Variance reduction techniques: Experimental comparison and analysis for single systems

Ihsan Sabuncuoglu , Mehmet Murat Fadiloglu & Sabri Çelik

Variance reduction techniques: Experimental comparison and analysis for single systems

IHSAN SABUNCUOGLU[1,*], MEHMET MURAT FADILOGLU[1] and SABRI ÇELIK[2]

[1] *Department of Industrial Engineering, Bilkent University, Ankara 06800, Turkey*
E-mail: sabun@bilkent.edu.tr
[2] *Department of Industrial Engineering and Operations Research, Columbia University, New York, NY 10027, USA*

We provide a thorough analysis of the effectiveness of different Variance Reduction Techniques (VRTs). We consider both stand-alone and combined applications of two input techniques, Antithetic Variates (AV) and Latin Hypercube Sampling (LHS), and two output techniques, Control Variates (CV) and Poststratified Sampling (PS). Previous research in the area mainly focuses on asymptotic variance reduction. In this experimental study, we measure the performance of VRTs under finite simulation run lengths and analyze their effects. Our findings show that the asymptotic variance reduction results do not readily apply to finite-length simulations. We consider three different types of systems (M/M/1, serial production line and (s, S) inventory control systems) and compare the VRTs under various experimental conditions. We observe that a variance reduction cannot be guaranteed for every instance a VRT is applied. Our results also indicate that the output VRTs (CV, PS) are better than input VRTs (AV, LHS) on the average for the single systems considered in this study. More interestingly, the less-sophisticated techniques (AV, CV) often perform better than the relatively more-complex techniques (LHS, PS). A comprehensive bibliography is also provided.

Keywords: Simulation, variance reduction techniques, antithetic variates, Latin hypercube sampling, control variates, poststratified sampling

1. Introduction

Simulation is used extensively in analyzing complex real-life systems (e.g., distribution, production and communication systems). For many situations, it is the only available analysis technique (Law and Kelton, 2000). Many simulation models are extremely large, highly complex, require extensive development times and use a large amount of computer resources to run. Since running these models is expensive, it is important that their output is properly analyzed. In general, simulation output is stochastic in nature due to random input elements, and proper analysis requires the use of appropriate statistical methods (i.e., constructing confidence intervals, determining the number of simulation runs, selecting the best system, etc.).

A typical characteristic of the problems studied by simulation is that a variety of system configurations are examined under various conditions, thus requiring numerous cases to be run. Due to the large number of cases to be analyzed in a limited amount of time, the minimum number of simulation runs should be taken to achieve the desired precision. Furthermore, even though simulation is tradi-

tionally used to deal with long-term strategic decision problems (e.g., design problems, etc.), recent simulation applications have been directed to analyze short-term operational problems (e.g., real-time scheduling and dispatching problems). In these daily applications, the simulation cycle is even shorter due to the need for a quick response to frequently changing operating conditions.

Variance Reduction Techniques (VRTs) are able to decrease the simulation time by reducing the number of simulation runs required to achieve a specified estimation precision. Specifically, VRTs are developed to reduce the variance in output from simulation models. This may be especially useful for complex systems for which a reduction in the long computational times required would be significant. However, the potential benefits of VRTs are not easy to realize, because implementing VRTs is usually difficult and requires considerable advance planning for large simulation models. Analysts need to carefully consider how to build these capabilities into their models at the development stage, since their subsequent incorporation into the model can be extremely difficult (if not impossible).

Note that the theoretical results on variance reduction provided by VRTs are often of an asymptotic nature. However, all simulation studies use finite simulation runs and lengths. Thus, the effectiveness of VRTs under finite

---

*Corresponding author

simulation conditions needs to be investigated. One of the objectives of our paper is to address this issue.

There are essentially two main types of VRTs in the literature. The first type is useful for the analysis of single systems whereas the other type handles multiple systems. The purpose of simulation studies for single systems is to estimate certain performance measures for the given system under a given configuration as accurately as possible. In contrast, in multiple system studies, the emphasis is on the relative performance of two or more systems or configurations of the same system. The most popular techniques for single systems include: (i) Antithetic Variates (AV); (ii) Control Variates (CV); (iii) indirect estimation; (iv) stratified sampling; and (v) conditional sampling. The only widely used VRT of the second type is Common Random Numbers (CRN) (see Law and Kelton 2000, p. 582). In general, the techniques use one or more of the following strategies for reducing variances: (i) induce positive correlation; (ii) induce negative correlation; or (iii) control randomness.

In this paper, we consider the VRTs for single systems (i.e., the first category), as they constitute a logical starting point. Specifically, we analyze the performance of four VRTs: (i) AV; (ii) CV; (iii) Latin Hypercube Sampling (LHS); and (iv) Poststratified Sampling (PS). These are the most popular methods in the literature. Their distinct natures require their effectiveness on different settings to be examined individually.

We apply these techniques individually (stand-alone applications) as well as in combined (hybrid) applications. In the latter case, we aim at understanding their interactions in various problem domains. Specifically, we apply these four VRTs to: (i) an M/M/1 queueing system; (ii) a serial line production system; and (iii) an inventory system so that their performances in different system configurations can be measured in order to learn more about their effectiveness in reducing variance.

The rest of the paper is organized as follows. We give a brief overview of the literature in Section 2. The VRTs considered herein are discussed in Section 3. This is followed by a discussion on experimental settings in Section 4. The simulation results are presented in Section 5 for stand-alone VRTs and in Section 6 for combined VRTs. Concluding remarks and future research directions are given in Section 7.

2. Literature review

The existing research work on VRTs can be summarized in two categories: (i) review papers on VRTs and their classifications; and (ii) comparative studies to test some of the VRTs under various experimental conditions. A comprehensive list of these studies is given in Table 1.

The work in the first category provides a general guideline for users to select the appropriate VRT. It also describes the characteristics of the VRTs and their relationships. For example, L'Ecuyer (1994) presents an overview of the various techniques in the VRT literature by giving a number of examples; he discusses CRN, AV, CV, importance sampling, indirect estimators, stratification, LHS, conditioning, descriptive sampling, hybrid methods and virtual measures. In another survey, Nelson and Schmeiser (1983) develop a classification framework by which VRTs are discussed in this context. According to the authors, VRTs transform the simulation models into related models in which the simulation yields more-precise estimates of the parameters of interest. These transformations modify the inputs of a simulation model through distribution replacement or dependence induction. They also modify the statistics through auxiliary information or equivalent information. This basic transformation idea was extended to make a taxonomy of VRTs by Nelson (1990).

In the second category, we find analytical and empirical studies to compare VRTs. Glynn and Whitt (1989) investigate the asymptotic efficiency of estimators for average queue length and average waiting time measures. They show, for instance, that estimating the average queue length indirectly via Little's law is more efficient than doing so directly. In another study, Carson and Law (1980) focus on the efficient estimation of the mean delay in a queue, mean time-in-system, average number in queue and in system for GI/G/s queueing systems. Minh (1989) develops a partial conditional expectation technique. Sullivan *et al.* (1982) investigate the efficiency of AV for estimating the expected completion time of a stochastic activity network. Their results indicate that AV can produce the same precision as independent simulations but with approximately one-quarter of the computational effort. In another study, Avramidis and Wilson (1996) propose multiple sample quantile estimators based on AV and LHS. The results of their simulation experiments indicate that the proposed methods yield significant reduction in bias and variance.

Wilson and Pritsker (1984) conduct an empirical study in which CV and PS are used separately. Their results indicate that for analytically tractable models of closed and mixed machine repair systems, postratification produces variance reductions of between 10 and 40%, whereas CV achieves a variance reduction of between 20 and 90%. Ross (2001) also demonstrates how certain VRTs can be used efficiently in the analysis of queueing models. He considers three techniques: dynamic stratified sampling, utilization of multiple control variates and the replacement of random variables by their conditional expectations.

There are also integrated VRT application studies in the literature. Schruben and Margolin (1978) give conditions under which the AV and CRN techniques produce guaranteed efficiency improvements (c.f. Song and Chiu (2007)). Kleijnen (1975) combines AV and CRN to compare two alternative systems. The results indicate that the combined technique can be inferior to AV and to CRN in this particular application. He also develops a new combined scheme, which is superior to the stand-alone application of VRTs.

Table 1. Summary of the VRT literature

Authors	Contribution	CRN techniques proposed/tested	Systems studied
L'Ecuyer (1994)	Survey	—	—
Nelson (1987)	Survey	—	—
Glynn and Whitt (1989)	Comparison	Indirect vs. direct estimation	Queueing systems; specifically GI/G/s
Carson and Law (1980)	Comparison	Indirect vs. direct estimation	Queueing systems; specifically GI/G/s
Minh (1989)	Comparison	Partial conditional expectation	—
Sullivan *et al.* (1982)	Comparison	AV	Stochastic activity networks
Avramidis and Wilson (1998)	Comparison	AV and LHS	Stochastic activity networks
Wilson and Pritsker (1984)	Comparison	CV and PS	Queueing systems
Wilson and Pritsker (1994)	Comparison	CV and PS	Queueing systems
Ross (2001)	Comparison	Dynamic stratified sampling, multiple CV and conditional expectation	Queueing systems
Schruben and Margolin (1978)	Comparison	Integration of AV and CRN	—
Kleijnen (1975)	Comparison	Integration of AV and CRN	—
Yang and Liou (1996)	Comparison	Integration of AV and CV	—
Kwon and Tew (1994)	Comparison	Integration of AV and CV	Stochastic activity networks
Burt and Gaver (1970)	Comparison	Integration of AV and CV	—
Nelson (1990)	Comparison	Integration of AV and CV	Inventory system and M/M/1 queue
Tew and Wilson (1994)	Comparison	Integration of AV with CV and CRN	—
Avramidis and Wilson (1996)	Comparison	Integration of conditional expectation, AV, LHS and CV	Stochastic activity networks

In another study, Yang and Liou (1996) show that the combined use of CV and AV yields a smaller variance than the conventional control variate estimator applied without AV. Kwon and Tew (1994) present three methods to combine AV and CV. Burt and Gaver (1970) combine AV and CV and obtain better results than either method applied individually.

Nelson (1990) analyzes the efficiency of CV and AV in improving the performance of point and interval estimators when initial bias is present. Tew and Wilson (1994) incorporate CV into an AV and CRN scheme and investigate the conditions under which the combination scheme performs better than AV, CRN and CV used on their own, and direct simulation.

Avramidis and Wilson (1996) examine all combinations of conditional expectation, correlation induction (AV and LHS) and CV. They derive sufficient conditions under which this strategy yields a smaller variance. In their simulation experiments with stochastic activity networks, an integrated technique of conditional expectation and LHS performs better than other studied techniques.

In summary, even though the VRT literature is quite rich, the integrated methods are not adequately studied and most work is devoted to specific models and specific VRTs. In this paper, we examine the relative performances of four VRTs (AV, LH, CV and PS) in terms of their hybrid as well as their stand-alone applications. Comparisons are provided for three systems: (i) M/M/1 queue; (ii) a serial production line; and (iii) (s, S) inventory system. Our aim is to make a thorough study testing the selected VRTs and their combined applications under various settings. The study should be useful to researchers and practitioners who are confronted with the choice of a VRT to apply to their finite-length simulation.

3. VRTs

We consider four different VRTs in this comparative study. These techniques are AV, LHS, CV and PS. Although there are other VRTs available, these are certainly the most prominent techniques that can be applied in any simulation study without any model-specific structural knowledge. Such knowledge is required in other techniques such as variance reduction by conditioning, stratified sampling, and importance sampling. Such a requirement prohibits generic application, limiting the use of these techniques. In addition, CRN is not included in our study as it is only applicable to multiple systems. We classify the VRTs into two categories.

1. *Input techniques*: These techniques induce negative correlations among output random variables in simulation runs by using negatively correlated input random variables. AV and LHS fall into this category. These techniques require manipulation of the input streams used in the simulation.

2. *Output techniques*: These techniques employ auxiliary variables in an attempt to correct the output variables. CV and PS fall into this category. These techniques do not require any change in the way the simulation experiment is performed. They only modify the output obtained from the simulation experiment. In using these techniques one needs to provide some information about the distribution of the auxiliary variables.

Even though each one of these techniques can be applied on their own, it is also possible to combine input techniques with output techniques. The combinations result in what are called combined techniques. In this study, we consider four individual and two combined VRTs. The combination of PS with input techniques is not considered because the implementation requirements of PS make this combination impossible (as discussed later). Before delving into details of the experimental study, we first provide a short description of how each technique works.

3.1. *AV*

When using the AV technique, instead of taking statistically independent runs, we introduce a statistical dependence. Namely, for each independent run, we also take an antithetic run using random numbers that are complementary with the ones in the other run. Thus, if U_k is a particular uniform $(0,1)$ random number used for a particular purpose in the first run, then $1 - U_k$ is used for the same purpose in the antithetic run. Thus, in these two runs we use two streams of random numbers that are perfectly negatively correlated. The outputs of these two runs are expected to be negatively correlated given that the output random variable is a monotone function of the numbers in the input stream and synchronization is achieved (Ross, 1990). In order to illustrate the monotonicity of an output random variable as a function of the input stream, consider an exponential random variable Y with a mean of one that is generated via the inverse transform method, $Y = -\log U$, using a sequence of uniform random numbers, $U[0, 1]$. Notice that Y (the output random variable) is a monotone decreasing function of U (the input stream).

Let us call the output variable obtained from the first run X_1, and the one obtained from the antithetic run X_2. Since U_k and $1 - U_k$ are both uniform between zero and one, the expected values of both X_1 and X_2 are the same and they are equal to what we want to estimate, i.e., $E[X] = E[X_1] = E[X_2]$. Now we use $(X_1 + X_2)/2$ as our estimator of $E[X]$.

Clearly this is an unbiased estimator. Moreover, since:

$$\mathrm{Var}\left(\frac{X_1 + X_2}{2}\right) = \frac{1}{4}[\mathrm{Var}(X_1) + \mathrm{Var}(X_2) + 2\,\mathrm{Cov}(X_1, X_2)]$$
$$= \frac{1}{2}[\mathrm{Var}(X) + \mathrm{Cov}(X_1, X_2)],$$

the variance of this estimator is less than that of the estimator obtained by averaging independent observations, given that the covariance between X_1 and X_2 is negative.

In the application of AV, we perform n independent macro-replications, each consisting of two correlated micro-replications. Thus, we have a total of $2n$ replications that are pairwise negatively correlated. Given that the monotonicity assumption holds, the variance of our estimator is less than the variance of the independent runs case.

3.2. *LHS*

LHS is based on the idea of inducing a negative correlation among input random number streams used for different simulation replications. While in AV each macro-replication consists of two correlated micro-replications, in LHS the number of micro-replications is a parameter (k) that can be selected. The interval $[0, 1]$ is partitioned into k equal length subintervals. Each of the k random numbers—used for the same purpose across the micro-replications—is mapped to a different subinterval selected at random. This creates a negative correlation among the input streams within a macro-replication. The average of the k output variables from micro-replications corresponding to a macro-replication is used as the estimator (McKay *et al.*, 1979). Furthermore, given that the monotonicity assumption holds, the negative correlation between input random number streams gives rise to a negative correlation between the output variables of the micro-replications. The expressions given for AV in the previous section are also valid for LHS with $k = 2$. The only difference between AV and LHS with $k = 2$ is the way the negative correlation is induced among the input streams within macro-replications. However, the correlation inducement technique used in LHS generalizes to macro-replications with more than two micro-replications. Therefore, a variance reduction is expected just as in the case of AV, but this method is computationally more expensive than AV. This is due to the fact that random numbers have to be assigned to different subintervals using a random permutation, which slows the simulation run.

3.3. *CV*

In the CV technique, the simulation experiment is performed as usual. There is no intervention in the way the input random numbers are generated. The only additional requirement in the experiment is that one needs to collect

statistics for an additional random variable—along with the output variable—which is called the control variate. The control variate, which has to be correlated with the output variable, is used to correct the output variable in such a way that the corrected variable shows less variation around the estimated value.

Let us call the output variable X, the control variate Y and the mean value of the control variate μ_Y. The corrected output variable is obtained using $X_c = X - a(Y - \mu_Y)$. This is again an unbiased estimator of $E[X]$. The variance of the corrected variable is minimized when $a = \text{Cov}(X, Y)/\text{Var}(Y)$ (Ross, 1990). Since $\text{Cov}(X, Y)$ is almost never known beforehand, it needs to be estimated during the simulation. The corrected statistics (the statistics corresponding to corrected output variables) are calculated after the value of a is determined. Then, the corrected statistics are averaged and a confidence interval is constructed accordingly. It is possible to correct the output variable further by using more than one control variate (Yang and Nelson, 1992).

3.4. *PS*

The PS strategy, which is an output technique first proposed as a VRT by Wilson and Pritsker (1984)—even though Sethi (1963) discussed the underlying idea before—also makes use of an auxiliary variable that is called the stratification variate. However, compared to CV, more information is required about the auxiliary variable; one needs to know its distribution exactly.

In PS, the range of the stratification variate is partitioned into L strata. The strata lengths are not usually equal. In fact, one way to construct the strata is to partition the range into equal-probability intervals. Another scheme for the case of a normal stratification variate is suggested by Sethi (1963). In the current study, we experiment with both schemes.

Let N_h denote the number of stratification variates that fall in the hth stratum, $h = 1, 2, \ldots, L$, X_{hj}, $1 \le j \le N_h$, denote the response variables for those replications whose stratification variates fall in the hth stratum, and denote by π_h the probability that the stratification variate is in the hth stratum. The poststratified estimator of $E[X]$ is

$$\overline{X}_{ps} = \sum_{h=1}^{L} \pi_h \left(\frac{1}{N_h} \sum_{j=1}^{N_h} X_{hj} \right).$$

This is again an unbiased estimator and the variance of this estimator can be estimated using:

$$\hat{\text{Var}}(\overline{X}_{ps}) = \sum_{h=1}^{L} \frac{1}{N_h} \pi_h^2 S_{yh}^2,$$

where S_{yh}^2 is the sample variance for the response variables for those replications whose stratification variates fall in the hth stratum. In order to construct an interval estimator one

also needs the degrees of freedom of the variance estimator, which is approximately (Wilson and Pritsker, 1984);

$$\upsilon_e = \frac{\left(\sum_{k=1}^{L} (1/N_h) \pi_h^2 S_{yh}^2 \right)^2}{\sum_{k=1}^{L} (\pi_h^4 S_{yh}^4 / N_h^2 (N_h + 1))} - 2.$$

Although both PS and CV are output techniques, their implementations differ considerably. In CV, we average independent corrected output variables, whereas in PS we use the formula above to obtain $\overline{X}_{ps}$. This is not an average in the usual sense and its variance estimator has to be calculated in a different fashion. All other techniques use the classical interval estimator (as in raw simulation) based on averaging. The implementation of the PS technique is somewhat more complicated.

3.5. *Combined techniques*

It is possible to combine one of the input techniques with one of the output techniques with the prospect of obtaining an additional variance reduction. The input techniques create micro-replication groups whose output variables are expected to be negatively correlated. Then the output variables of the micro-replication groups are averaged to come up with the macro-replication outputs. The macro-replication outputs are statistically independent.

In order to combine an input technique with an output technique, the output technique can be applied either before or after this averaging. If it is applied before, then the output of each micro-replication is corrected using the auxiliary variable for that replication and then the corrected variables are averaged to obtain the macro-replication output. This is the first combination scheme. The second combination scheme starts with averaging to obtain the macro-replications output as in the case of the stand-alone application of the input method. The method continues by correcting the macro-replication output making use of the macro-replication auxiliary variable, which is the average of the micro-replication auxiliary variables. Yang and Liou (1996) propose a third scheme specifically for AV + CV that uses corrected micro-replication outputs without averaging. Pilot runs indicate that the second scheme, which is also the most straightforward one, works the best. In this paper we only report the results of this second scheme, which involves the application of the input technique first, and then the output technique.

Since we are considering two input and two output techniques in this study, we can obtain four different combinations. The first two combinations are AV + CV and LHS + CV, on which we report in this study. The other two possible combinations are AV + PS and LHS + PS, which are not feasible. For the first combination scheme, it is not possible to obtain the exact distribution of the average of correlated micro-replication variables, which is needed for the application of PS. The second combination scheme is
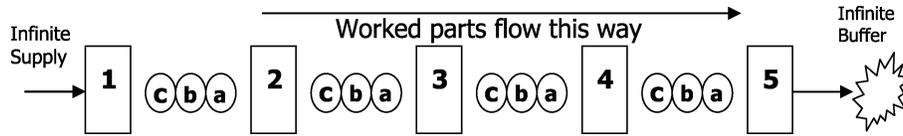
Fig. 1. Serial production line.

not considered either, since PS needs many statistically independent output statistics, which do not exist in a micro-replication. Thus, PS cannot be readily applied in conjunction with any other VRT.

4. Experimental setting

We use three classical discrete-event simulation models to perform a comparative analysis of the effectiveness of the VRTs we presented in the previous section. Our intention is to observe what kind of variance reductions those VRTs actually yield when applied to these simulation models. We also want to see if one can advocate the application of certain techniques over others or if their relative performances depend on the model at hand. The three models we consider are all well-studied, classical simulation models: (i) the M/M/1 queue; (ii) a serial production line; (iii) and an inventory system. Since most simulation studies in the literature relate to either inventory or queueing models, the results we obtain from these three models should be applicable to many simulation settings.

First, we consider the simple M/M/1 model with two traffic intensities (ρ), 0.5 and 0.9, by setting the service rate $\mu = 10$ and the arrival rates $\lambda = 5$ and $\lambda = 9$, respectively. We simulate this M/M/1 model for 410 000 entities in total; however, we discard the first 10 000 entities as the transient period at both congestion levels. Even though a much shorter run length is enough to estimate a first moment, we need to take longer simulation runs to estimate higher moments. The output variable is the time-in-system parameter for the entities.

Second, we consider a model for a serial production line system consisting of five workstations with limited buffers between stations. The stations are coupled with each other by "blockage" and "starvation" phenomena. The serial line production system is illustrated in Fig. 1, where the rectangles denote the stations, with the circles corresponding to the buffer spaces between those stations.

In our experiments, we assume a balanced line of five stations such that all stations have the same processing time distribution—a lognormal distribution with a mean of one and a standard deviation of 0.3. Three buffer spaces are available between subsequent stations. The output variable is again the time-in-system parameter for the parts. We calculate the corresponding statistics for 40 000 parts after discarding the first 800 parts, since this line with limited part capacity quickly reaches steady-state operation.

Finally, we consider an inventory system utilizing an (s, S) policy. We assume weekly demands having a normal distribution with mean 19.23 and standard deviation 5.658. The orders will arrive at the beginning of each week with a constant lead time of 2 weeks. The maximum inventory level and the reorder point are determined to be 150 and 20, respectively. If the inventory on hand drops under 20 in week 7, say 15, an order of $150 - 15 = 135$ is placed. Then this order arrives at the beginning of week 9. The inventory on hand is selected as the output variable. We discard the first 800 weeks and calculate statistics for the next 52 000 weeks (1000 years) of the system operation.

For each of these systems, interval estimates for the mean of the output variable are calculated using 60 replications. This procedure is repeated applying the discussed individual and combined VRTs, which means that all techniques are compared with the same number of replications. When input techniques are used, the 60 replications are divided into correlated micro-replication groups. Since the improvements in variance estimators and half-lengths vary when the same procedure is repeated with different random number seeds, improvement statistics are calculated by repeating the entire procedure, consisting of the application of the VRTs on 60 replications, for ten times in each case. By calculating interval estimates for the improvements, based on the ten super-replications, the effectiveness of each VRT is assessed.

In each super-replication, the half-length for the performance measure of interest using 60 independent replications and the corresponding half-lengths with VRTs are obtained. The percent improvement for the super-replication is calculated for each VRT with respect to the independent replications case. The average and the standard deviation for the ten super-replications are calculated and the corresponding half-lengths on the percent improvements due to VRTs are reported in tables.

5. Computational results of stand-alone VRT applications

5.1. *AVs*

For the M/M/1 model, we consider three cases. We induce negative correlation on: (i) service times; (ii) interarrival times; and (iii) both interarrival and service times. As the standard implementation of AV, we use U values in odd-numbered replications and $1 - U$ values in even-numbered

Table 2. Half-length improvements with AV

		Average (%)	Std. dev. (%)	Lower limit (%)	Upper limit (%)	Half-length (%)
AV applied to interarrival times for the M/M/1 system	$\rho = 0.9$	11.9	3.1	4.9	19.0	7.0
	$\rho = 0.5$	8.6	3.7	0.1	17.0	8.5
AV applied to service times for the M/M/1 system	$\rho = 0.9$	9.3	5.1	−2.3	20.8	11.6
	$\rho = 0.5$	10.5	4.8	−0.4	21.4	10.9
AV applied to service & interarrival times for the M/M/1 system	$\rho = 0.9$	14.0	3.8	5.5	22.5	8.5
	$\rho = 0.5$	15.5	2.3	10.2	20.7	5.2
AV applied to the serial line system		7.2	4.8	−3.7	18.1	10.9
AV applied to the inventory system		10.7	4.1	1.5	19.9	9.2

replications. In order to achieve full synchronization, we dedicate a different random number stream to each input variable. This also prevents correlation of the output random variable across even and odd-numbered replications.

The results of the ten different simulation experiments (each consisting of 60 replications) are summarized in Table 2. The second and third columns represent the average and the standard deviation of half-length improvements due to the application of AV over the independent replications case (i.e., no VRT case). Columns 4 through 6 report the interval estimates of half-length improvements. The other tables in the manuscript follow the same format.

The results in Table 2 indicate that the application of AV to both interarrival and service times outperforms its application solely to service times and to interarrival times regardless of the utilization level, due to the higher negative correlation induced by the two input variables. Negative improvements are also observed in some of the ten different simulation experiments when AV is applied to service times. However, overall the average half-length improvement (or reduction) is positive. In theory, the variance reduction is guaranteed in an asymptotic sense (Ross, 1990). In our case, however, we take a finite number of samples from the simulation model (even though the run length is 400 000 observations and 60 independent runs are taken at each design point). Thus, in practice one should not be surprised when an improvement is not achieved in some VRT applications. Our experiments indicate that the success rate (i.e., the number of times the half-length is reduced due to AV) is more than 90% when AV is applied to the two input variables simultaneously.

In the serial line case, we apply AV to the service times of all five stations simultaneously. In general, AV provides more than a 7% improvement in the half-length. However, this improvement is not significant because the confidence interval of the half-length includes zero. The success rate of AV turns out to be 80% (i.e., positive improvements are observed in eight out of ten confidence intervals) in our experiments. This indicates that a variance reduction cannot

be guaranteed at every instance AV is applied. However, in order to obtain a precise estimate for the success rate, one would need to use more than ten samples.

For the inventory system, we apply AV to the demand variable. The average improvement is 10.7% for the half-length. Note that the results are statistically significant.

5.2. *LHS*

We consider $k = 2$ and 3 stratification levels for all three systems. We do not run the system for $k = 4$ or beyond, since pilot runs indicated no improvement, yet additional computational burden. Similarly to AV, we consider three cases for the M/M/1 system. We induce negative correlation among: (i) service times; (ii) interarrival times; and (iii) both interarrival and service times. The results, listed in Table 3, indicate that the use of LHS yields a significant improvement in the half-length at the high utilization rate (0.9) when applied to both interarrival and service times. In the other cases, even though the average half-length reductions are usually positive, the corresponding confidence intervals contain zero.

For the M/M/1 model, we note that LHS is more effective when applied to both input random variables as in the case of AV. A significant difference in half-length improvements is not observed for the $k = 2$ and $k = 3$ cases. Once again there is no guarantee that the variance is reduced by LHS. Moreover, the probability of success is much smaller with LHS than with AV. Thus, we do not generally recommend LHS for the M/M/1 system unless it is applied to both input variables in cases with a high utilization rate.

We have similar results for the serial production line. The results indicate that $k = 2$ and $k = 3$ stratification levels do not differ from each other in terms of the half-length reduction. LHS with $k = 2$ fails in two instances out of ten whereas LHS with $k = 3$ fails in three instances, although this difference is not statistically significant. One should note that in the case of $k = 3$, part of the variance reduction

Table 3. Half-length improvements with LHS

		k	Average (%)	Std. dev. (%)	Lower limit (%)	Upper limit (%)	Half-length (%)
LHS applied to interarrival times in the M/M/1 system	$\rho = 0.9$	2	0.5	5.2	−11.1	12.2	11.7
	$\rho = 0.5$		−3.6	3.6	−11.8	4.6	8.2
	$\rho = 0.9$	3	9.2	4.5	−0.9	19.4	10.2
	$\rho = 0.5$		2.8	5.5	−9.6	15.2	12.4
LHS applied to service times in the M/M/1 system	$\rho = 0.9$	2	2.9	4.1	−6.3	12.1	9.2
	$\rho = 0.5$		−1.6	4.7	−12.1	9.0	10.5
	$\rho = 0.9$	3	2.4	5.4	−9.8	14.6	12.2
	$\rho = 0.5$		5.5	3.5	−2.3	13.3	7.8
LHS applied to interarrival & service times in the M/M/1 system	$\rho = 0.9$	2	15.5	3.4	7.9	23.2	7.7
	$\rho = 0.5$		9.0	4.8	−1.8	19.8	10.8
	$\rho = 0.9$	3	15.4	5.1	3.8	26.9	11.6
	$\rho = 0.5$		8.3	4.6	−2.0	18.6	10.3
LHS applied to the serial line system		2	6.6	4.3	−3.2	16.4	9.8
		3	7.0	4.4	−2.9	16.8	9.8
LHS applied to the inventory system		2	6.1	4.1	−3.2	15.4	9.3
		3	12.2	5.9	−1.1	25.6	13.3

improvement is lost due to the decrease in the degrees of freedom compared to the $k = 2$ case.

For the inventory system, increasing the stratification level positively contributes to the reduction of the half-length. However, these improvements are not statistically significant. In conclusion, based on our set of experiments, we do not recommend LHS for this type of inventory problem.

5.3. *CVs*

We use the average service time and the average interarrival time as the control variates for the M/M/1 system. We report the improvements when these two variates are used individually and together. Unlike the previous two input techniques (AV and LHS), we have positive improvements in all of the ten experiments. It seems that CV (as an output technique) performs better than these output techniques by achieving greater variance reductions (Table 4).

The performance of CV depends on the correlation between the selected control variate and the output random variable. Note that the correlation is larger when the utilization rate is smaller if the service time is used as the control variate. This can be explained by queueing theory. In general, the time-in-system is the sum of two components: service time and time-in-queue. In the M/M/1, service time constitutes only 10% of the time-in-system for the high-loaded system whereas it constitutes 50% for the low-loaded case. For that reason, service time as a control variate performs well at the low-utilization rate. Similarly, using interarrival times works better at lower utilization levels than at higher levels. However, the service times are in general more effective than the interarrival times as control variates. This is due to the fact that there is a higher correlation between service times and time-in-system compared to the correlation between interarrival times and time-in-system. When both control variates are used simultaneously, the improvement is even more pronounced. Thus, we conclude that all control variates should be used simultaneously to achieve more variance reduction. This strategy, i.e., using all possible control variates, works considerably better than the input techniques.

In the serial line, there are five input variables (service times for each station), which are the candidates for the control variates. In Table 4, we present the results on the half-length improvements created by the application of CV when each service time is used as the control variate one at a time and then all the service times are used together. Results indicate that the use of different service times as the control variate yields different improvements. Since the middle station is the critical resource in a serial line due to the "bowl" phenomenon, the control variates associated with the middle stations yield more variance reduction than those from the other stations. Note that CV should be applied to all the service times since the amount of variance reduction is the greatest (28.4–36.2%). Finally, when CV is compared to AV and LHS, CV provides greater half-length improvements.

For the (s, S) inventory system, the weekly demand variable is used as the control variate. The results indicate that CV consistently provides significant improvements in the half-length. A comparison of CV with AV and LHS reveals the fact that CV gives relatively larger lower limits for the half-lengths. Thus, CV should be the preferred method.

Table 4. Half-length improvements with CV

	Control variates		Average (%)	Std. dev. (%)	Lower limit (%)	Upper limit (%)	Half-length (%)
M/M/1 system	Interarrival times	$\rho = 0.9$	12.5	1.4	9.2	15.8	3.2
		$\rho = 0.5$	5.8	2.1	−0.4	12.0	6.2
	Service times	$\rho = 0.9$	17.0	2.2	12.0	21.9	4.9
		$\rho = 0.5$	32.0	2.2	37.0	27.0	5.0
	Interarrival & service times	$\rho = 0.9$	29.4	2.1	24.9	34.1	4.6
		$\rho = 0.5$	37.2	1.9	32.9	41.5	4.3
Serial line system	Service 1		4.3	1.7	0.4	8.3	3.9
	Service 2		5.8	1.3	2.9	8.8	3.0
	Service 3		8.9	2.5	3.3	14.6	5.6
	Service 4		7.8	2.8	1.6	14.0	6.2
	Service 5		4.9	2.3	−0.2	10.1	5.2
	All		32.3	1.7	28.4	36.2	3.9
Inventory system	Demand		12.3	2.2	10.7	13.8	1.5

5.4. *PS*

We consider four different stratification levels, $L = 2, 3, 4,$ and 5. In order to make strata in PS, there are two possibilities: (i) equal probability strata; and (ii) the allocation scheme by Sethi (1963). When a steady-state simulation is conducted, the average of the input variables can be approximated by a normal distribution due to the central limit theorem. Therefore, the strata allocation scheme suggested by Sethi for a normal stratification variate is expected to be effective in steady-state simulations. In the M/M/1 system, we use the sample average service time as the stratification variate.

We present the results of the allocation scheme by Sethi and the equal probability scheme at two different utilizations, corresponding to each stratification level (Table 5). The results indicate that PS provides improvements in the ranges 18.9–26.2% and 10.8–14.0% for low and high utilizations, respectively. As the number of strata increases, the improvement generally tends to increase. Note that the service time as the stratification variate performs much better at the low utilization due to the reasons discussed in the CV case. Examining the results for various strata, PS with $L = 5$ provides the best improvement at the low utilization whereas $L = 4$ achieves the best performance at the high utilization rate. Thus, $L = 4$ and $L = 5$ are recommended for the M/M/1 case. Since PS does not require much computational effort, simulation output data can easily be analyzed at different stratification levels and the best performing stratum can be recommended for that particular application. We also observe that construction of the strata is as important a factor in the resulting improvement as the selection of the stratification variate. Results indicate that PS is the second best VRT at low utilizations, following CV; its average improvements are better than AV and LHS. At high utilizations, even though PS provides better average improvements, these are not statistically significant.

In the serial line system, we select the variable that has the highest correlation with the throughput. Unlike the case of CV, we can choose only one variable as the stratification variate in PS. Since the bottleneck in the serial line determines the throughput of the system when the service times of the stations are equal, the critical variable for PS is the service time of the third station (i.e., the middle station). The results of the simulation experiments indicate that none of the confidence intervals include zero (i.e., PS provides positive improvements in the precision and the variance). Recall that in the M/M/1 case, CV outperforms PS when the same random variable is used as both the control and the stratification variate. The same observation is also made in this case when the service time of the third station is used as both the control and the stratification variate. Note that CV can employ more than one random variate, whereas PS can only employ a single variate. We observe that CV using all five average service times as control variates outperforms PS.

For the inventory system, we employ the average weekly demand as the stratification variate. Clearly, weekly demand has a strong correlation with the Inventory On Hand (IOH). This is confirmed by our experimental results. The correlation between the weekly demand and the IOH is −0.47, whereas it is −0.28 in the backlogged demand case. Thus, our choice of employing the weekly demand to stratify the IOH values is well justified. The simulation results given in Table 5 indicate that Sethi's allocation scheme is marginally better than the equal probability scheme. Examining the results of the two schemes, we see that Sethi's allocation scheme produced the best result (10.2% for $L = 5$). Confidence intervals for the average improvements include zero for $L = 2$ regardless of the stratification scheme. However, lower limits are larger than zero with higher stratification levels. This means that PS is expected to provide improvements in terms of half-length. Of course, once again,

Table 5. Half-length improvements with PS

	L	Sethi's scheme					Equal probability scheme				
		Average (%)	Std. dev. (%)	Lower limit (%)	Upper limit (%)	Half length (%)	Average (%)	Std. dev. (%)	Lower limit (%)	Upper limit (%)	Half length (%)
M/M/1 system ρ = 0.5	2	18.9	2.8	12.4	25.3	6.4	18.9	2.8	12.4	25.3	6.4
	3	22.6	2.6	16.7	28.5	5.9	21.4	2.6	15.6	27.2	5.8
	4	26.0	2.8	19.8	32.2	6.2	25.1	2.3	19.9	30.3	5.2
	5	26.2	2.1	21.4	31.0	4.8	26.4	2.8	20.3	32.6	6.2
ρ = 0.9	2	10.9	2.7	4.9	16.8	6.0	10.9	2.7	4.9	16.8	6.0
	3	10.8	2.9	4.3	17.3	6.5	11.4	2.9	4.8	18.0	6.6
	4	14.0	2.4	8.6	19.4	5.4	12.6	2.8	6.3	18.9	6.3
	5	13.8	2.7	7.7	20.0	6.1	13.8	2.3	8.7	19.0	5.2
Serial line system	2	6.0	2.4	0.6	11.4	5.4	6.0	2.4	0.6	11.4	5.4
	3	6.3	2.6	0.5	12.2	5.8	6.9	2.3	1.7	12.2	5.2
	4	7.2	2.5	1.5	12.8	5.6	6.8	2.5	1.1	12.5	5.7
	5	7.1	2.2	2.0	12.1	5.1	6.4	2.5	0.8	11.9	5.5
Inventory system	2	6.2	3.2	−1.1	13.5	7.3	6.2	3.2	−1.1	13.5	7.3
	3	8.3	2.4	2.9	13.6	5.4	8.7	2.6	2.8	14.5	5.9
	4	9.6	2.8	3.3	15.9	6.3	9.0	2.5	3.3	14.8	5.7
	5	10.2	2.1	5.5	14.9	4.7	8.9	2.6	3.0	14.8	5.9

Table 6. Half-length improvements with AV + CV

	Control variates		Average (%)	Std. dev. (%)	Lower limit (%)	Upper limit (%)	Half-length (%)
M/M/1 system	Interarrival times	$\rho = 0.9$	17.8	3.6	9.7	25.9	8.1
		$\rho = 0.5$	16.1	2.4	11.5	22.3	5.4
	Service times	$\rho = 0.9$	24.8	2.9	18.2	31.4	6.6
		$\rho = 0.5$	34.8	3.0	28.0	41.7	6.9
	Interarrival & service times	$\rho = 0.9$	33.0	5.4	21.0	45.1	12.1
		$\rho = 0.5$	35.4	2.8	29.0	41.9	6.4
Serial line system	Service 1		10.7	4.6	0.4	21.0	10.3
	Service 2		11.0	5.0	−0.2	22.2	11.2
	Service 3		11.6	4.4	1.7	21.5	9.9
	Service 4		10.1	4.2	0.7	19.5	9.4
	Service 5		12.4	4.0	3.3	21.5	9.1
	All		23.6	4.6	13.1	34.1	10.5
Inventory system	Demand		12.8	4.1	22.0	3.7	9.2

this improvement cannot be guaranteed. Similar to the previous cases (the M/M/1 and the serial line systems), we observe better performance of CV over PS. On the other hand, PS yields positive improvements ranging over a larger spectrum than CV.

6. Experimental results for combined techniques

In this section, we consider combined applications of input and output VRTs. The issue we are seeking to address here is whether combined use brings further improvement over the stand-alone applications.

6.1. *AV + CV*

The half-length improvements in Table 6 are for the M/M/1 system, when AV is applied to both interarrival and service times and when the average interarrival and service times are used as control variates in CV. We report results on the use of these two control variates individually and simultaneously.

Clearly, AV + CV (using the second combination scheme as discussed in Section 3.5) outperforms the individual applications of the four considered VRTs. Moreover, it is again better to use both control variates simultaneously. The effect of using both variates is especially pronounced at higher utilizations. One should note that the combined scheme works better than the two stand-alone techniques independent of the utilization level and irrespective of the control variable used. Thus, it is best to use the combined scheme and use as many control variates as possible in the M/M/1 model.

In the serial production line system, AV is applied to each service time simultaneously, while CV uses as the control variate the service time of choice or all service times together. We observe that combining AV with CV degrades the improvement in terms of half-length compared to stand-

alone application of CV. CV alone produces an improvement of 32.3% on the average whereas its combination with AV produces 23.6%. In contrast, when CV is applied to service times individually, combining CV with AV with the second scheme contributes positively to the improvements. Thus, the combination of AV and CV performs worse than the stand-alone application of CV. It is possible to explain this phenomenon within a framework we propose. First, the total gain in the combination of AV and CV is equal to the gain obtained with the stand-alone application of AV plus the additional gain obtained with the application of CV on top of AV, loosely speaking. It is possible to express this as follows:

$$Total\ Gain\ of\ AV + CV = Gain\,(AV) + Gain\,(CV|AV).$$

Our observations during this study all suggest that *Gain* $(CV) > Gain\,(CV|AV) > 0$. This means that CV produces more variance reduction when applied to unmodified independent output variables. When it is applied to modified output variables (macro-replication outputs obtained by AV), it provides a smaller additional variance reduction, although this reduction is still positive. This suggests that the combined use performs better than the stand-alone use of CV, if the variance reduction due to AV is greater than the loss in the reduction of CV due to its interaction with AV, i.e., *Gain* $(AV) > Gain(CV) - Gain(CV|AV)$. In the M/M/1 system, using the average service time as the control variate, the *Gain(CV|AV)* values are $24.8 - 14.0 = 10.8\%$ and $34.8 - 15.5 = 19.3\%$ for the 0.9 and 0.5 utilizations, respectively. Hence, *Gain* $(CV) = 17.0\% > 10.8\%$ for the 0.9 utilization system and *Gain* $(CV) = 32.0\% > 19.4\%$ for the 0.5 utilization. Thus, we would expect that the combination performs better, irrespective of the used control variate. Nevertheless, in the serial line case, *Gain(CV|AV)* $= 32.3 - 7.6 = 24.7\%$ is larger than *Gain* $(AV) = 7.6\%$. Therefore, the combination performs worse than CV alone.

Table 7. Half-length improvements with LHS + CV

	k	Control variates		Average (%)	Std. dev. (%)	Lower limit (%)	Upper limit (%)	Half-length (%)
M/M/1 system	2	Interarrival & service times	$\rho = 0.9$	22.0	3.2	14.8	29.3	7.2
			$\rho = 0.5$	28.8	3.5	20.9	36.7	7.9
	3	Interarrival & service times	$\rho = 0.9$	21.4	4.7	10.8	32.1	10.7
			$\rho = 0.5$	29.3	2.7	23.3	35.3	6.0
Serial line system	2	Service 1		11.0	4.0	2.0	20.0	9.0
		Service 2		16.8	4.6	6.5	27.2	10.4
		Service 3		14.3	4.1	5.0	23.6	9.3
		Service 4		16.2	3.5	8.3	24.1	7.9
		Service 5		10.4	4.0	1.5	19.4	9.0
		All		19.1	4.3	9.3	28.8	9.8
	3	Service 1		10.7	4.9	−0.3	21.8	11.1
		Service 2		13.5	4.6	3.1	23.9	10.4
		Service 3		14.9	4.9	3.9	25.9	11.0
		Service 4		13.6	3.8	5.1	22.1	8.5
		Service 5		13.4	4.0	4.5	22.4	9.0
		All		21.9	4.7	11.4	32.4	10.5
Inventory system	2	Demand		15.1	3.9	6.3	23.9	8.8
	3	Demand		23.6	3.3	16.2	31.1	7.5

For the inventory system, the second combination scheme cannot be applied. This is due to the fact that the only input variable for the model (weekly demand), which has a symmetric distribution, has to be the control variate. When AV is applied before CV (the second combination scheme), the macro-replication averages for the control variate are always zero (due to synchronization and the symmetry of the demand distribution). This prevents the use of CV to correct the output variables. Hence, we apply the first combination scheme for the inventory system. The simulation results indicate that the scheme is successful in achieving variance reduction. Comparing the results with the AV and CV applied individually, it is obvious that the combined use of these two techniques contributes to the average improvement in half-length. This observation is in line with our previous discussion.

6.2. *LHS + CV*

We combine CV and LHS by correcting with CV the macro-replication outputs obtained using LHS. Since we obtain the best results with the stand-alone application of LHS when we apply it to both interarrival and service times, we also use this setting for the combined application. We apply LHS to both variables and stratify input variables into $k = 2, 3$ stratification levels in the M/M/1 system. We take the service time as the control variate. The results for both utilizations are presented in Table 7.

The results indicate that the use of the service time as the control variate and induction of negative correlation among the replications by LHS performs almost as well as AV + CV. A careful examination of the results indicates that the differences between the average improvements are signifi-

cant at both the two and the three micro-replications cases. Furthermore, lower limits for the average improvements are far from zero so this combination produces a considerable reduction in the variance. At low utilization, this combination results in smaller improvements than the stand-alone application of CV. Increasing the micro-replication number does not significantly affect the improvement. Interestingly, this combination provides more consistent improvements at low utilization whereas the opposite is true at high utilization values. Comparing the average improvements of the combined policies with the single application of LHS indicates the superiority of the combination. All of these findings are again in agreement with our explanations in the previous section.

In the serial line case, we apply LHS to all service times at $k = 2, 3$ and consider using different control variates during the application of CV. Since all lower limits are greater than zero, LHS + CV provides an improvement in terms of half-length. Average improvements indicate that applying CV with LHS produces poorer results than CV alone when all service times are used as control variates. This is again due to the fact that for this case the stand-alone application of CV results in a very large improvement and the introduction of LHS seems to degrade this improvement obtained via CV. We conclude that any combined application yields improved results in single control variate cases, while it gives inferior improvements when all service times are used as control variates.

For the inventory system, the only input variable, demand, is assigned to be the control variate and stratified on $k = 2$ and $k = 3$ levels. According to the results, this combination appears to be very effective in reducing the variance and the results at $k = 3$ are especially commendable. We also

observe that increasing the number of micro-replications significantly contributes to the performance in terms of the average improvements, lower limits and half-lengths. Thus, the additional improvement resulting from LHS is significant.

7. Discussion and conclusions

In this paper, we analyze four variance reduction techniques (AV, LHS, CV and PS) for three systems (M/M/1, serial production line inventory systems). From our extensive simulation experiments we make the following conclusions.

As stated in the literature (Law and Kelton, 2000), VRTs cannot guarantee variance reduction in each simulation application. In our simulation experiments, we observe that the amount of variance reduction by VRTs can vary substantially from one simulation run to another run for each system considered. In some cases, we even observe that VRTs can backfire. The variance reductions promised by VRTs can only be guaranteed as the number of simulation runs tends to infinity.

Among the stand-alone applications of the four VRTs, CV stands out as the best technique. This is followed by PS, AV and LHS, even though there are some settings in which this order changes. In the experiments we conducted, we found that output techniques (CV and PS) outperformed input techniques (AV, LHS). In addition, output techniques do not require any change in the way the simulation experiment is conducted—they only modify output obtained from the simulation experiments. Thus, their application is much easier compared to input techniques, which require controlling randomness and achieving synchronization. In our simulation experiments, we also observe that the performance of input techniques deteriorates as systems become more complex. The negative correlation induced among input variables often produces a less-significant correlation among output variables when filtered through complex systems.

Among the output VRTs, CV is better than PS when the same output variable is used as an auxiliary variable in both techniques. Moreover, CV has an inherent advantage since it can be used with more than one control variate, resulting in a greater potential for variance reduction. Among the input techniques we recommend AV over LHS because AV usually produces more improvement and it is also much simpler to implement than LHS.

Even though the stand-alone applications of VRTs (especially CV) provide significant variance reduction, further improvement is possible with their combined (hybrid) application. These additional variance reductions may be important for certain practical applications when simulation runs are expensive. Our experiences with the combined techniques suggest that the AV + CV combination is the best in terms of both variance reduction and ease of application. It is also interesting to observe that, in our examples, this combined VRT yields superior results compared

to other more complicated and computationally more expensive techniques.

The results presented in this paper are valid under the current experimental conditions for the selected four VRTs. There is a need for further research along the following directions. First, this study can be extended to multiple system comparison where CRN would be considered in combined VRT applications. Second, one would also look into the possibility of combining PS with input techniques. Third, we believe that it is more beneficial to concentrate future research efforts on output techniques rather than input techniques.

References

Avramidis, A.N. and Wilson, J.R. (1996) Integrated variance reduction strategies for simulation. *Operations Research,* **44**, 327–346.

Avramidis, A.N. and Wilson, J.R. (1998) Correlation induction techniques for estimating quantiles in simulation experiments. *Operations Research,* **46**, 574–591.

Burt, J.M. and Gaver, D.P. (1970) Simple stochastic networks: some problems and procedures. *Naval Research Logistics Quarterly,* **17**, 439–459.

Carson, J.S. and Law, A.W. (1980) Conservation equations and variance reduction in queueing simulations. *Operations Research,* **28**, 535–546.

Glynn, P.W. and Whitt, W. (1989) Indirect estimation of $L = \lambda W$. *Operations Research,* **37**, 82–103.

Kleijnen, J.P.C. (1975) Antithetic variates, common random numbers and optimal computer time allocation in simulation. *Management Science,* **21**, 1176–1185.

Kwon, C. and Tew, J.D. (1994) Strategies for combining antithetic variates and control variates in designed simulation experiments. *Management Science,* **40**, 1021–1034.

Law, A.M. and Kelton W.D. (2000) *Simulation Modeling and Analysis*, McGraw-Hill, New York, NY.

L'Ecuyer, P. (1994) Efficiency improvement and variance reduction, in *Proceedings of the 1994 Winter Simulation Conference*, Tew, J.D., Manivannan, S., Sadowski, D.A. and Seda, A.F. (eds), IEEE, Piscataway, NJ, pp. 122–132.

McKay, M.D., Beckman, R.J. and Conover, W.J. (1979) A comparison of three methods for selecting input random variables in the analysis of output from a computer code. *Technometrics,* **21**(2), 239–245.

Minh, D.L. (1989) A variant of the conditional expectation variance reduction technique and its application to the simulation of the $GI/G/1$ queues. *Management Science,* **11**, 1334–1340.

Nelson, B.L. (1987) A perspective on variance reduction in dynamic simulation experiments. *Communications in Simulations and Statistics-Simulation and Computation*, **B16**, 385–426.

Nelson, B.L. (1990) Control variate remedies. *Operations Research,* **38**(6), 974–992.

Nelson, B.L. and Schmeiser, B.W. (1983) Variance reduction: basic transformations, in *Proceedings of the 1983 Winter Simulation Conference*, Roberts, S., Banks, J. and Schmeise, B. (eds), IEEE, Piscataway, NJ, pp. 255–258.

Ross, S.M. (1990) *A Course in Simulation*, Macmillan, New York, NY.

Ross, S.M. (2001) Applying variance reduction ideas in queueing simulations. *Probability in the Engineering and Informational Sciences,* **15**(4), 481–494.

Schruben, L.W. and Margolin, B.H. (1978) Pseudo-random number assignment in statistically designed simulation and distribution sampling experiments. *Journal of the American Statistical Association,* **73**, 504–525.

Sethi, Y.K. (1963). A note on optimum stratification of populations for estimating population means. *Australian Journal of Statistics,* **5**, 20–33.

Song, W. and Chiu, W. (2007) A five-class variance swapping rule for simulation experiments: A correlated-blocks design. *IIE Transactions,* **39**, 713–722.

Sullivan, R.S., Hayya, J.C. and Schaul, R. (1982) Efficiency of the antithetic variate method for simulating stochastic networks. *Management Science,* 28, 563–572.

Tew, J.D. and Wilson, J.R. (1994) Estimating simulation meta-models using combined correlation-based variance reduction techniques. *IIE Transactions,* **26**, 2–16.

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

Wilson, J.R. and Pritsker, A.A.B. (1994) Experimental evaluation of variance reduction techniques for queueing simulation using generalized concomitant variables. *Management Science,* **30**, 1459–1472.

Yang, W.N. and Liou, W.W. (1996) Combining antithetic variates and control variates in simulation experiments. *ACM Transactions on Modeling and Computer Simulation,* **6**, 243–260.

Yang, W.N. and Nelson, B.L. (1992) Multivariate batch means and control variates. *Management Science,* **38**, 1415–1431.

Biographies

Ihsan Sabuncuoglu is a Professor and Chair of Industrial Engineering at Bilkent University. He received the B.S. and M.S. degrees in Industrial Engineering from the Middle East Technical University and a Ph.D. degree in Industrial Engineering from Wichita State University. He teaches and conducts research in the areas of simulation, scheduling, and manufacturing systems. He has published papers in *IIE Transactions*, *Interfaces*, *Decision Sciences*, *Simulation*, *International Journal of Production Research*, *International Journal of Flexible Manufacturing Systems*, *Journal of Manufacturing Systems*, *European Journal of Operational Research*, *Journal of Operational Research Society*, *Journal of Military Operations Research*, *OMEGA* and *Computers and Industrial Engineering*. He is on the Editorial Board of *International Journal of Computer Integrated Manufacturing*, *International Journal of Operations and Quantitative Management Journal of Operations Management* and *International Journal of Systems Science*. He is an associate member of the Institute of Industrial Engineering and the Institute for Operations Research and the Management Sciences.

Mehmet Murat Fadiloglu is an Assistant Professor of Industrial Engineering at Bilkent University. He received his B.S. in Electrical Engineering from Bilkent University, and his Ph.D. in Industrial and Systems Engineering from the University of Florida. His research interests include stochastic modeling, models for production and inventory systems and simulation. He has published papers in *European Journal Operational Research, Mathematical and Computer Modeling* and *Naval Research Logistics.* He is a member of the Institute for Operations Research and the Management Sciences.

Sabri Celik is a Ph.D. candidate in the Department of Industrial Engineering and Operations Research at Columbia University. He received his B.S. and M.S. degrees from Bilkent University, both in Industrial Engineering. His research interests include stochastic models, dynamic pricing and revenue management and supply chain management.