Combined Screening and Selection of the Best with Control Variates

Shing Chih Tsai
Barry L. Nelson
Jeremy Staum

Department of Industrial Engineering & Management Sciences
Northwestern University
2145 Sheridan Road
Evanston, IL 60208-3119

December 13, 2006

Abstract

Nelson and Staum derived ranking-and-selection (R&S) procedures that employ control-variate (CV) estimators instead of sample means to obtain greater statistical efficiency. However, control-variate estimators require more computational effort than sample means, and effective controls must be identified. In this paper, we present a new CV screening procedure to avoid much of the computation cost along with a better paired CV model than Nelson and Staum. 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. Some guidelines about control-variate selection and an empirical evaluation are 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 output variances or the number of systems are 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 outputs are used as the estimators of the expected performance. Nelson and Staum (2006) derived R&S procedures that 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.

One of our goals is to propose a new control variate (CV) screening procedure to decrease the computation cost and still obtain the statistical efficiency. A superior paired CV model is provided and compared to the paired model in Nelson and Staum (2006). We also propose a two-stage procedure which captures the ability to screen out inferior systems and the statistical efficiency of CVs for selection: We use a screening procedure with CVs 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 a smaller sample size than Rinott's (1978) procedure even when the correlation between the output and control is modest. Therefore 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.

The 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 present the improved paired CV model and a new CV screening procedure in Section 4. Section 6 contains some guidance for selecting control variates in this context. In Section 7, we present the CV combined procedure in detail. The paper ends with an empirical evaluation, including a queueing example, performed to compare the two combined procedures (Sections 8 and 9), and conclusions in Section 10. All proofs are relegated to the Appendix.

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 Procedure 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 control variates 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

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

where the $q_i \times 1$ vector $\mathbf{C}_{ij}$ is called the *control* and is assumed multivariate normal, while $\{\eta_{ij}, i = 1, 2, \ldots, k, j = 1, 2, \ldots, n\}$ are mutually independent and $\{\eta_{ij}, j = 1, 2, \ldots, n\}$ is a set of i.i.d. $\mathrm{N}(0, \tau_i^2)$ random variables. For each system $i = 1, 2, \ldots, k$, the controls $\{\mathbf{C}_{ij}, j = 1, 2, \ldots, n\}$ are also i.i.d., are independent of $\{\eta_{ij}, j = 1, 2, \ldots, n\}$ and have known expected value $\boldsymbol{\xi}_i$. The X_{ij} are therefore i.i.d. $\mathrm{N}(\mu_i, \sigma_i^2)$ random variables, with both μ_i and σ_i^2 unknown and (perhaps) unequal. The multiplier $\boldsymbol{\beta}_i$ is a $q_i \times 1$ vector of unknown constants that captures the relationship between the output X_{ij} and the control $\mathbf{C}_{ij}$, while η_{ij} represents that part of the variability in X_{ij} that is not explained by the controls.

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 outputs 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.

To define the CV point estimator, let

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

Then the CV point estimator of μ_i is

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

(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 $\widehat{\mu}_i(n)$ and an estimator of its variance to derive R&S procedures. Let

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

and define

$$\begin{aligned}\widehat{\tau}_i^2(n) &= \frac{1}{n-q_i-1}\mathbf{X}_i(n)'\left[\mathbf{I}-\mathbf{A}_i(n)\left(\mathbf{A}_i'(n)\mathbf{A}_i(n)\right)^{-1}\mathbf{A}_i'(n)\right]\mathbf{X}_i(n)\\ &= \frac{1}{n-q_i-1}\sum_{j=1}^{n}\left[X_{ij}-\widehat{\mu}_i(n)-(\mathbf{C}_{ij}-\boldsymbol{\xi}_i)'\widehat{\boldsymbol{\beta}}_i(n)\right]^2 \qquad (2)\end{aligned}$$

as the residual variance estimator. Further, let

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

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 and Hsu (1993), Theorem 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* $i=1,2,\ldots,k$, *where* $\chi^2_{n-q_i-1}$ *is a chi-squared random variable with* $n-q_i-1$ *degrees of freedom.*

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 but relax the assumption that $\mathbf{C}_{ij}$ has to be multivariate normal. Let $t_{p,\nu}$ represent the p quantile of the t distribution with ν degrees of freedom.

Procedure 1 (Individual CV Screening Procedure).

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

$$I_{\text{Indiv}} = \{i : \widehat{\mu}_i(n_i) - \widehat{\mu}_\ell(n_\ell) \geq -W_{i\ell}, \forall \ell \neq i\}, \tag{4}$$

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 $\Pr\{k \in I_{\text{Indiv}}\} \geq 1 - \alpha$ when Model (1) holds even if we relax the assumption that $\mathbf{C}_{ij}$ is multivariate normal. Nelson and Staum (2006) also showed that very little correlation between the output and control is required for the subset size with CVs to be smaller than that with sample means.

The advantage of this procedure is that we just need to compute k CVs. Its disadvantage is that the assumption in Model (1) that there is no dependence between residuals across systems induced by CRN will not hold in practice. Therefore Nelson and Staum (2006) proposed a screening procedure with paired control variates, which we improve upon in the next section.

4 Screening Procedures with Paired Control Variates

In this section we briefly review the paired control variate model of Nelson and Staum (2006) and propose a more general model on which a new procedure is based. An adjustment is also provided to reduce the computation cost and retain the benefit of paired CV estimators.

Nelson and Staum (2006) use a paired CV model to avoid the assumption that the controls entirely explain the dependence induced by CRN. To do this, they 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 $\boldsymbol{\xi}_{i\ell} = \boldsymbol{\xi}_i - \boldsymbol{\xi}_\ell$, for $i \neq \ell$. Since they need the outputs and the controls to be paired across systems, the number of observations must be equal for each system in the same pair, and the number of controls for each system in the same pair should also be equal. For convenience we let n be the common number of replications and q be the common number of controls for each system. Then they assume that a model like Model (1) holds:

$$X_j(i, \ell) = \mu_{i\ell} + (\mathbf{C}_j(i, \ell) - \boldsymbol{\xi}_{i\ell})'\mathbf{B}(i, \ell) + \varepsilon_j(i, \ell), \tag{5}$$

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 $\boldsymbol{\xi}_{i\ell}$.

Unlike Model (1), Model (5) can hold even when η_{ij} and $\eta_{\ell j}$ are dependent. However, this model may break down when CRN causes $\mathbf{C}_{ij} = \mathbf{C}_{\ell j}$ for all j, which cancels the controls.

To avoid this, we present a different model to explain the relationship between the controls and the outputs. We assume the following new model holds:

$$X_j(i,\ell) = \mu_{i\ell} + (\mathbf{C}_{ij} - \boldsymbol{\xi}_i)'\boldsymbol{\beta}_i - (\mathbf{C}_{\ell j} - \boldsymbol{\xi}_\ell)'\boldsymbol{\beta}_\ell + \eta_j(i,\ell), \tag{6}$$

where $\{\eta_j(i,\ell), j = 1, 2, \ldots, n\}$ is a set of i.i.d. $\mathrm{N}(0, \sigma^2_{i\ell})$ random variables. The $(q_i+q_\ell)\times 1$ vector $(\mathbf{C}'_{ij}, \mathbf{C}'_{\ell j})'$ is assumed multivariate normal. For each pair of systems $i, \ell = 1, 2, \ldots, k, i \neq \ell$ the controls $\{\mathbf{C}_{ij}, j = 1, 2, \ldots, n\}$ and $\{\mathbf{C}_{\ell j}, j = 1, 2, \ldots, n\}$ are also i.i.d., are independent of $\{\eta_j(i,\ell), j = 1, 2, \ldots, n\}$ and have known expected values $\boldsymbol{\xi}_i$ and $\boldsymbol{\xi}_\ell$. Like Model (5), Model (6) can also 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 $\mu_{i\ell}$ under Model (6), and define $\widehat{\tau}^2_{i\ell}(n)$ and $\widehat{\Delta}^2_{i\ell}(n)$ in analogy to Equations (2) and (3).

We now assume that Model (6) holds to execute the all-pair screening procedure. Before describing the procedure, we present an argument for the superiority of Model (6) over Model (5):

- Model (6) is a more general model which is equivalent to Model (5) when $\boldsymbol{\beta}_i = \boldsymbol{\beta}_\ell$.
- Suppose Model (5) holds, but we compute $\widehat{\mu}_{i\ell}(n)$ assuming Model (6) holds. Then $\widehat{\mu}_{i\ell}(n)$ is still unbiased. However, $\mathrm{Var}[\widehat{\mu}_{i\ell}(n)]$ will be inflated because of the loss of degrees of freedom (from $n-q-1$ to $n-2q-1$). The inflation of variance will not be substantial when n is not too small.
- Suppose Model (6) holds with $\boldsymbol{\beta}_i \neq \boldsymbol{\beta}_\ell$, but we compute $\widehat{\mu}_{i\ell}(n)$ assuming Model (5) holds. Then $\widehat{\mu}_{i\ell}(n)$ will be biased and $\mathrm{Var}[\widehat{\mu}_{i\ell}(n)]$ will be increased especially when $\boldsymbol{\beta}_i$ is very different from $\boldsymbol{\beta}_\ell$ (see the Appendix).
- Under Model (6) the number of controls for each system in the same pair is not required to be equal. Therefore, we gain potential benefits in terms of CV selection (notice that the degrees of freedom is $n - q_i - q_\ell - 1$ in general).
- Model (6) makes the all-pair screening procedure below more compatible with the CV selection procedure (see Section 5) which relies on Model (1). In fact Model (1) implies Model (6). Therefore we do not have to be concerned about any incongruity in the CV combined procedure.

We form the following all-pair screening procedure based on Model (6).

Procedure 2 (All-Pair Screening Procedure).

1. Choose the confidence level $1 - \alpha > 1/k$.
2. Obtain $n > \max_{i\neq\ell}(q_i + q_\ell + 2)$ 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 $t_{i\ell} = t_{1-\alpha/(k-1),n-q_i-q_\ell-1}$ and create the subset

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

Nelson and Staum (2006) proved that $\Pr\{k \in I_{\text{AllPair}}\} \geq 1-\alpha$ when Model (5) holds. This procedure is also valid when Model (6) pertains and $(\mathbf{C}'_{ij}, \mathbf{C}'_{\ell j})'$ is not required to be multivariate normal. The advantage of this procedure is that we do not have to be concerned about the dependence remaining in the 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. Therefore, we propose a new procedure which requires less computation and creates a subset $I \supseteq I_{\text{AllPair}}$, 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 use Model (6) and denote the system with the largest $\widehat{\mu}_i(n)$ as K^*.

Procedure 3 (Best Bet Screening Procedure).

1. Choose the confidence level $1-\alpha > 1/k$.
2. Obtain $n > \max_{i\neq\ell}(q_i + q_\ell + 2)$ 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, $K^* = \underset{i}{\operatorname{argmax}}\, \widehat{\mu}_i(n)$, and then form the $k-1$ paired CV estimators $\widehat{\mu}_{iK^*}(n)$ for all $i \neq K^*$.
4. Let $t_{iK^*} = t_{1-\alpha/(k-1),n-q_i-q_{K^*}-1}$ and create the subset

$$I_{\text{Best Bet}} = \left\{ i : \widehat{\mu}_{iK^*}(n) \geq -t_{iK^*}\, \widehat{\Delta}_{iK^*}(n)\widehat{\tau}_{iK^*}(n) \right\} \cup \{\ K^*\}.$$

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 $\widehat{\mu}_i(n) - \widehat{\mu}_\ell(n)$. The 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 the all-pair screening procedure ($k(k-1)/2$), when the number of alternatives is large, and it avoids the assumption that CVs explain all the dependence induced by CRN.

Remark 4.1. The system with the largest sample mean is also a potential best system, so we could let $K^* = \underset{i}{\operatorname{argmax}}\, \bar{X}_i(n)$, and then do screening with paired CVs just against K^*. This procedure can save a lot of computation cost because we only need to compute $k-1$ paired CV estimators. Unfortunately, the subset formed by this procedure may be much larger than that formed by the all-pair screening procedure, because there is not much correlation between $\widehat{\mu}_{i\ell}(n)$ and $\bar{X}_i(n) - \bar{X}_\ell(n)$.

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). The following description is based on 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 $n_0 > q + 2$ observations and calculate $\widehat{\tau}_i^2(n_0)$.

3. For each system $i = 1, 2, \ldots, k$, set the total sample size
$$N_i = \min_{n \geq n_0} \left\{ n : \left(\frac{n-q}{q}\right)\left(\frac{n\delta^2}{h^2 \widehat{\tau}_i^2(n_0)} - 1\right) \geq \mathcal{F}_{q,n-q}^{(\gamma)} \right\}$$
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
$$\gamma = \begin{cases} (1-\alpha_0)^{\frac{1}{k}}, & \text{if the systems are simulated independently} \\ 1 - \alpha_0/k, & \text{otherwise.} \end{cases}$$

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
$$\mu_i - \max_{\ell \neq i} \mu_\ell \in \left[-\left(\widehat{\mu}_i(N_i) - \max_{\ell \neq i} \widehat{\mu}_\ell(N_\ell) - \delta\right)^-, \left(\widehat{\mu}_i(N_i) - \max_{\ell \neq i} \widehat{\mu}_\ell(N_\ell) + \delta\right)^+ \right] \quad (7)$$
for $i = 1, 2, \ldots, k$. Furthermore,
$$\Pr\left\{ \mu_B - \max_{\ell=1,\ldots,k} \mu_\ell \geq -\delta \right\} \geq 1 - \alpha, \quad (8)$$
that is, with high confidence, the mean of the selected system is within δ of the mean of the truly best system.

Nelson and Staum (2006) proved that $\Pr\{B = k\} \geq 1-\alpha$ whenever Model (1) holds and $\mu_k - \mu_{k-1} \geq \delta$. Regardless of the configuration of the true means, the confidence intervals (7) have coverage probability at least $1-\alpha$ by Theorem 1 of Nelson and Matejcik (1995), while Inequality (8) follows from Corollary 1 of Nelson and Goldsman (2001).

6 Control-Variate Selection for Screening

In this section we provide some guidance for selecting control variates. When CRN is involved in the screening procedure, can we take advantage of CRN when choosing control variates, or should we just select favorable control variates to minimize the variance of each CV point estimator individually? For the screening procedure with individual control variates, we use the screening threshold $W_{i\ell}$ in Equation (4) as the measure of chosen subset size. The smaller $W_{i\ell}$ is, the more difficult it is for system i to survive in the subset. To make the expected subset size as small as possible, we select favorable control variates to minimize $\mathrm{E}[W_{i\ell}^2]$ because $W_{i\ell}$ is nonnegative. To simplify the analysis, suppose we choose first-stage sample sizes and controls for each system such that $n_i = n$ and $q_i = q, \forall i$. Consequently, for all $i = 1, 2, \ldots, k$, $t_i = t$, and we know that

$$\begin{aligned}
\mathrm{E}[W_{i\ell}^2] &= \mathrm{E}\left[t_i^2\,\widehat{\Delta}_i^2(n_i)\widehat{\tau}_i^2(n_i) + t_\ell^2\,\widehat{\Delta}_\ell^2(n_\ell)\widehat{\tau}_\ell^2(n_\ell)\right] \\
&= \mathrm{E}\left[t^2\,\widehat{\Delta}_i^2(n)\widehat{\tau}_i^2(n)\right] + \mathrm{E}\left[t^2\,\widehat{\Delta}_\ell^2(n)\widehat{\tau}_\ell^2(n)\right] \\
&= t^2\,\mathrm{Var}[\widehat{\mu}_i(n)] + t^2\,\mathrm{Var}[\widehat{\mu}_\ell(n)] \\
&= t^2\,\left(\mathrm{Var}[\widehat{\mu}_i(n)] + \mathrm{Var}[\widehat{\mu}_\ell(n)]\right). \\
&= t^2\,\left(\frac{n-2}{n-q-2}\right)\left(\frac{\mathrm{Var}[\eta_{ij}] + \mathrm{Var}[\eta_{\ell j}]}{n}\right).
\end{aligned}$$

Clearly we would like to minimize the variance of each CV point estimator. In other words, we should choose control variates for each system to obtain the greatest variance reduction individually. Notice that $E[W_{i\ell}^2]$ is unaffected by CRN, therefore CRN is irrelevant with respect to the individual CV screening procedure. Añonuevo and Nelson (1988), Nelson (1989) and Bauer and Wilson (1992) give some algorithms for selecting good control variates individually. However, CRN can affect $\mathrm{Cov}[\eta_{ij}, \eta_{\ell j}]$ which represents the benefits of CRN that we cannot capture in the individual CV screening procedure. This is the disadvantage inherent in the screening procedure with individual control variates.

For the screening procedure with paired control variates under Model (6), the expectation of the screening threshold from Step 3 of Procedure 2 is

$$\begin{aligned}
\mathrm{E}\left[t_{i\ell}^2\,\widehat{\Delta}_{i\ell}(n)^2\widehat{\tau}_{i\ell}(n)^2\right] &= t_{i\ell}^2\,\mathrm{Var}[\widehat{\mu}_{i\ell}(n)] \\
&= t_{i\ell}^2\,\left(\frac{n-2}{n-2q-2}\right)\frac{\mathrm{Var}[\eta_j(i,\ell)]}{n} \\
&= t_{i\ell}^2\,\left(\frac{n-2}{n-2q-2}\right)\frac{\mathrm{Var}[\eta_{ij} - \eta_{\ell j}]}{n} \\
&= t_{i\ell}^2\,\left(\frac{n-2}{n-2q-2}\right)\left(\frac{\mathrm{Var}[\eta_{ij}] + \mathrm{Var}[\eta_{\ell j}] - 2\mathrm{Cov}[\eta_{ij}, \eta_{\ell j}]}{n}\right)
\end{aligned}$$

which directly incorporates the reduced variance by applying CVs to the paired observations. The more positive correlation that remains in the residuals across systems induced by CRN,

the larger $\text{Cov}[\eta_{ij}, \eta_{\ell j}]$ will be. The paired CV procedure exploits this dependence so that it can perform better than the individual CV screening procedure especially when n is much larger than q. Therefore, we should choose control variates that can be paired across systems i and ℓ to minimize the variance of $\widehat{\mu}_{i\ell}(n)$, and these may be different than we would choose to minimize the variance of each individual CV estimator. As a result it is possible that the CVs chosen for system i could be different when it is paired with each system $\ell = 1, 2, \ldots, k, \ell \neq i$.

7 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

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

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

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

- 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 or with CRN.

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 7.1. Following is a procedure which combines the individual CV screening procedure with the CV selection-of-the-best procedure.

Procedure 5 (Individual CV Combined Procedure).

1. Select overall confidence level $1-\alpha > 1/k$, 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

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

where

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

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

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

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

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

Notice: $1 - \alpha = (1 - \alpha_0)(1 - \alpha_1 - \alpha_2)$ (*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 $B = \text{argmax}_i \, \widehat{\mu}_i(N_i)$ as best from all systems $i \in I$.

Theorem 1. *If Model (1) holds, then the individual CV combined procedure selects a system B such that $\Pr\{B = k\} \geq 1 - \alpha$ whenever $\mu_k - \mu_{k-1} \geq \delta$. For any configuration of the means, the following hold with probability greater than or equal to $1 - \alpha$:*

- *For all $i \in I$,*

$$\mu_i - \max_{\ell \in I, \ell \neq i} \mu_\ell \in \left[- \left(\widehat{\mu}_i(N_i) - \max_{\ell \in I, \ell \neq i} \widehat{\mu}_\ell(N_\ell) - \delta \right)^- , \left(\widehat{\mu}_i(N_i) - \max_{\ell \in I, \ell \neq i} \widehat{\mu}_\ell(N_\ell) + \delta \right)^+ \right]. \tag{9}$$

- *The mean of the system we select will be within δ of the mean of the truly best system in I with probability $\geq 1-\alpha$, that is*

$$\Pr\left\{\mu_B - \max_{\ell \in I, \ell \neq B} \mu_\ell \geq -\delta\right\} \geq 1-\alpha. \tag{10}$$

Remark 7.1. 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 7.2. We can combine the paired CV screening procedure with the CV selection-of-the-best procedure. When we use the best bet screening procedure, we need to change Step 2 to the following:

2. Obtain $n > \max_{i \neq \ell}(q_i + q_\ell + 2)$ 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, $K^* = \operatorname{argmax}_i \widehat{\mu}_i(n)$, and then form the $k-1$ paired CV estimators $\widehat{\mu}_{iK^*}(n)$ for all $i \neq K^*$. Then we let $t_{iK^*} = t_{1-\alpha_0/(k-1), n-q_i-q_{K^*}-1}$ and create the subset

$$I_{\text{Best Bet}} = \left\{i : \widehat{\mu}_{iK^*}(n) \geq -t_{iK^*}\, \widehat{\Delta}_{iK^*}(n) \widehat{\tau}_{iK^*}(n)\right\} \cup \{\ K^*\}\,.$$

An additive approach is applied ($1 - \alpha = 1 - \alpha_0 - \alpha_1 - \alpha_2$).

Theorem 2. *If Model (1) holds, then the paired CV combined procedure selects a system B such that $\Pr\{B = k\} \geq 1 - \alpha$ whenever $\mu_k - \mu_{k-1} \geq \delta$, and statistical inferences (9) and (10) still hold regardless of the configuration of the true means.*

We prove that $\Pr\{B = k\} \geq 1 - \alpha$ with independence among $\{\eta_{ij}, i = 1, 2, \ldots, k, j = 1, 2, \ldots, n\}$ in the Appendix. However, experiments showed that this paired CV combined procedure can perform very well even when $\{\eta_{ij}, i = 1, 2, \ldots, k\}$ are positively dependent.

8 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 TNS-I, and the paired CV combined procedure which we call TNS-P. These procedures allow for unknown and unequal variances across systems and the use of CRN although TNS-I does not exploit CRN.

The systems are represented by various configurations of k normal distributions; in all cases, system k was the best (had the largest true mean). Let X_i be a simulation observation from system i, for $i = 1, 2, \ldots, k$. For simplicity, we assume that there is $q = 1$ control variate. 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 and independent of $\{\eta_i, i = 1, 2, \ldots, k\}$. The correlation between controls C_i and C_ℓ for $i \neq \ell$ is ρ_c. The correlation between residuals η_i and η_ℓ for $i \neq \ell$ is ρ_η. The squared correlation coefficient between X_i and C_i is $\rho^2_{(x,c)}$.

We evaluated each procedure on different configurations of the systems, examining factors including the number of systems k, the practically significant difference δ, the initial sample size n_0, the variance 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 of the variability in outputs can be explained by the controls. When $\rho_\eta \neq 0$ then Model (6) holds but Model (1) does not hold. 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.

8.1 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. These experiments with the slippage configuration illustrated that CVs can make the screening procedure more efficient even under the most difficult situation.

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).

Table 1: Effect of Number of Systems for NSGS and TNS-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
	TNS-I	1	12	0.41
k=5	NSGS	0.98	186	0.96
	TNS-I	1	19	0.76
k=10	NSGS	0.98	234	0.97
	TNS-I	1	27	0.86
k=25	NSGS	0.98	306	0.99
	TNS-I	1	34	0.92
k=100	NSGS	0.99	430	0.99
	TNS-I	1	49	0.98

8.2 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 was superior to the combined sample-mean-based procedure under any configuration we examined.

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 Model (6) and the best bet screening procedure in TNS-P.

8.2.1 Effect of Number of Systems

See Table 1 for an illustration. Systems are simulated independently since NSGS and TNS-I do not exploit CRN. The goal is to compare NSGS with TNS-I when we have different numbers of systems. As k increases, the average number of samples per system increases greatly in NSGS compared to TNS-I. The percentage of systems that received second-stage sampling is smaller in TNS-I than in NSGS.

8.2.2 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 the 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. Thus, very little $\rho^2_{(x,c)}$ is required for the CV

Table 2: Effect of Control Variates for TNS-I in 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
$\rho_{(x,c)}^2$=0.2	Individual CV	0.97	241	0.98
$\rho_{(x,c)}^2$=0.4	Individual CV	1	181	0.99
$\rho_{(x,c)}^2$=0.6	Individual CV	1	129	0.97
$\rho_{(x,c)}^2$=0.8	Individual CV	1	68	0.99

Table 3: Effect of Correlation for TNS-I and TNS-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	26	0.55
ρ_c=0	Individual CV	1	35	0.90
ρ_η=0.8	Paired CV	1	13	0.10

combined procedure to beat NSGS. Larger $\rho_{(x,c)}^2$ means the CVs can explain more variability of the outputs, and thereby makes the CV combined procedure more efficient.

8.2.3 Effect of Correlation

See Table 3 for an illustration. Here we compare TNS-I and TNS-P under different ρ_η. When the correlation between residuals is larger, TNS-P performs better and beats TNS-I easily. In Table 3, we see that the PSS of TNS-P is as low as 0.10 which shows the high efficiency of TNS-P when ρ_η is large. Notice that CRN does not affect the screening threshold for TNS-I, but it does affect the point estimator, which is why performance of TNS-I in Table 3 varies when we have different ρ_η.

9 Illustration

In this section we use a queueing example to illustrate the application of TNS-I, TNS-P and NSGS. We use the $M/M/s/c$ model which represents a queueing system with Poisson arrivals, exponentially distributed service times, s servers, a capacity of c customers, and first-come, first-served queueing discipline. The customers arrive with arrival rate λ. The service rate for an individual server is μ. We perform each procedure on five different

Table 4: The Five Queueing Systems and their Expected Waiting Times in Steady State

System i	λ	s	μ	$E[W]$
1	4	1	5	0.88
2	4	2	5/2	0.98
3	4	3	5/3	1.10
4	4	4	5/4	1.24
5	4	5	1	1.38

Table 5: Results for NSGS, TNS-I and TNS-P in 100 Trials with $\delta = 0.1$, $n_0 = 10$, and $1 - \alpha = 0.95$

Procedure	PCS	ANS	$\widehat{\text{se}}$(ANS)	PSS
NSGS	0.97	462	13.4	1
TNS-I	1	301	10.8	0.81
TNS-P with CRN	1	207	10.4	0.68

configurations of the systems in which $\lambda/s\mu = 4/5$ where the performance measure is the steady-state mean of the waiting time in system. The capacity c is set as 15. The five configurations are shown in Table 4 along with their true expected waiting times. System 1 is obviously the best system.

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

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

where W_{ijm} is the waiting time in system for customer m of replication j from system i and S_{ijm} is the service time for customer m of replication j from system i. The initial sample size n_0 is set as 10 for each system. We choose the indifference zone δ to be 0.1 and CRN is applied.

Table 5 shows the results of the TNS-I, TNS-P and NSGS with 100 macroreplications and confidence level $1 - \alpha = 0.95$. We also provide the estimated standard error of ANS to show that there is a significant difference.

These three procedures all exceed the desired probability of correct selection. NSGS is unable to screen out inferior systems in the first stage, therefore its ANS is much larger than the other procedures. We can eliminate more systems in TNS-P than in TNS-I to further

reduce the average number of samples needed by using CVs with CRN.

10 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 proposed a more general paired CV model and a new paired CV screening procedure to reduce the computation cost and retain the benefits of paired CVs as well. As we showed in Sections 8 and 9, TNS-I is superior to NSGS for all the scenarios we examined. NSGS is based on the assumption that all systems are simulated independently, and TNS-I assumes that the dependence induced by CRN is entirely explained by the controls. On the other hand, TNS-P is significantly more efficient than TNS-I when the CVs do not explain all dependence due to CRN. However, computational experiments show that the advantage of TNS-P over TNS-I is diminishing with larger numbers of systems and TNS-P requires more computation cost than TNS-I. As a rough rule of thumb, we use TNS-P when CRN is involved, but use TNS-I when all the systems are simulated independently.

A Appendix

For the individual and paired CV combined procedures, the proofs in this appendix assume that Model (1) holds. We also assume $\mu_k - \mu_{k-1} \geq \delta$.

A.1 Model (5) and Model (6)

Here we compare Model (5; Nelson and Staum paired CV model) and Model (6; our new paired CV model) in terms of $\mathrm{E}[\widehat{\mu}_{i\ell}(n)]$ and $\mathrm{Var}[\widehat{\mu}_{i\ell}(n)]$. We know that Model (6) tends not to lose much when the true underlying model is Model (5) because $\widehat{\mu}_{i\ell}(n)$ is still unbiased and the inflation of $\mathrm{Var}[\widehat{\mu}_{i\ell}(n)]$, due to the loss of degrees of freedom from $n-q-1$ to $n-2q-1$, will not be substantial when n is not too small. Therefore, we focus on the consequences of assuming that Model (5) holds when in fact the true model is Model (6).

We first compute $\widehat{\mu}_{i\ell}(n)$ under the assumption that Model (5) holds while Model (6) is actually the true model with $\boldsymbol{\beta}_i \neq \boldsymbol{\beta}_\ell$. The case $q_i = 1,\ \forall i$ is sufficient to illustrate the point. For convenience, let

$$\mathbf{C} = \{C_{ij}, j = 1, 2, \ldots, n, i = 1, 2, \ldots, k\}$$

be the collection of all observed controls.

We know that $\widehat{\mu}_{i\ell}(n) = \bar{X}(i,\ell) - (\bar{C}(i,\ell) - \xi_{i\ell})\widehat{\mathrm{B}}(i,\ell)$, and $\mathrm{E}[\widehat{\mu}_{i\ell}(n)|\mathbf{C}] = \mathrm{E}[\bar{X}(i,\ell)|\mathbf{C}] - (\bar{C}(i,\ell) - \xi_{i\ell})\mathrm{E}[\widehat{\mathrm{B}}(i,\ell)|\mathbf{C}]$. Therefore,

$$\mathrm{E}[\widehat{\mu}_{i\ell}(n)] = \mathrm{E}\left[\mathrm{E}[\widehat{\mu}_{i\ell}(n)|\mathbf{C}]\right] = \mu_{i\ell} - \mathrm{E}\left[(\bar{C}(i,\ell) - \xi_{i\ell})\mathrm{E}[\widehat{\mathrm{B}}(i,\ell)|\mathbf{C}]\right]$$

where

$$\widehat{\mathrm{B}}(i,\ell) = \frac{\sum_{j=1}^{n}(X_j(i,\ell) - \bar{X}(i,\ell))(C_j(i,\ell) - \bar{C}(i,\ell))}{\sum_{j=1}^{n}(C_j(i,\ell) - \bar{C}(i,\ell))^2}$$

so that

$$\mathrm{E}[\widehat{\mathrm{B}}(i,\ell)|\mathbf{C}] = \frac{\sum_{j=1}^{n}\mathrm{E}[X_j(i,\ell) - \bar{X}(i,\ell)|\mathbf{C}](C_j(i,\ell) - \bar{C}(i,\ell))}{\sum_{j=1}^{n}(C_j(i,\ell) - \bar{C}(i,\ell))^2}.$$

Because Model (6) holds, we know that

$$\mathrm{E}[X_j(i,\ell) - \bar{X}(i,\ell)|\mathbf{C}] = (C_{ij} - \xi_i)\beta_i - (C_{\ell j} - \xi_\ell)\beta_\ell - (\bar{C}_i - \xi_i)\beta_i + (\bar{C}_\ell - \xi_\ell)\beta_\ell = (C_{ij} - \bar{C}_i)\beta_i - (C_{\ell j} - \bar{C}_\ell)\beta_\ell.$$

Thus, we obtain

$$\begin{aligned}
\mathrm{E}[\widehat{\mathrm{B}}(i,\ell)|\mathbf{C}] &= \frac{\sum_{j=1}^{n}\left((C_{ij} - \bar{C}_i)\beta_i - (C_{\ell j} - \bar{C}_\ell)\beta_\ell\right)(C_j(i,\ell) - \bar{C}(i,\ell))}{\sum_{j=1}^{n}(C_j(i,\ell) - \bar{C}(i,\ell))^2} \\
&= \frac{\sum_{j=1}^{n}\left(\beta_i(C_{ij} - C_{\ell j} - \bar{C}_i + \bar{C}_\ell) + (\beta_i - \beta_\ell)(C_{\ell j} - \bar{C}_\ell)\right)(C_j(i,\ell) - \bar{C}(i,\ell))}{\sum_{j=1}^{n}(C_j(i,\ell) - \bar{C}(i,\ell))^2} \\
&= \beta_i + (\beta_i - \beta_\ell)\frac{\sum_{j=1}^{n}\left(C_{\ell j} - \bar{C}_\ell\right)(C_j(i,\ell) - \bar{C}(i,\ell))}{\sum_{j=1}^{n}(C_j(i,\ell) - \bar{C}(i,\ell))^2}. \qquad (11)
\end{aligned}$$

Therefore, the bias is

$$\begin{aligned}
\mathrm{E}[\widehat{\mu}_{i\ell}(n)] - \mu_{i\ell} &= -\mathrm{E}\left[(\bar{C}(i,\ell) - \xi_{i\ell})\mathrm{E}[\widehat{\mathrm{B}}(i,\ell)|\mathbf{C}]\right] \\
&= -\mathrm{E}\left[\left((\beta_i - \beta_\ell)\frac{\sum_{j=1}^{n}\left(C_{\ell j} - \bar{C}_\ell\right)(C_j(i,\ell) - \bar{C}(i,\ell))}{\sum_{j=1}^{n}(C_j(i,\ell) - \bar{C}(i,\ell))^2}\right)\left(\bar{C}(i,\ell) - \xi_{i\ell}\right)\right] \qquad (12)
\end{aligned}$$

which is not equal to 0 in general if $\beta_i \neq \beta_\ell$.

We now examine the impact on variance. Notice that

$$\begin{aligned}
\mathrm{Var}[\widehat{\mu}_{i\ell}(n)] &= \mathrm{Var}\left[\mathrm{E}\left[\widehat{\mu}_{i\ell}(n)|\mathbf{C}\right]\right] + \mathrm{E}\left[\mathrm{Var}[\widehat{\mu}_{i\ell}(n)|\mathbf{C}]\right] \\
&= \mathrm{Var}\left[(\bar{C}_i - \xi_i)\beta_i - (\bar{C}_\ell - \xi_\ell)\beta_\ell - (\bar{C}(i,\ell) - \xi_{i\ell})\mathrm{E}[\widehat{\mathrm{B}}(i,\ell)|\mathbf{C}]\right] + \mathrm{E}\left[\mathrm{Var}[\widehat{\mu}_{i\ell}(n)|\mathbf{C}]\right].
\end{aligned}$$

Since

$$\widehat{\mathrm{B}}(i,\ell) = \frac{\sum_{j=1}^{n}(X_j(i,\ell) - \bar{X}(i,\ell))(C_j(i,\ell) - \bar{C}(i,\ell))}{\sum_{j=1}^{n}(C_j(i,\ell) - \bar{C}(i,\ell))^2} = \frac{\sum_{j=1}^{n}X_j(i,\ell)(C_j(i,\ell) - \bar{C}(i,\ell))}{\sum_{j=1}^{n}(C_j(i,\ell) - \bar{C}(i,\ell))^2},$$

we have

$$\mathrm{Var}[\widehat{\mathrm{B}}(i,\ell)|\mathbf{C}] = \mathrm{Var}[X_j(i,\ell)|\mathbf{C}]\frac{\sum_{j=1}^{n}(C_j(i,\ell) - \bar{C}(i,\ell))^2}{\left(\sum_{j=1}^{n}(C_j(i,\ell) - \bar{C}(i,\ell))^2\right)^2} = \frac{\mathrm{Var}[X_j(i,\ell)|\mathbf{C}]}{\sum_{j=1}^{n}(C_j(i,\ell) - \bar{C}(i,\ell))^2}.$$

Now

$$\begin{aligned}
\mathrm{Var}[\widehat{\mu}_{i\ell}(n)|\mathbf{C}] &= \mathrm{Var}[\bar{X}(i,\ell) - (\bar{C}(i,\ell) - \xi_{i\ell})\widehat{\mathrm{B}}(i,\ell)|\mathbf{C}] \\
&= \mathrm{Var}[\bar{X}(i,\ell)|\mathbf{C}] + (\bar{C}(i,\ell) - \xi_{i\ell})^2\mathrm{Var}[\widehat{\mathrm{B}}(i,\ell)|\mathbf{C}] - 2(\bar{C}(i,\ell) - \xi_{i\ell})\mathrm{Cov}[\bar{X}(i,\ell), \widehat{\mathrm{B}}(i,\ell)|\mathbf{C}] \\
&= \mathrm{Var}[X_j(i,\ell)|\mathbf{C}]\left(\frac{1}{n} + \frac{(\bar{C}(i,\ell) - \xi_{i\ell})^2}{\sum_{j=1}^n (C_j(i,\ell) - \bar{C}(i,\ell))^2}\right) \\
&\quad -2(\bar{C}(i,\ell) - \xi_{i\ell})\mathrm{Cov}[\bar{X}(i,\ell), \widehat{\mathrm{B}}(i,\ell)|\mathbf{C}].
\end{aligned}$$

And,

$$\begin{aligned}
\mathrm{Cov}[\bar{X}(i,\ell), \widehat{\mathrm{B}}(i,\ell)|\mathbf{C}] &= \mathrm{Cov}[\mu_i - \mu_\ell + (\bar{C}_i - \xi_i)\beta_i - (\bar{C}_\ell - \xi_\ell)\beta_\ell + \bar{\eta}(i,\ell), \widehat{\mathrm{B}}(i,\ell)|\mathbf{C}] \\
&= \mathrm{Cov}[\bar{\eta}(i,\ell), \widehat{\mathrm{B}}(i,\ell)|\mathbf{C}].
\end{aligned}$$

Now since

$$\begin{aligned}
\widehat{\mathrm{B}}(i,\ell) &= \frac{\sum_{j=1}^n \left(\left(C_{ij} - \bar{C}_i\right)\beta_i - \left(C_{\ell j} - \bar{C}_\ell\right)\beta_\ell + \eta_j(i,\ell) - \bar{\eta}(i,\ell)\right)(C_j(i,\ell) - \bar{C}(i,\ell))}{\sum_{j=1}^n (C_j(i,\ell) - \bar{C}(i,\ell))^2} \\
&= \beta_i + (\beta_i - \beta_\ell)\frac{\sum_{j=1}^n \left(C_{\ell j} - \bar{C}_\ell\right)(C_j(i,\ell) - \bar{C}(i,\ell))}{\sum_{j=1}^n (C_j(i,\ell) - \bar{C}(i,\ell))^2} \\
&\quad + \frac{\sum_{j=1}^n (\eta_j(i,\ell) - \bar{\eta}(i,\ell))(C_j(i,\ell) - \bar{C}(i,\ell))}{\sum_{j=1}^n (C_j(i,\ell) - \bar{C}(i,\ell))^2}
\end{aligned}$$

we have

$$\begin{aligned}
\mathrm{Cov}[\bar{\eta}(i,\ell),\widehat{\mathrm{B}}(i,\ell)|\mathbf{C}] &= \mathrm{Cov}\left[\bar{\eta}(i,\ell),\frac{\sum_{j=1}^{n}\left(\eta_j(i,\ell)-\bar{\eta}(i,\ell)\right)\left(C_j(i,\ell)-\bar{C}(i,\ell)\right)}{\sum_{j=1}^{n}(C_j(i,\ell)-\bar{C}(i,\ell))^2}\Big|\mathbf{C}\right] \\
&= \frac{1}{\sum_{j=1}^{n}(C_j(i,\ell)-\bar{C}(i,\ell))^2}\mathrm{Cov}\left[\bar{\eta}(i,\ell),\sum_{j=1}^{n}\left(\eta_j(i,\ell)-\bar{\eta}(i,\ell)\right)C_j(i,\ell)\Big|\mathbf{C}\right] \\
&= \frac{1}{\sum_{j=1}^{n}(C_j(i,\ell)-\bar{C}(i,\ell))^2}\left(\mathrm{Cov}\left[\bar{\eta}(i,\ell),\sum_{j=1}^{n}\eta_j(i,\ell)C_j(i,\ell)\Big|\mathbf{C}\right]\right. \\
&\quad \left.-\ \mathrm{Cov}\left[\bar{\eta}(i,\ell),\bar{\eta}(i,\ell)\sum_{j=1}^{n}C_j(i,\ell)\Big|\mathbf{C}\right]\right) \\
&= \frac{1}{\sum_{j=1}^{n}(C_j(i,\ell)-\bar{C}(i,\ell))^2}\left(\sum_{j=1}^{n}\frac{C_j(i,\ell)}{n}\mathrm{Var}\left[\eta_j(i,\ell)\right]-\sum_{j=1}^{n}C_j(i,\ell)\mathrm{Var}\left[\bar{\eta}(i,\ell)\right]\right) \\
&\quad \Big(\text{Notice}:\mathrm{Var}\left[\eta_j(i,\ell)\right]=\sigma^2_{i\ell},\forall j\Big) \\
&= \frac{1}{\sum_{j=1}^{n}(C_j(i,\ell)-\bar{C}(i,\ell))^2}\left(\sum_{j=1}^{n}\frac{C_j(i,\ell)}{n}\sigma^2_{i\ell}-\sum_{j=1}^{n}C_j(i,\ell)\frac{\sigma^2_{i\ell}}{n}\right) \\
&= 0.
\end{aligned}$$

Then we take the expectation of $\mathrm{Var}[\widehat{\mu}_{i\ell}(n)|\mathbf{C}]$ to yield

$$\mathrm{E}\left[\mathrm{Var}[\widehat{\mu}_{i\ell}(n)|\mathbf{C}]\right]=\mathrm{Var}[\eta_j(i,\ell)]\left(\frac{n-2}{n(n-3)}\right)$$

(Lavenberg and Welch 1981). Further, from Equation (11)

$$\begin{aligned}
&\mathrm{Var}\left[(\bar{C}_i-\xi_i)\beta_i-(\bar{C}_\ell-\xi_\ell)\beta_\ell-(\bar{C}(i,\ell)-\xi_{i\ell})\mathrm{E}[\widehat{\mathrm{B}}(i,\ell)|\mathbf{C}]\right] \\
&= \mathrm{Var}\Bigg[(\bar{C}_i-\xi_i)\beta_i-(\bar{C}_\ell-\xi_\ell)\beta_\ell-\left(\bar{C}(i,\ell)-\xi_{i\ell}\right)\beta_i \\
&\qquad -(\beta_i-\beta_\ell)\left(\bar{C}(i,\ell)-\xi_{i\ell}\right)\frac{\sum_{j=1}^{n}\left(C_{\ell j}-\bar{C}_\ell\right)\left(C_j(i,\ell)-\bar{C}(i,\ell)\right)}{\sum_{j=1}^{n}(C_j(i,\ell)-\bar{C}(i,\ell))^2}\Bigg].
\end{aligned}$$

To simplify it, let

$$\left(\bar{C}(i,\ell)-\xi_{i\ell}\right)\frac{\sum_{j=1}^{n}\left(C_{\ell j}-\bar{C}_\ell\right)\left(C_j(i,\ell)-\bar{C}(i,\ell)\right)}{\sum_{j=1}^{n}(C_j(i,\ell)-\bar{C}(i,\ell))^2}=\Lambda(\mathbf{C}),$$

which is a function of $\mathbf{C}$. Then,

$$\begin{aligned}
\mathrm{Var}\left[\mathrm{E}\left[\widehat{\mu}_{i\ell}(n)|\mathbf{C}\right]\right] &= \mathrm{Var}\left[(\bar{C}_i-\xi_i)\beta_i-(\bar{C}_\ell-\xi_\ell)\beta_\ell-(\bar{C}(i,\ell)-\xi_{i\ell})\mathrm{E}[\widehat{\mathrm{B}}(i,\ell)|\mathbf{C}]\right] \\
&= \mathrm{Var}\left[(\bar{C}_i-\xi_i)\beta_i-(\bar{C}_\ell-\xi_\ell)\beta_\ell-\left(\bar{C}(i,\ell)-\xi_{i\ell}\right)\beta_i-(\beta_i-\beta_\ell)\Lambda(\mathbf{C})\right] \\
&= \mathrm{Var}\big[(\bar{C}_i-\xi_i)\beta_i-(\bar{C}_\ell-\xi_\ell)\beta_i+(\bar{C}_\ell-\xi_\ell)\beta_i-(\bar{C}_\ell-\xi_\ell)\beta_\ell \\
&\qquad -\left(\bar{C}(i,\ell)-\xi_{i\ell}\right)\beta_i-(\beta_i-\beta_\ell)\Lambda(\mathbf{C})\big] \\
&= \mathrm{Var}\big[(\bar{C}_\ell-\xi_\ell)\beta_i-(\bar{C}_\ell-\xi_\ell)\beta_\ell-(\beta_i-\beta_\ell)\Lambda(\mathbf{C})\big] \\
&= \mathrm{Var}\big[(\beta_i-\beta_\ell)(\bar{C}_\ell-\xi_\ell-\Lambda(\mathbf{C}))\big] \\
&= (\beta_i-\beta_\ell)^2\mathrm{Var}\left[\bar{C}_\ell-\Lambda(\mathbf{C})\right].
\end{aligned}$$

So, when we assume that Model (5) holds but the true model is Model (6) the variance of the CV estimator is

$$\mathrm{Var}[\widehat{\mu}_{i\ell}(n)] = (\beta_i-\beta_\ell)^2\mathrm{Var}\left[\bar{C}_\ell-\Lambda(\mathbf{C})\right]+\mathrm{Var}[\eta_j(i,\ell)]\left(\frac{n-2}{n(n-3)}\right). \tag{13}$$

On the other hand, when we assume Model (6) holds and Model (6) is indeed the true model

$$\mathrm{Var}[\widehat{\mu}_{i\ell}(n)] = \mathrm{Var}[\eta_j(i,\ell)]\left(\frac{n-2}{n(n-4)}\right). \tag{14}$$

In summary, if we compute $\widehat{\mu}_{i\ell}(n)$ assuming Model (5) holds while Model (6) is the true model with $\boldsymbol{\beta}_i \neq \boldsymbol{\beta}_\ell$, then Equation (12) shows us that $\widehat{\mu}_{i\ell}(n)$ will be biased. Further, from Equation (13) and (14) we see that $\mathrm{Var}[\widehat{\mu}_{i\ell}(n)]$ will be increased, especially when $\boldsymbol{\beta}_i$ is very different from $\boldsymbol{\beta}_\ell$. This illustrates the inferiority of Model (5) relative to Model (6).

A.2 CV Combined Procedure with Individual Screening

We prove the multiplicative approach for Theorem 1,

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

In the multiplicative approach, we assume that Model (1) holds. 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

$$Z_i = \frac{\widehat{\mu}_k(n_k)-\widehat{\mu}_i(n_i)-(\mu_k-\mu_i)}{\sqrt{\widehat{\Delta}_k^2(n_k)\tau_k^2+\widehat{\Delta}_i^2(n_i)\tau_i^2}},\ V_i = \frac{\widehat{\mu}_k(N_k)-\widehat{\mu}_i(N_i)-(\mu_k-\mu_i)}{\sqrt{\widehat{\Delta}_k^2(N_k)\tau_k^2+\widehat{\Delta}_i^2(N_i)\tau_i^2}},$$

$$A_i = \widehat{\Delta}_k^2(n_k)\tau_k^2 + \widehat{\Delta}_i^2(n_i)\tau_i^2,\ D_i = \widehat{\Delta}_k^2(N_k)\tau_k^2 + \widehat{\Delta}_i^2(N_i)\tau_i^2.$$

The probability of correct selection is

$$\begin{aligned}
&\Pr\left\{\widehat{\mu}_k(n_k) - \widehat{\mu}_i(n_i) \geq -W_{ki}, \forall i \neq k;\ \widehat{\mu}_k(N_k) > \widehat{\mu}_i(N_i), \forall i \in I\right\} \\
&= \Pr\left\{\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}} \geq \frac{-W_{ki} - (\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;\right. \\
&\qquad \left.\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}} > \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 \in I\right\} \\
&\geq \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}},\ V_i < \frac{\delta}{\sqrt{\widehat{\Delta}_k^2(N_k)\tau_k^2 + \widehat{\Delta}_i^2(N_i)\tau_i^2}}, \forall i \neq k\right\} \qquad (15) \\
&= \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}},\ V_i < \frac{\delta}{\sqrt{\widehat{\Delta}_k^2(N_k)\tau_k^2 + \widehat{\Delta}_i^2(N_i)\tau_i^2}}, \forall i \neq k \middle| \mathbf{C}, \widehat{\tau}^2\right\}\right] \\
&\geq \mathrm{E}\left[\Pr\left\{Z_i \leq \frac{W_{ki}}{\sqrt{A_i}}, \forall i \neq k \middle| \mathbf{C}, \widehat{\tau}^2\right\} \Pr\left\{V_i < \frac{\delta}{\sqrt{D_i}}, \forall i \neq k \middle| \mathbf{C}, \widehat{\tau}^2\right\}\right] \qquad (16) \\
&= \mathrm{E}\left[\mathrm{E}\left[\Pr\left\{Z_i \leq \frac{W_{ki}}{\sqrt{A_i}}, \forall i \neq k \middle| \mathbf{C}, \widehat{\tau}^2\right\} \Pr\left\{V_i < \frac{\delta}{\sqrt{D_i}}, \forall i \neq k \middle| \mathbf{C}, \widehat{\tau}^2\right\} \middle| \mathbf{C}\right]\right] \qquad (17) \\
&\geq \mathrm{E}\left[\mathrm{E}\left[\Pr\left\{Z_i \leq \frac{W_{ki}}{\sqrt{A_i}}, \forall i \neq k \middle| \mathbf{C}, \widehat{\tau}^2\right\} \middle| \mathbf{C}\right] \mathrm{E}\left[\Pr\left\{V_i < \frac{\delta}{\sqrt{D_i}}, \forall i \neq k \middle| \mathbf{C}, \widehat{\tau}^2\right\} \middle| \mathbf{C}\right]\right]. \qquad (18)
\end{aligned}$$

Inequality (15) holds because $\mu_k - \mu_i \geq \delta$, $\{\forall i \in I\}$ is a smaller set than $\{\forall i \neq k\}$ which makes the condition tougher, and because of the symmetry of the normal distribution. And Inequality (16) is an application of Slepian's inequality (e.g., Tong 1980). Then since the first term in Inequality (17) is a nonnegative, real-valued function and increasing in each of $\{\widehat{\tau}_1^2(n_1), \widehat{\tau}_2^2(n_2), \ldots, \widehat{\tau}_k^2(n_k)\}$, and the second term in Inequality (17) is 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 1 $\{\widehat{\tau}_1^2(n_1), \widehat{\tau}_2^2(n_2), \ldots, \widehat{\tau}_k^2(n_k)\}$ are conditionally independent given $\mathbf{C}$, we can apply Lemma 2.4 in Tamhane (1977) to get Inequality (18). From Nelson and Staum (2006), we know that

$$\begin{aligned}
\mathrm{E}\left[\Pr\left\{Z_i \leq \frac{W_{ki}}{\sqrt{A_i}}, \forall i \neq k \middle| \mathbf{C}, \widehat{\tau}^2\right\} \middle| \mathbf{C}\right] &= \Pr\left\{Z_i \leq \frac{W_{ki}}{\sqrt{A_i}}, \forall i \neq k \middle| \mathbf{C}\right\} \geq 1 - \alpha_0 \quad \text{and} \\
\mathrm{E}\left[\mathrm{E}\left[\Pr\left\{V_i < \frac{\delta}{\sqrt{D_i}}, \forall i \neq k \middle| \mathbf{C}, \widehat{\tau}^2\right\} \middle| \mathbf{C}\right]\right] &= \Pr\left\{V_i < \frac{\delta}{\sqrt{D_i}}, \forall i \neq k\right\} \geq 1 - \alpha_1 - \alpha_2.
\end{aligned}$$

So we can conclude from Inequality (18) that the probability of correct selection is

$$\Pr\{\widehat{\mu}_k(n_k)-\widehat{\mu}_i(n_i) \geq -W_{ki}, \forall i \neq k;\ \widehat{\mu}_k(N_k) > \widehat{\mu}_i(N_i), \forall i \in I\} \geq (1-\alpha_0)\,(1-\alpha_1-\alpha_2) = 1-\alpha.$$

We need to verify that Slepian's inequality can be applied for Inequality (16). It is easy to show that the $\mathrm{Cov}[Z_i, Z_j|\mathbf{C}, \widehat{\tau}^2]$ and $\mathrm{Cov}[V_i, V_j|\mathbf{C}, \widehat{\tau}^2]$ are nonnegative for any system $i \neq j$ (Nelson and Staum 2006). So here we only need to examine $\mathrm{Cov}[Z_i, V_j|\mathbf{C}, \widehat{\tau}^2]$.

When $i \neq j$,

$$\begin{aligned}
\mathrm{Cov}[Z_i, V_j|\mathbf{C}, \widehat{\tau}^2] &= \mathrm{Cov}\left[\frac{\widehat{\mu}_k(n_k) - \widehat{\mu}_i(n_i) - (\mu_k - \mu_i)}{\sqrt{\widehat{\Delta}_k^2(n_k)\tau_k^2 + \widehat{\Delta}_i^2(n_i)\tau_i^2}}, \frac{\widehat{\mu}_k(N_k) - \widehat{\mu}_j(N_j) - (\mu_k - \mu_j)}{\sqrt{\widehat{\Delta}_k^2(N_k)\tau_k^2 + \widehat{\Delta}_j^2(N_j)\tau_j^2}} \middle| \mathbf{C}, \widehat{\tau}^2\right] \\
&= \frac{1}{a_i d_j}\mathrm{Cov}\left[\widehat{\mu}_k(n_k), \widehat{\mu}_k(N_k)|\mathbf{C}, \widehat{\tau}^2\right]
\end{aligned}$$

where $a_i = \sqrt{A_i}$, $d_j = \sqrt{D_j}$. We can factor out a_i and d_j since they are both constants when we condition on $\mathbf{C}$. And we know

$$\begin{aligned}
\widehat{\mu}_k(n_k) &= \left[\frac{1}{n_k}\mathbf{1}'_{n_k\times 1} - \left(\bar{\mathbf{C}}_k(n_k) - \boldsymbol{\xi}_k\right)'\left(\mathbf{L}'_k(n_k)\mathbf{L}_k(n_k)\right)^{-1}\mathbf{L}'_k(n_k)\right]\mathbf{X}_k(n_k) \\
&= \mathbf{a}'\mathbf{X}_k(n_k)
\end{aligned}$$

$$\begin{aligned}
\widehat{\mu}_k(N_k) &= \left[\frac{1}{N_k}\mathbf{1}'_{N_k\times 1} - \left(\bar{\mathbf{C}}_k(N_k) - \boldsymbol{\xi}_k\right)'\left(\mathbf{L}'_k(N_k)\mathbf{L}_k(N_k)\right)^{-1}\mathbf{L}'_k(N_k)\right]\mathbf{X}_k(N_k) \\
&= \mathbf{b}'\mathbf{X}_k(N_k).
\end{aligned}$$

It follows that

$$\begin{aligned}
\mathrm{Cov}[Z_i, V_j|\mathbf{C}, \widehat{\tau}^2] &= \frac{1}{a_i d_j}\mathrm{Cov}\left[\widehat{\mu}_k(n_k), \widehat{\mu}_k(N_k)|\mathbf{C}, \widehat{\tau}^2\right] \\
&= \frac{1}{a_i d_j}\mathbf{a}'\Omega\mathbf{b}
\end{aligned}$$

where

$$\Omega = \begin{pmatrix} \mathbf{E} & \mathbf{F} \end{pmatrix},\ \mathbf{E} = \mathrm{Var}[\mathbf{X}_k(n_k)],\ \mathbf{F} = [\mathbf{0}]_{n_k\times(N_k-n_k)}.$$

Since $\mathbf{E}$ is a diagonal matrix with positive elements and $a_i d_j$ is positive, we can conclude that $\mathrm{Cov}[Z_i, V_j|\mathbf{C}, \widehat{\tau}^2]$ is nonnegative if $\mathbf{a}'\mathbf{B}$ is nonnegative, where the vector $\mathbf{B}$ is composed of the first n_k elements of the vector $\mathbf{b}$. We know

$$\mathbf{B}' = \left[\frac{1}{N_k}\mathbf{1}'_{n_k\times 1} - \left(\bar{\mathbf{C}}_k(N_k) - \boldsymbol{\xi}_k\right)'\left(\mathbf{L}'_k(N_k)\mathbf{L}_k(N_k)\right)^{-1}\left(\mathbf{L}'_k(n_k) + \mathbf{m}_k(n_k)\right)\right]$$

where

$$\mathbf{m}_k(n_k) = \{\bar{\mathbf{C}}_k(n_k) - \bar{\mathbf{C}}_k(N_k), \bar{\mathbf{C}}_k(n_k) - \bar{\mathbf{C}}_k(N_k), \ldots, \bar{\mathbf{C}}_k(n_k) - \bar{\mathbf{C}}_k(N_k)\}_{1\times n_k}.$$

Then

$$\begin{aligned}
\mathbf{a}'\mathbf{B} &= \frac{1}{N_k} - \frac{1}{n_k}\mathbf{1}'_{n_k\times 1}\left[\mathbf{L}_k(n_k)\left(\mathbf{L}'_k(N_k)\mathbf{L}_k(N_k)\right)^{-1}\left(\bar{\mathbf{C}}_k(N_k) - \boldsymbol{\xi}_k\right)\right] \\
&\quad -\frac{1}{n_k}\mathbf{1}'_{n_k\times 1}\left[\mathbf{m}'_k(n_k)\left(\mathbf{L}'_k(N_k)\mathbf{L}_k(N_k)\right)^{-1}\left(\bar{\mathbf{C}}_k(N_k) - \boldsymbol{\xi}_k\right)\right] \\
&\quad -\left(\bar{\mathbf{C}}_k(n_k) - \boldsymbol{\xi}_k\right)'\left(\mathbf{L}'_k(n_k)\mathbf{L}_k(n_k)\right)^{-1}\mathbf{L}'_k(n_k)\frac{1}{N_k}\mathbf{1}_{n_k\times 1} \\
&\quad +\left(\bar{\mathbf{C}}_k(n_k) - \boldsymbol{\xi}_k\right)'\left(\mathbf{L}'_k(n_k)\mathbf{L}_k(n_k)\right)^{-1}\mathbf{L}'_k(n_k)\left[\mathbf{L}_k(n_k)\left(\mathbf{L}'_k(N_k)\mathbf{L}_k(N_k)\right)^{-1}\left(\bar{\mathbf{C}}_k(N_k) - \boldsymbol{\xi}_k\right)\right] \\
&\quad +\left(\bar{\mathbf{C}}_k(n_k) - \boldsymbol{\xi}_k\right)'\left(\mathbf{L}'_k(n_k)\mathbf{L}_k(n_k)\right)^{-1}\mathbf{L}'_k(n_k)\left[\mathbf{m}'_k(n_k)\left(\mathbf{L}'_k(N_k)\mathbf{L}_k(N_k)\right)^{-1}\left(\bar{\mathbf{C}}_k(N_k) - \boldsymbol{\xi}_k\right)\right] \\
&= \frac{1}{N_k} - \frac{1}{n_k}\mathbf{1}'_{n_k\times 1}\left[\mathbf{m}'_k(n_k)\left(\mathbf{L}'_k(N_k)\mathbf{L}_k(N_k)\right)^{-1}\left(\bar{\mathbf{C}}_k(N_k) - \boldsymbol{\xi}_k\right)\right] \\
&\quad +\left(\bar{\mathbf{C}}_k(n_k) - \boldsymbol{\xi}_k\right)'\left[\left(\mathbf{L}'_k(N_k)\mathbf{L}_k(N_k)\right)^{-1}\left(\bar{\mathbf{C}}_k(N_k) - \boldsymbol{\xi}_k\right)\right] \qquad (19) \\
&= \frac{1}{N_k} + \left(\bar{\mathbf{C}}_k(N_k) - \boldsymbol{\xi}_k\right)'\left[\left(\mathbf{L}'_k(N_k)\mathbf{L}_k(N_k)\right)^{-1}\left(\bar{\mathbf{C}}_k(N_k) - \boldsymbol{\xi}_k\right)\right] \qquad (20) \\
&= \frac{1}{N_k} + \frac{\left(\bar{\mathbf{C}}_k(N_k) - \boldsymbol{\xi}_k\right)'\mathbf{S}^{-1}\left(\bar{\mathbf{C}}_k(N_k) - \boldsymbol{\xi}_k\right)}{N_k - 1} \qquad (21) \\
&= \frac{1}{N_k} + \frac{\mathbf{T}^2}{N_k(N_k-1)} > 0 \qquad (22)
\end{aligned}$$

where $\mathbf{S}$ is the sample covariance matrix of controls from system k. That is, $\mathbf{L}'_k(N_k)\mathbf{L}_k(N_k) = (N_k - 1)\mathbf{S}$. And $\mathbf{T}^2$ is the generalized $\mathbf{T}^2$-statistic of controls from system k, which means $\mathbf{T}^2 = N_k\left(\bar{\mathbf{C}}_k(N_k) - \boldsymbol{\xi}_k\right)'\mathbf{S}^{-1}\left(\bar{\mathbf{C}}_k(N_k) - \boldsymbol{\xi}_k\right)$ (Anderson, 1984). Equality (19) holds because $\mathbf{1}'_{n_k\times 1}\mathbf{L}_k(n_k) = 0$, $\left(\mathbf{L}'_k(n_k)\mathbf{L}_k(n_k)\right)^{-1}\mathbf{L}'_k(n_k)\mathbf{L}_k(n_k) = \mathbf{I}$, where $\mathbf{I}$ is identity matrix, and $\mathbf{L}'_k(n_k)\mathbf{m}'_k(n_k) = 0$. Equality (20) holds because $(1/n_k)\mathbf{1}'_{n_k\times 1}\mathbf{m}'_k(n_k) = (\bar{\mathbf{C}}_k(n_k) - \bar{\mathbf{C}}_k(N_k))'$. Equality (21) and Equality (22) hold because of the definition of $\mathbf{S}^{-1}$ and $\mathbf{T}^2$. Therefore, $\mathrm{Cov}[Z_i, V_j|\mathbf{C}, \widehat{\tau}^2]$ is positive when $i \neq j$.

When $i = j$,

$$\mathrm{Cov}[Z_i, V_j|\mathbf{C}, \widehat{\tau}^2] = \frac{1}{a_i d_j}\Big[\mathrm{Cov}\left[\widehat{\mu}_k(n_k), \widehat{\mu}_k(N_k)|\mathbf{C}, \widehat{\tau}^2\right] + \mathrm{Cov}\left[\widehat{\mu}_j(n_j), \widehat{\mu}_j(N_j)|\mathbf{C}, \widehat{\tau}^2\right]\Big].$$

We can obtain $\mathrm{Cov}\left[\widehat{\mu}_j(n_j), \widehat{\mu}_j(N_j)|\mathbf{C}, \widehat{\tau}^2\right] > 0$ similarly. Therefore, $\mathrm{Cov}[Z_i, V_j|\mathbf{C}, \widehat{\tau}^2]$ is also positive when $i = j$.

A.3 CV Combined Procedure with Paired Screening

Theorem 2 is proven here. We can apply the additive approach to the CV combined procedure with paired screening. We assume that Model (1) holds. Define

$$V_i = \frac{\widehat{\mu}_k(N_k) - \widehat{\mu}_i(N_i) - (\mu_k - \mu_i)}{\sqrt{\widehat{\Delta}_k^2(N_k)\tau_k^2 + \widehat{\Delta}_i^2(N_i)\tau_i^2}}.$$

The probability of correct selection is

$$\begin{aligned}
&\Pr\left\{\widehat{\mu}_{ki}(n) \geq -t_{ki}\,\widehat{\Delta}_{ki}(n)\widehat{\tau}_{ki}(n), \forall i \neq k;\ \widehat{\mu}_k(N_k) > \widehat{\mu}_i(N_i), \forall i \in I\right\} \\
&\geq \Pr\left\{ \widehat{\mu}_{ki}(n) \geq -t_{ki}\,\widehat{\Delta}_{ki}(n)\widehat{\tau}_{ki}(n), \forall i \neq k; \right. \\
&\qquad \left. \frac{\widehat{\mu}_k(N_k) - \widehat{\mu}_i(N_i) - (\mu_k - \mu_i)}{\sqrt{\widehat{\Delta}_k^2(N_k)\tau_k^2 + \widehat{\Delta}_i^2(N_i)\tau_i^2}} > \frac{-(\mu_k - \mu_i)}{\sqrt{\widehat{\Delta}_k^2(N_k)\tau_k^2 + \widehat{\Delta}_i^2(N_i)\tau_i^2}}, \forall i \in I \right\} \\
&\geq \Pr\left\{ \widehat{\mu}_{ki}(n) \geq -t_{ki}\,\widehat{\Delta}_{ki}(n)\widehat{\tau}_{ki}(n), \forall i \neq k;\ V_i > \frac{-\delta}{\sqrt{\widehat{\Delta}_k^2(N_k)\tau_k^2 + \widehat{\Delta}_i^2(N_i)\tau_i^2}}, \forall i \in I \right\} \quad (23) \\
&\geq \Pr\left\{ \widehat{\mu}_{ki}(n) \geq -t_{ki}\,\widehat{\Delta}_{ki}(n)\widehat{\tau}_{ki}(n), \forall i \neq k;\ V_i < \frac{\delta}{\sqrt{\widehat{\Delta}_k^2(N_k)\tau_k^2 + \widehat{\Delta}_i^2(N_i)\tau_i^2}}, \forall i \neq k; \right. \\
&\qquad \left. \widehat{\Delta}_i^2(N_i) \leq \frac{\delta^2}{h^2\,\widehat{\tau}_i^2(n_i)}, \forall i \right\} \quad (24) \\
&\geq \Pr\left\{ \widehat{\mu}_{ki}(n) \geq -t_{ki}\,\widehat{\Delta}_{ki}(n)\widehat{\tau}_{ki}(n), \forall i \neq k;\ V_i < \frac{h}{\sqrt{\frac{\tau_k^2}{\widehat{\tau}_k^2(n_k)} + \frac{\tau_i^2}{\widehat{\tau}_i^2(n_i)}}}, \forall i \neq k; \right. \\
&\qquad \left. \widehat{\Delta}_i^2(N_i) \leq \frac{\delta^2}{h^2\,\widehat{\tau}_i^2(n_i)}, \forall i \right\} \\
&\geq 1 - p_0 - p_1 - p_2 \quad (25)
\end{aligned}$$

where

$$p_0 = 1 - \Pr\left\{\widehat{\mu}_{ki}(n) \geq -t_{ki}\,\widehat{\Delta}_{ki}(n)\widehat{\tau}_{ki}(n), \forall i \neq k\right\} \quad \text{and}$$

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

$$p_2 = 1 - \Pr\left\{\widehat{\Delta}_i^2(N_i) \leq \frac{\delta^2}{h^2\,\widehat{\tau}_i^2(n_i)}, \forall i\right\}.$$

By Lemma 1, the conditional distribution of V_i given $\{\widehat{\tau}^2, \mathbf{C}\}$ is standard normal. Inequality (23) holds because $\mu_k - \mu_i \geq \delta$. Inequality (24) holds because of the symmetry of the normal distribution and the event is smaller if we require a bound on the value of $\widehat{\Delta}_i^2(N_i)$, while Inequality (25) is an application of the Bonferroni inequality.

We know that $p_0 \leq \alpha_0$, $p_1 \leq \alpha_1$, $p_2 \leq \alpha_2$ (Nelson and Staum 2006). So we can conclude that the probability of correct selection is

$$\Pr\{\widehat{\mu}_{ki}(n) \geq -t_{ki}\,\widehat{\Delta}_{ki}(n)\widehat{\tau}_{ki}(n), \forall i \neq k;\ \widehat{\mu}_k(N_k) > \widehat{\mu}_i(N_i), \forall i \in I\} \geq 1 - \alpha_0 - \alpha_1 - \alpha_2 = 1 - \alpha.$$

References

Anderson, T. W. 1984. *An Introduction to Multivariate Statistical Analysis*, 2nd Edition. New York: John Wiley.

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

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

Bechhofer, R. E., T. J. Santner and D. Goldsman. 1995. *Design and Analysis of Experiments for Statistical Selection, Screening and Multiple Comparisons*. 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. 1989. Batch size effects on the efficiency of control variates in simulation. *European Journal of Operational Research* 43:184–196.

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

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

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.

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*. New York: Academic Press.

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