Likelihood based inference for diffusion driven models

SIDDHARTHA CHIB

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

`chib@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`

August 2004

Abstract

This paper provides methods for carrying out likelihood based inference for diffusion driven models, for example discretely observed multivariate diffusions, continuous time stochastic volatility models and counting process models. The diffusions can potentially be non-stationary. Although our methods are sampling based, making use of Markov chain Monte Carlo methods to sample the posterior distribution of the relevant unknowns, our general strategies and details are different from previous work along these lines. The methods we develop are simple to implement and simulation efficient. Importantly, unlike previous methods, the performance of our technique is not worsened, in fact it improves, as the degree of latent augmentation is increased to reduce the bias of the Euler approximation. In addition, our method is not subject to a degeneracy that afflicts previous techniques when the degree of latent augmentation is increased. We also discuss issues of model choice, model checking and filtering. The techniques and ideas are applied to both simulated and real data.

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

1.1 The diffusion driven class

Diffusions play a substantial role in modern economics and finance. Recently there has been a significant increase in interest in methods for estimating and testing parametric diffusion models using discrete data, such as daily or monthly observations. In this paper we show how to carry out likelihood inference on a problem where the observations are driven by an underlying diffusion. This *diffusion driven* class covers a large number of models which are frequently employed in financial economics.

We suppose we observe Y_i at a non-stochastic time τ_i for $i = 1, 2, ..., n$, where we have ordered the data in time so that $0 \leq \tau_1 \leq \tau_2 \leq ... \leq \tau_n \leq T$. We assume that Y is related to

an underlying $d-$dimensional continuous time process $\alpha(t)$ for $t \geq 0$. Our central assumption is that conditionally on the sample path of α, then Y is Markovian. This is written abstractly as

$$(Y_{i-j} \perp\!\!\!\perp Y_{i+1}) \,|\, Y_i, \{\alpha(t); t \in [0, T]\}\,,$$

where $j > 0$ and $\perp\!\!\!\perp$ is the symbol for independence. The model is completed by assuming α is a diffusion which satisfies the stochastic differential equation (SDE)

$$d\alpha(t) = \mu\{\alpha(t); \theta\}dt + \Theta\{\alpha(t); \theta\}dW(t), \quad t \geq 0, \tag{1}$$

where μ and Θ are drift and volatility functions, respectively, depending on α and an unknown parameter vector θ, and $W = (W_1, ..., W_d)'$ is a $d-$dimensional vector of independent standard Brownian motions. Sometimes we write $\Sigma = \Theta\Theta'$ to ease the exposition. We also assume that μ satisfies the Novikov condition (e.g. Karatzas and Shreve (1991, pp. 198–201)).

The diffusion driven class covers many of the common diffusion based models in financial economics. Here we list some examples and discuss the variety of methods researchers have developed to handle these models.

Example 1 Discretely observed multivariate diffusions. $Y_i = \alpha(\tau_i)$, which means that the multivariate Y_i are Markovian. Such models are often used to model interest rates in various ways. The econometric difficulty with this class of models is that we do not in general know the distribution of $Y_{i+1}|Y_i$. In the 1990s the task of estimating parametric non-linear SDEs prompted the development of some original moments based strategies for these models (see, for example, the reviews by Gallant and Tauchen (2004), Bibby, Jacobsen, and Sørensen (2004) and Aït-Sahalia, Hansen, and Scheinkman (2004)). Although such methods are attractive, in principle it is natural to base parametric inference for diffusions on the likelihood function. This is not straightforward as the likelihood is typically not immediately available, although the density of $Y_{i+1}|Y_i$ can be computed using numerically intensive methods (e.g. Pedersen (1995), Elerian, Chib, and Shephard (2001), Durham and Gallant (2002), Aït-Sahalia (2002), Aït-Sahalia (2003)). This paper develops novel simulation based Bayesian inference methods to optimally learn about diffusions through the likelihood function. These methods rely on the idea of data augmentation, which was introduced in this context by Elerian, Chib, and Shephard (2001), Roberts and Stramer (2001) and Eraker (2001), and is explained in detail in the next Section. Related likelihood based approaches are outlined by Durham (2003), Brandt and Santa-Clara (2002), Nicolau (2002) and Hurn, Lindsay, and Martin (2003).

Example 2 Stochastic volatility (e.g. Ghysels, Harvey, and Renault (1996) and Shephard (2004)). Suppose a univariate log-price process P follows the following diffusive stochastic

volatility (SV) process

$$\mathrm{d}P(t) = \left\{\mu_P + \beta\sigma^2(t)\right\}\mathrm{d}t + \sigma(t)\mathrm{d}B(t), \tag{2}$$

where $\sigma^2(t)$ is some function of the univariate $\alpha(t)$. We assume that B and W are standard Brownian motions with correlation coefficient ρ and that we record $Y_i = P(\tau_i)$. Then

$$Y_{i+1}|Y_i, \{\alpha(t); t \in [0, T]\} \sim N(Y_i + \mu_P(\tau_{i+1} - \tau_i) + \beta\sigma_{i+1}^2 + \rho Z_{i+1}, (1 - \rho^2)\sigma_{i+1}^2),$$

where

$$\sigma_{i+1}^2 = \int_{\tau_i}^{\tau_{i+1}} \sigma^2(u)\mathrm{d}u, \quad Z_{i+1} = \int_{\tau_i}^{\tau_{i+1}} \sigma(u)\mathrm{d}W(u).$$

Multivariate and multifactor SV models can also be written in this setup. The problem of carrying out for this kind of hidden factor model was an early motivation for the classic papers on indirect inference and efficient method of moments by Smith (1993), Gourieroux, Monfort, and Renault (1993) and Gallant and Tauchen (1996). The use of Markov chain Monte Carlo methods to carry out inference on Euler approximations to this type of model was developed by, for example, Jacquier, Polson, and Rossi (1994) and Kim, Shephard, and Chib (1998), while the idea of using these methods to overcome the effect of the discretisation bias was discussed by Kim, Shephard, and Chib (1998) and Eraker (2001).

Example 3 Counting process. Suppose N is a one dimensional counting process obtained by time changing a standard, homogeneous Poisson process N^*. We assume the time change is of the form $\int_0^t \lambda(u)\mathrm{d}u$ where we assume $\lambda \perp\!\!\!\perp N^*$ and λ is some function of the univariate α. Then

$$\mathrm{E}(N(t + dt) - N(t)|\alpha(t)) = \lambda(t)\mathrm{d}t,$$

so we can think of λ as the spot intensity of N. N is a Cox (1955) process (they are also called doubly stochastic processes) and provides a viable alternative to the Engle and Russell (1998) and Bowsher (2003) approaches to modelling the dynamics of the times of a sequence of events. They are used as the basis for much of the empirically realistic modelling developed in the credit risk literature. See for example Lando (1998) and Duffie and Singleton (1999). We show that likelihood based inference can be based on either $Y_i = N(\tau_i)$ or on the entire sample path of N, as in Section 5.1.

1.2 Augmentation and inference

Our approach to inference is Bayesian, so we start with a prior on θ, which we combine with the likelihood of the observations $Y_1, ..., Y_n|\theta$ to produce a posterior

$$\theta|Y_1, ..., Y_n. \tag{3}$$

Unfortunately the likelihood function for diffusion driven model is not known, except for in the simplest linear models. We side step this issue by employing augmentation. This approach emerges from the work of Elerian, Chib, and Shephard (2001), Eraker (2001) and Roberts and Stramer (2001), but our strategy and details are different.

In principle we augment θ with the entire path of α from time 0 to T. Then our approach is to simulate from

$$\theta, \{\alpha(t); t \in [0, T]\} | Y_1, ..., Y_n. \tag{4}$$

If we do this many times and just record the values of θ, we get a sample from (3) and thereby estimates of any posterior quantity of interest, e.g. posterior means, quantiles and covariances.

We show how it is possible to design Markov chain Monte Carlo (MCMC) samplers which draw from (4). For an extensive discussion of MCMC methods see the review of Chib (2001). Our methods are based on a block strategy of iterating through the following basic algorithm. Later we develop a straightforward reparameterisation of the blocking method in order to overcome a degeneracy in Algorithm 1.2. However, for the moment we ignore this issue.

Algorithm 1.2.

1. Sample from $\{\alpha(t); t \in [0, T]\} | \theta, Y_1, ..., Y_n$ by updating the subsets of the sample path.

   (a) Randomly split time from 0 to T into $K + 1$ sections. We write these *subsampling times* as

   $$0 \leq t_1 \leq t_2 \leq ... t_K \leq T.$$

   (b) For $i = 1, ..., K - 1$ sample the *subpath*

   $$\{\alpha(t); t \in [t_i, t_{i+1}]\} | \theta, Y_1, ..., Y_n, \alpha(t_i), \alpha(t_{i+1}). \tag{5}$$

   (c) Sample

   $$\{\alpha(t); t \in [0, t_1]\} | \theta, Y_1, ..., Y_n, \alpha(t_1)$$

   and

   $$\{\alpha(t); t \in [t_K, T]\} | \theta, Y_1, ..., Y_n, \alpha(t_K).$$

2. Draw from

   $$\theta | \{\alpha(t); t \in [0, T]\}.$$

3. Goto 1.

4

The resulting draws obey a Markov chain whose equilibrium marginal distribution is (4). All inferences are based on these sequences, beyond a suitable burn-in. Of course, the draws are serially correlated and therefore not as informative as i.i.d. draws from (4). We measure this dependence by the so-called inefficiency factor (autocorrelation time) of each posterior estimate. This measure, written $INF(L)$, is defined as $1 + 2\sum_{i=1}^{L}\rho(i)$, where $\rho(i)$ is the autocorrelation at lag i and L is a truncation point. See also Geweke (1989) who prefers to report the inverse of this number. 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 $INF(L)$ times as many iterations, beyond the transient phase of the Markov chain.

In implementing this strategy the main issue is how to sample from variables of the type (5).

1.3   Outline of the paper

In Section 2 we proceed to a discussion of the implementation of the augmentation approach through an Euler discretisation and illustrate the effectiveness of the method via an example. Following this we move onto methodology in Section 2.4 for partially observed diffusions, the principle example being the stochastic volatility model. The method for parameter sampling is introduced in Section 3. Simulation experiments are conducted to see how discretisation effects the performance of the algorithms. The application to volatility model with leverage for the Standard and Poor's index is examined in Section 4. Discussion of some related literature as well as extensions of our work are given in Section 5. Concluding remarks are in Section 6.

2   Augmentation and inference

2.1   The basic framework

In this Section we show how to sample from (5). To do this we first think about solving the slightly simpler problem of sampling the subpath from

$$\{\alpha(t); t \in [t_i, t_{i+1}]\} \,|\theta, \alpha(t_i), \alpha(t_{i+1}), \tag{6}$$

which is a multivariate non-linear bridge diffusion. This removes the data from (5). To ease the notation we suppress the dependence on θ in this Section.

Remark 1 *The case of discretely observed multivariate diffusions $Y_i = \alpha(\tau_i)$ can be immediately handled directly through (6) by a tactical choice of the subsampling time $\{t_i; \ i = 1, 2, ..., K\}$. All we have to do is to ensure that*

$$\{\tau_i; \ i = 1, 2, ..., n\} \subseteq \{t_i; \ i = 1, 2, ..., K\}. \tag{7}$$

The reason for this is that (5) simplifies to (6) when (7) holds. This strategy of choosing $\{t_i; \ i = 1, 2, ..., K\}$ to ensure (7) holds is, implicitly, an important simplifying feature of the approach used by Elerian, Chib, and Shephard (2001) and Roberts and Stramer (2001).

Carrying out the simulation from (6) directly is difficult due to the fact that we are conditioning on $\alpha(t_{i+1})$. Instead, we simulate from a rather similar process relatively easily — rejecting some of these proposals in order to correct for the resulting error. Consider the alternative diffusion α^*, which is constructed to have the following four properties

- It only exists on the time interval $[t_i, t_{i+1}]$.

- It starts at $\alpha^*(t_i) = \alpha(t_i)$.

- It finishes $\alpha^*(t_{i+1}) = \alpha(t_{i+1})$.

- It has the same volatility function as the α process.

We are rather free to select the drift function of α^*, so we use the simple form

$$\begin{aligned} \mathrm{d}\alpha^*(t) &= \{t_{i+1} - t\}^{-1} \{\alpha(t_{i+1}) - \alpha^*(t)\} \, \mathrm{d}t + \Theta \{\alpha^*(t)\} \, \mathrm{d}W(t) & (8) \\ &= \mu^*\{\alpha^*(t)\}\mathrm{d}t + \Theta \{\alpha^*(t)\} \, \mathrm{d}W(t), & (9) \end{aligned}$$

where we have suppressed the dependence of the drift on t. It is easy to accurately sample from this process using a high frequency Euler approximation (e.g. Kloeden and Platen (1992) and Jacod and Protter (1998)). This is crucial in our work and is discussed in general terms below.

Due to the common drift function the models (1) and (8) deliver locally equivalent measures P and Q, respectively. The resulting likelihood ratio is given by the Girsanov's formula (Øksendal (1998, p. 147)) for the path $\{\alpha(t); t \in [t_i, t_{i+1}]\}$. Ignoring constants it is given by

$$\begin{aligned} \log L_{P,Q}(\alpha) &= \int_{t_i}^{t_{i+1}} \left(\mu\left(\alpha\right) - \mu^*\left(\alpha\right)\right)' \Sigma^{-1}\left(\alpha\right) \mathrm{d}\alpha \\ &\quad - \frac{1}{2} \int_{t_i}^{t_{i+1}} \left(\mu\left(\alpha\right) - \mu^*\left(\alpha\right)\right)' \Sigma^{-1}\left(\alpha\right) \left(\mu\left(\alpha\right) - \mu^*\left(\alpha\right)\right) \mathrm{d}u. \end{aligned}$$

Beskos and Roberts (2004) show that it is sometimes possible to use this likelihood ratio inside a rejection algorithm to sample from the tied down version of (1) by making proposals from (8). However, in general this is not possible and we have to resort to MCMC methods.

We iterate to a path from P by using proposals from Q with the help of the following Metropolis-Hastings algorithm (for details of the algorithm see Chib and Greenberg (1995)).

Algorithm 2.1.

1. Set $j = 1$. Calculate some initial stretch $\{\alpha^{(0)}(t); t \in [t_i, t_{i+1}]\}$ which obeys the end point constraints $\alpha(t_i), \alpha(t_{i+1})$.

2. Propose the subpath $\{\alpha^{(j)}(t); t \in [t_i, t_{i+1}]\}$ by sampling from (8).

3. Accept proposal with probability

$$\min \left[1, L_{P,Q}(\alpha^{(j)}) / L_{P,Q}(\alpha^{(j-1)}) \right],$$

   otherwise write

$$\left\{ \alpha^{(j)}(t); t \in [t_i, t_{i+1}] \right\} = \left\{ \alpha^{(j-1)}(t); t \in [t_i, t_{i+1}] \right\}.$$

4. Set $j = j + 1$. Goto 2.

2.2 Numerical implementation

It is easy to accurately sample from the crucial (8) using a high frequency Euler approximation (e.g. Kloeden and Platen (1992) and Jacod and Protter (1998)). The Euler approximation of (8) is

$$\alpha_{i,j+1} | \alpha_{i,j} \sim N\left(\alpha_{i,j} + \delta\mu^*\left(\alpha_{i,j}\right), \delta\Sigma_{i,j}\right), \tag{10}$$

where $M_i \geq 1$ is a large positive integer, $\delta = (t_{i+1} - t_i)/M_i$, $\Sigma_{i,j} = \Sigma\left(\alpha_{i,j}\right)$ and

$$\alpha_{i,j} = \alpha(t_i + \delta j), \quad j = 0, 1, 2, ..., M_i.$$

This Euler approximation contrasts with an Euler approximation of the true process (1), which has

$$\alpha_{i,j+1} | \alpha_{i,j} \sim N\left(\alpha_{i,j} + \delta\mu\left(\alpha_{i,j}\right), \delta\Sigma_{i,j}\right). \tag{11}$$

We write the conditional density of (11) as $p_N(\alpha_{i,j+1}|\alpha_{i,j})$, while using $q_N(\alpha_{i,j+1}|\alpha_{i,j})$ for the corresponding one for (10).

Remark 2 *In the univariate case this is the same as the proposal process for the importance sampler of Durham and Gallant (2002). An extensive explanation of this importance sampler is given in Chib and Shephard (2002). In neither of these references is this link to the continuous time model in (8) made explicit, although the connection is more easily seen through Chib and Shephard (2002).*

Writing $\mu_{i,j} = \mu\left(\alpha_{i,j}\right)$ and $\mu_{i,j}^* = \mu^*\left(\alpha_{i,j}\right)$ then $\log L_{P,Q}(\alpha)$ is approximated by

$$\log \widehat{L}_{P,Q}(\alpha) = \sum_{j=0}^{M_i} \left(\mu_{i,j} - \mu_{i,j}^*\right)' \Sigma_{i,j}^{-1} \left(\alpha_{i,j+1} - \alpha_{i,j}\right) - \frac{1}{2}\delta \sum_{j=0}^{M_i} \left(\mu_{i,j} - \mu_{i,j}^*\right)' \Sigma_{i,j}^{-1} \left(\mu_{i,j} - \mu_{i,j}^*\right),$$

which we express as

$$\log \widehat{L}_{P,Q}(\alpha) = \sum_{j=0}^{M_i} \left(\log p_N(\alpha_{i,j+1}|\alpha_{i,j}) - \log q_N(\alpha_{i,j+1}|\alpha_{i,j}) \right). \tag{12}$$

Clearly $\log \widehat{L}_{P,Q}(\alpha)$ converges in probability to $\log L_{P,Q}(\alpha)$ as $M_i \to \infty$ using standard properties of the Euler approximation.

Remark 3 *For a fixed M_i, equation (12) shows the connection between this method and those of Elerian, Chib, and Shephard (2001) and Eraker (2001). Both these papers tried to sample from the log-density*

$$\sum_{j=0}^{M_i} \log p_N(\alpha_{i,j+1}|\alpha_{i,j}),$$

tied down by $\alpha_{i,0} = \alpha(t_i)$ and $\alpha_{i,M_i} = \alpha(t_{i+1})$. Elerian, Chib, and Shephard (2001) made multivariate proposals using a Laplace type approximation, which was computationally intensive but effective. Instead we now adopt the simpler approach of proposing blocks using (10), which is easier to code and numerically faster and better behaved as $M_i \to \infty$. Further, our analysis shows that a well designed proposal process should produce an excellent MCMC algorithm as the Metropolis-Hastings acceptance probability does not go to zero as $M_i \to \infty$. Eraker (2001) advocated a rather different approach. He favoured running MCMC chains inside these blocks, updating a single $\alpha_{i,j}$ at a time, conditional on its neighbours. Although this algorithm is as simple to code as our new approach and runs as quickly, it produces output that is more serially correlated. Indeed, Elerian (1999) proved that the rate of convergence of the Eraker (2001) algorithm worsens linearly with M_i.

When M is small it is quite possible that our proposal is not an outstandingly good one, due to the non-linearity in the diffusion. Thus, it may be beneficial to replace the Gaussian assumption in the proposal (10) with a heavier tailed alternative such as the multivariate-t. This requires a corresponding, but simple, adjustment in (12).

2.3 Numerical example

Before moving on, it is helpful to illustrate some of these points on a toy numerical example. We use the log of a square root process. Ito's lemma implies that it obeys the SDE

$$d\alpha = \left\{ \kappa(\mu - e^\alpha)e^{-\alpha} - \frac{1}{2}\sigma^2 e^{-\alpha} \right\} dt + \sigma e^{-\alpha/2} dW. \tag{13}$$

In this example a unit of time represents a month. We follow Durham and Gallant (2002) and Aït-Sahalia (2002, Table 3) in fixing $\kappa = 0.5$, $\mu = 0.06$ and $\sigma = 0.15$ to reflect parameters

calibrated to the US monthly treasury bill rate. We take $T = 2$, $n = 2$, $\tau_1 = 0$ and $\tau_2 = 2$, conditioning on $\alpha(0) = \log(0.05)$ and $\alpha(2) = \log(0.25)$. The task now is to draw paths from a non-linear diffusion which starts at $\log(0.05)$ and finishes at $\log(0.25)$. It is 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.
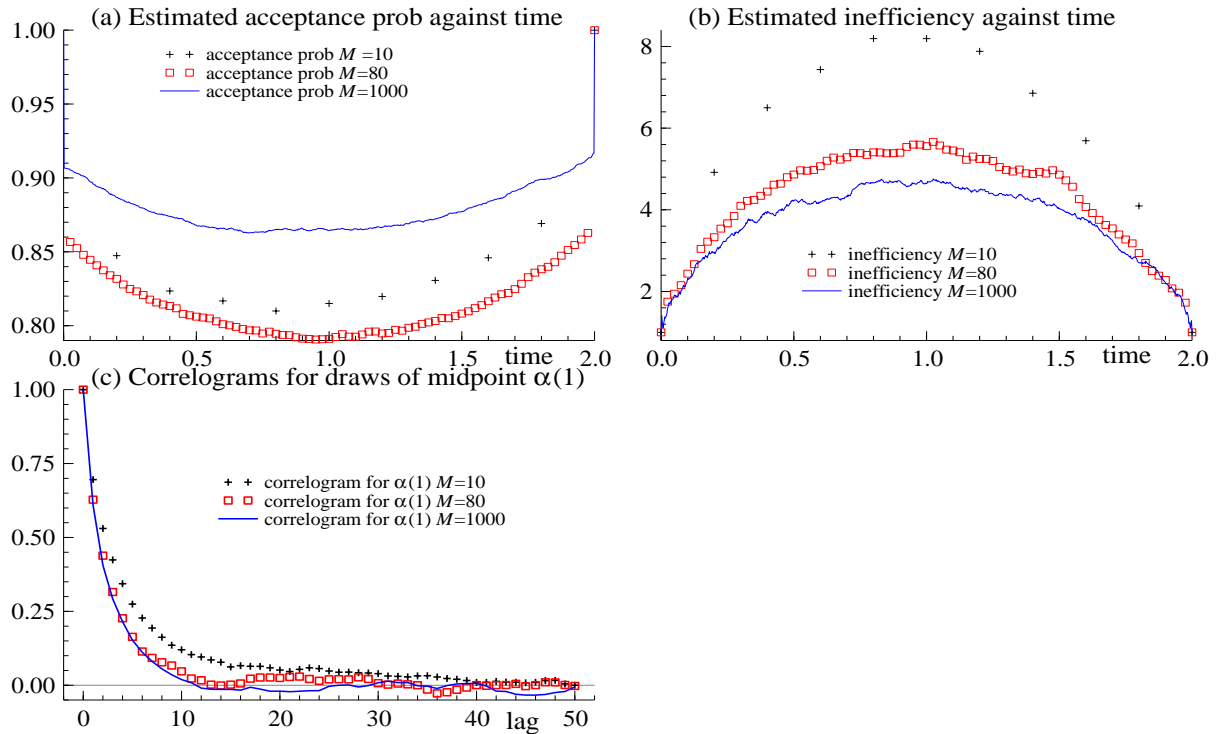

Figure 1: *Analysis of CIR model with $K = 2$, and M*, the number of points in the Euler discretisation per unit of time, set to 10, 80 and 1000. *(a) estimated average acceptance probability, (b) estimated inefficiency both against t, time. (c): correlogram of the sampled midpoint $\alpha(1)$.*

Throughout we use an Euler approximation based on M observations per unit of time. We report results from experiments run with M taking the values 10, 80 and 1,000. We selected two subsampling times t_1 and t_2 randomly without replacement from the set of times $\{j/M; \ j \in 1, 2, ..., 2M - 1\}$ and then sampled subpaths using the algorithm we have just discussed. Our results are based on 10,000 MCMC draws collected after a burn-in of a 100 cycles. The computation time is basically proportional to M.

Figure 1(a) shows the average acceptance rate from the M-H step for algorithms with the displayed subsampling times on the interval $(0, 2)$. The acceptance rate is usually over 80% and is not affected by M — the degree of augmentation. This is the expected result, although it is highly unusual in the context of MCMC algorithms. In Figure 1(b) we record the inefficiency factors for each augmented point, computed using 50 lags, while 1(c) shows the autocorrela-

tion function of the interpolated $\alpha(1)$. These inefficiency factors (with the maximum less than 8) indicate that the output from this sampler mixes extremely well. Interestingly, the inefficiency factors fall as M is increased. Overall the evidence reveals that the sampler has reliable performance and improves with larger values of M.

2.4 Reinstating the data

So far we have discussed the problem of sampling from

$$\{\alpha(t); t \in [t_i, t_{i+1}]\} \,|\alpha(t_i), \alpha(t_{i+1}),$$

rather than

$$\{\alpha(t); t \in [t_i, t_{i+1}]\} \,|Y_1, ..., Y_n, \alpha(t_i), \alpha(t_{i+1}). \tag{14}$$

As we have shown, consideration of the former expression is sufficient to deal with discretely observed diffusions. We assume that the probability density of (14) which we write as

$$\mathrm{d}F\left(Y_2, ..., Y_n|Y_1, \{\alpha(t); t \in [0, T]\}\right) = \prod_{i=1}^{n-1} \mathrm{d}F\left(Y_{i+1}|Y_i, \{\alpha(t); t \in [\tau_i, \tau_{i+1}]\}\right)$$

can be computed. Associated with the strip of time from t_i to t_{i+1} and the path of the process $\{\alpha(t); t \in [t_i, t_{i+1}]\}$ are the observations Y which are influenced by this subpath. Let c_i and c_{i+1} be, respectively, the largest and smallest integer such that

$$\tau_{c_i} \le t_i \le t_{i+1} \le \tau_{c_{i+1}}.$$

Then the MCMC algorithm given in 2.1 changes just by a modification of Step 3. It is restated to become:

Algorithm 2.4.

1. Set $j = 1$. Calculate some initial stretch $\{\alpha^{(0)}(t); t \in [t_i, t_{i+1}]\}$ which obeys the end point constraints $\alpha(t_i), \alpha(t_{i+1})$.

2. Propose the subpath $\{\alpha^{(j)}(t); t \in [t_i, t_{i+1}]\}$ by sampling from (8).

3. Accept proposal with probability

$$\min \left[1, \frac{L_{P,Q}(\alpha^{(j)})}{L_{P,Q}(\alpha^{(j-1)})} \frac{\displaystyle\prod_{k=c_i}^{c_{i+1}-1} \mathrm{d}F\left(Y_{k+1}|Y_k, \alpha^{(j)}\right)}{\displaystyle\prod_{k=c_i}^{c_{i+1}-1} \mathrm{d}F\left(Y_{k+1}|Y_k, \alpha^{(j-1)}\right)} \right], \tag{15}$$

otherwise write

$$\left\{\alpha^{(j)}(t); t \in [t_i, t_{i+1}]\right\} = \left\{\alpha^{(j-1)}(t); t \in [t_i, t_{i+1}]\right\}.$$

4. Set $j = j + 1$. Goto 2.

Remark 4 *It is often natural to choose $t_i = \tau_i$ for all $i = 1, 2, ..., n$ then (15) simplifies to*

$$\min \left[1, \frac{L_{P,Q}(\alpha^{(j)})}{L_{P,Q}(\alpha^{(j-1)})} \frac{\mathrm{d}F\left(Y_{i+1}|Y_i, \alpha^{(j)}\right)}{\mathrm{d}F\left(Y_{i+1}|Y_i, \alpha^{(j-1)}\right)} \right],$$

as $c_i = i$ for all i.

Example 4 (continued from Example 2). In the SV case

$$Y_{i+1}|Y_i, \alpha \sim N(Y_i + \mu_P \left(\tau_{i+1} - \tau_i\right) + \beta\sigma_{i+1}^2 + \rho Z_{i+1}, \left(1 - \rho^2\right)\sigma_{i+1}^2),$$

where

$$\sigma_{i+1}^2 = \int_{\tau_i}^{\tau_{i+1}} \sigma^2(u)\mathrm{d}u, \quad Z_{i+1} = \int_{\tau_i}^{\tau_{i+1}} \sigma(u)\mathrm{d}W(u).$$

Example 5 (continued from Example 3). In the Cox process case

$$Y_{i+1} - Y_i|Y_i, \alpha \sim Po(\lambda_{i+1}), \qquad \lambda_{i+1} = \int_{\tau_i}^{\tau_{i+1}} \lambda(u)\mathrm{d}u,$$

where $Po(\mu)$ denotes a Poisson random variable with mean μ.

3 Algorithm for parameter sampling

3.1 Basic approach

The sole remaining problem for handling diffusion driven models is the step for sampling from

$$\theta|\left\{\alpha(t); t \in [0, T]\right\}.$$

At this point it is helpful to reparameterise the diffusion component as

$$\mathrm{d}\alpha(t) = \mu\{\alpha(t); \psi\}\mathrm{d}t + \Theta\{\alpha(t); \omega\}\mathrm{d}W(t), \quad t \geq 0, \tag{16}$$

where $\theta = \left(\psi', \omega'\right)'$.

We initially think of ω as known, when the log-likelihood for the path of $\{\alpha(t); t \in [0, T]\}$ by Girsanov's formula is

$$
\begin{aligned}
\log L(\alpha; \psi) \quad = \quad & const + \int_0^t \left(\mu\left(\alpha\right) - \mu^*\left(\alpha\right)\right)' \Sigma^{-1}\left(\alpha\right) \mathrm{d}\alpha \\
& -\frac{1}{2} \int_0^t \left(\mu\left(\alpha\right) - \mu^*\left(\alpha\right)\right)' \Sigma^{-1}\left(\alpha\right) \left(\mu\left(\alpha\right) - \mu^*\left(\alpha\right)\right) \mathrm{d}u.
\end{aligned}
$$

We can use this inside a M-H algorithm to appropriately sample from

$$\psi|\omega, \alpha.$$

11

The difficulty with this method is that the sample path of α exactly gives us the integral of the volatility function through the quadratic variation of the diffusion

$$[\alpha](t) = \int_0^t \Theta\{\alpha(u); \omega\}\Theta\{\alpha(u); \omega\}' \mathrm{d}u.$$

In many cases this is likely to mean that we can deduce ω from α and therefore $\omega|\alpha, \psi$ may well be degenerate, which implies that the MCMC method will not converge. This feature was pointed out in the context of univariate diffusions by Roberts and Stramer (2001) and we call it the *Roberts-Stramer critique* of this method.

Of course, in practice the sampling schemes above are implemented via an Euler scheme, applied in conjunction with M augmented points. The critique is less binding when M is small, which explains why writers like Eraker (2001) and Eraker, Johannes, and Polson (2003) have not really remarked on it. However, when M is substantially large, as in our algorithm, the possibility of degeneracy of the conditional distribution of ω becomes increasingly likely, regardless of what happens in step 1 of Algorithm 1.2. It is therefore important to have a generic solution to this problem, which we now supply.

3.2   A reparameterisation

In the discussion thus far we have focused on ways to sample from $\alpha, \theta | Y_1, ..., Y_n$, however, we can alternatively sample from

$$\theta, \{W(t); t \in [0, T]\} | Y_1, ..., Y_n,$$

the posterior of the parameters and the driving Brownian process. We write the prior for θ as $\pi(\theta)$. At first sight this looks like the same problem because the addition of θ to the path of W yields the path of α. In fact, the resulting MCMC algorithm (which we refer to as the innovation scheme) is subtly different.

Algorithm 3.2.

1. Sample from $\{W(t); t \in [0, T]\} | \theta, Y_1, ..., Y_n$ by updating the subsets of the sample path.

2. Draw from

$$\theta | \{W(t); t \in [0, T]\}, Y_1, ..., Y_n.$$

3. Goto 1.

Step 1 is carried out using Algorithm 2.4. Conditional upon θ we convert $\{\alpha(t); t \in [0, T]\}$ to obtain $\{W(t); t \in [0, T]\}$. This is not a standard Gibbs-type method when applied to W.

However, since, conditional on θ, there is a one-to-one relationship between W and α over $t \in [0, T]$ it is valid to take a Gibbs sample of α and convert it into W. The remaining task is to sample from $\theta|W, Y_1, ..., Y_n$. Generically this could take the following form.

2a. Combine $\theta^{(j-1)}$ and W to construct a path $\alpha^{(j-1)}$.

2b. Propose $\theta^{(j)}$ from some density $g(\theta)$ which could depend upon Y, W and $\theta^{(j-1)}$. Use $\theta^{(j)}$ and W to construct a path $\alpha^{(j)}$. Accept the proposal with probability

$$\min \left[1, \frac{g(\theta^{(j-1)})}{g(\theta^{(j)})} \frac{\prod_{k=1}^{n-1} \mathrm{d}F\left(Y_{k+1}|Y_k, \alpha^{(j)}, \theta^{(j)}\right) \pi(\theta^{(j)})}{\prod_{k=1_i}^{n-1} \mathrm{d}F\left(Y_{k+1}|Y_k, \alpha^{(j-1)}, \theta^{(j-1)}\right) \pi(\theta^{(j-1)})} \right].$$

2c. If the proposal is rejected write $\theta^{(j)} = \theta^{(j-1)}$.

The innovation scheme algorithm is rather simple and overcomes the Roberts-Stramer critique. In practice we have tended to propose θ using a Laplace approximation to the conditional posterior

$$\left\{ \prod_{k=1}^{n-1} \mathrm{d}F\left(Y_{k+1}|Y_k, \alpha^{(j)}, \theta\right) \right\} \pi(\theta).$$

3.3 Simulation experiment

Our aim is to compare the effectiveness of finite augmentation implementations of samplers based on $\theta, W|Y$ to those which use $\alpha, \theta|Y$. We know that in the limit of M the latter algorithm will collapse. The former scheme should, however, not be affected by this problem. We develop a simple simulation experiment, letting M take the values 1, 4, 10 and 50 and T the values 60 and 600. We return to Example 4 but now set $\mu_P = \beta = \rho = 0$ to simplify the exposition and use a log-OU volatility process with $\sigma^2(t) = \exp(\alpha(t))$ where

$$\mathrm{d}\alpha = \kappa(\mu - \alpha)\mathrm{d}t + \omega \mathrm{d}W$$

and $\theta = (\kappa, \mu, \omega)'$ which we set to $(0.03, -0.6, \sqrt{0.03})'$ in generating the data. We let $\alpha(0) \sim N(\mu, \omega^2/2\kappa)$, the stationary density of α, and assume that the log-price process is observed every unit of time, so $Y_i = P(i)$ for $i = 1, 2, ..., n$. The augmentation is carried out with fixed blocks, setting $t_j = 10j$ for $j = 1, 2, ..., n/10$.

We assess the effectiveness of the two algorithms by reporting the INF measure, see Section 2. The results are shown in Table 1. We see that the conclusions are similar. The inefficiency factors for μ and the middle volatility state, $\alpha_{T/2}$, appear largely unaffected by the size of T

and M under either scheme. This reflects the robustness of our algorithms in sampling the latent diffusions. For the state based scheme the inefficiency factors for κ and ω rise with T and particularly with M. For $T = 600$ and $M = 50$ the state scheme for ω is more than $2,000$ times less efficient than a hypothetical independent sampler. This inefficiency is almost certainly severely underestimated as we can only run the sampler for 20,000 iterations due to the computational burden when $M = 50$. Additional graphical evidence, not displayed here, shows that the Markov chain output for ω hardly moves from its initial position when M is 50. For the innovation scheme, our recommended approach, the inefficiency factors for all parameters are low (at most around 35) and do not rise with M. The overall efficiency of the innovation scheme is due to the low posterior correlation between α and ω.

$T = 60$	κ		μ		ω		α_{30}	
	$\alpha, \theta\|Y$	$\theta, W\|Y$	$\alpha, \theta\|Y$	$\theta, W\|Y$	$\alpha, \theta\|Y$	$\theta, W\|Y$	$\alpha, \theta\|Y$	$\theta, W\|Y$
$M = 1$	5.27	2.31	16.6	9.19	33.6	2.35	17.1	5.42
$M = 4$	6.09	2.41	14.9	8.22	146	2.20	15.7	7.62
$M = 10$	4.38	2.12	13.0	8.19	581	2.67	15.4	6.05
$M = 50$	8.04	3.06	30.2	8.83	1526	3.09	16.1	6.56
$T = 600$	κ		μ		ω		α_{300}	
	$\alpha, \theta\|Y$	$\theta, W\|Y$	$\alpha, \theta\|Y$	$\theta, W\|Y$	$\alpha, \theta\|Y$	$\theta, W\|Y$	$\alpha, \theta\|Y$	$\theta, W\|Y$
$M = 1$	61.1	12.2	2.13	1.95	230	33.8	9.23	8.21
$M = 4$	175	13.4	1.95	1.72	731	35.6	10.7	10.8
$M = 10$	105	12.7	2.09	1.62	605	31.7	13.0	10.7
$M = 50$	67.4	15.8	2.17	1.94	2169	36.8	22.3	11.9

Table 1: Results for sampler based upon states and upon innovations for parameters in stochastic volatility model. True parameters $(\kappa, \mu, \omega)' = (0.03, -0.6, \sqrt{0.03})'$. Reported are inefficiency factors based upon 20,000 MCMC samples with block size of 10 units of time and M Euler augmented points per unit of time.

4 Application: S & P 500

In this section we estimate the volatility with leverage model of (2) to daily returns data on the closing prices of the Standard and Poor's 500 index from 5/5/1995 to 14/4/2003 ($T = 2,000$). We consider three different models for the volatility $\sigma^2 = \exp(\alpha)$, with

$$
\begin{aligned}
\mathrm{d}\alpha &= \kappa(\mu - \alpha)\mathrm{d}t + \omega\mathrm{d}W, & \text{OU} \\
\mathrm{d}\alpha &= \left\{\kappa(\mu - e^\alpha)e^{-\alpha} - \omega^2 e^{-\alpha}/2\right\}\mathrm{d}t + \omega e^{-\alpha/2}\mathrm{d}W, & \text{SQRT} \\
\mathrm{d}\alpha &= \left\{\kappa(\mu - e^\alpha)e^{-\alpha} - \omega^2/2\right\}\mathrm{d}t + \omega\mathrm{d}W. & \text{GARCH diffusion.}
\end{aligned}
\tag{17}
$$

The parameters κ, ω and μ thus have different interpretations across the different models, while μ_P and β from (2) are common. The sixth parameter is the leverage coefficient ρ.

We run 20,000 iterations of the non-centered MCMC algorithm using the $\theta, W|Y$ parameterisation for $M = 1, 4$ and 10 where $\theta = (\kappa, \mu, \omega, \mu_P, \beta, \rho)'$. We sample, on average, 10 days

of the volatility diffusion at a time, regardless of M. The log-likelihood (estimated at the mean of $\theta|y$) is computed using particle filter methods, which are discussed in Section 5.2 of this paper. The correlograms from the MCMC output are shown in Figure 2 (for $M = 1$ for brevity). The posterior kernel density estimates are given in Figure 3 for the volatility parameters and Figure 4 for the volatility parameters. To calibrate these numbers, when we fit the Gaussian-GARCH(1,1) by maximum likelihood with the inclusion of μ_y and β the corresponding maximised log-likelihood was $-3,074$. Numerical summaries of the marginal posterior densities for each parameter, including the inefficiency factors, are provided in Table 2 for $M = 1$, 4, 10 and 20.
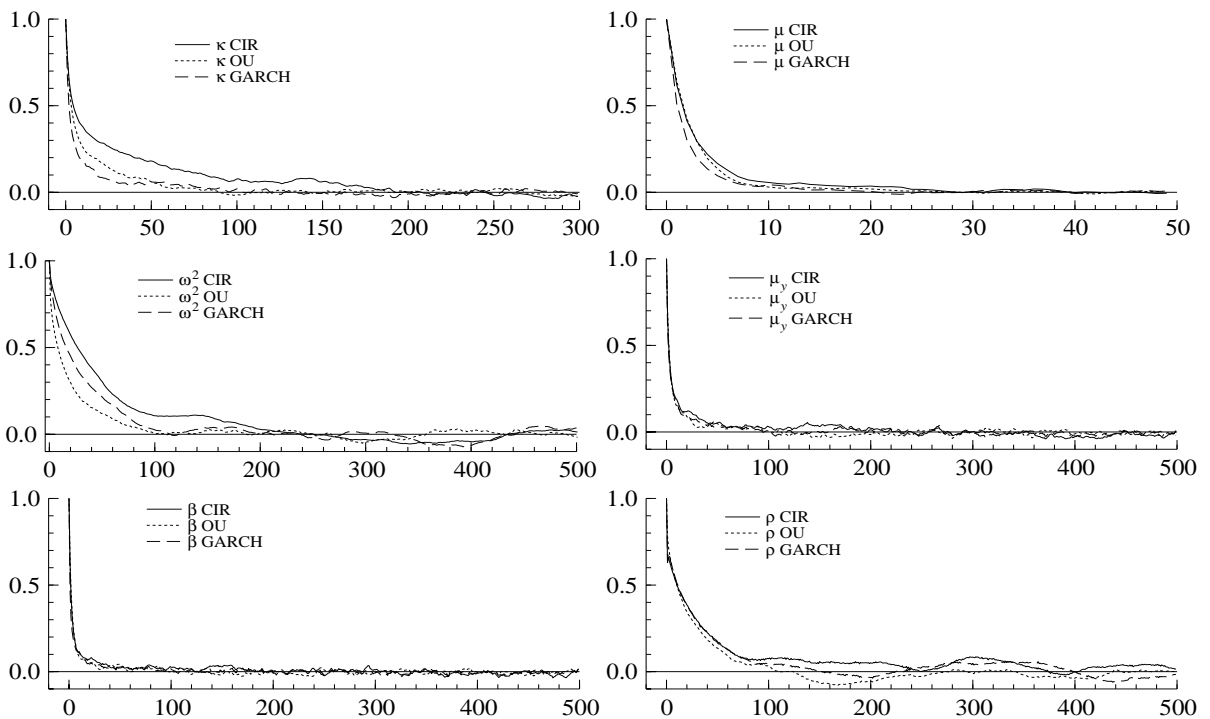

Figure 2: *Correlograms resulting from* $20,000$ *iterations of the MCMC algorithm. Data is S &P 500 continuously componded returns. T=2,000, $M = 1$.*

Leverage is important as we find ρ to be negative in all cases. When ρ is constrained to 0 we find that μ_P and β are different from 0 (results not reported here) and the maximized log-likelihoods are $-3,040.2$, $-3,059.3$, $-3,037.3$ for the SQRT, OU and GARCH diffusion models, respectively. Notice these are much higher than the Gaussian-GARCH model mentioned in the previous paragraph. When ρ is left unconstrained we see from Table 2 that both parameters are close to 0 and that the maximized log-likelihood value increases considerably for each of the three models. Quantitatively there is a degree of robustness associated with the estimation of ρ as we obtain similar results across models. We also see from Figure 3 that increasing M has
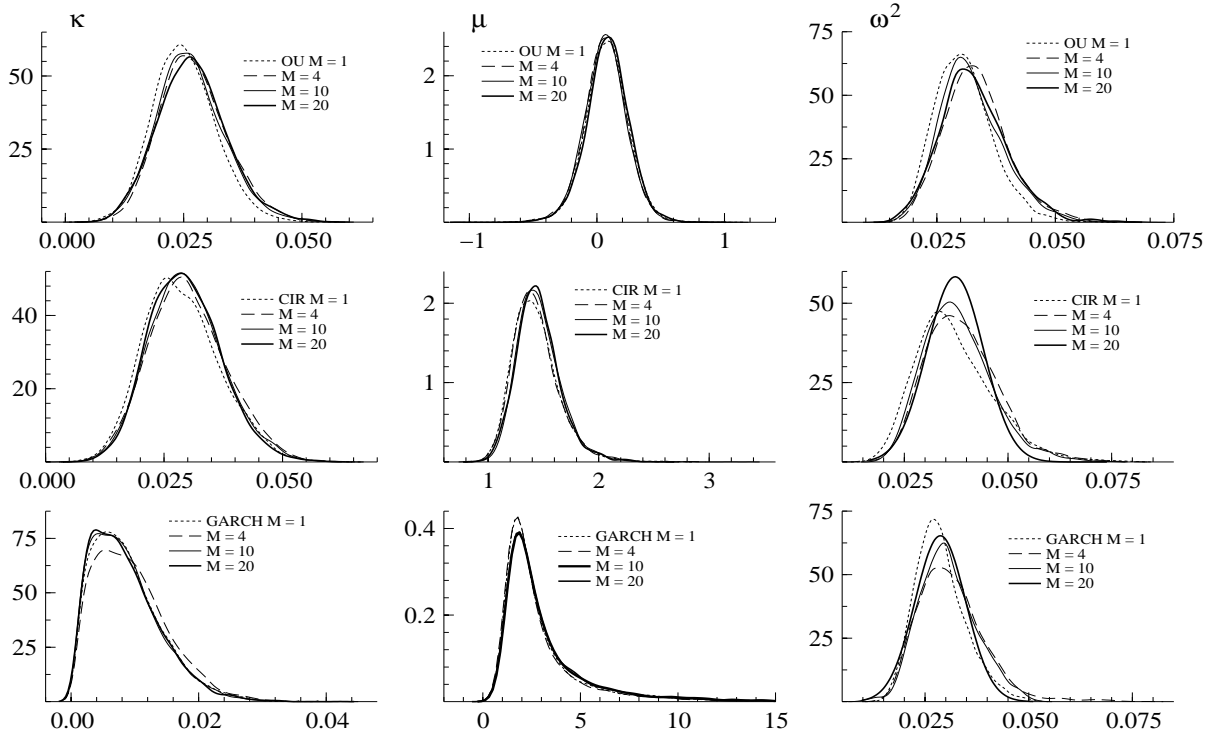
Figure 3: *Kernel density estimates for κ, μ and ω^2 (from left to right) for the OU, SQRT and Nelson SV models (top to bottom) resulting from $20,000$ iterations of the MCMC algorithm. Data is S &P 500 continuously componded returns. T=2,000, $M = 1, 4, 10, 20$.*

little effect on the marginal posterior densities of the volatility parameters. However, the effect on the return parameters (Figure 4) is more substantial, and the choice $M = 1$ is inadequate. Increasing M from 1 to 4 is sufficient for the parameters μ_P and β, whereas for ρ inferences based on $M = 4$ and $M = 10$ are slightly different in the SQRT and GARCH diffusion models. The reason that M needs to be moderately large for ρ is due to the dependence on the Ito integral Z_{i+1} in (2); in contrast, for the case of σ^2_{i+1} the integrator is time.

It can also be observed that as we increase the value of M from 1, ρ decreases quite substantially (almost by one standard deviation) to less than -0.8 in each model. In addition, the posterior standard deviation of ρ falls from more than 0.05 when $M = 1$ to around 0.04 when $M = 10$. Further increases in M do not affect the results. The inefficiency factors displayed in Table 2 are a little higher than those in Table 1. This is probably due to the posterior correlation between ω^2 and ρ. However, in all cases the inefficiency factors are less than 100 and vary little with M. This indicates that our methods are invariant to M in terms of the efficiency factors (although computationally linear in M). Thus the choice $M = 10$ is adequate in this example. The resulting dimension of the discretised volatility path is $20,000$.

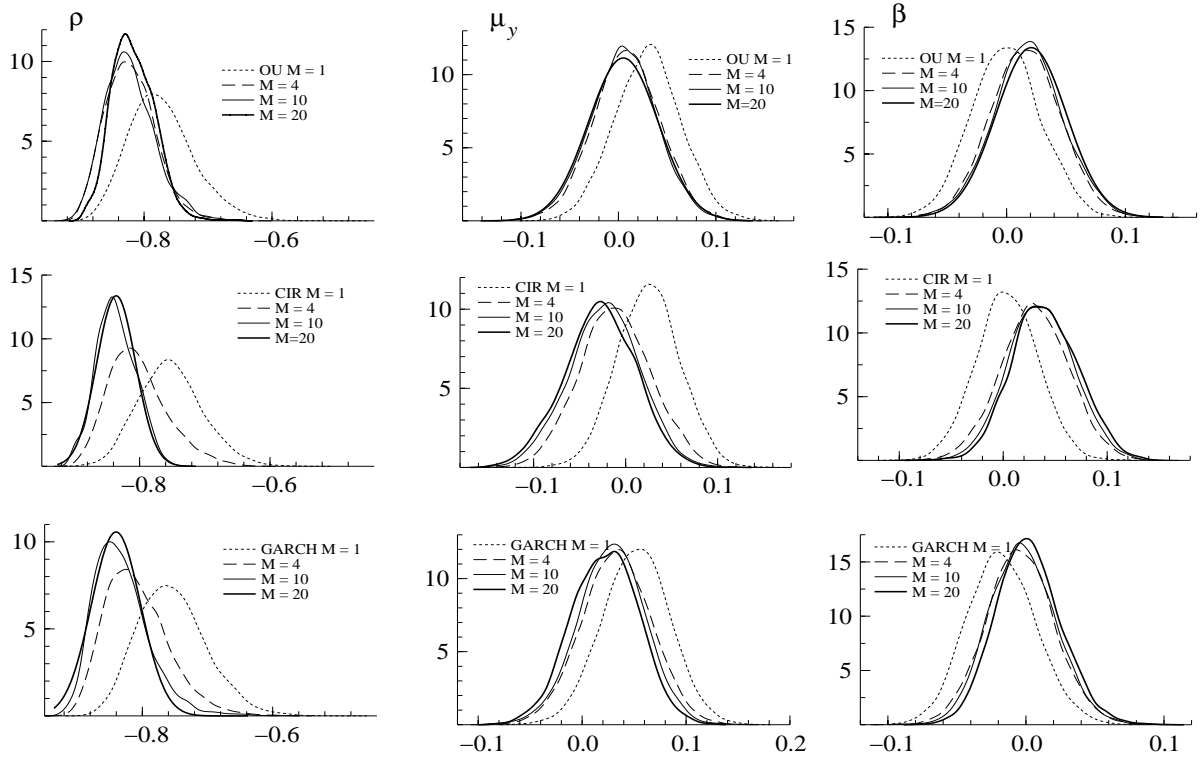Whilst the differences between the three models, in terms of log-likelihood, are large when

Figure 4: *Kernel density estimates for ρ, μ_y and β (from left to right) for the OU, SQRT and Nelson SV models (top to bottom) resulting from $20,000$ iterations of the MCMC algorithm. Data is S&P 500 continuously componded returns. T=2,000, M = 1, 4, 10, 20.*

leverage is not included, the differences disappear when leverage is included. When leverage is excluded the log-likelihood indicates that the GARCH diffusion model is preferable. When leverage is included, the SQRT model for volatility appears to provide the best fit. The log likelihoods of these models are all high relative to a competing GARCH(1, 1) model.

5 Discussion and extensions

5.1 Intensity models

In Examples 3 and 5 we have analysed counting processes which are observed at irregularly spaced time intervals. Perhaps more interestingly we can also carry out inference directly on the entire sample path of the counting process N. These type of data are now common in financial econometrics where we have records of quote changes and transactions (for example Engle (2000) for an overview). Such situations also arise in the field of credit risk. The interarrival times are exponentially distributed conditional upon the λ path. Let us define the arrival time of events as $0 \leq a_1 < a_2 < ... < a_{Y_T} \leq T$ where N_T is the total number of events over $[0, T]$ then we

CIR model: log lik $(\bar{\theta}) = -2997.86$								
	k			μ			ω^2		
	mean.	st. dev	INF (300)	mean	st. dev.	INF (50)	mean	st. dev.	INF (500)
M=1	0.028	0.0079	33.0	1.43	0.2132	5.8	0.0363	0.0088	86.8
M=4	0.030	0.0080	37.9	1.43	0.2096	7.3	0.0388	0.0087	71.9
M=10	0.029	0.0075	22.3	1.45	0.1968	8.2	0.0368	0.0076	68.6
M=20	0.029	0.0075	32.6	1.46	0.1982	9.2	0.0377	0.0074	70.7
	μ_y			β			ρ		
	mean	st. dev.	INF (50)	mean	st. dev.	INF	mean	st. dev.	INF (500)
M=1	0.0274	0.0348	18.7	0.0022	0.0301	14.3	-0.746	0.0515	55.1
M=4	-0.0115	0.0388	28.9	0.0305	0.0325	18.6	-0.802	0.0452	97.7
M=10	-0.0227	0.0384	17.1	0.0368	0.0320	14.1	-0.830	0.0371	66.0
M=20	-0.0277	0.0383	27.1	0.0392	0.0315	18.0	-0.835	0.0352	82.3

log OU model: log lik $(\bar{\theta}) = -2999.44$								
	k			μ			ω^2		
	mean	st. dev.	INF (300)	mean	st. dev.	INF (50)	mean	st. dev.	INF (500)
M=1	0.024	0.0066	20.9	0.0736	0.1744	5.15	0.0302	0.0060	32.0
M=4	0.027	0.0071	27.7	0.0672	0.1642	6.21	0.0333	0.0068	60.2
M=10	0.026	0.0071	12.4	0.0714	0.1622	7.78	0.0319	0.0068	16.3
M=20	0.027	0.0072	31.5	0.0721	0.1671	5.69	0.0328	0.0070	54.8
	μ_y			β			ρ		
	mean	st. dev.	INF (50)	mean	st. dev.	INF (50)	mean	st. dev.	INF (500)
M=1	0.0313	0.0342	12.1	-0.0722	0.0280	2.53	-0.774	0.0515	34.1
M=4	0.0085	0.0344	14.3	0.0159	0.0292	9.89	-0.821	0.0392	51.8
M=10	0.0046	0.0344	10.9	0.0189	0.0295	7.79	-0.824	0.0397	71.6
M=20	0.0051	0.0358	13.6	0.0206	0.0297	8.65	-0.825	0.0365	67.8

Nelson model: log lik $(\bar{\theta}) = -2999.54$								
	k			μ			ω^2		
	mean	st. dev.	INF (300)	mean	st. dev.	INF (50)	mean	st. dev.	INF (500)
M=1	0.0088	0.0054	15.5	2.96	2.24	4.10	0.0289	0.0063	44.3
M=4	0.0095	0.0057	22.7	2.93	2.33	6.18	0.0313	0.0081	87.1
M=10	0.0087	0.0054	18.3	3.06	2.27	2.72	0.0300	0.0059	30.6
M=20	0.0085	0.0054	23.5	3.09	2.32	7.13	0.0287	0.0061	56.3
	μ_y			β			ρ		
	mean	st. dev.	INF (50)	mean	st. dev.	INF (50)	mean	st. dev.	INF (500)
M=1	0.0538	0.0325	14.25	-0.0730	0.0287	2.42	-0.746	0.0541	47.1
M=4	0.0330	0.0331	17.47	-0.0041	0.0247	2.76	-0.807	0.0499	51.5
M=10	0.0271	0.0309	22.57	-0.0014	0.0236	3.67	-0.832	0.0430	68.0
M=20	0.0220	0.0317	19.32	0.0017	0.0235	4.52	-0.840	0.0391	59.6

Table 2: *Parameter estimates resulting from different volatility models applied to S & P 500 returns. $T = 2000$.*

obtain the log-likelihood as,

$$-\int_0^T \lambda(u)\mathrm{d}u + \int \log \lambda \{a(u)\} \, \mathrm{d}N_u = -\int_0^T \lambda(u)\mathrm{d}u + \sum_{i=1}^{Y_T} \log \lambda(a_i).$$

We can update blocks of intensities and use a fraction of this log-likelihood inside a M-H algorithm, in precisely the manner for SV models.

5.2 Filtering

5.2.1 Basics of particle filtering

Economists are often interested in

$$\alpha(\tau_i)|Y_1, ..., Y_i, \theta, \qquad i = 1, 2, ..., n. \tag{18}$$

Gordon, Salmond, and Smith (1993), Pitt and Shephard (1999), Jacod and Moral (2001), Jacod, Moral, and Protter (2001) and Johannes, Polson, and Stroud (2002) have developed a method called particle filters which can sequentially approximate this density. The last three references are particularly relevant as they develop particle filtering methods in the context of partially observed SDEs, while the second, third and sixth deal with examples in financial economics. A radical alternative to filtering, called backprojection, has been developed by Gallant and Tauchen (1998).

We suppress dependence on θ. The basic approach of particle filtering is to start with a sample from $\alpha(\tau_i)|Y_1, ..., Y_i$, which we write as

$$\alpha^{(j)}(\tau_i), \quad j = 1, 2, ..., J.$$

We use this sample to produce a sample of size J from $\alpha(\tau_{i+1})|Y_1, ..., Y_{i+1}$. Having propagated the particles one step forward we can repeat the process sequentially through time. The propagation can be carried out in a number of ways. A simple algorithm is given below.

1. For each $j = 1, 2, ..., J$ produce a sample of size K subpaths

$$\{\alpha(t); t \in [\tau_i, \tau_{i+1}]\} | \alpha^{(j)}(\tau_i).$$

Call this subpath $\alpha^{(j,k)}$, $k = 1, 2, ..., K$.

2. Compute
$$w_{j,k} = dF(Y_{i+1}|Y_i, \left\{\alpha^{(j,k)}(t); t \in [\tau_i, \tau_{i+1}]\right\}).$$

3. Resample subpaths with probability proportional to $w_{j,k}$ to produce a sample of size J. Call these subpaths $\left\{\alpha^{(j)}(t); t \in [\tau_i, \tau_{i+1}]\right\}$ for $j = 1, 2, ..., J$. Placed into memory $\alpha^{(j)}(\tau_{i+1})$.

The above references show that as $J, K \to \infty$ this algorithm sequentially produces valid samples from the filtering density (18).

5.2.2 Model fit and model checking

The output from the particle filter can allow us to compute the likelihood for the model using the prediction decomposition

$$\mathrm{d}F(Y_2, ..., Y_n | Y_1, \theta) = \prod_{i=1}^{n-1} \mathrm{d}F(Y_{i+1} | Y_1, Y_2, ..., Y_i, \theta),$$

where we can consistently estimate $\mathrm{d}F(Y_{i+1} | Y_1, Y_2, ..., Y_i, \theta)$ as

$$\frac{1}{JK} \sum_{j=1}^{J} \sum_{k}^{K} w_{j,k}.$$

In discrete time SV models Kim, Shephard, and Chib (1998) used these types of calculations to compare SV models to GARCH processes.

The estimate of the likelihood can be used as an input into the marginal likelihood of the model

$$\int dF(Y_2, ..., Y_n | Y_1, \theta) \pi(\theta) \mathrm{d}\theta,$$

which from Chib (1995), can be more usefully expressed as

$$\frac{dF(Y_2, ..., Y_n | Y_1, \theta^*) \pi(\theta^*)}{\pi(\theta^* | y)},$$

where θ^* is any value in the support of the posterior. A method for estimating the marginal likelihood based on the proceeding identity has been developed by Chib (1995) who recommends numerically approximating the ordinates $dF(Y_2, ..., Y_n | Y_1, \theta^*)$ and $\pi(\theta^* | Y_1, ..., Y_n)$ taking θ^* to be a high-density point such as the posterior mean of θ. This approach to comparing models was implemented for discrete time SV models by Elerian, Chib, and Shephard (2001). No real new issues arise here.

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

$$u_{i+1} = F(Y_{i+1} | Y_1, ..., Y_i; \theta).$$

Such a time series of distribution functions should be i.i.d. 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 Diebold, Gunther, and Tay (1998) 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 challenge is actually computing u_{i+1}. However, the output from the particle filter solves this problem for we just have to average

$$F(Y_{i+1}|Y_i, \left\{ \alpha^{(j,k)}(t); t \in [\tau_i, \tau_{i+1}] \right\}$$

over the superscript j, k recorded in Step 2 of the previous algorithm.

6   Conclusion

This paper has provided a unified likelihood based approach for inference in diffusion driven models. This is based on a effective proposal scheme for sampling subpaths of the diffusive process and a reparameterisation of the model to overcome degeneracies in the MCMC algorithm. 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 provide a consistent way of dealing with other issues such as model choice, model checking and filtering.

7   Acknowledgments

References

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

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

Aït-Sahalia, Y., L. P. Hansen, and J. Scheinkman (2004). Discretely sampled diffusions. In Y. Aït-Sahalia and L. P. Hansen (Eds.), *Handbook of Financial Econometrics*. Amsterdam: North Holland. Forthcoming.

Beskos, A. and G. Roberts (2004). Exact simulation of diffusions. Unpublished paper: Mathematics Department, Lancaster University.

Bibby, B. M., M. Jacobsen, and M. Sørensen (2004). Estimating functions for discretely samples diffusion-type models. In Y. Aït-Sahalia and L. P. Hansen (Eds.), *Handbook of Financial Econometrics*. Amsterdam: North Holland. Forthcoming.

Bowsher, C. G. (2003). Modelling security market events in continuous time: intensity based, multivariate point process models. Unpublished paper: Nuffield College, Oxford.

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.

Chib, S. (1995). Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association 90*, 1313–21.

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. and N. Shephard (2002). Comment on 'Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes,' by Durham and Gallant. *Journal of Business and Economic Statistics 20*, 325–327.

Cox, D. R. (1955). Some statistical models connected to a series of events (with discussion). *Journal of the Royal Statistical Society, Series B 17*, 129–64.

Diebold, F. X., T. A. Gunther, and T. S. Tay (1998). Evaluating density forecasts with applications to financial risk management. *International Economic Review 39*, 863–883.

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

Duffie, D. and K. Singleton (1999). Modeling term structures of defaultable bonds. *Review of Financial Studies 12*, 687–720.

Durham, G. (2003). Likelihood-based specification analysis of continuous-time models of the short-term interest rate. *Journal of Financial Economics 70*, 463–487.

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.

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.

Engle, R. F. (2000). The econometrics of ultra-high frequency data. *Econometrica 68*, 1–22.

Engle, R. F. and J. R. Russell (1998). Forecasting transaction rates: the autoregressive conditional duration model. *Econometrica 66*, 1127–1162.

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. G. Polson (2003). The impact of jumps in returns and volatility. *Journal of Finance 53*, 1269–1300.

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 (2004). 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*, 1317–39.

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.

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

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

Jacod, J. and P. Protter (1998). Asymptotic error distributions for the Euler method for stochastic differential equations. *Annals of Probability 26*, 267–307.

Jacquier, E., N. G. Polson, and P. E. Rossi (1994). Bayesian analysis of stochastic volatility models (with discussion). *Journal of Business and Economic Statistics 12*, 371–417.

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

Karatzas, I. and S. E. Shreve (1991). *Brownian Motion and Stochastic Calculus* (2 ed.), Volume 113 of *Graduate Texts in Mathematics*. Berlin: Springer–Verlag.

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

Kloeden, P. E. and E. Platen (1992). *Numerical Solutions to Stochastic Differential Equations.* New York: Springer.

Lando, D. (1998). On Cox processes and credt risk securities. *Review of Derivatives Research 2*, 99–120.

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

Øksendal, B. (1998). *Stochastic Differential Equations. An Introduction with Financial Applications* (5 ed.). Berlin: Springer.

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. 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.

Shephard, N. (1994). Partial non-Gaussian state space. *Biometrika 81*, 115–131.

Shephard, N. (2004). *Stochastic Volatility: Selected Readings.* Oxford: Oxford University Press. Forthcoming.

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. *Journal of Forecasting 4*, 283–91.