J. R. Statist. Soc. B (2006)
68, *Part* 3, *pp.* 333–382

Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes

Alexandros Beskos, Omiros Papaspiliopoulos, Gareth O. Roberts and Paul Fearnhead

Lancaster University, UK

[*Read before* The Royal Statistical Society *at a meeting organized by the* Research Section *on Wednesday, October 12th, 2005*, Professor T. J. Sweeting *in the Chair*]

Summary. The objective of the paper is to present a novel methodology for likelihood-based inference for discretely observed diffusions. We propose Monte Carlo methods, which build on recent advances on the exact simulation of diffusions, for performing maximum likelihood and Bayesian estimation.

Keywords: Cox–Ingersoll–Ross model; EM algorithm; Graphical models; Markov chain Monte Carlo methods; Monte Carlo maximum likelihood; Retrospective sampling

1. Introduction

Diffusion processes are extensively used for modelling continuous time phenomena in many scientific areas; an incomplete list with some indicative references includes economics (Black and Scholes, 1973; Chan *et al.*, 1992; Cox *et al.*, 1985; Merton, 1971), biology (McAdams and Arkin, 1997), genetics (Kimura and Ohta, 1971; Shiga, 1985), chemistry (Gillespie, 1976, 1977), physics (Obuhov, 1959) and engineering (Pardoux and Pignol, 1984). Their appeal lies in the fact that the model is built by specifying the instantaneous mean and variance of the process through a stochastic differential equation (SDE). Specifically, a diffusion process V is defined as the solution of an SDE of the type

$$\mathrm{d}V_s = b(V_s;\theta)\,\mathrm{d}s + \sigma(V_s;\theta)\,\mathrm{d}B_s, \tag{1}$$

driven by the scalar Brownian motion B. The functionals $b(\cdot;\theta)$ and $\sigma(\cdot;\theta)$ are called the *drift* and the *diffusion coefficient* respectively and are allowed to depend on some parameters $\theta \in \Theta$. They are presumed to satisfy the regularity conditions (locally Lipschitz, with a linear growth bound) that guarantee a weakly unique, global solution of equation (1); see chapter 4 of Kloeden and Platen (1995). In this paper we shall consider only one-dimensional diffusions, although multivariate extensions are possible.

For sufficiently small time increment $\mathrm{d}t$ and under certain regularity conditions (see Kloeden and Platen (1995)), $V_{t+\mathrm{d}t} - V_t$ is *approximately* Gaussian with mean and variance given by the so-called *Euler* (or Euler–Maruyama) approximation

$$V_{t+\mathrm{d}t} \approx V_t + b(V_t;\theta)\,\mathrm{d}t + \sigma(V_t;\theta)Y\sqrt{\mathrm{d}t}, \qquad \text{where } Y \sim N(0,1),$$

Address for correspondence: Gareth O. Roberts, Department of Mathematics and Statistics, Fylde College, Lancaster University, Lancaster, LA1 4YF, UK.
E-mail: g.o.roberts@lancaster.ac.uk

© 2006 Royal Statistical Society 1369–7412/06/68333

though higher order approximations are also available. The *exact* dynamics of the diffusion process are governed by its transition density

$$p_t(v, w; \theta) = P(V_t \in \mathrm{d}w | V_0 = v; \theta)/\mathrm{d}w, \qquad t > 0,\ w, v \in \mathbf{R}. \tag{2}$$

We shall assume that the process is observed without error at a given collection of time instances,

$$\mathbf{v} = \{V_{t_0}, V_{t_1}, \ldots, V_{t_n}\}, \qquad 0 = t_0 < t_1 < \ldots < t_n;$$

this justifies the notion of a *discretely observed diffusion process*. The time increments between consecutive observations will be denoted $\Delta t_i = t_i - t_{i-1}$ for $1 \leqslant i \leqslant n$.

The log-likelihood of the data set $\mathbf{v}$ is

$$l(\theta|\mathbf{v}) = \sum_{i=1}^{n} l_i(\theta), \qquad l_i(\theta) := \log\{p_{\Delta t_i}(V_{t_{i-1}}, V_{t_i}; \theta)\}.$$

Unfortunately, in all except a few special cases the transition density of the diffusion process and thus its likelihood are not analytically available. Therefore, it is already well documented that deriving maximum likelihood estimates (MLEs) for discretely observed diffusion processes is a very challenging problem. None-the-less, theoretical properties of such MLEs are now well known in particular under ergodicity assumptions; see for example Kessler (1997) and Gobet (2002).

Inference for discretely observed diffusions has been pursued in three main directions. One direction considers estimators that are alternative to the MLE. Established methods within this paradigm include techniques that are based on estimating functions (Bibby *et al.*, 2002), indirect inference (Gourieroux *et al.*, 1993) and efficient methods of moments (Gallant and Long, 1997). Another direction involves numerical approximations to the unknown likelihood function. Aït-Sahalia (2002) advocated the use of closed form analytic approximations to the unknown transition density; see Aït-Sahalia (2004) for multidimensional extensions. An alternative strategy has been to estimate an approximation to the likelihood by using Monte Carlo (MC) methods. The approximation is given by Euler-type discretization schemes, and the estimate is obtained by using importance sampling. The strategy was put forward by Pedersen (1995) and Santa-Clara (1995) and was considerably refined by Durham and Gallant (2002). The third direction employs Bayesian imputation methods. The idea is to augment the observed data with values at additional time points so that a satisfactory complete-data likelihood approximation can be written down and to use the Gibbs sampler or alternative Markov chain Monte Carlo (MCMC) schemes; see Roberts and Stramer (2001), Elerian *et al.* (2001) and Eraker (2001). An excellent review of several methods of inference for discretely observed diffusions is given in Sørensen (2004).

The approach that is introduced in this paper follows a different direction, which exploits recent advances in simulation methodology for diffusions. Exact simulation of diffusion sample paths has become feasible since the introduction of the exact algorithm (EA) in Beskos *et al.* (2004a). The algorithm is reviewed in Section 2 and relies on a technique called *retrospective sampling* which was developed originally in Papaspiliopoulos and Roberts (2004). To date there are two versions of the algorithm: EA1, which can be applied to a rather limited class of diffusion processes, which we call $\mathcal{D}_1$, and EA2, which is applicable to the much more general $\mathcal{D}_2$-class; all definitions are given in Section 2. The greater applicability of EA2 over EA1 comes at the cost of higher mathematical sophistication in its derivation, since certain results and techniques from stochastic analysis are required. However, its computer implementation is similar to that of EA1.

In this paper we show how to use the EA to produce a variety of methods that can be used for maximum likelihood and Bayesian inference. We first discuss three unbiased MC estimators of the transition density (2) for a fixed value of θ: the bridge method (Section 4; first proposed in Beskos *et al.* (2004a)), the acceptance method (AM) (Section 5) and the Poisson estimator (Section 6; first proposed in Wagner (1988a)). The last two estimators are evolved in Sections 5.1 and 6 to yield unbiased estimators of the transition density *simultaneously* for all $\theta \in \Theta$. Thus, the simultaneous estimators can readily be used in conjunction with numerical optimization routines to estimate the MLE and other features of the likelihood surface.

We proceed by introducing a Monte Carlo expectation–maximization (MCEM) algorithm in Section 8. The construction of the algorithm crucially depends on whether there are unknown parameters in the diffusion coefficient σ. The simpler case where only drift parameters are to be estimated is treated in Section 8.1, whereas the general case necessitates the path transformations of Roberts and Stramer (2001) and it is handled in Section 8.2.

Section 9 presents an MCMC algorithm which samples from the joint posterior distribution of the parameters and of appropriately chosen latent variables. Unlike currently favoured methods, our algorithm is not based on imputation of diffusion paths but instead on what we call a *hierarchical simulation model*. In that way, our MCMC method circumvents computing the likelihood function.

Therefore, all our methodology is simulation based, but it has advantages over existing methods of this type for two reasons.

(a) The methods are *exact* in the sense that no discretization error exists, and the MC estimation provides the only source of error in our calculations. Specifically, as the number of MC samples increases, the estimated MLE converges to the true MLE and, as the number of iterations in our MCMC algorithm increases, the samples converge to the true posterior distribution of the parameters.
(b) Our methods are computationally efficient. Whereas approximate methods require rather fine discretizations (and consequently a number of imputed values which greatly exceeds the observed data size) to guarantee sufficient accuracy, our methodology suffers from no such restrictions.

A limitation of the methods that are introduced here is that their applicability is generally attached to that of the EA. However, on-going advances on the EA itself (Beskos *et al.*, 2005a) will weaken further the required regularity conditions so that a much larger class of diffusions than $\mathcal{D}_2$ can be effectively simulated. It is expected that these enhanced simulation algorithms will be of immediate use to the methods that are presented in this paper.

Our methods are illustrated on three different diffusion models. The first is the periodic drift model, which belongs to $\mathcal{D}_1$ and, although it is quite interesting in its own right since its transition density is unavailable, it is used primarily for exposition. However, we also consider two more substantial and well-known applications: the logistic diffusion model for population growth and the Cox–Ingersoll–Ross (CIR) model for interest rates. The former belongs to the $\mathcal{D}_2$-class, whereas the latter is a diffusion process that is outside the $\mathcal{D}_2$-class, and it is used to illustrate how our exact methods can be extended for processes for which the EA2 algorithm is not applicable. Moreover, since we can calculate analytically the likelihood for this model, we have a bench-mark to test the success of our approach. We fit the CIR model to a well-studied data set, which contains euro–dollar rates (recorded every 10 days) between 1973 and 1995, to allow for comparisons with existing methods.

All the algorithms that are presented in this paper are coded in C and have been executed on a Pentium IV 2.6 GHz processor. We note that our methods are not computationally demanding

according to modern statistical computing standards, and in the examples that we have considered the computing times (which are reported explicitly in the following sections) were in the magnitude of seconds, or at worst minutes.

The structure of the paper is as follows. Section 2 reviews the EA. Section 3 sets up the context of transition density estimation, Sections 4–6 present the three different estimators and Section 7 compares them theoretically and empirically. Section 8 introduces the MCEM algorithm and Section 9 the MCMC algorithm. We finish with some general conclusions and directions for further research in Section 10. Background material and proofs are collected in a brief appendix.

2. Retrospective exact sampling of diffusions

Understanding the statistical methodology to be presented in later sections presupposes the introduction of the simulation techniques that are central to our approaches. In particular, we need to understand the form of the output that is provided by the EA. The main references for the material of this section are Beskos *et al.* (2004a) and Beskos and Roberts (2005). For this paper it suffices to analyse the EA for simulating diffusion paths conditional on their ending point (also known as diffusion bridges). In particular we shall show how to simulate a diffusion path starting from $V_0 = v$ and ending at $V_t = w$, for any $t > 0$, $v, w \in \mathbf{R}$. Simulation of unconditioned paths follows in the same lines and is sketched in Section 2.3. When necessary we characterize the EA as *conditional* or *unconditional* to emphasize the type of diffusion simulation that it is used for.

The EA performs rejection sampling by proposing paths from processes that we can simulate and accepting them according to appropriate probability density ratios. The novelty lies in the fact that the paths proposed are unveiled only at finite (but random) time instances and the decision whether to accept the path or not can be easily taken.

It is essential that we first transform the diffusion process (1) into an SDE of unit diffusion coefficient by applying the 1–1 transformation $V_s \to \eta(V_s;\theta) =: X_s$, where

$$\eta(u;\theta) = \int^u \frac{1}{\sigma(z;\theta)}\,\mathrm{d}z \tag{3}$$

is any antiderivative of $\sigma^{-1}(\cdot;\theta)$. Assuming that $\sigma(\cdot;\theta)$ is continuously differentiable, we apply Itô's rule to find that the SDE of the transformed process writes as

$$\mathrm{d}X_s = \alpha(X_s;\theta)\,\mathrm{d}s + \mathrm{d}B_s, \qquad X_0 = x = \eta(V_0;\theta), \quad s \in [0,t], \tag{4}$$

where

$$\alpha(u;\theta) = \frac{b\{\eta^{-1}(u;\theta);\theta\}}{\sigma\{\eta^{-1}(u;\theta);\theta\}} - \frac{\sigma'\{\eta^{-1}(u;\theta);\theta\}}{2}, \qquad u \in \mathbf{R};$$

η^{-1} denotes the inverse transformation and σ' denotes the derivative with respect to the space variable. In what follows we shall make the following standard assumptions for any $\theta \in \Theta$.

(a) $\alpha(\cdot;\theta)$ is continuously differentiable.
(b) $(\alpha^2 + \alpha')(\cdot;\theta)$ is bounded below.
(c) Girsanov's formula for X that is given in expression (32) in Appendix B.1 is a martingale with respect to Wiener measure.

We define

$$A(u;\theta) := \int^u \alpha(z;\theta)\,\mathrm{d}z$$

to be any antiderivative of α. The transition density of X is defined as

$$\tilde{p}_t(x, y; \theta) = P(X_t \in \mathrm{d}y | X_0 = x; \theta)/\mathrm{d}y, \qquad t > 0, \quad x, y \in \mathbf{R}. \tag{5}$$

Before proceeding, we require the following preliminary notation. Let $C \equiv C([0, t], \mathbf{R})$ be the set of continuous mappings from $[0, t]$ to $\mathbf{R}$, $\mathcal{C}$ the corresponding cylinder σ-algebra and $\omega = (\omega_s, s \in [0, t])$ a typical element of C. Let $\mathbb{Q}_\theta^{(t,x,y)}$ denote the distribution of the process X conditioned to start at $X_0 = x$ and to finish at $X_t = y$, for some fixed x and y, and $\mathbb{W}^{(t,x,y)}$ be the probability measure for the corresponding Brownian bridge (BB). The notation highlights the dependence of the measure $\mathbb{Q}_\theta^{(t,x,y)}$ on θ.

The objective is to construct a rejection sampling algorithm to draw from $\mathbb{Q}_\theta^{(t,x,y)}$. The following lemma proved in Appendix B.1 is central to the methodology. By $\mathcal{N}_t(u)$ we denote the density of the normal distribution with mean 0 and variance t evaluated at $u \in \mathbf{R}$.

Lemma 1. Under conditions (a)–(c) above, $\mathbb{Q}_\theta^{(t,x,y)}$ is absolutely continuous with respect to $\mathbb{W}^{(t,x,y)}$ with density

$$\frac{\mathrm{d}\mathbb{Q}_\theta^{(t,x,y)}}{\mathrm{d}\mathbb{W}^{(t,x,y)}}(\omega) = \frac{\mathcal{N}_t(y-x)}{\tilde{p}_t(x, y; \theta)} \exp\left\{ A(y; \theta) - A(x; \theta) - \int_0^t \frac{1}{2}(\alpha^2 + \alpha')(\omega_s; \theta)\, \mathrm{d}s \right\}.$$

Let

$$l(\theta) \leqslant \inf_{u \in \mathbf{R}} \{(\alpha^2 + \alpha')(u; \theta)/2\},$$
$$r(\omega, \theta) \geqslant \sup_{s \in [0,t]} \{(\alpha^2 + \alpha')(\omega_s; \theta)/2 - l(\theta)\}, \qquad \omega \in C.$$

In general r is a positive random variable. However, in special cases it can be chosen independently of ω, e.g. if $(\alpha^2 + \alpha')(\cdot; \theta)$ is bounded above. We now define the non-negative function $0 \leqslant \phi \leqslant 1$ which transforms a given path ω as follows:

$$\phi(\omega_s; \theta) = \frac{1}{r(\omega, \theta)} \left\{ \frac{(\alpha^2 + \alpha')(\omega_s; \theta)}{2} - l(\theta) \right\}, \qquad s \in [0, t]. \tag{6}$$

It is now clear that

$$\frac{\mathrm{d}\mathbb{Q}_\theta^{(t,x,y)}}{\mathrm{d}\mathbb{W}^{(t,x,y)}}(\omega) \propto \exp\left\{ -r(\omega, \theta) \int_0^t \phi(\omega_s; \theta)\, \mathrm{d}s \right\} \leqslant 1, \qquad \mathbb{W}^{(t,x,y)}\text{—almost surely.} \tag{7}$$

Thus we have managed in expression (7) to bound the density ratio. The key to proceed to a feasible rejection sampler is to recognize expression (7) as a specific Poisson process probability.

Theorem 1. Let $\omega \in C$, Φ be a homogeneous Poisson process of intensity $r(\omega, \theta)$ on $[0, t] \times [0, 1]$ and N be the number of points of Φ below the graph $s \mapsto \phi(\omega_s; \theta)$. Then

$$P(N = 0 | \omega) = \exp\left\{ -r(\omega, \theta) \int_0^t \phi(\omega_s; \theta)\, \mathrm{d}s \right\}.$$

Theorem 1 suggests rejection sampling by means of an auxiliary Poisson process as follows.

Step 1: simulate a sample path $\omega \sim \mathbb{W}^{(t,x,y)}$.
Step 2: calculate $r(\omega, \theta)$; generate a marked Poisson process $\Phi = \{\Psi, \Upsilon\}$, with points $\Psi = \{\psi_1, \ldots, \psi_\kappa\}$ that are uniformly distributed on $[0, t]$ and marks $\Upsilon = \{\upsilon_1, \ldots, \upsilon_\kappa\}$ that are uniformly distributed on $[0, 1]$, where $\kappa \sim \mathrm{Po}\{r(\omega, \theta)t\}$.

Step 3: compute the acceptance indicator

$$I := \prod_{j=1}^{\kappa} \mathbb{I}[\phi(\omega_{\psi_j}; \theta) < \upsilon_j]. \tag{8}$$

Step 4: if $I = 1$, i.e. $\{N = 0\}$ has occurred, then accept ω; otherwise return to step 1.

Unfortunately, this 'algorithm' is impossible to implement since it requires the simulation of complete BBs on $[0, t]$. However, it might be possible to determine I on the basis of only partial information about the path proposed. For instance, when r is only a function of θ (see Section 2.1) we can actually reverse the order in which steps 1 and 2 are carried out. Specifically, we would first simulate Φ, and afterwards, *retrospectively*, we would realize ω at the time instances that are determined by Ψ, since this is sufficient for determining I. The technique of exchanging the order of simulation to implement in finite time simulation of infinite dimensional random variables has been termed *retrospective sampling* in Papaspiliopoulos and Roberts (2004).

The general framework under which the EA operates assumes that it is possible to write

$$r(\omega, \theta) = r\{k(\omega), \theta\}$$

where $k(\omega)$ has the following properties.

(a) $k(\omega)$ is finite dimensional.
(b) The law of $k(\omega)$ can be simulated (under $\mathbb{W}^{(t,x,y)}$).
(c) The finite dimensional distributions of $\mathbb{W}^{(t,x,y)}$ given $k(\omega)$ can be simulated.

Under these conditions, the following retrospective implementation of the rejection sampler can be carried out in finite time: the EA.

Step 1: simulate $k(\omega)$.
Step 2: generate a realization of the marked Poisson process $\Phi = \{\Psi, \Upsilon\}$ of rate $r\{k(\omega), \theta\}$.
Step 3: simulate the skeleton $\{\omega_{\psi_1}, \ldots, \omega_{\psi_\kappa}\}$, conditionally on $k(\omega)$.
Step 4: compute the acceptance indicator I.
Step 5: if $I = 1$, then accept the skeleton proposed, and return $k(\omega)$ and $S(\omega) := \{(0, x), (\psi_1, \omega_{\psi_1}), \ldots, (\psi_\kappa, \omega_{\psi_\kappa}), (t, y)\}$; otherwise return to step 1.

$S(\omega)$ is an *exact* draw from a finite dimensional distribution of $\mathbb{Q}_\theta^{(t,x,y)}$, which can be filled in at any required times afterwards using simply BB interpolation (see Section 2.3). The technical difficulty of finding or simulating $k(\omega)$, and consequently simulating the process at some time points given $k(\omega)$, imposes some restrictions on the applicability of the EA. We now describe the two cases where the algorithm can be easily applied.

2.1. Exact algorithm 1 (EA1)

Implementation of the EA is straightforward when r does not depend on ω. This will be true within the following diffusion class.

Definition 1. We say that a diffusion process V with SDE (1) belongs to $\mathcal{D}_1$, and write $V \in \mathcal{D}_1$, if the drift of the transformed process $X_s = \eta(V_s; \theta)$, $s \in [0, t]$, satisfies conditions (a)–(c) below equation (4) and $(\alpha^2 + \alpha')(\cdot; \theta)$ is bounded above.

In this case

$$\begin{aligned} r(\omega, \theta) &\equiv r(\theta) \\ &= \sup_{u \in \mathbf{R}} \{(\alpha^2 + \alpha')(u; \theta)/2 - l(\theta)\}. \end{aligned}$$

Within $\mathcal{D}_1$, step 1 of the EA is unnecessary, since no information about the path is *a priori* needed for determining the Poisson rate. Moreover, step 3 entails simulation from a finite dimensional distribution of the BB (see Appendix A).

2.1.1. Example 1 (periodic drift)

Consider the SDE

$$\mathrm{d}X_s = \sin(X_s - \theta)\,\mathrm{d}s + \mathrm{d}B_s.$$

Though apparently simple, the SDE cannot be solved analytically. However, the EA1 algorithm can be applied since $X \in \mathcal{D}_1$ with $l(\theta) = -\frac{1}{2}$ and $r(\theta) = 9/8$. Our proposed methods will be tested on the SINE data set simulated from X with the unconditional version of EA1 (Beskos and Roberts, 2005) under the specifications $n = 1000$, $\Delta t_i = 1$, $X_0 = 0$ and $\theta = \pi$; see also Fig. 1. When θ is to be estimated, we take $\Theta = [0, 2\pi]$ for identifiability.

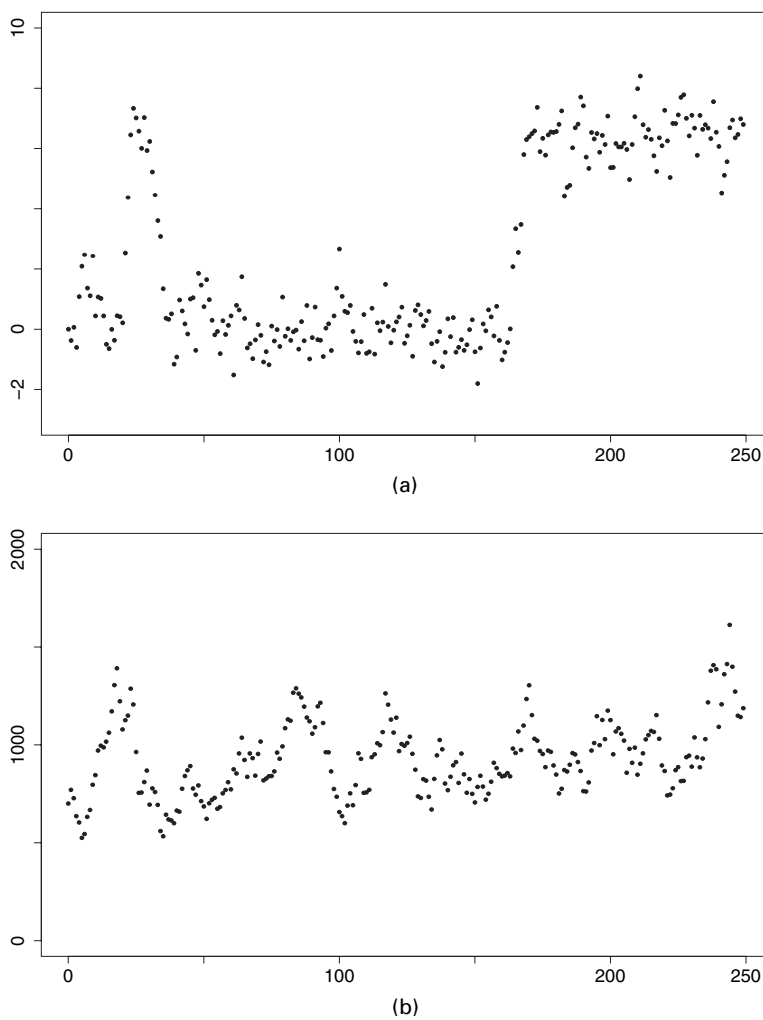

Fig. 1. First 250 values of (a) the SINE (using the EA1 algorithm) and (b) the LOG-GROWTH (using the EA2 algorithm) data sets that are defined in example 1 and example 2 respectively

2.2. Exact algorithm 2 (EA2)

The EA2 algorithm applies to the wider class of diffusion processes that is defined below.

Definition 2. We say that a diffusion process V with SDE (1) belongs to the class $\mathcal{D}_2$, and write $V \in \mathcal{D}_2$, if the drift of the transformed process $X_s = \eta(V_s; \theta), s \in [0, t]$, satisfies conditions (a)–(c) below equation (4) and

$$\limsup_{u \to \infty} \{(\alpha^2 + \alpha')(u; \theta)\} < \infty \tag{9}$$

or $\limsup_{u \to -\infty} \{(\alpha^2 + \alpha')(u; \theta)\} < \infty$.

Owing to symmetry, we shall study only case (9). We define the following elements of the path space $C = C([0, t], \mathbf{R})$:

$$m = \inf\{\omega_s; s \in [0, t]\},$$
$$\tau = \sup\{s \in [0, t] : \omega_s = m\};$$

thus m is the minimum value of a path ω, and τ is the time that the minimum is attained. Within $\mathcal{D}_2$, $k(\omega) = m(\omega)$ and

$$\begin{aligned} r(\omega, \theta) &\equiv r(m, \theta) \\ &= \sup_{u \in [m, \infty]} \{(\alpha^2 + \alpha')(u; \theta)/2 - l(\theta)\} < \infty, \qquad \omega \in C. \end{aligned}$$

As shown in Beskos *et al.* (2004a) and summarized in Appendix A, m satisfies the three requirements that are stated just above the EA. Simulation of (τ, m) can be done by using simple transformations of elementary random elements. It is known (Asmussen *et al.*, 1995) that the BB conditionally on (τ, m) can be derived in terms of two independent *Bessel bridges*, each operating on either side of (τ, m). The Bessel bridge is defined as a BB that is constrained to be positive and its simulation can be carried out by means of independent BBs; see Appendix A for details.

2.2.1. Example 2 (the logistic growth model)

A popular model for describing the dynamics of a population which grows at a geometric rate in an environment with limited feeding resources is given by the SDE

$$\mathrm{d}V_s = RV_s(1 - V_s/\Lambda)\,\mathrm{d}s + \sigma V_s\,\mathrm{d}B_s, \qquad \theta = (R, \Lambda, \sigma).$$

R is the growth rate per individual, Λ the maximal population that can be supported by the resources of the environment and σ a noise parameter; for more details see chapter 6 of Goel and Richter-Dyn (1974). Related models have been investigated in the context of financial economics; see for example Gourieroux and Jasiak (2003). The transition density of V is not known analytically. The modified process $X_s = -\log(V_s)/\sigma$ solves the SDE

$$\mathrm{d}X_s = \left\{ \frac{\sigma}{2} - \frac{R}{\sigma} + \frac{R}{\sigma\Lambda} \exp(-\sigma X_s) \right\} \mathrm{d}s + \mathrm{d}B_s.$$

It can be verified that $V \in \mathcal{D}_2$ and that the EA2 algorithm is applicable with $l(\theta) = \sigma^2/8 - R/2$ and

$$r(m, \theta) = \max\{(\alpha^2 + \alpha')(m; \theta)/2 - l(\theta), R^2/2\sigma^2\}.$$

Our proposed methods will be tested on the LOG-GROWTH data set simulated from V by first simulating from X by using the unconditional EA2 algorithm and then transforming $X_s \to V_s$

(Beskos *et al.*, 2004a). We took $n = 1000$, $\Delta t_i = 1$, $V_0 = 700$ and $(R, \Lambda, \sigma) = (0.1, 1000, 0.1)$; see Fig. 1.

2.3. Unconditional exact algorithm and path reconstruction

The EA can be applied in a similar fashion when the diffusion is conditioned only on its initial point $X_0 = x$. The main difference lies in that the final value of the path proposed is distributed according to the density

$$h(u) \propto \exp\{A(u;\theta) - (u-x)^2/2t\}$$

(which is assumed to be integrable). Simulation from h can be done efficiently by using an adaptive rejection sampling approach. Conditionally on this final point, the rest of the path is a BB, and the EA proceeds as already described.

The output of the EA is a skeleton $S(\omega)$ and possibly the collection $k(\omega)$ of variables related to the path. However, we can afterwards fill in this finite representation of an accepted path ω according to the dynamics of the proposed BB and without further reference to the target process. For the EA1 algorithm, the path reconstruction requires the simulation of the BBs that connect the currently unveiled instances of an accepted path ω. For the EA2 algorithm, the extra conditioning of the proposed BB on its minimum m implies that the filling in will be driven by two independent Bessel bridges; see Appendix A.

3. Unbiased transition density estimators—preliminaries

As already discussed in Section 1 one of our primary aims is to provide unbiased MC estimators of the transition density equation (2). The first step is to express equation (2) in terms of the density (5) of the transformed process X. A change of variables argument yields

$$p_t(v, w; \theta) = \tilde{p}_t\{\eta(v;\theta), \eta(w;\theta); \theta\}\, |\eta'(w;\theta)|, \qquad \text{for all } t, v, w, \theta.$$

Thus, we shall demonstrate how to estimate $\tilde{p}_t(x, y; \theta)$ for arbitrary x, y and θ, using the *unconditional* EA (bridge method) and the *conditional* EA (AM, Poisson estimator). Automatically, our methods yield unbiased estimators of the likelihood $\exp\{l(\theta|\mathbf{v})\}$, for any fixed $\theta \in \Theta$, so coupled with grid or more elaborate stochastic search algorithms they can be used for locating the MLE. However, this *pointwise* exploration of the likelihood surface is expected to be computationally inefficient owing to the introduction of independent MC error at each likelihood evaluation, and it is not guaranteed to provide consistent MC estimators of the MLE. We overcome these drawbacks by modifying the two methods based on the conditional EA to give estimates of the likelihood function *simultaneously* for all $\theta \in \Theta$. The simultaneous methods achieve estimation of the complete function

$$\theta \mapsto \tilde{p}_t\{\eta(v;\theta), \eta(w;\theta); \theta\}, \qquad \text{for any fixed data points } v, w, \tag{10}$$

using a single stream of random elements which are independent of θ. Therefore, numerically efficient optimization algorithms can be used to estimate features of the likelihood surface, such as the MLE and level sets. The simultaneous AM has appealing consistency properties, as we discuss in Section 5.1.

A fundamental result which the AM and the Poisson estimator are based on is the following corollary to lemma 1. This important result is also contained in lemma 1 of Dacunha-Castelle and Florens-Zmirou (1986).

Corollary 1. Let X be the diffusion process in equation (4) with transition density $\tilde{p}_t(x, y; \theta)$. Then

$$\tilde{p}_t(x, y; \theta) = \mathcal{N}_t(y - x)\, \mathbb{E}_{\mathbb{W}^{(t,x,y)}}\left[\exp\left\{A(y; \theta) - A(x; \theta) - \int_0^t \frac{1}{2}(\alpha^2 + \alpha')(\omega_s; \theta)\, \mathrm{d}s\right\}\right]. \tag{11}$$

4. The bridge method

The bridge method that was introduced in Beskos *et al.* (2004a) uses explicitly the output of the *unconditional* EA to produce unbiased estimates of $\tilde{p}_t(x, y; \theta)$ for any fixed x, y and θ. We enrich the notation for the transition density to allow for conditioning on further random elements: let

$$\tilde{p}_t(x, y|\cdot; \theta) = P(X_t \in \mathrm{d}y | X_0 = x, \cdot; \theta)/\mathrm{d}y$$

be the conditional density of X_t given the starting value and any other random elements.

Consider some $\delta > 0$. Suppose that $S(\omega)$ and $k(\omega)$ are the output of the unconditional EA for simulating X started at $X_0 = x$ on the extended time interval $[0, t + \delta]$. By conditional expectation properties, we have that

$$\tilde{p}_t(x, y; \theta) = \mathbb{E}_{S(\omega), k(\omega)}[\tilde{p}_t\{x, y | S(\omega), k(\omega); \theta\}]. \tag{12}$$

The joint distribution of $S(\omega)$ and $k(\omega)$ is intractable but can be easily simulated by using the EA on $[0, t + \delta]$. Recall that the value of the path at times other than those specified by $S(\omega)$ can be obtained according to BB dynamics (in EA1) or Bessel bridge dynamics (in EA2). As a result, the conditional distribution of X_t given $S(\omega)$ and $k(\omega)$, and thus density on the right-hand side of equation (12), can be easily identified. By construction $S(\omega)$ contains at least one time point on either side of t, since $(0, x), (t + \delta, \omega_{t+\delta}) \in S(\omega)$, and let t_- and t_+ denote the adjacent time points, $t_- < t < t_+$. By the Markov property $\tilde{p}_t\{x, y | S(\omega), k(\omega); \theta\}$ depends only on the points (t_-, ω_{t_-}) and (t_+, ω_{t_+}) of $S(\omega)$ and it will be equal either to a BB or a Bessel bridge density, both of which are analytically available and can be computed explicitly.

Thus, the main attraction of the method is that the density of X_t given the simulated variables (the skeleton) is explicitly known, unlike for example the MC method by Pedersen (1995), where it is approximated.

5. The acceptance method

In this section we show how to use a simple identity that is related to the *conditional* EA to derive an unbiased estimator of the transition density $\tilde{p}_t(x, y; \theta)$. Let $a(x, y, \theta)$ be the acceptance probability of the EA for simulating from $\mathbb{Q}_\theta^{(t,x,y)}$. Directly, expression (7) implies that

$$a(x, y, \theta) = \mathbb{E}_{\mathbb{W}^{(t,x,y)}}\left[\exp\left\{-r(\omega, \theta) \int_0^t \phi(\omega_s; \theta)\, \mathrm{d}s\right\}\right]. \tag{13}$$

Substituting this expression in equation (11), and rearranging terms according to equation (6), yields the following identity which relates the acceptance probability of the EA to the transition density of the diffusion:

$$\tilde{p}_t(x, y; \theta) = \mathcal{N}_t(y - x)\, \exp\{A(y; \theta) - A(x; \theta) - l(\theta)t\}\, a(x, y, \theta). \tag{14}$$

Recall the definition of the acceptance indicator I in expression (8), which we shall now rewrite as $I(x, y, \theta, \Phi, \omega)$ to emphasize the elements which determine its value: the Poisson process Φ,

and the proposed bridge $\omega \sim \mathbb{W}^{(t,x,y)}$, unveiled at the times that are determined by Φ. By definition,

$$a(x, y, \theta) = \mathbb{E}\{I(x, y, \theta, \Phi, \omega)\}, \tag{15}$$

where the expectation is taken with respect to the joint distribution of Φ and ω. Thus, a simple unbiased estimator of $a(x, y, \theta)$ for any fixed θ, x and y can be obtained by recording the number of times that the EA accepts a proposed skeleton in K, say, number of trials.

Estimation of the transition density through equations (14) and (15) is referred to as the AM. An essential feature of the resulting estimator is that it is almost surely bounded, since $I \leqslant 1$. As a result, all the moments of the estimator are finite.

5.1. The simultaneous acceptance method

In this section we upgrade the AM to yield an estimator of the complete map in expression (10). The method is now referred to as the simultaneous AM (SAM). We emphasize that equations (13)–(15) are applied for $x = x(\theta) = \eta(v; \theta)$ and $y = y(\theta) = \eta(w; \theta)$. The simultaneous estimator is obtained by expressing $a(x, y, \theta)$ as an expectation of a function of θ, x, y and random variables, none of which depends on θ, x and y. The method is presented for $V \in \mathcal{D}_1$, since the derivation of the SAM for $V \in \mathcal{D}_2$ is technically much more difficult, and it is contained in Beskos *et al.* (2005b), along with more theoretical results about the methods.

In the AM, the Poisson process Φ and the proposed path ω both depend on θ, since the former is of rate $r(\theta)$ and the latter is a BB from $X_0 = x$ to $X_t = y$. As a result $a(x, y, \theta)$ can be estimated only by running a separate MC experiment for each $\theta \in \Theta$. Below we show how the thinning property of the Poisson process (see for example Section 5.1 of Kingman (1993)) and the relocation invariance property of BBs (see Appendix A) can be exploited to decouple the dependence between Φ, ω and θ.

The relocation invariance property of BBs suggests that we can rewrite the acceptance indicator in terms of a standard BB, $\omega \sim \mathbb{W}^{(t,0,0)}$, as follows:

$$I(x, y, \theta, \Phi, \omega) = \prod_{j=1}^{\kappa} \mathbb{I}[\phi\{\omega_{\psi_j} + (1 - \psi_j/t)x + (\psi_j/t)y; \theta\} < \upsilon_j]. \tag{16}$$

Suppose that we can find an $r_{\max} < \infty$, such that

$$r(\theta) \leqslant r_{\max}, \qquad \text{for all } \theta \in \Theta.$$

Let $\Phi_{\max} = \{\Psi_{\max}, \Upsilon_{\max}\}$ be a marked Poisson process on $[0, t] \times [0, 1]$ with rate $r_{\max}$ and κ number of points, $\kappa \sim \mathrm{Po}(r_{\max} t)$. The thinning property for Poisson processes implies that the process that is obtained by deleting each point of $\Phi_{\max}$ with probability $1 - r(\theta)/r_{\max}$ is a Poisson process with rate $r(\theta)$. Conditionally on κ, let $U = (U_1, \ldots, U_\kappa)$ be a collection of independent and identically distributed Un[0, 1] selection random variables. The thinning property implies that we can rewrite equation (16) in terms of $\Phi_{\max}$ as

$$I(x, y, \theta, \Phi_{\max}, \omega, U) = \prod_{j=1}^{\kappa} \mathbb{I}[\mathbb{I}[U_j > r(\theta)/r_{\max}]\, \phi\{\omega_{\psi_j} + (1 - \psi_j/t)x + (\psi_j/t)y; \theta\} < \upsilon_j];$$

note that only Poisson points for which $U_j < r(\theta)/r_{\max}$ determine the value of I.

Theorem 2. Let $\omega \sim \mathbb{W}^{(t,0,0)}$, and $\Phi_{\max} = \{\Psi_{\max}, \Upsilon_{\max}\}$ be an independent marked Poisson process of rate $r_{\max}$ with κ number of points. Then

$$\begin{aligned}\mathbb{E}\{I(x,y,\theta,\Phi_{\max},\omega,U)|\Psi_{\max},\omega\} &= \prod_{j=1}^{\kappa}\left[1-\frac{r(\theta)}{r_{\max}}\phi\left\{\omega_{\psi_j}+\left(1-\frac{\psi_j}{t}\right)x+\frac{\psi_j}{t}y;\theta\right\}\right]\\ &= r_{\max}^{-\kappa}\prod_{j=1}^{\kappa}\left[r_{\max}+l(\theta)-\frac{1}{2}(\alpha^2+\alpha')\right.\\ &\quad\left.\times\phi\left\{\omega_{\psi_j}+\left(1-\frac{\psi_j}{t}\right)x+\frac{\psi_j}{t}y;\theta\right\}\right]. \qquad (17)\end{aligned}$$

$a(x,y,\theta)$ is the expected value of equation (17) with respect to the joint distribution of $\Psi_{\max}$ and $\omega\sim\mathbb{W}^{(t,0,0)}$; thus equation (17) together with equation (14) determine an unbiased *simultaneous* estimator of $\theta\mapsto\tilde{p}_t\{\eta(v;\theta),\eta(w;\theta);\theta\}$, for any fixed data points v and w. An MC estimator of $a(x,y,\theta)$ is obtained by averaging over independent realizations of equation (17).

5.1.1. *Example: periodic drift*

We applied the SAM to the SINE data set, choosing $r_{\max}=9/8$. For two different MC sample sizes, $K=100$ and $K=1000$, we used Brent's optimization algorithm (which combines parabolic interpolation with the golden section search algorithm; see section 10.2 of Press *et al.* (1992)) for finding the maximum of the log-likelihood. The MLE was estimated as 3.116, when $K=100$ (in 6 s), and 3.112 when $K=1000$ (in 57 s). Using numerical differentiation, the standard error was estimated as 0.04 when both $K=100$ and $K=1000$. Fig. 2(a) shows the estimate of the log-likelihood function based on the two different MC sample sizes.

When the model assumed is in $\mathcal{D}_2$, it is mathematically much more dificult to derive a formula that is analogous to equation (17) which allows for simultaneous estimation of the likelihood function. In EA2 the minimum m and the time τ when the minimum occurs both depend on θ. Since the Poisson rate r is a function of m as well as θ, and ω is simulated conditionally on (τ,m), the dependence of ω and Φ on θ is considerably more difficult than in EA1 to decouple and the method requires novel couplings and results from stochastic analysis. Such constructions are beyond the scope of this paper and are presented in Beskos *et al.* (2005b), where several other issues related with the SAM are also tackled. In particular it is proved that the simultaneous estimator for diffusions in $\mathcal{D}_1$ and $\mathcal{D}_2$ converges uniformly in θ to the likelihood function as the number of MC samples increases. Under standard assumptions, this guarantees consistency of the MC estimate of the MLE.

5.1.2. *Example: logistic growth*

We used the SAM in conjunction with the downhill simplex optimization method (section 10.4 of Press *et al.* (1992)) to locate the MLE of the LOG-GROWTH data set, for MC sample sizes $K=100$ and $K=1000$. The estimates corresponding to (R,Λ,σ) were $(0.1098, 1014.89, 0.10057)$ for $K=100$ (in 2 min) and $(0.1097, 1014.78, 0.10057)$ for $K=1000$ (in 20 min). Using numerical differentiation the curvature of the estimated log-likelihood at the estimated MLE was found:

$$\begin{pmatrix} 4567 & 0.072 & -9601 \\ & 0.011 & -0.05 \\ & & 198464 \end{pmatrix}.$$

The corresponding standard errors are (0.016, 9.5, 0.002). Fig. 2(b) shows the estimates of the profile log-likelihood of R for the two MC sample sizes.

6. The Poisson estimator

Corollary 1 relates the transition density of the diffusion to an expectation over the BB measure. Thus, any unbiased estimator of the expectation on the right-hand side of equation (11)

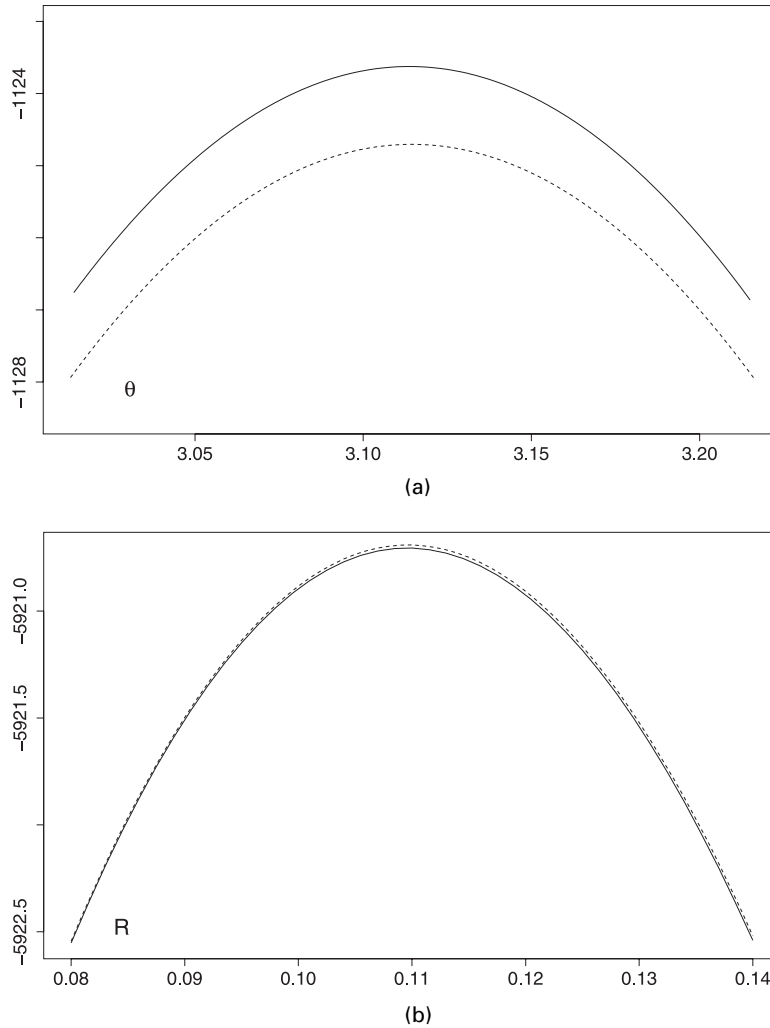

Fig. 2. Simultaneous estimate of (a) the log-likelihood θ for the SINE data set and (b) the profile log-likelihood R for the LOG-GROWTH data set, using the SAM with $K = 100$ (-------) and $K = 1000$ (———) MC samples

corresponds to an unbiased estimator of $\tilde{p}_t(x, y; \theta)$. Generalizing, suppose that the expectation

$$\mathbb{E}_{\mathbb{P}^{(t,x,y)}}\left[\exp\left\{-\int_0^t f(\omega_s)\,\mathrm{d}s\right\}\right] < \infty \tag{18}$$

is to be estimated for arbitrary continuous function f and diffusion bridge measure $\mathbb{P}^{(t,x,y)}$. For any $c \in \mathbf{R}$, $\lambda > 0$, and path ω, we write

$$\begin{aligned}\exp\left\{-\int_0^t f(\omega_s)\,\mathrm{d}s\right\} &= \exp(-ct)\sum_{j=0}^{\infty}\frac{1}{j!}\left\{\lambda t\int_0^t \frac{c - f(\omega_s)}{\lambda t}\,\mathrm{d}s\right\}^j \\ &= \exp\{(\lambda - c)t\}\,\mathbb{E}\left[\left\{\int_0^t \frac{c - f(\omega_s)}{\lambda t}\,\mathrm{d}s\right\}^K\right],\end{aligned}$$

where the expectation is taken with respect to $\kappa \sim \mathrm{Po}(\lambda t)$, conditionally on ω. If $\psi \sim \mathrm{Un}[0,t]$, then $\{c - f(\omega_\psi)\}/\lambda$ is an unbiased estimator of $\int_0^t \{c - f(\omega_s)\}/\lambda t \, \mathrm{d}s$. Let $\Psi = \{\psi_1, \ldots, \psi_\kappa\}$ be a Poisson process of rate λ on $[0,t]$, and $\omega \sim \mathbb{P}^{(t,x,y)}$. Then, we obtain the following simple unbiased estimator of expectation (18), which we call the *Poisson estimator*:

$$\exp\{(\lambda - c)t\}\lambda^{-\kappa} \prod_{j=1}^{\kappa} \{c - f(\omega_{\psi_j})\}. \tag{19}$$

This estimator was introduced in the context of statistical physics by Wagner (1988a, 1989). It can be derived from first principles that its second moment is

$$\exp\{(\lambda - 2c)t\} \, \mathbb{E}_{\mathbb{P}^{(t,x,y)}}\left(\exp\left[\frac{1}{\lambda}\int_0^t \{c - f(\omega_s)\}^2 \, \mathrm{d}s\right]\right), \tag{20}$$

which is not guaranteed to be finite. The choice of c and λ with the view of improving the efficiency of the algorithm is discussed in Section 7.

Taking $\mathbb{P}^{(t,x,y)} \equiv \mathbb{W}^{(t,x,y)}$, and $f = (\alpha^2 + \alpha')/2$, the Poisson estimator can be used with equation (11) to estimate $\tilde{p}_t(x, y; \theta)$. This transition density estimator will also be referred to as the Poisson estimator. A simultaneous estimation of the complete map $\theta \mapsto \tilde{p}_t\{\eta(v;\theta), \eta(w;\theta); \theta\}$ for any data points v and w merely requires decoupling of ω in expression (19) from x and y, since Ψ clearly is independent of θ. Since in the current context $\mathbb{P}^{(t,x,y)} \equiv \mathbb{W}^{(t,x,y)}$, it suffices to exploit the relocation invariance property of BBs and to rewrite expression (19) as

$$\exp\{(\lambda - c)t\}\lambda^{-\kappa} \prod_{j=1}^{\kappa} [c - f\{\omega_{\psi_j} + (1 - \psi_j/t)x + (\psi_j/t)y\}], \qquad \omega \sim \mathbb{W}^{(t,0,0)}. \tag{21}$$

This estimator will be referred to as the *simultaneous* Poisson estimator. For more general diffusion bridge measures, decoupling of ω from x and y can be cumbersome, since the intuitive relocation invariance property does not hold in general. Such a case is treated in the following section.

6.1. Inference for the Cox–Ingersoll–Ross model

We apply the Poisson estimator to infer about a diffusion process outside $\mathcal{D}_2$, the CIR diffusion model, which solves

$$\mathrm{d}V_s = \rho(\mu - V_s)\,\mathrm{d}s + \sigma\sqrt{V_s}\,\mathrm{d}B_s, \qquad \theta = (\rho, \mu, \sigma).$$

It is assumed that all parameters are positive and $2\rho\mu > \sigma^2$, which guarantees that V does not hit zero (see page 391 of Cox *et al.* (1985)). We fit the CIR model to a well-studied data set (used among others by Aït-Sahalia (1996), Elerian *et al.* (2001) and Roberts and Stramer (2001)) which contains daily euro–dollar rates between 1973 and 1995, to allow for comparisons with existing methods. We take a subsample of 550 values, corresponding to time intervals of 10 days, since the diffusion model seems more appropriate on that scale. We call this subsample the EURODOLLAR data set and plot it in Fig. 3(a).

In this case $\eta(u;\theta) = 2\sqrt{u}/\sigma$; thus the transformation $X_s = \eta(V_s;\theta)$ solves the SDE

$$\mathrm{d}X_s = \left\{\left(\frac{2\rho\mu}{\sigma^2} - \frac{1}{2}\right)\frac{1}{X_s} - \frac{\rho}{2}X_s\right\}\mathrm{d}s + \mathrm{d}B_s.$$

Setting $\alpha(u;\theta) = (2\rho\mu/\sigma^2 - 0.5)/u - \rho u/2$, it is easy to check that the CIR model is outside $\mathcal{D}_2$. X is almost surely positive and its distribution is absolutely continuous with respect to that of the Bessel process. Let $\mathbb{Q}_\theta^{(t,x,y)}$ and $\mathbb{R}^{(t,x,y)}$ be the bridge measures of X and the Bessel process

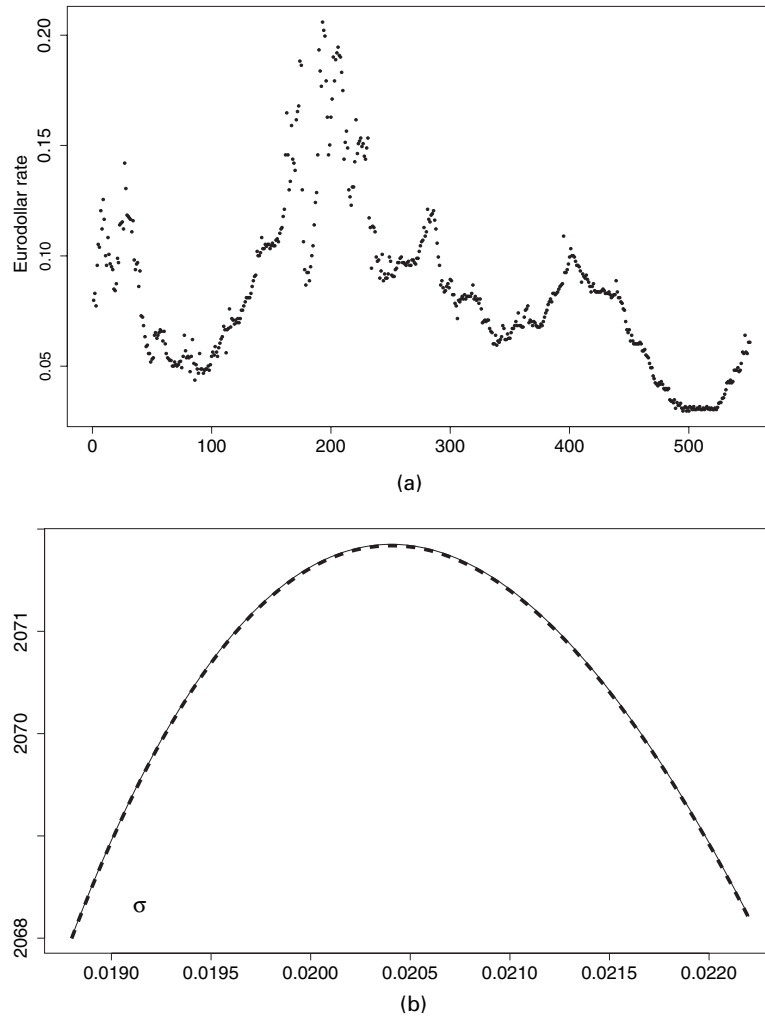

Fig. 3. (a) EURODOLLAR data set and (b) real (———) profile log-likelihood of σ and its estimation (– – –) by using the simultaneous Poisson estimator with $\lambda = c = 1$ and $K = 100$ MC samples

respectively, and $\tilde{p}_t(x, y; \theta)$ and $q_t(x, y)$ be the corresponding transition densities. Direct application of lemma 1 yields an analogue of equation (11) for positive diffusions:

$$\tilde{p}_t(x, y; \theta) = \frac{y}{x} q_t(x, y)\, \mathbb{E}_{\mathbb{R}^{(t,x,y)}}\left[\exp\left\{A(y;\theta) - A(x;\theta) - \int_0^t \frac{1}{2}(\alpha^2 + \alpha')(\omega_s;\theta)\,\mathrm{d}s\right\}\right]. \tag{22}$$

Thus, we can readily use the Poisson estimator to estimate the transition density of the CIR model for a fixed $\theta \in \Theta$. However, simultaneous estimation for all $\theta \in \Theta$ is non-trivial, since the decoupling of $\omega \sim \mathbb{R}^{(t,x,y)}$ from x and y is not straightforward. The appropriate construction is in fact a variation of the approach that we have devised for the SAM for diffusions in $\mathcal{D}_2$ and is contained in Beskos *et al.* (2005b). This stems from the characterization of the Bessel bridge as a BB that is conditioned to remain positive. As in the EA2 algorithm, we consider first the minimum m of the Bessel bridge and subsequently reconstruct the path given m. The distribu-

tion of m is the restriction on $(0, \infty)$ of the distribution of the minimum of the corresponding BB. Given m, the path can be reconstructed exactly as if it were a BB with known minimum m.

We applied the simultaneous Poisson estimator to the EURODOLLAR data set taking, after some pilot tuning, $\lambda = c = 1$. In Fig. 3(b) we show the real and the estimated profile log-likelihood for σ by using the downhill simplex optimization algorithm. Even with 100 MC samples the estimated curve coincides with the true curve, although it can be shown that the variance of the estimator is infinite in this case.

7. Comparison of different transition density estimators

We have introduced three different methods which can be used to estimate the diffusion transition density. Whereas the AM (and SAM) were devised solely for this purpose, it is important to recognize that the other two methods have greater scope. The Poisson estimator can estimate expectations of general diffusion exponential functionals (18). The bridge method is based on equation (12), which is a form of *Rao–Blackwellization* that can be used to estimate other conditional probabilities for diffusions, e.g. hitting probabilities.

In the context of transition density estimation, the weakness of the AM and the bridge method over the Poisson estimator is that their applicability is determined by that of the EA. On the contrary, the Poisson estimator can be used without assuming that the drift α is suitably bounded.

A significant advantage of the AM over the competitors is that it is guaranteed to have finite polynomial moments. For the bridge method little is known since we have not yet derived analytic expressions for its variance. Checking whether the variance of the Poisson estimator is finite is in general tedious. A notable exception is when the diffusion is in $\mathcal{D}_1$, where, since $f = (\alpha^2 + \alpha')/2$ is bounded, it is easy to see from expression (20) that the variance will be finite for any finite c and λ. Two examples outside $\mathcal{D}_1$ for which we know that this will not be the case are the CIR and the logistic growth models.

A feature of the bridge method which is likely to make it generally less efficient than the competitors is that the simulated paths ignore the final data point, since the method is founded on the *unconditional* EA.

When the aim is to explore features of the likelihood surface, it is imperative that the estimators yield estimates of the transition density simultaneously for all $\theta \in \Theta$. The bridge method is the only method for which we have been unable to achieve that.

Although derived from very different perspectives, the SAM and the simultaneous Poisson estimator are related. Taking $\lambda = r_{\max}$, $c = \lambda + l(\theta)$ and contrasting equation (17) with expression (21) reveals that, in $\mathcal{D}_1$, the SAM is a special case of the simultaneous Poisson estimator. Moreover, there are certain optimality properties of this choice. For any c such that $c > r(\theta) + l(\theta)$, expression (20) is bounded above by $\exp([-2c + \lambda + \{c - l(\theta)\}^2/\lambda]t)$. This quantity is minimized by any pair (λ, c) such that $c = \lambda + l(\theta)$, where $\lambda \geqslant r(\theta)$. Requiring that the Poisson estimator yields estimates simultaneously for all $\theta \in \Theta$, the computationally most efficient bound on the variance is achieved by the choice $\lambda = r_{\max}$ and $c = \lambda + l(\theta)$, under which the Poisson estimator and the SAM coincide. Note that, in this choice, λ is the range and c the maximum of the functional $(\alpha^2 + \alpha')(u; \theta)/2$ over all $u \in \mathbf{R}, \theta \in \Theta$. It is not obvious whether choosing $c > r(\theta) + l(\theta)$ is optimal. The connection between the two methods is less transparent outside $\mathcal{D}_1$. The rate of the Poisson process that is used in the SAM for diffusions in $\mathcal{D}_2$ will depend on the minimum of the path proposed, and it is precisely because of this dependence that the estimator is almost surely bounded. The Poisson process and the path proposed in the Poisson estimator are inherently independent. Guided by our findings for $\mathcal{D}_1$, we propose to

Table 1. Summary of 500 independent estimations of the MLE of the LOG-GROWTH data set†

Parameter	*Estimator*	*Minimum*	*1st quartile*	*Median*	*3rd quartile*	*Maximum*
R	SAM	0.1051	0.1083	0.1092	0.1100	0.1130
	Simultaneous Poisson	0.1030	0.1072	0.1087	0.1099	0.1139
Λ	SAM	1010.13	1013.91	1014.89	1016.07	1021.07
	Simultaneous Poisson	1008.88	1013.48	1014.93	1016.50	1022.80
σ	SAM	0.10045	0.10053	0.10056	0.10058	0.10065
	Simultaneous Poisson	0.10040	0.10050	0.10054	0.10058	0.10069

†Each estimation is based on $K = 10$ MC samples. For each parameter, the first row corresponds to the SAM and the second to the simultaneous Poisson estimator, with $\lambda = 1$ and $c = l(\theta) + \lambda$. The average time for each experiment was around five times higher in the SAM compared with the simultaneous Poisson estimator.

choose λ according to an estimate of the range of $(\alpha^2 + \alpha')(u; \theta)/2$ over all $u \in \mathbf{R}$, $\theta \in \Theta$, and to take $c = \lambda + l(\theta)$. This choice has proved successful empirically.

A brief study of the performance of the SAM and the simultaneous Poisson estimator in estimating the MLE of the LOG-GROWTH data set is summarized in Table 1. The parameters λ and c were chosen as suggested above. Taking into account its smaller computational cost, the simultaneous Poisson estimator is more efficient in this example. A feature that is not depicted in Table 1 is the sensitivity of the simultaneous Poisson estimator on the choice of c and λ. The method can produce unreliable estimates, especially with few MC samples, for certain choices of the tuning parameters. In contrast the SAM is fully automatic.

8. A Monte Carlo expectation–maximization approach

The EM algorithm (Dempster *et al.*, 1977) is suitable for locating the MLE when the likelihood of the observed data is intractable but the joint likelihood of the observed and some appropriately defined missing data is of a simple form. At the expense of some additional effort the EM algorithm can also be used to obtain an estimate of the observed information matrix. To fix ideas, let $\mathbf{v}_{\text{obs}} = \{V_0 = v, V_t = w\}$ be the observed data, which consist of discrete time observations from equation (1) (without loss of generality we assume two data points). It is assumed throughout this section that $V \in \mathcal{D}_2$. When the diffusion coefficient does not depend on unknown parameters, i.e. when $\sigma(\cdot; \theta) = \sigma(\cdot)$, we can treat the paths between consecutive observations as missing data and use Girsanov's formula (32) to form the *complete likelihood*. This naturally leads to an EM algorithm whose implementation is described in the next section. However, when the diffusion coefficient depends on θ, the missing paths should be appropriately transformed, as shown in Section 8.2.

8.1. Monte Carlo expectation–maximization for diffusions with known diffusion coefficient

We first transform the process $V_s \to X_s = \eta(V_s)$, where η is defined in equation (3) and, since by assumption $\sigma(\cdot; \theta) = \sigma(\cdot)$, η does not depend on θ. Thus, we define $\mathbf{x}_{\text{obs}} = \{X_0 = x, X_t = y\}, x = \eta(v)$ and $y = \eta(w)$; $\mathbf{x}_{\text{obs}}$ contains discrete time observations from the diffusion X with SDE (4). The observed log-likelihood is $l(\theta|\mathbf{x}_{\text{obs}}) = \log\{\tilde{p}_t(x, y; \theta)\}$. Let $\mathbf{x}_{\text{mis}}$ denote the missing

path, $(X_s, s \in [0,t])$, and $\mathbf{x}_{\text{com}} = \{\mathbf{x}_{\text{obs}}, \mathbf{x}_{\text{mis}}\}$ denote the complete data, which is the continuous diffusion path starting from $X_0 = x$ and finishing at $X_t = y$. The complete log-likelihood is given by Girsanov's formula (32),

$$l(\theta|\mathbf{x}_{\text{com}}) = A(y;\theta) - A(x;\theta) - \tfrac{1}{2}\int_0^t (\alpha^2 + \alpha')(X_s;\theta)\,\mathrm{d}s, \tag{23}$$

and

$$\mathbf{x}_{\text{mis}}|\mathbf{x}_{\text{obs}};\theta \sim \mathbb{Q}_\theta^{(t,x,y)}.$$

The E-step of the EM algorithm requires analytic evaluation of

$$Q(\theta,\theta') = \mathbb{E}_{\mathbf{x}_{\text{mis}}|\mathbf{x}_{\text{obs}};\theta'}\{l(\theta|\mathbf{x}_{\text{com}})\}, \qquad \theta \in \Theta.$$

However, note that we can introduce the random variable $U \sim \text{Un}[0,t]$, which is independent of both $\mathbf{x}_{\text{mis}}$ and $\mathbf{x}_{\text{obs}}$, and rewrite

$$Q(\theta,\theta') = A(y;\theta) - A(x;\theta) - \frac{t}{2}\,\mathbb{E}_{(\mathbf{x}_{\text{mis}},U)|\mathbf{x}_{\text{obs}};\theta'}\{(\alpha^2 + \alpha')(X_U;\theta)\}. \tag{24}$$

We cannot perform analytically the expectation above. However, since we can simulate U, and X_u for any u by using the EA, we adopt an MC implementation of the EM algorithm. The MCEM algorithm was introduced in Wei and Tanner (1990); convergence and implementation issues were tackled in Chan and Ledolter (1995), Sherman *et al.* (1999) and Fort and Moulines (2003). It is well documented (see for example Fort and Moulines (2003), and references therein) that the number of MC samples that are used to approximate the expectation should increase with the EM iterations. We have followed this approach in our examples. Several methods have been proposed for using the EM algorithm to estimate the observed information matrix; see for example Louis (1982), Jamshidian and Jennrich (2000) and Meng and Rubin (1991). We have used the method that was suggested by Louis (1982), where we use MC estimations of the required expectations, as in Wei and Tanner (1990).

8.1.1. Example: periodic drift

We applied MCEM to the SINE data set; the EM iterations are presented in the second column of Table 2. The M-step was implemented by using Brent's optimization method. The convergence is very rapid, essentially in one iteration, and the estimated MLE is in perfect agreement with the estimate that is obtained with the SAM in Section 5.1. The standard error was estimated to be 0.04.

8.2. The general case: missing information and a path transformation

An important result from stochastic analysis, which has found numerous applications in statistical inference for diffusion processes (see for instance Barndorff-Nielsen and Shephard (2002)), is that a continuous time diffusion path on an interval of time $[0,t]$ can be used to estimate perfectly the parameters that are involved in $\sigma(\cdot;\theta)$. A similar result does not hold for parameters in the drift $b(\cdot;\theta)$, where perfect estimation from continuous time data typically holds asymptotically as $t \to \infty$. However, a finite number of discrete time observations can contain only finite information about any of the parameters. The computational implication is that we cannot construct an EM algorithm as in Section 8.1 where the paths between the observed data are treated as missing data. According to the EM terminology, this augmentation scheme leads to a fraction of missing information equal to 1. This problem was first encountered in an MCMC framework by Roberts and Stramer (2001) and Elerian (1999).

Table 2. MCEM iterations for the SINE (second column) and the LOG-GROWTH (last three columns) data sets†

Iteration j	$\theta^{(j)}$	$R^{(j)}$	$\Lambda^{(j)}$	$\sigma^{(j)}$
0	0.500	0.7000	1500.00	0.7000
1	3.119	0.1099	1015.32	0.1006
2	3.114	0.1097	1014.80	0.1006
3	3.115	0.1097	1014.83	0.1006
4	3.115	0.1097	1014.51	0.1006
5	3.114	0.1097	1014.66	0.1006
6	3.113	0.1097	1014.76	0.1006
7	3.113	0.1097	1014.67	0.1006
8	3.113	0.1097	1014.75	0.1006
9	3.113	0.1097	1014.71	0.1006
10	3.113	0.1095	1014.78	0.1006

†The first five iterations were performed using 200 MC samples and the last five using 2000 samples. The execution times were 177 s for the SINE data and 18 min for the LOG-GROWTH data.

One remedy to this type of problems is to find a suitable invertible transformation of the missing data to reduce the augmented information. In particular, we want to transform the process so that the diffusion coefficient is independent of θ, exploiting the result that continuous time paths contain only finite information about the drift parameters. Our approach is in the spirit of Roberts and Stramer (2001). We start by transforming the process, $V_s \to X_s = \eta(V_s;\theta)$, where η is the transformation in equation (3); therefore $X_0 = x(\theta) = \eta(v;\theta)$ and $X_t = y(\theta) = \eta(w;\theta)$. Unlike in Section 8.1, $(X_s, s \in [0,t])$ is not directly observed; instead it is a function of the unknown parameters. The second level of path transformation is from $X_s \to \dot{X}_s$ where

$$\dot{X}_s := X_s - \left(1 - \frac{s}{t}\right) x(\theta) - \frac{s}{t} y(\theta), \qquad s \in [0,t]. \tag{25}$$

$\dot{X}$ is a diffusion bridge starting from $\dot{X}_0 = 0$ and finishing at $\dot{X}_t = 0$. Its dynamics depend on θ and are typically intractable; nevertheless it is easy to simulate $\dot{X}_s$ at any time $s \in [0,t]$, conditionally on $\mathbf{v}_{\text{obs}}$ and a specific value of θ, following the procedure

(a) find $x = x(\theta)$ and $y = y(\theta)$,
(b) simulate X_s from the bridge diffusion measure $\mathbb{Q}_\theta^{(t,x,y)}$ by using the EA and
(c) apply the transformation in expression (25).

The inverse transformation from $\dot{X}$ to X is

$$g_\theta(\dot{X}_s) := \dot{X}_s + \left(1 - \frac{s}{t}\right) x(\theta) + \frac{s}{t} y(\theta), \qquad s \in [0,t].$$

We propose an augmentation scheme where $\mathbf{v}_{\text{mis}} = (\dot{X}_s, s \in [0,t])$ and $\mathbf{v}_{\text{com}} = \{\mathbf{v}_{\text{obs}}, \mathbf{v}_{\text{mis}}\}$. The following lemma, which is proved in Appendix B, gives the complete log-likelihood.

Lemma 2. The log-likelihood of the complete data $\mathbf{v}_{\text{com}}$ is given by

$$l(\theta|\mathbf{v}_{\text{com}}) = \log|\eta'(w;\theta)| + \log[\mathcal{N}_t\{y(\theta) - x(\theta)\}] + A\{y(\theta);\theta\} - A\{x(\theta);\theta\}$$
$$- \int_0^t \tfrac{1}{2}(\alpha^2 + \alpha')\{g_\theta(\dot{X}_s);\theta\}\,\mathrm{d}s.$$

Arguing as in Section 8.1, we have that

$$Q(\theta,\theta') = \log|\eta'(w;\theta)| + \log[\mathcal{N}_t\{y(\theta)-x(\theta)\}] + A\{y(\theta);\theta\} - A\{x(\theta);\theta\} - \frac{t}{2}\,\mathbb{E}_{(\mathbf{v}_{\text{mis}},U)|\mathbf{v}_{\text{obs}};\theta'}[(\alpha^2+\alpha')\{g_\theta(\dot{X}_U);\theta\}].$$

We can now apply MCEM in a similar fashion to that in Section 8.1.

8.2.1. Example: logistic growth

We used MCEM to estimate the MLE and the observed information for the LOG-GROWTH data set. The results in Table 2 indicate rapid convergence. We used the Fletcher–Reeves version of the conjugate gradient method for the M-step of the algorithm (see section 10.5 of Press *et al.* (1992)). The standard errors are estimated using $K = 2000$ as (0.015, 9.5, 0.002).

9. Hierarchical simulation models and inference using Markov chain Monte Carlo methods

In this section we develop an MCMC algorithm for Bayesian inference for discretely observed diffusions. The stationary distribution of our MCMC algorithm is the *exact* posterior distribution of the parameters. However, it is not only the *exactness* which distinguishes our approach from competitive existing MCMC algorithms. Existing methods follow a path augmentation approach, as we did in the EM algorithm in Section 8. The missing continuous paths are approximated by a fine discrete time Markov chain, whose transition is assumed to follow an Euler-type approximation to the true diffusion transition. The joint distribution of the observed and missing data is given by an appropriate approximation of Girsanov's formula. Then, a Gibbs sampler (or more general componentwise updating MCMC algorithm) is used to sample from the approximate posterior distribution of the parameters and the missing paths. Often, for any two data points several points need to be imputed in between them to obtain a good approximation to the true posterior of θ. However, the performance of basic MCMC schemes can severely deteriorate as the amount of imputation increases; see Roberts and Stramer (2001) for details.

The approach that is introduced here is not based on augmentation of paths. Instead, we construct a graphical model which involves the variables that are used in the EA, and we show that the posterior distribution of the parameters is obtained as a marginal in this graphical model. Thus, we use an appropriate Metropolis–Hastings algorithm to sample from the joint posterior of all variables in the graph. One of the steps of the sampler involves running the *conditional* EA. The state space of our MCMC algorithm typically has much smaller dimension than the competing augmentation methods and as a result it can be computationally more efficient. However, a comparison between alternative MCMC methods is not carried out in this paper and will be reported elsewhere.

We describe in detail the MCMC algorithm when it is assumed that $V \in \mathcal{D}_1$. The derivation of the algorithm when $V \in \mathcal{D}_2$ is technically much more difficult, and it can be found in Beskos *et al.* (2004b). Nevertheless, we present simulation results for both cases. An essential ingredient of the algorithm is the following lemma which derives the density of the output of the conditional EA1 algorithm.

Lemma 3. Consider any two fixed points x and y. Let $\Phi = \{\Psi, \Upsilon\}$ be the marked Poisson process on $[0,t]\times[0,1]$ with rate $r(\theta)$ and number of points $\kappa \sim \text{Po}\{r(\theta)t\}$, which is used in EA1 for simulating from $\mathbb{Q}_\theta^{(t,x,y)}$. Let $\omega \sim \mathbb{W}^{(t,0,0)}$, and I be the acceptance indicator (16) which decides whether $(\omega_s + (1-s/t)x + (s/t)y, s \in [0,t])$ is accepted as a path from $\mathbb{Q}_\theta^{(t,x,y)}$. Then, the conditional density of ω and Φ given $\{I=1\}$, $\pi(\omega,\Phi|\theta,x,y,I=1)$, is

$$\frac{\exp[t\{1-r(\theta)\}]r(\theta)^{\kappa}}{a(x,y,\theta)}\prod_{j=1}^{\kappa}\mathbb{I}\left[\phi\left\{\omega_{\psi_j}+\left(1-\frac{\psi_j}{t}\right)x+\frac{\psi_j}{t}y;\theta\right\}<\upsilon_j\right], \tag{26}$$

with respect to the product measure $\Phi\times\mathbb{W}^{(t,0,0)}$, where Φ is the measure of a unit rate Poisson process on $[0,t]\times[0,1]$, and $a(x,y,\theta)$ is the acceptance probability of the EA1 algorithm.

To clarify the notation, in the remainder of the section (ω^*,Φ^*) represent the accepted random elements in EA1, with density $\pi(\omega^*,\Phi^*|\theta,x,y)$ given in expression (26); Φ^*, and ω^* at any collection of times, can be easily sampled by using the conditional EA1 algorithm.

Let $\pi(\theta)$ denote an appropriately specified prior density for θ. The aim is to sample from the posterior distribution of θ given diffusion observations $\mathbf{v}=\{V_{t_0},\ldots,V_{t_n}\}$, say $\pi(\theta|\mathbf{v})$, where $V\in\mathcal{D}_1$. It is convenient to define

$$x_i=x_i(\theta):=\eta(V_{t_i};\theta),\qquad i=0,\ldots,n, \tag{27}$$

where η is the transformation in equation (3). For a given θ, the x_is are discrete time samples from SDE (4). For $1\leqslant i\leqslant n$, let $\omega_i^*=(\omega_{i,s}^*,s\in[t_{i-1},t_i])$ and $\Phi_i^*=\{\Psi_i^*,\Upsilon_i^*\}$ be the accepted elements of the conditional EA1 algorithm applied on the time interval $[t_{i-1},t_i]$; $S(\omega_i^*)$ will denote the corresponding output skeleton. We construct the following hierarchical model, which is defined through the densities:

$$\begin{aligned}\theta&\sim\pi(\theta),\\ V_{t_i}|V_{t_{i-1}},\theta&\sim p_{\Delta t_i}(V_{t_{i-1}},V_{t_i};\theta),\\ (\omega_i^*,\Phi_i^*)|\theta,x_{i-1}(\theta),x_i(\theta)&\sim\pi\{\omega_i^*,\Phi_i^*|\theta,x_{i-1}(\theta),x_i(\theta)\}.\end{aligned} \tag{28}$$

Note that in this hierarchical model the observed data are in the middle of the hierarchy. We term expression (28) a *hierarchical simulation model*, to emphasize that it represents the order in which variables must be simulated to ensure that the output of the EA1 algorithm, ω_i^*, is indeed from $\mathbb{Q}_\theta^{(\Delta t_i,x_{i-1},x_i)}$. The posterior density of interest, $\pi(\theta|\mathbf{v})$, is a marginal of the joint posterior density of θ and the latent variables, $\pi(\theta,\{\omega_i^*,\Phi_i^*,1\leqslant i\leqslant n\}|\mathbf{v})$. We aim at sampling from this density via the Gibbs sampler. The conditional density of the latent variables is

$$\pi(\{\omega_i^*,\Phi_i^*,1\leqslant i\leqslant n\}|\theta,\mathbf{v})=\prod_{i=1}^{n}\pi\{\omega_i^*,\Phi_i^*|\theta,x_{i-1}(\theta),x_i(\theta)\}; \tag{29}$$

thus the pairs (ω_i^*,Φ_i^*) are conditionally independent with density given in expression (26). The key property of this hierarchical simulation model is that θ is independent of $\{(\omega_i^*,\Phi_i^*),1\leqslant i\leqslant n\}$ given the collection of skeletons $\{S(\omega_i^*),1\leqslant i\leqslant n\}$. Specifically, the conditional density of θ given the observed data and the latent variables is given in the following theorem.

Theorem 3. θ is conditionally independent of $\{(\omega_i^*,\Phi_i^*),1\leqslant i\leqslant n\}$ given $\mathbf{v}$ and $\{S(\omega_i^*),1\leqslant i\leqslant n\}$, with density $\pi(\theta|\{S(\omega_i^*),1\leqslant i\leqslant n\},\mathbf{v})$ proportional to

$$\begin{aligned}&\pi(\theta)\,r(\theta)\sum_{i=1}^{n}\kappa_i^*\exp\left[A\{x_n(\theta);\theta\}-A\{x_0(\theta);\theta\}-\{l(\theta)+r(\theta)\}\sum_{i=1}^{n}\Delta t_i\right]\\ &\quad\times\prod_{i=1}^{n}|\eta'\{x_i(\theta);\theta\}|\mathcal{N}_{\Delta t_i}\{x_i(\theta)-x_{i-1}(\theta)\}\\ &\quad\times\prod_{i=1}^{n}\left(\prod_{j=1}^{\kappa_i^*}\left[1-\phi\left\{\omega^*_{i,\psi^*_{i,j}}+(1-\psi^*_{i,j}/\Delta t_i)\,x_{i-1}(\theta)+(\psi^*_{i,j}/\Delta t_i)\,x_i(\theta);\theta\right\}\right]\right),\qquad\theta\in\Theta.\end{aligned} \tag{30}$$

Note that expression (30) is given in a simple computable form, and the intractable terms $a(x_{i-1}, x_i, \theta)$ are not involved when conditioning on $\{S(\omega_i^*), 1 \leqslant i \leqslant n\}$. Our proposed MCMC algorithm is a componentwise updating algorithm, simulating iteratively the skeletons $\{S(\omega_i^*), 1 \leqslant i \leqslant n\}$ and θ from their posterior conditional distributions. The skeletons $\{S(\omega_i^*), 1 \leqslant i \leqslant n\}$ are conditionally independent given θ and can be easily simulated by using the EA1 algorithm. In some cases it might be possible to sample from density (30) directly; otherwise a Metropolis–Hastings step could be used. When a random-walk Metropolis algorithm is being used, it will generally make sense to scale the proposal variance to be proportional to n^{-1}.

9.1. Example: periodic drift

We implemented the algorithm for the SINE data set. We used a uniform prior on $[0, 2\pi]$. It took 47 s to run the algorithm for 10 000 iterations, and some summaries are shown in the rightmost column of Fig. 4. The posterior mean is estimated as 3.1127 and the posterior standard deviation as 0.04. We used a Metropolis step to update θ, which had acceptance probability 0.49. The algorithm mixes very rapidly, and essentially the autocorrelation in the θ-series is because a Metropolis step is used rather than direct simulation from its conditional distribution. The dependence between θ and the latent variables is very weak.

When the assumed model is in $\mathcal{D}_2$, the construction of the MCMC algorithm is more complicated. In particular, derivation of the joint density of $(\theta, \omega^*, \Phi^*)$ is challenging owing to the more complex structure of the EA2 algorithm. This is done in Beskos *et al.* (2004b), where also other important issues related to the implementation of the MCMC algorithm are tackled. The algorithm necessitates recent non-centring reparameterization methodology for hierarchical models as described in Roberts *et al.* (2004) and Papaspiliopoulos *et al.* (2003).

9.2. Example: logistic growth

Using the MCMC algorithm that is described in Beskos *et al.* (2004b), we obtained 50 000 samples from the posterior distribution of the parameters (R, Λ, σ) for the LOG-GROWTH data set. The computing time was 20 min, and a summary of the results is given in Fig. 4. The posterior means were estimated as (0.1075, 1017.4, 0.1007) and the posterior precision matrix as

$$\begin{pmatrix} 4589 & 0.065 & -9408 \\ & 0.001 & 0.116 \\ & & 196060 \end{pmatrix}$$

from which the posterior standard deviations read as (0.015, 31.13, 0.002).

10. Conclusions

In this paper we have introduced a variety of methods which can be used for likelihood-based inference for discretely observed diffusions. The methods rely on recent advances in the exact simulation of diffusions. The computational efficiency of the methods was illustrated in a collection of examples. However, an exhaustive simulation study which tests the relative performance of our methods and existing approaches, under various model specifications and parameter settings, is not given here. Such a detailed empirical investigation is currently taking place. However, some qualitative remarks are made below.

In general, the computing time that is required for our methods critically depends on the rate of the Poisson process. *Ceteris paribus*, this rate is a function of the time increment Δt between

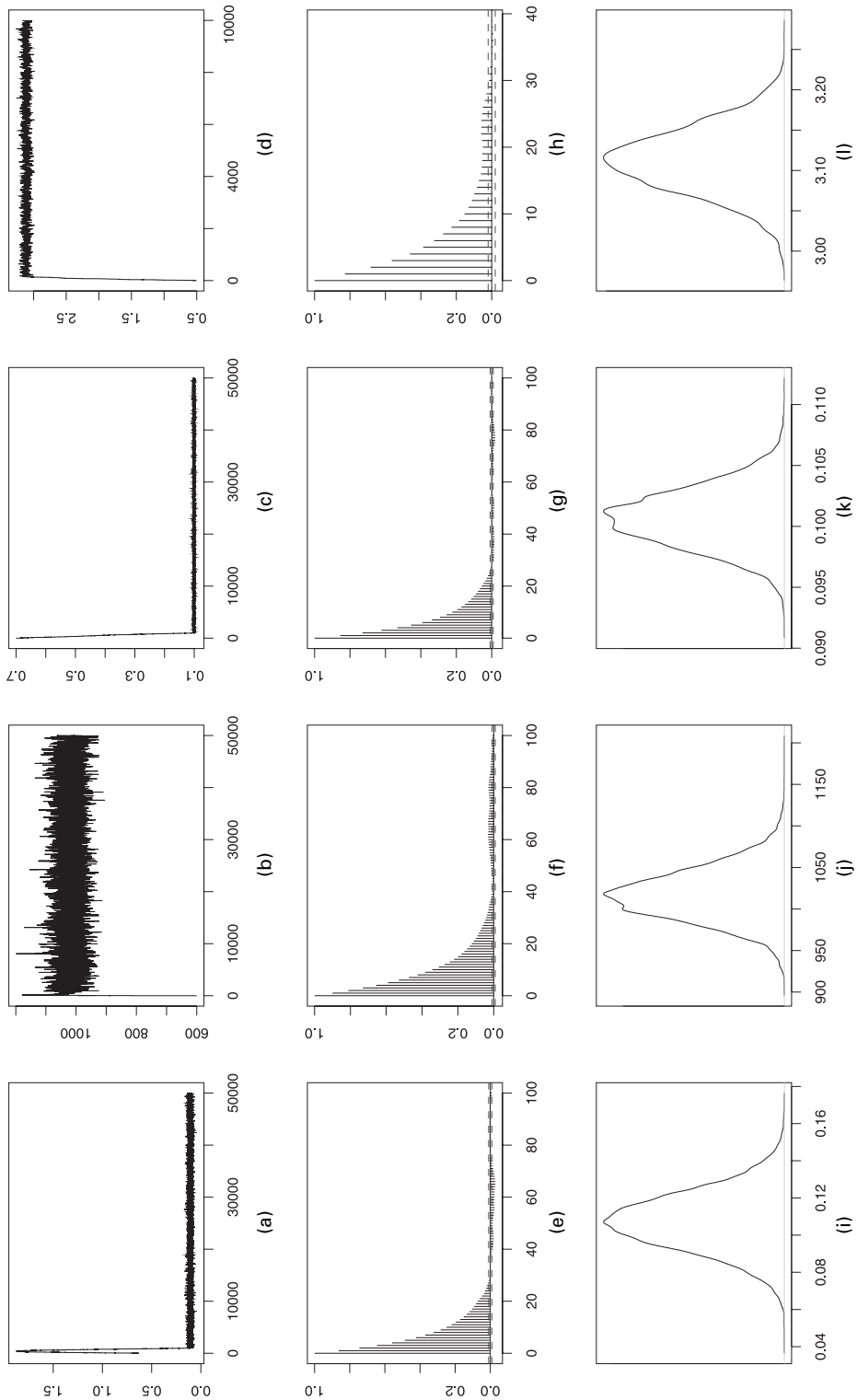

Fig. 4. Summary of MCMC results—(a)–(d) trace plots, (e)–(h) autocorrelation plots and (i)–(l) posterior density estimates: (d), (h), (l) 10 000 MCMC iterations for the SINE data set (the first 500 samples were removed as burn-in for autocorrelation and density estimation); (a)–(c), (e)–(g), (i)–(k) 50 000 MCMC iterations for the LOG-GROWTH data set (with a burn-in of 2000 iterations); (a), (e), (i) R; (b), (f), (j) Λ; (c), (g), (k) σ; (d), (h), (l) θ

the observations. The performance (measured either in computing time or MC error) of our methods will be very strong for small Δt (high frequency data). However, the performance of the methods as presented here deteriorates for sparser data sets (Δt large). Specifically, the problem stems from the following two characteristics of the conditional EA:

(a) the Poisson rate is linear in Δt (in EA2 it can increase even faster);
(b) the acceptance rate typically decreases exponentially to 0 in Δt.

The computing time that is required for the AM and SAM increases because of (a) and for the MCEM and MCMC algorithms because of (a) and (b). In the AM and SAM, the variance will also increase with Δt, since MC averages would be heavily dominated by the terms (of very small probability) corresponding to Poisson configurations with very few points. Similarly, expression (20) suggests that the variance of the Poisson estimator increases exponentially with Δt. In MCMC sampling, the computing cost can be transformed to be linear in Δt. This is achieved by augmenting any two observed data with additional points in between. This is implemented by applying the conditional EA on intervals of length a fraction of Δt. Of course, the additional augmentation will affect the MCMC mixing.

Some extensions of our methodology are possible. One important direction is the extension of our methods for diffusions outside the $\mathcal{D}_2$-class. As we have shown, the Poisson estimator can be readily used for likelihood inference for more general diffusion processes. However, current progress on the EA itself, which attempts to remove the boundedness conditions on the drift (Beskos *et al.* 2005a), is expected to broaden the applicability of the estimation methods. It is worth noting that MCEM and the AM remain unaltered irrespective of which version of the EA is being used; thus they can readily accommodate extensions in the EA. The other methods use explicitly the structure of the EA and will have to be modified appropriately.

We have concentrated here on the case where the diffusion is observed without error at a finite set of times. However, data often occur in different forms. For instance, data might be subject to observation error. All the methods for evaluating MLEs are difficult to extend to this case. However, the MCMC approach can be extended in a straightforward manner. An alternative interesting data form is when we observe a one-dimensional component of a higher dimensional diffusion, as for instance in continuous time filtering models. We are currently working on a collection of such filtering problems, and we have found that our methods can be extended to this case, although there are significant additional implementation challenges in this approach.

Our methodology extends some way to time inhomogeneous and multivariate diffusions. In carrying out these extensions, there are two steps in the arguments that are used in this paper that need to be generalized. Firstly, it is necessary to generalize the transformation (3) which eliminates the diffusion coefficient. This is straightforward under mild smoothness conditions on σ in the time inhomogeneous extension. For multivariate extensions, however, the generalized version of transformation (3) involves the solution of an appropriate vector differential equation which is often intractable or insolvable (see for example Aït-Sahalia (2004)). This imposes restrictions on the class of multivariate diffusions to which the methods that are presented here can currently be applied. Secondly, we need to eliminate the stochastic integral from Girsanov's formula (32) to derive lemma 1. Again this is fairly routine in the time inhomogeneous case, whereas an extra condition is needed in the multivariate case, requiring that the multivariate drift be the gradient of a suitable potential function. This is a well-known condition in stochastic analysis and for ergodic, unit diffusion coefficient diffusions corresponds to reversibility.

It is natural to ask whether the ideas of this paper extend to SDEs that are driven by Lévy processes. However, the Cameron–Martin–Girsanov formula, providing a closed form likelihood

function, is critical to all our methodology. Unfortunately, there is not an analogous expression for the Radon–Nikodym derivative of an infinite activity Lévy-driven SDE with respect to a measure that is tractable and easy to simulate from. Nevertheless, it is straightforward to extend our methods to incorporate SDEs with jumps according to a finite activity Lévy process (as for example considered in Roberts *et al.* (2004)).

We hope that the collection of techniques that are described in this paper will have further applications. We are currently working on MC estimation of derivative hedge ratios in finance. The methodology builds on Rao–Blackwellization techniques such as those devised for the bridge method.

Although we recognize that the mathematical details of this work are complex for those without a working knowledge of diffusion theory, we firmly believe that our methods have the potential to influence applied statistical work. This is because the algorithms that we use are relatively simple and easy to code, and the methods are not computationally demanding and are capable of handling long time series. In addition, motivated by the desire to make our work as accessible as possible, we have started to develop generic software for the implementation of some of our methods, beginning with the EA.

Acknowledgements

The authors thank Mike Pitt and Neil Shephard for helpful suggestions. We are grateful to all the referees for their comments which contributed to the improvement of the paper. The first author acknowledges financial support from the Greek State Scholarships Foundation and the second author is funded by Engineering and Physical Sciences Research Council grant GR/S61577/01.

Appendix A: Background material

A.1. Brownian motion and Brownian bridge

Standard Brownian motion $\{B_s; 0 \leqslant s \leqslant t\}$, $B_0 = 0$, can be simulated on a collection of time instances $0 = s_0 < s_1 < \ldots < s_k \leqslant t$ in the following recursive way:

$$B_{s_j} = B_{s_{j-1}} + Y_j, \qquad Y_j \sim N(0, s_j - s_{j-1}), \quad 1 \leqslant j \leqslant k.$$

A standard BB with distribution $\mathbb{W}^{(t,0,0)}$ can be obtained as a transformation of B:

$$B_s - \frac{s}{t} B_t, \qquad s \in [0, t].$$

An arbitrary BB with distribution $\mathbb{W}^{(t,x,y)}$ can be simulated via the relocation invariance property of BBs: if $\omega \sim \mathbb{W}^{(t,0,0)}$, then $\omega_s + (1 - s/t)x + (s/t)y$ is a path from $\mathbb{W}^{(t,x,y)}$.

A.2. Decomposition of a Brownian path at its minimum

The following theorem, which was stated in Beskos *et al.* (2004a), shows how to simulate a BB path together with its minimum m and the time τ when the minimum is attained.

Theorem 4. A path $\omega \sim \mathbb{W}^{(t,x,y)}$ can be simulated in the following way.

Step 1: simulate m and τ according to their joint density

$$\pi(m, \tau) = \sqrt{(2\pi t)} \frac{(y-m)(x-m)}{\sqrt{\{\tau^3 (t-\tau)^3\}}} \exp\left\{ -\frac{(x-m)^2}{2\tau} - \frac{(y-m)^2}{2(t-\tau)} + \frac{(y-x)^2}{2t} \right\}, \tag{31}$$

for $m \leqslant \min(x, y)$ and $\tau \in [0, t]$.

Step 2: the rest of the path is a transformation of six independent standard BBs, $b_j \sim \mathbb{W}^{(1,0,0)}$, $j=1,\ldots,6$. Analytically, for $0 \leqslant s \leqslant \tau$,

$$\omega_s = m + \sqrt{\tau}\sqrt{\left[\left\{\frac{(x-m)(\tau-s)}{\tau^{3/2}} + b_{1,(1-s/\tau)}\right\}^2 + b^2_{2,(1-s/\tau)} + b^2_{3,(1-s/\tau)}\right]},$$

and, for $\tau < s \leqslant t$,

$$\omega_s = m + \sqrt{(t-\tau)}\sqrt{\left[\left\{\frac{(y-m)(s-\tau)}{(t-\tau)^{3/2}} + b_{4,(s-\tau)/(t-\tau)}\right\}^2 + b^2_{5,(s-\tau)/(t-\tau)} + b^2_{6,(s-\tau)/(t-\tau)}\right]}.$$

In step 1, m is drawn from its marginal distribution, which is a transformed Rayleigh distribution and, conditionally on m, τ is drawn by using the inverse cumulative distribution function method.

A.3. Bessel process and Bessel bridge

The (three-dimensional) Bessel process or bridge can be defined as a scalar Brownian motion or bridge that is constrained to be positive; for the main results see chapter 11 of Revuz and Yor (1994). Recall that an arbitrary Bessel bridge measure is denoted by $\mathbb{R}^{(t,x,y)}$. Simulation of $\omega \sim \mathbb{R}^{(t,0,y)}$ can be carried out by means of three independent BBs (see for example Bertoin and Pitman (1994)):

$$\omega_s = \sqrt{\{(ys + b_{1,s})^2 + b^2_{2,s} + b^2_{3,s}\}}, \qquad 0 \leqslant s \leqslant t, \quad b_j \sim \mathbb{W}^{(t,0,0)}.$$

Simulation of $\omega \sim \mathbb{R}^{(t,x,y)}$ is more involved and follows the lines of theorem 4. We first simulate its minimum m, whose distribution is the restriction on $(0,\infty)$ of the distribution of the minimum of the corresponding BB. Given m, the path can be reconstructed exactly as if it were a BB with known minimum m.

Appendix B: Proofs of main results

B.1. Proof of lemma 1

Girsanov's formula, which gives the density of the law of X with respect to Wiener measure, is

$$\exp\left\{\int_0^t \alpha(\omega_s;\theta)\,\mathrm{d}\omega_s - \frac{1}{2}\int_0^t \alpha^2(\omega_s;\theta)\,\mathrm{d}s\right\}. \tag{32}$$

Thus, using Itô's lemma

$$\begin{aligned}\frac{\mathrm{d}\mathbb{Q}_\theta^{(t,x,y)}}{\mathrm{d}\mathbb{W}^{(t,x,y)}}(\omega) &= \frac{\mathcal{N}_t(y-x)}{\tilde{p}_t(x,y;\theta)}\exp\left\{\int_0^t \alpha(\omega_s;\theta)\,\mathrm{d}\omega_s - \frac{1}{2}\int_0^t \alpha^2(\omega_s;\theta)\,\mathrm{d}s\right\}\\ &= \frac{\mathcal{N}_t(y-x)}{\tilde{p}_t(x,y;\theta)}\exp\left\{A(y;\theta) - A(x;\theta) - \int_0^t \frac{1}{2}(\alpha^2+\alpha')(\omega_s;\theta)\,\mathrm{d}s\right\}.\end{aligned}$$

B.2. Proof of lemma 2

We shall find the density $\pi(V_t, \dot{X}|\theta)$ of the law of $\mathbf{v}_{\mathrm{com}}$ with respect to Leb $\times\, \mathbb{W}^{(t,0,0)}$, where Leb denotes the Lebesgue measure on $\mathbf{R}$. Clearly,

$$\pi(V_t, \dot{X}|\theta) = p_t(V_0, V_t;\theta)\,\pi(\dot{X}|V_t,\theta).$$

The law of the 1–1 transformation $g_\theta(\dot{X})$ given V_t is just $\mathbb{Q}_\theta^{(t,x,y)}$, where $x = \eta(V_0;\theta)$ and $y = \eta(V_t;\theta)$. Thus,

$$\pi(\dot{X}|V_t,\theta) = \frac{\mathrm{d}\mathbb{Q}_\theta^{(t,x,y)}}{\mathrm{d}\mathbb{W}^{(t,x,y)}} g_\theta(\dot{X}).$$

Use of lemma 1 concludes the proof.

B.3. Proof of lemma 3

Note that the Radon–Nikodym derivative of a homogeneous Poisson process with rate $r(\theta)$ on $[0,t]\times[0,1]$ with respect to Φ is

$$\exp\{-r(\theta)t\}\frac{r(\theta)^{\kappa}t^{\kappa}}{\kappa!}\left\{\exp(-t)\frac{t^{\kappa}}{\kappa!}\right\}^{-1},$$

which simplifies to $\exp[t\{1-r(\theta)\}]r(\theta)^{\kappa}$.

B.4. Proof of theorem 3

The joint density of θ and the latent variables conditionally on $\mathbf{v}$ can be decomposed as

$$\pi(\theta, \{\omega_i^*, \Phi_i^*, 1 \leqslant i \leqslant n\}|\mathbf{v}) = \pi(\theta|\mathbf{v})\,\pi(\{\omega_i^*, \Phi_i^*, 1 \leqslant i \leqslant n\}|\theta, \mathbf{v}).$$

To simplify the formulae, we write x_i instead of $x_i(\theta)$. Clearly,

$$\pi(\theta|\mathbf{v}) \propto \pi(\theta)\prod_{i=1}^{n}\tilde{p}_{\Delta t_i}(x_{i-1}, x_i; \theta)|\eta'(x_i;\theta)|.$$

Using equation (14), the acceptance probability of EA1 $a(x_{i-1}, x_i, \theta)$ can be expressed as

$$\tilde{p}_{\Delta t_i}(x_{i-1}, x_i; \theta)\,[\mathcal{N}_{\Delta t_i}(x_i - x_{i-1})\,\exp\{A(x_i;\theta) - A(x_{i-1};\theta) - l(\theta)\Delta t_i\}]^{-1}.$$

Combining this expression with expressions (29) and (26) we find that $\pi(\theta, \{\omega_i^*, \Phi_i^*, 1 \leqslant i \leqslant n\}|\mathbf{v})$ is proportional to

$$\pi(\theta)\,r(\theta)\sum\kappa_i^*\exp\left[A(x_n;\theta) - A(x_0;\theta) - \{l(\theta) + r(\theta)\}\sum_{i=1}^{n}\Delta t_i\right]\prod_{i=1}^{n}|\eta'(x_i;\theta)|\mathcal{N}_{\Delta t_i}(x_i - x_{i-1})$$
$$\times\prod_{i=1}^{n}\left(\prod_{j=1}^{\kappa_i^*}\mathbb{I}\left[\phi\left\{\omega^*_{i,\psi^*_{i,j}} + (1-\psi^*_{i,j}/\Delta t_i)x_{i-1} + (\psi^*_{i,j}/\Delta t_i)x_i;\theta\right\} < \upsilon^*_{i,j}\right]\right).$$

It is now straightforward to integrate out $\Upsilon_i^* = \{\upsilon_{i,j}, 1 \leqslant j \leqslant \kappa_i^*\}$ and to obtain expression (30).

References

Aït-Sahalia, Y. (1996) Testing continuous-time models of the spot interest rate. *Rev. Finan. Stud.*, **9**, 385–426.

Aït-Sahalia, Y. (2002) Maximum likelihood estimation of discretely sampled diffusions: a closed-form approximation approach. *Econometrica*, **70**, 223–262.

Aït-Sahalia, Y. (2004) Closed-form likelihood expansions for multivariate diffusions. *Working Paper*. Princeton University, Princeton. (Available from `http://www.princeton.edu/~yacine/research.htm`.)

Asmussen, S., Glynn, P. and Pitman, J. (1995) Discretization error in simulation of one-dimensional reflecting Brownian motion. *Ann. Appl. Probab.*, **5**, 875–896.

Barndorff-Nielsen, O. E. and Shephard, N. (2002) Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *J. R. Statist. Soc.* B, **64**, 253–280.

Bertoin, J. and Pitman, J. (1994) Path transformations connecting Brownian bridge, excursion and meander. *Bull. Sci. Math.*, **118**, 147–166.

Beskos, A., Papaspiliopoulos, O. and Roberts, G. O. (2004a) Retrospective exact simulation of diffusion sample paths with applications. To be published. (Available from `http://www.maths.lancs.ac.uk/~papaspil/research.html`.)

Beskos, A., Papaspiliopoulos, O. and Roberts, G. O. (2004b) Markov chain Monte Carlo for exact inference for discretely observed diffusions. To be published.

Beskos, A., Papaspiliopoulos, O. and Roberts, G. O. (2005a) A new factorisation of diffusion measure with view towards simulation. To be published.

Beskos, A., Papaspiliopoulos, O. and Roberts, G. O. (2005b) Monte carlo maximum likelihood estimation for discretely observed diffusion processes. To be published.

Beskos, A. and Roberts, G. O. (2005) Exact simulation of diffusions. *Ann. Appl. Probab.*, **15**, in the press.

Bibby, B. M., Jacobsen, M. and Sørensen, M. (2002) Estimating functions for discretely sampled diffusion-type models. In *Handbook of Financial Econometrics* (eds Y. Aït Sahalia and L. P. Hansen). Amsterdam: North-Holland.

Black, F. and Scholes, M. S. (1973) The pricing of options and corporate liabilities. *J. Polit. Econ.*, **81**, 637–659.

Chan, K., Karolyi, A. G., Longstaff, F. A. and Sanders, A. B. (1992) An empirical comparison of alternative models of the short-term interest rate. *J. Finan.*, **47**, 1209–1227.

Chan, K. S. and Ledolter, J. (1995) Monte Carlo EM estimation for time series models involving counts. *J. Am. Statist. Ass.*, **90**, 242–252.

Cox, J. C., Ingersoll, Jr, J. E. and Ross, S. A. (1985) A theory of the term structure of interest rates. *Econometrica*, **53**, 385–407.

Dacunha-Castelle, D. and Florens-Zmirou, D. (1986) Estimation of the coefficients of a diffusion from discrete observations. *Stochastics*, **19**, 263–284.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977) Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J. R. Statist. Soc.* B, **39**, 1–38.

Durham, G. B. and Gallant, A. R. (2002) Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes (with discussion). *J. Bus. Econ. Statist.*, **20**, 297–338.

Elerian, O. (1999) Simulation estimation of continuous time series models with applications to finance. *PhD Thesis*. Nuffield College, Oxford.

Elerian, O., Chib, S. and Shephard, N. (2001) Likelihood inference for discretely observed nonlinear diffusions. *Econometrica*, **69**, 959–993.

Eraker, B. (2001) MCMC analysis of diffusion models with application to finance. *J. Bus. Econ. Statist.*, **19**, 177–191.

Fort, G. and Moulines, E. (2003) Convergence of the Monte Carlo expectation maximization for curved exponential families. *Ann. Statist.*, **31**, 1220–1259.

Gallant, A. R. and Long, J. R. (1997) Estimating stochastic differential equations efficiently by minimum chi-squared. *Biometrika*, **84**, 125–141.

Gillespie, D. T. (1976) A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Computnl Phys.*, **22**, 403–434.

Gillespie, D. T. (1977) Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, **81**, 2340–2361.

Gobet, E. (2002) LAN property for ergodic diffusions with discrete observations. *Ann. Inst. H. Poincaré Probab. Statist.*, **38**, 711–737.

Goel, N. S. and Richter-Dyn, N. (1974) *Stochastic Models in Biology*. New York: Academic Press.

Gourieroux, C. and Jasiak, J. (2003) Multivariate Jacobi process and smooth transitions with applications. *J. Econ.*, to be published.

Gourieroux, C., Monfort, A. and Renault, E. (1993) Indirect inference. *J. Appl. Econometr.*, **8**, 85–118.

Jamshidian, M. and Jennrich, R. I. (2000) Standard errors for EM estimation. *J. R. Statist. Soc.* B, **62**, 257–270.

Kessler, M. (1997) Estimation of an ergodic diffusion from discrete observations. *Scand. J. Statist.*, **24**, 211–229.

Kimura, M. and Ohta, T. (1971) *Theoretical Aspects of Population Genetics*. Princeton: Princeton University Press.

Kingman, J. F. C. (1993) *Poisson Processes*. Oxford: Clarendon.

Kloeden, P. and Platen, E. (1995) *Numerical Solution of Stochastic Differential Equations*. New York: Springer.

Louis, T. A. (1982) Finding the observed information matrix when using the EM algorithm. *J. R. Statist. Soc.* B, **44**, 226–233.

McAdams, H. and Arkin, A. (1997) Stochastic mechanisms in gene expression. *Proc. Natn. Acad. Sci. USA*, **94**, 814–819.

Meng, X.-L. and Rubin, D. B. (1991) Using EM to obtain asymptotic variance-covariance matrices: the SEM algorithm. *J. Am. Statist. Ass.*, **86**, 899–909.

Merton, R. C. (1971) Optimum consumption and portfolio rules in a continuous-time model. *J. Econ. Theory*, **3**, 373–413.

Obuhov, A. M. (1959) Description of turbulence in terms of Lagrangian variables. In *Advances in Geophysics*, vol. 6, pp. 113–116. New York: Academic Press.

Papaspiliopoulos, O. and Roberts, G. O. (2004) Retrospective MCMC methods for Dirichlet process hierarchical models. To be published.

Papaspiliopoulos, O., Roberts, G. O. and Sköld, M. (2003) Non-centered parameterizations for hierarchical models and data augmentation (with discussion). In *Bayesian Statistics 7* (eds J. M. Bernardo, M. S. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith and M. West), pp. 307–326. New York: Oxford University Press.

Pardoux, É. and Pignol, M. (1984) Étude de la stabilité de la solution d'une EDS bilinéaire à coefficients périodiques: application au mouvement des pales d'hélicoptère. *Lect. Notes Control Inform. Sci.*, **63**, 92–103.

Pedersen, A. R. (1995) Consistency and asymptotic normality of an approximate maximum likelihood estimator for discretely observed diffusion processes. *Bernoulli*, **1**, 257–279.

Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. P. (1992) *Numerical Recipes in C: the Art of Scientific Computing*, 2nd edn. Cambridge: Cambridge University Press.

Revuz, D. and Yor, M. (1994) *Continuous Martingales and Brownian Motion*, 2nd edn. Berlin: Springer.

Roberts, G. O., Papaspiliopoulos, O. and Dellaportas, P. (2004) Bayesian inference for non-Gaussian Ornstein–Uhlenbeck stochastic volatility processes. *J. R. Statist. Soc.* B, **66**, 369–393.

Roberts, G. O. and Stramer, O. (2001) On inference for partially observed nonlinear diffusion models using the Metropolis-Hastings algorithm. *Biometrika*, **88**, 603–621.

Santa-Clara, P. (1995) Simulated likelihood estimation of diffusions with an application to the short-term interest rate. *Working Paper*. National Bureau of Economic Research, Cambridge.

Sherman, R. P., Ho, Y.-Y. K. and Dalal, S. R. (1999) Conditions for convergence of Monte Carlo EM sequences with an application to product diffusion modeling. *Econ. J.*, **2**, 248–267.
Shiga, T. (1985) Mathematical results on the stepping stone model in population genetics. In *Population Genetics and Molecular Evolution*, pp. 267–279. Tokyo: Japan Scientific Society Press.
Sørensen, H. (2004) Parametric inference for diffusion processes observed at discrete points in time: a survey. *Int. Statist. Rev.*, **72**, 337–354.
Wagner, W. (1988a) Unbiased multi-step estimators for the Monte Carlo evaluation of certain functional integrals. *J. Computnl Phys.*, **79**, 336–352.
Wagner, W. (1988b) Monte Carlo evaluation of functionals of solutions of stochastic differential equations: variance reduction and numerical examples. *Stochast. Anal. Appl.*, **6**, 447–468.
Wagner, W. (1989) Unbiased Monte Carlo estimators for functionals of weak solutions of stochastic differential equations. *Stochast. Stochast. Rep.*, **28**, 1–20.
Wei, G. C. G. and Tanner, M. A. (1990) A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. *J. Am. Statist. Ass.*, **85**, 699–704.

Discussion on the paper by Beskos, Papaspiliopoulos, Roberts and Fearnhead

Eric Moulines (*Ecole Nationale Supérieure des Télécommunications, Paris*)

The problem of estimation and inference for discretely observed diffusion has been studied extensively in recent years. The classical approach to the problem is based on a first-order approximation. In many applications, this approximation is not sufficiently accurate for sampling instants at which data are available. This paper belongs to the class of Monte Carlo maximum likelihood methods which was introduced by Pedersen (1995) and later considerably refined by many researchers.

Suppose that we wish to approximate the transition density of the diffusion over an interval. The first-order approximation is accurate if this interval is sufficiently short. If this is not so, we may partition the interval by introducing fictitious observations in such a way that the first-order approximation is sufficiently accurate on each subinterval. These fictitious points are then marginalized. This is most often carried out by using importance sampling techniques, much of the effort being spent on the design of an appropriate proposal distribution (see for instance Elerian *et al.* (2001) and the references therein). Although this approach can come arbitrarily close to the true transition density, its implementation remains computationally cumbersome.

The present paper takes a different direction. There are two main contributions in it. First, the authors propose an exact algorithm to sample trajectories of non-linear diffusion. Using the exact algorithm they propose several novel Monte Carlo estimators of the likelihood functions. These methods are extremely ingenious and have considerable appeal; however, they are still limited in scope, because the exact algorithm is available only under restrictive conditions on the drift terms of an appropriately transformed version of the diffusion. Second, the authors propose a novel estimator of the likelihood function which replaces the discretization and importance sampling estimates by a clever Poisson approximation. This method can be applied to a much wider class of diffusions, alleviating some of the problems that are linked to the discretization issue (see the discussion below). These are both potentially very useful pieces of practical statistical methodology and the authors are to be congratulated for these findings. I would like to raise two points.

(a) My first question concerns the potential benefits of the Poisson approximation compared with the classical importance sampling estimator. The transition density of the diffusion

$$\mathrm{d}X_s = \alpha(X_s;\theta)\,\mathrm{d}s + \mathrm{d}B_s$$

may be expressed as

$$\begin{aligned}\tilde{p}_t(x,y;\theta) &= \mathcal{N}_t(y-x)\,\mathbb{E}_{\mathbb{W}^{(t,x,y)}}\left[\exp\left\{A(y;\theta)-A(x;\theta)-\int_0^t \tfrac{1}{2}(\alpha^2+\alpha')(\omega_s;\theta)\,\mathrm{d}s\right\}\right]\\ &= \mathcal{N}_t(y-x)\,\mathbb{E}_{\mathbb{W}^{(t,x,y)}}\left[\exp\left\{\int_0^t \alpha(\omega_s;\theta)\,\mathrm{d}\omega_s - \tfrac{1}{2}\int_0^t \alpha^2(\omega_s;\theta)\,\mathrm{d}s\right\}\right]\end{aligned}$$

where $A(u;\theta)=\int^u \alpha(s;\theta)\,\mathrm{d}s$ and $\mathbb{W}^{(t,x,y)}$ denotes the probability of a Brownian bridge pinned at x and y. This estimator immediately suggests a 'plain' Monte Carlo estimator of the transition density,

$$\hat{p}_t(x,y;\theta) = \frac{\mathcal{N}_t(y-x)}{K}\sum_{k=1}^{K}\exp\left\{\int_0^t \alpha(\omega_s^{(k)};\theta)\,\mathrm{d}\omega_s^{(k)} - \tfrac{1}{2}\int_0^t \alpha^2(\omega_s^{(k)};\theta)\,\mathrm{d}s\right\},$$

where $\omega^{(k)} \sim \mathbb{W}^{(t,x,y)}$. Of course this estimator cannot be computed. The classical approach is to approximate the inner integral by Riemann sums.

Despite their appearance, the Poisson estimator shares similarities with the Riemann approximation, with the inverse of the intensity λ^{-1} (the mean distance between two successive samples) playing a role that is similar to the discretization interval δ. It would be of interest to understand why, for the same average number of samples ($\delta = 1/\lambda$), the Poisson approximation should dominate the 'natural' Riemann estimator (at least in the asymptotic regime when $\delta \to 0$ and $\lambda \to \infty$).

(b) My second question deals with the Monte Carlo EM algorithm. I focus here on the case where the diffusion term does not depend on the parameter for simplicity. When using the Monte Carlo EM algorithm, we must estimate conditional expectations

$$\mathbb{E}_{\mathbb{Q}_{\theta'}^{(t,x,y)}}\left\{\int_0^t f(X_s;\theta)\,\mathrm{d}s\right\},$$
$$f(x;\theta) = \tfrac{1}{2}\{\alpha^2(x;\theta) + \alpha'(x;\theta)\},$$

where $\mathbb{Q}_{\theta'}^{(t,x,y)}$ denotes the distribution of the process conditioned to start at $X_0 = x$ and to finish at $X_t = y$. We may estimate these quantities by using a plain Monte Carlo technique, but this supposes that we may obtain exact samples from $\mathbb{Q}_{\theta'}^{(t,x,y)}$, which is not always possible. Moreover, this is presumably not a very clever choice from the computational point of view. Assume that we are willing to estimate $\int \Phi(x)\,\pi(\mathrm{d}x)$. If the samples for π are generated from an instrumental distribution q by using the accept–reject algorithm, then after K draws from the *instrumental distribution* the Monte Carlo estimate is given by

$$\sum_{k=1}^{K} \frac{\mathbb{1}_{\{U_k \leqslant \pi(\xi_k)/M_q(\xi_k)\}}}{\sum_{l=1}^{K} \mathbb{1}_{\{U_l \leqslant \pi(\xi_l)/q(\xi_l)\}}}\Phi(\xi_k), \qquad \xi_k \sim q, \quad U_k \sim U[0,1],$$

where M is some constant satisfying $M\,q(x) > \pi(x)$ for all x. It is in general more sensible to use an importance sampling estimator,

$$\sum_{k=1}^{K} \frac{\pi(\xi_k)/q(\xi_k)}{\sum_{l=1}^{K} \pi(\xi_l)/q(\xi_l)}\Phi(\xi_k).$$

As suggested by Gelman (1995) and Quintana *et al.* (1999), nothing prevents us from approximating the intermediate quantities in the Monte Carlo EM algorithm by using the importance sampling estimator. The situation is more complicated here, because the importance weights and the function Φ cannot be computed exactly. Indeed, if the Brownian bridge is chosen to be the importance distribution, the importance sampling estimate of $\mathbb{E}_{\mathbb{Q}_{\theta'}^{(t,x,y)}}\{\int_0^t f(X_s;\theta)\,\mathrm{d}s\}$ takes the form

$$\sum_{k=1}^{K} \int_0^t f(\omega_s^{(k)};\theta)\,\mathrm{d}s \frac{\exp\left\{\int_0^t f(\omega_s^{(k)})\,\mathrm{d}s\right\}}{\sum_{l=1}^{K} \exp\left\{\int_0^t f(\omega_s^{(k)})\,\mathrm{d}s\right\}},$$

which involves integrals. Nevertheless, this expression can be approximated numerically by replacing integrals by Riemann sums or by the Poisson approximation. It would be of interest to compare these two approaches.

As with any good paper read to the Society, this one contains much interesting material, both on the theoretical foundations of the methodology and on numerous aspects of practical applications. There is still much work to be done, but plenty has already been done by Alexandros Beskos, Omiros Papaspiliopoulos, Gareth Roberts and Paul Fearnhead. It gives me great pleasure to propose a vote of thanks.

Dan Crisan (*Imperial College London*)

The paper extends classical results related to exact simulations of random variables to an infinite dimensional set-up. Here random processes, in particular, solutions of one-dimensional stochastic differential equations, are exactly sampled and various maximum likelihood and Bayesian inference methods are

developed on the basis of the two versions EA1 and EA2 of the exact sampling algorithms.

It is very easy to obtain a sample path from a Brownian motion or from its conditioned version, the Brownian bridge. This observation is the basis of the exact simulation of diffusions: if a diffusion X has an absolutely continuous distribution with respect to that of a Brownian motion or a Brownian bridge, then the *retrospective sampling* method that was developed by Papaspiliopoulos and Roberts can generate an exact simulation of X by using the perfect Brownian sampling.

Unfortunately the class of one-dimensional stochastic differential equations with solutions that are absolutely continuous with respect to a Brownian motion is very small. In effect we can only have equations of the form

$$\mathrm{d}X_t = \alpha(X_t)\,\mathrm{d}t + \mathrm{d}B_t.$$

(I shall not make explicit the coefficients' dependence of the unknown parameter θ as it is not relevant to the arguments that are presented.) X is simply a Brownian motion plus a drift term. This stems from the fact that the Brownian path $t \to B_t$ has quadratic variation $\langle B\rangle_t = t$ and X must have the same property if it wants to have an absolute continuous distribution with respect to that of B. In what follows I shall call these processes $\mathcal{B}$-diffusions.

The methods that are presented in the paper are applicable to a larger class of processes. Under additional assumptions, all diffusions that can be reduced to a $\mathcal{B}$-diffusion via a suitable change of space co-ordinates can be simulated exactly. These are the diffusions V for which there is a diffeomorphism $\eta : \mathbb{R} \to \mathbb{R}$ such that $\eta(V)$ is a $\mathcal{B}$-diffusion. From a pathwise perspective, the change of co-ordinates amounts to stretching or squashing the paths so that they obtain the right quadratic variation. This can be done, more often than not, at the expense of ending up with unbounded or, even worse, exploding drifts α. As a result, the simulation procedures and the inference methodology become either more difficult or not possible at all.

It is far more natural to start with a diffusion X that has an explicit distribution and the same quadratic variation structure as V and then to check whether it has an absolutely continuous distribution with respect to that of V. There is a generic method that finds such diffusions X and it stems from a classical result by Doss (1977) and Sussmann (1978). The result states that, if b is Lipschitz continuous and that σ is twice differentiable with bounded first and second derivatives, then the one-dimensional stochastic differential equation

$$\mathrm{d}V_t = b(V_t)\,\mathrm{d}t + \sigma(V_t)\,\mathrm{d}B_t$$

has a unique strong solution which can be written in the form

$$V_t(\omega) = u\{B_t(\omega), Y_t(\omega)\}, \qquad t \geqslant 0, \quad \omega \in \Omega,$$

where $u : \mathbb{R}^2 \to \mathbb{R}$ is the solution of the ordinary differential equation

$$\begin{aligned} &\frac{\partial u}{\partial x} = \sigma(u), \\ &u(0, y) = y \end{aligned} \tag{33}$$

and the function $t \to Y_t(\omega)$ solves an ordinary differential equation for every $\omega \in \Omega$. However, if we choose a *deterministic* or a *constant* Y, the resulting process

$$X_t(\omega) = u\{B_t(\omega), Y_t\}, \qquad t \geqslant 0, \quad \omega \in \Omega,$$

will satisfy the equation

$$\mathrm{d}X_t = b\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}B_t$$

where

$$\bar{b} = \tfrac{1}{2}\,\sigma(X_t)\,\sigma'(X_t) + \frac{\partial u}{\partial y}(B_t, Y_t)\frac{\mathrm{d}Y_t}{\mathrm{d}t^-}.$$

In particular, if Y is constant, then

$$\begin{aligned} X_t &= x + \tfrac{1}{2}\int_0^t \sigma(X_s)\,\sigma'(X_s)\,\mathrm{d}t + \int_0^t \sigma(X_s)\,\mathrm{d}B_s \\ &= x + \int_0^t \sigma(X_s) \circ \mathrm{d}B_s, \end{aligned}$$

where the last integral is a Stratonovitch integral. Applying this strategy to the logistic growth model (the second example in the paper), the corresponding process X can be chosen to be the log-normal diffusion

$$\mathrm{d}X_t = RX_t\,\mathrm{d}t + \sigma X_t\,\mathrm{d}B_t,$$

with the explicit solution

$$X_t = x\,\exp\{\sigma B_t + t(R - \sigma^2/2)\}.$$

The density of the (unconditional) distribution of V with respect to the distribution of X (on the path space) will then be proportional to

$$\exp\left[-d\left\{X_t^2(\omega) - x^2 + \frac{t}{2}\right\} - \frac{d^2}{2}\int_0^t X_s^2(\omega)\,\mathrm{d}s\right], \qquad d = \frac{R}{2\sigma\Lambda}.$$

There are three apparent advantages of this method over the change of spatial co-ordinates approach:

(a) it applies to a wide class of sets of coefficients (b, σ);
(b) it usually leads to simpler expressions for the Radon–Nikodym derivative, hence perhaps simplifying the ensuing estimation procedures;
(c) it extends naturally to multivariate diffusions and time inhomogeneous diffusions. We need to consider the multidimensional and, respectively, the time inhomogeneous version of expression (33).

For the Cox–Ingersoll–Ross model (the third example in the paper) though, the resulting process X is *not* absolutely continuous with respect to V. However, in the same vein, the Cox–Ingersoll–Ross diffusion is absolutely continuous with respect to the squared Bessel process

$$\mathrm{d}X_t = \delta\,\mathrm{d}t + 2\sqrt{X_t}\,\mathrm{d}B_t.$$

I have two further comments on the paper. It would be desirable to see a theoretical analysis of the computational effort that is involved for the various inference methods presented as a function of the number of time instances when the data are acquired. Also no comparison or analogy is made between the methods presented and the sequential Monte Carlo methods. However, the authors say that a comparison study between their methods and other existing methods is in progress.

In conclusion, the authors have already done a large amount of work on the subject, though I would venture to say that there is still much to be done on this line of research and I look forward to seeing the follow-ups. It gives me great pleasure in congratulating the authors on their paper and in seconding the vote of thanks.

The vote of thanks was passed by acclamation.

Barry Rowlingson (*Lancaster University*)
We are developing a package for the R programming environment to follow the theoretical developments that are described in the paper as well as to implement basic functionality for diffusion processes.

At present we can simulate a range of diffusion processes using the Euler method, and also we can use the EA1 algorithm for those that can be simulated that way. Methods for plotting realizations of diffusion processes have been implemented.

The code contains contributions from the paper's authors as well as from Duncan Murdoch. Eventually we hope to release the package into the CRAN archive of R packages, but for now it is available on request from the authors of the paper. We welcome any offers of help in the development of this package.

N. Chopin (*University of Bristol*)
This is an impressive piece of work; the methodology is very elegant, and the range of applications is promising. I shall raise three points.

(a) Given the first expression in lemma 1, the boundedness assumption on $\alpha^2 + \alpha'$—condition (b) in Section 2—is natural from a standard accept–reject perspective. One may naïvely remark, however, that, during a small time interval $[0, t]$, a Brownian bridge is confined to a small interval with very high probability. Thus I wonder whether, in some specific settings (for instance when computing the likelihood between two close observations x and y), an arbitrary truncation of function α, to force condition (b), would not produce an acceptable bias. Would this be a reasonable way to extend

slightly the class of diffusions that can be simulated from the simulation techniques that are presented here?

(b) I believe that the methodology developed in this paper can also be used in sequential problems, and within particle filtering algorithms. I shall not give exact details, for lack of space, but I shall illustrate this point with a toy example: the sequential filtering of a diffusion with known parameters, but observed at discrete times with noise. The simulation of the diffusion trajectory would then replace the simulation of the hidden Markov chain in the standard case.

(c) The authors mention some shortcomings of Euler discretization schemes, and corresponding Markov chain Monte Carlo samplers: in particular, the trade-off between discretization bias (coarse grids) and computational inefficiency (fine grids) is difficult to handle; 'computational inefficiency' refers here both to the cost of each iteration and to the number of iterations that are required to reach convergence.

I wonder whether this trade-off could not be dealt with in an automatic fashion through some adaptive algorithm. Consider some sequence p_n of posterior distributions (of parameters and imputed values) corresponding to a decreasing sequence of discretization steps Δ_n, and apply a sequential Monte Carlo algorithm to this artificial sequence. It then seems possible to design some stopping rule that detects when the discretization bias has become small, relatively to the Monte Carlo error. This alternative approach may be useful for cases where the Markov chain Monte Carlo algorithm that is developed in this paper cannot be applied.

S. B. Connor and W. S. Kendall (*University of Warwick, Coventry*)

We enjoyed and are greatly impressed by the elegance of this exact simulation method. This is one of an increasing number of simulation options which involve striking applications of probabilistic ideas. We take the opportunity here to sketch some recent work on another such option, that of coupling from the past (CFTP). It is natural to wonder whether these ideas might connect together in productive and useful ways.

CFTP was first introduced by Propp and Wilson (1996) as a method for sampling from the exact stationary distribution of an ergodic Markov chain. However, as shown by Foss and Tweedie (1998), this algorithm is possible (in an impractical sense) if and only if the Markov chain is uniformly ergodic. More recently, Kendall (2004) showed that all geometrically ergodic chains have (again impractical) dominated CFTP algorithms (as introduced in Kendall (1998)). This poses the question: what about when the chain X is *polynomially ergodic*?

We are currently investigating conditions under which a dominated CFTP algorithm exists for a polynomially ergodic chain X. Specifically, we consider chains for which there are constants $\alpha \in (0, 1)$ and $b, c \in (0, \infty)$, a small set C and a scale function $V : \mathcal{X} \to [1, \infty)$, bounded on C, such that

$$\mathbb{E}\{V(X_{n+1}) | X_n = x\} \leqslant V(x) - c\, V^{\alpha}(x) + b\, \mathbf{1}_{[X_n \in C]}. \tag{34}$$

This drift condition implies that the chain is polynomially ergodic, as shown in Jarner and Roberts (2002).

In Connor and Kendall (2006) we introduce the concept of a *tame* chain. A chain X satisfying the drift condition (34) is tame if we can subsample X (in a specific, adaptive way) to produce a geometrically ergodic chain X'. In this case we prove that we can produce a suitable dominating process for X by slowing down the dominating process for X' that is shown to exist in Kendall (2004).

Thus the question becomes: when is a polynomially ergodic chain tame? We have some sufficient conditions for tameness, but also examples to show that none of these are necessary. It is unclear at present whether all chains are tame, and this is an issue that we shall be investigating as part of our future research in this area.

John T. Kent (*University of Leeds*)

Let me focus on the authors' first model, which can be recast in the form

$$\mathrm{d}X_t = -\lambda \sin(X_t - \mu)\, \mathrm{d}t + \sigma\, \mathrm{d}W_t, \qquad \lambda > 0.$$

As a diffusion on the circle, it can be termed a 'von Mises diffusion' because in equilibrium it follows a von Mises distribution, with probability density function proportional to $\exp\{\kappa \cos(x - \mu)\}$, $0 \leqslant x < 2\pi$, $\kappa = 2\lambda/\sigma^2$ (e.g. Kent (1978)). It has some nice applications.

(a) *Position model*: imagine a liquid in a magnetic field containing tiny iron filings whose orientations are subject to the same sorts of molecular bombardments that motivate the standard Brownian motion model.

(b) *Velocity model*: consider a moving bacterium with a preferential direction of motion. Suppose that the magnitude of the velocity of the bacterium is constant (or at least the process for magnitude is independent of orientation) and that the orientation of the velocity follows the von Mises diffusion (Hill *et al.*, 1997). The data consist of locations of the bacterium at regularly spaced times.

Several questions come to mind for the von Mises process and more generally.

(i) Can the methods of this paper be used for applications such as the bacterium example, where the velocity follows a diffusion, but where the data are based on the integrated velocity?
(ii) Suppose that we wish to estimate μ for data from the von Mises process. The simplest statistical approach is to compute the maximum likelihood estimate (MLE) by treating the data as independent and identically distributed from a von Mises distribution. We might expect that this naïve MLE will be close to the proper MLE, but that we shall underestimate the true variance of the estimator. This latter problem might be overcome by allowing for the autocorrelation in the data. How does this simple statistical approach compare with the proper analysis that is proposed in this paper?
(iii) The efficiency of the authors' simulation technique presumably decreases as either of two quantities increases: the time step $\delta t/\sigma$ and the concentration parameter κ. Is this correct?
(iv) It is an elementary observation that the distribution of a Brownian bridge does not depend on the presence of any (constant) drift in the underlying Brownian motion. Does this property help to explain the unexpectedly high efficiency of the authors' simulation techniques in general?

Bruno Casella (*Università Bocconi, Milan, and Lancaster University*)
I shall report on on-going research joint with Gareth Roberts that is strictly connected to the paper. In the paper the *exact algorithm* is the key tool for the construction of an unbiased Monte Carlo estimator of the transition densities to be used for inference on the parameters of the diffusion. The exact algorithm framework can be successfully applied in many other Monte Carlo problems involving the simulation of functionals $\xi(\cdot)$ of diffusion processes. A natural area where these problems arise is option pricing.

Our main idea is the following. Given a process $X = \{X_t : 0 \leqslant t \leqslant T\} \in \mathcal{D}_1$ with starting-point x and unknown probability measure $\mathbb{Q}$, we simulate an exact skeleton

$$\mathcal{S}_1 = \{(0, x), (\psi_1, \omega_{\psi_1}), \ldots, (\psi_k, \omega_{\psi_k}), (T, \omega_T)\}$$

by the (unconditional) EA1 algorithm. Given $\mathcal{S}_1$, $\mathbb{Q}$ is the product of independent Brownian bridges. Using this characterization, we can simulate ξ conditionally on S_1 *exactly* (i.e. without discretization error). We named this methodology the *exact Monte Carlo method*. We applied the method to the problem of the estimation of

$$\begin{aligned}\nu &= \mathbb{E}_{\mathbb{Q}}\{\xi(\omega)|\omega_0 = x\}\\ &= \mathbb{E}_{\mathbb{Q}}\{g(\omega_T)\,\mathbb{I}_{\{\tau<T\}} + h(\omega_T)\,\mathbb{I}_{\{\tau\geqslant T\}}|\omega_0 = x\}\end{aligned}$$

where τ is the *first exit time* of ω from a set $H := (a, b)$ such that $x \in (a, b)$, and $g(\cdot)$ and $h(\cdot)$ are two measurable functions. This problem is relevant in finance, e.g. in barrier options pricing or credit risk modelling. First, we call algorithm EA1 and output S_1. Given S_1, all that we need for the simulative of ξ is to sample the crossing event of a product of independent Brownian bridges. In the single-barrier case ($\alpha = -\infty$ or $b = \infty$) we apply the well-known Bachelier–Levy theory which provides an explicit formula for the *one-sided* crossing probability of a Brownian bridge. Notably, this also allows us to construct a Rao–Blackwellized Monte Carlo estimator of ν, which is similar in spirit to the one that was used for the bridge method, by sampling directly $\mathbb{E}_{\mathbb{Q}}(\xi|S_1)$ (Fig. 5). In the double-barrier case we do not have an explicit formula for the (*two-sided*) crossing probability of a Brownian bridge: we circumvent the problem by developing a suitable procedure to simulate from it. The procedure is based on Doob's representation of the crossing probability of the Brownian motion as a telescopic sum.

Simulation studies show that our estimator is

(a) *unbiased*—Euler-based estimators converge to our estimator as the length of the discretization interval decreases—and
(b) *efficient*—for small discretization intervals, our estimator is more efficient than Euler's.

Extensions of this methodology to processes $X \in \mathcal{D}_2$ and jump diffusion processes are possible. We are currently working on the application of the exact Monte Carlo method to Monte Carlo estimation of derivative hedge ratios in finance.

The following contributions were received in writing after the meeting.

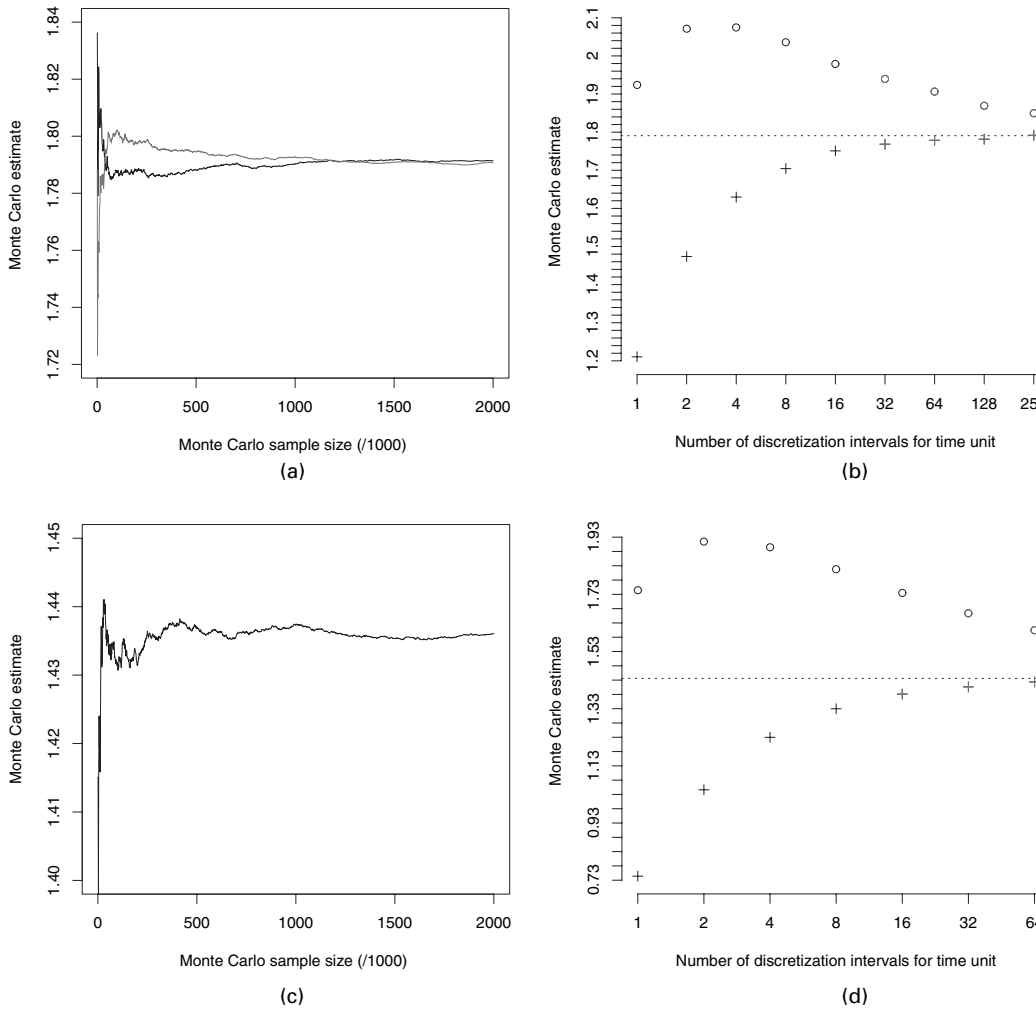

Fig. 5. Monte Carlo estimate of $\nu = \mathbb{E}(X_T \mathbb{1}_{\{\tau > T\}})$ where X is the SINE model (single-barrier case: $x = 1.5$; $a = -\infty$; $b = 4.5$; $T = 5$): (a) Monte Carlo convergence of the exact Monte Carlo estimator (not Rao–Blackwellized and Rao–Blackwellized); (b) convergence of the estimators based on a discrete Euler scheme (∘) and a continuous Euler scheme (+) to the exact Monte Carlo estimate (·····), double-barrier case ($x = 2$; $a = 1$; $b = 4.5$; $T = 5$); (c) Monte Carlo convergence of the exact Monte Carlo estimator; (d) convergence of the estimators based on a discrete Euler scheme (∘) and a continuous Euler scheme (+) to the exact Monte Carlo estimate (·····)

Frank Ball, Ian Dryden and Mousa Golalizadeh (*University of Nottingham*)

We congratulate the authors for their inspirational paper. Our particular interest concerns extensions of the retrospective exact simulation work to compact spaces. Ball *et al.* (2004) have considered Ornstein–Uhlenbeck (OU) processes in Kendall's (1984) shape space for $k \geqslant 3$ points in two dimensions. In particular, a family of OU processes has been defined with Itô stochastic differential equations for the Riemannian shape distance ρ_t to a reference configuration at time $t > 0$ that is given by

$$\mathrm{d}\rho_t = \{\cot(2\rho_t) + (k-3)\cot(\rho_t) - g'(\rho_t)\}\,\mathrm{d}t + \mathrm{d}B_t, \qquad 0 \leqslant \rho_t \leqslant \pi/2,$$

where $g'(\rho) = \mathrm{d}g(\rho)/\mathrm{d}\rho$ is the infinitesimal OU drift term, B_t is standard Brownian motion on the real line and the process starts at ρ_0 at time $t = 0$. In the case $k = 3$ the planar shape space for triangles of points is equivalent to a sphere of radius $\frac{1}{2}$ in three dimensions, and ρ_t is the colatitude on this sphere.

Since the diffusion process for ρ_t is restricted to the closed interval $[0, \pi/2]$, simulation of the ρ_t-process is challenging. For example the Euler scheme is problematic near the end points 0 or $\pi/2$ and needs adaptation (see Ball *et al.* (2004)).

We are currently exploring the retrospective exact simulation methods of the present paper for simulating ρ_t, and the methods need to cope with the facts that the domain is a closed interval, $\alpha^2 + \alpha'$ is not bounded below at $\pi/2$ and not bounded above at 0.

We have used an adaptation of algorithm EA2 for simulating ρ_t, where the decomposition is carried out with respect to the minimum, as follows.

(a) The function $\alpha^2 + \alpha'$ is replaced with a differentiable bounded function which is identical to $\alpha^2 + \alpha'$ except in the intervals $[0, \varepsilon_1)$ and $(\pi/2 - \varepsilon_2, \pi/2]$, where $\varepsilon_1, \varepsilon_2 > 0$ are small.
(b) The time length T of a simulation is chosen such that the probability that the maximum is outside the interval $[0, \pi/2]$ is negligible. This is practical unless ρ_0 is very close to $\pi/2$, in which case the decomposition should be taken with respect to the maximum and the probability that the minimum is outside the interval should be negligible.
(c) Additional criteria for rejecting a proposed path are if
   (i) the simulated end point from density h in Section 2.3 is outside the interval $[0, \pi/2]$ or
   (ii) the simulated minimum m of the path is below zero.

Consider an example with $g'(\rho) = \kappa \sin(2\rho)$. This process was introduced by Kent (1975) for the sphere ($k = 3$). We use the adapted retrospective exact simulation method for a single long run consisting of 100000 individual time steps of length $T = 0.02$, where the end point from one time step becomes the start-point for the next step. In Fig. 6 we plot simulated values at unit intervals from the OU process, and a histogram of the simulated values evaluated at time intervals of length 0.02 with the density of the theoretical equilibrium distribution overlaid. The histogram and density are extremely similar, indicating that approximations to retrospective exact sampling can also be useful in more general situations.

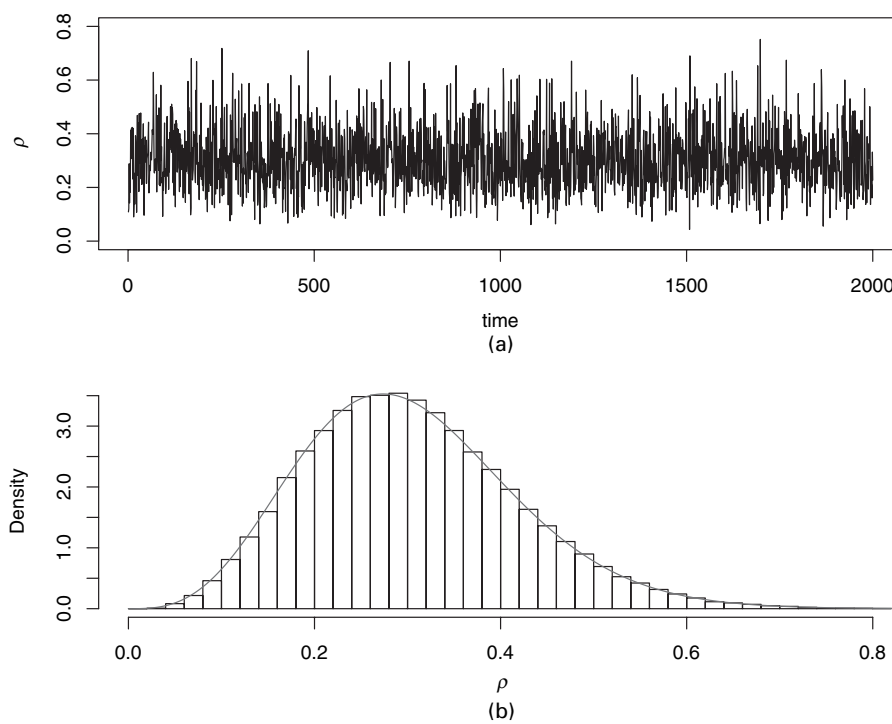

Fig. 6. (a) Simulated points at unit time steps from the diffusion process for ρ_t with $g'(\rho) = \kappa \sin(2\rho)$, $k = 4$, $\kappa = 10$ and $\rho_0 = 0.2$ and (b) density scaled histogram of simulated points from the OU process at intervals of 0.02, with the density of the theoretical equilibrium distribution $f(\rho) \propto \sin^{2k-5}(\rho) \cos(\rho) \exp\{\kappa \cos(2\rho)\}$ (———)

N. H. Bingham (*University of Sheffield*)

I congratulate the authors on this interesting paper. I confine myself to a few comments on theoretical and financial aspects.

Curse of dimensionality

The multidimensional extensions that are mentioned in the final section are intriguing from the point of view of mathematical finance. There, by Markowitzian diversification, one holds a large number of assets; the danger here is the *curse of dimensionality*. Methods are available to counter this which reduce the effective dimensionality from the number or *assets* to the number of industrial *sectors*; see for example Bingham *et al*. (2003).

Ensemble models

An alternative approach to estimating diffusion coefficients by discrete methods arises where we have (or can obtain) an *ensemble* of paths, or diffusing particles say, rather than one. We can then extract information from the *counts* of the number of diffusing particles in some window of observation, as a function of time. The technique is called *number fluctuation spectroscopy*. It was used by Smoluchowski, following Einstein's work, to estimate Avogadro's number, and by Rothschild to study mobility of spermatozoa. For details and references, see for example Bingham and Dunham (1997) and Bingham and Pitts (1998).

Drift and diffusion

As observed in Section 2, granted a sample path over a time interval of positive length, we can determine the diffusion coefficient with certainty (e.g. in the Brownian case from the quadratic variation). But this involves (uncountably) infinite sample sizes, and we can obtain and handle only finite samples in reality.

Nevertheless, the suggestion is that in principle we can discriminate between different diffusion coefficients with certainty; if both are the same, we can form likelihood ratios and, for example, test hypotheses comparing the drifts.

Diffusions, or Itô processes, being locally Gaussian, relevant here is the split between mutual absolute continuity and singularity of Gaussian processes, for which see for example Ibragimov and Rozanov (1978), chapter III.

Splitting times

The use of random times such as when the minimum is (last) attained—which is far from being a stopping time—to split the path into independent fragments is a powerful technique. Such times are called *splitting times*; for general theory, see Rogers and Williams (1994), section III.49.

Multidimensional diffusions

The irreducibility condition that was used by Aït-Sahalia (2004) to reduce to the case of unit diffusion coefficient predates the standard work on multidimensional diffusions by Stroock and Varadhan (1979)—which does not contain it. It is good to see practice driving theory here.

Peter Clifford (*Oxford University*)

I have just a few comments on an enjoyable paper, Firstly note that the Cox–Ingersoll–Ross diffusion

$$\mathrm{d}Z_t = p\sqrt{Z_t}\,\mathrm{d}B_t + (qZ_t + r)\,\mathrm{d}t$$

is the square of an Ornstein–Uhlenbeck process in a specific dimension. The model is discussed in Karlin and Taylor (1981), pages 333–334, where the transition density is given explicitly. The transitions of the diffusion can be readily simulated by using the representation of a non-central χ^2-distribution as a Poisson mixture of gamma variates.

My second comment is more substantial. I am uneasy that a novel method of solving stochastic and partial differential equations has not being rigorously compared with established methods in numerical analytic terms. At the very least it would have been helpful to see additions to Fig. 1 in Aït-Sahalia (2002), but ideally we should be given comparative analyses of the growth of error and the order of accuracy.

Finally, where can I see a real diffusion? Stock-market transactions and price movements, for example, are essentially discrete. Perhaps it is better to think of diffusion as modelling latent sentiment, which is continuously fluctuating but not directly observable. The observables are then driven by this underlying process. A simple model is that market events (trades etc.) are a Poisson process with intensity that is proportional to the diffusion. For the Cox–Ingersoll–Ross class of diffusions we have recently shown how

to calculate likelihoods explicitly for such a doubly stochastic process. The method exploits the identity between discrete processes of this type and the death epochs of the standard immigration, birth and death process (Clifford and Wei, 1993).

Valentine Genon-Catalot (*Université Paris 5 René Descartes*)
For diffusion having ergodic properties, complete results are available when estimating both drift and diffusion parameters by maximum likelihood (consistency, asymptotic normality as the sampling interval tends to 0 and the total length time interval of observation tends to ∞). Without ergodicity assumptions, exact maximum likelihood estimators of diffusion parameters have now well-known properties (consistency, asymptotic mixed normality, as the sampling interval tends to 0 within a fixed length time interval). Because, for most models, the exact likelihood is not analytically tractable, these results are of a theoretical nature and researchers have all been faced with the same problem: it was not possible to produce exact simulated data.

Now, let me express my enthusiasm and congratulate the authors for their impressive papers. Retrospective exact simulation of diffusions sample paths is a revolution for the study of diffusion processes and especially for the statistics of diffusion processes. This revolution is just at its first steps of development.

The authors focus on producing unbiased estimation of the exact likelihood. A crucial step is to obtain unbiased estimators of the transition density $p_t(x, y, \theta)$ of the model. I am especially interested in three of the methods proposed to achieve this goal: the acceptance and the simultaneous acceptance methods and the Poisson estimator. For these, the authors use a representation formula of the transition density which was already known but had only been used for theoretical purposes. This formula contains the expectation of a functional under a Brownian bridge distribution. I am particularly impressed by the simultaneous acceptance method which produces a very simple unbiased estimator of this expectation for all values of the parameter to be estimated. Also, the Poisson estimator is interesting because it can be applied to a wider framework.

Numerous problems remain to solve. In particular, the paper makes no connection between the simulations and theoretical results. For instance, the sine model is a null recurrent diffusion whereas logistic growth may be positive recurrent. Asymptotic results on maximum likelihood estimators are thus different. Joint estimation of drift and diffusion parameters is not investigated here. In diffusion models, using simulation based on Euler schemes, we can observe that some parameters are badly estimated and others are not: for instance, when the drift is mean reverting ($b(x) = \alpha(\beta - x)$), α is badly estimated (a good estimator requires a very large time interval), and β is always well estimated. The parameters in the diffusion coefficient are usually well estimated (even when the total time interval is small). This area is probably under study by the authors and this paper suggests plenty of new and interesting topics.

Konstantinos Kalogeropoulos (*Athens University of Economics and Business*)
I add my congratulations to the authors for their wonderful contribution to the challenging problem of inference for diffusions.

My comments focus on two assumptions that hinder some desirable extensions. The first assumption requires the Markovianity of the observed diffusion. Although not stated explicitly, it is the Markov property that allows us to write down the log-likelihood as

$$l(\theta|\mathbf{V}) = \sum_{i=1}^{n} l_i(\theta), \qquad l_i(\theta) := \log\{p_{\Delta t_i}(V_{t_{i-1}}, V_{t_i}; \theta)\}.$$

Non-Markovian diffusion models are being used extensively. Famous examples include stochastic volatility and continuous time autoregressive moving average models.

The second assumption stems from the transformation in equation (3). To find such a transformation $h(\cdot)$ for a d-dimensional diffusion we need to solve the following vector differential equation:

$$\nabla h(V_t, \theta)\, \sigma(V_t, \theta)\, \sigma(V_t, \theta)'\, \nabla h(V_t, \theta)' = I_d, \tag{35}$$

where now $\sigma(V_t, \theta)\, \sigma(V_t, \theta)'$ denotes the diffusion matrix and I_d is the identity matrix of dimension d. Since equation (35) is not always solvable, this limits the extensions to the multivariate case. Aït-Sahalia (2004) termed the diffusions with solvable equation (35) *reducible*. He also showed that the diffusions that are involved in ordinary stochastic volatility models do not belong to this class.

An alternative likelihood-based approach uses data augmentation schemes; see Roberts and Stramer (2001), Elerian *et al.* (2001) and Eraker (2001). These schemes overcome the barrier of the Markov property and, although they do not provide exact inference, their discretization error may become arbitrarily small.

The existence of the transformation in equation (3) is a major issue for data augmentation schemes as well. Under a fairly thin discretization the quadratic covariation process determines exactly the parameters in $\sigma(V_t, \theta)$ and the algorithm degenerates. Nevertheless, there is some on-going work that handles some cases of multivariate diffusions, Kalogeropoulos *et al.* (2006a) propose a reparameterization for *reducible* diffusions, whereas Kalogeropoulos (2006), Kalogeropoulos *et al.* (2006b), Chib *et al.* (2004) and Golightly and Wilkinson (2005) present algorithms for stochastic volatility models.

Mathieu Kessler (*Universidad Politécnica de Cartagena*)
The authors are to be congratulated for providing an impressive collection of ingenious and innovative ways to implement likelihood-based estimation for discretely observed diffusions. Their paper combines original expressions for the transition density of the process and the use of exact simulation algorithms in the corresponding Monte Carlo procedures.

Like several other recent papers, it focuses on obtaining the best possible numerical approximation to the likelihood, for the derived estimators to inherit the nice properties, e.g. asymptotic efficiency, of the likelihood-based estimators. In particular, the authors emphasize the fact that, through the use of exact simulation algorithms, they can obtain unbiased estimations of the transition density. However, I am missing insight about the consequences of the Monte Carlo approximation on the behaviour of the estimators. Why, for example, is it an improvement with respect to the existing methods to be able to use unbiased estimators of the transition density? Would it be possible to assess, through a theoretical result, the rate at which the number N of Monte Carlo replicates should grow with the number n of observations, for the asymptotic efficiency to be maintained under the assumption of ergodicity? The existing results for approximation-based estimators that have been proposed in the literature can only guarantee the existence of such an increasing sequence N_n. In particular, what would be the asymptotic properties of the proposed estimators, for a fixed Monte Carlo effort, when the number of observations tends to ∞?

Unfortunately, such results do not seem to be easy to obtain. It would be different if the Monte Carlo approximation did not concern the transition density itself but the derivative of its logarithm with respect to the parameter, i.e. the building-block of the score function. Indeed, if an unbiased Monte Carlo approximation of $\partial_\theta p_t / p_t(u, v; \theta)$ were available, it would be possible to prove the asymptotic normality of the deduced approximate estimator, even for a fixed Monte Carlo effort. Moreover, it would be easy to characterize precisely the corresponding loss of efficiency in the asymptotic variance; this kind of result was obtained for estimating functions in Kessler and Paredes (2002). My last question is, therefore, could some of the expressions that are proposed for the transition density be used to derive unbiased simulation-based approximations to $\partial_\theta p_t / p_t(u, v; \theta)$?

Claudia Klüppelberg (*Munich University of Technology*)
I congratulate the authors for an immensely impressive and path breaking paper. They use a very rich methodology embracing deep results from stochastic analysis, Markov chain Monte Carlo simulation and algorithmic methods. Their simultaneous acceptance method is based on the Cameron–Martin–Girsanov formula, which is the basis of dealing with stochastic differential equations driven by Brownian motion. Of course, this is a very large class of models, although they must restrict it further for technical reasons.

As modelling issues quite naturally often run ahead of the necessary development of the corresponding statistical inference, the next question related to their work is already waiting for its solution. High frequency financial data are modelled by continuous time models; stylized facts require, among other features, the estimation of processes with jumps. These features are covered by the meanwhile well-known Barndorff-Nielsen and Shephard (2002) model that was mentioned in the paper, and also by a new continuous time generalized autoregressive conditional heteroscedasticity (COGARCH) model, which is defined as follows. For a Lévy process L and parameters $\beta, \eta, \varphi > 0$ the COGARCH(1,1) process is defined as the solution to the equations (for $t > 0$)

$$\begin{aligned} \mathrm{d}S_t &= \sqrt{V_t}\,\mathrm{d}L_t, \\ \mathrm{d}V_{t+} &= (\beta - \eta V_t)\,\mathrm{d}t + \varphi V_t\,\mathrm{d}[L, L]_t^{(d)}, \end{aligned}$$

(Klüppelberg *et al.*, 2004) where V is left continuous and

$$[L, L]_t^{(d)} = \sum_{0 < s \leqslant t} (\Delta L_s)^2$$

is the discrete part of the quadratic variation process of L, i.e. it is a subordinator.

This process has been estimated in Müller (2005) by methods that were used also in Roberts *et al*. (2004) for the Barndorff-Nielsen and Shephard model, but both papers are restricted to finite activity Lévy processes. It would indeed be interesting to see over the next few years a development of such clever methods as presented in the present paper to models that are driven by infinite activity Lévy processes.

Hans R. Künsch (*Eidgenössische Technische Hochschule Zürich*)
The authors are to be congratulated on their impressive and highly original paper. I believe that diffusion processes will become increasingly more important in statistical applications because they are the obvious extension of deterministic differential equations which are widespread. It is nice that with the methods of this paper it is not necessary to rely on a time discretization. However, one hopes that the estimation procedures do not depend on the fine structure that is implied by the diffusion model since this is the most doubtful part.

As a more technical comment, I would like to add that the exact simulation method that is presented here allows us to implement the accept–reject version of the particle filter (see Hürzeler and Künsch (1998)) for partially observed diffusions. Assume that the observations $Y_{1:n} = (Y_1, \ldots, Y_n)$ are conditionally independent given (X_t) and Y_i depends on X_{t_i} only. The particle filter approximates the conditional density of X_{t_i} given $Y_{1:i}$ by a sample of particles $(x_{t_i}^{(1)}, \ldots, x_{t_i}^{(N)})$. To generate particles at time t_{i+1} recursively, we select an index K with probabilities proportional to $\exp\{A(x_{t_i}^{(k)})\}\, Z(x_{t_i}^{(k)})$; then given $K = k$ we generate a proposal at time t_{i+1} according to the density

$$\frac{1}{Z(x_{t_i}^{(k)})} \exp\left\{ A(u) - \frac{(u - x_{t_i}^{(k)})^2}{2(t_{i+1} - t_i)} \right\} p(y_{i+1} | X_{t_{i+1}} = u)$$

and finally we decide whether to accept this proposal by the exact algorithm in this paper. If the normalizing constant $Z(x_{t_i}^{(k)})$ cannot be computed in closed form, we can replace it by a suitable approximation at the cost of an additional factor in the acceptance probability.

T. J. Lyons (*Oxford University*)
This paper is focused on developing earlier and interesting work on methods for simulating processes (and particularly bridges) into numerical tools for distinguishing between a parameterized family of Markov models on the basis of discrete time observations of a process. A key point in this is to understand, in a computationally effective way, the relative likelihood of the Markov transition $x_n \to x_{n+1}$ given the different parameter values. In effect, we need a parameterized solution to the backward equations.

One could certainly attempt this analysis in a variety of ways, including the use of partial differential equations. It is clear that the benefits and complexity of different approaches will mean that some approaches are better than others.

To make comparisons, it seems essential that a clear quantitative approach is adopted, and one regret that I had about the paper is the lack of analytic details with which to judge the quality of the approach. In particular, it would be helpful to have precise classes of models for which we would apply this approach, and estimates for the errors after a given number of function evaluations etc. in terms of the smoothness of the coefficients in the models studied etc.

It would have been very helpful to have precise results (in terms of the number of function evaluations per unit error) setting out when this method was more effective than, for example, partial differential equation methods. The latter might be expected to perform quite well in low dimensions.

Rogemar Mamon and Keming Yu (*Brunel University, Uxbridge*)
The applied theme of the paper on likelihood-based inference for discretely observed diffusions is very interesting. However, in financial modelling the specification of a diffusion process V via the stochastic differential equation (SDE) in equation (1) is too simple to capture the stylized features of the economy and markets. Nowadays, the use of regime switching models to understand the behaviour of financial variables (e.g. interest rates) is becoming more common, motivated by empirical evidence and owing to economic grounds. See, for example, Bansal and Zhou (2002). In particular, both the drift and the diffusion coefficients of the SDE in equation (1) can have dynamics of their own. For instance, they can be allowed to switch between economic regimes and the shift is modulated by a finite state Markov chain. This will certainly enrich standard models such as the Cox–Ingersoll–Ross model that was mentioned in the paper. Guo and Zhang (2004) and Wu and Zheng (2004), among others, have provided examples for this type of model formulation. Assuming that an SDE describing the evolution of a financial phenomenon does not have an analytic solution, especially if it has periodic drift and therefore the log-likelihood is not available,

we wonder how the exact algorithm can be extended into the setting where the process has a mixture of Gaussian and Markov chain dynamics. It would be of interest to determine how the exact algorithm can be employed efficiently in the estimation of transition matrices, the state of the Markov chain and other model parameters.

Also, the authors pointed out that the ideas and methods of the paper can accommodate SDEs with jumps provided that they are based on a finite activity Lévy process. In general, what class of integrators (i.e properties and characterization) are possible for the stochastic integral in equation (1) so that either the EA1 or EA2 algorithms can be extended successfully? In other words, what kind of semimartingales can be included in this class?

Finally, it is not clear to us how the proposed transition density estimation technique compares with nonparametric estimation methods such as the kernel estimation of transition density. The latter is much simpler in mathematical details, and computationally easier to implement as well as distribution free. For example, given the observations $V = \{V_{t_0}, V_{t_1}, \ldots, V_{t_n}\}$ of a diffusion process V, we can estimate quickly the transition density (2) via

$$\hat{p}_t = \frac{1}{n} \sum_{i=1}^{n} K_h(V_{t_1} - t),$$

where $K_h(\cdot)$ is a kernel function with bandwidth h.

Jesper Møller (*Aalborg University*)

I wonder how successfully the ideas of this interesting and stimulating paper can be extended to the case of a Cox process X_t, where V_t is a non-negative diffusion process and X_t conditional on V_t is a Poisson process with intensity function V_t. Suppose that $t_1, \ldots, t_n$ are the events of the Cox process observed on a finite time interval $[0, a]$ and, as in the paper, we observe the diffusion process only at the times $0 = t_0 < t_1, \ldots, < t_n$. Using the notation in the paper, the likelihood is

$$L(\theta|\mathbf{v}, t_1, \ldots, t_n) = E_\theta\left\{ \exp\left(-\int_0^a V_t \, \mathrm{d}t \right) \prod_{i=1}^{n} V_{t_i} \times p_{\Delta t_i}(V_{t_{i-1}}, V_{t_i}; \theta) \middle| \mathbf{v} \right\}$$

where the conditional expectation is with respect to the diffusion process given $(V_{t_0}, \ldots, V_{t_n}) = \mathbf{v}$. How efficiently would the methods in the paper apply when the likelihood is approximated and maximized by using a Markov chain Monte Carlo missing data approach?

In passing it may be worth noting that, if the diffusion is a Cox–Ingersoll–Ross model, Srinivasan (1988) and Clifford and Wei (1993) have established the equivalence between the Cox process and the process of death times of a simple immigration, birth and death process, which is easy to simulate. Incidentally, this model is a special case of the permanent process that is introduced in McCullagh and Møller (2005).

Sara Pasquali and Fabrizio Ruggeri (*Istituto di Matematica Applicata e Tecnologie Informatiche, Milan*)

We express our gratitude to the authors for an enlightening paper which provides an outbreaking new method for likelihood-based inference for discretely observed diffusions. The paper provides many details about the method and discusses many of its aspects, leaving a thorough analysis of some of them to forthcoming papers. We are looking forward to reading them since they will address some of our concerns. Here we just want to comment on possible extensions of the method.

The most general stochastic differential equation is given by

$$\mathrm{d}V_s = b(s, V_s; \theta)\, \mathrm{d}s + \sigma(s, V_s; \theta)\, \mathrm{d}B_s,$$

as in the following case about population dynamics:

$$\mathrm{d}V_s = R(s) - \alpha \exp(\beta T) V_s - \sigma \exp(\beta T) V_s \, \mathrm{d}w_t$$

where $\theta = (\alpha, \beta, \sigma)$, w_t is Gaussian white noise, T is the temperature (possibly dependent on the time s) and $R(s)$ is a known recruitment function.

In the paper, the time s does not appear directly in the drift and in the diffusion, but only through V_s. We wonder whether the method proposed is valid also in this case.

As a second comment, we think of practical problems in which parallel, multivariate diffusion processes are to be analysed all together. We suppose that the diffusion processes might slightly differ in drift and/or

diffusion and/or the prior on the parameters. Is there a way to estimate all of them together, sharing as many simulated points as possible, or should they be analysed separately?

Another related comment is about evolving systems in which new data are made available as time goes by. Is it possible to analyse the process on $[0, s+t]$, starting from estimation in $[0, s]$ and adding the data that are observed in $(s, s+t]$, or should the estimation on $[0, s+t]$ restart from scratch?

It would be interesting to explore the relationship between parameter estimation and stochastic stability, e.g. in the logistic growth model (example 2) where the equilibrium state 0 is stable for $\sigma^2 > 2R$ and unstable otherwise. In the latter case, which is considered in the paper, it is known that the process V_s fluctuates around the value Λ. Once data have been simulated from a logistic model, we would like to know what happens about the stability properties of the equilibrium, especially when parameters are close to the threshold $\sigma^2 = 2R$.

Harry Pavlopoulos (*Athens University of Economics and Business*)

The authors are to be congratulated for their ground breaking work on likelihood-based parametric inference for discretely observed diffusions, as well as for the clarity with which they have summarized its mathematical and computational aspects (indeed of highly technical nature) in the paper. The purpose of this comment is to raise two interrelated points.

First, adding to the possible extensions of the methods that were considered by the authors, I suggest also the consideration of discretely observed diffusion with boundary condition(s) imposed on either one or both end points of its interval state space, provided that these are regular accessible boundaries. In other words, is it possible to extend the methodology to estimate parameters imposed through boundary condition(s), specifying transition probabilities from the interior of the state space to its boundaries and vice versa, in addition to parameters introduced through the drift and diffusion coefficients? This question is really of much broader intent, pertaining also to nonparametric methods of inference for discretely observed diffusions, such as those proposed by Aït-Sahalia (1996), Jiang and Knight (1997), Stanton (1997) and more recently by Fan and Zhang (2003) and Bandi and Phillips (2003). This consideration is of importance in geophysical and other environmental applications where the observed process is intermittent. An example of this situation is the diffusion model that was proposed by Pavlopoulos and Kedem (1992) for processes of spatially averaged rain rate over a sufficiently large region, where the state space of the diffusion is the closed interval $[0, \infty)$ and intermittency between wet and dry states of the region is modelled via a sticky boundary condition at $\{0\}$.

Second, I should like to share that throughout my reading of the paper I was under the impression that the spacing between time instants at which the diffusion is observed is quite general, as implied by the inequality $0 = t_0 < t_1 < \ldots < t_n$ when the log-likelihood of the observations $\mathbf{v}$ was defined. Regular spacing in the EURODOLLAR data and in the simulations of the SINE and LOG-GROWTH data did not change this impression, until reading the conclusions of the paper where the authors address the performance of their methods by implying a uniform sampling frequency Δt, and thus regular spacing. This point remains subtle, at least in my understanding, and needs clarification. However, if the methodology holds under irregular spacing, then in the case of regular accessible boundaries it might be still implemented for conditional inference, at least, using observations only from the interior of the state space and ignoring those on the boundaries.

Christian P. Robert (*Université Paris Dauphine*)

The authors have achieved in this paper a *tour de force*, bypassing the usual and awkward discretization that is required for processing diffusions. The example of the Cox–Ingersoll–Ross model is quite revealing in this respect since, although a closed form representation via non-central χ^2-distributions (or ARG(1) chains) does exist, Euler's discretization of the diffusion is biased because of the positivity constraint on the discretized process. This paper is thus bound to have long-term consequences in the way that we can handle and do inference on stochastic processes.

Obviously (and happily so!), much remains to be done in this new direction for the processing of parameterized diffusions. The boundedness condition on $\alpha^2 + \alpha'$ is quite restrictive, and the construction of the bound $r(\omega, \theta)$ seems to be another *tour de force*. It would be of considerable interest to achieve a relaxation of the boundedness condition as well as a more constructive approach to the derivation of the bound $r(\omega, \theta)$. Given that the method at the core of the paper is intrinsically an accept–reject method, I wonder whether or not importance sampling alternatives could be available, in that they do not require (in principle) a bound on the target distribution that is to be simulated. In particular and in parallel with standard importance sampling theory on finite dimensional distributions, is it possible to come up with variance

comparison and optimality goals in a family of different importance functions? Furthermore, in classical Monte Carlo approximations, using an indicator I as in equation (12) in the estimation of $p_t(x, y; \theta)$ often leads to estimators with high variance. A choice of an importance function that would increase the probability of getting a 1 for this indicator would thus be of interest. For instance, in the simulation of the Cox–Ingersoll–Ross model, importance functions that introduce a drift in the variance of expression (30) lead to a significant improvement in the estimation of some quantities (Douc *et al.*, 2005). Similarly, any kind of Rao–Blackwellization applied to this indicator, if possible, would decrease the variance.

Mathias Rousset (*University of Toulouse*) **and Arnaud Doucet** (*University of British Columbia, Vancouver*) The authors are to be congratulated for this impressive paper which solves many problems and opens many avenues of investigation. We present here a direct application of their methodology to time discretization error-free filtering of partially observed diffusions. Consider the following diffusion where $X_0 \sim \pi_0$ and for $t > 0$:

$$\mathrm{d}X_t = \alpha(X_t)\,\mathrm{d}t + \mathrm{d}B_t.$$

This diffusion is partially observed at times $\{t_k\}_{k \geqslant 1}$ (where $t_k > t_{k-1}$) and the conditional density of the kth observation Y_{t_k} given by $g_k(y_{t_k}|x_{t_k})$ is known analytically. We are interested in estimating sequentially the distributions

$$p(x_{t_0:t_k}|y_{t_1:t_k})\,\mathrm{d}x_{t_0:t_k} = \mathbb{P}(X_{t_0:t_k} \in \mathrm{d}x_{t_0:t_k}|Y_{t_1:t_k} = y_{t_1:t_k})$$

where $t_0 = 0$, $x_{t_0:t_k} = (x_{t_0}, x_{t_2}, \ldots, x_{t_k})$ and $y_{t_1:t_k} = (y_{t_1}, y_{t_2}, \ldots, y_{t_k})$. To achieve this we propose to use a sequential Monte Carlo (SMC) algorithm (Doucet *et al.*, 2001). The distributions are approximated by a large number N of weighted random samples $\{X_{t_0:t_k}^{(i)}\}$ $(i = 1, \ldots, N)$ termed particles. The particles are sampled by using

$$X_{t_0:t_k}^{(i)} \sim q_k(x_{t_0:t_k}|y_{t_1:t_k})$$

with

$$q_k(x_{t_0:t_k}|y_{t_1:t_k}) = \pi_0(x_{t_0}) \prod_{n=1}^{k} q_n(x_{t_n}|y_{t_n}, x_{t_{n-1}})$$

where $\{q_n(x_{t_n}|y_{t_n}, x_{t_{n-1}})\}$ are importance distributions known pointwise. In the standard SMC framework, these particles should be reweighted according to normalized weights proportional to

$$\begin{aligned} w_k(X_{t_0:t_k}^{(i)}) &= \frac{p(X_{t_0:t_k}^{(i)}, y_{t_1:t_k})}{q_k(X_{t_0:t_k}^{(i)}|y_{t_1:t_k})} \\ &= w_{k-1}(X_{t_0:t_{k-1}}^{(i)}) \frac{\tilde{p}_{\Delta t_k}(X_{t_{k-1}}^{(i)}, X_{t_k}^{(i)})\, g(y_{t_k}|X_{t_k}^{(i)})}{q_k(X_{t_k}^{(i)}|y_{t_k}, X_{t_{k-1}}^{(i)})} \end{aligned}$$

where

$$\begin{aligned} \tilde{p}_{\Delta t_k}(x_{t_{k-1}}, x_{t_k}) &= \mathbb{P}(X_{t_k} \in \mathrm{d}x_{t_k}|X_{t_{k-1}} = x_{t_{k-1}})/\mathrm{d}x_{t_{k-1}} \\ &= \mathcal{N}_{\Delta t_k}(x_{t_k} - x_{t_{k-1}}) \exp\{A(x_{t_k}) - A(x_{t_{k-1}})\}\, a(x_{t_{k-1}}, x_{t_k}) \end{aligned}$$

with

$$A(u) = \int^{u} \alpha(z)\,\mathrm{d}z$$

and

$$a(x_{t_{k-1}}, x_{t_k}) = \mathbb{E}_{\mathbb{W}^{(x_{t_{k-1}}, x_{t_k})}}\left[\exp\left\{-\tfrac{1}{2}\int_{t_{k-1}}^{t_k} (\alpha^2 + \alpha')(\omega_s)\,\mathrm{d}s\right\}\right].$$

The particles are resampled whenever the variance of $\{w_k(X_{t_0:t_k}^{(i)})\}$ is too large. Clearly this SMC algorithm cannot be implemented as $\tilde{p}_{\Delta t_k}(x_{t_{k-1}}, x_{t_k})$ does not admit a closed form expression. However, a straightforward argument shows that it is not necessary to know $w_k(X_{t_0:t_k}^{(i)})$ exactly. Only an unbiased positive estimate $\hat{w}_k(X_{t_0:t_k}^{(i)})$ of $w_k(X_{t_0:t_k}^{(i)})$ is necessary to obtain asymptotically consistent SMC estimates under weak assumptions. Hence all the techniques that were developed by the authors to estimate $\tilde{p}_{\Delta t_k}(x_{t_{k-1}}, x_{t_k})$ unbiasedly

can be applied straightforwardly. The need for positive estimates restricts us to diffusions that are similar to those of the EA1 or EA2 algorithms.

To be efficient, the SMC method requires the design of 'good' importance distributions and to obtain estimates of the importance weights with low variance. To design the importance distributions, approximate analytically $\tilde{p}_{\Delta t_k}(x_{t_{k-1}}, x_{t_k})$ by a Gaussian distribution using a local linearization technique (Durham and Gallant, 2002) and combine this approximate prior with $g(y_{t_k}|x_{t_k})$ or a linearized version of it to obtain $q_k(x_{t_k}|y_{t_k}, x_{t_{k-1}})$.

It is also crucial to reduce the variance of the estimates of the weights given by

$$\mathrm{var}\{\hat{w}_k(X_{t_0:t_k})\} = \mathrm{var}\{w_k(X_{t_0:t_k})\} + \mathbb{E}[\mathrm{var}\{\hat{w}_k(X_{t_0:t_k})|X_{t_0:t_k}\}].$$

To achieve this, if the Poisson estimator of Section 6 is used, we could sample for each particle P Poisson random variables but use the same Brownian bridge to sample retrospectively for computational savings. However, a large Poisson parameter λ and a large P may be needed to obtain a reasonable variance.

Osnat Stramer (*University of Iowa, Iowa City*)

This paper proposes a new, thorough and interesting approach for calculating the likelihood of discretely observed diffusions. The log-likelihood of the data set v is

$$l(\theta|v) = \sum_{i=1}^{n} \log\{p_{\Delta t_i}(V_{t_{i-1}}, V_{t_i}; \theta)\},$$

where the transition density of the diffusion process V is $p_{\Delta t_i}(V_{t_{i-1}}, V_{t_i}; \theta)$. I thank the authors for providing a mine of ideas for modelling continuous time models.

In my comment, I shall also refer to two other existing approaches for estimating the transition density. One approach is the simulation methods that were proposed by Pedersen (1995) and improved substantially by Durham and Gallant (2002). This approach involves two types of approximations or errors:

(a) due to discretization and
(b) due to averaging.

Increasing the number of time intervals per Δ_{t_i} reduced the bias, but at the cost of increasing the number of simulations (see Stramer and Jun (2005a)). The second error will not disappear even if we use exact simulation. The methodology that is introduced in this paper avoids the first error but its efficiency depends heavily on the efficiency of the exact simulation of conditional diffusions. As pointed out by the authors, the acceptance rate of the Poisson process typically decreases exponentially to 0 in the length of the interval Δt_i between adjacent observations.

Another approach for estimating the transition probability is the closed form method that was introduced in Aït-Sahalia (2002a, 2004). It has been shown in Aït-Sahalia (2002b) and Stramer and Jun (2005b) that the closed form methods can be much faster and more accurate than the Durham and Gallant (2002) method in many interesting examples, though in theory it is assumed that Δt_i is small. It would be interesting to see a numerical comparison of the methods that were presented in this paper with the simulation-based methods and the closed form methods.

The methodology that is presented in this paper requires that the diffusion process be transformed to a new one with unit volatility. The simulation approach is in general more amenable when this transformation can be done (see Durham and Gallant (2002)). As noted by the authors a multivariate extension of their methods is therefore only possible for a limited class of models. More investigation in this direction will be necessary.

The **authors** replied later, in writing, as follows.

Firstly we thank all the discussants for their many contributions, insights and thought-provoking questions. The area of inference for partially observed diffusions is rapidly developing and hopefully our paper and the subsequent discussion will add further momentum to this exciting field.

The scope of the exact algorithm

A number of discussants (including Ball and his colleagues, Bingham, Chopin, Crisan, Kalogeropoulos, Moulines, Pasquali and Ruggeri, and Robert) asked about the scope of the exact algorithm (EA) methodology. Our paper introduces a general framework for inference which is illustrated mainly within the quite restrictive EA1 context and to some extent the more general EA2 situation.

We recall that the application of algorithm EA2 requires conditions (a)–(c) just below equation (4) to be satisfied, and the additional condition (9). Whereas (a)–(c) are largely innocuous, condition (9) is a serious constraint on the applicability of the method. For instance, many natural diffusion models violate this condition (including for instance the Ornstein–Uhlenbeck process). There are two ways in which the applicability of our methodology can be extended significantly beyond these constraints.

(a) As outlined in the paper, the Poisson estimator permits unbiased estimation for models outside $\mathcal{D}_2$. This permits the construction of a Monte Carlo EM algorithm for classes of diffusions that are outside $\mathcal{D}_2$. The potential problem with this approach is that finite variances of estimators are not guaranteed to exist.
(b) We have recently developed an extension of the EA called the EA3 algorithm (see Beskos *et al.* (2005a)) which is applicable to all diffusions in a class $\mathcal{D}_3$ which is described as diffusions which satisfy only conditions (a)–(c). Since $\mathcal{D}_3$ is not constained by restriction (9), it covers a very broad class of models.

Algorithm EA3 is mathematically more complex than either EA1 or EA2, though not necessarily less efficient. It is based on the idea that, once upper and lower bounds for the trajectory are known, the usual rejection sampling algorithm is easy to implement. For this, it introduces a different decomposition of Brownian motion, which we call *layered Brownian motion*. This decomposition permits upper and lower bounds on the diffusion sample path to be imposed once the sample path *layer* has been constructed. What makes algorithm EA3 more complicated than algorithm EA2 is the fact that the joint law of Brownian motion and its maximum modulus up to a fixed time t is not easily tractable, and exact simulation requires the construction of specific retrospective simulation techniques for probabilities that are expressible as infinite sums of alternating terms.

A minor modification of algorithm EA3 can also be used to cover other cases. For instance, by a judicious choice of the layers that are used in the construction, diffusions with finite entrance boundaries (and which are therefore not reached with probability 1) such as the shape space Ornstein–Uhlenbeck model of Ball and his colleagues can be simulated exactly. However, we also appreciate that the approximate approach that is adopted in their contribution represents sound methodology, since it seems easy to ensure that the approximation that is involved is arbitrarily accurate.

Several discussants enquired about other extensions. Beskos *et al.* (2005a) also shows that extensions of the EA to the time inhomogeneous case (as queried by Pasquali and Ruggeri) is straightforward, although the function ϕ must be modified accordingly. Moreover extensions to the multivariate case (as mentioned by Kalogeropoulos and Bingham) is routine provided that the following two (non-trivial) conditions are met.

(a) The multivariate version of the transformation that is given in equation (3) needs to be possible. Although equation (3) is always possible for scalar diffusions, its multivariate counterpart (translating a diffusion with arbitrary diffusion coefficient to a diffusion with unit diffusion coefficient) may not exist. Conditions for this are given in Aït-Sahalia (2004) as mentioned in the contributions of Kalogeropoulos and Bingham.
(b) Assuming that condition (3) is possible, the resulting multivariate drift function must be of gradient type, i.e. there is a function $V(\mathbf{x})$ such that the drift $\boldsymbol{\alpha}(\mathbf{x})$ satisfies

$$\boldsymbol{\alpha}(\mathbf{x}) = \nabla A(\mathbf{x}).$$

This is a very common and natural condition for diffusions which corresponds (in the stationary case) to the reversibility of the diffusion.

However, when these two conditions are not met, it seems difficult to see how the EA can be applied at all.

Kalogeropoulos mentions the interesting extension to non-Markov models. Stochastic differential equations are very naturally formulated in non-Markov settings, and existing data augmentation techniques can in principle handle this generalization. Furthermore this generalization is important for certain application areas such as finance where Markovianity can be doubtful. (See Dellaportas *et al.* (2004) for an example of a Bayesian analysis of a non-Markov model by data augmentation.) Except in special cases, the EA cannot be applied for non-Markov models.

Pavlopoulos discusses two generalizations. Firstly the case of regular accessible boundaries for diffusions, for instance with reflection at the boundary, is complicated. Where diffusion drifts are locally bounded and converging to 0 at the boundary, and volatilities are bounded away from 0, pure reflecting

boundaries can easily be dealt with by *unfolding* the reflecting boundary producing mirror image dynamics either side of the boundary. This approach works for two-sided boundaries also. However, it is not clear how to extend this to the case where the drift is not converging to 0 at the boundary since now the *unfolded* diffusion has discontinuous drift.

Secondly, Pavlopoulos asks about varying Δ_t. In fact this causes no problem for our methodology since all the Monte Carlo procedures are carried out independently between each two consecutive observations.

Kluppelberg and Mamon and Yu ask the intriguing question about whether our methodology can be extended to deal with processes that are driven by Lévy process noise. To date, we have no idea how this can be accomplished in the infinite activity Lévy case. However, we certainly agree that this is an important open problem arising from our work here.

Numerical investigations and comparisons

Several discussants (Clifford, Crisan, Lyons and Stramer) request further comparisons with existing methodology and we agree completely that this is needed.

There are two (not entirely mutually exclusive) stages to this work. Firstly are numerical investigations into the effectiveness of the EA for different diffusions in comparison with competing discretization schemes. Important progress towards this goal has been achieved in Casella's comment and in more detail in his doctoral thesis.

Furthermore, there are tight analytic bounds on the computational cost of the EA in Beskos and Roberts (2005) and Beskos *et al.* (2004). For instance in the recast von Mises example in the interesting discussion of Kent, Beskos and Roberts (2005) demonstrated that to simulate the diffusion over a time interval of length T requires computational effort of order $\lambda^2 T/\sigma^2$.

The second stage in a comparison study will involve comparisons of the methods of this paper for likelihood-based inference with competitor methodologies. One important issue here will be that, although we would expect most (or maybe all) methods to deteriorate for increasingly sparse data sets, it is of great interest to know winch methods show the greatest level of robustness to this phenomenon. For the Markov chain Monte Carlo data augmentation methodology for instance this involves studying the effects of block updating strategies for data sets where imputing the entire missing data between observations is inefficient. Some interesting conclusions in this direction appear in Chib *et al.* (2004).

For this comparison work, we emphasize again the important distinction between methods that can smoothly estimate whole likelihood surfaces and those which are focused on pointwise estimation. The former methods are highly likely to be the more effective in most statistical contexts. Smoothness (more precise) of the estimated likelihood surface is required for the proof of consistency (for large Monte Carlo samples) of maximum likelihood estimates as demonstrated in Beskos *et al.* (2005b). Lyons suggests comparisons with numerical methods for solving the partial differential equations that are satisfied by the diffusion transition density. This approach was explored in Lo (1988) although it appears difficult to estimate entire likelihood surfaces simultaneously by using this method.

Theoretical foundations

Kessler asks interesting questions about the properties of estimators that are produced by our Monte Carlo methodology. We have only partial answers to this question, showing for instance in Beskos *et al.* (2005b) that, as the Monte Carlo sample size increases, the Monte Carlo maximum likelihood estimate converges almost surely to the true maximum likelihood estimate. Furthermore, the Monte Carlo EM method can give an unbiased estimate of the score function, and thus we would expect that, even for fixed Monte Carlo sample size per observation, the Monte Carlo EM maximum likelihood estimate would be consistent in the limit as the sample size increases. These are both important consequences of unbiasedness in the estimators. We certainly concur that more quantitative rate results would be desirable.

However, although our methods are based on Monte Carlo methods, as pointed out by Robert, the transparency of our methods readily permits variance reduction ideas which mean that, in the examples that we have considered so far, very robust estimates of whole likelihood surfaces are obtained using minimal computational effort.

Filtering and the Poisson estimator

Robert and Stramer ask for guidance about the optimal implementation of the Poisson estimator. In recent work (Fearnhead *et al.*, 2005) we have generalized the Poisson estimator to permit a general distribution for the number of time points at which the diffusion is evaluated, and we have produced optimality criteria for this distribution. Motivated by these criteria, a sensible proposal distribution appears to be a negative binomial distribution whose mean is close to $\int_0^t \{c - f(\omega_s)\}\,\mathrm{d}s$. In the EA1 setting of the SINE example,

such a choice can lead to a reduction in variance by two or three orders of magnitude over the Poisson estimator with parameters that are chosen as in Section 7.

There are several comments concerning how our work can be applied to filtering problems, e.g. where we make partial observations of the underlying diffusion at discrete time points. Chopin points out that the EA enables a simple implementation of the basic particle filter (Gordon *et al.*, 1993) as the EA enables us to simulate exactly the value of the state at the next time point given the current value of the state. Künsch suggests that the rejection-sampling-based particle filter of Hürzeler and Künsch (1998) could be applied, and Doucet suggests a more general particle filter where the Poisson estimator is used to generate random weights for each particle, with the mean of these random weights being equal to the true (analytic) weight.

We have independently been investigating how the EA and related ideas can be applied to particle filters, and in Fearnhead *et al.* (2005) we have developed a general framework for *unbiased* particle filters. Here we give a brief summary of results in the case of observing the SINE data set with normal error (variance σ^2). For this application, the rejection sampling idea of Künsch can be applied by

(a) choosing a current particle with probability proportional to

$$\exp\left\{-A(x_{t_i}^{(k)}) - \frac{(y_{i+1} - x_{t_i}^{(k)})^2}{2(\sigma^2 + t_{i+1} - t_i)}\right\},$$

(b) proposing $X_{t_{i+1}} = u$ conditional on $x_{t_i}^{(k)}$ from a normal distribution with mean

$$\frac{x_{t_i}^{(k)}\sigma^2 + (t_{i+1} - t_i)y_{i+1}}{\sigma^2 + t_{i+1} - t_i},$$

and variance $\sigma^2(t_{i+1} - t_i)/(\sigma^2 + t_{i+1} - t_i)$, and

(c) performing two accept–reject steps, the first with acceptance probability $\exp\{A(u) - 1\}$, and the second being the EA1 accept–reject step for the path.

This algorithm has the same overall acceptance probability as a simple implementation of algorithm EA1 for simulating from the SINE state diffusion, but it has the important advantage of simulating from the particle filter's target distribution at time t_{i+1} and avoiding the need for an importance sampling correction by the likelihood. This is particularly useful if σ^2 is small relative to $t_{i+1} - t_i$. However, an efficient *random-weight* particle filter (similar to the ideas of Rousset and Doucet) has computational advantages over even this rejection algorithm. The advantages depend on the value of the states at time t_i and t_{i+1}—but in extreme cases (where both state values are close to a multiple of 2π) computational gains of over a factor of 10 are possible.

As Rousset and Doucet point out, to implement a random-weight particle filter requires the random weights to be positive with probability 1. Although this is straightforward to achieve in EA1 situations, we have adapted EA3 ideas to ensure this for general diffusions. The methodology that we have developed is trivially extended to inference for diffusion-driven Cox processes (Møller and Clifford), and it should be possible to extend to stochastic volatility models (Kalogeropoulos) and velocity models where positions are observed (Kent). For full details see Fearnhead *et al.* (2005).

Mathematical formulation

Crisan suggests that it is more natural to avoid the step which transforms the observed diffusion to one of unit diffusion coefficient. It is clear that this is mathematically equivalent to our approach, and the Radon–Nikodym derivative that is obtained by using both approaches should be almost surely equivalent. In fact, the solution to the differential equation that he gives in his equation (32) turns out to define our unit diffusion coefficient transformation (3) in the time homogeneous case, and the time inhomogeneous case results in a simple generalization of transformation (3) which we describe in detail in Beskos *et al.* (2005a).

In fact, from a statistical perspective, it seems to us to be more natural to work with the transformed unit volatility process since this can be thought of as an infinitesimal standardization of the missing data.

Implicit in Crisan's comment is the useful idea that different dominating measure proposal distributions can be used as the basis for the rejection sampling approach. There are several different tractable unit diffusion coefficient proposals which could be used in this way: for example in the Cox–Ingersoll–Ross example we used a Bessel process, and in more recent work we have used Ornstein–Uhlenbeck processes. However, Crisan's approach offers no additional generality since any general diffusion which has explicit

finite dimensional distributions corresponds (through transformation(3)) to one which has unit diffusion coefficient.

Miscellaneous

We concur with the view of Clifford, Künsch, and Mamon and Yu, who mention both the fact that diffusions that are used in statistical modelling are often rather approximate, but that precise macroscopic properties of models are likely to be unimportant in many practical modelling contexts. For instance in financial data, stylized facts about financial asset power variation are at odds with the characteristic finite non-zero quadratic variation which diffusions are bound to exhibit.

Clifford points out that the exact likelihood in the Cox–Ingersoll–Ross model is explicit. This is precisely why we used this example so that the analytic solution could be explicitly compared with our results.

Chopin raises the possibility of substantially increasing efficiency at the cost of a certain bias by truncating drifts in regions which the diffusion is unlikely to visit in the time period of interest. This seems a sensible practical suggestion. It turns out that algorithm EA3 using the layered Brownian motion construction can be motivated in a similar way, and moreover allows unbiasedness to be retained.

Pasquali and Ruggeri and Genon-Catalot ask about the relationship between parameter estimation and diffusion stability. In the asymptotic case, much is known about this through the elegant theory of mixed asymptotic normality (see Basawa and Rao (1980)). In the finite sample case, it is difficult to make general statements. However, as the discussants point out, even our simple logistic growth model is sufficiently simple for interesting structure to emerge. In this example for instance, for $\sigma^2 > 2R$, the diffusion will eventually converge to 0 so presumably K will be badly estimated in this case.

Mamon and Yu ask about the comparisons of our methods with nonparametric methods for stationary diffusions. Kent mentions a similar issue, pointing out that serial correlation of the data can be taken into account by standard time series methods. Although this idea requires stationarity, an appealing feature of such methodology is the fact that it will provide an increasingly good approximation to an exact likelihood-based approach for sparse data—exactly the situation which is problematic for all currently existing exact likelihood-based methods.

References in the discussion

Aït-Sahalia, Y. (1996) Nonparametric pricing of interest rate derivative securities. *Econometrica*, **64**, 527–560.

Aït-Sahalia, Y. (2002a) Maximum likelihood estimation of discretely sampled diffusions: a closed form approximation approach. *Econometrica*, **70**, 223–262.

Aït-Sahalia, Y. (2002b) Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes: comment. *J. Bus. Econ. Statist.*, **20**, 317–321.

Aït-Sahalia, Y. (2004) Closed-form likelihood expansions for multivariate diffusions. *Working Paper*. Princeton University, Princeton. (Available from `http://www.princeton.edu/~yacine/research.htm`.)

Ball, F. G., Dryden, I. L. and Golalizadeh, M. (2004) Brownian motion and Ornstein-Uhlenbeck processes in planar shape space. *Technical Report 04-11*. Division of Statistics, University of Nottingham, Nottingham.

Bandi, F. M. and Phillips, P. C. B. (2003) Fully nonparametric estimation of scalar diffusion models. *Econometrica*, **71**, 241–283.

Bansal, R. and Zhou, H. (2002) Term structure of interest rates with regime shifts. *J. Finan.*, **57**, 1997–2042.

Barndorff-Nielsen, O. E. and Shephard, N. (2002) Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *J. R. Statist. Soc.* B, **64**, 253–280.

Basawa, I. V. and Rao, B. L. S. P. (1980) *Statistical Inference for Stochastic Processes*. New York: Academic Press.

Beskos, A., Papaspiliopoulos, O. and Roberts, G. O. (2004) Retrospective exact simulation of diffusion sample paths with applications. To be published. (Available from `http://www.maths.lancs.ac.uk/~papaspil/research.html`.)

Beskos, A., Papaspiliopoulos, O. and Roberts, G. O. (2005a) A new factorisation of diffusion measure with view towards simulation. To be published.

Beskos, A., Papaspiliopoulos, O. and Roberts, G. O. (2005b) Monte Carlo maximum likelihood estimation for discretely observed diffusion processes. To be published.

Beskos, A. and Roberts, G. O. (2005) Exact simulation of diffusions. *Ann. Appl. Probab.*, **15**, in the press.

Bingham, N. H. and Dunham, B. (1997) Estimating diffusion coefficients from count data. *Ann. Inst. Statist. Math.*, **49**, 667–679.

Bingham, N. H., Kiesel, R. and Schmidt, R. (2003) A semi-parametric approach to risk management. *Quant. Finan.*, **3**, 426–441.

Bingham, N. H. and Pitts, S. M. (1999) Non-parametric estimation for the M/G/infinity queue. *Ann. Inst. Statist. Math.*, **51**, 71–97.

Chib, S., Pitt, M. K. and Shephard, N. (2004) Likelihood-based inference for diffusion models. *Working Paper*. Nuffield College, Oxford. (Available from `http://www.nuff.ox.ac.uk/users/shephard/`.)

Clifford, P. and Wei, G. (1993) The equivalence of the Cox process with squared radial Ornstein-Uhlenbeck intensity and the death process in a simple population model. *Ann. Appl. Probab.*, **3**, 314–337.
Connor, S. B. and Kendall, W. S. (2006) Perfect simulation for a class of positive recurrent Markov chains. Submitted to *Ann. Appl. Probab*.
Dellaportas, P., Friel, N. and Roberts, G. O. (2004) Bayesian model selection for partially observed diffusion models. To be published.
Doss, H. (1977) Liens entre équations différentielles et ordinaires. *Ann. Inst. H. Poincaré* B, **13**, 99–125.
Douc, R., Guillin, A., Marin, J.-M. and Robert, C. P. (2005) Minimum variance importance sampling via Population Monte Carlo. *Cahiers du Ceremade*. Université Paris Dauphine, Paris.
Doucet, A., de Freitas, N. and Gordon, N. J. (eds) (2001) *Sequential Monte Carlo Methods in Practice*. New York: Springer.
Durham, G. B. and Gallant, A. R. (2002) Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes (with discussion). *J. Bus. Econ. Statist.*, **20**, 297–338.
Elerian, O. S., Chib, S. and Shephard, N. (2001) Likelihood inference for discretely observed non-linear diffusions. *Econometrica*, **69**, 959–993.
Eraker, B. (2001) Markov chain Monte Carlo analysis of diffusion models with application to finance. *J. Bus. Econ. Statist.*, **19**, 177–191.
Fan, J. and Zhang, C. (2003) A reexamination of diffusion estimators with applications to financial model validation. *J. Am. Statist. Ass.*, **98**, 118–134.
Fearnhead, P., Papaspiliopoulos, O. and Roberts, G. O. (2005) Particle filters for partially observed diffusions. To be published.
Foss, S. G. and Tweedie, R. L. (1998) Perfect simulation and backward coupling. *Stochast. Mod.*, **14**, 187–203.
Gelman, A. (1995) Methods of moments using monte-carlo simulation. *J. Comput. Graph. Statist.*, **4**, 36–54.
Golightly, A. and Wilkinson, D. (2005) Bayesian sequential inference for nonlinear multivariate diffusions. *Working Paper*. University of Newcastle, Newcastle upon Tyne. (Available from `http://www.staff.ncl.ac.uk/d.j.wilkinson/pub.html`.)
Gordon, N., Salmond, D. and Smith, A. F. M. (1993) Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE. Proc.* F, **140**, 107–113.
Guo, X. and Zhang, Q. (2004) Closed-form solutions for perpetual American put options with regime-switching. *SIAM J. Appl. Math.*, **64**, 2034–2049.
Hill, N. A. and Häder, D.-P. (1997) A biased random walk model for the trajectories of swimming micro-organisms. *J. Theoret. Biol.*, **186**, 503–526.
Hürzeler, M. and Künsch, H. R. (1998) Monte Carlo approximations for general state space models. *J. Comput. Graph. Statist.*, **7**, 175–193.
Ibragimov, I. A. and Rozanov, Y. A. (1978) *Gaussian Random Processes*. Berlin: Springer.
Jarner, S. F. and Roberts, G. O. (2002) Polynomial convergence rates of Markov chains. *Ann. Appl. Probab.*, **12**, 224–247.
Jiang, G. J. and Knight, J. L. (1997) A nonparametric approach to the estimation of diffusion processes, with an application to a short-term interest rate model. *Econometr. Theory*, **13**, 615–645.
Kalogeropoulos, K. (2006) Likelihood-based inference for a class of multivariate diffusions with unobserved paths. To be published.
Kalogeropoulos, K., Dellaportas, P. and Roberts, G. (2006a) Likelihood based inference for multidimensional diffusion models using data augmentation. *Working Paper*.
Kalogeropoulos, K., Roberts, G., and Dellaportas, P. (2006b) Bayesian inference on stochastic volatility diffusion models using time change transformations. *Working Paper*.
Karlin, S. and Taylor, H. M. (1981) *A Second Course in Stochastic Processes*. New York: Academic Press.
Kendall, D. G. (1984) Shape manifolds, Procrustean metrics and complex projective spaces. *Bull. Lond. Math. Soc.*, **16**, 81–121.
Kendall, W. S. (1998) Perfect simulation for the area-interaction point process. In *Probability Towards 2000* (eds L. Accardi and C. Heyde), pp. 218–234. New York: Springer.
Kendall, W. S. (2004) Geometric ergodicity and perfect simulation. *Electron. Communs Probab.*, **9**, 140–151.
Kent, J. T. (1975) Discussion on 'Statistics of directional data' (by K. V. Mardia). *J. R. Statist. Soc.* B, **37**, 377–378.
Kent, J. T. (1978) Time reversible diffusions. *Adv. Appl. Probab.*, **10**, 819–835.
Kessler, M. and Paredes, S. (2002) Computational aspects related to martingale estimating functions for a discretely observed diffusion. *Scand. J. Statist.*, **29**, 425–440.
Klüppelberg, C., Lindner, A. and Maller, R. (2004) A continuous time GARCH process driven by a Lévy process: stationarity and second order behaviour. *J. Appl. Probab.*, **41**, no. 3, 1–22.
Lo, A. W. (1988) Maximum likelihood estimation of generalized ito processes with discretely sampled data. *Econometr. Theory*, **4**, 231–247.
McCullagh, P. and Møller, J. (2005) The permanent process. *Research Report R-2005-29*. Department of Mathematical Sciences, Aalborg University, Aalborg.
Müller, G. (2005) MCMC estimation of the COGARCH(1,1) model. To be published.

Pavlopoulos, H. and Kedem, B. (1992) Stochastic modeling of rain rate processes: a diffusion model. *Stochast. Mod.*, **8**, 397–420.

Pedersen, A. R. (1995) Consistency and asymptotic normality of an approximate maximum likelihood estimator for discretely observed diffusion processes. *Bernoulli*, **1**, 257–279.

Propp, J. G. and Wilson, D. B. (1996) Exact sampling with coupled Markov chains and applications to statistical mechanics. *Rand. Struct. Algrthms*, **9**, 223–252.

Quintana, F. A., Liu, J. and Del Pino, G. (1999) Monte-Carlo EM with importance reweighting and its applications in random effects models. *Comput. Statist. Data Anal.*, **29**, 429–444.

Roberts, G. O., Papaspiliopoulos, O. and Dellaportas, P. (2004) Bayesian inference for non-Gaussian Ornstein–Uhlenbeck stochastic volatility processes. *J. R. Statist. Soc.* B, **66**, 369–393.

Roberts, G. and Stramer, O. (2001) On inference for partial observed non-linear diffusion models using the metropolis-hastings algorithm. *Biometrika*, **88**, 603–621.

Rogers, L. C. G. and Williams, D. (1994) *Diffusions, Markov Processes and Martingales*, vol. 1, *Foundations*, 2nd edn. New York: Wiley.

Srinivasan, S. K. (1988) *Point Process Models of Cavity Radiation and Detection*. London: Griffin.

Stanton, R. (1997) A nonparametric model of term structure dynamics and the market price of interest rate risk. *J. Finan.*, **52**, 1973–2002.

Stramer, O. and Jun, Y. (2005a) Asymptotics of an efficient Monte Carlo estimation for the transition density of diffusion processes. To be published.

Stramer, O. and Jun, Y. (2005b) Parametric inference for partially observed diffusion processes: a comparison study. To be published.

Stroock, D. and Varadhan, S. R. S. (1979) *Multidimensional Diffusions*. New York: Springer.

Sussmann, H. J. (1978) On the gap between deterministic and stochastic ordinary differential equations. *Ann. Probab.*, **6**, 19–41.

Wu, S. and Zeng, Y. (2004) Affine regime-switching for interest rate term structure. In *Mathematics of Finance* (eds G. Yin and Q. Zhang), pp. 375–386. Providence: American Mathematical Society.