Control variates for variance reduction in indirect inference: interest rate models in continuous time

Calzolari, Giorgio and Di Iorio, Francesca and Fiorentini, Gabriele

Universita' di Firenze, CEIBS - Jiaotong University - Shanghai, ISTAT - Roma, Universidad de Alicante

November 1996

Control Variates for Variance Reduction in Indirect Inference: Interest Rate Models in Continuous Time*

Giorgio CALZOLARI§
Università di Firenze
and
CEIBS, Jiaotong University, Shanghai

Francesca DI IORIO¶
ISTAT, Roma

Gabriele FIORENTINI‖
Universidad de Alicante

November 1996

KEYWORDS. Efficient Monte Carlo, Variance reduction techniques, Control variates, Indirect inference, Euler discretization, Short-term interest rate, Stochastic differential equation.

*For presentation at the seventh edition of the *European Conferences of the Econometrics Community EC*², Firenze, 12-14 December, 1996. We have benefitted from discussions with Carlo Bianchi, Eugene M. Cleur, and Eric Renault, but retain full responsibility for the contents of this paper.

§Università di Firenze, Dipartimento di Statistica "G. Parenti", Viale Morgagni 59, I-50134 Firenze, Italy. Fax (+39) 55 4223560.
CEIBS, China Europe International Business School, Jiaotong University, Minhang, 800 Dong Chuan Road, Shanghai 200240, P. R. China. Fax: (+86) 21 64358928.

¶ISTAT, Istituto Nazionale di Statistica, Via C. Balbo 16, I-00184 Roma, Italy. Fax (+39) 6 46734173.

‖Universidad de Alicante, Departamento de Fundamentos del Analisis Economico, Apdo 99, E-03080 Alicante, Spain. Fax. (+34) 6 5903614.

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. There is, in fact, an additional component of the variance, which depends on the stochastic simulation involved in the estimation procedure. To reduce this undesirable effect, one should properly increase the number of simulations (or the length of each simulation) and thus the computation cost. Alternatively, this paper shows how variance reduction can be achieved, at virtually no additional computation cost, by use of control variates. The Ornstein-Uhlenbeck equation, used by Vasicek to model the short term interest rate in continuous time, and the so called *square root* equation, used by Cox, Ingersoll and Ross, are explicitly considered and experimented with. Monte Carlo experiments show that, for some parameters of interest, a global efficiency gain about 35%-45% over the simplest indirect estimator is obtained at about the same computation cost.

摘要

诸如间接推理或模拟极大似然之类的模拟估计量，被成功地用来估计随机差分方程。它们调整连续时间内内在随机过程的离散化所带来的偏差(不相容性)。因此而付出的代价是被估计的参数增加方差。事实上，方差还有一个组成部分，它依赖于估计过程中涉及的随机模拟。减少这一不利影响，应该适当地增加模拟的数目(或每个模拟的长度)，并因此增加计算成本。本文提出的另一方案是如何应用控制变量来达到方差缩减，并且实际上不增加计算成本。被凡西塞克(Vasicek)用作连续时间内的短期利率模型的欧斯坦－乌伦贝克(Ornstein-Uhlenbeck)方程，以及被考克斯(Cox)、英格索尔(Ingersoll)和罗斯(Ross)应用的所谓的"平方根"方程，在文中有详细的讨论并用实验验证。蒙特卡罗(Monte Carlo)实验表明，对利率的一些参数，用大约相同的计算成本，能获得比最简单的间接估计量高出约35%－45%的总体效率。

关键词汇:

Simulation	模拟	stochastic process	随机过程
estimator	估计量	continuous time	连续时间
inference	推理	variance	方差
maximum likelihood	极大(最大)似然	parameter	参数
stochastic differential equation	随机差分方程	variance reduction	方差缩减
bias	偏差	control variate	控制变量
inconsistency	不相容性	starting point	起始值
discretization	离散化		

1 Introduction

Simulation estimators, such as indirect inference (Gourieroux, Monfort and Renault, 1993), simulated maximum likelihood (Smith, 1993) or efficient method of moments (Gallant and Tauchen, 1992) 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, in fact, the variance is due to the intrinsic stochastic nature of the data and to the models adopted; on the other hand, it also 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 computation effort. Therefore a trade-off arises between variance reduction and computation 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 computation cost. There is a wide literature on efficient Monte Carlo techniques, such as stratified 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), Geweke (1994) and Richard (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 implementation requirements, the method of *control variates* proved to be even more effective (Sterbenz and Calzolari, 1990). This method also is suitable for evaluating variances, where antithetic variates fail (Calzolari and Sterbenz, 1986).

This paper shows how control variates can be profitably used to reduce the variance of indirect estimators when applied to stochastic differential equations. The Ornstein-Uhlenbeck process, used in Vasicek (1977) to model the short term interest rate in continuous time, and the so called *square root* process, used in Cox, Ingersoll and Ross (1985), are explicitly considered and experimented with.

Results of the Monte Carlo experiments show that, for some parameters of interest, the variance component due to simulation can be reduced 4-5 times at about the same computation cost. This implies a global efficiency gain of 35%-45% over the simplest indirect estimator (reduction of the global variance).

2 Intuitive introduction to control variates for indirect estimators in the just-identified case

For the models considered in this paper, introducing control variates turns out to be particularly simple. We consider, in fact, the Vasicek (1977) and the Cox, Ingersoll and Ross (1985) models for short term interest rates in continuous time, each of which is based on a particular stochastic differential equation (the *econometric* model). The former is based on the Ornstein-Uhlenbeck process, the latter is based on the so called *square root* process. Each model is usually approximated by a discretized equation whose parameters maintain a close one-to-one correspondence with the parameters of the continuous time econometric model. Examples can be found in Bianchi and Cleur (1996), Broze, Scaillet and Zakoïan (1994, 1995), Cleur (1995), Di Iorio (1996), Pastorello, Renault and Touzi (1994).

For simplicity, we omit from our notation exogenous variables (which are not included in the particular models considered), initial values (which are supposed to be asymptotically not influent), and the distribution of the error terms (which is supposed to be known, for example i.i.d. standard normal).

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

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

and 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, like

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

which can be easily estimated (for example, by maximum likelihood).

We assume that, for any $\theta \epsilon \Theta$, and a given probability distribution for e, values of y can be generated from (2.1) and that the estimation of (2.2) with these values of y leads to an estimator of β, say $\beta(\theta, e)$ (that is a function of θ and of the random errors). We assume that some fairly standard regularity conditions ensure that this estimator converges, for $T \to \infty$, to a well defined and regular *binding function* $b(\theta)$, for any $\theta \epsilon \Theta$, as in Gourieroux, Monfort and Renault (1993). In finite samples, obviously the estimator will differ from its limit value, the difference being the finite sample estimation error of parameters. Thus, in the finite sample case, for any $\theta \epsilon \Theta$ an estimator of the parameters in (2.2) will be equal to the binding function $b(\theta)$ plus the parameter estimation error (PEER), which is a random vector due to the particular finite sample of error terms e. These error terms are assumed

to have a known distribution, i.i.d. standard normal in our experiments, variance being already included in the parameter vector θ

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

The parameter estimation error $\text{PEER}(\theta, e)$ is a random vector asymptotically vanishing and we may assume that regularity conditions ensure for $\sqrt{T}\ \text{PEER}(\theta, e)$ an asymptotic zero mean normal distribution with variance-covariance matrix that will be denoted[1] Σ

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

Since the covariance matrix is obtained from the misspecified model in (2.2), it is well known from White (1982) that an expression for Σ (as well as its estimate) would involve both the Hessian and the matrix of outer products of the first derivatives of the likelihoods.

By entering the historically observed values of y, we estimate from equation (2.2) a vector of parameters, say $\hat{\beta}$. If the model of interest (2.1) really is the data generating process, the historically observed y are a function of the *true* vector of parameters, say 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 vector of parameters θ and pseudo-random error terms $\tilde{e}$ into (2.1), we generate by simulation pseudo-random values $\tilde{y}$ that are introduced into (2.2). Model (2.2) is estimated and a vector of parameters, say $\tilde{\beta}$, is produced. The sample period can be, of course, of any length, being data produced by simulation, but we keep for the moment the same sample length as for the historically observed data, say T. Notice that $\tilde{e}$ are generated from "the same" distribution as the unobservable historical error terms e. Thus, $\tilde{\beta}$ is a function (the same function as before) of the tentative parameters θ and of the pseudo-random error terms, say

[1] An example where indirect inference is completely unnecessary might be helpful to fix ideas. Let us suppose one is dealing with a linear regression model with nonrandom exogenous regressors, under standard textbook conditions

$$y = X\theta + e$$

with the additional condition that the variance of the i.i.d. e's is known = 1 (not helpful for estimation of θ, as well known).

The econometric and auxiliary models are coincident, so using OLS we get

$$\beta(\theta, e) = (X'X)^{-1}X'y = \theta + (X'X)^{-1}X'e$$

thus $b(\theta) = \theta$ (the binding function is the identity function), $\text{PEER}(\theta, e) = (X'X)^{-1}X'e$ (not a function of θ, and asymptotically vanishing), and asymptotically $\sqrt{T}\ \text{PEER}(\theta, e) \rightarrow N[0, \lim(X'X/T)^{-1}]$.

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

In this case of exact identification, indirect inference procedures can take advantage of the one-to-one correspondence between θ and β parameters. We calibrate the θ parameters (keeping $\tilde{e}$ fixed) till we find $\tilde{\beta} = \hat{\beta}$. In other words we look for the values of θ that solve the system of equations

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

The calibration procedure thus aims at solving the system of equations (2.7). These equations are only implicitly defined, and cannot be expressed in closed form.

The solution vector will be called $\hat{\theta}$; this is the 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 Gourieroux, Monfort and Renault (1993) in a more general context. In our context, if estimation of the auxiliary model (2.2) is performed by quasi-maximum-likelihood, the estimator turns out to be identical to the simulated QML (Smith, 1993).

We can write

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

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

The left hand sides will be equal upon convergence of the indirect estimation procedure, thus equating the right hand sides and multiplying by $\sqrt{T}$

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

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 $T \rightarrow \infty$), regularity conditions ensure that, asymptotically, the random vector $\sqrt{T}\ \text{PEER}(\hat{\theta}, \tilde{e})$ is the same as $\sqrt{T}\ \text{PEER}(\theta_0, \tilde{e})$. 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, as the former are the unobservable errors in the historical process, while the latter are generated by simulation.

Thus variances must be summed. As the distribution of e and $\tilde{e}$ is the same by assumption (i.i.d. N(0,1), in our experiments), the variance turns out to be simply double. Thus the right hand side of (2.12) will be $N(0, 2\Sigma)$, asymptotically.

Applying the "δ-method" (e.g. Rao, 1973, p.388) the left hand side of (2.11) has asymptotically the same distribution as

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

We take advantage of the one-to-one correspondence between the θ and β parameters; thus, the Jacobian R_0 is a square matrix. Assuming that it is nonsingular in some neighbourhood of θ_0, we invert the Jacobian obtaining, asymptotically

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

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

2.1 Reducing variance by means of replicated simulations, with larger computation cost

It is well known that this variance can be reduced, with a larger computation cost.

In fact, let us replace the single simulation-calibration of $\hat{\theta}$ with the average of H replicated simulations-calibrations, say $\hat{\theta}_h$, $h = 1, ..., H$. Each $\hat{\theta}_h$ is the value of θ that solves the system

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

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

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

[2] In the simple example given in the previous footnote, the Jacobian is the unit matrix. Thus, asymptotically

$$\sqrt{T}(\hat{\theta} - \theta_0) \simeq \sqrt{T}(X'X)^{-1}X'e - \sqrt{T}(X'X)^{-1}X'\tilde{e}$$

The two components are independent random vectors, each of which has variance-covariance matrix $(X'X/T)^{-1}$. So the asymptotic variance-covariance matrix of the indirect estimator will simply be the double of that of the OLS estimator, that is $2lim(X'X/T)^{-1}$.

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

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

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

Instead of equation (2.14), we have in this case, asymptotically

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

where the asymptotic variance-covariance matrix of the term in square brackets is now $(1 + \frac{1}{H})\Sigma$, being the $\tilde{e}_h$ 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.[3] Notice that the same result would be obtained, at the same computational cost, if the H procedures with T data were replaced by one procedure with HT simulated data (Gourieroux, Monfort and Renault, 1993, section 2.3).

Equations like (2.14) or (2.19) quite clearly evidence 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 used. The second component on the right hand sides of (2.14) and (2.19) depends entirely on simulation, and can be made arbitrarily small, at the cost of a large simulation effort.

2.2 Reducing variance by means of control variates, without additional computation cost

We now push further ahead the strict one-to-one correspondence between the θ and the β parameters, by considering that, for the models at hand, they are *the same* parameters, plugged into two different models.

Suppose that we generate values of the y's from simulation of the approximated model (2.2) instead of the *econometric* model of interest (2.1). Then model (2.2) will no longer be misspecified for these new simulated values of y. If we estimate model (2.2) using these new simulated values of y, the estimator will be consistent and asymptotically normal.

[3] Still using the example found in the previous footnotes, the asymptotic variance-covariance matrix of the indirect estimator will be $(1 + \frac{1}{H})$ times that of the OLS estimator, $(1 + \frac{1}{H})lim(X'X/T)^{-1}$.

without need of further assumptions. There is no need here of introducing a binding function, as it is simply the identity function.

We may also use different auxiliary models, instead of just the approximated model (2.2), all sharing the feature of being at the same time simulable and estimable, as well as being approximations to the *econometric* model of interest (2.1). We shall call them *the control variate models*. To avoid introduction of new symbols, we write one of these control variate models directly using the θ parameters

$$y = p(\theta, e) \tag{2.20}$$

This model will be simulated as well as estimated with simulated data (if the control variate model is the model (2.2), the function p will be the function g in (2.2), but re-written with θ instead of β).

The necessary features of any control variate model (2.20) are the following.

1) The parameters must be the same as those of the *econometric* model, θ.

2) It must be possible to plug into the model *the same* pseudo-random errors $\tilde{e}$ already introduced into model (2.1).

3) It must be possible to simulate and estimate the model.

For any $\theta \epsilon \Theta$ and any error terms $\tilde{e}$ generated from the usual distribution, if we simulate model (2.20) and re-estimate its parameters, the estimated parameters will differ from θ by a *new* estimation error

$$\tilde{\theta} = \theta + \text{NPEER}(\theta, \tilde{e}) \tag{2.21}$$

Under suitable regularity conditions, the new parameter estimation error NPEER will be asymptotically vanishing and $\sqrt{T}$ NPEER$(\theta, \tilde{e})$ will be asymptotically zero mean normal. Of course it will not be equal to the parameter estimation error PEER$(\theta, \tilde{e})$ of equation (2.6), as simulations of the y's are made in different ways. However, suppose that the control variate model (2.20) is a close approximation of the *econometric* model, and that the $\tilde{e}$ pseudo-random errors are the same used in the calibration procedure; it is quite reasonable to expect that $\sqrt{T}$ PEER$(\theta, \tilde{e})$ and $\sqrt{T}$ NPEER$(\theta, \tilde{e})$ are random vectors with variance-covariance matrices sufficiently close to each other and with a strong positive correlation for any $\theta \epsilon \Theta$.

This additional simulation-estimation is performed just once, at the end of the calibration procedure that has produced the 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. As a value for θ, we adopt the converged value $\hat{\theta}$; as pseudo-random errors $\tilde{e}$, we use the same already used in the calibration procedure.

The estimated vector will be called $\hat{\tilde{\theta}}$. It is used to compute, by difference

$$\text{NPEER}(\hat{\theta}, \tilde{e}) = \hat{\tilde{\theta}} - \hat{\theta} \tag{2.22}$$

and we use this to compute the control variate indirect estimator as

$$\hat{\theta}_{cv} = \hat{\theta} + R_0^{-1} \text{ NPEER}(\hat{\theta}, \tilde{e}) \tag{2.23}$$

As $\hat{\theta}$ converges to θ_0 (for $T \to \infty$), regularity conditions may ensure that the random vector $\sqrt{T}$ NPEER$(\hat{\theta}, \tilde{e})$ is asymptotically the same as $\sqrt{T}$ NPEER$(\theta_0, \tilde{e})$. Thus, from (2.14) and (2.23) we get, asymptotically

$$\sqrt{T}(\hat{\theta}_{cv} - \theta_0) = \sqrt{T}\left[(\hat{\theta}_{cv} - \hat{\theta}) + (\hat{\theta} - \theta_0)\right] \tag{2.24}$$

$$\simeq R_0^{-1}\sqrt{T}\ \text{PEER}(\theta_0, e) + R_0^{-1}\left[\sqrt{T}\ \text{NPEER}(\theta_0, \tilde{e}) - \sqrt{T}\ \text{PEER}(\theta_0, \tilde{e})\right] \tag{2.25}$$

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 is exactly the same as the irreducible part in the simple indirect estimator, as in equation (2.14). The second component, in square brackets, 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 strong variance reduction in this second component, when compared with the second term on the right hand side of (2.14).[4]

[4] In the footnote example, we have

$$\hat{\tilde{\theta}} = \hat{\theta} + (X'X)^{-1}X'\tilde{e}$$

$$\text{NPEER}(\hat{\theta}, \tilde{e}) = (X'X)^{-1}X'\tilde{e}$$

This is exactly equal to PEER$(\hat{\theta}, \tilde{e})$.

$$\hat{\theta}_{cv} = \hat{\theta} + \text{NPEER}(\hat{\theta}, \tilde{e}) = \hat{\theta} + (X'X)^{-1}X'\tilde{e}$$

$$\sqrt{T}(\hat{\theta}_{cv} - \theta_0) = \sqrt{T}\left[(\hat{\theta}_{cv} - \hat{\theta}) + (\hat{\theta} - \theta_0)\right]$$

$$= \sqrt{T}(X'X)^{-1}X'e + \left[\sqrt{T}(X'X)^{-1}X'\tilde{e} - \sqrt{T}(X'X)^{-1}X'\tilde{e}\right]$$

The term in square brackets is zero. Therefore, the control variates indirect estimator turns out to have the same variance as the OLS estimator; that is, half the variance of the simple indirect estimator

2.3 Remark

It is difficult to predict the efficiency gain produced by the control variates. Intuitively, if the *econometric* model (2.1) and the two approximations (2.2) and (2.20) are quite close to each other, the last component in square brackets on the right hand side of equation (2.25) should give very little contribution to the variance of the estimator. In the extreme case of the three 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; for the simple indirect estimator the variance would be double. Of course, this extreme case is an example where indirect inference is completely useless, like the example in the footnotes.

For cases of practical interest, the efficiency gain can be evaluated by means of Monte Carlo experiments. This will be undertaken later in this paper.

3 The econometrics of continuous time interest rate models

Generally, the behaviour of short-term interest rates is assumed to be explained by a stochastic differential equation defined by

$$dr_t = \mu(r_t, \theta)dt + q(r_t, \theta)d\mathcal{W}_t \tag{3.26}$$

where r_t is the spot interest rate, $\mathcal{W}_t$ is a Wiener process, and θ is a vector of parameters. The functions $\mu(\cdot)$ and $q(\cdot)$ are denominated "drift" and "diffusion" respectively, and satisfy the uniqueness and existence conditions of the solution. These conditions are of the Lipschitz type and are given, for example, in Arnold (1975) and Ikeda and Watanabe (1989).

Recently, the interest of financial analysts and researchers in the empirical analysis of continuous time interest rate models has increased considerably. The main problems with the analysis of an equation of the type defined above are that the Markov process, solution of the equation, is known exactly only for some particular cases, and that interest rates are actually observed at discrete time intervals. In general, it is possible to obtain a (computationally tractable) approximate solution based on some discretization scheme. Particularly simple is the so-called Euler scheme. Other types of approximations are those proposed by Mihlstein and by Talay (see Kloeden and Platen, 1992). However, if estimation is conducted through a discrete time econometric model, these approximations induce a (discretization) bias (or inconsistency) in the resulting estimators of the parameters of interest. Indirect inference adjusts for the discretization bias (inconsistency).

In an attempt to compare different models, Chan, Karolyi, Longstaff and Schwartz (1992) first recognize that, regardless of the approach (arbitrage or equilibrium), most term

structure models imply dynamics for the short-term riskless rate that can be embedded in the following expression

$$dr_t = k(a - r_t)dt + \sigma r_t^{\gamma} d\mathcal{W}_t \tag{3.27}$$

This process bears several important economic features. The specification for the drift implies a mean reverting effect toward the long-run equilibrium level a; the value of k gives the importance of this effect (the speed of adjustment). The diffusion allows the volatility of interest rate changes to be sensitive to the level of the short-term rate. The variance elasticity parameter γ gives a measure of this sensitivity. Finally, σ is a scale parameter.

When $\gamma = 0$, we obtain the Ornstein-Uhlenbeck omoskedastic process employed by Vasicek (1977)

$$dr_t = k(a - r_t)dt + \sigma d\mathcal{W}_t \tag{3.28}$$

For $\gamma = 1/2$, we obtain the square root process of the Cox, Ingersoll and Ross (1985) model

$$dr_t = k(a - r_t)dt + \sigma\sqrt{r_t}d\mathcal{W}_t \tag{3.29}$$

When $\gamma = 1$, we obtain the Brennan and Schwartz (1979) model (that, however, will not be considered in our experiments).

With the notation of the previous section, the vector of parameters of interest is $\theta = (a, k, \sigma^2)'$.

Empirical estimation is usually performed on a discretized version of the model. Discretization in most available applications is based on the following Euler scheme

$$r_t - r_{t-1} = ka - kr_{t-1} + \epsilon_t \tag{3.30}$$

where $E_{t-1}(\epsilon_t) = 0$ and $E_{t-1}(\epsilon_t^2) = \sigma^2$ (Ornstein-Uhlenbeck) or $E_{t-1}(\epsilon_t^2) = (\sigma\sqrt{r_{t-1}})^2$ (square root process). The parameters of the discretized model are the same as for the continuous time model: $\beta = (a, k, \sigma^2)'$ (strict correspondence between the θ and the β parameters).

A *naive* estimator of the discretized model (3.30) is immediately available, since maximum likelihood = least squares for this model. For the square root process, data must be first divided by $\sqrt{r_{t-1}}$.

The main drawback of this econometric approach is that the discrete time approximation can have a non negligible cost in terms of statistical properties of the estimators. The

so-called convexity effect leads, in fact, to a biased (inconsistent) estimation of some parameters of interest.

To apply indirect estimation, the *econometric* model(s) must be simulable. Equations (3.28) and (3.29) are not immediately suitable for this purpose, so also for simulation they are replaced by a discretized version. This, however, does not raise the same problems as discretization for the estimable model. Simulation in fact can be made with arbitrarily small time intervals δ_t, and regularity conditions can ensure that the discretized model, with a conveniently small δ_t, exhibits negligible differences from the corresponding continuous time model. For our purposes, a value $\delta_t \leq 0.1$ proved to be sufficiently accurate. Of course, the approximation can be made even more accurate with the implementation of higher order discretization schemes. An example is the explicit order 2 weak scheme of Gallant and Tauchen (1995). However, our experiments and findings from previous studies (see Bianchi, Cesari and Panattoni, 1995, Broze, Scaillet and Zakoïan, 1995, and Cleur, 1995) suggest that, for the models considered here, the first order scheme is acceptable, provided that δ_t is at most equal to 1/10 (we have assumed $\delta_t = 0.05$).

The discretized models used for simulation instead of (3.28) and (3.29) are, respectively

$$r_t - r_{t-\delta_t} = ka\delta_t - kr_{t-\delta_t}\delta_t + \sqrt{\delta_t \sigma^2}\; e_t \tag{3.31}$$

$$r_t - r_{t-\delta_t} = ka\delta_t - kr_{t-\delta_t}\delta_t + \sqrt{r_{t-\delta_t}}\;\sqrt{\delta_t \sigma^2}\; e_t \tag{3.32}$$

A time unit corresponds to the frequency of historical data. Thus if data are daily, t and $t-1$ refer to consecutive days, and $\delta_t = 0.05$ (or $1/\delta_t = 20$) means that 20 data are generated to produce one daily simulated value. So t and $t-\delta_t$ refer to consecutive generated data, while consecutive days will be t and $t-20\delta_t$.

4 The indirect inference procedure

The indirect inference procedure for the stochastic differential equations considered here takes advantage of the one-to-one correspondence between the θ and the β parameters. As already stated, it is a case of exact identification, where the results are unaffected by the choice of the matrix of weights usually involved in this type of estimator. Moreover, it is identical to the simulated QML (Smith, 1993).

We have implemented our procedure in the following steps.

1) The available series of observations for the dependent variable r_t, $t = 1, 2, ..., T$ is assumed to have been generated as in equation (3.31) (or 3.32). In applied work, this is the series of historical data. In Monte Carlo studies, this will be one replication of pseudo-historical data.

2) Naive estimation of the parameters $\hat{\beta} = (\hat{a}, \hat{k}, \hat{\sigma^2})'$ is obtained, for the auxiliary model (3.30).

3) A tentative value for the *true* model parameters $\theta = (a, k, \sigma^2)'$ is chosen. It is used as a starting point for the iterative calibration procedure.

4) A sample of pseudo-random error terms $\tilde{e}_t$ i.i.d. $N(0,1)$ is generated. In all our experiments we have adopted a sample length T, equal to the length of the observable time series. Since we use $\delta_t = 0.05$, $T/\delta_t = 20T$ values must be generated to produce a simulated series of T values. Possible lengths multiple of T (that is HT) can be adopted

5) The value of $\theta = (a, k, \sigma^2)'$ is plugged into equation (3.31) (or 3.32). The equation is solved recursively, then one value of r out of 20 ($= 1/\delta_t$) values is chosen. This produces the pseudo-random series $\tilde{r}_t$.

6) Naive estimation of equation (3.30) is performed on the series of pseudo-random $\tilde{r}_t$, obtaining a vector of parameters $\check{\beta}$.

7) The two vectors of parameters $\hat{\beta}$ and $\check{\beta}$ are compared. If they are equal (or very close to each other) the estimation procedure has come to its end, otherwise the tentative values of the parameters $\theta = (a, k, \sigma^2)'$ are modified (calibrated) and a new iteration of the procedure starts again from step 5. Notice that the pseudo-random errors $\tilde{e}_t$ generated at step 4 must not be re-generated, they must remain fixed in all iterations until convergence of the procedure. The values of the series $\tilde{r}_t$ change across iterations only as an effect of changing θ.

8) When convergence is achieved, the last value of the tentative parameters is the *simple indirect estimate* of the parameters of interest. This vector will be called $\hat{\theta} = (\hat{a}, \hat{k}, \hat{\sigma^2})'$.

4.1 The control variates

Various control variates could be introduced, with different choices of the control variate model (2.20). In this paper we use only one control variate model for the Ornstein-Uhlenbeck process (model 3.31), and only one control variate model for the square root process (model 3.32).

Both models obviously share all the features required for control variate models, as they were listed just after equation (2.20).

After the simple indirect estimator $\hat{\theta}$ has been computed (step 8 of the previous section), the last simulated series produced by equation (3.31) (or 3.32) is used to re-estimate the same equation (thus using the entire simulated series whose length is $T/\delta_t = 20T$). If $\tilde{\hat{\theta}}$ is the estimate, then $\tilde{\hat{\theta}} - \hat{\theta}$ is the new parameter estimation error, $\text{NPEER}(\hat{\theta}, \tilde{e})$, that produces the control variates. Thus, the indirect estimator with control variates will be

$$\hat{\theta}_{cv} = \hat{\theta} + R_0^{-1}\;\text{NPEER}(\hat{\theta}, \tilde{e}) = \hat{\theta} + R_0^{-1}(\tilde{\hat{\theta}} - \hat{\theta}) \tag{4.33}$$

4.2 Some computational aspects

Concerning the solution of the implicit system of equations (which are not written in closed form) that yields the indirect estimator, little is reported in the literature. Since an analytic solution does not exist, the problem must be solved numerically. We have adopted the following updating equation

$$\tilde{\theta}_{(j)} = \tilde{\theta}_{(j-1)} + \lambda A^{-1}_{(j-1)}(\tilde{\beta}_{(j-1)} - \hat{\beta})$$

where $\tilde{\theta}_{(j)}$ is the value of the calibrated parameters after j iterations, $A_{(j-1)}$ is a matrix that determines the direction of the jth step, and λ is a real number (scalar) which determines the stepsize in the given direction. In our applications we first perform a few "simple" iterations (typically four to five) taking A equal to the identity matrix, and then we switch to one "complicated" iteration taking A equal to the Jacobian matrix of derivatives of the auxiliary parameters with respect to the parameters of interest (an approximation of matrix R_0 of equation (2.13)). Thus we alternate iterations of a Jacobi solution method with iterations of a Newton solution method of the system (2.7).

This heuristic switching rule is maintained until convergence is reached. A reasonable starting point for the procedure can be $\tilde{\theta}_{(1)} = \hat{\beta}$.

5 The Monte Carlo experiment

The design of the Monte Carlo is as follows. The *econometric* structural models and their *true* parameters are kept fixed in all experiments. Specifically, they are, respectively

$$dr_t = 0.5\,(\,0.1 - r_t\,)dt + 0.1\,dW_t \tag{5.34}$$

$$dr_t = 0.5\,(\,0.1 - r_t\,)dt + 0.1\,\sqrt{r_t}\,dW_t \tag{5.35}$$

These model specifications imply a moderate mean reversion effect (the mean-lag and the half-life are both equal to 1) toward an average level of 10% for the annualized spot interest rate.

While a more complete and detailed set of experiments is still in progress, the results presented in this paper are related to a sample period length $T = 4000$. This also is the length of the simulated series ($H = 1, HT = T$); this allows us to appreciate the benefits due to the use of the control variates. The simulation step is kept fixed ($\delta_t = 0.05$, thus $T/\delta_t = 80000$).

Table 1 is related to the Ornstein-Uhlenbeck process (O-U), used in Vasicek (1977) to model the short term interest rate. Table 2 is related to the square root process, adopted in the Cox, Ingersoll and Ross (1985) model.

Table 1: O-U – mean estim. param. and (var.), [Monte-Carlo var.]

Par.	*True*	Least.Sqr.	Ind. Inf H=1	Ind. Inf Cnt. Var.
a	.1	.1001 ($10^{*}10^{-4}$) [$.11^{*}10^{-4}$]	.1001 ($.20^{*}10^{-4}$) [$.20^{*}10^{-4}$]	.1001 [$.11^{*}10^{-4}$]
k	.5	.3978 ($.16^{*}10^{-3}$) [$.14^{*}10^{-3}$]	.5012 ($.81^{*}10^{-3}$) [$.84^{*}10^{-3}$]	.5019 [$.53^{*}10^{-3}$]
σ^2	.01	.0064 ($.19^{*}10^{-7}$) [$.29^{*}10^{-7}$]	.0100 ($.15^{*}10^{-6}$) [$.15^{*}10^{-6}$]	.0100 [$.14^{*}10^{-6}$]
T=4000				Replications= 10000

Each row contains the following values:

1) The *true* value of a parameter (used in all Monte Carlo replications to generate the pseudo-historical data).

2) The Monte Carlo mean of the naive estimates of the parameter in the discretized model, computed across 10000 replications.

3) The Monte Carlo mean of the simple indirect estimates of the parameter.

4) The Monte Carlo mean of the control variate indirect estimates.

Under each mean, in square brackets, we display the Monte Carlo variance of the parameter, computed across the 10000 replications. In parentheses, we display the mean of the estimated variance of the parameter, computed across the same replications.

The variance of the naive estimator is computed in a straightforward way.

The estimated variance of the simple indirect estimator is computed as in Gourieroux, Monfort and Renault (1993, section 3).

For the indirect estimator with control variates, only the Monte Carlo variance is displayed in square brackets, as an estimator of the variance is not available.

Table 2: Sq.Rt.– mean estim. param. and (var.), [Monte-Carlo var.]

Par.	*True*	Least.Sqr.	Ind. Inf H=1	Ind. Inf Cnt. Var.
a	.1	.1000 ($10^{*}10^{-5}$) [$.10^{*}10^{-5}$]	.1000 ($.19^{*}10^{-5}$) [$.20^{*}10^{-5}$]	.1000 [$.11^{*}10^{-5}$]
k	.5	.3979 ($.15^{*}10^{-3}$) [$.17^{*}10^{-3}$]	.5015 ($.84^{*}10^{-3}$) [$.89^{*}10^{-3}$]	.5021 [$.57^{*}10^{-3}$]
σ^2	.01	.0066 ($.24^{*}10^{-7}$) [$.24^{*}10^{-7}$]	.0100 ($.16^{*}10^{-6}$) [$.16^{*}10^{-6}$]	.0100 [$.15^{*}10^{-6}$]
T=4000				Replications= 10000

For the simple indirect estimator, the mean of the estimated variances is remarkably close to the Monte Carlo variance.

For the naive estimator, the bias (inconsistency) is quite evident for the parameters k and σ^2.

Indirect inference (with or without control variates) adjusts for the bias (inconsistency): the mean estimated parameter is practically equal to the *true* value.

Control variates produce a remarkable reduction of the variance of some parameter estimates, with respect to simple indirect estimates. For the parameter a, the variance is nearly the same as for the naive estimator, while for the simple indirect estimator it is much larger. This, however, is of no interest given that this naive direct estimator is practically unbiased (consistent) for parameter a.

For the parameter σ^2, the reduction of variance over the simple indirect estimator is negligible. Use of these control variates is therefore not interesting for the estimation of parameter σ^2. Other control variates should be searched for, but this will not be attempted in this paper.

The great benefit produced by the use of control variates is evident for the parameter k. We see in fact for the Ornstein-Uhlenbeck process (Table 1) that the Monte Carlo variance of the simple indirect estimate of this parameter is 0.84×10^{-3}; that reduces to 0.53×10^{-3} if control variates are used. Thus, there is a reduction of almost 40% in the global variance of the estimator; that means an efficiency gain of 40% over the simple indirect estimator, at about the same computation cost.

We may think of the variance 0.84×10^{-3} as being composed of two equal parts (as it comes from equation 2.14): 0.42×10^{-3} (the irreducible component of the variance) and 0.42×10^{-3} (the component of the variance due to simulation). The variance with control variates is 0.53×10^{-3}; that we may attribute to the irreducible component (the same as before, therefore, 0.42×10^{-3}) and to the component due to simulation (thus 0.11×10^{-3}). Therefore, the component of the variance due to simulation is about 4 times smaller than for the simple indirect estimator. The same variance would be obtained by the simple indirect estimator using $H = 4$, thus at a considerably higher computational cost.

The results for the square root models (Table 2) are quite similar, and the same comments apply.

6 Conclusion

We have shown in this paper why and how control variates can help in improving the efficiency of indirect estimators. 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 computation cost (that is, computation time), an indirect estimator with control

variates can be as efficient as a simple indirect estimator which is based on much longer simulated series and, therefore, involves a much higher cost.

Different control variates could be introduced and adopted. Some simple 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 effective for some parameters of interest. We will need study other types of control variates, and a more complete and detailed set of experimental results should be available in the near future.

References

Arnold, L. (1973): *Stochastic Differential Equations*. New York: John Wiley & Sons, Inc.

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, forthcoming): "Indirect Estimation of Stochastic Differential Equation Models: Some Computational Experiments," *Computational Statistics*.

Brennan, M. J. and E. S. Schwartz (1979): "A Continuous Time Approach to the Pricing of Bonds," *Journal of Banking and Finance* 3, 135-155.

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

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.

Chan K.C., G.A. Karolyi, F.A. Longstaff and A.B. Sanders (1992): "An Empirical Comparison of Alternative Models of the Short-Term Interest Rate," *The Journal of Finance* 47, 1209-1227.

Cleur E. M. (1995): "A Comparison of Alternative Discrete Approximations to the Cox Ingersoll Ross Model." Università di Pisa, Department of Statistics, mimeo.

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

Di Iorio, F. (1996): "Metodo Generalizzato dei Momenti e Modelli Markoviani per la Valutazione delle Attività Finanziarie." Università di Firenze, Department of Statistics, *PhD Thesis*.

Gallant, A. R. and G. Tauchen (1992): "Which Moment to Match?" Duke University, Department of Economics, mimeo.

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

Geweke, J. (1994): "Simulation and Numerical Integration." Federal Reserve Bank of Minneapolis, Working Paper No. 526.

Gourieroux, 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.

Ikeda, N. and S. Watanabe (1989): *Stochastic Differential Equations and Diffusion Processes*. Amsterdam: North-Holland Publishing Company.

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.

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 Laszlo Matya and Patrick 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.