The University of Adelaide

Essays on Continuous Time Diffusion Models

A DISSERTATION

SUBMITTED TO THE GRADUATE CENTRE
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Economics

By

Di Yuan

November 2013

Declaration

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution in my name and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I give consent to this copy of my thesis when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

The author acknowledges that copyright of published works contained within this thesis resides with the copyright holder(s) of those works. I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library catalogue and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

_____

Signature of Author

ABSTRACT

Essays on Continuous Time Diffusion Models

Di Yuan

During the past few decades, continuous time diffusion models have become an integral part of financial economics. Especially, in certain core areas in finance, such as interest rate, asset pricing, option pricing, portfolio selection and volatility modelling, continuous time diffusion models have proved to be a very attractive way to conduct research and gain economic intuition. This thesis makes three main contributions to the field of continuous time diffusion models.

First, we propose regime-switching Heston, GARCH, and CEV stochastic volatility models where all parameters are allowed to vary depending on the state of the economy. Then we apply these models to describe the dynamics of S&P 500 and VIX. We find strong evidence of regime shifts for all models. The CEV model is statistically preferred to other two nested models in explaining dynamics of data.

Second, because the true transition density functions of regime-switching stochastic volatility models are unknown, the standard maximum likelihood estimation cannot be conducted. We first conduct the maximum likelihood estimation with closed-form likelihood expansions for regime-switching continuous time stochastic volatility models.

Third, to approximate a continuous time diffusion process, researchers often use the Euler approximation in the literature. Theoretically, the smaller the discretization interval is, the more accurate the Euler approximation is expected to be. However, even when the discretization interval is too small, the accuracy of the Euler approximation can get worse because of the roundoff error and random number generator bias. A variety of univariate and multivariate diffusion models from the literature are considered. We use the solution of a diffusion process when it is available and

usable as a benchmark. The Milstein approximation is also adopted to compare the accuracy of the Euler approximation. Depending on the problem of interest, different criteria are used to measure accuracy of approximation. The percentage error and strong convergence can be examined when a good approximation of sample path of a diffusion model is required. The weak convergence is preferred for the cases where approximation of moments of the process matters. In our Monte Carlo simulation studies of diverse diffusion models, we measure accuracy of the Euler approximation not only by using those criteria but also by looking at end point of the approximation. The simulation results show that an appropriate discretization interval must be picked for different diffusion models when applying the Euler approximation.

Acknowledgements

I would like to take this opportunity to express my sincere gratitude to a number of people, who have provided me with useful assistance and support during my PhD study.

First and foremost, I want to thank my principle supervisor Dr Seungmoon Choi. It has been an honour to be his first PhD student. He provided unreserved support during my PhD study and generously paved the way for my development as a research economist.

I would like to thank my co-supervisor Professor Jiti Gao for his kind assistance and support on academic and various other issues. I also wish to thank Associate Professor Ralph Bayer, who helped me book a computer lab to run simulations. I also thank Dr Nicholas Sim for his helpful suggestions when he attended my presentation. Special thanks go to Professor Christopher Findlay, Dr Mandar Oak, Dr Jacob Wong, Dr Tatyana Chesnokova and Mrs Anne Arnold for their assistance with administrative issues on various occasions.

I would like to thank all my friends for all the emotional support, camaraderie, entertainment, and caring they provided.

I wish to thank my brother Wang Yuan, who has been my best friend all my life and thank him for all his advice and support.

This thesis is dedicated to my mother Li Liu, who has provided me with absolute and unconditional support and encouragement throughout my PhD study. Without her, this thesis would not have been completed.

Table of Contents

List of Tables

List of Figures

CHAPTER 1

Introduction

During the past few decades, continuous time diffusion models have become an integral part of financial economics. Especially, in certain core areas in finance, such as interest rate, asset pricing, option pricing, portfolio selection and volatility modelling, continuous time diffusion models have proved to be a very attractive way to conduct research and gain economic intuition (Sundaresan (2000)).

This thesis makes three main contributions to the field of continuous time diffusion models by addressing the following research questions.

(1) Could we combine regime shifts with continuous time stochastic volatility models to explain the dynamics of multivariate diffusion models?
(2) How do we conduct maximum likelihood estimation to estimate regime-switching stochastic volatility models where the true transition density functions are unknown?
(3) How do we choose an appropriate time interval for a continuous time diffusion model when approximating a continuous time diffusion process using the Euler approximation method?

The following part will explain the motivation behind each research question in order.

The motivation of combining regime shifts with continuous time stochastic volatility models originally comes from the plots of the first difference of daily S&P 500 and VIX from 1990 to 2012. They tend to exhibit periods of high and low volatility, which is a stylized phenomenon called volatility clustering in the financial time series. Most periods of high volatility clustering can be associated with some significant historic events. Regarding S&P 500 and VIX data, there seems to be at least two regimes, namely, a high regime and a low regime. Furthermore, since Hamilton (1989)'s work, a substantial number of econometric literature have shown the evidence of regime shifts existing in financial series data and regime-switching models outperforming regular single regime models. For example, Hamilton and Sunmel

(1994) propose the regime-switching ARCH models, Fornari and Mele (1996) develop the regime-switching GARCH models, Calvet and Fisher (2004) introduce discrete-time regime-switching stochastic volatility models, Choi (2009) suggests the regime-switching univariate diffusion models. However, there still remains a big gap to combine regime shifts with multivariate continuous time diffusion models. To fill this gap, we propose regime-switching Heston, GARCH, and CEV continuous time stochastic volatility models. What is more, our models are distinguished from other recently proposed regime-switching stochastic volatility models (Mitra (2010), Durham and Park (2013), Papanicolaou and Sircar (2013)) by allowing all parameters of the drift and the volatility components to vary depending on the state of the economy.

Only a few of literature combine regime shifts with multivariate continuous time stochastic volatility models because of the substantial difficulties coming from estimation of multivariate stochastic volatility models. The biggest challenge is that the transition probability density of a diffusion process is hardly ever known in closed form so that the standard maximum likelihood estimation cannot be conducted. Many different methods have been proposed to approximate the transition probability densities. Examples are numerical solution of the Kolmogorov partial differential equation (Lo (1988)), Monte Carlo simulation (Pedersen (1995)), binomial or other trees (Jensen and Poulsen (2002)), and the Euler approximation (Kleppe, Yu, and Skaug (2010)). However, neither of these methods produces a closed form density function. To solve this issue, Aït-Sahalia (2002) and Aït-Sahalia (2008) propose a specific closed-form approximation approach for the univariate and multivariate continuous time diffusion models, respectively. Furthermore, his work has been extended to univariate time homogeneous diffusion models driven by Lėvy processes (Schaumburg (2001)), to univariate time inhomogeneous diffusion models(Egorov, Li, and Xu (2003)), to regime-switching time homogeneous univariate diffusion models for short term interest rates (Choi (2009)), to multivariate time homogeneous jump diffusion models (Yu (2007)), to multivariate time inhomogeneous diffusion models (Choi (2013)), to a Bayesian setting (Stramer, Bognar, and Schneider (2010)) and to damped diffusion (Li (2010)). Jensen and Poulsen (2002) compare most commonly used techniques, including Hermite expansion, Euler approximation, simulation-based methods, binomial approximations and numerical solution of Kolmogorov partial differential equations, and find that Aït-Sahalia (2002)'s method is the best in terms of

accuracy and speed trade-off. Because of the outstanding performance of Aït-Sahalia (2008)'s method to obtain approximate transition probability density function of the stochastic volatility models of our interest, we carry out the maximum likelihood estimation for regime-switching continuous time stochastic volatility models by using Hamilton algorithm (Hamilton (1989)).

Although several methods have been proposed to approximate the transition probability densities, the Euler approximation is still the simplest and most commonly used approaches in the literature (Kristensen (2010), Song (2011), Chen and Hong (2011), Zhao (2011), Koo and Linton (2012) and Chen and Song (2013)). Theoretically, the accuracy of the Euler approximation depends on the time discretization step Δ. To be more precise, as Δ goes to 0, the Euler approximation is expected to be closer to the true diffusion process. In a simulation, researchers set different time discretization step Δ for the Euler approximation in order to achieve two main objectives. First, when the objective is to simulate a diffusion process or test statistical estimators, a good pathwise approximation is required. For example, Christensen, Kinnebrock, and Podolskij (2010) first generate a complete high-frequency record of N (N denoting the number of seconds in 6.5 hours) equidistant observations of the efficient price by using the Euler approximation with $\Delta = 1/23400$. Second, when the objective is to compute moments, probabilities and other functions of the diffusion process, a good endpoint approximation is required. For example, Song (2011) uses the Euler approximation to generate data from univariate and multivariate continuous time diffusion models by setting $\Delta = 1/(252 \times 1500)$ for daily frequency, $\Delta = 1/(12 \times 1500)$ for monthly frequency, $\Delta = 1/(4 \times 1500)$ for quarterly frequency and $\Delta = 1/(1 \times 1500)$ for yearly frequency respectively, and then uses these data to analyse the impact of the numerical integral approximation. Chen and Hong (2011) set $\Delta = 1/(12 \times 120)$ at monthly frequency to examine the sizes of specification tests for some multivariate continuous time diffusion models. Kristensen (2010) chooses $\Delta = 1/(252 \times 10)$ at daily frequency for the evaluation of the likelihood. Zhao (2011) approximates the conditional log-likelihood by using $\Delta = 1/252$. Koo and Linton (2012) set $\Delta = 1/101$ to investigate the robustness of estimators under locally stationary diffusion processes. Chen and Song (2013) use subintervals of $\Delta = 1/(52 \times 5)$ for weekly frequency to approximate Lévesque integral. Aït-Sahalia and Kimmel (2007) and Choi (2013) generate data at daily frequency by setting

$\Delta = 1/(252 \times 30)$ to carry out MLE for continuous time diffusion models. To our best knowledge, so far, there has been no literature to study whether these above time intervals are effective enough for the Euler scheme to provide accurate approximation. Theoretically, the smaller the discretization interval is, the more accurate the Euler approximation is expected to be. However, even when the discretization interval is too small, the accuracy of the Euler approximation can get worse because of the roundoff error (Peter and Platen (1992)) and random number generator bias (Komori, Saito, and Mitsui (1994)). The random number generator bias primarily results from lack of mutual independence in the samples from a random number generator when Δ is too small. Therefore, Chapter 4 intends to provide some guidelines on the choice of the discretization interval when approximating diffusion processes using the Euler approximation method.

This thesis is organized as follows.

In Chapter 2, based on the work of Aït-Sahalia (2008) and Choi (2009), we conduct the maximum likelihood estimation with closed-form likelihood expansions for regime-switching continuous time stochastic volatility models. To begin with, we review the reducible and irreducible methods of Aït-Sahalia (2008). Since our regime-switching continuous time stochastic volatility models proposed in Chapter 3 not satisfy the definition of reducible diffusions in Aït-Sahalia (2008), we resort to the irreducible method to find an approximate log-transition density of the process in each regime. Then, we use Hamilton algorithm (Hamilton (1989)) to compute the likelihood of the regime-switching continuous time stochastic volatility models and to conduct the maximum likelihood estimation. Furthermore, we successfully apply this methodology to estimate the regime-switching stochastic volatility models, which are proposed to describe the behaviour of S&P 500 and VIX in Chapter 3.

In Chapter 3, we combine the regime shift with three stochastic volatility models, the Heston model, the GARCH model and the CEV model. According to the number of regimes, the initial probability and the transition probability matrix specifications, we estimate four models for each case. For Heston model, we compare H-R1 (single-regime), H-R2-1 (two regimes, time-constant transition matrix, unconditional probability for the probability of the initial state), H-R2-2 (two regimes, time-constant transition matrix, additional parameter for the probability of the initial state) and

H-R2TVTP (two regimes, time-varying transition matrix with a logistic function, additional parameter for the probability of the initial state). For the GARCH model and the CEV model, We estimate G-R1, G-R2-1, G-R2-2, G-R2TVTP and C-R1, C-R2-1, C-R2-2, C-R2TVTP, separately. All parameters in these models are allowed to vary depending on the state of the economy. The maximum likelihood estimation, which is developed in Chapter 2, is applied to estimate these regime-switching models. Using S&P 500 and VIX data for the stock price and volatility proxy, respectively, we report estimates for each model, as well as AIC and BIC to compare different models. Another metric we employed to compare the performance of different regime-switching models is called Regime Classification Measure (RCM). It is first proposed by Ang and Bekaert (2002) and then applied by Choi (2009). According to the inferred probabilities of staying in a particular regime, the closer the RCM value is to 0, the better the regime classification of a model is. We also use LR statistic to compare regime-switching continuous time diffusion models with single regime models. Furthermore, we investigate the regime-switching probabilities of the regime-switching models with time varying transition matrix and additional parameter for the probability of initial state. We also analyse the corresponding approximate conditional transition density functions with their 95% confidence intervals in order to find the evidence of regime shift. Our estimation results show four main findings. First, the regime-switching models are significantly different from the single regime models. Second, there are strong evidences for the existence of the high and low volatility regimes, for the time varying transition probability of the regime variable, and for high persistence of the high regime. Third, the time varying transition probability mainly depends on the stock market index S&P 500 rather than its volatility. Fourth, the regime-switching CEV model with time varying transition matrix and additional parameter for the probability of initial state performs better than other regime-switching models.

Chapter 4 provides some guidelines on the choice of the discretization interval when approximating diffusion processes using the Euler approximation method. We use the solution of a diffusion process when it is available and usable as a benchmark. The Milstein approximation is also adopted to compare the accuracy of the Euler approximation. Depending on the problem of interest, different criteria are used to measure the accuracy of approximation. The percentage error and strong convergence can be examined when a good approximation of sample path of a diffusion model is

required. The weak convergence is preferred for the cases where approximation of moments of the process matters. On one hand, we analyses the effective time discretization step Δ for univariate stochastic diffusion models, including Black-Scholes model, Vasicek model, CIR model, Inverse of Feller's Square Root model, Linear Drift CEV model and Nonlinear Mean Reversion model. On the other hand, we move to multivariate stochastic diffusion models, such as Heston, GARCH, CEV, Stein, Scott, Hull-White, Hagan and SABR process. For the univariate diffusion continuous time models, our numerical tests suggest an appropriate time interval should be set for each model in order to satisfy different error requirements when the objective is to simulate a diffusion process or test statistical estimators. When the objective is to compute moments, probabilities or other functions of the diffusion process, we suggest considering the turning point of the corresponding weak convergence and choosing Δ that does not exceed this turning point. Considering the bivariate continuous time diffusion models, we report guidelines for both of state variables. When a good approximation of sample path of a diffusion model is required, Δ could be set according to the state variable of interest. However, for the cases where approximation of moments of the process matters, our Monte Carlo simulation results recommend to consider the volatility state variable in the first place.

Chapter 5 concludes this thesis with a discussion of the results as well as outlining future research directions.

CHAPTER 2

Maximum Likelihood Estimation for Regime-Switching Continuous Time Stochastic Volatility Models

2.1. Introduction

To estimate continuous time diffusion models, researchers have proposed a variety of econoometric methods. These methods include simulation method of Duffie and Singleton (1993), maximum likelihood estimation (MLE) from Pearson and Sun (1994) and Durham and Gallant (2002), generalized method of moment (GMM) by Hansen and Scheinkman (1995), nonparametric method of Aït-Sahalia (1996a) and Aït-Sahalia (1996b), efficient method of moment (EMM) by Gallant and Tauchen (1998) and Bayesian method from Elerian, Chib, and Shephard (2001). The maximum likelihood estimation (MLE) method has not been used much because the transition density function, hence the log-likelihood function, is not known for most multivariate diffusions. Several methods have been proposed to approximate the transition probability densities. Examples are the Kolmogorov partial differential equation (Lo (1988)), Monte Carlo Simulation (Pedersen (1995)), binomial or other trees (Jensen and Poulsen (2002)), and the Euler approximations (Kleppe, Yu, and Skaug (2010)). However, neither of these methods produces a closed form density function. To solve this issue, Aït-Sahalia (2002) and Aït-Sahalia (2008) proposed a specific closed-form approximation approach for the univariate and multivariate continuous time stochastic volatility models, respectively. Then this approximation has been successfully applied by Choi (2009) for univariate models and Aït-Sahalia and Kimmel (2007) for multivariate models. Furthermore, his work has been extended to univariate time homogeneous diffusion models driven by Lėvy processes (Schaumburg (2001)), to univariate time inhomogeneous diffusion models (Egorov, Li, and Xu (2003)), to regime-switching time homogeneous univariate diffusion models for short term interest rates (Choi (2009)), to multivariate time homogeneous jump diffusion models (Yu (2007)), to multivariate time inhomogeneous diffusion models (Choi (2013)), to a Bayesian setting (Stramer, Bognar, and Schneider (2010)) and to damped diffusion (Li

(2010)). Jensen and Poulsen (2002) compare most commonly used techniques, including Hermite expansion, Euler approximations, simulation-based methods, binomial approximations and numerical solution of Kolmogorov partial differential equations, and find that Aït-Sahalia (2002)'s method is the best in terms of accuracy and speed trade-off. Because of the outstanding performance of Aït-Sahalia (2002)'s method for univariate continuous time models, we try to adopt Aït-Sahalia (2008)'s method for multivariate continuous time diffusion models.

Since our regime-switching continuous time stochastic volatility models proposed in Chapter 3 dissatisfy the conception of reducible diffusions in Aït-Sahalia (2008), we resort to the irreducible method to find an approximate log-transition density for each regime. Then, gaining enlightenment from Choi (2009), we use Hamilton algorithm (Hamilton (1989)) to obtain the likelihood function of the entire regime-switching continuous time stochastic volatility models and maximize the likelihood function to conduct the maximum likelihood estimation.

This chapter is organized as follows. Section 2 reviews Aït-Sahalia (2008)'s closed-form likelihood expansions in terms of reducible and irreducible cases for continuous time diffusion models. Section 3 develops the closed form maximum likelihood estimation for regime-switching continuous time stochastic volatility models by using Hamilton algorithm (Hamilton (1989)).

2.2. Closed-form Likelihood Expansions

In each regime, the dynamics of the state vector X_t follow a stochastic differential equation. Using the Bayes' Rule, the joint probability density function of the data $\left(x_{n\Delta}, x_{(n-1)\Delta}, \cdots, x_{\Delta}, x_0\right)$ can be written as

$$\begin{aligned} & p\left(x_{n\Delta}, x_{(n-1)\Delta}, \cdots, x_{\Delta}, x_0; \theta\right) \\ = \; & p\left(x_{n\Delta} | x_{(n-1)\Delta}, \cdots, x_{\Delta}, x_0; \theta\right) \times p\left(x_{(n-1)\Delta} | x_{(n-2)\Delta}, \cdots, x_{\Delta}, x_0; \theta\right) \\ & \cdots \times p\left(x_{2\Delta} | x_{\Delta}, x_0; \theta\right) \times p\left(x_{\Delta} | x_0; \theta\right) \times p\left(x_0; \theta\right). \end{aligned}$$

The sample conditional log-likelihood is represented as

$$l_n\left(\theta\right) \equiv \ln(p\left(x_0; \theta\right)) + \sum_{i=1}^{n} \ln\left[p\left(x_{i\Delta} | I_{(i-1)\Delta}; \theta\right)\right]. \tag{2.1}$$

Ignoring the asymptotically irrelevant density of the initial observation x_0, equation 2.1 is represented as

$$l_n(\theta) \equiv \sum_{i=1}^{n} \ln\left[p\left(x_{i\Delta}|I_{(i-1)\Delta};\theta\right)\right], \tag{2.2}$$

where $I_t = \{X_s|s \leq t\}$ is the information set consisting of data $\{t = i\Delta|i = 0, 1, \cdots, n\}$. If the transition density function of a diffusion process is available, conditional likelihood functions $p\left(x_{i\Delta}|I_{(i-1)\Delta};\theta\right)$ for all $i = 0, 1, 2, \cdots, n$ can be calculated by using the Hamilton algorithm (Hamilton (1989)) shown in Section 2.3. However, in practice, the transition density function, hence the log-likelihood function, is not known in closed form for most multivariate diffusions. Therefore, we resort to method of Aït-Sahalia (2008) to find an approximate log-transition density.

2.2.1. Reducible Method

If a multivariate diffusion is reducible, the log-likelihood function can be found by using reducible method that includes Hermite expansion and Kolmogorov method. However, if it is not reducible (irreducible), we have to use the irreducible method. In this section, we are going to discuss Aït-Sahalia (2008)'s work by introducing the concept of reducibility for diffusions first.

Definition 1. *The diffusion process* X_t

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

is said to be reducible to unit diffusion (or reducible, in short) if and only if there exists a one-to-one transformation of the diffusion process X_t *into a diffusion process* Y_t *whose dispersion matrix* σ_Y *is the identity matrix. That is, there exists an invertible function* $\gamma(x;\theta)$*, infinitely differentiable in* x *on* S_X *such that* $Y_t \equiv \gamma(X_t;\theta)$ *satisfies the stochastic differential equation*

$$dY_t = \mu_Y(Y_t;\theta)\,dt + dW_t$$

on the domain S_Y*.*

By applying the Itô lemma, the drift function of the i-th component of Y_t is

$$\begin{aligned}\mu_{Y_i}(y;\theta) &= \sum_{p=1}^{m}\mu_p\left(\gamma^{inv}(y;\theta);\theta\right)\frac{\partial\gamma_i(y;\theta)}{\partial x_p}|_{x=\gamma^{inv}(y;\theta)} \\ &+\frac{1}{2}\sum_{p=1}^{m}\sum_{q=1}^{m}\sum_{r=1}^{m}\sigma_{pr}\left(\gamma^{inv}(y;\theta);\theta\right)\sigma_{qr}\left(\gamma^{inv}(y;\theta);\theta\right)\frac{\partial^2\gamma_i(y;\theta)}{\partial x_p\partial x_q}|_{x=\gamma^{inv}(y;\theta)}\end{aligned}$$

and

$$\nabla\gamma(x;\theta)=\sigma^{-1}(x;\theta)$$

where $\nabla\gamma(x;\theta)$ represents the Jacobian matrix of $\gamma(x;\theta)=[\gamma_1(x;\theta),\cdots,\gamma_m(x;\theta)]'$ with respect to $x\in R^m$ such that $\nabla\gamma(x;\theta)=[\partial\gamma_i(x;\theta)/\partial x_j]_{i=1,\cdots,m;j=1,\cdots,m}$. We can tell whether a diffusion process is reducible or not by checking the conditions for $\sigma(x;\theta)$ given below.

Proposition 2. *(Necessary and sufficient condition for reducibility) The diffusion process X_t is reducible if and only if*

$$\frac{\partial\sigma_{ij}^{-1}(x;\theta)}{\partial x_k}=\frac{\partial\sigma_{ik}^{-1}(x;\theta)}{\partial x_j}$$

for each x in S_t and triplet $\{i,j,k\}\subset\{1,2,\cdots,m\}$ such that $k>j$. Here $\partial\sigma_{ij}^{-1}(x;\theta)$ is the (i,j) element of the inverse matrix of $\sigma(x;\theta)$.

If a multivariate diffusion process is reducible, an approximate transition density can be obtained by using the Hermite expansion. The Hermite polynomials for the case of multivariate variables are

$$H_h(z)=\frac{(-1)^{|h|}}{\phi(z)}\frac{\partial^{|h|}\phi(z)}{\partial z_1^{h_1}\cdots\partial z_m^{h_m}},$$

where $|h|=\sum_{i=1}^{m}h_i$ and $\phi(z)$ is the pdf of an m-dimensional standard multivariate normal distribution with mean 0 and covariance identity matrix. The Hermite polynomials are orthogonal in the following sense

$$\int_{R^m}H_h(z)H_k(z)\phi(z)dz=\begin{cases}h_1!\cdots h_m! & \text{if } h=k\\ 0 & \text{otherwise}\end{cases}.$$

The J-th order Hermite series expansion, $p_Y^{(J)}$ for a multivariate time homogeneous unit diffusion process Y_t is of the form

$$p_Y^{(J)}(\Delta,y|y_0;\theta)=\Delta^{-m/2}\phi\left(\frac{y-y_0}{\sqrt{\Delta}}\right)\sum_{h\in N^m:|h|\leq J}\eta_h(\Delta,y_0;\theta)H_h\left(\frac{y-y_0}{\sqrt{\Delta}}\right).$$

Using the orthogonality of the Hermite polynomials,

$$\eta_h(\Delta, y_0) = \frac{1}{h_1! \cdots h_m!} E\left[H_h\left(\frac{y - y_0}{\sqrt{\Delta}}\right) | Y_{t_0} = y_0\right]. \tag{2.3}$$

If the coefficient of equation 2.3 is approximated up to K-th order by using the infinitesimal generator A_Y corresponding to the diffusion process Y_t, i.e,

$$A_Y \circ f(\Delta, y, y_0) = \sum_{i=1}^{m} \mu_{Y_i}(y; \theta) \frac{\partial f(\Delta, y, y_0)}{\partial y_i} + \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \frac{\partial^2 f(\Delta, y, y_0)}{\partial y_i \partial y_j}.$$

Then, we get

$$\begin{aligned} & p_Y^{(J,K)}(\Delta, y|y_0; \theta) \\ = \; & \Delta^{-m/2} \phi\left(\frac{y - y_0}{\sqrt{\Delta}}\right) \\ & \times \left\{ \sum_{h \in N^m : |h| \leq J} \frac{1}{h_1! \cdots h_m!} \left[\sum_{i=0}^{K} \frac{\Delta^i}{i!} A_Y^i \circ H_h\left(\frac{y - y_0}{\sqrt{\Delta}}\right)|_{y=y_0}\right] H_h\left(\frac{y - y_0}{\sqrt{\Delta}}\right) \right\}. \end{aligned} \tag{2.4}$$

We can send J to ∞ and rearrange the terms of equation 2.4 in ascending orders of Δ to acquire an alternative expansion $p_Y^{(K)} = p_Y^{(\infty,K)}$ such that

$$\begin{aligned} & p_Y^{(K)}(\Delta, y|y_0; \theta) \\ = \; & \Delta^{-m/2} \phi\left(\frac{y - y_0}{\sqrt{\Delta}}\right) \\ & \times \exp\left[\sum_{i=0}^{m} (y_i - y_{0i}) \int_0^1 \mu_{Y_i}(y_0 + u(y - y_0); \theta) \, du\right] \sum_{k=0}^{K} c_Y^{(k)}(\Delta, y|y_0; \theta) \frac{\Delta^k}{k!}. \end{aligned}$$

Taking logarithm of it and Taylor-expanding it in Δ around zero, we get the approximate log-transition density of Y_t in the following theorem, which is proposed by Aït-Sahalia (2008), as a result of applying the Kolmogorov-equation method.

Theorem 3. *The K-th order log-transition density expansion of a time-homogeneous multivariate unit diffusion process Y_t is*

$$l_Y^{(K)}(\Delta, y|y_0; \theta) = -\frac{m}{2} \ln(2\pi\Delta) + \frac{C_Y^{(-1)}(\Delta, y|y_0; \theta)}{\Delta} + \sum_{k=0}^{K} c_Y^{(k)}(\Delta, y|y_0; \theta) \frac{\Delta^k}{k!} \tag{2.5}$$

with

$$C_Y^{(-1)}(\Delta, y|y_0; \theta) = -\frac{1}{2} \sum_{i=0}^{m} (y_i - y_{0i})^2$$

$$C_Y^{(0)}(\Delta, y|y_0;\theta) = \sum_{i=0}^{m}(y_i - y_{0i})\int_0^1 \mu_{Y_i}(y_0 + u(y-y_0);\theta)\,du$$

and for $k \geq 1$

$$C_Y^{(0)}(\Delta, y|y_0;\theta) = k\int_0^1 G_Y^{(k)}(\Delta, y_0 + u(y-y_0)|y_0;\theta)\,u^{k-1}du,$$

where

$$\begin{aligned} G_Y^{(1)}(\Delta, y|y_0;\theta) &= -\sum_{i=0}^{m}\frac{\partial\mu_{Y_i}(y;\theta)}{\partial y_i} - \sum_{i=0}^{m}\mu_{Y_i}(y;\theta)\frac{\partial C_Y^{(0)}(\Delta, y|y_0;\theta)}{\partial y_i} \\ &\quad + \frac{1}{2}\sum_{i=0}^{m}\left\{\frac{\partial^2 C_Y^{(0)}(\Delta, y|y_0;\theta)}{\partial y_i^2} + \left[\frac{\partial C_Y^{(0)}(\Delta, y|y_0;\theta)}{\partial y_i}\right]^2\right\} \end{aligned}$$

and for $k \geq 2$

$$\begin{aligned} G_Y^{(k)}(\Delta, y|y_0;\theta) &= \frac{1}{2}\sum_{i=0}^{m}\frac{\partial^2 C_Y^{(k-1)}(\Delta, y|y_0;\theta)}{\partial y_i^2} - \sum_{i=0}^{m}\mu_{Y_i}(y;\theta)\frac{\partial C_Y^{(k-1)}(\Delta, y|y_0;\theta)}{\partial y_i} \\ &\quad + \frac{1}{2}\sum_{i=0}^{m}\sum_{h=0}^{k-1}\binom{k-1}{h}\frac{\partial C_Y^{(h)}(\Delta, y|y_0;\theta)}{\partial y_i}\frac{\partial C_Y^{(k-1-h)}(\Delta, y|y_0;\theta)}{\partial y_i}. \end{aligned}$$

The transition density $p_Y(\Delta, y|y_0;\theta)$ *satisfies the Kolmogorov forward and backward equations, respectively,*

$$\frac{\partial p_Y(\Delta, y|y_0;\theta)}{\partial\Delta} = -\sum_{i=0}^{m}\frac{\partial\left\{\mu_{Y_i}(y;\theta)\,p_Y(\Delta, y|y_0;\theta)\right\}}{\partial y_i} + \frac{1}{2}\sum_{i=0}^{m}\frac{\partial^2 p_Y(\Delta, y|y_0;\theta)}{\partial y_i^2}$$

and

$$\frac{\partial p_Y(\Delta, y|y_0;\theta)}{\partial\Delta} = \sum_{i=0}^{m}\mu_{Y_i}(y;\theta)\frac{\partial p_Y(\Delta, y|y_0;\theta)}{\partial y_{0i}} + \frac{1}{2}\sum_{i=0}^{m}\frac{\partial^2 p_Y(\Delta, y|y_0;\theta)}{\partial y_{0i}^2}.$$

If we plug the log-likelihood expansion 2.5 into the Kolmogorov forward equation for $I_Y(\Delta, y|y_0;\theta)$*:*

$$\begin{aligned} \frac{\partial I_Y(\Delta, y|y_0;\theta)}{\partial\Delta} &= -\sum_{i=0}^{m}\frac{\partial\mu_{Y_i}(y;\theta)}{\partial y_i} - \sum_{i=0}^{m}\mu_{Y_i}(y;\theta)\frac{\partial I_Y(\Delta, y|y_0;\theta)}{\partial y_i} \\ &\quad + \frac{1}{2}\sum_{i=0}^{m}\frac{\partial^2 I_Y(\Delta, y|y_0;\theta)}{\partial y_i^2} + \frac{1}{2}\sum_{i=0}^{m}\left[\frac{\partial I_Y(\Delta, y|y_0;\theta)}{\partial y_i}\right]^2, \end{aligned} \tag{2.6}$$

and equate the coefficients of Δ^k*,* $k \geq -2$ *on both sides of equation 2.6 then we can get the PDEs for* $C_Y^{(k)}(t, y|t_0, y_0)$*,* $k \geq -1$ *by solving the equations given in the above*

theorem. The Kolmogorov backward equation for $I_Y(\Delta, y|y_0;\theta)$ *is*

$$(2.7)\quad \frac{\partial I_Y(\Delta, y|y_0;\theta)}{\partial \Delta} = \sum_{i=0}^{m} \mu_{Y_i}(y;\theta) \frac{\partial I_Y(\Delta, y|y_0;\theta)}{\partial y_{0i}} + \frac{1}{2}\sum_{i=0}^{m} \frac{\partial^2 I_Y(\Delta, y|y_0;\theta)}{\partial y_{0i}^2} + \frac{1}{2}\sum_{i=0}^{m} \left[\frac{\partial I_Y(\Delta, y|y_0;\theta)}{\partial y_{0i}}\right]^2.$$

In this way, the log-density, $I_Y(\Delta, y|y_0;\theta)$ of the original process X_t can be retrieved from $I_Y(\Delta, y|y_0;\theta)$ by the change of variable as

$$\begin{aligned} I_Y(\Delta, y|y_0;\theta) &= \ln\left(Det\left[\nabla\gamma(x;\theta)\right]\right) + I_Y\left(\Delta, \gamma(x;\theta)\left|\gamma(x_0;\theta);\theta\right.\right) \\ &= -D_v(x;\theta) + I_Y\left(\Delta, \gamma(x;\theta)\left|\gamma(x_0;\theta);\theta\right.\right), \end{aligned}$$

where $D_v(x;\theta) \equiv \frac{1}{2}\ln\left(Det\left[v(x;\theta)\right]\right)$, because $Det\left[\nabla\gamma(x;\theta)\right] = Det\left[\sigma^{-1}(x;\theta)\right] = Det\left[v(x;\theta)\right]^{-1/2}$ with $v(x;\theta) = \sigma(x;\theta)\sigma(x;\theta)^T$. Replacing I_Y with $I_Y^{(K)}$ found above and define $I_X^{(K)}$ as

$$\begin{aligned} I_Y^{(K)}(\Delta, y|y_0;\theta) &= -D_v(x;\theta) + I_Y\left(\Delta, \gamma(x;\theta)\left|\gamma(x_0;\theta);\theta\right.\right) \\ &= -D_v(x;\theta) - \frac{m}{2}\ln(2\pi\Delta) \\ &\quad + \frac{C_Y^{(-1)}\left(\Delta, \gamma(x;\theta)\left|\gamma(x_0;\theta);\theta\right.\right)}{\Delta} \\ &\quad + \sum_{k=0}^{K} C_Y^{(k)}\left(\Delta, \gamma(x;\theta)\left|\gamma(x_0;\theta);\theta\right.\right)\frac{\Delta^k}{k!}. \end{aligned}$$

By construction, $I_X^{(K)}$ solves the Kolmogorov equations 2.6 and 2.7 for X_t at the same order as $I_Y^{(K)}$.

However, in fact, most multivariate time-inhomogeneous diffusion is irreducible. If X_t is irreducible, none of Hermite-expansion and Kolmogorov-equation methods can be adopted since we cannot transform it into a unit diffusion process. As a result, we have to resort to the irreducible method.

2.2.2. Irreducible Method

The main idea of finding approximate log-transition density of irreducible diffusions is to postulate the form of the log-likelihood expansion of X_t as the one found from

the reducible case:

$$
\begin{aligned}
(2.8) \qquad l_X^{(K)}(\Delta, x|x_0;\theta) &= -\frac{m}{2}\ln(2\pi\Delta) - D_v(x;\theta) \\
&\quad + \frac{C_X^{(-1)}(\Delta, x|x_0;\theta)}{\Delta} \\
&\quad + \sum_{k=0}^{K} C_X^{(k)}(\Delta, x|x_0;\theta)\frac{\Delta^k}{k!},
\end{aligned}
$$

where $k \geq -1$, $D_v(x;\theta) = \frac{1}{2}\ln(\det[v(x;\theta)])$ and $v(x;\theta) = \sigma(x)\sigma'(x)$.

The Kolmogorov forward and backward equations for $l_X(\Delta, x|x_0;\theta)$ are, respectively,

$$
\begin{aligned}
(2.9) \quad \frac{\partial l_X(\Delta, x|x_0;\theta)}{\partial\Delta} &= -\sum_{i=1}^{m}\frac{\partial\mu_i(x;\theta)}{\partial x_i} + \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\frac{\partial^2 v_{ij}(x;\theta)}{\partial x_i\partial x_j} \\
&\quad - \sum_{i=1}^{m}\mu_i(t,x)\frac{\partial l_X(\Delta, x|x_0;\theta)}{\partial x_i} \\
&\quad + \sum_{i=1}^{m}\sum_{j=1}^{m}\frac{\partial v_{ij}(x;\theta)}{\partial x_i}\frac{\partial l_X(\Delta, x|x_0;\theta)}{\partial x_j} \\
&\quad + \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} v_{ij}(x;\theta)\frac{\partial^2 l_X(\Delta, x|x_0;\theta)}{\partial x_i\partial x_j} \\
&\quad + \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\frac{\partial l_X(\Delta, x|x_0;\theta)}{\partial x_i} v_{ij}(x;\theta)\frac{\partial l_X(\Delta, x|x_0;\theta)}{\partial x_j}
\end{aligned}
$$

and

$$
\begin{aligned}
(2.10) \frac{\partial l_X(\Delta, x|x_0;\theta)}{\partial\Delta} &= -\sum_{i=1}^{m}\mu_i(x;\theta)\frac{\partial l_X(\Delta, x|x_0;\theta)}{\partial x_{0i}} \\
&\quad + \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} v_{ij}(x;\theta)\frac{\partial^2 l_X(\Delta, x|x_0;\theta)}{\partial x_{0i}\partial x_{0j}} \\
&\quad + \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\frac{\partial l_X(\Delta, x|x_0;\theta)}{\partial x_{0i}} v_{ij}(x;\theta)\frac{\partial l_X(\Delta, x|x_0;\theta)}{\partial x_{0j}}.
\end{aligned}
$$

Replacing l_X with $l_X^{(K)}$ in the Kolmogorov equations 2.9 and 2.10 for the log-transition density of X_t and matching the terms with the same orders of Δ yields partial differential equations of the coefficients $\sum_{k=0}^{K} C_X^{(k)}(\Delta, x|x_0;\theta)$, $k \geq -1$. By equating the terms of order Δ^{-2} in 2.9, the equation that determines the coefficient $C_X^{(-1)}$ is

obtained as

$$C_X^{(-1)}(\Delta, x|x_0;\theta) = -\frac{1}{2}\left(\frac{\partial C_X^{(-1)}(\Delta, x|x_0;\theta)}{\partial x}\right)^T v(x;\theta)\left(\frac{\partial C_X^{(-1)}(\Delta, x|x_0;\theta)}{\partial x}\right).$$

However, it is still a challenge to get an explicit solution of this equation (Varadhan (1967)). As a result, we could not get the coefficients of the expansion 2.8. In order to get the coefficients $C_X^{(k)}$, we can Taylor-expand each coefficient around x_0 up to j_k-th order. In this way, the same approximation error of $O_p\left(\Delta^{K+1}\right)$ can be achieved for each coefficient. Let j_k-th order Taylor expansion of $C_X^{(k)}(\Delta, x|x_0;\theta)$ be $C_X^{(k,j_k)}(\Delta, x|x_0;\theta)$, which is in the form of

$$C_X^{(k,j_k)}(\Delta, x|x_0;\theta) = \sum_{i\in I_k}\beta_i^{(k)}(x_0)(x_1 - x_{01})^{i_1}(x_2 - x_{02})^{i_2}\cdots(x_m - x_{0m})^{i_m}.$$

Given the original term $C_X^{(j_{-1},-1)}(\Delta, x|x_0;\theta)$, the next term $C_X^{(j_0,0)}(\Delta, x|x_0;\theta)$ could be calculated explicitly. Once $C_X^{(j_{-1},-1)}(\Delta, x|x_0;\theta)$ and $C_X^{(j_0,0)}(\Delta, x|x_0;\theta)$ are given, the next term $C_X^{(j_1,1)}(\Delta, x|x_0;\theta)$ could be determined, and so on. To be more precise, the coefficients have to found recursively from low order term to high order term one by one in $C_X^{(j_{-1},-1)}(\Delta, x|x_0;\theta)$. The order j_k [in $(x - x_0)$] corresponding to a given order k (in Δt) could be decided as follows. To begin with,

$$\begin{aligned}\left|C_X^{(K)}(\Delta, x|x_0;\theta)\Delta^k - C_X^{(k,j_k)}(\Delta, x|x_0;\theta)\Delta^k\right| &= O_p\left(\|X_\Delta - X_0\|^{j_k}\Delta^k\right)\\ &= O_p\left(\Delta^{j_k/2+k}\right),\end{aligned}$$

because $X_\Delta - X_0 = O_p\left(\Delta^{1/2}\right)$. In consequence, setting

$$j_k = 2(K - k + 1)$$

will provide an approximation error $O_p\left(\Delta^{K+1}\right)$. For example, if $K = 2$, $j_{-1} = 8$, $j_0 = 6$, $j_1 = 4$, $j_2 = 2$, $j_3 = 0$. In order to state the final result pertaining to the closed-form solutions $C_X^{(k,j_k)}(\Delta, x|x_0;\theta)$, we need to solve a system of linear equations

$$f_X^{(k-1)}(\Delta, x|x_0;\theta) = 0,$$

where $k = -1, 0, ..., K$,

$$\begin{aligned} f_X^{(-2)}(\Delta, x|x_0;\theta) &= -2C_X^{(-1)}(\Delta, x|x_0;\theta) \\ &\quad -\sum_{i=1}^{m}\sum_{j=1}^{m} v_{ij}(x;\theta)\frac{\partial C_X^{(-1)}(\Delta, x|x_0;\theta)}{\partial x_i}\frac{\partial C_X^{(-1)}(\Delta, x|x_0;\theta)}{\partial x_j}, \end{aligned}$$

$$\begin{aligned} f_X^{(-1)}(\Delta, x|x_0;\theta) &= -\sum_{i=1}^{m}\sum_{j=1}^{m} v_{ij}(x;\theta)\frac{\partial C_X^{(-1)}(\Delta, x|x_0;\theta)}{\partial x_i}\frac{\partial C_X^{(0)}(\Delta, x|x_0;\theta)}{\partial x_j} \\ &\quad -G_X^{(0)}(\Delta, x|x_0;\theta), \end{aligned}$$

and for $k \geq 1$,

$$\begin{aligned} f_X^{(k-1)}(\Delta, x|x_0;\theta) &= C_X^{(k)}(\Delta, x|x_0;\theta) \\ &\quad -\frac{1}{k}\sum_{i=1}^{m}\sum_{j=1}^{m} v_{ij}(x;\theta)\frac{\partial C_X^{(-1)}(\Delta, x|x_0;\theta)}{\partial x_i}\frac{\partial C_X^{(k)}(\Delta, x|x_0;\theta)}{\partial x_j} \\ &\quad -G_X^{(K)}(\Delta, x|x_0;\theta). \end{aligned}$$

$G_X^{(k)}(\Delta, x|x_0;\theta)$ is for $k = 0$ and 1 respectively

$$G_X^{(0)}(\Delta, x|x_0;\theta) = G_X^{(0,1)}(\Delta, x|x_0;\theta) + G_X^{(0,3)}(\Delta, x|x_0;\theta),$$

where

$$\begin{aligned} G_X^{(0,1)}(\Delta, x|x_0;\theta) &= -\sum_{i=1}^{m}\mu_i(x;\theta)\frac{\partial C_X^{(-1)}(\Delta, x|x_0;\theta)}{\partial x_i} \\ &\quad +\sum_{i=1}^{m}\sum_{j=1}^{m}\frac{\partial v_{ij}(x;\theta)}{\partial x_i}\frac{\partial C_X^{(-1)}(\Delta, x|x_0;\theta)}{\partial x_j} \\ &\quad -\sum_{i=1}^{m}\sum_{j=1}^{m} v_{ij}(x;\theta)\frac{\partial C_X^{(-1)}(\Delta, x|x_0;\theta)}{\partial x_i}\frac{\partial D_v(x;\theta)}{\partial x_j} \\ &\quad +\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} v_{ij}(x;\theta)\frac{\partial^2 C_X^{(-1)}(\Delta, x|x_0;\theta)}{\partial x_i \partial x_j} \end{aligned}$$

and

$$G_X^{(0,3)}(\Delta, x|x_0;\theta) = \frac{m}{2},$$

and

$$G_X^{(1)}(\Delta, x|x_0;\theta) = G_X^{(1,1)}(\Delta, x|x_0;\theta) + G_X^{(1,2)}(\Delta, x|x_0;\theta) + G_X^{(1,3)}(\Delta, x|x_0;\theta),$$

where

$$\begin{aligned} G_X^{(1,1)}(\Delta, x|x_0;\theta) &= -\sum_{i=1}^{m}\mu_i(x;\theta)\frac{\partial C_X^{(0)}(\Delta, x|x_0;\theta)}{\partial x_i} \\ &\quad +\sum_{i=1}^{m}\sum_{j=1}^{m}\frac{\partial v_{ij}(x;\theta)}{\partial x_i}\frac{\partial C_X^{(0)}(\Delta, x|x_0;\theta)}{\partial x_j} \\ &\quad -\sum_{i=1}^{m}\sum_{j=1}^{m}v_{ij}(x;\theta)\frac{\partial C_X^{(0)}(\Delta, x|x_0;\theta)}{\partial x_i}\frac{\partial D_v(x;\theta)}{\partial x_j} \\ &\quad +\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}v_{ij}(x;\theta)\frac{\partial^2 C_X^{(0)}(\Delta, x|x_0;\theta)}{\partial x_i \partial x_j}, \end{aligned}$$

$$G_X^{(1,2)}(\Delta, x|x_0;\theta) = \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}v_{ij}(x;\theta)\frac{\partial C_X^{(0)}(\Delta, x|x_0;\theta)}{\partial x_i}\frac{\partial C_X^{(0)}(\Delta, x|x_0;\theta)}{\partial x_j},$$

$$\begin{aligned} G_X^{(1,3)}(\Delta, x|x_0;\theta) &= -\sum_{i=1}^{m}\frac{\partial \mu_i(x;\theta)}{\partial x_i}+\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\frac{\partial^2 v_{ij}(x;\theta)}{\partial x_i \partial x_j} \\ &\quad +\sum_{i=1}^{m}\mu_i(x;\theta)\frac{\partial D_v(x;\theta)}{\partial x_i} \\ &\quad -\sum_{i=1}^{m}\sum_{j=1}^{m}\frac{\partial v_{ij}(x;\theta)}{\partial x_i}\frac{\partial D_v(x;\theta)}{\partial x_j} \\ &\quad -\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}v_{ij}(x;\theta)\left[\frac{\partial^2 D_v(x;\theta)}{\partial x_i \partial x_j}-\frac{\partial D_v(x;\theta)}{\partial x_i}\frac{\partial D_v(x;\theta)}{\partial x_j}\right]. \end{aligned}$$

When $k \geq 2$,

$$G_X^{(k)}(\Delta, x|x_0;\theta) = G_X^{(k,1)}(\Delta, x|x_0;\theta) + G_X^{(k,2)}(\Delta, x|x_0;\theta),$$

where

$$\begin{aligned} G_X^{(k,1)}(\Delta, x|x_0;\theta) &= -\sum_{i=1}^{m}\mu_i(x;\theta)\frac{\partial C_X^{(k-1)}(\Delta, x|x_0;\theta)}{\partial x_i} \\ &\quad +\sum_{i=1}^{m}\sum_{j=1}^{m}\frac{\partial v_{ij}(x;\theta)}{\partial x_i}\frac{\partial C_X^{(k-1)}(\Delta, x|x_0;\theta)}{\partial x_j} \\ &\quad -\sum_{i=1}^{m}\sum_{j=1}^{m}v_{ij}(x;\theta)\frac{\partial C_X^{(k-1)}(\Delta, x|x_0,\theta)}{\partial x_i}\frac{\partial D_v(x;\theta)}{\partial x_j} \\ &\quad +\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}v_{ij}(x;\theta)\frac{\partial^2 C_X^{(k-1)}(\Delta, x|x_0;\theta)}{\partial x_i \partial x_j} \end{aligned}$$

and

$$\begin{aligned} G_X^{(k,2)}(\Delta, x|x_0;\theta) &= \sum_{i=1}^{m}\sum_{j=1}^{m} v_{ij}(x;\theta)\frac{\partial C_X^{(0)}(\Delta, x|x_0;\theta)}{\partial x_i}\frac{\partial C_X^{(k-1)}(\Delta, x|x_0;\theta)}{\partial x_j} \\ &\quad + \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} v_{ij}(x;\theta) \\ &\quad \times \left[\sum_{h=1}^{k-2}\binom{k-1}{h}\frac{\partial C_X^{(h)}(\Delta, x|x_0;\theta)}{\partial x_i}\frac{\partial C_X^{(k-1-h)}(\Delta, x|x_0;\theta)}{\partial x_j}\right]. \end{aligned}$$

Replacing $C_X^{(k)}(\Delta, x|x_0;\theta)$ with $C_X^{(k,j_k)}(\Delta, x|x_0;\theta)$ in 2.9, we get

$$\begin{aligned} &\widetilde{l}_X^{(K)}(\Delta, x|x_0;\theta) \\ &= -\frac{m}{2}\ln(2\pi\Delta) - D_v(x;\theta) + \frac{C_X^{(j_{-1},-1)}(\Delta, x|x_0;\theta)}{\Delta} + \sum_{k=0}^{K} C_X^{(j_k,k)}(\Delta, x|x_0;\theta)\frac{\Delta^k}{k!}. \end{aligned}$$

Therefore, $C_X^{(-1)}$ could be determined by solving the equation $f_x^{(-2)} = 0$. Given $C_X^{(-1)}$, $G_X^{(0)}$ becomes known and then $C_X^{(0)}$ could be obtained by solving the equation $f_x^{(-1)} = 0$. Given $C_X^{(0)}$, $G_X^{(1)}$ becomes known and then $C_X^{(1)}$ could be obtained by solving the equation $f_x^{(0)} = 0$, and so on.

2.3. Hamilton Algorithm

As mentioned at the beginning of Section 2, given the approximate transition density function, conditional likelihood functions $p\left(x_{i\Delta}|I_{(i-1)\Delta};\theta\right)$ for all $i = 0, 1, 2, \cdots, n$ can be calculated using the algorithm developed by Hamilton (1989). Given a two state, $\{L, H\}$, a new state variable R_t^* is defined as follows:

$$R_t^* = 1 \text{ if } R_{t-\Delta} = L \text{ and } R_t = L$$

$$R_t^* = 2 \text{ if } R_{t-\Delta} = L \text{ and } R_t = H$$

$$R_t^* = 3 \text{ if } R_{t-\Delta} = H \text{ and } R_t = L$$

$$R_t^* = 4 \text{ if } R_{t-\Delta} = H \text{ and } R_t = H$$

where Δ denotes a time period. The state $R_t^* = 1$ ($R_t^* = 4$) describes the process keeps staying in the regime L (H) at time $t - \Delta$ and t, while the state $R_t^* = 2$ ($R_t^* = 3$) describes the process leaves regime L (H) and enters regime H (L) at time t. Therefore, R_t^* follows a continuous time Markov chain with four states and the

corresponding generator matrix is

$$P^* = \begin{pmatrix} p_{LL} & 0 & p_{LL} & 0 \\ p_{LH} & 0 & p_{LH} & 0 \\ 0 & p_{HL} & 0 & p_{HL} \\ 0 & p_{HH} & 0 & p_{HH} \end{pmatrix}.$$

At time $t+\Delta$, the input of the algorithm are $\widehat{\zeta}_{t+\Delta|t}$ and $\eta_{t+\Delta}$,

$$\widehat{\zeta}_{t+\Delta|t} = \begin{pmatrix} P\left(R^*_{t+\Delta}=1|I_t;\theta\right) \\ P\left(R^*_{t+\Delta}=2|I_t;\theta\right) \\ P\left(R^*_{t+\Delta}=3|I_t;\theta\right) \\ P\left(R^*_{t+\Delta}=4|I_t;\theta\right) \end{pmatrix}$$

and

$$\eta_{t+\Delta|t} = \begin{pmatrix} p\left(x_{t+\Delta}|R^*_{t+\Delta}=1,I_t;\theta\right) \\ p\left(x_{t+\Delta}|R^*_{t+\Delta}=2,I_t;\theta\right) \\ p\left(x_{t+\Delta}|R^*_{t+\Delta}=3,I_t;\theta\right) \\ p\left(x_{t+\Delta}|R^*_{t+\Delta}=4,I_t;\theta\right) \end{pmatrix},$$

where I_t is the information set which covers data up to time t so that $I_t = \{x_\tau|\tau \le t\}$. Then the joint likelihood of $X_{t+\Delta}$ and $R^*_{t+\Delta}$ is the results of $\widehat{\zeta}_{t+\Delta|t} \odot \eta_{t+\Delta}$, and the transition density function can be calculated as

$$\begin{aligned} p\left(x_{t+\Delta}|I_t;\theta\right) &= \sum_{j=1}^{4} p\left(x_{t+\Delta}, R^*_{t+\Delta}=j|I_t;\theta\right) \\ &= \sum_{j=1}^{4} P\left(R^*_{t+\Delta}=j|I_t;\theta\right) p\left(x_{t+\Delta}|R^*_{t+\Delta}=j,I_t;\theta\right). \end{aligned} \tag{2.11}$$

After it, $\widehat{\zeta}_{t+\Delta|t+\Delta}$, the optimal inference about $R^*_{t+\Delta}$, is computed by

$$\widehat{\zeta}_{t+\Delta|t+\Delta} = \frac{\widehat{\zeta}_{t+\Delta|t} \odot \eta_{t+\Delta}}{p\left(x_{t+\Delta}|I_t;\theta\right)} = \begin{pmatrix} P\left(R^*_{t+\Delta}=1|I_{t+\Delta};\theta\right) \\ P\left(R^*_{t+\Delta}=2|I_{t+\Delta};\theta\right) \\ P\left(R^*_{t+\Delta}=3|I_{t+\Delta};\theta\right) \\ P\left(R^*_{t+\Delta}=4|I_{t+\Delta};\theta\right) \end{pmatrix}.$$

Then $\widehat{\zeta}_{t+2\Delta|t+\Delta}$, which is the input of the next time period, is updated by premultiplying the transition matrix P^* by $\widehat{\zeta}_{t+\Delta|t+\Delta}$. Assuming $P\left(R_1 = L\right) = p$, the initial

state probabilities of R_t^* are

$$(\pi_1, \pi_2, \pi_3, \pi_4) = (p_{LL} \cdot p, p_{LH} \cdot p, p_{HL} \cdot (1-p), p_{HH} \cdot (1-p)).$$

Based on these starting values, equation 2.11, the conditional density function of each time period can be calculated by iterating the above procedure. Then the ML estimates can be attained by maximizing the log-likelihood function $l_n(\theta)$. However, the estimation results cannot identify exactly which state the process was in at each time point. Only the probability of staying in each regime can be inferred based on the observations of the stock market index S&P 500. By using the information set up to time t, we can get $P(R_t^* = i|I_t;\theta)$, the inferred probability of R_t^* being R. Since we are more interested in the state of R_t rather than R_t^*, we calculate the filtered probability as

$$P(R_t = L|I_t;\theta) = P(R_t^* = 1|I_t;\theta) + P(R_t^* = 3|I_t;\theta)$$

and

$$P(R_t = H|I_t;\theta) = P(R_t^* = 2|I_t;\theta) + P(R_t^* = 4|I_t;\theta),$$

where $P(R_t^* = i|I_t;\theta)$ is the i-th input of $\widehat{\zeta}_{t+\Delta|t+\Delta}$. Based on the whole information set, we also calculate the smoothed probability of the economy being in each state at time t as

$$P(R_t = L|I_T;\theta) = P(R_t^* = 1|I_T;\theta) + P(R_t^* = 3|I_T;\theta)$$

and

$$P(R_t = H|I_T;\theta) = P(R_t^* = 2|I_T;\theta) + P(R_t^* = 4|I_T;\theta),$$

where $T = n\Delta$ is the last time point of the data. Although the transition probability varies with asset prices, the above recursive procedure does not change. Therefore, we only need to replace the elements of P^* with the corresponding time dependent probabilities.

CHAPTER 3

Continuous Time Stochastic Volatility Models with Regime Shift

3.1. Introduction

Modelling and estimating the volatility of financial assets is one of the central questions in modern econometric research. The two main widely used methods are the univariate GARCH family of models and stochastic volatility models. Compared with discrete time models, continuous time models are closer to the stochastic behaviour of financial data. This is because, although financial data are available as discrete time observations, the economy does not cease to exist in between observations (Phillips (1988)). Actually, economic activities take place without a moment's pause and financial decision makers can make and change their dynamic trading strategies every second.

In early work, Black and Scholes (1973) propose that the stock market index S&P 500 follows a geometric Brownian motion process. The main drawback of this model is that the return process has constant volatility. In practice, instantaneous relative volatilities of the asset price are time-varying across time and strikes. Furthermore, since Mandelbrot (1963), there are many literatures show the asset returns have heavy-tailed character. In order to solve this problem, some two-factor stochastic volatility models are proposed and proved outperforming than Black-Scholes Model, such as Wiggins (1987), Hull and White (1987), Stein and Stein (1991), Heston (1993), Duffie, Pan, and Singleton (2000) and Jones (2003). These models allow not only the stock market index S&P 500 and its volatility follow stochastic processes, but also incorporate the correlation between log-returns and changes in variance. This is the main reason that this chapter chooses to use the continuous time stochastic volatility models to describe the behaviour of the stock market S&P 500 and its volatility.

Since Hamilton (1989), a substantial number of econometric literatures show the evidence of regime shift existing in financial series data and regime-switching models outperforming regular single regime models. In terms of discrete time models, for

example, Hamilton and Sunmel (1994) propose the regime-switching ARCH models, Fornari and Mele (1996) develop the regime-switching GARCH models. Relative to discrete time stochastic volatility models, Calvet and Fisher (2004) introduce discrete-time regime-switching stochastic volatility models. With respect to the univariate continuous time models, Choi (2009) proposes the regime-switching univariate diffusion models. However, there are scarcely any literatures combining regime shifts with multivariate continuous time stochastic volatility models because of the substantial difficulties coming from the estimating of multivariate stochastic volatility models.

On one hand, the transition density of the state vector is hardly ever known in closed form so that the standard maximum likelihood estimation cannot be conducted. On the other hand, the additional state variables which determine the level of volatility are usually unobserved in practise. Kleppe, Yu, and Skaug (2010) consider the Euler approximations for the transition probability density functions and employ a modified efficient importance sampling (EIS) algorithm, which is developed by Richard and Zhang (2007), to integrate out a latent volatility process. Compared to the Euler approximation, Aït-Sahalia and Kimmel (2007) and Aït-Sahalia (2008) provide an improved closed-form approximation for the true joint transition density and propose an adjusted Black-Scholes proxy. Following their work, Choi (2013) finds closed-form likelihood expansions for multivariate time-inhomogeneous diffusions and conducts Monte Carlo to produce all state variables. In this chapter, we apply the method, which is developed in Chapter 2, to estimate continuous time stochastic volatility models with regime shifts.

This chapter is organized as follows. Section 2 motivates the econometric specifications by plotting two series of financial data. Section 3 introduces three families of continuous time stochastic volatility models with regime shift. Section 4 summarizes and discusses the results, including parameter estimates, filtered and smoothed probabilities, approximate conditional transition density functions for each regime, respectively. Then conclusion follows.

3.2. Data and Motivation

Daily Standard & Poor's 500 (S&P 500) and the Chicago Board Options Exchange (CBOE) Volatility Index (VIX), from January 2, 1990 to June 12, 2012, are downloaded from DataStream for the estimation. The VIX is calculated as the square root

of the risk neutral expectation of the S&P 500 variance for a 30 days term initiated today. The descriptive statistics of S&P 500 and VIX series are shown in Table 3.1.

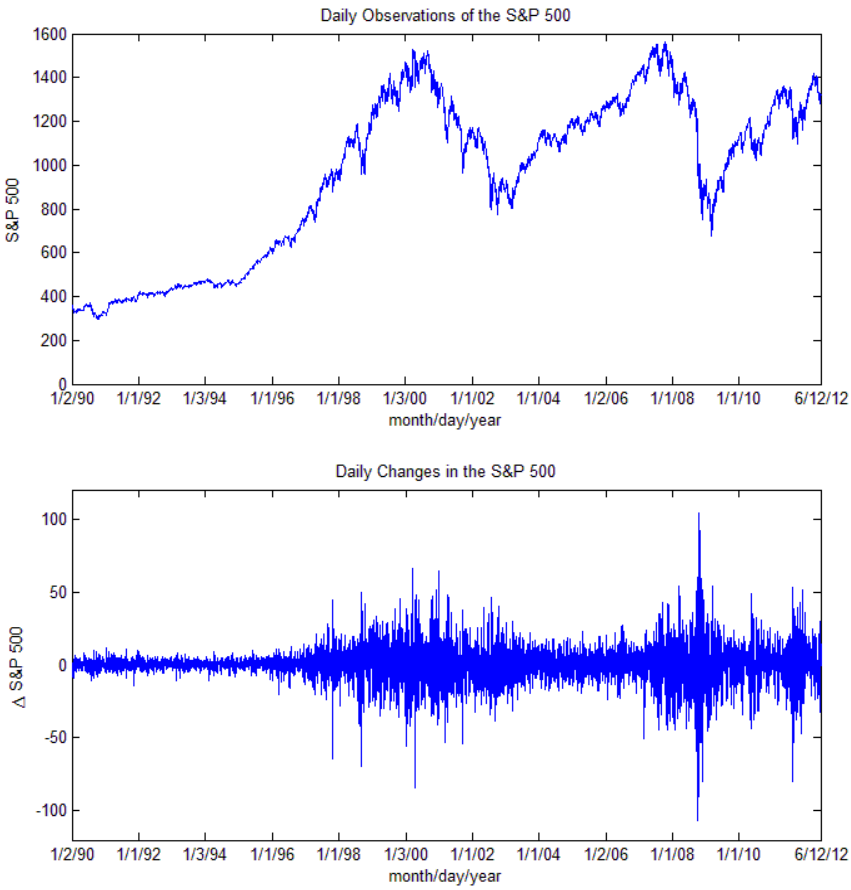
Table 3.1. Summary Statistics of S & P 500 and VIX

	S&P 500	VIX
Observations	5856	5856
Minimum	295.460	9.310
Maximum	1565.150	80.860
Mean	954.314	20.518
Std. Dev	370.450	8.188
Skewness	-0.365	1.944
Kurtosis	1.737	9.763

Plots of the two time series data and the first difference of them illustrate that S&P 500 and VIX behave quite differently in different time periods. They tend to exhibit periods of high and low volatility, which is a stylized phenomenon called volatility clustering in financial time series. Most periods of high volatility clustering can be associated with some significant historic events. With respect to the variance of S&P 500 and/or VIX, there seems to be at least two regimes. This characteristic of the two time series data motivates us to combine regime-switching with continuous time stochastic volatility models.

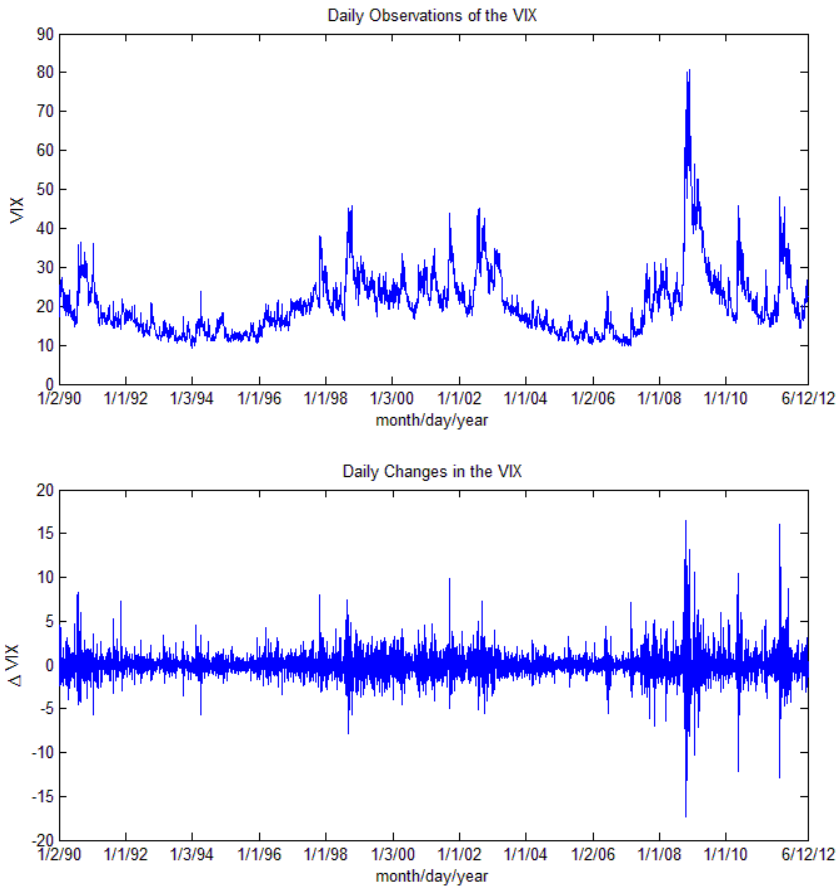
The data set covers the period of 1990 to 2012 during which a number of significant events caused strong changes in the behaviours of S&P 500 and VIX. The Gulf War started from 2 August 1990 and last to 28 February 1991. The Union of Soviet Socialist Republics officially ceased to exist on December 31, 1991. The British Conservative government withdrawing the pound from the European Exchange Rate Mechanism caused 1992 Black Wednesday, which led to estimated £800 million trading losses in August and September. The 1994 Northridge earthquake in the Los Angeles area caused an estimated $20 billion in damage, making it one of the costliest natural disasters in U.S. history. At the end of 1995, the severe budget crisis forced the federal government to shut down for several weeks. Then the Asian financial crisis started in Thailand in July 1997 and then triggered the 1998 Russian financial crisis. The 1998 collapse of long-term capital management leads to an agreement among 14 financial institutions for a $3.6 billion recapitalization under the supervision of Federal Reserve. During the period of 1997 to 2000, in industrialized nations, the stock prices of Internet companies shoot up and then the dot-com bubble finally

Figure 3.1. Daily Observations and Changes of S&P 500



burst because of an inflation report in April of 2000. September 11 attacks in 2001 had immediate and far-ranging economic effects. After the October 2001 invasion in Afghanistan, the Iraq War was launched on March 19, 2003. The 2004 Atlantic hurricane season impacted Florida, Charley, Frances, Ivan and Jeanne, and produced over $50 billion in damage. What's even worse, two powerful hurricanes, Hurricane Katrina and Hurricane Rita hit the Gulf Coast region in August and September 2005. The Chinese Correction plunge of February 27, 2007 caused the Dow Jones Industrial Average dropped by 3.29%, which is the biggest one-day slide since the September 11. The subprime mortgage crisis started in late 2006. It triggered a global financial crisis through 2007 and 2008 and then rapidly evolved into a global financial crisis. On September 15, 2008 Lehman Brothers filed for bankruptcy protection. In the late of 2009, a Dubai debt standstill caused a sharp drop in global stock markets. On April 27, 2010, Standard & Poor's downgrades Greece's sovereign credit rating, triggering another decline in global stock markets and furthering the European sovereign-debt crisis. The 2010 Flash Crash led the Dow Jones Industrial Average drop nearly 1000

Figure 3.2. Daily Observations and Changes of VIX



points, which is the worst intra-day point loss in history. Due to the fears of contagion of the European sovereign debt crisis to Spain and Italy and downgrading of the US's credit rating, stock prices dropped sharply in August 2011 in stock exchanges across the United States, the Middle East, Europe and Asia. After the August 2011 stock markets fall, severe volatility of stock market indexes is continuing until now.

3.3. Continuous Time Stochastic Volatility Models with Regime Shift

In a stochastic volatility model, the stock market index S&P 500 is a function of a vector of state variables X_t that follows an m-dimensional multivariate diffusion process,

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

where state variables X_t and the drift functions $\mu^P(X_t;\theta)$ are $m \times 1$ vectors, $\sigma(X_t;\theta)$ is an $m \times m$ volatility matrix, and W_t^P is an $m \times 1$ vector of independent Brownian motions under the objective probability measure P. Both $\mu^P(X_t;\theta)$ and $\sigma(X_t;\theta)$ depend on X_t and they are known up to a parameter vector $\theta \in \Theta$, which is a

compact subset of R^p. To avoid loss of generality, any correlation structures between the state variables can be modelled by using the off-diagonal terms in the volatility matrix $\sigma(X_t;\theta)$, which could be asymmetric. In our model, state variables X_t is a 2×1 vector, including S_t and Y_t representing the stock market index S&P 500 and its variance respectively.

Followed Harrison and Kreps (1979) and Harrison and Pliska (1981), the physical measure P can be re-stated in terms of the risk-neural measure denoted by Q,

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

where $\mu^Q(X_t;\theta)$ is an m-dimensional function of X_t and W_t^Q is an $m \times 1$ vector of independent Brownian motions under Q.

Our models combine regime switching with different stochastic volatility models under the objective probability measure P, in which the dynamics of the stock market index S&P 500 and its volatility are assumed to follow a bivariate continuous time regime-shifting Markov process:

$$dX_t = \mu^P(X_t, R_t;\theta)\,dt + \sigma(X_t, R_t;\theta)\,dW_t^P. \tag{3.1}$$

Under the equivalent martingale measure Q, equation 3.1 is rewritten as

$$dX_t = \mu^Q(X_t, R_t;\theta)\,dt + \sigma(X_t, R_t;\theta)\,dW_t^Q,$$

where the regime index R_t follows a two-state continuous time Markov chain. Strictly speaking, the future state depends on the previous one state. Motivated by the behaviour of S&P 500 and VIX in Figures 3.1 and 3.2, low (L) and high (H) volatility regimes are assumed to identify two different economic situations. Therefore, the stock market index S&P 500 and its volatility can evolve in an L regime or an H regime at any given point. However, the true state of the economy has to be inferred all the times because it cannot be observed directly.

The two-regime specification allows each parameter has two different estimates depending on the two different economic states. Therefore, the two-regime framework can help explain some significant properties of asset prices. First, a two-regime model performs better than a single-regime model when explaining volatility clustering. The strong persistence of the high volatility regime and the distinct long-run average price variances of different regimes are able to capture high volatility clustering. Second,

a stochastic volatility model with two regimes can help analyse leverage effect better in different volatility situations. Third, the high variance regime inferred by our regime-switching models can identify most important incidents in financial history.

Given a two state, $\{L, H\}$, we assume the shifts are controlled by the two state continuous time Markov chain with the generator matrix

$$\Lambda = \begin{pmatrix} \lambda_{LL} & \lambda_{HL} \\ \lambda_{LH} & \lambda_{HH} \end{pmatrix} = \begin{pmatrix} -\lambda_{LH} & \lambda_{HL} \\ \lambda_{LH} & -\lambda_{HL} \end{pmatrix},$$

where λ_{ij} is the transition rate representing the rate of transitioning from the state i to the state j as time goes to zero and $\pi_{ij} > 0$ for $i \neq j$. We assume, during each time period Δ, regime switching can occur at most once and at most one regime shift can occur every day. The corresponding transition matrix $P(t)$ can be obtained by solving Kolmogorov backwards equations or Kolmogorov forwards equations.

$$\begin{aligned} P(t) &= \begin{pmatrix} P_{LL}(t) & P_{HL}(t) \\ P_{LH}(t) & P_{HH}(t) \end{pmatrix} \\ &= \frac{1}{\lambda_{LH} + \lambda_{HL}} \begin{pmatrix} \lambda_{HL} + \lambda_{LH} e^{-(\lambda_{LH}+\lambda_{HL})t} & \lambda_{HL}\left(1 - e^{-(\lambda_{LH}+\lambda_{HL})t}\right) \\ \lambda_{LH}\left(1 - e^{-(\lambda_{LH}+\lambda_{HL})t}\right) & \lambda_{LH} + \lambda_{HL} e^{-(\lambda_{H}+\pi_{HL})t} \end{pmatrix}, \end{aligned}$$

in which the intensity λ_{ij} and the time period t affect the transition probabilities together. Additionally, the unconditional probabilities of the above Markov being in the regime L and H are

$$(\lambda_L, \lambda_H) = \left(\frac{\lambda_{HL}}{\lambda_{LH}+\lambda_{HL}}, \quad \frac{\lambda_{LH}}{\lambda_{LH}+\lambda_{HL}} \right),$$

respectively at any time. When estimating our model, we assume the regime index R_t follows a discrete time Markov chain of order 2. The transition matrix P can be written as

$$P = \begin{pmatrix} p_{LL} & p_{HL} \\ p_{LH} & p_{HH} \end{pmatrix},$$

where p_{xy} is the transition probability that the chain enters state y after leaving state x and $p_{xy} = P(R_t = x | R_{t-1} = y), x, y = L, H$. With respect to p_{LL} and p_{HH}, we restate the stationary distribution of Markov chain as

$$(\pi_L, \pi_H) = \left(\frac{1-p_{HH}}{2-p_{LL}-p_{HH}}, \quad \frac{1-p_{LL}}{2-p_{LL}-p_{HH}} \right).$$

In order to make the generator matrix vary with the state variables s_t and Y_t, we model the transition probabilities

$$\begin{aligned} p_{LL} &= P\left(R_t = L | R_{t-\Delta} = L\right) = F\left(c_L + d_L s_{t-\Delta} + e_L Y_{t-\Delta}\right) \\ p_{LH} &= P\left(R_t = H | R_{t-\Delta} = L\right) = 1 - F\left(c_L + d_L s_{t-\Delta} + e_L Y_{t-\Delta}\right) \\ p_{HH} &= P\left(R_t = H | R_{t-\Delta} = H\right) = F\left(c_H + d_H s_{t-\Delta} + e_H Y_{t-\Delta}\right) \\ p_{HL} &= P\left(R_t = L | R_{t-\Delta} = H\right) = 1 - F\left(c_H + d_H s_{t-\Delta} + e_H Y_{t-\Delta}\right), \end{aligned}$$

where Δ denotes a time period. Following previous literatures, $F(x)$ could be the cumulative Normal distribution function, $\Phi(x)$ (Gray (1996)) or the logistic function, $\exp(x)/[1+\exp(x)]$ (Diebold, Lee, and Weinbach (1994) and Dai, Singleton, and Yang (2007)), because both of them always yield nonnegative real numbers results, which are between 0 and 1. When $d_L = d_H = 0$ and $e_L = e_H = 0$, the transition probabilities become constant and the time-varying transition matrix is reduced to the time-constant transition matrix. If $d_i > 0$ $(d_i < 0)$, the probability of the regime index keeping in the same state in the next time period has a positive (negative) relationship with the stock market index S&P 500. If $e_i > 0$ $(e_i < 0)$, the probability of the regime index keeping in the same state in the next period has a positive (negative) relationship with the volatility of the stock market index S&P 500. Since Choi (2009) compared the two different transition probability functions when estimating two-regime diffusion models of the short-term interest rate and showed there is a negligible difference, we choose to use the logistic function.

When estimating our model, two methods are considered to set the initial state probability. One is to treat the unconditional probability of the original state, $P(s_0 = L)$ as an additional parameter p. Another is setting the original state probability equal the unconditional probability $\left(\frac{1-p_{HH}}{2-p_{LL}-p_{HH}}, \frac{1-p_{LL}}{2-p_{LL}-p_{HH}} \right)$.

We combine the regime shift with three stochastic volatility models, including the Heston model, the GARCH model and the CEV model. According to the number of regimes, the initial probability and the transition probability matrix specifications, we will estimate four models for each group. For Heston model, I will compare H-R1 (single-regime), H-R2-1 (two regimes, time-constant transition matrix, unconditional probability for the probability of the initial state), H-R2-2 (two regimes, time-constant transition matrix, additional parameter for the probability of the initial state) and

H-R2TVTP(two regimes, time-varying transition matrix with a logistic function, additional parameter for the probability of the initial state). For the GARCH model and the CEV model, We will estimate G-R1, G-R2-1, G-R2-2, G-R2TVTP and C-R1, C-R2-1, C-R2-2, C-R2TVTP, separately.

3.3.1. Heston Stochastic Volatility Model with regime shift

The Heston model, named after Heston (1993), is a commonly used stochastic volatility model. Under the equivalent martingale measure Q, the asset price S_t and its volatility Y_t follow the dynamics

$$dX_t = d\begin{bmatrix} S_t \\ Y_t \end{bmatrix} = \begin{bmatrix} (r-d)\,S_t \\ \kappa'(\gamma' - Y_t) \end{bmatrix} dt + \begin{bmatrix} \sqrt{(1-\rho^2)\,Y_t}S_t & \rho\sqrt{Y_t}S_t \\ 0 & \sigma\sqrt{Y_t} \end{bmatrix} d\begin{bmatrix} W_1^Q(t) \\ W_2^Q(t) \end{bmatrix}, \tag{3.2}$$

where $dW_1^Q(t)$, $dW_2^Q(t)$ are Wiener processes with correlation ρdt, r is risk free interest rate, d is dividend yield of the stock, γ' is the long run average price variance, κ' is the rate at which Y_t reverts to γ', σ is the volatility of volatility. Because Y_t follows the square root process, it's lower bound is zero. To avoid reaching the boundary value, the Feller condition $2\kappa'\gamma' \geq \sigma^2$ should be satisfied.

With respect to the logarithmic asset price $s_t = \ln S_t$, the dynamics 3.2 are then rewritten as

$$dX_t = d\begin{bmatrix} s_t \\ Y_t \end{bmatrix} = \begin{bmatrix} r - d - \frac{1}{2}Y_t \\ \kappa'(\gamma' - Y_t) \end{bmatrix} dt + \begin{bmatrix} \sqrt{(1-\rho^2)\,Y_t} & \rho\sqrt{Y_t} \\ 0 & \sigma\sqrt{Y_t} \end{bmatrix} d\begin{bmatrix} W_1^Q(t) \\ W_2^Q(t) \end{bmatrix}$$

by Itô lemma.

The market price of risk is $\left[\lambda_1\sqrt{(1-\rho^2)\,Y_t}, \lambda_2\sqrt{Y_t}\right]'$. Under the objective measure P, the joint dynamics of s_t and Y_t is

$$dX_t = d\begin{bmatrix} s_t \\ Y_t \end{bmatrix} = \begin{bmatrix} a + bY_t \\ \kappa(\gamma - Y_t) \end{bmatrix} dt + \begin{bmatrix} \sqrt{(1-\rho^2)\,Y_t} & \rho\sqrt{Y_t} \\ 0 & \sigma\sqrt{Y_t} \end{bmatrix} d\begin{bmatrix} W_1^P(t) \\ W_2^P(t) \end{bmatrix},$$

where $a = r - d$, $b = \lambda_1(1-\rho^2) + \lambda_2\rho - \frac{1}{2}$, $\kappa = \kappa' - \lambda_2\sigma$, $\gamma = \left(\frac{k+\lambda_2\sigma}{\kappa}\right)\gamma'$.

Since we treated the volatility as observed when estimating, we cannot identify the market price of risk of the stochastic volatility variable λ_2. As Aït-Sahalia and Kimmel (2007) did, we assume $\lambda_2 = 0$ for all of the specifications. Therefore, the parameter b is reduced to $\lambda(1-\rho^2) - \frac{1}{2}$ and λ is used instead of λ_1. Our two-regime Heston model assumes that both the dynamics of logarithmic S&P 500 ,s_t, and its

volatility, Y_t, follow continuous time regime-switching Markov processes

$$dX_t = d\begin{bmatrix} s_t \\ Y_t \end{bmatrix} = \begin{bmatrix} a_{R_t} + b_{R_t} Y_t \\ \kappa_{R_t}\left(\gamma_{R_t} - Y_t\right) \end{bmatrix} dt + \begin{bmatrix} \sqrt{\left(1-\rho_{R_t}^2\right) Y_t} & \rho_{R_t}\sqrt{Y_t} \\ 0 & \sigma_{R_t}\sqrt{Y_t} \end{bmatrix} d\begin{bmatrix} W_1^P(t) \\ W_2^P(t) \end{bmatrix},$$

where the regime index R_t follows a two states continuous time first order Markov chain.

3.3.2. GARCH Stochastic Volatility Model with regime shift

Under the risk-neutral measure Q, the stock market index S&P 500, S_t, and its volatility Y_t follow the dynamics

$$dX_t = d\begin{bmatrix} S_t \\ Y_t \end{bmatrix} = \begin{bmatrix} (r-d)\, S_t \\ \kappa'\left(\gamma' - Y_t\right) \end{bmatrix} dt + \begin{bmatrix} \sqrt{\left(1-\rho^2\right) Y_t} S_t & \rho\sqrt{Y_t} S_t \\ 0 & \sigma Y_t \end{bmatrix} d\begin{bmatrix} W_1^Q(t) \\ W_2^Q(t) \end{bmatrix}, \tag{3.3}$$

where the parameters obey the conditions $\kappa'\gamma' \geq 0$. $dW_1^Q(t)$, $dW_2^Q(t)$ are Wiener processes with correlation ρdt. We could see the difference between the Heston and GARCH model is the volatility term of Y_t. In terms of $s_t = \ln S_t$, the joint dynamics 3.3 of s_t and Y_t is

$$dX_t = d\begin{bmatrix} s_t \\ Y_t \end{bmatrix} = \begin{bmatrix} r - d - \frac{1}{2} Y_t \\ \kappa'\left(\gamma' - Y_t\right) \end{bmatrix} dt + \begin{bmatrix} \sqrt{\left(1-\rho^2\right) Y_t} & \rho\sqrt{Y_t} \\ 0 & \sigma Y_t \end{bmatrix} d\begin{bmatrix} W_1^Q(t) \\ W_2^Q(t) \end{bmatrix}$$

by Itô lemma.

Since the same reason as mentioned in the Heston model, the market price of risk is reduced to $\left[\lambda_1\sqrt{\left(1-\rho^2\right) Y_t}, 0\right]'$. The dynamics of the state variables under the measure P are

$$dX_t = d\begin{bmatrix} s_t \\ Y_t \end{bmatrix} = \begin{bmatrix} a + bY_t \\ \kappa\left(\gamma - Y_t\right) \end{bmatrix} dt + \begin{bmatrix} \sqrt{\left(1-\rho^2\right) Y_t} & \rho\sqrt{Y_t} \\ 0 & \sigma Y_t \end{bmatrix} d\begin{bmatrix} W_1^P(t) \\ W_2^P(t) \end{bmatrix},$$

where $a = r - d$, $b = \lambda\left(1-\rho^2\right) - \frac{1}{2}$, $\kappa = \kappa'$, $\gamma = \gamma'$.

The logarithmic S&P 500 s_t and its volatility Y_t is modelled as

$$dX_t = d\begin{bmatrix} s_t \\ Y_t \end{bmatrix} = \begin{bmatrix} a_{R_t} + b_{R_t} Y_t \\ \kappa_{R_t}\left(\gamma_{R_t} - Y_t\right) \end{bmatrix} dt + \begin{bmatrix} \sqrt{\left(1-\rho_{R_t}^2\right) Y_t} & \rho_{R_t}\sqrt{Y_t} \\ 0 & \sigma_{R_t} Y_t \end{bmatrix} d\begin{bmatrix} W_1^P(t) \\ W_2^P(t) \end{bmatrix},$$

where the regime R_t follows a continuous time first order Markov chain with two states.

3.3.3. CEV Stochastic Volatility Model with regime shift

Jones (2003) proposed a more general model, which nests the Heston model ($\beta = 1/2$) and the GARCH model ($\beta = 1$). Under the risk-neutral measure Q, the stock market index S&P 500 S_t and its volatility Y_t follow the dynamics

$$(3.4) \quad dX_t = d\begin{bmatrix} S_t \\ Y_t \end{bmatrix} = \begin{bmatrix} (r-d)\, S_t \\ \kappa'(\gamma' - Y_t) \end{bmatrix} dt + \begin{bmatrix} \sqrt{(1-\rho^2)\, Y_t} S_t & \rho\sqrt{Y_t} S_t \\ 0 & \sigma Y_t^{\beta} \end{bmatrix} d\begin{bmatrix} W_1^Q(t) \\ W_2^Q(t) \end{bmatrix},$$

where $1/2 \leq \beta \leq 1$. In terms of the logarithmic stock price, the process 3.4 is rewritten as

$$dX_t = d\begin{bmatrix} S_t \\ Y_t \end{bmatrix} = \begin{bmatrix} r - d - \frac{1}{2} Y_t \\ \kappa'(\gamma' - Y_t) \end{bmatrix} dt + \begin{bmatrix} \sqrt{(1-\rho^2)\, Y_t} & \rho\sqrt{Y_t} \\ 0 & \sigma Y_t^{\beta} \end{bmatrix} d\begin{bmatrix} W_1^Q(t) \\ W_2^Q(t) \end{bmatrix}$$

by Itô lemma.

Because of $\Lambda = \left[\lambda_1 \sqrt{(1-\rho^2)\, Y_t}, 0\right]'$, P-measure dynamics of the state variables are then

$$dX_t = d\begin{bmatrix} s_t \\ Y_t \end{bmatrix} = \begin{bmatrix} a + bY_t \\ \kappa(\gamma - Y_t) \end{bmatrix} dt + \begin{bmatrix} \sqrt{(1-\rho^2)\, Y_t} & \rho\sqrt{Y_t} \\ 0 & \sigma Y_t^{\beta} \end{bmatrix} d\begin{bmatrix} W_1^P(t) \\ W_2^P(t) \end{bmatrix},$$

where $a = r - d$, $b = \lambda(1-\rho^2) - \frac{1}{2}$, $\kappa = \kappa'$, $\gamma = \gamma'$.

Our regime-switching model is

$$dX_t = d\begin{bmatrix} s_t \\ Y_t \end{bmatrix} = \begin{bmatrix} a_{R_t} + b_{R_t} Y_t \\ \kappa_{R_t}\left(\gamma_{R_t} - Y_t\right) \end{bmatrix} dt + \begin{bmatrix} \sqrt{\left(1-\rho_{R_t}^2\right) Y_t} & \rho_{R_t}\sqrt{Y_t} \\ 0 & \sigma_{R_t} Y_t^{\beta_{R_t}} \end{bmatrix} d\begin{bmatrix} W_1^P(t) \\ W_2^P(t) \end{bmatrix},$$

where the regime R_t follows a continuous time first order markov chain with two states.

3.4. Estimation Results

The transition density of the state vector is hardly ever known in closed form so that the standard maximum likelihood estimation cannot be conducted. To estimate the continuous time stochastic volatility models with regime shifts, we apply the maximum likelihood estimation method, which is developed in Chapter 2.

In our three different families of stochastic volatility models, there are just two elements in the state vector X_t. The first element S_t representing daily Standard & Poor's 500 Index (S&P 500) can be directly observed, however, the state variable Y_t

Table 3.2. Summary Statistics of S&P 500, VIX, ln(S&P 500) and IV

	S&P 500	VIX	ln(S&P 500)	IV
Observations	5856	5856	5856	5856
Minimum	295.460	9.310	5.689	0.003
Maximum	1565.150	80.860	7.356	0.725
Mean	954.314	20.518	6.763	0.048
Std. Dev	370.450	8.188	0.474	0.055
Skewness	-0.365	1.944	-0.745	4.627
Kurtosis	1.737	9.763	2.083	36.633

which determines the level of volatility is unobserved. As mentioned in Aït-Sahalia and Kimmel (2007), two different methods can be applied to deal with this problem in estimation. One method is to employ a filtering technique (Bates (2006)), the alternative one is to assume an option price C_t is observed, then the time series of Y_t can be extracted from C_t. Since the former method cannot fully identify all of the parameters under the equivalent martingale measure Q, we choose the latter approach and use the Chicago Board Options Exchange Volatility Index (VIX) as C_t. In order to infer the value of Y_t from C_t, there are three approaches can be applied. One approach is to compute the value of Y_t from a function of the stock price S_t and an option price C_t. The second method is proposed by Ledoit, Santa-Clara, and Yan (1998), in which the Black-Scholes implied volatility of an at-the-money short-maturity option is treated as a proxy for the instantaneous volatility state variable S_t. Motivated by Hull and White (1987), Aït-Sahalia and Kimmel (2007) proposed an adjusted Black-Scholes proxy, which can avoid the significant bias caused by using the simple unadjusted Black-Scholes proxy in the estimation of the elasticity of volatility parameter. Therefore, we choose to employ the modified Black-Scholes proxy for the state variable Y_t.

If the Q-measure drift of Y_t is of the form $a + bY_t$, then the adjusted proxy, IV_t, is

$$IV_t = \frac{b \cdot \tau \cdot Y_t^{imp} + a \cdot \tau}{\exp(b \cdot \tau) - 1} - \frac{a}{b}, \tag{3.5}$$

where Y_t^{imp} is the implied volatility of the short-maturity at-the-money option and τ is time to maturity of the option.

We first estimate parameters of the following univariate CEV model for $(VIX_t/100)^2$

$$dX_t = \kappa(\gamma - X_t)\,dt + \sigma X_t^{\beta} dW_t$$

to get $a = \kappa\gamma$ and $b = -\kappa$.

Because the true log-likelihood function of the CEV model is unknown, the approximate log-likelihood function was obtained using the irreducible method to conduct maximum likelihood estimation. We use the daily S&P 500 and VIX downloaded from DataStream between January 2, 1990 and June 12, 2012 and get the estimation results

$$\begin{aligned} dY_t &= 2.61^{**}\left(0.054^{**} - Y_t\right)dt + 1.28^{**}X_t^{0.84^{**}}dW_t \\ &\qquad (0.726)\ \ (0.010) \qquad\qquad\qquad (0.011)\ \ (0.002) \end{aligned}$$

with standard errors in parentheses.

We use $\left(VIX_t/100\right)^2$ for Y_t^{imp} and estimates of a and b in equation 3.5 to compute the adjusted volatility such that

$$IV_t = -0.0064 + 1.1182\left(VIX_t/100\right)^2.$$

In Table 3.2, we give the descriptive statistics of S&P 500, VIX, logarithmic S&P 500 and IV series from January 2,1990 to June 12, 2012.

In Table 3.3, we summarize the ML estimation results for the four different Heston stochastic volatility models. The first column shows estimation outcomes for the single-regime model H-R1. The second and third columns report the estimation results for two regime-switching models with time invariant transition matrix. The difference between them is the approach we use to deal with the probability of the original state. We keep the unconditional probability at the value of 0.5 for the H-R2-1 model, but set an additional parameter to estimate for the model H-R2-2. The last column presents ML estimation outcomes for the regime-switching model with varying transition matrix and an additional parameter for the probability of the original state. The asterisk by the parameter estimate implies that, at the 5% significance level, it is different from zero.

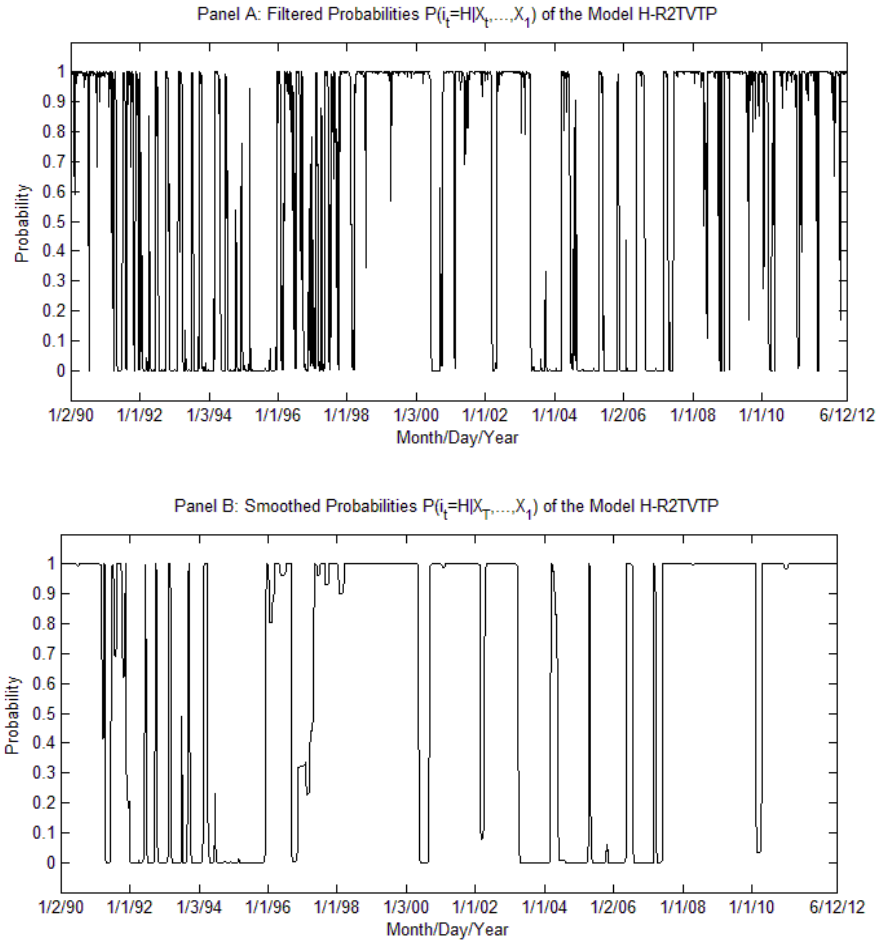
In the single regime model H-R1, besides the market price of risk λ and the rate of return rd, the rest estimates are statistically significant. The long run average volatility $\gamma^{1/2}$ is estimated to be approximately 21.45% per year with a rate of mean reversion coefficient κ of 6.817. The correlation coefficient ρ between the innovations to the stock price and stochastic volatility is strongly negative, -0.799. The variance of volatility σ is close to 0.726.

Table 3.3. Maximum Likelihood Estimation Results of Main Heston Models(01/02/1990-06/12/2012)

Model	H-R1	H-R2-1	H-R2-2	H-R2TVTP
Regime	Single Regime	Two Regimes		
		Time Invariant Transition Matrix		Time Varying Transition Matrix
$P(R_1 = L)$	—	$\frac{1-p_{HH}}{2-p_{LL}-p_{HH}}$		p
$P(R_{t+\Delta} = i \vert R_t = i)$ with $i = L$ and H	—	p_{ii}	p_{ii}	$\frac{exp(c_i+d_i s_t+e_i Y_t)}{1+exp(c_i+d_i s_t+e_i Y_t)}$
Log-likelihood	42231.867	43094.026	43094.026	43293.294
Parameters	Maximum Likelihood Estimates(Standard Error)			
λ_L	6.927	6.000(7.718)	6.000(7.737)	6.000(4.897)
λ_H	(9.911)	6.000(6.313)	6.000(7.872)	6.000(4.148)
κ_L	6.817**	4.999**(1.049)	4.999**(1.043)	4.999**(0.763)
κ_H	(1.786)	6.999**(0.929)	6.999**(1.148)	6.999**(0.749)
γ_L	0.046**	0.018**(0.002)	0.018**(0.002)	0.019**(0.002)
γ_H	(0.017)	0.042**(0.005)	0.042**(0.006)	0.040**(0.005)
ρ_L	−0.799**	−0.600**(0.010)	−0.600**(0.010)	−0.623**(0.002)
ρ_H	(0.002)	−0.799**(0.002)	−0.799**(0.004)	−0.792**(0.004)
σ_L	0.726**	0.249**(0.006)	0.249**(0.006)	0.234**(0.001)
σ_H	(0.002)	0.558**(0.001)	0.558**(0.004)	0.633**(0.003)
rd_L	0.0002	0.0002(0.090)	0.0002(0.091)	0.0002(0.015)
rd_H	(0.234)	0.0002(0.099)	0.0002(0.138)	0.0002(0.008)
p_{LL}	–	0.099**(0.011)	0.099**(0.012)	–
p_{HH}	–	0.900**(0.011)	0.900**(0.011)	–
c_L	–	–	–	−0.199(2.915)
c_H	–	–	–	0.169(0.514)
d_L	–	–	–	−1.457**(0.447)
d_H	–	–	–	1.237**(0.083)
e_L	–	–	–	−0.014(9.554)
e_H	–	–	–	0.010(0.483)
p	–	–	0.500(1.010)	0.50(0.661)
AIC	−14.421	−14.713	−14.712	−14.779
BIC	−14.414	−14.697	−14.695	−14.757
RCM	–	27.031	27.027	6.388

Note: Maximum likelihood estimates with standard errors in parentheses for the period 01/02/1990-06/12/2012 for the four different models with the general diffusion specification are presented in this table. The standard errors are calculated using sample mean of the outer product of the score functions evaluated at the ML estimates. According to the number of regimes, the starting probability and the transition probability matrix specifications, the four models are: H-R1 (single regime),H-R2-1 (two regimes, time invariant transition matrix,unconditional probability for the probability of initial state), H-R2-2 (two regimes, time invariant transition matrix, additioanl parameter for the probability of initial state), H-R2TVTP (two regimes,time varying transition matrix with logistic function,additional parameter for the probability of initial state).

Figure 3.3. Regime-Switching Probabilities of the Model H-R2TVTP

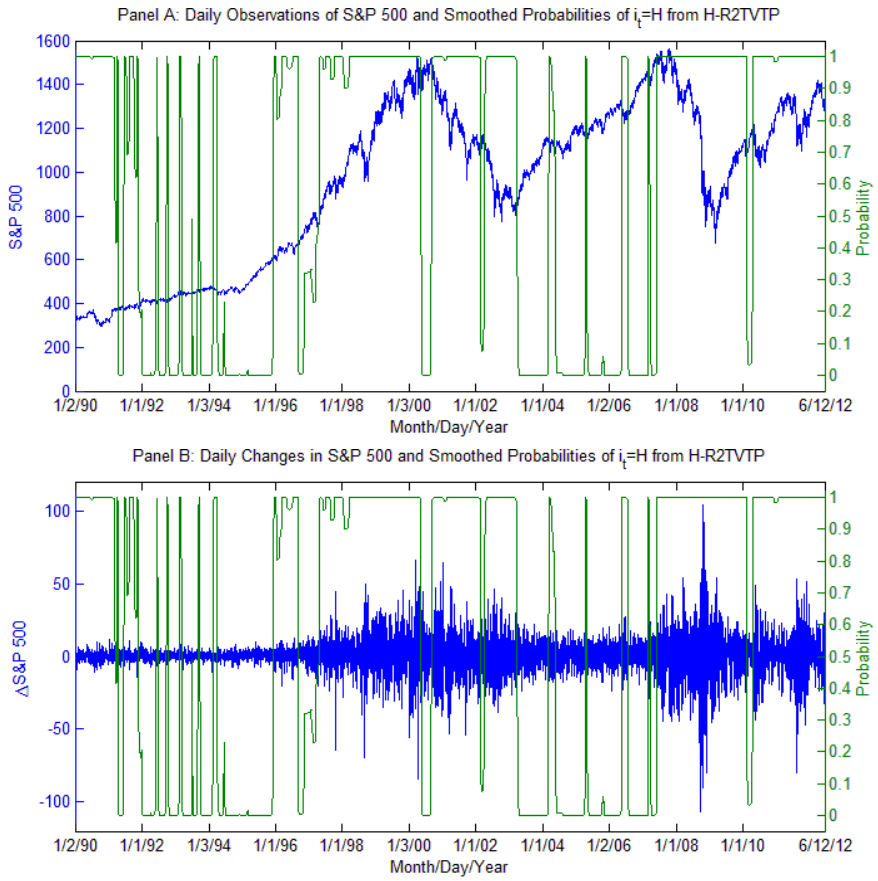


Comparing three different regime-switching models, we find that the rate of mean reversion of Y_t is higher in Regime H, $\kappa_H > \kappa_L$. The long-run average value of Y_t is bigger in Regime H, $\gamma_H > \gamma_L$. The correlation coefficient ρ between the innovations to stock price and stochastic volatility is stronger in Regime H, $\rho_H > \rho_L$. The variance of Y_t is more volatile in Regime H, $\sigma_H > \sigma_L$. rd_L, rd_H, λ_L and λ_H are statistically insignificant. Regime H is very persistent $p_{HH} \approx 0.900$ but Regime L is not so persistent $p_{LL} \approx 0.099$. In the model of H-R2TVTP, the variable s_t appears to be much more important in explaining the time varying transition probabilities of i_t.Because $d_L < 0$ ($d_H > 0$), p_{LL} (p_{HH}) decreases (increases) as s_t increases. Both c and e are statistically insignificant.

.

Analysing the log-likelihood values from the four different models, it illustrates that the single regime model H-R1 reports the smallest value 42231.867 while the

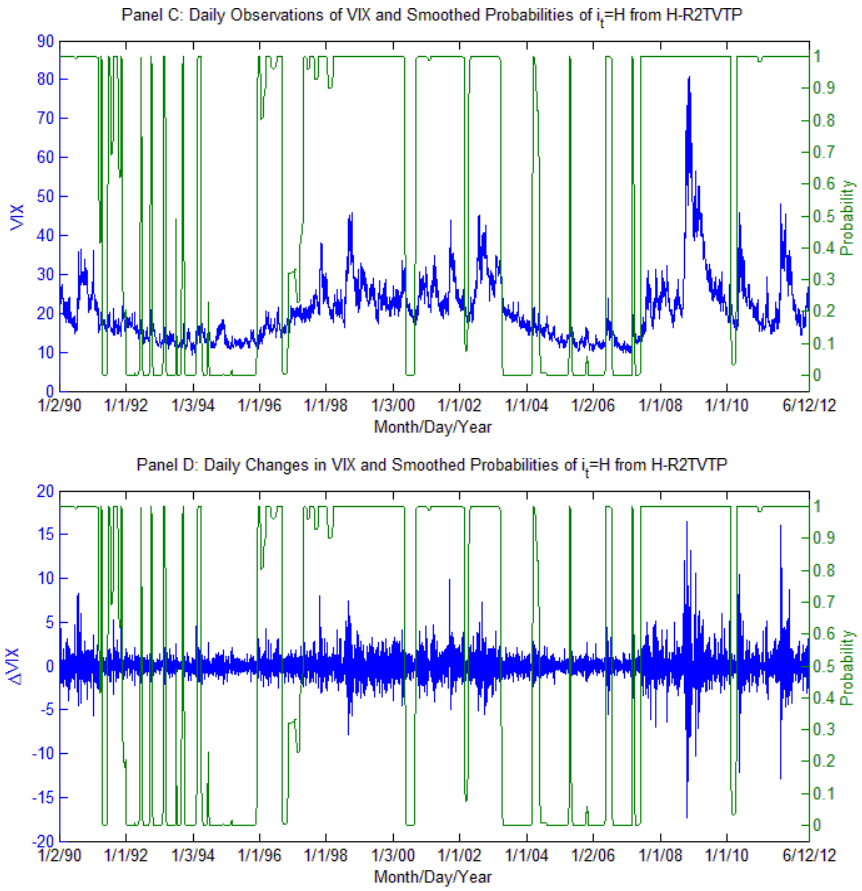
Figure 3.4. Regime-Switching Probabilities of the Model H-R2TVTP with S&P 500 and its First Difference



model H-R2TVTP owns the biggest one 43293.294. The models of H-R2-1 and H-R2-2 display the same log-likelihood value with very similar ML estimates. There seems to be not much difference between using unconditional probability and employing a new parameter for the original state. However, there is strong evidence of existence of high and low volatility regimes for the time varying transition probability of the regime variable.

Since the parameters related to the second regime of the process are unidentified under the null hypothesis of single regime, traditional test statistics cannot be used to test whether there is one regime or not. Take the model of H-R2-1 for example, given $P(s_1 = L) = 1$, if the null hypothesis is $p_{LL} = 1$, then λ_H, κ_H, γ_H, ρ_H, σ_H, rd_H, and p_{HH} cannot be identified under the null of no regime switching. In this case, standard asymptotic distribution theory cannot be applied so that standard likelihood ratio tests and Wald tests cannot be conducted. Although some literatures (Davis (1987), Hansen (1992), Hansen (1996) and Cho and White (2007)) try to address this problem, testing for multiple regimes is still particularly challenging. Therefore, we resort to

Figure 3.5. Regime-Switching Probabilities of the Model H-R2TVTP with VIX and its First Difference



calculate Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) to compare different specifications. Although the difference is very narrow, the H-R2TVTP model reports the smallest AIC and BIC values. Another metric we employed to compare the performance of different regime-switching models is called Regime Classification Measure (RCM). It is first proposed by Ang and Bekaert (2002) and then applied by Choi (2009) for R regime case as

$$RCM\left(R\right)=100R^{R}\frac{1}{n}\sum_{t=1}^{n}\left(\prod_{i=1}^{R}p_{i,t}\right),$$

where $p_{i,t}=P\left(s_t=i|I_T\right)$, $R=2$, $100R^R$ is used to normalize the statistic to be between 0 and 100. A model that does a good (bad) job of distinguishing between regimes will make an inference about s_t being in a regime close to 1 or 0 $(1/R)$. Therefore, according to the inferred probabilities of staying in a particular regime, the closer the RCM value is to 0, the better the regime classification of a model is. It can be seen from Table 3.3, the H-R2-1 and H-R2-2 models display very similar

RCM, but there is a great improvement in the model H-R2TVTP. In the term of RCM values, it also shows strong evidence for the existence of varying transition matrix.

In order to analyse the probability of being in state H, we draw the time series of filtered probabilities and smoothed probabilities over the sample for the model H-R2TVTP in Figure 3.3. To make it more clearly, we plot the time series of smoothed probabilities with S&P 500, VIX and their first difference separately in Figures 3.4 and 3.5.

These illustrations display that the H-R2TVTP model identifies most high volatility periods, which can be connected with some significant events in U.S. history and financial history since 1990. The beginning high volatility regime corresponds to the Gulf War, which began from 2 August 1990 and lasted to 28 February 1991. The next state of H reflects the Black Wednesday in 1992. The high volatility regime around 1994 matches the 1994 Northridge earthquake in the Los Angeles area, caused an estimated $20 billion in damage, making it one of the costliest natural disasters in U.S. history. The following H state throughout 1996 is linked to the severe budget crisis, which forced the federal government to shut down for several weeks at the end of 1995. The 1998-2000 high volatility regime is related to the Asian financial crisis, started in Thailand in July 1997 and then triggered the 1998 Russian financial crisis. It is also coincident with the 1998 collapse of long-term capital management, leading to an agreement among 14 financial institutions for a $3.6 billion recapitalization under the supervision of the Federal Reserve. Furthermore, the dot-com bubble covered roughly the period of 1995 to 2000, and finally burst because of an inflation report in April of 2000. The next three successive years of H state is caused by the September 11, 2001 terrorist attacks and the October 2001 invasion in Afghanistan. After that, the Iraq War was lanched on March 19, 2003. The 2004 high volatility regime is linked to the 2004 Atlantic hurricane season, which impacted Florida, Charley, Frances, Ivan and Jeanne, and produced over $50 billion in damage. What's even worse, in August and September 2005, two powerful hurricanes, Hurricane Katrina and Hurricane Rita hit the Gulf Coast region. The last long H state is associated with the result of a succession of financial events. The Chinese Correction plunge of February 27, 2007 caused the Dow Jones Industrial Average dropped by 3.29%, which is the biggest one-day slide since the September 11. The subprime mortgage crisis, which has its roots in the closing years of the 20th century, became apparent between 2007 and 2008, and

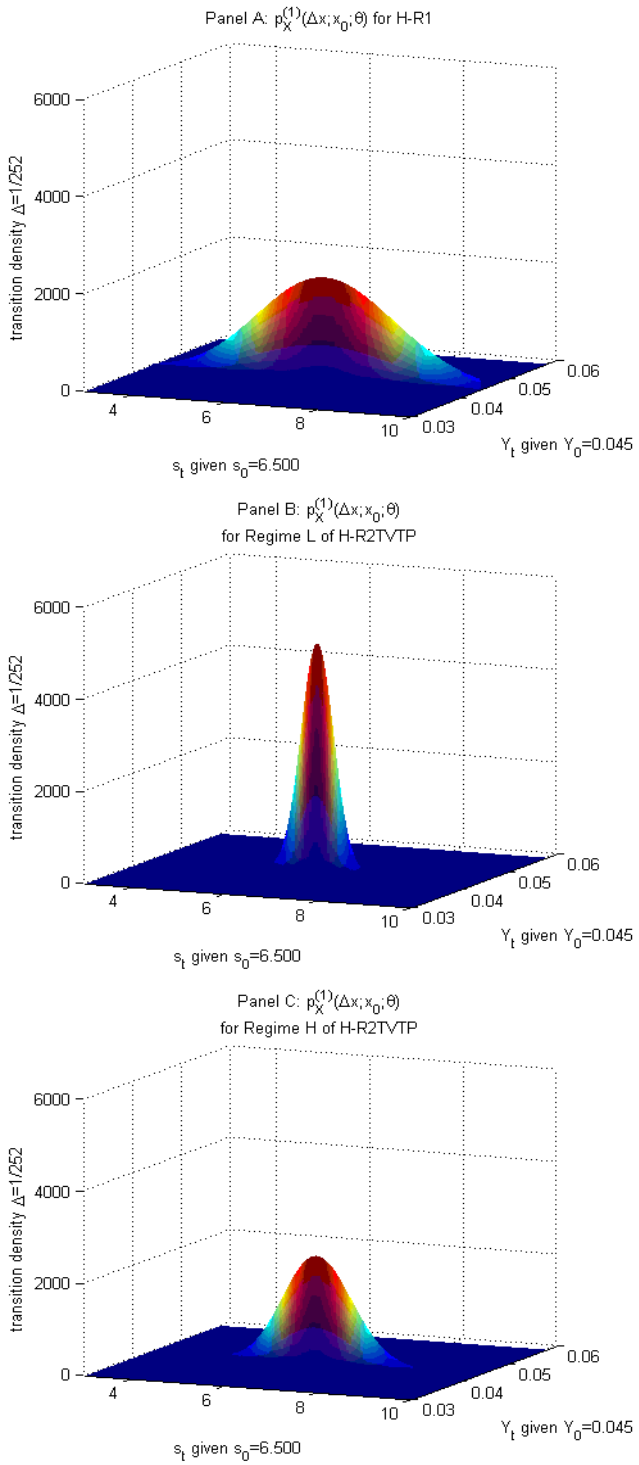
then rapidly evolved into a global financial crisis. On September 15, 2008 Lehman Brothers filed for bankruptcy protection. The European sovereign-debt crisis started late 2009, followed by the 2010 Flash Crash and the August 2011 stock markets fall across the world, furthering severe volatility of stock market indexes until now.

Using the ML estimates given in Table 3.3, Figure 3.6 plot the approximate conditional transition density functions of the stochastic differential equations given $x_0 = [s_0, Y_0]' = [6.500, 0.045]'$ for the models of H-R1 and H-R2TVTP, respectively. Since the corresponding graphs of the model H-R2-1 and H-R2-2 are very similar to those of the H-R2TVTP model, we omit them and compare only the model H-R1 and H-R2TVTP. At around the given asset price $s_0 = 6.500$ and its volatility $Y_0 = 0.045$, the conditional density function of regime L for the model H-R2TVTP is topped out with the peak of 4925.969 while that of regime H is flattened out with the top of 2316.343. Compared with those of the H-R2TVTP model, the conditional density function of the H-R1 modal is centred on a height of 2071.961. We also calculate the 95% confidence interval by delta method for each case and find that the 3D 95% confidence interval surfs form an upper layer and a lower layer just around the density functions.

To make it more clear, we split the 3D shapes into two pieces and take the cross sections at $Y_0 = 0.045$. Then Figure 3.7 plot conditional transition density functions and their 95% confidence bands for the model H-R1 and H-R2TVTP separately. The different approximate conditional transition density functions show strong evidences for the existence of high and low volatility regimes and for the time varying transition probability of the regime variable.

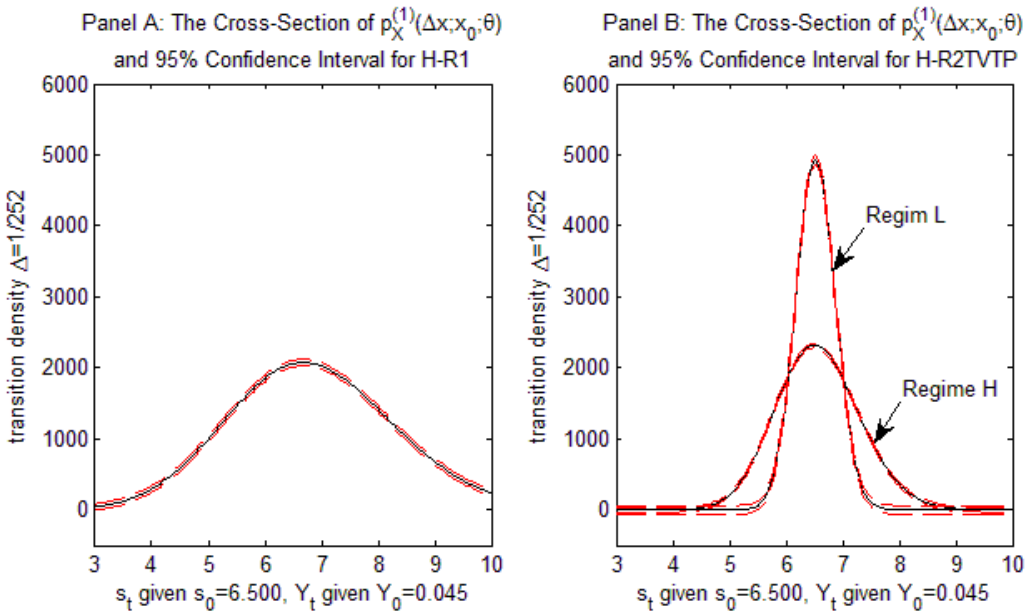
Table 3.4 reports the ML estimation results for the family of GARCH stochastic volatility models. The first column shows ML estimation outcomes for the single-regime model, G-R1. The rest three columns display the three different two-regime models G-R2-1 (two regimes, time invariant transition matrix, unconditional probability for the probability of initial state), G-R2-2 (two regimes, time invariant transition matrix, additional parameter for the probability of initial state) and G-R2TVP (two regimes, time varying transition matrix with logistic function,additional parameter for the probability of initial state) respectively. The asterisk by the parameter estimate implies that, at the 5% significance level, it is different from zero.

Figure 3.6. Conditional Transition Density Functions for the Model H-R1 and H-R2TVTP



In the term of the single regime model G-R1, the long term value of the volatility $\gamma^{1/2}$ is approximately 17.89% per year, which is smaller than that in the H-R1 model. The speed of mean reversion coefficient κ is estimated to be 2.859, much smaller than 6.817 in the H-R1 model. The leverage effect ρ is -0.799 with the variance of

Figure 3.7. Conditional Transition Density Functions and 95% Condidence Bands for the Model H-R1 and H-R2TVTP



volatility 2.319. However, the market price of risk λ and the rate of return rd are statistically insignificant.

To analyse the three different two-regime models, we find very similar conclusions as the class of Heston stochastic volatility models. $\kappa_H > \kappa_L$, $\gamma_H > \gamma_L$, $\rho_H > \rho_L$, $\sigma_H > \sigma_L$, rd_L, rd_H, λ_L and λ_H are still not statistically significant. . Although κ_L is not significant, in fact, the p-value is 0.051, which is a bit bigger than 0.05. H State is very persistent while L state is not so steady. In the model of G-R2TVTP, time varying transition probabilities of i_t are mainly affected by the variable s_t, other than the constant term and Y_t. When $d_L < 0$ ($d_H > 0$), p_{LL} (p_{HH}) and s_t have a negative (positive) relationship. The ML estimates c and e are not significant.

Comparing four log-likelihood values, the single regime model G-R1 reports the smallest value 42061.618. What we should pay attention to is that, in the Heston stochastic volatility model, the parameters have to satisfy the Feller condition $2\kappa\gamma \geq \sigma^2$, which is much more restricted than the condition $\kappa\gamma \geq 0$ in the G-R1 model. With Feller condition, κ, γ and σ can't vary so much as they do in the GARCH stochastic volatility model. We can find that in the G-R1 model, Feller condition is violated by $2 \times 2.859 \times 0.032 < 2.319^2$. Like the model of H-R2-1 and H-R2-2, the model of G-R2-1 and G-R2-2 also report very closed log-likelihood values and similar

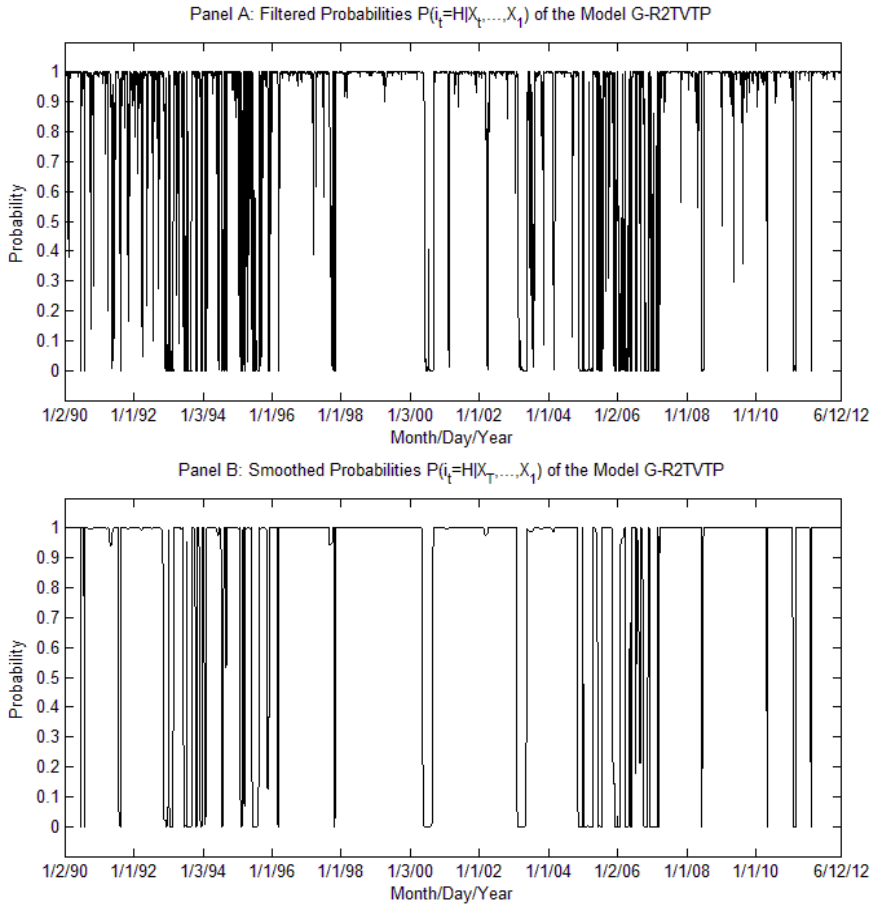
ML estimation outcomes. The log-likelihood value of G-R2TVTP is 43274.182 that is much bigger than the log-likelihood value of single regime model G-R1.

Table 3.4. Maximum Likelihood Estimation Results of Main GARCH Models(01/02/1990-06/12/2012)

Model	G-R1	G-R2-1	G-R2-2	G-R2TVTP
Regime	Single Regime	Two Regimes		
		Time Invariant Transition Matrix		Time Varying Transition Matrix
$P(R_1 = L)$	—	$\frac{1-p_{HH}}{2-p_{LL}-p_{HH}}$		p
$P(R_{t+\Delta} = i \mid R_t = i)$ with $i = L$ and H	—	p_{ii}	p_{ii}	$\frac{exp(c_i+d_i s_t+e_i Y_t)}{1+exp(c_i+d_i s_t+e_i Y_t)}$
Log-likelihood	42061.618	42957.084	43063.939	43274.182
Parameters	Maximum Likelihood Estimates(standard error)			
λ_L	3.020	2.000(7.403)	2.000(6.802)	1.999(6.271)
λ_H	(3.777)	3.999(5.212)	3.999(5.381)	3.999(3.172)
κ_L	2.859**	2.002(1.027)	2.002**(0.936)	1.999**(0.478)
κ_H	(0.657)	4.998**(1.211)	4.998**(1.278)	4.999**(0.641)
γ_L	0.032**	0.030**(0.012)	0.037**(0.014)	0.030**(0.005)
γ_H	(0.006)	0.063**(0.013)	0.081**(0.018)	0.112**(0.013)
ρ_L	−0.799**	−0.760**(0.000)	−0.756**(0.000)	−0.703**(0.000)
ρ_H	(0.000)	−0.805**(0.000)	−0.802**(0.000)	−0.801**(0.000)
σ_L	2.319**	1.786**(0.006)	1.792**(0.005)	0.999**(0.003)
σ_H	(0.001)	2.600**(0.002)	2.599**(0.003)	2.504**(0.003)
rd_L	0.0002	0.0002(0.084)	0.0002(0.074)	0.0002(0.009)
rd_H	(0.049)	0.0002(0.021)	0.0002(0.044)	0.0002(0.001)
p_{LL}	–	0.099**(0.013)	0.09**(0.009)	–
p_{HH}	–	0.899**(0.014)	0.90**(0.010)	–
c_L	–	–	–	−0.199(1.655)
c_H	–	–	–	0.169(0.220)
d_L	–	–	–	−1.459**(0.245)
d_H	–	–	–	1.239**(0.032)
e_L	–	–	–	−0.014(4.280)
e_H	–	–	–	0.010(0.197)
p	–	–	0.49(0.842)	0.500(0.535)
AIC	−14.363	−14.666	−14.702	−14.772
BIC	−14.356	−14.650	−14.685	−14.751
RCM	–	47.584	30.707	2.385

Note: Maximum likelihood estimates with standard errors in parentheses for the period 01/02/1990-06/12/2012 for the four different models with the general diffusion specification are presented in this table. The standard errors are calculated using sample mean of the outer product of the score functions evaluated at the ML estimates. According to the number of regimes, the starting probability and the transition probability matrix specifications, the four models are: G-R1 (single regime),G-R2-1 (two regimes, time invariant transition matrix,unconditional probability for the probability of initial state), G-R2-2 (two regimes, time invariant transition matrix, additioanl parameter for the probability of initial state), G-R2TVTP (two regimes,time varying transition matrix with logistic function,additional parameter for the probability of initial state).

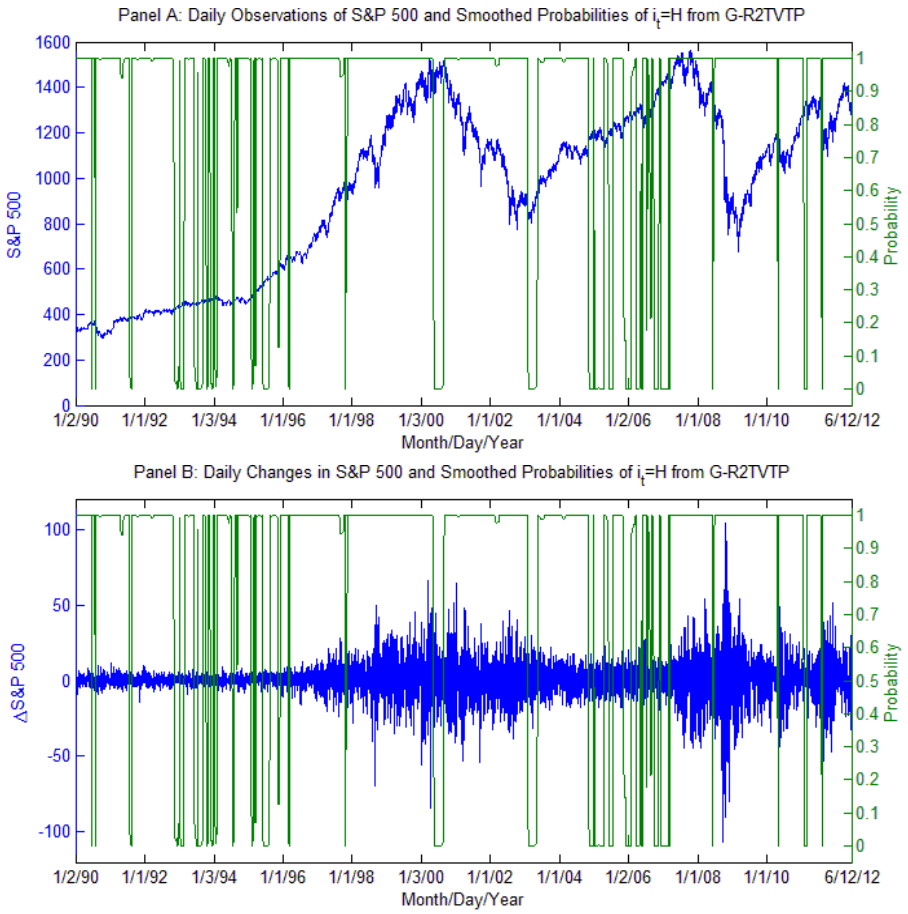
Figure 3.8. Regime-Switching Probabilities of the Model G-R2TVTP



AIC and BIC are employed to compare the single regime model with three different two-regime models. The G-R1 model reports the biggest AIC -14.363 and BIC -14.356 while the G-R2TVTP model shows the smallest AIC -14.772 and BIC -14.751. The models G-R2-1 and G-R2-2 display the similar AIC and BIC values. In the term of RCM, we can find strong advantage of the two regimes model with time varying transition matrix.

The filtered probabilities and smoothed probabilities of staying in H state over the sample for the model of G-R2TVTP are displayed in Figure 3.8. Then Figures 3.9 and 3.10 plot the time series of smoothed probabilities with S&P 500, VIX and their first difference separately. Although the smoothed probabilities of G-R2TVTP are not so smooth as that of the model H-R2TVTP, it still identifies the main high volatility periods of 1990-1991, 1994-1995, 1998-2000, 2008-2010 and 2010-2012. Using the ML estimates provided in Table 3.4, Figure 3.12 plots the approximate conditional transition density functions of the stochastic differential equations given $x_0 = [s_0, Y_0]' = [6.500, 0.045]'$ for the model of G-R1 and G-R2TVTP, respectively.

Figure 3.9. Regime-Switching Probabilities of the Model G-R2TVTP with S&P 500 and its First Difference



At around the point x_0, the conditional density function of regime L for the model H-R2TVTP has a peak of 1318.050 while that of regime H is with the top of 641.312 and the model G-R1 is centred on a height of 716.815.

In order to show 95% confidence bands, Figure 3.11 plot the cross sections at $Y_0 = 0.045$. These 95% confidence interval surfs are much closed to the conditional density function for each pecification.

In Table 3.5, we display the estimation results for the model of C-R1, C-R2-1, C-R2-2 and C-R2TVTP respectively. The instantaneous standard deviation of S&P 500 $\gamma^{1/2}$ in the model C-R1 is 23.45% per year. The rate of mean reversion coefficient κ is estimated to be 6.049 with the correlation coefficient $\rho = -0.809$. The variance of volatility σ is close to 0.994, which is between 0.726 from the model G-R1 and 2.319 from the model H-R1. Of particular interest for the CEV model is the exponent β, which is estimated at 0.54, above the Heston value of 0.5 but below the GARCH value of 1. The market price of risk λ and the rate of return rd are statistically insignificant.

Figure 3.10. Regime-Switching Probabilities of the Model G-R2TVTP with VIX and its First Difference

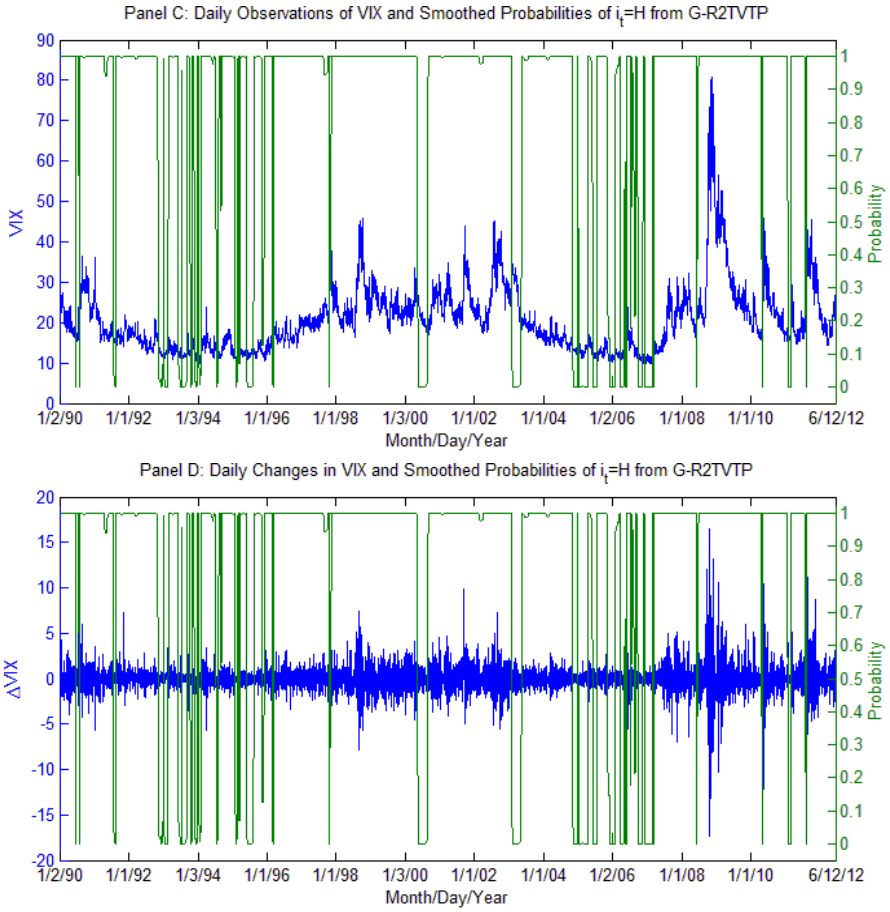


Figure 3.11. Conditional Transition Density Functions and 95% Condidence Bands for the Model G-R1 and G-R2TVTP

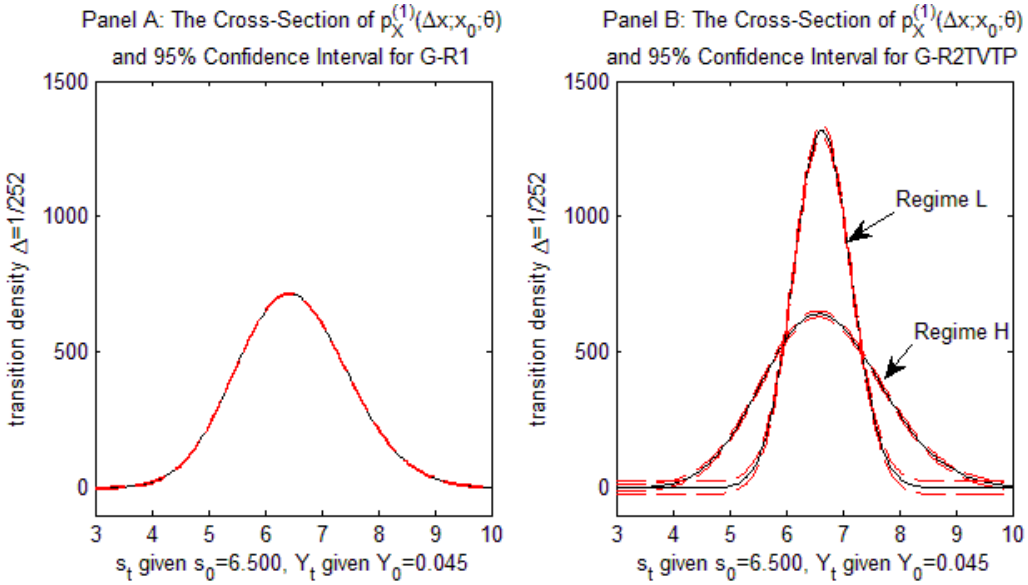
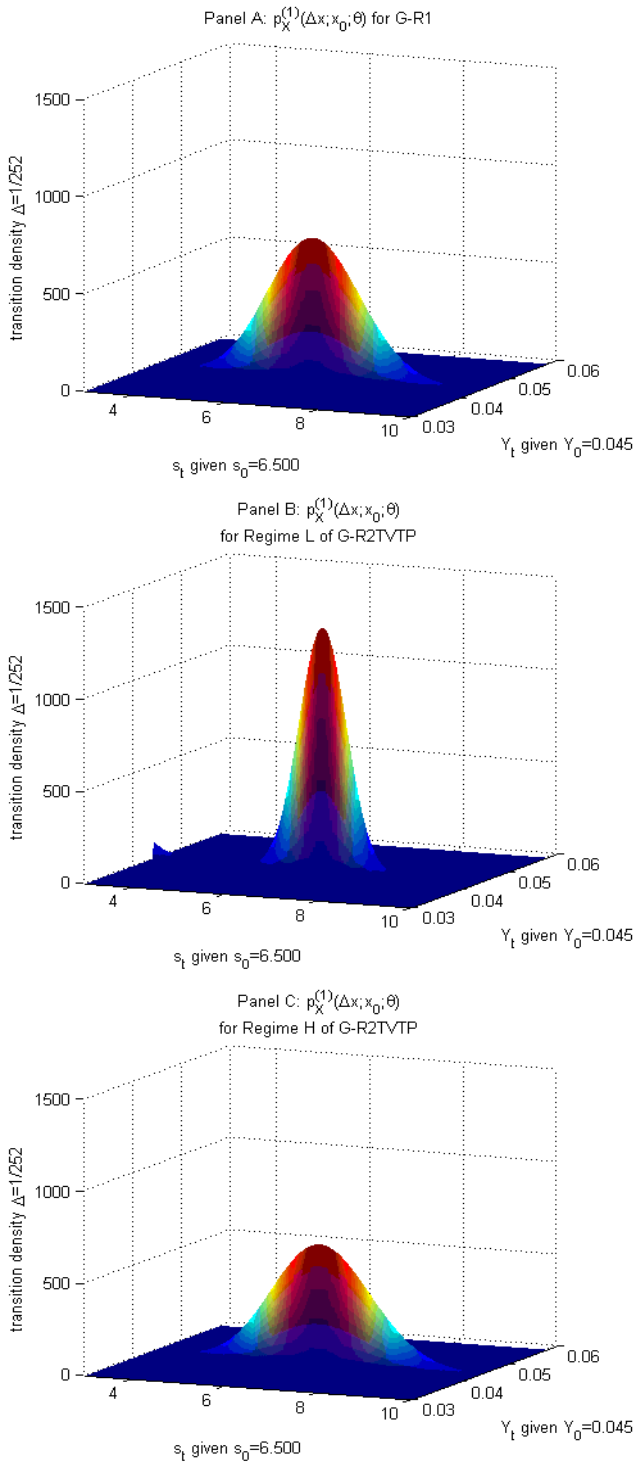


Figure 3.12. Conditional Transition Density Functions for the Model G-R1 and G-R2TVTP



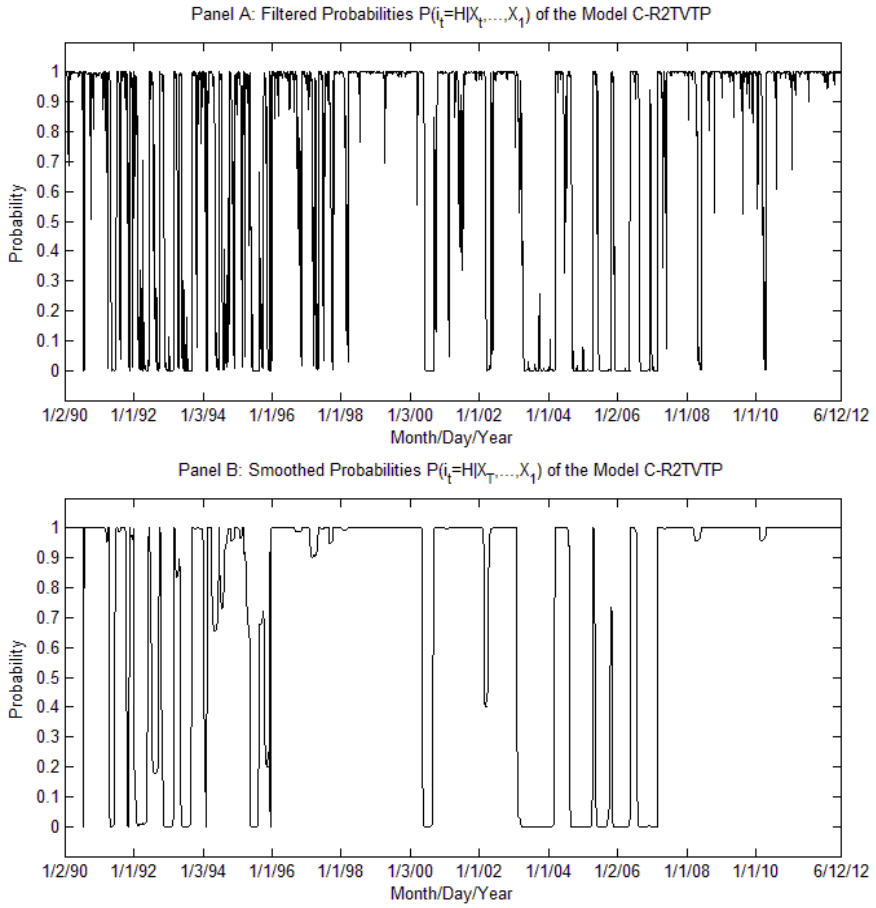
Comparing three different regime-switching models, we find similar results as the Heston and GARCH stochastic volatility models. $\kappa_H > \kappa_L$, $\gamma_H > \gamma_L$, $\rho_H > \rho_L$, $\sigma_H > \sigma_L$, however, the exponent $\beta_H < \beta_L$. λ_L and rd_L are statistically significant at the 5% level, but λ_H and rd_H are still insignificant.

Table 3.5. Maximum Likelihood Estimation Results of Main CEV Models(01/02/1990-06/12/2012)

Model	C-R1	C-R2-1	C-R2-2	C-R2TVTP
Regime	Single Regime	Two Regimes		
		Time Invariant Transition Matrix		Time Varying Transition Matrix
$P(R_1 = L)$	—	$\frac{1-p_{HH}}{2-p_{LL}-p_{HH}}$		p
$P(R_{t+\Delta} = i \mid R_t = i)$ with $i = L$ and H	—	p_{ii}	p_{ii}	$\frac{exp(c_i+d_i s_t+e_i Y_t)}{1+exp(c_i+d_i s_t+e_i Y_t)}$
Log-likelihood	42397.267	43096.667	43197.402	43734.808
Parameters	Maximum Likelihood Estimates(standard error)			
λ_L	6.069	3.002(8.543)	3.002(8.632)	2.000(6.632)
λ_H	(5.018)	4.995(5.750)	4.995(5.532)	5.999**(4.145)
κ_L	6.049**	2.197**(0.962)	2.177**(0.976)	3.000**(0.853)
κ_H	(1.295)	7.027**(1.377)	7.025**(1.272)	5.000**(1.284)
γ_L	0.055**	0.020**(0.006)	0.020**(0.006)	0.020**(0.004)
γ_H	(0.011)	0.045**(0.007)	0.045**(0.008)	0.067**(0.016)
ρ_L	−0.809**	−0.705**(0.004)	−0.705**(0.004)	−0.701**(0.002)
ρ_H	(0.001)	−0.804**(0.002)	−0.804**(0.002)	−0.826**(0.001)
σ_L	0.994**	0.599**(0.014)	0.599**(0.014)	0.614**(0.009)
σ_H	(0.004)	1.115**(0.006)	1.117**(0.007)	1.320**(0.007)
rd_L	0.0002	0.0002(0.091)	0.0002(0.091)	0.0002(0.017)
rd_H	(0.082)	0.0002(0.077)	0.0002(0.084)	0.0002**(0.000)
β_L	0.54	0.744**(0.004)	0.744**(0.004)	0.748**(0.002)
β_H	(0.004)	0.645**(0.002)	0.645**(0.002)	0.645**(0.001)
p_{LL}	–	0.079**(0.007)	0.079**(0.007)	–
p_{HH}	–	0.900**(0.008)	0.900**(0.008)	–
c_L	–	–	–	−0.189(2.434)
c_H	–	–	–	0.159(0.411)
d_L	–	–	–	−1.437**(0.366)
d_H	–	–	–	1.217**(0.063)
e_L	–	–	–	−0.013(7.350)
e_H	–	–	–	0.009(1.299)
p	–	–	0.501(0.798)	0.500(0.728)
AIC	−14.477	−14.713	−14.747	−14.929
BIC	−14.469	−14.695	−14.728	−14.905
RCM	–	34.583	33.113	7.781

Note: Maximum likelihood estimates with standard errors in parentheses for the period 01/02/1990-06/12/2012 for the four different models with the general diffusion specification are presented in this table. The standard errors are calculated using sample mean of the outer product of the score functions evaluated at the ML estimates. According to the number of regimes, the starting probability and the transition probability matrix specifications, the four models are: C-R1, C-R2-1, C-R2-2, C-R2TVTP

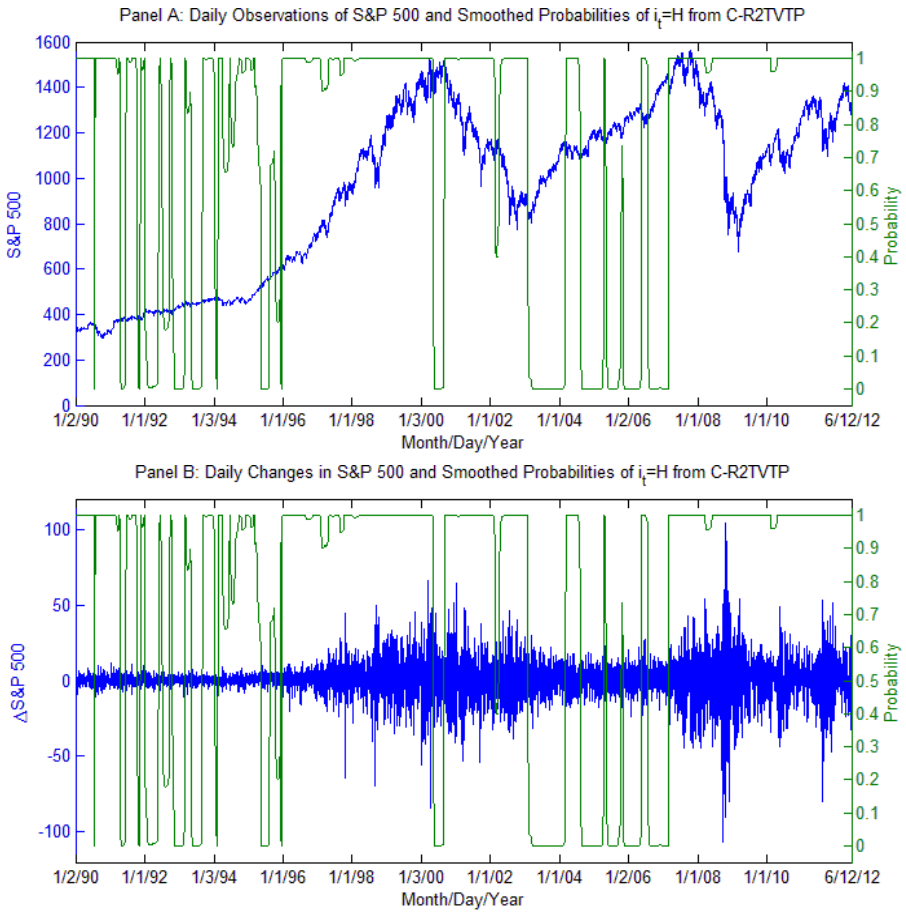
Figure 3.13. Regime-Switching Probabilities of the Model C-R2TVTP



Considering the log-likelihood values from the four different models, the single regime model C-R1 exhibits the smallest value 42397.267, which is bigger than 42061.618 from the model G-R1 and 42061.618 of the model H-R1. The main reason is still Feller condition $2\kappa\gamma \geq \sigma^2$. The models C-R2-1 and C-R2-2 show log-likelihood values of 43096.667 and 43197.402, respectively. The largest log-likelihood value 43734.808 belongs to C-R2TVTP with time varying transition matrix.When we move from the single regime model C-R1 to C-R2-1, C-R2-2 and C-R2TVTP, the AIC values decreases gradually from -14.477 to -14.929 and BIC declines from -14.469 to -14.905.The models of C-R2-1 and C-R2-2 report RCM 34.583 and 33.113 while the C-R2TVTP model shows RCM 7.781 that is much smaller than the previous two models. It is definitely a strong evidence for the existence of two regimes with varying transition matrix.

To analyse the probability of staying in state H, we draw the time series of filtered probabilities and smoothed probabilities over the sample for the model C-R2TVTP in 3.13.

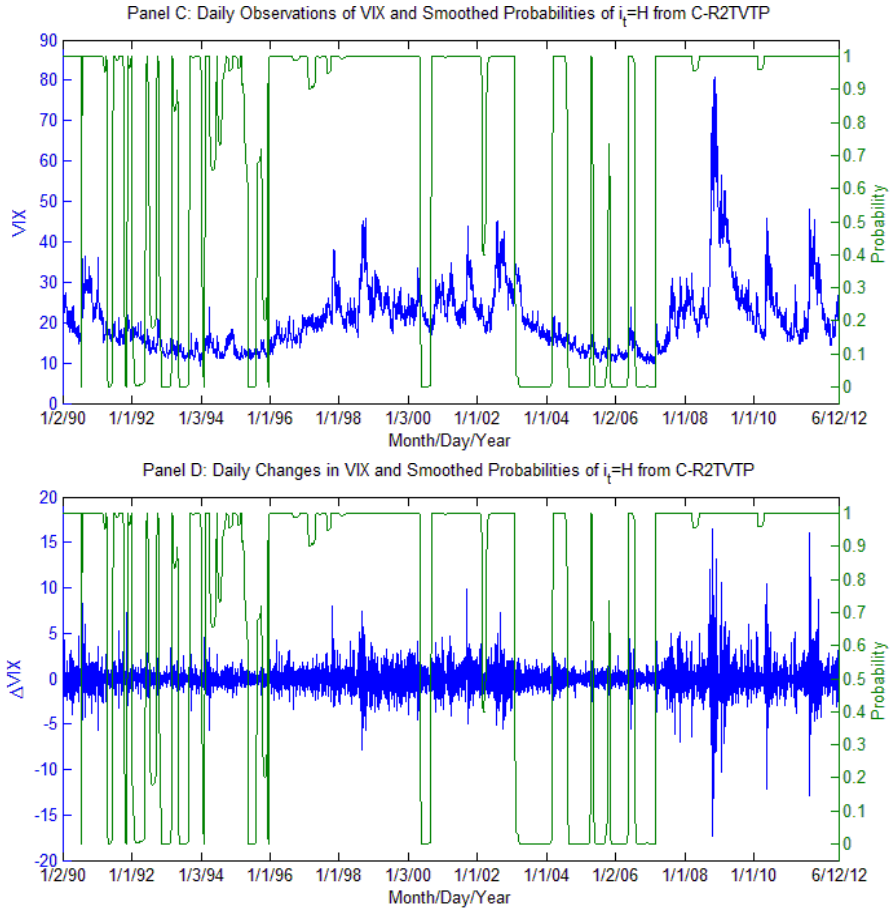
Figure 3.14. Regime-Switching Probabilities of the Model C-R2TVTP with S&P 500 and its First Difference



Then Figures 3.14 and 3.15 plot the time series of smoothed probabilities with S&P 500, VIX and their first difference separately.

The smoothed probabilities are very smooth and therefore we can figure out most high volatility periods more clearly than the models of H-R2TVTP and G-R2TVTP. The 1990-1991 high regime state connects with the Gulf War, the 1992 period reflects the Black Wednesday, the high volatility regime around 1994 is linked to he 1994 Northridge earthquake in the Los Angeles area. The successive 1996-2003 H state is associated with the budget crisis in U.S., Asian financial crisis, Russian financial crisis, long-term capital management collapse, the dot-com bubble, September 11, 2001 terrorist attacks, as well as the October 2001 invasion in Afghanistan and 2003 Iraq War. The 2004 and 2005 high volatility regime is coincident with the 2004 Atlantic hurricane season and 2005 Hurricane Katrina and Hurricane Rita, respectively. The H state regime around 2007 matches the Chinese Correction plunge. The final long high volatility period of 2008-2012 is associated with the subprime mortgage crisis, Lehman

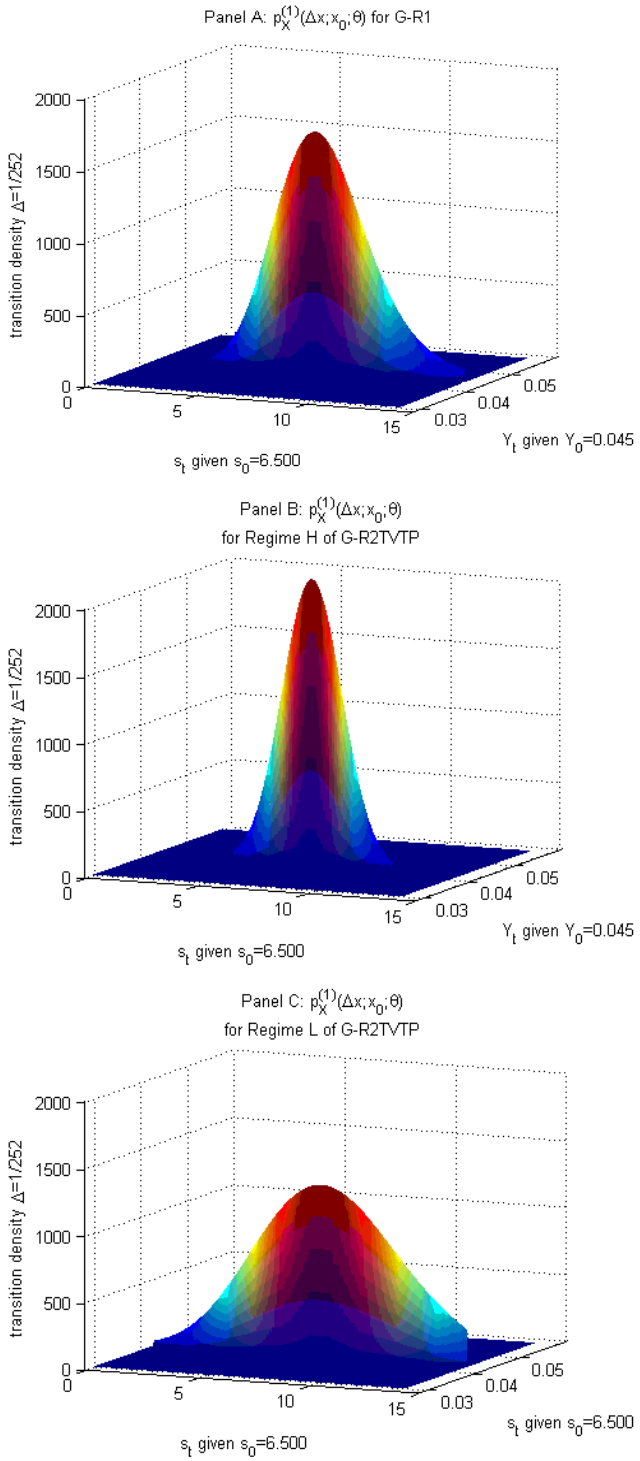
Figure 3.15. Regime-Switching Probabilities of the Model C-R2TVTP with VIX and its First Difference



Brothers Bankruptcy protection, the European sovereign-debt crisis, the 2010 Flash Crash and the August 2011 stock markets fall, which is still going on until now.

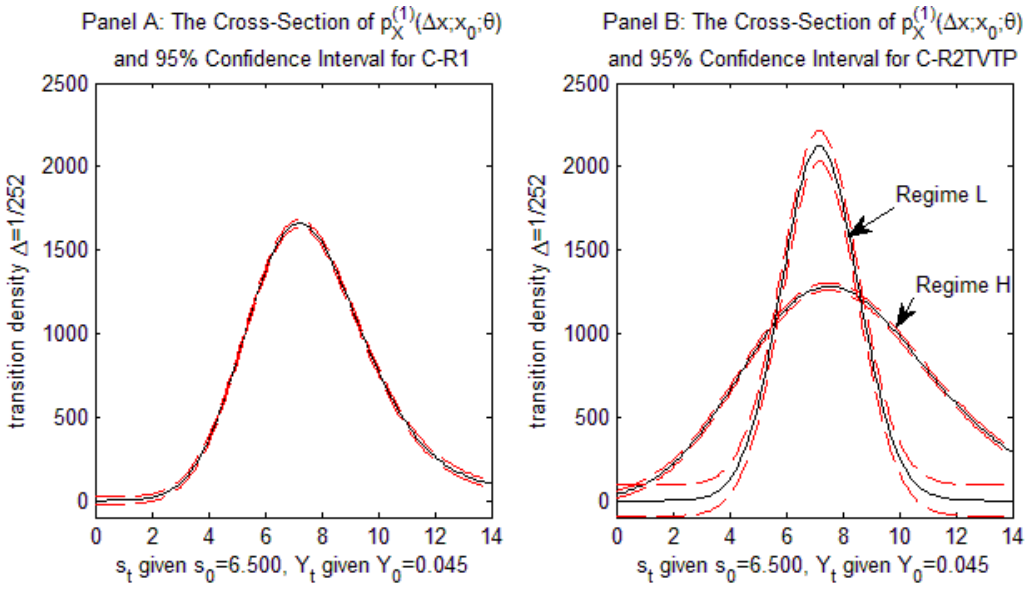
Using the ML estimates provided in Table 3.5, given $x_0 = [s_0, Y_0]' = [6.500, 0.045]'$, the approximate conditional transition density functions of the stochastic differential equations for the C-R1 and C-R2TVTP models are plotted respectively in Figure 3.16. At around x_0 point, the conditional density function of regime L for the model C-R2TVTP peaks at the height of 2122.339, that of regime H is flattened out with the top of 1280.951 and the model C-R1 is centred on a height of 1658.754. Compared to the H-R2TVTP and G-R2TVTP models, the conditional transition density functions of the model C-R2TVTP for the two regimes are more close to each other as well as that of the single regime model.The cross sections at $Y_0 = 0.045$ of conditional transition density functions and their 95% confidence bands for the C-R1 model and the C-R2TVTP model are showed in Figure 3.17.

Figure 3.16. Conditional Transition Density Functions for the Model C-R1 and C-R2TVTP



In order to compare three families of continuous time stochastic volatility models, we calculate likelihood ratio statistics for the nested models as Aït-Sahalia and Kimmel (2007) do. The H-R1 and G-R1 models are rejected in favour of the model C-R1 by reporting LRT statistic of 330.800 and 671.298, respectively. This result

Figure 3.17. Conditional Transition Density Functions and 95% Condidence Bands for the Model C-R1 and C-R2TVTP



is cohering with Aït-Sahalia and Kimmel (2007). The LRT statistic for the model H-R1-1 against the model C-R1-1 is 5.282 and the corresponding p-value is 0.071. Hence, the H-R1-1 model can't be rejected. However, the statistic for the model G-R1-1 is 279.168 and cannot reject the hypothesis of the model C-R1-1. Considering the H-R1-2 and the G-R1-2 models, the LRT statistics against the C-R1-2 model are 206.752 and 266.926 respectively. So the C-R1-2 model is rejected. Moving to the regime-switching models with time varying transition matrix, the H-R2TVTP and G-R2TVTP models are rejected in favour of the model C-R2TVTP by reporting LRT statistic of 833.028 and 921.251, respectively. To conclude, the CEV model outperforms the Heston and GARCH models not only for the single regime models but also for the two-regime models.

3.5. Conclusions

We combine the regime shift with three stochastic volatility models, including the Heston model, the GARCH model and the CEV model. According to the number of regimes, the initial probability and the transition probability matrix specifications, we estimate four models for each group. For Heston model, we compare H-R1 (single-regime), H-R2-1 (two regimes, time-constant transition matrix, unconditional probability for the probability of the initial state), H-R2-2 (two regimes, time-constant transition matrix, additional parameter for the probability of the initial state) and

H-R2TVTP (two regimes, time-varying transition matrix with a logistic function, additional parameter for the probability of the initial state). For the GARCH model and the CEV model, We estimate G-R1, G-R2-1, G-R2-2, G-R2TVTP and C-R1, C-R2-1, C-R2-2, C-R2TVTP, separately. What's more, all parameters in these models are allowed to vary depending on the state of the economy. The maximum likelihood estimation, which is developed in Chapter 2, is applied to estimate these models. Using S&P 500 and VIX data for the stock price and volatility proxy, respectively, we report estimators for each model, as well as AIC, BIC, RCM, LR statistics to compare different models. Furthermore, we investigate the regime-switching probabilities of the regime-switcing models with time varying transition matrix and additional parameter for the probability of initial state. We also analyse the corresponding approximate conditional transition density functions with their 95% confidence intervals in order to find the evidence of regime shift. Our estimation results show four main findings. First, the regime-switching models are significantly different from the single regime models. Second, there are strong evidences for the existence of the high and low volatility regimes, for the time varying transition probability of the regime variable, and for high persistence of the high regime. Third, the time varying transition probability mainly depends on the stock market index S&P 500 rather than its volatility. Fourth, the regime-switching CEV model with time varying transition matrix and additional parameter for the probability of initial state performs better than other regime-switching models.

CHAPTER 4

Accuracy of Euler Approximation for Continuous Time Diffusion Models

4.1. Introduction

To estimate continuous time diffusion models, researchers have proposed a variety of econometric methods. These methods include simulation method of Duffie and Singleton (1993), maximum likelihood estimation (MLE) from Pearson and Sun (1994) and Durham and Gallant (2002), generalized method of moment (GMM) by Hansen and Scheinkman (1995), nonparametric method of Aït-Sahalia (1996a) and Aït-Sahalia (1996b), efficient method of moment (EMM) by Gallant and Tauchen (1998) and Bayesian method from Elerian, Chib, and Shephard (2001). The maximum likelihood estimation (MLE) method has not been used much because the transition density functions are unknown in closed form for most continuous time diffusion models. Since Sargan (1974), several methods have been proposed to approximate the transition probability densities. These include the Kolmogorov partial differential equation of Lo (1988), Monte Carlo simulation from Pedersen (1995), binomial or other trees by Jensen and Poulsen (2002) and approximate closed form likelihood expansions of time homogeneous univariate of Aït-Sahalia (2002) and multivariate models of Aït-Sahalia (2008).

However, to approximate a diffusion process, the Euler approximation is still the simplest and most commonly used approaches in recent literatures (Christensen, Kinnebrock, and Podolskij (2010), Kristensen (2010), Song (2011), Chen and Hong (2011), Zhao (2011), Koo and Linton (2012) and Chen and Song (2013)). Theoretically, the accuracy of the Euler approximation depends on the time discretization step Δ. To be more precise, as Δ goes to 0, the Euler approximation is expected to be closer to the true diffusion process.

In a practical simulation, researchers set different time discretization step Δ for the Euler approximation in order to achieve two main objectives. First, when the

objective is to simulate a diffusion process or test statistical estimators, a good pathwise approximation is required. For example, Christensen, Kinnebrock, and Podolskij (2010) first generate a complete high-frequency record of N (N denoting the number of seconds in 6.5 hours) equidistant observations of the efficient price by using the Euler approximation with $\Delta = 1/23400$. Second, when the objective is to compute moments, probabilities and other functions of the diffusion process, a good endpoint approximation is required. For example, Song (2011) uses the Euler approximation to generate data from univariate and multivariate continuous time diffusion models by setting $\Delta = 1/(252 \times 1500)$ for daily frequency, $\Delta = 1/(12 \times 1500)$ for monthly frequency, $\Delta = 1/(4 \times 1500)$ for quarterly frequency and $\Delta = 1/(1 \times 1500)$ for yearly frequency respectively, and then uses these data to analyse the impact of the numerical integral approximation. Chen and Hong (2011) set $\Delta = 1/(12 \times 120)$ at monthly frequency to examine the sizes of specification tests under multivariate continuous time diffusion models. Kristensen (2010) chooses $\Delta = 1/(252 \times 10)$ at daily frequency for the evaluation of the likelihood. Zhao (2011) approximates the conditional log-likelihood by using $\Delta = 1/252$. Koo and Linton (2012) set $\Delta = 1/101$ to investigate the robustness of estimators under locally stationary diffusion processes. Chen and Song (2013) use subintervals of $\Delta = 1/(52 \times 5)$ for weekly frequency to approximate Lévesque integral. Aït-Sahalia and Kimmel (2007) and Choi (2013) generate data at daily frequency by setting $\Delta = 1/(252 \times 30)$ to carry out MLE for continuous time diffusion models. To our best knowledge, so far, there has been no literature to study whether these above time intervals are effective enough for the Euler scheme to provide accurate approximation. Theoretically, the smaller the discretization interval is, the more accurate the Euler approximation is expected to be. However, even when the discretization interval is too small, the accuracy of the Euler approximation can get worse because of the roundoff error (Peter and Platen (1992)) and random number generator bias (Komori, Saito, and Mitsui (1994)). The random number bias primarily results from lack of mutual independence in the samples from a random number generator when Δ is quite small.

The purpose of this chapter is to provide a guideline of effective time steps Δ for the Euler approximation to different diffusion processes over a finite interval $[0, T]$. In order to find out an effective time step Δ, the quality of a discrete time approximation

is usually judged according to two basic criterions, including convergence (or consistency) and stability (Peter and Platen (1992)). Convergence concerns the accuracy of an Euler approximation over a finite interval $[0, T]$ for small time discretization steps Δ, while stability investigates the quality of an approximation in a long term, $T \to \infty$. There have been some literatures studying the stability of different discrete time approximations (Saito and Mitsui (1996), Higham (2000a), Higham (2000b), Higham (2001) and Buckwar and Sickenberger (2011)). Since the purpose of this chapter is to find out a small enough time discretization step Δ over a finite interval $[0, T]$, our numerical tests examine convergence rather than stability.

According to the objective of a practical simulation, two criterions could be applied to assess the convergence of an Euler approximation. When the objective is to simulate a diffusion process or test statistical estimators, a good pathwise approximation is required. To measure how good a pathwise approximation is, the first step of our numerical tests is to exam the Euler approximation by calculating the percentage errors at the end of time interval $[0, T]$ for different time step Δ. However, with only a few exceptions, the true diffusion process are unknown for many diffusion processes. Without the true diffusion process, a relatively more accurate diffusion process is needed to be a benchmark for judging the accuracy of the Euler approximation. The benchmark we choose is the Milstein approximation, which is a second-order approximation proposed by Milstein (1975). Comparing with the Euler approximation, the Milstein approximation is a combination of a normal distribution and a chi square distribution. This could be one reason why researchers prefer to use the Euler approximation. Based on the benchmark, the second step of our numerical tests is to calculate the absolute error by generating 100 paths for different time step Δ and then illustrate the strong convergence. When the objective is to compute moments, probabilities and other functions of the diffusion process, a good endpoint approximation is required. Following previous literatures (Peter and Platen (1992) and Higham (2001)), the last step of our numerical tests is to calculate the mean errors by generating 100 paths for different time step Δ and then display the weak convergence. In fact, we generated 100, 1000, 10000 and 100000 paths in order to compare the corresponding results. Then we found these results are very similar. This is the reason why we generate 100 paths to analyse both of the strong convergence and weak convergence. All the numerical tests about univariate diffusion models are

done in MATLAB R2011b on a computer with 32-bit operating system and all the numerical tests about multivariate diffusion models are done in MATLAB R2011b on a computer with 64-bit operating system. On one hand, our numerical tests suggest an appropriate time interval should be set in order to satisfy different error requirements for the case of simulating a diffusion process or testing statistical estimators. On the other hand, the turning point should be considered when computing moments, probabilities or other functions of the diffusion process.

Section 2 reviews the Euler and Milstein approximations and introduces the criterions to investigate the accuracy of the Euler approximation. Section 3 analyses the effective time discretization step Δ for univariate stochastic diffusion models, including Black-Schole-Merton model, Vasicek model, CIR model, Inverse of Feller's Square Root model, Linear Drift CEV model and Nonlinear Mean Reversion model. Section 4 moves to multivariate stochastic diffusion models, such as Heston, GARCH, CEV, Stein, Scott, Hull-White, Hagan and SABR process. Section 5 concludes this chapter.

4.2. Euler and Milstein Approximations

Given an diffusion process $\{X_t, 0 \leq t \leq T\}$ solution of the stochastic differential equation

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

with initial deterministic value $X_{t_0} = X_0$ and the discretization $\Pi_N = \Pi_N([0,T])$ of the interval $[0,T]$, $0 = t_0 < t_1 < \cdots < t_N = T$, the Euler approximation of X is a continuous stochastic process Y satisfying the iterative scheme

$$\begin{aligned} Y_{i+1} &= Y_i + \mu(t_i, Y_i)(t_{i+1} - t_i) + \sigma(t_i, Y_i)(W_{i+1} - W_i) \\ &= Y_i + \mu \cdot \Delta t + \sigma \cdot \Delta W_i \end{aligned}$$

for $i = 0, 1, \ldots, N-1$, with $Y_0 = X_0$, $\Delta t = t_{i+1} - t_i$ and $\Delta W_i = W_{i+1} - W_i$. In order to increase the accuracy of the approximation, Milstein (1975) proposed the Milstein approximation, which adds the second-order term by using Itô lemma. Denoting by $\sigma_x(t,x)$ the partial derivative of $\sigma(t,x)$ with respect to x, the Milstein approximation becomes

$$\begin{aligned} Z_{i+1} &= Z_i + \mu(t_i, Z_i)(t_{i+1} - t_i) + \sigma(t_i, Z_i)(W_{i+1} - W_i) \\ &\quad + \frac{1}{2}\sigma(t_i, Z_i)\,\sigma_x(t_i, Z_i)\left[(W_{i+1} - W_i)^2 - (t_{i+1} - t_i)\right], \end{aligned} \tag{4.1}$$

where $i = 0, 1, \ldots, N-1$, with $Z_0 = X_0$. We could simplify 4.1 as

$$Z_{i+1} = Z_i + \mu \cdot \Delta t + \sigma \cdot \Delta W_i + \frac{1}{2}\sigma\sigma_x \left[(\Delta W_i)^2 - \Delta t \right],$$

where $\Delta t = t_{i+1} - t_i$ and $\Delta W_i = W_{i+1} - W_i$.

For a N-dimensional stochastic differential equation

$$dX_t = M(t, X_t)\, dt + \Sigma(t, X_t)\, dW_t, \tag{4.2}$$

where $M(t, X_t)$ is a $n \times 1$ vector, $\Sigma(t, X_t)$ is a $n \times n$ matrix and W is a n-dimensional uncorrelated Brownian motion. The m-th component of the multidimensional Euler scheme for equation 4.2 is

$$y_{i+1} = y_i + M(t_i, Y_i)(t_{i+1} - t_i) + \Sigma(t_i, Y_i)(W_{i+1} - W_i)$$

for $i = 0, 1, \ldots, N-1$, with $Y_0 = X_0$. The m-th component of the multidimensional Milstein scheme is

$$z_{i+1} = z_i + M_m \cdot \Delta t + \sum_{j=1}^{M} \Sigma_{ij} \cdot \Delta W_j + \sum_{j=1}^{M}\sum_{k=1}^{M}\sum_{l=1}^{M} \Sigma_{jk}\left(\partial_{x_j}\Sigma_{ml}\right) I_{(k,l)},$$

wherein all of the coefficient functions M_m and Σ_{mj}, etc., are to be evaluated with Z_t. The double Itô integral $I_{(k,l)}$ is defined as

$$I_{(k,l)} = \int_{s=t}^{t+\Delta t} \int_{u=t}^{s} dW_k(u)\, dW_l(s). \tag{4.3}$$

For $k = l$, the double integrals 4.3 simplifies to

$$I_{(k,k)} = \frac{1}{2}\left(\Delta W_k^2 - \Delta t\right).$$

For $k \neq l$, Abe (2004) presents five methods to simulate the double integrals

$$I_{(1,2)} = \int_{t_n}^{t_{n+1}} \int_{t_n}^{s_1} dW_{1,s_2} dW_{2,s_1}$$

$$I_{(2,1)} = \int_{t_n}^{t_{n+1}} \int_{t_n}^{s_1} dW_{2,s_2} dW_{1,s_1}$$

for the 2D Milstein approximation, and shows the Real Variance formulae is the best approximation of the double integral among the five methods. Therefore, we applies

the Real Variance formulae to calculate $I_{(1,2)}$ and $I_{(2,1)}$.

$$\begin{aligned} I_{(2,1)} &= \frac{1}{2}\left(I_{(1,2)} + I_{(2,1)}\right) - \frac{1}{2}\left(I_{(1,2)} - I_{(2,1)}\right) \\ &= \frac{1}{2}\left(dW_1 dW_2\right) - \frac{1}{2}\left(L\grave{e}vy\ Area\right) \\ &= \frac{1}{2}dtZ_1Z_2 - \frac{1}{2}\left(\sqrt{\frac{dt^2}{3}}Z_1Z_2 + \sqrt{\frac{dt^2r^2}{3}}Z_2Z_3\right) \end{aligned}$$

and

$$I_{(1,2)} = dtZ_1Z_2 - I_{(2,1)},$$

where $r^2 = Z_1^2 + Z_2^2$, Z_1 and Z_2 are the original random numbers of the Brownian motion and Z_3 is a random number from the standard normal distribution.

To find out an effective time discretization step Δ, the quality of a time discrete approximation is usually judged according to two basic standards, including convergence (or consistency) and stability (Peter and Platen (1992)). Convergence concerns the accuracy of an approximation over a finite interval $[0, T]$ for small time discretization steps Δ, while stability investigates the quality of an approximation in a long term, $t \to \infty$. There have been some literatures studying the stability of different discrete time approximations (Saito and Mitsui (1996), Higham (2000a), Higham (2000b), Higham (2001) and Buckwar and Sickenberger (2011)). Since the purpose of this chapter is to find out a small enough time step Δ over a finite interval $[0, T]$, the numerical tests examine convergence rather than stability.

According to the objective of a practical simulation, two criteria could be applied to assess the convergence of an Euler or Milstein approximation. When the objective is to simulate a diffusion process or test statistical estimators, a good pathwise approximation is required. To measure how good the pathwise approximation is, this chapter examines the absolute error criterion, which is defined as

$$\epsilon = E\left(\left|X_T - Y\left(T\right)\right|\right),$$

the expectation of the absolute value of the difference between the diffusion process X and the approximation Y at the end of time interval $[0, T]$. The approximation Y with a time step Δ converges strongly to X at time T if

$$\lim_{\Delta \to 0} E\left(\left|X_T - Y^{\Delta}\left(T\right)\right|\right) = 0.$$

The approximation Y is said to converge strongly with order γ, if there exists a constant C subject to

$$E\left(\left|X_T - Y\left(T\right)\right|\right) \leq C\Delta^{\gamma}$$

for any fixed $\Delta \in (0, T)$. In a practical simulation, literatures (Kristensen (2010), Song (2011), Chen and Hong (2011), Zhao (2011), Koo and Linton (2012) and Chen and Song (2013)) usually generate a discretized Brownian path and then get an Euler or a Milstein scheme to approximate a specific diffusion process rather than to generate many schemes and then take average of them to approximate such diffusion process. For this reason, the first step of our numerical tests is to exam an Euler approximation and a Milstein approximation separately by calculating the percentage error

$$PE = \left|Y\left(T\right) - X_T\right| / X_T \times 100\%$$

at the end of time interval $[0, T]$ for different time step Δ, if the true diffusion process or a reference solution could be generated. Then moving to the next step, the absolute error could be computed by generating 100 different Euler or Milstein approximation for different time step Δ and then the strong convergence could be illustrated.

Besides simulating a good pathwise approximation, another important objective is to compute moments, probabilities and other functions of the diffusion process. In this case, the approximation of the first moment of a particular diffusion process X is investigated by considering the mean error

$$\mu = \left|E\left(Y\left(T\right)\right) - E\left(X_T\right)\right|$$

at a finite terminal time T. The approximation Y with a time step Δ converges weakly to X at time T if

$$\lim_{\Delta \to 0} \left|E\left(g\left(X_T\right)\right) - E\left(g\left(Y^{\Delta}\left(T\right)\right)\right)\right| = 0.$$

The functions g are continuously differentiable and have polynomial growth. To simplify our calculation, this chapter assumes $g(X) = X$ as Higham (2001) did. In a practical simulation, the mean error equals to the absolute error when a single Euler or Milstein approximation is generated. The approximation Y is said to converge strongly with order γ, if there exists a constant C subject to

$$\left|E\left(Y\left(T\right)\right) - E\left(X_T\right)\right| \leq C\Delta^{\gamma}$$

for any fixed $\Delta \in (0, T)$. Our numerical tests calculate the mean error of 100 different Euler or Milstein approximation for different time step Δ and plot them to display the weak convergence. All the numerical tests about univariate diffusion models are done in MATLAB R2011b on a computer with 32-bit Operating System and all the numerical tests about multivariate diffusion models are done in MATLAB R2011b on a computer with 64-bit Operating System.

4.3. Univariate Stochastic Volatility Models

4.3.1. The Black-Scholes-Merton(BSM) or Geometric Brownian Motion Model

This process is introduced by Black and Scholes (1973) and Merton (1973) to model asset prices. It is the solution to the stochastic differential equation

$$dX_t = \theta_1 X_t dt + \theta_2 X_t dW_t$$

with initial deterministic value $X_0 = x_0$ and $\theta_2 > 0$. W_t is a standard Brownian motion. The parameters θ_1 and θ_2 are constants, representing the interest rate and the volatility of risky activities respectively. The explicit solution (Lacus (2008)) to this equation is

$$X_t = X_0 e^{\left(\theta_1 - \frac{1}{2}\theta_2^2\right)t + \theta_2 W_t}.$$

The Euler approximation of the process X is a continuous stochastic process Y satisfying the iteration

$$Y_{t+1} = Y_t + \theta_1 Y_t \cdot \Delta t + \theta_2 Y_t \cdot \Delta W_t$$

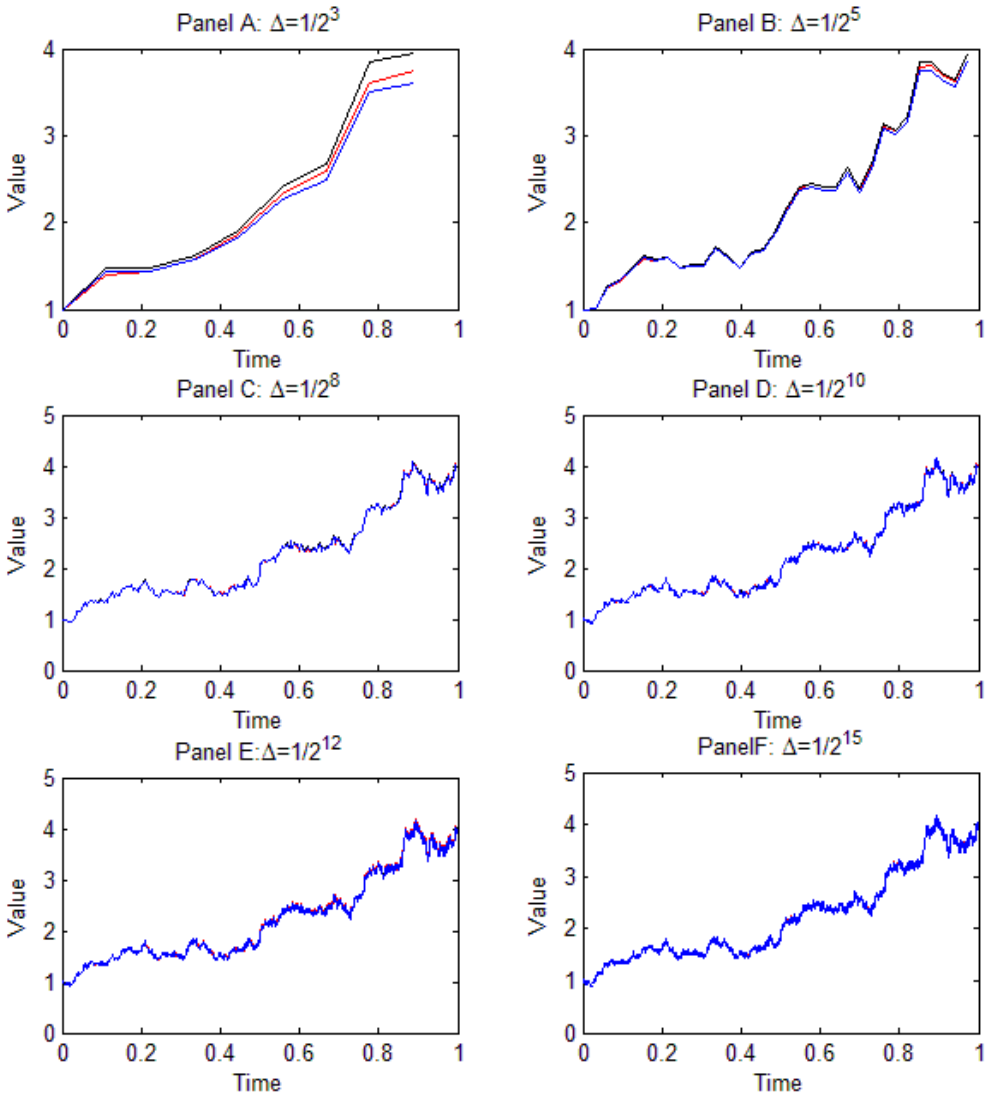
with initial value $Y_0 = X_0$ for $t = 0, 1, \ldots N-1$. The Milstein approximation of X is a continuous stochastic process Z satisfying the iterative scheme

$$Z_{t+1} = Z_t + \theta_1 Z_t \cdot \Delta t + \theta_2 Z_t \cdot \Delta W_t + \frac{1}{2}\theta_2^2 Z_t \left(\Delta W_t^2 - \Delta t\right)$$

with original deterministic value $Z_0 = X_0$ for $t = 0, 1, \ldots, N-1$.

A BSM model with $\theta_1 = 1$, $\theta_2 = 1/2$ and $X_0 = 1$ is taken as an example for the following numerical tests. A Brownian bridge from 0 to 1 is used as W_t. To check whether the Euler and Milstein schemes could provide a good pathwise approximation, we plot the Euler approximation Y and the Milstein approximation Z with the explicit solution X for different time intervals Δ.

Figure 4.1. The Euler and Milstein Approximations against the Explicit Solution to the BSM Model



Panel A of Figure 4.1 shows that the Euler approximation approaches the explicit solution more closely than the Milstein approximation when $\Delta = 1/2^3$. Panel B and Panel C display that the Euler and Milstein schemes are crossing and overlapping each other around the true solution when $\Delta = 1/2^5$ and $\Delta = 1/2^8$. The other panels exhibit the two approximations almost completely copy the true solution when Δ decreases to $1/2^{10}$, $1/2^{12}$ and $1/2^{15}$. Moving from Panel A to Panel F, it shows that the Euler and Milstein approximations for a finer time discretization would be closer to the explicit solution. To be more precise, we display maximum absolute

error between the explict solution and its approximations in Table 4.1 for different discretization steps Δ separately.

Table 4.1. Maximum Absolute Error between the Exact Solution and Approximations to the BSM Model

Δ	Euler	Milstein	Δ	Euler	Milstein
$1/2^1$	0.8950	1.0006	$1/2^{11}$	0.0213	0.0017
$1/2^2$	0.3946	0.6167	$1/2^{12}$	0.0336	0.0009
$1/2^3$	0.2509	0.3455	$1/2^{13}$	0.0020	0.0004
$1/2^4$	0.0354	0.1858	$1/2^{14}$	0.0097	0.0002
$1/2^5$	0.0524	0.0951	$1/2^{15}$	0.0045	0.0001
$1/2^6$	0.0222	0.0550	$1/2^{16}$	0.0037	0.0001
$1/2^7$	0.0196	0.0292	$1/2^{17}$	0.0014	0.0000
$1/2^8$	0.0205	0.0148	$1/2^{18}$	0.0011	0.0000
$1/2^9$	0.0398	0.0074	$1/2^{19}$	0.0004	0.0000
$1/2^{10}$	0.0126	0.0035	$1/2^{20}$	0.0004	0.0000

In Table 4.1, maximum absolute error of the Euler approximation varies for different Δ, whereas maximum absolute error of the Milstein approximation keeps decreasing as Δ decreases. Compared with the Milstein approximation, the Euler approximation reports smaller maximum absolute error when $\Delta \geq 1/2^7$. However, the Milstein approximation shows smaller maximum absolute error than the Euler approximation when $\Delta < 1/2^7$.

With respect to the endpoint, Tables 4.2 and 4.3 examine the absolute errors and percentage errors between the explicit solution and its approximations for different Δ. It seems that the Euler scheme has smaller absolute errors when $\Delta \geq 1/2^8$ while the Milstein scheme has smaller absolute errors when Δ decreases to $1/2^9$ and even smaller. This could be another point of view to explain why previous literatures prefer to use the Euler scheme rather than the Milstein scheme to approximate the true diffusion process when Δ is simply set at data collected frequency, such as daily ($\Delta = 1/252$), weekly ($\Delta = 1/52$), monthly ($\Delta = 1/12$), quarterly ($\Delta = 1/4$), and yearly ($\Delta = 1$).

Figure 4.2 illustrates the endpoint values of the two approximations with the explicit solution for different step sizes Δ. To put it another way, it displays the speed of convergence of the Euler and Milstein schemes to the true value as a function of the discretization step $\Delta = 1/N$. Undoubtedly, the Milstein approximation converges

Table 4.2. Absolute Error between the Exact Solution and Approximations for the BSM Model at the Endpoint

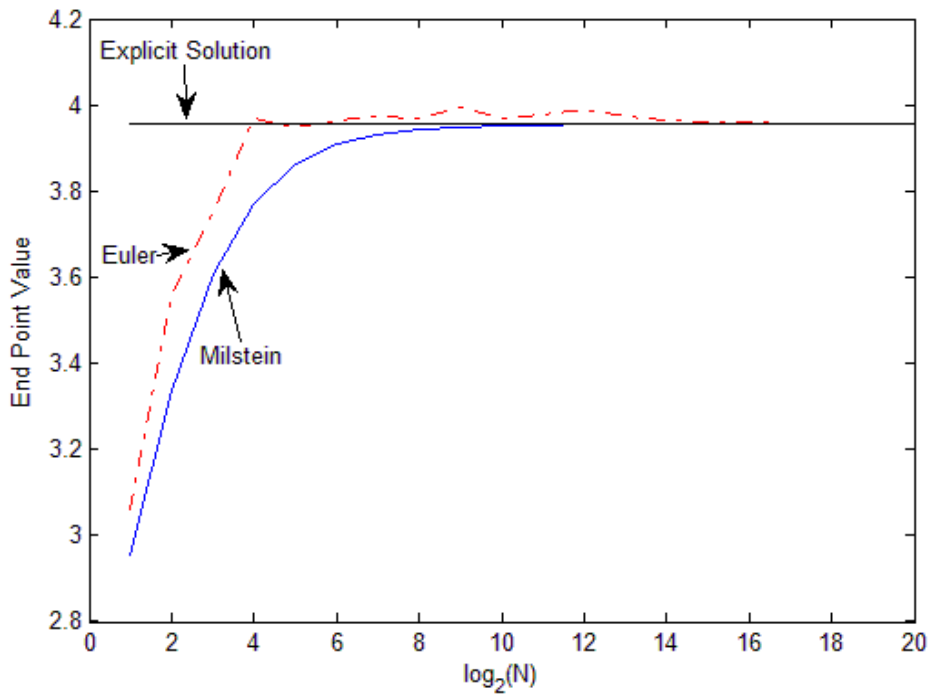
Δ	Euler	Milstein	Δ	Euler	Milstein
$1/2^1$	0.8950	1.0006	$1/2^{11}$	0.0209	0.0014
$1/2^2$	0.3946	0.6167	$1/2^{12}$	0.0330	0.0008
$1/2^3$	0.2031	0.3448	$1/2^{13}$	0.0190	0.0003
$1/2^4$	0.0129	0.1845	$1/2^{14}$	0.0094	0.0002
$1/2^5$	0.0096	0.0944	$1/2^{15}$	0.0041	0.0001
$1/2^6$	0.0086	0.0473	$1/2^{16}$	0.0035	0.0000
$1/2^7$	0.0186	0.0254	$1/2^{17}$	0.0010	0.0000
$1/2^8$	0.0122	0.0129	$1/2^{18}$	0.0010	0.0000
$1/2^9$	0.0395	0.0064	$1/2^{19}$	0.0001	0.0000
$1/2^{10}$	0.0126	0.0029	$1/2^{20}$	0.0001	0.0000

to the true value at a higher speed than the Euler approximation does. This outcome is very similar to Lacus (2008)'s result. It seems that the Euler approximation reaches the true value at $\Delta = 1/2^{15}$ and the Milstein approximation realizes the true value when $\Delta = 1/2^{10}$. Table 4.3 shows the percentage error could be controlled below 1.00% when $\Delta \leq 1/2^4$ for the Euler approximation and $\Delta \leq 1/2^7$ for the Milstein approximation. In comparison, the percentage error could be controlled below 0.10% when $\Delta \leq 1/2^{16}$ for the Euler approximation and $\Delta \leq 1/2^{10}$ for the Milstein approximation. Actually, starting from $\Delta = 1/2^{16}$, the percentage errors of the Milstein scheme reaches 0.00%, which means the Milstein scheme could be used as a benchmark instead of the true diffusion process.

Table 4.3. Percentage Errors of the Euler and Milstein Approximations to the BSM Model at Endpoint

Δ	Euler	Milstein	Δ	Euler	Milstein
$1/2^1$	22.629%	25.299%	$1/2^{11}$	0.528%	0.035%
$1/2^2$	9.977%	15.593%	$1/2^{12}$	0.834%	0.020%
$1/2^3$	5.135%	8.718%	$1/2^{13}$	0.480%	0.008%
$1/2^4$	0.326%	4.665%	$1/2^{14}$	0.238%	0.005%
$1/2^5$	0.243%	2.387%	$1/2^{15}$	0.104%	0.003%
$1/2^6$	0.217%	1.196%	$1/2^{16}$	0.088%	0.000%
$1/2^7$	0.470%	0.642%	$1/2^{17}$	0.025%	0.000%
$1/2^8$	0.308%	0.326%	$1/2^{18}$	0.025%	0.000%
$1/2^9$	0.999%	0.162%	$1/2^{19}$	0.003%	0.000%
$1/2^{10}$	0.316%	0.073%	$1/2^{20}$	0.003%	0.000%

Figure 4.2. Convergence Speed of the Euler and Milstein Approximations for the BSM Model at Endpoint



To get the strong convergence of the Euler and Milstein approximations, we generate 100 different discretized Brownian paths over $[0, 1]$. For each path, the Euler and Milstein approximations are applied for 20 different stepsizes $\Delta = 1/2^N$, $N = 1, 2, \ldots, 20$. Figure 4.3 depicts the strong convergence of the Euler and Milstein approximations for different time steps. The first right point of each approximation is the mean absolute different value for $\Delta = 1/2$, the second right point is for $\Delta = 1/2^2$, the third right point is for $\Delta = 1/2^3$ etc. The Euler approximation has strong order of convergence $\gamma = 0.550$, which is close to its theoretical strong order of convergence $\gamma = 0.5$ (Peter and Platen (1992)). Compared with the Euler approximation, the Milstein approximation has a greater strong order of convergence $\gamma = 0.973$, which is very similar to its theoretical strong order of convergence $\gamma = 1$ (Peter and Platen (1992)). Figure 4.3 confirms that the Milstein approximation converges to the true values faster than the Euler approximation. In other words, the Milstein approximation could be employed as a benchmark process for the Euler approximation as Δ goes to 0.

Figure 4.3. Strong Convergence of the Euler and Milstein Approximations to the BSM Model

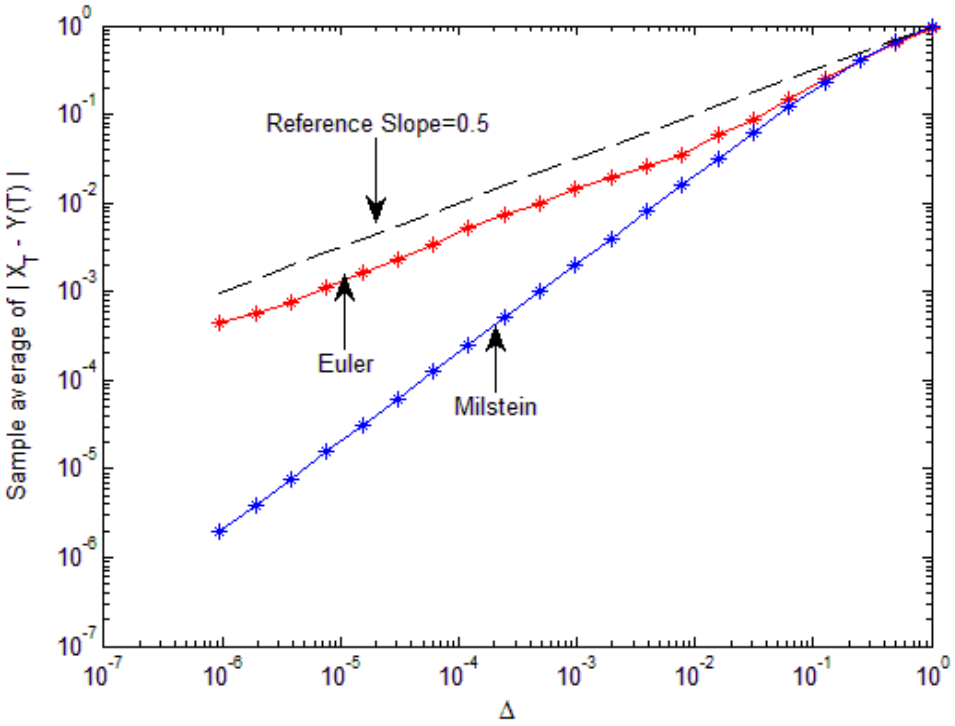
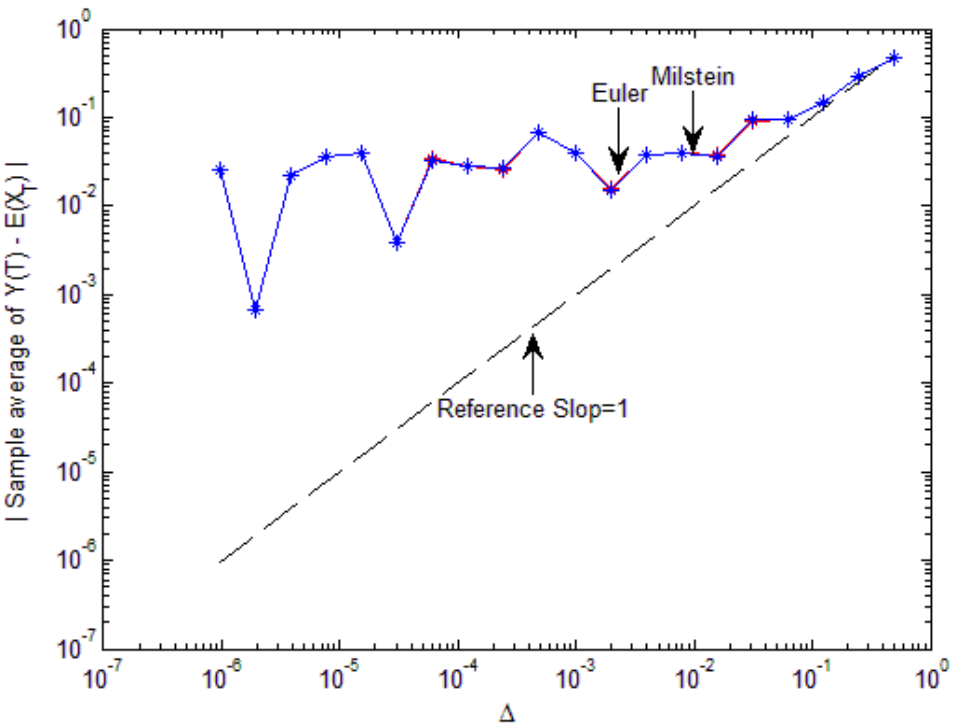


Figure 4.4. Weak Convergence of the Euler and Milstein Approximations to the BSM Model



In terms of weak convergence, we simulate 100 discretized Brownian paths over $[0, 1]$ and use 20 different discretization time steps $\Delta = 1/2^N$, $N = 1, 2, \ldots, 20$ for each path. Figure 4.4 displays that the Euler and Milstein approximations are crossing and overlapping each other. From the first right point where $\Delta = 1/2$ to the ninth right point where $\Delta = 1/2^9$, the error of the means for the two approximations keep declining. Furthermore, the slope of the two approximations are close to their theoretical weak order of convergence $\gamma = 1$ (Peter and Platen (1992)). However, starting from the tenth point where $\Delta = 1/2^{10}$, the error of the means for each approximation starts to increase and keeps varying around 0.035 when Δ gets more smaller. What is even worse, the weak order of convergence decreases to $\gamma = 0.250$, which is a quite large deviation from their theoretical value $\gamma = 1$. In fact, Komori, Saito, and Mitsui (1994) find that, besides roundoff error, this large deviation of weak convergence rate exists in a broader class of SDEs because of random number bias. This bias primarily results from lack of mutual independence in the samples from a random number generator when Δ is quite small. To solve this problem, a very time-consuming "sieving" process proposed by Komori, Saito, and Mitsui (1994) should be carried out at every single time step. Therefore, the presence of weak convergence is a strong evidence to prove that finer discretization time step does not always upgrade the Euler and Milstein approximations. Based on above analysis, for the Black-Scholes-Merton Model, our numerical tests suggest that $\Delta = 1/2^9$ is the turning point for the Euler and Milstein approximations. It seems that, in terms of weak convergence, the accuracy of the Euler and Milstein approximations couldn't be improved, even be lower after the turning point.

To sum up, if a good pathwise approximation is required, the accuracy of the Euler approximation seems to be improved as Δ goes to 0. However, if a good approximation of moments is needed, $\Delta = 1/2^9$ is probably a nice option for the Euler approximation.

4.3.2. Vasicek Model

Vasicek model is proposed by vasicek (1977) to specify that the instantaneous interest rate follows the Ornstein-Uhlenbeck process

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

where initial deterministic value $X_0 = x_0$ and $\sigma > 0$. W_t is a Brownian motion. μ represents the long-run equilibrium value of the process, θ stands for the speed of reversion and σ acts as the volatility. The explicit solution (Lacus (2008)) to this equation is

$$X_t = \mu + (X_0 - \mu)\, e^{-\theta t} + \sigma \int_0^t e^{-\theta(t-\mu)} dW_u.$$

The Euler approximation of the process X is a continuous stochastic process Y satisfying the iteration

$$Y_{t+1} = Y_t + \theta\,(\mu - Y_t) \cdot \Delta t + \sigma \cdot \Delta W_t$$

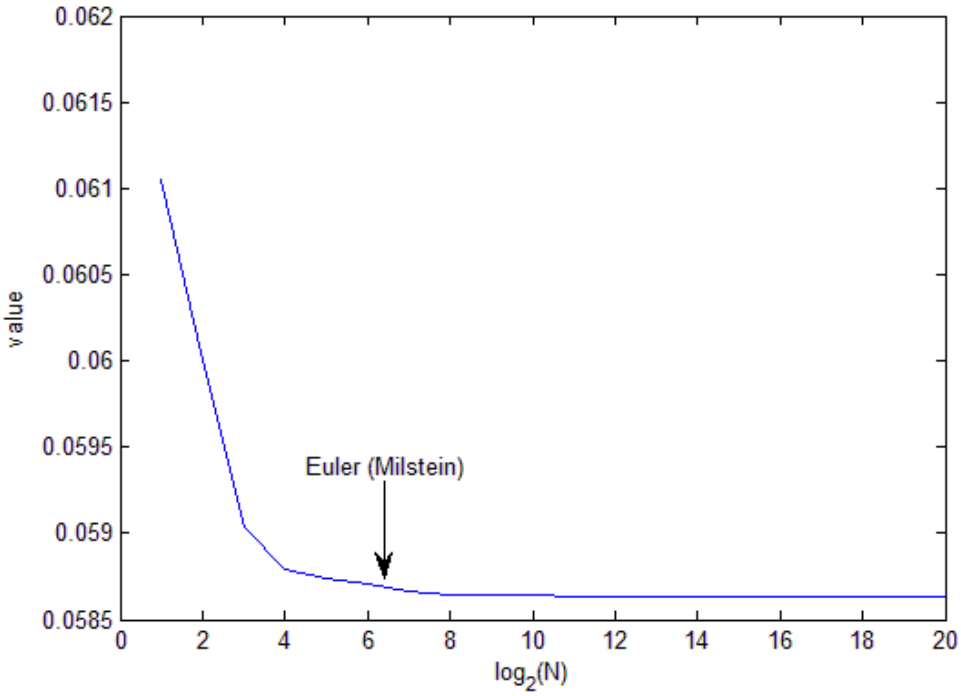
with initial value $Y_0 = X_0$ for $t = 0, 1, \ldots N-1$. The Milstein approximation of X is a continuous stochastic process Z satisfying the iterative scheme

$$\begin{aligned} Z_{t+1} &= Z_t + \theta\,(\mu - Z_t) \cdot \Delta t + \sigma \cdot \Delta W_t + \frac{1}{2}\sigma \cdot 0 \cdot \left(\Delta W_t^2 - \Delta t\right) \\ &= Z_t + \theta\,(\mu - Z_t) \cdot \Delta t + \sigma \cdot \Delta W_t \end{aligned}$$

with original deterministic value $Z_0 = X_0$ for $t = 0, 1, \ldots N-1$. Because the partial derivative of σ with respect to x is equal to 0, the Milstein approximation reduces to the Euler approximation. We take a Vasicek model with $\theta = 0.258$, $\mu = 0.0717$, $\sigma = 0.02213$ and $X_0 = 0.03$ as an example. These values are taken from Aït-Sahalia (1999). A Brownian bridge from 0 to 1 is used as W_t.

Although Itô integral could be approximated to get the true solution for different time intervals Δ, it is still a big challenge to interpret the absolute error between the explicit solution and its approximations. This is because Itô integral depends on which point in each of the small time intervals is used to compute the value of the function. Different time intervals lead to quite different results. When we measure the difference between the Euler (Milstein) approximation and the true solution for different stepsizes, it will be very hard to judge whether this difference results from the approximation method or the error from Itô integral. In the light of Figure 4.2, the Euler and Milstein approximations to the BSM model converges to the exact solution when Δ goes to 0. Accordingly, our numerical tests resort to Figure 4.5 of the convergence speed at the endpoint. It shows, from $\Delta = 1/2^9$, the Euler (Milstein) approximation converges to a stable value of 0.058634, which could be used as a reference solution. In Table 4.4, the Euler (Milstein) approximation of the endpoint

Figure 4.5. Convergence Speed of the Euler (Milstein) Approximation for the Vasicek Model at Endpoint



with the percentage errors are calculated. When the time stepsize Δ is set at $1/2^3$, the percentage error of the Euler (Milstein) approximation could be below 1.00%. When the time stepsize Δ is set at $1/2^7$, the percentage error of the Euler (Milstein) approximation could be below 0.10%.

Table 4.4. Euler (Milstein) Approximation and Percentage Errors to the Vasicek Model at Endpoint

Δ	Euler (Milstein)	PE	Δ	Euler (Milstein)	PE
$1/2^1$	0.061049	4.119%	$1/2^{11}$	0.058636	0.003%
$1/2^2$	0.060001	2.331%	$1/2^{12}$	0.058635	0.002%
$1/2^3$	0.059036	0.686%	$1/2^{13}$	0.058634	0.000%
$1/2^4$	0.058787	0.261%	$1/2^{14}$	0.058634	0.000%
$1/2^5$	0.058743	0.186%	$1/2^{15}$	0.058634	0.000%
$1/2^6$	0.058706	0.123%	$1/2^{16}$	0.058634	0.000%
$1/2^7$	0.058668	0.058%	$1/2^{17}$	0.058634	0.000%
$1/2^8$	0.058643	0.015%	$1/2^{18}$	0.058634	0.000%
$1/2^9$	0.058641	0.012%	$1/2^{19}$	0.058634	0.000%
$1/2^{10}$	0.058638	0.007%	$1/2^{20}$	0.058634	0.000%

Figure 4.6. Strong Convergence of the Euler (Milstein) Approximation to the Vasicek Model

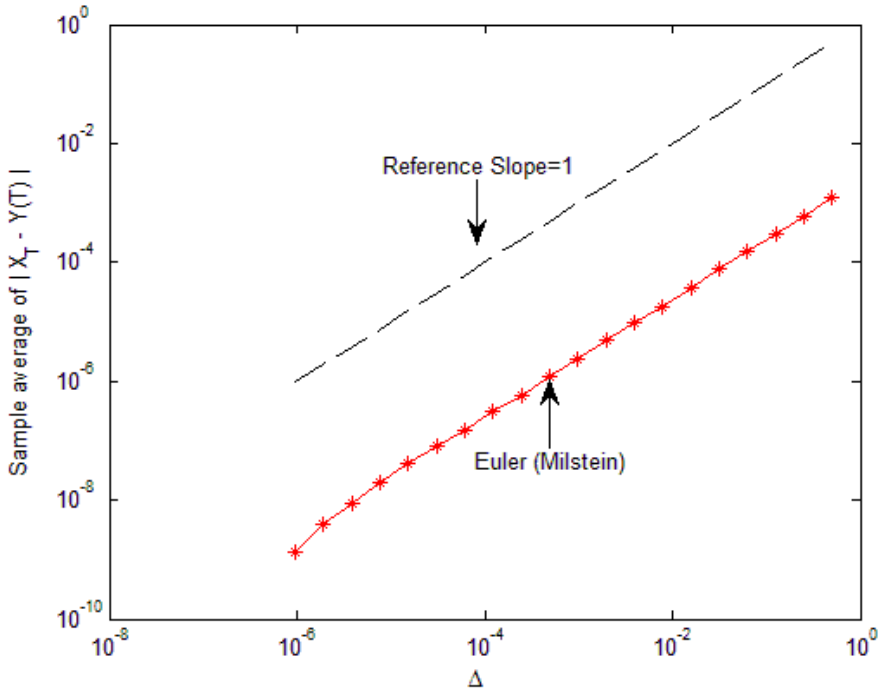
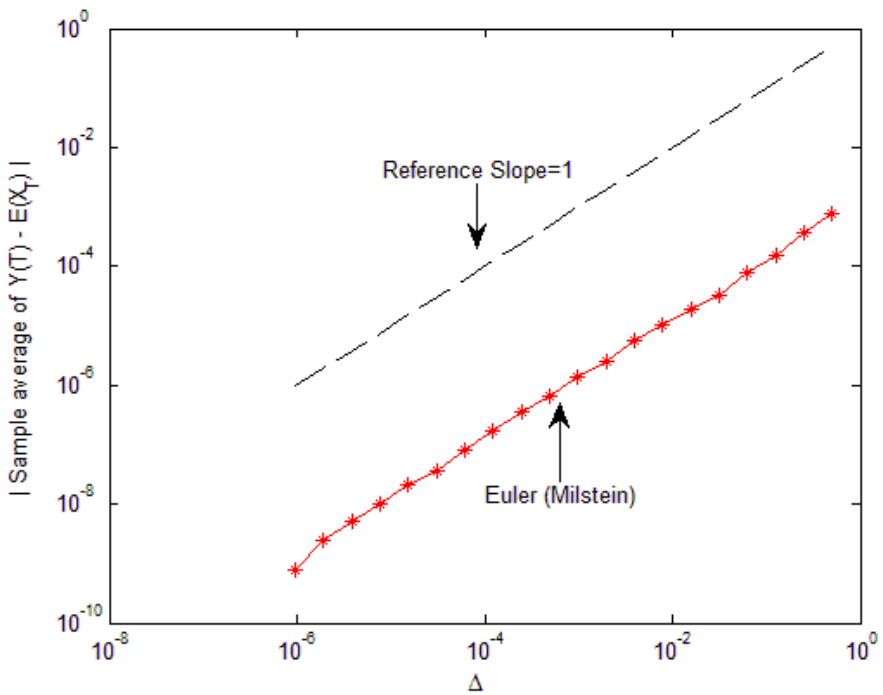


Figure 4.7. Weak Convergence of the Euler (Milstein) Approximation to the Vasicek Model



To get the strong and weak convergence of the Euler (Milstein) approximation, we generate 100 different discretized Brownian paths over $[0, 1]$. For each path, the Euler and Milstein approximations are calculated for 20 different stepsizes $\Delta = 1/2^N$, $N = 1, 2, \ldots, 20$. Figures 4.6 and 4.7 display the strong and weak convergence of the Euler (Milstein) approximation, respectively. Although the weak convergence is not so smoothed as the strong convergence, its value of 1.009 is still very close to its theoretical value $\gamma = 1$. The main reason seems that the random number bias is relieved because the volatility term σdW_t does not contains the state variable X_t. Overall, for the Vasicek model, there is no turning point for strong and weak convergence. The Euler approximation to the Vasicek process could be upgraded by shrinking Δ.

4.3.3. CIR Model

The CIR process (Cox, Ingersoll, and Ross (1985)) is the solution to the stochastic differential equation

$$dX_t = \theta(\mu - X_t)\,dt + \sigma\sqrt{X_t}dW_t,$$

where initial deterministic value $X_0 = x_0$ and $\sigma > 0$. W_t is a Brownian motion. The stochastic differential equation has the explicit solution

$$X_t = (X_0 - \mu)\,e^{-\theta t} + \sigma e^{-\theta t}\int_0^t e^{\theta\varsigma}\sqrt{X_\varsigma}dW_\varsigma.$$

The Euler approximation of the process X is a continuous stochastic process Y satisfying the iteration

$$Y_{t+1} = Y_t + \theta(\mu - Y_t)\cdot\Delta t + \sigma\sqrt{Y_t}\cdot\Delta W_t$$

with initial value $Y_0 = X_0$ for $t = 0, 1, \ldots N-1$.The Milstein approximation of X is a continuous stochastic process Z satisfying the iterative scheme

$$Z_{t+1} = Z_t + \theta(\mu - Z_t)\cdot\Delta t + \sigma\sqrt{Z_t}\cdot\Delta W_t + \frac{1}{4}\sigma^2\cdot\left(\Delta W_t^2 - \Delta t\right)$$

with original deterministic value $Z_0 = X_0$ for $t = 0, 1, \ldots N-1$. Using the parameters of Aït-Sahalia (1999), we set $\theta = 0.145$, $\mu = 0.0732$, $\sigma = 0.06521$ and $X_0 = 0.03$. A Brownian bridge from 0 to 1 is used as W_t. Since the explicit solution contains Itô integral, we could not get a stable solution for different discretization time steps

$\Delta = 1/2^N$, $N = 1, 2, \ldots, 20$. Figure 4.8 directly draws the Euler and Milstein approximations at the endpoint. It implies that the Euler approximation converges to a stable value at $\Delta = 1/2^{16}$ while the Milstein approximation reaches to the same value when $\Delta = 1/2^9$. Then Tables 4.5 and 4.6 report endpoint approximated values and their percentage errors separately. The percentage error of the Euler and Milstein approximations could be easily controlled less than 1.00% when $\Delta \leq 1/2^2$. However, the percentage error decreases to 0.10% when $\Delta \leq 1/2^{11}$ for the Euler approximation and $\Delta \leq 1/2^3$ for the Milstein approximation.

Table 4.5. Euler and Milstein Approximations to the CIR Model at Endpoint

Δ	Euler	Milstein	Δ	Euler	Milstein
$1/2^1$	0.047549	0.047036	$1/2^{11}$	0.046480	0.046493
$1/2^2$	0.046727	0.046849	$1/2^{12}$	0.046473	0.046493
$1/2^3$	0.046322	0.046507	$1/2^{13}$	0.046465	0.046493
$1/2^4$	0.046300	0.046537	$1/2^{14}$	0.046481	0.046493
$1/2^5$	0.046479	0.046500	$1/2^{15}$	0.046487	0.046493
$1/2^6$	0.046385	0.046493	$1/2^{16}$	0.046488	0.046493
$1/2^7$	0.046374	0.046500	$1/2^{17}$	0.046489	0.046493
$1/2^8$	0.046421	0.046492	$1/2^{18}$	0.046489	0.046493
$1/2^9$	0.046374	0.046492	$1/2^{19}$	0.046492	0.046493
$1/2^{10}$	0.046443	0.046493	$1/2^{20}$	0.046493	0.046493

Table 4.6. Percentage Errors of Euler and Milstein Approximations to the CIR Model at Endpoint

Δ	Euler	Milstein	Δ	Euler	Milstein
$1/2^1$	2.271%	1.168%	$1/2^{11}$	0.028%	0.000%
$1/2^2$	0.050%	0.766%	$1/2^{12}$	0.043%	0.000%
$1/2^3$	0.368%	0.030%	$1/2^{13}$	0.060%	0.000%
$1/2^4$	0.415%	0.095%	$1/2^{14}$	0.026%	0.000%
$1/2^5$	0.030%	0.015%	$1/2^{15}$	0.013%	0.000%
$1/2^6$	0.232%	0.000%	$1/2^{16}$	0.011%	0.000%
$1/2^7$	0.256%	0.015%	$1/2^{17}$	0.009%	0.000%
$1/2^8$	0.155%	0.002%	$1/2^{18}$	0.009%	0.000%
$1/2^9$	0.256%	0.002%	$1/2^{19}$	0.002%	0.000%
$1/2^{10}$	0.108%	0.000%	$1/2^{20}$	0.000%	0.000%

Figure 4.8. Convergence Speed of the Euler and Milstein Approximations to the CIR Model at Endpoint

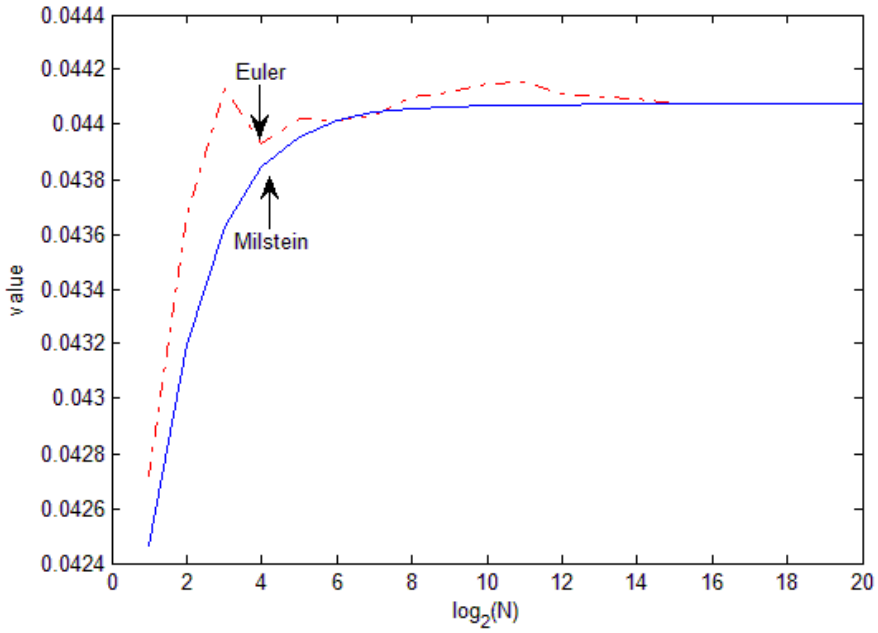
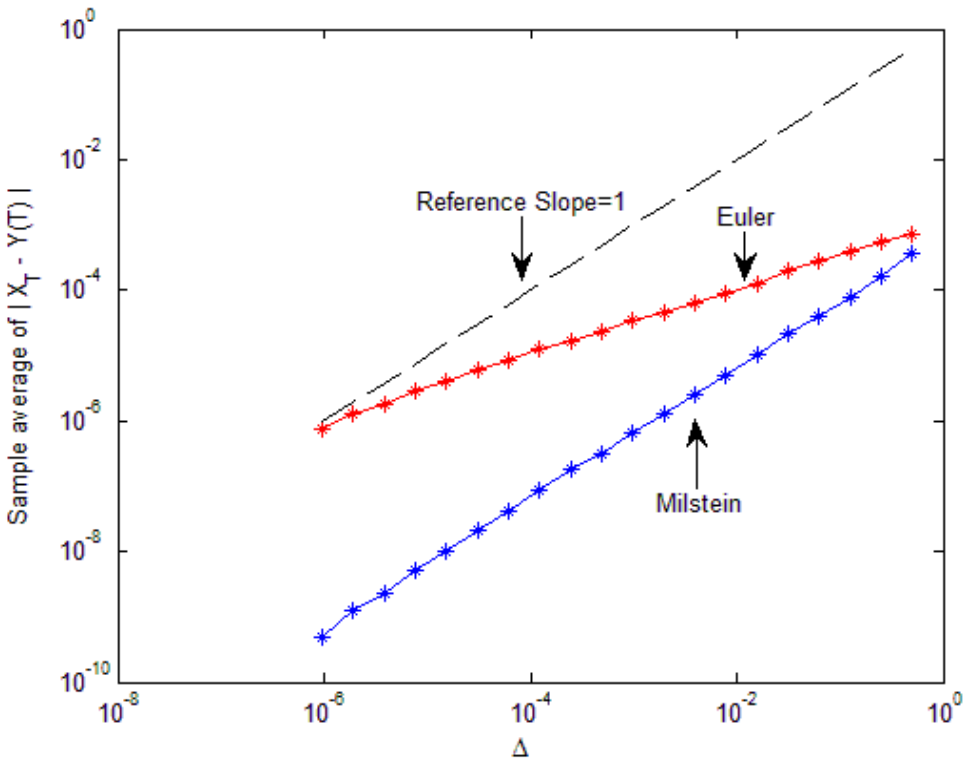
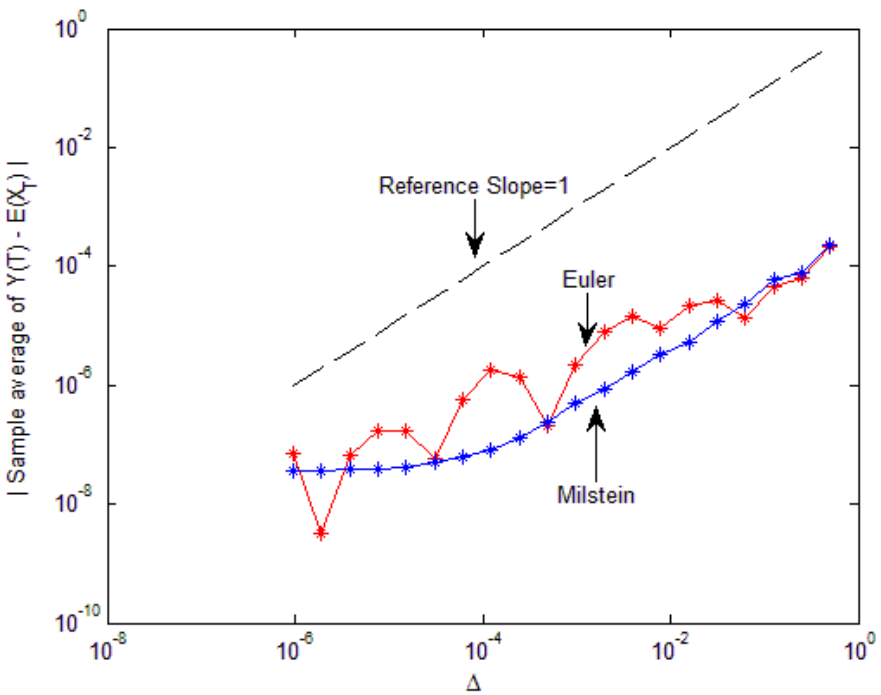


Figure 4.9. Strong Convergence of the Euler and Milstein Approximations to the CIR Model



To get the strong and weak convergence of the Euler and Milstein approximations, 100 different discretized Brownian paths over $[0, 1]$ are generated. For each path, the Euler and Milstein approximations are applied for 20 different stepsizes $\Delta = 1/2^N$, $N = 1, 2, \ldots, 20$. The strong and weak convergence of the Euler and Milstein approximations are analysed in Figures 4.9 and 4.10. Similar to the BSM Model and the Vasicek Model, the Euler and Milstein approximations converges with strong order of 0.5 and 1, respectively. Considering the weak convergence, it implies $\Delta = 1/2^{11}$ is a turning point for the Euler scheme Therefore, if a good approximation of moments is needed, $\Delta = 1/2^{11}$ is probably an effective option for the Euler approximation to the CIR model.

Figure 4.10. Weak Convergence of the Euler and Milstein Approximations to the CIR Model



4.3.4. Inverse of Feller's Square Root (IFSR) Model

This specification of the interest rate process is defined by Ahn and Gao (1999) as

$$dX_t = X_t \left(\kappa - \left(\sigma^2 - \kappa\alpha\right) X_t\right) dt + \sigma X_t^{3/2} dW_t$$

with unknown explicit solution. The Euler scheme of the process X is a continuous stochastic process Y satisfying the iterative scheme.

$$Y_{t+1} = Y_t + Y_t\left(\kappa - \left(\sigma^2 - \kappa\alpha\right) Y_t\right) \cdot \Delta t + \sigma Y_t^{3/2} \cdot \Delta W_t$$

with original deterministic value $Y_0 = X_0$ for $t = 0, 1, \ldots N-1$. The Milstein approximation of the process X is a continuous stochastic process Z satisfying the iteration

$$Z_{t+1} = Z_t + Z_t\left(\kappa - \left(\sigma^2 - \kappa\alpha\right) Z_t\right) \cdot \Delta t + \sigma Z_t^{3/2} \cdot \Delta W_t + \frac{3}{4}\sigma^2 Z_t^2\left(\Delta W_t^2 - \Delta t\right)$$

with initial value $Z_0 = X_0$ for $t = 0, 1, \ldots N-1$. We take an IFSR model with $\kappa = 0.177$, $\sigma = 0.8059$, $\alpha = 15.019$ and $X_0 = 0.03$ as an example. These values are taken from Aït-Sahalia (1999). A Brownian bridge from 0 to 1 is used as W_t. With respect to the endpoint, Figure 4.11 displays the convergence speed of the two approximations. It implies the Euler approximation reaches the reference solution at $\Delta = 1/2^{15}$ and the Milstein approximation comes to the reference solution at $\Delta = 1/2^{10}$. The percentage error could be controlled smaller than 0.10% by setting $\Delta \leq 1/2^3$ for the Euler approximation and $\Delta \leq 1/2^4$ for the Milstein approximation, based on the data reported by Tables 4.7 and 4.8.

Table 4.7. Euler and Milstein Approximations to the IFSR Model at Endpoint

Δ	Euler	Milstein	Δ	Euler	Milstein
$1/2^1$	0.042800	0.042507	$1/2^{11}$	0.044044	0.044091
$1/2^2$	0.043606	0.043251	$1/2^{12}$	0.044066	0.044092
$1/2^3$	0.043797	0.043644	$1/2^{13}$	0.044073	0.044092
$1/2^4$	0.044116	0.043857	$1/2^{14}$	0.044080	0.044092
$1/2^5$	0.043911	0.043974	$1/2^{15}$	0.044085	0.044093
$1/2^6$	0.044099	0.044033	$1/2^{16}$	0.044086	0.044093
$1/2^7$	0.044074	0.044061	$1/2^{17}$	0.044089	0.044093
$1/2^8$	0.044142	0.044077	$1/2^{18}$	0.044092	0.044093
$1/2^9$	0.044086	0.044086	$1/2^{19}$	0.044094	0.044093
$1/2^{10}$	0.044051	0.044089	$1/2^{20}$	0.044093	0.044093

To get the strong and weak convergence of the Euler (Milstein) approximation, 100 different discretized Brownian paths over $[0, 1]$ are generated. For each path, the Euler and Milstein approximations are calculated for 20 different stepsizes $\Delta = 1/2^N$, $N = 1, 2, \ldots, 20$. The Euler and Milstein approximations converges with strong order of 0.5 and 1, respectively in Figure 4.12. In comparison, the weak convergence of

Table 4.8. Percentage Errors of Euler and Milstein Approximations to the IFSR Model at Endpoint

Δ	Euler	Milstein	Δ	Euler	Milstein
$1/2^1$	0.293%	0.360%	$1/2^{11}$	0.011%	0.000%
$1/2^2$	0.110%	0.191%	$1/2^{12}$	0.006%	0.000%
$1/2^3$	0.067%	0.102%	$1/2^{13}$	0.005%	0.000%
$1/2^4$	0.005%	0.054%	$1/2^{14}$	0.003%	0.000%
$1/2^5$	0.041%	0.027%	$1/2^{15}$	0.002%	0.000%
$1/2^6$	0.001%	0.014%	$1/2^{16}$	0.002%	0.000%
$1/2^7$	0.004%	0.007%	$1/2^{17}$	0.001%	0.000%
$1/2^8$	0.011%	0.004%	$1/2^{18}$	0.000%	0.000%
$1/2^9$	0.002%	0.002%	$1/2^{19}$	0.000%	0.000%
$1/2^{10}$	0.010%	0.001%	$1/2^{20}$	0.000%	0.000%

the Euler and Milstein schemes are not stick to their theoretical values in Figure 4.13. From the first right point where $\Delta = 1/2$ to the twelfth right point where $\Delta = 1/2^{12}$, the error of the means for the Euler approximation keep declining. However, starting from the thirteenth point where $\Delta = 1/2^{13}$, the error of the means increases dramatically. As a consequence, our numerical tests suggest $\Delta = 1/2^{12}$ seems to be an effective time discretization steps for the Euler schemes.

Figure 4.11. Convergence Speed of the Euler and Milstein Approximations to the IFSR Model at Endpoint

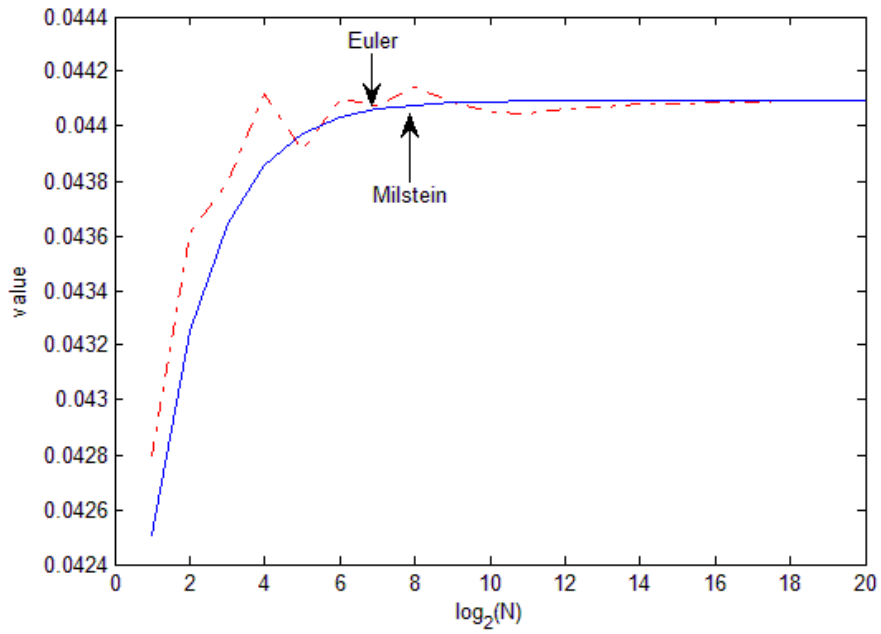


Figure 4.12. Strong Convergence of the Euler and Milstein Approximations to the IFSR Model

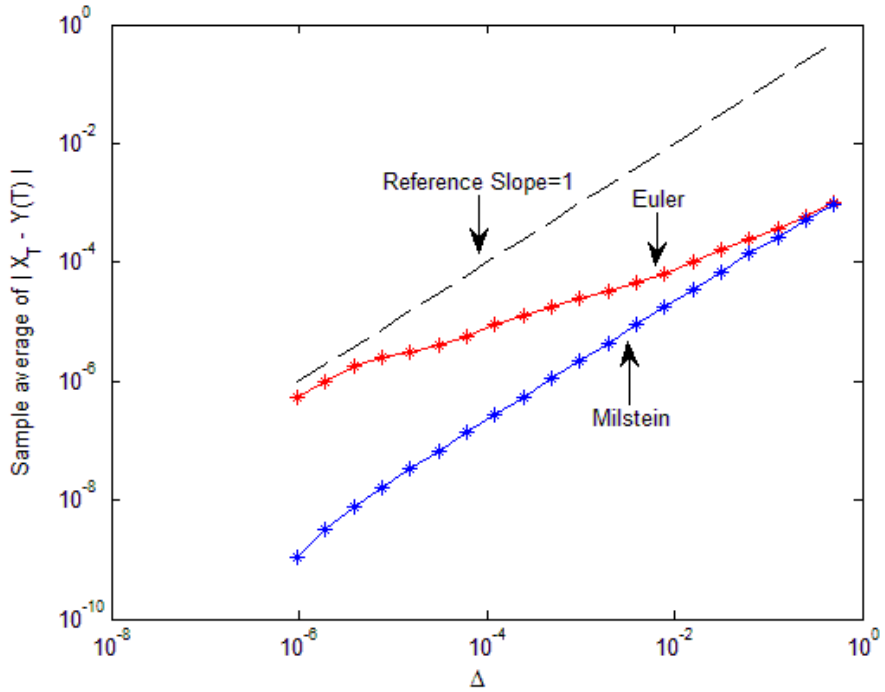
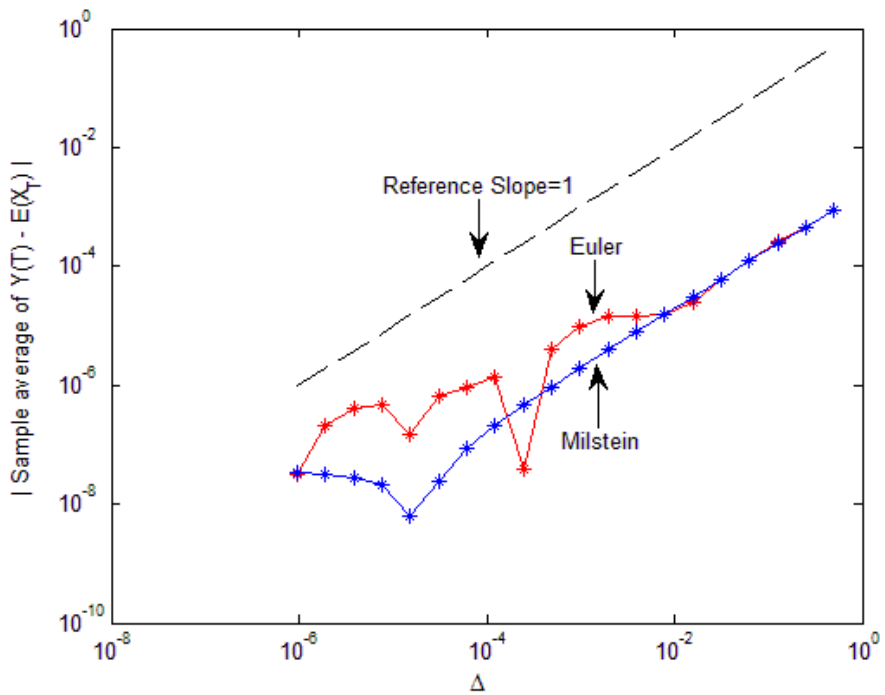


Figure 4.13. Weak Convergence of the Euler and Milstein Approximations to the IFSR Model



4.3.5. Linear Drift CEV (LDCEV) Model

This process is introduced by Chan, Karolyi, Longstaff, and Sanders (1992) to model the short term interest rate. It is the solution to the stochastic differential equation

$$dX_t = \kappa\left(\alpha - X_t\right)dt + \sigma X_t^{\rho} dW_t$$

with initial deterministic value $X_0 = x_0$. The process X is positive when $\alpha > 0$, $\kappa > 0$, $\rho > 1/2$. W_t is a Brownian motion. The Euler approximation of the process X is a continuous stochastic process Y satisfying the iteration

$$Y_{t+1} = Y_t + \kappa\left(\alpha - X_t\right)\cdot\Delta t + \sigma Y_t^{\rho}\cdot\Delta W_t$$

with initial value $Y_0 = X_0$ for $t = 0, 1, \ldots N-1$. The Milstein approximation of X is a continuous stochastic process Z satisfying the iterative scheme

$$Z_{t+1} = Z_t + \kappa\left(\alpha - Z_t\right)\cdot\Delta t + \sigma Z_t^{\rho}\cdot\Delta W_t + \frac{1}{2}\sigma^2\rho Z_t^{2\rho-1}\left(\Delta W_t^2 - \Delta t\right)$$

with original deterministic value $Z_0 = X_0$ for $t = 0, 1, \ldots N-1$. Using the parameters of Aït-Sahalia (1999), we set $\kappa = 0.0972$, $\alpha = 0.0808$, $\sigma = 0.7224$, $\rho = 1.46$ and $X_0 = 0.03$. A Brownian bridge from 0 to 1 is used as W_t. Figure 4.14 plots the Euler and Milstein approximations at the endpoint. Tables 4.9 and 4.10 report the endpoint values and percentage errors. The Euler and Milstein approximations overlaps at $\Delta = 1/2^7$ and then converge to a stable value at $\Delta = 1/2^{14}$. Both of the percentage errors are below 0.10% when $\Delta \leq 1/2^8$.

Table 4.9. Euler and Milstein Approximations to LDCEV Model at Endpoint

Δ	Euler	Milstein	Δ	Euler	Milstein
$1/2^1$	0.039063	0.038739	$1/2^{11}$	0.039032	0.039032
$1/2^2$	0.039286	0.039106	$1/2^{12}$	0.039046	0.039046
$1/2^3$	0.039323	0.039238	$1/2^{13}$	0.039042	0.039042
$1/2^4$	0.039182	0.039142	$1/2^{14}$	0.039027	0.039027
$1/2^5$	0.039071	0.039079	$1/2^{15}$	0.039027	0.039027
$1/2^6$	0.039056	0.039045	$1/2^{16}$	0.039026	0.039026
$1/2^7$	0.039108	0.039102	$1/2^{17}$	0.039026	0.039026
$1/2^8$	0.039039	0.039038	$1/2^{18}$	0.039028	0.039028
$1/2^9$	0.039059	0.039058	$1/2^{19}$	0.039027	0.039027
$1/2^{10}$	0.039055	0.039055	$1/2^{20}$	0.039027	0.039027

Figure 4.14. Convergence Speed of the Euler and Milstein Approximations to the LDCEV Model at Endpoint

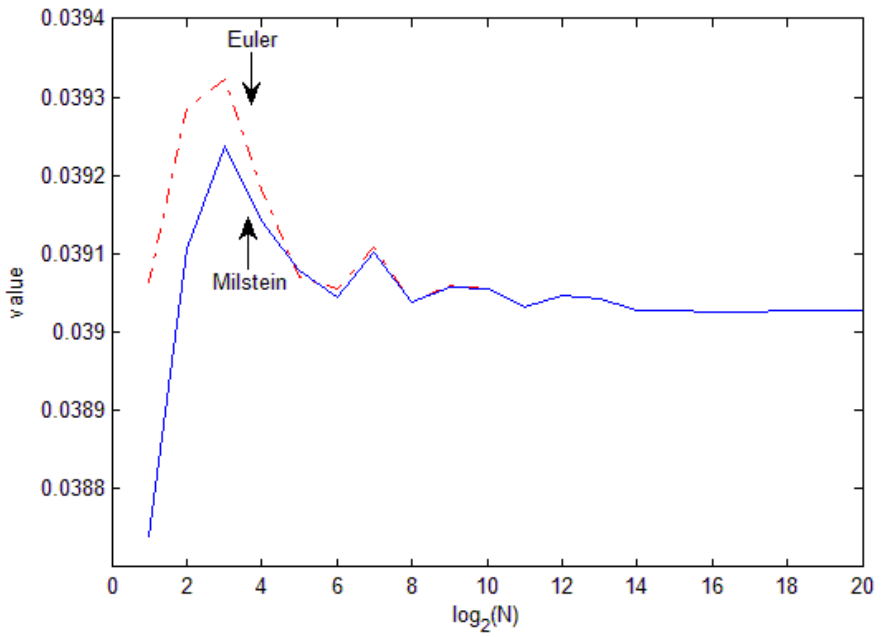


Figure 4.15. Strong Convergence of the Euler and Milstein Approximations to the LDCEV Model

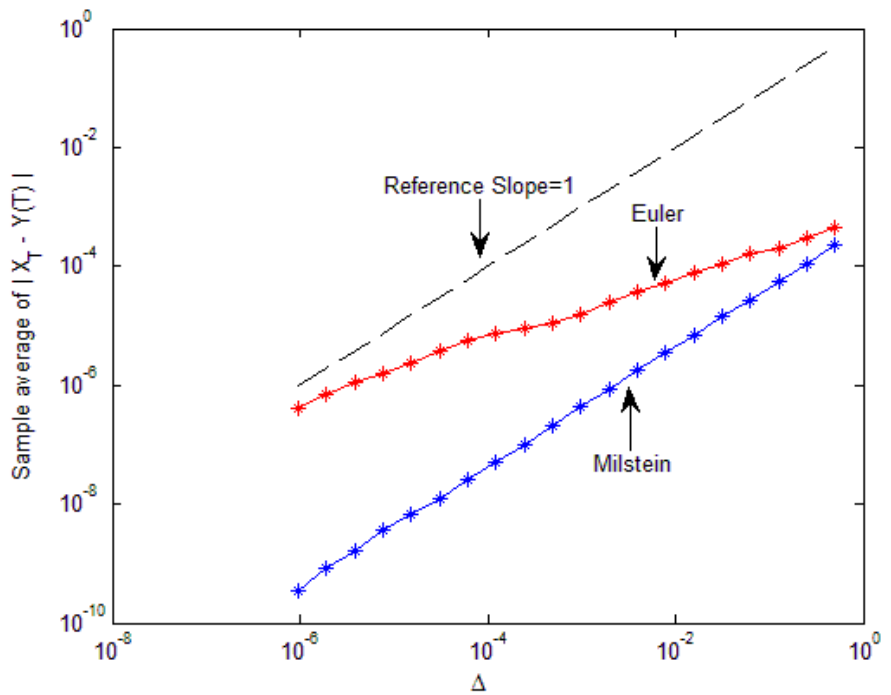
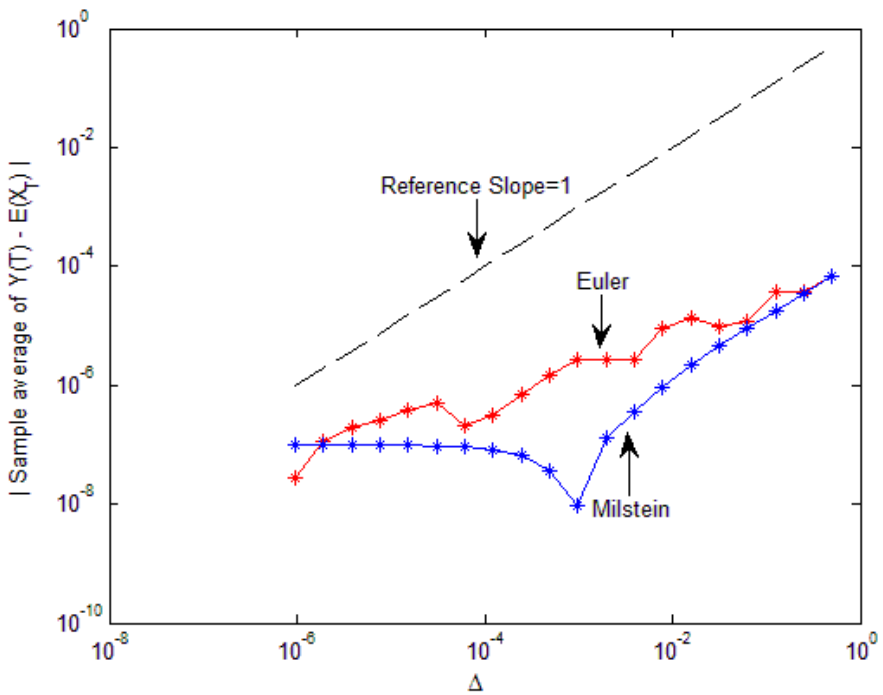


Table 4.10. Percentage Errors of Euler and Milstein Approximations to LDCEV Model at Endpoint

Δ	Euler	Milstein	Δ	Euler	Milstein
$1/2^1$	0.092%	0.738%	$1/2^{11}$	0.013%	0.013%
$1/2^2$	0.664%	0.202%	$1/2^{12}$	0.049%	0.049%
$1/2^3$	0.758%	0.541%	$1/2^{13}$	0.038%	0.038%
$1/2^4$	0.397%	0.295%	$1/2^{14}$	0.000%	0.000%
$1/2^5$	0.113%	0.133%	$1/2^{15}$	0.000%	0.000%
$1/2^6$	0.074%	0.046%	$1/2^{16}$	0.003%	0.003%
$1/2^7$	0.208%	0.192%	$1/2^{17}$	0.003%	0.003%
$1/2^8$	0.031%	0.028%	$1/2^{18}$	0.003%	0.003%
$1/2^9$	0.082%	0.079%	$1/2^{19}$	0.000%	0.000%
$1/2^{10}$	0.072%	0.072%	$1/2^{20}$	0.000%	0.000%

Figure 4.16. Weak Convergence of the Euler and Milstein Approximations to the LDCEV Model



To get the strong and weak convergence of the Euler and Milstein approximations, 100 different discretized Brownian paths over $[0, 1]$ are generated. For each path, the Euler and Milstein approximations are applied for 20 different stepsizes $\Delta = 1/2^N$, $N = 1, 2, \ldots, 20$. The strong and weak convergece of the Euler and Milstein approximationes are analysed in Figures 4.15 and 4.16. Similar to previous models,

the Euler and Milstein approximationes converges with strong order of 0.5 and 1, respectively. Considering the weak convergence, it implies $\Delta = 1/2^{14}$ is the turning point for the Euler scheme and $\Delta = 1/2^{10}$ is the turning point for the Milstein scheme.

4.3.6. Nonlinear Mean Reversion (NMR) Model

Table 4.11. Euler and Milstein Approximations to the NMR Model at Endpoint

Δ	Euler	Milstein	Δ	Euler	Milstein
$1/2^1$	0.039802	0.039559	$1/2^{11}$	0.038502	0.038491
$1/2^2$	0.038881	0.038694	$1/2^{12}$	0.038489	0.038492
$1/2^3$	0.038843	0.038677	$1/2^{13}$	0.038484	0.038491
$1/2^4$	0.038475	0.038586	$1/2^{14}$	0.038484	0.038491
$1/2^5$	0.038466	0.038534	$1/2^{15}$	0.038490	0.038491
$1/2^6$	0.038456	0.038509	$1/2^{16}$	0.038493	0.038491
$1/2^7$	0.038552	0.038496	$1/2^{17}$	0.038493	0.038491
$1/2^8$	0.038541	0.038490	$1/2^{18}$	0.038494	0.038491
$1/2^9$	0.038531	0.038491	$1/2^{19}$	0.038491	0.038491
$1/2^{10}$	0.038500	0.038490	$1/2^{20}$	0.038491	0.038491

Table 4.12. Percentage Errors of Euler and Milstein Approximations to NMR Model at Endpoint

Δ	Euler	Milstein	Δ	Euler	Milstein
$1/2^1$	3.406%	2.775%	$1/2^{11}$	0.029%	0.000%
$1/2^2$	1.013%	0.527%	$1/2^{12}$	0.005%	0.003%
$1/2^3$	0.914%	0.483%	$1/2^{13}$	0.018%	0.000%
$1/2^4$	0.042%	0.247%	$1/2^{14}$	0.018%	0.000%
$1/2^5$	0.065%	0.112%	$1/2^{15}$	0.003%	0.000%
$1/2^6$	0.091%	0.047%	$1/2^{16}$	0.005%	0.000%
$1/2^7$	0.158%	0.013%	$1/2^{17}$	0.005%	0.000%
$1/2^8$	0.130%	0.003%	$1/2^{18}$	0.008%	0.000%
$1/2^9$	0.104%	0.000%	$1/2^{19}$	0.000%	0.000%
$1/2^{10}$	0.023%	0.003%	$1/2^{20}$	0.000%	0.000%

In order to produce different mean reversion when interest rate values in different part of the domain, the NMR model is proposed by Aït-Sahalia (1996b) to specify that the spot interest rate follows

$$dX_t = \left(\alpha_{-1}X_t^{-1} + \alpha_0 + \alpha_1 X_t + \alpha_2 X_t^2\right) dt + \sigma X_t^{\rho} dW_t,$$

where initial deterministic value $X_0 = x_0$. The Euler approximation of the process X is a continuous stochastic process Y satisfying the iteration

$$Y_{t+1} = Y_t + \left(\alpha_{-1} Y_t^{-1} + \alpha_0 + \alpha_1 Y_t + \alpha_2 Y_t^2\right) \cdot \Delta t + \sigma Y_t^{\rho} \cdot \Delta W_t$$

with initial value $Y_0 = X_0$ for $t = 0, 1, \ldots N-1$. The Milstein approximation of X is a continuous stochastic process Z satisfying the iterative scheme

$$Z_{t+1} = Z_t + \left(\alpha_{-1} Z_t^{-1} + \alpha_0 + \alpha_1 Z_t + \alpha_2 Z_t^2\right) \cdot \Delta t + \sigma Z_t^{\rho} \cdot \Delta W_t + \frac{1}{2}\sigma^2 \rho Z_t^{2\rho-1} \left(\Delta W_t^2 - \Delta t\right)$$

with original deterministic value $Z_0 = X_0$ for $t = 0, 1, \ldots N-1$. We take an NMR model with $\alpha_{-1} = 0.00107$, $\alpha_0 = -0.0517$, $\alpha_1 = 0.877$, $\alpha_2 = -4.604$, $\sigma = 0.8047$, $\rho = 3/2$ and $X_0 = 0.03$ as an example. These values are taken from Aït-Sahalia (1999). A Brownian bridge from 0 to 1 is used as W_t. With respect to the endpoint, Figure 4.17 displays the convergence speed of the two approximations. It implies the Euler approximation reaches the reference solution at $\Delta = 1/2^{15}$ and the Milstein approximation comes to the reference solution at $\Delta = 1/2^8$. According to Tables 4.11 and 4.12, the percentage error could be controlled under 0.10% by setting $\Delta \leq 1/2^{10}$ for the Euler approximation.

To get the strong and weak convergence of the Euler and Milstein approximations, 100 different discretized Brownian paths over $[0, 1]$ are generated. For each path, the Euler and Milstein approximations are calculated for 20 different stepsizes $\Delta = 1/2^N$, $N = 1, 2, \ldots, 20$. The Euler and Milstein approximations converge respectively with strong order of 0.5 and 1 in Figure 4.18. In terms of the weak convergence, Figure 4.19 shows the error of the means for the Euler approximation keep declining until the turning point $\Delta = 1/2^{13}$ while the Milstein approximation stops declining from $\Delta = 1/2^{16}$.

Figure 4.17. Convergence Speed of the Euler and Milstein Approximations to the NMR Model at Endpoint

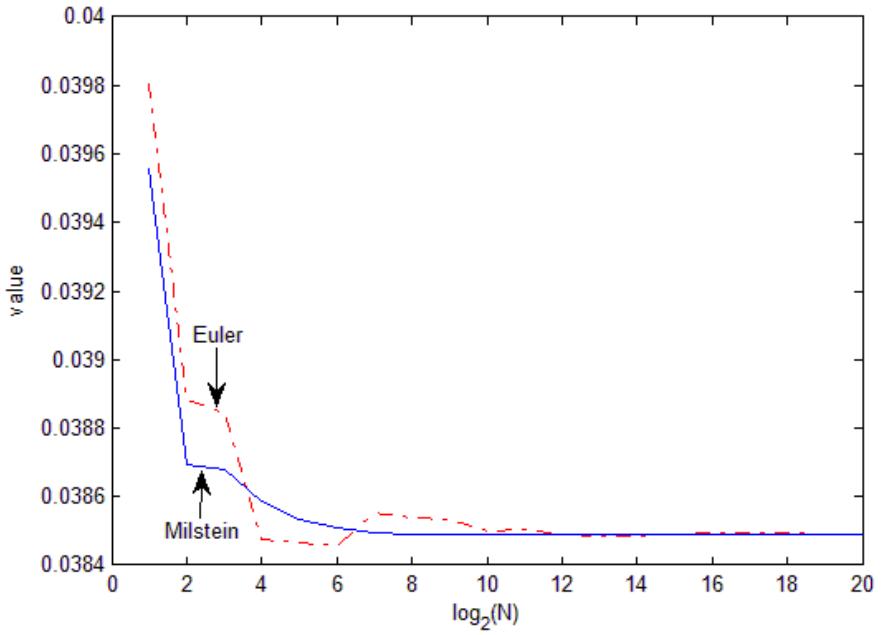


Figure 4.18. Strong Convergence of the Euler and Milstein Approximations to the NMR Model

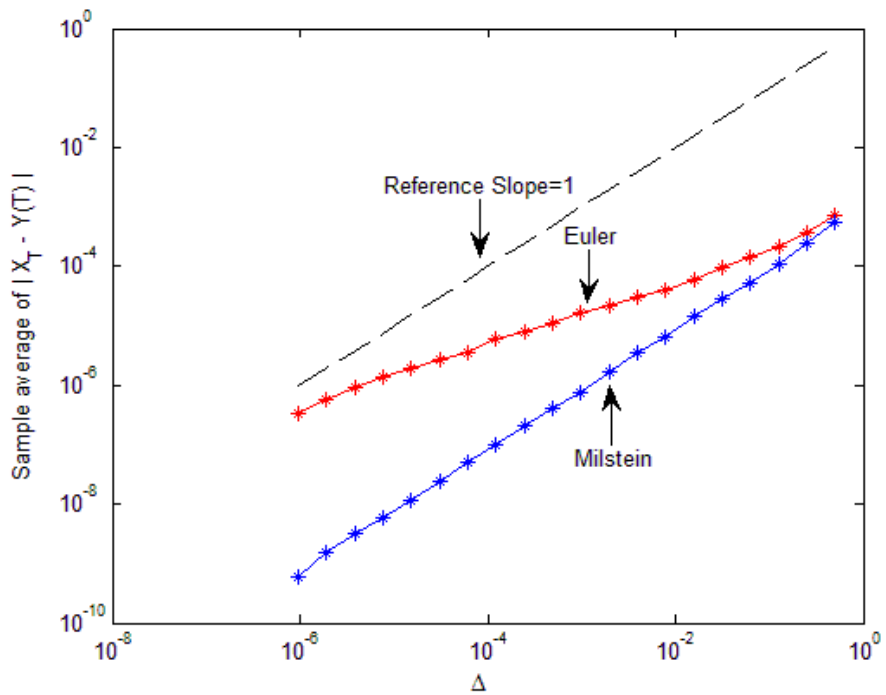
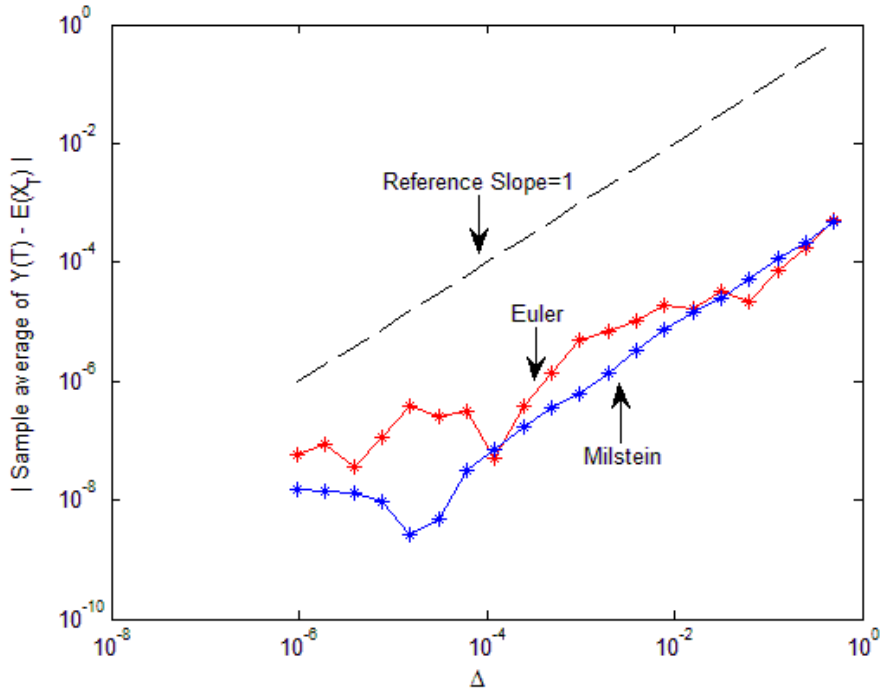


Figure 4.19. Weak Convergence of the Euler and Milstein Approximations to the NMR Model



4.4. Multivariate Stochastic Volatility Models

4.4.1. Heston Model

The Heston model, named after Heston (1993), is a commonly used stochastic volatility model. Under the objective measure P, the joint dynamics of S_t and V_t is

$$d\begin{bmatrix} S_t \\ V_t \end{bmatrix} = \begin{bmatrix} \mu + bV_t \\ \kappa(\gamma - V_t) \end{bmatrix} dt + \begin{bmatrix} \sqrt{(1-\rho^2)V_t} & \rho\sqrt{V_t} \\ 0 & \sigma\sqrt{V_t} \end{bmatrix} d\begin{bmatrix} W_1(t) \\ W_2(t) \end{bmatrix},$$

where $W_1(t)$ and $W_2(t)$ are two uncorrelated standard Brownian motions. The Euler approximations of S and V are continuous stochastic processes X and U satisfying the iterative schemes

$$\begin{aligned} X_{t+1} &= X_t + (\mu + bU_t)\cdot\Delta t + \sqrt{(1-\rho^2)U_t}\cdot\Delta W_1(t) + \rho\sqrt{U_t}\cdot\Delta W_2(t) \\ U_{t+1} &= U_t + k(\gamma - U_t)\cdot\Delta t + \sigma\sqrt{U_t}\cdot\Delta W_2(t) \end{aligned}$$

with initial value $X_0 = S_0$, $U_0 = V_0$ for $t = 0, 1, \ldots N-1$. Since the volatility process V is not influenced directly by the dynamics for S, the Milstein approximation for V is given by the standard one-dimensional formula. Then the fact that $\Sigma_{21} = 0$

and $\partial_{x_1}\Sigma(x) = 0$ simplifies the calculation of the Milstein term for V as a stochastic process Z,

$$Z_{t+1} = Z_t + k(\gamma - Z_t)\cdot\Delta t + \sigma\sqrt{Z_t}\cdot\Delta W_2(t) + \frac{1}{4}\sigma^2\left[(\Delta W_2(t))^2 - \Delta t\right].$$

Consequently, the Milstein approximation for X is a continuous stochastic process Y satisfying the iterative schemes

$$\begin{aligned} Y_{t+1} &= Y_t + (\mu + bZ_t)\cdot\Delta t + \sqrt{(1-\rho^2)Z_t}\cdot\Delta W_1(t) + \rho\sqrt{Z_t}\Delta W_2(t) \\ &\quad + \frac{1}{2}\sigma\sqrt{1-\rho^2}I_{(2,1)} + \frac{1}{2}\sigma\rho I_{(2,2)} \end{aligned}$$

with initial value $Y_0 = S_0$, $Z_0 = V_0$ for $t = 0, 1, \ldots N-1$. Using the parameters of Aït-Sahalia and Kimmel (2007), we set $\kappa = 5.07$, $\gamma = 0.0457$, $\mu = 0.0002$, $\rho = -0.767$, $\sigma = 0.48$, $b = 1.106$ with initial values $X_0 = 6.5$, $V_0 = 0.045$. Two Brownian bridges from 0 to 1 are used as $W_1(t)$ and $W_2(t)$ separately.

Figure 4.20. Convergence Speed of the Euler and Milstein Approximations to the Heston Model at Endpoint

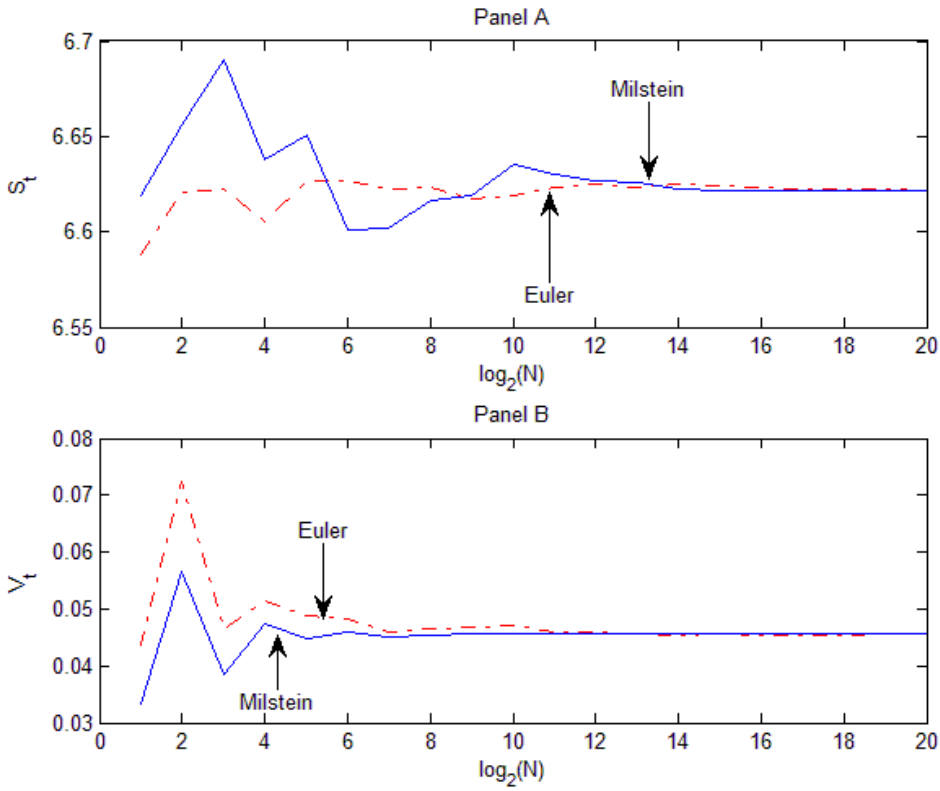


Figure 4.20 plots the Euler and Milstein approximations to the state variables S and V at the endpoint. The Euler and Milstein approximations to the state variable V seems to overlap at $\Delta = 1/2^{11}$, while the Euler and Milstein approximations to the state variable S crosses at $\Delta = 1/2^{7}$ and then appears to converge to a stable value at $\Delta = 1/2^{14}$. The Milstein approximated values to state variable S and V at the time interval $\Delta \leq 1/2^{20}$ still could be considered as the benchmark for the Euler approximations. Then Tables 4.13, 4.14, 4.15, and 4.16 report the endpoint values with percentage errors for the state variable S and V separately. The percentage errors of the Euler approximation to the state variable S could be easily controlled under 0.10% when $\Delta \leq 1/2^{5}$. In comparison, the percentage error of the Euler approximation to the state variable V could be controlled below 1.00% only when $\Delta \leq 1/2^{11}$ and below 0.10% when $\Delta \leq 1/2^{19}$. Hence it seems that the state variable V is much harder to control than state variable S. Then our numerical tests focus on the state variable V rather than S.

Table 4.13. Euler and Milstein Approximations to S of the Heston Model at Endpoint

Δ	Euler	Milstein	Δ	Euler	Milstein
$1/2^{1}$	6.588023	6.619029	$1/2^{11}$	6.623445	6.629996
$1/2^{2}$	6.620444	6.655676	$1/2^{12}$	6.624835	6.627112
$1/2^{3}$	6.622374	6.690201	$1/2^{13}$	6.622975	6.626153
$1/2^{4}$	6.605708	6.637825	$1/2^{14}$	6.625161	6.622605
$1/2^{5}$	6.626960	6.650608	$1/2^{15}$	6.624526	6.621842
$1/2^{6}$	6.626891	6.601293	$1/2^{16}$	6.623693	6.621480
$1/2^{7}$	6.622450	6.601614	$1/2^{17}$	6.622851	6.621252
$1/2^{8}$	6.623689	6.616302	$1/2^{18}$	6.622599	6.621525
$1/2^{9}$	6.616881	6.619136	$1/2^{19}$	6.622221	6.621864
$1/2^{10}$	6.618738	6.635436	$1/2^{20}$	6.621962	6.621908

Table 4.14. Percentage Errors of the Euler and Milstein Approximations to S of the Heston Model at Endpoint

Δ	Euler	Milstein	Δ	Euler	Milstein
$1/2^1$	0.512%	0.738%	$1/2^{11}$	0.023%	0.013%
$1/2^2$	0.022%	0.202%	$1/2^{12}$	0.044%	0.049%
$1/2^3$	0.007%	0.541%	$1/2^{13}$	0.016%	0.038%
$1/2^4$	0.245%	0.295%	$1/2^{14}$	0.049%	0.000%
$1/2^5$	0.076%	0.133%	$1/2^{15}$	0.040%	0.000%
$1/2^6$	0.075%	0.046%	$1/2^{16}$	0.027%	0.003%
$1/2^7$	0.008%	0.192%	$1/2^{17}$	0.014%	0.003%
$1/2^8$	0.027%	0.028%	$1/2^{18}$	0.010%	0.003%
$1/2^9$	0.076%	0.079%	$1/2^{19}$	0.005%	0.000%
$1/2^{10}$	0.048%	0.072%	$1/2^{20}$	0.001%	0.000%

Table 4.15. Euler and Milstein Approximations to V of the Heston Model at Endpoint

Δ	Euler	Milstein	Δ	Euler	Milstein
$1/2^1$	0.043626	0.033309	$1/2^{11}$	0.046021	0.045558
$1/2^2$	0.072311	0.056541	$1/2^{12}$	0.045914	0.045572
$1/2^3$	0.046425	0.038409	$1/2^{13}$	0.045652	0.045578
$1/2^4$	0.051314	0.047392	$1/2^{14}$	0.045404	0.045579
$1/2^5$	0.048815	0.044718	$1/2^{15}$	0.045564	0.045582
$1/2^6$	0.048271	0.045927	$1/2^{16}$	0.045468	0.045581
$1/2^7$	0.046018	0.045213	$1/2^{17}$	0.045507	0.045581
$1/2^8$	0.046562	0.045300	$1/2^{18}$	0.045512	0.045582
$1/2^9$	0.046900	0.045609	$1/2^{19}$	0.045540	0.045581
$1/2^{10}$	0.047082	0.045594	$1/2^{20}$	0.045558	0.045581

Figure 4.21. Strong Convergence of the Euler and Milstein Approximations to the Heston Model

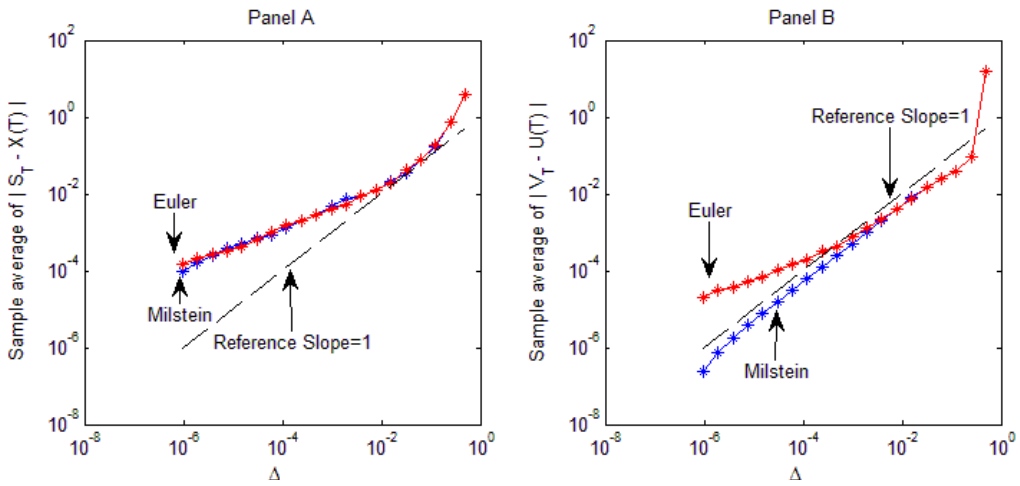
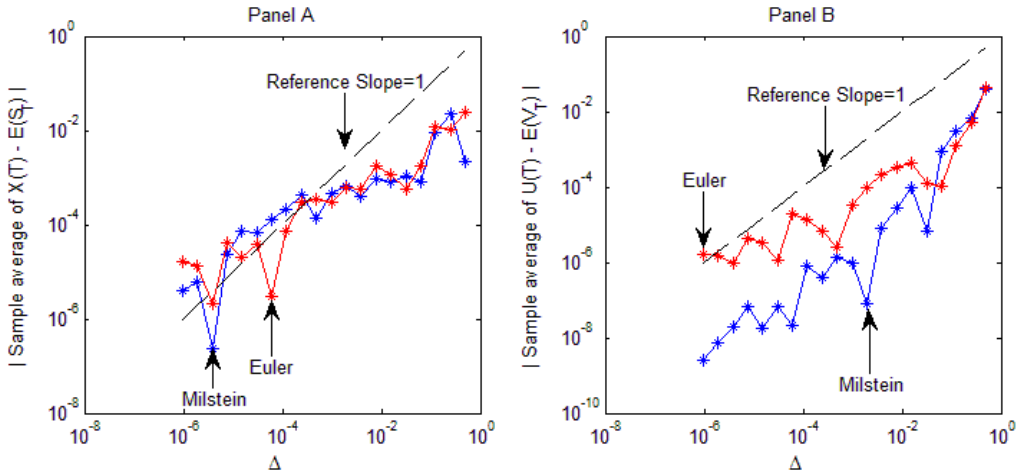


Table 4.16. Percentage Errors of the Euler and Milstein Approximations to V of the Heston Model at Endpoint

Δ	Euler	Milstein	Δ	Euler	Milstein
$1/2^1$	4.290%	26.924%	$1/2^{11}$	0.966%	0.051%
$1/2^2$	58.643%	24.045%	$1/2^{12}$	0.729%	0.020%
$1/2^3$	1.851%	15.736%	$1/2^{13}$	0.154%	0.007%
$1/2^4$	12.576%	3.973%	$1/2^{14}$	0.390%	0.005%
$1/2^5$	7.093%	1.895%	$1/2^{15}$	0.039%	0.002%
$1/2^6$	5.901%	0.759%	$1/2^{16}$	0.248%	0.001%
$1/2^7$	0.958%	0.809%	$1/2^{17}$	0.163%	0.000%
$1/2^8$	2.151%	0.616%	$1/2^{18}$	0.152%	0.001%
$1/2^9$	2.893%	0.062%	$1/2^{19}$	0.091%	0.000%
$1/2^{10}$	3.292%	0.028%	$1/2^{20}$	0.050%	0.000%

Figure 4.22. Weak Convergence of the Euler and Milstein Approximations to the Heston Model



Figures 4.21 and 4.22 investigate the strong and weak convergence of the Euler and Milstein approximations by generating 100 different discretized Brownian paths over $[0, 1]$. In the following models, including the GARCH, CEV, Stein, Scott, Hull-White, Hagan and SABR process, we also generate 100 paths to analyse the strong and weak convergence of the Euler and Milstein approximations. From Figure 4.21, we could see the mean absolute different values keep declining although the strong convergence of the state variable S and V are not equal to their theoretical values. However, based on the weak convergence, $\Delta = 1/2^{11}$ seems to be the turning point of the Euler approximation to the state variable V. To conclude, $\Delta = 1/2^{11}$ could

be an effective time step for the Heston model by considering both of the pathwise approximation and moments approximation requirements.

4.4.2. GARCH Model

In the GARCH stochastic volatility model, the joint dynamics of S_t and Y_t is defined as

$$d\begin{bmatrix} S_t \\ V_t \end{bmatrix} = \begin{bmatrix} \mu + bV_t \\ \kappa(\gamma - V_t) \end{bmatrix} dt + \begin{bmatrix} \sqrt{(1-\rho^2)V_t} & \rho\sqrt{V_t} \\ 0 & \sigma V_t \end{bmatrix} d\begin{bmatrix} W_1^P(t) \\ W_2^P(t) \end{bmatrix}$$

under the objective measure P, where $W_1(t)$ and $W_2(t)$ are two uncorrelated standard Brownian motions. The Euler approximations of S and V are continuous stochastic processes X and U satisfying the iterative schemes

$$\begin{aligned} X_{t+1} &= X_t + (\mu + bU_t)\cdot\Delta t + \sqrt{(1-\rho^2)U_t}\cdot\Delta W_1(t) + \rho\sqrt{U_t}\Delta W_2(t) \\ U_{t+1} &= U_t + k(\gamma - U_t)\cdot\Delta t + \sigma U_t\cdot\Delta W_2(t) \end{aligned}$$

with initial value $X_0 = S_0, U_0 = V_0$ for $t = 0, 1, \ldots N-1$. The Milstein approximations of S and V are in the form of

$$\begin{aligned} Y_{t+1} &= Y_t + (\mu + bZ_t)\cdot\Delta t + \sqrt{(1-\rho^2)Z_t}\cdot\Delta W_1(t) + \rho\sqrt{Z_t}\Delta W_2(t) \\ &\quad + \frac{1}{2}\sigma\sqrt{(1-\rho^2)Z_t}I_{(2,1)} + \frac{1}{2}\sigma\rho\sqrt{Z_t}I_{(2,2)} \\ Z_{t+1} &= Z_t + k(\gamma - Z_t)\cdot\Delta t + \sigma Z_t\cdot\Delta W_2(t) + \frac{1}{2}\sigma^2 Z_t\left[(\Delta W_2(t))^2 - \Delta t\right] \end{aligned}$$

with initial value $Y_0 = S_0$, $Z_0 = V_0$ for $t = 0, 1, \ldots N-1$. Using the parameters of Aït-Sahalia and Kimmel (2007), we set $\kappa = 1.62$, $\gamma = 0.074$, $\mu = 0.0002$, $\rho = -0.754$, $\sigma = 2.204$, $b = 0.5356$ with initial values $X_0 = 6.5$, $V_0 = 0.045$. Two Brownian bridges from 0 to 1 are used as $W_1(t)$ and $W_2(t)$ separately. Panel A of Figure 4.23 displays that the Euler and Milstein approximations to the state variables S at the endpoint coincide when $\Delta = 1/2^9$. Similarly, the Euler and Milstein approximations to the state variables V overlaps at $\Delta = 1/2^9$ in Panel B of Figure 4.23. Tables 4.17, 4.18, 4.19 and 4.20 report the endpoint values with percentage errors for the state variable S and V separately. The percentage errors of the Euler approximations to the state variable S fluctuate with different time interval Δ and then could be controlled under 1.00% until $\Delta = 1/2^{13}$. On the contrary, the percentage error of the Euler approximations to the state variable V seems very difficult to manage even

with a much finer Δ, such as $1/2^{18}$. The key to examine the accuracy of the Euler approximation to the GARCH model still depends on the variable V.

Figure 4.23. Convergence Speed of Euler and Milstein Approximations to the GARCH Model at Endpoint

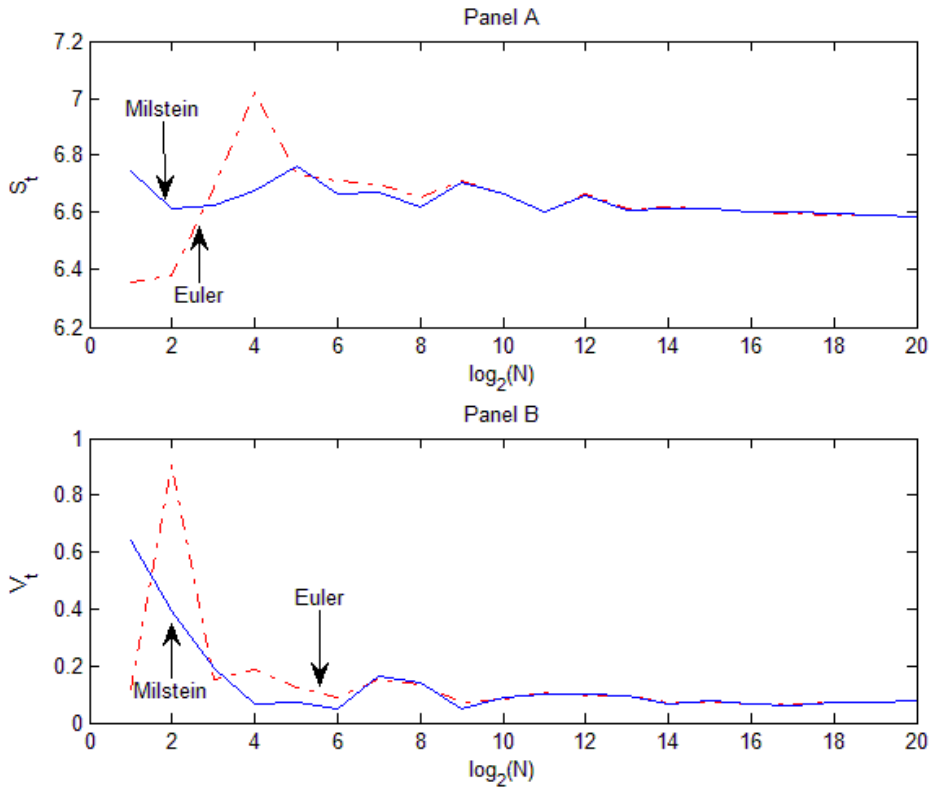


Table 4.17. Euler and Milstein Approximations to S of the GARCH Model at Endpoint

Δ	Euler	Milstein	Δ	Euler	Milstein
$1/2^1$	6.358827	6.745628	$1/2^{11}$	6.603617	6.600676
$1/2^2$	6.381579	6.615489	$1/2^{12}$	6.662975	6.657750
$1/2^3$	6.690366	6.627843	$1/2^{13}$	6.612067	6.609285
$1/2^4$	7.019722	6.678406	$1/2^{14}$	6.622412	6.614624
$1/2^5$	6.737023	6.761393	$1/2^{15}$	6.615832	6.613851
$1/2^6$	6.710716	6.664585	$1/2^{16}$	6.603039	6.603400
$1/2^7$	6.701118	6.672786	$1/2^{17}$	6.598838	6.600938
$1/2^8$	6.652367	6.617784	$1/2^{18}$	6.592789	6.595184
$1/2^9$	6.709254	6.705222	$1/2^{19}$	6.591800	6.594109
$1/2^{10}$	6.664480	6.666364	$1/2^{20}$	6.586996	6.586554

Table 4.18. Percentage Errors of the Euler and Milstein Approximations to S of the GARCH Model at Endpoint

Δ	Euler	Milstein	Δ	Euler	Milstein
$1/2^1$	3.457%	2.415%	$1/2^{11}$	0.259%	0.214%
$1/2^2$	3.112%	0.439%	$1/2^{12}$	1.160%	1.081%
$1/2^3$	1.576%	0.627%	$1/2^{13}$	0.387%	0.345%
$1/2^4$	6.577%	1.395%	$1/2^{14}$	0.544%	0.426%
$1/2^5$	2.284%	2.654%	$1/2^{15}$	0.445%	0.414%
$1/2^6$	1.885%	1.185%	$1/2^{16}$	0.250%	0.256%
$1/2^7$	1.739%	1.309%	$1/2^{17}$	0.186%	0.218%
$1/2^8$	0.999%	0.474%	$1/2^{18}$	0.095%	0.131%
$1/2^9$	1.863%	1.802%	$1/2^{19}$	0.080%	0.115%
$1/2^{10}$	1.183%	1.212%	$1/2^{20}$	0.007%	0.000%

Table 4.19. Euler and Milstein Approximations to V of the GARCH Model at Endpoint

Δ	Euler	Milstein	Δ	Euler	Milstein
$1/2^1$	0.119116	0.641979	$1/2^{11}$	0.109113	0.103050
$1/2^2$	0.902997	0.394653	$1/2^{12}$	0.097299	0.098670
$1/2^3$	0.152487	0.192744	$1/2^{13}$	0.097572	0.095362
$1/2^4$	0.185309	0.067453	$1/2^{14}$	0.069757	0.068635
$1/2^5$	0.126186	0.069022	$1/2^{15}$	0.072404	0.078131
$1/2^6$	0.088246	0.049968	$1/2^{16}$	0.065336	0.066824
$1/2^7$	0.153296	0.165749	$1/2^{17}$	0.066440	0.060608
$1/2^8$	0.137912	0.139083	$1/2^{18}$	0.070234	0.069961
$1/2^9$	0.069735	0.049900	$1/2^{19}$	0.073030	0.073572
$1/2^{10}$	0.086056	0.090139	$1/2^{20}$	0.074920	0.075070

Figure 4.24. Strong Convergence of the Euler and Milstein Approximations to the GARCH Model

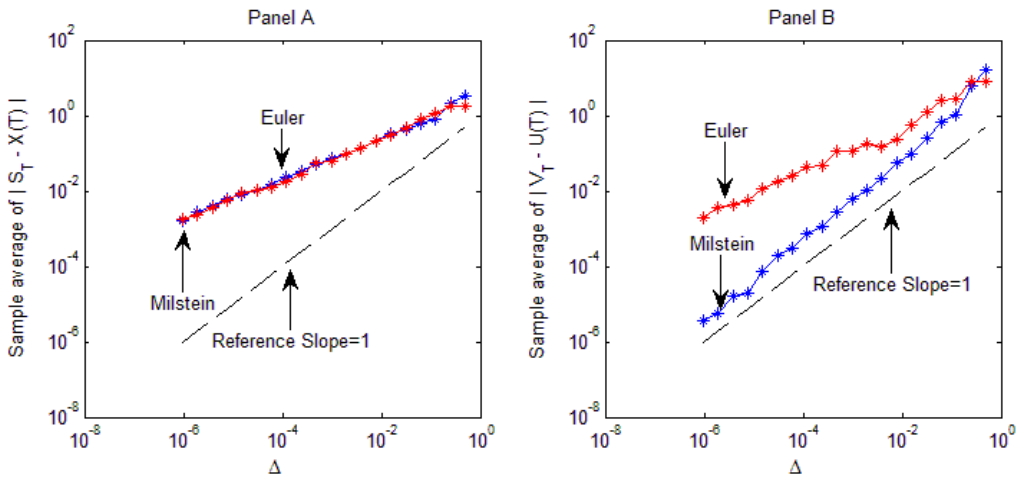
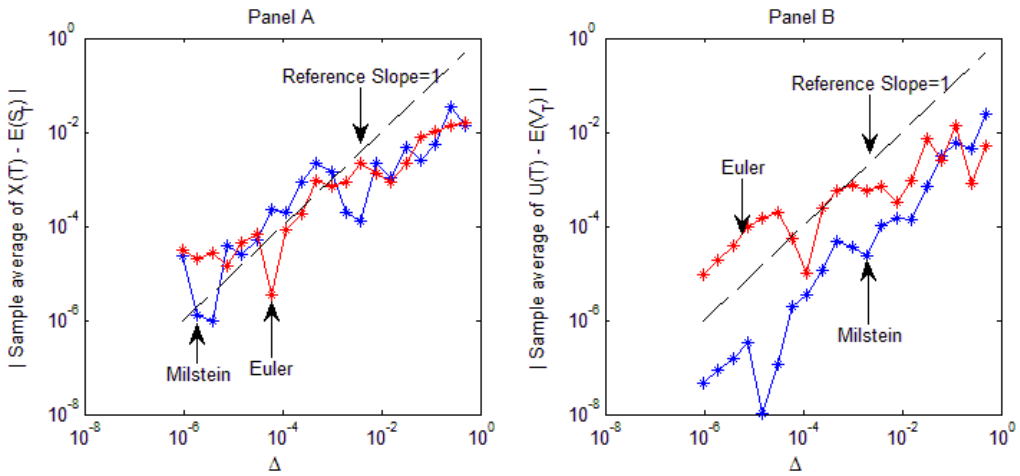


Table 4.20. Percentage Errors of the Euler and Milstein Approximations to V of the GARCH Model at Endpoint

Δ	Euler	Milstein	Δ	Euler	Milstein
$1/2^1$	58.674%	755.175%	$1/2^{11}$	45.349%	37.272%
$1/2^2$	1102.876%	425.715%	$1/2^{12}$	29.611%	31.438%
$1/2^3$	103.127%	156.753%	$1/2^{13}$	29.748%	27.031%
$1/2^4$	146.849%	10.146%	$1/2^{14}$	7.078%	8.572%
$1/2^5$	68.092%	8.056%	$1/2^{15}$	3.551%	4.078%
$1/2^6$	17.552%	33.438%	$1/2^{16}$	12.967%	10.985%
$1/2^7$	104.204%	120.793%	$1/2^{17}$	11.496%	19.264%
$1/2^8$	83.712%	85.271%	$1/2^{18}$	6.442%	6.806%
$1/2^9$	7.106%	33.529%	$1/2^{19}$	2.717%	0.020%
$1/2^{10}$	14.635%	20.074%	$1/2^{20}$	0.199%	0.000%

Figure 4.25. Weak Convergence of the Euler and Milstein Approximations to the GARCH Model



The strong convergence of the Euler and Milstein approximations are shown is Figure 4.24 and the weak convergence is displayed in Figure 4.25. Similar to the Heston model, the strong convergence performs quite well with some small volatilities from $\Delta = 1/2^1$ to $\Delta = 1/2^{20}$. In terms of the weak convergence, $\Delta = 1/2^{13}$ seems to be the turning point of the Euler approximation to the state variable V. Overall, if a good pathwise approximation is required, it's better to set Δ as small as possible. While, if a good approximation of moments is needed, $\Delta = 1/2^{13}$ is probably a relatively good choice for the Euler approximation.

4.4.3. CEV Model

Under the objective measure P, the CEV model is in the form of

$$d\begin{bmatrix} S_t \\ V_t \end{bmatrix} = \begin{bmatrix} \mu + bV_t \\ \kappa(\gamma - V_t) \end{bmatrix} dt + \begin{bmatrix} \sqrt{(1-\rho^2)V_t} & \rho\sqrt{V_t} \\ 0 & \sigma V_t^{\beta} \end{bmatrix} d\begin{bmatrix} W_1^P(t) \\ W_2^P(t) \end{bmatrix},$$

where $W_1(t)$ and $W_2(t)$ are two uncorrelated standard Brownian motions. The Euler approximations of S and V are continuous stochastic processes X and U satisfying the iterative schemes

$$\begin{aligned} X_{t+1} &= X_t + (\mu + bU_t)\cdot\Delta t + \sqrt{(1-\rho^2)U_t}\cdot\Delta W_1(t) + \rho\sqrt{U_t}\Delta W_2(t) \\ U_{t+1} &= U_t + k(\gamma - U_t)\cdot\Delta t + \sigma U_t^{\beta}\cdot\Delta W_2(t) \end{aligned}$$

with initial value $X_0 = S_0$, $U_0 = V_0$ for $t = 0, 1, \ldots N-1$. The Milstein approximations of S and V are continuous stochastic processes Y and Z satisfying the iterative schemes

$$\begin{aligned} Y_{t+1} &= Y_t + (\mu + bZ_t)\cdot\Delta t + \sqrt{(1-\rho^2)Z_t}\cdot\Delta W_1(t) + \rho\sqrt{Z_t}\Delta W_2(t) \\ &\quad + \beta\sigma\sqrt{(1-\rho^2)}Z_t^{\beta-\frac{1}{2}}I_{(2,1)} + \beta\sigma\rho Z_t^{\beta-\frac{1}{2}}I_{(2,2)} \\ Z_{t+1} &= Z_t + k(\gamma - Z_t)\cdot\Delta t + \sigma Z_t^{\beta}\cdot\Delta W_2(t) + \frac{1}{2}\beta\sigma^2 Z_t^{2\beta-1}\left[(\Delta W_2(t))^2 - \Delta t\right] \end{aligned}$$

with initial value $Y_0 = S_0$, $Z_0 = V_0$ for $t = 0, 1, \ldots N-1$. Using the parameters of Aït-Sahalia and Kimmel (2007), we set $\kappa = 4.1031$, $\gamma = 0.0451$, $\mu = 0.0002$, $\rho = -0.760$, $\sigma = 0.8583$, $b = 1.1474$ with initial values $X_0 = 6.5$, $V_0 = 0.045$. Two Brownian bridges from 0 to 1 are used as $W_1(t)$ and $W_2(t)$ separately. Figure 4.26 plots the Euler and Milstein approximations to the state variables S and V at the endpoint. The Euler and Milstein approximations to the state variable V seems to overlap at $\Delta = 1/2^{14}$, while the Euler and Milstein approximations to the state variable S appears to converge to a stable value at $\Delta = 1/2^{11}$. Tables 4.21, 4.22, 4.23 and 4.24 report the endpoint values with percentage errors for the state variable S and V separately. The percentage errors of the Euler approximations to the state variable S could be controlled under 0.10% when $\Delta \leq 1/2^{10}$ and under 0.01% when $\Delta \leq 1/2^{18}$, whereas the percentage error of the Euler approximations to the state variable V could be controlled below 1.00% when $\Delta \leq 1/2^{13}$ and below 0.10% when $\Delta \leq 1/2^{20}$. The key to examine the accuracy of the Euler approximation to the CEV

model does not only depend on the state variable V but also depend on the variable S.

Figure 4.26. Convergence Speed of the Euler and Milstein Approximations to the CEV Model at Endpoint

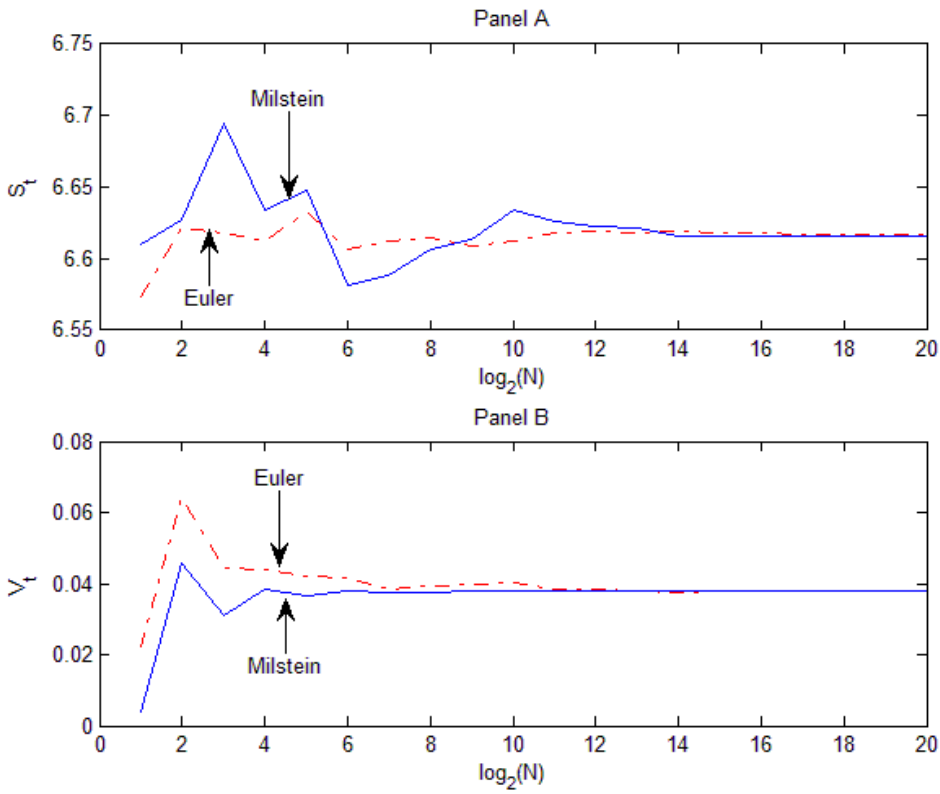


Table 4.21. Euler and Milstein Approximations to S of the CEV Model at Endpoint

Δ	Euler	Milstein	Δ	Euler	Milstein
$1/2^1$	6.573591	6.610135	$1/2^{11}$	6.618049	6.625998
$1/2^2$	6.621955	6.627164	$1/2^{12}$	6.619145	6.622116
$1/2^3$	6.618450	6.694917	$1/2^{13}$	6.617617	6.621000
$1/2^4$	6.612742	6.633994	$1/2^{14}$	6.619256	6.616181
$1/2^5$	6.632596	6.648246	$1/2^{15}$	6.618511	6.615546
$1/2^6$	6.606149	6.580954	$1/2^{16}$	6.617783	6.615172
$1/2^7$	6.611966	6.587901	$1/2^{17}$	6.617035	6.615143
$1/2^8$	6.614067	6.607018	$1/2^{18}$	6.616855	6.615518
$1/2^9$	6.609126	6.613624	$1/2^{19}$	6.616497	6.616021
$1/2^{10}$	6.612020	6.633781	$1/2^{20}$	6.616289	6.616168

Table 4.22. Percentage Errors of the Euler and Milstein Approximations to S of the CEV Model at Endpoint

Δ	Euler	Milstein	Δ	Euler	Milstein
$1/2^1$	0.644%	0.091%	$1/2^{11}$	0.028%	0.149%
$1/2^2$	0.087%	0.166%	$1/2^{12}$	0.045%	0.090%
$1/2^3$	0.034%	1.190%	$1/2^{13}$	0.022%	0.073%
$1/2^4$	0.052%	0.269%	$1/2^{14}$	0.047%	0.000%
$1/2^5$	0.248%	0.485%	$1/2^{15}$	0.035%	0.010%
$1/2^6$	0.151%	0.532%	$1/2^{16}$	0.024%	0.015%
$1/2^7$	0.064%	0.427%	$1/2^{17}$	0.013%	0.015%
$1/2^8$	0.032%	0.138%	$1/2^{18}$	0.010%	0.010%
$1/2^9$	0.106%	0.038%	$1/2^{19}$	0.005%	0.002%
$1/2^{10}$	0.063%	0.266%	$1/2^{20}$	0.002%	0.000%

Table 4.23. Euler and Milstein Approximations to V of the CEV Model at Endpoint

Δ	Euler	Milstein	Δ	Euler	Milstein
$1/2^1$	0.022157	0.003924	$1/2^{11}$	0.038465	0.037813
$1/2^2$	0.064295	0.045518	$1/2^{12}$	0.038312	0.037834
$1/2^3$	0.044484	0.030846	$1/2^{13}$	0.037888	0.037845
$1/2^4$	0.043486	0.038398	$1/2^{14}$	0.037592	0.037846
$1/2^5$	0.042191	0.036550	$1/2^{15}$	0.037847	0.037848
$1/2^6$	0.041837	0.037844	$1/2^{16}$	0.037708	0.037847
$1/2^7$	0.038289	0.037552	$1/2^{17}$	0.037759	0.037847
$1/2^8$	0.039479	0.037487	$1/2^{18}$	0.037759	0.037847
$1/2^9$	0.039758	0.037874	$1/2^{19}$	0.037789	0.037847
$1/2^{10}$	0.040025	0.037857	$1/2^{20}$	0.037812	0.037847

Figure 4.27. Strong Convergence of the Euler and Milstein Approximations to the CEV Model

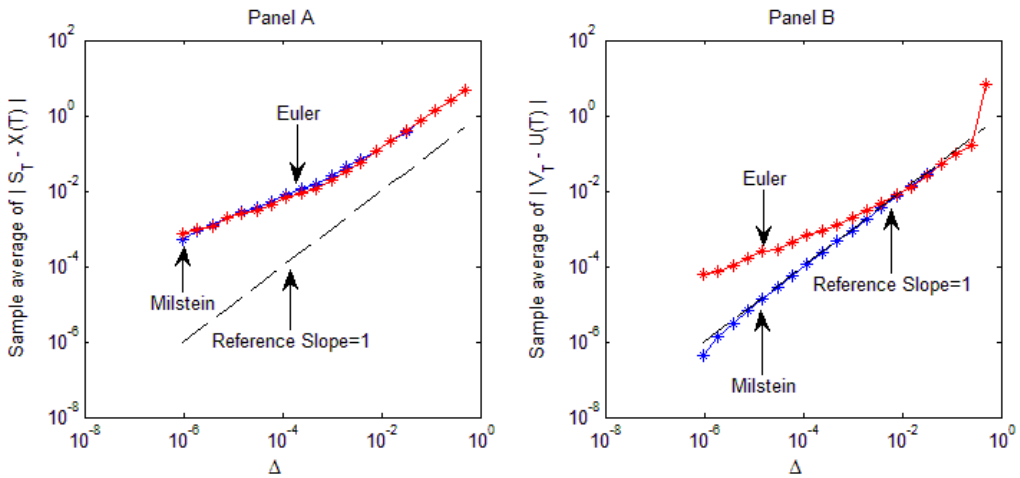
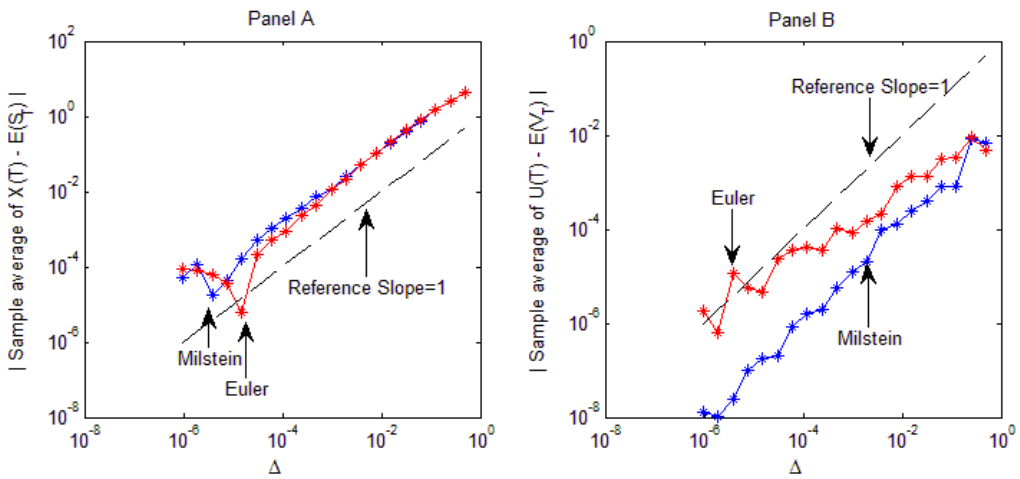


Table 4.24. Percentage Errors of the Euler and Milstein Approximations to V of the CEV Model at Endpoint

Δ	Euler	Milstein	Δ	Euler	Milstein
$1/2^1$	41.456%	89.633%	$1/2^{11}$	1.633%	0.090%
$1/2^2$	69.883%	20.269%	$1/2^{12}$	1.229%	0.034%
$1/2^3$	17.538%	18.499%	$1/2^{13}$	0.108%	0.005%
$1/2^4$	15.694%	1.456%	$1/2^{14}$	0.672%	0.004%
$1/2^5$	11.478%	3.426%	$1/2^{15}$	0.000%	0.003%
$1/2^6$	10.543%	0.006%	$1/2^{16}$	0.367%	0.000%
$1/2^7$	1.169%	0.780%	$1/2^{17}$	0.233%	0.000%
$1/2^8$	4.312%	0.950%	$1/2^{18}$	0.231%	0.001%
$1/2^9$	5.048%	0.071%	$1/2^{19}$	0.153%	0.000%
$1/2^{10}$	5.756%	0.027%	$1/2^{20}$	0.092%	0.000%

Figure 4.28. Weak Convergence of the Euler and Milstein Approximations to the CEV Model



Figures 4.27 and 4.28 illustrate the strong and weak convergence of the Euler and Milstein approximation. From Figure 4.27, we could see the mean absolute different values keep declining although the strong convergence of the state variable S and V are not equal to their theoretical values. However, based on the weak convergence, there seems no turning points for the Euler approximation to the state variable V. Moving to the state variable S, the turning point of the Euler approximation appears at $\Delta = 1/2^{16}$. To conclude, $\Delta = 1/2^{16}$ could be an effective time step for the CEV model by considering the moments approximation requirement.

4.4.4. Hull-White (HW) Model

The HW stochastic volatility model, proposed by Hull and White (1987), is defined as

$$d\begin{bmatrix} S_t \\ V_t \end{bmatrix} = \begin{bmatrix} \mu_\mu - \frac{1}{2}V_t \\ \nu_\mu V_t \end{bmatrix} dt + \begin{bmatrix} \sqrt{(1-\rho^2)\,V_t} & \rho\sqrt{V_t} \\ 0 & \sigma V_t \end{bmatrix} d\begin{bmatrix} W_1^P(t) \\ W_2^P(t) \end{bmatrix}$$

under the objective measure P, where $W_1(t)$ and $W_2(t)$ are two uncorrelated standard Brownian motions.

The Euler approximations of S and V are continuous stochastic processes X and U satisfying the iterative schemes

$$\begin{aligned} X_{t+1} &= X_t + \left(\mu_\mu - \frac{1}{2}U_t\right)\cdot\Delta t + \sqrt{(1-\rho^2)\,U_t}\cdot\Delta W_1(t) + \rho\sqrt{U_t}\Delta W_2(t) \\ U_{t+1} &= U_t + v_\mu U_t\cdot\Delta t + \sigma U_t\cdot\Delta W_2(t) \end{aligned}$$

with initial value $X_0 = S_0$, $U_0 = V_0$ for $t = 0, 1, \ldots N-1$.

The Milstein approximation of S and V are in the form of

$$\begin{aligned} Y_{t+1} &= Y_t + \left(\mu_\mu - \frac{1}{2}Z_t\right)\cdot\Delta t + \sqrt{(1-\rho^2)\,Z_t}\cdot\Delta W_1(t) + \rho\sqrt{Z_t}\Delta W_2(t) \\ &\quad +\sigma\sqrt{(1-\rho^2)\,Z_t}I_{(2,1)} + \sigma\rho\sqrt{Z_t}I_{(2,2)} \\ Z_{t+1} &= Z_t + \mu_v Z_t\cdot\Delta t + \sigma Z_t\cdot\Delta W_2(t) + \frac{1}{2}\sigma^2 Z_t\left[(\Delta W_2(t))^2 - \Delta t\right] \end{aligned}$$

with initial value $Y_0 = S_0$, $Z_0 = V_0$ for $t = 0, 1, \ldots N-1$.

Using the parameters of Hull and White (1987), we set $\mu_\mu = 0$, $v_\mu = 0$, $\rho = -0.5$, $\sigma = 1$ with initial values $X_0 = 1$, $V_0 = 0.5$. Two Brownian bridges from 0 to 1 are used as $W_1(t)$ and $W_2(t)$ separately.

Panel A of Figure 4.29 displays that the Euler and Milstein approximations to the state variables S at the endpoint coincide when $\Delta = 1/2^{12}$. Similarly, the Euler and Milstein approximations to the state variables V overlaps at $\Delta = 1/2^{13}$ in Panel B of Figure 4.29.

Tables 4.25, 4.26, 4.27 and 4.28 report the endpoint values with percentage errors for the state variable S and V separately. The percentage errors of the Euler approximations to the state variable S could be easily controlled under 0.10% even when $\Delta \leq 1/2^7$. In comparison, the percentage error of the Euler approximations to the state variable V could be controlled below 1.00% only when $\Delta \leq 1/2^{13}$. For this reason, the volatility variable V is probably the key of our numerical tests.

Figure 4.29. Convergence Speed of the Euler and Milstein Approximations to the HW Model at Endpoint

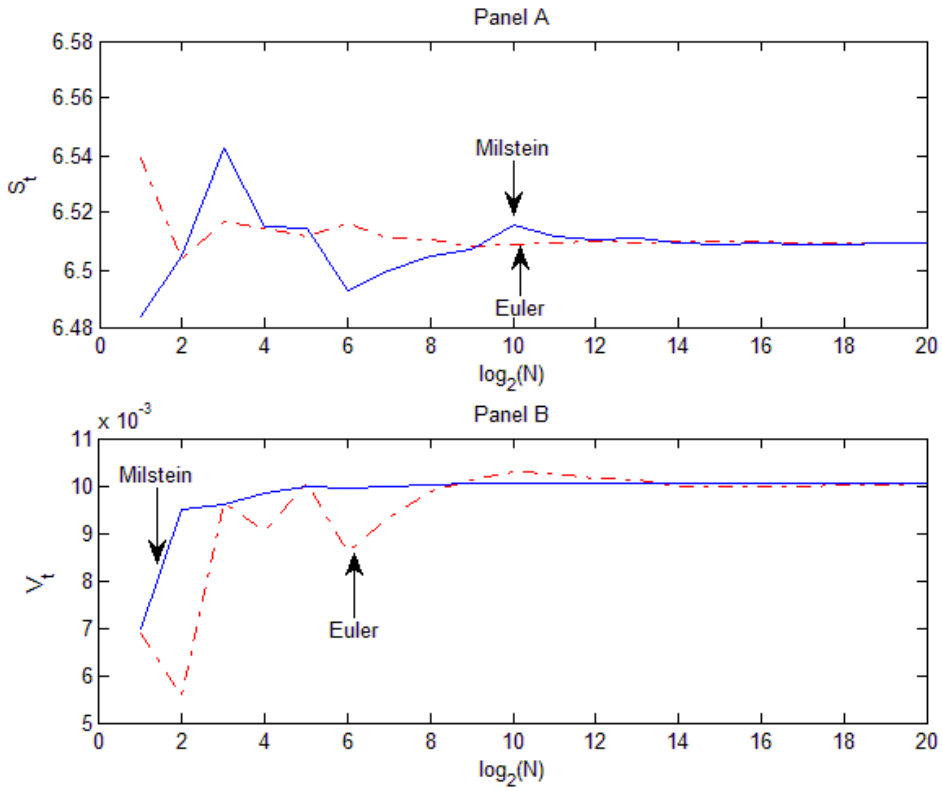


Table 4.25. Euler and Milstein Approximations to S of the HW Model at Endpoint

Δ	Euler	Milstein	Δ	Euler	Milstein
$1/2^1$	6.539388	6.483894	$1/2^{11}$	6.509479	6.511973
$1/2^2$	6.503988	6.504705	$1/2^{12}$	6.510221	6.510707
$1/2^3$	6.517041	6.542775	$1/2^{13}$	6.509567	6.511215
$1/2^4$	6.514346	6.515121	$1/2^{14}$	6.510167	6.509433
$1/2^5$	6.511866	6.514692	$1/2^{15}$	6.510072	6.509074
$1/2^6$	6.516269	6.492866	$1/2^{16}$	6.509821	6.509172
$1/2^7$	6.511361	6.499614	$1/2^{17}$	6.509568	6.508955
$1/2^8$	6.510735	6.504708	$1/2^{18}$	6.509496	6.509118
$1/2^9$	6.508178	6.506863	$1/2^{19}$	6.509409	6.509264
$1/2^{10}$	6.508671	6.515461	$1/2^{20}$	6.509324	6.509336

Table 4.26. Percentage Errors of the Euler and Milstein Approximations to S of the HW Model at Endpoint

Δ	Euler	Milstein	Δ	Euler	Milstein
$1/2^1$	0.462%	0.391%	$1/2^{11}$	0.002%	0.040%
$1/2^2$	0.082%	0.071%	$1/2^{12}$	0.014%	0.021%
$1/2^3$	0.118%	0.514%	$1/2^{13}$	0.004%	0.029%
$1/2^4$	0.077%	0.089%	$1/2^{14}$	0.013%	0.001%
$1/2^5$	0.039%	0.082%	$1/2^{15}$	0.011%	0.004%
$1/2^6$	0.106%	0.253%	$1/2^{16}$	0.007%	0.003%
$1/2^7$	0.031%	0.149%	$1/2^{17}$	0.004%	0.006%
$1/2^8$	0.021%	0.071%	$1/2^{18}$	0.002%	0.003%
$1/2^9$	0.018%	0.038%	$1/2^{19}$	0.001%	0.001%
$1/2^{10}$	0.010%	0.094%	$1/2^{20}$	0.001%	0.000%

Table 4.27. Euler and Milstein Approximations to V of the HW Model at Endpoint

Δ	Euler	Milstein	Δ	Euler	Milstein
$1/2^1$	0.022157	0.006867	$1/2^{11}$	0.038465	0.010269
$1/2^2$	0.064295	0.005563	$1/2^{12}$	0.038312	0.010159
$1/2^3$	0.044484	0.009654	$1/2^{13}$	0.037888	0.010135
$1/2^4$	0.043486	0.009041	$1/2^{14}$	0.037592	0.010000
$1/2^5$	0.042191	0.010038	$1/2^{15}$	0.037847	0.009993
$1/2^6$	0.041837	0.008613	$1/2^{16}$	0.037708	0.009983
$1/2^7$	0.038289	0.009325	$1/2^{17}$	0.037759	0.009998
$1/2^8$	0.039479	0.009874	$1/2^{18}$	0.037759	0.010015
$1/2^9$	0.039758	0.010120	$1/2^{19}$	0.037789	0.010031
$1/2^{10}$	0.040025	0.010307	$1/2^{20}$	0.037812	0.010038

Figure 4.30. Strong Convergence of the Euler and Milstein Approximations to the HW Model

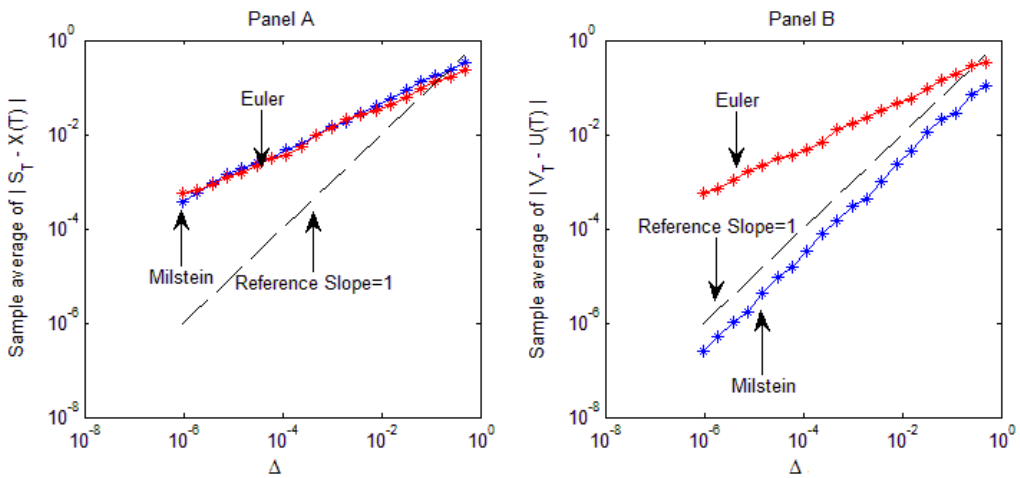
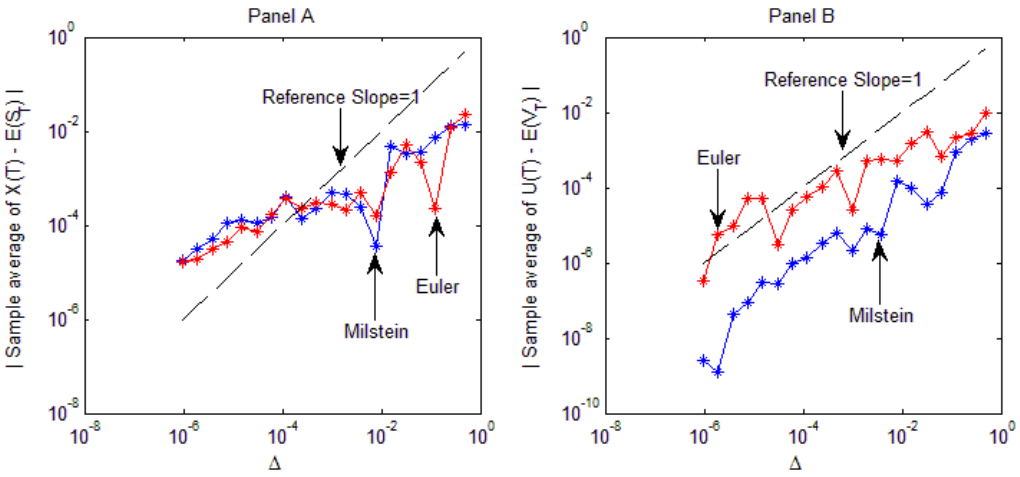


Table 4.28. Percentage Errors of the Euler and Milstein Approximations to V of the HW Model at Endpoint

Δ	Euler	Milstein	Δ	Euler	Milstein
$1/2^1$	31.610%	30.457%	$1/2^{11}$	2.270%	0.019%
$1/2^2$	44.597%	5.348%	$1/2^{12}$	1.174%	0.016%
$1/2^3$	3.855%	4.462%	$1/2^{13}$	0.934%	0.004%
$1/2^4$	9.960%	1.843%	$1/2^{14}$	0.412%	0.001%
$1/2^5$	0.029%	0.594%	$1/2^{15}$	0.480%	0.001%
$1/2^6$	14.216%	0.909%	$1/2^{16}$	0.579%	0.001%
$1/2^7$	7.132%	0.415%	$1/2^{17}$	0.429%	0.000%
$1/2^8$	1.664%	0.204%	$1/2^{18}$	0.257%	0.000%
$1/2^9$	0.785%	0.034%	$1/2^{19}$	0.100%	0.000%
$1/2^{10}$	2.674%	0.035%	$1/2^{20}$	0.026%	0.000%

Figure 4.31. Weak Convergence of the Euler and Milstein Approximations to the HW Model



The strong convergence of the Euler and Milstein approximations are shown is Figure 4.30 and the weak convergence is displayed in Figure 4.31. The strong convergence keeps smooth from $\Delta = 1/2^1$ to $\Delta = 1/2^{20}$. In terms of the weak convergence, $\Delta = 1/2^{15}$ seems to be the turning point of the Euler approximation to the state variable V. Overall, if a good pathwise approximation is required, it's better to set Δ as small as possible. While, if a good approximation of moments is needed, $\Delta = 1/2^{15}$ is probably a relatively good selection for the Euler approximation.

4.4.5. Hagan Model

Hagan, Kumar, and Lesniewski (2002) introduced the Hagan model, in which instantaneous volatility is a martingale but its variance grows unbounded. This model could be reduced to the HW model when $\mu_v = \alpha^2$ and $\sigma = 2\alpha$. Under the objective measure P, the joint dynamics of S_t and V_t is

$$d\begin{bmatrix} S_t \\ V_t \end{bmatrix} = \begin{bmatrix} \mu - \frac{1}{2}V_t^2 \\ 0 \end{bmatrix} dt + \begin{bmatrix} \sqrt{(1-\rho^2)}V_t & \rho V_t \\ 0 & \alpha V_t \end{bmatrix} d\begin{bmatrix} W_1^P(t) \\ W_2^P(t) \end{bmatrix},$$

where $W_1(t)$ and $W_2(t)$ are two uncorrelated standard Brownian motions.

The Euler approximations of S and V are continuous stochastic processes X and U satisfying the iterative schemes

$$\begin{aligned} X_{t+1} &= X_t + \left(\mu - \frac{1}{2}U_t^2\right) \cdot \Delta t + \sqrt{(1-\rho^2)}U_t \cdot \Delta W_1(t) + \rho U_t \Delta W_2(t) \\ U_{t+1} &= U_t + \alpha U_t \cdot \Delta W_2(t) \end{aligned}$$

with initial value $X_0 = S_0$, $U_0 = V_0$ for $t = 0, 1, \ldots N-1$.

The Milstein approximation of S and V are in the form of

$$\begin{aligned} Y_{t+1} &= Y_t + \left(\mu - \frac{1}{2}Z_t^2\right) \cdot \Delta t + \sqrt{(1-\rho^2)}Z_t \cdot \Delta W_1(t) + \rho Z_t \Delta W_2(t) \\ &\quad + \sigma\sqrt{(1-\rho^2)}Z_t I_{(2,1)} + \sigma\rho Z_t I_{(2,2)} \\ Z_{t+1} &= Z_t + \sigma Z_t \cdot \Delta W_2(t) + \frac{1}{2}\sigma^2 Z_t \left[(\Delta W_2(t))^2 - \Delta t\right] \end{aligned}$$

with initial value $Y_0 = S_0$, $Z_0 = V_0$ for $t = 0, 1, \ldots N-1$. We set $\mu = 1$, $\rho = -0.8$, $\alpha = 0.5$, with initial values $X_0 = 6.5$, $V_0 = 0.045$. Two Brownian bridges from 0 to 1 are used as $W_1(t)$ and $W_2(t)$ separately.

Figure 4.32 illustrates that both of the Euler and Milstein approximations to the state variables S and V seem to converge to a stable value at $\Delta = 1/2^{14}$. Tables 4.29, 4.30, 4.31 and 4.32 report the endpoint values with percentage errors for the state variable S and V separately. The percentage errors of the Euler approximations to the state variable S could be easily controlled around 0.01% even for big Δ. However, the percentage error of the Euler approximations to the state variable V could be controlled below 1.00% when $\Delta \leq 1/2^8$. The key to examine the accuracy of the Euler approximation to the Hagan model is mainly determined by the state variable V.

Figure 4.32. Convergence Speed of the Euler and Milstein Approximations to the Hagan Model at Endpoint

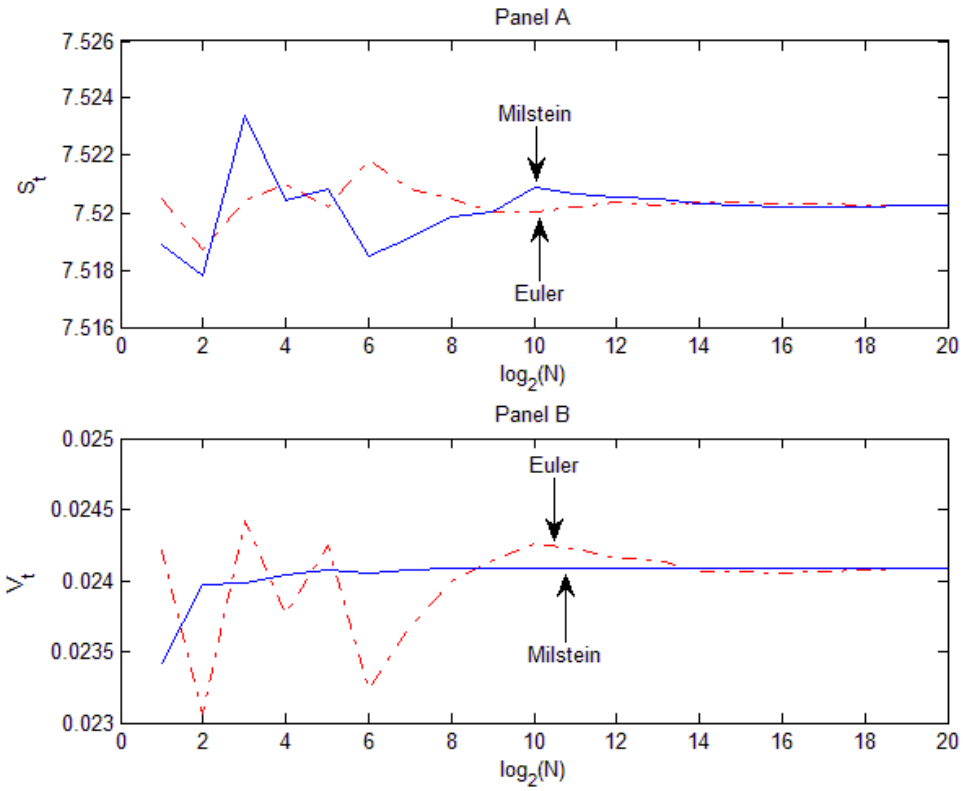


Table 4.29. Euler and Milstein Approximations to S of the Hagan Model at Endpoint

Δ	Euler	Milstein	Δ	Euler	Milstein
$1/2^1$	7.520494	7.518883	$1/2^{11}$	7.520207	7.520672
$1/2^2$	7.518730	7.517824	$1/2^{12}$	7.520348	7.520526
$1/2^3$	7.520409	7.523382	$1/2^{13}$	7.520252	7.520493
$1/2^4$	7.520981	7.520454	$1/2^{14}$	7.520390	7.520291
$1/2^5$	7.520191	7.520834	$1/2^{15}$	7.520369	7.520240
$1/2^6$	7.521874	7.518475	$1/2^{16}$	7.520334	7.520216
$1/2^7$	7.520824	7.519103	$1/2^{17}$	7.520292	7.520196
$1/2^8$	7.520497	7.519879	$1/2^{18}$	7.520277	7.520218
$1/2^9$	7.520042	7.520055	$1/2^{19}$	7.520259	7.520236
$1/2^{10}$	7.520054	7.520896	$1/2^{20}$	7.520243	7.520241

Table 4.30. Percentage Errors of the Euler and Milstein Approximations to S of the Hagan Model at Endpoint

Δ	Euler	Milstein	Δ	Euler	Milstein
$1/2^1$	0.003%	0.018%	$1/2^{11}$	0.000%	0.006%
$1/2^2$	0.020%	0.032%	$1/2^{12}$	0.001%	0.004%
$1/2^3$	0.002%	0.042%	$1/2^{13}$	0.000%	0.003%
$1/2^4$	0.010%	0.003%	$1/2^{14}$	0.002%	0.001%
$1/2^5$	0.001%	0.008%	$1/2^{15}$	0.002%	0.000%
$1/2^6$	0.022%	0.023%	$1/2^{16}$	0.001%	0.000%
$1/2^7$	0.008%	0.015%	$1/2^{17}$	0.001%	0.001%
$1/2^8$	0.003%	0.005%	$1/2^{18}$	0.000%	0.000%
$1/2^9$	0.003%	0.002%	$1/2^{19}$	0.000%	0.000%
$1/2^{10}$	0.002%	0.009%	$1/2^{20}$	0.001%	0.000%

Table 4.31. Euler and Milstein Approximations to V of the Hagan Model at Endpoint

Δ	Euler	Milstein	Δ	Euler	Milstein
$1/2^1$	0.024217	0.023425	$1/2^{11}$	0.024224	0.024086
$1/2^2$	0.023065	0.023974	$1/2^{12}$	0.024158	0.024086
$1/2^3$	0.024422	0.023980	$1/2^{13}$	0.024143	0.024087
$1/2^4$	0.023769	0.024045	$1/2^{14}$	0.024062	0.024087
$1/2^5$	0.024245	0.024074	$1/2^{15}$	0.024058	0.024087
$1/2^6$	0.023241	0.024057	$1/2^{16}$	0.024052	0.024087
$1/2^7$	0.023673	0.024074	$1/2^{17}$	0.024061	0.024087
$1/2^8$	0.023999	0.024081	$1/2^{18}$	0.024071	0.024087
$1/2^9$	0.024148	0.024088	$1/2^{19}$	0.024081	0.024087
$1/2^{10}$	0.024253	0.024088	$1/2^{20}$	0.024085	0.024087

Figure 4.33. Strong Convergence of the Euler and Milstein Approximations to the Hagan Model

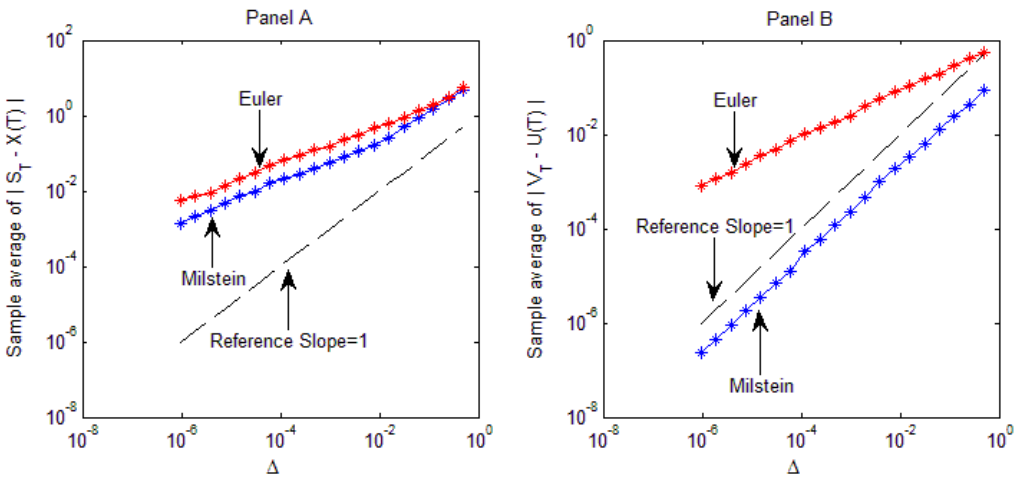
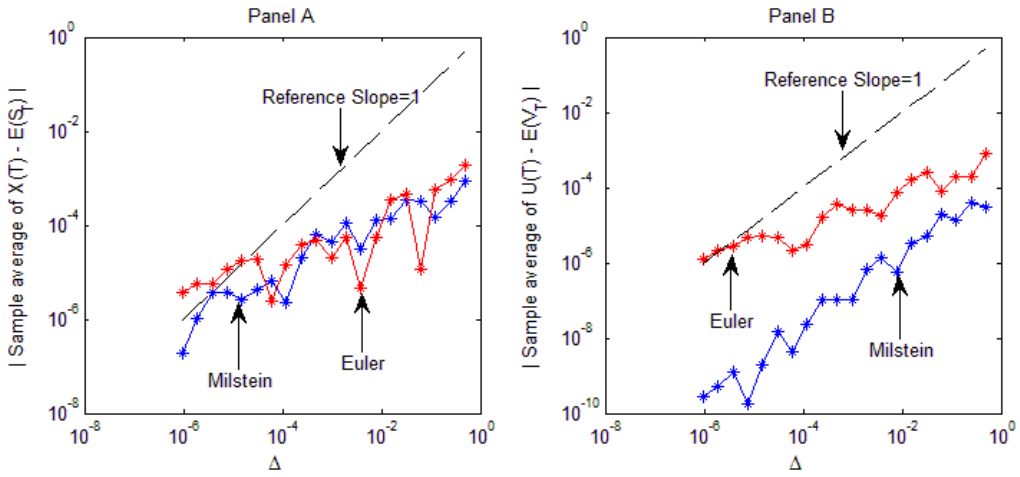


Table 4.32. Percentage Errors of the Euler and Milstein Approximations to V of the Hagan Model at Endpoint

Δ	Euler	Milstein	Δ	Euler	Milstein
$1/2^1$	0.540%	2.748%	$1/2^{11}$	0.571%	0.002%
$1/2^2$	4.242%	0.467%	$1/2^{12}$	0.295%	0.002%
$1/2^3$	1.390%	0.441%	$1/2^{13}$	0.235%	0.000%
$1/2^4$	1.321%	0.173%	$1/2^{14}$	0.101%	0.000%
$1/2^5$	0.658%	0.051%	$1/2^{15}$	0.120%	0.000%
$1/2^6$	3.513%	0.125%	$1/2^{16}$	0.145%	0.000%
$1/2^7$	1.717%	0.053%	$1/2^{17}$	0.107%	0.000%
$1/2^8$	0.365%	0.025%	$1/2^{18}$	0.064%	0.000%
$1/2^9$	0.256%	0.004%	$1/2^{19}$	0.025%	0.000%
$1/2^{10}$	0.690%	0.004%	$1/2^{20}$	0.006%	0.000%

Figure 4.34. Weak Convergence of the Euler and Milstein Approximations to the Hagan Model



Figures 4.33 and 4.34 investigate the strong and weak convergence of the Euler and Milstein approximation. Figure 4.33 shows the mean absolute different values keep declining and the strong convergence of the state variable S and V are very close to their theoretical values. However, based on the weak convergence, $\Delta = 1/2^{14}$ seems to be the turning point of the Euler approximation to the state variable V. To sum up, $\Delta = 1/2^{14}$ could be an effective time step for the Hagan model by considering moments approximation requirements.

4.4.6. SABR Volatility Model

The name SABR stands for "stochastic alpha, beta, rho" referring to the parameters of the model. It original introduced by Hagan, Kumar, and Lesniewski (2002) and is widely used by practitioners in the interest rate derivative markets. In the SABR stochastic volatility model, the joint dynamics of S_t and Y_t is defined as

$$d\begin{bmatrix} S_t \\ V_t \end{bmatrix} = \begin{bmatrix} \frac{1}{2}V_t^2 S_t^{2\beta-2} \\ 0 \end{bmatrix} dt + \begin{bmatrix} \sqrt{(1-\rho^2)}V_t S_t^{\beta-1} & \rho V_t S_t^{\beta-1} \\ 0 & \alpha V_t \end{bmatrix} d\begin{bmatrix} W_1^P(t) \\ W_2^P(t) \end{bmatrix}$$

under the objective measure P, where $W_1(t)$ and $W_2(t)$ are two uncorrelated standard Brownian motions. The Euler approximations of S and V are continuous stochastic processes X and U satisfying the iterative schemes

$$\begin{aligned} X_{t+1} &= X_t + \left(\frac{1}{2}U_t^2 X_t^{2\beta-2}\right)\cdot\Delta t + \sqrt{(1-\rho^2)}U_t X_t^{\beta-1}\cdot\Delta W_1(t) + \rho U_t X_t^{\beta-1}\Delta W_2(t) \\ U_{t+1} &= U_t + \alpha U_t \cdot \Delta W_2(t) \end{aligned}$$

with initial value $X_0 = S_0$, $U_0 = V_0$ for $t = 0, 1, \ldots N-1$. The Milstein approximations of S and V are in the form of

$$\begin{aligned} Y_{t+1} &= Y_t + \left(-\frac{1}{2}Z_t^2 Y_t^{2\beta-2}\right)\cdot\Delta t + \sqrt{1-\rho^2}Z_t Y_t^{\beta-1}\cdot\Delta W_1(t) + \rho Z_t Y_t^{\beta-1}\Delta W_2(t) \\ &\quad + \left[(\beta-1)\left(1-\rho^2\right)Z_t^2 Y_t^{2\beta-3} + \rho\sqrt{1-\rho^2}Z_t Y_t^{2\beta-2}\right]\cdot\left[(\Delta W_1(t))^2 - \Delta t\right] \\ &\quad + \left[\rho^2(\beta-1)Z_t^2 Y_t^{2\beta-3} + \alpha\rho Z_t Y_t^{\beta-1}\right]\cdot\left[(\Delta W_2(t))^2 - \Delta t\right] \\ &\quad + \left[\left(1-\rho^2\right)Z_t^2(\beta-1)Y_t^{2\beta-3} + \alpha\sqrt{1-\rho^2}Z_t Y_t^{\beta-1}\right]I_{(2,1)} \\ &\quad + \rho\sqrt{1-\rho^2}(\beta-1)Z_t^2 Y_t^{2\beta-3} I_{(1,2)} \\ Z_{t+1} &= Z_t + \alpha Z_t\cdot\Delta W_2(t) + \frac{1}{2}\alpha^2 Z_t\left[(\Delta W_2(t))^2 - \Delta t\right] \end{aligned}$$

with initial value $Y_0 = S_0$, $Z_0 = V_0$ for $t = 0, 1, \ldots N-1$. We set $\rho = -0.415$, $\alpha = 0.514$, $\beta = 0.5$ with initial values $X_0 = 6.5$, $V_0 = 0.045$. Two Brownian bridges from 0 to 1 are used as $W_1(t)$ and $W_2(t)$ separately. Panel A of Figure 4.35 displays that the Euler and Milstein approximations to the state variables S at the endpoint coincide when $\Delta = 1/2^{14}$. On the other hand, in Panel B of Figure 4.35, the Euler and Milstein approximations to the state variables V overlaps at $\Delta = 1/2^8$ and converges to a stable value at $\Delta = 1/2^{14}$. Tables 4.33, 4.34, 4.35 and 4.36 report the endpoint values with percentage errors for the state variable S and V separately.

Figure 4.35. Convergence Speed of the Euler and Milstein Approximations to the SABR Model at Endpoint

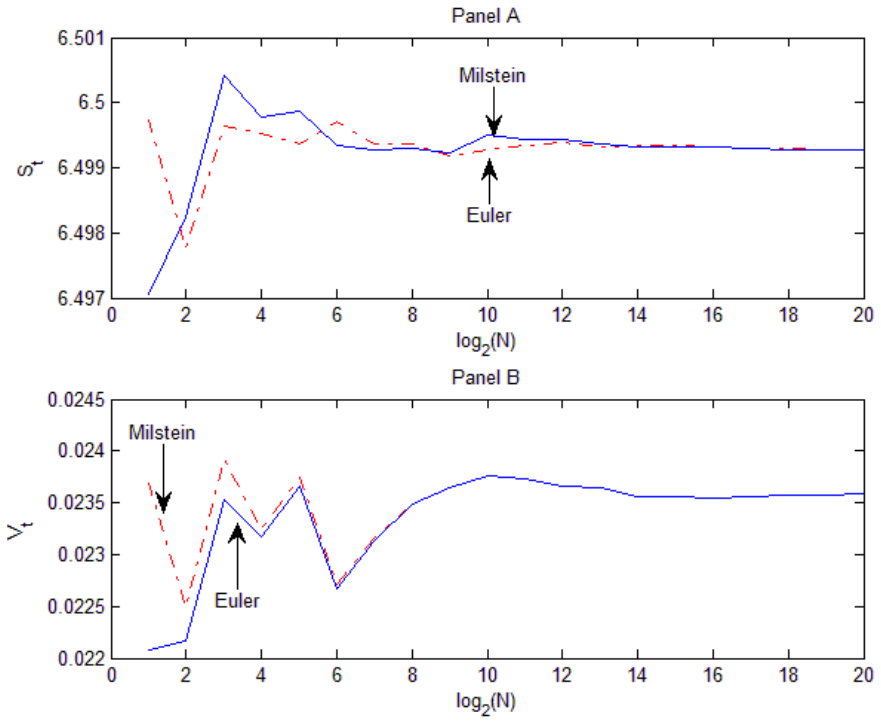


Table 4.33. Euler and Milstein Approximations to S of the SABR Model at Endpoint

Δ	Euler	Milstein	Δ	Euler	Milstein
$1/2^1$	6.499724	6.497063	$1/2^{11}$	6.499342	6.499431
$1/2^2$	6.497770	6.498237	$1/2^{12}$	6.499385	6.499426
$1/2^3$	6.499638	6.500417	$1/2^{13}$	6.499319	6.499371
$1/2^4$	6.499530	6.499773	$1/2^{14}$	6.499351	6.499323
$1/2^5$	6.499358	6.499864	$1/2^{15}$	6.499334	6.499321
$1/2^6$	6.499716	6.499340	$1/2^{16}$	6.499310	6.499311
$1/2^7$	6.499361	6.499268	$1/2^{17}$	6.499290	6.499290
$1/2^8$	6.499374	6.499287	$1/2^{18}$	6.499288	6.499278
$1/2^9$	6.499194	6.499227	$1/2^{19}$	6.499283	6.499280
$1/2^{10}$	6.499268	6.499510	$1/2^{20}$	6.499276	6.499277

Table 4.34. Percentage Errors of the Euler and Milstein Approximations to S of the SABR Model at Endpoint

Δ	Euler	Milstein	Δ	Euler	Milstein
$1/2^1$	0.007%	0.034%	$1/2^{11}$	0.001%	0.002%
$1/2^2$	0.023%	0.016%	$1/2^{12}$	0.002%	0.002%
$1/2^3$	0.005%	0.018%	$1/2^{13}$	0.000%	0.001%
$1/2^4$	0.004%	0.008%	$1/2^{14}$	0.001%	0.001%
$1/2^5$	0.001%	0.009%	$1/2^{15}$	0.001%	0.001%
$1/2^6$	0.007%	0.001%	$1/2^{16}$	0.001%	0.001%
$1/2^7$	0.001%	0.000%	$1/2^{17}$	0.000%	0.000%
$1/2^8$	0.001%	0.000%	$1/2^{18}$	0.000%	0.000%
$1/2^9$	0.001%	0.000%	$1/2^{19}$	0.000%	0.000%
$1/2^{10}$	0.000%	0.004%	$1/2^{20}$	0.000%	0.000%

Table 4.35. Euler and Milstein Approximations to V of the SABR Model at Endpoint

Δ	Euler	Milstein	Δ	Euler	Milstein
$1/2^1$	0.023684	0.022075	$1/2^{11}$	0.023726	0.023735
$1/2^2$	0.022492	0.022166	$1/2^{12}$	0.023657	0.023658
$1/2^3$	0.023919	0.023526	$1/2^{13}$	0.022364	0.023643
$1/2^4$	0.023248	0.023169	$1/2^{14}$	0.023559	0.023559
$1/2^5$	0.023744	0.023656	$1/2^{15}$	0.023554	0.023554
$1/2^6$	0.022708	0.022670	$1/2^{16}$	0.023548	0.023548
$1/2^7$	0.023156	0.023132	$1/2^{17}$	0.023557	0.023557
$1/2^8$	0.023493	0.023481	$1/2^{18}$	0.023568	0.023568
$1/2^9$	0.023647	0.023643	$1/2^{19}$	0.023578	0.023578
$1/2^{10}$	0.023756	0.023754	$1/2^{20}$	0.023582	0.023582

Figure 4.36. Strong Convergence of the Euler and Milstein Approximations to the SABR Model

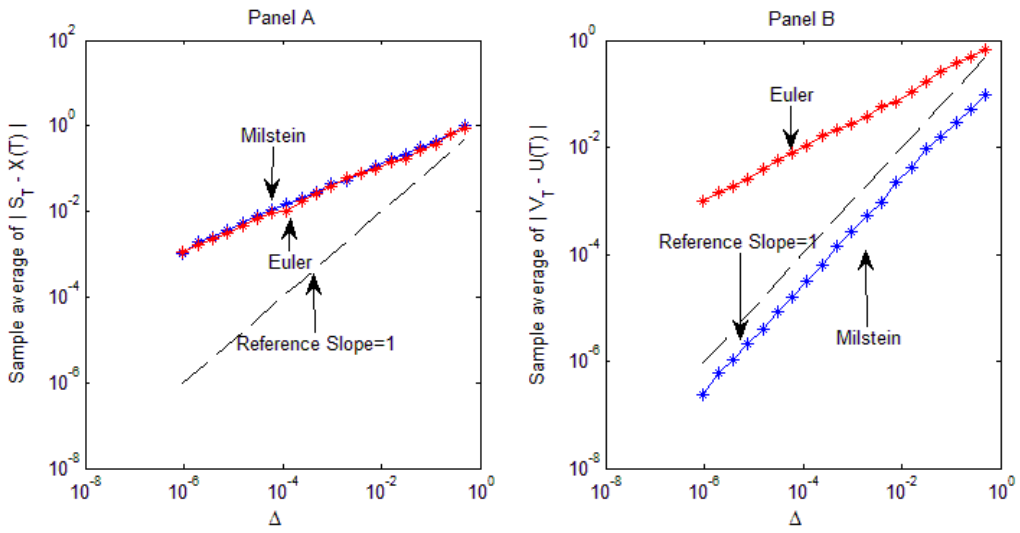
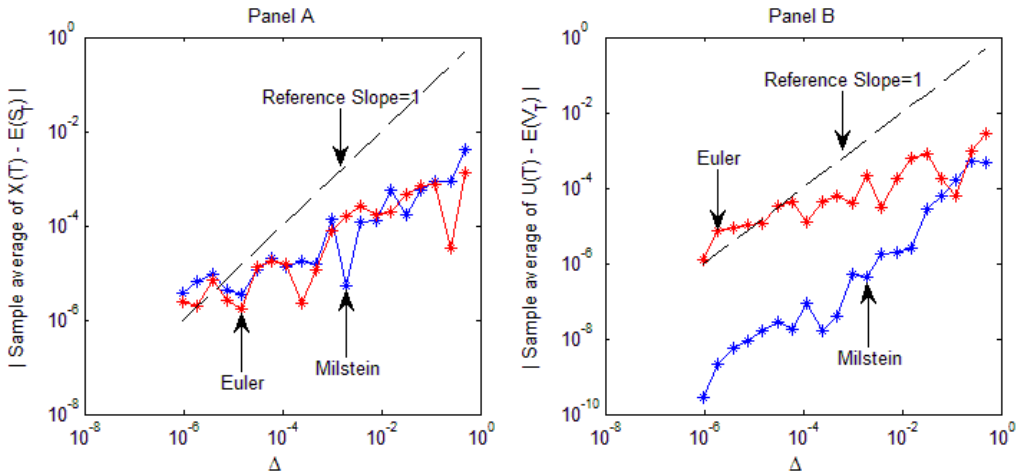


Table 4.36. Percentage Errors of the Euler and Milstein Approximations to V of the SABR Model at Endpoint

Δ	Euler	Milstein	Δ	Euler	Milstein
$1/2^1$	0.432%	6.392%	$1/2^{11}$	0.610%	0.604%
$1/2^2$	4.625%	6.004%	$1/2^{12}$	0.318%	0.319%
$1/2^3$	1.426%	0.240%	$1/2^{13}$	0.255%	0.257%
$1/2^4$	1.417%	1.752%	$1/2^{14}$	0.100%	0.101%
$1/2^5$	0.686%	0.314%	$1/2^{15}$	0.120%	0.120%
$1/2^6$	3.706%	3.868%	$1/2^{16}$	0.146%	0.146%
$1/2^7$	1.810%	1.911%	$1/2^{17}$	0.107%	0.107%
$1/2^8$	0.380%	0.432%	$1/2^{18}$	0.061%	0.061%
$1/2^9$	0.276%	0.260%	$1/2^{19}$	0.020%	0.020%
$1/2^{10}$	0.735%	0.727%	$1/2^{20}$	0.000%	0.000%

Figure 4.37. Weak Convergence of the Euler and Milstein Approximations to the SABR Model



The percentage errors of the Euler approximations to the state variable S could be easily controlled under 0.10%, even 0.01% for big time steps Δ. On the contrary, the percentage errors of the Euler approximations to the state variable V fluctuate with different time interval Δ and could be controlled under 1.00% until $\Delta = 1/2^8$. Undoubtedly, the key to examine the accuracy of the Euler approximation to the SABR model mainly depends on the variable V.

Figures 4.36 and 4.37 illustrate the strong and weak convergence of the Euler and Milstein approximation. Figure 4.36 displays that the order of the strong convergence of the state variable S and V are very close to their theoretical values. However, based on the weak convergence, $\Delta = 1/2^{13}$ seems to be the turning point for the Euler

approximation to the state variable V. To conclude, $\Delta = 1/2^{12}$ could be an effective time step for the SABR model if a good approximation of moments is needed.

4.4.7. Stein Model

The Stein Model in proposed by Stein and Stein (1991) and Schöbel and Zhu (1999), in which the joint dynamics of S_t and Y_t is in the form of

$$d\begin{bmatrix} S_t \\ V_t \end{bmatrix} = \begin{bmatrix} \mu - \frac{1}{2}V_t^2 \\ \kappa(\theta - V_t) \end{bmatrix} dt + \begin{bmatrix} \sqrt{(1-\rho^2)}V_t & \rho V_t \\ 0 & \sigma \end{bmatrix} d\begin{bmatrix} W_1^P(t) \\ W_2^P(t) \end{bmatrix}$$

under the objective measure P, where $W_1(t)$ and $W_2(t)$ are two uncorrelated standard Brownian motions. The Euler approximations of S and V are continuous stochastic processes X and U satisfying the iterative schemes

$$\begin{aligned} X_{t+1} &= X_t + \left(\mu - \frac{1}{2}U_t^2\right) \cdot \Delta t + \sqrt{(1-\rho^2)}U_t \cdot \Delta W_1(t) + \rho U_t \Delta W_2(t) \\ U_{t+1} &= U_t + k(\theta - U_t) \cdot \Delta t + \sigma \cdot \Delta W_2(t) \end{aligned}$$

with initial value $X_0 = S_0$, $U_0 = V_0$ for $t = 0, 1, \ldots N-1$. Because the partial derivative of σ with respect to x is equal to 0, the Milstein approximation reduces to the Euler approximation. We take a Stein model with $\rho = -0.8$, $\mu = 0.25$, $\kappa = 3.5$, $\theta = 0.25$, $\sigma = 0.40$ with initial values $X_0 = 6.5$, $V_0 = 0.045$ as an example. Two Brownian bridges from 0 to 1 are used as $W_1(t)$ and $W_2(t)$ separately. In the light of above multivariate models, the Euler and Milstein approximations to the state variables S and V coverge to a stable value when Δ goes to 0. Therefore, our numerical tests resort to Figure 4.38 of the convergence speed at the endpoint. Panel B shows, from $\Delta = 1/2^7$, the Euler (Milstein) approximation to the state variable V reaches a stable value of 0.269129, which could be used as a reference solution of V. In Panel A, the Euler (Milstein) approximation to the state variable S goes to a stable value of 6.837018, which could be used as a benchmark of S. Then Tables 4.37 and 4.38 report the Euler (Milstein) approximations to the state variables S and V with their percentage errors at the endpoint. When the time stepsize Δ is set at $1/2^3$, the percentage error of the Euler (Milstein) approximation to S could be below 1.00%. In comparision, the percentage error of the Euler (Milstein) approximation to V could be controlled under 1.00% when the time stepsize Δ is set at smaller than

$1/2^7$. Because of this, the key to examine the accuracy of the Euler approximation to the Stein model still depends on the variable V.

Figure 4.38. Convergence Speed of Euler (Milstein) Approximation to the Stein Model at Endpoint

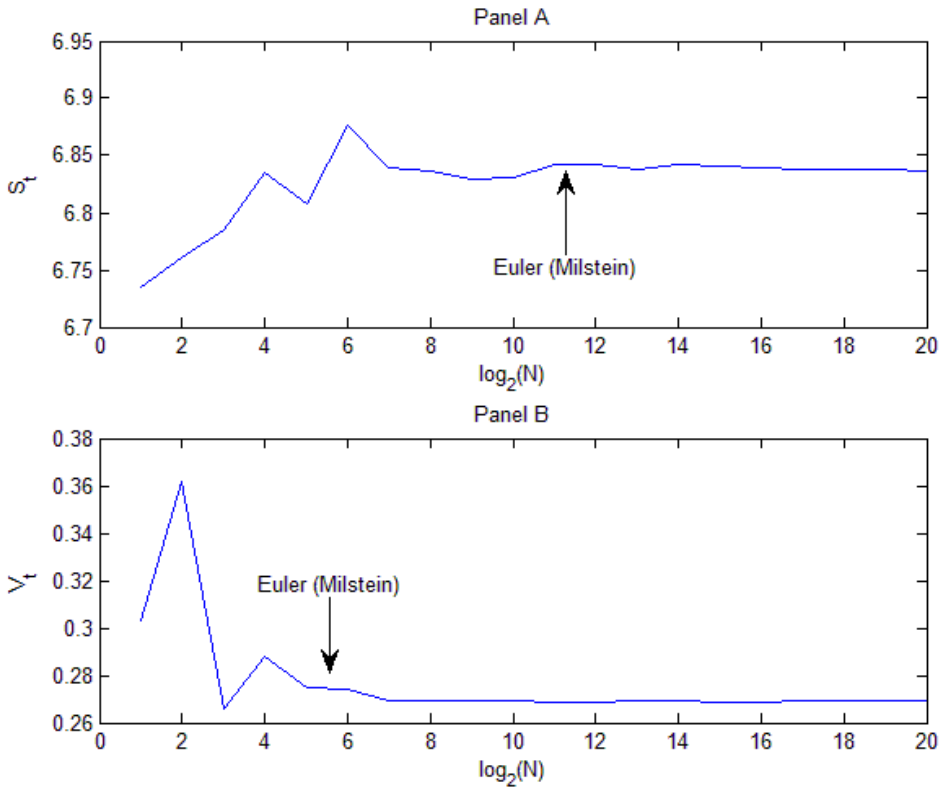


Table 4.37. Euler (Milstein) Approximation and Percentage Errors for S of the Stein Model at Endpoint

Δ	Euler (Milstein)	PE	Δ	Euler (Milstein)	PE
$1/2^1$	6.734333	1.502%	$1/2^{11}$	6.841704	0.069%
$1/2^2$	6.760353	1.121%	$1/2^{12}$	6.842354	0.078%
$1/2^3$	6.785546	0.753%	$1/2^{13}$	6.837188	0.002%
$1/2^4$	6.835201	0.027%	$1/2^{14}$	6.841598	0.067%
$1/2^5$	6.808187	0.422%	$1/2^{15}$	6.840925	0.057%
$1/2^6$	6.876405	0.576%	$1/2^{16}$	6.839561	0.037%
$1/2^7$	6.839542	0.037%	$1/2^{17}$	6.837952	0.014%
$1/2^8$	6.836660	0.005%	$1/2^{18}$	6.837610	0.009%
$1/2^9$	6.829615	0.108%	$1/2^{19}$	6.837431	0.006%
$1/2^{10}$	6.831336	0.083%	$1/2^{20}$	6.837018	0.000%

Table 4.38. Euler (Milstein) Approximation and Percentage Errors for V of the Stein Model at Endpoint

Δ	Euler (Milstein)	PE	Δ	Euler (Milstein)	PE
$1/2^1$	0.303152	12.642%	$1/2^{11}$	0.268674	0.169%
$1/2^2$	0.361950	34.489%	$1/2^{12}$	0.268541	0.219%
$1/2^3$	0.265782	1.244%	$1/2^{13}$	0.268970	0.059%
$1/2^4$	0.287744	6.917%	$1/2^{14}$	0.268974	0.058%
$1/2^5$	0.275001	2.182%	$1/2^{15}$	0.268846	0.105%
$1/2^6$	0.274070	1.836%	$1/2^{16}$	0.268853	0.103%
$1/2^7$	0.269179	0.018%	$1/2^{17}$	0.269256	0.047%
$1/2^8$	0.269226	0.036%	$1/2^{18}$	0.269182	0.019%
$1/2^9$	0.269615	0.180%	$1/2^{19}$	0.269138	0.003%
$1/2^{10}$	0.269011	0.044%	$1/2^{20}$	0.269129	0.000%

Figure 4.39. Strong Convergence of the Euler (Milstein) Approximation to the Stein Model

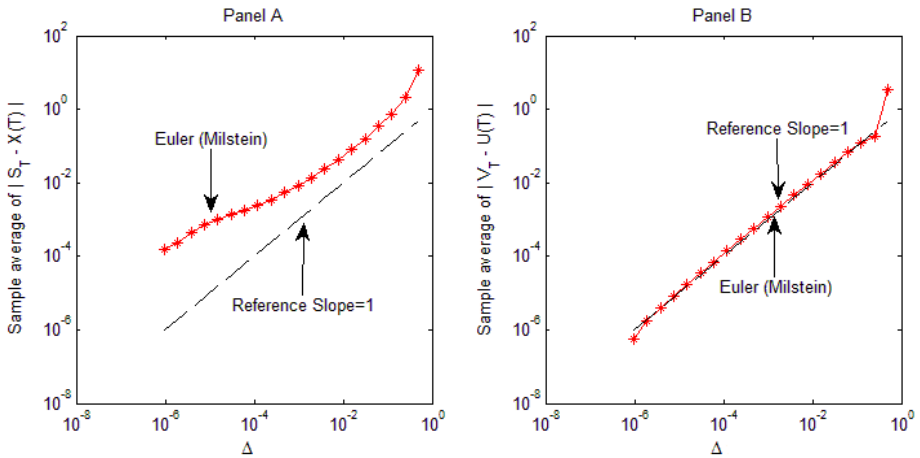
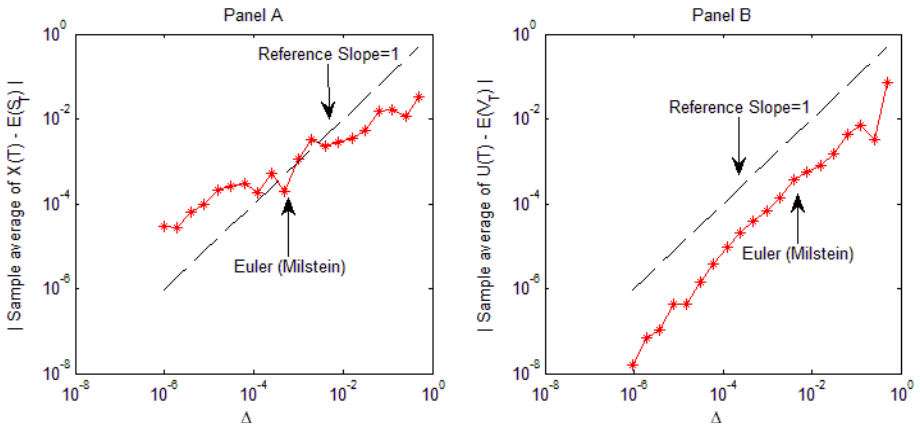


Figure 4.40. Weak Convergence of the Euler (Milstein) Approximation to the Stein Model



The strong convergence of the Euler (Milstein) approximations to the state variables S and V are shown is Figure 4.39 and the weak convergence are displayed in Figure 4.40. The strong convergence of the state variable V keeps smooth from $\Delta = 1/2^1$ to $\Delta = 1/2^{20}$, while the strong convergence of the state variable S departs slightly from its theoretical value from $\Delta = 1/2^9$. In terms of the weak convergence, there seems no turning point for the Euler (Milstein) approximations to the state variables V. Even so, moving to the state variables V, the turning point appears at $\Delta = 1/2^{11}$. To sum up, if a good pathwise approximation is required, the accuracy of the Euler (Milstein) approximation seems to be improved as Δ goes to 0. However, if a good approximation of moments is needed, $\Delta = 1/2^{11}$ is probably a nice option for the Euler (Milstein) approximation.

4.4.8. Scott Model

The Scott Model, proposed by Scott (1987) assumes the the joint dynamics of S_t and Y_t is in the form of

$$d\begin{bmatrix} S_t \\ V_t \end{bmatrix} = \begin{bmatrix} \mu - \frac{1}{2}e^{2V_t} \\ \kappa(\theta - V_t) \end{bmatrix} dt + \begin{bmatrix} \sqrt{(1-\rho^2)}e^{V_t} & \rho e^{V_t} \\ 0 & \sigma \end{bmatrix} d\begin{bmatrix} W_1^P(t) \\ W_2^P(t) \end{bmatrix}$$

under the objective measure P, where $W_1(t)$ and $W_2(t)$ are two uncorrelated standard Brownian motions. The Euler approximations of S and V are continuous stochastic processes X and U satisfying the iterative schemes

$$\begin{aligned} X_{t+1} &= X_t + \left(\mu - \frac{1}{2}e^{2U_t}\right) \cdot \Delta t + \sqrt{(1-\rho^2)}e^{U_t} \cdot \Delta W_1(t) + \rho e^{U_t} \Delta W_2(t) \\ U_{t+1} &= U_t + k(\theta - U_t) \cdot \Delta t + \sigma \cdot \Delta W_2(t) \end{aligned}$$

with initial value $X_0 = S_0$, $U_0 = V_0$ for $t = 0, 1, \ldots N-1$. Because the partial derivative of σ with respect to x is equal to 0, the Milstein approximation reduces to the Euler approximation. We take a Scott model with $\rho = -0.8$, $\mu = 6$, $\kappa = 0.5$, $\theta = 0.25$, $\sigma = 0.05$ with initial values $X_0 = 6.5$, $V_0 = 0.045$ as an example. Two Brownian bridges from 0 to 1 are used as $W_1(t)$ and $W_2(t)$ separately. Similar to the Stein model, our numerical tests resort to Figure 4.41 of the convergence speed at the endpoint to find the benchmarks for the Euler (Milstein) approximations to the state variable S and V. Panel B shows, from $\Delta = 1/2^8$, the Euler (Milstein) approximation to the state variable V reaches a stable value of 0.094310, which could be used as a

reference solution of V. In Panel A, the Euler (Milstein) approximation to the state variable S goes to a stable value of 12.157735, which could be used as a benchmark of S. Then Tables 4.39 and 4.40 report the Euler (Milstein) approximations to the state variables S and V with their percentage errors at the endpoint. The percentage error of the Euler (Milstein) approximation to S could be easily below 0.10%. In comparision, the percentage error of the Euler (Milstein) approximation to V could be controlled under 0.10% when the time stepsize Δ is set at smaller than $1/2^8$. Because of this, the key to examine the accuracy of the Euler approximation to the Scott model depends on the variable V as well.

Figure 4.41. Convergence Speed of the Euler (Milstein) Approximation to the Scott Model at Endpoint

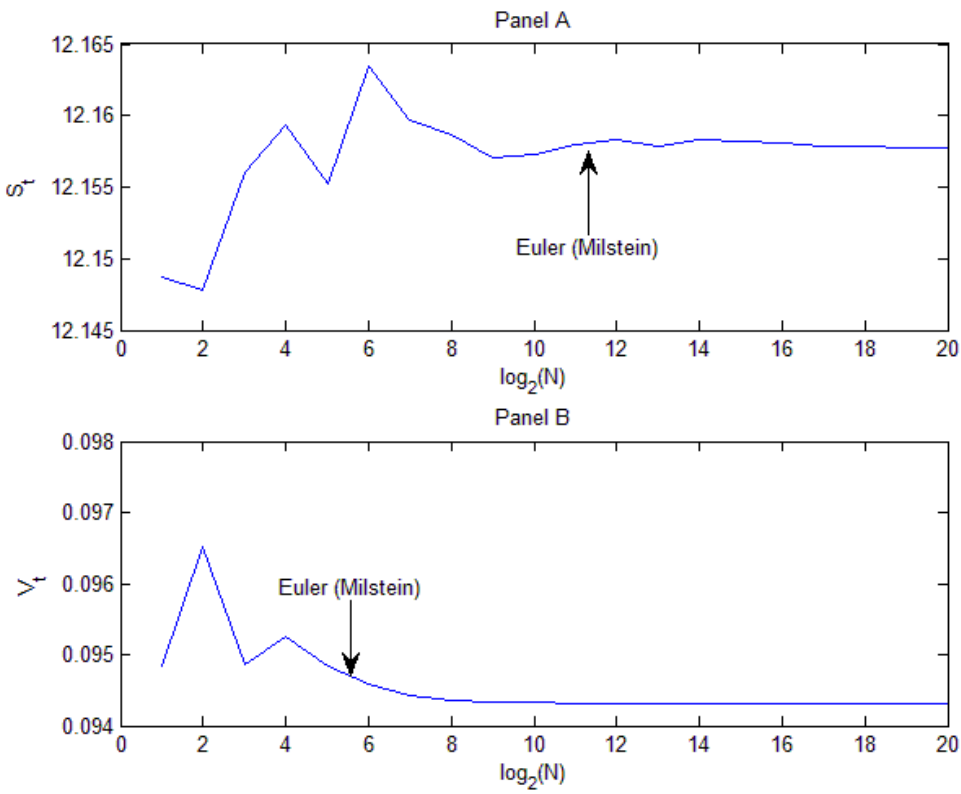


Table 4.39. Euler (Milstein) Approximation and Percentage Errors for S of the Scott Model at Endpoint

Δ	Euler (Milstein)	PE	Δ	Euler (Milstein)	PE
$1/2^1$	12.148770	0.074%	$1/2^{11}$	12.157938	0.002%
$1/2^2$	12.147759	0.082%	$1/2^{12}$	12.158262	0.004%
$1/2^3$	12.155987	0.014%	$1/2^{13}$	12.157837	0.001%
$1/2^4$	12.159303	0.013%	$1/2^{14}$	12.158325	0.005%
$1/2^5$	12.155264	0.020%	$1/2^{15}$	12.158192	0.004%
$1/2^6$	12.163418	0.047%	$1/2^{16}$	12.158049	0.003%
$1/2^7$	12.159739	0.016%	$1/2^{17}$	12.157896	0.001%
$1/2^8$	12.158620	0.007%	$1/2^{18}$	12.157849	0.001%
$1/2^9$	12.157107	0.005%	$1/2^{19}$	12.157791	0.000%
$1/2^{10}$	12.157241	0.004%	$1/2^{20}$	12.157735	0.000%

Table 4.40. Euler (Milstein) Approximation and Percentage Errors for V of the Scott Model at Endpoint

Δ	Euler (Milstein)	PE	Δ	Euler (Milstein)	PE
$1/2^1$	0.094839	0.561%	$1/2^{11}$	0.094318	0.008%
$1/2^2$	0.096526	2.350%	$1/2^{12}$	0.094312	0.002%
$1/2^3$	0.094862	0.586%	$1/2^{13}$	0.094311	0.001%
$1/2^4$	0.095247	0.994%	$1/2^{14}$	0.094310	0.000%
$1/2^5$	0.094839	0.561%	$1/2^{15}$	0.094310	0.000%
$1/2^6$	0.094578	0.285%	$1/2^{16}$	0.094310	0.000%
$1/2^7$	0.094414	0.110%	$1/2^{17}$	0.094310	0.000%
$1/2^8$	0.094361	0.054%	$1/2^{18}$	0.094310	0.000%
$1/2^9$	0.094341	0.033%	$1/2^{19}$	0.094310	0.000%
$1/2^{10}$	0.094327	0.018%	$1/2^{20}$	0.094310	0.000%

Figure 4.42. Strong Convergence of the Euler (Milstein) Approximation to the Scott Model

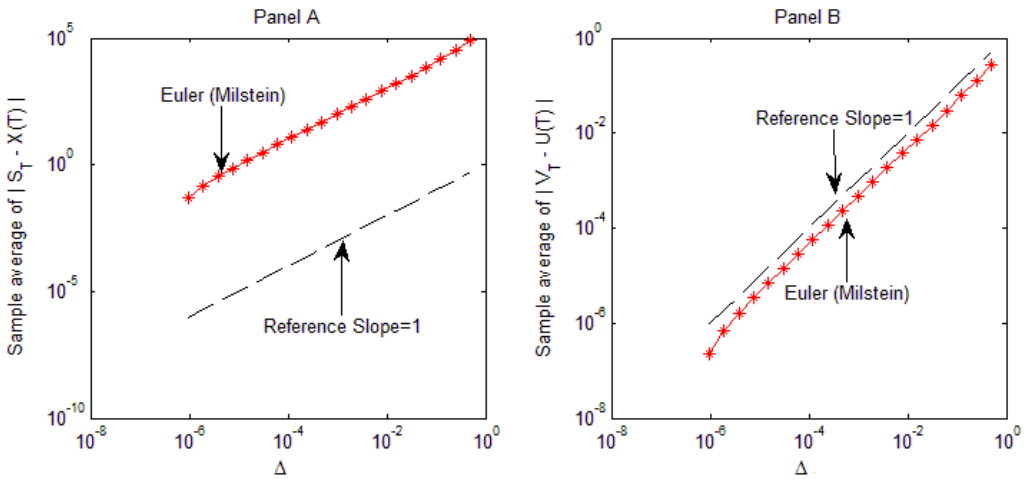
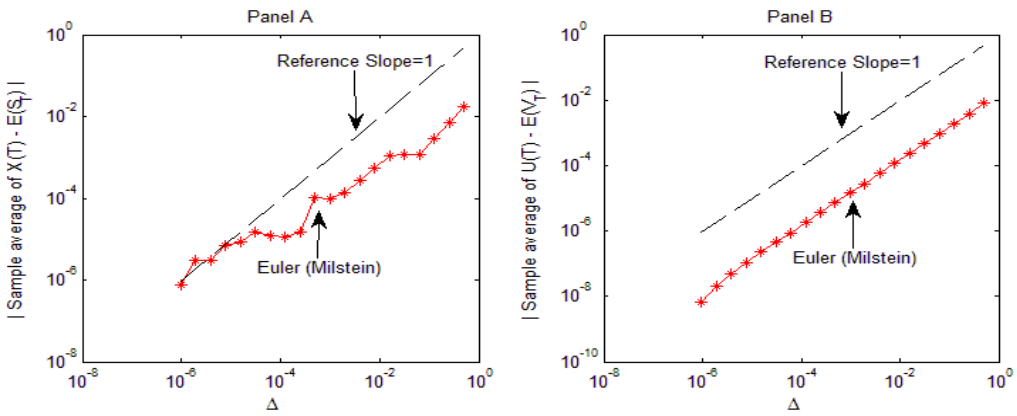


Figure 4.43. Weak Convergence of the Euler (Milstein) Approximation to the Scott Model



The strong convergence of the Euler (Milstein) approximations to the state variables S and V are shown is Figure 4.42 and the weak convergence is displayed in Figure 4.43. The strong convergence of the state variable V and S show their theoretical value. In terms of the weak convergence, obviously, there are no turning points for the Euler (Milstein) approximations to the state variables S and V. Therefore, the time step size Δ is probably set as small as possible by considering either the pathwise approximation or the moments approximation requirement.

4.5. Conclusions

This chapter compares the accuracy of the Euler approximation to a variety of univariate and multivariate diffusion processes for different time intervals by checking different criteria. The percentage error and strong convergence are examined when a good approximation of sample path of a diffusion model is required. The weak convergence is preferred for the cases where approximation of moments of the process matters. Besides these, our Monte Carlo simulations investigate convergence speed of the end point for different time intervals as well. Tables 4.41 and 4.42 provide guidelines for univariate and multivariate diffusion models, respectively.

Normally, when the objective is to simulate a diffusion process or test statistical estimators, our numerical tests suggest an appropriate time interval should be set in order to satisfy different error requirements. Taking the first row of Table 4.41 as an example, the time step Δ could be set at $1/2^4$ at the longest for the Euler approximation to the BSM model to meet the error requirement of less than 1.00%.

Table 4.41. Guidelines on the Choice of the Discretization Interval for Univariate Diffusion Models

	PE≤ 1.00%	PE≤ 0.10%	PE≤ 0.01%	Turning Point
BSM	$1/2^4$	$1/2^{16}$	$1/2^{19}$	$1/2^9$
Vasicek	$1/2^3$	$1/2^7$	$1/2^{10}$	–
CIR	$1/2^2$	$1/2^{11}$	$1/2^{17}$	$1/2^{11}$
IFSR	–	$1/2^3$	$1/2^{12}$	$1/2^{12}$
LDCEV	–	$1/2^8$	$1/2^{14}$	$1/2^{14}$
NMR	$1/2^3$	$1/2^{10}$	$1/2^{15}$	$1/2^{13}$

Table 4.42. Guidelines on the Choice of the Discretization Interval for Multivariate Diffusion Models

		PE≤ 1.00%	PE≤ 0.10%	PE≤ 0.01%	Turning Point
Heston	S	–	$1/2^5$	$1/2^{18}$	$1/2^{14}$
	V	$1/2^{11}$	$1/2^{19}$	–	$1/2^{11}$
GARCH	S	$1/2^{13}$	$1/2^{18}$	$1/2^{20}$	$1/2^{14}$
	V	$1/2^{20}$	–	–	$1/2^{13}$
CEV	S	–	$1/2^{10}$	$1/2^{18}$	$1/2^{16}$
	V	$1/2^{13}$	$1/2^{20}$	–	$1/2^{19}$
HW	S	–	$1/2^7$	$1/2^{16}$	$1/2^7$
	V	$1/2^{13}$	$1/2^{19}$	–	$1/2^{15}$
Hagan	S	–	–	$1/2^7$	$1/2^8$
	V	$1/2^8$	$1/2^{18}$	$1/2^{20}$	$1/2^{14}$
SABR	S	–	–	$1/2^3$	$1/2^{12}$
	V	$1/2^8$	$1/2^{18}$	$1/2^{20}$	$1/2^{13}$
Stein	S	$1/2^3$	$1/2^{10}$	$1/2^{18}$	$1/2^{11}$
	V	$1/2^7$	$1/2^{17}$	$1/2^{19}$	–
Scott	S	–	–	$1/2^8$	–
	V	$1/2^3$	$1/2^8$	$1/2^{11}$	–

Δ is suggested to be set no bigger than $1/2^{16}$ to satisfy the error requirement of less than 0.10%, and Δ is better to set less than $1/2^{19}$ to satisfy the error requirement of no more than 0.01%. The turning point column provides an effective time step Δ for the Euler approximation to each diffusion process to compute moments, probabilities or other functions of the diffusion process. When Δ is smaller than its corresponding turning point, the error of the means keeps varying around some value and even increases sometimes.

Since the state variables of multivariate diffusion processes behave in different pattern, we report guidelines for both of them. When a good approximation of

sample path of a diffusion model is required, Δ could be set according to which state variable is interested of. However, for the cases where approximation of moments of the process matters, our Monte Carlo simulation results recommend to consider the state variable V in the first place.

CHAPTER 5

Conclusions

This thesis makes three main contributions to the field of continuous time diffusion models.

First, we combine the regime shift with three stochastic volatility models, including the Heston model, the GARCH model and the CEV model. According to the number of regimes, the initial probability and the transition probability matrix specifications, we estimate four models for each group. For Heston model, we compare H-R1 (single-regime), H-R2-1 (two regimes, time-constant transition matrix, unconditional probability for the probability of the initial state), H-R2-2 (two regimes, time-constant transition matrix, additional parameter for the probability of the initial state) and H-R2TVTP (two regimes, time-varying transition matrix with a logistic function, additional parameter for the probability of the initial state). For the GARCH model and the CEV model, We estimate G-R1, G-R2-1, G-R2-2, G-R2TVTP and C-R1, C-R2-1, C-R2-2, C-R2TVTP, separately. What's more, all parameters in these models are allowed to vary depending on the state of the economy. The maximum likelihood estimation, which is developed in Chapter 2, is applied to estimate these models. Using S&P 500 and VIX data for the stock price and volatility proxy, respectively, we report estimates for each model, as well as AIC, BIC, RCM, LR statistics to compare different models. Furthermore, we investigate the regime-switching probabilities of the regime-switching models with time varying transition matrix and additional parameter for the probability of initial state. We also analyse the corresponding approximate conditional transition density functions with their 95% confidence intervals in order to find the evidence of regime shift. Our estimation results show four main findings. First, the regime-switching models are significantly different from the single regime models. Second, there are strong evidences for the existence of the high and low volatility regimes, for the time varying transition probability of the regime variable, and for high persistence of the high regime. Third, the time varying transition probability mainly depends on the stock market index S&P 500 rather than its volatility. Fourth, the regime-switching CEV model with time varying transition

matrix and additional parameter for the probability of initial state performs better than the other regime-switching models.

Second, we first develop maximum likelihood estimation with closed-form likelihood expansions for multivariate regime-switching continuous time diffusion models. Furthermore, we apply it successfully to estimate our regime-switching stochastic volatility models.

Third, we provide guidelines for the choice of discretization interval when approximating a continuous time diffusion process using the Euler approximation method. On one hand, we analyses the effective time discretization step Δ for univariate stochastic diffusion models, including Black-Scholes-Merton model, Vasicek model, CIR model, Inverse of Feller's Square Root model, Linear Drift CEV model and Nonlinear Mean Reversion model. On the other hand, we move to multivariate stochastic diffusion models, such as Heston, GARCH, CEV, Stein, Scott, Hull-White, Hagan and SABR process. Regarding univariate diffusion continuous time models, our numerical tests suggest an appropriate time interval should be set for each model in order to satisfy different error requirements when the objective is to simulate a diffusion process or test statistical estimators. When the objective is to compute moments, probabilities or other functions of the diffusion process, we suggest considering the turning point of the corresponding weak convergence and control Δ does not exceed this turning point. Considering the bivariate continuous time diffusion models, we report guidelines for both of state variables. When a good approximation of sample path of a diffusion model is required, Δ could be set according to which state variable is interested of. However, for the cases where approximation of moments of the process matters, our Monte Carlo simulation results recommend to consider the volatility state variable in the first place.

In the future study, we intend to extend our work in four directions.

First, we intend to add jumps or more regimes in these models in order to propose a more general stochastic volatility model to describe the behaviour of S&P 500 and VIX. For example, the stock market index S&P 500 could be modelled as a function of a vector of state variables X_t that follows a bivariate diffusion process,

$$dX_t = \mu^P\left(X_t;\theta\right)dt + \sigma\left(X_t;\theta\right)dW_t^P + J_t^P dN_t^P,$$

where the pure jump process N^P has stochastic intensity $\lambda(X_t;\theta)$ and jump size 1. The state variables X_t and the drift functions $\mu^P(X_t;\theta)$ are 2×1 vectors, $\sigma(X_t;\theta)$ is an 2×2 volatility matrix, and W_t^P is an 2×1 vector of independent Brownian motions under the objective probability measure P. Both $\mu^P(X_t;\theta)$ and $\sigma(X_t;\theta)$ depend on X_t and they are known up to a parameter vector $\theta\in\Theta$, which is a compact subset of R^p. Additionally, our regime-switching stochastic volatility models could be used for option pricing, according to the suggestions we got during our presentations.

Second, using the experience of Chapter 2 for reference, we could develop maximum likelihood estimation with closed-form likelihood expansions for continuous time stochastic volatility models with jumps. As Aït-Sahalia and Kimmel (2007) mentioned, the expression will be in the form of

$$\begin{aligned} p_X^{(J)}(\Delta,x|x_0;\theta) &= \exp\left(-\frac{m}{2}\ln(2\pi\Delta)-D_v(x;\theta)+\frac{c_X^{(-1)}(x|x_0;\theta)}{\Delta}\right) \\ &\quad\times\sum_{k=0}^{J}c_X^{(k)}(x|x_0;\theta)\frac{\Delta^k}{k!}+\sum_{k=1}^{J}d_X^{(k)}(x|x_0;\theta)\frac{\Delta^k}{k!}, \end{aligned}$$

where $d_x^{(k)}$ are the new terms needed to capture the presence of the jumps in the transition probability density function. Then we could follow the same procedure in Chapter 2 to conduct this estimation methodology to estimate regime-switching stochastic volatility models with jumps.

Third, our regime-switching stochastic volatility models or other stochastic volatility models could be extended to the time inhomogeneous case. To estimate these models, we intend to apply Choi (2013)'s method, which is an extension of Aït-Sahalia (2008). Choi (2013) finds the irreducible method does not work like the time homogeneous case, because the recursive way of getting the coefficients $C_X^{(j_k,k)}$ breaks down. However, he shows that those indeterminate terms are cancelled out in $\widetilde{I}_X^{(K)}(t,x|t_0,x_0;\theta)$ even if they cannot be found from the PDEs of $C_X^{(-1)}(t,x|t_0,x_0;\theta)$ and $C_X^{(0)}(t,x|t_0,x_0;\theta)$.

Fourth, more continuous time diffusion models need to be investigated in Chapter 4. We intend to consider some high-dimensional continuous time diffusion models,

such as the stochastic skew model,

$$\begin{aligned} dS_t &= \mu S_t dt + \sigma e^{\gamma_t\left(\frac{S_t}{H}-1\right)} S_t dW_t^{(s)} \\ d\ln\sigma_t &= k_\sigma\left(\ln\sigma_\infty - \ln\sigma_t\right) + \alpha_\sigma dW_t^\sigma \\ d\gamma_t &= \kappa_\gamma\left(\gamma_\infty - \gamma\right) dt + \alpha_\gamma dW_t^\gamma, \end{aligned}$$

where $E[dW_t^\sigma dW_t^\gamma] = E[dW_t^\sigma dW_t^S] = E[dW_t^\gamma dW_t^S] = 0$. In addition, as we mentioned in Chapter 4, convergence concerns the accuracy of an Euler approximation over a finite interval $[0, T]$ for small time discretization steps Δ, while stability investigates the quality of an approximation in a long term, $T \to \infty$. We intend to analyse the stability of the Euler approximation to different continuous time diffusion models as well.

References

Abe, A. (2004): "Strong Taylor Schemes for Stochastic Volatility," *Working Paper*, The Oxford Centre for Industrial and Applied Mathematics, the University of Oxford.

Ahn, D.-H., and B. Gao (1999): "A Parametric Nonlinear Model of Term Structure Dynamics," *Review of Financial Studies*, 12, 721–762.

Aït-Sahalia, Y. (1996a): "Nonparametric Pricing of Interest Rate Derivative Securities," *Econometrica*, 64, 527–560.

——— (1996b): "Testing Continuous-Time Models of the Spot Interest Rate," *Review of Financial Studies*, 9, 385–426.

——— (1999): "Transition Densities for Interest Rate and Other Nonlinear Diffusions," *Journal of Finance*, 54, 1361–1395.

——— (2002): "Maximum Likelihood Estimation of Discretely Sampled Diffusions: A Closed-Form Approximation Approach," *Econometrica*, 70, 223–262.

——— (2008): "Closed-Form Likelihood Expansions for Multivariate Diffusions," *Annals of Statistics*, 36, 906–937.

Aït-Sahalia, Y., and R. Kimmel (2007): "Maximum Likelihood Estimation of Stochastic Volatility Models," *Journal of Financial Economics*, 83, 413–452.

Ang, A., and G. Bekaert (2002): "Regime Switches in Interest Rate," *Journal of Business and Economic Statistics*, 20, 163–182.

Bates, D. S. (2006): "Maximum Likelihood Estimation of Latent Affine Processes," *The Review of Financial Studies*, 19, 909–965.

Black, F., and M. Scholes (1973): "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy*, 81, 637–654.

Buckwar, E., and T. Sickenberger (2011): "A Comparative Linear Mean-Square Stability Analysis of Maruyama- and Milstein-Type Methods," *Mathematics and Computers in Simulation*, 81, 1110–1127.

Calvet, L. E., and A. j. Fisher (2004): "How to Forecast Long-Run Volatility: Regime Switching and the Estimation of Multifractal Processes," *Journal of Financial Econometrics*, 2, 49–83.

Chan, K., G. Karolyi, F. A. Longstaff, and A. B. Sanders (1992): "An Empirical Comparison of Alternative Models of the Short-Term Interest Rate," *Journal of Finance*, 47, 1209–1227.

Chen, B., and Y. Hong (2011): "Generalized Spectral Testing for Multivariate Continuous-Time Models," *Journal of Econometrics*, 164, 268–293.

Chen, B., and Z. Song (2013): "Testing Whether the Underling Continuous-Time Process Follows a Diffusion: An Infinitesimal Operator-Based Approach," *Journal of Econometrics*, 173, 83–107.

Cho, J. S., and H. White (2007): "Testing for Regime Switching," *Econometrica*, 75, 1671–1720.

Choi, S. (2009): "Regime-Switching Univariate Diffusion Models of the Short-Term Interest Rate," *Studies in Nonlinear Dynamics & Econometrics*, 13, Article 4.

——— (2013): "Closed-Form Likelihood Expansions for Multivariate Time-Inhomogeneous Diffusions," *Journal of Econometrics*, 172, 45–65.

Christensen, K., S. Kinnebrock, and M. Podolskij (2010): "Pre-averaging Estimators of the Ex-post Covariance Matrix in Noisy Diffusion Models with Non-synchronous Data," *Journal of Econometrics*, 159, 116–133.

Cox, J., J. Ingersoll, and S. Ross (1985): "A Theory of the Term Structure of Interest Rates," *Econometrica*, 53, 385–407.

Dai, Q., K. J. Singleton, and W. Yang (2007): "Regime Shifts in a Dynamic Term Structure Model of U.S. Treasury Bond Yield," *The Review of Financial Studies*, 20, 1669–1706.

Davis, R. B. (1987): "Hypothesis Testing When a Nuisance Parameter is Present Only under the Alternative," *Biometrika*, 74, 33–43.

Diebold, F., J.-H. Lee, and G. Weinbach (1994): *Regime Switching with Time-Varying Transition Probabilities in C. Hargreaves, ed., Nonstationary Time Series Analysis and Cointegration*. Oxford University Press. Oxford.

Duffie, D., J. Pan, and K. Singleton (2000): "Transform Analysis and Asset Pricing for Affine Jump-Diffusions," *Econometrica*, 68, 1343–1376.

Duffie, D., and K. Singleton (1993): "Simulated Moments Estimation of Markov Models of Asset Prices," *Econometrica*, 61, 929–952.

Durham, G., and Y.-H. Park (2013): "Beyond Stochastic Volatility and Jumps in Returns and Volatility," *Journal of Business and Economic Statistics*, 31, 107–121.

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

Egorov, A. V., H. Li, and Y. Xu (2003): "Maximum Likelihood Estimation of Time Inhomogeneous Diffusions," *Journal of Econometrics*, 114, 107–139.

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

Fornari, F., and A. Mele (1996): "Medeling the Changing Asymmetry of Conditional Volatilities," *Economic Letters*, 50, 197–203.

Gallant, A. R., and G. Tauchen (1998): "Reprojecting Partially Observed Systems with Application to Interest Rate Diffusions," *Journal of the American Statistical Association*, 93, 10–24.

Gray, S. (1996): "Modelling the Conditional Distribution of Interest Rates as a Regime-Switching Process," *Journal of Financial Economics*, 42, 27–62.

Hagan, P., D. Kumar, and A. Lesniewski (2002): "Managing Smile Risk," *Wilmott*, pp. 84–108.

Hamilton, J., and R. Sunmel (1994): "A Conditional Heteroskedasticity and Change in Regimes," *Journal of Econometrics*, 64, 307–333.

Hamilton, J. D. (1989): "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, 57, 357–384.

Hansen, B. (1992): "The Likelihood Ratio Test Under Nonstandard Conditions: Testing the Markov Switching Model of GNP," *Journal of Applied Econometrics*, 7, 61–82.

——— (1996): "Erratum: The likelihood Ratio Test Under Nonstandard Conditions: Testing the Markov Switching Model of GNP," *Journal of Applied Econometrics*, 11, 195–198.

Hansen, L. P., and J. A. Scheinkman (1995): "Back to the Future: Generating Moment Implications for Continuous-Time Markov Processes," *Econometrica*, 63, 767–804.

Harrison, M., and D. Kreps (1979): "Martingales and Arbitrage in Multiperiod Securities Markets," *Journal of Economic Theory*, 20, 381–408.

Harrison, M., and S. Pliska (1981): "Martingales and Stochastic Volatility Integrals in the Theory of Continuous Trading," *Stochastic Processes and Their Applications*, 11, 215–260.

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

Higham, D. J. (2000a): "A-Stability and Stochastic Mean-Square Stability," *BIT Numerical Mathematics*, 40, 404–409.

——— (2000b): "Mean-Square and Asymptotic Stability of the Stochastic Theta Method," *SIAM Journal on Numerical Analysis*, 38, 753–769.

——— (2001): "An Algorithmic Introduction to Numerical Simulation of Stochastic Differential Equations," *SIAM Review*, 43, 525–546.

Hull, J., and A. White (1987): "The Pricing of options on assets with stochastic volatility," *Journal of Finance*, 42, 281–300.

Jensen, B., and R. Poulsen (2002): "Transition Densities of Diffusion Processes: Numerical Comparison of Approximation Techniques," *Journal of Derivatives*, 9, 18–32.

Jones, C. (2003): "The Dynamics of Stochastic Volatility: Evidence from underlying and Volatility Models," *Journal of Econometrics*, 116, 181–224.

Kleppe, T. S., J. Yu, and H. Skaug (2010): "Simulated Maximum Likelihood Estimation of Continuous Time Stochastic Volatility Models," *Advances in Econometrics*, 26, 137–161.

Komori, Y., Y. Saito, and T. Mitsui (1994): "Some Issues in Discrete Approximate Solution for Stochastic Differential Equations," *Computers and Mathematics with Applications*, 28, 269–278.

Koo, B., and O. Linton (2012): "Estimation of Semiparametric Locally Stationary Diffusion Models," *Journal of Econometrics*, 170, 210–233.

Kristensen, D. (2010): "Pseudo-Maximum Likelihood Estimation in Two Classes of Semiparametric Diffusion Models," *Journal of Econometrics*, 156, 239–259.

Lacus, S. M. (2008): *Simulation and Inference for Stochastic Differential Equations*. Springer.

Ledoit, O., P. Santa-Clara, and S. Yan (1998): "Relative Pricing of Options with Stochastic volatility," *Working Paper*, University of California-Los Angeles Finance.

Li, M. (2010): "A Damped Diffusion Framework for Financial Modelling and Closed-Form Maximum Likelihood Estimation," *Journal of Economic Dynamics and Control*, 34, 132–157.

Lo, A. W. (1988): "Maximum Likelihood Estimation of Generalized Ito Processes with Discretely Sampled Data," *Econometric Theory*, 4, 231–247.

Mandelbrot, B. (1963): "The Variation of Certain Speculative Prices," *Journal of Business*, XXXVI, 392–417.

Merton, R. C. (1973): "Theory of Rational Option Pricing," *Bell Journal of Economics and Management Science*, 4, 141–183.

Milstein, G. N. (1975): "Approximate Integration of Stochastic Differential Equations," *Theory of Probability and its Applications*, 9, 557–562.

Mitra, S. (2010): "Regime Switching Stochastic Volatility Option Pricing," *Journal of Financial Markets and Derivatives*, 1, 213–242.

Papanicolaou, A., and R. Sircar (2013): "A Regime-Switching Heston model for VIX and S&P 500 Implied Volatilities," *Quantitative Finance*, pp. 1–7.

Pearson, N. D., and T.-S. Sun (1994): "Exploiting the Conditional Density in Estimating the Term Structure: An Application to the Cox, Ingersoll, and Ross Model," *Journal of Finance*, 49, 1279–1304.

Pedersen, A. R. (1995): "A New Approach to Maximum-Likelihood Estimation for Stochastic Differential Equations Based on Discrete Observations," *Scandinavian Journal of Statistics*, 22, 55–71.

Peter, E., and E. Platen (1992): *Numerical Solution of Stochastic Differential Equations*. Springer-Verlag.

Phillips, P. C. (1988): "An Interview with Professor A. R. Bergstrom," *Econometric Theory*, 4, 301–328.

Richard, J.-F., and W. Zhang (2007): "Efficient High-Dimensional Importance Sampling," *Journal of Econometrics*, 141, 1385–1411.

Saito, Y., and T. Mitsui (1996): "Stability Analysis of Numerical Schemes for Stochastic Differential Equations," *SIAM Journal on Numerical Analysis*, 33, 2254–2267.

Sargan, J. (1974): "Some Discrete Approximations to Continuous Time Stochastic Models," *Journal of the Royal Statistical Society*, 36, 74–90.

Schaumburg, E. (2001): "Maximum Likelihood Estimation of Jump Processes," *Ph.D. thesis*, Princeton University.

Schöbel, R., and J. Zhu (1999): "Stochastic Volatility with an Ornstein Uhlenbeck Process: An Extension," *European Finance Review*, 3, 23–46.

Scott, L. (1987): "Option Pricing When the Variance Changes Randomly: Theory, Estimation and an Application," *Journal of Financial and Quantitative Analysis*, 22, 419–438.

Song, Z. (2011): "A Martingale Approach for Testing Diffusion Models Based on Infinitesimal Operator," *Journal of Econometrics*, 162, 189–212.

Stein, E., and J. Stein (1991): "Stock Price Distribution with Stochastic Volatility," *The Review of Financial Studies*, 4, 727–752.

Stramer, O., M. Bognar, and P. Schneider (2010): "Bayesian Inference for Discretely Sampled Markov Processes with Closed-Form Likelihood Expansions," *Journal of Financial Economics*, 8, 450–480.

Sundaresan, S. M. (2000): "Continuous-Time Methods in Finance: A Review and an Assessment," *Journal of Finance*, 4, 1569–1622.

VARADHAN, S. (1967): "Diffusion Processes in a Small Time Interval," *Communications on Pure and Applied Mathematics*, 20, 659–685.

VASICEK, O. (1977): "An Equilibrium Characterisation of the Term Structure," *Journal of Financial Econometrics*, 5, 177–188.

WIGGINS, J. (1987): "Option Values under Stochastic Volatility: Theory and Empirical Estimate," *Journal of Financial Economics*, 19, 351–372.

YU, J. (2007): "Closed-Form Likelihood Approximation and Estimation of Jump-Diffusions with an Application to the Realignment Risk of the Chinese Yuan," *Journal of Econometrics*, 141, 1245–1280.

ZHAO, Z. (2011): "Nonparametric Model Validations for Hidden Markov Models with Applications in Financial Econometrics," *Journal of Econometrics*, 162, 225–239.