Università
Ca'Foscari
Venezia

Master's Degree
in Models and Methods of
Quantitative Economics

Final Thesis

Bayesian Inference for Diffusion Processes

Supervisor
Prof. Roberto Casarin

Graduand
Ngoc Truong Nguyen
MN.: 890521

Academic Year
2022/2023

Abstract

This thesis explores the use of Bayesian inference and Markow Chain Monte Carlo (MCMC) methods for diffusion processes in asset pricing, including Geometric Brownian Motion (GBM), Multivariate Merton's model, and term structure models.

In asset pricing, the GBM serves as a fundamental framework, and antithetic techniques are employed to compare the efficiency of standard MCMC and the antithetic MCMC. Moreover, we expand upon the GBM by incorporating jump processes through the Multivariate Merton's model. Bayesian inference is applied to estimate structure and state variables such as jump size and time intensity.

Moving to the term structure modelling, the Vasicek model is employed to capture the dynamics of interest rates. Here, we propose the band matrix techniques for latent variables and the Adaptive Metropolis-Hastings algorithm for variables. We also test the validity and efficiency of the MCMC approximation using simulated data. Furthermore, application to real data from economics and finance is provided.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Diffusion processes, distinguished by stochastic movements and the progressive dispersion of quantities across temporal and spatial dimensions, have demonstrated significant utility across diverse fields. Some significant examples where this occurs include bio-informatics (Arkin et al., 1998; McAdams and Arkin, 1999) , traffic flows (Zhang et al., 2019), image processing and computer vision (Weickert et al., 1998), and economic & finance (Merton, 1990; Dixit, 1993).

Consider an Itô stochastic process that satisfies the form's stochastic differential equation (SDE):

$$dk\left(t\right) = m\left\{k\left(t\right), t, \theta\right\} dt + n\left\{k\left(t\right), t, \theta\right\} dW\left(t\right) \tag{1.0.1}$$

where:

- $m\left\{k\left(t\right), t, \theta\right\}$ and $n\left\{k\left(t\right), t, \theta\right\}$ are the drift and diffusion functions, respectively, which depends on $k\left(t\right), t$, and unknown vector of parameters θ
- $dW\left(t\right)$ is the increment of a Wiener process, with $W\left(0\right) = 0$.

Assume that all conditions necessary for solving the SDE for a diffusion $k(t)$ have been met (see Oksendal (2013), p.65). Assume we have $k_t = k\left(\tau_t\right)$ measurements at times $\{\tau_1, \tau_2, ..., \tau_T\}$ and $\Delta_t = \tau_{t+1} - \tau_t \geq 0$ for $t \leq T$. The objective is to estimate θ given the observations $K = (k_1, k_2, ..., k_T)$. The estimation of θ in the likelihood context is derived from the log-likelihood

function $\log L(k_2, k_3, ..., k_T|\theta, y_1) = \sum_{t=1}^{T-1} \log h(k_{t+1}|k_t, \theta)$, where k_t is obtained from the SDE:

$$k(t) = k(0) + \int_0^t m\{k(s), s, \theta\} ds + \int_0^t n\{k(s), s, \theta\} dW(s) \quad (1.0.2)$$

and the Markov transition kernels $h(k_{t+1}|k_t, \theta)$ are used in this process.

Closed-form solutions for $t \leq T$ can be obtained for (1.0.2). If a strong solution of the underlying process is available, then the closed form of $h(k_{t+1}|k_t, \theta)$ allows for straightforward likelihood inference. However, analytic solutions for SDEs are infrequently obtainable, and the parameters governing their dynamics are often difficult to estimate from discrete-time data. Estimation poses a challenge due to the discrepancy between the continuous time formulation of the model and the discrete frequency nature of the available sample data. Moreover, naive discretizations of diffusion processes may result in discretization bias in the estimates.

The interest in estimating diffusion parameters from discretely sampled measurements has increased and can be classified into three main areas. First, alternative estimators to maximum likelihood estimation (MLE) have been proposed in the literature. These include the use of estimating functions (Bibby, 1996), the analytic approximation to the likelihood function (Ait-Sahalia, 1998), and generalized method of moments (GMM)-based estimators, which are discussed in Duffie and Glynn (2004). Second, Pedersen (1995) have proposed Simulated Maximum Likelihood Estimation (SMLE) approaches, Durham and Gallant (2002), further improving the computational efficiency of SMLE by alternate random schemes and higher-order subtransition densities. Finally, MCMC techniques were suggested by Elerian et al. (2001), Roberts and Stramer (2001), and Eraker (2001). Bayesian and likelihood methods often use a first-order Euler discretization to estimate the true transition densities, which was first introduced in Kloeden et al. (1992). Elerian et al. (1998) examined the Milstein scheme, an extended version of the Euler-Maruyama scheme that includes a correction term with second-order derivatives of the drift and diffusion functions for better accuracy. For instance, for the general Itô process in (1.0.2), the discretization implied by the Euler-Maruyama scheme is:

$$k_{t+1} = k_t + m\{k(t), t, \theta\} \Delta_t + n\{k(t), t, \theta\} (W_{t+1} - W_t) \quad (1.0.3)$$

and by the Milstein scheme is:

$$\begin{aligned} k_{t+1} = k_t + m\left\{k\left(t\right), t, \theta\right\}\Delta_t + n\left\{k\left(t\right), t, \theta\right\}\left(W_{t+1} - W_t\right) \\ + \frac{1}{2} n\left\{k\left(t\right), t, \theta\right\} \frac{dn\left\{k\left(t\right), t, \theta\right\}}{dk_t}\left[\left(W_{t+1} - W_t\right)^2 - \Delta_t\right]. \end{aligned} \tag{1.0.4}$$

Typically, the Euler-Maruyama scheme cannot use inter-observation times as a time step due to their large size. To address this, latent data points are augmented between each pair of observations by dividing the time interval into more minor equidistant points and replacing missing data with simulated data. However, extensive augmentation of data points can pose challenges for the Bayesian computation approach. In particular, the strong correlation between unknown parameters related to the diffusion or volatility coefficient and the missing data can result in slow convergence rates of naive sampling methods (Bernardo et al., 2003). Roberts and Stramer (2001) addressed the issue by transforming the SDE to achieve a constant diffusion coefficient, which resolved the dependence problem. Golightly and Wilkinson (2006) employed a sequential simulation filter that is not scalable for significant augmentation. Still, large data sets can cause degeneration of posterior samples, resulting in inadequate mixing of MCMC schemes like the Gibbs sampler. On the other hand, the innovation scheme proposed by Golightly and Wilkinson (2008) using analysis of a partially and discretely observed SDE as a missing data problem does not suffer from this problem. Pieschner and Fuchs (2020) incorporated the Milstein scheme into Bayesian data augmentation and determined that impractical constraints limit its application to multidimensional processes.

This thesis will focus on applying Markov Chain Monte Carlo (MCMC) to estimate the diffusion process in dynamic asset pricing. This approach is grounded in arbitrage and equilibrium arguments, which establish the functional relationship between asset prices and a range of economic fundamentals, including state variables, structural parameters, and market prices of risk.

The inference of asset pricing using the Bayesian approach typically requires determining the conditional distribution of observed prices, denoted by Y, given the state variables, X, and the parameters, θ. The combination of information from the model and the observed prices is essential for inferring parameters and state variables, which is facilitated by the posterior distribution $p\left(\theta, X|Y\right)$. Properly characterizing the joint probability distri-

bution $\mathrm{p}(\theta, X|Y)$ in continuous-time asset pricing models presents several challenges.

1. The observed prices are discrete, whereas the theoretical model assumes continuous evolution of prices and state variables.

2. The researcher's perspective often needs to capture the latent nature of state variables.

3. The dimensionality of the joint posterior distribution $\mathrm{p}(\theta, X|Y)$ is frequently high, leading to the inadequacy of conventional sampling techniques.

4. Several continuous-time models generate transition distributions for prices and state variables that are non-normal and nonstandard, thereby rendering the application of standard estimation methods such as maximum likelihood estimation (MLE) more intricate.

Examining the Heston model as an example in Heston (1993) is beneficial to contextualize the matter. Let S_t and ν_t be a pair of variables that simultaneously determine the price of an asset and its stochastic volatility.

$$dS_t = \mu S_t dt + \sqrt{\nu_t} S_t dW_t^s \left(\mathbb{P}\right) \tag{1.0.5}$$
$$d\nu_t = \kappa\left(\theta - \nu_t\right) + \xi\sqrt{\nu_t} dW_t^\nu \left(\mathbb{P}\right). \tag{1.0.6}$$

where $W_t^s(\mathbb{P})$ and $W_t^\nu(\mathbb{P})$ are Brownian motions under the physical measure $\mathbb{P}$, μ is the average price of the asset, θ is the long-run average variance of the price, κ is the rate at which ν_t revers to θ and ξ is the volatility of the volatility, which determines the variance of ν_t.

The price of derivatives that include options is a common observation among researchers. In the context of derivative pricing, assuming the presence of a probability measure denoted by $\mathbb{Q}$ under the no-arbitrage condition is common.

$$\begin{aligned} dS_t &= \mu S_t dt + \sqrt{\nu_t} S_t dW_t^s \left(\mathbb{Q}\right) \\ d\nu_t &= \kappa^{\mathbb{Q}}\left(\theta^{\mathbb{Q}} - \nu_t\right) + \xi\sqrt{\nu_t} dW_t^\nu \left(\mathbb{Q}\right). \end{aligned} \tag{1.0.7}$$

where $W_t^s(\mathbb{Q})$ and $W_t^\nu(\mathbb{Q})$, are defined under the risk-neutral measure $\mathbb{Q}$. The parameters $\kappa_\nu^{\mathcal{Q}}$ and $\theta^{\mathbb{Q}}$ capture the diffusive "price of volatility risk". Under

the risk-neutral measure $\mathbb{Q}$, the valuation of a call option on the underlying asset S_t at maturity T and strike price K can be expressed as follows:

$$C_t = C(S_t, \nu_t, \theta) = \mathrm{E}^{\mathbb{Q}} \left[\exp - \int_t^T r_s ds \left(S_T - K \right)_+ | \nu_t, S_t, \theta \right]. \tag{1.0.8}$$

where $\theta = \left(\theta^{\mathbb{P}}, \theta^{\mathbb{Q}} \right)$ are the physical and risk-neutral parameters, respectively, r_s refers to the instantaneous interest rate. The state variable X consists of the volatility.

In this case, empirical asset pricing aims to learn about the volatility-related state variables, risk-neutral and objective parameters, and model specifications from the observed stock returns and option prices. The joint posterior distribution $p(\theta, X|Y)$ denotes the information available from the sample regarding the objective and risk-neutral parameters in the context of parameters, which measures the estimation risk or the inherent uncertainty in estimating parameters. The marginal distribution $p\left(X|Y\right)$ provides a consistent approach for estimating stochastic volatility over time by integrating the model and data of the state variables, which is crucial for empirical issues requiring volatility estimations, such as option pricing or portfolio applications. Applying traditional approaches to this model is difficult since volatility is latent and the transition density for observed prices is unknown. MCMC techniques might address all these issues.

The development of MCMC algorithms for exploring $p(\theta, X|Y)$ begins with employing state space models to represent asset pricing problems, following Duffie (1996), which facilitates the construction of MCMC algorithms by emphasizing the modular structure of asset pricing models. The observation equation represents the conditional distribution of asset prices given the state variables and parameters. The evolution equation describes the state variable's dynamics based on the parameters. Equations (1.0.5) and (1.0.6) constitute the observation equations, while (1.0.8) represent the evolution equation, as previously stated. Nonlinear and non-Gaussian state space models can be used to describe asset pricing models, for which MCMC techniques are highly appropriate due to multiple factors:

1. Continuous-time asset models involve stochastic differential equations (SDEs) that incorporate Brownian motions, Poisson processes, and other independent and identically distributed (i.i.d.) shocks with easily characterizable distributions to determine the prices and state variables. When discretized over a finite time interval, the models can be

represented as time series models with standard error distributions such as the Normal, Discrete Mixture of Normal, or t-Students. Bayesian inference can be directly applied to these models using standard tools. Bayesian inference can be directly applied to these models using standard tools.

2. MCMC is a comprehensive method that concurrently estimates parameters and latent variables. MCMC calculates the distribution of latent variables and parameters based on observed data, presenting a distinct alternative to the common practice of utilizing approximate filters or latent variable proxies with noise in the existing literature. Discretely observed data enables researchers to distinguish between the impacts of jumps and stochastic volatility in models of interest rates or equity prices.

3. MCMC techniques facilitate researchers' evaluation of estimation and model risk. Estimation risk arises from the uncertainty in estimating parameters or state variables, while model risk arises from the uncertainty in specifying the model. The assessment of estimation risk is crucial in real-world scenarios and requires quantifying its effects.

The thesis is structured as follows. Chapter 2 presents the fundamental principles of Markov Chain Monte Carlo (MCMC) techniques. Chapter 3 demonstrates the application of MCMC techniques in asset pricing, which includes the creation of posterior approximation algorithms for GBM and Multivariate Merton's model. Chapter 4 introduces a term-structure model, specifically the Vasciek model, and its parameter estimation using MCMC techniques. Chapter 5 provides concluding remarks and proposes ideas for further research.

Chapter 2

MCMC methods

2.1 Gibbs Sampling

The Gibbs sampler is a basic MCMC algorithm described in Chapter 7 of Robert et al. (1999), which involves iterative direct sampling from all complete conditional distributions of parameters and latent variables. The subsequent statement defines a Gibbs sampler: given $\left(\theta^{(0)}, X^{(0)}\right)$

$$\begin{aligned} &1.\ \text{Draw } \theta^{(1)} \sim p\left(\theta | X^{(0)}, Y\right) \\ &2.\ \text{Draw } X^{(1)} \sim p\left(X | \theta^{(1)}, Y\right). \end{aligned} \tag{2.1.1}$$

The Gibbs sampler produces a sequence of random variables $\left\{\theta^{(h)}, X^{(H)}\right\}_{h=1}^{H}$ with an empirical distribution which converges to the joint posterior distribution $p\left(\theta, X | Y\right)$. The algorithm is iterated until convergence under the researcher's control of H.

If direct draw from $p\left(\theta | X, Y\right)$ and $p\left(X | \theta, Y\right)$ are not feasible, these distributions can be simplified. For instance, let us examine the Gibbs sampler: given $\left(\theta^{(0)}, X^{(0)}\right)$

$$\begin{aligned} &1.\ \text{Draw } \theta_1^{(1)} \sim p\left(\theta_1 | \theta_2^{(0)}, \theta_3^{(0)}, \theta_h^{(0)}, X^{(0)}, Y\right) \\ &2.\ \text{Draw } \theta_2^{(1)} \sim p\left(\theta_2 | \theta_1^{(1)}, \theta_3^{(0)}, \theta_h^{(0)}, X^{(0)}, Y\right) \\ &\vdots \\ &h.\ \text{Draw } \theta_h^{(1)} \sim p\left(\theta_h | \theta_1^{(1)}, \theta_2^{(1)}, \theta_3^{(1)}, \ldots \theta_{h-1}^{(1)}, X^{(0)}, Y\right) \end{aligned} \tag{2.1.2}$$

and then draw the states $(X|\theta, Y)$. If block drawing is not possible for the states, $p(X|\theta, Y)$ can be factorized into lower-dimensional distributions using similar reasoning.

The Gibbs sampler necessitates drawing from the entire set of conditional distributions conveniently. It often requires drawing random variables from standard continuous distributions and discrete distributions, such as Normal, t-Students, Beta, Gamma, Binomial, Dirichlet, etc.

2.2 The Griddy Gibbs Sampler

The Griddy Gibbs sampler approximates the conditional distribution using a discrete set of points. Assuming the continuous univariate distribution of θ, $p(\theta|X, Y)$ can be evaluated point-wise. However, direct draws from the nonstandard distribution of $p(\theta|X, Y)$ are not feasible. The Griddy Gibbs sample discretizes the continuous distribution of θ into M-points, $\{\theta_j\}_{j=1}^{M}$. Ritter and Tanner (1992) propose an algorithm based on this approximation:

$$\begin{aligned}
&\text{Step 1 : Compute } p\left(\theta_j|X, Y\right) \text{ and set } w_j = \frac{p\left(\theta_j|X, Y\right)}{\sum_{j=1}^{M} p\left(\theta_j|X, Y\right)}.\\
&\text{Step 2 : Approximate the inverse CDF of } p\left(\theta_j|X, Y\right).\\
&\text{Step 3: Generate a uniform on [0, 1] and invert the approximate CDF.}
\end{aligned} \tag{2.2.1}$$

The authors also address grid point selection issues and demonstrate the algorithm's ability to characterize conditional distributions in some cases effectively. The algorithm exhibits good performance with low parameter discretization. High-dimensional systems may exhibit poor algorithmic performance.

2.3 Metropolis-Hastings

The Gibbs sampler may only be applicable when it is convenient to sample one or more conditional distributions. In some cases, nonlinear models may have unrecognizable parameter conditional distributions. The distribution is identifiable in other cases, yet efficient sampling algorithms do not exist. The Metropolis-Hastings algorithm is commonly used in such scenarios.

Assuming a single parameter case, let us consider the task of sampling from a one-dimensional distribution $\pi(\theta)$ for simplicity. To produce samples θ^* from $\pi(\theta)$ using a Metropolis-Hastings algorithm, a proposal or candidate density $q\left(\theta^{(h+1)}|\theta^{(h)}\right)$ must be specified by the researcher. Typically, this conditional distribution highly depends on the other parameters and state variables. As in Metropolis et al. (1953), it is sufficient to evaluate the density ratio $\frac{\pi(\theta^*)}{\pi\left(\theta^{(h)}\right)}$. The general steps of the Metropolis-Hastings algorithm can be described as follows:

Step 1. : Draw θ^* from the proposal density $q\left(\theta^{(h+1)}|\theta^{(h)}\right)$

Step 2. : Accept θ^* with probability $\alpha\left(\theta^{(h)}, \theta^*\right)$ or reject it $\left(\text{i.e. } \theta^{(h+1)} = \theta^{(h)}\right)$ (2.3.1)

where:

$$\alpha\left(\theta^{(h)}, \theta^*\right) = \min\left(\frac{\pi\left(\theta^*\right)/q\left(\theta^*|\theta^{(h)}\right)}{\pi\left(\theta^{(h)}\right)/q\left(\theta^{(h)}|\theta^*\right)}, 1\right). \tag{2.3.2}$$

Implementing the Metropolis-Hastings algorithm involves generating random variables from a proposal distribution and evaluating an acceptance criterion. This algorithm decomposes the conditional distribution into a component easily sampled to generate candidate points and another component that determines the acceptance criteria. The acceptance criterion ensures the accuracy of the algorithm's stationary distribution. By iteratively applying the algorithm, it produces a sequence of samples $\left\{\theta^{(h)}\right\}_{h=1}^{H}$ that converge to the target distribution $\pi\left(\theta\right)$.

While the theory does not impose limitations on the proposal density, it is crucial to acknowledge that the choice of the proposal density heavily influences the algorithm's efficiency. Too thin tails in the proposal density can result in slow convergence of the algorithm. The algorithm may fail to converge if it gets trapped by a particular region of the parameter space.

Two specific instances of the Metropolis-Hastings algorithm deserve particular consideration: Independence Metropolis-Hastings (IMH) and Random Walk Metropolis-Hastings (RWMH).

2.3.1 Independence Metropolis–Hastings

The Metropolis-Hastings algorithm employs a proposal density, $q\left(\theta|\theta^{(h)}\right)$, which depends on the previous chain $\theta^{(h)}$, as well as other parameters and

states variables. An alternative approach is to sample the candidate θ^* from a distribution that is independent of the previous state, denoted as $q\left(\theta^*|\theta^{(h)}\right) = q^{(h+1)}$, which corresponds to an IMH algorithm described as follows:

$$\begin{aligned} &\text{Step 1. : Draw } \theta^* \text{ from the proposal density } q\left(\theta^*\right) \\ &\text{Step 2. : Accept } \theta^* \text{ with probability } \alpha\left(\theta^{(h)}, \theta^*\right) \end{aligned} \tag{2.3.3}$$

where:

$$\alpha\left(\theta^{(h)}, \theta^*\right) = \min\left(\frac{\pi\left(\theta^*\right)/q\left(\theta^*\right)}{\pi\left(\theta^{(h)}\right)/q\left(\theta^{(h)}\right)}, 1\right). \tag{2.3.4}$$

Although the candidate draws, θ^*, are independent of the previous state, the dependence of the acceptance probability on previous draws implies that the sequence $\left\{\theta^{(h)}\right\}_{h=1}^{H}$ is not independent. In IMH, the choice of the proposal distribution is often tailored to closely match particular characteristics of the target distribution.

2.3.2 Random-Walk Metropolis-Hastings

The RWMH algorithm, first introduced by Metropolis et al. (1953), uses a random-walk model to generate candidate points. Specifically, the candidate θ^* is obtained by adding an independent error term ϵ_t to the current state $\theta^{(h)}$, i.e., $\theta^* = \theta^{(h)} + \epsilon_t$. The error term ϵ_t is normally assumed to have a symmetric density function with fat tails, such as a t-Students distribution. In the RWMH, the choice of the proposal distribution is generic and does not consider the specific structural characteristics of the target distribution.

The equation $q\left(\theta|\theta^{(h)}\right) = q\left(\theta^{(h)}|\theta\right)$ illustrates the symmetry in the proposal density of the RWMH algorithm. This symmetry allows for simplification of the algorithm, leading to the following expression:

$$\begin{aligned} &\text{Step 1. : Draw } \theta^* \text{ from the proposal density } q\left(\theta^*|\theta^{(h)}\right) \\ &\text{Step 2. : Accept } \theta^* \text{ , i.e. } \theta^{(h+1)} = \theta^*, \text{ with probability } \alpha\left(\theta^{(h)}, \theta^*\right) \end{aligned} \tag{2.3.5}$$

where:

$$\alpha\left(\theta^{(h)}, \theta^*\right) = \min\left(\frac{\pi\left(\theta^*\right)}{\pi\left(\theta^{(h)}\right)}, 1\right). \tag{2.3.6}$$

The error term's variance is under the researcher's control in RWMH. The algorithm's error term variance should be adjusted to achieve an acceptable level of accepted draws, typically between 20-40%. A comprehensive discussion of this will be provided in the following part.

2.4 Adaptive Metropolis-Hastings

Continuing with RWMH from the previous section, we want to choose the scale of the proposal σ to optimize the results of the MCMC algorithm. An initial observation reveals that when the value of σ is minimal, the acceptance rate of proposed moves approaches unity. However, these accepted moves represent minor changes, leading to poor mixing of the Markov chain (Figure 2.2). Conversely, when σ is excessively large, most moves are rejected, resulting in the limited overall movement of the chain (Figure 2.1). Optimal performance requires a moderate value of σ, allowing for reasonable proposals and maintaining a reasonably high acceptance rate (Figure 2.3).

One practical approach to addressing this issue involves monitoring the acceptance rate of the algorithm, i.e., the proportion of accepted proposed moves. If the acceptance rate is close to 1, σ is too tiny (Figure 2.2). Conversely, if the acceptance rate is close to 0, σ is too large (Figure 2.1). We can balance exploration and exploitation in the Markov chain by selecting a value of σ that yields an acceptance rate far from both extremes (Figure 2.3) (see Brooks et al. (2011), chapter 4).

This straightforward guideline is a rule of thumb for scaling RWMH or Adaptive Metropolist-Hastings (Adaptive MH) algorithms. However, it allows for a wide range of choices, and further insights can be obtained under specific conditions. Even when we understand the criteria defining an optimal MCMC algorithm, the challenge remains to achieve this optimality, specifically, how to design and implement a Markov chain exhibiting (approximately) optimal characteristics. Gelman et al. (1997) demonstrated that multivariate proposal distribution yields an optimal acceptance rate of 0.234 for target distributions with a specific structure. According to Roberts and Stramer (2001) recommendation, a univariate proposal distribution should have a value of 0.44. Even though these theoretical outcomes were formulated based on relatively stringent assumptions, they have demonstrated their efficacy in addressing broader issues. A general approach in MCMC is adjusting acceptance probabilities by manually tuning the scale parameter to attain the above-mentioned rates. The manual adjustment of scale parameters can pose challenges in practical applications, mainly when the MCMC sampler involves numerous proposal distributions. Additionally, the presence of parameter correlations can exacerbate these difficulties.

We propose using the Robbins-Monro process, a stochastic search algorithm, to automatically tune scale parameters. The adaptive sampler adjusts

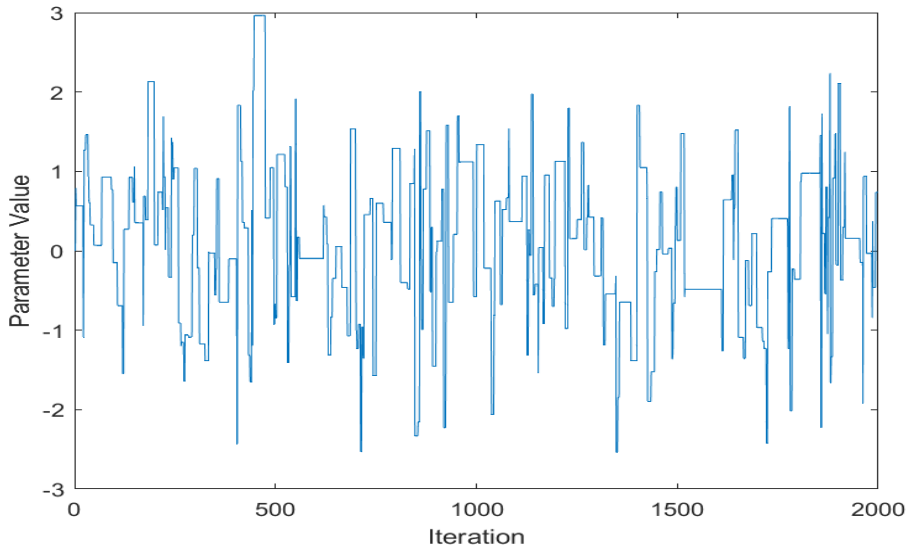

Figure 2.1: MCMC draws from the posterior distribution: large σ, small acceptance rate, and poor mixing.

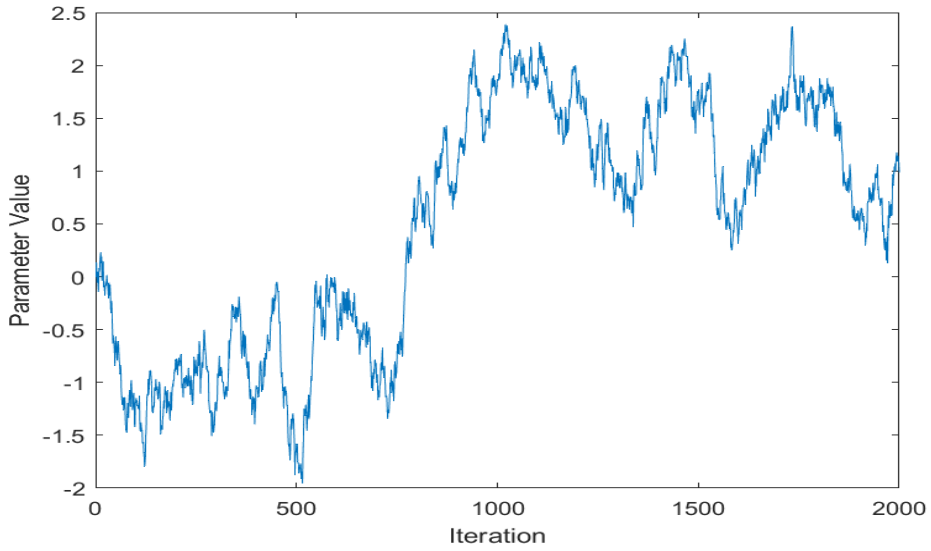

Figure 2.2: MCMC draws from the posterior distribution: small σ, large acceptance rate, and poor mixing.

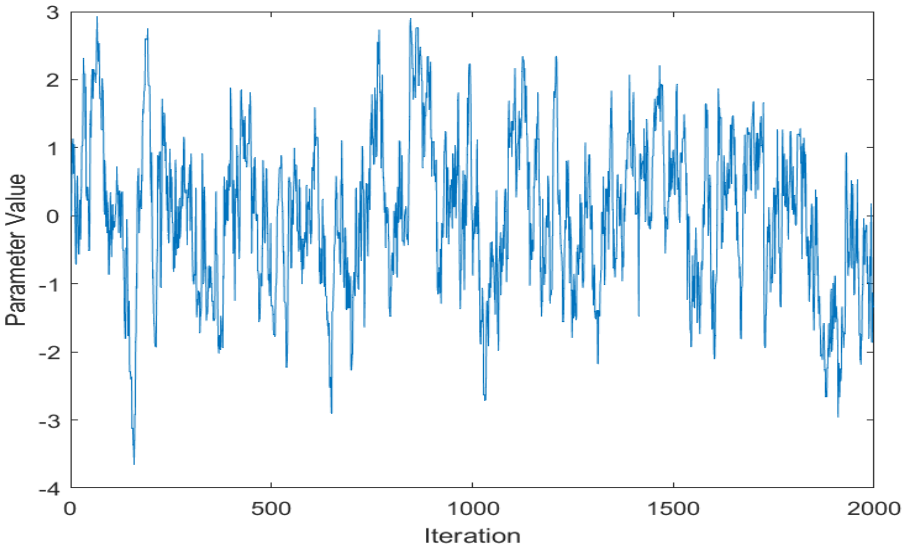

Figure 2.3: MCMC draws from the posterior distribution: medium σ, good acceptance rate, and good mixing.

the value of σ based on the acceptance or rejection of the previous MCMC proposed move. Specifically, it increases σ after an acceptance and decreases σ after a rejection. The step size, or change in σ, exhibits a linear decrease as the number of iterations in the Markov chain increases.

2.5 The Robbins-Monro process

In the context of this thesis, we only consider the algorithm in a univariate case. The Robbins-Monro process was initially developed for a typical scenario that is described in the following:

The probability of success for a binary response is denoted by $p(\sigma)$, with σ being a controllable parameter. The function $p(\sigma)$ is assumed to decrease monotonically concerning σ. In RWMH, it is generally valid to assume that a minor scale parameter σ corresponds to a higher acceptance rate $p(\sigma)$, and vice versa.

The objective is to determine the standard deviation, denoted by σ, corresponding to a predetermined probability of success, denoted by p^*. Let σ^* be the designated value such that $p(\sigma^*) = p^*$. The Robbins-Monro procedure

performs a probabilistic exploration through the execution of a series of attempts. During each trial, the value of σ is equated to the present estimation of σ^*. Suppose the trial yields a favourable outcome, σ^* increases, while an unfavourable outcome decreases. The notation σ_k represents the estimation of the true value σ^* obtained during the kth iteration, where $k = 1, 2, \ldots$. Robbins and Monro (1951) introduced the process that involved transitioning from σ_k to σ_{k+1} based on the following rule:

$$\sigma_{k+1} = \begin{cases} \sigma_k + \frac{c(1-p^*)}{k} & \text{if the } k\text{th trial is a success} \\ \sigma_k - \frac{cp^*}{k} & \text{if the } k\text{th trial is a failure} \end{cases} \tag{2.5.1}$$

- The size of steps is controlled by c, where the optimal value for c is $c = \frac{\sigma_k}{p^*(1-p^*)}$, which is suggested by Garthwaite et al. (2016)
- The optimal value of p^* has been determined to be 0.44 for the univariate case.

This fundamental approach involves iteratively enhancing the estimation of σ^* during each iteration of the Markov chain.

The initial values for a search process can either be randomly selected (e.g., $\sigma_1 = 1$) or deliberately chosen based on an estimated standard deviation of the original function. Following Garthwaite (1996), it is optional for σ_1 to be appropriately selected, given that the Robbins-Monro process can be supervised and a new search initiated if the initial value appears inadequate. Convergence of $\sigma_k \to \sigma^*$ may be delayed as the step size decreases with k.

Chapter 3

Asset Pricing

The following two chapters' modelling and estimation methods are based on Johannes and Polson (2010).

3.1 Geometric Brownian Motion

The Geometric Brownian Motion (GBM) is a fundamental asset pricing model for an asset's price, in which the stock price S_t satisfies a well-known SDE:

$$d \log S(t) = \mu dt + \sigma dW_{(t)}^{\mathbb{P}} \tag{3.1.1}$$

where μ is the continuously-compounded expected return and σ is the volatility. Prices are discretely recorded at equally spaced time intervals. A closed-form solution exists for continuously-compounded returns in this model:

$$Y_t = \log\left(\frac{S_t}{S_{t-1}}\right) = \mu + \sigma\epsilon_t, \tag{3.1.2}$$

where $\epsilon_t \sim N(0,1)$, i.i.d, $t = 1, 2, \ldots T$. The model for the continuously compounded return vector produces a conditional likelihood:

$$\begin{aligned} p\left(Y|\mu,\sigma^2\right) &= \left|2\pi\sigma^2 I_t\right|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma^2}\left(Y-\mu\right)'\left(Y-\mu\right)\right\} \\ &= \left(2\pi\sigma^2\right)^{-\frac{T}{2}} \exp\left\{-\frac{1}{2\sigma^2}\left(Y-\mu\right)'\left(Y-\mu\right)\right\}. \end{aligned} \tag{3.1.3}$$

To fully specify the model, we make the following assumption of independent priors for μ and σ^2:

$$\mu \sim \left(\mu_0, \sigma_0^2\right), \quad \sigma^2 \sim IG\left(\nu_0, S_0\right). \tag{3.1.4}$$

The Gibbs sampler is used to obtain samples from the joint posterior distribution of the parameters and iterates the following steps:

1. Draw μ from the Normal distribution $\mathcal{N}\left(\widehat{\mu}, K_\mu\right)$ where:

$$K_\mu = \frac{1}{\frac{1}{\sigma^2} + \frac{T}{\sigma_0^2}}, \quad \widehat{\mu} = K_\mu \left(\frac{1_T' Y}{\sigma^2} + \frac{\mu_0}{\sigma_0^2}\right). \tag{3.1.5}$$

2. Draw σ^2 from the Inverse Gamma distribution:

$$\mathcal{IG}\left(\nu_0 + \frac{T}{2}, S_0 + \frac{1}{2}\left(Y - \mu\right)'\left(Y - \mu\right)\right). \tag{3.1.6}$$

See **Appendix A** for the derivation of the full conditional distributions.

In the following, we provide an antithetic implementation of the Gibbs sampler. Antithetic techniques can enhance the efficiency of a Gibbs sampler, an iterative MCMC algorithm. Antithetic variables can produce negatively correlated sample pairs, resulting in quicker convergence and decreased variance. The Antithetic techniques with Gibbs, as described by Holmes and Jasra (2009) and Casarin et al. (2023), will be applied to the GBM in (3.1.1).

The first and second steps of the Antithetic Gibbs sampler are the same as (3.1.5) and (3.1.6), whereas the remaining steps are outlined below:

3. Using the same random standard normal number but a different sign with (3.1.5) in each draw.

4. Draw antithetic mean μ_1 from the Normal distribution $\mathcal{N}\left(\widehat{\mu_1}, K_{\mu_1}\right)$ where:

$$K_{\mu_1} = \frac{1}{\frac{1}{\sigma_1^2} + \frac{T}{\sigma_0^2}}, \quad \widehat{\mu_1} = K_{\mu_1} \left(\frac{1_T' Y}{\sigma_1^2} + \frac{\mu_0}{\sigma_0^2}\right). \tag{3.1.7}$$

5. Draw the antithetic variance σ_1^2 from the Inverse Gamma distribution with the antithetic mean μ_1:

$$\mathcal{IG}\left(-\nu_0 + \frac{T}{2}, S_0 + \frac{1}{2}\left(Y - \mu_1\right)'\left(Y - \mu_1\right)\right). \tag{3.1.8}$$

We will demonstrate the methodology using a GBM applied to the S&P500 equity index returns between 03/01/2007 and 08/12/2022.

	Posterior		
	M	**Std**	**CI**
	(a) Standard Gibbs Sampler		
μ	0.0253	1.699	[-0.0151,0.066]
σ^2	1.698	0.0375	[1.626,1.772]
	(b) Antithetic Gibbs Sampler		
μ_1	0.0256	0.00019	[0.0253,0.026]
σ_1^2	1.698	0.0371	[1.647,1.75]

Table 3.1: Estimated Geometric Brownian Motion on S&P 500 daily returns from 03/01/2007 to 08/12/2022: parameter estimates (M), their standard deviation (Std) and 95% Credible Interval (CI). Posterior quantities are approximated by standard Gibbs Sampler (panel a) and Antithetic Gibbs Sampler (panel b).

The Gibbs and Antithetic Gibbs samplers have been coded in Matlab (see **GBM_code**). We consider 6250 iterations for each sampler. The first 1250 iterations were discarded as a burn-in period of the MCMC chain. The posterior distributions were summarised using the last 5000 draws, resulting in the parameter estimates reported in Table 3.1, their standard deviations, and credible intervals.

The posterior histogram is presented in Figure 3.1, while additional graphs can be found in **Appendix B**. From all the graphs, we can see that:

- With the mean parameter (μ, μ_1), the antithetic techniques have a much lower standard deviation than the standard techniques.
- However, the variance parameter (σ^2, σ_1^2) is nearly identical in both techniques.
- Looking at the progressive mean plot in Figure B.2, we can see that with the antithetic techniques, these two parameters converged faster than the standard techniques.

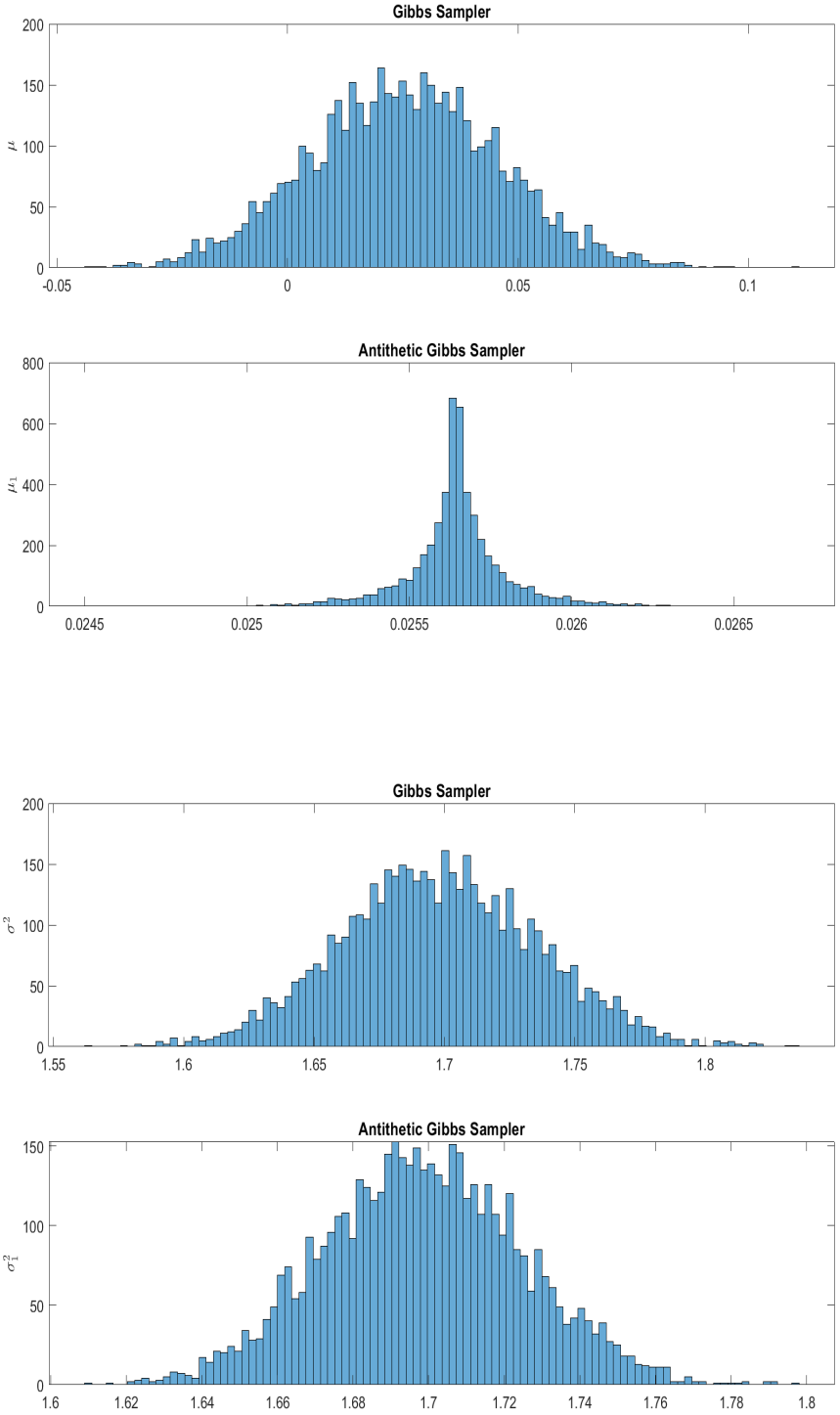

Figure 3.1: MCMC approximation of the posterior histograms for μ, μ_1, σ^2 and σ_1^2 (from top to bottom).

3.2 A Multivariate Version of Merton's Model

Matsuda (2004) introduced a jump-diffusion model as an extension of the Geometric Brownian Motion, specifically a multivariate version of the Merton model (1976). The asset price M-vector is a solution to the following stochastic differential equation (SDE):

$$dS(t) = \mu S(t)dt + \sigma S(t)dW_{(t)}^{\mathbb{P}} + d\left(\sum_{j=1}^{N_{(t)}^{\mathbb{P}}} S\left(\tau_j\right) - \left(e^{Z_j^{\mathbb{P}}} - 1\right)\right), \tag{3.2.1}$$

The given variables are a vector standard Brownian motion denoted as $W_{(t)}^{(\mathbb{P})}$, a diffusion matrix $\Sigma = \sigma\sigma'$, a Poisson process $N_t^{(\mathbb{P})}$ with constant intensity λ, and multivariate normal jump sizes $Z_j \in R^M$ with mean μ_z and variance-covariance matrix V_z. The model assumes correlated jumps in prices and allows for the inclusion of an idiosyncratic jump process. Its objective is to solve the SDE for continuously compounded equity returns within a daily interval, as follows:

$$\log\left(S_t/S_{t-1}\right) = \mu + \sigma\left(W_t^{\mathbb{P}} - W_{t-1}^{\mathbb{P}}\right) + \sum_{j=N_t^{\mathbb{P}}+1}^{N_{t+1}^{\mathbb{P}}} Z_j\left(\mathbb{P}\right), \tag{3.2.2}$$

To incorporate the variance correction, we construct the drift vector similar to the GBM equation (3.1.1). The Euler-Maruyama scheme in (1.0.3) is then used to discretize the jump component, assuming a single jump per time interval.

$$Y_t \equiv \log\left(S_t/S_{t-1}\right) = \mu + \sigma\epsilon_t + J_t Z_t \tag{3.2.3}$$

The probability of J_t being equal to 1 is denoted by λ, where $\lambda \in (0,1)$. Additionally, the structure of the jump sizes remains unchanged. The MCMC algorithm is used to sample from the joint distribution $p(\theta, X|Y)$, where θ represents a set of structure parameters including μ, Σ, λ, μ_z, and V_z, and X represents the state parameters including vectors of jump times and jump sizes denoted by J and Z, respectively. The observed prices are represented by the vector Y. The full likelihood for the continuous-compound returns is:

$$p\left(Y|\theta, J, Z\right) \propto |\Sigma|^{-\frac{Tm}{2}} \times \exp\left\{-\frac{1}{2}\left(Y - \mu - ZJ\right)'\left(I_T \otimes \Sigma^{-1}\right)\left(Y - \mu - ZJ\right)\right\}. \tag{3.2.4}$$

with m as the number of return series, T being the number of observations, I_T is an $T \times T$ identity matrix and 1_{Tm} is an $Tm \times 1$ vectors of one.

Next, we select standard independent conjugate priors for five structure parameters and two state parameters, conditioned on the remaining parameters, that is:

$$\begin{aligned} &\mu \sim \mathcal{N}_m(\mu_0, V_0), \quad V \sim \mathcal{IW}_m(\nu_0, S_0), \quad \mu_z \sim \mathcal{N}_m(\mu_{z_0}, V_{z_0}), \\ &V_z \sim \mathcal{IW}_m(\nu_{z_0}, S_{z_0}), \quad \lambda \sim \mathcal{B}(\alpha_0, \beta_0). \end{aligned} \tag{3.2.5}$$

The Gibbs sampler is employed for sampling from the joint posterior distribution of parameters. The sampling process involves the following steps:

1. Draw μ from the Normal distribution $\mathcal{N}_m\left(\widehat{\mu}, K_\mu^{-1}\right)$ where:

$$\begin{aligned} K_\mu &= \Sigma_0^{-1} + 1_{Tm}'\left(I_T \otimes \Sigma^{-1}\right) 1_{Tm} \\ \widehat{\mu} &= K_\mu^{-1}\left(V_0^{-1}\mu_0 + 1_{Tm}'\left(I_T \otimes \Sigma^{-1}\right)(Y - ZJ)\right). \end{aligned} \tag{3.2.6}$$

2. Draw V from the Inverse Wishart distribution:

$$\mathcal{IW}_m\left(\nu_0 + T, S_0 + \sum_{t=1}^{T}\left(Y_t - \mu - J_t Z_t\right)'\left(Y_t - \mu - J_t Z_t\right)\right). \tag{3.2.7}$$

3. Draw μ_z from the Normal distribution $\mathcal{N}_m\left(\widehat{\mu_z}, K_{\mu_z}^{-1}\right)$ where:

$$\begin{aligned} K_{\mu_z} &= V_{z0}^{-1} + 1_{Tm}'\left(I_T \otimes V_z^{-1}\right) 1_{Tm} \\ \widehat{\mu_z} &= K_{\mu_z}^{-1}\left(V_{z0}^{-1}\mu_{z0} + 1_{NT}'\left(I_T \otimes V_z^{-1}\right) Z\right). \end{aligned} \tag{3.2.8}$$

4. Draw V_z from the Inverse Wishart distribution:

$$\mathcal{IW}_m\left(\nu_{z_0} + T, S_0 + (Z - \mu_z)'(Z - \mu_z)\right). \tag{3.2.9}$$

5. Draw λ from the Beta distribution:

$$\mathcal{B}\left(\alpha + \sum_{t=1}^{T} J_t, T - \sum_{t=1}^{T} J_t\right). \tag{3.2.10}$$

6. Draw Z_t from the Normal distribution $\mathcal{N}(k_t, V_t)$ where:

$$\begin{aligned} V_t &= \left(J_t \Sigma^{-1} + V_z^{-1}\right)^{-1} \\ k_t &= V_z^{-1}\left(J_t \Sigma^{-1}(Y_t - \mu) + V_z^{-1}\mu_z\right). \end{aligned} \tag{3.2.11}$$

7. Draw J_t from the Bernoulli Distribution:

$$\mathcal{B}in(1, \varphi_t). \tag{3.2.12}$$

See **Appendix C** for the derivation of the full conditional distributions.

	Prior		**Posterior**		
	M	**Std**	**M**	**Std**	**CI**
μ_1	0	10	-0.0252	6.9746	[-13.9837,13.4947]
μ_2	0	10	0.0587	6.9743	[-13.4647,14.0015]
σ_1^2	1.67	0.248	2.4568	0.0546	[2.3525,2.5649]
σ_2^2	1.67	0.248	2.9051	0.0655	[2.7814,3.0346]
ρ_{12}	0	1.5	0.7369	0.001	[0.7228,0.7508]
μ_{z1}	0	10	0.0398	6.9876	[-13.4694,13.9581]
μ_{z2}	0	10	-0.0244	6.988	[-14.0236,13.5765]
σ_{z1}^2	1.67	0.248	2.6642	4.3961	[0.008,14.0372]
σ_{z2}^2	1.67	0.248	2.6347	4.3877	[0.008,13.9987]
ρ_{z12}	0	1.5	-0.7701	0.4169	[-0.9801,0.2241]
λ	0.5	0.29	0.1448	0.2168	[0.0002,0.5068]

Table 3.2: Bi-variate Merton's model estimated on S&P 500 and NASDAQ 100 daily returns from 03/01/2007 to 08/12/2022. Parameter estimates (M), their standard deviation (Std), and 95% Credible Interval (CI).

.

We will demonstrate the methodology using a bi-variate Merton's model applied to returns of the S&P500 and NASDAQ 100 equity indexes covering the period from 03/01/2007 to 08/12/2022. The sampler Matlab code for the Gibbs sampler ran for 6,250 iterations with a burn-in period of 1,250, as shown in **MMerton_code**. The posterior distributions were summarized using the last 5000 draws, resulting in the parameter estimates reported in Table 3.2, their standard deviations and credible intervals.

The posterior parameters histograms and the posterior jump times are presented in Figures 3.2-3.4. MCMC trace plots and progressive mean plots can be found in **Appendix D**. From all the tables and graphs, we can see that:

- The jump size characteristics of both returns are nearly identical and have a high negative correlation.
- Despite having identical standard deviations, the NASDAQ 100 index has more than twice the absolute value of posterior average returns compared to the S&P 500 index.
- Figures 3.2-3.4 present the posterior trace plot for all parameters and highlight the Gibbs sampler's movement within the posterior distribution after burn-in.
- The informative prior on λ indicates rare jump events, as Figure 3.4 confirms.
- The posterior estimates for the parameter jump size are quite not informative:
  - The observed posterior standard deviation shows a marked increase compared to the prior assumption.
  - The jump size correlation has a different sign from the returns correlation, which is quite unusual in the case of two major indexes.
  - The percentage of simulation step that correlates smaller than -0.8 is the highest.
- The sampler Matlab code in **MMerton_code** exhibits low efficiency due to the numerous parameter computations and updates executed in each iteration, especially the jump size and time parameters.

- The model's constant volatility results in clustered jumps in some periods, thereby capturing volatility time-variation that is not inherently incorporated in the model.

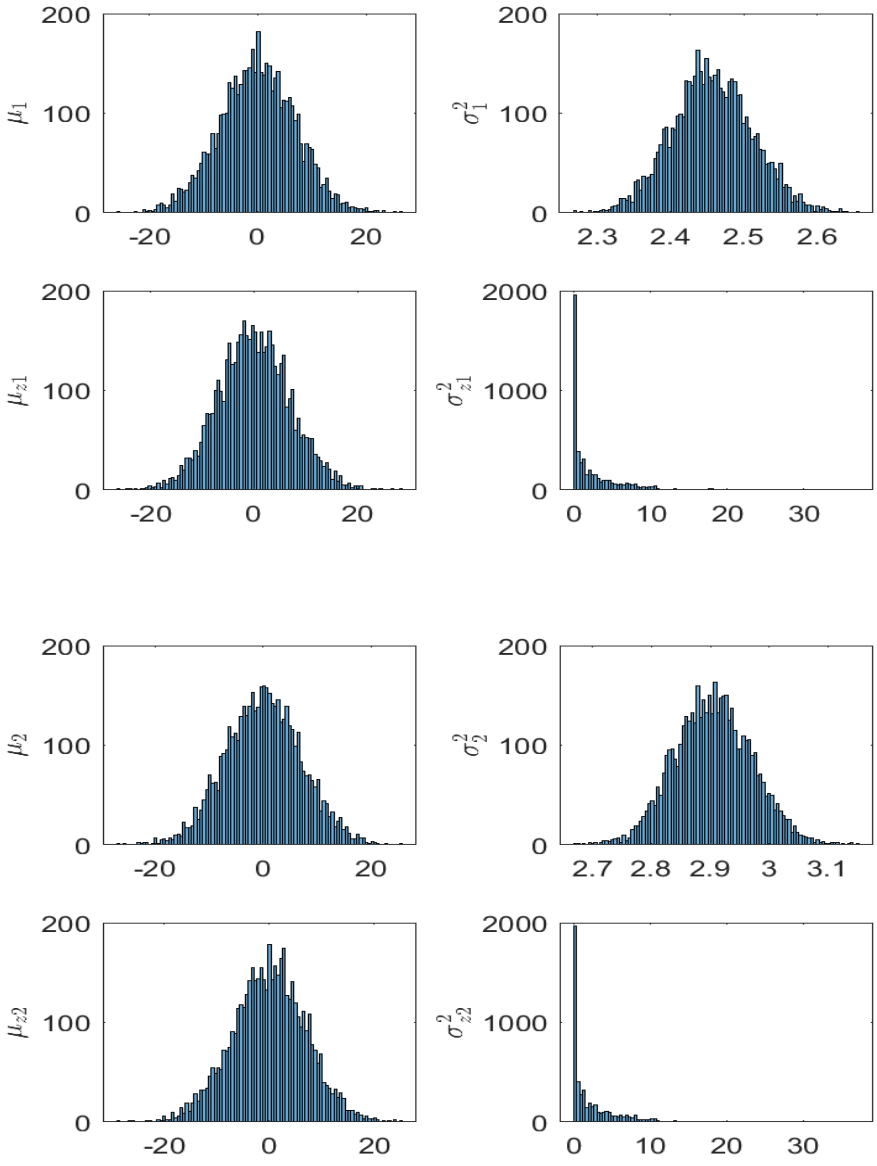

Figure 3.2: MCMC approximation of the posterior histograms for S&P 500 parameters and NASDAQ 100 parameters (from top to bottom).

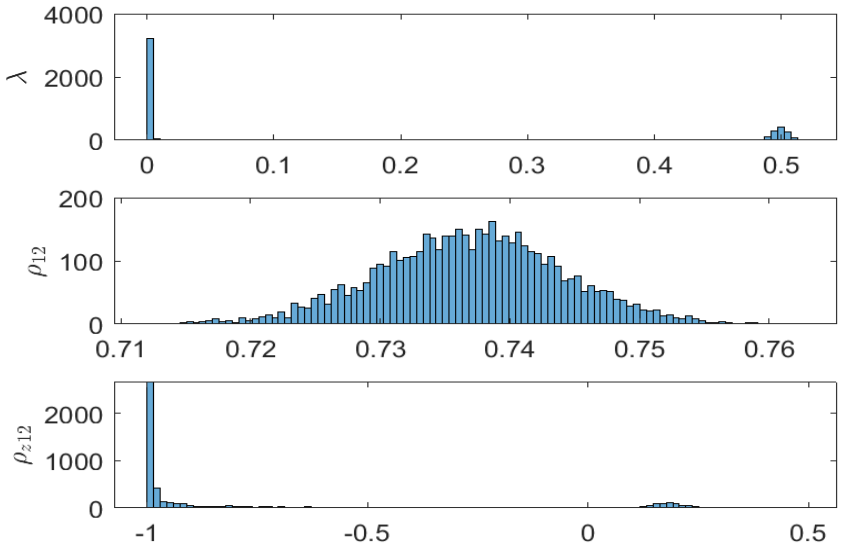

Figure 3.3: MCMC approximation of the posterior histograms for other parameters (jump intensity, return correlation, and jump size correlation).

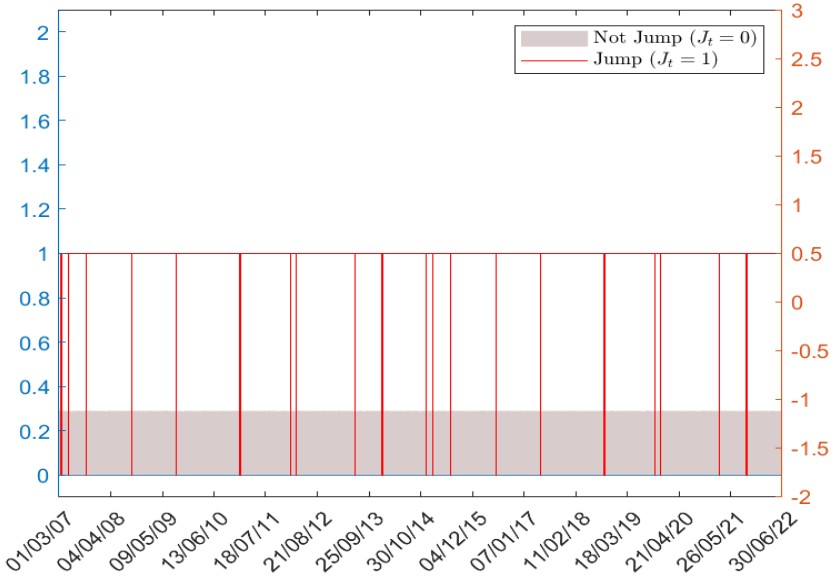

Figure 3.4: MCMC approximation of the posterior of the jump times

Chapter 4

Term Structure of Interest Rates

This Chapter illustrates Bayesian inference and MCMC for term structure models. This application's novelty consists of using an efficient sampling procedure for the posterior of the hidden spot rate. We replaced the standard Forward filtering Backward sampling with the band matrix technique proposed by Chan and Jeliazkov (2009). Moreover, the Adaptive MH method is used for the parameters instead of the RWMH usually employed.

4.1 Term structure model

The first term structure model is the uni-variate, Gaussian model of Vasicek (1977). Let $(\mathbb{R}, \mathcal{F}, \mathbb{P})$ a probability space and $(W_t^{\mathbb{P}})_{t\geq 0}$ be a Wiener (or standard Brownian motion) process under the physical measure $\mathbb{P}$. In Vasicek model, the spot rate $r_{(t)}$ is the Itô's process, which satisfies:

$$r_{(t)} = r_0 \int_0^t \left(a_1 - b_1 r_{(s)}\right) ds + \int_0^t \sigma_r dW_{(s)}^{\mathbb{P}},\ t \geq 0 \tag{4.1.1}$$

or equivalently, the SDE

$$dr_{(t)} = \left(a_1 - b_1 r_{(t)}\right) dt + \sigma_r dW_{(t)}^{\mathbb{P}},\ t \geq 0 \tag{4.1.2}$$

with given initial value r_0, where a_1 and b_1 are the means and the speed of reversion parameters, respectively. The parameters are measured under the physical measure $\mathbb{P}$.

The spot rate evolves under the equivalent martingale measure (risk-neutral measure) $\mathbb{Q}$ via:

$$dr_{(t)} = \left(a_2 - b_2 r_{(t)}\right) dt + \sigma_r dW^{\mathbb{Q}}_{(t)}, \tag{4.1.3}$$

where a_2 and b_2 are measured under $\mathbb{Q}$ and $(W^{\mathbb{Q}}_{(t)})_{t\geq 0}$ is a standard Brownian motion under $\mathbb{Q}$. In the following, we define the parameter vectors $\theta^{\mathbb{P}} = (a_1, b_1, \sigma_r) \in \Theta$ and $\theta^{\mathbb{Q}} = (a_2, b_2, \sigma_r) \in \Theta$, where Θ is the parameter space.

Let $\mathbb{F} = (\mathcal{F}_t)$ be the natural filtration generated by $(r_{(t)})_{t\geq 0}$ and denote with $\mathbb{E}^{\mathbb{Q}}_t(\cdot) = \mathbb{E}^{\mathbb{Q}}(\cdot|\mathcal{F}_t)$ the conditional expectation, where $\mathbb{Q}$ is the risk-neutral measure. The price of a zero coupon bond at time t is the expected value of the value of 1\$ discounted by the risk-free instantaneous rate between t and $t+\tau$:

$$\begin{aligned} P\left(r_{(t)}, \tau\right) &= \mathrm{E}^{Q}_t \left[\exp\left(\int_t^{t+\tau} r_{(s)} ds\right)\right] \\ &= \exp\left[\alpha\left(a_2, b_2, \sigma_r, \tau\right) + \beta\left(b_2, \sigma_r, \tau\right) r_{(t)}\right] \end{aligned} \tag{4.1.4}$$

where the loading functions are:

$$\begin{aligned} \beta\left(b_2, \sigma_r, \tau\right) &= \frac{\exp\left(-b_2\tau\right) - 1}{b_2} \\ \alpha\left(a_2, b_2, \sigma_r, \tau\right) &= \frac{1}{2}\left[\frac{\sigma_r^2}{b_2^2} - \frac{a_2}{b_2}\right]\left(\tau - \beta\left(b_2, \sigma_r, \tau\right)\right) - \frac{\sigma_r^2}{4b_2}\beta\left(b_2, \sigma_r, \tau\right)^2. \end{aligned} \tag{4.1.5}$$

We assume that there exists a panel of zero coupon bonds, continuously-compounded yields $Y_t = [Y_{t,\tau_1}, ..., Y_{t,\tau_m}]$, where $Y_{t,\tau} = -\ln P\left(r_{(t)}, \tau\right)$ and $\tau = [\tau_1, ..., \tau_k]$ are the yield to maturity and the set of maturities, respectively. Given the maturity, we will write the loading functions in the vector form: $\tilde{\alpha}\left(a_2, b_2, \sigma_r, \tau\right) = \left(\alpha_{t,\tau_1}, ..., \alpha_{t,\tau_m}\right)'$ and $\tilde{\beta}\left(b_2, \sigma_r, \tau\right) = \left(\beta_{t,\tau_1}, ..., \beta_{t,\tau_m}\right)'$.

4.2 Inference

To estimate all of the parameters, we use a discrete-time approximation of the first order (Euler-Maryama approximation, in (1.0.3)) of the SDEs and obtain:

$$\begin{aligned} Y_t &= \tilde{\alpha}\left(a_2, b_2, \sigma_r, \tau\right) + \tilde{\beta}\left(b_2, \sigma_r, \tau\right) r_t + \epsilon_t \\ r_{t+1} &= r_t + a_1 - b_1 r_t + \sigma_r \eta_{t+1} \end{aligned} \tag{4.2.1}$$

with $\epsilon_t \sim \mathcal{N}(0, \Sigma)$ and $\eta_t \sim \mathcal{N}(0, 1)$, i.i.d, $t = 1, 2, \ldots T$.

Prior to going into MCMC estimation, we will provide a brief overview of parameter identification. Equations (4.1.4) and (4.1.5) show that:

- The identification of a_1 and b_1 is based exclusively on the short rate dynamics.
- $\tilde{\alpha}(a_2, b_2, \sigma_r, \tau)$ is linear in a_2, and both $\tilde{\alpha}(a_2, b_2, \sigma_r, \tau)$ and $\tilde{\beta}(b_2, \sigma_r, \tau)$ are non-linear in b_2;
- The identification of Σ is based only on the bond yields cross-section.
- The joint identification of σ_r is achieved by analyzing both the cross-section of bond yields and the latent short-rate dynamics.

Next, we choose independent standard conditionally conjugate priors for each objective measure parameter given the rest, that is:

$$\begin{gathered}(a_1, b_1) \sim \mathcal{N}_2(z_0, V_0), \quad a_2 \sim \mathcal{N}\left(m_{2a}, \sigma_{2a}^2\right), \quad b_2 \sim \mathcal{N}\left(m_{2b}, \sigma_{2b}^2\right), \\ \Sigma \sim \mathcal{IW}_m(\nu_0, S_0), \quad \sigma_r^2 \sim \mathcal{IG}(\nu_r, S_r), \quad r_0 \sim \mathcal{N}\left(c_0, d_0^2\right).\end{gathered} \tag{4.2.2}$$

. Assuming stationary, we can ensure $b_1 > 0$ by eliminating any MCMC draws with $b_1 < 0$, as we anticipate a non-explosive latent short rate. The posterior distribution is $p(\theta, r|Y)$ where $\theta = (a_1, b_1, a_2, b_2, \sigma_r, \Sigma, r_0)$ is the parameter vector and $Y = (Y_1, ..., Y_T)'$ and $r = (r_1, ..., r_T)'$ are the observation vectors.

The Gibbs sampler is used to obtain samples from the joint posterior distribution of the parameters. The sampler iterates the following steps:

1. Generate r from the Normal distribution $\mathcal{N}(\widehat{r}, V_r^{-1})$ where:

$$V_r = w_1, \quad \widehat{r} = V_r^{-1} w_2 = \frac{w_2}{w_1}. \tag{4.2.3}$$

2. Generate $z = (a_1, b_1)$ from the Normal distribution $\mathcal{N}_2(\hat{z}, K_z^{-1})$ where:

$$K_z = \left(V_0^{-1} + \frac{\tilde{r}'\tilde{r}}{\sigma_r^2}\right), \quad \hat{z} = K_z^{-1}\left(V_0^{-1} z_0 + \frac{\tilde{r}'\Delta r}{\sigma_r^2}\right). \tag{4.2.4}$$

3. Generate Σ from the Inverse Wishart distribution:

$$\mathcal{IW}_m\left(\nu_0 + T, \sum_{t=1}^{T}\left(Y_t - \tilde{\alpha} - \tilde{\beta} r_t\right)'\left(Y_t - \tilde{\alpha} - \tilde{\beta} r_t\right)\right). \tag{4.2.5}$$

4. Generate a_2 from the Normal distribution $\mathcal{N}\left(\widehat{a_2}, D_{2a}\right)$ where:

$$D_{2a} = \left(0.25Tg_1^{'}\Sigma^{-1}g_1 + \frac{1}{\sigma_{2a}^2}\right)^{-1}, \quad \widehat{a_2} = D_{2a}\left(u_1 + u_2 + \frac{m_{2a}}{\sigma_{2a}^2}\right). \tag{4.2.6}$$

5. Generate b_2 from Adaptive MH with proposal distribution $\mathcal{N}\left(b_2, \sigma_k^2\right)$ and σ_k is adjusted at each step following the formula from (2.3.5), (2.3.6) and (2.5.1).

6. Generate σ_r^2 from Adaptive MH algorithm with truncated proposal distribution $\mathcal{N}\left(0, \sigma_r^2, \sigma_j^2\right)$ and σ_j is adjusted at each step following the formula from (2.3.5), (2.3.6) and (2.5.1).

7. Generate r_0 from the Normal distribution $\mathcal{N}\left(\widehat{r_0}, D_{r_0}\right)$ where:

$$D_{r_0} = \left(\frac{n_1^2}{\sigma_r^2} + \frac{1}{d_0^2}\right)^{-1}, \quad \widehat{r_0} = D_{r_0}\left(\frac{n_1\left(r_1 - a_1\right)}{\sigma_r^2} + \frac{c_0}{d_0^2}\right). \tag{4.2.7}$$

See **Appendix E** for the derivation of the full conditional distributions.

We conduct a simulation experiment to assess the posterior approximation's performance. The simulated yields are generated using specific parameter values based on findings from the empirical term structure literature, as reported by Pearson and Sun (1989). In particular, the mean reversion parameter b_1 in the latent short rate process is set to 0.1036 for monthly observation data. The long-run mean of r_t, denoted as a_1, is assigned a value of 0.0518. Similarly, the risk-neutral parameters b_2 and a_2 are set to 0.0529 and 0.0518, respectively. The volatility parameter σ_r^2 is set at 0.0131 ($\sigma_r = 0.1145$).

The correlation matrix represents the variance-covariance matrix of the pricing errors Σ in the observation equation,

$$\rho = \begin{bmatrix} 1 & 0.2 & 0.2 & 0.2 \\ 0.2 & 1 & 0.2 & 0.2 \\ 0.2 & 0.2 & 1 & 0.2 \\ 0.2 & 0.2 & 0.2 & 1 \end{bmatrix}$$

and the diagonal matrix with the maturity-specific standard deviations on the main diagonal

$$D = \begin{bmatrix} 0.2 & 0 & 0 & 0 \\ 0 & 0.3 & 0 & 0 \\ 0 & 0 & 0.4 & 0 \\ 0 & 0 & 0 & 0.5 \end{bmatrix}$$

which correspond to standard deviations in short rates of 20%, 30%, 40% and 50% for the respective terms to maturity, τ_j, $j = 1, 3, 6, 12$.

We first simulate a time series for the latent factor r_t monthly to generate the observed yield curves, using the specified parameter values and the initial condition $r_0 = 1\%/12$. Based on these simulated values of r_t, we then generate 192 monthly observed yield curves from 03/01/2007 to 08/12/2022. The yield curves are generated for four different terms to maturity, namely $\tau_j = 1, 3, 6, 12$ months. The first equation in (4.2.1) is used for this simulation process. To calculate the values of $\tilde{\alpha}(a_2, b_2, \sigma_r, \tau)$ and $\tilde{\beta}(b_2, \sigma_r, \tau)$ in (4.1.5), the parameter values mentioned above are used.

	Posterior			
	True value	**M**	**Std**	**CI**
a_1	0.0518	0.0534	0.0191	[0.0160,0.0685]
b_1	0.1036	0.126	0.0392	[0.0706,0.2027]
a_2	0.0529	0.0552	0.0017	[0.0518, 0.0586]
b_2	0.0529	0.0574	0.0044	[0.0500,0.0670]
σ_r^2	0.0131	0.0146	0.0007	[0.0132,0.0161]

Table 4.1: Simulation experiments for Term-structure model from 03/01/2007 to 08/12/2022. Parameter estimates (M), their standard deviation (Std), and 95% Credible Interval (CI).

The simulation exercise aims to determine how well the Bayesian MCMC method performs in estimating the true parameters of the Vasicek model

in (4.2.1), in particular, as compared with estimation parameters using our assumption. The sampler Matlab code in **Vasicek_code** is executed using Simulation data for 6250 iterations, with the first 1250 iterations being discarded as a burn-in period. The parameter estimate presented in Table 4.1 was obtained by summarizing the posterior distribution using the most recent 5000 draws. From the table, we can see that:

- The MCMC method quite accurately estimates the parameters of the true process, as evidenced by the closeness of the point estimates obtained from the posterior means to the actual parameter values.
- However, the estimated parameters exhibit an upward bias towards the true value, indicating the necessity of augmenting the data to capture the dynamics of the underlying process better and reduce biases in parameter estimation.

4.3 US interest rate application

Using a term structure model and the MCMC methodologies, we estimate the latent spot rate from the actual monthly data for the US Treasury Rates for One Month, Three Months, Six Months, and One Year from 03/01/2007 to 08/12/2022. Using the same sampler Matlab code from **Vasicek_code** to this actual data with the same number of iterations (6250) and burn-in period (1250) resulted in the parameters estimate that is shown in Tables 4.2 and 4.3.

	Posterior		
	M	**Std**	**95% CI**
a_1	0.5217	0.1034	[0.3202,0.7282]
b_1	0.9096	0.1484	[0.619,1.1973]
σ_r^2	0.4841	0.0131	[0.4502,0.4993]

Table 4.2: Parameters estimates for state equation from 03/01/2007 to 08/12/2022. Parameter estimates (M), their standard deviation (Std), and 95% Credible Interval (CI).

	Posterior		
	M	**Std**	**CI**
a_2	0.6466	0.0792	[0.5404,0.8681]
b_2	0.4739	0.0247	[0.4023,0.4987]
σ_1^2	2.5824	0.2642	[2.1098,3.1389]
σ_2^2	2.603	0.2685	[2.1245,3.1673]
σ_3^2	2.5438	0.2682	[2.0786,3.1165]
σ_4^2	2.4134	0.2733	[1.9313,3.01]
ρ_{12}	0.7909	0.0014	[0.7879,0.7932]
ρ_{13}	0.7674	0.0051	[0.7563,0.7761]
ρ_{14}	0.7507	0.0085	[0.7318,0.7653]
ρ_{23}	0.7852	0.0023	[0.7803,0.7892]
ρ_{24}	0.774	0.0047	[0.7636,0.7821]
ρ_{34}	0.7927	0.0013	[0.7898,0.7949]

Table 4.3: Parameter estimates for observation equation from 03/01/2007 to 08/12/2022. Parameter estimates (M), their standard deviation (Std) and 95% Credible Interval (CI).

The latent short rate vs. actual rate, average acceptance rate, and posterior histogram graphs are shown in Figures 4.1-4.6, and additional graphs can be found in **Appendix F**. From all of the tables and graphs, we can see that:

- The variances and correlation parameters for four maturities have a similar pattern (mean, standard deviation, and posterior distribution).
- The mean and reversion parameters in the observation and state equations have quite the same behaviour as the simulated parameters set up in the previous section but have a different posterior distribution.

  - With the state equation, both parameters a_1 and b_1 have normal distributions.

  - With the observation equation, parameters a_2 and b_2 might have left-skewed and right-skewed distributions, respectively.

  $\Rightarrow$ This might reflect the pattern of actual financial data about interest rates.

- The average acceptance rates of b_2 and σ_r^2 are around 24.5% and 17.5%, respectively (Figure 4.1), which is quite far from the target acceptance rate of 44% proposed for Adaptive MH for the univariate case, which might be due to the problem of identification in the loading functions in (4.1.5), leading to an unsuitable shape for the likelihood function of both parameters.

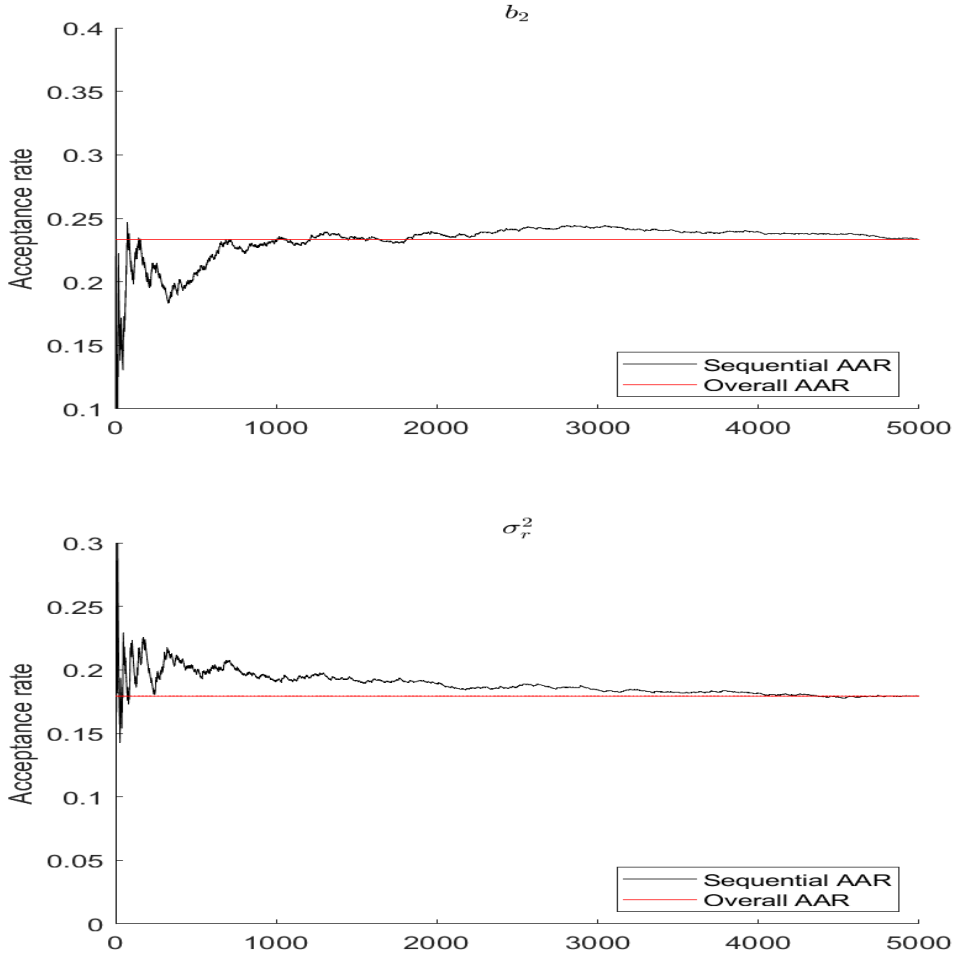

Figure 4.1: MCMC average acceptance rate for b_2 and σ_r^2 with Adaptive MH (from top to bottom).

- The positive pair correlation with nearly equal magnitude (Table 4.3) among the four maturities reflects the observed co-movements in interest rates, as depicted in Figure 4.2. However, applying the one-factor Vasicek model to this pattern might not align with the behaviour above.

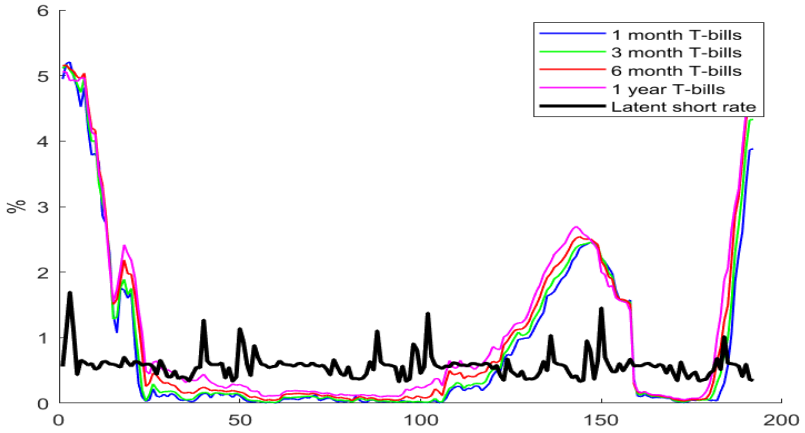

Figure 4.2: 1 month T-bill, 3 months T-bill, 6 months T-bill, 1 year T-bill, and the estimated latent short rate path, resulting from fitting a one-factor Vasicek model to monthly observations of the T-bill rate from 03/01/2007 to 08/12/2022.

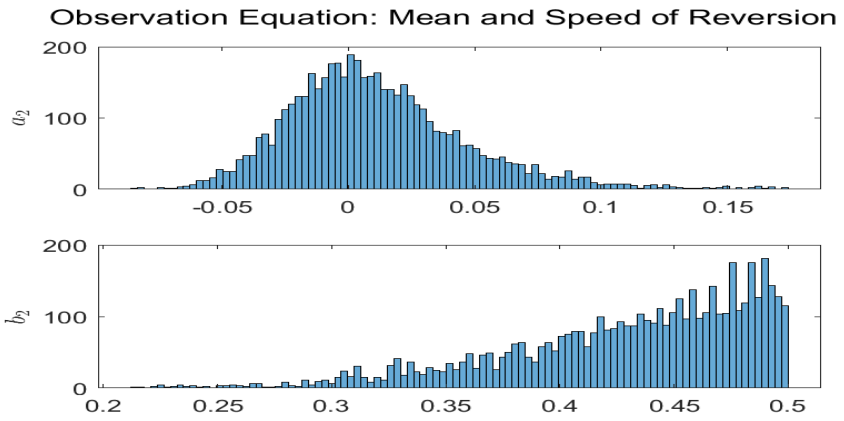

Figure 4.3: MCMC approximation of the posterior histograms for the parameters of the observation equation.

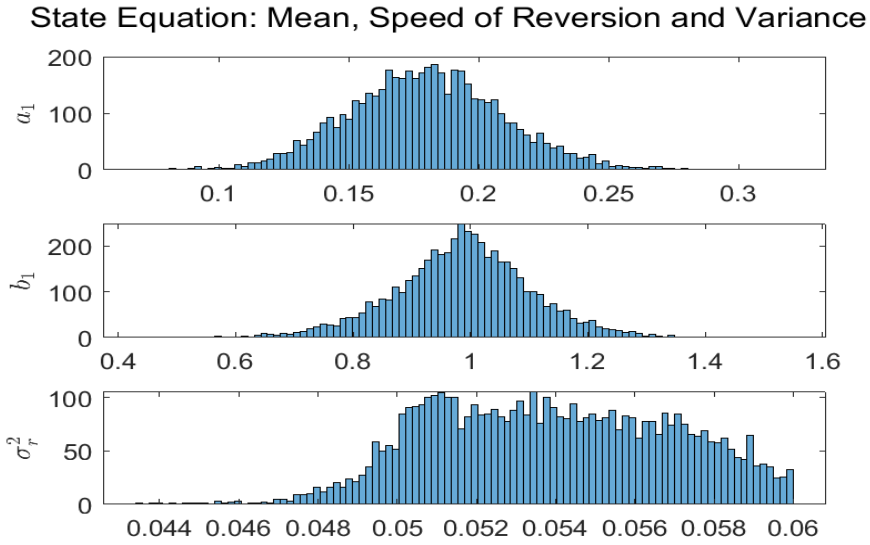

Figure 4.4: MCMC approximation of the posterior histograms for the parameters of the state equation.

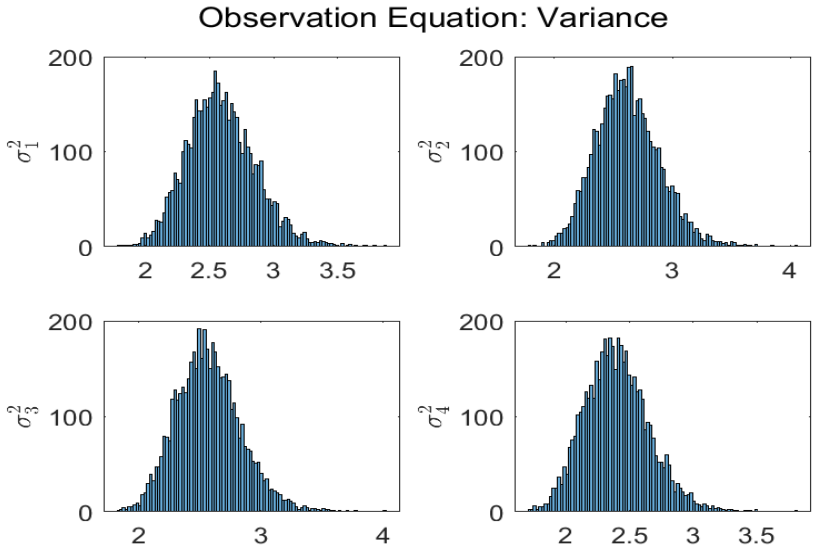

Figure 4.5: MCMC approximation of the posterior histograms for the variance parameters of the observation equation.

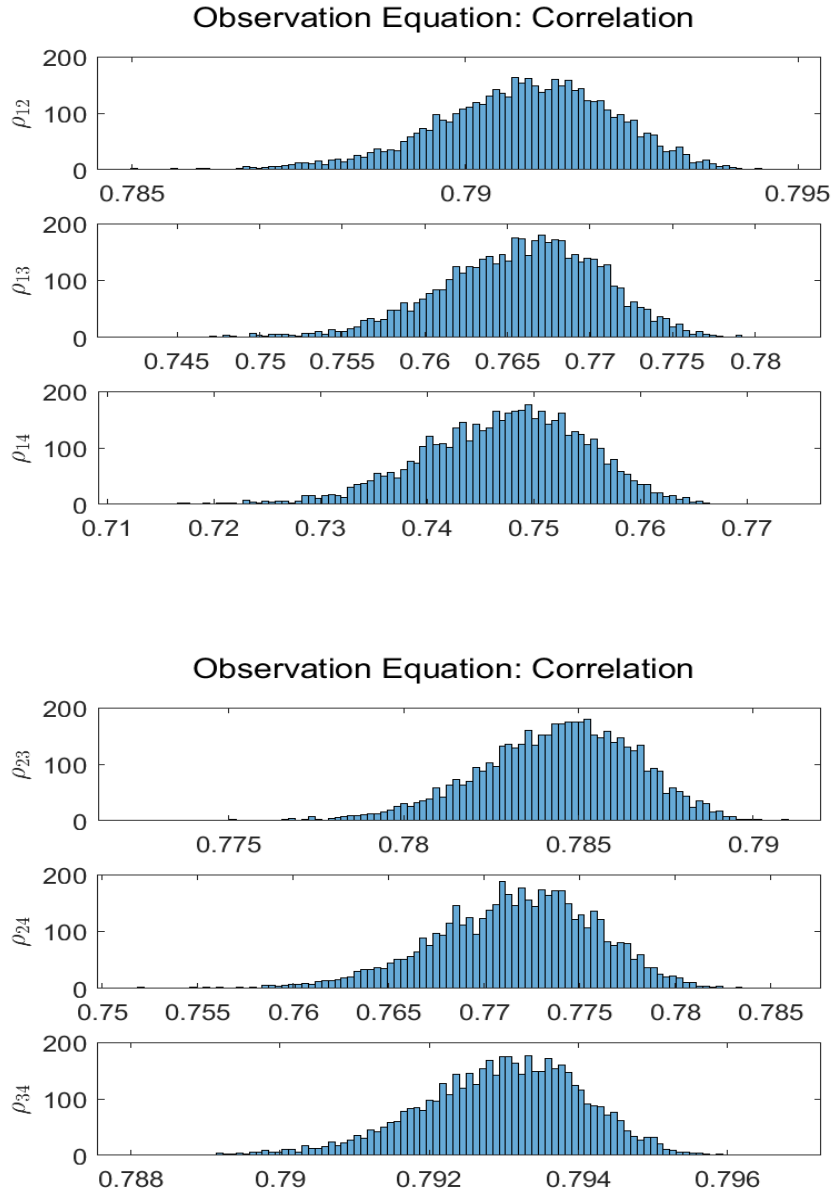

Figure 4.6: MCMC approximation of the posterior histograms for the correlation parameters of the observation equation.

Chapter 5

Conclusion

This thesis deals with Bayesian inference methods for the diffusion processes. We review some numerical methods for posterior approximation and develop posterior inference for standard asset pricing models and interest rate term structures.

In asset pricing, we applied Bayesian inference and antithetic techniques to estimate the parameters of the GBM, which assumes constant volatility and log-normal returns. By employing Bayesian inference, we incorporated prior knowledge and updated it with observed data to obtain posterior distributions for the mean and variance. In addition, through antithetic techniques, we efficiently generated pairs of correlated samples, enabling us to improve the accuracy and convergence rate of our estimation process, which allowed us to make more reliable inferences about the dynamics of the GBM and quantify the uncertainty associated with our parameter estimates.

Furthermore, we extended our analysis to the Multivariate Merton's model, which incorporates jump processes into asset price dynamics. Including price jumps in the process enables modelling of non-Gaussian behaviour and extreme events in financial markets. The application of the Bayesian framework facilitated the estimation of parameters related to jump intensity, return correlation, and jump size correlation, thereby offering valuable insights into the dynamics of these variables. Our findings revealed the importance of considering jumps in asset pricing models and their impact on risk assessment. However, there still exist some limits regarding the jump parameters and computational times, and further research in this area could explore more complex diffusion processes, such as stochastic volatility models or jump-diffusion models, and investigate advanced Bayesian techniques, such as ran-

dom scan, in which the variables in the joint distribution are updated one at a time in a random order, instead of updating them sequentially to reduce computational times. Additionally, expanding the application of Bayesian inference to other areas of finance, such as option pricing or portfolio optimization, would further improve our understanding of financial markets and enhance risk management practices.

In the interest rate term structure, we focused on applying the Vasicek model, which characterizes the evolution of interest rates. We obtained a comprehensive view of the term structure dynamics by considering the posterior distributions of the parameters using MCMC methods. From there, we could make informed predictions about interest rate movements. Nevertheless, one typical limitation of the Vasicek model is that we do not control for the negative value of the interest rate, which might affect our estimation and prediction of the term structure. One possible alternative is to draw the posterior parameters of the Cox-Ingersoll-Ross (CIR) model or the Hull-White model, which includes additional parameters to capture the potential for negative interest rate levels. Another alternative is to consider term structure models with regime switches, in which the long-run mean and volatility can change over time. Last, we can extend the model suggested above with the jump component and more hidden factors or consider more maturity with augmentation techniques to better capture the complex behaviour of the interest rate term structure.

In summary, Bayesian inference is valuable for estimating and modelling diffusion processes in asset pricing and interest rate term structure. Integrating Bayesian techniques and financial models presents novel opportunities for research and practical implementation, facilitating enhanced comprehension and decision-making within the finance domain.

Bibliography

Ait-Sahalia, Y. (1998). Maximum Likelihood Estimation of Discretely Sampled Diffusions: A closed-form approximation approach.

Arkin, A., Ross, J., and McAdams, H. H. (1998). Stochastic Kinetic Analysis of Developmental Pathway Bifurcation in Phage λ-Infected Escherichia coli Cells. *Genetics*, 149(4):1633–1648.

Bernardo, J., Bayarri, M., Berger, J., Dawid, A., Heckerman, D., Smith, A., and West, M. (2003). Non-centered parameterisations for hierarchical models and data augmentation. In *Bayesian Statistics 7: Proceedings of the Seventh Valencia International Meeting*, volume 307. Oxford University Press London.

Bibby, B. M. (1996). On estimation for discretely observed diffusions: A review. *Theory of Stochastic Processes*, 2:49–56.

Brooks, S., Gelman, A., Jones, G., and Meng, X.-L. (2011). *Handbook of Markov Chain Monte Carlo*. CRC press.

Casarin, R., Frattarolo, L., Radu, C., and Robert, C. P. (2023). Living on the Edge: An Unified Approach to Antithetic Sampling. *Statistical Science, forthcoming*.

Chan, J. C. and Jeliazkov, I. (2009). Efficient simulation and integrated likelihood estimation in state space models. *International Journal of Mathematical Modelling and Numerical Optimisation*, 1(1-2):101–120.

Dixit, A. K. (1993). *The Art of Smooth Pasting*, volume 55. Taylor & Francis.

Duffie, D. (1996). State-Space Models of the Term Structure of Interest Rates. In *Stochastic Analysis and Related Topics V*, pages 41–67. Springer.

Duffie, D. and Glynn, P. (2004). Estimation of Continuous-Time Markov Processes Sampled at Random Time Intervals. *Econometrica*, 72(6):1773–1808.

Durham, G. B. and Gallant, A. R. (2002). Numerical Techniques for Maximum Likelihood Estimation of Continuous-Time Diffusion Processes. *Journal of Business & Economic Statistics*, 20(3):297–338.

Elerian, O., Chib, S., and Shephard, N. (2001). Likelihood Inference for Discretely Observed Nonlinear Diffusions. *Econometrica*, 69(4):959–993.

Elerian, O. et al. (1998). A note on the existence of a closed form conditional transition density for the Milstein scheme. *Economics discussion paper*, page W18.

Eraker, B. (2001). MCMC Analysis of Diffusion Models With Application to Finance. *Journal of Business & Economic Statistics*, 19(2):177–191.

Garthwaite, P. H. (1996). Confidence Intervals from Randomization Tests. *Biometrics*, pages 1387–1393.

Garthwaite, P. H., Fan, Y., and Sisson, S. A. (2016). Adaptive optimal scaling of Metropolis-Hastings algorithms using the Robbins–Monro process. *Communications in Statistics-Theory and Methods*, 45(17):5098–5111.

Gelman, A., Gilks, W. R., and Roberts, G. O. (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms. *The Annals of Applied Probability*, 7(1):110–120.

Golightly, A. and Wilkinson, D. J. (2006). Bayesian Sequential Inference for Stochastic Kinetic Biochemical Network Models. *Journal of Computational Biology*, 13(3):838–851.

Golightly, A. and Wilkinson, D. J. (2008). Bayesian inference for nonlinear multivariate diffusion models observed with error. *Computational Statistics & Data Analysis*, 52(3):1674–1693.

Heston, S. L. (1993). A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency options. *The Review of Financial Studies*, 6(2):327–343.

Holmes, C. and Jasra, A. (2009). Antithetic Methods for Gibbs Samplers. *Journal of Computational and Graphical Statistics*, 18(2):401–414.

Johannes, M. and Polson, N. (2010). CHAPTER 13 - MCMC Methods for Continuous-Time Financial Econometrics. In *Handbook of financial econometrics: Applications*, pages 1–72. Elsevier.

Kloeden, P. E., Platen, E., Kloeden, P. E., and Platen, E. (1992). *Stochastic Differential Equations*. Springer.

Matsuda, K. (2004). Introduction to Merton Jump Diffusion Model. *Department of Economics, The Graduate Center, The City University of New York, New York*.

McAdams, H. H. and Arkin, A. (1999). It's a noisy business! Genetic regulation at the nanomolar scale. *Trends in Genetics*, 15(2):65–69.

Merton, R. (1990). *Continuous-Time Finance*. Blackwell.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics*, 21(6):1087–1092.

Oksendal, B. (2013). *Stochastic Differential Equations: An Introduction with Applications*. Springer Science & Business Media.

Pearson, N. D. and Sun, T.-S. (1989). A Test of the Cox, Ingersoll, and Ross Model of the Term Structure. Technical report, Working paper, MIT.

Pedersen, A. R. (1995). A New Approach to Maximum Likelihood Estimation for Stochastic Differential Equations Based on Discrete Observations. *Scandinavian Journal of Statistics*, pages 55–71.

Pieschner, S. and Fuchs, C. (2020). Bayesian inference for diffusion processes: using higher-order approximations for transition densities. *Royal Society Open Science*, 7(10):200270.

Ritter, C. and Tanner, M. A. (1992). Facilitating the Gibbs Sampler: The Gibbs Stopper and the Griddy-Gibbs Sampler. *Journal of the American Statistical Association*, 87(419):861–868.

Robbins, H. and Monro, S. (1951). A Stochastic Approximation Method. *The Annals of Mathematical Statistics*, pages 400–407.

Robert, C. P., Casella, G., and Casella, G. (1999). *Monte Carlo Statistical Methods*, volume 2. Springer.

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

Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5(2):177–188.

Weickert, J. et al. (1998). *Anisotropic Diffusion in Image Processing*, volume 1. Teubner Stuttgart.

Zhang, Y., Wang, S., Chen, B., Cao, J., and Huang, Z. (2019). Trafficgan: Network-Scale Deep Traffic Prediction With Generative Adversarial Nets. *IEEE Transactions on Intelligent Transportation Systems*, 22(1):219–230.

A Derivations for the Geometric Brownian Motion

A.1 Full conditional distribution of μ

$$\begin{aligned} p\left(\mu|Y,\sigma^2\right) &\propto p\left(Y|\mu,\sigma^2\right)p(\mu) \\ &\propto \exp\left\{-\frac{1}{2\sigma^2}\left(Y-\mu\right)'\left(Y-\mu\right)\right\} \times \exp\left\{-\frac{1}{2\sigma_0^2}\left(\mu-\mu_0\right)^2\right\} \\ &\propto \mathcal{N}\left(\widehat{\mu},K_\mu\right) \end{aligned} \tag{A.1}$$

with 1_T is an $T \times 1$ vectors of one and:

$$K_\mu = \frac{1}{\frac{1}{\sigma^2}+\frac{T}{\sigma_0^2}}, \quad \widehat{\mu} = K_\mu\left(\frac{1_T' Y}{\sigma^2}+\frac{\mu_0}{\sigma_0^2}\right). \tag{A.2}$$

A.2 Full conditional distribution of σ^2

$$\begin{aligned} p\left(\sigma^2|Y,\mu\right) &\propto p\left(Y|\mu,\sigma^2\right)p(\sigma^2) \\ &\propto \left(\sigma^2\right)^{\left(\frac{T}{2}+\nu_0+1\right)} \times \exp\left\{-\frac{1}{\sigma^2}\left(S_0+\frac{1}{2}\left(Y-\mu\right)'\left(Y-\mu\right)\right)\right\} \\ &\propto \mathcal{IG}\left(\nu_0+\frac{T}{2}, S_0+\frac{1}{2}\left(Y-\mu\right)'\left(Y-\mu\right)\right). \end{aligned} \tag{A.3}$$

B Results for Geometric Brownian Motion

B.1 Posterior trace plot

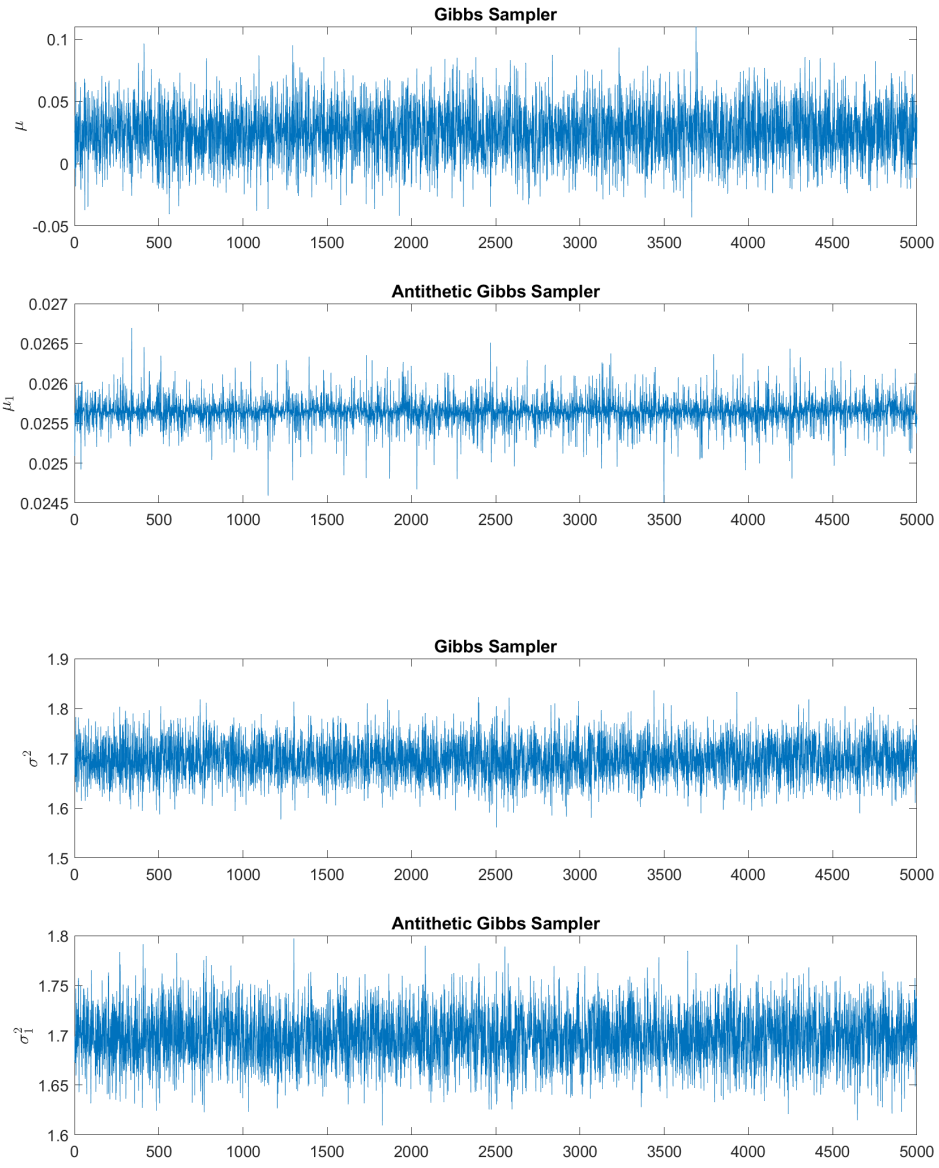

Figure B.1: MCMC posterior trace plots for μ, μ_1, σ^2 and σ_1^2 (from top to bottom).

B.2 Progressive mean plot

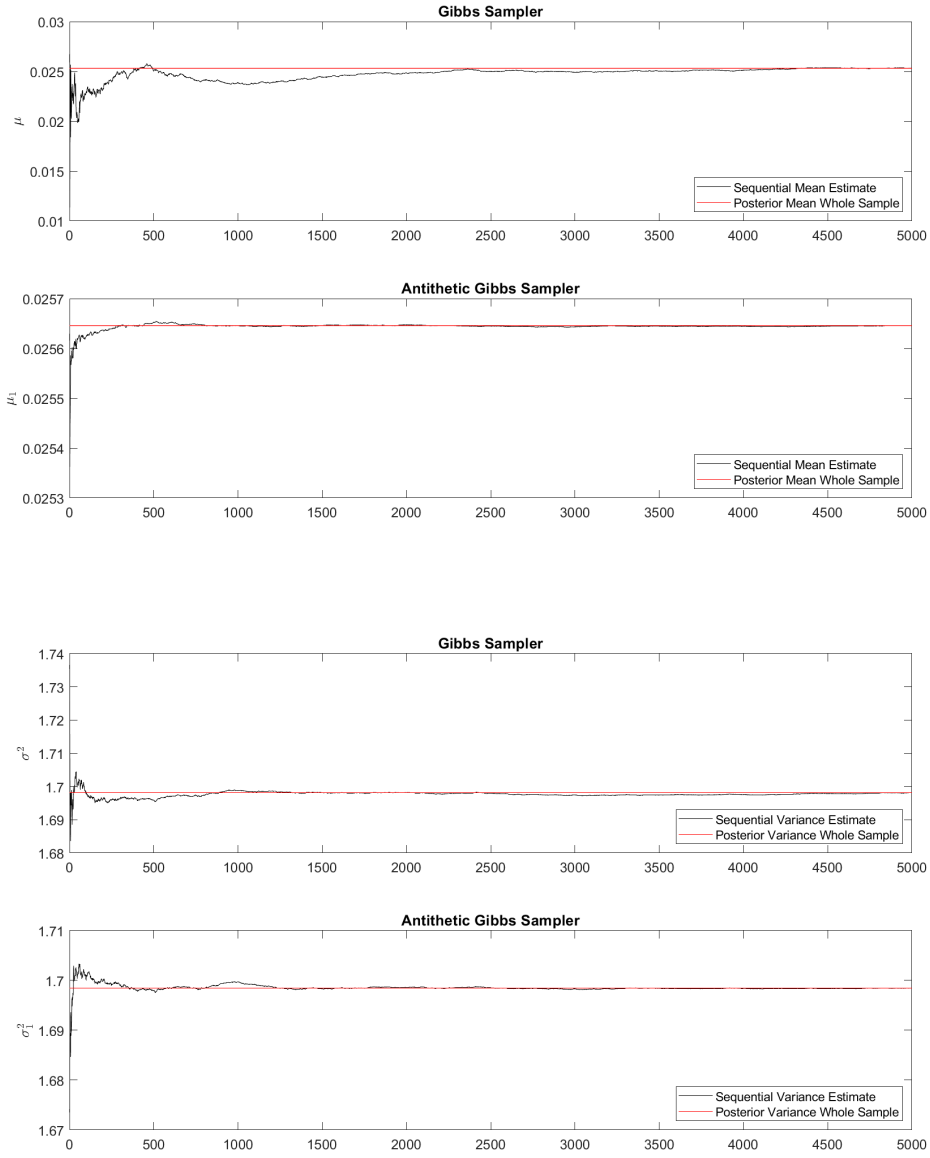

Figure B.2: MCMC progressive averages increasing the number of simulations for μ, μ_1, σ^2 and σ_1^2 (from top to bottom).

C Derivation for the Multivariate Merton's model

C.1 Full conditional distribution of μ

$$\begin{aligned} p(\mu|Z,J,Y) &\propto p(Y|\theta,Z,J)\,p(\mu) \\ &\propto |\Sigma|^{-\frac{Tm}{2}} \times \exp\left\{-\frac{1}{2}(Y-\mu-ZJ)'\left(I_T\otimes\Sigma^{-1}\right)(Y-\mu-ZJ)\right\} \\ &\quad \times \exp\left\{-\frac{1}{2}(\mu-\mu_0)'V_0^{-1}(\mu-\mu_0)\right\} \\ &\propto \exp\left\{-\frac{1}{2}\left[\mu' K_\mu^{-1}\mu - 2\mu'\widehat{\mu}\right]\right\} \\ &\propto \mathcal{N}_m\left(\widehat{\mu}, K_\mu^{-1}\right) \end{aligned} \tag{C.1}$$

with:

$$\begin{aligned} K_\mu &= \Sigma_0^{-1} + 1_{Tm}'\left(I_T\otimes\Sigma^{-1}\right)1_{Tm} \\ \widehat{\mu} &= K_\mu^{-1}\left(V_0^{-1}\mu_0 + 1_{Tm}'\left(I_T\otimes\Sigma^{-1}\right)(Y-ZJ)\right). \end{aligned} \tag{C.2}$$

C.2 Full conditional distribution of V

$$\begin{aligned} p(V|Z,J,Y) &\propto p(Y|\Theta,Z,J)\,p(V) \\ &\propto |\Sigma|^{-\frac{\nu_0+N+T+1}{2}} \times \exp\left\{-\frac{1}{2}tr\left(S_0 + (Y-\mu-ZJ)'(Y-\mu-ZJ)\,\Sigma^{-1}\right)\right\} \\ &\propto \mathcal{IW}_m\left(\nu_0+T, S_0+\sum_{t=1}^{T}(Y_t-\mu-J_tZ_t)'(Y_t-\mu-J_tZ_t)\right). \end{aligned} \tag{C.3}$$

C.3 Full conditional distribution of μ_z

$$\begin{aligned} p\left(\mu_z|Z,J,\right) &\propto p\left(Z|\theta,J\right)p\left(\mu_z\right) \\ &\propto \exp\left\{-\frac{1}{2}\left(Z-\mu_z\right)'\left(I_T\otimes V_z^{-1}\right)\left(Z-\mu_z\right)\right\}\times\exp\left\{-\frac{1}{2}\left(\mu_z-\mu_{z_0}\right)'V_{z0}^{-1}\left(\mu_z-\mu_{z_0}\right)\right\} \\ &\propto \exp\left\{-\frac{1}{2}\left[\mu_z'K_{\mu_z}^{-1}\mu_z-2\mu_z'\widehat{\mu_z}\right]\right\} \\ &\propto \mathcal{N}_m\left(\widehat{\mu_z},K_{\mu_z}^{-1}\right) \end{aligned} \tag{C.4}$$

with:

$$\begin{aligned} K_{\mu_z} &= V_{z0}^{-1}+1_{Tm}'\left(I_T\otimes V_z^{-1}\right)1_{Tm} \\ \widehat{\mu_z} &= K_{\mu_z}^{-1}\left(V_{z0}^{-1}\mu_{z0}+1_{NT}'\left(I_T\otimes V_z^{-1}\right)Z\right). \end{aligned} \tag{C.5}$$

C.4 Full conditional distribution of V_z

$$\begin{aligned} p\left(V_z|Z,J\right) &\propto p\left(Z|\theta,J\right)p\left(V_z\right) \\ &\propto \left|V_z\right|^{-\frac{\nu_{z0}+N+T+1}{2}}\times\exp\left\{-\frac{1}{2}tr\left(S_{z0}+\left(Z-\mu_z\right)'\left(Z-\mu_z\right)V_z^{-1}\right)\right\} \\ &\propto \mathcal{IW}_m\left(\nu_{z_0}+T,S_0+\left(Z-\mu_z\right)'\left(Z-\mu_z\right)\right). \end{aligned} \tag{C.6}$$

C.5 Full conditional distribution of λ

$$\begin{aligned} p\left(\lambda|J\right) &\propto p\left(J|\lambda\right)p\left(\lambda\right) \\ &\propto \left[\lambda^{\sum_{t=1}^{T}J_t}\left(1-\lambda\right)^{T-\sum_{t=1}^{T}J_t}\right]\lambda^{\alpha-1}\left(1-\lambda\right)^{\beta-1} \\ &\propto \mathcal{B}\left(\alpha+\sum_{t=1}^{T}J_t,\beta+T-\sum_{t=1}^{T}J_t\right). \end{aligned} \tag{C.7}$$

C.6 Full conditional distribution of Z_t

$$\begin{aligned} p\left(Z_t|Y_t, J_t, \theta\right) &\propto \exp\left\{-\frac{1}{2}\left(h_t'\Sigma^{-1}h_t + \left(Z_t-\mu_z\right)' V_z^{-1}\left(Z_t-\mu_z\right)\right)\right\} \\ &\propto \exp\left\{-\frac{1}{2}\left(Z_t-k_t\right)' V_t^{-1}\left(Z_t-k_t\right)\right\} \\ &\propto \mathcal{N}\left(k_t, V_t\right) \end{aligned} \tag{C.8}$$

with:

$$\begin{aligned} h_t &= Y_t - \mu - Z_t J_t \\ V_t &= \left(J_t\Sigma^{-1} + V_z^{-1}\right)^{-1} \\ k_t &= V_z^{-1}\left(J_t\Sigma^{-1}\left(Y_t-\mu\right) + V_z^{-1}\mu_z\right). \end{aligned} \tag{C.9}$$

C.7 Full conditional distribution of J_t

The conditional posterior distribution for jump times is Bernoulli due to the binary nature of J_t, which can only take values of 0 or 1.

$$\begin{aligned} p\left(J_t|\theta, Z_t, Y_t\right) &\propto p\left(Y_t|J_t, Z_t, \theta\right) p\left(J_t|\Theta\right) \\ &\propto \mathcal{B}in\left(1, \varphi_t\right) \end{aligned} \tag{C.10}$$

with:

$$\varphi_t = \frac{\lambda \exp\left\{-\frac{1}{2}\left(Y_t-\mu-Z_t\right)'\Sigma^{-1}\left(Y_t-\mu-Z_t\right)\right\}}{\lambda \exp\left\{-\frac{1}{2}\left(Y_t-\mu-Z_t\right)'\Sigma^{-1}\left(Y_t-\mu-Z_t\right)\right\} + (1-\lambda)\exp\left\{-\frac{1}{2}\left(Y_t-\mu\right)'\Sigma^{-1}\left(Y_t-\mu\right)\right\}}. \tag{C.11}$$

.

D Results for the Multivariate Merton's model

D.1 Posterior trace plot

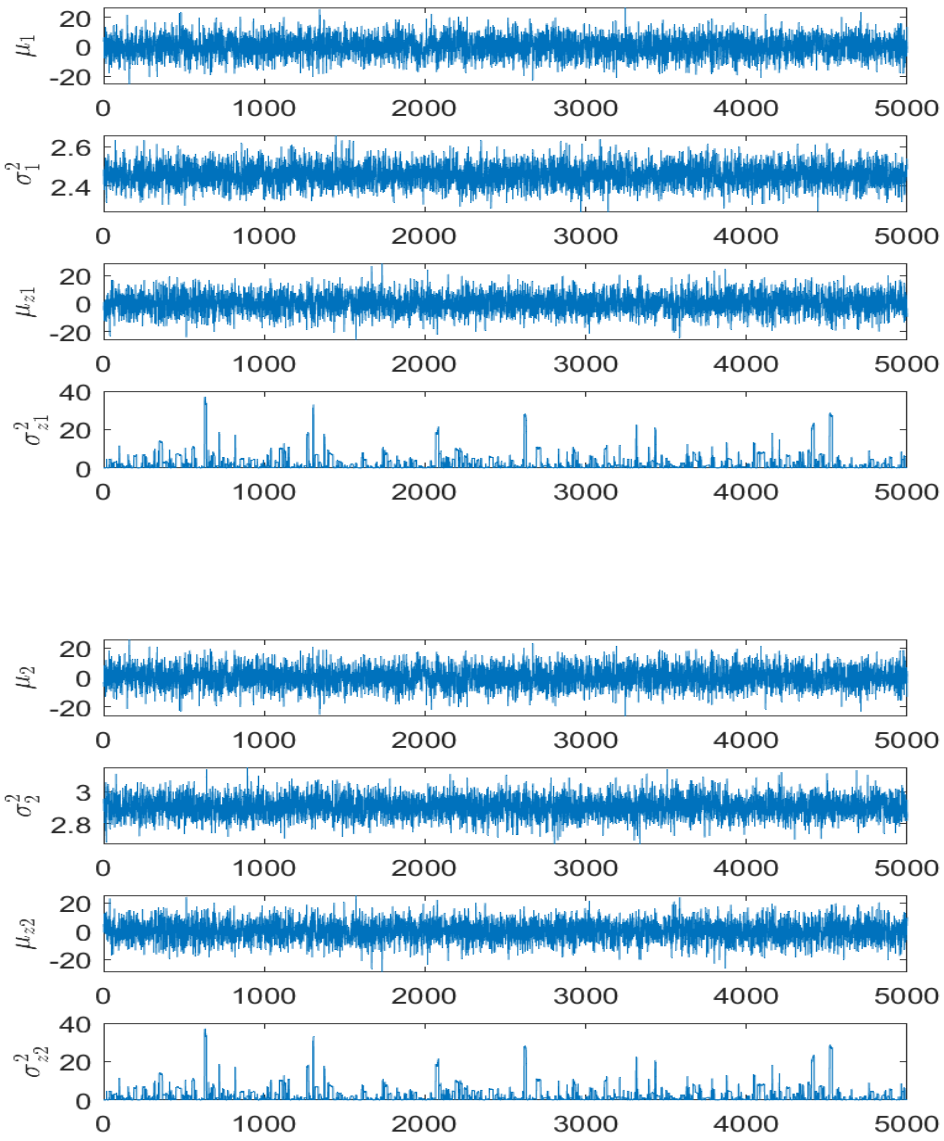

Figure D.1: MCMC posterior trace plots for S&P 500 and NASDAQ 100 parameters (from top to bottom).

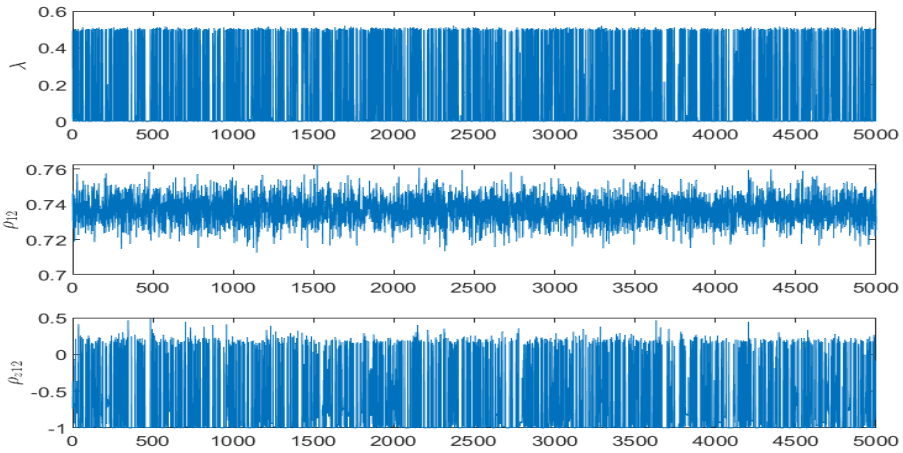

Figure D.2: MCMC posterior trace plots for other parameters (jump intensity, return correlation, and jump size correlation).

D.2 Progressive mean plot

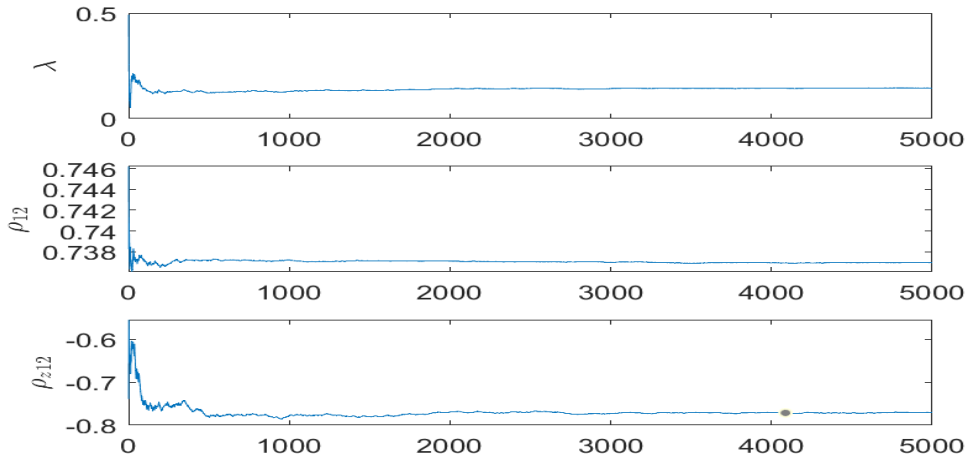

Figure D.3: MCMC progressive averages increasing the number of simulations for other parameters (jump intensity, return correlation, and jump size correlation).

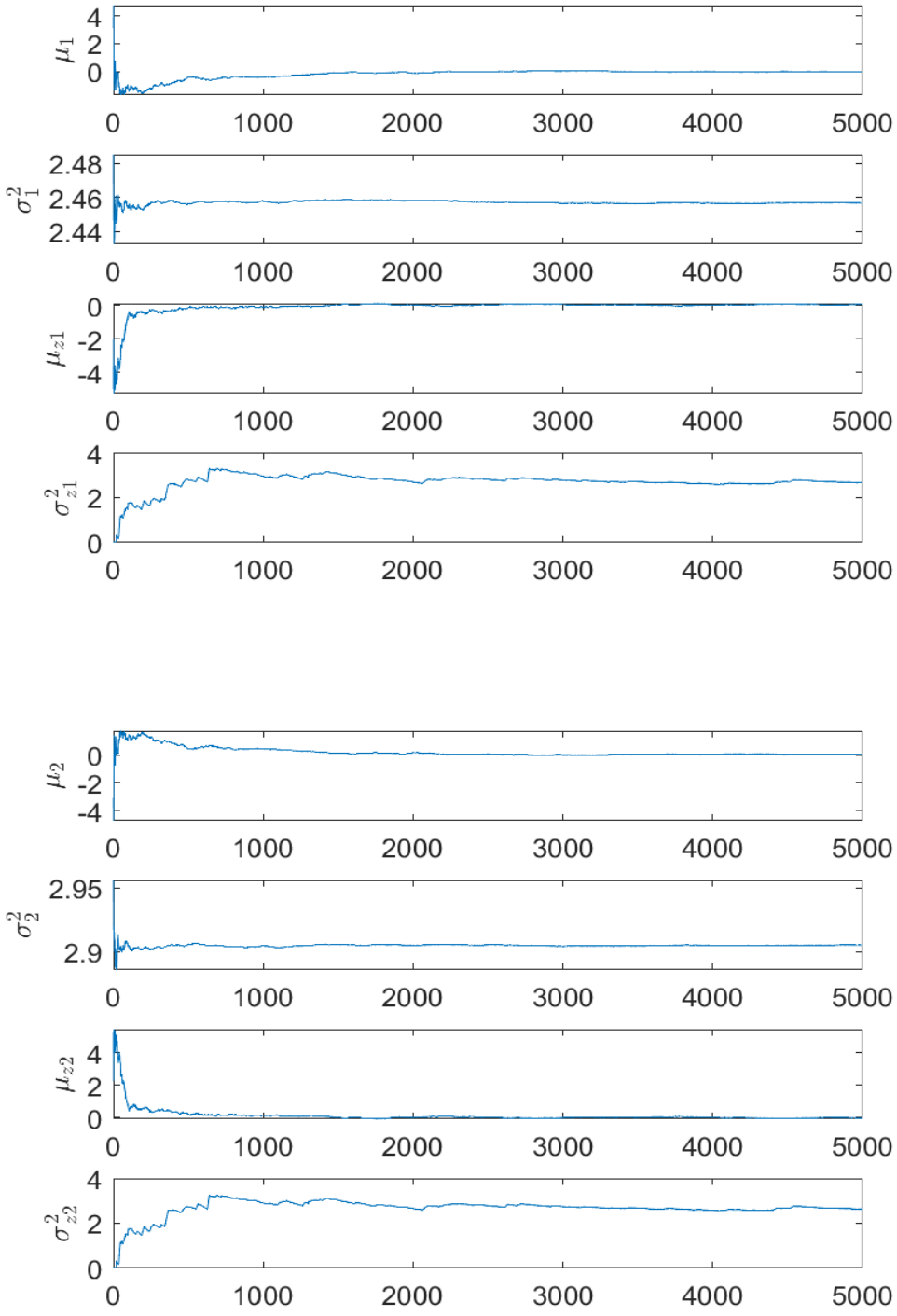

Figure D.4: MCMC progressive averages increasing the number of simulations for S&P 500 parameters and NASDAQ 100 parameters (from top to bottom).

E Derivation for the Vasicek model

E.1 Full conditional distribution of r

Rewrite the state equation in (4.2.1). We have:

$$\begin{aligned} r_{t+1} &= a_1 + (1 - b_1)\, r_t + \sigma_r \eta_{t+1} \\ &= a_1 + n_1 r_t + \sigma_r \eta_{t+1} \end{aligned} \tag{E.1}$$

$$\begin{aligned} \Rightarrow r_{t+1} - n_1 r_t &= a_1 + \sigma_r \eta_{t+1} \\ \Rightarrow Hr &= a_1 1_T + n_1 \tilde{r_0} + \sigma_r \eta \end{aligned}$$

where:

$$n_1 = 1 - b_1, \quad H = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ -n_1 & 1 & 0 & \dots & 0 \\ 0 & -n_1 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & -n_1 & 1 \end{bmatrix} \text{ is a } T \times T \text{ matrix}$$

$\tilde{r_0} = (r_0, 0, \dots, 0)', \eta = (\eta_1, \dots, \eta_T)'$ are $T \times 1$ vectors.

H is a band matrix with $|H| = 1 \neq 0$. Therefore, H is invertible. We claim that: $H^{-1}\tilde{r_0} n_1 = r_0 \left(n_1, n_1^2, \dots, n_1^T\right)' = A r_0$, with $A = \left(n_1, n_1^2, \dots, n_1^T\right)'$ Combine with (1.10), and we have:

$$r = a_1 H^{-1} 1_T + A r_0 + \sigma_r H^{-1} \eta. \tag{E.2}$$

thus:

$$p\left(r | a_1, b_1, \sigma_r, r_0\right) = \left(2\pi\sigma_r^2\right)^{-\frac{T}{2}} \exp\left\{ -\frac{1}{2\sigma_r^2} \left(r - a_1 H^{-1} 1_T - A r_0\right)' H' H \left(r - a_1 H^{-1} 1_T - A r_0\right) \right\}.$$

$$
\begin{aligned}
p\left(r|Y,\Theta\right) &\propto p\left(Y|\tilde{\alpha},\tilde{\beta},\Sigma\right)p\left(r|a_1,b_1,\sigma_r,r_0\right)\\
&\propto \exp\left\{-\frac{1}{2}\left(Y-1_T\otimes\tilde{\alpha}-\left(I_T\otimes\tilde{\beta}\right)r\right)'\left(I_T\otimes\Sigma^{-1}\right)\left(Y-1_T\otimes\tilde{\alpha}-\left(I_T\otimes\tilde{\beta}\right)r\right)\right\}\\
&\quad\times\exp\left\{-\frac{1}{2\sigma_r^2}\left(r-a_1H^{-1}1_T-Ar_0\right)'H'H\left(r-a_1H^{-1}1_T-Ar_0\right)\right\}\\
&\propto \exp\left\{-\frac{1}{2}\left(r'w_1r-2r'w_2\right)\right\}\\
&\propto \mathcal{N}_T\left(\widehat{r},V_r^{-1}\right)
\end{aligned}
\tag{E.3}
$$

with:

$$
\begin{aligned}
w_1 &= I_T\otimes\tilde{\beta}'\Sigma^{-1}\tilde{\beta}+\frac{H'H}{\sigma_r^2},\\
w_2 &= \left(I_T\otimes\tilde{\beta}'\Sigma^{-1}\right)Y-\left(1_T\otimes\tilde{\beta}'\Sigma^{-1}\tilde{\alpha}\right)+\frac{H'H}{\sigma_r^2}\left(a_1H^{-1}1_T+Ar_0\right)\\
V_r &= w_1,\quad \widehat{r}=V_r^{-1}w_2=\frac{w_2}{w_1}.
\end{aligned}
\tag{E.4}
$$

E.2 Full conditional distribution of $z=(a_1,b_1)$

Rewrite the state equation in (4.2.1), we have the following:

$$
\Delta r=\tilde{r}z+\sigma_r\eta_{t+1} \tag{E.5}
$$

with: $\Delta r=[r_1-r_0,...,r_T-r_{T-1}]'$ and $\tilde{r}=\begin{bmatrix}1 & -r_1\\ \vdots & \vdots\\ 1 & -r_T\end{bmatrix}$.

Since $\eta_t\sim\mathcal{N}(0,1)$, from (1.5), thus:

$$
\begin{aligned}
p\left(\Delta r|z,\sigma_r\right) &= \left|2\pi\sigma_r^2I_T\right|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2\sigma_r^2}\left(\Delta r-\tilde{r}z\right)'\left(\Delta r-\tilde{r}z\right)\right\}\\
&= \left(2\pi\sigma_r^2\right)^{-\frac{T}{2}}\exp\left\{-\frac{1}{2\sigma_r^2}\left(\Delta r-\tilde{r}z\right)'\left(\Delta r-\tilde{r}z\right)\right\}.
\end{aligned}
\tag{E.6}
$$

Combine with (4.2.2), and we have:

$$\begin{aligned} p\left(z|\Delta r,\sigma_r\right) &\propto p\left(\Delta r|z,\sigma_r\right)p\left(z\right) \\ &\propto \exp\left\{-\frac{1}{2\sigma_r^2}\left(\Delta r-\tilde{r}z\right)'\left(\Delta r-\tilde{r}z\right)\right\}\times\exp\left\{-\frac{1}{2}\left(z-z_0\right)'V_0^{-1}\left(z-z_0\right)\right\} \\ &\propto \exp\left\{-\frac{1}{2}\left[z'K_z^{-1}z-2z'\widehat{z}\right]\right\} \\ &\propto \mathcal{N}_2\left(\widehat{z},K_z^{-1}\right) \end{aligned} \tag{E.7}$$

with:

$$K_z=\left(V_0^{-1}+\frac{\tilde{r}'\tilde{r}}{\sigma_r^2}\right),\quad \widehat{z}=K_z^{-1}\left(V_0^{-1}z_0+\frac{\tilde{r}'\Delta r}{\sigma_r^2}\right). \tag{E.8}$$

E.3 Full conditional distribution of Σ

Rewriting the observation equation in (4.2.1), we have: $Y=1_T\otimes\tilde{\alpha}+\left(I_T\otimes\tilde{\beta}\right)r+\epsilon$, with $\epsilon\sim\mathcal{N}\left(0,I_T\otimes\Sigma\right)$ and 1_T is a $T\times 1$ vectors of one, thus:

$$Y|\tilde{\alpha},\tilde{\beta},\Sigma\sim\mathcal{N}_{Tm}\left(1_T\otimes\tilde{\alpha}+\left(I_T\otimes\tilde{\beta}\right)r,I_T\otimes\Sigma\right) \tag{E.9}$$

with probability density function:

$$\begin{aligned} &p\left(Y|\tilde{\alpha},\tilde{\beta},\Sigma\right) \\ &=(2\pi)^{-\frac{mT}{2}}\left|\Sigma\right|^{-\frac{T}{2}} \\ &\quad\times\exp\left\{-\frac{1}{2}\left(Y-1_T\otimes\tilde{\alpha}-\left(I_T\otimes\tilde{\beta}\right)r\right)'\left(I_T\otimes\Sigma^{-1}\right)\left(Y-1_T\otimes\tilde{\alpha}-\left(I_T\otimes\tilde{\beta}\right)r\right)\right\} \\ &=(2\pi)^{-\frac{mT}{2}}\left|\Sigma\right|^{-\frac{T}{2}} \\ &\quad\times\exp\left\{-\frac{1}{2}\sum_{t=1}^{T}\left(Y_t-\tilde{\alpha}-\tilde{\beta}r_t\right)'\Sigma^{-1}\left(Y_t-\tilde{\alpha}-\tilde{\beta}r_t\right)\right\}. \end{aligned} \tag{E.10}$$

Following the prior assumptions in equation (4.2.2), we have an Inverse Wishart conditional distribution:

$$\begin{aligned} p\left(\Sigma|Y,\tilde{\alpha},\tilde{\beta}\right) &\propto p\left(Y|\tilde{\alpha},\tilde{\beta},\Sigma\right)p\left(\Sigma\right) \\ &\propto |\Sigma|^{-\frac{\nu_0+T+m+1}{2}} \times \exp\left\{-\frac{1}{2}tr\left[\left(S_0+\sum_{t=1}^{T}\left(Y_t-\tilde{\alpha}-\tilde{\beta}r_t\right)'\left(Y_t-\tilde{\alpha}-\tilde{\beta}r_t\right)\right)\Sigma^{-1}\right]\right\} \\ &\propto \mathcal{IW}_m\left(\nu_0+T,\sum_{t=1}^{T}\left(Y_t-\tilde{\alpha}-\tilde{\beta}r_t\right)'\left(Y_t-\tilde{\alpha}-\tilde{\beta}r_t\right)\right). \end{aligned} \tag{E.11}$$

E.4 Full conditional distribution of a_2

Rewrite the second equation in (4.1.5), but in the vector form, we have:

$$\begin{aligned} \tilde{\alpha}\left(a_2,b_2,\sigma_r,\tau\right) &= \frac{1}{2}\left[\frac{\sigma_r^2}{b_2^2}-\frac{a_2}{b_2}\right]\left(\tau-\tilde{\beta}\left(b_2,\sigma_r,\tau\right)\right)-\frac{\sigma_r^2}{4b_2}\tilde{\beta}\left(b_2,\sigma_r,\tau\right)^2 \\ &= -\frac{1}{2}a_2\left(\frac{\tau-\tilde{\beta}}{b_2}\right)+\frac{\sigma_r^2}{4b_2}\left[2\left(\frac{\tau-\tilde{\beta}}{b_2}\right)-\tilde{\beta}^2\right] \\ &= -\frac{1}{2}a_2g_1+\frac{\sigma_r^2}{4b_2}\left(2g_1-\tilde{\beta}^2\right) \\ &= -\frac{1}{2}a_2g_1+h_1. \end{aligned} \tag{E.12}$$

with $g_1=\left(\frac{\tau-\tilde{\beta}}{b_2}\right)$ and $h_1=\frac{\sigma_r^2}{4b_2}\left(2g_1-\tilde{\beta}^2\right)$ are a $m\times 1$ vectors.

$$\begin{aligned} p\left(a_2|Y,b_2,\Sigma\right) &\propto p\left(Y|a_2,b_2,\Sigma\right)p\left(a_2\right) \\ &\propto \exp\left\{-\frac{1}{2}\left(Y-1_T\otimes\tilde{\alpha}-\left(I_T\otimes\tilde{\beta}\right)r\right)'\left(I_T\otimes\Sigma^{-1}\right)\left(Y-1_T\otimes\tilde{\alpha}-\left(I_T\otimes\tilde{\beta}\right)r\right)\right\} \\ &\quad\times\exp\left\{-\frac{1}{2\sigma_{2a}^2}\left(a_2-m_{2a}\right)^2\right\} \\ &\propto \exp\left\{-\frac{1}{2}\left[a_2^2D_{2a}^{-1}-2a_2\widehat{a_2}\right]\right\} \\ &\propto \mathcal{N}\left(\widehat{a_2},D_{2a}\right) \end{aligned} \tag{E.13}$$

with:

$$
\begin{aligned}
u_1 &= \frac{1}{4}\left[T g_1^{'}\Sigma^{-1}h_1 + T h_1^{'}\Sigma^{-1}g_1 + \left(1_T^{'}\otimes g_1^{'}\Sigma^{-1}\tilde{\beta}\right)r - \left(1_T^{'}\otimes g_1^{'}\Sigma^{-1}\ \right)Y\right] \\
u_2 &= \frac{1}{4}\left[r'\left(1_T\otimes\tilde{\beta}'\Sigma^{-1}g_1\right) - Y'\left(1_T\otimes\Sigma^{-1}g_1\right)\right] \\
D_{2a} &= \left(0.25T g_1^{'}\Sigma^{-1}g_1 + \frac{1}{\sigma_{2a}^2}\right)^{-1}, \quad \hat{a_2} = D_{2a}\left(u_1 + u_2 + \frac{m_{2a}}{\sigma_{2a}^2}\right).
\end{aligned}
\tag{E.14}
$$

E.5 Full conditional distribution of b_2

$$
\begin{aligned}
&p\left(b_2|Y,a_2,\Sigma\right) \propto p\left(Y|a_2,b_2,\Sigma\right)p\left(b_2\right) \\
&\propto \exp\left\{-\frac{1}{2}\left(Y - 1_T\otimes\tilde{\alpha} - \left(I_T\otimes\tilde{\beta}\right)r\right)^{'}\left(I_T\otimes\Sigma^{-1}\right)\left(Y - 1_T\otimes\tilde{\alpha} - \left(I_T\otimes\tilde{\beta}\right)r\right)\right\} \\
&\quad\times\exp\left\{-\frac{1}{2\sigma_{2b}^2}\left(b_2 - m_{2b}\right)^2\right\}.
\end{aligned}
\tag{E.15}
$$

This conditional distribution of b_2 is nonstandard. We use a Metropolis-Hasting algorithm to sample from this nonstandard density within the Gibbs sampler. Because b_2 only appears in the yield equation, generating a reasonable proposal for an independent Metropolis can be difficult. Thus we will use an RWMH step for b_2. In addition, the proposal distribution will be automatically adjusted during the simulation (Adaptive MH).

E.6 Full conditional distribution of σ_r^2

From (4.2.2) and (E.3), we have:

$$\begin{aligned} p\left(\sigma_r^2|Y,r,\Theta\right) &\propto p\left(Y|\tilde{\alpha},\tilde{\beta},\Sigma\right) p\left(r|a_1,b_1,\sigma_r,r_0\right) p\left(\sigma_r^2\right) \\ &\propto \mathcal{N}_T\left(\widehat{r},V_r^{-1}\right) \mathcal{IG}\left(\nu_r,S_r\right). \end{aligned}$$

This conditional density of σ_r^2 is nonstandard (multiplication of the Normal distribution and the Inverse Gamma distribution), so we will use the Metropolis-Hasting algorithm to sample this. Moreover, σ_r appears in both the yield equation and state equation, so we will use Adaptive MH for this to account for the complicated shape of the target distribution.

E.7 Full conditional distribution of r_0

Recall that r_0 only appears in the first state equation: $r_1 = a_1 + n_1 r_0 + \sigma_r \eta_1$, where $\eta_1 \sim \mathcal{N}(0,1)$. Given the normal prior r_0 in (4.2.2), we can again use standard linear regression results to get:

$$\begin{aligned} p\left(r_0|Y,r,\Theta\right) &\propto \exp\left\{-\frac{1}{2}\left[r_0^2\left(\frac{n_1^2}{\sigma_r^2}+\frac{1}{d_0^2}\right) - 2r_0\left(\frac{n_1\left(r_1-a_1\right)}{\sigma_r^2}+\frac{2c_0}{d_0^2}\right)\right]\right\} \\ &\propto \mathcal{N}\left(\left(\frac{n_1\left(r_1-a_1\right)}{\sigma_r^2}+\frac{2c_0}{d_0^2}\right)\left(\frac{n_1^2}{\sigma_r^2}+\frac{1}{d_0^2}\right)^{-1}, \left(\frac{n_1^2}{\sigma_r^2}+\frac{1}{d_0^2}\right)^{-1}\right). \end{aligned}$$

F Results for Vasicek model

F.1 Posterior trace plot

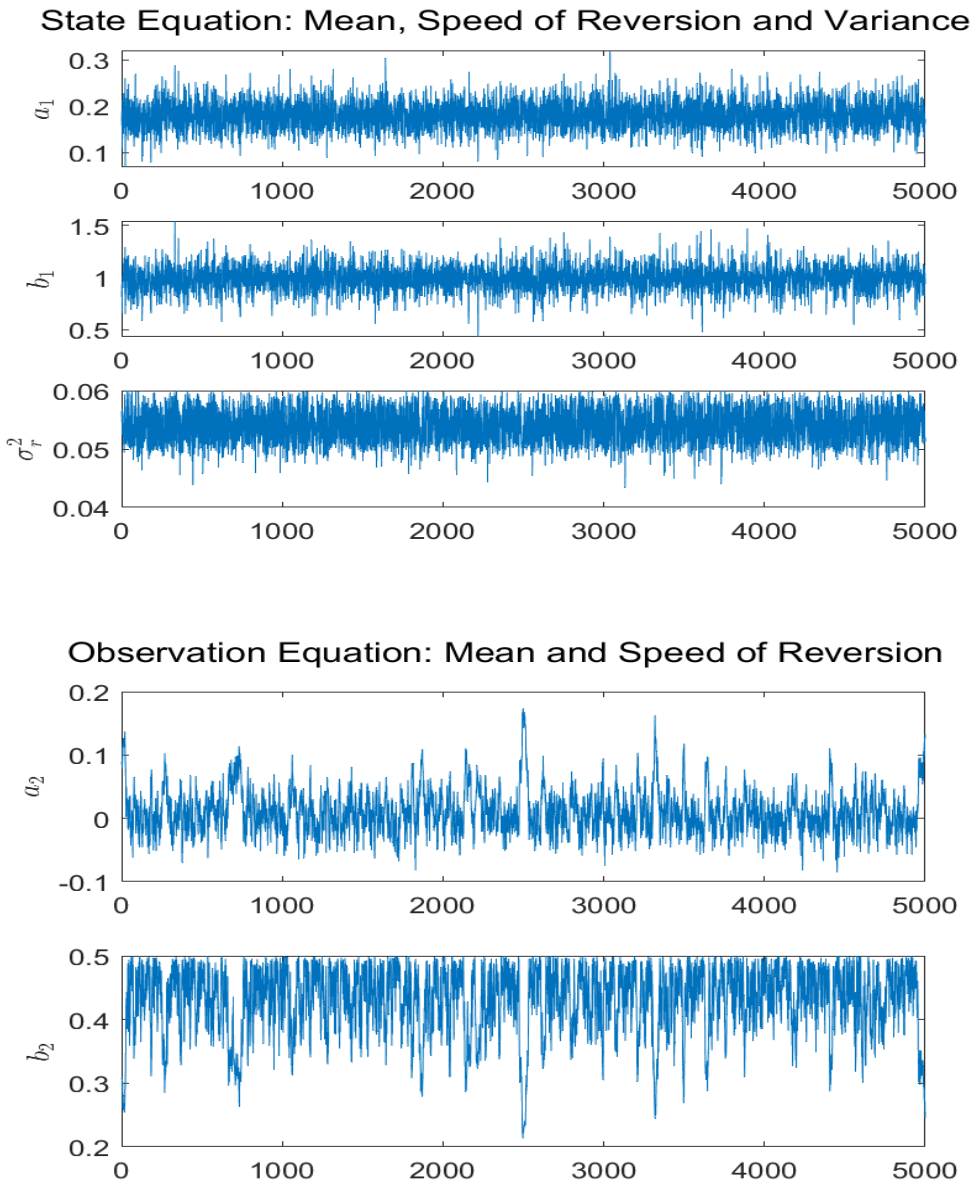

Figure F.1: MCMC posterior trace plots for the parameters of the state and observation equations (from top to bottom).

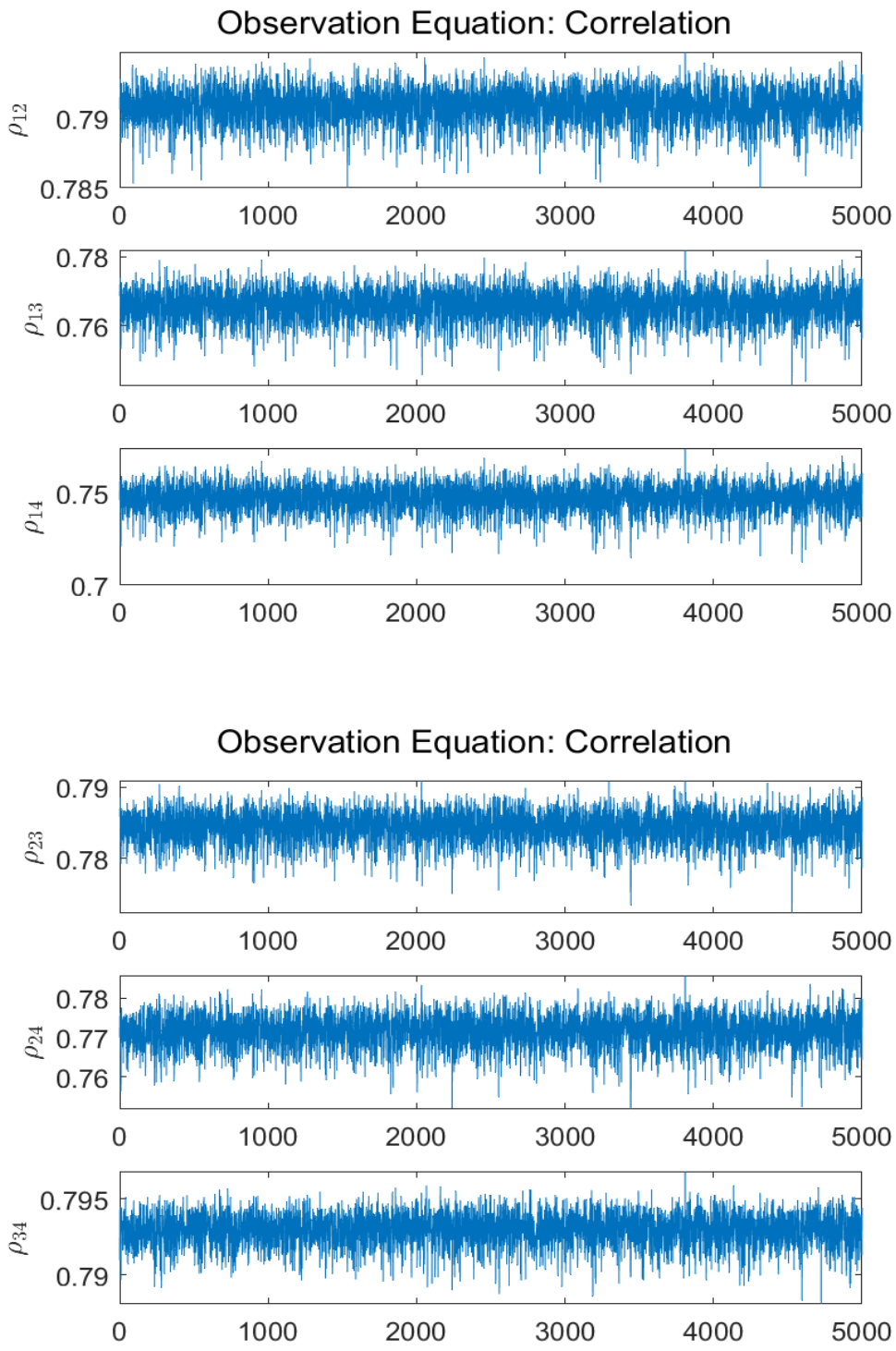

Figure F.2: MCMC posterior trace plots for the correlation parameters of the observation equation.

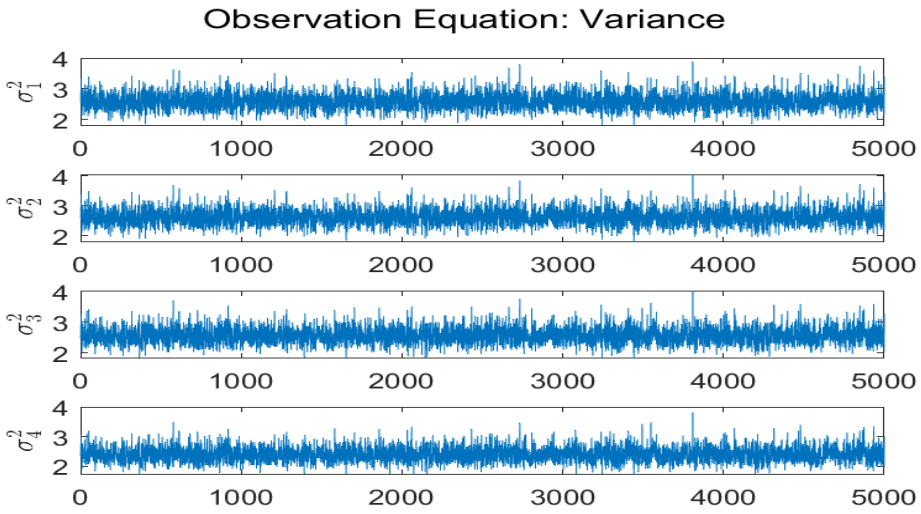

Figure F.3: MCMC posterior trace plots for the variance parameters of the observation equation.

F.2 Progressive mean plot

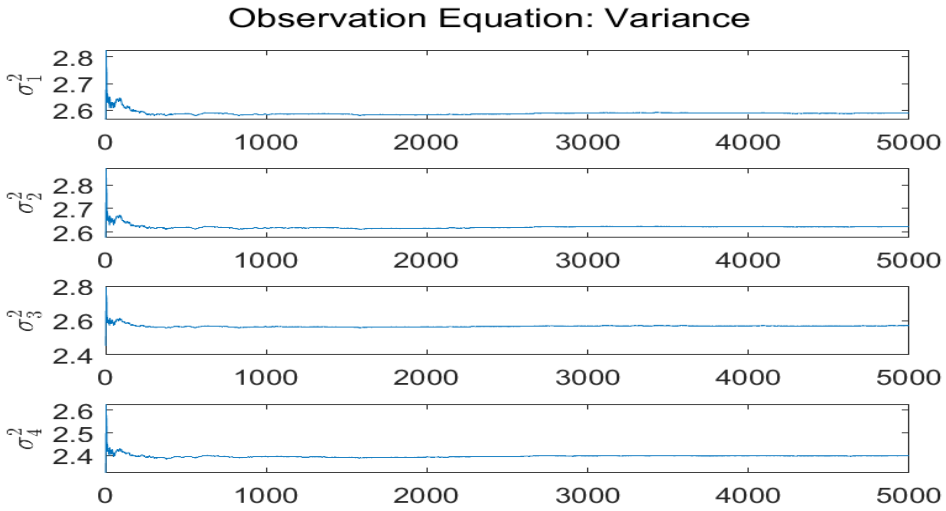

Figure F.4: MCMC progressive averages increasing the number of simulations for the variance parameters of the observation equation.

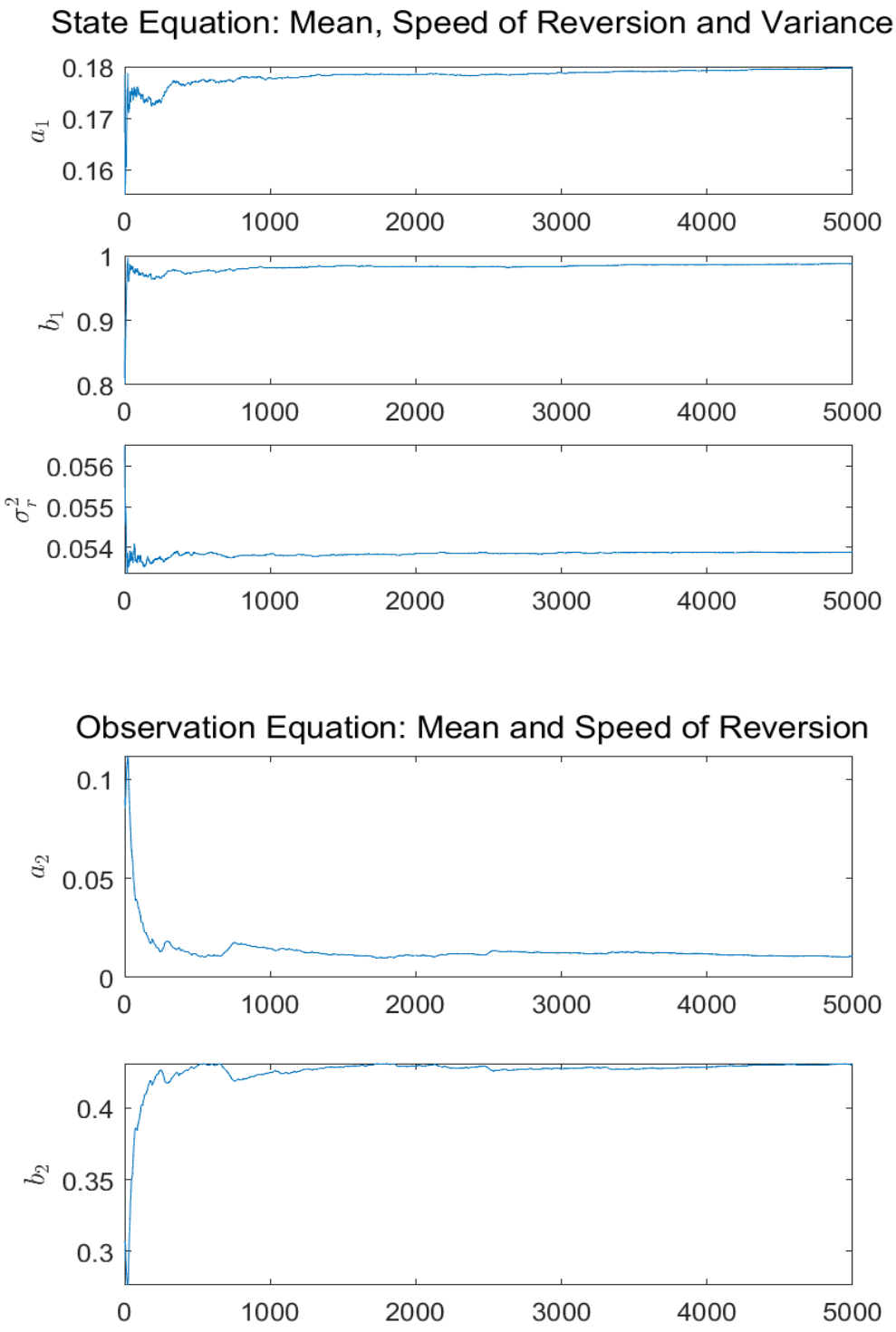

Figure F.5: MCMC progressive averages increasing the number of simulations for the state and the observation equations (from top to bottom).

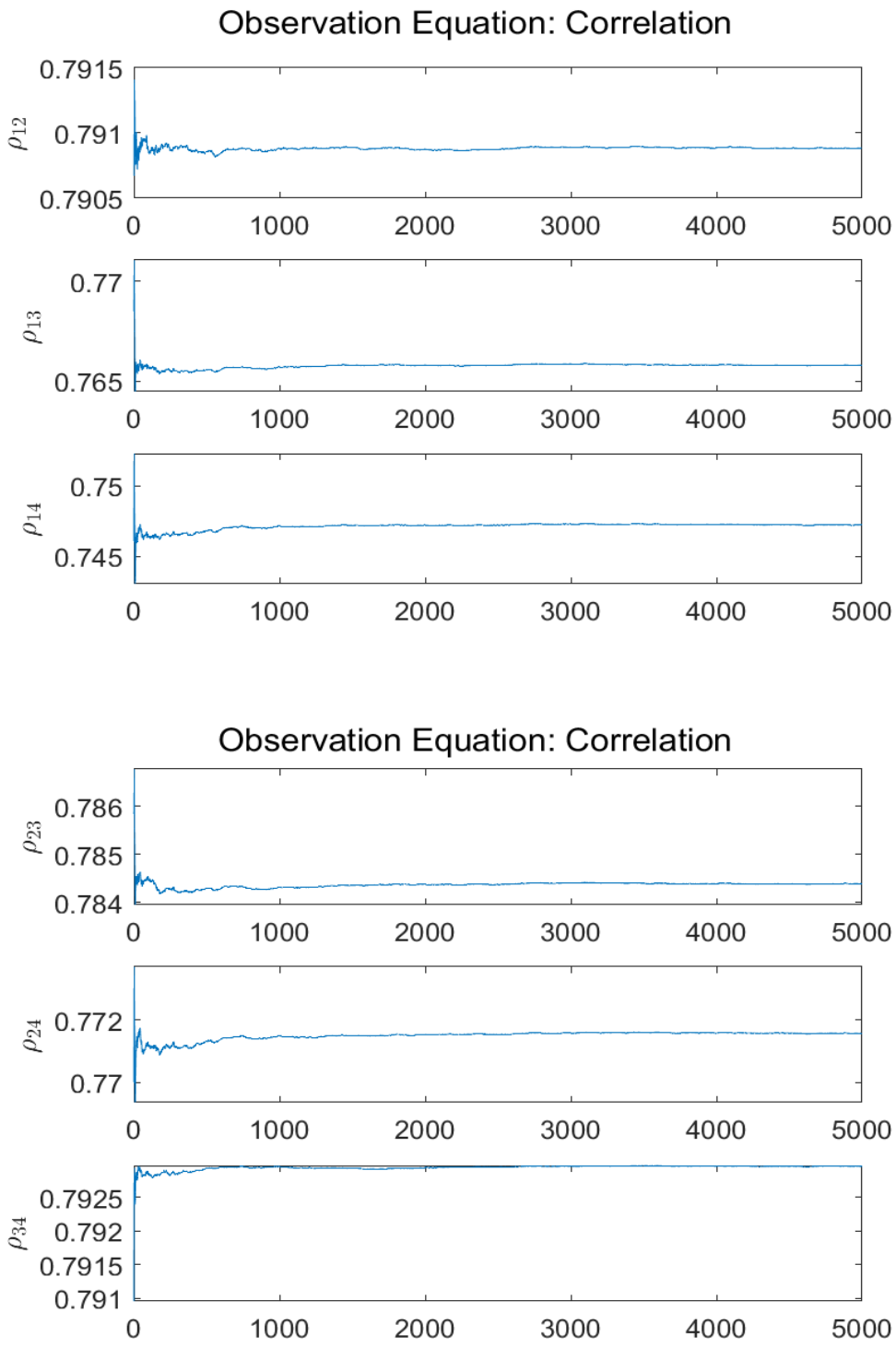

Figure F.6: MCMC progressive averages increasing the number of simulations for the correlation parameters of the observation equation.