CONTROL VARIATES FOR VARIANCE REDUCTION IN INDIRECT INFERENCE: INTEREST RATE MODELS IN CONTINUOUS TIME*

Giorgio Calzolari, Francesca Di Iorio and Gabriele Fiorentini**

WP-AD 98-09

*Some results of this paper were presented at the poster session of the seventh edition of the European Conferences of the Econometrics Community EC, Firenze (1996), and at the European Meeting of the Econometric Society, Toulouse (1997). A part of this research was carried out while the first author was visiting the China Europea International Business School, Shanghai. We have benefitted from discussions with M. Arellano, C. Bianchi, E.M. Cleur, G.M. Gallo, C. Gouriéroux, G.E. Mizon, J. Mora, E. Renault and E. Sentana. Two anonymous referees have also helped us greatly improve the paper. Of course the usual caveat applies. Financial supports from MURST, CNR, CICYT (project PB96-03399) and IVIE are gratefully acknowledeged.

** G. Calzolari: Université di Firenze; F. Di Iorio: ISTAT; G. Fiorentini: University of Alicante.

Editor: **Instituto Valenciano de Investigaciones Económicas, S.A.**
Primera Edición Abril 1998.
ISBN: 84-482-1771-3
Depósito Legal: V-1166-1998
Impreso por Key, S.A.
Av. Cardenal Benlloch, 69, 46021-Valencia.
Impreso en España.

CONTROL VARIATES FOR VARIANCE REDUCTION IN INDIRECT INFERENCE: INTEREST RATE MODELS IN CONTINUOUS TIME

Giorgio Calzolari, Francesca Di Iorio & Gabriele Fiorentini

ABSTRACT

Simulation estimators, such as indirect inference or simulated maximum likelihood, are successfully employed for estimating stochastic differential equations. They adjust for the bias (inconsistency) caused by discretization of the underlying stochastic process, which is in continuous time. The price to be paid is an increased variance of the estimated parameters. The variance suffers from an additional component, which depends on the stochastic simulation involved in the estimation procedure. To reduce this undesirable effect, one could properly increase the number of simulations (or the length of each simulation) and thus the computational cost. Alternatively, this paper shows how variance reduction can be achieved, at virtually no additional computational cost, by use of control variates. The Ornstein-Uhlenbeck process, used by Vasicek to model the short term interest rate in continuous time, and the square-root process, used by Cox, Ingersoll and Ross, are explicitly considered and experimented with. Monte Carlo experiments show a global efficiency gain of almost 50% over the simple indirect estimator.

KEYWORDS: Efficient Monte Carlo; Variance Reduction Techniques; Control Variates; Indirect Inference; Euler Discretization; Short-Term Interest Rate; Stochastic Differential Equation.

1 Introduction

Simulation estimators, such as indirect inference (Gouriéroux, Monfort and Renault, 1993), simulated maximum likelihood (Smith, 1993) or efficient method of moments (Gallant and Tauchen, 1996) are successfully employed for estimating stochastic differential equations. They adjust for the bias (inconsistency) caused by the discretization of the underlying stochastic process, which is in continuous time.

The price to be paid is an increased variance of the estimated parameters. On the one hand, the variance is due to the intrinsic stochastic nature of the data and to the models adopted; on the other hand, it depends on the stochastic simulation involved in the estimation procedure. This latter component is, in some sense, an undesirable additional experimental variance, which can be made arbitrarily small by properly enlarging the number of simulations, at the cost of a bigger computational effort. Therefore a trade-off arises between variance reduction and computational cost.

Efficient Monte Carlo techniques may be helpful in reducing experimental variance, thus providing a reduction of the global variance of the estimator and, therefore, an overall improvement of the efficiency, without increasing the computational cost. There is a wide literature on efficient Monte Carlo techniques, such as stratified sampling, systematic sampling, importance sampling, antithetic variates, control variates, etc., that started many years ago (e.g. Kahn, 1956, Moy, 1971, Simon, 1976), or Hendry (1984), and, more recently, Newton (1994), Richard (1996), and Geweke (1996). For instance, a simple method like *antithetic variates* proved to be effective in evaluating the small sample bias of estimators for simultaneous equations (e.g. Hendry and Harrison, 1974, or Mikhail, 1975), or the simulation bias in nonlinear macroeconometric models (e.g. Calzolari, 1979). With slightly more complex implementational requirements, the method of *control variates* proved to be even more effective (Calzolari and Sterbenz, 1986, Sterbenz and Calzolari, 1990).

This paper shows how control variates can be profitably used to reduce the variance of indirect estimators when applied to just-identified models. The method is illustrated with applications to stochastic differential equations. In particular, the Ornstein-Uhlenbeck process, used in Vasicek (1977) to model the short term interest rate in continuous time, and the square-root process, used in Cox, Ingersoll and Ross (1985), are explicitly considered. Results of the Monte Carlo experiments show that the variance component due to simulation can be strongly reduced at about the same computational cost of a simple indirect estimation. This implies a global efficiency gain of almost 50%, which would otherwise require a very large increase in the number of replications.

The rest of the paper is organized as follows. In section 2 we derive the indirect inference control variate estimator for the just identified case. In section 3 we apply the control variates to two well known continuous time models and analyze their computational benefits through a Monte Carlo experiment. Conclusions are drawn in the last section.

2 Indirect inference in the just-identified case

Let the *econometric* model (or model of interest) be represented parametrically as

$$y = f(\theta, e) \tag{2.1}$$

where y is the observable series, $\theta \in \Theta \subseteq \mathbb{R}^p$ is the vector of parameters and e is a white noise with known distribution function (see Gouriéroux, Monfort and Renault, 1993, p.S87). For simplicity, the notation omits exogenous variables and initial values (which are supposed to be asymptotically unimportant).
We assume that this model can be simulated; that is we can produce values of y conditional on the parameters θ by entering random values of e. However, this model may not be estimated, or estimation can be so complex and discouraging that econometricians replace it with an approximation, called the auxiliary model

$$y = g(\beta, \eta) \tag{2.2}$$

whose parameters, $\beta \in B \subseteq \mathbb{R}^q$, can be easily estimated (for example, by maximum likelihood). In addition to the standard assumptions of Gouriéroux, Monfort and Renault (1993) we require that $p = q$ so as to consider only just identified cases, and also that the auxiliary model (2.2) can be simulated by entering random values of η.
We assume that, for any $\theta \in \Theta$, the estimation of (2.2) with the observed series y leads to an estimator of β, say $\beta(\theta, e)$ (that is a function of θ and of the random errors). Under some fairly standard regularity conditions this estimator converges, for $T \to \infty$, to a well defined and regular *binding function* $b(\theta)$, as in Gouriéroux, Monfort and Renault (1993). In finite samples, the estimator will differ from its limit value, the difference being the estimation error of parameters. Thus, for any $\theta \in \Theta$ an estimator of the parameters in (2.2) will be equal to the binding function $b(\theta)$ plus the parameter estimation error (PEER), which depends on the error terms e

$$\beta(\theta, e) = b(\theta) + \text{ PEER}(\theta, e) \tag{2.3}$$

The parameter estimation error $\text{PEER}(\theta, e)$ is a random vector asymptotically vanishing and its precise form (decomposed in first and second order terms) can be found in Gouriéroux and Monfort (1996, section 4.4.1). We may assume that regularity conditions ensure that $\sqrt{T}\ \text{PEER}(\theta, e)$ has an asymptotic zero mean normal distribution with variance-covariance matrix that will be denoted $\Sigma(\theta)$

$$\sqrt{T}\ \text{PEER}(\theta, e) \longrightarrow N(0, \Sigma(\theta)) \tag{2.4}$$

and an expression for $\Sigma(\theta)$ (as well as its estimate) would involve both the Hessian and the matrix (or matrices) of outer products of the first derivatives of the loglikelihoods (e.g. White, 1982, Newey and West, 1987, Gouriéroux and Monfort, 1996, section 4A.2).

Using the actually observed series y, we estimate from equation (2.2) a vector of parameters estimates, say $\hat{\beta}$. If the model of interest (2.1) really is the data generating process, the observed y are a function of the *true* vector of parameters, denoted θ_0, as well as of the unobservable error terms, say e. Therefore, $\hat{\beta}$ turns out to be a function of such parameters θ_0 and of the unobservable error terms e

$$\hat{\beta} = \beta(\theta_0, e) = b(\theta_0) + \text{ PEER}(\theta_0, e) \tag{2.5}$$

By entering a tentative value for the vector of parameters θ, say $\tilde{\theta}$, and pseudo-random error terms $\tilde{e}$ into (2.1), we generate by simulation pseudo-random values of the dependent variable y that are introduced into (2.2). Model (2.2) is estimated and a vector of parameter estimates, say $\tilde{\beta}$, is produced. The sample period can be, of course, of any length, being data produced by simulation, (at least, in the absence of exogenous variables), but we keep for the moment the same sample length T, as for the observed data. Notice that $\tilde{e}$ are generated from "the same" distribution as the unobservable error terms e (apart from an approximation due to the generation of pseudo-random numbers with a computer algorithm). Thus, $\tilde{\beta}$ can be written as

$$\tilde{\beta} = \beta(\tilde{\theta}, \tilde{e}) = b(\tilde{\theta}) + \text{ PEER}(\tilde{\theta}, \tilde{e}) \tag{2.6}$$

By taking advantage of exact identification, we calibrate $\tilde{\theta}$ (keeping $\tilde{e}$ fixed) till we find $\tilde{\beta} = \hat{\beta}$. In other words we look for a value of the vector $\tilde{\theta}$ that solves the system of equations

$$\beta(\tilde{\theta}, \tilde{e}) = \hat{\beta} \tag{2.7}$$

It should be noticed that these equations are only implicitly defined, and cannot be expressed in closed form.

The solution vector will be called $\hat{\theta}$; this is the *simple* indirect estimator of the *econometric* model's parameter vector θ_0. Conditions that ensure consistency and asymptotic normality of this estimator can be found in Gouriéroux, Monfort and Renault (1993) in a more general context.

We can write

$$\hat{\beta} = b(\theta_0) + \text{ PEER}(\theta_0, e) \tag{2.8}$$

$$\tilde{\beta} = b(\hat{\theta}) + \text{ PEER}(\hat{\theta}, \tilde{e}) \tag{2.9}$$

Since, upon convergence of the indirect estimation procedure, $\tilde{\beta} = \hat{\beta}$, we can write

$$\sqrt{T}b(\theta_0) + \sqrt{T}\text{ PEER}(\theta_0, e) = \sqrt{T}b(\hat{\theta}) + \sqrt{T}\text{ PEER}(\hat{\theta}, \tilde{e}) \tag{2.10}$$

and therefore

$$\sqrt{T}\left[b(\hat{\theta}) - b(\theta_0)\right] = \sqrt{T}\ \text{PEER}(\theta_0, e) - \sqrt{T}\ \text{PEER}(\hat{\theta}, \tilde{e}) \tag{2.11}$$

As $\hat{\theta}$ converges to θ_0 (for $T \to \infty$), regularity conditions ensure that $[\sqrt{T}\ \text{PEER}(\hat{\theta}, \tilde{e}) - \sqrt{T}\ \text{PEER}(\theta_0, \tilde{e})] \to 0$. Thus, asymptotically

$$\sqrt{T}\left[b(\hat{\theta}) - b(\theta_0)\right] \simeq \sqrt{T}\ \text{PEER}(\theta_0, e) - \sqrt{T}\ \text{PEER}(\theta_0, \tilde{e}) \tag{2.12}$$

The random error terms e and $\tilde{e}$ are obviously independent and have the same distribution by assumption. Thus the right hand side of (2.12) will be asymptotically $N(0, 2\Sigma(\theta_0))$. Application of the "δ-method" (e.g. Rao, 1973, p.388) to the left hand side expression of (2.12) yields asymptotically

$$\sqrt{T}\left[b(\hat{\theta}) - b(\theta_0)\right] \simeq R_0\sqrt{T}(\hat{\theta} - \theta_0) \qquad R_0 = \left[\frac{\partial b(\theta)}{\partial \theta'}\right]_{\theta_0} \tag{2.13}$$

In our just-identification context the Jacobian R_0 is a square matrix. Assuming that it is nonsingular in some neighbourhood of θ_0, we invert it and obtain, asymptotically

$$\sqrt{T}(\hat{\theta} - \theta_0) \simeq {R_0}^{-1}\sqrt{T}\ \text{PEER}(\theta_0, e) - {R_0}^{-1}\sqrt{T}\ \text{PEER}(\theta_0, \tilde{e}) \tag{2.14}$$

Hence, the asymptotic covariance matrix of the indirect estimator is ${R_0}^{-1}2\Sigma(\theta_0){R_0'}^{-1}$. The fact that $\Sigma(\theta_0)$ is doubled is clearly due to the independence between e and $\tilde{e}$ in equation (2.14).

2.1 Reducing variance by means of replicated simulations

It is well known that the estimation variance can be reduced, with a larger computational cost (e.g. Gouriéroux, Monfort and Renault, 1993).
Let us replace the single simulation-calibration of $\hat{\theta}$ with the average of H replicated simulation-calibrations, say $\hat{\theta}_h$, $h = 1, ..., H$. Each $\hat{\theta}_h$ is the value of $\tilde{\theta}$ that solves the system

$$\beta(\tilde{\theta}, \tilde{e}_h) = \hat{\beta} \tag{2.15}$$

with $\tilde{e}_h$ independently drawn across different replications. Repeating the procedure above, we have

$$\hat{\beta} = b(\theta_0) + \ \text{PEER}(\theta_0, e) \tag{2.16}$$

$$\tilde{\beta}_h = b(\hat{\theta}_h) + \ \text{PEER}(\hat{\theta}_h, \tilde{e}_h) \qquad h = 1, 2, ..., H \tag{2.17}$$

where each $\hat{\theta}_h$ is calibrated till $\tilde{\beta}_h = \hat{\beta}$, and finally all $\hat{\theta}_h$ are averaged to produce

$$\hat{\theta} = \frac{1}{H}\sum_{h=1}^{H}\hat{\theta}_h \tag{2.18}$$

In this case we have, asymptotically

$$\sqrt{T}(\hat{\theta} - \theta_0) \simeq {R_0}^{-1}\left[\sqrt{T}\ \text{PEER}(\theta_0, e) - \frac{1}{H}\sum_{h=1}^{H}\sqrt{T}\ \text{PEER}(\theta_0, \tilde{e}_h)\right] \tag{2.19}$$

where the asymptotic variance-covariance matrix of the term in square brackets is now $(1 + \frac{1}{H})\Sigma(\theta_0)$, because the $\tilde{e}_h$ are independent of each other and of e.
The variance reduction corresponding to a multiplying factor $1 + \frac{1}{H}$ (instead of 2) is obtained at the cost of H calibration procedures instead of just one. In the absence of exogenous variables, the same variance reduction would be obtained if the H procedures with T simulated data were replaced by one procedure with HT simulated data.
Equations like (2.14) or (2.19) make clear the two components that contribute to the variance of the indirect estimator. The first component on the right hand side of both equations depends on e and R_0, thus it is irreducible, given the data, the estimation method, and the models. The second component on the right hand sides of (2.14) and (2.19) depends entirely on simulation, and can be made arbitrarily small by using a large value of H, at the cost of a higher computational effort.

2.2 Reducing variance by means of control variates

The control variate method that we propose in this paper is capable of reducing the variance with (almost) no additional computational cost. The control variates are based on the simulation-estimation of the auxiliary model (2.2). Such a procedure aims at producing an estimation error strongly correlated with $\sqrt{T}\ \text{PEER}(\hat{\theta}, \tilde{e})$ of equation (2.11). We need an additional assumption, beyond those required by simple indirect estimation. We assume that the auxiliary model (2.2) can be simulated using the same pseudo-random errors $\tilde{e}$ used to simulate the econometric model (2.1). If it is not possible to use directly $\tilde{e}$ inside model (2.2), it must be possible to use pseudo-random errors $\tilde{\eta}$ obtained by transformation of $\tilde{e}$, that is $\tilde{\eta} = \eta(\tilde{e})$.
This is not a strong assumption, as it is usually fulfilled in the cases of practical interest. Suppose now we simulate model (2.2) with a given value of β, denoted $\dot{\beta}$, and pseudo-random errors $\tilde{\eta} = \eta(\tilde{e})$, and then re-estimate its parameters. Since this is *direct* estimation of β, not indirect estimation of θ, the usual regularity conditions will ensure that the estimator, $\tilde{\dot{\beta}}$ say, will be asymptotically normally distributed around its pseudo-true value, without the need of further assumptions (there is no need here of introducing a binding function, as it is simply the identity function)

$$\tilde{\dot{\beta}} = \dot{\beta} + \ \text{NPEER}(\dot{\beta}, \eta(\tilde{e})) \tag{2.20}$$

The *new* parameter estimation error (NPEER) is such that $\sqrt{T}$ NPEER$(\dot{\beta}, \eta(\tilde{e}))$ is asymptotically zero mean normal. It is a random vector *controlled* by the experimenter, as it is simply the difference between the estimated parameters and the parameter values introduced into the model.

To construct the control variate estimator we start by taking $\dot{\beta} = \hat{\beta}$, the actual estimate. We then enter $\hat{\beta}$ and $\eta(\tilde{e})$ into the auxiliary model (2.2), and with the series y produced by simulation of (2.2) we re-estimate β and obtain $\tilde{\hat{\beta}}$. Recalling that $\tilde{\beta} = \hat{\beta}$ and using equation (2.9), the vector NPEER can be written as

$$\tilde{\hat{\beta}} - \hat{\beta} = \text{NPEER}(\hat{\beta}, \eta(\tilde{e})) = \text{NPEER}\left\{\left[b(\hat{\theta}) + \text{PEER}(\hat{\theta}, \tilde{e})\right], \eta(\tilde{e})\right\} \tag{2.21}$$

We finally use this NPEER to adjust the simple indirect estimator $\hat{\theta}$

$$\hat{\theta}_{cv} = \hat{\theta} + R^{-1}(\tilde{\hat{\beta}} - \hat{\beta}) = \hat{\theta} + R^{-1}\ \text{NPEER}(\hat{\beta}, \eta(\tilde{e})) \tag{2.22}$$

where $\hat{\theta}_{cv}$ is the control variate estimator of θ_0 and the Jacobian $R = \partial b(\theta)/\partial \theta'$ is evaluated at $\hat{\theta}$.

As $\hat{\theta}$ converges to θ_0 (for $T \to \infty$), regularity conditions ensure that asymptotically

$$\sqrt{T}\ \text{NPEER}(\hat{\beta}, \eta(\tilde{e})) \simeq \sqrt{T}\ \text{NPEER}[b(\theta_0), \eta(\tilde{e})] \tag{2.23}$$

Thus, from (2.14) and (2.22) we get, asymptotically

$$\sqrt{T}(\hat{\theta}_{cv} - \theta_0) = \sqrt{T}\left[(\hat{\theta}_{cv} - \hat{\theta}) + (\hat{\theta} - \theta_0)\right] \tag{2.24}$$
$$\simeq\ {R_0}^{-1}\sqrt{T}\ \text{PEER}(\theta_0, e) + {R_0}^{-1}\left\{\sqrt{T}\ \text{NPEER}[b(\theta_0), \eta(\tilde{e})] - \sqrt{T}\ \text{PEER}(\theta_0, \tilde{e})\right\}$$

Given the independence between e and $\tilde{e}$, the asymptotic variance-covariance matrix of the control variate estimator will be the sum of the covariance matrices of the two components. The first component in equation (2.24) is exactly the same as the irreducible part in the simple indirect estimator, as in equation (2.14). The second component in equation (2.24), in braces, is the difference between two random vectors quite close to each other (similar variance-covariance matrix and strong positive correlation). Therefore, it is quite reasonable to expect a large variance reduction in this second component, when compared with the second term on the right hand side of (2.14).

The additional simulation-estimation is performed just once, at the end of the calibration procedure that has produced the simple indirect estimator $\hat{\theta}$: no further parameter calibration is required. Therefore, the additional cost of the computation is quite small and almost negligible when compared with the cost of computing the indirect estimator.

The new code required to implement the control variate procedure is very little as it only performs simulation and estimation of the auxiliary model, without any further calibration. It should also be noted that the control variates can be applied even when

the indirect inference is performed with $H > 1$, and still it may produce some variance reduction with respect to the last term on the right hand side of equation (2.19).
In general, it is not straightforward to predict the efficiency gain produced by the control variates. Intuitively, if the *econometric* model (2.1) and the *auxiliary* model (2.2) are quite close to each other, the component in braces on the right hand side of equation (2.24) should give very little contribution to the variance of the estimator. In the extreme case of the two models being coincident, such a component disappears and the variance of the estimator would be exactly the same as the variance of the direct estimator. Of course, this is an extreme case where indirect inference would be completely useless.

2.2.1 Remark

Estimating a model with a sample of length HT could be faster than estimating H times the same model with samples of length T. This may happen when the estimation of the auxiliary model requires some iterative optimization procedure, since the number of estimation-iterations usually decreases with the sample length. Conversely, when the auxiliary model estimator has a closed form expression (e.g. OLS), the number of floating-point operations to be performed is nearly proportional to the length of the series. Moreover, on average, the number of calibration-iterations does not decrease with the length of the simulated series. Thus, we do not expect a noticeable difference in the computational speed of the two methods. For instance, in section 3 of this paper the auxiliary models are estimated by OLS, $X'X$ being a 2×2 matrix. If T is not very small, the number of floating-point operations required to compute $(X'X)^{-1}X'y$ with series of length HT is nearly proportional to H, the time to invert the 2×2 matrix being negligible. An accurate set of experiments confirmed that the computing time is roughly the same in the two cases.

3 Examples of control variates and Monte Carlo results

We apply the control variate procedure to two well known continuous time models based on stochastic differential equations, namely the Cox, Ingersoll and Ross (1985) model and the Vasicek (1977) model.
The main problems with the analysis of diffusion processes are that the Markov process solution of the equation is known exactly only for some particular cases, and that variables are actually observed at discrete time intervals. If estimation is conducted through a discrete time model, these approximations induce a (discretization) bias (or inconsistency) in the resulting estimators. Indirect inference adjusts for the discretization bias (inconsistency).
Indirect estimation of diffusion processes is widely exemplified in the recent econometric literature (e.g. Bianchi and Cleur, 1996, Broze, Scaillet and Zakoïan 1994, 1995,

Pastorello, Renault and Touzi 1994). So our discussion will be focussed on the implementation of the control variates.

3.1 Square-root process

The square-root process has been used by Cox, Ingersoll and Ross (1985) for modelling the behaviour of the short term interest rate, and it is defined as

$$dy_t = k(a - y_t)dt + \sigma\sqrt{y_t}\, d\mathcal{W}_t \tag{3.25}$$

where y_t is the spot interest rate and $\mathcal{W}_t$ is a Wiener process.
Our simulation is performed by resorting to a discretization based on the simple Euler scheme. Other types of approximations could be adopted, for instance those proposed by Mihlstein and by Talay (see, for example, Kloeden and Platen, 1992), or the explicit order 2 weak scheme of Gallant and Tauchen (1995). With the Euler scheme, the discretized process used for simulation is

$$y_t - y_{t-\delta} = ka\delta - ky_{t-\delta}\delta + \sqrt{y_{t-\delta}}\,\sqrt{\delta\sigma^2}\, e_t \tag{3.26}$$

Taking (3.25) as a valid characterization of the process generating interest rates, regularity conditions ensure that the discretized model, with a conveniently small discretization step δ, exhibits negligible differences from the corresponding continuous time model. For our purposes, a value $\delta \leq 0.1$ proved to be sufficiently accurate (we have used $\delta = 0.05$; see on this problem also the empirical applications in Bianchi, Cesari and Panattoni, 1995, Broze, Scaillet and Zakoïan, 1995, and Bianchi and Cleur, 1996).
A time unit corresponds to the frequency of the actually observed data. Thus if data are daily, t and $t-1$ refer to consecutive days, and $\delta = 0.05$ (or $\delta^{-1} = 20$) means that 20 data are generated to produce one daily observation.
The auxiliary model we use is

$$y_t - y_{t-1} = ml - my_{t-1} + \sqrt{y_{t-1}}\sqrt{\psi^2}\, \eta_t \tag{3.27}$$

The vector of parameters of interest is $\theta = (a, k, \sigma^2)'$. The corresponding parameters of the auxiliary model are $\beta = (l, m, \psi^2)'$.
A naive estimator $\hat{\beta} = (\hat{l}, \hat{m}, \hat{\psi}^2)'$ of the auxiliary model (3.27) is easily obtained by weighted least squares, that is least squares after data have been divided by $\sqrt{y_{t-1}}$.
To produce the simple indirect estimator $\hat{\theta} = (\hat{a}, \hat{k}, \hat{\sigma}^2)'$ we generate a pseudo-random vector of independent standard normal errors $\tilde{e}$ of length $T \times \delta^{-1}$.
To obtain the control variates, we simulate the auxiliary model (3.27), using $\hat{\beta}$ and a vector of pseudo-random errors $\tilde{\eta} = \eta(\tilde{e})$ of length T, whose tth element is $\tilde{\eta}_t = \sqrt{\delta}(\tilde{e}_{t-1+\delta} + \tilde{e}_{t-1+2\delta} + \cdots + \tilde{e}_{t-\delta} + \tilde{e}_t)$. Weighted least squares estimation yields $\tilde{\hat{\beta}} = (\tilde{\hat{l}}, \tilde{\hat{m}}, \tilde{\hat{\psi}}^2)'$. Finally

the control variate estimator from equation (2.22) is given by

$$\begin{pmatrix} \hat{a}_{cv} \\ \hat{k}_{cv} \\ \hat{\sigma}^2_{cv} \end{pmatrix} = \begin{pmatrix} \hat{a} \\ \hat{k} \\ \hat{\sigma}^2 \end{pmatrix} + R^{-1} \begin{pmatrix} \tilde{\hat{l}} - \hat{l} \\ \tilde{\hat{m}} - \hat{m} \\ \tilde{\hat{\psi}}^2 - \hat{\psi}^2 \end{pmatrix} \tag{3.28}$$

The results of a Monte Carlo exercise, presented in Table 1, are related to a sample period of length $T = 1000$. For the simulated series we use $H = 1$ and $H = 10$. Also control variates have been applied with $H = 1$ and $H = 10$. Sample means and variances of estimated parameters have been computed from 10000 Monte Carlo replications; about 0.1% replications have been discarded due to convergence problems in the calibrations. The last row of the table displays the computing times (in seconds of CPU on a Pentium-Pro 200).
For the variance parameter σ^2 (ψ^2 in the auxiliary model), results are also displayed in graphical form in Figure 1.

TABLE 1: Square root process

$$dy_t = 0.5\,(\,0.1 - y_t\,)dt + 0.1\,\sqrt{y_t}\;d\mathcal{W}_t$$

Monte-Carlo means and (variances) of estimated parameters

	Econometric Model				Auxiliary Model	
	H=1		H=10			
	Ind.Inf.	Cntr.Var.	Ind.Inf.	Cntr.Var.		Least Squares
a	0.1000 (8.0×10^{-6})	0.1000 (4.2×10^{-6})	0.1000 (4.5×10^{-6})	0.1000 (4.1×10^{-6})	l	0.1000 (4.1×10^{-6})
k	0.5030 (3.7×10^{-3})	0.5039 (2.1×10^{-3})	0.5053 (2.0×10^{-3})	0.5049 (1.9×10^{-3})	m	0.4001 (6.9×10^{-4})
σ^2	0.0101 (6.6×10^{-7})	0.0100 (3.7×10^{-7})	0.0100 (3.6×10^{-7})	0.0100 (3.3×10^{-7})	ψ^2	0.0066 (9.9×10^{-8})
CPU	4 989 s	5 250 s	49 750 s	52 515 s		

T=1000 Replications=10000

The least squares estimator of the auxiliary model shows a quite evident bias for the second and third parameters (mean reversion and variance). Indirect inference (with or without control variates) adjusts for the bias (inconsistency): the mean estimated parameter is very close to the *true* value.
Control variates produce a remarkable reduction in the variance of the parameter estimates, with respect to simple indirect inference. Let us first look at the estimates with $H = 1$. The variance of the long run mean parameter (4.2×10^{-6}) is nearly the same as for the least squares estimator, while for the simple indirect estimator it is almost the double (8.0×10^{-6}). This, however, is of no interest given that the direct estimator is practically unbiased (consistent) for this parameter.
Of great interest is the benefit produced by the use of control variates for the mean

reversion parameter (k) and for the variance parameter (σ^2). The Monte Carlo variance of the simple indirect estimate of k is 3.7×10^{-3}, and reduces to 2.1×10^{-3} if control variates are used. Thus, there is a reduction of about 43% in the global variance of the estimator at almost the same computational cost (only 5% more, on average). About the same gain is achieved for the σ^2 parameter, whose variance decreases from 6.6×10^{-7} to 3.7×10^{-7}.
To get an efficiency gain of this magnitude, the simple indirect estimator requires $H = 10$, as it is evident from the table. From our experiments the computing time for doing it was about 10 times larger than for $H = 1$ (more than 9 times larger than the control variates). Interestingly, if we decide that we can afford the computational cost of $H = 10$, the further application of control variates is quite worthwhile. At a small additional cost, it reduces the variance to approximately a half of the simple indirect estimator with $H = 1$: and obviously no further improvement could be possible.

3.2 Ornstein-Uhlenbeck process

A simpler model for the behaviour of short-term interest rates was proposed by Vasicek (1977), who employed the Ornstein-Uhlenbeck process

$$dy_t = k(a - y_t)dt + \sigma \, d\mathcal{W}_t \tag{3.29}$$

For this process a simple exact discretization is available (e.g. Gouriéroux, Monfort and Renault, 1993, p.S102)

$$y_t = a(1 - e^{-k}) + e^{-k} y_{t-1} + \sigma \left(\frac{1 - e^{-2k}}{2k} \right)^{1/2} e_t \tag{3.30}$$

The auxiliary model we use is

$$y_t - y_{t-1} = ml - m y_{t-1} + \sqrt{\psi^2} \, \eta_t \tag{3.31}$$

The parameters are the same as in the previous section, both for the econometric model and for the auxiliary model; and here also we obtain the estimator $\hat{\beta} = (\hat{l}, \hat{m}, \hat{\psi}^2)'$ of the auxiliary model (3.31) by ordinary least squares.
The only remarkable difference from the example of the previous section is in the simulation of the auxiliary model (3.31) to produce the control variates. Here in fact we do not need to transform the vector $\tilde{e}$, and we directly use $\tilde{\eta} = \tilde{e}$, which is of length T.
Table 2 reports the results of the Monte Carlo experiment. For the variance parameter, results are also displayed in graphical form in Figure 2. Sample means and variances of estimated parameters have been computed from 10000 Monte Carlo replications; no convergence problems occurred in the calibrations.
The benefit from the application of control variates is even more impressive than for previous model. With $H = 1$, an additional computational cost of about 5% reduces

TABLE 2: Ornstein-Uhlenbeck process

$dy_t = 0.5\,(\,0.1 - y_t\,)dt + 0.1\;dW_t$

Monte-Carlo means and (variances) of estimated parameters

	Econometric Model					Auxiliary Model	
	H=1		H=10		H=50		
	Ind.Inf.	Cntr.Var.	Ind.Inf.	Cntr.Var.	Ind.Inf		Least Squares
a	0.1000 (8.2×10^{-5})	0.1000 (4.1×10^{-5})	0.1000 (4.4×10^{-5})	0.1000 (4.0×10^{-5})	0.1000 (4.1×10^{-5})	l	0.1000 (4.0×10^{-5})
k	0.5029 (3.6×10^{-3})	0.5044 (1.8×10^{-3})	0.5057 (1.9×10^{-3})	0.5056 (1.7×10^{-3})	0.5053 (1.8×10^{-3})	m	0.3962 (6.4×10^{-4})
σ^2	0.0100 (6.5×10^{-7})	0.0100 (3.2×10^{-7})	0.0100 (3.5×10^{-7})	0.0100 (3.2×10^{-7})	0.0100 (3.3×10^{-7})	ψ^2	0.0063 (8.0×10^{-8})
CPU	539 s	566 s	5 371 s	5 648 s	26 235 s		

T=1000 Replications=10000

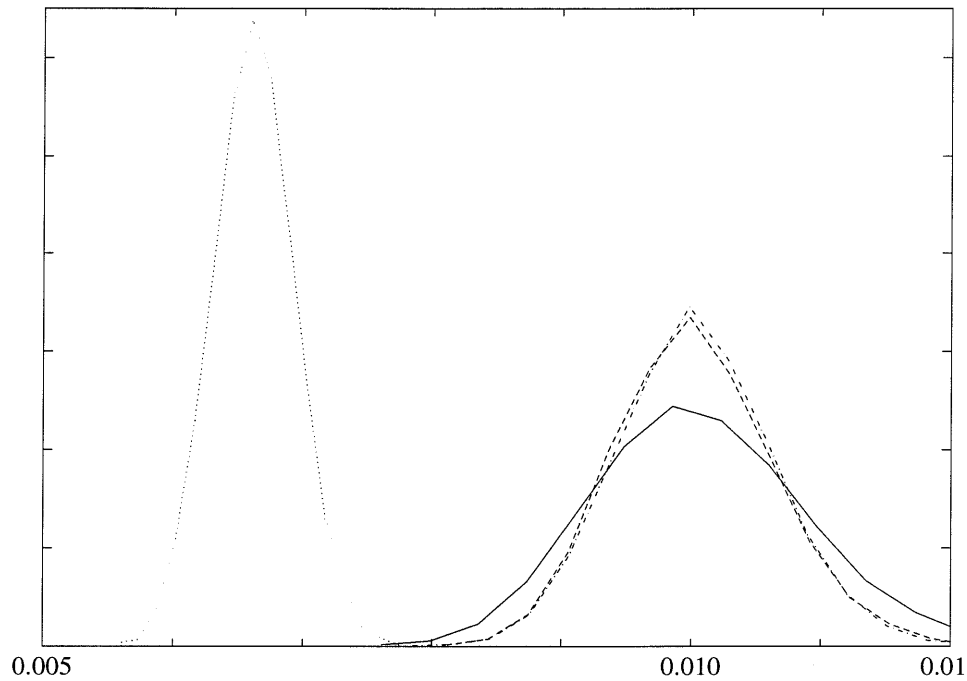
the variance by nearly 50%, which is the largest possible reduction. Simple indirect estimation does not reach the same efficiency either with $H = 10$, or with $H = 50$, and the computational cost is enormously higher.

4 Conclusion

We have shown in this paper that control variates can help in improving the efficiency of indirect estimators in the just-identified case. The paper has shown in some detail how the control variates can act on that part of the variance that depends on the simulation. At about the same computational cost (that is, computing time), an indirect estimator with control variates can be as efficient as a simple indirect estimator based on much longer simulated series, which involves a much higher cost.

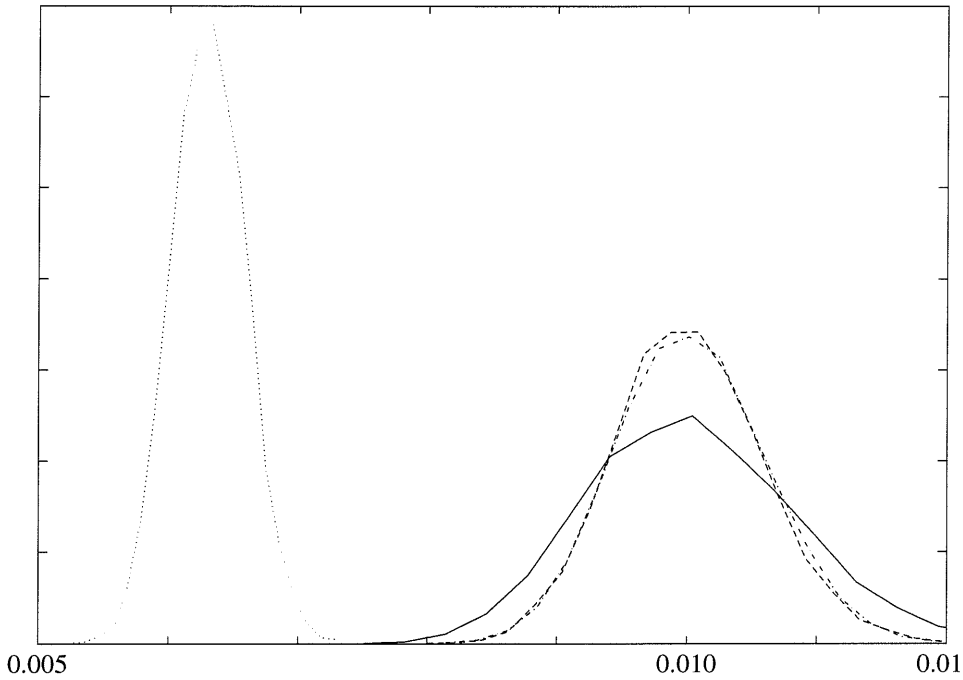
Control variates have been applied to a couple of well known models of the short term interest rate in continuous time, and proved to be quite effective in a large set of Monte Carlo experiments.

Figure 1: Square root process. Monte Carlo distribution of variance parameter



$\cdots$ LEAST S. —— I. I. H=1 $-\cdot-$ I. I. H=10 $---$ CNTR. VAR. $H = 1$

Figure 2: Ornstein-Uhlenbeck process. Monte Carlo distribution of variance parameter



$\cdots$ LEAST S. —— I. I. H=1 $-\cdot-$ I. I. H=50 $---$ CNTR. VAR. $H = 1$

References

Bianchi, C., R. Cesari, and L. Panattoni (1995): "Alternative Estimators of the Cox, Ingersoll and Ross Model of the Term Structure of Interest Rates: A Monte Carlo Study." Roma: Banca d'Italia, Temi di Discussione N.326.

Bianchi, C., and E. M. Cleur (1996): "Indirect Estimation of Stochastic Differential Equation Models: Some Computational Experiments," *Computational Economics* 9, 257-274.

Broze, L., O. Scaillet, and J. M. Zakoïan (1994): "Quasi Indirect Inference for Diffusion Processes." Paris: Crest, Document de Travail No. 9511, forthcoming in *Econometric Theory.*

Broze, L., O. Scaillet, and J.M. Zakoïan (1995): "Testing for Continuous Time Models of the Short-Term Interest Rate," *Journal of Empirical Finance* 2, 199-223.

Calzolari, G. (1979): "Antithetic Variates to Estimate the Simulation Bias in Non-Linear Models," *Economics Letters* 4, 323-328.

Calzolari, G., and F. P. Sterbenz (1986): "Control Variates to Estimate the Reduced Form Variances in Econometric Models," *Econometrica* 54, 1483-1490.

Cox, J. C., J. Ingersoll, and S. A. Ross (1985): "A Theory of Term Structure of Interest Rates," *Econometrica* 53, 385-408.

Gallant, A. R., and G. Tauchen (1995): "Estimation of Continuous Time Models for Stock Returns and Interest Rates." Duke University, Department of Economics, mimeo.

Gallant, A. R., and G. Tauchen (1996): "Which Moments to Match?" *Econometric Theory* 12, 657-681.

Geweke, J. (1996): "Monte Carlo Simulation and Numerical Integration." in *Handbook of Computational Economics*, ed. by H.M. Amman, D.A. Kendrick, and J. Rust. Amsterdam: Elsevier Science B.V. Vol. I, 731-800.

Gouriéroux, C., and A. Monfort (1996): *Simulation-Based Econometric Methods.* Oxford University Press.

Gouriéroux, C., A. Monfort, and E. Renault (1993): "Indirect Inference," *Journal of Applied Econometrics* 8, S85-S118.

Hendry, D. F. (1984): "Monte Carlo Experimentation in Econometrics," in *Handbook of Econometrics*, ed. by Z. Griliches, and M. D. Intriligator. Amsterdam: North-Holland Publishing Company, Vol. II, 937-976.

Hendry, D. F., and R. W. Harrison (1974): "Monte Carlo Methodology and the Small Sample Behaviour of Ordinary and Two-Stage Least Squares," *Journal of Econometrics* 2, 151-174.

Kahn, H. (1956): "Use of Different Monte Carlo Sampling Techniques," in *Symposium on Monte Carlo Methods*, ed. by H. A. Meyer. New York: John Wiley & Sons, Inc., 146-190.

Kloeden, P. E., and E. Platen (1992): *Numerical Solution of Stochastic Differential Equations*. Berlin: Springer-Verlag.

Mikhail, W. M. (1975): "A Comparative Monte Carlo Study of the Properties of Econometric Estimators," *Journal of the American Statistical Association* 70, 94-104.

Moy, W. A. (1971): "Variance Reduction," in *Computer Simulation Experiments with Models of Economic Systems*, ed. by T. H. Naylor. New York: John Wiley & Sons, Inc., 269-289.

Newey, W. K., and K. D. West (1987): "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica* 55, 703-708.

Newton, N. J. (1994): "Variance Reduction for Simulated Diffusions," *SIAM J. Appl. Math.* 54, 1780-1805.

Pastorello, S., E. Renault, and N. Touzi (1994): "Statistical Inference for Random Variance Option Pricing." Università di Padova, Dipartimento di Scienze Economiche, Quaderno No. 31.

Rao, C. R. (1973): *Linear Statistical Inference and its Applications*. New York: John Wiley & Sons, Inc.

Richard, J. F. (1996): "Simulation Techniques," in *The Econometrics of Panel Data*, ed. by L. Matya, and P. Sevestre. Dordrecht: Kluwer Academic Publishers, 613-638.

Simon, G. (1976): "Computer Simulation Swindles, with Applications to Estimates of Location and Dispersion," *Applied Statistics, Journal of the Royal Statistical Society, series C* 25, 266-274.

Smith, A. (1993): "Estimating Nonlinear Time-Series Models Using Simulated Vector Autoregressions," *Journal of Applied Econometrics* 8, S63-S84.

Sterbenz, F. P., and G. Calzolari (1990): "Alternative Specifications of the Error Process in the Stochastic Simulation of Econometric Models," *Journal of Applied Econometrics* 5, 137-150.

Vasicek, O.A. (1977): "An Equilibrium Characterization of the Term Structure," *Journal of Financial Economics* 5, 177-88.

White, H. (1982): "Maximum Likelihood Estimation of Misspecified Models," *Econometrica* 50, 1-25.