2015

Asset pricing based on stochastic delay differential equations

Yun Zheng
Iowa State University

Asset pricing based on stochastic delay differential equations

by

Yun Zheng

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Co-majors: Applied Mathematics

Statistics

Program of Study Committee:

L. Steven Hou, Co-major Professor

Huaiqing Wu, Co-major Professor

Scott Hansen

Daniel J. Nordman

Ananda Weerasinghe

Iowa State University

Ames, Iowa

2015

DEDICATION

I would like to dedicate this dissertation to my wife Yu Qiu, my daughter Grace Zheng, and my parents Hong Zheng and Shixuan Zhu, whose love and support sustained me throughout.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGMENTS

Completion of this work would not have been possible without the guidance, patience, and support of my advisers, Dr. L. Steven Hou and Dr. Huaiqing Wu. They have been helping me to grow as a thinker and to keep things in perspective. At the same time, they have taught me so many valuable things, not only about mathematics and statistics, but also about life. I am very fortunate to have met them.

I would like to thank my program of study committee members, Dr. Scott Hansen, Dr. Daniel Nordman, and Dr. Ananda Weerasinghe, for their valuable advice and help. I am also indebted to Dr. Alicia Carriquiry, Dr. Paul Sacks, and Dr. Wolfgang Kliemann for helping me in my graduate study. I am lucky to have taken several courses from Dr. Stephen Vardeman. His knowledge, view, and philosophy always inspire me.

Last but not least, I have met so many friends in this quiet small town and thanks for their help and support.

ABSTRACT

This dissertation studies stochastic delay differential equations (SDDEs), applies them to real market data, and compares them with classic models. In Chapter 2, we study the mathematical properties of stochastic differential equations with or without delay, and introduce the linear SDDE for several specific financial market behaviors we are interested in. Since it is hard to find an explicit solution of an SDDE, we introduce a numerical technique and use it to analyze the SDDE. In Chapter 3, we use the Euler-Maruyama method to discretize a continues-time stochastic system and show the convergence in different senses of the numerical scheme to the true solution of the linear SDDE. Furthermore, it is crucial to understand the quantitative behavior of the parameters for the stochastic system and the impact of introducing the delay term, but these parameters are unknown and hard to estimate. In Chapter 4, we propose a blocking method to group the price points and use the Bayesian methods to estimate the parameters in the linear SDDE. We then apply the model to real stock price data, estimate and calibrate all the parameters in the stochastic system, and compare them with the parameters obtained from the classic geometric Brownian motion model.

CHAPTER 1. OVERVIEW

1.1 Introduction

In recent years, stochastic differential equations (SDEs) have become an important tool for research on valuation of financial assets. The advances in the theory of continuous-time mathematical finance and the wide use of stochastic diffusion models have solved many sophisticated derivative pricing problems. The geometric Brownian motion (GBM) with constant drift and volatility is used in mathematical finance to model the asset price in the Black-Scholes model [Black and Scholes (1973) and Merton (1973)], which has become the most well-known mathematical model for option pricing.

Empirical studies have shown some weaknesses of using the GBM to approximate the underlying asset price evolution process. One significant flaw is the requirement of constant volatility, which is shown not to be the case in practice. Several approaches have been considered to overcome the assumption of constant volatility for the underlying GBM, including introducing local and stochastic volatility models. Those models have the Markov property and are only influenced by the immediate present data. Another requirement that is shown not to be the case in practical situations is the non-arbitrage market efficiency assumption. The absence of arbitrage opportunities is a fundamental principle for the modern theory of financial asset pricing. Studies of those issues can be found in Kaldor (1940), Stein and Stein (1991), Elsanosi et al. (2000), Brandt and Santa-Clara (2002), Jones (2003), Hu, Mohammed, and Yan (2004), Anh and Inoue (2005), and Stoica (2005).

The motivation for considering a stochastic delay differential equation is from the forward and backward looking of the financial market. In the financial market, the insider and elite trader use past information, such as the historical stock prices, to predict the market movement and make investment decisions. This feedback behavior is absent from the classic SDE model. By introducing delay terms to the standard GBM model, we take historical information into consideration and incorporate the feedback effect in the model. The SDDE models provide a more realistic mathematical formulation for the evolution of asset prices in an inefficient financial market than the GBM based models.

1.2 Stochastic Differential Equations

1.2.1 Introduction

This section briefly reviews the theory of SDEs. For proofs of the theorems and details of the definitions, the reader is referred to Karatzas and Shreve (1991), Evans (2013), and Øksendal (2013). The presentation of Itô integrals and Stratonovich integrals are mostly based on those references mentioned above. We will start with the definitions.

Definition 1. A **stochastic process** $\{X(t)\}_{t\geq 0}$ is a collection of random variables on a given probability space $(\Omega, \mathcal{F}, P)$ indexed by time t.

- For each $t \geq 0$, the function $w \to X(t; w)$ is a measurable function defined on the probability space $(\Omega, \mathcal{F}, P)$.

- For each $w \in \Omega$, the function $t \to X(t; w)$ is called a sample path or trajectory of the process.

The most well-known stochastic process, the Brownian motion, has been fundamental for the development of the SDE. The Brownian motion, also called the Wiener process, $W(t)$, is a stochastic process that satisfies the following three conditions:

- $W(0) = 0$ and the sample path $t \to W(t; w)$ is almost surely everywhere continuous.
- The increment $W(t) - W(s) \sim N(0, t-s)$, where $0 \leq s < t$.
- The increment $W(t_1) - W(s_1)$ and $W(t_2) - W(s_2)$ are independent for $0 \leq s_1 < t_1 \leq s_2 < t_2$.

Definition 2. Let $\{X(t)\}_{t \geq 0}$ be a stochastic process defined on a probability space $(\Omega, \mathcal{F}, P)$ with filtration $\{\mathcal{F}_t, t \geq 0\}$. Then this process has the **Markov property** if for any $s \geq 0$, $t \geq 0$, and $A \in \mathbb{B}$, where $\mathbb{B}$ is the set of all Borel sets,

$$P(X(t+s) \in A | \mathcal{F}_t) = P(X(t+s) \in A | X(t)).$$

The Markov property is called the memory-less property, and means that, given the current value $X(t)$, the probability distribution of any future value $X(t+s)$ $(s \geq 0)$ does not depend on any history of the process before time t. A Markov process is a stochastic process with the Markov property.

Definition 3. A **stochastic differential equation** (SDE) is a differential equation in which one or more of the terms are stochastic processes, thus resulting in a solution that is a stochastic process. A typical form is

$$dX(t) = \mu(X(t), t)dt + \sigma(X(t), t)dW(t),$$

where $\{W(t), t \geq 0\}$ is a Wiener process. This equation should be interpreted as an informal way of expressing the corresponding integral equation

$$X(t+s) = X(t) + \int_t^{t+s} \mu(X(u), u)du + \int_t^{t+s} \sigma(X(u), u)dW(u).$$

1.2.2 Stochastic Integrals

An SDE that describes a stochastic process $\{S(t), 0 \leq t \leq T\}$ is written in differential form as

$$dS(t) = f(S(t), t)dt + g(S(t), t)dW(t), \quad 0 \leq t \leq T,$$

or in integral form as

$$S(t) = S(0) + \int_0^t f(S(u), u)du + \int_0^t g(S(u), u)dW(u), \quad 0 \leq t \leq T,$$

where the first integral on the right-hand side is a Riemann integral and the second is a stochastic integral.

In order to define the stochastic integral $\int_0^t g(S(u), u)dW(u)$, let $\{t_k\}$ be a partition of $(0, t]$ with $0 = t_0 < t_1 < \cdots < t_m = t$. Then the stochastic integral can be written in L^2 sense as

$$\int_0^t g(S(u), u)dW(u) = \lim_{h \to 0} \sum_{k=0}^{m-1} g(S(\tau_k), \tau_k)(W(t_{k+1}) - W(t_k)),$$

where $h = \max(t_{k+1} - t_k)$ and $\tau_k = (1 - \lambda)t_k + \lambda t_{k+1}$, $k = 0, 1, \cdots, m-1$, $\lambda \in [0, 1]$. There are two well-known ways to choose λ:

- $\lambda = 0 \Rightarrow \tau_k = t_k$ (the Itô integral)
- $\lambda = 1/2 \Rightarrow \tau_k = (t_{k+1} + t_k)/2$ (the Stratonovich integral)

We will use the Itô integral as it is widely used in mathematical finance while the Stratonovich integral is frequently used in physics. It can be proved that the two integrals are actually equivalent. A detailed proof can be found in Øksendal (2013). Using Itô's lemma, we can find a "translation" between the Itô and Stratonovich calculus:

- Itô's SDE

$$dS(t) = f(S(t), t)dt + g(S(t), t)dW(t);$$

- Stratonovich's SDE

$$dS(t) = \tilde{f}(S(t), t)dt + g(S(t), t) \circ dW(t),$$

$$\tilde{f}(S(t), t) = f(S(t), t) - \frac{1}{2} g(S(t), t) \frac{\partial}{\partial s} g(s, t)|_{s=S(t)}.$$

1.2.3 Standard Models for Asset Prices

The basic assumption in the standard models for asset prices is that the history of asset prices is available and can be examined, but cannot be used to forecast future asset prices. However, it may be possible from examination of the past prices to predict likely jumps in the future asset prices, and to estimate their means, variances, and the probability distributions. Almost all option pricing models are based on one simple model for asset price movements, and involve parameters derived, for example, from historical or market data. It is assumed that asset prices must move randomly because of the efficient market hypothesis. This hypothesis can be briefly described by the following two statements:

- The history is fully reflected in the present price, and does not hold any further information.

- Markets respond immediately to any new information about an asset.

Hence, according to this hypothesis, the modeling of asset prices can be interpreted as the modeling of the arrivals of new information that affects the asset price. With these two assumptions, changes in the asset price define a Markov process. For the valuation of an asset, the absolute change in the asset price is not a useful quantity. Instead, most models associate a return to each change in the asset price, where the return is defined to be the change in the price divided by the original value. Informally this can be described as follows: if at time t the asset price is $S(t)$, and during a subsequent time interval dt the price changes to $S(t) + dS(t)$, then the goal is to model the return defined as $dS(t)/S(t)$.

The simplest model for valuing the return $dS(t)/S(t)$ decomposes it into two parts. One part is the predictable, deterministic, and anticipated return similar to the return on a risk-free investment. This part gives a contribution μdt, where μ is a measure of the average rate of growth of the asset price, also known as the drift. The second part models the random change in the asset price in response to external effects, such as unexpected events. This part is represented by a random sample drawn from a normal distribution with mean zero and adds a term $\sigma dW(t)$, where σ is the so-called volatility, which measures the standard deviation of the returns. The term $dW(t)$ contains the randomness of the asset process and is normally distributed with mean zero and variance dt, where $W(t)$ is a standard Wiener process.

If we put the drift term and the volatility term together, then we get the most important example, the GBM, which is a stochastic process $S(t)$ that satisfies the following SDE:

$$\frac{dS(t)}{S(t)} = \mu dt + \sigma dW(t),$$

or, as is commonly written in the literature,

$$dS(t) = \mu S(t)dt + \sigma S(t)dW(t). \tag{1.2.1}$$

We will examine the properties of the GBM. By Itô's lemma,

$$df(t, X(t)) = \frac{\partial f}{\partial t}dt + \frac{\partial f}{\partial x}dX(t) + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(dX(t))^2.$$

We can write the solution to (1.2.1) in an analytic format as

$$d\log(S(t)) = (\mu - \sigma^2/2)dt + \sigma dW(t),$$

$$S(t) = S(0)\exp\{(\mu - \sigma^2/2)t + \sigma W(t)\}.$$

Thus, the mean and variance of $S(t)$ are

$$E(S(t)) = S(0)e^{\mu t} \text{ and } \mathrm{Var}(S(t)) = S(0)^2 e^{2\mu t}(e^{\sigma^2 t - 1}).$$

The GBM has several desired properties for modeling asset prices:

- The expected returns of the GBM are independent of the value of the process (stock price).

- The GBM takes only positive values, just as real stock prices.
- The GBM shows the same kind of "roughness" in its paths as we see in real stock prices.
- Calculations with the GBM are relatively easy. It has an analytic solution, while most SDEs do not have closed-form solutions.

The GBM is commonly used to model the evolution of an underlying asset price in the Black-Scholes model, which is the most widely used mathematical model of a financial market containing certain derivative investment instruments. However, the GBM is not a completely realistic model. In particular, it falls short of reality in the following points:

- In the real stock market, volatility changes over time (possibly stochastically), but the GBM assumes constant volatility.
- In the real stock market, returns of an asset are usually not normally distributed. Instead, they have higher kurtoses ("fatter tails"). This means that there is a higher chance of large price movements.
- In the real stock market, returns of an asset are not necessarily independent of the historical prices of the underlying asset.

CHAPTER 2. STOCHASTIC DELAY DIFFERENTIAL EQUATIONS

2.1 Introduction

The importance of considering stochastic delay differential equations (SDDEs) is that they provide a framework under which one can deviate somewhat from the efficient market hypothesis. That is, all publicly available information is fully reflected in current asset prices. However, it is not realistic to make such an assumption, for example, to believe that all the market information is available to the public and digested uniformly across the market. Even news published on the annual reports of large corporations, such as investment banks, can not be absorbed and held by all investors simultaneously and immediately after the announcement. The insiders and elite traders are able to get and absorb information differently from the general public and then manage to influence the market movement.

To accommodate our understanding of the model formulation, instead of assuming that the underlying asset price follows the GBM and has the Markov property, we assume that the future asset price evolution depends on not only the current state, but also the historical states. Our approach is based on SDDEs, generalized from both deterministic delay differential equations and ordinary stochastic differential equations. We will explain our motivation in detail and study the mathematical properties of SDDEs. Related work can be found in Mohammed (1984), Mao (1997), and Stoica (2005).

2.2 Motivation

The stock market is full of information and price movements are based on how investors understand this information available to them. To illustrate possible issues with the efficient market hypothesis and to show the importance of historical information, we look at the stock prices of two major American financial institutions, Bank of America Corporation and Wells Fargo & Company, from July 2012 to January 2013. On July 12, 2012, shares of Bank of America Corporation were traded at \$7.48 while shares of Wells Fargo & Company were traded at \$32.85 (market closing prices). In the following two months, shares of those two financial institutions were traded within a similar range of about 5%.

The two institutions filed their quarterly reports in the middle of October 2012. On Friday, October 12, 2012, Wells Fargo & Company announced that the third quarter of 2012 was the company's best quarter ever in terms of net income, even though the company's revenue did not impress the market. The company earned 88 cents per share, compared with the average analyst estimate of 87 cents per share, and 72 cents per share the company earned in the same quarter a year ago. Even though the company's total revenue of \$21.2 billion grew by 8% compared with the same quarter a year ago, the analysts were expecting the company to generate \$21.5 billion in revenue.

Wells Fargo posted a record profit of \$4.9 billion for the third quarter as mortgage lending picked up, but the weaker-than-expected revenue left investors disappointed. Wells Fargo, the nation's largest mortgage lender, said it originated \$139 billion of mortgages during the third quarter, up 6% from a year ago, as record low interest rates drove home owners to refinance their mortgages. The company increased its market share in the mortgage origination and servicing markets, and continued to gain low-cost deposits, but the low rates also weighed on the interest income that Wells Fargo earned on its loans

Figure 2.1: Bank of America versus Wells Fargo: stock prices from July 2012 to January 2013.

and other investments. In fact, the bank's profit margin from lending and investing, a key financial measure for banks, fell to 3.66% from 3.88% a year ago.

In the short term, Wells Fargo faced challenges from its competitors and needed to deal with the pressure from net interest margin contraction. In the long term, Wells Fargo had the potential to earn between \$3.8 and \$4 in the year 2013, which would put the forward price-earnings (P/E) ratio on the low end at about 9. Thus, the forward earnings yield would be greater than 10%, allowing ample room for dividend increases and share buybacks moving forward.

Five days later, on Wednesday, October 17, 2012, Bank of America Corporation reported a net profit of \$340 million and a revenue of \$20.7 billion for the third quarter, compared with a profit of \$6.2 billion and a revenue of \$28.7 billion a year ago. Although its revenue decreased and missed the average expectation, its profit was better than expected.

This earnings report was extremely convoluted and noisy, but through a careful examination of the facts, it was clear that Chief Executive Officer Brian Moynihan had done a masterful job at building up capital and reducing risk. When Moynihan took over on January 2010, the company was heavily burdened by two massive and ill-fated acquisitions, and was forced to deal with these problems with some of the lowest capital ratios among large U.S. banks. Consequently, the share price went down to $5 from about $20. Then the company launched a broad cost-cutting program that aimed at eliminating $8 billion in annual expenses and 30,000 jobs. In August 2011, the company enhanced its capital structure and prospects through a strategic capital raise of $5 billion from Berkshire Hathaway; it also adjusted its business model to work better in the new regulatory and economic landscape. Non-interest expenses decreased by 6.6% year over year compared with the third quarter of 2011. While slower than its competitors, it was gradually gaining ground and is moving in the right direction.

Through this post-report analysis, we could draw the conclusion that the quarterly reports contained different information for the two companies and might impact the share prices in the future. After its quarterly report, on October 15, 2012, Wells Fargo was downgraded by an analyst from Buy to Neutral, while on November 19, Bank of America was upgraded by an analyst from Hold to Buy. Also, the market reaction to the quarterly reports from the two financial institutions was different in the short term. Consequently, the stock of Bank of America kept the good momentum while the stock of Wells Fargo was no longer the favorite pick in the financial sector at that moment; the Bank of America share price continued to soar while the Wells Fargo share price tumbled. After three months, at the end of 2012, the return of Bank of America shares was above 50% while the return of Wells Fargo shares was around 10%.

In retrospect, it seemed natural that when some relative calmness returned to the market in the end of 2012, investors would choose to invest more in companies that announced profits rather than losses. However, before the quarterly reports, the share price of Bank of America had already surged about 20% while the share price of Wells Fargo had not changed much. Those price movements reflected the prediction of investors for the two companies before and after the quarterly reports. One could observe the presence of a feedback effect after the announcement of the financial results for the two major American financial institutions. More important, historical information impacted both current and future price movement. This and other real examples motivated us to consider the following factors in our research.

- Backward/forward looking
  - Elite traders tend to get information from historical behaviors of the stock prices and perform fundamental analysis of the company to help them set realistic expectations.
  - Insiders are able to get information ahead of the public.
  - The market derives forward-looking projections for the financial assets.
- Insider information
  - Regular traders expect the dynamics of the stock price to follow a GBM.
  - Insiders know that the stock price is influenced by certain past events.

Therefore, we are specifically interested in modeling the following scenario:

- Stock price movements stay in a certain range for a certain time period, most often one to three months.
- Stock prices surge or plunge after the above stable movement stage.

We believe that it is important to explore a modeling approach based on both current and historical information for the above scenario. We would like to answer the following question: What is the most likely movement of the stock price after a significant movement?

2.3 Stochastic Delay Differential Equations (SDDEs)

2.3.1 Formulation

Definition 4. Let $(\Omega, \mathcal{F}, P)$ be a complete probability space with a filtration $\{\mathcal{F}_t\}_{t\geq 0}$ satisfying the usual conditions (increasing and right continuous while $\mathcal{F}_0$ contains all P-null sets), and $W(t)$ be a Wiener process defined on the probability space. The SDDE has the following form:

$$\begin{cases} dS(t) = f(S(t), S(t-\tau), t)dt + g(S(t), S(t-\tau), t)dW(t), \quad t \geq 0, \\ S(t) = \phi(t), \quad -\tau \leq t \leq 0, \end{cases} \tag{2.3.1}$$

where the delay $\tau > 0$ is fixed, $f : R \times R \times R_+ \to R$, and $g : R \times R \times R_+ \to R$, with $R_+ = [0, \infty)$. In order to solve equation (2.3.1) we need to know the initial data and assume that $\phi(t)$ is an $\mathcal{F}_0$ measurable random variable with $E(\sup_{-\tau\leq t\leq 0} |\phi(t)|^2) < \infty$, and that $\phi(\cdot)$ is continuous on $[-\tau, 0]$ for each $w \in \Omega$.

If the function g does not depend on S, then we say that equation (2.3.1) has an additive noise; otherwise, the equation has a multiplicative noise. If the functions f and g do not explicitly depend on t, then the equation is called autonomous, and we have

$$dS(t) = f(S(t), S(t-\tau))dt + g(S(t), S(t-\tau))dW(t).$$

It is natural to study the existence and uniqueness of the solution to equation (2.3.1). There are constraints imposed on the functions f and g to ensure that.

2.3.2 Existence and Uniqueness of the Solution

Definition 5. Let $(\Omega, \mathcal{F}, P)$ be a complete probability space with a filtration $\{\mathcal{F}_t\}_{t\geq 0}$. For the following SDDE

$$\begin{cases} dS(t) = f(S(t), S(t-\tau), t)dt + g(S(t), S(t-\tau), t)dW(t), \quad 0 \leq t \leq T, \\ S(t) = \phi(t), \quad -\tau \leq t \leq 0, \end{cases} \tag{2.3.2}$$

a real-valued stochastic process $S(t)\colon [-\tau, T] \times \Omega \to R$ is called a **strong solution**, if it is a measurable, sample-continuous, and $\mathcal{F}_t$ adapted process, and satisfies the SDDE (2.3.2) almost surely.

Definition 6. If any other strong solution $\hat{S}(t)$ is indistinguishable from a strong solution $S(t)$, namely

$$P\left(S(t) = \hat{S}(t) \text{ for all } -\tau \leq t \leq T\right) = 1,$$

then we say that the strong solution $S(t)$ is **pathwise unique**.

Definition 7. The functions f and g in equation (2.3.2) satisfy the **local Lipschitz condition**, if for each integer $i \geq 1$, there exists a positive constant K_i such that

$$|f(x_1, y_1, t) - f(x_2, y_2, t)| \vee |g(x_1, y_1, t) - g(x_2, y_2, t)| \leq K_i(|x_1 - x_2| + |y_1 - y_2|)$$

for x_1, y_1, x_2, and $y_2 \in R$ with $|x_1| \vee |y_1| \vee |x_2| \vee |y_2| \leq i$ and any $t \in R_+$, where $|x| \vee |y| = \max(|x|, |y|)$.

Definition 8. The functions f and g in equation (2.3.2) satisfy the **linear growth condition**, if there exists a positive constant K such that

$$|f(x, y, t)| \vee |g(x, y, t)| \ \leq K(1 + |x| + |y|)$$

for all $(x, y, t) \in R \times R \times R_+$.

We now state the existence and uniqueness theorem as follows.

Theorem: Under the local Lipschitz condition and the linear growth condition, the SDDE (2.3.1) has a pathwise unique strong solution. Moreover, the solution has the property that

$$E\left(\sup_{-\tau \leq t \leq T} |S(t)|^2\right) < \infty \ \text{ for any } 0 < T < \infty.$$

A detailed proof can be found in Mohammed (1984) and Mao (1997). This theorem can be proved by the standard technique of Picard's iteration. Note that on $[0, \tau]$, the SDDE (2.3.1) becomes

$$dS(t) = f(S(t), \phi(t-\tau), t)dt + g(S(t), \phi(t-\tau), t)dW(t),$$

with the initial value $S(0) = \phi(0)$. Note that this SDE has a unique solution if the linear growth condition holds and $f(x, y, t)$ and $g(x, y, t)$ are locally Lipschitz continuous in x (not necessarily in y). Moreover, the solution has the property that

$$E\left(\sup_{-\tau \le t \le \tau} |S(t)|^2\right) < \infty.$$

Once the solution $S(t)$ on $[0, \tau]$ is known, we can proceed the above argument on the intervals $[\tau, 2\tau]$, $[2\tau, 3\tau]$ and so on, and hence draw the conclusion about the solution on $[-\tau, \infty)$.

2.3.2.1 Example

Based on the motivation in Section 2.2, we model stock returns using an SDDE. We assume that trading takes place continuously over time. The stock returns respond to the information absorbed by speculators at the previous time point τ. Namely, the split of the trading asset could depend on the information arrival time point τ. In the continuous-time modeling, this feedback process is modeled by an SDDE with a linear delay, which has the following form:

$$\begin{cases} dS(t) = (a \cdot S(t) + b \cdot S(t-\tau))dt + (c \cdot S(t) + d \cdot S(t-\tau))dW(t), & 0 \le t \le T, \\ S(t) = \phi(t), \quad -\tau \le t \le 0, \end{cases} \tag{2.3.3}$$

where the delay $\tau > 0$ is fixed, $f : R \times R \times R_+ \to R$, and $g : R \times R \times R_+ \to R$. In order to solve equation (2.3.3) we need to know the initial data and assume that $\phi(t)$ is an $\mathcal{F}_0$ measurable random variable with $E(\sup_{-\tau \le t \le 0} |\phi(t)|^2) < \infty$, and that $\phi(\cdot)$ is continuous on $[-\tau, 0]$ for each $w \in \Omega$.

First we check that the SDDE (2.3.3) satisfies the local Lipschitz condition and the linear growth condition. For equation (2.3.3), $f(x, y, t) = ax + by$ and $g(x, y, t) = cx + dy$. Thus, we can have

$$\begin{aligned}|f(x_1, y_1, t) - f(x_2, y_2, t)| &= |a(x_1 - x_2) + b(y_1 - y_2)| \\ &\leq (|a| \vee |b|)(|x_1 - x_2| + |y_1 - y_2|),\end{aligned}$$

$$\begin{aligned}|g(x_1, y_1, t) - g(x_2, y_2, t)| &= |c(x_1 - x_2) + d(y_1 - y_2)| \\ &\leq (|c| \vee |d|)(|x_1 - x_2| + |y_1 - y_2|).\end{aligned}$$

Combining the above two inequalities, we can see that the global Lipschitz condition is satisfied. Hence the local Lipschitz condition and the linear growth condition are both satisfied. Then for equation (2.3.3), when $t \in [0, \tau]$, the SDDE with a linear delay becomes a linear SDE

$$dS(t) = (a \cdot S(t) + b \cdot \phi_1(t))dt + (c \cdot S(t) + d \cdot \phi_1(t))dW(t),$$

where $\phi_1(t) = \phi(t - \tau)$, with the initial value $S(0) = \phi(0)$. This linear SDE has the following explicit solution

$$\begin{aligned}S(t) = \Phi_1(t)\{\phi(0) &+ \int_0^t \Phi_1^{-1}(s)\,[b \cdot \phi_1(s) - c \cdot d \cdot \phi_1(s)]\,ds \\ &+ \int_0^t \Phi_1^{-1}(s) \cdot d \cdot \phi_1(s) dW(s)\},\end{aligned}$$

where

$$\Phi_1(t) = \exp\left((a - c^2/2)t + cW(t)\right).$$

Next, when $t \in [\tau, 2\tau]$, the SDDE with a linear delay (2.3.3) becomes a linear SDE

$$dS(t) = (a \cdot S(t) + b \cdot \phi_2(t))dt + (c \cdot S(t) + d \cdot \phi_2(t))dW(t),$$

where $\phi_2(t) = S(t-\tau)$ is known and the initial value $S(\tau)$ can be obtained from the previous step. This linear SDE has the following explicit solution

$$
\begin{aligned}
S(t) = \Phi_2(t)\{S(\tau) + \int_\tau^t \Phi_2^{-1}(s)\,[b \cdot \phi_2(s) - c \cdot d \cdot \phi_2(s)]\,ds \\
+ \int_\tau^t \Phi_2^{-1}(s) \cdot d \cdot \phi_2(s) dW(s)\},
\end{aligned}
$$

where

$$
\Phi_2(t) = \exp\left((a - c^2/2)t + cW(t)\right).
$$

Then, we can repeat this procedure over the intervals $[i\tau, (i+1)\tau]$, for $i = 2,\, 3, \cdots$, and obtain the explicit solution for this SDDE with a linear delay. We also get the following result

$$
E\left(\sup_{-\tau \le t \le T} |S(t)|^2\right) < \infty, \text{ for any } 0 < T < \infty.
$$

CHAPTER 3. NUMERICAL METHODS

3.1 Introduction

SDDEs generalize both deterministic delay differential equations and stochastic ordinary differential equations, incorporate memory to provide more realistic models for phenomena in many areas of science, and have a richer structure than the regular SDEs. However, an SDDE in general has no closed-form solution. Thus we have to use numerical techniques to approximate the solution and understand its quantitative behavior. Related work includes, for example, Kloeden and Platen (1992), Buckwar (2000), Higham (2001), Durham and Gallant (2002), and Mao (2003).

Numerical techniques are widely used to approximate complex systems. In this chapter, we will investigate a numerical technique to approximate the strong solutions of a linear SDDE. We will first study the consistency and show the convergence of the numerical scheme under different senses. We will then present numerical examples and give a motivation from the financial market.

3.2 Notation and Definitions

Let $(\Omega, \mathcal{F}, P)$ be a complete probability space with a filtration $\{\mathcal{F}_t\}_{t\geq 0}$ satisfying the usual conditions (increasing and right continuous while $\mathcal{F}_0$ contains all P-null sets), and $W(t)$ be a Wiener process defined on the probability space. We say a random variable $X \in L^p(\Omega, \mathcal{F}, P)$, where $0 < p < \infty$, if

$$\parallel X \parallel_p \equiv (E(|X|^p))^{1/p} < \infty.$$

From Jensen's inequality, we know that

$$\| X \|_p \leq \| X \|_q \quad \text{for all } 1 \leq p \leq q \text{ and } X \in L^q(\Omega, \mathcal{F}, P).$$

We now set up the framework of a numerical scheme for solving the following autonomous SDDE

$$\begin{cases} dS(t) = f(S(t), S(t-\tau))dt + g(S(t), S(t-\tau))dW(t), \;\; 0 \leq t \leq T, \\ S(t) = \phi(t), \;\; -\tau \leq t \leq 0. \end{cases} \tag{3.2.1}$$

- Let $t_0 < t_1 < \cdots < t_N$ be an equidistant partition of the interval $[0, T]$, where the n-th step-point $t_n = n \cdot h$, $n = 0, 1, ..., N$, and $h = T/N$ is the mesh of the partition.
- Assume that $N_\tau \equiv \frac{\tau}{h}$ and $N_0 \equiv \frac{T}{\tau}$ are both integers.
- Let $\Delta W_{n+1} = W(t_{n+1}) - W(t_n)$ be the increment of the Wiener process $W(t)$.

Suppose $\tilde{S}_n$ is an approximation of the strong solution to (3.2.1). Then we can construct an explicit single-step scheme with an increment function Ψ as the following:

$$\begin{cases} \tilde{S}_{n+1} = \tilde{S}_n + \Psi(h, \tilde{S}_n, \tilde{S}_{n-N_\tau}, \Delta W_{n+1}), \;\; 0 \leq n \leq N-1, \\ \tilde{S}_{n-N_\tau} = \phi(t_n - \tau), \;\; 0 \leq n \leq N_\tau. \end{cases} \tag{3.2.2}$$

Let $\bar{S}(t_{n+1})$ be the approximation of $S(t_{n+1})$ obtained after only one step of the scheme from t_n to t_{n+1}, that is,

$$\bar{S}(t_{n+1}) = S(t_n) + \Psi(h, S(t_n), S(t_{n-N_\tau}), \Delta W_{n+1}).$$

Note that $\tilde{S}_{n+1}$ is the approximation of $S(t_{n+1})$ using (3.2.2).

A numerical scheme for solving a stochastic system should satisfy certain conditions, which naturally depend on the form of the stochastic system. Following the above notation, we now consider two ways of measuring the accuracy of a numerical scheme to approximate the strong solution of an SDDE, the local and global error.

Definition 9. The **local error** is the sequence of random variables

$$\delta_n = S(t_n) - \bar{S}(t_n), \ \ n = 1, ..., N.$$

The local error measures the difference between the approximation and the exact solution on a sub-interval of the integration given an exact initial condition at the start of the sub-interval.

Definition 10. The **global error** is the sequence of random variables

$$\epsilon_n = S(t_n) - \tilde{S}_n, \ \ n = 1, ..., N.$$

The global error measures the buildup of the difference between the approximation and the exact solution over the entire integration range.

Throughout this chapter, the expectations for the local error are conditional on the current state $\mathcal{F}_{t_n}$ while the expectations for the global error are conditional on the initial data $\mathcal{F}_{t_0}$. For simplicity, we will write $E(\delta_{n+1}|\mathcal{F}_{t_n})$ as $E(\delta_{n+1})$ and so on, and write $E(\epsilon_{n+1}|\mathcal{F}_{t_0})$ as $E(\epsilon_{n+1})$ and so on. Also note that, for $t \geq 0$, $S(t)$ and $W(t)$ are $\mathcal{F}_t$-measurable, and for $-\tau \leq t \leq 0$, $S(t)$ is $\mathcal{F}_0$-measurable. In particular, $\mathcal{F}_0$ can be defined as the σ-algebra generated by $\{S(t), -\tau \leq t \leq 0\}$ and all P-null sets.

Definition 11. The numerical scheme (3.2.2) is **consistent** in the mean with order p_1 and in the mean square with order p_2 if the following inequalities hold with $p_2 \geq 1/2$ and $p_1 \geq p_2 + 1/2$:

$$\max_{1 \leq n \leq N} |E(\delta_n)| \leq Ch^{p_1} \text{ as } h \to 0,$$

$$\max_{1 \leq n \leq N} (E|\delta_n|^2)^{1/2} \leq Ch^{p_2} \text{ as } h \to 0,$$

where the constant C does not depend on the mesh h, but may depend on T and the initial conditions.

Definition 12. The numerical scheme (3.2.2) is **convergent** in the mean with order p if the following inequality holds:

$$\max_{1 \leq n \leq N} |E(\epsilon_n)| \leq Ch^p \text{ as } h \to 0.$$

The numerical scheme (3.2.2) is **convergent** in the mean square with order p if the following inequality holds:

$$\max_{1 \leq n \leq N} (E|\epsilon_n|^2)^{1/2} \leq Ch^p \text{ as } h \to 0.$$

It is very important to understand the convergence in both the mean and the mean square. The motivation for considering this is the computation of the expectation of the solution to the SDDE. For instance, in finance, the expected return is one of the most important characteristics of a financial asset to be calculated. Once the convergence is established, we are able to approximate the expected value $E(S(t_n))$ by the average of M realizations $\dfrac{1}{M}\sum_{i=1}^{M} \tilde{S}_n^{(i)}$ in the discretized scheme for fairly large M.

3.3 The Euler-Maruyama Scheme

In mathematics, the Euler-Maruyama scheme is a method for obtaining an approximate numerical solution of an SDE. This method is named after Leonhard Euler and Gisiro Maruyama, and is a generalization of the Euler method for ordinary differential equations (ODEs) to SDEs. Here we apply the Euler-Maruyama scheme to the linear SDDE (2.3.3). First, we make a partition $t_0 < t_1 < \cdots < t_N$ of the interval $[0, T]$, where $t_n = n \cdot h$, $n = 0, 1, \cdots, N$, and $h = T/N$. For the delay τ, assume that $\tau = N_\tau \cdot h$ for an integer N_τ. Then the Eular-Maruyama scheme has the following increment function:

$$\begin{aligned} \Psi(h, \tilde{S}_n, \tilde{S}_{n-N_\tau}, \Delta W_{n+1}) &= f(\tilde{S}_n, \tilde{S}_{n-N_\tau})h + g(\tilde{S}_n, \tilde{S}_{n-N_\tau})\Delta W_{n+1} \\ &= (a \cdot \tilde{S}_n + b \cdot \tilde{S}_{n-N_\tau})h + (c \cdot \tilde{S}_n + d \cdot \tilde{S}_{n-N_\tau})\Delta W_{n+1}, \end{aligned} \tag{3.3.1}$$

for $0 \leq n \leq N-1$, where $\Delta W_{n+1} = W(t_{n+1}) - W(t_n)$. Here, $\Delta W_1, \ldots, \Delta W_N$ are independent $N(0, h)$ random variables.

Theorem 3.3.1. If the functions f and g satisfy the local Lipschitz condition and the linear growth condition, then the Euler-Maruyama scheme is consistent in the mean with order $p_1 = 2$ and in the mean square with order $p_2 = 1$. Namely,

$$\max_{1\leq n\leq N} |E(\delta_n)| \leq Ch^2 \text{ as } h \to 0,$$
$$\max_{1\leq n\leq N} (E|\delta_n|^2)^{1/2} \leq Ch \text{ as } h \to 0,$$

where the constant C does not depend on the mesh h, but may depend on T and the initial conditions.

A proof of this theorem can be found in Buckwar (2000). To apply this theorem to the linear SDDE (2.3.3), we note that the functions $f(x, y) = ax + by$ and $g(x, y) = cx + dy$ both satisfy the global Lipschitz condition, which implies the local Lipschitz condition and the linear growth condition. Thus the Euler-Maruyama scheme for the linear SDDE (2.3.3) is consistent in the mean with order $p_1 = 2$ and in the mean square with order $p_2 = 1$. To prove the convergence of the scheme for the linear SDDE, both in the mean and in the mean square, we first establish the following lemma for the increment function Ψ in the scheme. Note that from (3.3.1),

$$\begin{aligned} d_n &\equiv \Psi(h, S(t_n), S(t_n-\tau), \Delta W_{n+1}) - \Psi(h, \tilde{S}_n, \tilde{S}_{n-N_\tau}, \Delta W_{n+1}) \\ &= \Big(ah\cdot S(t_n) + bh\cdot S(t_n-\tau)\Big) + \Big(c\cdot S(t_n) + d\cdot S(t_n-\tau)\Big)\Delta W_{n+1} \\ &- \Big(ah\cdot \tilde{S}_n + bh\cdot \tilde{S}_{n-N_\tau}\Big) - \Big(c\cdot \tilde{S}_n + d\cdot \tilde{S}_{n-N_\tau}\Big)\Delta W_{n+1} \\ &= \Big(ah\cdot \epsilon_n + bh\cdot \epsilon_{n-N_\tau}\Big) + \Big(c\cdot \epsilon_n + d\cdot \epsilon_{n-N_\tau}\Big)\Delta W_{n+1}. \end{aligned} \tag{3.3.2}$$

Lemma 3.3.1. For the linear SDDE (2.3.3), there exist positive constants K_1 and K_2

that depend only on a, b, c, and d, such that for any $0 < h < 1$,

$$\begin{aligned} &E(|d_n\epsilon_n||\mathcal{F}_{t_0}) \le K_1 h \left\{E(|\epsilon_n|^2|\mathcal{F}_{t_0}) + E(|\epsilon_{n-N_\tau}|^2|\mathcal{F}_{t_0})\right\}, \\ &E(|d_n|^2|\mathcal{F}_{t_0}) \le K_2 h^2 \left\{E(|\epsilon_n|^2|\mathcal{F}_{t_0}) + E(|\epsilon_{n-N_\tau}|^2|\mathcal{F}_{t_0})\right\}. \end{aligned} \tag{3.3.3}$$

In the linear SDDE (2.3.3), as we mentioned before, the functions $f(x, y) = ax + by$ and $g(x, y) = cx + dy$ satisfy the global Lipschitz condition. Then the proof is straightforward and omitted here.

Theorem 3.3.2. The Euler-Maruyama scheme for the linear SDDE (2.3.3) is convergent in the mean with order 1 and in the mean square with order 1/2, that is,

$$\begin{aligned} &\max_{1\le n\le N} |E(\epsilon_n)| \le C^* h \text{ as } h \to 0, \\ &\max_{1\le n\le N} (E|\epsilon_n|^2)^{1/2} \le C^* h^{1/2} \text{ as } h \to 0, \end{aligned} \tag{3.3.4}$$

where the constant C^* does not depend on the mesh h but may depend on T and the initial conditions.

Proof: We first rearrange the expression of the global error as the following:

$$\begin{aligned} \epsilon_{n+1} &= S(t_{n+1}) - \tilde{S}_{n+1} \\ &= \Big[S(t_{n+1}) - S(t_n) - \Psi(h, S(t_n), S(t_n - \tau), \Delta W_{n+1})\Big] + \Big[S(t_n) - \tilde{S}_n\Big] \\ &+ \Big[\Psi(h, S(t_n), S(t_n - \tau), \Delta W_{n+1}) - \Psi(h, \tilde{S}_n, \tilde{S}_{n-N_\tau}, \Delta W_{n+1})\Big] \\ &= \delta_{n+1} + \epsilon_n + d_n. \end{aligned} \tag{3.3.5}$$

Now we show the convergence in the mean with order 1. Applying the conditional expectation with respect to the σ-algebra $\mathcal{F}_{t_0}$ and using Theorem 3.3.1 and (3.3.2), we have the following inequalities holding almost surely, for $h > 0$ sufficiently small:

$$\begin{aligned}
|E(\epsilon_{n+1}|\mathcal{F}_{t_0})| &= |E(\delta_{n+1}|\mathcal{F}_{t_0}) + E(\epsilon_n|\mathcal{F}_{t_0}) + E(d_n|\mathcal{F}_{t_0})| \\
&\leq |E(\delta_{n+1}|\mathcal{F}_{t_0})| + |E(\epsilon_n|\mathcal{F}_{t_0})| + |E(d_n|\mathcal{F}_{t_0})| \\
&= \Big|E\big(E(\delta_{n+1}|\mathcal{F}_{t_0})|\mathcal{F}_{t_n}\big)\Big| + |E(\epsilon_n|\mathcal{F}_{t_0})| + |E(d_n|\mathcal{F}_{t_0})| \\
&= \Big|E\big(E(\delta_{n+1}|\mathcal{F}_{t_n})|\mathcal{F}_{t_0}\big)\Big| + |E(\epsilon_n|\mathcal{F}_{t_0})| + |E(d_n|\mathcal{F}_{t_0})| \\
&\leq Ch^2 + |E(\epsilon_n|\mathcal{F}_{t_0})| + \Bigg|E\Big\{\Big[\big(ah \cdot \epsilon_n + bh \cdot \epsilon_{n-N_\tau}\big) \\
&\quad + \big(c \cdot \epsilon_n + d \cdot \epsilon_{n-N_\tau}\big)\Delta W_{n+1}\Big]\Big|\mathcal{F}_{t_0}\Big\}\Bigg| \\
&\leq Ch^2 + (1 + |ah|) \cdot |E(\epsilon_n|\mathcal{F}_{t_0})| + |bh| \cdot |E(\epsilon_{n-N_\tau}|\mathcal{F}_{t_0})|.
\end{aligned} \tag{3.3.6}$$

To show that the expectation of the global error is bounded by a constant times the mesh h, we now use mathematical induction, recursively through the partition intervals of length τ in $[0, T]$.

Step 1: For $n = 0, 1, \cdots, N_\tau$, $t_n \in [0, \tau]$, and $t_{n-N_\tau} \in [-\tau, 0]$, corresponding to the initial conditions. Thus, $\epsilon_{n-N_\tau} = 0$ and

$$|E(\epsilon_{n+1}|\mathcal{F}_{t_0})| \leq (1 + |ah|) \cdot |E(\epsilon_n|\mathcal{F}_{t_0})| + Ch^2.$$

In particular, we have

$$\begin{aligned}
&|E(\epsilon_1|\mathcal{F}_{t_0})| \leq Ch^2, \\
&|E(\epsilon_2|\mathcal{F}_{t_0})| \leq Ch^2 + (1 + |ah|) \cdot |E(\epsilon_1|\mathcal{F}_{t_0})| \leq [1 + (1 + |ah|)]Ch^2, \\
&|E(\epsilon_3|\mathcal{F}_{t_0})| \leq Ch^2 + (1 + |ah|) \cdot |E(\epsilon_2|\mathcal{F}_{t_0})| \leq [1 + (1 + |ah|) + (1 + |ah|)^2]Ch^2, \\
&\qquad\vdots
\end{aligned}$$

Then we get the following general form

$$
\begin{aligned}
|E(\epsilon_{n+1}|\mathcal{F}_{t_0})| &\leq [1 + (1 + |ah|) + (1 + |ah|)^2 + \cdots + (1 + |ah|)^n]Ch^2 \\
&= \frac{(1 + |ah|)^{n+1} - 1}{(1 + |ah|) - 1}Ch^2 \\
&= [(1 + |ah|)^{n+1} - 1]\frac{C}{|a|}h^{2-1} \\
&\leq [(e^{|a|h})^{n+1} - 1]\frac{C}{|a|}h \\
&\leq [e^{|a|T} - 1]\frac{C}{|a|}h \\
&= \tilde{C}_1 h,
\end{aligned}
$$

where $\tilde{C}_1 = (e^{|a|T} - 1)\frac{C}{|a|}$ does not depend on the mesh h.

Step 2: Suppose that the above result holds for the jth interval of length τ in $[0, T]$, where $j \geq 1$, that is, for $n = (j-1)N_\tau, (j-1)N_\tau + 1, \cdots, jN_\tau$, $t_n \in [(j-1)\tau, j\tau]$ and

$$|E(\epsilon_{n+1}|\mathcal{F}_{t_0})| \leq \tilde{C}_j h.$$

Then for the $(j+1)$st interval $[j\tau, (j+1)\tau]$, $n = jN_\tau, jN_\tau + 1, \cdots, (j+1)N_\tau$, $t_n \in [j\tau, (j+1)\tau]$, and we have

$$
\begin{aligned}
|E(\epsilon_{n+1}|\mathcal{F}_{t_0})| &\leq Ch^2 + (1 + |ah|) \cdot |E(\epsilon_n|\mathcal{F}_{t_0})| + |bh| \cdot |E(\epsilon_{n-N_\tau})|\mathcal{F}_{t_0})| \\
&\leq Ch^2 + (1 + |ah|) \cdot |E(\epsilon_n|\mathcal{F}_{t_0})| + |bh| \cdot \tilde{C}_j h \\
&\leq (1 + |ah|) \cdot |E(\epsilon_n|\mathcal{F}_{t_0})| + (C + |b|\tilde{C}_j)h^2 \\
&= (1 + |ah|) \cdot |E(\epsilon_n|\mathcal{F}_{t_0})| + C^*_{j+1}h^2,
\end{aligned}
$$

where $C^*_{j+1} = C + |b|\tilde{C}_j$. Then using the same technique in the first step, we can derive that for $n = jN_\tau, jN_\tau + 1, \cdots, (j+1)N_\tau$, $t_n \in [j\tau, (j+1)\tau]$ and

$$|E(\epsilon_{n+1}|\mathcal{F}_{t_0})| \leq \tilde{C}_{j+1} h,$$

where

$$\tilde{C}_{j+1} = (e^{|a|T} - 1)\frac{C^*_{j+1}}{|a|} = (e^{|a|T} - 1)\frac{C + |b|\tilde{C}_j}{|a|}.$$

Then the result holds for the $(j+1)$st interval $[j\tau, (j+1)\tau]$. By mathematical induction, for $n = 0, 1, \cdots, N-1$, $t_n \in [0, T]$, we have

$$|E(\epsilon_{n+1}|\mathcal{F}_{t_0})| \leq \max\left(\tilde{C}_1, \cdots, \tilde{C}_{N_0}\right) \cdot h.$$

Next we show the numerical scheme's convergence in the mean square with order 1/2. By squaring both sides of (3.3.5) we get the following:

$$\begin{aligned}
|\epsilon_{n+1}|^2 &= |\delta_{n+1} + \epsilon_n + d_n|^2 \\
&\leq |\delta_{n+1}|^2 + |\epsilon_n|^2 + |d_n|^2 + 2|\delta_{n+1}||\epsilon_n| + 2|d_n||\delta_{n+1}| + 2|\epsilon_n||d_n| \\
&\leq 2|\delta_{n+1}|^2 + |\epsilon_n|^2 + 2|d_n|^2 + 2|\delta_{n+1}||\epsilon_n| + 2|\epsilon_n||d_n|.
\end{aligned}$$

Then applying the conditional expectation with respect to the σ-algebra $\mathcal{F}_{t_0}$, we get the following inequality holding almost surely.

$$\begin{aligned}
E(|\epsilon_{n+1}|^2|\mathcal{F}_{t_0}) \leq 2E(|\delta_{n+1}|^2|\mathcal{F}_{t_0}) + E(|\epsilon_n|^2|\mathcal{F}_{t_0}) + 2E(|d_n|^2|\mathcal{F}_{t_0}) \\
+ 2E(|\delta_{n+1}||\epsilon_n||\mathcal{F}_{t_0}) + 2E(|\epsilon_n||d_n||\mathcal{F}_{t_0}).
\end{aligned} \tag{3.3.7}$$

We will use Theorem 3.3.1, Lemma 3.3.1, the Cauchy-Schwartz inequality, and the properties of the conditional expectation to obtain the convergence rate for each term on the right hand side of the above inequality.

By Theorem 3.3.1, we have, for $h > 0$ sufficiently small,

$$\begin{aligned}
E(|\delta_{n+1}|^2|\mathcal{F}_{t_0}) &= E(E(|\delta_{n+1}|^2|\mathcal{F}_{t_0})|\mathcal{F}_{t_n}) \\
&= E(E(|\delta_{n+1}|^2|\mathcal{F}_{t_n})|\mathcal{F}_{t_0}) \\
&\leq (Ch)^2.
\end{aligned}$$

By Lemma 3.3.1, we have

$$E(|d_n|^2|\mathcal{F}_{t_0}) \leq K_2 h^2 \left\{E(|\epsilon_n|^2|\mathcal{F}_{t_0}) + E(|\epsilon_{n-N_\tau}|^2|\mathcal{F}_{t_0})\right\}.$$

Also for the cross-product terms, we have

$$\begin{aligned}
E(|\delta_{n+1}||\epsilon_n||\mathcal{F}_{t_0}) &= E\Big\{\Big[E\big(|\delta_{n+1}||\mathcal{F}_{t_n}\big)\cdot|\epsilon_n|\Big]|\mathcal{F}_{t_0}\Big\}\\
&\leq \Big(E\big(\left[E(|\delta_{n+1}||\mathcal{F}_{t_n})\right]^2|\mathcal{F}_{t_0}\big)\Big)^{1/2}\cdot\Big(E(|\epsilon_n|^2|\mathcal{F}_{t_0})\Big)^{1/2}\\
&\leq \Big(E\big(|Ch^2|^2|\mathcal{F}_{t_0}\big)\Big)^{1/2}\cdot\Big(E\big(|\epsilon_n|^2|\mathcal{F}_{t_0}\big)\Big)^{1/2}\\
&= \Big(E\big(|C^2h^3|\mathcal{F}_{t_0}\big)\Big)^{1/2}\cdot\Big(E\big(h|\epsilon_n|^2|\mathcal{F}_{t_0}\big)\Big)^{1/2}\\
&\leq \frac{C^2}{2}h^3+\frac{h}{2}E(|\epsilon_n|^2|\mathcal{F}_{t_0}),
\end{aligned}$$

and by Lemma 3.3.1,

$$E(|d_n||\epsilon_n||\mathcal{F}_{t_0})\leq K_1h\left[E(|\epsilon_n|^2|\mathcal{F}_{t_0})+E(|\epsilon_{n-N_\tau}|^2|\mathcal{F}_{t_0})\right].$$

Combining all the above inequalities, we have, for h sufficiently small,

$$E(|\epsilon_{n+1}|^2|\mathcal{F}_{t_0})\leq(1+\tilde{C}_1h)E(|\epsilon_n|^2|\mathcal{F}_{t_0})+\tilde{C}_2hE(|\epsilon_{n-N_\tau}|^2|\mathcal{F}_{t_0})+\tilde{C}_3h^2,$$

where $\tilde{C}_1=1+2(K_1+K_2)$, $\tilde{C}_2=2(K_1+K_2)$, and $\tilde{C}_3=3C^2$.

Similar to the convergence in the mean, we can use mathematical induction to show that

$$E(|\epsilon_n|^2|\mathcal{F}_{t_0})\leq C^*h,$$

for some constant C^*, thus proving the theorem.

Theorem 3.3.3. The Euler-Maruyama scheme for the linear SDDE (2.3.3) is convergent in the variance with order 1/2, that is,

$$\max_{1\leq n\leq N}\left|\mathrm{Var}(S(t_n))-\mathrm{Var}(\tilde{S}_n)\right|\leq C^{**}h^{1/2}\text{ as }h\to 0, \tag{3.3.8}$$

where the constant C^{**} does not depend on the mesh h but may depend on T and the initial conditions.

Proof: Note that

$$\begin{aligned}
\Big|\mathrm{Var}(S(t_n)) - \mathrm{Var}(\tilde{S}_n)\Big| &= \Big|\big(E(S^2(t_n)) - [E(S(t_n))]^2\big) - \Big(E(\tilde{S}_n^2) - \Big[E(\tilde{S}_n)\Big]^2\Big)\Big| \\
&= \Big|E\left(S^2(t_n) - \tilde{S}_n^2\right) + \Big(\Big[E(\tilde{S}_n)\Big]^2 - [E(S(t_n))]^2\Big)\Big| \\
&\leq \Big|E\Big[\left(S(t_n) - \tilde{S}_n\right) \cdot \left(S(t_n) + \tilde{S}_n\right)\Big]\Big| \\
&+ \Big|\Big(E(\tilde{S}_n) + E(S(t_n))\Big) \cdot \Big(E(\tilde{S}_n) - E(S(t_n))\Big)\Big| \\
&\leq \Big(E(S(t_n) - \tilde{S}_n)^2\Big)^{1/2} \cdot \Big(E(S(t_n) + \tilde{S}_n)^2\Big)^{1/2} \\
&+ \Big|\Big(E(\tilde{S}_n) + E(S(t_n))\Big) \cdot \Big(E(\tilde{S}_n) - E(S(t_n))\Big)\Big| \\
&\leq \max_{1\leq n\leq N} (E|\epsilon_n|^2)^{1/2} \cdot \Big(E(S(t_n) + \tilde{S}_n)^2\Big)^{1/2} \\
&+ \max_{1\leq n\leq N} |E(\epsilon_n)| \cdot \Big|E(\tilde{S}_n) + E(S(t_n))\Big|.
\end{aligned} \tag{3.3.9}$$

Furthermore, we have

$$\begin{aligned}
\Big(E(S(t_n) + \tilde{S}_n)^2\Big)^{1/2} &= \Big(E[\tilde{S}_n - S(t_n) + 2S(t_n)]^2\Big)^{1/2} \\
&= \Big(E\big(2S(t_n) - \epsilon_n\big)^2\Big)^{1/2} \\
&\leq \left(E|\epsilon_n|^2\right)^{1/2} + 2\left(E(S^2(t_n))\right)^{1/2},
\end{aligned}$$

and

$$\begin{aligned}
\Big|E(\tilde{S}_n) + E(S(t_n))\Big| &= \Big|E\left(\tilde{S}_n - S(t_n)\right) + 2E\left(S(t_n)\right)\Big| \\
&\leq E|\epsilon_n| + 2E(|S(t_n)|).
\end{aligned}$$

By the existence and uniqueness theorem of a strong solution to the linear SDDE, there exists a positive number M such that

$$E\left(\sup_{-\tau\leq t\leq T} |S(t)|^2\right) < M.$$

Thus, $E\left(|S(t_n)|^2\right) \leq M$, and $E\left(|S(t_n)|\right) \leq M^{1/2}$. Then by Theorem 3.3.2, for $h > 0$ sufficiently small, the above inequalities become

$$\left(E(S(t_n) + \tilde{S}_n)^2\right)^{1/2} \leq C^* h^{1/2} + 2M^{1/2},$$

and

$$\left|E(\tilde{S}_n) + E(S(t_n))\right| \leq C^* h + 2M^{1/2}.$$

Therefore, the inequality (3.3.9) implies that

$$\begin{aligned}\left|\mathrm{Var}(S(t_n)) - \mathrm{Var}(\tilde{S}_n)\right| &\leq C^* h^{1/2}\left[C^* h^{1/2} + 2M^{1/2}\right] + C^* h\left[C^* h + 2M^{1/2}\right] \\ &\leq C^{**} h^{1/2},\end{aligned}$$

for some constant C^{**} that does not depend on the mesh h.

3.4 Examples

We now consider the following linear SDDE

$$\begin{cases} dS(t) = (-3S(t) + 2e^{-1}S(t-1) + 3 - 2e^{-1})dt \\ \qquad\qquad + (cS(t) + dS(t-1))dW(t), \;\; 0 \leq t \leq 2, \\ S(t) = 1 + e^{-t}, \;\; -1 \leq t \leq 0, \end{cases} \tag{3.4.1}$$

where $W(t)$ is the Wiener process.

We shall use this linear SDDE to examine the convergence of the Euler-Maruyama scheme in the mean, and to illustrate the impact of delay terms on the behavior of the underlying stochastic process.

First, consider the mean function $m(t)$ of the stochastic process $S(t)$. Since $m(t) =$

$E(S(t))$, taking expectation on both sides of (3.4.1) gives

$$\begin{cases} m'(t) = -3m(t) + 2e^{-1}m(t-1) + 3 - 2e^{-1}, & 0 \le t \le 2, \\ m(t) = 1 + e^{-t}, & -1 \le t \le 0. \end{cases} \tag{3.4.2}$$

Now we solve the mean function by iteration.

When $t \in [0, 1]$,

$$m(t-1) = 1 + e^{-(t-1)}.$$

Then

$$\begin{aligned} m'(t) &= -3m(t) + 2e^{-1}(1 + e^{-(t-1)}) + 3 - 2e^{-1} \\ &= -3m(t) + 3 + 2e^{-t}. \end{aligned}$$

So for $t \in [0, 1]$, the delay differential equation (3.4.2) becomes a first-order ODE with the initial condition $m(0) = 1 + e^0 = 2$. That is,

$$\begin{cases} m'(t) = -3m(t) + 3 + 2e^{-t}, & 0 \le t \le 1, \\ m(0) = 2. \end{cases} \tag{3.4.3}$$

Note that

$$\begin{aligned} & m'(t) = -3m(t) + 3 + 2e^{-t} \\ \Leftrightarrow\ & \left[e^{3t}m(t)\right]' = 3e^{3t} + 2e^{2t} \\ \Leftrightarrow\ & e^{3t}m(t) = e^{3t} + e^{2t} + C, \end{aligned}$$

where $C = m(0) - e^0 - e^0 = 0$. Thus

$$m(t) = 1 + e^{-t}, \ \ 0 \le t \le 1.$$

Similarly, when $t \in [1, 2]$, we have

$$m(t) = 1 + e^{-t}, \ \ 1 \le t \le 2.$$

Thus, this linear SDDE (3.4.1) has the mean function

$$m(t) = 1 + e^{-t}, \ \ 0 \leq t \leq 2.$$

Figure 3.1 shows one sample path including the initial conditions and the mean function for this stochastic process after the initial conditions.

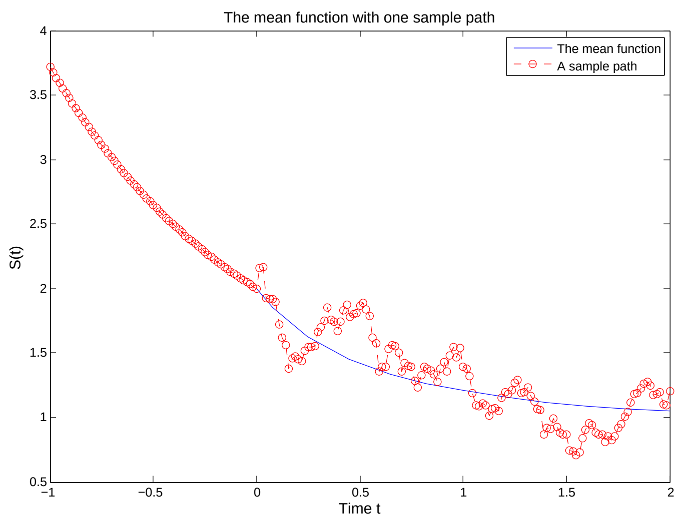


Figure 3.1: Graph of the mean function with one sample path for the stochastic process in (3.4.1).

Also, we can use this linear SDDE to examine the order of convergence in the mean as follows. Let $c = d = 0.1$ in (3.4.1). Then we simulated 50000 sample paths of the Brownian motion over the time period $[0, 2]$ with different meshes $h = 2^{-5}, ..., 2^{-9}$, and calculated the average global error at the end time point $T = 2$, namely, the difference

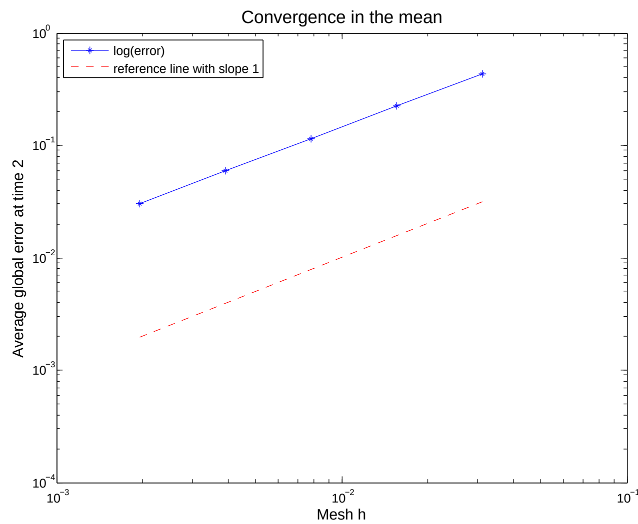


Figure 3.2: Illustration of convergence in the mean with order 1.

between the average of all sample realizations and the value of the mean function at 2. By Theorem 3.3.2,

$$\max_{1\leq n\leq N} |E(\epsilon_n)| \leq C^* h, \text{ as } h \to 0.$$

Thus,

$$\log|E(\epsilon_N)| \leq \log(C^*) + \log h, \text{ as } h \to 0.$$

So we may expect the plot of log(error) versus $\log(h)$ to show a line pattern with slope 1. In Figure 3.2, the blue line represents the $\log - \log$ scale graph of errors versus meshes and the red dashed line is the reference line with slope 1. We can see that the two lines are approximately parallel, indicating that the order of convergence in the mean

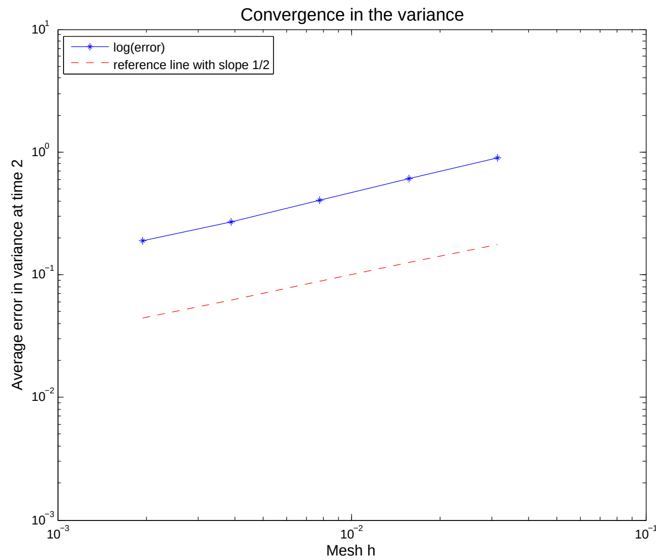


Figure 3.3: Illustration of convergence in the variance with order 1/2.

is approximately 1. Similarly, Figure 3.3 indicates that the order of convergence in the variance is approximately 0.5. Note that the average error in the variance at time $T = 2$ is the difference between the sample variance of all sample realizations and the true variance at $T = 2$.

Now we present more cases by varying the coefficient d and the length of delay τ in the above example to illustrate the impact of delay terms on the behavior of the stochastic process. First, we varied the coefficient of $S(t-\tau)$ in the diffusion portion, namely, let $d = 0.01$, 0.1, and 0.5 in (3.4.1), generating Figures 3.4, 3.5, and 3.6. In each graph, there are four colored curves, which start from the point (0, 2), after the initial condition, and represent mean functions for the stochastic process with various lengths of delay $\tau = 1, 1/2, 1/16$, and no-delay $\tau = 0$. The light blue curve located at the bottom of all four curves represents the mean function for the stochastic process with no-delay. There is a relatively sharp change before and after the point (0, 2), indicating little impact from the initial conditions on the mean function curve. The red curve, which represents the mean function for the stochastic process with length of delay $\tau = 1/16$, sits close to the light blue curve. This indicates that for this case, small perturbation in the length of delay would not result in big change in the mean function. When the length of delay is increased to $\tau = 1/2$, the corresponding green curve moves up and away from both the light blue and red curves. This indicates the significant impact of the initial conditions embedded in the delay term on the mean function afterwards. Furthermore, the blue curve located at the top of all four curves, corresponding to $\tau = 1$, is the farthest one away from the no-delay curve. These graphs show that as the length of delay $\tau \to 0$, the mean of the delay process approaches the mean of the no-delay process.

Also in these graphs, the red circles represent one sample realization of the stochastic process. By gradually increasing d in the diffusion portion, we observe that the stochastic process becomes more volatile, which is not surprising. Next, we restricted that $c + d = 0.2$ and generated several graphs in Figure 3.7 by plugging in different combinations of c and d. Apparently, even though the summation of c and d does not change, different combinations of these two coefficients could affect the volatility of the underlying stock price process. This also indicates the impact of the delay term in the diffusion portion of the process.

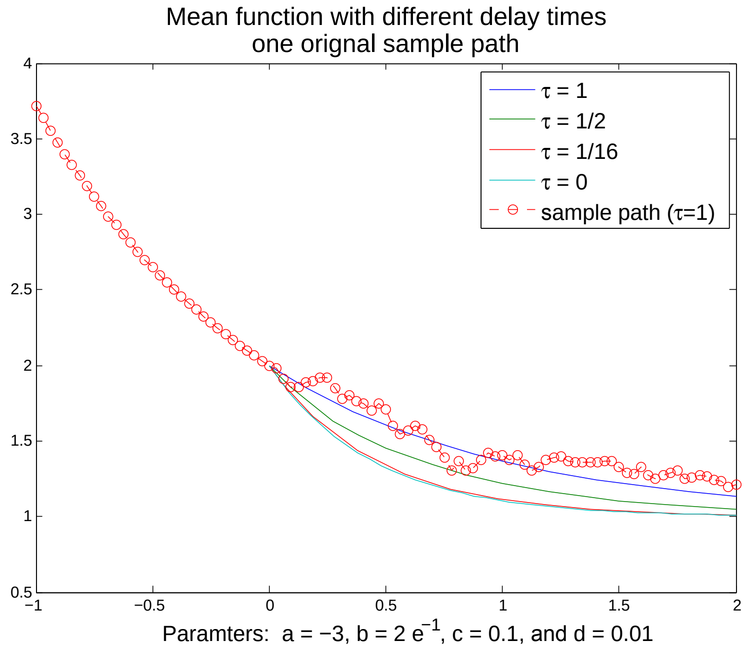


Figure 3.4: Graph of the mean function with various lengths of delay for the stochastic process in (3.4.1) with $d = 0.01$.

We now consider the impact of the delay term in the diffusion portion for simulated financial scenarios of our interest mentioned before:

- Stock price movements stay in a certain range for a certain time period, most often one to three months.
- Stock prices surge or plunge after the above stable movement stage.

In order to simulate this kind of price movements, we use a linear combination of several trigonometric functions as the initial conditions in the stochastic process (3.4.1). We are interested in the behavior of the price movement after the surge or plunge of the stock price. For example, by letting $c = d = 0.1$, we generated Figure 3.8; by letting $c = 0.1$ and $d = 0.01$, we generated Figure 3.9. Figure 3.10 was generated for two cases with $c + d = 0.2$. Similar to the previous example, the red circles represent one sample

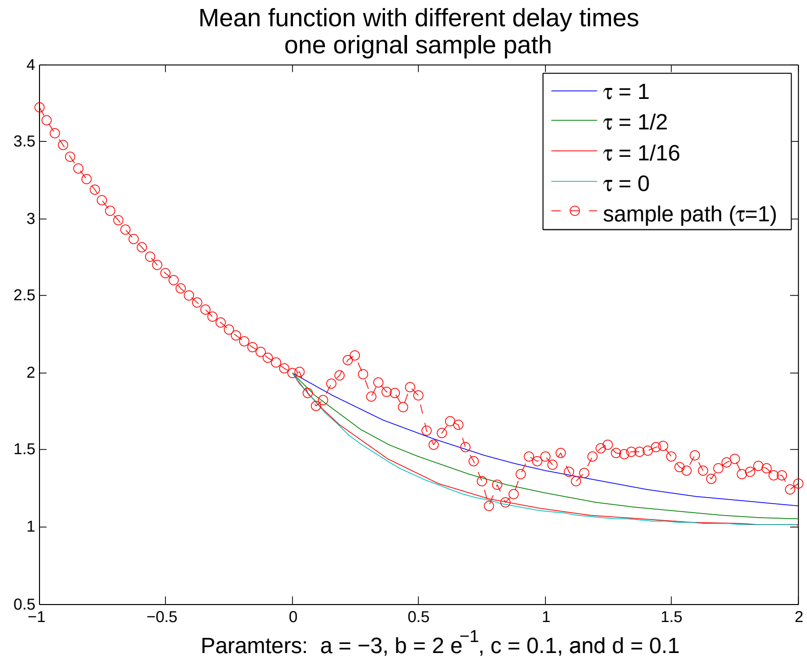


Figure 3.5: Graph of the mean function with various lengths of delay for the stochastic process in (3.4.1) with $d = 0.1$.

path, and the four colored curves represent the mean functions for the stochastic process with various lengths of delay $\tau = 1, 1/2, 1/16$, and no-delay $\tau = 0$. The light blue and red curves tend to follow a flat line, indicating that even after some significant price movements, the average stock price is expected to be relatively stable over time. On the other hand, the blue and green curves show a correction pattern, indicating that the underlying asset pricing is reacting to the significant price movement and absorbing the information from previous moves. In the first simulated financial scenario, the stock price is sinking from 1.5 to 1, about 33% decline in a short period. From the financial point of view, during such significant price plunge, usually the underlying asset could experience Overbought/Oversold. Thus, it is likely to see the market adjust the movement in the opposite direction, namely, the price will bounce back from the recent low. Both the blue and green curves exhibit such patterns supported by our financial intuitions, indicating

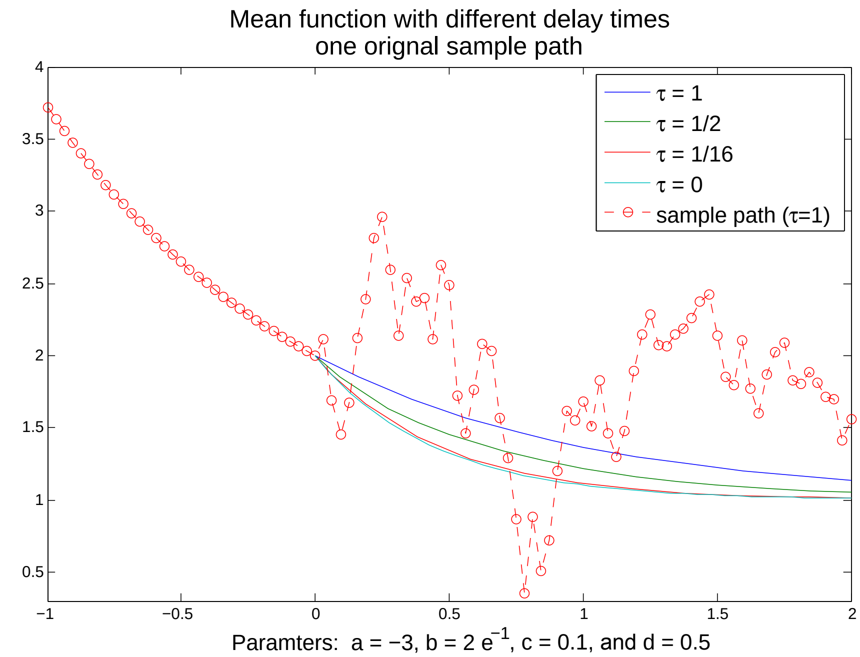


Figure 3.6: Graph of the mean function with various lengths of delay for the stochastic process in (3.4.1) with $d = 0.5$.

that in this simulated financial scenario, the stochastic process with delay terms behaves closer to reality than the no-delay process.

Furthermore, it is worth mentioning the financial meaning of Overbought/Oversold. In the financial markets, support and resistance levels are often used to determine efficient entry and exit points in the technical analysis. If the stock price is reaching a resistance level, it might be a good time to sell or sell short. If the stock price is reaching a support level, it might be a good time to cover or buy. There are many subjective decisions that an investor has to make when conducting the analysis. The Overbought/Oversold (OBOS) flag and associated quantitative indicator OBOS Index can be used to keep track of the current stock price position and determine whether and when to make any reaction to certain price movement. The above examples indicate that the stochastic process with delay terms could fit this financial phenomenon well.

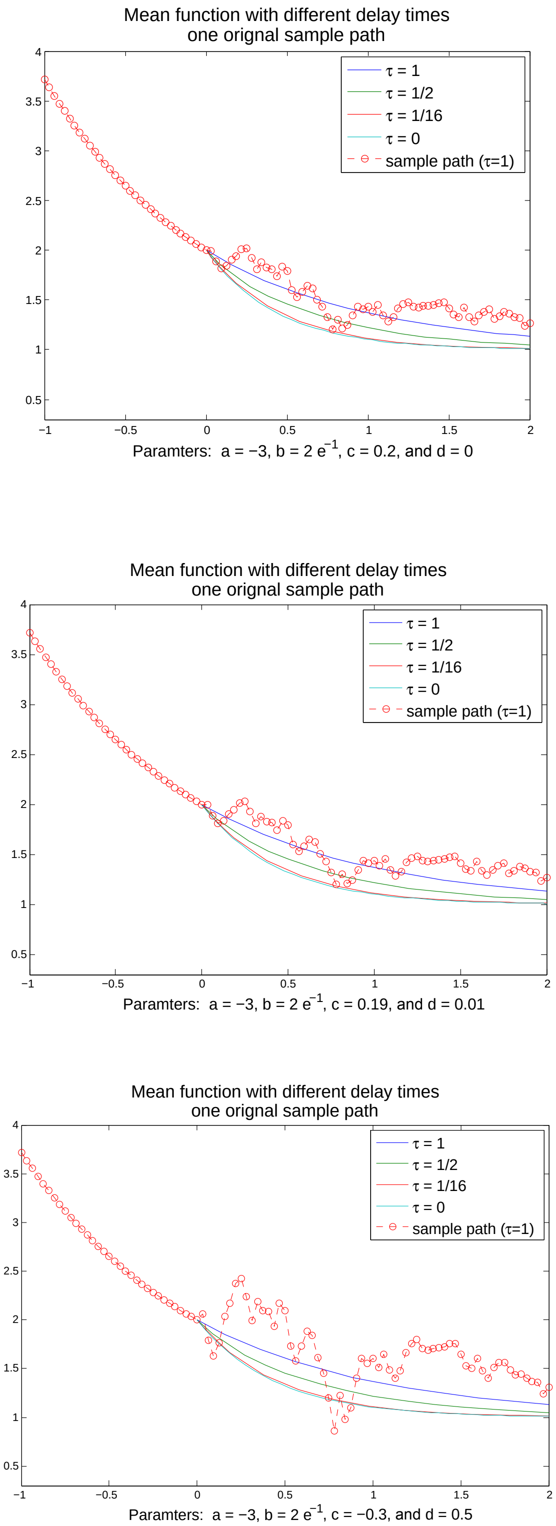
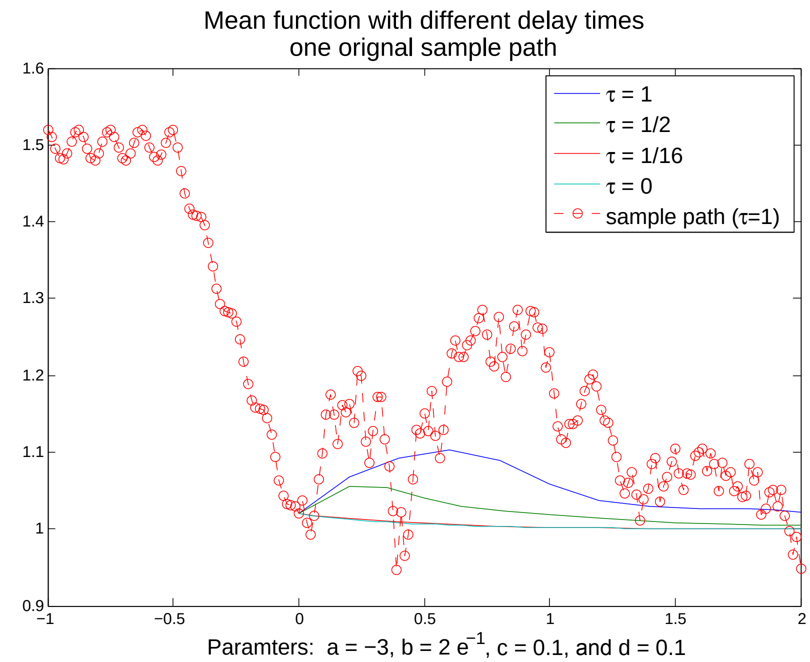


Figure 3.7: Graph of the mean function with various lengths of delay for the stochastic process in (3.4.1) with $c + d = 0.2$.



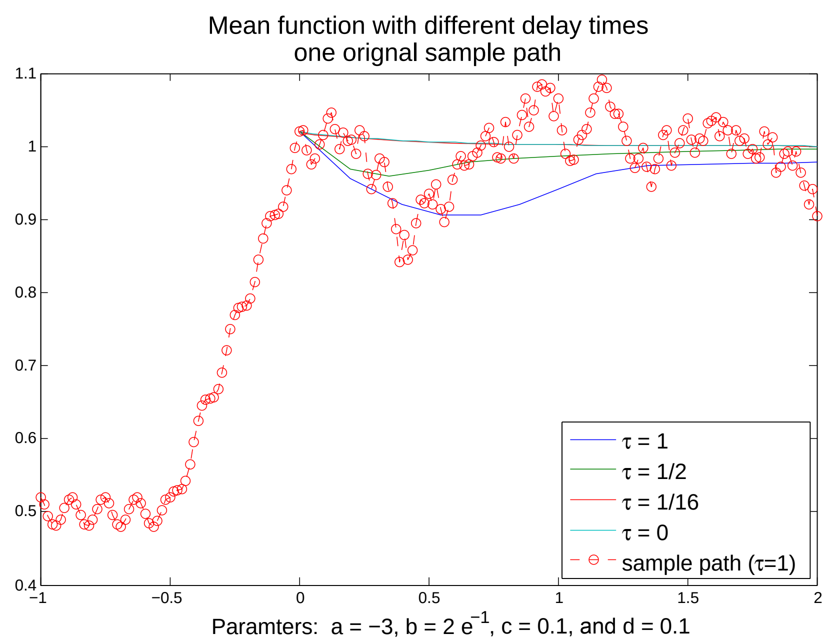
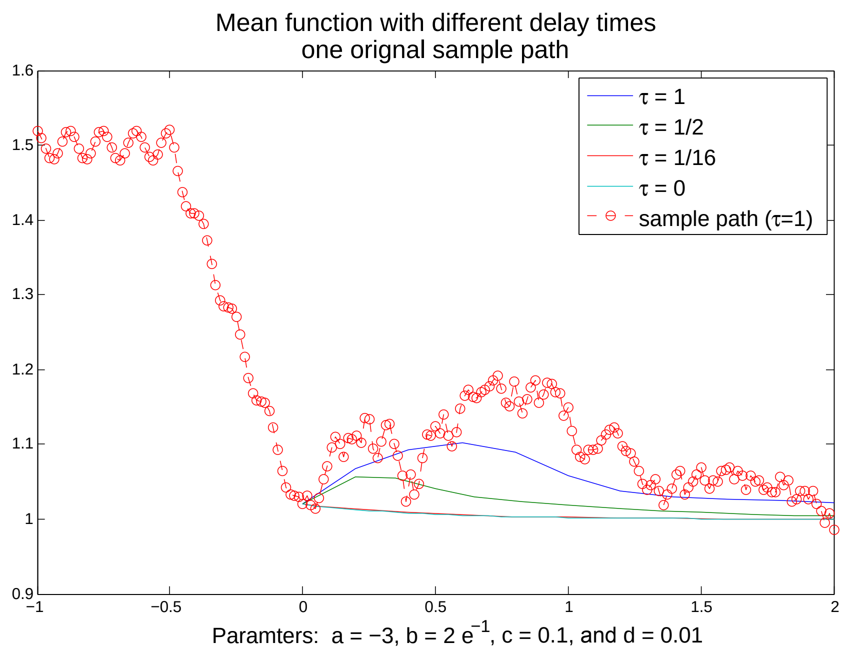


Figure 3.8: Simulated financial scenarios with $d = 0.1$.



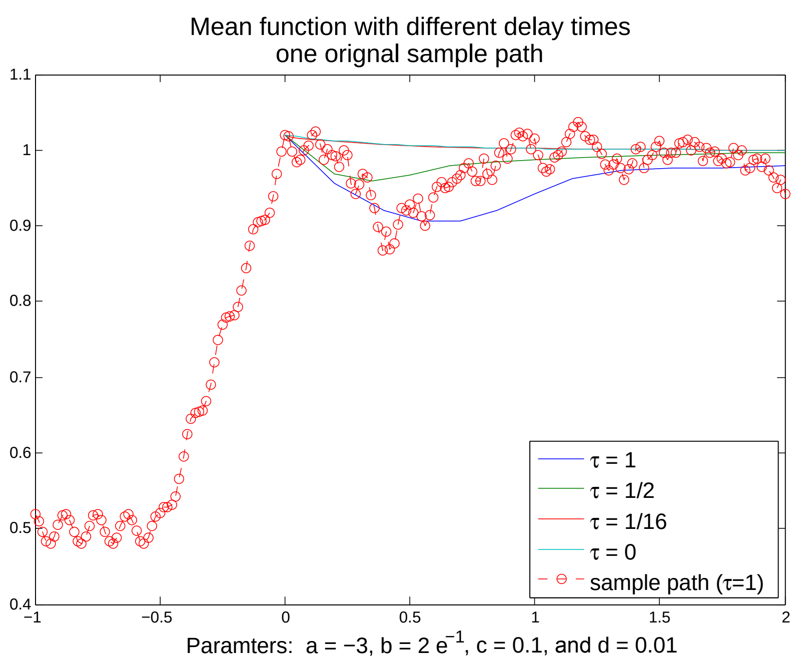


Figure 3.9: Simulated financial scenarios with $d = 0.01$.

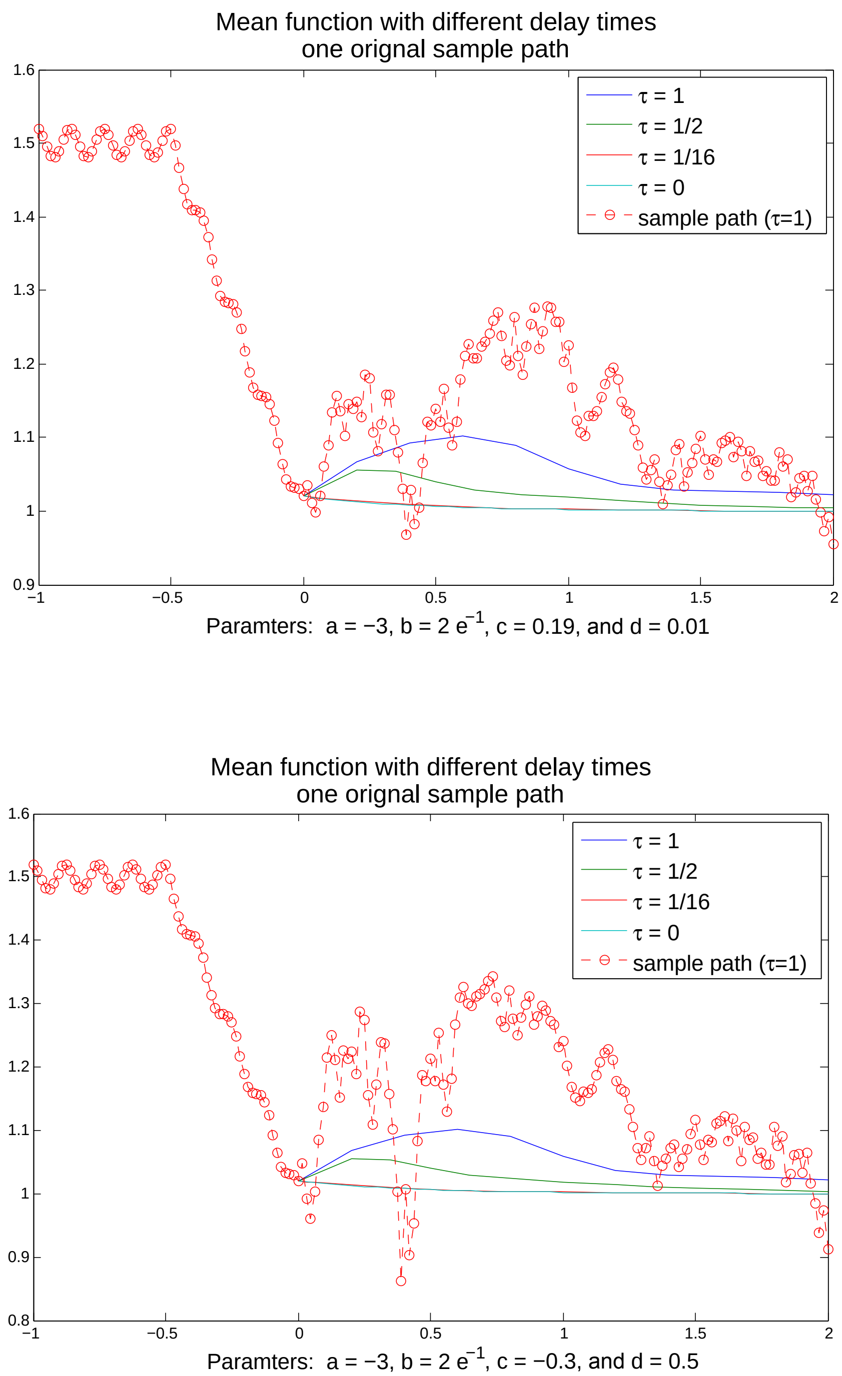


Figure 3.10: Simulated financial scenarios with $c + d = 0.2$.

CHAPTER 4. PARAMETER ESTIMATION

4.1 Introduction

The previous chapter discussed the mathematical properties of a numerical scheme to approximate the solution of a linear SDDE. This chapter aims to estimate the parameters in the SDDE and compare it with the traditional models.

Since transition densities of diffusion processes are usually unavailable in closed form, likelihood-related approaches are often difficult to implement. Early statistical approaches have used the method of moments to avoid the integration involved with the likelihood computation; see, for example, Taylor (1986) and Melino and Turnbull (1990). However, such methods depend heavily on large-sample asymptotics, making them unreliable in limited finite samples, especially when dealing with data within a certain period of time. In contrast to these approaches, Bayesian methods have been introduced by Carlin, Polson, and Stoffer (1992), Pedersen (1995), Jones (1998), and Yu and Meyer (2006). The Bayesian methods do not need large-sample approximations and the inference will be exact in finite samples. Also, stationarity is not required, although it may be imposed through the prior if desired.

4.1.1 Bayesian Inference

The purpose of Bayesian inference [Box and Tiao (1973), Bernardo and Smith (2000), and Gelman et al. (2014)] is to provide a mathematical method for modeling uncertainties and making decisions according to rational principles. The tools of this method are probability distributions and the rules of probability. The difference between the Bayesian and Frequentist statistical inferences is that in the Bayesian inference the probability of an event is the uncertainty of the event in a single trial, not the frequency of the event occurring as the number of trials goes to infinity, as in the Frequentist inference. Because models in the Bayesian inference are formulated in terms of probability distributions, the probability axioms and computation rules also apply in the Bayesian inference.

Bayes' theorem is the foundation of the Bayesian inference and named after Thomas Bayes. Mathematically, Bayes' theorem gives the relationship between the probabilities of A and B, $P(A)$ and $P(B)$, and the conditional probabilities of A given B and B given A, $P(A|B)$ and $P(B|A)$. In its most common form, this theorem states that

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}.$$

This provides an expression for the relationship between two conditional probabilities that are the reverse of each other.

In the model-based statistical inference, we usually start with two major components, data $\mathbf{Y}$ and parameter Θ, where $\mathbf{Y}$ represents the observable random vector, and the distribution of $\mathbf{Y}$ depends on the parameter Θ that is unknown or unobservable. The objective is to make plausible statements about the parameter Θ. The distribution of $\mathbf{Y}$ is specified by the probability density function (pdf)

$$f(\mathbf{y}|\Theta),$$

and this function will describe an entire family of probability models for $\mathbf{Y}$, depending on the parameter Θ. The Bayesian inference assumes that Θ is a random vector with the pdf $g(\Theta)$, called the **prior** pdf of Θ. Thus the joint pdf of $(\mathbf{Y}, \Theta)$ is given by

$$f(\mathbf{y}, \Theta) = f(\mathbf{y}|\Theta)g(\Theta) = L(\Theta)g(\Theta),$$

where

$$L(\Theta) = f(\mathbf{y}|\Theta),$$

as a function of Θ for given observed data $\mathbf{y}$, is called the likelihood function in the Frequentist inference. So the Bayesian statistical inference adds to the Frequentist model assumptions prior knowledge about the parameter Θ before data collection. The **posterior** pdf, $f(\Theta|\mathbf{y})$, is the pdf of the parameter Θ given the observed data $\mathbf{y}$. Using Bayes' theorem, we have

$$f(\Theta|\mathbf{y}) = \frac{f(\mathbf{y}|\Theta)g(\Theta)}{f(\mathbf{y})}.$$

So

$$\text{Posterior} \ \propto \ \text{Prior} \cdot \text{Likelihood.}$$

If $f(\Theta|\mathbf{y})$ and $g(\Theta)$ are in the same distribution family, the prior and posterior distributions are called conjugate distributions, and the prior is called a conjugate prior for the likelihood.

We seek the conditional distribution of the parameter of interest given the data and then make inference about the parameter. In order to achieve that, it is ideal to find an analytic form for the marginal posterior distribution for the parameters of interest and generate iid random samples from the marginal posterior distribution. However, finding the analytic form might not be feasible. Thus we use simulation strategies to sample from the posterior distribution. One of the most powerful developments in Bayesian statistics has been "Markov Chain Monte Carlo" (MCMC) simulation, which uses a suitable Markov chain whose empirical distribution of states visited for long runs of the chain approximates the posterior. Studies on MCMC simulation can be found in Casella and George (1992), Chib and Greenberg (1996), and Gelman et al. (2014). The typical simulation methods that are widely used in Bayesian computation include the following:

- Rejection sampling
- Gibbs sampling
- Slice sampling

- Metropolis-Hastings algorithm

- Metropolis-Hastings within Gibbs algorithm

Recent advances in Bayesian computation have made it possible to implement sensible Bayes solutions to statistical problems that are highly problematic when attacked from other vantage points. We will use the Metropolis-Hastings within Gibbs algorithm for parameter estimation.

4.2 Methodology

4.2.1 Model

As motivated by the previous chapters, we will use the following model to describe the financial phenomenon that we are interested in

$$\begin{cases} dX(t) = (a \cdot X(t) + b \cdot X(t-\tau))dt + (\tilde{c} \cdot X(t) + \tilde{d} \cdot X(t-\tau))dW(t), \ 0 \leq t \leq T, \\ X(t) = \phi(t), \ \tau \leq t \leq 0. \end{cases} \tag{4.2.1}$$

We will use Bayesian methods to estimate the parameters in this SDDE. To simplify the calculation, let $c = \tilde{c}$ and $d = \tilde{d}/\tilde{c}$, and rewrite the SDDE as

$$\begin{cases} dX(t) = (a \cdot X(t) + b \cdot X(t-\tau))dt + c \cdot (X(t) + d \cdot X(t-\tau))dW(t), \ 0 \leq t \leq T, \\ X(t) = \phi(t), \ \tau \leq t \leq 0. \end{cases} \tag{4.2.2}$$

The inference setup used by Bayesian methods based on time discretization is as follows:

- The stochastic process $\{X(t)\}_{t\geq 0}$ with parameter $\Theta = (a, b, c, d)$

- $\mathbf{y} = \{x_0, x_1, ..., x_N\}$, the observations of $\{X(t)\}_{t\geq 0}$ at time points $t_0, t_1, ..., t_N$

- The objective is to estimate $f(\Theta|\mathbf{y})$. If the stochastic process has the Markov property and the transition probability density of the stochastic process is known, from Bayes' theorem,

$$\begin{aligned} f(\Theta|\mathbf{y}) &\propto f(\mathbf{y}|\Theta) \cdot f(\Theta) \\ &\propto f(x_0|\Theta) \cdot \prod_{i=0}^{N-1} f(x_{i+1}|x_i, \Theta) \cdot f(\Theta). \end{aligned} \tag{4.2.3}$$

First, we apply the Euler-Maruyama scheme to approximate the SDDE (4.2.2), which can be rewritten as

$$\begin{aligned} X_{n+1} = X_n &+ (a \cdot X_n + b \cdot X_{n-N_\tau}) \cdot h \\ &+ c \cdot (X_n + d \cdot X_{n-N_\tau}) \cdot \Delta W_{n+1}, \end{aligned} \tag{4.2.4}$$

for $n = 0, 1, ..., N-1$, where $\Delta W_1, ..., \Delta W_N$ are independent $N(0, h)$ random variables, and $h = \sigma^2 = dt$. The initial condition is then described by the observations

$$(X_{-N_\tau}, X_{-N_\tau+1}, ..., X_{-1}, X_0).$$

Notice that the stochastic process governed by this SDDE does not have the Markov property, so we can not use the direct approach to find the transition density and write out the posterior pdf. Since here we are dealing with the fixed-length delay terms, we propose a blocking method to create a stochastic process with the Markov property. Let

$$\vec{Y}_n = (X_n, X_{n-1}, ..., X_{n-N_\tau})'.$$

Then the initial condition is

$$\vec{Y}_0 = (X_0, X_{-1}, ..., X_{-N_\tau})',$$

and we can show that the new stochastic process $\{\vec{Y}_n\}_{0\leq n\leq N}$ has the Markov property. Note that

$$
\begin{aligned}
f(\Theta|\vec{y}_0, \vec{y}_1, ..., \vec{y}_N) &\propto f(\vec{y}_1, \vec{y}_2, ..., \vec{y}_N|\vec{y}_0, \Theta) \cdot f(\Theta) \\
&= f(\Theta) \cdot \prod_{i=0}^{N-1} f(\vec{y}_{i+1}|\vec{y}_i, \Theta) \\
&= f(\Theta) \cdot f(x_1, x_0, ..., x_{-N_\tau+1}|x_0, ..., x_{-N_\tau+1}, x_{-N_\tau}, \Theta) \\
&\cdot f(x_2, x_1, ..., x_{-N_\tau+2}|x_1, ..., x_{-N_\tau+2}, x_{-N_\tau+1}, \Theta) \\
&\vdots \\
&\cdot f(x_N, x_{N-1}, ..., x_{N-N_\tau}|x_{N-1}, ..., x_{N-N_\tau}, x_{N-N_\tau-1}, \Theta) \qquad (4.2.5)\\
&= f(\Theta) \cdot f(x_1|x_0, ..., x_{-N_\tau+1}, x_{-N_\tau}, \Theta) \\
&\cdot f(x_2|x_1, ..., x_{-N_\tau+2}, x_{-N_\tau+1}, \Theta) \\
&\vdots \\
&\cdot f(x_N|x_{N-1}, ..., x_{N-N_\tau}, x_{N-N_\tau-1}, \Theta) \\
&= f(\Theta) \cdot f(x_1|x_{-N_\tau}, x_0, \Theta) \cdot f(x_2|x_{-N_\tau+1}, x_1, \Theta) \\
&... \cdot f(x_N|x_{N-N_\tau-1}, x_{N-1}, \Theta).
\end{aligned}
$$

From the above derivation, we have shown that the transition density of the new stochastic process $\{\vec{Y}_n\}_{0\leq n\leq N}$ satisfies the following formula

$$f(\vec{y}_n|\vec{y}_{n-1}, \Theta) = f(x_n|x_{n-1}, x_{n-N_\tau-1}, \Theta) = \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp\left(-\frac{(x_n-\mu_n)^2}{2\sigma_n^2}\right),$$

where

$$\mu_n = (ah+1)x_{n-1} + bhx_{n-N_\tau-1},$$

and

$$\sigma_n^2 = c^2(x_{n-1} + dx_{n-N_\tau-1})^2 h.$$

Thus, we can obtain the likelihood function as

$$f(\vec{y}_1, \vec{y}_2, ..., \vec{y}_n|\vec{y}_0, \Theta) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(x_i-\mu_i)^2}{2\sigma_i^2}\right), \qquad (4.2.6)$$

where $\mu_i = (ah+1)x_{i-1} + bhx_{i-N_\tau-1}$ and $\sigma_i^2 = c^2(x_{i-1} + dx_{i-N_\tau-1})^2 h$, for $i = 1, 2, ..., N$.

4.2.2 Parameters

An important step in Bayesian inference is to specify the prior distributions for the parameters. For the drift terms a and b in the SDDE, we will use the univariate normal distribution as the prior. For the volatility terms c and d, in a practical Bayesian analysis of variance component models, the most commonly used prior for a variance component is the inverse gamma distribution, which is the conjugate prior for the normal likelihood function [e.g., Gelman et al. (2014)]. However, here we have two parameters in the variance components. By setting up the likelihood function as in (4.2.6), we will choose the inverse gamma distribution as the prior for c^2.

Now we are able to find the fully conditional posterior pdf for the first three parameters in the model. For a,

$$
\begin{aligned}
f(a\,|\vec{y}_0, \vec{y}_1, ..., \vec{y}_N, b, c^2, d) &= f(a\,|x_1, x_2, ..., x_N, \vec{y}_0, b, c^2, d) \\
&\propto \exp\left\{-\sum_{i=1}^{N} \frac{[x_i - (ah+1)x_{i-1} - bhx_{i-N_\tau-1}]^2}{2\sigma_i^2}\right\} \cdot \exp\left\{-\frac{a^2}{2\sigma_a^2}\right\} \\
&\propto \exp\left\{-\sum_{i=1}^{N} \frac{[(x_i - x_{i-1} - bhx_{i-N_\tau-1}) - hx_{i-1}a]^2}{2\sigma_i^2} - \frac{a^2}{2\sigma_a^2}\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left(\frac{1}{\sigma_a^2} + \sum_{i=1}^{N} \frac{(hx_{i-1})^2}{\sigma_i^2}\right) a^2 + \right. \\
&\qquad \left. \sum_{i=1}^{N} \frac{hx_{i-1}(x_i - x_{i-1} - bhx_{i-N_\tau-1})}{\sigma_i^2} a\right\}.
\end{aligned}
\tag{4.2.7}
$$

Thus, the conditional posterior distribution of a is a univariate normal distribution with mean

$$
\left[\frac{1}{\sigma_a^2} + \sum_{i=1}^{N} \frac{(hx_{i-1})^2}{\sigma_i^2}\right]^{-1} \cdot \sum_{i=1}^{N} \frac{hx_{i-1}(x_i - x_{i-1} - bhx_{i-N_\tau-1})}{\sigma_i^2},
$$

and variance

$$
\left[\frac{1}{\sigma_a^2} + \sum_{i=1}^{N} \frac{(hx_{i-1})^2}{\sigma_i^2}\right]^{-1}.
$$

Similarly, for b,

$$
\begin{aligned}
f(b\,|\vec{y}_0, \vec{y}_1, ..., \vec{y}_N, a, c^2, d) &= f(b\,|x_1, x_2, ..., x_N, \vec{y}_0, a, c^2, d) \\
&\propto \exp\left\{-\sum_{i=1}^{N} \frac{[x_i - (ah+1)x_{i-1} - bhx_{i-N_\tau-1}]^2}{2\sigma_i^2}\right\} \cdot \exp\left\{-\frac{b^2}{2\sigma_b^2}\right\} \\
&\propto \exp\left\{-\sum_{i=1}^{N} \frac{[(x_i - x_{i-1} - ahx_{i-1}) - bhx_{i-N_\tau-1}]^2}{2\sigma_i^2} - \frac{b^2}{2\sigma_b^2}\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left(\frac{1}{\sigma_b^2} + \sum_{i=1}^{N} \frac{(hx_{i-N_\tau-1})^2}{\sigma_i^2}\right) b^2 + \right. \\
&\qquad \left. \sum_{i=1}^{N} \frac{hx_{i-N_\tau-1}(x_i - x_{i-1} - ahx_{i-1})}{\sigma_i^2} b\right\}.
\end{aligned}
\tag{4.2.8}
$$

Thus, the conditional posterior distribution of b is also a univariate normal distribution with mean

$$
\left[\frac{1}{\sigma_b^2} + \sum_{i=1}^{N} \frac{(hx_{i-N_\tau-1})^2}{\sigma_i^2}\right]^{-1} \cdot \sum_{i=1}^{N} \frac{hx_{i-N_\tau-1}(x_i - x_{i-1} - ahx_{i-1})}{\sigma_i^2},
$$

and variance

$$
\left[\frac{1}{\sigma_b^2} + \sum_{i=1}^{N} \frac{(hx_{i-N_\tau-1})^2}{\sigma_i^2}\right]^{-1}.
$$

For c^2, which has the inverse gamma distribution $IG(\alpha, \beta)$ as the prior, its pdf is

$$
f(c^2; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} (c^2)^{-\alpha-1} \exp\left(-\frac{\beta}{c^2}\right).
$$

Thus, the conditional posterior pdf of c^2 is

$$
\begin{aligned}
f(c^2|\vec{y}_0, \vec{y}_1, ..., \vec{y}_N, a, b, d) &= f(c^2|x_1, x_2, ..., x_N, \vec{y}_0, a, b, d) \\
&\propto \prod_{i=1}^{N} \frac{1}{\sqrt{c^2(x_{i-1} + dx_{i-N_\tau-1})^2}} \\
&\quad \cdot \exp\left\{-\sum_{i=1}^{N} \frac{(x_i - \mu_i)^2}{2c^2(x_{i-1} + dx_{i-N_\tau-1})^2 h}\right\} \\
&\quad \cdot (c^2)^{-\alpha-1} \exp\left(-\frac{\beta}{c^2}\right) \\
&\propto (c^2)^{-(N/2+\alpha)-1} \cdot \exp\left\{-\frac{\beta + \sum_{i=1}^{N} k_i}{c^2}\right\},
\end{aligned}
\tag{4.2.9}
$$

where

$$k_i = \frac{(x_i - (ah+1)x_{i-1} - bhx_{i-N_\tau-1})^2}{2h(x_{i-1} + dx_{i-N_\tau-1})^2}.$$

Thus, the conditional posterior distribution of c^2 is also an inverse gamma distribution

$$c^2|\vec{y}_0, \vec{y}_1, ..., \vec{y}_N, a, b, d \sim IG(N/2 + \alpha,\ \beta + \sum_{i=1}^{N} k_i).$$

In the above derivation, for computational convenience, conjugate priors are used for the first three model parameters. The full conditional posterior pdf's for the first three parameters are in closed form so we can use Gibbs sampling to generate samples. However, for the fourth parameter d, there is no closed form available for its conditional posterior pdf so the Metropolis-Hastings algorithm will be used to generate samples as it does not require sampling from any distribution defined directly, but rather only from a proposal distribution that can be selected in advance.

The Metropolis-Hastings algorithm requires the specification of a proposal distribution $J_i(d^*|d_i)$ governing the probability of d^*, which is chosen as a candidate to replace the existing ith iteration draw d_i. Notice that the proposal distribution may depend upon the iteration number and the current iterate, d_i. However, it is common practice to violate this restriction early in a run of an MCMC algorithm, letting the algorithm "adapt" for a while before beginning to save iterates as potentially representing the true underlying distribution.

Finding an appropriate proposal distribution is crucial to the success of the algorithm. The idea of tuning the algorithm early in a run is to get from the starting vector to the important part of the distribution and also to tune the parameters of the jumping distributions to make the algorithm efficient, as measured by the acceptance rate (the fraction of candidate draws that are accepted).

For the random walk Metropolis, a high acceptance rate means that most new samples occur right around the current data point, which indicates that the Markov chain is moving rather slowly and not exploring the parameter space fully. On the other hand, a low acceptance rate means that the chain is inefficient as it rejects too many candidate draws, and is hence not moving much and not mixing well. An efficient Metropolis sampler has an acceptance rate that is neither too high nor too low. In general, the acceptance rate is recommended to be between 0.2 and 0.4 for the random walk Metropolis, and close to 1 for the independent Metropolis.

Here we will let $J_i(d^*|d_i) \sim N(d_i, \sigma_d^2)$. Then we can try different variances of this proposal density to achieve good tuning results. If the variance is too small, the acceptance rate will be high. On the other hand, if the variance is too large, the acceptance rate will be low and the chain will converge slowly.

4.2.3 Sampling Algorithm

Overall, we use the Metropolis within Gibbs algorithm to draw samples and construct the Markov chain. The algorithm is, in the Gibbs sampling framework, with the "Metropolis step" substituting for the update of some particular parameters. The algorithm proceeds as follows, after initializing a_0, b_0, c_0^2, and d_0, for the $(i+1)$st step:

1. Draw $a_{i+1} \sim f(a_{i+1}|\vec{y_0}, \vec{y_1}, ..., \vec{y_N}, b_i, c_i^2, d_i)$.

2. Draw $b_{i+1} \sim f(b_{i+1}|\vec{y_0}, \vec{y_1}, ..., \vec{y_N}, a_{i+1}, c_i^2, d_i)$.

3. Draw $c_{i+1}^2 \sim f(c_{i+1}^2|\vec{y_0}, \vec{y_1}, ..., \vec{y_N}, a_{i+1}, b_{i+1}, d_i)$.

4. Draw d^* from the proposal distribution $J_i(d^*|d_i) \sim N(d_i, \sigma_d^2)$ and compute the acceptance ratio

$$r = \frac{f(d^*|x_1, x_2, ..., x_N, \vec{y_0}, a_{i+1}, b_{i+1}, c_{i+1}^2)/J_i(d^*|d_i)}{f(d_i|x_1, x_2, ..., x_N, \vec{y_0}, a_{i+1}, b_{i+1}, c_{i+1}^2)/J_i(d_i|d^*)}.$$

5. Accept d^* as the $(i+1)$st iteration d_{i+1} with probability $\min(r, 1)$. If d^* is not accepted, then $d_{i+1} = d_i$.

6. go back to step 1 and proceed the $(i+2)$nd step.

4.2.4 Case Study

We used the methodology in the previous sections to analyze the daily stock prices of Bank of America Corporation (BOA) from September 2012 to April 2013 based on both the linear SDDE (4.2.2) and the GBM model. During this time period, the stock of BOA was first traded between \$8 and \$9 per share for almost one month and then started to soar. We are interested in this kind of financial scenario and believe that historical information could affect the current return and guide investment decisions. From experience, we set the fixed delay τ as two months. Then we used Bayesian methods to estimate the parameters and compared the results of the linear SDDE (4.2.2) and the GBM

$$dX(t) = a \cdot X(t)dt + c \cdot X(t)dW(t), \ -\tau \leq t \leq T.$$

First, we assumed that the prior distributions for the parameters in the linear SDDE are $a \sim N(0, 9)$, $b \sim N(0, 9)$, $c^2 \sim IG(3, 1)$, and $d \sim N(0, 9)$. Then we performed the random walk Metropolis within Gibbs by creating Markov chains of length 10000 for each parameter. We chose the proposal density $J_i(d^*|d_i) \sim N(d_i, \sigma_d^2)$ with $\sigma_d = 0.1$ and kept the corresponding acceptance rates around 30%. After this parameter setup, we ran five chains with burn-in 1,000 iterations. The trace plots in Figure 4.1 suggest that the chains are mixing adequately. The Gelman-Rubin multiple-sequence diagnostic test [Gelman and Rubin (1992)] was run on all five chains for each parameter. All potential scale reduction factors from the Gelman-Rubin test were estimated to be less than 1.01; see Figure 4.2. This indicates that the chains for each parameter are all from the same distribution and have the run-length long enough to achieve the desired stationary distribution. Figure 4.3 shows the posterior and prior distributions for all parameters in the SDDE.

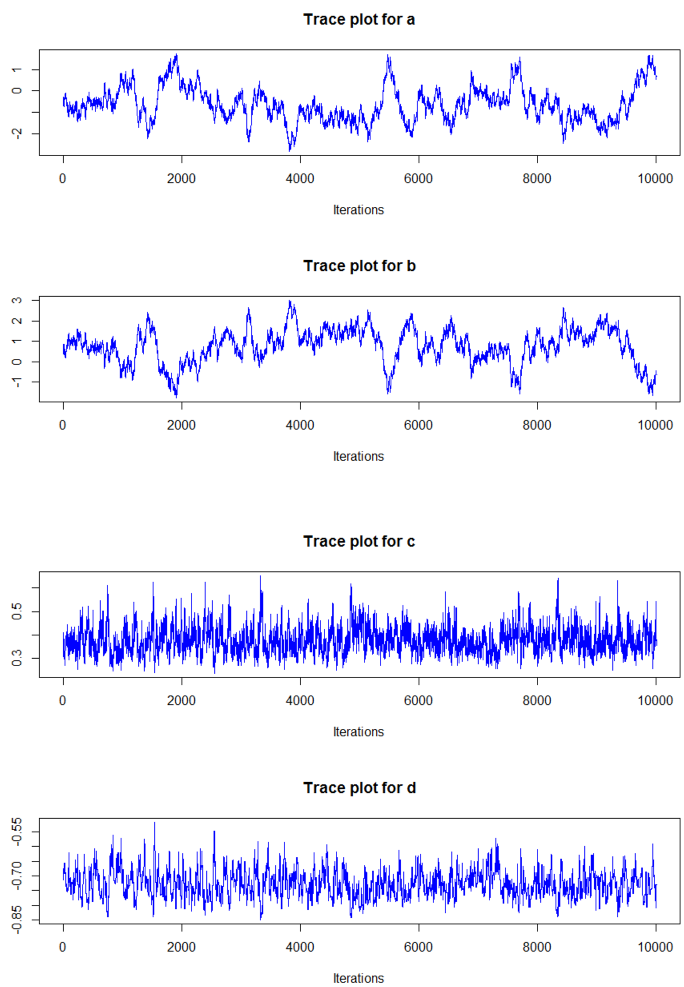


Figure 4.1: Trace plots for parameters a, b, c, and d.

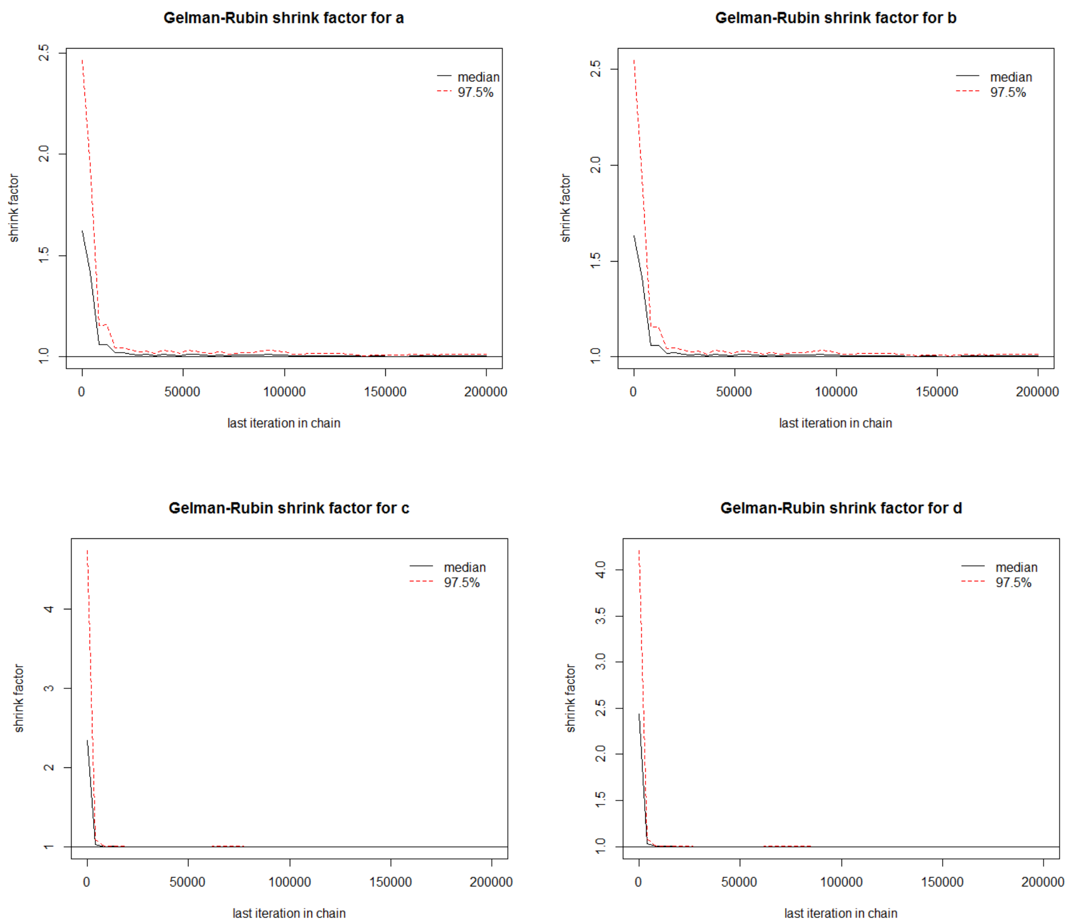


Figure 4.2: Gelman-Rubin plots for parameters a, b, c, and d.

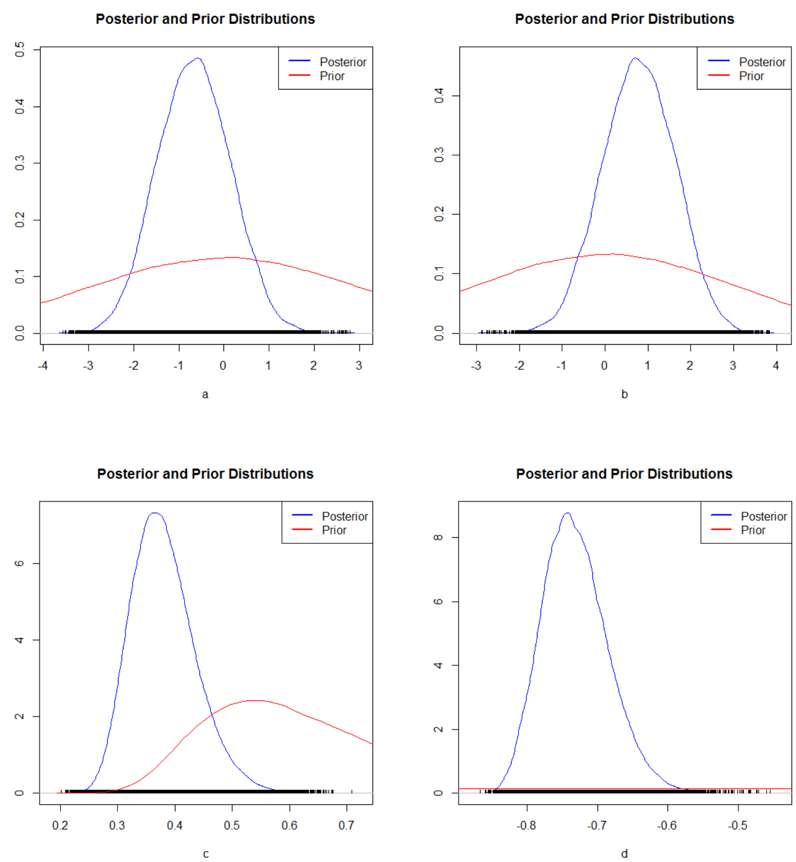


Figure 4.3: Posterior and prior distributions for parameters a, b, c, and d.

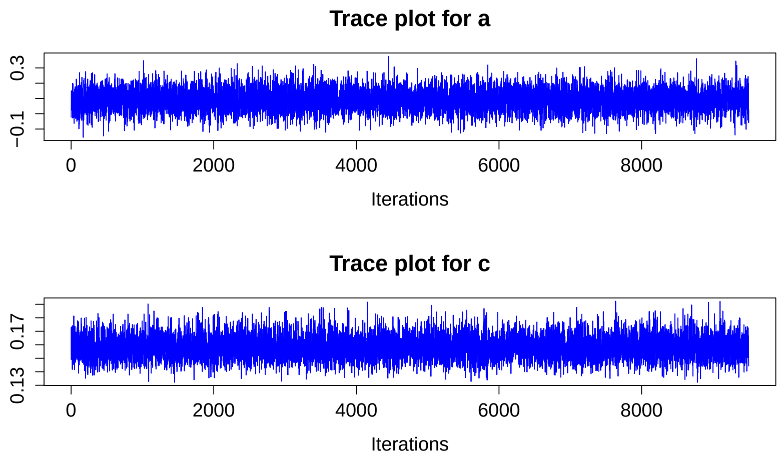


Figure 4.4: Trace plots for parameters a and c.

Notice that, if the delay parameters $b = d = 0$, the SDDE becomes the GBM. We could then employ the Gibbs sampling only to obtain the posterior distributions of the two parameters a and c. The trace plots in Figure 4.4 suggest that the chains are mixing adequately and the Gelman-Rubin multiple-sequence diagnostic in Figure 4.5 indicates that the chains for each parameter are sampled from the same distribution and have the run-length long enough to achieve the desired stationary distribution.

To compare the two models, we present the estimated parameters and the model fit statistics in Table 4.1. For each model, the first row in the table gives the point estimate of each parameter and the second row gives the corresponding standard deviation. The table also gives the DIC (deviance information criterion) value for each model. The DIC was introduced by Spiegelhalter et al. (2002) and used to compare stochastic models by Berg et al. (2004). It is a Bayesian version of the well-known Akaike information criterion (AIC), and is related also to the Bayesian (or Schwarz) information criterion (BIC). Similar to the AIC and the BIC, the DIC balances between model adequacy and model complexity. The DIC is also easy to calculate and is applicable to a wide range of statistical models. It is based on the posterior distribution of the log-likelihood or the deviance, which is defined by the following steps.

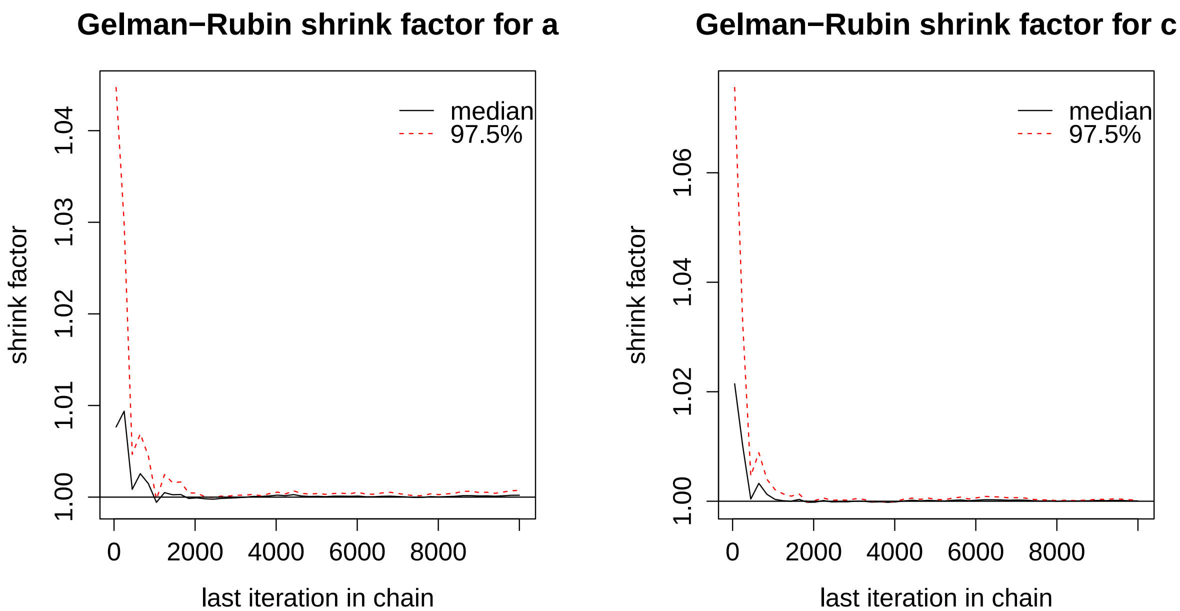


Figure 4.5: Gelman-Rubin plots for parameters a and c.

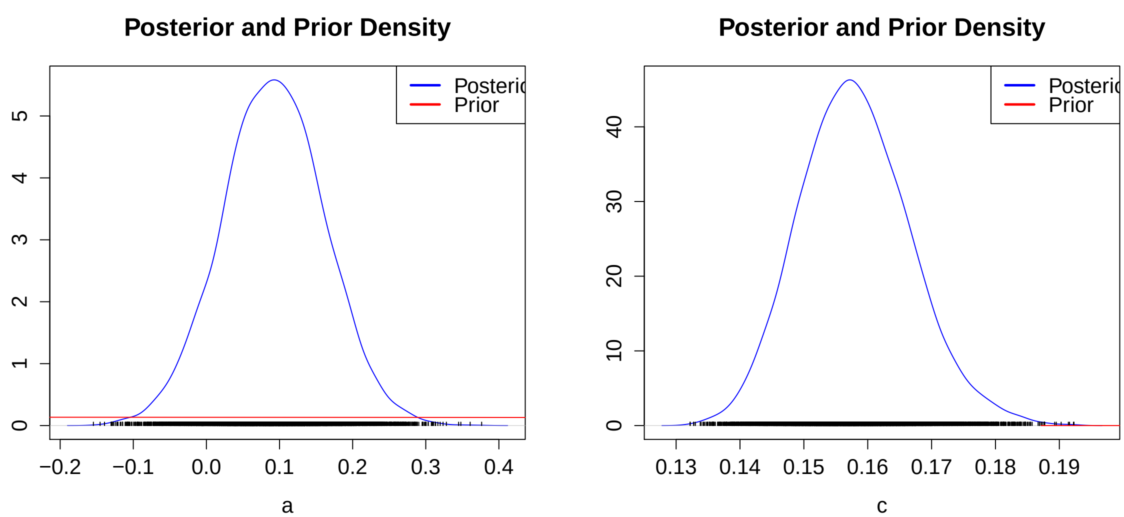


Figure 4.6: Posterior and prior distributions for parameters a and c.

The deviance is -2 times the log likelihood, that is,

$$D(\Theta) = -2\log(f(y|\Theta)).$$

Define the posterior mean of the deviance as

$$\bar{D}(\Theta) = E_{\Theta|y}\left[D(\Theta)\right],$$

and the deviance evaluated at the posterior mean of Θ as

$$D(\bar{\Theta}) = D\left(E_{\Theta|y}\left[\Theta\right]\right).$$

The DIC is then defined as

$$\text{DIC} = D(\bar{\Theta}) + 2\cdot(\bar{D}(\Theta) - D(\bar{\Theta})).$$

In this study of BOA stock price movement, based on Table 4.1, the SDDE model has a much lower DIC than the DIC of the GBM Model, which indicates that the SDDE fits the data better than the traditional GBM, as we expected.

Table 4.1: Parameter estimation and comparison

Model	a	b	c	d	DIC
SDDE	-0.7028	0.8343	0.3795	-0.730	31.52
	(0.6213)	(0.6735)	(0.0558)	(0.04914)	
GBM	0.0918	0	0.158	0	320.63
	(0.0699)	(0)	(0.00868)	(0)	

Recall that in this example the stock prices stayed low for several months and then started to move up quickly. If a stock has a large price range movement over a short-time period, it is considered highly volatile and may expose the investors to an increased risk of loss, but it also implies an increased potential for substantial short-term gains. In other words, it is a good opportunity to buy and hold the stock during the bottom-formation period and then wait for the signal of further price movement. The question we are trying to answer is, once the price starts to move, how do we track the movement and make a reasonable investment decision?

An interesting finding in this case study is that the estimated values of parameters a and b have different signs with similar absolute values, indicating that the current price has a negative impact on the return while the historical price has a positive impact on the return. So we could use the two parameters as indicators of asset allocation in certain period of time to guide the investment. When the stock stays in a relatively narrow and low price range, we should buy the stock as parameter b has more influence on the return, and sell the stock once the price moves up to certain range. To illustrate the idea and give guidance for investment, we can quantify the behavior of the expected return as the following:

$$aX(t) + bX(t-\tau) = X(t)\left(a + b\frac{X(t-\tau)}{X(t)}\right).$$

If $a + b\frac{X(t-\tau)}{X(t)} > 0$, namely, the average return is positive, then it is reasonable to hold the stock and we can keep holding it as long as the condition is satisfied. On the other hand, if $a + b\frac{X(t-\tau)}{X(t)} < 0$, the average return becomes negative, which triggers the sell signal. In our case here, the estimated parameters $a = -0.7028$ and $b = 0.8343$, so the critical value for the ratio between the present and past stock prices can be calculated by

$$\frac{X(t)}{X(t-\tau)} = -\frac{0.8343}{-0.7028} \approx 118.7\%.$$

This ratio of $\frac{X(t)}{X(t-\tau)}$ quantifies the range of price movement, and also defines the buy and sell signals in the price movement process. The ratio also answers our question: Once the price starts to move, how do we track the movement and make a reasonable investment decision? From the calculation for this case study, we can conclude that, when the price stays in the bottom and moves in a narrow price range, the ratio is relatively small, less than 119%, which indicates a positive average return, and the strategy is holding the stock. Once the price moves up quickly from the bottom, if the pull up from bottom is roughly over 20%, the strategy would be selling the stock, holding cash, and waiting for the next buy signal. This strategy does reflect some practical behaviors in the stock market. A relatively large pull up of the stock price is usually followed by a pull back, which is necessary for resuming the up trend. Our model gives a quantitative measurement of this market behavior and feasible guidance for stock trading.

BIBLIOGRAPHY

[1] Anh, V. and Inoue, A. (2005). Financial markets with memory. I: dynamic models. *Stochastic Analysis and Applications*, *23*, 275–300.

[2] Berg, A., Meyer, R., and Yu, J. (2004). Deviance information criterion for comparing stochastic volatility models. *Journal of Business and Economic Statistics*, *22*, 107–120.

[3] Bernardo, J. M. and Smith, A. F. M. (2000). *Bayesian Theory*, John Wiley and Sons, New York.

[4] Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, *81*, 637-654.

[5] Box, G. E. P. and Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis*, John Wiley and Sons, New York.

[6] Brandt, M. W. and Santa-Clara, P. (2002). Simulated likelihood estimation of diffusions with an application to exchange rate dynamics in incomplete markets. *Journal of Financial Economics*, *63*, 161–210.

[7] Buckwar, E. (2000). Introduction to the numerical analysis of stochastic delay differential equations. *Journal of Computational and Applied Mathematics*, *125*, 297–307.

[8] Carlin, B. P., Polson, N. G., and Stoffer, D. S. (1992). A Monte Carlo approach to nonnormal and nonlinear state-space modeling. *Journal of the American Statistical Association*, *87*, 493-500.

[9] Casella, G. and George, E. I. (1992). Explaining the Gibbs sampler. *American Statistician*, *46*, 167–174.

[10] Chib, S. and Greenberg, E. (1996). Markov chain Monte Carlo simulation methods in econometrics. *Econometric Theory*, *12*, 409–431.

[11] Durham, G. B. and Gallant, A. R. (2002). Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes. *Journal of Business and Economic Statistics*, *20*, 297-316.

[12] Elsanosi, I., Øksendal, B. J., and Sulem, A. (2000). Some solvable stochastic control problems with delay. *Stochastics and Stochastic Reports*, *71*, 69-89.

[13] Evans, L. C. (2013). *An Introduction to Stochastic Differential Equations*, American Mathematical Society.

[14] Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2014). *Bayesian Data Analysis*, third edition, CRC, Boca Raton.

[15] Gelman, A., and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*, 457–472.

[16] Higham, D. J. (2001). An algorithmic introduction to numerical simulation of stochastic differential equations. *SIAM Review*, *43*, 525–546.

[17] Hu, Y., Mohammed, S. E. A., and Yan, F. (2004). Discrete-time approximations of stochastic delay equations: the Milstein scheme. *Annals of Probability*, *32*, 265–314.

[18] Jones, C. S. (1998). Bayesian estimation of continuous-time finance models. Manuscript, University of Rochester.

[19] Jones, C. S. (2003). The dynamics of stochastic volatility: evidence from underlying and options markets. *Journal of Econometrics*, *116*, 181–224.

[20] Kaldor, N. (1940). A model of the trade cycle. *Economic Journal*, *50*, 78–92.

[21] Karatzas, I. and Shreve, S. E. (1991). *Brownian Motion and Stochastic Calculus*, second edition, Springer-Verlag, Berlin.

[22] Kloeden, P. E. and Platen, E. (1992). *Numerical Solution of Stochastic Differential Equations*, Springer-Verlag, Berlin.

[23] Mao, X. (1997). *Stochastic Differential Equations and Applications*, Horwood Publishing Limited, Chichester.

[24] Mao, X. (2003). Numerical solutions of stochastic functional differential equations. *LMS Journal of Computation and Mathematics*, *6*, 141–161.

[25] Melino, A. and Turnbull, S. M. (1990). Pricing foreign currency options with stochastic volatility. *Journal of Econometrics*, *45*, 239-265.

[26] Merton, R. C. (1973). Theory of rational option pricing. *Bell Journal of Economics and Management Science*, *4*, 141–183.

[27] Mohammed, S. E. A. (1984). *Stochastic Functional Differential Equations*, Research Notes in Mathematics, Pitman, London.

[28] Øksendal, B. (2013). *Stochastic Differential Equations: An Introduction with Applications*, sixth edition, Springer-Verlag, Berlin.

[29] Pedersen, A. R. (1995). A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. *Scandinavian Journal of Statistics*, *22*, 55–71.

[30] Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, *64*, 583–639.

[31] Stein, E. M. and Stein, J. C. (1991). Stock price distributions with stochastic volatility: an analytic approach. *Review of Financial Studies*, *4*, 727-752.

[32] Stoica, G. (2005). A stochastic delay financial model. *Proceedings of the American Mathematical Society*, *133*, 1837–1841.

[33] Taylor, S. J. (1986). *Modeling Financial Time Series*, John Wiley and Sons, New York.

[34] Yu, J. and Meyer, R. (2006). Multivariate stochastic volatility models: Bayesian estimation and model comparison. *Econometric Reviews*, *25*, 361–384.