Likelihood based inference for observed and partially observed diffusions

Siddhartha Chib

Olin School of Business, Washington University, St Louis, USA

chib@olin.wustl.edu

Michael K Pitt

Department of Economics, University of Warwick, Coventry CV4 7AL, UK

m.k.pitt@warwick.ac.uk

Neil Shephard

Nuffield College, University of Oxford, Oxford OX1 1NF, UK

neil.shephard@nuf.ox.ac.uk

April 24, 2003

Abstract

This paper provides methods for carrying out likelihood based inference on non-linear observed and partially observed non-linear diffusions. The diffusions can potentially be non-stationary. The methods are based on innovative Markov chain Monte Carlo methods combined with an augmentation strategy. We study the performance of the methods as the degree of augmentation goes to infinity and find that the methods are robust. Various extensions to the modelling framework are provided, while we also discuss issues of model choice, model checking and filtering.

Keywords: Bayes estimation, Brownian bridge, Non-linear diffusion, Euler approximation, Markov chain Monte Carlo, Metropolis-Hastings algorithm, Missing data, Simulation, Stochastic differential equation.

1 Introduction

Consider a $d-$dimensional stochastic process $\alpha(t) = (\alpha_1(t), ..., \alpha_d(t))'$ that satisfies an Ito stochastic differential equation (SDE) of the form,

$$\mathrm{d}\alpha(t) = \mu\{\alpha(t); \theta\}\mathrm{d}t + \Theta\{\alpha(t); \theta\}\mathrm{d}w(t), \tag{1}$$

where $\mu : d \times 1$ and $\Theta : d \times d$ are the non-anticipative drift and volatility functions, respectively, depending on α and an unknown parameter vector θ, and $w(t) = (w_1(t), ..., w_d(t))'$ is a vector of independent standard Brownian motions. Suppose that the stochastic process $\{\alpha(t)\}$ is *fully* or *partially* observed at the sequence of time points

$$0 < \tau_1 < \tau_2 < ... < \tau_T,$$

where, for simplicity, the time gap $\tau_{i+1} - \tau_i > 0$ is assumed constant and equal to Δ. Let $Z_i : k_i \times d$ $(k_i \leq d)$ denote a known selection matrix and assume that we only have the discrete time measurements

$$y_i = Z_i\alpha(\tau_i),\ i = 1, 2, ..., T.$$

The fully observed case arises when the rank of Z_i is d of which the leading example is $Z = I_d$. We get the partially observed case when the rank of Z_i is less than d. For simplicity of exposition, assume that Z_i does not vary over i.

The principal aim of the analysis is to estimate θ given the $T \times k$ matrix of measurements $y = (y_1, ..., y_T)'$. In the partially observed case we may also be interested in estimating the corresponding $T \times d$ matrix of unobserved states $(\alpha(\tau_0), ..., \alpha(\tau_T))'$.

In the likelihood context, the parameter vector θ is estimated by maximizing the conditional log-likelihood function

$$\log g(y|\theta) = \log g(y_2, ..., y_T|\mathcal{F}_1, \theta) = \sum_{i=1}^{T-1} \log g(y_{i+1}|\mathcal{F}_i, \theta), \tag{2}$$

where $\mathcal{F}_i = \sigma(y_1, ..., y_i)$ is the sigma field generated by the discrete time observations. When the vector diffusion is fully observed, y_i is Markov and the conditional log-likelihood reduces to the relatively simple form

$$\log g(y_2, ..., y_T|\mathcal{F}_1, \theta) = \sum_{i=1}^{T-1} \log g(y_{i+1}|y_i, \theta).$$

For most practical SDE's, the form of the conditional density $g(y_{i+1}|\mathcal{F}_i, \theta)$ is not available, even in the Markov case. In response to this difficulty, a number of authors have pursued

non-likelihood based methods for estimating SDE's, for example, the indirect inference method (Smith (1993) and Gourieroux, Monfort, and Renault (1993)), efficient method of moments (Gallant and Tauchen (1996), Gallant and Long (1997) and Gallant and Tauchen (2003)), kernel-based estimation of the transition density (Ait-Sahalia (1996a), Ait-Sahalia (1996b) and Jiang and Knight (1997)), and the method of estimating functions (Kessler and Sorensen (1999), Sørensen (1997), Sørensen (2000), Florens-Zmirou (1989) and Hansen and Scheinkman (1995)).

At the same time, there has been considerable work on overcoming the difficulties inherent in a likelihood analysis, primarily through the use of simulation-based methods. Pedersen (1995) introduced the important idea of estimating $g(y_{i+1}|y_i, \theta)$ by simulation in conjunction with an augmentation procedure to reduce the error of the Euler discretization. Ait-Sahalia (2002), Egorov, Li, and Xu (2003) and Ait-Sahalia (2003) explore the use of infinite series Edgeworth-based expansions[1] to approximate $g(y_{i+1}|y_i, \theta)$, and Durham and Gallant (2002), building on Pedersen (1995) and, more directly, Elerian, Chib, and Shephard (2001), use importance sampling to approximate the transition density. Other similar approaches are outlined by Brandt and Santa-Clara (2002), Nicolau (2002) and Hurn, Lindsay, and Martin (2003). To the best of our knowledge, none of these extend to the general non-Markov case.

An atttractive alternative to these techniques is the estimation of SDE's by modern Markov chain Monte Carlo Bayesian methods as principally developed by Elerian, Chib, and Shephard (2001) and Roberts and Stramer (2001). There is also some Bayesian work on the partially observed case, by Kim, Shephard, and Chib (1998) in the context of stochastic volatility processes and by Roberts and Stramer (2001) and Eraker (2001) for general non-linear diffusions. One reason for pursuing the Bayesian approach is that it appears to be tailor-made for partially observed diffusions, the setting in which other approaches and techniques are not very effective. For example, importance sampling as advanced by Durham and Gallant (2002) becomes infeasible because of a large increase in the dimension of the space of latent-variables (which now includes the entire missing states — Durham and Gallant (2002) do consider the non-Markov case but are forced to employ techniques other than importance sampling) while Edgeworth-based methods become difficult to generalize. On the other hand, Bayesian MCMC methods remain feasible because they can be used to split up the parameter space into a set of lower dimensional spaces, a sort of "divide-and conquer" strategy that (to our knowledge) has no parallel in other methods.

[1]Schaumburg (2002) extended this work to cover stochastic differential equations driven by Lévy processes.

1.1 Outline of the paper

The structure of this paper is as follows. In Section 2 we provide a brief overview of our augmentation procedure for estimating vector diffusions. Essentially, the approach amounts to a discretization of the SDE by the Euler method on a time scale that is finer than that of the observed data. In addition we outline the structure of our Markov chain Monte Carlo (MCMC) method for performing inference for both observed and partially observed SDE's. As is standard in the context of MCMC methods, the inferential approach does not require computation of the likelihood function. In Section 3 we present the implementation of the algorithm starting with the fully observed SDE and provide details on how the SDE can be sampled on our fine time scale conditioned on the data and the parameters. Our sampling technique is built around a distribution proposed by Durham and Gallant (2002) in the context of importance sampling. We find that a robustified version of that distribution when used as proposal density within a Metropolis-Hastings step leads to an algorithm which is highly effective even when the degree of augmentation is large.

In Section 4 we extend our ideas to partially observed diffusions. For this case, we develop a different proposal density for our Metropolis step. We illustrate the method with an example in which the SDE has an analytic solution. We also consider a more complex stochastic volatility diffusion.

In Section 5 we illustrate the methods on real data within the context of several stochastic volatility diffusions. Disussion of some related literature as well as extensions of our work are given in Section 6 while Section 7 has our conclusions. A proof of one of the results given in the paper is provided in the Appendix.

2 Augmentation and inference

We begin by describing the augmentation method that lies at the core of the Bayesian approach for fitting both the fully observed and partially observed diffusions. The link between the augmentation method and the rest of the prior-posterior Bayesian analysis is developed in Section 2.2. For the fully observed case, specific details related to implementation are contained in Section 3 whereas those for partially observed diffusions are in Section 4.

2.1 Augmentation

The main problem in fitting SDE's is that the the transtion density $g(y_{i+1}|\mathcal{F}_i, \theta)$, and consequently the conditional likelihood $g(y|\theta)$, are not generally available in closed form. Although

this may appear to be an insuperable problem, the computation of the conditional likelihood can be by-passed entirely by a simulation-based augmentation strategy that amounts to treating the continuous path of the stochastic process between each pair of observed measurements as "missing" and then iterating on a two-step process in which the that missing path is restored given the current values of the parameters, followed by inferences on the parameters given the restored missing measurements. The idea of simulation combined with augmentation in this context goes back to Pedersen (1995), Gourieroux, Monfort, and Renault (1993) and Gallant and Tauchen (1996) though our specific Bayesian implementation is based on Elerian, Chib, and Shephard (2001) and Eraker (2001).

To review some of the specifics, suppose that within successive pair of time points $\{\tau_i, \tau_{i+1}\}$ of length Δ, we introduce $M-1$ equally-spaced intermediate time points

$$\tau_i < \tau_i + \delta < \tau_i + 2\delta < ... < \tau_i + (M-1)\delta < \tau_i + M\delta = \tau_{i+1}$$

where $\delta = \Delta/M$. Now for each $i = 1, 2, ..., T$, let the values of the diffusion at this finer time scale be

$$\alpha_{(i-1)M+j} = \alpha(\tau_i + j\delta), \quad j = 0, 1, 2, ..., M,$$

where we let $\alpha(\tau_i) = \alpha_{(i-1)M}$. Now collect these values across i in a $\{(T-1)M+1\} \times d$ vector

$$\alpha^+ = (\alpha_0, \alpha_1, ..., \alpha_{(T-1)M})'.$$

The idea is to solve the SDE by considering the discretized Euler version of the Ito SDE that operates on the time scale of the augmented time points. This discretrized version is given by

$$\alpha_{j+1}|\alpha_j, \theta \sim N_d(\alpha_j + \mu(\alpha_j;\theta)\delta, \Sigma(\alpha_j;\theta)\delta), \quad j = 0, 1, 2 \ldots, \tag{3}$$

where $\Sigma(\cdot;\theta) = \Theta(\cdot;\theta)\}\Theta(\cdot;\theta)\}'$. From the transition density in (3) it is easy to see that by iterating M times and marginalising we get

$$f^M(\alpha_{iM}|\alpha_{(i-1)M}, \theta) = \int f(\alpha_{iM}|\alpha_i^*;\theta) f(\alpha_i^*|\alpha_{(i-1)M};\theta) \mathrm{d}\alpha_i^*, \tag{4}$$

where

$$\alpha_i^* = (\alpha_{(i-1)M+1}, \alpha_{(i-1)M+2}, ..., \alpha_{iM-1})'.$$

are the unobserved values of the SDE's on the intermediate time points in the interval $\{\tau_i, \tau_{i+1}\}$. This is the augmentation-based Euler approximation to the true density $g(\alpha(\tau_{i+1})|\alpha(\tau_i), \theta)$. It can be shown that

$$f^M(\alpha(\tau_{i+1})|\alpha(\tau_i), \theta) \to g(\alpha(\tau_{i+1})|\alpha(\tau_i), \theta), \quad \text{as} \quad M \to \infty,$$

see Pedersen (1995, Theorem 3) and Kohatsu-Higa and Ogawa (1997)). Hence, by operating on this fine time scale it is possible to reduce the bias of the error due to discretization and the gains can be substantial as documented by Elerian, Chib, and Shephard (2001) for univariate diffusions.

By way of additional notation, let us introduce the $M \times d$ block

$$\alpha_i^+ = (\alpha_{(i-1)M+1}, \alpha_{(i-1)M+2}, ..., \alpha_{iM-1}, \alpha_{iM})', \tag{5}$$

which means that α^+ can be expressed in the natural block form

$$\alpha^+ = (\alpha_0, \alpha_1^+, \alpha_2^+, ..., \alpha_{T-1}^+)'.$$

We may now write

$$f(\alpha^+|\theta) = f(\alpha_0|\theta) \prod_{i=1}^{T-1} f(\alpha_i^+|\alpha_{(i-1)M}, \theta).$$

where from (3)

$$f(\alpha_i^+|\alpha_{(i-1)M}; \theta) = \prod_{j=0}^{M-1} f(\alpha_{(i-1)M+j+1}|\alpha_{(i-1)M+j}, \theta).$$

The procedures that we develop in the sequel will base prior-posterior inferences on the quantity $f(\alpha^+|\theta)$ or, in effect, the likelihood built up from f^M instead of g. It is clear that if M is finite, as in any practical implementation, then all bias from discretization cannot be removed. However, this sort of error is unavoidable and also arises in (one form or another) in other simulation-based approaches such as the EMM, indirect inference and simulated maximum likelihood methods. Indeed, even the infinite series expansion approach of Ait-Sahalia (2002) cannot be exact since the infinte series expansion must be truncated at some point. Of course, the approximations are likely to be good and can be improved, in our case for example, by adjusting M.

2.2 Inference

The augmentation approach to dealing with diffusions means that the model can be expressed as

$$y_i = Z\alpha_{(i-1)M}, \quad i = 1, 2, ..., T,$$

where

$$\alpha_{j+1}|\alpha_j \sim N_d(\alpha_j + \mu(\alpha_j; \theta)\delta, \Sigma(\alpha_j; \theta)\delta), \quad j = 0, 1, 2, ...$$

On letting $\alpha^+ = (\alpha_0, \alpha_1, ..., \alpha_{(T-1)M})'$, Bayesian inference on $\theta|y$ can be carried out by simply sampling from the posterior[2]

$$f(\theta, \alpha^+|y),$$

and then discarding the α^+ draws, leading to samples from $\theta|y$. There are many ways in which this simulation can be carried out — some of these will involve enormous computational effort, others will be manageable. In the sequel we present those that are effective in the sense of being simulation efficient, applicable to both fully and partially observed diffusions, and scaleable in the dimension of the vector diffusion.

In the Bayesian context, the modern way of simulating from enormously large dimensional distributions, such as the posterior $\theta, \alpha^+|y$, is by MCMC methods. These methods proceed through the sampling of a Markov chain whose limiting (invariant) distribution is the posterior distribution of interest. The Markov chain is constructed by breaking up the parameter space into smaller, manageable blocks, and then recursively sampling each block conditioned on the data and the most recent values of the remaining blocks. A full discussion of the theory, along with the many issues related to implementation, are discussed by, for example, Chib (2001, pp. ?). In our problem, we construct the Markov chain simulation in the following way:

1. Update α^+ from

$$f(\alpha^+|y, \theta),$$

2. Update θ from

$$f(\theta|y, \alpha^+),$$

3. Goto 1.

It turns out that this simple sampling scheme can be used for both the observed and partially observed diffusions. Another point is that the sampling of the parameters in step 2 is relatively straightforward and model-specific. We shall therefore not have much to say about this step. Step 1 on the other hand is more crucial and involved and the viability of the Bayesian approach in this context depends on being able to sample α^+ in an effective way. We have developed such methods for both the fully and partially observed cases. We present the details as they relate to the fully observed diffusion in Section 3[3] and those for the partially observed case in Section 4, which naturally extends the observed case.

[2]Inevitably there are some degeneracies in these densities as $y_i = Z\alpha_{i,0}$, however we will ignore this in our notation for densities.

[3]Other cases where Z is of rank d can be dealt with by simply considering observations of the form $Z^{-1}y_i = \alpha_i$.

3 Observed diffusions

3.1 The big picture

The observed case of the diffusion (1) arises when $Z = I_d$, so that the sequence of discrete time observations form a Markov process with $y_i = \alpha_{(i-1)M}$. The general sampling method for Step 1 of the approach described in Section 2.2 becomes quite simple due to the Markovian nature of y_i. In particular the problem of sampling from $f(\alpha^+|y,\theta)$ becomes the task of drawing from

$$\prod_{i=1}^{T-1} f(\alpha_i^+|y_i, y_{i+1}, \theta), \quad \text{where} \quad \alpha_i^+ = (\alpha_{(i-1)M+1}, \alpha_{(i-1)M+2}, ..., \alpha_{iM-1}, \alpha_{iM})'. \tag{6}$$

Because in the fully observed case $\alpha_{iM} = y_{i+1}$, it is only necessary to sample a subset of α_i^+, specifically

$$\alpha_i^* = (\alpha_{(i-1)M+1}, \alpha_{(i-1)M+2}, ..., \alpha_{iM-1})'.$$

From (6) we see that the α_i^* are conditionally independent and the problem of interest therefore reduces to one of sampling the distributions

$$\alpha_i^*|y_i, y_{i+1}, \quad i = 1, 2, ..., T-1,$$

seperately for α_i^*. This means that Step 1 of Section 2.2 becomes:

1'. Update α_i^* from

$$f(\alpha_i^*|y_i, y_{i+1}, \theta), \tag{7}$$

for $i = 1, 2..., T-1$.

This reduction in the dimension is clearly very helpful as the task of sampling from the enormously high dimensional distribution $\alpha^+|y,\theta$ is now broken down to a series of $T-1$ independent sampling steps, each of dimension $M \times d$. Furthermore, the target distribution $\alpha_i^*|y_i, y_{i+1}, \theta$ is relatively easy to sample. Here we discuss a method based on the importance sampling distribution that appears in Durham and Gallant (2002) (henceforth DG). We utilize a version of that importance function to form a proposal density for our Metropolis-Hastings algorithm. We will see that this leads to an effective sampling procedure for these distributions.

3.2 Algorithm for observed diffusions

3.2.1 The core ideas

We now consider the sampling of the distribution $\alpha_i^*|y_i, y_{i+1}, \theta$ and since all such distributions have the same structure we simply set $i = 1$ and also suppress the parameters θ for notational

convenience. The two observed observations are then now $\alpha_0 = y_1$ and $\alpha_M = y_2$ and the augmented points are $\alpha^{*\prime} = (\alpha_1', ..., \alpha_{M-1}')'$. Thus, in this notation, the question of interest is the sampling of the distribution

$$\alpha^* | \alpha_0, \alpha_M.$$

From (3), we have that α_j follows the multivariate Markov process

$$\alpha_{j+1} | \alpha_j \sim N_d(\alpha_j + \mu(\alpha_j)\delta, \Sigma(\alpha_j)\delta), \quad j = 0, \ldots, M-1. \tag{8}$$

which implies by simple calculation that $\alpha^* | \alpha_0, \alpha_M$ has the density

$$f(\alpha_1, ..., \alpha_{M-1} | \alpha_0, \alpha_M) = \left\{ \prod_{j=1}^{M} f(\alpha_j | \alpha_{j-1}) \right\} / f(\alpha_M | \alpha_0) \tag{9}$$

$$= \prod_{j=1}^{M-1} f(\alpha_j | \alpha_{j-1}, \alpha_M), \tag{10}$$

where $f(\alpha_M | \alpha_0)$ is the unknown normalizing constant. This is the (target) density that we need to sample. It is easy to see that the (smoothed) densities in the product of (10) cannot be calculated nor can they be directly simulated due to the non-linearity in (8).

In their work, DG develop an importance function to integrate out $\alpha_1, ..., \alpha_{M-1}$, leading to an estimate of the likelihood contribution. Although our objectives are quite different, we can utilize this importance function as a proposal density for the Metropolis-Hastings algorithm. Later, in Section 4, we provide a generalisation of this proposal density that enables consideration of a broader class of models.

To describe the DG approach, consider the question of finding a close approximation to the target density

$$f(\alpha_j | \alpha_{j-1}, \alpha_M) \propto f(\alpha_M | \alpha_j) f(\alpha_j | \alpha_{j-1}),$$

where the density $f(\alpha_j | \alpha_{j-1})$ arises from (8). Now set

$$q(\alpha_j | \alpha_{j-1}, \alpha_M) \propto \widehat{f}(\alpha_M | \alpha_j) f(\alpha_j | \alpha_{j-1})$$

where

$$\widehat{f}(\alpha_M | \alpha_j) = N_d(\alpha_M | \alpha_j + \mu(\alpha_{j-1})\delta^*, \Sigma(\alpha_{j-1})\delta^*) \quad \text{where} \quad \delta^* = (M-j)\,\delta. \tag{11}$$

Note that $\widehat{f}(\alpha_M | \alpha_j)$ is the standard Euler approximation between α_M and α_j, but with α_{j-1} appearing in the drift and volatility functions. This form of $\widehat{f}(\alpha_M | \alpha_j)$, as a function of α_j, produces a closed form expression for the density $q(\alpha_j | \alpha_{j-1}, \alpha_M)$. DG utilize the resulting density as an importance function in their importance sampling procedure. The exact form of this density follows from this simple Proposition.

Proposition 1 *For some fixed integer $j \geq 0$, we assume, that for arbitrary integers $N > k > j \geq 0$ that*

$$\alpha_N|\alpha_k \sim N_d(\alpha_k + \mu(\alpha_{j-1})\,(N-k)\,\delta, \Sigma(\alpha_{j-1})\,(N-k)\,\delta),$$

then, writing

$$m_j = \frac{(k-j)}{N-j}\,(\alpha_N - \alpha_j) \quad \textit{and} \quad v_j = \delta\frac{(k-j)\,(N-k)}{(N-j)},$$

we have that

$$\alpha_k|\alpha_j, \alpha_N \sim N_d\left(\alpha_j + m_j, v_j\Sigma(\alpha_{j-1})\right). \tag{12}$$

□

Proof. Given in Appendix 8.1.

Note that, under this measure, the mean of $\alpha_k|\alpha_j, \alpha_N$ does not depend upon the drift μ. Instead α_k is simply interpolated between α_j and α_N.

The exact form of the DG importance function then follows trivially.

Definition 2 *DG Importance Function*

$$q(\alpha_1, ..., \alpha_{M-1}|\alpha_0, \alpha_M) = \prod_{j=1}^{M-1} q(\alpha_j|\alpha_{j-1}, \alpha_M), \tag{13}$$

where $q(\cdot|\alpha_{j-1}, \alpha_M)$ is a Gaussian density, $N_d(\alpha_{j-1} + m_j, v_j\Sigma(\alpha_{j-1}))$, with

$$m_j = \frac{\alpha_M - \alpha_{j-1}}{(M-j+1)}, \quad v_j = \delta\frac{(M-j)}{(M-j+1)}. \tag{14}$$

□

We have written the DG importance function in an equilibrium correction form (see, for example, Hendry (1995)). If the α_{j-1} is below (above) α_M, then the mean of the importance function is shifted upwards (downwards). As j increases this correction strengthens.

For our problem, we utilize this importance function as a proposal density for a Metropolis step. In line with the suggestion in Chib and Greenberg (1995), the distribution of the proposal density is taken to be multivariate-t, $\widetilde{q}(\alpha_j|\alpha_{j-1}, \alpha_M)$ say, which has the mean and variance given by (14). If we let $\alpha^{*(j)}$ denote the current value of α^* at the end of the jth MCMC cycle, then the Metropolis step is implemented as follows.

Algorithm 1: *Metropolis method: observed diffusions*

1. *Sample* $z = (z_1, ..., z_{M-1})' \sim \widetilde{q}(z|\alpha_0, \alpha_M) = \prod_{j=1}^{M-1} \widetilde{q}(z_j|z_{j-1}, \alpha_M)$.

2. *Evaluate*

$$p(z, \alpha^{*(j)}) = \min\left\{1, \frac{f(z|\alpha_0, \alpha_M)}{\widetilde{q}(z|\alpha_0, \alpha_M)} \frac{\widetilde{q}(\alpha^{*(j)}|\alpha_0, \alpha_M)}{f(\alpha^{*(j)}|\alpha_0, \alpha_M)}\right\}. \tag{15}$$

3. *With probability* $p(z, \alpha^{*(j)})$, *set* $\alpha^{*(j+1)} = z$; *otherwise set* $\alpha^{*(j+1)} = \alpha^{*(j)}$.

We can expect the efficiency of this method to be high because $\widetilde{q}(\alpha_j|\alpha_{j-1}, \alpha_M)$ is likely to be a good approximation to $f(\alpha_j|\alpha_{j-1}, \alpha_M)$. In addition, the parameters of the proposal density can be calculated quickly, more quickly than those in Elerian, Chib, and Shephard (2001). It may be seen that the the conditional mean and variance given by (14) are based on a linear interpolation and may therefore be thought of as a bridge between the two observations. Intuitively, we might expect this linear proposal to work well when the observations α_0, α_M are closely spaced in time, as the true conditional mean may be close to linear. When the observations are far apart in time (or equivalently, when there is low persistence) a proposal density based on the linear approximation may be less adequate.

3.2.2 Blocking

One of the advantages of Bayesian inference via MCMC over importance sampling, is that it allows one to tackle higher dimensional problems since the sampling can be done (in sequence) on small blocks of parameters conditioned on the rest. For example, instead of sampling $\alpha_1, ..., \alpha_{M-1}$ from $f(\alpha_1, ..., \alpha_{M-1}|\alpha_0, \alpha_M)$ in one block we may sample it in two.

1". At each sweep select an integer k at random such that $1 < k < M$

(a) Sample from $f(\alpha_1, ..., \alpha_{k-1}|\alpha_0, \alpha_k)$,

(b) Sample from $f(\alpha_k, ..., \alpha_{M-1}|\alpha_{k-1}, \alpha_M)$.

In order to implement this strategy we can apply the Metropolis algorithm of the previous subsection to $f(\alpha_1, ..., \alpha_{k-1}|\alpha_0, \alpha_k)$ and then seperately on $f(\alpha_k, ..., \alpha_{M-1}|\alpha_{k-1}, \alpha_M)$. The same approach can obviously be extended to more than two blocks. Blocking in this way can be effective when tailoring by linear approximation is less precise for the larger block and more precise when the blocks are smaller. There can be a problem, however, if the block size is taken to be too small because that leads to an increase in the dimension of the state space of the underlying Markov chain and can increase the serial correlation of the simulated output. It is possible to show that the dependence in the simulated output increases linearly with the number of blocks in Gaussian Ornstein-Uhlenbeck processes (see Elerian (1999) and Roberts and Stramer (2001)). Similar analytic calculations are not possible in more interesting diffusions,

however the same behavior can be observed empirically. Hence, it is important to consider both the block size and the adequacy of the proposal while tuning the MCMC algorithm. We will revisit this question in Section 4 when we discuss models for partially observed diffusions.

3.3 Analysis as $M \to \infty$

Before we progress to the examples, we show that the proposal density based on the DG importance function has potentially very good properties even as $M \to \infty$. The simulation scheme presented above can be productively thought of as a continuous time pure jump process $\alpha_M(t)$ on the interval $t \in [\tau_i, \tau_i + \Delta]$

$$\alpha_M(t) = \alpha(\tau_i) + \sum_{j=1}^{\lfloor (t-\tau_i)M/\Delta \rfloor} \left(m_j + v_j^{1/2} \Sigma^{1/2} \left[\alpha_M \left\{ \tau_i + \delta (j-1) \right\} \right] \varepsilon_j \right),$$

where ε_j is a vector of independent standard normals. Here the process jumps every Δ/M periods of continuous time. Then as $M \to \infty$ the law of this process converges to the law of the diffusion $z(t)$ which satisfies the SDE

$$\mathrm{d}z(t) = \{\tau_{i+1} - t\}^{-1} \{\alpha(\tau_{i+1}) - z(t)\} \, \mathrm{d}t + \Theta \{z(t)\} \, \mathrm{d}b(t), \tag{16}$$

where b is a vector of standard independent Brownian motions which are independent of the w process given in (1). This process is initialised at

$$z(\tau_i) = \alpha(\tau_i) = \alpha_0$$

and has the property that

$$z(\tau_{i+1}) = \alpha(\tau_{i+1}) = \alpha_M$$

with probability one. Hence this diffusion is a type of vector bridge process. It is important to note that (16) has exactly the same volatility function as (1) and so this approximation only differs via the drift. In particular we are using

$$\{\tau_{i+1} - t\}^{-1} \{\alpha(\tau_{i+1}) - z(t)\} \quad \text{instead of} \quad \mu \{z(t)\}.$$

From a theoretical viewpoint, it is insightful to think of the case where we have infinite computer power and allow $M \to \infty$. Then instead of making proposals to move sequences of the process, the algorithm will move sections of the entire continuous path of the process.

1. Propose the new path z from (16) to replace the existing path $\alpha^{*(j)}$. We write Q as the measure associated with this proposal, and G for the measure under (1).

2. Evaluate

$$p(z, \alpha^{*(j)}) = \min\left\{1, \frac{\mathrm{d}G(z)}{\mathrm{d}Q(z)} \frac{\mathrm{d}Q(\alpha^{*(j)})}{\mathrm{d}G(\alpha^{*(j)})}\right\}.$$

3. With probability $p(z, \alpha^{*(j)})$ we accept the path z, otherwise we reject it and continue to use $\alpha^{*(j)}$. Goto 1.

Here $\mathrm{d}G(z)/\mathrm{d}Q(z)$ is the Radon-Nikodym derivative, or likelihood ratio process, of moving from the approximating process Q to the true process G. As the volatility functions in (1) and (16) are the same, $\alpha(z, \alpha^{*(j)})$ is always properly defined[4]. This would not be true if the volatility functions differed. Therefore, we can expect to make proposals which are accepted even in this infinite dimensional case, which means that in practice it is possible to let M be large without causing the algorithm to get sticky. Of course, the practical performance of the procedure (in terms of the mixing of the generated output) is determined solely by the difference between the true and approximating drift functions. When the latter difference is small, the procedure works well, otherwise it works poorly.

3.4 Example 1: Cox-Ingersoll-Ross model

We now consider the Cox-Ingersoll-Ross (CIR) model given by the SDE

$$\mathrm{d}x = k(\mu - x)\mathrm{d}t + \sigma\sqrt{x}\mathrm{d}w(t).$$

Marginally we have $x \sim \mathrm{Ga}\left(\frac{2k\mu}{\sigma^2}, \frac{2k}{\sigma^2}\right)$. For this model, the transition density $g(x(\tau_{i+1})|x(\tau_i))$, and consequently the likelihood, can be evaluated exactly. We shall work with the process transformed to the real line by taking $\alpha = \log(x)$. By Ito's lemma we obtain

$$\mathrm{d}\alpha(t) = \left\{k(\mu - e^{\alpha(t)})e^{-\alpha(t)} - \frac{1}{2}\sigma^2 e^{-\alpha(t)}\right\}\mathrm{d}t + \sigma e^{-\alpha(t)/2}\mathrm{d}w(t). \tag{17}$$

Before turning to the results of the MCMC sampler, we mention that we will evaluate the performance of our sampling scheme in terms of the inefficiency factor[5], or the integrated autocorrelation time, of each posterior estimate. This measure is defined as

$$INF = 1 + 2\sum_{i=1}^{\infty}\rho(i),$$

[4]An implication of this analysis is that if we follow DG and use z as an importance sampler to estimate the density $g(y_{i+1}|y_i, \theta)$ in fully observed diffusions, then in the limit as $M \to \infty$ then the importance sampler weights will be $\mathrm{d}G(z)/\mathrm{d}Q(z)$, which should be, in theory, well behaved. This is very unusual in importance sampling, for as the dimension of the integration goes to infinity we usually expect the variance of the importance sampler to increase exponentially with the dimension while it often even goes to infinity, which can mean that its rate of convergence will become poor (see, for example, Koopman and Shephard (2002)).

[5]See also Geweke (1989) who prefers to report the inverse of this number

where $\rho(\cdot)$ is the autocorrelation function of the sampled draws. By way of interpretation, to make the variance of the posterior estimate the same as that from independent draws, the MCMC sampler must be run for INF times as many iterations, beyond the transient or burn-in phase.

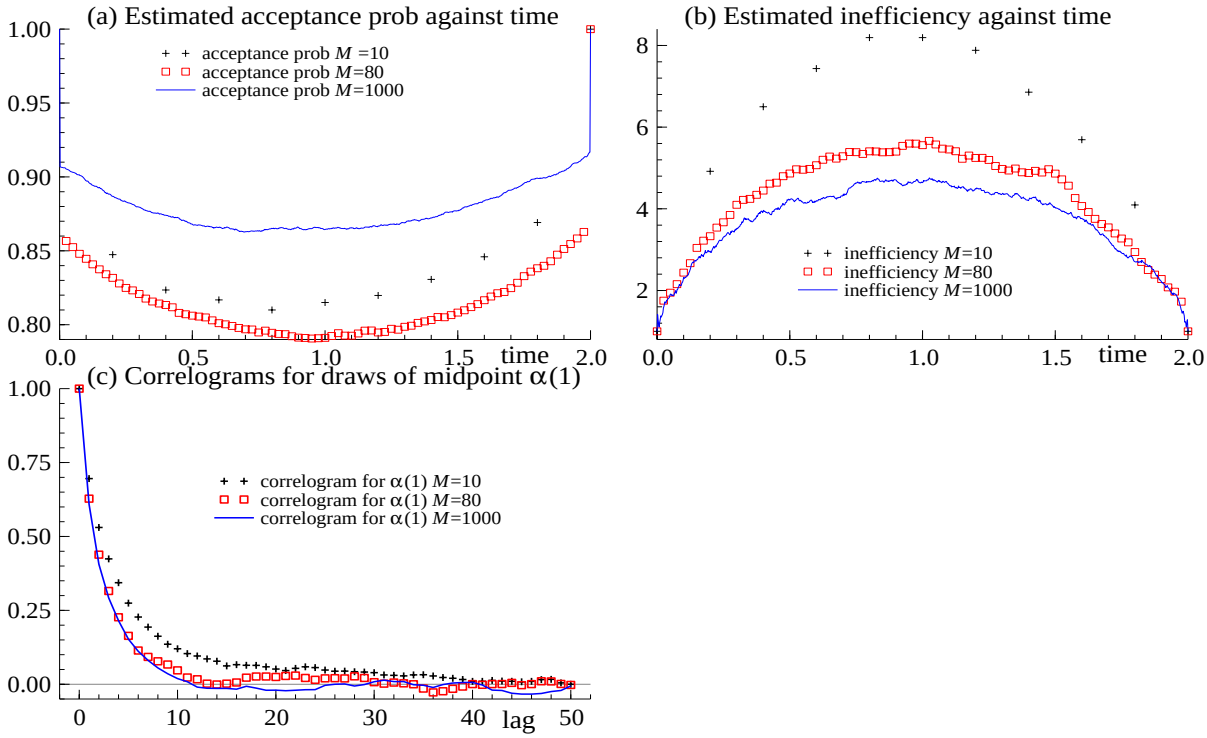

Figure 1: *Analysis of CIR model with* 3 *blocks,* $M = 10$, 80 and 1000. *(a) estimated average acceptance probability, (b) estimated inefficiency both against t, time. (c): correlogram of the sampled midpoint* $\alpha(1)$.

We first examine inference for the CIR process tied down by two observations with fixed parameters. We let $\Delta = 2$ and follow Durham and Gallant (2002) and Ait-Sahalia (2002, Table 3) in fixing $k = 0.5$, $\mu = 0.06$ and $\sigma = 0.15$. These parameters are calibrated to the US monthly treasury bill rate. Under these choices, $x = e^{\alpha}$ has marginal mean and variance of 0.06 and 0.00135 respectively. We also let the two observations be $\alpha(0) = \alpha_0 = \log(0.05)$ and $\alpha(2) = \alpha_M = \log(0.25)$. It is clearly unlikely that we would have a move in interest rates from 5% to 25% in 2 months. Nonetheless, this setting offers an interesting test of the viability of our method.

We implement our M-H algorithm with 3 blocks and with M taking the values 10, 80 and 1000. The choice $M = 1000$ is particularly interesting because it allows us to see if augmentation of high level is practically possible. In each case, the degrees of the freedom of the multivariate-t

proposal density is taken to be 50 and we base our results on 10000 MCMC draws collected after a burn-in of a 100 cycles. We mention that the computation time is basically propotional to M. We display our results in Figure 1. In this figure, in panel (a) we plot the acceptance rate from the M-H step for each augmented data point in the interval $(0, 2)$. The acceptance rate is usually over 80% and is not affected by the degree of augmentation. In panel (b) of the same plot, we record the inefficiency factors for each augmented point, computed using 50 lags. In panel (c) we report the autocorrelation function for the latent data in the mid-point of our interval. These inefficiency factors show that the method performs extremely well with the maximum inefficiency being less than 8. Interestingly, the inefficiency factors fall as M is increased. Since the number of blocks remains fixed at 3, these results show that the M-H step for the latent data is not particularly affected by the number of augmented points that are sampled simultaneously. This example provides a potent illustration of the fact that it is possible to use a fine discretization of the time interval without affecting the performance of the sampling scheme, consistent with the theoretical arguments advanced in the previous sub-section.

4 Partially observed diffusions

4.1 The big picture

We now turn to an analysis of vector SDE that are only partially observed. In the notation introduced in Section 1, the observed data is given by

$$y(\tau_i) = Z\alpha(\tau_i), \quad i = 1, ..., T,$$

where Z is a non-random $k \times d$ selection matrix. The key difference in approach for dealing with this (non-Markov) situation is that now it is not particularly helpful to consider blocks defined in terms of the latent α_i^+ between y_i and y_{i+1}. Instead it is necessary to work with the entire α^+ in terms of the natural complete ordering

$$\alpha^+ = (\alpha_0, \alpha_1, ..., \alpha_{M(T-1)})', \tag{18}$$

so that

$$y_i = Z\alpha_{M(i-1)}, \quad i = 1, \ldots, T-1.$$

To sample this very high-dimensional latent vector from its distribution conditioned on the observed data and the parameters, we subdivide α^+ randomly into B blocks

$$\alpha^+ = (\alpha^{+(1)}, \alpha^{+(2)}, ..., \alpha^{+(B)})'.$$

Corresponding to these blocks we write

$$y = (y^{+(1)}, y^{+(2)}, ..., y^{+(B)})',$$

the real observations which fall within these blocks. Each block $\alpha^{+(l)}$ may contain more or less than M vectors of the diffusion and so $y^{+(l)}$ could include single vectors of observations, multiple observations or could be the null set. Finally, it will be convenient to write the elements of the l-th block as

$$\alpha^{+(l)} = (\alpha_1^{+(l)}, \alpha_2^{+(l)}, ..., \alpha_{N_l}^{+(l)})',$$

where N_l is the number of elements in this block.

For each block we update from its full conditional density. Thus the sampling step becomes

1‴. (a) Subdivide α^+ randomly into B blocks, with the l-th block being $\alpha^{+(l)}$.

(b) Update from

$$f(\alpha^{+(l)}|\alpha_{N_{l-1}}^{+(l-1)}, \alpha_1^{+(l+1)}, y^{(l)}), \tag{19}$$

for $l = 1, 2, ..., B$.

The fact that we condition only on $\alpha_{N_{l-1}}^{+(l-1)}$, $\alpha_1^{+(l+1)}$ and $y^{(l)}$ is due to the Markovian nature of the diffusion.

4.2 Algorithm for partially observed diffusions

The only task that remains is to design simulators for $\alpha^{+(l)}|\alpha_{N_{j-1}}^{+(l-1)}, \alpha_1^{+(l+1)}, y^{(l)}$. In order to focus on the main ideas we will surpress the superscript l and write $\alpha^{*\prime} = (\alpha_1', ..., \alpha_{N-1}')$ and think about sampling from

$$\alpha^*|\alpha_0, \alpha_N, y^+.$$

Clearly if y^+ is the null set, then we can update samples from this distribution by using the DG sampler. Hence that case is of no new interest. Before considering the case of multiple observations it is helpful to think about y^+ containing only a single observation within this block, $\widetilde{y}_k$. Recall this notation means this observation corresponds to

$$\widetilde{y}_k = Z\alpha_k.$$

In this case we have to sample from the full conditional density

$$f(\alpha^*|\alpha_0, \alpha_N, \widetilde{y}_k) = I(\widetilde{y}_k = Z\alpha_k) \left\{ \prod_{j=1}^{N} f(\alpha_j|\alpha_{j-1}) \right\} / f(\widetilde{y}_k, \alpha_N|\alpha_0). \tag{20}$$

As in Section 3, the normalising term $f(\alpha_N, \widetilde{y}_k|\alpha_0)$ is unknown to us. However we can evaluate the numerator of (20), allowing the use of the Metropolis method.

In theory this can be rewritten as

$$\begin{aligned} f(\alpha^*|\alpha_0, \alpha_N, \widetilde{y}_k) &= \prod_{j=1}^{N-1} f(\alpha_j|\alpha_{j-1}, \widetilde{y}_k, \alpha_N) \quad (21) \\ &= \prod_{j=1}^{k} f(\alpha_j|\alpha_{j-1}, \widetilde{y}_k, \alpha_N) \prod_{j=k+1}^{N-1} f(\alpha_j|\alpha_{j-1}, \alpha_N). \quad (22) \end{aligned}$$

The $\alpha_j|\alpha_{j-1}, \alpha_N$ terms in (22) are essentially the expressions which appeared in Section 3.2, so we know how to update that block. Hence the only remaining issue is how to sample from $f(\alpha_j|\alpha_{j-1}, \widetilde{y}_k, \alpha_N)$.

In Section 3.2, we noted that the derivation of the approach of DG depends upon the approximation, $\widehat{p}(\alpha_M|\alpha_j)$, made in (11). We parallel this approach and work under the measures, for $N > k > j$,

$$\begin{aligned} \widehat{f}(\alpha_k|\alpha_j) &= \mathrm{N}(\alpha_k|\alpha_j + \mu(\alpha_{j-1})\delta^{(k-j)}, \Sigma(\alpha_{j-1})\delta^{(k-j)}), \\ \widehat{f}(\alpha_N|\alpha_k) &= \mathrm{N}(\alpha_N|\alpha_k + \mu(\alpha_{j-1})\delta^{(N-j)}, \Sigma(\alpha_{j-1})\delta^{(N-j)}), \quad (23) \\ \widetilde{y}_k &= Z\alpha_k. \end{aligned}$$

Then,

$$\alpha_j|\alpha_{j-1}, \widetilde{y}_k, \alpha_N \sim N_d\left(m_j^\dagger, V_j^\dagger\right),$$

where the mean and covariance are given in the following Proposition.

Proposition 3 *Writing* $\Sigma_{j-1} = \Sigma(\alpha_{j-1})$ *and*

$$m_j = \frac{\alpha_N - \alpha_{j-1}}{(N-j+1)}, \quad v_j = \delta\frac{(N-j)}{(N-j+1)},$$

and

$$c_j = 1 - \frac{(k-j)}{N-j},\ m_j^* = \frac{(k-j)}{N-j}\alpha_N, \quad v_j^* = \delta\frac{(k-j)\,(N-k)}{(N-j)}.$$

Then $\alpha_j|\alpha_{j-1}, \widetilde{y}_k, \alpha_N \sim N_d\left(m_j^\dagger, V_j^\dagger\right)$ *where*

$$\left(V_j^\dagger\right)^{-1} = v_j^{-1}\Sigma_{j-1}^{-1} + \frac{c_j^2}{v_j^*}Z'\left(Z\Sigma_{j-1}Z'\right)^{-1}Z$$

and

$$\left(V_j^\dagger\right)^{-1} m_j^\dagger = v_j^{-1}\Sigma_{j-1}^{-1}\left(\alpha_{j-1} + m_j\right) + \frac{c_j}{v_j^*}Z'\left(Z\Sigma_{j-1}Z'\right)^{-1}\left(\widetilde{y}_k - Zm_j^*\right).$$

□

Proof. Proposition 1 implies that

$$\alpha_j|\alpha_{j-1},\alpha_N \sim N\left(\alpha_{j-1}+m_j, v_j\Sigma_{j-1}\right),$$

and

$$\alpha_k|\alpha_j,\alpha_N \sim N\left(c_j\alpha_j+m_j^*, v_j^*\Sigma_{j-1}\right), \quad c_j=1-\frac{(k-j)}{N-j},\ m_j^*=\frac{(k-j)}{N-j}\alpha_N, \quad v_j^*=\delta\frac{(k-j)(N-k)}{(N-j)}.$$

Thus

$$\widetilde{y}_k|\alpha_j,\alpha_N \sim N\left(c_jZ\alpha_j+Zm_j^*, v_j^*Z\Sigma_{j-1}Z'\right),$$

Then, from standard Bayesian regression theory[6]

$$\left(V_j^\dagger\right)^{-1}=v_j^{-1}\Sigma_{j-1}^{-1}+\frac{c_j^2}{v_j^*}Z'\left(Z\Sigma_{j-1}Z'\right)^{-1}Z$$

and

$$\left(V_j^\dagger\right)^{-1}m_j^\dagger=v_j^{-1}\Sigma_{j-1}^{-1}\left(\alpha_{j-1}+m_j\right)+\frac{c_j}{v_j^*}Z'\left(Z\Sigma_{j-1}Z'\right)^{-1}\left(\widetilde{y}_k-Zm_j^*\right).$$

□

The proposal on the entire block is thus

$$q(\alpha_1,...,\alpha_{N-1}|\alpha_0,\widetilde{y}_k,\alpha_N)=\prod_{j=1}^{k}q(\alpha_j|\alpha_{j-1},\widetilde{y}_k,\alpha_N)\prod_{j=k+1}^{N-1}q(\alpha_j|\alpha_{j-1},\alpha_N). \tag{24}$$

We can now propose from $q(\alpha_1,...,\alpha_{N-1}|\alpha_0,\widetilde{y}_k,\alpha_N)$ and compare with the true density given by (20) in a Metropolis algorithm. This is a direct generalisation of the algorithm in Section 3.2. We can see that if there are no observations within the block then the true density given by (20) is exactly that of Section 3.2 and we have exactly the same proposal as in Section 3.2. Similarly, we obtain exactly the same setup when if observe α_k exactly, so that $Z=I_d$. Section 3.2, is when the we exactly observe α_k, so that $Z=I_d$. The computational complexity of the method, which involves simulating from (24) and evaluating the Metropolis acceptance probability, is of order $N-1$, as in Section 3.2.

We shall use this proposal for our general method, involving an arbitrary number of measurements in the block under consideration. The general method will be apparent by examining

[6]Recall that it is well known that if $\beta\sim N(b_0,S_0)$ and $y|\beta\sim N(X\beta,\Omega)$, then $\beta|y\sim N(b_p,S_p)$ where

$$\begin{aligned} S_p^{-1} &= S_0^{-1}+X'\Omega^{-1}X, \\ S_p^{-1}b_p &= S_0^{-1}b_0+X'\Omega^{-1}y. \end{aligned}$$

the case of two observations within the block, $\widetilde{y}_k$ and $\widetilde{y}_l$, where $0 < k < l < N$. In this case the full conditional density we wish to draw from becomes

$$f(\alpha_1, ..., \alpha_{N-1}|\alpha_0, \widetilde{y}_k, \widetilde{y}_l, \alpha_N) = \left\{ f(\widetilde{y}_k|\alpha_k) f(\widetilde{y}_l|\alpha_l) \prod_{j=1}^{N} f(\alpha_j|\alpha_{j-1}) \right\} / f(\widetilde{y}_k, \widetilde{y}_l, \alpha_N|\alpha_0)$$

$$= \prod_{j=1}^{k} f(\alpha_j|\alpha_{j-1}, \widetilde{y}_k, \widetilde{y}_l, \alpha_N) \prod_{j=k+1}^{l} f(\alpha_j|\alpha_{j-1}, \widetilde{y}_l, \alpha_N) \prod_{j=k+1}^{N-1} f(\alpha_j|\alpha_{j-1}, \alpha_N). \tag{25}$$

We can see that at the beginning of the block, before the two observations occur, the conditional densities involve both future observations. This is reflected in the terms of the first product of (25). In general, for several observations in a block, the initial states will depend upon all future measurements in the block. This can raise computational difficulties for proposals as the computational complexity of our algorithm could be non-linear in N. However, we avoid this problem by only conditioning upon the next observation rather than the other future observations in the block. This is illustrated by examining our proposal distribution,

$$q(\alpha_1, ..., \alpha_{N-1}|\alpha_0, \widetilde{y}_k, \widetilde{y}_l, \alpha_N) = \prod_{j=1}^{k} q(\alpha_j|\alpha_{j-1}, \widetilde{y}_k, \alpha_N) \prod_{j=k+1}^{l} q(\alpha_j|\alpha_{j-1}, \widetilde{y}_l, \alpha_N) \prod_{j=k+1}^{N-1} q(\alpha_j|\alpha_{j-1}, \alpha_N). \tag{26}$$

where the terms $q(\alpha_j|\alpha_{j-1}, \widetilde{y}, \alpha_N)$ are given by (**??**) and $q(\alpha_j|\alpha_{j-1}, \alpha_N)$ by the DG proposal of Section 3.2. The crucial aspect is that we only condition upon the next observation, as seen by inspecting the first product term of (26). This ensures that our algorithm is fast and linear in N in terms of computational complexity. In practice, sampling from (26) is hardly any more complicated computationally than sampling from the proposal of Section 3.2. As in Section 3.2 we use a multivariate student t distribution $\widetilde{q}(\alpha_j|\alpha_{j-1}, \widetilde{y}, \alpha_M)$ which has mean and variance, m_j and V_j respectively.

We shall now outline the MCMC method for the general case where the collection of observations within the block under consideration are denoted by $\widetilde{y}^*$. We shall assume that the MCMC method has been running for j iterations so that our current value for the block is $\alpha^{*(j)}$. We use the Metropolis method, detailed below, to perform the MCMC scheme.

Algorithm 2: *Metropolis method: partially observed diffusions*

1. *Sample* $z = (z_1, ..., z_{N-1})' \sim \widetilde{q}(z|\alpha_0, \widetilde{y}^*, \alpha_N)$.

2. *Evaluate*

$$p(z, \alpha^{*(j)}) = \min \left\{ 1, \frac{f(z|\alpha_0, \widetilde{y}^*, \alpha_N)}{\widetilde{q}(z|\alpha_0, \widetilde{y}^*, \alpha_N)} \frac{\widetilde{q}(\alpha^{*(j)}|\alpha_0, \widetilde{y}^*, \alpha_N)}{f(\alpha^{*(j)}|\alpha_0, \widetilde{y}^*, \alpha_N)} \right\}. \tag{27}$$

3. *With probability* $p(z, \alpha^{*(j)})$, *set* $\alpha^{*(j+1)} = z$; *otherwise set* $\alpha^{*(j+1)} = \alpha^{*(j)}$.

In the following section we consider the application of this method to a model which is of state space form, see Harvey (1993), when the Euler discretisation is applied. This allows an assessment of the scheme as we know the true posterior density of the discretised states, via Kalman recursion methods, conditional upon the observations. This is a model in which the state innovations are correlated. We find that our simple method works extremely well for this example. In Section 4.4, we examine the stochastic volatility (SV) model, estimating both the parameters and the discretised states.

4.3 Example 1: Gaussian factor structure

4.3.1 The model

We start with a simple model where analytic calculations are possible. Consider a bivariate, linear process

$$y(t) = \sigma_1 b_1(t) + \alpha_2(t),$$

where α_2 is a Gaussian OU process given by the solution to

$$\mathrm{d}\alpha_2(t) = k(\mu - \alpha_2)\mathrm{d}t + \sigma_2 \mathrm{d}b_2(t), \quad \text{with} \quad k > 0.$$

We assume that b_1 and b_2 are independent standard Brownian motions. Hence if t is large the Brownian motion component will dominate the variation in $y(t)$, but over the short run α_2 can have a substantial impact.

This means we can put this model into our model structure by writing

$$y(\tau_i) = Z\alpha(\tau_i), \quad \text{where} \quad Z = (1\ 0)\,,$$

and

$$\begin{aligned} \mathrm{d}\alpha_1(t) &= \kappa(\mu - \alpha_2)\mathrm{d}t + \sigma_2 \mathrm{d}b_2(t) + \sigma_1 \mathrm{d}b_1(t) \\ \mathrm{d}\alpha_2(t) &= \kappa(\mu - \alpha_2)\mathrm{d}t + \sigma_2 \mathrm{d}b_2(t), \quad \text{with} \quad k > 0. \end{aligned}$$

Importantly the first component of the system $\alpha(t) = \{\alpha_1(t), \alpha_2(t)\}'$, is exactly but infrequently observed. The second component can only be infered through the observation of the first.

The model may be placed into companion form, of (1), as

$$\mathrm{d}\alpha(t) = \mu(\alpha)\mathrm{d}t + \Omega \mathrm{d}w, \tag{28}$$

where, writing ι as a vector of ones,

$$\mu(\alpha) = \kappa(\mu - \alpha_2)\iota, \quad \Sigma = \Omega\Omega' = \begin{pmatrix} \sigma_1^2 + \sigma_2^2 & \sigma_2^2 \\ \sigma_2^2 & \sigma_2^2 \end{pmatrix}.$$

The Euler approximation to (28) is

$$\alpha_{j+1} = \alpha_j + \kappa\delta(\mu - \alpha_{2j})\iota + \sqrt{\delta}\Omega v_j, \tag{29}$$

where v_j are independent bivariate standard Gaussian variates. Together with the observations on the Euler scale

$$\widetilde{y}_{Mi} = Z\alpha_{Mi}, \quad i = 1, ..., T,$$

the expression (29) forms a Gaussian, linear model allowing exact analysis of the Euler discretisation of the model for any specific value of δ. In particular we can directly calculate the posterior of the complete data $f(\alpha^+|y;\theta)$, where $\theta = (\kappa, \mu, \sigma_1^2, \sigma_2^2)'$. The Gaussian densities can be efficiently handled using the Kalman filter and associated smoother (e.g. Harvey (1989) and Durbin and Koopman (2002)).

This model is chosen to test our simulation methodology as we have the smoothing density of our states for the Euler approximation. We can therefore compare our output to the true smoothed density. Overall this model should provide an information about the robustness of our efficient and relatively simple algorithm.

4.3.2 Results

We take $T = 12$ and a unit frequency of observation, so we have observations at actual times $0, 1, 2, ..., 11$. We take $M = 4$ ($\delta = 0.25$) resulting in a total of 45 states $\alpha_0, ..., \alpha_{44}$. We take our parameters to be $(\kappa, \theta, \sigma_2^2) = (0.3, 0.5, 0.03)$ and $\sigma_1^2 = 0.03$. The MCMC sampler is run using just 2 blocks with the degrees of freedom in our multivariate t set to 30. The number of simulations for our MCMC sampler is $20,000$, while we discarded the first ??? in order to take into account of the effect of burnin.

The true posterior mean of the states in displayed, in Figure 2, with the 95% intervals obtained from our MCMC results. Recorded is the true value, computed using the moments of a normal distribution via the Kalman smoother, together with the 95% intervals obtained from our MCMC results. Figure 2(a) shows the results for the observed component α_1, while Figure 2(b) shows the corresponding results for α_2. We would expect the true smoothed values to lie inside the bounds, which is what we see.

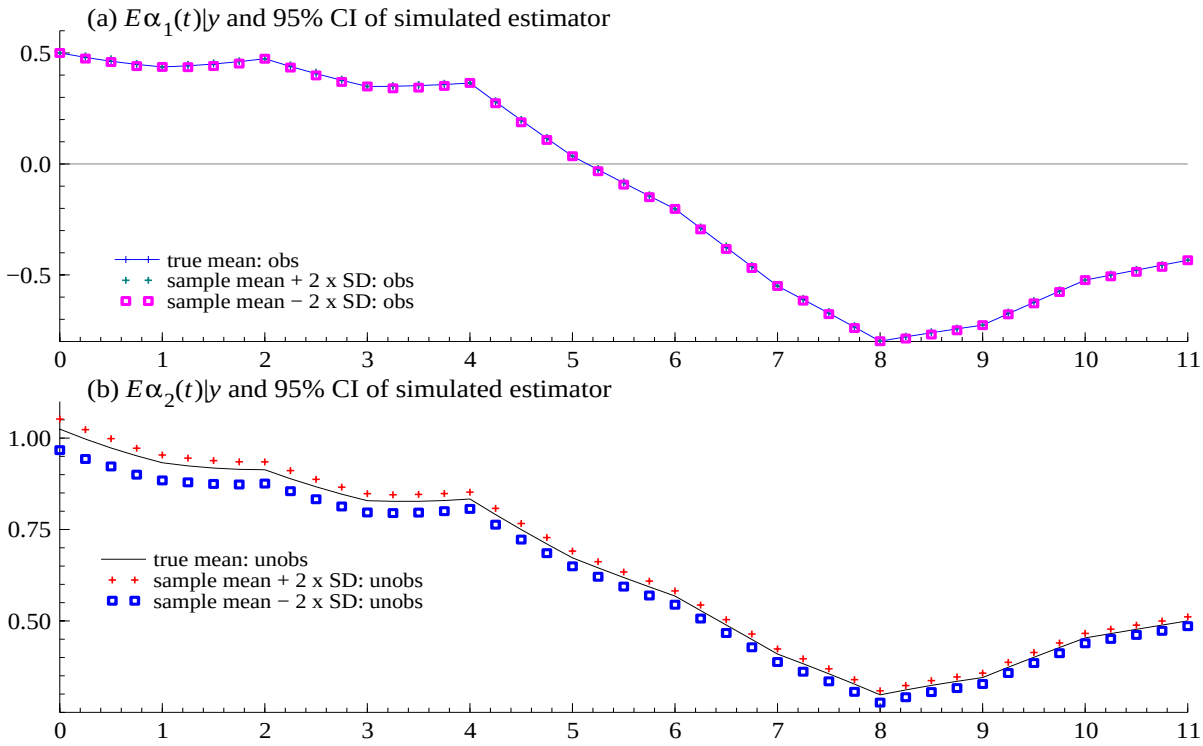

Figure 2: *OU plus Brownian motion example. We use a 2 block and* $\nu = 30$ *degree of freedom sampler. True and estimated means for posterior trajecties together with 95% confidence interval from MCMC results. (a) observed component* α_1*, (b) unobserved component* α_2*.*

Figure 3(a) shows the average acceptance rate of the MCMC sampler at different points in the sample. Figure 3(b) shows the MCMC trajectories and correlograms for two middle states $\alpha_{15} = (\alpha_{15,1}, \alpha_{15,2})'$ and $\alpha_{30} = (\alpha_{30,1}, \alpha_{30,2})'$, the first component of the state being the regularly observed part, the second the unobserved part. Clearly, the dependence is low as indicated by the trajectory plots and the correlograms. The inefficiencies, drawn in Figure 3(c), for all the states are very low. The inefficiencies, compared to a direct Monte Carlo sampler, over the states are a higher than in the case where we conditioned upon the end points, at around values of 4 for the observed part of the state and from 12 to 5 for the unobserved part of the state.

4.4 Example 2: Stochastic volatility example

Here we shall consider the standard stochastic volatility used for stock or exchange rate returns, see for instance Hull and White (1987), Ghysels, Harvey, and Renault (1996) and Shephard (1996). The continuous time setup has

$$y(t) = \alpha_1(t),$$

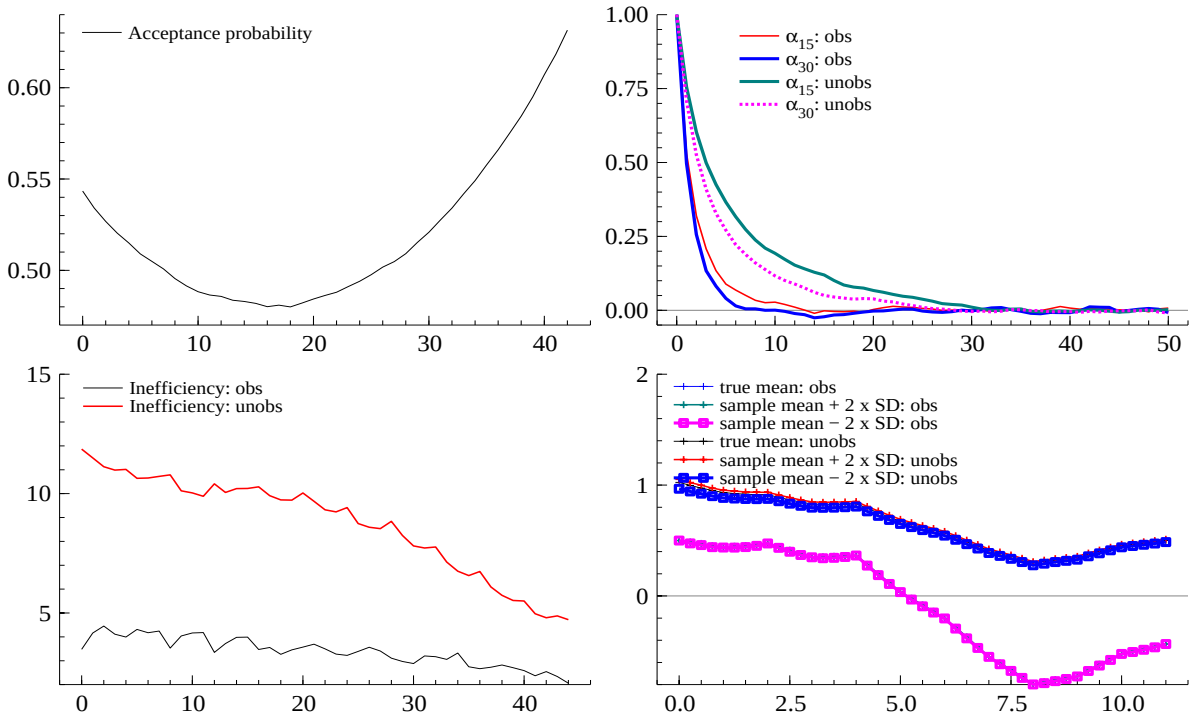

Figure 3: *Double OU example. (a) average acceptance probability against time. (b) Correlogram for observed and unobserved components of* α_{15} *and* α_{30}*. (c): Inefficiency for observed and unobserved components against index.*

at times $0 \leq \tau_1 < \tau_2 < ... < \tau_T$ with

$$\begin{pmatrix} \mathrm{d}\alpha_1(t) \\ \mathrm{d}\alpha_2(t) \end{pmatrix} = \begin{pmatrix} \mu + \beta \exp\{\alpha_2(t)\} \\ \kappa(\mu - \alpha_2) \end{pmatrix} \mathrm{d}t + \begin{pmatrix} \exp\{\alpha_2(t)/2\} & 0 \\ 0 & \omega \end{pmatrix} \begin{pmatrix} \mathrm{d}w_1(t) \\ \mathrm{d}w_2(t) \end{pmatrix},$$

where w_1 and w_2 are standard Brownian motions. Here

- α_1 represents the log-price process,
- α_2 the spot log-variance process,
- μ the drift, and β the risk premium,
- κ the persistence of the volatility process, while μ is the mean of the log–variance process and ω is the corresponding volatility of volatility,
- $\mathrm{Cov}(w_1(t), w_2(t)) = \rho t$, where ρ is viewed as a leverage parameter (e.g. Black (1976) and Nelson (1991)).

In typical applications both μ and β are hard to estimate and so in this experiment we will set them to be zero in this experiment.

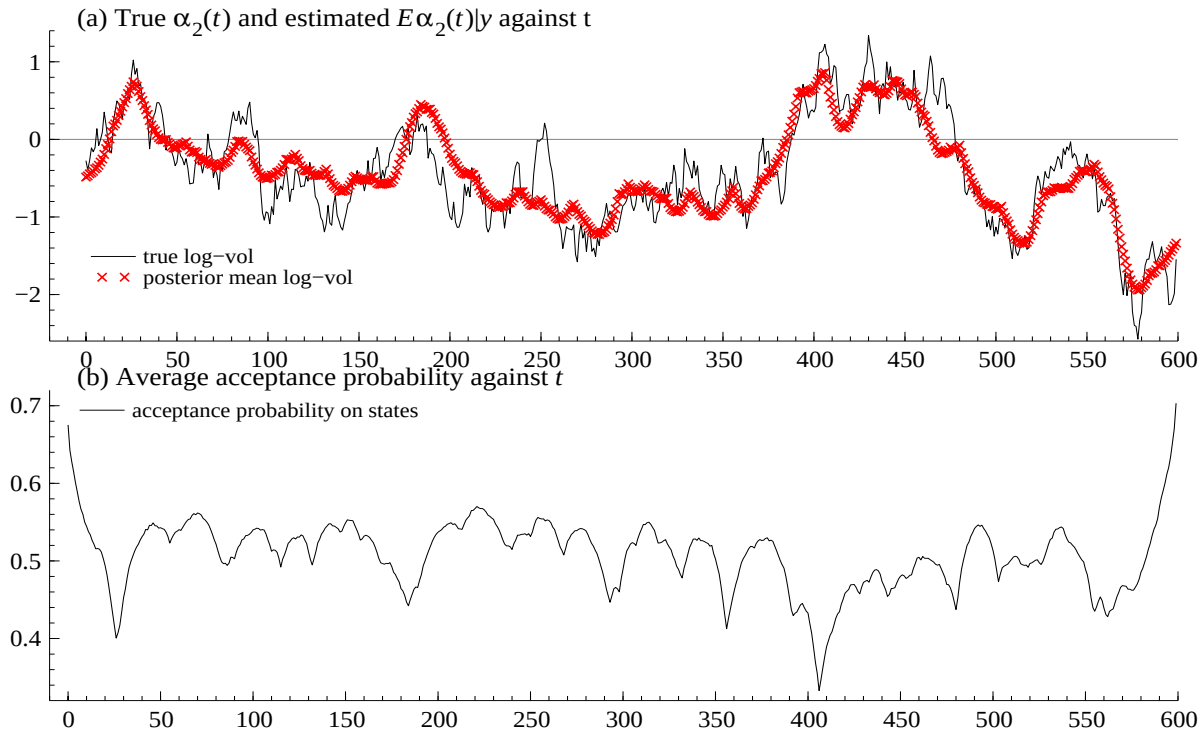

Figure 4: *Log-OU Volatility example. We use knots (15,5,10) and set* $\nu = 30$, $\delta = 1$ *and* $n = 600$. *(a) The true volatility path together with the posterior mean from the MCMC sumulations. (b)The Metropolis acceptance probabilities for all the states.*

Let us denote the parameters as $\theta = (\kappa, \mu, \sigma^2)'$. Representing in the usual manner, the complete data as α^+, we sample, using the approach of Section 4.2, from $f(\alpha^+|\theta; y)$ then from $f(\theta|\alpha^+)$. Sampling from $f(\theta|\alpha^+)$ is relatively straightforward as we essentially have a linear model, apart from the initial distribution term. The sampling scheme is given in the Appendix, section 8.2.

We shall illustrate the performance of our approach by simulating a time series of length $T = 600$, unit spacing in time between observations. We take $\kappa = 0.03$, $\mu = -0.6$ and $\omega = \sqrt{0.03}$. Initially we use $M = 1$. We vary the number of knots systematically as 15, 10 and 5 and set the degrees of freedom in our multivariate t to be $\nu = 30$. The results are given in Figures 4 and 5. From Figure 4(b), it is clear that the acceptance probabilities on the states (log-volatilities) vary between 0.4 and 0.6. The posterior mean for the log-volatility is clearly close to the true log-volatility. For the parameters, it is clear from Figures 4 and 5 that there is relatively little dependence across MCMC sweeps for μ and κ, whilst ω^2 exhibits more dependence.

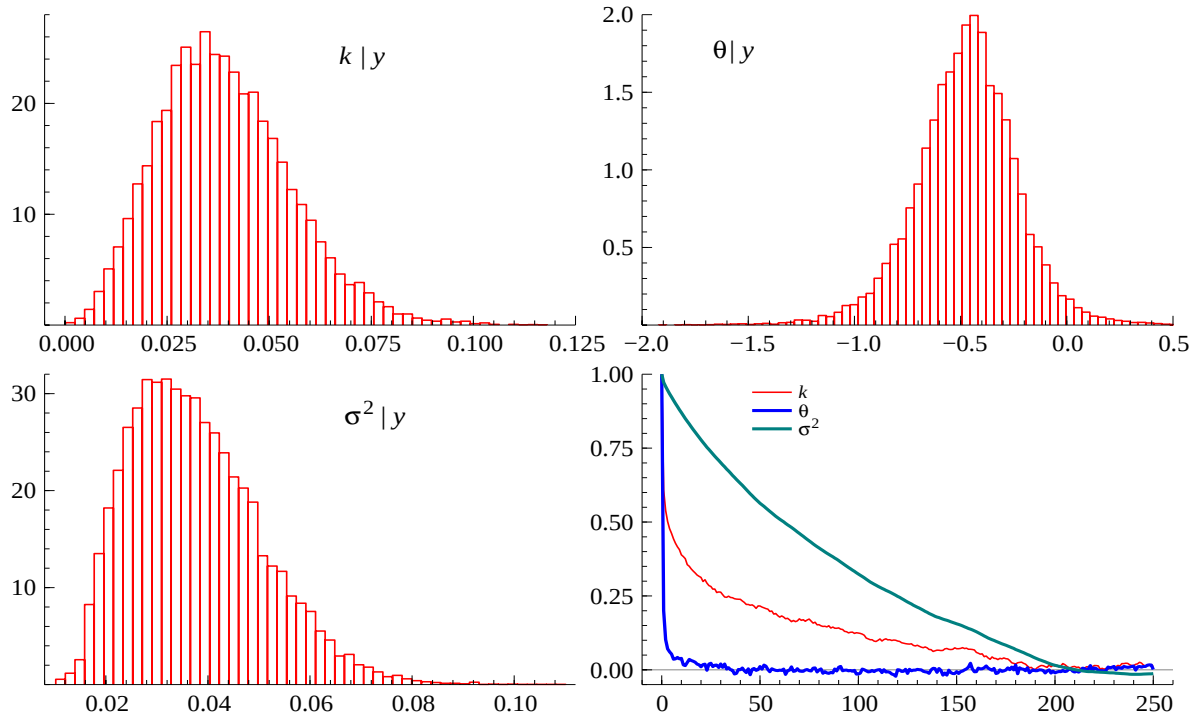

Figure 5: *Log-OU Volatility example with* $k = 0.03$, $\theta = -0.6$, $\sigma^2 = 0.03$. *We use knots (15,5,10) and set* $\nu = 30$, $\delta = 1$ *and* $n = 600$. *TOP: Histogram and correlogram for* k. *MIDDLE: Histogram and correlogram for* θ. *BOTTOM: Histogram and correlogram for* θ.

5 Major application

6 Discussion and extensions

6.1 Addition of jump processes

Jumps are often added to stochastic differential equations. In the SV literature they appear in the literature both in the price process and in the volatility process. A paper which looks at their empirical importance is Chernov, Gallant, Ghysels, and Tauchen (2002), who used EMM to carry out the estimation. Some recent stimulating papers on the MCMC analysis of continuous time models with jumps include Roberts, Papaspiliopoulos, and Dellaportas (2001), Fruhwirth-Schnatter and Soegner (2001) and Griffin and Steel (2002). These papers all focus on the Barndorff-Nielsen and Shephard (2001) OU volatility models, which are quite specialised.

For general diffusions we might add a vector jump by writing

$$\mathrm{d}\alpha(t) = \mu\{\alpha(t-);\theta\}\mathrm{d}t + \Theta\{\alpha(t-);\theta\}\mathrm{d}w(t) + \mathrm{d}z(t), \tag{30}$$

where $z(t)$ is a pure jump process. In principle this jump process can be quite intricate, allowing

infinite numbers of jumps in any finite amounts of time. Here we will follow the bulk of the literature in specialising the framework to the case where z is a compound Poisson process

$$z(t) = \sum_{j=1}^{N(t)} c_s,$$

where here c_j is assumed to be iid from some distribution $D(\alpha(t-);\theta)$. Here $N(t)$ is a counting process with, potentially, an intensity

$$\lambda(t) = \lambda\left\{\alpha(t-);\theta\right\}.$$

Recall $\lambda(t)$ has the interpretation as the derivative of the conditional expectation of the expectation of the counting process with respect to time.

From this structure it is clear that (30) is no longer a Markov chain, which is unfortunate from a number of viewpoints. It is more productive to wrap this model into a Markov companion form

$$\mathrm{d}\beta(t) = \left\{ \begin{array}{c} \mathrm{d}\alpha(t) \\ \mathrm{d}z(t) \end{array} \right\}.$$

We can use the augmentation idea directly on the augmented process $\beta(t)$. Now over small time intervals

$$z_{j+1} = z_j + n_j c_j, \quad n_j \sim Po\left\{\lambda\left\{\alpha_j;\theta\right\}\right\}, \quad c_j \sim D\left(\alpha_j;\theta\right).$$

Of course this process is not fully observed, but this yields no new issues as the previous Sections developed methods to handle this problem.

6.2 Discrete time analysis of partially observed systems

There is a considerable literature on partially observed systems in discrete time. Clearly this is highly related to the continuous case. When the system is Gaussian then in many situations the model can be placed into a Gaussian state space form and can be computationally efficiently handled using the Kalman filter and associated smoother. Discussions of this is given in, for example, Harvey (1989) and Durbin and Koopman (2002). Alternatively, unobserved components are often modelled as discrete state Markov chains, in which case the Baum, Petrie, Soules, and Weiss (1970) algorithm is well known in many areas of science. In economics this style of analysis is often associated with the work of Hamilton (1989).

Non-Gaussian discrete-time state space models have been studied by many authors. Early examples include Carlin, Polson, and Stoffer (1992), Carter and Kohn (1994) and Shephard (1994). Papers which focus on discrete time stochastic volatility models include Kim, Shephard,

and Chib (1998), Chib, Nardari, and Shephard (2002) and Eraker, Johannes, and Polson (2002), who also discuss some of the related literature on this subject.

6.3 Filtering

Economists are often interested in the sequence of items such as $\alpha_{i,0}|y_1,...,y_i,\theta$. Although MCMC can compute this, by simulating from $\alpha_{1,0},...,\alpha_{i,0}|y_1,...,y_i,\theta$ for each i (e.g. Shephard (1994)), it is unbearably slow as this is carried out from scratch for each i. Instead we focus on the recent literature on particle filters which focus directly on the sequential problem. Leading references in this regard, in the context of partially observed systems, are Gordon, Salmond, and Smith (1993), Kim, Shephard, and Chib (1998), Pitt and Shephard (1999), Jacod and Del Moral (2001), Jacod, Del Moral, and Protter (2001) and Johannes, Polson, and Stroud (2002). The last three references are particularly relevant as they develop these methods in the context of partially observed SDEs. A non-particle filter alternative to this approach has been developed by Gallant and Tauchen (1998), whose backprojection method approximates $\mathrm{E}\left(\alpha_{i,0}|y_1,...,y_i\right)$.

We can use the extended DG sampler to carry out particle filtering for partially observed SDEs. Here we briefly outline how this would work. Throughout we surpress dependence on θ.

The basic approach of particle filtering is to start with a sample from $\alpha_{i,0}|y_1,...,y_i$, which we write

$$\alpha_{i,0}^{(j)}, \quad j=1,2,...,J.$$

The approach of the particle filter is to use this sample to produce a sample of size J from $\alpha_{i+1,0}=\alpha_{i,M}|y_1,...,y_{i+1}$. Having reproduced the system one step forward we can repeat this through time, tackling the whole problem by a simple recursion. But how do we quickly simulate from $\alpha_{i,M}|y_1,...,y_{i+1}$ given samples the samples $\alpha_{i,0}^{(j)}$?

The problem is to sample from

$$f(\alpha_{i,1},...,\alpha_{i,M}|y_1,y_2,...,y_{i+1}) \simeq \sum_{j=1}^{J} f(\alpha_{i,1},...,\alpha_{i,M}|\alpha_{i,0}^{(j)},y_{i+1}),$$

where the approximation becomes ever more accurate as $J\to\infty$, for if we can do this we can discard all the elements of the sample except for the desired $\alpha_{i,M}|y_1,...,y_{i+1}$. Now

$$f(\alpha_{i,1},...,\alpha_{i,M}|\alpha_{i,0}^{(j)},y_{i+1}) \propto I(y_{i+1}=Z\alpha_{i,M}) \prod_{j=0}^{M-1} f(\alpha_{i,j+1}|\alpha_{i,j}),$$

so in principle we can sample from this object.

A simple approach to dealing with this problem is to make the following proposals, using L which is much larger than J.

1. Set $l = 1$.

2. Sample k randomly from $1, 2, ..., J$.

3. Use the extended DG sampler to make a proposal from

$$\alpha_{i,1}, ..., \alpha_{i,M} | \alpha_{i,0}^{(k)}, y_{i+1}.$$

   Call this sample $\alpha^{(l)}$.

4. Write the density of the DG sampler as g and compute the weight

$$w_l = \frac{f(\alpha^{(l)} | \alpha_{i,0}^{(k)}, y_{i+1})}{g(\alpha^{(l)} | \alpha_{i,0}^{(k)}, y_{i+1})}.$$

5. Let $l = l + 1$. Discard all the samples except $\alpha_{i,M}^{(l)}$, which is placed into memory. Goto 2 until $l = L$.

This produces a population of new particles

$$\alpha_{i,M}^{(1)}, ..., \alpha_{i,M}^{(L)}$$

and associated weights

$$w_1, ..., w_L.$$

We resample the new particles, using probabilities proportional to the weights, to produce a population of size J. As $L \to \infty$ this produces an asymptotically valid population from $\alpha_{i,M} | y_1, ..., y_{i+1}$ just using the theory of sampling, importance resampling (Rubin (1987) and Rubin (1988)). As $\alpha_{i,M} = \alpha_{i+1,0}$ this solves the task we set ourselves.

6.4 Model comparison

An advantage of the Bayesian MCMC approach to inference is that there is a principled way in which to compare the fit of various competing diffusion based models. Bayesian's compare models through their marginal likelihoods

$$\begin{aligned} f(y) &= \int f(y|\theta) f(\theta) \mathrm{d}\theta \\ &= \frac{f(y|\theta) f(\theta)}{f(\theta|y)}, \end{aligned}$$

for any value of θ where $f(\theta|y) > 0$. A discussion of this general problem can be found in Chib (2001) who recommends numerically approximating $f(y|\theta)$ and $f(\theta|y)$ with θ taken as

$\mathrm{E}(\theta|y)$, which can be found using the MCMC analysis. This approach was implemented for fully observed diffusions by Elerian, Chib, and Shephard (2001). No real new issues of principle arise here, however it is now harder to estimate $f(y|\theta)$. This quantity can be approximated a single θ using the particle filter discussed in the previous subsection.

6.5 Model checking

Model checking can be carried out via the one-step ahead forecast densities of the observables y_i. In this exposition we will assume this is a univariate series, although similar ideas can be developed for the multivariate case. Throughout we will work with

$$u_{i+1} = \Pr(y_{i+1}|y_1, ..., y_i; \theta).$$

Such a time series of distribution functions should be iid uniform on the interval 0 to 1 if the model is correct. This is discussed in the context of non-Gaussian time series models by, for example, Smith (1985), Shephard (1994) and Kim, Shephard, and Chib (1998), although earlier work goes back to at least Rosenblatt (1952). The later paper on this topic by X, Gunther, and Tay () has attracted some attention. Work on using this idea in the context of exactly observed diffusions is given in Elerian, Chib, and Shephard (2001).

In this context the main challange is actually computing u_{i+1}. However, the output from the particle filter solves this problem for it produces a sample from $\alpha_{i,0}|y_1, ..., y_i$. Hence we can use this sample to propergate the a sample from $\alpha_{i+1,0}|y_1, ..., y_i$, which gives us a sample from $y_{i+1}|y_1, ..., y_i; \theta$. Thus we can just estimate u_{i+1} by counting the number of simulations below the observed value.

7 Conclusion

This paper has provided a unified likelihood based approach for inference in observed and partially observed multivariate diffusions. This is based on a effective proposal scheme which is developed out of some work by Durham and Gallant (2002) and represents an extension of some earlier work of Kim, Shephard, and Chib (1998) and Elerian, Chib, and Shephard (2001). This MCMC method is rather robust and can, in principle, work even in the context of large dimensional diffusions or diffusions with many state variables.

We extend the basic analysis to talk about how to deal with jumps in the model. The framework also provides a consistent way of dealing with other issues such as model choice, model checking and filtering.

8 Appendix

8.1

We will show that for $N > k > j$ then under the Euler law

$$\alpha_k|\alpha_j, \alpha_N \sim N\left(\alpha_j + m_j^*, v_j^*\Sigma(\alpha_{j-1})\right), \quad m_j^* = \frac{(k-j)}{N-j}\left(\alpha_N - \alpha_j\right), \quad v_j^* = \delta\frac{(k-j)\,(N-k)}{(N-j)}.$$

Proof. Now

$$\begin{pmatrix} \alpha_k \\ \alpha_N \end{pmatrix}|\alpha_j \sim N\left\{\begin{pmatrix} 1 \\ 1 \end{pmatrix} \otimes \alpha_j + \delta\begin{pmatrix} k-j \\ N-j \end{pmatrix} \otimes \mu(\alpha_{j-1}), \delta\begin{pmatrix} k-j & k-j \\ k-j & N-j \end{pmatrix} \otimes \Sigma(\alpha_{j-1})\right\}.$$

This implies that

$$\alpha_k|\alpha_j, \alpha_N \sim N\left(\alpha_j + m_j^*, v_j^*\Sigma(\alpha_{j-1})\right), \quad m_j^* = \frac{(k-j)}{N-j}\left(\alpha_N - \alpha_j\right), \quad v_j^* = \delta\frac{(k-j)\,(N-k)}{(N-j)}.$$

as

$$v_j^* = \frac{(k-j)\,(N-j)}{(N-j)} - \frac{(k-j)^2}{N-j} = \frac{(k-j)\,(N-k)}{(N-j)}$$

and

$$m_j^* = \delta\,(k-j)\,\mu(\alpha_{j-1}) + \frac{\delta\,(k-j)}{\delta N - j}\left(\alpha_N - \alpha_j - \delta\,(N-j)\,\mu(\alpha_{j-1})\right).$$

8.2 Sampling $\theta|\alpha^+$ for the log OU-Vol model

In our OU volatility model, the discretisation of the volatility $x_t = \alpha_{2t}$ follows,

$$x_{t+1} = x_t + k(\mu - x_t)\delta + \sigma\delta^{\frac{1}{2}}u_t,$$

for $t = 1, ..., n = M(T-1)$. where $u_t \sim \mathsf{NID}(0,1)$. For ease of notation we may write this as,

$$y_t = (\beta_0\ \beta_1)(1\ x_t)' + \sigma^* u_t,$$

for $t = 1, ..., n$. where $y_t = \frac{x_{t+1}-x_t}{\delta}$, $\beta_0 = k\mu$, $\beta_1 = -k$, $\sigma^{*2} = \sigma^2/\delta$. We also have the initial condition,

$$x_1 \sim \mathsf{NID}\left(\mu, \frac{\sigma^2}{2k}\right) = \mathsf{NID}\left(\frac{\beta_0}{-\beta_1}, \frac{\delta\sigma^{*2}}{-2\beta_1}\right).$$

Our full conditional likelihood can now be written as,

$$f(\beta_0, \beta_1, \sigma^{*2} \mid x_1, x_2, ..., x_n) \propto f(y_1, ..., y_{n-1} \mid \beta_0, \beta_1, \sigma^{*2}) f(x_1 \mid \beta_0, \beta_1, \sigma^{*2}) p(\beta_0, \beta_1, \sigma^{*2}).$$

The term $p(\beta_0, \beta_1, \sigma^{*2})$ is the joint prior which is conjugate to the linear model likelihood $f(y_1, ..., y_{n-1} \mid \beta_0, \beta_1, \sigma^{*2})$. The initial condition term will not be conjugate and is therefore

ignored in our proposal. This is corrected in the Metropolis-Hastings step. We have the proposal,

$$g(\beta_0, \beta_1, \sigma^{*2} \mid x_1, x_2, ..., x_n) \propto f(y_1, ..., y_{n-1} \mid \beta_0, \beta_1, \sigma^{*2}) p(\beta_0, \beta_1, \sigma^{*2}),$$

where $g(\beta_0, \beta_1, \sigma^{*2} \mid y_1, ..., y_{n-1}) = g(\beta_0, \beta_1 \mid y_1, ..., y_{n-1}; \sigma^{*2}) g(\sigma^{*2} \mid y_1, ..., y_{n-1})$. Using standard statistical procedures, see for instance Bernardo and Smith (1994), we may proceed as

Our prior is

$$p(\beta_0, \beta_1, \sigma^{*2}) \propto p(\beta_0, \beta_1 |\ \sigma^{*2}) p(\sigma^{*2}) I(\beta_1 < 0),$$

where $p(\beta_0, \beta_1 |\ \sigma^{*2})$ is a bivariate Gaussian, $\mathrm{N}(\mu_\beta, \Omega_\beta^{-1})$ and $p(\sigma^{*2}) \propto \sigma^{*-2}$. The indicator term $I(\beta_1 < 0)$, restricts $\beta_1 < 0$, hence $k > 0$ ensuring stationarity. Letting $\widehat{\beta} = (X'X)^{-1}X'y$ and $S^2 = (y - X\widehat{\beta})'(y - X\widehat{\beta})/(n-3)$ we obtain,

$$\begin{aligned} g(\sigma^{*2} \mid y) &= \mathrm{Iga}\left(\frac{(n-3)}{2}; \frac{(n-3)}{2} S^2\right) \\ g(\beta_0, \beta_1 | y; \sigma^{*2}) &= \mathrm{N}\left(\mu_\beta^*; \Omega_\beta^{*-1}\right), \end{aligned}$$

where

$$\begin{aligned} \Omega_\beta^* &= \sigma^{*-2}(X'X) + \Omega_\beta \\ \mu_\beta^* &= \Omega_\beta^{*-1}\{\sigma^{*-2}(X'X)\widehat{\beta} + \Omega_\beta \mu_\beta\}. \end{aligned}$$

Having sampled from this joint density we decide whether or not to accept our proposal $\Psi^p = (\beta_0^p, \beta_1^p, \sigma^{*2p})'$ based upon the Metropolis probability, involving only the initial distribution,

$$\Pr(\Psi^c \longrightarrow \Psi^p) = \min\left[\frac{f(x_1 \mid \Psi^p)}{f(x_1 \mid \Psi^c)}\right],$$

where $\Psi^c = (\beta_0^c, \beta_1^c, \sigma^{*2c})'$ is the current parameter vector. This gives the probability of moving from the current parameter vector to the proposed value. If $\beta_1^p < 0$, we retain the current vector, Ψ^c. We convert our vector back to the original parameterisation $(k,\ \mu,\ \sigma^2)'$ in a simple manner.

9 Acknowledgments

Neil Shephard's research is supported by the UK's ESRC through the grant "Econometrics of trade-by-trade price dynamics," which is coded R00023839. All the calculations made in this paper are based on software written by the authors using the Ox language of Doornik (2001).

References

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

Ait-Sahalia, Y. (1996b). Testing continuous-time models of the spot interest rate. *Review of Financial Studies 9*, 385–426.

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

Ait-Sahalia, Y. (2003). Closed-form likelihood expansions for multivariate diffusions. Unpublished paper: Department of Economics, Princeton University.

Barndorff-Nielsen, O. E. and N. Shephard (2001). Non-Gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics (with discussion). *Journal of the Royal Statistical Society, Series B 63*, 167–241.

Baum, L. E., T. Petrie, G. Soules, and N. Weiss (1970). A maximization technique occuring in the statistical analysis of probabilistic functions of Markov chains. *Annals of Mathematical Statistics 41*, 164–171.

Bernardo, J. M. and A. F. M. Smith (1994). *Bayesian Theory*. Chichester: John Wiley.

Black, F. (1976). Studies of stock price volatility changes. *Proceedings of the Business and Economic Statistics Section, American Statistical Association*, 177–181.

Brandt, M. W. and P. Santa-Clara (2002). Simulated likelihood estimation of diffusions with an application to exchange rates dynamics in incomplete markets. *Journal of Financial Economics 63*, 161–210.

Carlin, B. P., N. G. Polson, and D. Stoffer (1992). A Monte Carlo approach to nonnormal and nonlinear state-space modelling. *Journal of the American Statistical Association 87*, 493–500.

Carter, C. K. and R. Kohn (1994). On Gibbs sampling for state space models. *Biometrika 81*, 541–53.

Chernov, M., A. R. Gallant, E. Ghysels, and G. Tauchen (2002). Alternative models of stock price dynamics. *Journal of Econometrics*. Forthcoming.

Chib, S. (2001). Markov chain Monte Carlo methods: computation and inference. In J. J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics*, Volume 5, pp. 3569–3649. Amsterdam: North-Holland.

Chib, S. and E. Greenberg (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistican 49*, 327–35.

Chib, S., F. Nardari, and N. Shephard (2002). Markov chain Monte Carlo methods for generalized stochastic volatility models. *Journal of Econometrics 108*, 281–316.

Doornik, J. A. (2001). *Ox: Object Oriented Matrix Programming, 3.0*. London: Timberlake Consultants Press.

Durbin, J. and S. J. Koopman (2002). A simple and efficient simulation smoother for state space time series analysis. *Biometrika 89*, 603–616.

Durham, G. and A. R. Gallant (2002). Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes (with discussion). *Journal of Business and Economic Statistics 20*, 297–338.

Egorov, A. V., H. Li, and Y. Xu (2003). Maximum likelihood estimation of time-inhomogeneous diffusions. *Journal of Econometrics 114*, 107–139.

Elerian, O. (1999). Simulation estimation of continuous-time models with applications to finance. Unpublished D.Phil. thesis, Nuffield College, Oxford.

Elerian, O., S. Chib, and N. Shephard (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. *Journal of Business and Economic Statistics 19*, 177–191.

Eraker, B., M. Johannes, and N. Polson (2002). The impact of jumps in returns and volatility. *Journal of Finance*. Forthcoming.

Florens-Zmirou, D. (1989). Approximate discrete-time schemes for statistics of diffusion processes. *Statistics 20*, 547–557.

Fruhwirth-Schnatter, S. and L. Soegner (2001). Bayesian estimation of the Barndorff-Nielsen-Shephard stochastic volatility model with marginal gamma laws. Statistical research report, University of Vienna.

Gallant, A. R. and J. R. Long (1997). Estimating stochastic differential equations efficiently by minimum chi-square. *Biometrika 84*, 125–41.

Gallant, A. R. and G. Tauchen (1996). Which moments to match. *Econometric Theory 12*, 657–81.

Gallant, A. R. and G. Tauchen (1998). Reprojection partially observed systems with applications to interest rate diffusions. *Journal of the American Statistical Association 93*, 10–24.

Gallant, A. R. and G. Tauchen (2003). Simulated score methods and indirect inference for continuous-time models. In Y. Ait-Sahalia and L. P. Hansen (Eds.), *Handbook of Financial Econometrics*. Amsterdam: North Holland. Forthcoming.

Geweke, J. (1989). Bayesian inference in econometric models using Monte Carlo integration. *Econometrica 57*, 1317–1339.

Ghysels, E., A. C. Harvey, and E. Renault (1996). Stochastic volatility. In C. R. Rao and G. S. Maddala (Eds.), *Statistical Methods in Finance*, pp. 119–191. Amsterdam: North-Holland.

Gordon, N. J., D. J. Salmond, and A. F. M. Smith (1993). A novel approach to nonlinear and non-Gaussian Bayesian state estimation. *IEE-Proceedings F 140*, 107–113.

Gourieroux, C., A. Monfort, and E. Renault (1993). Indirect inference. *Journal of Applied Econometrics 8*, S85–S118.

Griffin, J. E. and M. F. J. Steel (2002). Inference with non-Gaussian Ornstein-Uhlenbeck processes for stochastic volatility. Unpublished paper: Institute of Mathematics and Statistics, University of Kent at Canterbury.

Hamilton, J. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica 57*, 357–384.

Hansen, L. P. and J. Scheinkman (1995). Back to the future: generating moment implications for continuous-time Markov processes. *Econometrica 63*, 767–804.

Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.

Harvey, A. C. (1993). *Time Series Models* (2 ed.). Hemel Hempstead: Harvester Wheatsheaf.

Hendry, D. F. (1995). *Dynamic Econometrics*. Oxford: Oxford University Press.

Hull, J. and A. White (1987). The pricing of options on assets with stochastic volatilities. *Journal of Finance 42*, 281–300.

Hurn, A., K. A. Lindsay, and V. L. Martin (2003). On the efficacy of simulated maximum likelihood for estimating the parameters of stochastic differential equations. *Journal of Time Series Analysis 24*, 45–63.

Jacod, J. and P. Del Moral (2001). Interacting particle filtering with discrete observations. In A. Doucet, J. F. G. de Freitas, and N. J. Gordon (Eds.), *Sequential Monte Carlo Methods in Practice*, pp. 43–77. New York: Springer-Verlag.

Jacod, J., P. Del Moral, and P. Protter (2001). The Monte-Carlo method for filtering with discrete-time observations. *Probability Theory and Related Fields 120*, 346–368.

Jiang, G. L. and J. L. Knight (1997). A nonparametric approach to the estimation of diffusion processes, with an application to a short-term interest-rate models. *Econometric Theory 13*, 615–645.

Johannes, M., N. Polson, and J. Stroud (2002). Nonlinear filtering of stochastic differential equations with jumps. Unpublished paper: Graduate School of Business, Columbia University.

Kessler, M. and M. Sorensen (1999). Estimating equations based on eigenfunctions for a discretely observed diffusion process. *Bernoulli 5*, 299–314.

Kim, S., N. Shephard, and S. Chib (1998). Stochastic volatility: likelihood inference and comparison with ARCH models. *Review of Economic Studies 65*, 361–393.

Kohatsu-Higa, A. and S. Ogawa (1997). Weak rate of convergence for an Euler scheme of nonlinear SDE's. *Monte Carlo Methods and Applications 3*, 327–345.

Koopman, S. J. and N. Shephard (2002). Testing the assumptions behind the use of importance sampling. Unpublished paper: Nuffield College, Oxford.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset pricing: a new approach. *Econometrica 59*, 347–370.

Nicolau, J. (2002). A new technique for simulating the likelihood of stochastic differential equations. *Econometrics Journal 5*, 91–103.

Pedersen, A. R. (1995). A new approach to maximum likelihood estimation for stochastic differential equations on discrete observations. *Scandinavian Journal of Statistics 27*, 55–71.

Pitt, M. K. and N. Shephard (1999). Filtering via simulation: auxiliary particle filter. *Journal of the American Statistical Association 94*, 590–599.

Roberts, G., O. Papaspiliopoulos, and P. Dellaportas (2001). Bayesian inference for non-gaussian Ornstein-Uhlenbeck stochastic volatility processes. Submitted paper. Department of Mathematics and Statistics, Lancaster University.

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

Rosenblatt, M. (1952). Remarks on a multivariate transformation. *Annals of Mathematical Statistics 23*, 470–2.

Rubin, D. B. (1987). A noniterative sampling/importance resampling alternative to the data augmentation algorithm for creating a few imputations when the fraction of missing information is modest: the SIR algorithm. Discussion of **?**). *Journal of the American Statistical Association 82*, 543–546.

Rubin, D. B. (1988). Using the SIR algorithm to simulate posterior distributions. In J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith (Eds.), *Bayesian Statistics 3*, pp. 395–402. Oxford: Oxford University Press Press.

Schaumburg, E. (2002). Estimation of discretely sampled Markov processes with jumps. Unpublished paper: Finance Department, Northwestern University.

Shephard, N. (1994). Partial non-Gaussian state space. *Biometrika 81*, 115–31.

Shephard, N. (1996). Statistical aspects of ARCH and stochastic volatility. In D. R. Cox, D. V. Hinkley, and O. E. Barndorff-Nielsen (Eds.), *Time Series Models in Econometrics, Finance and Other Fields*, pp. 1–67. London: Chapman & Hall.

Smith, A. A. (1993). Estimating nonlinear time series models using simulated vector autoregressions. *Journal of Applied Econometrics 8*, S63–S84.

Smith, J. Q. (1985). Diagnostic checks of non-standard time series models. *J. Forecasting 4*, 283–91.

Sørensen, M. (1997). Estimating equations for discretely observed diffusions: a review. In I. V. Basawa, V. P. Godambe, and R. L. Taylor (Eds.), *Selected Proceedings of the Symposium on Estimating Functions*, pp. 305–325. IMS Lecture Notes. Monograph Series.

Sørensen, M. (2000). Prediction based estimating equations. *Econometrics Journal 3*, 123–147.

Tanner, M. A. and W. H. Wong (1987). The calculation of posterior distributions by data augmentation (with discussion). *Journal of the American Statistical Association 82*, 528–50.

X, D. F., T. A. Gunther, and T. S. Tay. Evaluating density forecasts with applications to financial risk management. *International Economic Review 1998*, 863–883.