Simulation Efficiency for Estimation of System Parameters in Computer Simulation

Chimyung Kwon

Department of Applied Statistics, Dong-A University

시뮬레이션 실험에서 시스템 성과에 대한 추정치의 정확도를 개선하기 위한 분산감소기법 (Variance Reduction Technique)은 입력영역과 출력영역에 대한 것으로 나누어 볼 수 있다. 본 연구에서는 시스템 성과 추정량이 단일 변량인 경우에, 분산감소기법으로 많이 사용되는 Antithetic Variates 방법과 Control Variates 방법을 결합하여 응용가능한 시뮬레이션 실험설계 기법을 제시하고 이 기법을 선택된 모형에 적용하여 시뮬레이션의 효율성을 분석하였다. 실험결과, 제안된 기법은 기존 방법들 보다 추정치의 분산을 5%-8% 더 감소시켰다. 비록 제한된 실험결과이지만 이러한 효과는 대형 시뮬레이션의 경우에 적지 않으리라 기대된다. 특히 효과적인 Control Variates의 수가 적은 경우, 제안된 기법은 매우 효율적이다.

I. Introduction

Through a controlled simulation experiment, often an experimenter is concerned with estimating the mean response of interest from the outputs of the simulation model. Frequently large-scale systems analysis through simulation requires extensive experimentation with a simulation model to obtain acceptable precision in the estimator of interest. If we can reduce the variance of the estimator of interest at little additional cost, we can obtain greater precision of the estimator with the same amount of simulation. In this work we propose a method of combining variance reduction techniques for improving the estimation on the mean response of interest.

For a single population model, usually antithetic variates and control variates are applied to reduce the error of the estimator for the mean response. Antithetic variates assigns complementary random numbers to pairs of simulation runs taken at a single design point to induce a negative correlation between the responses. If the covariance between two responses obtained by antithetic replicates is negative, then the variance of the estimator for the mean response is less than that obtained by two independent replicates.

In contrast to the approach of antithetic variates, the method of control variates attempts to exploit correlations between the response and selected control variates within a single run. Let y_i and c_i be the response of interest and the (sx1) vector of control variates, respectively, obtained from the ith simulation run with $E[c_i]=0$. In the context of performing 2h independent replications of the simulation, the normality assumption on the response of interest and control variates allows that the response is represented as the following linear model:

random numbers used for driving the jth stochastic component of the simulation model at the ith replication. Also let $\mathbf{R}_i$ be the set of g random number streams for the ith replication:

$$\mathbf{R}_i = (\mathbf{r}_{i1}, \mathbf{r}_{i2}, \ldots, \mathbf{r}_{ig}) \quad \text{for} \quad i = 1, 2, \ldots, 2h.$$

We now consider the random number assignment strategy of jointly utilizing antithetic variates and control variates for a simulation model which requires g such random number streams to drive all of its stochastic components at a single replication. To this end, we separate $\mathbf{R}_i$ into two mutually exclusive and exhaustive subsets of random number streams, $(\mathbf{R}_{i1}, \mathbf{R}_{i2})$ (i=1, 2, ..., 2h). The first subset, $\mathbf{R}_{i1}$, consisting of (g-s) random number streams is used to drive the non-control stochastic model components. The second subset, $\mathbf{R}_{i2}$, consists of s random number streams used to drive the control variate stochastic model components.

We consider the correlated replication strategy: use antithetic variates for all stochastic components except the control variates across 2h replications. Through statistical analysis and simulation experimentation, we will explore how this method may improve the simulation efficiency in reducing the variance of the estimator, and what conditions are necessary for this method to ensure an improvement in variance reduction. That is, within the ith paired replications, this method uses $(\mathbf{R}_{2i-1,1}, \mathbf{R}_{2i-1,2})$ and $(\bar{\mathbf{R}}_{2i-1,1}, \mathbf{R}_{2i,2})$ where $\mathbf{R}_{2i-1,1}$, $\mathbf{R}_{2i-1,2}$, and $\mathbf{R}_{2i,2}$ are sets of randomly selected random number streams, and $\bar{\mathbf{R}}_{2i-1,1}$ is antithetic to $\mathbf{R}_{2i-1,1}$. Across pairs of replications, this method uses independent streams. Thus, the ith pair of responses, y_{2i-1} and y_{2i} (i=1, 2,...,h), are negatively correlated by antithetic streams through the non-control stochastic components. However, through the 2h replications, the control variates $\mathbf{c}_i$ (i=1, 2, ..., 2h) are independently generated by the assignment of independent streams through the control variate stochastic components at each replication. Due to independent streams for the control variates, the response y_{2i-1} (y_{2i}) is independent of control variates $\mathbf{c}_{2i}$ ($\mathbf{c}_{2i-1}$) within a paired simulation output. Based on the above discussions, we establish the following assumptions:

1. $Var(y_i) = \sigma_y^2$, for i= 1, 2, ..., 2h (homogeneity of response variances across replicates),

2. $Cov(y_i, y_j) = -\rho\, \sigma_y^2$ $(\rho > 0)$, if j= i+1 (i=1, 3, ..., 2h-1) (homogeneity of induced negative correlations across replicates pairs). Otherwise, $Cov(y_i, y_j) = 0$,

3. $Cov(y_i, \mathbf{c}_i) = \sigma_{yc}'$ for i=1, 2, ..., 2h (homogeneity of control variates response covariance across replicates), and $Cov(y_i, \mathbf{c}_j) = \mathbf{0}'$, for $i = j$,

4. $Cov(\mathbf{c}_i) = \Sigma_c$, for i= 1, 2, ..., 2h (homogeneity of control variates covariance structure across replicates), and

5. $Cov(\mathbf{c}_i, \mathbf{c}_j) = \mathbf{0}_{s\times s}$, for $i = j$ (independence of control variates between replicates).

$$y = \mu_y 1_{2h} + C\alpha + \varepsilon \qquad (1)$$

where $y = (y_1, y_2, \ldots, y_{2h})'$, 1_{2h} is a (2hx1) vector of 1's, C is a (2hxs) control variate matrix whose ith row consists of c_i, μ_y is a parameter of the mean response, α is a (sx1) coefficient vector of control variates, and ε is the (2hx1) vector of error terms (see Lavenberg, Moeller and Welch 1982). The least squares estimators of α and μ_y in the linear model in (1) are given by, respectively,

$$\hat{\alpha} = (C'PC)^{-1}C'Py \quad \text{and} \quad \hat{\mu}_y = \bar{y} - \bar{c}\,\hat{\alpha}\ ,$$

where $\bar{y}$ and $\bar{c}$ are the mean response and mean control variate observations across 2h replications, and $P = I_{2h} - 1_{2h}1'_{2h}/2h$ (see Searle 1971 p. 341). Under the assumption that $\varepsilon \sim$ IID $N(0, \sigma^2_{y|c})$, the least squares estimator μ_y is an unbiased estimator for μ_y. Lavenberg, Moeller and Welch (1982) showed that the unconditional variance of μ_y is given by

$$\mathrm{Var}(\mu_y) = (2h - 2)/(2h - s - 2)\ (1 - R_{yc}^2)\sigma_y^2/2h, \qquad (2)$$

where $R_{yc}^2 = \sigma_y^{-2}\sigma_{yc}'\Sigma_{yc}^{-1}\sigma_{yc}$ is the square of the multiple correlation coefficient between y_i and c_i. They defined the quantity (2h-2)/(2h-s-2) as the loss factor due to the estimation of the unknown parameter α in (1), and ($1-R_{yc}^2$) as the minimum variance ratio which represents the potential for reducing the variance of the estimator by the control variates. Thus, the efficiency of control variates is measured by the product of the loss factor and the minimum variance ratio.

In this research, our main interest is to combine these two correlation methods that utilize correlations between simulation output either within a single run or across different replications in one simulation experiment for improving the estimation of the mean response of interest. Suppose that through correlated replications of simulation runs, we get a reduced variance of the estimator for the mean response and yet maintain the same correlation between the response and control variates as those obtained under independent replications. Then it is conjectured that we may take advantage of both antithetic variates and control variates together in one simulation run, and reduce the variance of the estimator further than by applying either antithetic variates or control variates separately.

Based on this conjecture, this research focuses on developing a method for combining antithetic variates and control variates for the estimation of the mean response of interest. For this purpose, we consider a method of utilizing induced correlations between: (a) the responses of interest, and (b) the response and a set of control variates obtained by an appropriate assignment of random numbers streams through the replications, and try to improve upon the simulation efficiency of the control variates method.

II. Simulation Efficiency of Combined Method

In computer simulation, random number streams that drive a simulation model are under the control of the experimenter and completely determine the simulation output. Let the random number stream r_{ij} denote the sequence of

Under these assumptions, the variances of the mean responses and mean control variates within the ith replication pair, $\bar{y}_i = (y_{2i-1} + y_{2i})/2$ and $\bar{c}_i = (c_{2i-1} + c_{2i})/2$ are given by, respectively,

$$Var(\bar{y}_i) = (1-\rho)\,\sigma_y^2/2, \quad \text{and} \quad Cov(\bar{c}_i) = \Sigma_c/2.$$

Also, the covariance between $\bar{y}_i$ and $\bar{c}_i$ is given by

$$Cov(\bar{y}_i, \bar{c}_i) = Cov(y_{2i-1} + y_{2i}, c_{2i-1} + c_{2i})/4 = \sigma_{yc}'/2.$$

Thus, the joint normality assumption of the response and control variates gives the joint distribution of $\bar{y}_i$ and $\bar{c}_i$ as follows:

$$\begin{bmatrix} \bar{y}_i \\ \bar{c}_i \end{bmatrix} \sim N_{s+1}\left[\begin{bmatrix} \mu_y \\ 0 \end{bmatrix}, \; 1/2\begin{bmatrix} (1-\rho)\sigma_y^2 & \sigma'_{yc} \\ \sigma_{yc} & \Sigma_c \end{bmatrix}\right] \tag{3}$$

Consequently, $\bar{y}_i$, given $\bar{c}_i$, is normally distributed with expectation $E[\bar{y}_i|\bar{c}_i] = \mu_y + \alpha\bar{c}_i$ and variance

$$Var(\bar{y}_i|\bar{c}_i) = [(1-\rho)\sigma_y^2 - \sigma_{yc}'\Sigma_c^{-1}\,\sigma_{yc}]/2 = \tau_1^2/2. \tag{4}$$

(see Theorem 2.5.1 in Anderson 1984). As with the case of the linear relationship in (1), the vector of the mean paired responses, y, can be represented as the following linear model:

$$\bar{y} = \mu_y 1_h + \bar{C}\alpha + \varepsilon^*, \tag{5}$$

where $\bar{C}$ is a (hxs) control variate matrix whose ith row is $\bar{c}'_i$. Regression analysis on this linear model yields the controlled estimator for the mean response as

$$\hat{\mu}_y = 1'_h[\bar{y} - \bar{C}(\bar{C}'Q\bar{C})^{-1}\bar{C}Q\bar{y}]/h = 1'_h[I_h - \bar{C}(\bar{C}'Q\bar{C})^{-1}\bar{C}'Q)]\bar{y}/h,$$

where $Q = I_h - 1_h1'_h/h$. Given $\bar{C}$, taking the operation of variance on the above equation yields

$$Var(\mu_y|\bar{C}) = 1/h^2\; 1'_h[I_h - \bar{C}(\bar{C}'Q\bar{C})^{-1}\bar{C}'Q]Var(\bar{y}|\bar{C})[I_h - Q\bar{C}((\bar{C}'Q\bar{C})^{-1}\bar{C}']1_h$$

Since $(\bar{y}_i, \bar{c}_i)$ of the ith pair of simulation output is independent of that of a different pair of replications, from equation (4), we have

$$Var(\bar{y}|\bar{C}) = [(1-\rho)\sigma_y^2 - \sigma_{yc}'\Sigma_c^{-1}\sigma_{yc}]I_h/2 = \tau_1^2 I_h/2.$$

Substituting for $Var(\bar{y}|\bar{C})$ into $Var(\hat{\mu}_y|\bar{C})$ gives

$$Var(\hat{\mu}_y|C) = \tau_1^2[h + 1'_h\bar{C}(\bar{C}'Q\bar{C})^{-1}\bar{C}'1_h]/(2h^2), \tag{6}$$

since $Q1_h = 1'_hQ = 0$. From the result of (3) and assumption 5, the (hxs) random matrix $\bar{C}$ has the matrix normal distribution: $\bar{C} \sim N_{h,s}(0, I_s, \Sigma_c/2)$, where 0 is a (hxs) matrix of zeroes. Thus, by definition of the Wishart distribution (see Section 17.3 in Arnold 1981), $\bar{C}(\Sigma_c/2)^{-1}\bar{C}' \sim W_h(s, I_h)$ and, by Theorem 17.7a in Arnold (1981), the (sxs) random matrix $(\bar{C}'Q\bar{C})$ follows the Wishart distribution: $(\bar{C}'Q\bar{C}) \sim W_s(h-1, \Sigma_c/2)$ since Q is an idempotent matrix with rank (h-1). We note that $(1'_h\bar{C})$ and $(\bar{C}'Q\bar{C})$ are independent. Thus, the expectation of the conditional variance in (6) can be written as

$$Var(\hat{\mu}_y)=E[Var(\hat{\mu}_y|\bar{C})]= \tau_1{}^2/(2h^2)E[h+1'_h\bar{C}\ E[(\bar{C}'Q\bar{C})^{-1}]\ \bar{C}'1_h]. \qquad (7)$$

Theorems 17.6a and 17.15d in Arnold (1981) give, respectively,

$$E[\bar{C}\Sigma_c/2)^{-1}\bar{C}'] = sI_h, \text{ and } E[(\bar{C}'P\bar{C})^{-1}] = [(\Sigma_c/2)^{-1}/(h-s-2)] \text{ if } h > (s+2)$$

Therefore, plugging the second equation of this equation into (7) finally yields

$$Var(\hat{\mu}_y)=\tau_1{}^2/(2h^2)[h+hs/(h-s-2)] = \sigma_y{}^2/(2h)x(1-\rho-R_{yc}{}^2)x(h-2)/(h-s-2), \qquad (8)$$

where R_{yc} is the multiple correlation coefficient between y_i and c_i (i=1, 2, ..., 2h). This result indicates that the minimum variance ratio of this method is $(1-\rho-R_{yc}{}^2)$, and the loss factor is (h-2)/(h-s-2).

III. Comparison of Combined Method and Control Variates Method

We compare Combined Method developed in the previous section and the method of control variates with respect to the unconditional variances of the estimators for the mean response, and summarize these results. A comparison of equations (2) and (8) yields that Combined Method is better than the control variates method if

$$(1-\rho-R_{yc}{}^2)\ (h-2)/(h-s-2) < (1-R_{yc}{}^2)(2h-2)/(2h-s-2)$$

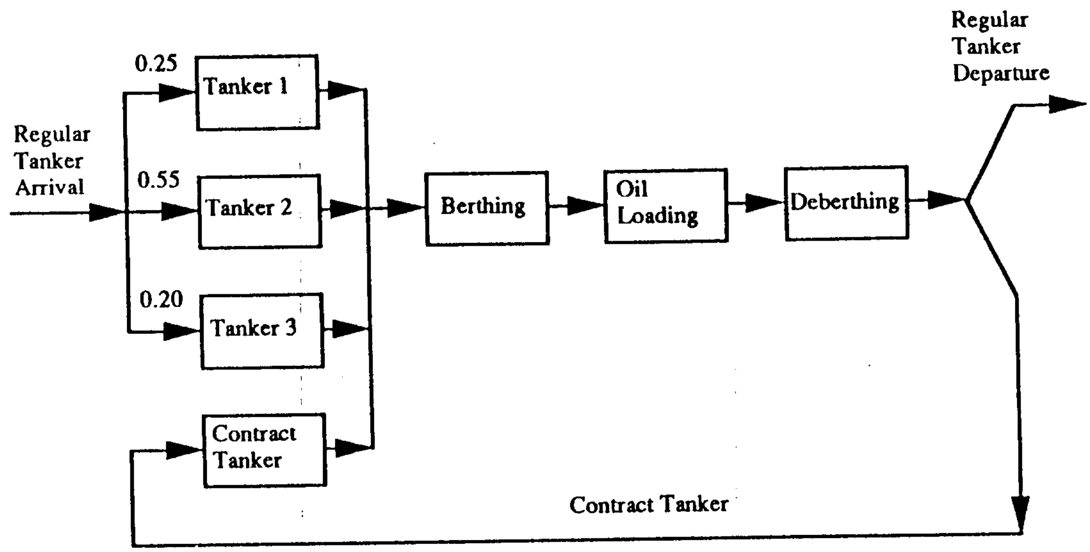
As shown in this equation, the loss factor of combined method is greater than that of the control variates method. Hence, for preference of the combined method to the control variates method, the minimum variance ratio of the combined method should, at least, compensate for an increase in the associated loss factor. As we see, the effects of antithetic variates and control variates to the minimum variance ratio for Combined Method is represented by an additive form in reducing the variance of the estimator for the mean response.

IV. Examples

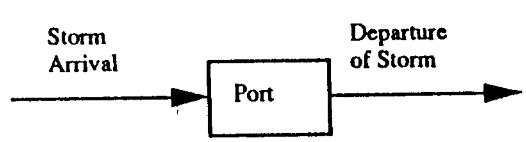
We conducted a set of simulation experiments on a system to evaluate the performance of the variance reduction techniques considered earlier. We offers brief descriptions of a system and the methods used to simulate it.

Figure 1 shows the port operations model (see p. 197 in Prisker 1986). A port in Africa is used to load tankers with crude oil for overwater shipment. The port has facilities for loading as many as three tankers simultaneously. The tankers, which arrive at the port according to a uniform distribution with range [4, 18] hours, are of three types The relative frequency of the various types, their loading time requirements, and their distributions of loading time as follows:

type	relative frequency	loading time(hours)	distribution
1	0.25	[16, 20]	uniform
2	0.55	[21, 27]	uniform
3	0.20	[32, 40]	uniform



(a) Tanker Arrival and Port Operation Segment



(b) Storm Segment

Figure 1. Port Operations Model

There is one tug at the port. Tankers of all types require the services of this tug to move into a berth, and later to move out of a berth. When the tug is available, any berthing or deberthing activity takes about one hour. Top priority is given to the berthing activity. A shipper is considering bidding on a contract to transfer oil from the port to the United Kingdom. He has determined that 5 tankers of a particular type would have to be committed to this task to meet contract specifications. These tankers would require [18, 24] hours, uniformly distributed, to load oil at the port. After loading and deberthing, they would travel to the United Kingdom, offload the oil, and return to the port for reloading. Their round-trip travel time, including offloading, is estimated to be [216, 264] hours with a uniform distribution. A complicating factor is that the port experiences storms. The time between the onset of storms is exponentially distributed with a mean of 48 hours, and a storm lasts [2, 6] hours, uniformly distributed. No tug can start an operation until a storm is over. Before the port authorities can commit themselves to accommodating the proposed 5 tankers, the effect of the additional port traffic on the in-port residence time of the current port users must be determined. It is desired to simulate the operation of the port over a two-year period (19,280 hours) under the proposed new commitment to measure in-port residence time of the proposed additional tankers, as well as the three types of tankers which already use the port.

The port operations model includes nine stochastic components to which nine separate random number streams are assigned. Direct simulation and antithetic variates, respectively, use the same assignment rules in selecting a set of nine random number streams through the replications as before. In using the control variates method, seven possible standardized control variates present themselves (see the definition of standardized control variate in Wilson and Pritsker 1984a and 1984b). That is, inter-arrival times of tankers of three different types which are already in the system, oil loading times of each tanker (three regular types tankers and tankers on a contract), round trip travel times of tankers on a contract, and duration of storm. We collected six control variates except the storm duration control variates since we expected that the frequency of storm is low and its in-port residence time is small. Table 1 shows the correlation matrix between the four responses of interest and the six collected control variates obtained by 200 independent replications. Based on this table, we employed the three control variates of interarrival times of tankers already in system and oil loading times of tankers of type 1 and 2 for implementing the three combined methods.

Table 1. Correltion Matrix between the Responses and Control Variates

	c_1	c_2	c_3	c_4	c_5	c_6
y_1	-0.689	0.133	0.288	-0.049	-0.029	-0.040
y_2	-0.675	0.113	0.278	-0.039	-0.015	-0.038
y_3	-0.639	0.108	0.252	-0.040	-0.028	-0.033
y_4	-0.698	0.114	0.267	-0.059	-0.011	-0.042

Combined Method employs (a) the same assignment rule as direct simulation for the first replicate within each pair of replication, and (b) a set of nine streams, those that correspond to the control variates (stream 1, 2 and 3) are randomly selected, and the others are set antithetic to their counterparts in the first replication for the second replication. However, across the pairs of replications, each of these methods randomly selects a set of nine random number streams. We coded this model in SLAM II and conducted a simulation of this system 200 times for each method. Each method simulated the model for 21000 hours, and collected statistics after clearing data for the first 1000 hours to reduce the initialization bias.

V. Experimental Results

This section provides a summary of simulation results obtained by employing antithetic variates, control variates and the combined method to the port operations model. To provide an assessment of the efficiency gain obtained by each estimation method, we calculated performance statistics of the percentage reduction in variance and width of a nominal 90% confidence interval for each applied method.

Tables 2 and 3, respectively, summarize the results on percentage reductions in variance and 90% half-length confidence intervals for each response of interest (control variates used the three most effective ones). In computing the efficiency of control variates method, regression analysis on all six control variates indicates reduction in variance for each response of interest in the range from 40% to 50%. When we chose the three most effective control variates (c_1, c_2, c_3) in Table 1, regression analysis showed an increment of reduction in variance for each response by around 3%.

Table 2. Percentage Reduction in Variance

Estimator (Sojourn Time in port)	Antithetic Variates	Contol Variates	Combined Method
Tanker 1	51.63	53.23	60.06
Tanker 2	51.16	50.37	56.80
Tanker 3	45.90	44.55	50.10
Tanker on Contract	54.00	53.03	61.15

Table 3. Percentage Reduction in 90% Confidence Interval

Estimator (Sojourn Time in port)	Antithetic Variates	Contol Variates	Combined Method
Tanker 1	29.70	31.61	36.10
Tanker 2	29.37	29.55	33.55
Tanker 3	25.66	25.54	28.58
Tanker on Contract	29.21	29.31	35.58

Based on the simulation results of this model, we provide inferences in applying variance reduction techniques as follows: (a) antithetic variates and control variates reduce the variance of the estimator for each response in the range from 45% to 55%, and their performances are similar; (b) the efficiency gain of Combined Method shows the additive effects of antithetic variates and control variates, and reduces the variance of each estimator more than antithetic variates and control variates in the range from 5% to 8%, and the 90% confidence interval in the range from 3% to 6%;

VI. Conclusions

From the simulation experiment on the selected model, we note that (a) Combined Method shows the additive effects of antithetic variates and control variates in reducing the variance of the estimator, and (b) the performance of Combined Method was better than those of control variates and antithetic variates.

In combining antithetic variates and control variates, we used a strategy using independent streams for driving the control variates. We may use an antithetic variates for driving the control variates for the case that synchronization of random number streams is easily achieved in the model. Generally, for a complex model, an effective set of control variates is small. Also, the marginal effect of including one more control variate is very small when there is a strong correlation between a set of control variates already used in the system and the control variates to be added (see the discussion of Beja 1967). Thus, the combined method which is based on using the effective control variates and additionally trying to reduce the variance of the estimator by the correlated replicates may yield better results than applying either the control variates or antithetic variates separately for a complexmodel when the number of replications is not small. We expect this result may be useful in the design of a large-scale simulation.

References

Arnold, S. F., The Theory of Linear Models and Multivariate Analysis, John Wiley & Sons, New York, 1981.

Anderson, T. W., An Introduction to Multivariate Statistical Analysis, John Wiley & Sons, New York, 1984.

Beja, A., "Multiple Control Variates in Monte Calro Simulation with Application to Queueing Systems with Priorities," Developments in Operations Research, 255-264, B. Aviltzhak (Ed.), Gordon and Breach, New York, 1969.

Graybill, F. A., Theory and Application of the Linear Model, Duxury Press, North Scituate, Massachusetts, 1976.

Lavenberg, S. S., T. L. Moeller and P. D. Welch, "Statistical Results on Control Variates with Application to Queuing Network Simulation," Operations Research, 30 (1982), 182-202.

Pritsker, A. A. B., Introduction to Simulation and SLAM II, John Wiley & Sons, New York, 1986.

Searle, S. R., Linear Models. John Wiley & Sons, New York, 1971.

Wilson, J. R., "Variance Reduction Techniques for Digital Simulation," Journal of Mathematical and Management Sciences, 4 (1984), 227-312.

Wilson, J. R. and A. A. B. Pritsker, "Variance Reduction in Queuing Simulation Using Generalized Concomitant Variables," Journal of Statistical Computation and Simulation, 19 (1984a), 129-153.

----- and -----, "Experimental Evaluation of Variance Reduction Techniques for

Queuing Simulation Using Generalized Concomitant Variables," Management Science, 30 (1984b), 1459-1472.