Queen's Economics Department Working Paper No. 781

Regression-Based Methods for Using Control and Antithetic Variates in Monte Carlo Experiments

Russell Davidson James G. Mackinnon

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

5-1990

DISCUSSION PAPER NO. 781

REGRESSION-BASED METHODS FOR USING CONTROL AND ANTITHETIC VARIATES IN MONTE CARLO EXPERIMENTS*

by

Russell Davidson and James G. MacKinnon

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

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

and

GREQE
Ecole des Hautes Etudes en Sciences
Universités d'Aix-Marseille II et III
Centre de la Vielle Charité
2 rue de la Charité
13002 Marseille, France

May 1990

* 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, two referees and seminar participants at UCSD, Berkeley, UCLA and USC for comments on the earlier paper. This research was supported, in part, by grants from the Socialciences and Humanities Research Council of Canada.

DISCUSSION PAPER NO. 781

REGRESSION-BASED METHODS FOR USING CONTROL AND ANTITHETIC VARIATES IN MONTE CARLO EXPERIMENTS*

by

Russell Davidson and James G. MacKinnon

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

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

and

GREQE
Ecole des Hautes Etudes en Sciences
Universités d'Aix-Marseille II et III
Centre de la Vielle Charité
2 rue de la Charité
13002 Marseille, France

May 1990

* 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, two referees and seminar participants at UCSD, Berkeley, UCLA and USC for comments on the earlier paper. This research was supported, in part, by grants from the Socialciences and Humanities Research Council of Canada.

Abstract

Methods based on linear regression provide a very easy way to use the information in control and antithetic 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. Ways to obtain approximately optimal control variates in many cases of interest are also proposed. These methods seem to work well in practice, and can greatly reduce the number of replications required to obtain a given level of accuracy.

1. Introduction

Monte Carlo methods are widely used to study the finite-sample properties of estimators and test statistics when analytic results are not available. Hendry (1984) provides a survey of Monte Carlo methods in econometrics, and Ripley (1987) and Lewis and Orav (1989) provide more modern treatments from the perspectives of statistics and operations research. One problem with Monte Carlo experiments is that the results 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, and despite the rapidly declining cost of computation in recent years, this means that a set of Monte Carlo experiments can be very costly.

In many cases, computational costs can be greatly reduced by the use of variance reduction techniques to improve the precision with which the quantities of interest in the experiment are estimated. One of the best-known such techniques is to make use of what are called control variates. A control variate is a random variable that is correlated with the estimator or test statistic the properties of which are being investigated, 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 and/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. A simple modification allows this regres-

sion to be used with antithetic variates instead of, or along with, control variates. Thus the regression procedure encompasses and improves upon both of the two most widely used methods of variance reduction. This procedure has been discussed in the operations research literature,[1] but is not covered in Hendry's survey and appears to be unfamiliar to most econometricians. In the next section we therefore discuss it in some detail. In Section 3 a modified version of the regression procedure is introduced that can be used to estimate quantiles. Then in Section 4 we discuss how to choose control variates in an (approximately) optimal fashion. In Section 5, the regression procedure is extended to situations where more than one estimate of θ is obtained on each replication, as in the case of antithetic variates. Finally, in Section 6, we present some Monte Carlo results which demonstrate how useful these procedures can be in practice.

Suppose that a Monte Carlo experiment involves N replications, on each of which we obtain 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's calculated during the experiment. Obvious examples of θ include the bias, variance, skewness or kurtosis of a particular parameter estimate or test statistic. If θ were the bias of some estimator, t_j would be the estimate obtained on the j^{th} replication, minus its true value; if it were the mean squared error of an estimator, it would be the squared difference between the estimate on the j^{th} replication and its true value; and so on. 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, under either the null hypothesis or some specified alternative. 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

---

[1] Ripley (1987) and Lavenberg and Welch (1981) are two good references. The operations research literature on the subject of control variates is very large, but most of it is concerned with simulation problems of a type that do not arise in econometrics.

mean of anything, they do not seem to fit into this general scheme. They require a slightly different treatment, as we explain in Section 3.

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

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

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 random quantity t_j.[2] The two primary requirements for τ_j are that it be correlated with t_j and that it have mean zero. If we can find such a control variate, we can always obtain a more efficient estimate than $\bar{\theta}$, at least when N is reasonably large. It is necessary only that the variance of τ_j, $V(\tau)$, and its covariance with t_j, $\mathrm{Cov}(t,\tau)$, be finite and non-zero.

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

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

where $\bar{\tau}$ is the sample mean of the τ_j's and λ is a scalar that has to be determined. On average, $\bar{\tau}$ will be zero, since τ_j has population mean zero, so that $\ddot{\theta}(\lambda)$ clearly has the same mean as $\bar{\theta}$. But in almost every actual sample, $\bar{\tau}$ will be non-zero. If, for example, $\bar{\tau}$ is positive, and if the control variates τ_j are positively correlated with the quantities t_j, it is likely that $\bar{\theta}$ will also exceed its true mean. In this case λ would be positive, so that $\ddot{\theta}(\lambda)$ would involve subtracting a multiple of $\bar{\tau}$ from $\bar{\theta}$.

The choice of λ is obviously crucial. It seems natural to choose it so as to minimize the variance of the CV estimator (1):

---

[2] There may of course be more than one possible choice for τ_j. We shall consider the possibility of using several control variates in the next section.

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

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

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

Substituting (3) into (2), the variance of $\ddot{\theta}(\lambda^*)$ is then seen to be

$$V(\ddot{\theta}(\lambda^*)) = N^{-1}\left[V(t) - \frac{\mathrm{Cov}(t,\tau)^2}{V(\tau)}\right] = (1-\rho^2)V(\bar{\theta}), \tag{4}$$

where ρ is the correlation between the control variate τ_j and the estimate t_j. We see from (4) that whenever this correlation is non-zero, $V(\ddot{\theta}(\lambda^*))$ will be less than $V(\bar{\theta})$, so that there will be some gain from using the control variate.

In the early literature on control variates (including Hammersley and Handscomb (1964) and even Hendry (1984)), λ was arbitrarily set to unity. This will be a good choice if $V(t_j)$ and $V(\tau_j)$ are similar in magnitude and ρ is close to one, but it is clearly not the best choice. Unless N is very small, we can generally do better by estimating λ^*. This may be done in several ways, some obvious and some not so obvious. One of the less obvious ones is to use a certain linear regression, which we will discuss at length in the next section.

2. Regression-based Methods for Using Control Variates

It is very easy to use a linear regression to calculate CV estimates. This regression may be written as

$$\mathbf{t} = \theta\boldsymbol{\iota} + \lambda\boldsymbol{\tau} + \text{residuals}, \tag{5}$$

where $\mathbf{t}$ and $\boldsymbol{\tau}$ are N-vectors with typical elements t_j and τ_j respectively, and $\boldsymbol{\iota}$ is an N-vector of ones. Thus the regressand is a vector of quantities with mean θ, and the regressors are a constant term and the vector of control variates. It is no accident

that the constant term in (5) is called θ and the coefficient on τ is called λ. Because the τ_j's have population mean zero, the constant term in (5) is simply the mean of the t_j's. Moreover, the OLS estimate of λ from (5) will provide an estimate of the optimal λ^* defined in (3). This estimate is

$$\hat{\lambda} = (\tau^T M_\iota \tau)^{-1} \tau^T M_\iota t ,$$

where M_ι is the matrix $I - \iota(\iota^T\iota)^{-1}\iota^T$ that takes deviations from the mean. It is easy to see that $\hat{\lambda}$ is just the sample covariance of t and τ, divided by the sample variance of τ, so that it is the empirical counterpart of λ^*.

The OLS estimate of θ from regression (5) is

$$\hat{\theta} = \bar{\theta} - \hat{\lambda}\bar{\tau} ;$$

this follows from standard results for linear regressions with a constant term. Thus it is clear from (1) that the OLS estimate $\hat{\theta}$ is equal to $\ddot{\theta}(\hat{\lambda})$. Since $\hat{\lambda}$ is consistent for λ^*, under rather weak assumptions, $\hat{\theta}$ will be asymptotically equivalent to $\ddot{\theta}(\lambda^*)$.

Running regression (5) not only yields the CV estimate of θ, but also an estimated variance for that estimate, which we need in order to gauge the accuracy of the results and decide whether N is sufficiently large. This estimated variance is

$$\hat{\omega}^2(\iota^T M_\tau \iota)^{-1}, \tag{6}$$

where $\hat{\omega}$ is the standard error of the regression. The second factor here must tend to N^{-1}, since τ asymptotically has no explanatory power for ι, so that we could simply use $N^{-1}\hat{\omega}^2$ rather than (6) to estimate the variance of $\hat{\theta}$. The residuals from (5) are the same as those from the regression of $t - \bar{t}\iota$ on $\tau - \bar{\tau}\iota$, which means that the estimate $\hat{\omega}^2$ tends for large N to

$$V(t) - \frac{(\mathrm{Cov}(t,\tau))^2}{V(\tau)} = V(t)(1 - \rho^2),$$

in accord with (4). 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. We shall explore the implications of this fact in Section 4.

This analysis makes it clear that, contrary to what much of the literature on control variates implies, the link between θ and the τ_j's need not be close. In particular, the older literature that sets $\lambda = 1$ implicitly assumes both that $V(t) = V(\tau)$ and that ρ is close to one, conditions that many potential control variates fail to satisfy. In fact, any random variable that can be calculated along with t_j, is correlated with it (either positively or negatively), and has mean zero, finite variance, and finite covariance with t_j, can be used as a control variate.

Since this is the case, 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

$$\mathbf{t} = \theta\boldsymbol{\iota} + \mathbf{T}\boldsymbol{\lambda} + \text{residuals}, \tag{7}$$

where $\mathbf{T}$ is an $N \times c$ matrix, each column of which consists of observations on one of c control variates. It is clear that the OLS estimate of θ from this regression will once again provide the estimate we are seeking, since all the columns of $\mathbf{T}$ have mean zero. The OLS estimate of θ from (7) is

$$\hat{\theta} = (\boldsymbol{\iota}^{\mathsf{T}}\mathbf{M}_{\mathbf{T}}\boldsymbol{\iota})^{-1}\boldsymbol{\iota}^{\mathsf{T}}\mathbf{M}_{\mathbf{T}}\mathbf{t},$$

where $\mathbf{M}_{\mathbf{T}} = \mathbf{I} - \mathbf{T}(\mathbf{T}^{\mathsf{T}}\mathbf{T})^{-1}\mathbf{T}^{\mathsf{T}}$. Since $N^{-1}\boldsymbol{\iota}^{\mathsf{T}}\mathbf{M}_{\mathbf{T}}\boldsymbol{\iota}$ tends to unity as N tends to infinity, it is easy to see that the variance of $\hat{\theta}$ is once again just $N^{-1}\omega^2$, where ω is the true standard error of regression (7). Thus our objective in choosing control variates is simply to make regression (7) fit as well as possible; see Section 4.

The residuals in regression (7) will often not be conditionally homoskedastic. Thus it might be thought better to use as an estimate of the variance of $\hat{\theta}$, not the OLS variance estimate, but rather a heteroskedasticity-consistent one of the sort proposed by Eicker (1963) and 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, although generally not in the sample), 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 (7) will not, in general, be consistent when the residuals of (7) are heteroskedastic, but these coefficients are not of interest.

One aspect of this 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 6, we describe some experimental results on how well this procedure works for finite N.

3. Quantile Estimation

The estimation of quantiles is very 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 the mean of anything, 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 (7) 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

$$F(q_\alpha) = \alpha . \tag{8}$$

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 (8) by the empirical distribution of the generated y's:

$$N^{-1} \sum_{j=1}^{N} I(\bar{q}_\alpha - y_j) = \alpha , \tag{9}$$

where the indicator function I is defined to be equal to one if its argument is posi-

tive, and zero otherwise. If $\bar{q}_\alpha$ is not uniquely defined by (9), we may take for $\bar{q}_\alpha$ the mean of the set of numbers that satisfy (9). If αN is not an integer, there will be no $\bar{q}_\alpha$ that satisfies (9), in which case we may take for $\bar{q}_\alpha$ the mean of the set of numbers that make the left-hand side of (9) as close as possible in absolute value to α. In the following discussion we shall for simplicity consider only values of α for which (9) can be exactly satisfied for the chosen N.

Consider for some possibly random $\ddot{q}_\alpha$ close to q_α the random variable

$$I(\ddot{q}_\alpha - y) - \alpha .$$

If we suppose, in the case in which $\ddot{q}_\alpha$ is random, that $\ddot{q}_\alpha$ and y are independent, the mean of this variable conditional on $\ddot{q}_\alpha$, using (8), is

$$E\big(I(\ddot{q}_\alpha - y)\big) - \alpha = F(\ddot{q}_\alpha) - F(q_\alpha). \tag{10}$$

Suppose also that $\ddot{q}_\alpha$ approaches q_α with a root-N rate of convergence:

$$\ddot{q}_\alpha - q_\alpha = O(N^{-\frac{1}{2}}).$$

If the density $f(q_\alpha) \equiv F'(q_\alpha)$ exists and is nonzero, the mean (10) becomes, by Taylor's Theorem,

$$f(\ddot{q}_\alpha)(\ddot{q}_\alpha - q_\alpha) + O(N^{-1}). \tag{11}$$

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)), and we have used it in the experiments described in Section 6. Since the density only has to be estimated at a single point, 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 (11) that

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

or, equivalently,

$$\ddot{q}_\alpha - E\Big(\big(I(\ddot{q}_\alpha - y)\big) - \alpha\big)/\ddot{f}\Big) = q_\alpha + o(N^{-\frac{1}{2}}). \tag{12}$$

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

$$\ddot{q}_\alpha - (I(\ddot{q}_\alpha - y_j)) - \alpha)/\ddot{f} . \tag{13}$$

The regression function 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 (12), be an estimate of q_α correct to the leading asymptotic order of $N^{-\frac{1}{2}}$. 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 (13) 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 (13) 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

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

For large N this tends to

$$\alpha(1 - \alpha)/(Nf^2(q_\alpha)), \tag{14}$$

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 (14).

An alternative approach has been used in the operations research literature (see 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 very little about how to choose the control variate(s) to be used as regressors in regression (7) or its analogue for quantile estimation. 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 (7),

$$t_j = \theta + g(T_j, \gamma) + \text{residuals}, \tag{15}$$

where the nonlinear regression function $g(T, \gamma)$ is restricted to have mean zero, and γ is a vector of parameters. Estimation of (15) by nonlinear least squares would yield $\hat{\gamma}$ 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 (15) 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 $E(t) = E(E(t|T))$. 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 $g(T, \gamma)$. In cases where not much is known about that relationship, it may make sense to use a number of control variates in an effort to make $g(T, \gamma)$ 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 often realistic, 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. Let x denote the control variate 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

$$y = \mu + \rho\sigma x + v, \quad v \sim N(0, (1 - \rho^2)\sigma^2), \tag{16}$$

where ρ is the correlation between y and x. Using (16), we can easily 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. It is clear that we want t_j to equal y_j in regression (7). From (16) 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. 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 estimate 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 estimate were used instead of the naive estimate.

Now suppose that we wish to estimate σ^2. The obvious choice for t is $(y - \hat{\mu})^2$, although any consistent estimate of μ could be used. From (16) we see that

$$(y - \mu)^2 = (\rho\sigma x + v)^2 = \rho^2\sigma^2x^2 + 2\rho\sigma xv + v^2. \tag{17}$$

This implies that

$$E((y - \mu)^2|x) = \rho^2\sigma^2x^2 + (1 - \rho^2)\sigma^2. \tag{18}$$

The optimal regressor, adjusted to have mean zero, is evidently $(x^2 - 1)$. From (17) and (18), the variance of the optimally-chosen CV estimate $\hat{\sigma}^2$ will be

$$N^{-1}E\left((\rho^2\sigma^2x^2 + 2\rho\sigma xv + v^2) - (\rho^2\sigma^2x^2 + (1 - \rho^2)\sigma^2)\right)^2 = 2N^{-1}(1 - \rho^4)\sigma^4. \tag{19}$$

In contrast, the variance of the naive estimate 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 $|\rho| < 1$. One can easily estimate σ by taking the square root of $\hat{\sigma}^2$. Using (19) and the standard formula for the variance of a function of a random variable, the variance of $\hat{\sigma}$ will be

$$\frac{1}{2}N^{-1}\sigma^2(1 - \rho^4),$$

which is again smaller than that of the naive estimate by a factor of $(1 - \rho^4)$.

Now suppose that we are interested in the size or power of a test statistic, so that 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's. Davidson and MacKinnon (1981) and Rothery (1982) independently studied this problem under the assumption that the control variate, like the t_j's, can take on only two possible values, and proposed a technique based on the method of maximum likelihood. 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

$$\tau_j = I(x - q_{1-\alpha}) - \alpha. \tag{20}$$

Thus τ_j is a binary variable that is equal to $1 - \alpha$ when x_j exceeds $q_{1-\alpha}$, and $-\alpha$

otherwise, for some α that should be as close to θ as possible. Since the probability that x_j will exceed $q_{1-\alpha}$ is α, (20) clearly has population mean zero.

The expectation of $I(y - y^c)$ conditional on x is not a step function like (20). 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

$$\begin{aligned} E(t|x) &= \text{Prob}\left(v > y^c - \mu - \rho\sigma x\right) \\ &= \Phi\left(\frac{\mu + \rho\sigma x - y^c}{\sigma(1 - \rho^2)^{1/2}}\right) \\ &= \Phi\left((\mu + \rho\sigma x - y^c)/\omega\right) = \Phi(a + bx), \end{aligned}$$

where $a = (\mu - y^c)/(\sigma(1 - \rho^2)^{\frac{1}{2}})$, $b = \rho/(1 - \rho^2)^{\frac{1}{2}}$, $\omega = \sigma(1 - \rho^2)^{\frac{1}{2}}$ 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's, 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:

$$\Phi(\hat{a} + \hat{b}x_j) - \Phi((\hat{\mu} - y^c)/\hat{\sigma}). \tag{21}$$

The second term in (21) is the unconditional mean of the first; if μ and σ were known, it would be $\theta \equiv E(t)$. Regressing the observations of t on a constant and (21) should give a reasonably good estimate of θ, even if the assumption that y is normally distributed is not quite correct. If we are interested in tail areas for two-tail tests, the optimal regressor will be the sum of (21) and its analogue for the other tail.

When all our assumptions are satisfied, we see that (for large N) the variance of $\hat{\theta}$ will be

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

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

$$N^{-1}E\Big(\Phi(a + bx)\big(1 - \Phi(a + bx)\big)\Big), \tag{22}$$

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 expression (13). Since the factor of $1/\bar{f}$ in (13) is just a constant, it is unnecessary to include it in the optimal regressor, which is therefore

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

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

$$\Phi\left(\frac{\bar{q}_\alpha - \hat{\mu} - \hat{\rho}\hat{\sigma}x}{\hat{\sigma}(1 - \hat{\rho}^2)^{1/2}}\right) - \Phi\left(\frac{\bar{q}_\alpha - \hat{\mu}}{\hat{\sigma}}\right). \tag{23}$$

This control variate is very similar to (21), the 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 (23) as a control variate when estimating q_α will produce the same proportional gain in efficiency as using (21) when estimating α.

It might seem that one could improve on the procedures we have suggested by using something other than OLS (or NLS) to estimate regressions (7) and (15). In many cases, notably for tail areas and quantiles, those regressions will have error terms that are conditionally heteroskedastic. It might therefore seem appropriate to use some sort of GLS procedure. Unfortunately, this turns out not to be the case. The problem is that GLS will in general be consistent only if the regression model is correctly specified, that is if $g(T, \gamma)$ in (15) is equal to τ^* for some γ. If one does not have the proper specification of $g(T, \gamma)$, the residuals will not be orthogonal to all possible functions of T. Since the GLS weights will necessarily be chosen as functions of T, they may well be correlated with the OLS residuals, thereby biasing the GLS estimate of the constant. In contrast, as we showed in Section 2, OLS yields a consistent estimate of θ under very weak conditions. Since it is unlikely in practice that one would know enough

about y to construct a regressor that is actually optimal, OLS appears to be the procedure of choice.

It is interesting to see how substantially the number of replications can be reduced by the use of control variates, while maintaining a given level of accuracy. Table 1 presents some illustrative results, for the simple 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 naive estimator exceeds the number needed by the control variate estimator. 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, since the 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.[3] Entries for the OCV estimator for tail areas (and quantiles) are the ratio of $\Phi(\alpha)(1 - \Phi(\alpha))$ to expression (22), 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 (see Section 6), there is always a gain of at least a fac-

---

3 The results for the binary control variate assume that the value of α used to construct it is the same as the probability θ that $y > y^c$. This assumption is trivial to satisfy when one is estimating quantiles, but somewhat harder to satisfy when estimating tail areas. If this assumption is not satisfied, the binary control variate will perform less well than these results suggest. See Davidson and MacKinnon (1981), where we suggested choosing α to be the naive estimate of θ.

tor of two when the optimally-chosen control variate is used. Thus it appears that it will often be worthwhile to use control variates.

5. Antithetic Variates

Up to this point we have assumed that each replication of each experiment yields only one estimate of θ. It may often be possible to obtain more than one, however. One technique that does so is the method of antithetic variates; see, among others, Hammersley and Handscomb (1964), Rubinstein (1981) and Ripley (1987). This technique generates 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, and then takes an average of their average. 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.

In general, we shall suppose that there are ℓ different estimates of θ on each replication. The estimates will be denoted t_{ji}, $j = 1, \ldots, N$, $i = 1, \ldots, \ell$. A simple procedure for estimating θ in this case is to run the artificial regression

$$t_1 = \theta\iota + \sum_{i=2}^{\ell} \gamma_i(t_i - t_1) + \text{residuals}, \tag{24}$$

where t_1 and t_i are vectors with typical elements t_{j1} and t_{ji} respectively. It is obvious from previous results that regression (24) is valid. Since t_{ji} and t_{j1} are both unbiased estimates of the same thing, their difference must have mean zero, and hence it is valid to use that difference as a control variate. If additional control variates are available, then of course they too can be included as additional regressors, so that (24) would become

$$t_1 = \theta\iota + \sum_{i=2}^{\ell} \gamma_i(t_i - t_1) + T\lambda + \text{residuals}, \tag{25}$$

where as usual $\mathbf{T}$ is a matrix of control variates. For quantile estimation the regressand would be modified as described in Section 3.[4]

It does not matter which estimate of θ is given the index 1 and put on the left-hand side of regressions (24) and (25). The estimate of θ from (25) is

$$\hat{\theta} = (\iota^T \mathbf{M} \iota)^{-1} \iota^T \mathbf{M} t_1 , \tag{26}$$

where $\mathbf{M}$ denotes the matrix that projects onto the orthogonal complement of the subspace spanned by $\mathbf{T}$ and $(t_i - t_1)$ for $i = 2, \ldots, \ell$. Provided only that the t_i are linearly independent, the subspace spanned by the vectors $t_i - t_1$, $i = 2, \ldots, \ell$, is of dimension $\ell - 1$ and is identical to the subspace spanned by, for example, the vectors $t_i - t_2$, $i = 1, 3, 4, \ldots, \ell$. In fact, for any i and j different from 1, $t_i - t_j = (t_i - t_1) - (t_j - t_1)$. Thus the projection $\mathbf{M}$ will be the same regardless of which t_i is the regressand, and so will be $(\iota^T \mathbf{M} \iota)^{-1}$, the first factor in (26). Hence the estimate of θ from (25) modified to have t_i as the regressand must for all i be

$$(\iota^T \mathbf{M} \iota)^{-1} \iota^T \mathbf{M} t_i . \tag{27}$$

But observe that $\mathbf{M}(t_i - t_j) = 0$ for all i, j by the way $\mathbf{M}$ is constructed, which implies that $\mathbf{M}t_i = \mathbf{M}t_j$. This then implies that the second factor in (27) is the same for all t_i, and so (27) will equal $\hat{\theta}$ regardless of which t_i is the regressand in (25).

Regression (24) turns the venerable technique of antithetic variates into a simple special case of the general procedure for using control variates, and also improves its performance. Suppose we generate t_{j1} and t_{j2} on each replication, arbitrarily treat t_{j1} as the regressand for (25), and treat $(t_{j2} - t_{j1})$ as a control variate. The classical approach of averaging the simple averages $\frac{1}{2}(t_{j1} + t_{j2})$ is equivalent to restricting γ to equal -.5 in the regression

$$t_1 = \theta\iota + \gamma(t_2 - t_1) + \text{residuals}. \tag{28}$$

---

[4] Lavenberg and Welch (1981) note that the difference between two antithetic variates can be treated as a control variate and used in a regression. The idea does not seem to have been used in econometrics, however.

This value of γ will be optimal if t_1 and t_2 have a correlation of -1, in which case a perfectly accurate answer could be obtained, but will not be optimal in general. In contrast, regression (28) automatically chooses a value of γ that will be (asymptotically) optimal. Of course, whether this technique will be useful at all depends on whether we can obtain pairs of estimates that are highly negatively correlated.

6. Simulation Evidence

Since the theoretical arguments of this paper have all been asymptotic, one may well wonder whether the CV estimators of various θ's that we have discussed are in fact reliable for reasonable values of N. By "reliable" we mean that δ% confidence intervals based on normal theory and the estimated standard errors for $\hat{\theta}$ should cover the true values approximately δ% 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 a number of 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. 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 did 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.

It is clear from the table that regression-based control variate methods 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.[5] 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 .5, 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 $d = 30$. All the estimators are about equally reliable. They all perform well, but do have a tendency to cover the true value less than 95% of the time when $N = 500$, especially when $d = 30$ (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. The BCV estimator works a good deal less well than the OCV one (even though the latter is not really optimal here), especially when $d = 5$. The BCV estimator would probably have worked better if the value of α used to construct the control variate had been estimated, rather than fixed at .025.

---

[5] For $\rho^2 = .850$, .940 and .982 respectively, $1/(1 - \rho^2)$ is 6.67, 16.67 and 55.56.

Finally, we come to quantile estimation. As with estimators of tails, the principal determinant of the performance of quantile estimators is $\min((\alpha, (1-\alpha))N$, the number of replications for which y lies below (or above, if $\alpha > .5$) the quantile being estimated. None of the estimators is very reliable when estimating quantiles far from .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 is clearly 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. It appears from the table that one simply should not use this estimator when $\min((\alpha, (1-\alpha))N$ is small. Part of the problem with the OCV estimator is that it is sometimes quite severely biased away from zero. The bias is most severe when $d = 5$, and it goes away rapidly as both N and α are increased. When $\min((\alpha, (1-\alpha))N$ is less than about 20, these results suggest that it is probably better to use the BCV estimator rather than the OCV one. Since the former has the additional advantage of being somewhat easier to compute, one may prefer to use it even when $\min((\alpha, (1-\alpha))N$ is greater than 20.

In the remainder of 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 Eicker (1963) and White (1980)). The model of interest is

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \quad E(\mathbf{u}\mathbf{u}^{\mathsf{T}}) = \boldsymbol{\Omega}, \tag{29}$$

where $\mathbf{X}$ is $n \times k$ and $\mathbf{\Omega}$ 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

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

There are various HCCME's for this model, which all take the form

$$\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}^\mathsf{T}\mathbf{X})^{-1}\mathbf{X}^\mathsf{T}\hat{\mathbf{\Omega}}\mathbf{X}(\mathbf{X}^\mathsf{T}\mathbf{X})^{-1}, \tag{30}$$

but differ in how $\hat{\mathbf{\Omega}}$ is calculated. For the present experiment, we define $\hat{\mathbf{\Omega}}$ as a diagonal matrix with typical diagonal element equal to $(n/(n-k))\hat{u}_t^2$, where $\hat{u}_t$ is the t^{th} residual from OLS estimation of (29). Other choices for $\hat{\mathbf{\Omega}}$ may work better in finite samples; see MacKinnon and White (1985).

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

$$\frac{\hat{\beta}_i - \beta_{i0}}{(\hat{V}_{ii})^{1/2}}, \tag{31}$$

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 (30). We shall assume that the u_t's are normally and independently distributed. Thus if $(\hat{V}_{ii})^{1/2}$ in (31) were replaced by $(V_{ii})^{1/2}$, the true standard error of $\hat{\beta}_i$, the statistic (31) would be distributed as $N(0, 1)$. This infeasible test statistic, which corresponds to what we called x in Section 4, will be used to generate control variates for the actual test statistic (31), 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 (29) with two regressors and a constant term, where the u_t's followed a particular pattern of heteroskedasticity related to the X_t's. 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. The results we shall discuss are in Table 3. 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 $n = 3200$. As we explain below, this type of experimental design makes sense when control variates are being used.

The first line of Table 2 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 only interested in means and standard deviations, we could make N proportional to $1/n$ and still obtain results for large sample sizes that were 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 $n^{-\frac{1}{2}}$, 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 esti-

mates of interest. It can also be used when more than one estimate of the quantity of interest is available on each replication, as in the case of antithetic variates.

The regression technique for using control and antithetic variates is not new, although it does not seem to have been discussed 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.

Eicker, F., 1963, Asymptotic normality and consistency of the least squares estimators for families of linear regressions, Annals of Mathematical Statistics, 34, 447-456.

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.

Jennrich, R. I., 1969, Asymptotic properties of non-linear least squares estimators, Annals of Mathematical Statistics, 40, 633-643.

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.

Rubinstein, R. Y., 1981, Simulation and the Monte Carlo Method, John Wiley & Sons, New York.

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

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
α=.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
α=.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
α=.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
α=.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
α=.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
α=.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 a control variate estimator. When there are two entries, the first is for the optimally-chosen control variate and the second 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% BCV	94.0	94.5	94.9	93.3	95.1	95.0	93.0	94.5	94.4
Ratio-OCV	1.67	1.69	1.71	2.00	1.99	1.99	2.91	2.93	2.95
Ratio-BCV	1.30	1.31	1.33	1.56	1.54	1.55	2.31	2.29	2.34
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
Bias-naive	-0005	-0007	-0018	+0006	+0007	+0005	+0006	+0005	+0015
Bias-OCV	-0053	-0033	-0024	-0028	-0020	-0005	-0008	-0008	-0009
Bias-BCV	-0030	-0022	-0021	-0011	-0014	-0000	+0001	-0006	-0002
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
Bias-naive	-0100	-0012	-0023	-0008	+0019	-0019	+0021	+0042	+0010
Bias-OCV	-0867	-0203	-0109	-0300	-0099	-0072	-0078	-0026	-0024
Bias-BCV	-0168	-0049	-0042	-0060	-0026	-0036	-0007	-0000	-0004

Continued ....

Notes to Table 2

OCV means the optimally-chosen control variate, and BCV means the binary control variate, for tail-area and quantile estimation.

Entries opposite "95% naive", "95% CV", "95% OCV" and "95% BCV" represent the percentage of the time that calculated 95% confidence intervals included the true value of the quantity being estimated.

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

To save space, the decimal point has been omitted from the entries for bias. Thus, for example, -0016 should be read as -0.0016 .

Table 3

Performance of Pseudo t-statistics Based on HCCME

Sample:	25	50	100	200	400	800	1600	3200
N:	20000	20000	10000	10000	5000	5000	2500	2500
ρ^2	.886	.938	.963	.980	.988	.994	.997	.998
mean:								
naive	-0.443	-0.176	-0.209	0.235	-1.131	-4.480	-0.299	0.798
	(0.928)	(0.822)	(1.092)	(1.049)	(1.446)	(1.416)	(2.023)	(1.989)
CV	0.167	0.361	-0.127	-0.229	0.525	0.084	0.194	0.018
	(0.312)	(0.205)	(0.209)	(0.147)	(0.158)	(0.114)	(0.118)	(0.080)
median:								
naive	0.072	-0.418	0.309	1.970	-1.779	-3.842	0.332	1.634
	(1.091)	(1.008)	(1.330)	(1.281)	(1.794)	(1.837)	(2.592)	(2.484)
CV	0.125	0.067	0.259	0.394	-0.286	0.173	-0.347	-0.101
	(0.378)	(0.302)	(0.357)	(0.297)	(0.365)	(0.322)	(0.350)	(0.321)
S.D.:								
naive	1.312	1.163	1.092	1.049	1.022	1.001	1.011	0.994
	(0.008)	(0.006)	(0.008)	(0.008)	(0.010)	(0.010)	(0.014)	(0.014)
CV	1.324	1.162	1.086	1.044	1.029	1.012	1.007	1.003
	(0.006)	(0.004)	(0.004)	(0.002)	(0.003)	(0.002)	(0.002)	(0.001)
test size:								
naive	9.845	7.480	6.540	5.980	5.160	5.560	4.880	5.440
	(0.211)	(0.186)	(0.247)	(0.237)	(0.313)	(0.324)	(0.431)	(0.454)
CV	9.919	7.535	6.325	6.005	5.125	5.197	5.099	5.186
	(0.143)	(0.117)	(0.145)	(0.126)	(0.155)	(0.130)	(0.130)	(0.111)
c. value:								
naive	2.601	2.318	2.149	2.042	1.982	1.956	2.008	1.936
	(0.021)	(0.017)	(0.021)	(0.019)	(0.025)	(0.024)	(0.037)	(0.032)
CV	2.615	2.305	2.136	2.051	2.005	1.986	1.975	1.973
	(0.019)	(0.014)	(0.014)	(0.012)	(0.014)	(0.012)	(0.015)	(0.012)

Notes: Estimated standard errors are in parentheses.

Entries for mean and median are 100 times the actual values.

"test size" is estimated size of 5% one-tail test (in per cent).

"c. value" is estimated critical value for 5% two-tail test.