Fully Sequential Selection Procedures with Control Variates

Shing Chih Tsai [1]

Department of Industrial and Information Management

National Cheng Kung University

No. 1, University Road, Tainan City, Taiwan

Barry L. Nelson

Department of Industrial Engineering & Management Sciences

Northwestern University

Evanston, IL 60208-3119, USA

June 12, 2008

[1]Corresponding author

Abstract

Fully sequential selection procedures have been developed in the field of stochastic simulation to find the simulated system with the best expected performance when the number of alternatives is finite. Kim and Nelson proposed the $\mathcal{KN}$ procedure to allow for unknown and unequal variances and the use of common random numbers. $\mathcal{KN}$ approximates the *raw sum* of differences between observations from two systems as a Brownian motion process with drift and uses a triangular continuation region to decide the stopping time of the selection process. In this paper we derive new fully sequential selection procedures that employ a more effective sum of differences which we called a *controlled sum*. Two provably valid procedures and an approximate procedure are described. Empirical results and a realistic illustration are provided to compare the efficiency of these procedures with other procedures that solve the same problem.

1 Introduction

In the stochastic simulation community, ranking and selection (R&S; see for instance Bechhofer, Santner and Goldsman 1995) procedures have been proposed to select the simulated system with the largest or smallest expected performance measure when the number of alternative designs is finite. There are many R&S procedures providing different types of guarantees. In this paper we focus on extending one well-known and effective procedure due to Kim and Nelson (2001, procedure $\mathcal{KN}$) which is a fully sequential R&S procedure. Such procedures take a single basic observation from each system still in contention at each stage, and eliminate systems whenever they are statistically inferior. This sort of elimination has been shown to greatly reduce the computational effort required to find the best system relative to two-stage procedures.

$\mathcal{KN}$ guarantees to select the best system with a pre-specified probability of correct selection (PCS) when the true expectations satisfy an indifference-zone requirement. To accomplish this, $\mathcal{KN}$ approximates the sum of differences between observations from each pair of systems as a Brownian motion process with drift, and uses a triangular continuation region to decide the stopping time of the selection process. Figure 1 shows the continuation region for this procedure for systems i and k. Either system i or system k will be eliminated depending on which direction the sum of differences exits this region. $\mathcal{KN}$ requires independent and identically distributed (i.i.d.) normal data but it allows unknown and unequal variances and the use of common random numbers (CRN). CRN is a technique that tries to generate a positive correlation between the outputs of different systems, and therefore reduce the variance of the difference between them, by using the same pseudorandom numbers to simulate each system.

Recently, more adaptive or cost-effective procedures have been derived from $\mathcal{KN}$ to address a variety of situations that are encountered in the stochastic simulation context. For instance, Goldsman et al. (2001) and Kim and Nelson (2006) proposed two R&S procedures

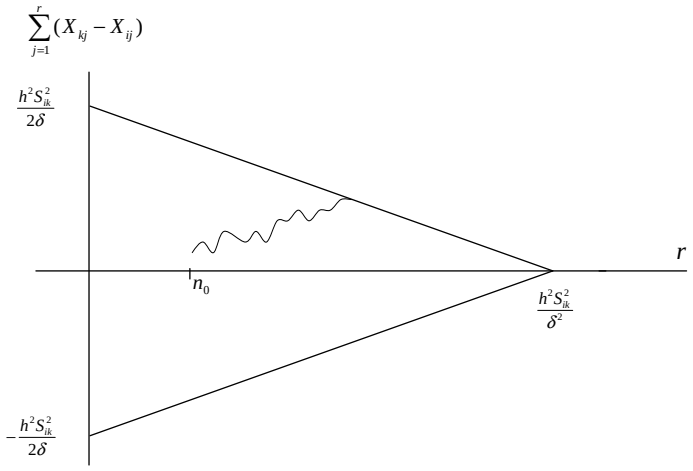

Figure 1: Continuation region for the fully sequential, indifference-zone procedure $\mathcal{KN}$

that extend $\mathcal{KN}$ to steady-state simulation experiments. Many simulation optimization algorithms try to move from a current solution to an improved solution on each iteration by choosing the best from a set of neighbors. Pichitlamken et al. (2006) proposed fully sequential procedures that can provide a statistical guarantee on each iteration of an optimization even if the initial sample sizes are unequal. Although $\mathcal{KN}$ can decrease the expected total number of samples necessary to achieve a decision, it needs to repeatedly switch among the different simulated systems to reduce the cost of sampling. Hong and Nelson (2005) proposed sequential procedures that attempt to balance the cost of sampling and switching to minimize the total computational cost. In practice it is possible that the number of systems is not fixed at the beginning of the experiment; instead the systems are revealed sequentially during the experiment. Hong and Nelson (2007) presented procedures to select the best each time new systems are revealed and provide the desired statistical guarantee whenever the experiment terminates. All of these fully sequential selection procedures mentioned above are based on the *raw sum* of differences between two systems' outputs.

Control variates (CV) is a variance reduction technique. Controls are random variables in the simulation that are correlated with the output of interest, but whose expected values

are known (Lavenberg and Welch 1981). Nelson and Staum (2006) derived two-stage R&S procedures that employ CV estimators, and Tsai, Nelson and Staum (2008) added screening in the first stage. These CV procedures can be more statistically efficient than their sample-mean-based counterparts since the CV estimator has a smaller variance than the conventional sample-mean estimator.

Our goal is to develop new fully sequential R&S procedures by employing a more effective sum of differences, which we called a *controlled sum*, instead of the raw sum of differences used in all previous work. A controlled sum of differences can be more statistically efficient than a raw sum of differences because the Brownian motion process based on it has reduced variance and the continuation region for the selection process has smaller area, leading to fully sequential procedures that are correspondingly more efficient. In a controlled sum of differences, the raw sum of differences is adjusted by a multiple $\boldsymbol{\beta}$ of the centered sum of control variates that is correlated with the raw sum of differences. This reduces the variance without changing the drift. The vector of coefficients, $\boldsymbol{\beta}$, is critical to the effectiveness of the controlled sum. When the optimal $\boldsymbol{\beta}$ is known, then our procedure is equivalent to $\mathcal{KN}$ applied to a lower-variance output process. In practice, however, the optimal $\boldsymbol{\beta}$ is not known and choosing $\boldsymbol{\beta}$ arbitrarily may degrade our procedure. Therefore, the key issue is estimating the optimal $\boldsymbol{\beta}$.

The paper is organized as follows: In Section 2, we present our CV model and a Generic Procedure from which specific procedures are derived. In Section 3 we introduce a fully sequential procedure assuming the optimal $\boldsymbol{\beta}$, which we denote as $\boldsymbol{\beta}^*$, is known and use this procedure to show the potential benefits of the controlled sum. Section 4 provides a statistically valid procedure when $\boldsymbol{\beta}^*$ is unknown. In Section 5 we present a procedure that combines the procedure in Section 4 and $\mathcal{KN}$. An approximate procedure which can not be proven to obtain the PCS guarantee but may require a smaller sample size is discussed in Section 6. Empirical results and a realistic illustration are provided in Sections 7 and 8, respectively. The paper ends with conclusions in Section 9. All proofs and most details of

the procedures are relegated to the Appendix.

2 The Generic Procedure

In this section we present the CV model on which our procedures are based, provide the definitions and notation that will be used throughout the paper and introduce the Generic Procedure utilizing a controlled sum. The description in Section 2.1 is based on Nelson and Staum (2006).

2.1 Control-Variate Model

Let X_{ij} be the jth simulation observation from system i, for $i = 1, 2, \ldots, k$. We assume that it can be represented as

$$X_{ij} = \mu_i + (\mathbf{C}_{ij} - \boldsymbol{\xi}_i)^T \boldsymbol{\beta}_i^* + \eta_{ij}, \tag{1}$$

where the $q_i \times 1$ vector $\mathbf{C}_{ij}$ is called the *control* and is assumed multivariate normal, while $\{\eta_{ij}, j = 1, 2, \ldots, n\}$ is a set of i.i.d. $\mathrm{N}(0, \tau_i^2)$ random variables. For each system $i = 1, 2, \ldots, k$, the controls $\{\mathbf{C}_{ij}, j = 1, 2, \ldots, n\}$ are also i.i.d., are independent of $\{\eta_{ij}, j = 1, 2, \ldots, n\}$ and have known expected value $\boldsymbol{\xi}_i$. The X_{ij} are therefore i.i.d. $\mathrm{N}(\mu_i, \sigma_i^2)$ random variables, with both μ_i and σ_i^2 unknown and (perhaps) unequal. Furthermore, for each pair of systems $i, \ell = 1, 2, \ldots, k, i \neq \ell$ the controls $(\mathbf{C}_{ij}^T, \mathbf{C}_{\ell j}^T)$ and the $(\eta_{ij}, \eta_{\ell j})$ are assumed multivariate normal. The multiplier $\boldsymbol{\beta}_i^*$ is a $q_i \times 1$ vector of unknown constants that captures the relationship between the output X_{ij} and the control $\mathbf{C}_{ij}$, while η_{ij} represents that part of the variability in X_{ij} that is not explained by the controls. As a consequence of these assumptions, $\boldsymbol{\beta}_i^* = \mathrm{Var}\left[\mathbf{C}_{ij}\right]^{-1} \mathrm{Cov}\left[\mathbf{C}_{ij}, X_{ij}\right]$ and $\tau_i^2 = (1 - R_i^2)\sigma_i^2$, where $R_i^2 = \mathrm{Cov}\left[X_{ij}, \mathbf{C}_{ij}\right] \mathrm{Var}\left[\mathbf{C}_{ij}\right]^{-1} \mathrm{Cov}\left[\mathbf{C}_{ij}, X_{ij}\right] / \sigma_i^2$, the square of the multiple correlation coefficient between X_{ij} and $\mathbf{C}_{ij}$ (Lavenberg and Welch 1981). Model (1) can be justified when $(X_{ij}, \mathbf{C}_{ij}^T)$ are themselves averages or standardized averages of some input random variables

(Wilson and Pritsker 1984), but of course it is always an approximation.

Remark 2.1. We could have arrived at the same conclusions by starting with the assumption that $(X_{ij}, \mathbf{C}_{ij}^T)$ is multivariate normal which implies the linear model (1), but we prefer to emphasize Model (1) as the starting point.

A control-variate estimator of μ_i can be much more statistically efficient than the sample mean of the X_{ij}. We review some basic properties of the CV estimator under Model (1) below.

Define the sample mean of the outputs and controls as

$$\bar{X}_i(n) = \frac{1}{n}\sum_{j=1}^{n} X_{ij} \text{ and } \bar{\mathbf{C}}_i(n) = \frac{1}{n}\sum_{j=1}^{n} \mathbf{C}_{ij}.$$

We append "(n)" to quantities defined across n observations.

Let

$$\widehat{\boldsymbol{\beta}}_i(n) = \mathbf{S}_{\mathbf{C}_i}^{-1}(n)\mathbf{S}_{\mathbf{C}_i X_i}(n)$$

where $\mathbf{S}_{\mathbf{C}_i}(n)$ is the sample variance-covariance matrix of $\mathbf{C}_{ij}$ and $\mathbf{S}_{\mathbf{C}_i X_i}(n)$ is the sample covariance vector between $\mathbf{C}_{ij}$ and X_{ij}.

Then the CV point estimator of μ_i is

$$\begin{aligned}
\widehat{\mu}_i(n) &= \bar{X}_i(n) - \left(\bar{\mathbf{C}}_i(n) - \boldsymbol{\xi}_i\right)^T \widehat{\boldsymbol{\beta}}_i(n) \\
&= \frac{1}{n}\sum_{j=1}^{n}\left[X_{ij} - (\mathbf{C}_{ij} - \boldsymbol{\xi}_i)^T\widehat{\boldsymbol{\beta}}_i(n)\right].
\end{aligned}$$

It is known that under Model (1)

$$\mathrm{E}[\widehat{\mu}_i(n)] = \mu_i \quad \text{and} \quad \mathrm{Var}[\widehat{\mu}_i(n)] = \left(\frac{n-2}{n-q_i-2}\right)\frac{\tau_i^2}{n}.$$

The term $(n-2)/(n-q_i-2)$ is known as the loss ratio, and it quantifies the impact of the estimation of $\boldsymbol{\beta}_i^*$.

The standard unbiased estimator of $\mathrm{Var}\left[\widehat{\mu}_i(n)\right]$ is $\widehat{\tau}_i^2(n)\widehat{\Delta}_i^2(n)$, where

$$\widehat{\tau}_i^2(n) = \frac{1}{n-q_i-1}\sum_{j=1}^{n}\left[X_{ij} - \widehat{\mu}_i(n) - (\mathbf{C}_{ij} - \boldsymbol{\xi}_i)^T\widehat{\boldsymbol{\beta}}_i(n)\right]^2$$

is the residual variance estimator and

$$\widehat{\Delta}_i^2(n) = \frac{1}{n} + \frac{1}{n-1}\left(\bar{\mathbf{C}}_i(n) - \boldsymbol{\xi}_i\right)^T \mathbf{S}_{\mathbf{C}_i}^{-1}(n)\left(\bar{\mathbf{C}}_i(n) - \boldsymbol{\xi}_i\right).$$

2.2 The Procedure

Suppose that a larger mean is better, and unknown to us $\mu_k \geq \mu_{k-1} \geq \cdots \geq \mu_1$. We want a procedure that guarantees to select system k with PCS $\geq 1-\alpha$ whenever $\mu_k \geq \mu_{k-1}+\delta$, where $\delta > 0$ is a user-specified parameter representing the smallest difference worth detecting. In this section we present a Generic Procedure that, under certain conditions, provides such guarantees.

For each system $i = 1, 2, \ldots, k$, any non-negative integers a, b, with $b > a+1$, and $q_i \times 1$ vector $\boldsymbol{\beta}_i$, the controlled sum from the $(a+1)$st sample to the bth sample is defined as

$$X_i\left[a, b, \boldsymbol{\beta}_i\right] = \sum_{j=a+1}^{b}\left[X_{ij} - (\mathbf{C}_{ij} - \boldsymbol{\xi}_i)^T\boldsymbol{\beta}_i\right].$$

We use the following *controlled sum of differences* between systems i and ℓ to construct the tracking process in our procedure:

$$\begin{aligned} X_i\left[a, b, \boldsymbol{\beta}_i\right] - X_\ell\left[a, b, \boldsymbol{\beta}_\ell\right] &= \sum_{j=a+1}^{b}\left[X_{ij} - (\mathbf{C}_{ij} - \boldsymbol{\xi}_i)^T\boldsymbol{\beta}_i - X_{\ell j} + (\mathbf{C}_{\ell j} - \boldsymbol{\xi}_\ell)^T\boldsymbol{\beta}_\ell\right] \\ &= \sum_{j=a+1}^{b}\left[X_{ij} - X_{\ell j}\right] - \sum_{j=a+1}^{b}\left[(\mathbf{C}_{ij} - \boldsymbol{\xi}_i)^T\boldsymbol{\beta}_i - (\mathbf{C}_{\ell j} - \boldsymbol{\xi}_\ell)^T\boldsymbol{\beta}_\ell\right], \end{aligned} \quad (2)$$

where the first sum on the right-hand side of the Equation (2) is the raw sum of differences

used in $\mathcal{KN}$, and the second sum is a correction that depends on the centered controls and $\boldsymbol{\beta}_i$ and $\boldsymbol{\beta}_\ell$. The correction term is employed to obtain a variance reduction. Although the continuation region is usually presented in terms of the (controlled) sum of differences $X_i\left[a, b, \boldsymbol{\beta}_i\right] - X_\ell\left[a, b, \boldsymbol{\beta}_\ell\right]$, we will interchangeably work with the sample mean difference $\bar{X}_i[a, b, \boldsymbol{\beta}_i] - \bar{X}_\ell[a, b, \boldsymbol{\beta}_\ell]$, where

$$\bar{X}_i[a, b, \boldsymbol{\beta}_i] = \frac{1}{b-a} X_i[a, b, \boldsymbol{\beta}_i].$$

For all $i \neq \ell$, define the *controlled sample variance*, $S^2_{i\ell}\left[a, b, \boldsymbol{\beta}_i, \boldsymbol{\beta}_\ell\right]$, as

$$\frac{1}{b-a-1} \sum_{j=a+1}^{b} \left\{ X_{ij} - \left(\mathbf{C}_{ij} - \boldsymbol{\xi}_i\right)^T \boldsymbol{\beta}_i - X_{\ell j} + \left(\mathbf{C}_{\ell j} - \boldsymbol{\xi}_\ell\right)^T \boldsymbol{\beta}_\ell - \bar{X}_i[a, b, \boldsymbol{\beta}_i] + \bar{X}_\ell[a, b, \boldsymbol{\beta}_\ell] \right\}^2.$$

Based on the notation above we present the Generic Procedure from which the procedures in Sections 3, 4, and 6 can be derived.

The Generic Procedure

Setup: Select confidence level $1/k < 1 - \alpha < 1$, indifference-zone parameter $\delta > 0$, preliminary-stage sample size $m_0 > q + 2$ (or $m_0 = 0$ when there is no preliminary stage), and first-stage sample size n_0 such that $n_0 - m_0 \geq 2$. Let $\lambda = \delta/2$ and $h^2 = 2\eta \times (n_0 - m_0 - 1)$, where

$$\eta = \frac{1}{2}\left[\left(\frac{2\alpha}{k-1}\right)^{-2/(n_0 - m_0 - 1)} - 1\right].$$

Initialization: Let $I = \{1, 2, \ldots, k\}$ be the set of systems still in contention.

If $m_0 > 0$, then

obtain $(X_{ij}, \mathbf{C}_{ij}), i = 1, 2, \ldots, k, j = 1, 2, \ldots, m_0$ (preliminary stage),

compute estimator $\widehat{\boldsymbol{\beta}}_i(m_0)$ of $\boldsymbol{\beta}_i^*$ and set $\boldsymbol{\beta}_i = \widehat{\boldsymbol{\beta}}_i(m_0), i = 1, 2, \ldots, k$.

Else

set $\boldsymbol{\beta}_i = \boldsymbol{\beta}_i^*$ or 0 or to an arbitrary value as desired, for $i = 1, 2, \ldots, k$.

Endif

Obtain $(X_{ij}, \mathbf{C}_{ij}), i = 1, 2, \ldots, k, j = m_0 + 1, m_0 + 2, \ldots, n_0$ (first stage).

Compute $S^2_{i\ell}\,[m_0, n_0, \boldsymbol{\beta}_i, \boldsymbol{\beta}_\ell]$, for all $i \neq \ell$.

Set the observation counter $r = n_0$ and go to **Screening**.

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

$$I = \left\{ i : i \in I^{\text{old}} \text{ and } \bar{X}_i[m_0, r, \boldsymbol{\beta}_i] \geq \bar{X}_\ell[m_0, r, \boldsymbol{\beta}_\ell] - W_{i\ell}[m_0, n_0, \boldsymbol{\beta}_i, \boldsymbol{\beta}_\ell, r], \forall \ell \in I^{\text{old}}, \ell \neq i \right\},$$

where

$$W_{i\ell}[m_0, n_0, \boldsymbol{\beta}_i, \boldsymbol{\beta}_\ell, r] = \max\left\{0, \frac{h^2 S^2_{i\ell}[m_0, n_0, \boldsymbol{\beta}_i, \boldsymbol{\beta}_\ell]}{2\delta(r - m_0)} - \lambda\right\}.$$

Stopping Rule: If $|I| = 1$, then stop and select the system whose index is in I as the best.

Otherwise, take one additional observation $(X_{i,r+1}, \mathbf{C}_{i,r+1})$ from each system $i \in I$, set $r = r + 1$ and go to **Screening**.

Remark 2.2. There is a family of triangular continuation regions indexed by an integer parameter that is called c by Kim and Nelson (2001). We have only presented the region obtained when $c = 1$, which Kim and Nelson showed to be a good compromise.

In the sections that follow we specialize the Generic Procedure in various ways, prove its validity where possible, and examine design parameters that can affect its performance.

3 A Procedure for Known $\boldsymbol{\beta}^*$

In this section we assume that $\boldsymbol{\beta}^*$ is known to demonstrate the potential benefits of the controlled sum in fully sequential selection procedures. The case of unknown $\boldsymbol{\beta}^*$ is discussed

in Sections 4–6. Notice that the Generic Procedure presented above becomes $\mathcal{KN}$ when $m_0 = 0$ and $\boldsymbol{\beta}_i = 0, \forall i$. Assuming Model (1) holds, when $m_0 = 0$ and we know $\boldsymbol{\beta}^*$ or use arbitrary but fixed $\boldsymbol{\beta}$, then this procedure is $\mathcal{KN}$ applied to a new normal random variable, $X'_{ij} = X_{ij} - (\mathbf{C}_{ij} - \boldsymbol{\xi}_i)^T \boldsymbol{\beta}_i$. Therefore, the test statistic in the **Screening** step is exactly the sum of differences between two normal random variables with means μ_i and μ_ℓ, so it satisfies the requirements for the validity of $\mathcal{KN}$ and the desired PCS guarantee is still provided.

We next show that the procedure for known $\boldsymbol{\beta}^*$ is better than $\mathcal{KN}$. Notice that this procedure will tend to have smaller S^2_{ik} which narrows and shortens the continuation region. We expect the continuation region of the procedure with known $\boldsymbol{\beta}^*$ to be completely contained within the region of $\mathcal{KN}$, therefore the *area* of the continuation region is a good basis for comparison because we have more opportunities to eliminate inferior systems earlier when the procedure has a smaller continuation region, provided the process drift is unchanged, as it is here. To simplify the result for the purpose of illustration, we assume that Model (1) holds and the systems are simulated independently. For $\mathcal{KN}$ we know the area of the continuation region A in Figure 1 is (ignoring rounding)

$$\left(\frac{h^2 S^2_{ik}}{2\delta}\right)\left(\frac{h^2 S^2_{ik}}{\delta^2}\right) = \frac{h^4 S^4_{ik}}{2\delta^3}$$

(Kim and Nelson, 2001). For normally distributed data,

$$\mathrm{E}[S^4_{ik}] = \sigma^4_{ik} \times \frac{(n_0+1)}{(n_0-1)}.$$

Therefore, we obtain

$$\mathrm{E}[A]_{\mathcal{KN}} = \mathrm{E}\left[\frac{h^4 S^4_{ik}}{2\delta^3}\right] = \frac{h^4}{2\delta^3} \times \frac{(n_0+1)}{(n_0-1)} \times \sigma^4_{ik},$$

where $\sigma^2_{ik} = \mathrm{Var}[X_{kj} - X_{ij}] = \sigma^2_k + \sigma^2_i$.

For the procedure with known $\boldsymbol{\beta}^*$,

$$\mathrm{E}[A]_{\boldsymbol{\beta}^*} = \mathrm{E}\left[\frac{h^4 S_{ik}^4[0, n_0, \boldsymbol{\beta}_k^*, \boldsymbol{\beta}_i^*]}{2\delta^3}\right] = \frac{h^4}{2\delta^3} \times \frac{(n_0+1)}{(n_0-1)} \times \tau_{ik}^4,$$

where

$$\begin{aligned}
\tau_{ik}^2 &= \mathrm{Var}\left[X_{kj} - (\mathbf{C}_{kj} - \boldsymbol{\xi}_k)^T \boldsymbol{\beta}_k^* - X_{ij} + (\mathbf{C}_{ij} - \boldsymbol{\xi}_i)^T \boldsymbol{\beta}_i^*\right] \\
&= \mathrm{Var}\left[\eta_{kj} - \eta_{ij}\right] \\
&= \left(1 - R_k^2\right)\sigma_k^2 + \left(1 - R_i^2\right)\sigma_i^2,
\end{aligned}$$

where R_i^2 is the square of the multiple correlation coefficient between X_{ij} and $\mathbf{C}_{ij}$. For simplicity, we assume that $R_i^2 = R^2$, for $i = 1, 2, \ldots, k$. Then,

$$\mathrm{E}[A]_{\boldsymbol{\beta}^*} = \mathrm{E}[A]_{\mathcal{KN}} \times \left(1 - R^2\right)^2.$$

Thus, the expected area of the continuation region for the Generic Procedure with known $\boldsymbol{\beta}^*$ is smaller than that for $\mathcal{KN}$, so we expect to be more efficient by applying controlled sums instead of raw sums. Notice that larger correlation between the outputs and the controls leads to a smaller continuation region, so choosing effective control variates is important. See Añonuevo and Nelson (1988), Nelson (1989) and Bauer and Wilson (1992) for a general discussion of selecting good control variates.

Also notice that an arbitrarily chosen $\boldsymbol{\beta}$ could *increase* the expected area of the continuation region. This is easiest to see if we simplify the analysis even further by assuming that systems i and k are identical, but still independently simulated. Then $\sigma_{ik}^2 = 2\mathrm{Var}[X]$ and $\tau_{ik}^2 = 2\left(\mathrm{Var}[X] + \boldsymbol{\beta}^T\mathrm{Var}[\mathbf{C}]\boldsymbol{\beta} - 2\boldsymbol{\beta}^T\mathrm{Cov}[\mathbf{C}, X]\right)$. Therefore, $\tau_{ik}^2 \geq \sigma_{ik}^2$ if $\boldsymbol{\beta}^T\mathrm{Var}[\mathbf{C}]\boldsymbol{\beta} \geq 2\boldsymbol{\beta}^T\mathrm{Cov}[\mathbf{C}, X]$, which occurs, for instance, if $\boldsymbol{\beta} = 3\boldsymbol{\beta}^*$. This motivates the need for the procedures in Sections 4–6 that estimate $\boldsymbol{\beta}^*$.

4 A Procedure for Unknown $\boldsymbol{\beta}^*$

In most stochastic simulation experiments $\boldsymbol{\beta}^*$ is not known in advance. Therefore, we need to spend some effort estimating $\boldsymbol{\beta}^*$ which is not required for $\mathcal{KN}$. In this section we introduce a fully sequential procedure that allows unknown $\boldsymbol{\beta}^*$ and also guarantees the PCS; some guidelines for the design of this procedure are also discussed in the Appendix.

The Controlled Sequential Selection procedure ($\mathcal{CSS}$) collects preliminary-stage samples $(X_{ij}, \mathbf{C}_{ij}), j = 1, 2, \ldots, m_0$ to compute $\widehat{\boldsymbol{\beta}}_i(m_0)$ for each system, and then obtains an additional $n_0 - m_0$ first-stage samples which are used to compute the controlled sample variance, $S^2_{i\ell}[m_0, n_0, \widehat{\boldsymbol{\beta}}_i(m_0), \widehat{\boldsymbol{\beta}}_\ell(m_0)]$. Figure 2 shows the continuation region for $\mathcal{CSS}$.

Why do we need a preliminary stage, since it would seem to be more efficient to set $m_0 = 0$ and compute both $\widehat{\boldsymbol{\beta}}_i(n_0)$ and $S^2_{i\ell}[0, n_0, \widehat{\boldsymbol{\beta}}_i(n_0), \widehat{\boldsymbol{\beta}}_\ell(n_0)]$ from the first-stage data? The reason is that the sample variance will not have the statistical properties we need to prove the validity of $\mathcal{CSS}$, in particular it is not an unbiased estimator of

$$\frac{1}{r}\ \mathrm{Var}\left[X_i[0, r, \widehat{\boldsymbol{\beta}}_i(n_0)] - X_\ell[0, r, \widehat{\boldsymbol{\beta}}_\ell(n_0)]\right]$$

nor does it have a scaled chi-squared distribution conditional on $\widehat{\boldsymbol{\beta}}_i(n_0)$ and $\widehat{\boldsymbol{\beta}}_\ell(n_0)$.

We already know the procedure is valid for known $\boldsymbol{\beta}^*$ or arbitrary, fixed $\boldsymbol{\beta}$. Therefore, it is natural to expect that it is valid when $\widehat{\boldsymbol{\beta}}_i(m_0)$ is estimated from an independent sample.

Theorem 1. *If Model (1) holds, then the $\mathcal{CSS}$ procedure selects system k with probability $\geq 1 - \alpha$ whenever $\mu_k - \mu_{k-1} \geq \delta$.*

The proof is provided in the Appendix.

Remark 4.1. This procedure is statistically valid with or without the use of CRN. If the systems are simulated independently, then we can replace the value of η in the **Setup** step with

$$\eta = \frac{1}{2}\left[\left[2 - 2(1-\alpha)^{1/(k-1)}\right]^{-2/(n_0 - m_0 - 1)} - 1\right].$$

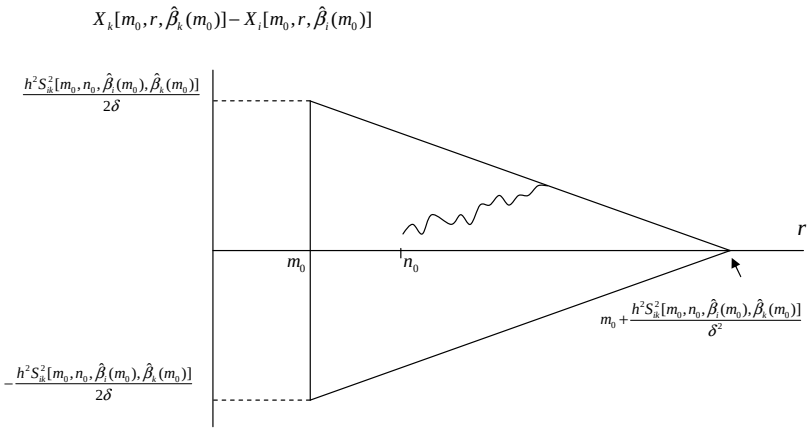

Figure 2: Continuation region for $\mathcal{CSS}$

In this case, the value of η is decreased which makes the continuation region smaller, therefore the procedure may be terminated more quickly. This result follows directly from Theorem 2 of Kim and Nelson (2001).

It is important to notice that the variance of $\bar{X}_i[m_0, r, \widehat{\boldsymbol{\beta}}_i(m_0)]$ is different from the variance of the usual CV estimator $\widehat{\mu}_i(r)$. Under Model (1) it is known that

$$\mathrm{E}\left[\bar{X}_i[m_0, r, \widehat{\boldsymbol{\beta}}_i(m_0)]\right] = \mu_i \quad \text{and} \quad \mathrm{Var}\left[\bar{X}_i[m_0, r, \widehat{\boldsymbol{\beta}}_i(m_0)]\right] = \left(\frac{m_0 - 2}{m_0 - q_i - 2}\right)\frac{\tau_i^2}{r - m_0}$$

(Ripley 1987). Thus, the loss ratio contains the preliminary-stage sample size m_0 instead of the overall sample size r. Clearly it is important that m_0 may not be too small to make the loss insignificant.

The advantage of $\mathcal{CSS}$ is that we can exploit CVs and preserve the required PCS, but the disadvantage is that we need to collect some preliminary samples to estimate $\widehat{\boldsymbol{\beta}}_i(m_0)$ before the screening process is initiated. These preliminary-stage samples are acquired only for the purpose of estimating $\boldsymbol{\beta}^*$ and we can not employ them to eliminate systems. Furthermore, a suitable size for the preliminary-stage samples needs to be determined. In the appendix we provide an analysis that suggests appropriate preliminary-stage sample sizes; for instance,

$16 \leq m_0 \leq 20$ when there are 3 controls. In the appendix we also show that if we follow these guidelines then there is little potential benefit from updating the estimator $\widehat{\boldsymbol{\beta}}$ after the preliminary stage.

5 A Controlled Sequential Selection Procedure Combined with $\mathcal{KN}$

$\mathcal{CSS}$ is a statistically valid procedure that requires taking preliminary-stage samples to calculate $\widehat{\boldsymbol{\beta}}_i(m_0)$ before we enter the screening process; therefore these m_0 samples are wasted. In this section we propose a controlled sequential procedure in which the preliminary-stage samples can be *exploited* while still securing the required PCS. This procedure is basically the combination of $\mathcal{KN}$ and $\mathcal{CSS}$, so we call it $\mathcal{CSS}$-$\mathcal{C}$.

In $\mathcal{CSS}$-$\mathcal{C}$ the m_0 preliminary-stage samples are collected to compute $\widehat{\boldsymbol{\beta}}_i(m_0)$ and the *raw* sample variance is utilized to set up the continuation region for $\mathcal{KN}$. The $\mathcal{KN}$ procedure is then performed from observation m_0 to observation n_0 (first stage); meanwhile, the controlled sample variance $S^2_{ik}[m_0, n_0, \widehat{\boldsymbol{\beta}}_i(m_0), \widehat{\boldsymbol{\beta}}_k(m_0)]$ is obtained and both $\mathcal{KN}$ and $\mathcal{CSS}$ are implemented in parallel after the first stage (observation n_0). A system is eliminated when either $\mathcal{KN}$ or $\mathcal{CSS}$ eliminates it, and the procedure terminates when there is only one system remaining. Figure 3 illustrates the continuation region for $\mathcal{CSS}$-$\mathcal{C}$.

The advantage of $\mathcal{CSS}$-$\mathcal{C}$ is that it gives us an opportunity to eliminate inferior systems in the first stage. The disadvantage is that it uses the conservative Bonferroni inequality to combine $\mathcal{KN}$ and $\mathcal{CSS}$ to guarantee the overall PCS. This procedure is more desirable when there are a large number of systems whose means are expected to differ widely. In that case, we expect the savings gained through eliminating inferior systems in the first stage to more than offset the losses incurred in applying the Bonferroni inequality. We present $\mathcal{CSS}$-$\mathcal{C}$ in detail in the Appendix.

In $\mathcal{CSS}$-$\mathcal{C}$ we apply $\mathcal{KN}$ to screen out noncompetitive systems in the first stage. Then

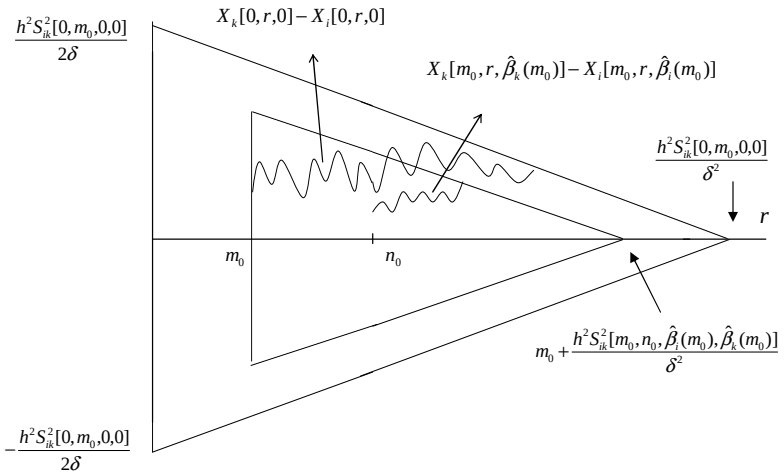

Figure 3: Continuation region for $\mathcal{CSS}$-$\mathcal{C}$

after the first stage $\mathcal{KN}$ and $\mathcal{CSS}$ are applied to the surviving systems to select the best system. We spend α_0 of the overall allowable probability of incorrect selection α on $\mathcal{KN}$, and the other $\alpha - \alpha_0$ on $\mathcal{CSS}$ (typically we take $\alpha_0 = \alpha/2$). The following analysis investigates the effect of splitting α in $\mathcal{CSS}$-$\mathcal{C}$.

Let $n_1 = n_0 - m_0$, be the number of samples between the preliminary and first stages, and let N denote the maximum number of samples the could be required for system i to eliminate system k or vice versa for a fully sequential procedure (see Figure 2). Then

$$\mathrm{E}[N]_{\mathcal{CSS}} = m_0 + \frac{2\eta_1(n_1-1)\left(\frac{m_0-2}{m_0-q-2}\right)\tau_{ik}^2}{\delta^2},$$

where

$$\eta_1 = \frac{1}{2}\left[\left(\frac{2\alpha}{k-1}\right)^{-2/(n_1-1)} - 1\right].$$

For $\mathcal{CSS}$-$\mathcal{C}$ the maximum number of samples that the controlled sum can take until the procedure stops is larger than $\mathcal{CSS}$ because a smaller probability of incorrect selection is

specified. If we take $\alpha_0 = \alpha/2$, then

$$\mathrm{E}[N]_{\mathcal{CSS}\text{-}\mathcal{C}} = m_0 + \frac{2\eta_2(n_1-1)\left(\frac{m_0-2}{m_0-q-2}\right)\tau_{ik}^2}{\delta^2},$$

where

$$\eta_2 = \frac{1}{2}\left[\left(\frac{\alpha}{k-1}\right)^{-2/(n_1-1)} - 1\right].$$

We can show that

$$\frac{\mathrm{E}[N]_{\mathcal{CSS}\text{-}\mathcal{C}}}{\mathrm{E}[N]_{\mathcal{CSS}}} \approx \frac{\eta_2}{\eta_1}.$$

Table 1 gives the value of η_2/η_1 as a function of n_1 for different values of k when $\alpha = 0.05$ and $\alpha_0 = \alpha/2$. This table shows that the ratio η_2/η_1 decreases as the number of systems and the number of samples in the first stage increase, and it also illustrates that once we reach, say, 15 to 30 observations in the first stage, there is not much potential reduction in this ratio from increasing n_1 further within a realistic range of first-stage sample sizes ($n_1 \leq 100$). Therefore, 15 to 30 samples are recommended to take in the first stage to moderate the disadvantage of employing the Bonferroni inequality.

6 A Controlled Sequential Selection Procedure with Approximate Variance Estimator

To avoid wasting samples in the preliminary stage, we derive the Controlled Sequential Selection Procedure with Approximate Variance Estimator ($\mathcal{CSS}$-$\mathcal{A}$). $\mathcal{CSS}$-$\mathcal{A}$ is similar to $\mathcal{CSS}$ except that it only requires a first stage (sample size $n_0 > q+2$) to estimate the $\widehat{\boldsymbol{\beta}}_i(n_0)$ and an approximate variance estimator. For the Generic Procedure, we set $m_0 = 0$ in the **Setup** step, set $\boldsymbol{\beta}_i = \widehat{\boldsymbol{\beta}}_i(n_0)$ and replace $S^2_{i\ell}[0, n_0, \widehat{\boldsymbol{\beta}}_i(n_0), \widehat{\boldsymbol{\beta}}_\ell(n_0)]$ with $n_0\left(\widehat{\Delta}_i^2(n_0)\widehat{\tau}_i^2(n_0) + \widehat{\Delta}_\ell^2(n_0)\widehat{\tau}_\ell^2(n_0)\right)$ in the **Initialization** step.

The internal variance estimator, $r\left(n_0\widehat{\Delta}_i^2(n_0)\widehat{\tau}_i^2(n_0)\right)$, is a biased estimator of $\mathrm{Var}\left[X_i[0, r, \widehat{\boldsymbol{\beta}}_i(n_0)]\right]$

Table 1: η_2/η_1 as a function of $n_1 = n_0 - m_0$ when $1 - \alpha = 0.95$ and $\alpha_0 = \alpha/2$.

	k				
n_1	2	10	50	100	500
2	4.03	4.00	4.00	4.00	4.00
5	1.60	1.46	1.43	1.43	1.42
10	1.42	1.26	1.22	1.21	1.20
15	1.37	1.22	1.18	1.17	1.15
20	1.35	1.20	1.16	1.15	1.13
30	1.33	1.18	1.14	1.13	1.11
40	1.32	1.18	1.13	1.12	1.10
50	1.32	1.17	1.13	1.12	1.09
60	1.32	1.17	1.13	1.11	1.09
70	1.31	1.17	1.12	1.11	1.09
80	1.31	1.16	1.12	1.11	1.09
90	1.31	1.16	1.12	1.11	1.09
100	1.31	1.16	1.12	1.11	1.09

for any observation counter $r > n_0$, leading to the unfavorable consequence that the PCS guarantee may not be attained. However, we expect the bias to be mild when n_0 is not too small, because

$$\begin{aligned} \widehat{\boldsymbol{\beta}}(n_0) &\xrightarrow{\text{w.p.1}} \boldsymbol{\beta}^* \text{ as } n_0 \to \infty \text{ (Avramidis and Wilson 1993), and} \\ n_0\widehat{\Delta}^2(n_0)\widehat{\tau}^2(n_0) &\xrightarrow{p} \tau^2 \text{ as } n_0 \to \infty \text{ (Nelson 1990),} \end{aligned}$$

where $\xrightarrow{\text{w.p.1}}$ denotes convergence with probability 1 and $\xrightarrow{p}$ denotes convergence in probability. Thus, if n_0 is not too small, then $X_{ij} - (\mathbf{C}_{ij} - \boldsymbol{\xi}_i)^T \widehat{\boldsymbol{\beta}}_i(n_0)$ behaves like $X_{ij} - (\mathbf{C}_{ij} - \boldsymbol{\xi}_i)^T \boldsymbol{\beta}_i^*$ and $r\left(n_0\widehat{\Delta}_i^2(n_0)\widehat{\tau}_i^2(n_0)\right)$ is approximately $r\tau_i^2$ which is equal to $\text{Var}\big[X_i[0, r, \boldsymbol{\beta}_i^*]\big]$. Therefore, $\mathcal{CSS}$-$\mathcal{A}$ should perform like a valid procedure with known $(\boldsymbol{\beta}_i^*, \tau_{i\ell}^2)$ if n_0 is not too small, a procedure for which we provide details in the Appendix. It is worth noting that $S^2[0, n_0, \widehat{\boldsymbol{\beta}}(n_0)]$ is also a strongly consistent estimator of τ^2. However, we still choose to use $n_0\widehat{\Delta}^2(n_0)\widehat{\tau}^2(n_0)$ because it is an unbiased estimator of $1/n_0 \text{Var}\left[X[0, n_0, \widehat{\boldsymbol{\beta}}(n_0)]\right]$ and has a scaled chi-squared distribution, which allows us to select an appropriate number of degrees of freedom.

$\mathcal{CSS}$-$\mathcal{A}$ is likely to be conservative so that the required PCS is still maintained even with a biased variance estimator. Empirical evaluation shows that $\mathcal{CSS}$-$\mathcal{A}$ performs well in most configurations.

If we look at the expected value of the largest possible sample size N and assume that CVs explain all effects of CRN, we find that

$$\mathrm{E}[N]_{\mathcal{CSS}-\mathcal{A}} = \mathrm{E}\left[\frac{h^2\left[n_0\widehat{\Delta}_i^2(n_0)\widehat{\tau}_i^2(n_0) + n_0\widehat{\Delta}_k^2(n_0)\widehat{\tau}_k^2(n_0)\right]}{\delta^2}\right] = \frac{2\eta \times (n_0 - 1)\left(\frac{n_0-2}{n_0-q-2}\right)\tau_{ik}^2}{\delta^2}.$$

To facilitate the analysis we assume that $n_0(\mathcal{CSS}\text{-}\mathcal{A}) = m_0(\mathcal{CSS}) = n_0(\mathcal{CSS})/2$; then the following result holds:

$$\mathrm{E}[N]_{\mathcal{CSS}-\mathcal{A}} = \mathrm{E}[N]_{\mathcal{CSS}} - m_0,$$

which shows less sampling is expended for $\mathcal{CSS}$-$\mathcal{A}$ by the amount of the preliminary-stage sample size m_0.

7 Empirical Evaluation

In this section we perform an empirical evaluation to compare the CV procedures presented in this paper to $\mathcal{KN}$ and to each other.

The systems are represented by various configurations of k normal distributions; in all cases, system k was the best (had the largest true mean). Let X_i be a simulation observation from system i, for $i = 1, 2, \ldots, k$. For simplicity, we assume that there is $q = 1$ control variate and the output can be represented as

$$X_i = \mu_i + (C_i - \xi_i)\beta_i + \eta_i,$$

where $\{\eta_i, i = 1, 2, \ldots, k\}$ are $\mathrm{N}(0, \sigma_\eta^2)$ random variables. The $\{C_i, i = 1, 2, \ldots, k\}$ are $\mathrm{N}(\xi_i, \sigma_c^2)$ random variables and independent of $\{\eta_i, i = 1, 2, \ldots, k\}$. The squared correlation

coefficient between X_i and C_i is $R^2_{(x,c)}$.

We evaluated each procedure on different configurations of the systems, examining factors including the number of systems k, the practically significant difference δ, the preliminary-stage sample size m_0, the first-stage sample size $n_1 = n_0 - m_0$, the variance of controls σ_c^2, the variance of residuals σ_η^2, and the configuration of the means μ_i. We also examined the impact of $R^2_{(x,c)}$ on the performance of each procedure. Note that $R^2_{(x,c)} = \beta^2\sigma_c^2/(\beta^2\sigma_c^2 + \sigma_\eta^2)$. The configurations, the experiment design, and the results are described below.

7.1 Configurations and Experiment Design

We used the slippage configuration (SC) of the true means of the systems to investigate a difficult scenario for fully sequential selection procedures. In the SC, the mean of the best system μ_k was set to exactly δ or a multiple of δ, while $\mu_1 = \mu_2 = \cdots = \mu_{k-1} = 0$. Subsequent experiments were performed with $\mu_k = \delta$. These experiments with the slippage configuration showed that CVs can make the fully sequential procedure more efficient even under the most difficult situation. To examine the efficiency of these procedures in eliminating inferior systems, monotone-decreasing means (MDM) were also used. In the MDM configuration, the differences between the means of any two adjacent systems, $\mu_i - \mu_{i-1}$ were set to δ/γ where γ was a constant in the experiment design. For later experiments, the value $\gamma = 1$ was used.

We chose the first-stage sample size to be $n_1 = 20$, for $i = 1, 2, \ldots, k$. The mean of the controls, ξ_i, was set to be 0, and β_i was set to be 1, for $i = 1, 2, \ldots, k$. The number of systems in each experiment varied over $k = 2, 5, 10, 25, 100$. The indifference-zone parameter, δ, was set to $\delta = \sqrt{(\sigma_c^2 + \sigma_\eta^2)/n_1}$, where σ_c^2 is the variance of controls and σ_η^2 is the variance of residuals; therefore, δ is one standard deviation of the first-stage sample mean. For each configuration, 500 macroreplications (complete repetitions) of the entire fully sequential procedure were carried out. In all experiments, the nominal probability of correct selection was set at $1 - \alpha = 0.95$. We took $\alpha_0 = \alpha/2$ in $\mathcal{CSS}$-$\mathcal{C}$. To compare the performance of

the procedures we recorded the estimated probability of correct selection (PCS), the average number of samples per system (ANS), and the percentage of systems that survived after first-stage sampling (PSS) which is only meaningful for $\mathcal{CSS}$-$\mathcal{C}$. Notice that ANS is used to gauge a procedure's overall efficiency, while PSS gauges the effectiveness of the first-stage $\mathcal{KN}$ component in screening out noncompetitive systems for $\mathcal{CSS}$-$\mathcal{C}$.

7.2 Summary of Results

We do not attempt to present comprehensive results from such a large simulation study. Instead, we present details of some typical examples that emphasize the key conclusions.

7.2.1 Effect of m_0

We examined the effect of different values of the preliminary-stage sample size m_0 for $\mathcal{CSS}$ and $\mathcal{CSS}$-$\mathcal{C}$ and compared them to $\mathcal{KN}$ and $\mathcal{CSS}$-$\mathcal{A}$ with the same first-stage sample size $n_1 = 20$. Notice that the derivation in Appendix A.2 shows that the optimal preliminary-stage sample size is $m_0^* = 10$ when $q = 1$, $n_1 = 20$. As shown in Table 2, all the procedures achieve the required PCS with different m_0, since the validity has nothing to do with the value of $\widehat{\boldsymbol{\beta}}_i(m_0)$. However, if m_0 deviates much from m_0^*, in this case less than 6 or greater than 30, then $\mathcal{CSS}$ and $\mathcal{CSS}$-$\mathcal{C}$ will be degraded in terms of ANS; their performance may even be worse than $\mathcal{KN}$. They tend to have lower ANS than $\mathcal{KN}$ when m_0 is in the range of 8 to 20.

7.2.2 Effect of Control Variates

We evaluated the effect of control variates for the three CV procedures and compared them to $\mathcal{KN}$ under the slippage configuration. The results in Appendix A.2 suggest that $R^2_{(x,c)}$ needs to be no less that 0.13 and no more than 0.26 for $\mathcal{CSS}$ to be more efficient than $\mathcal{KN}$ when $n_1 = 20$, $m_0 = 10$, and $q = 1$. As Table 3 shows, $\mathcal{CSS}$ and $\mathcal{CSS}$-$\mathcal{A}$ outperform $\mathcal{KN}$ easily with $R^2_{(x,c)} \geq 0.2$ in this example. However, $\mathcal{CSS}$-$\mathcal{C}$ needs to have a larger

Table 2: Effect of m_0 for $\mathcal{CSS}$ and $\mathcal{CSS}$-$\mathcal{C}$ in the SC with $\mu_k = \delta$, $k = 10$, $n_1 = 20$, $R^2_{(x,c)} = 0.4$, and nominal PCS = 0.95.

m_0	Procedure	PCS	ANS	PSS
	$\mathcal{KN}$	0.96	151	
	$\mathcal{CSS}$-$\mathcal{A}$	0.98	98	
4	$\mathcal{CSS}$	0.96	186	
	$\mathcal{CSS}$-$\mathcal{C}$	0.96	184	0.98
6	$\mathcal{CSS}$	0.96	127	
	$\mathcal{CSS}$-$\mathcal{C}$	0.97	137	0.98
8	$\mathcal{CSS}$	0.97	118	
	$\mathcal{CSS}$-$\mathcal{C}$	0.98	125	0.99
10	$\mathcal{CSS}$	0.96	114	
	$\mathcal{CSS}$-$\mathcal{C}$	0.98	120	0.99
12	$\mathcal{CSS}$	0.97	112	
	$\mathcal{CSS}$-$\mathcal{C}$	0.97	124	0.99
14	$\mathcal{CSS}$	0.97	113	
	$\mathcal{CSS}$-$\mathcal{C}$	0.97	122	0.99
30	$\mathcal{CSS}$	0.97	124	
	$\mathcal{CSS}$-$\mathcal{C}$	0.97	128	0.98
40	$\mathcal{CSS}$	0.96	133	
	$\mathcal{CSS}$-$\mathcal{C}$	0.98	129	0.97
50	$\mathcal{CSS}$	0.97	143	
	$\mathcal{CSS}$-$\mathcal{C}$	0.98	136	0.96

multiple correlation coefficient ($R^2_{(x,c)} \geq 0.4$) to outperform $\mathcal{KN}$, since the SC configuration is especially undesirable for $\mathcal{CSS}$-$\mathcal{C}$. Of course, a larger $R^2_{(x,c)}$ makes these procedures even more efficient because the CVs can explain more variability of the outputs. Notice that PSS of $\mathcal{CSS}$-$\mathcal{C}$ is not affected by $R^2_{(x,c)}$ because the CVs are not involved in the first-stage screening process.

7.2.3 Comparisons Across All Configurations of the Means

The conclusions in this section are based on Table 4. Here we compare the four procedures under the slippage configuration and monotone-decreasing means configuration when $k = 2, 5, 10$, and 100. Notice that the SC and MDM are identical when $k = 2$. The experiments

Table 3: Effect of control variates for CV procedures in comparison with $\mathcal{KN}$ under the SC with $\mu_k = \delta$, $k = 10$, $m_0 = 10$, $n_1 = 20$, and nominal PCS = 0.95.

$R^2_{(x,c)}$	Procedure	PCS	ANS	PSS
	$\mathcal{KN}$	0.96	151	
0.2	$\mathcal{CSS}$	0.97	146	
	$\mathcal{CSS}$-$\mathcal{A}$	0.98	134	
	$\mathcal{CSS}$-$\mathcal{C}$	0.98	160	0.99
0.4	$\mathcal{CSS}$	0.96	114	
	$\mathcal{CSS}$-$\mathcal{A}$	0.98	98	
	$\mathcal{CSS}$-$\mathcal{C}$	0.98	120	0.99
0.6	$\mathcal{CSS}$	0.96	80	
	$\mathcal{CSS}$-$\mathcal{A}$	0.97	67	
	$\mathcal{CSS}$-$\mathcal{C}$	0.98	89	0.99
0.8	$\mathcal{CSS}$	0.96	46	
	$\mathcal{CSS}$-$\mathcal{A}$	0.97	35	
	$\mathcal{CSS}$-$\mathcal{C}$	0.97	52	0.99

showed that all procedures require greater ANS as the number of systems increases in the SC. However, for the MDM configuration these procedures need fewer samples from each system when the number of systems increases since the additional systems are far from the best.

We find that $\mathcal{CSS}$-$\mathcal{A}$ is superior to the other procedures across most of the configurations we examined in terms of ANS. Results in Section 6 suggest that $\text{E[ANS]}_{\mathcal{CSS}-\mathcal{A}} = \text{E[ANS]}_{\mathcal{CSS}} - m_0$, and this is consistent with what we have seen in the experiments. Notice that the estimated PCS of $\mathcal{CSS}$-$\mathcal{A}$ is greater than 0.95 even under the SC with $k = 2$ and $\mu_2 - \mu_1 = \delta$, which is the most difficult case to deliver the desired PCS since the inequalities used to extend to $k > 2$ tend to make all of these procedures conservative as k increases.

The performance of $\mathcal{CSS}$-$\mathcal{C}$ is not better than the other CV procedures under the slippage configuration because of the Bonferroni inequality, and the ratio of $\text{E[ANS]}_{\mathcal{CSS}-\mathcal{C}}$ to $\text{E[ANS]}_{\mathcal{CSS}}$ is very close to the result derived in Section 5. We see that $\mathcal{CSS}$-$\mathcal{C}$ dominates under the MDM configuration with $k = 100$; the PSS value indicates that the procedure is able to eliminate many inferior systems in the first stage ($\mathcal{KN}$), thus reducing the overall ANS

Table 4: Comparisons among all configurations of the means with $m_0 = 10$, $n_1 = 20$, $R^2_{(x,c)} = 0.4$, and nominal PCS = 0.95.

		k=2	k=5		k=10		k=100	
Procedure	Measure	SC	SC	MDM	SC	MDM	SC	MDM
$\mathcal{KN}$	PCS	0.95	0.95	0.98	0.96	0.99	1	1
	ANS	67	127	81	151	72	210	41
$\mathcal{CSS}$	PCS	0.96	0.96	0.99	0.97	0.99	1	1
	ANS	58	94	65	113	56	149	36
$\mathcal{CSS}$-$\mathcal{A}$	PCS	0.96	0.96	0.99	0.98	0.99	1	1
	ANS	45	81	52	98	45	162	27
$\mathcal{CSS}$-$\mathcal{C}$	PCS	0.97	0.98	0.99	0.99	1	1	1
	ANS	65	107	72	120	58	160	21
	PSS	0.96	0.98	0.93	0.99	0.79	0.99	0.16

dramatically. Therefore, $\mathcal{CSS}$-$\mathcal{C}$ could be more or less efficient than the other procedures, depending on how much we gain from screening out systems in the first stage.

8 An Illustrative Example

In this section we provide a queueing example to compare our procedures with NSGS, TNS-I and $\mathcal{KN}$. NSGS is a combined sample-mean-based procedure due to Nelson et al. (2001) that uses a screening procedure with sample means to eliminate uncompetitive systems after the first stage of sampling, and then applies Rinott's IZ selection procedure (Rinott 1978) in the second stage. TNS-I is a two-stage CV combined procedure presented in Tsai, Nelson and Staum (2008). Both NSGS and TNS-I allow unknown and unequal variances, but CRN is not exploited.

Consider the $M/M/s/c$ queue with Poisson arrivals, exponentially distributed service times, s servers, a capacity of c customers, and first-come, first-served queueing discipline. The customers arrive with rate λ. The service rate for an individual server is μ. Each

procedure is performed on ten different configurations of the systems in which $\lambda/s\mu = 4/5$ where the performance measure is the steady-state mean of the waiting time in system. The capacity c is set to 15. The ten configurations along with their true expected waiting times, which can be analytically computed, are given in Table 5. System 1 is obviously the best system.

To mitigate the initial transient bias, we initialize the simulation in steady state. That is, for each replication we sample the initial condition in accordance with that steady-state distribution of the number of customers in the system. An average waiting time for thirty customers is used as the output on each replication. We use the average service time as the control on replication j, which means

$$X_{ij} = \frac{\sum_{m=1}^{30} W_{ijm}}{30} \text{ and } C_{ij} = \frac{\sum_{m=1}^{30} S_{ijm}}{30},$$

where W_{ijm} is the waiting time in system for customer m of replication j from system i and S_{ijm} is the service time for customer m of replication j from system i. The preliminary-stage sample size m_0 for $\mathcal{CSS}$ and $\mathcal{CSS}$-$\mathcal{C}$ is set as 20, and we set the first-stage sample size $n_1 = 10$ for all procedures. We choose the indifference-zone parameter δ to be 0.1 and CRN is not applied.

Table 6 gives the results of the simulation study with 100 complete macroreplications and nominal PCS = 0.95. We also provide the estimated standard error of ANS to illustrate that there is a significant difference.

The observed PCS for all procedures is greater than 0.95 except for $\mathcal{CSS}$-$\mathcal{A}$. Comparing to the experimental results of Section 7 in which $\mathcal{CSS}$-$\mathcal{A}$ works very well with the linear and normal assumption (Model (1)), the bias problem appears to be exacerbated in the queueing example. To be more specific, $\bar{X}[0, n_0, \widehat{\boldsymbol{\beta}}(n_0)]$ and $n_0\widehat{\Delta}^2(n_0)\widehat{\tau}^2(n_0)$ will be biased when linearity fails; therefore, the observed PCS for $\mathcal{CSS}$-$\mathcal{A}$ deviates from the nominal PCS greatly. Figure 4 shows a scatter plot of X_{ij} and C_{ij} for system 1 which illustrates the

Table 5: The ten queueing systems and their expected waiting times in steady state.

System i	λ	s	μ	$E[W]$
1	4	1	5	0.88
2	4	2	5/2	0.98
3	4	3	5/3	1.10
4	4	4	5/4	1.24
5	4	5	1	1.38
6	4	6	5/6	1.52
7	4	7	5/7	1.67
8	4	8	5/8	1.81
9	4	9	5/9	1.96
10	4	10	5/10	2.10

Table 6: Results for NSGS, TNS-I, $\mathcal{KN}$, $\mathcal{CSS}$, $\mathcal{CSS}$-$\mathcal{C}$ and $\mathcal{CSS}$-$\mathcal{A}$ in 100 trials with $\delta = 0.1$, $m_0 = 20$, $n_1 = 10$ and $1 - \alpha = 0.95$.

Procedure	PCS	ANS	$\widehat{\text{se}}$(ANS)	PSS
NSGS	0.98	158	6.1	0.9
TNS-I	0.99	90	5.4	0.4
$\mathcal{KN}$	0.99	89	2.9	
$\mathcal{CSS}$	0.96	37	0.4	
$\mathcal{CSS}$-$\mathcal{C}$	0.96	36	0.4	0.7
$\mathcal{CSS}$-$\mathcal{A}$	0.90	42	2.1	

nonlinear relationship. Notice that the unbiasedness of the controlled sum and the controlled sample variance for $\mathcal{CSS}$ and $\mathcal{CSS}$-$\mathcal{C}$ do *not* depend on linearity. The efficiencies of the three CV procedures are similar and superior to NSGS, TNS-I and $\mathcal{KN}$ in terms of ANS. There is not much difference between the performance of TNS-I and $\mathcal{KN}$. TNS-I can eliminate more systems than $\mathcal{CSS}$-$\mathcal{C}$ in the first stage (PSS=0.4 vs. PSS=0.7); however, $\mathcal{CSS}$-$\mathcal{C}$ has smaller ANS than TNS-I because our CV fully sequential selection procedure is much more efficient than the CV selection-of-the-best procedure in Nelson and Staum (2006).

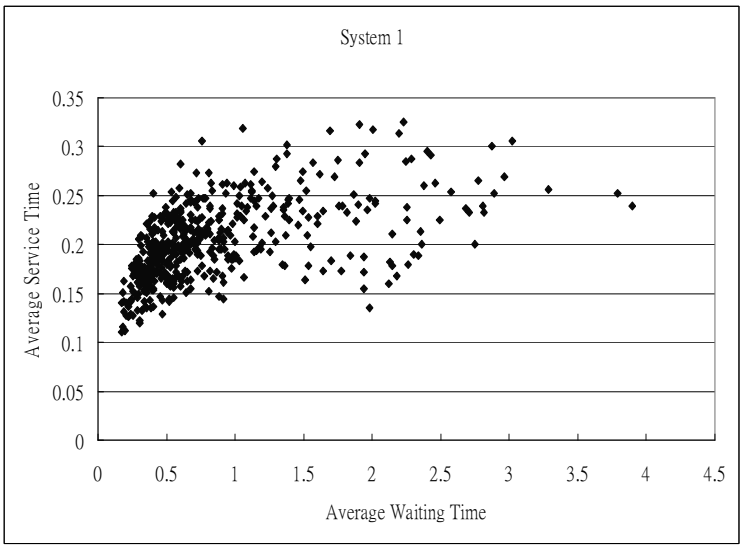

Figure 4: Scatter plot for average waiting time vs. average service time from system 1.

9 Conclusions

In this paper we have proposed a general methodology and several specific procedures for applying control variates in a fully sequential indifference-zone selection procedure. We recommend using $\mathcal{CSS}$-$\mathcal{C}$ only when there are a very large number of widely spaced systems so that the benefits of screening during the first stage are realized. For general use we recommend $\mathcal{CSS}$. We showed that these two procedures reduce the required sample size with respect to $\mathcal{KN}$ while still delivering the PCS guarantee. An approximate procedure called $\mathcal{CSS}$-$\mathcal{A}$ may require fewer observations, and the experiments showed that it performed well when all assumptions are satisfied even though we cannot prove its validity. On the other hand, when the linearity assumption is violated, it seems risky to use $\mathcal{CSS}$-$\mathcal{A}$.

Additional refinement of these procedures may be possible. For instance, the variance-dependent sampling approach in the fully sequential procedure of Hong (2006) could be adapted to the controlled sum.

Acknowledgment

The authors acknowledge the helpful comments and suggestions of L. Jeff Hong, the department editor and referees.

A Appendix

A.1 Proof of $\mathcal{CSS}$

We first prove Theorem 1. For random $\boldsymbol{\beta}$ independent of first-stage and later data,

$$\begin{aligned} \text{PCS} &= \int \text{PCS}(\boldsymbol{\beta})\ \mathrm{d}F_{\boldsymbol{\beta}} \\ &\geq \int (1-\alpha)\ \mathrm{d}F_{\boldsymbol{\beta}} \qquad (3) \\ &= 1-\alpha. \end{aligned}$$

Inequality (3) is true because $\text{PCS}(\boldsymbol{\beta}) \geq 1-\alpha$ for arbitrary but fixed $\boldsymbol{\beta}$ as a direct consequence of Kim and Nelson (2001).

A.2 Choice of m_0 for $\mathcal{CSS}$

In $\mathcal{CSS}$, the experimenter needs to specify the preliminary-stage sample size m_0. The more data we demand in the preliminary stage, the smaller the expected value of the maximum additional sample size is because of the loss ratio, but the more data are wasted before the screening process begins. In this section we try to provide guidelines for selecting m_0 so that $\mathcal{CSS}$ will be efficient with respect to $\mathcal{KN}$. We do this by looking at the area of the continuation region, and also the expected value of the maximum sample size, for $\mathcal{CSS}$ vs. $\mathcal{KN}$, and make rough approximations. The usefulness of these approximations is demonstrated in the empirical results.

To simplify the analysis, we assume that the systems are simulated independently, focus on systems i and k, and assume that $q_i = q$ and $R_i^2 = R^2$, for $i = 1, 2, \ldots, k$. Let n_0' represent the first-stage sample size in $\mathcal{KN}$. Then

$$\mathrm{E}[A]_{\mathcal{KN}} = \frac{h^4}{2\delta^3} \times \frac{(n_0'+1)}{(n_0'-1)} \times \sigma_{ik}^4 = \frac{2\eta'^2(n_0'-1)(n_0'+1)\sigma_{ik}^4}{\delta^3}, \tag{4}$$

where

$$\eta' = \frac{1}{2}\left[\left(\frac{2\alpha}{k-1}\right)^{-2/(n_0'-1)} - 1\right].$$

For convenience, let $n_1 = n_0 - m_0$ be the first-stage sample size for $\mathcal{CSS}$ and let $S_{CV}^2 = S_{ik}^2[m_0, n_0, \widehat{\boldsymbol{\beta}}_i(m_0), \widehat{\boldsymbol{\beta}}_k(m_0)]$. Then

$$\begin{aligned}
\mathrm{E}[A]_{\mathcal{CSS}} &= \mathrm{E}\left[\left(m_0 + \frac{2\eta(n_1-1)S_{CV}^2}{\delta^2}\right) \times \frac{\eta(n_1-1)S_{CV}^2}{\delta}\right] \\
&= \frac{\eta m_0(n_1-1)\mathrm{E}\left[S_{CV}^2\right]}{\delta} + \frac{2\eta^2(n_1-1)^2\mathrm{E}\left[S_{CV}^4\right]}{\delta^3} \\
&\leq \frac{\eta m_0(n_1-1)\mathrm{E}\left[S_{CV}^2\right]}{\delta} \times \frac{2\eta\mathrm{E}\left[S_{CV}^2\right]}{\delta^2} + \frac{2\eta^2(n_1-1)^2\mathrm{E}\left[S_{CV}^4\right]}{\delta^3} \qquad (5) \\
&\quad \text{(when } \delta \text{ is small enough to make } 2\eta\mathrm{E}\left[S_{CV}^2\right]/\delta^2 \geq 1) \\
&= \frac{2\eta^2 m_0(n_1-1)\mathrm{E}^2\left[S_{CV}^2\right]}{\delta^3} + \frac{2\eta^2(n_1-1)^2\mathrm{E}\left[S_{CV}^4\right]}{\delta^3} \\
&\leq \frac{2\eta^2(m_0+n_1-1)(n_1-1)\mathrm{E}\left[S_{CV}^4\right]}{\delta^3} \qquad (6) \\
&\approx \frac{2\eta^2(m_0+n_1-1)(n_1-1)\left(\frac{m_0-2}{m_0-q-2}\right)^2(1-R^2)^2\sigma_{ik}^4}{\delta^3}, \qquad (7)
\end{aligned}$$

where

$$\eta = \frac{1}{2}\left[\left(\frac{2\alpha}{k-1}\right)^{-2/(n_1-1)} - 1\right].$$

We know that there always exists a δ small enough so that Inequality (5) holds, and the choice of m_0 is most critical when δ is small, so we assume this to be the case for the purpose

of analysis. Inequality (6) holds because $\mathrm{E}\left[S_{CV}^4\right] \geq \mathrm{E}^2\left[S_{CV}^2\right]$, and Approximation (7) follows because

$$\mathrm{E}\left[S_{CV}^4\right] = \mathrm{Var}\left[S_{CV}^2\right] + \mathrm{E}\left[S_{CV}^2\right]^2 \approx \mathrm{E}\left[S_{CV}^2\right]^2 = \left(\frac{m_0 - 2}{m_0 - q - 2}\right)^2 (1 - R^2)^2 \sigma_{ik}^4,$$

since $\mathrm{Var}\left[S_{CV}^2\right]$ is decreasing on the order of $\tau_{ik}^4/(m_0)^2$ when n_1 is not too small, whereas $\mathrm{E}\left[S_{CV}^2\right]^2$ is closer to τ_{ik}^4 with increasing m_0.

Our goal is to make the area of the continuation region of $\mathcal{CSS}$ smaller than $\mathcal{KN}$. To put the procedures on an equal footing, we choose first-stage sample sizes for $\mathcal{CSS}$ and $\mathcal{KN}$ such that $n_0' = n_1$; therefore $\eta' = \eta$. We find (7) $\leq$ (4) if

$$(1 - R^2)^2 \leq \frac{(m_0 - q - 2)^2 (n_1 + 1)}{(m_0 - 2)^2 (m_0 + n_1 - 1)}. \tag{8}$$

To maximize the RHS of (8), the optimal m_0 is (ignoring rounding)

$$m_0^* = \frac{3q + 4 + \sqrt{q(9q + 8n_1 + 8)}}{2},$$

which is an increasing function of q and n_1, and increasing faster in q. Consequently, if we use preliminary sample size m_0^*, then $\mathcal{CSS}$ will tend to be more efficient than $\mathcal{KN}$ when

$$R^2 \geq 1 - \sqrt{\frac{(m_0^* - q - 2)^2 (n_1 + 1)}{(m_0^* - 2)^2 (m_0^* + n_1 - 1)}} \equiv R_A^2(n_1, m_0^*, q).$$

Notice that m_0^* is also a function of n_1 and q, and R_A^2 is decreasing in n_1, but increasing in q.

This measure is unfair to $\mathcal{CSS}$ because a larger area of the continuation region is utilized from Inequalities (5) and (6). Therefore, the threshold R_A^2 represents a very conservative lower bound on the required R^2 to make $\mathcal{CSS}$ superior to $\mathcal{KN}$. On the other hand, if we look at the maximum number of observations until the procedure terminates, N, and use

m_0^*, then

$$\begin{aligned}
\mathrm{E}[N]_{\mathcal{CSS}} &= \mathrm{E}\left[m_0^* + \frac{2\eta(n_1-1)S_{CV}^2}{\delta^2}\right] \\
&\geq \mathrm{E}\left[\frac{2\eta(n_1-1)S_{CV}^2}{\delta^2}\right] \qquad (9) \\
&= \frac{2\eta(n_1-1)\mathrm{E}\left[S_{CV}^2\right]}{\delta^2} \\
&= \frac{2\eta(n_1-1)\left(\frac{m_0^*-2}{m_0^*-q-2}\right)(1-R^2)\sigma_{ik}^2}{\delta^2}, \qquad (10)
\end{aligned}$$

and

$$\mathrm{E}[N]_{\mathcal{KN}} = \frac{2\eta'(n_0'-1)\sigma_{ik}^2}{\delta^2}. \qquad (11)$$

Under the same assumptions, we can show that (10) $\leq$ (11) if

$$R^2 \geq 1 - \frac{m_0^*-q-2}{m_0^*-2} \equiv R_B^2(m_0^*, q),$$

which demonstrates that the threshold R_B^2 is just the standard loss ratio. Notice that R_B^2 is not as conservative as R_A^2 since a smaller maximum sample size is used by ignoring m_0^* in Inequality (9). This approximation makes the most sense when there are a number of very close systems so that the procedure may reach the largest observation N.

We expect the lower bound on R^2 such that $\mathcal{CSS}$ is more efficient that $\mathcal{KN}$ to be between R_A^2 and R_B^2, provided we use m_0^* preliminary stage samples. Table 7 lists the value of m_0^*, R_A^2, and R_B^2 as a function of the first-stage sample size $n_1 = n_0 - m_0$ for $q = 3$ controls. Notice that in the typical first-stage sample size range of 10–30, the range of m_0^* is 16–21, which is our recommended setting. Notice also that if the first-stage sample size is 20 or greater, then the controls do not have to be exceptionally effective (large R^2) for $\mathcal{CSS}$ to be more effective than $\mathcal{KN}$, again provided we use m_0^*. Our empirical results support these

Table 7: m_0^*, R_A^2 and R_B^2 as a function of the first-stage sample size n_1 when $q = 3$.

	$q = 3$		
n_1	m_0^*	R_A^2	R_B^2
2	13	0.66	0.27
3	13	0.62	0.27
5	14	0.57	0.25
8	15	0.51	0.23
10	16	0.48	0.21
15	17	0.43	0.20
20	19	0.39	0.18
25	20	0.36	0.17
30	21	0.34	0.16
35	22	0.32	0.15
40	23	0.30	0.14
45	24	0.29	0.14
50	25	0.28	0.13
60	26	0.26	0.13
80	29	0.23	0.11
100	32	0.21	0.10

conclusions.

A.3 Why not update $\widehat{\boldsymbol{\beta}}$ in $\mathcal{CSS}$?

In $\mathcal{CSS}$ we calculate $\widehat{\boldsymbol{\beta}}(m_0)$ from preliminary-stage samples, and then use that $\widehat{\boldsymbol{\beta}}(m_0)$ in the controlled sum until the procedure terminates. Alternatively, we could update $\widehat{\boldsymbol{\beta}}(r)$ every time we acquire an additional observation, or at least update $\widehat{\boldsymbol{\beta}}(r)$ once in a while. In the limit, such a procedure should perform similarly to the procedure with known $\boldsymbol{\beta}^*$ because $\widehat{\boldsymbol{\beta}}(r)$ is getting closer and closer to $\boldsymbol{\beta}^*$. However, we cannot prove the small-sample validity of updating, and even if we could establish its asymptotic validity, the computational effort to implement it would be significant. Moreover, there is not much potential benefit from doing so. The best case for updating $\widehat{\boldsymbol{\beta}}(r)$ is that it attains $\boldsymbol{\beta}^*$ essentially right after the preliminary stage (first update); call this the $\boldsymbol{\beta}^*(m_0)$ procedure. We compare this case with $\mathcal{CSS}$ by looking at the expected value of the largest possible terminal sample size N that

can be reached, namely

$$\mathrm{E}[N]_{\mathcal{CSS}} = \mathrm{E}\left[m_0 + \frac{h^2 S_{ik}^2[m_0, n_0, \widehat{\boldsymbol{\beta}}_i(m_0), \widehat{\boldsymbol{\beta}}_k(m_0)]}{\delta^2}\right] = m_0 + \frac{h^2\left(\frac{m_0-2}{m_0-q-2}\right)\tau_{ik}^2}{\delta^2}$$

$$\mathrm{E}[N]_{\boldsymbol{\beta}^*(m_0)} = \mathrm{E}\left[m_0 + \frac{h^2 S_{ik}^2[m_0, n_0, \boldsymbol{\beta}_i^*, \boldsymbol{\beta}_k^*]}{\delta^2}\right] = m_0 + \frac{h^2\tau_{ik}^2}{\delta^2}.$$

We find the only difference is the usual loss factor. Therefore, we can argue that there is not much advantage to obtain from updating provided m_0 is not too small. For instance, we can choose $m_0 = 20$ when $q = 1$, which implies that the worst case for $\mathcal{CSS}$ is $\mathrm{E}[N]_{\mathcal{CSS}} \approx 1.06$ $\mathrm{E}[N]_{\boldsymbol{\beta}^*(m_0)}$.

However, updating the variance estimator could be helpful since asymptotically the procedure would converge to a known-variance case; see Kim and Nelson (2006) for a version of $\mathcal{KN}$ with variance updating. If we were to try to adapt the same approach to $\mathcal{CSS}$ then both $\widehat{\boldsymbol{\beta}}$ and S_{ik}^2 would need to be updated.

A.4 Proof of $\mathcal{CSS}$-$\mathcal{C}$

We prove the validity of $\mathcal{CSS}$-$\mathcal{C}$. Let ICS denote the event that an incorrect selection is made when the procedure terminates. Then

$$\begin{aligned}
\Pr\{\text{ICS}\} &= \Pr\{\text{ICS with exit through } \mathcal{KN} \text{ or } \mathcal{CSS}\} \\
&\leq \Pr\{\text{ICS with exit through } \mathcal{KN}\} + \Pr\{\text{ICS with exit through } \mathcal{CSS}\} \quad (12) \\
&\leq \alpha_0 + (\alpha - \alpha_0) \quad (13) \\
&= \alpha.
\end{aligned}$$

The Bonferroni inequality justifies Inequality (12), while Inequality (13) follows because of the validity of $\mathcal{KN}$ and Theorem 1.

A.5 $\mathcal{CSS}$-$\mathcal{C}$ Procedure

Based on the notation in Section 2, we describe $\mathcal{CSS}$-$\mathcal{C}$ in detail.

Procedure $\mathcal{CSS}$-$\mathcal{C}$

Setup: Select confidence level $1/k < 1-\alpha < 1$, $\alpha = \alpha_1 + \alpha_2$, indifference-zone parameter $\delta > 0$ and preliminary-stage sample size $m_0 > q+2$. Let $\lambda = \delta/2$ and $h_1^2 = 2\eta\left[\alpha_0, m_0\right] \times (m_0 - 1)$, where

$$\eta\left[\alpha_0, m_0\right] = \frac{1}{2}\left[\left(\frac{2\alpha_0}{k-1}\right)^{-2/(m_0-1)} - 1\right].$$

$\mathcal{KN}$ **Initialization:** Let $I = \{1, 2, \ldots, k\}$ be the set of systems still in contention.

Obtain $(X_{ij}, \mathbf{C}_{ij}), i = 1, 2, \ldots, k, j = 1, 2, \ldots, m_0$ (preliminary stage).

Compute estimator $\widehat{\boldsymbol{\beta}}_i(m_0)$ of $\boldsymbol{\beta}_i^*$ and set $\boldsymbol{\beta}_i = \widehat{\boldsymbol{\beta}}_i(m_0), i = 1, 2, \ldots, k$.

Compute $S_D^2(i, \ell) = S_{i\ell}^2\left[0, m_0, 0, 0\right]$, for all $i \neq \ell$.

Select the first-stage sample size n_0 such that $n_0 - m_0 \geq 2$, set the observation counter $r = m_0$ and go to $\mathcal{KN}$ **Screening**.

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

$$I = \left\{i : i \in I^{\text{old}} \text{ and } \bar{X}_i[0, r, 0] \geq \bar{X}_\ell[0, r, 0] - W_{i\ell}^C\left[h_1^2, S_D^2(i, \ell), r\right], \forall \ell \in I^{\text{old}}, \ell \neq i\right\},$$

where

$$W_{i\ell}^C\left[h_1^2, S_D^2(i, \ell), r\right] = \max\left\{0, \frac{h_1^2 S_D^2(i, \ell)}{2\delta r} - \lambda\right\}.$$

Keep adding data and performing screening until the observation counter $r = n_0$.

$\mathcal{CSS}$ **Initialization:** Let $h_2^2 = 2\eta\,[\alpha_1, n_0 - m_0] \times (n_0 - m_0 - 1)$, and for all $i \neq \ell$ still in I, retrieve the $n_0 - m_0$ observations $X_{ij}, \mathbf{C}_{ij}, j = m_0 + 1, m_0 + 2, \ldots, n_0$ maintained in the first stage to compute $S^2_{CV}(i,\ell) = S^2_{i\ell}[m_0, n_0, \boldsymbol{\beta}_i, \boldsymbol{\beta}_\ell]$.

Go to $\mathcal{KN} + \mathcal{CSS}$ **Screening**.

$\mathcal{KN} + \mathcal{CSS}$ **Screening:** Set $I^{\text{old}} = I$. Let

$$
\begin{aligned}
I^{\mathcal{KN}} &= \left\{ i : i \in I^{\text{old}} \text{ and } \bar{X}_i[0,r,0] \geq \bar{X}_\ell[0,r,0] - W^C_{i\ell}\left[h_1^2, S^2_D(i,\ell), r\right], \forall \ell \in I^{\text{old}}, \ell \neq i \right\}, \\
I &= \left\{ i : i \in I^{\mathcal{KN}} \text{ and } \bar{X}_i[m_0,r,\boldsymbol{\beta}_i] \geq \bar{X}_\ell[m_0,r,\boldsymbol{\beta}_\ell] - W^C_{i\ell}\left[h_2^2, S^2_{CV}(i,\ell), r\right], \forall \ell \in I^{\mathcal{KN}}, \ell \neq i \right\}.
\end{aligned}
$$

Stopping Rule: If $|I| = 1$, then stop and select the system whose index is in I as the best.

Otherwise, take one additional observation $(X_{i,r+1}, \mathbf{C}_{i,r+1})$ from each system $i \in I$, set $r = r + 1$ and go to $\mathcal{KN} + \mathcal{CSS}$ **Screening**.

A.6 Procedure for known ($\boldsymbol{\beta}^*$, τ^2)

We present a procedure with known $\boldsymbol{\beta}^*$ and τ^2 which helps justify the approximate validity of $\mathcal{CSS}$-$\mathcal{A}$.

Procedure for known ($\boldsymbol{\beta}^*$, τ^2)

Setup: Select confidence level $1/k < 1 - \alpha < 1$, indifference-zone parameter $\delta > 0$. Let $\lambda = \delta/2$ and $h^2 = 2\eta$, where

$$\eta = -\ln\left(\frac{2\alpha}{k-1}\right).$$

Initialization: Let $I = \{1, 2, \ldots, k\}$ be the set of systems still in contention.

Obtain $(X_{i1}, \mathbf{C}_{i1}), i = 1, 2, \ldots, k$.

Set the observation counter $r = 1$ and go to **Screening**.

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

$$I = \left\{ i : i \in I^{\text{old}} \text{ and } \bar{X}_i[0, r, \boldsymbol{\beta}_i^*] \geq \bar{X}_\ell[0, r, \boldsymbol{\beta}_\ell^*] - W_{i\ell}(r), \forall \ell \in I^{\text{old}}, \ell \neq i \right\},$$

where

$$W_{i\ell}(r) = \max\left\{0, \frac{h^2 \tau_{i\ell}^2}{2\delta r} - \lambda\right\}.$$

Stopping Rule: If $|I| = 1$, then stop and select the system whose index is in I as the best.

Otherwise, take one additional observation $(X_{i,r+1}, \mathbf{C}_{i,r+1})$ from each system $i \in I$, set $r = r + 1$ and go to **Screening**.

The PCS guarantee is due to Kim and Nelson (2006).

References

Añonuevo, R. and B. L. Nelson. (1988) Automated estimation and variance reduction via control variates for infinite-horizon simulations. *Computers and Operations Research*, **15**, 447–456.

Avramidis, A. N. and J. R. Wilson. (1993) A splitting scheme for control variates. *Operations Research Letters*, **14**, 187–198.

Bauer K. W. and J. R. Wilson. (1992) Control-variate selection criteria. *Naval Research Logistics*, **39**, 307–321.

Bechhofer, R. E., T. J. Santner and D. Goldsman. (1995) *Design and Analysis of Experiments for Statistical Selection, Screening and Multiple Comparisons*, John Wiley & Sons, New York, NY.

Goldsman, D., S.-H. Kim, W. Marshall, and B. L. Nelson. (2001) Ranking and selection procedures for steady-state simulation: Perspectives and procedures. *INFORMS Journals on Computing*, **14**, 2–19.

Hong, L. J. and B. L. Nelson. (2005) The tradeoff between sampling and switching: New sequential procedures for indifference-zone selection. *IIE Transactions*, **37**, 623–634.

Hong, L. J. (2006). Fully sequential indifference-zone selection procedures with variance-dependent sampling. *Naval Research Logistics*, **53**, 464–476.

Hong, L. J. and B. L. Nelson. (2007) Selecting the best system when systems are revealed sequentially. *IIE Transactions*, **39**, 723–734.

Kim, S.-H. and B. L. Nelson. (2001) A fully sequential procedure for indifference-zone selection in simulation. *ACM Transactions on Modeling and Computer Simulation*, **11**, 251–273.

Kim, S.-H. and B. L. Nelson. (2006) On the asymptotic validity of fully sequential selection procedures for steady-state simulation. *Operations Research*, **54**, 475–488.

Lavenberg, S. S. and P. D. Welch. (1981) A perspective on the use of control variables to increase the efficiency of Monte Carlo simulations. *Management Science*, **27**, 322–335.

Nelson, B. L. (1989) Batch size effects on the efficiency of control variates in simulation. *European Journal of Operational Research*, **43**, 184–196.

Nelson, B. L. (1990) Control-variate remedies. *Operations Research*, **38**, 974–992.

Nelson, B. L., J. Swann, D. Goldsman and W. Song. (2001) Simple procedures for selecting the best simulated system when the number of alternatives is large. *Operations Research*, **49**, 950–963.

Nelson, B. L. and J. Staum. (2006) Control variates for screening, selection, and estimation of the best. *ACM Transactions on Modeling and Computer Simulation*, **16**, 52–75.

Pichitlamken, J., B. L. Nelson and L. J. Hong. (2006) A sequential procedure for neighborhood selection-of-the-best in optimization via simulation. *European Journal of Operational Research*, **173**, 283–298.

Rinott, Y. (1978) On two-stage selection procedures and related probability-inequalities. *Communications in Statistics–Theory and Methods*, **A7**, 799–811.

Ripley, B. D. (1987) *Stochastic Simulation*. John Wiley & Sons, New York, NY.

Tsai, S. C., Nelson, B. L. and J. Staum. (2008) Combined screening and selection of the best with control variates. *Advancing the Frontiers of Simulation: A Festschrift in Honor of George S. Fishman*, Kluwer, forthcoming.

Wilson, J. R. and , A. A. B. Pritsker. (1984) Variance reduction in queueing simulation using generalized concomitant variables. *Journal of Statistical Computation and Simulation*, **19**, 129–153.

Biographies

SHING CHIH TSAI is an Assistant Professor of Industrial and Information Management at National Cheng Kung University. He received his B.S. degree in Industrial Management from National Taiwan University, and his M.S. and Ph.D. degrees in Industrial Engineering and Management Sciences at Northwestern University. His research interests are in experiment design and analysis of stochastic simulation experiments.

BARRY L. NELSON is the Charles Deering McCormick Professor of Industrial Engineering and Management Sciences at Northwestern University and is Editor in Chief of *Naval Research Logistics*. His research centers on the design and analysis of computer simulation experiments on models of stochastic systems.