USING COMMON RANDOM NUMBERS AND CONTROL VARIATES IN MULTIPLE-COMPARISON PROCEDURES

WEI-NING YANG and BARRY L. NELSON
Ohio State University, Columbus, Ohio

(Received September 1989; revision received September 1990; accepted October 1990)

This paper considers the determination of the relative merits of two or more system designs via stochastic simulation experiments by constructing simultaneous interval estimates of certain differences in expected performance. Tukey's all-pairwise-comparisons procedure, Hsu's multiple-comparisons-with-the-best procedure, and Dunnett's multiple-comparisons-with-a-control procedure are standard methods for making such comparisons. We propose refinements for all three procedures through the use of two variance reduction techniques: common random numbers and control variates. We show that the proposed procedures are better than the standard multiple-comparison procedures in the sense that they have a larger probability of containing the true difference and, at the same time, not containing zero when a difference exists.

This paper considers the determination of the relative merits of two or more system designs via stochastic simulation experiments; this is undoubtedly the most frequent use of simulation in practice. Specifically, we compare the expected values of univariate performance measures that are common to each system design by constructing simultaneous interval estimates of certain differences in expected performance. This approach is known in statistics as *multiple comparisons* (see, for instance, Miller 1981, or Hochberg and Tamhane 1987). Applications include, but are not limited to, comparing the expected cost per period of several inventory policies, determining the allocation of buffer space that maximizes throughput in an assembly line, and evaluating the improvement in the response time of a computer system if new hardware is added.

In the simulation literature, ranking and selection procedures have often been recommended for comparing system designs, particularly when the goal is to select the best design (e.g., Gray and Goldsman 1988). The difference between ranking and selection and multiple comparisons is analogous to the difference between hypothesis testing and interval estimation. The former results in a decision, rather than an estimate, so it is less informative. In typical ranking and selection procedures the resulting inference pertains only to the design selected as the best. Also, since ranking and selection procedures always result in a decision, two-stage sampling is needed to obtain a prespecified probability of selecting the best system design. In contrast, multiple-comparison procedures provide inference about relationships among *all* system designs and can be implemented in a single stage of sampling, but they do not guarantee a decision.

Standard methods for making simultaneous comparisons are Tukey's (1953) all-pairwise-multiple-comparisons procedure, Hsu's (1984) multiple-comparisons-with-the-best procedure, and Dunnett's (1955) multiple-comparisons-with-a-control procedure. In this paper, we propose refinements for all three procedures that are applicable in stochastic simulation experiments, where refinement means greater sensitivity to differences in system performance. The refinement is achieved through the use of two variance reduction techniques: common random numbers and control variates. Common random numbers is a well known variance reduction technique that is used to sharpen estimators of differences. Unfortunately, statistical analysis under common random numbers for more than a single difference has been difficult to derive (an exception is Clark and Yang (1986) who developed a conservative ranking and selection procedure based on the Bonferroni inequality). The primary contribution of this paper is extending the use of common random numbers to simultaneous estimation of several differences.

After reviewing some background, three new multiple-comparison procedures are presented and shown to be superior to existing procedures under certain conditions. An example and some discussion end the paper.

Subject classifications. Simulation, statistical analysis: variance reduction. Statistics: multiple comparisons.

0030-364X/91/3904-0583 $01.25

1. BACKGROUND

Assume that $r \geqslant 2$ system designs are to be compared in terms of their mean performance, and denote the mean performance of the ith system by θ_i, $i = 1, 2, \ldots, r$. If the difference between the performance of each system design and every other system design is of interest, then Tukey's method of all pairwise multiple comparisons (MCA) provides simultaneous confidence intervals for $\theta_i - \theta_j$ for all $i \neq j$. However, if identification of the system with the largest mean performance is of interest, then $\theta_i - \max_{j \neq i} \theta_j$ for $i = 1, 2, \ldots, r$, are the appropriate parameters to estimate, because if $\theta_i - \max_{j \neq i} \theta_j > 0$, then system i has the largest mean; otherwise, $\theta_i - \max_{j \neq i} \theta_j \geqslant -\Delta$ indicates that the mean performance of system i is within Δ of the largest. Hsu's method of multiple comparisons with the best (MCB) provides simultaneous confidence intervals for $\theta_i - \max_{j \neq i} \theta_j$ for all i (see Hsu and Nelson 1988 and Yang and Nelson 1989 for examples of using MCB for optimization via simulation). Finally, if the difference between the mean performance of a designated system design (say, design r) and every other alternative system design is of interest, then Dunnett's method of multiple comparisons with a control (MCC) provides simultaneous confidence intervals for $\theta_i - \theta_r$ for all $i \neq r$.

Let Y_{ij} be the jth simulation output from the ith system design and suppose that $\theta_i = \mathrm{E}[Y_{ij}]$ for all j. The multiple-comparison procedures just cited are all applicable if the balanced one-way model (1) pertains:

$$Y_{ij} = \theta_i + \epsilon_{ij} \tag{1}$$

for $i = 1, 2, \ldots, r$ and $j = 1, 2, \ldots, n$, where $\epsilon_{11}, \epsilon_{12}, \ldots, \epsilon_{rn}$ are independent $N(0, \sigma^2)$ random variables with σ^2 unknown. Let $\theta_1, \theta_2, \ldots, \theta_r$ be estimated by the sample means

$$\bar{Y}_i = \frac{1}{n} \sum_{j=1}^{n} Y_{ij}$$

for $i = 1, 2, \ldots, r$, and let σ^2 be estimated by the pooled sample variance

$$\hat{\sigma}^2 = \frac{1}{r(n-1)} \sum_{i=1}^{r} \sum_{j=1}^{n} (Y_{ij} - \bar{Y}_i)^2.$$

The constants n and r, and the random variables $\bar{Y}_1, \ldots, \bar{Y}_r$ and $\hat{\sigma}^2$, are the inputs to all three of the multiple-comparison procedures described in the sections that follow.

In stochastic simulation experiments the response variable, Y_{ij}, often has a strong linear relationship with certain input random variables that drive the simulation experiment. Suppose that the response variable can be described by the following more detailed model:

$$Y_{ij} = \theta_i + \boldsymbol{\beta}_i'(\mathbf{X}_j - \boldsymbol{\mu}) + \eta_{ij} \tag{2}$$

for $i = 1, 2, \ldots, r$ and $j = 1, 2, \ldots, n$, where $\eta_{11}, \eta_{12}, \ldots, \eta_{rn}$ are independent $N(0, \tau^2)$ random variables with τ^2 unknown; $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ are i.i.d. $q \times 1$ vectors of input random variables that are independent of η_{ij} and have known mean vector $\boldsymbol{\mu}$; and $\boldsymbol{\beta}_i$ is a $q \times 1$ unknown constant vector. In contrast to model (1), where $\mathrm{Var}[Y_{ij}] = \sigma^2$ for all i and j, in model (2) $\mathrm{Var}[Y_{ij}] = \tau^2 + \boldsymbol{\beta}_i'\boldsymbol{\Sigma}_{\mathrm{X}}\boldsymbol{\beta}_i$, where $\boldsymbol{\Sigma}_{\mathrm{X}} = \mathrm{Var}[\mathbf{X}_j]$, which is not necessarily the same for all i.

Let $\theta_1, \theta_2, \ldots, \theta_r$ be estimated by the *control-variate estimators*

$$\hat{\theta}_i = \bar{Y}_i - \hat{\boldsymbol{\beta}}_i'(\bar{\mathbf{X}} - \boldsymbol{\mu})$$

for $i = 1, 2, \ldots, r$, and let τ^2 be estimated by

$$\hat{\tau}^2 = \frac{1}{r(n-q-1)} \sum_{i=1}^{r} \sum_{j=1}^{n} [Y_{ij} - \hat{\theta}_i - \hat{\boldsymbol{\beta}}_i'(\mathbf{X}_j - \boldsymbol{\mu})]^2$$

where

$$\bar{\mathbf{X}} = \frac{1}{n} \sum_{j=1}^{n} \mathbf{X}_j$$

and

$$\hat{\boldsymbol{\beta}}_i = \mathbf{S}_{\mathrm{XX}}^{-1}\mathbf{S}_{\mathrm{X}Y_i} = \left[\frac{1}{n-1} \sum_{j=1}^{n} (\mathbf{X}_j - \bar{\mathbf{X}})(\mathbf{X}_j - \bar{\mathbf{X}})'\right]^{-1} \times \left[\frac{1}{n-1} \sum_{j=1}^{n} (\mathbf{X}_j - \bar{\mathbf{X}})(Y_{ij} - \bar{Y}_i)\right].$$

Notice that we have assumed that the $\mathbf{X}_j$, called the *control vectors*, are *identical* across system designs. This means that the same control vectors must be available in each system design, and that they can be forced to take identical values, typically by generating them using common random numbers. For example, in an inventory system different inventory policies result in different system designs, but the demand on each system is independent of the inventory policy simulated. Thus, if the total demand during the planning horizon is the control variate, then common random numbers results in an *identical* total demand for each inventory policy.

The approach that we have taken, which is central to the analysis that follows, is to account for the dependence resulting from common random numbers through a linear model (2) with unknown parameters. We discuss the consequences of this assumption further in the final section. The other key assumption—that the control variates are identical across systems—restricts the simulation input processes that can be used as control variates. Inputs not used as control variates should be independent across systems to satisfy model (2); in practice, this means that different random number streams must be assigned to each system design for those input processes.

Let $\hat{\delta}^2 = n^{-1} + (n-1)^{-1}(\bar{\mathbf{X}} - \boldsymbol{\mu})'\mathbf{S}_{\mathbf{XX}}^{-1}(\bar{\mathbf{X}} - \boldsymbol{\mu})$. The constants n, q and r, and the random variables $\hat{\theta}_1, \ldots, \hat{\theta}_r$, $\hat{\tau}^2$ and $\hat{\delta}^2$, are the inputs to all three new multiple-comparison procedures derived below.

In the next three sections, we construct confidence intervals for multiple comparisons based on control-variate estimators. The following two lemmas are critical in those derivations.

Lemma 1. *Let $\vec{\mathbf{X}} = \{\mathbf{X}_j, j = 1, 2, \ldots, n\}$. Assume that model (2) holds. Then conditional on $\vec{X}$, the random variables $\hat{\theta}_i$ are independent $N(\theta_i, \hat{\delta}^2\tau^2)$ random variables for $i = 1, 2, \ldots, r$.*

Proof. Notice that

$$
\begin{aligned}
\hat{\theta}_i &= \bar{Y}_i - \hat{\boldsymbol{\beta}}_i(\bar{\mathbf{X}} - \boldsymbol{\mu}) \\
&= \bar{Y}_i - [\mathbf{S}_{\mathbf{XX}}^{-1}\mathbf{S}_{\mathbf{X}Y_i}]'(\bar{\mathbf{X}} - \boldsymbol{\mu}) \\
&= \bar{Y}_i - [(\mathbf{L}'\mathbf{L})^{-1}\mathbf{L}'\mathbf{Y}_i]'(\bar{\mathbf{X}} - \boldsymbol{\mu}) \\
&= n^{-1}\mathbf{1}_{1\times n}\mathbf{Y}_i - [\mathbf{Y}_i'\mathbf{L}(\mathbf{L}'\mathbf{L})^{-1}](\bar{\mathbf{X}} - \boldsymbol{\mu}) \\
&= \{n^{-1}\mathbf{1}_{1\times n} - (\bar{\mathbf{X}} - \boldsymbol{\mu})'(\mathbf{L}'\mathbf{L})^{-1}\mathbf{L}'\}\mathbf{Y}_i \qquad (3)
\end{aligned}
$$

where $\mathbf{L}' = [(\mathbf{X}_1 - \bar{\mathbf{X}}), (\mathbf{X}_2 - \bar{\mathbf{X}}), \ldots, (\mathbf{X}_n - \bar{\mathbf{X}})]$; $\mathbf{1}_{1\times n}$ is a $1 \times n$ matrix with each component 1, and $\mathbf{Y}_i' = [Y_{i1}, Y_{i2}, \ldots, Y_{in}]$. Thus, conditional on $\vec{\mathbf{X}}$, $\hat{\theta}_i$ is normally distributed because it is a linear combination of independent normally distributed random variables. Also, $\hat{\theta}_i$ is independent of $\hat{\theta}_j$ for all $i \neq j$ since, conditional on $\vec{\mathbf{X}}$, $\mathbf{Y}_i$ is independent of $\mathbf{Y}_j$.

The conditional mean and variance of $\hat{\theta}_i$ can be derived from (3) by noting that $\mathrm{E}[Y_{ij} | \vec{\mathbf{X}}] = \theta_i + \boldsymbol{\beta}_i'(\mathbf{X}_j - \boldsymbol{\mu})$ and $\mathrm{Var}[Y_{ij}\vec{\mathbf{X}}] = \tau^2$ under model (2).

Lemma 2. *Assume that model (2) holds. Then conditional on $\vec{\mathbf{X}}$, $\hat{\tau}^2$ is independent of $\hat{\theta}_i$ for all i, and $r(n - q - 1)\hat{\tau}^2/\tau^2$ has a χ^2 distribution with $r(n - q - 1)$ degrees of freedom.*

Proof. Let

$$
\mathbf{A} = \begin{pmatrix} 1 & (\mathbf{X}_1 - \boldsymbol{\mu})' \\ 1 & (\mathbf{X}_2 - \boldsymbol{\mu})' \\ \vdots & \vdots \\ 1 & (\mathbf{X}_n - \boldsymbol{\mu})' \end{pmatrix}
$$

and

$$
\hat{\mathbf{Y}}_i = \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}\mathbf{Y}_i = \begin{pmatrix} \hat{\theta}_i + \hat{\boldsymbol{\beta}}_i'(\mathbf{X}_1 - \boldsymbol{\mu}) \\ \hat{\theta}_i + \hat{\boldsymbol{\beta}}_i'(\mathbf{X}_2 - \boldsymbol{\mu}) \\ \vdots \\ \hat{\theta}_i + \hat{\boldsymbol{\beta}}_i'(\mathbf{X}_n - \boldsymbol{\mu}) \end{pmatrix}.
$$

Then

$$
\begin{aligned}
Q_i &= \sum_{j=1}^{n} [Y_{ij} - \hat{\theta}_i - \hat{\boldsymbol{\beta}}_i'(\mathbf{X}_j - \boldsymbol{\mu})]^2 \\
&= (\mathbf{Y}_i - \hat{\mathbf{Y}}_i)'(\mathbf{Y}_i - \hat{\mathbf{Y}}_i) \\
&= (\mathbf{Y}_i - \hat{\mathbf{Y}}_i)'\mathbf{Y}_i \\
&= [\mathbf{Y}_i - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{Y}_i]'\mathbf{Y}_i \\
&= \mathbf{Y}_i'[\mathbf{I} - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}']\mathbf{Y}_i.
\end{aligned}
$$

Hence, conditional on $\vec{\mathbf{X}}$, Q_i is independent of Q_l for all $i \neq l$ because $\mathbf{Y}_i$ is conditionally independent of $\mathbf{Y}_l$ for all $i \neq l$ and $\mathbf{A}$ is fixed by conditioning. Also, conditional on $\vec{\mathbf{X}}$, Q_i/τ^2 has a chi-squared distribution with $n - q - 1$ degrees of freedom (Rao 1973, 3b.5 i), and is conditionally independent of $\hat{\theta}_i$ (Rao, 3b.4 viii). Notice that $r(n - q - 1)\hat{\tau}^2 = \sum_{i=1}^{r} Q_i$, which is the sum of conditionally independent chi-squared random variables. Thus, conditional on $\vec{\mathbf{X}}$, $r(n - q - 1)\hat{\tau}^2/\tau^2$ has a chi-squared distribution with $r(n - q - 1)$ degrees of freedom and is independent of $\hat{\theta}_i$ for all i.

2. ALL PAIRWISE MULTIPLE COMPARISONS (MCA)

In a discrete-item inventory system, let s represent the reorder point for an item and S represent the maximum stock level. Each feasible (s, S) combination is an inventory policy, and the expected cost per period of different policies may be of interest. Let θ_i be the expected cost per period for policy i. In this section, we develop a procedure that could be used to compare each inventory policy with the others.

For model (1), Tukey gives the following $(1 - \alpha)100\%$ simultaneous confidence intervals:

$$
\theta_i - \theta_j \in \bar{Y}_i - \bar{Y}_j \pm q^{\alpha}_{r,r(n-1)} \cdot \hat{\sigma}/\sqrt{n} \quad \text{for all } i \neq j
$$

for the $r(r-1)/2$ differences in system performance, where $q^{\alpha}_{r,r(n-1)}$ is the upper α quantile of the Studentized range distribution with r systems and $r(n-1)$ degrees of freedom; extensive tables of $q^{\alpha}_{r,r(n-1)}$ can be found in Hochberg and Tamhane. Theorem 1 establishes interval estimators for the same parameters based on the control-variate estimators.

Theorem 1. *Assuming that model* (2) *holds*:

$$\theta_i - \theta_j \in \hat{\theta}_i - \hat{\theta}_j \pm q^{\alpha}_{r,r(n-q-1)} \cdot \hat{\delta}\hat{\tau}$$

for all $i \neq j$ *are* $(1-\alpha)100\%$ *simultaneous confidence intervals for* $\theta_i - \theta_j$ *for all* $i \neq j$.

Proof. From Lemmas 1 and 2, and the definition of the Studentized range distribution, the following probability statement holds conditional on $\bar{\mathbf{X}}$:

$$\Pr\{\theta_i - \theta_j \in \hat{\theta}_i - \hat{\theta}_j \pm q^{\alpha}_{r,r(n-q-1)} \cdot \hat{\delta}\hat{\tau} \text{ for all } i \neq j \mid \bar{\mathbf{X}}\} = 1 - \alpha.$$

Since the right-hand side does not depend on $\bar{\mathbf{X}}$, the probability is $1-\alpha$ unconditionally as well.

3. MULTIPLE COMPARISONS WITH THE BEST (MCB)

In a manufacturing system there may be a limited amount of buffer space available between workstations in an assembly line. If buffer space is measured in terms of the number of discrete items that can be stored in the space, then different allocations lead to a finite number of system designs. Let θ_i be the expected throughput of the line under allocation i. The allocation that maximizes throughput for the line may be of interest. In this section, we develop a procedure that could be used to select the best allocation.

For model (1), Hsu gives the following $(1-\alpha)100\%$ simultaneous confidence intervals:

$$\theta_i - \max_{j\neq i} \theta_j \in \left[\left(\bar{Y}_i - \max_{j\neq i} \bar{Y}_j - d^{\alpha}_{r-1,r(n-1)} \cdot \hat{\sigma}\sqrt{2/n}\right)^-, \left(\bar{Y}_i - \max_{j\neq i} \bar{Y}_j + d^{\alpha}_{r-1,r(n-1)} \cdot \hat{\sigma}\sqrt{2/n}\right)^+\right],$$

$$i = 1, 2, \ldots, r$$

where $x^+ = \max\{0, x\}$, $x^- = \min\{0, x\}$, and $d^{\alpha}_{r-1,r(n-1)}$ is the upper α quantile of a random variable that is the maximum of $r-1$ equally correlated multivariate-t random variables with correlation ½ and $r(n-1)$ degrees of freedom; extensive tables of $d^{\alpha}_{r-1,r(n-1)}$ can be found in Hochberg and Tamhane. Theorem 2 establishes interval estimators for the same parameters based on the control-variate estimators.

Theorem 2. *Assuming that model* (2) *holds*:

$$\theta_i - \max_{j\neq i} \theta_j \in \left[\left(\hat{\theta}_i - \max_{j\neq i} \hat{\theta}_j - d^{\alpha}_{r-1,r(n-q-1)} \cdot \sqrt{2}\hat{\delta}\hat{\tau}\right)^-, \left(\hat{\theta}_i - \max_{j\neq i} \hat{\theta}_j + d^{\alpha}_{r-1,r(n-q-1)} \cdot \sqrt{2}\hat{\delta}\hat{\tau}\right)^+\right]$$

for $i = 1, 2, \ldots, r$ *are* $(1-\alpha)100\%$ *simultaneous confidence intervals for* $\theta_i - \max_{j\neq i}\theta_j$ *for all* i.

The proof, which is analogous to that of Theorem 1, follows the steps in Hsu and Nelson.

4. MULTIPLE COMPARISONS WITH A CONTROL (MCC)

In a computer system, new hardware (e.g., disk drive, additional memory) could be added to improve system response. Let θ_r be the mean response time for the existing system, and let θ_i, $i \neq r$ be the mean response time for hardware upgrade proposal i. In this section, we develop a procedure that could be used to compare each upgrade proposal to the existing system.

For model (1), Dunnett gives the following $(1-\alpha)100\%$ simultaneous confidence intervals:

$$\theta_i - \theta_r \in \bar{Y}_i - \bar{Y}_r \pm |d|^{\alpha}_{r-1,r(n-1)} \cdot \hat{\sigma}\sqrt{2/n}, \quad i = 1, 2, \ldots, r-1$$

where $|d|^{\alpha}_{r-1,r(n-1)}$ is the upper α quantile of a random variable that is the maximum of the absolute values of $r-1$ equally correlated multivariate-t random variables with correlation ½ and $r(n-1)$ degrees of freedom; extensive tables of $|d|^{\alpha}_{r-1,r(n-1)}$ can be found in Hochberg and Tamhane. Theorem 3 establishes interval estimators for the same parameters based on the control-variate estimators.

Theorem 3. *Assuming that model* (2) *holds*

$$\theta_i - \theta_r \in \hat{\theta}_i - \hat{\theta}_r \pm |d|^{\alpha}_{r-1,r(n-q-1)} \cdot \sqrt{2}\hat{\delta}\hat{\tau}$$

for $i = 1, 2, \ldots, r-1$ *are* $(1-\alpha)100\%$ *simultaneous confidence intervals for* $\theta_i - \theta_r$ *for all* $i \neq r$.

The proof is analogous to that of Theorem 1.

5. EXPECTED WIDTH OF THE CONFIDENCE INTERVALS

The goal of multiple-comparison procedures is to identify the differences between systems' performance. The expected width of the confidence interval is an important criterion, given that the interval achieves its nominal coverage, because it indicates the size difference that the procedure can distinguish. Theorem 4 shows that the control-variate interval estimators have a shorter expected width in the special case when *both* models (1) and (2) hold simultaneously. This assumption is necessary to make a fair comparison because neither model (1) or model (2) implies the other.

For notation, let H and H_{CV} denote the width of the standard and the control-variate, multiple-comparison intervals, respectively. Define $c^{\alpha}_{r,df}$ to be the generic upper α quantile for MCA, or $\sqrt{2}$ times the upper α quantile for MCB and MCC, with r systems and df degrees of freedom. Then $H_{\text{CV}} = 2c^{\alpha}_{r,r(n-q-1)} \cdot \hat{\delta}\hat{\tau}$ and $H = 2c^{\alpha}_{r,r(n-1)} \cdot \hat{\sigma}/\sqrt{n}$ are the generic widths of the control-variate and the standard intervals, respectively, with the exception of MCB, where H_{CV} and H are the fundamental quantities that determine the width of these constrained intervals.

Let R_i^2 be the squared multiple correlation coefficient between the response variable, Y_{ij}, and the control vector, $\mathbf{X}_j$, in model (2), and let $\rightarrow_{\mathscr{P}}$ denote convergence in probability. If both models (1) and (2) hold simultaneously, then the multiple correlation coefficient in model (2) is the same across systems and $\tau^2 = (1 - R^2)\sigma^2$. Under these conditions we have the following result.

Theorem 4. *Assuming that both models* (1) *and* (2) *pertain simultaneously*:

$$\frac{\mathrm{E}[H_{\text{CV}} \mid \vec{\mathbf{X}}]}{\mathrm{E}[H]} \xrightarrow{\mathscr{P}} \sqrt{1 - R^2} \leqslant 1.$$

Proof. Conditional on $\vec{\mathbf{X}}$, the expected width of the control-variate, multiple-comparison intervals is

$$\mathrm{E}[H_{\text{CV}} \mid \vec{\mathbf{X}}] = 2c^{\alpha}_{r,r(n-q-1)} \frac{\sqrt{2}}{\sqrt{r(n-q-1)}} \cdot \frac{\Gamma((r(n-q-1)+1)/2)}{\Gamma(r(n-q-1)/2)} \cdot \hat{\delta}\tau.$$

Also, the expected width of the standard multiple-comparison intervals is

$$\mathrm{E}[H] = 2c^{\alpha}_{r,r(n-1)} \frac{\sqrt{2}}{\sqrt{r(n-1)}} \frac{\Gamma((r(n-1)+1)/2)}{\Gamma(r(n-1)/2)} \cdot \frac{\sigma}{\sqrt{n}}.$$

From Theorem 5.2.3 in Anderson (1984), $n(\bar{\mathbf{X}} - \boldsymbol{\mu})'\mathbf{S}^{-1}_{\mathbf{XX}}(\bar{\mathbf{X}} - \boldsymbol{\mu}) \rightarrow_{\mathscr{D}} \chi^2_q$, a chi-squared random variable with q degrees of freedom, where $\rightarrow_{\mathscr{D}}$ denotes convergence in distribution. Thus, $n\hat{\delta}^2 \rightarrow_{\mathscr{P}} 1$, by Slutsky's theorem and the fact that convergence in distribution to a constant implies convergence in probability as well. Then, $n\hat{\delta}^2\tau^2 \rightarrow_{\mathscr{P}} (1 - R^2)\sigma^2$, which implies $\sqrt{n}\hat{\delta}\tau/\sigma \rightarrow_{\mathscr{P}} \sqrt{1 - R^2}$

As n increases for fixed r, $c^{\alpha}_{r,r(n-q-1)} \rightarrow c^{\alpha}_{r,r(n-1)} \rightarrow c^{\alpha}_{r\infty}$ because the quantile decreases at a decreasing rate as the degrees of freedom increase. Also, $\Gamma(a + \frac{1}{2})/\Gamma(a) \rightarrow \sqrt{a}$ as a increases. Thus, the ratio

$$\frac{\mathrm{E}[H_{\text{CV}} \mid \vec{\mathbf{X}}]}{\mathrm{E}[H]} = \frac{c^{\alpha}_{r,r(n-q-1)}}{c^{\alpha}_{r,r(n-1)}} \cdot \frac{\sqrt{n}\hat{\delta}\tau}{\sigma} \xrightarrow{\mathscr{P}} \sqrt{1 - R^2}.$$

If both models (1) and (2) hold simultaneously and the response variable and the control vector are jointly normal, then we obtain the following stronger result.

Theorem 5. *Assuming that both models* (1) *and* (2) *hold and* $(Y_{ij}, \mathbf{X}_j)$, $j = 1, 2, \ldots, n$ *are* $(q + 1)$*-variate normally distributed for all* i

$$\frac{\mathrm{E}[H_{\text{CV}}]}{\mathrm{E}[H]} \rightarrow \sqrt{1 - R^2} \leqslant 1.$$

Proof. Let $W_n = (n/(n-1))(\bar{\mathbf{X}} - \boldsymbol{\mu})'\mathbf{S}^{-1}_{\mathbf{XX}}(\bar{\mathbf{X}} - \boldsymbol{\mu})$, so that $\hat{\delta}^2 = (1 + W_n)/n$. Under the normality assumption, $(n - 1)W_n$ has a Hotelling-T^2 distribution with $n - 1$ degrees of freedom and parameter q, which implies that $(n - q)W_n/q$ follows an F distribution with degrees of freedom q and $n - q$. By the relationship between the F distribution and the Beta distribution, we have

$$\frac{1}{1 + W_n} \sim \mathrm{Beta}\left(\frac{n-q}{2}, \frac{q}{2}\right)$$

where $\mathrm{Beta}((n-q)/2, q/2)$ is the Beta distribution with parameters $(n - q)/2$ and $q/2$. Then

$$\mathrm{E}[\hat{\delta}] = \frac{1}{\sqrt{n}} \cdot \mathrm{E}[(1 + W_n)^{1/2}] = \frac{1}{\sqrt{n}} \frac{\Gamma(n/2)\Gamma((n-q-1)/2)}{\Gamma((n-q)/2)\Gamma((n-1)/2)}$$

yielding

$$\begin{aligned}
&\mathrm{E}[H_{\mathrm{CV}}] \\
&= \mathrm{E}[\mathrm{E}[H_{\mathrm{CV}} \mid \tilde{\mathbf{X}}]] \\
&= 2c^{\alpha}_{r,r(n-q-1)} \frac{\sqrt{2}}{\sqrt{r(n-q-1)}} \\
&\quad \cdot \frac{\Gamma((r(n-q-1)+1)/2)}{\Gamma(r(n-q-1)/2)} \cdot \mathrm{E}[\hat{\delta}] \cdot \tau \\
&= c^{\alpha}_{r,r(n-q-1)} \frac{\sqrt{2}}{\sqrt{r(n-q-1)}} \\
&\quad \cdot \frac{\Gamma((r(n-q-1)+1)/2)}{\Gamma(r(n-q-1)/2)} \\
&\quad \times \frac{1}{\sqrt{n}} \frac{\Gamma(n/2)\Gamma((n-q-1)/2)}{\Gamma((n-q)/2)\Gamma((n-1)/2)} \cdot \tau.
\end{aligned}$$

Thus

$$\begin{aligned}
&\frac{\mathrm{E}[H_{\mathrm{CV}}]}{\mathrm{E}[H]} \\
&= \frac{c^{\alpha}_{r,r(n-q-1)}}{c^{\alpha}_{r,r(n-1)}} \frac{\sqrt{r(n-1)}}{\sqrt{r(n-q-1)}} \\
&\quad \cdot \frac{\Gamma((r(n-q-1)+1)/2)}{\Gamma(r(n-q-1)/2)} \frac{\Gamma(r(n-1)/2)}{\Gamma((r(n-1)+1)/2)} \\
&\quad \times \frac{\Gamma(n/2)\Gamma((n-q-1)/2)}{\Gamma((n-q)/2)\Gamma((n-1)/2)} \sqrt{1-R^2} \to \sqrt{1-R^2}.
\end{aligned}$$

Even though the confidence interval width is an important criterion when the interval estimator achieves its nominal coverage, it only reflects the performance of the variance estimator. Point estimator performance also affects the interval estimator. When model (2) holds and n is not too small, $\hat{\theta}_i$ is a better point estimator of θ_i than $\bar{Y}_i$ in terms of smaller variance (Lavenberg and Welch 1981).

A more comprehensive criterion for multiple-comparison procedures is the probability of identifying differences and the direction of differences, when differences exist, since it considers both point estimator and variance estimator performance simultaneously. We empirically compare the control-variate procedures and the standard procedures in terms of this probability in the next section.

6. EMPIRICAL EVALUATION

Multiple-comparison procedures construct simultaneous confidence intervals for selected differences in expected system performance. If $\mathscr{C}$ is the event that the intervals simultaneously contain all of the selected differences, either $\theta_i - \theta_j$, $\theta_i - \max_{j \neq i} \theta_j$ or $\theta_i - \theta_r$, then $\Pr\{\mathscr{C}\} = 1 - \alpha$ when the assumptions underlying the procedure are satisfied. The event $\mathscr{C}$ could be called correct inference because the procedure correctly identifies a region that contains all of the parameters. Let $\mathscr{U}$ be the event that the intervals exclude zero when the true difference is not zero. The event $\mathscr{U}$ could be called useful inference because the procedure distinguishes differences in expected performance. Given two multiple-comparison procedures for the same estimation problem when both have $\Pr[\mathscr{C}\} = 1 - \alpha$, the superior procedure is the one for which $\Pr\{\mathscr{C} \cap \mathscr{U}\}$ is larger; that is, the one that has a larger probability of *correct and useful inference*. Notice that the event $\mathscr{C} \cap \mathscr{U}$ implies that differences, and the direction of the differences, are correctly identified.

We estimated $\Pr\{\mathscr{C} \cap \mathscr{U}\}$ for an (s, S) inventory system similar to the one mentioned earlier. In an (s, S) inventory system some discrete item is periodically reviewed. If the inventory level is found to be below s units, then an order is issued to bring the inventory level up to S units; otherwise, no additional items are ordered. Different (s, S) inventory policies result in different inventory systems. Koenig and Law (1985) used this example to illustrate a two-stage subset selection procedure for determining a subset that contains the inventory policy with the minimum expected cost; Hsu and Nelson (1988) and Yang and Nelson (1989) used this example to illustrate MCB. See any of these papers for a detailed description of the model.

The performance measure of the (s, S) inventory system is the expected average cost per period of the inventory system for 30 periods. The only stochastic element in the model is the demand for inventory in each period, which is assumed to a sequence of i.i.d. Poisson random variables with a common mean of 25. The five policies considered, and their corresponding expected average cost per period (which can be calculated analytically for this example), are given in Table I.

The probability of correct inference and the probability of correct and useful inference were estimated for the standard and control-variate procedures for $n = 10, 20, 30, \ldots, 100$ by repeating the entire experiment 700 times, implying a maximum standard error of $(0.25/700)^{1/2} \approx 0.02$ for the probability estimates. The nominal confidence level was set at $\alpha = 0.05$, so that the estimated $\Pr\{\mathscr{C}\}$ should be approximately 0.95 if correct coverage is being achieved.

Table I
Inventory Policies and Expected Average Cost per Period

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

For the standard procedures each inventory policy was simulated independently, i.e., a different random number stream was used to generate demands under each (s, S) policy. For the control-variate procedures a single random number stream was used to make demands identical for each policy, and the control variate was the sum of the demands for all 30 periods. IMSL subroutine drgivn was used to compute the control-variate estimators via least-squares regression. Simulations included all five policies ($r = 5$) or policies 3, 4 and 5 only ($r = 3$).

For MCA, we are interested in simultaneous confidence intervals for $\theta_i - \theta_j$ for all $i \neq j$. For MCB, we construct simultaneous confidence intervals for $\theta_i - \min_{j \neq i} \theta_j$, $i = 1, 2, \ldots, r$, by constructing simultaneous confidence intervals for $(-\theta_i) - \max_{j \neq i}(-\theta_j)$, $i = 1, 2, \ldots, r$, since we are interested in the system with the minimum expected cost. For MCC, we use inventory policy (20, 80) as the control system and construct simultaneous confidence intervals for $\theta_i - \theta_5$ for all $i \neq 5$.

Tables II and III give the results for $r = 3$ and 5, respectively. They show that both procedures appear to have $\Pr\{\mathscr{C}\} \approx 0.95$, but the control-variate procedures dominate the standard procedures in terms of the $\Pr\{\mathscr{C} \cap \mathscr{U}\}$. Notice that in both cases the probability of correct and useful inference is significantly lower than the coverage probability unless n is large, but the control-variate procedures can more than double $\Pr\{\mathscr{C} \cap \mathscr{U}\}$ for small values of n. The nearly zero probability of correct and useful inference for MCA when $r = 5$ emphasizes the value of using MCB when only comparisons with the best system design are of interest, which is typically the case in optimization problems.

Yang (1989) performed an extensive empirical comparison of the standard and control-variate, multiple-comparison procedures using linear models so that factors such as the number of control variates, q, dependence between the controls and response, R^2, and the difference between the mean responses, $\theta_i - \theta_j$, could be controlled. The results from those experiments support the conclusions drawn from the inventory system example.

7. DISCUSSION

We expect multiple-comparison procedures based on control-variate estimators to be better simply because the control-variate estimators are more precise. However, because of the loss in degrees of freedom, they may not be better when the sample size is too small. This paper establishes that simultaneous confidence intervals based on common random numbers and control-variate estimators can be constructed, and proves that they are better when the sample size is not too small and the assumptions of both models (1) and (2) apply (Theorem 4). However, both (1) and (2) will not hold in general.

Table II
Estimated $\Pr\{\mathscr{C}\}$ and $\Pr\{\mathscr{C} \cap \mathscr{U}\}$ for MCA, MCB, MCC, Respectively, for $r = 3$, $\alpha = 0.05$ and Inventory Policies $i = 3, 4, 5$. (The standard error of the estimates is ≤ 0.02.)

	$\Pr\{\mathscr{C}\}$						$\Pr\{\mathscr{C} \cap \mathscr{U}\}$					
	Standard			Control Variate			Standard			Control Variate		
n	MCA	MCB	MCC	MCA	MCB	MCC	MCA	MCB	MCC	MCA	MCB	MCC
10	0.94	0.97	0.95	0.93	0.96	0.91	0.03	0.13	0.05	0.12	0.27	0.14
20	0.96	0.97	0.95	0.95	0.95	0.92	0.09	0.20	0.11	0.24	0.42	0.27
30	0.94	0.96	0.95	0.94	0.96	0.92	0.12	0.30	0.15	0.37	0.59	0.41
40	0.95	0.97	0.95	0.93	0.96	0.91	0.17	0.35	0.20	0.48	0.69	0.51
50	0.96	0.97	0.95	0.93	0.96	0.91	0.25	0.43	0.29	0.58	0.79	0.63
60	0.95	0.95	0.94	0.94	0.95	0.93	0.31	0.51	0.35	0.69	0.86	0.72
70	0.95	0.96	0.95	0.93	0.95	0.92	0.37	0.56	0.43	0.77	0.89	0.78
80	0.94	0.97	0.94	0.95	0.96	0.93	0.41	0.62	0.46	0.80	0.92	0.83
90	0.94	0.96	0.94	0.94	0.96	0.92	0.47	0.65	0.50	0.86	0.96	0.87
100	0.95	0.97	0.95	0.93	0.96	0.91	0.54	0.73	0.59	0.88	0.96	0.88

Table III
Estimated $\Pr\{\mathscr{C}\}$ and $\Pr\{\mathscr{C} \cap \mathscr{U}\}$ for MCA, MCB, MCC, Respectively, for $r = 5$ and $\alpha = 0.05$. (The standard error of the estimates is $\leqslant 0.02$.)

	$\Pr\{\mathscr{C}\}$						$\Pr\{\mathscr{C} \cap \mathscr{U}\}$					
	Standard			Control Variate			Standard			Control Variate		
n	MCA	MCB	MCC	MCA	MCB	MCC	MCA	MCB	MCC	MCA	MCB	MCC
10	0.94	0.97	0.93	0.96	0.97	0.95	0.00	0.09	0.03	0.00	0.19	0.11
20	0.95	0.96	0.93	0.96	0.97	0.95	0.00	0.17	0.10	0.00	0.31	0.19
30	0.95	0.95	0.93	0.95	0.96	0.95	0.00	0.26	0.14	0.00	0.46	0.33
40	0.94	0.96	0.93	0.97	0.98	0.95	0.00	0.31	0.18	0.00	0.58	0.43
50	0.95	0.96	0.94	0.95	0.96	0.94	0.00	0.39	0.27	0.00	0.68	0.54
60	0.93	0.93	0.93	0.96	0.97	0.95	0.00	0.47	0.33	0.00	0.77	0.65
70	0.93	0.94	0.93	0.95	0.96	0.93	0.00	0.52	0.40	0.00	0.84	0.73
80	0.93	0.94	0.92	0.96	0.98	0.95	0.00	0.57	0.44	0.00	0.88	0.77
90	0.94	0.94	0.92	0.96	0.98	0.95	0.00	0.60	0.47	0.00	0.92	0.83
100	0.94	0.94	0.93	0.95	0.98	0.93	0.00	0.68	0.56	0.00	0.95	0.87

When the linear relationship (2) holds, we suspect that the equal residual variance assumption of model (2) is often less severely violated than the corresponding equal variance assumption of model (1), because model (1) is a special case of model (2) with $\boldsymbol{\beta}_i = \boldsymbol{\beta}_l$ for all i and l. Thus, if the linear relationship holds, the assumption of model (1) is stronger than that of model (2) in the sense that it implies that the dependence between the response variable and the control variates is the same for all systems.

On the other hand, if the linear relationship does not hold, then the control-variate point estimators are biased (Nelson 1990). However, if the relationship between the response and the controls is similar for each system design, then all of the control-variate point estimators may be biased in the same direction, so that taking differences may partially cancel the bias. For this reason, multiple-comparison procedures based on control-variate estimators are expected to be robust to deviation from the linearity assumption.

Of course, the validity of multiple-comparison inference also depends on distributional assumptions: marginally normal responses, in the case of model (1), and conditionally normal responses, in the case of model (2). Appropriate caution should be exercised when using procedures based on such strong assumptions.

Assuming that model (2) holds, the control-variate estimators are statistically independent. Thus, common random numbers have no direct effect on the estimators of the differences, $\hat{\theta}_i - \hat{\theta}_j$; that is, there is no variance reduction beyond what is achieved by using control variates alone. However, common random numbers make it possible to calculate the appropriate quantiles to form confidence intervals. On the other hand, common random numbers tend to reduce the variance of $\bar{Y}_i - \bar{Y}_j$. Unfortunately, under common random numbers we cannot construct simultaneous confidence intervals based on sample means $\bar{Y}_i$, $i = 1, 2, \ldots, r$ because the covariance structure of the estimators is unknown.

A limitation in our formulation is that the control variates $\mathbf{X}_j$, $j = 1, 2, \ldots, n$ must assume identical values across systems. This assumption is not necessary. The assumption guarantees equal conditional variance of the control-variate estimators for different systems, which makes the appropriate quantiles (e.g., $d^{\alpha}_{r-1,r(n-q-1)}$) easy to calculate. In general, we only need to know the ratios of the variances of estimators for different systems; equality is not required. More precisely, we only need equal residual variances for different systems under model (2), and estimators with a diagonal covariance matrix, in order to compute appropriate quantiles (Hayter 1989, Edwards and Hsu 1983). Extensions in this direction are under investigation.

The empirical results reported here and in Yang (1989) suggest that the control-variate, multiple-comparison procedures are superior, in the sense of having larger probability of correct and useful inference, when the sample size is not too small. We conjecture that, provided $R^2 > 0$ and models (1) and (2) hold, there exists an n^* such that the probability of correct and useful inference for the control-variate procedures is strictly larger than the corresponding probability for the standard procedures for all $n > n^*$. However, it is difficult to prove this conjecture because the probability of correct and useful inference of both procedures converges to the nominal coverage.

ACKNOWLEDGMENT

This research was partially supported by National Science Foundation grant No. ECS-8707634. The authors acknowledge the helpful comments of Jason C. Hsu and two anonymous referees.

REFERENCES

ANDERSON, T. W. 1984. *An Introduction to Multivariate Statistical Analysis*. John Wiley, New York.

CLARK, G. M., AND W. YANG. 1986. A Bonferroni Selection Procedure When Using Common Random Numbers With Unknown Variances. *Proceedings of the 1986 Winter Simulation Conference*, 313–315.

DUNNETT, C. W. 1955. A Multiple Comparison Procedure for Comparing Several Treatments With a Control. *J. Am. Statist. Assoc.* **50,** 1096–1121.

EDWARDS, D. G., AND J. C. HSU. 1983. Multiple Comparisons With the Best Treatment. *J. Am. Statist. Assoc.* **78,** 965–971.

GRAY, G., AND D. GOLDSMAN. 1988. Indifference-Zone Selection Procedures for Choosing the Best Airspace Configuration. *Proceedings of the 1988 Winter Simulation Conference*, 445–450.

HAYTER, A. J. 1989. Pairwise Comparisons of Generally Correlated Means. *J. Am. Statist. Assoc.* **84,** 208–213.

HOCHBERG, Y., AND A. C. TAMHANE. 1987. *Multiple Comparison Procedures*. John Wiley, New York.

HSU, J. C. 1984. Constrained Simultaneous Confidence Intervals for Multiple Comparisons With the Best. *Ann. Statist.* **12,** 1136–1144.

HSU, J. C., AND B. L. NELSON. 1988. Optimization Over a Finite Number of System Designs With One-Stage Sampling and Multiple Comparisons With the Best. *Proceedings of the 1988 Winter Simulation Conference*, 451–457.

KOENIG, L. W., AND A. M. LAW. 1985. A Procedure for Selecting a Subset of Size m Containing the l Best of k Independent Normal Populations, With Applications to Simulation. *Commun. Statist.* **B14,** 719–734.

LAVENBERG, T. L., AND P. D. WELCH. 1981. A Perspective on the Use of Control Variables to Increase the Efficiency of Monte Carlo Simulations. *Mgmt. Sci.* **27,** 322–335.

MILLER, R. G. 1981. *Simultaneous Statistical Inference*, 2nd ed. Springer-Verlag, New York.

NELSON, B. L. 1990. Control-Variate Remedies. *Opns. Res.* **38,** 974–992.

RAO, C. R. 1973. *Linear Statistical Inference and Its Applications*. John Wiley, New York.

TUKEY, J. W. 1953. The Problem of Multiple Comparisons. Unpublished manuscript.

YANG, W. 1989. Multivariate Estimation and Variance Reduction for Terminating and Steady-State Simulation. Ph.D. Dissertation, Department of Industrial and Systems Engineering, Ohio State University, Columbus.

YANG, W., AND B. L. NELSON. 1989. Optimization Using Common Random Numbers, Control Variates and Multiple Comparisons With the Best. *Proceedings of the 1989 Winter Simulation Conference*, 444–449.