Likelihood-based Inference for a Class of Multivariate Diffusions with Unobserved Paths

Konstantinos Kalogeropoulos

May 18, 2007

Abstract

This paper presents a Markov chain Monte Carlo algorithm for a class of multivariate diffusion models with unobserved paths. This class is of high practical interest as it includes most diffusion driven stochastic volatility models. The algorithm is based on a data augmentation scheme where the paths are treated as missing data. However, unless these paths are transformed so that the dominating measure is independent of any parameters, the algorithm becomes reducible. The methodology developed in Roberts and Stramer (2001 Biometrika 88(3):603-621) circumvents the problem for scalar diffusions. We extend this framework to the class of models of this paper by introducing an appropriate reparametrisation of the likelihood that can be used to construct an irreducible data augmentation scheme. Practical implementation issues are considered and the methodology is applied to simulated data from the Heston model.

Keywords: Data augmentation; Markov chain Monte Carlo; Stochastic volatility; Heston model.

1 Introduction

Diffusion processes constitute a natural and useful tool for modelling phenomena evolving continuously in time. They find applications in many different fields including finance, biology, physics, engineering etc. A diffusion process V_t is defined through a stochastic differential equation (SDE):

$$dV_t = \mu(V_t, \theta)dt + \sigma(V_t, \theta)dB_t,\ V_0 = v_0,\ 0 \leq t \leq T,\ V_t \in \Re^d \tag{1}$$

where B_t is a standard Brownian motion. The drift μ and volatility σ of the diffusion should satisfy some regularity conditions (locally Lipschitz with a growth bound) to ensure that

the SDE will have a weakly unique solution; see chapter 4 of Kloeden and Platen (1995). Throughout this paper we assume that their functional form is known and we focus on the problem of parametric inference.

In practice we can only observe a discrete skeleton of the diffusion V. Depending on what kind of data we observe, we can further classify the diffusion models into two categories. The first category includes models where we have observations on every coordinate of the vector V. In the second category, the process is divided into two components, say X and α, and we only observe points of X. In this paper we focus on a subclass of the second category, henceforth denoted by $\mathbb{C}$, considering multivariate diffusions with unobserved paths that satisfy the following SDE ($0 \leq t \leq T$):

$$\begin{pmatrix} dX_t \\ d\alpha_t \end{pmatrix} = \begin{pmatrix} \mu_x(\alpha_t, \theta) \\ \mu_\alpha(\alpha_t, \theta) \end{pmatrix} dt + \begin{pmatrix} \sigma_x(\alpha_t, \theta) & 0 \\ 0 & \sigma_\alpha(\alpha_t, \theta) \end{pmatrix} \begin{pmatrix} dB_t \\ dW_t \end{pmatrix}, \tag{2}$$

where B and W denote standard Brownian motions that can potentially be correlated. The class $\mathbb{C}$ includes many interesting diffusions. For instance it contains most diffusion driven stochastic volatility models which are particularly useful in financial applications; see for instance Ghysels et al. (1996) and Shephard (2005). Famous examples of stochastic volatility models in $\mathbb{C}$ are the models introduced in Heston (1993), Stein and Stein (1991) and Hull and White (1987)*. Section 4 provides an example based on the model of Heston (1993).

A crucial difference between the models of these two categories is the availability of the Markov property. Suppose that we observe V (or X accordingly) at times $\{t_k,\ 0 \leq t_k \leq T,\ k = 0, 1, \ldots, n\}$ and let $Y = \{Y_k = V_{t_k}\}$. In the first category we can write the likelihood using the transition density of the diffusion (conditional on the initial point Y_0):

$$p_\theta(Y) = \prod_{k=1}^{n} p_\theta(Y_k | Y_{k-1}). \tag{3}$$

However, this is not always true for the models of the second category and consequently for the diffusions in $\mathbb{C}$. In both cases the likelihood is not generally available in closed form and the problem of inference is quite complicated. As a result of this, the literature contains various methodologies that may or not be based on the likelihood; see Sørensen (2004) for an extensive review. Likelihood based approaches are either analytical (Aït-Sahalia (2002), Aït-Sahalia (2005)), or they use simulations (Pedersen (1995), Durham and Gallant (2002)).

*It is easier to see this if we use $S = \exp(X)$ instead of X

They usually approximate the likelihood in a way so that the discretisation error can become arbitrarily small, although the methodology developed in Beskos et al. (2006) succeeds exact inference in the sense that it allows only for Monte Carlo error. Unfortunately, all of the above rely on the Markov property and therefore become hard to generalize to the non-Markovian case.

A natural way to proceed is via data augmentation, a methodology introduced by Tanner and Wong (1987). The idea is based on the fact that the likelihood can always be well approximated given the entire path of V or a sufficiently fine partition of it. Therefore, the unobserved paths of V are treated as missing data and a finite number of points, large enough to make the approximation error arbitrarily small, is imputed. Elerian et al. (2001), Eraker (2001) and Jones (2003) use Markov chain Monte Carlo (MCMC) approaches and approximate the posterior through the Euler-Maruyama approximation of the transition density (3). As noted in Roberts and Stramer (2001) however, there exists a strong dependence between the imputed sample paths and the volatility coefficients. In fact the algorithm becomes reducible as the number of imputed points increases.

Roberts and Stramer (2001) tackle the problem for scalar diffusions by a reparametrisation on the paths of V and Kalogeropoulos et al. (2007a) offer an extension for some multivariate diffusions. As explained in section 2.1 however, this framework does not cover the models in $\mathbb{C}$. This paper focuses on this class and we introduce a novel reparametrisation of the likelihood that may serve as the basis for data augmentation schemes. Alternative approaches to this problem can be found in Chib et al. (2005) and Golightly and Wilkinson (2005).

The paper is organized as follows. Next section elaborates on the need for a reparametrisation and provides a likelihood that may serve as the basis for inference purposes. Section 3 presents the details of a data augmentation scheme that can handle models in $\mathbb{C}$. In section 4, the methodology is applied to simulated data. We highlight the necessity of the reparametrisation in a simple stochastic volatility model and we perform a simulation study based on the model of Heston (1993) to illustrate the proposed methodology. Finally we conclude in section 5 with some relevant discussion.

2 Reparametrisation

2.1 The need for a reparametrisation

Without loss of generality we will consider a slightly simpler version of (1), i.e. a scalar diffusion with constant volatility:

$$dV_t = \mu(V_t, \theta)dt + \sigma dB_t,\ V_0 = y_0,\ 0 \leq t \leq T \tag{4}$$

In practice we have observations of V at a finite set of times and we want to draw inferences for θ and σ based on them. Under the hypothetical scenario that the entire path V is observed, we can identify σ through the quadratic variation process:

$$T\sigma^2 \ = \ \int_0^T (dV_t)^2 \ a.s. \tag{5}$$

We can then see the likelihood for θ as the Radon-Nikodym derivative of the law of V with respect to that of a Brownian motion with volatility σ^2. This is provided by Girsanov's formula (see for instance chapter 8 of Oksendal (2000)), yet in most cases a closed form solution is not available. Given observations on a sufficiently fine scale we can use approximations but such data are not available by definition for diffusions with unobserved paths.

An alternative option is to apply a data augmentation scheme and make inferences for θ, σ and the unobserved diffusion paths. Nevertheless, as the number of imputed points increases, the posterior density of σ^2 conditional on V is just a point mass, so that (5) is satisfied (Roberts and Stramer, 2001). Hence, any data augmentation scheme based on the diffusion V will result in a reducible MCMC algorithm; the sample paths of V will force σ to remain at its current value and never converge. Another way to see the problem is to note that the measures $\{\mathbb{W}_\sigma, \sigma \in \mathbb{R}\}$ are mutually singular and therefore so are $\{\mathbb{P}_\sigma, \sigma \in \mathbb{R}\}$. Section 4.1 demonstrates this problem using simulated data from a simple stochastic volatility model. Note that imputation of arbitrarily many points is vital for limiting the approximation error.

The problem may be resolved if we apply a transformation on V, so that the algorithm based on the transformed diffusion is no longer reducible. In the spirit of Papaspiliopoulos et al. (2003), this can be seen as a non-centering reparametrisation. Next subsection contains the relevant details and introduces a suitable reparametrisation for the diffusions in class $\mathbb{C}$.

2.2 A suitable likelihood parametrisation

Roberts and Stramer (2001) proposed a reparametrisation of the likelihood for general scalar

diffusions that breaks down the dependence between the sample paths and the volatility coefficients. They note that the likelihood is initially written with respect to a dominating measure that depends on parameters and they introduce a two-step transformation under which the dominating measure is parameter-free. The first step of this transformation, denoted by $h(V_t, \theta)$, is given by the solution to the following differential equation:

$$\frac{\partial h(V_t, \theta)}{\partial V_t} \;=\; \{\sigma(V_t, \theta)\}^{-1} \tag{6}$$

As noted in Kalogeropoulos et al. (2007a), in more than one dimensions the above extends to the following system of partial differential equations (PDEs):

$$\nabla h(V_t, \theta)\; \sigma(V_t, \theta)\, [\nabla h(V_t, \theta)\; \sigma(V_t, \theta)\}]' \;=\; I_d,$$

where I_d denotes the identity matrix of dimension d and A' is the transpose of a matrix A. It is not guaranteed that the above system of PDEs will have a solution. In fact, as proved in Aït-Sahalia (2005), the system is inconsistent for stochastic volatility models and therefore such a transformation does not exist for diffusions in $\mathbb{C}$. Hence we take a slightly different route.

Consider a diffusion in $\mathbb{C}$ and denote the observations of the process X at times $t_k,\ k = 1 \ldots, n$ by Y, $Y = \{X_{t_k},\ 0 \leq t_k \leq T\} = \{Y_k,\ k = 1, \ldots, n\}$. For simplicity we assume that the diffusion in (2) is two-dimensional (we discuss about extensions to higher dimensions in the next subsection):

$$dX_t = \mu_x(\alpha_t, \theta)dt + \rho\sigma_x(\alpha_t, \theta)dW_t + \sqrt{1-\rho^2}\sigma_x(\alpha_t, \theta)dB_t$$

$$d\alpha_t = \mu_\alpha(\alpha_t, \theta)dt + \sigma_\alpha(\alpha_t, \theta)dW_t$$

where B and W are independent Brownian motions. The parameter ρ is the correlation between the instantaneous increments of X and α and reflects what is termed as *leverage effect* in the stochastic volatility literature. Let $\mathbb{P}_\theta(X, \alpha)$ denote the law of $(X, \alpha)'$. Clearly we can write:

$$\mathbb{P}_\theta(X, \alpha) \;=\; \mathbb{P}_\theta(\alpha)\; \mathbb{P}_\theta(X|\alpha),$$

where $\mathbb{P}_\theta(\alpha)$ and $\mathbb{P}_\theta(X|\alpha)$ denote the laws of α and X given α respectively. Note that given the path of the unobserved process and its parameters, α and W become deterministic

functions of time. Looking again at the SDE of X we note that $\mathbb{P}_\theta(X|\alpha)$ depends only on the observations Y the density of which is

$$p_\theta(Y|\alpha) \ = \prod_{k=1}^{n} p_\theta\left(Y_k \,|Y_{k-1}, \{\alpha_t : t_{k-1} \le t \le t_k\}\right).$$

where

$$p_\theta\left(Y_k \,|Y_{k-1}, \{\alpha_t : t_{k-1} \le t \le t_k\}\right) \sim N\left(\mu_k,\ (1-\rho^2)\int_{t_{k-1}}^{t_k} \sigma_x(\alpha_s,\theta)^2 ds\right)$$

with

$$\mu_k = Y_{k-1} + \int_{t_{k-1}}^{t_k} \mu_x(\alpha_s,\theta)ds + \rho\int_{t_{k-1}}^{t_k} \sigma_x(\alpha_s,\theta)dW_s.$$

Denote by $\mathbb{Q}_\theta(\alpha)$ the distribution of the driftless version of α. Combining all of the above, we can attempt to write down the likelihood as

$$\frac{d\mathbb{P}_\theta}{d\mathbb{Q}_\theta}(X,\alpha) \ = \ p_\theta(Y|\alpha)\ \frac{d\mathbb{P}_\theta(\alpha)}{d\mathbb{Q}_\theta(\alpha)} \tag{7}$$

The dominating measure $\mathbb{Q}_\theta(\alpha)$ in the parametrisation of (7) corresponds to a diffusion with volatility $\sigma_\alpha(\alpha_t,\theta)$ and therefore clearly depends on θ. For this reason we are going to introduce a transformation $\beta_t = h(\alpha_t,\theta)$ that originates from a differential equation, similar to (6):

$$\frac{\partial h(\alpha_t,\theta)}{\partial \alpha_t} \ = \ \{\sigma_\alpha(\alpha_t,\theta)\}^{-1}$$

Applying Ito's lemma we note that the transformed process β_t has unit volatility and drift:

$$\mu_\beta(\beta,\theta) = \frac{\mu_\alpha\left[h^{-1}(\beta,\theta),\theta\right]}{\sigma_\alpha\left[h^{-1}(\beta,\theta),\theta\right]} - \frac{1}{2}\frac{\partial\sigma_\alpha\left[h^{-1}(\beta,\theta),\theta\right]}{\partial h^{-1}(\beta,\theta)}.$$

Given the initial point of β ($\beta_0 = h(\alpha_0,\theta)$), Girsanov's formula now provides the Radon-Nikodym derivative between the law of β and that of a Brownian motion starting at β_0. This is still problematic however, as this law depends on parameters (β_0 is a function of θ). For this reason we introduce a second transformation:

$$\gamma_t = \beta_t - \beta_0, \quad \beta_t = \eta(\gamma_t)$$

The process γ_t will have unit volatility and drift $\mu_\gamma(\gamma,\theta) = \mu_\beta(\eta(\gamma),\theta)$. Now we can use Girsanov's formula for the Radon-Nikodym derivative between the law of γ, denoted by $\mathbb{P}_\theta(\gamma)$, and that of a standard Brownian motion starting at 0 ($\mathbb{W}$):

$$\frac{d\mathbb{P}_\theta(\gamma)}{d\mathbb{W}} = G(\gamma,\theta) = \exp\left(\int_0^T \mu_\gamma(\gamma_s,\theta)d\gamma_s - \frac{1}{2}\int_0^T \mu_\gamma^2(\gamma_s,\theta)ds\right).$$

We are finally in a position to write down the likelihood with respect to a parameter-free dominating measure:

$$\frac{d\mathbb{P}_\theta}{d\mathbb{Q}_\theta}(X,\gamma) \;= p_\theta(Y|\gamma)\,\frac{d\mathbb{P}_\theta(\gamma)}{d\mathbb{W}} \;=\; p_\theta(Y|\gamma)\,\pi(\alpha_0)\,G(\gamma,\theta) \tag{8}$$

where $p_\theta(Y|\gamma) = p_\theta(Y|\alpha) = p_\theta\left\{Y|\eta\circ h^{-1}(\gamma,\theta)\right\}$ and $\pi(\alpha_0)$ has to be specified as part of the model. If available, the stationary distribution of α is a natural choice. Using this likelihood we can construct an irreducible data augmentation scheme that can be used for inference purposes. Given a sufficiently fine partition of the path of γ, the Ito and path integrals in the likelihood may be calculated numerically. The relevant discretisation error may become arbitrarily small as we impute more points. More details on this scheme and its practical implementation are provided in the next section.

2.3 Multifactor and Multivariate Stochastic Volatility Models

In the previous subsection for illustration processes we made the assumption that both processes X and α are one-dimensional. This assumption is not necessary; we can still use such a parametrisation of the likelihood in models where either or both X_t and α are vector processes. Famous examples of such models are that of multivariate and multifactor stochastic volatility used for example in Duffie and Kan (1996) and Chernov et al. (2003). We should keep in mind however, that this likelihood parametrisation requires the existence of a transformation of α, $h(\alpha,\theta)$, to a diffusion with identity volatility matrix. Therefore it is only applicable to models with such $\sigma_\alpha(\alpha,\theta)$, so that the system of PDEs:

$$\nabla h(\alpha_t,\theta)\;\sigma_\alpha(\alpha_t,\theta)\left[\nabla h(\alpha_t,\theta)\;\sigma_\alpha(\alpha_t,\theta)\right]' \;=\; I_d$$

has a solution. The methodology contains most cases of practical interest in multifactor and multivariate stochastic volatility. For example it contains models with a constant volatility matrix for α or with conditionally independent volatilities ($\alpha = (\alpha_1,\dots\alpha_d)'$, $\sigma_\alpha(\alpha,\theta) \equiv diag\left\{f(\alpha_i,\theta)\right\}$, $i = 1,\dots,d$). Extensions to more general models are possible by diagonalizing $\sigma_\alpha(.)$

3 Data augmentation scheme

This section presents a MCMC algorithm that can be used to sample from the posterior densities of the parameter vector θ and the unobserved paths of γ. The model is formulated in continuous time but in practice we impute the values of γ at a finite set of times. Denote by m the number of imputed points between any two observations. We have to choose a large enough m so that the error due to the discretisation of γ_t is sufficiently small. Roberts and Stramer (2001) use the stability of the likelihood estimate as a diagnostic for the fine-ness of the discretisation. The idea is that if the likelihood estimates for two different numbers of m are approximately the same, the discretisations are likely to be sufficiently fine.

Updating θ is relatively straightforward as we can get the relevant conditional posterior densities using the likelihood defined in (8). The parameter vector can be divided into two components $\theta = (\theta_1, \theta_2)$, depending on the part of likelihood they appear. θ_1 contains the parameter involved in the μ_x and σ_x as well as ρ and α_0. θ_2 contains the parameters in $\mu_\alpha(.)$ and $\sigma_\alpha(.)$. Let $\pi(\theta_i),\ i = 1, 2$ denote the corresponding priors. The conditional posterior densities $\pi(\theta_i|Y)$ will then be:

$$\pi(\theta_1|Y) \propto p_\theta(Y|\gamma)\pi(\theta_1)$$

$$\pi(\theta_2|Y) \propto p_\theta(Y|\gamma)G(\gamma, \theta)\pi(\theta_3)$$

For some models it may be possible to identify the conditional posterior densities above and apply Gibbs steps. Otherwise ordinary random walk Metropolis updates may be used.

In the remainder of this section we will focus on the updates of the diffusion paths of γ, a clearly more complicated task. We proceed using an independence sampler, proposing for instance from Brownian motion. Updating the entire path could lead to extremely low acceptance rates as the discrepancy between the proposed and current path is substantial. A way to circumvent this problem is to split the process into blocks and update each one in turn using diffusion bridges (i.e. diffusions conditioned to finish at a particular point). By doing so, we increase the acceptance rate of the moves since the discrepancy between the proposed and current moves is significantly lower. For more details see Elerian (1999). On the other hand, Elerian also notes that smaller blocks result in slower convergence and therefore some tuning is required to get the optimal results.

While it is rather clear that the path of γ should be divided into blocks, it is not straightforward how this should be done. Suppose that we observe Y at times t_k, as in section 2.2,

and that we split the path of γ into n blocks $\{b_k = \gamma_s, \quad t_{k-1} \leq s \leq t_k, \quad k = 1, 2, \ldots, n\}$. But under this formulation the endpoints of the blocks are not updated at all leading to a reducible MCMC chain; an alternative blocking scheme is needed. The blocking strategy adopted in this paper, uses block updating of overlapping segments. This scheme has been also used in Roberts and Stramer (2004) in a slightly different context. Under this procedure, we update γ_s for $t_i \leq s \leq t_{i+c}$ for $i = 0, 1, \ldots, n-c$, where c is an integer smaller than n. It is our experience that high values of c improve the mixing of the algorithm as long as the acceptance rate of the blocks is not too small. We use an independence sampler with a Brownian bridge as the proposal distribution. The MCMC algorithm becomes:

1. Set $i = 0$.
2. Propose a Brownian bridge starting at $\gamma(t_i)$ and finishing at $\gamma(t_{i+c})$. Denote it by γ_s^*, $t_i \leq s \leq t_{i+c}$.
3. Accept γ_s^* with probability:
$$min\left(1, \frac{G(\gamma_s^*, \theta)\pi_\theta(X_s|\gamma_s^*)}{G(\gamma_s, \theta)\pi_\theta(X_s|\gamma_s)}\right)$$
4. Set $i = i + 1$ (until $i \leq n - c$). Note that for $i = n - c$ the proposal should be just Brownian motion rather than a Brownian bridge.

An alternative blocking strategy uses random sized blocks. More specifically at each iteration the path is randomly split into blocks and the paths between the endpoints are updated, whereas the endpoints remain unchanged until the next iteration. See Elerian's thesis for more details. This scheme was adopted in Chib et al. (2005) and found to perform well.

4 Simulations

The simulations performed in this section aim to demonstrate two aspects of the problem. First, we highlight the necessity of the reparametrisation introduced in section 2. This is done in section 4.1 by exposing the problem in the case of a very simple stochastic volatility model. In section 4.2 we present the proposed methodology using simulated data from the Heston model.

4.1 Data augmentation without reparametrisation

We simulated data from the following stochastic volatility model of $\mathbb{C}$:

$$\begin{pmatrix} dX_t \\ d\alpha_t \end{pmatrix} = \begin{pmatrix} \exp(\alpha_t/2)dB_t \\ \sigma dW_t \end{pmatrix}, \quad X_0 = \alpha_0 = 0, \quad \sigma > 0, \quad 0 \leq t \leq 100.$$

where B and W are independent Brownian motions. Note that in this case $\mu_x \equiv \mu_\alpha \equiv 0$. Therefore, unless we apply a reparametrisation, Girsanov's formula is not useful. Alternatively we may use the Euler-Maruyama approximation; see chapter 9 of Kloeden and Platen (1995). Suppose that we observe X at times $\{t_k = k, k = 0, 1, \ldots, n\}$, $n = 100$ and that we impute the corresponding values for α. Furthermore we impute m values of α between every pair of successive times with observations. For simplicity we assume that the imputed points are equidistant and denote the time interval between them by $\delta = (m+1)^{-1}$. Let $V_t = (X_t, \alpha_t)'$ and $\Sigma = diag\{\exp(\alpha_t), \sigma^2\}$. Under the Euler-Maruyama approximation and given V_0 we get:

$$\pi(Y, \sigma^2, V_t) = \prod_{t=1}^{n(m+1)+1} \pi(V_t|V_{t-1}, \sigma^2), \quad \pi(V_t|V_{t-1}, \sigma^2) \sim N\left(V_{t-1},\ \delta\Sigma_{t-1}\right),$$

If we assign $\pi(\sigma^2) \propto \sigma^{-2}$ as the prior for σ^2 and assume that α_0 is known, we get that its conditional posterior density is an Inverse-Gamma distribution with parameters:

$$a = \frac{n(m+1)}{2},\ b = \frac{(m+1)\sum_{t=1}^{n(m+1)+1}(\alpha_t - \alpha_{t-1})^2}{2}$$

We ran a MCMC chain for 100,000 iterations for different numbers of imputed points ($m = 1, 10, 40, 100$), updating the paths as described in section 3 and using the Gibbs step for the updates σ^2. We used the overlapping blocks scheme for the updates of the paths. The length of the block was chosen as $c = 5$ to improve autocorrelation for the posterior draws of σ^2 and the acceptance rate of the blocks was approximately 78%. Figure 1*b* shows the autocorrellation of the posterior draws of σ^2 for each value of m. Clearly, the autocorrelation increases dramatically leading to an increasingly slower chain. An alternative way to see this is to note that the variance of the conditional posterior for σ^2 goes to 0 as we increase m.

The problem can be resolved if we apply this paper's proposed reparametrisation. Following the route of section 2.2, we set $\beta_t = \alpha_t/\sigma$ and $\gamma_t = \beta_t - \beta_0 = \beta_t$ and we get

$$dX_t = \exp(\sigma\beta_t/2)dBt,$$

where β is a standard Brownian motion independent of B. Note that the part with Girsanov's formula drops out of the likelihood which now simplifies to:

$$\pi(Y, \sigma^2, \beta_t) = \pi(\beta_0) \prod_{k=1}^{n} \pi(Y_k | Y_{k-1}, \sigma^2, \beta_t),$$

where

$$\pi(Y_k | Y_{k-1}, \sigma^2, \beta_t) \sim N \left\{ Y_{k-1}, \int_{t_{k-1}}^{t_k} \exp(\sigma \beta_s) ds \right\}.$$

Figure 1*a* contains the corresponding autocorrelation plots of the posterior draws of σ^2 taken from the reparametrised data augmentation scheme. Unlike the previous case (figure 1*b*) there is clearly no increase in the autocorrelation.

[Figure 1 about here.]

4.2 Reparametrised scheme applied to the Heston model

In this section we illustrate the proposed methodology to simulated data from the model of Heston (1993). We also demonstrate the immunity of the algorithm developed in sections 2 and 3 to the increase of the number of imputed points m. If we take the log of the observed process the Heston model can be written as:

$$dX_t = \mu_x dt + \left\{ (1 - \rho^2) \alpha_t \right\}^{1/2} dB_t + \rho \, {\alpha_t}^{1/2} dW_t$$

$$d\alpha_t = \kappa(\mu - \alpha_t) dt + \sigma {\alpha_t}^{1/2} dW_t$$

where B_t and W_t are independent Brownian motions and $Corr(dX_t, d\alpha_t) = \rho dt$ as before. In accordance to section 2.2 we apply the following 2-step transformation:

1. $\beta_t = h(\alpha_t) = 2{\alpha_t}^{1/2}/\sigma, \;\; \beta_t > 0$

2. $\gamma_t = \beta_t - 2{\alpha_0}^{1/2}/\sigma, \;\; \beta_t = \eta(\gamma_t)$

Then using Ito's lemma we get ($g := \eta \circ h^{-1}$):

$$dX_t = \mu_x dt + \left\{ (1 - \rho^2) g(\gamma_t) \right\}^{1/2} dB_t + \rho \, g(\gamma_t)^{1/2} dW_t$$

$$d\gamma_t = \left\{ \frac{2\kappa\mu - 0.5\sigma^2}{\sigma^2 \eta(\gamma_t)} - 0.5\kappa\eta(\gamma_t) \right\} dt + dW_t$$

We can now proceed by writing down the likelihood as in section 2.2 and implementing a data augmentation scheme as in section 3. Using Brownian bridges as proposals for the paths of γ is not the best choice given the constraint $\beta_t > 0$. Alternatively we may choose to update the paths of α instead, since they are linked with a deterministic function with the paths of γ given σ. We may propose from the diffusion Z that satisfies the SDE:

$$dZ_t = \frac{\sigma^2}{4}dt + \sigma {Z_t}^{1/2} dB_t$$

To simulate bridges from Z we can first simulate a Brownian bridge BB_t with volatility $\sqrt{\sigma/2}$ and then set $Z = \frac{1}{2}\sigma {BB_t}^2$. The parameter values used in the simulation are similar to those obtained from the analysis of the closing prices of Standard and Poor' 500 index in Chib et al. (2005) and Aït-Sahalia and Kimmel (2005).

We simulated 1008 data points (excluding the initial point) from the Heston model corresponding to 1008 working days or 4 years of data. In accordance with the relevant literature, we set the initial values to $X_0 = log(100)$ and $\alpha_0 = \mu = 0.9$. As in the previous section, we ran a MCMC chain for 80,000 iterations for different numbers of imputed points ($m = 20, 40, 80$). We chose the value of $c = 8$ to achieve lower autocorrelation on the parameter posterior draws. The acceptance rate for the blocks was approximately 32%. Figure 2 shows the autocorrelation plot for the posterior draws of σ for different values of m. There is no evidence of any increase whatsoever, even for $m = 80$. In figure 3, we see the posterior densities of the log-likelihood and σ, again for different values of m. All the densities are similar, providing strong evidence that the discretisation is sufficiently fine. Finally table 1 provides the posterior means and standard deviations of the parameters. We see that these estimates are in good agreement with the values we simulated the data from.

[Table 1 about here.]

[Figure 2 about here.]

[Figure 3 about here.]

5 Discussion

The methodology developed in this paper allows for likelihood-based inference in a class of multivariate diffusions with unobserved paths. This class includes diffusion driven stochastic volatility models where the problem of inference is particularly difficult due to the lack of

the Markov property. This is accomplished via a MCMC algorithm that is relatively simple and fast to implement as illustrated in sections 3 and 4. Under the proposed framework it is quite straightforward to limit the discretisation error by simply increasing the number of imputed points. The reparametrised scheme of this paper achieves that leaving both the autocorrelation of the posterior draws and the irreducibility of the chain, intact.

Alternative data augmentation schemes for the diffusions in $\mathbb{C}$ can be found in Chib et al. (2005) and Golightly and Wilkinson (2005). Apart from the different blocking strategy, the methodology in Chib et al. (2005) is based on a reparametrisation under which the paths of the unobserved process are transformed to W rather than γ. This scheme also contains Laplace approximation proposals for the updates of the parameters of α. The work in Golightly and Wilkinson (2005) uses bayesian sequential techniques and joint updates of the diffusion paths and the parameters in the volatility functions.

As in most data augmentation schemes for diffusions, an independence sampler was used for the block updates of the paths. A disadvantage of this method is that the acceptance rate of the sampler will be small for larger blocks and consequently the algorithm will deteriorate. The use of smaller blocks is an option, but it may slow down the mixing of the chain. A more sophisticated choice for the proposal of the sampler, for instance a linear diffusion bridges, may improve the performance of the MCMC chain.

The class of diffusions considered in this paper does not include cases, where the drift and the volatility of the observed process depend on the process itself; for instance stochastic volatility models for interest rates with mean reverting drift. The reason for this is that the distribution of the observed process given the unobserved is no longer available in closed form. The data augmentation scheme of Kalogeropoulos et al. (2007b) handle such models using time change transformations of the observed process. An alternative option is provided by Golightly and Wilkinson (2005).

6 Acknowledgments

This work was supported by the project Irakleitos - Fellowships for Research of the Athens University of Economics and Business (which is co-financed by the European Union) and by the EPSRC grant GR/S61577/01. The author would particularly like to thank Gareth Roberts, Petros Dellaportas, Antonio Pievatolo and the referees for their very helpful and constructive comments.

References

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

Aït-Sahalia, Y. (2005). Closed form likelihood expansions for multivariate diffusions. In preparation.

Aït-Sahalia, Y. and Kimmel, R. (2005). Maximum likelihood estimation for stochastic volatility models. To appear in the Journal of Financial Economics.

Beskos, A., Papaspiliopoulos, O., Roberts, G., and Fearnhead, P. (2006). Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(3):333–382.

Chernov, M., Gallant, A., Ghysels, E., and Tauchen, G. (2003). Alternative models for stock price dynamics. *Journal of Econometrics*, 116:225–258.

Chib, S., Pitt, M. K., and Shephard, N. (2005). Likelihood based inference for diffusion models. Submitted.

Duffie, D. and Kan, R. (1996). A yield factor model of interest rates. *Mathematical Finance*, 6:379–406.

Durham, G. B. and Gallant, A. R. (2002). Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes. *J. Bus. Econom. Statist.*, 20(3):297–316. With comments and a reply by the authors.

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

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

Eraker, B. (2001). Markov chain Monte Carlo analysis of diffusion models with application to finance. *J. Bus. Econom. Statist.*, 19(2):177–191.

Ghysels, E., Harvey, A., and Renault, E. (1996). Stochastic volatily, in. *Handbook of Statistics 14, Statistical Methods in Finance*. G.S. Maddala and C.R. Rao (eds), North Holland, Amsterdam.

Golightly, A. and Wilkinson (2005). Bayesian sequential inference for nonlinear multivariate diffusions. Submitted.

Heston, S. (1993). A closed-form solution for options with stochastic volatility. with applications to bonds and currency options. 6:327–343.

Hull, J. C. and White, A. D. (1987). The pricing of options on assets with stochastic volatilities. *Journal of Finance*, 42(2):281–300.

Jones, C. S. (2003). Nonlinear mean reversion in the short-term interest rate. *The review of financial studies*, 16:793–843.

Kalogeropoulos, K., Dellaportas, P., and Roberts, G. (2007a). Bayesian inference for correllated diffusions observed discretly in time. In submision.

Kalogeropoulos, K., Roberts, G., and Dellaportas, P. (2007b). A unified inference framework for diffusion driven stochastic volatility models. In submission.

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

Oksendal, B. (2000). *Stochastic differential equations*. Springer, 5th edition.

Papaspiliopoulos, O., Roberts, G., and Skôld (2003). Non-centered parametrisations 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. R. (1995). A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. *Scand. J. Statist.*, 22(1):55–71.

Roberts, G. and Stramer, O. (2001). On inference for partial observed nonlinear diffusion models using the metropolis-hastings algorithm. *Biometrika*, 88(3):603–621.

Roberts, G. and Stramer, O. (2004). On bayesian analysis of non-linear continuous time autoregressive models. Submitted.

Shephard, N. (2005). *Stochastic Volatility: Selected Readings*. Oxford University Press.

Sørensen, H. (2004). Parametric inference for diffusion processes observed at discrete points in time: a survey. *International Statistical Review*, 72(3):337–354.

Stein, E. M. and Stein, J. C. (1991). Stock proce distributions with stochastic volatility: an analytic approach. *Review of Financial Studies*, 4(4):727–752.

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

List of Figures

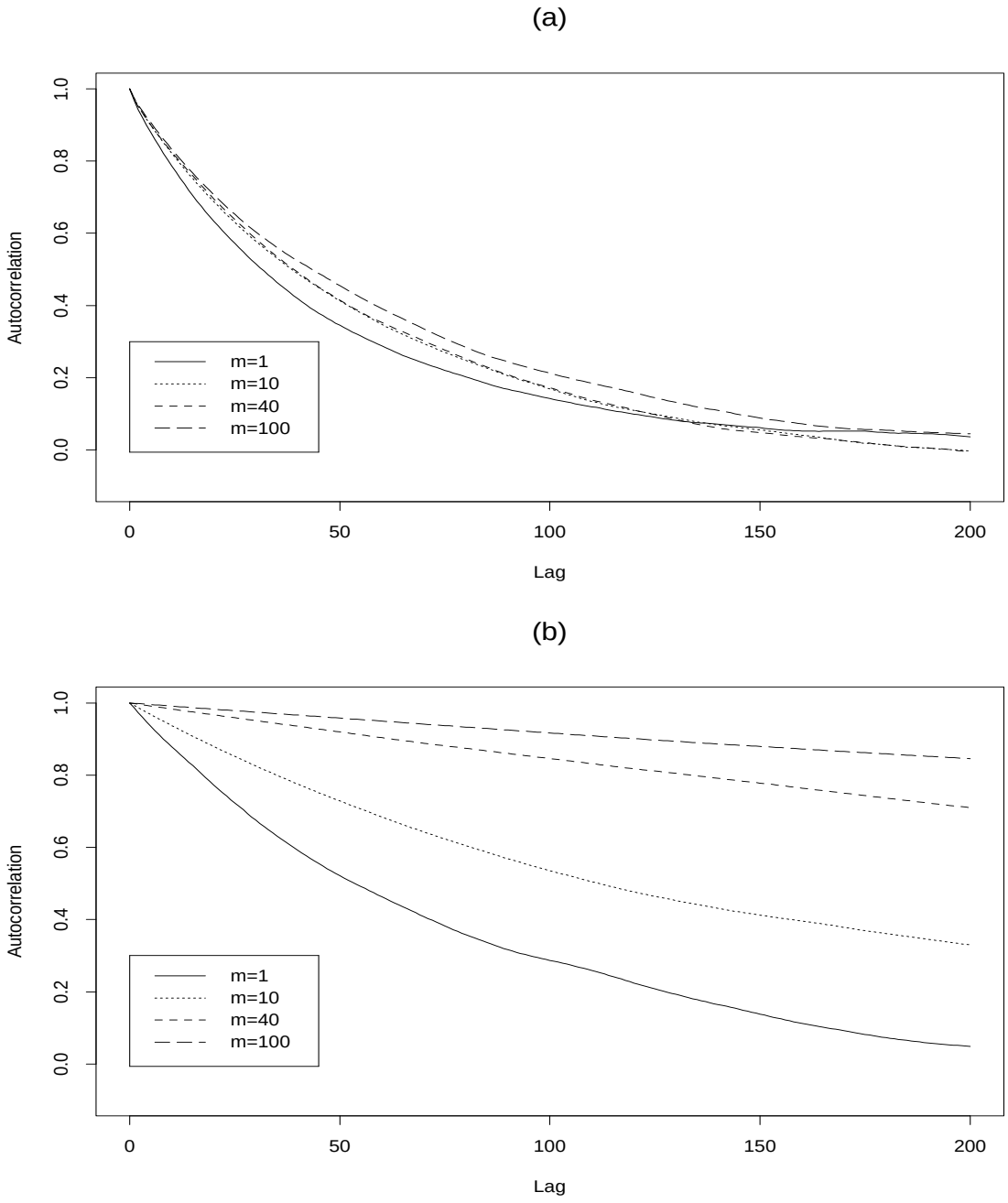

Figure 1: Autocorrelation plots of posterior draws of σ^2 for different values of imputed points between observations (m) for the simple stochastic volatility model. The draws in (a) correspond to the reparametrised scheme scheme and in (b) to the scheme without transformation.

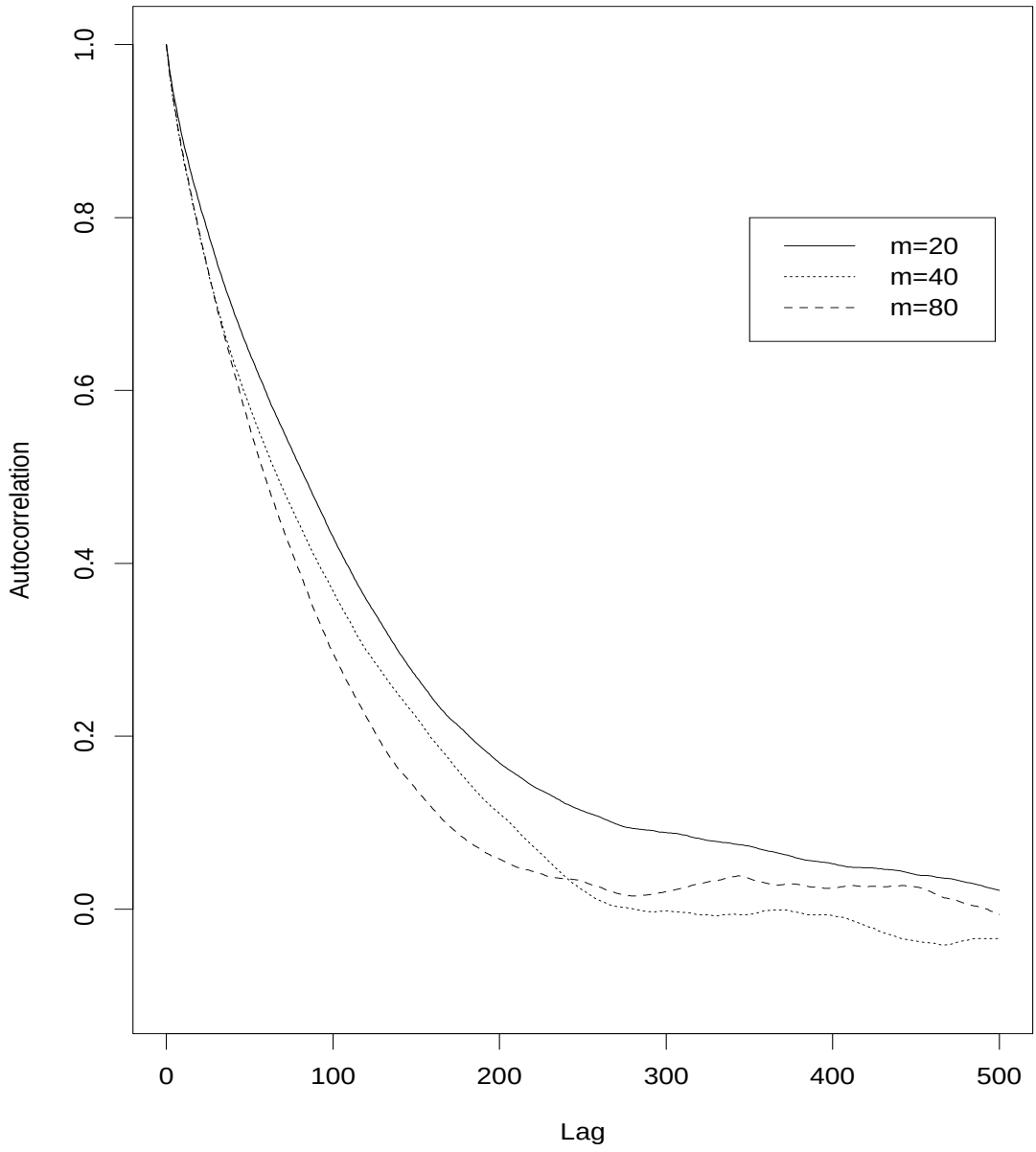

Figure 2: Autocorrelation plots of posterior draws of σ for different values of imputed points between observations (m) for the Heston model.

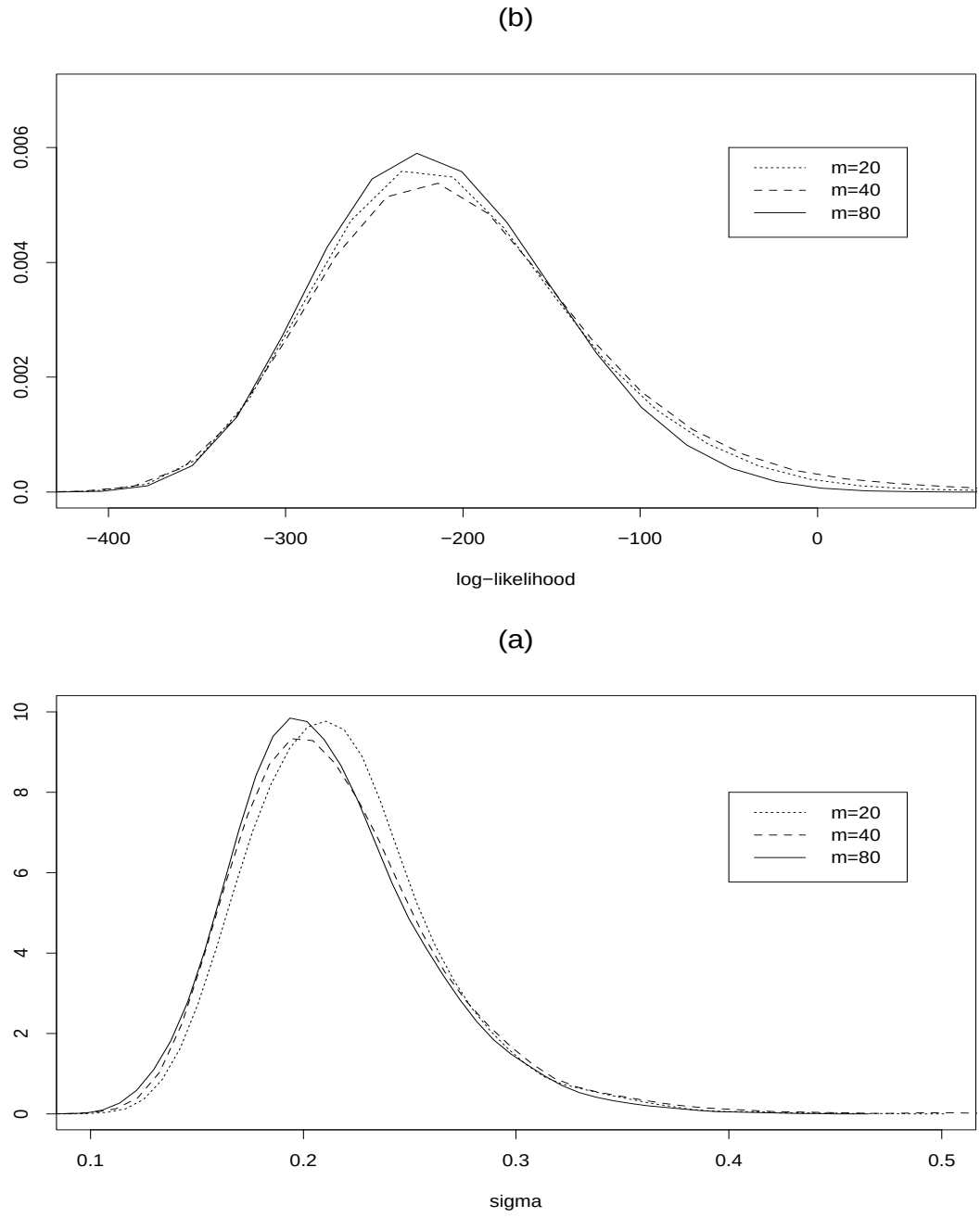

Figure 3: Posterior densities of (a) log-likelihood and (b) σ for different values of imputed points between observations (m) for the Heston model.

List of Tables

Parameter	κ	μ	σ	ρ	μ_x
True value	0.1	0.9	0.2	-0.5	0
Posterior mean	0.0915	0.8712	0.2126	-0.479	0.0187
Posterior SD	0.0296	0.0863	0.0436	0.1134	0.0278

Table 1: Posterior means and standard deviations of the parameters versus their true values.