

**Efficient Importance Sampling Maximum Likelihood Estimation
of Stochastic Differential Equations**

Sergio Pastorello
(Università di Bologna)

Eduardo Rossi
(Università di Pavia)

171 (12-04)

Dipartimento di economia politica
e metodi quantitativi
Università degli studi di Pavia
Via San Felice, 5
I-27100 Pavia

Dicembre 2004

Efficient Importance Sampling Maximum Likelihood Estimation of Stochastic Differential Equations

Sergio Pastorello* Eduardo Rossi†
Università di Bologna Università di Pavia

First draft: November 5, 2004

This draft: December 14, 2004

Abstract

This paper considers ML estimation of a diffusion process observed discretely. Since the exact loglikelihood is generally not available, it must be approximated. We review the most efficient approaches in the literature, and point to some drawbacks. We propose to approximate the loglikelihood using the EIS strategy (Richard and Zhang, 1998), and detail its implementation for univariate homogeneous processes. Some Monte Carlo experiments evaluate its performance against an alternative IS strategy (Durham and Gallant, 2002), showing that EIS is at least equivalent, if not superior, while allowing a greater flexibility needed when examining more complicated models.

JEL codes: C13, C15, C22

Keywords: Diffusion process, Stochastic differential equation, Transition density, Importance sampling, Simulated maximum likelihood

1 Introduction

In the last thirty years, diffusion processes described by stochastic differential equations have become an increasingly common tool used to describe the evolution over time of economic and financial data. Although the process is defined in continuous time, the available data are always sampled in discrete time. This gives rise to an issue when considering the estimation of the parameters of the process.

*Dipartimento di Scienze Economiche, Piazza Scaravilli, 2, 40126 Bologna. Tel: +39 0512098144; fax: +39 0512098040; email: pastorel@economia.unibo.it.

†Dipartimento di Economia Politica e Metodi Quantitativi, Via S.Felice 5, 27100 Pavia. Tel: +39 0382506207; fax: +39 0382304226; email: erossi@eco.unipv.it. Financial support from COFIN 2002 is gratefully acknowledged.

There exist by now several approaches to the estimation of the parameters of stochastic differential equations. Some of traditional strategies provide consistent results only under the assumption that the length of the interval between consecutive observations shrinks to zero. This assumption is usually deemed unappealing for several reasons. First, high frequency observations are not well described by diffusion processes, e.g. because the latter have continuous trajectories, while high frequency financial observations exhibit discontinuous piecewise constant trajectories. Second, reducing the length of the interval between observations introduces in the problem a variety of undesirable microstructure effects. Finally, for some variables (e.g. non financial ones) it may simply be impossible to increase the frequency of observation.

More recently, several alternatives have been developed which provide consistent estimates of the parameters while keeping constant the time distance between observations. Their degree of complexity varies considerably, but traditional efficiency considerations suggest to focus on maximum likelihood estimation. The latter, however, is difficult because in general the transition density function of these processes is not known in closed form, and it has to be approximated somehow. Even in this case, however, different solutions have been proposed in the literature, albeit there is a widespread consensus that two of them are clearly superior in terms of performance: the analytical closed-form expansion put forth by Aït-Sahalia (2002), and the refined simulated importance sampling strategy developed by Durham and Gallant (2002) (see e.g. Jensen and Poulsen, 2002, for a numerical comparison).

Importance sampling is simply a tool which provides an estimate of the value of a high dimensional integral. The main idea is to see the integral as the expected value w.r.t. of a certain auxiliary density of a function defined by the ratio of the original integrand and of the auxiliary density itself. This expectation is then approximated by averaging over simulations drawn from the auxiliary density; hence, it is stochastic in nature. Obviously, the properties of this approximation (such as its unbiasedness and dispersion) depend on tuning parameters such as the number of simulations, but the most important ingredient for a successful implementation of importance sampling is by far the choice of the auxiliary density. Durham and Gallant (2002) have suggested a very simple, yet quite satisfactory, auxiliary density labelled Modified Brownian Bridge.

In this paper, we suggest, as an alternative to MBB, to implement the Efficient Importance Sampling technique developed by Richard and Zhang (1998), which has already been applied successfully in a variety of discrete time models. We start in the next section by briefly reviewing Aït-Sahalia (2002) and Durham and Gallant (2002) strategies. This allows us to fix the notation, and to point to some weak points of these approaches. Although many (but not all) of these issues become apparent only in multivariate models, in this paper we compare the performance of EIS with that of MBB in the same simple univariate setups considered in the literature on the subject. We outline the EIS approach in section 3, and provide in section 4 some details about its implementation in scalar diffusion processes. Finally, section 5 reports the results of

a variety of Monte Carlo experiments comparing EIS to MBB, and exploring issues such as the numerical error in approximating the loglikelihood or in ML estimating the parameters, when the statistical uncertainty intrinsic in the estimation procedure is or is not taken into account. The last section concludes.

2 The problem

In what follows, we consider the scalar parametric SDE:

$$dX_t = \mu(X_t; \boldsymbol{\theta})dt + \sigma(X_t; \boldsymbol{\theta})dW_t \quad (1)$$

where W_t is a standard Brownian Motion. Time homogeneity is assumed only for simplicity; adapting the following discussion to time inhomogeneity requires some straightforward passages. Extensions to multivariate as well as to jump diffusion processes are currently under investigation. We assume that (1) has a nonexploding, unique solution. Explosive solutions are excluded because in their case no transition density exists; notice that stationarity is not required. Among the various alternative sets of sufficient conditions ensuring that this assumption actually holds, the simplest one is that $\mu(\cdot; \boldsymbol{\theta})$ and $\sigma(\cdot; \boldsymbol{\theta})$ satisfy global Lipschitz and linear growth conditions. Karatzas and Shreve (1991) and Aït-Sahalia (2002) discuss alternative sets of sufficient conditions which can be better suited to deal with SDEs which do not meet the previous requirements and are frequently encountered in finance

Let Δ be the length of the interval between two consecutive observations, labelled x_0 and x_Δ , and let $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^p$ denote a vector of unknown parameters. We consider the problem of approximating the exact transition density, denoted by $p(x_\Delta|x_0; \Delta, \boldsymbol{\theta})$. In the following, we adopt a simplified notation in which the dependence of the transition densities (or their logs) on Δ is dropped; the length of the interval between the observations is unambiguously identified by the indexes of the latter. This function is unknown, apart from a few, very simple cases, but its evaluation is essential to compute maximum likelihood estimates of $\boldsymbol{\theta}$ given a sample of discrete observations of the process.

A variety of strategies have been proposed to estimate $\boldsymbol{\theta}$, including simulation-based approaches such as Indirect Inference (Gouriéroux, Monfort and Renault, 1993) or the Efficient Method of Moments (Gallant and Tauchen, 1996), Generalized Method of Moments approaches (Carrasco, Chernov, Florens and Ghysels, 2002; Duffie and Glynn, 2001), nonparametric (Aït-Sahalia, 1996a,b; Bandi and Phillips, 2003; Stanton, 1997) and Bayesian strategies (Eraker, 2001; Jones, 1999), among many others. A few others have been advanced to approximate the unknown transition density, and hence to allow efficient (albeit approximate) maximum likelihood estimation: among these, numerically solving the Fokker-Planck-Kolmogorov partial differential equation (Lo, 1988), closed-form analytic approximation based on Hermite polynomials expansion (Aït-Sahalia, 1999, 2002), and the simulation-based, Monte Carlo integration

strategy suggested by Pedersen (1995) and Brandt and Santa-Clara (2002), and further explored by Durham and Gallant (2002).

The purpose of this paper is to suggest an efficient and very flexible alternative strategy to approximate the transition density of the process. Although the approach we advocate is numerical in nature, it is convenient for future reference to start by briefly outlining the closed-form analytic approximation developed by Aït-Sahalia (2002). A subsequent section will briefly review the literature on the existing numerical approximation techniques.

2.1 Aït-Sahalia (2002) closed-form analytical approximation

Aït-Sahalia (2002) has derived a closed form analytical approximation of the transition density of the SDE (1) using its Hermite polynomials expansion. The approximation can be made arbitrarily good by choosing a sufficiently high order of the expansion, without any need to assume that $\Delta \rightarrow 0$. To guarantee that the expansion actually converges, Aït-Sahalia (2002) has shown that (1) must be transformed into another SDE, with the density sufficiently close to Gaussian. This can be achieved by defining the new variable:

$$Y = \gamma(X; \boldsymbol{\theta}) = \int^X \frac{1}{\sigma(u; \boldsymbol{\theta})} du$$

where the constant of integration is irrelevant. An application of the Itô's Lemma shows that the SDE governing the dynamics of y is given by:

$$dY_t = \mu_Y(Y_t; \boldsymbol{\theta})dt + dW_t \quad (2)$$

where:

$$\mu_Y(y; \boldsymbol{\theta}) = \frac{\mu[\gamma^{-1}(y; \boldsymbol{\theta}); \boldsymbol{\theta}]}{\sigma[\gamma^{-1}(y; \boldsymbol{\theta}); \boldsymbol{\theta}]} - \frac{1}{2} \sigma'[\gamma^{-1}(y; \boldsymbol{\theta}); \boldsymbol{\theta}]$$

This transformation is known as Lamperti transform, and it plays a crucial role in the effectiveness of both the analytical approximation approach of this section Aït-Sahalia (2002), and the numerical IS strategy of Durham and Gallant (2002), to be reviewed in the following one.

Let the Hermite polynomials be defined as

$$H_j(z) = e^{\frac{1}{2}z^2} \frac{d^j}{dz^j} e^{-\frac{1}{2}z^2}, \quad \text{for } j = 0, 1, 2, \dots$$

The Hermite expansion of $p_Y(y_\Delta | y_0; \boldsymbol{\theta})$ is then given by:

$$p_Y^{(J)}(y_\Delta | y_0; \boldsymbol{\theta}) = \Delta^{-1/2} \phi[\Delta^{-1/2}(y_\Delta - y_0)] \sum_{j=0}^J \eta_j(\Delta, y_0, \boldsymbol{\theta}) H_j[\Delta^{-1/2}(y_\Delta - y_0)] \quad (3)$$

where $\phi(\cdot)$ is the density of a $\mathcal{N}(0, 1)$ random variable. The coefficients $\eta_j(\Delta, y_0, \boldsymbol{\theta})$ in (3) are given by the conditional expectation of the corresponding Hermite polynomial. They are

usually unknown, but straightforward to approximate using a Taylor expansion based on the infinitesimal generator of the process:

$$\eta_z^{(j)}(\Delta, y_0, \boldsymbol{\theta}) = \frac{1}{j!} \mathbb{E}\{H_j[\Delta^{-1/2}(y_\Delta - y_0)]|y_0; \boldsymbol{\theta}\} \approx \frac{1}{j!} \sum_{k=0}^K \frac{\Delta^k}{k!} \mathcal{A}^k H_j[\Delta^{-1/2}(y_\Delta - y_0)] \Big|_{y_\Delta=y_0}$$

where, for any function $g(\cdot)$ (subject to some regularity conditions):

$$\mathcal{A}g(y) = \mu_Y(y; \boldsymbol{\theta}) \frac{\partial g(y)}{\partial y} + \frac{1}{2} \frac{\partial^2 g(y)}{\partial y^2}$$

When this approximation is replaced in (3), the result is a double expansion $p_Y^{(J,K)}(y_\Delta|y_0; \boldsymbol{\theta})$ of the unknown transition density. By collecting together terms involving the same powers of Δ , it is possible to let $J \rightarrow \infty$, and to obtain an approximation which only depends on K . Furthermore, the terms of the Taylor expansion in Δ of the log of this expression can be collected in much the same way in order to derive a similar expansion for the log of the transition density given by:

$$\ell_Y^{(K)}(y_\Delta|y_0; \boldsymbol{\theta}) = -\frac{1}{2} \log(2\pi\Delta) + \frac{C_Y^{(-1)}(y_\Delta|y_0; \boldsymbol{\theta})}{\Delta} + \sum_{k=0}^K C_Y^{(k)}(y_\Delta|y_0; \boldsymbol{\theta}) \frac{\Delta^k}{k!} \quad (4)$$

The most convenient way to obtain the expression of the coefficients in (4) is to consider the Kolmogorov forward and backward equations for ℓ_Y , respectively given by:

$$\begin{aligned} \frac{\partial \ell_Y(y_\Delta|y_0; \boldsymbol{\theta})}{\partial \Delta} &= -\frac{\partial \mu_Y(y_\Delta)}{\partial y_\Delta} - \mu_Y(y_\Delta) \frac{\partial \ell_Y(y_\Delta|y_0; \boldsymbol{\theta})}{\partial y_\Delta} + \frac{1}{2} \frac{\partial^2 \ell_Y(y_\Delta|y_0; \boldsymbol{\theta})}{\partial y_\Delta^2} + \frac{1}{2} \left[\frac{\partial \ell_Y(y_\Delta|y_0; \boldsymbol{\theta})}{\partial y_\Delta} \right]^2 \\ \frac{\partial \ell_Y(y_\Delta|y_0; \boldsymbol{\theta})}{\partial \Delta} &= \mu_Y(y_0) \frac{\partial \ell_Y(y_\Delta|y_0; \boldsymbol{\theta})}{\partial y_0} + \frac{1}{2} \frac{\partial^2 \ell_Y(y_\Delta|y_0; \boldsymbol{\theta})}{\partial y_0^2} + \frac{1}{2} \left[\frac{\partial \ell_Y(y_\Delta|y_0; \boldsymbol{\theta})}{\partial y_0} \right]^2 \end{aligned}$$

By substituting (4) in these expressions and collecting terms in powers of Δ , we obtain a sequence of differential equations in the coefficients $C_Y^{(k)}(y_\Delta|y_0; \boldsymbol{\theta})$ which can be solved explicitly. See Aït-Sahalia (2003) for more details. Finally, given $\ell_Y^{(K)}(y_\Delta|y_0; \boldsymbol{\theta})$, a sequence of approximations of the log transition density $\ell(x_\Delta|x_0; \boldsymbol{\theta})$ of the original X process can be constructed with an application of the Jacobian formula.

The previous Hermite expansion requires to compute the Lamperti transform of the original diffusion process. This task, however, is not always straightforward. Bakshi and Ju (2003), for example, point out that for some choices of the diffusion coefficient $\sigma(\cdot; \boldsymbol{\theta})$ the integral in $\gamma(\cdot; \boldsymbol{\theta})$ or its inverse $\gamma^{-1}(\cdot; \boldsymbol{\theta})$ may not be available analytically. Nevertheless, in the univariate case these quantities can always be evaluated using numerical quadrature methods at the cost of increasing the computational burden. It should be noticed, however, that for many interesting multivariate diffusions the Lamperti transform may not exist at all (see Aït-Sahalia, 2003, Proposition 1). Processes with this property are said to be irreducible.

An alternative way to overcome the unavailability of a closed-form expression for the Lamperti transform of the process is suggested by Aït-Sahalia (2003). Basically, the idea is to

postulate a functional form for the expansion for $\ell(x_\Delta|x_0; \boldsymbol{\theta})$, and then compute its coefficients using the Kolmogorov forward and backward equations, as seen above. By analogy with (4), a convenient assumption is:

$$\ell^{(K)}(x_\Delta|x_0; \boldsymbol{\theta}) = -\frac{1}{2} \log(2\pi\Delta) - \frac{1}{2} \log \sigma^2(x_\Delta; \boldsymbol{\theta}) + \frac{C_X^{(-1)}(x_\Delta|x_0; \boldsymbol{\theta})}{\Delta} + \sum_{k=0}^K C_X^{(k)}(x_\Delta|x_0; \boldsymbol{\theta}) \frac{\Delta^k}{k!} \quad (5)$$

where the last term is the log of the Jacobian of the Lamperti transform, which is obviously missing from (4). Since generally the coefficients in (5) do not have closed-form expressions, it is convenient to approximate them using a Taylor expansion w.r.t. x_Δ in x_0 . The coefficients of these expansions can be identified by substituting (5) in the Kolmogorov equations of the process, and separately considering the coefficients of increasing powers in Δ . From a computational point of view, this requires to solve a sequence of systems of equations; an explicit solution can be obtained using symbolic mathematical software.

Formulas (4) and (5) provide two alternative closed-form approximations for the log of the transition density of the univariate diffusion (1). While it can be shown that the former converges for fixed Δ to the true loglikelihood as the order of approximation K increases to infinity, for the latter to converge it is necessary to assume that $\Delta \rightarrow 0$. Another drawback is that frequently (depending on the SDE under scrutiny) both approaches yield loglikelihood approximations which are polynomial functions of a subset of parameters. As an example, assume that $\boldsymbol{\theta}$ can be split in two subsets $(\boldsymbol{\theta}'_\mu, \boldsymbol{\theta}'_\sigma)'$ whose elements respectively appear in the drift and the diffusion coefficients, and that the former is a polynomial in $\boldsymbol{\theta}_\mu$. Using the formulae for the reducible case in Aït-Sahalia (2003), it is then easy to check that $\mu_Y(\cdot; \boldsymbol{\theta})$ and $C_Y(k)(y_\Delta|y_0; \boldsymbol{\theta})$, for $k = 0, \dots, K$, are also polynomials in $\boldsymbol{\theta}_\mu$. This property may be troublesome when some of the parameters for which it is valid can increase to infinity, as in this case the loglikelihood approximation may grow unbounded.

Furthermore, both approximations may not be well suited to estimate models in which the X process is not observed directly, but rather only through a one-to-one transform which also depends on $\boldsymbol{\theta}$. This framework characterizes applications which use the price of a derivative asset to recover an otherwise unobservable state variable, such as stochastic volatility option pricing models or term structure models driven by unobservable factors, such as the affine class (see Aït-Sahalia and Kimmel, 2003, for an application of the latter kind). The reason for this is that the coefficients of the various powers of Δ in (4) and (5) may become singular for specific values of x_Δ and x_0 , e.g. when $x_\Delta \rightarrow 0$, $x_0 \rightarrow 0$ or $x_\Delta \rightarrow x_0$, depending on the diffusion process under scrutiny. If direct observations for both quantities are available, this is not an issue (apart from the latter case, which may be encountered due to tick pricing procedures commonly adopted in financial markets), but when they are not, the loglikelihood approximation becomes infinite for any value of $\boldsymbol{\theta}$ which sets the implied values of the latent variables equal to some singularity point. Finally, no extension of this approach to the case of multivariate SDE with partial observability of the state vector is available as of this writing. For example, Aït-Sahalia

and Kimmel (2004) estimate a continuous time stochastic volatility model, but are forced to approximate the unobservable volatility state using the implied volatility of a short dated at-the-money option.

2.2 Approximations based on Importance Sampling integration

In this paper, we shall focus on loglikelihood approximations based a numerical approach which evaluates the unknown transition density of the process using Monte Carlo Importance Sampling (IS hereafter) techniques. To briefly outline the strategy, start by considering an approximate discretization scheme of the SDE (1), and denote with $p^a(x_\Delta|x_0; \boldsymbol{\theta})$ the transition density implied by this choice. For example, for the Euler scheme we know that:

$$p^a(x_\Delta|x_0; \boldsymbol{\theta}) = \phi[x_\Delta; x_0 + \mu(x_0; \boldsymbol{\theta})\Delta, \sigma(x_0; \boldsymbol{\theta})^2\Delta] \quad (6)$$

where $\phi(\cdot; m, v^2)$ denotes the pdf of a Gaussian r.v. with expected value m and variance v^2 . It is well known that $p^a(x_\Delta|x_0; \boldsymbol{\theta})$ provides an acceptable approximation to $p(x_\Delta|x_0; \boldsymbol{\theta})$ only when Δ is sufficiently small, which is often not the case for most available samples. Let us consider a partition of the interval $[0, \Delta]$ in M subintervals delimited by $0 = \tau_0 < \tau_1 < \dots < \tau_M = \Delta$, and let for simplicity $x_m = x_{\tau_m}$. Notice that $x_M = x_\Delta$. Without losing in generality, and in order to simplify the notation, we assume that $\tau_m - \tau_{m-1} = \Delta/M = \delta$. Since, by the Chapman-Kolmogorov property:

$$p(x_M|x_0; \boldsymbol{\theta}) = \int_{\mathbb{R}^{M-1}} \prod_{m=1}^M p(x_m|x_{m-1}; \boldsymbol{\theta}) dx_1 \dots dx_{M-1}$$

we can approximate $p(x_M|x_0; \boldsymbol{\theta})$ with:

$$p_M(x_M|x_0; \boldsymbol{\theta}) = \int_{\mathbb{R}^{M-1}} \prod_{m=1}^M p^a(x_m|x_{m-1}; \boldsymbol{\theta}) dx_1 \dots dx_{M-1} \quad (7)$$

where, following Durham and Gallant (2002, Assumption 4), we implicitly assume that this integral exists. Under this condition, the approximation can be made as precise as desired by simply increasing the number of subintervals M . Sufficient conditions ensuring that $p_M(x_M|x_0; \boldsymbol{\theta})$ exhibits some desirable theoretical properties (existence and convergence towards $p(x_M|x_0; \boldsymbol{\theta})$ as $M \rightarrow \infty$) have been advanced by Pedersen (1995). The major difficulty in this approach is represented by the evaluation of the $(M-1)$ -dimensional integral in (7). Monte Carlo integration is in general the only feasible approach. Its most basic implementation, as suggested by Pedersen (1995) and Brandt and Santa-Clara (2002) is based on the observation that (7) can be reformulated as the expectation of $p^a(x_M|x_{M-1}; \boldsymbol{\theta})$ over x_{M-1} and with respect to the distribution of the latter induced by $\prod_{m=1}^{M-1} p^a(x_m|x_{m-1}; \boldsymbol{\theta})$, which is a pdf on \mathbb{R}^{M-1} . Let $\{(\tilde{x}_1^{(s)}, \dots, \tilde{x}_{M-1}^{(s)}), s = 1, \dots, S\}$ be S independent trajectories drawn from this density. Notice that by construction all these simulated trajectories start at x_0 . The integral in (7) can now

be approximated with:

$$\tilde{p}_M^{(S)}(x_M|x_0; \boldsymbol{\theta}) = \frac{1}{S} \sum_{s=1}^S p^a(x_M|\tilde{x}_{M-1}^{(s)}; \boldsymbol{\theta}) \quad (8)$$

Durham and Gallant (2002) have pointed out that this strategy is in general highly inefficient, because the probability that a draw $\tilde{x}_{M-1}^{(s)}$ from $\prod_{m=1}^{M-1} p^a(x_m|x_{m-1}; \boldsymbol{\theta})$ belongs to the region effectively contributing to the value of $p_M(x_M|x_0; \boldsymbol{\theta})$ is infinitesimal, so that (8) is almost always heavily downward biased and affected by a huge variance. Although there are techniques that can improve the numerical accuracy of natural MC estimates (such as antithetic and control variates), these are often unable to compensate for the initial selection of an inherently inefficient sampler.

For these reasons, Durham and Gallant (2002) explored a number of alternatives based on IS. To illustrate this approach, let $r(x_1, \dots, x_{M-1})$ denote an auxiliary pdf on \mathbb{R}^{M-1} , and rewrite (7) as:

$$p_M(x_M|x_0; \boldsymbol{\theta}) = \int_{\mathbb{R}^{M-1}} \frac{\prod_{m=1}^M p^a(x_m|x_{m-1}; \boldsymbol{\theta})}{r(x_1, \dots, x_{M-1})} r(x_1, \dots, x_{M-1}) dx_1 \dots dx_{M-1} \quad (9)$$

If we now let $\{(\tilde{x}_1^{(s)}, \dots, \tilde{x}_{M-1}^{(s)}), s = 1, \dots, S\}$ be S independent trajectories drawn from r , an alternative approximation of $p_M(x_M|x_0; \boldsymbol{\theta})$ can be obtained using:

$$\tilde{p}_M^{(S)}(x_M|x_0; \boldsymbol{\theta}) = \frac{1}{S} \sum_{s=1}^S \frac{\prod_{m=1}^M p^a(\tilde{x}_m^{(s)}|\tilde{x}_{m-1}^{(s)}; \boldsymbol{\theta})}{r(\tilde{x}_1^{(s)}, \dots, \tilde{x}_{M-1}^{(s)})} \quad (10)$$

Geweke (1996) shows that $\tilde{p}_M^{(S)}(x_M|x_0; \boldsymbol{\theta})$ converges towards $p_M(x_M|x_0; \boldsymbol{\theta})$ by a Strong Law of Large Numbers. To assess the accuracy of the IS approximation (10), a \sqrt{S} -Central Limit Theorem can be used if:

$$\int_{\mathbb{R}^{M-1}} \frac{\prod_{m=1}^M p^a(x_m|x_{m-1}; \boldsymbol{\theta})^2}{r(x_1, \dots, x_{M-1})} dx_1 \dots dx_{M-1} < \infty \quad (11)$$

which ensures that (10) has a finite variance. It is easy to check that condition (11) holds if

$$\left| \frac{\prod_{m=1}^M p^a(x_m|x_{m-1}; \boldsymbol{\theta})}{r(x_1, \dots, x_{M-1})} \right| < \infty \quad (12)$$

on the domain of integration. Hence, it is essential to choose a sampling density $r(x_1, \dots, x_{M-1})$ such that its tails do not decline faster than those of $\prod_{m=1}^M p^a(x_m|x_{m-1}; \boldsymbol{\theta})$.

Apart from the choice of M and S , the IS strategy requires the specification of the approximate density $p^a(x_m|x_{m-1}; \boldsymbol{\theta})$, hereafter labelled “subtransition density”, and the sampling density $r(x_1, \dots, x_{M-1})$. Durham and Gallant (2002) have shown that the former component mainly affects the bias of $\tilde{p}_M^{(S)}(x_M|x_0; \boldsymbol{\theta})$, while the latter is connected to its Monte Carlo variance. After intensive investigations on a large number of possible choices, and focussing on a specific benchmark process (a square root process with parameters chosen to fit the time series behavior of US interest rate data), they come to the following twofold conclusion.

First, the most efficient bias reduction technique is based on a two step procedure. The first one is the Lamperti transform discussed in the previous section. The second one provides the approximation of the transition pdf of y on the subintervals (τ_{m-1}, τ_m) , for $m = 1, \dots, M$. Durham and Gallant (2002) suggest to apply to the transformed SDE the Shoji and Ozaki (1998) local linearization method, which approximates (2) with an Ornstein-Uhlenbeck process, characterized by a readily available Gaussian transition density:

$$p_{s\&o}(y_m|y_{m-1}; \boldsymbol{\theta}) = \phi[y_m; y_{m-1} + \mu_{s\&o}(y_{m-1}; \boldsymbol{\theta}), \sigma_{s\&o}^2(y_{m-1}; \boldsymbol{\theta})]$$

with:

$$\begin{aligned} \mu_{s\&o}(y_{m-1}; \boldsymbol{\theta}) &= \frac{\mu_Y(y_{m-1}; \boldsymbol{\theta})\kappa_{m-1}}{\mu'_Y(y_{m-1}; \boldsymbol{\theta})} + \frac{\mu''_Y(y_{m-1}; \boldsymbol{\theta})}{2[\mu'_Y(y_{m-1}; \boldsymbol{\theta})]^2} [\kappa_{m-1} - \mu'_Y(y_{m-1}; \boldsymbol{\theta})\delta] \\ \sigma_{s\&o}^2(y_{m-1}; \boldsymbol{\theta}) &= \frac{\exp[2\mu'_Y(y_{m-1}; \boldsymbol{\theta})\delta] - 1}{2\mu'_Y(y_{m-1}; \boldsymbol{\theta})} \\ \kappa_{m-1} &= \exp[\mu'_Y(y_{m-1}; \boldsymbol{\theta})\delta] - 1 \end{aligned}$$

Second, the most efficient variance reduction technique is based on an auxiliary density labelled Modified Brownian Bridge, in which each point of the simulated trajectories are drawn from a Gaussian density which approximates the pdf of x_m given x_{m-1} and x_M , given by:

$$p_{mbb}(x_m|x_M, x_{m-1}; \boldsymbol{\theta}) = \phi[x_m|x_{m-1} + \mu_{mbb}(x_{m-1}), \sigma_{mbb}^2(x_{m-1}; \boldsymbol{\theta})] \quad (13)$$

with:

$$\mu_{mbb}(x_{m-1}) = \frac{x_M - x_{m-1}}{\Delta - \tau_{m-1}} \delta \quad \text{and} \quad \sigma_{mbb}^2(x_{m-1}; \boldsymbol{\theta}) = \frac{M - m}{M - m + 1} \sigma^2(x_{m-1}; \boldsymbol{\theta}) \delta$$

This strategy is quite efficient for the benchmark process considered by Durham and Gallant (2002), but its performance in different contexts, such as those characterized by a volatility higher than that usually observed in interest rate processes, is still unknown. Moreover, in extending it to multivariate contexts, it should be noted that not all multivariate diffusions can be transformed in such a way that the transformed process has a constant diffusion process (Aït-Sahalia, 2003), so that the first point in the Durham and Gallant (2002) strategy loses its generality. An alternative bias reduction strategy is based on extrapolation, which is a well known technique to numerically evaluate conditional expectations of diffusion processes with higher order accuracy (see Kloeden and Platen, 1992, chapter 15). Extrapolation techniques are usually very effective in removing the bias, but at the cost of a large increase in the variance. For this reason Durham and Gallant (2002) do not suggest its use for univariate diffusions.

Finally, an extension to multivariate settings with latent variables is even more problematic, because the unavailability of observations of the latter makes it strictly impossible to apply the MBB based variance reduction technique. For example, to extend the MBB approach to the case of latent variables, Durham and Gallant (2002) and Durham (2003) suggest to (i) use the Pedersen (1995) and Brandt and Santa-Clara (2002) “blind” approach outlined earlier to

draw the trajectories of the latent components of the process, and conditionally on these (ii) use the MBB sampler for the observed components. Although the Monte Carlo experiments in Durham and Gallant (2002) are encouraging, it is not clear whether such a strategy can represent a general solution, and in particular whether it effectively overcomes the critiques originally advanced to the basic Pedersen (1995) and Brandt and Santa-Clara (2002) sampling strategy. In this paper, though, we are only concerned with the univariate case; extensions to the multivariate case with partial observability are currently under investigation.

In the following section we shall briefly outline an alternative and very promising variance reduction technique labelled Efficient IS in the context of the numerical evaluation of the integral (8). In a later section, we will investigate its properties and compare them to the preferred Durham and Gallant (2002) strategy.

3 Efficient Importance Sampling

The Efficient Importance Sampling (hereafter EIS) procedure has been proposed by Richard and Zhang (1998) and applied to maximum likelihood estimation of the parameters of a variety of settings, including dynamic latent variable models in discrete time (Liesenfeld and Richard, 2003a,b), and dynamic discrete choice panel models (Zhang and Lee, 2004). EIS is essentially a strategy to build a sampling density $r(x_1, \dots, x_{M-1})$ containing a huge number of parameters and ideally suited to provide extremely accurate estimates of some high dimensional integrals. Moreover, under some reasonable approximating hypothesis, the strategy exploits a recursive decomposition of the original problem, and it can be described as a recursive sequence of auxiliary low-dimensional least squares problems. For a thorough presentation of the EIS strategy in its most general form, see Richard and Zhang (1998). In this section, we illustrate the basic idea underlying the EIS algorithm in the context of continuous time diffusion processes like (1). For the sake of simplicity, we consider two versions of the algorithm, labelled “one-step EIS” and “sequential EIS”.

3.1 One-step EIS

To simplify the notation, let $\boldsymbol{\lambda} = (x_1, \dots, x_{M-1})'$ denote the $(M - 1) \times 1$ vector of integration variables in the target integral (7). $\boldsymbol{\lambda}$ should be interpreted as “latent” variables, while the observable variables are x_0 and x_M . To start with, EIS requires the selection of a parametric class of sampling densities, among which the optimal one (in a sense to be specified below) will be chosen. In this respect, it is clearly important that this family exhibits sufficient flexibility, but also that condition (11) holds. In practice, the family of candidate sampling densities is usually dictated by the problem under scrutiny; for example, it may consist of straightforward and/or mathematically convenient parametric extensions of the distribution of the latent variables given the past observable variables. The latter is labelled “natural sampler” by Richard and Zhang

(1998), and is usually provided by the model itself; in the framework we are considering, it is given by $p^a(\boldsymbol{\lambda}|x_0; \boldsymbol{\theta}) = \prod_{m=1}^{M-1} p^a(x_m|x_{m-1}; \boldsymbol{\theta})$. Let $\mathcal{R} = \{r(\boldsymbol{\lambda}; \mathbf{a}); \mathbf{a} \in A\}$ denote the class of auxiliary sampling densities indexed by a parameter vector $\mathbf{a} \in A$. Finally, let $p^a(x_M, \boldsymbol{\lambda}|x_0; \boldsymbol{\theta}) = \prod_{m=1}^M p^a(x_m|x_{m-1}; \boldsymbol{\theta})$ denote the joint distribution of the latent variables and of the forward observable variable x_M conditional on the backward observable variable x_0 and the parameters. The integral in (9) can then be rewritten as:

$$p_M(x_M|x_0; \boldsymbol{\theta}) = \int \frac{p^a(x_M, \boldsymbol{\lambda}|x_0; \boldsymbol{\theta})}{r(\boldsymbol{\lambda}; \mathbf{a})} r(\boldsymbol{\lambda}; \mathbf{a}) d\boldsymbol{\lambda} \quad (14)$$

where the integral is taken over the support of $\boldsymbol{\lambda}$ (omitted for simplicity).

Suppose now for a moment that there exists a value $\mathbf{a}(x_M, x_0; \boldsymbol{\theta})$ of the auxiliary parameters such that:

$$\forall \boldsymbol{\lambda}, \quad p^a(x_M, \boldsymbol{\lambda}|x_0; \boldsymbol{\theta}) \propto r[\boldsymbol{\lambda}; \mathbf{a}(x_M, x_0; \boldsymbol{\theta})] \quad (15)$$

If this property holds, then the correct result can be computed with zero variance, because substitution of (15) into (14) immediately yields that $p_M(x_M|x_0; \boldsymbol{\theta})$ equals the proportionality factor in (15), since $r(\boldsymbol{\lambda}; \mathbf{a})$ is a density w.r.t. $\boldsymbol{\lambda}$ for every \mathbf{a} . By consequence, it must necessarily be that:

$$r[\boldsymbol{\lambda}; \mathbf{a}(x_M, x_0; \boldsymbol{\theta})] = p^a(\boldsymbol{\lambda}|x_M, x_0; \boldsymbol{\theta}) = \prod_{m=1}^{M-1} p^a(x_m|x_M, x_{m-1}; \boldsymbol{\theta})$$

This observation intuitively explains why Durham and Gallant (2002) variance reduction strategy works so well. Each conditional distribution $p^a(x_m|x_M, x_{m-1}; \boldsymbol{\theta})$ is generally unknown, but it can be approximated by $p_{mbb}(x_m|x_M, x_{m-1}; \boldsymbol{\theta})$ defined in (13). The small but generally nonzero approximation error introduces some Monte Carlo variance in the IS evaluation of (14), but this strategy is overall quite efficient in the simple benchmark example considered by Durham and Gallant (2002).

EIS suggests an alternative strategy to make (15) hold as closely as possible. Let $\{\tilde{\boldsymbol{\lambda}}^{(s)}, s = 1, \dots, S\}$ denote S independent trajectories drawn from $r(\boldsymbol{\lambda}; \mathbf{a})$. The idea is simply to estimate $\mathbf{a}(x_M, x_0; \boldsymbol{\theta})$ by solving the following least squares problem:

$$\min_{c, \mathbf{a}} \sum_{s=1}^S \left[\log p^a(x_M, \tilde{\boldsymbol{\lambda}}^{(s)}|x_0; \boldsymbol{\theta}) - c - \log r(\tilde{\boldsymbol{\lambda}}^{(s)}; \mathbf{a}) \right]^2 \quad (16)$$

An alternative, and probably superior strategy would be to consider a weighted least squares objective function:

$$\min_{c, \mathbf{a}} \sum_{s=1}^S \left[\log p^a(x_M, \tilde{\boldsymbol{\lambda}}^{(s)}|x_0; \boldsymbol{\theta}) - c - \log r(\tilde{\boldsymbol{\lambda}}^{(s)}; \mathbf{a}) \right]^2 \frac{p^a(x_M, \tilde{\boldsymbol{\lambda}}^{(s)}|x_0; \boldsymbol{\theta})}{r(\tilde{\boldsymbol{\lambda}}^{(s)}; \mathbf{a})} \quad (17)$$

because the inclusion of weights allows to focus the estimation of c and \mathbf{a} on the regions of values of $\boldsymbol{\lambda}$ for which the ratio $p^a(x_M, \boldsymbol{\lambda}|x_0; \boldsymbol{\theta})/r(\boldsymbol{\lambda}; \mathbf{a})$ is higher, thus contributing more to the target integral. Problem (17) is precisely the one suggested by Richard and Zhang (1998), although

their discussion is based on the inspection of an approximation of the Monte Carlo variance of the IS estimate of (14). In the following, we shall consider the simpler problem (16), but many of the following observations carry over to (17) without difficulty. It is also important to notice that, independently from the objective function, the approximation of the integrand provided by EIS is *global* in nature, as opposed to methods providing local approximations, such as, among others, Taylor series expansions and Laplace approximations. This is the main reason why this technique has proven so successful in empirical applications.

The derivation above suggests that the estimate of the intercept parameter c is also an estimate of $\log p_M(x_M|x_0; \boldsymbol{\theta})$. It should be noted, however, that in (16) the estimates of the parameters c and \mathbf{a} are based on observations which themselves depend on \mathbf{a} , being IID draws from $r(\boldsymbol{\lambda}; \mathbf{a})$. An additional difficulty is given by the fact that the candidate sampling density $r(\boldsymbol{\lambda}; \mathbf{a})$ is generally a (possibly complicated) nonlinear function of \mathbf{a} . Both features are annoying, because they make problem (16) a highly nonlinear one. To circumvent the first issue, it is possible to consider the following iterative strategy. Given an initial sampling density, draw the S trajectories $\{\tilde{\boldsymbol{\lambda}}^{(s)}, s = 1, \dots, S\}$, and use them to compute the first step estimates $\hat{c}_{(1)}$ and $\hat{\mathbf{a}}_{(1)}$. Now use the fitted sampling density $r(\boldsymbol{\lambda}; \hat{\mathbf{a}}_{(1)})$ to redraw the trajectories of the latent variables, and reestimate the parameters on them. These steps might be iterated until convergence in \mathbf{a} is attained, but in practice very few iterations are sufficient to obtain an extremely low objective function, independently from the initial sampling density. An important practical point is that the simulated trajectories must be based on the same set of random numbers. This is necessary both for the convergence of the iterative procedure illustrated above, and for the maximization of an objective function (such as a loglikelihood, or a quadratic form built from a set of moments) computed by simulation, and for the convergence of such estimates for fixed Monte Carlo sizes.

To overcome the issue related to the nonlinearity of problem (16) in \mathbf{a} , let:

$$r(\boldsymbol{\lambda}; \mathbf{a}) = \frac{r_K(\boldsymbol{\lambda}; \mathbf{a})}{\rho(\mathbf{a})}, \quad \text{with } \rho(\mathbf{a}) = \int r_K(\boldsymbol{\lambda}; \mathbf{a}) d\boldsymbol{\lambda}$$

where $r_K(\boldsymbol{\lambda}; \mathbf{a})$ and $\rho(\mathbf{a})$ denote the kernel and the associated constant of integration of $r(\boldsymbol{\lambda}; \mathbf{a})$, respectively. Using these definitions, problem (16) can be rewritten as:

$$\min_{c, \mathbf{a}} \sum_{s=1}^S \left[\log p^a(x_M, \tilde{\boldsymbol{\lambda}}^{(s)}|x_0, \boldsymbol{\theta}) - c - \log r_K(\tilde{\boldsymbol{\lambda}}^{(s)}; \mathbf{a}) \right]^2 \quad (18)$$

where the integrating constant $\log \rho(\mathbf{a})$ has been absorbed by c . Notice also that any multiplicative factor in $p^a(x_M, \boldsymbol{\lambda}|x_0; \boldsymbol{\theta})$ which does not depend on either $\boldsymbol{\lambda}$ may be collected in the intercept coefficient, although we do not pursue this possibility here to keep the notation as simple as possible. This version of the problem allows to estimate \mathbf{a} using only the kernel $r_K(\boldsymbol{\lambda}; \mathbf{a})$ of the auxiliary sampling density $r(\boldsymbol{\lambda}; \mathbf{a})$. For most distributions, the former is a much simpler function of the parameters \mathbf{a} than the latter. Moreover, if $r(\boldsymbol{\lambda}; \mathbf{a})$ belongs to the exponential family, then $\log r_K(\boldsymbol{\lambda}; \mathbf{a})$ can be written as a linear function of \mathbf{a} , in which case (18) becomes

a *linear* least squares problem. Finally, the efficient IS estimate of $p_M(x_M|x_0; \boldsymbol{\theta})$ can be easily derived by applying (10), with auxiliary density given by $r(\boldsymbol{\lambda}; \hat{\mathbf{a}})$.

3.2 Sequential EIS

The main drawback of the strategy just outlined is that it requires to simultaneously estimate all the parameters \mathbf{a} appearing in the sampling density of a $(M-1)$ -dimensional vector of latent variables, $\boldsymbol{\lambda}$. For $r(\boldsymbol{\lambda}; \mathbf{a})$ to exhibit the degree of flexibility necessary to closely approximate the target integrand $p^a(x_M, \boldsymbol{\lambda}|x_0; \boldsymbol{\theta})$, a large number of parameters in \mathbf{a} is generally required. Simultaneous estimation of all the elements of \mathbf{a} is not impossible, but it would likely require restrictive hypothesis on the class of candidate sampling densities, such as multivariate normality, which would largely limit the effectiveness of EIS and would probably violate condition (11). In contrast, a recursive strategy could provide a much more flexible approach.

To illustrate this point, we consider the following factorization of the sampling density:

$$r(\boldsymbol{\lambda}; \mathbf{a}) = \prod_{m=1}^{M-1} r(x_m|x_{m-1}; \mathbf{a}_m) \quad (19)$$

where $\mathbf{a} = (\mathbf{a}'_1, \dots, \mathbf{a}'_{M-1})'$. We also assume that, for all m :

$$r(x_m|x_{m-1}; \mathbf{a}_m) = \frac{r_K(x_m; x_{m-1}, \mathbf{a}_m)}{\rho(x_{m-1}, \mathbf{a}_m)}, \quad \text{where } \rho(x_{m-1}, \mathbf{a}_m) = \int r_K(x_m; x_{m-1}, \mathbf{a}_m) dx_m$$

We start by illustrating the main points of the discussion in a simple example, corresponding to $M = 3$. A subsequent section extends the results to a generic higher value of M , and outlines the recursive algorithm.

3.2.1 Two-steps sequential EIS

If $M = 3$, the target integral may be rewritten as:

$$\begin{aligned} p_3(x_3|x_0; \boldsymbol{\theta}) &= \int \frac{\prod_{m=1}^3 p^a(x_m|x_{m-1}; \boldsymbol{\theta})}{\prod_{m=1}^2 r(x_m|x_{m-1}; \mathbf{a}_m)} \prod_{m=1}^2 r(x_m|x_{m-1}; \mathbf{a}_m) dx_2 dx_1 \\ &= \int h_2(x_3, x_1; \boldsymbol{\theta}, \mathbf{a}_2) \frac{p^a(x_1|x_0; \boldsymbol{\theta}) \rho(x_1, \mathbf{a}_2)}{r(x_1|x_0; \mathbf{a}_1)} r(x_1|x_0; \mathbf{a}_1) dx_1 \end{aligned} \quad (20)$$

where:

$$h_2(x_3, x_1; \boldsymbol{\theta}, \mathbf{a}_2) = \int \frac{p^a(x_3, x_2|x_1; \boldsymbol{\theta})}{r_K(x_2; x_1, \mathbf{a}_2)} r(x_2|x_1; \mathbf{a}_2) dx_2 = \mathbb{E} \left[\frac{p^a(x_3, x_2|x_1; \boldsymbol{\theta})}{r_K(x_2; x_1, \mathbf{a}_2)} \middle| x_1 \right]$$

where the expectation is taken w.r.t. $r(x_2|x_1; \mathbf{a}_2)$. As in the previous section, let us suppose that it is possible to find a value $\mathbf{a}_2(x_3, \boldsymbol{\theta})$ such that:

$$\forall x_1, x_2, \quad p^a(x_3, x_2|x_1; \boldsymbol{\theta}) \propto r_K[x_2; x_1, \mathbf{a}_2(x_3, \boldsymbol{\theta})] \quad (21)$$

If (21) is valid, then the inner integral can be computed exactly. As this is not generally possible, a strategy similar to that outlined in the previous paragraph can be followed. Given S independent trajectories $\{\tilde{\boldsymbol{\lambda}}^{(s)}, s = 1, \dots, S\}$, $\mathbf{a}_2(x_3, \boldsymbol{\theta})$ can be estimated by minimizing the following sum of squares:

$$\min_{c_2, \mathbf{a}_2} \sum_{s=1}^S \left[\log p^a(x_3, \tilde{x}_2^{(s)} | \tilde{x}_1^{(s)}; \boldsymbol{\theta}) - c_2 - \log r_K(\tilde{x}_2^{(s)}; \tilde{x}_1^{(s)}, \mathbf{a}_2) \right]^2 \quad (22)$$

which, under a suitable choice for \mathcal{R} , can be characterized as a linear least squares problem. Let us denote with $\hat{\mathbf{a}}_2(x_3, \boldsymbol{\theta})$ the solution of (22) w.r.t. \mathbf{a}_2 . Using this value, we can estimate the inner integral in (20) using the standard IS formula:

$$\hat{h}_2[x_3, x_1; \boldsymbol{\theta}, \hat{\mathbf{a}}_2(x_3, \boldsymbol{\theta})] = \frac{1}{S} \sum_{s=1}^S \frac{p^a(x_3, \tilde{x}_2^{(s)} | \tilde{x}_1^{(s)}; \boldsymbol{\theta})}{r_K[\tilde{x}_2^{(s)}; \tilde{x}_1^{(s)}, \hat{\mathbf{a}}_2(x_3, \boldsymbol{\theta})]}$$

where each $\tilde{x}_2^{(s)}$ is drawn independently from $r[x_2 | \tilde{x}_1^{(s)}; \hat{\mathbf{a}}_2(x_3, \boldsymbol{\theta})]$. Let us assume that $\hat{\mathbf{a}}_2(x_3, \boldsymbol{\theta})$ makes (21) approximately valid, so that $\hat{h}_2[x_3, x_1; \boldsymbol{\theta}, \hat{\mathbf{a}}_2(x_3, \boldsymbol{\theta})]$ is extremely precise:

$$\text{Var} \left[\hat{h}_2[x_3, x_1; \boldsymbol{\theta}, \hat{\mathbf{a}}_2(x_3, \boldsymbol{\theta})] \middle| x_1 \right] = \frac{1}{S} \text{Var} \left[\frac{p^a(x_3, x_2 | x_1; \boldsymbol{\theta})}{r_K[x_2; x_1, \hat{\mathbf{a}}_2(x_3, \boldsymbol{\theta})]} \middle| x_1 \right] \approx 0 \quad (23)$$

Let us now examine the evaluation of the outer integral in (20) using these results. An IS estimate based on S independent trajectories has variance equal to:

$$\text{Var}[\tilde{p}_3^{(S)}(x_3 | x_0; \boldsymbol{\theta})] = \frac{1}{S} \text{Var} \left[\frac{p^a(x_3, x_2 | x_1; \boldsymbol{\theta})}{r_K[x_2; x_1, \hat{\mathbf{a}}_2(x_3, \boldsymbol{\theta})]} \frac{p^a(x_1 | x_0; \boldsymbol{\theta}) \rho[x_1, \hat{\mathbf{a}}_2(x_3, \boldsymbol{\theta})]}{r(x_1 | x_0; \mathbf{a}_1)} \right]$$

where the variance is taken w.r.t. $r(x_1 | x_0; \mathbf{a}_1) r[x_2 | x_1; \hat{\mathbf{a}}_2(x_3, \boldsymbol{\theta})]$. Using (23), we can rewrite this as:

$$\text{Var}[\tilde{p}_3^{(S)}(x_3 | x_0; \boldsymbol{\theta})] \approx \frac{1}{S} \text{Var} \left[\frac{p^a(x_1 | x_0; \boldsymbol{\theta}) \rho[x_1; \hat{\mathbf{a}}_2(x_3, \boldsymbol{\theta})] h_2[x_3, x_1; \boldsymbol{\theta}, \hat{\mathbf{a}}_2(x_3, \boldsymbol{\theta})]}{r(x_1 | x_0; \mathbf{a}_1)} \right]$$

where the variance is now taken only w.r.t. $r(x_1 | x_0; \mathbf{a}_1)$. If we were to follow the same steps above, we should now look for a value $\mathbf{a}_1(x_3, x_0, \boldsymbol{\theta})$ such that:

$$\forall x_1, \quad p^a(x_1 | x_0; \boldsymbol{\theta}) \rho[x_1, \hat{\mathbf{a}}_2(x_3, \boldsymbol{\theta})] h_2[x_3, x_1; \boldsymbol{\theta}, \hat{\mathbf{a}}_2(x_3, \boldsymbol{\theta})] \propto r[x_1 | x_0; \mathbf{a}_1(x_3, x_0, \boldsymbol{\theta})] \quad (24)$$

This strategy would fully extend to a recursive framework the strategy outlined in section 3.1. It is easy to see, however, that its implementation requires either an analytical expression, or a Monte Carlo approximation as a functional of x_1 , for $h_2[x_3, x_1; \boldsymbol{\theta}, \hat{\mathbf{a}}_2(x_3, \boldsymbol{\theta})]$. The first is unavailable, and the second is impracticable, because it would need to keep x_1 fixed at some value, while $\{\tilde{\boldsymbol{\lambda}}^{(s)}, s = 1, \dots, S\}$ are full trajectory draws. Notice moreover that the number of such conditional simulation steps would also grow exponentially with M .

For this reason, Richard and Zhang (1998) suggest to consider the alternative, and sub-optimal, strategy obtained by simply dropping $h_2[x_3, x_1; \boldsymbol{\theta}, \hat{\mathbf{a}}_2(x_3, \boldsymbol{\theta})]$ from the proportionality condition above:

$$\forall x_1, \quad p^a(x_1 | x_0; \boldsymbol{\theta}) \rho[x_1, \hat{\mathbf{a}}_2(x_3, \boldsymbol{\theta})] \propto r[x_1 | x_0; \mathbf{a}_1(x_3, x_0, \boldsymbol{\theta})]$$

To make the problem of estimating $\mathbf{a}_1(x_3, x_0, \boldsymbol{\theta})$ a linear least squares one, let us again consider the decomposition of the auxiliary density in the ratio of its kernel to an integrating constant. Since the latter may be included in the proportionality coefficient, $\mathbf{a}_1(x_3, x_0, \boldsymbol{\theta})$ may be estimated by solving the following problem:

$$\min_{c_1, \mathbf{a}_1} \sum_{s=1}^S \left[\log \left[p^a(\tilde{x}_1^{(s)} | x_0; \boldsymbol{\theta}) \rho[\tilde{x}_1^{(s)}, \hat{\mathbf{a}}_2(x_3, \boldsymbol{\theta})] \right] - c_1 - \log r_K(\tilde{x}_1^{(s)}; x_0, \mathbf{a}_1) \right]^2 \quad (25)$$

As before, any multiplicative factor in $p^a(\tilde{x}_1^{(s)} | x_0; \boldsymbol{\theta}) \rho[\tilde{x}_1^{(s)}, \hat{\mathbf{a}}_2(x_3, \boldsymbol{\theta})]$ which does not depend on x_1 may be collected in the intercept coefficient c_1 .

What makes the strategy composed of (22) and (25) a fully operational one is precisely dropping $h_2[x_3, x_1; \boldsymbol{\theta}, \hat{\mathbf{a}}_2(x_3, \boldsymbol{\theta})]$ from the proportionality condition (24). An obvious situation in which this key simplification would be fully valid would be when $h_2[x_3, x_1; \boldsymbol{\theta}, \hat{\mathbf{a}}_2(x_3, \boldsymbol{\theta})]$ did not depend on x_1 . Notice however that this is very unlikely, as it is easy to check that:

$$h_2[x_3, x_1; \boldsymbol{\theta}, \hat{\mathbf{a}}_2(x_3, \boldsymbol{\theta})] = \frac{p_2(x_3 | x_1; \boldsymbol{\theta})}{\rho[x_1, \hat{\mathbf{a}}_2(x_3, \boldsymbol{\theta})]}$$

Richard and Zhang (1998) justify the deletion of $h_2[x_3, x_1; \boldsymbol{\theta}, \hat{\mathbf{a}}_2(x_3, \boldsymbol{\theta})]$ on the basis of their Assumption 1, which states that the Monte Carlo sampling variance of the latter term is negligible w.r.t. that of the product $p^a(x_1 | x_0; \boldsymbol{\theta}) \rho[x_1, \hat{\mathbf{a}}_2(x_3, \boldsymbol{\theta})]$, to the extent that dropping $h_2[x_3, x_1; \boldsymbol{\theta}, \hat{\mathbf{a}}_2(x_3, \boldsymbol{\theta})]$ in problem (25) has virtually no effect on the estimate $\hat{\mathbf{a}}_1(x_3, x_0, \boldsymbol{\theta})$. They also correctly point out that a direct verification of this assumption is impossible, because it would require a numerical evaluation of the neglected term. However, such a direct verification is in a certain sense superfluous, because its validity is easily checked ex post: if it did not hold, sequential EIS in high dimensional integrals would necessarily fail miserably. Moreover, the R^2 coefficients in both (22) and (25) provide a quick measure of how close the proportionality conditions (21) and (24) are to being valid; if unsatisfactory results emerge, a close inspection of the residuals estimated in (22) and (25) may help in suggesting fruitful extensions of the class \mathcal{R} of sampling densities.

3.2.2 Generic M -steps sequential EIS

In this section we outline the sequential EIS algorithm for arbitrary M . We assume that a suitable set of canonical random numbers, to be used throughout the IS evaluation and maximization of the loglikelihood, has been drawn, and that an initial sampling density has been chosen. The latter can be the member of \mathcal{R} corresponding to a specified vector of parameters \mathbf{a} , the Durham and Gallant (2002) MBB density, or the natural sampler. We also assume that \mathcal{R} is chosen in such a way that the minimization problems reduce to linear least squares programs. Under these assumptions the generic M -steps sequential EIS algorithm to evaluate the transition density of diffusion processes can be outlined as follows.

- **Step 0** [*Simulation of trajectories*]: Use the initial sampling density to draw a set $\{\tilde{\lambda}^{(s)}, s = 1, \dots, S\}$ of independent high-frequency trajectories of the SDE.
- **Step 1** [*Auxiliary regression at subinterval $M - 1$*]: Estimate by OLS \mathbf{a}_{M-1} by solving the program:

$$\min_{c_{M-1}, \mathbf{a}_{M-1}} \sum_{s=1}^S \left[\log p^a(x_M, \tilde{x}_{M-1}^{(s)} | \tilde{x}_{M-2}^{(s)}; \boldsymbol{\theta}) - c_{M-1} - \log r_K(\tilde{x}_{M-1}^{(s)}; \tilde{x}_{M-2}^{(s)}, \mathbf{a}_{M-1}) \right]^2$$

Let us denote with $\hat{\mathbf{a}}_{M-1}$ this estimate of \mathbf{a}_{M-1} .

- **Step $n = M - m$, $m = M - 2, \dots, 1$** [*Auxiliary regression at step m*]: Given the solution obtained at the previous steps, estimate by OLS \mathbf{a}_m by solving the program:

$$\min_{c_m, \mathbf{a}_m} \sum_{s=1}^S \left[\log \left[p^a(\tilde{x}_m^{(s)} | \tilde{x}_{m-1}^{(s)}; \boldsymbol{\theta}) \rho(\tilde{x}_m^{(s)}, \hat{\mathbf{a}}_{m+1}) \right] - c_m - \log r_K(\tilde{x}_m^{(s)}; \tilde{x}_{m-1}^{(s)}, \mathbf{a}_m) \right]^2$$

Let us denote with $\hat{\mathbf{a}}_m$ this estimate of \mathbf{a}_m .

- **Step M** [*EIS evaluation of the transition density*]: Given the estimate of the full vector of parameters $\hat{\mathbf{a}} = (\hat{\mathbf{a}}'_1, \dots, \hat{\mathbf{a}}'_{M-1})'$ obtained in the previous steps, draw S independent trajectories $\{\tilde{\lambda}^{(s)}, s = 1, \dots, S\}$ from $r(\boldsymbol{\lambda}; \hat{\mathbf{a}})$, and approximate the target integral (7) with:

$$\tilde{p}_M^{(S)}(x_M | x_0; \boldsymbol{\theta}) = \frac{1}{S} \sum_{s=1}^S \frac{p^a(x_M, \tilde{\lambda}^{(s)} | x_0; \boldsymbol{\theta})}{r(\tilde{\lambda}^{(s)}; \hat{\mathbf{a}})}$$

The algorithm consist in $M + 1$ steps, indexed from 0 to M . To avoid the risk that a badly inefficient initial sampler prevent EIS from determining the most efficient sampling density in \mathcal{R} , Richard and Zhang (1998) suggest to iterate steps 0 to $M - 1$ before evaluating the desired transition density in step M . Also, as in the one-step version of EIS introduced in section 3.1, it is possible to consider weighted versions of the least squares minimization programs in steps 1 to $M - 1$.

4 Implementing EIS for scalar diffusion processes

The implementation of the EIS algorithm to evaluate the transition density of a diffusion process requires to specify the approximate transition density $p^a(x_m | x_{m-1}; \boldsymbol{\theta})$ and the class of sampling densities \mathcal{R} . For the latter choice, other applications of the EIS strategy generally consider a class \mathcal{R} of candidate sampling densities which is a parametric extension of the chosen subdensity. For the problem we consider, this approach implies that both the approximate subdensity and the auxiliary sampling density are Gaussians. This approach is discussed in the following subsection. However, it is by no means the only one available; in particular, bias reduction

considerations suggest to use as subdensity the Aït-Sahalia (2002) closed form analytical approximation described in section 2.1. While this is in general not Gaussian, it is still possible to evaluate the multidimensional integral using a Gaussian sampling density. We outline this alternative approach in subsection 4.2. In both cases, the associated kernels and integrating constants defined in section 3.2 are easily computed.

4.1 EIS implementation with Gaussian subdensities

In their extensive analysis, Durham and Gallant (2002) have shown that the choice of the subdensity is a crucial ingredient to control the bias of the IS estimate of the transition density. They discuss a large number of available alternatives, among which the most effective is the Shoji and Ozaki (1998) local linearization approach applied to the Lamperti transform of the original diffusion. Anyway, all but one of them are basically Gaussian transition densities. In this section, we consider the implementation of the EIS strategy when the chosen subdensity has this property.

Let for simplicity the M subintervals be of equal length $\delta = \Delta/M$, and consider the following Gaussian subdensity:

$$p^a(x_m|x_{m-1}; \boldsymbol{\theta}) = \phi(x_m; \eta_{m-1}, v_{m-1}^2)$$

In this case:

$$\begin{aligned} p^a(x_M, x_{M-1}|x_{M-2}; \boldsymbol{\theta}) &\propto \frac{1}{v_{M-1} v_{M-2}} \exp \left[-\frac{1}{2} \left(\frac{(x_M - \eta_{M-1})^2}{v_{M-1}^2} + \frac{(x_{M-1} - \eta_{M-2})^2}{v_{M-2}^2} \right) \right] \\ p^a(x_m|x_{m-1}; \boldsymbol{\theta}) &\propto \frac{1}{v_{m-1}} \exp \left[-\frac{1}{2} \frac{(x_m - \eta_{m-1})^2}{v_{m-1}^2} \right] \end{aligned}$$

Let us now turn to the specification of the class \mathcal{R} of candidate auxiliary densities. A convenient one defines the kernels of the conditional densities as the product of the natural sampler and the kernel of a Gaussian density:

$$r_K(x_m; x_{m-1}, \mathbf{a}_m) = p^a(x_m|x_{m-1}; \boldsymbol{\theta}) \zeta(x_m; \mathbf{a}_m), \quad \text{where } \zeta(x_m; \mathbf{a}_m) = \exp(a_{m1}x_m + a_{m2}x_m^2)$$

Under this assumption, $r_K(x_m; x_{m-1}, \mathbf{a}_m)$ is itself a Gaussian kernel, and the associated constant of integration is easily computed:

$$\begin{aligned} r_K(x_m; x_{m-1}, \mathbf{a}_m) &\propto \exp \left(-\frac{1}{2} \frac{x_m^2 - 2\nu_{m-1}x_m}{\omega_{m-1}^2} \right) \\ \rho(x_{m-1}, \mathbf{a}_m) &\propto \left(\frac{\omega_{m-1}^2}{v_{m-1}^2} \right)^{1/2} \exp \left(\frac{1}{2} \frac{\nu_{m-1}^2}{\omega_{m-1}^2} - \frac{1}{2} \frac{\eta_{m-1}^2}{v_{m-1}^2} \right) \end{aligned}$$

where:

$$\nu_{m-1} = \omega_{m-1}^2 \left(\frac{\eta_{m-1}}{v_{m-1}^2} + a_{m1} \right) \quad \text{and} \quad \omega_{m-1}^2 = \left(\frac{1}{v_{m-1}^2} - 2a_{m2} \right)^{-1}$$

Notice that this choice implies that $r(x_m|x_{m-1}; \mathbf{a}_m) = \phi(x_m; \nu_{m-1}, \omega_{m-1}^2)$, which makes it very simple to draw trajectories from the sampling densities. Moreover, in the minimization programs of steps 1 to $M - 1$ the approximate subdensity $p^a(x_m|x_{m-1}; \boldsymbol{\theta})$ cancels out, leaving us with a recursive sequence of linear least squares problems. More specifically, the program in step 1 becomes:

$$\min_{c_{M-1}, a_{M-1,1}, a_{M-1,2}} \sum_{s=1}^S \left[\log p^a(x_M|\tilde{x}_{M-1}^{(s)}; \boldsymbol{\theta}) - c_{M-1} - a_{M-1,1}\tilde{x}_{M-1}^{(s)} - a_{M-1,2}(\tilde{x}_{M-1}^{(s)})^2 \right]^2$$

while those in steps 2 to $M - 1$ become:

$$\min_{c_m, a_{m1}, a_{m2}} \sum_{s=1}^S \left[\log \rho(\tilde{x}_m^{(s)}, \hat{\mathbf{a}}_{m+1}) - c_m - a_{m1}\tilde{x}_m^{(s)} - a_{m2}(\tilde{x}_m^{(s)})^2 \right]^2$$

Notice that, should this specification of \mathcal{R} be unable to provide an adequate fit, additional flexibility can be introduced at a minor cost by allowing the parameters a_{m1} and a_{m2} to be themselves linear combinations of “deep” parameters and (functions of) lagged values of x_m . As a simple example, consider $a_{m1} = \mathbf{k}'_{1,m-1}\boldsymbol{\beta}_1$ and $a_{m2} = \mathbf{k}'_{2,m-1}\boldsymbol{\beta}_2$, where $\mathbf{k}_{1,m-1}$ and $\mathbf{k}_{2,m-1}$ are two vectors of K_1 and K_2 , respectively, functions of x_{m-1} , and $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are conformable vectors of parameters. Under such an assumption, the problems above remain linear in $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, and the sampling density is still $\mathcal{N}(x_m; \nu_{m-1}, \omega_{m-1}^2)$, conditionally on x_{m-1} . In the empirical analysis reported below, we found this extension to be useless for the subdensities considered in this section. However, it will become more important for those considered in the following one.

4.2 EIS implementation with non Gaussian subdensities

Although most candidate subdensities are Gaussian, alternative approximate transition densities may be considered. An example is provided by Elerian (1998) suggested a noncentral Chi-squared transition density based on the Milstein discretization scheme. Durham and Gallant (2002) investigated its properties, and concluded that the improvement over the simpler Gaussian subdensity based on the Euler discretization scheme was marginal.

We also extensively experimented with a subdensity derived using one of Ait-Sahalia (2002) closed form analytical loglikelihood approximations. Although at first this choice might seem puzzling, it is based on the observation that validity of the most general approximation approach (i.e. the one not based on the Lamperti transform of the diffusion process) requires that $\Delta \rightarrow 0$, which can not be met in real data, but can be naturally verified in the simulated likelihood framework by letting M be sufficiently high. EIS may still be useful even when an exact closed form analytical approximation of the log transition density exists, e.g. in multivariate contexts when some elements of the state vector are not observable. In such a case, the sample loglikelihood may be evaluated by integrating out the latent variables. This requires to compute an integral whose dimension is proportional to the number of observations, a task similar to that characterizing stochastic volatility models in discrete time.

As the focus of this paper is on univariate diffusions, we do not investigate further this possibility, but defer it instead to future research. Nevertheless, it is interesting to provide a sketch of the implementation of EIS to non Gaussian densities, because it allows to better appreciate the flexibility of the approach we advocate. If $p^a(x_m|x_{m-1}; \boldsymbol{\theta})$ is not Gaussian, the strategy set out in the previous section is no more convenient, because it is in general impossible to determine explicitly the auxiliary density r and to simulate from it. However, repeated experimentation revealed that in most cases to match the approximate subdensity it is sufficient to choose

$$r_K(x_m; x_{m-1}, \mathbf{a}_m) \propto \exp(a_{m1}x_m + a_{m2}x_m^2)$$

provided that the coefficients $\mathbf{a}_m = (a_{m1}, a_{m2})'$ are allowed to depend linearly on appropriate functions of lagged (usually up to the order 1) values of x_m . The specific form of these functions depends on the chosen subdensity, but a quick inspection of the functional form of the latter usually suggests some low order (rational) polynomial. The choice above for the kernel of the auxiliary density obviously implies that

$$\rho(x_{m-1}, \mathbf{a}_m) \propto (-2a_{m2})^{-1/2} \exp\left(-\frac{1}{2} \frac{a_{m1}^2}{2a_{m2}}\right)$$

and that

$$r(x_m|x_{m-1}; \mathbf{a}_m) = \phi\left(x_m; -\frac{a_{m1}}{2a_{m2}}, -\frac{1}{2a_{m2}}\right)$$

so that it is again very simple to draw trajectories from the sampling densities. As before, the (deep) parameters in \mathbf{a}_m may be computed by solving a recursive sequence of linear least squares problem.

Recall that the existence of a finite variance for the IS estimate of the transition density requires that the tails of the auxiliary density in the denominator do not decline faster than those of density in the numerator. In practice, this condition may be difficult to check analytically, in particular when the transition subdensity is not Gaussian. Some guidance, however, may be obtained using the test statistics developed by Koopman and Shephard (2003) using extreme value theory; see section 5.2 for further details.

5 Numerical experiments

5.1 Setup

This section reports the results of several numerical experiments devoted to the comparison of the performances of EIS and MBB approaches on simulated and real data sets. Although aimed at investigating different properties of the two approaches, the numerical experiments that follow share a basic setup whose description is the subject of this section.

Following the literature, we chose to focus on univariate interest rate diffusion processes, and took as benchmarks the same models (with one irrelevant exception) considered by Aït-Sahalia

(1999) to explore the quality of its closed form transition density approximation. These are the following:

1. *Ornstein - Uhlenbeck process.* Vasicek (1977) assumed that the dynamics of the short term interest rate could be described by the following SDE:

$$dx_t = \theta_1(\theta_2 - x_t)dt + \theta_3 dW_t$$

with $\theta_1 > 0$ and $\theta_3 > 0$. The support of x_t is the whole real line. It is well known that this process can be exactly discretized, leading to a Gaussian transition density with mean $\theta_2 + (x_0 - \theta_2) \exp(-\theta_1 \Delta)$ and variance $[1 - \exp(-2\theta_1 \Delta)] / (2\theta_1)$. This property, which is also shared by the following two benchmark processes, allows us to compare the results obtained using the two importance sampling approximations with those based on the true transition density. Obviously, the Lamperti transform for this process is given by $y_t = x_t / \theta_3$, and

$$\mu_Y(y_t; \theta) = \frac{\theta_1(\theta_2 - \theta_3 y_t)}{\theta_3}$$

2. *Square root process.* Cox, Ingersoll and Ross (1985) modelled the short term interest rate using the following SDE:

$$dx_t = \theta_1(\theta_2 - x_t)dt + \theta_2 \sqrt{x_t} dW_t$$

where the three parameters are positive, and $2\theta_1\theta_2/\theta_3^2 \geq 1$ in order to make the origin inaccessible; under these assumptions the domain of x_t is $(0, +\infty)$. The transition density for this process is noncentral Chi-squared with a non integer number of degrees of freedom; see Cox, Ingersoll and Ross (1985) for details. Its Lamperti transform is given by $y_t = 2\sqrt{x_t}/\theta_3$, and

$$\mu_Y(y_t; \theta) = -\frac{\theta_1 y_t}{2} - \frac{1}{2y_t} \left(1 - \frac{4\theta_1\theta_2}{\theta_3^2}\right)$$

3. *Inverse square root process.* Ahn and Gao (1998) suggested that the interest rate process could be defined as the reciprocal of a square root process. An application of the Itô's Lemma provides the SDE of the resulting specification:

$$dx_t = x_t[\theta_2 + (\theta_2^2 - \theta_2\theta_1)x_t]dt + \theta_3 x_t^{3/2} dW_t$$

for positive θ_1 , θ_2 and θ_3 . The domain of the process is $(0, +\infty)$. By definition, the transition density for this process can be computed by applying the Jacobian formula to that of the Square root process. The Lamperti transform is given by $y_t = 2/(\theta_3\sqrt{x_t})$, and $\mu_Y(y_t; \theta)$ is the same as in the square root process.

4. *Linear drift, CEV diffusion.* This specification originates from Chan, Karolyi, Longstaff and Sanders (1992). The dynamics is governed by the following SDE:

$$dx_t = \theta_2(\theta_1 - x_t)dt + \theta_3 x_t^{\theta_4} dW_t$$

where again all the parameters are positive, and $\theta_4 > 1/2$, so that x_t is distributed on $(0, +\infty)$. For this process no explicit transition density is known, unless unrealistic constraints are placed on the parameters. We assume that $\theta_4 > 1$. Under this assumption, the Lamperti transform is given by $y_t = x_t^{1-\theta_4}/[\theta_3(\theta_4 - 1)]$, and

$$\mu_Y(y_t; \boldsymbol{\theta}) = \frac{\theta_4}{2y_t(\theta_4 - 1)} - \theta_2(1 - \theta_4)y_t + \frac{\theta_2\theta_1}{\theta_3}[\theta_3(1 - \theta_4)y_t]^{\theta_4/(\theta_4-1)}$$

5. *Nonlinear mean reversion.* This model was suggested (among others) by Aït-Sahalia (1996b) as a way to describe a situation in which the strength of mean reversion was much higher at the end of the domain than in its central part. The SDE is given by

$$dx_t = \left(\frac{\theta_1}{x_t} + \theta_2 + \theta_3x_t + \theta_4x_t^2 \right) dt + \theta_5x_t^{1/2}dW_t$$

where $\theta_5 > 0$, and $(0, +\infty)$ as domain of x_t . Notice that a more flexible specification could be considered, with one more parameter θ_6 instead of $1/2$ as the power of x_t in the diffusion coefficient. Although this would be straightforward, we follow Aït-Sahalia (1999) and choose the constrained specification for simplicity. As the diffusion coefficient is the same, the Lamperti transform is the same as in the Square root process, and

$$\mu_Y(y_t; \boldsymbol{\theta}) = -\frac{64\theta_4 + \theta_5^2(-48 + 16\theta_3y_t^2 + 4\theta_2\theta_5^2y_t^4 + \theta_1\theta_5^4y_t^6)}{32\theta_5^2y_t}$$

For each of the diffusion processes above, we set the values of the parameters at their estimates reported in Aït-Sahalia (1999). These were obtained on a sample of 432 monthly observations of the Federal Funds rate between January 1963 and December 1998 using exact ML (when possible, i.e. for processes 1, 2 and 3), or approximate ML, using a closed form expansion of the log transition density. Similarly, whenever the numerical experiments required to choose a value of the process at the beginning (x_0) and at the end of the transition (x_Δ), we adopted the same values as Aït-Sahalia (1999).

In the results that follow the two IS techniques were applied with the same set of tuning parameters. In all cases, $M = 8$ subintervals and $S = 32$ trajectories were considered. We experimented somewhat with higher values of both, but ended up confirming Durham and Galant (2002) results that very little is lost by focussing the attention on the simpler setting. We enhanced variance reduction by using throughout antithetic variates and standardized innovations, i.e. sets of random numbers normalized in order to set their sample mean and variance to zero and one, respectively. After using the Shoji and Ozaki (1998) Gaussian subdensity, no need for extrapolating techniques in order to further reduce the bias emerged.

Additional tuning parameters must be specified in order to implement EIS estimation of the transition density. Given that the chosen subdensity is Gaussian, we adopted the implementation outlined in section 4.1. The algorithm outlined in section 3.2.2 was iterated twice, and the iterations were started at the MBB subdensity. This choice allowed to limit the computational

burden of the problem, but it should be noticed that it is not an essential ingredient for the results that follow. Specifically, we experimented with the alternative strategy of increasing the number of iterations and start them from the auxiliary density associated to the basic Euler subdensity of the process, and obtain very similar results. Moreover, notice that if the MBB auxiliary density could not be used to start the iterations (e.g. because some of the state variables are not observed), an alternative density could be derived from a second order Taylor series expansion of $\log p^a(x_M, \boldsymbol{\lambda}|x_0; \boldsymbol{\theta})$ w.r.t. $\boldsymbol{\lambda}$, as suggested by Liesenfeld and Richard (2003b).

All the computations in this paper were carried out in Fortran, using several NAG subroutines for specific tasks. For example, we remarked that the performance of EIS is significantly improved if the solutions of the least squares problems are computed with higher precision, presumably because this allows a more accurate numerical evaluation of the derivatives. For this reason we used NAG subroutine F04AMF which uses an iterative refinement to provide (approximate) full machine precision estimate of the auxiliary density parameters. Since EIS requires more computations than IS based on the MBB auxiliary density, the evaluation of the likelihood takes more time (about three to five times more, provided that the EIS algorithm is coded efficiently). Even with the increase of computational burden, however, ML estimation is still a matter of a few minutes for the typical sample, as witnessed by the Monte Carlo experiments that follow. In general, the total computational burden of EIS grows linearly with T and M , but more slowly with S and the dimension of the state variable (which is fixed at one in this paper). Extending the methods in this paper to multivariate frameworks, hence, should keep the problem quite tractable.

5.2 Loglikelihood approximation

A first set of experiments studies the quality of the loglikelihood approximation provided by the two approaches. To this end, for each of the processes above we considered the evaluation of the log transition density for fixed backward variable x_0 , and forward variables x_Δ ($\Delta = 1/12$) inside a given range. For each couple (x_0, x_Δ) , we repeated EIS and MBB evaluation 1,000 times. The results are depicted in Figure 1, for the three processes whose exact transition density is known in closed form, and Figure 2, for the remaining two. In both cases the distribution of IS loglikelihood estimates is described using box-and-whiskers plots where the boxes give the centered 50% percentile range, while the vertical segments give the centered 95% percentile range, with a horizontal segment denoting the median. Notice that in all cases the diagrams reporting the results of the EIS and the MBB approaches share the same vertical scale. Moreover, in Figure 1 the vertical axis measures the (*scaled*) *approximation error* w.r.t. the true log transition density, while in Figure 2 it measures the dispersion *of the estimates* around their medians. Finally, and again following Aït-Sahalia (1999), in Figure 2 we considered two starting points x_0 for each of the two processes.

The message in these diagrams is clear: while EIS and MBB are essentially equivalent in

Figure 1: Approximation errors for EIS and MBB estimates of the log transition density for 3 processes for which the latter is known. The boxes (resp. vertical segments) give the centered 50% (resp. 95%) percentile range, with a horizontal segment denoting the median, over 1,000 replications. In all cases: $\Delta = 1/12$; both methods use the same set of normalized random numbers, the Shoji and Ozaki (1998) subdensity and no extrapolation.

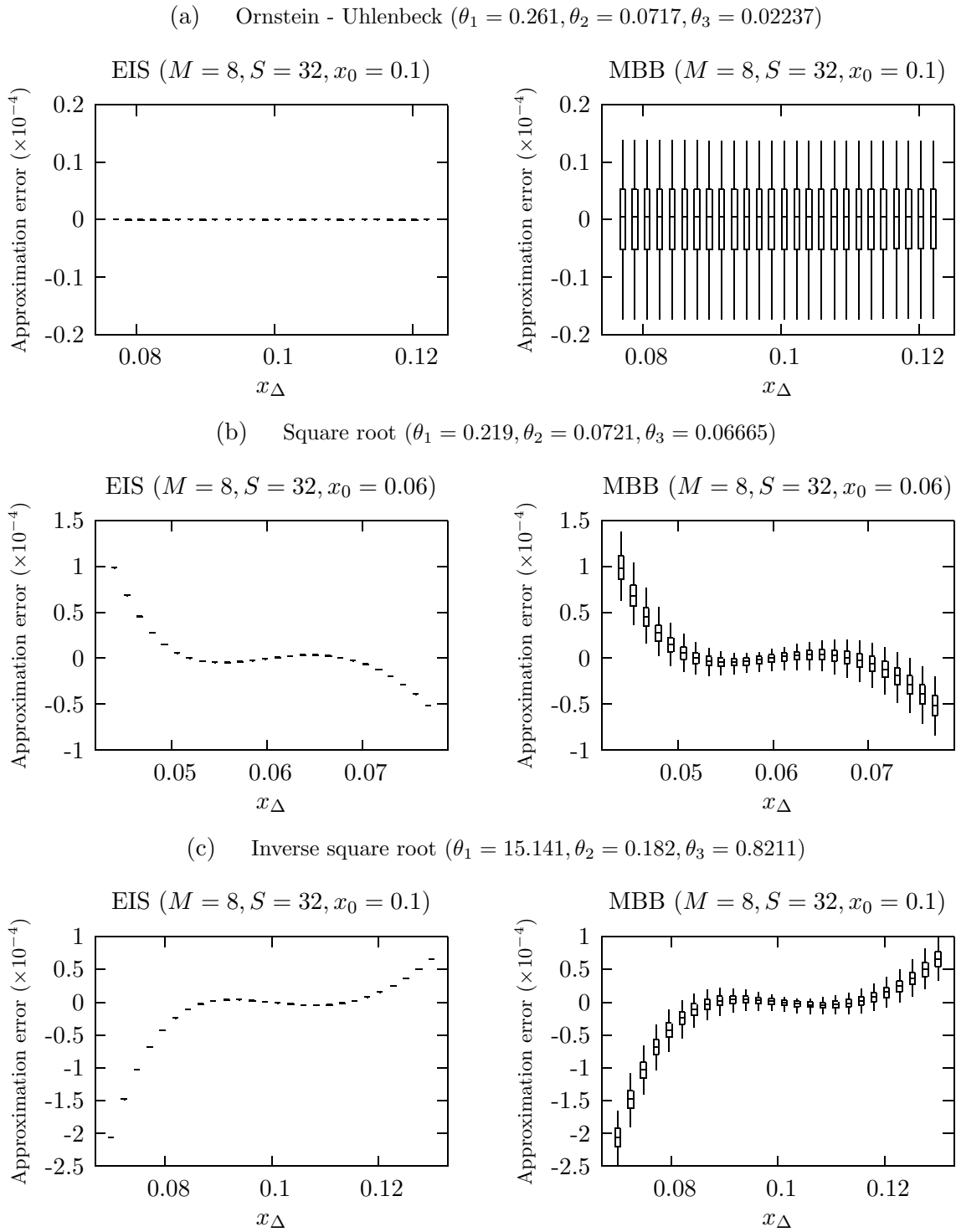
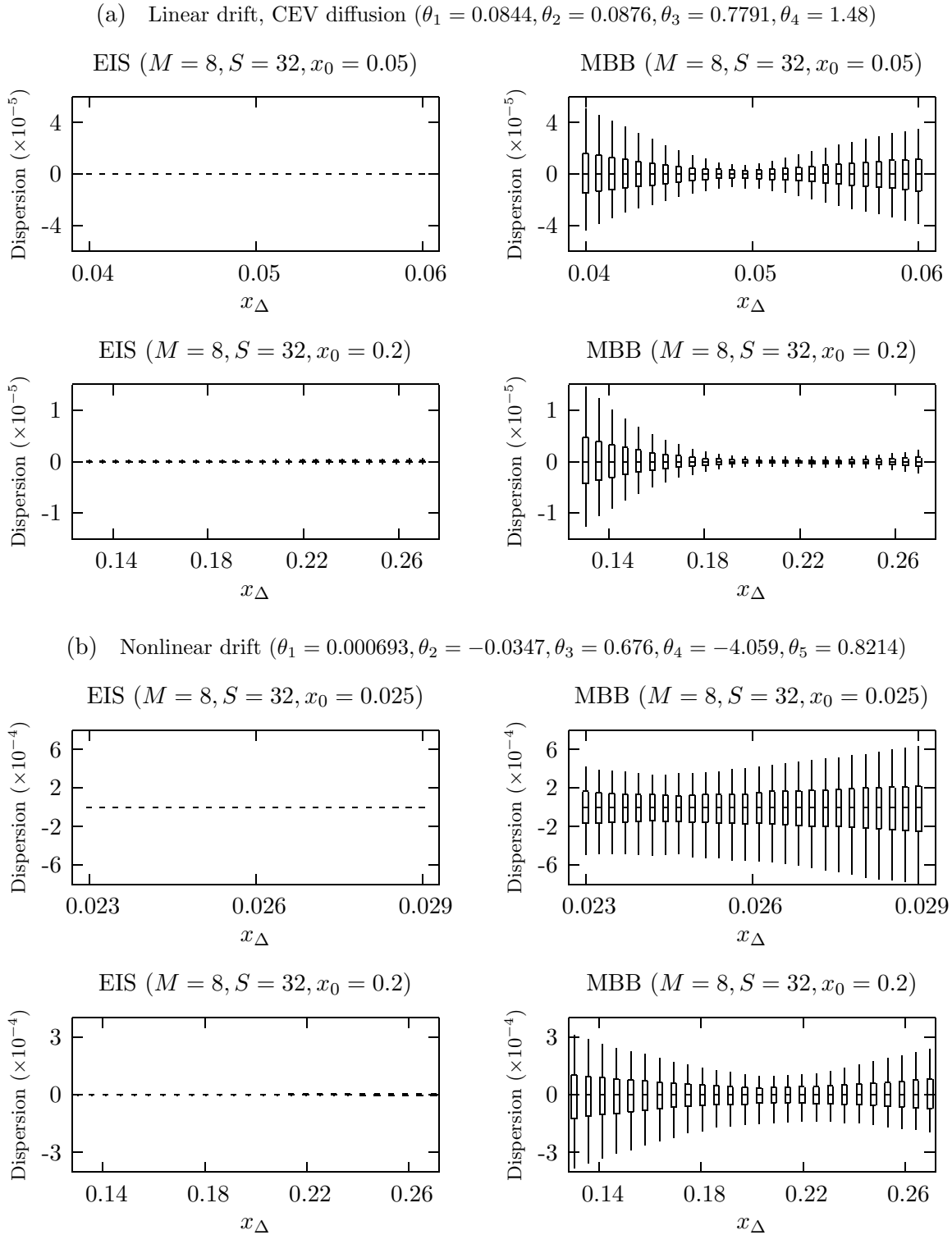


Figure 2: Dispersion of EIS and MBB estimates of the log transition density for processes for which the latter is unknown. The boxes (resp. vertical segments) give the 50% (resp. 95%) percentile range centered on the median, over 1,000 replications. In all cases: $\Delta = 1/12$; both methods use the same set of normalized random numbers, the Shoji and Ozaki (1998) subdensity and no extrapolation.



term of bias (at least for the first three processes, for which this conclusion can be drawn), the former are much less volatile than the latter. The reduction in variance associated to EIS can be measured by a factor ranging from 10 to 1,000, or more; typically, however, the reduction in variance increases rapidly with the size of the transition, so a conservative estimate of the actual reduction of variance should probably be situated somewhere around the lower end of this range. It is more interesting to observe that EIS estimates have virtually no dispersion, which suggests that, when necessary, bias reduction could be obtained by extrapolation, instead of using the Lamperti transform of the process (which is not always possible for multivariate diffusion processes), or increasing the number of subintervals M (which significantly increases the computational burden). The results of some numerical experiments in this direction (not reported for the sake of brevity) confirm this intuition.

Also, notice that for an Ornstein - Uhlenbeck diffusion process the product of subdensities $\prod_{m=1}^M p^a(x_m|x_{m-1}; \boldsymbol{\theta})$ is proportional to a Gaussian density kernel for $\boldsymbol{\lambda} = (x_1, x_2, \dots, x_{M-1})'$, as each individual subdensity $p^a(x_m|x_{m-1}; \boldsymbol{\theta})$ is conditionally normal with mean linear in x_{m-1} and constant variance. This property allows to attain a perfect fit for the corresponding auxiliary regressions, and zero MC variance for the EIS likelihood estimate.

As already outlined in section 2.2, the validity of Importance Sampling approximations rests on the assumption that condition 12 holds. In most cases this is quite difficult to check, and frequently this condition is simply assumed to hold (see e.g. Durham and Gallant, 2002, Assumption 4, p. 300). Recently, however, Koopman and Shephard (2003) have suggested to use extreme value theory to empirically assess the appropriateness of this assumption. Their idea is the following. Smith (1987) argues that for a I.I.D. population the limit distribution of the random variables higher than an ever increasing threshold value is generalized Pareto (see Smith, 1987). This law has two parameters, which can be estimated using standard maximum likelihood; moreover, an interesting property is that the number of finite moments equals the reciprocal of one of the parameters, labelled ξ . This allows to check for the existence of the variance of the IS estimate by restating the null hypothesis as an inequality restriction on ξ , which can be verified using the standard Wald, Lagrange Multipliers and Likelihood Ratio test statistics. This approach has been applied by Lee and Koopman (2004) to two alternative IS estimates of the sample density for some discrete time stochastic variance models.

We implemented the three (Wald, LM and LR) test statistics suggested by Koopman and Shephard (2003) both for EIS and MBB transition density estimation procedures, along with some of the graphic diagnostic tools they propose to detect outliers and analyze the stabilization of the recursive standard error of IS estimates. We do not report the results for the sake of brevity (they are available upon request), but they were overwhelmingly supporting the null hypothesis of the existence of the variance in all cases. Clearly, this result could be expected a priori, given the extremely low variance of both EIS and MBB loglikelihood estimates outlined above.

5.3 Monte Carlo investigation of the numerical error in SML parameter estimates

In this section, we address the issue of simulation error on the ML estimates of the parameters by repeatedly estimating the models on the same sample, while using different sets of random numbers in the computation of IS loglikelihood estimates. For each of the diffusion processes we consider, ML estimation was repeated 1,000 times, using the same Federal Funds rate time series used in Aït-Sahalia (1999). We summarize the 1,000 parameter estimates using their average and standard deviation; the latter statistic provides a direct measure of the simulation error of the EIS and MBB approaches. To put this quantity in perspective, we also report the average (over the 1,000 replications) asymptotic standard errors of the ML estimates, computed using the outer product of gradients estimate of the Fisher information matrix. These standard errors provide information on the *statistical* uncertainty associated to ML estimation of the parameters, whereas MC standard errors are informative about their *numerical* uncertainty. If the two IS approximation techniques are to work adequately, we expect the latter to be much smaller than the former.

The results are summarized in Tables 1 and 2, which refer to the first three and last two processes, respectively. Both Tables consider the numerical uncertainty surrounding the estimation of the individual parameters and of the loglikelihood at the optimum. In the former case, however, exact ML estimation is possible; hence, Table 1 also reports these results, and, for both EIS and MBB SML estimates, “Average” refers to the average bias (multiplied by 10^5). In Table 2, exact ML estimation is impossible, and consequently “Average” reports the average SML estimate for each parameter. Notice that in all panels the Monte Carlo standard errors are multiplied by 10^5 , whereas the asymptotic standard errors are reported in levels.

As expected, there is no Monte Carlo uncertainty in EIS results for the Ornstein - Uhlenbeck process (apart some tiny numerical roundoff error). In all but one case (parameter θ_3 for the linear drift, CEV diffusion process in Table 2) the MC standard errors of EIS based estimates are smaller than those corresponding to MBB, and in most cases significantly so. Also, the asymptotic standard errors are always much higher than the MC standard errors, suggesting that the numerical uncertainty is virtually irrelevant w.r.t. to the statistical uncertainty intrinsic in ML estimation. Table 1 shows that the bias of SML w.r.t. exact ML estimates is very low, and generally smaller for EIS. Notice that both EIS and MBB ML estimates of the parameters reported in panel (b) of Table 2 are fairly different from those obtained in Aït-Sahalia (1999) (cfr. panel (b) in Figure 2). This is probably due to the fact that his estimates are based on a low order (i.e., not sufficiently precise) closed form approximation of the loglikelihood.

Inspection of the two tables shows that, while essentially always supporting the superiority of EIS w.r.t. MBB, the evidence in favor of the former is much more clear for the processes in Table 1 than for those in Table 2. We argue that this result is due to the weak identification of the two relatively more complicated processes in the sample we consider. Specifically, we

Table 1: Approximation errors for parameter estimates for processes for which the log transition density is known. The results were obtained by estimating 1,000 times each model on 432 monthly observations of the Federal Funds rate between January 1963 and December 1998 (the same sample considered in Ait-Sahalia, 1999) using different sets of pseudo random numbers. EIS denotes SML estimates with EIS importance sampler. MBB denotes SML estimates with Durham and Gallant (2002) importance sampler. MLE denotes true ML estimates. EIS and MBB use the same set of normalized random numbers, the Shoji and Ozaki (1998) subdensity and no extrapolation. “Average” is the average bias, “MC S.E.” is the Monte Carlo standard error, and “Asy S.E.” is the average asymptotic standard error.

(a) Ornstein - Uhlenbeck					
		θ_1	θ_2	θ_3	$\ell(\hat{\theta})$
MLE		0.26100	0.07171	0.02237	3.63447
EIS - MLE	Average ($\times 10^5$)	0.00000	0.00017	0.00000	0.00000
	MC S.E. ($\times 10^5$)	0.00000	0.00000	0.00021	0.00000
	Asy S.E.	0.10151	0.02642	0.00020	n.a.
MBB - MLE	Average ($\times 10^5$)	0.56121	-0.78345	-0.03387	1.67724
	MC S.E. ($\times 10^5$)	0.18895	0.64343	0.01980	0.26149
	Asy S.E.	0.10150	0.02642	0.00020	n.a.
(b) Square root					
		θ_1	θ_2	θ_3	$\ell(\hat{\theta})$
MLE		0.21895	0.07206	0.06665	3.91820
EIS - MLE	Average ($\times 10^5$)	4.82821	3.43030	0.44053	0.00811
	MC S.E. ($\times 10^5$)	0.00263	0.00113	0.00080	0.01831
	Asy S.E.	0.07937	0.01705	0.00075	n.a.
MBB - MLE	Average ($\times 10^5$)	5.57039	3.63197	0.40523	1.57009
	MC S.E. ($\times 10^5$)	0.11147	0.03006	0.00883	0.22459
	Asy S.E.	0.07945	0.01705	0.00075	n.a.
(c) Inverse square root					
		θ_1	θ_2	θ_3	$\ell(\hat{\theta})$
MLE		15.14005	0.18205	0.82115	4.15813
EIS - MLE	Average ($\times 10^5$)	-33.79762	-0.69390	0.33351	-0.15814
	MC S.E. ($\times 10^5$)	2.05434	0.09596	0.02106	0.02590
	Asy S.E.	2.91027	0.07160	0.01796	n.a.
MBB - MLE	Average ($\times 10^5$)	184.57059	28.46553	1.38271	0.72442
	MC S.E. ($\times 10^5$)	32.56223	4.43794	0.15065	0.13019
	Asy S.E.	2.91548	0.07173	0.01799	n.a.

investigated in detail the likelihood surface for the linear drift, CEV diffusion and the nonlinear drift model, and we remarked that the objective function is essentially flat over a subset of the parameter space. We suspect that, somewhat paradoxically, the higher precision of the EIS estimate of the loglikelihood helps the optimization algorithm in detecting several local maxima inside this region; in turn, this leads to a slight increase in the Monte Carlo dispersion of the estimates of the parameters and of the loglikelihood optimum.

We conclude this section with a word of caution about the computation of the asymptotic standard errors of the SML estimates of the parameters. In this paper, we chose to base this computation on the outer product of the gradients estimate of the Fisher information matrix be-

Table 2: Approximation errors for parameter estimates for processes for which the log transition density is unknown. The results were obtained by estimating 1,000 times each model on 432 monthly observations of the Federal Funds rate between January 1963 and December 1998 (the same sample considered in Ait-Sahalia, 1999) using different sets of pseudo random numbers. EIS denotes SML estimates with EIS importance sampler. MBB denotes SML estimates with Durham and Gallant (2002) importance sampler. EIS and MBB use the same set of normalized random numbers, the Shoji and Ozaki (1998) subdensity and no extrapolation. “Average” is the average estimate, “MC S.E.” is the Monte Carlo standard error, and “Asy S.E.” is the average asymptotic standard error.

		(a) Linear drift, CEV diffusion				
		θ_1	θ_2	θ_3	θ_4	$\ell(\hat{\theta})$
EIS	Average	0.08417	0.08862	0.77921	1.48120	4.15818
	MC S.E. ($\times 10^5$)	0.07230	0.15914	0.31419	0.14468	0.00302
	Asy S.E.	0.05201	0.10493	0.07693	0.03671	n.a.
MBB	Average	0.08408	0.08890	0.77920	1.48119	4.15819
	MC S.E. ($\times 10^5$)	1.30063	4.33585	0.21425	0.15677	0.13566
	Asy S.E.	0.05175	0.10493	0.07693	0.03672	n.a.

		(b) Nonlinear drift					
		θ_1	θ_2	θ_3	θ_4	θ_5	$\ell(\hat{\theta})$
EIS	Average	0.00066	-0.03281	0.64546	-3.91304	0.82136	4.15860
	MC S.E. ($\times 10^5$)	0.10516	6.15896	107.71796	551.72084	0.15269	0.01930
	Asy S.E.	0.00148	0.08030	1.31134	6.22026	0.01831	n.a.
MBB	Average	0.00067	-0.03302	0.64825	-3.92416	0.82137	4.15862
	MC S.E. ($\times 10^5$)	0.87727	51.30561	896.80166	4593.98309	1.24484	0.27890
	Asy S.E.	0.00148	0.08036	1.31238	6.22472	0.01832	n.a.

cause extensive experimentations showed that this strategy provides results which are essentially identical (replication by replication) between EIS and MBB. On the contrary, we remarked that the EIS and MBB estimate of the Hessian matrix of the loglikelihood provide results which are again very similar on average, but much more dispersed in the former case than in the latter. We argue that the loss in precision of EIS w.r.t. MBB is due to the tiny numerical discontinuities introduced by the solution of the least squares problems which are an essential ingredient of the first approach, but which are absent from the second one. In general, the *absolute* size of the approximation error for the loglikelihood gradient is essentially independent of the size of the derivatives, so that its *relative* importance is higher when the derivatives are small, i.e. in the proximity of the optimum. The use of a Fortran routine providing high precision solutions of the least squares problems essentially eliminates this problem in the case of the loglikelihood gradients, but seems to be inadequate in the case of second order derivatives. Anyway, this issue is still under investigation.

5.4 Monte Carlo investigation of the statistical error in SML parameter estimates

The numerical experiments in the previous sections showed that there is evidence that EIS results, both the estimate of the loglikelihood and the SML estimates of the parameters, are superior (slightly less biased and much less dispersed) than those associated to MBB IS. Nevertheless, it should be noticed that the numerical differences, although fairly apparent, are relatively small in size, and certainly tiny w.r.t. the statistical uncertainty surrounding the true values of the parameters in samples of typical size. The set of Monte Carlo experiments we discuss in this section sheds some light on this point. For each of the five diffusion processes, we simulated 1,000 samples of 512 monthly observations using the same set of parameter values used to draw Figures 1 and 2. For each sample, we estimated the parameters and the value of the loglikelihood at the optimum using both EIS and MBB based SML. For the first three processes, moreover, we also computed the exact ML estimates; this allows us to gauge whether the distance from SML and exact ML estimates is small relative to the distance between exact ML estimates and true values of the parameters. For the last two processes this analysis can not be carried out because exact MLE is impossible; hence we simply focus on the descriptive statistics of the two sets of SML estimates, and compare them with the true values of the parameters.

The results of these experiments are reported in Tables 3 and 4. In the first one, we report the true values of the parameters (labelled TRUE) together with three descriptive statistics (average, standard error, and RMSE) for each one of three sets of differences: exact ML estimates minus true values (MLE - TRUE), EIS SML minus exact ML estimates (EIS - MLE), and MBB SML minus exact ML estimates (MBB - MLE). In the second one, the first set of differences can not be computed, and the descriptive statistics are provided directly for the differences EIS - TRUE and MBB - TRUE.

Inspection of the first Table shows that the distance between SML and exact ML estimates is a small fraction of that between the latter and the true values. As it had to be expected, this is especially true for the parameters in the diffusion coefficient, and for those in the drift not measuring the speed of mean reversion. It should be noted that, to keep the computational burden of these Monte Carlo experiments reasonable, we fixed relatively low values for the number of subintervals M and of simulated trajectories S . In real data applications both tuning parameters could be increased, and this would certainly lead to SML results even closer to exact ML ones.

Finally, both Tables show that the performances of EIS and MBB based SML are extremely close. This confirms that, for the simple univariate processes we consider in this paper, the tiny numerical advantages provided by EIS essentially disappear when the statistical uncertainty implicit in ML parameter estimation estimation is taken into account. Nevertheless, we believe that EIS should still be seen as an interesting alternative to MBB importance sampling, for reasons we already recalled in the previous sections: (*i*) it does not require the knowledge of

Table 3: Approximation errors for parameter estimates for processes for which the log transition density is known. The results were obtained in Monte Carlo experiments with 1,000 simulated samples of 512 monthly observations started at the respective unconditional means for panel (a) and (b), and from $1/\theta_1$ for panel (c). EIS denotes SML estimates with EIS importance sampler. MBB denotes SML estimates with Durham and Gallant (2002) importance sampler. AS denotes the Ait-Sahalia (2002) closed-form approximated ML estimates (order 3, reducible case). MLE denotes true ML estimates. TRUE means true values. EIS and MBB use the same set of normalized random numbers, the Shoji and Ozaki (1998) subdensity and no extrapolation.

(a) Ornstein - Uhlenbeck				
		θ_1	θ_2	θ_3
TRUE		0.26100	0.07170	0.02237
MLE - TRUE	Average	0.09153	-0.00042	0.00001
	S.E.	0.15370	0.01344	0.00072
	RMSE	0.17889	0.01344	0.00072
EIS - MLE	Average	0.00000	0.00000	0.00000
	S.E.	0.00013	0.00000	0.00000
	RMSE	0.00013	0.00000	0.00000
MBB - MLE	Average	0.00211	0.00000	0.00000
	S.E.	0.01012	0.00005	0.00001
	RMSE	0.01034	0.00005	0.00001
(b) Square root				
		θ_1	θ_2	θ_3
TRUE		0.21900	0.07210	0.06665
MLE - TRUE	Average	0.11223	-0.00023	0.00009
	S.E.	0.14125	0.01264	0.00213
	RMSE	0.18041	0.01264	0.00214
EIS - MLE	Average	-0.00574	0.00018	-0.00001
	S.E.	0.02353	0.00100	0.00006
	RMSE	0.02422	0.00101	0.00006
MBB - MLE	Average	-0.00421	0.00018	-0.00001
	S.E.	0.02390	0.00100	0.00006
	RMSE	0.02427	0.00101	0.00006
(c) Inverse square root				
		θ_1	θ_2	θ_3
TRUE		15.14100	0.18200	0.82110
MLE - TRUE	Average	0.18270	0.11308	0.00108
	S.E.	3.71740	0.13847	0.02625
	RMSE	3.72188	0.17877	0.02627
EIS - MLE	Average	0.03205	-0.00517	-0.00013
	S.E.	0.17822	0.02045	0.00057
	RMSE	0.18018	0.02110	0.00059
MBB - MLE	Average	0.03084	-0.00373	-0.00009
	S.E.	0.17815	0.02005	0.00058
	RMSE	0.18080	0.02039	0.00058

the value of the entire state variables vector at the beginning and at the end of the transitions, which makes it clearly superior to MBB in multivariate models with latent variables, and (ii) its

Table 4: Approximation errors for parameter estimates for processes for which the log transition density is unknown. The results were obtained in Monte Carlo experiments with 1,000 simulated samples of 512 monthly observations started at the unconditional mean for the process in panel (a), and from 0.06 for the process in panel (b). EIS denotes SML estimates with EIS importance sampler. MBB denotes SML estimates with Durham and Gallant (2002) importance sampler. AS denotes the Ait-Sahalia (2002) closed-form approximated ML estimates (order 1, reducible case). TRUE means true values. EIS and MBB use the same set of normalized random numbers, the Shoji and Ozaki (1998) subdensity and no extrapolation.

		(a) Linear drift, CEV diffusion				
		θ_1	θ_2	θ_3	θ_4	
TRUE		0.08440	0.08760	0.77910	1.48000	
EIS - TRUE	Average	0.01205	0.09611	0.03946	0.00387	
	S.E.	0.06995	0.13323	0.23576	0.10310	
	RMSE	0.07098	0.16428	0.23904	0.10317	
MBB - TRUE	Average	0.01185	0.09741	0.04028	0.00442	
	S.E.	0.06957	0.13415	0.23435	0.10187	
	RMSE	0.07057	0.16579	0.23778	0.10196	

		(b) Nonlinear drift				
		θ_1	θ_2	θ_3	θ_4	θ_5
TRUE		0.00069	-0.03470	0.67600	-4.05900	0.82140
EIS - TRUE	Average	0.00181	-0.07226	1.05104	-5.72522	0.00157
	S.E.	0.00364	0.14978	2.23131	12.03658	0.02641
	RMSE	0.00406	0.16630	2.46646	13.32882	0.02646
MBB - TRUE	Average	0.00166	-0.06300	0.87259	-4.65231	0.00157
	S.E.	0.00372	0.14759	2.10913	10.89357	0.02637
	RMSE	0.00407	0.16047	2.28250	11.84542	0.02641

higher numerical precision (lower simulation induced variance) allows to reduce the bias without increasing M , but by using extrapolation methods. We are already exploiting these edges in parallel research, focussing on more complicated univariate and multivariate contexts, with very encouraging results.

6 Conclusions

In this paper we considered the issue of ML estimation of the parameters of a diffusion process whose dynamics is described by a univariate homogeneous stochastic differential equation. We reviewed the two most efficient techniques advanced so far in the literature, i.e. those based on analytic closed-form loglikelihood approximations and simulated importance sampling loglikelihood estimation. Although their performance is quite satisfactory in simple contexts, their implementation in more complicated setups, such as those characterized by multivariate processes and latent variables, poses several challenging issues. To overcome them, we suggested to apply an alternative importance sampling strategy put forth by Richard and Zhang (1998), labelled Efficient Importance Sampling, whose performance is equivalent, if not superior, to that

of the importance sampling approaches already appeared in the literature, but whose application requires to meet less stringent conditions. The price to pay is in the form of a moderately higher computational burden, which do not preclude however the possibility to set up a Monte Carlo study to analyze the finite sample performance of the approximation strategy.

As benchmark cases we considered five stochastic processes commonly adopted in the financial literature to describe the evolution over time of the short term interest rate, and also used by Aït-Sahalia (1999) to study the performance of his closed-form approximation approach. A set of Monte Carlo experiments focussed on different aspects of EIS implementation, namely the properties of loglikelihood approximations, of the numerical error in simulated ML estimates, and of the interaction of numerical error with the statistical uncertainty intrinsic in ML parameter estimates. Overall, the comparison with the alternative, state-of-the-art importance sampling strategy suggested by Durham and Gallant (2002) suggests that EIS seems to provide superior results in terms of loglikelihood approximation and of numerical error in parameter estimation. While this edge is lost when the statistical uncertainty is taken into account, these results foresee promising developments in more complicated settings, as they allow to add to importance sampling loglikelihood estimation the degree of flexibility needed to overcome the issues posed by the unavailability of the Lamperti transform of the process, or of the existence of non observable state variables. We plan to explore these developments in future research.

References

- Ahn, D.-H. and B. Gao (1998), “A Parametric Nonlinear Model of the Term Structure Dynamics”, w.p. University of North Carolina at Chapel Hill.
- Aït-Sahalia, Y. (1996a), “Nonparametric Pricing of Interest Rate Derivative Securities”, *Econometrica*, 64, 527-560.
- Aït-Sahalia, Y. (1996b), “Testing Continuous-Time Models of the Spot Interest Rate”, *Review of Financial Studies*, 9, 385-426.
- Aït-Sahalia, Y. (1999), “Transition Densities for Interest Rate and Other Nonlinear Diffusion”, *Journal of Finance*, 54, 1361-1395.
- Aït-Sahalia, Y. (2002), “Maximum Likelihood Estimation of Discretely Sampled Diffusions: A Closed-Form Approximation Approach”, *Econometrica*, 70(1), 223-262.
- Aït-Sahalia, Y. (2003), “Closed-Form Likelihood Expansions for Multivariate Diffusions”, NBER Working Paper No. w8956.
- Aït-Sahalia, Y. and R. Kimmel (2003), “Estimating Affine Multifactor Term Structure Models Using Closed-Form Likelihood Expansions”, NBER Working Paper No. t0286.

- Aït-Sahalia, Y. and R. Kimmel (2004), “Maximum Likelihood Estimation of Stochastic Volatility Models”, NBER Working Paper No. w10579.
- Bakshi, G. and N. Ju (2003), “A Refinement to Aït-Sahalia’s (2002) ”Maximum Likelihood Estimation of Discretely Sampled Diffusions: A Closed-Form Approximation Approach””, *Journal of Business*, forthcoming.
- Bandi, F. and P. Phillips (2003), “Fully Nonparametric Estimation of Scalar Diffusion Models”, *Econometrica*, 71(1), 241-83.
- Brandt, M. W. and P. Santa-Clara (2002), “Simulated Likelihood Estimation of Diffusions with an Application to Exchange Rate Dynamics in Incomplete Markets”, *Journal of Financial Economics*, 63, 161-210.
- Carrasco, M., M. Chernov, J.-P. Florens and E. Ghysels (2002), “Estimating Diffusions with a Continuum of Moment Conditions”, unpublished document, Department of Economics, University of Carolina at Chapel Hill.
- 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”, *Journal of Finance*, 47, 1209-27.
- Cox, J. C., J. E. Ingersoll and S. A. Ross (1985), “A Theory of the Term Structure of Interest Rates”, *Econometrica*, 53, 385-407.
- Duffie, D. and P. Glynn (2001), “Estimation of Continuous-Time Markov Processes Sampled at Random Time Intervals”, working paper, Graduate School of Business, Stanford University.
- Durham, G. B. (2003), “Likelihood-Based Specification Analysis of Continuous-Time Models of the Short-Term Interest Rate”, *Journal of Financial Economics*, 70(3), 463-87.
- Durham, G. B. and R. A. Gallant (2002), “Numerical Techniques for Maximum Likelihood Estimation of Continuous-Time Diffusion Processes”, *Journal of Business and Economic Statistics*, 20, 297-316.
- Elerian, O. (1998), “A note on the existence of a closed form conditional transition density for the Milstein scheme”, w.p. Nuffiel College, Oxford University.
- Eraker, B. (2001), “MCMC Analysis of Diffusion Models with Application to Finance”, *Journal of Business and Economic Statistics*, vol. 19, n. 2, pp. 177-191.
- Gallant, R. and G. Tauchen (1996), “Which Moments to Match?”, *Econometric Theory*, 12, 657-681.
- Geweke, J. (1996), “Monte Carlo Simulation and Numerical Integration”, chap. 15 in Amman, H. M., D. A. Kendrick and J. Rust, *Handbook of Computational Economics*, Elsevier, Amsterdam.

- Gouriéroux, C., A. Monfort and E. Renault (1993), “Indirect Inference”, *Journal of Applied Econometrics*, 8, S85-S118.
- Jensen, B. and R. Poulsen (2002), “Transition Densities of Diffusion Processes: Numerical Comparison of Approximation Techniques”, *Journal of Derivatives*, 9(4), 18-32.
- Jones, C. (1999), “Bayesian Estimation for Continuous-Time Finance Models”, w.p., Simon School of Business, University of Rochester.
- Karatzas, I. and S. E. Shreve (1991), *Brownian Motion and Stochastic Calculus*, Springer, New York.
- Kloeden, P. E. and E. Platen (1992), *Numerical Solutions of Stochastic Differential Equations*, Springer Verlag, Berlin.
- Koopman S. J. and N. Shephard (2003), “Testing the assumptions behind the use of importance sampling”, mimeo, Nuffiel College, Oxford University.
- Lee, K. M. and S. J. Koopman (2004), “Estimating Stochastic Volatility Models: A Comparison of Two Importance Samplers”, *Studies in Nonlinear Dynamics and Econometrics*, 8(2), Article 5, 1-15.
- Liesenfeld, R. and J.-F. Richard (2003a), “Estimation of Dynamic Bivariate Mixture Models: Comments on Watanabe (2000)”, *Journal of Business and Economic Statistics*, 21(4), 570-6.
- Liesenfeld, R. and J.-F. Richard (2003b), “Univariate and Multivariate Stochastic Volatility Models: Estimation and Diagnostics”, *Journal of Empirical Finance*, 10(4), 505-31.
- Lo, A. W. (1988), “Maximum Likelihood Estimation of Generalized Itô Processes with Discretely Sampled Data”, *Econometric Theory*, 4, 231-47.
- Pedersen, A. R. (1995), “A New Approach to Maximum Likelihood Estimation for Stochastic Differential Equations Based on Discrete Observations”, *Scandinavian Journal of Statistics*, 22, 55-71.
- Richard, J.-F. and W. Zhang (1998), “Efficient High-Dimensional Monte Carlo Importance Sampling”, unpublished manuscript, University of Pittsburgh, Dept. of Economics.
- Shoji, I. and T. Ozaki (1998), “Estimation of Nonlinear Stochastic Differential Equations by a Local Linearization Method”, *Stochastic Analysis and Applications*, 16(4), 733-52.
- Smith, R. L. (1987), “Estimating Tails of Probability Distributions”, *Annals of Statistics*, 15, 1174-1207.
- Stanton, R. (1997), “A Nonparametric Model of Term Structure Dynamics and the Market Price of Interest Rate Risk”, *Journal of Finance*, 52, 1973-2002.

Vasicek, O. (1977), “An Equilibrium Characterization of the Term Structure”, *Journal of Financial Economics*, 5, 177-188.

Zhang, W. and L. Lee (2004), “Simulation Estimation of Dynamic Discrete Choice Panel Models with Accelerated Importance Sampling”, *Econometrics Journal*, 7, 120-42.