Essays on practical issues in asset pricing: estimation and simulation

Thesis

Submitted to the University of Warwick
for the degree of
Doctor of Philosophy

by
YAN WANG

supervised by
Dr. Xing Jin

27 July, 2015

To those who ever lightened my life

Acknowledgment

A large number of people contributed directly and indirectly to the accomplishment of this thesis. Most of all, I want to thank my supervisor Dr. Xing Jin. He introduced me to the subject of my research and gave me the opportunity that conducted to this thesis. His support and contagious enthusiasm were absolutely essential for me. I also thank Dr. Nick Webber who provided full support and drew me to this area at the very beginning.

I would like to thank both the University of Warwick and Warwick Business School, who supported me both financially and academically. I want to thank Kevin, for the support he has shown during the past many years. I could not have finalized this thesis without him.

At last, a special thanks goes to my new friends at the University of Warwick and those old yet lifetime ones spread over the world, who always encouraged me, helped me in dark times and made my life easier and happier.

Declarations

The work submitted in this thesis is the result of my own investigation, except where otherwise stated. It has not already been accepted for any degree at another university, and is also not being concurrently submitted for any other degree at another university.

Contents

List of Figures

List of Tables

Abstract

This thesis studies several practical issues in asset pricing, including MCMC estimation of time-changed Lévy processes, calibration techniques for stochastic volatility models, and a sampling scheme for the SABR model. First, a MCMC estimation approach is developed to estimate time-changed Lévy processes. Simulation-based experiments demonstrate good accuracy of the MCMC approach. An empirical study on its fitness of the return dynamics is provided, which shows that time-changed Lévy models can achieve excellent performance in capturing index returns. Second, a further study on MCMC estimation is applied to multivariate Lévy processes, in order to evaluate the efficiency and accuracy of the Bayesian technique for high-dimensional portfolio theory. Last, a new representation of the SABR model is proposed by adopting a coupling approach, based on which, the uncorrelated SABR is sampled from its density. Numerical experiments are implemented to compare the sampling scheme with the Euler discretization scheme and examine the accuracy of Hagan's popular formula for the implied Black-Scholes volatility.

Chapter 1

General Introduction

Modelling the return dynamics of individual stocks and indices is a core concern of academics and practitioners working on the asset pricing theory. A good pricing model is essential for portfolio allocation, derivatives pricing and risk modelling and management. How to choose and use a proper model is actually the mixed field of estimation, calibration and simulation. For instance, to price and hedge financial derivatives, the first move is to calibrate the model with the curvature of current market surface. With the calibrated model, simulation might be needed to price some exotic derivatives which cannot be priced explicitly. For the portfolio allocation problem, the underlying model needs to be estimated with a period of historical data. The optimal portfolio weights will be calculated according to the estimated model. Clearly, asset pricing models can only be used after either calibration or estimation.

The difficulty of proposing a proper asset pricing model is not trivial and involves following

the development of financial markets. Looking back to the history of finance, every financial crisis has brought some new features to the market such as the crash of October 1987, the burst of the dot-com bubble in 2000, and the recent financial crisis starting in 2007. Nowadays a pricing model has to be able to capture large-size price and volatility movements on the market. In addition, the model also needs to be able to account for well-documented stylized facts such as volatility clustering and fat tails in the return distribution, and must be capable of modeling the leverage effect, and produce the smile patterns observed in option data.

Jumps in capturing return dynamics along with diffusion were first introduced in Merton (1976), in order to model large and rare movements of returns. Stochastic volatility is another famous market feature that has significant impact on option pricing, which inspires the development of stochastic volatility models such as the celebrated Heston model proposed by Heston (1993). However, the development of modelling never stop as the market keeps exhibiting new features that cannot be captured by existing models. For instance, neither stochastic volatility models nor jump models can explain the market data in a quantitative sense. In particular, stochastic volatility models fail to explain large price drops that occur during a single day since it would require an unrealistically high volatility level before and after the crash. Models with jumps in returns can explain large price movements, but fail to explain volatility clustering over time.

Merton (1976) starts using jumps by introducing the compound Poisson jump. A more general choice is using a Lévy process. Lévy processes are related to Infinitely Divisible

distribution, which can provide a variety of non-Gaussian distributions. Nowadays, Lévy processes have become a popular alternative to diffusion, especially in derivative pricing. Jump risk that represents the sudden loss in the market cannot be modelled by diffusion models. Imitating the Black-Scholes model, many Geometric Lévy models have been proposed, such as the Variance Gamma (VG) model (see Madan et al. (1998)) and the CGMY model (see Carr et al. (2002)).

Simulation is probably the best way to derive fair prices when no explicit solution can be obtained. PDE method and lattice method all have limitations such as the curse of dimensionality. Naive simulation methods like the Euler scheme can generate an good approximation of the derivative price; however, a biased estimator of prices might cause significantly large losses. Exact-simulation is introduced in order to avoid pricing biases. The SABR model has been widely used by practitioners in the financial industry, especially in the interest rate derivative markets. The popularity of the SABR model might be due to its fast asymptotic solution to implied volatility. The simulation of the SABR model is very crucial as no explicit solution exists. The probability of touching boundary is a big issue for simulating the SABR model. Exact-simulation method for the SABR model is still an open question, which is of great practical value.

In this thesis, three topics associated with practical issues are discussed. Estimation, calibration and simulation are the most important practical topics of financial engineering. In chapter 2, a Bayesian Markov Chain Monte Carlo (MCMC) estimation method is developed for estimating time-changed Lévy models which admit both stochastic volatility

and jumps. The estimation examples show that the MCMC method is capable of estimating Lévy models with very good accuracy, based on simulation data. We also show that returns produced by time-changed infinite activity Lévy processes cannot be modelled by existing jump-diffusion models. Empirical results suggest that infinite activity Lévy processes can outperform jump-diffusion models in capturing the variation of index returns, even without diffusion.

In chapter 3, we consider multidimensional, continuous-time modelling problem where the observation process is a diffusion with drift and volatility coefficients being modeled as continuous-time, finite-state Markov chains with a common state process. For the econometric estimation for drift and volatility of the underlying Markov chain, we develop an discrete-time Markov chain Monte Carlo (MCMC) sampler and compare these approaches with maximum likelihood (ML) estimation. For simulated data, MCMC outperforms ML estimation for all scenarios. Finally, for real market index data, we apply the estimation approach and obtain fit results.

In chapter 4, we propose a new representation of the SABR model, in which, the SABR model can be seen as a time-changed CEV dynamics. We derive the implicit form of the time-change, make a guess of its explicit solution and prove the validity of our guess in the numerical sense. We then try to sample the SABR model based on its new representation. We first sample the CEV dynamics directly from its density, then sample the time-change by approximating its time-T distribution with a lognormal distribution by matching their first two central moments. With the sampling schemes for both the CEV dynamics and

the time-change, we sample the non-correlated SABR model, compare the performance of our sampling scheme with the Euler Monte Carlo scheme, and examine the accuracy of Hagan's popular asymptotic formula for the implied Black-Scholes volatility.

Chapter 5 concludes the thesis and discusses future research.

Chapter 2

An empirical study on MCMC estimation for return dynamics with time-changed Lévy processes

2.1 Introduction

Modelling returns series is a core issue in asset pricing theory. Many models including both continuous-time models and discrete-time models have been proposed and assessed for different purposes. The development of modelling always reflects the evolution of the market. The market keeps exhibiting new features that existing models cannot fit, and market practitioners and academic researchers try to modify existing models or propose

new models in order to follow the market. For instance, the celebrated Black-Scholes model is the first quantitative model that has a broad range of applications in asset pricing. The Black-Scholes model assumes that log returns follow Normal distribution; however, empirical literature documents that market returns do not exhibit either zero-skewness or low excess-kurtosis. Moreover, after experiencing several financial crises, practitioners and the academia started to consider and identify the existence of jumps as rare large movements of stock prices are not so rare. The booming development of the derivatives market also brings another issue that needs to be taken into account, which is stochastic volatility. Practitioners realize that volatility risk plays an essential role in pricing and hedging, which results in the development of Stochastic Volatility (SV) models.

Evaluating the performance of a pricing model seems to be a straightforward job; however, it is not always the case. For some simple models, there are many possible estimation methods that can work, such as the Maximum likelihood estimation (MLE) method, the filtering method, and Generalized method of moments (GMM). Each method has its advantages and disadvantages. There is no universal method that works for every circumstance. Estimating multi-dimensional stochastic processes is always a daunting task. Calibration has also been used as a special approach of estimation, which is still debatable now as serious statisticians do not take calibration as an estimation technique. The key difference is that estimation adopts the historical data sampled in a relatively long period while calibration is usually done on daily basis.

Estimation technique partially determines the popularity of modelling work. A similar

situation can also be found in the field of calibration. Affine Jump-diffusion (AJD) models have been popular for a long time, because the Fast Fourier Transform (FFT) method enables it to do the calibration procedure in an acceptable time. Despite the disadvantages, AJD models still have a variety of application in asset pricing. A powerful estimation technique can boom the development of models and also provide a way to evaluate and compare the performance of models. Existing estimation methods such as MLE and GMM have difficulty in dealing with high-dimensional processes, which limits the use of sophisticated multi-dimensional asset pricing models.

Lévy processes have become popular and increasingly useful recently. Many literature have documented the applications of Lévy processes in various fields such as derivatives pricing, risk management and credit risk modelling (see Barndorff-Nielsen (1998), Carr et al. (2002) and Carr and Wu (2003)). Lévy processes are closely related to Infinitely Divisible Distribution (see Sato (1999)). Hence, there are many possible choices for the innovation distribution, in order to capture some market features such as asymmetric skewness and high kurtosis. Brownian motion is the cornerstone of stochastic modelling and only a special case of Lévy processes. Generally speaking, a Lévy process is a càdlàg stochastic process that has independent stationary increments. If we impose a distribution on the increments, we will have a specific Lévy process. Lévy processes have good economic intuition and provide good flexibility for modeling return dynamics. Moreover, Lévy processes introduce jumps into modelling, which can capture large and rare movements of asset prices and produce similar levels of skewness and kurtosis exhibited by the

market.

The development of Lévy models brings many new choices for asset pricing theory. Madan et al. (1998) propose the Variance Gamma (VG) process which provides an asymmetric distribution. Carr et al. (2002) discuss the empirical performance of several popular Lévy processes including VG process and CGMY process, in terms of statistical estimation and risk-neutral estimation. Lévy processes have not been commonly used to solve the portfolio allocation problem. One of the main reasons is that it is difficult to estimate Lévy type models. Li et al. (2008) propose a MCMC method to estimate Lévy processes and assess the performance of Lévy processes with market data. Kallsen and Muhle-Karbe (2011) develop a method of moments in order to estimate time-changed Lévy processes. In this chapter, we extend the MCMC method proposed by Li et al. (2008) to deal with time-changed Lévy processes and investigate whether time-changed Lévy processes can be used to model return dynamics.

Lévy processes can be categorized by the property of activity rate. Activity rate describes the intensity of the appearance of jumps. If the activity is infinite, the corresponding Lévy is said to be an infinite activity Lévy process. In any finite time interval, there will be infinitely many jumps for an infinite activity Lévy process. In addition to large and rare jumps, an infinite Lévy process can capture the behaviour of all sizes of jumps, even without the diffusion component. Our focus is to estimate time-changed Lévy processes and investigate whether infinite activity Lévy models, despite their theoretical properties, can significantly outperform existing Jump-diffusion models with stochastic volatility.

In this chapter, we aim to develop a MCMC estimation method for time-changed Lévy models and assess the goodness of fit based on market data. The main contribution of this chapter can be understood in three aspects. First, we develop the estimation method with full details, which can be applied to the whole set of Lévy processes. We will demonstrate whether the MCMC method can jointly estimate both model parameters and latent variables, especially those small infinite activity jumps. The empirical results suggest that MCMC estimation can provide very accurate joint identification of Lévy jumps and stochastic volatility. we demonstrate the accuracy and stability of the MCMC method based on simulation data. This enables us to explore whether time-changed Lévy model can generate better fit, compared with existing stochastic Jump-diffusion models.

Second, we study whether it is necessary to adopt time-changed Lévy processes in the statistical sense. Li et al. (2008) have demonstrated that AJD models cannot capture returns generated by infinite activity Lévy jumps. Based on estimation results, we show that time-changed Lévy models have similar behaviour as Jump-diffusion models with Lévy jumps used in Li et al. (2008). The diffusion component only matters for generating the Leverage-effect. Infinite activity Lévy processes can capture both small and large movements of returns.

Third, we assess the goodness of fit with market data and provide a clear answer to whether time-changed Lévy processes should be adopted in modelling return dynamics. Li et al. (2008) suggest that Lévy jump models can outperform AJD models in capturing the time series dynamics of the S&P 500 index returns. We move a step further to show

time-changed Lévy models can achieve a similar level of fit without diffusion. A theoretical shortcoming of time-changed Lévy models used in our experiment is the absence of the Leverage-effect. To weaken the impact of the Leverage-effect, we particularly choose an index data from the MSCI database. Further work should be done by using Lévy models subordinated by pure jump processes which admits the Leverage-effect in estimation experiments, in order to examine whether pure Lévy models can outperform stochastic volatility models with Lévy jumps in capturing the dynamics of index returns.

The rest of this chapter is organized as follows. Section 2.2 describes the MCMC estimation method developed for time-changed Lévy processes and introduces the stochastic volatility models built up by time-changed Lévy processes. Section 2.3 provides an empirical study on the accuracy of MCMC estimation based on simulation data and also discusses the difference between time-changed processes and other canonical stochastic processes. Section 2.4 presents the empirical analysis on the performance of fitting the return dynamics with time-changed Lévy processes. Section 2.5 concludes this chapter and discusses future research.

2.2 MCMC estimation and Lévy type models

2.2.1 Markov Chain Monte Carlo estimation

Markov Chain Monte Carlo (MCMC) is a Bayesian inference technique. Roughly speaking, the MCMC method obtains the point estimator by sampling from a posterior distribution. Usually, the difficulty of MCMC comes from deriving a simple and easy posterior distribution. There is no generic rules on how to derive posterior density functions from prior density functions. It requires specific experience to implement an efficient MCMC estimation method.

There are many choices of estimation methods such as the Maximum Likelihood estimation method, the Filtering estimation method, and the Generalized method of moments; however, estimating Lévy processes is still an daunting task because of the unknown density. Moreover, Lévy models involve stochastic volatility, jump intensity and the distribution of jumps, which make it very difficult to jointly estimate all observable variables and latent variables. Each Lévy process is unique, and requires specific care for estimation. For instance, α-stable processes do not have finite moments of log returns, which means that the method of moments is not applicable.

Estimating stochastic volatility models is difficult due to existence of the latent variable: volatility. If dealing with a jump model, latent variables will consist of jumps as well. Hence, estimating stochastic volatility models involves a high dimensionality problem. The Markov Chain Monte Carlo method is originally developed to overcome the high

dimensionality problem, so it is applicable to this estimation issue. The task of estimation is to find estimates of the underlying model that best fit the sample data. Latent variables cannot be directly observed. Suppose Θ is a vector of parameters needed to be estimated, V consists of all latent variables, such as jumps, volatility, and Y is the observed sample drawn from the market. By using the MCMC method, we want to figure out the conditional distribution $p(\Theta|Y)$, which provides the information of parameters with respect to the sample data.

The basic idea of MCMC is that we assume some prior distribution on the parameter that needs to be estimated. Then, we find the posterior distribution of the parameter and draw samples from the posterior distribution. According to the Bayes' Theorem, the posterior distribution is proportional to the likelihood times the prior distribution:

$$\begin{aligned} p(\Theta, V, J, G|Y) &\propto p(Y|\Theta, V, J, G)p(\Theta, V, J, G) \\ &= p(Y|\Theta, V, J, G)p(V, J, G|\Theta)p(\Theta), \end{aligned} \tag{2.1}$$

where Θ is the parameter, V is the latent volatility variable, J is the latent jump variable, G is the latent additional variable and Y is the sample data drawn from the market. If we can directly sample the distribution in (2.1), the estimation procedure will be fairly easy; however, in our case, conditional density is extremely complicated and direct sampling is not feasible. Alternatively, we can sample by constructing a Markov Chain over the parameters and latent variables whose equilibrium transition density converges to the desired posterior distribution. The sampling procedure is done by iterations. The best estimate is the empirical mean of samples simulated. Before taking the mean value,

we also need to cut off some samples from the beginning as it takes time to reach the stationary state. This is referred to as the 'burn-in' sample.

For each model, we can obtain the conditional density of parameters based on the sample data. The MCMC method adopts an iteration approach which iteratively generates samples from respective conditional posterior distributions. The algorithm is

1. Update all variables including parameters and latent variables given the prior distributions

$$\Theta^{(0)} : \pi(\Theta)$$

$$V^{(0)} : \pi(V)$$

$$J^{(0)} : \pi(J)$$

$$G^{(0)} : \pi(G)$$

2. for $k = 1, \ldots, M$

   the kth round of updating model parameters and latent variables iteratively

$$\theta_i^{(k)} : p\left(\theta_i^{(k)} | \Theta_i^{(k-1)}, V^{(k-1)}, J^{(k-1)}, G^{(k-1)}, Y\right), \quad i = 1, \ldots, m$$

$$V_i^{(k)} : p\left(V_i^{(k)} | \Theta^{(k)}, V_i^{(k-1)}, J^{(k-1)}, G^{(k-1)}, Y\right), \quad i = 1, \ldots, n$$

$$J_i^{(k)} : p\left(J_i^{(k)} | \Theta^{(k)}, V^{(k)}, J_i^{(k-1)}, G^{(k-1)}, Y\right), \quad i = 1, \ldots, n$$

$$G_i^{(k)} : p\left(G_i^{(k)} | \Theta^{(k)}, V^{(k)}, J^{(k)}, G_i^{(k-1)}, Y\right), \quad i = 1, \ldots, n$$

3. cut off the first N samples as the 'burn-in' samples and generate the best estimates

$\widehat{\Theta}$ by:

$$\widehat{\Theta}_i = \frac{1}{M-N} \sum_{j=N+1}^{M} \Theta_i^{(j)}, \quad i = 1, \ldots, n$$

$$\sigma^2(\Theta_i) = \frac{1}{M-N-1} \sum_{j=N+1}^{M} \left(\Theta_i^{(j)} - \widehat{\Theta}_i\right)^2, \quad i = 1, \ldots, n$$

The details of estimating each model are presented in Appendix A.1.

2.2.2 Lévy processes and Lévy type models

A Lévy process is a stochastic process with independent, stationary increments. Lévy processes have been widely used to model financial asset returns as Lévy processes can generate a variety of distributions. The most well known Lévy processes are Brownian motion and the Poisson process that are the most essential processes used in asset pricing.

Fix complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ endowed with a standard complete filtration $\{\mathcal{F}_t\}$ satisfying the usual conditions. A scaler stochastic process $\{L_t\}_{t\geq 0}$ is said to be a Lévy process if the following conditions are satisfied:

- $L_0 = 0$, a.s.;
- $L_t - L_s \perp L_s$, for any $t > s$;
- $L_t - L_s$ is equal in distribution to L_{t-s}, for any $t > s$.

If we impose a specific distribution on the increments, we can have a specific Lévy process. For instance, if $L_{t-s} \sim N(0, t-s)$, $\{L_t\}_{0\leq t\leq T}$ will be a Brownian motion.

Lévy processes can be fully characterized by the characteristic function. By the Lévy-Khintchine Theorem, the characteristic function of L_t has the representation:

$$\mathbb{E}\left[e^{iuL_t}\right] = \Phi(u) = \exp\left(iu\mu t - \frac{1}{2}u^2\sigma^2 t + t\int_{\mathbb{R}\setminus\{0\}}\left(e^{iux} - 1 - iux\mathbf{1}_{|x|<1}\right)\pi(dx)\right), \quad (2.2)$$

where $\mu \in \mathbb{R}$ is the drift, $\sigma \in \mathbb{R}^+$ is the dispersion parameter, $\mathbf{1}$ is the indicator function and $\pi(\cdot)$ is the Lévy measure defined on $\mathbb{R}\setminus\{0\}$ with the condition that

$$\int_{\mathbb{R}\setminus\{0\}} \min(1, x^2)\pi(dx) < \infty.$$

The Lévy-Khintchine representation (2.2) provides an intuitive insight into Lévy processes. An arbitrary Lévy process can be decomposed into three parts: a linear drift, a diffusion and a pure-jump. The pure-jump component can be understood as a superposition of independent Poisson processes with different jump sizes, where $\pi(dx)$ measures the jump intensity of the Poisson process with jump size x. Thus the Lévy-Khintchine representation is fully determined by Lévy triplet (μ, σ^2, π). The only continuous Lévy process is a Brownian motion with drift. If a Lévy process has no diffusion part, it is said to be a pure-jump process.

Lévy processes still belong to semimartingales and are directly related to Indefinitely Divisible Distribution. For each infinitely divisible probability distribution $\mathcal{F}$, there exists a unique Lévy process L_t such that L_1 follows $\mathcal{F}$. In fact, L_t can be described by L_1, due to its Markov property. Moreover, Lévy processes are natural to be used for modelling financial asset, because the property of 'independent stationary increments' coincides with the Efficient Market Hypothesis (see Fama (1970)).

Lévy processes can be categorized into two types: finite activity processes and infinite activity processes, based on activity rate. The Lévy measure $\pi(dx)$ determines the expected activity rate of jumps with size x per unit of time. Namely, the integral

$$\int_{\mathbb{R}\setminus\{0\}} \pi(dx) = \lambda \tag{2.3}$$

is the corresponding activity rate. If λ is finite, the process belongs to finite activity processes. Within any finite time interval, the number of jumps is finite. If the integral in (2.3) is not integrable, the process is an infinite activity process. This means that there will be infinitely many jumps in any finite time interval. The Poisson process is a finite activity process, due to its finite λ. The compound Poisson process defined as

$$Y_t = \sum_{k=1}^{N_t} \xi_k, \tag{2.4}$$

where N_t is a Poisson process with intensity λ and $(\xi_k)_{k\geq 1}$ are i.i.d. random variables, is another example of finite activity Lévy processes. The most common compound Poisson process is the Merton's jump process introduced in Merton (1976), where ξ_k follows $N(\mu, \sigma^2)$. The corresponding Lévy measure is

$$\pi_{MJ}(dx) = \lambda dF(x) = \lambda \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx,$$

where $F(\cdot)$ is the cumulative density function (CDF) of Normal distribution.

Popular examples of infinite activity Lévy processes include the Variance Gamma process, the CGMY process and the Log-stable process. These processes have different Lévy triplets, but have the same property that the activity of jumps is infinite. The reason that we pay attention to infinite activity processes is that infinite activity processes can model

both frequent-but-small and infrequent-but-large returns. The Merton's jump-diffusion model has a similar philosophy such that the diffusion component captures the movements of frequent-but-small returns and the jump component captures the movements of infrequent-but-large returns. Infinite activity processes can generate jumps with all sizes from small to large. A pure-jump process with infinite activity can approximate the behaviour of Jump-diffusion model, which will be demonstrated in Section 2.3.

The Variance Gamma (VG) process is proposed by Madan et al. (1998), which is an alternative to Gaussian innovation. The VG process can capture some market features such as non-zero skewness, high kurtosis and fat tails. The VG distribution can be obtained by constructing the difference of two independent Gamma random variables. For the sake of sampling, we use another representation of the VG distribution, that is mixing a Normal distribution with a Gamma random variate. Suppose G_t is a Gamma process with unit mean rate and variance rate of ν. A stochastic process X_t following:

$$X_t = \theta G_t + \sigma W_{G_t}, \tag{2.5}$$

where W_t is a standard Brownian motion, is said to be a Variance Gamma process. According to (2.5), a VG process is a time-changed drifted Brownian motion. A VG process has three parameters including $\Theta = (\sigma, \theta, \nu)$. The characteristic function of VG process X_t is

$$\phi(u) = \left(\frac{1}{1 - i\theta\nu u + \sigma^2 \nu u^2/2}\right)^{t/\nu}. \tag{2.6}$$

There is another representation of parameters for VG process, but it is not convenient for sampling and estimation. Carr et al. (2002) introduce a new Lévy process known as

the CGMY process. The VG process is a special case of the CGMY process; however, the CGMY process usually shows a similar performance as the VG process. Due to the difficulty of sampling, we only consider the VG process in this chapter.

The Log-stable (LS) process is another useful Lévy process, which belongs to the category of α-stable processes. α-stable processes have four parameters that control the behaviour of skew, tail, scale and drift. The characteristic function of an α-stable process L_t is presented as

$$\phi_\alpha(u) = \mathbb{E}[e^{iuL_t}] = \exp\left[iu\theta t - t|u|^\alpha\sigma^\alpha\left(1 - i\beta(sgnu)\tan\frac{\pi\alpha}{2}\right)\right],$$

where $\alpha \in (1, 2]$, $\theta \in \mathbb{R}$, $\sigma > 0$, $\beta \in [-1, 1]$. Usually, modelling with the α-stable processes cannot guarantee finite moments for spot prices, which is very essential. Carr and Wu (2003) introduce a Finite Moment Log Stable (LS) process that is a special case of α-stable processes where $\beta = -1$, $\alpha \in (1, 2)$, $\sigma = 1$, and $\theta = 0$. The LS process is the only type of α-stable process that ensures the existence of all moments. The LS process has only one degree of freedom that is the parameter of α. If $\alpha = 2$, the LS process becomes a Brownian motion. The corresponding characteristic function of the LS process is

$$\phi_{LS}(u) = \exp\left(-t(iu)^\alpha \sec\frac{\pi\alpha}{2}\right).$$

The LS process only has negative jumps as the associated Lévy measure is defined in the domain of $\mathbb{R}^+$; it has a positive drift that can compensate negative jumps, which still guarantees the Martingale property. Carr and Wu (2003) show that the LS process has a better fit of index options due to its ability to model the property that the implied

volatility smirk does not flatten out as maturity increases. This is the reason why we adopt it in our experiments.

Proposing a Lévy model is very trivial. Based on some Lévy process, mimicking the Black-Scholes model will result in a Geometric Lévy model. For instance, replacing the Brownian motion in the Black-Scholes model with a VG process we can obtain the VG model

$$S_t = S_0 \exp\left(\mu t + X_t\right), \tag{2.7}$$

where X_t is a VG process. If model (2.7) is used to price options, we also need to add up a Martingale correction parameter. Similarly, we can also get a LS model by replacing X_t with a LS process L_t. The VG model and the LS model provide better fit than the Black-Scholes model, but stochastic volatility is still missing. It is believed that volatility is not constant and exhibits stochastic behaviour. To capture stochastic volatility, many advanced models have been proposed with different features. A naive method is employing one more stochastic process to model the movements of instantaneous volatilities. Lévy type models cannot apply this method as it has no volatility parameter. The solution is the time-change technique.

Time-change technique can be applied to Lévy processes to generate stochastic volatility. The intuition is that we can randomize the clock on which the stochastic process is run. Hence, the number of transactions in a given time interval is also random. Since a high number of transactions causes high return volatility, time-changed Lévy processes can produce stochastic volatility. There are many choices for randomizing the business time,

such as the Ornstein-Uhlenbeck (OU) process and the CIR process. Lévy Subordinators are also good candidates. Defining a time-change is trivial as we can simply model the activity on which time runs. Let $t \to T_t (t \geq 0)$ be an increasing càdlàg process satisfying the usual conditions; a new process can be generated by evaluating L at T:

$$Y_t = L_{T_t}, \qquad t \geq 0.$$

As proved by Monroe (1978), every semimartingale can be represented by a time-change Brownian motion. Hence, Y_t is a very general specification for financial modelling. The random time T_t must be a nondecreasing process, and can be represented by

$$T_t = \int_0^t v_{s^-} ds,$$

where v_t is the instantaneous activity rate. An important point is that we want to impose dependence between the return innovations in L_t and the activity rate v_t, which can generate the leverage effect. The Heston model can also be represented as a time-changed Black-Scholes model with a CIR activity rate process. Carr et al. (2003) firstly investigate the behaviour of time-changed Lévy models by implementing and comparing the performance of a variety of time-changed Lévy models such the Stochastic Volatility Variance Gamma (SVVG) model. We adopt the same procedure to construct time-changed Lévy models as that in Carr et al. (2003).

The purpose of this chapter is to find out whether time-changed Lévy processes can outperform existing models including the Jump-diffusion model and simple Lévy models. For comparison, we choose several popular models including the Heston model with compound

Poisson jumps, the Heston model with Lévy jumps, the Time-changed Variance Gamma model (SVVG) and the Time-changed Log Stable model (SVLS) in our experiments. This chapter actually covers the gap left by Li et al. (2008) as it has been demonstrated that stochastic volatility Jump-diffusion models with Lévy jumps can outperform stochastic volatility Jump-diffusion models without/with compound Poisson jumps. We investigate whether pure time-changed Lévy models can do even better in terms of fitting the market.

Here we present model specifications of all models used in our experiments along with corresponding discretized forms.

- Heston-Jump model (HJ)

  The celebrated Heston model is a popular stochastic volatility model, yet empirical literature has documented that it cannot generate enough skew or smile (see Li et al. (2008)). The absence of jumps might be a reason. The Heston model with jumps will be a possible solution to this problem. Equipped with a compound Poisson jump, the Heston-Jump model can be presented as:

$$\begin{aligned} \frac{dS_t}{S_t} &= \mu dt + \sqrt{v_t} dW_t + dJ_t \\ dv_t &= \kappa(\eta - v_t)dt + \sigma_v \sqrt{v_t} dZ_t, \end{aligned} \tag{2.8}$$

  where $\mathbb{E}[dWtdZ_t] = \rho dt$. The discretized representation of HJ model is

$$\begin{aligned} Y_{t+1} &= Y_t + \mu\Delta + \sqrt{v_t\Delta}\epsilon^y_{t+1} + J_{t+1} \\ v_{t+1} &= v_t + \kappa(\eta - v_t)\Delta + \sigma_v\sqrt{v_t\Delta}\epsilon^v_{t+1}, \end{aligned} \tag{2.9}$$

where Δ is the discretization step, $J_{t+1} = \xi_{t+1}N_{t+1}$, ϵ^y_{t+1} and ϵ^y_{t+1} are correlated standard Normal distributed random variables with correlation ρ. The observable samples are $(Y_t)_{t=0}^T$, and latent variables include the instantaneous variance levels $(v_t)_{t=0}^T$, jump times $(N_t)_{t=1}^T$, and jump sizes $(\xi_t)_{t=1}^T$. And the $N_t \sim Poisson(\lambda)$ and the $\xi_t \sim N(\mu_J, \sigma_J^2)$. Parameters to be estimated are $\mathbf{\Theta} = \{\mu, \kappa, \eta, \sigma_v, \rho, \lambda_J, \mu_J, \sigma_J\}$.

- Heston with VG jump (HVG)

  The Heston model can also be merged with an infinite activity Lévy jump. If a VG process is adopted, we can obtain a Heston-VG model with the representation:

$$\begin{aligned} \frac{dS_t}{S_t} &= \mu dt + \sqrt{v_t} dW_t + dJ_t \\ dv_t &= \kappa(\eta - v_t)dt + \sigma_v \sqrt{v_t} dZ_t, \end{aligned} \tag{2.10}$$

  where dJ_t is the infinitesimal increment of a VG process. The discretized representation of the HVG model is

$$\begin{aligned} Y_{t+1} &= Y_t + \mu\Delta + \sqrt{v_t\Delta}\epsilon^y_{t+1} + J_{t+1} \\ v_{t+1} &= v_t + \kappa(\eta - v_t)\Delta + \sigma_v\sqrt{v_t\Delta}\epsilon^v_{t+1}, \end{aligned} \tag{2.11}$$

  where both ϵ^y and ϵ^v follow a standard Normal distribution with correlation ρ. The jump increment J_t can be decomposed as

$$J_{t+1} = \theta G_{t+1} + \sigma\sqrt{G_{t+1}}\epsilon^J_{t+1},$$

  where ϵ^J_{t+1} follows $N(0,1)$, and G_{t+1} follows $\Gamma\left(\frac{\Delta t}{\nu}, \nu\right)$. For the HVG model, we have latent variables $(v_t)_{t=0}^T$, jump sizes $(J_t)_{t=1}^T$, and random time $(G_t)_{t=1}^T$. The model

parameters are $\mathbf{\Theta} = \{\mu, \kappa, \eta, \sigma_v, \rho, \theta, \sigma, \nu\}$. In Li et al. (2008), the HVG is referred to as the SVVG model.

- Time-changed Variance Gamma model (SVVG)

  As suggested by empirical literature including Madan et al. (1998) and Carr and Wu (2003), the diffusion component might not be necessary if an infinite activity Lévy process is used. Hence, we want to remove the diffusion component in the HVG model and keep the VG process only. To produced stochastic volatility, we time-change the VG process with a CIR process. The time-changed Variance Gamma model can be represented as:

$$\begin{aligned} \log S_t &= \log S_0 + \mu t + X_{T_t} \\ T_t &= \int_0^t v_s ds \\ dv_t &= \kappa(\eta - v_t)dt + \sigma_v \sqrt{v_t} dZ_t. \end{aligned} \tag{2.12}$$

  The discretized representation of SVVG model is

$$\begin{aligned} Y_{t+1} &= Y_t + \mu\Delta + X_{t+1} \\ v_{t+1} &= v_t + \kappa(\eta - v_t)\Delta + \sigma_v \sqrt{v_t \Delta} \epsilon^v_{t+1}, \end{aligned} \tag{2.13}$$

  where X_{t+1} follows $\Gamma\left(\frac{\Delta v_{t+1}}{\nu}, \nu\right)$. An obvious drawback is that the SVVG model does not admit the Leverage-effect as there is no dependence between the spot price process and the variance process. The latent variables consist of $(v_t)_{t=0}^T$, jump sizes $(X_t)_{t=1}^T$, and random time $(G_t)_{t=1}^T$. The model parameters are $\mathbf{\Theta} = \{\mu, \kappa, \eta, \sigma_v, \theta, \sigma, \nu\}$.

- Time-changed Log-Stable model(SVLS)

  Similar to the SVVG model, we choose another infinite activity Lévy jump which is the LS process to construct a time-changed Lévy model. The representation of the SVLS model is:

$$\begin{aligned} \log S_t &= \log S_0 + \mu t + \sigma L_{T_t} \\ T_t &= \int_0^t v_s ds \\ dv_t &= \kappa(\eta - v_t)dt + \sigma_v \sqrt{v_t} dZ_t, \end{aligned} \tag{2.14}$$

  where L_t is a LS process with parameter α. The discretized representation of SVLS model is

$$\begin{aligned} Y_{t+1} &= Y_t + \mu \Delta t + \sigma X_{t+1} \\ v_{t+1} &= v_t + \kappa(\eta - v_t)\Delta t + \sigma_v \sqrt{v_t \Delta} \epsilon^v_{t+1}, \end{aligned} \tag{2.15}$$

  where X_{t+1} follows $F((\Delta v_{t+1})^{\frac{1}{\alpha}})$ where $F(\cdot)$ is the LS distribution. The latent variables consist of $(v_t)_{t=0}^T$ and jump sizes $(L_t)_{t=1}^T$. The model parameters are $\mathbf{\Theta} = \{\mu, \kappa, \eta, \sigma_v, \sigma, \alpha\}$.

2.3 An empirical study of time-changed Lévy processes with MCMC

After figuring out how to estimate time-changed Lévy processes, we want to provide some numerical examples in order to demonstrate how accurate the MCMC method can be.

With simulation data, we also investigate whether it is necessary to use time-changed Lévy processes to model asset prices. In others words, we try to show whether basic Lévy processes such as the VG process and the CGMY process can approximate the behaviour of time-changed Lévy processes. This procedure is similar to the work in Li et al. (2008). Our focus is to distinguish between time-changed Lévy processes and basic Lévy processes. The reason why we do not use compound Poisson process in experiments is that it has been discussed in Li et al. (2008). In the first subsection, we presents examples of estimation based on simulation data. Then we investigate whether time-changed Lévy processes can approximate the behaviour of 'simple' models such as the Merton's Jump-diffusion model and the VG model.

2.3.1 Performance of MCMC estimation

In this subsection, we provide a simulation-based experiment to show that the MCMC estimation method can provide accurate estimates for a variety of models. Models used in our experiments include HJ, HVG, SVVG and SVLS. In fact, estimating the HJ model has been fully discussed in existing literature such as Eraker et al. (2002) and Li et al. (2008). It is still a challenging job to estimate time-changed Lévy models. Following the same procedure in Li et al. (2008), we generate 100 samples of 20 years' daily data. The sampling algorithm of the LS model is provided in Appendix A.2. For each sample, we apply the MCMC estimation method and produce the best estimates.

The first thing we need to do is to check if the MCMC method proposed is reliable. A

common approach is estimating several benchmark models based on simulation data as the true values of parameters are known. We select four benchmark models including HJ, HVG, SVVG and SVLS. The HJ model is a Jump-diffusion model with stochastic volatility which is popular in derivatives pricing. The other three models are the real challenge we want to solve. For each model, we generate 100 samples consisting of daily data with the horizon of 20 years. Simulating the HJ model is simple as it only involves Normal distribution and Poisson distribution. Simulating the HVG model and the SVVG model are also straightforward as VG distribution can be understood as a time-changed drifted Brownian motion. It is not so straightforward how to simulate the SVLS model; we provides the detailed algorithm in Appendix A.2. For each sample simulated, we apply the MCMC method and obtain the best estimates. We use $40,000$ iterations for each MCMC estimation and discard the first $20,000$ iterations as the 'burn-in' time. The overall optimal estimate is the mean of all 100 best estimates. The dispersion of estimation results is measured in Root-Mean-Square Deviation (RMSE). All results are reported in Table 2.1.

The estimation results presented in Table 2.1 are very accurate as each estimate is close to the associated true value. In fact, the estimation results of the SVVG model and SVLS model are slightly weaker than that of the HJ model. It might be due to the fact that estimating time-changed Lévy processes involves sampling with uncommon distributions. Hence, the convergence is not as good as that which only involves sampling from the Normal distribution or the Poisson distribution.

Table 2.1: *Estimation results on simulation data for three models*

This table presents the estimation results of HJ, HVG, SVVG, SVLS model based on simulation data. For each model, 100 samples of daily data are simulated with the horizon of 20 years. For each sample, the estimates of parameters are generated with the MCMC method. The best estimate is the mean value of all estimates. For each panel, the true parameters predetermined are reported in the first row. The mean values of estimates of parameters are reported in the second row. The dispersion of estimates is measured by RMSE and reported in the third row.

Panel A: HJ model

	μ	κ	η	σ_v	ρ	μ_y	σ_y	λ_y
True	0.05	1.50	0.36	0.25	-0.60	-3.00	2.50	0.02
Estimate	0.0510	1.5231	0.3589	0.2520	-0.5982	-3.0334	2.2353	0.0221
RMSE	0.0112	0.026	0.0231	0.0018	0.0101	0.0432	0.0582	0.0034

Panel B: HVG model

	μ	κ	η	σ_v	ρ	θ	σ	ν
True	0.05	1.50	0.36	0.25	-0.60	-0.05	0.25	2.5
Estimate	0.0504	1.5331	0.3643	0.2487	-0.5896	-0.0489	0.2513	2.5021
RMSE	0.0089	0.0342	0.0533	0.0135	0.0343	0.0065	0.0034	0.0532

Panel C: SVVG model

	μ	κ	η	σ_v	θ	σ	ν
True	0.05	1.50	0.36	0.25	-0.05	0.25	2.5
Estimate	0.0491	1.5213	0.3598	0.2481	-0.503	0.2479	2.4891
RMSE	0.0079	0.0283	0.0481	0.0123	0.0068	0.0362	0.0682

Panel D: SVLS model

	μ	κ	η	σ_v	α	σ
True	0.05	1.50	0.36	0.25	1.85	0.25
Estimate	0.0493	1.5236	0.3581	0.2513	1.8483	0.2498
RMSE	0.0203	0.0483	0.0086	0.0068	0.0103	0.0048

It is believed that the MCMC estimation method developed in this chapter is reliable and accurate for the purpose of estimating time-change Lévy processes. We use two representative time-changed Lévy models: the SVVG model and the SVLS model. The results are accurate since the MCMC method can generate accurate samples of all latent variables. The multi-dimensional problem does not seem to be a burden. We demonstrate how to obtain accurate estimates of time-changed Lévy models with the MCMC method. Hence, it becomes feasible to investigate whether time-changed Lévy models can provide a better performance in terms of modelling asset returns.

2.4 Empirical results of time-changed Lévy processes

This section will provide a series of empirical results, to show the advantages of using time-changed Lévy processes in modelling returns. We carefully select the data set used in this estimation experiment. MSCI indices are managed by Morgan Stanley, and are often used as common benchmark portfolios for assessing funds' performance. We choose the World Index which captures large and mid cap representations across 23 developed Markets and has been maintained since 1969.

Table 2.4 reports the summary statistics of daily log returns of the MSCI World index. The sample data is collected from 2nd May 2001 to 31st December 2012. There are 3043 observations of returns reported in percentages in Table 2.4. The sampling period chosen consists of both booming times and the credit crunch. Figure 2.1 depicts the

time series plot of the MSCI World Index. We can observe the bearish period after the dot com bubble and the credit crunch period since 2008. There are also two booming periods such as the recovering time after the crunch. The time series plot of log returns presented in Figure 2.2 reveals the fact that the sampled returns are very volatile and exhibit volatility clustering. The largest daily return is 9.10% on 13th October 2008 and the biggest negative daily return is 7.32% on 15th October 2008.

Table 2.2: *A summary of statistics of daily log returns of the MSCI World Index*

This table reports the summary statistics of log returns of the MSCI World Index from 2nd May 2001 to 31st December 2012. The continuously compounded returns are calculated as the log difference successive index levels.

	Mean	Std	Skewness	Kurtosis	Max	Min
MSCI World	0.0141	1.1139	-0.3219	10.4325	9.0975	-7.3169

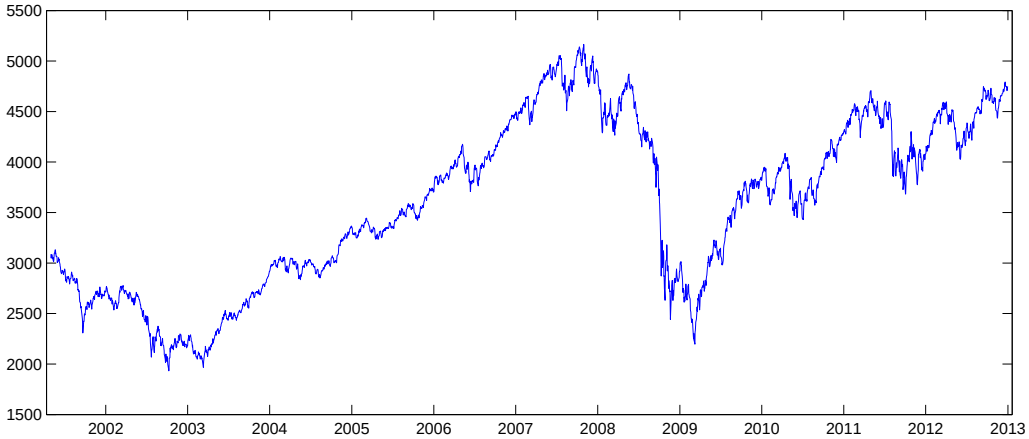

Figure 2.1: *Time series plot of the MSCI World Index from 02/05/2001 to 31/12/2012*

In the experiment, besides HJ and HVG model, we estimate two time-changed Lévy

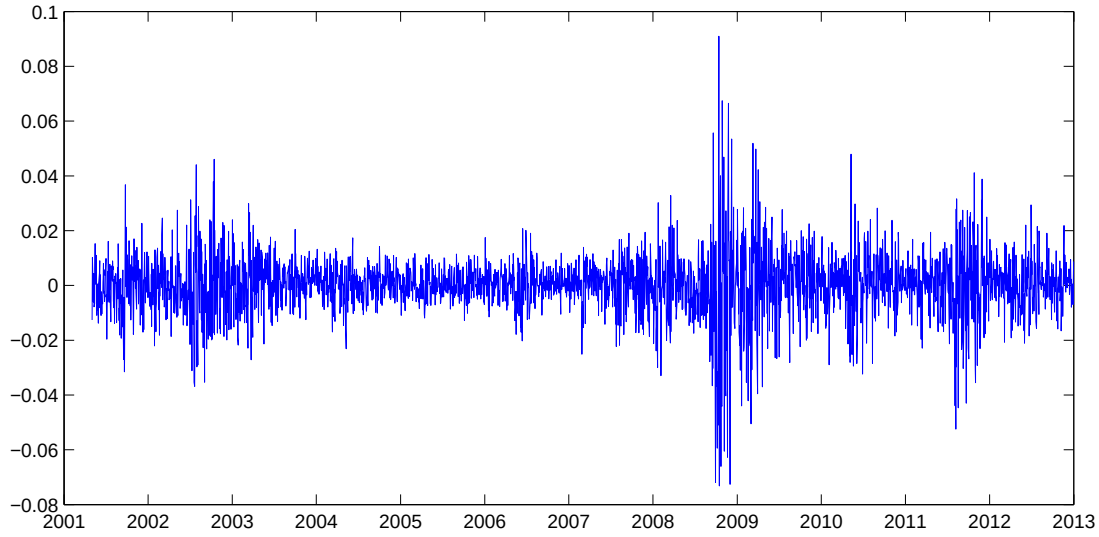

Figure 2.2: *Time series plot of log returns of the MSCI World Index from 02/05/2001 to 31/12/2012*

models, SVVG and SVLS, with the MCMC method presented in Section 2.2. The sample data is the log returns of daily MSCI World index. As described before, the best estimate is the expectation of the posterior distribution of the underlying parameter. We will sample from the posterior distribution and use the empirical mean as a good approximation to the expected value. For each estimation procedure, the sample size is $20,000$. There is a 'burn-in' problem for the MCMC estimation method. This is because the MCMC method relies on the fact that the stationary distribution can be obtained by using the transition distribution. Hence, we need to 'wait' for some time before reaching the stable state. We use $20,000$ as the 'burn-in' sample size, which means we will generate $40,000$ samples and only use the last $20,000$ samples to take the mean and standard deviation as the point estimate and standard error.

We report the estimated parameters in Table 2.3. Model used in the estimation include the Heston-Jump (HJ) model, the Heston-VG Jump-diffusion (HVG) model, the time-changed Variance Gamma (SVVG) model, and the time-changed Log Stable (SVLS) model. Estimates of parameters are reported with the associated standard errors which are presented in parentheses. The expected mean estimated has a similar level as the empirical mean which is 0.0141 reported in Table 2.4. It should not be the same as the innovation distribution does not guarantee the Martingale property. The volatility level also exhibits a similar level. It is interesting to check the jump intensity, based on the estimation results. The jump intensity of the HJ model is about $0.0056 * 252 = 1.4112$ per year. Apparently, the jump intensity is far away from reality. This indicates the necessity of using infinite activity Lévy processes. The average level of jump parts tends to be negative as we observe $\mu_y = -2.3140$. It is suggested that negative jumps are more frequent and important. Hence, it also suggests that we should use the LS process as this type of Lévy process only admits negative jumps with positive drift. The standard errors obtained are very small except for those parameters of the jump structure of the HJ model. This problem has also been reported in Eraker et al. (2002) and Li et al. (2008). The estimates of Lévy models are more accurate, which demonstrates that the MCMC method has a very good performance on estimating infinite activity Lévy processes.

To assess the performance of the estimation, we collect residuals from the last 100 iterations of each round of the MCMC estimation. The residuals should follow the standard Normal distribution if the model can fit the data accurately. Hence, the performance of

Table 2.3: *Estimation results of the MSCI World Index*

This table reports the estimation results given by the MCMC estimates method. The sample data is the MSCI World Index collected from 02/05/2001 to 31/12/2012. Model used in the estimation include the Heston-Jump (HJ) model, the Heston-VG Jump-diffusion (HVG) model, the time-changed Variance Gamma (SVVG) model, and the time-changed Log Stable (SVLS) model. Estimates of parameters are reported with the associated standard errors which are presented in parentheses.

HJ		HVG		SVVG		SVLS	
μ	0.0135	μ	0.0217	μ	0.0327	μ	0.0311
	(0.0010)		(0.0009)		(0.0012)		(0.0013)
κ	0.3541	κ	0.3125	κ	0.6421	κ	0.5982
	(0.0055)		(0.0036)		(0.0052)		(0.0063)
η	0.8461	η	0.8762	η	0.9874	η	0.9012
	(0.0832)		(0.0810)		(0.1032)		(0.0961)
σ_v	0.1235	σ_v	0.1323	σ_v	0.1459	σ_v	0.1682
	(0.0013)		(0.0016)		(0.0021)		(0.0091)
ρ	-0.3132	ρ	-0.3562	σ	0.5434	α	1.7654
	(0.0201)		(0.0108)		(0.0042)		(0.0014)
μ_y	-2.3140	σ	0.4325	θ	-0.1754	σ	0.7658
	(0.8931)		(0.0138)		(0.0132)		(0.0047)
σ_y	3.1596	θ	-0.1340	ν	4.9812		
	(1.1310)		(0.0063)		(0.1302)		
λ_y	0.0056	ν	5.3225				
	(0.0021)		(0.0391)				

estimation can be evaluated by testing how close the residuals are away from $N(0,1)$. In Figure 2.3, the Kernel Density estimator of the HJ model is presented with the corresponding QQ plot. Obviously, the HJ residuals are not normally distributed. The main differences appear to be the high moments and tail behaviour. According to the QQ plot in Figure 2.3, the left tail is extremely away from the Normal predication, which suggests that the HJ model has a poor performance on fitting the negative jumps exhibited from the market. Figure 2.4, 2.5 and 2.6 present the Density estimators and QQ plots for HVG, SVVG and SVLS models, respectively. Actually, the HVG model provides the best fit; the other two time-changed Lévy models also show good fit, but the tail behaviours are slightly different from Normal distribution. As discussed before, the time-changed Lévy models do not admit the leverage-effect. Hence, the tail behaviour might be slightly different. The correlation between the spot price and the volatility is about -0.3562, which is not very strong. Our estimation results of time-changed Lévy models might become worse if a different data set with high leverage-effect is used.

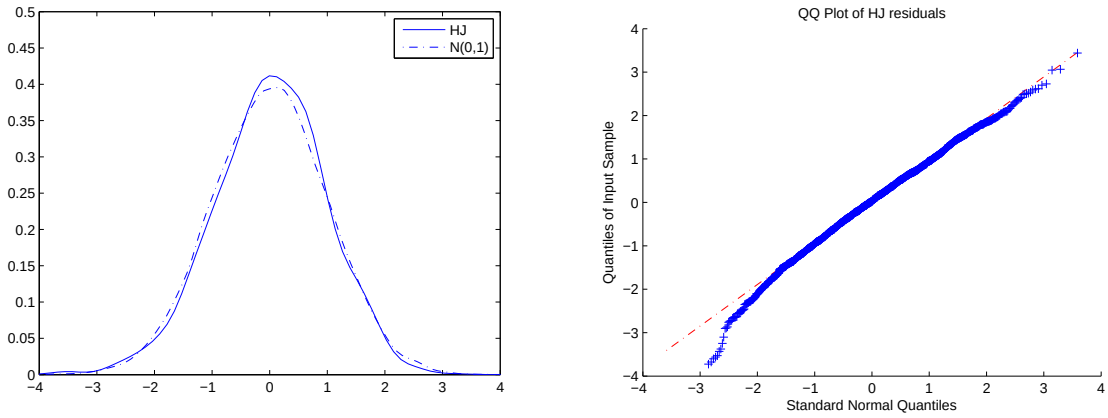

Figure 2.3: *Kernel density and QQ plot of the residuals of the HJ model*

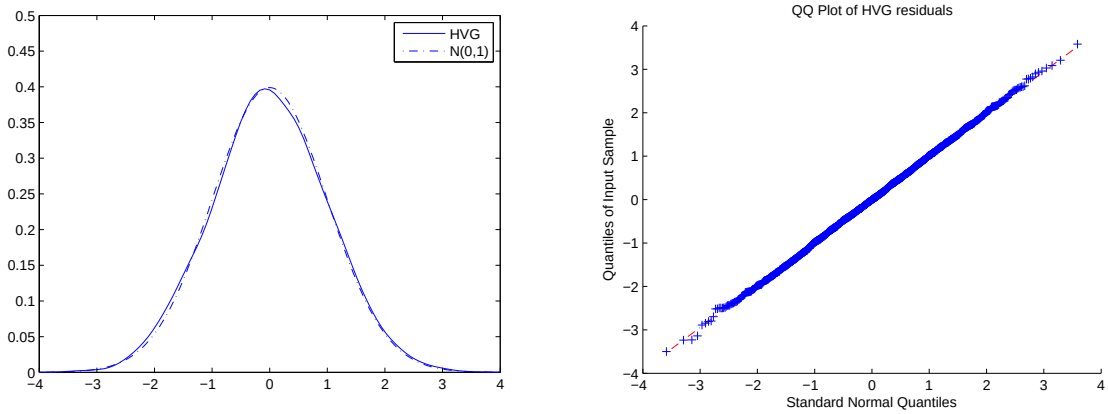

Figure 2.4: *Kernel density and QQ plot of the residuals of the HVG model*

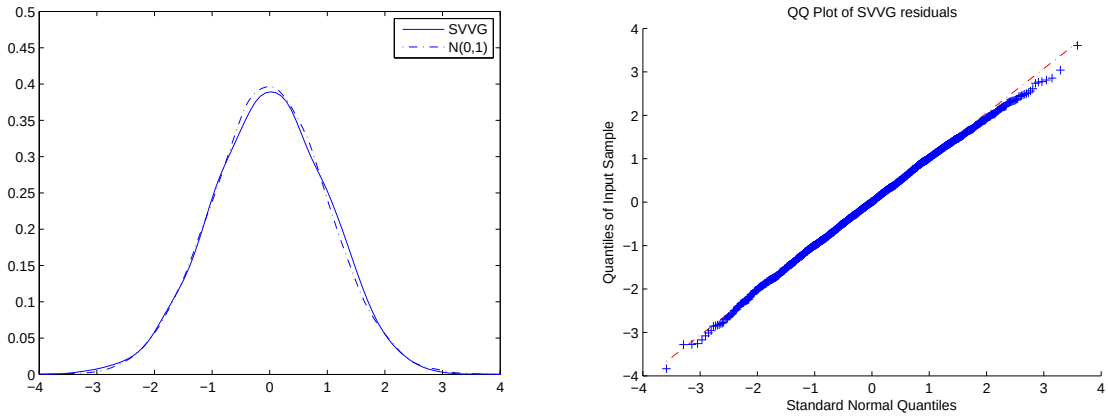

Figure 2.5: *Kernel density and QQ plot of the residuals of the SVVG model*

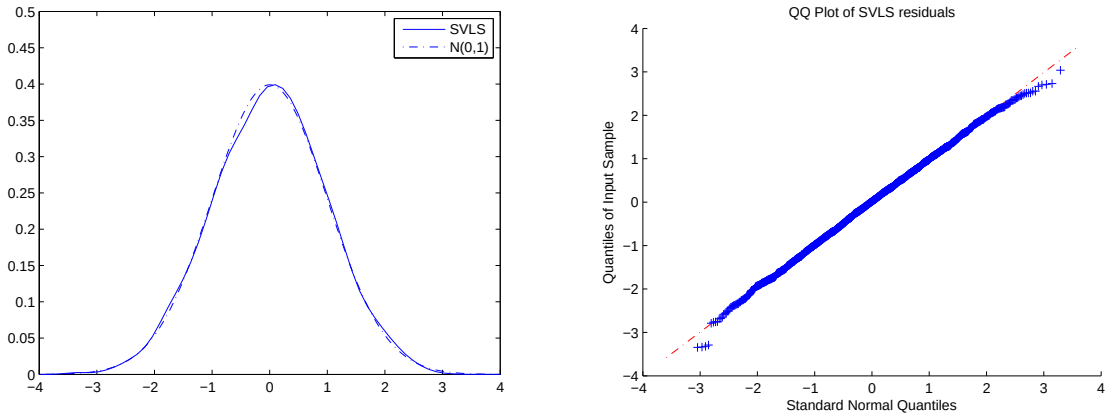

Figure 2.6: *Kernel density and QQ plot of the residuals of the SVLS model*

Apart from the graphical evaluation, a statistical test is also needed. A simple choice is the One-sample Kolmogorov-Smirnov test. We apply the Kolmogorov-Smirnov test to the final 100 iterations to check whether residuals follow the standard Normal distribution. In Figure 2.4, percentage of rejections is reported in the second column and the associated mean value of p-value is reported in the third column. For the HJ model, the KS test reject 96% of sets of residuals. The average p-value is only 0.0086. This result is similar to that in Li et al. (2008). Apparently, the HJ model cannot capture the behaviour of index returns. This is consistent to the results observed in Figure 2.3. The winner, as shown in Table 2.4, is the HVG model which has only 11 sets of residuals rejected in the KS test. The residuals produced by the HVG model is very close to standard Normal. The two time-changed Lévy models also provide good fit. For instance, 31 sets of residuals produced by the SVVG model are rejected while only 23 sets are rejected for the SVLS model. The corresponding mean p-values are 0.2781 and 0.3539, respectively. The above results demonstrate that infinite activity Lévy models outperform Jump-diffusion model

in capturing return dynamics. Pure time-changed Lévy models are slightly weaker than the stochastic volatility models with infinite Lévy jumps. This might be due to the absence of the Leverage-effect. If time-changed Lévy models subordinated by pure Lévy jumps are used, it is possible to achieve better performance. It is also suggested that the LS process can provide good performance in modelling index returns.

Table 2.4: *Empirical results of Kolmogorov-Smirnov Test*

We test four models including HJ, HSV, SVVG, and SVLS, based on the daily data of the MSCI World index. For each of the final 100 iterations, the model residuals are collected based on the estimated model parameters. We apply the Kolmogorov-Smirnov test to the final 100 iterations to check whether residuals follow the standard Normal distribution. Percentage of rejections is reported in the second column and the associated mean value of p-value is reported in the third column.

	percentage of rejection(%)	mean p-value
HJ	96	0.0086
HVG	11	0.4675
SVVG	31	0.2781
SVLS	23	0.3539

Due to the advantage that MCMC estimation can simulate latent variables, we can also have a look at the estimated latent variables such as volatility to check the accuracy of fit. In Figure 2.7, the estimated volatility plots are depicted for the four models used in our experiment. According to the return time series plot of the index shown in Figure 2.2, the most volatile period is between 2008 and 2009 when the credit crunch happened. Our estimated volatilities coincide with the observation. All the four models provide similar

variation of volatilities. We can observe extreme large jumps between 2008 and 2009 as well as in late 2011.

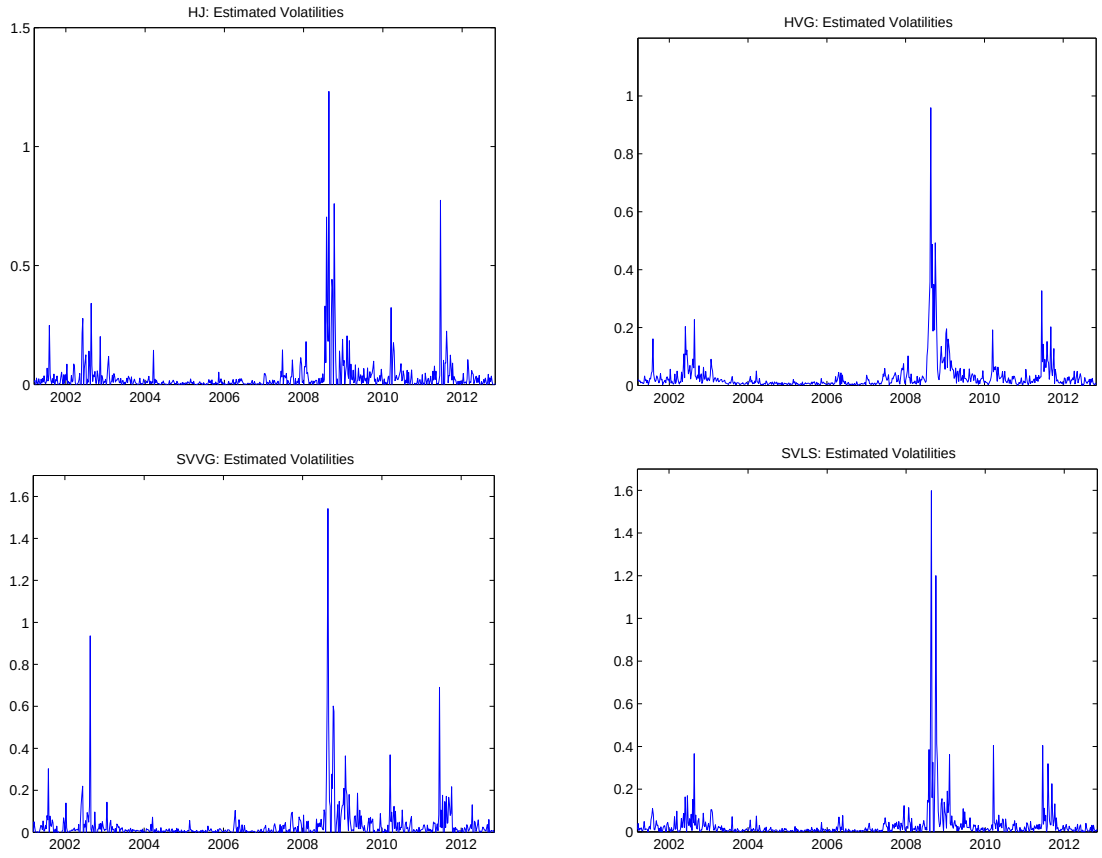

Figure 2.7: *Estimated volatilities of HJ, HVG, SVVG, and SVLS models*

This empirical experiment shows the benefit of adopting infinite activity Lévy processes in modelling index returns. The HVG, SVVG and SVLS model all show very good accuracy of fit. Although time-changed Lévy models do not outperform the stochastic volatility model with infinite activity Lévy jumps (HVG), this might be due to the absence of the Leverage-effect. For the index return data we choose, the correlation between the spot price and the volatility is not very strong. Hence, the difference on fitting between

the HVG model and other two time-changed Lévy models is not very significant. It seems that the only reason we should keep diffusion is because we want to generate the Leverage-effect. If proper time-changed Lévy models that admit the Leverage-effect can be developed, the diffusion component can be dropped without losing good fit.

2.5 Conclusion

Continuous-time models have been widely used in capturing the return dynamics of individual stocks and indices. The jump-diffusion model plays an important role for portfolio allocation problem as it can model both small and large movements of return series. However, the jump-diffusion models cannot achieve accurate fit of market returns. Time-changed Lévy processes provide a much broader and more flexible class of models for capturing asset price dynamics. This chapter investigates the empirical advantages of using Lévy processes for modeling index returns. The MCMC techniques developed in this chapter is a further extension of the MCMC method for stochastic volatility models in Li et al. (2008). The advantage of infinite activity Lévy processes is allowing an infinite number of jumps within any finite time interval. Hence, all jumps or movements of returns with different sizes can be captured by time-changed Lévy processes, especially for highly frequent discontinuous movements in stock prices. The MCMC estimation method can approximate model parameters, latent variables and jump variables of time-changed Lévy processes. Based on empirical analysis, we show that time-changed Lévy models can outperform traditional Jump-diffusion models in capturing both returns and volatility. The

diffusion component can be omitted as infinite activity Lévy processes can capture all sizes of movements including the small ones that diffusion can model.

Chapter 3

An Empirical Study of Multivariate MCMC Estimation on Lévy processes

In Chapter 2, we have provide a detailed investigation on estimating the one-dimensional time-changed Lévy processes with the Markov Chain Monte Carlo (MCMC) technique; However there is still a gap in this field. There is no efficient method that has been applied for multi-dimensional processes. The extreme complexity of high-dimensional estimation is the huge obstacle. Wether or not Lévy processes should be used for portfolio problem can only be answered with the development of estimation technique.

An efficient estimation procedure for multivariate Lévy processes is proposed and im-

plemented. The interest in multidimensional asset models based on Lévy processes is motivated by the importance of capturing market shocks using more refined distribution assumptions compared to the standard Gaussian framework, incorporating skewness and kurtosis. From a risk management perspective, in fact, the focus is specifically on the tails of the stock return distribution, and commonly used risk measure such as Value at Risk and intra-horizon Value at Risk aim at quantifying the economic impact of rare events. Further, for regulatory purposes these risk measures are usually obtained for short time horizons (i.e. 10 days), over which the effects of stochastic volatility are in general negligible (mainly due to the diffusive nature of the processes used for the modelling of volatility trends). In this respect, Lévy processes can offer a natural and robust approach to model distribution tails compared to the Brownian motion, especially over the short period, as they allow for extreme outcomes to happen more frequently. However, consistent and efficient estimation procedures, which are essential part of the calculation of relevant risk measures, can be problematic for Lévy processes as extensively documented in Cont and Tankov (2004). For example, these issues are exacerbated by increasing the dimension of the parameter space, which would be necessary in order to accommodate for the multivariate modelling required at portfolio level.

Linear transformations have been used extensively in the literature to build multivariate Lévy processes as these processes are invariant under such a transformation, and therefore their characteristic function and characteristic triplet can be obtained in a straightforward manner (see Cont and Tankov (2004) for example). Thus, the standard approach is to

model each risk driver as a linear combination of two independent processes representing respectively the systematic factor and the idiosyncratic shock, so that dependence between assets in a given portfolio is originated by the common component of the overall risk. Contributions based on linear transformations started with Vasicek (1987) for the case of Brownian motions; for the extension to Lévy processes we mention, Ballotta and Bonfiglioli (2014) and Luciano and Semeraro (2010). In more details, Luciano and Semeraro (2010) apply the factor approach to the asset log-returns process, which allows to choose any Lévy process as factor processes, and encompasses any class of Lévy processes, from subordinated Brownian motions to jump-diffusion processes. Linear transformations have also been used in the literature to build multivariate subordinators and therefore alternative multivariate versions of subordinated Brownian motions; this is the case of Luciano and Semeraro (2010) who offer a general construction for subordinated Brownian motions, such as the Normal Inverse Gaussian (NIG) and the Carr-Geman-Madan-Yor (CGMY) processes. Extensions to a factor-based subordinated Brownian motion are proposed by Luciano (2013) in order to incorporate additional dependence properties.

A common trait to all these contributions is the presence of (either explicit or implicit) convolution conditions, which allow to separate the dependence structure from the distribution of the margin processes. However, as argued by several authors such as Eberlein et al. (2008), this feature, although intuitive, leads to a biased view of the dependence in place as it reduces the flexibility of the factor model, and fails to recognize the different tail-behaviour shown by any portfolio component. In this respect, we notice that in the

model of Ballotta and Bonfiglioli (2014) these convolution conditions are not necessary for the model to retain its mathematical tractability, its relative flexibility in accommodating a wide range of dependence structures, positive and negative linear correlation and a parsimonious number of parameters.

As we try to fill the gap, we propose an efficient estimation procedure for the multivariate (exponential) Lévy processes model of Ballotta and Bonfiglioli (2014) with risk management applications type in view, specifically the computation of Value at Risk and intra-horizon Value at Risk for portfolios of dependent assets. Thus, we focus on the model estimation under the physical probability measure, which is in fact non-trivial as the common and the idiosyncratic factors driving the margins are not directly observable in the market. In order to simplify this problem, although other approaches are possible, here we follow standard market practice and assume that the common factor representing systematic risk can be well-proxied by the returns on a broad-based index.

The first main contribution of this chapter is the robust MCMC estimation procedure for the multivariate Lévy processes model. Simulation study shows that this approach can provide very accurate joint identification of Lévy jumps and stochastic volatility. we demonstrate the accuracy and stability of the MCMC method based on simulation data. This enables us to explore whether time-changed Lévy model can generate better fit, compared with existing stochastic Jump-diffusion models. To assess this estimation procedure, we also implement a standard one-step maximum likelihood approach in which all parameters of the multivariate Lévy process are estimated in a single step. Another

contribution is that we assess the goodness of fit with market data and provide an answer to whether Lévy processes should be adopt in order to modelling return dynamics. Li et al. (2008) suggest that Lévy jump models can outperform AJD models in capturing the time series dynamics of the S&P 500 index returns. Using two widely used multivariate Lévy models, we apply the estimation procedure proposed and achieve very good results. A further work should be done by using Lévy models subordinated by pure jump processes which admits the Leverage-effect in estimation experiment, in order to show whether pure Lévy models can outperform stochastic volatility models with Lévy jumps in capturing the dynamics of index returns.

The outline of the chapter is as follows. In Section 3.1, we review and specify the most relevant features of the multivariate Lévy model under consideration. In Section 3.2 we discuss the estimation method of the models in details. In Section 3.3 we assess the MCMC estimation procedure via simulation for two particular specifications of the model (the NIG and the Merton jump diffusion process), comparing the results with those obtained via a one-step maximum likelihood estimation. Section 3.4 concludes.

3.1 Lévy Model Specification

A Lévy process on a filtered probability space is a stochastic process characterized by independent and stationary increments whose distribution is infinitely divisible. Lévy processes have attracted attention in the financial literature due to the fact that they ac-

commodate distributions with non-zero higher moments (skewness and kurtosis), therefore allowing a more realistic representation of the stylized features of market quantities such as assets returns. Further, they represent a class of processes with known characteristic function via the celebrated Lévy-Khintchine representation; this feature in particular allows the development of efficient numerical schemes for the approximation of potentially unknown distribution functions and derivatives prices based on Fourier inversion techniques.

Let us denote by S_t the price of a financial asset. In the class of exponential-Lévy models, the price S_t is represented as

$$S_t = S_0 \exp(L_t)$$

where L is a Lévy process, with characteristic function $E(\exp(iuL_t)) = \exp(t\psi(u))$, where ψ denotes the so-called characteristic exponent. Assuming that we observe the price process on an equally-spaced time grid $t = 1, 2, \ldots, T$, the log-returns, defined as

$$X_t = \log\left(\frac{S_t}{S_{t-1}}\right) = L_t - L_{t-1}$$

are i.i.d. infinitely divisible random variables distributed as L_1.

A convenient representation of multivariate Lévy processes can be obtained via linear transformation of a vector of independent Lévy processes, each representing the idiosyncratic risk, and another independent Lévy process modelling the common risk component. The construction of Ballotta and Bonfiglioli (2014) is based on this principle and is summarized in the following.

Proposition 1 Let Z, $Y^{(j)}$, $j = 1, \ldots, n$ be independent Lévy processes, with characteristic function $\phi_Z(u,t)$ and $\phi_Y(u,t)$, for $j = 1, \ldots, n$, respectively. Then for $a_j \in R$, $j = 1, \ldots, n$

$$X_t = (X_t^{(1)}, \ldots, X_t^{(n)})' = (Y_t^{(1)} + a_1 Z_t, \ldots, Y_t^{(n)} + a_n Z_t)' \tag{3.1}$$

is a Lévy process on R_n with characteristic function

$$\phi_X(u,t) = \phi_Z \left(\sum_{j=1}^{n} a_j u_j, t \right) \prod_{j=1}^{n} \phi_Y(j)(u_j, t), u \in R^n \tag{3.2}$$

Further, the joint probability density function of the multivariate Lévy process X_t is

$$f_X(x_t^{(1)}, \ldots, x_t^{(n)}) = \int_{-\infty}^{\infty} f_{Y^{(1)}} \cdot \ldots \cdot f_{Y^{(n)}}(x_t^{(n)-a_n z}) f_Z(z) dz \tag{3.3}$$

We note that as the given multivariate Lévy model admits computable characteristic function, the joint distribution is always available (at least numerically), even when the components' distributions, $f_Y(1), \ldots, f_Y(n)$, f_Z, are not known analytically.

It follows from Proposition 1 that for each $X^{(j)}, j = 1, \ldots, n$, the process Z captures the systematic part of the risk originated by sudden changes affecting the whole market, while the process $Y^{(j)}$ represents the idiosyncratic shocks generated by company specific issues. Due to the presence of the common factor Z, the components of X are dependent and may jump together. In particular, for each $t \geq 0$, the components of $X(t)$ are positive associated if the loading factors a_j for $j = 1, \ldots, n$ are all positive or negative; otherwise the components of $X(t)$ are negative quadrant dependent. In any case, the dependence

between components is correctly described by the pairwise linear correlation coefficient

$$\rho^X_{j,l} = Corr(X^{(j)}_t, X^{(l)}_t) = \frac{a_j a_l Var(Z_1)}{\sqrt{Var(X^{(j)}_1)}\sqrt{Var(X^{(l)}_1)}} \tag{3.4}$$

as, for fixed $a_j, a_l \neq 0, \rho^X_{j,l} = 0$ if and only if Z is degenerate and the components are independent, whilst $|\rho^X_{j,l}| = 1$ if and only if $Y^{(j)}$ and $Y^{(l)}$ are degenerate, i.e. there is no idiosyncratic factor in the components $X^{(j)}$ and $X^{(l)}$. Further, $\text{sign}(\rho^X_{j,l}) = \text{sign}(a_j a_l)$, therefore both positive and negative correlations can be accommodated. Finally, we mention that the resulting multivariate model shows non-zero indices of tail dependence, the sign being controlled by the loading parameters. For fuller details on the characteristic triplet of the multivariate process and the dependence structure, we refer to Ballotta and Bonfiglioli (2014) and Ballotta et al. (2015).

We note the following. In first place, this construction is relatively parsimonious in terms of number of parameters involved as this grows linearly with the number of assets. Further, the adopted modeling approach is quite flexible as it can be applied to any Lévy process; Proposition 1 allows to specify any univariate Lévy process for Y_t and Z_t. In this respect, we note that differently from Ballotta et al. (2015), in this work we do not impose any convolution condition on the components aimed at recovering a known distribution for the margin processes, hence allowing for a more realistic portrayal of the asset log-return features and the dependence structure in place.

Finally, the model is particularly tractable as the full description of the multivariate vector X_t only requires information on the univariate processes Y_t and Z_t. For the purpose of the testing of the estimation procedure introduced in the next sections, we select two

alternative classes of Lévy processes commonly used for financial applications: a subordinated Brownian motion, represented by the NIG process, and a jump diffusion process with Gaussian severities as in Merton (1976), which we briefly review for completeness.

3.1.1 The Normal inverse Gaussian process (NIG)

The NIG model, introduced by Barndorff-Nielsen (1997), is a normal tempered stable process obtained by subordinating a Brownian motion by an (unbiased) independent Inverse Gaussian process. Its characteristic function reads

$$\phi_X(u,t) = \exp\left(i\mu t + \frac{t}{k}(1 - \sqrt{1 - 2iu\theta k + u^2\sigma^2 k})\right), u \in R \tag{3.5}$$

It follows by differentiation of the (log of the) characteristic function that the first four cumulants of X_t are

$$\begin{aligned} c_1 &= (\mu + \theta)t, \qquad c_2 = (\sigma^2 + \theta^2 k)t, \\ c_3 &= 3\theta k(\sigma^2 + \theta^2 k)t, \qquad c_4 = 3k(\sigma^4 + 6\sigma^2\theta^2 k + 5\theta^4 k^2)t \end{aligned} \tag{3.6}$$

From the above, we observe that θ controls the sign of the skewness of X_t, σ affects the overall variability and k controls the kurtosis of the distribution. The drift parameter μ affects the mean of the distribution, which otherwise would be concordant with the skewness, allowing to model return distributions with positive mean and negative skewness as well (and vice versa). Finally, the tails for the distribution are characterized by a power-modified exponential decay, or semi-heavy tail (see Cont and Tankov (2004), for example).

As the density function is known in semi-closed form (as it is expressed in terms of the modified Bessel function of the second kind, see Cont and Tankov (2004) for example), the parameters of the NIG model can be estimated directly using Maximum Likelihood (ML) estimation, initialized via the method of moments based on the first four theoretical cumulants derived above.

3.1.2 The Merton jump-diffusion process (MJD)

A Lévy jump-diffusion process has the form

$$X_t = \mu t + \sigma W_t + \sum_{i=1}^{N_t} J_i$$

where W is a standard Brownian motion, N is a Poisson process with rate $\lambda > 0$ counting the jumps of X, and $\{J_i\}_{i \in N}$ are i.i.d. random variables describing the jump sizes. All the random quantities involved, W, N and J_i (for all i), are assumed to be mutually independent. In the Merton's jump-diffusion model (Merton, 1976) jump sizes are all normally distributed, i.e. $J_i \in N(\nu, \tau^2)$ for all i.

It follows that the characteristic function is

$$\phi_X(u) = \exp\left(iu\mu t - \frac{u^2\sigma^2}{2} + \lambda t\left(e^{iu\nu - \frac{\tau^2 u^2}{2}} - 1\right)\right), u \in R \tag{3.7}$$

The first four cumulants of X_t are

$$c_1 = (\mu + \lambda\nu)t, c_2 = (\sigma^2 + \lambda(\nu^2 + \tau^2))t,$$

$$c_3 = \lambda\nu(3\tau^2 + \nu^2)t, c_4 = \lambda(3\tau^4 + 6\tau^2\nu^2 + \nu^4)t \tag{3.8}$$

We can observe how the parameters λ, ν and τ control the non-Gaussian part of the process; in particular, λ controls the sign of skewness (the density function is symmetric when $\nu = 0$), whilst λ governs the jumps frequency and therefore the level of excess kurtosis. We note that X_t has an infinite Gaussian mixture distribution with mixing coefficients given by a Poisson distribution with parameter λ; hence, the probability density function can be expressed as a fast converging series. Further the tails are heavier than in the pure Gaussian case (see Cont and Tankov (2004), for example).

We note that the estimation of the MJD model is far from trivial as the ML method requires a careful numerical optimization, as discussed in Honoré (1998). Consequently, in the numerical study we implement the Expectation Maximization (EM) algorithm in the formulation proposed by Duncan et al. (2009), which has simple closed form solutions for the M-step.

3.2 Model Estimation

For the purpose of the approach to the estimation of the given multivariate Lévy model, we distinguish between whether the common factor is observable or not. The latter case is considered in Section 3.3, where we show how the computation of the sample likelihood function is possible once we integrate out the common factor. However, the maximization of the likelihood in this case turns out to be feasible only if we consider a limited number of assets in our portfolio. A second possibility would be to consider

the unobservable common factor as a latent factor whose dynamic is assigned, so that the estimation procedure can be reduced to a (in general) non-Gaussian Kalman filtering problem. However, the application of these techniques is in general not straightforward and, in any case, they do not solve the dimensionality problem.

In this section, we describe an MCMC algorithm for the continuous-time model to estimate the parameters Q, B, and Σ given return data, $V = (V_m)_{m=1,\ldots,N}$ observed at fixed observation times Δt, $2\Delta t$, ..., $N\Delta t = T$. This method is easily extended to deal with non-equidistant observations; We allow for jumps of the hidden state process at any time and especially for any number of jumps within each observation interval.

3.2.1 Data Augmentation

The state process Y, which is allowed to jump any time, is described by the process of jump times, $J = (J_h)_{h=0,\ldots,H}$, and the sequence of states visited, $Z = (Z_h)_{h=0,\ldots,H}$, where H is the number of jumps of Y in $[0, T]$, i.e., $J_0 = 0, Z_0 = Y_0$, J_h is the time of the hth jump, and Z_h is the state Y jumps at the hth jump. Hence the inter-arrival time $\Delta J_h = J_h - J_{h-1}$ is exponentially distributed with parameter λZ_{h-1}. Note that J_{h+1} and Z_{h+1} are independent given J_h and Z_h. For parameter estimation, we augment the parameter space by adding the state process Y, and determine the joint posterior distribution of Q, B, Σ, and Y given the observed data V.

3.2.2 Prior Distributions

Prior distributions have to be chosen for all the parameters Q, B, Σ, and Y_0. We consider two prior specifications for simplification purpose, differing in the prior assumptions concerning the initial state Y_0. One prior is based on assuming prior independence among all parameters, i.e.,

$$\mathbb{P}(Q, B, \Sigma, Y_0) = \mathbb{P}(Q)\mathbb{P}(B)\mathbb{P}(\Sigma)\mathbb{P}(Y_0) \tag{3.9}$$

where $\mathbb{P}(Y_0) = 1/d$. However, if we use time 0 as the beginning of our observations after the process has already run for some time, it may be reasonable to assume that the state process starts from its ergodic probability ω, making Q and Y_0 dependent a priori:

$$\mathbb{P}(Q, B, \Sigma, Y_0) = \mathbb{P}(Q)\mathbb{P}(B)\mathbb{P}(\Sigma)\mathbb{P}(Y_0|Q) \tag{3.10}$$

where $\mathbb{P}(Y_0|Q) = \omega$. Under the second prior, Y given Q is a stationary Markov chain, i.e., $\mathbb{P}(Y_t|Q) = \omega$ for all $t \in [0, T]$.

Concerning the remaining parameters, we assume that the off-diagonal elements of Q as well as the elements of B are a priori mutually independent as are the volatility matrices. Furthermore, for $i = 1, \ldots, n$, we assume

$$Q_{kl} \sim \Gamma(f_{kl}, g_{kl}), mu_i^{(k)} \sim N(m_{ik}, \sigma_{ik}^2), C^{(k)} \sim IW(\Sigma^k, \nu_k) \tag{3.11}$$

With Γ, N and IW we refer to the Gamma or Normal distributions, respectively. We use the notation $m = (m_{1k}, \ldots, m_{nk})$ and $\sigma^2 = (\sigma_{1k}^2, \ldots, \sigma_{nk}^2)$ to denote the vectors of prior means and prior variances of $\mu^{(k)}$.

3.2.3 Complete-Data Likelihood Function

As given B, Σ, and Y, the price process S is Markov and the returns $(V_m)_{m=1,\dots,N}$ are independent, the complete-data likelihood function is given by

$$\mathbb{P}(V|Q,B,\Sigma,Y) = \mathbb{P}(V|B,\Sigma,Y)) = \prod \psi(V_m,\mu_m,\sigma_m^2) \tag{3.12}$$

where ψ denotes the density of a multivariate normal distribution with mean vector and covariance matrix given by

$$\begin{aligned} \bar{\mu}_m &= \int_{(m-1)\Delta t}^{m\Delta t} \mu^Y ds, \\ \bar{C}_m &= \int_{(m-1)\Delta t}^{m\Delta t} C^Y ds \end{aligned} \tag{3.13}$$

3.2.4 Proposal Distributions

To sample from the joint posterior distribution of Q, B, Σ, and Y given the observed data V, we partition the unknowns into the blocks Q, $\mu^{(k)}$, $C^{(k)}$, Y, and draw each block from the appropriate conditional distribution.

3.2.5 Choosing the Prior

Although, asymptotically, the hyperparameters of the prior distributions have vanishing influence on the results, they should be chosen with care, as we are dealing with a limited number of observations, in order not to introduce some bias or predetermine the results

too strongly. lightly data- dependent priors can be used to define the prior for the drift and volatility parameters.

Drifts: For the update of (k) for each state k, a Gibbs step can be performed as follows. First, we introduce the notation $B^{-k} = (\mu^{(1)}, \ldots, \mu^{(k?1)}, \mu^{(k+1)}, \ldots, \mu^{(d)})$ and

$$\begin{aligned} \sigma_m^k &= \int_{(m-1)\delta t}^{m\Delta t} I_{Y_s=k} ds \\ \mu_m^{-k} &= \sum_{l=1, l\neq k}^{d} \mu^{(l)} o_m^l \end{aligned} \tag{3.14}$$

Then we have

$$\mathbb{P}(\mu^{(k)}|V, B^{-k}, \Sigma, Y) \sim \psi(\mu^{(k)}, m_{.k}, Diag(s_{.k}^2))\Pi\psi(V_m - \mu_m^{-k}; \mu^{(k)} o_m^k, C_m)$$

and hence $\mu^{(k)}|V, B^{-k}, \Sigma, Y \sim N(a^{(k)}, S^{(k)})$, where

$$\begin{aligned} S^{(k)} &= \left(Diag(s_{.k}^2)^{-1} + \sim_{m=1}^{N} C_m^{-1}(o_m^k)^2\right)^{-1} \\ a^{(k)} &= S^{(k)} \left(Diag(s_{.k}^2)^{-1} m_{.k} + \sum_{m=1}^{N} C_m^{-1}(V_m - \mu_m^{-k}) o_m^k \right) \end{aligned} \tag{3.15}$$

State process: We first consider the full conditional probability distribution $\mathbb{P}(Y|V, Q, B, \Sigma)$. The prior distribution of the state process Y_t for $t > 0$ is determined by the distribution of Y_0 and is independent of B and Σ. Therefore, we obtain

$$\mathbb{P}(Y|V, Q, B, \Sigma) \sim \mathbb{P}(V|B, \Sigma, Y)\mathbb{P}(Y|Q)$$

The probability of Y given Q equals

$$\begin{aligned} \mathbb{P}(Y|Q) &= P(Y_0|Q)\Pi_{h=l}^{H} \left(\lambda Z_{h-1} e^{-\lambda Z_{h_1} \Delta J_h} \frac{Q_{Z_{h-1}}}{\lambda Z_{h-1}} \right) e^{-\lambda Z_H (T - J_H)} \\ &= \mathbb{P}(Y_0|Q)\Pi_{k=1}^{d}\Pi_{l=1, l\neq k}^{d} \left(e^{-Q_{kl} Q_T^k} Q_{kl}^{N_{kl}} \right) \end{aligned} \tag{3.16}$$

where Q_T^k denotes the occupation time of state k, and N_{kl} denotes the number of jumps from k to l,

$$Q_T^k = \int_0^T I_{Y_t=k} dt,$$
$$N_{kl} = \sum_{h=l}^{H} I_{Z_{h-1}=k, Z_h=l}$$

For the update of Y, we draw from the conditional distribution given Q, which simplifies the acceptance probability. To obtain good rates of acceptance, we do not update the whole process at one time but break it into a number of blocks of approximately exponentially distributed length, which are updated independent of each other.

For updating the block $(Y_t)_{t\in[t_0,t_1]}, 0 < t_0 < t_1 < T$, we generate a proposal $(Y'_t)_{t\in[0,t']}, t' = t_1 - t_0$, as follows: First, we set $Z'_0 = Y_{t_0}$. Then we simulate the waiting time until the next jump time and the state the chain jumps. This is repeated until the jump time is greater than t', which is assumed to happen after $H'+1$ steps, i.e., there are H' jumps in $[0, t']$. In order to fit the proposal Y' to Y, we have to consider three cases. If $Z'_{H'} = Y_{t_1}$ we are done. If $Z'_{H'} \neq Y_{t_1}$ and $H' > 0$, we enforce $Z'_{H'} = Y_{t_1}$, possibly removing the last jump, if the chain was in state Y_{t_1} before the jump.

3.3 Empirical Results

After figuring out how to estimate multivariate Lévy processes, we want to provide some numerical examples in order to demonstrate how accurate the MCMC method can be.

3.3.1 Estimation assessment

To assess the effectiveness of the two approaches presented in Section 3.3, we test them through simulation studies in two particular specifications of the multivariate model: the case in which all the involved processes are pure jump processes modelled according to Normal inverse Gaussian processes with drift ("all-NIG"); and the case in which all the involved processes are jump-diffusion processes of the Merton jump-diffusion kind ("all-MJD"). All required densities are generated via numerical inversion of the corresponding characteristic functions, using the Fast Fourier Transform (FFT) algorithm; alternatively the COS method suggested by Fang and Oosterlee (2008) can be adopted.

In this section we present the results of a simulation study aimed at assessing the estimation procedure. To this purpose we consider daily log-returns of the $SP500$ index and a selection of its constituents stocks; the observation period ranges from September 10, 2007 to May 20, 2013, for a total of 1434 observations per series.These data are extracted from Bloomberg database and adjusted for dividends.

We first estimate the chosen multivariate model using the index returns as proxy for Z. Then, we use the estimated parameters to generate series of the returns of the assets under consideration, to which the estimation procedure is re-applied. This allows us to recover the distribution of each parameter. We assess the estimation procedure in several cases, varying the length of the simulated series from one year up to four years of daily observations $T = [250, 500, 750, 1000]$ days and varying the number of components, considering up to 30 assets in the simulated portfolios ($n = [5, 10, 15, 30]$). For each of the

16 cases taken into account we repeat the simulation and estimation $S = 10000$ times, obtaining 10000 sets of parameters, denoted by Given the large number of parameters (if $n = 5$ the total number of parameters is 29 for the "all-NIG" model, 35 for the "all-MJD" model; if $n = 30$ there are 154 parameters for the "all-NIG" model, and 185 parameters for the "all-MJD" model) we cannot display detailed results for each parameter; hence, for illustrative purpose, we show only the assessment results for the estimation of the common factor Z, the first idiosyncratic factor $Y^{(1)}$ and average results relative to the loadings a_j,$j = 1, \dots n$. Complete results are available upon request. The assessment is made in terms of root mean square error, bias and inefficiency, defined as

$$RMSE(\hat{\theta}) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{\theta}^2 - \theta)^2} \tag{3.17}$$

$$\text{bias}(\hat{\theta}) = \sqrt{(E[\hat{\theta}] - \theta)^2} \tag{3.18}$$

$$\text{ineff}(\hat{\theta}) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\hat{\theta} - E[\theta]\right)^2} \tag{3.19}$$

where $\hat{\theta}$ indicates the estimates of the true parameter R used in the simulation step, and $E[\hat{\theta}] = \frac{1}{N}\sum_{i=1}^{N}\hat{\theta}_s$. We stress that the focus of our simulation studies is on investigating the effectiveness of splitting 10 the estimation procedure of the multivariate model in the two steps presented in Section 3.2. As a positive signal in this direction, we expect the errors obtained in the assessment of the final step, which depend on the loadings estimates and indirectly on the number of components, to be comparable with those obtained in the assessment of the first step, which consists in a plain univariate estimation.

Table 3.1 displays root mean square error, bias and inefficiency of the estimators for

the 'all-NIG' and 'all-MJD' models, as the length of the simulated series varies in $T = [250; 500; 750; 1000]$. The true value of the parameter, i.e. the parameter used in the simulation, and the mean of the estimator are very close to the true values as we can see the errors in different measures are all very small. This result suggests the good estimation of our procedure and also we observe that the estimators obtained almost unbiased.

Results are presented in Table 3.2 in which we report the average root mean square error, bias and inefficiency of the loadings a as the number of assets varies in $n = [5; 10; 15; 30]$ and the length of the simulated series for the estimation varies in $T = [250; 500; 750; 1000]$. We observe that the accuracy of the estimates improves as the sample size T increases, this feature being an indicator of consistency. Further, the accuracy is independent of the number of assets n as expected: the proposed 'observe, divide and conquer' strategy requires, in fact, $n + 1$ independent maximization procedures; therefore the accuracy of the resulting estimates is not affected by the number of assets considered. In order to analyze in more depth the behavior of the loadings estimators as the number of assets varies, we simulate datasets all made of series of fixed length T, and with number of assets, n, spanning the interval $[2; 60]$. For each n we simulate a dataset and we estimate the loadings, repeating the simulation-estimation procedure $10,000$ times. We then compute the average error, average bias, average standard error and average interquartile range of the loadings in correspondence of each n, meaning that, given n, we compute these measures for all a_j, $j = 1, \ldots n$, and then we take the average. The computations are repeated for simulated series of increasing length: $T = [250; 500; 750; 1000]$. The estimates

Table 3.1: *Common factors: Estimation errors expressed in absolute terms. (RMSE, Bias, Inefficiency)*

	$T = 250$	$T = 500$	$T = 750$	$T = 1000$
NIG Model				
$\mu = 0.0014$				
RMSE	9.85E-04	6.72E-04	5.42E-04	4.65E-04
Bias	4.33E-05	1.21E-05	1.84E-05	6.73E-06
Inefficiency	9.84E-04	6.71E-04	5.41E-04	4.65E-04
$\theta = -0.0014$				
RMSE	1.47E-03	1.02E-03	8.20E-04	7.12E-04
Bias	3.09E-05	2.42E-05	1.87E-05	4.90E-06
Inefficiency	1.47E-03	1.02E-03	8.20E-04	7.12E-04
$\rho = 0.0168$				
RMSE	1.76E-03	1.23E-03	1.01E-03	8.77E-04
Bias	1.77E-04	8.60E-05	6.41E-05	4.70E-05
Inefficiency	1.75E-03	1.22E-03	1.01E-03	8.75E-04
$k = 3.32$				
RMSE	1.30E+00	8.97E-01	7.26E-01	6.32E-01
Bias	1.91E-02	8.17E-03	5.85E-04	5.79E-03
Inefficiency	1.30E+00	8.97E-01	7.26E-01	6.32E-01
MJG Model				
$\mu = 0.0012$				
RMSE	8.24E-04	5.83E-04	4.66E-04	4.05E-04
Bias	2.80E-05	1.83E-05	2.49E-05	2.66E-05
Inefficiency	8.23E-04	5.83E-04	4.65E-04	4.04E-04
$\rho = 0.0075$				
RMSE	1.17E-03	8.90E-04	7.41E-04	7.41E-04
Bias	7.25E-05	1.31E-04	1.28E-04	1.47E-04
Inefficiency	1.17E-03	8.80E-04	7.30E-04	7.27E-04
$\nu = -0.0025$				
RMSE	3.14E-03	1.90E-03	1.51E-03	1.28E-03
Bias	1.35E-04	1.72E-04	6.60E-05	6.62E-05
Inefficiency	3.13E-03	1.89E-03	1.51E-03	1.28E-03
$\tau = 0.0210$				
RMSE	3.56E-03	2.82E-03	2.39E-03	2.36E-03
Bias	5.31E-04	5.89E-04	5.13E-04	5.59E-04
Inefficiency	3.52E-03	2.76E-03	2.33E-03	2.30E-03
$\lambda = 0.47$				
RMSE	1.50E-01	1.07E-01	8.72E-02	7.88E-02
Bias	1.48E-02	1.94E-02	2.33E-02	2.01E-02
Inefficiency	1.49E-01	1.05E-01	8.40E-02	7.62E-02

of the loadings appear to be consistent, as all the average error measures decrease when estimation is performed on longer time series.

Table 3.2: *Average results relative to the estimated parameters. Estimation errors expressed in absolute terms. RMSE, Bias, Inefficiency*

$T = 250$	RMSE	6.05E-02	5.23E-02	5.28E-02	6.05E-02
	Bias	Bias	1.83E-03	1.65E-03	1.64E-03
	Inefficiency	6.05E-02	5.23E-02	5.28E-02	6.05E-02
$T = 500$	RMSE	4.21E-02	3.65E-02	3.69E-02	4.23E-02
	Bias	9.18E-04	7.85E-04	8.14E-04	1.00E-03
	Inefficiency	4.21E-02	3.65E-02	3.68E-02	4.23E-02
$T = 750$	RMSE	3.42E-02	2.96E-02	2.99E-02	3.42E-02
	Bias	3.76E-04	5.80E-04	6.67E-04	7.28E-04
	Inefficiency	3.42E-02	2.96E-02	2.99E-02	3.42E-02
$T = 1000$	RMSE	2.96E-02	2.57E-02	2.60E-02	2.97E-02
	Bias	5.21E-04	3.85E-04	4.31E-04	3.77E-04
	Inefficiency	2.96E-02	2.57E-02	2.60E-02	2.97E-02

The results of the estimation of the idiosyncratic process are presented in Table 3, for the case of the first asset. In particular, the left-hand side of Table 3.2 displays root mean square error, bias and inefficiency of the estimators when the total number of assets is fixed ($n = 30$) and the length of the simulated series varies in $T = [250, 500, 750, 1000]$. On the right-hand side of the same table, we show the assessment results for a fixed T = 500, varying the number of assets. Although the estimation of each $Y^{(j)}$, $j = 1, \ldots n$, is

performed in a univariate way, the number of assets plays in principle a key role in the calibration of the loadings, which affects the estimation of the $Y^{(j)}$ parameters. Table 3 reveals almost uniform estimation errors for $n = [5, 10, 15, 30]$, showing that the number of assets has only a minimal impact on the estimation errors of the idiosyncratic terms for both the specifications we tested. As noted above, in this section we only discussed results relative to the first asset; similar conclusions hold for all assets considered.

In order to compare the efficiency of the two approaches, both in terms of estimation errors and computational time, we modify the efficiency gain index used in Monte Carlo simulation analysis and defined for example in Glasserman (2004). Given a specification of the model described, characterized by K parameters, we compute the efficiency gain, E_{21}, of the two-step procedure to the one-step maximum likelihood approach as

$$E_{21} = \frac{MSEt_1}{MSEt_2}$$

where MSE denotes the average mean square error

$$MSE = \frac{\sum_{i=1}^{K} MSE(\hat{\theta}_i)}{K} \tag{3.20}$$

of the parameters estimated by the MCMC method and the Maximum Likelihood approach respectively. MSE($\hat{\theta}_i$) is the mean square error (i.e. the square of the RMSE defined). In particular, we compute the efficiency gain index in correspondence of the 'all-NIG' and 'all-MJD' models with 5 and 15 components. In the $n = 5$ case, for each of the two approaches, we consider the mean square errors based on 1000 simulations; for $n = 15$ we rely on 100 simulations.

Results are reported in Table 3.3: we observe that the MCMC approach is significantly more efficient in terms of computational time. Moreover, for $n = 5$ the average mean square errors attained with the two-step approach are lower than those given by the one-step maximum likelihood almost the same for $n = 15$ (about 10% for the "all-NIG" model, 0.16% for "all-MJD"). According to efficiency index, in our experiment the MCMC procedure performed 3496 times more efficiently than the one step approach in the worst case ('all-MJD', $n = 15$) and 8139 times more efficiently in the best one ('all-NIG', $n = 5$). For a further comparison, we simulate $1,000$ samples, each made of 500 observations from the "all-NIG" and the "all-MJD" model with 5 components, and we estimate the parameters with both methods. For each simulated sample we then compare the maximum log-likelihood achieved using both approaches.

Table 3.3: *Average MSE, computation times (measured in seconds) and efficiency gains of the two-step approach to the maximum likelihood method.*

		MCMC		ML	
		MSE	Time	MSE	Time
n=5	NIG	0.0857	0.7	0.1407	3688.5
	MJD	0.0014	1.1	0.0028	3756.3
n=15	NIG	0.1403	1.9	0.0983	10273
	MJD	0.0016	3.4	0.0017	11180

3.3.2 Estimating multivariate stochastic dynamics

In this subsection, we examine daily data ranging from the most widely trade market indices. The indices used in the estimation are the MSCI indices of the United States, the United Kingdom, the Europe ex UK, and the Asia. In all, we have used four data sets including both domestic portfolios and international portfolios. A summary of statistical description of all data used in this experiment is provided in Table 3.4. For the international indices, the number of observations is 608. All returns are annualized log-returns in excess of the annualized 1-month T-bill rate. The international index data are sampled from May 2001 to December 2012.

Table 3.4: *A statistical description for international indices*

	Mean	Standard deviation	Correlation
US	0.0114	0.1918	0.7323
Europe	0.0308	0.2475	0.8275
UK	0.0291	0.2283	0.8129
Asia	0.0984	0.2368	0.7230

MCMC estimation of a joint distribution for the three indices is carried out with an increasing number d of states. For a fixed number of states, $30,000$ MCMC draws were generated after a burn-in of 5000 draws using MCMC procedure proposed. Estimation is based on the prior distributions that are invariant to relabeling the states. The prior for the state-specific variance-covariance matrix is $C^{(k)} \sim IW(\Theta, \nu)$ with $\nu = 2.5 + \frac{n-1}{2}$. Since Θ is likely to be influential, we consider a hierarchical prior where Θ is a random

parameter with a prior of its own $\Theta \sim W(G_0, g_0)$, where $g_0 = 0.5 + \frac{n-1}{2}$ and $G_0 = \frac{100 g_0}{\nu} Diag(1/s_1^2, ..., 1/s_n^2)$. Under this prior, an additional step has to be added to the MCMC scheme to sample $\Theta \sim W(G_0 + \sum_{k=1}^{d} C^(k-1), g_0) + d\nu$. Finally, the rows X_k of X are assumed to be independent and follow a Dirichlet distribution with $g_{kk} = 4$ and $g_{kj} = 1/(d-1)$ for $j \neq k$. This choice is based on Fruhwirth-Schnatter (2006) and leads to a prior that is invariant to relabeling the states.

We report the estimated parameters for 4-dimensional MJD model in Table 3.5. To show the accuracy of estimation results, we also report the comparison of moments between model and market data in Table 3.6.

Table 3.5: *The optimal estimates of the index data: MJD model*

μ	0.2531	0.2547	0.2437	0.2394	0.1948
Σ_b^0	0.0254	0.0664	0.0249	0.0669	0.1042
	0.0392	0.0749	0.0555	0.0899	0.0676
	0.0745	0.0786	0.0446	0.0632	0.0631
	0.0588	0.0740	0.0788	0.0476	0.0546
	0.0812	0.0154	0.0869	0.0396	0.0498
Σ_q^0	0.6249	0.6674	0.6411	0.6893	0.6244
	1.0460	1.0256	0.9686	0.9704	0.8350
η	0.0878	0.0210			
λ	0.0540	8.8976			

The empirical results clearly demonstrate the advantages the appropriateness of using

Table 3.6: *Comparison of moments for the index data: MJD model*

	data	model		data	model					
mean	0.0012	0.0012	kurtosis	9.1567	9.1853					
	0.0014	0.0014		8.4552	8.4310					
	0.0013	0.0014		7.9922	8.0219					
	0.0013	0.0013		9.3103	9.3036					
	0.0009	0.0009		7.8203	7.8258					
			data					model		
covariance (1e-3)	0.5891	0.5860	0.5391	0.5049	0.4017	0.5950	0.5802	0.5427	0.5070	0.4007
	0.5860	0.6314	0.5890	0.5583	0.4565	0.5802	0.6365	0.5940	0.5557	0.4611
	0.5391	0.5890	0.5766	0.5506	0.4615	0.5427	0.5940	0.5813	0.5487	0.4569
	0.5049	0.5583	0.5506	0.5532	0.4733	0.5070	0.5557	0.5487	0.5587	0.4723
	0.4017	0.4565	0.4615	0.4733	0.4743	0.4007	0.4611	0.4569	0.4723	0.4790

multivariate Lévy jump models modeling the joint dynamics of multi-assets. Lévy models can capture the many small movements in index returns that cannot be captured by pure-diffusion models. The Lévy jump models also have shown the goodness of fit in terms of high moments. If we allow jumps to follow different Lévy processes, Lévy jump models are likely to have even better performances in capturing the joint dynamics of index returns. Therefore, our analysis points out the great potentials of Lévy processes for continuous-time finance modeling and strongly suggests that we can enrich existing Łé models by incorporating infinite-activity Lévy jumps. Remember, the complexity of the MCMC method will not increase too much as the increasing of the number of assets. This is suggesting that the MCMC estimation method is capable of handling real portfolio problems since in reality the number of assets in a portfolio is usually bigger than 50. Such a large number makes existing classic estimation methods such as ML and GMM useless

for practice estimation problems.

3.4 Conclusion

We propose an estimation procedure for multivariate asset models based on linear transformation of Lévy processes as in Ballotta and Bonfiglioli (2014), allowing to extend the use of multivariate Lévy models for risk and portfolio management applications. The M-CMC estimation procedure proposed in this chapter basically can deal with multivariate estimation problem. It is fast to implement and its complexity does not increase with the number of components of the multivariate model. Our simulation study reveals that this approach is almost as effective as a more traditional direct maximum likelihood estimation of the whole set of parameters, as long as proper univariate estimation methods are used.The proposed approach is flexible with respect to the number of assets included in the portfolio and does not impose any convolution condition on the factors. Although in the numerical studies presented in this chapter we make the convenient assumption that all factors are modelled using the same type of process, this assumption can be relaxed as to allow any Lévy process for the idiosyncratic part across all the names included in the portfolio in order to accommodate different tail behaviours.

Applying the algorithms to financial time series, stable results are obtained for data with moderate volatility like stock indices. In general, achieving good fits for highly volatile data definitely requires a more refined model for the volatility. Our estimation result has

demonstrated the potential use of multivariate Lévy models. A possibility to extend this model to multiple dimensions might be to employ Markov switching in the correlation matrix as proposed by Pelletier (2006), and using Hobson and Rogers (1998) for the marginal volatilities. A good result obtained is that such a model increases the number of parameters only slightly increase the difficulty of estimation. Detailed investigations into that direction must however remain as future work. Also extensions to several common factors can also be considered by adopting the Independent Component Analysis (ICA) approach along the lines considered. This is left to future research.

Chapter 4

Sampling a special case of the SABR model

4.1 Introduction

In the financial industry, there are vanilla and exotic derivatives. The vanilla ones are also known as flow derivatives, as they are actively traded. Due to good liquidity of vanilla derivatives, they are used as the basis to price and hedge exotic derivatives, which are tailored to meet special client needs and inactively traded. The standard procedure to price an exotic derivative is that, first we should have a model; here we use the stochastic Alpha Beta Rho model, namely the SABR model, which is proposed by Hagan et al. (2002) and has dynamics as shown in equation (4.1). Secondly, we calibrate the model given prices of vanilla derivatives at a specified time; we numerically search for an optimal

set of model parameters in the sense of minimizing the aggregate distance between the given traded vanilla option prices and their corresponding model prices. At last, given the model with calibrated parameters, we sample the price of the target exotic derivative.

$$\begin{aligned} dF_t &= \sigma_t F_t^{\beta} dW_{1,t} \\ d\sigma_t &= \alpha \sigma_t dW_{2,t} \\ dW_{1,t} dW_{2,t} &= \rho dt \end{aligned} \tag{4.1}$$

The SABR model has gained its popularity in the financial industry due to the availability of a simple asymptotic connection (4.2) between model parameters and the implied Black-Scholes volatility. This makes fast calibration possible. However, the closed-form formula for the market implied Black-Scholes volatility is only an asymptotic one; it is more accurate only when the option expiry date T is small and the option strike price is not too far away from the at-the-money level.

$$\sigma_{implied} = f(\alpha, \beta, \rho, F_0, \sigma_0, K, T) \tag{4.2}$$

A good sampling scheme for the SABR model is not only important for pricing exotics, but can also helps us evaluate the goodness of the well-known asymptotic formula.

There are two natural numerical choices for sampling a model. One is the Monte Carlo method; thin time interval and large sample size have to be chosen to ensure the simulated distribution converges to the true distribution with good speed. The other choice is sampling the target distribution from its transition density; for the SABR model, we need to have the explicit joint density of (F_t, σ_t).

In this chapter, we try to sample the SABR model from its density. In the model (4.1), the forward asset price F_t can be seen as a constant elasticity of variance model, namely the CEV model, if its constant volatility component σ is replaced by a stochastic σ_t. The CEV dynamics, introduced by Cox (1975) has a special property, which is that it has a boundary at zero and its density at the boundary is not smooth. The boundary can be naturally killing for some parameter sets, which means that whenever F_t touches the boundary, it will stay there thereafter; or we have to set the boundary to be killing or reflecting or with some other kind of behavioural property to ensure a unique density.

Our endeavor starts from an alternative representation of the SABR model. According to the stochastic volatility framework in Hobson (2010)[1], (X_{A_t}, Y_{A_t}) is a weak solution to (F_t, σ_t), where the dynamics of (X_t, Y_t) is given by

$$\begin{aligned} dX_t &= X_t^{\beta} dZ_{1,t} \\ dY_t &= \alpha dZ_{2,t} \\ dZ_{1,t} dZ_{2,t} &= \rho dt. \end{aligned} \tag{4.3}$$

and $X_0 = F_0$, $Y_0 = \sigma_0$. A_t is the inverse of Γ_t, where $\Gamma_t = \int_0^t Y_s^{-2} ds$. Therefore, sampling the time-T distribution of F_T is equivalent to sampling X_{A_T}.

The representation of the SABR model in equation (4.3), disentangles the original complexity between F_t and σ_t in equation (4.1), yet adds it to the time-change A_t instead. In the new representation, X_t has a standard CEV dynamics, with constant volatility 1, while A_t follows some complex dynamics. The CEV dynamics can be sampled by

[1]The SABR model is a special case of the stochastic volatility framework proposed in Hobson (2010).

extending the sampling scheme proposed in Makarov and Glew (2010). For A_t, we find an alternative whose density is close to that of A_t in the numerical sense; the density of the alternative is available but hard to be sampled directly; instead, we use another distribution to approximate it by matching their first two central moments.

If β is zero, sampling the SABR X_{A_t} will be trivial as we have sampling schemes for both X_t and A_t respectively. In this chapter, we start from an alternative representation of the SABR model, provide a sampling scheme for the SABR model when β is zero and then analyze goodness of the asymptotic implied Black-Scholes formula under this setting.

This chapter will be organized as below. Section 1 introduces our target and a concrete procedure. In section 2, we will extend Makarov and Glew (2010)'s work and document how to sample the CEV dynamics X_t. Section 3 will focus on analyzing A_t; its numerical alternative is introduced. In section 4, we give dynamics of the new representation of the SABR model. Section 5 is on sampling X_{A_t} when β is zero. In section 6, we conclude this chapter.

4.2 Sampling the CEV dynamics

In this section, we first analyse the properties of the CEV dynamics, then we extend Makarov and Glew (2010)'s work and derive the procedure to sample the CEV dynamics from its true transition density. At last, we examine the goodness of this sampling.

4.2.1 Properties of the CEV dynamics

Here we consider $(F_t)_{t\geq 0}$ whose dynamics is of the CEV type,

$$dF_t = \sigma F_t^{\beta} dW_t. \tag{4.4}$$

By Ito's lemma, it smooth and monotonous transformation $X_t = \left(\frac{F_t^{1-\beta}}{1-\beta}\right)^2$ follows the dynamics

$$dX_t = \delta\sigma^2 dt + 2\sigma\sqrt{X_t} dW_t, \tag{4.5}$$

where $\delta = \frac{1-2\beta}{1-\beta}$. When $\sigma = 1$, $(X_t)_{t\geq 0}$ is called a $\delta-$dimensional squared Bessel process. We care about the behavior of $(F_t)_{t\geq 0}$ when β lies in $(0, 1)$, which is the parameter set that the SABR model is calibrated in practice.

Table 4.1: *Relation between parameters*

β	$\delta = \frac{1-2\beta}{1-\beta}$	$v = \frac{2(\delta\sigma^2)}{(2\sigma)^2} = \frac{\delta}{2} - 1$
0	1	-0.5
0.1	0.8889	-0.5556
0.3	0.5714	-0.7143
0.5	0	-1
0.7	-1.3333	-1.6667
0.9	-8	-5
1	$-\infty$	$-\infty$

When $0 < \beta < 1$, a crucial property of the CEV dynamics $(F_t)_{t\geq 0}$ is that its probability of touching zero is positive. According to the relationship between F_t and X_t, F_t is

a monotonous function of X_t and $F_t = 0$ is equivalent to $X_t = 0$, thus we study the properties of $(F_t)_{t\geq 0}$ from the properties of $(X_t)_{t\geq 0}$.

The density of $(X_t)_{t\geq 0}$ is available when $\beta \in (0,1)$. According to Borodin and Salminen (2002), when $\frac{1}{2} \leq \beta < 1$ ($\delta \leq 0$, $v \leq -1$), X_t has a unique solution and the left-hand boundary $l = 0$ is a natural killing boundary[2]. The density of X_t is

$$p(t;x,y) = \frac{P\left(X_t \in dy | X_0 = x\right)}{dy} = \frac{1}{2\sigma^2 t}\left(\frac{y}{x}\right)^{\frac{v}{2}} e^{-\frac{x+y}{2\sigma^2 t}} I_{|v|}\left(\frac{\sqrt{xy}}{\sigma^2 t}\right). \tag{4.6}$$

When $0 < \beta < \frac{1}{2}$ ($0 < \delta < 1$, $-1 < v < -\frac{1}{2}$), there is no unique solution for X_t. If the left-hand boundary $l = 0$ is set to be killing, the density of X_t is

$$p(t;x,y) = \frac{1}{2\sigma^2 t}\left(\frac{y}{x}\right)^{\frac{v}{2}} e^{-\frac{x+y}{2\sigma^2 t}} I_{|v|}\left(\frac{\sqrt{xy}}{\sigma^2 t}\right), \tag{4.7}$$

which coincides with the case when $\frac{1}{2} \leq \beta < 1$. If a reflecting left-hand boundary is set, the density of X_t is

$$p(t;x,y) = \frac{1}{2\sigma^2 t}\left(\frac{y}{x}\right)^{\frac{v}{2}} e^{-\frac{x+y}{2\sigma^2 t}} I_{v}\left(\frac{\sqrt{xy}}{\sigma^2 t}\right). \tag{4.8}$$

As is shown in Table 4.1, when $0 < \beta < 1$, v is always negative. The only difference between densities (4.6), (4.7) and (4.8) is the parameter for the modified Bessel function of the first kind $I_{\cdot}\left(\frac{\sqrt{xy}}{\sigma^2 t}\right)$. To make X_t has one expression of its density for $\beta \in (0,1)$, a killing boundary is set for X_t when $\beta \in (0, \frac{1}{2})$. Figure 4.1 gives plots of the densities of X_t for different βs, with parameters $t = 1$, $X_0 = 0.5$, $\sigma = 0.2$.

[2]A killing boundary is also called an absorbing boundary, which means that when X_t ever touches 0, it will stay at 0 thereafter.

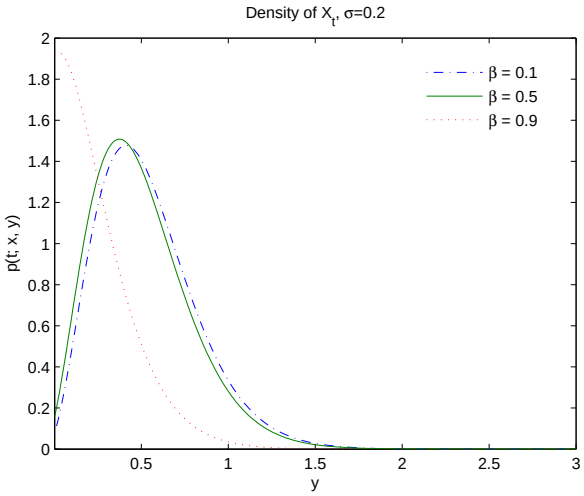

Figure 4.1: *Density of X_t, $p(t;x,y)$ $(y>0)$*

As β increases from 0 to 1, the drift coefficient of X_t, δ, decreases from positive to negative infinity, thus an increasing probability of killing at the left-hand boundary $l=0$ is expected intuitively. The probability of killing at $l=0$ up to time t is

$$P_{killing}(x;t) = 1 - \int_0^\infty p(t;x,y)dy = 1 - \frac{\gamma(|v|, \frac{x}{2\sigma^2 t})}{\Gamma(|v|)}. \tag{4.9}$$

Figure 4.2 gives a plot of the killing probability which the same set of parameters is used as in Figure 4.1. $P_{killing}$ increases with increasing β; it is consistent with our intuition.

4.2.2 Sampling the CEV dynamics from its density

In this chapter, instead of Monte Carlo simulation, we try to sample the CEV dynamics then the SABR model from their densities respectively. To sample F_t, we first extend the scheme proposed in Makarov and Glew (2010), sample X_t whose density is available and unified when a killing boundary is set for $\beta \in (0,1)$, and then obtain F_t by applying the

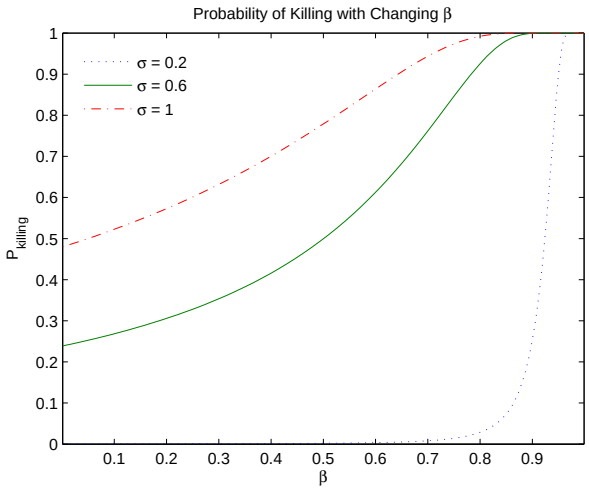

Figure 4.2: *Probability of Killing with Changing β*

transformation $F_t = \left((1-\beta)X_t^{\frac{1}{2}}\right)^{\frac{1}{1-\beta}}$.

As the probability of X_t killing at boundary $l = 0$ up to time t is positive, the transition density as shown in equation (4.6) does not integrate to one. The surviving probability at time t is $P_{surviving}(x;t) = \int_0^\infty p(t;x,y)dy = \frac{\gamma(|\mu|,\frac{x}{2\sigma^2 t})}{\Gamma|v|}$, and the killing probability is $P_{killing}(x;t) = 1 - P_{surviving}(x;t) = \frac{\Gamma(|v|,\frac{x}{2\sigma^2 t})}{\Gamma(|v|)}$. The actual transition probability of X_t is

$$p^*(t;x,y) = P_{surviving}(x;t) \cdot \left(\frac{p(t;x,y)}{P_{surviving}(x;t)}\right) + P_{killing}(x;t) \cdot \mathbf{1}_{y=0}. \tag{4.10}$$

Density of the first hitting time of $(X_t)_{t\geq 0}$

The first hitting time, τ, at zero for the squared Bessel process $(X_t)_{t\geq 0}$ is defined by $\tau = \inf\{t : X_t = 0 | X_0 = x\}$. Its density is

$$q(x;\tau) = \frac{1}{\sigma^2\tau\Gamma(|v|)}\left(\frac{x}{2\sigma^2\tau}\right)^{|v|} e^{-\frac{x}{2\sigma^2\tau}}. \tag{4.11}$$

As $d\left(\frac{x}{2\sigma^2\tau}\right) = -\frac{x}{2\sigma^2}\frac{1}{\tau^2}d\tau$,

$$\begin{aligned} P_{killing}(x;t) &= \int_0^t q(x;\tau)d\tau = \int_0^t \frac{1}{\tau\Gamma(|v|)}\left(\frac{x}{2\sigma^2\tau}\right)^{|v|} e^{-\frac{x}{2\sigma^2\tau}}d\tau \\ &= \int_{\frac{x}{2\sigma^2 t}}^{\infty} \frac{1}{\Gamma(|v|)}\left(\frac{x}{2\sigma^2\tau}\right)^{|v|-1} e^{-\frac{x}{2\sigma^2\tau}} d\left(\frac{x}{2\sigma^2\tau}\right) \\ &= \int_{\frac{x}{2\sigma^2 t}}^{\infty} g(\frac{x}{2\sigma^2\tau};|v|,1)d\left(\frac{x}{2\sigma^2\tau}\right) \end{aligned} \tag{4.12}$$

where $g(y;\alpha,\beta)$ is the density of Gamma distribution $G(\alpha,\beta)$. Therefore, the first hitting time, τ, can be sampled with $\tau = \frac{x}{2\sigma^2 H}$, where $H \sim G(|v|,1)$.

Density of X_t conditional on surviving

Randomized Gamma distribution $G(\alpha+Z,\beta)$

Suppose Z is a discrete random variable with density $p_n = P(Z=n)$, $G(\alpha,\beta)$ is Gamma distribution with shape parameter α and rate parameter β, then $G(\alpha+Z,\beta)$ is a randomized Gamma distribution. Its density has the form $g(y) = \sum_{n=0}^{\infty} p_n \frac{\beta^{\alpha+n}}{\Gamma(\alpha+n)} y^{\alpha+n-1}e^{-\beta y}$. If Z has an incomplete Gamma distribution with positive parameter θ and λ, $Z \sim I\Gamma(\theta,\lambda)$, whose density is

$$p_n = P(Z=n) = e^{-\lambda}\frac{\lambda^{n+\theta}}{\Gamma(n+\theta+1)}\frac{\Gamma(\theta)}{\gamma(\theta,\lambda)}, \tag{4.13}$$

then the density for the randomized Gamma $G(1+Z,\beta)$ (when $\alpha=1$) is

$$g(y) = \beta\frac{\Gamma(\theta)}{\gamma(\theta,\lambda)}\left(\frac{\beta}{\lambda}\right)^{-\frac{\theta}{2}} y^{-\frac{\theta}{2}}e^{-\lambda-\beta y}I_\theta(\sqrt{4\beta\lambda y}). \tag{4.14}$$

Relationship between $p(t;x,y)$ and $G(1+Z,\beta)$

The time-t transition density of a squared Bessel process conditional on surviving is

$$\frac{p(t;x,y)}{P_{surviving}(x;t)} = \frac{1}{2\sigma^2 t}\frac{\Gamma(|v|)}{\gamma(|v|,\frac{x}{2\sigma^2 t})}\left(\frac{y}{x}\right)^{\frac{v}{2}} e^{-\frac{x+y}{2\sigma^2 t}} I_{|v|}(\frac{\sqrt{xy}}{\sigma^2 t}). \tag{4.15}$$

We can find that $\frac{p(t;x,y)}{P_{surviving}(x;t)} = g(y)$ if we set $\theta = |v|$, $\lambda = \frac{x}{2\sigma^2 t}$ and $\beta = \frac{1}{2\sigma^2 t}$. Thus the conditional density $\frac{p(t;x,y)}{P_{surviving}(x;t)}$ has a randomized Gamma distribution $G(1+Z,\frac{1}{2\sigma^2 t})$, where $Z \sim I\Gamma(|v|,\frac{x}{2\sigma^2 t})$.

Exact sampling of F_T

In this section, we first give the procedure on how to sample X_T from its density as show in equation (4.10), given the sampling scheme of the first hitting time of $(X_t)_{t\geq 0}$ at 0 and the sampling scheme of the density of X_T conditional on its not touching the absorbing boundary.

$H \sim G(|v|,1)$, $\tau \leftarrow \frac{X_0}{2\sigma^2 H}$

if $\tau > T$ **then**

$Z \sim I\Gamma(|v|,\frac{x}{2\sigma^2 T})$

$X_T \sim G(1+Z,\frac{1}{2\sigma^2 T})$

else

$X_T = 0$

end if

To get an X_T for a specific sample path of X_t up to time T, we first sample one first hitting time τ of $(X_t)_{t\geq 0}$ at the zero absorbing boundary by sampling an $H \sim G(|v|,1)$

and getting the first hitting time from the transformation $\tau \leftarrow \frac{X_0}{2\sigma^2 H}$. Then if $\tau > T$, which means that for this specific path of X_t, it doesn't touch boundary 0 before T, then we sample X_T based on the scheme shown above; otherwise, $\tau \leq T$, X_t touches 0 before T and stays thereafter, thus $X_T = 0$.

We can sample enough sample paths of X_T, and get samples of F_T by the transformation $F_T = \left((1-\beta)X_T^{\frac{1}{2}}\right)^{\frac{1}{1-\beta}}$.

4.2.3 Numerical experiments

In this section, we want to check whether the sampling scheme for the CEV dynamics $(F_t)_{t\geq 0}$ works and can achieve good precision. As X_t is a smooth and monotonous transformation of F_t and the true density for X_T is available for comparison, here we test the sampling scheme for X_T directly instead of that of F_T.

We will check whether the sampled distribution of X_T at time $T = 1$ is close to its true density, the starting point $X_0 = 0.5$. Number of samples paths $N = 10^5$. Besides, there are two parameters β and σ for the CEV dynamics $(F_t)_{t\geq 0}$ in equation (4.4) which influence model behaviours, while in the dynamics of X_t (equation (4.5)), parameter $\delta = \frac{1-2\beta}{1-\beta}$. The distribution of X_T is not smooth at the zero boundary; for a sample path X_t $(0 \leq t \leq T)$, X_t might ever touch zero and stays at zero thereafter up until time T. The probability of killing at the zero boundary is increasing with increasing β (equivalent to increasing δ when $\beta \in (0, 1)$) and σ. In the experiments, to check whether the sampling scheme works

at the boundary, we choose different levels of $\beta \in [0.1, 0.5, 0.9]$ and $\sigma \in [0.2, 0.6, 1]$.

Table 4.2: *Probability of X_T killing at 0, $X_0 = 0.5$, $T = 1$*

	$\sigma = 0.2$		$\sigma = 0.6$		$\sigma = 1$	
	True	Sampled	True	Sampled	True	Sampled
$\beta = 0.1$	5.65E-04	5.80E-04	0.2457	0.2769	0.5052	0.5251
$\beta = 0.5$	0.0017	0.0022	0.4834	0.4949	0.7707	0.7698
$\beta = 0.9$	0.2570	0.2536	0.9993	0.9991	1.0000	1.0000

Table 4.2 gives both the true and the sampled probabilities of X_T killing at zero for nine designed parameter sets, from which, we can see that increasing β or σ causes increasing killing probability at the zero boundary, just the same as our intuition. Besides, sampled probabilities of killing at 0 are close to the true killing probabilities.

In Figure 4.3, sampled densities (solid lines) of X_T are plotted against true densities (dashed lines) using the same parameter sets. From these plots, we can find that the sampled densities of X_T are unbiased estimates of the true densities. When the probability of killing at the zero boundary is small, which is that both β and σ is small ($\beta = 0.1, 0.5$ and $\sigma = 0.2$), the density of X_T is hump-shaped and the absolute difference between the sampled density and the true density is large around the hump, small at the zero boundary and positive infinity. When the probability of killing is increasing, the shape of the density of X_T changes from hump-shaped to strictly decreasing. When the probability of killing at the zero boundary is approaching 1, the probability of not killing is approximately 0,

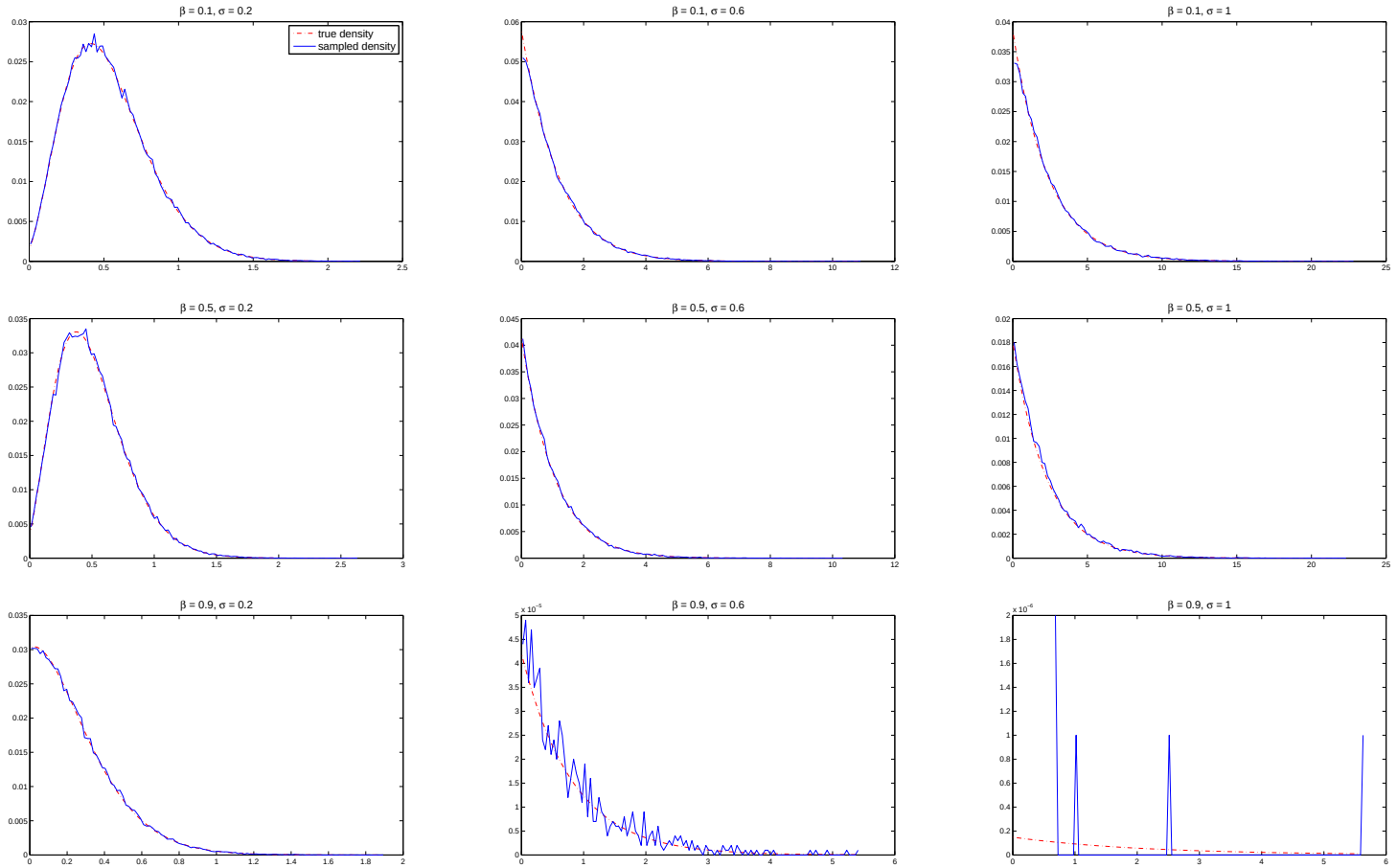

Figure 4.3: *Plots of sampled densities of* X_T *against true densities, while* $\beta = 0.1, 0.5, 0.9$, $\sigma = 0.2, 0.6, 1$, $X_0 = 0.5$, $T = 1$

the sampled density is not even smooth, e.g. when $\beta = 0.9$ and $\sigma = 0.6, 1$, $P_{killing} > 99.9\%$. Besides, we find that when β is small yet σ is large, e.g. $\beta = .1$, $\sigma = 0.6, 1$, the absolute difference between the sampled density and the true density is large around the zero boundary. The convergence rate of the sampled density to the true density can be improved if we increase the number of sample paths N.

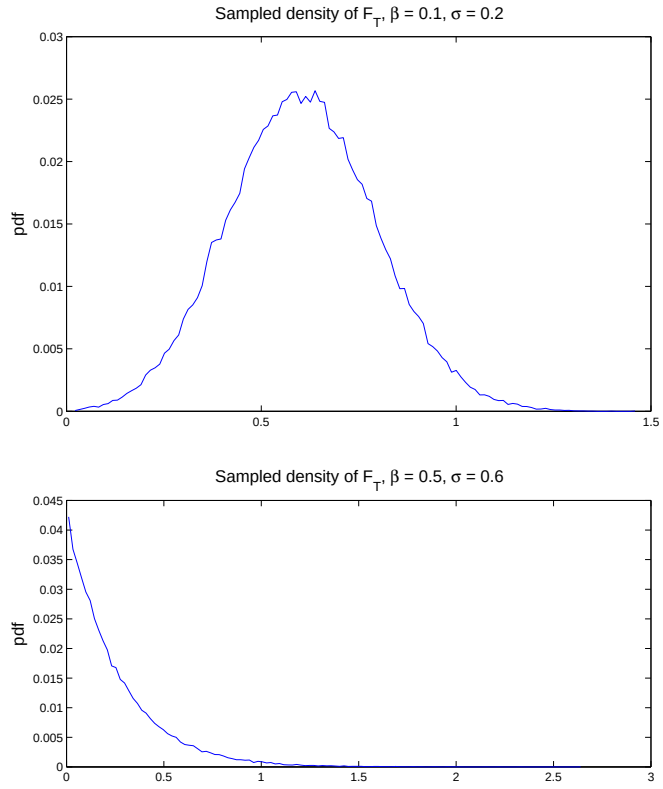

Figure 4.4: *Plots of sampled densities of F_T, while $\beta = 0.1$, $\sigma = 0.2$, or $\beta = 0.5$, $\sigma = 0.6$, $X_0 = 0.5$, $T = 1$*

4.3 Sampling the time-change A_t

As (X_{A_t}, Y_{A_t}) is a weak solution to the SABR pair (F_t, σ_t), sampling the forward asset price F_T in the SABR model is equivalent to sampling X_{A_T}. The SABR model (F_t, σ_t) has dynamics as shown in equation (4.1), while the dynamics of (X_t, Y_t) is given by

$$\begin{aligned} dX_t &= X_t^{\beta} dZ_{1,t} \\ dY_t &= \alpha dZ_{2,t} \\ dZ_{1,t} dZ_{2,t} &= \rho dt, \end{aligned} \tag{4.16}$$

$X_0 = F_0$, $Y_0 = \sigma_0$, and $\Gamma_t = \int_0^t Y_s^{-2} ds$, where A_t is the inverse of Γ_t.

In equation (4.16), $(X_t)_{t \geq 0}$ is a standard CEV dynamics whose exact sampling scheme is discussed in the last section. Before considering sampling X_{A_T}, we focus on how to sample A_T in this section. We first start from the dynamics of A_t

$$\begin{aligned} dY_t &= \alpha dZ_{2,t} \\ \Gamma_t &= \int_0^t Y_s^{-2} ds \\ A_t &= \Gamma_t^{-1}, \end{aligned} \tag{4.17}$$

and derive the implicit solution of A_t. Secondly, we propose an alternative of A_t, A_t^*, whose density is indistinguishable to that of A_t in the numerical sense. At last, a quick sampling scheme of A_t^* is given.

4.3.1 The implicit solution of A_t

In this part, we want to derive a simpler expression of A_t. Here we use $\Gamma(t)$ and $A(t)$ instead of Γ_t and A_t respectively, and will change them back at the end of this derivation. As by definition

$$\Gamma(t) = \int_0^t Y(s)^{-2} ds,$$

integrate $\Gamma(t)$ from 0 to 0, we get $\Gamma(0)$=0. Take derivatives w.r.t. t, it gives

$$\Gamma'(t) = Y(t)^{-2}.$$

and $\Gamma'(\Gamma^{-1}(t)) = Y(\Gamma^{-1}(t))^{-2}$. As $\Gamma(t)$ is a strictly increasing function, its inverse exists and is unique, and $\Gamma(\Gamma^{-1}(t)) = t$. Taking derivatives w.r.t. t on both sides, we get

$$\Gamma'(\Gamma^{-1}(t)) \left(\Gamma^{-1}(t)\right)' = 1,$$

then

$$\begin{aligned}
(\Gamma^{-1}(t))' &= \frac{1}{\Gamma'(\Gamma^{-1}(t))} \\
&= \frac{1}{Y(\Gamma^{-1}(t))^{-2}} \\
&= Y(\Gamma^{-1}(t))^2. \qquad (4.18)
\end{aligned}$$

By definition, $A(t)$ is the inverse of $\Gamma(t)$,

$$\begin{aligned} A(t) &= \Gamma^{-1}(t) \\ &= \int_0^t (\Gamma^{-1}(s))' ds \\ &= \int_0^t Y(\Gamma^{-1}(s))^2 ds \qquad \text{By equation (4.18)} \\ &= \int_0^t Y(A(s))^2 ds, \end{aligned}$$

Thus we get the implicit solution of A_t

$$A_t = \int_0^t Y_{A_s}^2 ds \tag{4.19}$$

$$dY_t = \alpha dZ_{2,t},$$

where Y_t is a Brownian motion with $Y_0 = \sigma_0$.

4.3.2 A numerically indistinguishable alternative, A_t^*

Although we already have the solution of A_t as in equation (4.19), it is implicit and complicated. We can get an idea of how it behaves by simulating it with Monte Carlo schemes. However, for analytical properties of A_t, we need to have an explicit solution or an explicit asymptotics of A_t or its density. As Y_{A_t} is a weak solution to σ_t in equation (4.1), we will make an attempt to find the explicit solution of A_t from this point.

σ_t is lognormal, and its explicit solution is $\sigma_t = \sigma_0 e^{-\frac{1}{2}\alpha^2 t + \alpha W_{2,t}}$. Here we make a guess of the explicit solution of A_t by substituting Y_{A_s} in equation (4.19) with $\sigma_s = \sigma_0 e^{-\frac{1}{2}\alpha^2 s + \alpha Z_{2,s}}$.

We name the guess as A_t^*.

$$A_t^* = \sigma_0^2 \int_0^t e^{-\alpha^2 s + 2\alpha Z_{2,s}} ds. \tag{4.20}$$

Can we use A_t^* in equation (4.20) instead of A_t in equation (4.19)? The Euler Monte Carlo simulation schemes for both A_t and A_t^* are implemented to test this. The details of the numerical implementation are written in Appendix B.1, from which we find that, A_t^* is pathwise identical to A_t; they are pathwise indistinguishable from the perspective of their Euler Monte Carlo schemes. Thus A_t is explicitly solved at least in the numerical sense, and we will use A_t^* instead of A_t in the remaining part of this chapter.

4.3.3 Sampling A_T^*

In the literature, $N_t = \int_0^t e^{2(\mu s + W_s)} ds$ is known as the integral of geometric Brownian motion (IGBM). Its closed-form density is available in the survey Matsumoto and Yor (2005), which is

$$P(N_t \in du | \mu t + W_t = x) = \frac{\sqrt{2\pi t}}{u} e^{\frac{x^2}{2t} - \frac{1+e^{2x}}{2u}} \theta_{\frac{e^x}{u}}(t) du,$$

where

$$\theta_r(t) = \frac{r e^{\frac{\pi^2}{2t}}}{\sqrt{2\pi^3 t}} \int_0^\infty e^{-\frac{y^2}{2t}} e^{-r\cosh y} (\sinh y) \sin\left(\frac{\pi y}{t}\right) dy.$$

If we make a constant time and space change, A_t^* could be re-parameterized as $A_t^* = \frac{\sigma_0^2}{\alpha^{\frac{1}{2}}} \int_0^{\alpha^{\frac{1}{2}} t} e^{2(-\frac{\alpha^{\frac{3}{2}}}{2} s + Z_{2,s})} ds$, which is of the form of a standard integral of geometric Brownian motion. However, its density is complicated which is hard to be sampled directly from. As

we already have the explicit solution, here instead of simulating it with 1-order Euler, 2-order Milstein Monte Carlo schemes, we choose a quick alternative. We will use a tractable distribution to approximate the distribution of A_t^* by matching their lower moments.

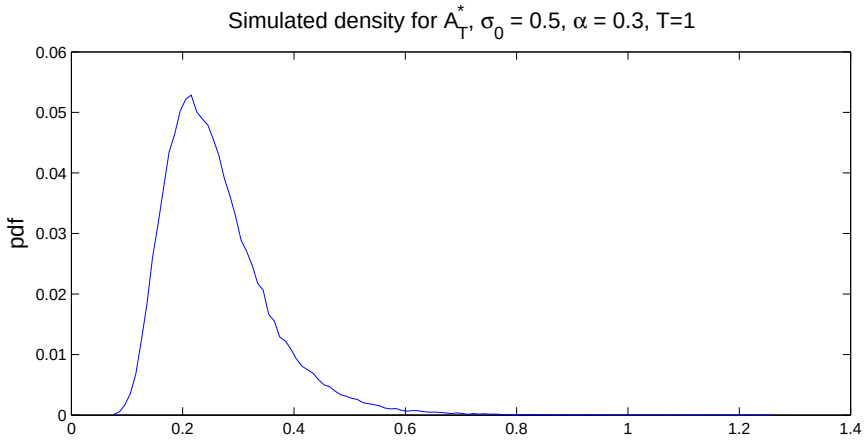

Figure 4.5: *Simulated density for A_T^*, while $\sigma_0 = 0.5$, $\alpha = 0.3$, $T = 1$, number of sample paths $N = 10^5$, number of discretization steps $M = 200$*

In Figure 4.5, we plot the simulated density for A_T^*, while the Euler scheme in Appendix B.1 is used and model parameters are $\sigma_0 = 0.5$, $\alpha = 0.3$, $T = 1$, number of sample paths $N = 10^5$, and number of discretization steps $M = 200$. We find from Figure 4.5 that, the density for A_T^* is bell-shaped yet skewed to the right. We can use a distribution with the same properties to approximate it, e.g. a lognormal distribution. The lognormal distribution is a simple and tractable candidate, with only two parameters controlling the shape of its distribution. Suppose we have a random variable $X \sim \text{LN}(\mu, \sigma^2)$, where

$\log(X) \sim \mathrm{N}(\mu, \sigma^2)$. Its unconditional mean $\mathrm{E}[X]$ and variance $\mathrm{Var}[X]$ are explicit,

$$\mathrm{E}[X] = \mathrm{e}^{\mu+\frac{\sigma^2}{2}}$$
$$\mathrm{Var}[X] = \mathrm{e}^{2\mu+\sigma^2}\left(\mathrm{e}^{\sigma^2-1}\right).$$

The conditional mean and variance of A_T^* can be calculated if perturbation is applied to parameter α,

$$\mathrm{E}[A_T^*|Z_{2,T}] = \sigma_0^2 T\left(1+\alpha Z_{2,T}+\frac{1}{3}\alpha^2(2Z_{2,T}^2-\frac{T}{2})+\frac{1}{3}\alpha^3(Z_{2,T}^3-Z_{2,T}T)\right)+\mathrm{O}(\alpha^4) \tag{4.21}$$

$$\mathrm{Var}[A_T^*|Z_{2,T}] = \frac{1}{3}\sigma_0^4\alpha^2T^3+\mathrm{O}(\alpha^4). \tag{4.22}$$

Derivation details can be found in Kahl (2007). These results are only accurate when both α and T are small, with which, we have the sampling scheme for A_T^*:

1. Simulate a vector of size N of $Z_{2,T}$ from the normal distribution $\mathrm{N}(0,T)$;

2. Given parameters α, σ_0, and sampled $Z_{2,T}$, calculate the conditional mean of A_T^*, $m = \mathrm{E}[A_T^*|Z_{2,T}]$, by equation (4.21) and the conditional variance of A_T^*, $s^2 = \mathrm{Var}[A_T^*|Z_{2,T}]$, by equation (4.22);

3. Equate the conditional mean and variance of A_T^* with unconditional mean and variance of X respectively, and calculate parameters μ and σ^2 for r.v. X;

$$\mu = \ln(m)-\frac{1}{2}\ln(1+\frac{s^2}{m^2})$$
$$\sigma^2 = \ln(1+\frac{s^2}{m^2})$$

4. Given vectors of μ and σ^2 obtained in step 3, simulate a vector of size N of $\epsilon \sim N(0,1)$ and get the samples of X by $e^{\mu+\sigma\epsilon}$.

The sampled distribution of X has the same first two central moments as that of A_T^*, and we use this distribution to represent the distribution of A_T^*. However, as we truncate residuals while calculating explicit moments of A_T^*, we need to choose small T and α to ensure that A_T^* and X have the same first two central moments.

Table 4.3: *First two central moments of A_T^* and X, while $\sigma_0 = 0.5$, $T = 1$, $\alpha = 0.1$ or $\alpha = 0.3$*

	$\alpha = 0.1$		$\alpha = 0.3$		$\alpha = 0.5$	
	Mean	Variance	Mean	Variance	Mean	Variance
A_T^*	0.2513	8.4578E-04	0.2618	0.0091	0.2831	0.0360
X	0.2513	8.4725E-04	0.2622	0.0088	0.2891	0.0322

We test three cases to check the performance of this moment-matching method to approximate a distribution with another tractable distribution. Common parameters used in the experiment are $\sigma_0 = 0.5$, $T = 1$, number of sample paths $N = 10^5$. As the discrepancies between the first two central moments of A_T^* and X increase with *alpha*, we check three choices of α, which are 0.1, 0.3 and 0.5. A_T^* is simulated by the Euler scheme with 200 discretization steps. The means and variances for both A_T^* and X in all three cases are provided in Table 4.3, and for each case, the simulated A_T^* is plotted against X, as shown in Figure 4.6. We can see that increasing α implies increasing differences between the first

two central moments between A_T^* and X and increasing differences between the simulated distribution of A_T^* and the approximated lognormal distribution of X.

4.4 A new representation of the SABR model

As A_t^* is a pathwise indistinguishable alternative to A_t from the perspective of Euler Monte Carlo scheme, we substitute A_t with A_t^* and get a new representation of the SABR model

$$\begin{aligned} dX_{A_t} &= X_{A_t}^{\beta} dZ_{1,A_t} \\ A_t &= \sigma_0^2 \int_0^t e^{-\alpha^2 s + 2\alpha Z_{2,s}} ds \\ dZ_{1,t} dZ_{2,t} &= \rho dt, \end{aligned} \tag{4.23}$$

where $X_0 = F_0$.

Sampling the forward asset price F_T in the SABR model is then equivalent to sampling X_{A_T} in equation (4.23), where $(X_t)_{t\geq 0}$ has a CEV-type dynamics whose exact sampling scheme is available in section 2, and A_t is the integral of a lognormal process whose time-t distribution can be approximated by a moment-matched lognormal distribution.

If the correlation coefficient ρ between Brownian increments $Z_{1,t}$ and $Z_{2,t}$ is zero, explicit density for X_{A_t} can be derived by the definition of conditional density, as explicit densities for both X_t and A_t is available. However, if $\rho \neq 0$, we don't know the density for X_{A_t}.

When $\rho \neq 0$, we can only incorporate this non-zero correlation into sampling through

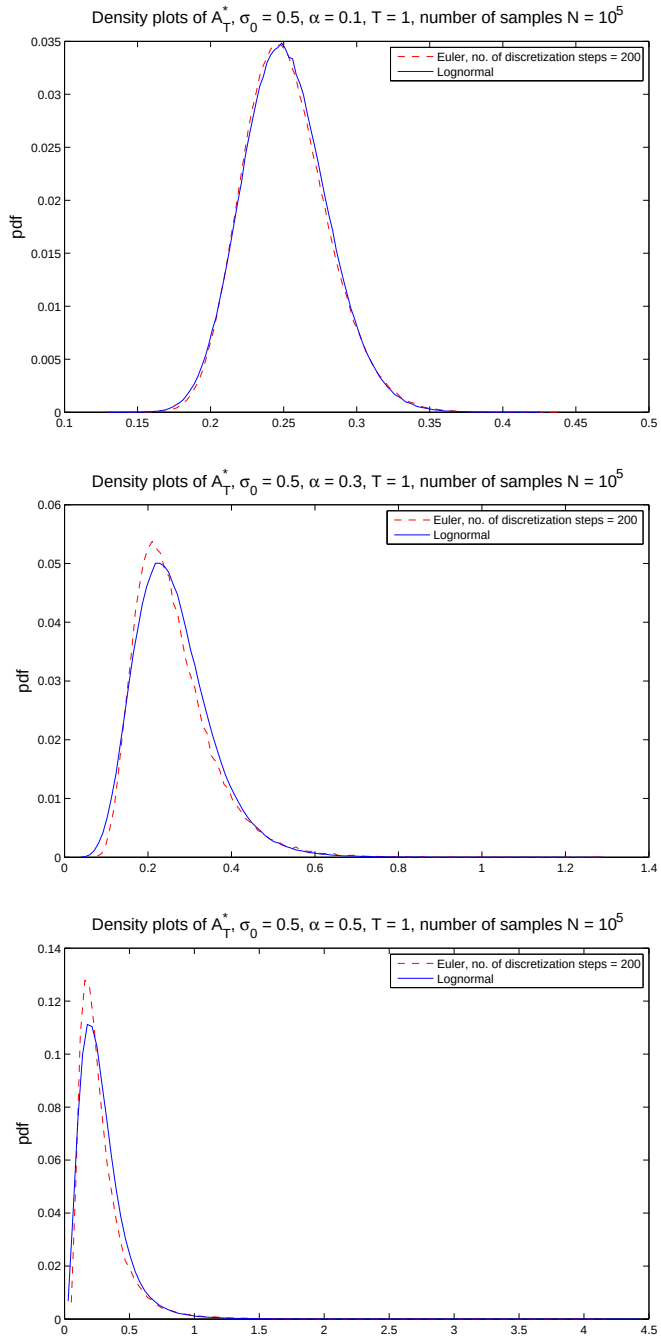

Figure 4.6: *Density plots of* A_T^**, while* $\sigma_0 = 0.5$*,* $\alpha = [0.1, 0.3, 0.5]$*,* $T = 1$*, number of sample paths* $N = 10^5$

pathwise numerical schemes. It seems that we cannot avoid discretization schemes like Monte Carlo to get a sample of X_{A_t}. Thus for the remaining part of this chapter, we will focus on sampling X_{A_t} when $\rho = 0$.

When $\rho = 0$, X_{A_t} can be seen as a CEV-type dynamics time-changed by an independent stochastic clock, which is the integral of some geometric Brownian motion. The sampling scheme is trivial. For a specific sample path of X_{A_T}, we first simulate the stochastic clock A_T and then given the simulated A_T, we sample from the density of X_{A_T}.

4.5 Numerical experiments

The SABR model is widely used by practitioners in the financial industry, especially in the interest rate derivative markets. One crucial reason for its popularity is that, for a European call, an explicit formula of the implied Black-Scholes volatility is available. When the strike K is not too far from the current forward asset price F_0, the implied Black-Scholes volatility has the form, as in Hagan et al. (2002),

$$\sigma_B(K, F0) \approx \frac{\sigma_0}{F_0^{1-\beta}} \left\{ 1 - \frac{1}{2}(1 - \beta - \rho\lambda) \ln(\frac{K}{F_0}) + \frac{1}{12} \left[(1-\beta)^2 + (2 - 3\rho^2)\lambda^2 \right] \ln^2(\frac{K}{F_0}) \right\}, \tag{4.24}$$

where $\lambda = \frac{\alpha}{\sigma_0} F_0^{1-\beta}$.

In this section, we want to check the performance of our sampling scheme for the SABR model when $\rho = 0$ against its Euler simulation scheme and the goodness of fit of the

asymptotic implied Black-Scholes volatility (4.24). To achieve these two targets simultaneously, we compare their implied Black-Scholes volatilities.

Suppose we want to price a series of European call options contingent on an asset, whose forward price at time-0 is $F_0 = X_0 = 0.5$. These options are with strikes $[0.1, 0.2, \cdots, 0.9]$ and 1-year maturity. For simplicity, the short rate is set to 0. As the SABR model has a natural absorbing boundary at 0 or we set the boundary to be absorbing, $(F_t)_{t\geq 0}$ can have quite different time-T distributions with different parameter sets. Besides $\alpha = 0.3$, we choose six sets of model parameters, as shown in Table 4.4, in experiments.

Table 4.4: *Parameter sets used in experiments*

	$\sigma_0 = 0.2$	$\sigma_0 = 0.6$	$\sigma_0 = 1$
$\beta = 0.1$	case 1	case 2	case 3
$\beta = 0.5$	case 4	case 5	–
$\beta = 0.9$	case 6	–	–

To obtain the implied Black-Scholes volatility of a European call option by simulation or sampling, we first simulated F_T with the Euler Monte Carlo scheme given the dynamics of the SABR model in equation (4.1) or sample X_{A_T} (equation (4.23)) with the sampling scheme proposed in this chapter; secondly, given sampled F_T or X_{A_T}, we price the European option; then the corresponding Black-Scholes volatility can be numerically implied based on option price obtained in the last step and parameters of the option[3]. In our ex-

[3]In Matlab, the command for calculating the implied volatility is blsimpv(asset price, strike, rate, maturity, option price).

periments, the number of samples for both the Euler simulation scheme and our sampling scheme is 10^6, the number of discretization steps for the Euler scheme is 200.

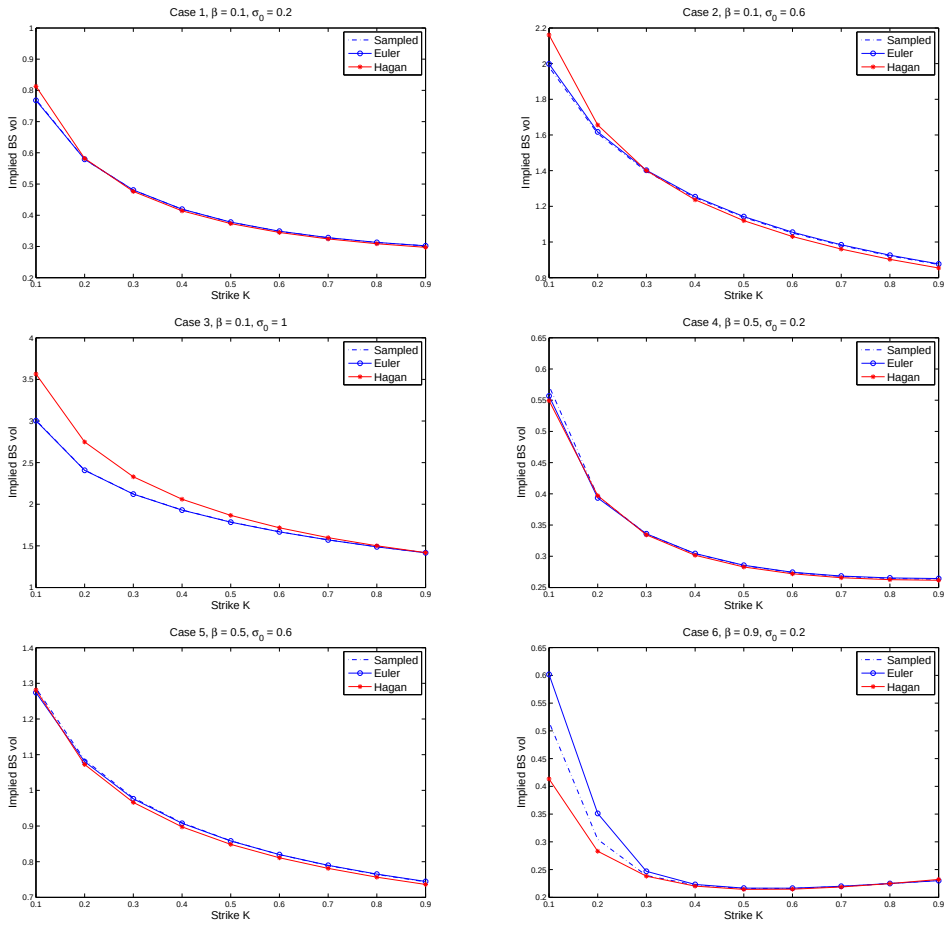

Figure 4.7: *Plots of sampled, simulated and asymptotic implied Black-Scholes volatilities*

Implied Black-Scholes volatilities, including the sampled ones by the scheme we propose, those simulated by the Euler Monte Carlo scheme and the ones calculated using the asymptotic formula in equation (4.24), are presented in Table 4.5, given six sets of parameters. For each parameter set, the sampled curve of implied volatilities is plotted against the simulated curve (Euler) and the asymptotic one (Hagan). From these plots

Table 4.5: *Sampled, simulated and asymptotic implied Black-Scholes volatilities*

			K=0.1	K=0.2	K=0.3	K=0.4	K_{ATM}=0.5	K=0.6	K=0.7	K=0.8	K=0.9	Option price at K_{ATM}
$\beta = 0.1$	$\sigma_0 = 0.2$	Sampled	0.7708	0.5789	0.4804	0.4191	0.3776	0.3485	0.3276	0.3126	0.3019	
		Euler	0.7681	0.5791	0.4806	0.4193	0.3779	0.3489	0.3280	0.3129	0.3021	0.0749
		Hagan	0.8129	0.5820	0.4761	0.4139	0.3732	0.3448	0.3241	0.3087	0.2971	
	$\sigma_0 = 0.6$	Sampled	1.9812	1.6079	1.3956	1.2483	1.1377	1.0505	0.9801	0.9220	0.8733	
		Euler	1.9988	1.6171	1.4021	1.2536	1.1422	1.0548	0.9841	0.9258	0.8768	0.2160
		Hagan	2.1610	1.6560	1.4002	1.2365	1.1196	1.0307	0.9602	0.9025	0.8542	
	$\sigma_0 = 1$	Sampled	3.0087	2.4104	2.1233	1.9312	1.7862	1.6697	1.5726	1.4898	1.4181	
		Euler	3.0042	2.4081	2.1217	1.9298	1.7847	1.6682	1.5713	1.4886	1.4170	0.3139
		Hagan	3.5647	2.7480	2.3300	2.0601	1.8661	1.7174	1.5987	1.501	1.4188	
$\beta = 0.5$	$\sigma_0 = 0.2$	Sampled	0.5729	0.3960	0.3357	0.3040	0.2851	0.2740	0.2677	0.2644	0.2632	
		Euler	0.5568	0.3935	0.3360	0.3044	0.2856	0.2744	0.2683	0.2653	0.2644	0.0568
		Hagan	0.5493	0.3971	0.3343	0.3016	0.2828	0.2719	0.2657	0.2626	0.2616	
	$\sigma_0 = 0.6$	Sampled	1.2862	1.0857	0.9791	0.9092	0.8586	0.8198	0.7891	0.7642	0.7434	
		Euler	1.2741	1.0805	0.9763	0.9076	0.8580	0.8200	0.7898	0.7651	0.7446	0.1660
		Hagan	1.2815	1.0726	0.9661	0.8976	0.8485	0.8110	0.7812	0.7566	0.7361	
$\beta = 0.9$	$\sigma_0 = 0.2$	Sampled	0.5161	0.3044	0.2393	0.2221	0.2160	0.2161	0.2195	0.2248	0.2306	
		Euler	0.6018	0.3512	0.2469	0.2233	0.2166	0.2165	0.2200	0.2249	0.2301	0.0431
		Hagan	0.4133	0.2831	0.2381	0.2202	0.2144	0.2147	0.2187	0.2248	0.2323	

we find that:

1. If the volatility of volatility component $\alpha = 0$, the SABR model is reduced to a CEV dynamics, then probabilities of F_T or X_{A_T} killing at 0 for all six cases are given in Table 4.2. For all six cases in our experiments, positive *alpha* increases the probability of F_T or X_{A_T} killing at 0.

2. For each case, all implied volatility curves have the same trend, e.g. in the plot for case 1, 2, 3, 4 or 5, these three curves are strictly decreasing, while for case 6, these three curves are smile shaped.

3. The implied volatility curves produced by the proposed sampling scheme are close to those produced by the Euler scheme for all cases except for case 6; this suggests that the sampled distribution of X_{A_T} is a good approximation to its true density thus our sampling scheme has good performance.

4. Differences between curves produced by two numerical schemes and Hagan's asymptotic formula are not neglectable. For most cases, their discrepancies are larger when option strikes are small, which implies that Hagan's asymptotic implied volatility formula is not accurate at the in-the-money region for most of the cases we test.

4.6 Conclusion

In this chapter, we propose a new representation of the SABR model, in which, the SABR model can be seen as a time-changed CEV dynamics. We derive the implicit form of the time-change, make a guess of its explicit solution and prove the validity of our guess in the numerical sense. The time-change is the integral of a lognormal process.

We then try to sample the SABR model based on its new representation. We first discuss properties of the CEV dynamics and how to sample it directly from its density. We then sample the time-change by approximating its time-T distribution with a lognormal distribution by matching their first two central moments. With the sampling schemes for both the CEV dynamics and the time-change, we sample the non-correlated SABR model, compare the performance of our sampling scheme with the Euler Monte Carlo scheme, and examine the accuracy of Hagan's popular asymptotic formula for the implied Black-Scholes volatility. Sampling a correlated SABR is left for further research in the future.

Chapter 5

General conclusions, contributions and further research

In this thesis, three independent research topics are presented and discussed, including exact-simulation of SABR model, estimating time-changed Lévy type models and the fast-calibration problem. These topics cover a wide range of interests of both academic researchers and market practitioners. Essential, this thesis provides three applications on how to choose, implement and use stochastic volatility models to solve problems in asset pricing.

The MCMC estimation method developed in Chapter 2 can be applied to any time-changed Lévy models and mixtures of diffusion-Lévy type models. Time-changed Lévy models, especially those employed with infinite activity processes, haven been widely

used for solving portfolio allocation problem. The main burden is that estimating Lévy models is a daunting task. Chapter 2 provides the details of estimating time-changed Lévy processes with the MCMC method and demonstrate that time-changed Lévy models can outperform Jump-diffusion models in capturing return dynamics.

A further investigation of the MCMC estimation method for multivariate Lévy processes is presented and evaluated in in Chapter 3. We assess the estimation approach via simulations, comparing the results with those obtained through a standard but more computationally intensive one-step maximum likelihood estimation. Much better estimation efficiency obtained by using MCMC approach is demonstrated in the empirical results and it is suggested that Lévy can be a good alternative for portfolio theory.

A new representation of the SABR model is proposed in Chapter 4, in which the SABR model can be seen as a time-changed CEV dynamics. The implicit form of the time-change is derived, with which, its numerically explicit alternative is found. Based on this new representation of the SABR model, a sampling scheme for the uncorrelated SABR is proposed. Numerical experiments suggest that our sampling scheme for the uncorrelated SABR has comparable performance with the Euler Monte Carlo scheme.

Future research can be concluded in three aspects. The MCMC estimation method has been discussed only with one-dimensional stochastic processes. It should be extended to multi-dimensional processes to make it applicable in practical problems, e.g. the portfolio choice problem. We will also investigate the performance of multi-dimensional Lévy processes in order to fit stock indices. The new representation of the SABR model discussed

in Chapter 4 leads to a different point of view of the SABR model. As the correlation between the spot price and the volatility is crucial and cannot be omitted, follow-up work on how to sample the SABR model with nonzero ρ should be done in future.

Bibliography

Ballotta, L. and Bonfiglioli, E. (2014). Multivariate asset models using lévy processes and applications. *The European Journal of Finance*, Forthcoming.

Ballotta, L., Deelstra, G., and Rayee, G. (2015). Quanto implied correlation in a multi-lévy framework. *SSRN Working Paper*.

Barndorff-Nielsen, O. (1998). Processes of normal inverse Gaussian type. *Finance and Stochastics*, 2(4):41–68.

Borodin, A. N. and Salminen, P. (2002). *Handbook of Brownian Motion - Facts and Formula*. Birkhäuser, 2 edition.

Carr, P., Geman, H., Madan, D., and Yor, M. (2002). The fine structure of asset returns: an empirical investigation. *The Journal of Business*, 75(2):305–332.

Carr, P., Geman, H., Madan, D. B., and Yor, M. (2003). Stochastic volatility for Lévy processes. *Mathematical Finance*, 13(3):345–382.

Carr, P. and Wu, L. (2003). The finite moment log stable process and option pricing. *Journal of Finance*, 58(2):753–777.

Cont, R. and Tankov, P. (2004). *Financial Modelling with Jump Processes*. Chapman and Hall/CRC, 2nd edition.

Cox, J. C. (1975). Note on option pricing I: constant elasticity of variance diffusions. Unpublished draft, Stanford University.

Duncan, J., Randal, J., and Thomson, P. (2009). Fitting jump diffiusion processes using the em algorithm. *Contributed talk at the Australasian Meeting of the Econometric Society,*.

Eberlein, E., Frey, R., and von Hammerstein, E. (2008). *Advanced credit portfolio modeling and CDO pricing, in: Mathematics: Key Technology for the Future*. Springer.

Eraker, B., Johannes, M., and Polson, N. (2002). The impact of jumps in volatility and returns. *Journal of Finance*, 58:1269–1300.

Fama, E. (1970). Efficient capital markets: a review of theory and empirical work. *Journal of Finance*, 25(2):383–417.

Fang, F. and Oosterlee, C. W. (2008). A novel pricing method for European options based on Fourier-cosine series expansions. *SIAM Journal on Scientific Computing*, 31(2):826–848.

Hagan, P. S., Kumar, D., Lesniewski, A. S., and Woodward, D. E. (2002). Managing smile risk. *Wilmott Magazine*, pages 84–108.

Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, 6:327–343.

Hobson, D. (2010). Comparison results for stochastic volatility models via coupling. *Finance and Stochastics*, 14(1):129–152.

Hobson, D. G. and Rogers, L. C. G. (1998). Complete models with stochastic volatility. *Mathematical Finance*, 8(27-48).

Honoré (1998). Pitfalls in estimating jump-diffiusion models. SSRN Working Paper.

Jacquier, E., Polson, N., and Rossi, P. (1994). Bayesian analysis of stochastic volatility models. *Journal of Business and Economic Statistics*, 12:371–389.

Kahl, C. (2007). Modelling and simulation of stochastic volatility in finance. PhD thesis.

Kallsen, J. and Muhle-Karbe, J. (2011). Method of moment estimation in time-changed Lvy models. *Statistics and Risk Modelling*, 28(2):169C194.

Li, H., Wells, M. T., and Yu, C. L. (2008). A bayesian analysis of return dynamics with Lévy jumps. *Review of Financial Studies*, 21(5):2345–2378.

Luciano, E. and Semeraro, P. (2010). Multivariate time changes for lévy asset models: Characterization and calibration. *Journal of Computational Applied Mathematics*, 233(1937-1953).

Luciano, E.and Marena, M. S. P. (2013). Dependence calibration and portfolio fit with factor-based time changes. *Carlo Alberto Notebooks*.

Madan, D., Carr, P., and Chang, E. (1998). The variance gamma process and option pricing. *European Finance Review*, 2:79–105.

Makarov, R. N. and Glew, D. (2010). Exact simulation of bessel diffusions. *Monte Carlo Methods and Applications*, 16(3-4):283–306.

Matsumoto, H. and Yor, M. (2005). Exponential functionals of brownian motion, i: Probability laws at fixed time. *Probability Surveys*, 2:312–347.

Merton, R. (1976). Option pricing when the underlying stock returns are discontinuous. *Journal of Financial Economics*, 3:125–44.

Monroe, I. (1978). Processes that can be embedded in Brownian motion. *Annals of Probability*, 6(1):42–56.

Pelletier, D. (2006). Regime switching for dynamic correlations. *Journal of Econometrics*, 131(445-473).

Sato, K. (1999). *Lévy Processes and Infinitely Divisible Distributions*. Cambridge University Press.

Vasicek, O. (1987). Probability of loss on loan portfolio. Memo.

Appendix A

A.1 Detailed description of MCMC algorithm

A.1.1 Prior distributions

In this subsection, we suggest the proper choices of prior distributions of all models. Standard priors are described which can ease the way to derive associated posteriors. Similar choices can be found in Eraker et al. (2002) and Li et al. (2008).

- Priors for return drift: $\mu \sim N(0, 1)$
- Priors for the variance processes: $\kappa \sim N(0, 1)$ (truncated at zero), $\eta \sim N(0, 1)$ (truncated at zeros). For the parameter σ_v and ρ, we follow Jacquier et al. (1994) and use the reparametrization (ϕ_v, ω_v) where $\phi_v = \sigma_v \rho$ and $\omega_v = \sigma_v^2(1 - \rho^2)$. The priors for ϕ_v and ω_v are $\phi_v|\omega_v \sim N\left(0, \frac{\omega_v^2}{2}\right)$ and $\omega_v^2 \sim IG(2, 200)$.
- Priors for jump specifications: $\lambda_y \sim Beta(2, 40)$, $\mu_y \sim N(0, 100)$, $\sigma_y \sim IG(5, 1/20)$, $\theta \sim N(0, 1)$, $\sigma \sim IG(2.5, 1/5)$, $\nu \sim IG(10, 1/10)$ and $\alpha \sim U(1, 2)$.

A.1.2 Posterior distributions

Good choices on priors can easy the job to find standard complete conditionals of posteriors. The only parameters for which we have to rely on a Metropolis-Hastings step are the latent variables v_t. In the following, we restrict the discussion to the estimation of the HVG model. Suppose m and M are the associated hyper-parameters of priors.

- Posterior of $\mu \sim N\left(\frac{S}{W}, \frac{1}{W}\right)$, where

$$W = \frac{\Delta}{(1-\rho^2)} \sum_{t=0}^{T-1} \frac{1}{v_t} + \frac{1}{M^2}$$

$$S = \frac{1}{(1-rho^2)} \sum_{t=0}^{T-1} \frac{1}{v_t} \left(C_{t+1} - \rho \frac{D_{t+1}}{\sigma_v} \right) + \frac{m}{M^2}$$

$$C_{t+1} = Y_{t+1} - Y_t - N_{t+1}\xi_{t+1}$$

$$D_{t+1} = v_{t+1} + (\kappa\Delta - 1)v_t - \kappa\eta\Delta$$

- Posterior of $\eta \sim N\left(\frac{S}{W}, \frac{1}{W}\right) \mathbf{1}_{\eta>0}$, where

$$W = \frac{\kappa^2\Delta}{\sigma_v^2(1-\rho^2)} \sum_{t=0}^{T-1} \frac{1}{v_t} + \frac{1}{M^2}$$

$$S = \frac{\kappa}{(1-\rho^2)\sigma_v} \sum_{t=0}^{T-1} \left(\frac{D_{t+1}/\sigma_v - \rho C_{t+1}}{v_t} \right) + \frac{m}{M^2}$$

$$C_{t+1} = Y_{t+1} - Y_t - \mu\Delta - N_{t+1}\xi_{t+1}$$

$$C_{t+1} = v_{t+1} + (\kappa\Delta - 1)v_t$$

- Posterior of $\kappa \sim N\left(\frac{S}{W}, \frac{1}{W}\right) \mathbf{1}_{\kappa>0}$, where

$$W = \frac{\Delta}{(1-\rho^2)\sigma_v^2} \sum_{t=0}^{T-1} \frac{(\eta - v_t)^2}{v_t} + \frac{1}{M^2}$$

$$S = \frac{1}{\sigma_v(1-\rho^2)} \sum_{t=0}^{T-1} \frac{(\eta - v_t)(D_{t+1}/\sigma_v - \rho C_{t+1}}{v_t} + \frac{m}{M^2}$$

$$C_{t+1} = Y_{t+1} - Y_t - \mu\Delta - N_{t+1}\xi_{t+1}$$

$$D_{t+1} = v_{t+1} - v_t$$

- Posterior of $\mu_y \sim N\left(\frac{S}{W}, \frac{1}{W}\right)$, where

$$W = \frac{T}{\sigma_y^2} + \frac{1}{M^2}$$

$$S = \frac{1}{\sigma_y^2} \sum_{t=0}^{T-1} \xi_{t+1} + \frac{m}{M^2}$$

- Posterior of σ_y follows an inverse Gamma distribution

$$\sigma_y^2 \sim IG\left(\frac{T}{2} + m, \frac{1}{\frac{1}{2}\sum_{t=0}^{T-1}(\xi_{t+1} - \mu_y)^2 + \frac{1}{M}}\right)$$

- Posterior of λ_y follows a Beta distribution

$$\lambda_y \sim Beta\left(\sum_{t=0}^{T-1} N_{t+1} + m, T - \sum_{t=0}^{T-1} N_{t+1} + M\right)$$

- Posteriors of σ_v and ρ can be obtained by a reparametrization (ϕ_v, ω_v), where $\phi_v = \sigma_v\rho$ and $\omega_v = \sigma_v^2(1-\rho^2)$. The posteriors of (ϕ_v, ω_v) are

$$\omega_v \sim IG\left(\frac{T}{2} + m, \frac{1}{\frac{1}{2}\sum_{t=0}^{T-1} D_{t+1}^2 + \frac{1}{M} - \frac{S^2}{2W}}\right)$$

$$\phi_v|\omega_v \sim N\left(\frac{S}{W}, \frac{\omega_v}{W}\right)$$

where

$$W = \sum_{t=0}^{T-1} C_{t+1}^2 + 2$$
$$S = \sum_{t=0}^{T-1} C_{t+1} D_{t+1}$$
$$C_{t+1} = (Y_{t+1} - Y_t - \mu\Delta - N_t\xi_t)/\sqrt{v_t\Delta}$$
$$D_{t+1} = (v_{t+1} - v_t - \kappa(\eta - v_t)\Delta)/\sqrt{v_t\Delta}$$

- Posteriors of $\xi_{t+1} \sim N\left(\frac{S}{W}, \frac{1}{W}\right)$, where

$$W = \frac{N_{t+1}^2}{(1-\rho^2)v_t\Delta} + \frac{1}{\sigma_y^2}$$
$$S = \frac{N_{t+1}}{(1-\rho^2)v_t\Delta}\left(C_{t+1} - D_{t+1}\frac{\rho}{\sigma_v}\right) + \frac{\mu_y}{\sigma_y^2}$$
$$C_{t+1} = Y_{t+1} - Y_t - \mu\Delta$$
$$D_{t+1} = v_{t+1} - v_t\kappa(\eta - v_t)\Delta$$

- Posteriors of $N_{t+1} \sim Bernoulli\left(\frac{\alpha_1}{\alpha_1+\alpha_2}\right)$, where

$$\alpha_1 = \exp\left(-\frac{1}{2(1-\rho^2)}(A_1^2 - 2\rho A_1 B)\right)\lambda_y$$
$$\alpha_2 = \exp\left(-\frac{1}{2(1-\rho^2)}(A_2^2 - 2\rho A_2 B)\right)(1-\lambda_y)$$
$$A_1 = (Y_{t+1} - Y_t - \mu\Delta - \xi_{t+1})/\sqrt{v_t\Delta}$$
$$A_2 = (Y_{t+1} - Y_t - \mu\Delta)/\sqrt{v_t\Delta}$$
$$B = (v_{t+1} - v_t - \kappa(\eta - v_t)\Delta)/(\sigma_v\sqrt{v_t\Delta})$$

- v_{t+1} has no explicit density and we rely on a Metropolis-Hastings step for this draw.

The complete conditional distribution for Vt is proportional to

$$p(Y_{t+1}, v_{t+1}|v_t, \xi_t, N_t, \Theta)p(v_t|Y_t, v_{t-1}, \xi_t, J_t, \Theta).$$

The proposal density can be drawn by using a rejection algorithm with probability

$$\min\left(\frac{p(Y_{t+1}, v_{t+1}|\Theta, v_t^{(g)}, J_t, \xi_t, N_t)}{p(Y_{t+1}, v_{t+1}|\Theta, v_t^{(g-1)}, J_t, \xi_t, N_t)}, 1\right) \tag{A.1.1}$$

where $v_t^{(g)}$ represents the g-th round iteration.

A.2 Sampling the LS distribution

Let Θ and W be two independent random variables. $\Theta \sim U(-\frac{\pi}{2}, \frac{\pi}{2})$ and $W \sim \exp(1)$. Define $\theta_0 = arctan(\beta \tan(\pi\alpha/2))/\alpha$, then

$$Z = \begin{cases} \frac{\sin\alpha(\theta_0+\Theta)}{(\cos\alpha\theta_0\cos\Theta)^{1/\alpha}}\left[\frac{\cos(\alpha\theta_0+(\alpha-1)\Theta)}{W}\right]^{(1-\alpha)/\alpha} \\ \frac{2}{\pi}\left[\left(\frac{\pi}{2}+\beta\Theta\right)\tan\Theta - \beta\log\left(\frac{\frac{\pi}{2}W\cos\Theta}{\frac{\pi}{2}+\beta\Theta}\right)\right] \end{cases} \tag{A.2.1}$$

Appendix B

B.1 The Euler schemes for A_t and A_t^*

Here we write down the Euler simulation schemes for A_t and our guess of A_t, A_t^*, then we compare these simulation schemes and see whether our guess makes sense.

The Euler scheme for A_t

The stochastic differential equation of A_t can be written as

$$dY_{A_t} = \alpha dZ_{2,A_t}$$

$$dA_t = Y_{A_t}^2 dt,$$

and $Y_0 = \sigma_0$, $A_0 = 0$. Its Euler Monte Carlo scheme is trivial

$$A_t = A_{t-\Delta} + Y_{A_{t-\Delta}}^2 \Delta \tag{B.1.1}$$

$$Y_{A_t} = Y_{A_{t-\Delta}} + \alpha \sqrt{A_t - A_{t-\Delta}} \epsilon_1 \tag{B.1.2}$$

with $A_0 = 0$, $Y_{A_0} = Y_0 = \sigma_0$. $\epsilon_1 \sim N(0,1)$. If we substitute equation (B.1.1) in equation (B.1.2), the expression of Y_{A_t} becomes

$$\begin{aligned} Y_{A_t} &= Y_{A_{t-\Delta}} + \alpha\sqrt{Y^2_{A_{t-\Delta}}\Delta}\epsilon_1 \\ &= Y_{A_{t-\Delta}} + \alpha|Y_{A_{t-\Delta}}|\sqrt{\Delta}\epsilon_1. \end{aligned} \tag{B.1.3}$$

The Euler scheme for A_t^*

The dynamics of A_t^* has the form

$$\begin{aligned} d\sigma_t &= \alpha\sigma_t dZ_{2,t} \\ A_t^* &= \int_0^t \sigma_s^2 ds, \end{aligned}$$

whose Euler scheme is

$$\sigma_t = \sigma_{t-\Delta} + \alpha\sigma_{t-\Delta}\sqrt{\Delta}\epsilon_2 \tag{B.1.4}$$

$$A_t^* = A_{t-\Delta}^* + \sigma_{t-\Delta}^2\Delta. \tag{B.1.5}$$

with $A_0^* = 0$. $\epsilon_2 \sim N(0,1)$. As σ_t is lognormal, it is non-negative for all $t \geq 0$.

Comparison

If we choose the same set of ϵ ($\epsilon = \epsilon_1 = \epsilon_2$) in simulating sample paths for A_t and A_t^* respectively, then given equation (B.1.3) in the Euler scheme for A_t and equation (B.1.4) in the Euler scheme for A_t^*, we can find that the simulated Y_{A_t} and σ_t are identical for all

$t = \Delta, 2\Delta, ..., T$.

$$\begin{aligned} \sigma_0 &= Y_{A_0} \\ \sigma_\Delta &= Y_{A_\Delta} \\ \sigma_{2\Delta} &= Y_{A_{2\Delta}} \\ ... &= ... \\ \sigma_{T-\Delta} &= Y_{A_{T-\Delta}} \\ \sigma_T &= Y_{A_T} \end{aligned}$$

By induction, the simulated A_T in equation (B.1.1) can be written as

$$\begin{aligned} A_T &= A_0 + \Delta \left(Y_{A_0}^2 + Y_{A_\Delta}^2 + ... + Y_{A_{T-2\Delta}}^2 + Y_{A_{T-\Delta}}^2 \right) \\ &= \Delta \left(Y_{A_0}^2 + Y_{A_\Delta}^2 + ... + Y_{A_{T-2\Delta}}^2 + Y_{A_{T-\Delta}}^2 \right), \end{aligned}$$

while the simulated A_T^* in equation (B.1.5) has the form

$$\begin{aligned} A_T^* &= A_0^* + \Delta \left(\sigma_0^2 + \sigma_\Delta^2 + ... + \sigma_{T-2\Delta}^2 + \sigma_{T-\Delta}^2 \right) \\ &= \Delta \left(\sigma_0^2 + \sigma_\Delta^2 + ... + \sigma_{T-2\Delta}^2 + \sigma_{T-\Delta}^2 \right) \\ &= \Delta \left(Y_{A_0}^2 + Y_{A_\Delta}^2 + ... + Y_{A_{T-2\Delta}}^2 + Y_{A_{T-\Delta}}^2 \right) \\ &\equiv A_T. \end{aligned} \tag{B.1.6}$$

Therefore, the Euler schemes for A_t and A_t^* are pathwise identical, which is that they are indistinguishable in the numerical sense.