Proceedings of the 2006 Winter Simulation Conference
L. F. Perrone, F. P. Wieland, J. Liu, B. G. Lawson, D. M. Nicol, and R. M. Fujimoto, eds.

COMBINED RANKING AND SELECTION WITH CONTROL VARIATES

Shing Chih Tsai
Barry L. Nelson

Department of Industrial Engineering & Management Sciences
Northwestern University
Evanston, IL 60208-3119, U.S.A.

ABSTRACT

Nelson and Staum derived R&S procedures that employ control-variate estimators (CVs) instead of sample means to obtain more statistical efficiency. However, control-variate estimators require more computational effort than sample means, and effective controls must be identified. We present a new CV screening procedure to avoid much of the computation cost. We also present a two-stage CV combined procedure which captures the ability to eliminate inferior systems in the first stage and the statistical efficiency of control variates for selection in the second stage. An empirical evaluation is provided.

1 INTRODUCTION

In simulation research and applications, ranking-and-selection procedures (R&S; see for instance Bechhofer et al. 1995) have proven to be quite useful for finding the system design that is the best, or near the best, where the "best" system is the one with the largest or smallest expected performance measure. However, R&S procedures are only recommended when the number of alternative designs is relatively small and the designs are not functionally related. For instance, the typical indifference-zone (IZ) selection procedure will require large numbers of observations to deliver the desired correct-selection guarantee when the number of systems is large. To solve this problem, Nelson et al. (2001) proposed a combined procedure which uses the subset selection approach to eliminate some uncompetitive systems in the first stage; it then applies a standard IZ selection procedure in the second stage. In this way, sampling cost can be saved while still maintaining the ease of implementation and statistical efficiency.

In almost all R&S procedures sample means of the responses are used as the estimators of the expected performance. Nelson and Staum (2006) derived R&S procedures which employ control-variate estimators instead of sample means. 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). These control-variate procedures can be more statistically efficient than the sample-mean-based procedures. However, control-variate estimators require more computational effort than sample means, and effective controls must be identified.

Our goal is to propose a two-stage procedure which captures the ability to screen out inferior systems and the statistical efficiency of control variates (CVs). We will use a screening procedure with control variates to eliminate obviously noncompetitive systems in the first stage and then apply a selection-of-the-best-with-control-variates procedure to the surviving subset of systems in the second stage. Nelson and Staum (2006) showed that the screening threshold with CVs is expected to be tighter than with sample means when the correlation between the output and control is not too small. Therefore, the expected subset size is correspondingly smaller. For the selection-of-the-best-with-control-variate procedure, Nelson and Staum (2006) also showed that we can expect savings of sample size compared with Rinott's (1978) procedure even when the correlation between the output and control is modest. So the sample size of the CV selection procedure is typically smaller than that of Rinott's (1978) procedure, which is based on sample means. Since the CV screening procedure is better than the standard screening procedure based on sample means, and the CV selection procedure is better than the selection procedure based on sample means, we can expect that a combined CV procedure is better than a combined procedure based on sample means. In this paper we develop the theory and procedures to support such a combined approach.

This paper is organized as follows: In Section 2, we outline the generic combined procedure. Sections 3–5 review CV estimators, and several CV R&S procedures. We also present a new CV screening procedure in Section 4. In Section 6, we present the CV combined procedure in detail. The paper ends with an empirical evaluation performed to compare the two combined procedures in Section 7, and

1-4244-0501-7/06/$20.00 ©2006 IEEE

conclusions in Section 8. The proofs of all theorems can be found in Tsai (2006).

2 GENERIC COMBINED PROCEDURE

In the CV combined procedure, we apply the CV selection-of-the-best procedure to the subset of systems chosen by the CV screening procedure to acquire both statistical and computational efficiency. The generic combined procedure is as follows. In the remainder of the paper we fill in specific pieces of this procedure.

1. For each system, obtain a small number of observations of the system performance measure and the controls. Then form CV estimators of each system's mean and calculate an estimator of the variance of each CV estimator.
2. Apply a CV screening procedure to eliminate inferior systems based on the information acquired in the first step.
3. If only one system survives, then stop and return that one as the best system. Otherwise, calculate the total number of observations needed for each system to detect a specified practically significant difference in performance with the desired confidence level.
4. Take additional observations from each surviving system and form CV estimators. Then select the system with the best CV estimator.

3 SCREENING WITH INDIVIDUAL CONTROL VARIATES

In this section we briefly provide the definitions and notation that will be used throughout the paper and review the screening procedure with individual controls in Nelson and Staum (2006). The following description is based on Nelson and Staum (2006).

3.1 Individual Control-Variate Estimators

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

$$X_{ij} = \mu_i + (\mathbf{C}_{ij} - \xi_i)'\beta_i + \eta_{ij} \tag{1}$$

where $\{\eta_{ij}, j = 1,2,\ldots,n\}$ is a set of i.i.d. $\mathrm{N}(0,\tau_i^2)$ random variables. The $q_i \times 1$ vector $\mathbf{C}_{ij}$ is called the *control* and is assumed multivariate normal. 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 ξ_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. The multiplier β_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.

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. The development is based on Nelson (1990), Nelson and Hsu (1993), and Nelson and Staum (2006).

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

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

We append "(n)" to quantities defined across n observations. Then the CV point estimator of μ_i is

$$\widehat{\mu}_i(n) = \bar{X}_i(n) - \left(\bar{\mathbf{C}}_i(n) - \xi_i\right)'\widehat{\beta}_i$$

(Nelson 1990). It is known that under Model (1)

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

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 $\mathbf{C}_{ij}$ (Lavenberg and Welch 1981).

We need to know the distribution of $\hat{\mu}_i(n)$ and an estimator of its variance to derive R&S procedures. Let

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

and define the residual variance estimator $\widehat{\tau}_i^2(n)$ as

$$\frac{1}{n-q_i-1}\mathbf{X}_i(n)'\left[\mathbf{I}-\mathbf{A}_i(n)\left(\mathbf{A}_i'(n)\mathbf{A}_i(n)\right)^{-1}\mathbf{A}_i'(n)\right]\mathbf{X}_i(n)$$
$$= \frac{1}{n-q_i-1}\sum_{j=1}^{n}\left[X_{ij}-\widehat{\mu}_i(n)-(\mathbf{C}_{ij}-\xi_i)'\widehat{\beta}_i(n)\right]^2. \quad (2)$$

Further, let

$$\widehat{\Delta}_i^2(n)=\frac{1}{n}+\frac{1}{n-1}\left(\bar{\mathbf{C}}_i(n)-\xi_i\right)'\mathbf{S}_{\mathbf{C}_i}^{-1}(n)\left(\bar{\mathbf{C}}_i(n)-\xi_i\right) \quad (3)$$

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 1 (Nelson & Hsu (1993), Thm. 4.1) *If Model (1) pertains, then conditional on* $\mathbf{C}_1(n),\mathbf{C}_2(n),\ldots,\mathbf{C}_k(n)$ *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.*

Property P3 requires that the η_{ij} are independent for all systems i as well as for all observations j. In practice, P3 will hold either if all systems are simulated independently, or if common random numbers (CRN) are used but the dependence due to CRN is entirely explained by the controls. CRN is a technique that tries to induce a positive correlation between the outputs of different systems by using the same pseudorandom numbers to simulate each alternative system and therefore reduce the variance of the difference between them.

3.2 Screening with Individual Control Variates

We will assume that unknown to us $\mu_k\geq\mu_{k-1}\geq\cdots\geq\mu_1$ and that bigger is better. The goal of the procedure is to find a subset I that contains system k with prespecified confidence. We also 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 t distribution with ν degrees of freedom.

Procedure 1 (Individual CV Screening)

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

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

where

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

Nelson and Staum (2006) proved that $\Pr\{k\in I_{Indiv}\}\geq 1-\alpha$ when Model (1) holds with independence among $\{\eta_{ij},i=1,2,\ldots,k,j=1,2,\ldots,n\}$. The advantage of this procedure is that we just need to compute k CVs. Its disadvantage is that the assumption that there is no dependence between residuals across systems induced by CRN will not hold in practice. So Nelson and Staum (2006) proposed a screening procedure with paired control variates.

4 SCREENING WITH PAIRED CONTROL VARIATES

In this section we review the screening procedure with paired control variates in Nelson and Staum (2006) and propose a new procedure to reduce the computation cost and retain the benefit of paired CVs as well. The description in Section 4.1 is based on Nelson and Staum (2006).

4.1 All-Pair Screening Procedure

Nelson and Staum (2006) use the paired controls model to avoid the assumption that the controls entirely explain the dependence induced by CRN. We form 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 $\xi_{i\ell}=\xi_i-\xi_\ell$. We need the observations and the controls to be paired across systems, so the number of observations should be equal for each system in the same pair, and also the number of controls for each system in the same pair is equal. For convenience we let n be the common number of replications and q be the common number of controls for each system. Then we assume that a model like Model (1) holds:

$$X_j(i,\ell)=\mu_{i\ell}+(\mathbf{C}_j(i,\ell)-\xi_{i\ell})'\mathbf{B}(i,\ell)+\varepsilon_j(i,\ell) \quad (4)$$

where $\{\varepsilon_j(i,\ell),j=1,2,\ldots,n\}$ is a set of i.i.d. $\mathrm{N}(0,\tau_{i\ell}^2)$ random variables. The $q\times 1$ vector $\mathbf{C}_j(i,\ell)$ is assumed multivariate normal. For each pair of systems $i,\ell=1,2,\ldots,k,i\neq\ell$ the controls $\{\mathbf{C}_j(i,\ell),j=1,2,\ldots,n\}$ are also i.i.d., are independent of $\{\varepsilon_j(i,\ell),j=1,2,\ldots,n\}$ and have known expected value $\xi_{i\ell}$. Unlike Model (1), Model (4) can hold even when η_{ij} and $\eta_{\ell j}$ are dependent.

For all $i\neq\ell$, we let $\widehat{\mu}_{i\ell}(n)$ be the corresponding CV estimator of $\mu_{i\ell}$, and define $\widehat{\tau}_{i\ell}^2(n)$ and $\widehat{\Delta}_{i\ell}^2(n)$ in analogy

to Equations (2) and (3), but applying CVs to differences between systems' output instead of to each system's output.

We call the following procedure proposed by Nelson and Staum (2006) the "All-Pair" screening procedure.

Procedure 2 (All-Pair Screening)

1. *Choose the confidence level* $1-\alpha > 1/k$.
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*

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

Nelson and Staum (2006) proved that $\Pr\{k \in I_{AllPair}\} \geq 1-\alpha$ when Model (4) holds. The advantage of this procedure is that we do not have to be concerned about the dependence remaining in residuals due to CRN. Its disadvantages are that we have to compute $k(k-1)/2$ CV estimators and that the procedure uses the conservative Bonferroni inequality.

4.2 "Best Bet" Screening Procedure

Nelson and Staum (2006) proved that $\Pr\{\, k \in I_{AllPair} \,\} \geq 1-\alpha$. However, the All-Pair screening procedure requires calculating $k(k-1)/2$ CV estimators which can be a large computation cost. In this subsection, we propose a new procedure which requires less computation and creates a subset $I \supseteq I_{AllPair}$, and therefore we can guarantee that $\Pr\{\, k \in I \,\} \geq 1-\alpha$. To accomplish this we choose some system K^* which is very likely to be the best system, and then perform screening with paired CVs just against K^*.

In the following "Best Bet" screening procedure, we denote the system with the largest $\widehat{\mu}_i(n)$ as K^*.

Procedure 3 (Best Bet Screening)

1. *Choose the confidence level* $1-\alpha > 1/k$.
2. *Obtain* $n > q+2$ *observations from each system and form the k CV estimators* $\widehat{\mu}_i(n), i = 1,2,\ldots,k$.
3. *Let* K^* *be the index of the system with the largest* $\widehat{\mu}_i(n)$*, that is,* $K^* = \arg\max_i \widehat{\mu}_i(n)$*, and then form the* $k-1$ *paired CV estimators* $\widehat{\mu}_{iK^*}(n)$ *for all* $i \neq K^*$.
4. *Let* $t = t_{1-\alpha/(k-1),n-q-1}$ *and create the subset*

$$I_{BestBet} = \left\{ i : \widehat{\mu}_{iK^*}(n) \geq -t\,\widehat{\Delta}_{iK^*}(n)\widehat{\tau}_{iK^*}(n), \right.$$
$$\left. \forall i \neq K^* \right\} \cup \{\, K^* \,\}.$$

The advantage of this procedure is that it can decrease the computation cost and achieve the desired statistical efficiency as well. The subset size will be close to that of the All-Pair screening procedure, because there is a large correlation between $\widehat{\mu}_{i\ell}(n)$ and $\widehat{\mu}_i(n) - \widehat{\mu}_\ell(n)$. And it also avoids the assumption that CVs explain all the dependence induced by CRN. Its disadvantage is that it needs to compute $2k-1$ CV estimators which is more than the individual CV screening procedure (k). However, it still saves computation cost compared with All-Pair screening procedure $(k(k-1)/2)$, when the number of alternatives is large.

5 SELECTING THE BEST WITH CONTROL VARIATES

In this section we briefly review the selection-of-the-best-with-control-variates procedure in Nelson and Staum (2006). Under Model (1), we adopt the indifference-zone (IZ) 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. We also assume that all systems have the same number of controls q. The procedure is as follows:

Procedure 4 (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 *observations and calculate* $\widehat{\tau}_i^2(n_0)$.
3. *For each system* $i = 1,2,\ldots,k$*, set the total sample size*

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

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

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

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* $B = \arg\max_i\ \widehat{\mu}_i(N_i)$.

Nelson and Staum (2006) proved that $\Pr\{B = k\} \geq 1-\alpha$ whenever $\mu_k - \mu_{k-1} \geq \delta$.

6 COMBINED PROCEDURE

In the combined procedure, we apply a screening procedure with control variates to eliminate noncompetitive systems in the first stage. Then in the second stage the CV selection-

of-the-best procedure is applied to the surviving systems to pick the best system, while still gaining the desired overall confidence level. Here are some key observations:

- We spend α_0 of the overall allowable error α for incorrect selection on the first screening stage, and $\alpha_1 + \alpha_2$ on the second selection-of-the-best stage.
- If we use the individual CV screening procedure in the first stage, then a multiplicative approach is applied:

$$1-\alpha \quad = (1-\alpha_0)(1-\alpha_1-\alpha_2).$$

- If we use the paired CV screening procedure in the first stage, then an additive approach is applied:

$$1-\alpha \quad = 1-\alpha_0-\alpha_1-\alpha_2.$$

- We set the appropriate critical constant t_i of each system $i = 1,2,\ldots,k$ in the CV screening procedure for k systems, n_i first stage samples, q_i control variates, and confidence level $1-\alpha_0$.
- We set the appropriate critical constant h of each system $i = 1,2,\ldots,k$ in the CV selection-of-the-best procedure for k systems, n_i first stage samples, q_i control variates, and confidence level $1-\alpha_1$.
- We set the appropriate critical constant γ in the CV selection-of-the-best procedure for k systems, confidence level $1-\alpha_2$, and depending on whether or not the systems are simulated independently.

In the procedure below we assume that $n_i - q_i$ is the same for each system $i = 1,2,\ldots,k$ and mention the necessary adjustment for unequal $n_i - q_i$ in Remark 6.1. Following is a procedure which combines the individual CV screening procedure with the CV selection-of-the-best procedure.

Procedure 5 (Individual CV Combined)

1. *Select overall confidence level $1/k < 1-\alpha < 1$, indifference-zone parameter $\delta > 0$, number of systems k, and first-stage sample size $n_i > q_i + 2$ from system $i = 1,2,\ldots,k$. Set $t_i = t_{(1-\alpha_0)^{1/(k-1)},n_i-q_i-1}$ and $h = h_{k,1-\alpha_1,n_i-q_i-1}$ which is Rinott's constant (see Wilcox 1984 or Bechhofer et al. 1995 for tables).*
2. *Obtain n_i observations from each system and calculate $\widehat{\mu}_i(n_i)$, $\widehat{\Delta}_i^2(n_i)$ and $\widehat{\tau}_i^2(n_i), i = 1,2,\ldots,k$. We also create the subset*

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

where

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

3. *If I contains a single index, then stop and return that system as the best. Otherwise, for all $i \in I$, compute the second-stage sample size*

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

where $\mathcal{F}^{(\gamma)}_{q_i,n_i-q_i}$ is the γ quantile of the F distribution with $(q_i, n_i - q_i)$ degrees of freedom, and

$$\gamma = \begin{cases} (1-\alpha_2)^{\frac{1}{k}}, & \text{if systems are independent} \\ 1-\alpha_2/k, & \text{otherwise.} \end{cases}$$

Notice that $1-\alpha = (1-\alpha_0)(1-\alpha_1-\alpha_2)$, the multiplicative approach.

4. *Take $N_i - n_i$ additional observations from all systems $i \in I$ and form the CV estimators $\widehat{\mu}_i(N_i)$ for $i \in I$.*
5. *Select the system $B = \arg\max_i \widehat{\mu}_i(N_i)$ as best from all systems $i \in I$.*

Theorem 1 *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 the individual CV combined procedure selects a system B such that $\Pr\{B = k\} \geq 1-\alpha$ whenever $\mu_k - \mu_{k-1} \geq \delta$.*

Remark 6.1 *Suppose that $n_i - q_i$ is different across systems. This causes the first-stage residual-variance estimators $\widehat{\tau}_1^2(n_1), \widehat{\tau}_2^2(n_2), \ldots, \widehat{\tau}_k^2(n_k)$ to have different degrees of freedom. One approach is to use the adjusted constant*

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

which is valid when degrees of freedom are unequal (Boesel, Nelson and Kim 2003).

Remark 6.2 *We can combine the paired CV screening procedure with the CV selection-of-the-best procedure to be the paired CV combined procedure. When we use Best Bet screening procedure, we need to change Step 2 to the following:*

2. *Obtain $n > q+2$ observations from each system and form the k CV estimators $\widehat{\mu}_i(n), i = 1,2,\ldots,k$. Let K^* be the index of the system with the largest $\widehat{\mu}_i(n)$, that is, $K^* = \arg\max_i \widehat{\mu}_i(n)$, and then form the $k-1$ paired CV estimators $\widehat{\mu}_{iK^*}(n)$ for all $i \neq K^*$. Then we let $t = t_{1-\alpha_0/(k-1),n-q-1}$ and*

create the subset

$$I_{BestBet} = \left\{ i : \widehat{\mu}_{iK^*}(n) \geq -t\,\widehat{\Delta}_{iK^*}(n)\widehat{\tau}_{iK^*}(n), \quad \forall i \neq K^* \right\} \cup \{\, K^* \,\}.$$

In Step 3, an additive approach is applied $(1-\alpha = 1-\alpha_0-\alpha_1-\alpha_2)$.

Theorem 2 *If Model (1) and Model (4) both hold 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 the paired CV combined procedure selects a system B such that* $\Pr\{B=k\} \geq 1-\alpha$ *whenever* $\mu_k - \mu_{k-1} \geq \delta$.

We prove this PCS guarantee assuming independence among $\{\eta_{ij}, i=1,2,\ldots,k, j=1,2,\ldots,n\}$ in Tsai (2006), however, the experiments showed that this paired CV combined procedure can perform very well even when $\{\eta_{ij}, i=1,2,\ldots,k, j=1,2,\ldots,n\}$ are not independent.

7 EMPIRICAL RESULTS

In this section we summarize the results of an empirical evaluation performed to compare the following procedures:

1. The combined sample-mean-based procedure (NSGS) 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 in the second stage. This procedure allows for unknown and unequal variances across systems, but CRN is not exploited.
2. The individual CV combined procedure which we call TN-I, and the paired CV combined procedure which we call TN-P. These procedures allow for unknown and unequal variances across systems and the use of CRN although TN-I does not exploit CRN.

The systems were 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. Then we assume 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 assumed to be $\mathrm{N}(\xi_i,\sigma_c^2)$ random variables which are independent of $\{\eta_i, i=1,2,\ldots,k\}$. The correlation between controls C_i and C_j for $i \neq j$ is ρ_c. The correlation between residuals η_i and η_j for $i \neq j$ is ρ_η. The squared correlation coefficient between X_i and C_i is $\rho^2_{(x,c)}$.

We evaluated each procedure on different variations of the systems, examining factors including the number of systems k, the practically significant difference δ, the initial sample size n_0, the variances of controls σ_c^2, the variance of residuals σ_η^2, the correlation of the controls ρ_c, and the correlation of residuals ρ_η. The larger σ_c^2 is compared with σ_η^2, the more the variability in outputs can be explained by the controls. Larger ρ_η means more dependence due to CRN is accounted for by the residuals. 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, in which μ_k was set to δ, while $\mu_1 = \mu_2 = \cdots = \mu_{k-1} = 0$. This is a difficult scenario for screening procedures because all the inferior systems are close to the best system. Notice that we do not need to examine more favorable configurations since our goal is to compare NSGS with the TN procedures. The slippage configuration is sufficient for this purpose.

We chose the initial sample size to be $n_0 = 10$, for $i=1,2,\ldots,k$. The mean of the controls, ξ_i, is set to be 0, for $i=1,2,\ldots,k$. We also set β_i 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, δ, was set to $\delta = \sqrt{(\sigma_c^2+\sigma_\eta^2)/n_0}$, where σ_c^2 is the variance of controls and σ_η^2 is the variance of residuals. For each configuration, 500 macroreplications (complete repetitions) of the entire combined procedure were performed. In all experiments, the nominal probability of correct selection was set at $1-\alpha = 0.95$. We took $\alpha_0 = \alpha_1 = \alpha_2 = \alpha/3$ in paired CV screening cases and took $\alpha_0 = \alpha/3$, $\alpha_1 = \alpha_2 = \alpha/(3-\alpha)$ in individual CV screening cases. For NSGS, we set $\alpha_0 = \alpha_1 = \alpha/2$. 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 received second-stage sampling (PSS).

7.2 Summary of Results

The PCS of the CV combined procedure was over 0.95 in all configurations. The overall experiments showed that the CV combined procedure is superior to the combined sample-mean-based procedure under any configuration we examined. The ANS increased much more slowly for the CV combined procedure than for the combined sample-mean-based procedure NSGS as k increased.

7.3 Some Specific Results

We do not try to present comprehensive results from such a large simulation study. Instead, we present selected results

that highlight the key conclusions. Notice that we apply the Best Bet screening procedure in TN-P.

7.3.1 Effect of Number of Systems

See Table 1 for an illustration. Systems are simulated independently since NSGS and TN-I do not exploit CRN. So the key factor is to compare NSGS with TN-I when we have different numbers of systems. As k increases, the average number of samples per system increases greatly in NSGS. However, the ANS increases much slower in TN-I than in NSGS as k increases. The percentage of systems that received second-stage sampling is also smaller in TN-I than in NSGS, which is not surprising. When the number of systems increases, TN-I can get more advantage than NSGS.

7.3.2 Effect of Control Variates

See Table 2 for an illustration. We know that $\rho^2_{(x,c)} = \sigma_c^2/\sigma_x^2 = \sigma_c^2/(\sigma_c^2+\sigma_\eta^2)$ which represents how good this CV is. In our experiments, we fix σ_x^2 to be 16. For example, $\rho^2_{(x,c)}$ = 0.2 means $\sigma_x^2 = 16$ and $\sigma_c^2 = 3.2$. We find that the performance of individual CV combined procedure is almost the same as NSGS when $\rho^2_{(x,c)}$ is 0.2. When $\rho^2_{(x,c)}$ is larger than 0.2, the CV combined procedure can outperform NSGS. Very little $\rho^2_{(x,c)}$ is required for the CV combined procedure to beat NSGS. Certainly, larger $\rho^2_{(x,c)}$ means the CVs can explain more variability of the outputs, and thereby makes the CV combined procedure more efficient.

7.3.3 Effect of Correlation

See Table 3 for an illustration. Here we compare TN-I and TN-P under different ρ_η. The performance of TN-I is not affected by the correlation between residuals. On the other hand, when the correlation between residuals is larger, TN-P performs better and beats TN-I easily. In Table 3, we see that the PSS of TN-P is as low as 0.09 which shows the high efficiency of TN-P when ρ_η is large.

Table 1: Effect of Number of Systems for NSGS and TN-I when $\sigma_c = 4, \sigma_\eta = 1, \rho_c = \rho_\eta = 0$

Number of systems	Procedure	PCS	ANS	PSS
k=2	NSGS	0.98	98	0.86
	TN-I	1	12	0.41
k=5	NSGS	0.98	186	0.96
	TN-I	1	19	0.76
k=10	NSGS	0.98	234	0.97
	TN-I	1	27	0.86
k=25	NSGS	0.98	306	0.99
	TN-I	1	34	0.92
k=100	NSGS	0.99	430	0.99
	TN-I	1	49	0.98

Table 2: Effect of Control Variates for TN-I comparison with NSGS when $\rho_c = 0, \rho_\eta = 0$, and $k = 10$

	Procedure	PCS	ANS	PSS
σ_x^2=16	NSGS	1	235	0.97
$\rho^2_{(x,c)}$=0.2	Individual CV	0.97	241	0.98
$\rho^2_{(x,c)}$=0.4	Individual CV	1	181	0.99
$\rho^2_{(x,c)}$=0.6	Individual CV	1	129	0.97
$\rho^2_{(x,c)}$=0.8	Individual CV	1	68	0.99

Table 3: Effect of Correlation for TN-I and TN-P when $\sigma_c = 4, \sigma_\eta = 1$, and $k = 10$

Correlation	Procedure	PCS	ANS	PSS
ρ_c=0	Individual CV	1	34	0.80
ρ_η=0.2	Paired CV	1	30	0.74
ρ_c=0	Individual CV	1	34	0.90
ρ_η=0.5	Paired CV	1	24	0.53
ρ_c=0	Individual CV	1	35	0.90
ρ_η=0.8	Paired CV	1	12	0.09

8 CONCLUSIONS

In this paper we presented a CV combined procedure which captures the ability to screen out inferior systems and the statistical efficiency of control variates. We also presented a new paired CV screening procedure to reduce the computation cost and retain the benefits of paired CV as well. As we showed in Section 7, TN-I is superior to NSGS for all the scenarios we examined. NSGS is based on the assumption that all systems are simulated independently, and TN-I assumes that the dependence induced by CRN is entirely explained by the controls. On the other hand, TN-P is significantly more efficient than TN-I when the correlation between residuals induced via CRN is large. However, the advantage of TN-P over TN-I is diminishing with larger numbers of systems and TN-P requires more computation cost than TN-I. As a rough rule of thumb, we use TN-P when CRN is involved, but use TN-I when all the systems are simulated independently.

REFERENCES

Bechhofer, R. E., T. J. Santner, and D. Goldsman. 1995. *Design and analysis of experiments for statistical selection, screening and multiple comparisons*. New York: John Wiley.

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

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. 1990. Control-variate remedies. *Operations Research* 38:974–992.

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

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

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.

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

Tsai, S. C. 2006. Ranking and selection using control-variate estimators. Proposal to Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, Illinois.

Wilcox, R. R. 1984. A table for Rinott's selection procedure. *Journal of Quality Technology* 16:97–100.

AUTHOR BIOGRAPHIES

SHING CHIH TSAI is a doctoral student in the Department of Industrial Engineering and Management Sciences at Northwestern University. His research interests are in experiment design and analysis of stochastic simulation experiments. He received his undergraduate degree in Industrial Management from National Taiwan University in 2000. His e-mail address is <s-tsai4@northwestern.edu>.

BARRY L. NELSON is the Charles Deering McCormick Professor of Industrial Engineering and Management Sciences at Northwestern University, and is Director of the Master of Engineering Management Program there. His research centers on the design and analysis of computer simulation experiments on models of stochastic systems. He has published numerous papers and two books. Nelson has served the profession as the Simulation Area Editor of *Operations Research* and President of the INFORMS (then TIMS) College on Simulation and is currently Editor in Chief of *Naval Research Logistics*. He has held many positions for the Winter Simulation Conference, including Program Chair in 1997 and current membership on the Board of Directors. His e-mail and web addresses are <nelsonb@northwestern.edu> and <www.iems.northwestern.edu/~nelsonb/>.