Control Variates for Screening, Selection, and Estimation of the Best

BARRY L. NELSON and JEREMY STAUM
Northwestern University

Ranking and selection procedures (R&S) were developed by statisticians to search for the best among a small collection of populations or treatments, where the "best" treatment is typically the one with the largest or smallest expected (long-run average) response. R&S procedures have been successfully extended to address situations that are encountered in stochastic simulation of alternative system designs, including unequal variances across alternatives, dependence both within the output of each system and across the outputs from alternative systems, and large numbers of alternatives to compare. In nearly all cases the estimator of the expected response is a (perhaps generalized) sample mean of the output of interest. In this paper we derive R&S procedures that employ *control-variate estimators* instead of sample means. Control variates can be much more statistically efficient than sample means, leading to R&S procedures that are correspondingly more efficient. We also consider the related problem of estimating the expected value of the best (as opposed to the selected) system design.

Categories and Subject Descriptors: I.6.8 [**Simulation and Modeling**]: Types of Simulation—*Monte Carlo*

General Terms: Algorithms, Performance, Theory

Additional Key Words and Phrases: Control variates, ranking and selection

1. INTRODUCTION

In the field of stochastic simulation, procedures from the branch of statistics known as ranking and selection (R&S: see for instance Bechhofer et al. [1995]) have proven to be quite useful. A standard application for R&S procedures is identifying the best among alternative system designs. There are published applications of R&S in simulation (e.g., Stallard and Owen [1998]), and commercial simulation software has incorporated R&S procedures (e.g., Automod's AutoStat, AweSim's Scenario Selector and Arena's Process Analyzer). Not surprisingly, there has been a correspondingly high interest within the simulation research community on extending R&S procedures, which were originally designed for physical experiments, to address situations that are encountered in stochastic simulation of alternative system designs. These situations include unequal variances across alternatives, dependence both within the output of each system (in steady-state simulation problems) and across the outputs from alternative systems (induced by the use of common random numbers, discussed below), and large numbers of alternatives to compare. See, for instance, Swisher et al. [2003].

In this paper we consider searching among a finite number of alternatives for the simulated system with the largest or smallest expected value of some designated output performance measure. For instance, we

may be interested in which inventory policy has the smallest expected cost, which scenario produces the largest conditional expected loss for a portfolio, or which system design yields the largest expected time to system failure. Nearly all R&S procedures for such problems employ sample means, or generalized sample means, as estimators of the expected response. Here we derive new R&S procedures that employ *control-variate estimators* instead of sample means. Control variates (CVs) can be much more statistically efficient than sample means, leading to R&S procedures that are correspondingly more efficient. We also consider the related problem of estimating the expected value of the best system, which arises when Nature selects the best (or worst) system for us, as opposed to the more frequently considered problem of estimating the expected value of the system we select because we think it is best.

This paper contains two contributions: First, we provide new R&S procedures that are no more difficult to employ than CV estimators themselves; as a side benefit the procedures for screening and selection also exploit the variance-reduction technique of common random numbers. Of at least equal importance, however, is that we take the first steps toward developing a theory of R&S based on CV estimators that is a companion to the existing theory based on sample means. Specifically, we provide tools for proving the small-sample validity of such procedures under normal-theory assumptions and indicate when the procedures can be expected to be robust to nonnormality in large samples. Since both R&S and CVs have had a long and successful history in simulation research and applications, it seems well worthwhile to develop a framework for reaping the benefits of both.

The paper is organized as follows: In Section 2, we review CV estimators and their basic properties. Sections 3–5 present new procedures for screening (finding a subset of systems that contains the best), selection of the best, and estimating the best, respectively. Section 6 provides a numerical illustration of the procedures. The validity of the procedures is proven in the Appendix.

2. CONTROL-VARIATE ESTIMATORS

This section provides definitions, notation and results for CVs that will be used throughout the remainder of the paper. We use the following model repeatedly. In it, X_{ij} is the jth simulated observation of the output of system i. The $q_i \times 1$ vector $\mathbf{C}_{ij}$ is called the *control*, and it has known expected value $\boldsymbol{\xi}_i$.

MODEL 1. *For each system* $i = 1, 2, \ldots, k$,

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

where $\boldsymbol{\beta}_i$ *is a* $q_i \times 1$ *vector of unknown constants and* $\{\eta_{ij}, j = 1, 2, \ldots, n\}$ *is a set of independent and identically distributed (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. and independent of* $\{\eta_{ij}, j = 1, 2, \ldots, n\}$.

The multiplier $\boldsymbol{\beta}_i$ 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. In practice, the controls are often averages or standardized averages of the input random variables that drive the simulation (see, for instance, Wilson and Pritsker [1984a; 1984b]). Model 1 is an approximation that makes the most sense when $\{(X_{ij}, \mathbf{C}'_{ij}), j = 1, 2, \ldots, n\}$ are obtained from n independent observations and are themselves averages of some more basic random variables.

CVs are designed to provide a more precise estimator of μ_i than the sample mean of $\{X_{ij}, j = 1, 2, \ldots, n\}$. We define the CV estimator and review basic properties under Model 1 below. The development is based on Nelson [1990] and Nelson and Hsu [1993].

For each system $i = 1, 2, \ldots, k$, let

$$\mathbf{X}_i(n) = \begin{pmatrix} X_{i1} \\ X_{i2} \\ \vdots \\ X_{in} \end{pmatrix} \text{ and } \mathbf{C}_i(n) = \begin{pmatrix} \mathbf{C}'_{i1} \\ \mathbf{C}'_{i2} \\ \vdots \\ \mathbf{C}'_{in} \end{pmatrix}$$

be vectors of the output and controls across all n observations from system i. Define the sample mean of the outputs and controls as

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

The standard or "crude" estimator of μ_i that CVs try to beat is $\bar{X}_i(n)$ which is unbiased and has variance σ_i^2/n, where $\sigma_i^2 = \mathrm{Var}[X_{ij}]$. We append "$(n)$" to quantities defined across n observations, a convention that will be important when we derive CV R&S procedures in Sections 3–5 below.

To define the CV point estimator, for each system $i = 1, 2, \ldots, k$, let

$$\mathbf{L}_i'(n) = \left[(\mathbf{C}_{i1} - \bar{\mathbf{C}}_i(n)), (\mathbf{C}_{i2} - \bar{\mathbf{C}}_i(n)), \ldots, (\mathbf{C}_{in} - \bar{\mathbf{C}}_i(n))\right].$$

If $\mathbf{1}_{n\times 1}$ is a column vector whose n elements all equal one, then the CV point estimator of μ_i is

$$\widehat{\mu}_i(n) = \left[\frac{1}{n}\mathbf{1}_{n\times 1}' - \left(\bar{\mathbf{C}}_i(n) - \boldsymbol{\xi}_i\right)' \left(\mathbf{L}_i'(n)\mathbf{L}_i(n)\right)^{-1} \mathbf{L}_i'(n)\right] \mathbf{X}_i(n) \tag{2}$$

$$= \bar{X}_i(n) - \left(\bar{\mathbf{C}}_i(n) - \boldsymbol{\xi}_i\right)' \widehat{\boldsymbol{\beta}}_i(n) \tag{3}$$

[Nelson 1990]. Equation (3) is the more traditional presentation because it shows that $\widehat{\mu}_i(n)$ is the intercept estimator in a least-squares regression of X_{ij} on $\mathbf{C}_{ij} - \boldsymbol{\xi}_i$. Likewise, $\widehat{\boldsymbol{\beta}}_i(n)$ is the usual slope estimator in this regression. Equation (2) is useful for seeing that the CV estimator is a weighted average of the outputs X_{ij} where the (random) weights depend only on the controls.

Assuming $(X_{ij}, \mathbf{C}_{ij}')$ are i.i.d. multivariate normal, which is sufficient for Model 1 to hold, Lavenberg and Welch [1981] established that

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

where $\tau_i^2 = (1 - R_i^2)\sigma_i^2$ and R_i^2 is the square of the multiple correlation coefficient between X_{ij} and $\mathbf{C}_{ij}$. Nelson [1990] showed that these results are asymptotically valid under much milder conditions. Thus, strong correlations lead to a variance reduction.

To derive R&S procedures we need to know the joint distribution of $\widehat{\mu}_i(n)$, $i = 1, 2, \ldots, k$, and we need an estimator for the variances. For each system $i = 1, 2, \ldots, k$, let

$$\mathbf{A}_i(n) = \begin{pmatrix} 1 & (\mathbf{C}_{i1} - \boldsymbol{\xi}_i)' \\ 1 & (\mathbf{C}_{i2} - \boldsymbol{\xi}_i)' \\ \vdots & \vdots \\ 1 & (\mathbf{C}_{in} - \boldsymbol{\xi}_i)' \end{pmatrix}.$$

If $\mathbf{I}$ is the identity matrix of rank q_i, then

$$\widehat{\tau}_i^2(n) = \frac{1}{n-q_i-1}\mathbf{X}_i(n)' \left[\mathbf{I} - \mathbf{A}_i(n)\left(\mathbf{A}_i'(n)\mathbf{A}_i(n)\right)^{-1}\mathbf{A}_i'(n)\right]\mathbf{X}_i(n)$$

$$= \frac{1}{n-q_i-1}\sum_{j=1}^{n}\left[X_{ij} - \widehat{\mu}_i(n) - (\mathbf{C}_{ij} - \boldsymbol{\xi}_i)'\widehat{\boldsymbol{\beta}}_i(n)\right]^2 \tag{4}$$

is the residual variance estimator. Further, let

$$\widehat{\Delta}_i^2(n) = \frac{1}{n} + \frac{1}{n-1}\left(\bar{\mathbf{C}}_i(n) - \boldsymbol{\xi}_i\right)' \mathbf{S}_{\mathbf{C}_i}^{-1}(n)\left(\bar{\mathbf{C}}_i(n) - \boldsymbol{\xi}_i\right) \tag{5}$$

where $\mathbf{S}_{\mathbf{C}_i}(n)$ is the sample variance-covariance matrix of $\mathbf{C}_{ij}$. Then we have the following key result:

LEMMA 2.1 [NELSON AND HSU 1993, THEOREM 4.1]. *If Model 1 pertains, then conditional on* $\mathbf{C}_1(n)$, $\mathbf{C}_2(n)$, ..., $\mathbf{C}_k(n)$ *the following properties hold:*

P1. $\widehat{\mu}_i(n) \sim \mathrm{N}(\mu_i, \widehat{\Delta}_i^2(n)\tau_i^2)$, $i = 1, 2, \ldots, k$.

P2. $\widehat{\tau}_i^2(n) \sim \dfrac{\tau_i^2 \chi^2_{n-q_i-1}}{n-q_i-1}$ *and is independent of* $\widehat{\mu}_i(n)$, *for* $i = 1, 2, \ldots, k$.

P3. If $\{\eta_{ij}, i = 1, 2, \ldots, k, j = 1, 2, \ldots, n\}$ *are mutually independent, then* $\{\widehat{\mu}_i(n), \widehat{\tau}_i^2(n), i = 1, 2, \ldots, k\}$ *are mutually independent.*

The additional condition specified in property P3, that the η_{ij} are independent for all systems i as well as for all observations j, will hold either if all k systems are simulated independently, or if common random numbers (CRN) are used but the dependence among the controls $\mathbf{C}_{ij}$ and $\mathbf{C}_{\ell j}$ entirely explains the dependence between the responses X_{ij} and $X_{\ell j}$ induced by CRN. To use CRN means to drive the simulation of each alternative system with the same pseudorandom numbers, and the goal is to induce a positive correlation between different systems' outputs, thereby reducing the variance of the estimated differences. See Law and Kelton [2000] for a general discussion of CRN, and Nelson and Hsu [1993] for CRN and CVs. Although it may seem implausible that the controls capture all of the dependence due to CRN, Yang and Nelson [1991] and Nelson and Hsu [1993] showed that multiple-comparisons inferences derived under this assumption still provide the desired coverage and take advantage of at least some of the potential benefit of employing CRN. Therefore, we will also make use of property P3 to formulate R&S procedures using CRN. For an approach to modeling any remaining dependence among the residuals $(\eta_{1j}, \eta_{2j}, \ldots, \eta_{kj})$, see Nelson [1993].

3. SCREENING

We will assume from here on that $\mu_k \geq \mu_{k-1} \geq \cdots \geq \mu_1$ and that a larger mean is better. Thus, unknown to the investigator, system k is the best. In screening our goal is to produce a subset $I \subseteq \{1, 2, \ldots, k\}$ such that the event $\{k \in I\}$ occurs with prespecified confidence. Further, we would like the subset to be as small as

possible. The purpose of using CVs is to try to reduce the size of the subset relative to a screening procedure based on sample means.

We next present a procedure similar to the "screen-to-the-best" procedure of Boesel et al. [2003], but using CV estimators instead of sample means. Subsequently, we present a simpler CV screening procedure based on a more complicated model, which avoids the assumption that the controls explain all of the dependence due to CRN.

3.1 Individual Controls

Within this subsection, we assume that Model 1 holds with independence among $\{\eta_{ij}, i = 1, 2, \ldots, k, j = 1, 2, \ldots, n\}$. Let $t_{p,\nu}$ represent the p quantile of the Student t-distribution with ν degrees of freedom.

Procedure 1 *Screening with Individual Controls.*

1. Choose the confidence level $1 - \alpha > 1/2$.
2. For $i = 1, 2, \ldots, k$, obtain $n_i > q_i + 2$ observations from system i and form the CV estimator $\widehat{\mu}_i(n_i)$.
3. Let $t_i = t_{(1-\alpha)^{1/(k-1)}, n_i - q_i - 1}$ and create the subset

$$I = \{i : \widehat{\mu}_i(n_i) - \widehat{\mu}_\ell(n_\ell) \geq -W_{i\ell}, \forall \ell \neq i\}, \tag{6}$$

where

$$W_{i\ell} = \sqrt{t_i^2 \, \widehat{\Delta}_i^2(n_i)\widehat{\tau}_i^2(n_i) + t_\ell^2 \, \widehat{\Delta}_\ell^2(n_\ell)\widehat{\tau}_\ell^2(n_\ell)}.$$

Proposition 1. *If Model 1 holds with independence among* $\{\eta_{ij}, i = 1, 2, \ldots, k, j = 1, 2, \ldots, n\}$, *then Procedure 1 produces a subset* I *such that* $\Pr\{k \in I\} \geq 1 - \alpha$.

As a point of comparison, consider the "screen-to-the-best" procedure of Boesel et al. [2003], which sets

$$I = \left\{ i : \bar{X}_i(n_i) - \bar{X}_\ell(n_\ell) \geq -\sqrt{\tilde{t}_i^2 \, \frac{S_i^2(n_i)}{n_i} + \tilde{t}_\ell^2 \, \frac{S_\ell^2(n_\ell)}{n_\ell}}, \forall \ell \neq i \right\}, \tag{7}$$

where $S_i^2(n_i)$ is the sample variance of $\{X_{i1}, X_{i2}, \ldots, X_{in_i}\}$ and $\tilde{t}_i = t_{(1-\alpha)^{1/(k-1)}, n_i - 1}$. Typically, the smaller the screening threshold, the smaller the subset. So, as a rough comparison, we investigate conditions under which we can conclude that

$$\mathrm{E}\left[t_i^2 \widehat{\Delta}_i^2(n_i)\widehat{\tau}_i^2(n_i) + t_\ell^2 \widehat{\Delta}_\ell^2(n_\ell)\widehat{\tau}_\ell^2(n_\ell)\right] \leq \mathrm{E}\left[\tilde{t}_i^2 \, \frac{S_i^2(n_i)}{n_i} + \tilde{t}_\ell^2 \, \frac{S_\ell^2(n_\ell)}{n_\ell}\right],$$

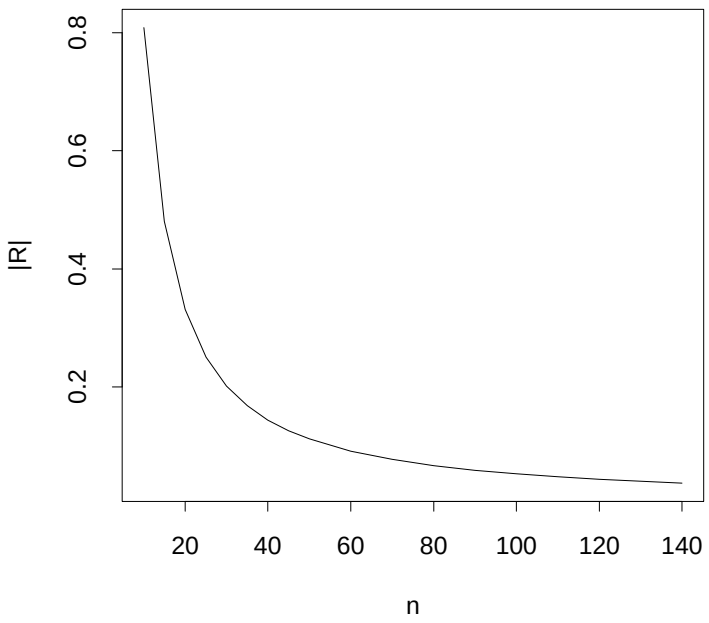

Fig. 1. Minimum multiple correlation required for improved screening with CVs according to Procedure 1 when there are $k = 10$ systems, $q = 5$ controls, and confidence level $1 - \alpha = 0.95$.

that is, that the screening threshold is expected to be smaller with CVs than without them. To simplify the result for the purpose of intuition, suppose $\sigma_i^2 = \sigma^2$, $n_i = n$, $q_i = q$ and $R_i^2 = R^2$ for all $i = 1, 2, \ldots, k$. Consequently, there exist t and $\tilde{t}$ such that $t_i = t$ and $\tilde{t}_i = \tilde{t}$ for all $i = 1, 2, \ldots, k$. Then, assuming that the response and the controls are jointly multivariate normal, the results of Lavenberg and Welch [1981] can be used to evaluate these expectations, yielding

$$2\,t^2 \left(\frac{n-2}{n-q-2} \right) (1 - R^2) \frac{\sigma^2}{n} \leq 2\,\tilde{t}^2 \frac{\sigma^2}{n}.$$

Thus, screening with CVs according to Procedure 1 will tend to be sharper when

$$R^2 \geq 1 - \frac{t^2_{(1-\alpha)^{1/(k-1)}, n-1}}{t^2_{(1-\alpha)^{1/(k-1)}, n-q-1}} \left(\frac{n-q-2}{n-2} \right). \tag{8}$$

This suggests that, if n is not too small, very little correlation between the output and control is required for improved screening. For instance, Figure 1 shows a plot of the bound on $|R|$ implied by Inequality (8) as a function of n for $k = 10$ systems, $q = 5$ controls, and confidence level $1 - \alpha = 0.95$.

3.2 Paired Controls

To avoid the assumption that the dependence due to CRN is entirely explained by the controls (i.e., $\{\eta_{ij}, i = 1, 2, \ldots, k, j = 1, 2, \ldots, n\}$ are mutually independent), we start by forming pairwise differences across systems, $X_j(i,\ell) = X_{ij} - X_{\ell j}, \mathbf{C}_j(i,\ell) = \mathbf{C}_{ij} - \mathbf{C}_{\ell j}, \mu_{i\ell} = \mu_i - \mu_\ell$ and $\boldsymbol{\xi}_{i\ell} = \boldsymbol{\xi}_i - \boldsymbol{\xi}_\ell$. We can do this only when the observations and the controls can each be paired across systems, so let n be the common number of replications and q be the common number of controls for each system. We then assume that an equation similar to Equation (1) holds for the differences.

MODEL 2. *For each system* $i = 1, 2, \ldots, k$,

$$X_j(i,\ell) = \mu_{i\ell} + (\mathbf{C}_j(i,\ell) - \boldsymbol{\xi}_{i\ell})'\boldsymbol{\beta}(i,\ell) + \eta_j(i,\ell), \tag{9}$$

and the controls $\{\mathbf{C}_{ij}, j = 1, 2, \ldots, n\}$ *and residuals* $\{\eta_{ij}, j = 1, 2, \ldots, n\}$ *are both i.i.d. sets and independent of each other.*

Unlike Model 1, Model 2 can hold even if some residuals η_{ij} and $\eta_{\ell j}$ are dependent. This dependence can arise because not all of the simulation input random variables affected by CRN are used as controls, for instance. In short, in this subsection we switch from using controls to explain system performance to using differences between controls to explain differences between systems' performances, which allows us to replace the assumption that residuals are independent across all systems with the assumption that for each pair of systems the residual and the difference between controls are independent.

For all $i \neq \ell$, let $\widehat{\mu}_{i\ell}(n)$ be the corresponding CV estimator of $\mu_{i\ell}$, and define $\widehat{\tau}^2_{i\ell}(n)$ and $\widehat{\Delta}^2_{i\ell}(n)$ in analogy to Equations (4) and (5), but applying CVs to differences between systems' output instead of to each system's output. In general, $\widehat{\mu}_{i\ell}(n) \neq \widehat{\mu}_i(n) - \widehat{\mu}_\ell(n)$, but $\widehat{\mu}_{i\ell}(n) = -\widehat{\mu}_{\ell i}(n)$.

Procedure 2 *Screening with Paired Controls.*

1. Choose the confidence level $1 - \alpha > 1/2$.
2. Obtain $n > q + 2$ observations from each system and form the $k(k-1)/2$ CV estimators $\widehat{\mu}_{i\ell}(n)$ for all $i \neq \ell$.

3. Let $t = t_{1-\alpha/(k-1),n-q-1}$ and create the subset

$$I = \left\{ i : \widehat{\mu}_{i\ell}(n) \geq -t\, \widehat{\Delta}_{i\ell}(n)\widehat{\tau}_{i\ell}(n), \forall \ell \neq i \right\}. \tag{10}$$

PROPOSITION 2. *If Model 2 holds, then Procedure 2 produces a subset I such that* $\Pr\{k \in I\} \geq 1-\alpha$.

The advantage of this alternative approach is that it does not assume that all of the dependence due to CRN is explained by the controls. Its disadvantages are that it requires computing $k(k-1)/2$ CV estimators rather than only k of them, and that it employs the very conservative Bonferroni inequality rather than the tighter Banjeree inequality on which the first screening procedure is based (see the Appendix).

Screening according to Procedure 2 will be effective when the *difference* between controls from systems i and ℓ, $\mathbf{C}_{ij}$ and $\mathbf{C}_{\ell j}$, is strongly correlated with the difference between outputs of the systems, X_{ij} and $X_{\ell j}$. Strong correlation between the controls $\mathbf{C}_{ij}$ and $\mathbf{C}_{\ell j}$ themselves is not essential. In fact, if CRN causes $\mathbf{C}_{ij} = \mathbf{C}_{\ell j}$ for all j then the controls simply cancel in Equation (9), and thus produce no benefit. In this case, which indeed occurs in the inventory planning example of Section 6, Procedure 2 should be modified not to use CVs, but simply to work with the differences $X_j(i,\ell) = X_{ij} - X_{\ell j}$. The relative merits of Procedures 1 and 2 depend on how much correlation CRN induces between systems, and how much of this dependence is captured by the controls.

4. SELECTING THE BEST

We now turn our attention to selecting the best system under Model 1 when the assumption of independence among the residuals $\{\eta_{ij}, i = 1,2,\ldots,k, j = 1,2,\ldots,n\}$ applies. We adopt the indifference-zone formulation in which we require a guaranteed probability of selecting system k whenever the difference $\mu_k - \mu_{k-1} \geq \delta$, where the indifference-zone parameter $\delta > 0$ is set to the smallest difference the analyst feels is worth detecting. As we will see in Inequality (13), the procedure also guarantees that, with high probability, the true mean of the system we select is within δ of the best system's mean, regardless of the configuration. Our procedure parallels Rinott's [1978] procedure, which is based on sample means. For simplicity we assume that all systems have the same number of controls, q, and mention the required adjustment for unequal numbers of controls in Remark 4.2.

The following procedure is based on decomposing the allowable error α into a part α_0 that is spent on controlling the dispersion of the controls and a part α_1 that is spent on controlling the dispersion of the CV estimators, given that the controls are not too far from their means.

Procedure 3 *Selecting the Best with Controls.*

1. Choose the indifference-zone parameter $\delta > 0$, confidence level $1 - \alpha > 1/k$ and choose $\alpha_0, \alpha_1 > 0$ such that $\alpha = \alpha_0 + \alpha_1$.

2. For each system $i = 1, 2, \ldots, k$, obtain $n_0 > q + 2$ observations and calculate $\widehat{\tau}_i^2(n_0)$.

3. For each system $i = 1, 2, \ldots, k$, set the total sample size

$$N_i = \min_{n \geq n_0} \left\{ n : \left(\frac{n-q}{q} \right) \left(\frac{n\delta^2}{h^2 \widehat{\tau}_i^2(n_0)} - 1 \right) \geq \mathcal{F}_{\gamma,q,n-q} \right\} \tag{11}$$

   where $h = h_{k,1-\alpha_1,n_0-q-1}$ is Rinott's [1978] constant, $\mathcal{F}_{\gamma,q,n-q}$ is the γ quantile of the F distribution with $(q, n-q)$ degrees of freedom, and

$$\gamma = \begin{cases} (1-\alpha_0)^{\frac{1}{k}}, & \text{if the systems are simulated independently} \\ 1 - \alpha_0/k, & \text{otherwise.} \end{cases}$$

4. Collect $N_i - n_0$ observations from system i and form the CV estimators $\widehat{\mu}_i(N_i)$ for $i = 1, 2, \ldots, k$.

5. Select system $B = \text{argmax}_i \, \widehat{\mu}_i(N_i)$, and form the $(1-\alpha)100\%$ simultaneous confidence intervals

$$\mu_i - \max_{\ell \neq i} \mu_\ell \in \left[\left(\widehat{\mu}_i(N_i) - \max_{\ell \neq i} \widehat{\mu}_\ell(N_\ell) - \delta \right)^-, \left(\widehat{\mu}_i(N_i) - \max_{\ell \neq i} \widehat{\mu}_\ell(N_\ell) + \delta \right)^+ \right] \tag{12}$$

   for $i = 1, 2, \ldots, k$.

PROPOSITION 3. *If Model 1 holds with independence among* $\{\eta_{ij}, i = 1, 2, \ldots, k, j = 1, 2, \ldots, n\}$ *and the distribution of each control* $\mathbf{C}_{ij}$ *is multivariate normal, then Procedure 3 produces a selected system* B *and* $(1-\alpha)100\%$ *simultaneous confidence intervals in Equation (12) such that*

1. *If the difference between the means of the best and second-best systems is at least* δ*, then the selection is correct with high confidence:* $\Pr\{B = k\} \geq 1 - \alpha$ *if* $\mu_k - \mu_{k-1} \geq \delta$.

2. *With high confidence, the mean of the selected system is within* δ *of the mean of the truly best system:*

$$\Pr\left\{ \mu_B - \max_{\ell=1,\ldots,k} \mu_\ell \geq -\delta \right\} \geq 1 - \alpha. \tag{13}$$

3. The confidence intervals have at least nominal coverage:

$$\Pr\left\{\mu_i - \max_{\ell\neq i}\mu_\ell \in \left[\left(\widehat{\mu}_i(N_i) - \max_{\ell\neq i}\widehat{\mu}_\ell(N_\ell) - \delta\right)^-, \left(\widehat{\mu}_i(N_i) - \max_{\ell\neq i}\widehat{\mu}_\ell(N_\ell) + \delta\right)^+\right], \forall i\right\} \geq 1-\alpha.$$

Remark 4.1. Readers familiar with indifference-zone ranking procedures might have anticipated a formula for N_i, the total sample size from system i, of the form

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

We prove the correctness of Equation (11) in the Appendix, but we provide some intuition here. The need for the complication in Equation (11) arises because the variance of the final CV estimator $\widehat{\mu}_i(N_i)$, conditional on the controls, is $\widehat{\Delta}_i^2(N_i)\tau_i^2$, not τ_i^2/N_i, and $\widehat{\Delta}_i^2(N_i)$ depends on values of the controls not yet observed after the first stage of sampling. To account for this, we spend α_0 of the allowable error α on bounding $\widehat{\Delta}_i^2(N_i)$. We can do this because, as shown in the Appendix, $\widehat{\Delta}_i^2(N_i)$ has a shifted and scaled F distribution. In the Appendix we also show that an approximation for N_i is

$$\max\left\{n_0, \left\lceil \frac{h^2\widehat{\tau}_i^2(n_0)}{\delta^2} + \chi^2_{\gamma,q}\right\rceil\right\}, \tag{14}$$

where $\chi^2_{\gamma,q}$ is the γ quantile of the chi-squared distribution with q degrees of freedom. This formula illustrates the inflation in sample size needed to account for the unknown value of $\widehat{\Delta}_i^2(N_i)$. Equation (14) is also useful as an initial value for solving the minimization in Equation (11).

Remark 4.2. Suppose that the number of controls differs across systems. This causes the first-stage residual-variance estimators $\widehat{\tau}_1^2(n_0), \widehat{\tau}_2^2(n_0), \ldots, \widehat{\tau}_k^2(n_0)$ to have different degrees of freedom. One fix is to select first-stage sample sizes for each system, say n_{0i}, such that $n_{01} - q_1 = n_{02} - q_2 = \cdots = n_{0k} - q_k$. Another approach is to use the adjusted constant

$$h' = h_{2,(1-\alpha_1)^{1/(k-1)},\min_i(n_0-q_i-1)}$$

which is valid when degrees of freedom are unequal [Boesel et al. 2003].

To gain some sense of when this new procedure will be effective we provide a rough sample-size comparison with Rinott's [1978] procedure (see Section 6 for a numerical example). Suppose that we choose α_0 very small and n_0 large enough that $h = h_{k,1-\alpha_1,n_0-q-1} \approx h_{k,1-\alpha,n_0-1}$, where the latter is the constant that

would be used in Rinott's [1978] procedure. Further, suppose δ is small enough that, using the approximate sample size formula (14) and ignoring rounding,

$$N_i \approx \frac{h^2 \widehat{\tau}_i^2(n_0)}{\delta^2} + \chi^2_{\gamma,q}.$$

The sample size of Rinott's [1978] procedure for each system $i = 1, 2, \ldots, k$ is

$$N_i^R \approx \frac{h^2\, S_i^2(n_0)}{\delta^2}.$$

Then the expected savings from using the CV procedure is approximately

$$\begin{aligned} \mathrm{E}[N_i^R - N_i] &\approx \frac{h^2\,\sigma_i^2}{\delta^2} - \frac{h^2\,\sigma_i^2}{\delta^2}(1 - R_i^2) - \chi^2_{\gamma,q} \\ &= \frac{h^2\,\sigma_i^2}{\delta^2} R_i^2 - \chi^2_{\gamma,q}. \end{aligned}$$

Therefore, an approximate criterion for the CV procedure to be more efficient in simulating system i is

$$R_i^2 > \left(\frac{\delta^2}{h^2 \sigma_i^2}\right) \chi^2_{\gamma,q} \approx \frac{\chi^2_{\gamma,q}}{\mathrm{E}[N_i^R]}$$

The correlation between the response and the controls must overcome the fixed cost $\chi^2_{\gamma,q}$ of using CVs. Specifically, the square of the multiple correlation coefficient must be larger than the ratio of the fixed cost $\chi^2_{\gamma,q}$ to the expected sample size of Rinott's [1978] procedure. This threshold is easier to overcome for problems with large marginal variance σ_i^2 or small indifference level δ. In other words, the harder the problem is, the more likely it is that selection of the best via Procedure 3 will be more efficient.

5. ESTIMATING THE BEST

We now consider a problem that is closely related to selection of the best system: estimating the value of μ_k. This is different from estimating μ_B, the expected value of the selected system B. Notice that the natural estimator $\widehat{\mu}_B(N_B)$ is biased high since $\widehat{\mu}_B(N_B) \geq \widehat{\mu}_k(N_k)$. Lesnevski et al. [2005] provide a set of conditions sufficient for estimators to provide a $(1-\alpha)100\%$ fixed-width confidence interval for μ_k. Here we restate the conditions and show how to form CV estimators that satisfy them when Model 1 applies. The confidence interval will have the form $[\widehat{\mu}_B(N_B) - a, \widehat{\mu}_B(N_B) + b]$ where $a + b = L$, the desired fixed width of the confidence interval.

To restate the conditions of Lesnevski et al. [2005], let $\widehat{\mu}_i$ denote a generic estimator of μ_i. To obtain a $(1-\alpha)100\%$ confidence interval of fixed width L, Lesnevski et al. [2005] show that it is sufficient that $a+b=L$, $\alpha_a+\alpha_b=\alpha$, and

$$\Pr\{\mu_k > \widehat{\mu}_k + b\} \leq \alpha_b, \quad \text{and} \tag{15}$$

$$\Pr\{\mu_i < \widehat{\mu}_i - a\} \leq \alpha_a/k,\ i=1,2,\ldots,k. \tag{16}$$

Remark 5.1. Lesnevski et al. [2005] split the error α into $\alpha_a+\alpha_b$ because it may be useful to spend unequal amounts of error on the lower and upper confidence bounds. If errors in one direction are no more important than in the other, $\alpha_a=\alpha_b=\alpha/2$ is a good choice.

Lesnevski et al. [2005] show that Inequalities (15)–(16) will be satisfied if there are increasing functions G_a and G_b defined on the positive part of the real line, such that, for all $i=1,2,\ldots,k$ and $x>0$,

$$\Pr\{\widehat{\mu}_i - \mu_i > x\} \leq 1-G_a(cx) \quad \text{and} \quad \Pr\{\widehat{\mu}_i - \mu_i < -x\} \leq 1-G_b(cx), \tag{17}$$

where

$$c = \frac{1}{L}\left(G_a^{-1}(1-\alpha_a/k) - G_b^{-1}(1-\alpha_b)\right), \tag{18}$$

$$a = \frac{1}{c}G_a^{-1}(1-\alpha_a/k), \quad \text{and} \tag{19}$$

$$b = \frac{1}{c}G_b^{-1}(1-\alpha_b). \tag{20}$$

They further require that $G_a(0) < 1-\alpha_a/k < \lim_{x\to\infty} G_a(x)$ and $G_b(0) < 1-\alpha_b < \lim_{x\to\infty} G_b(x)$, which guarantee that a and b exist and are positive. Then by Inequality (17) and Equation (19), $\Pr\{\widehat{\mu}_i - \mu_i > a\} \leq \alpha_a/k$, while by Inequality (17) and Equation (20), $\Pr\{\widehat{\mu}_i - \mu_i < -b\} \leq \alpha_b$. These inequalities, for all $i=1,2,\ldots,k$, are sufficient for Inequalities (15)–(16).

The following procedure is also given in Lesnevski et al. [2005]; our contribution is to prove that it provides estimators that satisfy Inequality (17). We state the procedure under the assumption that all systems have the same number of controls, q, relaxing this assumption in Remark 5.2.

Procedure 4 *Estimating the Best with Controls.*

1. Choose the fixed interval width $L > 0$ and confidence level $1-\alpha$. Also choose α_0, α_a, and α_b all positive such that $\alpha_a + \alpha_b = \alpha$, $\alpha_0 < \alpha_a/k$, and $\alpha_0 < \alpha_b$.

2. Obtain $n_0 > q+2$ observations from each system and calculate $\widehat{\tau}_i^2(n_0)$ for $i = 1, 2, \ldots, k$.

3. For each system $i = 1, 2, \ldots, k$, compute the final sample size

$$N_i = \min_{n \geq n_0} \left\{ n : \left(\frac{n-q}{q} \right) \left(\frac{n}{c^2 \widehat{\tau}_i^2(n_0)} - 1 \right) \geq \mathcal{F}_{1-\alpha_0, q, n-q} \right\} \tag{21}$$

where c is given by Equation (18) with $G_a(x) = G_b(x) = F_{t_{n_0-q-1}}(x) - \alpha_0$ and F_{t_ν} is the cumulative distribution function of the Student t-distribution with ν degrees of freedom.

4. Collect $N_i - n_0$ observations from system i and form the CV estimators $\widehat{\mu}_i(N_i)$ for $i = 1, 2, \ldots, k$.

5. Let $B = \text{argmax}_i \, \widehat{\mu}_i(N_i)$, and form the $(1-\alpha)100\%$ confidence interval

$$\mu_k \in \left[\widehat{\mu}_B(N_B) - a, \widehat{\mu}_B(N_B) + b \right].$$

Proposition 4. *If Model 1 holds and the distribution of each control* $\mathbf{C}_{ij}$ *is multivariate normal, then Procedure 4 produces CV estimators* $\widehat{\mu}_1(N_1)$, $\widehat{\mu}_2(N_2)$, …, $\widehat{\mu}_k(N_k)$ *that satisfy Inequality (17) with* $G_a(x) = G_b(x) = F_{t_{n_0-q-1}}(x) - \alpha_0$.

Notice that we have not assumed independence of residuals across systems. The joint distribution of quantities simulated from different systems is irrelevant. In particular, using CRN in Procedure 4 poses no problem, even if CVs do not explain all the dependence induced by CRN; on the other hand, CRN also produces no benefit for this procedure.

Applying the same argument as in the Remark in Appendix B, we can show that a good approximation for N_i is

$$N_i = \max \left\{ n_0, \left\lceil c^2 \widehat{\tau}_i^2(n_0) + \chi^2_{1-\alpha_0, q} \right\rceil \right\}. \tag{22}$$

Remark 5.2. If the systems have different numbers of controls, then the only adjustment required is to use

$$G_a(x) = G_b(x) = F_{t_{n_0-d-1}}(x) - \alpha_0 \tag{23}$$

where $d = \max_i\{q_i\}$. The approximation for N_i then becomes

$$N_i = \max\left\{n_0, \left\lceil c^2\widehat{\tau}_i^2(n_0) + \chi^2_{1-\alpha_0,d}\right\rceil\right\},$$

where c is obtained using Equation (23) in Equation (18).

6. ILLUSTRATION

In this section we use a small example to illustrate the use of the screening procedure of Section 3, the selection procedure of Section 4, and the estimation procedure of Section 5. See Lesnevski et al. [2005] for additional financial applications of the procedure for estimating the best.

For the purpose of illustration we use an (s, S) inventory planning example from Koenig and Law [1985]. In an (s, S) inventory system some discrete item is periodically reviewed. If the inventory level is found to be below s units then an order is placed to bring the inventory level up to S units; otherwise no additional items are ordered. Ordering, unit, holding and shortage costs are incurred. The only stochastic element is the demand for inventory in each period, which is i.i.d. Poisson with a mean of 25 items per period. We will consider $k = 5$ different (s, S) inventory policies where the performance measure is the expected value of the average cost over 30 periods. The five policies are shown in Table I along with their true expected average costs (thus, policy 2 is actually the best). See Koenig and Law [1985] for a detailed description of the model.

Following the suggestion of Wilson and Pritsker [1984a; 1984b]), we take as the control on replication j the standardized average of the demands over the 30 periods:

$$C_j = \frac{\sqrt{30}(\bar{D}_j - 25)}{\sqrt{25}}$$

where $\bar{D}_j = \sum_{m=1}^{30} D_{mj}/30$, and D_{mj} is the Poisson demand in period m of replication j.

Because the policy does not affect the distribution of demand, using CRN makes the controls identical for each inventory policy. As discussed at the end of Section 3.2, this makes CVs ineffective in Procedure 2. We report the performance of Procedures 1, 3 and 4 for this example.

Because we are interested in the minimum expected cost we apply our procedures to $-X_{ij}$, where X_{ij} is the observed average cost of policy i on replication j. The impact of CRN can be observed in the estimated

Table I. The Five (s, S) Inventory Policies and their Expected Average Costs per Period in Thousands of Dollars

Policy i	s	S	μ_i
1	20	40	114.18
2	20	80	112.74
3	40	60	130.55
4	40	100	130.70
5	60	100	147.38

Table II. Fraction of Time that Each Inventory Policy was in the Retained Subset Based on 1000 Trials

	Policy				
Procedure	1	2	3	4	5
Boesel et al. [2003]	0.89	1.00	0.00	0.00	0.00
Procedure 1	0.59	1.00	0.00	0.00	0.00

correlation matrix of $(X_{1j}, X_{2j}, \ldots, X_{5j})$, which is

$$\begin{pmatrix} 1.00 & 0.56 & 0.90 & 0.48 & 0.65 \\ & 1.00 & 0.51 & 0.95 & 0.43 \\ & & 1.00 & 0.43 & 0.61 \\ & & & 1.00 & 0.36 \\ & & & & 1.00 \end{pmatrix}.$$

Similarly, the vector of estimated correlation coefficients

$$(R_1, R_2, \ldots, R_5) = (0.87, 0.68, 0.81, 0.58, 0.66)$$

documents the dependence between the output (cost) and the control (demand).

Table II shows the results from 1000 applications of the screening procedure of Boesel et al. [2003] and our Procedure 1 with $n = 30$ replications per application and confidence level $1 - \alpha = 0.95$. Although screening turns out not to be particularly difficult in this problem, screening under Procedure 1 is much less likely to retain policy 1 than the procedure of Boesel et al. [2003]; in other words, the expected size of the subset is smaller.

In the next experiment, we applied Procedure 3 and Rinott's [1978] procedure to the problem of selecting the inventory policy with the minimum expected cost. We set $\delta = 1000$ dollars, the initial sample size to $n_0 = 10$ replications, the probability of correct selection to $1-\alpha = 0.95$, and split the allowable $\alpha = 0.05$ error into $\alpha_1 = 0.048$ for correct selection and $\alpha_0 = 0.002$ for controlling the control. We used the approximate sample size in Equation (14). In 100 trials, both Rinott's [1978] procedure and our Procedure 3 selected

Table III. Average Sample Size per Policy in 100 Trials of Selecting the Best with CRN

	Policy					
Procedure	1	2	3	4	5	Total
Rinott [1978]	203	267	254	221	114	1059
Procedure 3	68	175	111	192	87	623

the true best (policy 2) 100 times. Table III shows the average sample sizes per policy over the 100 trials. Both procedures achieve the required probability of correct selection, but our procedure is able to reduce the sample size 41% by using CVs to capture the strong positive correlation due to CRN.

To illustrate estimating the best we alter the problem by assuming that our inventory is vendor managed, meaning that the vendor will set the inventory policy. We are interested in knowing how costly this could be for us. Therefore, we want to estimate the expected value of the worst (maximum inventory cost) policy and thus apply Procedure 4 to the costs X_{ij} rather than $-X_{ij}$.

We applied Procedure 4 both with and without CVs. Without CVs the point estimators become sample means and the procedure is similar to one by Chen and Dudewicz [1976]. We set the desired confidence interval width to $L = 1000$ dollars, the initial sample size to $n_0 = 10$ replications, and the coverage probability to $1-\alpha = 0.95$. We split the allowable $\alpha = 0.05$ error into $\alpha_a = \alpha_b = 0.025$ for the upper and lower confidence limits. When using the single control, we allocate $\alpha_0 = 0.002$ for controlling the control; without a control, $\alpha_0 = 0$. The critical values are

$$
\begin{aligned}
c &= \frac{1}{L}\left(t_{1-\alpha_a/k+\alpha_0,n_0-q-1} + t_{1-\alpha_b+\alpha_0,n_0-q-1}\right), \\
a &= \frac{1}{c} t_{1-\alpha_a/k+\alpha_0,n_0-q-1}, \quad \text{and} \\
b &= \frac{1}{c} t_{1-\alpha_b+\alpha_0,n_0-q-1},
\end{aligned}
$$

where $q = 0$ or 1 is the number of controls used. For example, when using a single control, $c = (t_{0.997,8} + t_{0.977,8})/L$. We used the approximate sample size in Equation (22).

In 100 trials, the confidence interval formed by Procedure 4 without CVs contained the largest expected value (147.38) 99 times, while Procedure 4 with CVs covered on 98 trials. Table IV shows the average sample size per policy over the 100 trials. Both procedures achieve the required coverage probability of 95%, but

Table IV. Average Sample Size per Policy in 100 Trials of Estimating the Expected Value of the Most Expensive Policy with CRN

	Policy					
Procedure	1	2	3	4	5	Total
Procedure 4 without CV	454	556	566	472	266	2314
Procedure 4 with CV	147	400	239	422	204	1411

employing CVs reduces the sample size 39%.

Remark 6.1. Although we have presented procedures for screening, selecting the best and estimating the best separately, Nelson et al. [2001]—for selecting the best—and Lesnevski et al. [2005]—for estimating the best—show that even more efficient procedures can be obtained by screening before selecting or estimating. Screening reduces the number of systems that receive the total sample needed for selecting or estimating the best, which is often large.

7. CONCLUSIONS

In this paper we have proposed a new class of R&S procedures based on control-variate estimators. R&S procedures based on sample means are good general-purpose tools, while these new procedures will be particularly useful in contexts where computationally intensive problems are solved repeatedly so that it pays to invest some effort in finding good controls. Financial engineering is one such context: see Glasserman [2004] for a general discussion of CVs in financial engineering applications and Lesnevski et al. [2005] for a specific application. Fortunately, implementing our CV R&S procedures requires only standard tools (least-squares regression) and readily available critical values.

APPENDIX

A. SCREENING

A.1 Screening with Individual Controls

We first prove Proposition 1. The proof is similar to Nelson et al. [2001, pp. 961–962], to which we refer the reader for a somewhat more detailed development. We will need the following three lemmas.

LEMMA A.1 [BANERJEE 1961]. *Let Z be a $N(0,1)$ random variable that is independent of $Y_1, Y_2, \ldots, Y_k$, which are independent chi-squared random variables with Y_i having degrees of freedom ν_i. Let $\gamma_1, \gamma_2, \ldots, \gamma_k$*

be arbitrary weights such that $\sum_{i=1}^{k} \gamma_i = 1$ *and all* $\gamma_i \geq 0$. *Then*

$$\Pr\left\{ Z^2 \leq \sum_{i=1}^{k} t_i^2 \gamma_i \frac{Y_i}{\nu_i} \right\} \geq 1 - \alpha,$$

where $t_i = t_{1-\alpha/2,\nu_i}$.

LEMMA A.2 SLEPIAN'S INEQUALITY [TONG 1980]. *Let* $(Z_1, Z_2, \ldots, Z_k)$ *have a* k*-variate normal distribution with zero mean vector, unit variances, and correlation matrix* $\mathbf{\Sigma} = \{\rho_{ij}\}$. *Let* $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_k$ *be some constants. If all the* $\rho_{ij} \geq 0$, *then*

$$\Pr\left\{ \bigcap_{i=1}^{k} (Z_i \leq \varepsilon_i) \right\} \geq \prod_{i=1}^{k} \Pr\{Z_i \leq \varepsilon_i\}.$$

LEMMA A.3 [TAMHANE 1977]. *Let* $V_1, V_2, \ldots, V_k$ *be independent random variables, and let* $g_j(v_1, v_2, \ldots, v_k)$, $j = 1, 2, \ldots, p$, *be nonnegative, real-valued functions, each one nondecreasing in each of its arguments. Then*

$$\mathrm{E}\left[\prod_{j=1}^{p} g_j(V_1, V_2, \ldots, V_k) \right] \geq \prod_{j=1}^{p} \mathrm{E}\left[g_j(V_1, V_2, \ldots, V_k) \right].$$

For convenience, let

$$\mathbf{C} = \{\mathbf{C}_{ij}, j = 1, 2, \ldots, n_i, i = 1, 2, \ldots, k\}$$

be the collection of all observed controls, and let

$$\widehat{\tau}^2 = \{\widehat{\tau}_1^2(n_1), \widehat{\tau}_2^2(n_2), \ldots, \widehat{\tau}_k^2(n_k)\}$$

be the collection of all observed residual-variance estimators. Define

$$Z_i = \frac{\widehat{\mu}_k(n_k) - \widehat{\mu}_i(n_i) - (\mu_k - \mu_i)}{\sqrt{\widehat{\Delta}_k^2(n_k)\tau_k^2 + \widehat{\Delta}_i^2(n_i)\tau_i^2}}. \tag{24}$$

LEMMA A.4. *If Model 1 holds with all residuals* $\{\eta_{ij}, i = 1, 2, \ldots, k, j = 1, 2, \ldots, n\}$ *independent, then the conditional covariance of* Z_i *and* Z_ℓ *given* $\mathbf{C}$ *is nonnegative for any systems* $i \neq \ell$.

Proof: By assumption, for any systems $i \neq \ell$ and observations j, j', the outputs X_{ij} and $X_{\ell j'}$ are conditionally independent given all observed controls $\mathbf{C}$. Equation (2) shows that the CV estimator $\widehat{\mu}_i(n_i)$ is a weighted average of the observations $X_{i1}, X_{i2}, \ldots, X_{in_i}$, where the weights are a function of $\mathbf{C}$ only. Thus, the CV estimators $\widehat{\mu}_i(n_i)$ and $\widehat{\mu}_\ell(n_\ell)$ are conditionally independent given $\mathbf{C}$. Similarly, $\widehat{\Delta}_i^2(n_i)$ defined in Equation (5)

is only a function of $\mathbf{C}$. Therefore, conditional on $\mathbf{C}$, the covariance of Z_i and Z_ℓ has the same sign as the conditional covariance between $\widehat{\mu}_k(n_k) - \widehat{\mu}_i(n_i)$ and $\widehat{\mu}_k(n_k) - \widehat{\mu}_\ell(n_\ell)$, which is nonnegative.

We are interested in

$$\begin{aligned}
\Pr\{k \in I\} &= \mathrm{E}\left[\Pr\{k \in I | \widehat{\tau}^2, \mathbf{C}\}\right] \\
&= \mathrm{E}\left[\Pr\left\{\widehat{\mu}_k(n_k) - \widehat{\mu}_i(n_i) \geq -W_{ki}, \forall i \neq k | \widehat{\tau}^2, \mathbf{C}\right\}\right] \\
&= \mathrm{E}\left[\Pr\left\{Z_i \geq \frac{-(\mu_k - \mu_i) - W_{ki}}{\sqrt{\widehat{\Delta}_k^2(n_k)\tau_k^2 + \widehat{\Delta}_i^2(n_i)\tau_i^2}}, \forall i \neq k \middle| \widehat{\tau}^2, \mathbf{C}\right\}\right] \\
&\geq \mathrm{E}\left[\Pr\left\{Z_i \geq \frac{-W_{ki}}{\sqrt{\widehat{\Delta}_k^2(n_k)\tau_k^2 + \widehat{\Delta}_i^2(n_i)\tau_i^2}}, \forall i \neq k \middle| \widehat{\tau}^2, \mathbf{C}\right\}\right] \qquad (25) \\
&= \mathrm{E}\left[\Pr\left\{Z_i \leq \frac{W_{ki}}{\sqrt{\widehat{\Delta}_k^2(n_k)\tau_k^2 + \widehat{\Delta}_i^2(n_i)\tau_i^2}}, i = 1, 2, \ldots, k-1 \middle| \widehat{\tau}^2, \mathbf{C}\right\}\right] \qquad (26)
\end{aligned}$$

where Inequality (25) follows because $\mu_k - \mu_i \geq 0$. By Lemma 2.1, the conditional distribution of Z_i given $\{\widehat{\tau}^2, \mathbf{C}\}$ is standard normal, so Equation (26) follows because of the symmetry of the normal distribution.

Because of Lemma A.4, Lemma A.2 applies to Equation (26), yielding

$$\begin{aligned}
&\Pr\left\{Z_i \leq \frac{W_{ki}}{\sqrt{\widehat{\Delta}_k^2(n_k)\tau_k^2 + \widehat{\Delta}_i^2(n_i)\tau_i^2}}, i = 1, 2, \ldots, k-1 \middle| \widehat{\tau}^2, \mathbf{C}\right\} \\
&\quad \geq \prod_{i=1}^{k-1} \Pr\left\{Z_i \leq \frac{W_{ki}}{\sqrt{\widehat{\Delta}_k^2(n_k)\tau_k^2 + \widehat{\Delta}_i^2(n_i)\tau_i^2}} \middle| \widehat{\tau}^2, \mathbf{C}\right\}. \qquad (27)
\end{aligned}$$

Further, since

$$\frac{W_{ki}}{\sqrt{\widehat{\Delta}_k^2(n_k)\tau_k^2 + \widehat{\Delta}_i^2(n_i)\tau_i^2}} = \sqrt{\frac{t_k^2\,\widehat{\Delta}_k^2(n_k)\widehat{\tau}_k^2(n_k) + t_i^2\,\widehat{\Delta}_i^2(n_i)\widehat{\tau}_i^2(n_i)}{\widehat{\Delta}_k^2(n_k)\tau_k^2 + \widehat{\Delta}_i^2(n_i)\tau_i^2}}$$

is nonnegative, real-valued, and nondecreasing in each of $\{\widehat{\tau}_1^2(n_1), \widehat{\tau}_2^2(n_2), \ldots, \widehat{\tau}_k^2(n_k)\}$, and by Lemma 2.1 $\{\widehat{\tau}_1^2(n_1), \widehat{\tau}_2^2(n_2), \ldots, \widehat{\tau}_k^2(n_k)\}$ are conditionally independent given $\mathbf{C}$, Lemma A.3 implies

$$\mathrm{E}\left[\prod_{i=1}^{k-1} \Pr\left\{Z_i \leq \frac{W_{ki}}{\sqrt{\widehat{\Delta}_k^2(n_k)\tau_k^2 + \widehat{\Delta}_i^2(n_i)\tau_i^2}} \middle| \widehat{\tau}^2, \mathbf{C}\right\} \middle| \mathbf{C}\right] \geq \prod_{i=1}^{k-1} \Pr\left\{Z_i \leq \frac{W_{ki}}{\sqrt{\widehat{\Delta}_k^2(n_k)\tau_k^2 + \widehat{\Delta}_i^2(n_i)\tau_i^2}} \middle| \mathbf{C}\right\}.$$

Combining this with Inequalities (26) and (27),

$$\Pr\{k \in I\} \geq \mathrm{E}\left[\prod_{i=1}^{k-1} \Pr\left\{ Z_i \leq \frac{W_{ki}}{\sqrt{\widehat{\Delta}_k^2(n_k)\tau_k^2 + \widehat{\Delta}_i^2(n_i)\tau_i^2}} \middle| \mathbf{C} \right\}\right]. \tag{28}$$

We next analyze a typical factor in this product.

First, notice that

$$\begin{aligned} &\left(\frac{W_{ki}}{\sqrt{\widehat{\Delta}_k^2(n_k)\tau_k^2 + \widehat{\Delta}_i^2(n_i)\tau_i^2}}\right)^2 \qquad (29)\\ &= \frac{t_k^2\,\widehat{\Delta}_k^2(n_k)\widehat{\tau}_k^2(n_k) + t_i^2\,\widehat{\Delta}_i^2(n_i)\widehat{\tau}_i^2(n_i)}{\widehat{\Delta}_k^2(n_k)\tau_k^2 + \widehat{\Delta}_i^2(n_i)\tau_i^2}\\ &= t_k^2\left(\frac{\widehat{\Delta}_k^2(n_k)\tau_k^2}{\widehat{\Delta}_k^2(n_k)\tau_k^2 + \widehat{\Delta}_i^2(n_i)\tau_i^2}\right)\frac{\widehat{\tau}_k^2(n_k)}{\tau_k^2} + t_i^2\left(\frac{\widehat{\Delta}_i^2(n_i)\tau_i^2}{\widehat{\Delta}_k^2(n_k)\tau_k^2 + \widehat{\Delta}_i^2(n_i)\tau_i^2}\right)\frac{\widehat{\tau}_i^2(n_i)}{\tau_i^2}. \end{aligned}$$

By Lemma 2.1, conditional on $\mathbf{C}$, $\widehat{\tau}_k^2(n_k)/\tau_k^2$ and $\widehat{\tau}_i^2(n_i)/\tau_i^2$ are distributed as $\chi^2_{n_k-q_k-1}/(n_k-q_k-1)$ and $\chi^2_{n_i-q_i-1}/(n_i-q_i-1)$ respectively, independent of each other, and independent of Z_i, which is standard normal. Thus, Equation (29) is a weighted average suitable for the application of Lemma A.1. Making use of the symmetry of the normal distribution, we obtain

$$\begin{aligned} &\Pr\left\{ Z_i \leq \frac{W_{ki}}{\sqrt{\widehat{\Delta}_k^2(n_k)\tau_k^2 + \widehat{\Delta}_i^2(n_i)\tau_i^2}} \middle| \mathbf{C} \right\}\\ &= \frac{1}{2} + \frac{1}{2}\Pr\left\{ Z_i^2 \leq \left(\frac{W_{ki}}{\sqrt{\widehat{\Delta}_k^2(n_k)\tau_k^2 + \widehat{\Delta}_i^2(n_i)\tau_i^2}}\right)^2 \middle| \mathbf{C} \right\}\\ &\geq \frac{1}{2} + \frac{1}{2}\left(1 - 2\left(1 - (1-\alpha)^{\frac{1}{k-1}}\right)\right) \qquad (30)\\ &= (1-\alpha)^{\frac{1}{k-1}} \end{aligned}$$

where Inequality (30) follows from Lemma A.1. This holds for all $\mathbf{C}$, so from Inequality (28) we get $\Pr\{k \in I\} \geq 1-\alpha$.

A.2 Screening with Paired Controls

Proving Proposition 2 is more straightforward:

$$\Pr\{k \in I\} \;=\; \Pr\{\, \widehat{\mu}_{ki}(n) \geq -t\,\widehat{\Delta}_{ki}(n)\widehat{\tau}_{ki}(n), \forall i \neq k\}$$

$$\geq 1 - \sum_{i=1}^{k-1} \Pr\left\{ \widehat{\mu}_{ki}(n) < -t\, \widehat{\Delta}_{ki}(n)\widehat{\tau}_{ki}(n) \right\} \tag{31}$$

$$= 1 - (k-1)(\alpha/(k-1)) = 1 - \alpha. \tag{32}$$

The Bonferroni inequality justifies Inequality (31), while Equation (32) is true because Lemma 2.1 implies that, conditional on $\mathbf{C}$,

$$\frac{\widehat{\mu}_{ki}(n) - \mu_{ki}}{\sqrt{\widehat{\Delta}_{ki}^2(n)\widehat{\tau}_{ki}^2(n)}}$$

has a Student t-distribution with $n - q - 1$ degrees of freedom; because this is so whatever the level of $\mathbf{C}$, it also has this unconditional distribution.

B. SELECTING THE BEST WITH CONTROLS

To establish Proposition 3, we will prove that $\Pr\{B = k\} \geq 1 - \alpha$ if $\mu_k - \mu_{k-1} \geq \delta$. The fact that the confidence intervals (12) cover with probability at least $1 - \alpha$, regardless of the configuration of the true means, then follows from Theorem 1 of Nelson and Matejcik [1995], while Inequality (13) is true as a consequence of Corollary 1 of Nelson and Goldsman [2001].

We will need the following lemma, which is a direct consequence of Corollary 5.2.1 of Anderson [1984]:

LEMMA B.1. *If the distribution of $\mathbf{C}_{ij}$ is q-variate normal, then*

$$\left(\frac{n-q}{q}\right)\left(n\widehat{\Delta}_i^2(n) - 1\right)$$

has an F distribution with $(q, n-q)$ degrees of freedom.

Although we have assumed that the distribution of $\mathbf{C}_{ij}$ is q-variate normal, see Remark B.2.

To simplify the notation we let $h = h_{k,1-\alpha_1,n_0-q-1}$ and define Z_i as in Equation (24). The probability of correct selection is

$$\begin{aligned}
&\Pr\{\widehat{\mu}_k(N_k) > \widehat{\mu}_i(N_i), \forall i \neq k\} \\
&\geq \Pr\left\{\widehat{\mu}_k(N_k) > \widehat{\mu}_i(N_i), \forall i \neq k;\ \widehat{\Delta}_i^2(N_i) \leq \frac{\delta^2}{h^2\,\widehat{\tau}_i^2(n_0)}, \forall i\right\} \\
&= \Pr\left\{Z_i > \frac{-(\mu_k - \mu_i)}{\sqrt{\widehat{\Delta}_k^2(N_k)\tau_k^2 + \widehat{\Delta}_i^2(N_i)\tau_i^2}}, \forall i \neq k;\ \widehat{\Delta}_i^2(N_i) \leq \frac{\delta^2}{h^2\,\widehat{\tau}_i^2(n_0)}, \forall i\right\}
\end{aligned}$$

$$\geq \Pr\left\{ Z_i > \frac{-\delta}{\sqrt{\frac{\delta^2}{h^2\,\widehat{\tau}_k^2(n_0)}\tau_k^2 + \frac{\delta^2}{h^2\,\widehat{\tau}_i^2(n_0)}\tau_i^2}}, \forall i \neq k;\ \widehat{\Delta}_i^2(N_i) \leq \frac{\delta^2}{h^2\,\widehat{\tau}_i^2(n_0)}, \forall i \right\} \tag{33}$$

$$= \Pr\left\{ Z_i > \frac{-h}{\sqrt{\frac{\tau_k^2}{\widehat{\tau}_k^2(n_0)} + \frac{\tau_i^2}{\widehat{\tau}_i^2(n_0)}}}, \forall i \neq k;\ \widehat{\Delta}_i^2(N_i) \leq \frac{\delta^2}{h^2\,\widehat{\tau}_i^2(n_0)}, \forall i \right\}$$

$$\geq 1 - p_1 - p_0 \tag{34}$$

where

$$p_1 = 1 - \Pr\left\{ Z_i > \frac{-h}{\sqrt{\frac{\tau_k^2}{\widehat{\tau}_k^2(n_0)} + \frac{\tau_i^2}{\widehat{\tau}_i^2(n_0)}}}, \forall i \neq k \right\} \quad \text{and}$$

$$p_0 = 1 - \Pr\left\{ \widehat{\Delta}_i^2(N_i) \leq \frac{\delta^2}{h^2\,\widehat{\tau}_i^2(n_0)}, \forall i \right\}.$$

Inequality (33) holds because $\mu_k - \mu_i \geq \delta$ and because of the bound on the value of $\widehat{\Delta}_i^2(N_i)$, while Inequality (34) is an application of the Bonferroni inequality.

First consider p_0. The probability

$$\begin{aligned}
&\Pr\left\{ \widehat{\Delta}_i^2(N_i) \leq \frac{\delta^2}{h^2\,\widehat{\tau}_i^2(n_0)} \right\} \\
&= \mathrm{E}\left[\Pr\left\{ \widehat{\Delta}_i^2(N_i) \leq \frac{\delta^2}{h^2\,\widehat{\tau}_i^2(n_0)} \middle| \widehat{\tau}_i^2(n_0) \right\} \right] \\
&= \mathrm{E}\left[\Pr\left\{ \left(\frac{N_i - q}{q}\right)\left(N_i\widehat{\Delta}_i^2(N_i) - 1\right) \leq \left(\frac{N_i - q}{q}\right)\left(\frac{N_i\delta^2}{h^2\,\widehat{\tau}_i^2(n_0)} - 1\right) \middle| \widehat{\tau}_i^2(n_0) \right\} \right].
\end{aligned}$$

Conditional on $\widehat{\tau}_i^2(n_0)$, N_i is constant. By Lemma 2.1, $\widehat{\tau}_i^2(n_0)$ and $\{\mathbf{C}_{i1}, \mathbf{C}_{i2}, \ldots\}$ are independent, so the conditional distribution of $\{\mathbf{C}_{i1}, \mathbf{C}_{i2}, \ldots\}$ given $\widehat{\tau}_i^2(n_0)$ is the same as its unconditional distribution, namely q-variate normal. Therefore Lemma B.1 implies that the conditional distribution of $((N_i - q)/q)(N_i\widehat{\Delta}_i^2(N_i) - 1)$ given $\widehat{\tau}_i^2(n_0)$ is F with $(q, N_i - q)$ degrees of freedom. Consequently,

$$\Pr\left\{ \left(\frac{N_i - q}{q}\right)\left(N_i\widehat{\Delta}_i^2(N_i) - 1\right) \leq \left(\frac{N_i - q}{q}\right)\left(\frac{N_i\delta^2}{h^2\,\widehat{\tau}_i^2(n_0)} - 1\right) \middle| \widehat{\tau}_i^2(n_0) \right\} \geq \gamma,$$

as defined in Procedure 3. If systems are simulated independently, $\gamma = (1 - \alpha_0)^{1/k}$, and apply Slepian's inequality (Lemma A.2). If systems are simulated with CRN, $\gamma = 1 - \alpha_0/k$, and apply Bonferroni's inequality. In either case, the result is $p_0 \leq \alpha_0$.

Now consider p_1. Given $\mathbf{C}$ and $\widehat{\tau}^2$, the conditional distribution of Z_i is standard normal, by Lemma 2.1. By Lemma A.4, the conditional covariance of Z_i and Z_ℓ is nonnegative. Therefore,

$$\mathrm{E}\left[\Pr\left\{Z_i \geq \frac{-h}{\sqrt{\frac{\tau_k^2}{\widehat{\tau}_k^2(n_0)}+\frac{\tau_i^2}{\widehat{\tau}_i^2(n_0)}}}, \forall i \neq k \,\middle|\, \mathbf{C}, \widehat{\tau}^2\right\}\right] \geq \mathrm{E}\left[\prod_{i=1}^{k-1} \Pr\left\{Z_i \geq \frac{-h}{\sqrt{\frac{\tau_k^2}{\widehat{\tau}_k^2(n_0)}+\frac{\tau_i^2}{\widehat{\tau}_i^2(n_0)}}} \,\middle|\, \mathbf{C}, \widehat{\tau}^2\right\}\right]$$

by Slepian's inequality (Lemma A.2). Conditional on $\mathbf{C}$, the expectation over $\widehat{\tau}^2$ is Rinott's [1978] integral, which is at least $1-\alpha_1$ because of the procedure's choice of N_i. This is true for all $\mathbf{C}$, so it is true unconditionally as well.

Thus, the probability of correct selection $\Pr\{\widehat{\mu}_k(N_k) > \widehat{\mu}_i(N_i), \forall i \neq k\} \geq 1-p_0-p_1 \geq 1-\alpha_0-\alpha_1 = 1-\alpha$.

Remark B.2. Even if the distribution of $\mathbf{C}_{ij}$ is not multivariate normal, Theorem 5.2.3 of Anderson [1984] can be used to show that for large n,

$$(n-1)\left(n\widehat{\Delta}_i^2(n)-1\right)$$

has approximately the chi-squared distribution with q degrees of freedom. Take N_i to be the smallest integer greater than n_0 such that

$$(N_i-1)\left(\frac{N_i\delta^2}{h^2\widehat{\tau}_i^2(n_0)}-1\right) \geq \chi^2_{\gamma,q}.$$

Therefore, the procedure will be approximately valid when N_i is large, without the requirement that the distribution of the controls be multivariate normal. Notice that N_i will be large when the variance of the CV estimator is large, or when the difference worth detecting δ is small, the two most difficult situations in which to select the best. To obtain a convenient sample size formula, we solve the equation

$$N_i\left(\frac{N_i\delta^2}{h^2\widehat{\tau}_i^2(n_0)}-1\right) = \chi^2_{\gamma,q}.$$

Its positive solution is

$$N_i = \left(\frac{h^2\widehat{\tau}_i^2(n_0)}{2\delta^2}\right)\left(1+\sqrt{1+\frac{4\delta^2}{h^2\widehat{\tau}_i^2(n_0)}\chi^2_{\gamma,q}}\right). \tag{35}$$

Applying the inequality $\sqrt{1+x} \leq 1+x/2$ yields Equation (14) as an upper bound for Equation (35).

Giving a rule of thumb for how large the sample size N_i must be for an approximation such as Equation (14) or (35) to be adequate when the distribution of $\mathbf{C}_{ij}$ is not multivariate normal is not straightforward. Not only

does the adequacy of the approximation depend on the actual distribution of $\mathbf{C}_{ij}$, but it also depends on the number of controls q and the tail probability γ at which we hope that the quantile of $(n-1)\left(n\widehat{\Delta}_i^2(n)-1\right)$ is close to a chi-squared quantile. However, we can address the issue of when Equation (14) is a good approximation for Equation (11) by comparing Equations (14) and (35) to Equation (11) while varying q, γ, and $h^2\widehat{\tau}_i^2(n_0)/\delta^2$. In all cases, we found that Equation (11), which is exactly correct if $\mathbf{C}_{ij}$ is multivariate normal, is no less than Equation (35) and no more than Equation (14). We recommend using Equation (14) as a conservative approach. For each system in the example reported in Section 6, Equation (14) was either exactly equal to Equation (11) or larger by one. Equation (14) can be substantially larger in other cases, leading to a loss of efficiency, just as Equation (35) can be substantially less, leading to a violation of $p_0 \le \alpha_0$. After numerical experimentation guided by the fact that when N_i is specified by Equation (14),

$$\left(\frac{N_i-q}{q}\right)\left(\frac{N_i\delta^2}{h^2\widehat{\tau}_i^2(n_0)}-1\right)=\frac{\chi^2_{\gamma,q}}{q}\left(1+\frac{(\chi^2_{\gamma,q}-q)\delta^2}{h^2\widehat{\tau}_i^2(n_0)}\right),$$

we found that Equation (14) was no more than 1% larger than Equation (11) when $h^2\widehat{\tau}_i^2(n_0)/\delta^2$ is at least ten times larger than $\chi^2_{\gamma,q}-q$.

C. ESTIMATING THE BEST WITH CONTROLS

To prove Proposition 4 we will use the tools developed in Appendix B. As in that section, we present a proof based on the assumption that the distribution of $\mathbf{C}_{ij}$ is q-variate normal, bearing in mind that Remark B.2 demonstrates approximate validity if the sample size is large, even if $\mathbf{C}_{ij}$ is not normal.

First notice that Inequalities (15)–(16) are all *marginal* requirements, involving only one system at a time, as is the sufficient condition (17). Thus, all we need to do is to show that under our procedure, Inequality (17) holds for each system i when $G_a(x)=G_b(x)=F_{t_{n_0-q-1}}(x)-\alpha_0$. We will need the following lemma, which is a consequence of Theorem 1 of [Nelson 1990]:

LEMMA C.1. *If Model 1 holds then*

$$\frac{\widehat{\mu}_i(N_i)-\mu_i}{\sqrt{\widehat{\Delta}_i^2(N_i)\widehat{\tau}_i^2(n_0)}}$$

has a t distribution with n_0-q-1 degrees of freedom.

Showing that $\Pr\{\widehat{\mu}_i-\mu_i>x\}\le 1-(F_{t_{n_0-q-1}}(cx)-\alpha_0)$ follows steps similar to the proof in Appendix B,

which we give in abbreviated form here:

$$\begin{aligned}
\Pr\{\widehat{\mu}_i(N_i) - \mu_i \le x\} &\ge \Pr\left\{\widehat{\mu}_i(N_i) - \mu_i \le x;\ \widehat{\Delta}_i^2(N_i) \le \frac{1}{c^2\widehat{\tau}_i^2(n_0)}\right\} \\
&= \Pr\left\{\frac{\widehat{\mu}_i(N_i) - \mu_i}{\sqrt{\widehat{\Delta}_i^2(N_i)\widehat{\tau}_i^2(n_0)}} \le \frac{x}{\sqrt{\widehat{\Delta}_i^2(N_i)\widehat{\tau}_i^2(n_0)}};\ \widehat{\Delta}_i^2(N_i) \le \frac{1}{c^2\widehat{\tau}_i^2(n_0)}\right\} \\
&\ge \Pr\left\{\frac{\widehat{\mu}_i(N_i) - \mu_i}{\sqrt{\widehat{\Delta}_i^2(N_i)\widehat{\tau}_i^2(n_0)}} \le \frac{x}{\sqrt{\frac{\widehat{\tau}_i^2(n_0)}{c^2\widehat{\tau}_i^2(n_0)}}};\ \widehat{\Delta}_i^2(N_i) \le \frac{1}{c^2\widehat{\tau}_i^2(n_0)}\right\} \\
&= \Pr\left\{\frac{\widehat{\mu}_i(N_i) - \mu_i}{\sqrt{\widehat{\Delta}_i^2(N_i)\widehat{\tau}_i^2(n_0)}} \le cx;\ \widehat{\Delta}_i^2(N_i) \le \frac{1}{c^2\widehat{\tau}_i^2(n_0)}\right\} \\
&\ge F_{t_{n_0-q-1}}(cx) - \alpha_0,
\end{aligned}$$

where the final step comes from an application of the Bonferroni inequality and Lemmas B.1 and C.1. The other half of Inequality (17) follows from this result and the symmetry of the distribution of $\widehat{\mu}_i - \mu_i$.

ACKNOWLEDGMENT

This research was partially supported by National Science Foundation grant numbers DMI-0217690 and DMS-0202985 and by National Security Agency grant number H98230-04-1-0047. The authors gratefully acknowledge the aid of the Editor and three anonymous referees.

REFERENCES

Anderson, T. W. 1984. *An Introduction to Multivariate Statistical Analysis*, 2nd ed. John Wiley & Sons, New York.

Banerjee, S. 1961. On confidence interval for two-means problem based on separate estimates of variances and tabulated values of t-table. *Sankhyā A23*, 359–378.

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

Boesel, J., Nelson, B. L., and Kim, S. 2003. Using ranking and selection to 'clean up' after simulation optimization. *Operations Research 51*, 814–825.

Chen, H. J. and Dudewicz, E. J. 1976. Procedures for fixed-width interval estimation of the largest normal mean. *Journal of the American Statistical Association 71*, 752–756.

Glasserman, P. 2004. *Monte Carlo Methods in Financial Engineering*. Springer-Verlag, New York.

Koenig, L. W. and Law, A. M. 1985. A procedure for selecting a subset of size m containing the l best of k independent normal populations, with applications to simulation. *Communications in Statistics B14*, 719–734.

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

Law, A. M. and Kelton, W. D. 2000. *Simulation Modeling and Analysis*, 3rd ed. McGraw-Hill, New York.

Lesnevski, V., Nelson, B. L., and Staum, J. 2005. Simulation of coherent risk measures based on generalized scenarios. Working paper, Dept. of Industrial Engineering and Management Sciences, Northwestern University.

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

Nelson, B. L. 1993. Robust multiple comparisons under common random numbers. *ACM Transactions on Modeling and Computer Simulation 3*, 225–243.

Nelson, B. L. and Goldsman, D. 2001. Comparisons with a standard in simulation experiments. *Management Science 47*, 449–463.

Nelson, B. L. and Hsu, J. C. 1993. Control-variate models of common random numbers for multiple comparisons with the best. *Management Science 39*, 989–1001.

Nelson, B. L. and Matejcik, F. J. 1995. Using common random numbers for indifference-zone selection and multiple comparisons in simulation. *Management Science 41*, 1935–1945.

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

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

Stallard, C. and Owen, L. E. 1998. Evaluating adaptive signal control using CORSIM. In *Proceedings of the 1998 Winter Simulation Conference*, D. J. Medeiros, E. F. Watson, J. S. Carson, and M. S. Manivannan, Eds. IEEE Press, Piscataway, N. J., 1147–1153.

Swisher, J. R., Jacobson, S. H., and Yücesan, E. 2003. Discrete-event simulation optimization using ranking, selection, and multiple comparison procedures: A survey. *ACM Transactions on Modeling and Computer Simulation 13*, 134–154.

Tamhane, A. C. 1977. Multiple comparisons in model I one-way ANOVA with unequal variances. *Communications in Statistics A6*, 15–32.

Tong, Y. L. 1980. *Probability Inequalities in Multivariate Distributions*. Academic Press, New York.

Wilson, J. R. and Pritsker, A. A. B. 1984a. Experimental evaluation of variance reduction techniques for queueing simulation using generalized concomitant variables. *Management Science 30*, 1459–1472.

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

Yang, W.-N. and Nelson, B. L. 1991. Using common random numbers and control variates in multiple-comparison procedures. *Operations Research 39,* 4, 583–591.