Proceedings of the 1988 Winter Simulation Conference
M. Abrams, P. Haigh, and J. Comfort (eds.)

Minimal-MSE linear combinations of variance estimators of the sample mean

Wheyming Tina Song
Bruce Schmeiser
School of Industrial Engineering
Purdue University
West Lafayette, IN 47907, U.S.A.

ABSTRACT

We continue our investigation of linear combinations of variance-of-the-sample-mean estimators that are parameterized by batch size. First we state the mse-optimal linear-combination weights in terms of the bias vector and the covariance matrix of the component estimators for two cases: weights unconstrained and weights constrained to sum to one. Then we report a small numerical study that demonstrates mse reduction of about 80% for unconstrained weights and about 30% for constrained weights. The mse's and the percent reductions are similar for all four estimator types considered. Such large mse reductions could not be achieved in practice, since they assume knowledge of unknown parameters, which would have to be estimated. Optimal-weight estimation is not considered here.

1. INTRODUCTION

Consider a covariance-stationary sequence of random variables $\{X_i\}_{i=1}^n$ having mean μ, positive variance R_0 , and finite fourth moment μ_4. Estimating var($\bar{X}$), the variance of the sample mean, is a prototype problem in the analysis of simulation output. Several estimation techniques have been proposed to estimate var($\bar{X}$), including regenerative (Crane and Iglehart 1975), ARMA time series models (Fishman 1978, Schriber and Andrew 1984), spectral (Heidelberger and Welch 1981), standardized time series (Goldsman 1984, Schruben 1983), nonoverlapping batch means (Schmeiser 1982), and overlapping batch means (Meketon and Schmeiser 1984).

All of these methods work well only when underlying assumptions hold. The statistical properties deteriorate as a function of the degree to which the assumptions are violated. Also, most of the variance estimators have one or two parameters that determine the estimators' properties. The most common parameter is batch size, which is used by batch-means and some standardized-time-series estimators. Batch-size estimation based on hypothesis testing has been the classical approach (Bratley, Fox, and Schrage 1987, Law and Kelton 1983); optimal batch size in terms of the covariance structure has been recently discussed by Goldsman and Meketon (1986) and Song and Schmeiser (1988b).

We hope to create an estimator with determined parameter values that works well for any stationary processes under weak assumptions. Schmeiser and Song (1987) reported some Monte Carlo results about linearly combining some known (component) estimators with low degrees of freedom and creating a new estimator with more degrees of freedom. Under their assumption that correlations among each pair of some known estimators are independent of the underlying process, the linear combinations provide a better (than the component estimators) estimator for var($\bar{X}$) without estimating parameter values.

The component estimators in Schmeiser and Song (1987) have negligible biases, therefore reducing the variance is the only issue. They consider a vector of p component estimators, $\hat{\underline{V}} = (\hat{V}_1, \hat{V}_2, \cdots, \hat{V}_p)$ with variance-covariance matrix $\Sigma_{\hat{\underline{V}}}$. The linear combination with minimal-variance weights, $\underline{\alpha} = (\Sigma_{\hat{\underline{V}}}^{-1}\underline{1})/(\underline{1}^t\Sigma_{\hat{\underline{V}}}^{-1}\underline{1})$ retains negligible bias since the weight are constrained to sum to one.

If estimators do have non-negligible biases, those optimal-weights yield linear combinations with non-negligible bias, even though the variance is still minimized. In such case, the linear combination may have a mean-squared-error (mse), the sum of variance and squared bias, that is no better than that of some of the component estimators.

In this paper, we further investigate the idea of line combinations by no longer restricting the component estimators to negligible bias. The objective is to form a minimal-mse linear combination by choosing optimal weights. Section 2 is a description of the minimal-mse problem. Section 3 is a numerical study of linear combinations of nonoverlapping-batch-means (NBM), overlapping-batch-means (OBM), standardized-time-series-area (STS.A), and the nonoverlapping-batch-means combined with standardized-time-series-area (NBM+STS.A) estimators. Section 4 is a discussion.

2. MINIMAL-MSE PROBLEMS

Although we are primarily interested in mse, we give optimal weights for the more general risk function $\beta_1 \text{bias}^2 + \beta_2 \text{var}$, where $\beta_1 \geq 0$ and $\beta_2 > 0$. Mse corresponds to $\beta_1 = 1$ and $\beta_2 = 1$. We consider two cases: constrained and unconstrained. The constrained case, in which we require the weights to sum to one, arises naturally when the component estimators are unbiased. The unconstrained case allows any combinations of weights. These two cases are discussed in Section 2.1 and 2.2 respectively.

2.1. Constrained Weights

Let $\underline{\hat{V}} = (\hat{V}_1, \cdots, \hat{V}_p)^t$ be a linearly independent vector of estimators. Let $\Sigma_{\underline{\hat{V}}}$ be the $p \times p$ dispersion matrix of $\underline{\hat{V}}$, and let $\Lambda_{\underline{\hat{V}}}$ be the $p \times p$ matrix with the (i,j) entry $b_i b_j$, where $b_i \equiv \text{bias}(\hat{V}_i)$. Let $\underline{\alpha}$ be a weighting vector containing p real values α_i which is the weight associated with $\hat{V}_i$ used to form a linear combination. The value of $\underline{\alpha}$ then can be chosen to minimize the risk. This problem can be formulated as

$$\text{(P1)} \quad \text{minimize: } \beta_1 \text{bias}^2(\underline{\alpha}^t \underline{\hat{V}}) + \beta_2 \text{var}(\underline{\alpha}^t \underline{\hat{V}})$$

$$\text{subject to} : \underline{\alpha}^t \underline{1} = 1 ,$$

where $\beta_1 \geq 0$, $\beta_2 > 0$, and $\Sigma_{\underline{\hat{V}}}$ has full rank.

Proposition 1: For any covariance-stationary process X, the optimal solution of (P1) is the minimal-risk linear combination with the weights $\underline{\alpha} = (\Delta_{\underline{\hat{V}}}^{-1} \underline{1}) / (\underline{1}^t \Delta_{\underline{\hat{V}}}^{-1} \underline{1})$ and the resulting minimal mse is $(\underline{1}^t \Delta_{\underline{\hat{V}}}^{-1} \underline{1})^{-1}$, where $\Delta_{\underline{\hat{V}}}$ is defined to be $\beta_1 \Lambda_{\underline{\hat{V}}} + \beta_2 \Sigma_{\underline{\hat{V}}}$.

In our application, bias is defined with respect to $\text{var}(\overline{X})$.

We now state three lemmas necessary for proving Proposition 1. They are proven in Appendices A, B, and C, respectively.

Lemma 1: (P1) can be written as

$$\text{(P1}') \quad \text{minimize} : \underline{\alpha}^t \Delta_{\underline{\hat{V}}} \underline{\alpha}$$

$$\text{subject to} : \underline{\alpha}^t \underline{1} = 1 .$$

Lemma 2: $\Delta_{\underline{\hat{V}}}$ in the objective function of (P1′) is positive definite.

Lemma 3: Let $\mathbf{W}$ be a $p \times p$ matrix that can be decomposed into $\mathbf{T}\mathbf{T}^t$, where $\mathbf{T}$ is a nonsingular matrix and let $\underline{\alpha}$ be any $p \times 1$ scalar vector. Then

$$(\underline{\alpha}^t \mathbf{W} \underline{\alpha}) \geq \frac{(\underline{\alpha}^t \underline{1})^2}{\underline{1}^t \mathbf{W}^{-1} \underline{1}} .$$

Proof of Proposition 1: From Lemma 1, problem (P1′) and problem (P1) are equivalent, so we focus on problem (P1′). From Lemma 2 and the Cholesky decomposition, there exists a nonsingular matrix $\mathbf{T}$ such that $\Delta_{\underline{\hat{V}}} = \mathbf{T}\mathbf{T}^t$. Applying the Lemma 3, we have

$$(\underline{\alpha}^t \Delta_{\underline{\hat{V}}} \underline{\alpha}) \geq \frac{(\underline{\alpha}^t \underline{1})^2}{\underline{1}^t \Delta_{\underline{\hat{V}}}^{-1} \underline{1}} = \frac{1}{\underline{1}^t \Delta_{\underline{\hat{V}}}^{-1} \underline{1}}$$

since $\underline{\alpha}^t \underline{1} = 1$. Finally, when

$$\underline{\alpha} = \frac{(\Delta_{\underline{\hat{V}}}^{-1} \underline{1})}{(\underline{1}^t \Delta_{\underline{\hat{V}}}^{-1} \underline{1})} ,$$

$\underline{\alpha}^t \Delta_{\underline{\hat{V}}} \underline{\alpha}$ reaches its lower bound. ∎

When both β_1 and β_2 are 1, problem (P1) becomes the optimal-mse problem

$$\text{(P1.1)} \quad \text{minimize: } \text{mse}(\underline{\alpha}^t \underline{\hat{V}})$$

$$\text{subject to: } \underline{\alpha}^t \underline{1} = 1 .$$

Corollary 1: Problem (P1.1) has optimal weights $\underline{\alpha} = (\Delta_{\underline{\hat{V}}}^{-1} \underline{1}) / (\underline{1}^t \Delta_{\underline{\hat{V}}}^{-1} \underline{1})$ and the resulting minimal mse is $(\underline{1}^t \Delta_{\underline{\hat{V}}}^{-1} \underline{1})^{-1}$, where $\Delta_{\underline{\hat{V}}}$ is defined to be $\Lambda_{\underline{\hat{V}}} + \Sigma_{\underline{\hat{V}}}$.

The special case of (P1.1) with two component estimators is

$$(P1.1.1) \quad \text{minimize: } \text{mse}(\alpha_1 \hat{V}_1 + \alpha_2 \hat{V}_2)$$

$$\text{subject to: } \alpha_1 + \alpha_2 = 1 .$$

Corollary 2: Problem (P1.1.1) has optimal weights

$$\alpha_1 = \frac{(\sigma_2^2 + b_2^2) - (\sigma_{12} + b_1 b_2)}{(\sigma_1^2 + b_1^2) + (\sigma_2^2 + b_2^2) - 2(\sigma_{12} + b_1 b_2)}$$

and $\alpha_2 = 1 - \alpha_1$. The resulting minimal mse is

$$\sigma_2^2 + b_2^2 - \frac{[(\sigma_2^2 + b_2^2) - (\sigma_{12} + b_1 b_2)]^2}{(\sigma_1^2 + b_1^2) + (\sigma_2^2 + b_2^2) - 2(\sigma_{12} + b_1 b_2)},$$

where σ_1^2, σ_2^2, and σ_{12} are variance of $\hat{V}_1$, $\hat{V}_2$, and the covariance of $\hat{V}_1$ and $\hat{V}_2$, respectively; b_1 and b_2 are bias of $\hat{V}_1$ and $\hat{V}_2$, respectively.

Corollary 2 follows from Corollary 1 by writing

$$\Delta_{\hat{\underline{V}}}^{-1} = \frac{\begin{bmatrix} \sigma_2^2 + b_2^2 & -(\sigma_{12} + b_1 b_2) \\ -(\sigma_{12} + b_1 b_2) & \sigma_1^2 + b_1^2 \end{bmatrix}}{(\sigma_1^2 + b_1^2)(\sigma_2^2 + b_2^2) - (\sigma_{12} + b_1 b_2)^2}$$

and

$$\underline{1}^t \Delta_{\hat{\underline{V}}}^{-1} \underline{1} = \frac{\sigma_1^2 + \sigma_2^2 + b_1^2 + b_2^2 - 2\sigma_{12} - 2b_1 b_2}{(\sigma_1^2 + b_1^2)(\sigma_2^2 + b_2^2) - (\sigma_{12} + b_1 b_2)^2}.$$

Alternatively Corollary 2 can be obtained directly by taking first derivatives.

2.2. Unconstrained Weights

In this section, we formulate and solve

$$\text{(P2)} \quad \text{minimize: } \beta_1 \text{bias}^2(\underline{\alpha}^t \hat{\underline{V}}) + \beta_2 \text{var}(\underline{\alpha}^t \hat{\underline{V}}),$$

which is (P1) without the constraint on the weights $\underline{\alpha}$. We still assume that $\Sigma_{\hat{\underline{V}}}$ has full rank. Here the optimal linear combination may be biased even when all of the component estimators are unbiased.

Proposition 2. For any covariance-stationary process X, the optimal solution of (P2) is

$$\underline{\alpha} = \beta_1 \text{var}(\bar{X})(\beta_1 \mathscr{E}_{\hat{\underline{V}}} + \beta_2 \Sigma_{\hat{\underline{V}}})^{-1} \text{E}(\hat{\underline{V}}),$$

where $\text{E}(\hat{\underline{V}})$ is the vector of the component-estimator means and $\mathscr{E}_{\hat{\underline{V}}}$ is the matrix of cross products of means, $\text{E}(\hat{\underline{V}})\text{E}(\hat{\underline{V}})^t$.

Proposition 2 is obtained by taking first derivatives after expressing the bias as the difference of expected value and $\text{var}(\bar{X})$.

When both β_1 and β_2 are 1, problem (P2) becomes the optimal-mse problem

$$\text{(P2.1)} \quad \text{minimize: } \text{mse}(\underline{\alpha}^t \hat{\underline{V}}).$$

Corollary 3: Problem (P2.1) has optimal weights

$$\underline{\alpha} = \text{var}(\bar{X})(\mathscr{E}_{\hat{\underline{V}}} + \Sigma_{\hat{\underline{V}}})^{-1} \text{E}(\hat{\underline{V}}).$$

Restriction to two component estimations lead us to

$$(P2.1.1) \quad \text{minimize: } \text{mse}(\alpha_1 \hat{V}_1 + \alpha_2 \hat{V}_2)$$

and Corollary 4.

Corollary 4. Problem (P2.1.1) has optimal weights

$$\alpha_1 = \frac{\text{var}(\bar{X})\,(e_1 \sigma_2^2 - e_2 \sigma_{12})}{e_1^2 \sigma_2^2 + e_2^2 \sigma_1^2 + \sigma_1^2 \sigma_2^2 - 2 e_1 e_2 \sigma_{12} - \sigma_{12}^2}$$

and

$$\alpha_2 = \frac{\text{var}(\bar{X})\,(e_2 \sigma_1^2 - e_1 \sigma_{12})}{e_1^2 \sigma_2^2 + e_2^2 \sigma_1^2 + \sigma_1^2 \sigma_2^2 - 2 e_1 e_2 \sigma_{12} - \sigma_{12}^2},$$

where e_1 and e_2 are the expected values of $\hat{V}_1$ and $\hat{V}_2$, respectively.

3. NUMERICAL STUDY

The objective of the numerical study in this section is to see how much the use of linear combinations can improve (reduce) the mse and to study how the optimal component estimators behave as a function of estimator type, batch size, and statistical properties (bias, variance, and mse).

Specifics of the study are as follows. The component estimators are NBM, OBM, STS.A, and NBM+STS.A. The sample size n is 30. The data process is steady-state first-order autoregressive (AR(1)) with the parameter $\phi_1 = 0.8082$. This value of ϕ_1 corresponds to $\gamma_0 \equiv \sum_{h=-\infty}^{\infty} \rho_h = 10$, where ρ_h denotes the lag-h autocorrelation corr (X_i, X_{i+h}). The arbitrarily value 1 is used for $\text{var}(\bar{X})$. The biases, variances, and mse's are calculated numerically, using the approach discussed in Song and Schmeiser (1988a).

Before considering linear combinations in Section 3.2, we investigate the optimal mse for individual component estimators in Section 3.1.

3.1. Optimal Mse: Component Estimators

For the four component estimator types, we numerically compute their optimal batch sizes and optimal mse's, as well as associated biases and variances. These statistical properties are reported here primarily for comparison to the statistical properties of the linear combinations studied in Section 3.2. Goldsman and Meketon (1986) and Song and Schmeiser (1988b) contain a more complete study of the four component estimator types for this process.

Let m denote the batch size and $\hat{V}(m)$ denote the estimator type of NBM, OBM, STS.A, and NBM+STS.A with batch size m. Let m^* denote the mse-optimal batch size; i.e., mse($\hat{V}(m^*)$) $\leq$ mse($\hat{V}(m)$) for all batch sizes m. The numerically determined optimal batch size, squared bias, variance, and mse for $\hat{V}(m^*)$ of the four estimator types are shown in Table 1.

Table 1: Optimal Batch Size and Their Properties

Property	Estimator Type			
	NBM	OBM	STS.A	NBM+STS.A
m^*	7	8	15	15
$\text{bias}^2(\hat{V}(m^*))\times 10^2$	33.6	33.6	39.3	28.9
$\text{var}(\hat{V}(m^*))\times 10^2$	13.7	14.1	15.0	16.6
$\text{mse}(\hat{V}(m^*))\times 10^2$	47.3	47.7	54.3	45.5

The order of the estimator types in increasing values of mse($\hat{V}(m^*)$) shown in the last row, is NBM+STS.A, NBM, OBM, and STS.A, which differs from the results in Song and Schmeiser (1988b) where the order is OBM, NBM, NBM+STS.A, and STS.A for the same process and $n = 500$. We emphasize this difference to discourage the reader from making type comparisons based on the insufficient information in this study. Our only current purpose is to demonstrate mse-reduction with linear combinations.

3.2. Optimal Mse: Linear Combinations

In this section we numerically demonstrate the mse-reduction obtained with linear combinations of two component estimators for the sample size and process described in Section 3.1. The minimal-mse problems considered are (P1.1.1), weights constrained to sum to one, and (P2.1.1), unconstrained weights. The numerical results for these two problems are discussed in Section 3.2.1 and Section 3.2.2, respectively.

In each linear combination, let $\hat{V}_1(m_1)$ and $\hat{V}_2(m_2)$ denote the two component estimators with batch size m_1 and m_2, respectively. Let $\hat{V}^* \equiv \alpha_1{}^* \hat{V}_1(m_1{}^*) + \alpha_2{}^* \hat{V}_2(m_2{}^*)$ denote the mse-optimal linear combination of $\hat{V}_1(m_1)$ and $\hat{V}_2(m_2)$ with corresponding weights $\alpha_1{}^*$ and $\alpha_2{}^*$ and corresponding optimal batch sizes $m_1{}^*$ and $m_2{}^*$. The weights and batch sizes are simultaneously optimal in the sense that for any values of α_1, α_2, m_1, and m_2

$$\text{mse}(\hat{V}^*) \equiv \text{mse}(\alpha_1{}^* \hat{V}_1(m_1{}^*) + \alpha_2{}^* \hat{V}_2(m_2{}^*))$$
$$\leq \text{mse}(\alpha_1 \hat{V}_1(m_1) + \alpha_2 \hat{V}_2(m_2)).$$

The values of squared bias, variance, covariance, and mse in Table 2 through Table 4 are multiplied by 10^2 because some values are very small.

3.2.1. Constrained Weights Summing to One.

Consider the minimal-mse problem (P1.1.1) where the optimal weights sum to one. In this section we use α and $1-\alpha$ rather than α_1 and α_2, since the constraint makes this an inherently one-weight problem. The component estimators to be combined can be of common type or of different types. The results of combining common types are shown in Table 2 and different types in Table 3.

In Table 2 we combine common types of estimators (with different batch sizes), allowing us to temporarily drop the subscript on $\hat{V}$.

Table 2: Optimal Linear Combinations:
Common Types of Component Estimators.
Optimal Weights Summing to One.

Property	Estimator Type			
	NBM	OBM	STS.A	NBM+STS.A
$(m_1{}^*, m_2{}^*)$	(1, 2)	(1, 2)	(2, 5)	(2, 3)
$\text{bias}^2(\hat{V}(m_1{}^*))\times 10^2$	82.9	82.9	95.7	82.9
$\text{bias}^2(\hat{V}(m_2{}^*))\times 10^2$	70.3	70.5	83.4	76.0
$\text{var}(\hat{V}(m_1{}^*))\times 10^2$	0.227	0.227	0.00643	0.227
$\text{var}(\hat{V}(m_2{}^*))\times 10^2$	0.940	0.935	0.272	0.525
$\text{cov}(\hat{V}(m_1{}^*), \hat{V}(m_2{}^*))\times 10^2$	0.460	0.458	0.0112	0.345
$\text{corr}(\hat{V}(m_1{}^*), \hat{V}(m_2{}^*))$	0.996	0.996	0.268	0.999
α^*	-7.25	-7.33	-8.37	-15.0
$1-\alpha^*$	8.25	8.33	9.37	16.0
$\text{bias}^2(\hat{V}^*)\times 10^2$	9.80	10.3	13.7	8.45
$\text{var}(\hat{V}^*)\times 10^2$	20.9	21.1	22.6	20.1
$\text{mse}(\hat{V}^*)\times 10^2$	30.7	31.4	36.2	28.6
$1 - \frac{\text{mse}(\hat{V}^*)}{\text{mse}(\hat{V}(m^*))}$	35%	34%	33%	37%

Now observe the individual component estimators that comprise the optimal linear combinations in Table 2. The optimal pairs of batch sizes $m_1{}^*$ and $m_2{}^*$ have relative small values, as shown in the first row. Such batch sizes correspond to large biases, as shown in the second and third rows; and small variances, as shown in the fourth and fifth rows.

Now observe the relationships among the component estimators that comprise the optimal linear combinations. The covariances and correlations of the two optimal component estimators, shown in the sixth and seventh rows, are positive. Except for the STS.A estimator, the correlations are close to one.

The optimal weights α^* and $1-\alpha^*$ have different signs, which both partially cancels the large biases of the two component estimators and uses the positive correlation to obtain small variance. These effects can be seen in the formulas for the bias and variance of the linear combination, which are

$$\text{bias}(\hat{V}^*) = \alpha^* \,\text{bias}(\hat{V}_1{}^*) + (1-\alpha^*)\,\text{bias}(\hat{V}_2{}^*)$$

and

$$\text{var}(\hat{V}^*) = (\alpha^*)^2 \text{var}(\hat{V}_1{}^*) + (1-\alpha^*)^2 \text{var}(\hat{V}_2{}^*) + 2\alpha^*(1-\alpha^*)\,\text{cov}(\hat{V}_1{}^*, \hat{V}_2{}^*).$$

Now observe the mean squared error. Comparing mse($\hat{V}^*$) to mse($\hat{V}(m^*)$), which is obtained in Section 3.1, we see the mse-reduction is about 35% for any of the four estimator types. This consistency across types of estimators is reminiscent of Schmeiser and Song (1987), where the linear combinations considered resulted in variance decreases that were similar across types of estimators.

To provide some insight into the nature of the performance of the optimal linear combination compared to that of component estimators, consider Figure 1, which shows the statistical properties (bias, variance, and mse) of the NBM estimators corresponding to various batch sizes and of the optimal linear combination. The small variances and large biases associated with batch sizes $m_1{}^* = 1$ and $m_2{}^* = 2$, which are used in the optimal linear combination, are shown on the left side of the figure. The squared bias, variance, and mse of the optimal linear combination are shown as horizontal lines through the figure. We see that in addition to the 35% mse reduction, the bias is less than the bias of NBM with any batch size and the variance is less than NBM variances for batch sizes greater than seven. Although not shown, the analogous figures for OBM, STS.A, and NBM+STS.A show similar results.

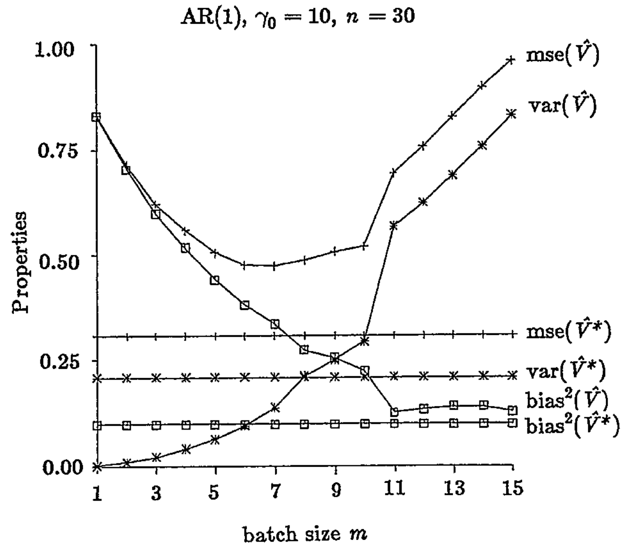

Figure 1: Statistical Properties of the NBM Estimators and of the Optimal Linear Combination

Table 3 is similar to Table 2, except that it contains the results for linear combinations composed of two different types of component estimators. The second, third and fourth columns of Table 3 show results for the linear combinations of (NBM, STS.A), (OBM, STS.A) and (NBM, OBM), respectively. Again, we see that the batch sizes in the optimal linear combination are small, with corresponding large biases and small variances. The covariances and correlations of the two component estimators are positive. The optimal weights α^* and $1-\alpha^*$ have different signs, again allowing the biases to partially cancel and the variance to be small. The mse-reduction is about 30% compared to mse($\hat{V}_{NBM+STS.A}(m^*)$) from Table 2. We use $\hat{V}_{NBM+STS.A}(m^*)$ as the basis to compute the mse reduction because it has the smallest mse in the last row of Table 1.

3.2.2. Unconstrained Weights. Consider the minimal-mse problem (P2.1.1). Since the optimal weights are unrestricted, the mse($\hat{V}^*$) must be smaller than or equal to the optimal mse's obtained in Section 3.2.1.

Using the same process and sample size, we show the optimal linear combination for common component types in Table 4.

The batch sizes $m_1{}^*$ and $m_2{}^*$ shown in the first row are again small. These batch sizes still correspond to large biases and small variances.

Table 3: Optimal Linear Combinations: Different Component Estimators Types. Optimal Weights Summing to One.			
Property	Estimator Types ($\hat{V}_1$, $\hat{V}_2$)		
	(NBM,STS.A)	(OBM,STS.A)	(NBM,OBM)
(m_1*, m_2*)	(2, 2)	(1, 2)	(1, 2)
bias^2($\hat{V}_1(m_1^*)$)x10^2	70.2	82.9	82.9
bias^2($\hat{V}_2(m_2^*)$)x10^2	95.7	95.7	70.5
var($\hat{V}_1(m_1^*)$)x10^2	0.940	0.227	0.227
var($\hat{V}_2(m_2^*)$)x10^2	0.00643	0.00643	0.935
cov($\hat{V}_1(m_1^*), \hat{V}_2(m_2^*)$)x10^2	0.0114	0.00882	0.458
corr($\hat{V}_1(m_1^*), \hat{V}_2(m_2^*)$)	0.147	0.231	0.996
α^*	4.75	9.83	-7.33
$1-\alpha^*$	-3.75	-8.83	8.33
bias^2($\hat{V}^*$)x10^2	9.83	9.87	10.3
var($\hat{V}^*$)x10^2	20.9	20.9	21.1
mse($\hat{V}^*$)x10^2	30.7	30.8	31.4
$1-\frac{\text{mse}(\hat{V}^*)}{\text{mse}(\hat{V}_{NBM+STS.A}(m^*))}$	33%	32%	31%

Table 4: Optimal Linear Combinations: Common Component Estimators Types. Unconstrained Weights.				
Property	Estimator Type			
	NBM	OBM	STS.A	NBM+STS.A
(m_1*, m_2*)	(1, 3)	(1, 2)	(2, 3)	(2, 5)
bias^2($\hat{V}(m_1^*)$)x10^2	82.9	82.9	95.7	82.9
bias^2($\hat{V}(m_2^*)$)x10^2	59.9	70.5	92.2	58.8
var($\hat{V}(m_1^*)$)x10^2	0.227	0.227	0.00643	0.227
var($\hat{V}(m_2^*)$)x10^2	2.18	0.935	0.0327	1.54
cov($\hat{V}(m_1^*), \hat{V}(m_2^*)$)x10^2	0.696	0.458	0.00568	0.588
corr($\hat{V}(m_1^*), \hat{V}(m_2^*)$)	0.990	0.996	0.392	0.995
α_1*	39.2	56.2	27.8	51.3
α_2*	-11.5	-25.8	7.44	-18.4
bias^2($\hat{V}^*$)x10^2	0.980	1.06	1.03	0.827
var($\hat{V}^*$)x10^2	8.92	9.25	9.11	8.27
mse($\hat{V}^*$)x10^2	9.90	10.3	10.1	9.09
$1-\frac{\text{mse}(\hat{V}^*)}{\text{mse}(\hat{V}(m^*))}$	79%	78%	81%	80%

The optimal weights α_1* and α_2* do not sum to one. For the first time, we see an example where the signs are both positive — STS.A. The bias of the optimal linear combination is

$$\text{bias}(\hat{V}^*) = \alpha_1 \,\text{bias}(\hat{V}_1) + \alpha_2 \,\text{bias}(\hat{V}_2) + (\alpha_1+\alpha_2-1)\,\text{var}(\bar{X}) .$$

Because of the last term, the signs of the optimal weights need not be different for the biases to partially offset each other.

The bias^2, var, and mse shown in the fourth-to-last, third-to-last, and next-to-last row, respectively, are much smaller than the corresponding values in Table 2. The mse reduction is about 80%, which is far larger than 35% in the last row of Table 2 and the 30% in the last row of Table 3.

4. DISCUSSION

We have given formulas for mse-optimal linear combinations and numerically studied estimator performance for a particular AR(1) process. The formulas, which are functions of the biases, variances, and covariances of the component estimators, consider the constrained and unconstrained cases. The numerical study demonstrated mse reductions of about 30% for all constrained linear combinations and about 80% for all unconstrained linear combinations.

The various estimator types all lead to similar optimal mse's. The tentative conclusion is that the choice of component estimator type is relatively unimportant.

In practice, the variance of the sample mean, the bias vector, and the covariance matrix are unknown. Estimation of the batch sizes and optimal weights, which is not studied here, may be difficult. We do not yet know the relative difficulties of estimating batch sizes for component estimators, single weights for constrained linear combinations, and double weights for unconstrained linear combinations. Therefore, the extent to which we can benefit from the mse reductions demonstrated here is unknown.

Some results necessary for estimating properties of component estimators are available. Schmeiser and Song (1987) discuss correlations among estimators with large batch size and hypothesize that the asymptotic correlations are not a function of the process and are good approximations for even small sample sizes. Goldsman and Meketon (1986) and Song and Schmeiser (1988b) discuss optimal batch sizes for component estimators. The component estimator S^2/n; which is NBM with batch size one, OBM with batch size one, or NBM+STS.A with batch size 2; arises repeatedly in our numerical results and has easily computed bias (e.g., David (1985)). We know of no other component estimator with known bias.

Keep in mind that the numerical results are for a single sample size and process. The sample size is quite small; $n = 30$ is only three times the sum of the correlations $\gamma_0 = 10$. However, the example reported is consistent with a limited amount of other experience, so we think the example is not misleading. The small sample size was used because (1) it allowed numerical (rather than Monte Carlo) mse calculations and (2) our experience indicates that even such a small sample size yield results consistent with large sample sizes. Nevertheless, results should be extrapolated only with caution. All that can be claimed with certainty is that this one example has demonstrably large potential mse reduction.

ACKNOWLEDGMENT

This material is based upon work supported by the National Science Foundation under Grant No. DMS-8717799. The Government has certain rights in this material.

APPENDIX A: PROOF OF LEMMA 1

Define bias($\hat{\underline{V}}$) to be the bias vector, $\underline{b} = (b_1, \cdots, b_p)^t$, where $b_i \equiv \text{bias}(\hat{V}_i) \equiv \text{E}(\hat{V}_i) - \text{var}(\overline{X})$. Then

$$\begin{aligned}\text{bias}(\underline{\alpha}^t \hat{\underline{V}}) &= \underline{\alpha}^t \underline{b} + \text{var}(\overline{X})(\underline{\alpha}^t \underline{1} - 1)\\ &= \underline{\alpha}^t \underline{b}\end{aligned}$$

since $\underline{\alpha}^t \underline{1} = 1$. Let $\Lambda_{\hat{\underline{V}}} \equiv \underline{b}\,\underline{b}^t$. Then

$$\begin{aligned}\beta_1 \text{bias}^2(\underline{\alpha}^t \hat{\underline{V}}) &= \beta_1(\underline{\alpha}^t \underline{b}\,\underline{b}^t \underline{\alpha})\\ &= \beta_1(\underline{\alpha}^t \Lambda_{\hat{\underline{V}}} \underline{\alpha})\\ &= \underline{\alpha}^t (\beta_1 \Lambda_{\hat{\underline{V}}}) \underline{\alpha}.\end{aligned}$$

Also

$$\begin{aligned}\beta_2 \text{var}(\underline{\alpha}^t \hat{\underline{V}}) &= \beta_2(\underline{\alpha}^t \Sigma_{\hat{\underline{V}}} \underline{\alpha})\\ &= \underline{\alpha}^t (\beta_2 \Sigma_{\hat{\underline{V}}}) \underline{\alpha}.\end{aligned}$$

Therefore

$$\begin{aligned}\beta_1 \text{bias}^2(\underline{\alpha}^t \hat{\underline{V}}) + \beta_2 \text{var}(\underline{\alpha}^t \hat{\underline{V}}) &= \underline{\alpha}^t [\beta_1 \Lambda_{\hat{\underline{V}}} + \beta_2 \Sigma_{\hat{\underline{V}}}] \underline{\alpha}\\ &= \underline{\alpha}^t \Delta_{\hat{\underline{V}}} \underline{\alpha}. \blacksquare\end{aligned}$$

APPENDIX B: PROOF OF LEMMA 2

We show that the objective function of (P1′), $\Delta_{\hat{\underline{V}}} \equiv \beta_1 \Lambda_{\hat{\underline{V}}} + \beta_2 \Sigma_{\hat{\underline{V}}}$, is positive definite, where $\beta_1 \geq 0$ and $\beta_2 > 0$.

By definition, $\Lambda_{\hat{\underline{V}}} \equiv \underline{b}\,\underline{b}^t$, where $\underline{b} = [b_1, b_2, \cdots, b_p]^t$, and

$$\Sigma_{\hat{\underline{V}}} \equiv \text{E}\left\{[\hat{\underline{V}} - \text{E}(\hat{\underline{V}})][\hat{\underline{V}} - \text{E}(\hat{\underline{V}})]^t\right\}.$$

For any nonzero real vector $\underline{s}$,

$$\begin{aligned}\underline{s}^t \Delta_{\hat{\underline{V}}} \underline{s} &= \underline{s}^t (\beta_1 \Lambda_{\hat{\underline{V}}} + \beta_2 \Sigma_{\hat{\underline{V}}}) \underline{s}\\ &= \underline{s}^t (\beta_1 \Lambda_{\hat{\underline{V}}}) \underline{s} + \underline{s}^t (\beta_2 \Sigma_{\hat{\underline{V}}}) \underline{s}\\ &= \beta_1 (\underline{s}^t \underline{b}\,\underline{b}^t \underline{s}) + \beta_2 \text{E}\left(\underline{s}^t [\hat{\underline{V}} - \text{E}(\hat{\underline{V}})][\hat{\underline{V}} - \text{E}(\hat{\underline{V}})]^t \underline{s}\right)\\ &= \beta_1 (\underline{s}^t \underline{b})^2 + \beta_2 \text{E}\left([\underline{s}^t(\hat{\underline{V}} - \text{E}(\hat{\underline{V}}))]^2\right) > 0,\end{aligned}$$

since $\beta_1 \geq 0$; $(\underline{s}^t \underline{b})^2 \geq 0$; $\beta_2 > 0$; and since $\Sigma_{\hat{\underline{V}}}$ has full rank, which implies that $\underline{s}^t(\hat{\underline{V}} - \text{E}(\hat{\underline{V}})) \neq 0$, therefore, $\text{E}\left([\underline{s}^t(\hat{\underline{V}} - \text{E}(\hat{\underline{V}}))]^2\right) > 0$. So, $\Delta_{\hat{\underline{V}}}$ is positive definite. $\blacksquare$

APPENDIX C: PROOF OF LEMMA 3

$$\begin{aligned}(\underline{\alpha}^t \underline{1})^2 &= (\underline{\alpha}^t \mathbf{I} \underline{1})^2\\ &= (\underline{\alpha}^t (\mathbf{T}\mathbf{T}^{-1}) \underline{1})^2\\ &= [(\underline{\alpha}^t \mathbf{T})(\mathbf{T}^{-1} \underline{1})]^2\\ &\leq (\underline{\alpha}^t \mathbf{T})(\underline{\alpha}^t \mathbf{T})^t (\mathbf{T}^{-1} \underline{1})^t (\mathbf{T}^{-1} \underline{1}) \quad \text{(A.C.1)}\\ &= (\underline{\alpha}^t \mathbf{T}\mathbf{T}^t \underline{\alpha})(\underline{1}^t ((\mathbf{T}^{-1})^t \mathbf{T}^{-1}) \underline{1})\\ &= (\underline{\alpha}^t \mathbf{W} \underline{\alpha})(\underline{1}^t \mathbf{W}^{-1} \underline{1}) \quad \text{(A.C.2)}\end{aligned}$$

Inequality (A.C.1) follows by applying the Cauchy-Schwartz Inequality $(\underline{X}^t \underline{Y})^2 \leq (\underline{X}^t \underline{X})(\underline{Y}^t \underline{Y})$. From (A.C.2) we have

$$(\underline{\alpha}^t \underline{1})^2 \leq (\underline{\alpha}^t \mathbf{W} \underline{\alpha})(\underline{1}^t \mathbf{W}^{-1} \underline{1}),$$

which implies

$$(\underline{\alpha}^t \mathbf{W} \underline{\alpha}) \geq \frac{(\underline{\alpha}^t \underline{1})^2}{\underline{1}^t \mathbf{W}^{-1} \underline{1}}. \blacksquare$$

REFERENCES

Bratley, P., Fox, B.L. and Schrage, L. (1987). *A Guide to Simulation.* Springer-Verlag.

Crane, M.A. and Iglehart, D.L. (1975). Simulating stable stochastic systems, III: regenerative process and discrete-event simulations. *Operations Research* **23**, 33-45.

David, H.A. (1985). Bias of S^2 under dependence. *American Statistician* **39**, 201.

Fishman, G.S. (1978). Grouping observations in digital simulations. *Management Science* **24**, 510-521.

Goldsman, D. (1984). On using standardized time series to analyze stochastic processes. Ph.D. Dissertation, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, New York.

Goldsman, D. and Meketon, M. (1986). A comparison of several variance estimators. Technical Report J-85-12, Department of Operations Research, AT&T Bell Laboratories, Holmdel, NJ 07733.

Heidelberger, P. and Welch, P.D. (1981). A spectral method for confidence interval generation and run length control in simulation. *Communications of the ACM* **24**, 233-245.

Law, A.M. and Kelton, W.D. (1983). *Simulation Modeling and Analysis.* McGraw-Hill.

Meketon, M.S. and Schmeiser, B. (1984). Overlapping batch means: something for nothing? In: *Proceedings of the 1984 Winter Simulation Conference* (S. Sheppard, U. Pooch, and D. Pegden, eds.), 227-230.

Schmeiser, B. (1982). Batch-size effects in the analysis of simulation output. *Operations Research* **30**, 556-568.

Schmeiser, B. and Song, W.-M.T. (1987). Correlation among estimators of the variance of the sample mean. In: *Proceedings of the 1987 Winter Simulation Conference* (A. Thesen, H. Grant, W. David Kelton, eds.), 309-316.

Schriber, T.J. and Andrews, R.W. (1984). ARMA-based confidence interval procedures for simulation output analysis. *American Journal of Mathematical and Management Science* **4**, 345-373.

Schruben, L.W. (1983). Confidence interval estimation using standardized time series. *Operations Research* **31**, 1090-1108.

Song, W.-M.T. and Schmeiser, B. (1988a). Estimating variance of the sample mean: quadratic-forms and cross-product covariances. *Operations Research Letters* **7** (1988a), in press.

Song, W.-M.T. and Schmeiser, B. (1988b). Estimating standard errors: Empirical behavior of asymptotic mse-optimal batch sizes. In: *20th Symposium on the Interface: Computing Science and Statistics,* forthcoming.

AUTHOR'S BIOGRAPHIES

Wheyming Tina Song is a Ph.D. candidate in the School of Industrial Engineering at Purdue University, majoring in simulation and applied probability. She received her undergraduate degree in statistics and master's degree in industrial management at Cheng-Kung University in Taiwan. She then received her master's degrees in applied mathematics and industrial engineering from the University of Pittsburgh.

Wheyming Tina Song
School of Industrial Engineering, Purdue Univerisity
West Lafayette, IN 47907, U.S.A.
(317) 743-8275
wheyming@gb.ecn.purdue.edu (arpanet)

Bruce Schmeiser is a professor in the School of Industrial Engineering at Purdue University. He received his undergraduate degree in mathematical sciences and master's degree in industrial and management engineering at The University of Iowa. His Ph.D. is from the School of Industrial and Systems Engineering at the Georgia Institute of Technology. He is the *Operations Research* area editor in simulation and has served in editorial positions of *IIE Transactions, Communications in Statistics, B: Simulation and Computation, Journal of Quality Technology, American Journal of Mathematical and Management Sciences,* and the *Handbook of Industrial Engineering.* He is the past chairman of the TIMS College on Simulation and Gaming. He represents ORSA on the Winter Simulation Conference Board of Directors, currently serving as chairman. His research interests are the probabilistic and statistical aspects of digital-computer stochastic simulation, including input modeling, random-variate generation, output analysis, and variance reduction.

Bruce Schmeiser
School of Industrial Engineering
Purdue University
West Lafayette, IN 47907, U.S.A.
(317) 494-5422
schmeise@pink.ecn.purdue.edu (arpanet)