Robust Multiple Comparisons Under Common Random Numbers

BARRY L. NELSON
The Ohio State University

In this article we show how a result used in the analysis of repeated-measures experiments can aid in the analysis of simulation experiments that employ common random numbers. We specifically consider the statistical procedure known as multiple comparisons with the best. We first establish when the proposed procedure provides exact inference, and then show that it is typically robust when it is not exact. The method is easy to apply in practice.

Categories and Subject Descriptors: G.3 [**Probability and Statistics**]; I.6.1 [**Simulation Theory**]

General Terms: Experimentation, Theory

Additional Key Words and Phrases: Common random numbers, computer simulation, multiple-comparison procedures, statistics, variance reduction

1. INTRODUCTION

Common random numbers (CRN) is a variance reduction technique that decreases the variance of estimators of the differences among the expected performances of two or more systems. CRN works by inducing positive dependence across the simulation outputs (responses) from all systems. Unfortunately, when there are more than two systems, accounting for the induced dependence in the statistical inference is difficult and has been a longstanding problem.

In Yang and Nelson [1991] and Nelson and Hsu [1993] we proposed control-variate models of CRN. These models account for the effects of CRN via a linear regression of the simulation outputs on functions of the simulation inputs called control variates. When these models pertain they facilitate exact multiple-comparison inference under CRN for two or more systems; by "exact" we mean that no conservative probabilistic inequalities, such as the Bonferroni inequality, are required. Nelson [1992] provides implementation details for control-variate-based multiple comparisons.

In some situations the control-variate models are not adequate, in the sense that some dependence across systems remains after regressing on the

---

This work is based on work supported by the National Science Foundation grant DDM-8922721.
Author's address: Department of Industrial and Systems Engineering, The Ohio State University, Columbus, OH 43210.

control variates. Also, the control-variate approach is difficult to apply in practice because of the need to select appropriate control variates, record their values, and perform a least-squares regression. In this article we show that a result used in the analysis of repeated-measures experiments provides an easy-to-apply, robust procedure that can be used with or without the control-variate model.

We first review models of CRN, and then establish sufficient conditions for multiple-comparison inference (Sections 2 and 3). Section 4 defines the property of sphericity, which is used in the analysis of repeated-measures experiments, and which we exploit to model CRN. The new multiple-comparison procedures that account for CRN are given in Section 5, and Section 6 establishes the robustness of these procedures.

2. MODELS OF CRN

We assume that the goal of the simulation experiment is to compare elements of the vector parameter $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_r)'$, where θ_i is the expected performance of system i. Suppose that larger expected performance implies a better system. For system i, we consider the parameter $\theta_i - \max_{\ell \neq i} \theta_\ell$, which is system i performance minus the best of the other systems' performance. In optimization problems, the parameters $\theta_i - \max_{\ell \neq i} \theta_\ell$, for $i = 1, 2, \ldots, r$, are often the parameters of primary interest. If $\theta_i - \max_{\ell \neq i} \theta_\ell > 0$, then system i is the best. But even if $\theta_i - \max_{\ell \neq i} \theta_\ell < 0$, if $\theta_i - \max_{\ell \neq i} \theta_\ell > -\epsilon$, for $\epsilon > 0$, then system i is within ϵ of the best. Simultaneous statistical inference on $\theta_i - \max_{\ell \neq i} \theta_\ell$, for $i = 1, 2, \ldots, r$, is termed *multiple comparisons with the best* (MCB).

We focus on MCB because of its close relationship to indifference-zone ranking and subset selection (see Hsu and Nelson [1988]). However, all of the results in this article apply to one-sided multiple comparisons with a control ($\theta_i - \theta_r, \forall i \neq r$), and most apply to all-pairwise multiple comparisons ($\theta_i - \theta_\ell$, $\forall i, \ell; i \neq \ell$).

Let Y_{ij} be the output from system i on the jth independent replication of the simulation, for $i = 1, 2, \ldots, r$ and $j = 1, 2, \ldots, n$. We assume that $E[Y_{ij}] = \theta_i$. Let $\mathbf{C}_{ij}$ be a $q_i \times 1$ vector of (possibly functions of) simulation input random variables from the jth replication of system i, and let $\boldsymbol{\mu}_i \equiv \mathrm{E}[\mathbf{C}_{ij}]$; we assume that $\boldsymbol{\mu}_i$ is known since the simulator specifies the distribution of the simulation inputs.

We approximate the relationship between the simulation inputs and outputs by a linear model with unknown parameters, specifically

$$Y_{ij} = \theta_i + (\mathbf{C}_{ij} - \boldsymbol{\mu}_i)' \boldsymbol{\beta}_i + \eta_{ij} \tag{1}$$

where $\boldsymbol{\beta}_i$ is a $q_i \times 1$ vector of unknown parameters, and $\eta_{i1}, \eta_{i2}, \ldots, \eta_{in}$ are i.i.d. normal random variables. The $\mathbf{C}_{ij}$ are called *control variates* in the variance reduction literature (e.g., Nelson [1990]). A point estimator of θ_i is obtained by regressing the outputs Y_{ij} on the control variates $\mathbf{C}_{ij} - \boldsymbol{\mu}_i$.

Let $\mathbf{I}_r$ denote the $r \times r$ identity matrix. If $\boldsymbol{\eta}_{\cdot j} \equiv (\eta_{1j}, \eta_{2j}, \ldots, \eta_{rj})'$ is distributed $\mathrm{N}(\mathbf{0}, \tau^2 \mathbf{I}_r)$—in other words, the residuals across systems are independent with common variance—then exact MCB inference can be derived

(Yang and Nelson [1991], Nelson and Hsu [1993]). This condition implies that all of the dependence across systems due to CRN is explained by the control variates, which is plausible since CRN directly induces dependence between the inputs that is translated to the outputs, and model (1) approximates the relationship between inputs and outputs.

In this article we extend our results to the case where $\boldsymbol{\eta}_{\cdot 1}, \boldsymbol{\eta}_{\cdot 2}, \ldots, \boldsymbol{\eta}_{\cdot n}$ are distributed i.i.d. $\mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\eta}})$, where $\boldsymbol{\Sigma}_{\boldsymbol{\eta}}$ is not necessarily $\tau^2 \mathbf{I}_r$; that is, the control variates do not capture all of the dependence due to CRN. In addition to the general model (1), we consider several special cases:

(1) No control variates are employed ($q_i = 0$, $\forall i$).
(2) The control variates are common across all systems ($\mathbf{C}_{ij} = \mathbf{C}_j$, $\forall i$).
(3) The control variates are common across all systems ($\mathbf{C}_{ij} = \mathbf{C}_j$, $\forall i$) and have a common relationship to the simulation response ($\boldsymbol{\beta}_i = \boldsymbol{\beta}$, $\forall i$).

Case 1 is of particular importance since no control variates are required. In cases 2 and 3, we assume that the effect of CRN is to cause the control variates to take identical values across all systems. When the control variates are not common, we assume that the effect of CRN is to make then dependent across systems, but not necessarily identical.

3. SUFFICIENT CONDITIONS FOR MCB

Suppose that $\hat{\boldsymbol{\theta}}$ is a point estimator of θ. Critical to the derivation of MCB inference is the distribution of

$$\mathbf{D}^{(i)}\hat{\boldsymbol{\theta}} = \begin{bmatrix} \hat{\theta}_1 - \hat{\theta}_i \\ \hat{\theta}_2 - \hat{\theta}_i \\ \vdots \\ \hat{\theta}_{i-1} - \hat{\theta}_i \\ \hat{\theta}_{i+1} - \hat{\theta}_i \\ \vdots \\ \hat{\theta}_r - \hat{\theta}_i \end{bmatrix}$$

where $\mathbf{D}^{(i)}$ is the $(r-1) \times r$ matrix obtained by inserting the $r \times 1$ column vector of -1's, denoted $-\mathbf{1}_{\mathbf{r}-\mathbf{1}}$, between the $(i-1)$st and the ith columns of $\mathbf{I}_{\mathbf{r}-\mathbf{1}}$. A sufficient condition for constructing MCB intervals is that

$$\mathbf{D}^{(i)}\hat{\boldsymbol{\theta}} \sim N\left(\begin{bmatrix} \theta_1 - \theta_i \\ \theta_2 - \theta_i \\ \vdots \\ \theta_{i-1} - \theta_i \\ \theta_{i+1} - \theta_i \\ \vdots \\ \theta_r - \theta_i \end{bmatrix}, \tau^2 \boldsymbol{\Omega}^{(i)} \right) \tag{2}$$

with $\boldsymbol{\Omega}^{(i)}$ known, and that there exists an estimator $\hat{\tau}^2$ of τ^2 such that $\hat{\tau}^2 \sim \tau^2\chi_\nu^2/\nu$ and is independent of $\hat{\boldsymbol{\theta}}$, where χ_ν^2 denotes the chi-squared distribution with ν degrees of freedom. See Nelson and Hsu [1993] for details.

Let $\boldsymbol{\Xi}^{(i)}$ be the correlation matrix associated with $\tau^2\boldsymbol{\Omega}^{(i)}$. From a computational standpoint, an additional condition for the formation of MCB intervals is that $\boldsymbol{\Xi}^{(i)}$ has so-called structure l (Tong [1980]):

$$\boldsymbol{\Xi}^{(i)} = \begin{bmatrix} 1 & \lambda_1^{(i)}\lambda_2^{(i)} & \cdots & \lambda_1^{(i)}\lambda_{r-1}^{(i)} \\ \lambda_2^{(i)}\lambda_1^{(i)} & 1 & \cdots & \lambda_2^{(i)}\lambda_{r-1}^{(i)} \\ & & \ddots & \\ \lambda_{r-1}^{(i)}\lambda_1^{(i)} & \lambda_{r-1}^{(i)}\lambda_2^{(i)} & \cdots & 1 \end{bmatrix} \tag{3}$$

where $\lambda_\ell^{(i)} \in (-1, 1)$. This condition insures that it is possible to numerically determine certain critical constants used in forming the intervals; see Nelson and Hsu [1993] for details.

4. SPHERICITY

We say that $\boldsymbol{\Sigma}_\eta$, the variance-covariance matrix $\boldsymbol{\eta}_{\cdot j}$, has the property of *sphericity* if it can be represented as

$$\boldsymbol{\Sigma}_\eta = \begin{bmatrix} 2\psi_1 + \tau^2 & \psi_1 + \psi_2 & \cdots & \psi_1 + \psi_r \\ \psi_2 + \psi_1 & 2\psi_2 + \tau^2 & \cdots & \psi_2 + \psi_r \\ & & \ddots & \\ \psi_r + \psi_1 & \psi_r + \psi_2 & \cdots & 2\psi_r + \tau^2 \end{bmatrix} \tag{4}$$

where $\tau^2 > \sqrt{r\sum_{i=1}^r \psi_i^2} - \sum_{i=1}^r \psi_i$ (the last condition insures that the matrix is positive definite). Sphericity is a generalization of *compound symmetry*, in which

$$\boldsymbol{\Sigma}_\eta = \sigma^2 \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{bmatrix}. \tag{5}$$

Several researchers have assumed that the effect of CRN is to induce a covariance structure satisfying compound symmetry with $\rho > 0$ (for example, Schruben and Margolin [1978], Nozari et al. [1987], and Tew and Wilson [1993]). Notice that sphericity includes independence (no CRN) as a special case. The assumption of sphericity is frequently used to account for the dependence among repeated measurements on a single subject in settings such as clinical trials.

Let $\bar{\eta}_{i\cdot} = n^{-1}\sum_{j=1}^{n}\eta_{ij}$, $\bar{\eta}_{\cdot j} = r^{-1}\sum_{i=1}^{r}\eta_{ij}$, and $\bar{\eta}_{\cdot\cdot} = (rn)^{-1}\sum_{i=1}^{r}\sum_{j=1}^{n}\eta_{ij}$, where a · subscript denotes summing with respect to that subscript. Define

$$\bar{\tau}^2 = \frac{\sum_{i=1}^{r}\sum_{j=1}^{n}\left(\eta_{ij} - \bar{\eta}_{i\cdot} - \bar{\eta}_{\cdot j} + \bar{\eta}_{\cdot\cdot}\right)^2}{(r-1)(n-1)}.$$

We will exploit and extend the following result, which is often used in the analysis of repeated-measures experiments (see, for example, Hochberg and Tamhane [1987]):

THEOREM 4.1. *If $\boldsymbol{\eta}_{\cdot 1}, \boldsymbol{\eta}_{\cdot 2}, \ldots, \boldsymbol{\eta}_{\cdot n}$ are distributed i.i.d. $N(\mathbf{0}, \boldsymbol{\Sigma}_\eta)$, and $\boldsymbol{\Sigma}_\eta$ has the property of sphericity* (4), *then $\bar{\tau}^2 \sim \tau^2\chi^2_{(r-1)(n-1)}/((r-1)(n-1))$ and is independent of $\bar{\eta}_{1\cdot}, \bar{\eta}_{2\cdot}, \ldots, \bar{\eta}_{r\cdot}$.*

5. SPHERICITY AND MCB UNDER CRN

In the following we derive exact MCB procedures under the assumption that $\boldsymbol{\Sigma}_\eta$ has the property of sphericity (4). In Section 6 we argue that these procedures should be robust to departures from sphericity.

5.1 Sphericity and the One-Way Model

For ease of presentation, and because of its practical importance, we first consider the case of no control variates ($q_i = 0$, $\forall i$), which reduces the control-variate model (1) to the one-way analysis-of-variance model

$$Y_{ij} = \theta_i + \eta_{ij}. \tag{6}$$

Suppose that the effect of CRN is to cause $\boldsymbol{\Sigma}_\eta$ to have a structure satisfying sphericity (4). We show that exact MCB inference can be based on the sample mean vector, $\bar{\boldsymbol{Y}} \equiv (\bar{Y}_{1\cdot}, \bar{Y}_{2\cdot}, \ldots, \bar{Y}_{r\cdot})'$.

First notice that

$$\frac{\sum_{i=1}^{r}\sum_{j=1}^{n}\left(Y_{ij} - \bar{Y}_{i\cdot} - \bar{Y}_{\cdot j} + \bar{Y}_{\cdot\cdot}\right)^2}{(r-1)(n-1)} = \frac{\sum_{i=1}^{r}\sum_{j=1}^{n}\left(\eta_{ij} - \eta_{i\cdot} - \eta_{\cdot j} + \eta_{\cdot\cdot}\right)^2}{(r-1)(n-1)} = \bar{\tau}^2$$

so that $\bar{\tau}^2 \sim \tau^2\chi^2_{(r-1)(n-1)}/((r-1)(n-1))$ and is independent of $\bar{\boldsymbol{Y}}$ by Theorem 4.1. By direct calculation we can show that

$$\mathbf{D}^{(i)}\bar{\mathbf{Y}} \sim \mathrm{N}\left(\begin{bmatrix} \theta_1 - \theta_i \\ \theta_2 - \theta_i \\ \vdots \\ \theta_{i-1} - \theta_i \\ \theta_{i+1} - \theta_i \\ \vdots \\ \theta_r - \theta_i \end{bmatrix}, \tau^2\boldsymbol{\Omega}^{(i)}\right)$$

where $\mathbf{\Omega}^{(i)} = 2/n\mathbf{\Xi}^{(i)}$, and

$$\mathbf{\Xi}^{(i)} = \begin{bmatrix} 1 & \frac{1}{2} & \cdots & \frac{1}{2} \\ \frac{1}{2} & 1 & \cdots & \frac{1}{2} \\ & & \ddots & \\ \frac{1}{2} & \frac{1}{2} & \cdots & 1 \end{bmatrix}. \tag{7}$$

Thus, $\mathbf{\Xi}^{(i)}$ has structure l with $\lambda_{\ell}^{(i)} = 1/\sqrt{2}$, and the sufficient conditions for MCB inference are satisfied by $\bar{\mathbf{Y}}$ and $\tilde{\tau}^2$. The corresponding MCB procedure is a simple modification of standard MCB (see Appendix A).

The obvious limitation of this result is that CRN must induce a covariance matrix $\mathbf{\Sigma}_\eta$ that satisfies sphericity. Empirical measures exist to check whether a sample covariance matrix satisfies sphericity (for example, Grieve and Ag [1984]), but we argue in Section 6 that the assumption of sphericity is robust whenever the off-diagonal elements of $\mathbf{\Sigma}_\eta$ are positive, the standard assumption behind CRN.

5.2 Sphericity and Control Variates

We now assume that the control-variate model (1) holds with $q_i > 0$, except that $\mathbf{\Sigma}_\eta = \mathrm{Var}[\boldsymbol{\eta}_{\cdot j}] \equiv \|\sigma_{i\ell}\|$ might not equal $\tau^2 \mathbf{I}_r$. In this way we attempt to capture any residual dependence due to CRN that is not explained by the control variates, and thereby extend the range of applicability of the control-variate model.

Let $\hat{\boldsymbol{\theta}} \equiv (\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_r)'$ denote the $r \times 1$ vector of control-variate estimators of $\boldsymbol{\theta}$, and let $\mathbf{C}$ denote the collection of all the control variates $\mathbf{C}_{ij}$, $i = 1, 2, \ldots, r$ and $j = 1, 2, \ldots, n$. For convenience of exposition we assume that each system has the same number of control variates ($q_1 = q_2 = \cdots = q_r = q$), but unless explicitly stated the results that follow do not depend on this assumption. The calculation of $\hat{\boldsymbol{\theta}}$ is described in Appendix A.

The MCB intervals associated with the control-variate model are constructed by conditioning on the control variates, $\mathbf{C}$, then showing that the resulting inference is independent of $\mathbf{C}$ (Nelson and Hsu [1993]). Therefore, we need to investigate the conditional distribution of $\mathbf{D}^{(i)}\hat{\boldsymbol{\theta}}$ given $\mathbf{C}$.

From Nelson [1990], we find that $\hat{\theta}_i = \mathbf{M}_i'\mathbf{Y}_{i\cdot}$, where $\mathbf{Y}_{i\cdot} \equiv (Y_{i1}, Y_{i2}, \ldots, Y_{in})'$ and $\mathbf{M}_i$ is an $n \times 1$ vector that depends only on $\mathbf{C}_{i1}, \mathbf{C}_{i2}, \ldots, \mathbf{C}_{in}$. This result can be used to show that $\mathrm{E}[\hat{\boldsymbol{\theta}}] = \boldsymbol{\theta}$, and that $\hat{\boldsymbol{\theta}}$ is conditionally normally distributed. The $\mathrm{Var}[\mathbf{D}^{(i)}\hat{\boldsymbol{\theta}}|\mathbf{C}]$ is given in the following lemma:

LEMMA 5.1. *If model* (1) *pertains then* $\mathrm{Var}[\hat{\boldsymbol{\theta}}|\mathbf{C}] = \|\mathbf{M}_i'\mathbf{M}_\ell \sigma_{i\ell}\|$.

As a consequence of Lemma 5.1, if the control variates are common across all systems ($\mathbf{C}_{ij} = \mathbf{C}_j$, $\forall i$), so that $\mathbf{M}_i = \mathbf{M}$ for all i, then $\mathrm{Var}[\hat{\boldsymbol{\theta}}|\mathbf{C}] = \mathbf{M}'\mathbf{M}\mathbf{\Sigma}_\eta$. Thus, $\mathrm{Var}[\hat{\boldsymbol{\theta}}|\mathbf{C}]$ will satisfy the conditions of sphericity exactly if $\mathbf{\Sigma}_\eta$ does. By direct calculation we can show that $\text{Var}[\mathbf{D}^{(i)}\hat{\boldsymbol{\theta}}|\mathbf{C}] = \tau^2\boldsymbol{\Omega}^{(i)} = \tau^2(2\mathbf{M}'\mathbf{M})\boldsymbol{\Xi}^{(i)}$, where $\boldsymbol{\Xi}^{(i)}$ is given in (7).

For the control-variate model the natural generalization of $\bar{\tau}^2$ as an estimator of τ^2 is

$$\bar{\tau}^2 = \frac{1}{r-1}\sum_{i=1}^{r}\frac{1}{n-q_i-1}\sum_{j=1}^{n}\left(\hat{\eta}_{ij} - \bar{\hat{\eta}}_{i\cdot} - \bar{\hat{\eta}}_{\cdot j} + \bar{\hat{\eta}}_{\cdot\cdot}\right)^2$$

where the $\hat{\eta}_{ij}$ are the estimated residuals from the least-squares regression (see Appendix B). The following theorem shows that this is exactly the right modification in the case of common control variates.

THEOREM 5.1. *If model* (1) *pertains, the control variates are common across all systems, and* $\boldsymbol{\Sigma}_\eta$ *satisfies sphericity* (4), *then, conditional on* $\mathbf{C}$, $\bar{\tau}^2 \sim \tau^2\chi^2_{(r-1)(n-q-1)}/((r-1)(n-q-1))$ *and is independent of* $\hat{\boldsymbol{\theta}}$.

Thus, conditional on $\mathbf{C}$, $\hat{\boldsymbol{\theta}}$ and $\bar{\tau}^2$ satisfy the sufficient conditions for MCB when the control variates are common across all systems.

Lemma 5.1 shows how the $\text{Var}[\hat{\boldsymbol{\theta}}|\mathbf{C}]$ depends on $\boldsymbol{\Sigma}_\eta$ in general. The following theorem establishes that, for large n, $\text{Var}[\hat{\boldsymbol{\theta}}|\mathbf{C}] \approx 1/n\boldsymbol{\Sigma}_\eta$ even if the control variates are not common; that is, for n large the conditional variance of $\hat{\boldsymbol{\theta}}$ approximately satisfies sphericity if $\boldsymbol{\Sigma}_\eta$ does.

THEOREM 5.2. *For all i and ℓ, $n\mathbf{M}_i'\mathbf{M}_\ell\sigma_{i\ell} \xrightarrow{P} \sigma_{i\ell}$ as $n \to \infty$, where $\xrightarrow{P}$ denotes convergence in probability.*

Although Theorem 5.1 does not hold if the control variates differ across systems, Theorem 5.2 implies that $\bar{\tau}^2$ will be approximately correct for large n.

5.3 The Special Case of Common Controls and Multiplier

Suppose that the control variates are common across all systems ($\mathbf{C}_{ij} = \mathbf{C}_j$, $\forall i$), have common relationship to the simulation response ($\boldsymbol{\beta}_i = \boldsymbol{\beta}, \forall i$), and $\boldsymbol{\Sigma}_\eta = \tau^2\mathbf{I}_r$. Nelson and Hsu [1993] showed that MCB inference can be derived based on the control-variate estimator $\hat{\boldsymbol{\theta}}$ associated with this model, and a variance estimator $\dot{\tau}^2$ with $rn - r - q$ degrees of freedom. A key difference in this case is that the parameters $(\boldsymbol{\theta}, \boldsymbol{\beta})$ are estimated from a single overall regression, rather than from r individual regressions for each system (see Appendix A).

When these conditions hold they imply that

$$\text{Var}[\mathbf{Y}_{\cdot j}] = \tau^2\mathbf{I}_r + \phi\mathbf{1}_r\mathbf{1}_r' \tag{8}$$

where $\mathbf{Y}_{\cdot j} \equiv (Y_{1j}, Y_{2j}, \ldots, Y_{rj})'$ and $\phi \equiv \text{Var}[\mathbf{C}_j'\boldsymbol{\beta}]$. Since (8) satisfies the conditions of sphericity with $\psi_j = \phi/2$ for all j, we could also construct MCB intervals based on the one-way model (6) using the sample mean $\bar{Y}$ and $\bar{\tau}^2$ with $(r-1)(n-1)$ degrees of freedom. Since both the control-variate approach and the sample-mean approach apply in this special case, which one is better?

Nelson and Hsu [1993] showed that, under these conditions, $\mathrm{Var}[\bar{Y}_{i\cdot} - \bar{Y}_{\ell\cdot}] = \mathrm{Var}[\hat{\theta}_i - \hat{\theta}_\ell] = 2\tau^2/n$, so the point estimators are of equal quality. The correlation structures of $\hat{\boldsymbol{\theta}}$ and $\bar{\mathbf{Y}}$ are also identical, so the critical values for MCB inference will differ only because the degrees of freedom differ. Notice that $rn - r - q > (r-1)(n-1)$, provided $q < n - 1$; since $q < (n-1)$ is required to use the control-variate model, the control-variate model is preferred in this special case where both are applicable because it provides more degrees of freedom for inference.

6. ROBUSTNESS

In this section we argue that we can expect MCB intervals formed under the assumption of sphericity to be robust to departures from sphericity. The discussion will be primarily in terms of the one-way model (6), but the results carry over directly to the control-variate models.

6.1 Mathematical Justification

As a basis for comparison, let

$$\ddot{\tau}^2 = \frac{\sum_{i=1}^{r}\sum_{j=1}^{n}\left(Y_{ij} - \bar{Y}_{i\cdot}\right)^2}{r(n-1)}$$

which is the usual pooled variance estimator used in multiple-comparison procedures, such as MCB (see Appendix A). This is the estimator we would use if we simulated the systems independently or if we ignored CRN.

Let $\varsigma = r^{-1}\sum_{i=1}^{r}\sigma_{ii}$, the average marginal variance of the observations across systems, and let $\varrho = (r(r-1))^{-1}\sum_{i \neq \ell}\sigma_{i\ell}$, the average marginal covariance across systems.

THEOREM 6.1. *For* $\boldsymbol{\Sigma}_\eta = \|\sigma_{i\ell}\|$,

$$E[\tilde{\tau}^2] = \varsigma\left(1 - \frac{\varrho}{\varsigma}\right) \tag{9}$$

and

$$E[\ddot{\tau}^2] = \varsigma. \tag{10}$$

Theorem 6.1 shows that the usual variance estimator, $\ddot{\tau}^2$, estimates the average marginal variance, ς, while $\tilde{\tau}^2$ adjusts by something like an average correlation, ϱ/ς. In fact, if $\sigma_{ii} = \sigma^2$ for all i, then ϱ/ς is precisely the average correlation. Therefore, the more nearly equal the covariances of pairs of systems, the better $\tilde{\tau}^2$ will approximate the true variance. Notice also that, when the covariances are positive, $\tilde{\tau}^2$ will be smaller—and the corresponding confidence intervals shorter and less conservative—than if we use $\ddot{\tau}^2$.

Of course, if $\boldsymbol{\Sigma}_\eta$ does not exactly satisfy sphericity then the proofs of confidence interval validity in Section 5 break down. Next we demonstrate empirically that the intervals can be expected to be robust.

6.2 Empirical Justification

Before presenting the results of a controlled study, we describe a system simulation example to serve as motivation. Nelson and Hsu [1993] described the simulation of five (s, S) inventory policies to determine the one with minimum expected cost/period over a planning horizon of 30 periods (notice that this is a minimization problem, rather than a maximization problem). The demand for product was the only stochastic input process, so the standardized average demand was used as a control variate; the standardized average is the sample average divided by the standard deviation. An experiment consisted of $n = 30$ replications, each of 30 periods, after which 95% MCB intervals were formed. The entire experiment was repeated 1000 times to estimate the probability of coverage (which should be 0.95), and the probability of correct and useful inference, where "correct and useful" means that the intervals cover $\theta_i - \min_{\ell \neq i} \theta_\ell$ and reveal differences when they really exist (which should be as close to 0.95 as possible). The desired effect of CRN is to raise the probability of correct and useful inference without degrading coverage (increasing n will also increase the probability of correct and useful inference, but with increased computing cost).

Results from Nelson and Hsu [1993] are displayed on the first two lines of Table I. The one-way model ignores the effect of CRN, leading to over coverage and a low probability of correct and useful inference. The control-variate model assumes that the single control variate entirely explains the effect of CRN; it works well in this example, nearly tripling the probability of correct and useful inference.

The last two lines of Table I display results for the one-way and control-variate models using $\bar{\tau}^2$ and $\bar{\tau}^2$, respectively; that is, assuming that $\mathbf{\Sigma}_\eta$ satisfies sphericity. The performance is also good for these models, meaning that the nominal coverage is maintained or nearly maintained along with a large probability of correct and useful inference. This is particularly satisfying for the one-way model because it is very easy to apply relative to the control-variate model.

Was this example particularly well suited for the assumption of sphericity? We estimated $\mathbf{\Sigma}_\eta$ for each model from 5000 replications, then calculated Grieve and Ag's [1984] ε measure of sphericity; ε takes values between 0 and 1, with 1 indicating perfect conformance to sphericity. The values of ε were 0.53 and 0.50 for the one-way and control-variate models, respectively (recall that for the control-variate model $\mathbf{\Sigma}_\eta$ is the variance-covariance matrix of the residuals after removing the effect of the control variates). Both values indicate significant departures from sphericity. This example is representative of what we have observed in many system simulation examples: despite significant departures from sphericity, multiple-comparison procedures based on $\bar{\tau}^2$ show robust performance.

From Theorem 6.1, we expect confidence intervals based on $\bar{\tau}^2$ to be shorter, and therefore sharper, than those based on $\ddot{\tau}^2$ in the presence of positive dependence. However, it is not obvious that these shorter intervals will maintain the nominal coverage probability, because the joint distribution

Table I. Results for the Inventory Model

model	coverage	correct and useful inference
one way with $\ddot{\tau}^2$	1.00	0.15
control variate with $\hat{\tau}^2$	0.95	0.43
one way with $\tilde{\tau}^2$	0.94	0.48
control variate with $\tilde{\bar{\tau}}^2$	0.93	0.51

of $\bar{\mathbf{Y}}$ and $\tilde{\bar{\tau}}^2$ depends critically on the assumption of sphericity. Since it is very difficult to find system simulation examples that systematically depart from this assumption, we designed the distribution-sampling experiment described below to obtain an idea of the robustness of the approximation over the range of possible cases.

Let a subscript (i) represent the unknown index of the ith smallest θ_i; e.g., $\theta_{(r)}$ is the largest system performance parameter. Define the event

$$\mathscr{E} \equiv \left\{ \bar{Y}_{i\cdot} - \bar{Y}_{(r)\cdot} - (\theta_i - \theta_{(r)}) \leq d^{1-\alpha}_{r-1,(r-1)(n-1)} \tilde{\bar{\tau}} \sqrt{\frac{2}{n}}, \ \forall i \neq (r) \right\}$$

where $d^{1-\alpha}_{r-1,(r-1)(n-1)}$ is the appropriate critical value for $(1-\alpha)100\%$ MCB confidence intervals under the assumption of sphericity (see Appendix A). The event $\mathscr{E}$ is analogous to one-sided multiple comparisons with a control, where the control system is the unknown best system. Let $\mathscr{C}$ be the event that the MCB intervals cover $\theta_i - \max_{\ell \neq i} \theta_\ell, \forall i$. Hsu and Nelson [1988] showed the $\Pr\{\mathscr{C}\} \geq \Pr\{\mathscr{E}\}$. Our experiments estimate $\Pr\{\mathscr{E}\}$ over the space of correlation matrices Ξ with positive off-diagonal elements. We chose this performance measure because $\Pr\{\mathscr{E}\}$ does not depend on $\boldsymbol{\theta}$, while $\Pr\{\mathscr{C}\}$ does.

The experiments were conducted as follows:

(1) Fix the number of systems, r, number of replications from each system, n, and the confidence level, $1-\alpha$. We considered $r = 3$, 5, and 10 systems; $n = 10$ and 30 replications; and $1 - \alpha = 0.95$.

(2) Generate a random r-dimensional correlation matrix Ξ using the method of Marsaglia and Olkin ([1984], p. 471). This method transforms a randomly generated point on the r-dimensional unit sphere into a correlation matrix. We modified the method to generate a point on the unit sphere with all positive coordinates, which leads to a correlation matrix with all positive elements.

(3) Generate n i.i.d. random vectors $\mathbf{Y}_{\cdot j} \sim \mathrm{N}(\mathbf{0}, \Xi)$, $j = 1, 2, \ldots, n$, and score a "hit" if $\mathscr{E}$ occurs; specifically, if $\{\bar{Y}_{i\cdot} - \bar{Y}_{r\cdot} \leq d^{1-\alpha}_{r-1,(r-1)(n-1)} \tilde{\bar{\tau}} \sqrt{2/n}, \ \forall i \neq r\}$.

(4) Repeat step 3 a total of 5000 times to obtain an estimate of $\Pr\{\mathscr{E}\}$, denoted $\widehat{\Pr}\{\mathscr{E}\}$ (this gives two significant digits of precision).

(5) Repeat steps 2–4 a total of 1000 times to estimate the distribution of $\Pr\{\mathscr{E}\}$ over a random sample of correlation matrices, Ξ.

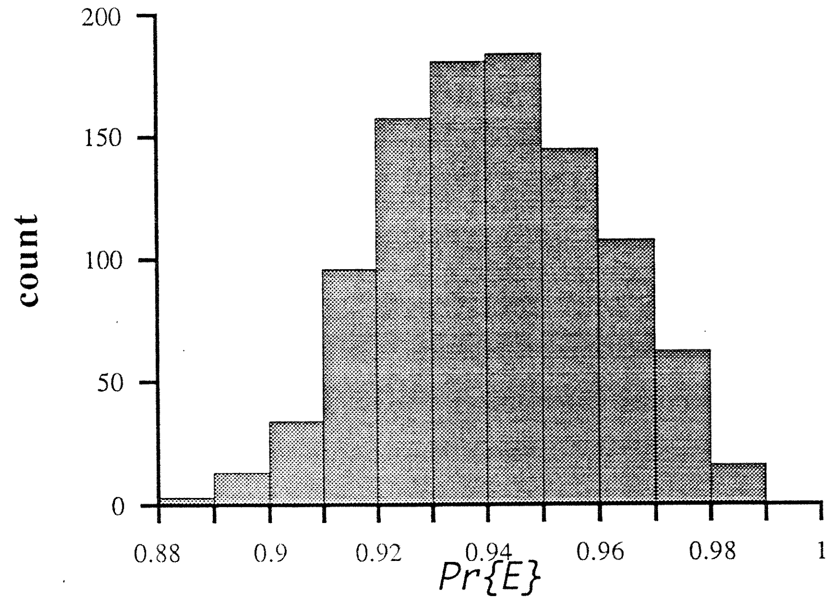

Fig. 1. Estimated coverage probability $\widehat{\Pr}\{\mathscr{E}\}$ for $r = 5$ systems and $n = 30$ replications.

This experiment bypasses two problems that affect all parametric multiple-comparison procedures—nonnormal data and heteroscedastic data—and focuses on the effect of positive correlation. The results are therefore optimistic in the same way that any parametric multiple-comparison procedure is optimistic with regard to these assumptions. However, the results are pessimistic in the sense that typically $\Pr\{\mathscr{C}\} > \Pr\{\mathscr{E}\}$, but we are estimating $\Pr\{\mathscr{E}\}$.

Since the results were nearly identical for all cases of r and n, we only present the single case $r = 5$ and $n = 30$. Over the 1000 generated correlation matrices, the minimum, mean, and maximum values of $\widehat{\Pr}\{\mathscr{E}\}$ were 0.89, 0.94, and 0.99, respectively; a histogram is given in Figure 1. Most of the randomly generated correlation matrices departed significantly from the assumption of sphericity; a histogram of the ε values for the 1000 correlation matrices is given in Figure 2. Figure 3 is a scatter plot of ε and $\widehat{\Pr}\{\mathscr{E}\}$. Notice that a very large value of ε is associated with coverage of about 0.95, but other values of ε can be associated with undercoverage, overcoverage, or correct coverage.

Our experience with system simulation examples indicates that coverage as low as 0.89 when the nominal level is 0.95 is rather pathological, provided the normal-theory assumptions are not significantly violated. And we know that typically $\Pr\{\mathscr{C}\} > \Pr\{\mathscr{E}\}$. Therefore, Figure 1 encourages us to believe that the procedure is robust enough to be used in practice. Since the performance of the procedure was not affected by the number of systems, r, we could inflate the nominal-coverage probability somewhat (say 0.97 when we

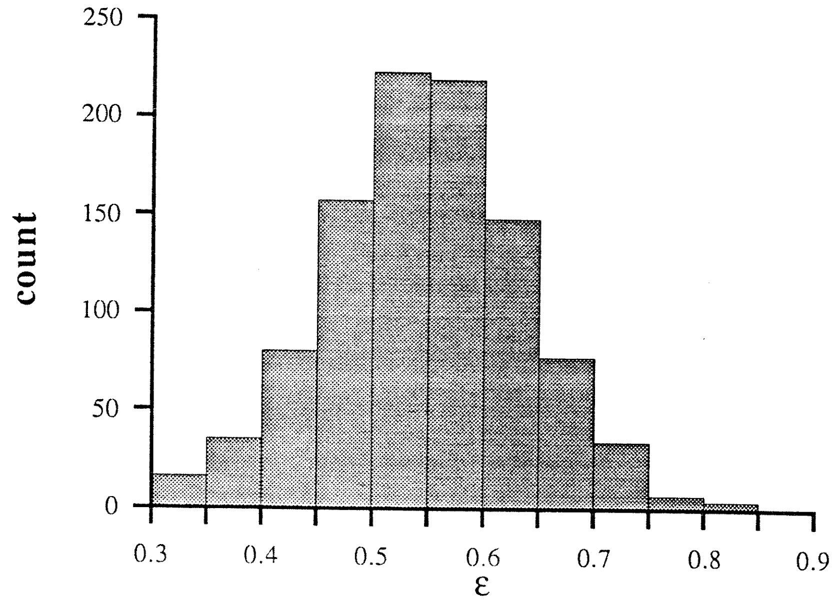

Fig. 2. Sphericity measure ε for the correlation matrices for $r = 5$ systems.

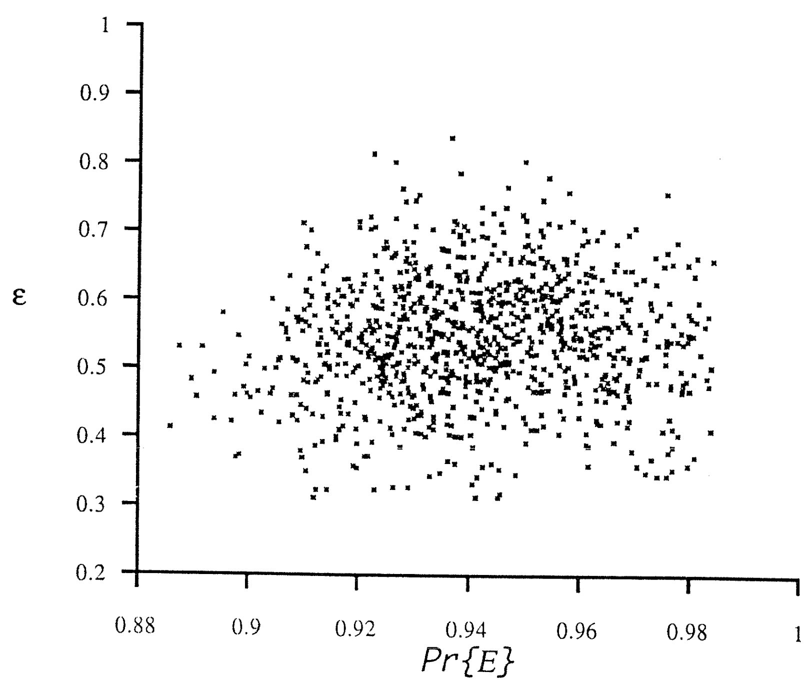

Fig. 3. Scatter plot of ε and $\widehat{\Pr}\{\mathscr{E}\}$ for $r = 5$ systems and $n = 30$ replications.

want 0.95) and still do better than procedures based on the Bonferroni inequality, where the necessary inflation is an increasing function of r. Although we had hoped that the sphericity measure ε would act as an indicator of pathological cases, this does not appear to be true. We continue to look for such an indicator.

APPENDIX A

This appendix defines notation and procedures. Let $\otimes$ denote the right Kronecker product (also called direct product) of two matrices, and let x^- denote $\min\{x, 0\}$ and x^+ denote $\max\{x, 0\}$.

MCB for the One-Way Model

If the simulation data satisfy the one-way model (6) with $\boldsymbol{\Sigma}_\eta = \tau^2 \mathbf{I}_r$, then a set of $(1-\alpha)100\%$ simultaneous confidence intervals for $\theta_i - \max_{\ell \neq i} \theta_\ell$ is

$$\left[\left(\bar{Y}_{i\cdot} - \max_{\ell \neq i} \bar{Y}_{\ell\cdot} - d^{1-\alpha}_{r-1, r(n-1)} \ddot{\tau} \sqrt{\frac{2}{n}} \right)^-, \right. \qquad \left. \left(\bar{Y}_{i\cdot} - \max_{\ell \neq i} \bar{Y}_{\ell\cdot} + d^{1-\alpha}_{r-1, r(n-1)} \ddot{\tau} \sqrt{\frac{2}{n}} \right)^+ \right] \tag{11}$$

for $i = 1, 2, \ldots, r$, where $d^{1-\alpha}_{r-1, r(n-1)}$ is the $1-\alpha$ quantile of an $(r-1)$-dimensional multivariate t random variable with $r(n-1)$ degrees of freedom and correlation matrix (7) (see, for instance, Hochberg and Tamhane [1987]), and $\ddot{\tau}$ is defined in Section 6.1.

When $\boldsymbol{\Sigma}_\eta$ satisfies sphericity, the only adjustment to (11) is to replace $\ddot{\tau}$ with $\tilde{\tau}$ and use a multivariate t quantile with $(r-1)(n-1)$ degrees of freedom.

Regression Formulations for the Control-Variate Models

Since the expected value of the control variate from system i, $\boldsymbol{\mu}_i$, is known, we can take $\boldsymbol{\mu}_i = \mathbf{0}$ without loss of generality.

Let

$$\mathbf{Y} \equiv \begin{bmatrix} \mathbf{Y}_{1\cdot} \\ \mathbf{Y}_{2\cdot} \\ \vdots \\ \mathbf{Y}_{r\cdot} \end{bmatrix}$$

where $\mathbf{Y}_{i\cdot} \equiv (Y_{i1}, Y_{i2}, \ldots, Y_{in})'$ denotes the responses across all n replications from system i.

Similarly, let

$$\mathbf{C} \equiv \begin{bmatrix} \mathbf{C}_{1\cdot} \\ \mathbf{C}_{2\cdot} \\ \vdots \\ \mathbf{C}_{r\cdot} \end{bmatrix}$$

where

$$\mathbf{C}_{i\cdot} \equiv \begin{bmatrix} \mathbf{C}'_{i1} \\ \mathbf{C}'_{i2} \\ \vdots \\ \mathbf{C}'_{in} \end{bmatrix}.$$

Organize the unknown parameters as

$$\gamma \equiv \begin{bmatrix} \theta_1 \\ \boldsymbol{\beta}_1 \\ \vdots \\ \theta_r \\ \boldsymbol{\beta}_r \end{bmatrix}.$$

Then the least-squares estimator of γ is $\hat{\gamma} = (\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}'\mathbf{Y}$, where

$$\mathbf{G} \equiv \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 & \cdots & \mathbf{0} \\ & & \vdots & \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X}_r \end{bmatrix}$$

and $\mathbf{X}_i \equiv [\mathbf{1}_n, \mathbf{C}_{i\cdot}]$. This representation shows that the elements of $\hat{\gamma}$ are obtained from r individual regressions, e.g., $\hat{\theta}_i$ is the first element of $(\mathbf{X}'_i\mathbf{X}_i)^{-1}\mathbf{X}'_i\mathbf{Y}_{i\cdot}$.

For the case when $\boldsymbol{\beta}_i = \boldsymbol{\beta}$ for all i, let $\mathbf{G} = [\mathbf{X},\mathbf{C}]$, where $\mathbf{X} = \mathbf{1}_n \otimes \mathbf{I}_r$, and $\mathbf{C}$ is as defined above. Then

$$\gamma \equiv \begin{bmatrix} \boldsymbol{\theta} \\ \boldsymbol{\beta} \end{bmatrix}$$

and $\hat{\gamma} = (\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}'\mathbf{Y}$.

APPENDIX B

PROOF OF LEMMA 5.1. Notice that

$$\begin{aligned} \mathrm{Cov}\left[\hat{\theta}_i, \hat{\theta}_\ell | \mathbf{C}\right] &= \mathrm{Cov}[\mathbf{M}'_i\mathbf{Y}_{i\cdot}, \mathbf{M}'_\ell\mathbf{Y}_{\ell\cdot} | \mathbf{C}] \\ &= \mathbf{M}'_i \mathrm{Cov}[\mathbf{Y}_{i\cdot}, \mathbf{Y}_{\ell\cdot} | \mathbf{C}] \mathbf{M}_\ell. \end{aligned}$$

But

$$\begin{aligned}
\operatorname{Cov}\left[Y_{ij}, Y_{\ell k} | \mathbf{C}\right] &= \operatorname{Cov}\left[\theta_i + (\mathbf{C}_{ij} - \boldsymbol{\mu}_i)'\boldsymbol{\beta}_i + \eta_{ij},\right. \\
&\qquad \left.\theta_\ell + (\mathbf{C}_{\ell k} - \boldsymbol{\mu}_\ell)'\boldsymbol{\beta}_\ell + \eta_{\ell k} | \mathbf{C}\right] \\
&= \operatorname{Cov}[\eta_{ij}, \eta_{\ell k}] \\
&= \begin{cases} 0, & j \neq k \\ \sigma_{i\ell}, & j = k. \end{cases}
\end{aligned}$$

Thus, $\operatorname{Cov}[\mathbf{Y}_{i\cdot}, \mathbf{Y}_{\ell\cdot} | \mathbf{C}] = \sigma_{i\ell}\mathbf{I}_n$, and the result follows by substitution. □

Let $\mathbf{Z}$ be distributed as a $t \times 1$-variate normal random variable with mean vector $\boldsymbol{\nu}$ and variance-covariance matrix $\boldsymbol{\Sigma}$. Let $\mathbf{Q}$ be a real (constant) $t \times t$ matrix of rank $u \leq t$, and let $\mathbf{P}$ be a real (constant) $t \times 1$ vector. The proof of Theorem 5.1 is based on the following results:

(1) Under model (1), the joint distribution of $\mathbf{Y}$ given $\mathbf{C}$ is normal with mean vector $\mathbf{G}\boldsymbol{\gamma}$ and variance-covariance matrix $\boldsymbol{\Sigma} = \mathbf{I}_n \otimes \boldsymbol{\Sigma}_\eta$. (This is easily verified using standard results.)

(2) (Box [1954], Theorem 2.1): The random variable $(\mathbf{Z} - \boldsymbol{\nu})'\mathbf{Q}(\mathbf{Z} - \boldsymbol{\nu})$ is distributed as $\sum_{j=1}^{u} \lambda_j \chi_1^2(j)$ where $\chi_1^2(1), \chi_1^2(2), \ldots, \chi_1^2(u)$ are i.i.d. chi-squared random variables with 1 degree of freedom, and $\lambda_1, \lambda_2, \ldots, \lambda_u$ are the u nonzero eigenvalues of $\boldsymbol{\Sigma}\mathbf{Q}$.

(3) (Rao [1973], Theorem 3b.4(viii)): A necessary and sufficient condition for $\mathbf{P}'\mathbf{Z}$ to be independent of $(\mathbf{Z} - \boldsymbol{\nu})'\mathbf{Q}(\mathbf{Z} - \boldsymbol{\nu})$ is that $\boldsymbol{\Sigma}\mathbf{Q}\boldsymbol{\Sigma}\mathbf{P} = \mathbf{0}$.

Additionally, Theorem 8.8.6 of Graybill [1969] will be used repeatedly: If $\mathbf{R}$ is an $m_1 \times m_2$ matrix, $\mathbf{S}$ an $m_3 \times m_4$ matrix, $\mathbf{T}$ an $m_2 \times m_5$ matrix, and $\mathbf{U}$ and $m_4 \times m_5$ matrix, then $(\mathbf{R} \otimes \mathbf{S})(\mathbf{T} \otimes \mathbf{U}) = (\mathbf{RT}) \otimes (\mathbf{SU})$.

PROOF OF THEOREM 5.1. From standard least-squares results

$$\hat{\boldsymbol{\eta}} = \begin{pmatrix} \hat{\eta}_{11} \\ \vdots \\ \hat{\eta}_{1n} \\ \vdots \\ \hat{\eta}_{r1} \\ \vdots \\ \hat{\eta}_{rn} \end{pmatrix} = \mathbf{HY} \tag{12}$$

where $\mathbf{H} \equiv \mathbf{I}_{rn} - \mathbf{G}(\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}'$ is symmetric and idempotent. But since the regression decomposes into r individual regressions (Appendix A) and since all the control variates are common,

$$\mathbf{H} = \mathbf{H}_1 \otimes \mathbf{I}_r \tag{13}$$

where $\mathbf{H}_1 = \mathbf{I}_n - \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'$, which is also symmetric and idempotent.

Notice that

$$\begin{aligned}(r-1)&(n-1-1)\bar{\tau}^2\\ &= \sum_{i=1}^{r}\sum_{j=1}^{n}\left(\hat{\eta}_{ij} - \bar{\hat{\eta}}_{i\cdot} + \bar{\hat{\eta}}_{\cdot j} + \bar{\hat{\eta}}_{\cdot\cdot}\right)^2\\ &= \sum_{i=1}^{r}\sum_{j=1}^{n}\left(\hat{\eta}_{ij} - \bar{\hat{\eta}}_{\cdot j}\right)^2\end{aligned}$$

since $\sum_{j=1}^{n}\eta_{ij} = 0$ by the properties of least square. Let $\mathbf{F} = \mathbf{I}_n \otimes 1/r\mathbf{1}_r\mathbf{1}_r'$, a symmetric, idempotent matrix. Then

$$\sum_{i=1}^{r}\sum_{j=1}^{n}\left(\hat{\eta}_{ij} - \bar{\hat{\eta}}_{\cdot j}\right)^2 = \hat{\boldsymbol{\eta}}'(\mathbf{I}_{rn} - \mathbf{F})(\mathbf{I}_{rn} - \mathbf{F})\hat{\boldsymbol{\eta}} = \hat{\boldsymbol{\eta}}'(\mathbf{I}_{rn} - \mathbf{F})\hat{\boldsymbol{\eta}}. \tag{14}$$

Combining (12)–(14)

$$\begin{aligned}(r-1)(n-q-1)\bar{\tau}^2 &= \mathbf{Y}'\mathbf{H}(\mathbf{I}_{rn} - \mathbf{F})\mathbf{H}\mathbf{Y}\\ &= \mathbf{Y}'\left(\mathbf{H}_1 \otimes \mathbf{I}_r - (\mathbf{H}_1 \otimes \mathbf{I}_r)\left(\mathbf{I}_n \otimes \frac{1}{r}\mathbf{1}_r\mathbf{1}_r'\right)(\mathbf{H}_1 \otimes \mathbf{I}_r)\right)\mathbf{Y}\\ &= \mathbf{Y}'\left(\mathbf{H}_1 \otimes \mathbf{I}_r - \mathbf{H}_1 \otimes \frac{1}{r}\mathbf{1}_r\mathbf{1}_r'\right)\mathbf{Y}\\ &= \mathbf{Y}'\left(\mathbf{H}_1 \otimes \left(\mathbf{I}_r - \frac{1}{r}\mathbf{1}_r\mathbf{1}_r'\right)\right)\mathbf{Y} \equiv \mathbf{Y}'\mathbf{Q}\mathbf{Y}.\end{aligned}$$

To find the distribution of $\mathbf{Y}'\mathbf{Q}\mathbf{Y}$, we first notice that $\mathbf{Y}'\mathbf{Q}\mathbf{Y} = (\mathbf{Y} - \mathbf{G}\boldsymbol{\gamma})'\mathbf{Q}(\mathbf{Y} - \mathbf{G}\boldsymbol{\gamma})$, so that Theorem 2.1 of Box [1954] applies. Therefore we need the eigenvalues of

$$\boldsymbol{\Sigma}\mathbf{Q} = (\mathbf{I}_n \otimes \boldsymbol{\Sigma}_\eta)\left(\mathbf{H}_1 \otimes \left(\mathbf{I}_r - \frac{1}{r}\mathbf{1}_r\mathbf{1}_r'\right)\right) = \mathbf{H}_1 \otimes \left(\boldsymbol{\Sigma}_\eta\left(\mathbf{I}_r - \frac{1}{r}\mathbf{1}_r\mathbf{1}_r'\right)\right). \tag{15}$$

From Nelson [1990], $\mathbf{H}_1$ has rank $n - q - 1$; therefore, since $\mathbf{H}_1$ is idempotent it has $n - q - 1$ nonzero eigenvalues all equal to 1 (Theorem 12.3.2, Graybill [1969]). Tedious algebra shows that $1/\tau^2\boldsymbol{\Sigma}_\eta(\mathbf{I}_r - 1/r\mathbf{1}_r\mathbf{1}_r')$ is idempotent and has trace $r - 1$; therefore $1/\tau^2\boldsymbol{\Sigma}_\eta(\mathbf{I}_r - 1/r\mathbf{1}_r\mathbf{1}_r')$ has $r - 1$ nonzero eigenvalues all equal to 1 (Theorem 12.6.12, Graybill [1969]), and $\boldsymbol{\Sigma}_\eta(\mathbf{I}_r - 1/r\mathbf{1}_r\mathbf{1}_r')$ has $r - 1$ nonzero eigenvalues all equal to τ^2.

Applying Theorem 8.8.13 of Graybill [1969], the eigenvalues of $\mathbf{H}_1 \otimes (\boldsymbol{\Sigma}_\eta(\mathbf{I}_r - 1/r\mathbf{1}_r\mathbf{1}_r'))$ are the product of all pairs of eigenvalues of $\mathbf{H}_1$ and $(\boldsymbol{\Sigma}_\eta(\mathbf{I}_r - 1/r\mathbf{1}_r\mathbf{1}_r'))$; that is, τ^2 with multiplicity $(r-1)(n-q-1)$ and 0 with multiplicity $nr - (r-1)(n-q-1)$. Thus,

$$(r-1)(n-q-1)\bar{\tau}^2 \sim \tau^2\chi^2_{(r-1)(n-q-1)}$$

from Box [1954, Theorem 2.1].

To show that $\hat{\boldsymbol{\theta}}$ and $\bar{\tau}^2$ are independent, we establish a stronger result, namely, that $\hat{\boldsymbol{\gamma}} = (\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}'\mathbf{Y} \equiv \mathbf{P}'\mathbf{Y}$ is independent of $\mathbf{Y}'\mathbf{Q}\mathbf{Y}$.

From Rao ([1973], 3b.4(viii)), they are independent if $(\mathbf{I}_n \otimes \boldsymbol{\Sigma}_\eta)\mathbf{Q}(\mathbf{I}_n \otimes \boldsymbol{\Sigma}_\eta)\mathbf{P} = \mathbf{0}$. Notice that $\mathbf{P}' = (\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}' = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1' \otimes \mathbf{I}_r \equiv \mathbf{T}_1 \otimes \mathbf{I}_r$ since the control variates are common across systems. Repeated application of Graybill's Theorem 8.8.6 [1969] gives

$$
\begin{aligned}
(\mathbf{I}_n \otimes \boldsymbol{\Sigma}_\eta)\left(\mathbf{H}_1 \otimes \left(\mathbf{I}_r - \frac{1}{r}\mathbf{1}_r\mathbf{1}_r'\right)\right)&(\mathbf{I}_n \otimes \boldsymbol{\Sigma}_\eta)(\mathbf{T}_1' \otimes \mathbf{I}_r) \\
&= (\mathbf{H}_1\mathbf{T}_1') \otimes \left(\boldsymbol{\Sigma}_\eta\left(\mathbf{I}_r - \frac{1}{r}\mathbf{1}_r\mathbf{1}_r'\right)\boldsymbol{\Sigma}_\eta\right).
\end{aligned}
$$

But

$$
\mathbf{H}_1\mathbf{T}_1' = \left(\mathbf{I}_n - \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\right)\left(\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\right) = \mathbf{0}
$$

which completes the proof. □

Proof of Theorem 5.2. Nelson ([1990, Appendix B]) showed that

$$
\mathbf{M}_i' = \frac{1}{n}\mathbf{1}_n' - \overline{\mathbf{C}}_{i\cdot}'(\mathbf{L}_i'\mathbf{L}_i)^{-1}\mathbf{L}_i'
$$

where $\overline{\mathbf{C}}_{i\cdot} = \mathbf{C}_i'\mathbf{1}_n/n$ and

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

Direct calculation gives

$$
n\mathbf{M}_i'\mathbf{M}_\ell = 1 + n\overline{\mathbf{C}}_{i\cdot}'(\mathbf{L}_i'\mathbf{L}_i)^{-1}\mathbf{L}_i'\mathbf{L}_\ell(\mathbf{L}_\ell'\mathbf{L}_\ell)^{-1}\overline{\mathbf{C}}_{\ell\cdot}.
$$

The proof follows by showing that the second term on the right-hand side converges in probability to 0 by repeated applications of Slutsky's theorem. □

Proof of Theorem 61. The $\mathrm{E}[\ddot{\tau}^2]$ is obtained by standard arguments. To compute $\mathrm{E}[\bar{\tau}^2]$, reorganize the η_{ij} by replication into $\boldsymbol{\eta} \equiv (\eta_{11}, \eta_{21}, \ldots, \eta_{r1}, \ldots, \eta_{1n}, \eta_{2n}, \ldots, \eta_{rn})'$. Then $\mathrm{Var}[\boldsymbol{\eta}] = \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_\eta \otimes \mathbf{I}_n$, and

$$
(r-1)(n-1)\bar{\tau}^2 = \boldsymbol{\eta}'(\mathbf{I}_{rn} - \mathbf{A})(\mathbf{I}_{rn} - \mathbf{B})\boldsymbol{\eta}
$$

where $\mathbf{A} \equiv 1/n\mathbf{I}_r \otimes \mathbf{1}_n\mathbf{1}_n'$ and $\mathbf{B} \equiv 1/r\mathbf{1}_r\mathbf{1}_r' \otimes \mathbf{I}_n$ (Wang [1992]). Thus,

$$
\begin{aligned}
\mathrm{E}[(r-1)(n-1)\bar{\tau}^2] &= \mathrm{tr}[(\mathbf{I}_{rn} - \mathbf{A})(\mathbf{I}_{rn} - \mathbf{B})\boldsymbol{\Sigma}] \\
&\quad + \mathrm{E}[\boldsymbol{\eta}'](\mathbf{I}_{rn} - \mathbf{A})(\mathbf{I}_{rn} - \mathbf{B})\mathrm{E}[\boldsymbol{\eta}] \\
&= \mathrm{tr}[(\mathbf{I}_{rn} - \mathbf{A} - \mathbf{B} + \mathbf{AB})\boldsymbol{\Sigma}]
\end{aligned}
$$

since $E[\boldsymbol{\eta}] = \mathbf{0}$. Noting that $\mathbf{AB} = 1/nr \mathbf{1}_r \mathbf{1}'_r \otimes \mathbf{1}_n \mathbf{1}'_n$, and working term by term

$$\begin{aligned}
\mathrm{tr}[\boldsymbol{\Sigma}] &= n\,\mathrm{tr}[\boldsymbol{\Sigma}_\eta] = nr\varsigma \\
\mathrm{tr}[\mathbf{A}\boldsymbol{\Sigma}] &= \mathrm{tr}\left[\left(\frac{1}{n}\mathbf{I}_r \otimes \mathbf{1}_n \mathbf{1}'_n\right)(\boldsymbol{\Sigma}_\eta \otimes \mathbf{I}_n)\right] = \frac{1}{n}\mathrm{tr}[\boldsymbol{\Sigma}_\eta \otimes \mathbf{1}_n \mathbf{1}'_n] = \frac{n}{n}\mathrm{tr}[\boldsymbol{\Sigma}_\eta] = r\varsigma \\
\mathrm{tr}[\mathbf{B}\boldsymbol{\Sigma}] &= \mathrm{tr}\left[\left(\frac{1}{r}\mathbf{1}_r \mathbf{1}'_r \otimes \mathbf{I}_n\right)(\boldsymbol{\Sigma}_\eta \otimes \mathbf{I}_n)\right] = \frac{1}{r}\mathrm{tr}[\mathbf{1}_r \mathbf{1}'_r \boldsymbol{\Sigma}_\eta \otimes \mathbf{I}_n] = \frac{n}{r}\mathrm{tr}[\mathbf{1}_r \mathbf{1}'_r \boldsymbol{\Sigma}_\eta] \\
\mathrm{tr}[\mathbf{AB}\boldsymbol{\Sigma}] &= \mathrm{tr}\left[\left(\frac{1}{nr}\mathbf{1}_r \mathbf{1}'_r \otimes \mathbf{1}_n \mathbf{1}'_n\right)(\boldsymbol{\Sigma}_\eta \otimes \mathbf{I}_n)\right] = \frac{1}{nr}\mathrm{tr}[\mathbf{1}_r \mathbf{1}'_r \boldsymbol{\Sigma}_\eta \otimes \mathbf{1}_n \mathbf{1}'_n] \\
&= \frac{n}{nr}\mathrm{tr}[\mathbf{1}_r \mathbf{1}'_r \boldsymbol{\Sigma}_\eta].
\end{aligned}$$

Putting these terms together, and noticing that $\mathrm{tr}[\mathbf{1}_r \mathbf{1}'_r \boldsymbol{\Sigma}_\eta] = r\varsigma + r(r-1)\varrho$, gives

$$\begin{aligned}
E[(r-1)(n-1)\tilde{\tau}^2] &= (n-1)r\varsigma - \frac{n-1}{r}(r\varsigma + r(r-1)\varrho) \\
&= (n-1)(r-1)(\varsigma - \varrho).
\end{aligned}$$

REFERENCES

BOX, G. E. P. 1954. Some theorems on quadratic forms applied in the study of analysis of variance problems, I. Effect of inequality of variance in the one-way classification. *An. Math. Stat. 25*, 290–302.

GRAYBILL, F. A. 1969. *Matrices with Applications in Statistics*. Wadsworth, Belmont, Calif.

GRIEVE, A. P., AND AG, C.-G. 1984. Tests of sphericity of normal distributions and the analysis of repeated measures designs. *Psychometrika 49*, 2, 257–267.

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

HSU, J. C., AND NELSON, B. L. 1988. Optimization over a finite number of system designs with one-stage sampling and multiple comparisons with the best. In *Proceedings of the 1988 Winter Simulation Conference*, M. A. Abrams, P. L. Haigh, J. C. Comfort, Eds. ACM/IEEE, New York, 451–457.

MARSAGLIA, G., AND OLKIN, I. 1984. Generating correlation matrices. *SIAM. Sci. Stat. Comput. 5*, 470–475.

NELSON, B. L. 1992. Common random numbers and multiple comparisons in simulation analysis. In *Proceedings of the 1st Industrial Engineering Research Conference*, G. Klutke, D. A. Mitta, B. O. Nnaji, L. M. Seiford, Eds. IIE, 463–466.

NELSON, B. L. 1990. Control variate remedies. *Oper. Res. 38*, 6, 974–992.

NELSON, B. L., AND HSU, J. C. 1993. Control-variate models of common random numbers for multiple comparisons with the best. *Manage. Sci*. To be published.

NOZARI, A., ARNOLD, S., AND PEGDEN, C. 1987. Statistical analysis for use with the Schruben and Margolin correlation induction strategy. *Oper. Res. 35*, 1, 127–139.

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

SCHRUBEN, L., AND MARGOLIN, B. 1978. Pseudorandom number assignment in statistically designed simulation and distribution sampling experiments. *J. Am. Stat. Assoc. 73*, 504–525.

TEW, J. D., AND WILSON, J. R. 1993. Estimating simulation metamodels using combined correlation-based variance reduction techniques. *IIE Trans.* To be published.

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

WANG, H.-L. 1992. Multiple comparisons procedures in repeated measures designs. Ph.D. dissertation, Dept. of Statistics, The Ohio State Univ., Columbus, Ohio.

YANG, W., AND NELSON, B. L. 1991. Using common random numbers and control variates in multiple-comparison procedures. *Oper. Res. 39*, 4, 583–591.

Received August 1992; revised March 1993; accepted June 1993