Australian National University

Stochastic Differential Equations: Simulation, Parameter Estimation and Applications

Zeyi Wang

Supervised by Prof Andrew Wood
and Prof Ross Maller

July 12, 2021

A thesis submitted in partial fulfilment of the requirements for the degree of Bachelor of Actuarial Studies with Honours in Statistics at the Australian National University.

Declaration

This thesis contains no material which has been accepted for the award of any other degree or diploma in any University, and, to the best of my knowledge and belief, contains no material published or written by another person, except where due reference is made in the thesis.

Zeyi Wang

July 12, 2021

Acknowledgements

The honours year has been the most challenging and rewarding. First of all, I would like to express my sincere gratitude to my supervisors, Prof Andrew Wood and Prof Ross Maller, without whom this thesis would not have been possible. It is Prof Andrew's immense knowledge in statistical concepts and patiently and systematically guiding me on various statistical and general problems that support me throughout the research. Meanwhile, Prof Ross provided constructive advice from a big-picture perspective to ensure my research is understandable and meaningful and gave me detailed and practical suggestions for my thesis writing. It is their continuous encouragement and tireless instructions that motivate me throughout the honours. Thank you for making time every Thursday afternoon to meet and discuss the statistical and thesis problems. I have appreciated your encouragement and support during this year. I also want to thank Dr Ding Ding for providing supports in looking for the data. Special thanks to the Honours convener, Dr Bronwyn Loong, for her thoughtful arrangements and academic help during the honours year.

I would also like to thank my friends and family for their accompany, support and love along the way. My parents' encouragement and understanding supported me during the nights that I struggled for the mathematical proof. A warm word for my friend Rui and Aarfa, who showed great support and help all the way. Special thanks to Aaron, with whom I studied in the Chifley library day and night. I would also want to thank Eric, for his support in proofreading techniques. I would acknowledge the academic staff in RSFAS for their valuable comments and suggestions made during the draft review. Finally, I would like to acknowledge the financial support given by Australian National University, College of Business and Economics, Research School of Finance, Actuarial Studies and Applied Statistics (RSFAS) Honours Scholarship.

Abstract

Stochastic differential equations (SDEs), including time-homogeneous Itô diffusion processes, play an essential role in modelling phenomena in various fields, including physics, biology and finance. The parameters of the stochastic model are usually unknown in reality. Statistical inference on the unknown parameters of an Itô diffusion process has continued to attract increasing attention in the last decades. Because in general, the maximum likelihood estimation is not directly applicable to the Itô diffusion process, sue to the transition density usually not being available in closed form, an approximation to the transition density is developed. We aim to formulate a **skew-normal approximation method** motivated by the fact that the well-known Gaussian approximation method [Kessler, 1997] is inadequate in a skewed situation.

The solution of an SDE, also known as the numerical method for solving the SDE, is crucial to model various phenomena. We built a simulation scheme of the two commonly used numerical methods for a **general Itô diffusion process** across various grid widths in R. In addition to the numerical method simulation scheme, we extended the existing parameter estimation scheme [Lu et al., 2021] to the skew-normal method, and can be applied to a general Itô diffusion process.

In the practical implementation of our parameter estimation scheme, we applied the Gaussian approximation method and the skew-normal approximation method to estimate the parameters of two commonly used interest rate models, the Cox–Ingersoll–Ross model and the Vasicek model, for a 3-year Australian government bond yield data set. The accuracy is verified by simulating the sample paths of the estimated models using the numerical method simulation scheme for the general Itô diffusion processes. The Vasicek model is demonstrated to exhibit a better performance as a model for the bond yield data under parametric bootstrap hypothesis testing.

Contents

List of Figures

List of Tables

1 Introduction

A stochastic differential equation (SDE) incorporates one or more random elements in the form of a stochastic process. SDEs, particularly diffusion processes, are widely used in diverse fields, including physics, biology and mathematical finance. However, all the models consist of unknown parameters that need to be estimated from existing observations. Consider a 1-dimensional Itô diffusion process with the form:

$$dX(t) = a(X(t), \sigma)\, dt + b(X(t), \theta)\, dW(t), \quad X(0) = x_0; \tag{1.1}$$

where W is the standard Brownian motion, a and b are real-valued functions. The parameter estimation for (θ, σ) is vital in practice, but also difficult since the maximum likelihood estimation is not directly applicable to the process X in (1.1), which leads to unknown transition density and likelihood function.

This estimation problem has been studied by many researchers in the past decades and improvements of the approximation method have been of interest. Dacunha-Castelle and Florens-Zmirou [1986] and Prakasa Rao [1983] demonstrated the existence of asymptotically normal and consistent maximum likelihood estimators under certain conditions. However, they applied a direct likelihood estimation on the true likelihood, which is not practicable for computation. Another method that many researchers has considered is the approximation to the likelihood function. Aït-Sahalia [2002] provided the closed-form approximations using Hermite

polynomials, which tended to converge to the true likelihood function. A Gaussian approximation to the transition density was considered by Kessler [1997] based on previous results and assumptions, for instance, [Dacunha-Castelle and Florens-Zmirou, 1986]. The Gaussian distribution is a symmetric distribution, which is less efficient in the case of skewness. This inspired us to consider another distribution of approximation to the transition density that captures skewness in the process. The **skew-normal** (SN) distribution is analysed and the scheme of SN approximation to the transition density is constructed.

Another interesting feature is the solution of SDEs, which is widely studied as it is essential to model various phenomena. Most SDEs do not have an explicit analytical solution, and the development of numerical methods for solving SDEs is crucial in practice. A widely used approach is the simulation of the sample paths of discrete-time approximations. The commonly used numerical methods are the Euler-Maruyama method [Maruyama, 1955] and the Milstein method [Milstein, 1975]. We will provide an elementary review and focus on approximating accuracy and comparison between the two common numerical methods across various grid widths.

A number of computation packages in R have been developed for the simulation and parameter estimation of SDEs using different methods, for instance, the *sde* package with the *sde.sim* function [Iacus, 2009], the *yuima* project package [Brouste et al., 2014], and the *Sim.DiffProc* package [Guidoum and Boukhetala, 2020] with the *fitsde* function. However, there is no package available for direct approximation of the transition density, and we develop the parameter estimation scheme based on the numerical experiments by Lu et al. [2021]. the parameter estimation algorithm is modified so that it can be applied to a process with an unknown distribution. The second (partial) derivatives (Hessian matrix) of the estimated parameters are derived using *fminunc* function in MATLAB. The existing functions for the simulation of SDEs are created for specific SDE models using the Euler method and the Milstein method. We build up the numerical method simulation algorithm for **general Itô diffusion processes**.

Diffusion processes play an important role in mathematical finance, particularly with respect to some dynamics of the financial assets, such as asset prices and interest rates. To transfer our theoretical analysis to practical applications, we aim to model certain phenomena with a one-dimensional Itô diffusion process using the parameter estimation scheme and the numerical method algorithm presented in this paper. Specifically, we are interested in modelling a 3-year Australian government bond yield using 5-year data set (24/03/2014 to 25/03/2019). Two commonly used models for interest rates, the Vasicek model and the Cox–Ingersoll–Ross (CIR) model, are considered in the data analysis. We then compare the goodness of fit of these two models.

In the remainder of this chapter, we will state the background of stochastic differential equations, especially the applications in finance, in Section 1.1. Section 1.2 outlines the structure of this thesis.

1.1 Background

Stochastic differential equations, in particular, time-homogeneous Itô diffusion processes, are important in a wide variety of application areas, particularly when random effects play a significant role. Its effect in modelling different phenomena has been applied to various fields, including finance ([Merton, 1973], [Merton, 1975] and [Shreve, 2004]), physics [Papanicolaou, 1995] and biology [Fogelson, 1984]. Specifically, in the financial area, SDEs are crucial for modelling various financial quantities, such as derivative prices and risk measures. Statistical inference for the unknown parameters of the diffusion process in (1.1) is crucial in all fields. Many studies have contributed to the conditions and restrictions on the likelihood estimation, such as [Dacunha-Castelle and Florens-Zmirou, 1986] and [Yoshida, 1992]. Some other important researchers contributed to this estimation problem, such as [Genon-Catalot,

1990], [Ozaki, 1992], [Yoshida, 1992], [Kessler, 1997], [Shoji and Ozaki, 1998], [Ait-Sahalia, 2001], [Aït-Sahalia, 2002], and Sørensen [2012]. The improvement of the estimation methods is still of great interest.

SDEs can be the basis of future marketing models, where stochastic modelling and quantitative analysis are valuable tools for commercial and economic events. Numerical methods for the solutions of SDEs are required for these purposes. Two commonly used numerical methods are the Euler-Maruyama method [Maruyama, 1955] and the Milstein method [Milstein, 1975]. A simulation algorithm constructed from these numerical methods can be tailored to fit into general Itô diffusion processes with various discretisation steps. Therefore, it can be helpful to evaluate and model a real data application.

Black and Scholes [1973] and Merton [1973] introduced the price modelling of risky assets using stochastic calculus, and subsequently, the mathematical finance field incorporating the applied probability theory emerged. An important example in financial mathematics is the geometric Brownian motion (GBM), acting as the dynamics of the price of a stock in the Black–Scholes options pricing model. Though GBM acts as one of the most widely used models of stock price behaviour, it has certain drawbacks, such as the constant volatility and continuous paths. Several alternative models have been developed to cater for realistic financial needs, for instance the Merton's mixed jump-diffusion model [Merton, 1976], the variance-gamma model [Madan et al., 1998] and the exponentially weighted moving average model (EWMA) [Engle, 1982]. Other financial metrics, such as interest rate and volatility, have also attracted research interests. The Vasicek model [Vasicek, 1977] is a well-known interest rate model in finance, which captures the mean reversion feature of the interest rate. A drawback of the Vasicek model is that it allows a negative interest rate. This is fixed in many other models, such as the CIR model [Cox et al., 1985] and the Black–Karasinski model [Black and Karasinski, 1991]. Real application problems using the CIR model and the Vasicek model incorporating the numerical method and parameter estimation scheme from applied fi-

nance standpoint are of interest to us.

1.2 Thesis Outline

The specific objectives of this thesis include:

1. to develop an algorithm for simulating the solutions to an Itô diffusion process.

2. to estimate the parameter (vector) for SDE models.

3. to use SDE models to investigate a real data application.

The next chapter introduces the numerical solutions of SDEs by reviewing the studies in [Kloeden and Platen, 1992] and [Øksendal, 2003]. In particular, we introduce the SDE and time-homogeneous Itô diffusion process, as well as some general properties of SDEs, including the Itô stochastic integral and Itô's lemma. Then, we formulate a method to solve the SDE, which is a discrete-time approximation of sample paths on a fine grid. Two commonly used numerical methods, the Euler-Maruyama method and the Milstein method, are introduced and we build the simulation scheme of the solutions of the general Itô diffusion process using these two methods in R and compare their accuracy on various grid widths.

Chapter 3 focuses on the estimation of the parameters of 1-dimensional time-homogeneous Itô diffusion processes. A general method in statistical inference is the use of the maximum likelihood estimation. The difficulty of this approach is that the transition density is unknown for most diffusion processes. We define the likelihood estimation equation with some assumptions, including the Markov property and ergodicity. Then, we show and verify a higher-order Itô-Taylor stochastic expansion result and review some previous literature results, particularly a Gaussian approximation method introduced by Kessler [1997]. We consider the skew-normal approximation to the transition density based on previous analyses and assumptions; this takes into account the skewness of the process. Finally, the parameter es-

timation scheme is constructed in MATLAB based on the programming result in [Lu et al., 2021], and the skew-normal method and the Kessler method are compared using the median absolute difference and root mean square error (RMSE).

In Chapter 4, we focus on a real data application and fit a proper model using the parameter estimation scheme developed in Chapter 3. Our interest is in a 3-year Australian government bond yield data set from 24/03/2014 to 25/03/2019, and we fit the data to both the Vasicek model and the CIR model. The two estimated models are compared by simulating some sample paths using the simulation scheme developed in Chapter 2. Then, we conduct a parametric bootstrap hypothesis test to select the best model from these two models.

Finally, Chapter 5 presents a summary of our derivations and results, as well as some recommendations for further research.

2 Review of stochastic calculus

This chapter provides a brief review of stochastic calculus concepts needed for our analysis. In the first four sections, we review stochastic differential equations and Itô diffusion processes, as well as the stochastic Itô integral and Itô formula. Some definitions from [Øksendal, 2003] and [Kloeden and Platen, 1992] are presented and proved. Then, we introduce two basic interest rate models taking the form of stochastic differential equations, which will be continually revisited for the rest of this paper. In Section 2.6, we introduce the numerical method for SDEs. Specifically, we provide a derivation of the Itô-Taylor stochastic expansion result as referenced from [Kloeden and Platen, 1992]. Two commonly used numerical methods, the Euler-Maruyama method [Maruyama, 1955] and the Milstein method [Milstein, 1975] are introduced. Consequently, we construct a simulation scheme in R based on these two numerical methods for a general Itô diffusion process. Finally, we compare the performance of the two methods on various grid widths.

2.1 Stochastic Differential Equations

In this section, we formally introduce the stochastic differential equations and give definitions to 1-dimensional Itô diffusion process.

Two distinct types of differential equations can be generated by adding random terms into a

deterministic differential equation. One type is obtained when an ordinary differential equation has a random initial value, $X(0)$. This type of process has the sample paths that are differentiable functions if the original model is sufficiently smooth. This process is called a random differential equation:

$$\frac{dX(t)}{dt} = a(t, X(t)), \quad t \geq 0,\ X(0) = X_0; \tag{2.1}$$

where $X(t)$ is a stochastic process with a random initial condition X_0, and $a(.)$ is a real valued function satisfied certain conditions that will be defined later.

Another type is known as a stochastic differential equation (SDE), which has one or more terms that are stochastic processes. We consider processes of the form

$$X(t) = X(0) + \int_0^t a(s, X(s))\mathrm{d}s + \int_0^t b(s, X(s))\,\mathrm{d}W(s), \tag{2.2}$$

with a given initial value $X(0) = x_0$. A short-hand form of (2.2) can be written as the SDE:

$$dX(t) = a(t, X(t))\,dt + b(t, X(t))\,dW(t). \tag{2.3}$$

The randomness of the stochastic differential equation is introduced via a noise term, $W = (W(t), t \geq 0)$, which denotes a standard Brownian motion. The coefficient $a(t, X(t))$ is a drift term and $b(t, X(t))$ is a diffusion term, and the formal definitions are given later on.

The randomness of the first type of differential equation (2.1) is extrinsic to the dynamics of the process, in which the only randomness is introduced by the initial value $X(0)$. The extrinsic here indicates that once the initial point $X(0) = X_0$ is known, the solution of the differential equation is determined. In other words, the behaviour of process $(X(t), t \geq 0)$ is then deterministic. However, the randomness of the SDE (2.2) is intrinsic to the dynamics in the sense that the random term $W(t)$ in the differential equation gives rise to the randomness

included in any solution of SDEs. Unlike random differential equations with inherit differentiability, the solutions of SDEs driven by Brownian motion generally do not have differentiable sample paths.

The one-dimensional time-homogeneous Itô diffusion process is a solution to a specific type of SDE, where its drift and diffusion terms depend only on the process X, but not on a specific time t. First, we define the standard Brownian motion (SBM) and introduce some of its nice features.

Definition 2.1 (1-dimensional Standard Brownian Motion). A standard Brownian motion (or a standard Wiener process) is a stochastic process $(W(t), t \geq 0)$, defined on a common probability space $(\Omega, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ with the following properties:

(1) $W(0) = 0$.

(2) W has stationary and independent increments: the term independent increments means that for every $t > 0$, the future increments $W(t+u) - W(u)$, $u \geq 0$, are independent of the past values $W(s)$, $s \leq t$; the term stationary increments means that for any $0 < u, t < \infty$, the distribution of the increment $W(t+u) - W(u)$ has the same distribution as $W(t) - W(0) = W(t)$.

(3) W has Gaussian increments: $W(t+u) - W(u)$ is normally distributed with mean of 0 and variance of u, thus $W(t+u) - W(u) \sim \mathcal{N}(0, t)$.

(4) W has continuous paths: $W(t)$ is continuous in t.

Definition 2.2 (Time-homogeneous Itô Diffusion). Let $W(t)$ be a 1-dimensional standard Brownian motion on the probability space $(\Omega, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$, defined in Definition 2.1. The σ-field $\mathcal{F}_t$ is generated by $W(s)$ for $s \leq t$. Then a 1-dimensional time-homogeneous Itô

diffusion process $X(t) = (X(t, \omega), t \geq 0)$ on $(\Omega, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ takes the form

$$X(t) = X(0) + \int_0^t a\left(X(s)\right)\, ds + \int_0^t b\left(X(s)\right)\, dW(s), \tag{2.4}$$

where the given initial condition $X(0) = x_0$, $a(X(t))$ and $b(X(t))$ are the drift and diffusion terms, respectively, and $a : \mathbb{R} \to \mathbb{R}$, $b : \mathbb{R} \to \mathbb{R}$, satisfying the usual Lipschitz continuity condition:

$$|a(x) - a(y)| + |b(x) - b(y)| \leq C|x - y|, \quad \text{all } x, y \in \mathbb{R}; \tag{2.5}$$

for some constant C. Here, $\omega \in \Omega$ is an element of the sample space. All random variables under consideration may be thought of as functions of ω, and when convenient we shall make the dependence on ω explicit.

The drift and diffusion terms in (2.4) depend only on the state of the process, $X(t)$, which indicates the time-homogeneous property of the diffusion process. In other words, the transition densities depend on the time differences rather than the exact values of time. This property will be revisited in Chapter 3 when we discuss the likelihood function. Equation (2.4) has a short-hand differential form given as

$$dX(t) = a(X(t))\, dt + b(X(t))\, dW(t). \tag{2.6}$$

An important example of SDEs is the geometric Brownian motion (GBM), which can be solved explicitly.

Example 2.3 (Geometric Brownian Motion). The process $X(t)$ is called a Geometric Brownian Motion if it satisfies the following SDE:

$$dX(t) = aX(t)\, dt + bX(t)\, dW(t), \tag{2.7}$$

where a and b are constants, W is the SBM (Definition 2.1).

The equation (2.7) can be solved explicitly:

$$X(t) = X(0)\, e^{(a-\frac{1}{2}b^2)\,t+b\,W(t)},$$

where $X(0) = x_0$ is taken to be independent of $(W(t), t \geq 0)$.

However, most SDEs do not have an explicit solution formula and numerical methods are needed to approximate the solutions.

2.2 Itô Stochastic Integral

The standard Brownian motion $W(t)$ is almost surely nowhere differentiable with respect to time (proved by Kloeden and Platen [1992, Chapter. 2, Section. 4]); therefore, calculating the integral with respect to $W(t)$ requires another type of integration procedure known as the Itô integral. Øksendal [2003] provided the detailed conditions and a rigorous proof of the Itô integral, and we refer to some basic definitions to construct the Itô integral.

Definition 2.4 (Elementary Functions). A function ϕ on probability space $(\Omega, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ from the class of measurable functions is said to be elementary if it satisfies

$$\phi(t,\omega) = \sum_{i=1}^{m} g_i(\omega)\, \mathbf{I}_{[t_i,t_{i+1})}(t), \tag{2.8}$$

where $\mathbf{I}$ is the indicator function, for $i = 1, 2, \cdots m$, and each function g_i depends only on the information up to time t_i, in other words, g_i is $\mathcal{F}_{t_i}$ measurable.

Lemma 2.5 (Øksendal [2003], Lemma 3.1.5 , the Itô Isometry).

If $\phi(t,\omega)$ is bounded and elementary then

$$E\left[\left(\int_S^T \phi(t,\omega)\, dW(t,\omega)\right)^2\right] = E\left[\int_S^T \phi(t,\omega)^2 dt\right], \tag{2.9}$$

where $W(t,\omega)$ is a standard Brownian motion on probability space $(\Omega, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$, defined in Definition 2.1 .

Proof of Lemma 2.5. First, from equation (2.8), we can evaluate

$$\begin{aligned}\phi(t,\omega)^2 &= \sum_{i=1}^{m} \left(g_i(\omega)\right)^2 \left(\mathbf{I}_{[t_i,t_{i+1})}(t)\right)^2 \\ &= \sum_{i=1}^{m} \left(g_i(\omega)\right)^2 \mathbf{I}_{[t_i,t_{i+1})}(t).\end{aligned} \tag{2.10}$$

Then, we prove the equality in (2.9) from the left hand side:

$$\begin{aligned}&E\left[\left(\int_S^T \phi(t,\omega)\, dW(t,\omega)\right)^2\right] \\ =&E\left[\left(\sum_{i=1}^{m} g_i(\omega)\Big[W(t_{i+1},\omega) - W(t_i,\omega)\Big]\right)^2\right] \quad \text{(given equation (2.10))} \\ =&E\left[\sum_{i,j=1}^{m} g_i(\omega)\, g_j(\omega)\, \widetilde{\Delta}W(i)\, \widetilde{\Delta}W(j)\right] \quad (\widetilde{\Delta}W(i) = W(t_{i+1},\omega) - W(t_i,\omega)) \\ =&\sum_{i,j=1}^{m} E\Big[g_i(\omega) g_j(\omega) \widetilde{\Delta}W(i) \widetilde{\Delta}W(j)\Big] \quad \text{(Fubini's theorem)}.\end{aligned} \tag{2.11}$$

We evaluate the equation (2.11) by dividing the summation with respect to the relationship of sizes between i and j.

First recall that the increment of SBM $\widetilde{\Delta}W(i)$ is normally distributed:

$$\widetilde{\Delta}W(i) = W(t_{i+1},\omega) - W(t_i,\omega) \sim N(0, \widetilde{\Delta}t_i), \tag{2.12}$$

where $\widetilde{\Delta}t_i = t_{i+1} - t_i$, when $i < j$, the terms $g_i(\omega)g_j(\omega)\widetilde{\Delta}W(i)$ and $\widetilde{\Delta}W(j)$ in (2.11) are independent. So the expectation term equals zero in this case. Similar logic, when $i > j$, gives that the terms $g_i(\omega)g_j(\omega)\widetilde{\Delta}W(j)$ and $\widetilde{\Delta}W(i)$ are independent, and again the expectation term equals zero. Thus, we only need to focus on the case $i = j$ of (2.11).

Given Definition 2.1, the increment of SBM $\widetilde{\Delta}W(i)$ possesses independent increments, and hence it depends only on the information at $\widetilde{\Delta}t_i = t_{i+1} - t_i$ and independent of the past values $W(t_i)$. Moreover, from Definition 2.4, the function g_i depends only on the information up to time t_i. Therefore, we show that $\widetilde{\Delta}W_i$ and g_i are independent. Consequently, continue on equation (2.11), equation (2.9) is evaluated as follows:

$$
\begin{aligned}
& E\left[\left(\int_S^T \phi(t,\omega)\, dW(t,\omega)\right)^2\right] \\
=& \sum_{i=1}^{m} E\big[g_i(\omega)^2(\widetilde{\Delta}W(i))^2\big] \\
=& \sum_{i=1}^{m} E\left[g_i(\omega)^2\right]\widetilde{\Delta}t_i \qquad \Big(E\big[\big(\widetilde{\Delta}W(i)\big)^2\big] = Var\big(\widetilde{\Delta}W(i)\big) = \widetilde{\Delta}t_i\Big) \\
=& E\left[\int_S^T \phi(t,\omega)^2 dt\right] \quad \text{(Fubini's theorem and equation (2.10))}. \qquad \square
\end{aligned}
$$

Given the property of Itô isometry, we can extend the definition of the Itô integral of an elementary function ϕ to a function $f(t,\omega)$, which is $\mathcal{F}_t$ measurable, and satisfies

$$
E\left[\int_s^T f(t,\omega)^2 dt\right] < \infty. \tag{2.13}
$$

Other assumptions related to the measure space of f are defined in [Øksendal, 2003, Definition. 3.1.4]. We will now introduce an Itô integral for a general function f:

Definition 2.6 (Øksendal [2003], Definition 3.1.6, the Itô Integral).

For a function f over a time interval $[S, T]$, we define the Itô integral :

$$I(f)(\omega) = \int_t^T f(s,\omega)\, dW(s,\omega) \tag{2.14}$$

$$= \lim_{n\to\infty} \int_t^T \phi_n(s,\omega)\, dW(s,\omega), \tag{2.15}$$

where $\{\phi_n\}$ is a sequence of elementary functions such that as $n \to \infty$,

$$E\left[\int_S^T \left(f(t,\omega) - \phi_n(t,\omega)\right)^2 dt\right] \to 0. \tag{2.16}$$

Remark 2.7.

1. The existence of a sequence of elementary functions $\{\phi_n\}$ satisfying Definition 2.4 and (2.16) is proved by Øksendal [2003].

2. The existence of the limit in (2.14) as an element of L^2 is guaranteed by the Itô isometry (Lemma 2.5), where elements in L^2 are the limits of the sequences $\{\phi_n\}$ such that $\sum_{n=1}^{\infty} |\phi_n|^2 < \infty$.

2.3 Ito's Lemma

In the previous section, we introduced the Itô stochastic integral and its definition. However, analogous to the classical calculus, which uses other techniques such as the chain rule to evaluate the integral rather than the general definition, we also need additional results to handle the stochastic integrals. An Itô's Lemma (or Itô formula) is the stochastic version of the chain rule.

Theorem 2.8 (1-dimensional Itô Formula). *Let $W(t,\omega)$ be the 1-dimensional standard Brownian motion on probability space $(\Omega, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ (Definition 2.1). $X(t,\omega)$ is a 1-*

dimensional stochastic process on $(\Omega, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ *taking the form*

$$X(t,\omega) = X(0,\omega) + \int_0^t a(s,\omega)\, ds + \int_0^t b(s,\omega)\, dW(s,\omega). \tag{2.17}$$

Consider $F(t,x) = g(t,x)$ *for* $0 \leq t \leq T$ *where* g *is a twice continuously differentiable function on* $[0,\infty) \times \mathbb{R}$. *Let* $a := a(t,\omega)$ *and* $b := b(t,\omega)$, *then we have*

$$\begin{aligned} dF = &\left(\frac{\partial}{\partial t}\, g(t, X(t,\omega)) + a\, \frac{\partial}{\partial x}\, g(t, X(t,\omega)) + \frac{1}{2}\, b^2\, \frac{\partial^2}{\partial x^2}\, g(t, X(t,\omega)) \right) dt \\ &+ b\, \frac{\partial}{\partial x}\, g(t, X(t,\omega))\, dW(t,\omega), \end{aligned} \tag{2.18}$$

where $\dfrac{\partial}{\partial t}$, $\dfrac{\partial}{\partial x}$ *and* $\dfrac{\partial^2}{\partial x^2}$ *are partial derivatives.*

The proof of Theorem 2.8 is provided in [Kloeden and Platen, 1992, Theorem. 3.3.2]. We will show an example of the use of the Itô formula to solve an Itô integral.

Example 2.9. Let $W(t)$ be the standard Brownian motion with $W(0) = 0$, then

$$\int_0^t W(s)\, dW(s) = \frac{1}{2} W(t)^2 - \frac{1}{2} t. \tag{2.19}$$

Proof of Example 2.9. Let $F(t,x) = g(t,x) = x^2$ and apply Itô formula on $F(t,x)$ for $x = W(t)$. Then, given the SDE equation (2.3), the drift and diffusion terms in our case are $a\big(t, W(t)\big) = 0$ and $b\big(t, W(t)\big) = 1$, and hence, we apply equation (2.18) and obtain

$$dF = dx^2 = dW(t)^2 = dt + 2W(t)\, dW(t), \tag{2.20}$$

or equivalently,

$$W(t)^2 = t + 2 \int_0^t W(s)\, dW(s). \tag{2.21}$$

Therefore, we evaluate the Itô integral

$$\int_0^t W(s)\,dW(s) = \frac{1}{2}W(t)^2 - \frac{1}{2}t. \qquad \square$$

2.4 Infinitesimal Generator

In this section, we introduce a second order partial differential operator $\mathcal{L}$ of an Itô diffusion process $X(t)$, known as the generator of the process $X(t)$. A general definition of generator $\mathcal{L}$ of a n-dimensional Itô diffusion process is given in [Øksendal, 2003]. We will focus on a 1-dimensional Itô diffusion process and define the generator as follows:

Definition 2.10 (Øksendal [2003], Definition 7.3.1, the Infinitesimal Generator). Let $X(t)$ be an Itô diffusion process in $\mathbb{R}$ taking the form in equation (2.4). The infinitesimal generator $\mathcal{L}$ of the process $X(t)$ is,

$$\mathcal{L}f(x) = \lim_{t \downarrow 0} \frac{E^x[f(X(t))] - f(x)}{t}; \quad x \in \mathbb{R}, \tag{2.22}$$

where f is a set of functions for which the limit exists at x for all $x \in \mathbb{R}$.

We will show a more applicable result of the generator $\mathcal{L}$ for a 1-dimensional Itô diffusion process, and the version for a n-dimensional Itô diffusion is derived in [Øksendal, 2003].

Definition 2.11 (Øksendal [2003], Theorem 7.3.3, Generator of Itô diffusion). Assume an Itô diffusion process $X(t)$, defined in (2.4), for a twice continuously differentiable function f for which the limit exists at all $x \in \mathbb{R}$, then the infinitesimal generator $\mathcal{L}$ is given by

$$\mathcal{L}f(x) := a(x)\,\partial_x f(x) + \frac{1}{2}b(x)^2\,\partial_x^2 f(x), \tag{2.23}$$

where $\partial_x := \frac{\partial}{\partial x}$ and $\partial_x^2 := \frac{\partial^2}{\partial x^2}$ are partial derivatives.

Suppose f is a function with k continuous derivatives, that is $f \in C^k$. The kth iterate of $\mathcal{L}$ is calculated as:

$$\mathcal{L}^k(f) = \mathcal{L}(\mathcal{L}^{k-1}f) \text{ and } \mathcal{L}^0(f) = \mathbb{1}, \tag{2.24}$$

where $\mathbb{1}$ is the identity function.

2.5 Some Stochastic Differential Equation Models

Many models derived from stochastic differential equations are widely used in physics and financial fields. In this section, we introduce two models (the Vasicek model and the Cox-Ingersoll-Ross model) that satisfy the 1-dimensional Itô diffusion process (2.6):

$$dX(t) = a(X(t))\,dt + b(X(t))\,dW(t), \tag{2.25}$$

where $W(t)$ is the standard Brownian motion. Both models are well-known interest rate models and have inherently good properties in modelling financial quantities of interests.

2.5.1 Vasicek Model

The Vasicek model [Vasicek, 1977] specifies that the interest rate $r(t)$ follows the stochastic differential equation

$$dr(t) = \theta(\mu - r(t))\,dt + \sigma\,dW(t), \tag{2.26}$$

where $W(t)$ is the standard Brownian motion, $r(0) = r_0$, and θ, μ, σ, r_0 are positive constants.

The drift term $\theta(\mu - r(t))$ allows the interest rate process to exhibit mean reversion property, which is an essential characteristic of real interest rates in the financial market. The parameter μ is the long term mean and θ is the speed of mean reversion. The diffusion term σ represents

the instantaneous volatility of the interest rate.

The limitation of the Vasicek model is that it allows the interest rate to be negative, and this problem is addressed in the Cox-Ingersoll-Ross model.

2.5.2 Cox-Ingersoll-Ross (CIR) Model

The Cox-Ingersoll-Ross model was first introduced by Cox et al. [1985] as an extension to the Vasicek model. It specifies that the instantaneous interest rate $r(t)$ follows the stochastic differential equation:

$$dr(t) = \theta(\mu - r(t))\, dt + \sigma\, \sqrt{r(t)}\, dW(t), \tag{2.27}$$

where $W(t)$ is the standard Brownian motion, $r(0) = r_0$, and θ, μ, σ, r_0 are positive constants.

The drift term of the CIR model is exactly the same with the Vasicek model, which means the CIR model also has an inherently mean reversion property with a long run mean μ and speed of mean reversion θ. The diffusion term $\sigma\sqrt{r(t)}$ guarantees the interest rate to be non-negative, which eliminates the main drawback of the Vasicek model.

2.6 Numerical Methods for SDEs

An effective way to approximate the solution of a SDE is to approximate sample paths of discrete-time approximations on a fine grid. In this procedure, a discrete-time difference equation is analysed in place of the continuous-time differential equation. This requires step by step generation of approximated values of the sample paths of a process at each discretisation time, which can be realised with the Itô-Taylor stochastic expansion. Furthermore, we introduced

two widely used numerical methods, the Euler-Maruyama method [Maruyama, 1955] and the Milstein method [Milstein, 1975], differing in the number of expansion terms. We analysed the accuracy of these two numerical methods by developing the corresponding simulation schemes on general Itô diffusion processes across various discretisation steps.

2.6.1 Itô-Taylor Stochastic Expansion

The deterministic Taylor formula provides a method to approximate a sufficiently smooth function in a neighbourhood of a given point to any desired order of accuracy. Similar to the deterministic Taylor formula that acts as a helpful tool in evaluating ordinary calculus, the Itô-Taylor stochastic expansion is universally used in stochastic calculus, which was first introduced in [Wagner and Platen, 1978]. The truncated Itô-Taylor stochastic expansion after specific orders can be used to create a discrete-time Taylor approximation.

Our analysis is restricted to a 1-dimensional Itô diffusion process $X(t)$ introduced in Definition 2.2 and satisfying

$$dX(t) = a(X(t))\,dt + b(X(t))\,dW(t), \tag{2.28}$$

where the solution $X(t)$ exists if $a(X(t))$ and $b(X(t))$ are sufficiently smooth real-valued functions satisfying a linear growth bound.

The Itô-Taylor stochastic expansion allows a twice continuously differentiable function F to be expanded about $F(X(t))$ in terms of multiple stochastic integrals weighted by some coefficients evaluated at $X(t_0)$. Because our analysis is restricted to a 1-dimensional Itô diffusion process, $F(t, x)$ depends only on x with $F(t, x) = F(x)$. The coefficients are constructed from the drift and diffusion components of the Itô diffusion process after some specified orders of these terms.

Theorem 2.12 (Kloeden and Platen [1992], Chapter 5.1, Itô-Taylor Stochastic Expansion). *Suppose $X(t)$ is the process defined in equation (2.28) and $b'(x) = \frac{\partial}{\partial x}b(x)$. The Itô-Taylor stochastic expansion with a time discretization, $0 = t_0 < t_1 < \cdots < t_N = T$ on a time interval $[0, T]$ is:*

$$\begin{aligned} X(t_{i+1}) =& X(t_i) + a(X(t_i))\,\Delta + b(X(t_i))\,\widetilde{\Delta}W(i) \\ &+ \frac{1}{2}\,b(X(t_i))\,b'(X(t_i))\,[(\widetilde{\Delta}W(i))^2 - \Delta] + \tilde{\mathcal{R}}, \end{aligned} \tag{2.29}$$

where equidistant time step $\Delta = t_{i+1} - t_i$, $\widetilde{\Delta}W(i) = W(t_{i+1}) - W(t_i)$ for $i = 0, 1, 2, ..., N-1$, a given initial value $X(t_0) = X(0) = x_0$, and $\tilde{\mathcal{R}}$ is the remaining term defined as:

$$\begin{aligned} \tilde{\mathcal{R}} =& \int_{t_i}^{t_{i+1}} \int_{t_i}^{s} \mathcal{L}^1 a(X(z))\,dz\,ds + \int_{t_i}^{t_{i+1}} \int_{t_i}^{s} b(X(z)) \frac{\partial}{\partial x} a(X(z))\,dW(z)\,ds \\ &+ \int_{t_i}^{t_{i+1}} \int_{t_i}^{s} \mathcal{L}^1 b(X(z))\,dz\,dW(s) + \int_{t_i}^{t_{i+1}} \int_{t_i}^{s} \int_{t_0}^{z} \mathcal{L}^1 b(X(u)) b'(X(u))\,du\,dW(z)\,dW(s) \\ &+ \int_{t_i}^{t_{i+1}} \int_{t_i}^{s} \int_{t_0}^{z} b(X(u))(b'(X(u)))^2\,dW(u)\,dW(z)\,dW(s), \end{aligned} \tag{2.30}$$

where $\mathcal{L}$ is the operator function defined in Definition 2.11.

Remark 2.13. The Itô-Taylor stochastic expansion formula shown above is for the 1-D time-homogeneous Itô diffusion process. The expansion formula for a general SDE is shown and proved in [Kloeden and Platen, 1992, Section. 5].

We first define some notations and provide some general results. Then we prove Theorem 2.12.

Notation 2.14.

1. Suppose F is a twice continuous differentiable function, so $F \in C^2$.

2. Given the computational formula of the infinitesimal generator in Definition 2.11,

$\mathcal{L}^1 F(x)$ is:

$$\mathcal{L}^1 F(x) = a(x)\frac{\partial}{\partial x}F(x) + \frac{1}{2}b(x)^2\frac{\partial^2}{\partial x^2}F(x). \tag{2.31}$$

3. $\int_{t_i}^{t_{i+1}} dt = t_{i+1} - t_i = \widetilde{\Delta} t_i$.

4. $\int_{t_0}^{t} dW(s) = \sum_{i=0}^{N-1} (W(t_{i+1}) - W(t_i)) = W(t) - W(t_0)$, for any points $t_0 \leq \cdots \leq t_i \leq t_{i+1} \leq \cdots \leq t_N = t$, and $\widetilde{\Delta} W(i) = W(t_{i+1}) - W(t_i)$.

Lemma 2.15. *The double Itô integral is evaluated as*

$$\int_{t_i}^{t_{i+1}} \int_{t_i}^{s} dW(z) dW(s) = \frac{1}{2}\left((\widetilde{\Delta} W(i))^2 - \widetilde{\Delta} t_i\right). \tag{2.32}$$

Proof of Lemma 2.15. Evaluate the double Itô integral from the inside to outside and realize that the first integral is a definite integral:

$$\begin{aligned}
&\int_{t_i}^{t_{i+1}} \int_{t_i}^{s} dW(z)\, dW(s) \\
&= \int_{t_i}^{t_{i+1}} [W(s) - W(t_i)]\, dW(s) \\
&= \int_{t_i}^{t_{i+1}} W(s)\, dW(s) - W(t_i) \int_{t_i}^{t_{i+1}} dW(s) \\
&= \frac{1}{2}W(t_{i+1})^2 - \frac{1}{2}t_{i+1} - (\frac{1}{2}W(t_i)^2 - \frac{1}{2}t_i) \quad \text{(given Example 2.9)} \\
&\quad - W(t_i)[W(t_{i+1}) - W(t_i)] \\
&= \frac{1}{2}\left(W(t_{i+1}) - W(t_i)\right)^2 - \frac{1}{2}(t_{i+1} - t_i) \\
&= \frac{1}{2}\left((\widetilde{\Delta} W(i))^2 - \widetilde{\Delta} t_i\right). \quad \text{(given notations defined in Notation 2.14)} \qquad \square
\end{aligned}$$

Proof of Theorem 2.12. To generate the Itô-Taylor stochastic expansion, we first apply the Itô formula (2.18) on a SDE taken an integral form (2.4). For a twice continuously differentiable

function $F(X(t))$, we obtain:

$$\begin{aligned} F(X(t)) = F(X(t_0)) &+ \int_{t_0}^{t} \left(a(X(s)) \frac{\partial}{\partial x} F(X(s)) + \frac{1}{2} b^2(X(s)) \frac{\partial^2}{\partial x^2} F(X(s)) \right) ds \\ &+ \int_{t_0}^{t} b(X(s)) \frac{\partial}{\partial x} F(X(s))\, dW(s) \\ &= F(X(t_0)) + \int_{t_0}^{t} \mathcal{L}^1 F(X(s))\, ds + \int_{t_0}^{t} b(X(s)) \frac{\partial}{\partial x} F(X(s))\, dW(s), \end{aligned} \tag{2.33}$$

for $s \in [t_0, t]$, and applying the definition of operators $\mathcal{L}^1 F(x)$ in (2.31).

We then apply Itô formula to equation (2.33) and approximate the SDE by truncating the higher-order multiple integrals, representing by the remaining term of the Itô-Taylor stochastic expansion. Then given equation (2.33), let $F_1(x) = x, F_2(x) = a(x), F_3(x) = b(x)$, where F_1, F_2 and F_3 are twice continuous differentiable functions, we obtain

$$\begin{cases} F_1(X(t)) = X(t) = X(t_0) + \displaystyle\int_{t_0}^{t} a(X(s))\, ds + \int_{t_0}^{t} b(X(s))\, dW(s) & (2.34) \\ F_2(X(t)) = a(X(t)) = a(X(t_0)) + \displaystyle\int_{t_0}^{t} \mathcal{L}^1 a(X(s))\, ds + \int_{t_0}^{t} b(X(s)) \frac{\partial}{\partial x} a(X(s))\, dW(s) & (2.35) \\ F_3(X(t)) = b(X(t)) = b(X(t_0)) + \displaystyle\int_{t_0}^{t} \mathcal{L}^1 b(X(s))\, ds + \int_{t_0}^{t} b(X(s)) \frac{\partial}{\partial x} b(X(s))\, dW(s). & (2.36) \end{cases}$$

Subsequently, we can write out the second equality in (2.34), and then substitute $a(X(s))$ and $b(X(s))$ by equations (2.35) and (2.36), which generates

$$\begin{aligned} X(t) &= X(t_0) + \int_{t_0}^{t} a(X(s))\, ds + \int_{t_0}^{t} b(X(s))\, dW(s) \\ &= X(t_0) + \int_{t_0}^{t} \left(a(X(t_0)) + \int_{t_0}^{s} \mathcal{L}^1 a(X(z))\, dz + \int_{t_0}^{s} b(X(z)) \frac{\partial}{\partial x} a(X(z))\, dW(z) \right) ds \\ &+ \int_{t_0}^{t} \left(b(X(t_0)) + \int_{t_0}^{s} \mathcal{L}^1 b(X(z))\, dz + \int_{t_0}^{s} b(X(z)) \frac{\partial}{\partial x} b(X(z))\, dW(z) \right) dW(s). \end{aligned} \tag{2.37}$$

Next, we collect the double integral terms with a remaining term $\mathcal{R}$ and obtain:

$$X(t) = X(t_0) + a(X(t_0)) \int_{t_0}^{t} ds + b(X(t_0)) \int_{t_0}^{t} dW(s) + \mathcal{R}, \tag{2.38}$$

where the remainder is:

$$\begin{aligned}\mathcal{R} = &\int_{t_0}^{t}\int_{t_0}^{s} \mathcal{L}^1 a(X(z))\, dzds + \int_{t_0}^{t}\int_{t_0}^{s} b(X(z))\frac{\partial}{\partial x}a(X(z))\, dW(z)ds \\ &+ \int_{t_0}^{t}\int_{t_0}^{s} \mathcal{L}^1 b(X(z))\, dzdW(s) + \int_{t_0}^{t}\int_{t_0}^{s} b(X(z))\frac{\partial}{\partial x}b(X(z))\, dW(z)\, dW(s).\end{aligned}$$

We can expand the last double integral term in $\mathcal{R}$ further by applying Itô formula (2.18) with $F(x) = b(x)\frac{\partial}{\partial x}b(x) = b(x)b'(x)$. The result presented below is the Itô-Taylor stochastic expansion for the SDE (2.28), with the remainder $\tilde{\mathcal{R}}$ that consists of higher-order stochastic integrals:

$$\begin{aligned}X(t) =& X(t_0) + a(X(t_0)) \int_{t_0}^{t} ds + b(X(t_0)) \int_{t_0}^{t} dW(s) \\ &+ b(X(t_0))b'(X(t_0)) \int_{t_0}^{t}\int_{t_0}^{s} dW(z)\, dW(s) + \tilde{\mathcal{R}},\end{aligned} \tag{2.39}$$

where the new remainder term $\tilde{\mathcal{R}}$ is

$$\begin{aligned}\tilde{\mathcal{R}} = &\int_{t_0}^{t}\int_{t_0}^{s} \mathcal{L}^1 a(X(z))\, dzds + \int_{t_0}^{t}\int_{t_0}^{s} b(X(z))\frac{\partial}{\partial x}a(X(z))\, dW(z)ds \\ &+ \int_{t_0}^{t}\int_{t_0}^{s} \mathcal{L}^1 b(X(z))\, dzdW(s) + \int_{t_0}^{t}\int_{t_0}^{s}\int_{t_0}^{z} \mathcal{L}^1 b(X(u))b'(X(u))du\, dW(z)\, dW(s) \\ &+ \int_{t_0}^{t}\int_{t_0}^{s}\int_{t_0}^{z} b(X(u))(b'(X(u)))^2 dW(u)\, dW(z)\, dW(s).\end{aligned} \tag{2.40}$$

The construction of a numerical method for a SDE, in other words, the discrete-time Itô-Taylor approximation, requires the evaluation of the integral terms in equation (2.39). Assuming equidistant step size $\widetilde{\Delta}t_i = t_{i+1} - t_i$ and $\widetilde{\Delta}W(i) = W(t_{i+1}) - W(t_i)$ for all $i = 0, 1, \cdots, N-1$,

we can rewrite the expansion (2.39) as

$$\begin{aligned} X(t_{i+1}) =& X(t_i) + a(X(t_i))\,\widetilde{\Delta}t_i + b(X(t_i))\,\widetilde{\Delta}W(i) \\ &+ \frac{1}{2}\,b(X(t_i))\,b'(X(t_i))\,[(\widetilde{\Delta}W(i))^2 - \widetilde{\Delta}t_i] + \tilde{\mathcal{R}}, \end{aligned} \tag{2.41}$$

where the last double Itô integral in (2.39) is solved in Lemma 2.15 and $\tilde{\mathcal{R}}$ is defined in (2.40) with $X(t_0) = x_0$. □

2.6.2 Euler-Maruyama Method

The Euler-Maruyama method [Maruyama, 1955] is one of the simplest numerical methods for SDEs. It is obtained when the Itô-Taylor approximation (2.41) is truncated after the first order terms:

$$X(t_{i+1}) = X(t_i) + a(X(t_i))\,\widetilde{\Delta}t_i + b(X(t_i))\,\widetilde{\Delta}W(i), \tag{2.42}$$

where $\widetilde{\Delta}t_i = t_{i+1} - t_i$, $\widetilde{\Delta}W(i) = W(t_{i+1}) - W(t_i)$ for $i = 0, 1, 2, ..., N-1$ and $X(t_0) = x_0$.

2.6.3 Milstein Method

The Milstein method [Milstein, 1975] keeps the second order terms of the Itô-Taylor approximation (2.41) and is constructed as:

$$\begin{aligned} X(t_{i+1}) =& X(t_i) + a(X(t_i))\,\widetilde{\Delta}t_i + b(X(t_i))\,\widetilde{\Delta}W(i) \\ &+ \frac{1}{2}\,b(X(t_i))\,b'(X(t_i))\,[(\widetilde{\Delta}W(i))^2 - \widetilde{\Delta}t_i], \end{aligned} \tag{2.43}$$

where $\widetilde{\Delta}t_i = t_{i+1} - t_i$, $\widetilde{\Delta}W(i) = W(t_{i+1}) - W(t_i)$ for $i = 0, 1, 2, ..., N-1$ and $X(t_0) = x_0$.

2.6.4 Simulations of SDEs by Euler-Maruyama and Milstein Methods

As discussed at the beginning of Section 2.6, numerical methods to approximate the solutions of stochastic differential equations are required since the solutions of most SDEs are not available in explicit form. This thesis aims to generate approximations to a realistic financial process that cannot be solved explicitly. Simulations for general Itô diffusion processes (2.6) using the Euler-Maruyama method (2.42) and the Milstein method (2.43) will be conducted in R and compared across various grid widths.

The simulation algorithm is constructed for general one-dimensional Itô diffusion processes. Both the Euler method and the Milstein method will be used in the simulation scheme, and comparisons between the two methods are analysed on various grid widths.

Algorithm Construction:

1. Gold standard simulation: the simulation from the Milstein method with grid width 10^{-7} ($N = 10^7$).

   Target simulations: simulations with the grid width set as $10^{-6}, 10^{-5}, 10^{-4}$ and 10^{-3} for both the Euler and the Milstein methods.

2. Generate N independent noise terms separately, $\widetilde{\Delta}W(i)$, where $\widetilde{\Delta}W(i) \sim \mathcal{N}(0, \widetilde{\Delta}t_i)$ independently, for $i = 0, 1, 2, ..., N-1$, so that the randomness introduced to the two methods are same.

3. Proposed stochastic model: the Cox-Ingersoll-Ross model (Section 2.5.2) for a process $R(t)$:

$$dR(t) = \theta(\mu - R(t))\, dt + \sigma\sqrt{R(t)}\, dW(t), \tag{2.44}$$

   where $R(0) = 1, T = 1, (\theta, \mu, \sigma) = (3, 2, 2)$ and $N = \{10^3, 10^4, 10^5, 10^6, 10^7\}$.

The approximation form using the Euler method is

$$R(t_{i+1}) = R(t_i) + 3 \times (2 - R(t_i))\,\widetilde{\Delta}t_i + 2 \times \sqrt{R(t_i)}\,\widetilde{\Delta}W(i), \tag{2.45}$$

and using the Milstein method is:

$$R(t_{i+1}) = R(t_i) + 3 \times (2 - R(t_i))\,\widetilde{\Delta}t_i + 2\sqrt{R(t_i)}\,\widetilde{\Delta}W(i) + [(\widetilde{\Delta}W(i))^2 - \widetilde{\Delta}t_i]. \tag{2.46}$$

4. Accuracy check: L^1 distance: $|R(t_N) - r(t_N)|$, where $R(t_N)$ represents the gold standard simulation using the Milstein method at time T with grid width 10^{-7} and $r(t_N)$ is any target simulation obtained by either the Euler method or the Milstein method at time T for various grid widths (from 10^{-6} to 10^{-3}).

Table 2.1 provides the L^1 distance analysis for the Euler method and the Milstein method across different grid widths from a single simulation of the CIR model.

L^1 Distance for a single simulation				
Grid width	10^{-6}	10^{-5}	10^{-4}	10^{-3}
Euler-Maruyama	0.00280	0.02975	0.03198	0.03773
Milstein	0.00193	0.02870	0.02277	0.02384

Table 2.1: L^1 Distance for one sample path discretised on various scales using the Euler-Maruyama method and the Milstein method on the CIR model with $R(0) = 1, T = 1$ and parameters $(\theta, \mu, \sigma) = (3, 2, 2)$.

Discussion of Table 2.1:

1. From Table 2.1, an overall declining trend of the L^1 distance as the grid width increases can be observed for both the Euler method and the Milstein method. Because all the target

simulations are compared against the gold standard simulation, which we assume is the most accurate simulation, this general trend indicates that a smaller grid width will lead to higher simulation accuracy.

2. Comparing the results of the Euler and the Milstein methods, it is clear that the Milstein method provides smaller error, thus higher accuracy compared to the Euler method across all grid widths. This is because the Milstein method includes an additional second-order term from the Itô-Taylor stochastic expansion compared to the Euler method.

The first 500 observations of the simulation sample are plotted for various grid widths to compare the accuracy of the Euler method and the Milstein method.

Discussion of Figure 2.1:

1. All four plots in Figure 2.1 show the comparison among the three simulation methods for the CIR process with parameters $(\theta, \mu, \sigma) = (3, 2, 2)$, $R(0) = 1$ and $T = 1$. The three simulation methods include the gold standard simulation using the Milstein method with grid width 10^{-7} (green lines), the target simulation using the Milstein method with grid width varying from 10^{-3} to 10^{-6} (red lines) and the target simulation using the Euler method with grid width the same as the Milstein target (blue lines). Because the Milstein method is more accurate than the Euler method, theoretically, we should expect that the plots of coarse scales (10^{-3} or 10^{-4}) show a relatively large difference between the Euler and Milstein simulations than the finer scales (10^{-5} or 10^{-6}).

2. From the top two plots of Figure 2.1, we can easily distinguish between the red lines (Milstein) and blue lines (Euler), which verifies our previous supposition. Bottom two plots compare the same simulation methods but with a change of the grid width to 10^{-5} and 10^{-6} respectively. The green and blue lines are easily seen, while the red lines are hardly visible, especially when the grid width decreases to 10^{-6}. The result is verified to be consistent via

swapping the colour of the Euler and the Milstein methods, from which we can rule out the probability due to the colour allocation effect.

Conclusion:

Summarising the error analysis shown in the table and the plots, we conclude that the Milstein method provides a more accurate approximation to the ”gold standard” than the Euler method, particularly when the grid width is relatively large. As the grid width becomes smaller, the difference between the Euler method and the Milstein method diminishes.

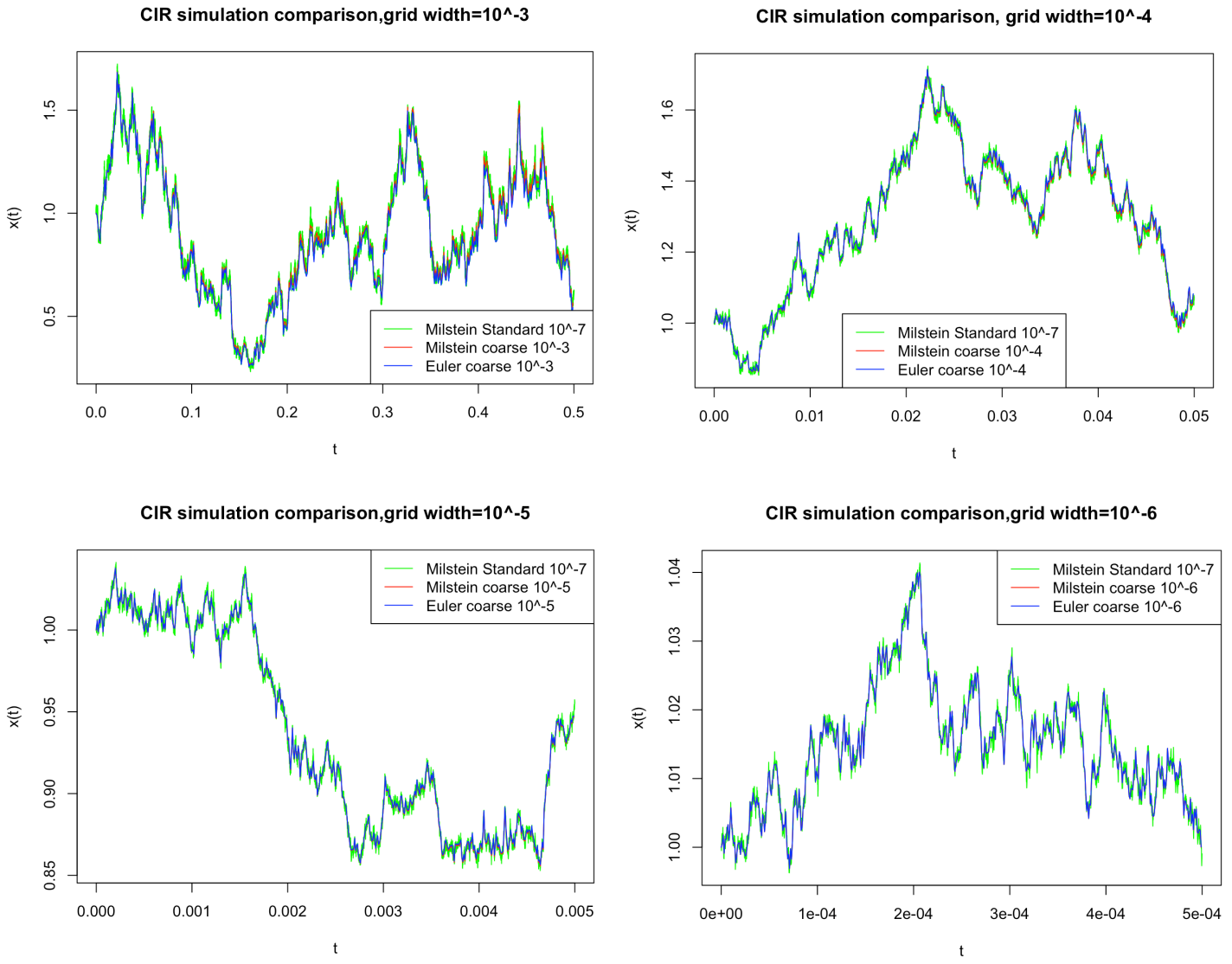

Figure 2.1: Sample path comparison plots of the first 500 observations of one simulation across various grid widths: (top-left): 10^{-3}, (top-right): 10^{-4}, (bottom-left): 10^{-5}, (bottom-right): 10^{-6}. The model parameters are $(\theta, \mu, \sigma) = (3, 2, 2)$, $R(0) = 1$ and $T = 1$.

3 Parameter Estimation of 1-D Itô Diffusion Process

In Chapter 2, we introduced the stochastic differential equation and some of its properties. We also derived the numerical method of diffusion processes using the Itô-Taylor stochastic expansion. This chapter focuses on the parameter estimation of time-homogeneous Itô diffusion processes with the Markov property.

Consider a one-dimensional Itô diffusion process $X(t)$ (Definition 2.2) on the interval $[0, T]$, which is the solution to the stochastic differential equation

$$dX(t) = a(X(t), \theta)\, dt + b(X(t), \sigma)\, dW(t),\ X_0 = x_0; \tag{3.1}$$

where $W(t)$ is the SBM (Definition 2.1), θ and σ are unknown parameters, and $a(x, \theta)$ and $b(x, \sigma)$ are the drift and diffusion coefficients defined on $\mathbb{R} \times \mathbb{R}_+$. From now on, the first argument of a and b is the state variable and the second argument is an unknown parameter. The existence of a unique solution of (3.1) is guaranteed by letting $a(x, \theta)$ and $b(x, \sigma)$ be continuously differentiable functions with respect to x.

A discretised trajectory $(X(t_i^n), 0 \leq i \leq n)$ with equidistant step

$$\Delta_n = t_{i+1}^n - t_i^n \tag{3.2}$$

is considered under discrete-time observations, where for each n, $[0, T]$ is divided into n equal

sub intervals, $0 = t_0^n \leq t_1^n \leq \cdots \leq t_{n-1}^n \leq t_n^n = T_n$. For notational convenience, we shall often drop the subscript n in Δ_n and T_n and drop the superscript in t_i^n. We are interested in estimating the unknown parameter vector $\alpha = (\sigma^\intercal, \theta^\intercal)^\intercal$.

A general method in statistical inference is using the likelihood function and maximum likelihood estimation. The difficulty of this approach is that the transition density is unknown for most stochastic processes. Many authors have contributed to the development of approaches to this estimation problem. Dacunha-Castelle and Florens-Zmirou [1986] provided a discussion that in ergodic case, the maximum likelihood estimator (vector) $\hat{\alpha}_{MLE}$ of the process (3.1) is consistent and asymptotically Gaussian under certain conditions. However, they proposed the exact discretisation of the transition density, which is not a practical method. A popular method conducted by many researchers ([Yoshida, 1992], [Kessler, 1997], [Shoji and Ozaki, 1998]) is to consider an approximation method to the transition density.

The aim of this chapter is to introduce and extend the parameter estimation scheme of 1-D Itô diffusion processes using approximation to the transition density methods. First, we review the likelihood function of Markov processes and present some necessary assumptions of the parameter estimation scheme. Then, we present and prove the higher-order Itô-Taylor expansion result reported in [Florens-Zmirou, 1989]. Some previous works, including the Gaussian approximation to the transition density [Kessler, 1997], are then revised. Most importantly, a new approximation method that uses the skew-normal density to approximate the transition density is introduced to handle the skewness in the process. Finally, the parameter estimation scheme is built in MATLAB and some approximation analyses for the Gaussian (Kessler) method and the skew-normal (SN) method are provided.

3.1 Likelihood approximation

3.1.1 Markov Property

First, we introduce some good properties of the diffusion process to assist the parameter estimation analysis. Kloeden and Platen [1992] introduced the Markov property of diffusion processes in a conditional probability format. We give the definition of the conditional probability of the diffusion process $X(t)$.

Definition 3.1. Suppose the diffusion process $X = (X(t), t \geq 0)$ defined in (3.1) taking continuous values in $\mathbb{R}$ has a joint density function $p_{X(t_0),\cdots,X(t_i)}\big(x_0, x_1, \cdots, x_i\big)$ for $i = 0, 1, 2\cdots$. Then let X be bounded Borel subsets $B \in \mathcal{B}(\mathbb{R})$, define its conditional probability as

$$\begin{aligned} &P\big(X(t_{i+1}) \in B | X(t_0) = x_0, X(t_1) = x_1, \cdots, X(t_i) = x_i\big) \\ &= \frac{\int_B p_{X(t_0),\cdots,X(t_i),X(t_{i+1})}\big(x_0, x_1, \cdots, x_i, y\,\big)dy}{\int_{-\infty}^{\infty} p_{X(t_0),\cdots,X(t_i),X(t_{i+1})}\big(x_0, x_1, \cdots, x_i, y\,\big)dy}, \end{aligned} \tag{3.3}$$

for all $0 \leq t_0 \leq t_1 \cdots \leq t_i \leq t_{i+1}$, and all $x_1, x_2, \cdots, x_{i+1} \in \mathbb{R}$, $i = 0, 1, 2\cdots$, and assuming the denominator is nonzero.

Theorem 3.2 (Markov Property, Kloeden and Platen [1992], Chapter 1 (6.17)). *For a diffusion process $X(t)$ possessing the Markov property, the future value $X(t_{i+1})$ only depends on the present value $X(t_i)$ and is not influenced by the past values $X(0), X(1), X(2), \cdots, X(t_{i-1})$. This is reflected in the formula*

$$\begin{aligned} &P\big(X(t_{i+1}) \in B | X(t_0) = x_0, X(t_1) = x_1, \cdots, X(t_i) = x_i\big) \\ &= P\big(X(t_{i+1}) \in B | X(t_i) = x_i\big), \end{aligned} \tag{3.4}$$

for all Borel subsets $B \in \mathcal{B}(\mathbb{R})$*, all* $0 \leq t_0 \leq t_1 \leq \cdots \leq t_{i+1}$ *and all* $x_1, x_2, \cdots, x_i \in \mathbb{R}$*,* $i = 0, 1, 2 \cdots$.

The diffusion process $X(t)$ (3.1) with the Markov property is also called a Markov process. Given Definition 3.1 and Theorem 3.2, the corresponding transition density $p_{X(t_{i+1})}\big(x|X(t_i) = x_i\big)$ is

$$P\big(X(t_{i+1}) \in B|X(t_i) = x_i\big) = \int_B p_{X(t_{i+1})}\big(x|X(t_i) = x_i\big)dx,$$

for all Borel subsets $B \in \mathcal{B}(\mathbb{R})$.

Given above definitions, we can decompose the joint density function into the product of transition densities. Denote the density of $X(t_i)$ as $p_{X(t_i)}(x_i)$, then the joint density of $X(t_0)$ and $X(t_1)$ is

$$p_{X(t_0),X(t_1)}\big(x_0, x_1\big) = p_{X(t_0)}(x_0)p_{X(t_1)}\big(x_1|X(t_0) = x_0\big),$$

for $0 \leq t_0 \leq t_1$. Consequently, we obtain the joint density for $x_0, x_1, \cdots, x_n$ using iteration

$$p_{X(t_0),\cdots,X(t_n)}\big(x_0, x_1, \cdots, x_n\big) = p_{X(t_0)}(x_0)\prod_{i=1}^{n} p_{X(t_i)}\big(x_i|X(t_{i-1}) = x_{i-1}\big), \tag{3.5}$$

for $0 \leq t_0 \leq t_1 \leq \cdots \leq t_n$.

3.1.2 Other Assumptions

The Markov property of diffusion processes performs a vital role in likelihood estimation; however, to build up the parameter estimation scheme, several additional assumptions need to be made, including a time-homogeneous Itô diffusion process with a stationary probability distribution, and hence the ergodic property.

When the Itô diffusion process was defined in Chapter 2, we briefly introduced an important feature, time-homogeneity, which indicates that the transition densities of the process $X(t)$ are dependent on the time difference $t_i - t_{i-1}$ only. We will state this property using the notation defined in the last section, that is

$$P\big(X(t_{i+1}) \in B | X(t_i) = x\big) = P\big(X(t_{i+1} - t_i) \in B | X(0) = x\big), \tag{3.6}$$

for any $0 \leq t_0 \leq \cdots \leq t_i \leq t_{i+1}$, $x \in \mathbb{R}$, and any Borel subsets $B \in \mathcal{B}(\mathbb{R})$.

Equation (3.6) shows that the transition probability has no relationship with the specific values t_{i-1} or t_i. Thus the Itô diffusion process is time-homogeneous. A formal proof of time-homogeneous property using the weak uniqueness of the solutions of SDEs is shown in [Øksendal, 2003, Definition. 7.1.1].

The assumption of time-homogeneity alone is not enough since this does not guarantee a stationary stochastic process. Gobet [2002] presents conditions under which there exists a unique stationary probability density $\bar{p}_{X(t_i)}(x_i)$ for a stationary diffusion process $X(t)$, such that

$$\bar{p}_{X(t_i)}(x_i) = \int_{-\infty}^{\infty} p_{X(t_i)}\big(x_i | X(t_{i-1}) = x_{i-1}\big)\, \bar{p}_{X(t_{i-1})}(x_{i-1})\, dx_{i-1}, \tag{3.7}$$

for all $x_1, \cdots, x_{i-1}, x_i \in \mathbb{R}$. This leads to a nice property called ergodicity, that connects the long run time average of the realisations and the spatial average of the stationary distribution.

Definition 3.3 (Ergodic Property). For a bounded measurable function $g(.) : \mathbb{R} \to \mathbb{R}$ and $\bar{p}_X(x)$ is the stationary probability density defined in (3.7), a diffusion process $X(t)$ possesses the ergodic property if

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T g(X(t))dt = \int_{-\infty}^{\infty} g(x)\, \bar{p}_X(x)\, dx, \quad \text{almost surely;} \tag{3.8}$$

that is, the limit of time average of its realisations exists and equals the spatial average with

respect to the stationary probability density.

The logic behind the ergodicity is the Law of Large Numbers, which will not be discussed here. Kloeden and Platen [1992] provided some detailed discussions on the ergodic property of discrete-time Markov chains. In this setting, the ergodicity condition guarantees a unique stationary probability density and thus a consistent estimator (vector), which are crucial properties in likelihood analysis.

3.1.3 Likelihood of Markov Processes

This section demonstrates the scheme of maximum likelihood estimation for the parameters of diffusion processes. Under the assumption of the Markov property of the diffusion process $X(t)$ (Theorem 3.2), we define the likelihood function based on the parameter vector α as a product of transition densities:

$$L_n(\alpha) = \prod_{i=1}^{n} p_{X(t_i)}(x_i|X(t_{i-1}) = x_{i-1}; \alpha), \tag{3.9}$$

where $p_{X(t_i)}(x_i|X(t_{i-1}) = x_{i-1}; \alpha)$ is the transition density of X given $X(t_{i-1}) = x_{i-1}$. The log-likelihood function with respect to the parameter vector α is then given by

$$l_n(\alpha) = \log L_n(\alpha) = \sum_{i=1}^{n} \log \left\{ p_{X(t_i)}(x_i|X(t_{i-1}) = x_{i-1}; \alpha) \right\}. \tag{3.10}$$

The score function is the first partial derivatives of the log-likelihood function with respect to the parameter vector α. When the transition density is differentiable, the score (vector) function is

$$U_n(\alpha) = \partial_\alpha \log L_n(\alpha) = \sum_{i=1}^{n} \partial_\alpha \log \left\{ p_{X(t_i)}(x_i|X(t_{i-1}) = x_{i-1}; \alpha) \right\}, \tag{3.11}$$

where $\partial_\alpha := \frac{\partial}{\partial \alpha}$ is the vector of partial derivatives.

The maximum likelihood estimators are found by setting $U_n(\alpha) = 0$. In general, the maximum likelihood estimation is an efficient technique for estimating unknown parameters. However, the transition density of the SDE is generally unknown. Therefore, an approximation to the likelihood function, and thus, the transition density, is required for maximum likelihood estimation.

3.2 High-order Itô-Taylor Expansion

To perform parameter estimation of the diffusion process with the maximum likelihood estimation method, we are interested in a practical method to approximate the transition densities. Dacunha-Castelle and Florens-Zmirou [1986] proposed an exact expression for the transition density, which is not practical for numerical calculation. However, they provided a valuable expansion result to approximate relevant characteristics of the diffusion process. They also proved consistency and the asymptotically Gaussian limit distribution of maximum likelihood estimators of the process defined in (3.1) as $\Delta_n \to 0$ and $T_n = n\Delta_n \to \infty$, where Δ_n is the equidistant time discretisation defined in (3.2) and n is number of observations. From now on, the analysis is on equidistant time discretisation and we will let Δ_n interchangeable with Δ. Moreover, the time interval T_n is not fixed and we allow the interchange between T_n and T. The high-order Itô-Taylor expansion result is introduced in this section, and some results of the Gaussian approximation to the distribution of the maximum likelihood estimators are discussed in the next section (Section 3.3).

First, we introduce some notation based on the Itô diffusion process $X(t)$ defined in (3.1).

Notation 3.4.

1. $\alpha = (\sigma^\intercal, \theta^\intercal)^\intercal$ and $c(x, \sigma) := b^2(x, \sigma)$.

2. $f : \mathbb{R} \to \mathbb{R}$: a $2(s+1)$-times continuously differential function.

3. For $0 \leq i \leq n$, $t_i^n = i\Delta_n$, where the equidistant step is assumed and Δ_n is interchangeable with Δ and the superscript n in t_i^n is allowed to be dropped.

4. θ_0, σ_0 and α_0 are the population values of θ, σ and α.

5. $\mathcal{L}_\alpha f$ is the infinitesimal generator based on parameter (vector) α defined in Definition 2.11. For simplicity, we will use $\mathcal{L}f := \mathcal{L}_\alpha f$ in most cases. The kth iterate of $\mathcal{L}$ is defined in equation (2.24).

In Chapter 2, we provided a simple version of the Itô-Taylor expansion (Theorem 2.12). A general Itô-Taylor expansion formula under a diffusion process $X(t)$ is given as

$$f(X(t)) = \sum_{j=0}^{s} \frac{\Delta_n^j}{j!} \mathcal{L}^s f(X(t_0)) + \int_{t_0}^{t} \int_{t_0}^{u_1} \cdots \int_{t_0}^{u_s} \mathcal{L}^{s+1} f(X(u_{s+1})) du_{s+1} \cdots du_1, \tag{3.12}$$

for $s \in [t_0, t]$. The Itô-Taylor expansion formula is important in numerical analysis and a modified form of the expansion is now considered, which is written in a conditional expectation form, conditioning on the σ-field $\mathcal{F}_i^n := \sigma(X(s), s \leq t_i^n)$ introduced by Dacunha-Castelle and Florens-Zmirou [1986].

Lemma 3.5 (Dacunha-Castelle and Florens-Zmirou [1986], Lemma 4). *For every function $f \in C^{2(s+1)}$ and the diffusion process $X(t)$ in the form*

$$dX(t) = a(X(t), \theta)\, dt + b(X(t), \sigma)\, dW(t), \tag{3.13}$$

where $\alpha = (\theta, \sigma)$ is unknown parameter vector and $W(t)$ is the SBM (Definition 2.1), we

define the higher-order Itô-Taylor expansion

$$E_{\alpha_0}[f(X(t_i))|\mathcal{F}_{i-1}^n] = \sum_{j=0}^{s} \frac{\Delta_n^j}{j!} \mathcal{L}_{\alpha_0}^j f(X(t_{i-1})) + \int_0^{\Delta_n} \int_0^{u_1} \cdots \int_0^{u_s} E_{\alpha_0}[\mathcal{L}_{\alpha_0}^{s+1} f(X(t_{i-1}+u_{s+1})|\mathcal{F}_{i-1}^n)] du_{s+1} \cdots du_1, \tag{3.14}$$

where α_0 is population value of α.

This result is used in many papers, but neither Dacunha-Castelle and Florens-Zmirou [1986] nor the subsequent literature has provided a proof of this expansion. We prove the expansion result using Dynkin's formula and induction on s.

Proof of Lemma 3.5. We first introduce Dynkin's formula and some lemmas, which will be helpful for the following induction proof procedure.

Theorem 3.6 (Dynkin's Formula, Øksendal [2003], Theorem 7.4.1). *For every function $f \in C^2$, τ is a stopping time, where $E[\tau] < \infty$. Then*

$$E[f(X(\tau))|X(0) = x] = f(x) + E\left[\int_0^\tau \mathcal{L}f(X(s))\, ds \middle| X(0) = x\right]. \tag{3.15}$$

Remark 3.7. Though Dynkin's formula is stated for stopping times, we will only need it for a fixed time.

Lemma 3.8 (Change the order of conditional expectation and integral). *Suppose $X(u)$ is a measurable function on probability space $(\Omega, \mathcal{F}, \mathbb{P})$, given two assumptions:*

1.

$$-\infty < a < b < \infty, \tag{3.16}$$

2.

$$\sup_{u\in[a,b]} E|X(u)| < \infty, \tag{3.17}$$

then

$$\int_a^b E[X(u)|\mathcal{F}]du = E\left[\int_a^b X(u)du\Big|\mathcal{F}\right]. \tag{3.18}$$

A rigorous proof is shown in [Schilling, 2017, Theorem. 27.17]. We provide a proof sketch of Lemma 3.8.

Proof of Lemma 3.8. Using conditional version of Fubini's theorem given by [Schilling, 2017], which holds under conditions (3.16) and (3.17), we have

$$\int_a^b E[X(u)|\mathcal{F}]du = E\left[\int_a^b X(u)du\Big|\mathcal{F}\right]. \tag{3.19}$$

Therefore, (3.18) holds. □

Lemma 3.9. *Let* $\mathbf{I}(.)$ *be the indicator function. The multiple integral is evaluated as*

$$\begin{aligned}&\int_0^{\Delta_n}\int_0^{u_1}\cdots\int_0^{u_k}\mathbf{I}\left(0\le u_{k+1}\le u_k\cdots\le u_1\le\Delta_n\right)\,du_{k+1}du_k\cdots du_1\\ =&\frac{\Delta_n^{k+1}}{(k+1)!},\end{aligned} \tag{3.20}$$

where Δ_n *is defined in Notation 3.4.*

Proof of Lemma 3.9. We first introduce an induction step:

$$\begin{aligned}&\int_0^{u_1}\cdots\int_0^{u_k}\mathbf{I}\left(0\le u_{k+1}\le u_k\cdots\le u_1\right)\,du_{k+1}\cdots du_2\\ =&\frac{u_1^k}{k!}.\end{aligned} \tag{3.21}$$

Elementary calculation shows that (3.21) holds for $k = 1$ and $k = 2$. Consequently, we can evaluate the multiple integral in (3.20) as

$$\begin{aligned}
&\int_0^{\Delta_n} \int_0^{u_1} \cdots \int_0^{u_k} \mathbf{I}\left(0 \le u_{k+1} \le u_k \cdots \le u_1 \le \Delta_n\right)\ du_{k+1} du_k \cdots du_1 \\
=&\int_0^{\Delta_n} \left[\int_0^{u_1} \cdots \int_0^{u_k} \mathbf{I}\left(0 \le u_{k+1} \le u_k \cdots \le u_1\right)\ du_{k+1} \cdots du_2 \right] du_1 \\
=&\int_0^{\Delta_n} \frac{u_1^k}{k!}\, du_1 \qquad \text{(by induction step (3.21))} \\
=&\frac{\Delta_n^{k+1}}{(k+1)!}. \quad \square
\end{aligned}$$

Remark 3.10. The multiple integral is calculated from inside to outside in this paper, which is opposite to the way Dacunha-Castelle and Florens-Zmirou [1986] did.

Given the above definitions and results, we provide the specific proof of equation (3.14) by induction on s.

Base step : Show that the statement holds for the smallest natural number when $s = 0$:

$$\begin{aligned}
E_{\alpha_0}[f(X(t_i))|\mathcal{F}_{i-1}^n] = &\sum_{j=0}^{s=0} \frac{\Delta_n^j}{j!} \mathcal{L}_{\alpha_0}^j f(X(t_{i-1})) \\
&+ \int_0^{\Delta_n} E_{\alpha_0}\left[\mathcal{L}_{\alpha_0}^{s+1} f(X(t_{i-1}+u_1)) \middle| \mathcal{F}_{i-1}^n \right] du_1 \qquad (3.22) \\
&= f(X(t_{i-1})) + E_{\alpha_0}\left[\int_0^{\Delta_n} \mathcal{L}_{\alpha_0} f(X(t_{i-1}+u_1)) du_1 \middle| \mathcal{F}_{i-1}^n \right]. \qquad (3.23)
\end{aligned}$$

The change of integral and conditional expectation from (3.22) to (3.23) is justified in Lemma 3.8. Because the last equality (3.23) satisfies Theorem 3.6, the base step is verified to be correct.

Induction step: Show that for any $s \ge 0$, if $E_{\alpha_0}[f(X(t_i))|\mathcal{F}_{i-1}^n]$ holds for $s = k$, then $E_{\alpha_0}[f(X(t_i))|\mathcal{F}_{i-1}^n]$ for $s = k+1$ also holds.

First, we calculate the expansion in (3.14) for $s = k$:

$$E_{\alpha_0}[f(X(t_i))|\mathcal{F}_{i-1}^n] = \sum_{j=0}^{s=k} \frac{\Delta_n^j}{j!} \mathcal{L}_{\alpha_0}^j f(X(t_{i-1})) + \int_0^{\Delta_n} \int_0^{u_1} \cdots \int_0^{u_k} E_{\alpha_0}\left[\mathcal{L}_{\alpha_0}^{k+1} f(X(t_{i-1} + u_{k+1})) \middle| \mathcal{F}_{i-1}^n \right] du_{k+1} \cdots du_1. \tag{3.24}$$

If we apply Theorem 3.6 on the expectation part of (3.24), it can be rewritten as

$$\begin{aligned} E_{\alpha_0}[f(X(t_i))|\mathcal{F}_{i-1}^n] = & \sum_{j=0}^{s=k} \frac{\Delta_n^j}{j!} \mathcal{L}_{\alpha_0}^j f(X(t_{i-1})) \\ & + \int_0^{\Delta_n} \int_0^{u_1} \cdots \int_0^{u_k} \left\{ \mathcal{L}_{\alpha_0}^{k+1} f(X(t_{i-1})) \right. \\ & \left. + E_{\alpha_0}\left[\int_0^{u_{k+1}} \mathcal{L}_{\alpha_0}^{k+2} f(X(t_{i-1} + u_{k+2})) du_{k+2} \middle| \mathcal{F}_{i-1}^n \right] \right\} du_{k+1} \cdots du_1. \end{aligned} \tag{3.25}$$

By Lemma 3.9, the first term in the multiple integral in (3.25) is

$$\int_0^{\Delta_n} \int_0^{u_1} \cdots \int_0^{u_k} \mathcal{L}_{\alpha_0}^{k+1} f(X(t_{i-1}))\, du_{k+1} \cdots du_1 = \frac{\Delta_n^{k+1}}{(k+1)!} \mathcal{L}_{\alpha_0}^{(k+1)} f(X(t_{i-1})). \tag{3.26}$$

Then, apply Lemma 3.8 on the expectation term in the multiple integral of (3.25) and combine with (3.26), we can expand $E_{\alpha_0}[f(X(t_i))|\mathcal{F}_{i-1}^n]$ in (3.25) to the case $s = k + 1$:

$$E_{\alpha_0}[f(X(t_i))|\mathcal{F}_{i-1}^n] = \sum_{j=0}^{s=k+1} \frac{\Delta_n^j}{j!} \mathcal{L}_{\alpha_0}^j f(X(t_{i-1})) + \int_0^{\Delta_n} \int_0^{u_1} \cdots \int_0^{u_{k+1}} E_{\alpha}\left[\mathcal{L}_{\alpha_0}^{k+2} f(X_{t_{i-1}} + u_{k+2}) \middle| \mathcal{F}_{i-1}^n \right] du_{k+2} \cdots du_1. \tag{3.27}$$

Consequently, the statement when $s = k + 1$ is also true from the induction.

We have proved that both the base step and the inductive step are true. By mathematical induction, the statement (3.14) holds for every natural number s and the proof is completed. □

To assist in further analysis, equation (3.14) is split into two parts:

$$E_\alpha[f(X(t_i))|\mathcal{F}_{i-1}^n] = m_s(\Delta_n, f(X(t_{i-1})); \alpha) + \mathcal{R}_{s+1}(\Delta_n, x; \alpha), \tag{3.28}$$

where we define based on the information of α

$$m_s(\Delta_n, f(X(t_{i-1})); \alpha) := \sum_{j=0}^{s} \frac{\Delta_n^j}{j!} \mathcal{L}_\alpha^j f(X(t_{i-1})) \tag{3.29}$$

and

$$\mathcal{R}_{s+1}(\Delta_n, x; \alpha) := \int_0^{\Delta_n} \int_0^{u_1} \cdots \int_0^{u_s} E_\alpha \left[\mathcal{L}_\alpha^{s+1} f(X(t_{i-1} + u_{s+1})) \middle| \mathcal{F}_{i-1}^n \right] du_{s+1} \cdots du_1.$$

3.3 Parameter Estimation with Gaussian Approximation to the transition density

In this chapter, we describe a Gaussian approximation to the transition density due to Kessler [1997]. The Gaussian approximation to the transition density has been analysed in many studies, since the increment of standard Brownian motion is Gaussian distributed with mean of 0 and variance of the time increment Δ_n (Definition 2.1). It is intuitive to expect that the diffusion process $X(t)$ (3.1) driven by the standard Brownian motion $W(t)$ is well approximated by the Gaussian characteristic. Dacunha-Castelle and Florens-Zmirou [1986] proved that under certain conditions, the maximum likelihood estimators of the diffusion process $X(t)$ are con-

sistent and asymptotically Gaussian distributed at the limit, $\Delta_n \to 0$ and $T_n = n\Delta_n \to \infty$. Kessler [1997] developed a method of the Gaussian approximation to the transition density based the work done in previous studies ([Dacunha-Castelle and Florens-Zmirou, 1986] and [Florens-Zmirou, 1989]). Kessler's contribution, which will be described in this section, is to provide higher-order approximations to the conditional mean and conditional variance in the Gaussian approximation.

First, an approximation to the likelihood function is required for maximum likelihood estimation of an Itô diffusion process. Florens-Zmirou [1989] considered the asymptotically exact scheme based on the higher-order Itô-Taylor expansion result presented in Lemma 3.5 for a Itô diffusion process $Y(t)$ taking the form in (3.1) with drift term $a(y, \theta)$ and diffusion term $b(y, \sigma) = \sigma$:

$$Y(t_i) = Y(t_{i-1}) + a(Y(t_{i-1}), \theta)\Delta_n + \sigma\widetilde{\Delta}W(i), \quad Y(t_0) = Y_0; \tag{3.30}$$

with equidistant time discretisation Δ_n and increment of standard Brownian motion $\widetilde{\Delta}W(i) = W(t_i) - W(t_{i-1})$. This is also known as the Euler-Maruyama method (Section 2.6.2). Florens-Zmirou [1989] proved that under certain conditions, as $\Delta_n \to 0$, $n\Delta_n \to \infty$, the process $Y(t)$ possesses ergodicity and the maximum likelihood estimator (vector) $\hat{\alpha}_{MLE}$ is consistently Gaussian distributed. Moreover, as $n\Delta_n^3 \to 0$, the derived estimators are asymptotically efficient. The likelihood function of the process $Y(t)$ can be used to construct the score function (3.11) and maximum likelihood estimators $\hat{\alpha}_{MLE}$.

Given that $Y(t)$ holds for ergodicity and generates consistent and asymptotically Gaussian estimators, Kessler [1997] constructed a likelihood function of $X(t)$ in place of $Y(t)$ (3.30) using the Gaussian approximation to the transition density. Instead of assuming a constant diffusion term σ, Kessler [1997] considered the general Itô diffusion process (3.1) with the

diffusion term $b(x, \sigma)$. Specifically, the exact scheme in (3.30) is modified to be

$$X(t_i) = X(t_{i-1}) + a(X(t_{i-1}), \theta)\,\Delta_n + b(X(t_{i-1}), \sigma)\,\widetilde{\Delta}W(i), \quad X(t_0) = X(0). \tag{3.31}$$

The increments $(X(t_i) - X(t_{i-1})) \to \mathcal{N}\left(a(X(t_{i-1}), \theta)\Delta_n, b(X(t_{i-1}), \sigma)^2\Delta_n\right)$ as $\Delta_n \to 0$ and $n\Delta_n \to \infty$, for $i = 1, 2, \cdots, n$. Then, the Gaussian transition density is expressed as:

$$p_{X(t_i)}(x_i|X(t_{i-1}) = x_{i-1}; \alpha) = \frac{1}{\sqrt{2\pi b(x_{i-1}, \sigma)^2\Delta_n}} \exp\left\{-\frac{\left((x_i - x_{i-1}) - a(x_{i-1}, \theta)\Delta_n\right)^2}{2b(x_{i-1}, \sigma)^2\Delta_n}\right\}, \tag{3.32}$$

where the parameter vector $\alpha = (\theta^\intercal, \sigma^\intercal)^\intercal$.

Given the formula (3.10), the log-likelihood function of $(X(t_i))_{0\le i\le n}$ is:

$$l_n(\alpha) = -\frac{1}{2}\sum_{i=1}^{n}\left(\frac{\left(x_i - x_{i-1} - a(x_{i-1}, \theta)\Delta_n\right)^2}{b(x_{i-1}, \sigma)^2\Delta_n} + n\log\left(2\pi\, b(x_{i-1}, \sigma)^2\Delta_n\right)\right). \tag{3.33}$$

To approximate the log-likelihood function (3.33), it is natural to consider the first and second conditional moments. Kessler [1997] approximated the moments using the Itô-Taylor expansion formula (3.28) introduced in the last section. We start the demonstration of the approximation procedure by giving some definitions and results.

Definition 3.11. Define the moment generating function of a random variable X

$$M_X(t) := E[e^{tX}], \tag{3.34}$$

then the nth non-central moment of X is

$$\mu'_n = E[X^n] = M_X^{(n)}(0), \tag{3.35}$$

and the nth central moment of X is

$$\mu_n = E[(X - E[X])^n]. \tag{3.36}$$

Given the Gaussian property of increments demonstrated in (3.32), Kessler [1997] approximated the transition density by matching its mean and variance with the first raw moment of $X(t_{i-1})$

$$\mu_1'(\alpha, x_{i-1}) \simeq x_{i-1} + a(x_{i-1}, \theta)\Delta_n, \tag{3.37}$$

and the second central moment of $X(t_{i-1})$

$$\mu_2(\alpha, x_{i-1}) \simeq b(x_{i-1}, \sigma)^2\Delta_n. \tag{3.38}$$

In other words, this leads to an approximation of the mean and variance of the transition density using the mean and variance of Gaussian distribution:

$$\begin{cases} \mu_1'(\alpha, x_{i-1}) &= E_\alpha[x_i | X(t_{i-1}) = x_{i-1}] \\ \mu_2(\alpha, x_{i-1}) &= E_\alpha[(x_i - \mu_1'(\alpha, x_{i-1}))^2 | X(t_{i-1}) = x_{i-1}] \end{cases}. \tag{3.39}$$

Consequently, the transition density is approximated using the density of $\mathcal{N}(\mu_1'(\alpha, x_{i-1}), \mu_2(\alpha, x_{i-1}))$. The resulting log-likelihood function of Gaussian approximation is

$$l_n(\alpha) = -\frac{1}{2}\sum_{i=1}^{n}\left(\frac{(x_i - \mu_1'(\alpha, x_{i-1}))^2}{\mu_2(\alpha, x_{i-1})} + n\log(2\pi\mu_2(\alpha, x_{i-1}))\right). \tag{3.40}$$

The maximum likelihood estimation $\hat{\alpha}_{MLE}$ is obtained by differentiating (3.40) with respect to the parameter vector $\alpha = (\theta^\intercal, \sigma^\intercal)^\intercal$ and equating the score (vector) function to zero.

3.4 Parameter Estimation Scheme with Skewness

In the previous section, we introduced the Gaussian scheme of parameter estimation (the Kessler method). However, the Gaussian distribution is known to be a symmetric distribution about 0, which means it is expected to be insufficient in handling the skewed case. We are interested in constructing a skew-normal approximation to the transition density based on the previous analysis in order to overcome the limitation of the Kessler method. The conditions introduced in the Kessler method to ensure consistent and asymptotically efficient estimators are also assumed in this section, which includes $\Delta_n \to 0$, $n\Delta_n \to \infty$ and $n\Delta_n^3 \to 0$.

3.4.1 Skew-Normal Distribution

This section introduces the skew-normal (SN) distribution first proposed in [Azzalini, 1985]. We also highlight the general features and some properties of the skew-normal distribution.

Definition 3.12 (Standard Normal PDF and CDF). Denote $\phi(z)$ by the standard normal probability density function

$$\phi(z) = \frac{e^{-z^2/2}}{\sqrt{2\pi}}, \tag{3.41}$$

where $z \in \mathbb{R}$, and the cumulative density function, $\Phi(z)$, by

$$\Phi(z) = \int_{-\infty}^{z} \phi(t)\, dt. \tag{3.42}$$

We now introduce a remarkable result from [Azzalini, 1985], which is essential for the construction of the SN distribution.

Lemma 3.13 (Azzalini [1985], Lemma 1). *Let h be a density function symmetric about 0, and G be an absolutely continuous distribution function such that its first derivative G' is*

symmetric about 0*. Then,*

$$2\,G(\lambda y)\,h(y) \tag{3.43}$$

is a density function for $-\infty < y < \infty$ *and* $\lambda \in \mathbb{R}$*.*

The proof of Lemma 3.13 is provided in [Azzalini, 1985]. The author used this result to directly construct the skew-normal distribution.

Definition 3.14 (Skew-normal Distribution). A random variable Z following the skew-normal distribution has the density function:

$$f_Z(z;\gamma) = 2\phi(z)\Phi(\gamma z), \quad (-\infty < z < \infty); \tag{3.44}$$

with asymmetry parameter $\gamma \in \mathbb{R}$.

A negative γ will cause the density (3.44) to have more weight on negative z, which indicates a negative skewness, whereas a positive γ indicates a positively skewed density. It follows easily from (3.44) that when $\gamma = 0$, the SN distribution turns out to be a Gaussian distribution.

The distribution function of skew-normal distribution is

$$F_Z(z;\gamma) = 2\int_{-\infty}^{z} \phi(t)\Phi(\gamma t)\,dt = 2\int_{-\infty}^{z}\int_{-\infty}^{\gamma t} \phi(t)\phi(u)\,du\,dt. \tag{3.45}$$

Our main target is to find the moments and cumulants of SN distribution and use them to construct an approximation to the transition density. The even moments of Z following SN distribution are equal to the even moments of standard normal distribution [Azzalini, 1985]. The moment generating function (mgf) of SN distribution is also derived in [Azzalini, 1985]:

$$M(t) = 2e^{t^2/2}\Phi(\delta t), \tag{3.46}$$

where $\delta = \frac{\gamma}{\sqrt{1+\gamma^2}}$ and $e^{t^2/2}$ is the mgf of the standard normal distribution.

Consequently, we calculate the first three non-central moments of Z using the formula (3.35) and (3.46):

$$\begin{cases} E[Z] = \sqrt{\frac{2}{\pi}}\delta \\ E[Z^2] = 1 \\ E[Z^3] = \sqrt{\frac{2}{\pi}}(3\delta - \delta^3) \end{cases} . \tag{3.47}$$

In the next section, we will develop the general scheme for approximations to the transition density using the first three cumulants of the distribution.

3.4.2 Scheme of Transition Density Approximation

In this section, we construct the estimation scheme using the first three cumulants and the higher-order Itô-Taylor expansion result in (3.29). The reason we include the first three cumulants rather than the first three moments is a matter of computational convenience. The standardised skewness of a distribution is used later to quantify skewness.

First, we will give the definitions of the cumulant generating function and the cumulants.

Definition 3.15 (Cumulant Generating Function). Denote $K(t)$ as the cumulant generating function, which is the natural logarithm of the moment generating function (3.34):

$$K(t) = \log M(t). \tag{3.48}$$

The cumulants κ_r are the coefficients in a power series expansion of the cumulant generating function (3.48):

$$K(t) = \sum_{r=1}^{\infty} \kappa_r \frac{t^r}{r!} = \kappa_1 t + \kappa_2 \frac{t^2}{2!} + \kappa_3 \frac{t^3}{3!} + \cdots . \tag{3.49}$$

The rth cumulant can be obtained by differentiating expansion (3.49) r times and setting $t = 0$:

$$\kappa_r = K^{(r)}(0). \tag{3.50}$$

Next, we will define some notation.

Notation 3.16.

1. $\kappa_1, \kappa_2, \kappa_3$ are the first three cumulants, and are calculated using (3.50).
2. μ'_1, μ'_2, μ'_3 are the first three raw (non-central) moments, defined in equation (3.35).
3. μ_1, μ_2, μ_3 are the first three central moments, defined in equation (3.36).
4. The standardised skewness of a distribution is defined as $\frac{\kappa_3}{\kappa_2^{3/2}}$.

Given the relationship between the moment generating function and the cumulant generating function (3.48), the first three cumulants can be expressed as functions of moments:

$$\begin{cases} \kappa_1 = \mu'_1 \\ \kappa_2 = \mu'_2 - {\mu'_1}^2 = \mu_2 \\ \kappa_3 = \mu'_3 - 2\mu'_1\mu'_2 + 2{\mu'_1}^3 = \mu_3 \end{cases}, \tag{3.51}$$

where κ_2 and κ_3 are equal to the second and third central moments, respectively.

Similar to the Kessler method (3.3), we aim to approximate the transition density using the density of a distribution that allows skewness. Specifically, the first three cumulants of the

distribution are required. The key idea is demonstrated as follows:

$$\begin{cases} \kappa_1(\alpha, x) & = E_\alpha[X(t_i)|X(t_{i-1}) = x] \\ \kappa_2(\alpha, x) & = E_\alpha[(X(t_i) - \kappa_1(\alpha, x))^2|X(t_{i-1}) = x] \\ \kappa_3(\alpha, x) & = E_\alpha[(X(t_i) - \kappa_1(\alpha, x))^3|X(t_{i-1}) = x] \end{cases}, \tag{3.52}$$

where $\kappa_r(\alpha, x)$ is the conditional rth cumulant of $X(t_i)$ given $X(t_{i-1}) = x$.

The closed form for the moments function is generally not available because the transition density of the diffusion process $X(t)$ is unknown. We apply the expansion result (3.28) to evaluate the first three cumulants. Specifically, the remaining term is omitted and only the summation term $m_s(\Delta, f(x); \alpha)$, defined in equation (3.29), is employed, where Δ is interchangeable with Δ_n. All the terms in the expansion of size $O(\Delta^3)$ are neglected.

To demonstrate the approximation procedure for the Itô diffusion process (3.1), we start by defining some notation.

Notation 3.17.

1. $a := a(x, \theta), c := c(x, \sigma) = b(x, \sigma)^2$.
2. $a' := \frac{\partial}{\partial_x} a$; $c' := \frac{\partial}{\partial_x} c$; $a'' := \frac{\partial^2}{\partial_{x^2}} a$; $c'' := \frac{\partial^2}{\partial_{x^2}} c$; $\partial_x := \frac{\partial}{\partial x}$; $\partial_x^2 := \frac{\partial^2}{\partial x^2}$.
3. Define $f_j(x) = x^j$, where $f_j(x)$ is twice continuously differentiable function for $j = \{1, 2, 3\}$.
4. $m_{j,s} := m_s(\Delta, f_j(x); \alpha)$, $m_{j,s}^2 := (m_{j,s})^2$, where $m_s(.)$ is the power expansion series defined in equation (3.29). We assume $s = 2$ since the terms of size $O(\Delta^3)$ are neglected.
5. $\mathcal{L}f(x) := \mathcal{L}_\alpha f(x)$, the infinitesimal generator defined in Definition 2.11.

Lemma 3.18 (Moment Estimation). *Apply the formula (3.29) on* $m_{j,s}$*, where* $j = \{1, 2, 3\}$ *and* $s = 2$*, to obtain the approximation to the first three non-central moments:*

for the first moment,

$$\begin{cases} m_{1,2} = x + \Delta a + \frac{1}{2}\Delta^2(aa' + \frac{1}{2}ca'') + O(\Delta^3) \\ m_{1,2}^2 = x^2 + 2\Delta ax + \Delta^2(a^2 + aa'x + \frac{1}{2}a''cx) + O(\Delta^3) \end{cases}; \tag{3.53}$$

for the second moment,

$$\begin{aligned} m_{2,2} =& x^2 + \Delta(2bx + c) \\ &+ \frac{1}{2}\Delta^2(2bb'x + b''cx + 2b^2 + bc' + 2b'c + \frac{cc''}{2}) + O(\Delta^3); \end{aligned} \tag{3.54}$$

and for the third moment,

$$\begin{aligned} m_{3,2} =& x^3 + 3\Delta(bx^2 + cx) \\ &+ \frac{1}{2}\Delta^2(3bb'x^2 + 6b^2x + 3bc'x + 6bc + \frac{3}{2}b''cx^2 + 6b'cx + \frac{3}{2}cc''x) + O(\Delta^3). \end{aligned} \tag{3.55}$$

Proof of Lemma 3.18. The logic underlying the calculation of the first three moments is the same, so we only demonstrate the calculations for the first moment. First, we find $\mathcal{L}^0 f(x)$, $\mathcal{L}^1 f(x)$ and $\mathcal{L}^2 f(x)$ for a general function $f(x)$.

The infinitesimal generator $\mathcal{L}^0 f(x)$ is calculated using the formula (2.24) defined in Definition 2.11

$$\mathcal{L}^0 f(x) = f(x). \tag{3.56}$$

The first iterate of operator $\mathcal{L}^1 f(x)$ is obtained by applying formula (2.23)

$$\mathcal{L}^1 f(x) = a\,\partial_x f(x) + \frac{c}{2}\,\partial_x^2 f(x), \tag{3.57}$$

where $c := c(x) = b(x)^2$. The second iterate of operator $\mathcal{L}^2 f(x)$ is

$$\begin{aligned}
\mathcal{L}^2 f(x) &= \mathcal{L}\big(\mathcal{L}^1(f(x)\big) \\
&= a\,\partial_x\big(\mathcal{L}^1 f(x)\big) + \frac{c}{2}\,\partial_x^2\big(L^1 f(x)\big) \\
&= a\,\partial_x\big(a\,\partial_x f(x) + \frac{c}{2}\,\partial_x^2 f(x)\big) + \frac{c}{2}\,\partial_x^2\big(a\,\partial_x f(x) + \frac{c}{2}\,\partial_x^2 f(x)\big) \\
&= a\left(a'\,\partial_x f(x) + a\,\partial_x^2 f(x) + \frac{c'}{2}\,\partial_x^2 f(x) + \frac{c}{2}\,\partial_x^3 f(x)\right) \\
&+ \frac{c}{2}\Bigg(a''\,\partial_x f(x) + a'\,\partial_x^2 f(x) + a'\,\partial_x^2 f(x) + a\,\partial_x^3 f(x) \\
&+ \frac{c''}{2}\,\partial_x^2 f(x) + \frac{c'}{2}\,\partial_x^3 f(x) + \frac{c'}{2}\,\partial_x^3 f(x) + \frac{c}{2}\,\partial_x^4 f(x)\Bigg).
\end{aligned} \tag{3.58}$$

In the last equality of equation (3.58), the derivatives of a and c depend on the diffusion process $X(t)$. The first and second derivatives of $f(x)$ are required. However, the terms of size $O(\Delta^3)$ will be ignored. Now, we evaluate those quantities.

When $f(x) = f_1(x) = x$, the first and second derivatives of $f_1(x)$ is

$$\partial_x f_1(x) = 1;\ \partial_x^2 f_1(x) = 0.$$

Then we find $\mathcal{L}^0 f_1(x)$, $\mathcal{L}^1 f_1(x)$ and $\mathcal{L}^2 f_1(x)$ using formulas (3.56), (3.57) and (3.58), respectively

$$\begin{cases}
\mathcal{L}^0 f_1(x) = x \\
\mathcal{L}^1 f_1(x) = a \\
\mathcal{L}^2 f_1(x) = aa' + \frac{1}{2}a''c
\end{cases}. \tag{3.59}$$

Therefore, we estimate the first non-central moment of $f(x) = f_1(x) = x$ using the formula

(3.29):

$$
\begin{aligned}
m_{1,2} &= \sum_{j=0}^{s=2} \frac{\Delta^j}{j!} \mathcal{L}^j f_1(x) + O(\Delta^3) \\
&= \mathcal{L}^0 f_1(x) + \Delta \mathcal{L}^1 f_1(x) + \frac{1}{2}\Delta^2 \mathcal{L}^2 f_1(x) + O(\Delta^3) \\
&= x + \Delta a + \frac{1}{2}\Delta^2 (aa' + \frac{1}{2} ca'') + O(\Delta^3). \qquad (3.60)
\end{aligned}
$$

Proof for the first moment is completed. The same logic applies to the second and third non-central moments. □

Given the estimations of the non-central moments $m_{j,s},\ j = \{1,2,3\}$, $s = 2$ (Lemma 3.18), we estimate the cumulants using the relationship between cumulants and moments (3.51). Lemma 3.19 provides the result of the estimation scheme for the first three cumulants.

Lemma 3.19. *The first three cumulants approximations are*

$$
\begin{cases}
\kappa_1 = x + \Delta a + \frac{\Delta^2}{2}(aa' + \frac{1}{2} ca'') + O(\Delta^3) \\
\kappa_2 = \Delta c + \Delta^2 (\frac{ac'}{2} + a'c + \frac{cc''}{4}) + O(\Delta^3) \\
\kappa_3 = \frac{3}{2}\Delta^2 cc' + O(\Delta^3)
\end{cases} . \qquad (3.61)
$$

Proof of lemma 3.19. In lemma 3.18, we define $m_{j,s}, j = \{1,2,3\}$ as approximations to non-central moments. The definition of μ'_j (3.35) indicates $m_{j,s} \simeq \mu'_j$, and hence, we can generate the relationship between κ_j and $m_{j,s}$ for $j = \{1,2,3\}$ given the relationship (3.51)

$$
\begin{cases}
\kappa_1 \simeq m_{1,2} \\
\kappa_2 \simeq m_{2,2} - m_{1,2}^2 \\
\kappa_3 \simeq m_{3,2} - 2m_{1,2}m_{2,2} + 2m_{1,2}^3
\end{cases} .
$$

Replacing $m_{j,s}$ with the approximation results in Lemma 3.18 and after some cancellations, we obtain the result in Lemma 3.19. □

In the next two sections, we will present some proposed distributions (skew-normal distribution in Section 3.4.3 and gamma distribution in Section 3.4.4) to approximate the transition density using the estimated cumulants of the distribution. Subsequently, maximum likelihood estimation is performed.

3.4.3 Skew-normal Approximation to Transition Density

The skew-normal distribution generalised from the normal distribution to allow for non-zero skewness is proposed in this study as an approximation method to the transition density. We have already introduced the essential information and some properties of the SN distribution in Section 3.4.1. One good feature of SN distribution is its **strict inclusion** of the normal at $\gamma = 0$. If the standardised third cumulant of the skew normal approaches 0 then the skew normal distribution approaches the normal distribution and, consequently, when the transition density has a small standardised third cumulant then the skew normal-based and normal based approximations can be expected to give similar results. We can then use the characteristic results (Section 3.4.1) provided earlier to perform the approximation procedure to transition density with SN distribution.

SN distribution overcomes the drawback in gamma distribution (see Section 3.4.4), which gives limited coverage of only the positive skewness portion. SN distribution can cover both the negative and positive skewness without a limiting case at the Gaussian ($\gamma = 0$). Moreover, it is reasonable to expect that, in the limit as the sampling interval Δ goes to 0 and $n\Delta$ goes to infinity, the first-order asymptotic properties of the two methods will be the same, because the standardised third cumulant of the transition density goes to 0 as Δ goes to 0; see formula

(3.61). The first three raw moments of SN distribution are already calculated in (3.47), hence, we can calculate the cumulants of SN given the relationship between moments and cumulants (3.51).

In this study, we consider a linear transformation of the skew-normal random variable $X = \mu + \sigma Z (\sigma > 0)$ with density function

$$f_X(x) = \frac{2}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)\Phi\left(\gamma\frac{x-\mu}{\sigma}\right), \tag{3.62}$$

introduced by Azzalini [1985]. We can calculate the first three cumulants of the transformed SN random variable X using the density function of Z in Definition 3.14 and equation (3.62). The results are presented in Lemma 3.20.

Lemma 3.20 (Cumulants of a Transformed SN Random Variable). *Given equation (3.47), for a random variable Z that follows the skew-normal distribution, the first three cumulants are*

$$\begin{cases} \kappa_{1,z} = E[Z] = \sqrt{\frac{2}{\pi}}\delta \\ \kappa_{2,z} = E[(Z - E[Z])^2] = Var[Z] = 1 - \frac{2}{\pi}\delta^2 \\ \kappa_{3,z} = E[(Z - E[Z])^3] = \sqrt{\frac{2}{\pi}}(\frac{4}{\pi} - 1)\delta^3 \end{cases}, \tag{3.63}$$

where $\kappa_{i,z}$ is the ith cumulant of an SN random variable Z. Then, for a transformed SN random variable $X = \mu + \sigma Z$, $\sigma > 0$, the first three cumulants are as follows:

$$\begin{cases} \kappa_1 = \mu + \sqrt{\frac{2}{\pi}}\delta\sigma \\ \kappa_2 = \sigma^2(1 - \frac{2}{\pi}\delta^2) \\ \kappa_3 = \sqrt{\frac{2}{\pi}}\delta^3\sigma^3(\frac{4}{\pi} - 1) \end{cases}, \tag{3.64}$$

where κ_i is the ith cumulant of the transformed SN random variable X.

Next, we interpret the three parameters (μ, σ^2, γ) of X in terms of the first three cumulants, and hence, we can build the SN approximation to the transition density, given the density function of a transformed SN random variable (3.62).

Lemma 3.21. *Given the results in Lemma 3.20, parameters of the transformed SN distribution are estimated as*

$$\begin{cases} \mu = \kappa_1 - \sqrt{\frac{2}{\pi}}\left(\frac{\kappa_3}{\sqrt{\frac{2}{\pi}}(\frac{4}{\pi}-1)}\right)^{\frac{1}{3}} \\ \sigma^2 = \left(\frac{\kappa_3}{2-\frac{\pi}{2}}\right)^{\frac{2}{3}} + \kappa_2 \\ \delta = \left\{\frac{\left(\frac{\kappa_3}{\sqrt{\frac{2}{\pi}}(\frac{4}{\pi}-1)}\right)^{\frac{2}{3}}}{\left(\frac{\kappa_3}{2-\frac{\pi}{2}}\right)^{\frac{2}{3}}+\kappa_2}\right\}^{\frac{1}{2}} \end{cases} . \tag{3.65}$$

Suppose the diffusion process $X(t)$ is of the form in (3.1) with parameter vector α, then, the transition density approximation based on the skew-normal distribution is

$$p_{X(t_i)}(x_i|X(t_{i-1}) = x_{i-1}; \alpha) = \frac{2}{\sigma}\phi\left(\frac{x_{i-1}-\mu}{\sigma}\right)\Phi\left(\gamma\frac{x_{i-1}-\mu}{\sigma}\right), \tag{3.66}$$

where μ, σ^2 and δ are estimated with first three cumulants in (3.65), $\gamma = \sqrt{\frac{\delta^2}{1-\delta^2}}$ when $\kappa_3 > 0$ and $\gamma = -\sqrt{\frac{\delta^2}{1-\delta^2}}$ when $\kappa_3 < 0$.

Finally, we approximate the log-likelihood function using Lemma 3.21:

$$l_n(\alpha) = \log L_n(\alpha) = \sum_{i=1}^{n} \log\left\{p_{X(t_i)}(x_i|X(t_{i-1}) = x_{i-1}; \alpha)\right\}, \tag{3.67}$$

and evaluate the maximum likelihood estimators $\hat{\alpha}_{MLE}$.

3.4.4 Gamma Approximation to Transition Density

An alternative to use the SN approximation to the transition density would be to use a gamma approximation to deal with skewness. Consider a random variable X, where $X - c$ for a con-

stant c follows the gamma distribution, with shape and scale parameters α and β respectively, where $x > 0, \alpha > 0, \beta > 0$. Then let a random variable $Y = (X - c) \sim Gam(\alpha, \beta)$, the probability density function of Y is

$$f(y; \alpha, \beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} y^{\alpha-1} e^{-\beta y}, \tag{3.68}$$

where $y \in [0, \infty)$ and the cumulative density function is

$$F(y; \alpha, \beta) = \int_0^y f(t; \alpha, \beta)\, dt. \tag{3.69}$$

We are interested in the first three moments of X given $Y = (X - c) \sim Gam(\alpha, \beta)$. The results are shown in the following lemma using the transformation method of random variables.

Lemma 3.22. *For $Y = (X - c) \sim Gam(\alpha, \beta)$, the relationship between the first three central moments of X and Y are*

$$\begin{cases} E[X] = E[Y] + c \\ E[(X - E[X])^2] = E[(Y - E[Y])^2] \\ E[(X - E[X])^3] = E[(Y - E[Y])^3] \end{cases} . \tag{3.70}$$

If we denote κ_i as the ith cumulant of X, then, we obtain the first three cumulants given the general cumulants results of gamma distribution and Lemma 3.22:

$$\kappa_1 = \frac{\alpha}{\beta} + c; \quad \kappa_2 = \frac{\alpha}{\beta^2}; \quad \kappa_3 = \frac{2\alpha}{\beta^3}. \tag{3.71}$$

Consequently, given the previous approximation results for the first three cumulants (Lemma

3.19) and equation (3.71), we can solve for the unknown parameters α, β and c

$$\begin{cases} \alpha = \frac{4\kappa_2^3}{\kappa_3^2} \\ \beta = \frac{2\kappa_2}{\kappa_3} \\ c = \kappa_1 - \frac{2\kappa_2^2}{\kappa_3} \end{cases} .$$

Although the estimations of the cumulants are achieved, there is a limitation of gamma distribution, that it covers the positive skewness portion only. To estimate the full range of the skewness, we need to introduce the negative gamma distribution to account for the negative skewness and the Gaussian distribution when the skewness goes to zero. This leads to an obvious difficulty in the parameter estimation process.

3.5 Parameter Estimation Computation Scheme

This section presents a computation scheme for the maximum likelihood estimator $\hat{\boldsymbol{\theta}}_{MLE} = (\theta_1, \theta_2, \theta_3)$ using the approximation to the transition density method introduced in the previous sections. The programming application is modified based on the simulation performed in MATLAB in [Lu et al., 2021]. The skew-normal approximation method proposed in this study is represented by SN. We also compare the SN result with that of the method of Kessler [1997] using the Gaussian approximation method, which is denoted by Kessler. The approximation of transition density is up to the order of Δ^2.

3.5.1 Assumptions and Models

The first model we select for this simulation analysis is the Cox-Ingersoll-Ross (CIR) model, introduced in Section 2.5.2. The CIR model has three parameters $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)$, and we

define the CIR model for an Itô diffusion process $X(t)$ as follows:

$$\textbf{Model 1}: \quad dX(t) = \theta_1(\theta_2 - X(t))\, dt + \theta_3\sqrt{X(t)}\, dW(t), \tag{3.72}$$

where $W(t)$ is the standard Brownian motion.

The potential advantage of the skew-normal approximation is the ability to deal with standardised skewness ($\frac{\kappa_3}{\kappa_2^{3/2}}$). When $\kappa_3 \neq 0$, the process exhibits either positive ($\kappa_3 > 0$) or negative ($\kappa_3 < 0$) skewness. Given the cumulant approximation formula derived in Lemma 3.19 and truncating the terms after $O(\Delta^2)$, κ_3 of the CIR model is approximated as

$$\kappa_3 \simeq \frac{3}{2}\,\Delta^2\,\theta_3^4\,x, \tag{3.73}$$

given equation (3.61) that $\kappa_3 = \frac{3}{2}\Delta^2 cc' + O(\Delta^3)$, and $cc' = \theta_3^4 x$ for a realisation point x . It is clear that the skewness of the CIR model depends on the discretisation step Δ, the diffusion coefficient θ_3 and the value of x, which are not allowed to be negative in the CIR model. And the skewness closes to zero as the discretisation step Δ decreases.

To further analyse the skewness (both positive and negative), we introduce another model:

$$\textbf{Model 2}: \quad dX(t) = \theta_1(\theta_2 - X(t))\, dt + \left(1 + \frac{\theta_3}{2}\sin\left(X(t)\right)\right)\, dW(t). \tag{3.74}$$

After introducing $\sin\left(X(t)\right)$ into the diffusion coefficient, the approximated value of the third cumulant becomes

$$\kappa_3 \simeq \frac{3}{2}\,\Delta^2\left(\frac{\theta_3^2}{4}\sin\left(2x\right) + \theta_3\cos\left(x\right)\right)\left(\frac{\theta_3^2}{2}\cos\left(2x\right) - \theta_3\sin\left(x\right)\right), \tag{3.75}$$

which may be positive or negative. Moreover, the smaller the discretisation step (Δ), the skewness is closer to 0. The sign of γ, the asymmetry parameter of the skew-normal distribution

defined in equation (3.66), is set according to the sign of κ_3 in each discretisation.

We use the Milstein approximation method (Section 2.6.3) to simulate the sample paths of the two models. Unlike the CIR model, which is known to have a non-central χ^2 transition density, the transition density of Model 2 (3.74) is unknown; therefore, a numerical simulation method is necessary to generate realisations of the process.

To visualise the performance of the maximum likelihood estimators, we examine the accuracy of approximations using two criteria: median absolute difference and root mean square error (RMSE). For both the Kessler and the SN methods, we compare the maximum likelihood estimators $\hat{\boldsymbol{\theta}}_{MLE}$ and the true parameter vector $\boldsymbol{\theta}_0$ that is used to generate the simulations. The maximum likelihood estimation scheme is performed using the functions *fminsearch* and *fminunc* in MATLAB. Both functions are optimisation procedures performed using the Nelder-Mead method, where the latter can be used to calculate the Hessian matrix when the variance estimation is needed. Each method (SN and Kessler) is assessed for various discretisation steps Δ. The observation window is unchanged as the value of Δ is varied. The observation window is fixed at 20, and the values of the discretisation step Δ considered here are 0.02, 0.01, 0.004 and 0.002. Therefore, the corresponding sample size N is varied across 1000, 2000, 5000 and 10000. There are 500 Monte-Carlo runs for all simulations.

A limitation of the SN method is that the standardised skewness is restricted to the interval $(-0.995, 0.995)$ as reported by Azzalini [1985]. In case there are some extreme skewness that moves out of this bound, we have modified the program to manually set the standardised skewness within the bound when it jumps outside the range.

3.5.2 Numerical results

As mentioned earlier, two parameters (median absolute difference and root mean square error (RMSE)) are calculated as the criteria to examine the errors of the estimated estimators. We are interested in examining the change in the accuracy of the approximation as the discretisation varies. In addition, we compare the approximation accuracy of the SN method and the Kessler method across the various discretisation scales. Specifically, we want to determine whether the two methods provide a significant difference in accuracy when there is skewness, either positive or negative. The average of the standardised skewness at each discretisation point is calculated as a reference.

Table 3.1 and Table 3.2 present the performances of various estimators using the Kessler method and the SN method for two models with a fixed observation window of 20, varying discretisation steps Δ and number of observations n. Two types of errors are calculated: the mean square error and the root mean square error. The standardised skewness at each discretisation point for each simulation $i = 1, \cdots, 500$ is calculated using formula $\frac{\kappa_3}{\kappa_2^{3/2}}$ and is denoted as $\rho_{3,i}$. The average standardised skewness of one simulation is $\bar{\rho}_{3,i} = \frac{1}{N}\sum \rho_{3,i}$, thus the average of 500 simulations is $\bar{\bar{\rho}}_3 = \frac{1}{500}\sum_{i=1}^{500} \bar{\rho}_{3,i}$.

Key findings:

It can be observed from the results of Model 1 (the CIR model) with parameters $\hat{\boldsymbol{\theta}}_0 = (2, 0.3, 1)$ presented in Table 3.1 that as the discretisation decreases, the trends of accuracy growth are not the same for the three parameters. θ_1 and θ_2 show no clear reduction in error estimations when Δ decreases from 0.02 to 0.002, whereas a significant reduction in the errors of θ_3 is realized. A similar result is also observed in Table 3.2 for Model 2 with parameters $\hat{\boldsymbol{\theta}}_0 = (0.5, 2, 3.5)$. This suggests that the diffusion coefficient, which measures the instantaneous volatility of the diffusion process X, is more accurately estimated when the discretisation is sufficiently

small. This is reasonable since X varies in a continuous scale in reality; as the discretisation step becomes sufficiently small ($\Delta \to 0$), the discretised observations become closer to the realisations in the continuous case. Consequently, variability in the process is better observed and leads to a smaller error. On the other hand, the drift coefficients estimate the long term properties of X and exhibit the mean-reversion; therefore, decreasing the discretisation step has no significant influence on the estimation accuracy of θ_1 and θ_2.

Further discussions:

1. From Table 3.1, the Kessler method and the SN method have almost the same approximation accuracy, especially when the discretisation step is small. The small skewness observed verifies that the SN distribution is approximately Gaussian distributed when the skewness is close to 0. The results of Model 2 in Table 3.2 are essentially the same. Though negative standardised skewness are observed at various discretisation steps, they are all small values close to zero. Therefore, as the discretisation decreases, the Kessler method and the SN method generate almost the same accuracy.

2. The approximated transition densities of the two models with different discretisation steps are plotted in Figure 3.1 and Figure 3.2 on top ($\Delta = 0.02$) and bottom ($\Delta = 0.01$). For the CIR model shown in Figure 3.1, the green line shows the density using the exact maximum likelihood estimation calculated given that the true transition density of the CIR process follows the non-central χ^2 distribution. From the top plot, it can be observed that when the exact transition density is positively skewed, both the Kessler and the SN methods are unable to exhibit the skewness. Given the formula of κ_3 (3.73), we expect that the skewness of the CIR model is close to zero as the discretisation decreases. In the bottom plot, the distribution is closer to normal, and the discrepancies between the Kessler, SN and true transition density are smaller, suggesting a higher accuracy. In Figure 3.2, the exact transition density is unknown for Model 2. The SN method exhibits a negatively skewed density compared with the Kessler

method in the top plot. In the bottom plot, similar to the CIR model, the distribution is closer to normal for both Kessler and SN. The discrepancy between Kessler and SN is minuscule, because κ_3 of Model 2 (3.75) also depends on Δ. The skew-normal tends to Gaussian as the discretisation (Δ) decreases and κ_3 closes to 0.

Conclusion:

Parameters are more accurately estimated as the discretisation decreases. In addition, the diffusion coefficients are more prone to changes in the discretisation step, providing a much higher accuracy at a small Δ. The SN method does not outperform the Kessler method nor the other way around. One reason is that the skewness are all close to zero in our analysis, making it difficult to verify the performance under skewed cases. Second, the values of the parameters also have substantial impacts.

	Δ	0.02		0.01		0.004		0.002	
	N	1000		2000		5000		10000	
	$\bar{\bar{\rho}}_3$	0.5282		0.3780		0.2442		0.1807	
	$\boldsymbol{\theta}$	***Kessler***	***SN***	***Kessler***	***SN***	***Kessler***	***SN***	***Kessler***	***SN***
θ_1	median $\lvert\hat{\theta} - \theta_0\rvert$	0.3451	0.3412	0.3300	0.3305	0.3426	0.3415	0.3293	0.3293
	RMS $\lvert\hat{\theta} - \theta_0\rvert$	0.6354	0.6432	0.5666	0.5705	0.5367	0.5383	0.5593	0.5605
θ_2	median $\lvert\hat{\theta} - \theta_0\rvert$	0.0406	0.0398	0.0334	0.0334	0.0456	0.0457	0.0427	0.0426
	RMS $\lvert\hat{\theta} - \theta_0\rvert$	0.0592	0.0582	0.0625	0.0615	0.0647	0.0643	0.0585	0.0584
θ_3	median $\lvert\hat{\theta} - \theta_0\rvert$	0.0180	0.0185	0.0104	0.0104	0.0078	0.0078	0.0053	0.0053
	RMS $\lvert\hat{\theta} - \theta_0\rvert$	0.0287	0.0289	0.0175	0.0175	0.0106	0.0106	0.0079	0.0079

Table 3.1: Error analysis of approximated estimators specified in Section 3.5.1 for the CIR model (Model 1 (3.72)) with parameters $\boldsymbol{\theta}_0 = (2, 0.3, 1)$ and $x_0 = 1$. $\hat{\theta}$ is the MLE using the proposed approximation method and applying on simulated sample paths, θ_0 is the true value of parameters used for simulating sample paths. Δ is the discretisation step and N is the number of observations. $\bar{\bar{\rho}}_3$ represents the average of the mean standardised skewness at each discretisation point of 500 Monte-Carlo runs.

	Δ	0.02		0.01		0.004		0.002	
	N	1000		2000		5000		10000	
	$\bar{\bar{\rho}}_3$	-0.0907		-0.0660		-0.0423		-0.0313	
	θ	***Kessler***	***SN***	***Kessler***	***SN***	***Kessler***	***SN***	***Kessler***	***SN***
θ_1	median $\lvert\hat{\theta} - \theta_0\rvert$	0.0833	0.0786	0.0785	0.0729	0.0769	0.0760	0.0839	0.0820
	RMS $\lvert\hat{\theta} - \theta_0\rvert$	0.3553	0.3681	0.3100	0.3111	0.2722	0.2728	0.3528	0.3548
θ_2	median $\lvert\hat{\theta} - \theta_0\rvert$	0.2730	0.2717	0.2621	0.2617	0.2485	0.2469	0.2686	0.2677
	RMS $\lvert\hat{\theta} - \theta_0\rvert$	0.4766	0.4813	0.4466	0.4481	0.4235	0.4239	0.4863	0.4867
θ_3	median $\lvert\hat{\theta} - \theta_0\rvert$	0.1008	0.0996	0.0551	0.0552	0.0275	0.0274	0.0197	0.0197
	RMS $\lvert\hat{\theta} - \theta_0\rvert$	0.1353	0.1340	0.0820	0.0818	0.0432	0.0432	0.0293	0.0293

Table 3.2: Error analysis of approximated estimators specified in Section 3.5.1 for Model 2 (3.74) with parameters $\boldsymbol{\theta}_0 = (0.5, 2, 3.5)$ and $x_0 = 0.6$. Other details are the same with Table 3.1.

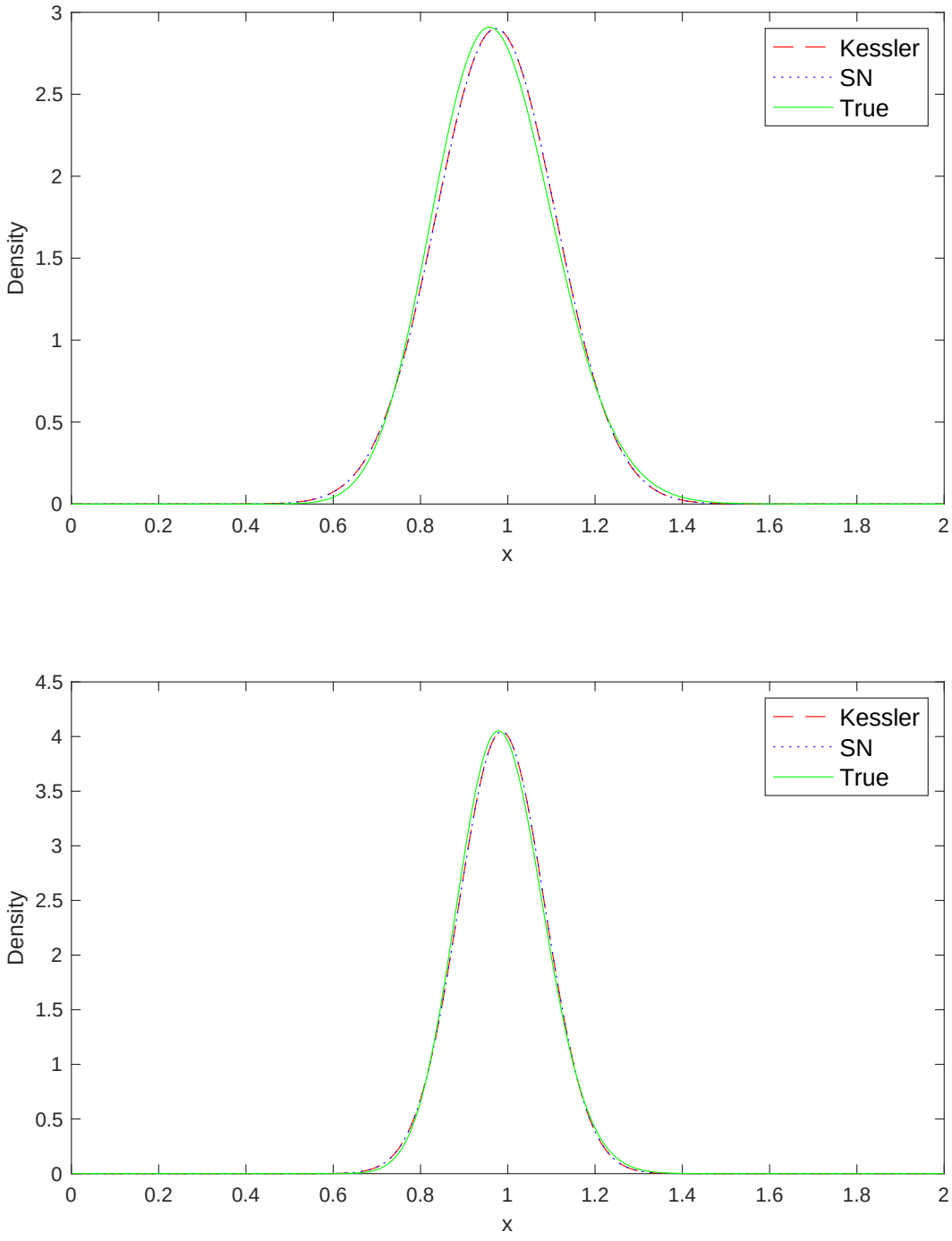

Figure 3.1: Approximations of the transition density specified in Section 3.5.1 of Model 1 (3.72). The model parameters are $\boldsymbol{\theta}_0 = (2, 0.3, 1)$, $x_0 = 1$, (Top) $\Delta = 0.02$, (Bottom) $\Delta = 0.01$.

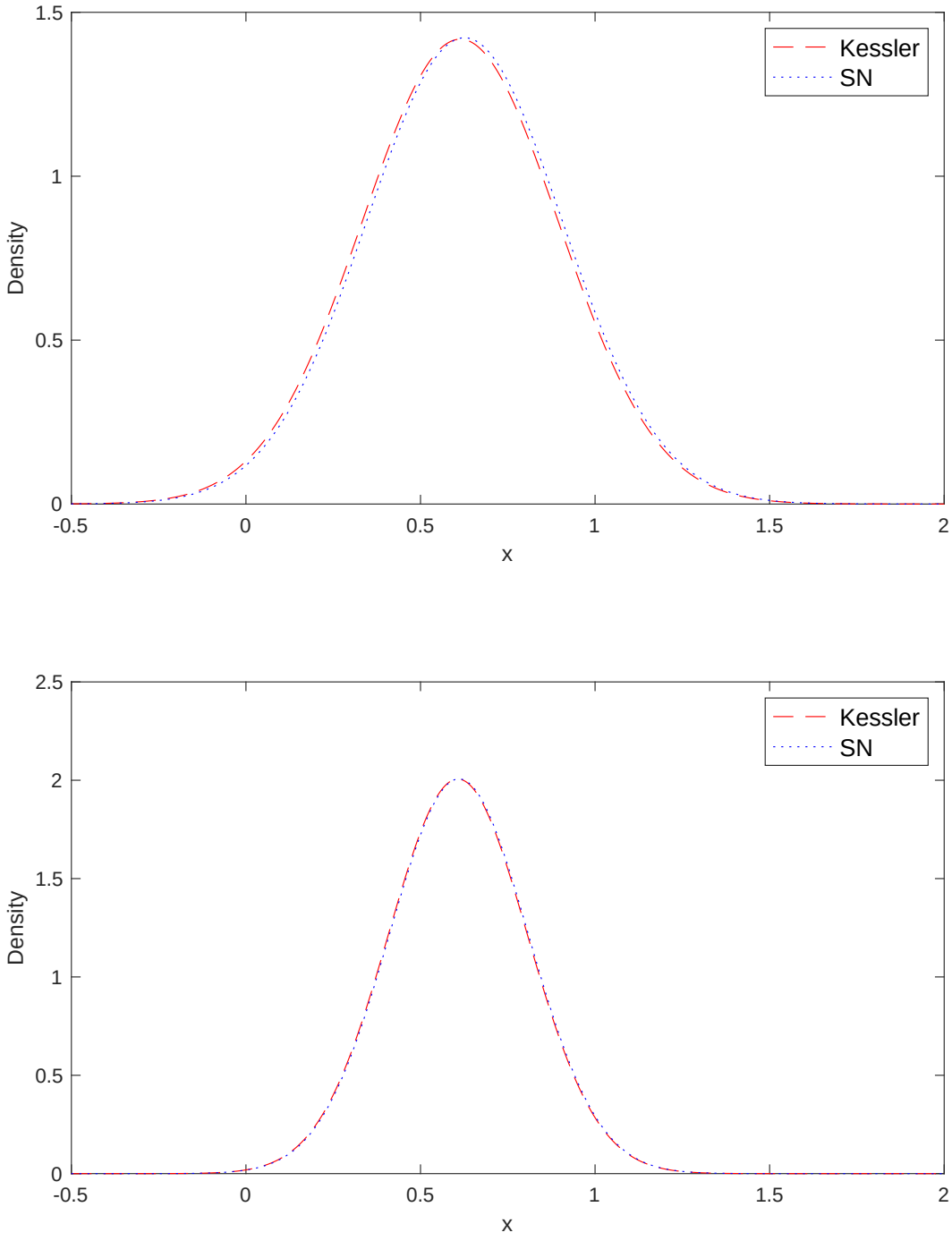

Figure 3.2: Approximations of the transition density specified in Section 3.5.1 of Model 2 (3.74). The model parameters are $\boldsymbol{\theta}_0 = (0.5, 2, 3.5)$, $x_0 = 0.6$, (Top) $\Delta = 0.02$, (Bottom) $\Delta = 0.01$.

4 Real Data Application

In the previous chapter, we introduced the maximum likelihood estimation scheme using approximations to the transition density. We discussed two different approximation methods in detail (SN method and Kessler method). In this chapter, we are interested in applying the parameter estimation procedure for Itô diffusion processes on a real data application, and hence, fitting a proper model with real data.

The data set used in this chapter is the daily yields (percent per annum) on the Australian government bond with 3 years maturity for the period 24/03/2014 to 05/04/2019 with 1261 observations (including the initial value) retrieved from Reserve Bank of Australia[1].

For this analysis, we focus on two models, the Vasicek model and the Cox-Ingersoll-Ross (CIR) model, introduced in Section 2.5.1 and Section 2.5.2. Both models exhibit mean-reverting features. It is of interest to estimate the parameters of the proposed two models. We have developed the parameter estimation scheme for the CIR model in Section 3.5, and the scheme for the Vasicek model is essentially the same, with a modification on the diffusion term. The two models for the interest rate $r(t)$ are expressed as

$$\textbf{Vasicek Model}: \; dr(t) = \theta_1(\theta_2 - r(t))\, dt + \theta_3\, dW(t), \tag{4.1}$$

$$\textbf{Cox-Ingersoll-Ross Model}: \; dr(t) = \theta_1(\theta_2 - r(t))\, dt + \theta_3\sqrt{r(t)}\, dW(t), \tag{4.2}$$

[1] data source: https://www.rba.gov.au/statistics/tables/xls/f02d.xls?v=2021-04-07-13-07-43

where parameter vector $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)$ and $W(t)$ is the standard Brownian motion.

First, we perform a general time series analysis on bond yield data in Section 4.1. In particular, we discuss the mean reversion property of the time series. A mean-reverting stochastic process is adopted as the primary assumption of the data. Then, in Section 4.2, we apply the parameter estimation scheme on the two mean-reverting models (the Vasicek model (4.3) and the CIR model (4.4)) using the optimisation function in MATLAB introduced in Chapter 3. Furthermore, we are interested in analysing the performance of the two models with the estimated parameters $\hat{\boldsymbol{\theta}}$. This is performed by simulating some sample paths generated using the numerical method simulation algorithm introduced in section 2.6. The Milstein method is used as the numerical method to achieve higher accuracy. There are 500 Monte-Carlo runs to generate 500 simulation paths. The average value of 500 Monte-Carlo runs is compared to the true data. To further examine the performance, we also investigate two types of confidence intervals: the one standard deviation from the mean of the 500 sample paths and the maximum and minimum values of the 500 sample paths at each time point.

Finally, in Section 4.3, we compare two estimated models using the parametric bootstrap hypothesis testing. This relies on the simulation-based calculation of the p-values, which is a technically important method when the distribution of the test statistic is unknown (non-parametric bootstrap hypothesis testing) or there exists a dubious approximate distribution (parametric bootstrap hypothesis testing). Previous studies ([Young, 1986], [Beran, 1988], [Hall, 1988], [Hall and Wilson, 1991]) suggested some guidelines for bootstrap hypothesis testing. The effectiveness of the bootstrap was demonstrated in [Efron and Tibshirani, 1994] and [Davison and Hinkley, 1997]. We conduct two bootstrap hypothesis tests with similar structures but reverse the null hypothesis. One test hypothesizes that the Vasicek model is correct, and we fit the simulated sample paths from both models into the Vasicek model with the estimated parameters $\hat{\boldsymbol{\theta}}$. Then, we calculate the maximum likelihood estimations using the optimisation function for each simulation and the true data. P-value is calculated as the

proportion of bootstrap samples with more extreme cases than the null hypothesis; in other words, the proportion of the test statistics of samples that are greater than the observed test statistic. The second test assumes that the CIR model is correct, and the rest of the test is a reverse of the role of data related to the Vasicek model and the CIR model. In the end, we analyse the approximation power of the final model through a simulation study of the accuracy of the constructed confidence intervals.

4.1 Daily Yield of 3-year Government Bond Data

We start by analysing some basic performances of the 3-year maturity Australian government bond yield data. The daily yield time series $r(t)$ across the investigation period (24/03/2014 to 05/04/2019) is shown in Figure 4.1. A bear run is observed from the beginning of the analysis period from approximately 3% in early 2014 to a low of 1.5% in early 2015, and subsequently, the trend fluctuates roughly between 1.5% to 2%. Debelle [2015] suspects that a global contributor of this yield decline is the global growth prospect; however, the change of the expectation for global growth over the medium term has not varied a lot. Another explanation may lie in the reduced supply of government bonds. Apart from the global factors, the declining trend of Australian government bond yield is also influenced by some domestic factors, including the February reduction in the cash rate target and consequently repricing of market expectations for future monetary policy.

The mean reversion is an essential characteristic of the interest rate. Unlike other financial prices, interest rates cannot rise or drop infinitely. They are forced to drop from very high levels since the economic activity would be hampered in that case. Similarly, it is not common to have extreme negative interest rates for a long time. Thus, interest rates tend to show mean-reverting in the long run and move in a limited range.

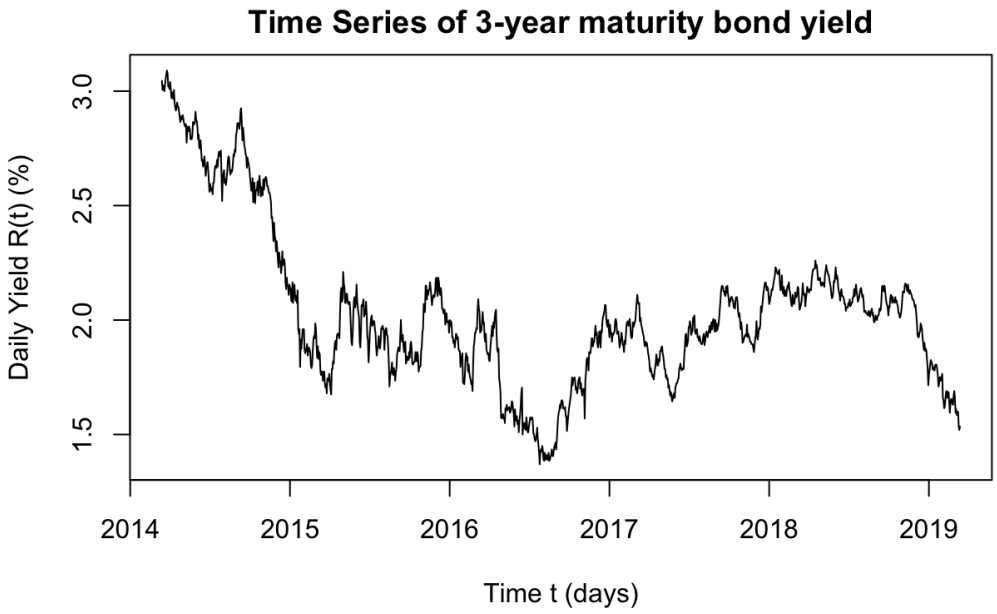

Figure 4.1: Time Series plot of daily yield of 3-year government bond data for the period 24/03/2014 to 05/04/2019. A significant decrease of bond yield is visualised from the beginning of 2014 to the beginning of 2015.

We test the mean reversion property using the Augmented Dickey-Fuller (ADF) Test, inspired by Chan [2013]. We use the *ur.df* function in R and specify that the test starts from lag 1. This is because the AR(1) auto-regressive feature in the discretised version of the relevant SDE is the reason for the mean-reverting feature in the stochastic models [Radkov, 2010].

The null hypothesis is that the process is mean-reverting is not rejected in the given time series. The ADF statistic obtained by the 3-year bond yield data is -2.6014. The fifth and tenth percentiles are -2.86 and -2.57. This means that the test fails to reject the null hypothesis at the 0.05 significance level, but the test rejects the null hypothesis at the 0.1 significance level. It seems reasonable to treat the 3-year bond yield data as mean-reverting.

Given the above analysis, we investigate the bond yield data using two mean-reverting models: Vasicek and CIR. The daily bond yield data is collected for calendar days in a year, excluding the weekends and public holidays. For simplicity, we denote the time discretisation as $\Delta = \frac{1}{252}$

given the general financial days in a year of 252 days.

4.2 Parameter Estimation

In this section, we estimate the parameters of the proposed models (the Vasicek model and the CIR model) using the parameter estimation computation scheme developed in Section 3.5. We also compare and evaluate the performance and goodness of fit of the two estimated models.

First, some modifications are made based on the parameter estimation scheme presented in Section 3.5 to tailor the algorithm to our application. The scheme for the Vasicek model is of the same logic as the parameter estimation scheme for the CIR model that we previously developed, but with a modification of the diffusion term. Now we present some general assumptions and modifications for the estimation scheme:

1. Define the time discretisation step $\Delta = \frac{1}{252}$ and let the initial value be the first value of the true data $r_0 = 3.045$.

2. Guarantee the consistency of the estimated parameters: to obviate the error due to the initial value selection and ensure that the choice of the initial value does not affect the final optimisation result, we run the algorithm 5000 times with various initial parameter values and verify the consistency of the estimated parameters.

3. Use both the Kessler (Gaussian) method and the skew-normal method introduced in Section 3.5 to approximate the parameters and compare the results.

4. The value of κ_3 of the CIR model was derived by equation (3.73) in Section 3.5, whereas κ_3 of the Vasicek model is 0 calculated by the cumulants approximation formula derived in equation (3.61) with a truncation of the terms after $O(\Delta^2)$. The average of

the skewness at each discretisation point is calculated for both models, denoted as $\bar{\rho}_3 = \frac{1}{N} \sum \frac{\kappa_3}{\kappa_2^{3/2}}$.

The estimated parameters for the Vasicek model and the CIR model are listed in Table 4.1 and Table 4.2. The estimations are verified to be consistent based on the 5000 optimisation results with various initial parameter values $\boldsymbol{\theta_0}$. The estimation results of the Kessler method and the SN method are exactly the same for the Vasicek model, whereas some distinctions between the two methods are detected for the CIR model.

Estimated Parameters for the Vasicek model (MLE)				
Method	θ_1	θ_2	θ_3	$\bar{\rho}_3$
Kessler (Gaussian)	2.02197	1.88665	0.56865	0
Skew-normal (SN)	2.02197	1.88665	0.56865	0

Table 4.1: Estimated parameters and skewness for the Vasicek model using the Kessler method and the skew-normal method, respectively. $\bar{\rho}_3$ is the average of the skewness at each discretisation point of the Vasicek model.

Estimated Parameters for the CIR model (MLE)				
Method	θ_1	θ_2	θ_3	$\bar{\rho}_3$
Kessler (Gaussian)	2.21824	1.89988	0.40457	0.06721
Skew-normal (SN)	2.21834	1.89973	0.40457	0.06721

Table 4.2: Estimated parameters and skewness for the CIR model using the Kessler method and the skew-normal method, respectively. $\bar{\rho}_3$ is the average of the skewness at each discretisation point of the CIR model.

Discussion on Table 4.1 and Table 4.2:

1. From Table 4.1, exactly the same parameter estimations from the two methods are expected for the Vasicek model. In Chapter 3, we introduce that when the skewness is 0, the skew-normal distribution is actually Gaussian distribution. The same parameter estimation result from the two methods for the Vasicek model verifies this conclusion.

2. From Table 4.2, few differences are observed between the two models, given a small positive standardized skewness 0.06721. Specifically, the estimations of diffusion coefficients θ_3 are exactly the same, whereas the differences are found in drift coefficients. This corresponds to our previous analysis, that the diffusion coefficient tend to be more consistent compared with the drift terms. Overall, the two estimations are consistent, and we will select the skew normal estimated parameters for the further analysis. The estimated models are shown below:

$$\textbf{Vasicek Model}: \ dr(t) = 2.02197(1.88665 - r(t))\ dt + 0.56865\ dW(t), \tag{4.3}$$

$$\textbf{CIR Model}: \ dr(t) = 2.21834(1.89973 - r(t))\ dt + 0.40457\sqrt{r(t)}\ dW(t). \tag{4.4}$$

Given the estimated parameters $\hat{\boldsymbol{\theta}}$ for the two models, it is worth analysing the goodness of fit of the estimated models by comparing them with the true data set. Since both the Vasicek model and the CIR model are the Itô diffusion process introduced in Chapter 2, we simulate the fitted models using the numerical methods and compare the simulated sample paths with true data. The numerical method of approximating the diffusion process scheme presented in Section 2.6 is modified and applied to the estimated models. A brief procedure for this simulation for the CIR model and the Vasicek model is as follows:

1. Simulate the estimated models $n = 500$ times with the same initial value $r_0 = 3.045$, but different white noise terms $\widetilde{\Delta}W$ for each simulation.

2. The total number of observations in each simulation: $N = 1260$, excluding the initial

value r_0.

3. Choose a numerical method: the Euler-Maruyama method and the Milstein method perform equally well when a finer scale is used. When the scale becomes coarse, the Milstein method performs better (Section 2.6). Given $\Delta = \frac{1}{252}$ and the operating times do not vary significantly in the two methods, we chose the Milstein method to simulate the sample paths.

4. Goodness of fit: generate and plot three types of evaluation criteria based on the simulated sample paths from the two models:

   (a) Average of 500 sample paths at each discretisation points (Mean).

   (b) Confidence interval with 1 standard deviation from the mean (*CI* of *1SD*).

   (c) Confidence interval with the maximum and minimum values among 500 simulations at each discretisation step (*CI* of *Maxmin*).

The plots of the three assessing criteria and the true yield data for the CIR model and the Vasicek model are shown in Figure 4.2. Next, we present some discussions on the simulation results.

Discussion on Figure 4.2:

Both plots in Figure 4.2 show that the overall trends of the true data (green line) are tracked closely by the average of the simulated processes (red line). Both red lines approach around 1.9% as time goes by, which demonstrates the mean-reverting properties of the Vasicek model and the CIR model. In the long run, the future trajectories of $r(t)$ evolves around the mean level θ_2 (1.88665 for Vasicek and 1.89973 for CIR). The *CI* of *1SD* (blue lines) include the true data most of the time, with some exceptions at extremely fluctuating intervals (mid-2014 and mid-2016). The *CI* of *Maxmin* of each discretisation point contains almost the full range

of true data. The Vasicek model generally provides a lower prediction bound compared with the CIR model. Overall, both models show good tracking of the true data, and hence, the goodness of fits are adequate. We are still interested in comparing the two models to select the best model for real data application. We then perform a parametric bootstrap hypothesis testing based on the two sets of simulated sample paths of the two models, as presented in Section 4.3.

Figure 4.2: (left) Comparison between the true data and simulations of sample paths, using the Vasicek model with parameters $\hat{\boldsymbol{\theta}} = (2.02197, 1.88665, 0.56865)$ and (right) comparison between the true data and simulations of sample paths, using the CIR model with parameters $\hat{\boldsymbol{\theta}} = (2.21834, 1.89973, 0.40457)$. The sample paths are all simulated with $r_0 = 3.045$ and $\Delta = \frac{1}{252}$.

4.3 Parametric Bootstrap Hypothesis Testing

In this section, we aim to decide whether it is better to use the Vasicek model or the Cox-Ingersoll-Ross model for the bond yield application. In particular, we conduct parametric bootstrap hypothesis testing for non-nested models to determine whether one model provides a better fit than the other.

Bootstrap hypothesis testing is different from traditional hypothesis testing, because it is usually simulation-based rather than based on exact or asymptotic distribution theory. Simulation-generated p-values are essential tools in data analysis. Efron and Tibshirani [1994] and Davison and Hinkley [1997] analysed the effectiveness of bootstrap both theoretically and numerically. Hall and Wilson [1991] provided two guidelines for the bootstrap hypothesis testing building on previous literature, such as [Young, 1986] and [Beran, 1988]. The first guideline states a preference to a re-sampling method that reflects the null hypothesis, even if the data might be drawn from a population that fails to satisfy the null hypothesis. Thus, it has the effect of increasing power. The second guideline states that the bootstrap hypothesis testing should focus on methods recognised as possessing good characteristics for confidence interval construction problems. This second guideline is closely related to the level error reduction and the accuracy of the test. There are two versions of bootstrap hypothesis testing: parametric and non-parametric. With the former, simulation is from a parametric distribution, while in the latter simulation is from the sample. In both cases, simulation should be conducted under the null hypothesis. Ventura [2010] summarised some practical procedures of parametric bootstrap hypothesis testing.

To perform bootstrap hypothesis testing, we use the simulated 500 samples paths of the estimated Vasicek model and the estimated CIR model in Section 4.2. This leads to the construction of **parametric** bootstrap samples, since we are simulating from the known estimated

models. Therefore, we build up the scheme of parametric bootstrap hypothesis testing using the simulated bootstrap samples and the true data, using the log-likelihood ratio as the test statistic.

Assume the null hypothesis that the Vasicek model is correct, to be tested against the alternative hypothesis that the CIR model is correct. The intuition under this test is: given the true performance of the two models, how extreme the CIR model outperforms the Vasicek model in bootstrap samples. First, we calculate the maximised log-likelihoods of the true data fitted under the estimated Vasicek model and the estimated CIR model. The re-sample used in this test is the 500 bootstrap samples generated from the estimated Vasicek model, as we believe the true model is the Vasicek model. We then evaluate the maximised log-likelihoods of the bootstrap samples fitted under the Vasicek model and the CIR model separately. Consequently, we can compare the ratio of the maximised log-likelihoods of the Vasicek and CIR models from each bootstrap sample with the ratio from the true data.

The test procedure is as following:

1. Null hypothesis H_0: the Vasicek model is correct.

   Alternative hypothesis H_A: the CIR model is correct.

2. Calculate test statistics: maximum log-likelihood ratio Λ.

   For each bootstrap simulated sample under H_0 (the Vasicek model is correct), $i = 1, 2, \cdots, 500$, calculate:

   $l_i(V)$: maximised log-likelihood fitted by the Vasicek model;

   $l_i(C)$: maximised log-likelihood fitted by the CIR model;

   $\Lambda_i = l_i(C) - l_i(V)$: maximum log-likelihood ratio of bootstrap sample i.

   Given the real data, calculate:

   $l_0(V)$: maximised log-likelihood fitted by the Vasicek model;

$l_0(C)$: maximised log-likelihood fitted by the CIR model;

$\Lambda_0 = l_0(C) - l_0(V)$: maximum log-likelihood ratio of real data (observed test statistic).

3. Calculate the p-value:

   Calculate i^*, the number of bootstrap samples, such that $\Lambda_i > \Lambda_0$. Then the p-value is

$$p = \frac{\text{Number of bootstraps such that } \{\Lambda_i > \Lambda_0\}}{(\text{Total number of bootstrap samples} + 1)} = \frac{i^*}{501}. \tag{4.5}$$

4. Draw conclusions: use the significance level $\alpha = 5\%$, when $p > \alpha$, there is no evidence to reject the null hypothesis, and we should conclude that the Vasicek model is correct. On the contrary, when $p < \alpha$, there is strong evidence to reject the null hypothesis and in favour of the CIR model.

The procedure is based on the null hypothesis that the Vasicek model is true. Since the procedure is not symmetric with respect to the Vasicek model and the CIR model, we repeat the procedure under the null hypothesis that the CIR model is correct. In the latter case, we use the sample paths simulated from the estimated CIR model to conduct the bootstrap hypothesis testing. The test statistic in this case is calculated by subtracting the maximised log-likelihoods fitted in the Vasicek model from the maximised log-likelihoods fitted in the CIR model.

The calculation procedure is completed in R and MATLAB. We first extract the simulated bootstrap sample paths from two estimated models generated in R in Section 4.2. Then we evaluate $l_i(V)$ and $l_i(C)$ for $i = 0, 1, \cdots, 500$ using the optimisation algorithm built in MATLAB. Based on those bootstrap samples, we construct two parametric bootstrap hypothesis tests with different null hypothesis: the Vasicek null hypothesis in the former case and the CIR null hypothesis in the latter case. The observed test statistic Λ_0 and test statistic of each simulation Λ_i are evaluated, and hence, the p-value is calculated based on the number of bootstrap samples i^*, such that $\Lambda_i > \Lambda_0$. The results are shown in Table 4.3:

Parametric bootstrap hypothesis tests		
H_0	Vasicek model is correct	CIR model is correct
$l_0(V)$	2410.090249	2410.090249
$l_0(C)$	2399.909485	2399.909485
Λ_0	-10.18076408	10.18076408
$i^* = \sum_i \{\Lambda_i > \Lambda_0\}$	331	0
Bootstrap sample size	500	500
p	0.662	0

Table 4.3: Construction and calculation of the parametric bootstrap hypothesis testing, where the second and third columns show results of the former hypothesis test given the Vasicek model with parameters $\hat{\boldsymbol{\theta}} = (2.02197, 1.88665, 0.56865)$ and latter hypothesis test given the CIR model with parameters $\hat{\boldsymbol{\theta}} = (2.21834, 1.89973, 0.40457)$, respectively.

Discussions of Table 4.3:

The maximised log-likelihoods of the true data $(l_0(V), l_0(C))$ are fixed in the two tests; however, the observed test statistics Λ_0 are opposite values of each other, since we reverse H_0 and H_A in the two tests. We are interested in the simulation-generated p-values, which are calculated as follows for the two tests :

$$p = Pr(\Lambda_i > -10.1808 | \text{ Vasicek model is true}) \tag{4.6}$$

and

$$p = Pr(\Lambda_i > 10.1808 | \text{ CIR model is true}). \tag{4.7}$$

The p-value indicates the probability of observing $\Lambda_i > \Lambda_0$ given that H_0 is true. We observe 331 samples with the larger log-likelihood ratios compared with the log-likelihood ratio of

the real data if we assume that the Vasicek model is true, and no sample is found to have a larger log-likelihood ratio compared with the log-likelihood ratio of the real data if the CIR model is assumed to be correct. This gives p-values of 0.662 and 0, respectively. For the former hypothesis test, the p-value of 0.662 is far greater than the significance level $\alpha = 5\%$, so we fail to reject H_0. The latter test has a p-value of 0, which is significantly smaller than 5% and suggests that we have strong confidence to reject the null hypothesis. Therefore, both hypothesis tests lead to the conclusion: the Vasicek model is better than the CIR model, and we should choose the Vasicek model as the model to approximate the bond yield data.

To better visualise the result, the densities of test statistics $\Lambda_i \, (i = 1, 2, \cdots, 500)$, calculated from 500 bootstrap samples, are plotted in Figure 4.3 and Figure 4.4. Some analysis procedures are inspired by [Ventura, 2010].

Discussion of Figure 4.3 and Figure 4.4:

These two figures are density plots of Λ_i of the two hypothesis tests when H_0 is true. They are actually the parametric bootstrap null distributions of the test statistics under H_0, which assumes either the Vasicek model is correct or the CIR model is correct. The red vertical lines represent the observed test statistics Λ_0. The area under the density curve to the right of Λ_0 is calculated as p-value. Moreover, it is evident that $p = 66.2\%$ and $p = 0$ for two hypothesis tests, respectively, which lead to the same conclusion at 5% significance level. We conclude that the Vasicek model is superior to the CIR model with this data set.

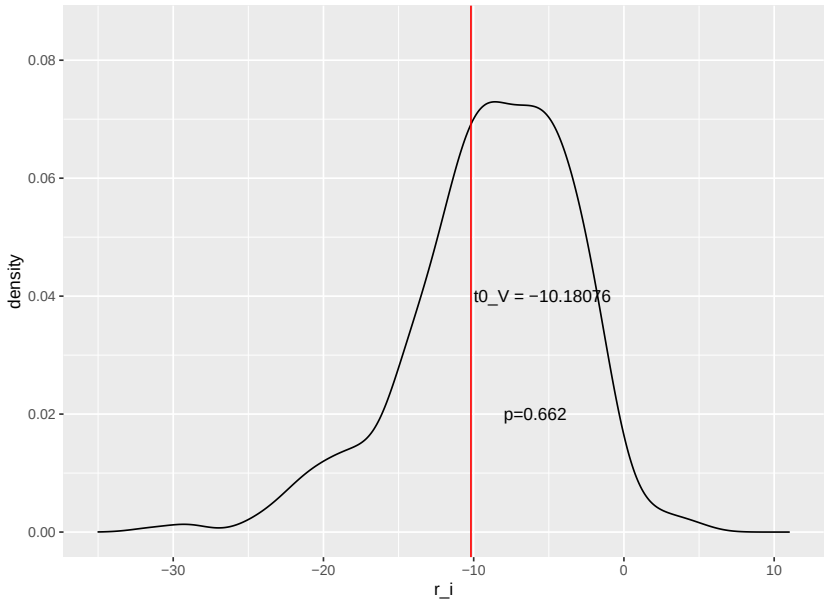

Figure 4.3: Density plot of the Monte-Carlo test statistics of parametric bootstrap samples where the null hypothesis is that the Vasicek model is true. The parameters of the estimated models from the original data set are $\hat{\boldsymbol{\theta}} = (2.02197, 1.88665, 0.56865)$, $r_0 = 3.045$ and $\Delta = \frac{1}{252}$.

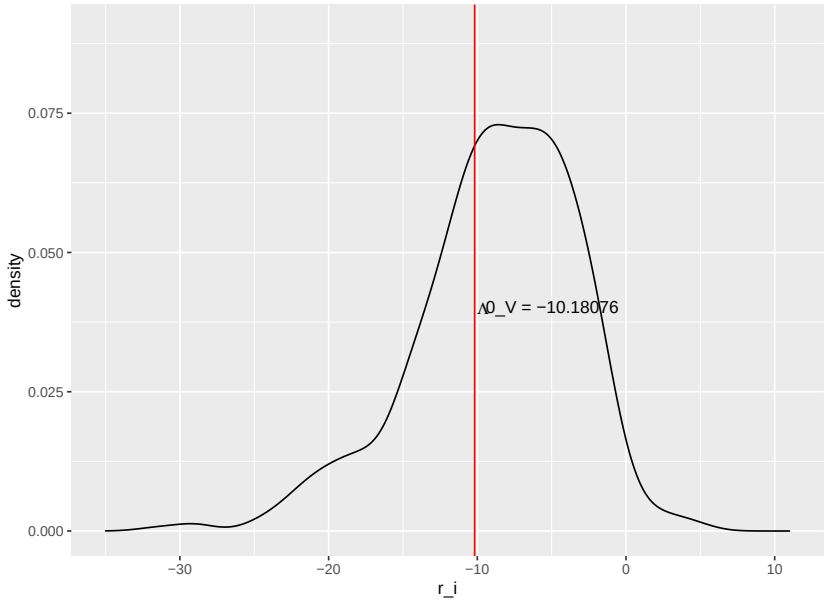

Figure 4.4: Density plot of the Monte-Carlo test statistics of parametric bootstrap samples where the null hypothesis is that the CIR model is true. The parameters of the estimated models from the original data set are $\hat{\boldsymbol{\theta}} = (2.21834, 1.89973, 0.40457)$, $r_0 = 3.045$ and $\Delta = \frac{1}{252}$.

We now look into the approximate confidence interval test of the Vasicek model applying the Empirical Rule in statistics. Since the parametric bootstrap hypothesis testing suggests that the Vasicek model is correct, we expect $\hat{p}$, the probability that the estimated true parameters lie within two standard deviations from the mean of the bootstrap estimated parameters is around 95%. In our case, it is expected to exhibit approximately 25 bootstrap samples (out of 500) that fail to include the true parameter estimations into their two standard deviations bounds. $\hat{p}$ is calculated based on the approximated true coverage from the bootstrap samples, and thus, a 95% confidence interval of this probability given its variance $\frac{\hat{p}(1-\hat{p})}{500}$ is obtained. It is crucial to make sure all values are calculated under the Vasicek model, which means all the bootstrap samples are simulated from the Vasicek model and then are fitted into the Vasicek model in the parameter estimation scheme. The variance of each bootstrap sample is calculated from the Hessian matrix of the maximum likelihood estimation using the *fminunc* function in MATLAB.

One issue with this test is, given $N = 1260$ observations in the original data set and the 500 bootstrap samples, the estimated parameters may not be sufficiently converged, which leads to a less accurate approximation. In other words, this suggests that the first-order asymptotic approximations need quite a large sample size before they attain good accuracy. And this phenomenon is more likely to be found in the drift terms, which represent the long term mean level. Recall the parameter estimation result in Table 3.1 and Table 3.2 in Section 3.5.2. When the sampling window is fixed, and the sampling interval is decreasing, the consistency of the drift terms is hardly achieved, in contrast to the significant improvement in the diffusion terms. This is because the drift terms are the parameters representing long term mean level behaviour, and they are less prone to instantaneous change. To analyse the asymptotic behaviour of the three parameters, we extend the number of observations to $N' = 5000$ and keep the sampling interval at $\Delta = \frac{1}{252}$ when simulating from the proposed Vasicek model.

Table 4.4 and Table 4.5 show the results of the confidence interval tests based on the bootstrap

samples with $N = 1260$ and $N' = 5000$ observations, respectively.

Discussion of Table 4.4 and Table 4.5

In table 4.4, less than 95% of the bootstrap samples manage to include the true parameters, especially θ_2 that only covers the true value with a probability 85.4%. This result is undesirable, but under our expectation. The relatively small sample size ($N = 1260$) and the fixed discretisation step Δ leads to a short observation window, which is not sufficient for the convergence of the estimated parameters.

A much more accurate approximation result is obtained by increasing the number of observations to $N' = 5000$ and fixing the sampling interval at $\Delta = \frac{1}{252}$, which leads to a larger sampling window. In table 4.5, $\hat{p}$ of the three parameters are close to the 95% bar, which suggests an increasing approximation accuracy as the sample size and sampling window increase. A dramatic increase in the approximation accuracy is visualised in θ_2, from 85.4% to 92%. This coincides with the previous analysis, that the approximation accuracy of the drift terms is more prone to the sample size and sampling window, rather than the sampling interval. The coverage of the diffusion parameter θ_3 also improves, from 90.8% to 94%. This indicates that the performance of the diffusion term is not only influenced by the sampling interval (likewise in Section 3.5.2), but also the sample size and sampling window.

Overall, the approximate confidence interval tests show an adequate estimation performance of the Vasicek model, in particular when the sample size and sampling window are large. This suggests that when using these approximate confidence intervals, some caution is required. We visualise that for smaller sample sizes, there tends to be under coverage and less accurate approximation result. In order to achieve the asymptotic approximation, the sample size needs to be quite large.

Confidence Interval Test ($N = 1260$)			
$\boldsymbol{\theta}$	θ_1	θ_2	θ_3
Counts of the true parameter outside the boundary of 500 bootstrap	36	72	45
Probability of the true parameter within the estimated bootstrap bound ($\hat{p}$)	92.6%	85.4%	90.8%
CI of the probability $\left(\hat{p} - 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{500}}, \hat{p} + 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{500}}\right)$	(90.3%,94.9%)	(82.3%,88.5%)	(88.3%,93.3%)

Table 4.4: Simulated confidence intervals for the probability that the true parameters lie within two standard deviations from the mean of the bootstrap estimated parameters, tested under 500 bootstrap samples with $N = 1260$ observations from the Vasicek model with parameters $\hat{\boldsymbol{\theta}} = (2.02197, 1.88665, 0.56865)$.

Confidence Interval Test ($N' = 5000$)			
$\boldsymbol{\theta}$	θ_1	θ_2	θ_3
Counts of the true parameter outside the boundary of 500 bootstrap	28	39	29
Probability of the true parameter within the estimated bootstrap bound ($\hat{p}$)	94.2%	92.0%	94.0%
CI of the probability $\left(\hat{p} - 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{500}}, \hat{p} + 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{500}}\right)$	(92.2%,96.3%)	(89.6%,94.4%)	(91.9%,96.1%)

Table 4.5: Simulated confidence intervals for the probability that the true parameters lie within two standard deviations from the mean of the bootstrap estimated parameters, tested under 500 bootstrap samples with $N' = 5000$ observations from the Vasicek model with parameters $\hat{\boldsymbol{\theta}} = (2.02197, 1.88665, 0.56865)$.

5 Summary, Conclusions and Future Research

This thesis proposes a new skew-normal approximation method for estimating parameters in SDE models, and develops a parameter estimation computation scheme and a numerical method simulation algorithm for general Itô diffusion processes, which has been applied to an interest rate application involves real data. Parameter estimation methods and the numerical methods to solutions of Itô diffusion processes are important because SDE models are widely used in modelling random quantities in finance and physics. The proposed skew-normal parameter estimation method achieved good estimates of the parameters, comparable to those of the Kessler method, at a small skewness. The estimation of the diffusion coefficient is also shown to be more accurate in the parameter estimation scheme as the discretisation becomes smaller. We also develop a simulation algorithm for the solution of general Itô diffusion processes and analysed the accuracy of two numerical methods across various grid widths. Finally, we modelled a 3-year Australian government bond yield data with the Vasicek model and the Cox-Ingersoll-Ross model using the parameter estimation algorithm and achieved some good model fitting results.

Section 5.1 presents the main contributions of this thesis. Then, in Section 5.2, we describe some possible future research ideas.

5.1 Main Contributions

It is not feasible in general to use exact maximum likelihood estimation to estimate the parameters of an Itô diffusion process because the transition density is unknown for most SDEs. We develop a skew-normal approximation method which extends the Gaussian approximation method [Kessler, 1997] by accommodating skewness. We modify the parameter estimation scheme proposed by Lu et al. [2021] for both the Kessler method and the SN method, and examine the accuracy of the estimated parameters. The accuracy of the diffusion coefficient is shown to be more prone to the changes in the discretisation step than the drift coefficients. We also build a simulation algorithm using the Euler method and the Milstein method under the general Itô diffusion processes. Further, we compare the accuracy of the two methods across various grid widths, where a preference to the Milstein method is generalised when the grid width is relatively large. Finally, we fit a proper model of a 3-year Australian government bond yield data using the parameter estimation scheme and the numerical method simulation scheme. The goodness of fit is verified by some visualisations and parametric bootstrap hypothesis testing.

5.2 Future Analysis

This thesis can be generalised in several ways. For the simulation scheme of numerical methods, one may enforce a boundary to the scheme to limit the value of realisations, e.g. if the process is non-negative, as is the case with CIR. Then, by specifying a certain stochastic model, this can be applied to some financial applications, such as the down-out option or the up-out option.

One may also consider an in-depth analysis of the skew-normal approximation to the transition

density for the parameter estimation. Specifically, the performance of the SN distribution compared with the Gaussian distribution for cases with a large skewness.

Furthermore, because the CIR model and Vasicek model both fit the bond yield data well (although the Vasicek model has been demonstrated to be better), one may be interested in fitting other interest rate models to this data set to compare the performance. Other parameter estimation methods can also be attempted, such as the methods introduced in [Aït-Sahalia, 2002] and [Shoji and Ozaki, 1998] to verify whether these methods can achieve a better fitting of the data.

Bibliography

Yacine Ait-Sahalia. Transition densities for interest rate and other nonlinear diffusions. In *Quantitative Analysis In Financial Markets: Collected Papers of the New York University Mathematical Finance Seminar (Volume II)*, pages 1–34. World Scientific, 2001.

A. Azzalini. A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, 12(2):171–178, 1985. ISSN 03036898, 14679469. URL http://www.jstor.org/stable/4615982.

Yacine Aït-Sahalia. Maximum likelihood estimation of discretely sampled diffusions: A closed-form approximation approach. *Econometrica*, 70(1):223–262, 2002. doi: https://doi.org/10.1111/1468-0262.00274. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/1468-0262.00274.

Rudolf Beran. Prepivoting test statistics: a bootstrap view of asymptotic refinements. *Journal of the American Statistical Association*, 83(403):687–697, 1988.

Fischer Black and Piotr Karasinski. Bond and option pricing when short rates are lognormal. *Financial Analysts Journal*, 47(4):52–59, 1991.

Fischer Black and Myron Scholes. The pricing of options and corporate liabilities. *The Journal of Political Economy*, pages 637–654, 1973.

Alexandre Brouste, Masaaki Fukasawa, Hideitsu Hino, Stefano M. Iacus, Kengo Kamatani, Yuta Koike, Hiroki Masuda, Ryosuke Nomura, Teppei Ogihara, Yasutaka Shimuzu, Masayuki Uchida, and Nakahiro Yoshida. The yuima project: A computational framework for simulation and inference of stochastic differential equations. *Journal of Statistical Software*, 57(4):1–51, 2014. doi: 10.18637/jss.v057.i04. URL http://www.jstatsoft.org/v57/i04/.

E. Chan. *Algorithmic Trading: Winning Strategies and Their Rationale*. Wiley Trading. Wiley, 2013. ISBN 9781118460146. URL https://books.google.com.au/books?id=WAlFDwAAQBAJ.

John C. Cox, Jonathan E. Ingersoll, and Stephen A. Ross. A theory of the term structure of interest rates. *Econometrica*, 53(2):385–407, 1985. ISSN 00129682, 14680262. URL http://www.jstor.org/stable/1911242.

Didier Dacunha-Castelle and Danielle Florens-Zmirou. Estimation of the coefficients of a diffusion from discrete observations. *Stochastics: An International Journal of Probability and Stochastic Processes*, 19(4):263–284, 1986.

A.C. Davison and D.V. Hinkley. *Bootstrap Methods and Their Application*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1997. ISBN 9780521574716. URL https://books.google.com.au/books?id=4aCDbm_t8jUC.

Guy Debelle. Global and domestic influences on the australian bond market photograph. [Online]. Available from: https://www.rba.gov.au/speeches/2015/sp-ag-2015-03-16.html, March 16 2015.

B. Efron and R.J. Tibshirani. *An Introduction to the Bootstrap*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis, 1994. ISBN 9780412042317. URL https://books.google.com.au/books?id=gLlpIUxRntoC.

Robert F Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the econometric society*, pages 987–1007, 1982.

Danielle Florens-Zmirou. Approximate discrete-time schemes for statistics of diffusion processes. *Statistics*, 20(4):547–557, 1989. doi: 10.1080/02331888908802205. URL https://doi.org/10.1080/02331888908802205.

Aaron L Fogelson. A mathematical model and numerical method for studying platelet adhesion and aggregation during blood clotting. *Journal of Computational Physics*, 56(1): 111–134, 1984.

V Genon-Catalot. Maximnm contrast estimation for diffusion processes from discrete observations. *Statistics*, 21(1):99–116, 1990.

Emmanuel Gobet. Lan property for ergodic diffusions with discrete observations. *Annales de l'Institut Henri Poincare (B) Probability and Statistics*, 38(5):711–737, 2002. ISSN 0246-0203. doi: https://doi.org/10.1016/S0246-0203(02)01107-X. URL https://www.sciencedirect.com/science/article/pii/S024602030201107X.

Arsalane Chouaib Guidoum and Kamal Boukhetala. Performing parallel monte carlo and moment equations methods for itô and stratonovich stochastic differential systems: R package Sim.DiffProc. *Journal of Statistical Software*, 96(2):1–82, 2020. doi: 10.18637/jss.v096.i02.

P. Hall and S. Wilson. Two guidelines for bootstrap hypothesis testing. *Biometrics*, 47:757–762, 1991.

Peter Hall. Theoretical comparison of bootstrap confidence intervals. *The Annals of Statistics*, pages 927–953, 1988.

S.M. Iacus. *Simulation and Inference for Stochastic Differential Equations: With R Examples*. Springer Series in Statistics. Springer New York, 2009. ISBN 9780387758398. URL https://books.google.com.au/books?id=ryCMlNVV8EAC.

Mathieu Kessler. Estimation of an ergodic diffusion from discrete observations. *Scandinavian Journal of Statistics*, 24(2):211–229, 1997. doi: https://doi.org/10.1111/1467-9469.00059. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-9469.00059.

P.E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*. Stochastic Modelling and Applied Probability. Springer Berlin Heidelberg, 1992. ISBN 9783540540625. URL https://books.google.com.au/books?id=BCvtssom1CMC.

Kevin W. Lu, Phillip J. Paine, Simon P. Preston, and Andrew T. A. Wood. Approximate discrete-time schemes for statistics of diffusion processes. *Invited revision for Scandinavian Journal of Statistics*, 2021.

Dilip B Madan, Peter P Carr, and Eric C Chang. The variance gamma process and option pricing. *Review of Finance*, 2(1):79–105, 1998.

Gisiro Maruyama. Continuous markov processes and stochastic equations. *Rendiconti del Circolo Matematico di Palermo*, 4(1):48, 1955.

Robert C Merton. Theory of rational option pricing. *The Bell Journal of economics and management science*, pages 141–183, 1973.

Robert C Merton. Optimum consumption and portfolio rules in a continuous-time model. In *Stochastic optimization models in finance*, pages 621–661. Elsevier, 1975.

Robert C Merton. Option pricing when underlying stock returns are discontinuous. *Journal of financial economics*, 3(1-2):125–144, 1976.

GN Milstein. Approximate integration of stochastic differential equations. *Theory of Probability & Its Applications*, 19(3):557–562, 1975.

B. Øksendal. *Stochastic Differential Equations: An Introduction with Applications*. Universitext (1979). Springer, 2003. ISBN 9783540047582. URL https://books.google.com.au/books?id=VgQDWyihxKYC.

Tohru Ozaki. A bridge between nonlinear time series models and nonlinear stochastic dynamical systems: a local linearization approach. *Statistica Sinica*, pages 113–135, 1992.

George C Papanicolaou. Diffusion in random media. In *Surveys in applied mathematics*, pages 205–253. Springer, 1995.

BLS Prakasa Rao. Asymptotic theory for non-linear least squares estimator for diffusion processes. *Statistics: A Journal of Theoretical and Applied Statistics*, 14(2):195–209, 1983.

Petar Radkov. The mean reversion stochastic processes applications in risk management. 07 2010.

René L Schilling. *Measures, integrals and martingales*. Cambridge University Press, 2017.

Isao Shoji and Tohru Ozaki. Estimation for nonlinear stochastic differential equations by a local linearization method. *Stochastic Analysis and Applications*, 16(4):733–752, 1998.

Steven E Shreve. *Stochastic calculus for finance II: Continuous-time models*, volume 11. Springer Science & Business Media, 2004.

Michael Sørensen. Estimating functions for diffusion-type processes. *Statistical methods for stochastic differential equations*, 124:1–107, 2012.

Oldrich Vasicek. An equilibrium characterization of the term structure. *Journal of financial economics*, 5(2):177–188, 1977.

Valérie Ventura. *Bootstrap Tests of Hypotheses*, pages 383–398. Springer US, Boston, MA, 2010. ISBN 978-1-4419-5675-0. doi: 10.1007/978-1-4419-5675-0_18. URL https://doi.org/10.1007/978-1-4419-5675-0_18.

Wolfgang Wagner and Eckhard Platen. *Approximation of It ô integral equations*. Zentralinst. für Mathematik u. Mechanik d. Akad. d. Wiss. d. DDR, 1978.

Nakahiro Yoshida. Estimation for diffusion processes from discrete observation. *Journal of Multivariate Analysis*, 41(2):220–242, 1992.

Alastair Young. Conditioned data-based simulations: some examples from geometrical statistics. *International Statistical Review/Revue Internationale de Statistique*, pages 1–13, 1986.