Bayesian Sequential Inference for Nonlinear Multivariate Diffusions

Andrew Golightly* and Darren J. Wilkinson

University of Newcastle upon Tyne, NE1 7RU, U.K.

24th May 2006

Abstract

In this paper, we adapt recently developed simulation-based sequential algorithms to the problem concerning the Bayesian analysis of discretely observed diffusion processes. The estimation framework involves the introduction of $m-1$ latent data points between every pair of observations. Sequential MCMC methods are then used to sample the posterior distribution of the latent data and the model parameters on-line. The method is applied to the estimation of parameters in a simple stochastic volatility model (SV) of the U.S. short-term interest rate. We also provide a simulation study to validate our method, using synthetic data generated by the SV model with parameters calibrated to match weekly observations of the U.S. short-term interest rate.

1 Introduction

Diffusion processes governed by stochastic differential equations (SDEs) can be a convenient tool for modelling economic data. As such, the use of diffusion processes in the area of Mathematical Finance is becoming common place and much effort has been spent searching for efficient ways to estimate diffusion parameters.

*a.golightly@ncl.ac.uk

In the context of likelihood-based inference, estimation of the parameters requires knowledge of the Markovian transition density for the SDE. However, as analytic solutions of SDEs are rarely available, we cannot obtain the transition density in closed form and are forced to approximate it. Various attempts to solve this problem have been made which include moment based estimation (Chan, Karolyi, Longstaff & Sanders 1992) and simulation based methods; see for example Durham & Gallant (2002).

Typically, since observations arrive at discrete times, yet the model is formulated in continuous time, it is natural to work with the first order Euler-Maruyama approximation. Unfortunately the inter-observation times are usually too large to be used as a time step with such an approximation. The resulting problem is the classic missing data problem extended by Pedersen (1995) whose treatment involves augmenting the observed low-frequency data with the introduction of $m-1$ latent data points in between every pair of observations. Markov chain Monte Carlo (MCMC) methods (Chib & Greenberg 1995) which sample the posterior distribution and model parameters have been proposed by Jones (1997), Elerian, Chib & Shephard (2001) and Eraker (2001). Such methods can be computationally expensive but can be easily applied to partially observed diffusions.

Here, we focus on the MCMC approach and address the problems arising if the amount of augmentation is large, that is, as m increases. In this case, high dependence between the parameters and missing data results in arbitrarily slow rates of convergence of basic algorithms such as single site Gibbs samplers (Eraker 2001). For univariate SDEs, Roberts & Stramer (2001) propose a transformation of the diffusion to break down this dependence. However, this technique cannot be applied to general multivariate diffusions such as those considered in this paper (Wilkinson 2003).

We propose a simulation filter, utilising the diffusion bridge construct suggested by Durham & Gallant (2002), which allows on-line estimation, can be effectively implemented for large sample sizes, and doesn't break down as either the degree of augmentation, m or the number of observations increases. The method relies on recently developed particle filtering methods. Such methods have been discussed extensively in the context of discrete time series; see for example Pitt & Shephard (1999), Carpenter, Clifford & Fearnhead (1999) and Doucet, Godsill & Andrieu (2000). Filtering for

SDEs has been discussed by Del Moral & Jacod (2001) and Johannes, Polson & Stroud (2006) among others. We apply the methodology to the estimation of parameters in a two component model with a latent stochastic volatility (SV) variable. To validate the method, we also provide a simulation study using the SV model as a test case, with parameters calibrated to match weekly observations of the U.S. short-term interest rate.

The remainder of this paper is organised as follows. In Section 2 we formulate the model and detail how MCMC methods can be used to analyse diffusions, focusing on the Gibbs sampler in Section 3 and our proposed simulation filter in Section 4. The simple Log-Gaussian Stochastic Volatility model is presented in Section 5 with an illustrative application and simulation study. Conclusions are drawn in Section 6.

2 Models

We consider inference for a d-dimensional Itô Diffusion that satisfies a stochastic differential equation of the form

$$dY_t = \mu(Y_t, \Theta)\, dt + \beta^{\frac{1}{2}}(Y_t, \Theta)\, dW_t\,, \tag{1}$$

where $\mu(Y_t, \Theta)$ and $\beta^{\frac{1}{2}}(Y_t, \Theta)$ are drift and diffusion functions (of dimension d and $d \times d$ respectively), depending on Y_t and an unknown parameter vector Θ of dimension p. dW_t is the increment of d-dimensional standard Brownian motion. We assume that the conditions under which the SDE can be solved for Y_t are satisfied (Øksendal 1995).

Often, Y_t will consist of both observable and unobservable components. To deal with this, we define $Y_t = (X_t, Z_t)'$, where X_t defines the observable part and Z_t the unobservable part of the system. Note that X_t and Z_t have dimensions d_1 and d_2 respectively such that Y_t has dimension $d = d_1 + d_2$.

We assume that the process X_t will be observed at a finite number of times and the objective is to conduct inference for the (unknown) parameter vector Θ on the basis of these partial and discrete observations. In practice it is necessary to work with the

discretized version of (1), given by the Euler-Maruyama approximation,

$$\Delta Y_t = \mu(Y_t, \Theta)\Delta t + \beta^{\frac{1}{2}}(Y_t, \Theta))\Delta W_t \,, \tag{2}$$

where ΔW_t is a d dimensional iid $\mathrm{N}(0, I\Delta t)$ random vector and Δt is a small scalar.

Now suppose we have measurements x_{τ_i} at evenly spaced times τ_0, τ_1, ..., τ_T with intervals of length $\Delta^* = \tau_{i+1} - \tau_i$. Then put $\Delta t = \Delta^*/m$ for some positive integer $m \geq 1$. By choosing m to be sufficiently large, we can ensure that the discretization bias is arbitrarily small, but this also introduces the problem of $m-1$ missing values in between every pair of observations.

We deal with these missing values by dividing the entire time interval $[\tau_0, \tau_T]$ into $mT+1$ equidistant points $\tau_0 = t_0 < t_1 < \ldots < t_n = \tau_T$. Altogether we have $d_1T(m-1) + d_2(mT+1)$ missing values which we substitute with simulations Y_{t_i}. We refer to the collection of simulated data and observations as the augmented data. Eraker (2001) denotes by $\hat{\mathrm{Y}}$ the $d \times (n+1)$ matrix obtained by stacking all elements of the augmented data, that is

$$\hat{\mathrm{Y}} = \begin{pmatrix} x_{1,t_0} & X_{1,t_1} & \cdots & x_{1,t_m} & X_{1,t_{m+1}} & \cdots & x_{1,t_n} \\ x_{2,t_0} & X_{2,t_1} & \cdots & x_{2,t_m} & X_{2,t_{m+1}} & \cdots & x_{2,t_n} \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ x_{d_1,t_0} & X_{d_1,t_1} & \cdots & x_{d_1,t_m} & X_{d_1,t_{m+1}} & \cdots & x_{d_1,t_n} \\ Z_{1,t_0} & Z_{1,t_1} & \cdots & Z_{1,t_m} & Z_{1,t_{m+1}} & \cdots & Z_{1,t_n} \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ Z_{d_2,t_0} & Z_{d_2,t_1} & \cdots & Z_{d_2,t_m} & Z_{d_2,t_{m+1}} & \cdots & Z_{d_2,t_n} \end{pmatrix}.$$

We now let $Y^k = (X^k, Z^k)$ denote the k^{th} column of $\hat{\mathrm{Y}}$. The joint posterior density of parameters and augmented data is given by

$$\pi(\hat{\mathrm{Y}}, \Theta) \propto \pi(\Theta)\pi(Z^0) \prod_{k=0}^{n-1} \pi(Y^{k+1}|Y^k, \Theta), \tag{3}$$

where $\pi(\Theta)$ is the prior density of the parameter vector, $\pi(Z^0)$ is the prior density of Z^0 and

$$\pi(Y^{k+1}|Y^k, \Theta) = \phi(Y^{k+1}\,;\, Y^k + \mu_k\Delta t\,,\, \beta_k\Delta t) \tag{4}$$

Here, $\mu_k = \mu(Y^k, \Theta)$, $\beta_k = \beta(Y^k, \Theta)$ and $\phi(\cdot\,;\, \psi, \Sigma)$ denotes the Gaussian density with mean ψ and variance matrix Σ. Note that $\pi(Y^{k+1}|Y^k, \Theta)$ is the transition density obtained from the Euler discretization.

We have formulated in (3) the joint posterior for the model parameters as well as observed and unobserved data, but real interest will usually be in the distribution $(\Theta, \hat{\mathrm{Y}} \backslash D_n | D_n)$ where $D_n = (x^0, x^m, \ldots x^n)$ denotes the observed data (up to time t_n). We now turn our attention to sampling this distribution, focusing on three MCMC schemes in particular. As discussed in Tanner & Wong (1987), inference may proceed by alternating between simulation of parameters conditional on augmented data, and simulation of the missing data given the observed data and the current state of the model parameters. As the number of unobservables is typically large, a Gibbs sampler is a particularly convenient way of sampling from (3). However, as augmentation increases, high dependence between missing data and parameters results in arbitrarily slow rates of convergence. Further, as new data become available, the sample of parameter values must be discarded, and a new sample must be created by restarting the Gibbs sampler from scratch to include new observations.

As each new observation arrives, our proposed simulation filter samples a new $(\Theta_*, (\hat{\mathrm{Y}} \backslash D_n)_*)$ in two stages: first Θ_* is sampled from a suitable proposal and then $(\hat{\mathrm{Y}} \backslash D_n)_*$ is sampled from a tractable approximation to $(\hat{\mathrm{Y}} \backslash D_n) | \Theta_*, D_n$. By simulating the latent data to be consistent with Θ_*, the dependence between them is overcome. For further discussion of MCMC methods in the context of Bayesian analysis of diffusions, see Roberts & Stramer (2001) and Elerian et al. (2001).

3 Gibbs Sampling

3.1 Single Site Gibbs Sampler

Here we consider the case in which the Gibbs sampler updates one column of $\hat{\mathrm{Y}}$ at a time and refer to this algorithm as the single site Gibbs sampler. We now adopt the notation where π denotes all proper densities and p denotes π in an un-normalised

form. For $k \neq 0, n$, the full conditional for Y^k is

$$\pi(Y^k|Y^{k-1}, Y^{k+1}, \Theta) \propto p(Y^k|Y^{k-1}, Y^{k+1}, \Theta) \tag{5}$$

where

$$p(Y^k|Y^{k-1}, Y^{k+1}, \Theta) = \phi(Y^k\,;\, Y^{k-1} + \mu_{i-1}\Delta t\,,\, \beta_{i-1}\Delta t)\,\phi(Y^{k+1}\,;\, Y^k + \mu_i\Delta t\,,\, \beta_i\Delta t)\,.$$

At iteration s of the Gibbs sampler one then draws

$$Y^k \sim \pi(Y^k|Y^{k-1}, Y^{k+1}, \Theta)$$

where Y^{k-1} is obtained at iteration s and Y^{k+1} at iteration $s-1$. For nonlinear diffusions Eraker (2001) suggests that when k is not a multiple of m, Y^k is updated using a Metropolis-Hastings (M-H) step with proposal density

$$q(\cdot\,|Y^{k-1}, Y^{k+1}, \Theta) = \phi\left(\cdot\,;\, \frac{1}{2}(Y^{k-1} + Y^{k+1})\,,\, \frac{1}{2}\Delta t\beta(Y^{k-1}, \Theta)\right)\,. \tag{6}$$

Eraker motivates this choice by proving that

$$q(Y^k|Y^{k-1}, Y^{k+1}, \Theta) \to \pi(Y^k|Y^{k-1}, Y^{k+1}, \Theta) \text{ as } \Delta t \to 0\,.$$

When k divides m we only need to simulate the d_2 elements corresponding to Z^k. Again, we use a M-H step and proposal density given by (6) but further conditioned on the observation x^k. This leaves the two special cases, namely $k = 0$ and $k = n$. Using (3), the full conditional of Z^0 is clearly proportional to $\pi(Z^0)\pi(Y^1|Y^0, \Theta)$ and a M-H step can be implemented to sample this distribution. When $k = n$, the full conditional of Z^n is given by $\pi(Y^n|Y^{n-1}, \Theta)$ further conditioned on the observation x^n and direct sampling from this distribution is possible. For a detailed discussion see Golightly & Wilkinson (2005).

3.2 Blocking Strategies

Elerian et al. (2001) show that updating one column of $\hat{\Upsilon}$ at a time leads to poor

mixing due to high correlation amongst the latent data and recommend updating missing values in blocks of random size. We now consider an algorithm where the Gibbs sampler updates latent values in blocks of size $2m+1$ and we refer to this algorithm as the block Gibbs sampler. Note that in the case of complete observation, there is a much simpler version of this algorithm using blocks of size $m-1$, but we do not present that here.

Consider times t_j, t_M and t_{M^+} where j is an integer multiple of m, $M = j+m$ and $M^+ = M+m$. Note that these times correspond to the observations, x^j, x^M and x^{M^+}. Treating Y^j and Y^{M^+} as fixed, the full conditional for the latent path in (t_j, t_{M^+}), $Y^{j+1}, \ldots, Z^M, \ldots, Y^{M^+-1}$, is

$$\pi(Y^{j+1}, \ldots, Z^M, \ldots, Y^{M^+-1} | Y^j, x^M, Y^{M^+}, \Theta) \propto \prod_{i=j}^{M^+-1} \pi(Y^{i+1}|Y^i, \Theta)\,. \tag{7}$$

By sampling this distribution for $j = 0, m, \ldots n-2m$, the use of overlapping blocks ensures that all of the Z^j (except Z^0 and Z^n) get updated sufficiently often. Naturally, the end-points of the process corresponding to the unobserved component, Z^0 and Z^n, must also be updated. For every sweep of the sampler we fix Y^m and draw from

$$\pi(Z^0, Y^1, \ldots, Y^{m-1} | x^0, Y^m, \Theta) \propto \prod_{i=0}^{m-1} \pi(Y^{i+1}|Y^i, \Theta) \tag{8}$$

and fix Y^{n-m} and draw from

$$\pi(Y^{n-m+1}, \ldots, Y^{n-1}, Z^n | Y^{n-m}, x^n, \Theta) \propto \prod_{i=n-m}^{n-1} \pi(Y^{i+1}|Y^i, \Theta)\,, \tag{9}$$

thus ensuring that Z^0 and Z^n are updated.

Unfortunately, sampling $Y^{j+1}, \ldots, Z^M, \ldots, Y^{M^+-1}$ directly is not possible. The Euler scheme allows handling of the likelihood between two consecutive values of the process (since we have a linear Gaussian structure in this case). However, obtaining the conditional density of missing values between two given observations that are $2m$ steps apart, under the non-linear structure of the underlying diffusion process, is complicated. To deal with this problem, we use a M-H step and construct a proposal distribution by adapting a method proposed by Durham & Gallant (2002), (see also Elerian et al.

(2001)) which they refer to as the "modified bridge" approach.

3.3 Modified Diffusion Bridge

Consider first the task of sampling $Y^{j+1}, \ldots, Y^{M-1}, Z^M$ conditional on Y^j, x^M and Θ. For a fully observed diffusion, X_t, Durham & Gallant (2002) draw $X^{j+1}, \ldots, X^{M-1}$ from a Gaussian density based on the Euler scheme conditional on x^j and x^M. We now adapt their proposal to our partially observed diffusion Y_t. That is, treating Y^j and x^M as fixed, we draw Y^{k+1}, for $k = j, \ldots, M-2$, from a Gaussian approximation to $\pi(Y^{k+1}|Y^k, x^M, \Theta)$,

$$\tilde{\pi}(Y^{k+1}|Y^k, x^M, \Theta) \quad \propto \quad \pi(Y^{k+1}|Y^k, \Theta)\tilde{\pi}(x^M|Y^{k+1}, \Theta) \,/\, \tilde{\pi}(x^M|Y^k, \Theta)\,. \qquad (10)$$

Here $\pi(Y^{k+1}|Y^k, \Theta)$ is the usual one step ahead Euler transition density given by (4), $\tilde{\pi}(x^M|Y^{k+1}, \Theta)$ and $\tilde{\pi}(x^M|Y^k, \Theta)$ are given by a much cruder Euler approximation,

$$\tilde{\pi}(x^M|Y^{k+1}, \Theta) = \phi(x^M\,;\, X^{k+1} + \mu^x(Y^{k+1}, \Theta)\Delta^+\,,\, \beta^{xx}(Y^{k+1}, \Theta)\Delta^+)\,, \qquad (11)$$

and

$$\tilde{\pi}(x^M|Y^k, \Theta) = \phi(x^M\,;\, X^k + \mu^x(Y^k, \Theta)\Delta^-\,,\, \beta^{xx}(Y^k, \Theta)\Delta^-) \qquad (12)$$

where $\Delta^+ = (M-k-1)\Delta t$, $\Delta^- = (M-k)\Delta t$ and we partition $\mu(Y^k, \Theta)$ and $\beta(Y^k, \Theta)$ as

$$\mu(Y^k, \Theta) = \begin{pmatrix} \mu^x(Y^k, \Theta) \\ \mu^z(Y^k, \Theta) \end{pmatrix}\,,\; \beta(Y^k, \Theta) = \begin{pmatrix} \beta^{xx}(Y^k, \Theta) & \beta^{xz}(Y^k, \Theta) \\ \beta^{zx}(Y^k, \Theta) & \beta^{zz}(Y^k, \Theta) \end{pmatrix}\,.$$

Although (11) is not linear Gaussian, we can approximate it further by noting that μ and β are locally constant (by assumption). We therefore estimate μ_{k+1} and β_{k+1} by μ_k and β_k respectively, to give

$$\tilde{\pi}(x^M|Y^{k+1}, \Theta) = \phi(x^M\,;\, X^{k+1} + \mu^x(Y^k, \Theta)\Delta^+\,,\, \beta(Y^k, \Theta)\Delta^+)\,. \qquad (13)$$

The approximate joint density of Y^{k+1} and x^M (conditional on Y^k and Θ) can then be found using standard multivariate Normal theory (see for example, page 21 of

Gamerman (1997)). We obtain,

$$\begin{pmatrix} Y^{k+1} \\ x^M \end{pmatrix} \sim \mathrm{N}_{d+d_1} \left\{ \begin{pmatrix} Y^k + \mu_k \Delta t \\ X^k + \mu_k^x \Delta^- \end{pmatrix}, \begin{pmatrix} \beta_k \Delta t & C_k \Delta t \\ C_k^{'} \Delta t & \beta_k^{xx} \Delta^- \end{pmatrix} \right\} \tag{14}$$

where $C_k^{'} = (\beta_k^{xx}, \beta_k^{xz})$. Conditioning (14) on x^M and simplifying yields,

$$\tilde{\pi}(Y^{k+1}|Y^k, x^M, \Theta) = \phi(Y^{k+1}\,;\, Y^k + \eta_k\,,\, \sigma_k)\,, \tag{15}$$

$$\eta_k = \left(\frac{1}{M-k}\right) \begin{pmatrix} x^M - X^k \\ \mu_k^{zz}(M-k)\Delta t + \beta_k^{zx}(\beta_k^{xx})^{-1}(x^M - [X^k + \mu_k^{xx}(M-k-1)\Delta t]) \end{pmatrix},$$

$$\sigma_k = \Delta t \left(\frac{1}{M-k}\right) \begin{pmatrix} (M-k-1)\beta_k^{xx} & (M-k-1)\beta_k^{xz} \\ (M-k-1)\beta_k^{zx} & (M-k)\beta_k^{zz} - \beta_k^{zx}(\beta_k^{xx})^{-1}\beta_k^{xz} \end{pmatrix}.$$

Hence, drawing recursively from (15) for $k = j, \ldots, M-2$ gives a diffusion bridge, $Y^{j+1}, \ldots, Y^{M-1}$ conditioned to start at Y^j and finish at x^M. Thus, (15) can be used as a proposal density in a M-H step inside the block Gibbs algorithm. However, a more efficient proposal density can be found by noting that in the block Gibbs algorithm we can also condition on Y^{M^+}. We therefore construct the Gaussian approximation, $\tilde{\pi}(Y^{k+1}|Y^k, x^M, Y^{M^+}, \Theta)$ by first writing the approximate joint density of Y^{k+1}, x^M and Y^{M^+}, conditional on Y^k and Θ,

$$\begin{pmatrix} Y^{k+1} \\ Y^{M^+} \\ x^M \end{pmatrix} \sim \mathrm{N}_{2d+d_1} \left\{ \begin{pmatrix} Y^k + \mu_k \Delta t \\ Y^k + \mu_k \Delta^{\sim} \\ X^k + \mu_k^x \Delta^- \end{pmatrix}, \begin{pmatrix} \beta_k \Delta t & \beta_k \Delta t & C_k \Delta t \\ \beta_k \Delta t & \beta_k \Delta^{\sim} & C_k \Delta t \\ C_k^{'} \Delta t & C_k^{'} \Delta t & \beta_k^{xx} \Delta^- \end{pmatrix} \right\} \tag{16}$$

where $\Delta^{\sim} = (M^+ - k)\Delta t$. Conditioning (16) on x^M and Y^{M^+} yields

$$\tilde{\pi}(Y^{k+1}|\, Y^k, x^M, Y^{M^+}, \Theta) = \phi(Y^{k+1}\,;\, Y^k + \eta_k^{'}\,, \sigma_k^{'}) \tag{17}$$

where

$$\eta_k^{'} = \mu_k \Delta t \,+\, \Delta t(\beta_k\,,\, C_k) \begin{pmatrix} \beta_k \Delta^{\sim} & C_k \Delta t \\ C_k^{'} \Delta t & \beta_k^{xx} \Delta^- \end{pmatrix}^{-1} \begin{pmatrix} Y^{M^+} - [Y^k + \mu_k^x \Delta^{\sim}] \\ x^M - [X^k + \mu_k^x \Delta^-] \end{pmatrix}.$$

$$\sigma_k^{'} = \beta_k \Delta t \; - \; \Delta t(\beta_k \, , \, C_k) \begin{pmatrix} \beta_k \Delta^{\sim} & C_k \Delta t \\ C_k^{'} \Delta t & \beta_k^{xx} \Delta^{-} \end{pmatrix}^{-1} \begin{pmatrix} \beta_k \\ C_k^{'} \end{pmatrix} \Delta t \, .$$

Naturally, Z^M must also be simulated and this is achieved by drawing from a Gaussian approximation,

$$Z^M \sim \tilde{\pi}(Z^M | Y^{M-1}, x^M, Y^{M^+}, \Theta) \, . \tag{18}$$

We use the density (17) as a proposal distribution inside the block Gibbs algorithm. First, we propose Y_*^{k+1} for $k = j, \ldots, M-2$ recursively from $\tilde{\pi}(Y_*^{k+1} | Y_*^k, x^M, \Theta)$. Next, we draw the mid-point, Z_*^M, from $\tilde{\pi}(Z_*^M | Y_*^{M-1}, x^M, Y^{M^+}, \Theta)$. We are then tasked with proposing $Y_*^{M+1}, \ldots, Y_*^{M^+-1}$. We achieve this by drawing Y_*^{k+1} for $k = M, \ldots, M^+-2$ from $\tilde{\pi}(Y_*^{k+1} | Y_*^k, Y^{M^+}, \Theta)$; that is, from the modified bridge density conditioned to start at Y_*^M, finish at Y^{M^+} and which is easily seen to be

$$\tilde{\pi}(Y_*^{k+1} | Y_*^k, Y^{M^+}, \Theta) = \phi \left(Y^{k+1} \, ; \, Y_*^k + \frac{Y^{M^+} - Y_*^k}{M^+ - k} \, , \, \frac{M^+ - k - 1}{M^+ - k} \beta(Y_*^k, \Theta) \Delta t \right) . \tag{19}$$

We let $q(Y_*^{j+1}, \ldots, Z_*^M, \ldots, Y_*^{M^+-1} | Y^j, x^M, Y^{M^+}, \Theta)$ denote the transition density for the move. This density is given by

$$\pi(Z_*^M | Y_*^{M-1}, x^M, Y^{M^+}, \Theta) \prod_{k=j}^{M-2} \tilde{\pi}(Y_*^{k+1} | Y_*^k, x^M, Y^{M^+}, \Theta) \prod_{k=M}^{M^+-2} \tilde{\pi}(Y_*^{k+1} | Y_*^k, Y^{M^+}, \Theta) \, . \tag{20}$$

Hence, if the chain is currently at $Y^{j+1}, \ldots, Z^M, \ldots, Y^{M^+-1}$, we accept a move to $Y_*^{j+1}, \ldots, Z_*^M, \ldots, Y_*^{M^+-1}$ with probability $\min\{1, A\}$ where,

$$A = \frac{\prod_{k=j}^{M^+-1} \pi(Y_*^{k+1} | Y_*^k, \Theta)}{\prod_{k=j}^{M^+-1} \pi(Y^{k+1} | Y^k, \Theta)} \times \frac{q(Y^{j+1}, \ldots, Z^M, \ldots, Y^{M^+-1} | Y^j, x^M, Y^{M^+} \Theta)}{q(Y_*^{j+1}, \ldots, Z_*^M, \ldots, Y_*^{M^+-1} | Y^j, x^M, Y^{M^+}, \Theta)} \, . \tag{21}$$

Finally, we consider the problem of sampling the remaining conditionals (8) and (9). A M-H step can be used to sample these densities — we draw Z_*^0 using a Metropolis random walk move and then propose Y_*^{k+1} for $k = 0, \ldots, m-2$ from $\tilde{\pi}(Y_*^{k+1} | Y_*^k, Y^m, \Theta)$ (given by (19) with M^+ replaced by m). We sample (9) by proposing Y_*^{k+1} for $k = n-m, \ldots, n-2$ from $\tilde{\pi}(Y_*^{k+1} | Y_*^k, x^n, \Theta)$, (given by (15) with M replaced by n) and

Z_*^n from $\pi(Z_*^n|Y_*^{n-1}, x^n, \Theta)$ — the one step Euler transition density conditioned on x^n. Acceptance probabilities for these moves are computed in the usual way. Further discussion of block Gibbs style algorithms can be found in Chib, Pitt & Shephard (2004).

3.4 Sampling from the Full Conditional for Θ

The remaining step in either the single site or block Gibbs sampler is to sample $\Theta_{(s)}$ conditional on the current state of Θ and the augmented data. The general form of $\pi(\Theta|\hat{\mathrm{Y}})$ is proportional to (3) and if the density cannot be recognised, a M-H step can be used. For the SV model discussed in Section 3, we use a Metropolis random walk update to sample Θ.

It may be the case that we require the last c components of $\Theta = (\theta_1, \ldots, \theta_p)'$ to be strictly positive. Here we set $\Lambda = (\lambda_1, \lambda_2, \ldots, \lambda_p)'$ where

$$\lambda_k = \begin{cases} \theta_k, & k = 1, \ldots, p-c, \\ \log(\theta_k), & k = p-c+1, \ldots, p \end{cases}$$

and assume independent proper Uniform priors for each λ_k. At iteration s, if our current value is Λ, we propose Λ_* by setting $\lambda_k^* = \lambda_k + w_k$, $k = 1, \ldots, p$ where $w_k \sim \mathrm{N}(0, \gamma_k)$ and the collection of tuning parameters, $\gamma = (\gamma_1, \ldots, \gamma_p)'$ determines the precise mixing properties of the resulting Markov chain.

Now set $\Theta = (\lambda_1, \ldots, \lambda_{p-c}, \exp(\lambda_{p-c+1}), \ldots, \exp(\lambda_p))'$. Then using (3) the likelihood of Θ (under the Euler scheme) is

$$L(\Theta|\hat{\mathrm{Y}}) = \prod_{k=0}^{n-1} \pi(Y^{k+1}|Y^k, \Theta)\,.$$

Hence at iteration s we accept the candidate value Θ_* with probability

$$\alpha(\Theta, \Theta_*|\hat{\mathrm{Y}}) = \min\left\{1, \frac{L(\Theta_*|\hat{\mathrm{Y}})}{L(\Theta|\hat{\mathrm{Y}})}\right\},$$

if Θ_* is consistent with the prior, and reject the move otherwise.

The single site Gibbs sampler then has the following algorithmic form:

1. Initialise all unknowns. Use linear interpolation to initialise X^k. Set the iteration counter to $s = 1$.

2. For all $k = 0, 1, \ldots, n$, at iteration s, draw $Y^k_{(s)}$ from its full conditional:-

   - For k not an integer multiple of m, draw $Y^k_{(s)}$ using a M-H step with proposal density given by (6).
   - When k is an integer multiple of m but $k \neq 0, n$, use a M-H step and propose Y^k_* from (6) further conditioned on x^k.
   - When $k = 0$, draw $Y^0_{(s)}$ using a M-H step.
   - When $k = n$, draw $Y^n_{(s)}$ directly from its full conditional.

3. Draw $\Theta_{(s)}$ using the Gaussian random walk update outlined above.

4. Increment s and return to step 2.

To implement the block Gibbs sampling strategy, step 2 above can be replaced by sampling from $\pi(Y^{j+1}, \ldots, Z^M, \ldots, Y^{M^+-1} | Y^j, x^M, Y^{M^+}, \Theta)$, given by (7), for $j = 0, m, \ldots, n - 2m$ using a M-H move with proposal density given by (20). The end points of the unobserved process, Z^0 and Z^n are updated by sampling from (8) and (9) with a M-H step.

3.5 Convergence Properties

As the amount of augmentation, m, increases, one can make very precise inference about the diffusion coefficient of the process via the quadratic variation (see Roberts & Stramer (2001)). It is this dependence (between the quadratic variation and diffusion coefficient) that results in slow mixing of MCMC algorithms such as the single site Gibbs sampler considered in Section 3.1 (see Amit (1991) and Eraker (2001)). Shephard & Pitt (1997) found that the chain converges faster by using random block sizes. Though this block updating method and also the blocking strategy of Section 3.2 are able to help overcome the dependence within the latent process, dependence between the parameters and latent process remains high and convergence will still become arbitrarily slow as either m or the number of observations increases. Roberts & Stramer (2001) overcome this dependence in the context of univariate diffusions by transforming the missing data, giving a partially non-centered parameterisation which leads to an

irreducible algorithm even in the limit $m \to \infty$. However, for a d-dimensional diffusion satisfying (1), finding such a transformation requires an invertible function, $g : \mathbf{R}^d \to \mathbf{R}^d$ such that

$$\nabla g(\nabla g)' = \beta^{-1} .$$

As stated by Papaspiliopolous, Roberts & Skôld (2003), this equation is almost always impossible to solve in practice for general non-linear multivariate diffusions such as the SV model considered here.

One possible solution to the dependence problem is to implement a joint update of the entire latent process and the parameters in a M-H step (Wilkinson 2003). This approach however, results in a very low acceptance rate as the number of observations increases. Although the proposed simulation filter relies on a joint update of Θ and the latent path, by updating sequentially as each observation becomes available, the sampler does not suffer from a low acceptance rate due to operating on the process one observation at a time. Furthermore, the joint update ensures that the latent path is consistent with the proposed diffusion coefficient and therefore dependence between them is overcome and the algorithm does not break down for large m or a large number of observations.

4 Simulation Filter

4.1 Introduction

Recent interest in sequential filtering has arisen due to the proposal of a class of filters known as 'particle filters'. Whereas the use of Monte Carlo filtering to produce estimates of posterior means and covariances can be traced back as far as Handschin & Mayne (1969), particle filters attempt to approximate the complete posterior. Such filters have been discussed extensively in the context of discrete time series with unobserved state variables, e.g., Berzuini, Best, Gilks & Larizza (1997), Pitt & Shephard (1999) and Doucet et al. (2000). Filtering for both parameters and state has been discussed by Liu & West (2001) and Stroud, Polson & Muller (2004) whilst applications involving SDEs have been examined by Del Moral, Jacod & Protter (2002) and Johannes et al. (2006). We start by outlining the general filtering approach (in the

context of discretely observed diffusions).

Consider data, $D_j = (x^0, x^m, \ldots, x^j)$, (where j is an integer multiple of m) arriving at times $t_0, t_m \ldots, t_j$ such that at time t_{j+m}, which we denote by t_M, new data x^M are accompanied by $m-1$ missing values, $X^{j+1}, \ldots, X^{M-1}$, and m values, $Z^{j+1}, \ldots, Z^{M-1}, Z^M$ corresponding to the unobserved component. As each observation becomes available we are interested in the on-line estimation of the unknown parameter vector, Θ.

We assume that we have a sample of size S from the distribution $\pi(\Theta, Z^j|D_j)$, which we denote by $\pi_j(\Theta, Z^j)$. At time t_M, we observe x^M. Assimilation of the information contained in x^M consists of generating a sample, $\{(\Theta_{(s)}, Z^M_{(s)}), s = 1, \ldots, S\}$ from the posterior $\pi_M(\Theta, Z^M)$ which can be found by formulating the posterior for parameters, latent and observed data, then integrating out the latent data. Using (3) we have

$$\pi_M(\Theta, Z^M) = \int_{\hat{\mathrm{Y}}_M} \pi(\Theta)\pi(Z^0) \prod_{k=0}^{M-1} \pi(Y^{k+1}|Y^k, \Theta) \tag{22}$$

where we define $\hat{\mathrm{Y}}_M = (Z^0, Y^1, \ldots, Y^{m-1}, Z^m, \ldots, Y^{M-1})$ and is simply the vector of latent values up to time t_M. Clearly, (22) can be written as

$$\pi_M(\Theta, Z^M) = \pi_j(\Theta, Z^j) \int_{\hat{\mathrm{Y}}_M \backslash \{\hat{\mathrm{Y}}_j\}} \prod_{k=j}^{M-1} \pi(Y^{k+1}|Y^k, \Theta) \tag{23}$$

where $\hat{\mathrm{Y}}_M \backslash \{\hat{\mathrm{Y}}_j\} = (Z^j, Y^{j+1}, \ldots, Y^{M-1})$ and is the vector of latent values in the interval $[t_j, t_M)$. So, at time t_M, our target is

$$\pi_M(\Theta, Z^M) \propto \pi_j(\Theta, Z^j) \prod_{k=j}^{M-1} \pi(Y^{k+1}|Y^k, \Theta) \tag{24}$$

with $Z^j, Y^{j+1}, \ldots, Y^{M-1}$ integrated out.

As $\pi_j(\Theta, Z^j)$ has no analytic form, the particle filter aims to recursively approximate this density by the the swarm of points or particles, $\{(\Theta_{(s)}, Z^j_{(s)}), s = 1, \ldots, S\}$ with each $\Theta_{(s)}, Z^j_{(s)}$ having a discrete probability mass of $1/S$. It is assumed that as $S \to \infty$, the particles approximate the filtering density, $\pi_j(\Theta, Z^j)$ increasingly well. Note that the filter treats the discrete support generated by the particles as the true (filtering) distribution. Various implementations of the particle filter are possible such as the

basic sampling/importance resampling (SIR) algorithm of Gordon, Salmond & Smith (1993) and rejection sampling (Pitt & Shephard 1999). We focus on a slightly different approach, sampling (24) by drawing $(Z^j, Y^{j+1}, \ldots, Y^{M-1}, Z^M, \Theta)$ via MCMC, then discarding all components except (Θ, Z^M). We refer to the algorithm as the simulation filter.

4.2 MCMC Sampling of Parameters and State

At time t_j we have a sample of size S, $\{(\Theta_{(s)}, Z^j_{(s)}), s = 1, \ldots, S\}$ from $\pi_j(\Theta, Z^j)$ and our goal, on observing x^M at time t_M, is to generate a sample, $\{(\Theta_{(s)}, Z^M_{(s)}), s = 1, \ldots, S\}$ from $\pi_M(\Theta, Z^M)$. Recall that $\pi_j(\Theta, Z^j)$ cannot be sampled directly and the particle filter treats the discrete support generated by the sample, $\{(\Theta_{(s)}, Z^j_{(s)}), s = 1, \ldots, S\}$, as the true density. Therefore, basic filtering algorithms initially propose (Θ_*, Z^j_*) from $\pi_j(\Theta, Z^j)$ by selecting an integer, u, from $\{1, \ldots, S\}$ and setting $(\Theta_*, Z^j_*) = (\Theta_{(u)}, Z^j_{(u)})$. Such an approach can, however, lead to sample impoverishment/depletion. Since parameters remain fixed through time, only a small number of distinct points proposed from $\pi_j(\Theta, Z^j)$ may be accepted. Since these points are propagated through to the next time point, this will result in a sample, $\{(\Theta_{(s)}, Z^M_{(s)}), s = 1, \ldots, S\}$ containing only a few distinct values of each $\Theta_{(s)}$.

Some attempts have been made to avoid impoverishment; Gordon et al. (1993) suggest adding random perturbations (or 'jitter') to state particles at each time step. Since then, the concept has been applied to fixed parameters. The idea is that if a particular value is replicated in the sample a number of times, it will be replaced by distinct (but similar) values. One first selects an integer, u, uniformly from the set $\{1, \ldots, S\}$ and then puts

$$(\Theta_*, Z^j_*)' \sim \mathrm{N}\{(\Theta_{(u)}, Z^j_{(u)})', h^2 V\} \tag{25}$$

where V is the Monte Carlo posterior variance and the overall scale of the kernel is a function of the smoothing parameter, h^2 usually dependent on the sample size, S. The

effect of (25) is to replace $\pi_j(\Theta, Z^j)$ in (24), with the smooth kernel density form,

$$\hat{\pi}_j(\Theta, Z^j) = \sum_{s=1}^{S} \phi\{(\Theta, Z^j)'; (\Theta_{(s)}, Z^j_{(s)})', h^2 V\}. \tag{26}$$

Note that this yields the filter's estimate of $\pi_M(\Theta, Z^M)$ as

$$\hat{\pi}_M(\Theta, Z^M) \propto \hat{\pi}_j(\Theta, Z^j) \prod_{k=j}^{M-1} \pi(Y^{k+1}|Y^k, \Theta). \tag{27}$$

The choice of h^2 in (26) is equivalent to the choice of smoothing parameter in kernel density estimation and a trade off between under and over smoothing should therefore be made. Standard rules of thumb for calculating a suitable h^2 can be found in Silverman (1986).

For large datasets however, Liu & West (2001) suggest that the random disturbances add up to give "information loss" over time (as the kernel density function is always over-dispersed relative to the posterior sample by a factor $1 + h^2$). To correct this, Liu & West (2001) employ a kernel shrinkage method by setting

$$(\Theta_*, Z^j_*)' \sim \mathrm{N}\{a(\Theta_{(u)}, Z^j_{(u)})' + (1-a)(\bar{\Theta}, \bar{Z}^j)', h^2 V\} \tag{28}$$

where $a^2 = 1 - h^2$, $h^2 = 1 - ((3\delta - 1)/2\delta)^2$, δ is a discount factor usually around 0.99 and $(\bar{\Theta}, \bar{Z}^j)'$ is the Monte Carlo posterior mean of $\pi_j(\Theta, Z^j)$. For the data considered in Sections 5.3 and 5.4 we find that using (25) works sufficiently well. See Liu & West (2001) and also West (1993) for further discussion on kernel smoothing.

Having proposed (Θ_*, Z^j_*), the final step in our MCMC strategy is to propose $Y^{j+1}_*, \ldots, Y^{M-1}_*, Z^M_*$ from a suitable proposal density. We use the modified bridge construct outlined in Section 3.3 and draw Y^{k+1}_* for $k = j, \ldots, M-2$ from the density, $\tilde{\pi}(Y^{k+1}_*|Y^k_*, x^M, \Theta_*)$ given by (15). We then propose Z^M_* from $\pi(Z^M_*|Y^{M-1}_*, x^M, \Theta_*)$ — that is, the one step ahead Euler transition density conditioned on x^M. Hence, if at some iteration, s, of our sampler we have current value

$$\Phi_{(s)} = (Z^j, Y^{j+1}, \ldots, Y^{M-1}, Z^M, \Theta),$$

then with probability $\min\{1, A\}$, where

$$A = \frac{\prod_{k=j}^{M-1} \pi(Y_*^{k+1}|Y_*^k, \Theta_*)}{\prod_{k=j}^{M-1} \pi(Y^{k+1}|Y^k, \Theta)} \times \frac{\pi(Z^M|Y^{M-1}, x^M, \Theta) \prod_{k=j}^{M-2} \tilde{\pi}(Y^{k+1}|Y^k, x^M, \Theta)}{\pi(Z_*^M|Y_*^{M-1}, x^M, \Theta_*) \prod_{k=j}^{M-2} \tilde{\pi}(Y_*^{k+1}|Y_*^k, x^M, \Theta_*)}, \quad (29)$$

we put $\Phi_{(s+1)} = \Phi_*$ and store $\Phi_{(s+1)}$ ready to be used at the next iteration. As with any MCMC sampler, the scheme can be modified by allowing a number of iterations to be discarded as "burn-in". After performing a further S iterations, the desired sample, $\{(\Theta_{(s)}, Z_{(s)}^M), s = 1, \ldots, S\}$ is obtained by dropping out all components of each $\Phi_{(s)}$ except $(\Theta_{(s)}, Z_{(s)}^M)$. Algorithmically, the simulation filter has the following form:

1. *Initialise* - Set $j = 0$. For $s = 1, \ldots, S$:
   - Draw $\Theta_{(s)} \sim \pi(\Theta)$ and $Z_{(s)}^0 \sim \pi(Z^0)$.
2. *MCMC* - Set $M = j + m$. Initialise $\Phi_{(0)}$. For $s = 1, \ldots, S$:
   - Propose (Θ_*, Z_*^j) using (25).
   - Propose $Y_*^{j+1}, \ldots, Y_*^{M-1}$ recursively from (15).
   - Draw $Z_*^M \sim \pi(Z_*^M|Y_*^{M-1}, x^M, \Theta_*)$.
   - Set $\Phi_* = (Z_*^j, Y_*^{j+1}, \ldots, Y_*^{M-1}, Z_*^M, \Theta_*)$ and put $\Phi_{(s)} = \Phi_*$ with probability $\min\{1, A\}$ where A is given by (29), otherwise put $\Phi_{(s)} = \Phi_{(s-1)}$.
   - Store $\Phi_{(s)}$.
3. *Pruning* - For $s = 1, \ldots, S$:
   - Discard all components of each $\Phi_{(s)}$ except $(\Theta_{(s)}, Z_{(s)}^M)$.
4. Set $j = j + m$ and return to step 2.

Thus step 2 performs the update for a given time point and we run our MCMC scheme as each observation becomes available, allowing on-line estimation of Θ. Further modifications may be made by thinning the MCMC output at the expense of running the sampler for longer; R iterations can be performed before thinning by a factor κ (such that $R = \kappa S$). This is done separately for each time point, with the final posterior sample of size S used as the prior for the next time point. Note that running our algorithm is no more computationally intensive than running an MCMC scheme for all

observations simultaneously. Computational cost of the simulation filter (and also of the single site and block Gibbs sampler) is reported in Section 5.4.

4.3 Convergence

At every iteration in step 2 of the simulation filter, the current path $Y^j, \ldots, Y^M$ is entirely consistent with the current parameter vector, since both the latent process and Θ are updated jointly. Hence the dependence between them is overcome. The appeal of using the modified bridge density to propose the path is two-fold. Firstly, (15) is Gaussian and therefore easy to simulate from and evaluate. Also, Chib & Shephard (2002) show that the estimate of the likelihood contribution under the modified bridge scheme is simply the Euler approximation of $\pi(Y^M|Y^j)$ times the expected value of a likelihood ratio of two predictive models.

Now let $\mathbf{Q}\{Y(m)|Y^j, x^{j+m}, \Theta\}$ denote the law of the stochastic process producing the proposed values $Y^{j+1}, \ldots, Y^{j+m-2}, Y^{j+m-1}$ in the finite interval $[t_j, t_{j+m}]$, and let $\mathbf{P}\{Y(m)|Y^j, x^{j+m}, \Theta\}$ denote the true process. As $m \to \infty$ (and hence $\Delta t \to 0$) the true likelihood of the underlying model is better approximated and

$$\frac{\prod_{k=j}^{j+m-1} \pi(Y_*^{k+1}|Y_*^k, \Theta_*)}{\prod_{k=j}^{j+m-2} \tilde{\pi}(Y_*^{k+1}|Y_*^K, x^{j+m}, \Theta_*)} \longrightarrow \frac{d\mathbf{P}_{Y|\Theta}}{d\mathbf{Q}_{Y|\Theta}}\{Y_*(m)|Y_*^j, x^{j+m}, \Theta_*\}, \tag{30}$$

the Radon-Nikodym derivative of the true process with respect to the proposal. Since the modified bridge is simply a discrete time approximation of the underlying diffusion, conditioned to start at Y^j and finish at x^{j+m}, (30) is non-singular as the volatility of the proposal matches that of the true process, in the limit as $m \to \infty$. Hence the acceptance probability (29) tends to a non-zero limit. For further discussion see Chib et al. (2004).

5 Stochastic Volatility

To illustrate our methodology, we examine a simple Log Gaussian Stochastic Volatility model of the form

$$\begin{aligned} dX_t &= \theta_1 X_t dt + X_t \exp(0.5 Z_t) dW_{1,t} \\ dZ_t &= (\theta_2 - \theta_3 Z_t) dt + \theta_4 dW_{2,t} \end{aligned} \tag{31}$$

where $dW_{1,t} \perp dW_{2,t}$ and $\theta_3, \theta_4 > 0$. Note that X_t represents stock price and Z_t corresponds to an unobserved volatility factor. Similar models have been examined by Andersen & Lund (1997), Eraker (2001) and Durham & Gallant (2002) whilst discrete time SV models have been examined by Kim, Shephard & Chib (1998), Stroud et al. (2004) and Shephard (2005). We now implement the (reducible) Gibbs sampling algorithms of Sections 3.1 and 3.2 and the (asymptotically irreducible) simulation filter of Section 4.

5.1 Gibbs Sampling

Expressing (31) in the form given by (1), we have

$$\mu(Y_t, \Theta) = \begin{pmatrix} \theta_1 X_t \\ \theta_2 - \theta_3 Z_t \end{pmatrix} , \beta(Y_t, \Theta) = \begin{pmatrix} X_t^2 \exp(Z_t) & 0 \\ 0 & \theta_4^2 \end{pmatrix}$$

where $\Theta = (\theta_1, \theta_2, \theta_3, \theta_4)'$. By assuming a standard non-informative prior for Θ, it is easily seen that $\pi(\Theta|\hat{\mathrm{Y}})$ can be factorised as

$$\pi(\Theta|\hat{\mathrm{Y}}) = \pi(\theta_1|\hat{\mathrm{Y}}) \pi(\theta_2, \theta_3, \theta_4|\hat{\mathrm{Y}})$$

where

$$\pi(\theta_1|\hat{\mathrm{Y}}) = \prod_{k=0}^{n-1} \phi\left(X^{k+1}\,;\, X^k + \theta_1 X^k \Delta t\,,\, (X^k)^2 \exp(Z^k) \Delta t\right)$$

and

$$\pi(\theta_2, \theta_3, \theta_4|\hat{\mathrm{Y}}) = \prod_{k=0}^{n-1} \phi\left(Z^{k+1}\,;\, Z^k + (\theta_2 - \theta_3 Z^k)\Delta t\,,\, \theta_4^2 \Delta t\right) .$$

As these expressions have forms similar to that of the likelihood function for linear regression problems, direct sampling of $\pi(\Theta|\hat{\mathrm{Y}})$ is possible (see Eraker (2001) for further details). We assume independent proper Uniform priors for θ_1, θ_2, $\log(\theta_3)$ and $\log(\theta_4)$ and sample $\pi(\Theta|\hat{\mathrm{Y}})$ using the Metropolis random walk update outlined in Section 3.4. Although sampling the full conditionals directly will result in parameter draws that are less autocorrelated (than when using the random walk update), mixing will still deteriorate for large m. Further, for more complicated models, full conditional distributions for the parameters are rarely available in closed form and methods such as the Metropolis random walk update considered here become important.

5.2 Simulation Filter

We now turn our attention to applying the simulation filter to the SV model. We first assume that at time t_j we have a sample $\{(\Theta_{(s)}, Z^j_{(s)}), s = 0, \ldots, S\}$ from $\pi_j(\Theta, Z^j)$. Then at time t_M, as the observation, x^M $(M = j + m)$ becomes available, we sample $\pi_M(\Theta, Z^M)$ via MCMC. We start by proposing Y_*^{k+1} for each $k = j, \ldots, M-2$ using recursive application of (15). Note that (15) can be simplified, as the SV model considered here is block diagonal and therefore the transition density, (4), can be factorised as

$$\pi(Y^{k+1}|Y^k, \Theta) = \pi(X^{k+1}|Y^k, \Theta)\pi(Z^{k+1}|Z^k, \Theta) \tag{32}$$

where

$$\pi(X^{k+1}|Y^k, \Theta) = \phi(X^{k+1}\,;\, \theta_1 X^k \Delta t\,,\, (X^k)^2 \exp(Z^k)\Delta t)\,, \tag{33}$$

$$\pi(Z^{k+1}|Z^k, \Theta) = \phi(Z^{k+1}\,;\, (\theta_2 - \theta_3 Z^k)\Delta t\,,\, \theta_4^2 \Delta t)\,. \tag{34}$$

Then, (15) can be written as

$$\tilde{\pi}(Y^{k+1}|Y^k, x^M, \Theta) = \tilde{\pi}(X^{k+1}|Y^k, x^M)\pi(Z^{k+1}|Z^k, \Theta)\,, \tag{35}$$

where

$$\tilde{\pi}(X^{k+1}|Y^k, x^M) = \phi\left(X^{k+1}\,;\, X^k + \left(\frac{x^M - X^k}{M-k}\right)\,,\, \left(\frac{M-k-1}{M-k}\right)\beta^{xx}(Y^k, \Theta)\Delta t\right) \tag{36}$$

and $\beta^{xx}(Y^k, \Theta) = (X^k)^2 \exp(Z^k)$.

Therefore, we can draw Z_*^{k+1}, (for $k = j, \ldots, M-1$) "blindly"; that is, given $(\Theta_*, Z_*^j) \sim \hat{\pi}_j(\Theta_*, Z_*^j)$, we simulate $Z_*^{k+1} \sim \pi(Z_*^{k+1}|Z_*^k, \Theta_*)$ using a recursive application of (34). Then for $k = j, \ldots, M-2$, we draw $X_*^{k+1} \sim \tilde{\pi}(X_*^{k+1}|Y_*^k, x^M)$ using (36). We then put $\Phi_{(s)} = \Phi_* = (Z_*^j, Y_*^{j+1}, \ldots, Y_*^{M-1}, Z_*^M, \Theta_*)$ with probability $\min\{1, A\}$ where A is given by (29). For the Stochastic Volatility model given by (31), the acceptance probability reduces to

$$A = \frac{\prod_{k=j}^{M-1} \pi(X_*^{k+1}|Y_*^k, \Theta_*)}{\prod_{k=j}^{M-1} \pi(X^{k+1}|Y^k, \Theta)} \times \frac{\prod_{k=j}^{M-2} \tilde{\pi}(X^{k+1}|Y^k, x^M)}{\prod_{k=j}^{M-2} \tilde{\pi}(X_*^{k+1}|Y_*^k, x^M)}, \tag{37}$$

Thus, after initialising with a sample from the prior, we run our MCMC scheme for each new time point, retaining $\{(\Theta_{(s)}, Z_{(s)}^j), s = 1, \ldots, S\}$ for all $j = m, 2m \ldots, n$.

5.3 An Empirical Application

To illustrate the methodology proposed in this paper, we present estimates of the parameters in the simple Log Gaussian Stochastic Volatility model. We use data consisting of 1509 weekly observations on the three-month U.S. Treasury bill rate (August 6, 1967 - August 30, 1996). The data were obtained from the Federal Reserve's weekly H.15 reports of market data (see URL *http://www.federalreserve.gov/releases/h15/data /wf/tbsm3m.txt*) and are plotted in Fig. 1.

[Figure 1 about here.]

We implement the Gibbs sampling algorithms and the simulation filter using the sampling strategies outlined in Sections 3.1, 3.2 and 4; both the single site Gibbs sampler and block Gibbs sampler are run for 7.5 million iterations with the first 2.5 million discarded as burn-in, thinned by a factor of 500, and the remaining 10,000 iterations are then used for the main monitoring runs. The simulation filter is run for 1 million iterations with a thin of 100 (to give $S = 10,000$ particles at each time point). Discretization bias is set by putting $m = 5$ (and therefore $\Delta t = 0.2$) giving a total of 13573 latent variables. Results for varying m are reported in Section 5.4.

The M-H scheme of Section 3.4 requires specification of the tuning parameters, γ. Although heuristic automated adaption procedures are possible, in this example, setting $\gamma = (0.0001, 0.005, 0.005, 0.005)'$ seems to produce satisfactory results. Table 1, Fig. 2, Fig. 3 and Fig. 4 provide posterior summaries obtained from each MCMC sheme.

[Table 1 about here.]

[Figure 2 about here.]

[Figure 3 about here.]

[Figure 4 about here.]

Naturally, the single site Gibbs sampler, block Gibbs sampler and simulation filter have the same unique equilibrium distribution and running each scheme until convergence will yield identical parameter estimates. For the U.S. T-bill data plotted in Fig. 1, both schemes lead to parameter estimates that are consistent with one another (though the Gibbs sampling algorithms both require a much longer run than the simulation filter).

As the (estimated) MCMC error is related to the autocorrelations within the chains, a useful way of assessing the relative performance of the three algorithms is to study the sample autocorrelation functions for each parameter. Fig. 4 shows the benefits of the simulation filter over each Gibbs strategy; in Fig. 2 and Fig. 3, the sampled values of θ_2, θ_3 and θ_4 are highly autocorrelated (though the blocking strategy seems to produce slightly less autocorrelated values than the single site Gibbs strategy) where as in Fig. 4 we see a clear reduction of the autocorrelations for each component of Θ despite a much shorter run.

5.4 Simulation Study

To validate the proposed MCMC scheme, this section presents evidence on the performance of the estimator of Θ in the SV model using synthetic data. Data were simulated from the SV model with $\theta_1 = 0.001$, $\theta_2 = -0.6$, $\theta_3 = 0.08$ and $\theta_4 = 0.5$ (calibrated to match the U.S. short-term interest rate) by using the Euler scheme with a sample

time interval of length 0.001. Every 1000th point was recorded to give a sample path of 500 observations on the interval $[0, 499]$.

For $m = 2, 10, 15$, we run the single site Gibbs sampler and the block Gibbs sampler for 5.5 million iterations with tuning parameters, $\gamma = (0.0001, 0.005, 0.005, 0.005)'$, thin by a factor 500 and discard the first 500,000 iterations as burn-in giving 10,000 iterations as the main monitoring run. We provide a much longer runs for $m = 5$; 10 million iterations with a thin of 1000. The simulation filter is run for $m = 2, 5, 10$ and 15 for 1 million iterations with a thin of 100 (giving $S = 10,000$ particles at each time point). Note that each algorithm is coded in C and executed on a Pentium IV 1.8 GHz processor. Computational times for the single site Gibbs sampler, block Gibbs sampler and simulation filter for a dataset of length 500, with $m = 15$ and 1 million iterations are 1370, 1410 and 399 minutes respectively. It is also important to bear in mind that should a new observation become available, both Gibbs sampling strategies must be started from scratch whereas the simulation need only be run for a further 48 seconds to assimilate the new information.

[Figure 5 about here.]

[Figure 6 about here.]

[Table 2 about here.]

[Table 3 about here.]

[Table 4 about here.]

Figures 5-6 and Tables 2-4 summarise the posterior distribution; Table 2 gives posterior means, medians and 95% probability intervals for Θ estimated from a single run of the single site Gibbs sampler. Similar descriptive statistics are provided in Tables 3 and 4, from the output of the block Gibbs sampler and the simulation filter. We see that all samplers produce estimates that are close to the true parameters that generated the sample data (given the estimated MCMC error). Fig. 5 shows that running each scheme until convergence yields estimates that are consistent with each other (though both Gibbs strategies required a much longer run than the simulation filter). Furthermore, each table demonstrates the clear advantage of including latent

variables in the estimation framework. As m increases, there is a notable decrease in discretization bias. For example, θ_2 has a true value of -0.6 while in Table 3 the simulation filter yields estimates of -0.5325, -0.5953 and -0.5999 for $m = 2, 10$ and 15 respectively. Note that there is little difference in results for $m = 10$ and $m = 15$.

Fig. 6 shows the advantage of the simulation filter. As we increase m, the single site Gibbs sampler and block Gibbs sampler both give a marked increase in the autocorrelations of θ_2 where as the autocorrelation plots for the simulation filter die down very quickly for all values of m. In addition, we find that there is little difference between the autocorrelations obtained under the single site scheme and those obtained under the block Gibbs scheme. Note that the comparison is essentially qualitative rather than quantitative and there are a number of ways in which each algorithm can be improved. The comparison is simply intended to illustrate the inherent problems with a Gibbs sampling approach, and that the simulation filter overcomes these.

6 Discussion

In this paper, we have provided a sequential, simulation-based approach to the problem of parameter estimation of partially and discretely observed multivariate diffusion processes.

By considering the analysis of a discretely observed SDE as a classic missing data problem (Pedersen 1995), inference can be problematic due to high dependence between the parameters and missing data (Roberts & Stramer 2001). In fact, as either m (the amount of augmentation) or the number of observations becomes large, we see arbitrarily slow rates of convergence of basic algorithms such as the Gibbs sampler considered in section 3. Further, if new data should arrive, conventional MCMC samplers must be started from scratch in order to obtain a new sample of parameter values.

Our simulation filter allows on-line estimation of the model parameters, Θ, which is an essential requirement for financial models where data arrives almost continuously. As each new observation arrives our algorithm overcomes the dependency between Θ and the latent data by simulating the latent data to be consistent with each proposed value of Θ.

Applications of the methodology included a Stochastic Volatility (SV) model, though

the simulation filter can be easily applied to general nonlinear multivariate diffusions. In particular, it should be noted that our methodology does not rely on the block-diagonal nature of the diffusion matrix in the SV example. Applications of the simulation filter to multivariate diffusions which are not block-diagonal can be found in Golightly & Wilkinson (2006).

Improving the efficiency of the simulation filter remains of great interest. For example, using a kernel density estimate to avoid sample impoverishment is somewhat ad-hoc. Although there appears to be no easy solution, this is the subject of ongoing research.

References

Amit, Y. (1991), 'On rates of convergence of stochastic relaxation for Gaussian and non-Gaussian distributions', *Journal of Multivariate Analysis* **38**, 82–99.

Andersen, T. G. & Lund, J. (1997), 'Estimating continuous-time stochastic volatility models of the short-term interest rate', *Journal of Econometrics* **77**, 343–377.

Berzuini, C., Best, N. G., Gilks, W. R. & Larizza, C. (1997), 'Dynamic conditional independence models and Markov chain Monte Carlo methods', *Journal of the American Statistical Association* **92**(440), 1403–1412.

Carpenter, J., Clifford, P. & Fearnhead, P. (1999), 'An improved particle filter for nonlinear problems', *IEE Procedings - Radar, Sonar and Navigation* **146**, 2–7.

Chan, K. C., Karolyi, G. A., Longstaff, F. A. & Sanders, A. B. (1992), 'An empirical comparison of alternative models of the short-term interest', *Journal of Finance* **47**, 1209–1228.

Chib, S. & Greenberg, E. (1995), 'Understanding the Metropolis-Hastings algorithms', *The American Statistician* **49**, 327–335.

Chib, S. & Shephard, N. (2002), 'Comment on 'Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes", *Journal of Business and Economic Statistics* **20**, 279–316.

Chib, S., Pitt, M. K. & Shephard, N. (2004), 'Likelihood based inference for diffusion driven models', *In submission*.

Del Moral, P. & Jacod, J. (2001), Interacting particle filtering with discrete observations, *in* A. Doucet, N. de Freitas & N. Gordon, eds, 'Sequential Monte Carlo Methods in Practice', Springer Verlag.

Del Moral, P., Jacod, J. & Protter, P. (2002), 'The Monte Carlo method for filtering with discrete-time observations', *Probability Theory and Related Fields* **120**, 346–368.

Doucet, A., Godsill, S. & Andrieu, C. (2000), 'On sequential Monte Carlo sampling methods for Bayesian filtering', *Statistics and Computing* **10**, 197–208.

Durham, G. B. & Gallant, R. A. (2002), 'Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes', *Journal of Business and Economic Statistics* **20**, 279–316.

Elerian, O., Chib, S. & Shephard, N. (2001), 'Likelihood inference for discretely observed nonlinear diffusions', *Econometrica* **69**(4), 959–993.

Eraker, B. (2001), 'MCMC analysis of diffusion models with application to finance', *Journal of Business and Economic Statistics* **19**, 177–191.

Gamerman, D. (1997), *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, Texts in Statistical Science, Chapman and Hall, London.

Golightly, A. & Wilkinson, D. J. (2005), 'Bayesian inference for stochastic kinetic models using a diffusion approximation', *Biometrics* **61**(3), 781–788.

Golightly, A. & Wilkinson, D. J. (2006), 'Bayesian sequential inference for stochastic kinetic biochemical network models', *Journal of Computational Biology* **13**(3), 838–851.

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

Handschin, J. E. & Mayne, D. Q. (1969), 'Monte Carlo techniques to estimate the conditional expectation in multi-stage non-linear filtering', *Journal of Control* **9**, 547–559.

Johannes, M. S., Polson, N. G. & Stroud, J. R. (2006), Optimal filtering of jump diffusions: Extracting latent states from asset prices.

Jones, C. (1997), A simple Bayesian approach to the analysis of Markov diffusion processes, Technical report, The Wharton School, University of Pennsylvania.

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

Liu, J. & West, M. (2001), Combined parameter and state estimation in simulation-based filtering, *in* A. Doucet, N. de Freitas & N. Gordon, eds, 'Sequential Monte Carlo Methods in Practice', Springer-Verlag.

Øksendal, B. (1995), *Stochastic Differential Equations: An Introduction with Applications*, 6th edn, Springer-Verlag.

Papaspiliopolous, O., Roberts, G. O. & Skôld, M. (2003), Non-centered parameterisations for hierarchical models and data augmentation, *in* J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith & M. West, eds, 'Bayesian Statistics 7', Oxford University Press, pp. 307–326.

Pedersen, A. (1995), 'A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations', *Scandinavian Journal of Statistics* **22**, 55–71.

Pitt, M. K. & Shephard, N. (1999), 'Filtering via simulation: Auxiliary particle filters', *Journal of the American Statistical Association* **446**(94), 590–599.

Roberts, G. O. & Stramer, O. (2001), 'On inference for partially observed nonlinear diffusion models using the Metropolis-Hastings algorithm', *Biometrika* **88**(4), 603–621.

Shephard, N. (2005), Are there discontinuities in financial prices?, *in* A. Davison, Y. Dodge & N. Wermuth, eds, 'Celebrating Statistics: Papers in honour of Sir David Cox on his 80th birthday', Oxford University Press, pp. 213–231.

Shephard, N. & Pitt, M. K. (1997), 'Likelihood analysis of non-Gaussian measurement time series', *Biometrika* **84**, 653–667.

Silverman, B. W. (1986), *Density Estimation for Statistics and Data Analysis*, Chapman and Hall: London.

Stroud, J. R., Polson, N. G. & Muller, P. (2004), Practical filtering for stochastic volatility models, *in* A. Harvey, S. J. Koopman & N. Shephard, eds, 'State Space and Unobserved Components Models', Cambridge Press, pp. 236–247.

Tanner, M. A. & Wong, W. H. (1987), 'The calculation of posterior distributions by data augmentation', *Journal of the American Statistical Association* **82**(398), 528–540.

West, M. (1993), 'Approximating posterior distributions by mixtures', *Journal of Royal Statistical Society, Series B: Statistical Methodology* **55**, 409–422.

Wilkinson, D. J. (2003), Disscussion to 'Non centred parameterisations for hierarchical models and data augmentation' by Papaspiliopoulos, Roberts and Skold, *in* 'Bayesian Statistics 7', Oxford Science Publications, pp. 323–324.

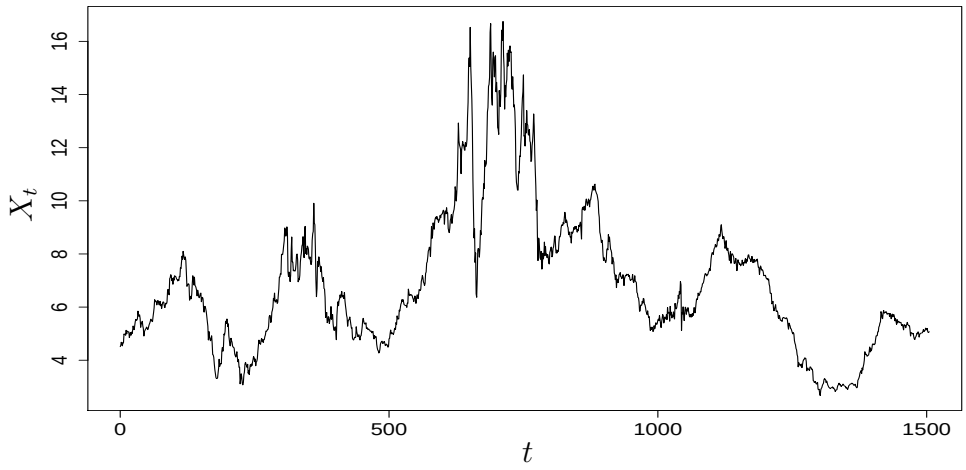

Figure 1: Weekly observations of the 3-month Treasury bill rate, 06/10/1967 – 30/08/1996, $n = 1509$.

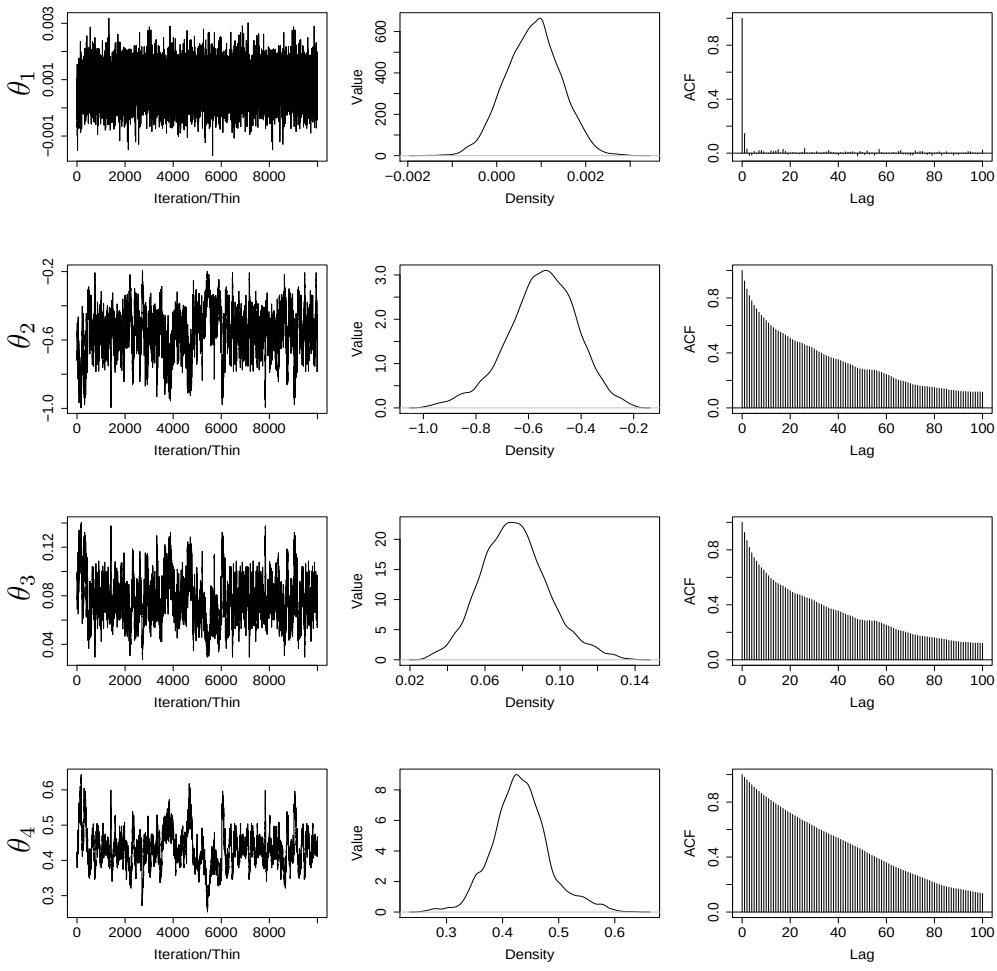

Figure 2: Trace, Density and Autocorrelation plots for Θ, obtained from the output of the single site Gibbs sampler using weekly observations of the 3 month Treasury bill rate, Oct 6, 1967 - Aug 30, 1996.

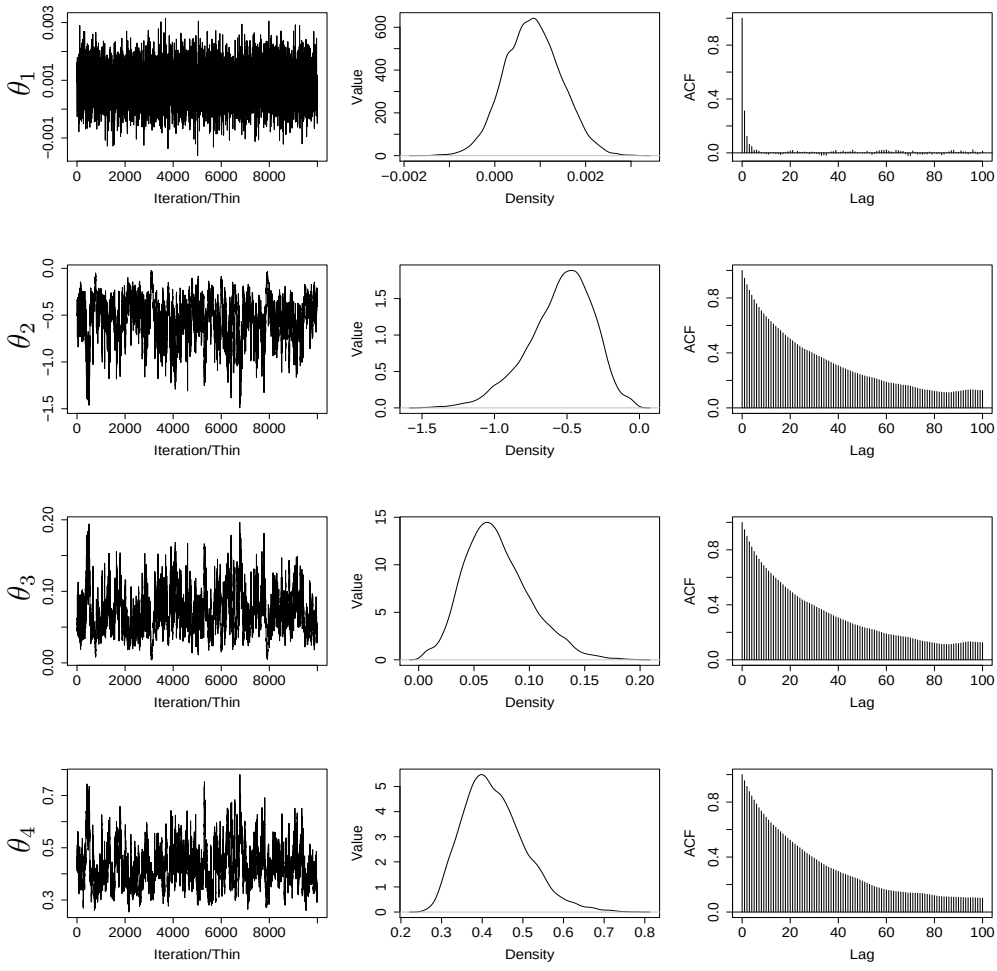

Figure 3: Trace, Density and Autocorrelation plots for Θ, obtained from the output of the block Gibbs sampler using weekly observations of the 3 month Treasury bill rate, Oct 6, 1967 - Aug 30, 1996.

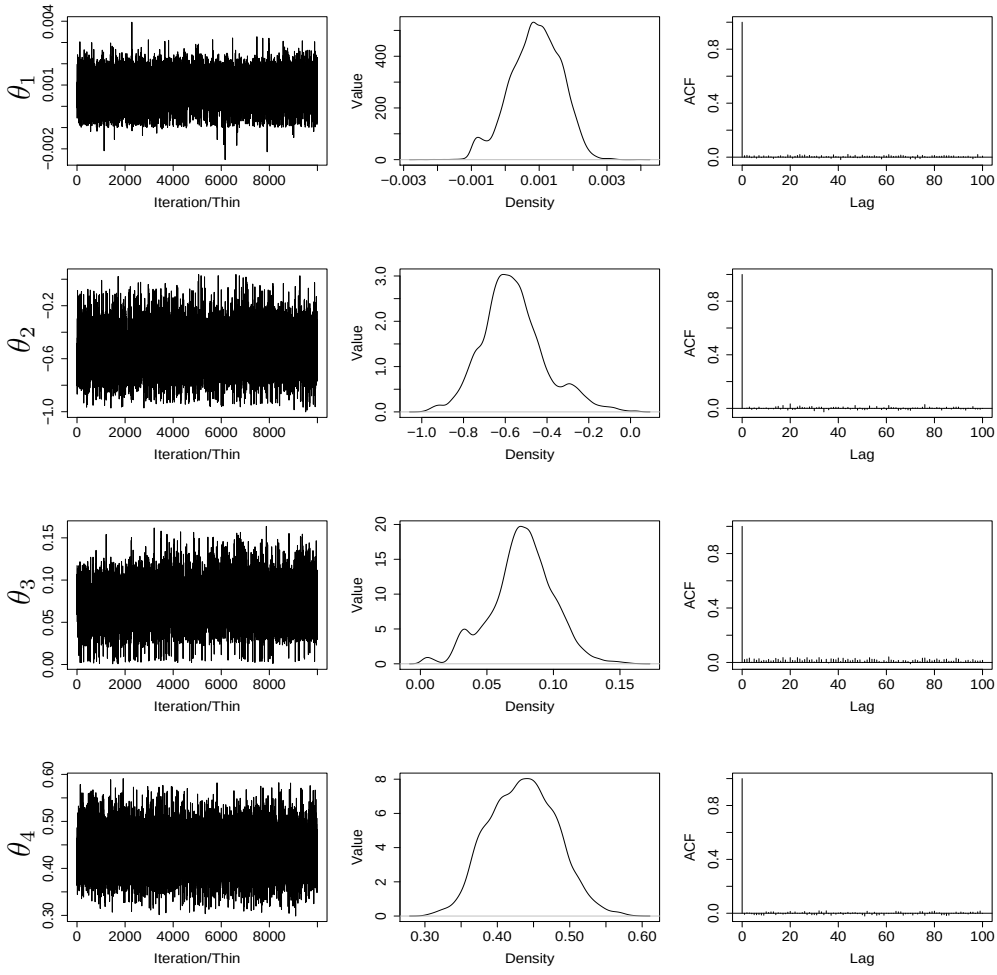

Figure 4: Trace, Density and Autocorrelation plots for Θ, obtained from the output of the simulation filter using weekly observations of the 3 month Treasury bill rate, Oct 6, 1967 - Aug 30, 1996.

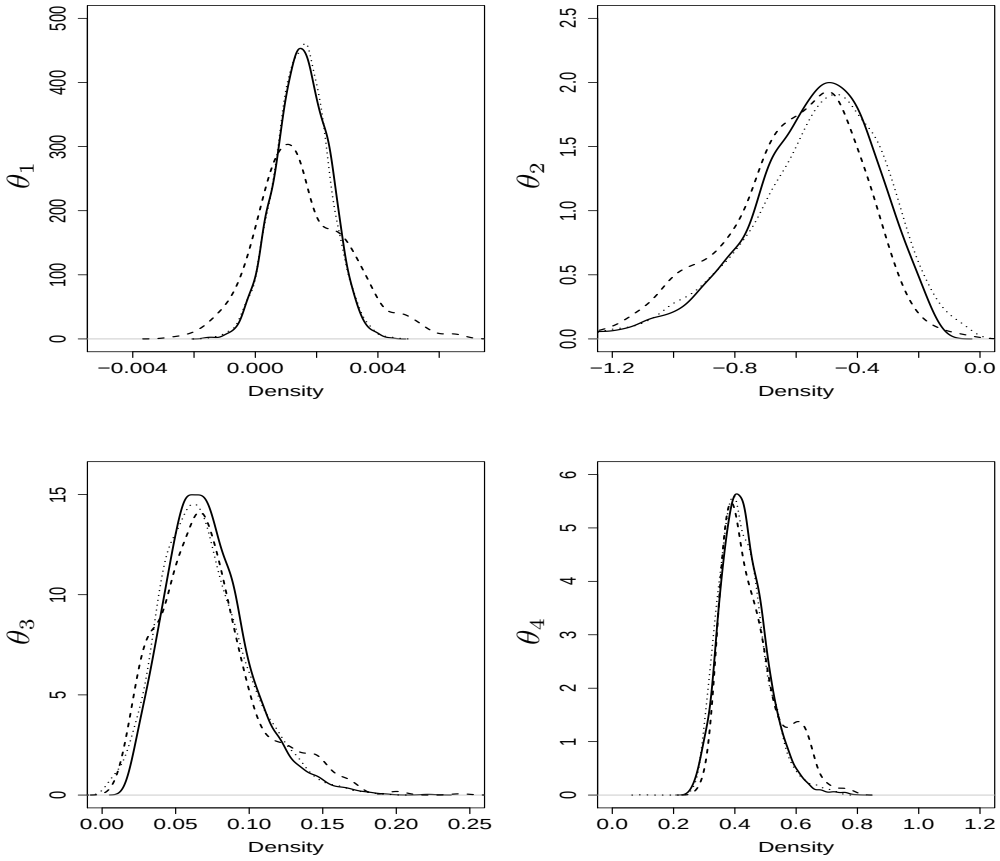

Figure 5: Density plots for Θ (estimated on 500 simulated observations) obtained from the output of 10 million iterations (with a thin of 1000) using the single site Gibbs sampler (solid line), the block Gibbs sampler (dotted line) and the simulation filter with 1 million iterations and a thin of 100 (dashed line). $m = 5$ in all cases.

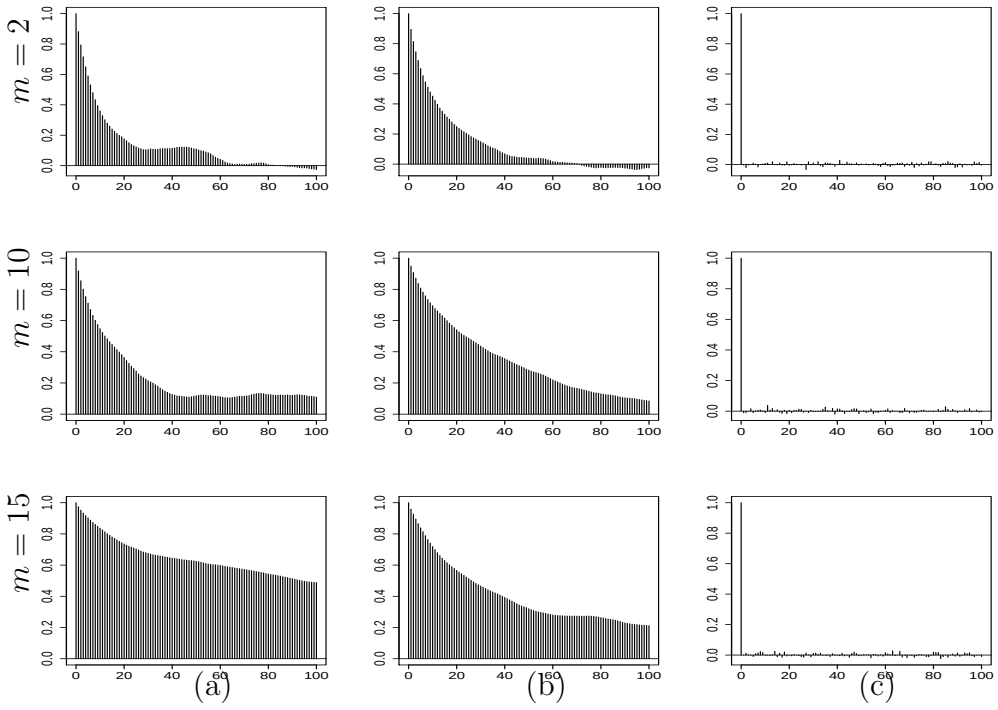

Figure 6: Autocorrelation plots for θ_2 and increasing m (estimated on 500 simulated observations), obtained from the output of the final 5.5 million iterations of a single run of 5.5 million iterations (with a thin of 500) from (a) the single site Gibbs sampler, (b) the block Gibbs sampler and (c) the simulation filter with 1 million iterations and a thin of 100.

	θ_1	θ_2	θ_3	θ_4
		Single site Gibbs sampler		
Mean	0.0008	-0.5489	0.0759	0.4319
Median	0.0008	-0.5423	0.0750	0.4300
95% Interval	(-0.0004,0.0020)	(-0.8421,-0.2883)	(0.0388,0.1154)	(0.3375,0.5498)
		Block Gibbs sampler		
Mean	0.0008	-0.5444	0.0710	0.4303
Median	0.0008	-0.5171	0.0674	0.4210
95% Interval	(-0.0003,0.0021)	(-1.0481,-0.1709)	(0.0219,0.1366)	(0.3061,0.6053)
		Simulation Filter		
Mean	0.0008	-0.5505	0.0766	0.4354
Median	0.0008	-0.5644	0.0774	0.4335
95% Interval	(-0.0008,0.0022)	(-0.8311,-0.2188)	(0.0280,0.1212)	(0.3513,0.5261)

Table 1: Posterior means, medians and 95% posterior probability intervals for Θ under the 3 MCMC schemes. Parameters are estimated using weekly observations of the 3 month Treasury bill rate, Oct 6, 1967 - Aug 30, 1996.

	θ_1	θ_2	θ_3	θ_4
		True Values		
	0.0010	-0.6000	0.0800	0.5000
		$m = 2$		
Mean	0.0015	-0.5073	0.0662	0.4163
Median	0.0015	-0.4955	0.0648	0.4119
95% Interval	(-0.0003,0.0032)	(-0.9014,-0.1865)	(0.0292,0.1177)	(0.2956,0.5510)
		$m = 5$		
Mean	0.0015	-0.5505	0.0718	0.4285
Median	0.0015	-0.5238	0.0687	0.4222
95% Interval	(-0.0002,0.0031)	(-1.0499,-0.2123)	(0.0277,0.1352)	(0.3004,0.5926)
		$m = 10$		
Mean	0.0015	-0.5807	0.0792	0.4297
Median	0.0015	-0.5687	0.0777	0.4272
95% Interval	(-0.0002,0.0034)	(-0.9015,-0.1820)	(0.0221,0.1186)	(0.3020,0.5741)
		$m = 15$		
Mean	0.0015	-0.5681	0.0776	0.4426
Median	0.0015	-0.5526	0.0760	0.4225
95% Interval	(-0.0003,0.0033)	(-0.8976,-0.1793)	(0.0227,0.1278)	(0.3110,0.5835)

Table 2: Posterior means, medians and 95% posterior probability intervals for Θ (estimated on 500 simulated observations), obtained from the output of the single site Gibbs sampler. For $m = 2, 10, 15$, results are based on the final 5 million iterations of a single run of 5.5 million (with a thin of 500). For $m = 5$, results are based on 10 million iterations (with a thin of 1000).

	θ_1	θ_2	θ_3	θ_4
		True Values		
	0.0010	-0.6000	0.0800	0.5000
		$m = 2$		
Mean	0.0015	-0.5385	0.0703	0.4246
Median	0.0015	-0.5234	0.0682	0.4182
95% Interval	(-0.0003,0.0032)	(-0.9628,-0.19959)	(0.0252,0.1267)	(0.3057,0.5813)
		$m = 5$		
Mean	0.0015	-0.5319	0.0694	0.4254
Median	0.0015	-0.5048	0.0657	0.4147
95% Interval	(-0.0002,0.0032)	(-1.0539,-0.1600)	(0.0203,0.1370)	(0.3036,0.6051)
		$m = 10$		
Mean	0.0015	-0.5300	0.0692	0.4187
Median	0.0015	-0.5029	0.0655	0.4096
95% Interval	(-0.0002,0.0033)	(-1.0360,-0.1940)	(0.0246,0.1347)	(0.3030,0.5822)
		$m = 15$		
Mean	0.0015	-0.5497	0.0718	0.4298
Median	0.0015	-0.5369	0.0700	0.4208
95% Interval	(-0.0002,0.0033)	(-0.9690,-0.1879)	(0.0236,0.1266)	(0.3242,0.5713)

Table 3: Posterior means, medians and 95% posterior probability intervals for Θ (estimated on 500 simulated observations), obtained from the output of the block Gibbs sampler. For $m = 2, 10, 15$, results are based on the final 5 million iterations of a single run of 5.5 million (with a thin of 500). For $m = 5$, results are based on 10 million iterations (with a thin of 1000).

	θ_1	θ_2	θ_3	θ_4
		True Values		
	0.0010	-0.6000	0.0800	0.5000
		$m = 2$		
Mean	0.0014	-0.5325	0.0676	0.3974
Median	0.0015	-0.5343	0.0688	0.3991
95% Interval	(-0.0008,0.0038)	(-0.8729,-0.2004)	(0.0285,0.1085)	(0.2502,0.5767)
		$m = 5$		
Mean	0.0016	-0.6113	0.0714	0.4522
Median	0.0014	-0.5719	0.0672	0.4301
95% Interval	(-0.0010,0.0024)	(-1.1113,-0.2594)	(0.0209,0.1513)	(0.3305,0.6567)
		$m = 10$		
Mean	0.0015	-0.5953	0.0844	0.4752
Median	0.0014	-0.6208	0.0865	0.4745
95% Interval	(-0.0010,0.0040)	(-0.9617,-0.1449)	(0.0227,0.1380)	(0.2949,0.6637)
		$m = 15$		
Mean	0.0014	-0.5999	0.0854	0.4526
Median	0.0014	-0.6017	0.0842	0.4275
95% Interval	(-0.0010,0.0050)	(-0.8750,-0.1353)	(0.0171,0.1533)	(0.2614,0.7315)

Table 4: Posterior means, medians and 95% posterior probability intervals for Θ (estimated on 500 simulated observations), obtained from the output of the simulation filter. All results are based on a single run of 1 million iterations with a thin of 100.