Proceedings of the 1993 Winter Simulation Conference
G.W. Evans, M. Mollaghasemi, E.C. Russell, W.E. Biles (eds.)

WEIGHTED BATCH MEANS AND IMPROVEMENTS IN COVERAGE

Diane P. Bischak

Department of Business Administration
University of Alaska
Fairbanks, Alaska 99775, U.S.A.

ABSTRACT

Weighted batch means is a procedure for producing a confidence interval for the mean of a covariance-stationary process. Weights placed on the observations within a batch are functions of the parameters of a fitted time-series model. Experiments show that the method works well in terms of achieved coverage when only a comparatively small number of observations is available, even for processes that display strong correlation. In theory the method should provide exact coverage for some processes. However, in practice the time-series identification procedure and estimation of the parameters and weights bring in bias. We investigate the sources of bias and suggest how coverage might be improved.

1 INTRODUCTION

The problem of constructing a confidence interval for the mean μ of a discrete-time, covariance-stationary stochastic process $\{X_i,\ i = 1,2,\ldots\}$ from a sample of n data values has received much attention in the simulation literature. The main reason for this attention is that it is difficult, in general, to produce a valid interval. That is, it is difficult to find a procedure that, when used to produce a nominal level $1-\alpha$ confidence interval in a very large number of experiments, will produce intervals that actually cover μ on $100(1-\alpha)\%$ of those experiments.

Validity is one of the most important attributes for a confidence-interval construction procedure to possess. Unfortunately, straightforward application of confidence-interval construction procedures that assume the observed data are independent and identically distributed (i.i.d.) does not provide validity when the data are autocorrelated. The problem lies in the fact that the usual estimator of the variance of the sample average $\overline{X} \equiv \sum_{i=1}^{n} X_i/n$,

$$\widehat{\mathrm{Var}}\left(\overline{X}\right) \equiv \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n(n-1)},$$

is biased when the data are correlated. If the process is positively autocorrelated, as is often the case in simulations of queuing systems, $\widehat{\mathrm{Var}}\left(\overline{X}\right)$ is biased low and confidence intervals created using it will, on average, have lower-than-nominal coverage.

Weighted batch means (Bischak, Kelton, and Pollock 1993) is a generalization of the method of confidence-interval construction known as batch means. Here we describe the method of (unweighted) batch means and our generalization, in which weights for observations are selected according to formulas for optimal weights developed for certain autoregressive and moving-average processes. Empirical results show that weighted batch means (WBM) provides coverage closer to nominal confidence than that of unweighted batch means (UBM) in the case of high autocorrelation and small n. In theory the method should provide valid intervals for certain processes; we examine why this is not the case in practice.

2 WEIGHTED BATCH MEANS

In the method of batch means (see Law and Kelton 1991) the n observations are grouped into k batches of m consecutive observations each. (We assume that k divides n, so that $n = km$.) The jth batch mean $\overline{X}_j \equiv \sum_{i=1}^{m} X_{ij}/m$, where X_{ij} denotes the ith observation in the jth batch, is then treated as a single "observation," and the confidence interval constructed from the k batch means is

$$\overline{\overline{X}} \pm t_{1-\alpha/2,k-1}\sqrt{\widehat{\mathrm{Var}}\left(\overline{\overline{X}}\right)}, \tag{1}$$

where

$$\overline{\overline{X}} \equiv \sum_{j=1}^{k} \overline{X}_j/k = \sum_{i=1}^{n} X_i/n$$

is the sample average of the batch means,

$$\widehat{\mathrm{Var}}\left(\overline{\overline{X}}\right) \equiv \frac{\sum_{j=1}^{k}(\overline{X}_j - \overline{\overline{X}})^2}{k(k-1)}$$

is its variance estimator, and $t_{1-\alpha/2,n-1}$ is the $100(1-\alpha/2)$th percentile of the t distribution with $n-1$ degrees of freedom.

Batch means formed from "large" batches will tend to be uncorrelated and normal and hence independent, making (1) a valid interval. For small samples, however, (1) will not be valid because of the remaining correlation of the $\overline{X}_j$'s. Positive process autocorrelation will bias the estimate of $\mathrm{Var}\left(\overline{\overline{X}}\right)$ downward; negative autocorrelation will result in overestimation of the point estimator variance.

If we use weights on the observations within each batch, the jth *weighted batch mean* is

$$\overline{Y}_j \equiv \sum_{i=1}^{m} w_i X_{ij},$$

where the weights w_i are constants. The resulting confidence interval is

$$\overline{\overline{Y}} \pm t_{1-\alpha/2,k-1}\sqrt{\widehat{\mathrm{Var}}\left(\overline{\overline{Y}}\right)}, \tag{2}$$

where $\overline{\overline{Y}} \equiv \sum_{j=1}^{k} \overline{Y}_j / k$ and

$$\widehat{\mathrm{Var}}\left(\overline{\overline{Y}}\right) \equiv \frac{\sum_{j=1}^{k}(\overline{Y}_j - \overline{\overline{Y}})^2}{k(k-1)}.$$

The weights to be used in WBM are determined by solving the following program (for details see Bischak, Kelton, and Pollock 1993):

$$(P1): \quad \min_{\mathbf{w}} \quad \mathrm{Var}\left(\overline{\overline{Y}}\right)$$
$$\text{s.t.} \quad \gamma_{\overline{Y}}(j) = 0, \qquad j = 1, \ldots, k-1$$
$$\sum_{i=1}^{m} w_i = 1,$$

where

$$\gamma_{\overline{Y}}(j) \equiv \sum_{p=1}^{m}\sum_{q=1}^{m} w_p w_q \gamma_X(jm + p - q)$$

is the weighted-batch-mean covariance at lag j with $\gamma_X(j) = \mathrm{Cov}(X_t, X_{t\pm j})$,

$$\mathrm{Var}\left(\overline{\overline{Y}}\right) = \sum_{j=1}^{k}\sum_{l=1}^{k} \gamma_{\overline{Y}}(j-l)/k^2,$$

and $\mathbf{w} = (w_1, w_2, \cdots, w_m)^{\mathrm{T}}$ is the vector of weights. Note that the solution weights may be negative and that for each vector of weights that solves $(P1)$, we can reverse the indices of the weights and get another solution.

Weights solving this program will provide an unbiased point estimator $\overline{\overline{Y}}$, will provide an unbiased variance estimator $\widehat{\mathrm{Var}}\left(\overline{\overline{Y}}\right)$, and will minimize the variance of the point estimator under the given conditions. The solution weights will not necessarily provide a valid confidence interval for μ, but if weights fulfilling these conditions are used and the resulting weighted batch means are normally distributed, they are also independent and (2) is valid.

In $(P1)$ the covariances $\gamma_X(l)$ for $l = 0, \ldots, n-1$ must be estimated. We use autoregressive-moving-average (ARMA) models to model the process, the advantage being that it is unnecessary to estimate explicitly all n covariances. Instead, we estimate a few parameters of which the covariances are functions. Although this restricts the applicability of the method somewhat, it is well known that many processes occurring in practice can be treated as though they were ARMA processes. For processes that display high autocorrelation that dies out slowly over many lags, the autocorrelation structure can be well represented by autoregressive processes.

The problem $(P1)$ has an analytic solution for certain ARMA processes. We briefly state results on AR(p) processes; details on these and on the optimal weights for the MA(1) process can be found at length in Bischak, Kelton, and Pollock (1993). When the parameter ϕ is known, there are exactly two optimal solutions to $(P1)$ for the AR(1) process

$$X_t = \mu + \phi(X_{t-1} - \mu) + \varepsilon_t, \tag{3}$$

where $|\phi| < 1$ and the errors $\{\varepsilon_t\}$ are i.i.d. $N\left(0, \sigma^2\right)$, $0 < \sigma^2 < \infty$. These solutions are

$$\mathbf{u} = \left(\frac{-\phi}{(m-1)(1-\phi)}, \frac{1}{m-1}, \cdots, \cdots, \frac{1}{m-1}, \frac{1}{(m-1)(1-\phi)}\right)^{\mathrm{T}}$$

and

$$\mathbf{v} = \left(\frac{1}{(m-1)(1-\phi)}, \frac{1}{m-1}, \cdots, \cdots, \frac{1}{m-1}, \frac{-\phi}{(m-1)(1-\phi)}\right)^{\mathrm{T}}.$$

The weights $\mathbf{u}$ and $\mathbf{v}$ are optimal for a given batch size m. Additional calculations show that for a number of batches $10 \le k \le 25$ the half-length is fairly insensitive to the choice of k.

If we rewrite (3) as

$$\varepsilon_t = X_t - \phi X_{t-1}$$

and take the sum of $m-1$ of these terms divided by the sum of the coefficients of the X's, we obtain

$$\frac{\sum_{i=t}^{t+m-2}\varepsilon_i}{(1-\phi)(m-1)} = \frac{1}{(1-\phi)(m-1)}X_{t+m-2} + \sum_{i=t}^{t+m-3}\frac{1}{m-1}X_i - \frac{\phi}{(1-\phi)(m-1)}X_{t-1}. \quad (4)$$

The right-hand side of (4) is precisely a batch mean weighted by $\mathbf{u}$. Since the ε_i's are normal i.i.d. random variables with variance σ^2, the distribution of $\overline{Y}$ using $\mathbf{u}$ (or $\mathbf{v}$) is normal with mean μ and variance $\sigma^2/[(1-\phi)^2(m-1)]$. The confidence interval (2) is therefore based on normal, uncorrelated, and hence i.i.d. data and will be valid for batches of any size $m \geq 2$.

It can be shown that for the AR(1) process the ratio of $\mathrm{Var}\left(\overline{\overline{X}}\right)$ and $\mathrm{Var}\left(\overline{\overline{Y}}\right)$ (using the optimal weights) is

$$\frac{\mathrm{Var}\left(\overline{\overline{Y}}\right)}{\mathrm{Var}\left(\overline{\overline{X}}\right)} = \frac{n^2\left(1-\phi^2\right)}{n^2\left(1-\phi^2\right) - nk\left(1-\phi^2\right) - 2\phi(n-k)\left(1-\phi^n\right)},$$

which is greater than one if $\phi > 0$. Hence weighting the batch means in this manner increases the variance of the point estimator, which increases the half-length of the confidence interval over that obtained with unweighted batch means.

For the AR(p) process

$$X_t = \mu + \phi_1(X_{t-1}-\mu) + \cdots + \phi_p(X_{t-p}-\mu) + \varepsilon_t,$$

where $\sum_{i=1}^{p}\phi_i < 1$, we can obtain weights from sums of AR(p) error terms, as in (4), which provide a feasible solution to ($P1$) and a local minimum; from extensive experimentation we postulate that they also form one of the optimal solutions, as was true in the AR(1) case. The derived weights are, for $0 \leq p \leq m/2$,

$$w_i = \begin{cases} -\sum_{j=p-i+1}^{p}\phi_j/C_{m,p} & \text{for } i = 1,\ldots,p, \\ \left(1-\sum_{j=1}^{p}\phi_j\right)/C_{m,p} & \text{for } i = p+1,\ldots,m-p, \\ \left(1-\sum_{j=1}^{m-i}\phi_j\right)/C_{m,p} & \text{for } i = m-p+1,\ldots,m, \end{cases}$$

and, for $m/2 < p < m$,

$$w_i = \begin{cases} -\sum_{j=p-i+1}^{p}\phi_j/C_{m,p} & \text{for } i = 1,\ldots,m-p, \\ -\sum_{j=p-i+1}^{m-i}\phi_j/C_{m,p} & \text{for } i = m-p+1,\ldots,p, \\ \left(1-\sum_{j=1}^{m-i}\phi_j\right)/C_{m,p} & \text{for } i = p+1,\ldots,m, \end{cases}$$

where

$$C_{m,p} = (m-p)\left(1-\sum_{j=1}^{p}\phi_j\right).$$

Since the errors are uncorrelated, each weighted batch mean is uncorrelated with the others at all lags. Using these weights

$$\mathrm{Var}\left(\overline{\overline{Y}}\right) = \frac{\sigma^2}{(n-kp)\left(1-\sum_{j=1}^{p}\phi_j\right)^2}.$$

The procedure for using the method of weighted batch means in practice is then as follows. After truncating the initial transient (see Law and Kelton 1991), we select $\tilde{p}$ and $\tilde{q}$, our best guess at the parameters of the ARMA(p,q) process from which the data was generated. For this identification an automatic procedure due to Gray, Kelley, and McIntire (1978) can be used. Choice of $\tilde{p}$ and $\tilde{q}$ are limited to the ARMA processes for which optimal analytic weights are currently known: the AR(p) and MA(1) processes. If $\tilde{p}$ and $\tilde{q}$ are both zero the data appear to be i.i.d. and unweighted batch means (for example) can be used. If either $\tilde{p}$ or $\tilde{q}$ is nonzero, the ϕ and θ vectors are estimated. A number of batches, k, is chosen, and weights are computed that are a function of m, $\tilde{p}$, $\tilde{q}$, $\hat{\phi}$, and $\hat{\theta}$. A $100(1-\alpha)\%$ confidence interval can then be produced from the weighted batch means using (2).

3 CONTROLLING SOURCES OF BIAS

From experimental results given in Bischak, Kelton, and Pollock (1993) it appears that in cases where the sample is small and the process displays strong autocorrelation, whether positive or negative, the method of weighted batch means provides better coverage than unweighted batch means. This is apparently due to the fact that the optimal weights produce a valid confidence interval if the process order and its parameters are known. However, when WBM is employed in practice the process order and parameters will not be known, and $\overline{\overline{Y}}$ and $\widehat{\mathrm{Var}}\left(\overline{\overline{Y}}\right)$ may be biased or correlated with each other; this may reduce coverage.

There are several potential sources of bias. In order to obtain the optimal weights for a given set of data, the order (p, q) of the ARMA process must be correctly identified. There are two reaons why bias in identification may be introduced. The first is that, so far, there is a limited set of models for which optimal weights have been determined, namely, the AR(p) and MA(1) processes. Due to computational difficulties with estimating the parameters, in our experiments we further restricted the choice of ARMA models to one of the AR(p) processes with $p \leq 4$ or the MA(1) process. There were a large number of experiments in which our identification procedure found that an ARMA model outside this set of processes was more suitable; we forced the procedure to reconsider and to select a model from the allowed set. The second problem with identification is that misspecification may occur, both because the data sets are small and also because the automatic identification procedure does not operate as well as would subjective judgment.

Even with correct identification of the process, poor estimation of $\phi_1, \ldots, \phi_p$ and $\theta_1, \ldots, \theta_q$ may bias $\overline{\overline{Y}}$ and $\widehat{\text{Var}}(\overline{\overline{Y}})$. We know, for example, that the usual estimators of ϕ for the AR(1) process are biased low. Once the parameters are estimated, the weights, which are functions of the order and the parameters, will be observations of the random variables $\hat{w}_i$. These estimators of the weights may be biased. In addition, if the weights are estimated from the data to which they will be applied, bias may result from the dependence of $\hat{w}_i$ and X_{ij}.

As an example of the effects of these sources of bias, consider the point estimator $\overline{\overline{Y}} \equiv \sum_{j=1}^{k} \overline{Y}_j / k$. This estimator will be unbiased if $\text{E}[\overline{Y}_j] = \mu$. If p, q, the ϕ's, and the θ's are known, the weights are constants and

$$\begin{aligned}
\text{E}[\overline{Y}_j] &= \text{E}[\sum_{i=1}^{m} w_i X_{ij}] \\
&= \sum_{i=1}^{m} w_i \text{E}[X_{ij}] \\
&= \mu \sum_{i=1}^{m} w_i \\
&= \mu,
\end{aligned}$$

where the last step follows from the fact that the weights sum to 1. In this case we have by a result similar to (4) that the weighted batch means are i.i.d. normal, the point estimator $\overline{\overline{Y}}$ and variance estimator $\widehat{\text{Var}}\left(\overline{\overline{Y}}\right)$ are unbiased and uncorrelated, and the confidence interval is valid.

If only the order of the process is known or if nothing is known about the process, the weights are estimated and

$$\begin{aligned}
\text{E}[\overline{Y}_j] &= \text{E}[\sum_{i=1}^{m} \hat{w}_i X_{ij}] \\
&= \sum_{i=1}^{m} \text{E}[\hat{w}_i X_{ij}].
\end{aligned}$$

If the order is known and the data and weights are dependent, $\text{E}[\hat{w}_i X_{ij}]$ must be derived in order to determine whether $\overline{Y}_j$ is unbiased. This expectation is difficult to derive analytically. If the order is known and the data and weights are independent, $\overline{Y}_j$ is unbiased since

$$\begin{aligned}
\sum_{i=1}^{m} \text{E}[\hat{w}_i X_{ij}] &= \sum_{i=1}^{m} \text{E}[\hat{w}_i]\text{E}[X_{ij}] \\
&= \mu \sum_{i=1}^{m} \text{E}[\hat{w}_i] \\
&= \mu \text{E}[\sum_{i=1}^{m} \hat{w}_i] \\
&= \mu,
\end{aligned}$$

where the last step follows because the weight estimators sum to 1. However, $\widehat{\text{Var}}\left(\overline{\overline{Y}}\right)$ may still be biased; one possible reason is that if non-optimal weights are used, the weighted batch means may not be uncorrelated. The point and variance estimators may also be correlated, with possible implications for coverage.

If the order of the process is unknown, whether or not the data and weights are independent, we may end up choosing the wrong order and hence the wrong form for the weights; for instance, weights which are optimal for an MA(1) process may be (incorrectly) used on observations from an AR(2) process. It is difficult to determine analytically the effect of this misspecification since the probability with which an incorrect process is chosen will depend both on the method used for identification and, since identification can be subjective, on the person performing the identification.

4 EMPIRICAL RESULTS ON IMPROVEMENTS IN COVERAGE

In order to examine the effects of each of these sources of bias, experiments were run on observations from the following processes: the AR(2) process with $\phi_1 = 0.6$, $\phi_2 = 0.3$, and normal errors with mean 0 and variance 1; the EAR(1) process with $\phi = 0.9$ and $\lambda = 1.0$; and the queue-delay process in an M/M/1 queueing system in steady-state with traffic intensity 0.8 and $\mu = 3.2$. The EAR(1) is defined as follows:

$$X_t = \begin{cases} \phi X_{t-1} & \text{w. p. } \phi \\ \phi X_{t-1} + \varepsilon_t & \text{w. p. } 1-\phi, \end{cases}$$

where ε_t is i.i.d. exponential with mean λ and $0 \leq \phi < 1$. The autocovariance function of the EAR(1) is analogous to that of the AR(1). Each of these processes exhibits strong positive correlation.

2,000 replications of each experiment were performed. Data from the processes were generated beginning in steady-state. Run lengths of 100 and 200 were used, and the data were batched into $k = 5$, 10, and 20 batches. All runs are independent across n and k, but runs on the same process use the same random numbers for each case so that the differences between cases can be more easily discerned. Thus in a given table, each row for a given n uses the same random numbers. Candidate models for identification were limited to a set of six: AR(p) processes with $p \leq 4$, the MA(1) process, and i.i.d. data.

The achieved coverage and average half-length for nominal 90% confidence intervals were calculated using UBM and WBM. (95% confidence interval half-lengths on coverage are in all cases no more than 0.022.) These experiments were then repeated with one or more sources of bias controlled, as follows:

1. weights independent of the data: a separate set of data independent of the original data was used to estimate the weights.

2. p known: $\tilde{p}$ is set to p (for the AR(2) only).

3. weights independent of the data and p known (for the AR(2) only).

4. p known and ϕ known: $\tilde{p}$ is set to p and $\hat{\phi}$ to ϕ (for the AR(2) only).

Also estimated were the relative bias of the UBM and WBM variance estimators for the AR(2) process and the correlation between the point and variance estimators for the EAR(1) and queue-delay processes. Bias was calculated only for the AR(2) because optimal weights, and hence Var $\left(\overline{\overline{Y}}\right)$, are not defined for the other two processes. The correlation of the point and variance estimators for the AR(2) process was found to be nearly zero for all cases. This is similar to the result of Kang and Goldsman (1990) that the correlation of $\overline{\overline{X}}$ and the UBM variance estimator is zero.

Table 1 shows that the achieved coverage for the AR(2) process is significantly greater for WBM than for UBM. Coverage is even greater if the weights for WBM are estimated from a separate set of data. Knowing more about the process improves coverage, as would be expected, but when the number of batches is small, estimating from a separate set of data improves coverage about the same amount as does knowing the order of the process. Coverage is improved further if we both know p and also estimate the weights independently of the data.

Table 1: Achieved Coverage for the AR(2) Process, $\alpha = 0.10$

		k		
n	Case	5	10	20
100	UBM	0.780	0.634	0.481
	WBM	0.818	0.737	0.693
	Weights independent	0.853	0.765	0.710
	p known	0.846	0.780	0.780
	p known, weights independent	0.879	0.817	0.793
	p, ϕ known	0.915	0.897	0.901
200	UBM	0.840	0.766	0.630
	WBM	0.861	0.845	0.782
	Weights independent	0.873	0.853	0.799
	p known	0.880	0.872	0.841
	p known, weights independent	0.891	0.884	0.851
	p, ϕ known	0.907	0.915	0.908

When everything is known about the AR(2) process from which the data are generated, the coverage is on average about 0.90, as expected. This is the only case in which coverage does not deteriorate as the number of batches increases. An interesting result in the AR(2) experiments is that if we have $n = 100$ data values and use another set of data of the same size to estimate the weights, we still do at least as well as UBM with $n = 200$. However, we would be better off using WBM on the full set of 200 data values.

Table 2 shows for each n and k estimates of the relative bias

$$\left(\mathrm{E}\left[\widehat{\mathrm{Var}}\left(\overline{\overline{Y}}\right)\right] - \mathrm{Var}\left(\overline{\overline{Y}}\right)\right) / \mathrm{Var}\left(\overline{\overline{Y}}\right)$$

for each case and the actual variance for both UBM and WBM. (For UBM $\overline{\overline{X}}$ replaces $\overline{\overline{Y}}$ in the above expression.) For the most part there is a close inverse relationship between relative bias in the variance estimator and coverage. The variance estimator is biased low for all cases except that it is unbiased when both p and ϕ are known, as expected. The relative bias is greatest for UBM. Estimating weights independently appears to *increase* the relative bias on average, even though this improves coverage.

Table 2: Estimated Relative Bias of Variance Estimators and True Variance for AR(2) Process

n	Case	k = 5	10	20
100	UBM	-0.529	-0.724	-0.844
	WBM	-0.320	-0.493	-0.586
	Weights independent	-0.373	-0.518	-0.563
	p known	-0.142	-0.309	-0.404
	p known,weights independent	-0.243	-0.356	-0.362
	p, ϕ known	0.013	-0.002	-0.007
	Var $\overline{\overline{X}}$	0.873	0.873	0.873
	Var $\overline{\overline{Y}}$	1.111	1.250	1.667
200	UBM	-0.318	-0.519	-0.703
	WBM	-0.144	-0.241	-0.368
	Weights independent	-0.171	-0.257	-0.392
	p known	-0.015	-0.054	-0.146
	p known,weights independent	-0.063	-0.099	-0.187
	p, ϕ known	0.021	0.022	-0.003
	Var $\overline{\overline{X}}$	0.468	0.468	0.468
	Var $\overline{\overline{Y}}$	0.526	0.556	0.625

Results on coverage in Table 3 for the EAR(1) process are similar to those for the AR(2). The best coverage is achieved with WBM and independent weights. However, the lowest correlation between $\overline{\overline{Y}}$ and $\widehat{\text{Var}}\left(\overline{\overline{Y}}\right)$ occurs when WBM is used with *dependent* weights, as shown in Table 4.

Figures 1 and 2 show graphically the achieved coverage and average half-length for the M/M/1 queue-delay process for $n = 100$ and $n = 200$, respectively. The two figures are very similar except that coverage improves somewhat for all three cases with larger n. Coverage with WBM is closer to the target value of 0.90, but the average half-length is greater. Independently estimating the weights decreases the average half-length slightly while also improving coverage slightly.

Table 5 shows that the estimated correlation of $\overline{\overline{Y}}$ and $\widehat{\text{Var}}\left(\overline{\overline{Y}}\right)$ for the M/M/1 process is less for WBM than for UBM. The correlation decreases further with independent weights. However, it *increases* with n. In this respect the correlation seems unrelated to the coverage.

5 CONCLUSIONS

From these experiments, we conclude that coverage with weighted batch means can be improved by using a separate run to estimate the weights, but the coverage from simply doubling the original sample size may be better. Improved identification and estimation for processes that can be well represented by an ARMA process will also improve coverage. The relationship among coverage, variance estimator bias, and the correlation of $\overline{\overline{Y}}$ and $\widehat{\text{Var}}\left(\overline{\overline{Y}}\right)$ when weights are estimated independently seems to go against intuition and should be explored further.

REFERENCES

Bischak, D. P., W. D. Kelton, and S. M. Pollock. 1993. Weighted batch means for confidence intervals in steady-state simulations. *Management Science*, to appear.

Gray, H. L., G. D. Kelley, and D. D. McIntire. 1978. A new approach to ARMA modeling. *Communications in Statistics, B—Simulation and Computation* 7:1–77.

Kang, K., and D. Goldsman. 1990. The correlation between mean and variance estimators in computer simulation. *IIE Transactions* 22:15–23.

Law, A. M., and W. D. Kelton. 1991. *Simulation modeling and analysis*. 2nd ed. New York: McGraw-Hill.

AUTHOR BIOGRAPHY

Diane P. Bischak is an Assistant Professor in the Department of Business Administration, University of Alaska Fairbanks. All her education was obtained at the University of Michigan, Ann Arbor. Before receiving the M.S. and Ph.D. degrees in industrial and operations engineering, she received the B. Music in violin performance. Her research interests include the statistical analysis of simulation output and the modeling of stochastic production processes. She is a member of the TIMS College on Simulation and is also a member of ORSA, TIMS, and IIE.

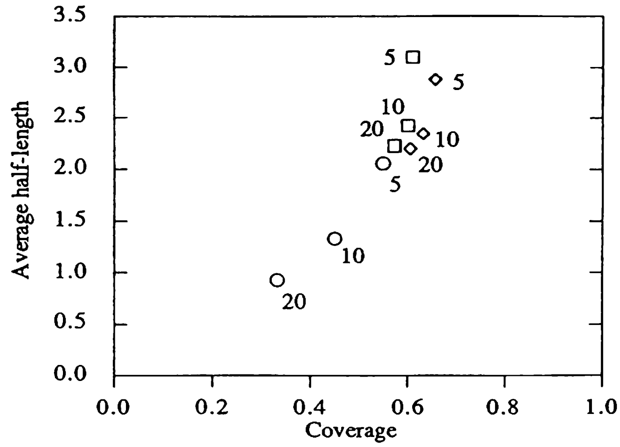

Figure 1: Achieved Coverage and Average Half-Length for the M/M/1 Queue-Delay Process, n = 100 and k = 5, 10, 20 (Circle=UBM, Square=WBM, Diamond=Independent Weights)

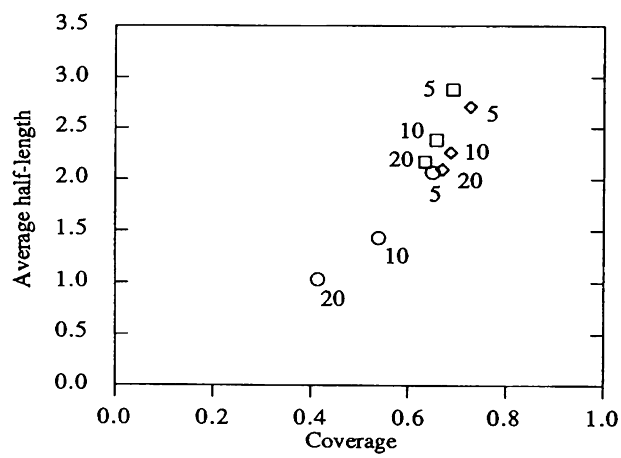

Figure 2: Achieved Coverage and Average Half-Length for the M/M/1 Queue-Delay Process, n = 200 and k = 5, 10, 20 (Circle=UBM, Square=WBM, Diamond=Independent Weights)

Table 3: Achieved Coverage for the EAR(1) Process, $\alpha = 0.10$

		k		
n	Case	5	10	20
100	UBM	0.768	0.666	0.548
	WBM	0.810	0.777	0.756
	Weights independent	0.832	0.793	0.767
200	UBM	0.840	0.774	0.683
	WBM	0.875	0.832	0.830
	Weights independent	0.883	0.849	0.839

Table 4: Estimated Correlation of Point and Variance Estimators for the EAR(1) Process

		k		
n	Case	5	10	20
100	UBM	0.586	0.682	0.707
	WBM	0.509	0.538	0.556
	Weights independent	0.586	0.622	0.633
200	UBM	0.512	0.636	0.653
	WBM	0.444	0.527	0.483
	Weights independent	0.503	0.606	0.627

Table 5: Estimated Correlation of Point and Variance Estimators for the M/M/1 Queue-Delay Process

		k		
n	Case	5	10	20
100	UBM	0.641	0.725	0.724
	WBM	0.522	0.462	0.511
	Weights independent	0.515	0.301	0.174
200	UBM	0.790	0.778	0.823
	WBM	0.611	0.508	0.542
	Weights independent	0.605	0.293	0.245