Bayesian Inference for Stochastic Differential Equations with Application to Finance

Aleksandra Svalova, Newcastle University
Supervisor: Dr Andrew Golightly

April 29, 2014

Abstract

Stochastic differential equations can be used to model many continuous time processes. We examine how these can be exploited in a financial context by considering the stock price process. Fischer Black and Myron Scholes proposed a model of this phenomenon in 1973, assuming that an asset's price follows a geometric Brownian motion. The latter has an analytically tractable stochastic differential equation, which is not the case for many diffusion processes describing financial models. Therefore we consider the Euler-Maruyama discretisation to extend the applicability of our work. We focus on parameter inference within the Bayesian school of thought: simulation from intractable posterior distributions is performed via the Markov chain Monte Carlo method. The discretisation accuracy is then assessed against the analytic solution. The developed model could be applied to sophisticated financial processes, including exotic path-dependent options.

Contents

1 Introduction **2**
1.1 Contextual Environment 2

2 Theoretical Background **4**
2.1 Brownian Motion 4
2.2 Itô Stochastic Processes 5
2.2.1 Generalised Brownian Motion as an SDE 7
2.3 Black-Scholes Model 7
2.3.1 Geometric Brownian Motion as an SDE 8
2.4 Euler-Maruyama Discretisation 10

3 Bayesian Inference **11**
3.1 Introduction 11
3.2 Markov Chain Monte Carlo 12
3.2.1 Background 12
3.2.2 Metropolis-Hastings Algorithm 13
3.2.3 MCMC Using the Analytic Solution 14
3.3 MCMC Using the Euler-Maruyama Discretisation 15
3.4 Imputation via Euler-Maruyama Discretisation 17
3.4.1 Single-site Updating 17
3.4.2 Block Updating 18

4 Applications **20**
4.1 Inference Using the Analytic Solution 20
4.1.1 A Real Data Example 23
4.2 Euler-Maruyama Discretisation 23
4.3 EM Imputation Sampler 28
4.3.1 E-M Block Imputation Sampler 32

5 Summary **41**

Chapter 1

Introduction

1.1 Contextual Environment

The stock price process (SPP) refers to price fluctuations of securities occurring on the stock market. Given the sophistication of its underlying environment the SPP is a highly-complex process, and we only scratch the surface of its non-mathematical context.

Stock markets are equivalent to money-generating vehicles where business shares are sold to investors [12](p.10). Traders (or professionals) sell securities and ensure market equilibrium by maintaining prices for stocks in lower demand, sometimes selling shares to themselves. A firm's occurrence on the exchange, or 'going public', marks a milestone in its development and is a sign of substantial growth. The market has a number of unusual properties. Firstly, a company has little if no direct benefit from changes in its share price. All trading profits are received by shareholders who then determine future re-investment. Secondly, the stock price of a company may change dramatically within a matter of minutes, making no impact on the worth of its assets or capital. These fluctuations are driven by investor demand and traders' (sometimes subjective) perception of the business value. Therefore, it is only plausible to examine this process from a Bayesian point of view.

Mathematically, we describe the SPP by Itô stochastic differential equations (SDEs) as outlined in the famous Black-Scholes (B-S) model [2](p.644). Assuming geometric Brownian motion (GBM) as a model for stock, the B-S theory allows to price European Vanilla-style options. As the associated SDE can be solved analytically we consider it as a starting point for our Bayesian analysis to follow. Within that context, the process parameters require empirical and theoretical information – the likelihood and prior distributions of the mean rate of return μ and volatility $\sigma^2 = 1/\tau$. From this we derive parameter posterior distributions up to proportionality. We then proceed with simulation – a Markov chain Monte Carlo (MCMC) scheme is developed and applied to the analytic solution. A particular choice of prior leads to a tractable full conditional distribution (FCD) of μ, enabling us to simulate from it directly. However, the τ posterior density has a non-standard form, requiring a Metropolis-Hastings update.

We then consider inference under the Euler-Maruyama (E-M) discretisation with a time increment given by the observation intervals. The analytic solution availability enables us to assess and verify the discretisation accuracy. For processes other than the SPP, this approximation may be crude unless the data are observed frequently in time. Thus, we consider a Bayesian imputation scheme where observed low frequency data are augmented by $k-1$ latent values between each pair of observations. We develop two approaches for updating the inter-observation latent path: a simple single-site scheme updating latent values one at a time, and a block updating scheme.

The structure of this report is as follows. Chapter 1 has introduced the contextual background of our research. Chapter 2 outlines the necessary theoretical base and Chapter 3 describes the Bayesian framework to be used in parameter inference. Chapter 4 is the core of this study, outlining the results of our simulation with applications of MCMC to real and synthetic data. Chapter 5 concludes this study and describes future directions to be taken.

Chapter 2

Theoretical Background

2.1 Brownian Motion

Definition 2.1
Standard Brownian Motion $\{W_t, t \geq 0\}$ is a continuous-time, continuous-value stochastic process characterized by the following properties.

1. W_t has continuous trajectories, $W_0 = 0$.
2. For all times $0 = t_0 < t_1 < t_2 < \infty$, the increments of BM are independent and stationary:
$$W_{t_1} - W_{t_0} \perp\!\!\!\perp W_{t_2} - W_{t_1}. \tag{2.1}$$
3. For all times $0 = t_0 < t_1 < \infty$, the non-overlapping increments have a **Gaussian density**:
$$W_{t_1} - W_{t_0} \sim N(0, t_1 - t_0). \tag{2.2}$$

From equation (2.2) one may obtain the conditional distributions of W_t, which can be used to simulate its trajectories at discrete times.

1. The distribution of W_t conditional on the process initial value is
$$W_t | W_0 = 0 \sim N(0, t). \tag{2.3}$$
2. The distribution of $\{W_t | W_s = w_s, s < t\}$ is obtained by noting that $W_t = W_t - W_s + W_s$, and therefore
$$W_t | W_s = w_s \sim N(w_s, t - s). \tag{2.4}$$

Figure 2.1 shows trajectories of standard BM for different values of the time increment Δt. As Δt decreases it is apparent that the process is not differentiable in analytic sense. Notice that W_t may be interpreted as a random walk, as proposed by Ross [9](p.631). For example, suppose X_t is a random walk taking a step of a size Δx each Δt time units, with an equal probability of moving to the right as to the left. Also consider

$$Y_i = \begin{cases} 1, & Pr(Y_i = 1) = 0.5, \\ -1, & Pr(Y_i = -1) = 0.5. \end{cases}$$

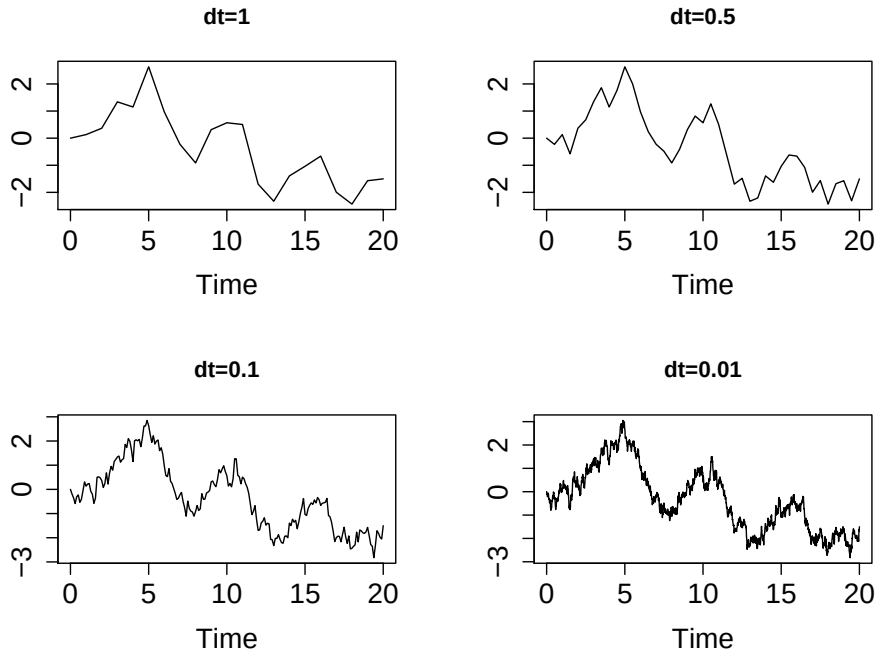

Figure 2.1: Plots of standard BM trajectories.

Therefore,

$$X_t = \Delta x(Y_{\Delta t} + Y_{2\Delta t} \ldots + Y_{n\Delta t}), \quad n = \frac{t}{\Delta t}, \tag{2.5}$$

and

$$E[X_t] = 0, \quad Var(X_t) = \frac{t(\Delta x)^2}{\Delta t}.$$

As n increases (or Δt decreases) the central limit theorem implies that

$$X_t \sim N\left(0, \frac{t(\Delta x)^2}{\Delta t}\right). \tag{2.6}$$

Allowing $\Delta x = \sqrt{\Delta t}$, such that $\Delta x \to 0$ as $\Delta t \to 0$,

$$X_t \sim N(0, t) \text{ as } \Delta t \to 0. \tag{2.7}$$

Therefore, as Δx and Δt to tend to 0 X_t tends to BM W_t implying that X_t is a continuous-value function [9] (p.632).

2.2 Itô Stochastic Processes

In order to formulate the SPP as a stochastic differential equation (analogous to the Black-Scholes model, Section 2.3), we consider the Itô stochastic processes. For $\{X_t, t \geq$

$0\}$, define $f(t)$ as a time-dependent stochastic process satisfying the Riemann integral as follows [1](p.335):

$$\int_a^b E(f^2(t))dt \;<\; \infty, \tag{2.8}$$

$$f(t) = f(t, X_t), \quad t \;\in\; [a, b]\,. \tag{2.9}$$

Given that (2.9) exists, assume

$$\begin{aligned} a &= t_1 < t_2 < \ldots < t_{k+1} = b \in [a, b], \\ \Delta t &= t_{i+1} - t_i = \frac{(b-a)}{k}, \\ \Delta W_{t_i} &= W_{t_{i+1}} - W_{t_i}. \end{aligned}$$

Under these conditions, the Itô stochastic integral of f is defined as

$$\int_a^b f(t)dW_t = \lim_{k\to\infty} \sum_{i=1}^{k} f(t_i)\Delta W_{t_i}. \tag{2.10}$$

Here, $\lim_{k\to\infty}$ denotes the mean square convergence [1](p.336), such that if $\mathbb{I} = \int_a^b f(t)dW_t$. Then

$$\lim_{k\to\infty} F_k = \mathbb{I} \quad \Rightarrow \quad \lim_{k\to\infty} E[(F_k - \mathbb{I})^2] = 0.$$

Following this, an Itô SDE satisfied by a stochastic process $\{X_t, t \geq 0\}$ driven by Brownian motion is defined as

$$dX_t = \alpha(X_t, \theta)dt + \sqrt{\beta(X_t, \theta)}dW_t, \tag{2.11}$$

where α is the drift and β is the diffusion coefficient (stochastic term). Both α and β depend on X_t and a parameter (vector) θ; dW_t is an increment of standard Brownian motion. Integrating both sides of dX_t we obtain

$$X_t = x_0 + \int_0^t \alpha(X_\tau, \theta)d\tau + \int_0^t \sqrt{\beta(X_\tau, \theta)}dW_\tau, \tag{2.12}$$

where x_0 is the initial value. The second and third terms of the X_t right hand side are deterministic and stochastic integrals respectively- removing the latter will make X_t an ordinary differential equation. Finally, in order to derive SDEs satisfied by a transformation of Itô process, we define the Itô chain rule:

$$dG(X_t) = \left(\alpha(X_t)G_x(X_t) + \frac{1}{2}\beta(X_t)G_{xx}(X_t)\right)\, dt + G_x(X_t)\sqrt{\beta(X_t)}\, dW_t. \tag{2.13}$$

Here, $G_x(X_t)$ denotes the derivative of X_t with respect to x; similarly, $G_{xx}(X_t)$ is the double-derivative. This result will be used in Section 2.3 to solve the Black-Scholes SDE.

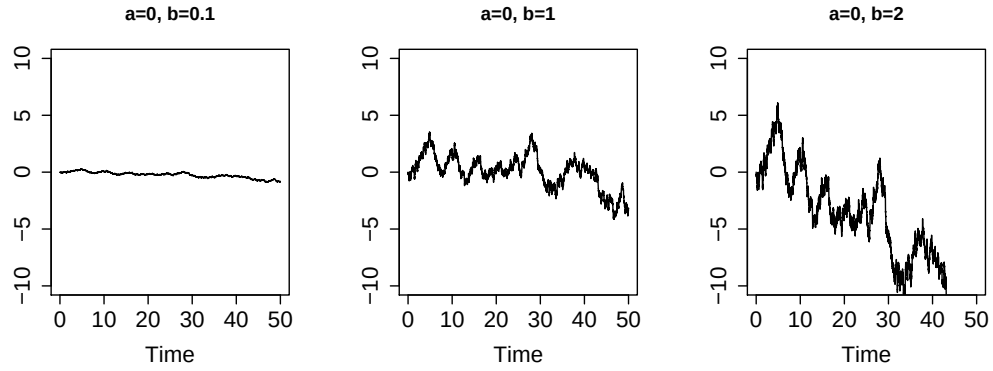

Figure 2.2: Generalised Brownian Motion

2.2.1 Generalised Brownian Motion as an SDE

Consider a generalised BM $\{X_t, t \geq 0\}$, whose SDE is defined as

$$\begin{aligned} dX_t &= adt + bdW_t, \quad X_0 = x_0, \\ X_t &= x_0 + at + bW_t, \quad x_0, a, b \in \mathbb{R}\,. \end{aligned} \tag{2.14}$$

This implies the following properties of X_t:

1. $X_0 = x_0$.
2. The increments of $\{X_t, t \geq 0\}$ are stationary and independent.
3. $X_t \sim N(x_0 + at, tb^2)$.

As before, the conditional distribution $X_t|X_s = x_s$ is:

$$\begin{aligned} X_t &= X_t - X_s + X_s = X_s + a(t-s) + b(W_t - W_s), \text{ therefore} \\ X_t|X_s &= x_s \sim N(x_s + a(t-s), b^2(t-s)). \end{aligned}$$

Figure 2.2 illustrates the behaviour of X_t under changes in b (a is fixed at zero). Larger b values cause trajectories to become more volatile due to an increase in the stochastic increment W_t. Two issues prohibit this form of BM from being used directly in our model. Firstly, X_t is defined on the entire real line which conflicts with the SPP data support. Crucially, the interval distribution of X_t is homogeneous which is unsatisfactory in our case. This motivates the use of geometric BM, defined in the next Section.

2.3 Black-Scholes Model

The Black-Scholes (B-S) model was first published in 1973, in a paper 'The Pricing of Options and Corporate Liabilities' [2](p.644). It described a model for the stock market and a method of pricing Vanilla-style options via SDEs. The authors, Myron Black and Fischer Scholes, proposed that in order to reduce risk of an option, an investor should

buy and sell its underlying asset- which is known as delta hedging. Although not without imperfections, the B-S model is still recognised as one of the most popular for pricing securities. Again, we will only consider the tip of the econometric iceberg that their work is, in particular the geometric BM representation of the SPP.

Definition 2.2
We define $\{S_t, t \geq 0\}$ as **geometric Brownian motion (GBM)** with parameters μ and σ^2, if S_t depends continuously on time t and the following assumptions hold for all times $0 \leq t_0 < t_1 < t_2 < \infty$:

1. $S_0 > 0$.

2. Ratios defined over non-overlapping time intervals are independent:

$$\frac{S_{t_2}}{S_{t_1}} \perp\!\!\!\perp \frac{S_{t_1}}{S_{t_0}}.$$

3. The process ratios follow a **Log-normal distribution**

$$\frac{S_{t_1}}{S_{t_0}} \sim LN\left(\left(\mu - \frac{\sigma^2}{2}\right)\Delta t, \sigma^2 \Delta t\right), \quad \Delta t = t_1 - t_0. \tag{2.15}$$

2.3.1 Geometric Brownian Motion as an SDE

Consider the task of formulating geometric Brownian motion as an (SDE). It can be shown that $S_t = e^{X_t}$ is a geometric BM, where

$$X_t = x_0 + \left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W_t. \tag{2.16}$$

Section 2.2.1 gives the SDE satisfied by X_t as

$$dX_t = \left(\mu - \frac{\sigma^2}{2}\right)dt + \sigma dW_t. \tag{2.17}$$

Hence, we derive an SDE for S_t by applying Itô formula with $G(t,x) = e^x$. Let

$$G_t = \frac{\partial G(t,x)}{\partial t} \quad \text{and} \quad G_x = \frac{\partial G(t,x)}{\partial x}.$$

Thus,

$$G_t(t, X_t) = 0\,, \qquad G_x(t, X_t) = e^{X_t}\,, \qquad G_{xx}(t, X_t) = e^{X_t}.$$

Applying Itô formula produces

$$\begin{aligned} dS_t = d\left(e^{X_t}\right) &= \left(\left[\mu - \frac{\sigma^2}{2}\right]e^{X_t} + \frac{1}{2}\sigma^2 e^{X_t}\right)dt + \sigma e^{X_t}\, dW_t, \\ \Rightarrow d\left(e^{X_t}\right) &= \mu e^{X_t}\, dt + \sigma e^{X_t}\, dW_t\,. \end{aligned} \tag{2.18}$$

Writing $S_t = e^{X_t}$ leads to the SDE for geometric Brownian motion:

$$dS_t = \mu S_t\, dt + \sigma S_t\, dW_t. \tag{2.19}$$

Note that we may solve this SDE explicitly, obtaining a closed-form expression for S_t. To perform this, apply Itô formula with $\boldsymbol{G(t,x) = \log(x)}$ and define

$$G_t(t, S_t) = 0\,, \qquad G_x(t, S_t) = \frac{1}{S_t}\,, \qquad G_{xx}(t, S_t) = -\frac{1}{S_t^2}\,.$$

Therefore,

$$\begin{aligned} d(\log(S_t)) &= \left(\mu S_t \frac{1}{S_t} - \frac{1}{2} S_t^2 \frac{\sigma^2}{S_t^2}\right) dt + \sigma S_t \frac{1}{S_t}\, dW_t \\ &= \left(\mu - \frac{1}{2}\sigma^2\right) dt + \sigma\, dW_t\,. \end{aligned} \tag{2.20}$$

Integrating both sides of (2.20) produces

$$\begin{aligned} \int_0^t d\left(\log(S_\tau)\right) &= \int_0^t \left(\mu - \frac{1}{2}\sigma^2\right) d\tau + \int_0^t \sigma\, dW_\tau, \\ \Rightarrow \log\left(\frac{S_t}{S_0}\right) &= \left(\mu - \frac{1}{2}\sigma^2\right) t + \sigma\, W_t, \\ \Rightarrow S_t &= S_0 \exp\left\{\left(\mu - \frac{1}{2}\sigma^2\right) t + \sigma\, W_t\right\}. \end{aligned} \tag{2.21}$$

Hence, when formulated as an SDE the Black-Scholes model of the SPP has a tractable analytic solution.

In what follows we consider the logarithmic transformation of the stock price ratios. This has the advantage of using Gaussian density within the MCMC sampler. It also leads to a more comprehensive understanding about the growth and decline of prices on the stable scale that the transformation allows.

Suppose that stock prices are observed every Δt time units. Therefore, by equation (2.21) the price ratios follow a **Log-normal** distribution:

$$S_t = S_0 \exp\left\{\left(\mu - \frac{1}{2}\sigma^2\right) \Delta t + \sigma\, W_t\right\}. \tag{2.22}$$

Suppose we have a set of data $\{s_{t_0}, s_{t_1}, \ldots s_{t_n}\}$, and thus define the log-ratios x_{t_i} and a data vector $\boldsymbol{x} = (x_{t_1}, x_{t_2}, \ldots x_{t_n})^T$ as a realisation of

$$X_{t_i} = \log\left(\frac{S_{t_i}}{S_{t_{i-1}}}\right) \sim N\left\{\left(\mu - \frac{\sigma^2}{2}\right)\Delta t, \sigma^2 \Delta t\right\}.$$

This result enables the convenience of the Gaussian density within data likelihood, which we make an extensive use of within Bayesian inference in Chapters 3 and 4.

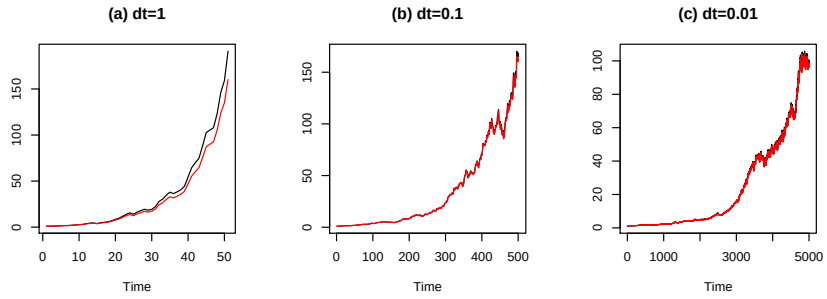

Figure 2.3: GBM (black) and E-M (red) trajectories, $t = 50, \mu = 0.1, \sigma = 0.1$.

2.4 Euler-Maruyama Discretisation

Although the SPP has a luxury of a closed-form solution to its underlying SDE, many financial processes are analytically intractable, an example of which is the stochastic volatility model. We, therefore, consider approximation methods, namely the Euler-Maruyama (E-M) discretisation.

Assume an SDE dX_t and the task of evaluating its integral on $[0, T]$, where

$$dX_t = a(t, X_t)dt + b(t, X_t)dW_t, \quad 0 = t_0 < t_1 < \ldots < t_n = T.$$

Approximated values of X_t, $\{y_0 < y_1 < \ldots < y_n, \quad y_i \equiv y_{t_i}\}$, are evaluated at their respective time points [10](p.4). Under this setup, the discretisation of X_t is as follows:

$$\begin{aligned} y_0 &= x_0, \\ y_{i+1} &= y_i + a(t_i, Y_i)\Delta t_{i+1} + b(t_i, y_i)\Delta W_{i+1}, \\ \Delta t_{i+1} &= t_{i+1} - t_i, \quad \Delta W_i = Z\sqrt{\Delta t_i}, \quad Z \sim N(0, 1). \end{aligned} \tag{2.23}$$

Therefore, the E-M method generates an approximation of the solution to dX_t as a set of values $\{y_0, \ldots, y_i\}$, which depends on random Z the quantities [10](p.5).

The following example verifies that the discretisation can be satisfactory for small Δt. Consider generating synthetic data from the Black-Scholes model with parameters

$$\mu = 0.1, \quad \sigma = 0.1, \quad t = 50.$$

We derive the E-M approximation and overlay the resulting trajectories, assessing accuracy against the analytic solution. Figure 2.3 shows simulation results fordifferent Δt values. Note that the approximation is reasonable for unit interval times, and as Δt decreases the two trajectories become almost identical. We perform equivalent analysis for the Bayesian parameter inference as the main subject of the next Chapter.

Chapter 3

Bayesian Inference

3.1 Introduction

In this Chapter we describe Bayesian inference for the parameters governing the SPP. Initially, we outline fundamental principles of the Bayesian approach and derive parameter posterior distributions. We then introduce the Markov chain Monte Carlo (MCMC) sampling method and describe our scheme using the analytic solution. An equivalent algorithm is developed based on the E-M discretisation. Furthermore, we introduce imputation within the E-M scheme tackling the inaccuracy issue associated with low observation frequency.

Definition 3.1
Bayes theorem allows combining information from empirical studies with prior expert knowledge to build inferences about underlying data parameters. For this, assume an expert-elicited *prior distribution* for parameter (vector) θ with an associated density $\pi(\theta)$, and a set of observations $\boldsymbol{x} = \{x_1, x_2, \ldots, x_n\}$ with a joint density $f(\boldsymbol{x}|\theta)$. We define the *likelihood* as

$$L(\theta|\boldsymbol{x}) = f(\boldsymbol{x}|\theta). \tag{3.1}$$

Then, by Bayes' theorem the *posterior distribution* of $\theta|\boldsymbol{x}$ is:

$$\pi(\theta|\boldsymbol{x}) = \frac{\pi(\theta)f(\boldsymbol{x}|\theta)}{f(\boldsymbol{x})} \propto \pi(\theta)f(\boldsymbol{x}|\theta). \tag{3.2}$$

In other words,

'the posterior is proportional to the prior times the likelihood'.

Here only the kernel forms of $\pi(\theta), L(\boldsymbol{x}|\theta)$ and $\pi(\theta|\boldsymbol{x})$ are considered, as computation of the posterior proportionality constant $f(\boldsymbol{x})$ can be difficult. This is well-suited to MCMC simulation due to its requirement of the equilibrium density form only (further discussed in the next Section).

Consider now a geometric BM corresponding to the SPP, governed by the mean rate of

return μ and precision $\tau = \sigma^{-2}$. Recalling that $x_{t_i} = \log(s_{t_i}/s_{t_{i-1}})$ with a corresponding Gaussian density, we define the model likelihood as:

$$\begin{aligned} L(\mu, \tau \,|\boldsymbol{x}) &= \prod_{i=1}^{n} \sqrt{\frac{\tau}{2\pi}} \exp\left\{-\frac{\tau}{2\Delta t}\left(x_{t_i} - \left(\mu - \frac{1}{2\tau}\right)\Delta t\right)^2\right\} \\ &= \left(\frac{\tau}{2\pi}\right)^{\frac{n}{2}} \exp\left\{\sum_{i=1}^{n}\left(x_{t_i} - \left(\mu - \frac{1}{2\tau}\right)\Delta t\right)^2\right\} \\ &\propto \tau^{\frac{n}{2}} \exp\left\{\sum_{i=1}^{n}\left(x_{t_i} - \left(\mu - \frac{1}{2\tau}\right)\Delta t\right)^2\right\}. \end{aligned} \tag{3.3}$$

We assign μ and τ (theoretical) prior distributions satisfying their data support and variability:

$$\pi(\mu) \equiv N(a, 1/b) \quad \text{and} \quad \pi(\tau) \equiv Ga(g, h). \tag{3.4}$$

As μ and τ are assumed to be independent *apriori* their joint prior distribution is the product of marginals:

$$\begin{aligned} \pi(\mu, \tau) &= \pi(\mu)\pi(\tau) \\ &= \frac{1}{\sqrt{2\pi b^2}} \exp\left\{-\frac{b(\mu - a)^2}{2}\right\} \frac{h^g \tau^{g-1} e^{-h\tau}}{\Gamma(g)} \\ &= \frac{h^g}{\Gamma(g)\sqrt{2\pi b^2}} \tau^{g-1} \exp\left\{-\frac{b(\mu - a)^2}{2} h\tau\right\} \\ &\propto \tau^{g-1} \exp\left\{-\frac{b(\mu - a)^2}{2} h\tau\right\}. \end{aligned} \tag{3.5}$$

Therefore, by Bayes theorem

$$\begin{aligned} \pi(\mu, \tau|\boldsymbol{x}) &\propto \pi(\mu)\pi(\tau)L(\mu, \tau|\boldsymbol{x}) \\ &\propto \tau^{g-1} \exp\left\{-\frac{b(\mu - a)^2}{2} h\tau + \sum_{i=1}^{n}\left(x_{t_i} - \left(\mu - \frac{1}{2\tau}\right)\Delta t\right)^2\right\}. \end{aligned} \tag{3.6}$$

The full conditional distribution of μ is of a standard form:

$$\mu|\tau, \boldsymbol{x} \sim N\left(\frac{2ab + 2n\overline{\boldsymbol{x}}\tau + n\Delta t}{2(b + n\Delta t\tau)}, \frac{1}{b + n\Delta t\tau}\right). \tag{3.7}$$

However, the joint parameter density (3.6) and full conditional posterior of τ are unavailable, thus simulation from them is non-trivial. We, therefore, proceed by considering the design of an MCMC scheme.

3.2 Markov Chain Monte Carlo

3.2.1 Background

Markov chain Monte Carlo (MCMC) is arguably the most essential tool in Bayesian statistics. It was developed soon after the ordinary Monte Carlo method (a framework of

repeated sampling via computer simulation) by Stanislaw Ulam and Nicholas Metropolis in 1946. Despite its glamorous connotations, this method has little to do with the destination – Metropolis came up with a code name after the casino where Ulam's uncle was gambling. Its initial use was in the Manhattan Project of building nuclear weapons for World War 2. Metropolis simulated thermodynamic equilibrium by implementing a particular algorithm until convergence. This led to a discovery that MCMC requires only some Markov chain with the same equilibrium distribution- the normalising constant $f(\boldsymbol{x})$ becomes redundant. This method was called the Metropolis algorithm, which Hastings generalised in 1970, obtaining the modern form of the **Metropolis-Hastings (M-H)** algorithm [3](p.3).
Later, the **Gibbs sampler** was introduced by Geman and Geman [7](p.721) describing a special case of M-H under the availability of the full conditional distribution. Gelfand and Smith made the wider Bayesian community aware of the Gibbs sampler, which up to that time had been known only in spatial statistics circles. 'It was rapidly realised that most Bayesian inference could be done by MCMC- and very little without it' [3](p.3).

3.2.2 Metropolis-Hastings Algorithm

Key to M-H algorithm is the **M-H update**. Suppose the target distribution of the MCMC $\pi(\cdot|\boldsymbol{x})$ is known up to a proportionality constant. Therefore, $\pi(\cdot|\boldsymbol{x})$ is a non-negative function that could be integrated (summed in a discrete case) to a finite, non-zero value. The M-H update proceeds as follows:

1. Given a Markov chain with the current state θ, draw a proposed move to ϕ from its conditional probability density $q(\phi|\theta)$.

2. Calculate the **Hastings ratio**:

$$r(\phi, \theta) = \frac{\pi(\phi|\boldsymbol{x})q(\theta|\phi)}{\pi(\theta|\boldsymbol{x})q(\phi|\theta)}. \tag{3.8}$$

3. Accept the proposed move ϕ with probability

$$\alpha(\phi, \theta) = \min\{1, r(\phi, \theta)\}. \tag{3.9}$$

Therefore, a M-H update makes a proposal move to the new state ϕ with probability $\alpha(\phi, \theta)$ [3](p.22-25).

The above can be seen as a special case of a more general algorithm, that successively updates the components of $\theta = (\theta_1, \theta_2, \ldots, \theta_p)^T$. In particular, a proposed value $\theta^{(j)}$ is obtained from $\pi(\theta^{(j-1)})$ by (successive) component-wise transitions using a M-H step:

- $\theta_1^{(j)} \sim \pi(\theta_1|\theta_2^{(j-1)}, \ldots, \theta_p^{(j-1)}, \boldsymbol{x})$
- $\theta_2^{(j)} \sim \pi(\theta_2|\theta_1^{(j)}, \theta_3^{(j-1)}, \ldots, \theta_p^{(j-1)}, \boldsymbol{x})$
  $\vdots$

- $\theta_p^{(j)} \sim \pi(\theta_1|\theta_1^{(j)}, \ldots, \theta_{p-1}^{(j)}, \boldsymbol{x})$

This is analogous to the original form of the Metropolis algorithm, which used component-wise transitions between successive draws from the proposal distributions $q_i(\phi_i|\theta_i^{(j-1)})$ (ϕ_i is the parameter of interest). In the case of Gibbs sampler, the proposal distribution is the FCD (which is available), therefore the acceptance probability is 1.

3.2.3 MCMC Using the Analytic Solution

Here we outline the design of an MCMC scheme simulating the SPP parameter posterior distribution based on the analytic solution. Recall that the full conditional posterior of μ is available, therefore we shall use a Gibbs sampler. However, the FCD of τ is non-trivial, demanding a M-H update and a suitable proposal distribution. Assume the proposed value of precision takes the form $\tau^* = \tau^{(j-1)}w$, where w is a random variable defined on the positive real line. In particular,

$$\tau^* = \tau^{(j-1)}w, \quad w \sim LN(0, v^2) \quad \Rightarrow \quad \log(w) \sim N(0, v^2). \tag{3.10}$$

Therefore,

$$\log(\tau^*) = \log(\tau^{(j-1)}) + \log(w) \quad \Rightarrow \quad \log(\tau^*) \sim N(\log(\tau^{(j-1)}), v^2). \tag{3.11}$$

Recall that

$$\mu|\tau, \boldsymbol{x} \sim N\left(\frac{2ab + 2n\overline{\boldsymbol{x}}\tau + n\Delta t}{2(b + n\Delta t\tau)}, \frac{1}{b + n\Delta t\tau}\right). \tag{3.12}$$

Then the M-H algorithm is as follows.

1. Set $j = 1$. Initialise μ and τ within their distribution support, normally at prior means. Therefore,
$$\mu^{(0)} = a, \quad \tau^{(0)} = g/h.$$
2. Draw $\mu^{(j)}$ from (3.12).
3. Draw τ^* from the M-H update.
   - By Equation (3.10), $\tau^* = \tau^{(j-1)}w$, where $w \sim LN(0, v^2)$.
   - Therefore, $\log(\tau^*) \sim N(\log(\tau^{(j-1)}), v^2)$ and
$$\begin{aligned} q(\log(\tau^*)|\log(\tau^{(j-1)})) &= \frac{1}{\sqrt{(2\pi v^2)}} \exp\left\{-\frac{(\log(\tau^*) - \log(\tau^{(j-1)}))^2}{2v^2}\right\} \\ &\propto \exp\left\{-\frac{(\log(\tau^*) - \log(\tau^{(j-1)}))^2}{2v^2}\right\}. \end{aligned} \tag{3.13}$$
   - We only require the proposal density up to proportionality as the normalising constants will cancel in the M-H ratio.

- Evaluate the acceptance probability $\alpha(\tau^*|\tau^{(j-1)})$:

$$\begin{aligned}\alpha(\tau^*|\tau^{(j-1)}) &= \min\left\{1, \frac{\pi(\tau^*|\mu^{(j)}, \boldsymbol{x})}{\pi(\tau^{(j-1)}|\mu^{(j)}, \boldsymbol{x})} \times \frac{q(\tau^{(j-1)}|\tau^*)}{q(\tau^*|\tau^{(j-1)})}\right\} \\ &= \min\left\{1, \frac{\pi(\tau^*|\mu^{(j)}, \boldsymbol{x})}{\pi(\tau^{(j-1)}|\mu^{(j)}, \boldsymbol{x})} \times \frac{\tau^*}{\tau^{(j-1)}}\right\}. \qquad (3.14)\end{aligned}$$

- Accept $\tau^* = \tau^{(j)}$ with probability $\alpha(\tau^*|\tau^{(j-1)})$, and retain the previous chain value otherwise.

4. Set $j \to j+1$ and return to Step 2.

Note that $Var(\log(\tau^*)) = v^2$ is likely to be unknown and would demand tuning. The next Chapter illustrates how this can be implemented within a practical application of the above scheme. However, we will now outline an equivalent sampling method of the SPP model based on the E-M discretisation.

3.3 MCMC Using the Euler-Maruyama Discretisation

Recalling Equation (2.23), let

$$\Delta S_t = \mu S_t \Delta t + \sigma S_t \Delta W_t, \quad \text{therefore} \qquad (3.15)$$

$$S_{t+\Delta t}|S_t = s_t \sim N(s_t + \mu s_t \Delta t, \sigma^2 S_t^2 \Delta t). \qquad (3.16)$$

Define the data vector of $(n+1)$ stock price observations as $\boldsymbol{s} = (s_{t_0}, s_{t_1}, \ldots s_{t_n})^T$, 'measured' at equally spaced time intervals $\Delta t = t_{i+1} - t_i$.
In order to reduce the computational burden associated with the S_t density we employ its linear transformation instead. Define $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)^T$, where

$$x_i = \frac{S_{t_i} - S_{t_{i-1}}}{S_{t_{i-1}}\sqrt{\Delta t}} \sim N(\mu\sqrt{\Delta t}, 1/\tau). \qquad (3.17)$$

We may derive the parameter likelihood by multiplying over the observations as follows:

$$\begin{aligned}L(\mu, \tau|\boldsymbol{x}) &\propto \tau^{\frac{n}{2}} \exp\left\{-\frac{\tau}{2}\sum_{i=1}^{n}\left(x_i - \mu\sqrt{\Delta t}\right)^2\right\} \\ &\propto \tau^{\frac{n}{2}} \exp\left\{-\frac{\tau}{2}\sum_{i=1}^{n}\left([x_i - \overline{\boldsymbol{x}}] + [\overline{\boldsymbol{x}} - \mu\sqrt{\Delta t}]\right)^2\right\} \\ &\propto \tau^{\frac{n}{2}} \exp\left\{-\frac{n\tau}{2}\left(s^2 + \left(\overline{\boldsymbol{x}} - \mu\sqrt{\Delta t}\right)^2\right)\right\}. \qquad (3.18)\end{aligned}$$

To proceed with Bayesian analysis, we assign μ and τ the following prior distributions (independent *apriori*):

$$\mu \sim N(a, 1/b), \quad \tau \sim Ga(g, h). \qquad (3.19)$$

Therefore, by Bayes' theorem

$$\begin{aligned}
\pi(\boldsymbol{\mu}, \boldsymbol{\tau}|\boldsymbol{x}) &\propto \pi(\mu,\tau)L(\mu,\tau|\boldsymbol{x}) \\
&\propto \sqrt{\frac{b}{2\pi}}\exp\left\{-\frac{b(\mu-a)^2}{2}\right\}\frac{h^g\tau^{g-1}\exp\{-\tau h\}}{G(g)}\times \\
&\times \tau^{\frac{n}{2}}\exp\left\{-\frac{n\tau}{2}\left(s^2+\left(\overline{\boldsymbol{x}}-\mu\sqrt{\Delta t}\right)^2\right)\right\} \\
&\propto \tau^{\frac{n}{2}+g-1}\exp\left\{-\tau h-\frac{b(\mu-a)^2}{2}-\frac{n\tau\left(s^2+(\overline{\boldsymbol{x}}-\mu\sqrt{\Delta t})^2\right)}{2}\right\}. \quad (3.20)
\end{aligned}$$

The posterior conditional distribution of τ is now trivially derived:

$$\pi(\tau|\mu,\boldsymbol{x}) \propto \tau^{\frac{n}{2}+g-1}\exp\left\{-\tau\left(h+\frac{n\left(s^2+(\overline{\boldsymbol{x}}-\mu\sqrt{\Delta t})^2\right)}{2}\right)\right\}, \text{ thus} \quad (3.21)$$

$$\tau|\mu,\boldsymbol{x} \sim Ga\left(\frac{n}{2}+g, h+\frac{n}{2}\left(s^2+\left(\overline{\boldsymbol{x}}-\mu\sqrt{\Delta t}\right)^2\right)\right), \quad (3.22)$$

where s^2 is the sample variance. For the full conditional of μ:

$$\begin{aligned}
\pi(\mu|\tau,\boldsymbol{x}) &\propto \exp\left\{-\frac{b(\mu-a)^2}{2}-\frac{n\tau\left(\overline{\boldsymbol{x}}-\mu\sqrt{\Delta t}\right)^2}{2}\right\} \\
&\propto \exp\left\{-\frac{b\mu^2-2ab\mu+a^2b+n\tau\overline{\boldsymbol{x}}^2-2n\tau\mu\overline{\boldsymbol{x}}\sqrt{\Delta t}+\mu^2n\tau\Delta t}{2}\right\} \\
&\propto \exp\left\{-\frac{\left(\mu^2(b+n\tau\Delta t)-2\mu\left(ab+n\tau\overline{\boldsymbol{x}}\sqrt{\Delta t}\right)+c\right)}{2}\right\} \\
&\propto \exp\left\{-\frac{(b+n\tau\Delta t)}{2}\left(\mu^2-\frac{2\mu\left(ab+n\tau\overline{\boldsymbol{x}}\sqrt{\Delta t}\right)}{b+n\tau\Delta t}+c\right)\right\}. \quad (3.23)
\end{aligned}$$

Therefore,

$$\mu|\tau,\boldsymbol{x} \sim N\left(\frac{ab+\overline{\boldsymbol{x}}n\tau\sqrt{\Delta t}}{b+n\tau\Delta t}, \frac{1}{b+n\tau\Delta t}\right). \quad (3.24)$$

Both parameters now have trivial posterior densities, therefore enabling the use of a Gibbs sampler (only) which is less burdensome to compute. We proceed as follows:

1. Initialise $\mu^{(0)}$ and $\tau^{(0)}$ at their prior means.
2. At iteration h:
   - Draw $\mu^{(h)} \sim \pi(\mu|\tau^{(h-1)},\boldsymbol{x})$ from (3.3).

- Draw $\tau^{(h)} \sim \pi(\tau|\mu^{(h)}, \boldsymbol{x})$ from (3.22).

3. Let $h := h + 1$ and go to Step 2.

3.4 Imputation via Euler-Maruyama Discretisation

In this section, we examine the effect of data imputation within Euler-Maruyama discretisation on parameter inference. Our analysis in Chapter 4 will illustrate that using observations at integer times provides a fairly consistent estimation of the μ posterior, given the particular underlying process. However, this approximation may be potentially unsatisfactory in practice, especially for the distribution of τ. To overcome this we consider data imputation using the following method. Suppose an interval $(i, i+1]$, partitioned in k equally-spaced sub-intervals:

$$i = t_{i,0} < t_{i,1} < \ldots t_{i,k} = i + 1, \text{ such that } t_{i,j} - t_{i,j-1} = \Delta t = \frac{1}{k}. \tag{3.25}$$

We **introduce $\boldsymbol{k-1}$ 'latent' time points** between every pair of observations (s_t, s_{t+1}). Therefore,

$$\begin{aligned} \boldsymbol{s} &= (s_0, s_1, \ldots, s_n)^T, \\ S_l &= \left(S_{t_{0,1}}, S_{t_{0,2}}, \ldots, S_{t_{0,k-1}}, S_{t_{1,1}}, \ldots S_{t_{1,k-1}}, \ldots, S_{t_{n-1,k-1}}\right), \\ \hat{S} &= \left(s_0, S_{t_{0,1}}, S_{t_{0,2}}, \ldots, S_{t_{0,k-1}}, s_1, S_{t_{1,1}}, S_{t_{1,2}}, \ldots S_{t_{1,k-1}}, s_2, \ldots, S_{t_{n-1,k-1}}, s_n\right)^T, \end{aligned} \tag{3.26}$$

where S_l denotes a vector of $\boldsymbol{n(k-1)}$ **latent** quantities and $\hat{S}$ is a vector combining $\boldsymbol{(n+1)}$ **observed** and $\boldsymbol{n(k-1)}$ **latent** data points. Under Bayesian analysis, the unobserved quantities are equivalent to unknown parameters, such as μ and τ, therefore require a distribution. As μ and τ are independent apriori, the (overall) joint conditional posterior density is

$$\pi(S_l, \mu, \tau|\boldsymbol{s}) \propto \pi(\mu)\pi(\tau) \prod_{i=0}^{n-1} \prod_{j=0}^{k-1} \pi\left(S_{t_{i,j+1}}|S_{t_{i,j}}, \mu, \tau\right), \tag{3.27}$$

where $\hat{S}_{t_{i,j}}$ denotes the approximate value of $\hat{S}$ with $\hat{S}_{t_{i,0}} \equiv s_{t_i}$. Therefore, the conditional distributions of the latent observations will be proportional to (3.27). There are many possible ways to update such latent parameters, and in the next two sections we will look at the single-site and block updating mechanisms. Note that the full conditional distributions of μ and τ are tractable with forms analogous to (3.22) and (3.24).

3.4.1 Single-site Updating

In order to simulate the latent data and SPP parameters, we define a Gibbs sampler as follows:

1. Initialise all unknown quantities, let $\Delta t = 1/k$.

2. At iteration h:

   - Draw $\mu^{(h)} \sim \pi(\mu|\tau^{(h-1)}, \hat{S}^{(h-1)})$.
   - Draw $\tau^{(h)} \sim \pi(\tau|\mu^{(h)}, \hat{S}^{(h-1)})$.
   - Define a block as two consecutive observations and $(k-1)$ latent values between them obtaining a vector of size $(k+1)$: $\{s_i, \hat{S}_{t_{i,1}}, \hat{S}_{t_{i,2}}, \ldots, \hat{S}_{t_{i,k-1}}, s_{i+1}\}$.
   - Propose $\hat{S}_{t_{i,j}} \sim N\left(\frac{1}{2}\left(\hat{S}^{(h)}_{t_{i,j-1}} + \hat{S}^{(h-1)}_{t_{i,j+1}}\right), \Delta t \hat{S}^2_{t_{i,j-1}}/(2\tau^{(h)})\right)$ [6](p.180).
   - Accept with probability $\alpha(\hat{S}^*_{t_{i,j}}|\hat{S}^{(h-1)}_{t_{i,j}})$:

$$\alpha = \min\left\{1, \frac{\pi\left(\hat{S}^*_{t_{i,j}}|\hat{S}^{(h)}_{t_{i,j-1}}, \hat{S}^{(h-1)}_{t_{i,j+1}}, \mu^{(h)}, \tau^{(h)}\right) q\left(\hat{S}^{(h-1)}_{t_{i,j}}|\hat{S}^{(h)}_{t_{i,j-1}}, \hat{S}^{(h-1)}_{t_{i,j+1}}, \mu^{(h)}, \tau^{(h)}\right)}{\pi\left(\hat{S}^{(h-1)}_{t_{i,j}}|\hat{S}^{(h)}_{t_{i,j-1}}, \hat{S}^{(h-1)}_{t_{i,j+1}}, \mu^{(h)}, \tau^{(h)}\right) q\left(\hat{S}^*_{t_{i,j}}|\hat{S}^{(h)}_{t_{i,j-1}}, \hat{S}^{(h-1)}_{t_{i,j+1}}, \mu^{(h)}, \tau^{(h)}\right)}\right\}.$$

   - Define $\hat{S}^{(h)}_{t_{i,j}} = \hat{S}^*_{t_{i,j}}$ with probability $\alpha(\hat{S}^*_{t_{i,j}}|\hat{S}^{(h-1)}_{t_{i,j}})$ and $\hat{S}^{(h-1)}_{t_{i,j}}$ otherwise.

3. Let $h := h + 1$ and go to Step 2.

Therefore, each of the proposed $\hat{S}^*_{t_{i,j}}$ latent quantities depend on their immediate neighbour points. Note that the posterior distribution of τ will be heavily influenced by imputation: both the mean and variance of $\pi(\tau|\mu, \boldsymbol{x})$ will change simultaneously with latent observations. The influence on μ is be less apparent as only the posterior mean is under the **direct** impact of imputation. Indeed, our simulation results in Chapter 4 show little sensitivity of $\pi(\mu|\tau, \hat{S})$ to an increase in k.

3.4.2 Block Updating

Instead of updating latent values one by one, these will now be considered as a block within two consecutive observations, and updated in a single procedure. The update algorithm is as follows:

1. Initialise all unknown quantities.

2. At iteration h,

   - Draw $\mu^{(h)} \sim \pi(\mu|\tau^{(h-1)}, \hat{S}^{(h-1)})$.
   - Draw $\tau^{(h)} \sim \pi(\tau|\mu^{(h)}, \hat{S}^{(h-1)})$.
   - Define a block as before: $\{s_i, \hat{S}_{t_{i,1}}, \hat{S}_{t_{i,2}}, \ldots, \hat{S}_{t_{i,k-1}}, s_{i+1}\}$. Here, we use the same proposal distribution for $\hat{S}^*_{t_{i,j}}$ as for the observations. Update the latent block as follows:

$$\begin{aligned}
\hat{S}^*_{t_{i,1}} &\sim N\left(s_i + \mu s_i \Delta t, \frac{s_i^2 \Delta t}{\tau}\right), \quad j = 1. \\
\hat{S}^*_{t_{i,j}} &\sim N\left(\hat{S}^*_{t_{i,j-1}} + \mu \hat{S}^*_{t_{i,j-1}} \Delta t, \frac{\hat{(S^*_{t_{i,j-1}})^2} \Delta t}{\tau}\right), \; j = 2, 3, \ldots, k-1.
\end{aligned}$$

3. Define the M-H acceptance ratio as:

$$\begin{aligned} A &= \frac{\prod_{j=1}^{k} \pi\left(\hat{S}^{*}_{t_{i,j}} | \hat{S}^{*}_{t_{i,j-1}}, \mu^{(h)}, \tau^{(h)}\right) \prod_{j=1}^{k-1} q\left(\hat{S}^{(h-1)}_{t_{i,j}} | \hat{S}^{(h-1)}_{t_{i,j-1}}, \mu^{(h)}, \tau^{(h)}\right)}{\prod_{j=1}^{k} \pi\left(\hat{S}^{(h-1)}_{t_{i,j}} | \hat{S}^{(h-1)}_{t_{i,j-1}}, \mu^{(h)}, \tau^{(h)}\right) \prod_{j=1}^{k-1} q\left(\hat{S}^{*}_{t_{i,j}} | \hat{S}^{*}_{t_{i,j-1}}, \mu^{(h)}, \tau^{(h)}\right)} \\ &= \frac{\pi(s_{i+1} | \hat{S}^{*}_{t_{i,j-1}}, \mu^{(h)}, \tau^{(h)})}{\pi(s_{i+1} | \hat{S}^{(h-1)}_{t_{i,k-1}}, \mu^{(h)}, \tau^{(h)}).} \end{aligned}$$

4. Accept the latent block with probability $\alpha(\hat{S}^{*}_{t_{i,j}} | \hat{S}^{(h-1)}_{t_{i,j}}) = \min\{1, A\}$, and retain the previous block otherwise.

5. Let $h := h + 1$ and go to Step 2.

Having outlined the technicalities of sampling procedures, we now apply our MCMC schemes to both real and synthetic data.

Chapter 4

Applications

4.1 Inference Using the Analytic Solution

To begin, we outline results of our MCMC sampling based on the analytic solution to the SPP. In order to verify model validity and tune v, we perform a 'pilot run' using synthetically generated data. The latter is simulated using the stock price formula from the B-S model with $\mu = 0.1, \tau = 100, \Delta t = 0.1$ and maximum observation time $T = 50$. The prior distributions of μ and τ are:

$$\mu \sim N(0, 100), \quad \tau \sim Ga(10, 0.1). \tag{4.1}$$

We implement the scheme for a short run of 2000 iterations in order to assess mixing under different v values. Figure 4.1 includes time series and autocorrelation plots of the τ chain. Observe that if v is too small the chain becomes 'cold'- it accepts the majority of values but explores the sample space slowly and does not converge. Similarly, if v is too large convergence is not achieved as well: the chain becomes 'hot' rejecting too many values and moving in 'leaps'. Observing the autocorrelation pattern of the proposal mechanism, a variance that is too small leads to non-independent draws. For larger values, these values have an oscillating behaviour and may increase for a large lag. Through trial and error it was estimated that the approximate optimal value for v is 0.12, also leading to minimal autocorrelation.

Having tuned the chain, we perform a longer run of 20,000 iterations, the output of which is displayed in Figure 4.2. The chain is mixing well for both variables and appears to be in equilibrium. The sample space is explored efficiently and with little divergence from the posterior mean. Figure 4.3 shows prior (red) and posterior (black) distributions of μ and τ. Starting with μ, we conclude that the MCMC has been informative. The prior mean has shifted from 0 to 0.1 (data mean) and there is a dramatic decrease in variance, so much so it was not reasonable to overlay the plots. Regarding τ, its mean is consistent with the theoretical prior value, and there is also a significant reduction in variance. Additionally, the independence between μ and τ is verified by plotting them against each other- Figure 4.4 indicates about (very) little linear relationship.
To finalize this section, we assess the 95% credible regions of μ and τ- refer to Tables 4.1 and 4.2. The behaviour of these regions will be observed under changes in Δt and

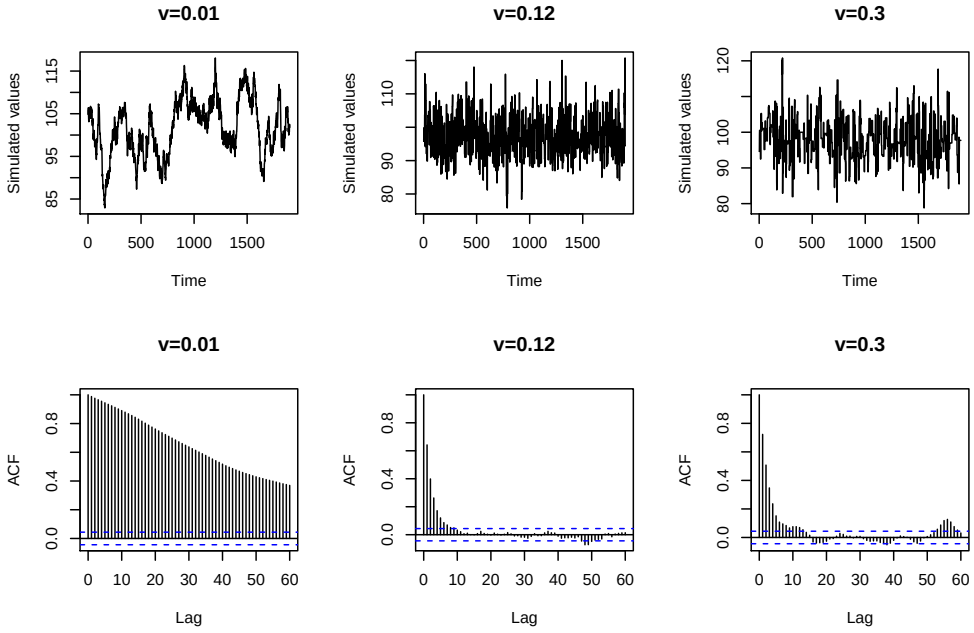

Figure 4.1: MCMC chain mixing and autocorrelation plots under different v.

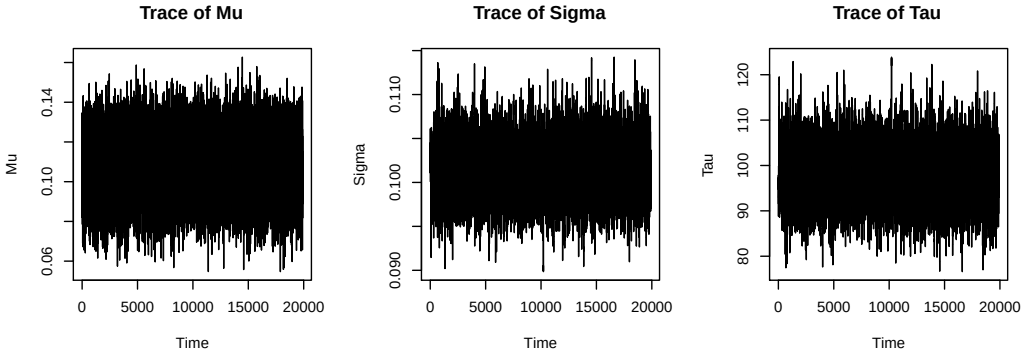

Figure 4.2: Run of the MCMC scheme based on the analytic solution, 20,000 iterations.

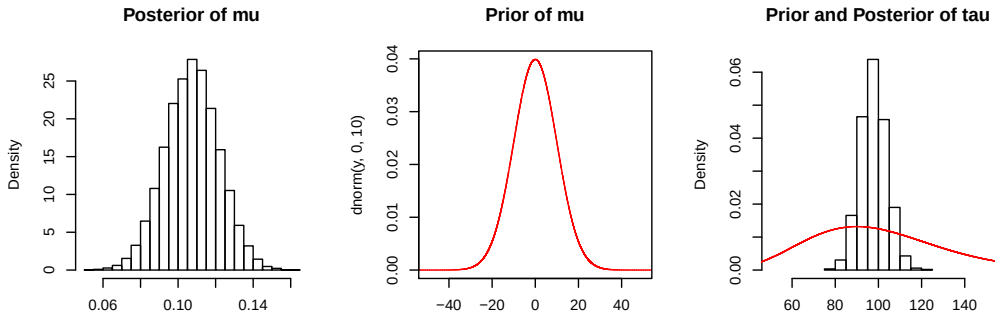

Figure 4.3: Prior and posterior distributions of μ and τ.

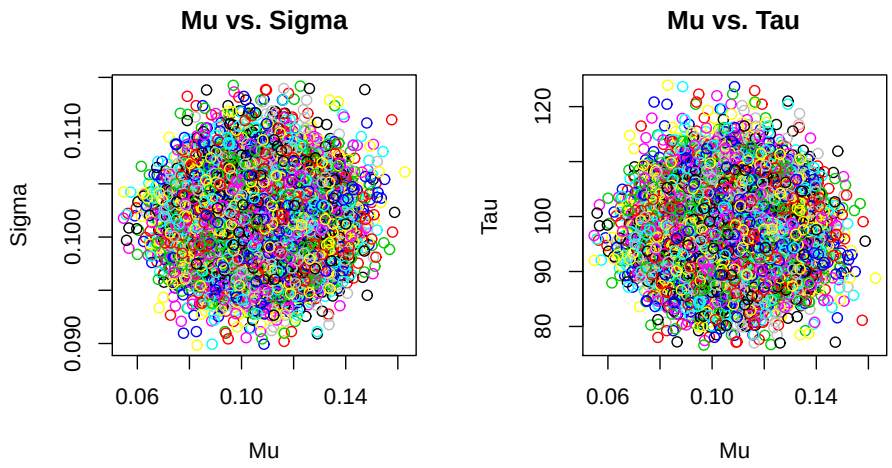

Figure 4.4: Plots of μ against σ and τ.

μ	$T = 1$	$T = 10$	$T = 100$
$\Delta t = 1$	(-0.17, 0.25)	(0.05, 0.17)	(0.09, 0.13)
$\Delta t = 0.1$	(0.05, 0.33)	(0.07, 0.19)	(0.08, 0.12)
$\Delta t = 0.01$	(0.02, 0.39)	(0.02, 0.15)	(0.07, 0.11)

Table 4.1: 95% credible intervals for μ.

τ	$T = 1$	$T = 10$	$T = 100$
$\Delta t = 1$	(48.20, 178.91)	(60.57, 185.63)	(90.44, 153.33)
$\Delta t = 0.1$	(62.35, 185.63)	(90.22, 153.51)	(85.32, 102.02)
$\Delta t = 0.01$	(90.91, 153.10)	(85.41, 101.89)	(94.57, 100.35)

Table 4.2: 95% credible intervals for τ.

T- time step and observation time length respectively. Let $c_{i,j}$ denote a cell in row i, column j. Starting with μ, it should come to a no surprise that as T increases and Δt decreases the regions become narrower. Intriguingly, the regions in $c_{1,3}, c_{2,3}$ and $c_{3,3}$ are similar despite a significant difference in the data amount (by a scale of 10 successively). Recall that μ is the mean interest rate of the SPP. Thus, a better understanding about μ is obtained by looking at a longer time period- focusing on the slope of a processes' trajectory rather than the variability. Conversely, to examine τ credible regions consider the diagonal entries of Table 4.2: $c_{3,1}, c_{2,2}$ and $c_{3,1}$ have the same number of data points (100), as do $c_{3,2}$ and $c_{3,2}$ (1000) and so on. Within these entries the regions are very similar. This implies that we achieve a better prediction about τ by looking at as many fluctuations within a given time period as possible (the slope becomes of secondary importance). This result is apparent throughout our simulation study, as sections to follow illustrate.

4.1.1 A Real Data Example

Having verified the validity of our MCMC scheme we will now use if for real data inference. Consider the daily closing prices of the Standard & Poor's 500 (S&P500) index. Standard & Poor's is an American financial services company producing financial analysis on stocks and bonds. S&P500 is an index of 500 U.S. leading companies (market cap in excess of $4 billion) whose stock is listed on New York Stock Exchange (NYSE) or National Association of Securities Dealers Automated Quotations (NASDAQ). Figures 4.5 (a) and (b) show trajectories of S&P500 daily closing prices the daily returns (ratio between prices at two consecutive days) respectively.

To proceed with Bayesian inference we assign μ and τ the following prior distributions:

$$\mu \sim N(0, 100), \quad \tau \sim Ga(10, 0.1). \tag{4.2}$$

For prior and posterior distributions of μ and τ refer to Figure 4.6, whose behaviour corresponds to the earlier results. The posterior mean values of μ and τ are now consistent with those of the data set, and there is a dramatic reduction in variance. Plots (c) and (d) show the MCMC output for μ and τ. Observe that the chains are mixing well for both parameters (as indicated by Figure 4.5 (c) and (d)), thus the sample space is explored efficiently and convergence is achieved.

4.2 Euler-Maruyama Discretisation

In this section we employ the E-M discretisation within MCMC sampling. The availability of the analytic solution enables to assess (and verify) its accuracy, allowing to employ this approximation to diffusions with intractable SDEs.
Recalling the Equation (2.23), the E-M method approximates a solution to an SDE

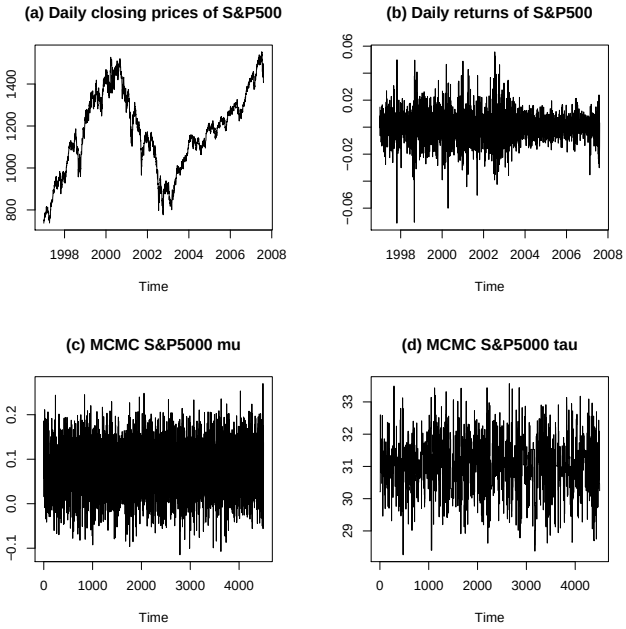

Figure 4.5: S&P500 data, times series plots and MCMC output.

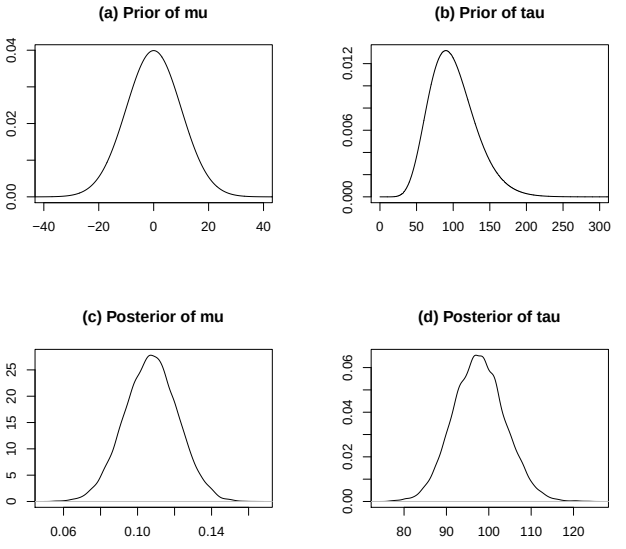

Figure 4.6: S&P500 data, prior and posterior densities for μ and τ.

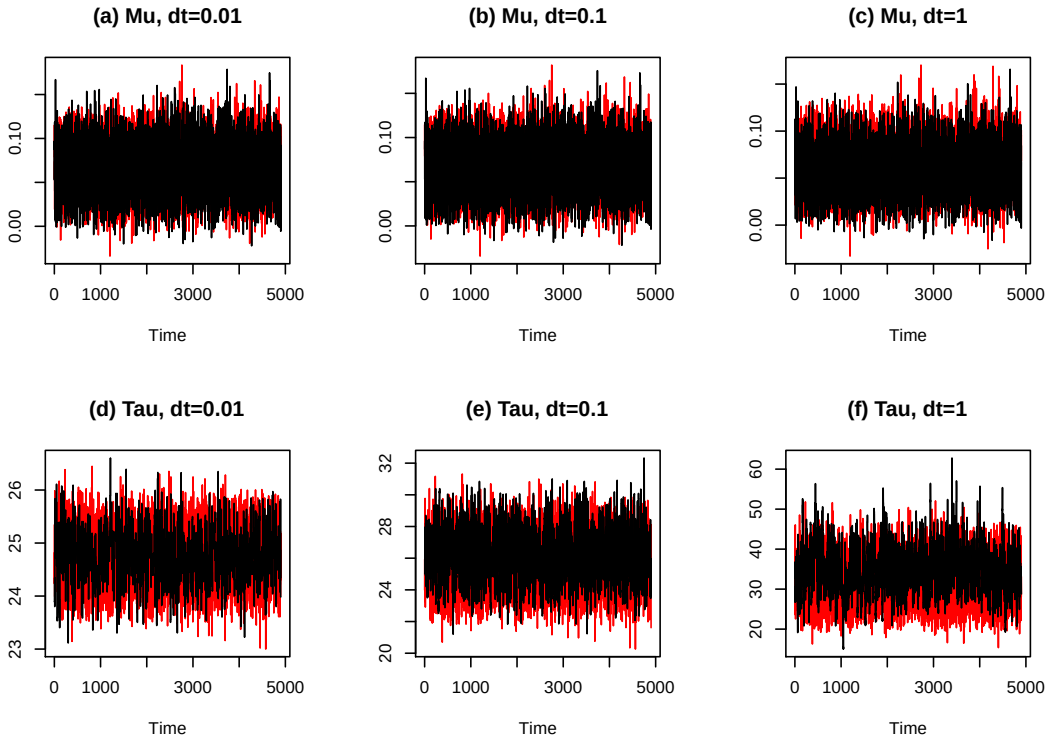

Figure 4.7: Trace plots from the analytic (black) and E-M (red) schemes.

$dX_t = a(t, X_t)dt + b(t, X_t)dW_t$ defined on $[0, T], 0 = t_0 < t_1 < \ldots t_n = T$ as follows:

$$\begin{aligned} y_0 &= x_0, \\ y_{i+1} &= y_i + a(t_i, Y_i)\Delta t_{i+1} + b(t_i, y_i)\Delta W_{i+1}, \\ \Delta W_i &= z_i\sqrt{\Delta t_i}, \quad z_i \sim N(0, 1). \end{aligned}$$

Now assume a Geometric Brownian Motion SDE,

$$dS_t = \mu S_t dt + \sigma S_t dW_t.$$

For notation ease, we shall further refer to the MCMC scheme using the analytic solution to the B-S equation as the **'analytic scheme'**, and that based on discretisation- **'E-M scheme'**. Figure 4.7 overlays realisations from the two samplers, from whom it is evident that the E-M scheme provides an accurate approximation to the process. The E-M trace has a marginally greater variation along the sample space implying an increase in variance (especially within τ), as shown by plots (d)-(f). However, the chains are mixing well and convergence is achieved. Indeed, the E-M chain explores the sample space more efficiently due to a higher acceptance rate than the analytic scheme.

Figure 4.8 shows trace plots from the E-M model for different Δt plotted within the same y-limits. For μ (plots (a)-(c)), the obtained results are consistent with those in Section 4.1. Realisations exhibit similar variation along the plot showing little sensitivity

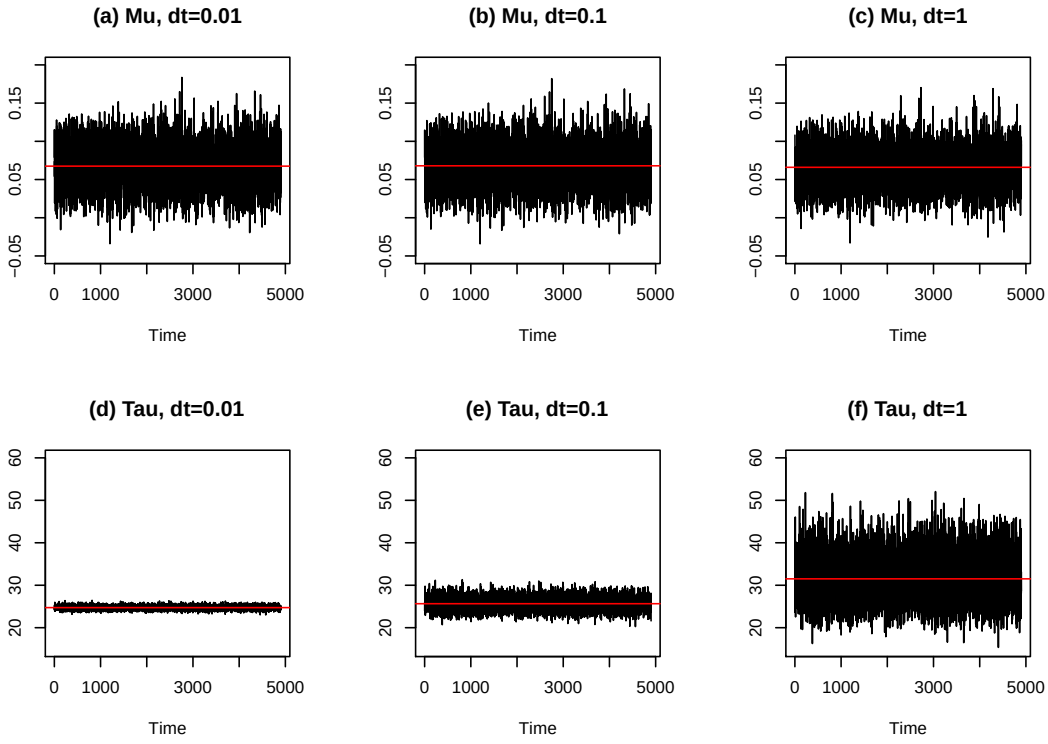

Figure 4.8: Traceplots of μ and τ posterior realisations from the E-M discretisation. Red line indicates the mean of times series.

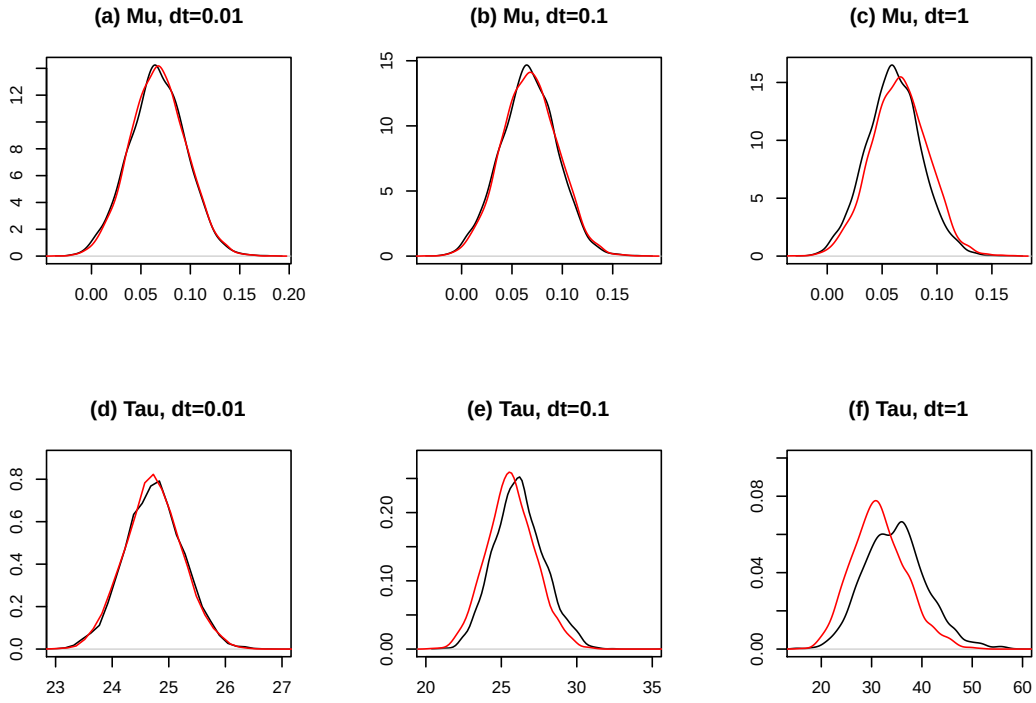

Figure 4.9: 'Analytic' (black) and E-M (red) posterior densities.

to changes in Δt. There is also little change in the posterior mean value, as indicated by red lines. Therefore, the posterior distribution of μ predicts the mean rate of return with similar certainty regardless the amount of provided data (for a fixed time horizon). The posterior of τ is more sensitive to a decrease in observation frequency. We follow a dramatic growth in variance, and increasing Δt beyond 0.1 has a significant negative impact on the posterior mean estimation which diverges from the 'actual' value of 0.25.

Figure 4.9 shows overlaid densities for the posterior distributions of μ and τ for the two samplers. As one would expect, the discretisation is closest to the 'actual' curves when the time step is minimal. However, plots (a)-(c) also suggest that in the case of μ the approximation is efficient for Δt as large as 1, beneficial when the deployed algorithm is highly time-consuming or observations- scarce. Regarding τ, the approximation becomes increasingly inefficient after Δt grows beyond 0.1. This, again, corresponds to the earlier results. To conclude, refer to Figure 4.10 showing overlaid densities of μ and τ on single plots. Whilst for μ no surprising results are found, the impact of low observation frequency is now evident for τ, as we observe significant deterioration in estimation accuracy. In particular, there is a positive drift of the posterior mean and an increase in variability.

Having verified the validity of the E-M model, we now apply the discretisation to the S&P500 data set. Figure 4.11 illustrates chain mixing and densities for the two models. Plots (a) and (b) compare the analytic (black) and E-M (red) samplers, performed

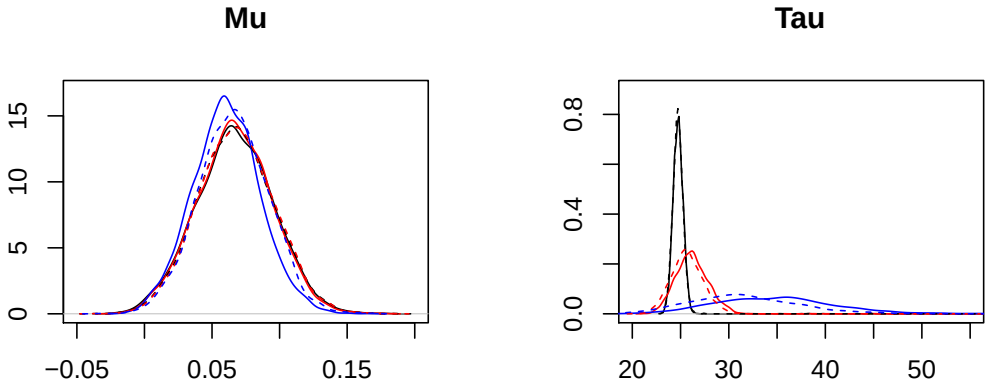

Figure 4.10: Overlaid densities of posterior distributions. Solid line- analytic sampler, dashed line- E-M sampler. Black: $\Delta t = 0.01$, Red: $\Delta t = 0.1$, Blue: $\Delta t = 1$.

on daily-observed data (for 10,000 iterations). Assuming 252 working days in a year, $\Delta t = 1/252 \approx 0.004$. Plot (a) indicates that the approximation is algorithm is efficient for the μ posterior, as chain mixing is superior to that of the analytic sampler. However, in terms of τ, we observe poor mixing. In order to optimise this, the data set is thinned by 3 and fed into the samplers, with a subsequent improvement in chain mixing for both parameters (refer to Figure 4.12). This is particularly evident in the case of τ where the sample space is now explored fairly efficiently. Note, that the mean values on both data sets are consistent, thus thinning has not biased our conclusions.

4.3 EM Imputation Sampler

In this section, we examine the effect of data imputation within Euler-Maruyama discretisation on parameter estimation. Although previous analysis has shown that using observations at integer times provides consistent estimation of the μ posterior, this approximation is potentially unsatisfactory in practice, especially for τ.

We employ the algorithm structure defined in Section 3.4. As before, firstly we verify model validity by performing a pilot run using synthetic 'observations' $\boldsymbol{s}$. We generate $\boldsymbol{s}$ from the B-S model with $\mu = 0.1$, $\sigma = 0.2$ thus $\tau = 25$, and define parameter priors:

$$\begin{gathered} \mu \sim N(a, 1/b), \quad \tau \sim Ga(g, h), \\ k = 2, \quad a = 0, \quad b = 0.01, \quad g = 0.1, \quad h = 0.1. \end{gathered} \tag{4.3}$$

Figure 4.13 shows posterior densities for the two parameters from the pilot run. Recall that $k = 2$ is equivalent to dividing the interval between two consecutive observations in two by imputing 1 latent variable, therefore $\Delta t = 0.5$. We note that $\pi(\mu|\tau, \boldsymbol{s})$ and $\pi(\tau|\mu, \boldsymbol{s})$ means (indicated in red) are fairly consistent with their prior values (0.1 and 25

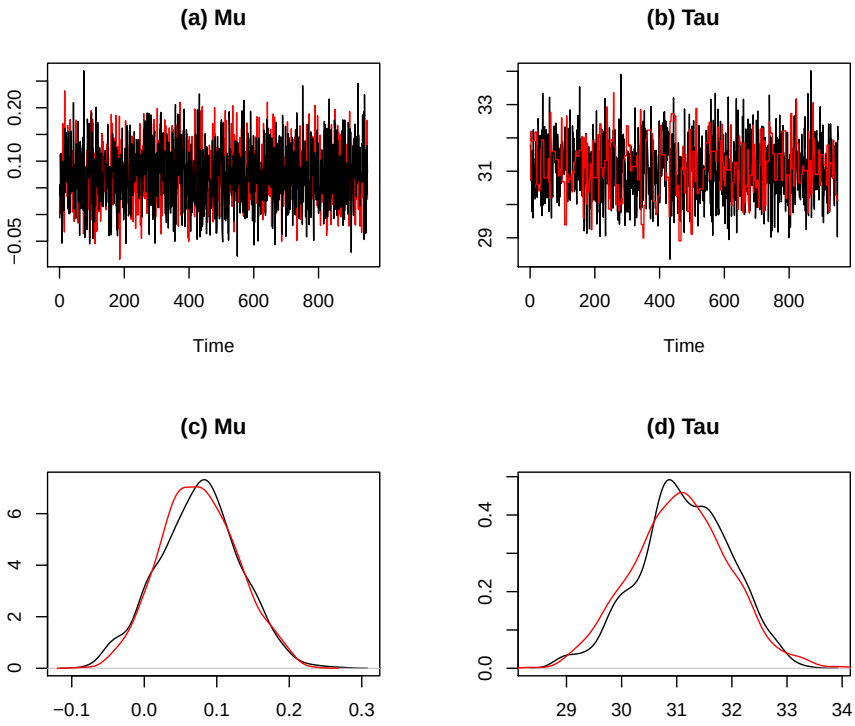

Figure 4.11: Trace plots of the analytic (black) and E-M (red) schemes using the S&P500 data. Plots (a) and (b) illustrate the first 5000 simulated values to highlight the mixing pattern.

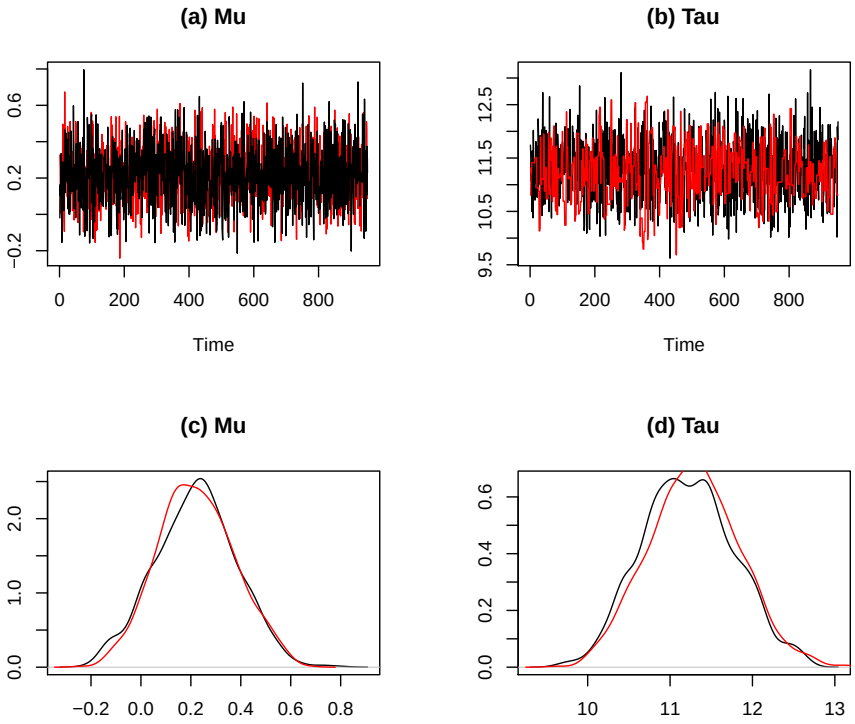

Figure 4.12: Trace plots of the analytic (black) and E-M (red) schemes using **thinned** S&P500 data, first 5000 observations.

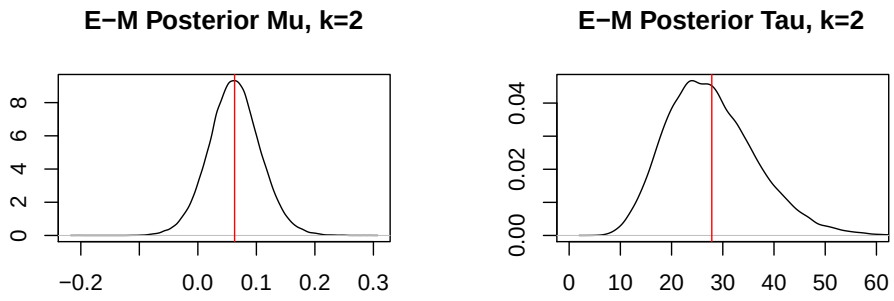

Figure 4.13: Posterior densities from the imputation E-M sampler, $k = 2$.

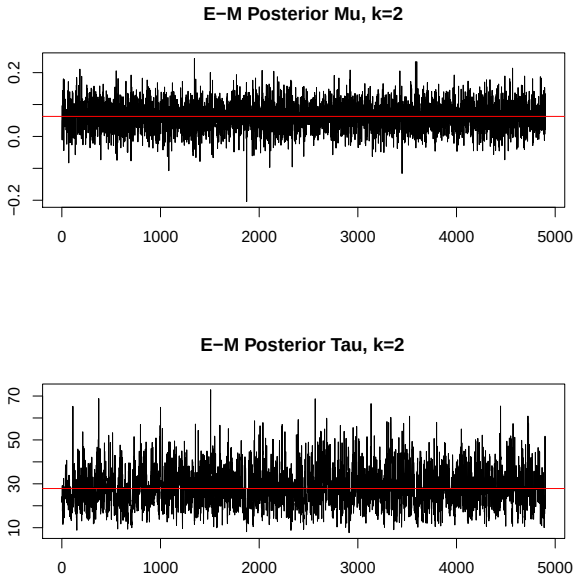

Figure 4.14: Chain mixing for imputation E-M sampler, $k = 2$, red indicates posterior mean values.

respectively). Referring to Figure 4.14, it is evident that the chain mixing is satisfactory for both variables as well. For μ, the sample space is explored efficiently with little variation along the plot, and the posterior mean is consistent with the true value that produced the data. For τ, the parameter space is also explored well but there is larger posterior uncertainty about the mean. This should not come to a surprise as recalling results from Section 4.1, the credible interval for τ is highly sensitive to the amount of provided data. Thus, a small k (i.e. 2) produces a discretisation that is rather crude.

Having verified that the imputation scheme is valid, we implement this algorithm for a longer run of 10,000 iterations. Figures 4.15 and 4.16 include densities for μ and τ (respectively) for $k = 2, 4, 5$ and 8 (cyan lines on trace plots indicate posterior means, red lines indicate prior means and blue- the 95% density regions). Starting with μ, the imputed E-M sampler is mixing well and has negligible autocorrelation for all k. The posterior means are fairly consistent with the prior value of 0.1 and are well within the 95% density region. Intriguingly, increasing frequency of data imputation does not significantly improve the posterior mean accuracy, whilst being computationally heavy. We note a decrease in posterior variance, although (again) this difference in negligible when $k = 5$ and 8 are compared. Looking at the τ plots in Figure 4.16, the sample space exploration deteriorates as k reaches 5, and the chain becomes 'cold' as $k = 8$. This occurs in parallel to an increase in autocorrelation and degradation in posterior density shape. In contrast, the posterior means are consistent with true values that produced the data for all k. Thus, for computational ease and optimal mixing use $k = 4$ or 5.

Our next step is to compare the 'latent data' (further called **'imputation E-M'**) sampler using $k = 2, 4, 5, 8$ with the former E-M algorithm with $\Delta t = 1$ (thus $k = 1$). This should indicate whether the introduction of latent variables has a positive impact on parameter estimation accuracy. For results, refer to Figure 4.17 showing density plots for both schemes, with overlaid prior mean values of μ and τ in red, blue lines indicate the 95% density regions. We again note that data imputation causes little improvement of μ density, although does offer some reduction in variability. Referring to τ, the consistency between the prior and posterior means is now optimised, thus employing latent variables improves the reliability of simulated results. We also obtain a significant reduction in variance. Figure 4.18 illustrates chain mixing for the two algorithms, with an indication of the posterior mean value in cyan. Again, the posterior mean of μ appears to be unaffected by an increase in imputation volume, whereas increasing the value of k leads to a noise reduction in τ posterior. On the other hand, smaller ks lead to higher acceptance rates within the algorithm, thus more efficient chain mixing. It appears optimal to choose $k = 4$ or 5, which satisfies the requirement of posterior variance reduction, chain mixing and computational efficiency.

We are now in a position to compare the efficiency of our 3 samplers. Figure 4.19 illustrates the densities for each of the schemes for different k values, from which it is evident that imputation improves parameter estimation. For both μ and τ, the analytic and imputation E-M posterior distributions (indicated in black and cyan respectively) match very closely for $k = 2$ and are almost identical for $k \geq 4$. The former E-M algorithm seems to be inferior in this case. For τ, there is a greater discrepancy between the analytic and imputed E-M samplers when $k = 2$, however taking $k = 4$ is sufficient to allow consistent estimation.

It is useful to analyse the effect of increasing k on chain mixing. From the above analysis it is evident that under conditions of single latent value updating, the mixing worsens as k grows. In particular, the correlation between algorithm draws increases simultaneously with k. This is especially the case for τ posterior due to the strong dependence between S_l and τ. Indeed, according to Eraker (2001) convergence of the Gibbs sampler worsens as Δt gets small, and does not converge at all as $\Delta t \to 0$ [6] (p.182). Nevertheless, it is also evident that the introduction of latent variables improves the consistency of the E-M algorithm with the analytic distribution (Figure 4.19). A possible solution to the convergence problem is described by Elerian et al. in the paper 'Likelihood Inference for Discretely Observed Nonlinear Diffusions' [5](p.971). It is proposed that the latent variables are updated in **blocks** between two consecutive observations, and the next section outlines the method and results of such a procedure.

4.3.1 E-M Block Imputation Sampler

In order to overcome the autocorrelation and computational inefficiency issues we now consider a **block E-M** scheme. Figure 4.20 includes chain mixing, autocorrelation and density plots for the block E-M scheme with $k = 5$, implemented for 10,000 iterations. Examining the posterior of μ we observe a good sample space exploration by the chain,

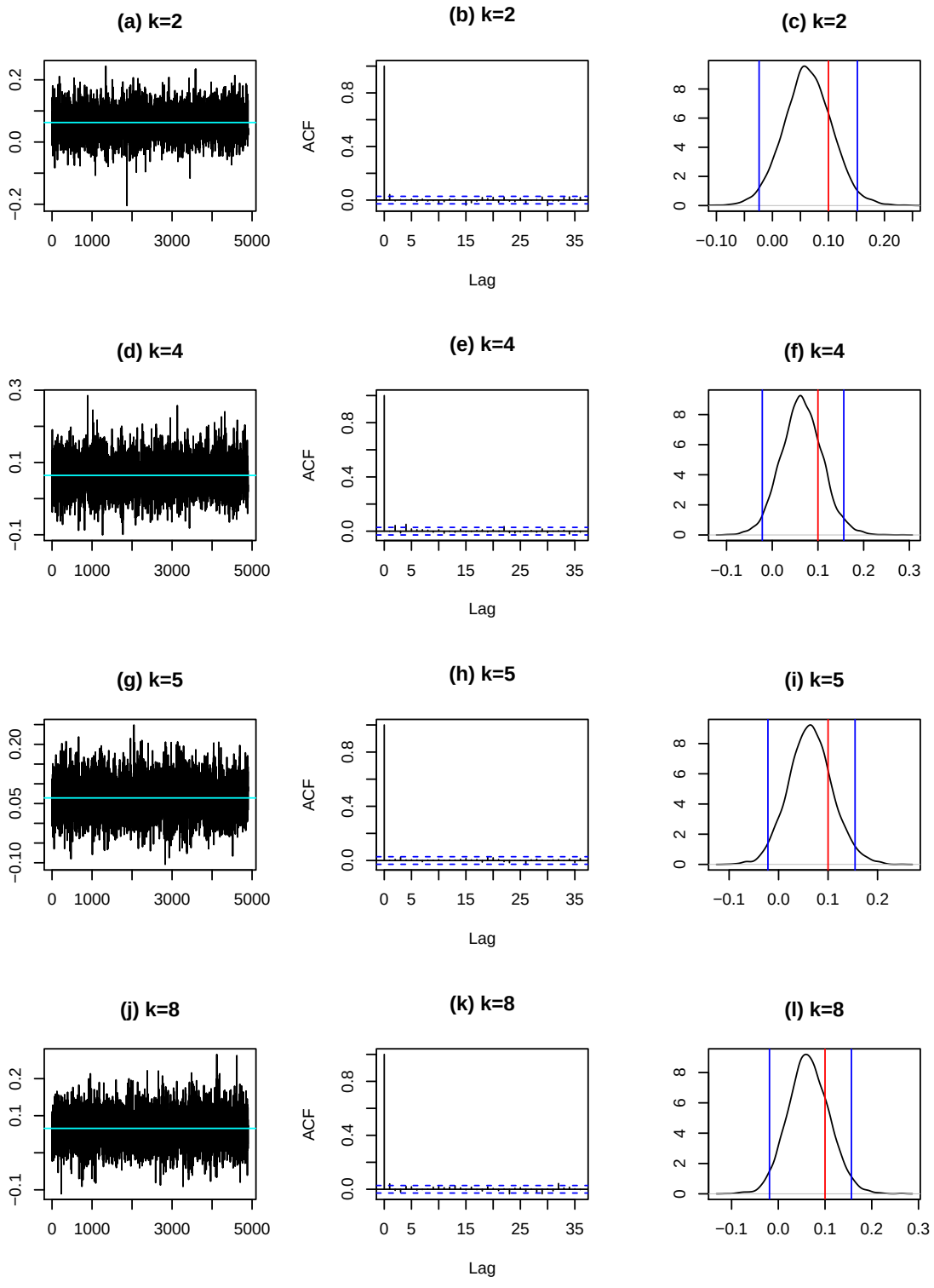

Figure 4.15: Posterior densities of μ using imputation E-M sampler, cyan- posterior mean value, red- prior mean values, blue- 95% density regions.

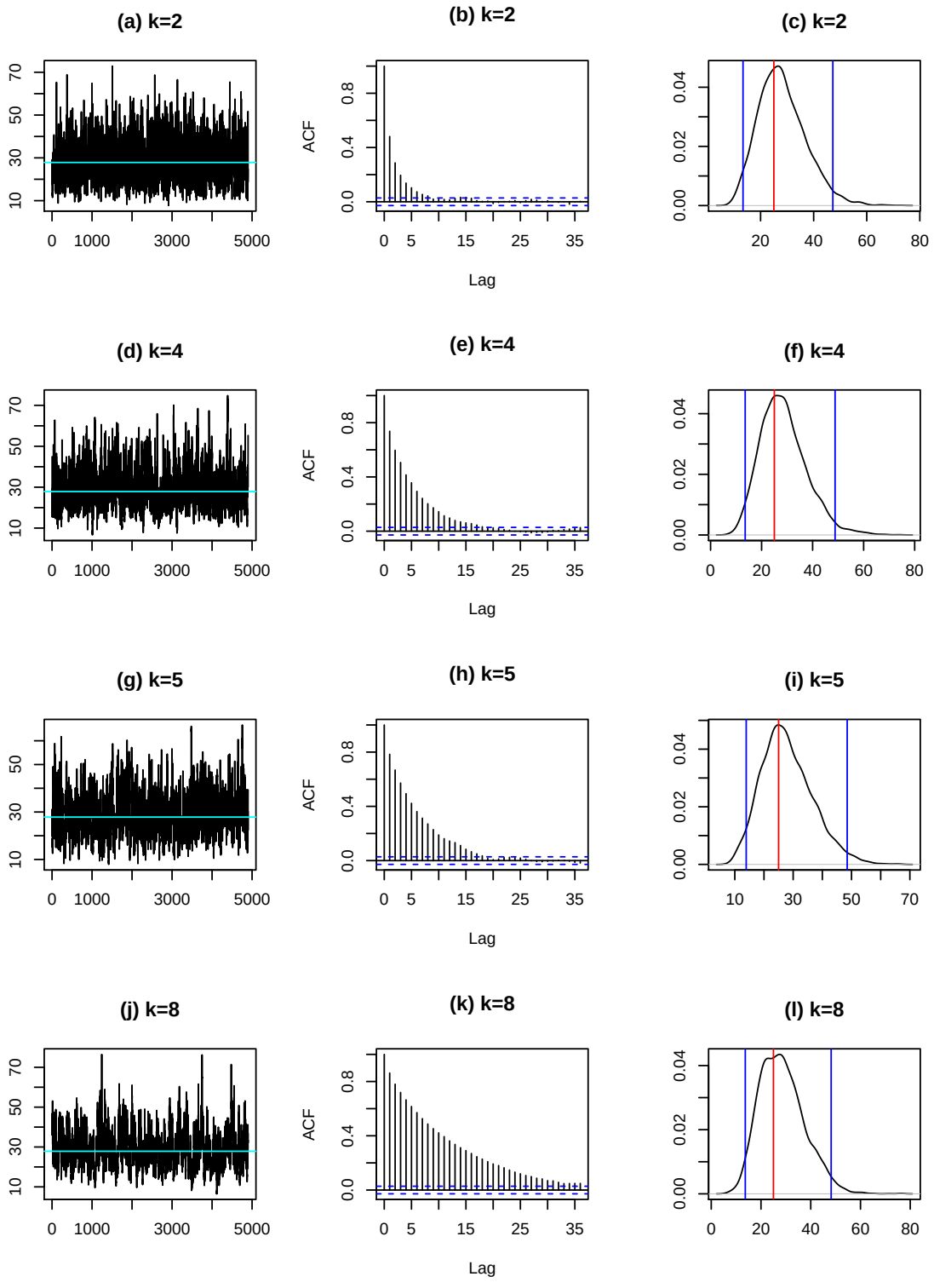

Figure 4.16: Posterior densities of τ using imputation E-M sampler, cyan- posterior mean value, red- prior mean values, blue- 95% density regions.

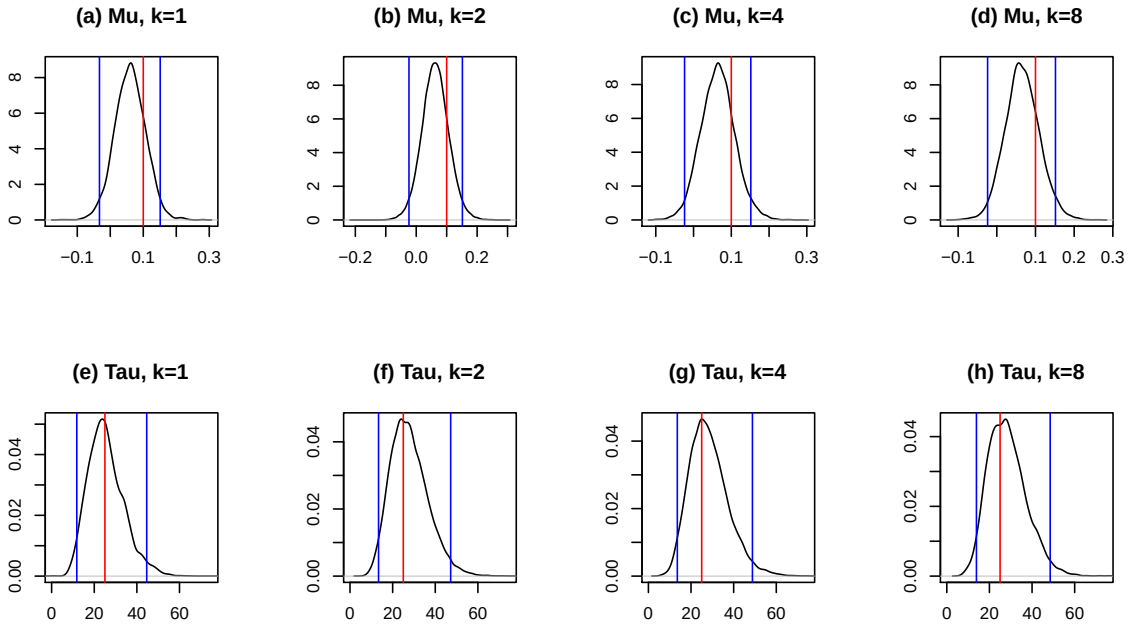

Figure 4.17: Density plots for original and imputation E-M schemes. Red indicates prior mean values, blue- 95% density regions.

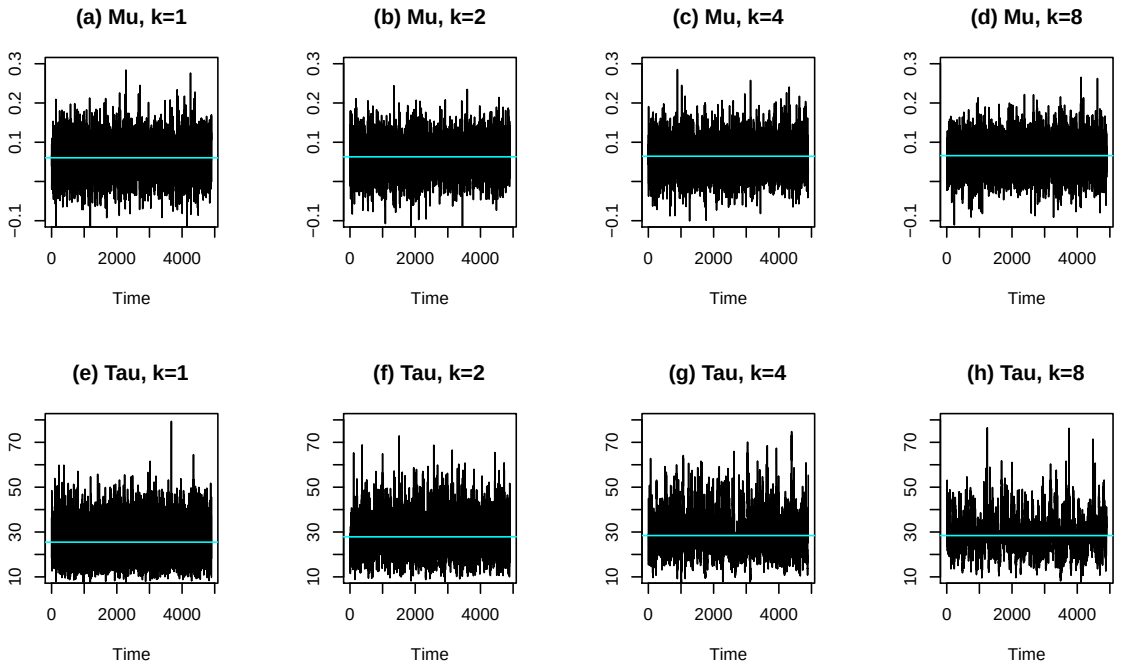

Figure 4.18: Chain mixing for two E-M schemes. Plots (a) and (e) are equivalent to the original (no imputation) E-M scheme. Cyan indicates true parameter mean.

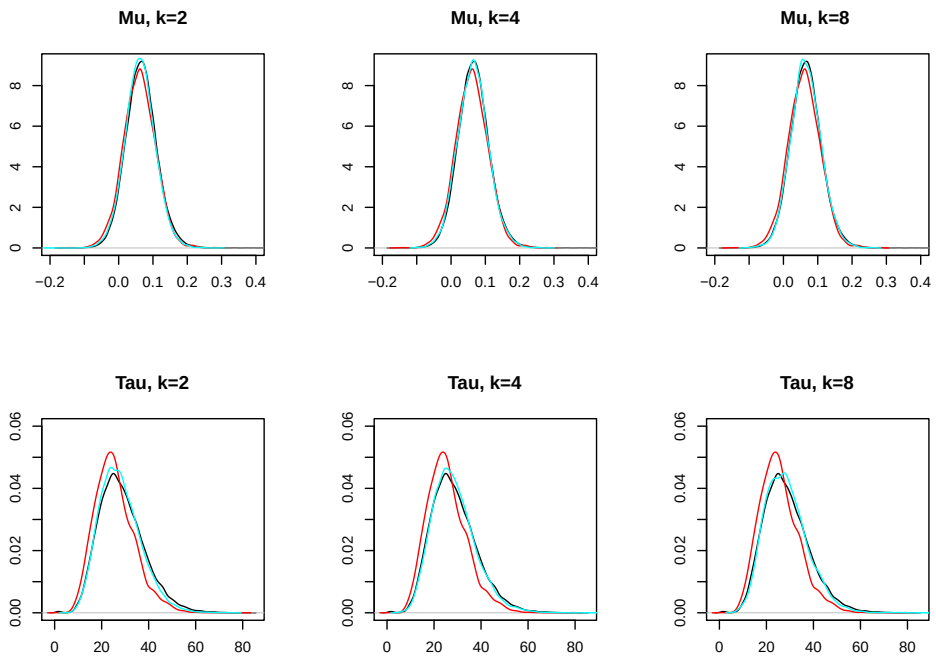

Figure 4.19: Sampler posterior densities for μ and τ, black indicates analytic scheme, red- original E-M, cyan- imputation E-M.

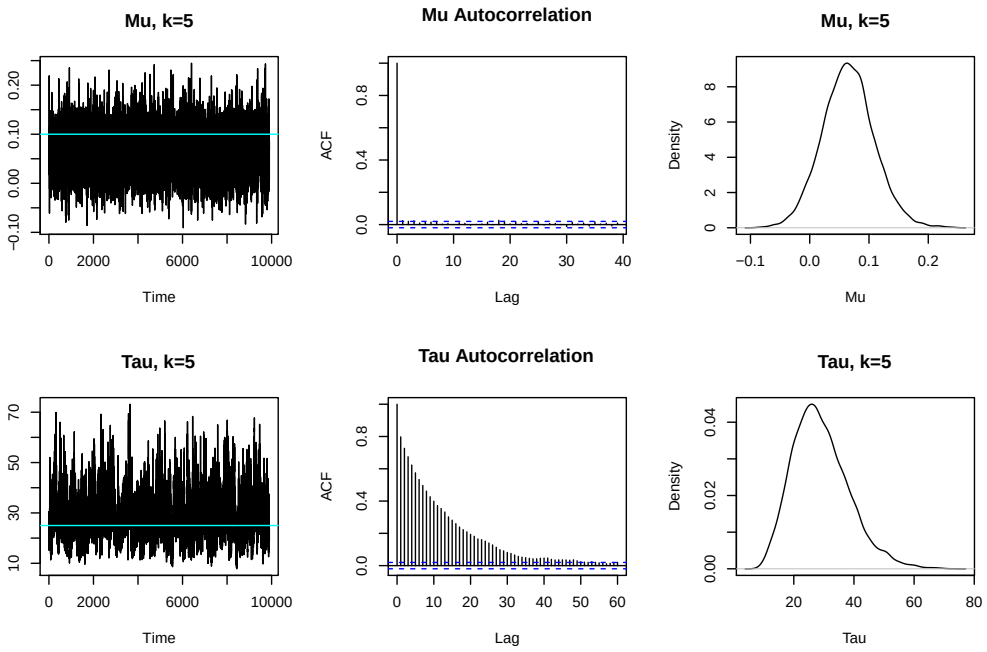

Figure 4.20: Block E-M scheme with $k = 5$, cyan indicates prior mean values.

and thus efficient mixing. The posterior density plot is centered close to the true mean (that produced the data) and the autocorrelation values are insignificant. Referring to τ, the chain mixing is adequate and the density is centered around the prior mean as well. However, autocorrelation values are high up to lag 40 indicating that draws from the posterior distribution are not independent.

Figure 4.21 includes analogous plots for $k = 8$ and 15, which indicate that increasing the imputation volume improves posterior mean estimation and leads to a decrease in variance for both parameters. However, as k increases, the chain mixing and autocorrelation pattern for τ deteriorate, therefore again presenting a dilemma of bias versus variance.

Figure 4.22 allows to compare the single-site and block E-M samplers for $k = 5$, from which we conclude that the two update mechanisms lead to (almost) identical posterior density estimation. Moreover, the single-site update sampler is superior for τ due to lower autocorrelation, whereas the block scheme is less time-consuming.

Finally, referring to Figure 4.23 we compare the densities generated by the block, single-site E-M and analytic samplers. We observe very little difference between the three schemes, especially in terms of μ. The accuracy obtained by $k = 5$ and $k = 8$ is equivalent, therefore μ does not require a large volume of imputation to be estimated efficiently. For τ, the block update is marginally more accurate when $k = 8$, and is also less time-consuming to implement than the single-site update. Thus, if the error cost of τ outweighs that of time one should choose a larger imputation volume due to an improved accuracy, and vice versa. Figure 4.24 illustrates the downfall of our current Block update mechanism. The Markov chain associated with the latent data vector $\{S_{t_{i,1}}, S_{t_{i,2}}, \ldots, S_{t_{i,k-1}}\}$ is only dependent on (or initialised at) its preceding observed value s_i, not the future s_{i+1}. Therefore, the discrepancy between the imputed $S_{t_{i,k-1}}$ and observed s_{i+1} will cause sampler inefficiency. This problem is targeted by introducing a **Brownian bridge**- Brownian motion conditioned to start at s_i and finish at s_{i+1} [4](p.305). This requires a suitable proposal density satisfying this dependence- such is outlined by Elerian et al. [5](p.969-970). The latent data values are drawn from a Gaussian density $S_{t_{i,j}}, S_{i+1}|S_{t_{i,j-1}}$ which can be constructed by approximating the joint density of $S_{t_{i.j}}$ and S_{i+1} as

$$\begin{pmatrix} S_{t_{i,j}} \\ S_{i+1} \end{pmatrix} \sim N_2 \left(\begin{pmatrix} S_{t_{i,j-1}} + \mu_i \Delta t \\ S_{t_{i,j-1}} + \mu_i (i - t_{i,j-1}) \Delta t \end{pmatrix}, \begin{pmatrix} \frac{\Delta t}{\tau} & \frac{\Delta t}{\tau} \\ \frac{\Delta t}{\tau} & \frac{(i - t_{i,j-1})\Delta t}{\tau} \end{pmatrix} \right). \tag{4.4}$$

We then condition on $S_{i+1} = s_{i+1}$ to give $S_{t_{i,j}}|S_{t_{i,j-1}}, s_{i+1}$ which can be sampled recursively. In fact, this a **modified diffusion bridge** – one of four Brownian bridge types proposed by Durham and Gallant [4](p.304). It has been shown that such type of sampling leads to a higher acceptance probability of imputed data (and subsequently, the process parameters), lowering computational cost and posterior autocorrelation. Pursuing the latter is intuitive in order to improve our current mechanisms, however (regretfully) is external to the current investigation.

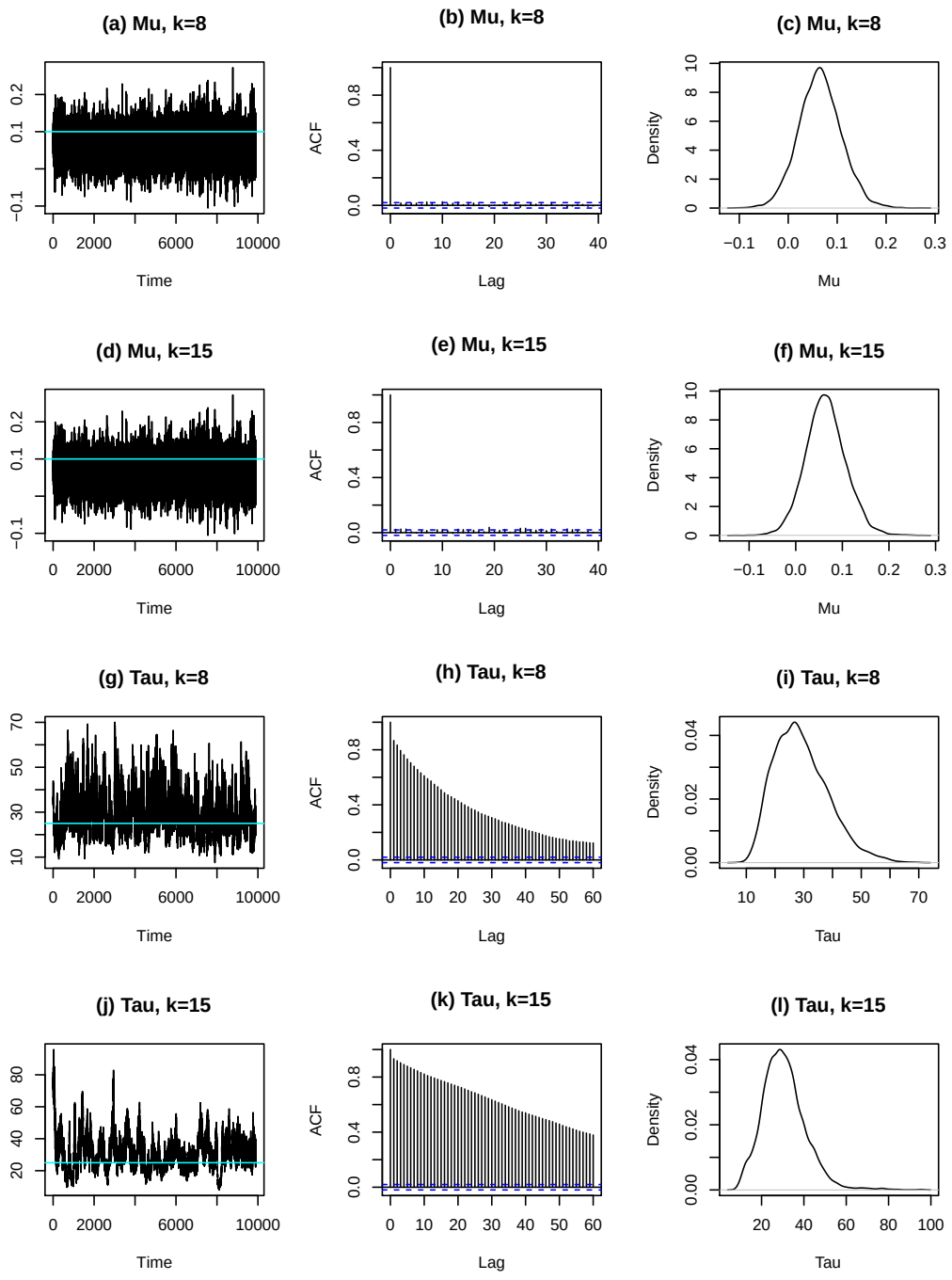

Figure 4.21: Block E-M scheme, k=8 and 15, cyan indicates prior mean values.

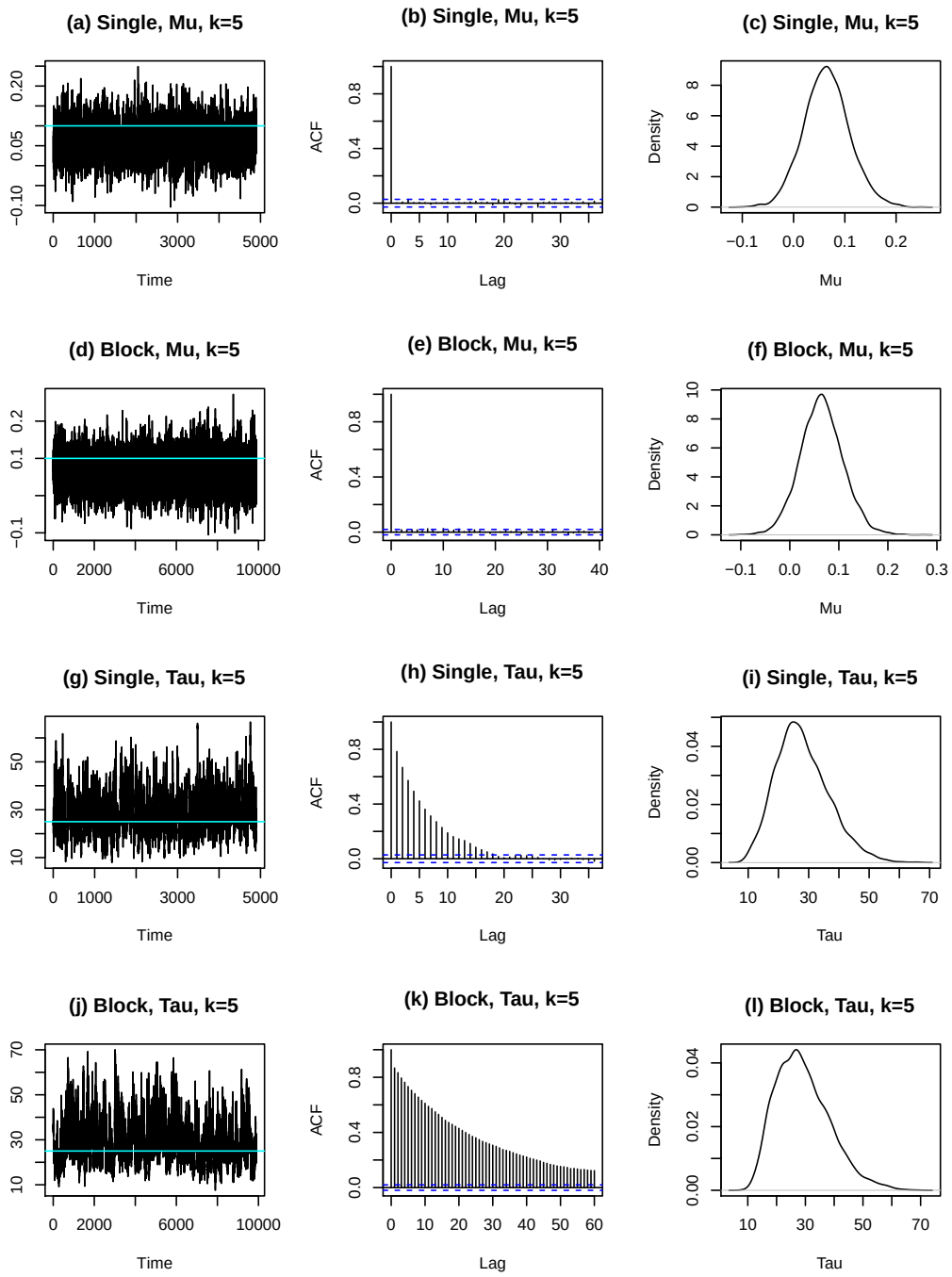

Figure 4.22: Comparison of Block and single-site (imputation) E-M schemes, k=5 and 15, cyan indicates prior mean values.

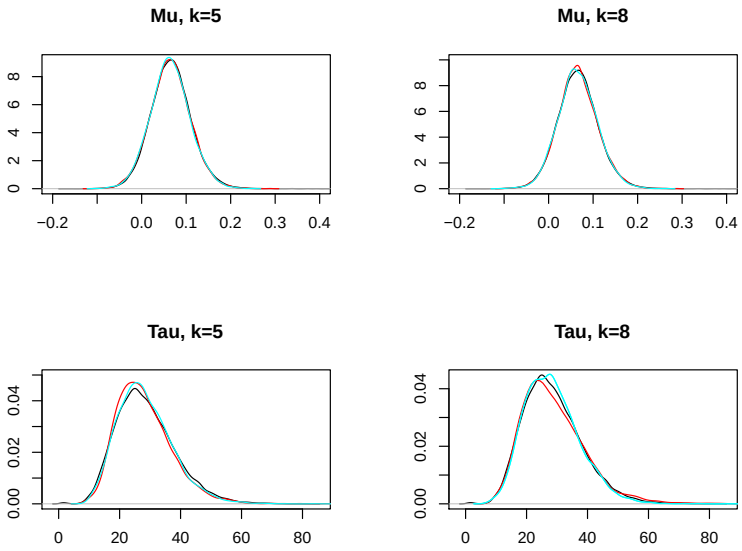

Figure 4.23: Densities for three schemes. Black indicates analytic, cyan- single-site E-M and red- block E-M schemes.

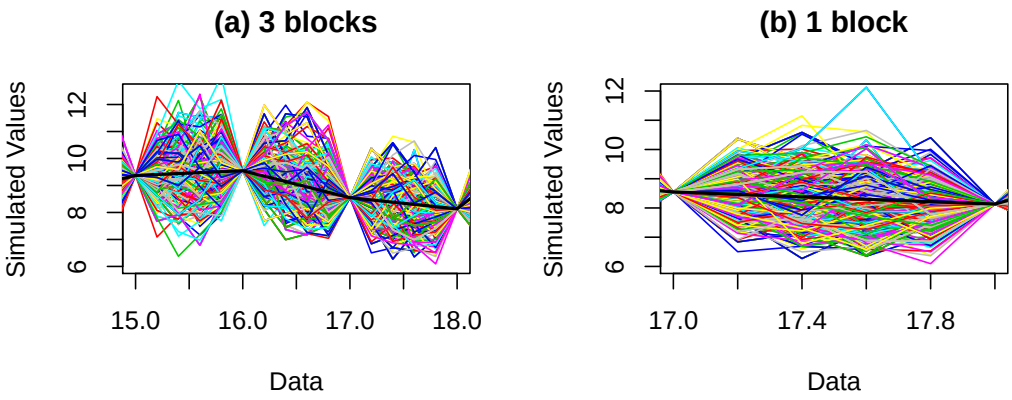

Figure 4.24: Block E-M scheme simulated (accepted) paths, black line is linear interpolation between consecutive observations.

Chapter 5

Summary

In practice, transition densities for many diffusion processes are unavailable in a closed form or are too burdensome to derive. The availability of the analytic solution to the SPP problem has allowed us to construct an 'exact' MCMC scheme and use it to evaluate the effectiveness of Euler-Maruyama discretisation in the context of simulation. The latter has shown to be sufficiently accurate to be used in practice, given that data are observed with medium-high frequency. When this is not satisfied we proposed a method of data imputation within an E-M sampler, using single-site and block updating. Again, these have shown to be sufficiently accurate leading to a marked improvement in τ estimation.

Two problems remain: the computational burden associated with data imputation and high posterior correlation of τ sampler paths. One would typically be faced with a choice between strong bias and autocorrelation, respectively caused by sampling error and non-independent draws of the latent variables. This indeed is unfortunate, as block updating should lead to optimised estimation and computation process. This issue can be overcome employing a Gaussian density proposal conditioning on the observation value preceding the latent block. This creates a 'Brownian bridge' of the imputed data ensuring that the proposed variables are consistent with 'observed' values [4](p.204). Durham and Gallant also introduce methods for 'acceleration' of Monte Carlo integration. Such include bias and variance reduction techniques, substantially optimising the computational efficiency [4](p.302).

Another possible issue of the current block update model is that the size of the latent value vector $(k-1)$ is fixed. Trivially, the proposed blocks always connect in the same place, 'which can foster dependencies in the MCMC sweeps' [5](p.971). Overcoming the consequent posterior τ autocorrelation is to transform k from scalar to a variable quantity that follows a Poisson distribution with rate λ. The above methods would be the next and exciting steps for our research, however are beyond the scope of this report.

Bibliography

[1] Allen, L.J.S. (2003), *An introduction to stochastic processes with applications to biology*, Upper Saddle River, N.J.: Pearson/Prentice Hall

[2] Black, F. & Scholes, M. (1973), *The Pricing of Options and Corporate Liabilities*, The Journal of Political Economy, Vol. 81, No. 3 (May - Jun), p. 637-654

[3] Brooks, S. (2010), *Handbook of Markov Chain Monte Carlo*, London: Chapman & Hall

[4] Durham, G.B. & Gallant, A.R., (2001), *Numerical Techniques for Maximum Likelihood Estimation of Continuous-Time Diffusion Processes*, Journal of Business & Economic Statistics, Vol. 20, No. 3, p. 297-316

[5] Elerian, O., Chib, S. & Shephard, N. (2001), *Likelihood Inference for Discretely Observed Nonlinear Diffusions*, Econometrica, Vol. 69, No. 4, p. 959-993

[6] Eraker, B. (2001), *MCMC Analysis of Diffusion Models with Application to Finance*, American Statistical Association, Journal of Business & Economic Statistics, Vol. 19, No. 2, p. 177-191

[7] Geman, S. & Geman, D. (1984), *Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 6, No. 6, p. 721741

[8] Rasmussen, C.E. & Williams, C.K.I. (2006), *Gaussian Processes for Machine Learning*, MIT Cambridge

[9] Ross, S.M. (2010), *Introduction to Probability Models*, 10th edition, Boston: Academic Press

[10] Sauer, T. (2012), *Numerical Solution of stochastic differential equations in finance*, Berlin-Heidelberg

[11] Steele, J. (2001), *Stochastic Calculus and Financial Applications*, Springer-Verlag

[12] Sutton, D.P. (1989), *Understanding the stock market*, Chicago, Ill.: Probus