

Konstantinos Kalogeropoulos **Inference for diffusion models using time change transformations**

**Article (Accepted version)
(Refereed)**

Original citation:

Kalogeropoulos, Konstantinos and Roberts, Gareth O. and Dellaportas, Petros (2010) Inference for diffusion models using time change transformations. [Annals of statistics](#), 38 (2). pp. 784-807.

DOI: [10.1214/09-AOS702](https://doi.org/10.1214/09-AOS702)

© 2007 [IMS](#)

This version available at: <http://eprints.lse.ac.uk/31421/>

Available in LSE Research Online: January 2011

LSE has developed LSE Research Online so that users may access research output of the School. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users may download and/or print one copy of any article(s) in LSE Research Online to facilitate their private study or for non-commercial research. You may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain. You may freely distribute the URL (<http://eprints.lse.ac.uk>) of the LSE Research Online website.

This document is the author's final manuscript accepted version of the journal article, incorporating any revisions agreed during the peer review process. Some differences between this version and the published version may remain. You are advised to consult the publisher's version if you wish to cite from it.

Inference for stochastic volatility models using time change transformations

Konstantinos Kalogeropoulos, Gareth O. Roberts and Petros Dellaportas

Abstract

We address the problem of parameter estimation for diffusion driven stochastic volatility models through Markov chain Monte Carlo (MCMC). To avoid degeneracy issues we introduce an innovative reparametrisation defined through transformations that operate on the time scale of the diffusion. A novel MCMC scheme which overcomes the inherent difficulties of time change transformations is also presented. The algorithm is fast to implement and applies to models with stochastic volatility. The methodology is tested through simulation based experiments and illustrated on data consisting of US treasury bill rates.

1 Introduction

Diffusion processes provide natural models for continuous time phenomena. They are used extensively in diverse areas such as finance, biology and physics. A diffusion process is defined through a stochastic differential equation (SDE)

$$dX_t = \mu(t, X_t, \theta)dt + \sigma(t, X_t, \theta)dW_t, \quad 0 \leq t \leq T, \quad (1)$$

where W is standard Brownian motion. The drift $\mu(\cdot)$ and volatility $\sigma(\cdot)$ reflect the instantaneous mean and standard deviation respectively. In this paper we assume the existence of a unique weak solution to (1), which translates into some regularity conditions (locally Lipschitz with a linear growth bound) on $\mu(\cdot)$ and $\sigma(\cdot)$; see chapter 5 of Rogers and Williams (1994) for more details.

The task of inference for diffusion processes is particularly challenging and has received remarkable attention in the recent literature; see Sørensen (2004) for an extensive review. The main difficulty is inherent in the nature of diffusions which are infinite dimensional objects. However, only a finite number of points may be observed and the marginal likelihood of these observations is generally unavailable in closed form. This has stimulated the development of various non-likelihood approaches which use indirect inference (Gouriéroux et al., 1993), estimating functions (Bibby and Sorensen, 1995), or the efficient method of moments (Gallant and Tauchen, 1996); see also Gallant and Long (1997).

Most likelihood based methods approach the likelihood function through the transition density of (1). Denote the observations by Y_k , $k = 0, \dots, n$, and with t_k their corresponding times. If the dimension of Y_k equals that of X (for each k) we can use the Markov property to write the likelihood, given the initial point Y_0 , as:

$$\mathcal{L}(Y, \theta | Y_0) = \prod_{k=1}^n p_k(Y_k | Y_{k-1}; \theta, \Delta), \quad \Delta = t_k - t_{k-1} \quad (2)$$

The transition densities $p_k(\cdot)$ are not available in closed form but several approximations are available. They may be analytical, see Aït-Sahalia (2002), Aït-Sahalia (2008), or simulation based, see Pedersen (1995), Durham and Gallant (2002). 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. A potential downside of these methods

may be their dependence on the Markov property. In many interesting multidimensional diffusion models the observation regime is different and some of their components are not observed at all.

A famous such example is provided by stochastic volatility models, used extensively to model financial time series such as equity prices (Heston, 1993; Hull and White, 1987; Stein and Stein, 1991), or interest rates (Andersen and Lund, 1998; Durham, 2002; Gallant and Tauchen, 1998). A stochastic volatility model is usually represented by a 2-dimensional diffusion

$$\begin{pmatrix} dX_t \\ d\alpha_t \end{pmatrix} = \begin{pmatrix} \mu_x(X_t, \alpha_t, \theta) \\ \mu_\alpha(\alpha_t, \theta) \end{pmatrix} dt + \begin{pmatrix} \sigma_x(\alpha_t, \theta) & 0 \\ 0 & \sigma_\alpha(\alpha_t, \theta) \end{pmatrix} \begin{pmatrix} dB_t \\ dW_t \end{pmatrix}, \quad (3)$$

where X denotes the observed equity (stock) log-price or the short term interest rate with volatility $\sigma_x(\cdot)$, which is a function of a latent diffusion α . Note that the observed process X in (3) is not Markov; the distribution of a future stock price depends (besides the current price) on the current volatility which in turn depends on the entire price history. Nevertheless, since the 2-dimensional diffusion is Markov, state space approaches are possible and may be implemented using sequential Monte Carlo techniques Pitt and Shephard (1999), Fearnhead et al (2008). While such approaches allow for online estimation of the volatility, various implications arise regarding inference for the diffusion parameters Liu and West (2001), Stroud et al (2004).

An alternative approach to the problem adopts Bayesian inference utilising Markov chain Monte Carlo (MCMC) methods. Adhering to the Bayesian framework, a prior $p(\theta)$ is first assigned on the parameter vector θ . Then, given the observations Y , the posterior $p(\theta|Y)$ can be explored through data augmentation (Tanner and Wong, 1987), treating the unobserved paths of X (paths between observations) as missing data. Note that the augmented diffusion satisfies the Markov property irrespectively of the observation regime. Hence data augmentation approaches are more general.

Initial MCMC schemes following this programme were introduced by Jones (1999); see also Jones (2003), Eraker (2001) and Elerian et al. (2001). However, as noted in the simulation based experiment of Elerian et al. (2001) and established theoretically by Roberts and Stramer (2001), any such algorithm's convergence properties will degenerate as the number of imputed points increases. The problem may be overcome with the reparametrisation of Roberts and Stramer (2001), and this scheme may be applied in all one-dimensional and some multi-dimensional contexts. However this framework does not cover general multidimensional diffusion models. Chib et al. (2007) and Kalogeropoulos (2007) offer appropriate reparametrisations but only for a class of stochastic volatility models. Alternative reparametrisations were introduced in Golightly and Wilkinson (2007); see also Golightly and Wilkinson (2006) for a sequential approach.

In this paper we introduce a novel reparametrisation that, unlike previous MCMC approaches, operates on the time scale of the observed diffusion rather than its path. This facilitates the construction of irreducible and efficient MCMC schemes, designed appropriately to accommodate the time change of the diffusion path. Being a data augmentation procedure, our approach does not rely on the Markov property and can be applied to a much larger class of diffusions than those considered in Aït-Sahalia (2002) and Beskos et al. (2006). Moreover it may be coupled with the approaches of Roberts and Stramer (2001) and Chib et al. (2007) to handle more general models, that is almost every stochastic volatility model used in practice. The paper is organized as follows: Section 2 elaborates on the need for a transformation of the diffusion to avoid problematic MCMC algorithms. In Section 3 we introduce a class of transformations based on changing the time scale of the diffusion process. Section 4 provides the details for the corresponding non-trivial MCMC implementation. The proposed methodology of the paper is tested and illustrated through numerical experiments in section 5, and on US treasury bill rates in section 6. Finally, section 7 concludes and provides some relevant discussion.

2 The necessity of reparametrisation

A Bayesian data augmentation scheme bypasses a problematic sampling from the posterior $\pi(\theta|Y)$ by introducing a latent variable \mathcal{X} that simplifies the likelihood $\mathcal{L}(Y; \mathcal{X}, \theta)$. It usually involves the following two steps:

1. Simulate \mathcal{X} conditional on Y and θ .
2. Simulate θ from the augmented conditional posterior which is proportional to $\mathcal{L}(Y; \mathcal{X}, \theta)\pi(\theta)$.

It is not hard to adapt our problem to this setting. Y represents the observations of the price process X . The latent variables \mathcal{X} introduced to simplify the likelihood evaluations are discrete skeletons of diffusion paths between observations or entirely unobserved diffusions. In other words, \mathcal{X} is a fine partition of multidimensional diffusion with drift $\mu_X(t, X_t, \theta)$ and diffusion matrix

$$\Sigma_X(t, X_t, \theta) = \sigma(t, X_t, \theta) \times \sigma(t, X_t, \theta)',$$

and the augmented dataset is $\mathcal{X}_{i\delta}$, $i = 0, \dots, T/\delta$, where δ specifies the amount of augmentation. The likelihood can be approximated via the Euler scheme

$$\mathcal{L}^E(Y; \mathcal{X}, \theta) = \prod_{i=1}^{T/\delta} p(\mathcal{X}_{i\delta} | \mathcal{X}_{(i-1)\delta}), \quad \mathcal{X}_{i\delta} | \mathcal{X}_{(i-1)\delta} \sim \mathcal{N}(\mathcal{X}_{(i-1)\delta} + \delta\mu_X(\cdot), \delta\Sigma_X(\cdot)),$$

which is known to converge to the true likelihood $\mathcal{L}(Y; \mathcal{X}, \theta)$ for small δ (Pedersen, 1995).

Another property of diffusion processes relates $\Sigma_X(\cdot)$ to the quadratic variation process. Specifically we know that

$$\lim_{\delta \rightarrow 0} \sum_{i=1}^{T/\delta} (\mathcal{X}_{i\delta} - \mathcal{X}_{(i-1)\delta}) (\mathcal{X}_{i\delta} - \mathcal{X}_{(i-1)\delta})^T = \int_0^T \Sigma_X(s, \mathcal{X}_s, \theta) ds \quad a.s.$$

The solution of the equation above determines the diffusion matrix parameters. Hence, there exists perfect correlation between these parameters and \mathcal{X} as $\delta \rightarrow 0$. This has disastrous implications for the mixing and convergence of the MCMC chain as it translates into reducibility for $\delta \rightarrow 0$. This issue was first noted by Roberts and Stramer (2001) for scalar diffusions and also confirmed by the simulation experiment of Elerian et al. (2001). In fact the convergence time is of $O(\frac{1}{\delta})$. This problem is not MCMC specific as it turns out that the convergence of its deterministic analogue, EM algorithm, is also problematic when the amount of information in the augmented data \mathcal{X} strongly exceeds that of the observations. In our case \mathcal{X} contains an infinite amount of information for $\delta \rightarrow 0$.

The problem may be resolved if we apply a transformation so that the algorithm based on the transformed diffusion is no longer reducible as $\delta \rightarrow 0$. Roberts and Stramer (2001) provide appropriate diffusion transformations for scalar diffusions. In a multivariate context this requires a transformation to a diffusion with unit volatility matrix; see for instance Kalogeropoulos et al. (2007). Ait-Sahalia (2008) terms such diffusions as reducible and proves the non-reducibility of stochastic volatility models that obey (3). The transformations introduced in this paper follow a slightly different route and target the time scale of the diffusion. Since the construction obeys the same principles of Roberts and Stramer (2001) the convergence time of the algorithm is independent of augmentation level controlled by δ . Another appealing features of such a reparametrisation is the generalisation to stochastic volatility models.

3 Time change transformations

For ease of illustration we first provide the time change transformation and the relevant likelihood function for scalar diffusion models with constant volatility. Nevertheless, one of the main advantages of this technique is the applicability to general stochastic volatility models.

3.1 Scalar diffusions

Consider a diffusion X defined through the following SDE:

$$dX_t = \mu(t, X_t, \theta)dt + \sigma dW_t^X, \quad 0 < t < 1 \quad \sigma > 0. \quad (4)$$

Without loss of generality, we assume a pair of observations $X_0 = y_0$ and $X_1 = y_1$. For more data, note that the same operations are possible for every pair of successive observations, and these pairs are linked together through the Markov property. We introduce the latent ‘missing’ path of X for $0 \leq t \leq 1$, denoted by X^{mis} , so that $X = (y_0, X^{mis}, y_1)$. In the spirit of Roberts and Stramer (2001), the goal is to write the likelihood for θ, σ with respect to a parameter-free dominating measure. Using Girsanov’s theorem we can get the Radon-Nikodym derivative between the law of the diffusion X , denoted by \mathbb{P}^X , and that of the driftless diffusion $dM_t = \sigma dW_t^X$ which represents Wiener measure and is denoted by \mathbb{W}^X . We write

$$\frac{d\mathbb{P}^X(X)}{d\mathbb{W}^X} = G(t, X, \theta, \sigma) = \exp \left\{ \int_0^1 \frac{\mu(t, X_t, \theta)}{\sigma^2} dX_t - \frac{1}{2} \int_0^1 \frac{\mu(t, X_t, \theta)^2}{\sigma^2} dt \right\}.$$

Consider the factorisation

$$\mathbb{W}^X = \mathbb{W}_y^X \times Leb(y_1) \times f(y_1|y_0, \sigma^2), \quad (5)$$

where y_1 is a Gaussian random variable, $y_1|y_0 \sim \mathcal{N}(y_0, \sigma^2)$, and $Leb(\cdot)$ denotes Lebesgue measure. This naturally factorises the measure of X as the Lebesgue density of y_1 under the dominating measure, multiplied by the conditional dominating measure \mathbb{W}_y^X . We can now write

$$\frac{d\mathbb{P}^X(X^{mis}, y_0, y_1)}{d\{\mathbb{W}_y^X \times Leb(y)\}} = G(t, X, \theta, \sigma) f(y_1|y_0, \sigma),$$

where clearly the dominating measure depends on σ since it contains \mathbb{W}_y^X which represents a Brownian bridge with volatility σ .

To remove this dependency on the parameter σ we consider the following time change transformation. Let a new time scale be

$$s = \eta_1(t, \sigma) = \int_0^t \sigma^2 d\omega = t\sigma^2, \quad (6)$$

and then define the new transformed diffusion U as

$$U_s = \begin{cases} X_{\eta_1^{-1}(s, \sigma)}, & 0 \leq s \leq \sigma^2, \\ M_{\eta_1^{-1}(s, \sigma)}, & s > \sigma^2. \end{cases}$$

The definition for $t > \sigma^2$ is needed to ensure that U is well defined for different values of $\sigma^2 > 0$ which is essential in the context of a MCMC algorithm. Using standard time change properties, see for example Oksendal (2000), the SDE for U is

$$dU_s = \begin{cases} \frac{1}{\sigma^2} \mu(s, U_s, \theta) dt + dW_s^U & 0 \leq s \leq \sigma^2, \\ dW_s^U, & s > \sigma^2, \end{cases}$$

where W^U is another Brownian motion at the time scale s . Applying Girsanov’s theorem again, the law of U , denoted by \mathbb{P}^U , is given through its Radon-Nikodym derivative with respect to the law \mathbb{W}^U of the Brownian motion W^U at the time scale s :

$$\begin{aligned} \frac{d\mathbb{P}^U(U^{mis}, y_0, y_1)}{d\mathbb{W}^U} = G(s, U, \theta, \sigma) &= \exp \left\{ \int_0^{+\infty} \frac{\mu(s, U_s, \theta)}{\sigma^2} dU_s - \frac{1}{2} \int_0^{+\infty} \frac{\mu(s, U_s, \theta)^2}{\sigma^4} ds \right\} \\ &= \exp \left\{ \int_0^{\sigma^2} \frac{\mu(s, U_s, \theta)}{\sigma^2} dU_s - \frac{1}{2} \int_0^{\sigma^2} \frac{\mu(s, U_s, \theta)^2}{\sigma^4} ds \right\}. \end{aligned} \quad (7)$$

By applying the factorisation of (5) on \mathbb{W}^U at the new time scale s , the likelihood can be written as

$$\frac{d\mathbb{P}^U(U^{mis}, y_0, y_1)}{d\{\mathbb{W}_y^U \times Leb(y)\}} = G(s, U, \theta, \sigma)f(y_1|y_0, \sigma),$$

The dominating measure still depends on σ as it contains \mathbb{W}_y^U which reflects a Brownian bridge with conditioning event $U_{\sigma^2} = y_1$. We therefore introduce a second transformation which applies to both the path and the time scale of the diffusion process U . Consider a new time scale

$$u = \eta_2(s, \sigma) = \frac{s}{\sigma^2(\sigma^2 - s)}, \text{ or } s = \eta_2^{-1}(u, \sigma) = \frac{u\sigma^4}{1 + u\sigma^2},$$

and define a new diffusion Z through

$$U_s = (\sigma^2 - s)Z_{\eta_2(s, \sigma)} + \left(1 - \frac{s}{\sigma^2}\right)y_0 + \frac{s}{\sigma^2}y_1, \quad 0 \leq t < \sigma^2. \quad (8)$$

Note that this transformation is a bijection. In the case of $y_0 = y_1 = 0$ its inverse is given by

$$Z_u = \frac{1 + u\sigma^2}{\sigma^2}U_{\eta_2^{-1}(u, \sigma)}, \quad 0 \leq u < +\infty.$$

Applying Ito's formula and using time change properties we can also obtain the SDE of Z based on another driving Brownian motion W^Z operating at the time scale u :

$$dZ_u = \frac{\mu\left(t, \frac{\sigma^2}{1+u\sigma^2}\nu(Z_u, \sigma), \theta\right) + \nu(Z_u, \sigma)\sigma^2}{1 + u\sigma^2} dt + dW_u^Z, \quad 0 \leq u < +\infty, \quad (9)$$

where $\nu(Z_u, \sigma) = U_s$. This operation essentially transforms to a diffusion that runs from 0 to $+\infty$ preserving the unit volatility. A third attempt to write the likelihood, now based on Z , again uses the Girsanov theorem and the factorisation of (5) to condition the dominating measure on y_1 . This writes

$$\frac{d\mathbb{P}^Z(Z, y_0, y_1)}{d\{\mathbb{W}_y^Z \times Leb(y)\}} = G(u, Z, \theta, \sigma)f(y_1|y_0, \sigma), \quad (10)$$

where \mathbb{P}^Z denotes the law of Z and \mathbb{W}_y^Z reflects the law of the driftless diffusion Z conditioned on y_1 . The integrals in $G(Z, \theta, \sigma)$ run up to $+\infty$, however the expression is finite being a bijection of the Radon-Nikodym derivative between \mathbb{P}^U and \mathbb{W}^U given by (7). Using the following lemma, we can prove that \mathbb{W}_y^Z is the law of the standard Brownian motion and hence the likelihood is written with respect to a dominating measure that does not depend on any parameters.

Lemma 1: Let W be a standard Brownian motion in $[0, +\infty)$. Consider the process defined for $0 \leq t \leq T$

$$B_t = (T - t)W_{t/\{T(T-t)\}} + \left(1 - \frac{t}{T}\right)y_0 + \frac{t}{T}y_1, \quad 0 \leq t < T$$

Then B is a Brownian bridge from y_0 at time 0 to y_1 at time T .

Proof. See (Rogers and Williams, 1994, IV.40.1) for the case $y_0 = 0, T = 1$. The extension for general y_0 and T is trivial. \square

Corollary 1: The process Z is standard Brownian motion under the dominating measure. In other words \mathbb{W}_y^Z is standard Wiener measure.

Proof. Note that \mathbb{W}_y^U reflects a Brownian bridge from y_0 at time 0 to y_1 at time T and we obtained \mathbb{W}_y^Z by using the transformation of Lemma 1. Since this transformation is a bijection, U is a Brownian bridge (under the dominating measure) if and only if Z is standard Brownian motion. \square

The likelihood, given by (10), allows the construction of an irreducible MCMC scheme that alternates between updating the parameters and the diffusion process Z . For the path updates we may use the fact that

$$\frac{d\mathbb{P}_y^Z}{d\mathbb{W}_y^Z}(Z|y_0, y_1) = G(t, Z, \theta, \sigma) \frac{f(y_1|y_0, \sigma)}{f^P(y_1|y_0, \theta, \sigma)} \propto G(t, Z, \theta, \sigma), \quad (11)$$

where \mathbb{P}_y^Z is the law of Z conditioned on y_1 and $f^P(\cdot)$ is the density of y_1 under \mathbb{P}^Z . Both \mathbb{P}_y^Z and $f^P(\cdot)$ are generally unknown but their calculation may be avoided in a MCMC algorithm which essentially only uses (10) and (11). Since these expressions only require the evaluation of $f(\cdot)$ and $G(\cdot)$, which are either known or may be approximated using the augmented diffusion path, the task of Bayesian inference is feasible. Details are presented in Section 4.

3.2 Stochastic volatility models

We will first demonstrate how the case of stochastic volatility models may be brought to a similar and equivalent form to that of the previous section. Consider the general class of stochastic volatility models with SDE given by (3) for $0 \leq t \leq t_1$. Without loss of generality, we may assume a pair of observations $(X_0 = y_0, X_1 = y_1)$ due to the Markov property of the 2-dimensional diffusion (X, α) . The likelihood can then be divided into two parts: The first contains the marginal likelihood of the diffusion α and the remaining part corresponds to the diffusion X conditioned on the path of α

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

Denote the marginal likelihood for α by $\mathcal{L}_\alpha(\alpha, \theta)$. To overcome reducibility issues arising from the paths of α one may use the reparametrisations of Chib et al. (2007) or Kalogeropoulos (2007). The relevant transformations of the latter are

$$\begin{aligned} \beta_t &= h_1(\alpha_t, \theta), \quad \frac{\partial h_1(\alpha_t, \theta)}{\partial \alpha_t} = \{\sigma_\alpha(\alpha_t, \theta)\}^{-1}, \\ \gamma_t &= \beta_t - \beta_0, \quad \beta_t = h_2(\gamma_t), \end{aligned}$$

and the marginal likelihood for the transformed latent diffusion γ becomes

$$\mathcal{L}_\gamma(\gamma, \theta) = \frac{d\mathbb{P}}{d\mathbb{W}}(\gamma) = G_\gamma(t, \gamma, \theta). \quad (12)$$

where \mathbb{W} denotes Wiener measure. By letting $\alpha_t = g_t^{\theta, \gamma} = h_1^{-1}(h_2(\gamma_t), \theta)$, the SDE of X conditional on γ becomes:

$$dX_t = \mu_x(X_t, g_t^{\theta, \gamma}, \theta)dt + \sigma_x(g_t^{\theta, \gamma}, \theta)dB_t, \quad 0 \leq t \leq 1.$$

Given the paths of the diffusion γ_t , the volatility function $\sigma_x(g_t^{\theta, \gamma}, \theta)$ may be viewed as a deterministic function of time and θ . The situation is similar to that of the previous section. We can now introduce a new time scale

$$s = \eta(t, \gamma, \theta) = \int_0^t \sigma_x^2(g_\omega^{\theta, \gamma}, \theta) d\omega.$$

Let T be the transformation of the ending time t_1 , $T = \eta(t_1, \gamma, \theta)$. We can then define U on the new time scale s as in (6):

$$U_s = \begin{cases} X_{\eta^{-1}(s, \gamma, \theta)}, & 0 \leq s \leq T, \\ M_{\eta^{-1}(s, \gamma, \theta)}, & s > T. \end{cases} \quad (13)$$

The SDE for U now becomes

$$dU_s = \left\{ \frac{\mu_x \left(U_s, g_{\eta^{-1}(s, \gamma, \theta)}^{\theta, \gamma}, \theta \right)}{\sigma_x^2(g_{\eta^{-1}(s, \gamma, \theta)}^{\theta, \gamma}, \theta)} \right\} dt + dW_s^U, \quad 0 \leq s \leq T.$$

We obtain the Radon Nikodym derivative between the distribution of U with respect to that of the Brownian motion W^U ,

$$\frac{d\mathbb{P}}{d\mathbb{W}^U} = G(s, U, \gamma, \theta),$$

and introduce \mathbb{W}_y^U as before. The density of y_1 under \mathbb{W}^U , denoted by $f(y_1|y_0, \gamma, \theta)$, is just

$$f(y_1|y_0, \gamma, \theta) \equiv N(y_0, T).$$

The dominating measure \mathbb{W}_y^U reflects a Brownian motion conditioned to equal y at a parameter depended time $T = \eta(t_1, \gamma, \theta)$. To remove this dependency we introduce another time scale

$$u = \eta_2(s) = \frac{s}{T(T-s)}, \quad \text{or } s = \eta_2^{-1}(u) = \frac{uT^2}{1+uT},$$

which defines a second time change similar to (8)

$$U_t = (T-s)Z_{s/\{T(T-s)\}} + \left(1 - \frac{s}{T}\right)y_0 + \frac{s}{T}y_1, \quad 0 \leq s < T. \quad (14)$$

Therefore, the SDE for Z is now given by

$$dZ_u = \frac{T}{1+uT} \left\{ \frac{\mu \left(\frac{T}{1+uT}\nu(Z_u), \gamma_{k(u, \gamma, \theta)}, \theta \right)}{\sigma_x^2(\gamma_{k(u, \gamma, \theta)}, \theta)} + \nu(Z_u) \right\} dt + dW_u^Z, \quad 0 \leq u < \infty,$$

where $k(u, \gamma, \theta)$ denotes the initial time scale of X , t , and $\nu(Z_u) = U_s$.

Conditional on γ , the likelihood can be written in a similar manner as in (10):

$$\frac{d\mathbb{P}}{d\{\mathbb{W}_y^Z \times Leb(y)\}}(Z|y_0, y_1, \gamma) = G(u, Z, \gamma, \theta)f(y_1|y_0, \gamma, \theta) \quad (15)$$

By using the exact same arguments of Section 3.1, we can show that \mathbb{W}_y^Z reflects a standard Wiener measure and therefore the dominating measure is independent of parameters. To obtain the full likelihood we need to multiply the two parts given by (12) and (15).

3.3 Incorporating leverage effect

In the previous section we made the assumption that the increments of X and γ are independent, in other words we assumed no leverage effect. This assumption can be relaxed in the following way: In the presence of a leverage effect ρ , the SDE of X conditional on γ can be written as

$$dX_t = \mu_x(X_t, g_t^{\gamma, \theta}, \theta)dt + \rho\sigma_x(g_t^{\gamma, \theta}, \theta)dW_t + \sqrt{1-\rho^2}\sigma_x(g_t^{\gamma, \theta}, \theta)dB_t, \quad 0 \leq t \leq t_1,$$

where W is the driving Brownian motion of γ). Note that given γ , W can be regarded as a function of γ and its parameters θ . Therefore, the term $\rho\sigma_x(g_t^{\gamma, \theta}, \theta)dW_t$ can be viewed as a deterministic function of time, and it can be treated as part of the drift of X_t . However, this operation introduces additional problems as the assumptions ensuring a weakly unique solution to the SDE of X are violated. To avoid this issue we introduce the infinitesimal transformation

$$X_t = \mathcal{H}(H_t, \rho, \gamma, \theta) = H_t + \int_0^t \rho\sigma_x(g_\omega^{\gamma, \theta}, \theta)dW_\omega,$$

which leads us to the following SDE for H :

$$dH_t = \mu_x \left\{ \mathcal{H}(X_t, \rho, \gamma, \theta), g_t^{\gamma, \theta}, \theta \right\} dt + \sqrt{1 - \rho^2} \sigma_x(g_t^{\gamma, \theta}, \theta) dB_t, \quad 0 \leq t \leq t_1.$$

We can now proceed as before, defining U and Z based on the SDE of H in a similar manner as in (13) and (14) respectively.

3.4 State dependent volatility

Consider the family of state dependent stochastic volatility models where conditional on γ , the SDE of X may be written as:

$$dX_t = \mu_x(X_t, g_t^{\gamma, \theta}, \theta) dt + \sigma_1(g_t^{\gamma, \theta}, \theta) \sigma_2(X_t, \theta) dB_t, \quad 0 \leq t \leq t_1.$$

This class contains among others, the models of Andersen and Lund (1998), Gallant and Tauchen (1998), Durham (2002), Eraker (2001). In order to apply the time change transformations of section 3.2, we should first transform X to \dot{X}_t , through $\dot{X}_t = l(X_t, \theta)$, so that it takes the form of (3). Such a transformation, which may be viewed as the first transformation in Roberts and Stramer (2001), should satisfy the following differential equation

$$\frac{\partial l(X_t, \theta)}{\partial X_t} = \frac{1}{\sigma_2(X_t, \theta)}.$$

The time change transformations for U and Z may then be defined on the basis of \dot{X} that will now have volatility $\sigma_1(g_t^{\gamma, \theta}, \theta)$. The transformation $l(\cdot)$ also applies to the observations $\dot{y}_0 = l(X_0, \theta)$ and $\dot{y}_1 = l(X_1, \theta)$ which may now be functions of the parameters in $\sigma_2(\cdot)$. These parameters enter the reparametrised likelihood in two ways: through the density $f(y_1 | \gamma, \theta)$ which now should include the relevant Jacobian term, and through the function $\nu(Z_u)$ in the drift of Z which according to (14) would depend on the transformed endpoints.

3.5 Multivariate stochastic volatility models

We may use the techniques of section 3.3 to define time change transformations for multidimensional diffusions. Consider a d -dimensional version of the SDE in (4) where σ now is a 2×2 matrix ($[\sigma]_{ij} = \sigma_{ij}$). As noted in Kalogeropoulos et al. (2007), the mapping between σ and the volatility matrix $\sigma\sigma^T$ should be 1-1 in order to ensure identifiability of the σ parameters. A way to achieve this, is by allowing σ to be the lower triangular matrix that produces the Cholesky decomposition of $\sigma\sigma^T$. For $d = 2$, the SDE of such a diffusion is given by

$$\begin{aligned} dX_t^{\{1\}} &= \mu(X_t^{\{1\}}, X_t^{\{2\}}, \theta) dt + \sigma_{11} dB_t, \\ dX_t^{\{2\}} &= \mu(X_t^{\{1\}}, X_t^{\{2\}}, \theta) dt + \sigma_{21} dB_t + \sigma_{22} dW_t. \end{aligned}$$

The time change transformations for $X^{\{1\}}$ will be exactly as in section 3.1. For $X^{\{2\}}$ note that given $X^{\{1\}}$ the term $\sigma_{21} dB_t$ is now a deterministic function of time and may be treated as part of the drift. Thus, we may proceed following the route of the section 3.3.

Similar transformations can be applied for diffusions that have, or may be transformed to have, volatility functions independent of their paths. For example we may assume two correlated price processes with correlation ρ_x :

$$\begin{aligned} [\sigma]_{11} &= \sigma_x^{\{1\}}(g_t^{\gamma, \theta}, \theta), \\ [\sigma]_{21} &= \rho_x \sigma_x^{\{2\}}(g_t^{\gamma, \theta}, \theta), \\ [\sigma]_{22} &= \sqrt{1 - \rho_x^2} \sigma_x^{\{2\}}(g_t^{\gamma, \theta}, \theta). \end{aligned}$$

We may proceed in a similar manner for multivariate stochastic volatility models of general dimension d .

4 MCMC implementation

The construction of an appropriate data augmentation algorithm involves several issues. We focus on describing how to update the latent diffusion paths and the parameters that drive the time change transformations. The updates of the remaining parameters include standard MCMC steps and are thus omitted. The change in the time scale introduces three interesting features: the presence of three time scales; the need to update diffusion paths that run from 0 to $+\infty$; and the fact that time scales depend on parameters. In this section we present the details of a MCMC scheme that addresses the above. For ease of illustration, we start with the simple case of a univariate diffusion with constant volatility

$$dX_t = \mu(t, X_t, \theta)dt + \sigma dB_t^X, \quad 0 \leq t \leq 1, \quad X_0 = y_0, \quad X_1 = y_1$$

and build up to the case of stochastic volatility models (3). Note that such extensions are not hard to implement since the time change transformations need only be applied to the paths of the observed diffusion process X . The updates of the transformed diffusion process γ , which drives the volatility, may be carried out using an overlapping scheme as in Kalogeropoulos (2007).

4.1 Three time scales

We introduce m intermediate points of X at equidistant times between 0 and 1, to give $X = \{X_{i/(m+1)}, i = 0, 1, \dots, m+1\}$. In addition, we make the assumption that m is large enough for accurate likelihood approximations and any error induced by the time discretisation is negligible for the purposes of our analysis. This assumption introduces no implications as the value of m may be set to an arbitrarily large value prior to the analysis.

Given a value of the time scale parameter σ , we can get the U -time points by applying (6) to each one of the existing points X , so that

$$U_{\sigma^2 i/(m+1)} = X_{i/(m+1)}, \quad i = 0, 1, \dots, m+1.$$

Note that it is only the times that change, the values of the diffusion remain intact. In a stochastic volatility model we would use the quantities

$$\int_{\frac{i}{m+1}}^{\frac{i+1}{m+1}} \sigma_x^2(\cdot) ds$$

for each pair of consecutive imputed points.

The points of Z are multiplied by a time factor which corrects the deviations from unit volatility. The times of the diffusion Z may be obtained by

$$t_i^Z = \frac{\sigma^2 i/(m+1)}{\sigma^2(\sigma^2 - \sigma^2 i/(m+1))}, \quad i = 0, 1, \dots, m.$$

Clearly this does not apply to the last point which occurs at time $+\infty$. The paths of X , or U , are thus more convenient for likelihood evaluations as they only contain finite time points. They may be used instead exploiting the fact that the relevant transformations are bijections. However, the component of the relevant Gibbs sampling scheme should be the diffusion process Z .

4.2 Updating the paths of Z

The paths of Z may be updated using an independence sampler with the reference measure as a proposal. Here \mathbb{W}_y^Z reflects a Brownian motion at the Z -time scale u , which is fixed given the current values of the time-scale parameters and the paths of γ in the case of stochastic volatility models. An appropriate algorithm is given by the following steps.

- **Step 1:** Propose a Brownian motion on the Z -time, say Z^* . The value at the endpoint (time $+\infty$) is not needed.

- Step 2: Transform back to X^* , using (8).
- Step 3: Accept with probability: $\min \left\{ 1, \frac{G(X^*, \theta, \sigma)}{G(X, \theta, \sigma)} \right\}$.

4.3 Updating time scale parameters

The updates of parameters that define the time scale, such as σ , are of particular interest. In almost all cases, their conditional posterior density is not available in closed form, and Metropolis steps are inevitable. However, every proposed value of these parameters will imply a different Z -time scale u . In other words, for each potential proposed value for σ there exists a different set of Z -points needed for accurate approximations of the likelihood the Metropolis accept-reject probabilities. In theory, this would pose no issues had we been able to store an infinitely thin partition of Z , but of course this is not possible.

We use retrospective sampling ideas; see Papaspiliopoulos and Roberts (2008) and Beskos and Roberts (2005) for applications in different contexts. Under the assumption of a sufficiently fine partition of the time scale of Z , all the non-recorded intermediate points contribute nothing to the likelihood and they are irrelevant in that respect. The set of recorded points is sufficient for likelihood approximation purposes. In other words, their distribution is given by the likelihood dominating measure \mathbb{W}_y^Z which reflects a Brownian motion. Hence, they can be drawn after the proposal of the candidate value of the time scale parameter. To ensure compatibility with the recorded partition of Z , it suffices to condition on their neighboring points. This is easily done using standard Brownian bridge properties: Suppose that we want to simulate the value of Z at time t_b which falls between the recorded values at times t_a and t_c , so that $t_a \leq t_b \leq t_c$. Denote by Z_{t_a} and Z_{t_c} the corresponding Z values. Under the assumption that Z is distributed according to \mathbb{W}_y^Z between t_a and t_c we have that

$$Z_{t_b} | Z_{t_a}, Z_{t_c} \sim N \left\{ \frac{(t_b - t_a)Z_{t_c} + (t_c - t_b)Z_{t_a}}{t_c - t_a}, \frac{(t_b - t_a)(t_c - t_b)}{t_c - t_a} \right\}. \quad (16)$$

We describe the algorithm for the case of more than two observations; denote by $X_0 = y_0$, $X_{t_1} = y_1$, \dots , $X_{t_n} = y_n$. The time change transformations should be applied to each pair of successive observations, thus giving n separate Z diffusion processes. As mentioned earlier, the diffusion processes X , or U are bijections of this collection of Z diffusion processes and may be used instead in a MCMC scheme. Our proposed algorithm for the σ -updates may be summarized through the following steps:

- Step 1: Propose a candidate value for σ , say σ^* .
- Step 2: Repeat for each pair of successive points:
 - Use (6) and (8) to get the new times associated with it.
 - Draw the values of Z retrospectively at the new times using (16).
 - Transform back to X^* (which corresponds to the time between the pair of successive points), using (8).
- Step 3: Form the entire path X^* by appropriately joining the bits between successive observations.
- Step 4: Accept with probability: $\min \left\{ 1, \frac{G(t, X^*, \theta, \sigma^*) \prod_{i=1}^n f(y_i | y_{i-1}, \sigma^*)}{G(t, X, \theta, \sigma) \prod_{i=1}^n f(y_i | y_{i-1}, \sigma)} \right\}$.

In a stochastic volatility model the updates of the paths of the transformed diffusion γ may be implemented using overlapping blocks. Note that these paths drive the time u , of the diffusion process Z , and therefore a similar algorithm as above should be embedded in their updates. For simplicity consider blocks of γ paths that correspond to times between non-successive observations. Each block is then further split into sub-blocks containing intervals between successive observations, thus providing a number of separate Z diffusion processes. Details are presented below:

- Step 1: Propose γ^* between the times of two non-successive observations, y_A and y_B , by a Brownian bridge connecting the relevant endpoints of γ .
- Step 2: Repeat for each pair of successive points between (and including) y_A and y_B :
 - Use (13) and (14) to get the new times associated with it.
 - Draw the values of Z retrospectively at the new times using (16).
 - Transform back to X^* (which corresponds to the time between the pair of successive points), using (14).
- Step 3: Join the bits of X^* to form its path between y_A and y_B .
- Step 4: Accept with probability: $\min \left\{ 1, \frac{G_\gamma(t, \gamma^*, \theta) G(t, X^*, \gamma^*, \theta) \prod_{i=1}^n f(y_i | y_{i-1}, \gamma^*, \theta)}{G_\gamma(t, \gamma, \theta) G(t, X, \gamma, \theta) \prod_{i=1}^n f(y_i | y_{i-1}, \gamma, \theta)} \right\}$.

5 Simulations

As discussed in section 2, appropriate reparametrisations are necessary to avoid issues regarding the mixing and convergence of the MCMC algorithm. In fact, the chain becomes reducible as the level of augmentation increases. This is also verified by the numerical examples performed in Kalogeropoulos (2007) even in very simple stochastic volatility models. In this section we present a simulation based experiment to check the immunity of MCMC schemes to increasing levels of augmentation, as well as the ability of our estimation procedure to retrieve the correct values of the diffusion parameters despite the fact that the series is partially observed and only at a finite number of points. We simulated data from the following stochastic volatility model

$$\begin{aligned} dX_t &= \kappa_x(\mu_x - X_t)dt + \rho \exp(\alpha_t/2)dW_t + \sqrt{1 - \rho^2} \exp(\alpha_t/2)dB_t, \\ d\alpha_t &= \kappa_\alpha(\mu_\alpha - \alpha_t)dt + \sigma dW_t, \end{aligned}$$

where B and W represent independent Brownian motions, and ρ reflects the correlation between the increments of X and α (also termed as leverage effect). A high frequency Euler approximating scheme with a step of 0.001 was used for the simulation of the diffusion paths. Specifically, 500,001 points were drawn and one value of X for every 1000 was recorded, thus forming a dataset of 501 observations of X at $0 \leq t \leq 500$. The parameter values were set to $\rho = -0.5$, $\sigma = 0.4$, $\kappa_x = 0.2$, $\mu_x = 0.1$, $\kappa_\alpha = 0.3$ and $\mu_\alpha = -0.2$

The transformations required to construct an irreducible data augmentation scheme are listed below. First α was transformed to γ through

$$\begin{aligned} \gamma_t &= \frac{\alpha_t - \alpha_0}{\sigma}, \quad 0 \leq t \leq 500, \\ \alpha_t &= g_t^{\gamma, \sigma} = \alpha_0 + \sigma \gamma_t. \end{aligned}$$

Given γ , and for each pair of consecutive observation times t_{k-1} and t_k ($k = 1, 2, \dots, 500$) on X, the following transformations were applied: First, we removed the term introduced from the leverage effect

$$H_t = X_t - \int_{t_{k-1}}^t \rho \exp \{g_s^{\gamma, \sigma} / 2\} dW_s, \quad t_{k-1} \leq t \leq t_k,$$

and consequently we set

$$\eta_1(t) = \int_{t_{k-1}}^t (1 - \rho)^2 \exp \{\nu(\gamma_s, \sigma, \alpha_0)\} ds.$$

Parameter	True value	Post. mean	Post. SD	Post 2.5%	Post median	Post 97.5%
κ_x	0.2	0.244	0.038	0.173	0.243	0.321
μ_x	0.1	0.313	0.174	-0.046	0.317	0.641
κ_α	0.3	0.304	0.148	0.110	0.277	0.672
μ_α	-0.2	-0.268	0.107	-0.484	-0.267	-0.059
σ	0.4	0.406	0.130	0.202	0.390	0.705
ρ	-0.5	0.477	0.138	-0.657	-0.491	-0.066

Table 1: Summaries of the posterior draws for the simulation example of Section 5 for $m = 50$.

Then, U and Z may be defined again from 13 and 14 respectively, but based on H rather on X . The elements of the MCMC scheme are Z , γ , α_0 and the parameters $(\kappa_x, \mu_x, \kappa_\alpha, \mu_\alpha, \rho, \sigma)$.

Vague priors were assigned to all of the parameters, subject to positivity constraints for κ_x , κ_α , σ and for ρ to be in $(-1, 1)$. The chain was run several times for 50,000 iterations on different levels of augmentation, by setting the number of imputed points to 2, 30, 40 and 50. As in Kalogeropoulos (2007), it was noted that a good choice of length of the overlapping blocks, needed for the updates of γ , may improve substantially the mixing of the chain. We used blocks with length corresponding to 8 observations. The acceptance rate rate for each block of γ was around 75% whereas the acceptance rate for each path of Z was around 95%. The time needed for such a MCMC run with $m = 40$ was roughly 4 hours in a mid-specification PC. We also noted a linear relationship between running times and m , which confirms the fact that the computational complexity of the algorithm is $O(m)$ (see discussion for more). Figure 1 shows autocorrelation plots for all parameters. There is no sign of any increase in the autocorrelation to raise suspicions against the irreducibility of the chain. This confirms the fact that convergence time of the algorithm is independent of m . Figure 2 shows density plots for all parameters and on all of m . These plots may be used to monitor the deterioration of the discretisation error. In this example, a choice of $m = 2$ may have been suboptimal whereas any value above 30 seems to perform well. Also, these plots reveal good agreement with the true values of the parameters, which is also supported by Table 1.

6 Application: US treasury bill rates

To illustrate the time change methodology we fit a stochastic volatility model to US treasury bill rates. The dataset consists of 1809 weekly observations (Wednesday) of the 3-month US Treasury bill rate from the 5th of January 1962 up to the 30th of August 1996. The data are plotted in Figure 3.

Previous analyses of these data include Andersen and Lund (1998), Gallant and Tauchen (1998), Durham (2002), Durham and Gallant (2002), Eraker (2001), and Golightly and Wilkinson (2006). Apart from some slight deviations the adopted stochastic volatility models consisted of the following SDE.

$$\begin{aligned} dr_t &= (\theta_0 - \theta_1 r_t)dt + r_t^\psi \exp(\alpha_t/2)dB_t, \\ d\alpha_t &= \kappa(\mu - \alpha_t)dt + \sigma dW_t, \end{aligned} \tag{17}$$

with independent Brownian motions B and W . In some cases the following equivalent model was used:

$$\begin{aligned} dr_t &= (\theta_0 - \theta_1 r_t)dt + \sigma_r r_t^\psi \exp(\alpha_t/2)dB_t, \\ d\alpha_t &= -\kappa\alpha_t dt + \sigma dW_t. \end{aligned} \tag{18}$$

The model in (17) was chosen, as posterior draws of its parameters exhibit substantially less autocorrelation. In line with Gallant and Tauchen (1998) and Golightly and Wilkinson (2006), we also set $\psi = 1$. Eraker (2001), Durham (2002) and Durham and Gallant (2002) assume general ‘elasticity of variance’ ψ but their estimates do not indicate a significant deviation from 1. By setting $X_t = \log(r_t)$, the volatility of X_t becomes $\exp(\alpha_t/2)$. Therefore the U -time for two consecutive observation times t_{k-1} and t_k is defined as

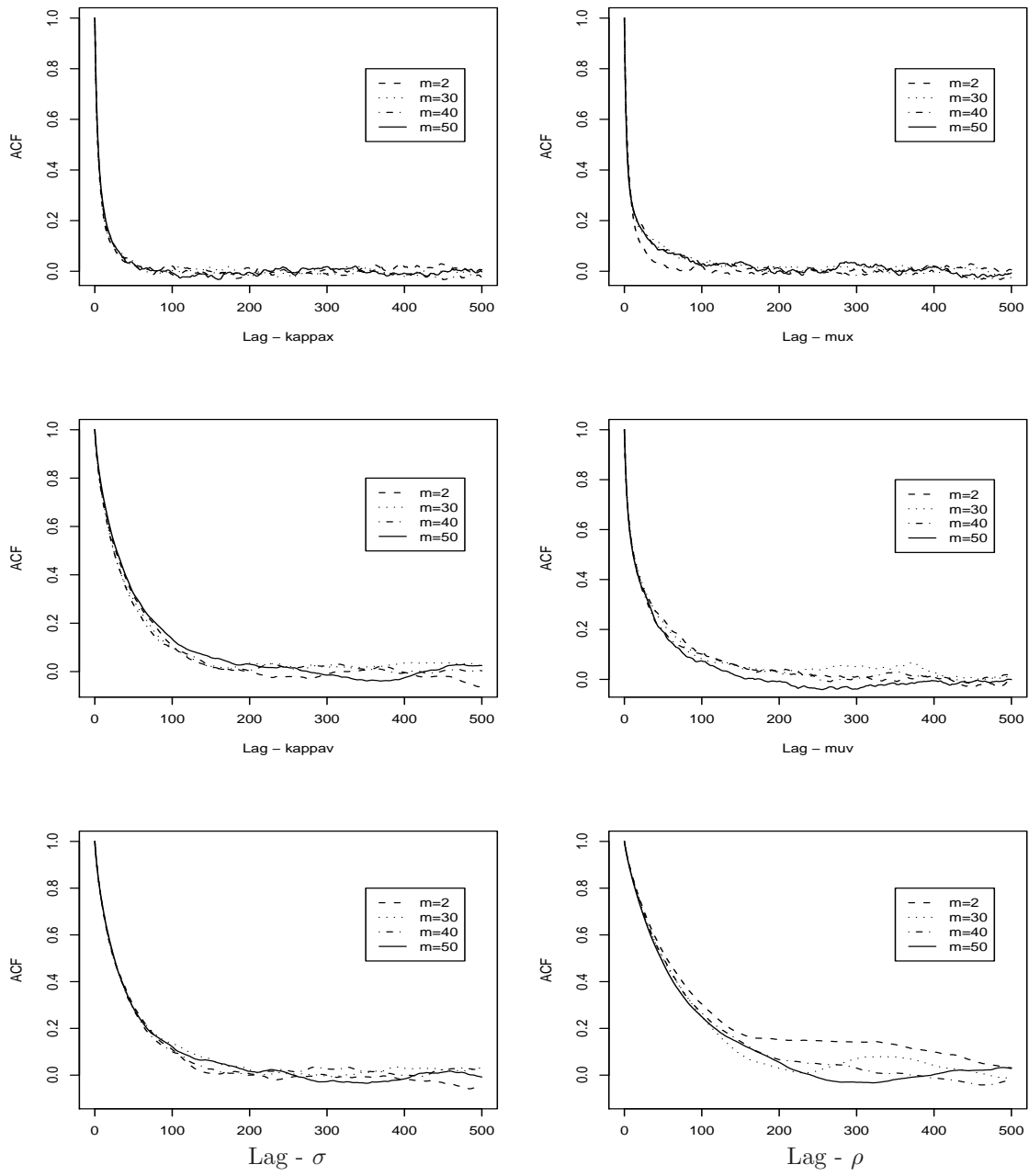


Figure 1: Autocorrelation plots of the parameter posterior draws for different numbers of imputed points $m = 2, 30, 40, 50$. Simulation example of Section 5.

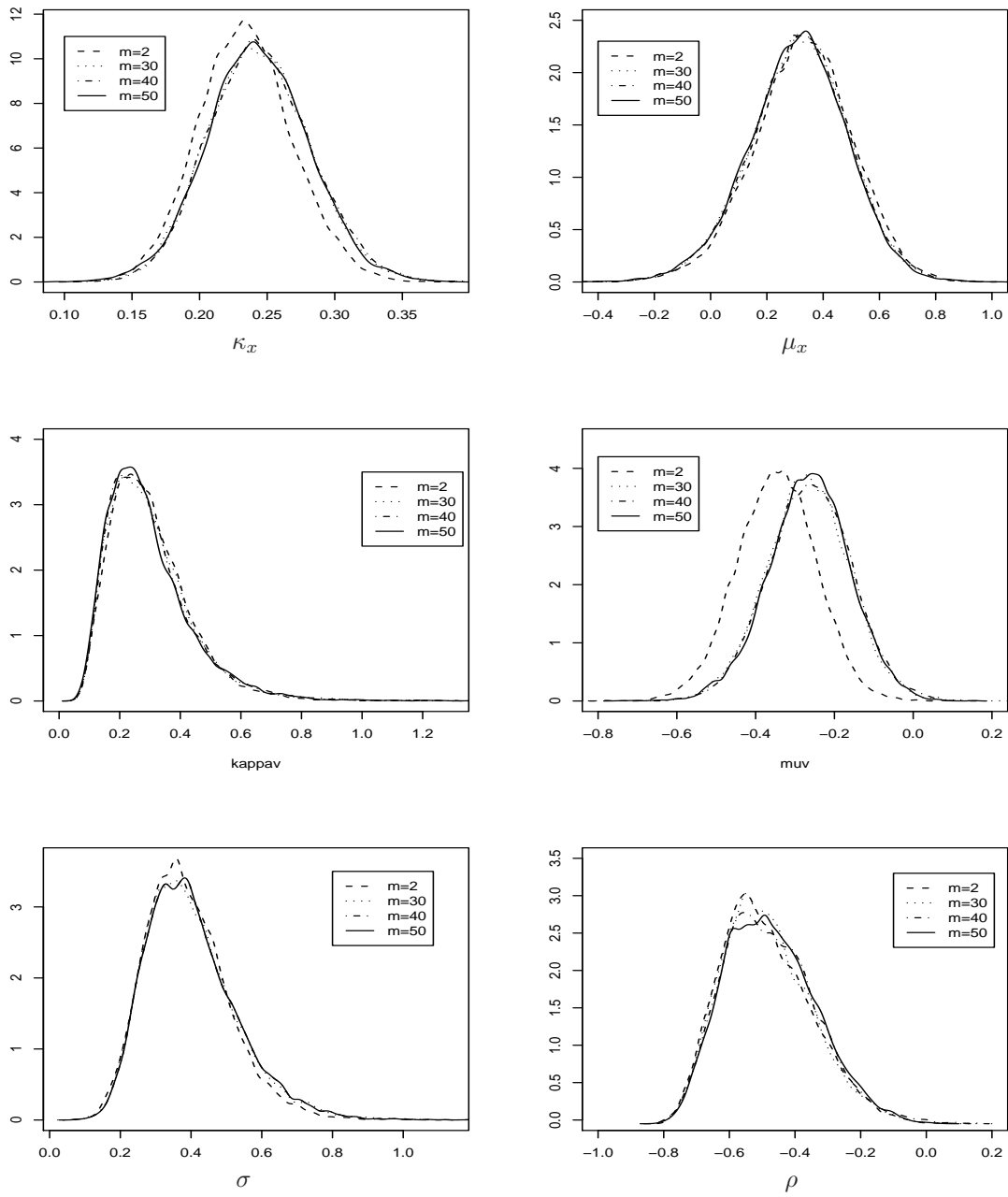


Figure 2: Kernel densities of the posterior draws of all the parameters for different numbers of imputed points $m = 2, 30, 40, 50$. Simulation example of Section 5.

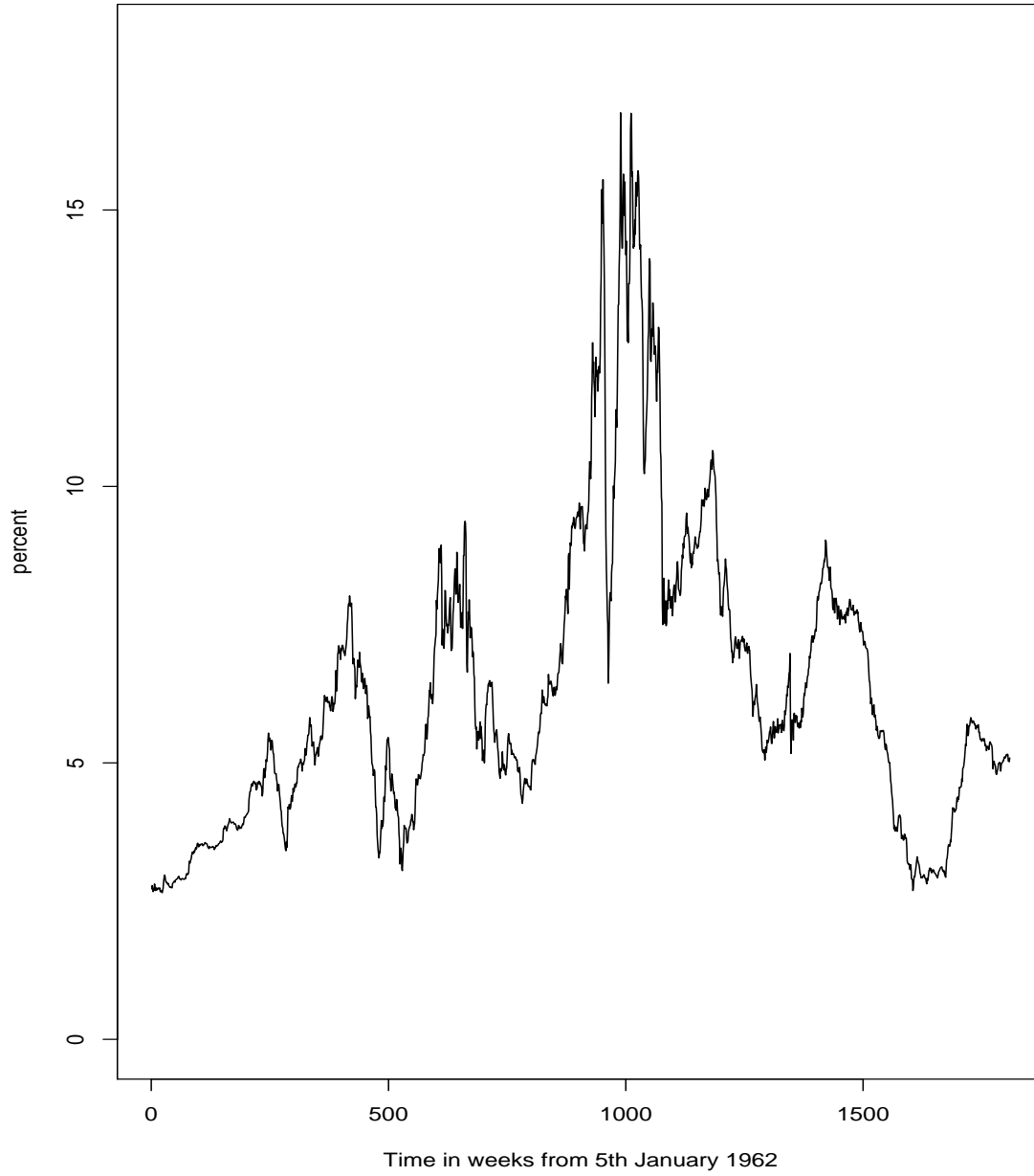


Figure 3: Weekly 3-month US Treasury bill rate from the 5th of January 1962 up to the 30th of August 1996.

Parameter	Post. mean	Post. SD	Post 2.5%	Post median	Post 97.5%
θ_0	0.130	0.238	-0.347	0.132	0.589
θ_1	0.013	0.057	-0.096	0.013	0.125
κ	2.403	0.620	1.319	2.360	3.745
μ	-3.966	0.211	-4.384	-3.964	-3.547
σ	2.764	0.311	2.199	2.750	3.420

Table 2: Summaries of the posterior draws for the stochastic volatility model of Weekly 3-month US Treasury bill rates.

$$\eta_1(t) = \int_{t_{k-1}}^t \exp(\alpha_s) ds,$$

and U and Z are given by (13) and (14) respectively. We also transformed α to γ as in section 5:

$$\gamma_t = \frac{\alpha_t - \alpha_0}{\sigma},$$

$$\alpha_t = g_t^{\gamma, \sigma} = \alpha_0 + \sigma \gamma_t.$$

We constructed appropriate MCMC schemes based on Z and γ to sample from the posterior of the parameters θ_0 , θ_1 , κ , μ and σ . The time was measured in years setting the distance between successive Wednesdays to 5/252. Non-informative priors were assigned to all the parameters, restricting κ and σ to be positive to ensure identifiability and eliminate the possibility of explosion. The algorithm was run for 50,000 iterations and for m equal to 2, 10 and 20. To optimize the efficiency of the chain we set the length of the overlapping blocks of γ to 10 which produced an acceptance rate of 51.9%. The corresponding acceptance rate for Z was 98.6%.

The kernel density plots of the posterior parameters and likelihood (Figure 4) indicate that a discretisation corresponding to an m of 10 or 20 provide reasonable approximations. A choice of $m = 2$ produces similar parameter posterior draws but the log-likelihood plot (bottom right) seems to be slightly off. The relevant autocorrelation plots of Figure 5 do not provide evidence of increasing autocorrelation in m . Finally, summaries of the posterior draws for all the parameters are provided in Table 2. The parameters κ , μ and σ are different from 0 verifying the existence of stochastic volatility. On the other hand, there is no evidence to support the existence of mean reversion on the interest rate process, as θ_0 and θ_1 are not far from 0. The results are in line with those of Durham (2002), Durham and Gallant (2002) and Golightly and Wilkinson (2006).

7 Discussion

Data augmentation MCMC schemes constitute a very useful tool for likelihood-based inference on diffusion models. They may not have the appealing properties of complete elimination of the time discretisation error (Beskos et al., 2006), or the closed form approximate likelihood expressions of Ait-Sahalia (2002), but nevertheless they give a satisfactory and very general solution to the problem. However, data augmentation schemes require careful construction to avoid the degeneracy issues described at the beginning of this paper.

Here, we introduce an innovative transformation which operates by altering the time axis of the diffusion. To accommodate the special features of time change transformations we also introduce a novel efficient MCMC scheme which mixes rapidly and is not prohibitively computationally expensive. Our method is also easy to implement and introduces no additional approximation error other than that included in methodologies based on a discretisation of the diffusion path. Moreover it has a broad range of applications which include general stochastic volatility models.

One clear advantage of the time change methodology is that in its pure form produces algorithms whose mixing time is bounded as m goes to infinity, as in Roberts and Stramer (2001). In addition, the computing

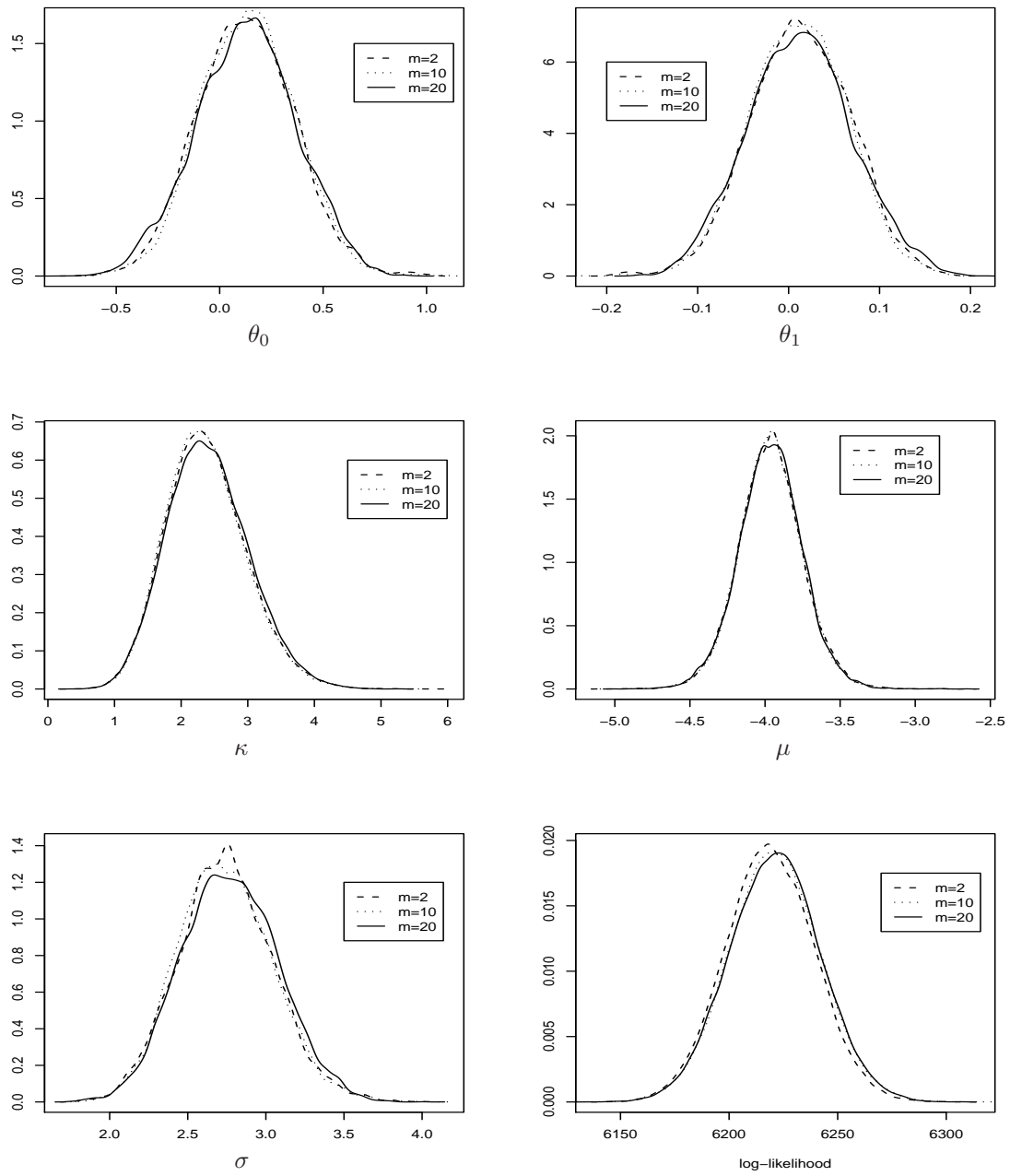


Figure 4: Kernel densities of the posterior draws of all the parameters and the log-likelihood for different values of imputed points $m = 2, 10, 20$. Example on Weekly 3-month US Treasury bill rates.

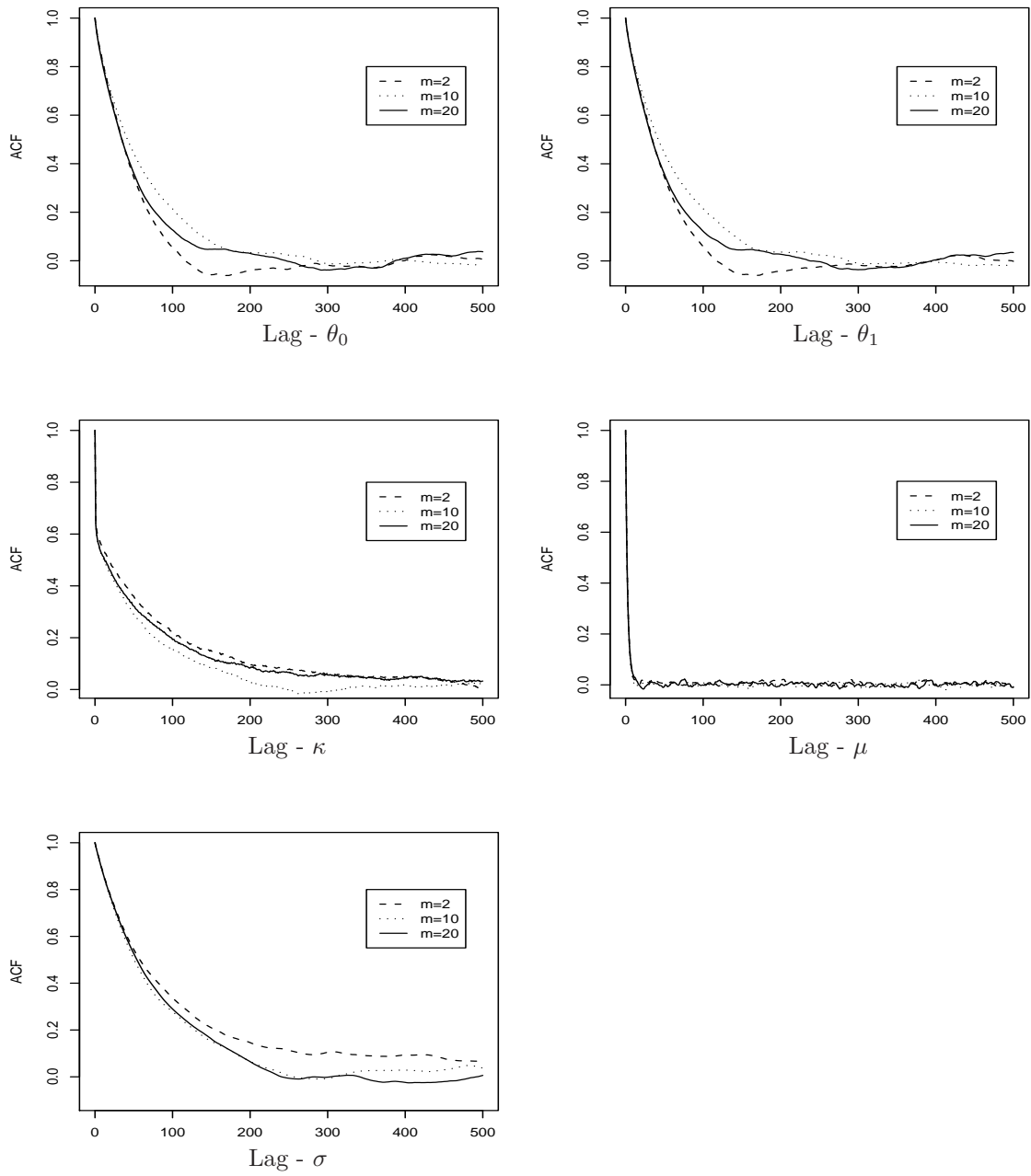


Figure 5: Autocorrelation plots for the posterior draws of the model parameters for different numbers of imputed points $m = 2, 10, 20$ for the analysis of Weekly 3-month US Treasury bill rates.

cost per iteration of our methods is $O(m)$ as with other competing methods. Thus the overall computing cost of our approach is $O(m)$ which compares favourably with competing methods that are typically $O(m^2)$. In our experience mixing properties of the methods introduced in this paper are good in comparison with competing methods for these types of models and data. Furthermore we have found out that implementation can routinely be carried out in a few hours on a mid-specification PC.

Further work will consider problems with state-dependent volatility and models which involve jump diffusions, to which the methodology introduced here can be easily applied. Fundamental to our approach here has been the introduction of a non-centered parametrisation to decouple dependence inherent in the model between missing data and volatility parameters. However non-centered constructions are not unique, as illustrated by the choice in the diffusion context between the state rescaling approaches of Golightly and Wilkinson (2007); Roberts and Stramer (2001) and the time-stretching strategy adopted here. Clearly, further work is required to investigate the relative merits of these approaches in different situations.

8 Acknowledgements

This work was supported by the EPSRC grant GR/S61577/01 and by the Marie Curie fellowships. The authors would also like to thank Omiros Papaspiliopoulos for helpful discussions.

References

- Aït-Sahalia, Y. (2002). Maximum likelihood estimation of discretely sampled diffusions: a closed form approximation approach. *Econometrica*, 70:223–262.
- Aït-Sahalia, Y. (2008). Closed form likelihood expansions for multivariate diffusions. *Annals of Statistics*, 36:906–937.
- Andersen, T. G. and Lund, J. (1998). Estimating continuous-time stochastic volatility models of the short term interest rate. *Journal of Econometrics*, 77:343–377.
- 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.
- Beskos, A. and Roberts, G. O. (2005). Exact simulation of diffusions. *Ann. Appl. Probab.*, 15(4):2422–2444.
- Bibby, B. and Sorensen, M. (1995). Martingale estimating functions for discretely observed diffusion processes. *Bernoulli*, 1:17–39.
- Chib, S., Pitt, M. K., and Shephard, N. (2007). Efficient likelihood based inference for observed and partially observed diffusions. Submitted.
- Durham, G. B. (2002). Likelihood based specification analysis of continuous time models of the short term interest rate. *Journal of Financial Economics*. Forthcoming.
- 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. 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.

- Fearnhead, P. and Papaspiliopoulos, O. and Roberts, G.O. (2008). Particle filters for partially observed diffusions. *Journal of Royal Statistical Society: Series B (Statistical Methodology)*., 70(4):755–777.
- Gallant, A. R. and Long, J. R. (1997). Estimating stochastic differential equations efficiently by minimum chi-squared. *Biometrika*, 84(1):125–141.
- Gallant, A. R. and Tauchen, G. (1996). Which moments to match? *Econometric Theory*, 12(4):657–681.
- Gallant, A. R. and Tauchen, G. (1998). Reprojecting partially observed systems with applications to interest rate diffusions. *Journal of American Statistical Association*, 93(441):10–24.
- Golightly, A. and Wilkinson, D. (2006). Bayesian sequential inference for nonlinear multivariate diffusions. *Statistics and Computing*, 16:323–338.
- Golightly, A. and Wilkinson, D. (2007). Bayesian inference for nonlinear multivariate diffusions observed with error. *Computational Statistics and Data Analysis*, 52(3), 1674–1693.
- Gouriéroux, C., Monfort, A., and Renault, E. (1993). Indirect inference. *Journal of Applied Econometrics*, 8:S85–S118.
- Heston, S. (1993). A closed-form solution for options with stochastic volatility. with applications to bonds and currency options. *Review of Financial Studies*, 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. (1999). Bayesian estimation of continuous-time finance models. Unpublished paper, Simon School of Business, University of Rochester.
- Jones, C. S. (2003). Nonlinear mean reversion in the short-term interest rate. *The review of financial studies*, 16:793–843.
- Kalogeropoulos, K. (2007). Likelihood based inference for a class of multidimensional diffusions with unobserved paths. *Journal of Statistical Planning and Inference*, 137:3092–3102
- Kalogeropoulos, K., Dellaportas, P., and Roberts, G. (2007). Likelihood-based inference for correlated diffusions. Submitted.
- Liu, J. and West, M. (2001). Combined parameter and state estimation in simulation-based filtering In *Sequential Monte Carlo Methods in Practice*, by A. Doucet, J.F.G. De Freitas and N.J. Gordon, Springer-Verlag, New York.
- Oksendal, B. (2000). *Stochastic differential equations*. Springer, 5th edition.
- Papaspiliopoulos, O. and Roberts, G. (2008). Retrospective MCMC for Dirichlet process hierarchical models. *Biometrika*, 95(1):169–186
- 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.
- Pitt, M.K. and Shephard, N. (1999). Filtering via Simulation: Auxiliary Particle Filters. *Journal of the American Statistical Association*, 94(446):590–599.
- Roberts, G. and Stramer, O. (2001). On inference for partial observed nonlinear diffusion models using the metropolis-hastings algorithm. *Biometrika*, 88(3):603–621.
- Rogers, L. C. G. and Williams, D. (1994). *Diffusions, Markov processes and martingales, 2, Ito calculus*. Wiley, Chicester.

- 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 price distributions with stochastic volatility: an analytic approach. *Review of Financial Studies*, 4(4):727–752.
- Stroud, J.R. and Polson, N.G. and Muller P. (2004). Practical filtering for stochastic volatility models. In *State Space and Unobserved Component Models*, (by Harvey et al.), Cambridge University Press 236–247.
- 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.