Dipartimento di Statistica
"Giuseppe Parenti"

Dipartimento di Statistica "G. Parenti" – Viale Morgagni 59 – 50134 Firenze - www.ds.unifi.it

QUADERNI 1999/46

Indirect Estimation of Just-Identified Models with Control Variates

G. Calzolari, F. Di Iorio,
G. Fiorentini

Università degli Studi
di Firenze

Indirect Estimation of Just-Identified Models with Control Variates*

Giorgio Calzolari[§] Francesca Di Iorio[¶] Gabriele Fiorentini[‖]
Università di Firenze Istat, Roma Universidad de Alicante

October 1999

Keywords: Efficient Monte Carlo, Variance reduction techniques, Control variates, Indirect inference, Arma models, Stochastic volatility, Stochastic differential equations, Logit model with random effects.

---

*This paper extends and generalizes results that, for some of the models discussed in section 5, were presented at the *European Meetings of the Econometric Society* (Toulouse, 1997 and Berlin, 1998), and at the seminars in the Universities of Alicante, Dortmund, Firenze, Udine, Wyoming (Laramie), at Cemfi (Madrid), Ceibs (Shanghai), Stockholm School of Economics, Rwi (Essen), and at the European University Institute (Firenze).

We have benefited from discussions with Manuel Arellano, Marco Barnabani, Patrizia Berti, Carlo Bianchi, Monica Billio, Fabrizio Cipollini, Eugene M. Cleur, Giampiero M. Gallo, Christian Gouriéroux, Fabrizia Mealli, Grayham E. Mizon, Chiara Monfardini, Jaime Mora, Carla Rampichini, Eric Renault, Pietro Rigo, Enrique Sentana, Elena Stanghellini, as well as participants at conferences and seminars, but retain full responsibility for the contents of this paper.

Financial support from Cicyt, Ivie, and Murst project "Modelli Statistici per l'Analisi delle Serie Temporali" is gratefully acknowledged.

[§]Università di Firenze, Dipartimento di Statistica "G. Parenti", Viale Morgagni 59, I-50134 Firenze, Italy. *Fax:* (+39) 055 4223560. *E-mail:* calzolar@ds.unifi.it.

[¶]Istat, Istituto Nazionale di Statistica, Via C. Balbo 16, I-00184 Roma, Italy. *Fax:* (+39) 06 47823304. *E-mail:* diiorio@istat.it

[‖]Universidad de Alicante, Departamento de Fundamentos del Analisis Economico, Apdo 99, E-03080 Alicante, Spain. *Fax:* (+34) 96 5903614. *E-mail:* gafi@merlin.fae.ua.es.

ABSTRACT

Simulation estimators, such as indirect inference or simulated maximum likelihood, are successfully employed for estimating models where the likelihood function does not have a simple analytical expression. They adjust for the bias (inconsistency) produced by the estimation of an *auxiliary* model that can be manageable, but is essentially misspecified. The price to be paid is an increased variance of the estimated parameters. A component of the variance 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 computational cost. Alternatively, this paper shows how variance reduction can be achieved, at virtually no additional computational cost, by use of control variates. This technique can be easily applied in the just-identified context, that is when the number of parameters is the same in the *econometric* model (the model of interest) and the *auxiliary* model. This is a case which often occurs in practical applications. Several models are explicitly considered and experimented with: moving average model, ARMA model, stochastic differential equations, dynamic Tobit model, discrete time stochastic volatility models, logit models with random effects. Monte Carlo experiments show, in some cases, a global efficiency gain up to almost 50% over the simplest indirect estimator, obtained at about the same computational cost.

1 Introduction

A wide part of the modern econometric models is associated with non trivial computational problems. Classical inference procedures, typically based on likelihood function, often lead to intractable numerical problems. To provide nevertheless a solution, several alternative procedures have been proposed, as the Pseudo-Maximum-Likelihood or the Generalized Method of Moments. In recent years, since the contribution of McFadden (1989), great interest has been shown in simulation-based procedures.

An interesting class of simulation-based estimation procedures includes the Simulated Maximum Likelihood (Smith, 1993), the Efficient Method of Moments (Gallant and Tauchen, 1996), the Indirect Inference (Gourieroux, Monfort and Renault, 1993). The basic idea of these techniques is to adjust the parameters of interest by a calibration procedure, in order to get similar characteristics for the observed endogenous variables and for the simulated ones.

There are many econometric fields where these methods can be usefully employed: continuous time finance (Gourieroux, Monfort and Renault, 1993; Broze, Scaillet and Zakoïan, 1995, 1998; Pastorello, Renault and Touzi, 1994; Bianchi, Cesari and Panattoni, 1994; Bianchi and Cleur 1996; Di Iorio and Fiorentini 1996; Calzolari, Di Iorio and Fiorentini 1998); limited dependent variables with non-spherical disturbances (Calzolari and Fiorentini, 1996; Gourieroux and Monfort, 1996; Mealli and Rampichini, 1999); latent variables dynamic models (Gourieroux, Monfort and Renault, 1993 ; Billio, Monfort and Robert, 1998); switching regimes models (for a survey see Gourieroux and Monfort, 1996).

Some aspects must be taken into account when indirect inference methods are used. The first aspect is the choice of a favourable calibration criterion in order to obtain consistent estimates. In fact, even if from the methodological point of view there are not many restrictions in the choice of the calibration criterion, the performance of indirect estimators is not independent of the calibration criterion chosen (see Di Iorio and Fiorentini, 1996; Monfardini 1998; Calzolari, Di Iorio and Fiorentini, 1998; Mealli and Rampichini, 1999).

The second aspect is an increased variance of the estimated parameters involved, in general, in all simulation-based methods. On the one hand, in fact, the variance is due to the intrinsic stochastic nature of the data and to the approximations 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 computational effort. Therefore a trade-off arises between variance reduction and computational cost.

The first goal of this paper is to provide a simple description of the indirect inference method through several examples (that can be used for teaching purposes), and to develop a simplified derivation of the asymptotic distribution of the estimators.

A second important goal is the implementation of an efficient software package.

The third and most important goal is to present a useful procedure to reduce the variance of the estimated parameters of just-identified models.

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, 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), Richard (1996), and others.

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).

An earlier paper of the authors (1998) applied a control variate technique to the indirect estimation of stochastic differential equations. Here, we develop the theoretical approach to cover in a more general way the just-identified case; that is when the number of parameters is the same for the model of interest and for the *auxiliary* model, even when no clear one-to-one correspondence between them can be established. Although not completely general, this is a case that occurs very often in practical applications.

Several models are explicitly considered and experimented with:

1. MA(1) model, with a simple AR(1) model used as *auxiliary* model;
2. ARMA(1,1) model, with an AR(2) *auxiliary* model;
3. the Ornstein-Uhlenbeck stochastic differential equation, used in Vasicek (1977) to model the short term interest rate in continuous time, with an AR(1) model on discrete data used as *auxiliary* model;
4. the *square root* stochastic differential equation, used in Cox, Ingersoll and Ross (1985), with an AR(1) *auxiliary* model applied to discrete data (after some transformation);
5. the stochastic differential equation *with unrestricted variance elasticity*, used in Chan et alii (1992), Di Iorio and Fiorentini (1996), and Nowman (1997), with a

nonlinear *auxiliary* model (applied to discrete data) that can be estimated by maximum likelihood iterating an instrumental variables procedure;

6. a dynamic Tobit model, with a simple AR(1) *auxiliary* model;

7. a discrete time stochastic volatility model, with two AR(10)-distributed-lag *auxiliary* models;

8. a two level logit model with random effects, with a linear probability *auxiliary* model.

The performance of the indirect estimator and the control variate procedures are verified by a set of Monte Carlo experiments.

Results of the Monte Carlo experiments show that, for some parameters of interest, the variance component due to simulation can be reduced from 3 up to 20 times at about the same computational cost. This implies a global efficiency gain up to almost 50% over the simplest indirect estimator (reduction of the global variance), at about the same computational cost.

For case 3 and 4, the control variates are the same already used in Calzolari, Di Iorio and Fiorentini (1998): we present here results related to a different choice of the parameter values. Moreover, for most of the models, we experiment also an estimator for the variance-covariance matrix.

The paper is organized as follows. In section two we present the indirect inference computational procedure. In section three we present the procedures to reduce the variance proposed by Gourieroux, Monfort and Renault (1993), based on replicated simulations. In section four we describe the Indirect inference procedure with control variates. In section five we report the results of the Monte Carlo experiments conducted for the cases described above. Conclusions and final remarks are in sections six and seven.

2 Indirect Inference

Indirect inference methods were first proposed by Gourieroux, Monfort and Renault (1993) and Gallant and Tauchen (1996).

The basic idea can be summarised as follows: suppose an *econometric* model be defined by:

$$y_t = f(y_{t-1}, x_t, e_t, \theta) \tag{2.1}$$

where f is a known function, x_t are exogenous variables, $\theta \in \Theta$ is the vector of parameters of interest, e_t are disturbances with known probability function, y_t are endogenous variables and y_{t-1} are lagged endogenous variables. We assume that this model can be

simulated; that is we can produce values of y_t conditional on x_t, e_t, θ and initial values. However, this model may not be estimated, or estimation can be so complex and discouraging that econometricians replace it with an approximation, like

$$y_t = g(y_{t-1}, x_t, \beta, \eta) \tag{2.2}$$

where g is a function, $\beta \in B$ is a parameter vector, η are random terms. Assume that parameters β can be easily estimated (for example, by least squares, or maximum likelihood). This will be called the *auxiliary* model.

A simple estimation of the *auxiliary* model, drawn on the observed variables y_t, leads to biased (inconsistent) estimates $\hat{\beta}$. Similar estimates, noted as $\tilde{\beta}$, can be obtained using simulated values of the endogenous variables, conditional on x_t, e_t and θ (let these simulated values be $\tilde{y}_t(\theta)$, $t = 1, 2, ...T$, and $\tilde{\beta}(\theta)$ the estimated parameters). If $\hat{\beta}$ and $\tilde{\beta}$ are not too far in some sense, we can assume that the θ values, that have produced the simulated endogenous $\tilde{y}_t$, are good estimates of the parameter of interest.

The indirect inference estimate of θ proposed by Gourieroux, Monfort and Renault (1993) is

$$\hat{\theta} = argmin_\theta \; [\hat{\beta} - \tilde{\beta}(\theta)]' {\Omega_1}^{-1} [\hat{\beta} - \tilde{\beta}(\theta)] \tag{2.3}$$

where Ω_1 is a positive definite matrix.

Gallant and Tauchen (1996) proposed a modified version of (2.3) defined by

$$\hat{\theta} = argmin_\theta \; \frac{\partial L}{\partial \beta'} \left[\tilde{y}(\theta); \hat{\beta}\right] {\Omega_2}^{-1} \frac{\partial L}{\partial \beta} \left[\tilde{y}(\theta); \hat{\beta}\right] \tag{2.4}$$

where Ω_2 is a positive definite matrix and $\frac{\partial L}{\partial \beta}(\tilde{y}(\theta); \hat{\beta})$ is the score function of the *auxiliary* model calculated on the simulated values $\tilde{y}(\theta)$.

Estimators (2.3) and (2.4), associated with optimal choice of the weighting matrices Ω_1 and Ω_2, have the same asymptotic efficiency (see, for example, Gourieroux and Monfort, 1996).

2.1 Simple indirect estimation of just-identified models

The *econometric* model and the *auxiliary* one can have a different number of parameters. When the number of parameters is the same, the indirect estimation procedure can take advantage of the equal dimension of the θ and the β parameters. This is a case of exact identification, and the results are unaffected by the choice of the matrix of weights usually involved in this type of estimator (moreover, they would be identical to those of

the simulated ML, Smith, 1993). Minimization of the quadratic forms (2.3) and (2.4) is in fact simply obtained when $\tilde{\beta} = \hat{\beta}$.

For simplicity, we omit from our notation exogenous variables, 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). So we represent the *econometric* model (or model of interest) as

$$y = f(\theta, e) \tag{2.5}$$

while the *auxiliary* model can be written as

$$y = g(\beta, \eta) \tag{2.6}$$

We assume that the *econometric* model can be easily simulated, that is we can produce values of y conditional on θ, by entering random values of e.

The procedure for the indirect inference can be implemented in the following steps.

1) The available series of observations for the dependent variable y_t, $t = 1, 2, ..., T$, is assumed to have been generated as in equation (2.5). In applied work, this is the series of observed data. In Monte Carlo studies, this will be really generated by simulation of (2.5), with *known true* parameters θ_0; we are so building one replication of pseudo-observed data.

2) *Naive* estimation (e.g. least squares, maximum likelihood, etc.) of the parameters $\hat{\beta}$ is obtained, using the observed data in the *auxiliary* model (2.6).

3) A tentative value for the *true* model parameters is chosen. We call it $\tilde{\theta}$. 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. Possible lengths multiple of T (that is HT) can be adopted (see section 3).

5) The $\tilde{e}_t$'s and $\tilde{\theta}$ are plugged into equation (2.5). The equation is solved to produce the pseudo-random series $\tilde{y}_t$.

6) *Naive* estimation of equation (2.6) is performed on the series of pseudo-random $\tilde{y}_t$, obtaining a vector of parameters $\tilde{\beta}$.

7) The two vectors of parameters $\hat{\beta}$ and $\tilde{\beta}$ 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 $\tilde{\theta}$ 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 $\tilde{y}_t$ change across iterations only as an effect of changing $\tilde{\theta}$.

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 $\hat{\theta}$.

2.2 The parameter estimation error

In order to study the properties of the simple indirect estimator, we need a more precise notation for the estimator of the β parameters of the *auxiliary* model (2.6). The y_t data are produced by simulation of the *econometric* model (2.5) (or supposed to have been generated by 2.5). So we can use the notation $\beta(\theta, e)$ for the estimator of the parameters of (2.6) conducted on y_t values generated by (2.5) for any $\theta \in \Theta$ and for a given probability distribution for e. 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 (2.6) 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 most of our experiments, a uniform distribution in (0,1) in some cases), as the distributional parameters (i.e. the variance) are already included in the parameter vector θ

$$\beta(\theta, e) = b(\theta) + \text{ PEER}(\theta, e) \tag{2.7}$$

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

$$\sqrt{T} \text{ PEER}(\theta, e) \longrightarrow N(0, \Sigma(\theta)) \tag{2.8}$$

Since the covariance matrix is obtained from the *auxiliary* model (2.6), which is misspecified, it is well known from White (1982) and Newey and West (1987) that an expression for $\Sigma(\theta)$ would involve both the Hessian and the matrix (or matrices) of outer products of the first derivatives of the log-likelihoods (that should be more properly called *quasi*-log-likelihoods, being the model misspecified)

$$\Sigma(\theta) = H_1^{-1} J_1 H_1^{-1} \tag{2.9}$$

In this equation, H_1 and J_1 are the probability limits of the *auxiliary* model's Hessian and outer-product matrices, computed at $b(\theta)$, when the values of y are produced by the *econometric* model (2.5).

Of course, an estimate of these matrices can be obtained performing the computation of derivatives in the available estimate of the *auxiliary* model's parameters. In particular, the estimate of J_1 involves the computation of products and cross products of the first derivatives of the log-likelihoods at different time lags, when data are autocorrelated (Newey and West, 1987).

Let's now consider the actually observed values of y. If the model of interest (2.5) really is the data generating process, the actually observed y are a function of the *true* vector of parameters, say θ_0, as well as of the unobservable error terms, say e_0. Therefore, the estimate of the β parameters (called $\hat{\beta}$ in the previous section) turns out to be a function of θ_0 and of the unobservable error terms e_0

$$\beta(\theta_0, e_0) = b(\theta_0) + \text{ PEER}(\theta_0, e_0) \tag{2.10}$$

By entering a tentative vector of parameters θ, say $\tilde{\theta}$, and pseudo-random error terms $\tilde{e}$ into the *econometric* model (2.5), we generate by simulation pseudo-random values $\tilde{y}$ that are introduced into the *auxiliary* model (2.6). The *auxiliary* model (2.6) is estimated, obtaining a vector of parameters $\beta(\tilde{\theta}, \tilde{e})$. The sample size can be, of course, of any length, being data produced by simulation (at least, in absence of exogenous variables), but we keep for the moment the same sample length as for the actually observed data, say T. Notice that $\tilde{e}$ are generated from "the same" distribution as the unobservable actual error terms e_0. Thus

$$\beta(\tilde{\theta}, \tilde{e}) = b(\tilde{\theta}) + \text{ PEER}(\tilde{\theta}, \tilde{e}) \tag{2.11}$$

We can now take advantage of the exact identification, that is when θ and β have the same number of parameters. We calibrate the $\tilde{\theta}$ parameters (keeping $\tilde{e}$ fixed) till we find the same vector of estimated β from both the simulated $\tilde{y}$ and the actually observed y. In other words we look for the values of $\tilde{\theta}$ that solve the system of equations

$$\beta(\tilde{\theta}, \tilde{e}) = \beta(\theta_0, e_0) \tag{2.12}$$

The calibration procedure thus aims at solving the system of equations (2.12). These equations are only implicitly defined, and usually cannot be expressed in closed form. It is usually possible to solve the system only in the just-identified case, because there is the same number of unknowns ($\tilde{\theta}$) as of equations (dimension of β).

The solution vector will be called $\hat{\theta}$; this is the *simple* indirect estimator of the *econometric* model's parameter vector θ_0. Thus, at solution, we have

$$\beta(\hat{\theta}, \tilde{e}) = \beta(\theta_0, e_0) \tag{2.13}$$

and, from eq. (2.11)

$$\beta(\hat{\theta}, \tilde{e}) = b(\hat{\theta}) + \ \textsc{peer}(\hat{\theta}, \tilde{e}) \tag{2.14}$$

Conditions that ensure consistency and asymptotic normality of this estimator can be found in Gourieroux, Monfort and Renault (1993) in the general context of overidentified models (in our context, if estimation of the *auxiliary* model (2.6) is performed by quasi-maximum-likelihood, the estimator turns out to be identical to the simulated ML, Smith, 1993, and to the efficient method of moments, Gallant and Tauchen, 1996).

2.3 The variance of the simple indirect estimator

We now observe that the left hand sides of equation (2.10) and (2.14) are equal (eq. 2.13); thus equating the right hand sides and multiplying by $\sqrt{T}$ we get

$$\sqrt{T} b(\theta_0) + \sqrt{T}\ \textsc{peer}(\theta_0, e_0) = \sqrt{T} b(\hat{\theta}) + \sqrt{T}\ \textsc{peer}(\hat{\theta}, \tilde{e}) \tag{2.15}$$

and therefore

$$\sqrt{T}\left[b(\hat{\theta}) - b(\theta_0)\right] = \sqrt{T}\ \textsc{peer}(\theta_0, e_0) - \sqrt{T}\ \textsc{peer}(\hat{\theta}, \tilde{e}) \tag{2.16}$$

As $\hat{\theta}$ converges to θ_0 (for $T \to \infty$), regularity conditions ensure that, asymptotically, we can replace the random vector $\sqrt{T}\ \textsc{peer}(\hat{\theta}, \tilde{e})$ with $\sqrt{T}\ \textsc{peer}(\theta_0, \tilde{e})$. Thus, asymptotically

$$\sqrt{T}\left[b(\hat{\theta}) - b(\theta_0)\right] \simeq \sqrt{T}\ \textsc{peer}(\theta_0, e_0) - \sqrt{T}\ \textsc{peer}(\theta_0, \tilde{e}) \tag{2.17}$$

The random error terms e_0 and $\tilde{e}$ are obviously independent, as the former are the unobservable errors in the process that produced the actually observed data, while the latter are generated by simulation. Thus variances must be summed. As the distribution of e_0 and $\tilde{e}$ is the same by assumption (i.i.d. N(0,1), in most of our experiments), the variance turns out to be simply double. Thus the right hand side of (2.17) will be $N(0, 2\Sigma(\theta_0))$, asymptotically.

Applying the "δ-method" (e.g. Rao, 1973, p.388) the left hand side of (2.16) is asymptotically

$$\sqrt{T}\left[b(\hat{\theta}) - b(\theta_0)\right] \ \simeq \ R_0\sqrt{T}(\hat{\theta} - \theta_0) \qquad\qquad R_0 = \left[\frac{\partial b(\theta)}{\partial \theta'}\right]_{\theta_0} \tag{2.18}$$

In our just-identification context there is the same number of 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

$$\sqrt{T}(\hat{\theta} - \theta_0) \quad \simeq \quad {R_0}^{-1}\sqrt{T}\ \text{PEER}(\theta_0, e_0) - {R_0}^{-1}\sqrt{T}\ \text{PEER}(\theta_0, \tilde{e}) \tag{2.19}$$

Thus, asymptotically the variance-covariance matrix of the simple indirect estimator is equal to

$$Var\left[\sqrt{T}(\hat{\theta} - \theta_0)\right] \ = \ {R_0}^{-1} 2\Sigma(\theta_0) {R_0'}^{-1} \tag{2.20}$$

The fact that $\Sigma(\theta_0)$ is doubled is clearly due to the independence between e_0 and $\tilde{e}$ in equation (2.19).

Somewhere in the following sections, we do not need the full notation $\beta(\theta_0, e_0)$ for the estimator of the *auxiliary* model parameters using the actually observed data. So, when the full notation is not strictly necessary, we shall resort to the simplified notation $\hat{\beta}$ (already used in sections 2 and 2.1), where

$$\hat{\beta} = \beta(\theta_0, e_0) \tag{2.21}$$

3 Reducing variance with replicated simulations

It is well known that the variance of indirect estimators can be reduced, with a larger computational 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 $\tilde{\theta}$ that solves the system

$$\beta(\tilde{\theta}, \tilde{e}_h) = \beta(\theta_0, e_0) \tag{3.22}$$

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

$$\beta(\theta_0, e_0) = b(\theta_0) + \ \text{PEER}(\theta_0, e_0) \tag{3.23}$$

$$\beta(\hat{\theta}_h, \tilde{e}_h) = b(\hat{\theta}_h) + \ \text{PEER}(\hat{\theta}_h, \tilde{e}_h) \qquad\qquad h = 1, 2, ..., H \tag{3.24}$$

where each $\hat{\theta}_h$ is calibrated till $\beta(\hat{\theta}_h, \tilde{e}_h) = \beta(\theta_0, e_0)$, that is till the left hand sides of (3.23) and (3.24) are equal.

Finally all $\hat{\theta}_h$ are averaged to produce

$$\hat{\theta} = \frac{1}{H} \sum_{h=1}^{H} \hat{\theta}_h \tag{3.25}$$

Instead of equation (2.19), we have in this case, asymptotically

$$\sqrt{T}(\hat{\theta} - \theta_0) \quad \simeq \quad {R_0}^{-1} \left[\sqrt{T} \ \text{PEER}(\theta_0, e_0) - \frac{1}{H} \sum_{h=1}^{H} \sqrt{T} \ \text{PEER}(\theta_0, \tilde{e}_h) \right] \tag{3.26}$$

where the asymptotic variance-covariance matrix of the term in square brackets is now $(1 + 1/H)\Sigma(\theta_0)$, being the $\tilde{e}_h$ independent of each other and of e_0.

The variance reduction corresponding to a multiplying factor $(1 + 1/H)$ instead of 2 is obtained at the cost of H calibration procedures instead of just one.

In absence of exogenous variables, the same result would be obtained if the H procedures with T data were replaced by one procedure with HT simulated data (Gourieroux, Monfort and Renault, 1993, section 2.3). The computational cost would be about the same, if the estimator of the *auxiliary* model has a closed form expression, and therefore the number of elementary arithmetic operations is roughly proportional to HT (as in the models considered in Calzolari, Di Iorio and Fiorentini, 1998). Otherwise, the computational cost of a single estimation with HT data would be usually lower than replicating H estimations with T data, when the *auxiliary* model needs an iterative estimation procedure (the number of iterations to convergence usually decreases as the sample length increases).

Equations like (2.19) or (3.26) quite clearly put into 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_0 and R_0. Thus, it is irreducible, given the data, the *auxiliary* model adopted and the estimation method used for the *auxiliary* model. The second component on the right hand sides of (2.19) and (3.26) depends entirely on simulation, and can be made arbitrarily small, at the cost of a large simulation effort.

4 Reducing variance with control variates

In a recent paper, the authors (1998) proposed a control variate method capable of reducing the variance with (almost) no additional computational cost. They applied the method to two financial models, based on stochastic differential equations. In this paper, we apply a similar technique to a wider class of models.

The control variate procedure is based on the simulation-estimation of the *auxiliary* model (2.6). It aims at producing an estimation error that *should be strongly* correlated with $\sqrt{T}$ PEER$(\hat{\theta}, \tilde{e})$ of equation (2.16).

We need an additional assumption, beyond those required by simple indirect estimation.

We assume that the *auxiliary* model (2.6) can be simulated using the same pseudo-random errors $\tilde{e}$ used to simulate the *econometric* model (2.5). If it is not possible to use directly $\tilde{e}$ inside model (2.6), 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 the *auxiliary* model (2.6) 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 $\dot{\beta}$, without the need of further assumptions (there is no need here of introducing a binding function, as it is simply the identity function)

$$\tilde{\dot{\beta}} = \dot{\beta} + \ \textsc{npeer}(\dot{\beta}, \eta(\tilde{e})) \tag{4.27}$$

The *new* parameter estimation error (NPEER) is such that $\sqrt{T}\ \textsc{npeer}(\dot{\beta}, \eta(\tilde{e}))$ 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. We must note that NPEER is the estimation error of a *correctly specified* model, thus its variance, asymptotically, is simply obtained from the Hessian of the *auxiliary* model, when the values of y are produced by the *auxiliary* model itself. We call this matrix H_2 to distinguish it from the H_1 of equation (2.9).

To construct the control variate estimator we start by taking $\dot{\beta} = \hat{\beta} = \beta(\theta_0, e_0)$, the actual estimate. We enter it into the *auxiliary* model (2.6), as well as $\eta(\tilde{e})$, and with the series y produced by simulation of (2.6) we re-estimate the β parameters and obtain $\tilde{\hat{\beta}}$. Thus, using equation (4.27) with (2.13), (2.14) and (2.21), the vector NPEER can be written as

$$\tilde{\hat{\beta}} - \hat{\beta} = \ \textsc{npeer}(\hat{\beta}, \eta(\tilde{e})) = \ \textsc{npeer}\left\{\left[b(\hat{\theta}) + \ \textsc{peer}(\hat{\theta}, \tilde{e})\right], \eta(\tilde{e})\right\} \tag{4.28}$$

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

$$\hat{\theta}_{cv} = \hat{\theta} + \hat{R}^{-1}(\tilde{\hat{\beta}} - \hat{\beta}) = \hat{\theta} + \hat{R}^{-1}\ \textsc{npeer}(\hat{\beta}, \eta(\tilde{e})) \tag{4.29}$$

where $\hat{\theta}_{cv}$ is called the *control variate* estimator of θ_0 and the Jacobian $\hat{R}$ is a *feasible* estimator of $\partial b(\theta)/\partial\theta'$, that is

$$\hat{R} = \left[\frac{\partial\beta(\theta, \tilde{e})}{\partial\theta'}\right]_{\hat{\theta}} \tag{4.30}$$

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

$$plim\ \hat{R} = R_0 = \left[\frac{\partial b(\theta)}{\partial \theta'}\right]_{\theta_0} \tag{4.31}$$

and that, asymptotically

$$\sqrt{T}\ \text{NPEER}(\hat{\beta}, \eta(\tilde{e})) \simeq \sqrt{T}\ \text{NPEER}[b(\theta_0), \eta(\tilde{e})] \tag{4.32}$$

Thus, from (2.19) and (4.29) we get, asymptotically

$$\begin{aligned} \sqrt{T}(\hat{\theta}_{cv} - \theta_0) &= \sqrt{T}\left[(\hat{\theta} - \theta_0) + (\hat{\theta}_{cv} - \hat{\theta})\right] \qquad (4.33) \\ \simeq\ & {R_0}^{-1}\sqrt{T}\ \text{PEER}(\theta_0, e_0) + {R_0}^{-1}\left\{\sqrt{T}\ \text{NPEER}[b(\theta_0), \eta(\tilde{e})] - \sqrt{T}\ \text{PEER}(\theta_0, \tilde{e})\right\} \end{aligned}$$

Given the independence between e_0 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 (4.33) is exactly the same as the irreducible part in the simple indirect estimator, as in equation (2.19). The second component in equation (4.33), in braces, is the difference between two random vectors *presumably* quite close to each other (strong positive correlation and similar variance-covariance matrix). 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 (2.19).

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 additional software code required to implement the control variate procedure is very little: just the simulation of the *auxiliary* model, being any other algorithm already required by the simple indirect estimator (estimation, calibration, etc.). 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 (3.26).

4.1 A clarifying example

A simple example might be helpful to fix ideas. For this example, estimation can be performed with the standard econometric methods, so that indirect inference would obviously be unnecessary. Let us suppose one is dealing with a linear regression model with nonrandom exogenous regressors, under standard textbook conditions

$$y = X\theta + e \tag{4.34}$$

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

The *econometric* and *auxiliary* models are coincident, so there is no need of distinguishing the η from the e error terms. Using OLS we get

$$\beta(\theta, e) = (X'X)^{-1}X'y = \theta + (X'X)^{-1}X'e \tag{4.35}$$

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 $\sqrt{T}\ \text{PEER}(\theta, e) \rightarrow N[0, lim(X'X/T)^{-1}]$.

In this simple example, the Jacobian is the unit matrix. Thus, asymptotically, if we perform simulation with $H = 1$ (see eq. 2.19)

$$\sqrt{T}(\hat{\theta} - \theta_0) \quad \simeq \quad \sqrt{T}(X'X)^{-1}X'e_0 - \sqrt{T}(X'X)^{-1}X'\tilde{e} \tag{4.36}$$

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

If simulation is performed with H replications, from eq. (3.25) it follows that the asymptotic variance-covariance matrix of the indirect estimator will be $(1 + 1/H)$ times that of the OLS estimator, $(1 + 1/H)lim(X'X/T)^{-1}$.

Let us now apply the control variate procedure, with $H = 1$.

$$\tilde{\hat{\beta}} = \hat{\beta} + (X'X)^{-1}X'\tilde{e} \tag{4.37}$$

$$\text{NPEER}(\hat{\beta}, \tilde{e}) = (X'X)^{-1}X'\tilde{e} \tag{4.38}$$

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

$$\hat{\theta}_{cv} = \hat{\theta} + \ \text{NPEER}(\hat{\beta}, \tilde{e}) = \hat{\theta} + (X'X)^{-1}X'\tilde{e} \tag{4.39}$$

$$\begin{aligned} \sqrt{T}(\hat{\theta}_{cv} - \theta_0) &= \sqrt{T}\left[(\hat{\theta} - \theta_0) + (\hat{\theta}_{cv} - \hat{\theta})\right] \\ &= \sqrt{T}(X'X)^{-1}X'e_0 + \left\{\sqrt{T}(X'X)^{-1}X'\tilde{e} - \sqrt{T}(X'X)^{-1}X'\tilde{e}\right\} \end{aligned} \tag{4.40}$$

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

4.2 The variance of the control variate estimator

In general, it is not straightforward to predict the efficiency gain produced by the control variates. Intuitively, if the *econometric* model (2.5) and the *auxiliary* model (2.6) are quite close to each other, the last component in braces on the right hand side of equation (4.33) 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; 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 previous section.

The asymptotic variance-covariance matrix of the control variate indirect estimator can be derived as follows.

From equation (4.33), given the independence between e_0 and $\tilde{e}$ and since $\sqrt{T}\textsc{peer}$ and $\sqrt{T}\textsc{npeer}$ have asymptotically zero expected value, we can write asymptotically

$$\begin{aligned} Var\left(\sqrt{T}(\hat{\theta}_{cv}-\theta_0)\right) = R_0^{-1}\Big\{Var\left[\sqrt{T}\textsc{peer}(\theta_0,e_0)\right] + Var\left[\sqrt{T}\textsc{peer}(\theta_0,\tilde{e})\right] \\ +Var\left[\sqrt{T}\textsc{npeer}(b(\theta_0),\eta(\tilde{e}))\right] - E\left[\sqrt{T}\textsc{peer}(\theta_0,\tilde{e}) * \sqrt{T}\textsc{npeer}'(b(\theta_0),\eta(\tilde{e}))\right] \\ -E\left[\sqrt{T}\textsc{npeer}(b(\theta_0),\eta(\tilde{e})) * \sqrt{T}\textsc{peer}'(\theta_0,\tilde{e})\right]\Big\} R_0'^{-1} \end{aligned} \tag{4.41}$$

The first two terms give the variance of the simple indirect estimator (cfr. equation 2.20).

The third term is easily computable, being the variance of a *correctly specified* model. We have already observed just after equation (4.27) that it is asymptotically

$$Var\left[\sqrt{T}\textsc{npeer}(b(\theta_0),\eta(\tilde{e}))\right] = H_2^{-1} \tag{4.42}$$

being H_2 the Hessian of the *auxiliary* model, when the values of y are produced by the *auxiliary* model itself.

Asymptotically, the expected value of the products which are the last terms of equation (4.41) is obtained combining the inverted Hessians and the outer product matrices of the auxiliary model when the y data are produced alternatively by the *econometric* model or by the *auxiliary* model. The final result will be

$$\begin{aligned} Var\left(\sqrt{T}(\hat{\theta}_{cv}-\theta_0)\right) \\ = R_0^{-1}\left\{2H_1^{-1}J_1H_1^{-1} + H_2^{-1} - H_1^{-1}J_{1,2}H_2^{-1} - H_2^{-1}J_{2,1}H_1^{-1}\right\} R_0'^{-1} \end{aligned} \tag{4.43}$$

H_1 and J_1 are the probability limits of the *auxiliary* model's Hessian and outer-product matrices, computed at $b(\theta_0)$, when the values of y are produced by the *econometric* model,

as in equation (2.9). H_2 is again the probability limit of the Hessian of the *auxiliary* model (2.6). when the y data have been produced by the *auxiliary* model itself.

$J_{1,2}$ is the probability limit of the *cross-products* matrix of first derivatives, when the y data have been first produced by the *econometric* model, and then by the *auxiliary* model. Finally, $J_{2,1}$ is the transpose of $J_{1,2}$.

The proof of equation (4.43) is completed in appendix A.

For cases of practical interest, the performance of the control variate estimator and of its variance-covariance matrix can be evaluated by means of Monte Carlo experiments. This will be done in the examples of section 5.

4.3 Computational procedure for control variates

After the simple indirect estimator $\hat{\theta}$ has been computed (step 8 of section 2.1), with the value $\hat{\beta}$ and the η-transformation of the pseudo-random error terms $\tilde{e}$ (step 4 of the section 2.1) we simulate the *auxiliary* model (2.6). With the simulated series of y we re-estimate the same model (2.6), obtaining an estimate that we call $\tilde{\hat{\beta}}$. Now, $\tilde{\hat{\beta}} - \hat{\beta}$ is the *new* parameter estimation error, $\text{NPEER}(\hat{\beta}, \eta(\tilde{e}))$, that produces the control variates. Thus, the indirect estimator with control variates will be

$$\hat{\theta}_{cv} = \hat{\theta} + \hat{R}^{-1} \ \text{NPEER}(\hat{\beta}, \eta(\tilde{e})) = \hat{\theta} + \hat{R}^{-1}(\tilde{\hat{\beta}} - \hat{\beta}) \tag{4.44}$$

where the Jacobian $\hat{R}$ is computed with numerical differentiation as

$$\hat{R} = \left[\frac{\partial \beta(\theta, \tilde{e})}{\partial \theta'} \right]_{\hat{\theta}} \tag{4.45}$$

The estimate of the variance-covariance matrix of $\hat{\theta}_{cv}$ can be obtained using equation (4.43), where R_0 is replaced by $\hat{R}$ of equation (4.45). The matrices H_1 and J_1 are computed in $\hat{\beta}$ with the y data produced by simulation of the *econometric* model, while H_2 is computed also in $\hat{\beta}$, but using the y data produced by simulation of the *auxiliary* model. Both sets of y data must be used to compute the *cross-products* matrix $J_{1,2}$ and its transpose $J_{2,1}$.

4.4 Remark

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 (see, for example, An and Liu, 1996). Since an analytic solution does not exist, the problem must be solved numerically. We have adopted the following updating equation

$$\tilde{\theta}_{(j)} = \tilde{\theta}_{(j-1)} + \lambda A^{-1}_{(j-1)}(\tilde{\beta}_{(j-1)} - \hat{\beta}) \tag{4.46}$$

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 most of our applications we alternate one "complicated" (Newton) iteration taking A equal to the Jacobian matrix of derivatives of the *auxiliary* parameters with respect to the parameters of interest ($\hat{R}$ of equation 4.30) with some "simple" iterations (typically three to four) without changing A. This heuristic switching rule is maintained until convergence is reached. Only for some models (when the θ and β parameters are essentially the same, even if plugged into different models) we could also perform iterations using the identity matrix (Jacobi solution method, or "fine tuning", using the terminology of An and Liu, 1996, or Mealli and Rampichini, 1999).

5 Models and Monte Carlo results

The design of the Monte Carlo is as follows. The *econometric* models and their *true* parameters are kept fixed in all experiments. For each of the models considered in this section we present a table of results, each table being related to a Monte Carlo experiment with 10 000 replications (only 2000 for one of the stochastic volatility models, where the time series are particularly long). In each table, a row displays:

1) The *true* value of a θ_0 parameter (used in all Monte Carlo replications to generate the pseudo-observed data).

2) The Monte Carlo mean of the simple indirect estimates of the parameter, computed across 10 000 replications.

3) The Monte Carlo mean of the control variate indirect estimate of the parameter.

4) The Monte Carlo mean of the naive estimates of a parameter in the *auxiliary* model (in some cases, it will be the β parameter that "reasonably" corresponds to the θ parameter).

Under each mean, in square brackets, we display the Monte Carlo variance of the parameter, computed across the 10 000 replications. In parentheses, we display the mean of the estimated variance of the parameter, computed across the same replications.

In each replication, the variance of the naive estimator is computed as in White (1982), being the *auxiliary* model misspecified. There are cases where White's estimator is inappropriate, for the presence of autocorrelation. In these cases, the Newey and West (1987) estimator has been used.

The estimated variance-covariance matrix of the simple indirect estimator is computed as in section 2.3.

For the indirect estimator with control variates, the estimate of the variance-covariance

matrix is computed as in section 4.2.

5.1 Moving average process

Let us consider a moving average process of order 1

$$y_t = \epsilon_t + \theta\epsilon_{t-1} \qquad \epsilon_t \sim N(0, \sigma^2) \tag{5.47}$$

The parameters of the *econometric* model are therefore θ and σ^2. With tentative values of the parameters, the model can be easily simulated generating pseudo-random errors $\tilde{\epsilon}_t = \tilde{\sigma}\tilde{e}_t$, where $\tilde{e}_t$ are $i.i.d.N(0,1)$.

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

Par.	*True*	Ind.Inf H=1	Ind.Inf Cntr.Var.	*Par.*	Least Sqr.
θ	0.5	0.501 $(.11*10^{-2})[.11*10^{-2}]$	0.500 $(.59*10^{-3})[.59*10^{-3}]$	β	0.400 $(.13*10^{-3})[.13*10^{-3}]$
σ^2	1.0	0.999 $(.99*10^{-3})[.10*10^{-2}]$	1.00 $(.53*10^{-3})[.53*10^{-3}]$	ψ^2	1.05 $(.47*10^{-3})[.46*10^{-3}]$
T=5000					Replications=10 000

As *auxiliary* model, we adopt a rough approximation based on an autoregressive model of order 1, where parameters are β and ψ^2.

$$y_t = \beta y_{t-1} + \psi\eta_t \qquad \eta_t \sim N(0,1) \tag{5.48}$$

The coefficient β and the residual variance ψ^2 are easily estimated by least squares. This is a simple *textbook* case that well exemplifies the performance of indirect inference (see Gourieroux and Monfort, 1996, pp.71-73). To produce the control variates, we use the *auxiliary* model (5.48). After the simple indirect estimates have been computed ($\hat{\theta}$, $\hat{\sigma}^2$), we can simulate model (5.48) using $\hat{\beta}$ and $\hat{\psi}^2$ and generating additive pseudo-random errors as $\hat{\psi}\tilde{e}_t$, where $\tilde{e}_t$ are the same pseudo-random errors $N(0,1)$ used to simulate (5.47). With these simulated data we re-estimate model (5.48), obtaining $\tilde{\hat{\beta}}$ and $\tilde{\hat{\psi}}^2$, and finally we use the differences $\tilde{\hat{\beta}} - \hat{\beta}$ and $\tilde{\hat{\psi}}^2 - \hat{\psi}^2$ to adjust the simple indirect estimates and produce the control variate estimates $\hat{\theta}_{cv}$ and $\hat{\sigma}^2_{cv}$.

Results are given in Table 1.

The least squares estimator of the *auxiliary* model shows a quite evident bias for both parameters. Indirect inference (with or without control variates) adjusts for the bias (inconsistency): the mean estimated parameter is very close to the *true* value.

For the simple indirect estimator, the mean of the estimated variances is remarkably close to the Monte Carlo variance; the Newey-West matrix has been used in the estimation.

Control variates produce a remarkable reduction of the variance of the two parameter estimates, with respect to simple indirect estimates. The result is almost striking. Variance is reduced by almost 50%. In other words, the enlargement of the variance due to simulation is practically negligible. A similar result could be obtained by simple indirect estimation, using $H = 20$ or more, thus at much higher computational cost.

5.2 Autoregressive moving average process

Let us consider an ARMA(1,1) process

$$y_t = \phi y_{t-1} + \epsilon_t + \theta \epsilon_{t-1} \qquad \epsilon_t \sim N(0, \sigma^2) \tag{5.49}$$

The *econometric* parameters of interest are ϕ, θ and σ^2. With tentative values of the parameters, the model can be easily simulated generating pseudo-random errors $\tilde{\epsilon}_t = \tilde{\sigma}\tilde{e}_t$, where $\tilde{e}_t$ are $i.i.d.N(0,1)$.

Table 2: ARMA(1,1) – mean estim. param. and (var.), [Monte-Carlo var.]

Par.	*True*	Ind.Inf H=1	Ind.Inf Cntr.Var.	*Par.*	Least Sqr.
ϕ	0.4	0.399 (.93*10^{-3})[.99*10^{-3}]	0.401 (.51*10^{-3})[.51*10^{-3}]	β_1	0.772 (.17*10^{-3})[.17*10^{-3}]
θ	0.4	0.402 (.14*10^{-2})[.14*10^{-2}]	0.399 (.74*10^{-3})[.74*10^{-3}]	β_2	-0.233 (.16*10^{-3})[.16*10^{-3}]
σ^2	1.0	1.00 (.84*10^{-3})[.83*10^{-3}]	1.00 (.42*10^{-3})[.42*10^{-3}]	ψ^2	1.014 (.41*10^{-3})[.39*10^{-3}]
T=5000					Replications=10 000

As *auxiliary* model, we adopt a rough approximation based on autoregressive model of order 2, whose coefficients β_1 and β_2 are quite simple to estimate by least squares, as well as the residual variance ψ^2

$$y_t = \beta_1 y_{t-1} + \beta_2 y_{t-2} + \psi \eta_t \tag{5.50}$$

Although slightly more complex than the previous case, this is again a rather simple example of the performance of indirect inference. To produce the control variates, we use the *auxiliary* model (5.50), which can be simulated using $\hat{\beta}_1$, $\hat{\beta}_2$ and generating additive pseudo-random errors $\hat{\psi}\tilde{e}_t$, where $\tilde{e}_t$ are the same $N(0,1)$ used to simulate (5.49).

Results are given in Table 2.

As for the previous model, the quite evident bias (inconsistency) of the least squares estimator is adjusted by Indirect inference (with or without control variates). The results are quite similar to those of the previous model. The reduction of the variance of all parameter estimates is even more striking. Control variates reduce by almost 50% the global variance of the simple indirect estimator. A similar result could be obtained by simple indirect estimation, using $H = 30$ or more, thus at much higher computational cost.

5.3 Ornstein-Uhlenbeck process

As *econometric* model let us now consider the stochastic differential equation employed by Vasicek (1977) to explain the behaviour of short-term interest rates (Ornstein-Uhlenbeck process)

$$dy_t = k(a - y_t)dt + \sigma d\mathcal{W}_t \tag{5.51}$$

where y_t is the spot interest rate, $\mathcal{W}_t$ is a Wiener process.

Table 3: O.U. – mean estim. param. and (var.), [Monte-Carlo var.]

Par.	*True*	Ind.Inf H=1	Ind.Inf Cntr.Var.	*Par.*	Least Sqr.
a	0.1	0.0998 $(.40*10^{-4})[.41*10^{-4}]$	0.1001 $(.22*10^{-4})[.22*10^{-4}]$	a	0.0999 $(.20*10^{-4})[.19*10^{-4}]$
k	0.5	0.5020 $(.16*10^{-2})[.17*10^{-2}]$	0.5009 $(.92*10^{-3})[.90*10^{-3}]$	k	0.3985 $(.35*10^{-3})[.34*10^{-3}]$
σ^2	0.01	0.0100 $(.31*10^{-6})[.30*10^{-6}]$	0.0100 $(.17*10^{-6})[.17*10^{-6}]$	σ^2	0.0064 $(.39*10^{-7})[.40*10^{-7}]$
T=2000					Replications=10 000

In many empirical works, simulation of model (5.51) is performed by resorting to a discretization, at very small time intervals. Examples can be found in Bianchi and Cleur (1996), Broze, Scaillet and Zakoïan (1995, 1998), Di Iorio (1996), Pastorello, Renault and Touzi (1994). Particularly simple is the so-called Euler scheme, which is adopted in this paper. Other types of approximations could be adopted, like 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). Using the Euler scheme, the discretized model used for simulation instead of (5.51) is

$$y_t - y_{t-\delta} = ka\delta - ky_{t-\delta}\delta + \sqrt{\delta\sigma^2}\, e_t \tag{5.52}$$

where e_t are Gaussian error terms. With the notation of the previous sections, the vector

of parameters of interest is $\theta = (a, k, \sigma^2)'$.

Simulation can be made with arbitrarily small time intervals δ, and regularity conditions can ensure that the discretized model, with a conveniently small 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, 1994, Broze, Scaillet and Zakoïan, 1998, and Bianchi and Cleur, 1996). So model (5.52) can be considered the *econometric* model of interest.

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

Since actual data are observed at discrete time intervals (unit time intervals), a simple empirical estimation in most available applications is performed on the following *auxiliary* model

$$y_t - y_{t-1} = ka - ky_{t-1} + \sigma\eta_t \tag{5.53}$$

where $E_{t-1}(\eta_t) = 0$ and $E_{t-1}(\eta_t^2) = 1$. There is such an obvious correspondence between the θ and the β parameters, that we can keep also for the parameters of the *auxiliary* model the same symbols as for the *econometric* model: $\beta = (a, k, \sigma^2)'$.

A *naive* estimator of the discretized model (5.53) is easily obtained by least squares.

The results presented in Table 3 are related to a sample period length $T = 2000$ (for analogous results related to $T = 1000$ see Calzolari, Di Iorio and Fiorentini, 1998). This also is the length of the simulated series ($H = 1, HT = T$). The simulation step is $\delta = 0.05$ (thus $T/\delta = 40\ 000$).

To obtain the control variates, we simulate the *auxiliary* model (5.53), using $\hat{\beta}$ and a vector of *i.i.d.* $N(0, 1)$ pseudo-random errors $\tilde{\eta} = \eta(\tilde{e})$ of length T, whose tth element is $\tilde{\eta}_t = \sqrt{\delta}(\tilde{e}_{t-1+\delta} + \tilde{e}_{t-1+2\delta} + \cdots + \tilde{e}_{t-\delta} + \tilde{e}_t)$.

5.4 Square-root process

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

$$dy_t = k(a - y_t)dt + \sigma\sqrt{y_t}d\mathcal{W}_t \tag{5.54}$$

where $\mathcal{W}_t$ is a Wiener process.

The discretized model used for simulation instead of (5.54) is

Table 4: Sq.rt. – mean estim. param. and (var.), [Monte-Carlo var.]

Par.	*True*	Ind.Inf H=1	Ind.Inf Cntr.Var.	*Par.*	Least Sqr.
a	0.1	0.1000 (.41*10^{-5})[.41*10^{-5}]	0.1000 (.21*10^{-5})[.22*10^{-5}]	a	0.1000 (.20*10^{-5})[.19*10^{-5}]
k	0.5	0.5003 (.17*10^{-2})[.18*10^{-2}]	0.5002 (.95*10^{-3})[.97*10^{-3}]	k	0.3986 (.35*10^{-3})[.34*10^{-3}]
σ^2	0.01	0.0100 (.30*10^{-6})[.32*10^{-6}]	0.0100 (.16*10^{-6})[.17*10^{-6}]	σ^2	0.0066 (.50*10^{-7})[.51*10^{-7}]
T=2000					Replications=10 000

$$y_t - y_{t-\delta} = ka\delta - ky_{t-\delta}\delta + \sqrt{y_{t-\delta}}\ \sqrt{\delta\sigma^2}\ e_t \tag{5.55}$$

where e_t are a Gaussian error terms. With the same considerations as in the previous subsection, (5.55) will be considered as the *econometric* model of interest.

The *auxiliary* model we use is

$$y_t - y_{t-1} = ka - ky_{t-1} + \sqrt{y_{t-1}}\ \sigma\eta_t \tag{5.56}$$

where $E_{t-1}(\eta_t) = 0$ and $E_{t-1}(\eta_t^2) = 1$.

The vector of parameters of interest is $\theta = (a, k, \sigma^2)'$. The parameters of the *auxiliary* model are the same as for the *econometric* model: $\beta = (a, k, \sigma^2)'$ (strict correspondence between the θ and the β parameters).

A *naive* estimator of the discretized model (5.56) is easily obtained by weighted least squares, that is least squares after data have been divided by $\sqrt{y_{t-1}}$.

To obtain the control variates, we simulate the *auxiliary* model (5.56), using $\hat{\beta}$ and a vector of *i.i.d.* $N(0,1)$ pseudo-random errors $\tilde{\eta} = \eta(\tilde{e})$ of length T, whose tth element is $\tilde{\eta}_t = \sqrt{\delta}(\tilde{e}_{t-1+\delta} + \tilde{e}_{t-1+2\delta} + \cdots + \tilde{e}_{t-\delta} + \tilde{e}_t)$.

Results are displayed in Table 4.

5.5 Stochastic differential equation with unrestricted variance elasticity

The model of interest is the same as in Chan, Karolyi, Longstaff and Sanders (1992; CKLS hereinafter):

$$dy_t = k(a - y_t)dt + \sigma y_t^{\gamma} d\mathcal{W}_t\ , \tag{5.57}$$

where $\mathcal{W}_t$ is a Wiener process and $\theta = (a, k, \sigma, \gamma)'$ is the vector of all parameters of interest.

Table 5: CKLS – mean estim. param. and (var.), [Monte-Carlo var.]

Par.	*True*	Ind.Inf H=1	Ind.Inf Cntr.Var.	*Par.*	NLFIML
a	0.1	.1001 $(.41*10^{-5})[.43*10^{-5}]$	.0999 $(.21*10^{-5})[.22*10^{-5}]$	a	.0999 $(.20*10^{-5})[.21*10^{-5}]$
k	0.5	.5021 $(.17*10^{-2})[.17*10^{-2}]$	.5024 $(.96*10^{-3})[.97*10^{-3}]$	k	.3988 $(.33*10^{-3})[.33*10^{-3}]$
γ	0.5	0.4992 $(.87*10^{-2})[.95*10^{-2}]$	0.5019 $(.54*10^{-2})[.59*10^{-2}]$	γ	0.3675 $(.27*10^{-2})[.25*10^{-2}]$
σ^2	0.01	0.0110 $(.28*10^{-4})[.27*10^{-4}]$	0.0105 $(.18*10^{-4})[.17*10^{-4}]$	σ^2	0.0036 $(.82*10^{-6})[.86*10^{-6}]$
T=2000					Replications=10 000

With the same considerations for the square root process and the Ornstein-Uhlenbeck process, the discretized model used for simulation instead of (5.57) is

$$y_t - y_{t-\delta} = ka\delta - ky_{t-\delta} + y_{t-\delta}^{\gamma}\ \sqrt{\delta\sigma^2}\ e_t \tag{5.58}$$

where e_t are i.i.d. standard normal disturbances. A natural choice for the *auxiliary* model is

$$y_t - y_{t-1} = ka - ky_{t-1} + y_{t-1}^{\gamma}\ \sigma\ \eta_t \tag{5.59}$$

where $E_{t-1}(\eta_t) = 0$ and $E_{t-1}(\eta_t^2) = 1$.

Even in this case the parameters of the *auxiliary* model are the same of the *econometric* model. In this case the performance of indirect inference estimators depends on the maximization criterion chosen for the *auxiliary* model. In fact, some experiments have shown that the main difficulty is the estimation of γ.

In this case, we follow an approach similar to that proposed by Calzolari and Fiorentini (1994) for systems of nonlinear simultaneous equations with ARCH errors. The *auxiliary* model can be seen as a nonlinear implicit equation, and written as

$$y_t y_{t-1}^{-\gamma}\ -\ ka y_{t-1}^{-\gamma}\ +\ (k-1)y_{t-1}^{1-\gamma} = \sigma\eta_t \tag{5.60}$$

where we also assume $\eta_t \sim N(0,1)$. Using the notation in Amemiya (1977), we can write

$$f(y_t, y_{t-1}, \alpha) = u_t$$

where $\alpha = (k, a, \gamma)'$ (and therefore $\beta = (\alpha', \sigma^2)'$ is the whole vector of *auxiliary* parameters).

The maximum likelihood estimation can be conducted by iterating to convergence an instrumental variables method (NLFIML, following Amemiya, 1977), that is

$$\hat{\alpha}_{(m+1)} = \hat{\alpha}_{(m)} - [\hat{G}'_{(m)} G_{(m)}]^{-1} \hat{G}'_{(m)} \hat{u}_{(m)} \qquad m = 1, 2... \tag{5.61}$$

where the matrix $G'_{(m)}$ has columns $g_{t,(m)} = \frac{\partial f}{\partial \alpha}|_{\hat{\alpha}_{(m)}}$, and the matrix $\hat{G}'_{(m)}$ has columns

$$\hat{g}_{t,(m)} = g_{t,(m)} - \left(\frac{1}{T} \sum_{s=1}^{T} \frac{\partial g_{s,(m)}}{\partial u_s} \right) \hat{u}_{t,(m)} \tag{5.62}$$

with all derivatives evaluated at $\hat{\alpha}_{(m)}$. The estimator of σ^2 is obtained from the residuals $\hat{u}_t = f(y_t, y_{t-1}, \hat{\alpha})$.

This procedure usually achieves convergence after very few iterations (typically 5-10).

Control variates are obtained from simulation of the *auxiliary* model, with a technique quite similar to what adopted for the previous differential equation models.

Results are displayed in Table 5.

5.6 Dynamic Tobit model

Let us consider the following dynamic Tobit model

$$\begin{aligned} y_t^* &= \theta_1 + \theta_2 \; y_{t-1}^* + \epsilon_t \qquad \epsilon_t \sim N(0, \sigma^2) \\ y_t &= max(y_t^*, 0) \end{aligned} \tag{5.63}$$

The *econometric* parameters of interest are θ_1, θ_2 and σ^2. With tentative values of the parameters, the model can be easily simulated generating pseudo-random errors $\tilde{\epsilon}_t = \tilde{\sigma} \tilde{e}_t$, where $\tilde{e}_t$ are $i.i.d.N(0, 1)$.

Table 6: Dyn.Tobit – mean estim. param. and [Monte-Carlo var.] – 50% censored

Par.	*True*	Ind.Inf H=1	Ind.Inf Cntr.Var.	*Par.*	Least Sqr.
θ_1	0.0	-0.104*10^{-2} [.29*10^{-2}]	-0.209*10^{-2} [.19*10^{-2}]	β_1	0.265 [.47*10^{-3}]
θ_2	0.5	0.501 [.29*10^{-2}]	0.499 [.25*10^{-3}]	β_2	0.423 [.14*10^{-2}]
σ^2	1.0	1.01 [.13*10^{-1}]	1.01 [.10*10^{-1}]	ψ^2	0.371 [.94*10^{-3}]
T=5000				Replications=10 000	

As *auxiliary* model, we adopt the autoregressive equation of order 1, treating the zeroes as regular values of the dependent variable

$$y_t = \beta_1 + \beta_2 \; y_{t-1} + \psi \eta_t \tag{5.64}$$

Coefficients β_1 and β_2 are quite simple to estimate by least squares, as well as the residual variance ψ^2.

To produce the control variates, we use the *auxiliary* model (5.64), which can be simulated using $\hat{\beta}_1$, $\hat{\beta}_2$ and generating additive pseudo-random errors $\hat{\psi}\tilde{e}_t$, where $\tilde{e}_t$ are the same $N(0,1)$ used to simulate (5.63).

Results are given in Table 6. In this example, the percentage of censored values is 50%. We display in the table only the Monte Carlo variances and not the average estimated variances (whose computation has encountered several numerical problems that need further investigation).

For results with different simulation-based estimators, see Billio, Monfort, and Robert (1998).

5.7 Discrete time stochastic volatility model

The *econometric* model is the simplest among the *dynamic* stochastic volatility models in discrete time, considered in Danielsson (1994). Its major properties are discussed in Taylor (1986, 1994). Detailed results on indirect inference for this model are presented in Monfardini (1998). We represent the model as

$$\begin{aligned} y_t &= e^{\frac{1}{2}h_t} \; r_t \\ h_t &= \theta_1 + \theta_2 h_{t-1} + \sigma \; v_t \end{aligned} \tag{5.65}$$

where y_t is the observed variable, h_t is unobservable, $e_t = (r_t, v_t)'$ are $i.i.d.$ $N(0, I)$ vectors.

Table 7: S.V. – mean estim. param. and [Monte-Carlo var.]

Par.	*True*	Ind.Inf H=1	Ind.Inf Cntr.Var.	*Par.*	Least Sqr.
θ_1	0.0	-.31*10^{-3} [.28*10^{-5}]	-.18*10^{-4} [.25*10^{-5}]	β_1	-.59*10^{-3} [.77*10^{-4}]
θ_2	0.9	.901 [.44*10^{-3}]	.899 [.25*10^{-3}]	β_2	.345 [.50*10^{-4}]
σ^2	0.1	0.099 [.82*10^{-3}]	.099 [.46*10^{-3}]	ψ^2	0.401 [.99*10^{-3}]
T=100 000-Aux.mod.1					Replications=2000

Transforming the observed variable as $log(y_t^2)$ we get the sum of an AR(1) process (h_t) and a non-Gaussian white noise ($log(r_t^2)$). Monfardini (1998) observes that, if both processes were Gaussian, the result would be Gaussian ARMA(1,1); being non-Gaussian, a Gaussian

Table 8: S.V. – mean estim. param. and [Monte-Carlo var.]

Par.	*True*	Ind.Inf H=1	Ind.Inf Cntr.Var.	*Par.*	Least Sqr.
θ_1	0.0	-0.24*10^{-3} [.15*10^{-4}]	0.41*10^{-4} [.13*10^{-4}]	β_1	0.26*10^{-2} [.33*10^{-3}]
θ_2	0.9	.899 [.72*10^{-3}]	.900 [.48*10^{-3}]	β_2	.343 [.36*10^{-3}]
σ^2	0.1	0.099 [.85*10^{-3}]	.100 [.54*10^{-3}]	ψ^2	0.402 [.86*10^{-3}]
T=20 000-Aux.mod.2				Replications=10 000	

ARMA(1,1) or some high order AR model (such as an AR(10)) could nevertheless be used as *auxiliary* model in indirect estimation.

We adopt an AR(10) *auxiliary* model, but to remain in a just-identified context we impose constraints on the distributed-lag coefficients. So, as *auxiliary* model, we use an AR(10)-distributed-lag model on $x_t = log(y_t^2) - E[log(r_t^2)]$ (where $E[log(r_t^2)]$ is known). The weights of the distributed lags are assumed linearly decreasing from lag-1 to lag-10.

$$x_t = \beta_1 + \beta_2[w_1 x_{t-1} + w_2 x_{t-2} + ... + w_{10} x_{t-10}] + u_t \tag{5.66}$$

We estimate by least squares the first two parameters of the *auxiliary* model, $\hat{\beta}_1$ and $\hat{\beta}_2$, and from the residual variance we subtract the variance of $log(r_t^2)$ (which is known) to produce the third parameter estimate, $\hat{\psi}^2$.

To produce the control variates, we use (5.66), simulated using $\hat{\beta}_1$ and $\hat{\beta}_2$, and pseudo--random errors $\tilde{u}_t = log(\tilde{r}_t^2) - E[log(r_t^2)] + \hat{\psi}\tilde{v}_t$, where $E[log(r_t^2)]$ is known, and $(\tilde{r}_t, \tilde{v}_t)' = \tilde{e}_t$ are the same pseudo-random errors used to simulate the *econometric* model (5.65). After simulation, we estimate by least squares the two coefficients, and the estimate of the third parameter is obtained from the residual variance, subtracting the variance of $log(r_t^2)$ (which is known).

Results are displayed in Table 7 for a *huge* sample size (100 000 observations). Such an unrealistically large number of observations is presented here only for comparison with an application to the same model, but using a slightly different *auxiliary* model.

We still use 10 lags in the distributed-lag OLS estimation, with weights that decrease linearly, so the estimation procedure for β_1 and β_2 is the same as before. But we derive the estimate $\hat{\psi}^2$ in a different way. To each least squares residual, we add $E[log(r_t^2)]$ (known), exponentiate the result, average over the sample, and finally take the double of the logarithm of the average.

The results in Table 8 show that, with 20000 observations, the variance of the simple indirect estimator is, for the second and third parameter, not much larger than in the

previous case (that had five times more observations). This well shows that even small improvements in the *auxiliary* model may lead to big improvements in the quality of the indirect estimates.

In both cases the gain due to the use of control variates is remarkable. We display in the tables only the Monte Carlo variances and not the average estimated variances (whose computation has encountered several numerical problems that need further investigation).

5.8 Two level logit model with random effects

We consider a two level logit model with random effects (also known as logit variance component model). Given a $N \times 1$ vector of binary responses y, with y_{ik} element being the response for the $i-$th individual in the $k-$th group ($i = 1, ..., n_k$; $k = 1, ..., K$; $N = \sum_{k=1}^{K} n_k$), we assume that, conditional on a $K \times 1$ vector of random effects ϵ, the elements of y are independent Bernoulli variables with probabilities[1]

$$P_{ik} = Pr(y_{ik} = 1) = \frac{exp(\theta_1 + \theta_2 x_{ik} + \epsilon_k)}{1 + exp(\theta_1 + \theta_2 x_{ik} + \epsilon_k)} \tag{5.67}$$

We assume $\epsilon_k = \sigma v_k$ *i.i.d.* $N(0, \sigma^2)$.

Table 9: Logit with random effects – mean estim. param. and [Monte-Carlo var.]

Par.	*True*	Ind.Inf H=1	Ind.Inf Cntr.Var.	*Par.*	Least Sqr.
θ_1	0.5	.502 [.65*10^{-2}]	.500 [.33*10^{-2}]	β_1	.607 [.13*10^{-3}]
θ_2	0.5	.501 [.12*10^{-2}]	.499 [.81*10^{-3}]	β_2	.098 [.22*10^{-4}]
σ^2	0.5	0.507 [.88*10^{-2}]	.497 [.50*10^{-2}]	ψ^2	.019 [.50*10^{-5}]
N=10 000;K=200					Replications=10 000

The model may be complicated by including a multiple level structure.

The conditional likelihood function is

$$L(\theta_1, \theta_2 | \epsilon) = \prod_{k=1}^{K} \prod_{i=1}^{n_k} P_{ik}^{y_{ik}} (1 - P_{ik})^{1 - y_{ik}} \tag{5.68}$$

To obtain the unconditional likelihood we need to integrate out the random effects ϵ, that is

[1] Alternatively, the model can be interpreted as a logit model for panel data with individual random effects. In that case y_{ik} would be the $k-$th observation of the $i-$th individual, with $i = 1, ..., n$; $k = 1, ..., T$; $N = nT$.

$$L(\theta_1, \theta_2, \sigma^2) = \int ... \int L(\theta_1, \theta_2|\epsilon)\phi(\epsilon)d\epsilon \tag{5.69}$$

where ϕ is the normal density.

Some difficulties arising in the computation of (5.69) have led to various approximated solutions, which make use of pseudo-likelihood functions, or first/second order Taylor expansion of model (5.67) (Goldstein, 1991, Longford, 1994). An indirect inference approach is presented in Mealli and Rampichini (1999).

As *auxiliary* model, here we use a simple linear probability model

$$y_{ik} = \beta_1 + \beta_2 x_{ik} + u_k + w_{ik} \tag{5.70}$$

where the coefficients are estimated by least squares, while the variance of u_k, say $\hat{\psi}$, is estimated as the mean of within-group covariances (the residual w_{ik} is ignored).

To produce the values of $\tilde{y}_{ik}$, we generate a $K \times 1$ vector $\tilde{v}$ of i.i.d. standard normals, and use the available estimate of σ^2 to produce $\tilde{\epsilon} = \sigma\tilde{v}$. Then we generate $\tilde{r}_{ik}$ uniform in (0,1), compare its value with the probability (5.67), and assign a 0 or 1 value to $\tilde{y}_{ik}$ accordingly. In the general notation used in this paper, the pseudo-random error $\tilde{e}_{ik}$, for each individual, is a bivariate random vector with independent components $\tilde{v}_k$ (normal, the same for all individuals of group k) and $\tilde{r}_{ik}$ (uniform).

To produce the control variates, we simulate the *auxiliary* model (5.70). We use $\tilde{u}_k = \hat{\psi}\tilde{v}_k$ (the same $\tilde{v}$ as before), compute the value on the right hand side of (5.70) (ignoring w_{ik}), compare the value with $\tilde{r}_{ik}$ (the same as before), and assign a 0 or 1 value to $\tilde{\hat{y}}_{ik}$ accordingly. If the linear probability value is less than zero or greater than one, its value is directly assigned to $\tilde{\hat{y}}_{ik}$, instead of 0 or 1. Finally, coefficients are computed by least squares, and the variance of u_k is estimated as the mean of within-group covariances.

In our Monte Carlo experiment, the covariates x_{ik} are generated as independent $N(.5, 1)$, fixed in all replications.

Results are displayed in Table 9. Indirect estimation adjusts the huge bias (inconsistency) of the *naive* least squares estimator of the *auxiliary* model. Control variates reduce considerably the variance (we display in the table only the Monte Carlo variances and not the average estimated variances, whose computation has encountered several numerical problems that need further investigation).

6 Summary comments

For those models where there is a strict correspondence between θ and β parameters, the bias (inconsistency) of the *naive* estimator is quite evident for most parameters. Indirect

estimation (with or without control variates) adjusts for the bias (inconsistency): the mean estimated parameter is practically equal to the *true* value. This last consideration holds also for those models where the strict correspondence between θ and β parameters is not too obvious (and so we could not properly speak of a bias for the *naive* estimator).

Control variates produce a remarkable reduction of the variance of some parameter estimates, with respect to simple indirect estimates. There are cases where the reduction is almost 50%, which means that the *additional* variance due to simulation is practically cancelled, as it would happen by performing simulations with a large value of H. But a large value of H would cause a big growth of the computationl costs. Thus, control variates strongly increase the computational efficiency.

There are case where the reduction is lower than 50%, which means that the improvement of the computational efficiency is not as large as before, but is still remarkable.

For most models we have also estimated, at each Monte Carlo replication, the variance of the estimated parameters, both for the simple indirect estimator and for the control variate estimator. The sample average of the estimated variances is usually quite close to the Monte Carlo variance. This is, first of all, a good guarantee of the correctness of numerical results. Moreover, the proposed estimator of the variances (not available in Calzolari, Di Iorio and Fiorentini, 1998) makes our control variates completely applicable also to problems with *real* data.

A general FORTRAN program has been implemented to perform the experiments of this paper.

Some specific subroutines must be prepared for each model, following the rules and constraints of the general framework. In particular, model specific subroutines must be provided for the simulation of the *econometric* model, for the estimation of the *auxiliary* model, for the simulation of the *auxiliary* model, and for computing the pseudo-likelihood function of the *auxiliary* model.

7 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 computational cost (that is, computation time), an indirect estimator with control variates can be as efficient as a simple indirect estimator based on much longer simulated series that, therefore, involves a much higher cost.

A large set of Monte Carlo experiments on a variety of different models has shown the magnitude of the efficiency gain.

A Appendix: The variance of the control variate indirect estimator

In this appendix we provide the proof for the control variate covariance matrix estimator described in section 4.2. For the control variate indirect estimator, the asymptotic covariance matrix is

$$\begin{aligned} Var\left(\sqrt{T}(\hat{\theta}_{cv}-\theta_0)\right) = R_0^{-1}\Big\{Var\left[\sqrt{T}\text{PEER}(\theta_0,e_0)\right] + Var\left[\sqrt{T}\text{PEER}(\theta_0,\tilde{e})\right] \\ +Var\left[\sqrt{T}\text{NPEER}(b(\theta_0),\eta(\tilde{e}))\right] - E\left[\sqrt{T}\text{PEER}(\theta_0,\tilde{e}) * \sqrt{T}\text{NPEER}'(b(\theta_0),\eta(\tilde{e}))\right] \\ -E\left[\sqrt{T}\text{NPEER}(b(\theta_0),\eta(\tilde{e})) * \sqrt{T}\text{PEER}'(\theta_0,\tilde{e})\right]\Big\} R_0'^{-1} \end{aligned} \tag{1.71}$$

The first three terms on the right hand side have been fully discussed in section 4.2. We prove here equation (4.43) as far as the last two terms are concerned.

As in Gourieroux, Monfort and Renault (1993), we assume that the *auxiliary* model's parameters (β) can be estimated by maximizing the function $Q_T(y,\beta)$ To simplify the exposition, we consider Q_T as ($1/T$ times) the log-likelihood of the *auxiliary* model (therefore misspecified or *quasi* log-likelihood). The estimator of the β parameters sets to zero the first derivatives of Q_T. We can now perform Taylor's expansion of the first derivatives around $b(\theta_0)$, analogously to the standard proof of ML asymptotic normality (e.g. Theil, 1971, p.395), and we get

$$\sqrt{T}\text{PEER}(\theta_0,\tilde{e}) \simeq -H_1^{-1}\sqrt{T}\frac{\partial Q_T}{\partial \beta}\left[y=f(\theta_0,\tilde{e}),\beta=b(\theta_0)\right] \tag{1.72}$$

where $H_1 = plim_{T\to\infty}\frac{\partial^2 Q_T}{\partial\beta\partial\beta'}$ (the Hessian matrix).

It must be noted that the above notation indicates that the y data are produced by the *econometric* model (2.5).

In the same way, when the y data are produced by the *auxiliary* model (2.6), we can expand the first derivatives of Q_T around $b(\theta_0)$ and obtain

$$\sqrt{T}\text{NPEER}(b(\theta_0),\eta(\tilde{e})) \simeq -H_2^{-1}\sqrt{T}\frac{\partial Q_T}{\partial \beta}\left[y=g(b(\theta_0),\eta(\tilde{e})),\beta=b(\theta_0)\right] \tag{1.73}$$

where $H_2 = plim_{T\to\infty}\frac{\partial^2 Q_T}{\partial\beta\partial\beta'}$. We must note that, this time, Q_T is the *true* log-likelihood function, as the y data are produced by the *auxiliary* model (2.6).

From (1.72) and (1.73) we have (asymptotically)

$$E\left[\sqrt{T}\text{PEER}(\theta_0,\tilde{e}) * \sqrt{T}\text{NPEER}'(b(\theta_0),\eta(\tilde{e}))\right] =$$

$$H_1^{-1} E\left\{\left[\sqrt{T}\frac{\partial Q_T}{\partial \beta}\left(y = f(\theta_0,\tilde{e}), \beta = b(\theta_0)\right)\right]\left[\sqrt{T}\frac{\partial Q_T}{\partial \beta'}\left(y = g(b(\theta_0),\eta(\tilde{e})), \beta = b(\theta_0)\right)\right]\right\} H_2^{-1} \tag{1.74}$$

The intermediate matrix involves cross-products of the first derivatives of the *quasi* log-likelihoods and log-likelihoods when the y data are produced by the *econometric* model (2.5) and *auxiliary* model (2.6), respectively. Asymptotically, we indicate it as $J_{1,2}$ and $J_{2,1}$ its transpose.

This completes the proof.

To estimate $J_{1,2}$ we use something similar to the Newey and West (1987) estimator of J_1, discussed in section 2.2. In particular, we need cross-products of first derivatives at different time-lags.

References

Amemiya, T. (1977):"The Maximum Likelihood and the Nonlinear Three-Stage Least Squares Estimator in the General Nonlinear Simultaneous Equation Model", *Econometrica* 45, 955-968.

An, Y. and M. Liu (1996): "Using Indirect Inference to Solve the Initial Conditions Problem", Duke University, Department of Economics, mimeo.

Bianchi, C., R. Cesari and L. Panattoni (1994): "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.

Billio, M., A. Monfort, and C. P. Robert (1998): "The Simulated Likelihood Ratio (SLR) Method". Paris: Crest, Document de Travail No. 9821.

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.

Broze, L., O. Scaillet and J. M. Zakoïan (1998): "Quasi Indirect Inference for Diffusion Processes", *Econometric Theory* 14, 161-186.

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

Calzolari, G., F. Di Iorio, and G. Fiorentini (1998): "Control Variates for Variance Reduction in Indirect Inference: Interest Rate Models in Continuous Time", *Econometrics Journal* 1, C100-C112.

Calzolari, G. and G. Fiorentini (1994): "Conditional Heteroskedasticity in Nonlinear Simultaneous Equations". Firenze: European University Institute, WP ECO n.94/44.

Calzolari, G., and G. Fiorentini (1996): "Indirect Inference for Censored Regression Models with Conditional Heteroskedasticity", presented at the *European Meeting of the Econometric Society.* Istanbul: Bogazici University, August 25-29.

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.

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

Danielsson, J. (1994): "Stochastic Volatility in Asset Prices: Estimation with Simulated Maximum Likelihood", *Journal of Econometrics* 64, 375-400.

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.*

Di Iorio, F. and G. Fiorentini (1996): "Indirect Inference for Continuous Time Interest Rate Models with Unrestricted Variance Elasticity", presented at the *Second Conference in Computing in Economics and Finance*, Geneve, 26-28 June.

Fiorentini, G, E. Sentana and N. Shephard (1998): "Exact Likelihood- Based Estimation of Conditionally Heteroskedastic Factor Models", Universidad de Alicante, Departamento de Fundamentos del Analisis Economico, discussion paper.

Gallant, A. R. and G. Tauchen (1995): "Estimation of Continuos 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. (1994): "Simulation and Numerical Integration", Federal Reserve Bank of Minneapolis, Working Paper No. 526.

Goldstein, H. (1991): "Nonlinear Multilevel Models, with an Application to Discrete Response Data", *Biometrika* 78, 45-51.

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

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.

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.

Longford, N. T. (1994): "Logistic Regression with Random Coefficients", *Journal of Computational Statistics and Data Analysis* 17, 1-15.

McFadden, D. (1989): "A Method of Simulated Method of Moments for Estimations of Discrete Response Models without Numerical Integration", *Econometrica*, 57, 995-1026.

Mealli, F. and C. Rampichini (1999): "Estimating Binary Multilevel Models through Indirect Inference", *Computational Statistics & Data Analysis* 29, 313-324.

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

Monfardini, C. (1998): "Estimating Stochastic Volatility Models through Indirect Inference", *Econometrics Journal* 1, C113-C128.

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.

Nowman, K. B. (1997): "Gaussian Estimation of Single-Factor Continuous Time Models of the Term Structure of Interest Rates", *Journal of Finance* 52, 1695-1706.

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. A. Jr. (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.

Taylor, S. J. (1986): *Modelling Financial Time Series*. Chichester: John Wiley & Sons Ltd.

Taylor, S. J. (1994): "Modelling Stochastic Volatility", *Mathematical Finance* 4, 183-204.

Theil, H. (1971): *Principles of Econometrics*. New York: John Wiley & Sons, Inc.

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.