



**ATHENS UNIVERSITY
OF ECONOMICS AND BUSINESS**

DEPARTMENT OF STATISTICS

**BAYESIAN INFERENCE FOR
MULTIDIMENSIONAL DIFFUSION PROCESSES**

By

Konstantinos P. Kalogeropoulos

DOCTORATE THESIS

Submitted to the Department of Statistics
of the Athens University of Economics and Business
in partial fulfilment of the requirements for
the degree of Doctor of Philosophy in Statistics

Athens, Greece
November 2006



**ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ
ΑΘΗΝΩΝ**

ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ

**ΣΤΑΤΙΣΤΙΚΗ ΚΑΤΑ ΒΑΥΕΣ ΓΙΑ
ΠΟΛΥΔΙΑΣΤΑΤΕΣ ΔΙΑΔΙΚΑΣΙΕΣ ΔΙΑΧΥΣΗΣ**

Κωνσταντίνος Π. Καλογερόπουλος

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

Που υποβλήθηκε στο Τμήμα Στατιστικής
του Οικονομικού Πανεπιστημίου Αθηνών
ως μέρος των απαιτήσεων για την απόκτηση
Διδακτορικού Διπλώματος στη Στατιστική

Αθήνα
Νοέμβριος 2006

Contents

1	Introduction	1
1.1	Diffusion models	1
1.2	Preliminaries	3
1.2.1	Brownian Motion	3
1.2.2	Diffusion processes	5
1.2.3	Bayesian Inference using Markov chain Monte Carlo	8
1.3	Statistical inference for Diffusion processes	9
1.3.1	Likelihood for Diffusions	10
1.3.2	Inference assuming ‘thin’ data	12
1.4	Review on likelihood based inference for diffusions	14
1.4.1	Inference using approximations of the transition density	15
1.4.2	Bayesian MCMC techniques using data augmentation	20
1.4.3	Sequential techniques for partially observed diffusions	24
1.4.4	Concluding remarks	25
1.5	Our contribution	26
2	Bayesian Inference for Discretely Observed Multidimensional Diffusions	29
2.1	Introduction - Notation	29
2.2	Data augmentation	30
2.3	Reducibility issues	32
2.3.1	A toy example	32
2.3.2	Measure theoretic probability framework	34

2.4	Transformation of the diffusion path	37
2.4.1	Reparametrisation	37
2.4.2	Which diffusions can we handle?	40
2.4.3	Stochastic volatility models	43
2.5	Markov chain Monte Carlo implementation	44
2.5.1	Updating the imputed paths	44
2.5.2	Updating the volatility parameters	46
2.6	Extensions	47
2.6.1	Change-points in volatility	47
2.6.2	Multivariate NLCAR models	50
2.7	Simulations	52
2.8	Real data example	56
2.9	Concluding remarks	61
3	Bayesian Inference for a Large Class of Stochastic Volatility Models	63
3.1	Introduction	63
3.2	Inference for stochastic volatility models	65
3.3	Reparametrisation	66
3.3.1	The need for a reparametrisation	66
3.3.2	A suitable likelihood parametrisation	67
3.4	Data augmentation scheme	70
3.5	Extensions	72
3.5.1	Multifactor and Multivariate Stochastic Volatility Models	72
3.5.2	Applications beyond stochastic volatility	73
3.6	Simulations	74
3.6.1	Data augmentation without reparametrisation	74
3.6.2	Reparametrised scheme applied to the Heston model	76
3.7	Connection with option pricing	81
3.7.1	Option pricing under stochastic volatility	81
3.7.2	Determining the option pricing parameters	83
3.7.3	Using option prices for inference purposes	84

3.8	Application: S&P 500 - VIX data	87
3.8.1	Data description	87
3.8.2	The models	88
3.8.3	Results	90
3.9	Concluding remarks	92
4	Bayesian Inference for General Stochastic Volatility Models Using Time Change Transformations	97
4.1	Introduction	97
4.2	The need for a time change transformation	99
4.3	Time change transformations for scalar diffusions	100
4.4	MCMC implementation	104
4.4.1	Three time scales	104
4.4.2	Updating the paths of X	105
4.4.3	Updating time scale parameters	107
4.5	Time change for stochastic volatility models	108
4.5.1	Main category	108
4.5.2	Incorporating leverage effect	112
4.5.3	Other extensions	113
4.6	Simulation example	113
4.7	Example:US treasury	115
4.8	Concluding remarks	120

List of Tables

2.1	Summaries of the posterior draws for the simulation example.	54
2.2	Summaries of the posterior draws for the real data example. Spanish-European rates dataset.	61
3.1	Posterior means and standard deviations of the parameters versus their true values for the Heston model.	78
3.2	Posterior means and standard deviations of parameters from the Heston, GARCH and log OU models for the SP500/VIX data.	92
4.1	Summaries of the posterior draws for the simulation example of Chapter 3 for $m = 50$	115
4.2	Summaries of the posterior draws for the stochastic volatility model of Weekly 3-month US Treasury bill rates.	120

List of Figures

2.1	Autocorrelation plots of posterior draws of σ^2 for different values of imputed points between observations (m) for the scalar diffusion toy example.	35
2.2	Graphical representation of the transformations of X to U and Z for scalar diffusions.	41
2.3	Graphical representation of the transformations of X to U and Z bottom in the case of changepoints in the volatility.	49
2.4	Kernel densities of the posterior draws of all the parameters for different values of imputed points $m = 20, 40, 50$	54
2.5	Autocorrelation plots for the posterior draws of all the parameters for different values of imputed points $m = 20, 40, 50$	55
2.6	Kernel densities of the posterior draws of all the parameters for different values of imputed points $m = 20, 40, 50$	57
2.7	Kernel densities of the posterior draws of all the parameters for different values of imputed points $m = 20, 40, 50$. Spanish-European rates dataset.	59
2.8	Kernel densities of the posterior draws of all the parameters for different values of imputed points $m = 20, 40, 50$. Spanish-European rates dataset.	60
3.1	Autocorrelation plots of posterior draws of σ^2 for different values of imputed points between observations (m) for the simple stochastic volatility model. The draws in (a) correspond to the reparametrised scheme and in (b) to the scheme without transformation.	77
3.2	Autocorrelation plots of posterior draws of σ for different values of imputed points between observations (m) for the Heston model.	79

3.3	Posterior densities of (a) log-likelihood and (b) σ for different values of imputed points between observations (m) for the Heston model.	80
3.4	Standard & Poor 500 values (up) and its volatility index (down).	89
3.5	Kernel density plots of the log-likelihood from the Heston (top), GARCH (middle) and log OU (bottom) models for the SP500/VIX data.	91
3.6	Autocorrelation plots of the posterior draws of σ (left) and ρ (right) of the Heston (top), GARCH (middle) and log OU (bottom) models for the SP500/VIX data.	93
4.1	Plots of a sample path for X , U and Z against their corresponding times for $\sigma = \sqrt{2}$ and $m = 7$. Z equals 0 at time $+\infty$	106
4.2	Updates of time scale parameters: For every proposed value of them, new points are required and should obtained conditional on the stored points.	109
4.3	Autocorrelation plots for the posterior draws of ρ and σ for different numbers of imputed points $m = 30, 50$. Simulation example of Chapter 3. . .	116
4.4	Kernel densities of the posterior draws of all the parameters for different numbers of imputed points $m = 30, 50$. Simulation example of Chapter 3.	117
4.5	Weekly 3–month US Treasury bill rate from the 5th of January 1962 up to the 30th of August 1996.	118
4.6	Kernel densities of the posterior draws of all the parameters and the log-likelihood for different values of imputed points $m = 10, 20$. Example on Weekly 3–month US Treasury bill rates.	121
4.7	Autocorrelation plots for the posterior draws of the model parameters for different numbers of imputed points $m = 10, 20$ for the analysis of Weekly 3–month US Treasury bill rates.	122

Chapter 1

Introduction

1.1 Diffusion models

The basis of a diffusion process is the celebrated Brownian motion. More than a century ago, Albert Einstein published his first paper on Brownian motion (Einstein, 1905). Nevertheless, he was not the first one to touch on this subject. In around 1827, the Scottish scientist Robert Brown observed the random behavior of pollen particles suspended in water. Naturally, this phenomenon came to be known after his name. Several years later (yet before Einstein) and in a completely different context, a young French mathematician, Louis Bachelier, defended a thesis at the Sorbonne in Paris, France, on a probabilistic model of the French bourse (Bachelier, 1900). Although his aim was to give a statistical description of financial transactions on the Paris stock market and he was obviously unaware of Brown's work, he ended up using the same process. The phrase with which he ended his thesis was 'the bourse, without knowing it, follows the laws of probability' is still a guiding principle in modern quantitative finance. None of these famous scientists however, did succeed to show that the model existed as a rigorous stochastic process; this was done later by Wiener (1923) who used ideas of Borel and Lebesgue from measure theory. As a result of that, Brownian motion is also now called Wiener process.

A diffusion process can be thought of as continuous time model for a continuous process, the noise of which is determined by Brownian motion. Such a model is formulated under a differential equation and therefore the construction of a 'stochastic integral' was

needed. This was invented several years later by the Japanese mathematician Ito. From 1942 to 1946, K. Ito published a sequence of papers (Stroock and Varadhan, 1987) defining a diffusion process X through a ‘stochastic differential equation’:

$$dX_t = \mu(t, X_t) + \sigma(t, X_t)dB_t.$$

Here $\mu(t, X_t)$ is the instantaneous mean increment, $\sigma(t, X_t)\sigma(t, X_t)'$ is its variance and B is a Brownian motion (Wiener process).

The intuitive definition of a diffusion process makes it appealing in many types of applications across a wide range of fields. In Finance, despite the fact that Bachelier’s work was forgotten for half a century, it was rediscovered in the 60’s (in part independently) when the statistician L. J. Savage showed it to the (eventual) Nobel Prize-winning economist Paul Samuelson. Samuelson modified it and proposed geometric Brownian motion as a model for stock prices. Interestingly enough, the role of probability in finance has since then only increased. Quite sophisticated tools of stochastic differential calculus are routinely used in pricing, hedging and risk assessment of financial products. However, diffusion processes find applications in other fields. Some examples include Biology (McAdams and Arkin, 1997), Genetics (Kimura and Ohta, 1971), Chemistry (Gillespie, 1976), Physics (Obuhov, 1959) etc.

The development of statistical inference for diffusion processes is thus of high practical importance. Moreover it is quite a challenging task from a mathematical perspective, since we can only finitely observe these infinite dimensional objects. In this thesis, we pursue likelihood based inference, which remains an open problem as we elaborate in this chapter. We adopt a Bayesian framework, utilizing Markov chain Monte Carlo (MCMC) techniques.

1.2 Preliminaries

In this section we provide some definitions that are essential for the upcoming material. Undoubtedly, this is an incomplete introduction geared towards the needs of this thesis. For a thorough and complete introduction with all the relevant details, the reader is referred to Oksendal (2000), Rogers and Williams (1994), Karatzas and Shreve (1991) and Kloeden and Platen (1995).

1.2.1 Brownian Motion

Definition

Consider the following general continuous time model for the continuous stochastic process X_t :

$$\frac{dX_t}{dt} = \mu(t, X_t) + \sigma(t, X_t) \times \text{'noise'}$$

The noise may be represented by another stochastic process W_t :

$$\frac{dX_t}{dt} = \mu(t, X_t) + \sigma(t, X_t) \times W_t$$

Some desirable properties for the noise process W_t are:

1. If $t_1 \neq t_2$, then W_{t_1}, W_{t_2} independent.
2. $\{W_t\}$ is stationary.
3. $E(W_t) = 0$

However, no stochastic process with continuous paths satisfies the first two properties. We thus set $t_0 < t_1 < \dots < t_k$ and consider a discrete time version of our model. Let

$$\begin{aligned}
X_{t_{k+1}} - X_{t_k} &= \mu(t, X_t)(t_{k+1} - t_k) + \sigma(t, X_t)W_t(t_{k+1} - t_k), \text{ or} \\
X_{t_{k+1}} - X_{t_k} &= \mu(t, X_t)(t_{k+1} - t_k) + \sigma(t, X_t)(B_{t_{k+1}} - B_{t_k})
\end{aligned}$$

It turns out that the only process that has zero mean, stationary and independent increments is Brownian motion (Knight, 1981).

A random variable B_t is called standard Brownian motion if it satisfies the following properties:

1. $B_0 = 0$
2. B_t is a continuous function of t .
3. B has independent, normally distributed increments: Let $Y_1 = B_{t_1} - B_{t_0}$, $Y_2 = B_{t_2} - B_{t_1}$, \dots , $Y_k = B_{t_k} - B_{t_{k-1}}$. Then
 - Y_1, \dots, Y_k are independent
 - $Y_k \sim N(0, t_k - t_{k-1})$

Brownian bridge

Let $\{B_t, t \geq 0\}$ denote a Brownian motion ($B_0 = 0$). Loosely speaking a Brownian bridge (between times 0 and 1) from 0 to a fixed real number b is the Brownian motion B conditioned to be b at time 1. Note that the conditioning on a zero probability event is not straightforward. Under a rigorous definition, a Brownian bridge is a continuous time Gaussian process $\{X_t, 0 \leq t \leq 1\}$ such that:

$$E(X_t) = bt, \quad \text{cov}(X_t, X_s) = \min(s, t) - st, \quad s, t \in [0, 1]$$

Alternatively we may define a Brownian bridge through either of the two following equations:

$$\begin{aligned} X_t &= B_t + t(b - B_1), \quad 0 \leq t \leq 1, \\ dX_t &= \frac{b - X_t}{1 - t} dt + dB_t, \quad 0 \leq t < 1, \quad X_1 = b. \end{aligned}$$

Note that the transformation from X to B is not 1 – 1. This is only true for another equivalent definition which we give in the Section 1.2.2.

1.2.2 Diffusion processes

Stochastic Differential Equations

A multidimensional diffusion is defined as the solution to the following vector stochastic differential equation (SDE):

$$dX_t = \mu(t, X_t, \theta)dt + \sigma(t, X_t, \theta)dB_t, \quad 0 \leq t \leq T, \quad (1.1)$$

where $B_t = (B_t^{\{1\}}, \dots, B_t^{\{d\}})'$ denotes a d -dimensional Brownian motion. The functions $\mu : [0, +\infty) \times \mathbb{S}_X \times \Theta \rightarrow \mathfrak{R}^d$ with $[\mu(\cdot)]_i = \mu^{\{i\}}(\cdot)$ and $\sigma(\cdot) : [0, +\infty) \times \mathbb{S}_X \times \Theta \rightarrow \mathfrak{R}^{d \times d}$ with $[\sigma(\cdot)]_{\{ij\}} = \sigma^{\{ij\}}(\cdot)$ correspond to the SDE's drift and dispersion matrix respectively (\mathbb{S}_X denotes the domain of the diffusion X). In coordinate form the above SDE writes:

$$dX_t^{\{i\}} = \mu^{\{i\}}(t, X_t, \theta)dt + \sum_j \sigma^{\{ij\}}(t, X_t, \theta)dB_t^{\{j\}}, \quad i = 1, \dots, d, \quad j = 1, \dots, d.$$

A less rigorous, yet more intuitive and easier to interpret, definition of X_t is the following: Suppose that X_t is a continuous Markov process and assign the following probability model to it:

$$X_{t+\delta} = X_t + \delta\mu(t, X_t, \theta) + \sigma(t, X_t, \theta)\epsilon,$$

$$\epsilon \sim N(0, \delta I_d),$$

where (I_d is the d -dimensional identity matrix and $'$ denotes the matrix transpose). If we take the limit as $\delta \rightarrow 0$ we get the same process of (1.1). Hence we can interpret $\mu(\cdot)$ (drift) and $\Sigma(\cdot) = \sigma(\cdot)\sigma(\cdot)'$ (volatility) as the mean and covariance matrix of the instantaneous increments of X_t .

For the purposes of this thesis, we require the existence of a weakly unique (in law) solution. This is ensured if (but not only if) the following assumptions hold (Rogers and Williams, 1994, vol 2; page 170):

1. $\Sigma(\cdot) := \sigma(\cdot)\sigma(\cdot)'$ is continuous
2. $\Sigma(x)$ is positive definite for each x (In other words $|\sigma(x)| \neq 0$ for each x).
3. There exists $K > 0$ such that $|\Sigma]_{ij}(x)| \leq K(1 + |x|^2)$ and $|\mu^{\{i\}}(x)| \leq K(1 + |x|)$ for all i, j and x .

Diffusion transformations - Ito's lemma

Let X be a d -dimensional diffusion as in Section 1.2.2:

$$\begin{aligned} dX_t &= \mu(t, X_t, \theta)dt + \sigma(t, X_t, \theta)dB_t, \text{ or, in coordinate form,} \\ dX_t^{\{i\}} &= \mu^{\{i\}}(t, X_t, \theta)dt + \sum_j \sigma^{\{ij\}}(t, X_t, \theta)dB_t^{\{j\}} \end{aligned}$$

Suppose that we are interested in a transformation $U_t = h(t, X_t, \theta) = (h_1(\cdot), \dots, h_p(\cdot))'$, where $h(\cdot)$ is a C^2 map from \mathbb{R}^d into \mathbb{R}^p . Using Ito's lemma we can get the following formula for the SDE of U :

$$dU_t^{\{k\}} = \frac{\partial h_k(t, X_t, \theta)}{\partial t} dt + \sum_{i=1}^d \frac{\partial h_k(t, X_t, \theta)}{\partial X_t^{\{i\}}} dX_t^{\{i\}} + \sum_{i,j=1}^d \frac{\partial^2 h_k(t, X_t, \theta)}{\partial X_t^{\{i\}} \partial X_t^{\{j\}}} dX_t^{\{i\}} dX_t^{\{j\}},$$

where $dB_t^{\{i\}} dB_t^{\{j\}} = \delta_{ij} dt$, $dB_t^{\{i\}} dt = dt dB_t^{\{i\}} = 0$ and $k = 1, \dots, p$.

Time change of a diffusion

As before let X be a diffusion satisfying the following general SDE:

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

Now consider a positive function $c(t, \theta, X_t)$ to \mathbb{R} and define:

$$\eta(t) = \int_0^t c(s, \theta, X_s) ds$$

We say that $\eta(t)$ is a (random) time change process with time change rate $c(t, \theta, X_t)$. Certainly $\eta(\cdot)$ is an invertible function. The diffusion X on the changed time, denote by $U_t = X_{\eta^{-1}(t, \theta, X_t)}$ has the following SDE:

$$dU_t = \frac{\mu(U_t, \theta)}{c(t, \theta, X_t)} dt + \frac{\sigma(Y_t, \theta)}{\sqrt{c(t, \theta, X_t)}} dB_t, \quad 0 \leq t \leq \eta(T).$$

Example: We may define a Brownian bridge as a 1 – 1 transformation of a Brownian motion using time change. More specifically a Brownian bridge X_t , from 0 to 0 and between times 0 and 1 can be defined as:

$$X_t = (1 - t)B_{\frac{t}{1-t}}$$

where B_t is a standard Brownian motion. This definition is equivalent with those of Section 1.2.1.

Quadratic variation

The quadratic covariation between two diffusions $[X_i(t), X_j(t)]$ is given by the limit (in probability) of cross-products of increments of the processes $X_i(t), X_j(t)$:

$$[X_i(t), X_j(t)] = \lim_{n \rightarrow \infty} \sum_{k=0}^{n-1} [X_i(t_{k+1}^n) - X_i(t_k^n)][X_j(t_{k+1}^n) - X_j(t_k^n)], \text{ a.s.}$$

where the partition $\{t_{k+1}^n\}$ is getting finer as $n \rightarrow \infty$. Hence the diffusion (volatility) matrix $\sigma(\cdot)$ can be determined from the quadratic covariation process Q of the path X_t . More specifically we have:

$$\lim_{n \rightarrow \infty} \sum_{k=0}^{n-1} (X_{t_{k+1}^n} - X_{t_k^n})(X_{t_{k+1}^n} - X_{t_k^n})' = \int_0^T \sigma(X_t, \theta)\sigma(X_t, \theta)' dt, \text{ a.s.} \quad (1.2)$$

1.2.3 Bayesian Inference using Markov chain Monte Carlo

Similarly to frequentist statistical inference, Bayesian inference requires a sampling model that produces the likelihood. The likelihood is provided by the probability model for the data and contains certain parameters which represent the unknown quantities of the problem. The main conceptual difference between frequentist and Bayesian frameworks regards the model parameters. In the Bayesian context these are described with the (potentially multi-dimensional) random variable θ and are indistinguishable from the data Y . Additionally, the Bayesian approach will place a *prior* distribution on the model parameters. The likelihood and the prior are then combined using Bayes' theorem to compute the *posterior* distribution. The posterior distribution is the conditional distribution of the unknown quantities given the observed data and is the object on which all Bayesian inference is based. The frequentist and Bayesian approaches, despite arising

from different principles do not necessarily give completely dissimilar answers. In fact, they can be connected in a decision-theoretic framework through *preposterior* evaluations (Rubin, 1984). In this thesis we will adopt the Bayesian paradigm which, while theoretically simple and more intuitive than the frequentist approach, requires evaluation of complex integrals even in fairly elementary problems.

The use of Bayesian methods in applied problems has exploded during the 1990s. The availability of fast computing machines was combined with a group of iterative simulation methods known as Markov chain Monte Carlo (MCMC) algorithms that greatly aided the use of realistically complex Bayesian models. The idea behind MCMC is to produce approximate samples from the posterior distribution of interest, by generating a Markov chain which has the posterior as its limiting distribution. This revolutionary approach to Monte Carlo was originated in the particle Physics literature in Metropolis et al. (1953). It was then generalised by Hastings (1970) to a more statistical setting. However, it was Gelfand and Smith (1990) that introduced MCMC methods to mainstream statistics and since then, the use of Bayesian methods for applied statistical modelling has increased rapidly. Relevant reviews can be found in Smith and Roberts (1993), Brooks (1998) and Dellaportas and Roberts (2003), whereas comprehensive accounts of MCMC-related issues are provided in Gilks et al. (1996). In an introductory technical level, Congdon (2001) describes the analysis of a wide range of statistical models using BUGS, freely available software for Bayesian Inference using MCMC, see Spiegelhalter et al. (1996). Many of these models, including generalised linear mixed models, can only be approximately analysed using classical statistical methodology. Conversely, it is straightforward to analyse models of this complexity using routine examples of BUGS.

1.3 Statistical inference for Diffusion processes

In this thesis we address the problem of performing statistical inference for the parameters of the diffusion processes. We proceed making the assumption that both the drift and the volatility functions have known functional forms. In other words we focus on parametric inference. Non parametric techniques are also available. For example Aït-Sahalia (1996a),

Florens-Zmirou (1993), Jacod (2000) use kernel based methods whereas Genon-Catalot et al. (1992), and Hoffmann (1999) use wavelets.

In theory diffusion process have continuous time paths, but in practice they can only be observed at finite number of times $0 = t_0 < t_1 < \dots < t_n = T$. We reserve $Y = \{Y_k = X_{t_k}, k = 0, 1, \dots, n\}$ to denote the relevant observations of the diffusion. We adopt a Bayesian framework, treating the parameters as random variables and expressing our uncertainty about them, before and after the observation of the diffusion, through suitable probability distributions. Our objective is the posterior distribution of the parameters which is obtained by combining the prior and the likelihood through Bayes theorem. As the posterior is almost always intractable, we aim in drawing samples from it utilizing MCMC techniques. In this section we provide some general information regarding the likelihood for diffusions and how it can be used for inference purposes in cases of observations with high frequency (sufficiently ‘fine’ data).

1.3.1 Likelihood for Diffusions

Dominating measure

Consider the probability of the observations (which are assumed to be random variables):

$$P_\theta(Y \in A) = \int_A L(Y, \theta) dx.$$

In the expression above $L(Y, \theta)$ provides the likelihood for the parameter vector θ . The integration is done with respect to the Lebesgue measure which in this in case is the likelihood’s dominating or reference measure. Lebesgue is not the only option for a reference measure. In fact, for any positive function $h(x)$ we can define the measure H by:

$$H(A) = \int_A h(x) dx,$$

and use this instead:

$$P_\theta(X \in A) = \int_A \frac{L(x, \theta)}{h(x)} h(x) dx = \int_A \frac{L(x, \theta)}{h(x)} dH.$$

The new likelihood function $L^*(x, \theta)$ will now be:

$$L^*(x, \theta) = \frac{L(x, \theta)}{h(x)}.$$

Note that both $L^*(x, \theta)$ and $L(x, \theta)$ are equivalent for estimation purposes. For example, they are maximized at the same value of θ .

Denote the probability measure of the random variable Y by \mathbb{P}_θ . An alternative way to see the likelihood L^* is as the Radon-Nikodym derivative between the measures \mathbb{P}_θ and H . We write:

$$\frac{d\mathbb{P}_\theta}{dH} = L^*(x, \theta), \quad \frac{d\mathbb{P}_\theta}{dLeb(x)} = L(x, \theta).$$

It is crucial that the dominating measure is independent of θ . It is this assumption that enables us to use $L^*(x, \theta)$ in the same way as $L(x, \theta)$ to perform inference for the parameter vector θ .

Likelihood using the Girsanov theorem

Consider the following d -dimensional diffusion

$$dX_t = \mu(X_t, \theta)dt + dB_t, \quad 0 \leq t \leq T \leq \infty.$$

Denote by \mathbb{P} the probability measure of X and with \mathbb{W} that of the driftless version of X , meaning just Brownian motion. Assume that X has a unique weak solution and that it satisfies the Novikov condition

$$E_{\mathbb{P}} \left[\exp \left(\frac{1}{2} \int_0^T \mu(X_t, \theta)' \mu(X_t, \theta) dt \right) \right] < \infty.$$

Then, according to the Girsanov theorem, \mathbb{P} is absolutely continuous with respect to \mathbb{W} with Radon-Nikodym derivative provided by the Cameron Martin Girsanov formula:

$$\frac{d\mathbb{P}}{d\mathbb{W}} = G(X_t, \theta) = \exp \left\{ \int_0^T \mu(X_t, \theta)' dX_t - \frac{1}{2} \int_0^T \mu(X_t, \theta)' \mu(X_t, \theta) dt \right\}$$

The quantity $G(X_t, \theta)$ may be seen as a likelihood for θ since its reference measure (Wiener measure) does not depend on any parameters. However, it contains Ito and diffusion path integrals which generally do not have an analytic solution.

Likelihood using transition density

Alternatively we may use the marginal density of the, finite dimensional, observations Y . Diffusions satisfy the Markov property, therefore we can express the likelihood as the product:

$$L(Y, \theta) = p(Y_0) \prod_{i=1}^n p(Y_i | Y_{i-1}; \theta).$$

But as before, the transition density is generally not available in closed form except for some specific cases like diffusion with linear drift and constant volatility (Ornstein-Uhlenbeck process), Geometric Brownian motion etc.

1.3.2 Inference assuming ‘thin’ data

We now return to the general version of the d - dimensional diffusion X :

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

Throughout this thesis we suppress the notation and write θ in both drift and volatility, although the parameters in these functions are different. The problem of likelihood based inference is simplified substantially under the assumption that the observation times $\{t_k, k = 0, 1, \dots, n\}$ provide a ‘sufficiently fine’ partition of the diffusion path of X . Girsanov theorem now provides the Radon-Nikodym derivative with respect to the law of the local martingale:

$$dM_t = \sigma(X_t, \theta)dB_t, \quad 0 \leq t \leq T.$$

More specifically we have the following quantity for the likelihood:

$$\frac{d\mathbb{P}}{d\mathbb{W}} = G(X_t, \theta) = \exp \left\{ \int_0^T [\Sigma^{-1}\mu(X_t, \theta)]' dX_t - \frac{1}{2} \int_0^T \mu(X_t, \theta)' \Sigma^{-1} \mu(X_t, \theta) dt \right\} \quad (1.3)$$

where $\Sigma = \sigma(X_t, \theta)\sigma(X_t, \theta)'$. The sufficiently fine observed path of X provides a solution to two problems. First, the reference measure of the likelihood depends on the volatility parameters. But these may be determined (rather than estimated) from the quadratic covariation process. For example consider a scalar diffusion X with constant volatility σ . Then using (1.2) we get the following relation with the quadratic variation process Q_X

$$Q_X = \sum_{k=0}^{n-1} (X_{t_{k+1}} - X_{t_k})^2 \approx \int_0^T \sigma^2 dt = T\sigma^2,$$

which gives us σ . Given the volatility parameters, the likelihood from (1.3) is well defined and we can use it to perform inference for the drift parameters. The second problem that we face is the evaluation of the Ito and path integrals. Again, the sufficiently fine observed path Y allows for accurate numerical evaluations:

$$G(X_t, \theta) \approx \exp \left\{ \sum_{k=0}^{n-1} [\Sigma_{\theta}^{-1}(Y_k)\mu_{\theta}(Y_k)]' \Delta Y_k - \frac{1}{2} \sum_{k=0}^{n-1} \mu_{\theta}(Y_k)' \Sigma_{\theta}^{-1}(Y_k)\mu_{\theta}(Y_k)\Delta t_k \right\}, \quad (1.4)$$

where $\Delta Y_k = Y_{k+1} - Y_k$ and $\Delta t_k = t_{k+1} - t_k$.

Alternatively, we may approximate the transition density using the *Euler-Maruyama* (Euler) approximation. Under this scheme we have

$$Y_{k+1}|Y_k \sim N \{Y_k + \mu_\theta(Y_k)\Delta t_k, \Sigma_\theta(Y_k)\Delta t_k\} \quad (1.5)$$

The Euler approximation essentially assumes constant drift and volatility function over each time interval between t_k and t_{k+1} , which is the same assumption made in the approximation of (1.4). See Kloeden and Platen (1995) for more details on its convergence properties and similar approximating schemes. In fact, one can check that the likelihood from (1.4) equals that of (1.5) divided by an euler scheme under the driftless version M . For more details on likelihood based inference for diffusions with sufficiently fine observations see Prakasa Rao (1999) and Polson and Roberts (1994) for a Bayesian approach.

But what if the observations are not so close to each other? The quadratic variation estimates are then biased and the likelihood approximations may become unacceptably poor. Moreover, how can we tell if this assumption is a realistic one? As we saw in Section 1.2.2 the notion of a diffusion's time is strongly related with its volatility parameters which we want to estimate. In the remainder of this thesis we no longer use this assumption and we deal with the problem in its absence.

1.4 Review on likelihood based inference for diffusions

As we already mentioned, the likelihood of a diffusion process is generally not available. This has stimulated the development of various interesting inference techniques for their parameters. Here we review the main likelihood-based methods, alternative approaches use indirect inference (Gouriéroux et al., 1993), estimating functions (Bibby and Sorensen, 1995) and the efficient method of moments (Gallant and Tauchen, 1996); see also Gallant

and Long (1997). We classify the likelihood based methods into 3 categories according to whether they use i) approximation of the transition density (Section 1.4.1), ii) Data augmentation through Bayesian MCMC techniques (Section 1.4.2) or iii) Sequential - online framework (Section 1.4.4). We focus on the ideas behind these techniques, rather than the technical details. For illustration purposes we refer to scalar diffusions of the general form

$$dX_t = \mu(X_t, \theta)dt + \sigma(X_t, \theta)dW_t,$$

and provide some relevant discussion about generalizations to the multivariate case. As always we assume a finite number of observations at times $\{t_k, k = 0, 1, \dots, n\}$ denoted by $Y = \{Y_k = X_{t_k}, k = 0, 1, \dots, n\}$. Multidimensional diffusions with unobserved paths are also termed partially observed diffusions. The following transformation to unit volatility is crucial in almost all cases:

$$U_t = h(X_t, \theta), \quad h(u, u_0, \theta) = \int_{u_0}^u \frac{1}{\sigma(z, \theta)} dz, \quad u, u_0 \in \mathbb{R} \quad (1.6)$$

Indeed, we get from Ito's lemma that U solves the SDE:

$$dU_t = \left\{ \frac{\mu(h^{-1}(U_t, \theta), \theta)}{\sigma(h^{-1}(U_t, \theta), \theta)} - \frac{1}{2} \frac{\partial \sigma(h^{-1}(U_t, \theta), \theta)}{\partial h^{-1}(U_t, \theta)} \right\} dt + dW_t \quad (1.7)$$

where $h^{-1}(\cdot)$ denotes the inverse of $h(\cdot)$.

1.4.1 Inference using approximations of the transition density

The methodologies of this section aim in estimating the likelihood through the product of the transition densities:

$$L(Y, \theta | Y_0) = \prod_{i=1}^n p(Y_i | Y_{i-1}; \theta, \Delta), \quad \Delta = t_i - t_{i-1}$$

Under such an approach, the likelihood is split into the smaller bits and it suffices to focus on $p(Y_k|Y_{k-1}; \theta, \Delta)$. Nevertheless, this complicates multivariate extensions, especially to partially observed diffusions where the observed diffusion component is no longer Markovian and the transition density depends on the entire history of the latent diffusion path. We present three different approximations, one of which provides a closed form analytical expression and the remaining two are based on Monte Carlo schemes.

A closed form analytical approximation

Aït-Sahalia (2002) develops an explicit deterministic estimate of $p(Y_k|Y_{k-1}; \theta, \Delta)$ using a Gram-Charlier series approximation (Kendall and Stuart, 1977, chapter 8), in other words advocating a Hermite series expansion around the standard Gaussian density. To ensure convergence and improve the behavior of the estimate, the diffusion path should be transformed appropriately to obtain a transition density which is closer to the standard Gaussian (the first term of Gram-Charlier series). For this reason Aït-Sahalia (2002) employs two transformations of the diffusion X . First, he aims for a diffusion with unit volatility which can be achieved using (1.6) to get U , and then he sets:

$$Z_t = \frac{h(X_t, \theta) - h(X_{t_{k-1}})}{\sqrt{t_k - t_{k-1}}}, \quad t_{k-1} \leq t \leq t_k, \quad X_t = \nu(Z_t, \theta)$$

The data are not actually transformed, this would not be possible under this framework as the transformation depends on θ , this is just part of the approximation procedure. Using standard change of variables arguments we get that

$$p(X_{t_k}|X_{t_{k-1}}; \theta, \Delta) = \frac{p(Z_{t_k}|Z_{t_{k-1}} = 0; \theta, \Delta)}{\sigma(\nu(Z_{t_k}, \theta), \theta)\sqrt{\Delta}} \quad (1.8)$$

Now consider the truncated after J terms Hermite series expansion $p^J(Z_{t_k}|Z_{t_{k-1}} = 0; \theta, \Delta)$ of $p(Z_{t_k}|Z_{t_{k-1}} = 0; \theta, \Delta)$ around the standard Gaussian density. One can plug it into (1.8) and obtain the following estimate of $p(X_{t_k}|X_{t_{k-1}}; \theta, \Delta)$:

$$p^J(X_{t_k}|X_{t_{k-1}}; \theta, \Delta) = \frac{p^J(Z_{t_k}|Z_{t_{k-1}} = 0; \theta, \Delta)}{\sigma(\nu(Z_{t_k}, \theta), \theta)\sqrt{\Delta}}$$

The Hermite series coefficients cannot be computed but they can be approximated analytically in terms of the diffusion infinitesimal generator (reference). This results into very complex expressions for the likelihood, yet explicit and with appealing properties. Aït-Sahalia (2002) proves that $p^J(\cdot)$ converges to the true density $p(\cdot)$ uniformly in Y_k and θ . Furthermore as J tends to infinity the resulting estimator is asymptotically equivalent to the Maximum Likelihood estimator (MLE) (Ait-Sahalia (2002), theorems 1 and 2). Also, his numerical experiments, reveal very good approximations even with $J = 1$ or 2.

Aït-Sahalia (2005) provides an extension for multidimensional diffusions. In this case the transformation $h(\cdot)$ (to a diffusion with the identity matrix of dimension d) does not always exist. When it does, one can proceed in a similar manner. If such a transformation is not available, an expansion in powers of $t_k - t_{k-1}$ may be used. Its coefficients may be determined using Kolmogorov's backward and forward equations and in terms of a Taylor expansion in $(Y_k - Y_{k-1})$. This may require a considerable amount effort to limit the approximation error in cases of very volatile models or sparse datasets.

Extensions to multidimensional diffusions with unobserved paths are not straightforward due to the lack of Markov property for the observed process. Regarding financial applications, Aït-Sahalia and Kimmel (2005) extend the methodology to stochastic volatility models where the volatility of an asset price is identified through additional data resources, namely option prices on this asset. see Section 3.7.3 for more.

Simulated likelihood

Pedersen (1995) and Santa Clara (1995), see also Brandt and Santa Clara (2001), introduced a Monte Carlo estimate of $p(Y_k|Y_{k-1}; \theta, \Delta)$. The main idea is to introduce a finite number of imputed points X_1, \dots, X_m , at times $\{t_j = t_{k-1} + \frac{j\Delta}{m+1}, j = 1, \dots, m\}$ (with $t_j - t_{j-1} = \Delta/(m+1) = \delta$), so that the resulting partition is sufficiently fine for the Euler approximation to be satisfactory. Then we can write (with $X_0 = Y_{k-1}$)

$$\begin{aligned}
p(Y_k|Y_{k-1}; \theta, \Delta) &= \int_{\mathbb{R}^m} \left\{ \prod_{j=1}^m p(X_j|X_{j-1} = x_{j-1}; \theta, \delta) \right\} p(Y_k|X_m = x_m; \theta, \delta) dx_1 \dots dx_m \\
&= \mathbb{E} [p(Y_k|X_m = x_m; \theta, \delta)]
\end{aligned}$$

If we attempt to approximate the densities over the shorter intervals of length δ using Euler, denote by $p^E(Y_k|X_m = x_m; \theta, \delta)$, we get the following estimate:

$$p^m(Y_k|Y_{k-1}; \theta, \Delta) = \mathbb{E} [p^E(Y_k|X_m = x_m; \theta, \delta)]$$

The expectation may be computed using Monte Carlo integration over independent realizations of X_m . Pedersen (1995) shows convergence in probability to the true likelihood for all θ under some boundedness conditions. To ensure the smoothness of the mapping $\theta \rightarrow p^m(Y_k|Y_{k-1}; \theta, \Delta)$, the same random elements, e.g. gaussian seeds, should be used for all θ . This, together with some further conditions, lead to weak convergence to the MLE.

This initial formulation to the problem has proven to perform poorly in practice: The Monte Carlo error of the estimate is increasing in the number of imputed points m . Durham and Gallant (2002) provide a substantial refinement using importance sampling. More specifically they estimate the likelihood through ($U_0 = Y_{k-1}$):

$$p^m(Y_k|Y_{k-1}; \theta, \Delta) = \mathbb{E} \left[\frac{\left\{ \prod_{j=1}^m p^E(U_j|U_{j-1}; \theta, \delta) \right\} p^E(Y_k|U_m; \theta, \delta)}{q(U_1, \dots, U_m)} \right],$$

where U_1, \dots, U_m are sample from some appropriate distribution q . Their choice of densities q , in contrast with the initial formulation, do take into account the ending point Y_k (diffusion bridges densities) and thus provides a vast improvement in limiting the Monte Carlo error. Under this importance sampling scheme the variance does not increase with m . Moreover, Durham and Gallant (2002) propose approximation schemes, more sophisticated than the Euler, to reduce the error due to time discretisation.

Equipped with the enhancements provided by Durham and Gallant (2002), this framework provides an alternative numerical option for likelihood based inference. Extension to multidimensional diffusions are straightforward but they come at an increasing computational cost, especially for high dimensional parameter vectors θ . Durham and Gallant (2002) also consider cases of partially observed diffusions using additional sequential Monte Carlo techniques, see Section 1.4.4 for more.

The Exact Algorithm: Eliminating the error due to time discretisation

Beskos et al. (2006b) introduced a novel Monte Carlo mechanism that achieves exact inference in the sense that the estimates of the transition density contain only Monte Carlo error. As in Aït-Sahalia (2002), the transformation (1.6) to a unit volatility diffusion U is necessary. The change of variables yields the following expression of the transition density

$$p(X_{t_k}|X_{t_{k-1}}; \theta, \Delta) = \frac{p(U_{t_k}|U_{t_{k-1}} = 0; \theta, \Delta)}{\sigma(h^{-1}(U_{t_k}, \theta), \theta)},$$

and $p(U_{t_k}|U_{t_{k-1}} = 0; \theta, \Delta)$ becomes the estimation target. Beskos et al. (2006b) offer 3 different estimation techniques: the bridge method (Beskos et al., 2005), the acceptance method and the Poisson estimator. The latter two are based on the following lemma, also found in Dacunha-Castelle and Florens-Zmirou (1986):

$$p(U_{t_k}|U_{t_{k-1}}; \theta, \Delta) = \mathbb{E} \left[\exp \left\{ \mathcal{A}(U_{t_k}, \theta) - \mathcal{A}(U_{t_{k-1}}, \theta) - \int \frac{1}{2} \mathcal{B}(U_s, \theta) ds \right\} \right] \times \mathcal{N}_\Delta(U_{t_k} - U_{t_{k-1}}) \quad (1.10)$$

where $(\mu_U(\cdot))$ denotes the drift of U as given by (1.7))

$$\begin{aligned} \mathcal{A}(x, \theta) &= \int^x \mu_U(s, \theta) ds, \\ \mathcal{B}(x, \theta) &= \mu_U^2(x, \theta) + \frac{\partial \mu_U(x, \theta)}{\partial x}, \end{aligned}$$

and $\mathcal{N}_\Delta(\cdot)$ denotes a normal distribution with mean 0 and variance Δ . The expectation is taken with respect to a Brownian bridge with endpoints $U_{t_{k-1}}, U_{t_k}$ and is intractable. The acceptance method and the Poisson estimator however, provide unbiased Monte Carlo estimates using the exact simulation framework developed in Beskos and Roberts (2005) and Beskos et al. (2005). As shown in Beskos et al. (2006b) it is also possible to construct Monte Carlo EM (Dempster et al., 1977) or MCMC schemes based on the likelihood estimates.

Apart from eliminating the bias due to the time discretisation of the diffusion path, this methodology has the appealing feature of optimal, and in a sense sufficient, imputation of random elements which results in computationally efficient algorithms. The exact simulation algorithms impose some restrictions, mainly on the functional form of $\mu_U(\cdot)$, which are relaxed with the further developments of Beskos et al. (2006a) and the framework covers almost all scalar diffusions. Extensions to multidimensional diffusions are possible when the transformation (1.6) exists. Papaspiliopoulos et al. (2006) consider partially observed diffusions under a sequential framework but the problem of inference for the parameter vector θ is not addressed.

1.4.2 Bayesian MCMC techniques using data augmentation

Adhering to the Bayesian framework, we first assign a prior on the parameter vector $p(\theta)$. Then, given the observations Y , we are interested in the posterior $p(\theta|Y)$ on which the Bayesian inference is based. This is not straightforward however, as the likelihood is generally intractable. To overcome this issue we may introduce a finite number of latent intermediate points between observations, so that the augmented partition allows for good likelihood approximations as in Section 1.3.2. This can be implemented through a data augmentation scheme Tanner and Wong (1987) where the imputed points, or some transformations of them, are treated as additional parameters. The methodology can be thought of as an adaptation of the simulated likelihood approach of Section 1.4.1 to the Bayesian MCMC context. In what follows, we provide the details for almost all the relevant MCMC schemes with or without transformations on the imputed points. For

ease of illustration we restrict to scalar diffusions and a pair of successive observations $Y_{k-1} = X_{t_{k-1}}$, $Y_k = X_{t_k}$ with $t_k - t_{k-1} = \Delta$, when there is no loss of generality.

Initial data augmentation schemes

As in Section 1.4.1, we can introduce a finite number of imputed points X_1, \dots, X_m between Y_{k-1} and Y_k , at times $\{t_j = t_{k-1} + \frac{j\Delta}{m+1}, j = 1, \dots, m\}$. The likelihood can then be approximated using the Euler-Maruyama scheme (see Section 1.5). Denote the full likelihood approximation by $p^m(Y, X, \theta)$. The number m should be chosen to be large enough do that $p^m(Y, X, \theta)$ is an accurate approximation of the true likelihood. Given a fixed value of m , we may proceed using a Gibbs sampling scheme with elements θ and X_j , $j = 1, \dots, m$. Based on some initial values θ^0 and X_j^0 , the algorithm updates from θ^i, X_j^i to θ^{i+1}, X_j^{i+1} by alternating between the following steps:

1. Update to θ^{i+1} using a Gibbs or a Metropolis - Hastings step on $p(\theta|Y, X)$.
2. For each pair of Y_{k-1}, Y_k , update to X_j^{i+1} for all j , using a Gibbs or a Metropolis - Hastings step on $p(X_j|Y, \theta)$.

The MCMC theory ensures that for $i \rightarrow \infty$, the draws θ^i, X_j^i are dependent samples from the true posterior $p(\theta, X_j|Y)$ and inference may be based on the marginal θ^i trajectory. In practice m may be chosen by repeating the algorithm over increasing number of imputed points, until a convergence on the likelihood and parameter estimates is reached.

This summarizes the framework of the data augmentation schemes of Jones (1999) - see also Jones (2003), Eraker (2001) and Elerian et al. (2001). Each of these methodologies use different strategies to carry out steps 1 and 2. The updates of θ are model-dependent under all frameworks. For the latent points X_j Jones (1999) and Eraker (2001) use a cyclic Metropolis - Hastings algorithm that updates each point conditional on its neighboring ones. To improve the mixing of the chain, Elerian et al. (2001) split the augmented diffusion path into blocks of random size and update each one in turn with a Metropolis Hastings algorithm.

The need for reparametrisation

In their simulation based experiment, Elerian et al. (2001) note an increase in the mixing of the chain as the number of imputed points increases which is confirmed by the theoretical results of Roberts and Stramer (2001). More specifically Roberts and Stramer (2001) note that as $m \rightarrow \infty$, the augmented diffusion path contains an infinite amount of information for the parameters in the volatility $\sigma(X_t, \theta)$ as these are determined exactly from the quadratic variation process (see Section 1.2). It is well known however, see Meng and van Dyk (1997) and Meng and van Dyk (1998), that the convergence of data augmentation schemes is problematic when the Fisher's information matrix in the augmented dataset strongly exceeds that in the observed data. In MCMC terms this translates into increasing autocorrelation in the parameter draws and reducibility for $m \rightarrow \infty$. This is not an MCMC specific issue, the same concerns exist into its deterministic analogue, the EM algorithm (Roberts and Sahu, 1999).

Under a measure theoretic probability framework, Roberts and Stramer (2001) note that the likelihood provided by Girsanov's theorem has reference measure that depend on parameters. To correct for that they use reparametrisation, see for instance Paspiliopoulos et al. (2003), through a 2-step transformation to a diffusion for which the likelihood (as given by Girsanov theorem) is written with respect to a reference measure that does not depend on any parameters. The first step of this transformation is no other than the diffusion U given by (1.6) which is also used in the methodologies of Aït-Sahalia (2002) and Beskos et al. (2006b). The second step is applied to each interval between the transformed observations $h(Y_{k-1}, \theta)$ and $h(Y_k, \theta)$ (at times t_{k-1} and t_k respectively):

$$Z_t = U_t - \frac{(t_k - t)h(Y_{k-1}, \theta)(t_{k-1}) + (t - t_{k-1})h(Y_k, \theta)}{t_k - t_{k-1}}, \quad t_{k-1} < t < t_k$$

The transformed diffusion Z , conditioned on the observations Y_{k-1} and Y_k , is absolutely continuous with respect to a Brownian bridge which equals to 0 at times t_{k-1} and t_k . Note that the likelihood, given by Girsanov formula, contains intractable stochastic and path integrals which can be evaluated numerically using the augmented path.

Based on this reparametrisation, Roberts and Stramer (2001) apply a data augmentation scheme with Z and θ as the elements of the Gibbs sampler. Note that the inference is now based on Z rather than X but if required, the inverse of this transformation may provide samples from the posterior of X . The updates of θ given Z are model dependent and may be implemented with random walk metropolis, or in some cases Gibbs steps. For the updates of Z , Roberts and Stramer (2001) split the path into blocks containing the segments between successive observations and update each one in turn. This step is implemented with an independence sampler proposing Brownian or Ornstein-Uhlenbeck bridges. Despite the fact that they consider only scalar diffusions, the methodology can be extended to multidimensional and partially observed models. In fact this is the research topic of this thesis.

Data augmentation for partially observed diffusions

Partially observed diffusion models are of particular interest as they are being used extensively in almost every application involving diffusions. In finance for example, stochastic volatility models play central role in various applications; see for instance Ghysels et al. (1996) and Shephard (2005). Such models can be seen as 2-dimensional diffusions which obey the dynamics of an SDE like the following:

$$\begin{pmatrix} dX_t \\ d\alpha_t \end{pmatrix} = \begin{pmatrix} \mu_x(X_t, \alpha_t, \theta) \\ \mu_\alpha(\alpha_t, \theta) \end{pmatrix} dt + \begin{pmatrix} \sigma_x(\alpha_t, \theta) & 0 \\ 0 & \sigma_\alpha(\alpha_t, \theta) \end{pmatrix} \begin{pmatrix} dB_t \\ dW_t \end{pmatrix}, \quad (1.11)$$

where B and W are standard Brownian motions, that can potentially be correlated, and the process α_t is not observed at all. Eraker (2001) and Jones (1999) consider such models but the reducibility issue raised by Roberts and Stramer (2001) is not addressed. On the other hand, as proved in Aït-Sahalia (2005), the transformation (1.6) does not exist for stochastic volatility models. Chib et al. (2005) provide a solution for the case where $\mu_x(\cdot) \equiv \mu_x(\alpha_t, \theta)$. They note that the density of $X|\alpha$ is tractable given the path of α and there is no need to impute points of C . The reparametrisation problem is thus reduced to

the diffusion α only. Taking advantage of the fact that α is an unconditioned diffusion, they introduce a novel reparametrisation using the driving Brownian motion of α , W . Their relevant MCMC scheme is based on W , θ , however α is sampled as a mechanism for updating W . This is possible because given the parameters of α and an appropriate time discretisation W is deterministic 1 – 1 function of α .

1.4.3 Sequential techniques for partially observed diffusions

The likelihood for stochastic volatility, and generally partially observed diffusion models, may be approximated in a sequential manner, in other words as we observe the data. In this thesis we only consider inference problems for fixed datasets (off-line inference). However, online techniques may prove useful for off-line problems where likelihood inference is not straightforward, e.g. partially observed diffusions. Given the initial points X_0, α_0 and further i observations of X at times t_1, \dots, t_i , all of them denoted by $\mathcal{F}_i = (Y_0, \dots, Y_i, \alpha_0)$, we want to approximate $p(Y_{i+1}|\mathcal{F}_i)$. If we knew the distribution of $\alpha_i|\mathcal{F}_i$ we could use

$$p(Y_{i+1}|\mathcal{F}_i) = \int p(Y_{i+1}|Y_i, x)dP_{\alpha_i|\mathcal{F}_i}(x)$$

This distribution can be calculated recursively by propagating forward that of $\alpha_{i-1}|\mathcal{F}_{i-1}$ in the following way

$$\begin{aligned} p(\alpha_{i-1}|\mathcal{F}_i) &= \frac{p(Y_i|\mathcal{F}_i, \alpha_{i-1})p(\alpha_{i-1}|\mathcal{F}_{i-1})}{p(Y_i|\mathcal{F}_{i-1})} \\ p(\alpha_i|\mathcal{F}_i) &= \int p(\alpha_i|Y_i, x)dP_{\alpha_{i-1}|\mathcal{F}_i}(x) \end{aligned}$$

This is the main idea of the particle filter, or else sequential Monte Carlo, algorithm which provides online estimates of the likelihood through an appropriate importance sampler; see for instance Pitt and Shephard (1999) and the references therein.

By adapting these ideas to the diffusion framework, Durham and Gallant (2002) extend their framework to cover stochastic volatility models. Their applications reveal no

particular efficiency issue. Nevertheless, it is well known that unless appropriate resampling techniques are applied, the variance of the importance weights can only increase in time, thus leading to degeneracy (Doucet et al., 2001).

On the other hand the use of resampling results in a poor approximation to the sequence of distributions $p_\theta(\alpha_0, \dots, \alpha_i | Y_0, \dots, Y_i)$, the particle filter may only provide a satisfactory approximation to the sequence $p_\theta(\alpha_{i-L+1}, \dots, \alpha_i | Y_0, \dots, Y_i)$, for a fixed lag $L > 0$; see for instance Andrieu et al. (2005) and Del Moral (2004). In practice the parameter space is only explored at the initialization of the algorithm and after a few iterations the marginal likelihood of the parameter is approximated by a single delta Dirac function. To deal with this problem several authors, for instance Gordon et al. (1993) substitute the degenerate likelihood with a Kernel approximation density.

Golightly and Wilkinson (2005) use this idea to construct data augmentation schemes for general diffusion models. They overcome the reducibility issue of Section 1.4.2 by jointly updating the volatility parameters and the diffusion paths through a Metropolis algorithm. In practice this can be implemented only in a sequential manner as the acceptance rate of joint Metropolis updates in a fixed dataset, for time up to T , is decreasing in T . The methodology of Golightly and Wilkinson (2005) is appealing even for off-line problems of inference because of its complete generality. The kernel based techniques however transform the fixed parameter θ into a slowly time varying one, whose dynamics is related to the width of the kernel. Furthermore, the choice of the kernel width and its effect is not always clear.

An alternative online estimation procedure on partially observed diffusions is introduced in Poyadjis (2006). This technique aims in maximum likelihood estimates which are obtained with the use of gradient methods based on $p_\theta(\alpha_{i-L+1}, \dots, \alpha_i | Y_0, \dots, Y_i)$.

1.4.4 Concluding remarks

The problem of inference on diffusion models is quite challenging and has stimulated a massive amount of research. For scalar diffusions in particular, there exist several

options for estimating the likelihood and the parameters of the diffusion given a finite number of observations. The exact algorithm of Beskos et al. (2006b) gives unbiased likelihood estimates whereas the remaining approaches can only reduce the bias due to time discretisation by imputing more intermediate points (Simulated likelihood and data augmentation methods) or using higher order approximations (Aït-Sahalia, 2002). On the other hand the analytical likelihood approximations of Aït-Sahalia (2002) provide the only option without Monte Carlo error.

Contrary to scalar diffusion models, the options for inference on multidimensional cases are limited. Assuming perfect observation of the diffusion, in other words data in all of its coordinates, the exact algorithm cannot be applied when the transformation of (1.6) is not available and the analytical approximations of Aït-Sahalia (2005) require additional Taylor expansions. Things are further complicated in the partially observed case, i.e. stochastic volatility models. The methods of Durham and Gallant (2002) and Beskos et al. (2006b) can only be used in a sequential context which complicates the task of inference, and the Ait-Sahalia framework is only applicable in some cases (Aït-Sahalia and Kimmel, 2005). Data augmentation schemes provide a natural alternative option but they require appropriate reparametrisations, as in Roberts and Stramer (2001), Chib et al. (2005), which are not always available.

1.5 Our contribution

In this thesis we focus on multidimensional diffusions using data augmentation - MCMC schemes. Our work can be seen as an extension to the framework of Roberts and Stramer (2001). In Chapter 2 we deal with cases of discretely observed diffusions (perfect observation) and develop appropriate MCMC techniques for inference purposes. The remaining two chapters focus on partially observed diffusions and in particular stochastic volatility models. We divide the stochastic volatility models into two categories, according to their application field in finance: Chapter 3 deals with models applied to equity prices whereas Chapter 4 examines interest rates. Chapters 2 and 3 use reparametrisations which may be viewed as natural extensions to the transformations of Roberts and Stramer (2001)

unlike the time change techniques of Chapter 4 which provide a quite different framework from both theoretical and MCMC implementation point of view. Conclusions are provided at the end of each chapter.

Chapter 2

Bayesian Inference for Discretely Observed Multidimensional Diffusions

2.1 Introduction - Notation

This chapter addresses the problem of modelling several continuous processes, say d , as a multidimensional diffusion. We denote each one of these processes by $X_t^{\{i\}}$, $i = 1, \dots, d$ and combine them together into $X_t = (X_t^{\{1\}}, \dots, X_t^{\{d\}})'$ so that X_t is a d -dimensional vector for each time t . The multidimensional diffusion is defined as the solution to the following vector stochastic differential equation (SDE) with as in Section 1.2.2.1:

$$dX_t = \mu(t, X_t, \theta)dt + \sigma(t, X_t, \theta)dB_t, \quad 0 \leq t \leq T, \quad (2.1)$$

The drift and volatility functions should satisfy some fairly general conditions to ensure that (1.1) has unique weak solution. See Section 1.2.2.1 for more details.

In practice we can only observe X_t at a finite set of points. In this chapter we will deal with the case where we have observations in all the coordinate processes $X_t^{\{1\}}, \dots, X_t^{\{d\}}$ or, in other words, the multidimensional diffusion is discretely (fully) observed. For il-

illustration purposes we further assume¹ that the diffusion is regularly observed (equal times between observations) and that the entire vector X_t is observed at each time. We denote the times of observations by t_k , $k = 1, \dots, n$, and the data with $Y = \{Y_k = X_{t_k} = (X_{t_k}^{\{1\}}, \dots, X_{t_k}^{\{d\}})', k = 1, \dots, n\}$.

Our goal is to perform Bayesian inference for the drift and volatility parameters θ using the information provided by the observations Y . Given an appropriate prior distribution on θ , we aim in sampling from the posterior utilizing MCMC techniques which are presented and illustrated in the remainder of this chapter.

The chapter is organized as follows: Section 2.2 introduces data augmentation as a natural framework to deal with the problem and Section 2.3 highlights potential reducibility issues which are illustrated through a simple example. In Section 2.4 we introduce appropriate transformations for an irreducible data augmentation scheme, the details of which are presented in Section 2.5. A discussion on some potential extensions is included in Section 2.6, whereas Sections 2.7 and 2.8 contain applications on simulated and real data respectively. Finally Section 2.9 concludes with some relevant discussion.

2.2 Data augmentation

The problem of likelihood based inference for diffusion parameters is particularly challenging. Except for some few cases (for instance diffusions with constant volatility and linear drift), it is impossible to solve the SDE analytically. Consequently the likelihood is not available in closed form, posing (setting) a great barrier to the inference problem.

Given sufficiently fine data, we can override the problem by the following two step procedure: (i) identify the volatility parameters using the quadratic covariation process and (ii) use suitable approximations for the likelihood for the drift parameters like the Euler scheme. However, when the data are not available on a fine enough scale, the estimates of the quadratic covariation process are strongly biased and the approximations

¹These assumptions can easily be relaxed even within the framework of this chapter.

of the likelihood can become unacceptably poor. Actually Florens-Zmirou (1989) shows inconsistency.

A natural way to proceed is via data augmentation, a methodology introduced by Tanner and Wong (1987) and used in the context of diffusions by Roberts and Stramer (2001), Elerian et al. (2001), Eraker (2001), Jones (1999), Jones (2003), Chib et al. (2005), Golightly and Wilkinson (2005) etc. As already mentioned, the main idea is that the likelihood can always be well approximated given the entire path of X or a sufficiently fine partition of it. The unobserved paths of X are treated as missing data and a finite number of points, large enough to make the approximation error arbitrarily small, is imputed. Hence we introduce m latent intermediate points between each pair of successive observation Y_k and Y_{k+1} . To simplify things we assume these points to be equidistant and we denote their distance by $\delta = 1/(m + 1)$. The augmented dataset between these successive points takes the following form:

$$\begin{array}{ccccccc}
 Y_k^{\{1\}} & X_{t_k+\delta}^{\{1\}} & X_{t_k+2\delta}^{\{1\}} & \dots & & & Y_{k+1}^{\{1\}} \\
 Y_k^{\{2\}} & X_{t_k+\delta}^{\{2\}} & X_{t_k+2\delta}^{\{2\}} & \dots & & & Y_{k+1}^{\{2\}} \\
 & \vdots & & & & & \vdots \\
 Y_k^{\{d\}} & X_{t_k+\delta}^{\{d\}} & X_{t_k+2\delta}^{\{d\}} & \dots & & & Y_{k+1}^{\{d\}}
 \end{array}$$

Similarly to the previous implementations of this scheme, the algorithm contains the following steps:

1. Initialize the parameter vector θ and the latent paths in an appropriate way (i.e. drawing from the prior distribution).
2. Update the paths given the current set of parameters θ .
3. Update θ conditional on the augmented path.

As noted in Roberts and Stramer (2001) however, there exists a strong dependence between the imputed sample paths and the volatility coefficients. In fact the algorithm becomes reducible as the number of imputed points increases. Similar issues arise for multidimensional diffusions. We provide the details in the next section.

2.3 Reducibility issues

We first demonstrate the problem using a rather simple example, which is nevertheless sufficient to expose the problem caused by the quadratic variation of the diffusion. Then, in Section 2.3.2, we adopt a more rigorous measure theoretic probability point of view, writing down the likelihood using Girsanov's theorem. We thus see the problem from a different angle, as in Roberts and Stramer (2001), which serves as a search-guide for appropriate reparametrisations that resolve the reducibility issue. These reparametrisations are introduced in Section 2.4.

2.3.1 A toy example

Let's consider the following, 2-dimensional diffusion:

$$dX_t = \sigma dB_t, \quad 0 \leq t \leq 1,$$

where B_t is a 2-dimensional standard Brownian motion. Let $X_0^{\{1\}} = X_0^{\{2\}} = 0$, $X_1^{\{1\}} = y_1$ and $X_1^{\{2\}} = y_2$ and assign a non-informative prior for σ^2

$$p(\sigma^2) \propto (\sigma^2)^{-1}.$$

We are after the posterior distribution of σ^2

$$p(\sigma^2|Y), \quad Y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}.$$

One may notice that under this simple diffusion, the marginal likelihood is available:

$$Y \sim N \left(\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix} \right) \right)$$

and therefore the posterior of σ^2 is the Inverse-Gamma distribution:

$$\sigma^2|Y \sim IG \left(1, \frac{y_1^2 + y_2^2}{2} \right).$$

However suppose now for illustration, that the full likelihood is unavailable and we use the augmentation scheme which imputes m equidistant points $X_\delta, X_{2\delta}, \dots, X_{m\delta}$ ($\delta = 1/(m+1)$) on both dimensions. Assuming that m is large enough to allow for a good approximation (which is in fact the exact transition density) of the likelihood we may use the Euler scheme:

$$X_{k\delta}|X_{(k-1)\delta} \sim N(X_{k-1}, \delta\sigma^2 I_2), \quad k = 1, \dots, m+1,$$

where I_2 is the identity matrix. We can now examine the full conditional distribution for σ^2 given the augmented path of X :

$$p(\sigma^2|X) \propto (\sigma^2)^{-(m+1)-1} \exp \left(\frac{(m+1)(Q_1 + Q_2)}{2\delta\sigma^2} \right),$$

where Q_1, Q_2 are the two scalar quadratic variation processes:

$$Q_1 = \sum_{k=1}^{m+1} \left(X_{k\delta}^{\{1\}} - X_{(k-1)\delta}^{\{1\}} \right)^2$$

$$Q_2 = \sum_{k=1}^{m+1} \left(X_{k\delta}^{\{2\}} - X_{(k-1)\delta}^{\{2\}} \right)^2$$

It is not difficult to recognize the kernel of an Inverse-Gamma distribution as before, but now with the parameters:

$$\sigma^2|X \sim IG\left(m+1, \frac{(m+1)(Q_1+Q_2)}{2}\right)$$

Taking the limit as $m \rightarrow +\infty$ we see that the mean of this $\sigma^2|X$ becomes

$$\lim_{m \rightarrow +\infty} E(\sigma^2|Y, X^{mis}) = \frac{E(Q_1+Q_2)}{2} = \sigma^2$$

as we would expect. Surprisingly enough though, the corresponding limit for the variance goes to 0. This implies that there is infinite amount of information (in the limit) on the imputed paths that force the volatility σ^2 to remain at its current value; the full conditional posterior is just a point mass.

This behavior appears only in the limit though. In practice we observe the following picture: Figure 2.1 displays autocorrelation plots of the posterior draws of σ^2 , obtained from a data augmentation scheme with m imputed points and σ^2 . The experiment was repeated for different values of m (1, 50, 100 and 1000) and the difference in the corresponding autocorrelations is substantial and steadily increasing. This is a major drawback as we would like to be able to impute as many points may be required to practically eliminate the discretisation error.

2.3.2 Measure theoretic probability framework

In this section, we examine the problem from a different point of view. Let us consider the general case of (1.1). Without loss of generality, we can assume that we observe only one point Y (apart from the initial one) as before ($X_0^{\{1\}} = X_0^{\{2\}} = 0$, $X_1^{\{1\}} = y_1$ and $X_1^{\{2\}} = y_2$). Using the Markov property we can get the likelihood for more observations by simply adding more product terms. In other words the path X_t , $1 \leq t \leq 2$ is independent from X_t , $0 \leq t \leq 1$ given X_1 and so on. Denote the probability law of X by \mathbb{P}_θ and that of its driftless version

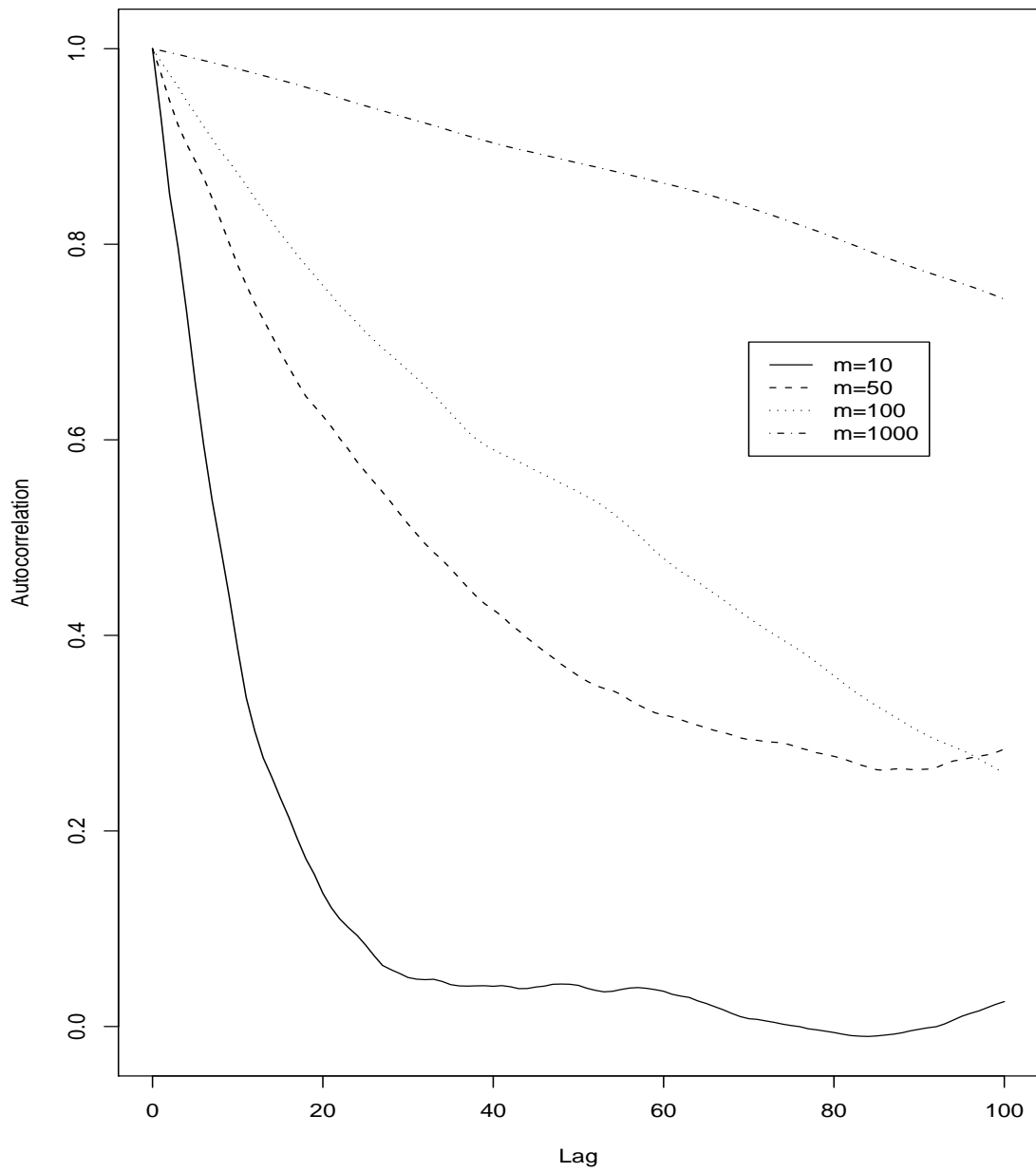


Figure 2.1: Autocorrelation plots of posterior draws of σ^2 for different values of imputed points between observations (m) for the scalar diffusion toy example.

$$dM_t = \sigma(X_t, \theta) dB_t$$

by \mathbb{Q}_θ . Under some fairly general conditions (see introduction), Girsanov's formula provides the Radon-Nikodym derivative of \mathbb{P}_θ with respect to \mathbb{Q}_θ ($\Sigma(\cdot) = \sigma(\cdot)\sigma(\cdot)'$):

$$\frac{d\mathbb{P}_\theta}{d\mathbb{Q}_\theta} = G(X, \mu, \sigma) = \exp \left\{ \int_0^T [\Sigma(X_s, \theta)^{-1} \mu(X_s, \theta)]' dX_s - \frac{1}{2} \int_0^T \mu(X_s, \theta)' \Sigma(X_s, \theta)^{-1} \mu(X_s, \theta) ds \right\}.$$

The integrals in the expression above have no analytic solution in general. Nevertheless, given an augmented path they can be evaluated numerically. The approximation error may become arbitrarily small by simply increasing the number of imputed points m .

Now assume for a moment that under \mathbb{Q}_θ the marginal density of Y with respect to the Lebesgue measure ($Leb_d(Y)$) is known, denote by $f_M(Y; \theta)$. For example if $\sigma(\cdot) = \sigma$ we have:

$$Y \sim N(0, \Sigma)$$

Clearly this is not always the case. However this assumption can be relaxed for the diffusions belonging to the framework developed in Section 2.4, where we provide the relevant details. With $f_M(Y; \theta)$ known we can factorise the dominating measure \mathbb{Q}_θ in the following way

$$\mathbb{Q}_\theta = \mathbb{Q}_\theta^y \otimes Leb_d(Y) \times f_M(Y; \theta).$$

We can now write:

$$\frac{d\mathbb{P}_\theta}{\mathbb{Q}_\theta^y \otimes Leb_d(Y) \times f_M(Y; \theta)}(X^{mis}, Y) = G(X, \mu, \sigma),$$

or else

$$\frac{d\mathbb{P}_\theta}{\mathbb{Q}_\theta^y \otimes Leb_d(Y)}(X^{mis}, Y) = G(X, \mu, \sigma) \times f_M(Y; \theta),$$

The expression above may be regarded as the likelihood for the paths of the diffusion between observations (X^{mis}) and the parameters θ . Nonetheless, a part of the likelihood's reference measure, in particular \mathbb{Q}_θ^y , depends on parameters as it reflects the distribution of a diffusion, conditioned on the event ($X_1 = Y$), with volatility given by $\sigma(X_t, \theta)$. One can also notice that since the volatility parameters are identified by the quadratic covariation process, the measure \mathbb{Q}_θ is just a point mass. Consequently, the measures \mathbb{Q}_θ are mutually singular (having different support) and therefore so are \mathbb{P}_θ . Thus, inference for both X^{mis}, μ, σ is not possible.

In the remainder of this chapter, and of this dissertation in general, we aim in finding appropriate reparametrisations, that correct this problem.

2.4 Transformation of the diffusion path

2.4.1 Reparametrisation

Girsanov's formula provides us the the Radon-Nikodym derivative with respect to a diffusion with zero drift and the same volatility as X . Hence, we look for a transformation of the diffusion path $U_t = h(X_t, \theta) = [h_1(X_t, \theta), \dots, h_d(X_t, \theta)]'$ so that the volatility matrix of the transformed diffusion does not contain parameters² or equivalently is the identity matrix. Using Ito's lemma, we obtain that $h(X_t, \theta)$ should satisfy the following vector differential equation

²being in a Bayesian framework we treat the parameters as random variables

$$dU_r = g(t, X, \theta)dt + \sum_i \left\{ \frac{\partial h_r}{\partial X_t^{\{i\}}}(X_t, \theta) \sum_j \sigma^{\{ij\}}(X_t, \theta) dB_t^{\{j\}} \right\},$$

for some function $g(\cdot)$ and $i, j, r \in \{1, \dots, d\}$. So we are looking for a function that will satisfy:

$$\sum_i \left\{ \frac{\partial h_r}{\partial X_t^{\{i\}}}(X_t, \theta) \sum_j \sigma^{\{ij\}}(X_t, \theta) dB_t^{\{j\}} \right\} = dB_t^{\{r\}}, \forall r \in \{1, \dots, d\},$$

or equivalently

$$\nabla h \Sigma (\nabla h)' = I_d. \quad (2.2)$$

Furthermore, the transformation inherits the Ito's lemma assumptions and has to be invertible. The SDE of the r -th coordinate of the transformed diffusion U will be given by:

$$dU_t^{\{r\}} = \sum_{i=1}^d \frac{\partial h_r(X_t, \theta)}{\partial X_t^{\{i\}}} dX_i, \quad r = 1, \dots, d,$$

which because of (2.2) becomes:

$$dU_t^{\{r\}} = \sum_{i=1}^d \frac{\partial h_r(X_t, \theta)}{\partial X_t^{\{i\}}} \mu^{\{i\}}(X_t, \theta) dt + dB_t^{\{r\}} = \mu_U^{\{r\}}(X_t, \theta) dt + dB_t^{\{r\}}, \quad r = 1, \dots, d.$$

Using Girsanov's theorem as before, we may now re-attempt to write down the likelihood:

$$\frac{d\mathbb{P}_\theta}{\mathbb{Q}^{h(y, \theta)} \otimes \text{Leb}_d(h(Y))} (U^{mis}, h(Y, \theta)) = G(U, \mu_U, I_d) f_M(Y; \theta),$$

or equivalently

$$\frac{d\mathbb{P}_\theta}{\mathbb{Q}^{h(y,\theta)} \otimes \text{Leb}_d(Y)} (U^{mis}, h(Y, \theta)) = G(U, \mu_U, I_d) \times n(h(Y); I_d) |J(Y, \theta)|^{-1},$$

where $J(Y, \theta)$ is the inverse of the Jacobian of the transformation $h(X_t, \theta)$ and $n(h(Y); I_d)$ is the normal density obtained under standard Wiener measure. This allows us to relax the assumption made in Section 2.3.2, that $f_M(Y; \theta)$ is known. Nevertheless, we have imposed the restriction of the existence of such a transformation $h(\cdot)$. We provide a comprehensive discussion on that in the next section.

Now let's have a look at the new reference measure $\mathbb{Q}^{h(y,\theta)}$. It reflects the law of d independent Brownian bridges conditioned by the observations $h(Y, \theta)$. Therefore it still depends on parameters. For this reason we introduce a second transformation. This transformation 'centers' the bridge, to start and finish at 0 in a way so that it will still have unit volatility. We define it below for the more general case of $t_{k-1} \leq t \leq t_k$, $(Y_{k-1}^{\{1\}}, \dots, Y_{k-1}^{\{d\}}) = (X_{t_{k-1}}^{\{1\}}, \dots, X_{t_{k-1}}^{\{d\}})$ and $(Y_k^{\{1\}}, \dots, Y_k^{\{d\}}) = (X_{t_k}^{\{1\}}, \dots, X_{t_k}^{\{d\}})$:

$$Z^{\{i\}}(s) = U^{\{i\}}(s) - \frac{(t_k - s)h(Y_{t_{k-1}}^{\{i\}}, \theta)(t_{k-1}) + (s - t_{k-1})h(Y_{t_k}^{\{i\}}, \theta)}{t_k - t_{k-1}}, \quad t_{k-1} < s < t_k, \quad \forall i, k. \quad (2.3)$$

Let $U = \eta(Z)$ be the inverse of this transformation:

$$U^{\{i\}}(s) = Z^{\{i\}}(s) + \frac{(t_k - s)h(Y_{t_{k-1}}^{\{i\}}, \theta)(t_{k-1}) + (s - t_{k-1})h(Y_{t_k}^{\{i\}}, \theta)}{t_k - t_{k-1}}, \quad t_{k-1} < s < t_k, \quad \forall i, k$$

The SDE for Z then becomes (again from Ito's lemma):

$$dZ_t = \mu_{U_t}(\eta(Z_t), \theta)dt + dB_t$$

We can now write the likelihood as:

$$\frac{d\mathbb{P}_\theta}{\mathbb{Q}^0 \otimes Leb_d(Y)}(Z^{mis}, h(Y, \theta)) = G(\eta(Z_t), \mu_U, I_d) \times n(h(Y); I_d) J^{-1}(Y, \theta).$$

The relevant dominating measure now reflects the distribution of d independent Brownian bridges that start and finish at 0.

A MCMC scheme based on the likelihood above is irreducible. The information contained in the quadratic covariation process does not affect the algorithm because all of the parameters have been moved to the drift function. Note also that as a result of these transformations, inference will now be based on Z , rather than X . However, since the transformations are invertible, it is straightforward to invert the MCMC output and obtain samples of X .

Figure 2.2 shows a graphical representation of the transformations on a coordinate of the process. The initial path is the one at the top. After applying the first transformation its volatility becomes equal to 1. Finally, the second transformation ‘centers’ the path so that it starts and finishes at 0.

2.4.2 Which diffusions can we handle?

The reparametrisations introduced in Section 2.4.1 do not cover all the diffusions defined by (1.1). The restrictions are imposed from the first transformation that stems from the solution of the system of partial differential equations of (2.2). Consequently the methodology of this chapter does not apply to diffusions for which this system is inconsistent.

Most known univariate diffusions have a solution as studied in Roberts and Stramer (2001). For multivariate diffusions the crucial aspect is the form of their volatility function $\sigma(X_t, \theta)$. Suppose that a diffusion has constant volatility $\sigma(X_t, \theta) = \sigma$ and general drift $\mu(X_t, \theta)$. Consider a transformation of the form $U_t = h(X_t) = \alpha X_t$, with α, σ being a $d \times d$ matrices. Using Ito’s formula we get

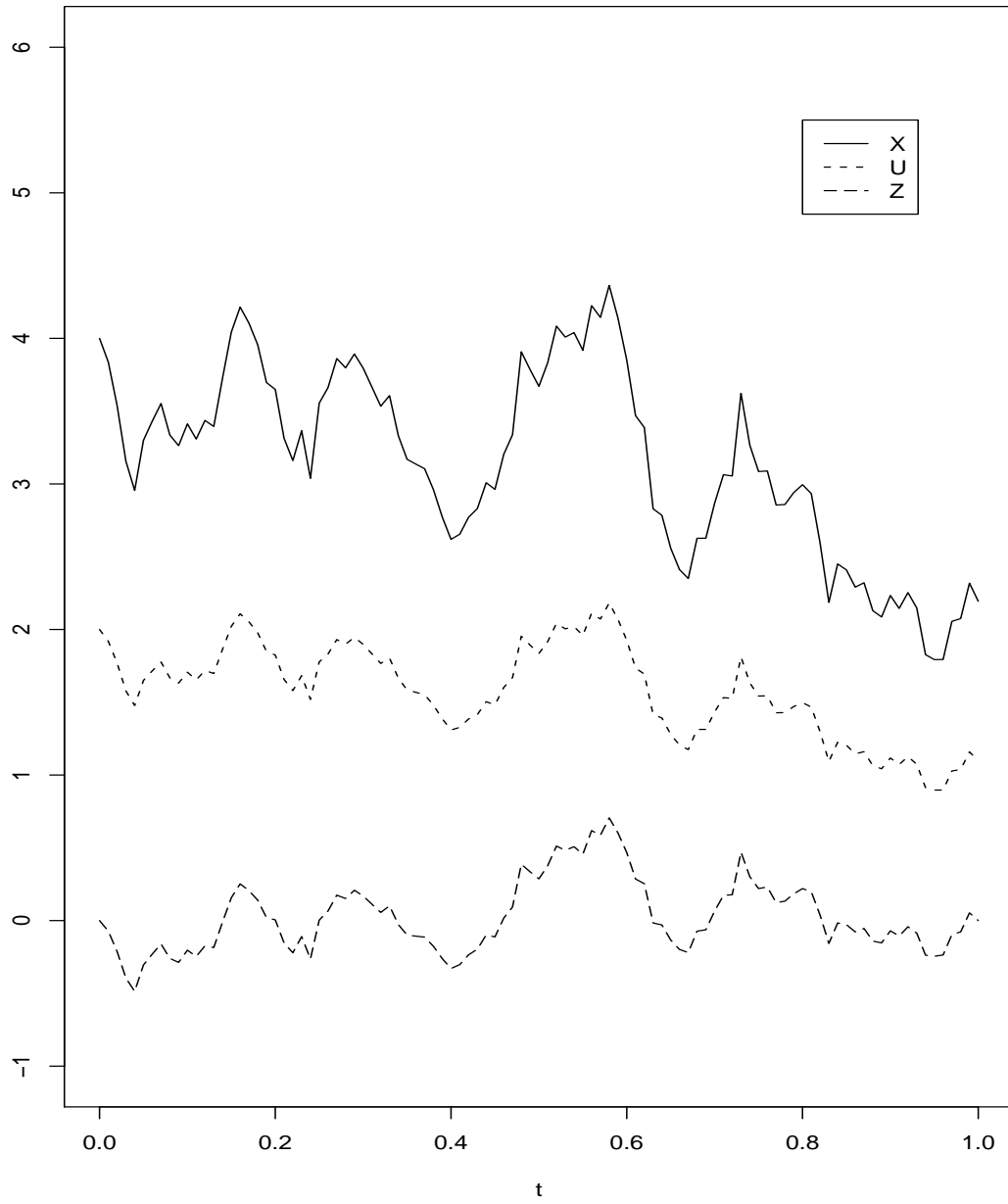


Figure 2.2: Graphical representation of the transformations of X to U and Z for scalar diffusions.

$$dU_t^{\{r\}} = \sum_{i=1}^d \frac{\partial h_r(X_t, \theta)}{\partial X^{\{i\}}} dX^{\{i\}}, \quad r = 1, \dots, d.$$

or

$$dU_t = \alpha \mu(X_t, \theta) + \alpha \sigma dB_t.$$

It is not difficult to see that the transformation with $\alpha = \sigma^{-1}$ ($U_t = h(X_t) = \sigma^{-1} X_t$) will lead us to unit volatility.

For another example consider diffusions with $\sigma(X_t, \theta)$ of the form:

$$\begin{pmatrix} \sigma_{11}(X_t^{\{1\}}, \theta) & 0 & \dots & 0 \\ 0 & \sigma_{22}(X_t^{\{2\}}, \theta) & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \sigma_{dd}(X_t^{\{d\}}, \theta) \end{pmatrix}$$

The coordinate processes $\{X_t^{\{1\}}, \dots, X_t^{\{d\}}\}$ are not necessarily independent as they may be related through the drift function. For such diffusion the system of (2.2) breaks into d ordinary differential equations and the problem becomes equivalent with that of the scalar diffusions case.

In other words the class of diffusions we can deal with, is the class of *reducible* diffusions introduced in Aït-Sahalia (2005) in the context of a multivariate extension to the framework developed in Aït-Sahalia (2002). Interestingly enough, the proposed analytical approximations of Aït-Sahalia (2005) have better properties for the class of reducible diffusions. Aït-Sahalia (2005) also offers a sufficient condition for reducibility (Aït-Sahalia, 2005, proposition 1): A diffusion X is reducible if $\sigma(X, \theta)$ satisfies

$$\frac{\partial [\sigma^{-1}(X, \theta)]_{ij}}{\partial X^{\{k\}}} = \frac{\partial [\sigma^{-1}(X, \theta)]_{ik}}{\partial X^{\{j\}}}, \quad (2.4)$$

for each $(i, j, k) = 1, \dots, d$ such that $k > j$.

2.4.3 Stochastic volatility models

Stochastic volatility models are quite famous models with a broad range of applications in finance. Their diffusion driven version can be seen as a 2–dimensional diffusion with a dispersion matrix of the following form:

$$\begin{pmatrix} \sigma_{11}(X_t^{\{2\}}, \theta) & 0 \\ 0 & \sigma_{11}(X_t^{\{2\}}, \theta) \end{pmatrix}, \quad (2.5)$$

which does not satisfy (2.4). The Ait-Sahalia condition is satisfied however, if $\sigma(X_t, \theta)$ has the form:

$$\begin{pmatrix} a(X_t^{\{1\}}, \theta) & a(X_t^{\{1\}}, \theta)b(X_t^{\{2\}}, \theta) \\ 0 & c(X_t^{\{2\}}, \theta) \end{pmatrix}.$$

For example if we set $a(X_t^{\{1\}}, \theta) = \sigma_{11}$, $b(X_t^{\{2\}}, \theta) = \sigma_{12} \exp(X_t^{\{2\}})$ and $c(X_t^{\{2\}}, \theta) = \sigma_{22}$ we obtain the following transformation:

$$h_1(X_t^{\{2\}}, X_t^{\{2\}}) = \frac{X_t^{\{1\}}}{\sigma_{11}} - \frac{\sigma_{12} \exp(X_t^{\{2\}})}{\sigma_{11} \sigma_{22}}$$

$$h_2(X_t^{\{2\}}) = \frac{X_t^{\{2\}}}{\sigma_{22}}$$

In the Chapters 2 and 3 of this dissertation, we focus on the usual, and easier to interpret, class of stochastic volatility models of (2.5). For these models a separate framework has been developed and alternative data augmentation irreducible MCMC schemes are available.

2.5 Markov chain Monte Carlo implementation

Using the likelihood of the previous section it is now possible to construct an irreducible data augmentation MCMC scheme. The algorithm can be described as a Gibbs sampling scheme which may be divided into 3 parts: the updates of the paths, the parameters of the volatility function and those of the drift. The updates for the drift parameters are rather straightforward as they can be carried out using standard random walk Metropolis techniques (see Section 1.4.2). We present the details for each of the other two parts in the following sections.

2.5.1 Updating the imputed paths

We proceed using an independence sampler. In other words we will propose from a proposal distribution and the proposed paths will be either accepted or rejected according to an appropriate probability that will ensure the convergence of the chain to the target posterior distribution. Suppose that the augmented path is divided into $n \times d$ diffusion bridges connecting the - say n - observed points (apart from the initial one). Under the Gibbs sampling framework we can further break the path into $n \times d$ parts, one per bridge, and update each one of them in turn. Using the results from the previous sections we see that:

$$\frac{d\mathbb{P}_\theta}{d\mathbb{Q}^0}(Z_{mis}|Y) = G(\eta(Z_t), \mu_U, I_d) \frac{f_M(Y; \sigma)}{f_X(Y; \sigma)} \propto G(\eta(Z_t), \mu_U, I_d), \quad (2.6)$$

where $f_X(Y; \sigma)$ is the density of Y with respect to the Lebesgue measure under \mathbb{P}_θ .

One simple choice for a proposal is the reference measure \mathbb{Q}^0 ; in other words we may propose Brownian bridges. Using (2.6), we end up with the following algorithm:

- **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 or reject with probability:

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

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

Updating the entire path at once will probably result in extremely low acceptance rates as the discrepancy between the proposed and current path would be substantial. By splitting into blocks, we correct for that increasing the acceptance rate; see Elerian (1999) for more details. On the other hand, Elerian also notes that smaller blocks result in slower convergence and mixing. Appropriate tuning is thus required to optimise this MCMC efficiency trade-off. An alternative way of increasing the acceptance is by a proposal distribution which is closer to the target \mathbb{P}_θ . Suppose that we propose from another distribution of diffusion bridges with drift λ denoted by \mathbb{L}^0 . We can now write using 2.6:

$$\frac{d\mathbb{P}_\theta}{d\mathbb{L}^0}(Z_{mis}|Y) = \frac{d\mathbb{P}_\theta/d\mathbb{Q}^0}{d\mathbb{L}^0/d\mathbb{Q}^0}(Z_{mis}|Y) \propto \frac{G(\eta(Z_t), \mu_U, I_d)}{G(\eta(Z_t), \lambda, I_d)} \quad (2.7)$$

Based on (2.7) above, the corresponding algorithm 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 or reject with probability:

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

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

A suitable choice for \mathbb{L}^0 is a linear diffusion bridge, chosen so that $\lambda^{\{i\}}$ approximates $\mu_U^{\{i\}}$.

A natural choice for λ , proposed by Ozaki (1992) is:

$$\lambda(L_t) = \mu_U(h(Y_{k-1})) + \frac{\partial \mu_U(h(Y_{k-1}))}{\partial U_t} (L_t - h(Y_{k-1}))$$

Setting $g = -2\frac{\partial \mu_U}{\partial U_t}(h(Y_{k-1}))$ and $b = \mu_U(h(Y_{k-1})) + 0.5gh(Y_{k-1})$, the SDE for the linear diffusion bridge L_t for $t_{k-1} \leq t \leq t_k$ (and with $L_{t_{k-1}} = h(Y_{k-1})$) is:

$$d\lambda(L_t) = dB_t + \left\{ b - \frac{gL_t}{2} - g^{1/2}e^{gt/2} \frac{g^{1/2}e^{gt/2}L_t - 2be^{gt/2}/g^{1/2} + (2b/g - h(Y_k))e^{gt_k/2}g^{1/2}}{e^{gt_k} - e^{gt}} \right\}$$

See Ozaki (1992) and Roberts and Stramer (2001) for more details. Note that such a bridge will start and finish at the corresponding values of $h(Y)$, therefore the transformation (2.3) should also be used. Alternatively one may work with U_t instead of Z_t keeping in mind that Z_t is the component of the Gibbs sampling scheme.

In the adverse case where the path has been split into $n \times d$ bridges and still both proposals give low acceptance rates, one may further split the path using overlapping bridges. See Section 3.4 for more details.

2.5.2 Updating the volatility parameters

In this section we assume that $\sigma(X_t, \theta)$ is a constant $d \times d$ matrix with $\sigma^{\{ij\}}(\cdot) = \sigma_{ij}$. The matrix $\Sigma = \sigma \times \sigma'$ is a covariance matrix and therefore can be treated using standard relevant MCMC techniques; see for instance Daniels and Kass (1999), Pinheiro and Bates (1996).

Note however that the full conditional of Σ will generally not be available. In this case the options for updating Σ are limited. A random walk on the space of positive definite matrices with an Inverse Wishart distribution is not the optimal choice, especially for high dimensions. Such a proposal provides only one tuning parameter (the degrees of freedom) and its flexibility is therefore limited. Similarly, the performance of an independence proposal is again questionable. Alternatively, we may attempt to update each one of its

components separately using a random walk. Such an approach will have no implications imposed by the multi-dimensionality. However it is not clear how the positive definiteness restrictions may be imposed under such a scheme.

We propose to update the matrix σ componentwise. Before proceeding any further let's recall one of the initial assumptions made in the introduction, that concerns the 1 – 1 correspondence of σ and Σ . Note that the likelihood contains only Σ and therefore such an assumption is necessary to ensure the identifiability of the components of σ . We adopt the Cholesky factorisation. In other words we assume a triangular σ of the form

$$\sigma = \begin{pmatrix} \sigma_{11} & 0 & 0 & \dots & 0 \\ \sigma_{21} & \sigma_{22} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ \sigma_{d1} & \sigma_{d2} & \sigma_{d3} & \dots & \sigma_{dd} \end{pmatrix},$$

with positive diagonal elements. Now note that:

- The entire space of symmetric positive definite matrices is covered since we are using the Cholesky decomposition.
- The matrix Σ will always be strictly positive definite since the diagonal entries of σ are positive and therefore $|\sigma| > 0$.

It is now safe to adopt random walk metropolis schemes for each one of the components of σ . Alternative factorisations are possible; see for instance Daniels and Kass (1999) and Pinheiro and Bates (1996) for more details.

2.6 Extensions

2.6.1 Change-points in volatility

The restrictions imposed by the system of partial differential equations of (2.2) may limit the flexibility in modelling diffusions with time varying volatilities. A solution to this is

given with the methodology developed in Chapters 2 and 3. Nevertheless, the following extension on the methodology of this chapter may prove particularly useful. Consider the following category of diffusions on $0 \leq t \leq T$ with general drift (as long as it ensures a weakly unique solution to the SDE) $\mu(X_t, \theta)$ and volatility $\sigma(X_t, \theta)$ of the following form:

$$\sigma(X_t, \theta) = \sigma_c(X_t, \theta), \quad t_{c-1} \leq t \leq t_c, \quad c = 1, \dots, l, \quad t_0 = 0$$

where the collections of $\{t_c\}$ is subset of times with observations $\{t_k, 1, \dots, n\}$ and $\sigma_c(\cdot)$ belongs to a reducible diffusion (i.e. the system of (2.2) has a solution) for all c . Since the volatility $\sigma(X_t, \theta)$ is not continuous, the existence of a unique solution on the entire SDE is not guaranteed and the Girsanov's theorem is not applicable. Nonetheless, $\sigma(X_t, \theta)$ is continuous in the intervals $[t_{k-1}, t_k]$ and therefore the likelihood is well defined in each one of them. If we use the Markov property we can then define the likelihood of the entire path by simply multiplying the likelihoods obtained from the separate bits.

Furthermore, the likelihood is not actually defined based on X but on Z . Note that

$$U_t = h_c(X_t, \theta), \quad t_{c-1} \leq t \leq t_c, \quad c = 1, \dots, l, \quad t_0 = 0$$

where $h_c(X_t, \theta)$ is the transformation that solves the system (2.2) for a diffusion with volatility $\sigma_c(X_t, \theta)$. Note also that U_t is now a discontinuous process as $h_c(Y_k, \theta)$ is not generally equal to $h_{c+1}(Y_k, \theta)$. Nevertheless, the likelihood will again be well defined, apart from the fact that it is written with respect to a parameter dependent dominating measure as before. The second transformation Z_t will be defined in exactly the same way after a careful choice of the appropriate transformed data points $h_c(Y_{k-1}, \theta)$ and $h_c(Y_k, \theta)$. Note that Z_t is now a continuous process with continuous (unit) volatility and therefore there is absolutely no concern whatsoever on the validity of the likelihood. These are illustrated on Figure 2.3 where sample paths from all X_t , U_t and Z_t are plotted.

In our discussion so far, both the location of the changepoints $\{t_c, c = 1, \dots, l\}$ and their number were assumed to be known and fixed. However, under an appropriate

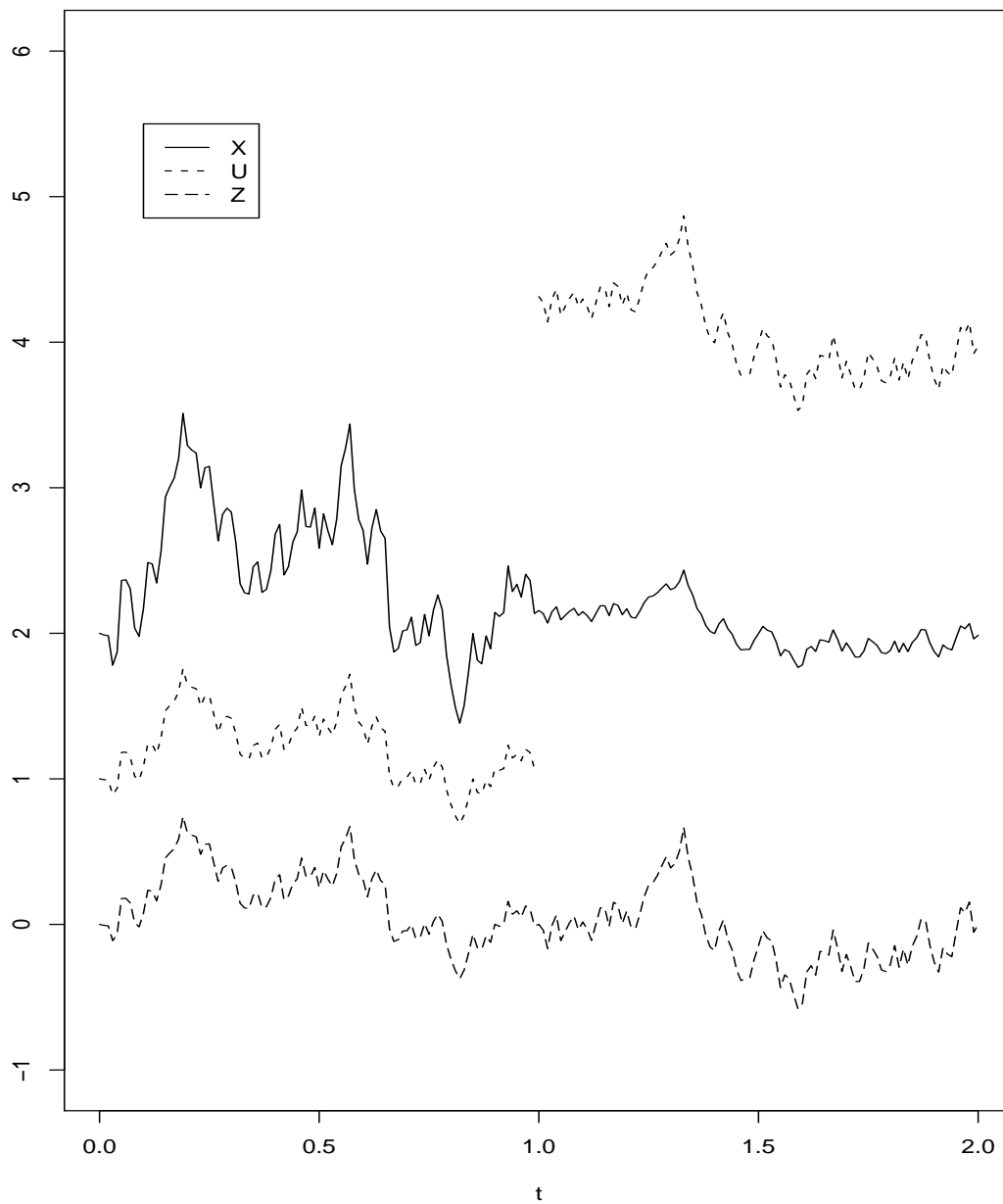


Figure 2.3: Graphical representation of the transformations of X to U and Z bottom in the case of changepoints in the volatility.

reversible jump MCMC methodology, introduced in Green (1995), they may be included in the model as parameters. In the context of diffusions a reversible jump was used in Dellaportas et al. (2006), but this issue was not addressed nor was the model choice for multivariate diffusions in general. The methodology of this chapter sets the basis for such an extension.

2.6.2 Multivariate NLCAR models

A particularly interesting class of continuous time processes with a wide range of applications contains the non linear continuous time autoregressive models, denoted by NLCAR(p). They were introduced in Tsai and Chan (2000) and include among others the continuous time linear autoregressive models CAR(p) and the continuous time threshold models CTAR(p) introduced by Brockwell (1994). They are defined by the following SDE:

$$\begin{aligned}
 dX_0(t) &= X_1(t)dt \\
 dX_1(t) &= X_2(t)dt \\
 &\vdots \\
 dX_{p-2}(t) &= X_{p-1}(t)dt \\
 dX_{p-1}(t) &= \mu(X(t), \theta)dt + \sigma dBt
 \end{aligned} \tag{2.8}$$

where $\mathbf{X}(t) = (X_0, X_1, \dots, X_{p-1})'$. In words, we assume that the $p - 1$ th derivative follows a diffusion with drift $g(\cdot)$ and volatility σ ; an NLCAR(1) is an ordinary diffusion process. The drift $\mu(\cdot)$ has to satisfy growth conditions in order for the equation above to have a solution; see Brockwell (1994). Note that under a fine discretisation and given the initial point, if either of X_0, X_1, \dots, X_{p-1} is available, the remaining ones can be obtained deterministically from it using numerical integration or differentiation.

In practice we only observe a discrete skeleton of X_0 and nothing on X_{p-1} . The same principles apply though: If we augment the path the path of X_{p-1} at a sufficiently thin scale, the effect of its quadratic variation process will cause again the same reducibility

issues. Roberts and Stramer (2004) proposed a 2-step reparametrisation in the spirit of Roberts and Stramer (2001).

The results of this chapter may be used for a multivariate extension of NL $\text{CAR}(p)$ models. To define multivariate NL $\text{CAR}(p)$ assume a univariate NL $\text{CAR}(p)$ on each dimension. Now $\mathbf{X}(t) = (x_j^i, i = 1, \dots, d, j = 0, 1, \dots, p - 1)$ and the process of the $p-1$ th derivatives can be written as in (2.8) but now σ is a $d \times d$ constant matrix. Following the same steps as in Brockwell (1994), with an application of the multivariate version of Girsanov's theorem, we require the drift to satisfy the same growth conditions to ensure the existence of a unique weak solution.

Now assume that $X_0(0) = 0$ and $X_0(1) = y$, both d -dimensional vectors. Denoting by \mathbb{P}_Θ the law of \mathbf{X} and with \mathbb{Q}_σ that of its driftless version, which we denote by \mathbf{M} , we can write (again $\Sigma = \sigma\sigma'$):

$$\frac{d\mathbb{P}_\Theta}{d\mathbb{Q}_\sigma} = G(\mathbf{X}, \mu, \sigma) = \exp \left\{ \int_0^T [\Sigma^{-1}\mu(\mathbf{X}(s), \theta)]' dX_{p-1}(s) - \frac{1}{2} \int_0^T \mu(\mathbf{X}(s), \theta)' \Sigma^{-1} \mu(\mathbf{X}(s), \theta) ds \right\}.$$

If we factorise the dominating measure in the same as in Section 2.3.2 we get the likelihood:

$$\frac{d\mathbb{P}_\Theta}{\mathbb{Q}_\sigma^y \otimes \text{Leb}_d(Y)}(\mathbf{X}^{mis}, Y) = G(\mathbf{X}, \mu, \sigma) \times f_M(y; \theta),$$

where $f_M(y; \theta)$ is the density of y under the dominating measure with respect to the Lebesgue measure. This would be the density of a Gaussian process, the mean and variance of which, would depend on the order p and Σ .

To remove the dependence of \mathbb{Q}_σ^y on σ we introduce the first transformation on X_{p-1} (and therefore on the entire \mathbf{X}) to be $U_{p-1} = h(X_{p-1}) = \sigma^{-1}X_{p-1}$, so that the volatility

of U_{p-1} is the identity matrix. The second transformation, applied on U_0 and therefore on the entire \mathbf{U} , is:

$$Z_0(t) = U_0(t) - E_{\mathbb{Q}_\sigma}(Z_0(t) | Z_0 t_k = \sigma^{-1} Y_k),$$

where the expectation is taken with respect to \mathbb{Q}_σ which represents a Gaussian process so it is relatively easy to compute. For $p = 1$ this transformation becomes the transformation of 2.4.1. It is easy to check that the likelihood in terms of \mathbf{Z} will have a reference measure that does not depend on σ or any other parameters.

Therefore it is possible to compute suitable MCMC scheme. Using the techniques of this chapter the multi-dimensionality should not introduce any additional problems as we can break down all of the MCMC components into small blocks.

2.7 Simulations

We simulated 500,001 observations from the following non-linear model using a high frequency approximating Euler scheme of step 0.001.

$$dX(t) = \begin{pmatrix} -\theta_1(X_t^{\{1\}})^3 \\ -\theta_2(X_t^{\{2\}})^3 \end{pmatrix} dt + \begin{pmatrix} \sigma_{11} & 0 \\ \sigma_{21} & \sigma_{22} \end{pmatrix} dB(t).$$

We only recorded one point every 1000, resulting thus in a dataset with 500 observations (excluding the initial point). Denote the data by $Y = \{Y_k = X_{t_k}, t_k = 0, 1, 2, \dots, 500\}$. The parameters were set to with $\theta_1 = 0.8$, $\theta_2 = 0.6$, $\sigma_{11} = 0.5$, $\sigma_{21} = 0.2$ and $\sigma_{22} = 0.3$.

Solving the vector differential equation of (2.2), we get the first transformed diffusion U_t from $U_t = h(X_t) = \sigma^{-1} X_t$. The second transformation Z_t is defined from (2.3). The density $f_M(Y; \sigma)$ is given by ($\Sigma = \sigma\sigma'$):

$$(|\Sigma|)^{-n/2} \exp \left\{ - \sum_{k=1}^{500} \frac{[Y_k - Y_{k-1}]' \Sigma^{-1} [Y_k - Y_{k-1}]}{2(t_k - t_{k-1})} \right\}$$

We put improper flat priors on all the parameters, restricting θ_1 , θ_2 , σ_{11} and σ_{22} to be positive. This was done to ensure first the non-explosivity of X (θ_1 and θ_2), and second Identifiability - 1-1 correspondence of σ and Σ (σ_{11} and σ_{22}). We then implemented the data augmentation scheme introduced in this chapter. A separate MCMC of 10,000 iterations was run for 3 different values of imputed points $m = 20, 40, 50$. The path proposals were Brownian bridges and since the acceptance rate was high (87.2 %), no other proposal was considered.

In Figure 2.4 at the bottom right, we see density plots of the log-likelihood for the 3 different values of m . Convergence of the likelihood may be used as an overall diagnostic check that the discretisation is sufficiently fine. The idea is that if two partitions give roughly the same likelihood approximation then the discretisation error is arguably sufficiently small. In our case the density plots of $m = 40$ and $m = 50$ are almost identical whereas that of $m = 20$ seems substantially different. One may reasonably argue that a discretisation corresponding to an $m = 40$ is enough. Otherwise we may just increase m at small computational cost.

Figure 2.5 contains autocorrelation plots for the posterior draws of all parameters. The picture we get is completely different than that of Figure 2.1. There is no sign of any increase of the autocorrelation for any of the parameters. Moreover, there is nothing to raise suspicion on the convergence of the chain.

Finally, Table 2.1 contains information on the posterior samples of the parameters for $m = 50$. The estimates seem to be in good agreement with the true values the data were simulated from. However, this is not the case regarding the quadratic variation estimates which are biased downwards. It seems that the imputation of the 50 intermediate points between each pair of successive observations has successfully restored the inadequacy of our sparse dataset.

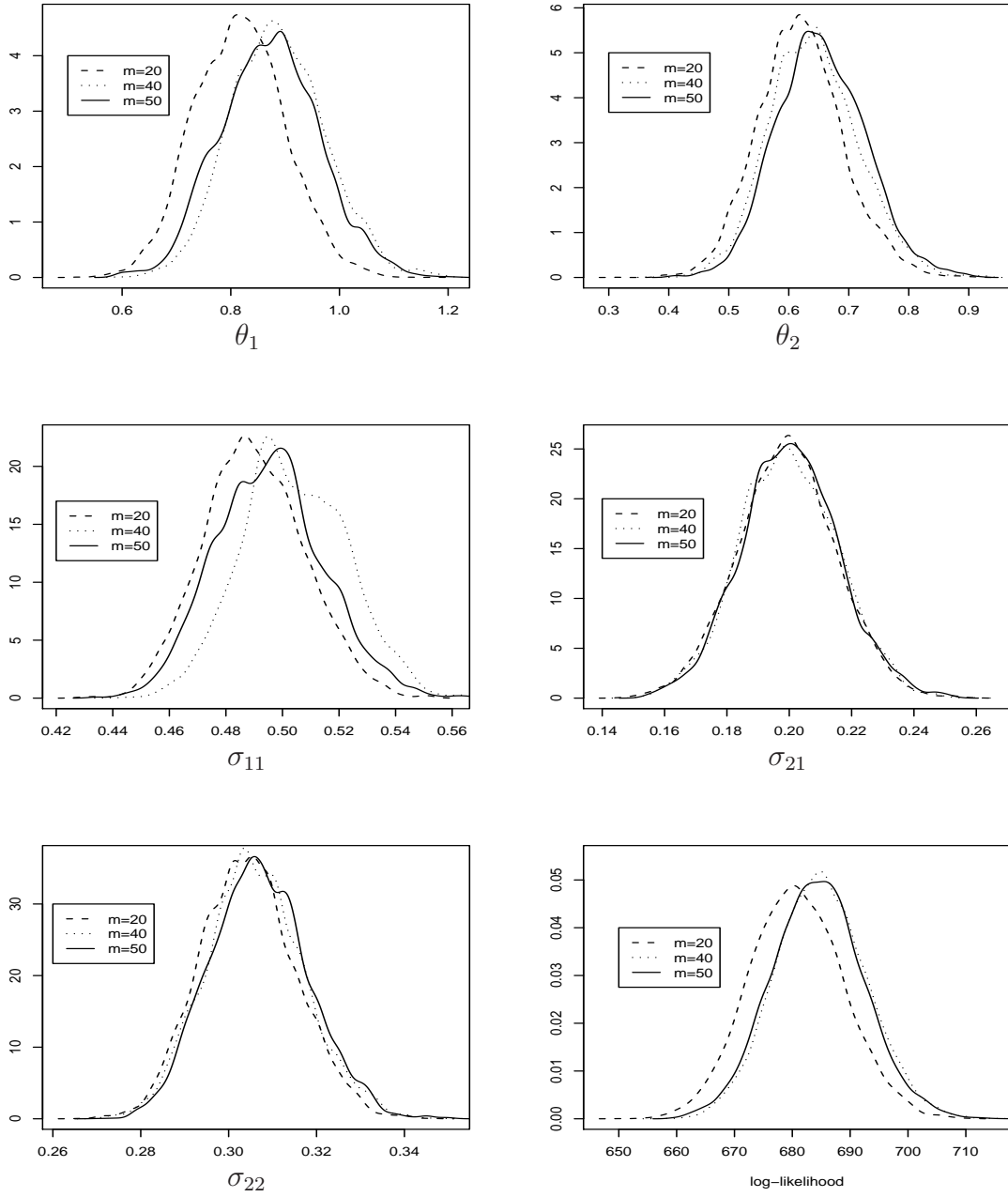


Figure 2.4: Kernel densities of the posterior draws of all the parameters for different values of imputed points $m = 20, 40, 50$.

Parameter	True value	Post. mean	Post. SD	Post 2.5%	Post median	Post 97.5%
θ_1	0.8	0.872	0.091	0.703	0.873	1.060
θ_2	0.6	0.654	0.074	0.515	0.651	0.809
σ_{11}	0.5	0.495	0.019	0.460	0.496	0.536
σ_{21}	0.2	0.200	0.016	0.170	0.200	0.232
σ_{22}	0.3	0.307	0.011	0.286	0.307	0.330

Table 2.1: Summaries of the posterior draws for the simulation example.

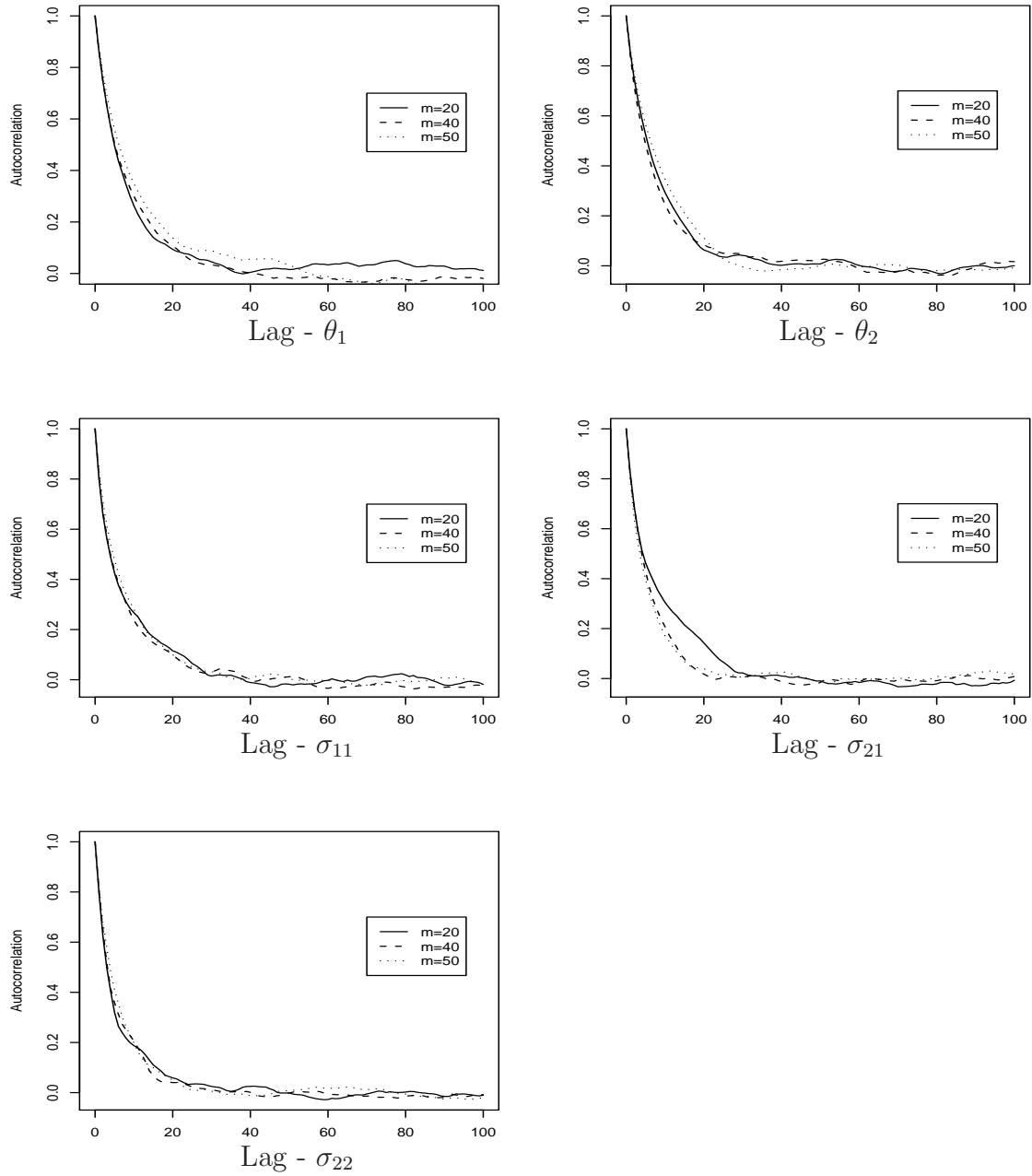


Figure 2.5: Autocorrelation plots for the posterior draws of all the parameters for different values of imputed points $m = 20, 40, 50$.

2.8 Real data example

In this section we illustrate our methodology on the European interest rates dataset of Corzo and Schwartz (2000). The data contain one-month interbank middle (between bid and ask) rate for the Spanish market and one-month deposits for the ECU interest rate. Weekly data from September 1990 to December 1997 (382 observations) were used. The data are plotted in Figure 2.6.

The primary aim of the analysis was to study the interest rate process and investigate the convergence, if any, of the Spanish interest rate to the European one. Thus, a natural diffusion model for the Spanish interest rate $X^{\{s\}}$ is the following:

$$dX_t^{\{s\}} = b_1 + b_2(X_t^{\{e\}} - X_t^{\{s\}})dt + \sigma_{11}dB_t$$

where $X^{\{e\}}$ is the European rate to which we assign the following linear drift model:

$$dX_t^{\{e\}} = (b_3 + b_4X_t^{\{e\}})dt + \rho\sigma_{22}dB_t + \sqrt{1 - \rho^2}\sigma_{22}dW_t.$$

The equation above implies that the interest rate increments may be correlated and their correlation is denoted by ρ . The presence of b_1 in the drift allows for a potential (minor) individual divergence of the Spanish rate whereas the parameter b_2 reflects the persistence of the convergence.

In a similar spirit with the analysis of Corzo and Schwartz (2000), we split the data in the middle (end of April 1994), and model the parameters separately, for two reasons:

- To account for the fact that the volatility is obviously higher in the first period for both rates.
- To examine if there were any changes in the behavior of the two processes between these two periods.

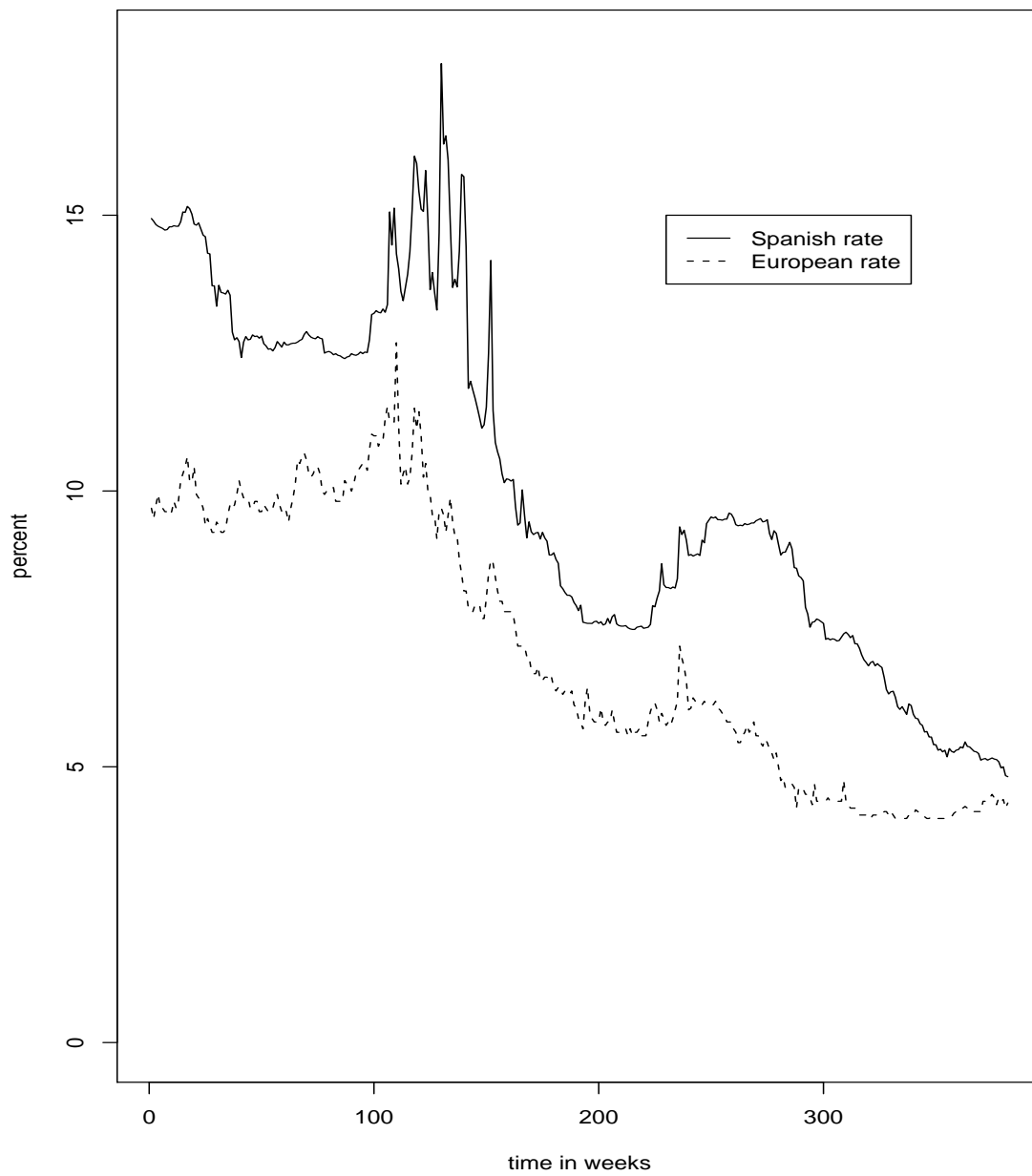


Figure 2.6: Kernel densities of the posterior draws of all the parameters for different values of imputed points $m = 20, 40, 50$

Hence, our model for the Spanish rate becomes

$$dX_t^{\{s\}} = b_1 + D_1 Z_t + (b_2 + D_2 Z_t)(X_t^{\{e\}} - X_t^{\{s\}})dt + (\sigma_{11} + D_3 Z_t)dB_t,$$

where Z_t equals 0 for the period up to April 1994, and 1 afterwards. Regarding the European rate, our preliminary analysis revealed that the parameters b_3 and b_4 were not significant. For simplicity and parsimony reasons we set them to 0:

$$dX_t^{\{e\}} = (\rho + D_4 Z_t)(\sigma_{22} + D_5 Z_t)dB_t + \sqrt{1 - (\rho + D_4 Z_t)^2}(\sigma_{22} + D_5 Z_t)dW_t.$$

We used the reparametrisations of Section 2.6.1 with one changepoint at the end of April 1994. We put improper flat priors on all the parameters imposing the necessary restrictions as in Section 2.5.2. Figure 2.7 depicts the autocorrelation plots for all the parameters for $m = 10, 20$ imputed points. There is no sign of increase in the autocorrelation with m and overall the efficiency of the chain is satisfactory.

From Figure 2.8 we can verify the fact that the discretisation error was reasonably controlled as it is really hard to distinguish the log-likelihood density of $m = 10$ from that of $m = 20$.

Note that this was achieved with much fewer imputed points than the simulation example. This is due to the fact that the deviation of the transformed diffusion paths from the reference measure was not substantial. This is also reflected in the acceptance rate of the Brownian bridge proposals which was extremely high (98%). Table 2.2 provides a summary of the MCMC output. We see that under our model, almost all the D -s are different from 0 with probabilities greater than 0.95 (0.912 for D_4), implying smaller variances and higher correlation for the second subperiod. Assuming that our model fits the data well, we may argue that the speed of convergence was higher during the first

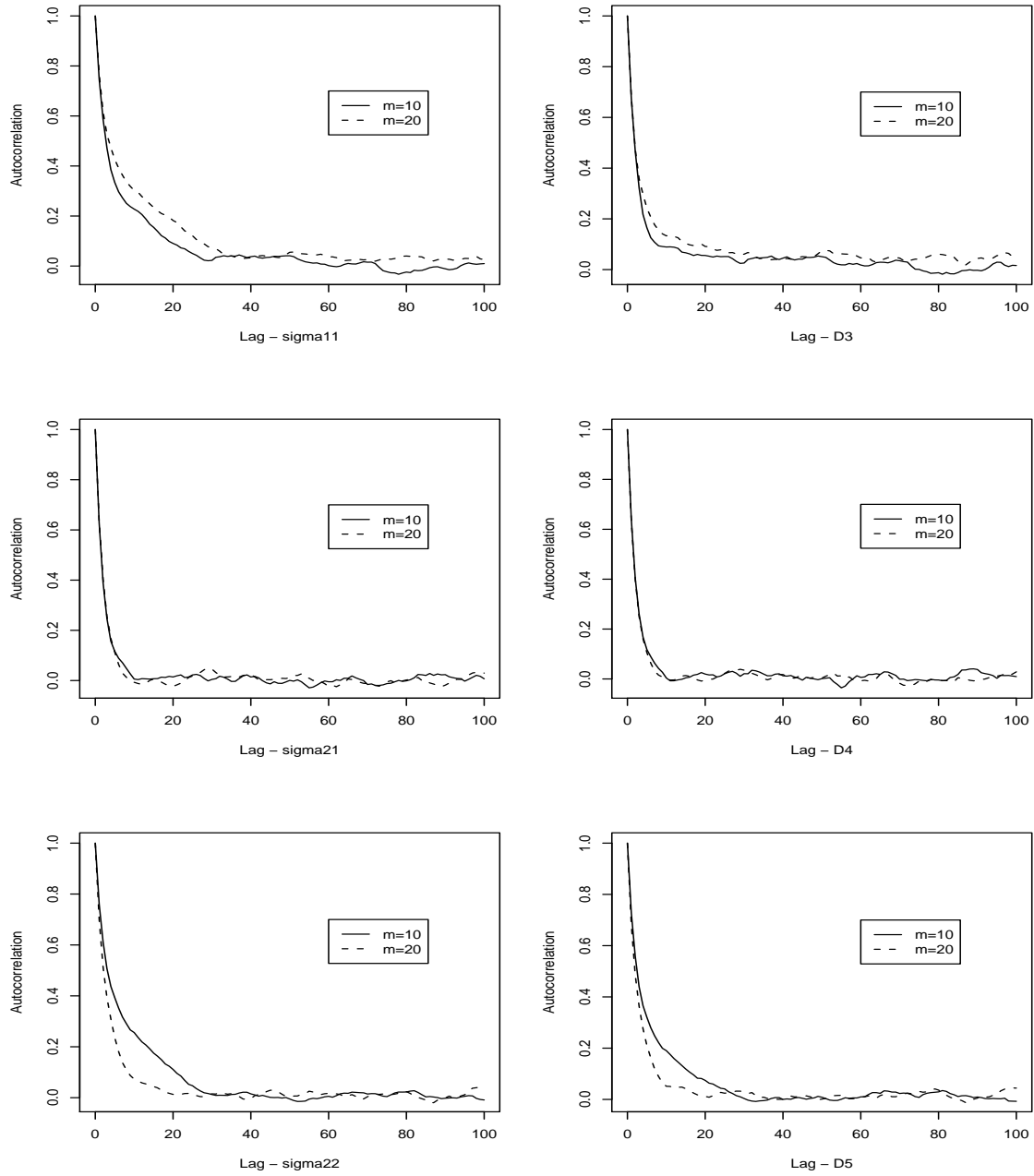


Figure 2.7: Kernel densities of the posterior draws of all the parameters for different values of imputed points $m = 20, 40, 50$. Spanish-European rates dataset.

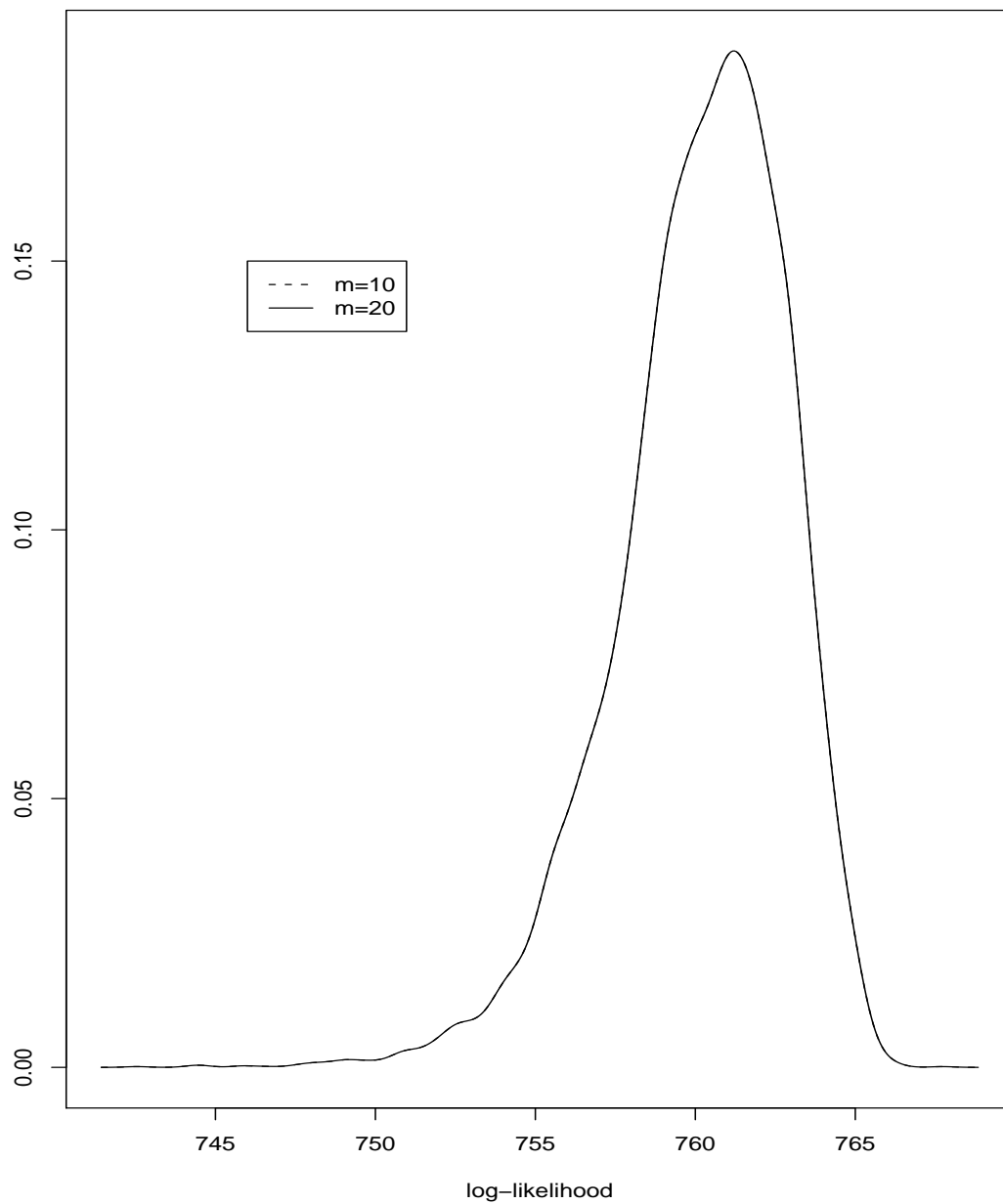


Figure 2.8: Kernel densities of the posterior draws of all the parameters for different values of imputed points $m = 20, 40, 50$. Spanish-European rates dataset.

Parameter	Post. mean	Post. SD	Post 2.5%	Post median	Post 97.5%
θ_1	0.280	0.108	0.073	0.276	0.497
θ_2	0.093	0.030	0.035	0.091	0.155
D_1	-0.262	0.107	-0.480	-0.258	-0.042
D_2	-0.080	0.030	-0.147	-0.080	-0.023
σ_1	0.537	0.030	0.480	0.536	0.599
σ_2	0.271	0.014	0.244	0.270	0.300
ρ	0.075	0.052	-0.029	0.077	0.176
D_3	-0.408	0.034	-0.473	-0.407	-0.342
D_4	0.142	0.072	-0.003	0.141	0.282
D_5	-0.125	0.016	-0.157	-0.125	-0.094

Table 2.2: Summaries of the posterior draws for the real data example. Spanish-European rates dataset.

subperiod September 1990 - April 1994. There exists a difference between 0.48 and 0.042 with probability greater than 0.95.

2.9 Concluding remarks

Multivariate diffusion models are of particular interest as they allow us to study beyond the individual characteristics of the univariate processes and explore potential interactions between them. In this chapter we provide a tool for Bayesian inference based on observations at a finite number of points, that covers a broad class of them. The methodology is a natural extension to the data augmentation framework developed in Roberts and Stramer (2001) and it includes likelihood approximations that may become arbitrarily small by simply increasing the augmentation.

The reparametrisation introduced in this chapter extends beyond the MCMC case and apply to any data augmentation scheme applied to a fixed dataset. As noted in Beskos et al. (2006b), this transformation is an essential to a multivariate extension to their framework and is also necessary for the convergence of an EM algorithm. In the MCMC context, the only way to avoid it is through the sequential techniques of Golightly and Wilkinson (2005) which contain the general online parameter estimation drawbacks.

The class of multivariate diffusions does not include some interesting cases like stochastic volatility models which are examined in the next two chapters. Nevertheless, it includes many interesting models in which the observed process may not necessarily be Markovian, like the multivariate NLCAR models.

Chapter 3

Bayesian Inference for a Large Class of Stochastic Volatility Models

3.1 Introduction

Volatility plays a prominent role in financial markets. Its importance is first noted in the early option pricing literature on the seminal work of Black and Scholes (1973) and Merton (1995). Under these formulations the equity prices followed a geometric Brownian motion implying a constant instantaneous relative volatility which drives option pricing through the famous Black-Scholes formula among practitioners. Despite the fact that today's evidence from almost all financial time series does not support this assumption, Black-Scholes remains the most widely used formula. Alternative formulations by Derman and Kani (1994), Dupire (1994) and Rubinstein (1995) consider state (price) dependent, time varying volatility. Although these models perform well in some cases, empirical studies (Dumas et al., 1998) show that they fail to explain the joint behavior of stocks and option prices.

Stochastic volatility models, such as Hull and White (1987), Stein and Stein (1991) and Heston (1993), where the volatility is a separate diffusion process, undoubtedly offer an improvement in that respect. Such models achieve to encompass some the empirical features of the data, as provided by both stocks and option prices. Although the problem was initially stated in a diffusion processes framework, the vast majority of the early

stochastic volatility literature is set in discrete time. These models focus on the skeleton of the process without worrying too much about its evolution over the intervals between the discrete points. See for instance Ghysels et al. (1996), Jacquier et al. (1994), Kim et al. (1998), Pitt and Shephard (1999) and their references therein. Continuous time stochastic volatility models aim in more realistic representation of the market. Also, unlike discrete time models, they can be fit on irregularly spaced, non-synchronized and multiple scale datasets which often appear in high frequency finance.

In this and the following chapter we consider all the diffusion driven stochastic volatility models that can be summarized or transformed to the following SDE:

$$\begin{pmatrix} dX_t \\ d\alpha_t \end{pmatrix} = \begin{pmatrix} \mu_x(X_t, \alpha_t, \theta) \\ \mu_\alpha(\alpha_t, \theta) \end{pmatrix} dt + \begin{pmatrix} \sigma_x(\alpha_t, \theta) & 0 \\ 0 & \sigma_\alpha(\alpha_t, \theta) \end{pmatrix} \begin{pmatrix} dB_t \\ dW_t \end{pmatrix}, \quad (3.1)$$

where X denotes the observed equity (stock) price, the volatility of which is driven by the latent diffusion α . To our knowledge, the SDE above contains almost all the models of the relevant literature and thus forms a very interesting family.

We divide the models of (3.1) into two categories. In the first category we assume that the drift of the observed process is independent of its paths. This covers the famous models of Hull and White (1987), Stein and Stein (1991) and Heston (1993)¹, and is the main focus of this chapter. In the next chapter we extend to the more general case by enhancing our machinery using time change transformations.

This chapter is organized as follows: Section 3.2 summarizes the methodologies available for likelihood based inference on stochastic volatility models. We approach the problem using data augmentation. Again appropriate reparametrisation and MCMC implementation techniques are essential. They are presented in Sections 3.3 and 3.4. Section 3.5 notes some possible extensions and Section 3.6 illustrates these on two simulation based examples. Then, in Section 3.7, we explore the relationship with option

¹It is easier to see this if we use $S = \exp(X)$ instead of X

pricing and we extend the framework to include cases where the latent diffusion α may be inferred from additional data (i.e. option prices). 3.8 provide a real data application on the Standard and Poor 500 - VIX data and finally Section 3.9 concludes with some potential extensions.

3.2 Inference for stochastic volatility models

Assume that the process we study can be divided into two components, say X and α , where we can only observe a discrete skeleton of the diffusion X . In this chapter, we focus on a subclass of the models included in (3.1), henceforth denoted by \mathcal{C} , considering multivariate diffusions with unobserved paths that satisfy the following SDE ($0 \leq t \leq T$):

$$\begin{pmatrix} dX_t \\ d\alpha_t \end{pmatrix} = \begin{pmatrix} \mu_x(\alpha_t, \theta) \\ \mu_\alpha(\alpha_t, \theta) \end{pmatrix} dt + \begin{pmatrix} \sigma_x(\alpha_t, \theta) & 0 \\ 0 & \sigma_\alpha(\alpha_t, \theta) \end{pmatrix} \begin{pmatrix} dB_t \\ dW_t \end{pmatrix}, \quad (3.2)$$

where B and W denote standard Brownian motions that may or may not be correlated. As already mentioned, the class \mathcal{C} includes many famous and widely used diffusion driven stochastic volatility models which are particularly useful in financial applications. Section 3.6.2 provides a simulation example based on the model of Heston (1993), henceforth referred to as Heston model, and Section 3.8 an analysis of the S&P 500 - VIX data. A crucial aspect of stochastic volatility models is the lack of the Markov property for the observed diffusion². In other words the volatility of a future stock price depends on the entire evolution of α rather than just the current value of X . Suppose that we observe X at times $\{t_k, 0 \leq t_k \leq T, k = 0, 1, \dots, n\}$ and let $Y = \{Y_k = X_{t_k}\}$. For scalar diffusions, or for the diffusions considered in Chapter 2, we can write the likelihood using the transition density of the diffusion (conditional on the initial point Y_0):

$$p_\theta(Y) = \prod_{k=1}^n p_\theta(Y_k | Y_{k-1}). \quad (3.3)$$

²The 2-dimensional diffusion still has it though

However, this is not always true for stochastic volatility models and consequently for the diffusions in \mathcal{C} . This complicates things even further, making inference for stochastic volatility models particularly challenging.

Likelihood approaches can be either analytical (Aït-Sahalia (2002), Aït-Sahalia (2005)), or simulation based (Pedersen (1995), Durham and Gallant (2002)). They usually approximate the likelihood in a way so that the discretisation error may become arbitrarily small, although the methodology developed in Beskos et al. (2006b) succeeds exact inference in the sense that it allows only for Monte Carlo error. Nevertheless, all of these methods rely on the Markov property and therefore become hard to generalize to stochastic volatility models. For Beskos et al. (2006b) and Durham and Gallant (2002) this can be done by adopting a sequential framework, whereas such an extension is possible under the Aït-Sahalia (2005) framework only when the volatility can be inferred from other financial instruments, i.e. options. See Sections 3.7, 3.8 and Aït-Sahalia and Kimmel (2005) for more details.

Consequently, data augmentation schemes are extremely useful for stochastic volatility models in particular, providing one of the few options for inference. By imputing both the latent paths of α and the paths between observations on X , they provide a likelihood approximation that can become arbitrarily precise. As before, a finite number of points is imputed that should be large enough to limit the error due to discretisation. Given the augmented path, it is then straightforward to get the likelihood using appropriate Euler schemes for the transition density; see for instance Elerian et al. (2001), Eraker (2001) and Jones (2003).

3.3 Reparametrisation

3.3.1 The need for a reparametrisation

A reparametrisation is necessary for the same reasons stated in the Section 2.4.1. Under a data augmentation scheme we will impute some ‘missing’ points and then we will update

the parameters θ conditional on the augmented diffusion paths. Nevertheless, as the number of imputed points increases, the conditional posterior density of the volatility parameters becomes a point mass, at the value indicated by the quadratic variation process of the augmented path. Hence, any data augmentation scheme of this kind, will result in a reducible MCMC algorithm in the limit: As the number of imputed points gets larger, the sample paths will increasingly ‘force’ the parameters to remain at their current value causing serious implications to the convergence and mixing of the chain. This is particularly inconvenient as the imputation of arbitrarily many points is vital for limiting the approximation error.

The problem may be resolved if we apply a transformation so that the algorithm based on the transformed diffusion is no longer reducible. In the spirit of Papaspiliopoulos et al. (2003), this can be seen as a non-centering reparametrisation. As already mentioned, Roberts and Stramer (2001) tackle the problem for scalar diffusions by a reparametrisation on the paths of V and in the Chapter 2 we introduced an extension for some multivariate diffusions. To apply this reparametrisation, we seek a transformation of the diffusion $(X, \alpha)'$ to one with unit volatility matrix. This requires the solution of the system 2.2. However, using the condition (2.4), we note that such a transformation does not exist. In other words the framework of Chapter 2, does not cover the models in \mathbb{C} . This chapter focuses on this class and introduces a novel reparametrisation of the likelihood that may serve as the basis for irreducible data augmentation schemes.

3.3.2 A suitable likelihood parametrisation

As before, unless we apply a suitable transformation, the likelihood using Girsanov theorem is written with respect to a dominating measure that depends on parameters θ . Consider a diffusion in \mathcal{C} and denote the observations of the process X at times t_k , $k = 1 \dots, n$ by Y , $Y = \{X_{t_k}, 0 \leq t_k \leq T\} = \{Y_k, k = 1, \dots, n\}$. For simplicity we assume that the diffusion in (3.2) is two-dimensional (we discuss about extensions to higher dimensions in the next section). Note that it can also be written as:

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

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

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

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

where $\mathbb{P}_\theta(\alpha)$ and $\mathbb{P}_\theta(X|\alpha)$ denote the laws of α and X given α respectively. Note that given the path of the unobserved process and its parameters, α and W become deterministic functions of time. Looking again at the SDE for X we note that $\mathbb{P}_\theta(X|\alpha)$ depends only on the observations Y the density of which is

$$p_\theta(Y|\alpha) = \prod_{k=1}^n p_\theta(Y_k | Y_{k-1}, \{\alpha_t : t_{k-1} \leq t \leq t_k\}). \quad (3.4)$$

where

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

with

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

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

$$\frac{d\mathbb{P}_\theta}{d\mathbb{Q}_\theta}(X, \alpha) = p_\theta(Y|\alpha) \frac{d\mathbb{P}_\theta(\alpha)}{d\mathbb{Q}_\theta(\alpha)} \quad (3.5)$$

The dominating measure $\mathbb{Q}_\theta(\alpha)$ in the parametrisation of (3.5) corresponds to a diffusion with volatility $\sigma_\alpha(\alpha_t, \theta)$ and therefore clearly depends on θ . For this reason we introduce a transformation $\beta_t = h(\alpha_t, \theta)$ similar to (1.7):

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

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

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

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

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

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

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

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

$$\frac{d\mathbb{P}_\theta}{d\mathbb{Q}_\theta}(X, \gamma) = p_\theta(Y|\gamma) \frac{d\mathbb{P}_\theta(\gamma)}{d\mathbb{W}} = p_\theta(Y|\gamma) p(\alpha_0) G(\gamma, \theta) \quad (3.6)$$

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

3.4 Data augmentation scheme

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

Updating θ is relatively straightforward as we can get the relevant conditional posterior densities using the likelihood defined in (3.6). The parameter vector can be divided into two components $\theta = (\theta_1, \theta_2)$, depending on the part of likelihood they appear. θ_1 contains the parameter involved in the μ_x and σ_x as well as ρ and α_0 . θ_2 contains the parameters in $\mu_\alpha(\cdot)$ and $\sigma_\alpha(\cdot)$. Let $p(\theta_i)$, $i = 1, 2$ denote the corresponding priors. The conditional posterior densities $p(\theta_i|Y)$ will then be:

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

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

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

In the remainder of this section we will focus on the updates of the diffusion paths of γ , a clearly more complicated task. We proceed using an independence sampler, proposing for instance from Brownian motion. As in the case of the previous chapter, the updating of the entire path could lead to extremely low acceptance rates due to the great discrepancy between the proposed and current path. A way to circumvent this problem is to split the process into blocks and update each one of them in turn, by proposing diffusion bridges.

While it is rather clear that the path of γ should be divided into blocks, it is not straightforward how this should be done. Suppose that we observe Y at times t_k , as in Section 3.3.2, and that we split the path of γ into n blocks $\{b_k = \gamma_s, \quad t_{k-1} \leq s \leq t_k, \quad k = 1, 2, \dots, n\}$. But under this formulation the endpoints of the blocks are not updated at all leading to a reducible MCMC chain; an alternative blocking scheme is needed.

The blocking strategy adopted in this paper, uses block updating of overlapping segments. This scheme has been also used in Roberts and Stramer (2004) in a slightly different context. Under this procedure, we update γ_s for $t_i \leq s \leq t_{i+c}$ for $i = 0, 1, \dots, n - c$, where c is an integer smaller than n . It is our experience that high values of c improve the mixing of the algorithm as long as the acceptance rate of the blocks is not too small.

We use an independence sampler with a Brownian bridge as the proposal distribution. The MCMC algorithm becomes:

1. Set $i = 0$.
2. Propose a Brownian bridge starting at $\gamma(t_i)$ and finishing at $\gamma(t_{i+c})$.
Denote it by γ_s^* , $t_i \leq s \leq t_{i+c}$.

3. Accept γ_s^* with probability:

$$\min \left(1, \frac{G(\gamma_s^*, \theta) p_\theta(X_s | \gamma_s^*)}{G(\gamma_s, \theta) p_\theta(X_s | \gamma_s)} \right)$$

4. Set $i = i + 1$ (until $i \leq n - c$). Note that for $i = n - c$ the proposal should be just Brownian motion rather than a Brownian bridge.

An alternative blocking strategy uses random sized blocks. More specifically, at each iteration the path is randomly split into blocks and the paths between the endpoints are updated, whereas the endpoints remain unchanged until the next iteration. See Elerian (1999) for more details. This scheme was adopted in Chib et al. (2005) and found to perform well.

3.5 Extensions

3.5.1 Multifactor and Multivariate Stochastic Volatility Models

So far, for illustration processes, we made the assumption that both processes X and α are one-dimensional. This assumption is not necessary; we can still use such a likelihood parametrisation in models where either or both X_t and α are vector processes. Famous examples of such models are that of multivariate and multifactor stochastic volatility used for example in Duffie and Kan (1996) and Chernov et al. (2003).

We should keep in mind however, that this likelihood parametrisation requires the existence of a transformation of α , $h(\alpha, \theta)$, to a diffusion with identity volatility matrix. Therefore it is only applicable to models with such $\sigma_\alpha(\alpha, \theta)$, so that the diffusion α is *reducible* or in other words if the following system of PDEs:

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

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

3.5.2 Applications beyond stochastic volatility

The methodology developed so far does not cover only stochastic volatility models. In fact, it can be applied to general partially observed diffusion settings in a state space framework. The likelihood can be specified in a similar manner with Section 3.3.2. One starts with the marginal distribution of the latent path which is provided by Girsanov's formula. The likelihood can then be completed by some appropriate density, like (3.4), for the observations conditional on any feature of the latent path.

One such example concerns diffusions observed with error. In this case, the density for the observation y at time t could be for instance gaussian with the value of the latent process (or some function of it) as the mean and some variance. Another example taken from finance is when we observe functions of the rates or equity prices. Under a typical scenario we have the bid and ask quotes but we are interested in a latent process that lies between these two values and can be thought of as the 'fair' or 'true' price. Again we can specify the likelihood assigning an appropriate distribution on some functional of these quotes that may depend on any element of the latent path of the unobserved 'true' price.

On some other occasions, the distribution of the observations could be a Poisson distribution, the rate of which could follow a diffusion process; see Kou et al. (2006) for an interesting application in genetics. In the same context, partially observed diffusions are also used in survival data analysis where they appear in the hazard function; see Aalen and Gjessing (2001), Aalen and Gjessing (2004) and Sangalli and Roberts (2006).

3.6 Simulations

The simulations performed in this section aim to demonstrate two aspects of the problem. First, we highlight the necessity of the reparametrisation introduced in Section 3.3. This is done by initially exposing the problem in the case of a very simple stochastic volatility model. Then we present the proposed methodology using simulated data from the Heston model.

3.6.1 Data augmentation without reparametrisation

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

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

where B and W are independent Brownian motions.

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

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

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

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

We ran a MCMC chain for 100,000 iterations for different numbers of imputed points ($m = 1, 10, 40, 100$), updating the paths as described in Section 3.4 and using the Gibbs step for the updates σ^2 . We used the overlapping blocks scheme for the updates of the paths. The length of the block was chosen as $c = 5$ to improve autocorrelation for the posterior draws of σ^2 and the acceptance rate of the blocks was approximately 78%. Figure 3.1b shows the autocorrelation of the posterior draws of σ^2 for each value of m . Clearly, the autocorrelation increases dramatically leading to an increasingly slower chain. An alternative way to see this is to note that the variance of the conditional posterior for σ^2 goes to 0 as we increase m .

The problem can be resolved if we apply this paper's proposed reparametrisation. Following the route of Section 3.3, we set $\beta_t = \alpha_t/\sigma$ and $\gamma_t = \beta_t - \beta_0 = \beta_t$ and we obtain

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

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

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

where

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

Figure 3.1a contains the corresponding autocorrelation plots of the posterior draws of σ^2 taken from the reparametrised data augmentation scheme. Unlike the previous case (Figure 3.1b) there is clearly no increase in the autocorrelation.

3.6.2 Reparametrised scheme applied to the Heston model

In this section we illustrate the proposed methodology to simulated data from the model of Heston (1993). We also demonstrate the immunity of the algorithm developed in Sections 3.3 and 3.4 to the increase of the number of imputed points m . If we take the log of the observed process the Heston model can be written as:

$$\begin{aligned} dX_t &= \mu_x dt + \{(1 - \rho^2)\alpha_t\}^{1/2} dB_t + \rho \alpha_t^{1/2} dW_t \\ d\alpha_t &= \kappa(\mu - \alpha_t)dt + \sigma\alpha_t^{1/2}dW_t \end{aligned}$$

where B_t and W_t are independent Brownian motions and $\text{Corr}(dX_t, d\alpha_t) = \rho dt$ as before. In line with Section 3.2.2 we apply the following 2-step transformation:

1. $\beta_t = h(\alpha_t) = 2\alpha_t^{1/2}/\sigma, \beta_t > 0$
2. $\gamma_t = \beta_t - 2\alpha_0^{1/2}/\sigma, \beta_t = \eta(\gamma_t)$

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

$$\begin{aligned} dX_t &= \mu_x dt + \{(1 - \rho^2)g(\gamma_t)\}^{1/2} dB_t + \rho g(\gamma_t)^{1/2} dW_t \\ d\gamma_t &= \left\{ \frac{2\kappa\mu - 0.5\sigma^2}{\sigma^2\eta(\gamma_t)} - 0.5\kappa\eta(\gamma_t) \right\} dt + dW_t \end{aligned}$$

We can now proceed by writing down the likelihood as in Section 3.3 and implementing a data augmentation scheme as in Section 3.4. Using Brownian bridges as proposals for the paths of γ is not the best choice given the constraint $\beta_t > 0$. Alternatively we may choose to update the paths of α instead, since they are linked with a deterministic function with the paths of γ given σ . We may propose from the diffusion Z that satisfies the SDE:

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

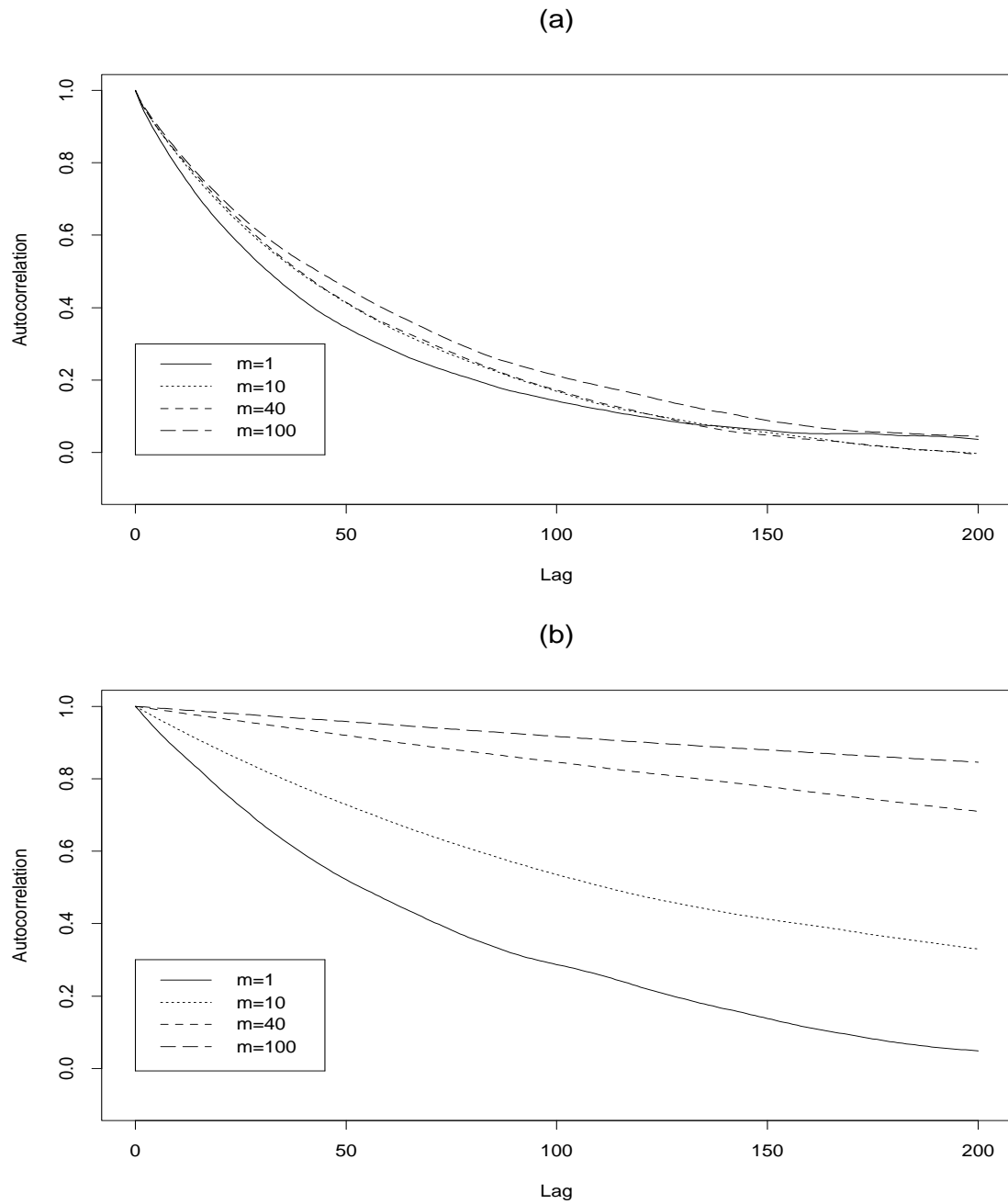


Figure 3.1: Autocorrelation plots of posterior draws of σ^2 for different values of imputed points between observations (m) for the simple stochastic volatility model. The draws in (a) correspond to the reparametrised scheme and in (b) to the scheme without transformation.

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

Table 3.1: Posterior means and standard deviations of the parameters versus their true values for the Heston model.

To simulate bridges from Z we can first simulate a Brownian bridge BB_t with volatility $\sqrt{\sigma/2}$ and then set $Z = \frac{1}{2}\sigma BB_t^2$. Regarding the updates of the parameter κ and μ , it turns out that they are almost Gibbs steps. Hence, we can use ‘clever’ proposals which are derived in the appendix.

The parameter values used in the simulation are similar to those obtained from the analysis of the closing prices of Standard and Poor’ 500 index in Chib et al. (2005) and Aït-Sahalia and Kimmel (2005). We simulated 1008 data points (excluding the initial point) from the Heston model corresponding to 1008 working days or 4 years of data. In accordance with the relevant literature, we set the initial values to $X_0 = \log(100)$ and $\alpha_0 = \mu = 0.9$. As before, we ran a MCMC chain for 80,000 iterations for different numbers of imputed points ($m = 20, 40, 80$). We chose the value of $c = 8$ to achieve lower autocorrelation on the parameter posterior draws. The acceptance rate for the blocks was approximately 32%.

Figure 3.1 shows the autocorrelation plot for the posterior draws of σ for different values of m . There is no evidence of any increase whatsoever, even for $m = 80$. In Figure 3.2, we see the posterior densities of the log-likelihood and σ , again for different values of m . All the densities are similar, providing strong evidence that the discretisation is sufficiently fine. Finally, Table 3.1 provides the posterior means and standard deviations of the parameters. We see that these estimates are in good agreement with the values we simulated the data from.

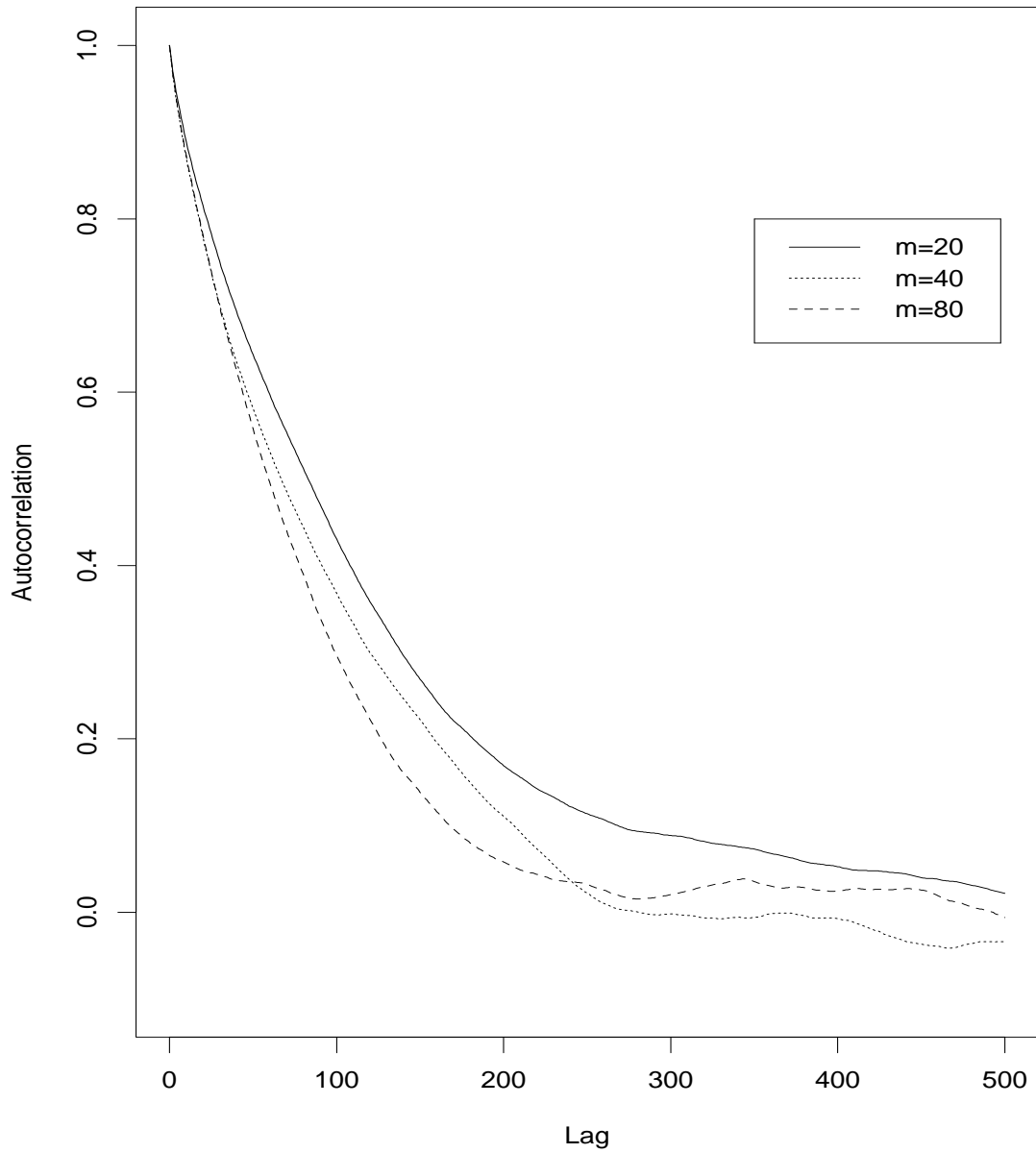


Figure 3.2: Autocorrelation plots of posterior draws of σ for different values of imputed points between observations (m) for the Heston model.

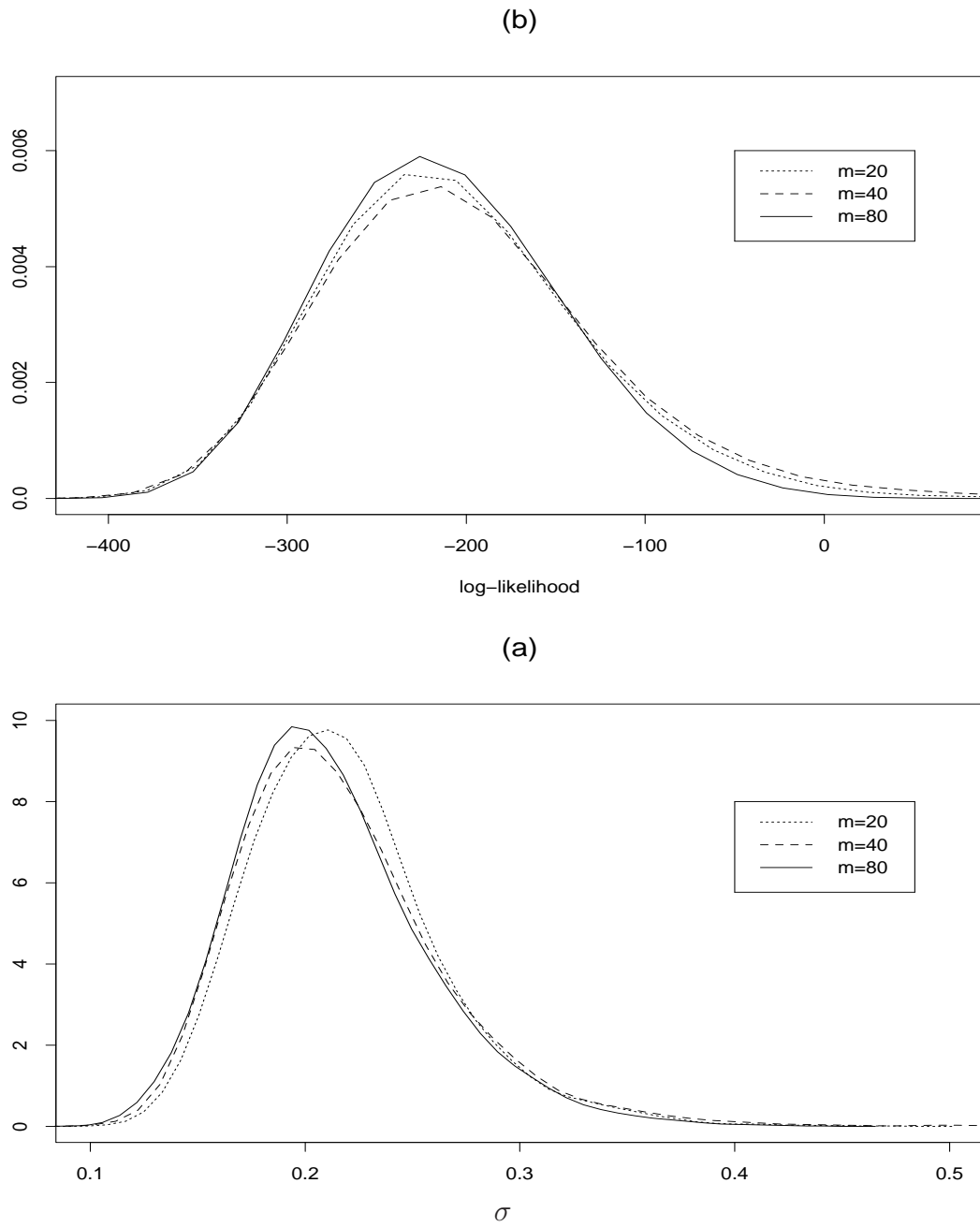


Figure 3.3: Posterior densities of (a) log-likelihood and (b) σ for different values of imputed points between observations (m) for the Heston model.

3.7 Connection with option pricing

The relation between option pricing and statistical inference can go in both directions. In some cases we need to estimate the parameters of the option pricing formula, for example the volatility in the Black-Scholes formula. On the other hand we may use the information contained in option prices to carry inference for the model parameters. In this section we provide a brief description of the option pricing problem and we describe how these ideas apply to our case.

3.7.1 Option pricing under stochastic volatility

Denote the log-price of an asset at time t by S_t . A European call/put option on the asset gives the right and not the obligation, to buy/sell the asset at a specific time $t + h$ (maturity) and at a specific price K (exercise or strike price). Assume that the dynamics of the asset process are determined by the following stochastic volatility model.

$$\begin{aligned}dS_t &= S_t \{ \mu_S^{\mathbb{P}}(\alpha_t, \theta_S, t) dt + \sigma_S(\alpha_t, \theta_S) dW_t^P \} \\d\alpha_t &= \mu_\alpha^{\mathbb{P}}(\alpha_t, \theta_\alpha, t) dt + \sigma_\alpha(\alpha_t, \theta_\alpha) dB_t^P\end{aligned}$$

where (W^S, W^U) are standard Brownian motions with correlation ρ . If we take $X_t = \log(S_t)$ it is not difficult to check that the model belongs in the class \mathcal{C} . The distribution of S , denoted by \mathbb{P} , is often called historical or physical measure. Under some fairly general assumptions, there exists a family of equivalent martingale measures \mathbb{Q} that can be used for finding a price for the option in a way that eliminates the possibility of arbitrage. Assuming a zero risk free rate, a fair price for the call option is just the the following expectation (with respect to \mathbb{Q}):

$$C_t = E^{\mathbb{Q}} [(S_t - K)^+ | \mathcal{F}_t]$$

Every such measure is called a risk neutral measure or a pricing measure. The existence of such measures is guaranteed if the Radon-Nikodym derivative of \mathbb{P} with respect to \mathbb{Q} has finite variance or alternatively by some boundedness restrictions; see Harrison and Krepps (1979) and Harrison and Krepps (1981) for more details. The dynamics of S under \mathbb{Q} would satisfy (again assuming zero risk free rate):

$$dS_t = S_t \sigma_S(\alpha_t, \theta_S) dW_t^{\mathbb{Q}}$$

$$d\alpha_t = \mu_{\alpha}^{\mathbb{Q}}(\alpha_t, \theta_{\alpha}) dt + \sigma_{\alpha}(\alpha_t, \theta_{\alpha}) dB_t^{\mathbb{Q}}$$

If the volatility of the process is independent of α (and therefore only the first of the equations above is relevant), then \mathbb{Q} is unique and the option pricing problem reduces on finding an appropriate value θ_S . Under such models the market is said to be *complete*. In our case however, an entire family of risk neutral measures \mathbb{Q} exists, each of which may lead to completely different prices; stochastic volatility models usually represent an *incomplete* market. The choice of \mathbb{Q} is not trivial. Henderson et al. (2005) prove that option prices with convex payoff structures are decreasing in the market price of volatility risk. This is in line with results of Bergman et al. (1996), Romano and Touzi (1997), Hobson (1993) and El Karoui et al. (1998) which show that the option price is increasing in volatility.

An rational idea is to choose the \mathbb{Q} which is ‘closer’ to \mathbb{P} (and therefore to reality in some sense). In an attempt to minimize the distance between the physical and pricing measure, Henderson et al. (2005) examine the class of q -optimal measures $H_q(\mathbb{P}, \mathbb{Q})$, where each q reflects a different measure distance:

$$H_q(\mathbb{P}, \mathbb{Q}) = \begin{cases} E \left[\frac{q}{q-1} \left(\frac{d\mathbb{Q}}{d\mathbb{P}} \right)^q \right], & \text{if } q \in \mathbb{R} \setminus \{0, 1\} \\ E \left[(-1)^{1+q} \left(\frac{d\mathbb{Q}}{d\mathbb{P}} \right)^q \log \left(\frac{d\mathbb{Q}}{d\mathbb{P}} \right) \right], & \text{if } q \in \{0, 1\} \end{cases}$$

The situation remains obscure however, as Henderson et al. (2005) also prove that option prices are decreasing in q . Nevertheless, things may be simplified if we assume

deterministic $\mu_S^{\mathbb{P}}(\cdot)$ which corresponds to a deterministic Sharpe ratio (mean over standard deviation) or equity risk premium and reflects an *almost complete* market. In this case all the q -optimal measures collapse to the case of $q = 0$, the minimal martingale measure, which also minimizes the reverse relative entropy; see Föllmer and Schweizer (1991) and Schweizer (1999). Under this measure we have $\mu_\alpha^{\mathbb{Q}}(\cdot) = \mu_\alpha^{\mathbb{P}}(\cdot)$.

3.7.2 Determining the option pricing parameters

Existing approaches for choosing the parameters of the pricing measure \mathbb{Q} we may be classified in two main categories. Under the first approach one ignores completely the real world measure \mathbb{P} and only defines \mathbb{Q} . Then, calibration methods may be used for the parameters, so that some sort of distance between model and observed option prices is minimized. This procedure is widely used as most of the times achieves a very good matching with of the option datasets. Nevertheless, the estimates may become unstable and there exist overfitting concerns. Moreover, as confirmed by empirical studies (Dumas et al., 1998), the joint stock - option prices implied from the model is not consistent with the observed price process. To overcome this problem such models require periodic recalibration. It is not straightforward to define the likelihood for the parameters and therefore the results of this chapter are not relevant with this approach.

An alternative procedure is to use both historical and pricing measures \mathbb{P} and \mathbb{Q} as in Aït-Sahalia and Kimmel (2005) and Chernov and Ghysels (2000). The main challenge of this approach is to define an appropriately link between them, which essentially translates in connecting $\mu_\alpha^{\mathbb{P}}(\cdot)$ and $\mu_\alpha^{\mathbb{Q}}(\cdot)$ with an appropriate function, say $g(\cdot)$. Once this is done, a way to go is by defining a likelihood based on \mathbb{P} and then obtain estimates for the \mathbb{Q} parameters via the transformation $g(\cdot)$. This approach has the appealing feature that it is possible to incorporate information from observed option prices, apart from the information contained in the price process, in an attempt to match the characteristic of both series (see next section for more details). On the other hand, the specification of the link between the real world and risk neutral measure is not straightforward. The problem is essentially the same as choosing \mathbb{Q} as discussed in the previous section and very often

the input of financial market experts is needed. In the application of Section 3.8, we simplified things by adopting a model that corresponds to an *almost complete* market. Doing so, a reasonable choice for \mathbb{Q} is the unique, in this case, q -optimal measure or in other words the minimal martingale measure with $\mu_\alpha^{\mathbb{Q}}(\cdot) \equiv \mu_\alpha^{\mathbb{P}}(\cdot)$.

The latter procedure fits very naturally with our framework. Having obtained samples from the posterior of \mathbb{P} parameters, we can just transform them to perform inference on those of \mathbb{Q} . Moreover, under the Bayesian setting it is straightforward to incorporate parameter, and potentially model (which may be viewed as another parameter), uncertainty using their posterior distribution (Kass and Raftery, 1995). The problem can then be included in a decision theoretic framework using appropriate utility functions for pricing options.

3.7.3 Using option prices for inference purposes

Obtaining α_t

So far, we addressed the problem of parameter estimation based on an unobserved volatility path. In financial applications however, apart from the equity prices, there exist large datasets of option prices that have been made upon them. As we demonstrate in this section, it is possible to infer all the corresponding points of the volatility path using these data. Then we can either treat this points as proper observations or do some adjustment as we discuss towards the end of this section.

Let us go back to the Black-Scholes model. Assuming that the process follows a geometric Brownian motion, and given the value of the price at time t , we can provide a ‘fair’ price for any type of option with any maturity h and strike price K . Conversely, for any option price (with a specific maturity, strike price and price value at time t , S_t), there exists a unique positive number which would have given the specific option price had it been the volatility in a Black-Scholes formula. This number is called (Black-Scholes) implied volatility, is denoted by $\sigma^{imp}(t, t+h)$ and is given by the solution of the following equation (again assuming zero interest rate).

$$\begin{aligned}
C_t &= S_t [\phi(d_1) - e^{-x_t} \phi(d_2)] \\
d_1 &= \frac{x_t}{\sigma^{imp}(t, t+h)\sqrt{h}} + \frac{\sigma^{imp}(t, t+h)\sqrt{h}}{2} \\
d_2 &= d_1 - \sigma^{imp}(t, t+h)\sqrt{h} \\
x_t &= \frac{\log S_t}{K}
\end{aligned}$$

Black-Scholes implied volatility is particularly useful even when the model assumes stochastic volatility $\sigma_s(\alpha_t)$. For small maturities h , it can be interpreted as implied average volatility (Ghysels et al., 1996):

$$\sigma^{imp}(t, t+h) \approx E_t \left[\frac{1}{h} \int_t^{t+h} \sigma_s(\alpha_t) ds \right]$$

Following these ideas it has become common practice in finance to construct volatility proxies by averaging many implied volatilities which are as close as possible to be ‘at the money’ and with short maturities. In Section 3.8 we analyze the volatility index (VIX) of the S&P 500 index which is provided by Chicago Board of Exchange and is constructed a similar algorithm, see Whaley (1993) and Whaley (2000) for more details. Having obtained the volatility of S by σ_t^{imp} , it is then straightforward to get α .

MCMC implementation

By setting $X_t = \log(S_t)$ and using Ito’s lemma, we obtain a model belonging to \mathcal{C} . The difference is that now we also have observations on α for each point of X . In other words, we have an at least 2-dimensional discretely (fully) observed diffusion as in Chapter 2. Nevertheless, we cannot apply the corresponding methodology because the diffusion is not *reducible* as mentioned earlier. Therefore we proceed by writing the likelihood as in this chapter, separating the bits of α and X given α . Regarding the marginal likelihood for α , we need apply the reparametrisation of Roberts and Stramer (2001) for univariate α , or its extension developed in Chapter 2 for the multivariate case, in order to construct

an irreducible MCMC algorithm. Note that this assumption is applied only on α and therefore we require only its reducibility (the joint diffusion is not reducible).

More specifically, β_t is defined in the same way as before and for each pair of successive observations, say times t_k and t_{k+1} , we set γ to be:

$$\gamma_t = \beta_t - \frac{(t_{k+1} - t)\beta_{t_k} + (t - t_k)\beta_{t_{k+1}}}{t_{k+1} - t_k} \quad (3.7)$$

Using Ito's lemma we can obtain the SDE for γ and apply the relevant parametrisation of Chapter 2. Doing so, we can then implement a similar MCMC scheme to that of Section 3.4. The only difference is in the updates of γ . Now there is no need to adopt neither a random sized nor an overlapping blocks scheme. We can just update the bridges between observations one by one using an independence sampler in a similar way as in Roberts and Stramer (2001) or Chapter 2 accordingly. See Section 3.8 for an application based on such an algorithm.

Adjusting the volatility proxy

Even though we use Black-Scholes implied volatility from options with short maturities, the approximation error due to the false assumption of constant volatility may not be negligible. The integrated volatility proxy of Ait-Sahalia and Kimmel (2005) corrects for the effect of mean reversion in volatility and in some cases provides an improvement. Denote the integral of the variance of X by $V(t, h)$, which is defined by

$$V(t, h) = \frac{1}{h} \int_t^{t+h} \sigma_S(\alpha_s) ds$$

and denote by $V^{imp}(t, h)$ the corresponding quantity obtained by the Black-Scholes implied volatility. To construct the volatility proxy we simply set $V^{imp}(t, h) = V(t, h)$, which can be seen as a rough averaging. This can be corrected in the case of a linear drift for $\sigma_S(\alpha_t)$, of the form $\theta_0 - \theta_1 \sigma_S(\alpha_t)$, under the pricing measure \mathbb{Q} . Also it provides an

improvement for any other \mathbb{Q} -drift that can be approximated better with a linear drift rather than no drift at all. The expected value of $V(t, h)$ is:

$$E[V(t, h)] = \frac{e^{\theta_1 h} - 1}{\theta_1} \left[\sigma_S(\alpha_t) + \frac{\theta_0}{\theta_1} \right] - \frac{\theta_0}{\theta_1}$$

Now if we set $E[V(t, h)] = V^{imp}(t, h)$, we get

$$\sigma_S(\alpha_t) = \frac{h\theta_1 V^{imp}(t, h) + \theta_0 h}{e^{\theta_1 h} - 1} - \frac{\theta_0}{\theta_1}$$

The parameters θ_0, θ_1 may be either estimated beforehand or incorporated in the MCMC scheme given their link with the corresponding \mathbb{P} ones.

Alternatively, or additionally, we can assume that the volatility proxy is just a noisy observation of the true volatility. For instance we may assume that:

$$\sigma_S^{imp}(\alpha_t) \sim N \{ \sigma_S(\alpha_t), \omega^2 \}$$

for some suitable noise variance ω^2 . Such an approach may be implemented with a data augmentation similar to that of Section 3.4. The likelihood will be slightly modified with the addition of the noise density for each extra observation $\sigma_S^{imp}(\alpha_{t_k}), k = 1, \dots, n$.

3.8 Application: S&P 500 - VIX data

3.8.1 Data description

The dataset consists of the closing values of the S&P 500 index provided by CBOE. CBOE also provides a very accurate volatility proxy, VIX. This is computed from European options with varying strike prices and maturity 30 calendar days. We used daily data

from the 2nd of January 1990 to the 30th of September 2003. This dataset has been used in many papers including Aït-Sahalia and Kimmel (2005), Chib et al. (2005) and Roberts and Stramer (2004). The data are plotted in Figure 3.4.

3.8.2 The models

We assigned the following model to the price S under the real world measure P .

$$dS_t = S_t \mu_S \alpha_t dt + S_t \alpha_t dB_t$$

This model reflects an *almost complete* market (Henderson et al., 2005). All the risk neutral measures \mathbb{Q} belonging to the family of q -optimal measures collapse to the minimal martingale measure under which $\mu_\alpha^{\mathbb{Q}}(\cdot) \equiv \mu_\alpha^P(\cdot)$. The choice of this measure can be justified as the one which is ‘closest’ to \mathbb{P} which represents the real world.

For the dynamics of the volatility process α_t , we considered 3 different models:

1. The Heston model (Heston, 1993):

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

2. The diffusion GARCH model (Andersen et al., 2006):

$$d\alpha_t = \kappa(\mu - \alpha_t)dt + \sigma\alpha_t dW_t$$

3. A log-normal volatility model under which the log volatility follows an Ornstein Uhlenbeck process:

$$d\log(\alpha_t) = \kappa(\mu - \log(\alpha_t))dt + \sigma dW_t$$

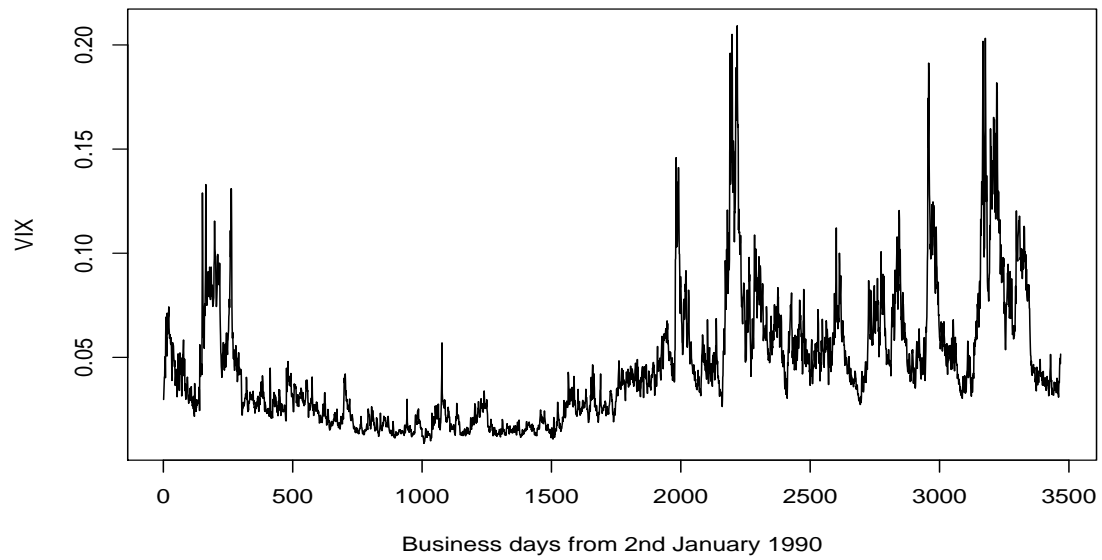
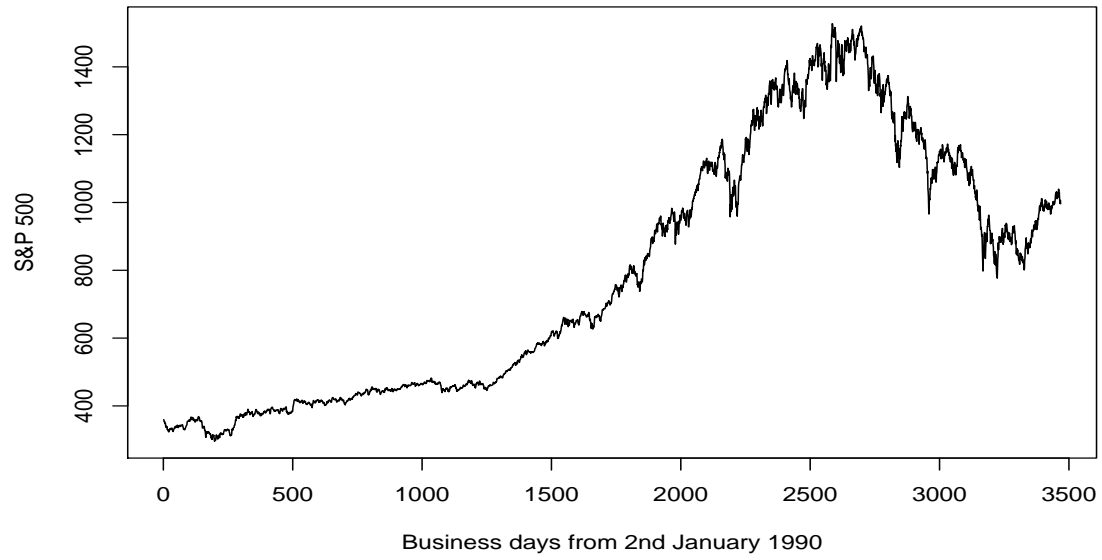


Figure 3.4: Standard & Poor 500 values (up) and its volatility index (down).

We also assumed a correlation ρ between the increments of the log-price $X_t = \log(S_t)$ and the volatility α (or $\log(\alpha)$ for the log-normal model), to capture the so called *leverage* effect. With our current formulation the Brownian motions B and W are correlated. alternatively if we use the following parametrisation for the SDE of X , they will be independent.

$$dX_t = (\mu_S \alpha_t + \frac{1}{4} \alpha_t^2) dt + \sqrt{1 - \rho^2} \alpha_t dB_t + \rho \alpha_t dW_t$$

In the appendix we provide the relevant transformations for these 3 models. Also as it turns out, the updates of κ and μ are almost Gibbs steps, which means that we construct efficient proposals for their updates. These are model specific and are also given in the appendix.

3.8.3 Results

A data augmentation scheme, suitable for stochastic volatility models with a volatility proxy, was applied for each model. The time was measured in calendar years, and we assumed equidistant observation times even between weekends. Parameters were updated either by random walk metropolis, or via ‘clever’ proposals as described in the appendix. The proposals for the paths of the transformed diffusion γ were Brownian bridges and the acceptance rate was particularly high in all cases (close to 95%). As before, different numbers of imputed points were used to assess the likelihood convergence over the resulting discretisations. Figure 3.5 plots kernel densities of posterior draws of the log likelihood for $m = 20, 40, 60$ for each model. In all cases the densities plots look similar indicating that the partitions of the augmented paths were reasonably fine to allow accurate likelihood approximations.

Next, we perform a check for increase in the autocorrelation of the posterior draws belonging to the parameters of the 2– dimensional diffusion’s volatility, σ and ρ . As we can see

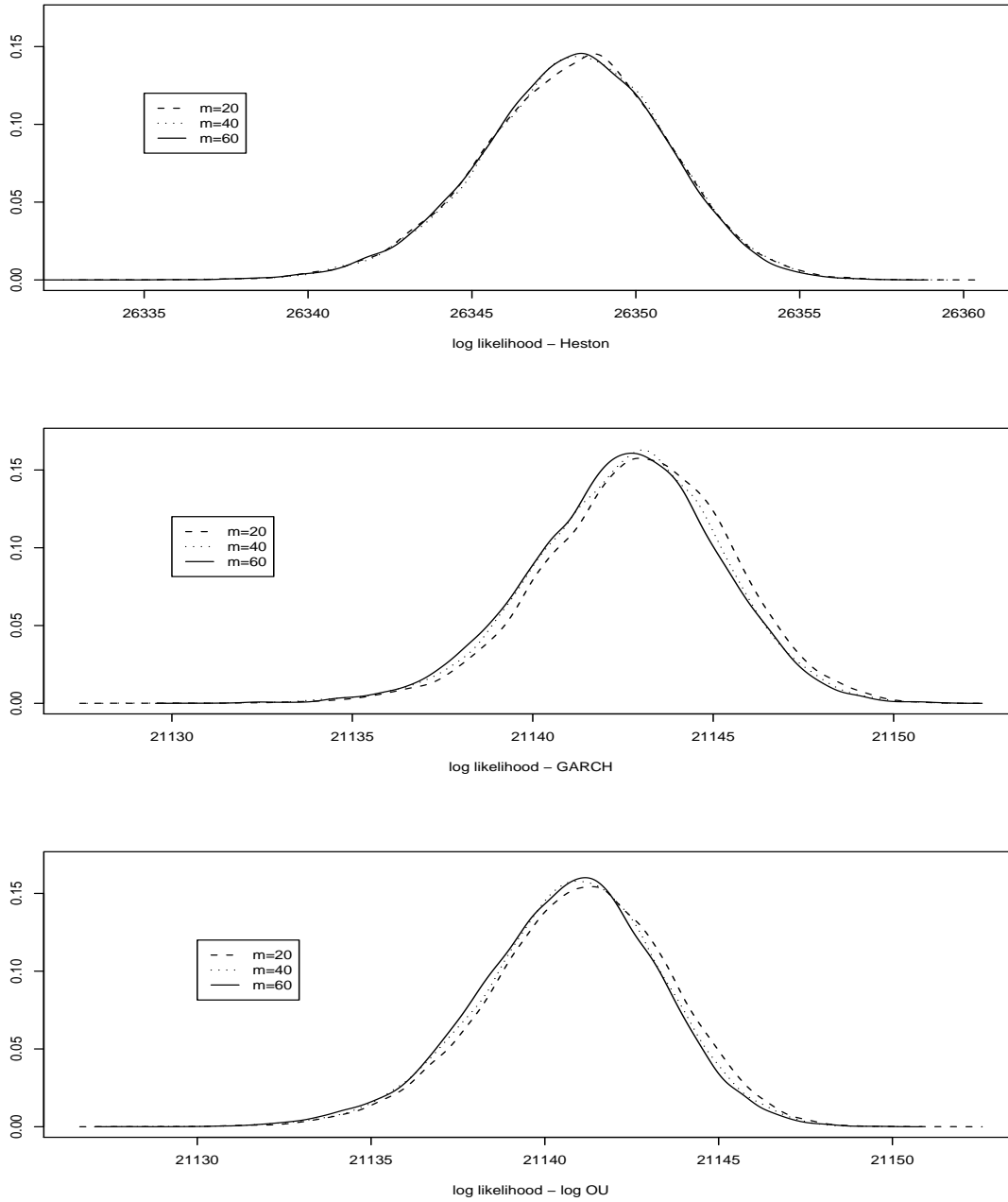


Figure 3.5: Kernel density plots of the log-likelihood from the Heston (top), GARCH (middle) and log OU (bottom) models for the SP500/VIX data.

Parameter	Heston model		GARCH model		log OU model	
	Post. mean	Post. SD	Post. mean	Post. SD	Post. mean	Post. SD
κ	4.558	0.751	2.791	0.695	3.969	0.608
μ	0.066	0.009	0.057	0.012	-3.286	0.149
σ	0.473	0.006	2.103	0.027	2.100	0.026
ρ	-0.766	0.007	-0.761	0.007	-0.761	0.007
μ_S	0.085	0.272	0.261	0.230	0.246	0.274

Table 3.2: Posterior means and standard deviations of parameters from the Heston, GARCH and log OU models for the SP500/VIX data.

from Figure 3.6 there is no evidence to warrant any suspicion regarding the reducibility in the limit of the chain.

The posterior samples of the model parameters are contained in Table 3.2 for the Heston, GARCH and log OU model respectively. The mean μ , persistence of reversion towards the mean κ and the volatility σ of the volatility (or log-volatility in the log OU model) take different values in each case reflecting the different model formulations. The correlation of the instantaneous log returns and volatility increments ρ is around -0.76 in all models, implying a strong *leverage* effect. Finally the parameter μ_S , which as previously mentioned represents the Sharpe ratio, is around 0 in all models.

As argued earlier, a rational choice for a ‘fair’ pricing measure \mathbb{Q} would be one with zero drift for X^3 and the \mathbb{P} SDE for α . Hence, the parameter estimates of κ , μ , σ and ρ may be used for option pricing purposes. Moreover, their posterior draws may be used in a suitable way to incorporate the uncertainty associated with them.

3.9 Concluding remarks

The methodology developed in this chapter allows for likelihood-based inference for most stochastic volatility models which are famous due to their massive use in finance. Generally it applies to a class of partially observed diffusions where the problem of inference is particularly difficult due to the lack of the Markov property. This is accomplished

³Under the assumption of zero risk-free rate

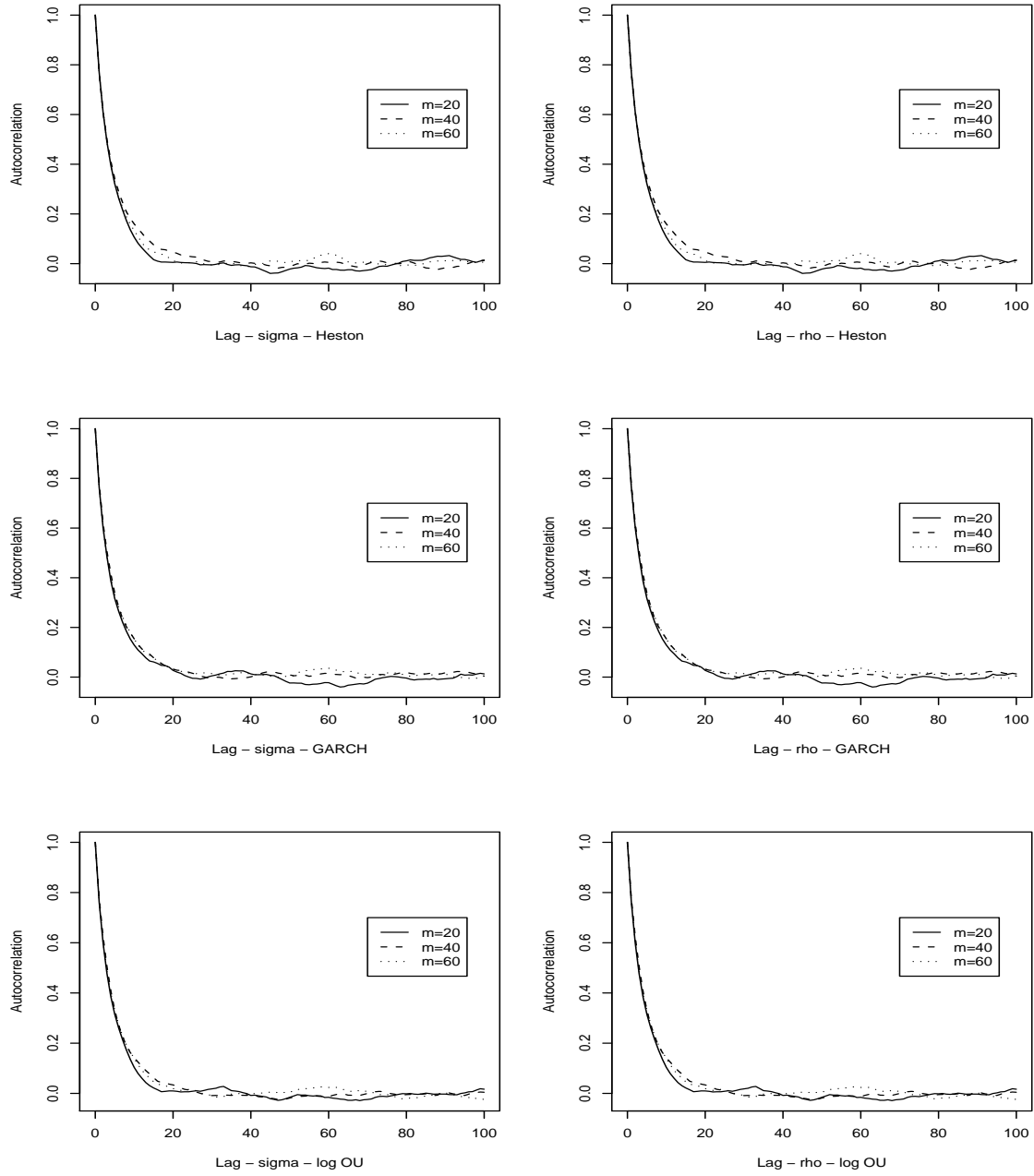


Figure 3.6: Autocorrelation plots of the posterior draws of σ (left) and ρ (right) of the Heston (top), GARCH (middle) and log OU (bottom) models for the SP500/VIX data.

via a MCMC algorithm that is relatively simple and fast to implement as illustrated in Sections 3.3 and 3.4. Under the proposed framework it is quite straightforward to limit the discretisation error by simply increasing the number of imputed points. The reparametrised scheme of this paper achieves that leaving both the autocorrelation of the posterior draws and the irreducibility of the chain, intact.

Furthermore, as discussed in Section 3.7 and illustrated in Section 3.8, the methodology of this chapter fits very well with issues related with stochastic volatility models in financial applications. It may be used to provide inputs to pricing formulae and at the same time it is possible to use information contained in relevant option price through a unified framework. The Bayesian setting can itself prove useful as it allows to incorporate model and parameter uncertainty to the option pricing problem.

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

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

The class of diffusions considered in this chapter does not include cases, where the drift and the volatility of the observed process depend on the process itself; for instance stochastic volatility models for interest rates with mean reverting drift. The reason for this is that the distribution of the observed process given the unobserved is no longer available in closed form. The data augmentation scheme of chapter 4 handle such models

using time change transformations of the observed process. An alternative option is provided by Golightly and Wilkinson (2005), but in a sequential setting only.

Chapter 4

Bayesian Inference for General Stochastic Volatility Models Using Time Change Transformations

4.1 Introduction

The class of diffusion driven stochastic volatility models studied in this thesis are described or can be transformed to the following SDE:

$$\begin{pmatrix} dX_t \\ d\alpha_t \end{pmatrix} = \begin{pmatrix} \mu_x(X_t, \alpha_t, \theta) \\ \mu_\alpha(\alpha_t, \theta) \end{pmatrix} dt + \begin{pmatrix} \sigma_x(\alpha_t, \theta) & 0 \\ 0 & \sigma_\alpha(\alpha_t, \theta) \end{pmatrix} \begin{pmatrix} dB_t \\ dW_t \end{pmatrix}, \quad (4.1)$$

where X is the observed process whose volatility is driven by the latent diffusion α and B, W denote standard Brownian motions. In the previous chapter, we introduced appropriate likelihood reparametrisations which are essential in the construction of irreducible MCMC schemes for stochastic volatility models. Nevertheless, all of these reparametrisations refer to diffusions with $\mu_x(X_t, \alpha_t, \theta) \equiv \mu_x(\alpha_t, \theta)$. Although this class is rich enough to include many of the stochastic volatility models used in practice, it does exclude some interesting cases. In this section we introduce a methodology that does not require this

assumption. Unlike the previous chapters, we use transformations that operate on the time scale of the diffusion rather than its path.

Apart from the flexibility of an enriched modelling framework, an additional motivation originates from econometric and financial applications concerning the modelling of the short term interest rate. The rate of interest over a short period of time is a fundamental economic and financial variable. Its behavior across time is strongly related with goods, services and wealth constituting a crucial figure for the economy in general. Also, as it affects the term structure directly, it has implications to the pricing of fixed income securities and derivatives. The finance literature normally treats the short term interest rate as a diffusion. For example under the famous (CIR) model of Cox et al. (1985), its dynamics are described by the following SDE:

$$dr_t = \kappa(\mu - r_t)dt + \sigma_r r_t^{1/2} dB_t.$$

Chan, K. C. et al. (1992) consider a generalization of the CIR model by setting the volatility as $\sigma_r r_t^\psi$, $\psi > 0$. The parameter ψ is often referred to as ‘elasticity of variance’. Note that for $\psi = 0$ we get the Vasicek model (Vasicek, 1977). Aït-Sahalia (1996b) notes evidence against linearity and proposes an alternative polynomial drift specification. However, this may be attributed to the inadequacy of a scalar diffusion model and therefore stochastic volatility is introduced: Andersen and Lund (1998), Gallant and Tauchen (1998), Durham (2002), Eraker (2001) etc. Their proposed stochastic volatility models can be summarized in the following SDE:

$$dr_t = \mu_\alpha(r_t, \alpha_t, \theta)dt + \sigma_r r_t^\psi \exp(\alpha_t/2)dB_t,$$

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

Estimation methods for these models usually include EMM, see for example Andersen and Lund (1998), Gallant and Tauchen (1998); or sequential Monte Carlo techniques, Durham and Gallant (2002), Golightly and Wilkinson (2005). For MCMC implementations see also Eraker (2001), Jones (2003) and Golightly and Wilkinson (2005).

Note that the models above fall belong to the diffusions with SDE as in (3.1) if we apply the transformation:

$$X_t = \begin{cases} r_t^{1-\psi}/\sigma_r(1-\psi), & \text{if } \psi \neq 1 \\ \log(r_t)/\sigma_r, & \text{if } \psi = 1 \end{cases}$$

Consequently, the methodology of this chapter is applicable to all of these formulations; see Section 4.5.3 for more details. The chapter is organized as follows: Section 4.2 summarizes the available likelihood based estimation techniques for stochastic volatility models and elaborates on the need for time change methodology. The time change reparametrisation is first demonstrated for univariate diffusions in Section 4.3 with some details on its MCMC implementation being provided in Section 4.4. However, its main appeal is its generalize-ability to stochastic volatility models of (3.1) and therefore Section 4.5 provides the details of its adaptation to them. Sections 4.6 and 4.7 contain two examples on simulated and real data respectively and Section 4.8 concludes.

4.2 The need for a time change transformation

Let us consider the general class of stochastic volatility models of this chapter defined by (3.1). In practice we observe X at a finite set of times and we want to draw inferences for the parameter vector θ based on these observations. Unfortunately, under the model above the observed process X_t is not Markovian in general. This rules out most of the likelihood based inference techniques that rely on the Markov property of the observed process; see Beskos et al. (2006b), Aït-Sahalia (2002) and Aït-Sahalia (2005), Pedersen (1995), Brandt and Santa Clara (2001) etc. The options for likelihood based inference are thus limited. Most of the remaining ones adopt a sequential framework which complicates the task of parameter estimation; Durham and Gallant (2002), Golightly and Wilkinson (2005).

Therefore, a data augmentation scheme becomes particularly valuable for stochastic volatility, since it provides a satisfactory solution to the problem through a unified

framework and does not rely on the Markov property. See Eraker (2001), Jones (1999) and Jones (2003) for relevant implementations. However, as noted in this thesis and in Roberts and Stramer (2001), any such scheme may result in a reducible MCMC algorithm unless appropriate reparametrisation is applied. For some diffusion models it is possible to tackle this problem using a 2-step transformation that enables us to write down the likelihood with respect to a parameter-free dominating measure as was done in Roberts and Stramer (2001) and Chapter 2 of this thesis. Nevertheless, such a transformation does not exist for stochastic volatility models since the corresponding diffusion is not *reducible* (Aït-Sahalia, 2005).

In the previous chapter we made the assumption $\mu_x(X_t, \alpha_t, \theta) \equiv \mu_x(\alpha_t, \theta)$, which is also necessary to the methodology of Golightly and Wilkinson (2005). Under this assumption, the marginal density of the observations given the volatility path is available in closed form as stated in Section 3.3.2, equation 3.4. Therefore there is no need to impute the paths of X and therefore no need to worry about reducibility issues associated with it; the problem essentially reduces to the diffusion of α only. When this is not the case however, the density if the marginal likelihood of the observations is generally no longer available and augmentation of the paths of X is inevitable. This raises reducibility issues for the parameters in $\sigma_x(\cdot)$.

As we present in this chapter, we can deal with such models using transformations that do not apply directly to the diffusion itself, but to its time scale. For ease of demonstration we introduce this reparametrisation for univariate diffusion first, and then we provide the necessary extensions to deal with the general stochastic volatility case.

4.3 Time change transformations for scalar diffusions

First we will define the likelihood with the use of time change transformations in scalar diffusion models with constant volatility. We will extend to stochastic volatility models in the Section 4.5. Consider a diffusion X defined through the following SDE:

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

Without loss of generality we assume that $X_0 = y_0$, $X_1 = y_1$. For more observations we may split the process into pairs of successive observations and repeat the following procedure for each one of them. We can then just multiply the likelihood bits using Markov property.

In order to draw inferences for the parameters (θ, σ) , one can treat the values of X for $(0 \leq t \leq 1)$ as missing values and adopt a data augmentation scheme. Girsanov's theorem provides the Radon-Nikodym derivative (and hence the likelihood for θ, σ) between the law of X and that of the driftless diffusion $M = \sigma dW(t)$. Unfortunately the dominating measure (that of M) depends on σ and this leads to an irreducible MCMC scheme Roberts and Stramer (2001).

As mentioned earlier, a two-step transformation that breaks down this dependence is available Roberts and Stramer (2001), but since this approach does not generalize to stochastic volatility models we will consider a time change transformation. Define a new time scale $\eta(t)$ to be:

$$\eta(t) = \int_0^t \sigma^2 ds = \sigma^2 t, \quad (4.2)$$

and set a new diffusion U in the following way:

$$U_t = \begin{cases} X_{\eta^{-1}(t)}, & 0 \leq t \leq \sigma^2 \\ M_{\eta^{-1}(t)}, & t > \sigma^2 \end{cases}$$

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 writes:

$$dU_t = \begin{cases} \frac{1}{\sigma^2}\mu(U_t, \theta)dt + dW_t^U & 0 \leq t \leq \sigma^2 \\ dW_t^U, & t > \sigma^2 \end{cases},$$

where W^U is another Brownian motion at the time scale of U . We can now use Girsanov's theorem as before. The law of U , denoted by \mathbb{P} , is given through its Radon-Nikodym derivative with respect to the law \mathbb{W}^U of the Brownian motion W^U at the U -time:

$$\begin{aligned} \frac{d\mathbb{P}}{d\mathbb{Q}} = G(U, \theta, \sigma) &= \exp \left\{ \int_0^{+\infty} \frac{\mu(U_s, \theta)}{\sigma^2} dU_s - \frac{1}{2} \int_0^{+\infty} \frac{\mu(U_s, \theta)^2}{\sigma^4} ds \right\} \\ &= \exp \left\{ \int_0^{\sigma^2} \frac{\mu(U_s, \theta)}{\sigma^2} dU_s - \frac{1}{2} \int_0^{\sigma^2} \frac{\mu(U_s, \theta)^2}{\sigma^4} ds \right\} \end{aligned} \quad (4.3)$$

We now introduce an intuitive factorisation of \mathbb{W}^U as the density of y_1 (under \mathbb{W}^U), multiplied by the dominating measure of the U -path, conditioned on y_1 . More specifically, we set

$$\mathbb{W}^U = \mathbb{W}_y^U \times \text{Leb}(y) \times f(y; \sigma),$$

where \mathbb{W}_y^U is the conditional dominating measure, $\text{Leb}(y)$ denotes Lebesgue measure and f denotes the density of the observed data with respect to the Lebesgue measure under the Wiener measure:

$$f(y_1; \sigma) \equiv N(y_0, \sigma^2)$$

Under this factorisation we can write down the likelihood with respect to $\mathbb{W}_y^U \times \text{Leb}(y)$:

$$d\mathbb{P}(U) = G(U, \theta, \sigma) f(y; \sigma) d\{\mathbb{W}_y^U \times \text{Leb}(y)\} \quad (4.4)$$

Note that the dominating measure in the likelihood specification of (4.4), $\mathbb{W}_y^U \times Leb(y)$, still depends on σ as it reflects a Brownian motion conditioned on the event $U_{\sigma^2} = y_1$. For this reason we introduce a second transformation which applies on both the diffusion's time and path:

$$\begin{aligned} U_t^0 &= (\sigma^2 - t)Z_{t/\sigma^2(\sigma^2-t)}, \quad 0 \leq t \leq \sigma^2, \\ U_t &= U_t^0 + \left(1 - \frac{t}{\sigma^2}\right)y_0 + \frac{t}{\sigma^2}y_1 \end{aligned} \tag{4.5}$$

The 1 – 1 property of this transformation is crucial. We can write

$$\frac{1}{\sigma^2 - t}U_t^0 = Z_{t/\sigma^2(\sigma^2-t)}, \quad 0 \leq t \leq \sigma^2,$$

and get the inverse of this transformation:

$$Z_t = \frac{1 + \sigma^2 t}{\sigma^2} U_{\sigma^4 t / (1 + \sigma^2 t)}^0, \quad 0 \leq t < +\infty$$

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 Z -time

$$dZ_t = \frac{\mu(\nu(Z_t), \theta)}{1 + \sigma^2 t} dt + dW_t^Z, \quad 0 \leq t < \infty, \tag{4.6}$$

where $\nu(Z_t) = U_t$. This transformation essentially transforms to a diffusion that runs from 0 to $+\infty$ in a way so that it will still have unit volatility. Interestingly enough, the conditioning event imposed by the observation y_1 occurs at time infinity and therefore it disappears. We can now write the likelihood as in (4.4) but using the SDE for Z , given by (4.6), and the corresponding reference measure \mathbb{W}_y^Z :

$$d\mathbb{P}(Z) = G(Z, \theta, \sigma) f(y; \sigma) d\{\mathbb{W}_y^Z \times Leb(y)\} \tag{4.7}$$

The Radon-Nikodym derivative between $\mathbb{P}(Z)$ and \mathbb{W}^Z will not go to infinity since it is an 1 – 1 transformation of the Radon-Nikodym derivative between $\mathbb{P}(U)$ and \mathbb{W}^U given by (4.3). The goal of the transformation is now achieved; it is not hard to see that the dominating measure is independent of σ .

Nevertheless, the implementation of a data augmentation scheme based on this likelihood is far from straightforward. For instance, our transformed process Z runs from 0 to $+\infty$ and its time scale depends on parameters that we want to estimate. We present the details of a novel MCMC data augmentation scheme which is practical and efficient in the next section.

4.4 MCMC implementation

The construction of an appropriate data augmentation algorithm is not trivial. We present the details for three important features introduced by the time change transformations: the three time scales, the updates of the imputed diffusion paths and the time scale parameters. As before, it suffices to assume observations $X_0 = y_0$ and $X_1 = y_1$. For more data we can just repeat the following procedure for each pair of successive observations.

4.4.1 Three time scales

We introduce m intermediate points of X at equidistant times between 0 and 1. More specifically, we augment the path which now consists of $m+2$ points: $X = \{X_{i/(m+1)}, i = 0, 1, \dots, m+1\}$. Throughout this chapter 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.

Given a value of the time scale parameter σ we can get the U –time points by applying (4.2) to each one of the existing points. Note that it is only the times that change, the values of the diffusion remain intact. Specifically, we set

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

The points of Z will be multiplied by a time factor which corrects deviations from unit volatility. The Z -time points 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 $+\infty$. For this reason it is more convenient to transform back to U or X to evaluate the likelihood. This is possible as the transformations were all 1 – 1. However we should always keep in mind that the component of the relevant Gibbs scheme is Z .

Figure 4.1 shows a graphical representation of X , U and Z plotted against their corresponding time scales for $\sigma = \sqrt{2}$ and $m = 7$. Although X and U have the same values, their volatilities are $\sqrt{2}$ and 1 respectively. The ending point of Z (which equals 0) does not appear on the graph as it occurs at time $+\infty$.

4.4.2 Updating the paths of X

We may proceed as in the previous chapter using an independence sampler with the reference measure as a proposal. In our case \mathbb{W}^Z reflects a Brownian motion at the Z -time which is fixed given the current values of the time-scale parameter(s). An appropriate algorithm will contain 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 U^* , using (4.5).

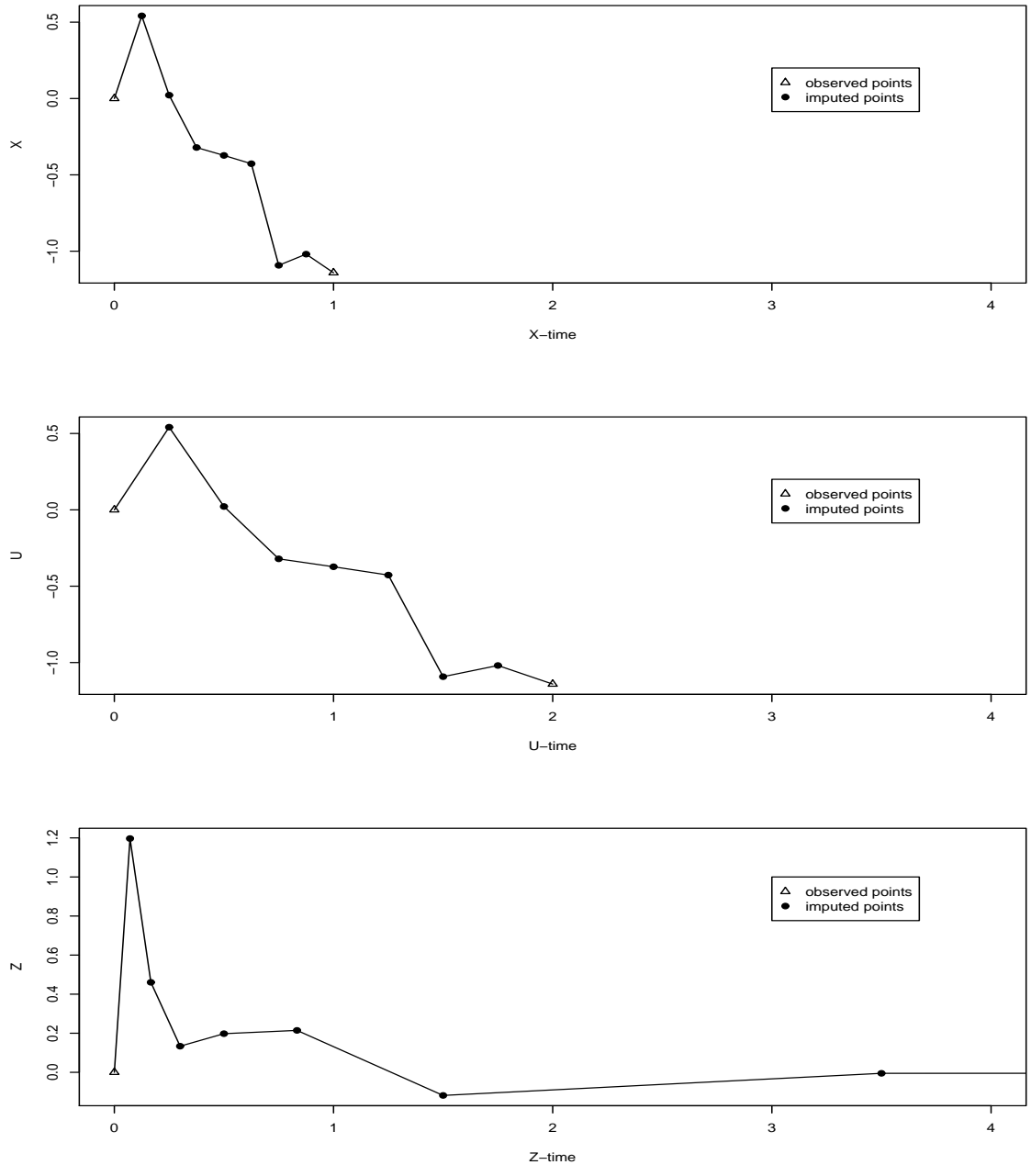


Figure 4.1: Plots of a sample path for X , U and Z against their corresponding times for $\sigma = \sqrt{2}$ and $m = 7$. Z equals 0 at time $+\infty$.

- Step 3: Accept or reject with probability:

$$\min \left\{ 1, \frac{G(U^*, \theta, \sigma)}{G(U^*, \theta, \sigma)} \right\}.$$

4.4.3 Updating time scale parameters

The updates of time scale parameters, in our case σ are of particular interest. In almost all of the cases, the conditional posterior of σ is not available in closed form and Metropolis steps are inevitable. However, the proposed values of these parameters will imply different Z -time scales. In other words, for each potential proposed value for σ there exist different set of Z -points needed for accurate approximations of the likelihood and consequently the Metropolis accept-reject probability. In theory, this would pose no issues if we could store an infinitely thin partition of Z , but of course this is not possible.

Alternatively we may use retrospective sampling ideas; see Papaspiliopoulos and Roberts (2005), Beskos and Roberts (2005) and ? for applications in different contexts. Under the assumption of a sufficiently fine partition 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. Therefore, they can be drawn AFTER the proposal of the candidate value of the time scale parameter. Since they contribute nothing to the likelihood, we may argue that their distribution is given by the likelihood's reference measure which reflects a Brownian motion. To ensure compatibility with the recorded partition of Z , it suffices to condition on the two 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 , and that two closest times of recorded values are 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}^U between t_a and t_c we have:

$$Z_{t_b} | Z_{t_a}, Z_{t_c} \sim N \left\{ \frac{(t_b - t_a)Z_{t_c} + (t_c - t_b)Z_{t_a}}{t_c - t_a}, \frac{(t_b - t_a)(t_c - t_b)}{t_c - t_a} \right\} \quad (4.8)$$

The situation is pictured in Figure 4.2, where the black bullets represent the stored points and the triangles the new points required for a proposed value of σ . The latter should be drawn retrospectively given the former via (4.8).

To sum up, a suitable algorithm for the σ -updates will consist of the following steps:

- Step 1: Propose a candidate value for σ , say σ^* .
- Step 2: Repeat for each pair of successive points:
 - Use (4.2) and (4.5) to get the new times associated with it.
 - Draw the values of Z at the new times using (4.8).
 - Transform back to U^* , using (4.5).

Form the entire path U^* by appropriately joining its bits.

- Step 3: Accept or reject with probability:

$$\min \left\{ 1, \frac{G(U^*, \theta, \sigma^*) f(y; \sigma^*)}{G(U^*, \theta, \sigma) f(y; \sigma)} \right\}.$$

4.5 Time change for stochastic volatility models

4.5.1 Main category

Let's return to the general class of stochastic volatility models of this chapter with SDE given by (3.1). As before, we assume one observation apart from the initial point ($X_0 = 0$, $X_{t_k} = y$), as the generalization to more data is straightforward due to the Markov property of the 2-dimensional diffusion. We may use the reparametrisations of Section 3.3.2 for α and obtain γ . Doing so, the likelihood for γ can be written with respect to a Brownian motion distribution \mathbb{W}^γ from 0 to t_k .

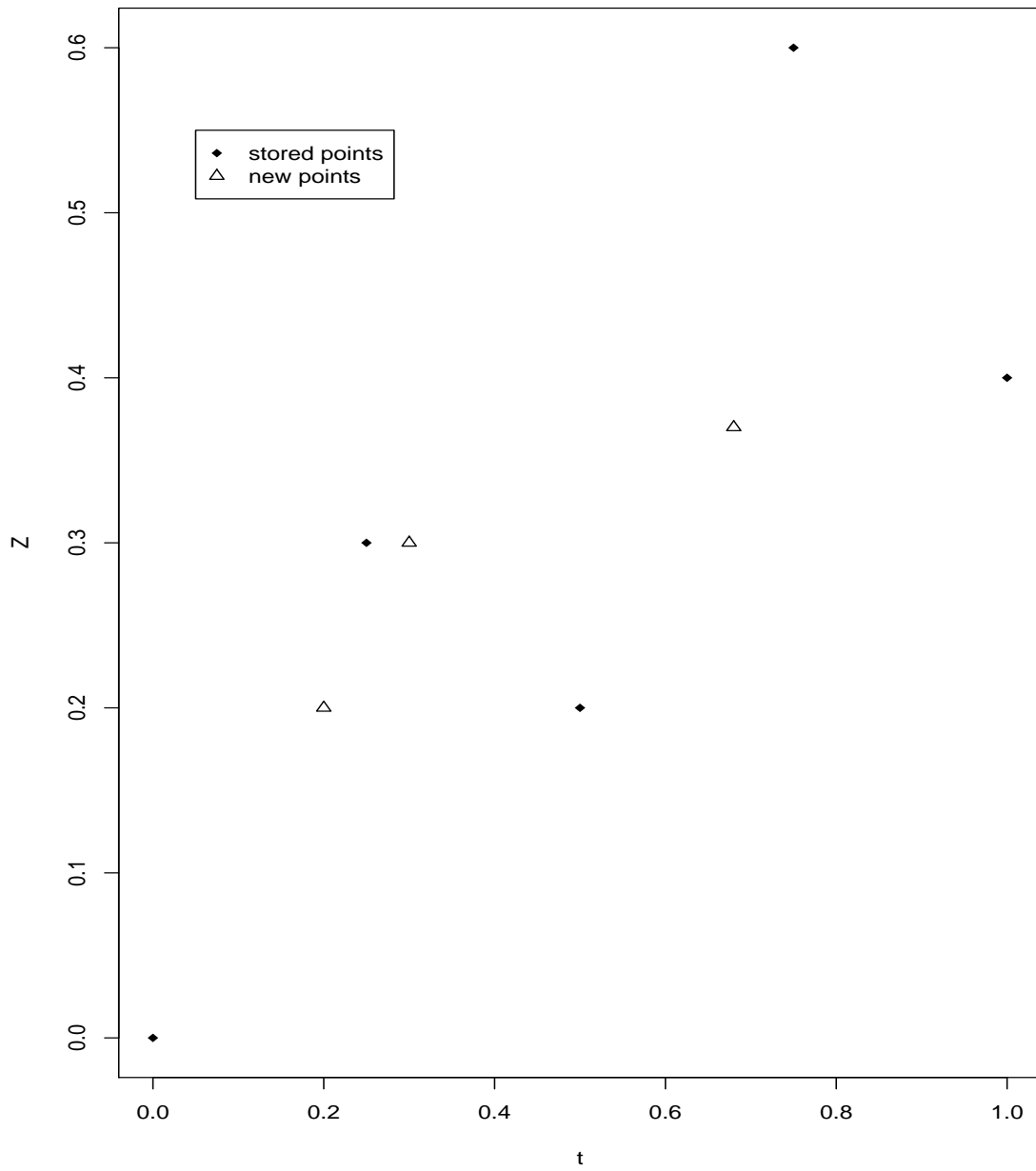


Figure 4.2: Updates of time scale parameters: For every proposed value of them, new points are required and should be obtained conditional on the stored points.

$$\frac{d\mathbb{P}}{d\mathbb{W}_\gamma}(\gamma) = G(\gamma, \theta) \quad (4.9)$$

Let $\alpha_t = g_t^\gamma = \eta \circ h^{-1}(\gamma_t, \theta)$. Then, conditional on γ , the SDE of X becomes:

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

Note that the situation is similar to Section 4.3. We can introduce a new time scale

$$\begin{aligned} \eta(t, \gamma, \theta) &= \int_0^t \sigma_x^2(g_s^\gamma, \theta)ds, \\ T &= \eta(t_k, \gamma, \theta), \end{aligned}$$

and define U with the new time scale as before (M is a Brownian motion on the U -time scale):

$$U_t = \begin{cases} X_{\eta^{-1}(t)}, & 0 \leq t \leq T \\ M_{\eta^{-1}(t)}, & t > T \end{cases} \quad (4.10)$$

The SDE for U now becomes:

$$dU_t = \left\{ \frac{\mu_x(U_t, \gamma_{\eta^{-1}(t, \gamma, \theta)}, \theta)}{\sigma_x^2(\gamma_{\eta^{-1}(t, \gamma, \theta)}, \theta)} \right\} dt + dW_t^U, \quad 0 \leq t \leq T.$$

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

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

,

and $\mathbb{W}^U = \mathbb{W}_y^U \times Leb(y) \times f(y, \gamma, \theta)$ with \mathbb{W}_y^U being the Wiener measure conditioned on the data, $Leb(y)$ the Lebesgue measure and f denotes the density of the observed data with respect to the Lebesgue measure under the Wiener measure:

$$f(y; \gamma, \theta) \equiv N(y_0, T)$$

The dominating measure \mathbb{W}_y^U reflects a Brownian motion conditioned to equal y at a parameter depended time $T = \eta(t_{k+1}, \gamma, \theta)$. To remove this dependency we introduce a second time change as before:

$$\begin{aligned} U_t^0 &= (T-t)Z_{\frac{t}{T(T-t)}}, \quad 0 \leq t \leq T, \\ U_t &= U_t^0 + (1 - \frac{t}{T})y_0 + \frac{t}{T}y_1 \end{aligned} \tag{4.11}$$

The SDE for Z is now given by:

$$dZ_t = \left\{ \frac{\mu(\nu(Z_t), \gamma_{k(t, \gamma, \theta)}, \theta)}{\sigma_x^2(\gamma_{k(t, \gamma, \theta)}, \theta)} \frac{T}{1+tT} \right\} dt + dW_t^Z, \quad 0 \leq t < \infty,$$

where $k(t, \gamma, \theta)$ denotes the initial time scale of X .

Conditional on γ , the likelihood can be written in a similar manner as in (4.7):

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

The full likelihood is thus provided by multiplying with the bit for γ from (4.9)

$$\frac{dP}{dW_\gamma}(\gamma) \frac{d\mathbb{P}}{d\{\mathbb{W}_y^Z \times Leb(y)\}}(Z|\gamma) = G(\gamma, \theta)G(Z, \gamma, \theta)f(y; \gamma, \theta) \tag{4.12}$$

Now we are in a position to construct a data augmentation scheme as in Section 4.4 with a suitable addition for the updates of γ . We propose an overlapping blocks scheme as in Section 3.4. Note that the Z -time depends on γ and therefore its paths should be treated as time scale parameters; see Section 4.4.3.

4.5.2 Incorporating leverage effect

So far, we made the assumption that the infinitesimal increments of X and γ are independent, in other words we assumed no *leverage* effect. However this is not always the case; the application of Section 3.8 provides such an example. This assumption can be relaxed in the following way: In the presence of a *leverage* effect ρ the SDE of X conditional on γ writes (W is the driving Brownian motion of γ):

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

Note that given γ , the term $\rho\sigma_x(g_t^\gamma, \theta)dW_t$ is deterministic since W can be regarded as a function of γ and its parameters θ . In that respect, it would be perfectly fine to treat it as part of the drift of X_t . Nevertheless if we do so, the assumptions ensuring a weakly unique solution to the SDE of X are violated. Therefore we define the following ‘intermediate’ infinitesimal transformation:

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

which lead us to the following SDE for H :

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

We can now proceed as before, defining U and Z based on the SDE and the volatility of H . For more details see the relevant example of Section 4.6.

4.5.3 Other extensions

In this section we note two additional extensions which may prove particularly useful for applications. First, we target stochastic volatility models where conditional on γ , the SDE of X may be written as:

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

This class contains the formulations of Andersen and Lund (1998), Gallant and Tauchen (1998), Durham (2002), Eraker (2001) etc. In order to be able to apply the time change transformations of the Section 4.5.1, we should first transform X , so that it takes the form of 3.1. A check on multivariate Ito's lemma properties indicates that this is essentially the same problem as transforming a diffusion with volatility $\sigma_2(X_t, \theta)$ to one with unit volatility; in other words, we can use the first transformation of Roberts and Stramer (2001). Doing so, the parameters of $\sigma_2(X_t, \theta)$ enter the likelihood in two ways: i) through the $f(y; \gamma, \theta)$ which now should include the relevant Jacobian term and ii) through the drift of Z which is centered at 0 based on the transformed observations. See Section 4.7 for an example.

The second extension refers to cases where the latent paths of α may be inferred from additional data resources, i.e. implied volatility from option prices, one may use the parametrisation of Section 3.7.3 for γ , and the relevant modification for its updates. The remaining likelihood specification and MCMC implementation would be exactly the same.

4.6 Simulation example

We simulated data from the following stochastic volatility model

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

using a high frequency Euler approximating scheme with a step of 0.001. We simulated 500,001 points and recorded one value of X for every 1000 forming a dataset of 500 observations of X (apart from the initial) and $0 \leq t \leq 500$. 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 $\mu_\alpha = -0.2$.

We now summarize the reparametrisations required to construct an irreducible data augmentation scheme. First we transform α to γ in the following way:

1. $\beta_t = h(\alpha_t) = \alpha_t/\sigma$ $0 \leq t \leq 500$
2. $\gamma_t = \beta_t - \alpha_0/\sigma$ $\alpha_t = \nu(\gamma_t, \sigma, \alpha_0)$

Given γ , we apply the following transformations on X for each pair of consecutive times t_{k-1} and t_k ($k = 1, 2, \dots, 500$). First, we remove the term introduced due to the *leverage* effect:

$$H_t = X_t - \int_{t_{k-1}}^t \rho \exp\{\nu(\gamma_s, \sigma, \alpha_0)/2\} dW_s, \quad t_{k-1} \leq t \leq t_k.$$

Then we set

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

and we define U and Z exactly as in Section 4.5.1, but based on H rather on X . The elements of the MCMC scheme are Z, γ, α_0 and the parameters.

We set flat priors on all parameters, restricting $\kappa_x, \kappa_\alpha, \sigma$ to be positive and ρ between -1 and 1 . We ran MCMC algorithms with numbers of imputed points being 30 and

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 4.1: Summaries of the posterior draws for the simulation example of Chapter 3 for $m = 50$.

50. The length of the overlapping blocks was 2 and the relevant acceptance rate 75% whereas the acceptance rate for X was 95%. Figure 4.3 shows autocorrelation plots for the 2-dimensional diffusion's $(X, \alpha)'$ volatility parameters ρ and σ . There is no sign of any increase to raise suspicions against the irreducibility of the chain. Figure 4.4 shows density plots for all parameters and both values of m . These plots indicate a sufficiently fine discretisation and a good agreement with true values of the parameters. The latter is also confirmed by Table 4.1.

4.7 Example:US treasury

To illustrate the methodology of this Chapter we fit a stochastic volatility model to US treasury bill rates. More specifically, the dataset consist 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 4.5.

This is a standard dataset for interest rate applications; see for instance Andersen and Lund (1998), Gallant and Tauchen (1998), Durham (2002), Durham and Gallant (2002), Eraker (2001), Golightly and Wilkinson (2005). In all these papers the adopted stochastic volatility models roughly had the following form:

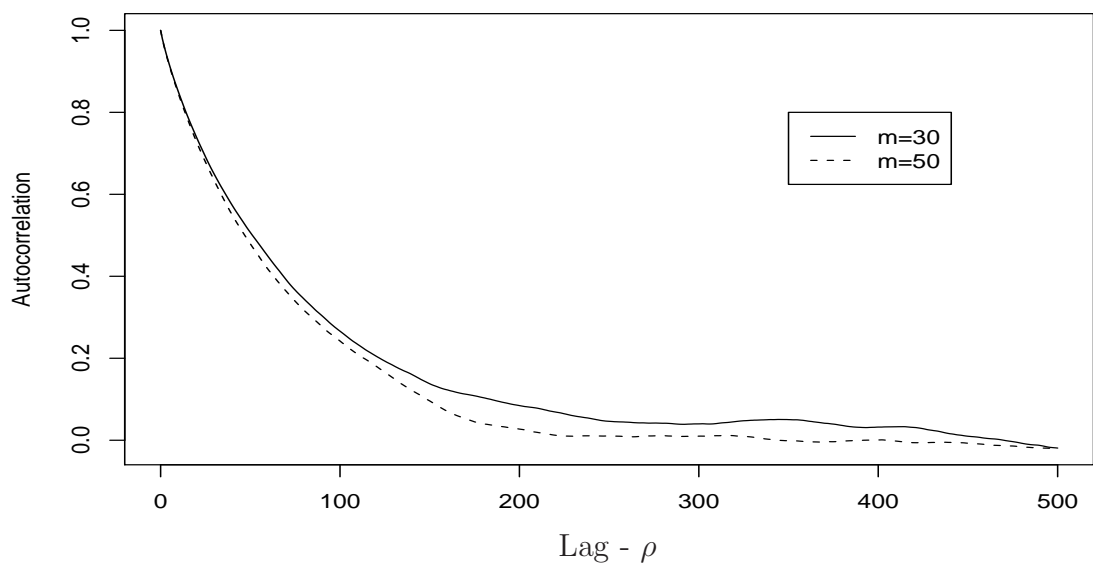
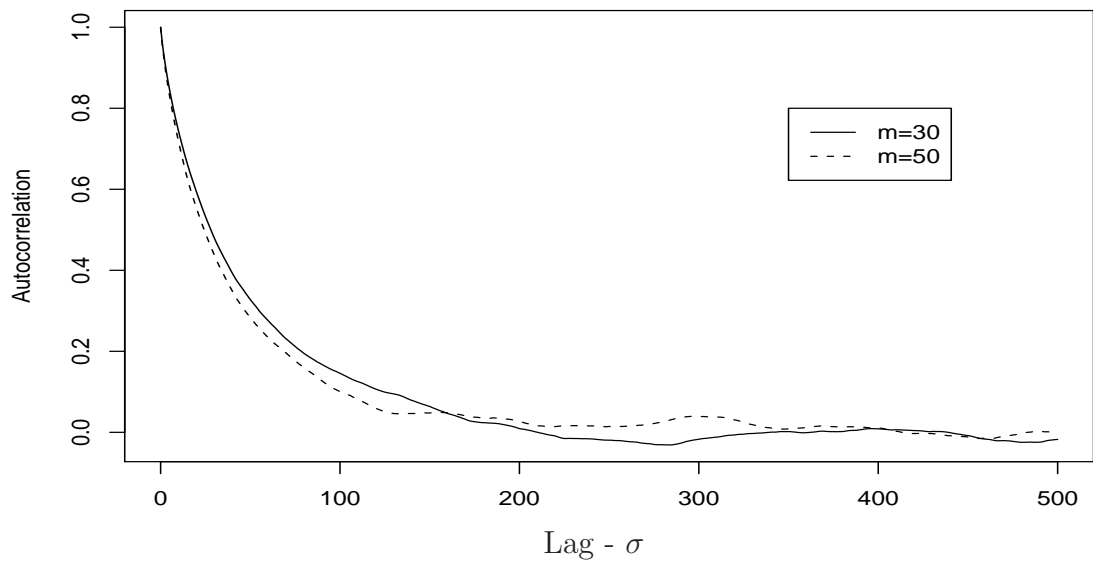


Figure 4.3: Autocorrelation plots for the posterior draws of ρ and σ for different numbers of imputed points $m = 30, 50$. Simulation example of Chapter 3.

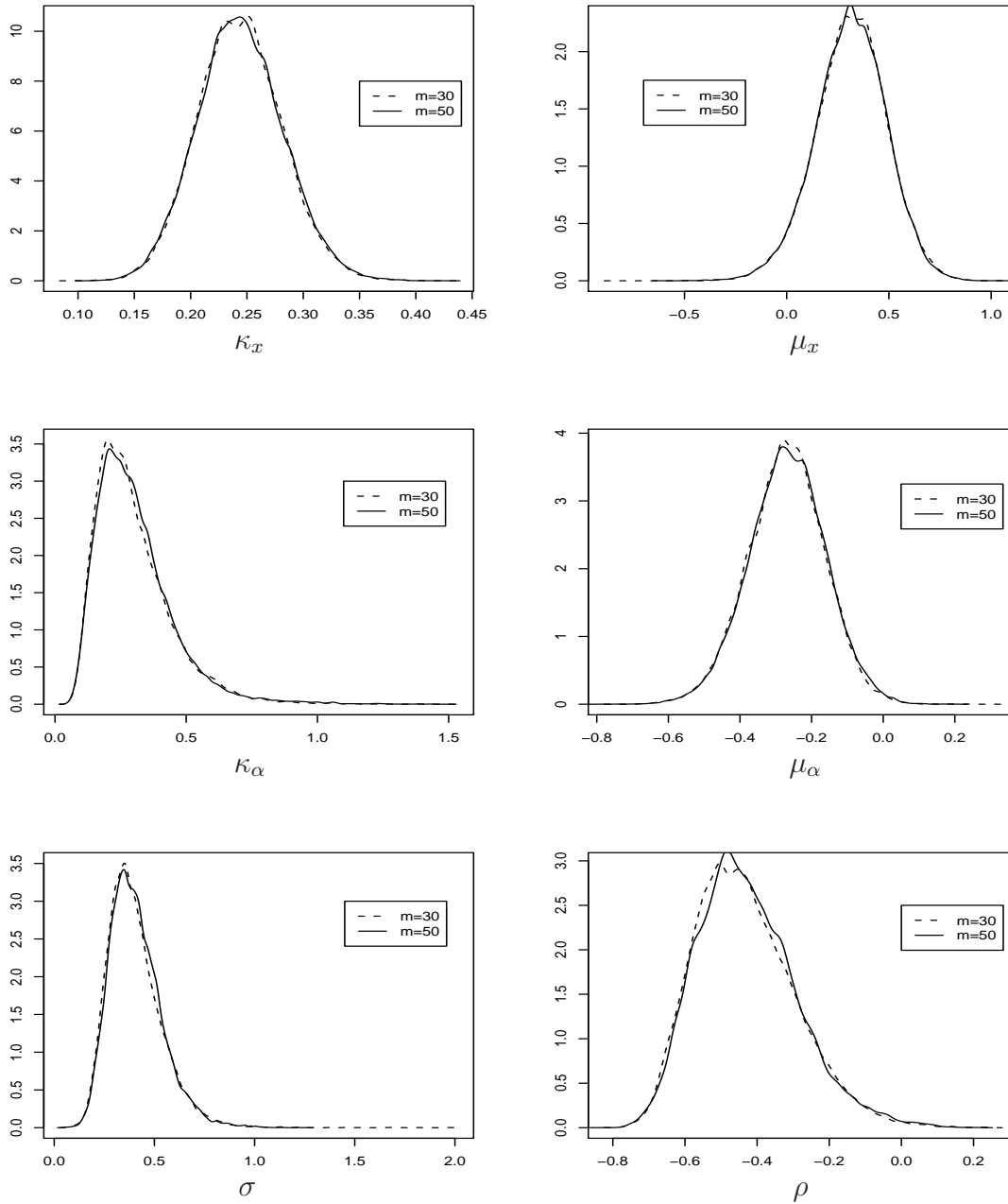


Figure 4.4: Kernel densities of the posterior draws of all the parameters for different numbers of imputed points $m = 30, 50$. Simulation example of Chapter 3.

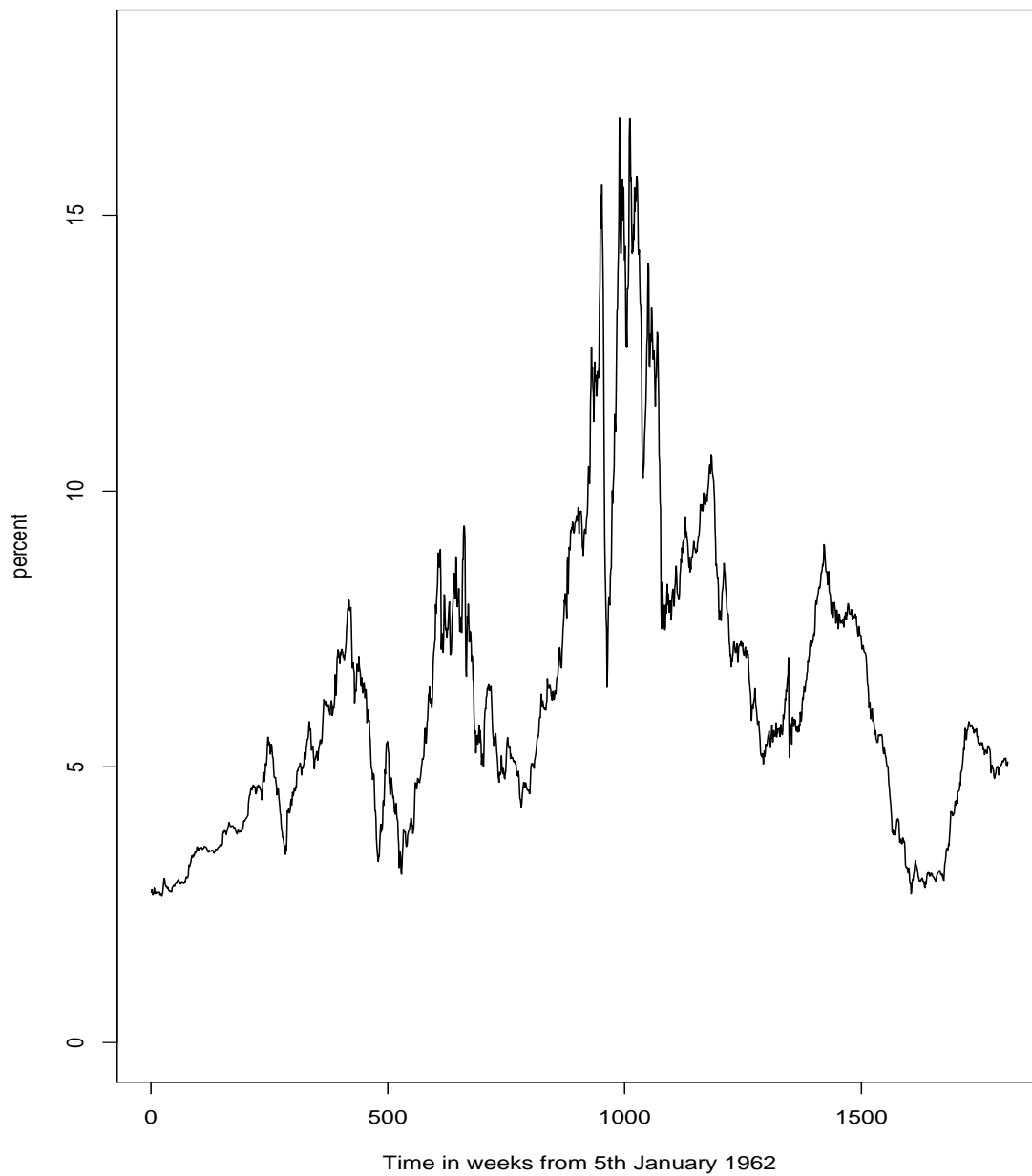


Figure 4.5: Weekly 3-month US Treasury bill rate from the 5th of January 1962 up to the 30th of August 1996.

$$\begin{aligned}
dr_t &= (\theta_0 - \theta_1 r_t)dt + r_t^\psi \exp(\alpha_t/2)dB_t, \\
d\alpha_t &= \kappa(\mu - \alpha_t)dt + \sigma dW_t,
\end{aligned}
\tag{4.13}$$

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

$$\begin{aligned}
dr_t &= (\theta_0 - \theta_1 r_t)dt + \sigma_r r_t^\psi \exp(\alpha_t/2)dB_t, \\
d\alpha_t &= -\kappa\alpha_t dt + \sigma dW_t,
\end{aligned}
\tag{4.14}$$

We just need to set $\sigma_r = \exp \mu$ to see that the models (4.13) and (4.14) are the same. The reason for choosing (4.13) was that the MCMC draws of μ were much less autocorrelated than the corresponding ones of σ_r . In line with Gallant and Tauchen (1998) and Golightly and Wilkinson (2005) 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. If we set $X_t = \log(r_t)$, the volatility of X_t becomes $\exp(\alpha_t/2)$. Hence we set the U -time, for two consecutive observation times t_{k-1} and t_k , as

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

and define U and Z accordingly. We also transform α to γ exactly as before:

1. $\beta_t = h(\alpha_t) = \alpha_t/\sigma$
2. $\gamma_t = \beta_t - \alpha_0/\sigma \quad \alpha_t = \nu(\gamma_t, \sigma, \alpha_0)$

We applied MCMC algorithms 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

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

between successive Wednesdays to 5/252. We chose non-informative priors for all the parameters, restricting κ and σ to be positive to ensure identifiability and eliminate the possibility of explosion respectively. The algorithm was run for 50,000 iterations and for m equal to 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 4.6) indicate that a discretisation from an m of 10 or 20 provide reasonable approximations. The corresponding autocorrelation plots of Figure 4.7 do not show increasing autocorrelation in m which would reveal reducibility (in the limit) issues. Finally summaries of the posterior draws for all the parameters are provided in Table 4.2. The parameters κ , μ and σ are different from 0 verifying the existence of stochastic volatility. However, this does not seem to be the case for the existence of mean reversion of the rate (at least not under this model), 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 (2005).

4.8 Concluding remarks

In this chapter we enriched the class of stochastic volatility models that can be handled using MCMC schemes in fixed datasets. The generalization relaxed the assumption made in the previous chapter that the drift of the observed process should be independent of it

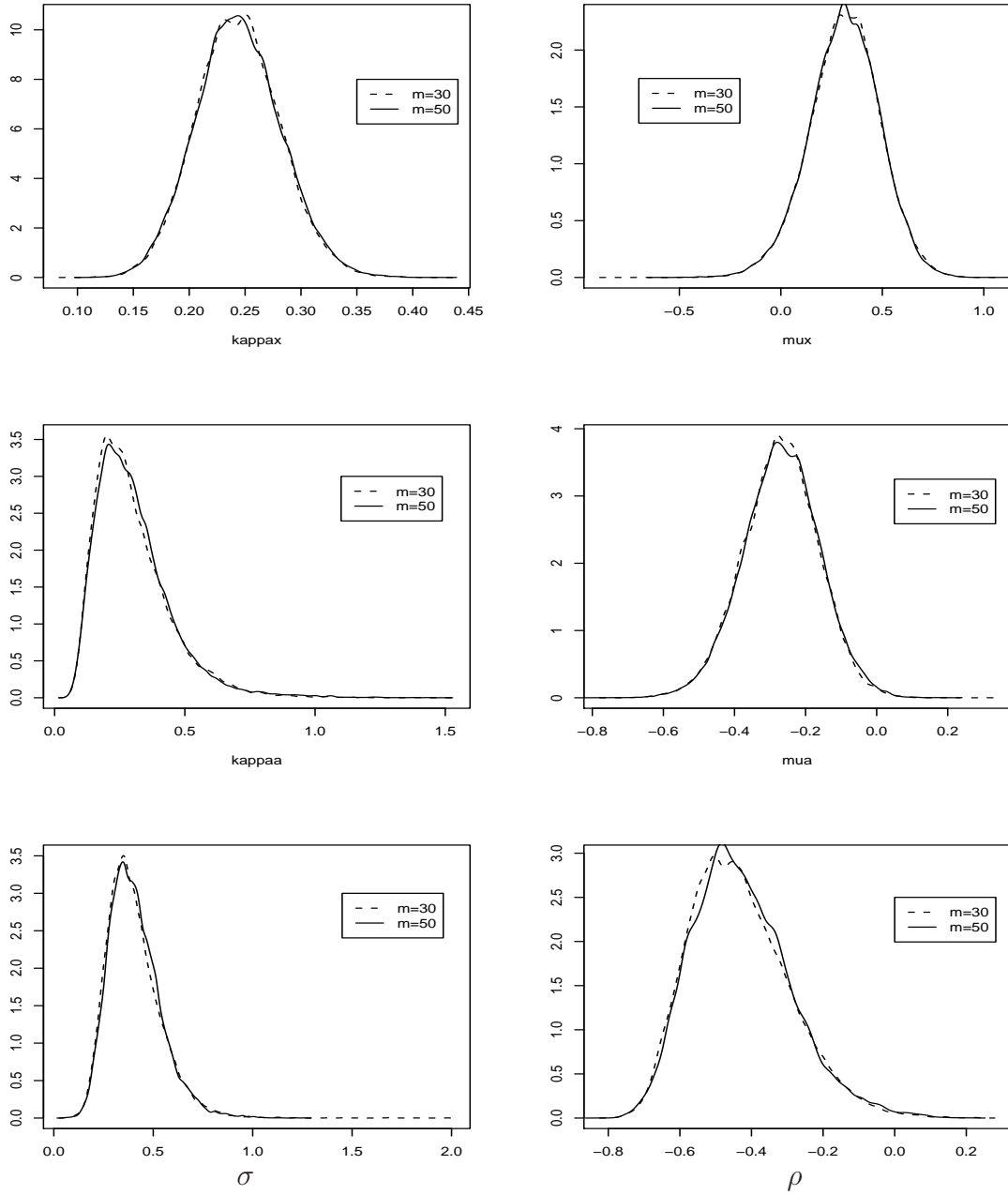


Figure 4.6: Kernel densities of the posterior draws of all the parameters and the log-likelihood for different values of imputed points $m = 10, 20$. Example on Weekly 3-month US Treasury bill rates.

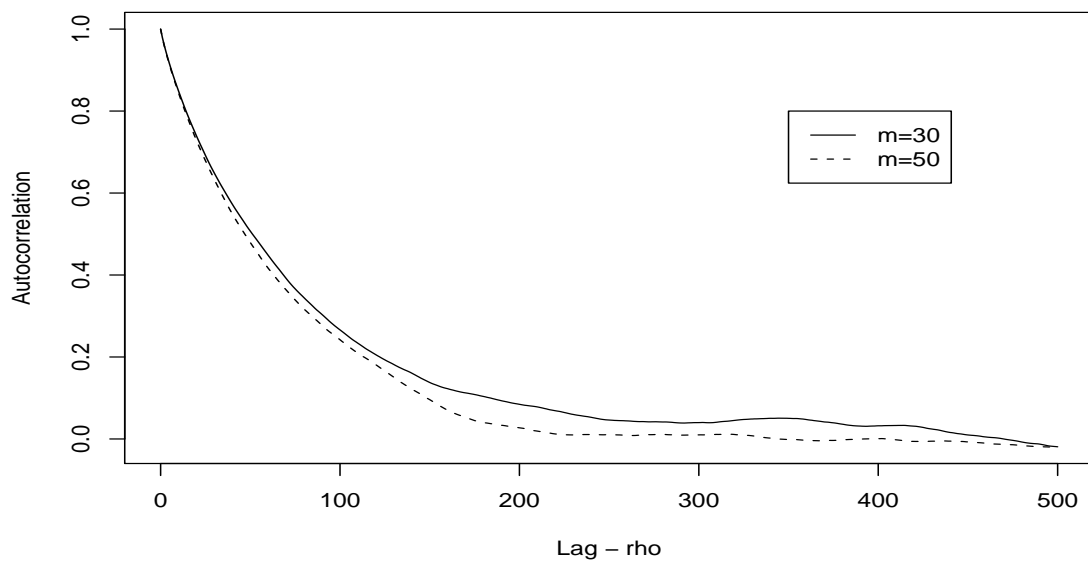
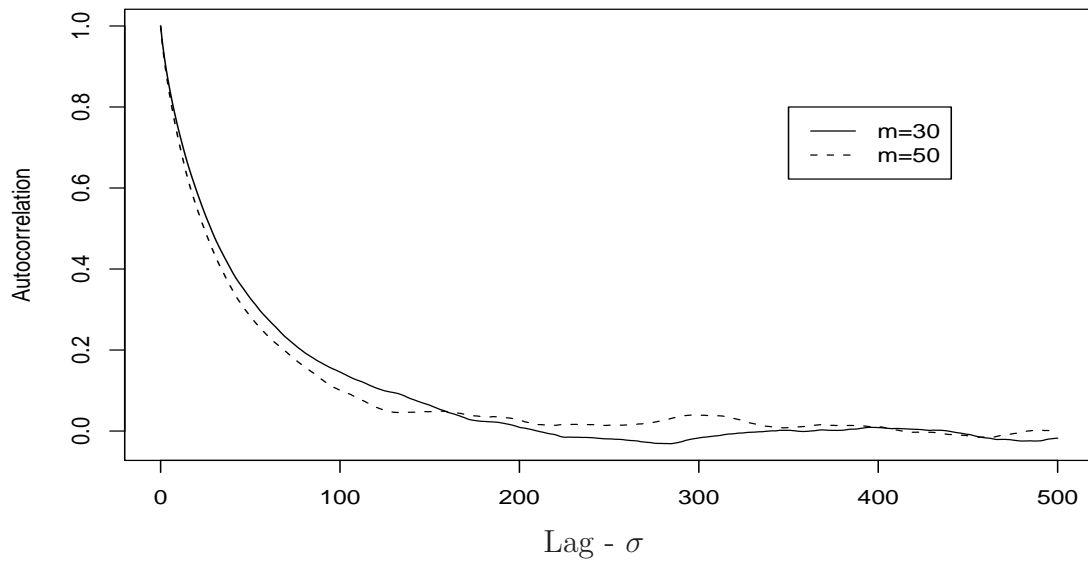


Figure 4.7: Autocorrelation plots for the posterior draws of the model parameters for different numbers of imputed points $m = 10, 20$ for the analysis of Weekly 3-month US Treasury bill rates.

and incorporated many interesting stochastic volatility models used in interest rates applications. This was achieved with the introduction of a reparametrisation which operates mainly on the time scale of the diffusion, rather than its path. In line with this innovative transformation we also introduced a novel efficient MCMC scheme which found to perform reasonably well and fast. Such a scheme introduces no additional approximation error other than that included in methodologies based on a discretisation of the diffusion path.

Overall, data augmentation 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., 2006b), or the closed form approximate likelihood expressions of Aït-Sahalia (2002), Aït-Sahalia (2005), but nevertheless they give a satisfactory and general solution to the problem. In fact, their main advantage is their unified general framework that covers discretely (fully) and partially observed diffusions and in particular stochastic volatility models. Therefore, the contribution of this thesis is of particular value in that respect.

There is still plenty room for improvement and extensions. The updates of the missing paths are done through an independence sampler whose performance is not guaranteed in all cases. The reparametrisations and sampling schemes introduced break the dependency between parameters and latent paths to a significant degree but they do not eliminate the problem; in some cases the posterior MCMC samples do exhibit considerable amount of autocorrelation which has implications to the efficiency of the algorithm. Also the framework is still not general enough to include jump diffusions and Levy processes. Finally, the problem of model selection has not been fully addressed. Marginal likelihood techniques as in Chib and Jeliazkov (2001) are available but they may become highly impractical and inefficient as number of latent variables (imputed paths) is particularly high. The development of a general reversible jump MCMC algorithm, in the spirit of Dellaportas et al. (2006) would provide a vast improvement.

Appendix

Reparametrisations

In order to apply the data augmentation scheme for the stochastic volatility models of 3.8, we need to apply suitable reparametrisations to the diffusions that correspond to the volatility α . These are model specific and they are provided below. In all the models the second transformation that leads to γ is defined in the same way as in (3.7) for two consecutive times t_k and t_{k+1} . Hence, the SDE of γ will have unit volatility and drift

$$\mu_\gamma(\gamma_t, \theta) = \mu_\beta \{ \eta(\gamma_t, \theta), \theta \},$$

where $\eta(\cdot)$ just inverts (3.7), $\beta_t = \eta(\gamma_t, \theta)$. Also the volatility of the SDE of β is always 1. Therefore we only give the first transformation and the SDE of the transformed diffusion β for each model. For the Heston model these are

$$\begin{aligned} \beta_t &= \frac{2\sqrt{\alpha_t}}{\sigma}, \\ d\beta_t &= \left\{ \frac{2\kappa\mu - 0.5\sigma^2}{\sigma^2\beta_t} - 0.5\kappa\beta_t \right\} dt + dW_t. \end{aligned}$$

For the GARCH diffusion model we have:

$$\begin{aligned}\beta_t &= \frac{\log(\alpha_t)}{\sigma}, \\ d\beta_t &= \left\{ \frac{\kappa(\mu - e^{\sigma\beta_t})}{\sigma e^{\sigma\beta_t}} - \frac{\sigma}{4} \right\} dt + dW_t.\end{aligned}$$

Finally, for the log-normal model we obtain:

$$\begin{aligned}\beta_t &= \frac{\log(\alpha_t)}{\sigma}, \\ d\beta_t &= \frac{\kappa(\mu - \sigma\beta_t)}{\sigma} dt + dW_t.\end{aligned}$$

Proposals for the volatility drift parameters

The conditional likelihood for the drift parameters of the volatility consists of two parts. The first is a Gaussian part which comes from the density of the observations on X . These parameters enter in this density through the driving Brownian motion of γ (W), which may be seen as a function of them and γ . The second bit is comes from Girsanov's formula. Had it been just the Girsanov part, these updates would be Gibbs steps. This is not the case here, nevertheless we can still use this full conditionals as proposals. The Metropolis accept-reject probability will then be determined from the priors (in the case of non-conjugate priors) and the first part of the likelihood only. We demonstrate how to obtain these proposals for the case of κ in the Heston model only, as the same procedure may be used for rest. Note that since only σ was involved in the transformation from α to γ (and vice versa), we can transform back to α in a deterministic way as we are operating conditional on σ and γ . We thus provide the proposal in α terms assuming an improper prior $p(\kappa) \propto 1$.

$$\begin{aligned}
p(\kappa \mid X, \alpha, \mu, \sigma, \rho, \mu_S) &\propto \exp \left\{ \int_0^T \frac{\kappa(\mu - \alpha_t)}{\sigma^2 \alpha_t} d\alpha_t - \frac{1}{2} \int_0^T \frac{\kappa^2(\mu - \alpha_t)^2}{\sigma^2 \alpha_t} dt \right\} \\
&= \exp \left\{ - \left(\frac{\kappa}{2\sigma^2} \int_0^T \frac{(\mu - \alpha_t)^2}{\alpha_t} dt - \frac{2\kappa}{2\sigma^2} \int_0^T \frac{\mu - \alpha_t}{\alpha_t} dt \right) \right\} \\
&= \exp \left\{ - \frac{\kappa^2 - 2I_2/I_1}{2\sigma^2/I_1} \right\} \propto \exp \left\{ - \frac{(\kappa - I_2/I_1)^2}{2\sigma^2/I_1} \right\}
\end{aligned}$$

where

$$\begin{aligned}
I_1 &= \int_0^T \frac{(\mu - \alpha_t)^2}{\alpha_t} dt \\
I_2 &= \int_0^T \frac{\mu - \alpha_t}{\alpha_t} dt.
\end{aligned}$$

Since $\kappa > 0$ for non-explosion, the proposal density can be recognized as truncated normal distribution kernel with mean I_2/I_1 and variance σ^2/I_1 . The integrals I_1, I_2 may be computed numerically given the augmented path of α .

Bibliography

- Aalen, O. O. and Gjessing, H. K. (2001). Understanding the shape of the hazard rate: a process point of view. *Statistical Science*, 16:1–22. With comments and a rejoinder by the authors.
- Aalen, O. O. and Gjessing, H. K. (2004). Survival models based on the Ornstein Uhlenbeck process. *Lifetime Data Analysis*, 10:407–423.
- Aït-Sahalia, Y. (1996a). Nonparametric pricing of interest rate derivative securities. *Econometrica*, 64(3):527–560.
- Aït-Sahalia, Y. (1996b). Testing continuous-time models of the spot interest rate. *The Review of Financial Studies*, 9:385–426.
- Aït-Sahalia, Y. (2002). Maximum likelihood estimation of discretely sampled diffusions: a closed form approximation approach. *Econometrica*, 70:223–262.
- Aït-Sahalia, Y. (2005). Closed form likelihood expansions for multivariate diffusions. In preparation.
- Aït-Sahalia, Y. and Kimmel, R. (2005). Maximum likelihood estimation for stochastic volatility models. To appear in the Journal of Financial Economics.
- Andersen, T., Bollerslev, T., and Meddahi, N. (2006). Analytic evaluation of volatility forecasts. Tech. rep., Northwestern University.
- Andersen, T. G. and Lund, J. (1998). Estimating continuous-time stochastic volatility models of the short term interest rate. 77:343–377.

- Andrieu, C., Doucet, A., and Tadic, V. B. (2005). On-line parameter estimation in general state space models. pages 332–337.
- Bachelier, L. (1900). Theory of speculation. *Ann. Sci. École Norm. Sup. (3)*, 17:21–86.
- Bergman, Y., Grundy, B., and Wiener, Z. (1996). General properties of option prices. *Journal of Finance*, 51(5):1573–1610.
- Beskos, A., Papaspiliopoulos, O., and Roberts, G. (2005). Retrospective exact simulation of diffusion sample paths with applications. To appear.
- Beskos, A., Papaspiliopoulos, O., and Roberts, G. (2006a). Monte carlo maximum likelihood estimation for discretely observed diffusion processes. Submitted.
- Beskos, A., Papaspiliopoulos, O., Roberts, G., and Fearnhead, P. (2006b). 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. 1:17–39.
- Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, 81:637–654.
- Brandt, M. W. and Santa Clara, P. (2001). Simulated likelihood estimation of diffusions with an application to exchange rate dynamics in incomplete markets. *Journal of Financial Economics*, 69:959–993.
- Brockwell, P. J. (1994). On continuous-time threshold ARMA processes. *Journal of Statistical Planning and Inference*, 39(2):291–303.
- Brooks, S. P. (1998). Markov chain monte carlo method and its application. 47:69–100.

- Chan, K. C., Karolyi, G. Andrew, Longstaff, Francis A., and Sanders, Anthony B. (1992). An empirical comparison of alternative models of the short-term interest rate. *The Journal of Finance*, 47(3):1209–1227.
- Chernov, M., Gallant, A., Ghysels, E., and Tauchen, G. (2003). Alternative models for stock price dynamics. *Journal of Econometrics*, 116:225–258.
- Chernov, M. and Ghysels, E. (2000). A study towards a unified approach to the joint estimation of objective and risk neutral measures for the purposes of options valuation. *Journal of Financial Economics*, 56:407–458.
- Chib, S. and Jeliazkov, I. (2001). Marginal likelihood from the Metropolis-Hastings output. 96(453):270–281.
- Chib, S., Pitt, M. K., and Shephard, N. (2005). Likelihood based inference for diffusion models. Submitted.
- Congdon, P. (2001). *Bayesian statistical modelling*. Wiley Series in Probability and Statistics. John Wiley & Sons Ltd., Chichester.
- Corzo, T. and Schwartz, E. S. (2000). Convergence within the european union: Evidence from interest rates. *Economic Notes*, 2:243–268.
- Cox, J. C., Ingersoll, J. E., and Ross, S. A. (1985). A theory of the term structure of interest rates. *Econometrica*, 53:385–407.
- Dacunha-Castelle, D. and Florens-Zmirou, D. (1986). Estimation of the coefficients of a diffusion from discrete observations. 19(4):263–284.
- Daniels, M. and Kass, R. (1999). Nonconjugate bayesian estimation of covariance matrices in hierarchical models. *Journal of the American Statistical Association*, 94:1254–1263.
- Del Moral, P. (2004). *Feynman-Kac formulae*. Probability and its Applications (New York). Springer-Verlag, New York. Genealogical and interacting particle systems with applications.

- Dellaportas, P., Friel, N., and Roberts, G. (2006). Bayesian model selection for partially observed diffusion models. *Biometrika*. To appear.
- Dellaportas, P. and Roberts, G. (2003). An introduction to MCMC. *in* Spatial Statistics and Computational methods, J. Muller (editor), 1-42. Springer-Verlag, NY.
- Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. 39(1):1–38. With discussion.
- Derman, E. and Kani, I. (1994). Riding on the smile. *Risk*, 7:32–39.
- Doucet, A., Godsill, S., and Andrieu, C. (2001). On sequential monte carlo sampling methods for bayesian filtering. 10:197–208.
- Duffie, D. and Kan, R. (1996). A yield factor model of interest rates. *Mathematical Finance*, 6:379–406.
- Dumas, B., Fleming, J., and Whaley, R. E. (1998). Implied volatility functions: Empirical tests. *Journal of Finance*, 53:2059–2106.
- Dupire, B. (1994). Pricing with a smile. *Risk*, 7:18–20.
- 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.
- Einstein, A. (1905). On the movement of small particles suspended in a stationary liquid demanded by the molecular kinetic theory of heat. *Ann. Physik*, 17.
- El Karoui, N., Jeanblanc, M., and Shreve, S. (1998). Robustness of the Black Scholes formula. *Mathematical Finance*, 8(2):93–126.
- Elerian, O. (1999). Simulation estimation of continuous-time models with applications to finance. D.Phil thesis, Nuffield College, Oxford.

- Elerian, O. S., Chib, S., and Shephard, N. (2001). Likelihood inference for discretely observed non-linear diffusions. *Econometrica*, 69:959–993.
- Eraker, B. (2001). Markov chain Monte Carlo analysis of diffusion models with application to finance. *J. Bus. Econom. Statist.*, 19(2):177–191.
- Florens-Zmirou, D. (1989). Approximate discrete schemes for statistics of diffusion processes. *Statistics*, 20:547–557.
- Florens-Zmirou, D. (1993). On estimating the diffusion coefficient from discrete observations. 30(4):790–804.
- Föllmer, H. and Schweizer, M. (1991). Hedging of contingent claims under incomplete information. In *Applied stochastic analysis (London, 1989)*, volume 5 of *Stochastics Monogr.*, pages 389–414. Gordon and Breach, New York.
- 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. 93(441):10–24.
- Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. 85(410):398–409.
- Genon-Catalot, V., Laredo, C., and Picard, D. (1992). Nonparametric estimation of the diffusion coefficient by wavelets methods. 19(4):317–335.
- Ghysels, E., Harvey, A., and Renault, E. (1996). Stochastic volatility, in. *Handbook of Statistics 14, Statistical Methods in Finance*. G.S. Maddala and C.R. Rao (eds), North Holland, Amsterdam.
- Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (1996). *Markov Chain Monte Carlo in Practice*. Chapman and Hall, London.

- Gillespie, D. (1976). A general method for numerically simulating: the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, 22:403–434.
- Golightly, A. and Wilkinson (2005). Bayesian sequential inference for nonlinear multivariate diffusions. Submitted.
- Gordon, N. J., Salmond, D., and Smith, A. (1993). A novel approach to nonlinear and non-Gaussian Bayesian state estimation. 140:107–113.
- Gouriéroux, C., Monfort, A., and Renault, E. (1993). Indirect inference. *Journal of Applied Econometrics*, 8:S85–S118.
- Green, P. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82(4):711–732.
- Harrison, M. and Krepps, D. (1979). Martingales and arbitrage in multiperiod securities markets. *Journal of Economic Theory*, 20:381–408.
- Harrison, M. and Krepps, D. (1981). Martingales and stochastic integrals in the theory of continuous trading. *Stochastic Processes and Their Applications*, 11:215–260.
- Hastings, W. K. (1970). Monte-carlo sampling methods using markov chains and their applications. 57:97–109.
- Henderson, V., Hobson, D., Howison, S., and Kluge, T. (2005). A comparison of q -optimal option prices in a stochastic volatility model with correlation. *Review of Derivatives Research*, 8:5–25.
- Heston, S. (1993). A closed-form solution for options with stochastic volatility. with applications to bonds and currency options. 6:327–343.
- Hobson, D. (1993). Volatility mis-specification, option pricing and super-replication. *Annals of Applied Probability*, 8(2):193–205.
- Hoffmann, M. (1999). Adaptive estimation in diffusion processes. 79(1):135–163.
- Hull, J. C. and White, A. D. (1987). The pricing of options on assets with stochastic volatilities. *Journal of Finance*, 42(2):281–300.

- Jacod, J. (2000). Non-parametric kernel estimation of the coefficient of a diffusion. *27(1):83–96*.
- Jacquier, E., Polson, N. G., and Rossi, P. (1994). Bayesian analysis of stochastic volatility models. *Journal of Business and Economic Statistics*, 12:371–389.
- 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.
- Karatzas, I. and Shreve, S. E. (1991). *Brownian motion and stochastic calculus*. Springer, New York.
- Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430):773–795.
- Kendall, M. and Stuart, A. (1977). *The advance theory of Statistics. Vol I: Distribution theory (4th edition)*. Charles Criffin & Co.
- Kim, S., Shephard, N., and Chib, S. (1998). Bayesian analysis of stochastic volatility models. *Review of Economic Studies*, 65:361–393.
- Kimura, M. and Ohta, T. (1971). *Theoretical aspects of population genetics*. Princeton University press.
- Kloeden, P. and Platen, E. (1995). *Numerical Solutions of Stochastic Differential Equations*. New York: Springer.
- Knight, F. B. (1981). *Essentials of Brownian motion and diffusion*, volume 18 of *Mathematical Surveys*. American Mathematical Society.
- Kou, S., Sunney Xie, X., and Liu, J. (2006). Bayesian analysis of single molecule experimental data. *Journal of Royal Statistical Society, series C*, 54(3):1–28.
- McAdams, H. and Arkin, A. (1997). Stochastic mechanisms for gene expressions. *Proc. of the National Academy of Science*, 94:814–819.

- Meng, X.-L. and van Dyk, D. (1997). The EM algorithm—an old folk-song sung to a fast new tune. 59(3):511–567. With discussion and a reply by the authors.
- Meng, X.-L. and van Dyk, D. (1998). Fast EM-type implementations for mixed effects models. 60(3):559–578.
- Merton, B. (1995). The theory of rational option pricing. *Bell Journal of Economics and Management Science*, 4:141–183.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A., and Teller, E. (1953). Equation of state calculation by fast computing machines. 21(6):1087–1092.
- Obuhov, A. M. (1959). Description of turbulence in terms of Lagrangian variables. In *Advances in Geophysics, Vol. 6 (Symposium on Atmospheric Diffusion and Air Pollution, Oxford, 1958)*, pages 113–116. Academic Press, New York.
- Oksendal, B. (2000). *Stochastic differential equations*. Springer, 5th edition.
- Ozaki, T. (1992). A bridge between nonlinear time series models and nonlinear stochastic dynamical systems. *Statist. Sinica*, 2:113–135.
- Papaspiliopoulos, O., Fearnhead, P., and Roberts, G. (2006). Particle filters for partially observed diffusions. Submitted.
- Papaspiliopoulos, O. and Roberts, G. (2005). Retrospective MCMC for Dirichlet process hierarchical models. Submitted.
- Papaspiliopoulos, O., Roberts, G., and Sköld (2003). Non-centered parametrisations for hierarchical models and data augmentation. in J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith & M. West, eds 'Bayesian Statistics 7', Oxford University Press, pp. 307-326.
- Pedersen, A. R. (1995). A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. *Scand. J. Statist.*, 22(1):55–71.
- Pinheiro, J. and Bates, D. (1996). Unconstrained parametrizations for variance-covariance matrices. *Statistics and Computing*, 6(3):289–296.

- Pitt, M. K. and Shephard, N. (1999). Filtering via simulation: Auxilliary particle filter. *Journal of American Statistical Association*, 446:590–599.
- Polson, N. and Roberts, G. (1994). Bayes factors for discrete observations from diffusion processes. *Biometrika*, 81(1):11–26.
- Poyadjis, G. (2006). Particle methods for parameter estimation in general state space models. D.Phil thesis, University of Cambridge.
- Prakasa Rao, B. (1999). *Statistical inference for diffusion type processes*. Oxford University press.
- Roberts, G. and Stramer, O. (2001). On inference for partial observed nonlinear diffusion models using the metropolis-hastings algorithm. *Biometrika*, 88(3):603–621.
- Roberts, G. and Stramer, O. (2004). On bayesian analysis of non-linear continuous time autoregressive models. Submitted.
- Roberts, G. O. and Sahu, S. (1999). On convergence of the em algorithm and the gibbs sampler. 9:55–64.
- Rogers, L. C. G. and Williams, D. (1994). *Diffusions, Markov processes and martingales, 2, Ito calculus*. Wiley, Chicester.
- Romano, M. and Touzi, N. (1997). Contingent claims and market completeness in a stochastic volatility model. *Math. Finance*, 7(4):399–412.
- Rubin, D. B. (1984). Bayesianly justifiable and relevant frequency calculations for the applied statistician. 12(4):1151–1172.
- Rubinstein, B. (1995). As simple as one, two, three. *Risk*, 8:44–47.
- Sangalli, L. and Roberts, G. (2006). Latent diffusion models for event hidtory analysis. In preparation.
- Santa Clara, P. (1995). Simulated likelihood estimation of diffusions with an application to the short term interest rate. Unpublished paper, Simon School of Business, University of Rochester.

- Schweizer, M. (1999). A minimality property of the minimal martingale measure. *42(1):27–31.*
- Shephard, N. (2005). *Stochastic Volatility: Selected Readings.* Oxford University Press.
- Smith, A. F. M. and Roberts, G. O. (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *55(1):3–23.*
- Spiegelhalter, D., Thomas, A., Best, N., and Gilks, W. (1996). *BUGS: Bayesian inference Using Gibbs Sampling, Version 0.5, (version ii).*
- 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.*
- Stroock, D. and Varadhan, S. (1987). *Kiyosi Ito selected papers.* Springer-Verlag.
- 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.*
- Tsai, H. and Chan, K. S. (2000). Testing for nonlinearity with partially observed time series. *Biometrika, 87(4):805–821.*
- Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics, 5:177–188.*
- Whaley, R. E. (1993). Derivatives on market volatility: Hedging tools long overdue. *Journal of Derivatives, 1:71–83.*
- Whaley, R. E. (2000). The investor fear gauge. *Journal of portfolio management, 26:12–17.*
- Wiener, N. (1923). Differential space. *2:131–174.*