Proceedings of the 2007 Winter Simulation Conference
S. G. Henderson, B. Biller, M.-H. Hsieh, J. Shortle, J. D. Tew, and R. R. Barton, eds.

DERIVATIVE ESTIMATION WITH KNOWN CONTROL-VARIATE VARIANCES

Jamie R. Wieland
Bruce W. Schmeiser

School of Industrial Engineering
315 North Grant Street
Purdue University
West Lafayette, IN 47907, U.S.A.

ABSTRACT

We investigate the conception that the sample variance of the control variate (CV) should be used for estimating the optimal linear CV weight, even when the CV variance is known. A mixed estimator, which uses an estimate of the correlation of the performance measure (Y) and the control (X) is evaluated. Results indicate that the mixed estimator has most potential benefit when no information on the correlation of X and Y is available, especially when sample sizes are small. This work is presented in terms of CV for familiarity, but its primary application is in derivative estimation. In this context, unlike CV, X and Y are not assumed to be correlated.

1 INTRODUCTION

In simulation experiments control variate (CV) estimators are used for variance reduction. Much work has been done in developing and analyzing CV estimators, including Lavenberg and Welch (1981), Rubinstein and Marcus (1985), Nelson (1989), Nelson and Richards (1991), and Szechtman and Glynn (2001).

In an experiment with an objective of estimating the expected value of performance measure Y, the linear CV estimator is $\bar{Y} - \alpha\left(\bar{X} - \mathrm{E}(X)\right)$, where $\bar{Y}$ is the sample mean of the performance measure, $\bar{X}$ is the sample mean of the control, and α is the CV weight. The choice of α that minimizes the variance of the CV estimator is $\alpha^* = \sigma_{XY} / \sigma_X^2$, where σ_{XY} is the covariance of X and Y, σ_X^2 is the variance of the control, and α^* is referred to as the optimal CV weight (Law and Kelton 2000). Assuming independent sampling, the variance of the CV estimator using α^* is

$$n^{-1}\sigma_Y^2\left(1-\rho_{XY}^2\right), \qquad (1)$$

which is less than the variance of $\bar{Y}$ when the correlation between X and Y, denoted ρ_{XY}, does not equate to zero. To use the optimal CV weight σ_{XY} and σ_X^2 must be known. Otherwise, these quantities are estimated, resulting in a variance reduction less than that achieved in (1) (Bauer 1987; and Bauer, Venkatraman, and Wilson 1987).

We consider the case where α^* must be estimated, but the control variance, σ_X^2, is known. In this case, there is a choice of using either the known variance, σ_X^2, or the sample variance, $\hat{\sigma}_X^2$, in estimating α^*.

Such a case may occur, for example, when the control is an input variable with a user-specified distribution. Cheng and Feast (1980) note that the majority of controls suggested in the literature do not have known variance; using standardized sums, they develop a method for converting a control with unknown variance into one with known variance.

Let $\hat{\alpha}^*$ be the estimate of α^* using least-squares estimates for σ_{XY} and σ_X^2 based on sample sizes of n observations,

$$\hat{\alpha}^* = \frac{\hat{\sigma}_{XY}}{\hat{\sigma}_X^2} = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2},$$

and $\hat{\alpha}_{KV}^*$ be the estimate of α^* using the least squares estimate of σ_{XY} with n observations and known variance σ_X^2,

$$\hat{\alpha}_{\mathrm{KV}}^* = \frac{\hat{\sigma}_{XY}}{\sigma_X^2} = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{(n-1)\sigma_X^2}.$$

Common conception is that even when σ_X^2 is known, $\hat{\alpha}^*$, rather than $\hat{\alpha}_{KV}^*$, should be used to estimate α^*. This conception could be attributed to results from the analysis of ratio estimators. For example, using the first order terms of a Taylor series expansion around the means to approximate the variances of $\hat{\alpha}^*$ and $\hat{\alpha}_{KV}^*$ (see Appendix A), we find that $\hat{\alpha}^*$ has lower variance than $\hat{\alpha}_{KV}^*$ when

$$\mathrm{Cov}\left(\hat{\sigma}_{XY},\hat{\sigma}_X^2\right) \geq \frac{\sigma_{XY}\,\mathrm{Var}\left(\hat{\sigma}_{XY}\right)}{2\sigma_X^2}. \qquad (2)$$

This expression indicates that positive correlation between the numerator and denominator of a ratio estimator can reduce variance. Due to commonality of terms in $\hat{\sigma}_{XY}$ and $\hat{\sigma}_X^2$, these quantities tend to have positive covariance, which is why it can be preferable to use the sample variance, rather than the known variance in estimating α^*. The applicability of expression (2) is limited, however, when σ_{XY} and $\mathrm{Var}\left(\hat{\sigma}_{XY}\right)$ are unknown.

In this work we develop a model, assuming that X and Y have a bivariate normal distribution, to quantify the decision of whether to use $\hat{\alpha}^*$ or $\hat{\alpha}^*_{KV}$ for estimating α^*. Results of our analysis indicate that the correlation between $\hat{\sigma}_X^2$ and $\hat{\sigma}_{XY}$ is almost equivalent to ρ_{XY}. This is constructive because we have solved for an upper bound on ρ_{XY} for which $\hat{\alpha}^*_{KV}$ is the preferred estimator. Furthermore, without much additional computational effort, when ρ_{XY} is unknown, it can be estimated and used to indicate whether $\hat{\alpha}^*$ or $\hat{\alpha}^*_{KV}$ is preferred.

Using an estimate of ρ_{XY}, denoted $\hat{\rho}_{XY}$, we evaluate a mixed estimator for α^*. Compared to $\hat{\alpha}^*$ and $\hat{\alpha}^*_{\mathrm{KV}}$, the mixed estimator's performance is more robust evaluated across all possible values of ρ_{XY}. As a result, the mixed estimator will provide the most benefit in contexts where ρ_{XY} is unknown.

Despite this work being presented in terms of CV, its applicability in this context is limited because controls are chosen such that ρ_{XY} is high in order to maximize the reduction in variance achieved.

Our results are more useful, however, in the context of gradient estimation, which is the primary motivation for this work. Wieland and Schmeiser (2006) propose using $\hat{\alpha}^*$ to estimate the derivative of the expected value of the performance measure with respect to the expected value of input X,

$$\frac{d\,\mathrm{E}(Y)}{d\,\mathrm{E}(X)}.$$

In this context, unlike that of CV, the objective is not to reduce variance relative to the sample mean, but only to obtain a point estimate of the derivative. Furthermore, X and Y are not assumed to be correlated as they are in CV. For example, if $d\,\mathrm{E}(Y)/d\,\mathrm{E}(X)$ is close to zero, then ρ_{XY} may also be close to zero. Thus, there is more potential benefit in using $\hat{\rho}_{XY}$ to indicate whether σ_X^2 or $\hat{\sigma}_X^2$ should be used when estimating α^*.

2 PROBLEM STATEMENT

Given performance measure $\mathrm{E}(Y)$; the expected value, μ_X, and variance, σ_X^2, of control/input X; sample size n; and the ability to obtain independent observations of both X and Y, denoted (X_i, Y_i), for $i = 1,2,\ldots,n$; we analyze the problem of whether to use $\hat{\alpha}^*$ or $\hat{\alpha}^*_{KV}$ for estimating α^*. There are, of course, other estimators that could be considered, one of which is evaluated in Section 5.

The metric used to compare estimators is

$$\frac{n \times \mathrm{E}\left(\left(\hat{\alpha}-\alpha^*\right)^2\right)}{\left|\alpha^*\right|+1}.$$

We refer to this metric as relative error, which is generalized MSE standardized by $\left|\alpha^*\right|+1$. In the denominator of this metric, the absolute value of α^* is used to eliminate differences in estimating positive and negative weights. Furthermore, one is added to $\left|\alpha^*\right|$ preventing division by zero for the cases where $\alpha^* = 0$.

3 PREVIOUS WORK

Bauer (1987b) first proposed using known CV variances for estimating α^*. Assuming a multivariate normal model, he shows that known variance can yield better estimators depending on ρ_{XY}. This work differs from our work in that he does not estimate ρ_{XY} to indicate which estimator is preferred.

Cheng and Feast (1980) use standardized sums to develop controls with known variances. They find that these controls yield better CV estimators.

Schmeiser and Taaffe (2000) investigate replacing the control-simulation mean with an approximation. The resulting control-variate estimator is biased.

4 COMPARING ESTIMATORS

To compare $\hat{\alpha}^*$ and $\hat{\alpha}^*_{\mathrm{KV}}$, we assume that X and Y have a bivariate normal (BVN) distribution with parameters μ_X, μ_Y, σ_X^2, σ_Y^2, and ρ_{XY}. The foundation of this assumption is that asymptotically, as sample sizes approach infinity, most estimators follow a multivariate central limit theorem.

Under the BVN assumption we find that the correlation of $\hat{\sigma}_X^2$ and $\hat{\sigma}_{XY}$, denoted ρ_σ, is almost equivalent to ρ_{XY}. (See Figure 1.) Since these quantities capture the same effects, we proceed with our analysis in terms of ρ_{XY} rather than $\mathrm{Cov}\left(\hat{\sigma}_{XY},\hat{\sigma}_X^2\right)$, which was the original concept presented in Section 1, because we have obtained expressions for variances of $\hat{\alpha}^*$ and $\hat{\alpha}^*_{\mathrm{KV}}$ in terms of ρ_{XY}.

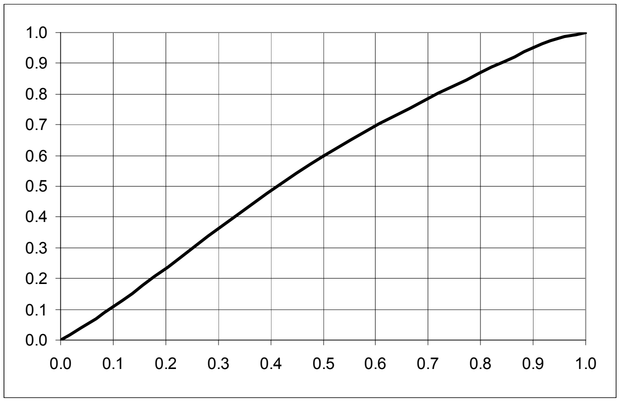

Figure 1: Estimated Monte Carlo results for ρ_σ versus ρ_{XY}. Note that these two quantities are almost equivalent.

Under the BVN assumption, both $\hat{\alpha}^*$ and $\hat{\alpha}^*_{KV}$ are unbiased estimators for α^* so we compare/contrast only the variances of these estimators, which assuming independent sampling are

$$\operatorname{Var}\left(\hat{\alpha}^*\right)=\frac{\sigma_Y^2}{\sigma_X^2}\cdot\frac{\left(1-\rho_{XY}{}^2\right)}{\left(n-3\right)}$$

and

$$\operatorname{Var}\left(\hat{\alpha}^*_{\mathrm{KV}}\right)=\frac{\sigma_Y^2}{\sigma_X^2}\cdot\frac{\left(1+\rho_{XY}{}^2\right)}{\left(n-1\right)}.$$

Refer to Appendix B for details. The means, μ_X and μ_Y, do not affect the variances. Furthermore, σ_X^2 and σ_Y^2 only re-scale the problem, affecting the variances of both estimators equally.

The better estimator, as measured by lower variance, depends on sample size n and ρ_{XY}, which is typically unknown. As $\left|\rho_{XY}\right|\to 1$, however, $\hat{\alpha}^*$ is the better estimator, because $\operatorname{Var}\left(\hat{\alpha}^*\right)\to 0$ regardless of sample size.

To further illustrate the dependence of the statistical performance of these estimators on ρ_{XY} and sample size, we plot the relative error metric versus ρ_{XY} for $\hat{\alpha}^*$ and $\hat{\alpha}^*_{KV}$. See Figure 2. For simplicity and without loss of generality, we fix $\sigma_X^2=\sigma_Y^2$ and consider only cases where ρ_{XY} is positive because the graph is symmetric along the vertical axis.

Figure 2 compares the relative error for $\hat{\alpha}^*$ and $\hat{\alpha}^*_{\mathrm{KV}}$. Error curves for $\hat{\alpha}^*_{\mathrm{KV}}$ are displayed with dotted lines, and solid lines are used to display the curves for $\hat{\alpha}^*$. Curves for samples sizes $n=4,5,10$ are shown with the darker gray lines representing larger sample sizes. The black curves indicate the limiting cases as $n\to\infty$.

For a given sample size, the better estimator, as measure by lower relative error, depends on ρ_{XY}. As sample sizes increase, $\hat{\alpha}^*$ is the better estimator across all ρ_{XY}.

As stated previously in this section, the BVN assumption is based on the concept that asymptotically, as sample sizes approach infinity, most estimators follow a multivariate central limit theorem. Despite the assumption being supported asymptotically, we have presented results that are dependent on sample size. These results should be interpreted in the context of batch means (Law

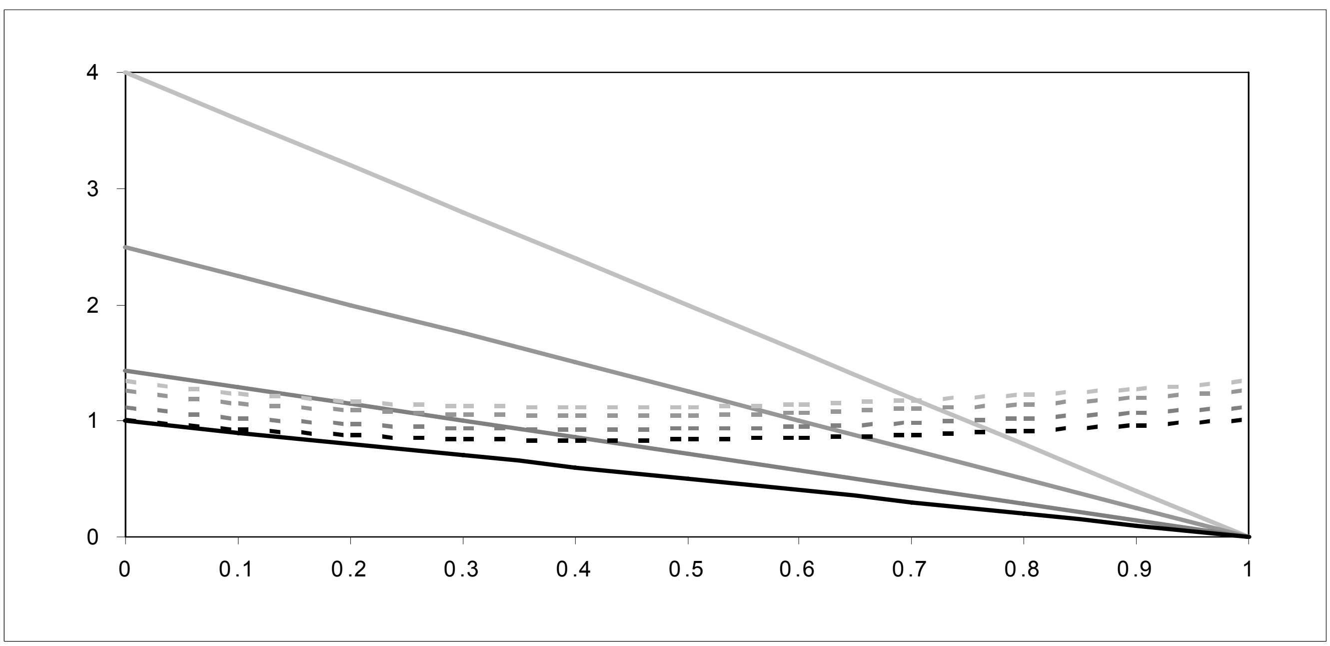

Figure 2: Relative error versus ρ_{XY} for both $\hat{\alpha}^*$ and $\hat{\alpha}^*_{\mathrm{KV}}$. Error curves for $\hat{\alpha}^*_{\mathrm{KV}}$ are displayed with dotted lines. Solid lines are used to display the curves for $\hat{\alpha}^*$. Curves for samples sizes $n=4,5,10$ are shown with the darker gray lines representing larger sample sizes. The black curves indicate the limiting cases as $n\to\infty$.

and Kelton 2000). We assume independent sampling in our analysis, but simulation output data can be autocorrelated. In such a case the number of dependent samples required to equate to one independent sample is

$$\frac{n}{1+2\sum_{h=1}^{n-1}\left(1-\frac{h}{n}\right)\rho_h},$$

where ρ_h is the lag-*h* autocorrelation. This expression indicates that when autocorrelation is present in output data, large samples of dependent data may be required to obtain the equivalent of only a few independent observations.

Given n independent observations, the upper bound on ρ_{XY} for which the relative error of $\hat{\alpha}^*_{KV}$ is less than that of $\hat{\alpha}^*$ is

$$\rho_{XY} < \left(\frac{1}{n-2}\right)^{1/2}. \tag{3}$$

Using this expression, if ρ_{XY} were known, we could determine which estimator to use. The decision rule would simply be to use $\hat{\alpha}^*_{KV}$ if ρ_{XY} is less than $(n-2)^{-1/2}$ and $\hat{\alpha}^*$ otherwise.

When ρ_{XY} is unknown, which is typically the case, it can be estimated, providing information as to which is the preferred estimator for α^*.

5 A MIXED ESTIMATOR

We now examine the problem of developing an estimator for α^* given not only the information listed in the original problem statement in Section 2, but also an estimate of ρ_{XY}.

Assuming that we are given $\hat{\rho}_{XY}$, we use expression (3) to develop a new estimator for α^* as

$$\hat{\alpha}^*_1 = \begin{cases} \hat{\alpha}^*, & \hat{\rho}_{XY} \le (n-2)^{-1/2} \\ \hat{\alpha}^*_{KV}, & \text{o.w.} \end{cases}.$$

Ideally, the boundary between $\hat{\alpha}^*$ and $\hat{\alpha}^*_{KV}$ would be chosen such that the relative error of $\hat{\alpha}^*_1$ is minimized. Using a boundary of $(n-2)^{-1/2}$, however, is a reasonable approximation because the error curves displayed in Figure 2 for $\hat{\alpha}^*$ and $\hat{\alpha}^*_{KV}$ are somewhat linear. This represents a constant loss function for minimizing the relative error of $\hat{\alpha}^*_1$. Therefore, any deviation in the optimal boundary away from $(n-2)^{-1/2}$ would be attributed ρ_{XY} being unknown. When the loss function is not constant, the optimal boundary leans towards the direction of lower loss in terms of relative error.

5.1 Experimental Results

Using Monte Carlo results we estimate $\text{MSE}(\hat{\alpha}^*_1)$ across sample sizes $n = 4, 5, 10, 100$ and correlations $\rho_{XY} = 0, 0.25, 0.5, 0.75, 1$.

Figure 3 displays the relative error of $\hat{\alpha}^*_1$ versus ρ_{XY}, which is indicated by the mixed solid/dotted line. The relative error curves for $\hat{\alpha}^*$ and $\hat{\alpha}^*_{\text{KV}}$ are also displayed in Figure 3, with solid curves for $\hat{\alpha}^*$ and dotted curves for $\hat{\alpha}^*_{\text{KV}}$. Two curves are displayed for each estimator. The

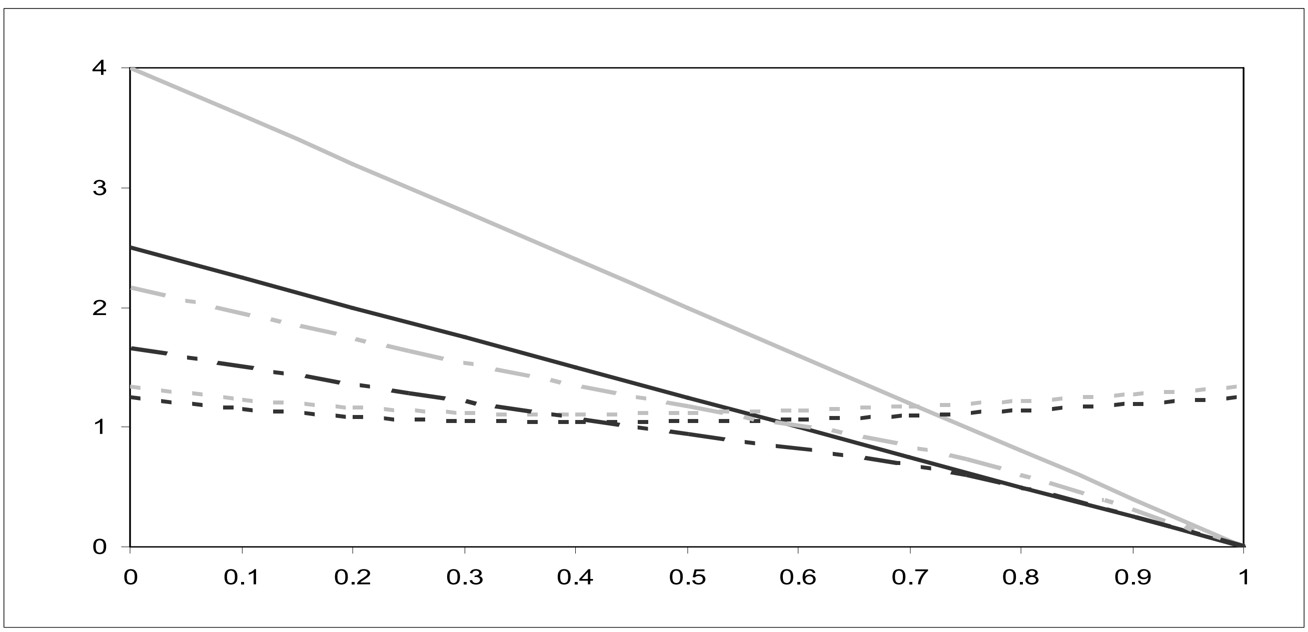

Figure 3: Relative error versus ρ_{XY} for $\hat{\alpha}^*_1$, which is indicated by the mixed solid/dotted line. The relative error curves for $\hat{\alpha}^*$ and $\hat{\alpha}^*_{\text{KV}}$ are also displayed, with solid curves for $\hat{\alpha}^*$ and dotted curves for $\hat{\alpha}^*_{\text{KV}}$. Two curves are displayed for each estimator. The lighter gray curves represent sample sizes of $n = 4$. The black curves represent sample sizes of $n = 5$.

gray curves represent sample sizes of $n = 4$. The black curves represent sample sizes of $n = 5$.

Curves for all three estimators for sample sizes of $n = 10$ and $n = 100$ are shown in Figure 4. Note that the scale of the vertical axis has been magnified so that differences between the estimators are visible. In Figure 4, the gray curves represent sample sizes of $n = 10$. The black curves represent sample sizes of $n = 100$.

Results displayed in Figures 3 and 4 show that $\hat{\alpha}_1^*$ has superior performance to $\hat{\alpha}^*$ for small sample sizes and almost equivalent performance for large sample sizes. Additionally, like the relative error for $\hat{\alpha}^*$, the error for $\hat{\alpha}_1^*$ is 0 when $\rho_{XY} = 1$ and it increases as $\rho_{XY} \to 0$. One reason that error increases as $\rho_{XY} \to 0$, especially for small sample sizes, is that the variance of $\hat{\rho}_{XY}$ increases as $\rho_{XY} \to 0$ and as sample size decreases.

Because the relative error of $\hat{\alpha}_1^*$ is not lower than that of $\hat{\alpha}_{\mathrm{KV}}^*$ across all values of ρ_{XY}, an alternative metric for comparison given sample size n would be area under the relative error curve. Using this metric, $\hat{\alpha}_1^*$ has slightly better performance than $\hat{\alpha}_{\mathrm{KV}}^*$ for $n = 4$, but much better performance as sample size increases.

6 CONCLUSIONS

We proposed a mixed estimator that uses an estimate of ρ_{XY} to decide between using σ_X^2 and $\hat{\sigma}_X^2$ in estimating α^*. Compared to $\hat{\alpha}^*$ and $\hat{\alpha}_{\mathrm{KV}}^*$, the mixed estimator's performance is more robust evaluated across all possible values of ρ_{XY}.

Our recommendation is to use the mixed estimator for estimating α^* in cases where no prior information about ρ_{XY} is known, regardless of sample size. When sample sizes are small (i.e. $n \le 10$), $\hat{\alpha}_1^*$ should be used unless ρ_{XY} is thought to be relatively close to zero (i.e. $\rho_{XY} \le 0.3$). In that case $\hat{\alpha}_{\mathrm{KV}}^*$ should be used. When sample sizes are moderate to large (i.e. $n > 10$), there is relatively little difference between $\hat{\alpha}_1^*$ and $\hat{\alpha}^*$. Either of these estimator are preferred to $\hat{\alpha}_{\mathrm{KV}}^*$ for all ρ_{XY}.

Our results indicate that the conception that even when σ_X^2 is known $\hat{\sigma}_X^2$ should be used for estimating α^* is true for large sample sizes. When sample sizes are small, the decision as to whether to use σ_X^2 or $\hat{\sigma}_X^2$ in estimating α^* is dependent on ρ_{XY}.

7 RELEVANCE AND FUTURE RESEARCH

This work was presented in the context of control variates, but its relevance to this area is limited because controls are often chosen such that ρ_{XY} is high. Therefore, improvement in using the proposed mixed estimator ($\hat{\alpha}_1^*$) over the traditional estimator ($\hat{\alpha}^*$) for the optimal control weight is marginal, especially for large sample sizes.

The proposed mixed estimator has most potential benefit when little information regarding ρ_{XY} is known. Such a case commonly occurs in gradient estimation,

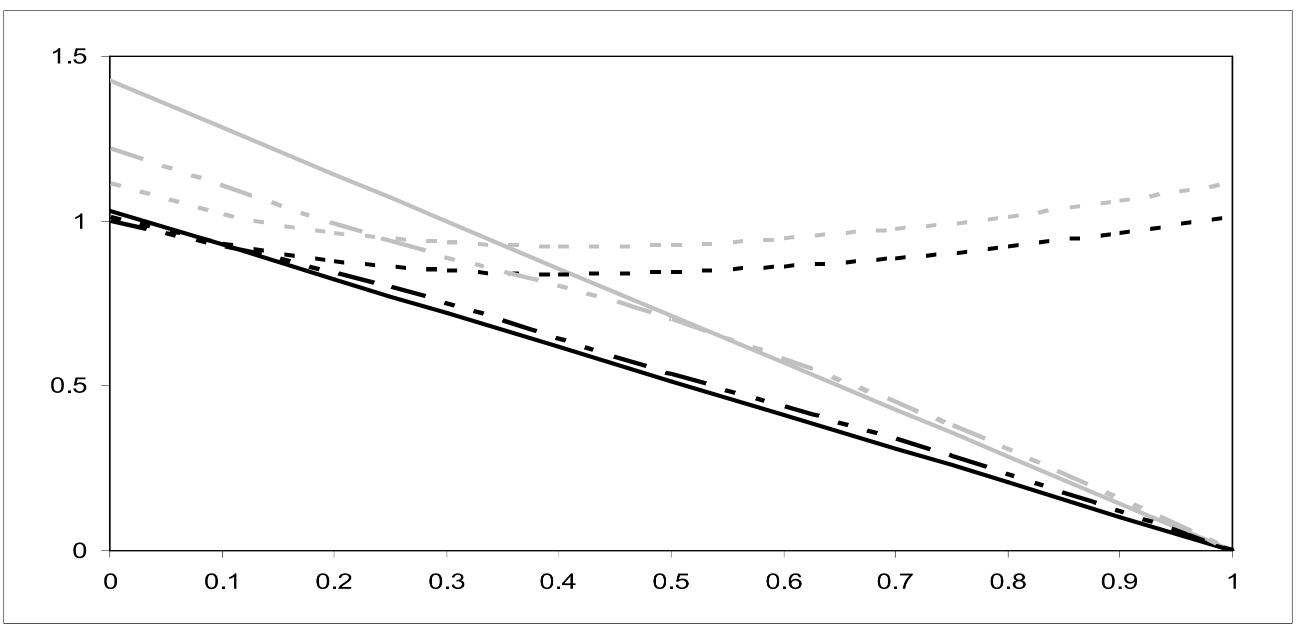

Figure 4: Relative error versus ρ_{XY} for $\hat{\alpha}_1^*$, which is indicated by the mixed solid/dotted line. The relative error curves for $\hat{\alpha}^*$ and $\hat{\alpha}_{\mathrm{KV}}^*$ are also displayed, with solid curves for $\hat{\alpha}^*$ and dotted curves for $\hat{\alpha}_{\mathrm{KV}}^*$. Two curves are displayed for each estimator. The lighter gray curves represent sample sizes of $n = 10$. The black curves represent sample sizes of $n = 100$. Note that the scale of the vertical axis has been magnified in this figure, compared to that in Figures 2 and 3, so that differences between estimators are visible.

which was the primary motivation for this work (refer to Wieland and Schmeiser 2006 for details). In this context, unlike CV, there are typically no prior assumptions regarding ρ_{XY}.

We focused on analyzing only three estimators, but there are others that could be considered. For example, in the CV context it is assumed that the expected value of the control is known. This information could be used in the estimator for the optimal CV weight by replacing the sample mean of the control with its expected value as

$$\frac{\sum_{i=1}^{n}\left(X_i - \mathrm{E}(X)\right)\left(Y_i - \bar{Y}\right)}{\sum_{i=1}^{n}\left(X_i - \mathrm{E}(X)\right)^2}.$$

Compared to the traditional CV estimator, $\hat{\alpha}^*$, this estimator has an additional degree of freedom. Another alternative would be to use a linear combination of $\hat{\alpha}^*$ and $\hat{\alpha}^*_{\mathrm{KV}}$,

$$\tau\hat{\alpha}^*_{\mathrm{KV}} + (1-\tau)\hat{\alpha}^*.$$

In this estimator the weight, τ, should be chosen to minimize relative error. Because $\hat{\alpha}^*$ and $\hat{\alpha}^*_{\mathrm{KV}}$ are dependent, the optimal weight is a function of $\mathrm{Cov}\left(\hat{\alpha}^*, \hat{\alpha}^*_{\mathrm{KV}}\right)$, which would need to be estimated. This is an area of future research.

Another area of future research is extending these results to higher dimensional problems. Such problems would incorporate multiple controls and extend derivative estimation to gradients.

APPENDIX A

Consider two ratio estimators $R_1 = Z/a$ and $R_2 = Z/A$, where a is $\mathrm{E}(A)$. The variance of R_1 is

$$\mathrm{Var}(Z)/a^2.$$

An approximation of the variance of R_2 is obtained from using the first-order terms of a Taylor series expansion around $\mathrm{E}(Z)$ and $\mathrm{E}(A)$, which is

$$\frac{\mathrm{Var}(Z)}{\mathrm{E}(A)^2} + \frac{\mathrm{E}(Z)^2\,\mathrm{Var}(Z)}{\mathrm{E}(A)^4} - \frac{2\,\mathrm{E}(Z)\mathrm{Cov}(Z,A)}{\mathrm{E}(A)^3}.$$

Comparing the variance of R_1 and R_2, we find that the variance of R_2 is less than that of R_1 when

$$\mathrm{Cov}(Z,A) \geq \frac{\mathrm{E}(Z)\mathrm{Var}(Z)}{2\,\mathrm{E}(A)}.$$

Putting this result in terms of our notation for the optimal linear CV weight, we have

$$\mathrm{Cov}\left(\hat{\sigma}_{XY}, \hat{\sigma}_X^2\right) \geq \frac{\sigma_{XY}\,\mathrm{Var}\left(\hat{\sigma}_{XY}\right)}{2\sigma_X^2}.$$

Therefore, when this inequality holds, using $\hat{\sigma}_X^2$, rather than σ_X^2, results in $\hat{\alpha}^*$ having lower variance.

APPENDIX B

Result 1: $\mathrm{Var}\left(\hat{\alpha}^*_{KV}\right) = \dfrac{\sigma_Y^2\left(1+\rho^2\right)}{\sigma_X^2\,(n-1)}$

Proof:

$$\mathrm{Var}\left(\hat{\alpha}^*_{KV}\right) = \underbrace{\mathrm{Var}_{X_i}\left(\mathrm{E}\left(\hat{\alpha}^*_{KV} \mid X_i = x_i,\ \forall i\right)\right)}_{\text{Part A}}$$
$$+\underbrace{\mathrm{E}_{X_i}\left(\mathrm{Var}\left(\hat{\alpha}^*_{KV} \mid X_i = x_i,\ \forall i\right)\right)}_{\text{Part B}}$$

Part A: Calculation for $\hat{\alpha}^*_{KV}$

$$\mathrm{Var}_{X_i}\left(\mathrm{E}\left(\hat{\alpha}^*_{KV} \mid X_i = x_i,\ \forall i\right)\right)$$
$$= \mathrm{Var}_{X_i}\left(\mathrm{E}\left(\left.\frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{(n-1)\sigma_X^2}\right| X_i = x_i,\ \forall i\right)\right)$$
$$= \mathrm{Var}_{X_i}\left(\frac{1}{(n-1)\sigma_X^2}\left(\sum_{i=1}^{n}\left(x_i - \bar{x}\right)\mathrm{E}\left(\left(Y_i - \bar{Y}\right) \mid X_i = x_i\right)\right)\right)$$
$$= \mathrm{Var}_{X_i}\left(\frac{1}{(n-1)\sigma_X^2}\left(\sum_{i=1}^{n}\left(x_i - \bar{x}\right)\left(\alpha\left(\bar{x} - x_i\right)\right)\right)\right)$$
$$= \frac{\alpha^2}{(n-1)^2\sigma_X^4}\mathrm{Var}_{X_i}\left(\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(\bar{X} - X_i\right)\right)$$

$*$ note that $\sum_{i=1}^{n}\left(\bar{X} - X_i\right)\left(\bar{X}\right) = 0$

$$= \frac{\alpha^2}{(n-1)^2\sigma_X^4}\mathrm{Var}_{X_i}\left(\sum_{i=1}^{n}\left(X_i^2 - X_i\bar{X}\right)\right)$$
$$= \frac{\alpha^2}{(n-1)^2\sigma_X^4}(n-1)^2\,\mathrm{Var}_{X_i}\left(S_X^2\right) = \frac{2\alpha^2}{(n-1)}$$

Part B: Calculation for $\hat{\alpha}^*_{KV}$

$$\mathrm{E}_{X_i}\left(\mathrm{Var}\left(\hat{\alpha}^*_{KV} \mid X_i = x_i, \forall i\right)\right)$$
$$= \mathrm{E}_{X_i}\left(\mathrm{Var}\left(\left.\frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{(n-1)\sigma_X^2}\right| X_i = x_i,\ \forall i\right)\right)$$
$$= \mathrm{E}_{X_i}\left(\frac{1}{(n-1)^2\sigma_X^4}\mathrm{Var}\left(\left.\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)\right| X_i = x_i,\ \forall i\right)\right)$$

assuming that Y_i are independent of each other $\forall\ i$

$*$ note that $\sum_{i=1}^{n}\left(x_i - \bar{x}\right)\left(\bar{Y}\right) = 0$

$$= \mathrm{E}_{X_i}\left(\frac{1}{(n-1)^2\sigma_X^4}\sum_{i=1}^{n}(x_i-\bar{x})^2 \,\mathrm{Var}(Y_i \mid X_i = x_i)\right)$$

$$= \mathrm{E}_{X_i}\left(\frac{1}{(n-1)^2\sigma_X^4}\sum_{i=1}^{n}(x_i-\bar{x})^2\left(\sigma_Y^2\left(1+\rho^2\right)\right)\right)$$

$$= \mathrm{E}_{X_i}\left(\frac{\left(\sigma_Y^2\left(1+\rho^2\right)\right)}{(n-1)^2\sigma_X^4}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2\right)$$

$$= \frac{\left(\sigma_Y^2\left(1+\rho^2\right)\right)}{(n-1)^2\sigma_X^4}(n-1)\,\mathrm{E}_{X_i}\left(S_X^2\right) = \frac{\left(\sigma_Y^2\left(1+\rho^2\right)\right)}{(n-1)\sigma_X^2}$$

□

Result 2: $\mathrm{Var}\left(\hat{\alpha}^*\right) = \dfrac{\sigma_Y^2\left(1-\rho^2\right)}{\sigma_X^2(n-3)}$

Proof:

$$\mathrm{Var}\left(\hat{\alpha}^*\right) = \underbrace{\mathrm{Var}_{X_i}\left(\mathrm{E}\left(\hat{\alpha}^* \mid X_i = x_i,\ \forall i\right)\right)}_{\text{Part A}} + \underbrace{\mathrm{E}_{X_i}\left(\mathrm{Var}\left(\hat{\alpha}^* \mid X_i = x_i,\ \forall i\right)\right)}_{\text{Part B}}$$

Part A: Calculation for $\hat{\alpha}^*$

$$\mathrm{Var}_{X_i}\left(\mathrm{E}\left(\hat{\alpha}^* \mid X_i = x_i,\ \forall i\right)\right)$$

$$= \mathrm{Var}_{X_i}\left(\mathrm{E}\left(\left.\frac{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)\left(Y_i-\bar{Y}\right)}{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2}\right| X_i = x_i,\ \forall i\right)\right)$$

$$= \mathrm{Var}_{X_i}\left(\frac{\sum_{i=1}^{n}(x_i-\bar{x})\,\mathrm{E}\left(\left(Y_i-\bar{Y}\right) \mid X_i = x_i\right)}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right)$$

$$= \mathrm{Var}_{X_i}\left(\frac{\sum_{i=1}^{n}(x_i-\bar{x})\,\mathrm{E}\left(\left(\begin{array}{l} \left(y_0+\alpha(x_i-\mu_X)\right) \\ -\frac{1}{n}\sum_{i=1}^{n}\left(y_0+\alpha(x_i-\mu_X)\right)\end{array}\right)\right)}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right)$$

$$= \mathrm{Var}_{X_i}\left(\frac{\alpha\sum_{i=1}^{n}(x_i-\bar{x})(x_i-\bar{x})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right) = \mathrm{Var}_{X_i}(\alpha) = 0$$

Part B: Calculation for $\hat{\alpha}^*$

$$\mathrm{Var}\left(\hat{\alpha}^* \mid X_i = x_i,\ \forall i\right)$$

$$= \mathrm{Var}\left(\left.\frac{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)\left(Y_i-\bar{Y}\right)}{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2}\right| X_i = x_i,\ \forall i\right)$$

assume that Y_i are independent of each other $\forall\ i$

$*$ note that $\sum_{i=1}^{n}(x_i-\bar{x})\left(\bar{Y}\right) = 0$

$$= \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2\,\mathrm{Var}(Y_i \mid X_i = x_i)}{\left(\sum_{i=1}^{n}(x_i-\bar{x})^2\right)^2}$$

$$= \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2\left(\sigma_Y^2\left(1-\rho^2\right)\right)}{\left(\sum_{i=1}^{n}(x_i-\bar{x})^2\right)^2}$$

$$= \frac{\left(\sigma_Y^2\left(1-\rho^2\right)\right)\sum_{i=1}^{n}(x_i-\bar{x})^2}{\left(\sum_{i=1}^{n}(x_i-\bar{x})^2\right)^2} = \frac{\left(\sigma_Y^2\left(1-\rho^2\right)\right)}{\sum_{i=1}^{n}(x_i-\bar{x})^2}.$$

Moreover,

$$\mathrm{E}_{X_i}\left(\mathrm{Var}\left(\hat{\alpha}_2 \mid X_i = x_i,\ \forall i\right)\right)$$

$$= \mathrm{E}_{X_i}\left(\frac{\left(\sigma_Y^2\left(1-\rho^2\right)\right)}{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2}\right) = \mathrm{E}_{X_i}\left(\frac{\left(\sigma_Y^2\left(1-\rho^2\right)\right)}{(n-1)S_X^2}\right)$$

$$= \frac{\left(\sigma_Y^2\left(1-\rho^2\right)\right)}{(n-1)}\mathrm{E}_{X_i}\left(S_X^{-2}\right) = \frac{\sigma_Y^2\left(1-\rho^2\right)}{\sigma_X^2(n-3)}$$

□

REFERENCES

Bauer, K. W. 1987. *Improved batching for confidence interval construction in steady-state simulation*. Ph.D. thesis, School of Industrial Engineering, Purdue University, West Lafayette, Indiana.

Bauer, K. W., Venkatraman, S., and J. R. Wilson. 1987. Estimation procedures based on control variates with known covariance matrix. In *Proceedings of the 1987 Winter Simulation Conference,* ed. A. Thesen, H. Grant, and W. D. Kelton, 334—341.

Cheng, R.C.H., and G. M. Feast. 1980. Control variables with known mean and variance. *The Journal of the Operational Research Society* 31: 51—56.

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

Law, A. M., and W. D. Kelton. 2000. *Simulation modeling & analysis.* 3rd ed. New York: McGraw-Hill, Inc.

Nelson, B. L. 1989. Batch size effects on the efficiency of control variates in simulation. *European Journal of Operational Research* 43:184—196.

Nelson, B. L., and D. Richards. 1991. Control variate remedies. *Operations Research* 38:974—992.

Rubinstein, R. Y., and R. Marcus. 1985. Efficiency of multivariate control variates in Monte Carlo simulation. *Operations Research* 33:661—677.

Schmeiser, B.W., Taaffe, M.R., and J. Wang. 2001. Biased control-variate estimation. *IIE Transactions* 33:219—228.

Szechtman, R., and P. W. Glynn. 2001. Constrained Monte Carlo and the method of control variates. In *Proceedings of the 2001 Winter Simulation Conference*, ed. B. A. Peters, J. S. Smith, D. J. Medeiros, and M. W. Rohrer, 394—400.

Wieland, J. R., and B. W. Schmeiser. 2006. Stochastic gradient estimation using a single design point. In *Proceedings of the 2006 Winter Simulation Conference,* ed. L. F. Perrone, B. G. Lawson, J. Liu, and F. P. Wieland, 390—397.

AUTHOR BIOGRAPHIES

JAMIE R. WIELAND is a Ph.D. student in the School of Industrial Engineering at Purdue University. She received a B.S. in Industrial Engineering & Management Sciences from Northwestern University in 2001 and a M.S. in Industrial Engineering from Purdue University in 2003. Her research interests are in stochastic operations research and economics. Her email address is <jwieland@purdue.edu>.

BRUCE W. SCHMEISER is a professor in the School of Industrial Engineering at Purdue University. His research interests center on developing methods for better simulation experiments. He is a Fellow of INFORMS, is a Fellow of IIE, and has been active within the Winter Simulation Conference for many years, including being the 1983 Program Chair and chairing the Board of Directors from 1988–1990. His email address is <bruce@purdue.edu>.