

ABSTRACT

Title of dissertation: A NEW LÉVY BASED SHORT-RATE MODEL
FOR THE FIXED INCOME MARKET AND
ITS ESTIMATION WITH PARTICLE FILTER

Bing Zhang, Doctor of Philosophy, 2006

Dissertation directed by: Professor Dilip Madan
Smith School of Business
University of Maryland, College Park

In this thesis two contributions are made to the area of mathematical finance. First, in order to explain the non-trivial skewness and kurtosis that is observed in the time-series data of constant maturity swap (CMS) rates, we employ the pure jump Lévy processes, i.e. in particular Variance Gamma process, to model the variation of unobservable economic factors. It is the first model to include Lévy dynamics in the short rate modeling. Specifically, the Vasicek [51] type of short rate framework is adopted, where the short rate is an affine combination of three mean-reverting state variables. Zero-coupon bonds and a few fixed income derivatives are developed under the model based on the transform method in Carr et al [13]. It is expected that the Lévy based short rate model would give more realistic explanations to the yield curve movements than Gaussian-based models.

Second, the model parameters are estimated by the *particle filter* (PF) technique. The PF has not seen wide applications in the field of financial engineering, partly due to its stringent requirement on the computing capability. However, given

cheap computing cost nowadays, the PF method is a flexible yet powerful tool in estimating state-space models with non-Gaussian dynamics, such as the Levy-based models. To customize the PF algorithm to our model, the continuous-time Lévy short rate model is cast into the discrete format by first-order forward Euler approximation. The PF technique is used to retrieve values of the unobservable factors by sequentially using readily available market prices. The optimal set of model parameters are obtained by invoking the quasi-maximum likelihood estimation.

A NEW LÉVY BASED SHORT-RATE MODEL
FOR THE FIXED INCOME MARKET AND ITS ESTIMATION
WITH PARTICLE FILTER

by

Bing Zhang

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2006

Advisory Committee:

Professor Dilip Madan, Chair/Advisor
Professor Michael Fu
Professor David Levermore
Professor Konstantina Trivisa
Professor Haluk Unal

© Copyright by
Bing Zhang
2006

DEDICATION

In memory of my father Zhang, Jianguo.

To my mother Bao, Xinhua.

ACKNOWLEDGMENTS

I want to first give my utmost gratitude to a respectful professor, an extraordinary scholar, a skillful professional, and a lovable friend — my advisor Dr. Dilip Madan. Since becoming his student, I have witnessed and have been deeply touched by his passion for knowledge, his diligence in working, his humbleness in learning, his intolerance for ambiguity, as well as his sense of humor. These observations have been the driving force that moved me forward with my PhD pursuance throughout these years. I can not recall in retrospect how many times I wandered outside, with a handful of questions, before entering his office, being nervous, hopeless and self-in-confident, but after talking to him my mind was always filled with clarity and order. I am greatly indebted to such a rare individual for his expertise, patience, and help. It is impossible to utter my entire gratitude in one page, so the rest is just left unsaid.

I want to thank Prof. Michael Fu for his excellent organization of our weekly financial seminar, where I acquired extensive information and knowledge on most recent development in mathematical finance. Also, his outstanding teaching and off-class help on stochastic processes and simulation would be my life-time benefit.

I would like to acknowledge the help from Dr. Peter Carr at Bloomberg LP in New York. It was a wonderful experience spending the 2005 summer as an intern in his quantitative financial research group. There I received close nurturing from

Peter with a project of pricing foreign exchange barrier options under a stochastic skew model. I also want to express my gratitude for his generous advice and help on my employment.

I thank Prof. Dave Levermore for funding my living during the doctoral education, otherwise I would not have completed the dissertation alive. I would like also to extend my gratitude to Prof. Konstantina Trivisa and Prof. Haluk Unal for agreeing to serve as the PhD committee members.

I would also like to say Thanks to my fellow classmates. Christian Silva guided me through my first finance project; Samvit Prakash initiated a math finance students gathering regarding our academic/career developments; Qing Xia and I spent numerous hours together discussing math finance problems; Huaqiang Ma has been a good company of mine to go to the academic conferences; etc.

I would like to express my grateful feelings to the staff members Ms. Alverda McCoy and Ms. Liz Wincek for the laughters and joy they brought to me on every each frustrating and endless research day.

I am sure this enumeration is far from being complete for people that I own debt to; any inadvertent missing in the list is my fault.

In the end, I want to thank my graceful wife Fei Lin. Without her love and support, this dissertation would be infeasible.

TABLE OF CONTENTS

List of Tables	vii
List of Figures	viii
1 Introduction	1
2 Lévy Processes in Finance and Monte Carlo Simulation	8
2.1 Background	8
2.2 Mathematics	10
2.3 Pure jump Lévy market models	12
2.3.1 Variance Gamma process	14
2.3.2 CGMY process	16
2.3.3 Market models	17
2.4 Monte Carlo simulation	19
2.4.1 Simulate VG and CGMY	21
2.4.2 Simulate VG and CGMY as time-change Brownian motion . .	23
2.4.3 Importance sampling under Lévy processes	25
2.4.4 Numerical results	32
3 The 3-Factor Lévy Based Short Rate Model	41
3.1 Background	41
3.2 Overview on fixed income modeling	43
3.2.1 Bond market and money account	43
3.2.2 Forward rate and Libor rate	46
3.2.3 Change of numeraire	48
3.3 Current short rate models	50
3.4 3-Factor Lévy short rate model dynamics	53
3.4.1 Motivation	53
3.4.2 Model dynamics	54
3.5 Pricing of bond and bond derivatives	57
3.5.1 Joint characteristic function	58
3.5.2 Bond	59
3.5.3 Swaps	61
3.5.4 Swaptions	63
3.5.5 Caps/Floors	68
4 Particle Filter and Parameter Estimation	73
4.1 Background	73
4.2 The problem and conceptual solution	76
4.3 Kalman filter	78
4.4 Particle filter	80
4.4.1 Importance sampling in particle filter	82
4.4.2 Sequential importance sampling	83

4.4.3	Selection of the importance density	85
4.4.4	Resampling	85
4.4.5	Generic particle filter algorithm	88
4.5	Estimate the 3-factor Lévy short rate model	89
4.5.1	Data description	89
4.5.2	Quasi-maximum likelihood estimation with particle filter	90
4.5.3	Estimation results and discussion	92
4.6	Yield curve analysis and pricing	94
4.6.1	Yield curve factor loading	100
4.6.2	Price caplet using simulation	102
5	Conclusion	106
A	108
A.1	Solution derivation of OU process, Eqn. (3.9)	108
A.2	Proof of the Bayes rule in the update stage for the conceptual solution of the filtering problem, Eqn. (4.4)	108
A.3	Proof of the posterior expansion, Eqn. (4.17)	109
	Bibliography	110

LIST OF TABLES

2.1	VG process simulation as the difference of two gamma random variables	21
2.2	CGMY process simulation as compound Poisson and Brownian motion approximation	24
2.3	VG process simulation as the time-change Brownian motion	25
2.4	CGMY process simulation as the time change Brownian motion	26
2.5	Chi-Square goodness-of-fit test numerical results for simulation of $CGMY(t; C, G, M, Y)$ process as compound Poisson process, 10^4 samples. <i>N.R.</i> stands for Not Rejected	36
2.6	Chi-Square goodness-of-fit test numerical results for simulation of $CGMY(t; C, G, M, Y)$ process as time-change Brownian motion, 10^4 samples. <i>N.R.</i> stands for Not Rejected	36
2.7	Parameters and specifications for the numerical illustration of VG measure change	39
3.1	ICA statistics for swap rates of 10 maturities	54
3.2	Numbers of model parameter for original model and reduced model	57
4.1	Sequential importance sampling algorithm	86
4.2	Resampling algorithm	88
4.3	Particle filter algorithm	88
4.4	In-sample maximum likelihood parameter estimates of 3-factor Lévy short rate model with particle filter using 10-year data from 04/25/1994 through 10/13/2004	95
4.5	In-sample estimation performance statistics of 3-factor Lévy short rate model with particle filter using 10-year data from 04/25/1994 through 10/13/2004	95
4.6	Implied volatility charts of model-generated caplet values of tenor 3-month on 10/13/2004 with maturity from 1 to 9 years and moneyness from at-the-money 1 to out-of-the-money 1.1	104

LIST OF FIGURES

2.1	Chi-Square goodness-of-fit test illustration for simulation of $CGMY(t; C, G, M, Y)$ process, as compound Poisson and time-change Brownian motion, with $t = 0.5, C = 0.8, G = 30, M = 30, Y = 0.5, 2 \times 10^4$ samples	33
2.2	Chi-Square goodness-of-fit test illustration for simulation of $CGMY(t; C, G, M, Y)$ process, as compound Poisson and time-change Brownian motion, with $t = 0.3, C = 0.5, G = 20, M = 10, Y = 0.5, 2 \times 10^4$ samples	34
2.3	CPU time comparison of compound Poisson and time-change Brownian motion for simulating $CGMY(t; C, G, M, Y)$ with $t = 0.2, C = 0.1, G = 20, M = 10$ and varying $Y, 10^4$ samples	38
2.4	Histogram comparison for $VG(t; C, G, M)$ and $VG(t; C', G', M')$ with parameters listed in Table (2.7), 2×10^4 samples	39
2.5	Convergence comparison of pricing out-of-money put options for $VG(t; C, G, M)$ and $VG(t; C', G', M')$ with importance sampling, parameters listed in Table (2.7)	40
3.1	US treasury yield curve on Feb. 09, 2005	46
4.1	Resampling scenario and mechanism in particle filter	87
4.2	Illustration of (quasi) maximum likelihood estimation using particle filter on Eurodollar futures and swap rates against market quotes from 04/25/1994 to 10/13/2004: Upper 2 panels for Eurodollar future on Libor(3m, 3m) and lower 2 panels for Eurodollar future on Libor(6m, 3m)	96
4.3	Illustration of (quasi) maximum likelihood estimation using particle filter on Eurodollar futures and swap rates against market quotes from 04/25/1994 to 10/13/2004: Upper 2 panels for Eurodollar future on Libor(12m, 3m) and lower 2 panels for 2-year swap rate	97
4.4	Illustration of (quasi) maximum likelihood estimation using particle filter on Eurodollar futures and swap rates against market quotes from 04/25/1994 to 10/13/2004: Upper 2 panels for 3-year swap rate and lower 2 panels for 5-year swap rate	98

4.5	Illustration of (quasi) maximum likelihood estimation using particle filter on Eurodollar futures and swap rates against market quotes from 04/25/1994 to 10/13/2004: Upper 2 panels for 10-year swap rate and lower 2 panels for 20-year swap rate	99
4.6	Illustration of (quasi) maximum likelihood estimation using particle filter on Eurodollar futures and swap rates against market quotes from 04/25/1994 to 10/13/2004: panels for 30-year swap rate	100
4.7	Yield Curve and Factor Loading	101
4.8	Implied volatility surface of model-generated caplet values, 10/13/2004	104

Chapter 1

Introduction

The past few decades has seen rapid development in the fixed income market worldwide. As bonds issued by governments and corporations are being actively traded between banks and funds, financial institutions on Wall Street have been aggressive in designing and trading bond derivatives to provide insurance to investors. This phenomenal progress has stimulated an exuberance of efforts, from both academia and industry, in mathematically modeling the dynamics exhibited from the fixed income market.

Modeling the dynamics of short rate, which is the instantaneous increment of a unit deposit, has been one most popular approach in modeling interest rate term structures. This stream of work was pioneered by Vasicek [51] and Cox-Ingersoll-Ross [18], and followed by Hull-White [36], etc. A common feature that all such models share is to describe the randomness in the interest rate term structure by Brownian motion. The use of Brownian motion implies that the conditional distribution of the modeled object is normal, and one only needs the mean and variance to characterize the distribution.

However, in the context of modeling financial assets, e.g. interest rates, such an assumption is in a contradiction to many aspects of the real market. For instance, the time series data of swap rates show that significant higher probabilistic moments, i.e.

skewness/kurtosis, persist (see [24]). In addition, the evidence from the fixed income derivative market indicates that the normality assumption breaks down too, as the Black-implied volatilities¹ backed from caps/floors/swaptions prices are markedly non-constant, which otherwise should have been flat if normality holds (see [9]).

Because the market clearly does not support the normality assumption, in this dissertation I seek to remedy these discrepancies in the existing fixed income models. I adopted an approach that has seen great success in modeling the equity market dynamics. This approach, as termed by pure-jump Lévy models, was originally proposed by Madan et al in [42], and some other researchers [50].

A Lévy process is a stochastic process with independent and stationary increments. It can be decomposed into three independent components: a deterministic drift process, a continuous path mean-zero diffusion and a jump process. A drifted Brownian motion is the special case of a Lévy process without jumps, while the general Lévy setting permits flexible jump structures in addition to the diffusion and drift.

In terms of building models for asset returns, it has been argued in Madan et al [42] that there is no absolute necessity to include a Brownian motion in the underlying asset movement, because in reality no asset's values move continuously (at least the movement is limited by the minimum tick size, e.g. 1 cent for traded stocks in New York Stock Exchange). Backed by this notion, the aforementioned

¹The Black-implied volatility is the market convention to quote the caps/floors/swaptions based on the Black formula [6]. The calculation of such a quantity is thus a reverse-engineering problem to back out a model parameter, the Black-implied volatility, from the known prices.

pure-jump Lévy market models use the infinite-active jump kernels, which make small jumps arrive infinitely frequent, in order to remove the diffusion component. Such models can flexibly incorporate the exhibited skewness and kurtosis in the market data by specifying the jump kernel. In specific, an uneven jump measure yields desired skewness, and a heavy tail weight in the jump measure provides adequate kurtosis. A few pure jump Lévy models for the equity market, including Variance Gamma (VG) model which is used in this work, have been proved sound to fit both the time-series of asset returns under physical measure and the smile/skew volatility surface under the risk-neutral measure. As an evidence of their market success, some of those models have been reportedly used by a couple of prestigious financial institutions on Wall Street to evaluate their equity option positions on a daily basis.

In this dissertation, I developed a 3-factor Lévy -based short rate model whose dynamics is

$$\begin{aligned} r(t) &= \alpha + \beta'x(t) , \\ dx(t) &= (a - \kappa x(t))dt + BdL(t) , \end{aligned} \tag{1.1}$$

where the short rate $r(t)$ is a scalar process expressed as an affine combination of factors $x(t)$ driven by Lévy dynamics $dL(t)$ in the background. The affine structure is specified by the scalar coefficient α and vector coefficient β , and the vector $x(t)$ follows a multidimensional mean-reverting Ornstein-Uhlenbeck (OU) process. The vector a is denoted as the long run mean-reverting level for the factors $x(t)$, κ as the mean-reverting strength, and B as the correlating matrix.

This is the first model utilizing the Lévy structure $dL(x)$ in the short rate,

as opposed to the existing Brownian motion-based models. Multi-factor models are more effective in explaining the interest rates movements, as one factor models can not decouple the correlations between bonds across different maturities. The 3-factor model is in line with the conclusion from the seminal paper by Litterman et al in [41] that found 3-factor is adequate to describe 99% of the yield curve data. The model follows the Vasicek equilibrium modeling framework (see [51]) by specifying the process under the physical measure. Because we are primarily concerned about pricing, the measure-change from the physical measure to risk-neutral pricing measure is developed. Under the risk-neutral measure, closed-form formulas for pricing bonds, swaps, caps/floors and swaptions via their characteristic functions are derived based on the transform method in Carr et al [13].

Parameter estimation is of critical importance in applying financial models. Although the Lévy -based model is structurally superior to Gaussian-based models, it is computationally more expensive and require more sophisticated algorithms for parameter estimation.

In order to acquire the stable estimation of the parameters, the model is first cast into the state-space format. In other words, the state variables are modeled as latent/unobservable factors while the model outputs observable measurements, i.e. the prices for the fixed income products. The idea of state-space model is formularized in Eqn. (1.2)

$$\begin{aligned} x_t &= (I - e^{-\kappa\Delta t})\frac{a}{\kappa} + e^{-\kappa\Delta t}x_{t-1} + B\Delta L_t, \\ z_t &= O(x_t; \Theta) + e_t, \end{aligned} \tag{1.2}$$

where the first equation approximates the OU process solution by using the first-order Euler method. Θ is the set of parameters to be estimated, and the $O(x_t; \Theta)$ in the second equation represents the non-linear pricing function w.r.t. the state variables x_t and Θ . z_t are the output prices and e_t are the normally distributed pricing errors.

We are given 10-year weekly-sampled time-series across 9 different asset values as data input. The maximum likelihood estimation (MLE) is used as the estimation algorithm, and the likelihood function is constructed for errors e_t between the market time-series quotes and model output prices. In the context of state-space model estimation, such a joint error likelihood function is built with the help from employing the filter technique such as Kalman filter or its likes. However, the widely-used Kalman filter requires strict theoretical assumptions, i.e. the linearity assumption on the measurement and propagation functions and normality assumption on the dynamics. Therefore, despite its power and convenience, Kalman filter (KF) is not immediately applicable in our model since the randomness is not Gaussian nor pricing functions linear.

We opt to use its counterpart, the particle filter (see [7, 20]), as our choice for estimating the model. Particle filter (PF) has not seen massive applications till recently with the computation cost being reduced tremendously. Contrasting to KF, which only updates two essential quantities i.e. the mean and variance, the rationale behind PF is to use a large amount of particles to represent the entire distribution. The movement of the distribution is carried out by applying forward Monte Carlo simulation sequentially. Therefore, by design, particle filter offers great flexibility as

it can accommodate arbitrary dynamic structures and measurement functions.

In a nutshell, at each time step a particle filter first makes an ex-ante prediction on the unobservable state variables \hat{x}_t . Once the new market observation z_t is received, the filter updates the state variables under the Bayes theorem. In theory, the updated state variables x_t converge to its *true* values, which are used to output the model prices by $O(x_t; \Theta)$. A filter technique recursively performs the prediction-and-update procedure at all time steps till it reaches the end of the time-series. The joint error likelihood function is then a multidimensional normal distribution across all sampled dates and assets, and the logarithm of such likelihood function is given as

$$L(\Theta) = -\frac{1}{2} \sum_{t=1}^M \sum_{i=1}^N ((z_t - \tilde{z}_t^i)'(\mathcal{R}^i)^{-1}(z_t - \tilde{z}_t^i)) , \quad (1.3)$$

where z_t is the market quote and \tilde{z}_t^i are model output. \mathcal{R} represents the covariance matrix for the errors. M and N are the number of sampling days and asset classes, respectively. Maximization of such likelihood function yields the optimal model parameters as follows

$$\Theta = \underset{\Theta}{arg \max} L(\Theta) . \quad (1.4)$$

In the dissertation, Chapter 2 introduces the Lévy process as a mathematical concept. It discusses a couple of popular pure-jump Lévy market models and their simulation algorithms. As an important tool to reduce simulation variance, importance sampling technique for jump processes is studied. Chapter 3 focuses on the development of 3-factor Lévy -based short rate. The measure change from physical measure to pricing measure (risk neutral) is derived. Under the risk-neutral mea-

sure, closed-form formulas for pricing bonds, swaps, caps/floors and swaptions via their characteristic functions are developed. In Chapter 4, the maximum likelihood estimation algorithm for the Lévy short rate model with particle filter is discussed in detail. The estimation results show that the model has certain forecasting power on the Libor and swap rates. In addition, the yield curve is analyzed using the factor loading methodology which facilitates hedging and risk management. We take caplet as an example of pricing fixed income derivatives. The model prices for caplets show qualitative improvement against the existing models, but it also indicates that the model is inadequate to price fixed income derivatives without further seasoning.

Chapter 2

Lévy Processes in Finance and Monte Carlo Simulation

2.1 Background

Mathematical finance has been an active research area after the birth of Black-Merton-Scholes (BMS) equation in 1973. However, the constant volatility assumption in BMS was obviously violated by the market observation of non-constant volatility across maturity and strikes. To explain this so-called “volatility smile/skew” phenomenon, diligent and astute researchers have proposed numerous alternatives to BMS theory. The attempts can be grouped into three classes:

1. Local volatility model: The rationale underlying the local volatility model is that future volatilities are deterministic functions of the underlying value and calendar time, and these functions are implied by the current vanilla option prices. The model was developed by Dupire [10], Rubinstein [48], Derman and Kani [19] in three independent efforts. The model retains the convenience of Black-Scholes type of hedging argument. Because of its simplicity, traders to price and hedge exotic options with the local volatility model after the model is calibrated to the vanilla option market.
2. Stochastic volatility model: The second alternative is to randomize the volatility by a second Markovian stochastic process. The stream of efforts was pio-

neered by Hull and White [36], Heston [34], etc. In those models, a non-zero correlation between the spot and volatility processes is assigned to reproduce the skewness across strikes and maturities. The Heston stochastic volatility model is one of the market standards for derivative pricing and risk management.

3. Jump model: The third approach is to keep the dynamics a one-dimensional Markovian process but add jumps to the underlying, and jump structures could be made to incorporate the skew/smile exhibited from the market. Merton [45] in 1976 proposed the first jump-diffusion model for the underlying. Recently, pure jump Lévy market models become increasingly popular, represented by Variance Gamma model by Madan et al [42], CGMY by Carr et al [12], etc. The Lévy models will be the building block for the work in this thesis and will be elaborated in great details herein.

The Lévy models have been proved effective in explaining the smile/skew in equity and foreign exchange (FX). People have built certain degree of belief in option prices under those models. As an evidence of the models' market success, a couple of prestigious financial houses on Wall Street have been using those models or their extensions to evaluate their option positions. In this chapter, I will first introduce the mathematics needed to understand Lévy processes. A few popular Lévy market models will then be described, including the VG and CGMY models. Monte Carlo (MC) simulation is the most widely used computational tool in derivative pricing, so the last section of this chapter will be dedicated to studying the simulation

algorithms for the VG and CGMY processes. Importance sampling technique is discussed as a variance reduction technique for MC simulation, and a couple of toy examples are presented in the end.

2.2 Mathematics

Lévy process is a stochastic process X_t that has independent and stationary increments. Consequently, it is a Markovian process with the marginal distribution of random variable X_t being infinitely divisible.

It is most common to study Lévy process by their characteristic functions. The characteristic function of a random variable X is defined as

$$\phi_X(u) = E[e^{iuX}] = \int_{-\infty}^{+\infty} e^{iux} f_X(x) dx, \quad (2.1)$$

where $f_X(x)$ is the probability density function of X and $u \in R$.¹ The characteristic function can be graphically viewed as the probability weighted average of a unit circle on the complex plane. If we denote $\phi_X(u) = e^{\psi_X(u)}$, $\psi_X(u)$ is then called the characteristic exponent of X .

A Lévy process X_t is infinitely divisible, which indicates that the characteristic function of marginal random variable X_t can be expressed as follows

$$\phi_{X_t}(u) = E[e^{iuX_t}] = e^{t\psi_{X_1}(u)} \quad (2.2)$$

where $\psi_{X_1}(u)$ is the characteristic exponent of the Lévy process at unit time. The property of infinite divisibility gives rise to a great convenience to study X_t , namely

¹ u can be extended to the complex plane.

one only needs to look at X_1 in order to investigate the distributional properties of X_t for any finite t .

A Lévy process can be decomposed into three independent components: the first is a deterministic drift with rate b , the second is a continuous path diffusion with volatility σ and the third is a jump process with the measure $\nu(dx)$. Hence, a Lévy process can be fully characterized by the combined Lévy triplet $(b, \sigma, \nu(dx))$ where $b \in R$, $\sigma \in R^+$ and $\nu(dx)$ is a measure defined on $R \setminus \{0\}$. The Lévy measure $\nu(dx)$ describes the arrival frequency of jumps with different sizes, and could be written in a functional form $\nu(dx) = k(x)dx$, where $k(x)$ is called the Lévy density.

For a one-dimensional Lévy process, the Lévy -Khintchine formula gives the expression for characteristic exponent $\psi_{X_1}(u)$ as follows

$$\psi_{X_1}(u) = bui - \frac{1}{2}\sigma^2u^2 + \int_{-\infty}^{+\infty} (e^{iux} - 1 - iux1_{[|x|<1]}) \nu(dx). \quad (2.3)$$

with

$$\int_{-\infty}^{+\infty} \min(1, x^2) \nu(dx) < \infty, \quad (2.4)$$

For a good reference of Lévy processes, please see Sato [49].

It is worth pointing out that many well-known stochastic processes are special cases of general Lévy settings. For instance, if we set $b = 0$ and let the jump density $k(x)$ vanish for all real x , a standard Brownian motion with variance σ^2 is then left. Or, if both b and σ are set zero but $k(x) = \lambda\delta(1)$, where $\delta(1)$ denotes the Dirac measure at 1, the Poisson process with arrival rate λ is restored.

The so-called pure jump Lévy models ignore the Brownian motion component but use tiny jumps to mimic the continuous movement. This could be realized by

tilting the Lévy density $k(x)$ sufficiently large as the jump size x approaches to zero.

For such pure jump processes, we can group the models into 3 exclusive categories.

Let's define

$$I = \int_{-\infty}^{+\infty} \nu(dx), \quad J = \int_{-\infty}^{+\infty} |x|\nu(dx)$$

where I and J denote the total arrival rate and total variation respectively. Thus,

a Lévy process would be called

1. Finite activity process if $I < \infty$ and $J < \infty$
2. Infinite activity but finite variation process if $I = \infty$ but $J < \infty$
3. Infinite activity and infinite variation process if $I = \infty$ and $J = \infty$

This classification is to distinguish the behavior of small jumps around the origin. Due to the high frequency of tiny jumps, infinite activity process implies that the total number of jump occurrence during any time interval is infinite. The infinite variation process, which encompasses the infinite activity process, indicates that beyond the infinite jump frequency, the summation of the absolute values of all occurred jumps goes to infinity too in any finite time interval. Both infinite activity and infinite variation processes could be used as the building block for the pure-jump Lévy market models developed in the next section.

2.3 Pure jump Lévy market models

After the 1987 equity market crash, investors flocked to buy out-of-the-money put options to protect their equity positions and this pushed up the out-of-the-

money volatility and made the negative skew more pronounced. Additionally, it is well known that the log daily returns have significant skewness and are fat-tailed. As the market evidence does not support the constant volatility Brownian motion models, people started looking for models with richer structures that yield realistic explanations. It is natural to think, after observing catastrophic market crash, that the market moves not only continuously, but jumps from time to time. As the first attempt in this regard, R. Merton in 1976 proposed a jump-diffusion model for equity market in [45] which used Brownian motion for the small movements and jumps for the large.

However, in real world trading never happens continuously but rather one trade after another, and the movement of the stock price path can not be absolutely continuous because it is at least limited by the minimum tick size (e.g. 1 cent for the traded stocks on New York Stock Exchange). So it leads to the suspicion that if the Brownian component should be absolutely needed in the model, especially in the sense of parsimonious modeling. A few researchers [42] have argued that in reality the continuous diffusion component is not statistically significant, as long as the small moves can be represented by alternative structures other than Brownian motion. Under such rationales, Brownian motion is excluded from the pure jump Lévy models.

In this section I will first describe the dynamics for a couple of popular Lévy processes such as VG and CGMY. The market models are to use these processes to describe the logarithm of the asset price. Consequently, the asset price itself follows an exponential Lévy process. Careful treatments should be given here, be-

cause to build a legitimate model for stock price, the model needs to satisfy the martingale condition under the appropriate measure and associated numeraire in order to prevent arbitrage opportunities. This section will discuss how to make the exponential Lévy process a martingale by correcting the convexity term caused by the exponentiation.

2.3.1 Variance Gamma process

Variance Gamma (VG) model developed by Madan, Carr and Chang in [42] is an elegant model that offers analytical tractability and straight forward simulation schemes, and it has been one of the most well-know pure-jump Lévy models in this area.

A VG random variable X follows a 3-parameter (σ, ν, θ) probability law, and its characteristic function is given by

$$\phi_{VG}(u; \sigma, \nu, \theta) = (1 - iu\theta\nu + \frac{1}{2}\sigma^2\nu u^2)^{-1/\nu}, \quad (2.5)$$

with $\sigma \in R^+, \nu \in R^+, \theta \in R$. The elegance of VG process lies in that its Lévy jump density can be expressed in a simple form as

$$k_{VG}(x) = \begin{cases} \frac{C \exp(Gx)}{|x|} & x < 0, \\ \frac{C \exp(-Mx)}{x} & x > 0, \end{cases} \quad (2.6)$$

where

$$\begin{aligned} C &= 1/\nu, \\ G &= (\sqrt{\frac{1}{4}\theta^2\nu^2 + \frac{1}{2}\sigma^2\nu} - \frac{1}{2}\theta\nu)^{-1}, \\ M &= (\sqrt{\frac{1}{4}\theta^2\nu^2 + \frac{1}{2}\sigma^2\nu} + \frac{1}{2}\theta\nu)^{-1}. \end{aligned} \quad (2.7)$$

One can recognize that each formula in Eqn.(2.6) is in fact the Lévy measure for a gamma random variable. This indicates that a VG random variable can be decomposed into two gamma random variables; one has positive jumps and the other has negative jumps. Under this representation, a VG random variable $X^{VG}(C, G, M)$ can be written as the difference between two gamma random variables

$$X^{VG}(C, G, M) = X^g(C, 1/M) - X^g(C, 1/G). \quad (2.8)$$

This fact leads to a straight-forward simulation algorithm that we will present in the following section.

The VG process can handle the skewness and excess kurtosis exhibited from the historical stock prices, and fit well the vanilla option volatility curve for single maturities. For instance, a negative parameter θ will result in a negative skewness, and the parameter ν provides the primary control for fat-tails in the empirical distribution. For the vanilla option market, a negative θ accounts for the negative slope in the volatility curve.

VG process can also be intuitively expressed as a time-change Brownian Motion, where the time-change process² is a gamma process. In specifics, a VG process can be obtained by substituting the deterministic time t with a gamma random variable $g(t)$ in a drifted Brownian motion $X(t) = bt + \sigma W(t)$. Under the (σ, ν, θ) parameterization we have the expression for X_{VG} as

$$X^{VG}(t) = \theta g(t) + \sigma W(g(t)), \quad (2.9)$$

where $g(t)$ follows gamma distribution $\text{gamma}(t/\nu, \nu)$.

²A time-change process is also called a subordinator in some literatures.

The concept of time-change Brownian motion has strong economics intuitions. It is clear that the market does not evolve identically every day; rather, some days the trading activities are more intensive, while other days the market is just quiet and has less trading going on. So the length of a market day is better measured by the concept of random “business time” rather than the calendar time. As a (technical) benefit, VG process could be simulated by generating the standard Brownian motion subordinated by gamma random time.

2.3.2 CGMY process

Variance Gamma process is of infinite activity but finite variation, but it can be extended to have a better control over the fine structure of asset return distribution by adding one parameter. The generalized model is called CGMY model as developed by Carr et al in [12]. Based on the parameterization (C, G, M) of VG, CGMY process adds the fourth parameter Y to the power of the denominator x as in Eqn.2.7. CGMY jump density is then given as —

$$k_{CGMY}(x) = \begin{cases} \frac{C \exp(Gx)}{|x|^{1+Y}} & x < 0, \\ \frac{C \exp(-Mx)}{x^{1+Y}} & x \geq 0, \end{cases} \quad (2.10)$$

where $C > 0, G > 0, M > 0, Y < 2$. ($Y < 2$ is to keep the process having finite second moment.)

By adding Y , we gain the flexibility to specify a finer structure. For $Y < 0$, the Lévy process is of finite activity; for $0 \leq Y < 1$, it is of infinite activity but finite variation; for $1 \leq Y < 2$ the process is of infinite activity and infinite variation. As special cases, VG process is recovered if $Y = 0$, and Kou’s double exponential model

in [40] is a finite activity process with $Y = -1$.

The characteristic function of CGMY random variable is given by

$$\phi_{CGMY}(u; C, G, M, Y) = \exp \left(C\Gamma(-Y) \{ (M - iu)^Y - M^Y + (G + iu)^Y - G^Y \} \right), \quad (2.11)$$

where the $\Gamma(\cdot)$ is the gamma function.

CGMY process also has a time-changed Brownian motion representation which has been recently discovered in Madan et al [43]. We will show it together with discussion of the CGMY simulation algorithm in the next section.

2.3.3 Market models

Lévy market models assume that the martingale component of the dynamics in the log price of X_t is given by a Lévy process, e.g. VG or CGMY. As we are most concerned about pricing derivatives, the model specification is skipped under the physical measure. Rather, the stock price dynamics as the exponential Lévy process under the risk neutral measure is given by

$$S_t = S_0 \exp \left((r + \omega)t + X_t \right), \quad (2.12)$$

where r is the risk-free interest rate and ω accounts for the “logarithm convexity correction”. The ω appears in the exponential because it is needed to make the stock process an exponential martingale, which is a general requirement in asset pricing theory in order to prevent arbitrage opportunities. It can be shown that ω is defined as

$$\exp(-\omega) = \phi(-i; \Theta)$$

where $\phi(\cdot; \cdot)$ is the characteristic function of the Lévy process and Θ represents the model parameters. For VG and CGMY models, their characteristic functions have been given in Eqn.(2.5) and (2.11) respectively.

Lévy market models have been applied to option pricing in the real world recently and the popularity has been increasing (see [17, 50] for details). Here we only want to briefly compare the Lévy models with the stochastic volatility models, in terms of their pros and cons in option pricing. Lévy models have the advantage of keeping themselves stay within the family of the one dimensional Markovian process, but the stochastic volatility models need an additional stochastic process that captures the volatility. Secondly, under stochastic volatility models, which have only continuous martingale components, are unable to generate the smile/skew volatility curve for short-dated options. This is caused by the fact that the path continuity prevents Brownian motions from generating enough variations in a short period of time dt . However, such a problem does not exist for infinite activity Lévy models because during the time interval of any length there are an infinite number of jumps occurring. Therefore, even for short-lived options such Lévy models can produce the exhibited skews.

However, pure jump Lévy models contradict the market in some other aspects. For instance, under Lévy models the implied volatility curve of the long-dated options flattens out, but the market shows significant skew in those options. This is because of the i.i.d. increment assumption that Lévy processes hold. The central limit theorem (CLT) states that the summation of a large number of i.i.d random variables approaches to the normal distribution, so for the long run the skewness

and kurtosis provided by the Lévy process are suppressed by CLT. Another contradicting example is that under the time-homogeneous Lévy models the implied volatility surface for the forward-start option ³ is theoretically an exact duplication of the current implied volatility surface, whereas the market indicates a significant difference between those two surfaces. A third unsatisfactory outcome with the Lévy models is the process's constant variance, which can not explain the volatility clustering and leverage effect observed in the historical data of the realized volatility.

To solve the above problems associated with basic Lévy models, Carr et al in [14] proposed the stochastic volatility Lévy models which essentially employ an additional Markovian process to time-change the Lévy process in the spot price dynamics. These models essentially combine the advantages of the basic Lévy and stochastic volatility models, thus simultaneously attack the aforementioned deficiencies successfully.

Although it would be very naive to claim the stochastic Lévy models are the perfect description of the real world, those models provide a closer look to help explain the option market phenomena.

2.4 Monte Carlo simulation

Monte Carlo simulation has been widely used in financial engineering with applications ranging from pricing, hedging, risk managing, etc. Compared to other methodologies in pricing derivatives, Monte Carlo simulation has the advantage of

³A plain vanilla option that comes into life on a specified future date.

being flexible and easy-to-implement. In fact, Monte Carlo simulation serves as the only approach when pricing derivatives with complicated structures such as exotics options or hybrids products that is becoming a booming business in recent years. For a full account with overviews in this regard, readers are referred to [26, 37].

It is a well-known drawback of Monte Carlo simulation that the convergence is slow, i.e. the confidence interval decreasing at a rate proportional only to the *square root* of the number of the random draws. In practise, one needs to seek for variance reduction techniques in order to speed up the convergence. Depending on specific scenarios, different variance reduction tools should be chosen. For instance, when pricing far-out-of-money options people use importance sampling to redirect the process to the interested area and then change it back by multiplying the *Radon-Nikodym* ratio. Variance reduction is especially important in derivative pricing business, because very often a customer wants a real-time quote on a product with however complicated structures.

In this section I will focus on discussing the simulation algorithms for VG and CGMY processes, and provide numerical comparisons between different simulation schemes. I will also talk about the importance sampling technique under the Lévy process, and numerical demonstrations will be presented.

2.4.1 Simulate VG and CGMY

Variance Gamma

We have learned from Eqn. (2.8) that a VG random variable could be decomposed into two gamma random variables. Based on this rationale, one could take advantage of the gamma random variable simulation and generate VG process as the difference of those two gamma random variables. The corresponding simulation algorithm, using the $\{C, G, M\}$ parameterization, is listed in Table (2.1)

Simulation of $X_t \sim \text{VG}(t; C, G, M)$

1. Generate $G_t^- \sim \text{gamma}(tC, 1/G)$
2. Generate $G_t^+ \sim \text{gamma}(tC, 1/M)$
3. Return $X_t = G_t^+ - G_t^-$

Table 2.1: VG process simulation as the difference of two gamma random variables

CGMY

Unfortunately, the elegant simulation algorithm, as in the above VG case, is not available for all Lévy processes. The most general approach to simulate an arbitrary Lévy process is to treat the jumps as compound Poisson process and sample from its Lévy density. For the compound Poisson simulation, one first calculate the jump arrival rate and simulate the point process to locate the random sequence of jump time epochs. Given a jump has occurred, we sample the jump size from the normalized Lévy density. A compound Poisson random variable $X_{CP}(t)$ is then the

summation of all jumps up to time t .

The problem of the above solution is that for infinite activity and infinite variation Lévy processes, the arrival rate is infinite for any time interval. However, one realizes that this *infinity* problem is caused by jumps of tiny sizes while the arrival rates for large jumps are always finite by definition. This provides a general way of simulating such processes — namely the small jumps are cut off and approximated while large jumps are simulated as compound Poisson. This general idea is tailored to the CGMY simulation algorithm as discussed below.

We first cut off the small jumps of absolute size less than ϵ , a small positive value⁴. Those small jumps will be approximated by Brownian motion with corresponding variance⁵. For the large jumps left, compound Poisson simulation is employed and we sample jumps using the acceptance-rejection method.

For large jumps, we first make a truncated jump density k_{CGMY}^ϵ in Eqn. (2.10) and then we have

$$k_{CGMY}^\epsilon(x) = \begin{cases} \frac{C \exp(Gx)}{|x|^{1+Y}} & x < -\epsilon, \\ \frac{C \exp(-Mx)}{x^{1+Y}} & x \geq \epsilon, \end{cases}$$

In order to make the best efficiency for the acceptance-reject method, we want to find a function $f^\epsilon(x)$ whose value is close to but always greater than $k_{CGMY}^\epsilon(x)$ at every x . For $x \geq \epsilon$, it can be shown that the function

$$f^\epsilon(x) = \frac{Y \epsilon^Y}{x^{Y+1}} 1_{[|x| \geq \epsilon]}, \quad Y < 2$$

⁴For practical purpose we pick $\epsilon = 10^{-4}$.

⁵Brownian motion provides a good approximation for small jumps as shown by Asmussen and Rusiński [1]

has an efficiency close to 1 as ϵ approaches to zero. The cumulative distribution function of $f^\epsilon(x)$ is obviously $F^\epsilon(x) = 1 - \frac{\epsilon^Y}{x^Y} 1_{[|x|>\epsilon]}$ whose inversion is simply $F_\epsilon^{-1}(x) = \epsilon u^{-1/Y}$ where u is a uniform random variable. The case of negative large jumps can be treated identically.

For jump sizes with absolute values smaller than ϵ , we calculate their variance as follows

$$\sigma_\epsilon^2 = \int_{-\epsilon}^{+\epsilon} x^2 \nu(dx) < \infty.$$

The small jumps are then approximated by a Brownian motion with drift zero and variance σ_ϵ^2 .

So far we have developed a simulation algorithm by decomposing the CGMY process into three components, i.e. the large positive jumps, the large negative jumps and small jumps, and generate the each component separately. The algorithm is listed in Table (2.2).

The I^- and I^+ are intensities for positive and negative jumps respectively, and they can be calculated as a gamma incomplete function.

2.4.2 Simulate VG and CGMY as time-change Brownian motion

The benefit of viewing a Lévy process time-changed Brownian Motion has two folds with respect to simulation. First, we avoid directly dealing with the Lévy jump density which might be difficult to sample from. Second, it provides a modularity benefit so as to fit into the existing simulation package.

Simulation of $Z_t \sim \text{CGMY}(t; C, G, M, Y)$ process

1. Calculate $I^- = C \int_{-\infty}^{-\epsilon} \frac{e^{Gx} dx}{|x|^{Y+1}}$ for $x < 0$, $I^+ = C \int_{\epsilon}^{\infty} \frac{e^{-Mx} dx}{x^{Y+1}}$ for $x > 0$
2. Simulate $N^- = \text{poisson}(tI^-)$ and $N^+ = \text{poisson}(tI^+)$, where N^- and N^+ refer to the number of negative and positive jumps respectively
3. For negative jumps, Loop $i = 1 : N^-$
 - Generate $U \sim \text{uniform}[0, 1]$
 - Repeat
 - Generate $\{W, V\} \sim \text{uniform}[0, 1]$
 - Set $X_i^- = \epsilon W^{-1/Y}$
 - set $T = \frac{C f^\epsilon(X_i^-) \epsilon^{-Y} e^{-\lambda \epsilon}}{Y k_{\text{CGMY}}^\epsilon(X_i^-)}$
 - Until $VT \leq 1$ and then store X_i^-
4. Do the same calculation for N^+ as in step (3), and store all X_i^+
5. Calculate $\sigma_\epsilon = \int_{-\epsilon}^{+\epsilon} x^2 k_{\text{CGMY}}(x) dx < \infty$
6. Simulate $X_\epsilon = \sigma_\epsilon \sqrt{t} B$ where $B \sim \text{Normal}(0, 1)$
7. Return $Z_t = - \sum_i^{N^-} X_i^- + \sum_i^{N^+} X_i^+ + X_\epsilon$

Table 2.2: CGMY process simulation as compound Poisson and Brownian motion approximation

Variance Gamma

As in Eqn.(2.9) VG is viewed as a Brownian motion subordinated by a gamma process. Below in Table (2.3) is the procedure for simulating VG process X_t with parameter set (σ, ν, θ) .

Simulation of $X_t \sim \text{VG}(t; \sigma, \nu, \theta)$

1. First step: generate $G_t \sim \text{gamma}(t/\nu, \nu)$
2. Second step: insert G_t into a Brownian motion
 - Generate a standard normal random variable $W \sim \text{Normal}(0, 1)$
 - Return $X_t = \theta G_t + \sigma \sqrt{G_t} W$

Table 2.3: VG process simulation as the time-change Brownian motion

CGMY

The expression of CGMY as a time-changed Brownian motion is recently discovered by Madan and Yor in [43] by using a random truncation upon a stable process. We skip the discussion here, but only list the simulation algorithm developed in [53] in Table (2.6).

2.4.3 Importance sampling under Lévy processes

Monte Carlo simulation is often used to evaluate integrals by sampling random points from the relevant probability distribution. However, the choice of sample distribution obviously makes a real-world difference to the efficiency of the method. For example, brutal-force simulation for a rare event under the original distribution creates a great deal of void points that contribute nothing to the calculation of the integral, and hence causes high variance and slow convergence. Importance sampling is a variance reduction technique. As inferred by the name, it samples from a different distribution under which most random draws would make non-zero contributions to the integral evaluation. The displacement caused by the distribution

Simulation of CGMY($t; C, G, M, Y$) process

1. $A = (G - M)/2, B = (G + M)/2$.
2. Take an ϵ a small value, say 0.0001. For jumps in the subordinator that is smaller than ϵ we use the expectation to replace it; for jumps bigger than ϵ we simulate it by inverting the CDF.
3. The expectation of the small jump is $d = \int_0^\epsilon y \frac{C}{y^{\frac{Y}{2}+1}} dy = \frac{C\epsilon^{1-\frac{Y}{2}}}{1-\frac{Y}{2}}$.
4. The arrival rate for jumps bigger than ϵ is $\lambda = \int_\epsilon^\infty \frac{C}{y^{1+\frac{Y}{2}}} dy = \frac{2C}{Y\epsilon^{\frac{Y}{2}}}$.
5. Let T be a pseudo-time $T = tC\sqrt{\pi}\Gamma(\frac{1+Y}{2})/2^{\frac{Y}{2}}$.
6. Generate a Poisson RV N with arrival rate $T\lambda$.
7. Generate $t_i, i = 1, \dots, N$ uniform distributed within $[0, t]$.
8. Generate jumps y_i at t_i given by $y_i = \frac{\epsilon}{(1-u_{1i})^{\frac{2}{Y}}}$, where u_{1i} is an independent uniform sequence. (Inversion of the Normalized Lévy density)
9. $S(t) = dt + \sum y_i 1_{h(y_i) > u_{2i}}$, where u_{2i} is another independent uniform sequence. (the calculation of $h(y_i)$ is presented below)
10. Once $S(t)$ is known, then the CGMY process $X_t = AS(t) + \sqrt{S(t)}W$, where W follows $\text{normal}(0, 1)$.

Calculation of truncation function $h(y)$

1. $h(y) = e^{-\frac{(B^2-A^2)y}{2}} \frac{\Gamma(\frac{Y+1}{2})}{\Gamma(Y)\Gamma(\frac{1}{2})} 2^Y (\frac{B^2y}{2})^{\frac{Y}{2}} I(Y, B^2y, \frac{B^2y}{2})$.
2. $I(Y, 2\lambda, \lambda) = \frac{H_{-Y}(\sqrt{2\lambda})\Gamma(Y)}{(2\lambda)^{\frac{Y}{2}}}$, where $H_\nu(\cdot)$ is the Hermite function.
3. Hermite function is explicitly known in terms of Confluent Hypergeometric Function ${}_1F_1$, where

$$H_\nu(z) = 2^{\nu/2}\pi \left[\frac{1}{\Gamma(\frac{1-\nu}{2})\Gamma(\frac{1}{2})} {}_1F_1\left(\frac{-\nu}{2}, \frac{1}{2}, z^2/2\right) - \frac{z}{\sqrt{2}\Gamma(\frac{-\nu}{2})\Gamma(\frac{3}{2})} {}_1F_1\left(\frac{1-\nu}{2}, \frac{3}{2}, z^2/2\right) \right]$$

Table 2.4: CGMY process simulation as the time change Brownian motion

change will be corrected by multiplying the each individual random evaluation by a ratio, which will be discussed in details below.

Suppose $p(x)$ is a probability distribution function (pdf) and one is interested in the integral

$$I = \int f(x)p(x)dx . \quad (2.13)$$

With a large number of random draws from the original $p(x)$, the integral I could be approximated as

$$I' = \sum f(x_i) . \quad (2.14)$$

However, the pdf function $p(x)$ and the evaluated function $f(x)$ are independent so $p(x)$ could concentrate its weights in the region where $f(x)$ has no significant values. Monte Carlo approximation then becomes inefficient because most draws will be wasted in this case. To increase the efficiency, we can equivalently rewrite the integral as

$$J = \int f(x) \frac{p(x)}{q(x)} q(x) dx , \quad (2.15)$$

such that it can be approximated by

$$J' = \sum f(x_j) \frac{p(x_j)}{q(x_j)} , \quad (2.16)$$

with points x_j drawn from the new distribution $q(x)$, termed the importance distribution. $q(x)$ shares the same sample space as $p(x)$ but can be chosen to concentrate in areas where the values of $f(x)$ are non-zero. The technique reduces the variance incurred in the integral approximation.

Above defined is the concept of importance sampling on random variables. In finance however, we are dealing with time series of asset returns which are processes,

i.e. a collection of random variables indexed by time. Nevertheless, a process can be considered as a random variable X measured by the probability law \mathbb{P} in the path space Ω . Then if we construct a new probability measure \mathbb{Q} such that it is equivalent to \mathbb{P} , namely

$$\mathbb{P}(\Omega) = 1 \iff \mathbb{Q}(\Omega) = 1,$$

then we call the new probability measure a *change of measure*.

The Radon-Nikodym derivative $Z_t = \frac{d\mathbb{P}}{d\mathbb{Q}}|_{\mathcal{F}_t}$ is what we need to perform the importance sampling. It is defined a \mathcal{F}_t measurable martingale process under measure \mathbb{Q} . The rigor in defining the general Radon-Nikodym derivative Z_t for Lévy processes can be found in Sato [49]. Here we discuss the special case — pure jump Lévy process without continuous martingale component, i.e. Brownian motion.

Proposition 1 *Let (X_t, P) and (X_t, P') be two pure jump Lévy processes with the triplets $(b, 0, \nu)$ and $(b', 0, \nu')$, then $P|_{\mathcal{F}_t}$ and $P'|_{\mathcal{F}_t}$ are equivalent for all t if and only if the following two conditions are satisfied*

1. *The Lévy measures are equivalent with*

$$\int_{-\infty}^{+\infty} (e^{\phi(x)/2} - 1)^2 \nu(dx) < \infty,$$

where $\phi(x) = \ln\left(\frac{d\nu}{d\nu'}\right)$

2. *We must also have*

$$b' - b = \int_{-1}^1 x(\nu' - \nu)(dx).$$

When they are equivalent, the Radon-Nikodym derivative is

$$\frac{dP'}{dP}|_{\mathcal{F}_t} = e^{U_t} \tag{2.17}$$

where U_t is given

$$U_t = \lim_{\epsilon \rightarrow 0} \left(\sum_{s \leq t, |\Delta X_s| > \epsilon} \phi(\Delta X_s) - t \int_{|x| > \epsilon} (e^{\phi(x)} - 1) \nu(dx) \right). \quad (2.18)$$

The definition of the Radon-Nikodym derivative in the Proposition 1 indicates that in general its calculation depends on the information of all jumps $\{\Delta X_s, \forall s < t\}$ along each entire path. This is an unwanted property because it could substantially increase the requirement for computer storage and computation. For example, suppose one is applying the importance sampling technique to price far out-of-money European option. However, simulating only the terminal stock prices under the new measure is no longer sufficient to price the option, and one has to generate the whole evolution for each sample path. As a matter of fact, the benefit from the variance reduction will be largely offset, if not totally lost, by the excessive CPU time spent on the path simulation.

However, the next two propositions show that in certain cases, i.e. if the ratio of two Lévy measures is strictly exponential, we could still rely on the terminal values to calculate the Radon-Nikodym derivative. And as a matter of fact, VG and CGMY processes fall in this convenient class.

Proposition 2 *Let (X_t, \mathcal{P}) and (X_t, \mathcal{P}') be two pure jump Lévy processes with the triplets $(b, 0, \nu)$ and $(b', 0, \nu')$, then the Radon-Nikodym derivative $\frac{d\mathcal{P}'}{d\mathcal{P}}|_{\mathcal{F}_t}$ only depends on the terminal value X_t instead of the whole path $\{\Delta X_s, \forall s \leq t\}$ if and only if the ratio of the two Lévy measures $\frac{d\nu}{d\nu'}$ is a strict exponential function, i.e. $\frac{d\nu}{d\nu'} = e^{cx}$ for all x , where c is a constant.*

Proof of Proposition 2. From Eqn.(2.17), we have

$$\begin{aligned}
\frac{d\mathcal{P}'}{d\mathcal{P}}|\mathcal{F}_t &= e^{U_t} \\
&= \exp\left(\lim_{\epsilon \rightarrow 0}\left(\sum_{s \leq t, |\Delta X_s| > \epsilon} \phi(\Delta X_s) - t \int_{|x| > \epsilon} (e^{\phi(x)} - 1)\nu(dx)\right)\right) \\
&= \exp\left(\lim_{\epsilon \rightarrow 0}\left(\sum_{s \leq t, |\Delta X_s| > \epsilon} \phi(\Delta X_s)\right)\right) \exp\left(\lim_{\epsilon \rightarrow 0}\left(-t \int_{|x| > \epsilon} (e^{\phi(x)} - 1)\nu(dx)\right)\right) \\
&= \exp\left(\lim_{\epsilon \rightarrow 0}\left(\sum_{s \leq t, |\Delta X_s| > \epsilon} c\Delta X_s\right)\right) \exp\left(\lim_{\epsilon \rightarrow 0}\left(-t \int_{|x| > \epsilon} (e^{\phi(x)} - 1)\nu(dx)\right)\right) \\
&= \exp(cX_t) \exp\left(\lim_{\epsilon \rightarrow 0}\left(-t \int_{|x| > \epsilon} (e^{\phi(x)} - 1)\nu(dx)\right)\right)
\end{aligned}$$

■

As shown in the proof, under the condition in Proposition 2 the calculation of the measure change solely depends on the terminal value X_t and no longer requires the path information. We show in the next Proposition 3 how this simplified measure change is applied to Variance Gamma process.

Proposition 3 *Let \mathcal{P} and \mathcal{P}' be the measures of Variance Gamma processes following $VG(t; C, G, M)$ and $VG(t; C', G', M')$. The Lévy measures are ν and ν' , respectively. X_t is the random variable under the measure of \mathcal{P} and can be decomposed into the difference of two gamma random processes, i.e.*

$$X_t = g_t^+ - g_t^-$$

where $g_t^- \sim \text{Gamma}(Ct, G)$ and $g_t^+ \sim \text{Gamma}(Ct, M)$. Then the Radon-Nikodym derivative $\frac{d\mathcal{P}'}{d\mathcal{P}}|\mathcal{F}_t$ only depends on the terminal value g_t^- and g_t^+ but not the entire path $\{\Delta g_s^\pm, \forall s \leq t\}$ only if $C = C'$ such that

$$\frac{d\mathcal{P}'}{d\mathcal{P}}|\mathcal{F}_t = \exp(-tZ) \phi^+(g_t^+) \phi^-(-g_t^-)$$

where

$$\begin{aligned} Z &= \int_{-\infty}^{\infty} (\nu' - \nu)(dx) \\ \phi^-(x) &= e^{-(G'-G)|x|}, x < 0 \\ \phi^+(x) &= e^{-(M'-M)x}, x > 0 \end{aligned}$$

The validity of Proposition 3 can be easily proved by Proposition 2, and the extension from VG to CGMY process is straight-forward too. We skip both proofs here.

The measure change technique is used widely in pricing complicated financial derivative products. For example, for far out-of-the-money European put option a naive Monte Carlo simulation will have significant variance in the price since only a small proportion of simulated paths will end up lower than the strike. Another example is the up-and-out barrier option in which the upper barrier is far away from the spot so the probability of breaching the barrier is slim. In those cases, a limited number of simulation paths will not be sufficient to make accurate estimation of the price.

To perform importance sampling, we need to alter the process and redirect it to the interested region. For example, if the underlying dynamics follows a $\text{VG}(t; C, G, M)$ law and the task is to price a up-and-out barrier call option with a high upper barrier. So we could sample from a new measure $\text{VG}(t; C', G', M')$ with $G' > G$ but $M' < M$ so as to make more upward moves and less downward moves for the new process than the original process. Note: the parameter C and C' should be kept the same in order to satisfy the condition in Proposition 2. By this

means, more paths will be favorably sampled to shoot toward the barrier; the bias caused by the change would be corrected by multiplying the results by the Radon-Nikodym derivative developed in Proposition 2. We will show a concrete numerical example in the next subsection.

2.4.4 Numerical results

Simulation validation

In this section we first present some simulation results for CGMY process using two different algorithms — as the compound Poisson and as the time-change Brownian motion (TCBM). We draw a comparison between the two algorithms on the CPU time spent on the simulation.

Figure 2.1 qualitatively illustrates the goodness-of-fit for the two algorithms for the symmetric case ($G = M$), and Figure 2.2 for the asymmetric one ($G \neq M$). In both plots, the blue lines represent the binned simulation data while the red lines are theoretical PDFs, which are obtained from Fourier inversion from characteristic function using algorithm developed in [44]. Both figures show good fitting qualities.

We also tabulate the chi-square goodness-of-fit numerical results in Table (2.5) and (2.6) ⁶ with a quantitative view of the algorithms performance. We conduct the experiment for various configurations of model parameters for both algorithms.

⁶ χ^2 is the chi-square test statistic, k is the number of freedom, p is the chi-square CDF value, and $\chi_{\alpha,k}^2$ is the critical value at confidence level α with freedom k . The null hypothesis \mathcal{H}_0 can not be rejected (denoted by " \mathcal{H}_0 N.R." in both tables) if χ^2 is less than $\chi_{\alpha,k}^2$.

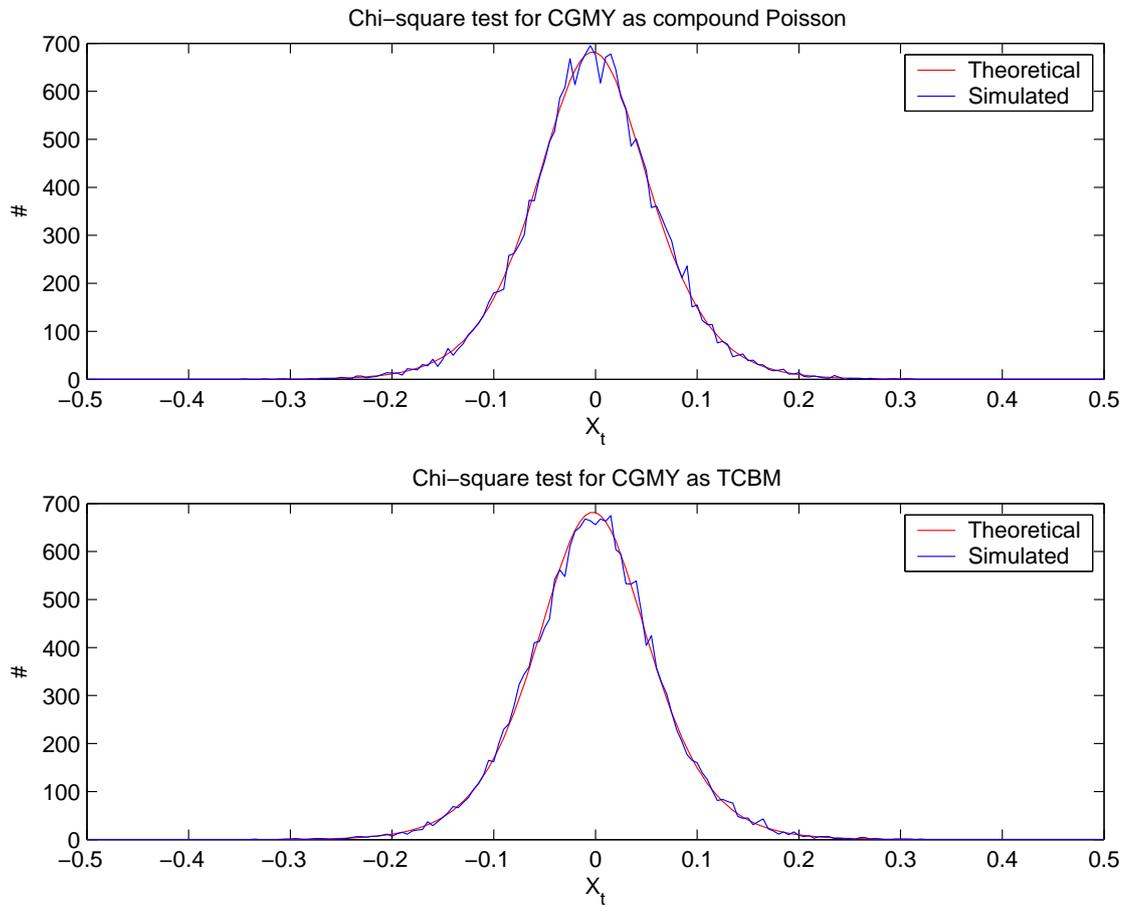


Figure 2.1: Chi-Square goodness-of-fit test illustration for simulation of $CGMY(t; C, G, M, Y)$ process, as compound Poisson and time-change Brownian motion, with $t = 0.5, C = 0.8, G = 30, M = 30, Y = 0.5, 2 \times 10^4$ samples

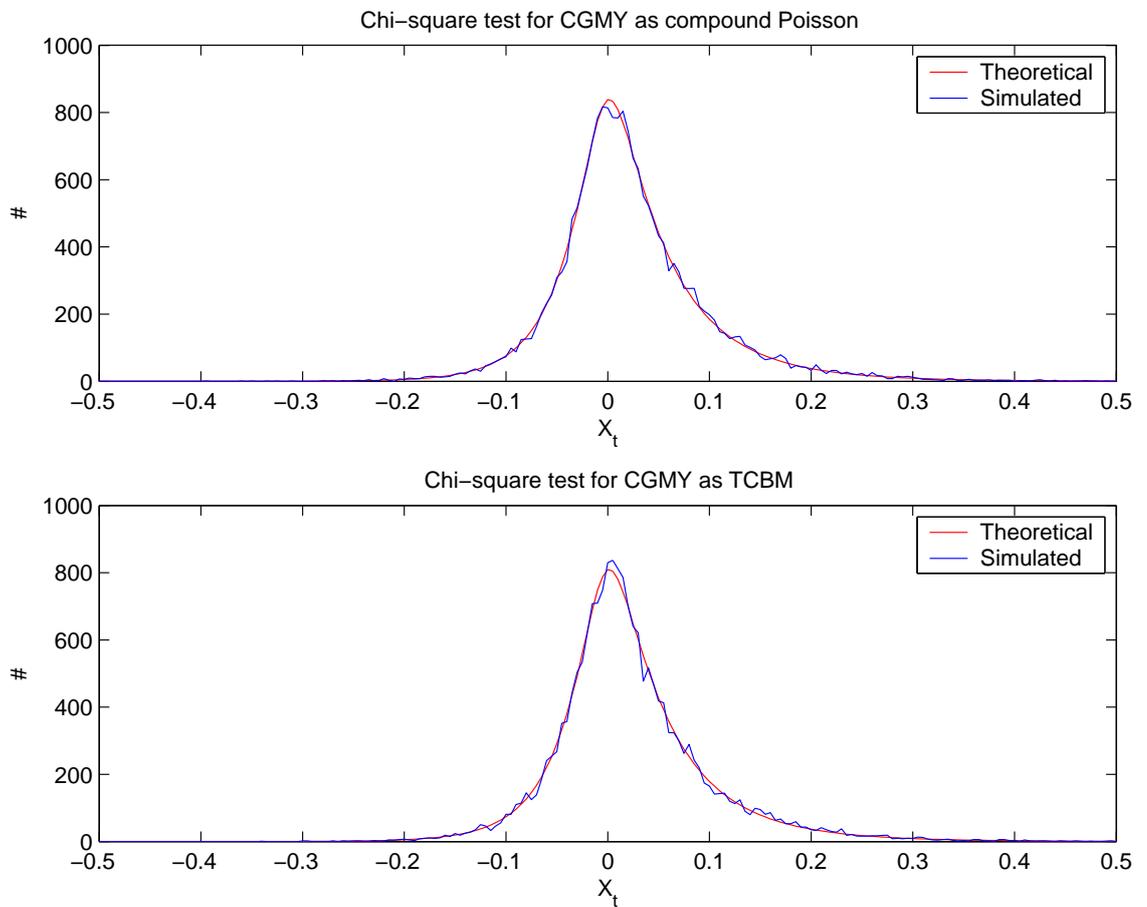


Figure 2.2: Chi-Square goodness-of-fit test illustration for simulation of $CGMY(t; C, G, M, Y)$ process, as compound Poisson and time-change Brownian motion, with $t = 0.3, C = 0.5, G = 20, M = 10, Y = 0.5, 2 \times 10^4$ samples

The tables show that in all cases the null-hypothesis \mathcal{H}_0 , which states that the simulated random variables are drawn from the specified CGMY distribution, can not be rejected for three levels of significance ($\alpha = 0.01, 0.05, 0.1$).

Efficiency comparison

Given the fact that both algorithms can successfully pass the goodness-of-fit test, we next compare the efficiencies in terms of CPU time. In CGMY process simulation, the value of parameter Y should be given special attention because it controls the behavior of small jumps, which cause most of the difficulties in simulation. For a large value of Y ($Y < 2$), the Lévy density increases sharply around the origin and hence tiny jumps happen more frequently than a smaller Y . Computation-wise, a larger Y value takes more simulation time for both algorithms. Although the compound Poisson method is more efficient for small Y , it suffers most from an increasing Y because the acceptance-rejection method would reject more and more draws as Y gets larger. The time-change Brownian motion takes more time than the compound Poisson for small Y since it involves the computation of complex functions such as the confluent hypergeometric function (the ${}_1F_1$ function), but it beats the counterpart for larger Y because the rejection rate becomes overwhelming for the compound Poisson's performance. The Figure (2.3) illustrates the following two aspects about the efficiency of the two algorithms —

1. both methods consume more CPU time with an increasing Y ,
2. the compound Poisson beats time-change Brownian motion (TCBM) for $Y \leq 1$

Parameter set					Chi-square statistics			Critical value & results		
t	C	G	M	Y	χ^2	p	k	$\chi_{0.01,k}^2$	$\chi_{0.05,k}^2$	$\chi_{0.10,k}^2$
0.5	0.8	30	30	0.5	95.77	0.42	98	133.476	122.108	116.315
								\mathcal{H}_0 <i>N.R.</i>	\mathcal{H}_0 <i>N.R.</i>	\mathcal{H}_0 <i>N.R.</i>
0.5	0.5	10	10	0.5	182.17	0.57	178	224.8	210.1	202.6
								\mathcal{H}_0 <i>N.R.</i>	\mathcal{H}_0 <i>N.R.</i>	\mathcal{H}_0 <i>N.R.</i>
0.5	0.5	30	20	0.5	80.64	0.13	95	130	118.8	113
								\mathcal{H}_0 <i>N.R.</i>	\mathcal{H}_0 <i>N.R.</i>	\mathcal{H}_0 <i>N.R.</i>
0.3	0.2	30	20	0.8	77.26	0.33	82	114.7	104.1	98.78
								\mathcal{H}_0 <i>N.R.</i>	\mathcal{H}_0 <i>N.R.</i>	\mathcal{H}_0 <i>N.R.</i>
0.3	0.2	30	20	1.2	114.24	0.09	135	176.1	163.1	156.4
								\mathcal{H}_0 <i>N.R.</i>	\mathcal{H}_0 <i>N.R.</i>	\mathcal{H}_0 <i>N.R.</i>
0.2	0.1	20	10	1.4	152.07	0.62	146	188.7	175.2	168.3
								\mathcal{H}_0 <i>N.R.</i>	\mathcal{H}_0 <i>N.R.</i>	\mathcal{H}_0 <i>N.R.</i>

Table 2.5: Chi-Square goodness-of-fit test numerical results for simulation of $CGMY(t; C, G, M, Y)$ process as compound Poisson process, 10^4 samples. *N.R.* stands for Not Rejected

Parameter set					Chi-square statistics			Critical value & results		
t	C	G	M	Y	χ^2	p	k	$\chi_{0.01,k}^2$	$\chi_{0.05,k}^2$	$\chi_{0.10,k}^2$
0.5	0.8	30	30	0.5	86.91	0.24	96	131.141	119.871	114.131
								\mathcal{H}_0 <i>N.R.</i>	\mathcal{H}_0 <i>N.R.</i>	\mathcal{H}_0 <i>N.R.</i>
0.5	0.5	10	10	0.5	186.69	0.68	178	224.8	210.1	202.6
								\mathcal{H}_0 <i>N.R.</i>	\mathcal{H}_0 <i>N.R.</i>	\mathcal{H}_0 <i>N.R.</i>
0.5	0.5	30	20	0.5	106.79	0.64	101	137	125.5	119.6
								\mathcal{H}_0 <i>N.R.</i>	\mathcal{H}_0 <i>N.R.</i>	\mathcal{H}_0 <i>N.R.</i>
0.3	0.2	30	20	0.8	62.66	0.07	80	112.3	101.9	96.58
								\mathcal{H}_0 <i>N.R.</i>	\mathcal{H}_0 <i>N.R.</i>	\mathcal{H}_0 <i>N.R.</i>
0.3	0.2	30	20	1.2	130.76	0.34	137	178.4	165.3	158.6
								\mathcal{H}_0 <i>N.R.</i>	\mathcal{H}_0 <i>N.R.</i>	\mathcal{H}_0 <i>N.R.</i>
0.2	0.1	20	10	1.4	169.60	0.85	150	193.2	179.6	172.6
								\mathcal{H}_0 <i>N.R.</i>	\mathcal{H}_0 <i>N.R.</i>	\mathcal{H}_0 <i>N.R.</i>

Table 2.6: Chi-Square goodness-of-fit test numerical results for simulation of $CGMY(t; C, G, M, Y)$ process as time-change Brownian motion, 10^4 samples. *N.R.* stands for Not Rejected

but underperforms TCBM for $Y > 1$.

Importance sampling

We next show how to improve the accuracy of Monte Carlo simulation using importance sampling. For the purpose of simplicity and convenience we take Variance Gamma process in this demonstration. We follow the Proposition 3 to perform the measure change for VG process. Assume that the underlying follows the symmetric process $\text{VG}(t; C, G, M)$ and our task is to price a pseudo far out-of-the-money put option with a deep-low strike. We pick a new $\text{VG}(t; C', G', M')$ process trending downward so that paths are more likely to breach the low strike. The model parameters and option specifications are listed in Table (2.7). With a smaller G' but larger M' we allow more negative jumps to happen than positive ones. As illustrated in Figure (2.4), the draws of random variables under the new measure is concentrated on the negative half axis, which is exactly what is needed to have more probability mass in the neighborhood of the deep-low strike. Figure (2.5) plots the averaged Monte Carlo price against the number of simulation runs for the naive Monte Carlo and Monte Carlo with importance sampling technique. Clearly, it shows the superior performance of applying importance sampling, as the option price under the new process with importance sampling converges much faster than that under the original one.

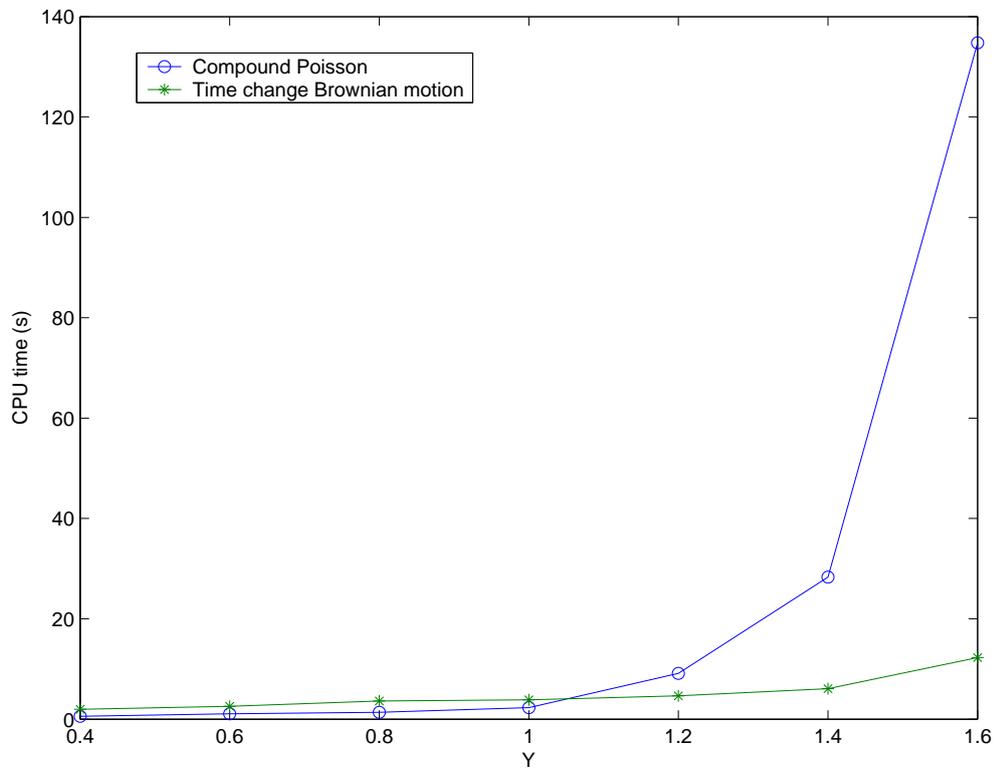


Figure 2.3: CPU time comparison of compound Poisson and time-change Brownian motion for simulating $CGMY(t; C, G, M, Y)$ with $t = 0.2, C = 0.1, G = 20, M = 10$ and varying Y , 10^4 samples

Original VG			New VG			Option specifications				
C	G	M	C'	G'	M'	Type	Strike	Maturity	Initial	Interest
5	10	10	5	7	12	Euro Put	$K = -1.5$	$T = 4$	$S_0 = 0$	$r = 0$

Table 2.7: Parameters and specifications for the numerical illustration of VG measure change

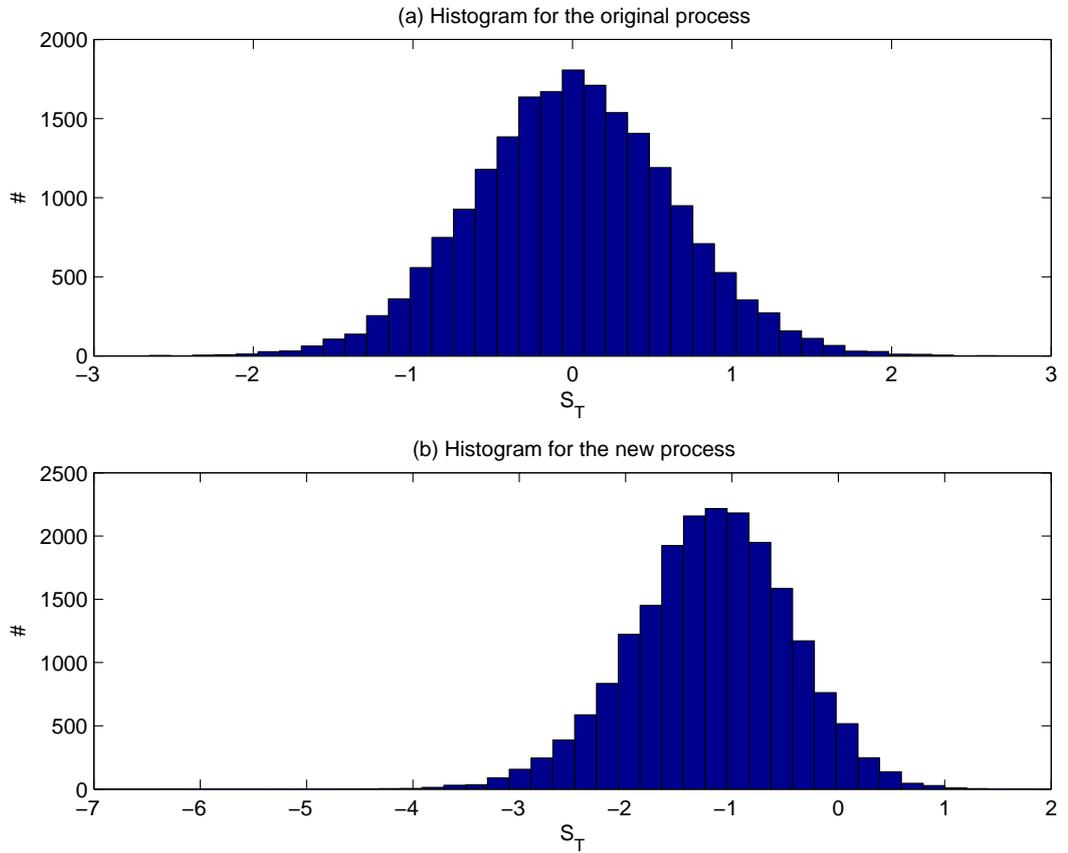


Figure 2.4: Histogram comparison for $VG(t; C, G, M)$ and $VG(t; C', G', M')$ with parameters listed in Table (2.7), 2×10^4 samples

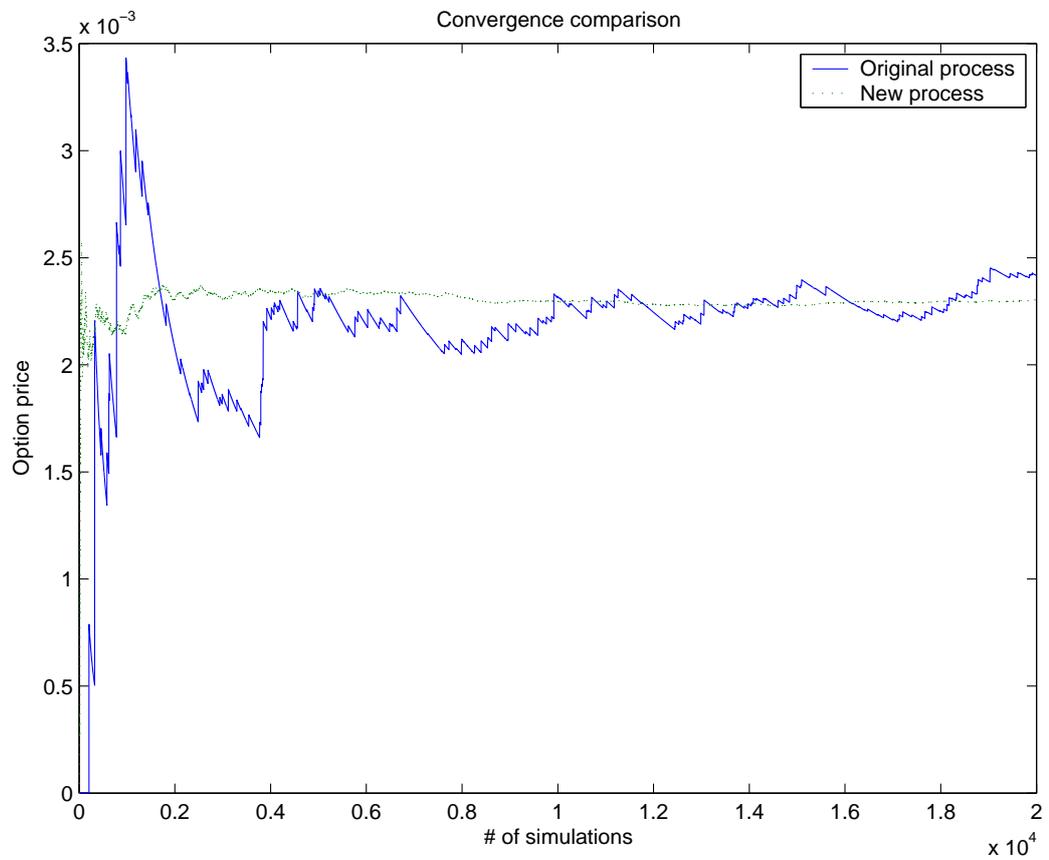


Figure 2.5: Convergence comparison of pricing out-of-the-money put options for $VG(t; C, G, M)$ and $VG(t; C', G', M')$ with importance sampling, parameters listed in Table (2.7)

Chapter 3

The 3-Factor Lévy Based Short Rate Model

3.1 Background

Whereas for modeling in equity markets one has the Black-Scholes theory arguably as a bench mark or market standard, the situation in fixed income market is more complicated. The mathematics involved in modeling the fixed income market dynamics is essentially infinite dimensional and no market standard has been established for more than special asset classes.

On the other hand, the fixed income market is much larger than equity market. It is so large that even bond derivatives are liquidly traded. On 04/25/2006, for instance, the outstanding notional for the CBOT futures contracts on government bonds with maturities of [30, 10, 5, 2] years was worth 180 billion dollars, which amounts to the 2004 GDP of the country Ireland ranked the 30th largest in the world in terms of GDP.

The research on modeling fixed income market has been active. The first attempt was the equilibrium modeling as pioneered by Vasicek [51] and Cox-Ingersoll-Ross [18] and followed by Hull-White [36], etc. Those models view the short rate, defined as the instantaneous increment of a unit deposit, as the only source of uncertainty in the economy. Short rate is then modeled by a mean-reverting process with Gaussian randomness. Short rate models are favored by hedge funds and invest-

ment banks' proprietary trading desks because those models supply the equilibrium prices which are presumably the prices the market will asymptotically converge to. However, fixed income derivative traders are not fond of this type of short rate models because they are not arbitrage free — i.e. differences between the model prices and the market quotes always exist under such models. Pricing derivatives with a model that can not make the underlying arbitrage free is dangerous and is avoided by practitioners.

No-arbitrage models came up in mid-90's represented by Heath-Jarrow-Morton (HJM) model [31]. Instead of modeling the short rate, the HJM model was to describe the forward rate dynamics. In the HJM framework, market models such as BGM/J model [8, 38] were developed for particular fixed income rates. No-arbitrage models give reasonable derivative prices, because the underlying values are perfectly matched to the market by design.

However, all the aforementioned models uses Brownian motion to describe the uncertainty. As we have discussed, the normality assumption associated with Brownian motion was barely supported by the market because Brownian motion lacks the structural properties to model many observed market phenomena.

Although the disagreement among the choice of models has not deterred banks from trading, a better model certainly helps understand the market better, and hence profit more. In this chapter we are going to develop a new short rate model based on the Vasicek framework yet without using Brownian motion. Rather, infinitely active Lévy jump structures are included in the short rate in order to reconcile the discrepancy between the market and the existing models.

As one factor models have been proved less effective in explaining the correlations between bonds across different maturities, three independent Lévy factors are employed in the model. This setup is in line with the conclusion as in [41, 32] which showed that 3 factors in the dynamics are adequate to explain more than 99% of the yield curve data historically.

In this chapter, the basic conventions and methodologies used in the fixed income modeling will be firstly introduced. A brief overview on the existing fixed income models will be provided. The degree of mathematics of general asset pricing theory will be kept minimal — only necessary concepts and notations will be included. We will spend most of this chapter developing the 3-factor Lévy based short rate model with a view on measure change. The formulas for fixed income derivatives under the model will be developed in the end.

3.2 Overview on fixed income modeling

3.2.1 Bond market and money account

Bonds are the basic element in the fixed income market. For all fixed income modeling it serves as the market input to determine model structures and estimate model parameters. In practice, the zero-coupon bond is mostly used as the reference instrument to infer the yield curve. For the sake of argument we hereafter assume that the zero-coupon bonds with continuous maturity in time are traded on the market.

A *bond* is a securitized form of loan; that is, a loan that can be traded. A zero-

coupon bond $P(t, T)$, issued at time t maturing at T , is a contract that guarantees the holder a dollar to be paid at terminal time T .¹ For brevity, such a coupon bond $P(t, T)$ is called a T -bond. In contrast to the T -bond, a *coupon bond* has intermediate coupon payments between t and T periodically.

We summarize the assumptions we have made on T -bond below:

- $P(T, T) = 1$ for all T ;
- $P(t, T) < 1$ for all $t < T$;
- There exists a market for T -bonds for every $T > 0$, and $P(t, T)$ is continuously differentiable in all T .

To complete the market, people usually assume there is a frictionless risk-free money account $B(t)$. Money account grows as follows

$$B(t) = \exp\left(\int_0^t r(s)ds\right), \quad (3.1)$$

where $B(0) = 1$ and the compounding is taken continuously. Here we have introduced the concept of a *spot rate* $r(t)$ which is defined as the instantaneous increment at time t of the money account.

Clearly $B(t)$ is a risk free asset insofar as its future value at $t + dt$ bears no uncertainty infinitesimally. $B(t)$ is also important to relate the amount of currency at different times: in order to have one dollar in the bank account at time T we need to have

$$\frac{B(t)}{B(T)} = \exp\left(-\int_t^T r(s)ds\right)$$

¹Here we assume the bond is default free.

dollars in the account at time $t \leq T$.

The relationship between a T -bond $P(0, T)$ and the money account $B(T)$ can be described as

$$P(0, T) = \exp \left(- \int_0^T r(s) ds \right) = \frac{1}{B(T)}. \quad (3.2)$$

We need another important yet straight-forward relationship linking T -bond $P(t, T)$ with spot rate $r(t)$ as given by

$$P(t, T) = \exp \left(- \int_t^T r(s) ds \right). \quad (3.3)$$

As the Eqn.(3.3) is not intuitive in knowing how much the bond grows during the period (t, T) , people designated the term *yield to maturity (or yield)* $R(t, T)$ to describe the average gain per unit time from the specific T -bond $P(t, T)$, where

$$\begin{aligned} P(t, T) &= \exp \left(- R(t, T)(T - t) \right), \\ R(t, T) &= - \frac{\log P(t, T)}{T - t}. \end{aligned} \quad (3.4)$$

It is clear that at time t , the yield is a function of maturity T given the short rate dynamics. The plot of $R(t, T)$ against the maturity T is referred as *yield curve*, the Figure 3.1 shows a upward sloping yield curve on February 9th, 2005. In theory, yield curve on different days can have all sorts of shapes, e.g. up-sloping, down-sloping or humped reflecting people's expectation on future interest rate level.

While the yield curve could be bootstrapped off from the bond prices by Eqn. (3.4), the spot rate is not directly observable. Available market proxies for the short rate include the Fed rate and short-dated, say 1-month, treasury rate, but the latter is considered better since the short-dated treasury bond is being liquidly traded.

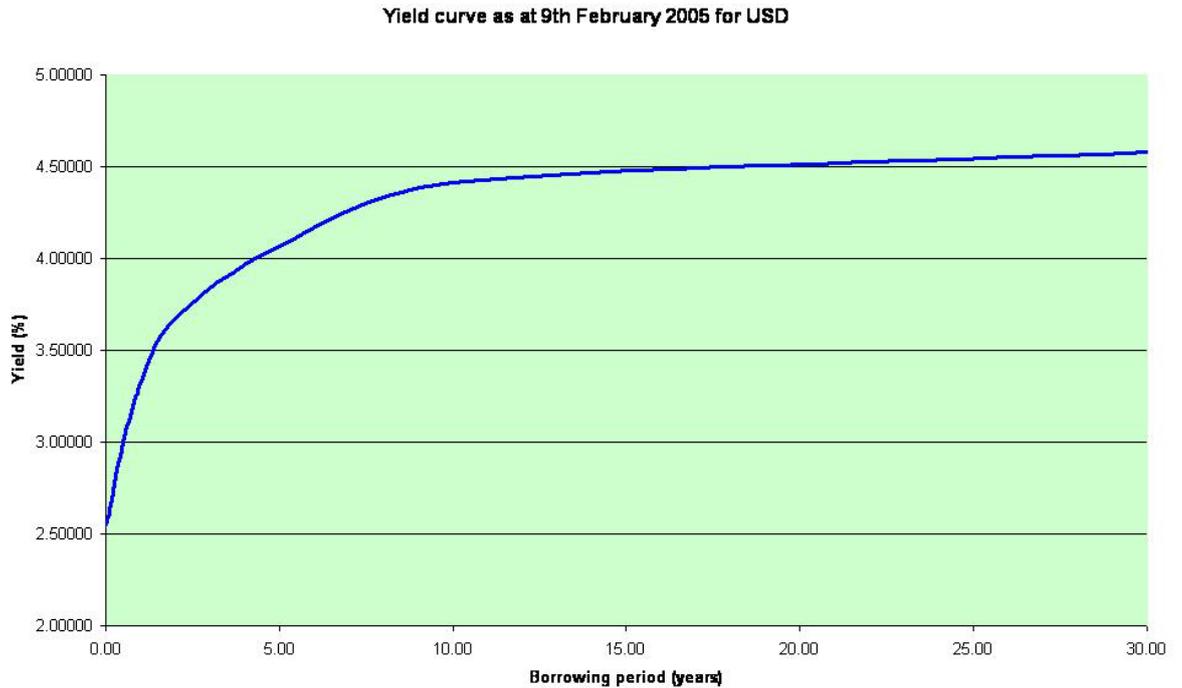


Figure 3.1: US treasury yield curve on Feb. 09, 2005

3.2.2 Forward rate and Libor rate

The term structure of zero-coupon bond prices does not contain very rich visual information, but a few better measures can be explained by implied interest rates. Below we list a variety of them including both the conceptual and the market-observable ones.

- The *discrete forward rate* $F(t; T_1, T_2)$ for period $[T_1, T_2]$ prevailing at t is defined as

$$F(t; T_1, T_2) = \frac{1}{T_2 - T_1} \left(\frac{P(t, T_1)}{P(t, T_2)} - 1 \right).$$

It can be regarded as the expected average return for future time period $[T_1, T_2]$, viewed at time t .

- The *simple spot rate* for the period $[t, T]$ is denoted by

$$F(t, T) = F(t; t, T)$$

- The *instantaneous forward rate* $f(t, T)$ with maturity T prevailing at t is defined as

$$f(t, T) = -\frac{\partial \log P(t, T)}{\partial T},$$

which determines the instantaneous gain of the continuously compounded T -bond at future time T .

- Libor (London InterBank Offer Rate) rates are the most important market observable rates and underly many interest rate derivative contracts such as swaps and caps/floors. In essence, Libor rates are simple-compounded forward rate of different tenors with the most used one being the 3-month Libor rate. A Libor rate with tenor τ is given as

$$L_t(T, T + \tau) = F(t; T, T + \tau).$$

Some simple yet important links between the bond prices and the above defined rates are summarized here. The T -bond price relates to the instantaneous forward rate in the following manner

$$P(t, T) = \exp\left(-\int_t^T f(t, u) du\right).$$

And the instantaneous forward rate approaches to the short rate as T nears t ,

$$r(t) = \lim_{T \rightarrow t} f(t, T).$$

3.2.3 Change of numeraire

Since the celebrated Black-Scholes work, the most popular pricing tool for computing asset prices has been the “risk-neutral pricing”. As argued in Harrison & Kreps [29] and Harrison & Pliska [30], the absence of arbitrage implies the existence of a risk-adjusted probability Q such that the current price of any security should equal its discounted expectation of future values. Under this equivalent measure Q the associated discounting factor, referred as the *numeraire*, is the riskless money account $B(t)$. Thus, the problem of derivative pricing is simply left as calculating the discounted expectation under risk neutral measure Q .

However, Geman et al [25] noted that the neither Q measure is necessarily the most natural choice for pricing a contingent claim nor the money account is the most convenient numeraire. In fact, under the measure Q many calculations of the expectation could be considerably complicated. In such cases, a change of numeraire can help to simplify the problem, and it has surprisingly helped reduce the complexity in pricing derivatives, especially in the fixed income market. In specific, Geman et al in [25] introduced the following definition.

Definition 4 *A numeraire is any positive non-dividend-paying asset.*

Intuitively, a numeraire is a unit asset chosen so as to normalize all other asset. With this definition, the following Proposition 5 holds.

Proposition 5 *Assume there exists a numeraire N and a probability measure Q^N , equivalent to the original measure Q^0 , such that the price of any traded asset X*

discounted by N is a (local) martingale under Q^N , i.e.,

$$\frac{X_t}{N_t} = E^N \left[\frac{X_T}{N_T} \middle| \mathcal{F}_t \right] \quad 0 \leq t \leq T.$$

Let U be an arbitrary numeraire associated with measure Q^U . Similar to (N, Q^N) , (U, Q^U) satisfies

$$\frac{X_t}{U_t} = E^U \left[\frac{X_T}{U_T} \middle| \mathcal{F}_t \right] \quad 0 \leq t \leq T.$$

Then the Radon-Nikodym derivative defining the measure Q^U is given by

$$\frac{dQ^U}{dQ^N} = \frac{U_T N_0}{U_0 N_T}. \quad (3.5)$$

By Proposition 5, the calculation of the discounted expectation of a traded asset X under an inconvenient measure could be translated under another convenient measure by

$$E^N \left[\frac{X_T}{N_T} \right] = E^U \left[\frac{X_T}{N_T} \frac{dQ^N}{dQ^U} \right] = E^U \left[\frac{U_0}{N_0} \frac{Z_T}{U_T} \right].$$

The change of numeraire technique is a useful pricing tool at one's disposal. It is typically employed when the money account discount factor under the risk-neutral measure makes the expectation calculation difficult, such as pricing caps/floors etc. To be specific we can make the following transformation

$$E^0 \left[\frac{h(X_T)}{B_T} \right] = S_0 E^S \left[\frac{h(X_T)}{S_T} \right]$$

whenever the following two properties are satisfied.

- $X_t S_t$ is a tradable asset ($0 \leq t \leq T$)
- $h(X_T)/S_T$ is conveniently simple

where $h(X_T)$ is the payoff function at time T . The first condition is to make the quantity $\frac{X_t S_t}{S_t} = X_t$ so that X_t could be modeled as a martingale under the measure Q^S ; and the second condition ensures that under the new numeraire the computation could be made simpler.

Change of numeraire is widely applied in fixed income and foreign exchange derivative pricing. As one can see in the sections where pricing of caps/floors and swaptions are discussed, the change of numeraire significantly reduced the complexity of pricing. For FX derivatives, change of numeraire is used when the derivative is priced in one currency but the underlying is dominated in another currency, e.g. a quanto option. Interested readers are referred to [3].

3.3 Current short rate models

The first seminal paper on short rate model was published by Vasicek [51] in 1977, and together with the development of interest rate market, fixed income models have been mushrooming ever since. As the first short rate model, Vasicek's work assumes the short rate evolves as an Ornstein-Uhlenbeck (OU) process with constant coefficients

$$dr(t) = \kappa(\theta - r(t))dt + \sigma dW^0(t), \quad (3.6)$$

where $dW^0(t)$ is the Brownian motion under real-world measure. As far as pricing is concerned, one needs to move the process from the physical measure to the risk neutral measure. Therefore, the concept of market price of risk is introduced as the compensation that the investors are paid to take the risk by entering a risky

contract. If we specify the market price of risk process $\lambda(t)$ to be linear with the short rate², i.e.

$$\lambda(t) = \lambda r(t)$$

with λ being a constant, the short rate process under risk-neutral measure would stay within the same OU process family. Thus, the risk-neutral short rate process is given as follows

$$dr(t) = (\kappa + \lambda\sigma)\left(\frac{\kappa\theta}{\kappa + \lambda\sigma} - r(t)\right)dt + \sigma dW(t),$$

where $dW(t)$ is the Brownian motion under risk neutral measure such that

$$dW(t) = dW^0(t) + \lambda(t)dt.$$

The solution to the SDE in Eqn.(3.6) suggests that at every time t , the random variable $r(t)$ is normally distributed. On the bright side, this property brings analytical solutions to pricing a variety of fixed income products. However, $r(t)$ being normally distributed implies the possibility of having negative short rate, which is counter-intuitive. For practical purposes, this drawback is tolerated sometimes by practitioners as they argue that, with reasonable parameter sets, the probability of $r(t)$ dropping below zero is slim. Nevertheless, it is an undesired property of having possible negative short rates.

Then there came the celebrated Cox-Ingersoll-Ross (CIR) [18] model that fixed this drawback. From its inception, CIR model has become the benchmark short rate

²In general there is no reason why the market price of risk process should take the linear form, but this form is taken to simplify the problem.

model for years. Under CIR model, the short rate $r(t)$ also follows a mean-reverting process, but it differs from Vasicek's by having the Brownian motion term multiplied by the square root of $r(t)$. The CIR dynamics is given as

$$dr(t) = \kappa(\theta - r(t))dt + \sigma\sqrt{r(t)}dW^0(t)$$

It can be shown that the solution to the SDE of CIR process follows a non-central Chi-square probability distribution, which is strictly non-negative³.

Analogous to Vasicek's, we move the process to the risk neutral measure by specifying the market price of risk under CIR model being linear in the square root of $r(t)$, i.e.

$$\lambda(t) = \lambda\sqrt{r(t)}$$

Under this formulation, the risk-neutral CIR process is given by

$$dr(t) = (\kappa + \lambda\sigma)\left(\frac{\kappa\theta}{\kappa + \lambda\sigma} - r(t)\right)dt + \sigma\sqrt{r(t)}dW(t).$$

$dW(t)$ is the Brownian motion under the risk neutral measure such that

$$dW(t) = dW^0(t) + \lambda(t)dt$$

Both Vasicek and CIR have closed-form formulas for pricing bonds and bond options under risk-neutral measure, whereas the solutions under Vasicek's are much simpler. Therefore, for the sake of simplicity, the Vasicek framework is adopted in our Lévy -based short rate model.

We have covered two of the milestone models, but a complete review about existing models is far beyond the scope of this thesis. Interested readers are referred to books by Brigo and Mercurio [9], Cairns [11], Rebonato [47] and Zagst [52]

³for a very small set of parameters, the process could hit zero.

3.4 3-Factor Lévy short rate model dynamics

The problem that puzzled the researchers for fixed income modeling resembles that for equity modeling. Historical data show strong non-normality, and the derivative market exhibits significant non-constant implied volatility. These markedly violations against the classic Black-Scholes or Black model require alternative structures to explain the market behavior.

We are encouraged by the success of pure jump Lévy models on equity modeling, and conjecture that similar problems on fixed income market could be tackled in the same manner. Therefore, in my dissertation a short rate model is developed using multi-factor Vasicek framework by replacing the Brownian motion with a pure jump Lévy dynamics.

3.4.1 Motivation

A preliminary investigation (see [24]) of changes in 10 swap rates over the period 04/25/1994 to 10/14/2004 using the methods of independent components analysis (ICA for short) revealed that they may be written as mixtures of factors with the following statistics in Table 3.1⁴.

We notice that the first three factors have highly non-zero skewness of 1.05, -1.52 , -1.12 and significant kurtosis 21.62, 132.28, 68.27 respectively. The results indicate that classic Gaussian-based short rate models such as Vasicek and CIR, regardless of the dimensionality, are inadequate to explain those time series data.

⁴We note that all ICA factors have zero means and unit variances by construction.

Factor	Skewness	Kurtosis
1	1.05	21.62
2	-1.52	132.28
3	-1.12	68.27
4	-.08	10.98
5	.15	11.72
6	.33	15.86
7	-.47	7.5
8	.37	6.5
9	.07	5.08
10	.0029	4.18

Table 3.1: ICA statistics for swap rates of 10 maturities

Besides, the statistics in Table 3.1 is confirming the conclusion as in Litterman [41] that 3-factor is sufficient to explain the variations in the yield curve.

Lévy processes, as described in Chapter 2, are ideal candidates for incorporating non-vanishing higher moments as revealed in the time series. We are then motivated to use a 3-factor Lévy dynamics to model the short rate in order to capture the characteristics shown in the table.

3.4.2 Model dynamics

Vasicek started modeling the short rate by studying the economic equilibrium, but this step is skipped here. We simply borrow the conclusion. The spot rate is written in a linear relationship with the 3 latent factors under the physical P measure as

$$r(t) = \alpha + \beta'x(t) \tag{3.7}$$

where $x(t)$ is a vector of three dimensional latent factor process, α is a real number and β is a 3-dimensional vector.

We specify $x(t)$ of following the multidimensional Ornstein-Uhlenbeck (O-U) process whereby we write the stochastic differential equation (SDE) of $x(t)$ is given as

$$dx(t) = (a - \kappa x(t))dt + BdL(t) \quad (3.8)$$

where a is a 3-dimensional vector, κ, B are full 3 by 3 matrices and L is a 3-dimensional, independent Lévy processes.

The solution of the SDE in Eqn.(3.8) takes the form

$$x(t) = e^{-\kappa t}x(0) + \int_0^t e^{-\kappa(t-s)} (ads + BdL(s)) \quad (3.9)$$

The detailed derivation can be found in Appendix A.1.

This specification determines the short rate dynamics under physical measure P . However, pricing bonds and other contingent claims must be done under a risk neutral measure Q . To change from physical measure to risk-neutral measure, we propose a specific form for the Radon-Nikodym derivative of Q with respect to P . The Radon-Nikodym derivative $\frac{dQ}{dP}$ is defined as

$$\frac{dQ}{dP} = \prod_{i=1}^3 \mathcal{E} \left((e^{\gamma_i x - \eta_i |x|} - 1) * (\mu_i - \nu_i) \right) \quad (3.10)$$

where μ_i is the random counting measure associated with the jumps of L_i , ν_i is the compensation measure for μ_i under P measure. γ_i, η_i are coefficients for risk pricing, with $(\gamma_i - \eta_i)$ being the risk premium of positive jumps and $(\gamma_i + \eta_i)$ the risk premium of negative jump risks. $\mathcal{E}(\cdot)$ denotes the stochastic exponential, which

makes the Radon-Nikodym derivative an exponential martingale. (For a complete treatment of stochastic exponential see Protter [46]) Under this proposed form of measure change, the risk neutral process and the physical process remain in the same parametric family.

Now that the model dynamics and measure change have been worked out, we choose the Variance Gamma (VG) process as the specific structure of the background driving Lévy process $L(t)$. The model parameters are listed in Table (3.2). As shown in the left column, the original model has totally 45 parameters. Apparently, this large number of parameters would overwhelm the computer program for model estimation and hence substantially slow down the optimization convergence. Therefore, we seek to simplify the model configuration by removing the redundancies in the model parameters. For instance, matrix B is to correlate $dL(t)$ and create dependence on factors $x(t)$, whereas the matrix κ has an identical effect. Hence, we can harmlessly collapse the matrix B to the constant identity matrix. Second, matrix κ does not need to be a full matrix, i.e. a lower triangular matrix (positive definite) suffices its functionality yet save 3 more parameters. Third, since we are primarily interested in pricing not the physical process of the short rate, we could assume the model parameters are under risk-neutral measure; this way we could eliminate the inclusion of six measure change parameters γ_i and η_i in Eqn. (3.10). The model parameterization is then reduced to only have 25 parameters as listed in the right column of Table 3.2. The model estimation is conducted in Chapter 4 for the reduced form.

Original Parameter Set (With physical parameters)		Reduced Parameter Set (With risk-neutral parameters)	
Symbol	Number	Symbol	Number
α	1	α	1
β	3	β	3
a	3	a	3
κ	9	κ	6, lower triangular matrix
B	9	B	0, constant identity matrix
$x(0)$	3	$x(0)$	3
$L_{1,2,3}$	9	$L_{1,2,3}$	9
$\gamma_{1,2,3}$	3	$\gamma_{1,2,3}$	0
$\eta_{1,2,3}$	3	$\eta_{1,2,3}$	0
<i>Total</i>	43	<i>Total</i>	25

Table 3.2: Numbers of model parameter for original model and reduced model

3.5 Pricing of bond and bond derivatives

In this section we discuss the analytical solutions for pricing fixed income products. It is a common practice to look at those problems by developing the characteristic function for the short rate (and its integral). As the building block to pricing derivatives, the bond pricing formula is first derived. Given the bond pricer, it is straight-forward to calculate the swap rate which can be represented by a bond portfolio. Derivatives pricing formulas, such as caps/floors and swaptions, are derived through Fourier transform.

3.5.1 Joint characteristic function

We first write the characteristic function of the multi-factor OU process $x(t)$ as follows

$$\begin{aligned}
\Phi_{x(t)}(u) &= E \left[e^{iu'x(t)} \right] = E \left[e^{iu'e^{-\kappa t}x(0) + iu' \int_0^t e^{-\kappa(t-s)}(ads + BdL(s))} \right] \\
&= \exp \left(iu'e^{-\kappa t}x(0) + iu' (I - e^{-\kappa t}) \kappa^{-1}a - \sum_{j=1}^3 \int_0^t \psi_j((u'e^{-\kappa(t-s)}B)_j)ds \right) \\
&= \exp(-\psi_{x(t)}(u)), \tag{3.11}
\end{aligned}$$

where the characteristic function for each individual Lévy process is denoted by

$$E \left[e^{ivL_j(1)} \right] = \exp(-\psi_j(v)).$$

We may then rewrite the characteristic exponent of $x(t)$ as

$$\psi_{x(t)}(u) = -iu'e^{-\kappa t}x(0) - iu' (I - e^{-\kappa t}) \kappa^{-1}a + \sum_{j=1}^3 \int_0^t \psi_j((u'e^{-\kappa(t-s)}B)_j)ds.$$

Recall the spot rate $r(t)$ as in Eqn. (3.7) is an affine combination of $x(t)$, thus the characteristic function of $r(t)$ can be straight-forwardly given as

$$\begin{aligned}
\Phi_{r(t)}(u) &= E[e^{iur(t)}] = E[e^{iu\alpha + iu\beta'x(t)}] \\
&= e^{iu\alpha} \exp(-\psi_x(u\beta)).
\end{aligned}$$

To find the bond price, we are interested in the characteristic function of the integral of $r(t)$. Below, we develop the joint characteristic function of the spot rate $r(t)$ and its integral of $\int_0^t r(s)ds$

$$\Phi_t(u, v) = E \left[\exp \left(iu \int_0^t r(s)ds + ivr(t) \right) \right].$$

We develop as follows.

$$\begin{aligned}
& E \left[\exp \left(iu \int_0^t r(s) ds + ivr(t) \right) \right] \\
= & E \left[\exp \left(iu \int_0^t (\alpha + \beta' x(s)) ds + iv(\alpha + \beta' x(t)) \right) \right] \\
= & e^{i\alpha(v+ut)} E \left[\exp \left(\int_0^t iu\beta' (e^{-\kappa s} x(0)) ds + iu\beta' \int_0^t \int_0^s e^{-\kappa(s-w)} (a dw + B dL(w)) ds \right) \right. \\
& \quad \left. \times \exp \left(iv\beta' (e^{-\kappa t} x(0)) + iv\beta' \int_0^t e^{-\kappa(t-w)} (a dw + B dL(w)) \right) \right] \\
= & e^{i\alpha(v+ut) + [iu\beta'(I - e^{-\kappa t})\kappa^{-1} + iv\beta'e^{-\kappa t}]x(0) + [iut\beta' - iu\beta'(I - e^{-\kappa t})\kappa^{-1} + iv\beta'(I - e^{-\kappa t})\kappa^{-1}]a} \\
& \times E \left[\exp \left(\int_0^t \int_0^s iu\beta' e^{-\kappa(s-w)} B dL(w) ds + \int_0^t iv\beta' e^{-\kappa(t-w)} B dL(w) \right) \right].
\end{aligned}$$

For the final expectation we note first that

$$\begin{aligned}
\int_0^t \int_0^s iu\beta' e^{-\kappa(s-w)} B dL(w) ds &= \int_0^t \int_w^t iu\beta' e^{-\kappa(s-w)} ds B dL(w) \\
&= \int_0^t iu\beta' (I - e^{-\kappa(t-w)}) \kappa^{-1} B dL(w).
\end{aligned}$$

The joint characteristic function is therefore given as

$$\begin{aligned}
\Phi_t(u, v) &= E \left[\exp \left(iu \int_0^t r(s) ds + ivr(t) \right) \right] \tag{3.12} \\
&= \exp \left(\begin{aligned} & i\alpha(v + ut) + [iu\beta'(I - e^{-\kappa t})\kappa^{-1} + iv\beta'e^{-\kappa t}]x(0) \\ & + [iut\beta' - iu\beta'(I - e^{-\kappa t})\kappa^{-1} + iv\beta'(I - e^{-\kappa t})\kappa^{-1}]a \end{aligned} \right) \\
&\times \exp \left(- \sum_j \int_0^t \psi_j \left([u\beta'(I - e^{-\kappa(t-w)})\kappa^{-1} + v\beta'e^{-\kappa(t-w)}] B_j \right) dw \right).
\end{aligned}$$

3.5.2 Bond

Equipped with the knowledge of the joint characteristic function, we could obtain the bond price $P(0,t)$, given the factors at level $x(0) = x$, by setting $u =$

$i, v = 0$ in Eqn.(3.12),

$$P(0, t) = E \left[\exp \left(- \int_0^t r(s) ds \right) \right] = \Phi_t(i, 0) = \exp(\mathcal{A}),$$

where

$$\begin{aligned} \mathcal{A} = & -\alpha t - \beta' (I - e^{-\kappa t}) \kappa^{-1} x \\ & + [\beta' (I - e^{-\kappa t}) \kappa^{-1} - t\beta'] \kappa^{-1} a \\ & - \sum_j \int_0^t \psi_j (i\beta' (I - e^{-\kappa(t-s)}) \kappa^{-1} B_j) ds. \end{aligned}$$

The matrix exponentials $\exp(-\kappa t)$ involved can be calculated by the following transformation

$$\kappa = T\Lambda T^{-1} = \sum_k \lambda_k Q_k,$$

where matrix Q_k is given by

$$Q_k = T_k R'_k,$$

and R'_k is the k^{th} row of T^{-1} . We may then expand \mathcal{A} as

$$\begin{aligned} \mathcal{A} = & -\alpha t - \beta' \kappa^{-1} x + \sum_j \beta' Q_j \kappa^{-1} x e^{-\lambda_j t} \\ & - t\beta' \kappa^{-1} a + \beta' \kappa^{-2} a - \sum_j \beta' Q_j \kappa^{-2} a e^{-\lambda_j t} \\ & - \sum_j \int_0^t \psi_j \left(i\beta' \kappa^{-1} B_j - i \sum_k \beta' Q_k \kappa^{-1} B_j e^{-\lambda_k(t-s)} \right) ds. \end{aligned}$$

Given Lévy processes are time-homogeneous, the more general bond prices

$P(s, t)$ can be expressed as

$$\begin{aligned}
P(s, t) &= \exp \left(\begin{aligned} & -\alpha(t-s) - \beta' (I - e^{-\kappa(t-s)}) \kappa^{-1} x(s) \\ & + [\beta' (I - e^{-\kappa(t-s)}) \kappa^{-1} - (t-s)\beta'] \kappa^{-1} a \end{aligned} \right) \\
&\quad \times \exp \left(- \sum_j \int_0^{t-s} \psi_j (i\beta' (I - e^{-\kappa(t-s-w)}) \kappa^{-1} B_j) dw \right) \\
&= \exp (a_{s,t} + b'_{s,t} x(s)), \tag{3.13}
\end{aligned}$$

where the coefficients $a_{s,t}$ and $b_{s,t}$ are defined as

$$\begin{aligned}
a_{s,t} &= -\alpha(t-s) + [\beta' (I - e^{-\kappa(t-s)}) \kappa^{-1} - (t-s)\beta'] \kappa^{-1} a \\
&\quad - \sum_j \int_0^{t-s} \psi_j (i\beta' (I - e^{-\kappa(t-s-w)}) \kappa^{-1} B_j) dw \\
&= -\alpha(t-s) - (t-s)\beta' \kappa^{-1} a + \beta' \kappa^{-2} a - \sum_j \beta' Q_j \kappa^{-2} a e^{-\lambda_j(t-s)} \\
&\quad - \sum_j \int_0^{t-s} \psi_j \left(i\beta' \kappa^{-1} B_j - i \sum_k \beta' Q_k \kappa^{-1} B_j e^{-\lambda_k(t-s-w)} \right) dw, \tag{3.14}
\end{aligned}$$

$$\begin{aligned}
b'_{s,t} &= -\beta' (I - e^{-\kappa(t-s)}) \kappa^{-1} \\
&= -\beta' \kappa^{-1} + \sum_j \beta' Q_j \kappa^{-1} e^{-\lambda_j(t-s)}. \tag{3.15}
\end{aligned}$$

3.5.3 Swaps

Interest rate swaps have been central in the fixed income derivative market (over-the-counter, OTC), and it is proven to be very successful in managing risks and arguably the most successful innovations in financial market. Many exotic fixed income derivatives are written with the swap rate underlying. The economic motivation behind swap is attributed to the comparative advantage; that is, different companies can borrow at different rates in different markets, but they would be better off if they are allowed to borrow from other markets that they can not access by

themselves. So the interest rate swap contract is designed to enable those companies that have different borrowing conduits to exchange the comparative benefits. See Hull in [35] for more economic explanations.

In technical terms, an interest rate swap is an agreement between two parties to exchange cash flows at a series of predefined future dates. In a vanilla payer(receiver) interest rate swap, one party agrees to pay(receive) a predetermined, fixed rate on agreed dates and receive(pay) a floating rate, often referred as Libor (London Inter-Bank Offer Rate) that is prevailing one period before. The payer swap can be specified as follows —

- notional N and a fixed rate K ;
- a number of future dates when to exchange cash flows, $T_0 < T_1 < \dots < T_n$ with $\alpha_i = T_i - T_{i-1}$ being the time fraction;⁵
- fixed-leg pays $NK\alpha_i$ and receives $N\alpha_i L_{T_{i-1}}(T_{i-1}, T_i)$ at time T_i , where the $L_{T_{i-1}}(T_{i-1}, T_i)$ is the Libor rate prevailing at T_{i-1} lasting for the period $[T_{i-1}, T_i]$.

Pricing interest rate swap is to find the fair value of the fixed rate K so that it is costless (with respect to the current yield curve) for both sides to enter such a contract at the initial time. To determine the swap contract value, let's have a closer look at a single payment at time T_i . From payer swap holder's point of view,

⁵There is no cash flow changing hands on the first date T_0 , but T_0 sets the first floating rate paid on T_1

the present value of the net cash flow at time T_i would be

$$\begin{aligned}\Pi_i(t, K) &= N\alpha_i P(t, T_i) (F(t; T_{i-1}, T_i) - K) \\ &= N (P(t, T_{i-1}) - P(t, T_i) - K\alpha_i P(t, T_i)) .\end{aligned}\quad (3.16)$$

Thus the total discounted cash flow $\Pi(t, K)$ will be the summation of the all the future discounted cash flows

$$\Pi(t, K) = \sum_i \Pi_i(t, K) = N \left(P(t, T_0) - P(t, T_n) - K \sum_{i=1}^n \alpha_i P(t, T_i) \right) .$$

The swap rate is obtained by setting $\Pi(t, K) = 0$ such that

$$K = k_t(T_0, \dots, T_n) \frac{P(t, T_0) - P(t, T_n)}{\sum_{i=1}^n \alpha_i P(t, T_i)} \quad (3.17)$$

for all $t < T_0$.

By the formula in Eqn.(3.17), the swap rate can be expressed as the ratio of two T -bond portfolios. Armed with the bond pricer developed in the previous section, the swap pricing is straight-forward following this formula.

3.5.4 Swaptions

Swaption by the name is an option written on the underlying swap. A European vanilla *payer/receiver* swaption with strike rate K is an option giving the holder the right but not obligation to enter a *payer/receiver* swap at a given future date, the maturity. Usually the maturity coincides with the first resetting date T_0 of the underlying swap, and the time from T_n to T_0 ⁶ is called the tenor of the swaption.

⁶See Section 3.5.3 for swap specification.

Recall that the value of a payer swap with fixed rate K at the first reset date T_0 is given by

$$\Pi(T_0, K) = N \sum_{i=1}^n \alpha_i P(T_0, T_i) (F(T_0; T_{i-1}, T_i) - K) ,$$

and the payoff of the swaption with strike K at maturity T_0 is

$$w(T_0, K) = N \left(\sum_{i=1}^n \alpha_i P(T_0, T_i) (F(T_0; T_{i-1}, T_i) - K) \right)^+ . \quad (3.18)$$

A swaption can be viewed as an option on a bond portfolios, and this makes it very difficult to be evaluated because it can hardly be decomposed into elementary payoffs.

Since $\Pi(T_0, K = k_{T_0})$ with k_{T_0} being the swap rate at time T_0 has to be made zero, one can show that the payoff in Eqn. (3.18) can also be written as

$$w(T_0, K) = N(k_{T_0} - K)^+ \sum_{i=1}^n \alpha_i P(T_0, T_i) . \quad (3.19)$$

Eqn. (3.19) can be easily extended to evaluate the expectation for the swaption at time $t < T_0$ that follows

$$w(t, K) = NE^Q \left[e^{-\int_t^{T_0} r_s ds} (k_{T_0} - K)^+ \sum_{i=1}^n \alpha_i P(T_0, T_i) | \mathcal{F}_t \right] , \quad (3.20)$$

where the expectation is taken on the risk-neutral measure Q .

We notice that the swaption formula in Eqn. (3.20) indicates the valuation problem can be translated as valuing a vanilla option written on the swap rate multiplied by the value of a traded bond portfolio. The difficulty lies in that, under the risk-neutral measure, the discounting factor and bond portfolio multiplier are dependent on the payoff such that they could not be taken outside the expectation operator.

To overcome this difficulty, we consider change of numeraire technique as discussed in Section 3.2.3. We take the new numeraire to be the bond portfolio value and call it *forward swap measure*, which is defined by

$$\begin{aligned} \frac{d\tilde{Q}}{dQ} = \Lambda_s &= e^{-\int_0^s r(w)dw} \frac{\sum_{i=1}^n \alpha_i P(s, T_i)}{\sum_{i=1}^n \alpha_i P(0, T_i)} \\ &= \mathcal{E} \left(\left(\tilde{Y}_s(y) - 1 \right) * (\mu - \nu) \right) \end{aligned} \quad (3.21)$$

for $0 < s < T_0$ where \tilde{Q} is denoted as the forward swap measure. Under the measure \tilde{Q} , the swap rate k_s is a positive martingale with

$$k_s = \mathcal{E} \left((Y_s^{swap}(y) - 1) * (\mu - \tilde{\nu}) \right), \quad (3.22)$$

where \mathcal{E} is the stochastic exponential (see [46]). In the above equations, μ stands for the random jump measure, and $\tilde{\nu}$ and ν the compensation jump measures associated with forward swap measure \tilde{Q} and the risk neutral measure Q , respectively.

The value of a swaption at time 0 with strike K is given by

$$w(K) = N \sum_{i=1}^n \alpha_i P(0, T_i) E^{\tilde{Q}} \left[(k_{T_0} - K)^+ \right].$$

The problem is now reduced to evaluate the expectation of $(k_{T_0} - K)^+$ under the measure \tilde{Q} . We learned from Carr and Madan [13] that the evaluation of such an expectation can be done using the transform method, i.e. the Fourier Transform, given the closed-form formula for the characteristic function of the swap rate k_s is readily known.

We then develop the characteristic function for k_s under the measure \tilde{Q} . To approach it, we need to evaluate $\tilde{Y}_s(y)$, which is the response of the measure change

process Λ_s to a jump shock in the Lévy factors. Consider a jump of size y_j in the j -th factor ($j = 1, 2, 3$). We have that

$$\frac{d\Lambda_s^j}{\Lambda_s^j} = \left(\tilde{Y}_s(y_j) - 1 \right) .$$

On the other hand, by the definition of Radon-Nikodym derivative Λ_s in Eqn. (3.21) we may observe that

$$\begin{aligned} \frac{d\Lambda_s^j}{\Lambda_s^j} &= \sum_{i=1}^n \frac{\alpha_i P(s, T_i)}{\sum_{i=1}^n \alpha_i P(s, T_i)} \left(e^{b'_{s, T_i} B_j y_j} - 1 \right) \\ &= \sum_{i=1}^n \omega_i(s) \left(e^{b'_{s, T_i} B_j y_j} - 1 \right) . \end{aligned} \quad (3.23)$$

To reduce the complexity in Eqn.(3.23), we adopt the idea of slow-varying martingale as in the BGM model [8], setting $\omega_i(s) \approx \omega(0)$. Such an approximation is empirically valid, because in reality the variability of ω_i is much less than $e^{b'_{s, T_i} B_j y_j} - 1$. Therefore, Eqn.(3.23) is simplified to be

$$\begin{aligned} \frac{d\Lambda_s^j}{\Lambda_s^j} &\approx \sum_{i=1}^n \omega_i(0) \left(e^{b'_{s, T_i} B_j y_j} - 1 \right) \\ &= \sum_{i=1}^n \frac{\alpha_i P(0, T_i)}{\sum_{i=1}^n \alpha_i P(0, T_i)} \left(e^{b'_{s, T_i} B_j y_j} - 1 \right) . \end{aligned}$$

Now we wish to see $\frac{d\Lambda_s^j}{\Lambda_s^j}$ expressed in the exponential form

$$\frac{d\Lambda_s^j}{\Lambda_s^j} = \left(e^{a_s(y_j)} - 1 \right) .$$

This implies that

$$a_s(y_j) = \ln \left(1 + \sum_{i=1}^n \frac{\alpha_i P(0, T_i)}{\sum_{i=1}^n \alpha_i P(0, T_i)} \left(e^{b'_{s, T_i} B_j y_j} - 1 \right) \right) ,$$

and taking a first order approximation gives

$$a_s(y_j) = \sum_{i=1}^n \frac{\alpha_i P(0, T_i)}{\sum_{i=1}^n \alpha_i P(0, T_i)} (b'_{s, T_i} B)_j y_j .$$

We then write

$$\tilde{Y}_s(y) = \exp \left(\sum_{j=1}^3 \sum_{i=1}^n \frac{\alpha_i P(0, T_i)}{\sum_{i=1}^n \alpha_i P(0, T_i)} (b'_{s, T_i} B)_j y_j \right).$$

We may then infer that

$$\begin{aligned} \tilde{\rho}_{s,j}(\tilde{y}_j) &= e^{c_{s,j}} \rho_j(\tilde{y}_j) \\ c_{s,j} &= \sum_{i=1}^n \frac{\alpha_i P(0, T_i)}{\sum_{i=1}^n \alpha_i P(0, T_i)} (b'_{s, T_i} B)_j, \end{aligned}$$

where $\tilde{\rho}_{s,j}$ and $\rho_{s,j}$ are Levy densities under \tilde{Q} and Q measures for the j -th factor.

Following Lévy -Khintchine theorem we have the characteristic exponent $\tilde{\psi}_{s,j}(u)$

for the j th Lévy factor at time s under the new forward swap measure,

$$\begin{aligned} \tilde{\psi}_{s,j}(u) &= - \int_{-\infty}^{\infty} (e^{iuy_j} - 1) \tilde{\rho}_{s,j}(y_j) dy_j \\ &= - \int_{-\infty}^{\infty} (e^{iuy_j} - 1) e^{c_{s,j}y_j} \rho_{s,j}(y_j) dy_j \\ &= - \int_{-\infty}^{\infty} (e^{(iu+c_{s,j})y_j} - e^{c_{s,j}y_j}) \rho_{s,j}(y_j) dy_j \\ &= - \int_{-\infty}^{\infty} (e^{(iu+c_{s,j})y_j} - 1) \rho_{s,j}(y_j) dy_j + \int_{-\infty}^{\infty} (e^{c_{s,j}y_j} - 1) \rho_{s,j}(y_j) dy_j \\ &= \psi_{s,j}^Q(u - ic_{s,j}) - \psi_{s,j}^Q(-ic_{s,j}), \end{aligned}$$

where $\psi_{s,j}^Q$ is the risk neutral characteristic exponent for Levy factor j at time s .

We now determine $Y_s^{swap}(y)$ in Eqn. (3.22). Consider now a jump of y_j in the evolution of the j -th factor. We have that

$$\frac{dk_s}{k_s} = (Y_s^{swap}(y_j) - 1).$$

Again, by direct computation

$$\frac{dk_s}{k_s} = - \frac{P(s, T_n)}{1 - P(s, T_n)} \left(e^{(b'_{s, T} B)_j y_j} - 1 \right) - \sum_{i=1}^n \frac{\alpha_i P(s, T_i)}{\sum_{i=1}^n \alpha_i P(s, T_i)} \left(e^{b'_{s, T_i} B_j y_j} - 1 \right).$$

Following the similar approximations (slow varying martingale and Taylor expanded exponential), we show the solution of $Y_s^{swap}(y)$ as

$$Y_s^{swap}(y) = \exp\left(\sum_{j=1}^3 \zeta_{s,j} y_j\right)$$

$$\zeta_{s,j} = -\frac{P(0, T_n)}{1 - P(0, T_n)} b'_{s, T_n} B_j - \sum_{i=1}^n \frac{\alpha_i P(0, T_i)}{\sum_{i=1}^n \alpha_i P(0, T_i)} b'_{s, T_i} B_j.$$

We are now able to build the characteristic function of the logarithm of swap rate process k_{T_0} at maturity T_0 under the forward swap measure by noting that

$$k_{T_0} = \exp\left(\sum_{j=1}^3 \zeta_{T_0,j} \Delta F_j - \sum_{j=1}^3 \int_0^{T_0} \int_{-\infty}^{\infty} (e^{\zeta_{s,j} y_j} - 1) \tilde{k}_{s,j}(y_j) dy_j\right).$$

We recognize from this expression that

$$E[e^{iu \ln k_{T_0}}] = \exp(-\psi_{T_0}^{swap}(u)),$$

where

$$\psi_{T_0}^{swap}(u) = -iu \sum_{j=1}^3 \int_0^{T_0} \tilde{\psi}_{s,j}(-i\zeta_{s,j}) ds + \sum_{j=1}^3 \int_0^{T_0} \tilde{\psi}_{s,j}(u\zeta_{s,j}) ds.$$

We realize at this time, with the knowledge of ψ_s^{swap} , we can fully utilize the FFT method in [13] to evaluate the swaption prices across different strikes.

3.5.5 Caps/Floors

A cap (floor) is a strip of caplets (floorlets) that gives the holder the protection over the rising (declining) rates. Thus, a cap contract would consist of the following elements —

- a notional of N and a fixed strike K ;

- a number of future dates $T_0 < T_1 < \dots < T_n$ with $T_i - T_{i-1} = \alpha_i$;
- at each time $T_i, i = 1, \dots, n$, the holder of the cap contract receives the cash flow $N\alpha_i (L_{T_{i-1}}(T_{i-1}, T_i) - K)^+$.

A floor contrasts a cap in that the floor contract holder receives the cash amount $N\alpha_i (K - L_{T_{i-1}}(T_{i-1}, T_i))^+$ at each time T_i . It is important to note the following equality, known as the cap/floor parity, holds

$$Cp(t, K) - Fl(t, K) = \Pi_p(t, K),$$

where $Cp(t, K)$ denotes the value of a cap at time t with strike K , $Fl(t, K)$ the corresponding value of a floor, and $\Pi_p(t, K)$ the value of a payer swap⁷. In this section, we show how to price the caplet under the model. The floorlet's valuation would be straight-forward using the parity relationship.

Because the present value of a cap is simply the summation of the present values of all caplets prior to maturity, a cap can therefore be viewed as a portfolio of options. This is an important distinction from a swaption contract which is an option on a portfolio, and makes the evaluation of caps is relatively easier than a swaption.

A single caplet with reset date T_{i-1} and settlement date T_i pays the holder the notional N multiplied by the difference between a Libor rate $L_{T_{i-1}}(T_{i-1}, T_i)$ and maturing at time T_i) and the strike K if $L_{T_{i-1}}(T_{i-1}, T_i) > K$, or zero otherwise. Therefore the value $w_i(K)$ of such a caplet at initial time maturing is

$$w_i(K) = N\alpha_i E^Q \left[e^{-\int_0^{T_i} r_s ds} (L_{T_{i-1}}(T_{i-1}, T_i) - K)^+ \right],$$

⁷The cap, floor and swap must have the exact coinciding dates for future cash flow exchanges.

where the expectation is taken under the risk-neutral measure Q .

Similar to the swaption case, the above expression is inconvenient for evaluation, so we seek for convenient measure change. The new measure, termed as T_i *forward measure*, is defined in the following way for $s < T_i$

$$\begin{aligned} \frac{d\tilde{Q}^{T_i}}{dQ} = \Lambda_{T_i}(s) &= e^{-\int_0^s r(w)dw} \frac{P(s, T_i)}{P(0, T_i)}, \\ &= \mathcal{E} \left(\left(\tilde{Y}_s^{T_i}(y) - 1 \right) * (\mu - \nu) \right). \end{aligned} \quad (3.24)$$

We show that

$$\frac{d\Lambda_{T_i}^j(s)}{\Lambda_{T_i}^j(s)} = \exp(b_{s,T} B_j y_j) - 1,$$

where y_j is the jump in the j -th jump component ($j = 1, 2, 3$). After identical algebraic manipulation and approximation as in the swaption case, we have

$$\tilde{Y}_s^{T_i}(y) = \exp \left(\sum_{j=1}^3 b_{s,T} B_j y_j \right).$$

The characteristic exponent for the j -th jump component under the forward measure is therefore given by

$$\begin{aligned} \tilde{\psi}_{s,j}^{T_i}(u) &= - \int_{-\infty}^{\infty} (e^{iuy_j} - 1) \tilde{k}_{s,j}(y_j) dy_j \\ &= - \int_{-\infty}^{\infty} (e^{iuy_j} - 1) e^{b_{s,T_i} B_j y_j} k_{s,j}(y_j) dy_j \\ &= - \int_{-\infty}^{\infty} (e^{(iu+b_{s,T_i} B_j)y_j} - e^{b_{s,T_i} B_j y_j}) k_{s,j}(y_j) dy_j \\ &= - \int_{-\infty}^{\infty} (e^{(iu+b_{s,T_i} B_j)y_j} - 1) k_{s,j}(y_j) dy_j + \int_{-\infty}^{\infty} (e^{b_{s,T_i} B_j y_j} - 1) k_{s,j}(y_j) dy_j \\ &= \psi_j^Q(u - ib_{s,T_i} B_j) - \psi_j^Q(-ib_{s,T_i} B_j), \end{aligned}$$

where the $\psi_j^{Q^0}$ denotes the characteristic exponent of the j -th jump component under the original risk-neutral measure Q^0 .

Now we look at the dynamics of the underlying forward rate $L_i(s) = F(s; s, T_i)$

which is the simple compounded spot rate prevailing at time s and T_i defined as

$$\begin{aligned} L_i(s) &= \frac{1}{\alpha_i} \left(\frac{1}{P(s, T_i)} - 1 \right) \\ &= \frac{1 - P(s, T_i)}{\alpha_i P(s, T_i)}, \end{aligned}$$

where $\alpha_i = T_i - s$. It is easy to see that under the forward measure \tilde{Q}^{T_i} , $L_i(s)$ is a martingale. Analogous to the swaption case, we define

$$\frac{dL_i(s)}{L_i(s)} = \left(Y_{i,s}^f(y_j) - 1 \right)$$

and we try to determine $Y_{i,s}^f(y)$ on the analysis of

$$\begin{aligned} \frac{dL_i(s)}{L_i(s)} &= -\frac{P(s, T_i)}{1 - P(s, T_i)} (e^{\sum_j b_{s, T_i} B_j y_j} - 1) - (e^{\sum_j b_{s, T_i} B_j y_j} - 1) \\ &= -\frac{1}{1 - P(s, T_i)} (e^{\sum_j b_{s, T_i} B_j y_j} - 1). \end{aligned}$$

Hence, we approximate

$$Y_{i,s}^f(y) = \exp \left(\sum_{j=1}^3 -\frac{b_{s, T_i} B_j y_j}{1 - P(0, T_i)} \right).$$

We then obtain the characteristic exponent of the logarithm of $L_i(T_{i-1}) = F(T_{i-1}, T_i)$

since the cap contract payoff is known at time T_{i-1} .

$$E \left[e^{iu \ln L_i(T_{i-1})} \right] = \exp(-\psi_{i-1}^i(u)),$$

and we have

$$\begin{aligned} \psi_{i-1}^i(u) &= -iu \sum_{j=1}^3 \int_0^{T_{i-1}} \tilde{\psi}_{s,j}^{T_i}(-i\xi_{s,j}) ds + \sum_{j=1}^3 \int_0^{T_{i-1}} \tilde{\psi}_{s,j}^{T_i}(u\xi_{s,j}) ds, \\ \xi_{s,j} &= -\frac{b_{s, T_i} B_j y_j}{1 - P(0, T_i)}. \end{aligned}$$

Once we have obtained the characteristic function for the logarithm of the forward rate $L_i(s)$ at time T_{i-1} , the rest calculation is similar to the swaption case, i.e. the FFT method is applied to calculate the value of a single caplet which pays off at each time T_i . The price of a cap is thus the summation of all caplets, and one can use the parity equation to deduce the value of a floor with the same structure as the cap.

Chapter 4

Particle Filter and Parameter Estimation

4.1 Background

The mission of assessing the empirical validity of interest models in general, and of our Lévy based model in specific, is of obvious importance. Before engaging in any performance comparison with other models, we should have a clear picture of how our model parameters influence the observable market quantities, such as the bond prices. That is, the model should first be sensibly estimated.

The problem has two aspects. From the theoretical point of view, we have developed a state-space model containing time-invariant parameters and a vector of latent state variables driven by a multi-factor Lévy process. The short rate is an affine combination of the latent factors. Bond prices and their derivative values are non-linear functions of the current state variables, given the model parameters known. From the empirical perspective, time series of market prices for the past 10 years are available. However, market prices are noisy data mixed with trading errors, bid/ask spread, market mis-specification, etc. Here, the task is to design an algorithm, which accounts for both the model structure and the disturbed data supply, to acquire a sound estimation of the model parameters.

Parameter estimation, also called “calibration” in financial engineering lingo, is an optimization procedure that minimizes the difference between the model outputs

and market inputs. The challenge in estimating the state-space model is that the model outputs depend on the knowledge of state variables which are not directly observable. Therefore, the critical step in estimating such models is to retrieve the values of latent state variables by filtering out the aforementioned noises using the filter technique. Once the state variables are determined, the optimization is carried out by minimizing the pricing errors. As a special case, if the errors are assumed to be white noises following multidimensional normal distribution, the maximum likelihood estimation could be invoked.

In specific, the model estimation under the state-space model with time series data consists of two steps.

- (1) Take the initial guess on the model parameters (or the returned parameter values from last optimization iteration); estimate sequentially the state variables through time by filtering; generate model prices based on the filtered state variables.
- (2) Maximize the joint likelihood function for the errors between the model prices and market prices

Step (1) is the focus of this chapter; that is, we will focus on designing algorithms that could sequentially retrieve the unobservable state variables from a large set of noisy signals. Kalman filter (KF) is the standard tool in this area, and under the assumption that the state variables are driven by Brownian motion and the measurement/propagation functions are linear, KF has been proved very capable of providing efficient, robust estimation. We have seen massive applica-

tions of this techniques in the area of financial engineering for model calibration or price predictions, etc. For instance, one investment strategy that hedge funds and banks' proprietary trading desks often use is to build an equilibrium model, cast it into state-space form, estimate it using the KF technique, make projections on price movements, and trade the difference between the model prediction and the market quotes. If the model indeed describes the correct market dynamics, market prices should converge (statistically) to the predicted price and traders pocket the difference. Performance of such strategies crucially depend on the quality of the implementation of the filter technique used for model estimation and prediction.

The KF-like filter would perform well if we live in a perfect linear and Gaussian world. Unfortunately, the real world is not even close to be perfect. As pointed out in the Chapter 3, non-Gaussianity exists significantly in the fixed-income world, and pricing functions are highly nonlinear. Whereas the extended Kalman filter (EKF) and unscented Kalman filter (UKF), as extensions to the original KF, were devised to cope with the non-linearity, the non-Gaussianity could not be rescued anyway in the restrictive Kalman filter world.

Therefore we resort to the particle filter (PF), a newly emerged filtering technique that has not been vastly applied due to its high demand in computing power. However, as this demand is being eased by the dramatic reduction in computing cost recently, particle filter has become an attractive alternative to handle the non-Gaussian and non-linear problems. Unlike KF where the distribution is completely determined by the variance-covariance specification, PF relies on a large number of simulated particles to represent the distribution. The distribution is moved for-

ward by propagating the entire set of particles, while under KF only the variance-covariance matrix needs to be updated. The propagation of the particles are realized by Monte Carlo simulation, and this is why PF is also referred by some literature as sequential Monte Carlo method.

In this chapter, Kalman filter and its extensions will be firstly discussed, given its prominent status in the area of state-space models estimation. Next we will focus on explaining the particle filter technique, and generic algorithms including the resampling technique will be studied. The generic particle filter algorithm is then customized to our model context and numerical results are presented. After the model is estimated, we perform factor analysis on the yield curve and experiment caplet pricing using Monte Carlo simulation.

4.2 The problem and conceptual solution

Let's introduce the filtering problem. Define $x_k \in \mathbb{R}^{n_x}$ the state vector where n_x is the dimension of the state variable and k is the time index. The evolution of x_k follows the below discrete-time stochastic model function

$$x_k = f_{k-1}(x_{k-1}, v_{k-1}) \quad (4.1)$$

where v_{k-1} is the randomness from the state variable evolution, and f_{k-1} is the propagation function perturbed by the randomness v_{k-1} , which is the unforeseeable disturbance in the state motion. The measurements are functions of state variables x_k and the measurement noises w_k

$$z_k = h_k(x_k, w_k), \quad (4.2)$$

where $z_k \in \mathbb{R}^{n_z}$, and h_k ¹ is a known, possibly non-linear function. The error w_k is often assumed as the white noise.

The challenge is to extract the values of unobservable x_k for all k based on the observations of $Z_k \triangleq \{z_i, i = 1, \dots, k\}$ which become readily available over time. Because x_k themselves are random variables, the problem essentially is to discover the posterior probability density function (PDF) $p(x_k|Z_k)$, conditioned on the observation Z_k . Assuming the initial $p(x_0) \triangleq p(x_0|Z_0)$ is known, we invoke Bayes theorem to recursively update the posterior distribution of x_k at each k . There are two steps in this task: prediction and update.

1. First we make the prediction about the PDF $p(x_k|Z_{k-1})$ by applying the Chapman-Kolmogorov equation

$$\begin{aligned} p(x_k|Z_{k-1}) &= \int p(x_k|x_{k-1}, Z_{k-1})p(x_{k-1}|Z_{k-1})dx_{k-1} \\ &= \int p(x_k|x_{k-1})p(x_{k-1}|Z_{k-1})dx_{k-1}. \end{aligned} \quad (4.3)$$

Note: from the first line to second line in Equation. (4.3) we used the Markov property, namely the information of Z_{k-1} has been included in x_{k-1} .

2. Second, after the observation z_k becomes available, Bayes rule is used to update the posterior

$$p(x_k|Z_k) = \frac{p(z_k|x_k)p(x_k|Z_{k-1})}{p(z_k|Z_{k-1})}, \quad (4.4)$$

where the normalizing factor in the denominator could be written as

$$p(z_k|Z_{k-1}) = \int p(z_k|x_k)p(x_k|Z_{k-1})dx_k.$$

¹It is noteworthy that for convenience in many cases both evolution function f_k and measurement function h_k are time-homogeneous such that the time index k could be dropped.

The conceptual solution as in Eqn. (4.4) unfortunately can not be determined optimally in general. It is because the storage of the entire PDF is equivalent to an infinite dimensional vector; therefore, one has to resort to sub-optimal algorithms to approximate the solution to the problem. However, this difficulty is considerably alleviated for models assuming linearity and Gaussian dynamics. In the next section we demonstrate how the Kalman filter leads to an optimal solution under the ideal assumptions.

4.3 Kalman filter

The Kalman filter assumes that the posterior PDF at every time step is Gaussian which can be completely characterized by the mean vector and covariance matrix. In other words, if $p(x_{k-1}|Z_{k-1})$ is normally distributed, it can be proved that $p(x_k|Z_k)$ is also normal if the following conditions about Eqn.(4.1) and (4.2) are satisfied —

- v_{k-1} and w_k are samples from normal distribution
- $x_k = F_{k-1}x_{k-1} + v_{k-1}$, with F_{k-1} being a $n_x \times n_x$ matrix such that function $f_{k-1}(x_{k-1}, v_{k-1})$ is linear in x_{k-1} and v_{k-1}
- $z_k = H_k x_k + w_k$, with H_k being a $n_z \times n_x$ such that function $h_k(x_k, w_k)$ is linear in x_k and w_k

We further denote Q_{k-1} and R_k the covariance matrix for the white noise v_{k-1} and w_k respectively, which are mutually independent.

Then the Kalman filter algorithm can be viewed in the following recursive manner —

$$p(x_{k-1}|Z_{k-1}) = \mathcal{N}(x_{k-1}; \hat{x}_{k-1|k-1}, P_{k-1|k-1}) \quad (4.5)$$

$$p(x_k|Z_{k-1}) = \mathcal{N}(x_k; \hat{x}_{k|k-1}, P_{k|k-1}) \quad (4.6)$$

$$p(x_k|Z_k) = \mathcal{N}(x_k; \hat{x}_{k|k}, P_{k|k}) \quad (4.7)$$

where $\mathcal{N}(x; m, P)$ is the density function of the normal distribution with mean m and covariance P , and $\hat{x}_{k|k}$ and $P_{k|k}$ denote the posterior estimate of the mean and covariance, and $\hat{x}_{k|k-1}$ and $P_{k|k-1}$ denote the prior estimate for the mean and covariance without the newest update.

With the notations above, the Kalman filter algorithm is carried out as follows,

$$\hat{x}_{k|k-1} = F_{k-1} \hat{x}_{k-1|k-1} \quad (4.8)$$

$$P_{k|k-1} = Q_{k-1} + F_{k-1} P_{k-1|k-1} F_{k-1}^T \quad (4.9)$$

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k (z_k - H_k \hat{x}_{k|k-1}) \quad (4.10)$$

$$P_{k|k} = P_{k|k-1} - K_k S_k K_k^T \quad (4.11)$$

where

$$S_k = H_k P_{k|k-1} H_k^T + R_k$$

is the covariance of the innovation term, and

$$K_k = P_{k|k-1} H_k^T S_k^{-1}$$

is the so-called Kalman gain.

By looking at the Eqn. (4.11) we find how Kalman filter helps reduce the covariance of the state variables with the reception of new information. The posterior

covariance $P_{k|k}$ equals the prior covariance $P_{k|k-1}$ less $K_k S_k K_k^T$, which is the amount of variance reduction coming from the new observation of z_k through Kalman gain.

From Eqn. (4.8) to (4.11) is the complete recipe of Kalman filter. The Kalman filter is the only optimal solution with the aforementioned assumptions held. However, in cases where the measurement functions are non-linear, sub-optimal solutions such as extended Kalman filter (EKF) and unscented Kalman filter (UKF) (see [20, 27, 28] for details) are devised to cope with the broken assumptions.

1. If the function(s) $f(x, v)$ or $h(x, w)$ (or both) are almost linear, we approximate them by the first-order Taylor expansion and replace those functions with their Jacobian matrices. This is the extended Kalman filter.
2. If non-linearity is high in functions $f(x, v)$ or $h(x, w)$, we use the *unscented transformation* that draws the deterministic *sigma* points to approximate the distribution. This method is the unscented Kalman filter or UKF.

Both EKF and UKF attempt to reconcile the non-linearity violation in the original KF, but neither could theoretically handle the problem if the state variables are non-Gaussian. However, it is crucial in our case to relax the Gaussianity restrictions since the state variables in our model are driven by pure jump Lévy processes. We hereby resort to the counterpart of KF — the particle filter.

4.4 Particle filter

Reality often manifests itself as being very complex: non-Gaussian, non-linear with continuous/discontinuous state space. Therefore, Kalman filter, albeit theoret-

ically sound, is not applicable in many practical situations. Particle filter emerged as an alternative algorithm to substitute Kalman filter in the non-Gaussian case. Instead of updating only the mean and covariance, PF uses a large number of points/particles to empirically approximate the posterior distribution in interest.

Admittedly in comparison to the optimal Kalman filter, particle filters are sub-optimal. Obviously, a finite number of simulated particles could not completely recover a continuous distribution, though more particles will help increase the precision. However, the number of particles could not go too large in the practice of state-space model estimation, since one needs to perform the simulation at every time step. This is a heavy load of computation which has prohibited from applying PF in the past when computing cost was high. Given enough computing power nowadays however, particle filter technique can be very helpful and offer great flexibility in estimating state-space models without restricted assumptions. Our 3-factor Lévy fixed income model is a perfect application in this regard.

In this section we will discuss in detail about particle filter in its generic format. Importance sampling again will be reviewed in the context of PF implementation. Sequential importance sampling (SIS), as the simplest PF implementation will be investigated. As naive SIS introduces the degeneracy problem, resampling technique is discussed to reduce the degree of degeneracy.

4.4.1 Importance sampling in particle filter

The mechanism of importance sampling has been demonstrated in Eqn.(2.13) to (2.16), and the idea behind it is to make the random draws from a more concentrated region that most concerns the calculation in interest. In the particle filter context, since we are interested in finding the posterior distribution $p_X(\cdot)$, we want to generate particles from a guessed importance distribution $q_X(\cdot)$ which preferably are close to the true posterior. In fact, with a more educated guess about $q_X(\cdot)$, the calculation on $p_X(\cdot)$ using Bayes theorem would be more accurate. It can be shown that, to some extent, the quality of the parameter estimation depends on the choice of the guessed importance distribution. The selection of the importance distribution will be addressed later on in this chapter.

Recall that in order to apply importance sampling, the Radon-Nikodym derivative $\tilde{\omega}(\cdot)$ needs to be calculated for each particle x_i as follows

$$\tilde{\omega}(x_i) = \frac{p_X(x_i)}{q_X(x_i)}. \quad (4.12)$$

Since the information about $p_X(\cdot)$ is not known beforehand, we can not rely on Eqn. (4.12) to compute the weights $\tilde{\omega}(x_i)$. To bypass this difficulty, the normalization is introduced so the new weights for each drawing i can be obtained by

$$\omega(x_i) = \frac{\tilde{\omega}(x_i)}{\sum \tilde{\omega}(x_i)}. \quad (4.13)$$

The normalization procedure will be discussed in details in the next section.

4.4.2 Sequential importance sampling

Now we discuss how to apply importance sampling with particle filter based on the conceptual solution as in Eqn. (4.4). We denote $X_k = \{x_j : j = 1, \dots, k\}$ the sequence of state variable x and $Z_k = \{z_j : j = 1, \dots, k\}$ the sequence of observations of z both up to time k . The the joint posterior distribution at time k is estimated as follows

$$p(X_k|Z_k) = \sum_{i=1}^N \omega_k^i \delta(X_k - X_k^i), \quad (4.14)$$

where $\delta(\cdot)$ is the *delta* function. The calculation of weights ω_k^i follows Eqn.(4.13). In particular, if samples are drawn from an importance density $q(X_k|Z_k)$ other than the true density, then the weights can be expressed as

$$\omega_k^i \propto \frac{p(X_k^i|Z_k^i)}{q(X_k^i|Z_k^i)}. \quad (4.15)$$

Suppose at time $k - 1$ we have samples that constituting an approximation of $p(X_{k-1}|Z_{k-1})$ and we are interested in having $p(X_k|Z_k)$ at time k with the reception of new observation z_k . If the importance distribution is chosen to have the following factorization

$$q(X_k|Z_k) = q(x_k|X_{k-1}, Z_k) q(X_{k-1}|Z_{k-1}), \quad (4.16)$$

one can then augment the existing samples $X_{k-1}^i \sim q(X_{k-1}|Z_{k-1})$ with the new state $x_k^i \sim q(x_k|X_{k-1}, Z_k)$ to obtain samples $X_k^i \sim q(X_k|Z_k)$.

To derive the weight update, the PDF $p(X_k|Z_k)$ can be expanded in the following manner — (For derivation see Appendix (A.3))

$$p(X_k|Z_k) \propto p(z_k|x_k) p(x_k|x_{k-1}) p(X_{k-1}|Z_{k-1}). \quad (4.17)$$

Now if we plug Eqn.(4.16) and (4.17) in the Eqn.(4.15), the weight update equation can be shown as

$$\begin{aligned}\omega_k^i &\propto \frac{p(z_k|x_k^i) p(x_k^i|x_{k-1}^i) p(X_{k-1}^i|Z_{k-1})}{q(x_k^i|X_{k-1}^i, Z_k) q(X_{k-1}^i|Z_{k-1})} \\ &= \omega_{k-1}^i \frac{p(z_k|x_k^i) p(x_k^i|x_{k-1}^i)}{q(x_k^i|X_{k-1}^i, Z_k)}.\end{aligned}\quad (4.18)$$

The relationship in Eqn.(4.18) can be simplified by the Markov property to drop the dependence on the history

$$q(x_k^i|X_{k-1}^i, Z_k) = q(x_k^i|x_{k-1}^i, z_k), \quad (4.19)$$

such that the importance density only depends on the most recent state and observation x_{k-1} and z_k . This reduction is particularly helpful when only the filtered posterior $p(x_k|Z_k)$ is estimated sequentially but large storage of past information is undesired. After the simplification, the weights update equation reads

$$\omega_k^i \propto \omega_{k-1}^i \frac{p(z_k|x_k^i) p(x_k^i|x_{k-1}^i)}{q(x_k^i|x_{k-1}^i, z_k)}.\quad (4.20)$$

And the updated posterior density is approximated as

$$p(x_k|Z_k) \approx \sum_{i=1}^N \omega_k^i \delta(x_k - x_k^i).\quad (4.21)$$

The Sequential Importance Sampling (SIS) algorithm gives rise to propagate the posterior distribution with new updates at each time step by Monte Carlo simulating a large number of particles. With the knowledge of the entire posterior distribution obtained sequentially we could be able perform any relevant calculations involving the distribution of the state variables, such as the likelihood function.

4.4.3 Selection of the importance density

The quality of the particle filter estimation crucially depends on the selection of the prior distribution $q(x_k^i|x_{k-1}^i, z_k)$ as in Eqn.(4.20). The optimal prior choice would minimize the variance of the importance weights. However, except in a few exceptions, finding the optimal choice is infeasible. Here we present the most widely-used suboptimal choice that is the so-called transitional prior,

$$q(x_k^i|x_{k-1}^i, z_k) = p(x_k^i|x_{k-1}^i). \quad (4.22)$$

This choice is very popular due to its simplicity. With this prior, the weight update function Eqn.(4.20) is conveniently reduced to

$$\omega_k^i \propto \omega_{k-1}^i p(z_k|x_k^i). \quad (4.23)$$

The SIS algorithm under this choice of prior distribution is summarized in Table (4.1).

The transitional prior would be found if the domain for most of prior's distribution mass is narrower than that of the likelihood function $p(z_k|x_k)$, otherwise problems such as fast degeneracy and impoverishment will arise. Improved algorithms to handle those problems have been developed in the past, and one of them, the resampling technique, will be discussed in the next section. Interested readers are suggested to read Chapter 3 in [27] for a more detailed account.

4.4.4 Resampling

Any suboptimal choice of prior distribution, such as the transitional prior discussed in the previous section, causes the degeneracy problem. In specific, it is

$[\{x_k^i, \omega_k^i\}_{i=1}^N] = \text{SIS}([\{x_{k-1}^i, \omega_{k-1}^i\}_{i=1}^N, z_k])$ <ul style="list-style-type: none"> • For $i = 1 : N$ <ul style="list-style-type: none"> – Draw x_k^i from the prior distribution $q(x_k^i x_{k-1}^i, z_k) = p(x_k^i x_{k-1}^i)$; – Evaluate the non-normalized weights according to Eqn.(4.23) $\tilde{\omega}_k^i = \omega_{k-1}^i p(z_k x_k^i)$ • End For • Calculate the normalizing factor $Z = \sum_{j=1}^N \tilde{\omega}_k^j$ • For $i = 1 : N$ <ul style="list-style-type: none"> – Normalize: $w_k^i = \frac{\tilde{\omega}_k^i}{Z}$ • End For
--

Table 4.1: Sequential importance sampling algorithm

found that the suboptimal prior could increase the variance in the importance weight over time. As a consequence, it is very likely that, after a few time steps, one particle picks up all the weights while other particles weigh little. The degeneracy problem leads to a large variance in the posterior distribution and harms the estimation. Unfortunately, it is theoretically inevitable when applying the suboptimal sequential importance sampling algorithm.

Whereas it is impossible to totally avoid the degeneracy, we could seek to reduce it. The degree of degeneracy could be measure by the *effective sample size* \hat{N}_{eff} is defined as

$$\hat{N}_{eff} = \frac{1}{\sum (\omega_k^i)^2}.$$

The smaller \hat{N}_{eff} is, the stronger degeneracy exists. \hat{N}_{eff} is then monitored over time; once a pre-defined threshold N_{thr} is breached, resampling is executed to lessen

the degeneracy degree. The resampling procedure is to split particles that carry large weights into small particles based on the probability, so technically it is to map a random measure $\{x_k^i, \omega_k^i\}$ into another random measure $\{x_k^{i*}, \frac{1}{N}\}$ of equal weights. The new random measure is generated by sampling with replacement N times from the approximated discrete posterior $p(x_k|Z_k)$ given by Eqn.(4.21) such that $P\{x_k^{i*} = x_k^j\}$ equals ω_k^j . Therefore, after resampling, a new set of particles with uniform weights is drawn and available for the next step of propagation.

The mechanism of resampling procedure is illustrated in Fig.(4.1), and the algorithm is summarized in Table 4.2.

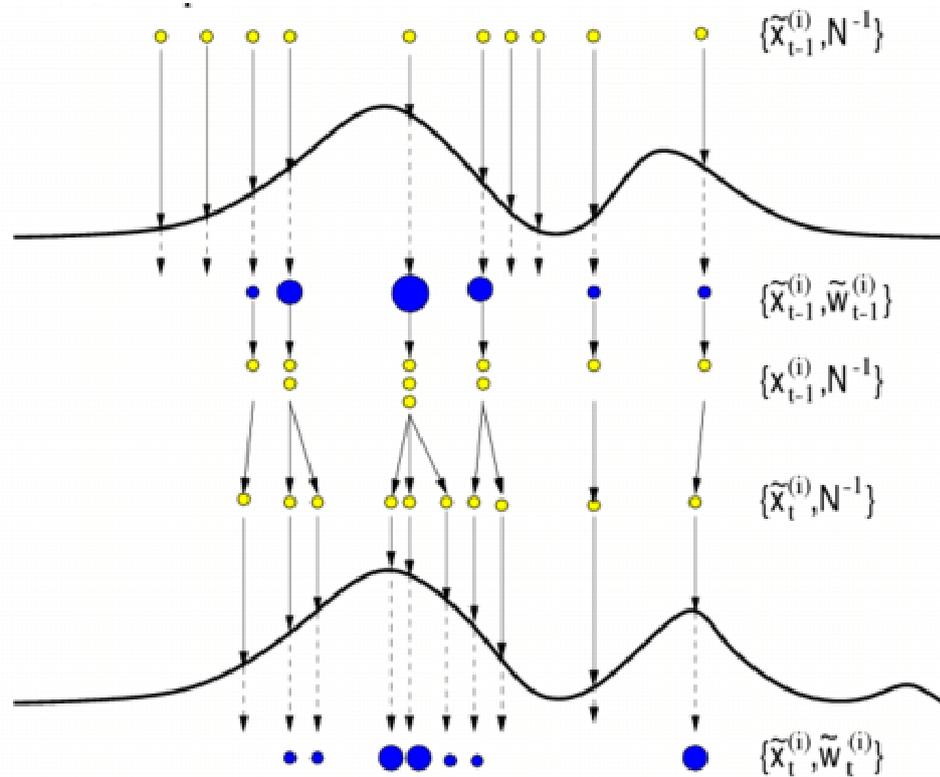


Figure 4.1: Resampling scenario and mechanism in particle filter

$[\{x_k^{i*}, \omega_k^i\}_{i=1}^N] = \text{RESAMPLE}([\{x_k^i, \omega_k^i\}_{i=1}^N, z_k])$ <ul style="list-style-type: none"> • Initialize $c_1 = \omega_k^1$ • For $i = 2 : N$ $c_i = c_{i-1} + \omega_k^i$ • End For • Draw random number $n_1 \sim \mathcal{U}[0, 1/N]$, set $i = 1$ • For $j = 1 : N$ <ul style="list-style-type: none"> - Let $u_j = u_1 + \frac{j-1}{N}$ - While $u_j > c_i$ $i = i + 1$ - Assign $x_k^{i*} = x_k^i$ and $\omega_k^j = \frac{1}{N}$ • End For

Table 4.2: Resampling algorithm

4.4.5 Generic particle filter algorithm

We conclude this section by summarizing the algorithm of the general particle filter which combines the sequential importance sampling (SIS) with resampling in Table 4.3. The algorithm is to generate the N new pairs of $\{x_k^i, \omega_k^i\}$ for $i \in \{1, \dots, N\}$ at the time step k with input of $\{x_{k-1}^i, \omega_{k-1}^i = \frac{1}{N}\}$.

$[\{x_k^i, \omega_k^i\}_{i=1}^N] = \text{PF}([\{x_{k-1}^i, \omega_{k-1}^i\}_{i=1}^N, z_k])$ <ul style="list-style-type: none"> • Filtering via SIS in Table 4.1 • Resample using the algorithm in Table (4.2) $[\{x_k^i, \omega_k^i = \frac{1}{N}\}_{i=1}^N] = \text{RESAMPLE}([\{x_k^i, \omega_k^i\}_{i=1}^N, z_k])$

Table 4.3: Particle filter algorithm

4.5 Estimate the 3-factor Lévy short rate model

In this section we will first describe the data set we use in the estimation and then cater the generic particle filter algorithm to our model estimation context. Numerical results will be presented and analyzed.

4.5.1 Data description

By courtesy of Caspian Capital Management LLP, we are provided with the over-the-counter Euro-dollar future rates and constant maturity swap (CMS) rates across different maturities and tenors.

In specific, the first 3 columns of data are rates reflected from the Eurodollar future contracts, which are written on the 3-month Libor rates for 3 different maturities of $T = 3, 6, 12$ months. The next 6 columns of data are the swap rates of 6 different maturities including $T_n = 2, 3, 5, 10, 15, 30$ years. In those swap contracts, the fixed-rate coupon payments are made semi-annually.

Therefore we have in total 9 columns of data. The data set contains 10 year daily market quotes that start from 04/25/1994 through 10/13/2004. To avoid weekday effects in the estimation, we only sample data weekly on every Wednesday. This makes the sampled data set have $9 \times 538 = 4842$ entries.

4.5.2 Quasi-maximum likelihood estimation with particle filter

Quasi-maximum likelihood estimation (MLE)² is applied to calibrate the model. The continuous-time dynamics of the short rate factors is first cast into a discrete version as the data supply is sampled weekly. Particle filter then helps to retrieve the values of state variables sequentially on every sampled time step. Based on the filtered state variables, pricing equations output the model prices, and pricing errors are simply the differences between market prices and the model prices. Errors are assumed to be white noises with pre-estimated variances, and the error likelihood functions are computed via the multidimensional normal distribution function. The concrete procedures will be given below.

We cast the SDE of the state variables $x(t)$ in Eqn. (3.8) and (3.9) by Euler approximation into discrete propagation equations here as

$$x_t = (I - e^{-\kappa\Delta t})\frac{a}{\kappa} + e^{-\kappa\Delta t}x_{t-1} + B\Delta L_t, \quad (4.24)$$

where I is the 3×3 identity matrix and ΔL_t is our Lévy randomness. With weekly sample frequency, we set $\Delta t = 7/365$.

We construct measurement functions for Eurodollar futures and swap rates of different kinds, assuming additive, normally distributed measurement errors:

$$z_t = O(x_t; \Theta) + e_t, \quad (4.25)$$

where z_t denotes the observable prices at time t , and $O(x_t; \Theta)$ denotes the model-implied values as a function of the parameter set Θ and the factor vector x_t . The

²It is called quasi-maximum likelihood method because the likelihood function is, strictly speaking, the conditional likelihood function.

term e_t denotes the normally distributed pricing errors with 0 mean and \mathcal{R} the co-variance matrix ³.

Recall the pricing formulas $O(x_t; \Theta)$ for Eurodollar futures and swap rates which are summarized below:

$$L_t(T, T + \alpha) = \frac{P(t, T) - P(t, T + \alpha)}{\alpha P(t, T + \alpha)}, \quad (4.26)$$

$$K_t(T_0, \dots, T_n) = \frac{P(t, T_0) - P(t, T_n)}{\sum_{i=1}^n \alpha P(t, T_i)}, \quad (4.27)$$

where the bond formulas, given the state variables, are shown in Eqn.(3.13). The $L_t(T, T + \alpha)$ in Eqn.(4.26) denotes the Eurodollar future rate prevailing at time t for the period $[T, T + \alpha]$ where $\alpha = 3$ months. And the $K_t(T_0, \dots, T_n)$ in Eqn.(4.27) stands for the rate of the swap contract with coupons being paid on T_1, \dots, T_n . In such a contract, the fixed rate leg makes the payment every $\alpha = 0.5$ year. To match the data, the pricing function $O(x_t; \Theta)$ is 9-dimensional, with the first 3 being Eurodollar futures maturing in $T = 3, 6, 12$ months and swap contracts maturing in $T_n = 2, 3, 5, 10, 20, 30$ years. And we define the pricing errors as the differences between \tilde{z}_t , the *rates* output by the model and z_t , the *rates* observed from the market.

The particle filter technique we developed in Section 4.4 can be straightforwardly customized in the context of our model estimation. In specific, at a particular time $t - 1$ one has a large number of particles with known values x_{t-1} that represent the posterior distribution at time $t - 1$. To propagate the poste-

³For the sake of convenience, in the estimation we take the covariance matrix \mathcal{R} to be the identity matrix I times a constant scalar σ^2

rior distribution to the next time step t , one first makes the ex ante predictions \tilde{x}_t according to Eqn. (4.24). Now at time t market quotes z_t become available and we now use the observable data to update the prediction. As shown in Table 4.1, this is equivalent to find the weight function ω_t^i for each particle i . Recall that the pricing error for each day t and each asset value i is modeled as white noise with variance \mathcal{R} so the log-likelihood function $l_t^i = \log p(z_t|\tilde{x}_t^i)$ is the exponent of the normal distribution

$$l_t^i = -\frac{1}{2} ((z_t - \tilde{z}_t^i)'(\mathcal{R})^{-1}(z_t - \tilde{z}_t^i)) . \quad (4.28)$$

To estimate the parameters, we build the joint log-likelihood function across the entire history for every sampled data. That is,

$$L(\Theta, z_{t=1}^M) = \sum_{t=1}^M \sum_{i=1}^N l_t^i, \quad (4.29)$$

where M is the number of days when we sample data from and N is the number of particles we simulate⁴. We choose parameters to maximize the log-likelihood of the data series, which is expressed as

$$\Theta = \mathit{arg} \max_{\Theta} L(\Theta, z_{t=1}^M) .$$

4.5.3 Estimation results and discussion

Assuming the Lévy process follows Variance Gamma, we estimate the reduced form of the model with risk-neutral parameterization as listed in the right column of Table 3.2 totalling 25 parameters. The program is written in Matlab and utilizes its optimization routines. The selection of initial values of the parameters are

⁴We take $N = 200$ in the estimation.

critical for the optimization, and an informed selection would substantially expedite the optimization convergence. The initial values are chosen via an trial-and-error procedure.

On one hand, a large amount of particles will make better precision for representing the distribution, and hence reduces the number of iterations used in the optimization. On the other hand however, having more particles substantially increases the workload in simulation. The optimal number of particles is beyond our knowledge, but in the Matlab code 200 particles are employed.

The estimation took considerable CPU time for convergence. It is not guaranteed that, due to the large-scale optimization along with randomness caused by simulation, the likelihood maximization stops at the global maximum. But as visualized by the plots from Figure 4.2 to 4.6, a reasonable convergence has been achieved. The estimated parameters are presented in Table 4.4.⁵

The performance statistics of the estimated model are presented in Table 4.5. Errors are the difference between market prices and model prices and quoted in basis points⁶. The statistics of pricing errors for both ex-ante prediction and ex-post update are presented in panel A and B respectively, and columns titled [Mean, Median, Std, MAD, Max, Min] refer to the [mean, median, standard deviation, mean-average-deviation, maximum, minimum] of the error series. EF stands for Eurodollar futures.

⁵We use L_1, L_2, L_3 to denote the parameter vectors for each factor. Since we are using VG process, all are 3-dimensional vectors.

⁶A basis point is one percent of a percentage.

Apparently, with the new market observation, the ex-post performance is much better than the ex-ante prediction in terms of smaller error standard deviations. Cross-sectionally, the results are indicating good predicting power as the mean and median values for all assets but EU(12m, 3m) are within the range of 10 basis points from zero. EU(12m, 3m) seems to be the least predictable. Also, we notice that the predictions for swap rates of shorter maturities have much smaller error standard deviations and MADs than other swap rates and Eurodollar futures. The whole time series of the model implied values against the market quotes are graphically illustrated in Figure 4.2 to 4.6 for each of the 9 asset classes.

Although the Lévy based models are structurally superior, the estimation procedures are, at the same time, considerably more complicated. It is expected that a better computing facility would enable us to use more particles in order to bring increased accuracy and enhanced performance.

4.6 Yield curve analysis and pricing

Based on the estimated model from the previous section, we perform the factor loading analysis on the yield curve and price the caplet using Monte Carlo simulation. The caplet volatility surface is constructed by converting the caplet values into Black implied volatility.

Model SDE: $r(t)=\alpha + \beta'x(t), dx(t) = (a - \kappa x(t))dt + dL(t)$				
α	κ			β
0.0762	$\begin{bmatrix} 0.0013 \\ 0.0903 \\ -0.3450 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0.6353 \\ -0.3927 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 0.9691 \end{bmatrix}$	$\begin{bmatrix} -0.0666 \\ -0.0568 \\ 0.1241 \end{bmatrix}$
a	L_1	L_2	L_3	x_0
$\begin{bmatrix} 0.0009 \\ 0.0353 \\ 0.4118 \end{bmatrix}$	$\begin{bmatrix} -0.0169 \\ 0.1200 \\ 0.1702 \end{bmatrix}$	$\begin{bmatrix} -0.0996 \\ 0.1028 \\ 0.1983 \end{bmatrix}$	$\begin{bmatrix} -0.0096 \\ 0.0828 \\ 0.1874 \end{bmatrix}$	$\begin{bmatrix} 2.1199 \\ -0.4353 \\ 1.0190 \end{bmatrix}$

Table 4.4: In-sample maximum likelihood parameter estimates of 3-factor Lévy short rate model with particle filter using 10-year data from 04/25/1994 through 10/13/2004

Asset	In-sample error statistics in basis points (bps)					
	Mean	Median	Std	MAD	Max	Min
A. Ex-ante						
EF(3m, 3m)	4.1811	4.1557	18.1331	14.0131	59.8101	-95.3415
EF(6m, 3m)	-7.7202	-6.7680	18.1461	13.7373	55.7387	-123.5331
EF(12m, 3m)	-12.7738	-10.5906	26.2126	21.0457	49.6241	-99.1613
Swap(2 year)	-2.9530	-3.7763	14.7992	11.4461	50.6473	-83.8178
Swap(3 year)	-4.9998	-5.6067	16.9883	13.3460	57.3284	-72.2385
Swap(5 year)	-5.7343	-6.5089	18.9978	14.8604	63.1169	-57.7497
Swap(10 year)	0.9012	-1.0923	20.4852	15.8901	93.8079	-50.6519
Swap(20 year)	7.7000	5.2903	22.2722	17.5681	110.7255	-39.3156
Swap(30 year)	2.6371	0.5487	23.0010	18.2084	106.6838	-47.9710
Average	-2.0846	-2.7053	19.8928	15.5684	71.9425	-74.4200
B. Ex-post						
EF(3m, 3m)	7.2434	5.9534	16.2113	12.5880	58.8404	-34.5549
EF(6m, 3m)	-4.8583	-4.2072	10.5541	8.2076	51.1436	-43.0244
EF(12m, 3m)	-10.3143	-7.5075	18.8791	15.4580	38.4103	-69.2283
Swap(2 year)	-0.2902	-0.0407	5.3054	4.1561	12.7163	-21.5128
Swap(3 year)	-2.6413	-3.0292	8.5221	6.6426	29.6837	-31.1299
Swap(5 year)	-3.8019	-4.6334	11.7396	8.9714	42.5053	-32.8672
Swap(10 year)	2.3866	1.1166	14.3203	10.9077	75.8342	-29.3619
Swap(20 year)	9.0442	7.4268	16.9991	13.2676	92.7829	-25.4996
Swap(30 year)	3.8909	1.6666	18.7174	14.6119	89.2396	-36.5544
Average	0.0732	-0.3616	13.4720	10.5345	54.5729	-35.9704

Table 4.5: In-sample estimation performance statistics of 3-factor Lévy short rate model with particle filter using 10-year data from 04/25/1994 through 10/13/2004

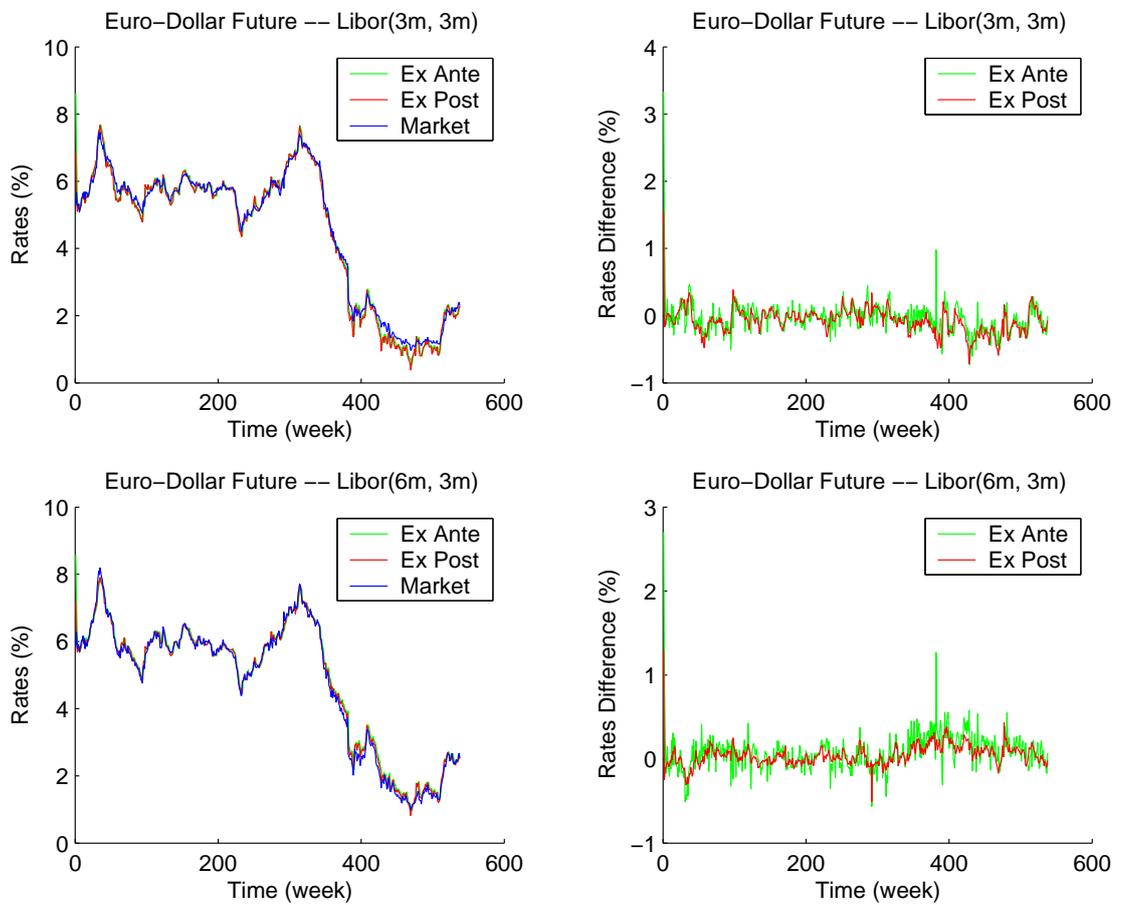


Figure 4.2: Illustration of (quasi) maximum likelihood estimation using particle filter on Eurodollar futures and swap rates against market quotes from 04/25/1994 to 10/13/2004: Upper 2 panels for Eurodollar future on Libor(3m, 3m) and lower 2 panels for Eurodollar future on Libor(6m, 3m)

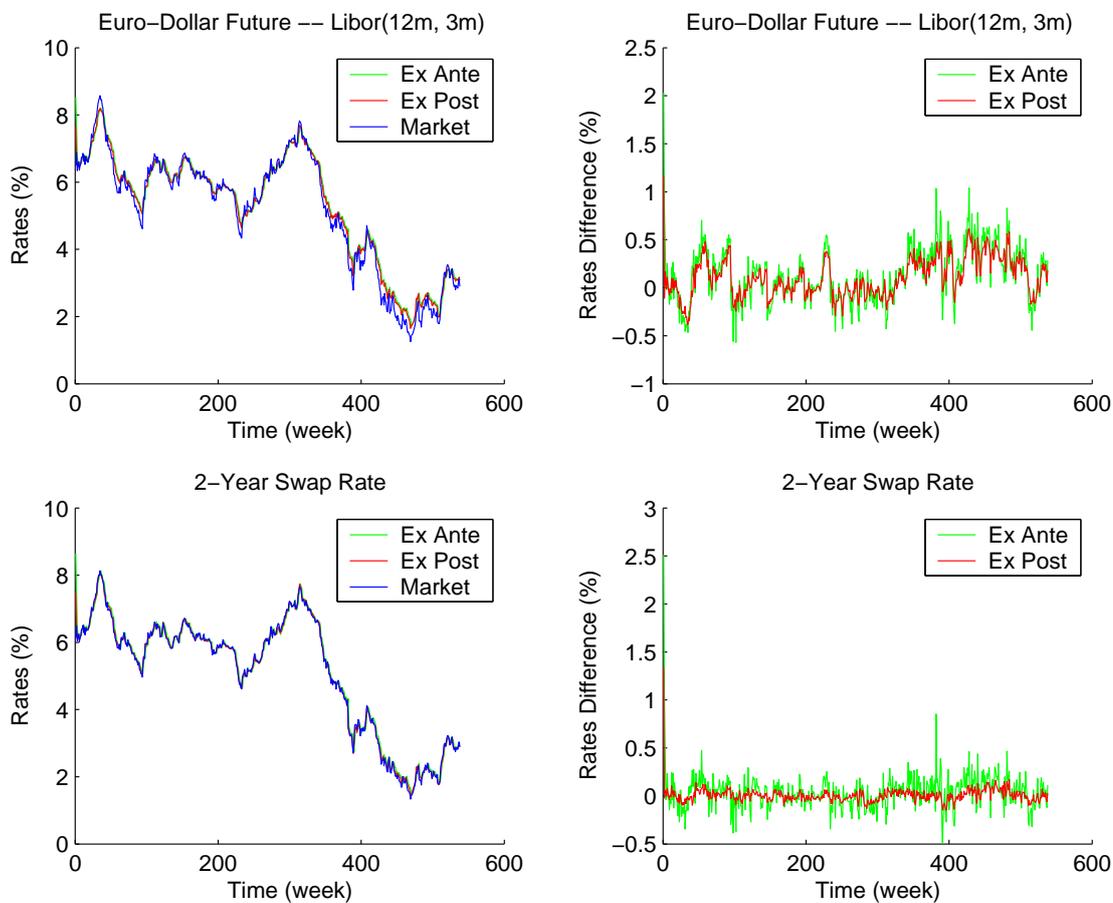


Figure 4.3: Illustration of (quasi) maximum likelihood estimation using particle filter on Eurodollar futures and swap rates against market quotes from 04/25/1994 to 10/13/2004: Upper 2 panels for Eurodollar future on Libor(12m, 3m) and lower 2 panels for 2-year swap rate

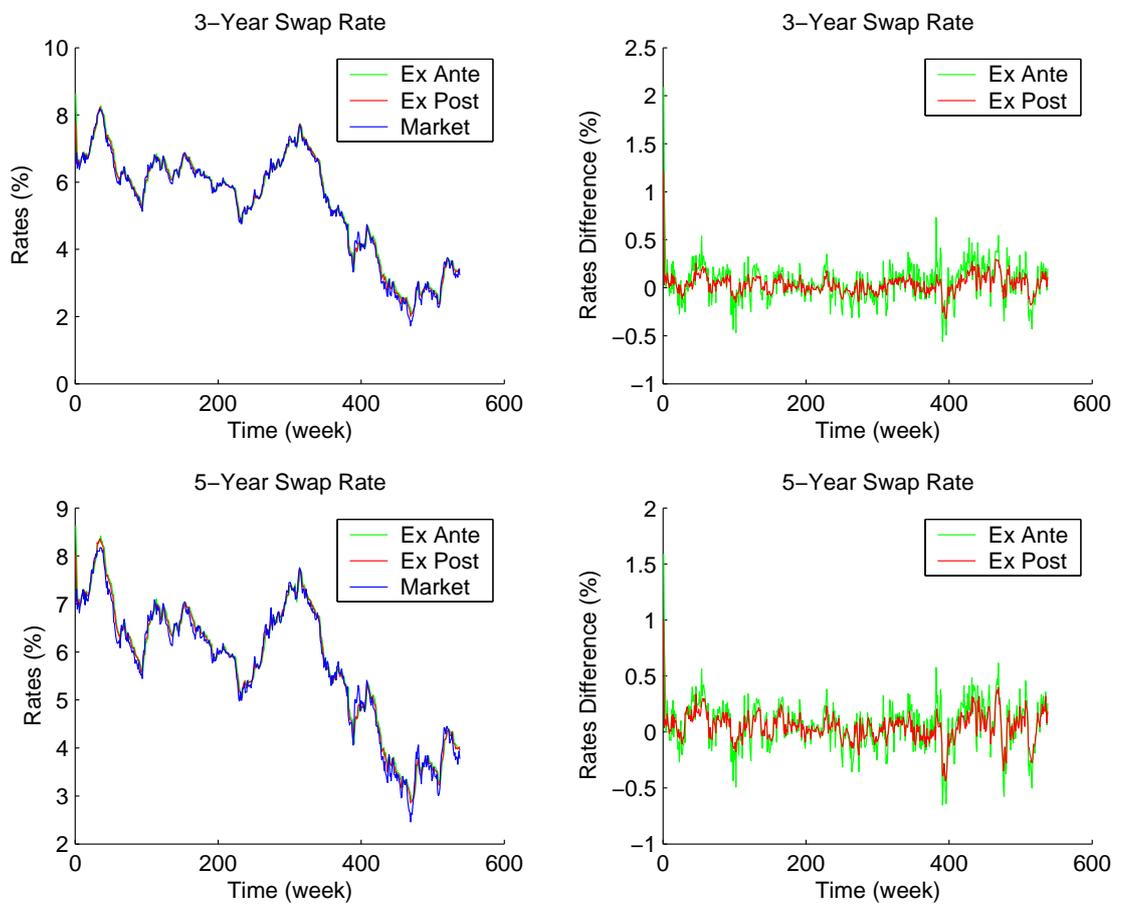


Figure 4.4: Illustration of (quasi) maximum likelihood estimation using particle filter on Eurodollar futures and swap rates against market quotes from 04/25/1994 to 10/13/2004: Upper 2 panels for 3-year swap rate and lower 2 panels for 5-year swap rate

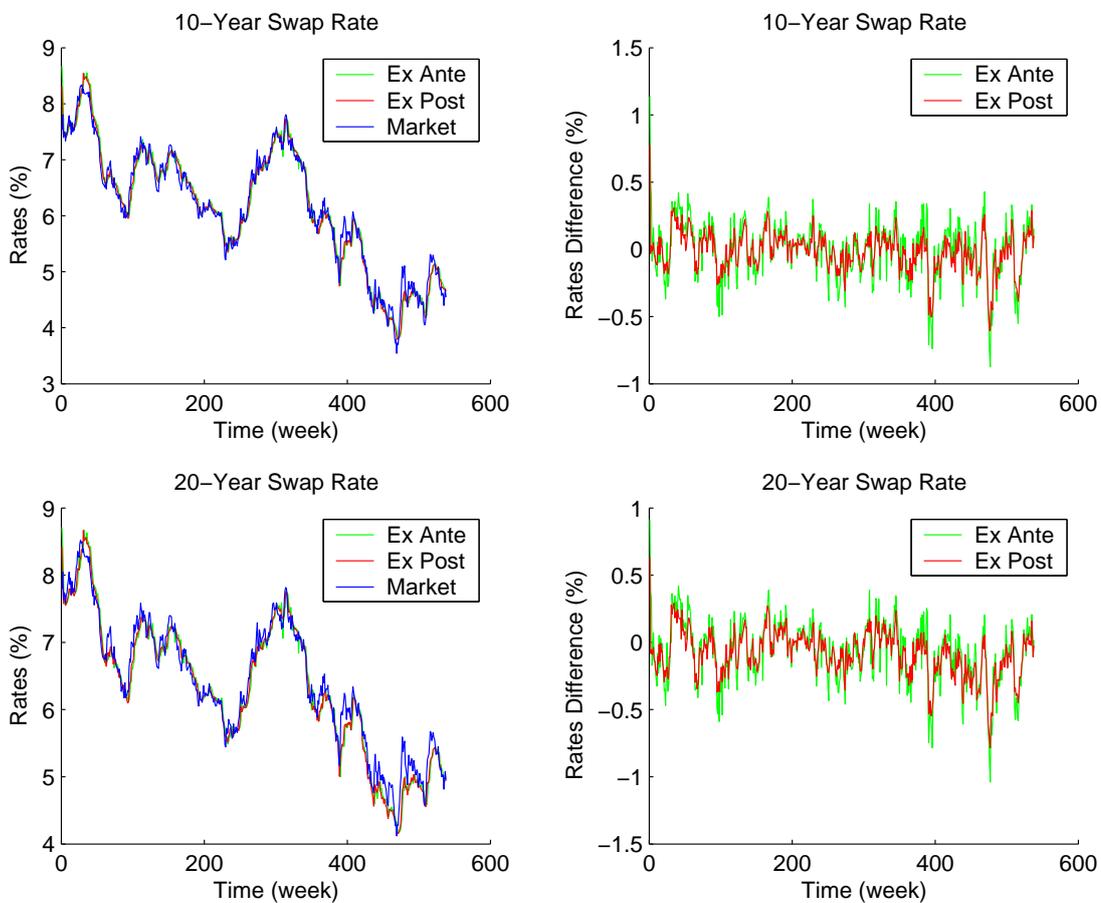


Figure 4.5: Illustration of (quasi) maximum likelihood estimation using particle filter on Eurodollar futures and swap rates against market quotes from 04/25/1994 to 10/13/2004: Upper 2 panels for 10-year swap rate and lower 2 panels for 20-year swap rate

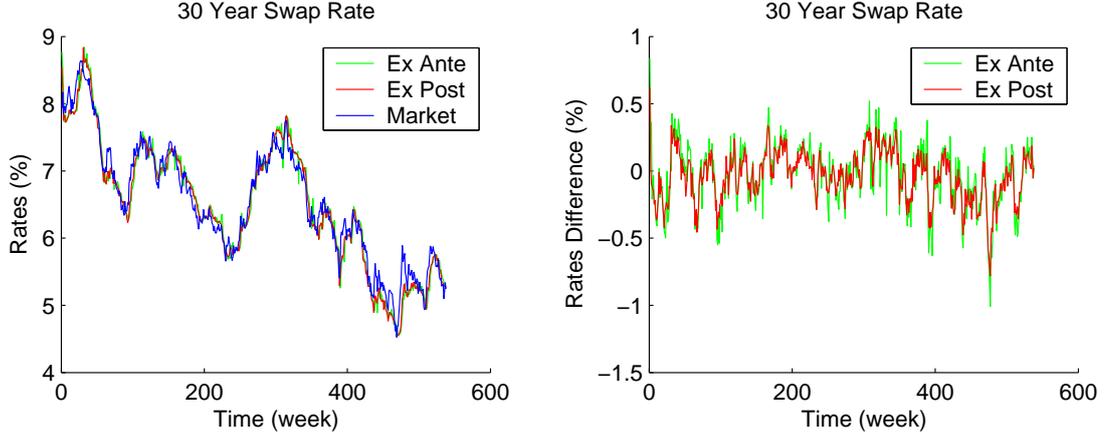


Figure 4.6: Illustration of (quasi) maximum likelihood estimation using particle filter on Eurodollar futures and swap rates against market quotes from 04/25/1994 to 10/13/2004: panels for 30-year swap rate

4.6.1 Yield curve factor loading

Factor loading analysis has been very popular in the field of fixed income research as pioneered by Litterman and Scheinkman in [41]. As a function of the maturity, the loading of a yield curve factor stands for the change in the yield of that maturity given a unit shock in the factor has occurred. The analysis on the factor loading helps to hedge yield curve related positions, as traders can neutralize the duration and convexity of each factor individually according to the loading functions.

We follow the parametric approach in [33]. According to Eqn. (3.13), the coefficient $a_{s,t}$ and $b_{s,t}$ determine the term structure of interest rates. The fair values $Y(s, t)$ of simply compounded spot rates are linked to the yield curve Lévy factors $x(s)$ by

$$Y(s, t) = - \left(\left[\frac{a_{s,t}}{t-s} \right] + \left[\frac{b_{s,t}}{t-s} \right]' x(s) \right) .$$

The slope coefficients $-\frac{b_{s,t}}{t-s}$ are the loading functions of the Lévy factors. We illustrate the factor loading in Figure 4.7.

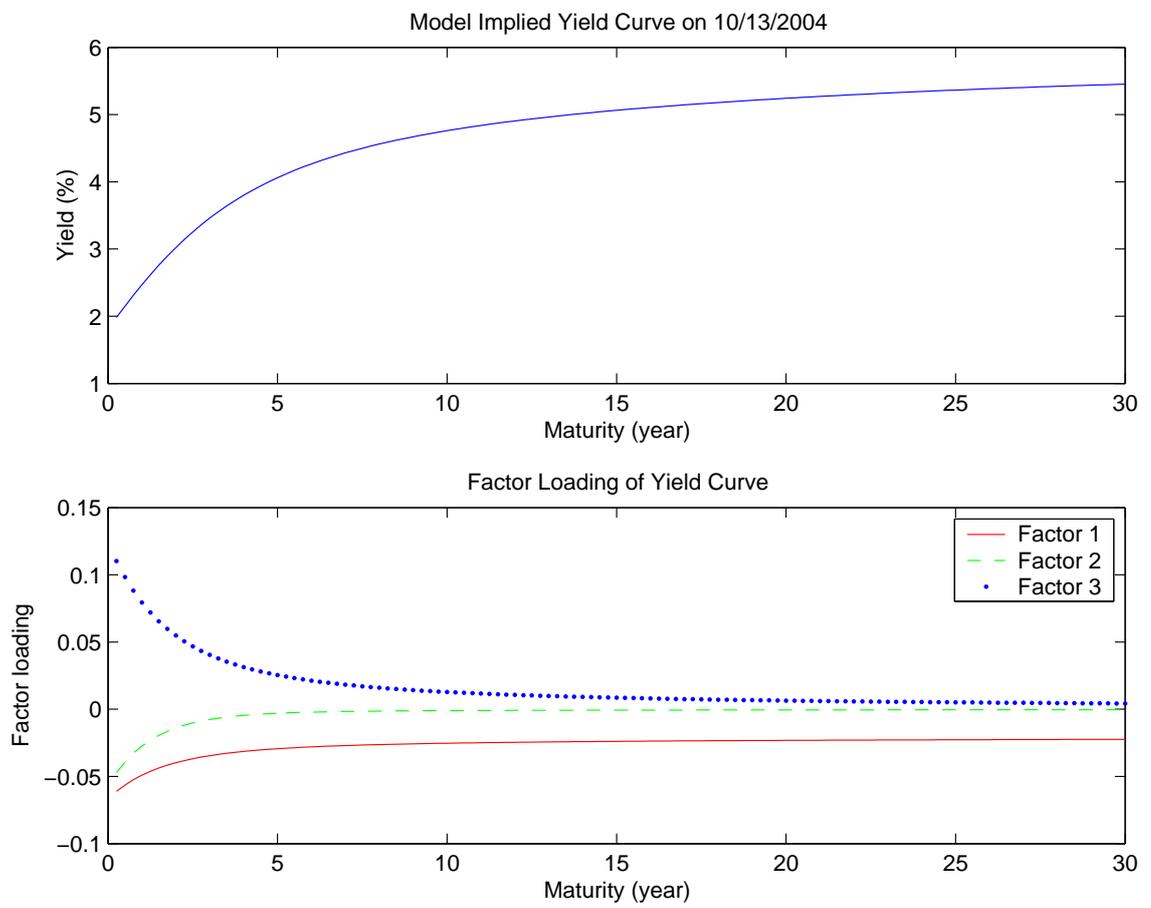


Figure 4.7: Yield Curve and Factor Loading

The upper panel of Figure 4.7 plots the model implied fair value of yield curve on the date of Oct. 13th, 2004 and the lower panel illustrates the factor loading, both as functions of maturity. The upper panel plot is consistent with the upward trending term structure of the market quote on the same day. The factor loading plot shows how the three factors control the variation of the yields at different maturities. The first factor is persistently significant across the spectrum of maturities; the second is least significant as it picks up slowly for short maturity but diminishes rapidly for long maturities; the third factor is the most significant factor for short maturities but dies out quickly after the maturity of 10 years.

4.6.2 Price caplet using simulation

With the estimated model we seek to use Monte Carlo simulation to price one of the mostly traded fixed income derivatives — caplet (see Section 3.5.5 for the payoff structures). Black formula is the market benchmark to price a caplet. Under the Black formula, the annualized caplet rate $Cpl(t; T, T + \Delta t)$ with strike K and volatility σ is

$$Cpl(t; T, T + \Delta t) = P(t, T) (F(t; T, T + \Delta t)\Phi(d^+) - K\Phi(d^-)) , \quad (4.30)$$

where

$$d^\pm = \frac{\log\left(\frac{F(t; T, T + \Delta t)}{K}\right) \pm \frac{1}{2}\sigma^2(T - t)}{\sigma\sqrt{T - t}} , \quad (4.31)$$

and $F(t; T, T + \Delta t)$ is the forward rate prevailing at time t for the future period $[T, T + \Delta t]$.

Obviously, if the price of the caplet is known one can invert by Eqn.(4.30)

and 4.31 to obtain the implied volatility σ . The market has adopted the convention of using σ instead of the rate itself to quote the caplet values, which gives a more intuitive measure about the caplet evaluation.

The Black implied volatilities are reported in the Table 4.6 and illustrated in the 3-D plot of Figure 4.8. As we can see, the implied volatility is increasing with the strike going farther out of the money for all maturities. And looking at the maturity direction, we find that the implied volatility is monotonically decreasing as the maturity gets longer. These results are consistent with some aspects of market observations qualitatively.

However, certain unrealistic characteristics about the model pricing caplet exist. First, the model can only generate a monotonic downward slope in the volatility surface with respect to maturity, while the market has historically shown a hump-shaped volatility shape with a peak between 2 years and 6 years. Second, the volatility is dropping too fast as the maturity goes up, where the 9-year ATM volatility drops below 5% which is way below the market observations of averagely 20%. This diminishing volatility against long maturity is possibly caused by the i.i.d. increment assumptions under the Lévy process.

These observations in fact are in agreement with the industry consensus that equilibrium models, like ours, are not the ideal candidates to price derivatives. It is because

1. The model is not arbitrage free and the underlying bond prices are not matched perfectly.

Black Implied Volatility of Caplet						
Maturity	Moneyness (K/F)					
	1.0	1.02	1.04	1.06	1.08	1.10
1Y	0.2267	0.2420	0.2557	0.2682	0.2796	0.2900
2Y	0.1359	0.1469	0.1566	0.1653	0.1732	0.1804
3Y	0.0992	0.1081	0.1160	0.1229	0.1292	0.1349
4Y	0.0789	0.0866	0.0934	0.0995	0.1049	0.1098
5Y	0.0686	0.0755	0.0816	0.0870	0.0918	0.0961
6Y	0.0607	0.0670	0.0726	0.0774	0.0818	0.0857
7Y	0.0546	0.0605	0.0656	0.0701	0.0742	0.0778
8Y	0.0511	0.0567	0.0614	0.0657	0.0694	0.0728
9Y	0.0492	0.0544	0.0589	0.0629	0.0665	0.0697

Table 4.6: Implied volatility charts of model-generated caplet values of tenor 3-month on 10/13/2004 with maturity from 1 to 9 years and moneyness from at-the-money 1 to out-of-the-money 1.1

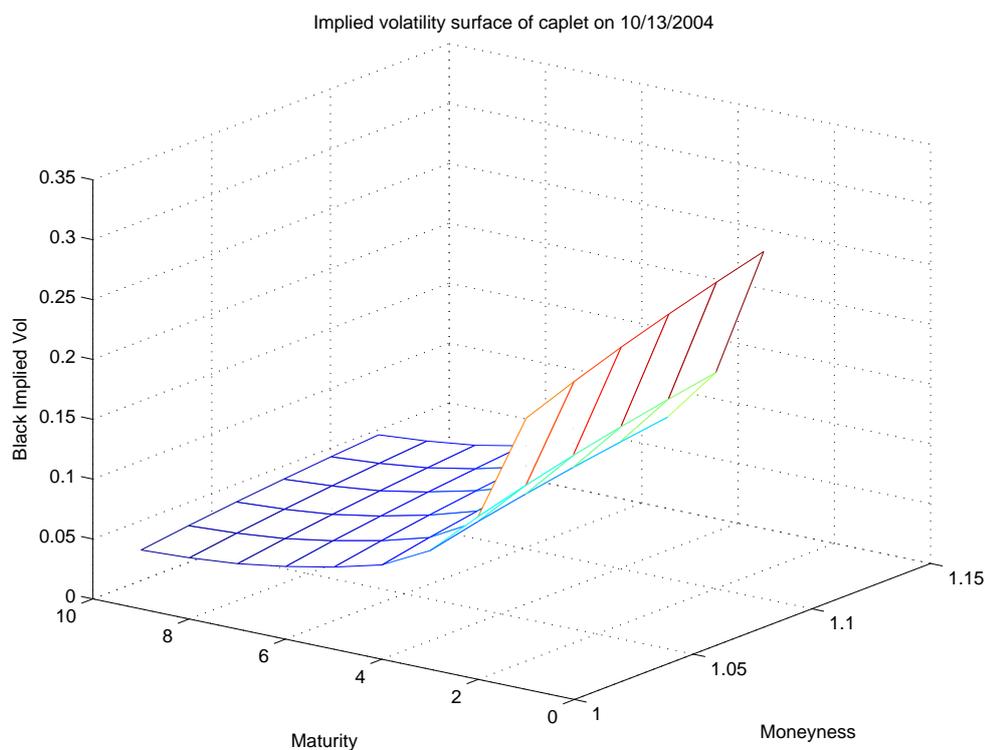


Figure 4.8: Implied volatility surface of model-generated caplet values, 10/13/2004

2. Term structure models can by no means capture certain characteristics that are unique to derivative markets, e.g. caplet market, as those characteristics are conceivably not priced in the yield curve.

There are three general methodologies to approach the problem of pricing derivatives based on yield curve models. First, additional factors besides the yield curve factors are included in the model and only dedicated to span the variations in the particular derivative market. The second is to add stochastic volatility to the Lévy factors in the original model to reflect the dynamics in the derivative market. The third approach is that people choose to specifically model the market-observable rate that underlies the derivative, such as Libor rate to the cap market or swap rate to the swaption market. By doing this, the underlying rates are automatically matched perfectly and hence arbitrage opportunities are eliminated.

Chapter 5

Conclusion

This work makes contributions to the mathematical finance area in two aspects

—

1. Developed a 3-factor Lévy -based model, which is *the first* attempt to use a non-Brownian motion structure in modeling the short rate dynamics. Bond, caplet/floorlets and swaption formulas are developed under the model.
2. Implemented a large-scale particle filter to estimate the short rate model. Particle filter methodology provides robust estimation for state-space models with non-Gaussian dynamics, and hence is an ideal candidate for studying Lévy based financial models that are becoming increasingly popular.

The estimation results show quality fittings to the historical data, and the estimated model demonstrated certain degree of forecasting power. However, the caplet pricing results indicate that the model is not adequate to price fixed income derivatives without further seasoning. This is in fact consistent with the consensus that equilibrium fixed income models such as the short rate models, regardless of the dynamics, are not able to capture certain characteristics in individual fixed income derivative market.

As of future work, two things can be done along this line.

1. First, the particle filter implementation can be improved by using more effi-

cient optimization routine, a better trial-and-error scheme, or generating more particles to increase the precision.

2. Second, based on the current framework, stochastic volatility can be added to the Lévy factors in order to explain the unique characteristics exhibited from different derivatives market and hence give reasonable prices for those products.

Appendix A

A.1 Solution derivation of OU process, Eqn. (3.9)

Proof. We notice from Eqn (3.8) that

$$\begin{aligned}
 d(e^{\kappa t}x(t)) &= e^{\kappa t}dx(t) + \kappa e^{\kappa t}x(t)dt \\
 \Rightarrow d(e^{\kappa t}x(t)) &= e^{\kappa t}(adt + BdL(s)) \\
 \Rightarrow \int_0^t d(e^{\kappa s}x(s)) &= \int_0^t e^{\kappa s}(ads + BdL(s)) \\
 \Rightarrow e^{\kappa t}x(t) - x(0) &= \int_0^t e^{\kappa s}(ads + BdL(s)) \\
 \Rightarrow x(t) &= e^{-\kappa t}x(0) + \int_0^t e^{-\kappa(t-s)}(ads + BdL(s)).
 \end{aligned}$$

■

A.2 Proof of the Bayes rule in the update stage for the conceptual solution of the filtering problem, Eqn. (4.4)

Proof.

$$\begin{aligned}
 p(x_k|Z_k) &= \frac{p(Z_k|x_k)p(x_k)}{p(Z_k)} \\
 &= \frac{p(z_k, Z_{k-1}|x_k)p(x_k)}{p(z_k, Z_{k-1})} \\
 &= \frac{p(z_k|Z_{k-1}, x_k)p(Z_{k-1}|x_k)p(x_k)}{p(z_k|Z_{k-1})p(Z_{k-1})} \\
 &= \frac{p(z_k|Z_{k-1}, x_k)p(x_k|Z_{k-1})p(Z_{k-1})p(x_k)}{p(z_k|Z_{k-1})p(Z_{k-1})p(x_k)} \\
 &= \frac{p(z_k|x_k)p(x_k|Z_{k-1})}{p(z_k|Z_{k-1})}.
 \end{aligned}$$

■

A.3 Proof of the posterior expansion, Eqn. (4.17)

Proof.

$$\begin{aligned} p(X_k|Z_k) &= \frac{p(z_k|X_k, Z_{k-1})p(X_k|Z_{k-1})}{p(z_k|Z_{k-1})} \\ &= \frac{p(z_k|X_k, Z_{k-1})p(x_k|X_{k-1}, Z_{k-1})p(X_{k-1}|Z_{k-1})}{p(z_k|Z_{k-1})} \\ &= \frac{p(z_k|x_k)p(x_k|x_{k-1})}{p(z_k|Z_{k-1})}p(X_{k-1}|Z_{k-1}) \\ &\propto p(z_k|x_k)p(x_k|x_{k-1})p(X_{k-1}|Z_{k-1}). \end{aligned}$$

We used the Markov properties in deriving those equalities.

■

BIBLIOGRAPHY

- [1] S. Asmussen and J. Rosiński, *Approximation of small jumps of Lévy processes with a view towards simulation*, Journal of Applied Probability, vol. 38, pp 482-493, 2001
- [2] B. Azimi-Sadjadi and P.S. Krishnaprasad, *A Particle Filtering Approach to Change Detection for Nonlinear Systems* EURASIP Journal on Applied Signal Processing, vol 15, pp.2295-2305, 2004
- [3] M. Baxter and A. Rennie, *Financial Calculus : An Introduction to Derivative Pricing*, Cambridge University Press, 1996
- [4] F. Black and P. Karasinski, *Bond and Option Pricing When Short Rates Are Lognormal*, Financial Analysts Journal, v.47, pp.52-59, 1991 Jul/Aug
- [5] F. Black, E. Derman and W. Toy, *A One-Factor Model of Interest Rates and Its Application to Treasury Bond Options*, Financial Analysts Journal, v.46, pp.33-39, Jan/Feb 1990
- [6] F. Black, *The pricing of commodity contracts*, Journal of Financial Economics, Vol.3, pp.167-179, 1976
- [7] M. Bolić, P. Djurić and S. Hong, *Resampling Algorithms and Architectures For Distributed Particles Filters*, Working paper, 2004
- [8] A. Brace, D. Gatarek and M. Musiela, *The Market Model of Interest Rate Dynamics*, Mathematical Finance, Vol.7, no.2, 1997, pp.127-155

- [9] D. Brigo and F. Mercurio, *Interest Rate Models: Theory and Practice*, Springer Finance, 2001
- [10] B. Dupire, *Pricing with a Smile*, Risk, vol.7, no.1, Jan. 1994, pp.18-20
- [11] A. Cairns, *Interest Rate Models*, Princeton University Press, 2004
- [12] P. Carr, H. Geman, D. Madan and M. Yor, *The Fine Structure of Asset Returns: An Empirical Investigation*, Journal of Business, vol.75, no.2, April 2002, pp.305-332
- [13] P. Carr and D. Madan, *Option valuation using fast Fourier Transform*, Journal of Computational Finance, v.2, pp.61-73, 1999
- [14] P. Carr, H. Geman, D. Madan and M. Yor, *Stochastic Volatility for Levy Processes*, Mathematical Finance, vol.13, No.3, July 2003, pp.345-382
- [15] P. Carr and B. Zhang, *Pricing FX Barrier Options under Stochastic Skew Model by Monte Carlo Simulation*, Bloomberg Technical Report, 2005
- [16] P. Carr, L. Wu and B. Zhang, *Finite Difference Method under Heston Stochastic Volatility Model and Stochastic Skew Model for Pricing Barrier Options*, Bloomberg Technical Report, 2005
- [17] R. Cont and P. Tankov, *Financial Modeling with Jump Processes*, Chapman & Hall/CRC, 2004
- [18] J. Cox, J. Ingersoll and S. Ross, *A Theory of the Term Structure of Interest Rates*, Econometrica, vol. 53, 1985, pp. 385-408

- [19] E. Derman and I. Kani, *Riding on a smile*, Risk, vol.7, no.2, 1994, pp.32-39
- [20] A. Doucet, *On Sequential Simulation Based Method for Bayesian Filtering*, Technical report CUEDF-INFENGTR.310, 1998
- [21] J.C. Duan and A. Fulop *Estimating the Structural Credit Risk Model When Equity Prices Are Contaminated by Trading Noises*, Working paper, 2005
- [22] D. Duffie, *Dynamic Asset Pricing Theory*, Princeton University Press, 1992
- [23] E. Eberlein and S. Raible, *Term Structure Models Driven By General Lévy Processes*, Mathematical Finance, vol.9, no.1, Jan. 1999, pp.31-53
- [24] H. Geman, E. Eberlein and D. Madan, *Formulation of Lévy Based Interest Rate Model*, working paper, Feb. 2005
- [25] H. Geman, N. Karoui and J. Rachev, *Changes of Numerair, Changes of Probability Measure and Option Pricing*, Journal of Applied Probability, vol.32, no.2, June 1995, pp.443-458
- [26] P. Glasserman, *Monte Carlo Methods in Financial Engineering*, Springer Finance, 2003
- [27] B. Ristic, S. Arulampalam and N. Gordon, *Beyond the Kalman Filter — Particle Filters for Tracking Applications*, Artech House, Boston and London, 2004
- [28] A. Doucet, N. de Freitas and N. Gordon, *Sequential Monte Carlo Methods in Practise*, Springer 2001

- [29] J. Harrison and D. Kreps, *Martingale and Arbitrage in Multiperiod Securities Markets*, Journal of Economic Theory, 20, 1979, pp.381-408
- [30] J. Harrison and S. Pliska, *Martingale and Stochastic Integrals in the Theory of Continuous Trading*, Stochastic Processes and Their Applications, 11, 1981, pp.215-260
- [31] D. Heath, R. Jarrow and A. Morton, *Bond Pricing and the Term Structure of Interest Rates: A New Methodology for Contingent Claims Valuation*, Econometrica, Vol. 60, No. 1, 1992, pp. 77-105
- [32] M. Heidari and L. Wu, *Are Interest Rate Derivatives Spanned by the Term Structure of Interest Rates?* Journal of Fixed Income, June 2003, pp. 75-86
- [33] M. Heidari and L. Wu, *Term Structure of Interest Rates, Yield Curve Residuals, and Consistent Pricing of Interest Rate Derivatives* Working paper, 2005
- [34] S. Heston, *A Closed-form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options*, Review of Financial Studies, vol.6, no.2, 1993, pp.327-343
- [35] J. Hull, *Options, Futures And Other Derviatives*, Prince Hall, 2005 July
- [36] J. Hull and A. White, *An Analysis of the Bias in Option Pricing Caused by a Stochastic Volatility*, Advances in Futures and Options Research, vol.3, 1988, pp.29-61
- [37] P. Jäckel, *Monte Carlo methods in finance*, John Wiley & Sons Ltd. 2002

- [38] R. Jamshidian, *Libor and Swap Market Models and Measures*, Finance and Stochastics, Vol.1, pp.290-330, 1997
- [39] A. Javaheri, *Inside Volatility Arbitrage*, Wiley Finance, 2005
- [40] S. Kou and H. Wang, *First passage times of a jump diffusion process*, Advanced Applied Probability, 35, 2003, pp. 504-531
- [41] R. Litterman, J. Scheinkman, *Common Factors Affecting Bond Returns*, Journal of Fixed Income, 1, 1991, pp. 54-61
- [42] D. Madan, P. Carr and E. Chang, *The Variance Gamma Process and Option Pricing*, European Finance Review, vol.2, no.1, June 1998, pp.79-105
- [43] D. Madan and M. Yor, *Representing CGMY and Meixner Lévy Processes as Time Change Brownian Motions*, Working paper, 2005
- [44] C. Menn and S. Rachev, *Calibrated FFT-based Density Approximations for α -stable Distributions*, working paper, 2004
- [45] R. Merton, *Option Pricing When Underlying Stock Returns Are Discontinuous*, Journal of Financial Economics, vol.3, 1976, pp.125-144
- [46] P. Protter, *Stochastic Integration and Differential Equations, Second Edition*, Springer-Verlag, Heidelberg, 2005
- [47] R. Rebonato, *Interest-Rate Option Models, second edition*, John Wiley & Sons, 1998

- [48] M. Rubinstein, *Implied Binomial Trees*, Journal of Finance, vol.49, no.3, Jan. 1994, pp.771-818
- [49] K. Sato, *Lévy Processes and Infinitely Divisible Distribution*, Cambridge Press, 1999
- [50] W. Schoutens, *Lévy Processes in Finance*, Wiley Finance, 2003
- [51] O. Vasicek, *An Equilibrium Characterization of the Term Structure*, Journal of Financial Econometric, v.5, 1977, pp.177-188
- [52] R. Zagst, *Interest Rate Management*, Springer Finance, 2002
- [53] B. Zhang, *Simulating CGMY as Time-Change Brownian Motion*, Bloomberg Technical Report, 2005
- [54] F. Zhou, *Black Smirks*, Risk, pp.87-91, May 2003