels for feasibility check $1-\alpha_1$ and comparison $1-\alpha_2$, where $\alpha_1+\alpha_2 = \alpha$. Use the **Setup** of the chosen feasibility check procedure, specifying $\beta_1 = (1-(1-\alpha_1)^{1/k})/s$ for independent systems and $\beta_1 = \alpha_1/(ks)$ for correlated systems.

Initialization: Use the **Initialization** of the chosen feasibility check procedure. In addition, obtain n_0 observations X_{in} from each system $i = 1, 2, \ldots, k$. For all i and $j \neq i$, compute the estimator $S^2_{X_{ij}}$.

Feasibility Check: Same as in the chosen feasibility check procedure.

Feasibility Stopping Rule: Same as in the chosen feasibility check procedure. In addition, for any system i receiving an additional sample, take X_{i,r_i+1}.

Setup for Comparison: If $|F| = 0$, conclude that there exist no feasible systems. If $|F| = 1$, then stop and select the system whose index is in F as the best. Otherwise, select $\delta > 0$. Let $\eta_2 = \frac{1}{2}((2\beta_2)^{-2/(n_0-1)} - 1)$, where $\beta_2 = \alpha_2/(|F| - 1)$, and $h_2^2 = 2\eta_2(n_0 - 1)$. Let $M = F$ now be the systems available for comparison. Set $r = n_0$.

Comparison: Considering any $i, j \in M$ such that $i \neq j$, if

$$\frac{r}{r_i}\sum_{n=1}^{r_i} X_{in} \leq \frac{r}{r_j}\sum_{n=1}^{r_j} X_{jn} - R(r; \delta, h_2^2, S_{X_{ij}}^2),$$

then eliminate i from M.

Comparison Stopping Rule: If $|M| = 1$, then stop and select the system whose index is in M as the best. Otherwise, for each system $i \in M$ with $r_i = r$, take one additional observation X_{i,r_i+1}, set $r_i = r_i + 1$ and $r = r + 1$. Then go to **Comparison**.

When representing the use of $\mathcal{F}_{\mathcal{B}}^{\mathcal{I}}$ or $\mathcal{F}_{\mathcal{A}}^{\mathcal{I}}$ with $\mathcal{HAK}$, we will denote the procedure as $\mathcal{HAK}(\mathcal{B})$ or $\mathcal{HAK}(\mathcal{A})$, respectively. The other combinations of procedures studied in this chapter (i.e., $\mathcal{HAK}+$ and $\mathcal{MD}_R$ with $\mathcal{F}_{\mathcal{B}}^{\mathcal{I}}$ and $\mathcal{F}_{\mathcal{A}}^{\mathcal{I}}$) are similarly denoted with this feasibility check marking.

5.2.2 A Simultaneously-running Procedure – $\mathcal{HAK}+$

Andradóttir and Kim [5] introduced the $\mathcal{AK}+$ procedure that performs feasibility check and comparison simultaneously after each additional stage of sampling. Thus systems are eliminated from contention after being found either infeasible or inferior to a feasible system. We now present our extension $\mathcal{HAK}+$. This simultaneously-running approach will show an improvement over $\mathcal{HAK}$ in configurations where feasibility check is slow to finish relative to comparison.

Procedure [$\mathcal{HAK}+$]

Setup: Select the overall confidence level $1/k \leq 1 - \alpha < 1$ and δ. Use the **Setup** of the chosen feasibility procedure. Let $\eta_2 = \frac{1}{2}((2\beta_2)^{-2/(n_0-1)} - 1)$.

Initialization: Use the **Initialization** of the chosen feasibility procedure. Also, let $SS_i = \emptyset$ be the set of superior systems to system i in terms of x_i. Let $h_2^2 =$

$2\eta_2(n_0-1)$. Obtain n_0 observations X_{in} from each system $i = 1, 2, \ldots, k$. For all i and $j \neq i$, compute the estimator $S^2_{X_{ij}}$. Set the observation counter $r = n_0$.

Feasibility Check: Same as in the chosen feasibility procedure. If found feasible, move i from M to F, and for all $j \in (M \cup F)$ with $i \in SS_j$, eliminate j from M or F and delete SS_j.

If found infeasible, eliminate i from M and any existing SS_j and delete SS_i.

Comparison: For each $i, j \in (M \cup F)$ such that $j \neq i$, $j \notin SS_i$, $i \notin SS_j$, and

$$\sum_{n=1}^{r} X_{in} \leq \sum_{n=1}^{r} X_{jn} - R(r; \delta, h_2^2, S_{X^2_{ij}}),$$

if $j \in F$, then eliminate i from M or F, delete SS_i, and remove i from any $SS_{j'}$; otherwise, if $j \notin F$, then add index j to SS_i.

Stopping Rule: If $|M| = 0$ and $|F| = 1$, then stop and select the system whose index is in F as the best. If $|M| = 0$ and $|F| = 0$ then stop and report that there is no feasible system. Otherwise, for all systems such that $i \in M \cup F$ and either $i \in M$ or $|SS_i| < |M|$, take one additional observation $(X_{i,r_i+1}, \boldsymbol{Y}_{i,r_i+1})$, set $r = r + 1$, and then $r_i = r$. Then go to **Feasibility Check**.

In Section 5.3, we will prove $\mathcal{HAK}+$ to be valid for independently simulated systems and correlated systems. The approach for choosing valid values of β_1 and β_2 is different for the two types of sampling, as we will detail further in Section 5.3 (see equations (13) and (15), as well as Remark 1 below).

5.2.3 A Simultaneously-running Procedure with Dormancy – $\mathcal{MD}_R$

The dormant with recall procedure, $\mathcal{D}_R$, of Chapter 4 is a more aggressive simultaneous constrained R&S procedure. Like $\mathcal{AK}+$, it can safely eliminate a system if it is found infeasible or inferior to another feasible system. The dormancy framework adds an additional condition, halting sampling from all systems found inferior to any system in contention with feasibility yet undetermined. This allows the procedure to avoid sampling from inferior systems and to compare and test for feasibility of the most promising systems first. A dormant system returns to contention if its superior system is eliminated.

The starting and stopping of sampling for dormant systems creates uneven sample sizes during the procedure, a difficulty overcome in $\mathcal{D}_R$ by storing past observations. The recall of past data allows the procedure to compare systems at an equal number of samples via the comparison procedure $\mathcal{KN}$ of Kim and Nelson [30]. In this section, we extend the statistically valid $\mathcal{D}_R$ procedure to multiple constraints, resulting in the $\mathcal{MD}_R$ procedure. While using summary statistics may save computational overhead (particularly memory needed for storage and time necessary to recall data), concerns about the validity of dormant algorithms with summary statistics were presented by Chapter 4. The heuristic procedures featuring dormancy, namely the dormant with catch-up and dormant with averages algorithms, can be extended in a similar fashion as $\mathcal{D}_R$, but this falls outside the scope of the current chapter.

Procedure [$\mathcal{MD}_R$]

Setup: Same as in $\mathcal{HAK}+$.

Initialization: Same as in $\mathcal{HAK}+$, except we also set $D = \emptyset$, where D denotes the set of dormant systems.

Feasibility Check: Same as in the chosen feasibility check procedure except feasibility is only checked for $i \in M \setminus D$ with $r_i = r$. If i is feasible, move i from M to F. For all $j \in M \cup F$ with $i \in SS_j$, eliminate j from M or F, delete SS_j, and remove j from D, if applicable. Else, if i is found infeasible, eliminate i from M and any existing SS_j and delete SS_i. If $i \in SS_j$ and $j \in D$, remove j from D and let $r = \min\{r, r_j\}$.

Comparison: For each $i, j \in (M \cup F) \setminus D$ such that $j \neq i$, r_i or r_j is equal to r, and

$$\sum_{n=1}^{r} X_{in} \leq \sum_{n=1}^{r} X_{jn} - R(r; \delta, h_2^2, S_{X_{ij}}^2),$$

if $j \in F$, then eliminate i from M or F, delete SS_i, and for all $j' \in D \setminus \{i, j\}$ with $i \in SS_{j'}$, eliminate i from $SS_{j'}$, remove j' from D, and let $r = \min\{r, r_{j'}\}$; otherwise if $j \notin F$, then add index j to SS_i and i to D.

Stopping Rule: If $|M| = 0$ and $|F| = 1$, then stop and select the system whose index is in F as the best. If $|M| = 0$ and $|F| = 0$, then stop and report that there is no feasible system. Otherwise, for all systems $i \in (M \cup F) \setminus D$ such that $r_i = r$, take one additional observation $(X_{i,r_i+1}, \boldsymbol{Y}_{i,r_i+1})$ and set $r_i = r_i + 1$. Set $r = r + 1$. Then go to **Feasibility Check**.

As for $\mathcal{HAK}+$, we will prove $\mathcal{MD}_R$ to be valid for both independently simulated systems and correlated systems. Valid choices of β_1, β_2 are discussed in Section 5.3 (see equations (13) and (15), as well as Remark 1 below).

5.3 Validity of Algorithms

We present $\mathcal{HAK}+$ and $\mathcal{MD}_R$ as statistically valid algorithms for general constrained R&S of independent or correlated systems. Sections 5.3.1 and 5.3.2 feature validity proofs for these two simultaneously-running procedures with independent and correlated systems, respectively. The proofs are presented while implementing $\mathcal{F}_{\mathcal{B}}^{\mathcal{I}}$ as the feasibility check procedure.

Remark 1. The use of $\mathcal{F}_{\mathcal{A}}^{\mathcal{I}}$ only requires an additional constraint within the proofs (i.e., $s + 1$ constraints rather than s constraints), as is clear from from Theorem 3 for independently simulated systems and from Lemma 2 of Batur and Kim [7] for correlated systems. Thus, Lemmas 1 and 2 and Theorems 4 through 7 hold for $\mathcal{F}_{\mathcal{A}}^{\mathcal{I}}$, as long as s is replaced by $s + 1$ in the statement of these results.

5.3.1 Validity of $\mathcal{HAK}+$ and $\mathcal{MD}_R$ for Independent Systems

To prove the validity of $\mathcal{HAK}+$ and $\mathcal{MD}_R$, we begin with the following lemma.

Lemma 1. *Under Assumptions 10, 11, 12, and 14, a simultaneously-running procedure for independently simulated systems guarantees*

$$\mathsf{Pr}\{CS\} \geq (1 - s\beta_1)^j + (1 - s\beta_1) + (1 - \beta_2)^{k-j-1} - 2 \tag{11}$$

when $|S_U| = j < k$ *and*

$$\mathsf{Pr}\{CS\} \geq (1 - s\beta_1)^k$$

when $S_U| = k$.

Proof: This proof is similar to the proof of Lemma 2 of Andradóttir and Kim [5]. Let A^* be the event that all systems $i \in S_U$ will eventually be eliminated by being declared infeasible. Let B^* be the event that system $[b]$ is declared feasible and all systems $i \in (S_D \cup S_A) \setminus \{b\}$ will be eventually eliminated by being declared inferior to system $[b]$. Then

$$\begin{aligned}
\mathsf{Pr}\{CS\} &= \mathsf{Pr}\{\text{all } i \in S_U \text{ and all } i \in (S_D \cup S_A) \text{ with } x_i \leq x_{[b]} - \delta \text{ are eliminated}\} \\
&\geq \mathsf{Pr}\{A^* \cap B^*\} \\
&\geq \mathsf{Pr}\{A^*\} + \mathsf{Pr}\{B^*\} - 1.
\end{aligned}$$

Now,

$$\begin{aligned}
\mathsf{Pr}\{A^*\} &= \mathsf{Pr}\{CD_i \text{ for all } i \in S_U\} \\
&\geq (1 - s\beta_1)^j \text{ (by Assumption 12).}
\end{aligned}$$

This proves the results when $|S_U| = k$. If $j = |S_U| < k$, then

$$\begin{aligned}
\mathsf{Pr}\{B^*\} &= \mathsf{Pr}\{CD_{[b]} \cap (CS_i \text{ for all } i \in (S_D \cup S_A) \text{ with } i \neq [b])\} \text{ (by Assumption 11)} \\
&\geq \mathsf{Pr}\{CD_{[b]}\} + \mathsf{Pr}\{\cap_{i \in (S_D \cup S_A) \setminus \{b\}} CS_i\} - 1 \\
&\geq (1 - s\beta_1) + \mathsf{Pr}\{\cap_{i \in (S_D \cup S_A) \setminus \{b\}} CS_i\} - 1 \text{ (by Assumption 12)} \\
&\geq (1 - s\beta_1) + (1 - \beta_2)^{k-j-1} - 1 \text{ (by Assumptions 11 and 14).}
\end{aligned}$$

All together, we have

$$\mathsf{Pr}\{CS\} \;\geq\; (1-s\beta_1)^j + (1-s\beta_1) + (1-\beta_2)^{k-j-1} - 2$$

when $j = |S_U| < k$, which concludes the proof. $\square$

Lemma 1 does not specify how to choose β_1 and β_2 for our procedure. There are many valid values of β_1 and β_2 that cause the right-hand side (RHS) of equation (11) to be greater than $1-\alpha$, but we would prefer the largest possible values for β_1 and β_2 to make our procedures efficient. Since $|S_U|$ may not be known at the time of initialization, we must also address how the RHS of equation (11) changes in j.

Remark 2. The lower bound $(1-s\beta_1)^k$ on $\mathsf{Pr}\{CS\}$ in Lemma 1 when $|S_U| = k$ satisfies

$$\begin{aligned}(1-s\beta_1)^k &= (1-s\beta_1)^{k-1} - (1-s\beta_1)^{k-1}s\beta_1\\ &\geq (1-s\beta_1)^{k-1} - s\beta_1,\end{aligned}$$

and $(1-s\beta_1)^{k-1} - s\beta_1$ is the value of the RHS of equation (11) when $j = k-1$. Therefore, the smallest lower bound on $\mathsf{Pr}\{CS\}$ in Lemma 1 is always achieved for $j = |S_U| < k$.

We provide one method that could be used to choose non-dominated values β_1 and β_2. The key to this approach is the choice of a parameter, e, that is the ratio of error for a complete feasibility check for one system to the error of a comparison between two systems, so that $e = s\beta_1/\beta_2$. For any choice of e, we can simplify the RHS of (11) and find a valid value of β_2. In particular, equation (11) now yields

$$\mathsf{Pr}\{CS\} \;\geq\; (1-e\beta_2)^j + (1-e\beta_2) + (1-\beta_2)^{k-j-1} - 2, \tag{12}$$

for $j \in |S_U| < k$. Since $j = |S_U|$ is unknown, we must find values of $\beta_2 \in [0, \min\{1, 1/e\}]$ such that the RHS of equation (12) is no smaller than $1-\alpha$ for all $j \in \{0, 1, \ldots, k-1\}$. Note that for a fixed value of $j \in \{0, 1, \ldots, k-1\}$, the RHS of (12) monotonically decreases from 1 to below 0 as β_2 increases from 0 to

$\min(1, 1/e)$. Thus, for any value of $j \in \{0, 1, \ldots, k-1\}$, there exists a value of β_2 such that the RHS of (12) is equal to $1 - \alpha$, which can be solved numerically.

Given that for all $j \in \{0, 1, \ldots, k-1\}$, a value of β_2 can be found to set the RHS of (12) equal to $1-\alpha$, one can iterate through all values of $j \in \{0, 1, \ldots, k-1\}$ to find the minimum β_2. The minimum β_2 would ensure that the lower bound on $\mathsf{Pr}\{CS\}$ exceeds $1 - \alpha$ for all $j \in \{0, 1, \ldots, k\}$, and then β_1 is calculated via the ratio e. This is one approach to supply values of β_1 and β_2 that satisfy Theorems 4 and 5. The choice of the parameter e will be addressed in Section 5.4.1 below.

We note that if $e = 1$, then $s\beta_1 = \beta_2$ and the value of $j \in [0, k-1]$ that minimizes the RHS of (12) is $j^* = (k-1)/2$. Therefore, a value of β_2 that guarantees the nominal PCS can be found by solving the equation $\beta_2 + 2[1 - (1 - \beta_2)^{(k-1)/2}] = \alpha$ (note that the left-hand side of this equation increases from 0 to 3 as β_2 increases from 0 to 1, so there is always a solution).

Theorem 4. *Under Assumptions 10 and 11 with independently simulated systems, $\mathcal{HAK}+$ implemented with $\mathcal{F}_{\mathcal{B}}^{\mathcal{I}}$ guarantees*

$$\mathsf{Pr}\{CS\} \geq 1 - \alpha$$

when

$$(1 - s\beta_1)^j + (1 - s\beta_1) + (1 - \beta_2)^{k-j-1} - 2 \geq 1 - \alpha \text{ for all } j \in \{0, 1, \ldots, k-1\}. \quad (13)$$

Proof: If the feasibility check procedure $\mathcal{F}_{\mathcal{B}}^{\mathcal{I}}$ is implemented under Assumption 10, Assumption 12 is satisfied as shown in the proof of Corollary 1 of Batur and Kim [7]. $\mathcal{HAK}+$ utilizes the $\mathcal{KN}$ procedure of Kim and Nelson [30] for comparison, so independence of the primary performance measure observations across systems, Assumptions 10 and 11, and the proof of Theorem 2 of Kim and Nelson [30] are sufficient to show that $\mathcal{HAK}+$ satisfies Assumption 14.

Since $\mathcal{HAK}+$ satisfies Assumptions 12 and 14, the result now follows from Lemma 1, Remark 2, and the fact that $|S_U|$ is unknown. $\square$.

Theorem 5. *Under Assumptions 10, 11, and 16 with independently simulated systems, $\mathcal{MD}_R$ implemented with $\mathcal{F}_{\mathcal{B}}^{\mathcal{I}}$ guarantees*

$$\mathsf{Pr}\{CS\} \geq 1 - \alpha$$

when equation (13) *is satified.*

Proof: Theorem 1 of Chapter 4 shows that when dormancy with recall is applied to a valid simultaneously-running procedure for constrained R&S and three conditions are satisfied, the resulting procedure with dormancy is also valid.

The first condition is that the validity of the original simultaneous procedure is proved by ensuring that with probability no smaller than $1 - \alpha$, the best system $[b]$ is declared feasible and all other systems in S would eventually be either declared infeasible or eliminated by that particular system (if they are not eliminated by another system first). This is true of $\mathcal{HAK}+$, see Theorem 4 and the proof of Lemma 1.

The second condition is identical to Assumption 16, so this condition is met. The third condition is that feasibility check and comparison parameters for both the original procedure and the new procedure under dormancy, such as indifference-zone parameters and variance estimates depend only on first-stage samples for each system and do not change as a function of the systems remaining in contention. $\mathcal{HAK}+$ with $\mathcal{F}_{\mathcal{B}}^{\mathcal{I}}$ satisfies this condition as well. Thus $\mathcal{MD}_R$ is valid and $\mathsf{Pr}\{CS\} \geq 1 - \alpha$. □

5.3.2 Validity of $\mathcal{HAK}+$ and $\mathcal{MD}_R$ for Correlated Systems

Correlation of data across systems requires a slightly different proof approach. While the feasibility check procedures of Batur and Kim [7] guarantee a desired probability of correct feasibility decision under correlation, the same is not true of all comparison techniques under correlation. Fortunately, the underlying comparison procedure of $\mathcal{HAK}+$ and $\mathcal{MD}_R$ is $\mathcal{KN}$ of Kim and Nelson [30], which is valid under correlation with certain parameter adjustments.

We present a lemma that will help prove the validity of $\mathcal{HAK}+$ and $\mathcal{MD}_R$.

Lemma 2. *Under Assumptions 10, 11, 13, and 15, a simultaneous procedure for correlated systems under s constraints guarantees*

$$\mathsf{Pr}\{CS\} \geq 1 - (j+1)s\beta_1 - (k-j-1)\beta_2 \tag{14}$$

when $|S_U| = j < k$ *and*

$$\mathsf{Pr}\{CS\} \geq 1 - ks\beta_1$$

when $|S_U| = k$.

Proof: Let A^* and B^* be defined as in the proof of Lemma 1. As in the proof of Lemma 1, we have

$$\mathsf{Pr}\{CS\} \geq \mathsf{Pr}\{A^*\} + \mathsf{Pr}\{B^*\} - 1,$$

and when $j = |S_U| < k$, then

$$\mathsf{Pr}\{B^*\} \geq \mathsf{Pr}\{CD_{[b]}\} + \mathsf{Pr}\{\cap_{i \in (S_D \cup S_A)\setminus\{[b]\}} CS_i\} - 1.$$

Moreover,

$$\begin{aligned} \mathsf{Pr}\{A^*\} &= \mathsf{Pr}\{CD_i \text{ for all } i \in S_U\} \\ &\geq (1 - js\beta_1) \text{ (Assumption 13).} \end{aligned}$$

This proves the result when $|S_U| = k$. When $j = |S_U| < k$, then

$$\begin{aligned} \mathsf{Pr}\{B^*\} &\geq (1 - s\beta_1) + \mathsf{Pr}\{\cap_{i \in (S_D \cup S_A)\setminus\{[b]\}} CS_i\} - 1 \text{ (Assumption 13)} \\ &\geq 1 - s\beta_1 - (k-j-1)\beta_2 \text{ (Assumptions 11 and 15).} \end{aligned}$$

Now, we have

$$\begin{aligned} \mathsf{Pr}\{CS\} &\geq (1 - js\beta_1) + (1 - s\beta_1 - (k-j-1)\beta_2) - 1, \\ &\geq 1 - (j+1)s\beta_1 - (k-j-1)\beta_2 \end{aligned}$$

when $j = |S_U| < k$, which concludes the proof. □

Remark 3. The lower bound $1-ks\beta_1$ on $\mathsf{Pr}\{CS\}$ in Lemma 2 when $|S_U|=k$ satisfies

$$1-ks\beta_1 \geq 1-(k+1)s\beta_1,$$

and $1-(k+1)s\beta_1$ is the value of the RHS of (14) when $j=k-1$. Therefore, the smallest lower bound on $\mathsf{Pr}\{CS\}$ in Lemma 2 is always achieved for $j=|S_U|<k$.

Since $j=|S_U|$ may be any integer between 0 and k, we must ensure $\mathsf{Pr}\{CS\} \geq 1-\alpha$ for any $j \in \{0,1,\ldots,k-1\}$. Recall that $e=s\beta_1/\beta_2$. We assume e is given. Then one can see easily that the value, $j^* \in \{0,1,\ldots,k-1\}$, that minimizes $1-[(j+1)e+(k-j-1)]\beta_2$ depends on e:

$$j^* = \begin{cases} k-1, & \text{if } e \geq 1, \\ 0, & \text{if } e < 1. \end{cases}$$

Note that for $e=1$, the RHS of (14) does not depend $j \in \{0,1,\ldots,k-1\}$. Thus, to achieve $\mathsf{Pr}\{CS\} \geq 1-\alpha$ for all values of j, a simultaneous procedure would require:

$$\beta_1 = \begin{cases} e\alpha/(sk), & \text{if } e \geq 1, \\ e\alpha/(se+s(k-1)), & \text{if } e < 1, \end{cases}$$

and

$$\beta_2 = \begin{cases} \alpha/(ek), & \text{if } e \geq 1, \\ \alpha/(e+(k-1)), & \text{if } e < 1. \end{cases}$$

This is one approach to provide values of β_1 and β_2 to satisfy Theorems 6 and 7 below.

Theorem 6. *Under Assumptions 10 and 11 with correlated systems such that $(X_{1n}, X_{2n}, \ldots, X_{kn})$ are iid multivariate normal with a positive definite covariance matrix, $\mathcal{HAK}+$ implemented with $\mathcal{F}_{\mathcal{B}}^{\mathcal{I}}$ guarantees*

$$\mathsf{Pr}\{CS\} \geq 1-\alpha$$

when

$$(j+1)s\beta_1+(k-j-1)\beta_2 \leq \alpha \text{ for all } j \in \{0,1,\ldots,k-1\}. \tag{15}$$

Proof: If the feasibility check procedure $\mathcal{F}_{\mathcal{B}}^{\mathcal{I}}$ is implemented under Assumption 10, Assumption 13 is satisfied as shown in the proof of Lemma 1 of Batur and Kim [7]. While independence is no longer assumed, $\mathcal{KN}$ can still be used to make valid decisions. The proof of Theorem 1 of Kim and Nelson [30] shows that Assumption 15 is met under Assumptions 10 and 11 in the presence of a positive definite covariance matrix. Since Assumptions 13 and 15 hold, the result now follows from Lemma 2, Remark 3, and the fact that $|S_U|$ is not known. □

Theorem 7. *Under Assumptions 10, 11, and 16 with correlated systems such that $X_{1n}, X_{2n}, \ldots, X_{kn}$ are iid multivariate normal with a positive definite covariance matrix, $\mathcal{MD}_R$ implemented with $\mathcal{F}_{\mathcal{B}}^{\mathcal{I}}$ guarantees*

$$\mathsf{Pr}\{CS\} \geq 1 - \alpha$$

when equation (15) is satisfied.

Proof: Again, we show $\mathcal{MD}_R$ and $\mathcal{HAK}+$ satisfy the conditions for Theorem 1 of Chapter 4. Theorem 6 and the proof of Lemma 2 ensure the statistical validity of $\mathcal{HAK}+$ and the first condition of Theorem 1. The second condition is satisfied by Assumption 16. The last condition holds as in Theorem 5. Thus, $\mathcal{MD}_R$ is statistically valid, and hence $\mathsf{Pr}\{CS\} \geq 1 - \alpha$. □

5.4 Efficient Design of Procedures for Constrained R&S

In this section, we consider some issues that directly affect the efficiency of $\mathcal{HAK}$, $\mathcal{HAK}+$, and MD_R, namely the choice of error parameters and use of CRN to induce a positive correlation between systems. These issues are addressed in Sections 5.4.1 and 5.4.2, respectively.

5.4.1 Error Allocation

The choice of parameters that govern the allowable error in the comparison and feasibility check phases of a constrained R&S procedure can be critical to efficiency.

For sequential procedures, the user chooses the parameters α_1 and α_2 as the total amount of error for the feasibility check and comparison phases, respectively. For simultaneous procedures, β_1 and β_2 equal the error of individual feasibility checks and comparisons, respectively. In this section, we provide experimental results that suggest efficient choices for α_1 and α_2 in sequentially-running procedures and β_1 and β_2 in simultaneously-running procedures.

If the relative difficulties of feasibility check and comparison were known, some efficiency could be gained by tuning the error allocation correctly. However, since details about the means and variances of the primary and secondary performance measures are often not known, robust strategies for error allocation are useful.

For our analysis of error allocation, we consider two procedures, $\mathcal{HAK}$ and $\mathcal{HAK}+$, as representatives of sequential and simultaneous constrained R&S procedures, respectively. $\mathcal{MD}_R$ is an application of the dormancy framework to $\mathcal{HAK}+$, so we expect these two procedures to produce similar results. Andradóttir and Kim [5] suggest an allocation for the procedures under one constraint, namely $\alpha_1 = \alpha_2 = \alpha/2$ for sequentially-running procedures and $\beta_1 = \beta_2$ for simultaneously-running procedures. However, when $s > 1$, it is unclear how this strategy should be extended. In particular, two reasonable choices are equal error allocation between feasibility check and comparison and equal error allocation for each (primary or secondary) performance measure tested (giving more error to the feasibility check phase to handle multiple constraints).

We use the $\mathcal{F}_{\mathcal{B}}^{\mathcal{I}}$ procedure for feasibility check, because it is a simple and valid approach. We discuss the advantages of $\mathcal{F}_{\mathcal{A}}^{\mathcal{I}}$ in Section 5.5, but do not want to add its complexity to the analysis of our results.

Section 5.4.1.1 details the setup featured in all of our numerical experiments. Section 5.4.1.2 provides the study of error allocation within sequential procedures. Section 5.4.1.3 investigates error allocation within simultaneous procedures.

5.4.1.1 Experimental Setup

To evaluate the relative performance of the allocations and our procedures, we tested the procedures under differing ratios of errors (α_1/α_2 or β_1/β_2) for various configurations of means and variances. These mean and variance configurations attempt to provide analogous results and analysis to the experimental studies of similar R&S studies, for example, Kim and Nelson [30] and Andradóttir and Kim [5] among others. We test the procedures for several combinations of means and variances under 10,000 macro-replications.

We set the first-stage sample size, n_0, to 20 and indifference-zone parameters and tolerance levels to $\delta = \epsilon_\ell = 1/\sqrt{20}$ for all $\ell = 1, 2, \ldots, s$, equal to the sample standard deviation of the initial average when samples have a variance of 1. We set a nominal PCS of $1 - \alpha = 0.95$. We include no acceptable systems, so that $S_A = \emptyset$, because Andradóttir and Kim [5] show that the presence of such systems does not significantly affect the experimental results. Finally, we set the constraint levels, q_ℓ, to zero.

We introduced an additional consideration for multiple constraints, specifically the number of violated constraints v for an infeasible system. The value of v is crucial to how quickly a feasibility check completes. For our tests, we will feature a varying number of constraints s and $v \in \{1, s\}$, with $v = 1$ implying a hard feasibility check and $v = s$ creating an easier feasibility check.

We now describe our mean configurations. The following monotone increasing configuration (MIM) of means, which emulates a common situation when many systems are either infeasible or inferior, was used:

$$x_i = E[X_{ij}] = (i - 1)\delta, \quad i = 1, 2, \ldots, k,$$

and

$$y_{i\ell} = E[Y_{i\ell j}] = \begin{cases} -(b-i+1)\epsilon, & i = 1, 2, \ldots, b, \\ (i-b)\epsilon & i = b+1, \ldots, k, \text{ and } \ell = 1, 2, \ldots, v, \\ -(i-b)\epsilon & i = b+1, \ldots, k, \text{ and } \ell = v+1, v+2, \ldots, s, \end{cases}$$

where b is the number of feasible systems.

In some of the experiments, we include the difficult means configuration (DM), which attempts to test the validity of the procedures by assigning system means in a challenging setup. In this configuration, there are $b-1$ feasible systems that are only slightly inferior (by an indifference-zone parameter) to the best system and the remaining superior systems are only slightly infeasible (by a tolerance level). More specifically, in the DM configuration,

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

and

$$y_{i\ell} = E[Y_{i\ell n}] = \begin{cases} -\epsilon_\ell, & i = 1, 2, \ldots, b, \\ \epsilon_\ell, & i = b+1, \ldots, k \text{ and } \ell = 1, 2, \ldots, v, \\ -\epsilon_\ell & i = b+1, \ldots, k \text{ and } \ell = v+1, v+2, \ldots, s, \end{cases}$$

where again δ is the indifference-zone parameter and ϵ_ℓ is the tolerance level.

We also examine a combination of variance configurations to test the robustness of the procedures when the relative difficulty of feasibility check and comparison varies. These configurations involve low (L) and high (H) variances $\sigma^2_{x_i}$ and $\sigma^2_{y_{i\ell}}$ of the primary and secondary performance measures. For simplicity, all secondary performance measures $\ell = 1, 2, \ldots, s$ are assigned identical variances. High variance results in either $\sigma^2_{x_i} = 5$ or $\sigma^2_{y_{i\ell}} = 5$, whereas low variance causes $\sigma^2_{x_i} = 1$ or $\sigma^2_{y_{i\ell}} = 1$. For our experiments, we consider three variance configurations, i.e., low $\sigma^2_{x_i}$ and low

$\sigma^2_{y_{i\ell}}$ (L/L), high $\sigma^2_{x_i}$ and low $\sigma^2_{y_{i\ell}}$ (H/L), and low $\sigma^2_{x_i}$ and high $\sigma^2_{y_{i\ell}}$ (L/H). Variances lower than 1 produce valid decisions quickly, for both feasibility check and comparison. For the sake of space, we do not consider other variance configurations.

Practically, correlation across primary and secondary performance measures should be expected, but Andradóttir and Kim [5] show that such correlation will not significantly affect the results of valid procedures. Hence, we will not revisit the topic in this chapter, and obtain primary and secondary performance measure samples independently for all systems.

Similarly, Batur and Kim [7] show that correlation across only secondary performance measures does not largely affect the performance of the feasibility check procedure $\mathcal{F}^{\mathcal{I}}_{\mathcal{B}}$. However, strong negative correlation across secondary performance measures can induce faster completion times in $\mathcal{F}^{\mathcal{I}}_{\mathcal{A}}$, while strong positive correlation reduces the effectiveness of the aggregate constraint. We expect similar conclusions would be found here. Therefore, we do not address this topic, and assume that secondary performance measure samples are independent of one another.

Additionally, the effects of correlation across systems should be considered. Unless expressed explicitly, we will consider independent systems, but CRN will be examined in Section 5.5.4 below.

5.4.1.2 Error Allocation for Sequential Procedures

We provide two tables addressing error allocation for $\mathcal{HAK}(\mathcal{B})$ in the MIM configuration with $k = 10$ and $v = 1$. Table 24 displays the average number of required observations for different error allocations for $\mathcal{HAK}(\mathcal{B})$ as we change the number of feasible systems b, while holding all other configuration settings steady. We consider a set of allocations, expressed by the ratio of α_1 to α_2. Table 24 shows that in a sequential procedure, as more systems are found feasible, more comparisons are necessary, requiring more error in the form of a higher α_2 to perform efficiently. In the

L/H case where the feasibility check is relatively difficult, additional error should be shifted towards α_1 for best performance. While the best allocation changes, the 1:1 allocation always appears to have performance close to the best, especially in the L/L and H/L configurations.

Table 24: Average number of required observations under the MIM configuration with $k = 10$ systems, $s = 2$ constraints, b feasible systems, and $v = 1$ infeasible constraints for the $\mathcal{HAK}(\mathcal{B})$ procedure with the given ratio of α_1 to α_2. (The best allocation is shown in bold and the recommended allocation is shown in a box.)

	L/L Variance Config.			H/L Variance Config.			L/H Variance Config.		
α_1/α_2	$b=1$	$b=5$	$b=9$	$b=1$	$b=5$	$b=9$	$b=1$	$b=5$	$b=9$
4	**484**	919	1018	762	922	1030	**2330**	**2947**	**2578**
3	491	892	979	747	895	988	2367	2991	2612
2	502	865	930	730	867	940	2434	3069	2681
3/2	513	852	904	727	853	912	2494	3144	2743
1	$\boxed{532}$	$\boxed{\mathbf{841}}$	$\boxed{875}$	$\boxed{\mathbf{726}}$	$\boxed{\mathbf{845}}$	$\boxed{883}$	$\boxed{2597}$	$\boxed{3276}$	$\boxed{2858}$
2/3	557	844	860	740	846	868	2727	3439	3000
1/2	578	874	**857**	754	856	**864**	2838	3577	3121
1/3	611	896	862	783	878	868	3015	3800	3313
1/4	638	916	872	808	899	876	3155	3975	3467

We present Table 25 where the number of feasible systems and parameters are fixed, but the number of constraints, s, varies. In this table, the larger numbers of constraints s tend to require more error devoted to feasibility check (α_1) for best performance. In the L/H configuration, it is advisable to allow more error for feasibility check. As in Table 24, we see that a 1:1 ratio is clearly advisable for low numbers of constraints, especially in the L/L and H/L variance configurations, but is still efficient for all cases. Table 25 also illustrates that the cost of additional constraints, measured by the number of required observations, grows sub-linearly with respect to s within each variance configuration and α_1/α_2 ratio.

Our experimental results suggest that for sequential procedures, an allocation rule that distributes error evenly between the feasibility check and comparison works well. While it makes sense to focus on L/L results as one may not know in advance

Table 25: Average number of required observations under the MIM configuration with $k = 10$ systems, s constraints, $b = 5$ feasible systems, and $v = 1$ infeasible constraints for the $\mathcal{HAK}(\mathcal{B})$ procedure with the given ratio of α_1 to α_2. (The best allocation is shown in bold and the recommended allocation is shown in a box.)

	L/L Variance Config.				H/L Variance Config.				L/H Variance Config.			
α_1/α_2	$s=1$	$s=2$	$s=4$	$s=8$	$s=1$	$s=2$	$s=4$	$s=8$	$s=1$	$s=2$	$s=4$	$s=8$
4	860	919	974	1091	867	922	980	1092	**2291**	**2947**	**3729**	**4671**
3	831	892	954	1077	839	895	961	1079	2318	2991	3781	4731
2	797	865	936	**1067**	804	867	942	**1072**	2377	3069	3878	4840
3/2	780	852	**930**	1069	784	853	**937**	1074	2435	3144	3966	4938
1	$\boxed{765}$	$\boxed{\mathbf{841}}$	$\boxed{933}$	$\boxed{1085}$	$\boxed{768}$	$\boxed{\mathbf{845}}$	$\boxed{940}$	$\boxed{1089}$	$\boxed{2544}$	$\boxed{3276}$	$\boxed{4113}$	$\boxed{5112}$
2/3	**759**	844	949	1114	**763**	846	956	1119	2678	3439	4300	5332
1/2	762	853	968	1143	**763**	856	974	1147	2788	3577	4458	5515
1/3	773	874	1004	1194	773	878	1011	1197	2973	3800	4713	5811
1/4	787	896	1036	1234	787	899	1042	1239	3121	3975	4916	6052

the relative difficulty of feasibility check versus comparison, the 1:1 rule is fairly robust to differing numbers of constraints, numbers of feasible systems, and variance configurations. We observe similar results for $v = s$ and in the DM mean configuration when $v \in \{1, s\}$, but omit these results. All displayed choices of allocation depart no more than 15% from the best.

5.4.1.3 Error Allocation for Simultaneous Procedures

In this section, we consider the simultaneously-running $\mathcal{HAK}+(\mathcal{B})$ procedure. Here we seek efficient and robust choices of β_1 and β_2. As in Section 5.4.1.2, we focus on performance, measured by the number of required observations, as the ratio of the two parameters changes.

Table 26 shows the average number of needed observations for a configuration with $k = 10$ systems, two constraints, one infeasible constraint for infeasible systems, and a varying number of feasible systems. We see that a ratio of $\beta_1/\beta_2 = 1/2$ is the best or close to the best for most scenarios. This result is analogous to our findings for $\mathcal{HAK}(\mathcal{B})$, as $\beta_1 = \beta_2/s$ corresponds to approximately equivalent error allocation for feasibility check and comparison.

Table 26: Average number of required observations under the MIM configuration with $k = 10$ systems, $s = 2$ constraints, b feasible systems, and $v = 1$ infeasible constraints for the $\mathcal{HAK}+(\mathcal{B})$ procedure with the given ratio of β_1 to β_2. (The best allocation is shown in bold and the recommended allocation is shown in a box.)

	L/L Variance Config.			H/L Variance Config.			L/H Variance Config.		
β_1/β_2	$b=1$	$b=5$	$b=9$	$b=1$	$b=5$	$b=9$	$b=1$	$b=5$	$b=9$
4	475	1171	1264	1003	1177	1281	2213	3018	2819
3	474	1126	1211	967	1132	1225	2213	2987	2774
2	473	1067	1136	884	1071	1150	2211	2943	2711
3/2	473	1026	1020	919	1029	1097	2212	2913	2670
1	**472**	971	957	841	974	1029	2211	2874	2615
2/3	473	920	**917**	800	924	964	**2210**	**2828**	2565
1/2	$\boxed{473}$	$\boxed{\mathbf{890}}$	$\boxed{944}$	$\boxed{\mathbf{773}}$	$\boxed{\mathbf{892}}$	$\boxed{\mathbf{922}}$	$\boxed{2215}$	$\boxed{3049}$	$\boxed{\mathbf{2538}}$
1/3	507	924	969	804	927	949	2412	3223	2719
1/4	534	953	989	831	955	972	2563	3369	2868

To test the performance of $\beta_1 = \beta_2/s$ as the number of constraints increases, again we use $k = 10$ systems, five feasible systems, and one infeasible constraint for each infeasible system. Table 27 shows that for s constraints, the appropriate allocation is $\beta_1/\beta_2 = 1/s$, except for the L/H variance configuration and $s = 8$ where it is close to optimal. So, again, even allocation between feasibility check and comparison is preferable. As for $\mathcal{HAK}(\mathcal{B})$, we see sublinear growth in the number of required observations as the number of constraints increases. The $\mathcal{HAK}+(\mathcal{B})$ procedure's performance depends heavily on the correct choice of error allocation ratio, however, the observed best allocation does not stray much from $\beta_1/\beta_2 = 1/s$. Again, we observe similar results in the DM mean configuration and when $v \in \{1, s\}$, but omit these results.

The sequential algorithm $\mathcal{HAK}(\mathcal{B})$, with poor choices of error allocation requiring 10% more samples than optimal in Table 24 appears to be less robust to poor error allocation than the simultaneous procedure, $\mathcal{HAK}+(\mathcal{B})$, where poor choices cost at most 33% more observations than the optimal in Table 26. However, if we were to translate the ratios of β_1 and β_2 into the scope of α_1 and α_2 (complete error of

Table 27: Average number of required observations under the MIM configuration with $k = 10$ systems, s constraints, $b = 5$ feasible systems, and $v = 1$ infeasible constraints for the $\mathcal{HAK}+(\mathcal{B})$ procedure with the given ratio of β_1 to β_2. (The best allocation is shown in bold and the recommended allocation is shown in a box.)

	L/L Variance Config.				H/L Variance Config.				L/H Variance Config.			
β_1/β_2	$s=1$	$s=2$	$s=4$	$s=8$	$s=1$	$s=2$	$s=4$	$s=8$	$s=1$	$s=2$	$s=4$	$s=8$
3/2	837	1026	1208	1459	847	1029	1216	1459	2261	2913	3667	4550
1	$\boxed{\mathbf{787}}$	971	1153	1401	$\boxed{\mathbf{793}}$	974	1161	1401	$\boxed{\mathbf{2222}}$	2874	3635	4523
2/3	812	920	1102	1346	818	924	1111	1344	2399	2843	3605	4500
1/2	834	$\boxed{\mathbf{890}}$	1067	1308	840	$\boxed{\mathbf{892}}$	1077	1308	2546	$\boxed{\mathbf{2828}}$	3585	4487
1/3	869	924	1023	1259	875	927	1031	1259	2757	3049	3565	4471
1/4	896	953	$\boxed{\mathbf{994}}$	1226	901	955	$\boxed{\mathbf{1004}}$	1225	2914	3223	$\boxed{\mathbf{3560}}$	4462
1/5	919	978	1016	1201	923	980	1026	1200	3044	3369	3702	4456
1/6	939	998	1036	1181	942	1000	1048	1181	3153	3491	3830	4452
1/7	956	1016	1055	1164	958	1019	1067	1165	3207	3597	3946	**4449**
1/8	971	1032	1072	$\boxed{\mathbf{1153}}$	974	1034	1084	$\boxed{\mathbf{1154}}$	3291	3694	4041	$\boxed{4454}$
1/9	984	1047	1088	1166	987	1050	1099	1167	3366	3778	4130	4537

feasibility check and comparison), the ratios of β_1 and β_2 correspond with much larger ratios of the overall feasibility check and comparison phases, α_1 and α_2. In particular, the overall error for each phase can be approximated by the values of $\alpha_1 = ks\beta_1$ and $\alpha_2 = (k-1)\beta_2$. Thus with $k = 10$ and $s = 2$, β_1/β_2 ratios from 4 to 1/4 correspond to α_1/α_2 ratios from roughly 80/9 to 5/9.

Ultimately, without knowing any properties of the systems ahead of time, the efficiency of an allocation that splits error evenly between feasibility check and comparison is relatively robust to the various possible configurations of feasible systems and number of constraints. This allocation takes slightly different forms in sequential and simultaneous procedures, but is either optimal or close to optimal in all of our experiments, especially in the L/L variance configuration. When implementing the recommended ratio into our simultaneous procedures, we also note that $\beta_1 = \beta_2/s$ corresponds to $e = 1$, a special case that leads to easily solvable valid values of β_1 and β_2 (see Section 5.3).

5.4.2 Considering Common Random Numbers

In this section, we discuss the use of CRN in constrained R&S procedures to improve the efficiency of comparison. CRN could be useful with our procedures, particularly $\mathcal{HAK}+$ and $\mathcal{MD}_\mathcal{R}$, proven to be valid under correlation across systems. This section also suggests why the implementation of CRN within procedures such as $\mathcal{HAK}$ that compare systems with unequal sample sizes may not provide valid PCS.

In Section 5.4.2.1, we take a closer look at a difficulty in comparing correlated systems with uneven sample sizes. Section 5.4.2.2 provides an analysis of the required correlation to make CRN advantageous in constrained R&S with comparison at even sample sizes.

5.4.2.1 Decisions under Correlation

While the independent simulation of systems is suitable for many problems, just a small amount of positive correlation can significantly improve the efficiency of fully-sequential R&S procedures. This positive correlation will reduce the variance of the difference of samples from two systems, allowing the comparison of the systems to be completed sooner. Usually, this positive correlation is created through the use of CRN (Law and Kelton [33]). The increase in efficiency comes at a cost, in that some comparison procedures may not make valid decisions for correlated systems, and Bonferroni bounds are used in setting up the procedures to ensure validity of the selection (see Theorems 6 and 7).

Two of the constrained R&S procedures for multiple constraints, i.e., $\mathcal{HAK}+$ and $\mathcal{MD}_R$, are valid with correlation across systems, as is shown in Section 5.3. The reason for this is that $\mathcal{HAK}+$ and $\mathcal{MD}_R$ always compare systems at equal sample sizes, and do so with the $\mathcal{KN}$ procedure. The $\mathcal{KN}$ procedure makes statistically valid decisions for both independent and correlated systems, as proven in Kim and Nelson [30], and thus satisfies Assumption 15.

However, procedures that compare systems at unequal sample sizes, like $\mathcal{HAK}$ and its underlying selection procedure $\mathcal{SSM}$ of Pichitlamken et al. [41], may not provide adequate PCS results. While $\mathcal{SSM}$ was proven to be statistically valid for comparison of independent systems, it does not ensure valid decisions under correlation.

The problem lies with obtaining good variability estimates of the process observed by $\mathcal{SSM}$ under correlation. Fully-sequential procedures, like $\mathcal{SSM}$, use the quantity

$$S^2_{X_{ij}} = \frac{1}{n_0 - 1} \sum_{n=1}^{n_0} (X_{in} - X_{jn} - [\bar{X}_i - \bar{X}_j])^2 \tag{16}$$

as an estimate of the variance of the difference between two systems, i and j, where $\bar{X}_i$ and $\bar{X}_j$ are the first-stage sample means for system i and j, respectively. This variance estimate allows the procedure to utilize the benefits of positive correlation, but only accurately represents the variability of the difference of sums under a common sample size r:

$$\sum_{n=1}^{r} (X_{in} - X_{jn}). \tag{17}$$

However, under unequal sample sizes $r_i < r_j$ with $r = \min(r_i, r_j)$ and high correlation, the statistic

$$\frac{r}{r_i} \sum_{n=1}^{r_i} X_{in} - \frac{r}{r_j} \sum_{n=1}^{r_j} X_{jn} \tag{18}$$

used by $\mathcal{SSM}$ can have a much higher variance than computed by $S^2_{X_{ij}}$, as the variability is driven by the lagging system's data points.

Let ρ_x denote the correlation across primary performance measure samples. Figure 7 draws sample paths of the difference of sums with equal sample sizes (17) and unequal sample sizes (18) under a high correlation, namely $\rho_x = 0.95$. It is clear that the unequal sums experience a much higher variability. The underestimation in $S^2_{X_{ij}}$ of the variability of the unequal sums (18) could lead to incorrect decisions. Thus, without adjustments to the comparison algorithm within $\mathcal{HAK}$, it is unclear that one can use this procedure to compare correlated systems. Further study of this topic falls outside the scope of this chapter.

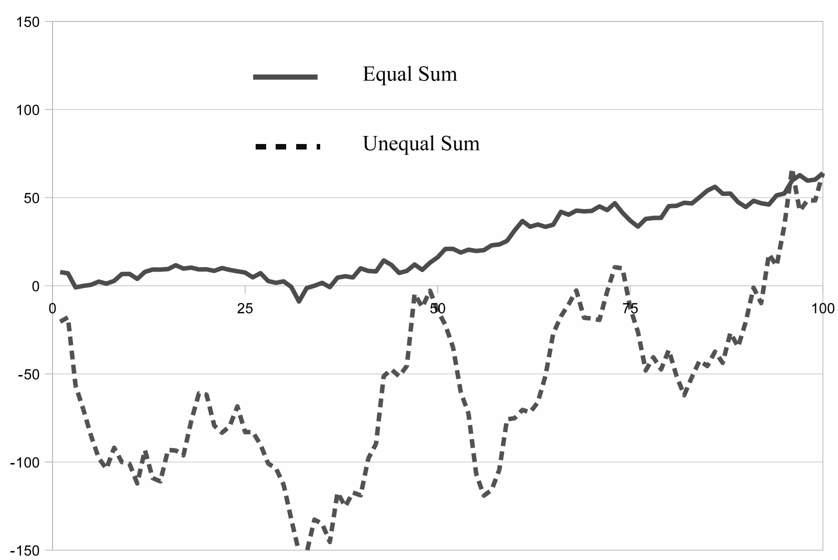

Figure 7: Sample paths of the difference of equal sums $\sum_{n=1}^{r}(X_{1n}-X_{2n})$ and unequal sums $\sum_{n=1}^{r} X_{1n} - \frac{r}{100}\sum\limits_{n=1}^{100} X_{2n}$ under $\rho_x = 0.95$.

5.4.2.2 Required Correlation

Having shown that $\mathcal{HAK}+$ and $\mathcal{MD}_R$ make valid decisions under correlation, we now look at the correlation necessary to overcome the conservative Bonferroni bound required for proving the validity of these procedures. The main difference between the independent and correlated cases in $\mathcal{HAK}+$ and $\mathcal{MD}_R$ lies in the selection of β_1 and β_2 and thus η_1 and η_2 (see equations (13) and (15)). If the positive correlation is not strong enough, our valid procedures with correlated systems may require more observations than with independently-simulated systems.

To analyze the difference between the independent and correlated systems, we consider the simple case of equal variances across systems for primary and secondary performance measures. We let the means be in the DM configuration with $b = k$ (so that $y_{i\ell} = -\epsilon_\ell$ for all $i = 1, \ldots, k$ and $\ell = 1, \ldots, s$ and $x_k = x_i + \delta$ for $i = 1, \ldots, k-1$). We also set the variance of the difference $X_{in} - X_{jn}$ to the same value σ_X^2 for all pairs of systems $i \neq j$ and set an equal variance $\sigma^2_{Y_\ell}$ across systems for all secondary performance measures $Y_{i\ell n}$. Thus, comparison and feasibility check will be based on the same expected continuation regions, and we can focus on two systems. One measure of the relative difficulty of constrained R&S would be a weighted sum of the expected maximum number of samples required to complete feasibility check and comparison, respectively, as a smaller sum would associate with quicker completion times than a larger sum.

Kim and Nelson [30] compute the expected maximum number of samples to be $2\eta_2(n_0-1)\left(\frac{\sigma^2_X}{\delta^2}\right)$ for comparison with a first-stage sample size of n_0. Similarly, we feature the average of the expected maximum number of samples required to determine the feasibility of the individual constraints, as $2\eta_1(n_0-1)\left(\frac{1}{s}\sum_{\ell=1}^{s}\frac{\sigma^2_{Y_\ell}}{\epsilon^2_\ell}\right)$, as a measure of the difficulty of the feasibility check. Therefore, when the measure of difficulty of feasibility check and comparison are averaged we obtain:

$$(n_0-1)\left(\eta_2\frac{\sigma^2_X}{\delta^2}+\frac{\eta_1}{s}\sum_{\ell=1}^{s}\frac{\sigma^2_{Y_\ell}}{\epsilon^2_\ell}\right). \tag{19}$$

Using equal weights for feasibility check and comparison is reasonable light of the results in Section 5.4.1.

If we let η_1 and η_2 be the values for independent systems and η'_1 and η'_2 be the values for correlated systems, then we can compute both values (the independent case can only be calculated numerically for general β_1 and β_2, see Section 5.3.1), and the ratio of the weighted averages of the expected sums of maximum number of required samples is

$$\frac{\text{Weighted average for independent systems}}{\text{Weighted average for correlated systems}}=\frac{\left(\eta_2\frac{\sigma^2_X}{\delta^2}+\frac{\eta_1}{s}\sum_{\ell=1}^{s}\frac{\sigma^2_{Y_\ell}}{\epsilon^2_\ell}\right)}{\left(\eta'_2\frac{\sigma^2_X}{\delta^2}(1-\rho_x)+\frac{\eta'_1}{s}\sum_{\ell=1}^{s}\frac{\sigma^2_{Y_\ell}}{\epsilon^2_\ell}\right)}.$$

We desire to have a smaller value in the correlated case, so we need

$$\rho_x>\left(1-\frac{\eta_2}{\eta'_2}-\frac{\frac{\eta_1-\eta'_1}{s}\sum_{\ell=1}^{s}\frac{\sigma^2_{Y_\ell}}{\epsilon^2_\ell}}{\eta'_2\frac{\sigma^2_X}{\delta^2}}\right). \tag{20}$$

For example, if we let $n_0 = 20$, $1-\alpha = 0.95$, $k = 5$, $s = 2$, $\beta_1 = \beta_2/s$, and $\frac{\sigma_X^2}{\delta^2} = \frac{\sigma_{Y_\ell}^2}{\epsilon_\ell^2}$ for $\ell = 1, 2$, then the resulting η_1, η_2, η_1', η_2' can be found. Using these settings, equation (20) shows that the weighted average of the expected maximum number of required samples under correlation becomes smaller than the weighted average of the expected maximum number of required samples under independent sampling with just $\rho_x > 0.002$. Thus, very little correlation may be needed to produce quicker overall completion times for our $\mathcal{HAK}+$ and $\mathcal{MD}_R$ procedures under CRN.

The relative difficulty of feasibility check and comparison is critical to this analysis, as CRN can only improve the efficiency of the comparison phase. Moreover, the above analysis uses the metric (19) to measure the difficulty of our constrained R&S procedures, and different results will be obtained for other measures. Nevertheless, this section suggests that even a small amount of correlation can overcome the conservative bounds required to ensure validity under CRN.

5.5 Experimental Evaluation and Comparison of Procedures

We now present experimental results to illustrate the comparative performance of our constrained R&S procedures. The experimental setup is as described in Section 5.4.1.1. The choice of $b = (k+1)/2$ was shown to minimize PCS of simultaneously-running procedures in Andradóttir and Kim [5], so we feature $b = (k+1)/2$ throughout. For the purpose of our experiments, we choose the ratios $\alpha_1 = \alpha_2$ and $\beta_1 = \beta_2/s$ as our error allocations, as featured in the previous section, selecting valid values of α_2 or β_2.

To demonstrate the validity of our procedures empirically, we discuss the observed PCS of the procedures in Section 5.5.1. Section 5.5.2 compares the procedures in terms of the number of required observations for multiple variance configurations. Section 5.5.3 shows how the number of required observations changes as the number of constrained performance measures increases. Finally, we provide results that show the effectiveness of CRN when coupled with a simultaneously-running procedure in Section 5.5.4.

5.5.1 PCS

We are interested in inspecting the PCS for both valid and heuristic procedures with both types of feasibility checks, $\mathcal{F}_{\mathcal{B}}^{\mathcal{I}}$ and $\mathcal{F}_{\mathcal{A}}^{\mathcal{I}}$. In Table 28, we present PCS results for a small number of systems ($k = 5$) with five constraints. We choose $v = 1$, because the feasibility check is easier when v is higher. Also, as k increases, the PCS of R&S procedures usually increase, so a setup with a small number of systems k and violated constraints v promises to be challenging in terms of validating PCS. We cover both the DM and MIM configurations with L/L, H/L, and L/H variances.

Table 28: Observed PCS for $k = 5$ systems with $s = 5$ constraints, $b = 3$ feasible systems, and $v = 1$ violated constraints.

	DM			MIM		
	L/L	H/L	L/H	L/L	H/L	L/H
$\mathcal{HAK}(\mathcal{B})$	0.983	0.973	0.994	0.990	0.996	0.995
$\mathcal{HAK}(\mathcal{A})$	0.981	0.973	0.993	0.990	0.996	0.994
$\mathcal{HAK}{+}(\mathcal{B})$	0.975	0.973	0.972	0.983	0.995	0.980
$\mathcal{HAK}{+}(\mathcal{A})$	0.973	0.971	0.970	0.982	0.995	0.978
$\mathcal{MD}_R(\mathcal{B})$	0.975	0.973	0.974	0.984	0.995	0.980
$\mathcal{MD}_R(\mathcal{A})$	0.974	0.971	0.971	0.982	0.995	0.979

We note that the observed PCS for all procedures, valid or heuristic, lies above the nominal 0.95. The heuristic $\mathcal{HAK}$ can be conservative for both DM and MIM configurations. The simultaneously-running procedures, $\mathcal{HAK}+$ and $\mathcal{MD}_R$, tend to be a little less conservative, except in the H/L case where the additional samples $\mathcal{HAK}$ takes for feasibility check are dominated by hard comparison. There is a small decrease in PCS between the valid feasibility check $\mathcal{F}_{\mathcal{B}}^{\mathcal{I}}$ and the heuristic version of $\mathcal{F}_{\mathcal{A}}^{\mathcal{I}}$, but not enough to discourage use. In fact, the observed PCS of all of our remaining configurations will lie above 0.95, so we will not feature PCS any further.

5.5.2 Required Number of Observations

We wish to compare the effectiveness of our procedures and feasibility check options in terms of the required number of observations. In Tables 29 and 30, we display the average number of required observations for all combinations of our procedures considering a large number of systems, $k = 101$, with $s = 5$ constraints. The number violated constraints for each infeasible system is $v = 1$ for Table 29 and $v = 5$ for Table 30.

Table 29: Average number of required observations for $k = 101$ systems with $s = 5$ constraints, $b = 51$ feasible systems, and $v = 1$ violated constraints.

	DM			MIM		
	L/L	H/L	L/H	L/L	H/L	L/H
$\mathcal{HAK}(\mathcal{B})$	27560	68856	131026	3769	9713	14845
$\mathcal{HAK}(\mathcal{A})$	27536	68835	130921	3768	9713	14843
$\mathcal{HAK}+(\mathcal{B})$	23552	67236	92782	4298	10343	13515
$\mathcal{HAK}+(\mathcal{A})$	23523	67208	92631	4298	10343	13514
$\mathcal{MD}_R(\mathcal{B})$	21208	67146	60506	3563	9667	8639
$\mathcal{MD}_R(\mathcal{A})$	21179	67118	60360	3562	9667	8637

Table 30: Average number of required replications for $k = 101$ systems with $s = 5$ constraints, $b = 51$ feasible systems, and $v = 5$ violated constraints.

	DM			MIM		
	L/L	H/L	L/H	L/L	H/L	L/H
$\mathcal{HAK}(\mathcal{B})$	23524	64905	110256	3457	9295	12593
$\mathcal{HAK}(\mathcal{A})$	19467	60836	90012	3217	9057	10293
$\mathcal{HAK}+(\mathcal{B})$	19975	63773	74080	4039	9957	11508
$\mathcal{HAK}+(\mathcal{A})$	16438	60226	56403	3858	9763	10556
$\mathcal{MD}_R(\mathcal{B})$	17740	63709	42150	3292	9290	6650
$\mathcal{MD}_R(\mathcal{A})$	14204	60162	24476	3095	9095	4692

Tables 29 and 30 show that $\mathcal{MD}_R$ outperforms $\mathcal{HAK}$ and $\mathcal{HAK}+$ in all cases, documenting the desirable effects of dormancy. Moreover, $\mathcal{HAK}+$ performs better than $\mathcal{HAK}$ in all cases, except for the MIM mean configuration under the L/L and

H/L variance configurations where feasibility check is relatively easy. The biggest difference in performance is seen when feasibility check is hard, where $\mathcal{MD}_R$ outperforms $\mathcal{HAK}$ and $\mathcal{HAK}+$ by at least 30%, sometimes more. When comparison is hard, $\mathcal{MD}_R$ and $\mathcal{HAK}$ again are the most promising, with one configuration that favors $\mathcal{HAK}$ ($v = 5$ and MIM).

The relative performance of the individual feasibility check options does not depend heavily on our general procedures, $\mathcal{HAK}$, $\mathcal{HAK}+$, and $\mathcal{MD}_R$, but is highly dependent on the number of violated constraints. Under $v = 1$, we see that the performance of procedures with $\mathcal{F}_{\mathcal{B}}^{\mathcal{I}}$ is similar to that of procedures with $\mathcal{F}_{\mathcal{A}}^{\mathcal{I}}$. This is expected, as aggregation is not very helpful when only one or two constraints are violated. This changes in Table 30. $\mathcal{F}_{\mathcal{A}}^{\mathcal{I}}$ is significantly superior in all cases when $v = 5$, and the savings over $\mathcal{F}_{\mathcal{B}}^{\mathcal{I}}$ ranges from 5% to 40%, depending on the relative difficulty of the feasibility check.

The performance of the procedures across the tables also indicates that a constrained R&S problem with $v = 5$ is easier than when $v = 1$, requiring at least 5% less observations. When the feasibility check is relatively more difficult than comparison, this effect is more pronounced with, with $v = 5$ requireing at least 19% fewer observations for $\mathcal{HAK}$, 8% fewer observations for $\mathcal{HAK}+$, and 30% fewer observations for $\mathcal{MD}_R$. When the infeasible systems violate $v = 5$ constraints, the feasibility check ends as soon as the first of these constraints is found infeasible (the minimum of the five screening completion times). If all but one of the measures is feasible, the feasibility check can be ended only by the one infeasible constrained performance measure.

5.5.3 Cost of Additional Constraints

In this section, we would also like to investigate the cost of additional constraints, as users may be interested in learning how many more (or less) observations would be needed to consider extra performance measures. In Section 5.4.1, a sublinear increasing trend in observations was seen as the number of constraints increases. Also, Section 5.5.2 indicates that a spectrum of results can be found, depending on the number of violated constraints, v, where large v usually indicates a quicker completion time than small v.

Therefore, our experiments consider the two cases, $v \in \{1, s\}$, for each number of constraints, $s \in \{1, \ldots, 5\}$, so that either all infeasible systems violate only one constrained performance measure, or and all infeasible systems violate every constrained performance measure. It is reasonable that most results will fall between these two cases. To increase the difference between the cases, we will implement $\mathcal{F}_{\mathcal{B}}^{\mathcal{I}}$ when $v = 1$ and $\mathcal{F}_{\mathcal{A}}^{\mathcal{I}}$ when $v = s$, as Table 30 shows $\mathcal{F}_{\mathcal{A}}^{\mathcal{I}}$ to be particularly efficient when v is high.

Figures 8 and 9 plot the required number of samples for each of our three procedures in the DM mean and L/L variance configuration and MIM mean and L/L variance configuration, respectively, where comparison and feasibility check have similar difficulties. The two lines plotted show the necessary observations under the favorable case where $v = s$ and $\mathcal{F}_{\mathcal{A}}^{\mathcal{I}}$ is implemented and the more difficult case where $v = 1$ and $\mathcal{F}_{\mathcal{B}}^{\mathcal{I}}$ is implemented. In the case when $v = 1$, we observe an increase in the number of required observations as s increases, but this increase is sublinear, as in Section 5.4.1.

Significantly, when $v = s$, for all of the procedures and configurations (except $\mathcal{HAK}+$ in MIM), we actually see an initial decrease in the number of observations. This is due to a much faster feasibility check, as screening stops once the first infeasible performance measure is identified. Figures 8 and 9 feature a growing difference between the cases (up to 20%) as the number of constraints grows. Thus, he introduction of additional constraints can influence the performance of the algorithms significally, but more constraints do not necessary mean more samples.

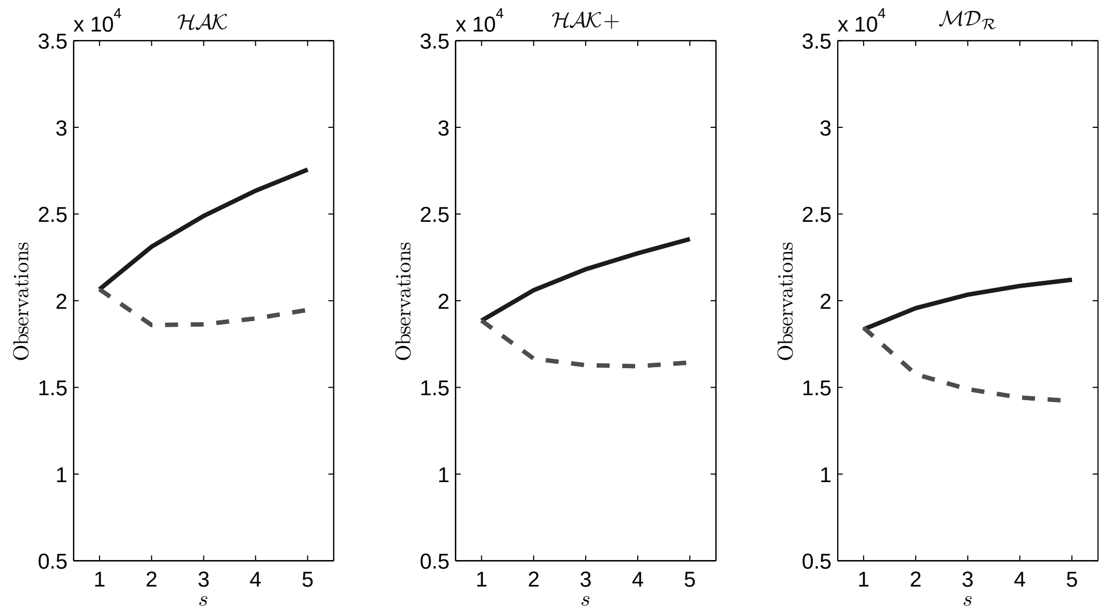

Figure 8: Required number of observations as a function of the number of constraints for the DM with L/L configuration considering $k = 101$ systems with $b = 51$. The top line corresponds to $\mathcal{F}_{\mathcal{B}}^{\mathcal{I}}$ under $v = 1$, while the bottom line corresponds to $\mathcal{F}_{\mathcal{A}}^{\mathcal{I}}$ under $v = s$.

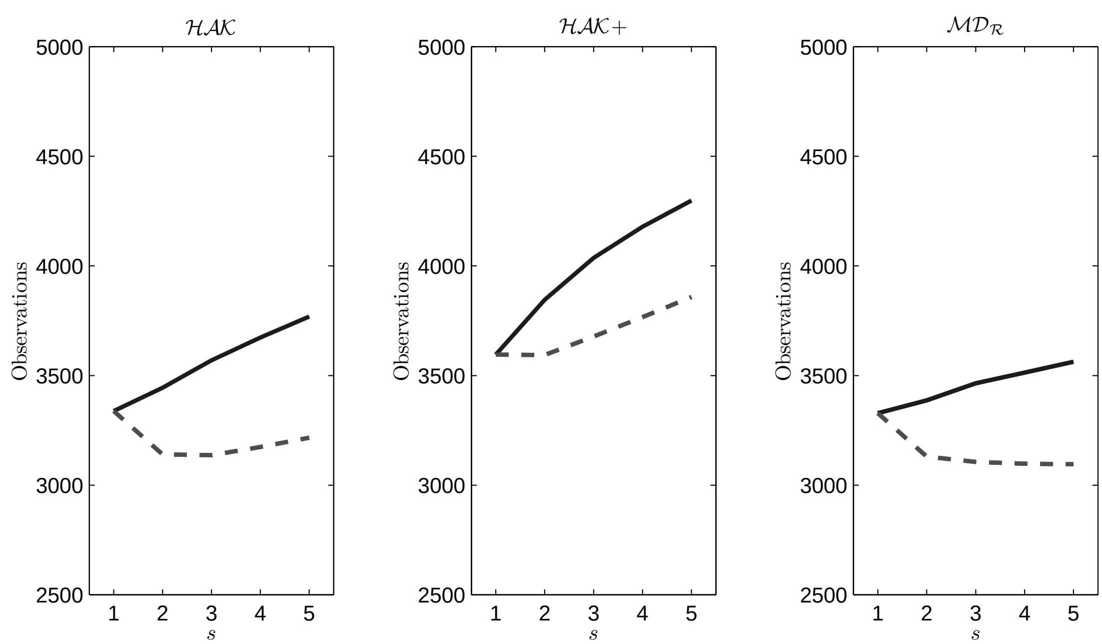

Figure 9: Required number of observations as a function of the number of constraints for the MIM with L/L configuration considering $k = 101$ systems with $b = 51$. The top line corresponds to $\mathcal{F}_{\mathcal{B}}^{\mathcal{I}}$ under $v = 1$, while the bottom value corresponds to $\mathcal{F}_{\mathcal{A}}^{\mathcal{I}}$ under $v = s$.

5.5.4 Common Random Numbers

The last topic for our experiments involves the savings experienced through the use of CRN. As the positive correlation induced by CRN reduces the variances involved in our comparison phase and not the feasibility check, we expect to see significant savings in the number of required samples, albeit smaller than that observed in the pure comparison of Kim and Nelson [30]. We inspect of use of CRN in the simultaneously running $\mathcal{MD}_R$ procedure, which has been shown to be valid under CRN.

Recall that ρ_x denotes the correlation across systems' primary performance measure samples, which we will test at varying levels to measure the effects of CRN. We let the secondary performances measure samples be independent across systems. Practically, correlation will occur and the values of β_1 will be different when considering independent or correlated systems, but the magnitude of correlation should not have a major impact as feasibility check procedures run separately for each system.

We present Table 31 as the experimental performance of the valid procedures $\mathcal{MD}_R$ under varying levels of correlations across systems. The configuration tested in Table 31 features $s = 5$ constraints with $v = 1$ violated constraints for infeasible systems, corresponding to a difficult feasibility check. We seek to show in this setup that CRN can provide considerable savings, but the presence of the feasibility check will limit the savings. The first line for each procedure indicates the number of required observations when systems are independent under valid parameters chosen according to equation (13), while all other procedures experience some level of correlation ρ_x as represented by a superscript, $\mathcal{MD}_R^{\rho_x}$, and operate under valid parameters chosen as in equation (15) to account for possible correlation.

Table 31 clearly shows that as ρ_x increases, the number of required observations for $\mathcal{MD}_R$ decreases. The amount of savings over the independent case are highly influenced by the difficulty of the comparison. Utilizing CRN can provide substantial savings under high correlation, reducing the number of required samples by 50%

Table 31: Average number of required replications for $k = 101$ systems with $s = 5$ constraints, $b = 51$ feasible systems, $v = 1$ violated constraints, and differing levels of correlation, ρ_x, across systems. All procedures utilize $\mathcal{F}_{\mathcal{B}}^{\mathcal{I}}$ for feasibility check.

	DM			MIM		
$\mathcal{MD}_R(\mathcal{B})^{\rho_x}$	L/L	H/L	L/H	L/L	H/L	L/H
$\mathcal{MD}_R(\mathcal{B})$	21208	67146	60506	3564	9667	8638
$\mathcal{MD}_R(\mathcal{B})^{0.00}$	21308	68423	60616	3571	9834	8649
$\mathcal{MD}_R(\mathcal{B})^{0.10}$	20260	62255	59494	3445	9162	8609
$\mathcal{MD}_R(\mathcal{B})^{0.25}$	19018	52086	57496	3299	7645	8330
$\mathcal{MD}_R(\mathcal{B})^{0.50}$	15794	37080	53647	3022	5659	8147
$\mathcal{MD}_R(\mathcal{B})^{0.75}$	13110	24254	50141	2855	3873	7852
$\mathcal{MD}_R(\mathcal{B})^{0.90}$	10860	15985	47500	2776	3060	7772

to 75% under hard comparison. We cannot save almost all observations for high correlation as standard comparison procedures may be able to (see Kim and Nelson [30]) because some samples must be used for feasibility check.

While $\mathcal{MD}_R(\mathcal{B})^0.00$ performs worse than $\mathcal{MD}_R$ in Table 31 due to the use of valid parameters values for correlated systems being applied to independent systems, the observed positive correlation required to provide savings in the number of required samples is low, less than $\rho_x = 0.1$ for all configurations we considered. This is consistent with our results in Section 5.4.2.2. Thus, CRN is an effective approach to improve the efficiency of simultaneously-running constrained R&S procedures.

5.6 Conclusion

In this chapter, we present and analyze three fully-sequential ranking and selection (R&S) procedures for finding the best simulated system that also satisfies constraints on multiple secondary performance measures. These procedures are combined with two valid feasibility check approaches, leading to six difference methods for solving the general constrained R&S problem. We show that two of the procedures, $\mathcal{HAK}+$ and $\mathcal{MD}_R$, are statistically valid considering independent or correlated systems, while the third procedure $\mathcal{HAK}$ may be a good heuristic option for independent systems.

With regards to experimental design and implementation, we identify two major issues, namely error allocation and use of common random numbers (CRN). In our experimental results, we find that allocating error evenly between feasibility check and comparison performs well in many configurations. We also show that CRN can effectively reduce the number of observations required for comparison for procedures that compare systems with equal sample sizes, even under a small amount of correlation.

Our experimental results also show that the number of required observations grows at most sub-linearly as the number of constraints increases, but in some cases, the number of observations could decrease due to an easier feasibility check. While all procedures have their advantages, we find the $\mathcal{MD}_R$ implemented with the $\mathcal{F}_{\mathcal{A}}^{\mathcal{I}}$ feasibility check is the best choice in many configurations.

CHAPTER VI

A MINIMAL SWITCHING PROCEDURE FOR CONSTRAINED RANKING AND SELECTION UNDER INDEPENDENT OR COMMON RANDOM NUMBERS

The procedures for constrained R&S discussed in the previous chapters of this thesis aim for efficiency in terms of observations required to find the best feasible system, but there are none that we know of that address the cost of switching between systems explicitly. While it is common to compare procedures based on the required number samples to achieve a nominal PCS, the possibly high cost (in both time and storage) of stopping and restarting complex simulations should also be considered. Hong and Nelson [28] and Osogami [39] present two fully-sequential procedures that perform valid comparison while limiting the number of switches.

As pointed out by Hong and Nelson [28] and Osogami [39], fully-sequential R&S procedures, such as the $\mathcal{KN}$ procedure of Kim and Nelson [30], may become inefficient if the penalties for switching are large. These costs would also be incurred by any constrained R&S procedure utilizing similar fully-sequential algorithms for comparison. Simulating systems in parallel could reduce switching costs, however, this is not always advisable as parallel computing involves its own complexities (including coordination among processors). Thus, we present a new fully-sequential indifference-zone procedure, named the Constrained Minimal Switching ($\mathcal{CMS}$) procedure, that addresses the concern of switching costs, while identifying the best feasible system.

Minimal switching procedures reduce the cost of stopping and restarting simulations, but often require extra samples to ensure that the number of switches in the comparison phase does not exceed the number of systems. We investigate the use of common random numbers (CRN), a variance reduction technique, to reduce the number of required samples for $\mathcal{CMS}$. Chapter 5 studied the use of CRN in constrained R&S. We proved the validity of two procedures that always compare systems with equal sample sizes, but expressed concerns about the validity of comparing systems with unequal sample sizes under CRN. In this chapter, we provide experimental results that show that PCS can be significantly degraded under high correlation. Since unequal sample sizes commonly occur within our new minimal switching procedure (and other procedures), we present four variance estimate modifications and show that their use within $\mathcal{CMS}$ under CRN captures savings, while still providing good PCS.

This chapter is organized as follows. Section 6.1 outlines the problem of constrained R&S, details notation, and sets assumptions for the validity of our procedure. Section 6.2 introduces the $\mathcal{CMS}$ procedure and includes a proof of its validity for independently simulated systems. In Section 6.3, we motivate the use of CRN, discuss its effects within our $\mathcal{CMS}$ procedure, and propose modifications to address its challenges. Section 6.4 features experimental results, followed by conclusions in Section 6.5.

6.1 Background

The goal of constrained R&S is the selection of the best system according to a primary performance measure out of a fixed number of alternatives, k, with constraints on s secondary performance measures. We outline the problem in Section 6.1.1, and introduce notation necessary for our algorithm and its proof in Section 6.1.2.

6.1.1 Problem Formulation

Let $(X_{in}, Y_{i1n}, \ldots, Y_{isn})$ be the nth observation of the ith system for the primary performance measure and s secondary performance measures. We consider the set of all possible systems $S = \{1, \ldots, k\}$. We let $x_i = \mathrm{E}[X_{in}]$ and $y_{i\ell} = \mathrm{E}[Y_{i\ell n}]$ be the mean values of the primary and secondary performance measures for each system $i \in S$ and constraint $\ell = 1, \ldots, s$. Therefore our objective is to determine which system has the best primary performance measure, while also satisfying all constraints:

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

We let $\sigma_{x_i}^2 = \mathrm{Var}[X_{in}]$ for all i and $\sigma_{y_{i\ell}}^2 = \mathrm{Var}[Y_{i\ell n}]$ for all i and ℓ. Moreover, the relationship between performance measures is governed by the following assumption.

Assumption 17. *For each $i = 1, 2, \ldots, k$,*

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

where $\overset{iid}{\sim}$ denotes independent and identically distributed, MN denotes multivariate normal, and Σ_i is the $(s+1) \times (s+1)$ covariance matrix of the vector $(X_{in}, Y_{i1n}, \ldots, Y_{isn})$.

The normality of data is a common assumption within ranking and selection, achieved through within-replication averages or batched means (Law and Kelton [33]). Furthermore, data points can be correlated across systems due to CRN and across performance measures.

The procedure detailed in this chapter utilizes the indifference-zone method for both the feasibility check and comparison phases. For all systems involved in the simulation, we designate the indifference-zone parameter, δ, as the smallest significant difference between systems' primary performance measures. So, we are "indifferent" between systems that have means within δ of each other.

Likewise, we consider the tolerance level ϵ_ℓ to be the smallest significant difference between $y_{i\ell}$ and q_ℓ. Therefore, we can place all systems into three sets in terms of feasibility. If system i is in S_D, the set of desirable systems, then $y_{i\ell} \leq q_\ell - \epsilon_\ell$ for all $\ell = 1, \ldots, s$. S_U is the set of undesirable systems where at least one secondary performance measure, $y_{i\ell}$, is infeasible, so that $y_{i\ell} > q_\ell + \epsilon_\ell$. All systems not in S_D or S_U fall into S_A, the set of acceptable systems.

Assumption 18. *Let $x_{[b]} \geq x_i + \delta$ for all $i \in S_D \cup S_A \setminus \{[b]\}$, where $[b]$ is the index of the best feasible system.*

Under Assumption 18, we let CS be the correct selection event that system $[b]$ is declared feasible and all systems in $S \setminus \{[b]\}$ are eliminated. If all systems are infeasible, then CS is the event that all systems in S are eliminated. We desire to ensure a nominal PCS at least $1 - \alpha$.

6.1.2 Notation and Assumptions

We present the following notation:

n_0 = the first-stage sample size;

$S^2_{X_{ij}}$ = the sample variance of the paired difference of $\{X_{i1}, \ldots, X_{in_0}\}$ and $\{X_{j1}, \ldots, X_{jn_0}\}$;

$S^2_{X_i}$ = the sample variance of $\{X_{i1}, \ldots, X_{in_0}\}$.

$S^2_{Y_{i\ell}}$ = the sample variance of $\{Y_{i\ell 1}, \ldots, Y_{i\ell n_0}\}$ the ℓth constraint of system i);

$\boldsymbol{Y}_{in} = (Y_{i1n}, Y_{i2n}, \ldots, Y_{isn})^T$;

$R(r; a, b, d) = \max\{0, \frac{bd}{2a} - \frac{a}{2}r\}$, for $a, b, d \in \mathbb{R}^+$ and $a \neq 0$;

CS_i = the event that a good selection is made in pairwise comparison of systems i and $[b]$, for any $i \in S_D \cup S_A$ with $x_{[b]} \geq x_i + \delta$;

CD_i = the event that correct decision is made on the feasibility of system $i \in S$ (when $i \in S_A$, CD_i can be infeasible or feasible);

β_1 = the error of an individual feasibility check for one performance measure of one system;

β_2 = the error of an individual comparison between two systems.

With this notation, we now present two assumptions that govern good feasibility check and comparison phases. Assumptions 19 and 20 ensure that feasibility check and comparison are handled in a valid manner.

Assumption 19. *The systems are simulated independently, and the feasibility check phase guarantees* $\mathsf{Pr}\{\cap_{i \in S'} CD_i\} \geq (1 - s\beta_1)^t$ *for any* $1 \leq t \leq k$ *and any subset* $S' \subseteq S$ *with cardinality* t*, (i.e.,* $|S'| = t$*) under* s *constraints.*

Assumption 20. *The systems are simulated independently, and the comparison phase guarantees* $\mathsf{Pr}\{\cap_{i \in S'} CS_i\} \geq (1 - \beta_2)^t$ *for any* $1 \leq t \leq k - 1$ *and any subset* S' *of* $\{i \in \{1, \ldots, k\} : x_i \leq x_{[b]} - \delta\}$ *with cardinality* t *(i.e.,* $|S'| = t$*).*

6.2 Constrained Minimal Switching Procedure – $\mathcal{CMS}$

In this section, we present a new approach for constrained R&S, namely $\mathcal{CMS}$, that minimizes the cost of switching from one system to another. This cost is often not factored into R&S studies, but it can comprise a large portion of the computation time.

We chose to feature two fully-sequential procedures for the feasibility check and comparison phases in $\mathcal{CMS}$, although many procedures satisfy Assumptions 19 and 20, including some two-stage procedures. Fully-sequential procedures have been shown to be efficient in many configurations, as comparison and feasibility check can be reevaluated after every stage of sampling, possibly with as little as one additional observation.

The feasibility check phase of $\mathcal{CMS}$ is performed by the $\mathcal{F}_{\mathcal{B}}^{\mathcal{I}}$ procedure of Batur and Kim [7] (with $c = 1$), a general, fully-sequential, and valid method for determining feasibility of multiple constrained performance measures. The comparison phase of $\mathcal{CMS}$ is performed by the $\mathcal{MSS}$ procedure of Hong and Nelson [28], modified as described in their Remark 3. While Hong and Nelson [28] proposed two fully-sequential R&S procedures that minimize the number of switches, utilizing the continuation region of the $\mathcal{SSM}$ method of Pichitlamken et al. [41], we chose the valid $\mathcal{MSS}$ procedure for implementation into our $\mathcal{CMS}$ switching procedure.

Our new constrained R&S procedure will retain the fully-sequential approaches of the feasibility check and comparison procedures, but requires an additional step to ensure minimal switching while also performing feasibility checks. The procedure will visit each system at most once after the first stage. To achieve this, at least one system must receive a large number of samples, the maximum necessary to complete comparison with all other systems. Therefore, we expect this algorithm to be conservative in terms of observations, but a good choice if switching costs are high.

The $\mathcal{CMS}$ procedure consists of three steps, namely sorting the systems by primary performance measure after the first-stage of sampling, performing feasibility check on systems according to their sorted order to find the initial guess for the best feasible system (B), and then comparing the current guess for the best feasible system (B) with the next best available system (A), until no systems remain. Sampling occurs for only the next best available system A. Each successive system A is simultaneously tested for feasibility and compared to B. System A can become the current guess for best feasible system only if it is found feasible and superior to system B. If one of these conditions is found not to be true, A is eliminated, a new A is chosen to be the next available system, and sampling shifts to the new system A. This proceeds until all available systems are eliminated by comparison or feasibility check.

Procedure [$\mathcal{CMS}$ for Multiple Constraints]

Setup: Select the overall confidence level $1/k \leq 1-\alpha < 1$ and first–stage sample size, $n_0 \geq 2$. Choose δ, ϵ_ℓ, and q_ℓ for $\ell = 1, 2, \ldots, s$. Let $\eta_1 = \frac{1}{2}((2\beta_1)^{-2/(n_0-1)} - 1)$ and $\eta_2 = \frac{1}{2}((2\beta_2)^{-2/(n_0-1)} - 1)$, where $\beta_1 = \beta_2/s$ and β_2 is the unique solution to the equation $\beta_2 + 2[1-(1-\beta_2)^{(k-1)/2}] = \alpha$.

Initialization: Let $h_1^2 = 2\eta_1(n_0-1)$ and $h_2^2 = 2\eta_2(n_0-1)$. Obtain n_0 observations X_{in} and $\boldsymbol{Y}_{in}$ from each system $i \in S$. For all i and ℓ, compute the estimators $S^2_{Y_{i\ell}}$.

Similarly, for all i and $j \neq i$, compute the estimator $S^2_{X_{ij}}$. Also compute N_{ij} for all $i, j \in S$ and $i \neq j$, where

$$N_{ij} = \max \left\{ n_0, \left\lceil \frac{h_2^2 S^2_{X_{ij}}}{\delta^2} \right\rceil \right\}$$

and $\lceil \cdot \rceil$ is the ceiling function. Let $SI_i = \emptyset$ be the set of systems inferior to system $i \in S$ in terms of the primary performance measure. Let $K_i = \emptyset$ be the set of constraints found to be feasible for system $i \in S$ and let the set of contending systems include all systems, $M = S$. The procedure will require the calculation of the maximum number of samples required for system i to complete comparison with all systems remaining in contention:

$$N_i = \max_{j \in M \setminus (SI_i \cup \{i\})} N_{ij}. \tag{21}$$

Set the observation counters $r_i = n_0$ for all $i \in S$.

Finding a Feasible System:

Initial Sorting: Sort the systems in M based on the first-stage sample means $\bar{X}_i = \frac{1}{n_0} \sum_{n=1}^{n_0} X_{in}$. Let B and A be the systems in M with the best and second-best first stage sample means.

Initial Screening for Comparison: Compare all systems $i \neq j$ in M based on n_0 samples. If

$$\sum_{n=1}^{n_0} X_{in} \geq \sum_{n=1}^{n_0} X_{jn} + R(n_0; \delta, h_2^2, S^2_{X_{Bj}}),$$

then add j to SI_i. Compute N_B using (21).

Initial Feasibility Check: For system B and $\ell \notin K_B$, if

$$\sum_{n=1}^{r_B} (Y_{B\ell n} - q_\ell) \geq R(r_B; \epsilon_\ell, h_1^2, S^2_{Y_{B\ell}}),$$

declare B to be infeasible. Else if

$$\sum_{n=1}^{r_B} (Y_{B\ell n} - q_\ell) \leq -R(r_B; \epsilon_\ell, h_1^2, S^2_{Y_{B\ell}}),$$

add ℓ to K_B. If $|K_B| = s$, declare B to be feasible, remove all systems in SI_B from M, and update A, if necessary.

Stopping Rule: If B is feasible and $|M| = 1$, declare B as the best feasible system. If B is infeasible and $|M| = 1$, then no feasible systems exist. If B is feasible and $|M| > 1$, proceed to **Feasibility and Comparison of A with B**. If B is infeasible and $|M| > 1$, then remove B from M, set $B = A$, compute N_B using (21), let A be the best system in $M \setminus \{B\}$ if $M \setminus \{B\} \neq \emptyset$, and proceed to **Initial Feasibility Check**. Otherwise, take an additional sample from system B, X_{B,r_B+1} and $\boldsymbol{Y}_{B,r_B+1}$, and set $r_B = r_B + 1$. If $r_B = N_B$, store $\sum_{n=n_0+1}^{N_B} X_{Bn}$. Go to **Initial Feasibility Check**.

Feasibility and Comparison of A with B

Sampling for Comparison: Find N_A using (21). If $r_B < N_B$, take an additional $N_B - r_B$ observations from system B and set $r_B = N_B$.

Comparison: If $B \notin SI_A$ and

$$\frac{r_A - n_0}{N_B - n_0} \sum_{n=n_0+1}^{N_B} X_{Bn} + \sum_{n=1}^{n_0} X_{Bn} \geq \sum_{n=1}^{r_A} X_{An} + R(r_A; \delta, h_2^2, S^2_{X_{BA}}),$$

then remove A from M and go to **Stopping Rule**.

If $B \notin SI_A$,

$$\frac{r_A - n_0}{N_B - n_0} \sum_{n=n_0+1}^{N_B} X_{Bn} + \sum_{n=1}^{n_0} X_{Bn} \leq \sum_{n=1}^{r_A} X_{An} - R(r_A; \delta, h_2^2, S^2_{X_{BA}}), \tag{22}$$

and A is feasible, then remove B from M. If $B \notin SI_A$, (22) is true, and A's feasibility is undetermined, add B to SI_A.

Feasibility: If the feasibility of A is unknown, use the same procedure as **Initial Feasibility Check**, except substitute A for B. If A is feasible, remove all system in SI_A from M. If A is infeasible, eliminate system A from M.

Stopping Rule: If $|M| = 1$, stop and declare the remaining system as the best. If $B \notin M$, then set $B = A$, update A, and go to **Sampling for Comparison**. If $A \notin M$, update A and go to **Sampling for Comparison**. Otherwise, take an additional sample from system A, X_{A,r_A+1} and $\boldsymbol{Y}_{A,r_A+1}$, and set $r_A = r_A + 1$. If $r_A = N_A$, store $\sum_{n=n_0+1}^{N_A} X_{An}$. Go to **Comparison**.

The step of taking N_B samples for the current guess for the best feasible system allows the procedure to make statistically valid decisions, while minimizing the number of switches. Each system is sampled at most twice, once for first-stage sampling and sorting and once for feasibility check and comparison. The procedure utilizes only N_B samples for comparison, even if more samples are obtained in a long feasibility check. This is desirable because Chapter 4 show that primary performance measure sample means may be biased at the completion of feasibility check if primary and secondary performance measures are correlated, so observations past N_B are possibly harmful.

To prove the validity of $\mathcal{CMS}$, we first require the following lemma for proving the validity of procedures for constrained R&S that perform feasibility check and comparison simultaneously:

Lemma 3. *(Chapter 5) Under Assumptions 17, 18, 19, and 20, a simultaneous procedure guarantees*

$$\mathsf{Pr}\{CS\} \geq (1 - s\beta_1)^j + (1 - s\beta_1) + (1 - \beta_2)^{k-j-1} - 2$$

when the number of undesirable systems is less than k, and $\mathsf{Pr}\{CS\} \geq (1 - s\beta_1)^k$ when the number of undesirable systems is equal to k.

Lemma 3 allows us to present the main result in this section. Note that for fixed k and α, $2(1 - \beta_2)^{(k-1)/2} - \beta_2 - 1$ monotonically decreases from 1 to -2 as β increases from 0 to 1, guaranteeing a unique solution to equation (23) below.

Theorem 8. *When the systems are simulated independently and Assumptions 17 and 18 hold, $\mathcal{CMS}$ guarantees*

$$\mathsf{Pr}\{CS\} \geq 1 - \alpha$$

when

$$2(1-\beta_2)^{(k-1)/2} - \beta_2 - 1 \geq 1 - \alpha. \tag{23}$$

Proof: We show that $\mathcal{CMS}$ satisfies the conditions of Lemma 3. $\mathcal{F}_{\mathcal{B}}^{\mathcal{I}}$ is proven to satisfy Assumption 19 under Assumption 17, see the proof of Corollary 1 of Batur and Kim [7]. Assumption 20 follows from the proof of validity of $\mathcal{MSS}$ using Fabian's bound, see Hong and Nelson [28] and Pichitlamken et al. [41].

As $\mathcal{CMS}$ satisfies Assumptions 19 and 20, we apply Lemma 3 and the fact that $\beta_1 = \beta_2/s$, and obtain that $\mathcal{CMS}$ guarantees

$$\begin{aligned} \mathsf{Pr}\{CS\} &\geq (1-s\beta_1)^j + (1-s\beta_1) + (1-\beta_2)^{k-j-1} - 2 \\ &= (1-\beta_2)^j + (1-\beta_2) + (1-\beta_2)^{k-j-1} - 2 \end{aligned} \tag{24}$$

when the number, j, of undesirable systems is less than k and $\mathsf{Pr}\{CS\} \geq (1-s\beta_1)^k$ when $j = k$. Remark 2 of Chapter 5 shows that the lowest $\mathsf{Pr}\{CS\}$ occurs when $j < k$, so any β_1 that guarantees $\mathsf{Pr}\{CS\} \geq 1-\alpha$ for $j < k$ also guarantees $\mathsf{Pr}\{CS\}$ when $j = k$.

Since $j = (k-1)/2$ minimizes the right-hand side of (24), the definition of β_2 yields

$$\begin{aligned} \mathsf{Pr}\{CS\} &\geq 2(1-\beta_2)^{(k-1)/2} - \beta_2 - 1 \\ &= 1-\alpha, \end{aligned}$$

which concludes the proof. □

6.3 CRN and Two-Sample Comparison

In our new switching procedure, we require the current best system B to take N_B samples, the maximum samples necessary to make a decision against any remaining system. To reduce this large number of observations, we turn to a popular variance technique, namely common random numbers (CRN), which Nelson and Matejcik [36], Chick and Inoue [18], Kim and Nelson [30], and Chapter 5 among others, show can be used to improve efficiency of both original and constrained R&S procedures.

Proper implementation of CRN can result in quicker decisions by inducing positive correlation across systems. Since $S^2_{X_{ij}}$ is defined as the sample variance of the difference of paired samples from systems i and j, positive correlation across systems can reduce the value of $S^2_{X_{ij}}$ significantly. In procedures that compare systems at even sample sizes, such as $\mathcal{KN}$ of Kim and Nelson [30], only a simple parameter adjustment to β_2 is needed make valid selection under CRN. Unfortunately, we cannot make valid decisions under CRN for two-sample procedures that compare systems with unequal sample sizes ($\mathcal{MSS}$ was proven valid for comparison of independent systems only).

We show how variance estimates in $\mathcal{CMS}$ under CRN negatively impact the validity of the procedure in Section 6.3.1, and propose four modifications to correct the estimates in Section 6.3.2.

6.3.1 Comparison with Positive Correlation

Two-sample procedures that estimate variance with $S^2_{X_{ij}}$ can underestimate the variability of the screening process observed. In our minimal switching procedure, screening is performed and a decision to continue sampling is based on a function of

$$\frac{r_A - n_0}{N_B - n_0} \sum_{n=n_0+1}^{N_B} X_{Bn} + \sum_{n=1}^{n_0} X_{Bn}, \sum_{n=1}^{r_A} X_{An},$$

and an estimate of the variance of the difference of these two sums. If $r_B = r_A = r$, $rS^2_{X_{BA}}$ is an estimator of

$$\mathrm{Var}\left[\sum_{n=1}^{r} X_{Bn} - \sum_{n=1}^{r} X_{An}\right] = r(\sigma^2_{x_B} + \sigma^2_{x_A} - 2\rho_x\sqrt{\sigma^2_{x_B}\sigma^2_{x_A}}), \tag{25}$$

where ρ_x is defined as the correlation between X_{Bn} and X_{An}.

However, if we assume that $r_A \ll r_B = N_B$, we create a situation commonly addressed in the $\mathcal{CMS}$ procedure. Here, the sum for system B is pegged at the random time N_B, a sample size many times larger than r_A. Thus, as we increment r_A

$$\begin{aligned}
&\mathrm{Var}\left[\frac{r_A - n_0}{N_B - n_0} \sum_{n=n_0+1}^{N_B} X_{Bn} + \sum_{n=1}^{n_0} X_{Bn} - \sum_{n=1}^{r_A} X_{An}\right] \\
&\approx \mathrm{Var}\left[\sum_{n=1}^{n_0} (X_{Bn} - X_{An})\right] + \mathrm{Var}\left[\sum_{n=n_0+1}^{r_A} X_{An}\right] \\
&= n_0(\sigma_{x_B}^2 + \sigma_{x_A}^2 - 2\rho_x \sqrt{\sigma_{x_B}^2 \sigma_{x_A}^2}) + (r_A - n_0)\sigma_{x_A}^2. \qquad (26)
\end{aligned}$$

When ρ_x is large, the quantity estimated by $r_A S^2_{X_{BA}}$ in (25) could be smaller than the variability observed by the process in (26). This underestimation of variability can cause premature decisions, hurting $\mathsf{Pr}\{CS_i\}$. The poor variance estimate creates a continuation region, $R(r; \delta, h_2^2, S^2_{X_{BA}})$, that is too small to make a valid decision.

We present an empirical study where we compare two systems, separated by the distance of the indifference-zone, δ, with system 1 being the preferable choice. Table 32 shows the observed $\mathsf{Pr}\{CS_2\}$ of a two-sample comparison under varying correlation $\rho_x \in \{0.0, 0.5, 0.6, 0.7, 0.8, 0.9\}$ and initial sample size differences. The two-sample procedure implemented for Table 32 is the $\mathcal{SSM}$ procedure of Pichitlamken et al. [41], the underlying approach for the more efficient version of $\mathcal{MSS}$ incorporating Fabian's bound that is implemented within $\mathcal{CMS}$. To simulate two-sample comparison, let $r_1 \in \{20, 30, 45, 70, 120, 200, 300, 500\}$ and $r_2 = 20$, so that we give system 1 more samples than system 2. Comparison is performed until a system is eliminated with a nominal $\mathsf{Pr}\{CS_2\}$ of 0.95.

Table 32 shows that for correlation, ρ_x, greater than 0.5, we can see degradation of $\mathsf{Pr}\{CS_2\}$ from the independent, even-sample case ($\rho_x = 0$ and $r_1 = r_2 = 20$). We also note that when $\rho_x > 0.5$ and the gap between the initial sample sizes or ρ_x increases, we observe even worse $\mathsf{Pr}\{CS_2\}$ values. For $\rho_x \geq 0.7$, we can no longer expect the $\mathsf{Pr}\{CS_2\}$ to meet nominal levels.

Table 32: Observed $\mathsf{Pr}\{CS_2\}$ with $x_1 = \delta = 1/\sqrt{20}, x_2 = 0, \sigma^2_{x_1} = \sigma^2_{x_2} = 1$, $r_2 = 20$, and varying correlation, ρ_x after 10,000 replications.

	$\rho_x = 0$	$\rho_x = 0.5$	$\rho_x = 0.6$	$\rho_x = 0.7$	$\rho_x = 0.8$	$\rho_x = 0.9$
$r_1 = 20$	0.976	0.978	0.981	0.981	0.985	0.991
$r_1 = 30$	0.977	0.978	0.978	0.974	0.965	**0.949**
$r_1 = 45$	0.976	0.976	0.974	0.970	0.951	**0.910**
$r_1 = 70$	0.979	0.977	0.974	0.963	**0.935**	**0.894**
$r_1 = 120$	0.981	0.978	0.972	0.955	**0.923**	**0.879**
$r_1 = 200$	0.988	0.981	0.968	**0.948**	**0.918**	**0.870**
$r_1 = 300$	0.990	0.979	0.967	**0.947**	**0.913**	**0.868**
$r_1 = 500$	0.994	0.981	0.969	**0.945**	**0.910**	**0.867**

6.3.2 Heuristic Modifications

We introduce four heuristic modifications to attempt to provide the desired $\mathsf{Pr}\{CS_i\}$ for two-sample comparisons. We will test the modifications within the $\mathcal{CMS}$ procedure and the $\mathcal{HAK}$ procedure of Chapter 5, but the modifications should also prove useful for any general R&S or constrained R&S procedure that utilizes a two-sample comparison. In Section 6.3.2.1, we describe a simple, but conservative modification. In Sections 6.3.2.2, 6.3.2.3, and 6.3.2.4, we introduce variations that will allow for the possibility to significantly benefit from CRNs, while still maintaining the nominal PCS within the constrained R&S procedures.

The approaches require the computation of the first-stage marginal sample variances for each system. Recall that for system i, this quantity is $S^2_{X_i}$. Also note that when incorporated in $\mathcal{CMS}$, these approaches will not only change variance estimates in comparison screening, but also the N_{ij} values that represent the maximum number of samples needed to complete comparison of systems i and j.

6.3.2.1 Two-Sample Modification 1: TS_1

The main concern with fully-sequential two-sample procedures under CRN lies in the underestimation of the variability of the comparison. Under positive correlation, we do have an upper bound for the variability, namely $S^2_{X_i} + S^2_{X_j}$. Therefore, for TS_1, we will replace $S^2_{X_{ij}}$ with

$$\hat{S}^2_{X_{ij}} = S^2_{X_i} + S^2_{X_j}$$

throughout the entire procedure. The savings due to the decrease in the variability of the even-sample process described in equation (25) is almost all lost, as the variance estimate is overly conservative. This modification restores the observed PCS, but will perform similarly to the case when systems are simulated independently.

6.3.2.2 Two-Sample Modification 2: TS_2

While TS_1 provides a valid decision, it does not utilize the positive correlation. Our second modification benefits from the reduction of variance CRN can provide, but only when sample sizes are equal. If sample sizes are not equal, we make decisions based on the conservative estimate of $S^2_{X_i} + S^2_{X_j}$. Thus, for TS_2, we will replace $S^2_{X_{ij}}$ with

$$\hat{S}^2_{X_{ij}} = \begin{cases} S^2_{X_i} + S^2_{X_j}, & \text{if } r_i \neq r_j, \\ S^2_{X_{ij}}, & \text{otherwise.} \end{cases}$$

Unfortunately, when $\mathcal{CMS}$ is implemented, $r_i \neq r_j$ for almost all samples. Therefore, the results obtained for TS_1 and TS_2, applied within $\mathcal{CMS}$, will be virtually identical. However, for other procedures such as $\mathcal{HAK}$, this may still be a desirable modification as shown in Section 6.4.2.2.

6.3.2.3 Two-Sample Modification 3: TS_3

The discussion in Section 6.3.1 suggests that the continuation region is corrupted when $S^2_{X_{ij}} < S^2_{X_i}$ and $r_i < r_j$. Instead of reverting to the conservative estimate of variability when sample sizes are not equal, we use $S^2_{X_i}$ as a bound on variability, when r_i is less than r_j. Therefore, for TS_3, we will replace $S^2_{X_{ij}}$ with

$$\hat{S}^2_{X_{ij}} = \begin{cases} \max\left\{S^2_{X_i}, S^2_{X_{ij}}\right\}, & \text{if } r_i < r_j, \\ \max\left\{S^2_{X_j}, S^2_{X_{ij}}\right\}, & \text{if } r_i > r_j, \\ S^2_{X_{ij}}, & \text{otherwise.} \end{cases}$$

6.3.2.4 Two-Sample Modification 4: TS_4

In a more aggressive modification, we note that when r_j is large and $r_i \approx 0$, $S^2_{X_i}$ would dominate the variability of the process in equation (26). However, when r_i and r_j are close, we would see variance closer to $S^2_{X_{ij}}$. This suggests that when $r_i < r_j$, the variance estimate $\hat{S}^2_{X_{ij}}$ should be close to $S^2_{X_i}$ for small r_i, and $\hat{S}^2_{X_{ij}}$ should approach $S^2_{X_{ij}}$ as r_i approaches r_j. Hence, in TS_4, we will replace $S^2_{X_{ij}}$ with

$$\hat{S}^2_{X_{ij}} = \begin{cases} \max\left\{\frac{r_j - r_i}{r_j} S^2_{X_i} + \frac{r_i}{r_j} S^2_{X_{ij}}, S^2_{X_{ij}}\right\}, & \text{if } r_i < r_j \\ \max\left\{\frac{r_i - r_j}{r_i} S^2_{X_j} + \frac{r_j}{r_i} S^2_{X_{ij}}, S^2_{X_{ij}}\right\}, & \text{if } r_i > r_j \\ S^2_{X_{ij}}, & \text{otherwise.} \end{cases}$$

Of all the proposed modifications, it is reasonable to expect TS_4 to approximate the variability of (26) the best, making it the most promising heuristic (i.e., we expect TS_4 to require the least number of observations to produce the desired PCS). We will present experimental results for all four modifications under differing levels of correlation ρ_x in Section 6.4.2.2 below.

6.4 Experimental Results

In this section, we evaluate the performance of our new $\mathcal{CMS}$ procedure compared to the performance of other constrained R&S procedures, namely $\mathcal{HAK}$, $\mathcal{HAK}+$, and $\mathcal{MD}_R$ of Chapter 5, in terms of the number of switches, number of required observations, and observed PCS. In Section 6.4.1, we discuss the experimental setup for all of our tests. We provide an analysis of $\mathcal{CMS}$ with and without the heuristic modifications to incorporate CRN in Section 6.4.2.

6.4.1 Setup

The mean and variance configurations for our experiments attempt to provide analogous results and analysis to the experimental studies of previous, related fully-sequential indifference-zone R&S studies, namely Kim and Nelson [30], Hong and Nelson [28], Pichitlamken et al. [41], Andradóttir and Kim [5], and Chapters 4 and 5. Our experiments will test the procedures in several different combinations of means and variances with 10,000 macro-replications. For all tests, we set $n_0 = 20$, and δ and ϵ_ℓ equal to the sample standard deviation $1/\sqrt{20}$ of the average when samples have a variance of 1 for all $\ell = 1, 2, \ldots, s$. We set a nominal PCS of $1 - \alpha = 0.95$. We set the number of acceptable system in S_A to be zero, as Andradóttir and Kim [5] show the existence of acceptable systems does not affect results significantly.

The difficult means configuration (DM) attempts to test the validity of the procedures by assigning system means in the most challenging setup. Systems are placed into two groups with respect to the best feasible system: some systems are only slightly inferior, but also feasible by a small amount, and some systems are vastly superior and also only slightly infeasible. In this setup, we define a slightly inferior system to be a distance of the indifference-zone parameter, δ, away from $x_{[b]}$. We also define slightly feasible (infeasible) to imply that a system's mean secondary performance measure ℓ lies a tolerance-level, ϵ_ℓ, below (above) the constraint, q_ℓ, for $\ell = 1, 2, \ldots, s$.

As an added consideration for multiple constraints, we recognize that the number of infeasible constraints of an infeasible system is important. Thus, in addition to considering the number of total systems, k, and the number of feasible systems, b, we will also look at the number of violated constraints, $v \in \{1, \ldots, s\}$, for infeasible systems. Hence, in the DM configuration,

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

and

$$y_{i\ell} = E[Y_{i\ell n}] = \begin{cases} -\epsilon_\ell, & i = 1, 2, \ldots, b, \\ \epsilon_\ell, & i = b+1, \ldots, k \text{ and } \ell = 1, 2, \ldots, v, \\ -\epsilon_\ell & i = b+1, \ldots, k \text{ and } \ell = v+1, v+2, \ldots, s. \end{cases}$$

We set the constraint levels, q_ℓ, to zero.

We also consider the MIM configuration, which will allow us to determine the efficiency at which the procedures determine the feasibility of clearly infeasible or feasible systems and compare substantially distant systems. In the MIM configuration,

$$x_i = E[X_{ij}] = (i-1)\delta, \quad i = 1, 2, \ldots, k,$$

and

$$y_{i\ell} = E[Y_{i\ell j}] = \begin{cases} -(b-i+1)\epsilon_\ell, & i = 1, 2, \ldots, b, \\ (i-b)\epsilon_\ell, & i = b+1, \ldots, k, \text{ and } \ell = 1, 2, \ldots, v, \\ -(i-b)\epsilon_\ell, & i = b+1, \ldots, k, \text{ and } \ell = v+1, v+2, \ldots, s, \end{cases}$$

where again we set $q_\ell = 0$.

For the experiments, we examine a combination of variance configurations to test the procedures under different difficulty of feasibility check and comparison. We consider a similar setup to Chapter 5, as we include low (L) and high (H) variances for the primary and secondary performance measures, $\sigma^2_{x_i}$ and $\sigma^2_{y_{i\ell}}$, respectively, but the H variance is larger than in Chapter 5 while the low variance remains the same. For simplicity, all secondary performance measures $\ell = 1, 2, \ldots, s$ are assigned identical variances. High variance results in either $\sigma^2_{x_i} = 10$ or $\sigma^2_{y_{i\ell}} = 10$ and low variance sets $\sigma^2_{x_i} = 1$ or $\sigma^2_{y_{i\ell}} = 1$.

As in Section 6.3, we let ρ_x be the correlation across systems primary performance measure samples. We will consider both independently simulated systems and systems with induced $\rho_x > 0$, modeling CRN. Andradóttir and Kim [5] and Chapter 4 present empirical results that show that the correlation across primary and secondary performance measures does note have a major impact on performance, so we will not revisit the topic in this chapter.

Similarly, Batur and Kim [7] show that correlation across only secondary performance measures does not largely affect the performance of the feasibility check procedure $\mathcal{F}^{\mathcal{I}}_{\mathcal{B}}$. We expect similar conclusions would be found here, and hence implement our procedures with independent secondary performance measure samples. Finally, we assume the constrained performance measures are not correlated across systems under CRN. In practice, secondary performance measures will likely be correlated across systems, but this correlation is unlikely to have a major impact on performance, since feasibility check is performed separately for individual systems.

6.4.2 Results

In our experimental results, we display the effectiveness of multiple constrained R&S procedures, with respect to observed PCS, average number of required samples, and average number of switches. We define a switch to be the initialization and resuming of sampling for a system. A two-stage procedure for k systems (all feasible) requires at most $2k$ switches, two sets of sampling for each system (one to gather first-stage samples and one to complete comparison). Fully-sequential procedures register a switch after each stage of sampling for every system remaining in contention. In Section 6.4.2.1, we will consider independent systems. We show how the use of CRN affects the performance of $\mathcal{HAK}$ of Chapter 5 and $\mathcal{CMS}$ in Section 6.4.2.2. We conclude Section 6.4.2.2 with an analysis of how our heuristic modifications can produce good PCS for procedures with high correlation.

6.4.2.1 Systems under Independent Sampling

To evaluate the performance of $\mathcal{CMS}$ under independent sampling of systems, we compare it to three procedures for constrained R&S, namely the $\mathcal{HAK}$, $\mathcal{HAK}$+, and $\mathcal{MD}_R$ procedures. $\mathcal{HAK}$ is a sequentially-running procedure, performing a complete feasibility check of all systems and then a comparison on the systems found feasible. This procedure is most efficient when feasibility check is quick to finish. $\mathcal{HAK}$+ and $\mathcal{MD}_R$ are simultaneously-running procedures that perform both feasibility check and comparison on all systems remaining in contention after each stage of sampling. These simultaneously-running procedures are preferable in cases when feasibility check is relatively difficult.

We operate the four procedures under similar setups. For example, we choose $\alpha_1 = \alpha_2$ in $\mathcal{HAK}$ and $\beta_1/s = \beta_2$ in $\mathcal{HAK}$+, $\mathcal{MD}_R$, and $\mathcal{CMS}$, so that error is allocated equally between feasibility check and comparison. This allocation was shown experimentally to be a robust choice in Chapter 5. All procedures are implemented with the feasibility check procedure, $\mathcal{F}_{\mathcal{B}}^{\mathcal{I}}$, although there are other methods that could be utilized (see, e.g., Batur and Kim [7]).

Tables 33, 34, and 35 display the observed PCS, average number of observations, and average number of switches, respectively, for 15 systems with 8 feasible and 101 systems with 51 feasible, in addition to three constraints and a combination of various mean and variance configurations. Each infeasible system violates only one of the constraints. We choose $b = \lceil \frac{k+1}{2} \rceil$ to minimize the PCS of our procedures. This setup challenges the PCS of the procedures, as shown by Andradóttir and Kim [5] and in Chapter 5. Half of the systems must be eliminated by comparison and half must be eliminated by feasibility check. The feasibility check is also difficult, as screening must catch the single violated constraint.

Table 33: Observed PCS for procedures with k independent systems, $s = 3$ constraints, b feasible systems, and $v = 1$ infeasible constraints.

	DM ($\sigma^2_{x_i}/\sigma^2_{y_{i\ell}}$)				MIM ($\sigma^2_{x_i}/\sigma^2_{y_{i\ell}}$)			
$k = 15, b = 8$	L/L	L/H	H/L	H/H	L/L	L/H	H/L	H/H
$\mathcal{HAK}$	0.985	0.998	0.977	0.977	0.994	0.999	0.992	0.996
$\mathcal{HAK}$+	0.980	0.981	0.979	0.963	0.990	0.993	0.990	0.992
$\mathcal{MD}_R$	0.980	0.985	0.979	0.963	0.990	0.993	0.990	0.992
$\mathcal{CMS}$	0.993	0.994	0.994	0.989	1.000	1.000	1.000	1.000
$k = 101, b = 51$								
$\mathcal{HAK}$	0.999	0.993	0.981	0.999	1.000	1.000	0.999	1.000
$\mathcal{HAK}$+	0.999	0.995	0.980	0.999	0.999	0.999	0.999	0.998
$\mathcal{MD}_R$	0.999	0.995	0.980	0.999	0.999	0.999	0.999	0.998
$\mathcal{CMS}$	1.000	0.996	0.999	0.999	1.000	1.000	1.000	0.998

Table 34: Average number of required samples for procedures with k independent systems, $s = 3$ constraints, b feasible systems, and $v = 1$ infeasible constraints.

	DM ($\sigma^2_{x_i}/\sigma^2_{y_{i\ell}}$)				MIM ($\sigma^2_{x_i}/\sigma^2_{y_{i\ell}}$)			
$k = 15, b = 8$	L/L	L/H	H/L	H/H	L/L	L/H	H/L	H/H
$\mathcal{HAK}$	2662	23684	14762	26607	1159	9753	7370	11654
$\mathcal{HAK}$+	2392	17416	14829	23734	1270	8279	7427	12646
$\mathcal{MD}_R$	2217	10775	14828	22101	1062	5600	7424	10633
$\mathcal{CMS}$	3472	12264	31645	38342	1833	6100	19163	22275
$k = 101, b = 51$								
$\mathcal{HAK}$	24695	225314	123517	246082	3566	25555	17470	29427
$\mathcal{HAK}$+	23097	152777	126224	232750	3986	22059	17873	34422
$\mathcal{MD}_R$	21783	99735	126193	220442	3425	13883	17545	26326
$\mathcal{CMS}$	30804	107019	281492	354863	5232	14717	43317	51674

Since $\mathcal{CMS}$ was proven valid, the performance in Table 33 is expected to be better than the nominal 0.95. We observe this to be true in all cases. Moreover, $\mathcal{CMS}$ commonly provides a higher PCS than the other procedures, which is a result of the extra samples needed to limit switches during the procedure's comparison phase. The observed PCS is much higher for $k = 101$ than when $k = 15$, for all four procedures.

The comparison phase of $\mathcal{CMS}$ can make this procedure less attractive than the other procedures in terms of the number of required observations. Still, Table 34 shows this is not always the case. When comparison and feasibility check phases are equally difficult (L $\sigma^2_{x_i}$/L $\sigma^2_{y_{i\ell}}$ and H $\sigma^2_{x_i}$/H $\sigma^2_{y_{i\ell}}$), $\mathcal{CMS}$ will require as much as 110% more observations. Under hard comparison (H/L), this extra percentage rises to about 160%. However, when only feasibility check is difficult (L/H), $\mathcal{CMS}$ can be relatively efficient, bettering the totals of all procedures except $\mathcal{MD}_R$. By determining feasibility only for systems with the most attractive primary performance measures, our switching procedure $\mathcal{CMS}$ spends fewer observations on the feasibility check. The results under $k = 15$ and $k = 101$ are similar, besides the larger number of required observations for $k = 101$.

Table 35 shows why $\mathcal{CMS}$ is a competitive procedure when the cost of switches is counted. For every configuration, $\mathcal{CMS}$ requires $2k$ or less switches when simulating k systems, including first-stage samples and the following feasibility checks and comparison. At times, systems will not require additional samples past the first-stage, resulting in less than $2k$ switches. The other simultaneously-running procedures, $\mathcal{HAK}+$ and $\mathcal{MD}_R$, can require thousands of switches, as every stage of sampling consists of as little as one observation from each system in contention. $\mathcal{HAK}$ is a special exception. When feasibility check is difficult and no additional samples are needed to complete comparison (L/H), $\mathcal{HAK}$ can also achieve the minimum number of possible switches. However, this performance is not seen in hard comparison configurations, where $\mathcal{CMS}$ clearly outperforms $\mathcal{HAK}$.

To illustrate the combined cost of sampling and switching for our systems, we present Tables 36 and 37 as the combined cost of observations in Table 34 and switches in Table 35. Hong and Nelson [28] perform an analysis of total costs when switching costs a factor of 1, 10, 100, and 1000 times larger than the sampling costs per observation. We feature experimental results for the first two factors, 1 and 10; the

Table 35: Average number of required switches for procedures with k independent systems, $s = 3$ constraints, b feasible systems, and $v = 1$ infeasible constraints.

	DM ($\sigma^2_{x_i}/\sigma^2_{y_{i\ell}}$)				MIM ($\sigma^2_{x_i}/\sigma^2_{y_{i\ell}}$)			
$k = 15, b = 8$	L/L	L/H	H/L	H/H	L/L	L/H	H/L	H/H
$\mathcal{HAK}$	320	30	12415	2934	213	30	6417	1904
$\mathcal{HAK}+$	2107	17131	14544	23449	985	7994	7142	12361
$\mathcal{MD}_R$	1726	2415	14535	19678	702	1578	7137	9515
$\mathcal{CMS}$	30	30	30	30	29	30	29	30
$k = 101, b = 51$								
$\mathcal{HAK}$	2350	202	101139	21311	451	201	14353	4131
$\mathcal{HAK}+$	21178	150858	104305	230831	2067	20140	15954	12393
$\mathcal{MD}_R$	17862	23473	103617	197744	1385	5350	15591	9547
$\mathcal{CMS}$	202	202	202	202	134	173	163	202

other two factors will yield results that are more favorable to $\mathcal{CMS}$. When switches and samples are weighted equally, Table 36 shows that the relative efficiency of $\mathcal{CMS}$ improves compared to the other procedures. In fact, even in this extreme case, our switching procedure is the best performer for the (L/H) variance configuration when feasibility check is hard.

Table 37 displays the resulting cost if the time switching between simulated systems takes ten times as long as obtaining an observation from a system. As costs tip towards switching, the results favor our switching procedure substantially. $\mathcal{CMS}$ is clearly the efficient choice under these conditions for all mean and variance configurations, significantly improving on the other procedures in all cases, and featuring as little as a quarter of the combined sampling and switching costs in the best case (H/L). Even when $\mathcal{HAK}$ requires $2k$ switches, we still find $\mathcal{CMS}$ to be the best performer, as $\mathcal{CMS}$ requires fewer *samples* in these cases. As the switching costs are multiplied by an even larger factor, we expect to see an even wider advantage in using $\mathcal{CMS}$ with any number of systems.

Table 36: Average total cost of switches and samples for procedures with k independent systems, $s = 3$ constraints, b feasible systems, and $v = 1$ infeasible constraints when switches and samples are equally costly.

	DM ($\sigma^2_{x_i}/\sigma^2_{y_{i\ell}}$)				MIM ($\sigma^2_{x_i}/\sigma^2_{y_{i\ell}}$)			
$k = 15, b = 8$	L/L	L/H	H/L	H/H	L/L	L/H	H/L	H/H
$\mathcal{HAK}$	2982	23714	27177	29541	1372	9783	13787	13558
$\mathcal{HAK}+$	4499	34547	29373	47183	2255	16273	14569	25007
$\mathcal{MD}_R$	3943	13190	29363	41779	1764	7178	14561	20148
$\mathcal{CMS}$	3502	12294	31675	38372	1862	6130	19192	22305
$k = 101, b = 51$								
$\mathcal{HAK}$	27045	225516	224656	267393	4017	25756	31823	33558
$\mathcal{HAK}+$	44275	303635	230529	463581	6053	42199	33827	46815
$\mathcal{MD}_R$	39645	123208	229810	418186	4810	19233	33136	35873
$\mathcal{CMS}$	31006	107221	281694	355065	5366	14890	43480	51876

Table 37: Average total cost of switches and samples for procedures with k independent systems, with $s = 3$ constraints, b feasible systems, and $v = 1$ infeasible constraints when switches are ten times as costly as samples.

	DM ($\sigma^2_{x_i}/\sigma^2_{y_{i\ell}}$)				MIM ($\sigma^2_{x_i}/\sigma^2_{y_{i\ell}}$)			
$k = 15, b = 8$	L/L	L/H	H/L	H/H	L/L	L/H	H/L	H/H
$\mathcal{HAK}$	5862	23984	138912	55947	3289	10053	71540	30694
$\mathcal{HAK}+$	23462	188726	160269	258224	11120	88219	78847	136256
$\mathcal{MD}_R$	19477	34925	160178	218881	8082	21380	78794	105783
$\mathcal{CMS}$	3772	12564	31945	38642	2123	6400	19453	22575
$k = 101, b = 51$								
$\mathcal{HAK}$	48195	227334	1134907	459192	8076	27565	161000	70737
$\mathcal{HAK}+$	234877	1661357	1169274	2541060	24656	223459	177413	158352
$\mathcal{MD}_R$	200403	334465	1162363	2197882	17275	67383	173455	121796
$\mathcal{CMS}$	32824	109039	283512	356883	6572	16447	44947	53694

6.4.2.2 Systems under CRN

In this section, we examine the performance of $\mathcal{HAK}$ and $\mathcal{CMS}$ with our new modified variance estimates under CRN. Tables 38, 39, 40, and 41 display the observed PCS and the average number of required observations for $k \in \{15, 101\}$, respectively, for a similar system setup as in Section 6.4.2.1, but with differing levels of induced correlation. We compare $\mathcal{HAK}$ and $\mathcal{CMS}$ applied to independent systems, with versions of $\mathcal{HAK}$ and $\mathcal{CMS}$ modified for correlated systems with induced correlation $\rho_x \in \{0, 0.1, 0.25, 0.5, 0.75, 0.9\}$ and adjusted parameters $\beta_2 = s\beta_1 = \alpha/k$, the parameters required for valid selection of the best feasible system under correlation in Lemma 2 in Chapter 5. We denote the procedures with these parameters as $\mathcal{HAK}(\rho_x)$ and $\mathcal{CMS}(\rho_x)$. The parameter adjustment produces slightly higher PCS and number of required observations than the independent case, but allows for valid feasibility check under correlation.

Table 38: Observed PCS for procedures with $k = 15$ systems, with $s = 3$ constraints, $b = 8$ feasible systems, and $v = 1$ infeasible constraints with induced correlation (ρ_x). PCS below $1 - \alpha = 0.95$ marked in bold.

	DM ($\sigma^2_{x_i}/\sigma^2_{y_{i\ell}}$)				MIM ($\sigma^2_{x_i}/\sigma^2_{y_{i\ell}}$)			
	L/L	L/H	H/L	H/H	L/L	L/H	H/L	H/H
$\mathcal{HAK}$	0.985	0.998	0.975	0.983	0.997	0.998	0.995	0.997
$\mathcal{HAK}(0)$	0.990	0.999	0.978	0.988	0.995	0.998	0.998	0.994
$\mathcal{HAK}(0.1)$	0.990	0.999	0.980	0.986	0.994	1.000	0.997	0.998
$\mathcal{HAK}(0.25)$	0.987	0.998	0.971	0.994	0.995	0.998	0.995	0.998
$\mathcal{HAK}(0.5)$	0.991	0.998	0.981	0.987	0.998	0.998	0.997	0.997
$\mathcal{HAK}(0.75)$	0.995	0.999	0.983	0.993	0.997	0.999	0.994	0.993
$\mathcal{HAK}(0.9)$	0.996	1.000	**0.940**	0.994	0.993	1.000	0.979	0.994
$\mathcal{CMS}$	0.993	0.994	0.994	0.989	1.000	1.000	1.000	1.000
$\mathcal{CMS}(0)$	0.994	0.997	0.995	0.997	1.000	1.000	1.000	1.000
$\mathcal{CMS}(0.1)$	0.996	0.997	0.993	0.993	1.000	1.000	1.000	1.000
$\mathcal{CMS}(0.25)$	0.991	0.991	0.992	0.992	1.000	1.000	0.999	1.000
$\mathcal{CMS}(0.5)$	0.988	0.995	0.979	0.982	1.000	1.000	0.998	1.000
$\mathcal{CMS}(0.75)$	0.986	0.980	**0.939**	0.978	1.000	1.000	1.000	1.000
$\mathcal{CMS}(0.9)$	0.984	0.981	**0.748**	**0.868**	1.000	1.000	0.970	0.999

Table 39: Observed PCS for procedures with $k = 101$ systems, with $s = 3$ constraints, $b = 50$ feasible systems, and $v = 1$ infeasible constraints with induced correlation (ρ_x). PCS below $1 - \alpha = 0.95$ marked in bold.

	DM ($\sigma^2_{x_i}/\sigma^2_{y_{i\ell}}$)				MIM ($\sigma^2_{x_i}/\sigma^2_{y_{i\ell}}$)			
	L/L	L/H	H/L	H/H	L/L	L/H	H/L	H/H
$\mathcal{HAK}$	0.999	0.993	0.981	0.999	1.000	1.000	0.999	1.000
$\mathcal{HAK}(0)$	0.993	1.000	0.985	0.990	0.999	0.999	0.999	1.000
$\mathcal{HAK}(0.1)$	0.990	1.000	0.984	0.991	0.999	1.000	1.000	1.000
$\mathcal{HAK}(0.25)$	0.993	0.999	0.986	0.996	1.000	1.000	0.998	0.997
$\mathcal{HAK}(0.5)$	0.994	1.000	0.990	0.994	1.000	1.000	0.999	0.999
$\mathcal{HAK}(0.75)$	0.995	0.999	0.987	0.995	0.998	0.999	1.000	1.000
$\mathcal{HAK}(0.9)$	0.998	0.999	**0.910**	0.998	0.995	1.000	0.996	0.998
$\mathcal{CMS}$	1.000	0.996	0.999	0.999	1.000	1.000	1.000	0.998
$\mathcal{CMS}(0)$	1.000	0.999	0.999	1.000	0.999	0.999	1.000	1.000
$\mathcal{CMS}(0.1)$	0.998	1.000	1.000	0.999	0.999	1.000	1.000	1.000
$\mathcal{CMS}(0.25)$	0.997	0.999	0.996	0.999	1.000	0.999	1.000	1.000
$\mathcal{CMS}(0.5)$	0.997	1.000	0.992	0.998	0.999	1.000	0.999	0.997
$\mathcal{CMS}(0.75)$	0.994	0.984	**0.932**	0.985	0.995	1.000	0.994	0.998
$\mathcal{CMS}(0.9)$	0.977	0.975	**0.717**	**0.891**	0.991	1.000	**0.940**	0.984

For most values of induced correlation, we see higher than nominal PCS for both procedures. However, PCS suffers in configurations with correlation over 0.9 in $\mathcal{HAK}$ and over 0.75 in $\mathcal{CMS}$ when comparison is difficult (H/L). In addition, we see degraded PCS for $\mathcal{CMS}$ with H/H variances and $\rho_x = 0.9$. In H/L, we see similar PCS results for $\mathcal{CMS}$ as for the $\mathcal{SSM}$ comparison procedure in Table 32. As comparison becomes relatively less difficult, Tables 38 and 39 show that the degradation in PCS for the constrained R&S becomes much less pronounced. Since the PCS is split between feasibility check and comparison, small losses in PCS due to poor comparison can be hidden by strong performance in the feasibility check. Also, since only small gaps in sample size develop in most configurations, due to low and equal variances, this effectively eliminates the worst cases seen in Table 32. Still, high correlation can cause poor selection when comparison is hard.

In terms of sampling, Tables 40 and 41 shows that CRN significantly reduce the number of observations needed. The new values of β_1 and β_2 used for correlated

Table 40: Average number of required samples for procedures with $k = 15$ systems with $s = 3$ constraints, $b = 8$ feasible systems, and $v = 1$ infeasible constraints with induced correlation (ρ_x).

	DM ($\sigma^2_{x_i}/\sigma^2_{y_{i\ell}}$)				MIM ($\sigma^2_{x_i}/\sigma^2_{y_{i\ell}}$)			
	L/L	L/H	H/L	H/H	L/L	L/H	H/L	H/H
$\mathcal{HAK}$	2675	23716	14739	26504	1165	9773	7174	11565
$\mathcal{HAK}(0)$	2679	23676	14686	26580	1177	9779	7179	11638
$\mathcal{HAK}(0.1)$	2601	23799	13343	25734	1119	9802	6454	11165
$\mathcal{HAK}(0.25)$	2496	23702	11254	24778	1064	9781	5588	10611
$\mathcal{HAK}(0.5)$	2403	23667	7765	23787	992	9750	3749	9914
$\mathcal{HAK}(0.75)$	2381	23663	4445	23627	980	9757	2097	9717
$\mathcal{HAK}(0.9)$	2373	23645	2640	23711	972	9759	1153	9760
$\mathcal{CMS}$	3472	12264	31645	38342	1833	6100	19163	22275
$\mathcal{CMS}(0)$	3524	12290	32860	38967	1861	6084	19715	22389
$\mathcal{CMS}(0.1)$	3297	12167	28867	36641	1722	6001	17300	20743
$\mathcal{CMS}(0.25)$	2840	11763	24121	31391	1481	5885	14125	17413
$\mathcal{CMS}(0.5)$	2105	11216	15978	22858	1100	5688	9326	12296
$\mathcal{CMS}(0.75)$	1435	10576	8196	15381	774	5438	4406	7768
$\mathcal{CMS}(0.9)$	1181	9923	3743	11875	685	5304	1942	5948

Table 41: Average number of required samples for procedures with $k = 101$ systems with $s = 3$ constraints, $b = 50$ feasible systems, and $v = 1$ infeasible constraints with induced correlation (ρ_x).

	DM ($\sigma^2_{x_i}/\sigma^2_{y_{i\ell}}$)				MIM ($\sigma^2_{x_i}/\sigma^2_{y_{i\ell}}$)			
	L/L	L/H	H/L	H/H	L/L	L/H	H/L	H/H
$\mathcal{HAK}$	24695	225314	123517	246082	3566	25555	17470	29427
$\mathcal{HAK}(0)$	24718	225297	124372	246004	3578	25583	17606	29570
$\mathcal{HAK}(0.1)$	24119	224838	112493	239726	3476	25547	15874	28435
$\mathcal{HAK}(0.25)$	23354	224854	95304	232397	3379	25618	13589	27110
$\mathcal{HAK}(0.5)$	22705	224589	66965	226104	3262	25605	9616	25893
$\mathcal{HAK}(0.75)$	22598	224696	38525	224828	3247	25580	5537	25573
$\mathcal{HAK}(0.9)$	22589	224992	24229	224680	3239	25572	3557	25564
$\mathcal{CMS}$	30804	107019	281492	354863	5232	14717	43317	51674
$\mathcal{CMS}(0)$	31276	107486	282612	358029	5293	14777	43980	52432
$\mathcal{CMS}(0.1)$	28825	105086	252079	338199	4929	14659	38478	47759
$\mathcal{CMS}(0.25)$	25136	102662	217250	294009	4492	14324	32018	40428
$\mathcal{CMS}(0.5)$	18708	97771	148440	219979	3716	13982	21048	29535
$\mathcal{CMS}(0.75)$	13179	92258	78773	142697	3074	13581	10805	19122
$\mathcal{CMS}(0.9)$	10530	88502	41223	105488	2809	13230	5347	14638

systems cause the procedures to perform slightly worse when applied to truly independent systems than procedures designed for independent systems. Discrepancies to this rule occur in Tables 40 and 41, but these results are well within the standard error of estimates. As correlation increases, the procedures exploiting CRN will require fewer observations. Even at modest levels of correlation, we can see significant improvement over the independent case. Savings due to CRN are restricted to the comparison phase, so (L/H) configurations feature only a small advantage for implementing CRN, while other variance configurations display larger savings. Not surprisingly, H/L configurations feature the largest savings. Difficult feasibility check configurations also require higher levels of correlation ($\rho_x > 0.25$) to see improvements. Percentage savings are similar for $k = 15$ and $k = 101$ systems. Since comparison dominates sampling for $\mathcal{CMS}$, we observe larger percentage gains under CRN for $\mathcal{CMS}$ than for $\mathcal{HAK}$.

Even with large possible savings under CRN, the PCS degradation for $\mathcal{HAK}$ and $\mathcal{CMS}$ may still cause concern. Tables 42 through 45 present the effectiveness of our heuristic variance modifications under three levels of correlation, $\rho_x \in \{0.1, 0.5, 0.9\}$ for $k = 15$ systems. Table 42 displays the observed PCS for our procedures with and without the heuristic modifications, for $\rho_x = 0.9$. Tables 43, 44, and 45 display the average number of required samples for our procedures, with and without the heuristic modifications, for $\rho_x = 0.9$, $\rho_x = 0.5$, and $\rho_x = 0.1$, respectively. The original procedures with independent systems are denoted $\mathcal{HAK}$ and $\mathcal{CMS}$, while correlation of ρ_x is induced in $\mathcal{HAK}(\rho_x)$ and $\mathcal{CMS}(\rho_x)$. $\mathcal{HAK}(\rho_x) + TS_i$ and $\mathcal{CMS}(\rho_x) + TS_i$ denote an implementation of $\mathcal{HAK}$ and $\mathcal{CMS}$ with the variance modification TS_i for $i = \{1, 2, 3, 4\}$. For the sake of brevity, we feature only configurations with $k = 15$ systems, but similar results were found for $k = 101$ systems.

Table 42 shows that all four variance modifications display a marked improvement in PCS, raising observed values above 0.988 in all configurations. The TS_4 modification tends to provide the smallest PCS, which experimentally confirms it to be the most aggressive modification.

Table 42: Observed PCS for procedures with $k = 15$ systems, with $s = 3$ constraints, $b = 8$ feasible systems, and $v = 1$ infeasible constraints with induced correlation $\rho_x = 0.9$.

	DM ($\sigma^2_{x_i}/\sigma^2_{y_{i\ell}}$)				MIM ($\sigma^2_{x_i}/\sigma^2_{y_{i\ell}}$)			
	L/L	L/H	H/L	H/H	L/L	L/H	H/L	H/H
$\mathcal{HAK}$	0.985	0.998	0.975	0.983	0.997	0.998	0.995	0.997
$\mathcal{HAK}(0.9)$	0.996	1.000	**0.940**	0.994	0.993	1.000	0.979	0.994
$\mathcal{HAK}(0.9) + TS_1$	0.997	1.000	0.998	0.996	0.997	1.000	0.998	0.996
$\mathcal{HAK}(0.9) + TS_2$	0.997	1.000	0.989	0.996	0.997	1.000	0.998	0.996
$\mathcal{HAK}(0.9) + TS_3$	0.997	1.000	0.989	0.996	0.997	1.000	0.998	0.996
$\mathcal{HAK}(0.9) + TS_4$	0.996	1.000	0.988	0.996	0.997	1.000	0.998	0.996
$\mathcal{CMS}$	0.993	0.994	0.994	0.989	1.000	1.000	1.000	1.000
$\mathcal{CMS}(0.9)$	0.984	0.981	**0.748**	**0.868**	1.000	1.000	0.970	0.999
$\mathcal{CMS}(0.9) + TS_1$	0.996	0.998	0.998	0.996	1.000	1.000	1.000	1.000
$\mathcal{CMS}(0.9) + TS_2$	0.996	0.998	0.998	0.996	1.000	1.000	1.000	1.000
$\mathcal{CMS}(0.9) + TS_3$	0.996	0.998	0.996	0.996	1.000	1.000	1.000	1.000
$\mathcal{CMS}(0.9) + TS_4$	0.996	0.998	0.988	0.992	1.000	1.000	1.000	1.000

While the PCS results in Table 42 may seem similar, we see in Tables 43, 44, and 45 that the additional observations required to secure PCS depends highly on the modifications implemented. We now discuss the results in more detail.

In $\mathcal{HAK}$, we notice two patterns. First, in almost all configurations, TS_4 is the most efficient, followed by TS_3, TS_2, and TS_1, in that order. In the special case (L/H) with difficult feasibility check, no additional observations are required for comparison. Therefore, all approaches perform equally well.

For $\mathcal{CMS}$, we observe different behavior. Since systems almost never reach equal sample sizes in $\mathcal{CMS}$, we see that the estimates TS_1 and TS_2 produce equivalent results. These modifications are very conservative, sometimes requiring five times the number of samples in Table 43 as the other adjustments. Under the low correlations of Tables 44 and 45, TS_1 and TS_2 do not require substantially more samples, but still are not preferable. Even though these modifications are conservative, they still outperform the independently sampled case in all instances, except for $\mathcal{HAK}$ with the TS_1 modification under the H/L variance configuration.

TS_3 and TS_4 fair much better than the first two modifications, but TS_4 is the superior choice of the variance estimate modifications for all configurations in Tables 43, 44, and 45. For only as little as 0.15% in Table 45 and at most 54% additional samples in Table 43 than the original procedure for correlated systems, TS_4 provides efficiency and good PCS. TS_4 is the most efficient modification for both $\mathcal{HAK}$ and $\mathcal{CMS}$. Utilizing this modification sacrifices only a small amount of samples to provide a good PCS and still significantly outperforms the independently sampled case.

Table 43: Average number of observations for procedures with $k = 15$ systems, with $s = 3$ constraints, $b = 8$ feasible systems, and $v = 1$ infeasible constraints with induced correlation $\rho_x = 0.9$.

	DM ($\sigma^2_{x_i}/\sigma^2_{y_{i\ell}}$)				MIM ($\sigma^2_{x_i}/\sigma^2_{y_{i\ell}}$)			
	L/L	L/H	H/L	H/H	L/L	L/H	H/L	H/H
$\mathcal{HAK}$	2675	23716	14739	26504	1165	9773	7174	11565
$\mathcal{HAK}(0.9)$	2373	23645	2640	23711	972	9759	1153	9760
$\mathcal{HAK}(0.9) + TS_1$	2591	23645	14992	25953	1158	9759	7456	11645
$\mathcal{HAK}(0.9) + TS_2$	2434	23645	2789	24327	1062	9759	1402	10652
$\mathcal{HAK}(0.9) + TS_3$	2380	23645	2787	23780	985	9759	1401	9895
$\mathcal{HAK}(0.9) + TS_4$	2373	23645	2735	23714	974	9759	1351	9778
$\mathcal{CMS}$	3472	12264	31645	38342	1833	6100	19163	22275
$\mathcal{CMS}(0.9)$	1181	9923	3743	11875	685	5304	1942	5948
$\mathcal{CMS}(0.9) + TS_1$	2717	11050	24328	31137	1527	5722	12893	15839
$\mathcal{CMS}(0.9) + TS_2$	2717	11050	24328	31137	1527	5722	12893	15839
$\mathcal{CMS}(0.9) + TS_3$	1825	10450	12913	19908	1000	5462	6991	10046
$\mathcal{CMS}(0.9) + TS_4$	1435	10414	5110	14796	773	5452	2983	7507

6.5 Conclusions

We present a procedure, $\mathcal{CMS}$, for constrained R&S that minimizes the number of switches between simulated systems while finding the best constrained system. This

Table 44: Average number of observations for procedures with $k = 15$ systems, with $s = 3$ constraints, $b = 8$ feasible systems, and $v = 1$ infeasible constraints with induced correlation $\rho_x = 0.50$.

	DM ($\sigma^2_{x_i}/\sigma^2_{y_{i\ell}}$)				MIM ($\sigma^2_{x_i}/\sigma^2_{y_{i\ell}}$)			
	L/L	L/H	H/L	H/H	L/L	L/H	H/L	H/H
$\mathcal{HAK}$	2675	23716	14739	26504	1165	9773	7174	11565
$\mathcal{HAK}(0.5)$	2403	23667	7765	23787	992	9750	3749	9914
$\mathcal{HAK}(0.5) + TS_1$	2633	23667	15259	26076	1160	9750	7447	11600
$\mathcal{HAK}(0.5) + TS_2$	2476	23667	7771	24520	1090	9750	3807	10893
$\mathcal{HAK}(0.5) + TS_3$	2410	23667	7769	23854	1001	9750	3788	10013
$\mathcal{HAK}(0.5) + TS_4$	2404	23667	7767	23802	996	9750	3767	9955
$\mathcal{CMS}$	3472	12264	31645	38342	1833	6100	19163	22275
$\mathcal{CMS}(0.5)$	2105	11216	15978	22858	1100	5688	9326	12296
$\mathcal{CMS}(0.5) + TS_1$	3183	11880	27956	35526	1663	5936	15873	19105
$\mathcal{CMS}(0.5) + TS_2$	3183	11880	27956	35526	1663	5936	15873	19105
$\mathcal{CMS}(0.5) + TS_3$	2266	11330	17436	24597	1185	5727	10191	13271
$\mathcal{CMS}(0.5) + TS_4$	2176	11316	16498	23491	1130	5725	9615	12614

Table 45: Average number of observations for procedures with $k = 15$ systems, with $s = 3$ constraints, $b = 8$ feasible systems, and $v = 1$ infeasible constraints with induced correlation $\rho_x = 0.10$.

	DM ($\sigma^2_{x_i}/\sigma^2_{y_{i\ell}}$)				MIM ($\sigma^2_{x_i}/\sigma^2_{y_{i\ell}}$)			
	L/L	L/H	H/L	H/H	L/L	L/H	H/L	H/H
$\mathcal{HAK}$	2675	23716	14739	26504	1165	9773	7174	11565
$\mathcal{HAK}(0.1)$	2601	23799	13343	25734	1119	9802	6454	11165
$\mathcal{HAK}(0.1) + TS_1$	2697	23799	15893	26779	1188	9802	7634	11856
$\mathcal{HAK}(0.1) + TS_2$	2614	23799	13342	25891	1149	9802	6462	11466
$\mathcal{HAK}(0.1) + TS_3$	2602	23799	13343	25749	1120	9802	6456	11179
$\mathcal{HAK}(0.1) + TS_4$	2602	23799	13343	25738	1120	9802	6455	11170
$\mathcal{CMS}$	3472	12264	31645	38342	1833	6100	19163	22275
$\mathcal{CMS}(0.1)$	3297	12167	28867	36641	1722	6001	17300	20743
$\mathcal{CMS}(0.1) + TS_1$	3531	12366	31254	39142	1842	6075	18470	21955
$\mathcal{CMS}(0.1) + TS_2$	3531	12366	31254	39142	1842	6075	18470	21955
$\mathcal{CMS}(0.1) + TS_3$	3301	12173	28913	36705	1724	6003	17320	20765
$\mathcal{CMS}(0.1) + TS_4$	3301	12172	28911	36666	1724	6003	17316	20759

is desirable, as the cost of switching can be expensive. We prove the validity of this procedure, guaranteeing a nominal probability of selecting the best feasible system for independently sampled systems.

To improve the efficiency of the procedure, we also wish to utilize common random numbers (CRN) to reduce variance within comparison. We show how strong positive correlation can adversely affect the probability of correct selection (PCS) for procedures, such as $\mathcal{CMS}$, that use two-sample comparison, because of the underestimation of the variance during the comparison. To achieve the nominal PCS while still increasing efficiency, we propose four variance modifications.

Our experiments show that $\mathcal{CMS}$ is an efficient option, if the cost of switching is larger than the cost of sampling or the feasibility check phase is difficult. Ensuring a minimal number of switches requires extra observations, but CRN can reduce the number of necessary samples. Our experiments show that PCS under high correlation is a concern, but the heuristic variance modifications provide good PCS, and some of them also preserve a large portion of the savings due CRN.

CHAPTER VII

CONTRIBUTIONS AND FUTURE RESEARCH

This thesis aims to advance the fields of steady-state output analysis and constrained ranking and selection (R&S), by providing new methods and procedures that improve efficiency. In this chapter, we summarize the main contributions of the thesis in Section 7.1 and present future research possibilities in Section 7.2.

7.1 Contributions

In Chapter 2, we introduce the overlapping modified jackknifed Durbin–Watson (OM) estimator and show that the OM estimator has several advantageous properties, including low bias, low variance, and an approximate χ^2 distribution. In Chapter 3, we investigate the use of overlapping variance estimators within steady-state R&S procedures. In our experimental study, we show that these variance estimators, including our new OM estimator, result in significant savings in the number of samples required to reach a decision.

In Chapter 4, we present a new framework for constrained R&S that allows certain systems to become dormant, halting sampling for those systems as the procedure continues. A system goes dormant when it is found inferior to another system whose feasibility has not been determined, and returns to contention only if its superior system is eliminated. This framework greatly reduces the number of required samples to choose the best feasible systems, especially when feasibility check is difficult. We provide three approaches to implement this framework within simultaneously-running procedures and show that one of them is statistically valid.

Chapter 5 presents three general procedures for constrained R&S. While previous procedures have been designed for one constraint, the three new procedures can incorporate any number of constrained performance measures. In addition, we show that the two simultaneously-running procedures can be extended to select the best feasible system under correlation. This extension allows the use of common random numbers (CRN), which have never been presented within constrained R&S, but are shown to be useful at improving the efficiency of the procedures.

In Chapter 6, we develop another procedure for constrained R&S with a different goal, namely minimizing the number of switches between simulated systems. We prove its validity for independently-simulated systems and demonstrate its usefulness when sampling plus switching costs are considered. However, we also show that CRN cannot be safely utilized within the procedure due to degradation of the probability of correct selection (PCS) (without substantial modification). To address this problem, we present four variance modifications, which preserve PCS in exchange for additional observations.

7.2 Future Research

There are a few topics that could be pursued within the subject areas of this thesis.

1. In Chapter 2, additional configurations of the overlapping area and overlapping CvM estimators should be inspected (analytically or experimentally) to find other configurations that provide good bias and low variance, besides the Durbin–Watson approach.

2. Error allocation strategies could be developed based on first-stage estimates of the relative difficulty of feasibility check and comparison for use within our procedures developed in Chapters 5 and 6.

3. In Chapter 5, a thorough investigation of the correlation required for $\mathcal{HAK}+$ and $\mathcal{MD}_R$ with CRN to improve on the independently sampled versions of the

same procedures would demonstrate the usefulness of CRN within our procedures.

REFERENCES

[1] Aktaran-Kalayci, T., Alexopoulos, C., Argon, N.T., Goldsman, D., and Wilson, J. R. Exact expected values of variance estimators for simulation. *Naval Research Logistics*, vol. 54, pp. 397–410, 2007.

[2] Alexopoulos, C., Argon, N. T., Goldsman, D., Steiger, N. M., Tokol, G., and Wilson, J. R. "Efficient computation of overlapping variance estimators for simulation," *INFORMS Journal on Computing*, vol. 19, pp. 314–327, 2007.

[3] Alexopoulos, C., Argon, N. T., Goldsman, D., Tokol, G., and Wilson, J. R. "Overlapping estimators for simulation," *Operations Research*, vol. 55, pp. 1090–1103, 2007.

[4] Alexopoulos, C. and Seila, A. F. "Output data analysis," in Handbook of Simulation, ed. Jerry Banks, pp. 225–272. Wiley: New York, 1998.

[5] Andradóttir, S., and Kim, S.-H. "Fully sequential procedures for comparing constrained systems via simulation," to appear in *Naval Research Logistics*, 2010.

[6] Batur, D., Goldsman, D., and Kim, S.-H. "An improved standardized time series Durbin–Watson variance estimator for steady-state simulation," *Operations Research Letters*, vol. 34, pp. 285–289, 2009.

[7] Batur, D., and Kim, S.-H. "Procedures for feasibility check when the number of systems or constraints is large," to appear in *ACM Transactions on Modeling and Computer Simulation*, 2010.

[8] Bechhofer, R. E., Santner, T. J., and Goldsman, D. *Design and Analysis of Experiments for Statistical Selection, Screening, and Multiple Comparison.* John Wiley & Sons, New York, 1995.

[9] Billingsley, P. *Convergence of Probability Measures.* John Wiley & Sons, New York, 1968.

[10] Borodin, A.N., and Salminen, P. *Handbook of Brownian Motion – Facts and Formulae*, second edition. Birkhäuser Verlag, Basel-Berlin-Boston, 2002.

[11] Butler, J., Morrice, D.J., and Mullarkey, P.W. "A multiple attribute utility theory approach to ranking and selection," *Management Science*, vol. 47, pp. 800–816, 2001.

[12] Chen, E.J., and Lee, L.H. "A multi-objective selection procedure of determining a Pareto set," *Computers & Operations Research*, vol. 36, pp. 1872–1879, 2009.

[13] Chen, E.J., and Kelton, W.D. "Sequential selection procedures: Using sample means to improve efficiency," *European Journal of Operational Research*, vol. 166, pp. 133–153, 2005.

[14] Chen, C.-H. "A lower bound for the correct subset-selection probability and its application to discrete event simulation," *IEEE Transactions on Automatic Control*, vol. 41, pp. 1227–1231, 1996.

[15] Chen, H.C., Chen, C.-H., and Yücesan, E. "Computing efforts allocation for ordinal optimization and discrete event simulation," *IEEE Transactions on Automatic Control*, vol 45, pp. 960–964, 2000.

[16] Chick, S.E. "Subjective probability and Bayesian methodology," in *Handbooks in Operations Research & Management Science: Simulation*, eds. S.G. Henderson and B.L. Nelson, pp. 225–257. North-Holland: Amsterdam, 2006.

[17] Chick, S.E., and Inoue, K. "New two-stage and sequential procedures for selection the best simulated system," *Operations Research*, vol. 49, pp. 732–743, 2001.

[18] Chick, S.E., and Inoue, K. "New procedures for identifying the best simulated system using common random numbers," *Management Science*, vol. 47, pp. 1133–1149, 2001.

[19] Dudewicz, E.J., and S.R. Dalal. "Allocation of observations in ranking and selection with unequal variances," *Sankhya*, vol. 37, pp. 28–78, 1975.

[20] Foley, R. D., and Goldsman, D. "Confidence intervals using orthonormally weighted standardized time series," *ACM TOMACS,* vol. 9, pp. 297–325, 1999.

[21] Glynn, P. W., and Iglehart, D. "Simulation output analysis using standardized time series," *Mathematics of Operations Research*, vol.15, pp. 1–16, 1990.

[22] Goldsman, D., Kang, K., Kim, S.-H., Seila, A. F., and Tokol, G. "Combining standardized time series area and Cramér–von Mises variance estimators," *Naval Research Logistics*, vol. 54, pp. 384–396, 2007.

[23] Goldsman, D., Kang, K., and Seila, A. F. "Cramér–von Mises variance estimators for simulations," *Operations Research*, vol. 47, pp. 299–309, 1999.

[24] Goldsman, D., Kim, S.-H., Marshall, W., and Nelson, B. L. 2002. "Ranking and selection procedures for steady-state simulation: Perspectives and procedures," *INFORMS Journal on Computing*, vol. 14, pp. 2–19, 2002.

[25] Goldsman, D., and Marshall, W. "Selection procedures with standardized time series variance estimators," in *Proceedings of the 1999 Winter Simulation Conference*, ed. P. A. Farrington, H. B. Nembhard, D. T. Sturrock, and G. W. Evans, pp. 382–388. Institute of Electrical and Electronics Engineers: Piscataway, New Jersey, 1999.

[26] Goldsman, D., Meketon, M. S., and Schruben, L. W. 1990. "Properties of standardized time series area variance estimators," *Management Science*, vol. 36, pp. 393–397, 1990.

[27] Hartmann, M. "An improvement on Paulson's procedure for selecting the population with the largest mean from k normal populations with a common unknown variance," *Sequential Analysis* vol. 10, pp. 1–16, 1991.

[28] Hong, L.J., and Nelson, B.L. "The tradeoff between sampling and switching: New sequential procedures for indifference-zone selection," *IIE Transactions*, vol. 37, pp. 623–634, 2005.

[29] Kabirian, A., and Ólafsson, S. "Selection of the best with stochastic constraints," in *Proceeding of the 2009 Winter Simulation Conference*, eds. M.D. Rossetti, R.R. Hill, B. Johansson, A. Dunkin and R.G. Ingalls, pp. 584–589. Institute of Electrical and Electronics Engineers, Piscataway, New Jersey, 2009.

[30] Kim, S.-H., and Nelson, B.L. "A fully sequential procedure for indifference-zone selection in simulation," *ACM TOMACS*, vol. 11, pp. 251–273, 2001.

[31] Kim, S.-H., and Nelson, B.L. "On the asymptotic validity of fully sequential selection procedures for steady-state simulation," *Operations Research*, vol. 54, pp. 475–488, 2006.

[32] Kim, S.-H., and Nelson, B.L. "Selecting the best system," In *Handbooks in Operations Research & Management Science: Simulation*, eds. S.G. Henderson and B.L. Nelson, pp. 501-534, North-Holland: Amsterdam, 2009.

[33] Law, A., and Kelton, D. *Simulation Modeling and Analysis*, third edition. McGraw Hill: New York, 2000.

[34] Lee, L.H., Chew, E.P., Teng, S.Y., and Goldsman, D. "Optimal computing budget allocation for multi-objective simulation models," in *Proceedings of the 2004 Winter Simulation Conference*, eds. R.G. Ingalls, M.D. Rossetti, J.S. Smith, and B.A. Peters, pp. 586–594. Institute of Electrical and Electronics Engineers: Piscataway, New Jersey, 2004.

[35] Lee, L.H., Chew, E.P., and Teng, S.Y. "Integration of statistical selection with search mechanism for solving multi-objective simulation-optimization problems," in *Proceedings of the 2006 Winter Simulation Conference*, eds. L.F. Perrone, F.P. Wieland, J. Liu, B.G. Lawson, D.M. Nicol, and R.M. Fujimoto, pp. 294–303. Institute of Electrical and Electronics Engineers: Piscataway, New Jersey, 2006.

[36] Matejcik, F.J., and Nelson, B.L. "Using common random numbers for indifference-zone selection and multiple comparisons in simulation," *Management Science*, vol. 41, pp. 1935–1945, 1995.

[37] Morrice, D.J. and Butler, J. "Ranking and selection with multiple 'targets'," in *Proceeding of the 2006 Winter Simulation Conference*, eds. L.F. Perrone, F.P. Wieland, J. Liu, B.G. Lawson, D.M. Nicol, and R.M. Fujimoto, pp. 222–230. Institute of Electrical and Electronics Engineers, Piscataway, New Jersey, 2006.

[38] Mukhopadhyay, N. "Some comments on two-stage selection procedures," *Communications in Statistics–Theory and Methods*, vol. A8, pp. 671–683, 1979.

[39] Osogami, T. "Finding Probably Best Systems Quickly via Simulations," *ACM TOMACS*, vol. 19(3), Article 12, pp. 1–12, 2009.

[40] Paulson, E. "A sequential procedure for selection the population with the largest mean from k normal populations," *Annals of Mathematical Statistics*, vol. 35, pp. 174-180, 1964.

[41] Pichitlamken, J., Nelson, B.L., and Hong, L.J. "A sequential procedure for neighborhood selection of the best in optimization via simulation," *European Journal of Operational Research*, vol. 173, pp. 283–298, 2006.

[42] Pujowidianto, N.A., Lee, L.H., Chen, C.-H., and Yap, C.M. "Optimal computing budget allocation for constrained optimization," in *Proceeding of the 2009 Winter Simulation Conference*, eds. M.D. Rossetti, R.R. Hill, B. Johansson, A. Dunkin and R.G. Ingalls, pp. 574–583. Institute of Electrical and Electronics Engineers, Piscataway, New Jersey, 2009.

[43] Rinott, Y. "On two-stage selection procedures and related probability-inequalities," *Communications in Statistics–Theory and Methods*, vol. A7, pp. 799–811, 1978.

[44] Santner, T.J., and Tamhane, A.C. "Designing experiments for selecting a normal population with a large mean and a small variance," in *Design of Experiments – Ranking and Selection: Essays in Honor of Robert E. Bechhofer*, eds. T.J. Santner and A.C. Tamhane, pp. 178-198. Marcel-Dekker: New York, 1984.

[45] Satterthwaite, F. E. "Synthesis of variance," *Psychometrika*, vol. 6, pp. 309–316, 1941.

[46] Schruben, L. W. "Confidence interval estimation using standardized time series," *Operations Research*, vol. 31, pp. 1090–1108, 1983.

[47] Szechtman, R. and Yücesan, E. "A new perspective on feasibility determination," in *Proceedings of the 2008 Winter Simulation Conference*, eds. S.J. Mason, R.R. Hill, L. Mönch, O. Rose, T. Jefferson, J.W. Fowler, pp. 273-280. Institute of Electrical and Electronics Engineers, Piscataway, New Jersey, 2008.