Queen's Economics Department Working Paper No. 803

Regression-based Methods for Using Control Variates in Monte Carlo Experiments

Russell Davidson
Queen's University and GREQE–EHESS

James G. MacKinnon
Queen's University

Department of Economics
Queen's University
94 University Avenue
Kingston, Ontario, Canada
K7L 3N6

(revised 1991-10)

Regression-based Methods for Using Control Variates in Monte Carlo Experiments*

Russell Davidson and James G. MacKinnon

Department of Economics
Queen's University
Kingston, Ontario, Canada
K7L 3N6

and

GREQE–EHESS
Centre de la Vieille Charité
2 Rue de la Charité
13002 Marseille, France

Department of Economics
Queen's University
Kingston, Ontario, Canada
K7L 3N6

Abstract

Methods based on linear regression provide an easy way to use the information in control variates to improve the efficiency with which certain features of the distributions of estimators and test statistics are estimated in Monte Carlo experiments. We propose a new technique that allows these methods to be used when the quantities of interest are quantiles. We also propose new ways to obtain approximately optimal control variates in many cases of interest. These methods seem to work well in practice, and can greatly reduce the number of replications required to obtain a given level of accuracy.

This paper appeared in *Journal of Econometrics*, **54**, 1992, 203–222. However, the online PDF is defective.

* This is a revised version of a paper originally entitled "A general procedure for using control and antithetic variates in Monte Carlo experiments". We are grateful to David Andrews, four referees, an Associate Editor, and seminar participants at UCSD, Berkeley, UCLA, USC, Queen's, McGill, Rochester and the University of Montreal for comments on earlier versions. This research was supported, in part, by grants from the Social Sciences and Humanities Research Council of Canada.

1. Introduction

Monte Carlo methods are widely used to study the finite-sample properties of estimators and test statistics. Hendry (1984) provides a survey of Monte Carlo methods in econometrics, and Ripley (1987) and Lewis and Orav (1989) provide recent treatments from the perspectives of statistics and operations research respectively. The results of Monte Carlo experiments are inevitably random, since they depend on the particular set of pseudo-random numbers used. To reduce this randomness to acceptable levels it is often necessary to perform many replications. A less costly way to reduce experimental randomness is to use variance reduction techniques, such as control variates. These are random variables that are correlated with the estimator or test statistic being studied, and of which certain properties of the distribution are known. Control variates can be calculated only in the context of Monte Carlo experiments, because they depend on things that cannot be observed in actual statistical investigations. The primary property that a control variate must have is a known (population) mean. The divergence between the sample mean of the control variate in the experiment and its known population mean is then used to improve the estimates from the Monte Carlo experiment. This works best if the control variate is highly correlated with the estimators or test statistics with which the experiment is concerned. For examples of control variates in econometric problems, see Hendry (1984) and Nankervis and Savin (1988).

In this paper we discuss a widely applicable, and yet very simple, procedure for using control variates to analyze the results of Monte Carlo experiments. The heart of this procedure is a least squares regression. This procedure has been discussed elsewhere—Ripley (1987) and Lavenberg and Welch (1981) are two good references—but it is not covered in Hendry's survey and appears to be unfamiliar to most econometricians. In the next section we therefore discuss it briefly. The principal results of the paper are then introduced in sections 3 and 4. In the former, a modified version of the regression procedure is introduced that can be used to estimate quantiles. In the latter, we discuss how to choose control variates in an (approximately) optimal fashion. In Section 5, we present some evidence on how well these techniques actually work, and in Section 6 we present an example of how useful they can be in practice.

2. Regression-based Methods for Using Control Variates

Suppose that a Monte Carlo experiment involves N replications, on each of which is obtained an estimate t_j, $j = 1, \ldots, N$, of some scalar quantity θ. Except when discussing quantile estimation, we shall suppose that θ is capable of being estimated as the mean of the N t_j calculated during the experiment. Obvious examples of θ include the bias, variance, skewness or kurtosis of a particular parameter estimate or test statistic. If, for example, θ were the bias of some estimator, t_j would be the estimate obtained on the j^{th} replication, minus its true value. Another possibility is that θ might be the size or power of a test statistic, that is, the probability that it exceeds a certain critical value. In such a case t_j would be unity if the test rejected the null hypothesis and zero otherwise. Since quantiles (such as medians, or critical values for test statistics) cannot be estimated as the mean of anything, they do not seem to fit into this general scheme. They require a somewhat different treatment, as we explain in Section 3.

It is always possible to estimate θ without using control variates. The obvious estimator is the sample mean of the t_j,

$$\bar{\theta} \equiv \frac{1}{N} \sum_{j=1}^{N} t_j,$$

which has variance $V(\bar{\theta}) = N^{-1}V(t)$. Now suppose that, for each replication, a control variate τ_j is computed along with the estimate t_j. The two primary requirements for τ_j are that it be correlated with t_j and that it have mean zero. Any random variable that is correlated with t_j and has known mean can be transformed so as to have mean zero, and could thus be used for τ_j. This suggests that there may often be several candidates for τ_j, but for the moment we shall assume that just one is to be used.

One way to write the control variate (CV) estimator for this case is

$$\ddot{\theta}(\lambda) \equiv \bar{\theta} - \lambda\bar{\tau}, \tag{1}$$

where $\bar{\tau}$ is the sample mean of the τ_j and λ is a scalar that has to be determined. The choice of λ is crucial. It seems natural to choose it so as to minimize the variance of (1):

$$V\big(\ddot{\theta}(\lambda)\big) = N^{-1}\big(V(t) + \lambda^2 V(\tau) - 2\lambda\mathrm{Cov}(t,\tau)\big). \tag{2}$$

Minimizing (2) with respect to λ, we find that the optimal value of λ is

$$\lambda^* = \frac{\mathrm{Cov}(t,\tau)}{V(\tau)}, \tag{3}$$

so that (1) becomes

$$\ddot{\theta}(\lambda^*) \equiv \bar{\theta} - \lambda^*\bar{\tau} = \bar{\theta} - \frac{\mathrm{Cov}(t,\tau)}{V(\tau)}\bar{\tau}. \tag{4}$$

In much of the literature on control variates (e.g. Hendry, 1984) λ is set to unity. This may seem reasonable if the control variate is based on an asymptotic expansion, and in fact $\lambda = 1$ will be a good choice if t_j and τ_j are highly correlated and have similar variances. But it is clearly not the best choice in general. Unless N is very small, so that $\hat{\lambda}$ (defined below) may be a poor estimate of λ^*, or λ^* is very close to unity, we are likely to be better off estimating λ^* than simply using $\lambda = 1$.

The naive estimator $\bar{\theta}$ could have been obtained by regressing $\boldsymbol{t}$, an $N-$vector with typical element t_j, on $\boldsymbol{\iota}$, an $N-$vector of ones. Let $\boldsymbol{\tau}$ denote an $N-$vector with typical element τ_j. Then a better way to estimate θ is to run the regression

$$\boldsymbol{t} = \theta\boldsymbol{\iota} + \lambda\boldsymbol{\tau} + \boldsymbol{u}. \tag{5}$$

This regression bears a striking resemblance to the artificial regression proposed by Tauchen (1985) for computing certain test statistics. The error terms u_j in (5) will be independent as long as each replication is independent. The estimate of λ from regression (5) is

$$\hat{\lambda} = (\boldsymbol{\tau}^\top \boldsymbol{M}_{\boldsymbol{\iota}} \boldsymbol{\tau})^{-1} \boldsymbol{\tau}^\top \boldsymbol{M}_{\boldsymbol{\iota}} \boldsymbol{t},$$

where $\boldsymbol{M}_{\boldsymbol{\iota}}$ is the matrix $\mathbf{I} - \boldsymbol{\iota}(\boldsymbol{\iota}^\top\boldsymbol{\iota})^{-1}\boldsymbol{\iota}^\top$ that takes deviations from the mean. Evidently $\hat{\lambda}$ is just the sample covariance of $\boldsymbol{t}$ and $\boldsymbol{\tau}$, divided by the sample variance of $\boldsymbol{\tau}$, so that it is

the empirical counterpart of λ^* defined in (3). Provided that $N^{-1}\boldsymbol{\tau}^\top \boldsymbol{M}_\iota \boldsymbol{\tau}$ and $N^{-1}\boldsymbol{\tau}^\top \boldsymbol{M}_\iota \boldsymbol{t}$ obey laws of large numbers, it is clear that $\hat{\lambda}$ will converge to λ^*.

What we are really interested in is the OLS estimate of θ from regression (5). From standard results for linear regressions with a constant term, this is $\hat{\theta} = \bar{\theta} - \hat{\lambda}\bar{\tau}$. Subtracting the true value θ_0 and multiplying by $N^{1/2}$, we find that

$$N^{1/2}(\hat{\theta} - \theta_0) = N^{1/2}(\bar{\theta} - \theta_0) - \hat{\lambda}(N^{1/2}\bar{\tau}).$$

Since $\hat{\lambda}$ converges to λ^*, $\hat{\theta}$ must be asymptotically equivalent to the optimal CV estimator defined in (4). Moreover, since $N^{1/2}(\bar{\theta} - \theta_0)$ and $N^{1/2}\bar{\tau}$ are asymptotically normally distributed, $\hat{\theta}$ must be as well.

The variance of $\hat{\theta}$ can be estimated in at least two ways. The usual OLS estimate is

$$\hat{\omega}^2(\boldsymbol{\iota}^\top \boldsymbol{M}_\tau \boldsymbol{\iota})^{-1},$$

where $\hat{\omega}$ is the standard error of the regression.[1] The second factor here must tend to N^{-1}, since $\boldsymbol{\tau}$ asymptotically has no explanatory power for $\boldsymbol{\iota}$, so that we could simply use $N^{-1}\hat{\omega}^2$ to estimate the variance of $\hat{\theta}$. This makes it clear that the better regression (5) fits (at least asymptotically), the more accurate $\hat{\theta}$ will be as an estimate of θ. Thus we want to choose control variates to make it fit as well as possible. In fact, the ratio of the asymptotic variance of $\hat{\theta}$ to the asymptotic variance of $\bar{\theta}$ is simply equal to $1 - R^2$, where R^2 is the asymptotic R^2 from regression (5).

As we remarked above, there may well be more than one natural choice for τ in many situations. Luckily, formulating the problem as a linear regression makes it obvious how to handle multiple control variates. The appropriate generalization of (5) is

$$\boldsymbol{t} = \theta\boldsymbol{\iota} + \boldsymbol{T}\boldsymbol{\lambda} + \boldsymbol{u}, \tag{6}$$

where $\boldsymbol{T}$ is an $N \times c$ matrix, each column of which consists of observations on one of c control variates. Since all the columns of $\boldsymbol{T}$ have expectation zero, it is clear that the OLS estimate of θ from this regression will once again provide the estimate we are seeking. If $\boldsymbol{M}_T$ denotes $\mathbf{I} - \boldsymbol{T}(\boldsymbol{T}^\top\boldsymbol{T})^{-1}\boldsymbol{T}^\top$, this estimate is

$$\hat{\theta} = (\boldsymbol{\iota}^\top \boldsymbol{M}_T \boldsymbol{\iota})^{-1}\boldsymbol{\iota}^\top \boldsymbol{M}_T \boldsymbol{t}.$$

The error terms in regressions (5) and (6) will often not be conditionally homoskedastic. Thus it might be thought better to use as an estimate of the standard error of $\hat{\theta}$, not the usual OLS estimate, but rather a heteroskedasticity-consistent one of the sort proposed by White (1980). In fact this is unnecessary, and may actually be harmful when N is small. We are concerned here only with the variance of the estimate of the constant. Since the other regressors are all by construction orthogonal to the constant in the population, this variance is in fact estimated consistently by the appropriate element of the ordinary OLS covariance matrix; see White (1980). Of course, the OLS standard errors for the remaining coefficients in (5) and (6) will not, in general, be consistent when the error terms are heteroskedastic, but these coefficients are generally not of interest.

[1] The reason we use ω rather than σ here will become apparent in Section 4.

The regression technique we have described can be used with antithetic variates as well as control variates. The idea of antithetic variates is to generate two different estimates of θ, say t_{j1} and t_{j2}, on each replication, in such a way that they will be highly negatively correlated; see Hammersley and Handscomb (1964) or Ripley (1987). For example, if one were interested in the bias of the least squares estimates of a nonlinear regression model, one could use each set of generated error terms twice, with all signs reversed the second time. This would create strong negative correlation between the two sets of least squares estimates, so that their average would have much less variance than either one alone.

By definition, t_{j1} and t_{j2} both have mean θ. Hence their difference must have mean zero, and can be treated like a control variate in regressions (5) or (6). One simply treats either $\boldsymbol{t}_1$ or $\boldsymbol{t}_2$ as the regressand (it does not matter which), and treats $\boldsymbol{t}_2 - \boldsymbol{t}_1$ as a control variate. However, the simpler approach of just averaging the t_{j1} and t_{j2}, which is equivalent to setting $\lambda = \frac{1}{2}$, is optimal whenever the two antithetic variates have the same variance, as will generally be the case; see Hendry and Trivedi (1972). Thus it probably does not make sense to employ the regression technique with antithetic variates unless one wishes to use both control and antithetic variates at the same time.

One aspect of the regression approach may be troubling in a few cases. It is that $\hat{\theta}$ is only asymptotically equal to $\ddot{\theta}(\lambda^*)$, and is only one of many asymptotically equivalent ways to approximate the latter. This should rarely be of concern, since in practice N will generally be quite large. In Section 5, we describe some experimental results on how well this procedure works for finite N.

3. Quantile Estimation

The estimation of quantiles is often one of the objects of a Monte Carlo experiment. For example, one may wish to characterize a distribution by its estimated quantiles, or to estimate critical values for a test statistic. Since a quantile cannot be expressed as a mean, control variates cannot be used directly in the way discussed above to improve the precision of quantile estimates. In this section we propose a control variate regression that does much the same for quantiles as regression (6) does for means, variances, tail areas and so on.

Suppose that we generate observations on a random variable y with a distribution characterized by the (unknown) distribution function $F(y)$, and wish to estimate the α quantile of the distribution F. By this we mean the value q_α that satisfies

$$F(q_\alpha) = \alpha. \tag{7}$$

In the absence of any information about F, the sample quantile $\bar{q}_\alpha$ is the most efficient consistent estimator of q_α available. It is defined by replacing F in (7) by the empirical distribution of the generated y:

$$N^{-1} \sum_{j=1}^{N} I(\bar{q}_\alpha - y_j) = \alpha, \tag{8}$$

where the indicator function I is defined to be equal to one if its argument is positive, and zero otherwise. If $\bar{q}_\alpha$ is not uniquely defined by (8), we may take for $\bar{q}_\alpha$ the mean of the set of numbers that satisfy (8). If αN is not an integer, there will be no $\bar{q}_\alpha$ that exactly satisfies (8). This problem can easily be dealt with, but since N is chosen by the investigator, we shall simply assume that αN is an integer.

Suppose that we have some estimate $\ddot{q}_\alpha$, independent of y, which approaches q_α at a rate proportional to $N^{-1/2}$. Now consider the random variable $I(\ddot{q}_\alpha - y) - \alpha$. The mean of this variable conditional on $\ddot{q}_\alpha$, using (7), is

$$E\big(I(\ddot{q}_\alpha - y)\big) - \alpha = F(\ddot{q}_\alpha) - F(q_\alpha). \tag{9}$$

If the density $f(q_\alpha) \equiv F'(q_\alpha)$ exists and is nonzero, the mean (9) becomes, by Taylor's Theorem,

$$f(\ddot{q}_\alpha)(\ddot{q}_\alpha - q_\alpha) + o(N^{-1/2}). \tag{10}$$

The quantity $f(\ddot{q}_\alpha)$ is not known in advance, but it may be estimated in a variety of ways; see Silverman (1986). One approach that seems to work well is kernel estimation (Rosenblatt, 1956). Since the density has to be estimated at a single point only, the calculations are not very demanding. Let us denote the estimate of $f(\ddot{q}_\alpha)$ by $\ddot{f}$. Provided it has the property that $\ddot{f} = f(q_\alpha) + o(1)$, we see from (10) that

$$(1/\ddot{f})E\big(I(\ddot{q}_\alpha - y) - \alpha\big) = \ddot{q}_\alpha - q_\alpha + o(N^{-1/2}),$$

or, equivalently,

$$\ddot{q}_\alpha - (1/\ddot{f})E\big(I(\ddot{q}_\alpha - y) - \alpha\big) = q_\alpha + o(N^{-1/2}). \tag{11}$$

This result allows us to construct a regression. The j^{th} observation on the regressand is

$$\ddot{q}_\alpha - (1/\ddot{f})\big(I(\ddot{q}_\alpha - y_j) - \alpha\big). \tag{12}$$

The regressors must include a constant and one or more control variates, of which it need be known merely that their expectations are zero. All of the arguments of the preceding section go through unaltered, and the estimated constant from the regression will, by (11), be an estimate of q_α correct to the leading asymptotic order of $N^{-1/2}$. We shall call this control variate estimator $\hat{q}_\alpha$.

The above analysis supposed the independence of the y_j and the preliminary estimate $\ddot{q}_\alpha$, but since the argument is an asymptotic one, it is admissible to replace strict independence with asymptotic independence. Thus it is possible in practice to use the ordinary quantile estimate $\bar{q}_\alpha$ for $\ddot{q}_\alpha$.

If the regressand (12) with $\ddot{q}_\alpha = \bar{q}_\alpha$ is regressed on a constant only, the estimate $\hat{q}_\alpha$ will be equal to $\bar{q}_\alpha$, because what is subtracted from $\bar{q}_\alpha$ in (12) is orthogonal to a constant. The estimated variance of $\hat{q}_\alpha$ will be N^{-1} times the estimated error variance from the regression, that is

$$\left(\frac{1}{N(N-1)\bar{f}^2}\right)\sum_{j=1}^{N}\Big(I(\bar{q}_\alpha - y_j) - \alpha\Big)^2.$$

For large N this tends to

$$\frac{\alpha(1-\alpha)}{Nf^2(q_\alpha)}, \tag{13}$$

which is the standard formula for the variance of a sample quantile. When the regression includes one or more control variates that have some explanatory power, the variance of $\hat{q}_\alpha$ will of course be less than (13).

An alternative approach has been used in the operations research literature (Lewis and Orav, 1989, Chapter 11). The N replications are sectioned into m groups, each consisting

of M replications. One then calculates quantile estimates, and control variates, for each of the m groups, and adjusts the average quantile estimate by regressing it on a constant and the average value of the control variate, using a regression with m observations. This approach avoids the need to estimate $1/\bar{f}$, but requires that both m and M (rather than just $N = mM$) be reasonably large. If M is too small, the individual quantile estimates may be seriously biased, while if m is too small, the estimates from the control variate regression may be unreliable.

4. Choosing Control Variates Optimally

So far we have said little about how to choose the control variate(s) to be used as regressors in regression (6) or its analog for quantile estimation. In this section we consider a special, but not unrealistic, case in which it is possible to obtain strong results on how to choose control variates optimally. These results are of considerable interest, especially for the estimation of tail areas, where they shed light on the properties of earlier procedures for the use of control variates.

Let t denote the random variable that has expectation θ, and T the set of random variables with known distributions from which control variates are to be constructed. Consider the following generalization of (6):

$$t_j = \theta + g(T_j, \boldsymbol{\gamma}) + \text{residuals}, \tag{14}$$

where the nonlinear regression function $g(T, \boldsymbol{\gamma})$ is restricted to have mean zero, and $\boldsymbol{\gamma}$ is a vector of parameters. Estimation of (14) by nonlinear least squares would yield $\hat{\boldsymbol{\gamma}}$ and $\hat{\theta}$, the latter being a consistent estimator of θ. One way of defining the conditional expectation $E(t|T)$ is as the function $g(T)$ that minimizes the variance of $t - g(T)$. Hence the variance of the residuals in (14) will be minimized by using a control variate that is proportional to $\tau^* \equiv E(t|T) - \theta$, and consequently the precision of the estimate of θ will be maximized by this choice. This argument implies that the theoretical lower bound to the variance of a control-variate estimate of θ is proportional to the variance of $t - \tau^*$.

Of course, if $E(t|T)$ were known there would be no need to estimate θ by Monte Carlo: it would be enough just to calculate $E(t) = E\big(E(t|T)\big)$. Further, in order to compute τ^* we need to know θ, which is precisely what the Monte Carlo experiment is trying to estimate! Thus in practice τ^* will not be available. Instead, we want to find functions of T that approximate τ^* as closely as possible. Thus we should make use of as much *a priori* information as possible about the relationship between t and T when specifying $g(T, \boldsymbol{\gamma})$. In cases where not much is known about that relationship, it may make sense to use several control variates in an effort to make $g(T, \boldsymbol{\gamma})$ provide a good approximation to τ^*.

In the remainder of this section we consider several related examples. We assume that the random variable of interest is normally distributed, and that another normally distributed random variable is available to provide control variates. These assumptions are not as unrealistic as they may seem, since in many cases asymptotic theory provides a control variate that is normally distributed and tells us that the variable of interest will be approximately normal; see Section 6 for an example. We shall let x denote the random variable from which we calculate control variates, and y denote the random variable of interest; various functions of x and y will later be denoted τ^* and t. We shall assume that x is distributed as $N(0, 1)$, that y is distributed as $N(\mu, \sigma^2)$, and that x and y are bivariate

normal. In this case T consists of all possible functions of x that have mean zero. Given our assumptions about x and y, we can write

$$y = \mu + \rho\sigma x + v, \quad v \sim N\big(0, (1-\rho^2)\sigma^2\big), \tag{15}$$

where ρ is the correlation between y and x. Using (15), we can find optimally-chosen control variates, as functions of x, for several cases of interest.

Suppose first that we wish to estimate μ, the mean of y. From (15) we see that $E(t|x) = \mu + \rho\sigma x$, which implies that $\tau^* = \rho\sigma x$, so that the optimal regressor in this case must be proportional to x. Thus we want t_j to equal y_j in regression (6). We also see that the variance of $\hat{\mu}$ will be $1/N$ times the variance of v, i.e. $N^{-1}(1-\rho^2)\sigma^2$. In contrast, the naive estimator of μ is the sample mean $\bar{y}$, and its variance is $N^{-1}\sigma^2$. Thus for any degree of accuracy, N could be smaller by a factor of $(1-\rho^2)$ if the optimally-chosen CV estimator were used instead of the naive estimator.

Now suppose that we wish to estimate σ^2. The obvious choice for t is $(y-\hat{\mu})^2$. From (15) we see that

$$(y-\mu)^2 = (\rho\sigma x + v)^2 = \rho^2\sigma^2x^2 + 2\rho\sigma xv + v^2, \tag{16}$$

which implies that

$$E\big((y-\mu)^2|x\big) = \rho^2\sigma^2x^2 + (1-\rho^2)\sigma^2. \tag{17}$$

The optimal regressor, adjusted to have mean zero, is evidently (x^2-1). From (16) and (17), the variance of the optimally-chosen CV estimator $\hat{\sigma}^2$ will be

$$N^{-1}E\Big(\big(\rho^2\sigma^2x^2 + 2\rho\sigma xv + v^2\big) - \big(\rho^2\sigma^2x^2 + (1-\rho^2)\sigma^2\big)\Big)^2 = 2N^{-1}(1-\rho^4)\sigma^4. \tag{18}$$

In contrast, the variance of the naive estimator is $2N^{-1}\sigma^4$. Thus for any degree of accuracy, N could be smaller by a factor of $(1-\rho^4)$ if the optimally-chosen CV estimate were used instead of the naive estimate. Note that the gain from using control variates is less when estimating the variance than when estimating the mean, since $(1-\rho^4)$ is greater than $(1-\rho^2)$ for all $|\rho| < 1$. This is true for estimating the standard deviation as well, since the variance of $\hat{\sigma}$ will be $\frac{1}{2}N^{-1}\sigma^2(1-\rho^4)$.

Now suppose that we are interested in the size or power of a test statistic. In this case θ is the probability that y exceeds a certain critical value, say y^c. Let $t_j = 1$ if y_j exceeds y^c and $t_j = 0$ otherwise. The naive estimate of θ is just the mean of the t_j. Davidson and MacKinnon (1981) and Rothery (1982) independently studied this problem under the assumption that the control variate, like the t_j, can take on only two possible values, and proposed a technique based on the method of maximum likelihood; see also Fieller and Hartley (1954). These authors did not use a regression framework, but the estimator they proposed turns out to be numerically identical to the OLS estimator of θ from regression (5), when τ_j is defined as

$$\tau_j = I(x - q_{1-\gamma}) - \gamma. \tag{19}$$

Thus τ_j is a binary variable that is equal to $1-\gamma$ when x_j exceeds $q_{1-\gamma}$, and $-\gamma$ otherwise, for some γ that should be as close to θ as possible. Since the probability that x_j will exceed $q_{1-\gamma}$ is γ, (19) clearly has population mean zero. There are various ways in which γ can be chosen. The simplest one is to pick it in advance, for example by setting $\gamma = 0.05$ when θ is the size of a test with an asymptotic size of 0.05. However, this method can be improved

upon, at least for large N, by choosing γ to be $\bar{\theta}$. The dependence of $\bar{\theta}$ on the data does not matter if N is large enough; see Davidson and MacKinnon (1981).

The expectation of $I(y - y^c)$ conditional on x is not a step function like (19). Thus the binary control variate proposed by ourselves and Rothery clearly cannot be optimal. In fact, when y is $N(\mu, \sigma^2)$ and x is $N(0, 1)$, we see that

$$E(t|x) = \text{Prob}(v > y^c - \mu - \rho\sigma x) = \Phi\left(\frac{\mu + \rho\sigma x - y^c}{\sigma(1-\rho^2)^{1/2}}\right)$$

$$= \Phi\big((\mu + \rho\sigma x - y^c)/\omega\big) = \Phi(a + bx),$$

where $a = (\mu - y^c)/\big(\sigma(1-\rho^2)^{1/2}\big)$, $b = \rho/(1-\rho^2)^{1/2}$, $\omega = \sigma(1-\rho^2)^{1/2}$ and Φ denotes the standard normal distribution function. There are many ways to estimate $\Phi(a + bx)$. The easiest is probably to recognize that the fitted values from the regression of y_j on x_j and a constant, which is the optimal CV regression for estimating the mean of the y_j, provide consistent estimates of $(\mu + \rho\sigma x)$, and the standard error provides a consistent estimate of ω. These estimates, along with the CV estimates of μ and σ, can then be used to construct a zero-mean control variate:

$$\Phi(\hat{a} + \hat{b}x_j) - \Phi\big((\hat{\mu} - y^c)/\hat{\sigma}\big). \tag{20}$$

The second term in (20) is an estimate of the unconditional mean of t; if μ and σ were known, it would be $\theta \equiv E(t)$. The fact that $\hat{a}$, $\hat{b}$, $\hat{\mu}$, and $\hat{\sigma}$ are estimates does not prevent expression (20) from having expectation zero if $\hat{a}$ and $\hat{b}$ are defined as above in terms of $\hat{\mu}$ and $\hat{\sigma}$, provided only that the x_j are normally distributed. Even if the assumption that y is normally distributed is not quite correct, (20) provides a valid control variate, and regressing the t_j on it and a constant should give a reasonably good estimate of θ. If we are interested in tail areas for two-tail tests, the approximately optimal regressor will be the sum of (20) and its analog for the other tail.

For large N, the variance of $\hat{\theta}$ will be

$$N^{-1}E\big(t_j - \Phi(a + bx_j)\big)^2.$$

Since the expectation of t_j conditional on x_j is $\Phi(a + bx_j)$, this expression reduces to

$$N^{-1}E\Big(\Phi(a + bx)\big(1 - \Phi(a + bx)\big)\Big), \tag{21}$$

which can easily be evaluated numerically. The corresponding expression for the naive estimator is $N^{-1}\theta(1 - \theta)$.

Finally, consider quantile estimation. The optimal regressor is the conditional expectation of $(1/\ddot{f})\big(I(\ddot{q}_\alpha - y_j) - \alpha\big)$. Since $1/\bar{f}$ is just a constant, it is unnecessary to include it in the optimal regressor, which is therefore

$$E\Big(\big(I(\bar{q}_\alpha - y) - \alpha\big)|x\Big).$$

This regressor would never be available in practice, but it may be approximated by

$$\Phi\left(\frac{\bar{q}_\alpha - \hat{\mu} - \hat{\rho}\hat{\sigma}x}{\hat{\sigma}(1-\hat{\rho}^2)^{1/2}}\right) - \Phi\left(\frac{\bar{q}_\alpha - \hat{\mu}}{\hat{\sigma}}\right). \tag{22}$$

This control variate is very similar to (20), the approximately optimal one for tail-area estimation, with $\bar{q}_\alpha$ replacing y^c and the signs of the arguments of $\Phi(\cdot)$ changed because we are now interested in a lower rather than an upper tail. Asymptotically, using (22) as a control variate when estimating q_α will produce the same proportional gain in efficiency as using (20) when estimating α.

In many cases, notably for tail areas and quantiles, regressions (6) and (14) will have error terms that are conditionally heteroskedastic. It might therefore seem appropriate to correct for this heteroskedasticity by using some form of weighted least squares. However, this turns out not to be the case. The problem is that weighted least squares will in general be consistent only if the regression model is correctly specified, that is if $g(T, \boldsymbol{\gamma})$ in (14) is equal to τ^* for some $\boldsymbol{\gamma}$. If $g(T, \boldsymbol{\gamma})$ is not specified correctly, the error terms will not be orthogonal to all possible functions of T. Since the weights will necessarily be chosen as functions of T, they may well be correlated with the error terms, thereby biasing the estimate of θ. In contrast, as we showed in Section 2, OLS yields a consistent estimate of θ under very weak conditions. Unless one knows enough about y to construct a regressor that is actually optimal, OLS appears to be the procedure of choice.

How much the number of replications can be reduced by the use of control variates, while maintaining a given level of accuracy, varies considerably. Table 1 presents some illustrative results for the case where x and y are both $N(0, 1)$. Each entry in the table is the ratio of the (asymptotic) variance for the naive estimator to that for a control variate estimator. This ratio is the factor by which the number of replications needed by the former exceeds the number needed by the latter. For μ and σ^2, only the ratio for the optimally-chosen control variate (OCV) estimator is reported, and these entries are simply $(1 - \rho^2)^{-1}$ and $(1 - \rho^4)^{-1}$ respectively. For the tail areas and quantiles (the same results apply to both), the ratio for the OCV estimator is reported first, followed by the ratio for the binary control variate (BCV) estimator discussed above.[2] Entries for the OCV estimator for tail areas (and quantiles) are the ratio of $\Phi(\alpha)\big(1-\Phi(\alpha)\big)$ to expression (21), which was evaluated numerically. Entries for the BCV estimator were calculated in a similar fashion.

It is evident from Table 1 that the gains from using control variates can be very substantial when y and x are highly correlated. They are greatest when estimating the mean and least when estimating small-probability tail areas and quantiles, where the OCV estimators do however always outperform the BCV ones quite handily. Provided that $\rho^2 \geq .9$, a level of correlation between the control variate and the variable of interest that is not unrealistic, there is always a gain of at least a factor of two when the optimally-chosen control variate is used. Thus it appears that it will often be worthwhile to use control variates.

---

[2] The results for the binary control variate assume that the value of γ used to construct it is the same as the probability θ that $y > y^c$. The binary control variate will work less well when this assumption is not satisfied. It is always approximately satisfied when estimating quantiles, and can be approximately satisfied when estimating tail areas if one sets $\gamma = \bar{\theta}$.

5. Simulation Evidence

Since the theoretical arguments of this paper have all been asymptotic, one may well wonder whether the various CV estimators that we have discussed are in fact reliable for reasonable values of N. By "reliable" we mean that $\delta\%$ confidence intervals based on normal theory and the estimated standard errors for $\hat{\theta}$ should cover the true values approximately $\delta\%$ of the time. One may also wonder whether the gains from the use of control variates implied by the theory of Section 4 can actually be realized in practice.

To investigate these questions, we performed several Monte Carlo experiments designed to simulate the commonly encountered situation in which y is an estimator or test statistic and x is the random variable to which it tends asymptotically. Each replication in one of these experiments corresponds to a Monte Carlo experiment with N replications, in which various properties of y are estimated using functions of x as control variates. In the experiments x was distributed as $N(0,1)$ and y was distributed as Student's t with numerator equal to x and number of degrees of freedom d equal to 5, 10 or 30. As d increases, the correlation between x and y increases, and the distribution of y becomes closer to $N(0,1)$. We performed 10,000 replications for $N = 500$ and 5000 replications for each of $N = 1000$ and $N = 2000$. Some of the results are shown in Table 2.

Regression-based control variate methods evidently work extremely well for estimation of the mean. The 95% confidence intervals always cover the true value just about 95% of the time, with the CV confidence intervals just as reliable as the naive ones. The CV estimates are however much more efficient than the naive ones, by factors that are almost exactly what the theory of Section 4 predicts given the observed correlations between x and y. (For $\rho^2 = .850$, .940 and .982 respectively, $1/(1-\rho^2)$ is 6.67, 16.67 and 55.56.) Thus in this case, the fact that y is Student's t rather than normal does not seem to matter at all.

The results for estimation of the variance are not so good. The 95% confidence intervals now tend to cover the true value somewhat less than 95% of the time, especially when d is small. This is true for both the naive and CV estimates, but is more pronounced for the latter. The efficiency gains are also much less than one would expect given the observed correlations between x and y. For example, if y were normally distributed, a ρ^2 of .982 (for the case where $d = 30$) should imply that the CV estimate of variance is about 28 times as efficient as the naive estimate, while in fact it is about 9.5 times as efficient. Nevertheless, the gains from using control variates are substantial except when $d = 5$.

The results for tail area estimation depend on what tail is being estimated. Every technique works better as θ gets closer to $\frac{1}{2}$, and no technique works well when θ is very close to zero or one, unless N is extremely large. The table shows results for what is called the 2.5% tail, but is really the tail area corresponding to the .025 critical value for the standard normal distribution. Thus the quantities actually being estimated are .0536 when $d = 5$, .0392 when $d = 10$, and .0297 when $d = 30$. In addition to the naive estimator, there are three different control variate estimators: BCV1 uses a binary control variate with $\gamma = .025$, BCV2 uses a binary control variate with $\gamma = \bar{\theta}$, and OCV uses the approximately optimal control variate (20). OCV and BCV1 are about equally reliable, and BCV2 slightly less so. They all have a tendency to cover the true value less than 95% of the time when $N = 500$, especially for $d = 30$, presumably because the tail area being estimated is smallest in that case. The gains from using control variates are not as great as Table 1 would suggest, but are by no means negligible. As expected, the OCV always works best, followed in all but one case by BCV2, then BCV1, and finally the naive estimator.

Finally, we come to quantile estimation. The principal determinant of the performance of quantile estimators seems to be $\min\big(\alpha, (1-\alpha)\big)N$, the number of replications for which y lies below (or above, if $\alpha > 0.5$) the quantile being estimated. None of the estimators is very reliable when estimating quantiles far from 0.5, perhaps because the kernel estimates of $\bar{f}$ are not very accurate. In every case shown in the table, the 95% confidence interval covers the true value less than 95% of the time, sometimes quite a lot less. However, the CV estimators are generally only a little bit worse than the naive estimator in this respect.

For the .05 quantile, the OCV estimator always outperforms the BCV and naive estimators, but not by as much as Table 1 suggests that it should. It may be worth using the OCV estimator in this case, but one must be a little cautious in drawing inferences even when N is several thousand. For the .01 quantile, the OCV estimator actually performs worse than the naive estimator in three cases, and worse than the BCV estimator in four. Remember that when $\alpha = .01$, $\min\big(\alpha, (1-\alpha)\big)N$ is only 5 when $N = 500$ and only 10 when $N = 1000$, so it is not surprising that asymptotic theory does not work very well. Part of the problem is that for small values of αN the OCV estimator is sometimes quite severely biased away from zero. Thus when $\min\big(\alpha, (1-\alpha)\big)N$ is less than about 20, these results suggest that it is probably better to use the BCV estimator instead of the OCV one. Since the former has the additional advantage of being easier to compute, some may prefer to use it all the time.

6. An Example

In this section, we illustrate some of the techniques discussed in this paper by using them in a small Monte Carlo experiment. The experiment concerns pseudo-t statistics for OLS estimators based on a heteroskedasticity-consistent covariance matrix estimator, or HCCME for short; see White (1980). The model considered is

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{u}, \quad E(\boldsymbol{uu}^\top) = \boldsymbol{\Omega} \tag{23}$$

where $\boldsymbol{X}$ is $n \times k$ and $\boldsymbol{\Omega}$ is an $n \times n$ diagonal matrix that is known to the experimenter but is treated as unknown for the purpose of estimation and inference. The true covariance matrix of the OLS estimates is

$$\boldsymbol{V}(\hat{\boldsymbol{\beta}}) = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{\Omega}\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}.$$

There are various HCCMEs for this model, which all take the form

$$\hat{\boldsymbol{V}}(\hat{\boldsymbol{\beta}}) = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\hat{\boldsymbol{\Omega}}\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}, \tag{24}$$

but differ in how $\hat{\boldsymbol{\Omega}}$ is calculated. For the present experiment, we define $\hat{\boldsymbol{\Omega}}$ as a diagonal matrix with typical diagonal element equal to $\big(n/(n-k)\big)\hat{u}_t^2$, where $\hat{u}_t$ is the t^{th} residual from OLS estimation of (23). Other choices for $\hat{\boldsymbol{\Omega}}$ may work better in finite samples; see MacKinnon and White (1985).

The statistics we examine are pseudo-t statistics of the form

$$\frac{\hat{\beta}_i - \beta_{i0}}{(\hat{V}_{ii})^{1/2}}, \tag{25}$$

where $\hat{\beta}_i$ is the OLS estimate of β_i for some i, β_{i0} is the value of β_i used to generate the data, and $(\hat{V}_{ii})^{1/2}$ is the square root of the i^{th} diagonal element of (24). We assume that

the u_t are normally and independently distributed. Thus if $(\hat{V}_{ii})^{1/2}$ in (25) were replaced by $(V_{ii})^{1/2}$, the true standard error of $\hat{\beta}_i$, the statistic (25) would be distributed as $N(0,1)$. This infeasible test statistic, which corresponds to what we called x in Section 4, was used to generate control variates for the actual test statistic (25), which corresponds to what we called y, and of which the finite-sample distribution is in general unknown.

We performed experiments for a particular case of (23) with two normally and independently distributed regressors and a constant term, with heteroskedasticity generated by a random coefficient model as in MacKinnon and White (1985). In Table 3 we report results only for one particular β_i. These results are of course specific to that parameter of the particular model and data generating process that we used. What we are primarily interested in is the relative performance of the CV and naive estimators as a function of the sample size n. We used 20,000 replications for $n = 25$ and $n = 50$, 10,000 for $n = 100$ and $n = 200$, 5000 for $n = 400$ and $n = 800$ and 2500 for $n = 1600$ and $n = 3200$. It makes sense to consider fairly large values of n, since HCCMEs are most often used in the context of cross-section data.

The first line of Table 3 shows that the correlation between the test statistic y and the control variate x increases very substantially as the sample size n increases. This is of course to be expected, since y is equal to x asymptotically. It means that the efficiency of the CV estimator increases relative to that of the naive estimator as n increases. That is why we can get away with reducing the number of replications by a factor of two every time n increases by a factor of four. In fact, if we were interested only in means and standard deviations, we could make N proportional to $1/n$ and still obtain results for large sample sizes that would be just as accurate as for small ones. Since the cost of a Monte Carlo experiment is in most cases roughly proportional to N times n, that would be very nice indeed. However, for estimating test sizes and quantiles it appears that we cannot reduce N that rapidly; instead, making N approximately proportional to $n^{-1/2}$, as we have done, seems to work quite well. For $n = 3200$, the case for which experimentation is most expensive and the control variates most useful, it would require between 7.5 and 618 times more replications to achieve the same accuracy using naive estimation as using control variates.

7. Conclusion

This paper has discussed a very simple, and yet very general, method for using the information in control variates to improve the efficiency with which quantities of interest are estimated in Monte Carlo experiments. The information in the control variates can be extracted simply by running a linear regression and recording the estimate of the constant term and its standard error. This technique can be used whenever one or more control variates with a known mean of zero can be computed along with the quantities of interest. It can also be used when two or more estimates of the latter are available on each replication, as in the case of antithetic variates.

This regression technique for using control and antithetic variates is not new, although it does not seem to have been used previously in econometrics. There are several new results in the paper, however. First of all, we have proposed a new way to estimate quantiles by modifying this regression procedure. Secondly, we have proposed ways to obtain approximately optimal control variates in many cases of interest, including the estimation of tail areas and quantiles. Finally, we have obtained a number of simulation results which suggest that these methods will generally work quite well in practice, provided the number of replications is not too small, and that they can dramatically reduce the number of replications required to obtain a given level of accuracy.

References

Davidson, R. and J. G. MacKinnon, 1981, Efficient estimation of tail-area probabilities in sampling experiments, Economics Letters 8, 73–77.

Fieller, E. C. and H. O. Hartley, 1954, Sampling with control variables, Biometrika 41, 494–501.

Hammersley, J. M. and D. C. Handscomb, 1964, Monte Carlo Methods, Methuen, London.

Hendry, D. F., 1984, Monte Carlo experimentation in econometrics, Ch. 16 in Handbook of Econometrics, Vol. II, eds. Z. Griliches and M. D. Intriligator, North-Holland, Amsterdam.

Hendry, D. F. and P. K. Trivedi, 1972, Maximum likelihood estimation of difference equations with moving average errors: a simulation study, Review of Economic Studies, 39, 117–145.

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

Lewis, P. A. W. and E. J. Orav, 1989, Simulation Methodology for Statisticians, Operations Analysts and Engineers, Wadsworth and Brooks/Cole, Pacific Grove, California.

MacKinnon J. G. and H. White, 1985, Some heteroskedasticity consistent covariance matrix estimators with improved finite sample properties, Journal of Econometrics, 29, 305–325.

Nankervis, J. C. and N. E. Savin, 1988, The Student's t approximation in a stationary first order autoregressive model, Econometrica, 56, 119–145.

Ripley, B. D., 1987, Stochastic Simulation, John Wiley & Sons, New York.

Rosenblatt, M., 1956. Remarks on some nonparametric estimates of density functions, Annals of Mathematical Statistics, 27, 832–837.

Rothery, P., 1982, The use of control variates in Monte Carlo estimation of power, Applied Statistics, 31, 125–129.

Silverman, B. W., 1986, Density Estimation for Statistics and Data Analysis, Chapman & Hall, New York.

Tauchen, G. E., 1985, Diagnostic testing and evaluation of maximum likelihood models, Journal of Econometrics 30, 415–443.

White, H., 1980, A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity, Econometrica, 48, 817–838.

Table 1

Potential Efficiency Gains from Control Variates

Quantity Estimated	ρ^2 : .50	.60	.70	.80	.90	.95	.99
μ	2.00	2.50	3.33	5.00	10.00	20.00	100.00
σ^2	1.33	1.56	1.96	2.78	5.26	10.26	50.25
$\alpha = .50$	1.50	1.69	1.97	2.44	3.48	4.95	11.10
	1.33	1.47	1.66	1.99	2.72	3.75	8.12
$\alpha = .25$	1.45	1.62	1.88	2.32	3.29	4.67	10.45
	1.29	1.41	1.59	1.90	2.58	3.55	7.62
$\alpha = .10$	1.33	1.48	1.69	2.06	2.89	4.08	9.09
	1.21	1.31	1.46	1.71	2.30	3.13	6.66
$\alpha = .05$	1.26	1.38	1.56	1.88	2.62	3.68	8.17
	1.16	1.24	1.36	1.59	2.11	2.85	6.05
$\alpha = .025$	1.20	1.30	1.46	1.74	2.40	3.35	7.41
	1.11	1.18	1.29	1.49	1.95	2.62	5.50
$\alpha = .01$	1.14	1.22	1.35	1.59	2.16	3.00	6.61
	1.08	1.13	1.22	1.38	1.78	2.37	4.91

Notes:

Each entry is the ratio of the (asymptotic) variance for the naive estimator to that for the control variate estimator. When there are two entries, the first is for the optimally-chosen control variate and the second is for the binary control variate.

Both x and y are assumed to be $N(0, 1)$.

Results for estimation of tail areas apply to quantile estimation as well.

Table 2

Performance of CV Estimators with Finite N

D.F.		5			10			30	
N	500	1000	2000	500	1000	2000	500	1000	2000
ρ^2	.853	.851	.850	.940	.940	.940	.982	.982	.982
Mean:									
95% naive	94.7	95.0	95.2	95.2	95.4	95.0	94.8	94.9	94.8
95% CV	95.3	94.6	95.2	95.3	95.8	95.3	94.6	95.0	95.5
Ratio	6.72	6.60	6.54	16.70	16.65	17.54	55.08	57.15	58.22
Variance:									
95% naive	90.9	92.4	92.9	94.1	94.9	94.9	94.6	94.5	94.6
95% CV	88.7	90.2	91.6	92.5	93.4	93.7	93.2	94.0	94.2
Ratio	1.36	1.28	1.40	2.92	2.91	2.88	9.22	9.59	9.46
2.5% Tail:									
95% naive	94.7	94.0	94.3	94.6	96.2	94.8	92.2	95.8	95.2
95% OCV	93.9	94.5	94.6	93.3	95.4	95.1	92.9	94.4	94.4
95% BCV1	94.0	94.5	94.9	93.3	95.1	95.0	93.0	94.5	94.4
95% BCV2	93.7	94.2	94.1	92.7	94.7	95.0	91.3	93.3	93.8
Ratio OCV	1.67	1.69	1.71	2.00	1.99	1.99	2.91	2.93	2.95
Ratio BCV1	1.30	1.31	1.33	1.56	1.54	1.55	2.31	2.29	2.34
Ratio BCV2	1.45	1.45	1.46	1.69	1.70	1.73	2.30	2.35	2.37
5% Quantile:									
95% naive	91.3	92.2	93.1	91.6	93.5	93.6	92.4	93.1	93.1
95% OCV	90.4	91.8	91.9	90.7	92.9	93.1	91.8	92.9	93.7
95% BCV	90.4	91.9	92.0	90.8	93.0	92.9	91.3	93.2	93.9
Ratio OCV	1.60	1.63	1.65	2.15	2.15	2.24	3.59	3.66	3.81
Ratio BCV	1.41	1.44	1.43	1.83	1.83	1.85	2.84	3.00	3.04
1% Quantile:									
95% naive	86.1	88.4	91.4	86.6	89.1	91.7	87.0	89.9	91.0
95% OCV	82.3	86.9	90.7	83.3	88.0	91.0	84.4	88.9	90.0
95% BCV	84.6	87.8	90.4	84.4	88.1	91.1	84.0	89.3	89.7
Ratio OCV	0.29	0.98	1.08	0.63	1.25	1.30	1.76	1.95	1.99
Ratio BCV	1.00	1.07	1.09	1.19	1.25	1.26	1.71	1.80	1.79

Notes:

OCV means the "optimally chosen" control variate and BCV means the binary control variate. For tail-area estimation, BCV1 means the binary control variate with γ equal to the nominal size of the tail, and BCV2 means the binary control variate with $\gamma = \bar{\theta}$.

Entries opposite "95% naive", "95% CV", and so on represent the percentage of the time that calculated 95% confidence intervals covered the true value.

Entries opposite "Ratio" are the ratios of the mean square error of the naive estimator to that of the specified CV estimator.

Table 3

Performance of Pseudo-*t* Statistics Based on HCCME

n N	25 20000	50 20000	100 10000	200 10000	400 5000	800 5000	1600 2500	3200 2500
ρ^2	0.886	0.938	0.963	0.980	0.988	0.994	0.997	0.998
Mean:								
Naive	−0.004 (.0093)	−0.002 (.0082)	−0.002 (.0109)	0.002 (.0105)	−0.011 (.0145)	−0.045 (.0142)	−0.003 (.0202)	0.008 (.0199)
CV	0.002 (.0031)	0.004 (.0021)	−0.001 (.0021)	0.002 (.0015)	0.005 (.0016)	0.001 (.0011)	0.002 (.0012)	0.000 (.0008)
Std. Dev.:								
Naive	1.312 (.0081)	1.163 (.0063)	1.092 (.0082)	1.049 (.0078)	1.022 (.0102)	1.001 (.0099)	1.011 (.0144)	0.994 (.0135)
CV	1.324 (.0058)	1.161 (.0035)	1.086 (.0037)	1.044 (.0025)	1.029 (.0026)	1.012 (.0019)	1.007 (.0020)	1.003 (.0013)
Test size:								
Naive	9.845 (0.211)	7.480 (0.186)	6.540 (0.247)	5.980 (0.237)	5.160 (0.313)	5.560 (0.324)	4.880 (0.431)	5.440 (0.454)
BCV	9.848 (0.158)	7.560 (0.129)	6.494 (0.159)	6.029 (0.139)	5.127 (0.173)	5.283 (0.148)	5.229 (0.154)	5.252 (0.130)
OCV	9.919 (0.143)	7.535 (0.117)	6.325 (0.145)	6.005 (0.126)	5.125 (0.155)	5.197 (0.130)	5.099 (0.130)	5.186 (0.111)
Crit. value:								
Naive	2.601 (.0213)	2.318 (.0172)	2.149 (.0205)	2.042 (.0192)	1.982 (.0255)	1.956 (.0240)	2.008 (.0373)	1.936 (.0325)
BCV	2.610 (.0193)	2.300 (.0144)	2.134 (.0155)	2.058 (.0128)	2.000 (.0158)	1.994 (.0141)	1.979 (.0172)	1.970 (.0132)
OCV	2.615 (.0186)	2.305 (.0137)	2.136 (.0144)	2.051 (.0118)	2.005 (.0144)	1.986 (.0124)	1.975 (.0152)	1.973 (.0119)

Notes:

Estimated standard errors are in parentheses.

"Test size" is the estimated size of a nominal 5% one-tail test, in per cent.

"Crit. value" is the estimated critical value for a 5% two-tail test.

"BCV" denotes estimates using the binary control variate; when estimating test size, the naive estimate of test size was used to construct the binary control variate. "OCV" denotes estimates using the approximately optimal control variate.