Likelihood based inference for correlated diffusions

Konstantinos Kalogeropoulos*

University of Cambridge, Department of Engineering - Signal Processing Laboratory

Petros Dellaportas

Athens University of Economics and Business, Department of Statistics

Gareth O. Roberts

University of Warwick, Department of Statistics

October 22, 2007

Abstract

We address the problem of likelihood based inference for correlated diffusion processes using Markov chain Monte Carlo (MCMC) techniques. Such a task presents two interesting problems. First, the construction of the MCMC scheme should ensure that the correlation coefficients are updated subject to the positive definite constraints of the diffusion matrix. Second, a diffusion may only be observed at a finite set of points and the marginal likelihood for the parameters based on these observations is generally not available. We overcome the first issue by using the Cholesky factorisation on the diffusion matrix. To deal with the likelihood unavailability, we generalise the data augmentation framework of Roberts and Stramer (2001 Biometrika 88(3):603-621) to $d-$dimensional correlated diffusions including multivariate stochastic volatility models. Our methodology is illustrated through simulation based experiments and with daily EUR /USD, GBP/USD rates together with their implied volatilities.

Keywords: Markov chain Monte Carlo, Multivariate stochastic volatility, Multivariate CIR model, Cholesky Factorisation.

*Address: Trumpington Street, Cambridge,UK,CB2 1PZ, Tel: +44 (0)1223 332766, E-mail: kk384@cam.ac.uk

1 Introduction

Diffusion processes provide a natural model for phenomena evolving continuously in time. One of their appealing features is that they are defined in terms of the instantaneous mean and variance of the process. Specifically, a diffusion x_t obeys the dynamics of the following stochastic differential equation (SDE)

$$dx_t = \mu(t, x_t, \theta)dt + \sigma(t, x_t, \theta)dw_t, \tag{1}$$

driven by standard Brownian motion w_t. The functions $\mu(.)$ and $\sigma(.)$ are termed as the drift and the volatility of the diffusion respectively. Throughout this paper we suppress the dependence on t to simplify the notation, but the methodology is also applicable to time inhomogeneous diffusions. The diffusion process x_t is well defined if (1) has a unique weak solution, which translates into some regularity conditions (locally Lipschitz with a linear growth bound) on $\mu(.)$ and $\sigma(.)$; see chapter 5 of Rogers and Williams (1994) for more details.

We address the problem of modelling several diffusions, denoted by $x_t^{\{i\}}, \quad i = 1, \ldots, d$. Each diffusion $x_t^{\{i\}}$ may have a drift $\mu^{\{i\}}(.)$ and volatility $\sigma^{\{i\}}(.)$ of general, yet known, form. We also allow for correlations, $corr(dx_t^{\{i\}}, dx_t^{\{j\}})$=$\rho_{ij} = \rho_{ji}$, $i \neq j$, on the instantaneous increments. The use of cross-correlations is quite common when modelling multivariate time series, as they may capture effects caused by common factors of the underlying stochastic processes. In this paper we illustrate our methodology through two examples of correlated diffusions. The first example targets interest rates and bond pricing. Such time series often exhibit strong inter-dependencies; for instance, interest rates may correspond to similar bonds but with different expiry dates, thus giving rise to correlations among them. In Section 5 we examine a multivariate version of the Cox et al. (1985) model (CIR), often used for such data. The second example considers currency pairs which are known to be correlated, possibly due to the common currencies they may represent. Section 6 contains an analysis on EUR/USD and GBP/USD data, based on multivariate versions of stochastic volatility diffusions, such as the model of Heston (1993). In both examples, the inclusion of correlations in the model is essential for two reasons. First, they may affect the parameter estimates of the individual diffusions, as well as their precision. Second, they reflect characteristics of the market which may be useful in the bond/option pricing procedure.

We proceed by combining the diffusions $x_t^{\{i\}}$ together into $X_t = (x_t^{\{1\}}, \ldots, x_t^{\{d\}})'$ (with $'$ denoting transposition), so that X_t is a $d-$dimensional vector for each time t. The diffusion matrix of X_t, A, denotes its instantaneous covariance and takes the following form:

$$A := \begin{pmatrix} \sigma^{\{1\}}(.)^2 & \rho_{12}\sigma^{\{1\}}(.)\sigma^{\{2\}}(.) & \ldots & \rho_{1d}\sigma^{\{1\}}(.)\sigma^{\{d\}}(.) \\ \rho_{12}\sigma^{\{1\}}(.)\sigma^{\{2\}}(.) & \sigma^{\{2\}}(.)^2 & \ldots & \rho_{2d}\sigma^{\{2\}}(.)\sigma^{\{d\}}(.) \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{d1}\sigma^{\{1\}}(.)\sigma^{\{d\}}(.) & \rho_{d2}\sigma^{\{2\}}(.)\sigma^{\{d\}}(.) & \ldots & \sigma^{\{d\}}(.)^2 \end{pmatrix} \tag{2}$$

The diffusion process X_t is defined through the following multi-dimensional SDE

$$dX_t = M(X_t, \theta)dt + \Sigma(X_t, \theta)dW_t, \tag{3}$$

where W_t is a $d-$dimensional Brownian motion with independent components, with vector valued drift $M : [0, +\infty) \times \mathcal{S}_X \times \Theta \to \Re^d$ with $[M(.)]_i = \mu^{\{i\}}(.)$, and matrix valued volatility (also termed as dispersion matrix) $\Sigma(\cdot) : [0, +\infty) \times \mathcal{S}_X \times \Theta \to \Re^{d\times d}$, where $\mathcal{S}_X$ and Θ denotes the domain of the diffusion X_t and the parameter vector θ respectively. The dispersion matrix Σ is a square root of the instantaneous covariance matrix $A = \Sigma\Sigma'$. To ensure a unique weak solution for X_t, we require a unique weak solution for each $x_t^{\{i\}}$ and the matrix A to be positive definite for all t, X_t, θ.

Each diffusion $x_t^{\{i\}}$ may be observed, with or without error, at a finite set of points, or may be entirely unobserved. The diffusion will be termed as directly observed in cases with exact observations on all $x_t^{\{i\}}$, and partially observed otherwise. For ease of exposition, the methodology of this paper is initially presented for directly observed diffusions, and adaptations to partial observation regimes, as in multivariate stochastic volatility models, are provided when necessary. Similarly, we consider observations of the entire vector of X_t at each time, although this assumption can easily be relaxed. We denote the times of observations by $t_k,\ k = 1, \ldots, n$, and the data with $Y = \left\{Y_k = X_{t_k} = (x_{t_k}^{\{1\}}, \ldots, x_{t_k}^{\{d\}})',\ k = 1, \ldots, n\right\}$. Our aim is to draw likelihood based inference for the parameter vector θ given these observations.

The task of inference on diffusions observed discretely in time is generally not trivial and has received a remarkable attention in the recent literature; see Sørensen (2004) for a recent review.

The main problem is that the likelihood is generally not available except for a few cases. This has stimulated various techniques based on likelihood approximations. Approximations may be analytical (Aït-Sahalia, 2005), or simulation based; see Pedersen (1995) or a refinement of this technique 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. (2006a) succeeds exact inference in the sense that it allows only for Monte Carlo error.

We shall adopt a Bayesian approach using Markov chain Monte Carlo (MCMC) method. Since diffusions are not completely observed, it is natural to use data augmentation (Tanner and Wong, 1987), treating the segments of diffusion sample path (or a suitably fine approximation to this) as missing data. Initial MCMC schemes of this type were introduced by Jones (1999), 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), the algorithms introduced in these initial implementations of MCMC in this context degenerate as the number of imputed points increases. The problem may be overcome for scalar diffusions with the reparametrisation of Roberts and Stramer (2001). An alternative reparametrisation is provided by Golightly and Wilkinson (2007), see also Golightly and Wilkinson (2006) for a sequential approach, which can in principle be applied in principle to any diffusion.

However, the adaptation of such MCMC scheme to multivariate diffusions introduces additional issues. The task of updating the covariance matrix A is generally not trivial, as its full conditional posterior is most of the times intractable, and the use of Metropolis steps is inevitable. It is therefore crucial, especially for high-dimensional diffusions, to update the covariance matrix componentwise as the discrepancy between proposed and current moves is increasing in d. This introduces the problem of preserving the positive definite structure of the diffusion matrix A. Note that drawing samples from the posterior of covariance matrices, which may not necessarily be diffusion matrices, is a general MCMC issue and usually requires appropriate matrix decompositions; see for example Pinheiro and Bates (1996) and Daniels and Kass (1999).

The contribution of this paper is two-fold. First, we introduce a natural and general framework for sampling diffusion matrices in a MCMC environment. This framework is based on the Cholesky factorisation of A and enables us to define Σ explicitly. The MCMC algorithm may then be appropriately designed to provide samples from the posterior of Σ, which can be transformed to A at any time through the Cholesky decomposition. This framework may be coupled with any of the previously mentioned likelihood approximation techniques, such as those of Beskos et al. (2006a) or Aït-Sahalia (2005), to perform Bayesian inference for the parameters of the multi-dimensional diffusion. Second, we offer a full and stand alone MCMC scheme which combines the Cholesky decomposition with the reparametrised data augmentation approach of Roberts and Stramer (2001). This scheme may be used for parameter estimation of several multivariate diffusion models including stochastic volatility. The use of data augmentation is justified by its convenient property to be applicable at both directly and partially observed diffusions.

The paper is organised as follows: Section 2 describes the structure of a data augmentation scheme and highlights potential problems regarding the irreducibility of the MCMC algorithm. These problems may be tackled with the reparametrisation of this paper which requires the Cholesky factorisation of the diffusion matrix, presented in Section 3. Specific MCMC implementation details are given in Section 4 and the methodology of this paper is illustrated through simulated data in Section 5, and on daily EUR/USD, GBP/USD currency pairs in Section 6. Finally, we summarise in Section 7 adding some discussion and links to some other relevant work.

2 Data augmentation and degeneracy issues

2.1 The problem in practice

Data augmentation scheme bypasses the problem of simulating directly from the posterior $\pi(\theta|Y)$, which is typically unavailable for discretely observed data. The idea is to introduce a latent variable $\mathcal{X}$ that simplifies the likelihood $\mathcal{L}(Y;\mathcal{X},\theta)$. We use 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` $\mathcal{L}(Y;\mathcal{X},\theta)\pi(\theta)$.

Our problem can easily be adapted to this setting. Y represents the observations of the price process X_t, and $\mathcal{X}$ contains discrete skeletons of the diffusion paths between Y. Thus, $\mathcal{X}$ and Y constitute the augmented dataset $X_{i\delta},\ i = 0, \ldots, T/\delta$, which is a fine partition of the multivariate diffusion X_t with δ controlling the amount of augmentation. Based on this partition the likelihood can be approximated, for example via the Euler-Maruyama approximation

$$\mathcal{L}^E(Y; \mathcal{X}, \theta) = \prod_{i=1}^{T/\delta} p(X_{i\delta}|X_{(i-1)\delta}),$$

$$X_{i\delta}|X_{(i-1)\delta} \sim \mathcal{N}\left(X_{(i-1)\delta} + \delta M(X_{(i-1)\delta}, \theta), \delta A(X_{(i-1)\delta}, \theta)\right), \tag{4}$$

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

Another property of diffusions relates $A(X_t, \theta)$ with the quadratic variation process. Specifically it is well-known that

$$\lim_{\delta \to 0} \sum_{i=1}^{T/\delta} \left(X_{i\delta} - X_{(i-1)\delta}\right)\left(X_{i\delta} - X_{(i-1)\delta}\right)' = \int_0^T A(X_s, \theta) ds \ \ a.s. \tag{5}$$

The solution of the equation above determines the diffusion matrix parameters exactly. Hence, there exists perfect correlation between these parameters and $\mathcal{X}$ as $\delta \to 0$. Thus for the theoretical algorithm which imputes the entire X path, the MCMC algorithm is reducible. In practice this means that as the proportion of imputed data points increases mixing problems for the MCMC chain become progressively worse This phenomenon was first noted in Roberts and Stramer (2001) and Elerian et al. (2001). As would be expected, the EM algorithm suffers from the same problem.

2.2 Measure theoretic probability viewpoint

In this section, we explore the problem from a different angle, through a slightly more rigorous look at the likelihood. Let X_t be a diffusion that satisfies (3) and assume $X_0 = Y_0$ and $X_1 = Y_1$, $Y = (Y_1, Y_2)$. Denote the probability law of X by $\mathbb{P}_\theta$ and that of its driftless version,

$$d\mathcal{M}_t = \sigma(X_t, \theta) dW_t,$$

by $\mathbb{Q}_\theta$. To write down the likelihood, we can use the Cameron-Martin-Girsanov formula which provides the Radon-Nikodym derivative of $\mathbb{P}_\theta$ with respect to $\mathbb{Q}_\theta$:

$$\begin{aligned}\frac{d\mathbb{P}_\theta}{d\mathbb{Q}_\theta} \;=\; G(X,M,A) \;&=\; \exp\left\{\int_0^T \left[A(X_s,\theta)^{-1}M(X_s,\theta)\right]' dX_s \right.\\ &\left. -\frac{1}{2}\int_0^T M(X_s,\theta)'A(X_s,\theta)^{-1}M(X_s,\theta)ds\right\}.\end{aligned}$$

Note that the expression above contains stochastic and path integrals for which an analytic solution is generally not available. However, given a sufficiently fine partition of the diffusion path, they can be evaluated numerically providing an approximation of the likelihood which is equivalent to (4).

Now assume for a moment that under $\mathbb{Q}_\theta$ the marginal density of Y with respect to $d-$dimensional Lebesgue measure $Leb_d(Y)$, is known and denote by $f_{\mathcal{M}}(Y;\theta)$. The dominating measure $\mathbb{Q}_\theta$ can be factorised in the following way

$$\mathbb{Q}_\theta = \mathbb{Q}_\theta^Y \times Leb_d(Y) \times f_{\mathcal{M}}(Y;\theta), \tag{6}$$

where $\mathbb{Q}_\theta^Y$ is the measure $\mathbb{Q}_\theta$ conditioned on the observations Y. We can now write

$$\frac{d\mathbb{P}_\theta}{\mathbb{Q}_\theta^Y \times Leb_d(Y)}(X^{mis},Y) \;=\; G(X,M,A) \times f_{\mathcal{M}}(Y;\theta). \tag{7}$$

The expression in (7) provides the likelihood for the latent diffusion paths X^{mis} and the parameters θ. However, this likelihood is not valid because its reference measure, $\mathbb{Q}_\theta^y$, depends on parameters. Furthermore, since the volatility parameters are identified by the quadratic covariation process, the measure $\mathbb{Q}_\theta$ is just a point mass. Consequently, the measures $\mathbb{Q}_\theta$ are mutually singular and therefore so are $\mathbb{P}_\theta$. Hence, inference for both X^{mis}, θ is not possible using a common $\sigma-$finite dominating measure. In the next section, we specify an appropriate transformation of the diffusion that allows a likelihood specification with respect to a parameter-free dominating measure. This transformation may be viewed as a generalisation of the one in Roberts and Stramer (2001). The transformed diffusion has unit volatility, thus the problems induced by the quadratic variation property of (5) are implicitly addressed.

3 Likelihood specification

3.1 A Cholesky factorisation of the diffusion matrix

Consider the multi-dimensional SDE of (3) with the diffusion matrix A of (2). The $d \times d$ matrices A and Σ are linked through $A = \Sigma\Sigma'$, therefore Σ is not unique. However, it is crucial to define Σ explicitly and establish a 1-1 mapping with A, as each one of these two matrices may be more convenient for different reasons. The likelihood, defined either through the Euler-Maruyama approximation in (4) or through Cameron-Martin-Girsanov's formula in (7), is expressed in terms of A, which is also the main target of inference. On the other hand A is a positive definite matrix, whereas the only assumption made on Σ requires its full rank. Hence it is generally more convenient to work with Σ in the context of a MCMC algorithm. Moreover, as mentioned in the previous section, the generalisation of the Roberts and Stramer (2001) reparametrisation involves a transformation to unit volatility which will naturally be based on Σ.

In this paper, we define Σ using the Cholesky decomposition of A. Let $S_x(X_t, \theta) = diag\{\sigma^{\{i\}}(X_t, \theta)\}$. The diffusion matrix may then be factorised in the following way

$$A(X_t, \theta) \;=\; S_x(X_t, \theta)\; R\; S_x(X_t, \theta),$$

where R is the correlation matrix. One may define Σ as the product of S_x with the Cholesky decomposition of R, say C. But the elements of C will not have the general Cholesky structure, since R has the additional property of being a correlation matrix. To eliminate such problems we write each $\sigma_i(X_t, \theta)$ as

$$\sigma^{\{i\}}(X_t, \theta) = c_i f^{\{i\}}(X_t, \theta), \;\; \forall i, \tag{8}$$

for some positive constants c_i. This imposes no restrictions as we can always set $f^{\{i\}}(X_t, \theta) = \sigma^{\{i\}}(X_t, \theta)/c_i$, see Section 3.4 for such an example. Now, based on $F_x(X_t, \theta) = diag\{f^{\{i\}}(X_t, \theta)\}$, we can use (8) to obtain an alternative decomposition of A,

$$A(X_t, \theta) \;=\; F_x(X_t, \theta)\; V\; F_x(X_t, \theta),$$

where V is a general symmetric positive definite matrix with

$$V_{ij} = \begin{cases} c_i^2, & i = j \\ \rho_{ij} c_i c_j, & i \neq j. \end{cases} \tag{9}$$

The Cholesky decomposition of V, denoted by C ($V = CC'$), may now be used. The dispersion matrix $\Sigma(X_t, \theta)$ is defined as

$$\Sigma(X_t, \theta) \;=\; F_x(X_t, \theta)\, C. \tag{10}$$

In coordinate form, Σ may be written as

$$[\Sigma(X_t, \theta)]_{ij} = \begin{cases} [C]_{ij} f_i(X_t, \theta), & j \leq i \\ 0, & j > i. \end{cases}$$

The only restriction on the constants C_{ij} requires compatibility with the Cholesky decomposition, which translates on positive diagonal entries C_{ii}. As we mention in 4.2, this is particularly convenient in a MCMC environment and specifically for componentwise updates of $\Sigma(X_t, \theta)$ parameters. The Cholesky decomposition establishes the 1-1 mapping between Σ and A and ensures that the entire space of diffusion matrices as A is covered.

3.2 Transformation to unit volatility

In Section 2, the need for a reparametrisation was highlighted in order to avoid degenerate MCMC algorithms. Roberts and Stramer (2001) provide a solution to the problem for scalar diffusions, which involves a transformation to unit volatility. However, in more than one dimensions such a transformation does not always exist, as noted Aït-Sahalia (2005). When such a transformation is available the diffusion is said to be reducible, a term introduced by Aït-Sahalia (2005) who also provides a necessary and sufficient condition for reducibility: diffusions with non-singular $\Sigma(X_t, \theta)$ are reducible if and only if

$$\frac{\partial [\Sigma(X_t, \theta)^{-1}]_{ij}}{\partial x_t^{\{k\}}} = \frac{\partial [\Sigma(X_t, \theta)^{-1}]_{ik}}{\partial x_t^{\{j\}}}, \; \forall\, i, j, k \in \{1, \ldots, d\}, \text{ with } j < k \tag{11}$$

Not all SDEs with diffusion matrix A as in (2) or dispersion matrix Σ as in (10) are reducible. In this section, we restrict our attention to diffusions with

$$\sigma^{\{i\}}(X_t, \theta) \equiv \sigma^{\{i\}}(x_t^{\{i\}}, \theta), \tag{12}$$

for which we prove the reducibility. This is established by the following proposition:

Proposition 3.1 *Let X be a d-dimensional diffusion which obeys the following SDE:*

$$dX_t = M(t, X_t, \theta)dt + \Sigma(t, X_t, \theta)dW_t.$$

Furthermore, assume that

$$\Sigma(X_t, \theta) \ = \ F_x(X_t, \theta)\ C,$$

where $F_x(X_t, \theta) = diag\{f^{\{i\}}(x_t^{\{i\}}, \theta)\}$ and C is a lower triangular matrix with positive diagonal elements. The diffusion X can then be transformed to one with identity diffusion matrix. In other words X is reducible.

Proof: See Appendix.

The next proposition provides explicitly a transformation to unit volatility. It may be viewed as an alternative proof of proposition 3.1

Proposition 3.2 *Consider the setting and the diffusion X_t of proposition 3.1. Suppose that there exist $g^{\{i\}}(x_t\{i\}, \theta)$ for $i = 1, \ldots, d$ with continuous second derivatives, so that*

$$\frac{\partial g^{\{i\}}(x_t^{\{i\}}, \theta)}{\partial x_t^{\{i\}}} = \frac{1}{f^{\{i\}}(x_t^{\{i\}}, \theta)}, \ j = 1, \ldots, d,$$

and let $G_x(X_t, \theta) = \Big(g^{\{1\}} x_t^{\{1\}}, \theta), \ldots, g^{\{d\}}(x_t^{\{d\}}, \theta)\Big)'$. Consider the transformation

$$H(X_t, \theta) = \Big(h^{\{1\}}(X_t, \theta), \ldots, h^{\{d\}}(X_t, \theta)\Big)' = C^{-1} G_x(X_t, \theta). \tag{13}$$

The diffusion $U_t = H(X_t, \theta)$ has then unit volatility.

Proof: See Appendix.

The transformation of (13) may be used to specify the likelihood under an appropriate reparametrisation which will ensure a non - decreasing efficiency, of the data augmentation MCMC scheme, in the level of augmentation. Notice that the transformation of (13) to unit volatility is not unique. This is not necessary for our methodology, in fact we only require its invertibility which is ensured as long as each $g_i(x_t\{i\}, \theta)$ is itself invertible. We present this reparametrisation in the Section 3.3, whereas in 3.4 we show how to relax the assumption of (12) to handle multivariate stochastic volatility models.

3.3 Reparametrised likelihood

Consider the diffusion that satisfies the SDE of (3) where the drift $M(.)$ and Σ satisfy the appropriate conditions so that X_t has a unique weak solution and Ito's lemma can be applied. Furthermore, assume that

$$\Sigma(X_t, \theta) \;=\; F_x(X_t, \theta)\, C,$$

where $F_x(X_t, \theta) = diag\{f^{\{i\}}(x_t^{\{i\}}, \theta)\}$ and C is a lower triangular matrix with positive diagonal elements. For ease of illustration let the entire vector of X_t be observed at each time and denote the times of observations by $t_k,\ k = 0, \ldots, n$, and the data with $Y = \Big\{Y_k = X_{t_k} = (x_{t_k}^{\{1\}}, \ldots, x_{t_k}^{\{d\}})',\ k = 1, \ldots, n\Big\}$. We will define the likelihood for a pair of successive observations, (Y_{k-1}, Y_k). Due to the Markov property of diffusions, the full likelihood is just given by the product of all pairs of consecutive observations. Without applying a reparametrisation, the likelihood can be defined through (7). However, as discussed in 2, this likelihood is problematic because it is written with respect to a dominating measure that depends on parameters. The aim of the reparametrisation is to obtain a likelihood with a parameter-free dominating measure.

The first step of the reparametrisation requires a transformation $U_t = H(X_t, \theta) = \big(u^{\{1\}}, \ldots, u^{\{d\}}\big)'$, so that the diffusion matrix of U_t is the $d-$dimensional identity matrix. As established by proposition 3.1, such a transformation does exist and can be obtained explicitly by (13). The SDE of the $r-$th coordinate of the transformed diffusion U will be given by:

$$du_t^{\{r\}} = \mu_U^{\{r\}}(U_t, \theta)dt + dw_t^{\{r\}},\ r = 1, \ldots, d,$$

with

$$\mu_U^{\{r\}}(U_t, \theta) = \sum_{i=1}^{d} \frac{\partial h_r(X_t, \theta)}{\partial x^{\{i\}}} \mu^{\{i\}}(X_t, \theta) + \sum_{i=1}^{d} \frac{\partial^2 h_r(X_t, \theta)}{\partial (x^{\{i\}})^2} [\Sigma(X_t, \theta)]_{ii}^2,$$

where X_t may replaced with $H^{-1}(U_t, \theta)$ so that the SDE is expressed in terms of U_t. If we use the Cameron-Martin-Girsanov formula in a similar manner as in Section 2.2, we can write the likelihood as

$$\frac{d\mathbb{P}_\theta}{\mathbb{W}^{Y^H} \times Leb_d(Y^H)} \big(U^{mis}, Y\big) \;=\; G(U, \mu_U, I_d) f_{\mathcal{M}}(Y; \theta),$$

or equivalently

$$\frac{d\mathbb{P}_\theta}{\mathbb{W}^{Y^H} \times Leb_d(Y)} \big(U^{mis}, Y\big) \;=\; G(U, \mu_U, I_d) \times \mathcal{N}\left(Y_k^H - Y_{k-1}^H, I_d\right) |J(Y, \theta)|,$$

where $\mathbb{W}^{Y^H}$ is just Wiener measure conditioned on the transformed observations $Y^H{=}H(Y,\theta)$, $\mathcal{N}(Y,V)$ denotes the Gaussian density of Y under 0 mean and covariance V, and $J(Y,\theta)$ is the Jacobian term from the transformation $H(Y,\theta)$. The dominating measure of the likelihood, $\mathbb{W}^{Y^H}$, reflects the distribution of d independent Brownian bridges with Y^H as endpoints and therefore depends on parameters. For this reason we introduce a second transformation

$$z^{\{i\}}(s) = u^{\{i\}}(s) - \frac{(t_k - s)H(y_{k-1}^{\{i\}},\theta)(t_{k-1}) + (s - t_{k-1})h(y_k^{\{i\}},\theta)}{t_k - t_{k-1}},\ t_{k-1} < s < t_k, \tag{14}$$

for all $i \in \{1,\ldots,d\}$, which centers the bridge to start and finish at 0 and preserves the unit volatility. Let $Z = \left(z^{\{1\}},\ldots,z^{\{d\}}\right)'$and the function $U = \eta(Z)$ to be the inverse of 14. The SDE for Z becomes

$$dz_t^{\{i\}} = \mu_{U_t}^{\{i\}}(\eta(Z_t),\theta)dt + dw_t^{\{i\}},\ \forall\, i \in \{1,\ldots,d\}$$

The likelihood may now be written as

$$\frac{d\mathbb{P}_\theta}{\mathbb{W}^{0)} \times Leb_d(Y)}\left(Z^{mis}, h(Y,\theta)\right) \ = \ G(\eta(Z_t), M_U, I_d) \times \mathcal{N}\left(Y_k^H - Y_{k-1}^H, I_d\right)|J(Y,\theta)|, \tag{15}$$

where

$$M_U = \left(\mu_{U_t}^{\{1\}}(\eta(Z_t),\theta),\ldots,\mu_{U_t}^{\{d\}}(\eta(Z_t),\theta)\right)'.$$

The dominating measure of the likelihood provided by 15 does not depend on any parameters, being the product of d independent Brownian bridges that start and finish at 0. The likelihood of (15) may be used to construct an irreducible MCMC scheme which will not degenerate as we increase the amount of augmentation. The stochastic and path integrals involved cannot be solved analytically but they can be evaluated numerically given a sufficiently fine partition of the diffusion path. Note also that, as a result of these transformations, inference will now be based on Z_t rather than X_t. However, the posterior draws of Z_t may be inverted to provide samples from the posterior of X_t.

3.4 Multivariate stochastic volatility models

In the previous subsection we assumed a diffusion with SDE that satisfies (12) so that the transformation of (13) is directly applicable. However, there exist interesting diffusion models outside of this class with a broad range of applications. One famous example of such models is provided by stochastic volatility; see for example Ghysels et al. (1996). Most diffusion driven stochastic volatility models, including those of Hull and White (1987), Stein and Stein (1991) and Heston (1993), belong to the following general class of $2-$dimensional SDEs

$$\begin{pmatrix} dx_t \\ \\ dv_t \end{pmatrix} = \begin{pmatrix} \mu_x(v_t, \theta) \\ \\ \mu_v(v_t, \theta) \end{pmatrix} dt + \begin{pmatrix} \sigma_x(v_t, \theta) & 0 \\ \\ 0 & \sigma_v(v_t, \theta) \end{pmatrix} \begin{pmatrix} db_t \\ \\ dw_t \end{pmatrix}, \tag{16}$$

where b_t and w_t are correlated standard Brownian motions, x_t usually denotes the log price, whose volatility is provided by another diffusion v_t.

Diffusions that satisfy SDEs as in (16) cannot generally be transformed to unit volatility (Aït-Sahalia, 2005), as the reparametrisation of 3.3 requires. Nevertheless, it is still possible to construct an irreducible data augmentation scheme to estimate their parameters. As noted in Chib et al. (2005) the conditional likelihood of x_t, given v_t, is available in closed form and therefore only the paths of v_t need to be imputed to approximate the likelihood. Consequently, as shown in Kalogeropoulos (2007), it suffices to transform v_t itself to unit volatility.

This idea may be coupled with the Cholesky factorisation to handle multivariate stochastic volatility models. We illustrate this for the case of a bivariate Heston model. The scalar Heston model can be written as

$$\begin{aligned} dx_t &= \left(\mu_x - \frac{1}{2}v_t^2\right) dt + \sqrt{v_t} db_t, \\ dv_t &= \kappa \left(\mu_v - v_t\right) dt + \sigma \sqrt{v_t} dw_t. \end{aligned}$$

where b_t and w_t are correlated. We can re-write the top equation, by setting $c = \sqrt{\mu_v}$, to

$$dx_t = \left(\mu_x - \frac{1}{2}v_t^2\right) dt + c\sqrt{\frac{v_t}{\mu_v}} dB_t.$$

Based on the formulation above, a bivariate Heston model may be written as a $4-$dimensional diffusion $X_t = \left(v_t^{\{1\}}, v_t^{\{2\}}, x_t^{\{1\}}, x_t^{\{2\}}\right)'$, with $x_t^{\{1\}}, x_t^{\{2\}}$ denoting the log-prices, and $v_t^{\{1\}}, v_t^{\{2\}}$ their volatilities. The diffusion matrix now has the general form of (2) all of the components of X_t may be correlated. Since (8) holds for each component of X_t, we can define the dispersion matrix of X_t as in (10)

$$\begin{pmatrix} dv_t^{\{1\}} \\ dv_t^{\{2\}} \\ dx_t^{\{1\}} \\ dx_t^{\{2\}} \end{pmatrix} = \begin{pmatrix} \kappa_1 \left(\mu_1 - v_t^{\{1\}}\right) \\ \kappa_2 \left(\mu_2 - v_t^{\{2\}}\right) \\ \mu_3 - \frac{1}{2}(v_t^{\{1\}})^2 \\ \mu_4 - \frac{1}{2}(v_t^{\{2\}})^2 \end{pmatrix} dt \; + \; F_x(X_t, \theta)\, C\, dB_t, \tag{17}$$

where now B_t is a 4−dimensional Brownian motion with independent components,

$$F_x(X_t, \theta) = diag \left\{ \sqrt{v_t^{\{1\}}}, \sqrt{v_t^{\{2\}}}, \frac{\sqrt{v_t^{\{1\}}}}{\mu_1}, \frac{\sqrt{v_t^{\{2\}}}}{\mu_2} \right\},$$

and C is the lower triangular Cholesky matrix whose entries C_{ij} may be seen as a 1-1 transformation of parameter vector containing the correlations ρ_{ij}, and also σ_1, σ_2, $\sqrt{\mu_1}$ and $\sqrt{\mu_2}$.

Regarding the likelihood, consider again a pair of successive observations, Y_{k-1}, Y_k with $Y_k = (y_k^{\{3\}}, y_k^{\{4\}})$, for $x_t^{\{1\}}, x_t^{\{2\}}$. Conditional on $v_t^{\{1\}}, v_t^{\{2\}}$, and therefore also on their corresponding Brownian components $b_t^{\{1\}}, b_t^{\{2\}}$, the likelihood for Y_k is a bi-variate Gaussian with mean

$$\begin{pmatrix} y_{k-1}^{\{3\}} + \int_{t_{k-1}}^{t_k} \left(\mu_3 - \frac{1}{2}(v_s^{\{1\}})^2\right) ds + C_{31} \int_{t_{k-1}}^{t_k} \sqrt{\frac{v_s^{\{1\}}}{\mu_1}} db_s^{\{1\}} + C_{32} \int_{t_{k-1}}^{t_k} \sqrt{\frac{v_s^{\{1\}}}{\mu_1}} db_s^{\{2\}} \\ y_{k-1}^{\{4\}} + \int_{t_{k-1}}^{t_k} \left(\mu_4 - \frac{1}{2}(v_s^{\{2\}})^2\right) ds + C_{41} \int_{t_{k-1}}^{t_k} \sqrt{\frac{v_s^{\{2\}}}{\mu_2}} db_s^{\{1\}} + C_{42} \int_{t_{k-1}}^{t_k} \sqrt{\frac{v_s^{\{2\}}}{\mu_2}} db_s^{\{2\}} \end{pmatrix},$$

and covariance matrix

$$\begin{pmatrix} \int_{t_{k-1}}^{t_k} C_{33}^2 \frac{v_s^{\{1\}}}{\mu_3^2} ds & \int_{t_{k-1}}^{t_k} C_{33} C_{43} \frac{\sqrt{v_s^{\{1\}} v_s^{\{2\}}}}{\mu_3 \mu_4} ds \\ \int_{t_{k-1}}^{t_k} C_{33} C_{43} \frac{\sqrt{v_s^{\{1\}} v_s^{\{2\}}}}{\mu_3 \mu_4} ds & \int_{t_{k-1}}^{t_k} (C_{43}^2 + C_{44}^2) \frac{v_s^{\{2\}}}{\mu_4^2} ds \end{pmatrix}.$$

The integrals above cannot be computed analytically, but the augmented path of $v_t^{\{1\}}, v_t^{\{2\}}$ enables accurate numerical approximations of them.

The remaining part of the likelihood may be obtained through the reparametrisation recipe of Section 3.3, modified according to the observation regime of the volatility. In some cases the volatility may be entirely unobserved, leading to a partially observed diffusion. Nevertheless alternative formulations are available, where information from option prices is used to construct exact or noisy volatility observations; see for example Aït-Sahalia and Kimmel (2005), Chernov and Ghysels (2000) and Kalogeropoulos et al. (2007). In the presence of exact observations the transformations of (13) and (14) may be used. Note that transformation to unit volatility refers to the 2-dimensional diffusion $(v_t^{\{1\}}, v_t^{\{2\}})'$, rather than the entire X_t. For the bivariate Heston model it takes the following form

$$U_t = H(X_t, D) = D^{-1} G_x(X_t),$$

where

$$G_x(X_t) = \left(2\sqrt{x_t^{\{1\}}}, 2\sqrt{x_t^{\{2\}}}\right)',$$

and D is a block of C containing the C_{ij} entries with $i, j = \{1, 2\}$. If the observations are noisy or they do not exist at all, the transformation of (14) may be replaced with

$$Z^{\{i\}}(s) = U^{\{i\}}(s) - U_0,\ 0 < s < t_n,$$

and the $\mathcal{N}\left(Y_k^H - Y_{k-1}^H, I_d\right)|J(Y, \theta)|$ part of the likelihood should be replaced with the relative noise density or removed accordingly.

The above likelihood specification can be applied to all multivariate stochastic volatility models that satisfy the SDE of 16. For more complex models, the framework of Golightly and Wilkinson (2007) or time change transformations of Kalogeropoulos et al. (2007) may be combined with the Cholesky factorisation.

4 MCMC implementation

Based on the likelihood specifications of the previous section, it is now possible to construct an irreducible data augmentation MCMC scheme. The algorithm may be divided into three parts: the updates of the diffusion paths Z^{mis}, the parameters of the dispersion matrix $\Sigma(X_t, \theta)$ and those of the drift $M(X_t, \theta)$. Generally, the updates of the drift parameters may be executed using standard random walk Metropolis techniques, although for some diffusion models the full conditionals may be analytically tractable and Gibbs steps may be used instead. Hence, in the next two subsections we provide some details regarding the updates of the diffusion paths and the volatility parameters.

4.1 Updating the imputed paths

There exist several options for carrying out this step and most of them are based on an independence sampler. For discretely observed diffusions the augmented path may be divided into $n \times d$ diffusion bridges connecting the observed points, and each one of them may be updated in turn. The full conditional of Z^{mis} may be written as

$$\frac{d\mathbb{P}_\theta}{d\mathbb{W}^0}(Z^{mis}|Y) = G(\eta(Z_t), M_U, I_d)\frac{f_{\mathcal{M}}(Y; A)}{f_{\mathcal{X}}(Y; A)} \propto G(\eta(Z_t), M_U, I_d), \tag{18}$$

where $f_{\mathcal{X}}(Y; A)$ is the density of Y with respect to the Lebesgue measure under $\mathbb{P}_\theta$. Note that this expression will be slightly different for stochastic volatility models.

The dominating measure of the likelihood $\mathbb{W}^0$, in other words a Brownian bridge, may be used as the proposal distribution for the independence sampler. Based on (18), the algorithm will then contain the following steps

- `Step 1: Propose a Brownian bridge from` t_{k-1} `to` t_k.
- `Step 2: Substitute into` i`-th dimension and form` Z_t^*.
- `Step 3: Accept with probability:`

$$\min\left\{1, \frac{G(\eta(Z_t^*), M_U, I_d)}{G(\eta(Z_t), M_U, I_d)}\right\}.$$

- `Repeat for all` $k = 1, \ldots n$ `and` $i = 1, \ldots, d$.

The algorithm above takes advantage of the transformation to unit volatility and splits the path into $n \times d$ independent, under the dominating measure, bridges. Alternative proposals are available such as the diffusion bridges introduced in Durham and Gallant (2002) and Delyon and Hu (2007), which can be adapted in a MCMC setting through the reparametrisation framework of Golightly and Wilkinson (2007). Another option is to propose local moves of the paths in the spirit of Beskos et al. (2006b). This approach may be viewed as a random walk metropolis in the space of diffusion bridges. Note however that this technique requires bridges with unit volatility, and therefore it can only be used for correlated diffusions through the reparametrisation framework of this paper.

Further increase in the acceptance rate may be achieved by choosing a proposal distribution which is closer to the target $\mathbb{P}_\theta$, for example a linear diffusion bridge. Suppose that we propose from another diffusion bridge distribution, denoted by $\mathbb{L}^0$, with drift L. We can now write:

$$\frac{d\mathbb{P}_\theta}{d\mathbb{L}^0}(Z_{mis}|Y) = \frac{d\mathbb{P}_\theta/d\mathbb{W}^0}{d\mathbb{L}_0/d\mathbb{W}^0}(Z_{mis}|Y) \propto \frac{G(\eta(Z_t), M_U, I_d)}{G(\eta(Z_t), L, I_d)} \tag{19}$$

Based on (19), the corresponding algorithm, termed as method B in Roberts and Stramer (2001), will consist of the following steps:

- `Step 1: Propose a Brownian bridge from` t_{k-1} `to` t_k.

- `Step 2:  Substitute into` i`-th dimension and form` Z_t^*`.`

- `Step 3:  Accept with probability:`

$$\min\left\{1, \frac{G(\eta(Z_t^*), M_U, I_d) G(\eta(Z_t), L, I_d)}{G(\eta(Z_t^*), L, I_d) G(\eta(Z_t), M_U, I_d)}\right\}.$$

- `Repeat for all` $k = 1, \ldots n$ `and` $i = 1, \ldots, d$`.`

However, low acceptance rates may still occur, especially in sparse datasets. In such cases, each bridge may be further split into smaller blocks and updating strategies based on overlapping or random sized blocks may be advocated; see Kalogeropoulos (2007) and Chib et al. (2005) for more details. These techniques may also be used in partially observed diffusions, for example in stochastic volatility models, where some components of the diffusion may be observed with error or not be observed at all.

4.2 Updating the volatility parameters

As mentioned earlier, the parameter updates of the diffusion matrix $A(X_t, \theta)$ are not trivial. Their full conditional posterior is generally not available in closed form, and Metropolis steps are inevitable. The construction of such steps has to ensure that the covariance matrix structure of $A(X_t, \theta)$ is preserved. At the same time, it is desirable to achieve a reasonably high acceptance rate of the proposed moves for a good mixing of the MCMC algorithm. While the former may be implemented by using an appropriate distribution for symmetric positive definite matrices, such as the Wishart distribution, it is extremely difficult to guarantee the latter, especially for high dimensional diffusions.

The Cholesky factorisation introduced in this paper may be of help in such cases. Specifically, the step of updating the constants c_i, and the correlations ρ_{ij}, with $i, j \in \{1, \ldots, d\}$ and $i < j$, may be replaced by componentwise updates of the Cholesky matrix C. In contrast with the correlations ρ_{ij}, the restrictions implied by the symmetric and positive definite diffusion matrix $A(X_t, \theta)$ may be enforced on the elements of C in a straightforward manner, as only the positivity of the diagonal entries is required.

Hence, the updates of C_{ij}'s may be implemented through standard random walk Metropolis steps.

Note that (c_i,ρ_{ij}) and C_{ij} are linked through

$$S_x(X_t,\theta)\; R\; S_x(X_t,\theta) \;=\; F_x(X_t,\theta)\; V\; F_x(X_t,\theta) \;=\; A(X_t,\theta), \tag{20}$$

where R is the correlation matrix and V is defined in (9). It is not hard to see that they are linked with an 1-1 mapping which is the solution of the system in (20) with $d(d+1)/2$ equations and unknowns. Hence, the draws from the posterior of C may be transformed back at any time, to obtain draws from the posterior of (c_i,ρ_{ij}).

5 Simulation based experiments

In this section we illustrate and test our data augmentation scheme on a $3-$dimensional CIR model. In other words, we consider a $3-$dimensional diffusion $X_t = (x_t^{\{1\}}, x_t^{\{2\}}, x_t^{\{3\}})'$ with linear drift for each component $\kappa_i(\mu_i - x_t^{\{i\}})$, the CIR formulation of the volatility, $\sigma_i\sqrt{x_t^{\{i\}}}$, and correlations between all the components, ρ_{ij}, $i = 1,2,3,\ j < i$. This model may be useful for the analysis of interest rates time series, where the cross-correlations may be substantial. Notice that our framework allows for more general drift and volatility formulations but the main focus of this simulation experiment lies mainly in the correlations ρ_{ij}. The dispersion matrix of the multi-dimensional diffusion X_t may be defined as in (10), with

$$F_x(X_t,\theta) = diag\left\{\sqrt{x_t^{\{1\}}}, \sqrt{x_t^{\{2\}}}, \sqrt{x_t^{\{3\}}}\right\},$$

and C being the lower triangular matrix from the Cholesky decomposition, whose entries C_{ij}, substitute the parameters σ_i and ρ_{ij}. The likelihood reparametrisation requires a transformation to unit volatility which is given by

$$U_t = H(X_t, C) = C^{-1} G_x(X_t),$$

with

$$G_x(X_t) = \left(2\sqrt{x_t^{\{1\}}}, 2\sqrt{x_t^{\{2\}}}, 2\sqrt{x_t^{\{3\}}}\right)'.$$

The second transformation is that of (14), and the likelihood may be obtained from (15). To complete the model formulation we assign non-informative priors: $p(\theta) \propto \theta^{-1}$ for the positive parameters κ_i, μ_i, C_{ii} and $p(\theta) \propto 1$ for the rest $(C_{ij}, i > j)$.

We simulated 500 equidistant observations (apart from the initial point) at times $\{t_k = k,\ k = 0\ldots, n\}$ with $t_n = 500$. Several MCMC runs, with different numbers of imputed points m=$\{20, 40, 60, 80\}$, were examined. This was done to monitor the autocorrelation as well as the approximation error of the likelihood in relation with the level of augmentation. The acceptance rate of the independence sampler used for the path updates was 98.14%, raising no concerns regarding its performance. Figure 1 shows autocorrelation plots for the posterior draws of the C matrix components. There is no sign of any increase to raise suspicions against the irreducibility of the chain. Figure 2 depicts density plots for some parameters as well as the log-likelihood which may be seen as an appropriate diagnostic plot for the quality of the approximations. Densities for $m = 60$ and $m = 80$ look similar and therefore the argument that their level of augmentation is sufficient appears to be plausible. The plots of Figure 2 and the results of Table 1, which contains summaries of the parameter posterior draws for $m = 80$, are in good agreement with the true values of the parameters.

[Figure 1 about here.]

[Figure 2 about here.]

[Table 1 about here.]

6 Application: EUR/USD and GBP/USD exchange rates

The dataset consists of roughly two years of daily exchange EUR/USD and GBP/USD rates, specifically from the 3rd of January 2005 to 22nd of December 2006. We denote these rates with $r^{eur/usd}$ and $r^{gbp/usd}$ and their logarithms with $Y^{eur/usd}$ and $Y^{gbp/usd}$ respectively. Our dataset also contains the corresponding month implied volatilities constructed from options made on the currency pairs. The data are plotted in Figure 3.

[Figure 3 about here.]

We use the implied volatilities of the currency pairs to construct proxies for their actual volatilities, denoted with $IV^{eur/usd}$ and $IV^{gbp/usd}$. For simplicity, these proxies are assumed to be exact observations of the volatilities. Alternative assumptions are possible, such as their adjustment (Aït-Sahalia and Kimmel, 2005), or a formulation with noisy observations. Table 2 provides several descriptive statistics including the correlation matrix of the 4−dimensional time series containing the implied volatilities and the log-exchange rates $Y = \left(IV^{eur/usd}, IV^{gbp/usd}, Y^{eur/usd}, Y^{gbp/usd}\right)$.

[Table 2 about here.]

Note that some correlations appear to be substantial and should be taken into account in the analysis of the data. Hence we fit the bivariate Heston model to the 4−dimensional time series Y using the MCMC data augmentation scheme of this paper. Section 3.4 provides details on the reparametrised likelihood for the data. For reasons of model parsimony, we only consider correlations between the pairs $\left(IV^{eur/usd}, IV^{gbp/usd}\right)$ and $\left(Y^{eur/usd}, Y^{gbp/usd}\right)$, and set the remaining ones $(\rho_{31},\rho_{32},\rho_{41},\rho_{42})$ to zero. This is in line with Table 2 and some preliminary analysis which considered all possible correlations. Note that the parameters of C that need to be updated are just C_{11}, C_{21}, C_{22} and C_{43}, as C_{33} and C_{44} are redundant and the remaining entries are equal to zero like the corresponding correlations. In other words, there exists a 1-1 mapping between the diffusion matrix elements $(\sigma_1,\sigma_2,\rho_{21},\rho_{43})$ and $(C_{11},C_{21},C_{22},C_{43})$. We complete the model by assigning non-informative priors as in the previous section: $p(\theta) \propto \theta^{-1}$ for the positive parameters $(\kappa_1,\kappa_2,\mu_1,\mu_2,C_{11},C_{22})$ and $p(\theta) \propto 1$ for the rest $(\mu_3,\mu_4,C_{21},C_{43})$.

As before, several MCMC runs with different numbers of imputed points $m=\{10, 20, 40\}$ were used. The data, referring to business days, were assumed to be equidistant and the time was measured in years. Again, the acceptance rate of the independence sampler used for the path updates was particularly high 99.16%. The autocorrelation plots of draws from the posterior of the parameters C_{11},C_{21},C_{22}, and C_{43}, in Figure 4, reveal no sign of any increase in the level of augmentation.

[Figure 4 about here.]

Regarding the approximation error due to the discretisation of the diffusion path, the density plots from the posterior draws of some parameters and the log-likelihood, in Figure 5, provide convergence evidence for the approximating sequence of the data augmentation scheme.

[Figure 5 about here.]

Table 3 contains summaries of the parameter posterior draws, where both correlations appear to be high. Note that the non-parametric estimates of Table 2 are based on the quadratic variation process and are therefore amenable to bias due to the discretisation of the diffusion path. On the other hand, the discretisation error of the model estimates may become arbitrary small. The posterior mean or median values provide point estimates of the parameters which may be used for option pricing purposes. Alternatively, the samples from their posterior of the parameters may be used in a Bayesian option pricing framework. In any case, it may be useful to take into account the correlated market structure of the log-exchange rate and their impled volatilities.

[Table 3 about here.]

7 Discussion

In this paper we introduced a parametrisation framework based on the Cholesky decomposition, for handling correlations of multi-dimensional diffusions in a Bayesian MCMC setting. This framework facilitates componentwise updates of the diffusion matrix, in a way so that its positive definite structure is preserved. It may therefore be of substantial value in high dimensional diffusion models. The Cholesky factorisation was used in connection with data augmentation and therefore applies to both directly and partially observed diffusions. In order to overcome degenerate MCMC algorithms, the likelihood reparametrisation of Roberts and Stramer (2001) was generalised to several multi-dimensional diffusions, including stochastic volatility models, thus providing a stand alone solution to the problem. Being a data augmentation scheme, our MCMC algorithm is based on an approximation of the likelihood, whose error may become arbitrarily small by simply increasing the level of augmentation.

Nonetheless, the Cholesky factorisation of the diffusion matrix may be coupled with alternative, to data augmentation, techniques for approximating the likelihood. The exact inference framework of Beskos et al. (2006a) and the analytic likelihood expansions of Aït-Sahalia (2005) provide such examples with appealing properties: the former eliminates entirely the error due to the discretisation of the diffusion path, whereas the latter provides closed form expressions of the likelihood. On the other hand, their generalisation to partially observed diffusion may present major difficulties.

Apart from the updates of the diffusion matrix parameters, our MCMC algorithm differs from other data augmentation schemes, such as those of Chib et al. (2005) and Golightly and Wilkinson (2007), in the proposal distribution of the independence sampler involved in the updates of the diffusion paths. Under these schemes, the proposal may either be the multi-dimensional bridge of the of Durham and Gallant (2002), or alternatively that of Delyon and Hu (2007), with the target diffusion matrix. Current work investigates the behavior of all existing approaches in different settings regarding the dimensionality of the diffusion, the amount of correlation, and the sparseness of the data.

8 Acknowledgements

Part of the work was carried out during a visit to Lancaster funded through the EU Marie Curie training scheme. The data of Section 6 were used with the kind permission of Citigroup.

References

Aït-Sahalia, Y. (2005). Closed form likelihood expansions for multivariate diffusions. *Annals of Statistics*. To appear.

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

Beskos, A., Papaspiliopoulos, O., Roberts, G., and Fearnhead, P. (2006a). 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., Roberts, G. O., Stuart, A., and Voss, J. (2006b). A MCMC method for diffusion bridges. Submitted.

Chernov, M. and Ghysels, E. (2000). A study towards a uinified approach to the joint estimation of objective and risk neutral measures for the purposes of options valuation. *Journal od Financial Economics*, 56:407–458.

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

Cox, J. C., Ingersoll, J. E., and Ross, S. A. (1985). A theory of the term structure of interest rates. *Econometrica*, 53:385–407.

Daniels, M. and Kass, R. (1999). Nonconjugate bayesian estimation of covariance matrices in hierarchical models. *Journal of the American Statistical Association*, 94:1254–1263.

Delyon, B. and Hu, Y. (2007). Simulation of conditioned diffusions and applications to parameter estimation. *Stochastic Processes and Application*. To appear.

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

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.

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., Roberts, G., and Dellaportas, P. (2007). Inference for stochastic volatility models using time change transformations. Submitted.

Pedersen, A. R. (1995). A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. *Scandinavian Journal of Statistics. Theory and Applications*, 22(1):55–71.

Pinheiro, J. and Bates, D. (1996). Unconstrained parametrizations for variance-covariance matrices. *Statistics and Computing*, 6(3):289–296.

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 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.

A Proofs of propositions

Proof of proposition 3.1:

The proof is based on he reducibility condition of (11), for which we need the inverse of $\Sigma(X_t, \theta)$

$$\Sigma(X_t, \theta)^{-1} \ = \ (F_x(X_t, \theta)\ C)^{-1} \ = \ C^{-1}\ F_x(X_t, \theta)^{-1}.$$

In coordinate form the above writes

$$[\Sigma(X_t, \theta)^{-1}]_{ij} = [C^{-1}]_{ij} f^{\{j\}}(x_t^{\{j\}}, \theta)^{-1},\ \forall\ i, j \in \{1, \ldots, d\}.$$

Hence, it is not hard to see that the reducibility condition of Aït-Sahalia (2005) holds because

$$\frac{\partial[\Sigma(X_t, \theta)^{-1}]_{ij}}{\partial x_t^{\{k\}}} = \frac{\partial[\Sigma(X_t, \theta)^{-1}]_{ik}}{\partial x_t^{\{j\}}} = 0,\ \forall\ i, j, k \in \{1, \ldots, d\},\ \text{with}\ j < k$$

Proof of proposition 3.2:

The diffusion matrix of U_t should be a $d-$dimensional identity matrix, therefore by Ito's lemma we get

$$\nabla H(X_t, \theta)\ A\ (\nabla H(X_t, \theta))' \ = \ I_d \tag{21}$$

Consider a transformation of the form

$$H(X_t, \theta) \ = \ B\ G_x(X_t, \theta),$$

where B is an arbitrary $d \times d$ matrix, independent of X_t.

We can write

$$\nabla H(X_t, \theta) \ = \ B\ D_G(X_t, \theta),$$

where $D_G(X_t, \theta)$ is a diagonal matrix with

$$[D_G(X_t, \theta)]_{ii} = f^{\{i\}}(x_t^{\{i\}}, \theta)^{-1},\ i = 1, \ldots, d.$$

Indeed, the $k-$th row of $\nabla H(X_t, \theta)$ equals

$$\nabla H(X_t, \theta) = \nabla \left(\sum_{j=1}^{d} B_{kj} g^{\{i\}}(x_t\{j\}, \theta) \right) = (B_{k1}, \ldots, B_{kd}) \ D_G(X_t, \theta).$$

If we substitute on (21), using also (10), we get

$$B \ D_G(X_t, \theta) \ F_x(X_t, \theta) \ C \ C' \ F_x(X_t, \theta) \ D_G(X_t, \theta) \ B' \ = \ I_d,$$

which since $D_G(X_t, \theta) \ F_x(X_t, \theta) \ = \ F_x(X_t, \theta) \ D_G(X_t, \theta) \ = \ I_d$ becomes

$$B \ C \ C' \ B' \ = \ I_d,$$

which is satisfied if we set $B = C^{-1}$.

List of Figures

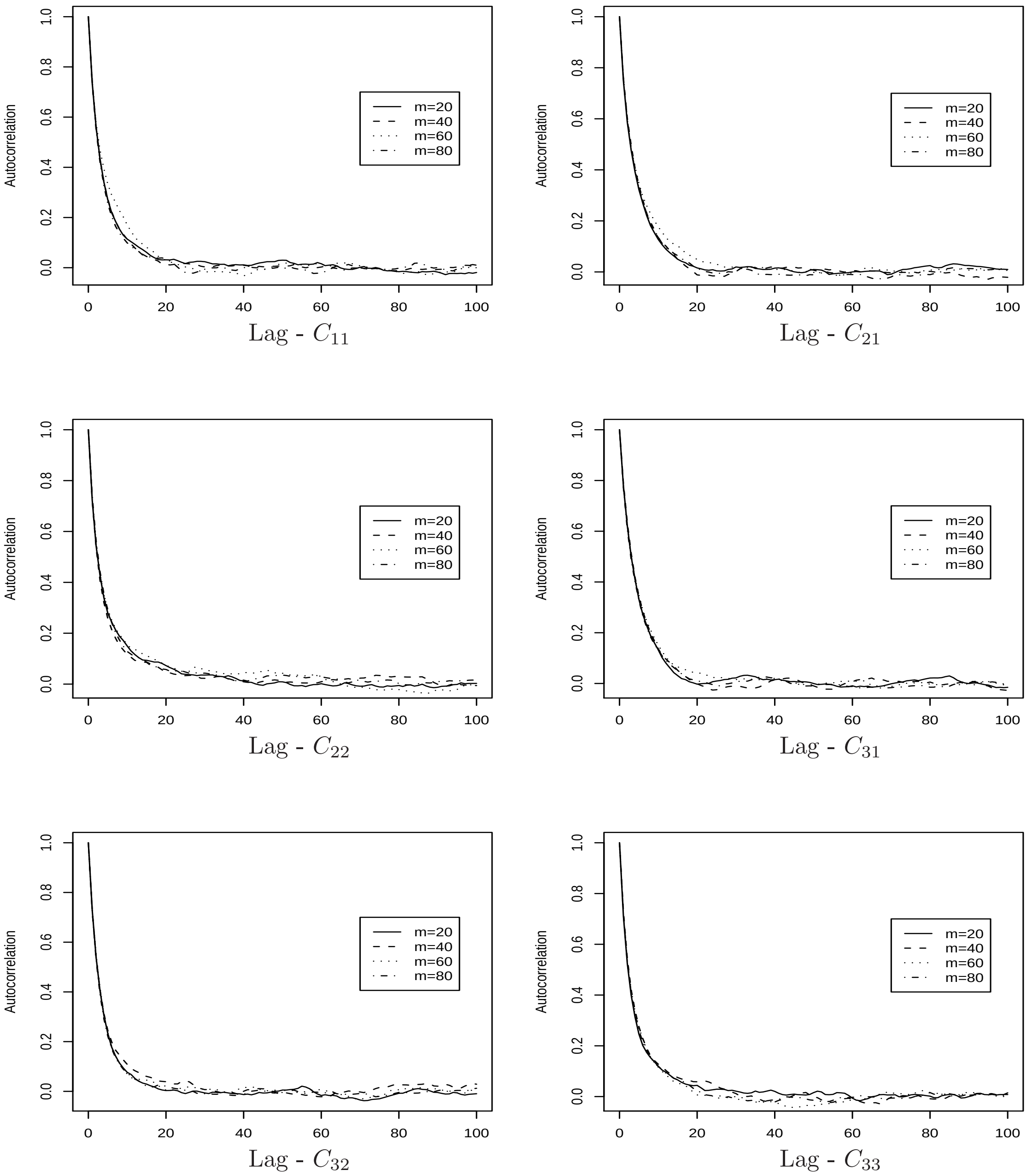


Figure 1: Autocorrelation plots for the posterior draws of the C matrix entries for different numbers of imputed points ($m = 20, 40, 60, 80$). Simulated data.

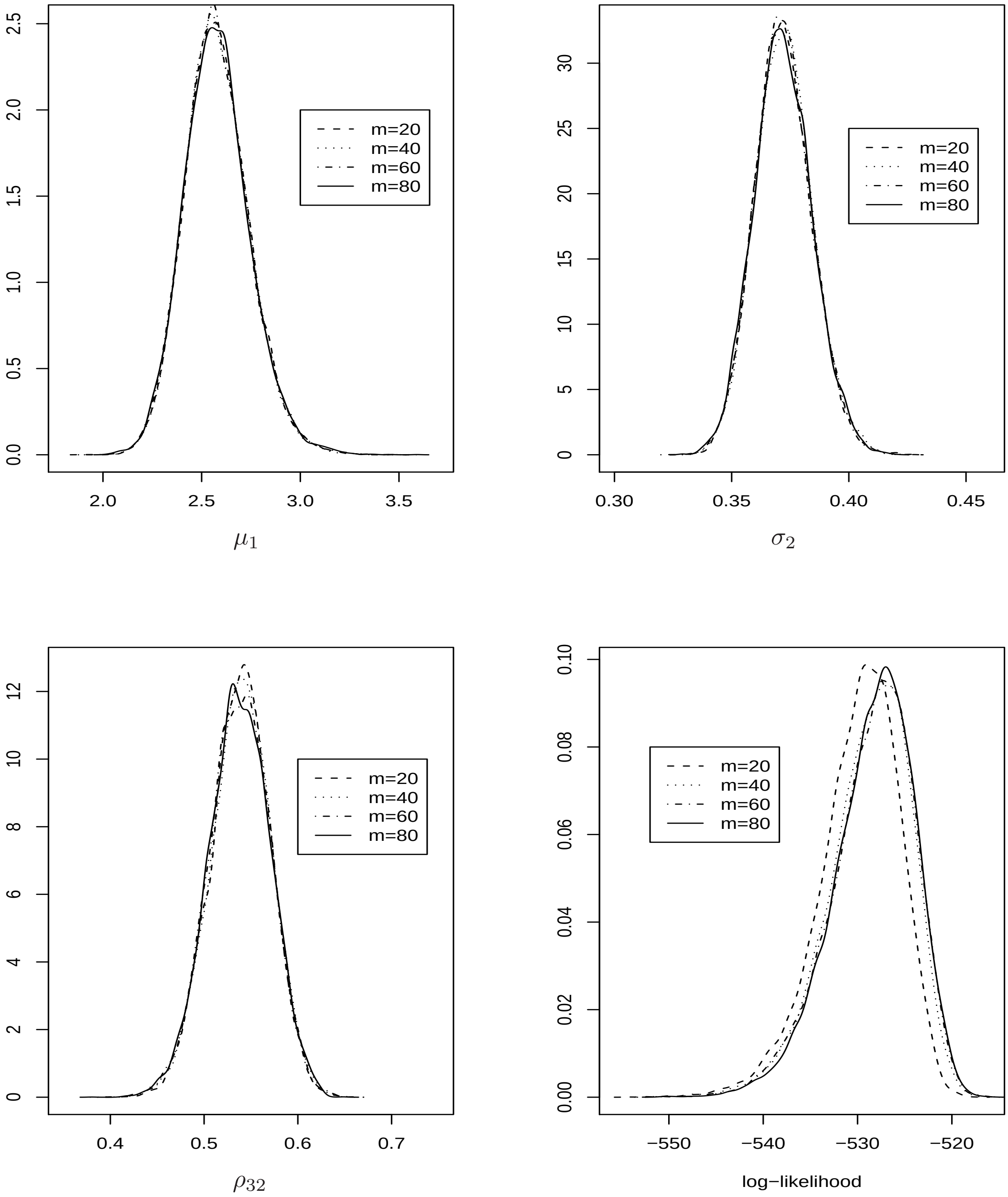
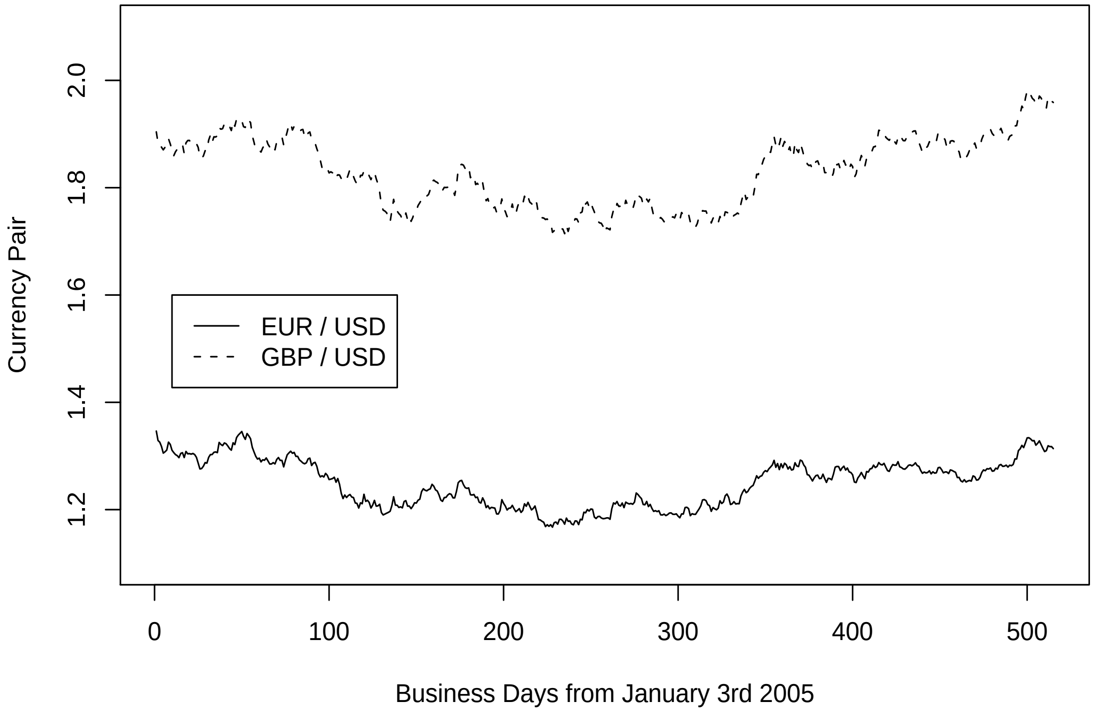


Figure 2: Kernel densities of the posterior draws for some parameters (μ_1, σ_2, ρ_{32}) and the log-likelihood, for different numbers of imputed points ($m = 20, 40, 60, 80$). Simulated data.



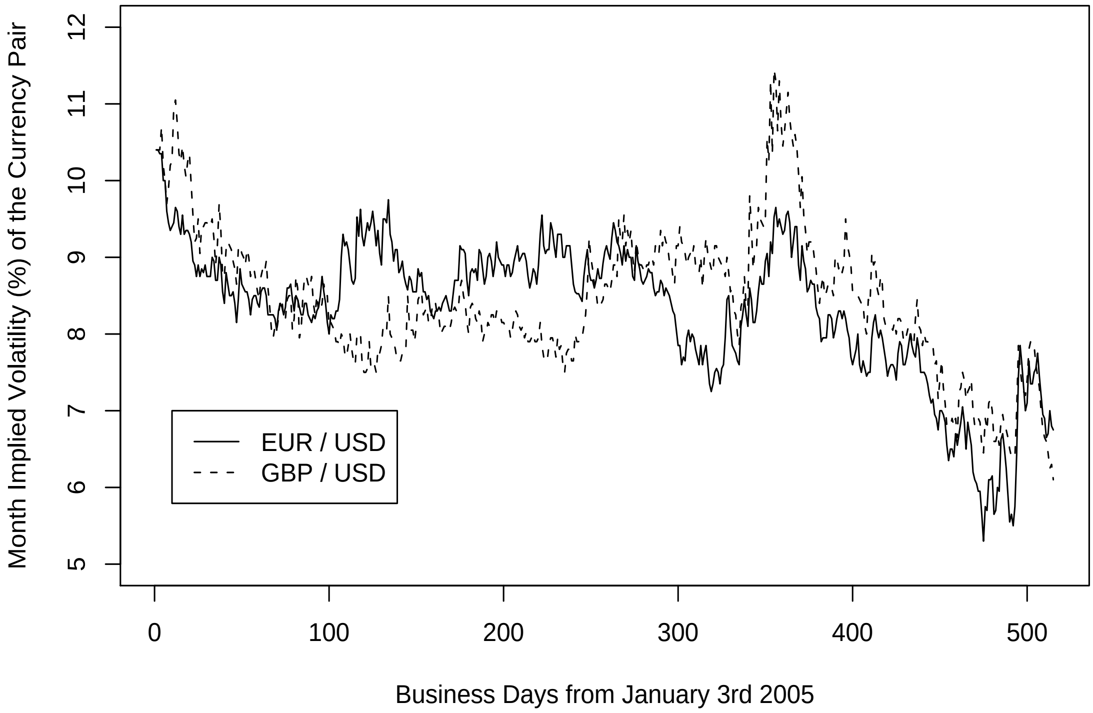


Figure 3: Daily EUR/USD and GBP/USD rates (up) and their month implied volatilities (%) (down) from 3rd of January 2005 to 22nd of December 2006.

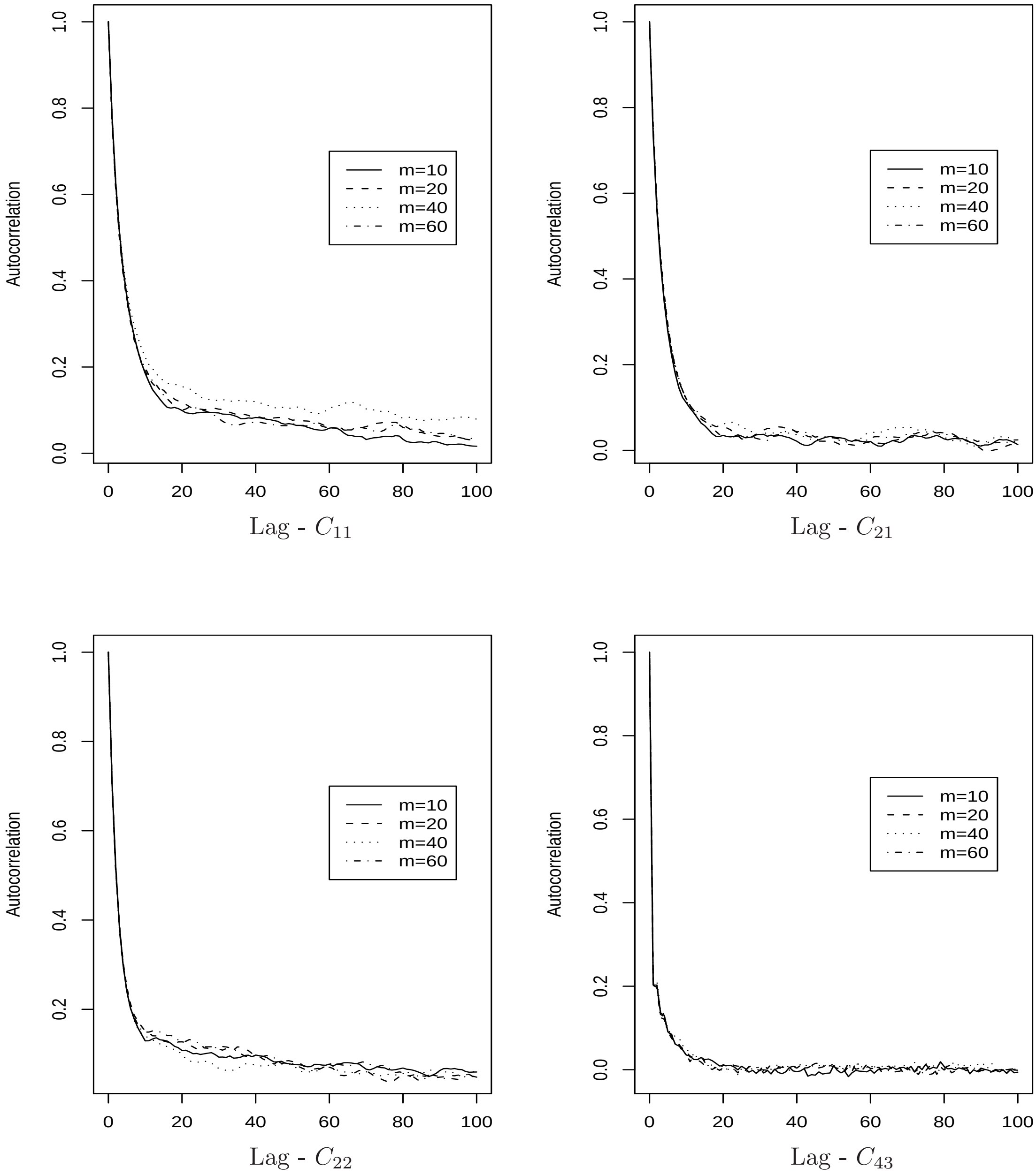


Figure 4: Autocorrelation plots for the posterior draws of the C matrix entries for different numbers of imputed points ($m = 10, 20, 40$). EUR/USD and GBP/USD exchange rates dataset.

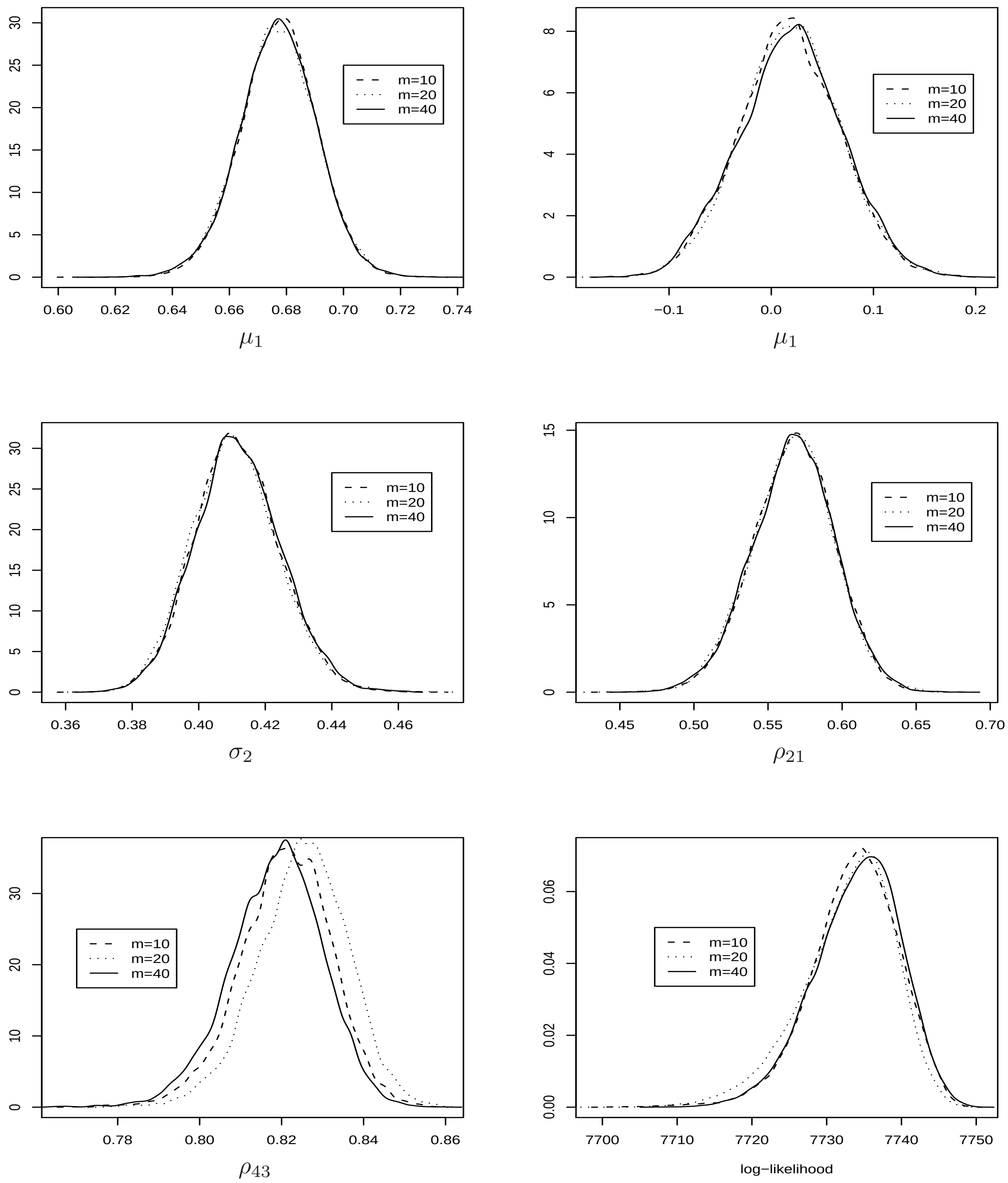


Figure 5: Kernel densities of the posterior draws for some parameters (μ_1, μ_1, σ_2, ρ_{21}, ρ_{43}) and the log-likelihood, for different numbers of imputed points ($m = 10, 20, 40$). EUR/USD and GBP/USD exchange rates dataset.

List of Tables

Parameter	True Value	Posterior mean	Posterior SD	Posterior median
κ_1	0.2	0.174	0.025	0.174
κ_2	0.15	0.123	0.031	0.121
κ_3	0.22	0.223	0.030	0.224
μ_1	2.5	2.578	0.167	2.571
μ_2	3.0	2.986	0.366	2.951
μ_3	2.0	1.908	0.094	1.905
σ_1	0.45	0.434	0.016	0.434
σ_2	0.35	0.372	0.012	0.372
σ_3	0.4	0.401	0.014	0.402
ρ_{21}	0.45	0.480	0.034	0.480
ρ_{31}	0.35	0.318	0.041	0.319
ρ_{32}	0.55	0.537	0.033	0.538

Table 1: Summaries of the posterior draws of the model parameters for $m = 80$. Simulated dataset.

	Mean	St. Deviation	Median	
$IV^{eur/usd} \times 100$	0.693	0.076	0.708	
$IV^{gbp/usd} \times 100$	0.704	0.078	0.696	
$r^{eur/usd}$	1.2499	0.045	1.2578	
$r^{gbp/usd}$	1.8304	0.066	1.8375	
	Correlation Matrix			
$\Delta IV^{eur/usd}$	1			
$\Delta IV^{gbp/usd}$	0.5551	1		
$\Delta Y^{eur/usd}$	0.0148	0.0101	1	
$\Delta Y^{gbp/usd}$	0.0119	0.0075	0.8093	1

Table 2: Descriptive statistics for EUR/USD and GBP/USD exchange rates and their implied volatilities.

Parameter	Posterior mean	Posterior SD	Posterior median
κ_1	0.153	0.023	0.153
κ_2	0.206	0.030	0.204
$\mu_1 \times 100$	0.677	0.014	0.677
$\mu_2 \times 100$	0.689	0.012	0.690
μ_3	0.001	0.053	0.001
μ_4	0.019	0.049	0.019
$\sigma_1 \times 100$	0.343	0.010	0.343
$\sigma_2 \times 100$	0.411	0.013	0.411
ρ_{21}	0.567	0.028	0.567
ρ_{43}	0.821	0.011	0.821

Table 3: Summaries of the posterior draws of the model parameters for $m = 60$. EUR/USD and GBP/USD exchange rates dataset.