SELECTION OF CONTROL VARIATES FOR VARIANCE REDUCTION IN A MULTIRESPONSE SIMULATION WITH A SMALL NUMBER OF REPLICATIONS

An Caris
Gerrit K. Janssens
Hasselt University - Campus Diepenbeek
Agoralaan - Building D
B-3590 Diepenbeek, Belgium
Phone: +32-11-268775, +32-11-268648
Fax: +32-11-268700
Email: {an.caris,gerrit.janssens}@uhasselt.be

KEYWORDS
variance reduction, multiresponse simulation, control variates, robust regression.

ABSTRACT

In this paper the selection of control variates for simulation experiments with multiple response variables is discussed. The aim is to reduce the variance of all response variables as much as possible. Therefore, response variables are weighted according to their coefficient of variation. A criterion is developed to determine the number of control variates and select the most appropriate ones. When the number of replications is small due to computational costs, outliers may seriously effect the estimation of coefficients for the control variates. Robust regression, based on the least median of squares criterion, is used to detect outliers.

1. INTRODUCTION

Using simulation, excessive run lengths or replications may be necessary to yield estimators with acceptable precision. A variety of variance reduction techniques (VRTs) have been developed to improve the efficiency of simulations. Efficiency is measured by the variance of the response variables from a simulation. Variance reduction techniques attempt to reduce the variance of an output random variable and thus obtain greater precision for the same amount of simulation time, or to achieve a desired precision with less simulation time (Law and Kelton 2000).

A comprehensive overview of five promising types of VRTs can be found in Law and Kelton (2000, Ch. 11). Investigations have been made to combine and integrate variance reduction strategies by Avramidis and Wilson (1996) and Tew (1995). Crawford and Gallwey (2000) study the effect of introducing a warm-up period on the use of VRTs. This paper focuses on one specific VRT, namely Control Variates (CV).

2. CONTROL VARIATES

The method of control variates attempts to take advantage of correlation between certain random variables to obtain a variance reduction. In the single variate case, a control variate C is a random variable whose expectation E(C) = μ_c is known and which is correlated with a random response variable Y. A constant *b*, which has the same sign as the correlation between Y and C, is used to scale the deviation between C and μ_c. The variance of Y is minimized for the optimal value β of *b*

$$\beta = \frac{Cov(Y,C)}{Var(Y)} \ . \qquad (1)$$

The minimum variance of Y is given by:

$$Var[Y(\beta)] = (1-\rho_{YC}^2)\cdot Var(Y) \qquad (2)$$

where ρ_{YC} is the correlation between Y and C. The stronger the correlation between Y and C, the greater the variance of Y is reduced. Lavenberg, Moeller and Welch (1982) call the quantity (1- ρ^2_{YC}) the *minimum variance ratio*. It is the factor by which the variance of Y could be reduced if the optimum coefficient β were known. In practice β needs to be estimated. This is done by replacing Cov(Y,C) and Var(Y) by their sample estimators, which is identical to calculating the least-squares estimate of the slope coefficient in the linear regression of Y on C.

Emsermann and Simon (2002) discuss the situation where the mean of a proposed control variate is unknown. The authors develop a method to estimate the mean of these quasi control variates so that they can still be used to improve the efficiency of a simulation.

Multiple control variates may be used. In this case **C**, $\boldsymbol{\mu_c}$, $\boldsymbol{b}$ and $\boldsymbol{\beta}$ represent vectors and ρ_{YC} is replaced by R_{YC}, which is the multiple correlation coefficient between Y and **C**. Estimates of $\boldsymbol{\beta}$ are obtained by performing a multiple linear regression of Y on **C**. Unfortunately, the estimation of the coefficients $\boldsymbol{\beta}$ causes a loss in variance reduction. Lavenberg et al. (1982) define the *loss factor*, the amount by which the variance is increased due to the use of estimated coefficients, as follows:

$$(K-2)/(K-Q-2) \qquad (3)$$

with Q the number of control variates and K the number of independent replications. This expression holds under the assumption that Y and **C** are normally distributed and $K > Q+2$.

In this paper the method of control variates is applied to multiresponse simulation models with a limited number of replications. **Y** represents a vector of n response variables Y_i (i=1, ..., n). The theory of CV assumes a normal distribution for both the response variables and control variates. However, this assumption may be violated when only a limited number of replications can be performed due to computational costs. Nelson (1990) evaluates a number of remedies for violations of the normality assumption, including jack-knifing, splitting, batching and bootstrapping. This study concentrates on the selection of control variates in case of a limited number of replications and proposes to use *least median of squares* (LMS) estimates instead of *ordinary least squares* (OLS) estimates.

2.1 Selection of Control Variates

According to Law and Kelton (2000) a good control variate should be strongly correlated with a response variable Y_i, in order to give a lot of information on Y_i and to make a good adjustment to it. Porta Nova and Wilson (1993) discuss the selection of control variates with the objective of estimating multiresponse metamodels. The authors examine specific covariance structures under the normality assumption.

Our aim is to determine a criterion to select a limited number of control variates for multiple response variables. Since we are confronted with a limited number of replications, the number of control variates should obviously be kept low. Rubinstein and Marcus (1985) ascertain that good variance reduction is achieved when the number of control variates is relatively small.

The objective is to reduce the variance of all response variables as much as possible. Therefore, more effort should be directed towards response variables with a higher variance relative to their average value. The relative variance of a response variable Y_i can be measured by its coefficient of variation (VarCoeff):

$$VarCoeff(Y_i) = \frac{Stdev(Y_i)}{E(Y_i)} \qquad (4)$$

For each available control variate the minimum variance ratio (1- ρ^2_{YC}) relative to each response variable can be estimated from the data. The minimum variance ratios are then weighted according to the coefficient of variation of each response variable. The weights w_i are defined as:

$$w_i = \frac{VarCoeff(Y_i)}{\sum_i VarCoeff(Y_i)} \qquad (5)$$

These weights can also be estimated from the simulation output. Then for each control variate the following criterion can be calculated:

$$\sum_i w_i(1-\rho^2_{Y_iC}) \qquad (6)$$

This results in an initial ranking of the control variates.

Next, we investigate the effect of using multiple control variates in the situation of a limited number of replications. In this case the loss factor also has to be taken into account. The expression (3) of Lavenberg et al. (1982) is used as an approximation because the normality assumption may not hold with a limited number of replications. The number and selection of control variates is based on the following criterion:

$$Min \quad \sum_i w_i(1-R^2_{Y_iC})\cdot(K-2)/(K-Q-2) \qquad (7)$$

Results are compared with the initial ranking of the control variates. Recommendations are made about the number of control variates.

2.2 Robust Regression

When the number of replications is low, Janssens, Deceuninck and Van Breedam (1995) suggest to use a robust regression method for estimating the coefficients $\boldsymbol{\beta}$. They propose to use the *least median of squares* (LMS) regression, developed by Rousseeuw (1984), instead of the *least sum of squares* criterion used in OLS regression. Classical least squares regression consists of minimizing the sum of squared residuals. The LMS method replaces the sum by the median of the squared residuals r_i^2. Its objective function can be written as:

$$Min \quad \underset{i}{med}(r^2{}_i) \qquad (8)$$

Because the median is used, this technique is more robust to outliers. Outliers can have a significant effect on the estimated coefficients when OLS regression is performed on a small number of observations. Looking at the OLS residuals is not sufficient for detecting outliers. The OLS fit often conceals bad data points. LMS can be used to detect outlying observations. The LMS method first fits a regression to the majority of the data and then discovers outliers as those points which possess large residuals from the robust regression. A disadvantage of the LMS method is its lack of efficiency. Therefore, LMS is used to purify the simulation sample from outliers. Finally, a re-weighted OLS regression is performed, in which outliers receive a zero weight.

Rousseeuw and Leroy (1987) describe PROGRESS (Program for RObust reGRESSion), an algorithm for implementing LMS regression. This program is used to estimate the coefficients $\boldsymbol{\beta}$**,** that measure the influence of control variates on response variables. First, the coefficients $\boldsymbol{\beta}$ are estimated by OLS-regression of **Y** on **C**. If all slope coefficients are significantly different from zero, $\mathbf{Y}(\boldsymbol{\beta})$ is calculated. The average and variance of $\mathbf{Y}(\boldsymbol{\beta})$ are used to construct confidence intervals. If not all coefficients $\boldsymbol{\beta}$ are significant, the LMS-method searches for outliers. Detected outliers are removed and OLS-regression is again applied. $\mathbf{Y}(\boldsymbol{\beta})$ is calculated, but now the median instead of the average of $\mathbf{Y}(\boldsymbol{\beta})$ is used to construct confidence intervals. It is not required to eliminate the outliers in the determination of confidence intervals for the response variables **Y**. Outliers are only discarded during the estimation of the coefficients $\boldsymbol{\beta}$.

Finally the results of using a single or multiple control variates in combination with OLS and LMS are compared, based on the *net variance ratio*. The net variance ratio is defined as the ratio of the variance of **Y** before and after variance reduction.

3. EXPERIMENTAL MODELS

The method described in section 2 is tested on several simulation models. A job shop system with priorities is studied in section 3.1. The method is applied to a production-inventory system with an unreliable production facility in section 3.2.

3.1 Job Shop System

The job shop under consideration consists of a single person who needs to pack regular jobs and rush jobs. The interarrival time of regular jobs follows an exponential distribution with an average of 3 minutes. Rush jobs also arrive according to an exponential distribution, with an average interarrival time of 10 minutes. Rush jobs have a higher priority than regular jobs and are therefore handled first. The time required to pack a job is exponentially distributed with an average of 5 minutes, irrespective of the job type. The job shop system is simulated during 480 minutes. Seven simulation runs are performed. The observed response variables are number of regular jobs packed (RV1), number of rush jobs packed (RV2), system time of regular jobs (RV3) and system time of rush jobs (RV4). Potential control variates to reduce the variance of the response variables are interarrival time of regular jobs (CV1), interarrival time of rush jobs (CV2) and service time of the packer (CV3). Due to the limited number of replications, the number of control variates used in variance reduction is restricted to two.

We first investigate which variable is most appropriate if only one control variate is used to reduce the variance of the four response variables. Table 1 shows the correlation matrix between response variables and control variates.

Table 1: Correlation Matrix

	CV1	*CV2*	*CV3*
RV1	-0.486	0.123	-0.859
RV2	0.079	-0.988	-0.476
RV3	0.025	0.437	0.313
RV4	-0.120	0.768	0.360

The control variates are evaluated according to the following criterion.

$$Min \quad \sum_i w_i (1-\rho^2_{Y_i C}) \cdot (K-2)/(K-3) \qquad (9)$$

Results are reported in table 2.

Table 2: Selection of a single Control Variate (Q=1)

Control Variate	*Criterion Value*
CV2	0.8461
CV3	1.0373
CV1	1.3922

The control variates in table 2 are ranked in increasing order of the criterion value. The interarrival time of rush jobs (CV2) is indicated as the best choice of control variate. Robust regression is applied to estimate the coefficients $\boldsymbol{\beta}$. The method enables us to obtain better estimates of the weights for the control variate by eliminating potential outliers. The variance of the response variables before and after variance reduction are given in table 3. The net variance ratio is calculated as column 3 divided by column 2. The last column of the table mentions the weights given to each response variable.

Table 3: Results of Variance Reduction with a single Control Variate (Q=1)

	Variance without CV	*Variance with CV*	*Net Variance Ratio*	w_i
RV1	93.62	93.62	1.0000	0.2750
RV2	33.48	0.79	0.0235	0.1703
RV3	440.53	440.53	1.0000	0.1547
RV4	16.56	6.80	0.4103	0.4000

The net variance ratio gives the factor by which the variance of a response variable is reduced. The system time of rush jobs (RV4) has been given the largest weight. Table 3 shows that its variance has been reduced by 59%. Interarrival time of rush jobs (CV2) is also a very good control variate for the number of rush jobs serviced (RV2). Its variance has notably decreased. It does not appear to be a good control variate for the number of regular jobs packed (RV1) or the system time of regular jobs (RV3). The coefficients $\boldsymbol{\beta}$ are not different from zero on a 5% level of significance, even after robust regression. A global net variance ratio can be calculated by multiplying the net variance ratio and the weight of each response variable and summing the results. The use of a single control variate gives a global net variance ratio of 0.5978.

Next, the potential advantage of using an additional control variate is investigated. Table 4 gives the multiple correlation coefficients between the response variables and all combinations of control variates.

Table 4: Multiple Correlation Coefficients of two Control Variates (Q=2)

	CV1 & CV2	*CV1 & CV3*	*CV2 & CV3*
RV1	0.492	0.885	0.985
RV2	0.988	0.539	0.994
RV3	0.442	0.324	0.465
RV4	0.769	0.441	0.772

A pair of variates is selected based on the criterion:

$$Min \quad \sum_i w_i (1 - R^2_{Y_i C}) \cdot (K-2)/(K-4) \qquad (10)$$

The results in table 5 are ranked in increasing order. The best combination of control variates is the interarrival time of rush jobs (CV2) and the service time of the packer (CV3). These are also the top two ranked control variates in the selection of a single control variate.

Table 5: Selection of two Control Variates (Q=2)

Control Variates	*Criterion Value*
CV2 & CV3	0.8792
CV1 & CV2	1.5012
CV1 & CV3	1.9233

Table 6 reports on the variance of the response variables before and after variance reduction with the use of two control variates.

Table 6: Results of Variance Reduction with two Control Variates (Q=2)

	Variance without CV	*Variance with CV*	*Net Variance Ratio*	w_i
RV1	93.62	2.75	0.0294	0.2750
RV2	33.48	0.53	0.0157	0.1703
RV3	440.53	440.53	1.0000	0.1547
RV4	16.56	6.80	0.4103	0.4000

A comparison of the results of table 3 with table 6 shows that most gain is obtained in the variance of the first response variable. The additional control variate 'service time' is capable of reducing the variance of the number of regular jobs packed by 97%. The gain in variance reduction of other response variables is limited. The global net variance ratio when applying variance reduction with two control variates equals 0.3295. So most variance reduction is realised when applying a single control variate (40%). An additional gain of 27% results from using a second control variate.

3.2 Production-Inventory system

Posner and Berg (1989) incorporate unreliability factors into the analysis of production-inventory systems. The authors allow for breakdowns of machines and resulting repair actions. We simulate a model consisting of a single production machine with a constant production rate of one unit per minute. The machine operates as long as the accumulated inventory is below a predetermined threshold and is idle otherwise. The operating time until failure is exponentially distributed with an average of 9 minutes. Repair time of a failed machine follows an exponential distribution with an average of 3 minutes. Customer demands are described as a compound Poisson process. Demands arrive according to a Poisson process at a rate of 0.3 customers per minute. Demand quantities are exponentially distributed with an average of 3 units. A demand that cannot be fully satisfied from the existing inventory takes the remaining units and the rest of the demand is lost.

The observed response variables are service level (RV1), machine utilization (RV2) and lost sales (RV3). Service level is defined as the ratio of the number of customers fully serviced on the total number of customers. Machine utilization represents the percentage of time the machine is operating. Lost sales measures the total number of units short during one replication. Potential control variates are interarrival time of customers (CV1), demand quantity (CV2), interarrival time of breakdowns (CV3) and repair time (CV4). Seven replications of each 720 minutes are performed.

First, the use of a single control variate is analyzed. Table 7 shows the correlation matrix between response variables and control variates.

Table 7: Correlation Matrix

	CV1	*CV2*	*CV3*	*CV4*
RV1	0.103	-0.724	-0.087	-0.188
RV2	-0.542	-0.085	-0.227	-0.336
RV3	-0.460	0.490	-0.203	-0.231

Criterion (9) is used to rank the control variates. In table 8 the quantity demanded by customers (CV2) appears to be the best control variate.

Table 8: Selection of a single Control Variate (Q=1)

Control Variate	*Criterion Value*
CV2	1.1420
CV1	1.1863
CV4	1.4170
CV3	1.4390

Coefficients $\boldsymbol{\beta}$ are estimated by making use of robust regression. Results of the variance reduction with demand quantity (CV2) as a single control variate are given in table 9.

Table 9: Results of Variance Reduction with a single Control Variate (Q=1)

	Variance without CV	Variance with CV	Net Variance Ratio	w_i
RV1	0.0010	0.0010	1.0000	0.0297
RV2	0.0012	0.0012	1.0000	0.0425
RV3	1182.46	931.03	0.7874	0.9278

The initial variation of the third response variable 'lost sales' is much larger then for the other two response variables. Therefore, lost sales receives the largest weight in the last column of table 9. Consequently, the selection of a control variate focuses on reducing the variance in lost sales. Demand quantity is selected because it has the highest correlation with lost sales. After robust regression the coefficients $\boldsymbol{\beta}$ for the first two response variables are not significant and their variance is not reduced. This results in a global net variance ratio of 0.8027.

Next, the potential benefit of applying two control variates is investigated. An overview of all possible combinations of control variates and the associated multiple correlation coefficients is given in table 10.

Table 10: Multiple Correlation Coefficients of two Control Variates (Q=2)

	CV1 & CV2	CV1 & CV3	CV1 & CV4	CV2 & CV3	CV2 & CV4	CV3 & CV4
RV1	0.733	0.237	0.531	0.724	0.727	0.242
RV2	0.546	0.574	0.589	0.235	0.338	0.359
RV3	0.679	0.483	0.550	0.555	0.653	0.231

Table 11 shows the ranking of these combinations of control variates according to criterion (10). The best combination of control variates consists of demand quantity (CV2) and interarrival time of customers (CV1). These control variates are also ranked first in the analysis of a single control variate in table 8.

Table 11: Selection of two Control Variates (Q=2)

Control Variates	Criterion Value
CV1 & CV2	1.6308
CV2 & CV4	1.7514
CV1 & CV4	2.0887
CV2 & CV3	2.0889
CV1 & CV3	2.3037
CV3 & CV4	2.8298

Variances of the response variables before and after variance reduction with two control variates are reported in table 12.

Variance of lost sales (RV3) is reduced by the same amount after controlling for demand quantity. The additional control variate 'interarrival time of customers' reduces only the variance of machine utilization (RV2).

Table 12: Results of Variance Reduction with Two Control Variates (Q=2)

	Variance without CV	Variance with CV	Net Variance Ratio	w_i
RV1	0.0010	0.0010	1.0000	0.0297
RV2	0.0012	0.0009	0.7799	0.0425
RV3	1182.46	920.57	0.7785	0.9278

The combination of demand quantity and interarrival time of customers does not lead to significant coefficients $\boldsymbol{\beta}$ for reducing the variance of the service level offered to customers (RV1). Finally a global net variance ratio of 0.7852 is obtained. In conclusion, a variance reduction of 20% is realised after applying a single control variate 'demand quantity'. An additional reduction of only 2% can be acquired by using the two control variates 'demand quantity' and 'interarrival time of customers'.

4. CONCLUSIONS

In this paper the selection of control variates in a multiresponse simulation with a small number of replications is investigated. This can be done by giving a weight to the response variables according to their coefficient of variation. In this way more effort is directed towards response variables with a higher relative variance. An initial ranking of control variates is attained by summing their weighted minimum variance ratios. When deciding how many control variates to use, the loss factor also has to be taken into account. The loss factor measures the loss in variance reduction due to the estimation of the coefficients $\boldsymbol{\beta}$.

In practice this comprehensible selection method proves to be effective and computationally efficient. In the analysis of experimental models the best combination of two control variates are the first two mentioned in the initial ranking. Most variance reduction is achieved after applying a single control variate. The use of a second control variate sometimes only leads to a restricted gain in variance reduction. This confirms the advise of limiting the number of control variates, especially with a small number of replications.

Robust regression is applied to eliminate the influence of outliers and achieve more significant estimates of the coefficients $\boldsymbol{\beta}$. When the number of replications is small, outliers can seriously distort the estimation of the optimal weights for the control variates. The least median of squares method proves to be highly effective for detecting outliers. A re-weighted OLS regression is performed, in which outliers receive a zero weight.

Further research can be done into testing the method on different simulation systems with other characteristics. A better approximation of the loss factor when the normality assumption does not hold, can be examined.

REFERENCES

Avramidis, A.N. and J.R. Wilson. 1996. "Integrated variance reduction strategies for simulation". *Operations Research* 44, No. 2, 327-346.

Crawford, J.W. and T.J. Gallwey. 2000. "Bias and variance reduction in computer simulation studies". *European Journal of Operational Research* 124, 571-590.

Emsermann, M. and B. Simon. 2002. "Improving simulation efficiency with quasi control variates". *Stochastic Models* 18, No. 3, 425-448.

Janssens, G.K.; W. Deceuninck; A. Van Breedam. 1995. "Opportunities of robust regression for variance reduction in discrete event simulation". *Journal of Computational and Applied Mathematics* 64, 163-176.

Lavenberg, S.S.; T.L. Moeller; P.D. Welch. 1982. "Statistical Results on Control Variables with Application to Queueing Network Simulation". *Operations Research* 30, No. 1, 182-202.

Law, A.M. and W.D. Kelton. 2000. *Simulation Modeling & Analysis*, 3rd edition. McGraw-Hill, New York.

Nelson, B.L. 1990. "Control variate remedies". *Operations Research* 38, No. 6, 974-992.

Posner, M.J.M. and M. Berg. 1989. "Analysis of a production-inventory system with unreliable production facility". *Operations Research Letters* 8, 339-345.

Porta Nova, A.M. 1993. "Selecting control variates to estimate multiresponse simulation metamodels". *European Journal of Operational Research* 71, 80-94.

Rousseeuw, P.J. 1984. "Least Median of Squares Regression". *Journal of the American Statistical Association* 79, 871-880.

Rousseeuw, P.J. and A.M. Leroy. 1987. *Robust Regression and Outlier Detection*. John Wiley & Sons, New York.

Rubinstein, R.Y. and R. Marcus. 1985. "Efficiency of Multivariate Control Variates in Monte Carlo Simulation". *Operations Research* 33, No. 3, 661-677.

Tew, J.D. 1995. "Simulation metamodel estimation using a combined correlation-based variance reduction technique for first and higher-order metamodels". *European Journal of Operational Research* 87, 349-367.

BIOGRAPHY

GERRIT K. JANSSENS (°1956) received degrees of M.Sc. in Engineering with Economy from the University of Antwerp (RUCA), Belgium, M.Sc. in Computer Science from the University of Ghent (RUG), Belgium, and Ph.D. from the Free University of Brussels (VUB), Belgium. After some years of work at General Motors Continental, Antwerp, he joined the University of Antwerp until the year 2000. Currently he is Professor of Operations Management and Logistics at the Hasselt University (UHasselt) within the Faculty of Business Administration. He also holds the CPIM certificate of the American Production and Inventory Control Society (APICS). During the last fifteen years he has been several times visiting faculty in universities in Thailand, Vietnam, Cambodia and Zimbabwe. His main research interests include the development and application of operations research models in production and distribution logistics.

AN CARIS (°1980) obtained the degree of Commercial Engineer at the Limburg University Centre (LUC) in 2003. As major she chose Operations Management and Logistics. After one year of practical experience in inventory management at Reynaers Aluminium, she started as an assistant for econometrics and operations research at the Hasselt University (UHasselt) in September 2004. In addition she prepares to acquire a doctoral degree in Applied Economic Sciences. She is a member of the research group Data Analysis and Modelling and the Transportation Research Institute (IMOB) of the Hasselt University. She takes a research interest in modelling intermodal freight transport networks with a focus on inland navigation.