Proceedings of the 1987 Winter Simulation Conference
A. Thesen, H. Grant, W. David Kelton (eds.)

METAMODEL ESTIMATION USING INTEGRATED CORRELATION METHODS

Jeffrey D. Tew
Department of IEOR
Virginia Polytechnic Institute and State University
Blacksburg, Virginia 24061

James R. Wilson
School of Industrial Engineering
Purdue University
West Lafayette, Indiana 47907

ABSTRACT

This paper develops a generalized approach for combining the use of the Schruben-Margolin correlation induction strategy and control variates in a simulation experiment designed to estimate a metamodel that is linear in the unknown parameters relating the response variable of interest to selected exogenous decision variables. This generalized approach is based on standard techniques of regression analysis. Under certain broad assumptions, the combined use of the Schruben-Margolin correlation induction strategy and control variates is shown to give a more efficient estimator of the metamodel coefficients than each of the following conventional correlation-based variance reduction techniques: independent streams, common random numbers, control variates, and the Schruben-Margolin strategy.

1. INTRODUCTION

In this section we present the notation used in this paper for describing simulation experiments, and we briefly review the Schruben-Margolin correlation induction strategy as well as the method of control variates.

1.1 Setup for Simulation Experiments

Consider a simulation experiment consisting of m design points, where each design point is identified by the settings of d factors or decision variables, denoted by ϕ, that are used as inputs to the simulation model. Let the response from the i'th design point be denoted by y_i and let the vector of responses from all m design points be denoted by $y = (y_1, y_2, \ldots, y_m)'$. Also, let ϕ_i be the setting of the d factors for the i'th design point and let $\{x_k: k = 1, 2, \ldots, p-1\}$ represent known functions of the factor settings. Then, assuming that the relationship between the response and the given functions of the factor settings is linear in the unknown parameters, we can write

$$y_i = \beta_0 + \sum_{k=1}^{p-1} \beta_k x_k(\phi_i) + \varepsilon_i$$

$$\text{for } i = 1, 2, \ldots, m, \qquad (1.1)$$

where $\{\beta_k: k = 0, 1, \ldots, p-1\}$ are the unknown model parameters and ε_i represents the inability of $\beta_0 + \sum_{k=1}^{p-1} \beta_k x_k(\phi_i)$ to determine y_i. Define $\mathbf{X}$ to be the $(m \times p)$ matrix whose first column is all ones and whose $(i, k+1)$ element is $x_k(\phi_i)$ (for $i = 1, 2, \ldots, m$ and $k = 1, 2, \ldots, p-1$). Thus, the relationship between the response and the functions of factor settings across all m design points can be written compactly as the following general linear model:

$$\mathbf{y} = \mathbf{X}\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}. \qquad (1.2)$$

On occasion later in this paper, we assume that the $\{x_k\}$ are chosen such that $\mathbf{X}$ is orthogonal, that is:

$$\mathbf{X}'\mathbf{X} = mI_p, \qquad (1.3)$$

and can be achieved by a simple reparameterization, or coding, of the functions of the factor levels.

We also assume that

$$\{\mathbf{y}_h: h = 1, 2, \ldots, r\} \text{ IID} \sim N_m(\mathbf{X}\underset{\sim}{\beta}, \underset{\sim}{\Sigma}), \qquad (1.4)$$

where y_{ih} denotes the response at the i'th design point on the h'th independent replication of the basic m-point experiment and $\underset{\sim}{\Sigma}$ is the $(m \times m)$ covariance matrix.

A simulation model is usually driven by randomly chosen <u>streams</u> of pseudorandom numbers. The streams are sequences of real numbers scaled to the interval [0,1] and constructed to appear random. For a single replication of the basic m-point experimental design, we represent the set of g pseudorandom number streams in the following way: (a) the (infinite) sequence of pseudorandom numbers available from the j'th stream at the i'th design point is

$$r_{ij} = (r_{ij1}, r_{ij2}, \ldots)$$

for $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, g$; (b) the set of streams for the i'th design point is

$$R_i = (r_{i1}, r_{i2}, \ldots, r_{ig}) \text{ for } i = 1, 2, \ldots, m; \quad (1.5)$$

and (c) the aggregate pseudorandom input for the basic m-point experimental design is

$$R = (R_1, R_2, \ldots, R_m)'.$$

Now at the i'th design point, R_i completely determines the events of the simulation so that we can write

$$y_i(R_i) = \beta_0 + \sum_{k=1}^{p-1} \beta_k x_k(\phi_i) + \varepsilon_i(R_i). \quad (1.6)$$

In conducting a simulation experiment, the simulation analyst must assign a set of random number streams to each experimental point. Three common methods of assigning the random number streams for the simulation experiment are: independent streams, common random numbers, and antithetic variates. Chapters 1 and 5 of Tew (1986) give a description of each of these methods in the context of metamodel estimation as well as references for further reading.

1.2 The Schruben-Margolin Correlation Induction Strategy

To facilitate the design of efficient simulation experiments, Schruben and Margolin (1978) devised a correlation induction strategy that utilizes the variance reduction techniques of common random numbers and antithetic variates in a scheme based on the concept of blocking. In addition to the assumptions (1.1) to (1.6), they assumed that the design matrix $X = (\underset{\sim}{1}_m T)$ is orthogonally blockable. A design matrix X that satisfies the properties of (1.3) is orthogonally blockable into two blocks if there exists an $(m \times 2)$ matrix W of zeros and ones such that $T'W = 0$ and $\underset{\sim}{1}_m'W = [m_1, m_2]$, where m_1 and m_2 are the respective block sizes. If we let $1 - r_{ij} \equiv (1 - r_{ij1}, 1 - r_{ij2}, \ldots)$ denote the complement of the random number stream r_{ij}, then the assignment rule of Schruben and Margolin can be expressed as follows:

Assignment Rule; If the m-point experimental design admits orthogonal blocking into two blocks of sizes m_1 and m_2, preferably chosen to be as nearly equal in size as possible, then for all m_1 design points in the first block, use a common set of pseudorandom numbers so that $R_i = R_1 = (r_{11}, r_{12}, \ldots, r_{1g})$, $i = 1, 2, \ldots, m$; and for all m_2 design points in the second block, use the antithetic (complementary) set of pseudorandom numbers so that $\bar{R}_i = \bar{R}_1 = (1-r_{11}, 1-r_{12}, \ldots, 1-r_{1g})$, $i = m_1+1, \ldots, m_1+m_2$.

Schruben and Margolin decomposed the error term ε_i into a <u>random block effect</u> b_i and a residual ε_i^o, both of which are functions of R. Thus the model in (1.1) can be written:

$$y_i(R) = \beta_0 + \sum_{k=1}^{p-1} \beta_k x_k(\phi_i) + b_i(R) + \varepsilon_i^o(R)$$

$$\text{for } i = 1, 2, \ldots, m. \quad (1.7)$$

In order to analyze the properties of this assignment rule, Schruben and Margolin made the following assumptions (for $1 \le i, j \le m$):

$$\left.\begin{array}{l}
\varepsilon_i(R) = b_i(R) + \varepsilon_i^o(R); \\
E[b_i(R)] = E[\varepsilon_i^o(R)] = 0; \\
\sigma_i^2 \equiv \mathrm{Var}(y_i) = \mathrm{Var}[y_i(R)] = \sigma^2; \\
\mathrm{Cov}[b_i(R), b_j(R)] = \rho_1\sigma^2, \text{ where } 0 < \rho_1 < 1; \\
\mathrm{Var}[\varepsilon_i^o(R)] = \sigma^2(1-\rho_1); \\
\mathrm{Cov}[b_i(R), b_j(\bar{R})] = \rho_2\sigma^2 \text{ where } -1 < \rho_2 < 0; \\
\mathrm{Cov}[b_i(R), \varepsilon_j^o(R)] = \mathrm{Cov}[b_i(R), \varepsilon_j^o(\bar{R})] = 0; \\
\mathrm{Cov}[\varepsilon_i^o(R), \varepsilon_j^o(R)] = \mathrm{Cov}[\varepsilon_i^o(R), \varepsilon_j^o(\bar{R})] = 0 \\
\qquad\qquad \text{for } i \ne j; \\
\underset{\sim}{\Sigma} = \|\mathrm{Cov}(y_i, y_j)\| \text{ is positive definite.}
\end{array}\right\} \quad (1.8)$$

These assumptions imply the following three properties:

1. The response variance is constant across all points in the design.
2. If y_i and y_j (for $i \ne j$) are realized from the same random number stream, then

$$\mathrm{Corr}[y_i(R), y_j(R)] = \rho_1, \; 0 < \rho_1 < 1. \quad (1.9)$$

3. If y_i and y_j (for $i \ne j$) are realized from antithetic (complementary) random number streams, R and $\bar{R}$ respectively, then

$$R, \bar{R} \text{ antithetic} \rightarrow \mathrm{Corr}[y_i(R), y_j(\bar{R})] = \rho_2,$$
$$-1 < \rho_2 < 0. \quad (1.10)$$

Under the Schruben-Margolin strategy with equal block sizes, the metamodel of (1.6) takes the following form:

$$y(R) = X\underset{\sim}{\beta} + WB(R) + \underset{\sim}{\varepsilon}^o(R), \quad (1.11)$$

where: $q \equiv \frac{m}{2} = m_1 = m_2$ is the common block size; $B' = [b(R_1), b(\bar{R}_1)]'$ is the (2×1) vector of random block effects; W is the $(m \times 2)$ <u>block incidence</u> matrix defined by Schruben and Margolin (1978) and Schruben (1979) and $\underset{\sim}{\varepsilon}^{o}$ is the (m × 1) vector of residual errors. Note that within each block, Schruben and Margolin assume a common block effect that does not depend on the design point. Let X_i (i = 1, 2) represent the design matrix for the i'th block. If the experimental points are so arranged that $\mathbf{X} = [\mathbf{X}_1' \ \mathbf{X}_2']'$, then we get

$$W = \left[\begin{array}{c|c} \underset{\sim}{1}_q & \underset{\sim}{0}_q \\ \hline \underset{\sim}{0}_q & \underset{\sim}{1}_q \end{array}\right], \qquad (1.12)$$

where each column of $\mathbf{W}$ contains q = m/2 ones. With the assumptions of (1.8), we get

$$\mathrm{Cov}[B] = \sigma^2 \begin{bmatrix} \rho_1 & \rho_2 \\ \rho_2 & \rho_1 \end{bmatrix}. \qquad (1.13)$$

Expressions (1.12) and (1.13), together with the assumptions of (1.8), result in the covariance structure of y given by the following:

$$\underset{\sim}{\Sigma}_0 = \sigma^2 \left[\begin{array}{ccc|ccc} 1 & \cdots & \rho_1 & \rho_2 & \cdots & \rho_2 \\ \cdot & \cdots & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot & \cdot & \cdots & \cdot \\ \rho_1 & \cdots & 1 & \rho_2 & \cdots & \rho_2 \\ \hline \rho_2 & \cdots & \rho_2 & 1 & \cdots & \rho_1 \\ \cdot & \cdots & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot & \cdot & \cdots & \cdot \\ \rho_2 & \cdots & \rho_2 & \rho_1 & \cdots & 1 \end{array}\right] = \sigma^2 \left[\begin{array}{c|c} \underset{\sim}{\Sigma}_{11} & \underset{\sim}{\Sigma}_{12} \\ \hline \underset{\sim}{\Sigma}_{21} & \underset{\sim}{\Sigma}_{22} \end{array}\right], \qquad (1.14)$$

where $\underset{\sim}{\Sigma}_{11}$ is $(m_1 \times m_1)$, $\underset{\sim}{\Sigma}_{12}$ is $(m_1 \times m_2)$, $\underset{\sim}{\Sigma}_{21}$ is $(m_2 \times m_1)$, and $\underset{\sim}{\Sigma}_{22}$ is $(m_2 \times m_2)$.

Based on experimental designs that admit orthogonal blocking, Schruben and Margolin proved the following theorem.

Theorem 1: If an experimental design admits orthogonal blocking, and if the assumptions of (1.13) hold, then under the assignment rule the ordinary least squares estimator of $\underset{\sim}{\beta}$ has a smaller generalized variance than it has under the following strategies: (a) the assignment of one common set of random numbers to all design points, or (b) the assignment of a different set of random numbers to each design point, provided

$$[1 + (m-1)\rho_1 - (2/m)(m_1)(m_2)(\rho_1-\rho_2)](1-\rho_1)^p < 1 \qquad (1.15)$$

in the latter case.

Corollary 1: Under the assumptions of Theorem 1, the assignment rule is superior to the use of common random numbers in estimating β_0 ; the two are equivalent in terms of dispersion for estimating $(\beta_1, \beta_2, \ldots, \beta_{p-1})'$. When compared to the use of a different random number stream at each point, both the assignment rule and common random numbers are superior in terms of dispersion for estimating $(\beta_1, \beta_2, \ldots, \beta_{p-1})'$.

Thus, Schruben and Margolin showed that their strategy is a successful means of combining the two correlation methods of common random numbers and antithetic variates for a large class of experimental designs.

1.3. The Method of Control Variates

The method of control variates involves identifying a vector of concomitant output variables, $\mathbf{c} = (c_1, c_2, \ldots, c_s)'$, having both a known mean $\underset{\sim}{\mu}_c$ and a strong linear relationship with the response of interest y. The basic idea is to predict and counteract the unknown deviation $y - \mu_y$ by subtracting from y an appropriate linear transformation of the known deviation $\mathbf{c} - \underset{\sim}{\mu}_c$. In the context of a simulation experiment as defined by equations (1.1)-(1.11), suppose that, along with the response from the i'th experimental point y_i (i = 1, 2, ..., m), we also observe an s-dimensional column vector of control variates $\mathbf{c}_i$. In this situation we may assume without loss of generality, that $E(\mathbf{c}_i) = \underset{\sim}{0}$ (i = 1, 2, ..., m). Moreover, we assume that at each experimental point, the response and the control variates are jointly normal.

If we let $\mathbf{c} = (\mathbf{c}_1', \mathbf{c}_2', \ldots, \mathbf{c}_m')'$, then we have

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{c} \end{bmatrix} \sim N_{m(s+1)}\left[\begin{bmatrix} \mathbf{X} \\ \underset{\sim}{0} \end{bmatrix}\underset{\sim}{\beta}, \ \underset{\sim}{\Sigma}^{cv}\right], \qquad (1.16)$$

where

$$\underset{\sim}{\Sigma}^{cv} = \begin{bmatrix} \sigma^2 \mathbf{I}_m & \mathbf{I}_m \otimes \underset{\sim}{\Sigma}_{yc} \\ \mathbf{I}_m \otimes \Sigma_{cy} & \mathbf{I}_m \otimes \underset{\sim}{\Sigma}_c \end{bmatrix}, \qquad (1.17)$$

$\underset{\sim}{\Sigma}_c$ is the (s × s) unknown covariance matrix of $\mathbf{c}_i$, and $\underset{\sim}{\Sigma}_{yc}$ is a (s × 1) unknown covariance matrix whose elements are the covariances between y_i and c_i. (Note that (1.16) indicates that Σ_c and Σ_{yc} do not depend on i.)

In the following development it will frequently be convenient to express results in terms of the right direct product (or Kronecker product) of two matrices. If G is a (t × u) matrix and H is a (v × z) matrix, then the right direct product of G and H is the (tv × uz) matrix

$$\mathbf{G} \otimes \mathbf{H} = \begin{bmatrix} g_{11}H & g_{12}H & \cdots & g_{1u}H \\ g_{21}H & g_{22}H & \cdots & g_{2u}H \\ \vdots & & \vdots & \vdots \\ g_{t1}H & g_{t2}H & \cdots & g_{tu}H \end{bmatrix}.$$

The model of (1.6) can be expanded to include the control variates by the method of additional regressors given in Section 3.7 of Seber (1977). Thus, we have

$$\mathbf{y}(\mathbf{R}) = \mathbf{X}\beta + \mathbf{C}(\mathbf{R})\alpha + \varepsilon(\mathbf{R}), \qquad (1.18)$$

where **R** is the selected set of random number streams for the experiment, α is the (s × 1) vector of control coefficients, and

$$\mathbf{C} = \begin{bmatrix} c_1' \\ c_2' \\ \vdots \\ c_m' \end{bmatrix}_{(m \times s)} . \qquad (1.19)$$

Often the (**R**) term in (1.18) is dropped where the dependence of y, **C**, and ε on **R** is understood.

Let $\mathbf{P} = \mathbf{I}_m - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ so that the least-squares estimator of α is $\hat{\alpha} = (\mathbf{C}'\mathbf{P}\mathbf{C})^{-1}\mathbf{C}'\mathbf{P}\mathbf{y}$. Then, by substituting $\hat{\alpha}$ for α and subtracting $\mathbf{C}\hat{\alpha}$ from both sides of (1.18), we get: $\mathbf{y} - \mathbf{C}\hat{\alpha} \cong \mathbf{X}\beta + \varepsilon$. Using the adjusted response vector $\mathbf{y} - \mathbf{C}\hat{\alpha}$ to find the least squares estimate of β yields:

$$\hat{\beta}^{cv} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{y} - \mathbf{C}\hat{\alpha}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{C}\hat{\alpha}. \qquad (1.20)$$

Next, we condition on **C**. Under the joint normality assumption (1.16), Nozari, Arnold, and Pegden (1984) showed that $E[\hat{\beta}^{cv}|\mathbf{C}] = \beta$,

$$\mathrm{Cov}[\hat{\beta}^{cv}|\mathbf{C}] = \tau^2(\mathbf{X}'\mathbf{X})^{-1} + \tau^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{C}(\mathbf{C}'\mathbf{P}\mathbf{C})^{-1}\mathbf{C}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}, \qquad (1.21)$$

where $\tau^2 = \sigma^2 - \Sigma_{yc}\Sigma_c^{-1}\Sigma_{cy}$, and $\hat{\beta}^{cv}|\mathbf{C} \sim N_p(\beta, \mathrm{Cov}[\hat{\beta}^{cv}|\mathbf{C}])$. This result can be used to construct conditional confidence regions and conditional simultaneous confidence intervals for β or its individual components. Assuming r independent replications of the basic m-point experiment, Nozari, Arnold, and Pegden also showed that $E[\hat{\beta}^{cv}] = \beta$ and

$$\mathrm{Cov}[\hat{\beta}^{cv}] = \frac{1}{r}\left(\frac{m-p-1}{m-p-s-1}\right)\tau^2(\mathbf{X}'\mathbf{X})^{-1} \text{ if } m-p-s-1 > 0. \qquad (1.22)$$

Nozari, Arnold, and Pegden used (1.22) to find conditions under which the use of control variates will yield a more efficient estimator of β.

2. THE COMBINED STRATEGY

Schruben and Margolin developed a strategy for effectively combining the two most popular correlation-based variance reduction techniques (common random numbers and antithetic variates) in one simulation experiment. Their results suggest that even more efficient simulation experiments may be obtained by further integration of variance reduction techniques into the experimental protocol. In this section we combine all of the correlation-based techniques (namely, common random numbers, antithetic variates, and control variates) into a unified strategy for the design and analysis of simulation experiments. This combined strategy parallels and extends the development of Schruben and Margolin (1978) and Nozari, Arnold, and Pegden (1984).

The model for the combined approach is

$$\mathbf{y}(\mathbf{R}) = \mathbf{X}\beta + \mathbf{C}(\mathbf{R})\alpha + \mathbf{W}\mathbf{B}(\mathbf{R}) + \varepsilon^o(\mathbf{R}), \qquad (2.1)$$

where y, **R**, **X**, β, **C**, α, **W**, **B**, and ε^o are defined in Section 1. We assume that the joint distribution of y and $\mathbf{c} = (c_1', c_2', \ldots, c_m')'$ is

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{c} \end{bmatrix} \sim N_{m(s+1)}\left(\begin{bmatrix} \mathbf{X} \\ \mathbf{0} \end{bmatrix}\beta, \Sigma^{cm}\right); \qquad (2.2)$$

where

$$\Sigma^{cm} = \begin{bmatrix} \Sigma_y^{cm} & \mathbf{I}_m \otimes \Sigma_{yc} \\ \mathbf{I}_m \otimes \Sigma_{cy} & \mathbf{I}_m \otimes \Sigma_c \end{bmatrix}; \qquad (2.3)$$

Σ_y^{cm} is the unconditional covariance matrix of y under the model (2.1); and Σ_c, Σ_{cy}, and Σ_{yc} are defined in Section 1. Let ρ_1^{cm} and ρ_2^{cm} be the analogs, respectively, of ρ_1 and ρ_2 under the model (2.1) and let

$$\underset{\sim}{\varepsilon} = \mathbf{W}\mathbf{B}(\mathbf{R}) + \underset{\sim}{\varepsilon}^{o}(\mathbf{R}). \qquad (2.4)$$

Now, as a result of (2.1) and (2.2), we know that

$$\underset{\sim}{\varepsilon} \sim N_m(\underset{\sim}{0}, \underset{\sim}{\Sigma}_{\varepsilon}^{cm}),$$

where $\underset{\sim}{\Sigma}_{\varepsilon}^{cm}$ is of the form given by (1.14) with σ^2, ρ_1, and ρ_2 replaced by $(\sigma^2 - \underset{\sim}{\Sigma}_{yc}\underset{\sim}{\Sigma}_c^{-1}\underset{\sim}{\Sigma}_{cy})$, ρ_1^{cm}, and ρ_2^{cm}, respectively. We also make the following two assumptions: (a) $\mathbf{B}$ and $\underset{\sim}{\varepsilon}^o$ are independent and (b) the components of $\underset{\sim}{\varepsilon}^o$ are independent. Assumption (a), coupled with the previous result about the distribution of $\underset{\sim}{\varepsilon}$, implies that both $\mathbf{B}$ and $\underset{\sim}{\varepsilon}$ have normal distributions (see Theorem 19 of Cramer (1970)). Assumption (b) implies that the covariance matrix of $\underset{\sim}{\varepsilon}^o$ is diagonal. In addition, by (2.1) and (2.2), we have

$$\underset{\sim}{\Sigma}_y^{cm} = (\underset{\sim}{\Sigma}_{yc}\underset{\sim}{\Sigma}_c^{-1}\underset{\sim}{\Sigma}_{cy})I_m + \underset{\sim}{\Sigma}_{\varepsilon}^{cm}. \qquad (2.5)$$

We now discuss methods of designing the simulation model (the experimental vehicle) as well as the overall simulation experiment that will ensure the validity of assumptions (2.1) through (2.5). Suppose that the simulation model has been structured so that the g random number streams driving the system can be segregated into two complementary, nonempty groups - the set of g_1 streams that do not affect the control vector $\mathbf{c}$, and the set of $g_2 = g - g_1$ streams that determine the value of $\mathbf{c}$. For example, suppose that in the simulation of a stochastic activity network with g arcs, a separate random number stream is dedicated to sampling each arc duration. If we use a path control vector $\mathbf{c}$ involving a total of g_2 arcs in the network where $g_2 < g$, then $\mathbf{c}$ is stochastically independent of the remaining $g_1 = g - g_2$ random number streams in the simulation. (See Venkatraman and Wilson (1985) for an elaboration of path controls.) As another example, suppose that a simulation model is driven by g random number streams and that a simplified version of this system with a known mean response can be driven by a subset consisting of g_2 of these streams. Then the response of the simplified system defines an external control variable for the original system that is stochastically independent of the remaining $g_1 = g - g_2$ streams used in the original system. (See Kleijnen (1974) for an elaboration of external control variates.)

At the i'th design point in a simulation experiment, let $\mathbf{R}_{i1}$ denote the set of g_1 random number streams that do not affect the control vector $\mathbf{c}_i$ and let $\mathbf{R}_{i2}$ denote the set of g_2 random number streams that determine the value of $\mathbf{c}_i$:

$$\mathbf{R}_i = (\mathbf{r}_{i1}, \mathbf{r}_{i2}, \ldots, \mathbf{r}_{ig_1}, \mathbf{r}_{i(g_1+1)}, \mathbf{r}_{i(g_1+2)}, \ldots, \mathbf{r}_{ig}) = (\mathbf{R}_{i1}, \mathbf{R}_{i2}) \quad \text{for } i = 1, 2, \ldots, m, \qquad (2.6)$$

where $g_1 + g_2 = g$. The $\{\mathbf{R}_{i1}\colon i = 1, 2, \ldots, m\}$ are selected according to the assignment rule of Schruben and Margolin. The R_{i2} $(i = 1, 2, \ldots, m)$ are randomly selected without restriction. This procedure allows $\mathbf{y}$ to have a covariance structure given by $\underset{\sim}{\Sigma}_y^{cm}$ and the $(m \times s)$ matrix C to have independent rows. Thus, we induce the desired covariance structure on $\mathbf{y}$ but not on the $\mathbf{c}_i$'s.

To summarize, if we take

$$R^{\circ} \equiv \begin{bmatrix} R_{11} \\ R_{21} \\ \cdot \\ \cdot \\ \cdot \\ R_{m1} \end{bmatrix} \text{ and } R^{*} \equiv \begin{bmatrix} R_{12} \\ R_{22} \\ \cdot \\ \cdot \\ \cdot \\ R_{m2} \end{bmatrix}$$

so that $R = (R^{\circ}, R^{*})$, then we have

$$\mathbf{y}(R^{\circ}, R^{*}) = X\underset{\sim}{\beta} + C(R^{*})\underset{\sim}{\alpha} + WB(R^{\circ}) + \underset{\sim}{\varepsilon}^{o}(R^{\circ}, R^{*}), \qquad (2.7)$$

with the following properties (for $1 < i, j < m$):

$$\left.\begin{aligned}
&\varepsilon_i(R^\circ, R^*) = b_i(R^\circ) + \varepsilon_i^o(R^\circ, R^*);\\
&E[b_i(R^\circ)] = E[\varepsilon_i^o(R^\circ, R^*)] = 0;\\
&\sigma_i^2 \equiv Var(y_i) = Var[y_i(R^\circ, R^*)] = \sigma^2;\\
&Cov[b_i(R^\circ), b_j(R^\circ)] = (\sigma^2 - \Sigma_{yc}\Sigma_c^{-1}\Sigma_{cy})\rho_1^{cm},\\
&\quad \text{where } 0 < \rho_1^{cm} < 1;\\
&Var[\varepsilon_i^o(R^\circ, R^*)] = (\sigma^2 - \Sigma_{yc}\Sigma_c^{-1}\Sigma_{cy})(1 - \rho_1^{cm});\\
&Cov[b_i(R^\circ), b_j(\bar{R}^\circ)] = (\sigma^2 - \Sigma_{yc}\Sigma_c^{-1}\Sigma_{cy})\rho_2^{cm}\\
&\quad \text{where } -1 < \rho_2^{cm} < 0;\\
&Cov[b_i(R^\circ), \varepsilon_j^o(R^\circ, R^*)]\\
&\quad = Cov[b_i(R^\circ), \varepsilon_j^o(\bar{R}^\circ, \bar{R}^*)] = 0;\\
&Cov[\varepsilon_i^o(R^\circ, R^*), \varepsilon_j^o(R^\circ, R^*)]\\
&\quad = Cov[\varepsilon_i^o(R^\circ, R^*), \varepsilon_j^o(\bar{R}^\circ, \bar{R}^*)] = 0\\
&\quad \text{for } i \neq j;\\
&Cov[b_i(R^\circ), c_j(R^*)]\\
&\quad = Cov[b_i(\bar{R}^\circ), c_j(R^*)] = 0;\\
&\Sigma_y^{cm} = Var[y(R)] \text{ is positive definite;}
\end{aligned}\right\} \quad (2.8)$$

and

$$B(R^\circ) \sim N_2(0, Cov[B]), \qquad (2.9)$$

where Cov[B] is given by (1.18) with σ^2, ρ_1, and ρ_2 replaced by $(\sigma^2 - \Sigma_{yc}\Sigma_c^{-1}\Sigma_{cy})$, ρ_1^{cm}, and ρ_2^{cm}, respectively.

2.2 Summary of Main Results

From (2.7), (2.8), and (2.9) we see that the least squares estimate of β is

$$\hat{\beta}^{cm} = (X'X)^{-1}X'(y - C\hat{\alpha}^{cm}) \qquad (2.10)$$

where

$$\hat{\alpha}^{cm} = (C'PC)^{-1}C'Py \qquad (2.11)$$

and P is given in Section 1.3. If X has rank p then $\hat{\beta}^{cm}$ is unique (see Seber (1977), p. 61). The conditional mean of $\hat{\beta}^{cm}$ given C is $E[\hat{\beta}^{cm}|C] = \beta$ and the conditional covariance of $\hat{\beta}$, given C, is

$$\begin{aligned}Cov[\hat{\beta}^{cm}|C] &= \tau_0^2(X'X)^{-1} + \tau_0^2(X'X)^{-1}X'C(C'PC)^{-1}C'X(X'X)^{-1}\\
&+ (X'X)^{-1}X'WCov[B]W'X(X'X)^{-1}\\
&+ (X'X)^{-1}XC(C'PC)^{-1}C'PWCov[B]W'PC(C'PC)^{-1}C'X(X'X)^{-1},\end{aligned} \qquad (2.12)$$

where $\tau_0^2 = (\sigma^2 - \Sigma_{yc}\Sigma_c^{-1}\Sigma_{cy})(1-\rho_1^{cm})$. These equations, in turn, yield the following unconditional mean and covariance of $\hat{\beta}^{cm}$ when the basic m-point experiment is independently replicated r times: $E[\hat{\beta}^{cm}] = \beta$ and

$$Cov[\hat{\beta}^{cm}] = \begin{bmatrix} Var[\hat{\beta}_0^{cm}] & 0' \\ 0 & Var[\hat{\beta}_1^{cm}] \end{bmatrix}, \qquad (2.13)$$

where

$$Var[\hat{\beta}_0^{cm}] = (\lambda_1^{cm} - \Sigma_{yc}\Sigma_c^{-1}\Sigma_{cy})\left(\frac{r-2}{r-s-2}\right)\left(\frac{1}{mr}\right) \qquad (2.14)$$

$$Cov[\hat{\beta}_1^{cm}] = (\lambda_3^{cm} - \Sigma_{yc}\Sigma_c^{-1}\Sigma_{cy})\left(\frac{m-p-2}{m-p-s-2}\right)\left(\frac{1}{r}\right)(T'T)^{-1}, \qquad (2.15)$$

with T defined in the first paragraph of Section 1.2,

$$\lambda_1^{cm} = (\sigma^2 - \Sigma_{yc}\Sigma_c^{-1}\Sigma_{cy})[1 + (q-1)\rho_1^{cm} + q\rho_2^{cm}], \qquad (2.16)$$

and

$$\lambda_3^{cm} = (\sigma^2 - \Sigma_{yc}\Sigma_c^{-1}\Sigma_{cy})(1-\rho_1^{cm}). \qquad (2.17)$$

If X is orthogonal then (2.15) becomes

$$Cov[\hat{\beta}_1^{cm}] = (\lambda_3^{cm} - \Sigma_{yc}\Sigma_c^{-1}\Sigma_{cy})\left(\frac{m-p-2}{m-p-s-2}\right)\left(\frac{1}{mr}\right)I_p. \qquad (2.18)$$

Next, we compare the combined strategy to the following 4 methods for conducting m-point simulation experiments: independent streams, common random numbers, control variates, and the Schruben-Margolin correlation induction strategy. In each case we assume that X is the same and that the overall experiment consists of r independent replications.

The comparison is based on the notion of domination that can be introduced between some pairs of positive semidefinite symmetric (PSDS) matrices. For such matrices P and Q, we write

$$P \gg Q \text{ if } P - Q \text{ is PSDS.} \qquad (2.19)$$

Further, if $Q \ll P$, then by definition $\Phi(Q) \ll \Phi(P)$ for all nondecreasing functions Φ. Now the determinant, trace, and maximum eigenvalue of a PSDS matrix are nondecreasing functions of that matrix;

thus if (2.19) holds for a given PSDS matrix Q and for all PSDS matrices P, then Q is D-, A-, and E-optimal. (For a more complete discussion of the dominance relationship see the comment by Kiefer in Schruben and Margolin (1978).)

Let $\hat{\beta}^{is}$, $\hat{\beta}^{cs}$, and $\hat{\beta}^{sm}$ be the least squares estimates of β, under the methods of independent streams, common random numbers, and Schruben-Margolin strategy, respectively. Then, we have

$$\text{Cov}[\hat{\beta}^{is}] = \frac{\sigma^2}{r}(\mathbf{X'X})^{-1}, \tag{2.20}$$

$$\text{Cov}[\hat{\beta}^{cs}] = \frac{\sigma^2}{r}\begin{bmatrix} \frac{1}{m}(1-\rho_1)+\rho_1 & 0' \\ 0 & (1-\rho_1)(\mathbf{T'T})^{-1} \end{bmatrix}, \tag{2.21}$$

and

$$\text{Cov}[\hat{\beta}^{sm}] = \frac{\sigma^2}{r}\begin{bmatrix} \frac{1}{2}(\rho_1+\rho_2)+\frac{1}{m}(1-\rho_1) & 0' \\ 0 & (1-\rho_1)(\mathbf{T'T})^{-1} \end{bmatrix}, \tag{2.22}$$

(see Schruben and Margolin (1978)).

From the discussion given above, we see that the covariance matrix in (2.22) compares to the covariance matrices in (2.20) and (2.21) as follows: $\text{Cov}[\hat{\beta}^{sm}] \ll \text{Cov}[\beta^{is}]$, and $\text{Cov}[\hat{\beta}^{sm}] \ll \text{Cov}[\beta^{cs}]$. These two results follow as a consequence of Theorem 1 of Schruben and Margolin (1978). Now, substituting (2.17) into (2.15) and comparing it to $\text{Var}[\hat{\beta}_1^{sm}]$ obtained from (2.25), we get

$$\text{Var}[\hat{\beta}_1^{cm}] \ll \text{Var}[\hat{\beta}_1^{sm}] \tag{2.23}$$

if the following condition is met:

$$[(1-R^2(y,\mathbf{c}))(1-\rho_1^{cm}) - R^2(y,\mathbf{c})](\frac{m-p-2}{m-p-s-2}) < (1-\rho_1), \tag{2.24}$$

where $R(y, c)$ is the coefficient of multiple correlation between y and **c**.

Similarly, working with (2.22), (2.14), and (2.16) yields

$$\text{Var}[\hat{\beta}_0^{cm}] < \text{Var}[\beta^{\wedge sm}_0] \tag{2.25}$$

if

$$\{(1-R^2(y,\mathbf{c}))[q(\rho_1^{cm}+\rho_2^{cm})+(1-\rho_1^{cm})] -R^2(y,\mathbf{c})\}(\frac{r-2}{r-s-2}) < [q(\rho_1+\rho_2)+(1-\rho_1)]. \tag{2.26}$$

Thus, under the conditions of (2.24) and (2.26), we have

$$\text{Var}[\hat{\beta}^{cm}] \ll \text{Var}[\hat{\beta}^{sm}] \tag{2.27}$$

since **X** is orthogonally blockable, which implies that $\hat{\beta}_0^{cm}$ and $\hat{\beta}_1^{cm}$ are independent and that $\hat{\beta}_0^{sm}$ and $\hat{\beta}_1^{sm}$ are independent. Also, in comparing (1.22) to (2.14) and (2.15), we get

$$\text{Var}[\hat{\beta}_0^{cm}] < \text{Var}[\beta_0^{cv}] \tag{2.28}$$

if

$$([q(\rho_1^{cm}+\rho_2^{cm})+(1-\rho_1^{cm})] - \frac{R^2(y,\mathbf{c})}{(1-R^2(y,\mathbf{c}))})(\frac{r-2}{r-s-2}) < (\frac{m-p-1}{m-p-s-1}), \tag{2.29}$$

and

$$\text{Cov}[\hat{\beta}_1^{cm}] \ll \text{Cov}[\beta_1^{cv}] \tag{2.30}$$

if

$$((1-\rho_1^{cm}) - \frac{R^2(y,\mathbf{c})}{(1-R^2(y,\mathbf{c}))})(\frac{m-p-2}{m-p-s-2})(\frac{m-p-s-1}{m-p-1}) < 1. \tag{2.31}$$

We can summarize these results with the following theorem:

Theorem 2.1: If the conditions of (2.24), (2.26), (2.30), and (2.31) are met, then with respect to D-, E-, and A-optimality in the estimation of β, the combined strategy (2.1) is superior to the following methods: (a) independent streams, (b) common random numbers, (c) Schruben-Margolin strategy, and (d) control variates.

Thus, the combined approach can give the best estimates of the coefficients in the metamodel of all methods considered in this chapter.

3. EXAMPLE

In this section we illustrate the implementation of combining the use of the Schruben-Margolin correlation induction strategy and control variates in a simulation experiment. We also compare the estimators of the metamodel coefficients under the combined strategy, control variates, and the Schruben-Margolin strategy.

3.1. The Job Shop System

Consider a job shop example similar to the one given by Nozari, Arnold, and Pegden (1987), and depicted in Figure 1. This example was chosen to maintain consistency with earlier work done by Tew and Wilson (1987) on validating the Schruben-Margolin strategy. Jobs arrive at this shop according to a Poisson process with an arrival rate of 10 per hour. All jobs enter the system through station 1. Upon completing service at station 1, 80% of the jobs go to station 2, 5% go to station 3, and 15% leave the system. A job at station 2, or station 3, leaves the system upon completion of service. The shop admits jobs from 8:00 A.M. to 4:00 P.M. every day. However, service at each station continues until all jobs admitted on one day leave the system. Service time at station 1 is a constant and service times at stations 2 and 3 are uniformly distributed over specified ranges.

The purpose of this example is to estimate the effects that different service time distributions have on some function of the expected system sojourn time for a job. Thus, the performance measure of interest is the daily average system sojourn time for all jobs entering the system. This estimation is done under the following three techniques for conducting a simulation experiment: control variates, the Schruben-Margolin strategy, and the combined use of control variates and the Schruben-Margolin strategy.

3.2. The Model of the Response

To study this system we employ a 2^3 factorial design with the following independent variables (factors): service time distribution at station 1 (x_1), service time distribution at station 2 (x_2), and service time distribution at station 3 (x_3). We consider a first order model without interactions given by

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon, \qquad (3.1)$$

where y is the performance measure of interest, x_i ($i = 1, 2, 3$) is defined above, $\underset{\sim}{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)'$ is the vector of unknown model parameters, and ε is the inability of $\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$ to determine y.

We also consider two standardized control variables based on the service times at station 2 and the service times at station 3. Let $U_k(k)$ denote the j'th service time sampled at station k and μ_k and σ_k be the mean and standard deviation, respectively, of the service time distribution at station k ($k = 2, 3$). Also, let $a(k,t)$ denote the number of service times that are sampled at station k during the (simulated) time period $[0,t]$. The standardized control variable accumulated at station k up to time t is

$$c_k(t) = [a(k,t)]^{-1/2} \sum_{j=1}^{a(k,t)} [U_j(k) - \mu_k]/\sigma_k \quad \text{for } k = 2, 3, \qquad (3.2)$$

(Wilson and Pritsker (1984)). Furthermore, from Wilson and Pritsker (1984), we have that

$$c(t) = (c_2(t), c_3(t))' \xrightarrow{L} N_2(\underset{\sim}{0}, I_2) \text{ as } t \to \infty, \qquad (3.3)$$

if the service times are sampled independently. Thus, the model in (3.1) becomes

$$(y - c\underset{\sim}{\alpha}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon. \qquad (3.4)$$

In simulating this system we dedicated a separate random number stream to each of the following four

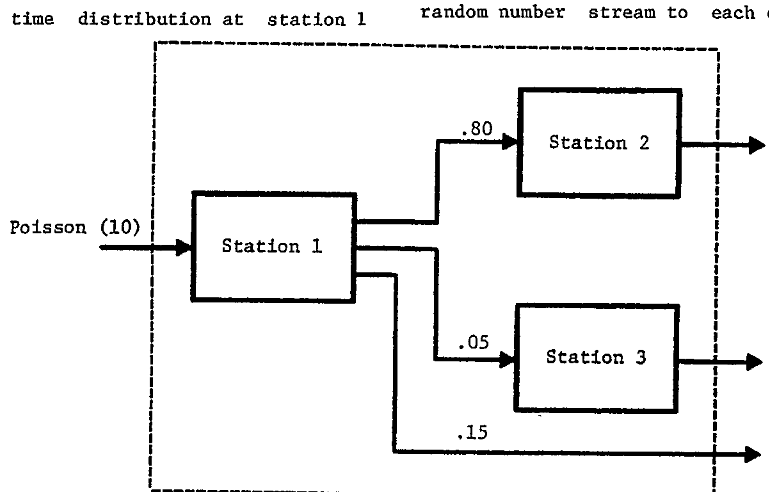

Figure 1. Job Shop System

random components in the model: interarrival times at station 1 (r_1), probabalistic branching upon completion of service at station 1 (r_2), service times at station 2 (r_3), and service times at station 3 (r_4). Under the Schruben-Margolin strategy all four random number streams are used for blocking whereas under the use of control variates and the combined strategy of control variates and the Schruben-Margolin strategy only r_1 and r_2 are used for blocking because r_3 and r_4 are used to generate the standardized control variables at stations 2 and 3, respectively.

Next, we consider the estimation of $\underset{\sim}{\beta}$ under each of the three techniques for conducting simulation experiments mentioned above. In each case, 20 independent estimates of $\underset{\sim}{\beta}$ are obtained by independently replicating the basic 8-point experiment 20 times. The sample covariance matrix based on these 20 estimates is used as an external estimate of the covariance matrix of the estimator of β.

3.3. Numerical Results

For the Schruben-Margolin strategy we get

$$\hat{Cov}(\hat{\underset{\sim}{\beta}}^{sm}) = \begin{bmatrix} 4.804 & .689 & -.271 & 1.472 \\ .689 & .810 & .025 & -.047 \\ -.271 & .025 & .070 & -.098 \\ 1.472 & -.047 & -.098 & .629 \end{bmatrix},$$

which yields $tr[\hat{Cov}(\hat{\underset{\sim}{\beta}}^{sm})] = 6.313$ and $det[\hat{Cov}(\hat{\underset{\sim}{\beta}}^{sm})] = .0138$. For the control variates technique we get

$$\hat{Cov}(\hat{\underset{\sim}{\beta}}^{cv}) = \begin{bmatrix} 29.498 & 8.452 & -4.277 & -.392 \\ 8.452 & 22.484 & 2.277 & .594 \\ -4.277 & 2.277 & 19.659 & 8.385 \\ -.392 & .594 & 8.385 & 24.181 \end{bmatrix},$$

which yields $tr[\hat{Cov}(\hat{\underset{\sim}{\beta}}^{cv})] = 95.823$ and $det[\hat{Cov}(\hat{\underset{\sim}{\beta}}^{cv})] = 222{,}990.360$. Finally, for the combined strategy of control variates and the Schruben-Margolin strategy we get

$$\hat{Cov}(\hat{\underset{\sim}{\beta}}^{cs}) = \begin{bmatrix} 25.675 & 8.513 & -8.188 & -6.594 \\ 8.513 & 25.700 & -3.817 & -5.749 \\ -8.188 & -3.817 & 21.176 & 6.979 \\ -6.594 & -5.749 & 6.979 & 21.919 \end{bmatrix},$$

which yields $tr[\hat{Cov}(\hat{\underset{\sim}{\beta}}^{cs})] = 24.471$ and $det[\hat{Cov}(\hat{\underset{\sim}{\beta}}^{cs})] = 195{,}610.348$.

4. CONCLUSIONS

Although any statistical comparison of the sample covariance matrices given in Section 3.3 will have low power due to the small number of replications the results suggest that the Schruben-Margolin strategy gave superior performance to the use of control variates and the combined use of control variates and the Schruben-Margolin strategy. We believe this is due to the large block effect induced under the Schruben-Margolin strategy brought about by the simple structure of the system and the use of all four random number streams for blocking. In effect, the simplicity of the system allows the block effect to account for most of the variability in the model. We expect that with a more complex system that this would not be the case and that the efficiency due to the control variables would surpass the efficiency due to the block effect. Currently, we are investigating the combined use of control variates and the Schurben-Margolin strategy for a more complex stochastic system.

REFERENCES

Cramer, H. (1970). Random Variables and Probability Distributions. Cambridge University Press, Cambridge.

Kleijnen, J.P.C. (1974). Statistical Techniques in Simulation, Part I. Marcel Dekker, New York.

Myers, R.H. (1976). Response Surface Methodology. Allyn and Bacon, Boston, Massachusetts.

Nozari, A., Arnold, S.F., and Pegden, C.D. (1984). Control variates for multipopulation simulation experiments. IIE Transactions 16, 159-169.

Nozari, A., Arnold, S. F., and Pegden, C. D., (1987). Statistical analysis for use with the Schruben and Margolin correlation induction strategy. Operations Research 35, 127-137.

Schruben, L.W. and Margolin, B.H. (1978). Pseudorandom number assignment in statistically designed simulation and distribution sampling experiments. Journal of American Statistical Association 73, 504-525.

Schruben, L.W. (1979). Designing correlation induction strategies for simulation experiments, Chapter 16 in Current Issues in Computer Simulation. Academic Press, New York, 235-255.

Seber, G.A.F. (1977). Linear Regression Analysis. John Wiley & Sons, New York.

Tew, J.D. (1986). Metamodel estimation under correlation methods for simulation experiments. Unpublished Ph.D. dissertation, School of Industrial Engineering, Purdue University, West Lafayette, Indiana.

Tew, J. D. and Wilson, J. R. (1987). Validation of correlation-induction strategies for simulation experiments. Technical Report VTR 8701, Department of Industrial Engineering and Operations Research, Virginia Polytechnic Institute and State University, Blacksburg, Virginia.

Venkatraman, S. and Wilson, J.R. (1985). Using path control variates in activity network simulation. In: Proceedings of the 1985 Winter Simulation Conference (D. T. Gantz, G. C. Blais, and S. L. Solomon, ed.). Institute of Electrical and Electronics Engineers, San Francisco, California 217-222.

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

AUTHORS' BIOGRAPHIES

JEFFREY D. TEW is an Assistant Professor in the Department of Industrial Engineering and Operations Research at Virginia Polytechnic Institute and State University. He received a B.S. in mathematics from Purdue University in 1979, a M.S. in statistics from Purdue University in 1981, and a Ph.D. in industrial engineering from Purdue University in 1986. His current research interests include simulation optimization and design of experiments. He is a member of ASA, IIE, IMS, ORSA, and SCS.

Jeffrey D. Tew
Department of IEOR
Virginia Polytechnic Institute and State University
Blacksburg, Virginia 24061
(703) 961-7099

JAMES R. WILSON is an Associate Professor in the School of Industrial Engineering at Purdue University. He received a B.A. in mathematics from Rice University in 1970, and M.S. and Ph.D. degrees in Industrial Engineering from Purdue University in 1977 and 1979 respectively. He has been involved in various simulation studies while working as a research analyst for the Houston Lighting & Power Company (1970-1972) and while serving as a U.S. Army officer (1972-1975). From 1979 to 1984, he was an Assistant Professor in the Operations Research Group of the Mechanical Engineering Department in The University of Texas at Austin. His current research interests include simulation output analysis, variance reduction techniques, ranking-and-selection procedures, and stopping rules. He is a member of ASA, IIE, ORSA, SCS, and TIMS.

James R. Wilson
School of Industrial Engineering
Purdue University
West Lafayette, Indiana 47907
(317) 494-5408