European Journal of Operational Research 43 (1989) 184–196
North-Holland

Theory and Methodology

Batch size effects on the efficiency of control variates in simulation *

Barry L. NELSON
Department of Industrial and Systems Engineering, The Ohio State University, 1971 Neil Avenue, Columbus, OH 43210-1271, U.S.A.

Abstract: This paper considers the combined use of control variates and batching for estimating the steady-state mean of an infinite-horizon process via simulation. Properties of the point and interval estimators from such a procedure are derived as functions of the number of batches and the number of control variates when the total sample size is fixed.

Keywords: Simulation, multivariate statistics, regression

1. Introduction

Variance reduction techniques (VRTs) are used to reduce the population variance of estimators derived from simulation experiments on models of stochastic processes. Recent surveys of variance reduction include those of Nelson (1987), Nelson and Schmeiser (1986), and Wilson (1984). Most VRTs are designed for finite-horizon (sometimes called 'transient' or 'terminating') processes for which the natural experiment design is to sample independent and identically distributed (i.i.d.) realizations of the process. They can be directly adapted to infinite-horizon (sometimes called 'steady-state') processes by sampling i.i.d. (and usually long) realizations of the process. However, this approach may be impractical because of the need to model or delete an initial-transient period from each realization.

* This research was partially supported by a Seed Grant from the Office of Research and Graduate Studies, The Ohio State University.

Received June 1988; revised November 1988

In simulation output analysis, the initial-transient problem has led to the development of methods for point and interval estimation based on a single realization; these include nonoverlapping batch means (Schmeiser, 1982), overlapping batch means (Meketon and Schmeiser, 1984), regeneration (Crane and Lemoine, 1977), autoregressive representation (Schriber and Andrews, 1984), spectral analysis (Heidelberger and Welch, 1981), and standardized time series (Schruben, 1983). Similarly, it is desirable to apply VRTs in single-realization designs. This paper develops one approach.

The nonoverlapping batch means method ('batch means' from here one) and the regeneration method (at least attempt to) find i.i.d. batch means and segments of simulation output, respectively, within the output from a single realization. Thus, they are good candidate methods to combine with VRTs designed for i.i.d. realizations. If an automated procedure is desired, then the regeneration method is less attractive because of the need to identify regeneration points. The batch means method, while only approximating independence, is a procedure that can be almost en-

tirely automated (see, for example, Fishman, 1978; Law and Carson, 1979; Mechanic and McKay, 1966; Schriber and Andrews, 1979).

In this paper we consider the combined use of the batch means method of output analysis and the control variates VRT. Previously, Iglehart and Lewis (1979), Lavenberg and Welch (1981), and Wilson and Pritsker (1984a, b) combined the regeneration method and control variates. Sharon and Nelson (1988) give some empirical results for using a single control variate with batch means.

2. Review of batch means and control variates

To review the batch means method and control variates VRT, let the output of the simulation experiment be represented by a sequence of identically distributed, but possibly dependent, random (column) vectors $Z_i = [Y_i, X_{1i}, X_{2i}, \ldots, X_{qi}]'$, $i = 1, 2, \ldots, n$. Assuming that the sequence is identically distributed implies that initial-transient effects have somehow been mitigated. Let $E[Z_i] = [\theta, \mu_1, \mu_2, \ldots, \mu_q]'$ and $\mathrm{Cov}[Z_i] = \Sigma$, where

$$\Sigma = \begin{bmatrix} \sigma_y^2 & \sigma_{yx}' \\ \sigma_{yx} & \Sigma_x \end{bmatrix}, \tag{1}$$

so that σ_y^2 is the scalar $\mathrm{Var}[Y_i]$, Σ_x is the $q \times q$ matrix of $\mathrm{Cov}[X_{ji}, X_{mi}]$, $j, m = 1, 2, \ldots, q$, and σ_{yx} is the $q \times 1$ vector of $\mathrm{Cov}[Y_i, X_{ji}]$, $j = 1, 2, \ldots, q$. Thus, the square of the multiple correlation coefficient between Y_i and $[X_{1i}, X_{2i}, \ldots, X_{qi}]$ is

$$R_{yx}^2 = \frac{\sigma_{yx}' \Sigma_x^{-1} \sigma_{yx}}{\sigma_y^2}. \tag{2}$$

For our purposes, θ is the unknown parameter of interest and $X_{1i}, X_{2i}, \ldots, X_{qi}$ are the q control variates. To be useful as a control variate, in the sense that we use the term, X_{ji} must be correlated with Y_i and $\mu_j = E[X_{ji}]$ must be known. For convenience, define the column vector

$$[X_i - \mu] = [X_{1i} - \mu_1, X_{2i} - \mu_2, \ldots, X_{qi} - \mu_q]',$$

which has expectation $[0, 0, \ldots, 0]'$ and covariance matrix Σ_x. Our convention is to use single subscripts to denote column vectors and double subscripts to denote scalar elements, with the exception of Y_i which is a scalar random variable.

The idea behind batch means is to transform the n dependent vectors $Z_1, Z_2, \ldots, Z_n$ into fewer (almost) independent and (almost) multivariate normally distributed batch vectors

$$\bar{Z}_j(k) = b^{-1} \sum_{i=(j-1)b+1}^{jb} Z_i$$

for $j = 1, 2, \ldots, k$; $b = n/k$ is called the batch size, k the number of batches, and vector addition is component-by-component. We use the convention that any random variable with a bar and argument k is a batch mean of $b = n/k$ observations; for example, $\bar{Y}_j(k)$ is the jth batch mean of $\{Y_i\}$ with batch size $b = n/k$. The batch means are expressed as functions of k rather than b because the number of batches will be the primary factor of interest later. Throughout this paper, the total sample size n is fixed, although our results are useful when n is determined sequentially as discussed in the last section.

Given k batch means, the control-variate estimator of θ is

$$\hat{\theta}(k, q) = \bar{Y} - \hat{\beta}(k, q)'[\bar{X} - \mu], \tag{3}$$

where

$$\bar{Y} = k^{-1} \sum_{j=1}^{k} \bar{Y}_j(k) = n^{-1} \sum_{i=1}^{n} Y_i,$$

$$[\bar{X} - \mu] = k^{-1} \sum_{j=1}^{k} [\bar{X}_j(k) - \mu] = n^{-1} \sum_{i=1}^{n} [X_i - \mu],$$

and

$$\hat{\beta}(k, q) = \hat{\Sigma}_x(k, q)^{-1} \hat{\sigma}_{yx}(k, q). \tag{4}$$

The quantities on the right-hand side of (4) are the sample versions of $\Sigma_x(k, q) = \mathrm{Cov}[\bar{X}_j(k)]$ and $\sigma_{yx}(k, q) = \mathrm{Cov}[\bar{Y}_j(k), \bar{X}_j(k)]$; specifically,

$$\hat{\Sigma}_x(k, q) = (k-1)^{-1} \sum_{j=1}^{k} [\bar{X}_j(k) - \bar{X}][\bar{X}_j(k) - \bar{X}]'$$

and

$$\hat{\sigma}_{yx}(k, q) = (k-1)^{-1} \sum_{j=1}^{k} [\bar{X}_j(k) - \bar{X}][\bar{Y}_j(k) - \bar{Y}].$$

Lavenberg and Welch (1981) considered the special case when $k = n$ (no batching), and the $\{Z_i\}$ are independent $(q+1)$-variate normal vectors. They showed that $E[\hat{\theta}(n, q)] = \theta$ and

$$\mathrm{Var}[\hat{\theta}(n, q)] = (1 - R_{yx}^2)(n-2) / (n-q-2)(\sigma_y^2/n).$$

This compares to $\mathrm{Var}[\bar{Y}] = \sigma_y^2/n$, showing that a variance reduction (relative to $\bar{Y}$) can be achieved if $1 - R_{yx}^2 < (n-q-2)/(n-2)$, and emphasizing the need to keep the sample size, n, large with respect to the number of control variates, q. In addition, they showed that

$$\mathrm{Var}[\hat{\theta}(n, q)] / \mathrm{Var}[\theta^*(n, q)] = (n-2)/(n-q-2),$$

where $\theta^*(n, q)$ is the same as (3) except that $\hat{\beta}(n, q)$ is replaced by $\beta^*(n, q) = \Sigma_x^{-1}\sigma_{yx}$. This ratio quantifies the penalty for having to estimate the optimal control-variate multiplier $\beta^*(n, q)$.

Schmeiser (1982) considered the effects of batching when $q = 0$ (no control variates). He derived properties of the confidence interval for θ formed from different numbers of batch means when the total sample size n is fixed and the batch means are actually i.i.d. normal; performance measures included the expectation, standard deviation, and coefficient of variation of the half width of the confidence interval, and the probability that the interval covers points other than θ. Schmeiser found that these performance measures do not improve significantly for $k > 30$ batches, no matter how large n is. Of course, the batch means are *not* i.i.d. normal in general. However, results from batching i.i.d. normal data are relevant for batching dependent, nonnormal data in the following sense: output analysis based on batch means assumes that at some number of batches (equivalently batch size) the dependence and nonnormality of the batch means can be ignored, and that this remains true for smaller numbers of batches (equivalently larger batch sizes). That is, for small enough k the results for i.i.d. normal batch means hold, at least approximately. Schmeiser showed that there is little benefit from using a large number of batches even if the batch means remain i.i.d. normal for larger k. This is encouraging because at larger numbers of batches (equivalently smaller batch size) the dependence and nonnormality of the batch means may be significant. Thus, there is potential harm from using a large number of batches when the data is dependent, but little harm from using a small number of batches when the data is independent.

In this paper we consider the effect on variance reduction and confidence interval performance of simultaneously applying control variates and batching. The results of Lavenberg and Welch (1981), and Schmeiser (1982) are special cases of these results when the number of batches is fixed at n and when $q = 0$, respectively. We have two effects and their interaction to consider: (1) the effect of using additional control variates (increasing q) relative to the number of batches k, and (2) the effect of using fewer, larger batches (decreasing k) than necessary to achieve approximate $(q+1)$-variate normality and independence for $q \geq 0$ control variates.

The results in this paper are derived by treating the output process $\{Z_i\}$ as i.i.d. $(q+1)$-variate normal vectors; specifically, we make the following three assumptions.

(i) Initial transient effects have been removed, i.e., the output process Z_i is covariance stationary.

(ii) For an output sequence of length n, the dependence and nonnormality of the batch means is negligible for all values of k.

(iii) The problem of $b = n/k$ not being integer is insignificant.

Assumption (ii) is made for convenience of exposition. In Appendix A we show that the results below hold under a weaker assumption that for *some* number of batches (equivalently batch size) the batch means process is essentially an i.i.d. normal process. This is the standard assumption of batch means analysis. Thus, our results apply to the steady-state simulation problem that motivated the research.

We have been implicitly assuming that the simulation output process can be represented by Z_i, $i = 1, 2, \ldots, n$, as defined above. In some simulation experiments it may not be the case that each output Y_i is naturally associated with exactly q control variates, or that the output process has a discrete-time index. Prebatching, possibly by time rather than count, is one approach that yields an output process of the form considered here. For example, if we have a continuous-time process $Z(t)$, $0 \leq t \leq \tau$, then

$$\bar{Z}_j(k) = b^{-1}\int_{(j-1)b}^{jb} Z(t)\,dt,$$

where $b = \tau/k$, and τ is fixed rather than n. We make this explicit by adding the following assumption.

(iv) The simulation output process can be represented by the $(q+1)$-variate random vectors Z_i, $i = 1, 2, \ldots, n$.

For the control variates with which we are most familiar—i.i.d. sequences of input random variables—batching by time is effective. Regardless, to apply control variates and batching together we must first obtain a process of the form assumed in (iv).

For example, Añonuevo and Nelson (1988) simulated a machine-repair system consisting of seven machines that are subject to failure. When a machine fails, it receives either a major or minor repair and an inspection before returning to service. The parameter of interest, θ, was the long-run expected number of functioning machines. Thus, the output $Y(t)$, the number of functioning machines at time t, is a continuous-time index variable. The control variates were X_{1i}, the ith machine lifetime, X_{2i}, the time to complete the ith major repair, X_{3i}, the time to complete the ith minor repair, and X_{4i}, the time to inspect the ith repaired machine; they are all discrete-time index variables. The system was simulated for $\tau = 7400$ time units, discarding results from the first 1000 time units, and batching was done by time. Thus, a batch mean is the sample mean of all values realized during a time interval, for example, $k = 50$ batches of size $b = 128$ time units.

In the next section, results analogous to those of Lavenberg and Welch (1981) for the variance of the control-variate estimator are presented. In Sections 4 and 5 we quantify the effect of changing k and q on confidence interval half width and coverage probability, analogous to Schmeiser (1982). Section 6 discusses the implications of these results.

3. Variance

Let $\Sigma(k, q) = \mathrm{Cov}[\bar{Z}_j(k)]$, the covariance matrix of a batch-mean vector. Under assumptions (i)–(iv), $\Sigma(k, q)/k = \Sigma(1, q)$, where $\Sigma(1, q)$ is the covariance matrix of $\bar{Z} = [\bar{Y}, \bar{X}']'$. Thus, the square of the multiple correlation coefficient between $\bar{Y}_j(k)$ and $\bar{X}_j(k)$, which we denote by $R^2_{yx}[\Sigma(k, q)]$, is not a function of k (we will retain the argument k, however, as a reminder that we are batching).

Under assumptions (i)–(iv) we have the following results; Result 1 is immediate from Lavenberg and Welch (1981, equation (34)), and Results 2 and 3 follow from the above discussion. When we say that q is 'fixed', we mean that not only the number of control variates, but also the particular random variables chosen as control variates, are fixed. Notice that, unlike the case that Schmeiser examined where batching does not affect the point estimator, batching does affect the variance of the control-variate point estimator.

Result 1. *For fixed q and $q < k$,*

$$\mathrm{Var}[\hat{\theta}(k, q)] / \mathrm{Var}[\bar{Y}] = \left(1 - R^2_{yx}[\Sigma(k, q)]\right)(k-2)/(k-q-2).$$

Result 2. *For fixed q and $q < k_1 < k_2$,*

$$\mathrm{Var}[\hat{\theta}(k_1, q)] / \mathrm{Var}[\hat{\theta}(k_2, q)] = \frac{(k_1-2)(k_2-q-2)}{(k_2-2)(k_1-q-2)} > 1.$$

Clearly, increasing k decreases variance relative to $\mathrm{Var}[\bar{Y}]$. Result 2 quantifies the loss in variance reduction from using fewer, larger batches for fixed q. The loss is very little when $0 \leqslant q \leqslant 5$ and $30 \leqslant k_1 \leqslant 60$, no matter how large k_2 is.

Investigation of the variance of $\hat{\theta}(k, q)$ when q is varied is more difficult, since $R^2_{yx}[\Sigma(k, q)]$ changes not only as q changes but also with the particular control variates selected. Consider two different sets of control variates containing q_1 and q_2 control variates (q_1 may equal q_2). We denote the associated covariance matrices as $\Sigma(k, q_1)$ and $\Sigma(k, q_2)$, and the control-variate estimators as $\hat{\theta}(k, q_1)$ and $\hat{\theta}(k, q_2)$, respectively. Then we have the following result.

Result 3. $\mathrm{Var}[\hat{\theta}(k, q_2)] < \mathrm{Var}[\hat{\theta}(k, q_1)]$ *if and only if*

$$\frac{1 - R^2_{yx}[\Sigma(k, q_2)]}{1 - R^2_{yx}[\Sigma(k, q_1)]} < \frac{k - q_2 - 2}{k - q_1 - 2}.$$

Since $1 - R^2_{yx}[\Sigma(k, q)]$ is the fraction of variation in $\overline{Y}$ not explained (controlled) by the control variates,

$$\left(1 - R^2_{yx}[\Sigma(k, q_2)]\right)/\left(1 - R^2_{yx}[\Sigma(k, q_1)]\right)$$

is the ratio of the unexplained variation in $\overline{Y}$ using the second set of control variates to the unexplained variation using the first set of control variates. As a special case of Result 3, consider *adding* control variates to a fixed set of q_1 control variates (thus, $q_2 > q_1$). In this case, $R^2_{yx}[\Sigma(k, q)]$ is nondecreasing in q, so Result 3 gives a bound on the increase in $R^2_{yx}[\Sigma(k, q)]$ necessary to insure that adding control variates leads to a variance reduction. To be even more specific, consider adding a single control variate to a fixed set of q_1 control variates (thus, $q_2 = q_1 + 1$). Table 1 gives values of $(k - q_2 - 2)/(k - q_1 - 2)$ for various values of k and q_2. For $k \geq 30$, the ratio is stable in the range of 1 to 5 control variates, and is close to 1, meaning that the addition of another control variate is not likely to degrade (increase) the variance of $\hat{\theta}(k, q)$, while it may reduce the variance.

The inequality in Result 3 holds if and only if

$$\frac{R^2_{yx}[\Sigma(k, q_2)] - R^2_{yx}[\Sigma(k, q_1)]}{1 - R^2_{yx}[\Sigma(k, q_1)]} > \frac{q_2 - q_1}{k - q_1 - 2}.$$

In the special case that we are adding control variates and $q_2 = q_1 + 1$, the term on the left-hand side is the increment of explained variation from the $(q_1 + 1)$st control variate as a fraction of the remaining unexplained variation with q_1 control variates. This ratio may seem more natural in some contexts. We have choosen the ratio in Result 3 because $1 - R^2_{yx}[\Sigma(k, q)]$ is the central term in the variance of $\hat{\theta}(k, q)$. However, all results in this section and the next could be expressed in terms of the alternate ratio.

In the next two sections we examine the effect on confidence interval performance of varying both k and q.

Table 1
For fixed sample size n and q_1 control variates, the ratio $(k - q_2 - 2)/(k - q_1 - 2)$ for adding a control variate

k	$q_2 = q_1 + 1$				
	1	2	3	4	5
8	0.83	0.80	0.75	0.67	0.50
10	0.88	0.86	0.83	0.80	0.75
20	0.94	0.94	0.94	0.93	0.93
30	0.96	0.96	0.96	0.96	0.96
41	0.97	0.97	0.97	0.97	0.97
51	0.98	0.98	0.98	0.98	0.98
61	0.98	0.98	0.98	0.98	0.98
121	0.99	0.99	0.99	0.99	0.99
∞	1.00	1.00	1.00	1.00	1.00

4. Properties of the half width

When assumptions (i)–(iv) hold, it follows from Lavenberg and Welch (1981, Appendix A) that a $(1 - \alpha)100\%$ confidence interval for θ is $\hat{\theta}(k, q) \pm H(\alpha/2, k, q)$, where

$$H(\alpha/2, k, q) = t(\alpha/2, k - q - 1)\left(\widehat{\mathrm{Var}}[\hat{\theta}(k, q)]\right)^{1/2}, \quad (5)$$

$$\widehat{\mathrm{Var}}[\hat{\theta}(k, q)] = \hat{\sigma}^2(k, q)S,$$

$$\hat{\sigma}^2(k, q) = (k - q - 1)^{-1} \sum_{j=1}^{k} \left(\overline{Y}_j(k) - \hat{\theta}(k, q) - \hat{\beta}(k, q)'\left[\overline{X}_j(k) - \mu\right]\right)^2,$$

$$S = k^{-1} + (k - 1)^{-1}\left[\overline{X} - \mu\right]'\hat{\Sigma}_x(k, q)^{-1}\left[\overline{X} - \mu\right],$$

and $t(\alpha/2, d)$ is the $1 - (\alpha/2)$ quantile of the t distribution with d degrees of freedom.

The random variable $H(\cdot)$ is called the half width of the confidence interval. We are interested in properties of $H(\alpha/2, k, q)$ as k, q, and α vary, and specifically in $E[H(\alpha/2, k, q)]$, $(\mathrm{Var}[H(\alpha/2, k, q)])^{1/2}$, and $\mathrm{CV}[H(\alpha/2, k, q)]$ (the coefficient of variation). When the arguments of $H(\cdot)$ are obvious, these performance measures are abbreviated as $E[H]$, $\sqrt{\mathrm{Var}[H]}$, and $\mathrm{CV}[H]$, respectively.

Since the confidence interval formed under assumptions (i)–(iv) achieves the nominal probability of coverage $1 - \alpha$, the smaller the values of all three performance measures the better. The coefficient of variation is a particularly useful measure since it scales $\sqrt{\mathrm{Var}[H]}$ by $E[H]$, does not depend on α, and is dimensionless. Of course, $\mathrm{CV}[H]$ should not be considered without reference to $E[H]$. However, for the confidence interval considered here, both measures increase and decrease together, so the coefficient of variation provides

an appropriate measure of confidence interval stability. The results below are based on expressions derived in Appendix B.

4.1. Fixed q

First consider the case when the control variates are fixed but the number of batches varies, remembering that n is also fixed. The results can be summarized as follows.

(1) As k increases, all three performance measures decrease but at a decreasing rate; in other words, the gain from more batches decrease as the number of batches increases.

(2) For large q, significant decreases in the performance measures occur at larger values of k; in other words, as q increases having k large is more valuable.

(3) For small α, the rate of decrease of the performance measures with increasing k is slower; in other words, the benefit from k large is greater for small α.

These summary conclusions are similar to those of Schmeiser (1982) for $q = 0$, which is a special case. However, while Schmeiser found little additional benefit from $k > 30$ batches, this upper limit increases as more control variates are used. For example, Table 2 shows the effect of increasing k for $\alpha = 0.05$ and $q = 0, 1, 5$. The numbers of batches displayed were selected to match Schmeiser's (1982). The units on $E[H]$ and $\sqrt{\text{Var}[H]}$ are $\{1 - R^2_{yx}[\Sigma(k, q)])\sigma^2_y/n\}^{1/2}$. Thus, the units differ from column to column in the table (since q is different), but not within a column. Since the units differ, it is important to realize that the performance of $H(\cdot)$ may be significantly improved by using more or different control variates provided $R^2_{yx}[\Sigma(k, q)]$ significantly increases, but comparison between the columns does not show this. However, we can compare the relative changes as k varies.

For example, at $\alpha = 0.05$ the decreases in $E[H]$, $\sqrt{\text{Var}[H]}$, and CV$[H]$ when going from $k = 30$ to $k = 61$ are 2, 32, and 30%, respectively, when $q = 0$; they are 3, 33, and 31% when $q = 1$; but they are 7, 42, and 37% when $q = 5$ (these percentage changes are based on three decimal places, while the tables only show two decimal places). In light of Lavenberg and Welch (1981), it is not surprising that larger k is desirable when q is large; the tables quantify this for the performance of $H(\alpha/2, k, q)$.

We cannot directly compare $E[H]$ and $\sqrt{\text{Var}[H]}$ for different values of q, but can directly compare CV$[H]$ since it is dimensionless. In Figure 1, CV$[H]$ is plotted as a function of the number of batches, k. Each curve represents a different number of control variates, q. One way to interpret the curves is to find the equivalent number of batches required to achieve the same coefficient of variation for different numbers of control variates. For example, when $k = 8$ and $q = 0$, CV$[H] = 0.272$. To achieve the same coefficient of variation with $q = 1$ or 5 control variates requires $k = 10$ or 16 batches, respectively. Thus, the addition of 5 control variates requires twice as many batches. However, to achieve a CV$[H]$ of 0.09, which occurs when $k = 61$ and $q = 0$, requires only $k = 63$ or 70 batches when $q = 1$ or 5, respectively, a much smaller percentage increase.

Table 2
For fixed sample size n and q control variates, the effect of number of batches, k, on $H(\alpha/2, k, q)$ when $\alpha = 0.05$

k	$q = 0$				$q = 1$				$q = 5$			
	$t(\alpha/2, k-q-1)$	$E[H]$	$\sqrt{\text{Var}[H]}$	CV$[H]$	$t(\alpha/2, k-q-1)$	$E[H]$	$\sqrt{\text{Var}[H]}$	CV$[H]$	$t(\alpha/2, k-q-1)$	$E[H]$	$\sqrt{\text{Var}[H]}$	CV$[H]$
8	2.365	2.28	0.62	0.27	2.447	2.55	0.82	0.32	4.303	7.77	7.12	0.92
10	2.262	2.20	0.53	0.24	2.306	2.38	0.64	0.27	2.776	4.05	2.04	0.50
20	2.093	2.07	0.34	0.16	2.101	2.13	0.37	0.17	2.145	2.47	0.54	0.22
30	2.045	2.03	0.27	0.13	2.048	2.07	0.28	0.14	2.064	2.25	0.36	0.16
41	2.021	2.01	0.23	0.11	2.023	2.04	0.23	0.12	2.030	2.16	0.28	0.13
51	2.009	2.00	0.20	0.10	2.010	2.02	0.21	0.10	2.014	2.11	0.23	0.11
61	2.000	1.99	0.18	0.09	2.001	2.01	0.19	0.09	2.004	2.08	0.21	0.10
121	1.980	1.98	0.13	0.06	1.980	1.98	0.13	0.07	1.981	2.02	0.14	0.07
∞	1.960	1.96	0	0	1.960	1.96	0	0	1.960	1.96	0	0

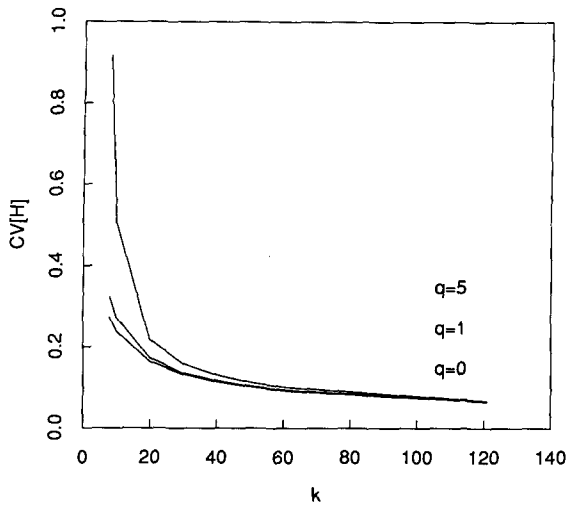

Figure 1. Comparison, by number of control variates q, of CV[$H(\alpha/2, k, q)$]

4.2. Fixed k

Consider adding control variates when the number of batches (equivalently the batch size) is fixed. Discussing the effect of varying q for fixed k is more difficult, since $R^2_{yx}[\Sigma(k, q)]$ changes as q changes; in fact, it is nondecreasing when we add control variates to a fixed set. Thus, we cannot directly compare the properties of the half width as q changes. However, we can make comparisons by considering

$$\big(1 - R^2_{yx}[\Sigma(k, q+1)]\big)/\big(1 - R^2_{yx}[\Sigma(k, q)]\big),$$

the ratio of the unexplained variation in $\overline{Y}$ after adding the $(q+1)$st control variate to the unexplained variation before adding it. This ratio, which we call the *marginal improvement ratio*, is always less than or equal to 1. However, if it is not enough less than 1, then confidence interval performance, in terms of the properties of the half width considered here, is degraded by adding the $(q+1)$st control variate. We let $r\{M\}$ be the breakeven point, or bound, such that

$$\frac{1 - R^2_{yx}[\Sigma(k, q+1)]}{1 - R^2_{yx}[\Sigma(k, q)]} \leqslant r\{M\}$$

implies that confidence interval performance is no worse after adding the $(q+1)$st control variate; performance is improved if the marginal improvement ratio is strictly less than $r\{M\}$. We let M stand for $E[H]$, or $\sqrt{\mathrm{Var}[H]}$, since the bound depends on the particular performance measure we consider. The results can be summarized as follows.

(1) As q increases, $r\{M\}$ decrease for all performance measures; in other words, the required improvement in $R^2_{yx}[\Sigma(k, q)]$ is greater for each additional control variate added to a fixed set.

(2) As k increases, $r\{M\}$ increases and stabilizes for all values of q; in other words, for large k the marginal improvement required for additional control variates is less and becomes constant.

(3) Varying α in the range 0.10, 0.05, and 0.01 does not significantly affect these results.

Table 3 shows $r\{M\}$ for $\alpha = 0.05$ and $k = 10$, 30 batches as $(q, q+1)$ goes from (0, 1) to (4, 5). When k reaches 30, $r\{M\}$ has stabilized and is close to 1, meaning it is easier to satisfy. However,

Table 4
For fixed sample size n and $k \geqslant 41$ batches, the marginal improvement ratio bound for adding a control variate

k	$r\{E[H]\}$	$r\{\sqrt{\mathrm{Var}[H]}\}$
41	0.97	0.92
51	0.98	0.94
61	0.98	0.95
121	0.99	0.98
∞	1.00	1.00

Table 3
For fixed sample size n and $k = 10$ batches, the marginal improvement ratio bound for adding a control variate when $\alpha = 0.05$

$(q, q+1)$	$k = 10$		$k = 30$	
	$r\{E[H]\}$	$r\{\sqrt{\mathrm{Var}[H]}\}$	$r\{E[H]\}$	$r\{\sqrt{\mathrm{Var}[H]}\}$
(0, 1)	0.86	0.66	0.96	0.92
(1, 2)	0.83	0.63	0.96	0.90
(2, 3)	0.80	0.60	0.96	0.89
(3, 4)	0.76	0.55	0.96	0.89
(4, 5)	0.69	0.47	0.96	0.89

comparing Table 3 to the $k = 30$ row in Table 1 shows that preserving confidence interval performance requires greater improvement in $R^2_{yx}[\Sigma(k, q)]$ (smaller marginal improvement ratio) than required to preserve point estimator variance. Table 4 summarizes the ratios for $k = 41$, 51, 61, 121, and ∞.

5. Probability of coverage

Under assumptions (i)–(iv),

$$p(\theta_1; \alpha, k, q) \equiv \Pr\{|\hat{\theta}(k, q) - \theta_1| \leq H(\alpha/2, k, q)\} = 1 - \alpha$$

for $\theta_1 = \theta$. In this section we examine the coverage function $p(\cdot)$ for $\theta_1 \neq \theta$. Ideally, $p(\theta_1; \alpha, k, q) = 0$ for $\theta_1 \neq \theta$, but since that is not possible we would like for $p(\cdot)$ to decrease rapidly as θ_1 gets farther away from θ. Let the deviation from θ be $\Delta = |\theta - \theta_1|$ in units of $\{(1 - R^2_{yx}[\Sigma(k, q)])\sigma_y^2/n\}^{1/2}$, since the coverage function $p(\cdot)$ is symmetric for positive and negative deviations from θ. The derivation of the results below is given in Appendix C.

5.1. Fixed q

For a fixed set of q control variates, $p(\theta_1; \alpha, k, q)$ is a decreasing function of k. Figures 2, 3, and 4 show the coverage function at various values of k for $q = 0$, 1, 5 and $\alpha = 0.05$. The units on Δ differ from figure to figure, but are the same for all curves in a figure. Again, since the units differ, it is important to realize that the performance of $p(\cdot)$ may be significantly improved by using more control variates provided $R^2_{yx}[\Sigma(k, q)]$ significantly increases, but comparison between the figures does not show this. However, we can compare the relative changes as k varies.

The sensitivity of the coverage function to k changes dramatically as q increases. For example, the difference between the $k = 30$ and $k = \infty$ curves when no control variates are used (Figure 2) is slight, as reported by Schmeiser (1982). However, when $q = 5$ control variates, the $k = 30$ and $k = 61$ curves differ by ≈ 0.1 for some values of Δ, and that much again from $k = \infty$. This means

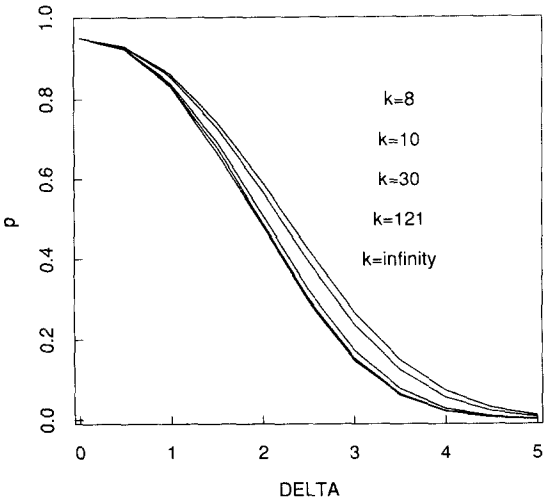

Figure 2. Comparison, by number of batches k, of $p(\theta + \Delta; \alpha, k, q)$ for $\alpha = 0.05$ and $q = 0$

that the probability that the confidence interval formed from 30 batches covers a value Δ distance from θ is 0.1 more than one formed from 60 batches for fixed n. Similar curves for $\alpha = 0.10$ and $\alpha = 0.01$ show the same changes in sensitivity, but slightly less or more dramatically, respectively.

5.2. Fixed k

For a fixed number of batches, we again look at the effect of adding control variates. Figures 5, 6, and 7 give $p(\cdot)$ for $k = 10$, 30, and 61, respectively, as q goes from 0 to 5 at the $\alpha = 0.05$ level.

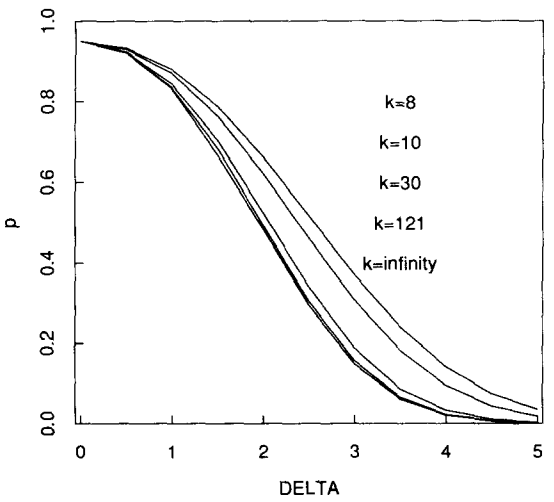

Figure 3. Comparison, by number of batches k, of $p(\theta + \Delta; \alpha, k, q)$ for $\alpha = 0.05$ and $q = 1$

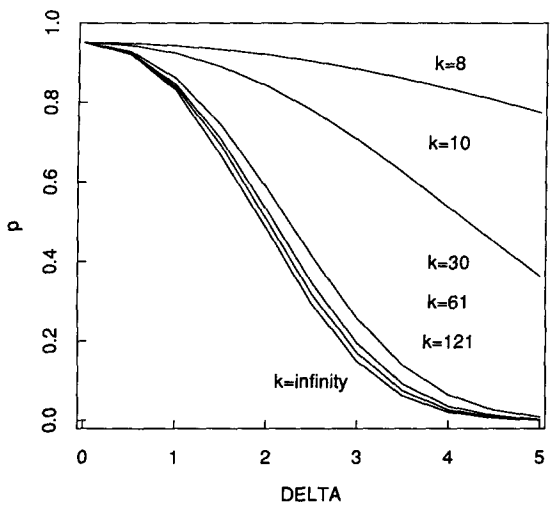

Figure 4. Comparison, by number of batches k, of $p(\theta + \Delta;\ \alpha,\ k,\ q)$ for $\alpha = 0.05$ and $q = 5$

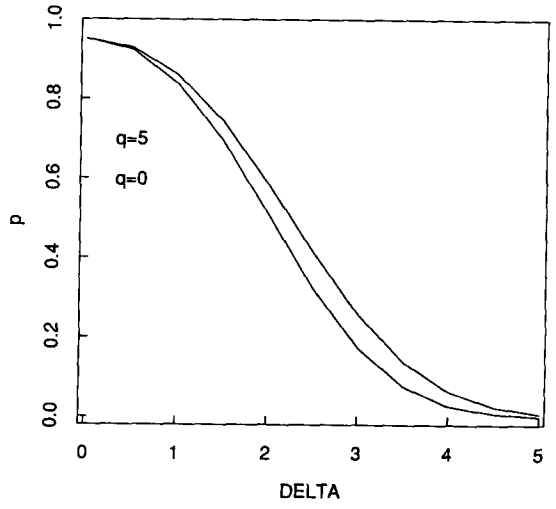

Figure 6. Comparison, by number of control variates q, of $p(\theta + \Delta;\ \alpha,\ k,\ q)$ for $\alpha = 0.05$ and $k = 30$

Since q changes for each curve in a figure, the units on each curve are different (recall that the units are $\{(1 - R^2_{yx}[\Sigma(k,\ q)])\sigma^2_y/n\}^{1/2}$, which changes with q). One way to interpret the curves is as the degradation in $p(\cdot)$ that would occur if the addition of another control variate resulted in only a negligible increase in $R^2_{yx}[\Sigma(k,\ q)]$.

As k increases, the coverage function becomes less sensitive to the number of control variates, even if the additional control variates contribute little. In fact, when $k = 61$, the curves for $q = 0$ and $q = 5$ are almost identical. However, since additional control variates can dramatically decrease $p(\cdot)$, another perspective on the figures is that as k increases, less is required of $R^2_{yx}[\Sigma(k,\ q)]$ to improve the coverage properties of the confidence interval. Similar behavior is observed for other values of α.

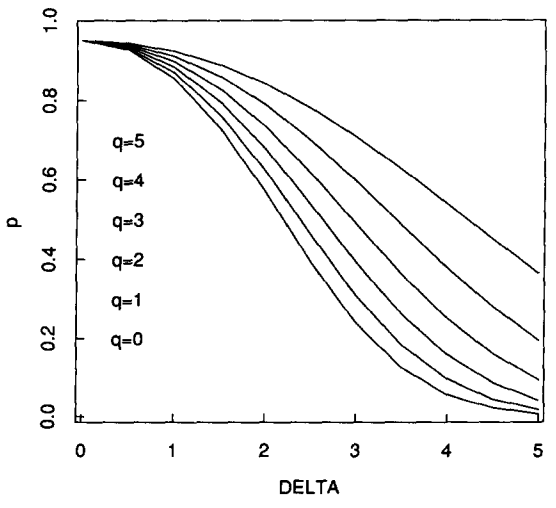

Figure 5. Comparison, by number of control variates q, of $p(\theta + \Delta;\ \alpha,\ k,\ q)$ for $\alpha = 0.05$ and $k = 10$

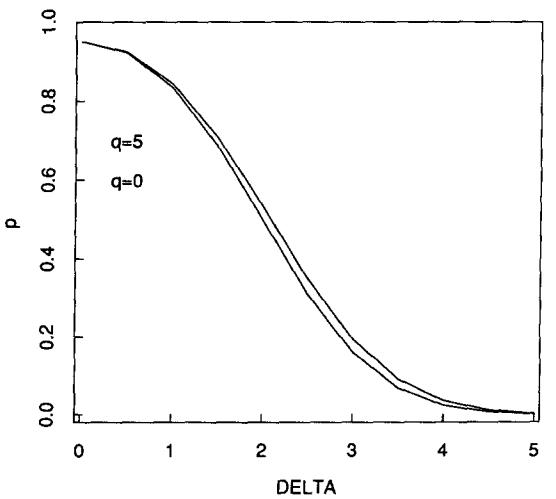

Figure 7. Comparison, by number of control variates q, of $p(\theta + \Delta;\ \alpha,\ k,\ q)$ for $\alpha = 0.05$ and $k = 61$

6. Implications

The results in Sections 3, 4, and 5 were derived under the assumption that the original output process $\{Z_i\}$ is an i.i.d. normal process. There are good reasons to batch a process that is already i.i.d., such as the outputs from replications in a finite-horizon experiment. The unbiasedness of the control-variate point estimator and the coverage probability of the control-variate confidence interval both depend on the normality of the output process. Even when the output process is i.i.d., it may not be normally distributed. Batching improves the approximation. Thus, point estimator bias can be reduced, and confidence interval coverage can be improved, by batching independent data. The results in this paper show that little is lost in estimator performance if 30 to 60 batches are used.

The results also apply to infinite-horizon experiments when independent realizations are employed and the total sampling budget, n, is fixed. In this case, the design decision is how to allocate the sampling budget to replications. To use the results above, each replication corresponds to a batch of size $b = n/k$, where k is the number of replications. Our results suggest dividing the budget into a modest number (30 to 60) of long replications. Since initial-transient effects are a concern, long replications are preferred from that standpoint as well.

The focus of this paper, however, is on the design of single-realization experiments. While Schmeiser (1982, p. 566) concluded that "$10 \leqslant k \leqslant 30$ is reasonable for most simulation situations", we modify those bounds to $10 \leqslant k \leqslant 60$ if from 1 to 5 control variates are employed, particularly if the number of control variates is greater than 1. As a general principle, the more control variates that are used the larger we would like k to be. Of course, k should not be so large that dependence and nonnormality of the batch means is significant, so $k = 60$ may not be possible. Our results do not account for the biases introduced when k is too large, but since the maximum allowable k is usually unknown (and ill-defined), the possibility that it is small is still more reason to use k no larger than necessary for good point and interval estimator performance. Our results also show that determining the maximum allowable k is not necessary in many cases, but rather determining that $k > 30$ or $k > 60$ is allowable is all that is important.

Ultimately, our findings should be incorporated into automated procedures for simulation output analysis and variance reduction. Typically, such procedures use statistical tests of hypothesis to determine if assumptions such as independence and normality are significantly violated (see, e.g., Chen and Seila, 1987). Because of the nature of statistical tests some erroneous decisions will be made, particularly since in many cases the hypotheses are known to be false, a priori. The smaller the number of batches k (equivalently, the larger the batch size b) the better the approximation of independence and normality, so an automated procedure should not insist on more batches than necessary to insure good performance of the point and interval estimators. In the range $30 \leqslant k \leqslant 60$, the properties of the half width, $H(\cdot)$, and the coverage function $p(\cdot)$, become increasingly less sensitive to k and q. The tests of hypothesis can be performed with values of k greater than 60, but even if a larger k is deemed acceptable further reduction of k can be made to gain more confidence in the approximation without sacrificing much in estimator performance.

Although there are frequently many potential control variates in a simulation experiment, it is our opinion that it is seldom possible to find more than five effective control variates. A subset of control variates may be selected from a larger set by choosing the ones that appear to make the most significant contribution to variance reduction. For example, Lavenberg, Moeller and Welch (1982), and Wilson and Pritsker (1984b) used forward selection and stepwise regression procedures, respectively, to select control variates in queueing network simulation. If batch means or independent realizations are used, our results suggest examining the estimated $R^2_{yx}[\Sigma(k, q)]$ values (computed automatically by most regression packages) for each fitted regression model and comparing them to the marginal improvement ratio bounds. As a rule of thumb, our results show that as the number of batches approaches 60, the penalty for adding even an ineffective control variate is slight, while the improvement from adding an effective control variate will be great. Thus, we recommend that an automated procedure initially search for an appropriate number of batches in the range $30 \leqslant k \leqslant 60$ to gain robustness from selection er-

rors. If at least 30 acceptable batches cannot be achieved, then serious consideration should be given to increasing the total sample size n. If the total sample size is increased, either to improve the independence and normality approximation or to achieve a confidence interval with prespecified half width, our results indicate that there is little reason to increase the number of batches beyond 60. Development of an algorithm based on these principles is reported by Añonuevo and Nelson (1988).

Acknowledgement

The author benefitted from discussions with John B. Neuhardt of The Ohio State University, Bruce Schmeiser, and James R. Wilson of Purdue University, and comments by two anonymous referees, one of whom suggested the alternate expression for Result 3.

Appendix A

The derivations below depend on assumptions (i)–(iv), which imply that Z_i, $i = 1, 2, \ldots, n$, is an i.i.d. $(q+1)$-variate normal process. However, assumption (ii) can be replaced by the weaker assumption (ii′), analogous to Schmeiser (1982):

(ii′) For an output process of length n, there exists a (usually unknown) number of batches $2 \leqslant k^* \leqslant n$ such that the batch means process $\bar{Z}_j(k)$, $j = 1, 2, \ldots, k$, is an i.i.d. $(q+1)$-variate normal process for any $k \leqslant k^*$.

The results stated above and derived below hold under assumption (ii′) when $k \leqslant k^*$. Of course, there is no k^* in general, but the central premise of batch means analysis is that for a small enough number of batches (equivalently, large enough batch size) the difference between the batch means process and an i.i.d. normal process, in terms of the properties of the point and interval estimators, is negligible.

To show that the derivations are unaffected by (ii′), recall that we defined $\Sigma(k, q) = \mathrm{Cov}[\bar{Z}_j(k)]$. In the special case when $\{Z_i\}$ is an i.i.d. process (assumption (ii)), $\Sigma(1, q) = n^{-1}\Sigma$. This relationship does not hold under (ii′) but, for $k \leqslant k^*$, $\Sigma(1, q) = \Sigma(k, q)/k$, since the batch means are independent. Thus, $R^2_{yx}[\Sigma(1, q)] = R^2_{yx}[\Sigma(k, q)]$ for $k \leqslant k^*$.

Results derived below under assumption (ii) also hold under (ii′) when $k \leqslant k^*$ provided they are stated in units of $\Sigma(1, q)$. Specifically, the units are

$$\left(1 - R^2_{yx}[\Sigma(k, q)]\right)\frac{\sigma^2_{\bar{Y}(k)}}{k} = \left(1 - R^2_{yx}[\Sigma(1, q)]\right)\sigma^2_{\bar{Y}},$$

where $\sigma^2_{\bar{Y}(k)} = \mathrm{Var}[\bar{Y}_j(k)]$. In the i.i.d. special case (ii), $\sigma^2_{\bar{Y}} = \sigma^2_y/n$. For simplicity, we derive the results under (ii), but a completely analogous derivation yields the same results under (ii′) for k restricted to be less than or equal to k^*.

Appendix B

Let $\{X = x\}$ represent the condition that $\{X_{1i} = x_{1i}, \ldots, X_{qi} = x_{qi}\}$, $i = 1, 2, \ldots, n$. The key to deriving properties of $H(\alpha/2, k, q)$ is that, under (i)–(iv), we can write $\mathrm{Var}[\hat{\theta}(k, q) \mid X = x] = (1 - R^2_{yx}[\Sigma(k, q)])(\sigma^2_{\bar{Y}(k)})s$, where

$$s = k^{-1} + (k-1)^{-1}[\bar{x} - \mu]'\hat{\Sigma}_x(k, q)^{-1}[\bar{x} - \mu]$$

(see Venkatraman and Wilson, 1986). Now the batch means $\bar{X}_j(k)$, $j = 1, 2, \ldots, k$, are i.i.d. q-variate normal vectors. Thus,

$$k[\bar{X} - \mu]'\hat{\Sigma}_x(k, q)^{-1}[\bar{X} - \mu] \sim T^2(k-1),$$

where $T^2(k-1)$ is a Hotelling T^2 random variable with $k-1$ degrees of freedom (Anderson, 1984). Using the relationship between the T^2 distribution and the F distribution, we immediately get $E[S] = (k-2)/[k(k-q-2)]$. These results are needed below.

Recall the definition of $H(\alpha/2, k, q)$ from equation (5). To compute $E[H]$ and $\mathrm{Var}[H]$ we first notice that

$$E\left[\hat{\sigma}^2(k, q) \mid X = x\right] = \left(1 - R^2_{yx}[\Sigma(k, q)]\right)\sigma^2_{\bar{Y}(k)}.$$

Also, conditional on $\{X = x\}$,

$$(k-q-1)\hat{\sigma}^2(k, q)/\sigma^2(k, q) \sim \chi^2(k-q-1),$$

where $\chi^2(k-q-1)$ is a chi-squared random variable with $k-q-1$ degrees of freedom (Rao, 1973). Then, analogous to Schmeiser (1982, Appendix A),

$$E\left[\hat{\sigma}(k, q) \mid X = x\right] = \sqrt{\frac{2}{k-q-1}}\,\frac{\Gamma((k-q)/2)}{\Gamma((k-q-1)/2)}\sigma(k, q),$$

where $\Gamma(\cdot)$ is the gamma function. Thus, in units of $\{(1 - R^2_{xy}[\Sigma(k, q)])\sigma_y^2/n\}^{1/2}$,

$$E\left[H(\alpha/2, k, q)\right] = t(\alpha/2, k-q-1)\sqrt{\frac{2k}{k-q-1}}\,\frac{\Gamma((k-q)/2)}{\Gamma((k-q-1)/2)}E[\sqrt{S}]. \tag{B.1}$$

Similarly, we have

$$E\left[H^2(\alpha/2, k, q)\right] = t^2(\alpha/2, k-q-1)((k-2)/(k-q-2))\left(1 - R^2_{yx}[\Sigma(k, q)]\right)\sigma_y^2/n.$$

Combining this with (B.1) gives

$$\sqrt{\mathrm{Var}[H(\alpha/2, k, q)]} = t(\alpha/2, k-q-1)\left(\frac{k-2}{k-q-2} - \frac{2k}{k-q-1}\,\frac{\Gamma^2((k-q)/2)}{\Gamma^2((k-q-1)/2)}E^2[\sqrt{S}]\right)^{1/2} \tag{B.2}$$

in units of $\{(1 - R^2_{yx}[\Sigma(k, q)])\sigma_y^2/n\}^{1/2}$.

The only term in (B.1) and (B.2) that could not be derived analytically as a function of k and q is $E[\sqrt{S}]$. We estimated $E[\sqrt{S}]$ via numerical integration by expressing $S = 1/k + T^2(k-1)/(k(k-1)) = 1/(k(1-B))$, where B is a random variable having a beta distribution and parameters $q/2$ and $(k-q)/2$. Then,

$$E[\sqrt{S}] = \frac{\gamma}{k}\int_0^1 (1-b)^{-1/2} b^{(q-2)/2}(1-b)^{(k-q-2)/2}\,\mathrm{d}b,$$

where γ is the normalizing constant for the beta distribution. The IMSL function DCADRE with relative error and absolute error set at 0.00001 was used to perform the integration on an IBM 3081-D computer. Double precision arithmetic was used to calculate all results presented in the tables, including the evaluation of $\Gamma(\cdot)$. A check on the coding of the numerical integration was provided by Monte Carlo sampling based on 200 000 realizations of $T^2(k-1)$ for each value of k and q. The IMSL subroutine GGBTR was used to generate beta variates that were transformed into T^2 variates, and the same starting seed was used for each estimate. The standard error of all the estimates was less than 0.001.

Appendix C

Under assumptions (i)–(iv) and conditional on $\{X = x\}$, we have

$$\Pr\left\{ \left|\hat{\theta}(k, q) - \theta\right| / \widehat{\mathrm{Var}}\left[\hat{\theta}(k, q)\right]^{1/2} < t(\alpha/2, k-q-1)\right\} = 1 - \alpha$$

(see Lavenberg and Welch, 1981, Appendix A). Since the right-hand side does not depend on X, the random variable in the expression has a t distribution with $k-q-1$ degrees of freedom unconditionally as well.

We are interested in

$$\Pr\left\{\left|\hat{\theta}(k, q)-\theta_1\right| < H(\alpha/2, k, q)\right\} = \Pr\left\{\frac{\left|\left(\hat{\theta}(k, q)-\theta\right)-\left(\theta_1-\theta\right)\right|}{\widehat{\operatorname{Var}}\left[\hat{\theta}(k, q)\right]^{1/2}} < t(\alpha/2, k-q-1)\right\}.$$

The random variable in the expression has a noncentral t distribution with noncentrality parameter

$$\delta = (\theta-\theta_1)/\operatorname{Var}\left[\hat{\theta}(k, q)\right]^{1/2} = (\theta-\theta_1)\sqrt{(k-q-2)/(k-2)}$$

in units of $\{(1-R_{yx}^2[\Sigma(k, q)])\sigma_y^2/n\}^{1/2}$. The IMSL subroutine MDTN was used to compute the noncentral t probabilities $p(\cdot)$ for various values of k, q, α, and $\Delta = \theta-\theta_1$.

References

Anderson, T.W. (1984), *An Introduction to Multivariate Statistical Analysis*, Wiley, New York.

Añonuevo, R., and Nelson, B.L. (1988), "Automated estimation and variance reduction via control variates for infinite-horizon simulations", *Computers & Operations Research* 15 (5), 447–456.

Chen, D.R., and Seila, A.F. (1987), "Multivariate inference in stationary simulation using batch means", *Proceedings of the Winter Simulation Conference*, 302–304.

Crane, M.A., and Lemoine, A.J. (1977), *An Introduction to the Regenerative Method of Simulation Analysis*, Lecture Notes in Control and Information Sciences, Springer, Berlin.

Fishman, G.S. (1978), "Grouping observations in digital simulation", *Management Science* 24, 510–521.

Heidelberger, P., and Welch, P.D. (1981), "A spectral method for confidence interval generation and run length control in simulations", *Communications of the ACM* 24 (4), 233–245.

Iglehart, D.L., and Lewis, P.A.W. (1979), "Regenerative simulation with internal controls", *Journal of the ACM* 26 (2), 271–282.

Lavenberg, S.S., Moeller, T.L., and Welch, P.D. (1982), "Statistical results on control variables with application to queueing network simulation", *Operations Research* 30 (1), 182–202.

Lavenberg, S.S., and Welch, P.D. (1981), "A perspective on the use of control variables to increase the efficiency of Monte Carlo simulations", *Management Science* 27 (3), 322–335.

Law, A.M., and Carson, J.S. (1979), "A sequential procedure for determining the length of a steady state simulation", *Operations Research* 27, 1011–1025.

Mechanic, H., and McKay, W. (1966), "Confidence intervals for averages of dependent data in simulations II", Technical Report ASDD 17-202, IBM Corporation, Yorktown Heights, NY.

Meketon, M.S., and Schmeiser, B.W. (1984), "Overlapping batch means: Something for nothing?", *Proceedings of the Winter Simulation Conference*, 227–230.

Nelson, B.L. (1987), "A perspective on variance reduction in simulation experiments", *Communications in Statistics* B16 (2), 385–426.

Nelson, B.L., and Schmeiser, B.W. (1986), "Decomposition of some well-known variance reduction techniques", *Journal of Statistical Computation and Simulation* 23, 183–209.

Rao, C.R. (1973), *Linear Statistical Inference and Its Applications*, Wiley, New York.

Schmeiser, B. (1982), "Batch size effects in the analysis of simulation output", *Operations Research* 30 (3), 556–568.

Schriber, T.J., and Andrews, R.W. (1979), "Interactive analysis of simulation output by the method of batch means", *Proceedings of the Winter Simulation Conference*, 513–525.

Schriber, T.J., and Andrews, R.W. (1984), "ARMA-based confidence intervals for simulation output analysis", *American Journal of Mathematical and Management Sciences* 4 (3 & 4), 345–375.

Schruben, L. (1983), "Confidence interval estimation using standardized time series", *Operations Research* 31 (6), 1090–1108.

Sharon, A.P., and Nelson, B.L. (1988), "Analytic and external control variates for queueing network simulation", *Journal of the Operational Research Society* 39 (6), 595–602.

Venkatraman, S., and Wilson, J.R. (1986), "The efficiency of control variates in multiresponse simulation", *Operations Research Letters* 5 (1), 37–42.

Wilson, J.R. (1984), "Variance reduction techniques for digital simulation", *American Journal of Mathematical and Management Sciences* 4 (3 & 4), 277–312.

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

Wilson, J.R., and Pritsker, A.A.B. (1984b), "Experimental evaluation of variance reduction techniques for queueing simulation using generalized concomitant variables", *Management Science* 30 (12), 1459–1472.