METAMODEL ESTIMATION USING INTEGRATED CORRELATION METHODS

Jeffrey D. Tow  
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, ..., y_m). Also, let ϕ_i be the setting of the d factors for the i'th design point and let \{k: k = 1, 2, ..., 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

\[ y_i = \beta_0 + \sum_{k=1}^{p-1} \beta_k x_k(\phi_i) + \epsilon_i \]

for i = 1, 2, ..., m,  \hspace{1cm} (1.1)

where \{\beta_k: k = 0, 1, ..., p-1\} are the unknown model parameters and \epsilon_i represents the inability of \beta_0 + \sum_{k=1}^{p-1} \beta_k x_k(\phi_i) to determine \epsilon_i. Define X to be the (m x p) matrix whose first column is all ones and whose (i, k+1) element is \{X_k(\phi_i)\} for i = 1, 2, ..., m and k = 1, 2, ..., 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:

\[ y = X\beta + \epsilon. \]

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

\[ X'X = \alpha I_p, \]

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

We also assume that

\[ (y_h: h = 1, 2, ..., r) \text{ IID } \sim N_m(\mu, \Sigma), \]

where \(y_h\) denotes the response at the i'th design point on the h'th independent replication of the basic m-point experiment and \Sigma is the (m x m) covariance matrix.

A simulation model is usually driven by randomly chosen streams 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 \(\Gamma_{ij} = (r_{ij1}, r_{ij2}, \ldots)\) for i = 1, 2, ..., m and j = 1, 2, ..., g; (b) the set
of streams for the i'th design point is
\[ R_i = (r_{i1}, r_{i2}, \ldots, r_{ig}) \quad \text{for} \quad i = 1, 2, \ldots, m; \quad (1.5) \]
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
\[ y_i(R_i) = \theta_0 + \sum_{k=1}^{p-1} b_k x_k(\phi_i) + e_i(R_i) . \quad (1.6) \]
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 \( X = (1_m) \) 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
\[ TW = 0 \quad \text{and} \quad W = [m_1, m_2], \]
where \( m_1 \) and \( m_2 \) are the respective block sizes. If we let
\[ 1 - r_{ij} \equiv (1 - r_{i1j}, 1 - r_{i2j}, \ldots) \]
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
\[ R_i = R_i = (r_{i1}, r_{i2}, \ldots, r_{ig}) , \]
where \( i = 1, 2, \ldots, m_1 \) and for all \( m_2 \) design points in the second block, use the antithetic (complementary) set of pseudorandom numbers so that
\[ R_i = R_i = (1 - r_{i1}, \ldots, 1 - r_{ig}) , \]
1. Schruben and Margolin decomposed the error term \( e_i \) into a random block effect \( b_i \) and a residual \( e_i^0 \), both of which are functions of \( R \). Thus the model in (1.1) can be written:
\[ y_i(R) = \theta_0 + \sum_{k=1}^{p-1} b_k x_k(\phi_i) + b_i(R) + e_i^0(R) \]
for \( i = 1, 2, \ldots, m \). (1.7)

In order to analyze the properties of this assignment rule, Schruben and Margolin made the following assumptions (for \( i < j < m \)):
\[ e_i(R) = b_i(R) + e_i^0(R); \]
\[ E[b_i(R)] = E[e_i^0(R)] = 0; \]
\[ \sigma_i^2 \equiv Var(y_i) = Var[y_i(R)] = \sigma^2; \]
\[ Cov(b_i(R), b_j(R)) = \rho_1 \sigma^2, \quad \text{where} \quad 0 < \rho_1 < 1; \]
\[ Var[e_i^0(R)] = \sigma^2(1 - \rho_1); \]
\[ Cov(b_i(R), e_j^0(R)) = \rho_2 \sigma^2 \quad \text{where} \quad -1 < \rho_2 < 0; \]
\[ Cov(e_i^0(R), e_j^0(R)) = Cov(e_i^0(R), e_j^0(R)) = 0; \]
\[ Cov(e_i^0(R), e_j^0(R)) = Cov(e_i^0(R), e_j^0(R)) = 0 \quad \text{for} \quad i \neq j; \]
\[ \Sigma = ICov(y_i, y_j) \] is positive definite.

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 \( i \neq j \)) are realized from the same random number stream, then
\[ \text{Corr}[y_i(R), y_j(R)] = \rho_1, \quad 0 < \rho_1 < 1. \]
3. If \( y_i \) and \( y_j \) (for \( i \neq j \)) are realized from antithetic (complementary) random number streams, \( R \) and \( \bar{R} \) respectively, then
\[ \text{Corr}[y_i(R), y_j(\bar{R})] = \rho_2, \]
\[ -1 < \rho_2 < 0. \]

Under the Schruben-Margolin strategy with equal block sizes, the metamodel of (1.6) takes the following form:
\[ y(R) = X\theta + Wb(R) + \varepsilon^0(R), \]
where: \( q \equiv \frac{m}{2} = m_1 = m_2 \) is the common block size; \( B' = [b(R_1), b(R_2)]' \) is the \((2 \times 1)\) vector of random block effects; \( W \) is the \((m \times 2)\) block incidence matrix.
defined by Schruben and Margolin (1978) and Schruben (1979) and \( \mathbf{e}^p \) is the \((m \times 1)\) 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 \( X = [X_1 X_2]' \), then we get

\[
W = \begin{bmatrix}
1 & 0 \\
0 & 1 \\
-1 & -1 \\
0 & 0
\end{bmatrix},
\]

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

\[
\text{Cov} [\mathbf{e}] = \sigma^2 \begin{bmatrix}
\rho_1 & \rho_2 \\
\rho_2 & \rho_1
\end{bmatrix},
\]

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

\[
\Sigma_0 = \sigma^2 \begin{bmatrix}
1 & \rho_1 & \rho_2 & \cdots & \rho_{i-1} \\
\rho_1 & 1 & \rho_2 & \cdots & \rho_{i-1} \\
\rho_2 & \rho_2 & 1 & \cdots & \rho_{i-1} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\rho_{i-1} & \rho_{i-1} & \rho_{i-1} & \cdots & 1
\end{bmatrix},
\]

where \( \Sigma_{11} \) is \((m_1 \times m_1)\), \( \Sigma_{12} \) is \((m_1 \times m_2)\), \( \Sigma_{21} \) is \((m_2 \times m_1)\), and \( \Sigma_{22} \) 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 \( \hat{\beta} \) 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

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

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 \( \beta_0 \); 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, \( \mathbf{e} = (e_1, e_2, \ldots, e_s)' \), having both a known mean \( \mu_e \) 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 \( c - \mu_c \). 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 \( Y_i \) \((i = 1, 2, \ldots, m)\), we also observe an \( s \)-dimensional column vector of control variates \( c_i \). In this situation we may assume without loss of generality, that \( E(c_i) = 0 \) \((i = 1, 2, \ldots, m)\). Moreover, we assume that at each experimental point, the response and the control variates are jointly normal.

If we let \( \mathbf{e} = (e_1, e_2, \ldots, e_s)' \), then we have

\[
\begin{bmatrix}
y \\
\mathbf{e}
\end{bmatrix} \sim N_{m+s+1}(X_0 \mathbf{e}, \Sigma_{cv}),
\]

where

\[
\Sigma_{cv} = \begin{bmatrix}
\sigma^2_{m, m} & \mathbf{I}_m \otimes \mathbf{C}_c \\
\mathbf{I}_m \otimes \mathbf{C}_y & \mathbf{I}_m \otimes \mathbf{I}_s
\end{bmatrix},
\]

\( \mathbf{C}_c \) is the \((s \times s)\) unknown covariance matrix of \( c_i \), and \( \mathbf{C}_y \) is a \((s \times 1)\) unknown covariance matrix whose
elements are the covariances between $y_i$ and $e_1$. (Note that (1.16) indicates that $\Sigma_c$ and $\Sigma_{yc}$ 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 $(t \times u)$ matrix and $H$ is a $(v \times s)$ matrix, then the right direct product of $G$ and $H$ is the $(tv \times us)$ matrix

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

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

$$y(R) = X\beta + C(R)\xi + \xi(R), \quad (1.18)$$

where $R$ is the selected set of random number streams for the experiment, $\beta$ is the $(s \times 1)$ vector of control coefficients, and

$$C =
\begin{bmatrix}
c_1' \\
c_2' \\
\vdots \\
c_m'
\end{bmatrix}
\quad (m \times s)
$$

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

Let $\hat{P} = \hat{X} - X(X'X)^{-1}X'$ so that the least-squares estimator of $\beta$ is $\hat{\beta} = (C'P'C)^{-1}C'P'y$. Then, by substituting $\hat{P}$ for $P$ and subtracting $C_0$ from both sides of (1.18), we get: $y - C_0 = \hat{X} + \xi$. Using the adjusted response vector $y - C_0$ to find the least squares estimate of $\beta$ yields:

$$\hat{\beta}_{cv} = (X'X)^{-1}X'y - C_0 = (X'X)^{-1}X'y - (X'X)^{-1}X'C_0.
$$

(1.20)

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

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

(1.21)

where $\tau^2 = \sigma^2 - E_{yc}^2 - E_{yc}^2$, and

$$\hat{\Sigma}_{cv} | C \sim \mathcal{N} \left( \beta, \text{Cov}[\hat{\beta}_{cv} | C] \right).$$

This result can be used to construct conditional confidence regions and conditional simultaneous confidence intervals for $\beta$ 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

$$\text{Cov}[\hat{\beta}_{cv}'] = \frac{1}{r} \left( \frac{m-1}{m-p-s-1} \right) \tau^2(X'X)^{-1}$$

if $m-p-s-1 > 0$.

(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 $\beta$.

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

$$y(R) = X\beta + C(R)\xi + WB(R) + \xi'(R), \quad (2.1)$$

where $y$, $R$, $X$, $\beta$, $C$, $\xi$, $W$, $B$, and $\xi'$ are defined in Section 1. We assume that the joint distribution of $y$ and $c = (c_1', c_2', \ldots, c_m')'$ is

$$
\begin{bmatrix}
y \\
c
\end{bmatrix}
\sim \mathcal{N}(m+c+1)
\begin{bmatrix}
X\beta \\
n \tau^2 \Sigma_{cm}
\end{bmatrix};
$$

(2.2)

where

$$\Sigma_{cm} =
\begin{bmatrix}
\rho_{cm} & I_m \otimes \Sigma_y \\
I_m \otimes \Sigma_{yc} & I_m \otimes \Sigma_c
\end{bmatrix},
$$

(2.3)

$\Sigma_y$ is the unconditional covariance matrix of $y$ under the model (2.1) and $\Sigma_{yc}$ and $\Sigma_{yc}$ are defined in Section 1. Let $\rho_1$ and $\rho_2$ be the anlogous, respectively, of $\rho_1$ and $\rho_2$ under the model (2.1) and
let
\[ \xi = \mathbf{W}\mathbf{B}(\mathbf{R}) + \xi^0(\mathbf{R}). \] (2.4)

Now, as a result of (2.1) and (2.2), we know that
\[ \xi \sim \mathbb{N}_m(0, \mathbf{\Sigma}^\text{cm}), \]
where \( \mathbf{\Sigma}^\text{cm} \) is of the form given by (1.14) with \( \sigma^2, \rho_c, \)
and \( \rho_2 \) replaced by \( (\sigma^2 - \xi_{yc}^{-1}_c \xi_{cy})^{-1}, \rho_1^c, \) and \( \rho_2^c \), respectively. We also make the following two assumptions: (a) \( \mathbf{B} \) and \( \xi^0 \) are independent and (b) the components of \( \xi^0 \) are independent. Assumption (a),
coupled with the previous result about the distribution of \( \xi \), implies that both \( \mathbf{B} \) and \( \xi \) have normal distributions (see Theorem 19 of Cramer (1970)). Assumption (b) implies that the covariance matrix of \( \xi^0 \) is diagonal. In addition, by (2.1) and (2.2), we have
\[ \mathbf{\Sigma}^\text{cm} = (\xi_{yc}^{-1}_c \xi_{cy})^{-1}_m + \mathbf{\Sigma}^\text{cm}. \] (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 Venkatarman 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 variables.)

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

\[ \mathbf{R}_{i1} = (R_{i11}, R_{i12}, \ldots, R_{i1(g_1+1)}, R_{i1(g_1+2)}, \ldots, R_{i1n}) = (R_{i11}, R_{i12}) \] for \( i = 1, 2, \ldots, m, \)

where \( g_1 + g_2 = g \). The \( (R_{i1i}: i = 1, 2, \ldots, m) \) are selected according to the assignment rule of Schrubin
and Margolin. The \( R_{i1j} (j = 1, 2, \ldots, m) \) are randomly selected without restriction. This procedure
allows \( y \) to have a covariance structure given by \( \mathbf{\Sigma}_y^c \)
and the \((m \times s)\) matrix \( \mathbf{C} \) to have independent rows. Thus, we induce the desired covariance structure on \( y \) but not on the \( c_i \)'s.

To summarize, if we take
\[ \mathbf{R}^c = \begin{bmatrix} R_{11} \\ R_{21} \\ \vdots \\ R_{m1} \end{bmatrix} \quad \text{and} \quad \mathbf{R}_y^c = \begin{bmatrix} R_{12} \\ R_{22} \\ \vdots \\ R_{m2} \end{bmatrix} \]
so that \( \mathbf{R} = (\mathbf{R}^c, \mathbf{R}_y^c) \), then we have
\[ y(\mathbf{R}^c, \mathbf{R}_y^c) = \mathbf{X}_y + \mathbf{C}(\mathbf{R}^c)\mathbf{R}_y + \mathbf{WB}(\mathbf{R}^c) + \xi^0(\mathbf{R}^c, \mathbf{R}_y^c), \] (2.7)
with the following properties (for \( 1 \leq i, j \leq m \)):
\[ e_{1}(R^{*}, R^{*}) = b_{1}(R^{*}) + e_{1}^{0}(R^{*}, R^{*}); \]
\[ E[b_{1}(R^{*})] = E[e_{1}^{0}(R^{*}, R^{*})] = 0; \]
\[ o_{1}^{2} = \text{Var}(y_{1}) = \text{Var}[y_{1}(R^{*}, R^{*})] = o^{2}; \]
\[ \text{Cov}[b_{1}(R^{*}), b_{j}(R^{*})] = (o^{2} - E_{yc}E_{cy})\rho_{1,j}^{cm}, \]

where \( 0 < \rho_{1}^{cm} < 1; \)
\[ \text{Var}[e_{1}^{0}(R^{*}, R^{*})] = (o^{2} - E_{yc}E_{cy})\left(1 - \rho_{1}^{cm}\right); \]
\[ \text{Cov}[b_{1}(R^{*}), b_{j}(R^{*})] = (o^{2} - E_{yc}E_{cy})\rho_{2}^{cm} \]

where \(-1 < \rho_{2}^{cm} < 0; \)
\[ \text{Cov}[\beta_1^{cm}, \beta_j^{cm}] = \left[ \begin{array}{cc} \text{Var}[\beta_1^{cm}] & 0 \\ 0 & \text{Var}[\beta_j^{cm}] \end{array} \right], \]

(2.8)

(2.9)

\[ \text{Cov}[\beta_1^{cm}] = \left(\lambda_{1}^{cm} - E_{yc}E_{cy}\right)\left(\frac{r-2}{r-\sigma^{2}}\frac{1}{mr}\right), \]

(2.14)

\[ \text{Cov}[\beta_2^{cm}] = \left(\lambda_{2}^{cm} - E_{yc}E_{cy}\right)\left(\frac{mp-2}{mp-\sigma^{2}}\right)\left(\frac{1}{mr}\right) \]

(2.15)

where \( T \) defined in the first paragraph of Section 1.2,
\[ \lambda_{1}^{cm} = (o^{2} - E_{yc}E_{cy})(1 + (q-1)p_{1}^{cm} + q_{2}^{cm}), \]

(2.16)

and
\[ \lambda_{3}^{cm} = (o^{2} - E_{yc}E_{cy})(1 - p_{1}^{cm}). \]

(2.17)

If \( X \) is orthogonal then (2.15) becomes
\[ \text{Cov}[\beta_1^{cm}] = \left(\lambda_{1}^{cm} - E_{yc}E_{cy}\right)\left(\frac{mp-2}{mp-\sigma^{2}}\right)\left(\frac{1}{mr}\right)X. \]

(2.18)

Next, we compare the combined strategy to the following four methods for conducting \( m \)-point simulation experiments: independent streams, common random numbers, control variates, and the Schruben-Hargolin 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
\[ P \gg Q \text{ if } P - Q \text{ is PSDS.} \]

(2.19)

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

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 \( \hat{\beta} \), under the methods of independent
streams, common random numbers, and Schruben-Margolin
strategy, respectively. Then, we have

\[
\text{Cov}[\hat{\beta}_{is}] = \frac{\sigma^2}{\tau} (X'X)^{-1},
\]

(2.20)

\[
\text{Cov}[\hat{\beta}_{cs}] = \frac{\sigma^2}{\tau} \begin{bmatrix}
\frac{1}{n} (1-\rho_1)^2 + \rho_1 & 0 \\
0 & (1-\rho_1) (Y'Y)^{-1}
\end{bmatrix},
\]

(2.21)

and

\[
\text{Cov}[\hat{\beta}_{sm}] = \frac{\sigma^2}{\tau} \begin{bmatrix}
\frac{1}{n} (\rho_1 + \rho_2) + \frac{1}{m} (1-\rho_1) & 0 \\
0 & (1-\rho_1) (Y'Y)^{-1}
\end{bmatrix},
\]

(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:
\( \text{Cov}[\hat{\beta}_{sm}] \ll \text{Cov}[\hat{\beta}_{is}] \), and \( \text{Cov}[\hat{\beta}_{sm}] \ll \text{Cov}[\hat{\beta}_{cs}] \). 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 \( \text{Var}[\hat{\beta}_{sm}] \)
obtained from (2.25), we get

\[
\text{Var}[\hat{\beta}_{sm}] \ll \text{Var}[\hat{\beta}_{is}]
\]

(2.23)

if the following condition is met:

\[
[(1-R^2(y, c))(1-\rho_1)^2 - R^2(y, c))(m-p-2)/(m-p-s-2)] < (1-\rho_1),
\]

(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

\[
\text{Var}[\hat{\beta}_{sm}] \ll \text{Var}[\hat{\beta}_{cs}]
\]

(2.25)

if

\[
[(1-R^2(y, c))(1-\rho_1^2) - R^2(y, c))(m-p-2)/(m-p-s-2)] < (1-\rho_1^2),
\]

(2.26)

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

\[
\text{Var}[\hat{\beta}_{sm}] \ll \text{Var}[\hat{\beta}_{is}]
\]

(2.27)

since \( X \) is orthogonally blockable, which implies that
\( \hat{\beta}_{cs} \) and \( \hat{\beta}_{sm} \) are independent and that \( \hat{\beta}_{sm} \) and \( \hat{\beta}_{is} \) are independent.

Also, in comparing (1.22) to (2.14) and
(2.15), we get

\[
\text{Var}[\hat{\beta}_{sm}] \ll \text{Var}[\hat{\beta}_{cs}]
\]

(2.28)

if

\[
\text{Var}[\hat{\beta}_{sm}] \ll \text{Var}[\hat{\beta}_{is}]
\]

(2.29)

if

\[
\text{Var}[\hat{\beta}_{sm}] \ll \text{Var}[\hat{\beta}_{cs}]
\]

(2.30)

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 \( \hat{\beta} \),
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
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 Nazari, Arnold, and Pagden (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

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

where $y$ is the performance measure of interest, $x_i$ ($i = 1, 2, 3$) is defined above, $\beta = (\beta_0, \beta_1, \beta_2, \beta_3)'$ is the vector of unknown model parameters, and $\epsilon$ 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 $U_k(t)$ denote the $j$th service time sampled at station $k$ and $\mu_k$ and $\sigma_k$ 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

$$ c_k(t) = [a_k(t)]^{-1/2} \sum_{j=1}^{a_k(t)} [U_j(t) - \mu_k(t)] / \sigma_k $$

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

$$ a(t) = (c_2(t), c_3(t))' \rightarrow N_x(0, I_x) \text{ as } t \rightarrow \infty, \quad (3.3) $$

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

$$ (y - c_0) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon. \quad (3.4) $$

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

![Figure 1. Job Shop System](image-url)
random components in the model: interarrival times at station 1 (\(r_1\)), probabilistic 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 variates at stations 2 and 3, respectively.

Next, we consider the estimation of \(\beta\) under each of the three techniques for conducting simulation experiments mentioned above. In each case, 20 independent estimates of \(\beta\) 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 \(\beta\).

### 3.3. Numerical Results

For the Schruben-Margolin strategy we get

\[
\text{Cov}(\hat{\beta}^{\text{SM}}) = \begin{bmatrix}
4.804 & 0.689 & -0.271 & 1.472 \\
0.689 & 0.801 & 0.025 & -0.047 \\
-0.271 & 0.025 & 0.070 & -0.098 \\
1.472 & -0.047 & -0.098 & 0.629
\end{bmatrix},
\]

which yields \(\text{tr}[\text{Cov}(\hat{\beta}^{\text{SM}})] = 6.313\) and \(\det[\text{Cov}(\hat{\beta}^{\text{SM}})] = 0.0138\). For the control variates technique we get

\[
\text{Cov}(\hat{\beta}^{\text{CV}}) = \begin{bmatrix}
29.498 & 8.452 & -4.277 & -0.392 \\
8.452 & 22.464 & 2.277 & 0.594 \\
-4.277 & 2.277 & 19.659 & 8.385 \\
-0.392 & 0.594 & 8.385 & 24.181
\end{bmatrix},
\]

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

\[
\text{Cov}(\hat{\beta}^{\text{CS}}) = \begin{bmatrix}
25.673 & 8.513 & -8.188 & -6.594 \\
8.513 & 25.700 & -3.817 & -5.749 \\
-8.188 & -3.817 & 21.176 & 6.979 \\
\end{bmatrix},
\]

which yields \(\text{tr}[\text{Cov}(\hat{\beta}^{\text{CS}})] = 24.471\) and \(\det[\text{Cov}(\hat{\beta}^{\text{CS}})] = 195,610.348\).

### 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 variates would surpass the efficiency due to the block effect. Currently, we are investigating the combined use of control variates and the Schruben-Margolin strategy for a more complex stochastic system.

### References


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, INSI, 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 at 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) 494-5409