

ABSTRACT

Title of dissertation: MODELING VIX AND VIX DERIVATIVES
 WITH MEAN REVERTING MODELS
 AND PARAMETER ESTIMATION
 USING FILTER METHODS

 GUOYUAN LIU, Doctor of Philosophy, 2012

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

In this thesis, we study the mean reverting property of the VIX time series, and use the VIX process as the underlying. We employ various mean reverting processes, including the Ornstein-Uhlenbeck (OU) process, the Cox-Ingersoll-Ross (CIR) process and the OU processes driven by Lévy processes (Lévy OU) to fit historical data of VIX, and calibrate the VIX option prices. The first contribution of this thesis is to use the Lévy OU process to model the VIX process, in order to explain the observed high kurtosis. To price the option using the Lévy OU process, we develop a FFT method.

The second contribution is to build a joint framework to consistently model the VIX and VIX derivatives together on the entire time series of market data. We choose multi-factor mean-reverting models, in which we model the VIX process as a linear combination of latent factors. To estimate the models, we use Euler approximation to find a discrete approximation for the VIX process. Based on this

approximate, we consider various filter methods, namely, the Unscented Kalman Filter (UKF), constrained UKF, mixed Gaussian UKF and Particle Filter (PF) for estimation. The performances of these models are compared and discussed. Radon Nikodym derivatives of the risk-neutral measure are discussed with respect to the physical measure for the jumps. A simple dynamic trading strategy was tested on these models.

MODELING VIX AND VIX DERIVATIVES WITH
MEAN REVERTING PROCESSES AND PARAMETER
ESTIMATION USING FILTER METHODS

by

Guoyuan Liu

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
2012

Advisory Committee:
Professor Dilip B. Madan, Chair/Advisor
Professor Eric V. Slud
Professor Paul J. Smith
Professor Victor M. Yakovenko
Professor Grace Yang

© Copyright by
Guoyuan Liu
2012

Acknowledgments

First of all, I would like to give my utmost gratitude to my advisor, Professor Dilip Madan. His perpetual energy and enthusiasm in research had motivated all his advisees including me. His advice, support and encouragement drove me forward throughout these years in my PhD pursuance. Without these, this dissertation would not have been possible.

I also owe my gratitude to Professor Paul Smith, Professor Eric Slud for their outstanding teaching in statistics course, their guidance in my PhD study and their agreeing to serve on my dissertation committee. I also want to thank Victor Yakovenko and Professor Grace Yang for agreeing to serve on my dissertation committee and spending much time reviewing my manuscript.

I want to thank Prof. Michael Fu for his excellent organization of our weekly finance research interaction team (RIT). I thank all participants in the RIT for helpful discussions. In addition, I would also like to say thank you to our math finance group members, Huaqiang Ma, Geping Liu and Lingyan Cao, from whom I have learned a lot.

I would like to express my grateful feelings to Ms. Alverda McCoy and her help in the administrative chores and her friendship.

Table of Contents

List of Tables	v
List of Figures	vi
1 Introduction	1
1.1 VIX and VIX Derivatives	1
1.2 VIX Models	2
1.3 Multi-factor Dynamic Model	5
2 Lévy Process and Mean Reverting Process	9
2.1 Lévy Process	9
2.1.1 Introduction	9
2.1.2 Variance Gamma Process	12
2.1.3 CGMY Process	14
2.1.4 Equivalence of Measures for Lévy processes	15
2.2 Ornstein Uhlenbeck (OU) Process	18
2.2.1 Definition	18
2.3 Cox-Ingersoll-Ross (CIR) Process	19
2.4 Lévy OU Process	20
3 One-Factor Model	23
3.1 Estimate the VIX Process in Physical Measure	23
3.1.1 Gaussian OU	23
3.1.2 CIR	25
3.1.3 Lévy OU	25
3.1.4 Results	26
3.2 Model Calibration using Option Price	30
3.2.1 CIR	30
3.2.2 Lévy OU and Pricing Option Using FFT	31
3.2.3 Numerical Results	33
3.3 Conclusions	37
4 Filter Methods	39
4.1 Filter Problem	39
4.2 Kalman Filter	41
4.3 Unscented Kalman Filter	42
4.3.1 Unscented Transformation	43
4.3.2 Algorithm	44
4.3.2.1 Constrained UKF	47
4.3.3 Mixed-Gaussian UKF	48
4.4 Particle Filter	50
4.4.1 Importance Sampling	51

5	Multifactor Model	55
5.1	Data	55
5.2	The Basic Model Structure	56
5.3	GOU3 Model with UKF	57
5.3.1	VIX Futures and Options	59
5.3.2	Parameter Estimation	59
5.3.3	In-Sample Test	61
5.3.4	Out-of-Sample Test	64
5.4	CIR3 Model with Constrained UKF	68
5.4.1	Results	69
5.5	CGMOU2 with Mixed-Gaussian UKF	71
5.6	CGMOU2 with Particle Filter	73
5.7	Tests on the Models	76
5.7.1	Radon-Nikodym Derivative	76
5.7.2	Trade Strategy Test	80
5.8	Conclusion	81
6	Conclusion and Future Study	83
A.1	The Derivation of VIX Call Price	85
	Bibliography	86

List of Tables

2.1	Cumulants of VG and $CGMY$	15
3.1	Result of estimation in physical measure. Parameters are estimated on VIX data from January 2, 1990 to October 27, 2010.	28
3.2	The average error statistics of various models in the estimation of VIX options. The error statistics are computed for five consecutive Wednesdays , February 28, 2006 through March 28, 2006, and then averaged.	34
3.3	The estimated parameters and the G/M ratios using CGMOU model. The parameters are estimated from VIX options for five consecutive Wednesdays.	37
5.1	In-sample maximum likelihood parameter estimates of GOU3 model with Unscented Kalman filter from March 30, 2004 to December 16, 2008	61
5.2	In-sample error statistics (GOU3 model) for VIX and VIX derivatives from March 30, 2004 to December 16, 2008	62
5.3	Out-of-sample error statistics (GOU3 model) for VIX and VIX derivatives for VIX from March 23, 2008 to October 27, 2009	65
5.4	In-sample maximum likelihood parameter estimates of CIR3 model with Constrained Unscented Kalman filter from March 30, 2004 to December 16, 2008	70
5.5	In-sample error statistics (CIR3) for VIX and VIX derivatives	70
5.6	Out-of-sample error statistic (CIR3) for VIX and VIX derivatives	71
5.7	In-sample maximum likelihood parameter estimates of CGMOU2 model with Mixed-Gaussian UKF from March 30, 2004 to December 16, 2008	72
5.8	In-sample error statistics (CGMOU2, Mixed Gaussian UKF) for VIX and VIX derivatives	73
5.9	Out-of-sample error statistic (CGMOU2, Mixed Gaussian UKF) for VIX and VIX derivatives	73
5.10	In-sample maximum likelihood parameter estimates of CGMOU2 model with Particle filter from March 30, 2004 to December 16, 2008	75
5.11	In-sample error statistic (CGMOU2, PF) for VIX and VIX derivatives	75
5.12	Out-of-sample error statistic (CGMOU2, PF)for VIX and VIX derivatives	76

List of Figures

3.1	Historical VIX path from January 1990 to September 2010.	24
3.2	Empirical PDF of \tilde{V} and model PDF of \tilde{U} for Gaussian OU, CIR, CGMOU, CGMYOU, CMOU and CMYOU models. After \tilde{V} are binned, the empirical PDF is given by the ratio of the number of points in each bin over the total number of points divided by the bin width. . .	29
3.3	Inverse Gaussian distribution with mean 10 and variance 6 and its call calculated by FFT	33
3.4	The Gaussian OU model fitted call price and the historical price on Feb 28, 2006 with maturity of 78 days	36
3.5	The CGMOU model fitted call price and the historical price on Feb 28, 2006 with maturity of 78 days	36
4.1	Illustration of estimation with state constraints	48
5.1	In-sample test (GOU3 model) for VIX from March 30, 2004 to December 16, 2008.	62
5.2	In-sample test (GOU3 model) for VIX Futures on October 21, 2008 and December 16, 2008.	63
5.3	In-sample test (GOU3 model) for VIX calls and puts on October 21, 2008 with maturities 36, 64, 92, 127, 155 days.	64
5.4	Out-of-sample test (GOU3 model) for VIX from March 23, 2008 to October 27, 2009.	66
5.5	Out-of-sample test (GOU3 model) for VIX Futures on June 23, 2009 and October 6, 2009.	67
5.6	Out-of-sample test (GOU3 model) on VIX calls and puts on October 6, 2009 with maturities 15, 43, 71, 106, 134 days	67
5.7	Radon-Nikodym derivatives for BDLP factor x_1 and x_2	79
5.8	Radon-Nikodym Derivative for BGLV factor x_1 and x_2	81

Chapter 1

Introduction

1.1 VIX and VIX Derivatives

The CBOE Volatility Index or VIX was first introduced by CBOE back in 1993 to measure the markets expectation of 30- day volatility implied by at-the-money S&P 100 option prices. VIX soon became the premier benchmark for U.S. stock market volatility and is often referred to as the "fear index" (CBOE 2003 [24]). In 2003, the CBOE revised the definition of VIX. The new VIX is the square root of one month par variance swap rate on S&P 500. A variance swap is a forward contract on annualized variance, the square of the realized volatility. The pay off of a one month variance swap (ignore the notional) is given by

$$Payoff = \frac{N}{n} \sum_{i=1}^n [\log(\frac{S_i}{S_{i-1}})]^2 - K^2 \quad (1.1)$$

where S_i is the stock price, N/n is the annualization factor, N and n are the number of business days in one year and on month respectively, and K is the quote. Therefore the value of VIX can be expressed as

$$VIX^2 = E^Q \frac{N}{n} \sum_{i=1}^n [\log(\frac{S_i}{S_{i-1}})]^2 \quad (1.2)$$

where E^Q stands for risk-neutral measure. The new VIX estimates expected volatility by averaging the weighted prices of S&P 500 puts and calls over a wide range of strike prices. The methodology is based on the finding of Carr and Mandan (1998)

[19] that a variance swap can be perfectly statistically replicated through vanilla puts and calls whereas a volatility swap requires dynamic hedging. Most data service providers (as well as the CBOE) now provide data for the new VIX calculated retroactively back to the early 1990s. But apart from its role as a risk indicator, nowadays it is possible to directly invest in volatility as an asset class by means of VIX derivatives. VIX futures were introduced by CBOE in March 2004. They are standard futures contracts on VIX. Further, European-style options on the VIX, VIX options were introduced in February 2006. VIX options and VIX futures are among the most actively traded contracts at CBOE and CBOE Futures Exchange (CFE). The negative correlation of volatility to stock market returns is well documented and suggests a diversification benefit to including volatility in an investment portfolio. VIX futures and options are designed to deliver pure volatility exposure in a single, efficient package(CBOE 2003 [24]).

1.2 VIX Models

In the early 1970s, Black, Scholes and Merton developed the famous Black Scholes model. The Black-Scholes model, which uses geometric Brownian motion to model the stock process, is the most celebrated model in the pricing of stock options. However empirical evidence suggests that the geometric Brownian motion assumption does not describe the statistical properties of financial time series very well, mainly in two aspects. In Cont (2001) [21] a more extended list of stylized features of financial data is given.

- The log returns do not behave according to a normal distribution.
- The estimated volatilities or the parameters of uncertainty estimated (or more generally the environment), for instance, the historical volatilities change stochastically over time and are clustered (Schoutens 2003[56]).

The first aspect has encouraged researchers to develop non-Gaussian models in the stock process. The Lévy models are one category of the successful models. The second aspect motivates the introduction of models where volatility is itself stochastic. Empirical data show non-Gaussianity in the VIX process too. In order to model the stochastic evolution of the VIX, it is reasonable to employ models developed for stocks and stock indexes. However, the major difference between the two areas is that VIX is mean-reverting while stocks or stock indexes are not. Numerous models that describe the general volatility or VIX processes have been implemented in the literature.

Whaley (1993) [55] used the geometric Gaussian model in the pricing of volatility options. Grunbichler and Longstaff (1996) [31] proposed the square root process, Cox-Ingersoll-Ross (CIR) to model VIX and price the VIX options. Assuming a stochastic volatility model for the stock process, Howison, Rafailidis, and Rasmussen (2004) [34] provided closed-form formulas for volatility-average and variance swaps for a variety of diffusion and jump-diffusion models for volatility, and described a general partial differential equation framework for volatility derivatives. Lin and Chang (2009) [45] derived a closed-form VIX option pricing model by using correlated jump diffusion processes to model both the S&P 500 index and the S&P 500

instantaneous volatility.

Due to incompleteness of the market, the VIX indexes cannot be replicated by the stocks or options. VIX derivatives are trading instruments. Historical data are available for the VIX and its derivatives, if we single out the market related to VIX, the only underlying process is the stochastic VIX index level. In this thesis, we concentrate on the mean-reverting characteristic of the VIX, using mean-reverting processes as the underlying process for models. Starting from assuming VIX is a mean-reverting process V_t , the VIX derivatives are priced by arbitrage-free theory. Let T be the maturity of the derivative contracts, let r be the interest rate and let $\tau = T - t$. The fair price of a VIX future is the time t risk-neutral expectation of the VIX at maturity T ,

$$F(t, V_t) = E_t^Q V_T, \quad (1.3)$$

Similarly, the prices of a European VIX call and put can be written as

$$\begin{aligned} C(t, V_t) &= e^{-r\tau} E_t^Q (V_T - K)^+ \\ P(t, V_t) &= e^{-r\tau} E_t^Q (K - V_T)^+. \end{aligned} \quad (1.4)$$

We compare several mean-reverting models, including the Ornstein-Uhlenbeck (OU) model, the Cox-Ingersoll-Ross (CIR) model and two Lévy OU models. We estimate their parameters and compare their performance both in physical measure and risk-neutral measure. The comparisons are done on both in-sample data and out-of-sample data. Two Lévy OU processes, Variance Gamma OU (or CGMOU) and CGMY OU (CGMY process see [18]), are used in this research, and have been proved to perform well. The uneven jump measure in the Lévy OU process yields

the desired skewness, and a heavy tail weight in the jump measure provides adequate kurtosis.

1.3 Multi-factor Dynamic Model

Option pricing models constructed for a variety of underlying assets are typically calibrated to derivative prices at a single time point. In interest rate modeling, there is an extensive literature on dynamic term structure models that estimate models across an entire time series of market prices of discount bonds and interest rate derivatives. Once the parameters are estimated from the historical data, the models are used to mark to market other over-the-counter derivatives or to make predictions for the market over a period of time without calibration. Examples include Duffee (2002) [28], Dai and Singleton (2000) [26]), Heidari and Wu (2003) [32] and Heidari and Wu (2009) [33].

Heidari and Wu (2009) propose a multi-factor model structure using Gaussian OU and CIR factors that successfully prices both interest rates and interest rate options. They employ two models; each has two orthogonal sets of mean reverting factors with the first set driving the yield curve and the second set driving options exclusively. The first model uses Gaussian affine or OU factor structure, while the second model uses a CIR factor structure, which allows for both stochastic central tendency and stochastic volatility. The estimation results show that the three yield curve factors explain over 99% of the variation of the yield curve, with three additional option factors improving the explained percentage variation of implied

volatilities to over 99%. Both models give comparably satisfactory results.

The VIX has the same mean-reverting character as the interest rate, and consistent records of time series data of VIX index and its derivatives are available going back more than four years. Inspired by Heidari and Wu, in the second part of the research, we use multi-factor mean-reverting models to build a consistent frame for the time series of index levels and the derivative prices. Under this framework, the factors capture both the systematic variation of VIX in the physical measure and the VIX derivative variation in the risk-neutral measure. We begin from the 3-factor Gaussian OU process (GOU3) and the 3-factor CIR process (CIR3), which are similar to the models in Heidari's paper, and then introduce the 2-factor CGMOU process (CGMOU2), which is a Lévy OU process driven by Variance Gamma factors.

In the multi-factor dynamic models, the VIX is modeled as a linear combination of mean-reverting factors. The factors are unobservable or latent, while the VIX price level and the price of the VIX futures and options are the observable measurements. The idea of the model is formulated below:

$$\begin{aligned}x_t &= F(x_{t-1}) + \epsilon_t \\y_t &= H(x_t; \Theta) + e_t\end{aligned}\tag{1.5}$$

where the first equation is the discretized approximation to the process of the state factors x_t , $F(x_{t-1})$ is the drift term, and ϵ_t is the uncertainty of the dynamic forward process. In the second equation, y_t are the observed prices, $H(x_t; \Theta)$ represents the measurement function with respect to the state variables, Θ is the parameter set to be estimated and e_t is the measurement noise.

The estimation of parameters in a dynamic setting is one of the major function of the filter technique. The non-linear measurement function $H(x_t; \Theta)$ excludes the use of the regular Kalman filter. The choice of different filter methods is based on the assumption about the ϵ_t . In the first two multi-factor dynamic models, GOU3 and CIR3, the background driving randomness is Gaussian, and we use Unscented Kalman filter (Julier 1997 [36]). For the CGMOU model, whose background driving randomness is a Variance Gamma process, we employ two filter methods to handle the non-Gaussianity. We can either employ a mixed-Gaussian density to approximate the non-Gaussian ϵ_t , or we can directly use the Particle Filter (PF) [17] to approximate the distribution of the states.

At each time step the filter time-updates the latent x_t and gives the prediction of the measurement mean \tilde{y}_t . Once the new market observation y_t is observed, the filter updates the state variables according to Bayes' theorem. The second step is called measurement update. A filter technique recursively performs time updating and measurement updating at each time step until it reaches the end of the time series. We further assume that the forecasting errors on the measurement series are normally distributed and define the weekly log likelihood function (ignoring the constant term) as

$$l_t(y_t; \Theta) = -\frac{1}{2}(y_t - \tilde{y}_t)^T(A_t)^{-1}(y_t - \tilde{y}_t), \quad (1.6)$$

where A_t denotes the conditional covariance matrix of the forecasts of the measurement series, y_t and \tilde{y}_t are market observations and model predictions respectively. The model parameters are estimated by maximizing the sum of log likelihoods de-

finned in Equation (1.6) over all time steps $t = 1, \dots, T$:

$$\hat{\Theta} = \arg \max_{\Theta} \sum_{t=1}^T l_t(y_t; \Theta). \quad (1.7)$$

In this dissertation, Chapter 2 introduces the Lévy process and three types of mean-reverting process: the Gaussian OU, CIR and general Lévy OU. Chapter 3 applies the mean-reverting process to estimate the parameters both in the physical measure and the risk-neutral measure and compares their performances. Chapter 4 briefly introduces filter techniques and focuses mainly on the filter methods employed in the dissertation. The algorithms of Unscented Kalman Filter, mixed-Gaussian Filter and Particle Filter are reviewed. In Chapter 5, using the mean-reverting processes described above, we build multifactor dynamic models for the joint physical measure and risk-neutral measure. Then we apply various filter techniques to estimate the parameters for those models. Their performances are compared.

Chapter 2

Lévy Process and Mean Reverting Process

2.1 Lévy Process

2.1.1 Introduction

Brownian motion has been the most widely studied stochastic process in mathematical finance since Bachelier [6] introduced the Brownian motion to model asset prices. Bachelier's model led to the celebrated Black-Scholes model (Black and Scholes 1973) [12], where the asset price follows geometric Brownian motion. However, empirical data does not agree with the Brownian models very well. For instance, the historical log returns do not have a normal distribution. They have fat tails or excess kurtosis. The Black-Scholes model cannot explain the so-called volatility smile. In order to overcome this imperfection, non-normal Lévy models were developed and have become increasingly popular in the last decade. Mandelbrot [43] studied the first non-normal exponential Lévy process in the 1960s and introduced the α -stable Lévy motion with index $\alpha < 2$. Later, models based on three or more general pure jump Lévy processes, such as variance gamma (VG), normal inverse Gaussian (NIG) and CGMY, were developed and studied.

The Lévy process, named after the French mathematician Paul Lévy, is a stochastic process which is continuous in probability and has independent and sta-

tionary increments.

The Lévy process can be thought of as an analogue of random walk in continuous time. Every Lévy process has a càdlàg, i.e. right continuous with left limits, modification which is itself a Lévy process. The formal definition can be written as follows:

Definition 1 (Lévy Process) *A càdlàg stochastic process $(x_t)_{t \geq 0}$ on $(\Omega, \mathcal{F}, \mathbb{P})$ with $x_0 = 0$ is called a Lévy process if it possesses the following properties citeSchoutens:*

- *Independent increments: for any $0 < t_0 < t_1 < \dots < t_n$, the random variables $x_{t_0}, x_{t_1} - x_{t_0}, x_{t_2} - x_{t_1}, \dots, x_{t_n} - x_{t_{n-1}}$ are independent.*
- *Stationary increments: the law of $x_{t+h} - x_t$ does not depend on t .*
- *Stochastic continuity: $\forall \epsilon > 0, \lim_{h \rightarrow 0} P(|x_{t+h} - x_t| \geq \epsilon) = 0$.*

The independent stationary increments property of a Lévy process leads to infinite divisibility of the marginal distribution. For any positive integer n , if we sample the path of a Lévy process with n equal time intervals, x_t can be written as a sum of n iid random variables. In other words, the characteristic function of x_t is the n th power of another characteristic function. This property of the distribution is called infinite divisibility, because of which the characteristic function $\phi_{x_t}(u)$ of Lévy process x_t can be expressed in a simple form

$$\phi_{x_t}(u) = E[e^{iux_t}] = e^{t\psi(u)},$$

where $\psi(u)$ is called the characteristic exponent of the Lévy process [22].

A Lévy process can be decomposed into three independent components: a deterministic drift with rate γ , a continuous path diffusion with volatility σ and a jump process with Lévy measure ν . Hence, a Lévy process can be fully characterized by (γ, σ^2, ν) . We call (γ, σ^2, ν) the characteristic triplet of the Lévy process.

Definition 2 (Lévy Measure) *Let $(X_t)_{t \geq 0}$ be a Lévy process on \mathbb{R}^d . The measure ν on \mathbb{R}^d defined by:*

$$\nu(A) = E[\#\{t \in [0, 1] : \Delta x_t \neq 0, \Delta x_t \in A\}], A \in \mathcal{B}(\mathbb{R}^d)$$

is called the Lévy measure of x : $\nu(A)$ is the expected number of jumps per unit time whose size belongs to A . If the Lévy measure is of the form $\nu(dx) = K(x)dx$, we call $K(x)$ the Lévy density.

For a one-dimensional Lévy process, the Lévy -Khintchine formula gives the expression for characteristic exponent.

Theorem 1 (Lévy-Khintchine Representation) *Let $(x_t)_{t \geq 0}$ be a Lévy process on \mathbb{R} . The characteristic exponent $\psi_{x_1}(u)$ can be decomposed as follows:*

$$\psi_{x_1}(u) = iu\gamma - \frac{1}{2}\sigma^2 u^2 + \int_{\mathbb{R} \setminus \{0\}} (e^{iux} - 1 - iux1_{|x| < 1})\nu(dx). \quad (2.1)$$

with $\int_{\mathbb{R} \setminus \{0\}} (1 \wedge x^2)\nu(dx) < \infty$.

The quantity $\int_{\mathbb{R} \setminus \{0\}} \nu(dx)$ denotes the total arrival rate or activity of a Lévy process. The Lévy process is of finite activity if $\int_{\mathbb{R} \setminus \{0\}} \nu(dx) < \infty$. Otherwise the Lévy process has infinite activity. The quantity $\int_{\mathbb{R} \setminus \{0\}} |x|\nu(dx)$ denotes the total variation. If $\int_{\mathbb{R} \setminus \{0\}} |x|\nu(dx) < \infty$, the Lévy process is of finite variation.

The Brownian motion and the Poisson process are special cases of the Lévy process. When $\nu = 0$, the Lévy process does not have jumps, the path is continuous, and the process becomes a Brownian motion. When the Lévy measure $\nu(dx) = \lambda\delta(1)$, λ is a constant, $\delta(1)$ is the Dirac function at 1, and the process becomes a Poisson process. When $\sigma^2 = 0$, the Lévy process has no diffusion part and becomes a pure jump process. Pure jump Lévy processes have no Brownian motion component but their tiny jumps mimic continuous movement. The Variance Gamma (VG) (Madan and Seneta 1987[42]) and CGMY (Carr et al 2002[18]) are two well known pure jump Lévy processes.

2.1.2 Variance Gamma Process

A Variance Gamma process can be expressed in two forms [41]. A VG process VG can be expressed as a Brownian motion $\theta t + \sigma W_t$ time-changed by a gamma process $g_t(1, \nu)$:

$$x_t(\theta, \sigma, \nu) = \theta g_t(1, \nu) + \sigma W(g_t(1, \nu)), \quad (2.2)$$

where $W = (W_t; t \geq 0)$ is a standard Brownian motion and the independent subordinator (i.e. an increasing, positive Lévy process), $g_t(1, \nu)$ is a gamma process with unit mean rate and variance rate ν . A gamma process is a random process with independent gamma distributed increments and it has a gamma marginal distribution [3].

Alternatively, a VG process VG can be represented as the difference of two

gamma random processes $g_t(C, 1/M)$ and $g_t(C, 1/G)$:

$$x_t(C, G, M) = g_t(C, 1/M) - g_t(C, 1/G) \quad (2.3)$$

where

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

The following equations give the characteristic exponent and Lévy density of a VG process:

- Characteristic exponent

$$\begin{aligned} \psi_{VG}(u) &= -1/\nu \log(1 - i\theta\nu u + (\sigma^2\nu/2)u^2) \\ &= -C(\log(1 - iu/M) + \log(1 + iu/G)) \end{aligned}$$

- Lévy density

$$K_{VG}(x) = \begin{cases} \frac{C e^{Gx}}{|x|} & x < 0 \\ \frac{C e^{-Mx}}{x} & x \geq 0 \end{cases} \quad (2.4)$$

The VG process has infinite activity. That is the VG process has infinitely many jumps in any finite time interval. The VG process also has paths of finite variation with no Brownian component. The popularity of the VG process lies in its flexibility in handling skewness and excess kurtosis. When $\theta = 0$, or $G = M$,

the distribution is symmetric. Negative values of θ or $G > M$ result in negative skewness in the distribution. Similarly, the parameter $\nu = 1/C$ primarily controls the kurtosis.

2.1.3 CGMY Process

Through introduction of a fourth parameter Y to the second parameterization $VG(C,G,M)$, the Variance Gamma process is extended to a new model. The generalized model is called the CGMY model as developed by Carr et al 2002 [18]. The parameter Y is added to the power of the denominator x in the expression for the Lévy density of the VG process

$$K_{CGMY}(x) = \begin{cases} \frac{C e^{Gx}}{|x|^{1+Y}} & x < 0 \\ \frac{C e^{-Mx}}{x^{1+Y}} & x \geq 0 \end{cases} \quad (2.5)$$

where $C, G, M > 0, Y < 2$. $Y < 2$ is needed to guarantee a finite second moment.

Adding the fourth parameter Y allows us to have more flexibility to control the behavior of the path. For $Y < 0$ the Lévy process is of finite activity; for $0 \leq Y < 1$, the process is of infinite activity but finite variation; for $1 \leq Y < 2$ the process is of infinite activity and infinite variation. The VG process is a special case of CGMY with $Y = 0$.

The characteristic exponent of CGMY is given by

$$\psi_{CGMY}(u) = C\Gamma(-Y) \left((M - iu)^Y - M^Y + (G + iu)^Y - G^Y \right) \quad (2.6)$$

where $\Gamma(\cdot)$ is the gamma function. The cumulants of $VG(\theta, \sigma, \nu)$ and $CGMY(C, G, M, Y)$ are summarized in Table 2.1:

Table 2.1: Cumulants of VG and $CGMY$

Cumulants	$VG(\theta, \sigma, \nu)$	$CGMY(C, G, M, Y)$
Mean	θt	$tC\Gamma(1 - Y)(M^{Y-1} - G^{Y-1})$
Variance	$\sigma^2 t + \nu \theta^2 t$	$tC\Gamma(2 - Y)(M^{Y-2} - G^{Y-2})$
c_3	$3\sigma^2 \theta \nu t + 2\theta^3 \nu^2 t$	$tC\Gamma(3 - Y)(M^{Y-3} - G^{Y-3})$
c_4	$3\sigma^4 \nu t + 6\theta^4 \nu^3 t + 12\sigma^2 \theta^2 \nu^2 t$	$tC\Gamma(4 - Y)(M^{Y-4} - G^{Y-4})$

2.1.4 Equivalence of Measures for Lévy processes

Definition 3 Let \mathbb{P} and \mathbb{Q} be two probability measures on the same measurable space (Ω, \mathcal{F}) . We say \mathbb{P} and \mathbb{Q} are equivalent if for all $F \in \mathcal{F}$

$$\mathbb{P}(F) = 0 \Leftrightarrow \mathbb{Q}(F) = 0.$$

The Radon-Nikodym derivative denoted by $\Lambda = d\mathbb{Q}/d\mathbb{P}$ is a positive random variable such that $\forall A \in \mathcal{F}$, $\mathbb{Q}(A) = \int_A \Lambda d\mathbb{P}$. One property of the Radon-Nikodym derivative is that for any random variable Z , we have

$$E^{\mathbb{Q}}[Z] = E^{\mathbb{P}} \left\{ Z \frac{d\mathbb{Q}}{d\mathbb{P}} \right\}.$$

Proposition 1 (Equivalence of measures for Brownian motions with drift)

Let (X, \mathbb{P}) and (X, \mathbb{Q}) be two Brownian motions on (Ω, \mathcal{F}_t) with volatilities $\sigma^{\mathbb{P}} > 0$ and $\sigma^{\mathbb{Q}} > 0$ and drifts $\mu^{\mathbb{P}}$ and $\mu^{\mathbb{Q}}$. The measures \mathbb{P} and \mathbb{Q} are equivalent if $\sigma^{\mathbb{P}} = \sigma^{\mathbb{Q}}$ and singular otherwise. When they are equivalent the Radon-Nikodym derivative is

$$\frac{d\mathbb{Q}}{d\mathbb{P}} \Big|_{\mathcal{F}_t} = \exp \left\{ \frac{\mu^{\mathbb{Q}} - \mu^{\mathbb{P}}}{\sigma^2} X_t - \frac{1}{2} \frac{\mu^{\mathbb{Q}} - \mu^{\mathbb{P}}}{\sigma^2} t \right\}. \quad (2.7)$$

A more general version of this result is given by Giranov theorem [?]

Proposition 2 (Equivalence of measures for general Lévy Processes [?]Sata 1999Sato

Let (X, \mathbb{P}) and (X, \mathbb{Q}) be two Lévy processes on \mathbb{R} with characteristic triplets (σ^2, ν, γ) and $(\sigma'^2, \nu', \gamma')$. Then \mathbb{P} and \mathbb{Q} are equivalent if and only if the three following conditions are satisfied:

1. $\sigma = \sigma'$;
2. the Lévy measures are equivalent with

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

where $\phi(x) = \log(d\nu'/d\nu)$;

3. if $\sigma = 0$ then we must in addition have

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

When \mathbb{P} and \mathbb{Q} are equivalent, the Radon-Nikodym derivative is

$$\frac{d\mathbb{Q}}{d\mathbb{P}} \Big|_{\mathcal{F}_t} = e^{U_t}$$

with

$$U_t = \eta X_t^c - \frac{\eta^2 \sigma^2 t}{2} - \eta \gamma t + \lim_{\epsilon \rightarrow 0} \left(\sum_{s \leq t, |\Delta X_s| > \epsilon} \phi(\Delta X_s) - t \int_{|x| > \epsilon} (e^{\phi(x)-1} \nu(dx)) \right).$$

Here (X_t^c) is the continuous part of (X_t) and η is such that

$$\gamma' - \gamma - \int_{-1}^1 x(\nu' - \nu)(dx) = \sigma^2 \eta$$

if $\sigma > 0$ and $\eta = 0$ if $\sigma = 0$.

From the Lévy measure, one can show that a CGMY process can be decomposed into the difference of a positive jump process and a negative jump process:

$$X_t = X_t^+ - X_t^-.$$

For a VG process X_t^+ and X_t^- are both gamma processes.

Proposition 3 (Change of Measure for CGMY process (Fu 2007[29])) *Let \mathbb{P} and \mathbb{Q} be two probability measures on the path space generated by two processes $CGMY(C, G, M, Y)$ and $CGMY(C', G', M', Y')$. The measures \mathbb{P} and \mathbb{Q} are equivalent if and only if $C = C'$ and $Y = Y'$. If they are equivalent, the Radon-Nikodym derivative $d\mathbb{Q}/d\mathbb{P}|_{\mathcal{F}_t}$ depends only on the accumulated positive jumps X_t^+ and negative jump X_t^- instead of the whole path $\Delta X_s^\pm, s \leq t$, and*

$$\frac{d\mathbb{Q}}{d\mathbb{P}}|_{\mathcal{F}_t} = e^{-tZ} \phi^+(X^+) \phi^-(X^-)$$

where

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

2.2 Ornstein Uhlenbeck (OU) Process

2.2.1 Definition

Mean reversion is a tendency for a stochastic process to remain near, or tend to return over time to a long-run mean. Many finance data exhibit mean-reverting properties, for instance interest rate and volatility. The simplest and most widely used mean reverting process is Ornstein Uhlenbeck (OU) Process.

Definition 4 (Ornstein Uhlenbeck (OU) Process) *A càdlàg stochastic process $(x_t)_{t \geq 0}$ on $(\Omega, \mathcal{F}, \mathbb{P})$ is called an Ornstein Uhlenbeck (OU) process if it possesses the following properties:*

- *Stationary: for any $0 < t_0 < t_1 < \dots < t_n$ and $h > 0$, the random variables $(x_{t_1}, x_{t_2}, \dots, x_{t_n})$ and $(x_{t_1+h}, x_{t_2+h}, \dots, x_{t_n+h})$ are identically distributed; that is, time shifts leave joint probabilities unchanged,*
- *Gaussian: for any $0 < t_0 < t_1 < \dots < t_n$, the random variables $(x_{t_1}, x_{t_2}, \dots, x_{t_n})$ are multivariate normally distributed,*
- *Markovian: for any $0 < t_0 < t_1 < \dots < t_n$, $P(x_{t_n} \in A | x_{t_{n-1}}, x_{t_{n-2}}, \dots, x_0) = P(x_{t_n} \in A | x_{t_{n-1}})$, $\forall A \in \mathcal{B}(\mathbb{R}^n)$*
- *Stochastic continuity: $\forall \epsilon > 0$, $\lim_{h \rightarrow 0} P(|x_{t+h} - x_t| \geq \epsilon) = 0$.*

Theorem 2 *A stochastic process is an OU process if it satisfies the following stochastic differential equation:*

$$dx_t = \kappa(m - x_t)dt + \sigma dW_t \quad (2.8)$$

where W_t is a standard Brownian motion, κ is the rate of mean reversion, and m is the long run mean.

Starting from x_0 , the solution of the SDE is given by

$$x_t = e^{-\kappa t}x_0 - (1 - e^{-\kappa t})m + e^{-\kappa t}(2\kappa)^{-\frac{1}{2}}W_{(e^{2\kappa t}-1)}. \quad (2.9)$$

Conditional on x_0 , x_t has a Gaussian marginal distribution. In the one dimensional case, the first and second moments are given below

$$\begin{aligned} E(x_t|x_0) &= e^{-\kappa t}x_0 - (1 - e^{-\kappa t})m, \\ Cov(x_s, x_t|x_0) &= \frac{\sigma^2}{2\kappa}(e^{-\kappa|s-t|} - e^{-\kappa|s+t|}), \end{aligned}$$

and the long term mean (i.e. as $t \rightarrow \infty$) is m .

2.3 Cox-Ingersoll-Ross (CIR) Process

Another widely used mean reverting process, especially in modeling interest rates, is the Cox-Ingersoll-Ross (CIR) Process, or square root diffusion process. It is the underlying process of the well-known Cox-Ingersoll-Ross term structure model (Cox et al 1985[25]).

The CIR process $x_t \in \mathbb{R}$ satisfies the stochastic differential equation

$$dx_t = \kappa(m - x_t)dt + \sigma\sqrt{x_t}dW_t$$

where m , σ , and κ are the parameters. Here κ is the rate of mean reversion, m is the long run mean, and σ corresponds to volatility.

Conditional on a positive initial state x_0 , x_t is distributed as $1/\gamma(t)$ times a non-central chi-square distribution with d degrees of freedom and non-centrality parameter $\lambda(t)$ (Cox et al 1985[25]):

$$x_t/\gamma(t) \sim \chi^2(d, \lambda(t)),$$

where

$$\begin{aligned} d &= 4\kappa m/\sigma^2, \\ \gamma(t) &= \frac{4\kappa}{\sigma^2(1 - e^{-\kappa t})}, \\ \lambda(t) &= x_0\gamma(t)e^{-\kappa t}. \end{aligned}$$

If m , σ , κ are all positive, and $2\kappa m > \sigma^2$, the CIR process is well-defined, always positive.

2.4 Lévy OU Process

In the OU process, the Brownian motion W_t is the *driving force*, and thus called the Background Driving Process. As we discussed, the randomness in empirical financial data most often shows non-Gaussianity. Barndorff-Nielsen and Shephard 2001 [10] introduced the Lévy OU process, which is driven by a more generalized Background Lévy Process (BDLP) z_t . The Lévy OU process is more flexible, and can be used to explain the non-Gaussianity in the observations.

The generalized Lévy OU process is a stochastic process satisfying the SDE

$$dx_t = \kappa(m - x_t)dt + \sigma dz_t \tag{2.10}$$

where z_t is a Lévy process. Given initial value x_0 , The SDE has solution

$$x_t = x_0 e^{-\kappa t} + m(1 - e^{-\kappa t}) + \int_0^t e^{\kappa(s-t)} dz_s. \quad (2.11)$$

The last term is the stochastic integral of the nonrandom function $e^{\kappa(s-t)}$. If z_t has characteristic triplet (γ, σ^2, ν) , then

$$\begin{aligned} \int_0^t e^{s-t} dz_s &= \gamma \int_0^t e^{\kappa(s-t)} ds + \sigma \int_0^t e^{s-t} dW_s \\ &\quad + \int_0^t \int_{|x| \leq 1} e^{\kappa(s-t)} \tilde{J}_z(ds \times dx) \\ &\quad + \int_0^t \int_{|x| > 1} e^{\kappa(s-t)} \tilde{J}_z(ds \times dx), \end{aligned} \quad (2.12)$$

where $J_z(A)$ is the jump measure, which counts the number of jumps with jump size in set A , and $\tilde{J}_z(A)$ is the compensated jump measure. If z is a Brownian motion, x_t is the Gaussian OU process. If z_t is a jump process with finite activity, the jumps of x_t are the same as the jumps of z_t . Between the jumps z_t , x_t decays exponentially due to the linear damping term. This results in a more realistic asymmetric behavior for volatility than in diffusion models: volatility jumps up suddenly but simmers down gradually (Tankov 2003[50]).

Let $\psi(u)$ be the characteristic exponent of the BDLP z_t . The characteristic function of X_t is given by

$$\phi_{x_t}(u) = \exp\{iu[(X_0 - m)e^{-\kappa t} + m] + \int_0^t \psi(ue^{\kappa(s-t)} ds)\}. \quad (2.13)$$

Assuming the characteristic triplet of BDLP (γ, σ^2, ν) defined as above, the distribution of x_t is infinitely divisible for every t and has characteristic triplet $(\gamma_t^y, (\sigma_t^y)^2, \nu_t^y)$, where

$$\begin{aligned}
\gamma_t^y &= \frac{\gamma}{\kappa}(1 - e^{-\kappa t}) + x_0 e^{-\kappa t} \\
(\sigma_t^y)^2 &= \frac{\sigma^2}{2\kappa}(1 - e^{-2\kappa t}) \\
\nu_t^y &= \int_1^{e^{\kappa t}} \nu(\epsilon B) \frac{d\epsilon}{\kappa\epsilon}.
\end{aligned} \tag{2.14}$$

Chapter 3

One-Factor Model

VIX process displays a mean reverting property. We use various mean reverting processes to model the historical movements of the VIX in physical measure. We also try various mean reverting processes to calibrate the VIX option prices in risk neutral measure. From the comparison the performances among those mean reverting processes, we choose candidates for the dynamic multifactor joint model.

3.1 Estimate the VIX Process in Physical Measure

VIX process is mean reverting in the physical measure. We can see this from the historical path in Figure 3.1. Most time the VIX fluctuates around its long-term mean; it jumps high during economic crises. For instance, during the last crisis that began in 2007 the VIX climbs to a peak in late 2008 and then tapers off. We use various mean reverting processes to model the historical movements of the VIX, estimate their parameters and compare their performances.

3.1.1 Gaussian OU

We first assume a Gaussian OU process for the VIX movement and estimate the Gauss OU model. We do ordinary least squares linear regression on daily changes of VIX using the following relationship. We obtain the estimated parameters and



Figure 3.1: Historical VIX path from January 1990 to September 2010.

the regression residuals $\sigma\epsilon_t$

$$V_{t+\Delta t} - V_t = -(1 - e^{-\kappa\Delta t})V_t + m(1 - e^{-\kappa\Delta t}) + \sigma\epsilon_t.$$

The PDF of the residuals is plotted in the top left panel of Figure 3.2. The residuals have a skewness of 0.9104 and excess kurtosis of 18.5009, which is large enough to invalidate the normal assumption in the Gaussian OU model. The Kolmogorov-Smirnov distance (K-S distance) for different models are summarized in Table 3.1. The large K-S distance distance also shows that the BDLP is badly fitted by normal distribution.

3.1.2 CIR

We next test the Cox-Ingersoll-Ross (CIR) model. According to the SDE of the CIR model below,

$$dV_t = -\kappa(V_t - m)dt + \sigma\sqrt{V_t}dW_t. \quad (3.1)$$

We do ordinary least squares linear regression on VIX daily data using

$$\frac{V_{t+\Delta t} - V_t}{\sqrt{V_t}} = \frac{\kappa m \Delta t}{\sqrt{V_t}} - \kappa\sqrt{V_t}\Delta t + \sigma\epsilon_t. \quad (3.2)$$

The residuals $\sigma\epsilon_t$ from the CIR model have a skewness of 1.0113, and excess kurtosis of 10.4275. The large excess kurtosis again casts a doubt on the assumption of normality. The large K-S distance shows that the CIR model does not fit either.

3.1.3 Lévy OU

To fit the VIX data, we introduce non-Gaussianity into the BDLP. The general Lévy OU models employ OU processes driven by non-Gaussian background Lévy processes (BDLP) z_t :

$$V_t = -\kappa(V_t - m)dt + dz_t.$$

We have tried various Lévy OU processes. The BDLP's in CGMOU, CGMYOU are the CGM (Variance Gamma), and the CGMY processes. Barndorff-Nielsen (2001) [10] proposes a positive non-Gaussian OU process to describe the movement of volatility. To test whether a Lévy OU process with only positive jumps can model the VIX process, we test the CMOU and CMYOU, whose BDLP are a Gamma process and a CGMY process with only positive jumps respectively.

To estimate the Lévy OU process, we first obtain V' by subtracting the decay term from the daily VIX value and then standardize V' into \tilde{V} . The density is fitted by \tilde{U} , or standardized \tilde{U} . The quantities V' , \tilde{V} , U and \tilde{U} are given below:

$$\begin{aligned} V' &= V_{t+\Delta t} - e^{-\kappa\Delta t}V_t \\ \tilde{V} &= \frac{S - \text{Mean}(V')}{S(V')}, \\ U &= \int_0^{\Delta t} e^{s-\Delta t} dZ_s \\ \tilde{U} &= \frac{U - E(U)}{\text{Stdv}(U)}, \end{aligned}$$

where $\text{Mean}(\cdot)$, $S(\cdot)$, $E(\cdot)$ and $\text{Stdv}(\cdot)$ stand for the sample mean, sample standard deviation, expectation and the standard deviation respectively. We choose a range from -6 to 6, then divide the range into bins of width 0.1. We run the MLE on the middle point of the bins b_i . The density p_i are given through FFT of the characteristic function of U . We count the frequencies n_i of \tilde{V} fallen within the i th bin. The parameters are estimated by maximizing the sum of Log Likelihood $n_i \log(p_i)$.

3.1.4 Results

The empirical distribution F_n for n *iid* observations X_i is defined as

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n I_{\{X_i \leq x\}},$$

where $I_{\{X_i \leq x\}}$ is the indicator function. The Kolmogorov-Smirnov distance (K-S distance) for a given CDF is defined as the supremum of the distance between the empirical and the hypothesized distribution

$$D_n = \sup_x |F_n(x) - F(x; \hat{\theta})|.$$

Smaller K-S distance indicates goodness of fit. The K-S distance and the estimated parameters of each model were computed for estimated VIX data and are tabulated in Table 3.1.

The estimations are based on the time series of daily VIX data from January 2, 1990 to October 27, 2010. The VIX data are obtained from Yahoo.com.

We can see from the results that neither the Gaussian OU nor the CIR model fits well. They both have large K-S distances. The positive Lévy OU processes, the CMOU and the CMYOU have large K-S distances. This shows that positive jumps are not enough to fit the time series of VIX data and negative jumps need to be included in the model. A large G value in the CGMOU and the CGMYOU model shows that it is necessary to include an occasional negative jump or a less significant diffusion in the background processes. Therefore, we will discard the positive OU processes from now on. The Lévy OU models with VG and CGMY as BDLP give satisfactory results, and CGMYOU does better at the price of adopting an extra parameter.

Figure 3.2 illustrates the K-S distance for different models. The solid lines represent the hypothesized PDFs calculated by MLE, while the circles represent the empirical PDFs.

The computation is done by a MATLAB program running on an AMD Quad-Core processor at 2.30 GHz. The estimation for Gaussian OU and CIR takes less than 1 second, while for the other models, there is no significant difference in the computation time, which is around 100 seconds.

Table 3.1: Result of estimation in physical measure. Parameters are estimated on VIX data from January 2, 1990 to October 27, 2010.

Model	Parameters	K-S distance
Gaussian OU	$\kappa : 0.1155, m : 20.4071, \sigma : 24.1118$	0.1152
CIR	$\kappa : 0.0135, m : 20.4080, \sigma : 4.5745$	0.0822
CGMOU	$\kappa : 0.0368, m : 11.3075, C : 0.8110,$ $G : 1.3961, M : 0.9669$	0.0572
CGMYOU	$\kappa : 0.2401, m : 11.0121, C : 2.3011,$ $G : 9.7808, M : 0.0436, Y : 1.1245$	0.0505
CMOU	$\kappa : 0.5468, m : 10.2103, C : 1.0487,$ $M : 0.0729$	0.1014
CMYOU	$\kappa : 0.5468, m : 10.5798, C : 1.0140,$ $M : 0.4932, Y : -1.7832$	0.0948

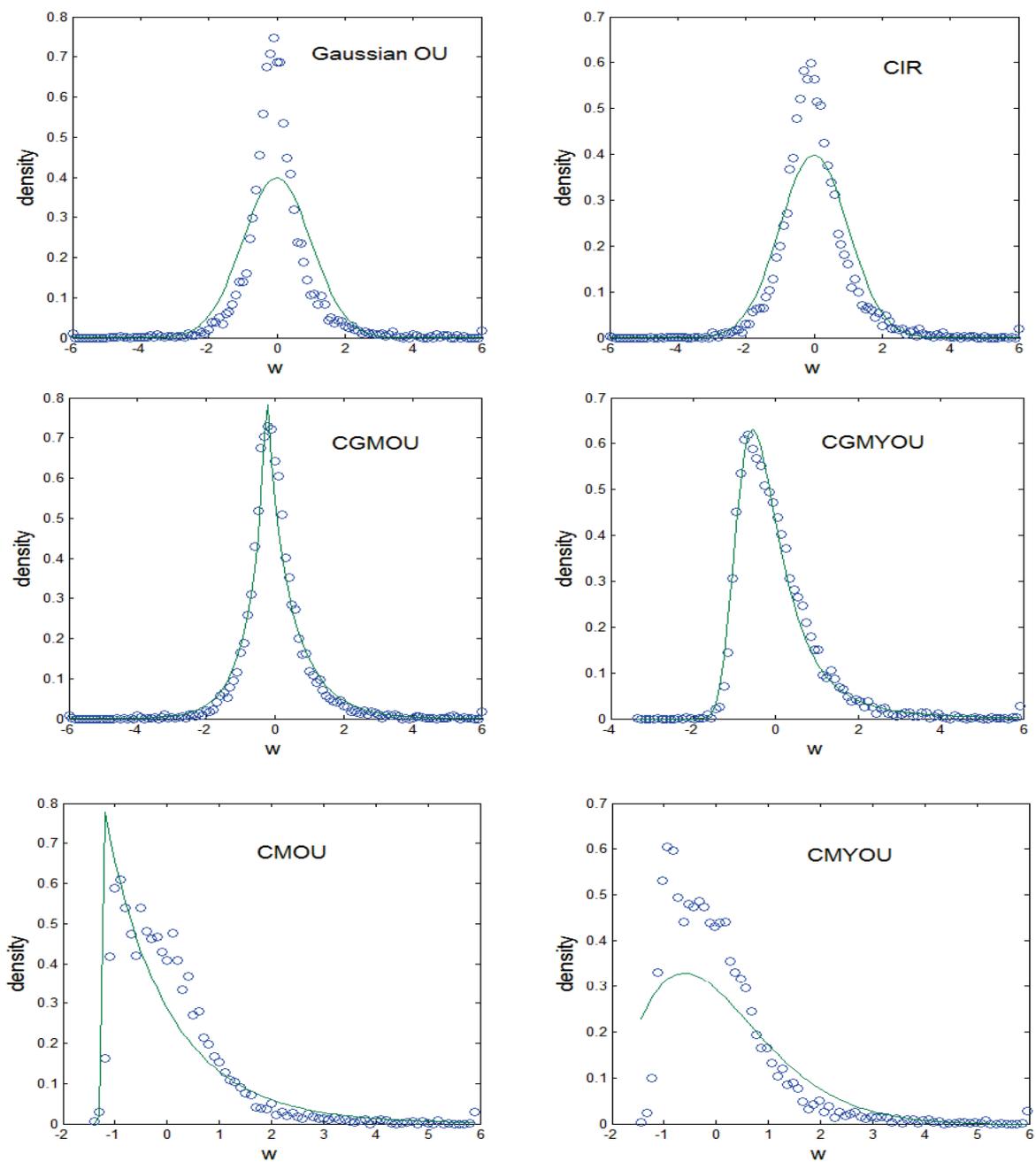


Figure 3.2: Empirical PDF of \tilde{V} and model PDF of \tilde{U} for Gaussian OU, CIR, CGMOU, CGMYOU, CMOU and CMYOU models. After \tilde{V} are binned, the empirical PDF is given by the ratio of the number of points in each bin over the total number of points divided by the bin width.

3.2 Model Calibration using Option Price

We assume a mean reverting stochastic Markovian process model for the VIX V_t at present time t . Let T be the maturity date of the derivative contract, r be the interest rate and $\tau = T - t$. The fair price of a VIX future is the time t risk-neutral expectation of VIX at maturity T :

$$F(t, V_t) = E_t^Q V_T.$$

Similarly, the prices of a European VIX call and put for given strike K can be written as

$$C(t, V_t) = e^{-r\tau} E_t^Q (V_T - K)^+$$

$$P(t, V_t) = e^{-r\tau} E_t^Q (K - V_T)^+.$$

From the VIX option and VIX future prices we can calibrate the risk-neutral measure using various models. The calibration is done by minimizing the sum of squared errors between the historical option prices on a certain day with all maturities and the model predicted prices.

3.2.1 CIR

Grunbichler and Longstaff (1996) [31] derived formulas for volatility options by assuming that the volatility index follows a CIR process under risk neutrality. According to Cox, Ingersoll and Ross (1985) [25], V_t follows a noncentral Chi-square distribution, which leads to the following pricing formula for the call option:

$$\begin{aligned}
C(V_t, K, T) &= e^{-r\tau} [e^{-\kappa\tau} V_t G(\gamma K; \nu + 4, \lambda) \\
&\quad + m(1 - e^{-\kappa\tau}) G(\gamma K; \nu + 2, \lambda) - KG(\gamma K; \nu, \lambda)] \quad (3.3)
\end{aligned}$$

with $G(\cdot; \nu + i\lambda)$ denoting the complementary distribution function for the noncentral Chi-square distribution $\chi^2(\nu + i, \lambda)$ with $\nu + i$ degrees of freedom and noncentral parameter λ . γ, ν, λ are given by

$$\begin{aligned}
\gamma &= \frac{4\kappa}{\sigma^2(1 - e^{-\kappa\tau})} \\
\nu &= \frac{4m\kappa}{\sigma^2} \\
\lambda &= \gamma e^{-\kappa\tau} V_t.
\end{aligned}$$

3.2.2 Lévy OU and Pricing Option Using FFT

When an analytical formula for option price is not available but the characteristic exponent of its BDLP is, FFT can be used to price the option.

Assuming that V_t (VIX) has a characteristic function Φ under the risk neutral measure, the general Fourier transform of a call price is given by Equation (3.5)

$$\Psi_c(u) = \int_0^\infty e^{iuK} C(K) dK \quad (3.4)$$

$$= e^{-r\tau} \left\{ \frac{1}{u^2} [1 - \Phi(u)] + \frac{i}{u} F_\tau \right\}, \quad (3.5)$$

where $u = a + bi$, $a \geq 0$ is a constant, $b \in \mathbb{R}$, and F_τ is the future price with maturity τ . By calculating the inverse Fourier transform of $\Psi_c(u)$ using the Fast Fourier Transform, we can obtain the option price. The derivation of Equation (3.5) is in Appendix A.

However, since the domain of V_t is the positive axis, the FFT result diverges near 0. When V_t approaches 0 or infinity, the truncation error of the numerical integral of Ψ becomes large, and the convergence is slow. The convergence can be dramatically improved by subtracting a pseudo gamma call option from the call option of interest.

We assume $X \sim \text{Gamma}(a, \theta)$, a call option on X price is given by

$$\begin{aligned} C_{\text{gamma}}(K) = E(X - K)^+ &= \int_K^\infty (x - K)x^{a-1}\theta^{-a}e^{-x/\theta}dx \\ &= a\theta G(K, a + 1, \theta) - KG(K, a, \theta), \end{aligned} \quad (3.6)$$

where K is the strike and G is the complementary CDF.

We make the underlying asset price of the pseudo Gamma call has the same mean and variance as the VIX. Then the call option of the VIX is given by

$$C(K) = C_{\text{gamma}}(K) + e^{-r\tau} \text{FFT}^{-1}\left(\frac{1}{u^2}[\Phi_{\text{gamma}}(K) - \Phi(K)]\right). \quad (3.7)$$

Here is an example for an imaginary VIX, which has an Inverse Gaussian (IG) distribution with mean 10 and variance 6. The PDFs of the IG and the Gamma distribution which approximates it are presented in the upper panel of Figure 3.3. The bottom figure gives the call option of the IG distributed VIX using Equation (3.7). We use 512 points in the FFT. The convergence is fast and the curve is smooth.

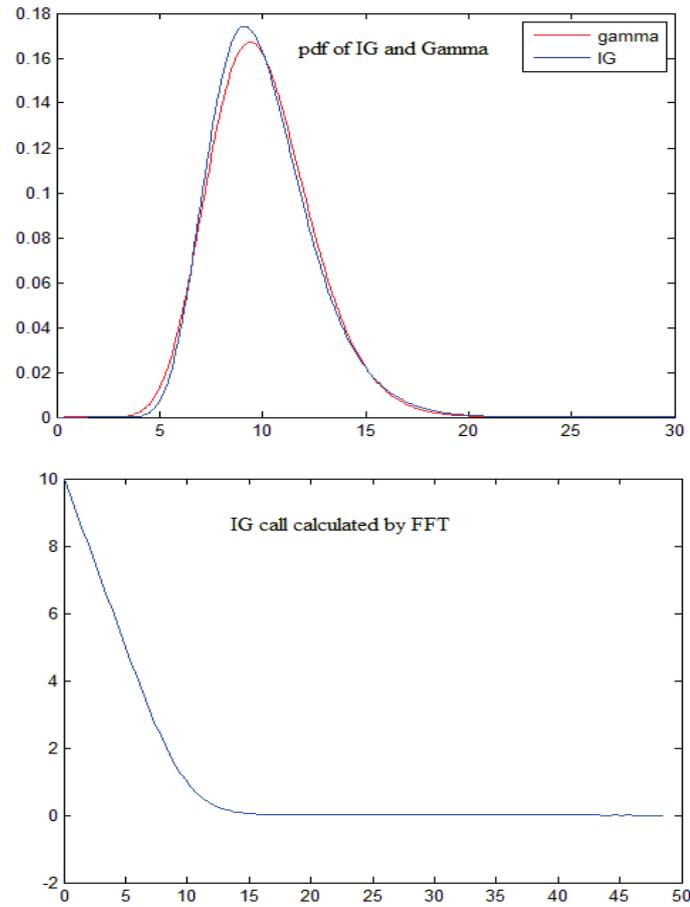


Figure 3.3: Inverse Gaussian distribution with mean 10 and variance 6 and its call calculated by FFT

3.2.3 Numerical Results

We have applied various models to historical option prices at five consecutive Wednesday, February 28, 2006 through March 28, 2006. We compute the following

error statistics to measure the quality of the fits:

$$\begin{aligned}
 APE &= \frac{\sum_{i=1}^N \frac{|A_i - M_i|}{N}}{\sum_{i=1}^N \frac{A_i}{N}}, \\
 AAE &= \frac{\sum_{i=1}^N \frac{|A_i - M_i|}{N}}{N}, \\
 ARPE &= \frac{1}{N} \sum_{i=1}^N \frac{|A_i - M_i|}{A_i}, \\
 RMSE &= \sqrt{\frac{1}{N} \sum_{i=1}^N \frac{(A_i - M_i)^2}{N}},
 \end{aligned}$$

where A_i is the actual value and M_i is the model predicted value. The average error statistics are presented in Table 3.2 below.

Table 3.2: The average error statistics of various models in the estimation of VIX options. The error statistics are computed for five consecutive Wednesdays , February 28, 2006 through March 28, 2006, and then averaged.

Model	Mean	AAE	RMSE	ARPE	APE
OU	-0.4555	0.5034	0.5910	0.6147	0.187
CIR	-0.0704	0.2683	0.3057	0.8841	0.096
CGMOU	-0.0240	0.2261	0.2692	0.2357	0.061
CGMYOU	-0.0233	0.1622	0.2592	0.1746	0.049
CMOU	0.0741	0.2731	0.3729	0.5131	0.098
CMYOU	0.0802	0.2800	0.3166	0.4806	0.110

The comparison shows that the Lévy OU models overall have smaller error

statistics than the CIR and the Gaussian OU models do, while the CGMYOU performs better than the CGMOU. However, given the fact that the CGMYOU has an extra parameter, the difference in their performances are reasonable. The Gauss OU model gives poor results. The comparison of performance between the Gaussian OU model and the CGMOU model is illustrated in Figure 3.4 and Figure 3.5. The two models are calibrated on the option price of March 7, 2006. The figures illustrate the model fitted call price and the historical price at the same maturity of 22 days. The positive Lévy CMYOU model performs worse than those allowing negative jumps, but not significant as its counterpart in the physical measure. This may be because, in the risk-neutral measure, the VIX process has more frequent big positive jumps than negative jumps. This can be explained as the buyers' greater concern about big positive jumps than negative jumps. This phenomenon of more frequent big jumps can be proved from the parameters estimated by CGMOU and CGMYOU through small M and big G , since small M means less damped positive jumps and small G means less damped positive jumps in the Variance Gamma process and the CGMY process. The parameters estimated vary significantly across dates. Thus the one-factor model is not suitable for the cross date estimation. The estimated parameters of CGMOU and their G/M ratios are presented in Table 3.3. The computation for the Gaussian OU and CIR models are fast while the other models need longer time.

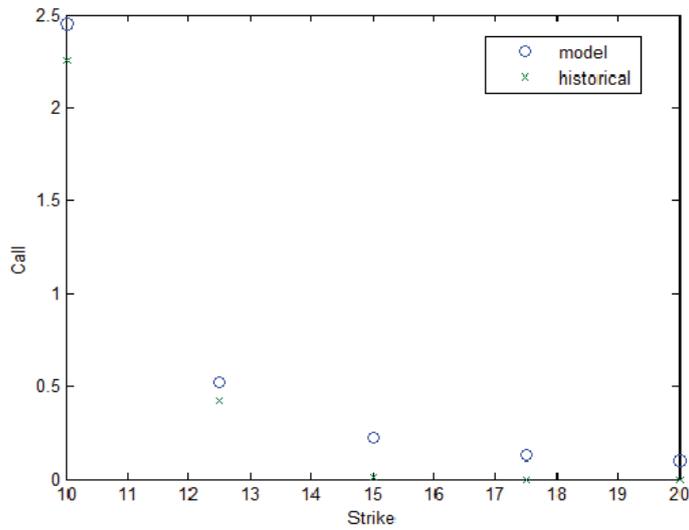


Figure 3.4: The Gaussian OU model fitted call price and the historical price on Feb 28, 2006 with maturity of 78 days

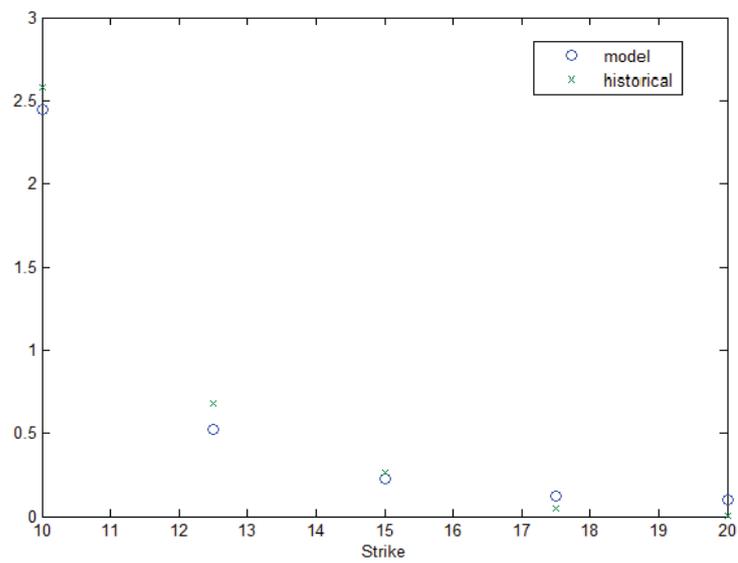


Figure 3.5: The CGMOU model fitted call price and the historical price on Feb 28, 2006 with maturity of 78 days

Table 3.3: The estimated parameters and the G/M ratios using CGMOU model.

The parameters are estimated from VIX options for five consecutive Wednesdays.

date	κ	m	C	G	M	G/M
Feb 28, 2006	18.5414	1.9698	226.6668	8.1209	0.9607	8.45
Mar 7, 2006	26.8292	0.1203	416.6761	22.1531	1.1445	19.36
Mar 14, 2006	52.2444	2.0948	906.7771	5.1851	1.2233	4.24
Mar 21, 2006	19.5233	5.8237	254.0834	32.7185	1.4621	22.38
Mar 28, 2006	8.9324	0.1160	226.3220	5.4397	1.2856	4.23

3.3 Conclusions

From the result of the tests for various mean reverting models both in physical measure and risk neutral measure, we make the following conclusions

1. The driving processes in VIX are better modeled as jump processes than Brownian motion both in the physical measure and the risk-neutral measure.
2. The Lévy OU process with only positive jumps does not fit the VIX process in physical measure. It is better to include negative jumps in the Lévy OU process.
3. In the risk neutral measure, for the Lévy OU processes with jumps, there are more frequent positive jumps than negative jumps.

4. Both in the physical measure and the risk neutral measure, the CGMOU and CGMYOU models perform better than other models. The CGMYOU model performs slightly better than the CGMOU at the cost of computing time.
5. The parameters estimate from VIX option calibration vary significantly across time. The one-factor model is not suitable for the cross date estimation

Given the facts that CGMOU and CGMYOU performs both in physical measure and risk neutral measure, that CGMYOU need longer computational time, and that one-factor model is not suitable for the cross date estimation. in later joint estimation of VIX and VIX derivatives, we choose multifactor CGMOU model as the multifactor Lévy OU model.

Chapter 4

Filter Methods

4.1 Filter Problem

Since 1960, when R.E. Kalman [38] published his famous paper describing a recursive solution to the discrete data linear filtering problem, filter methods have been the subject of extensive research and application. More filter methods were devised to deal with more general filtering problems with non-Gaussian noise and non-linear systems. Many filter methods found their applications in quantitative finance.

Filter methods recursively estimate the latent state of a dynamic system from a series of incomplete or noisy measurements. Let x_k be the states, y_k be the measurement at time k and $y_{1:k}$ be the sequence of measurement observations up to time k . We have the discrete dynamic system as follows:

$$x_k = f_k(x_{k-1}) + \nu_k \quad (4.1)$$

$$y_k = h_k(x_k) + w_k, \quad (4.2)$$

where f_k is the propagation function perturbed by an uncertainty, h_k is the measurement function, ν_k is the process uncertainty and w_k is the measurement noise. To further simplify the model, we assume the process x_t is Markovian and w_k is pure white noise.

The x_k 's are random variables, and the goal of the filter method is to find the posterior distribution $p(x_k|y_{1:k})$ of x_k given $y_{1:k}$. We assume the initial density $p(x_0) \triangleq p(x_0|y_0)$ is given and that we reach the distribution $p(x_{k-1}|y_{1:k-1})$ at time $k-1$. The estimation consists of two steps. First propagate the distribution to the time k , obtaining the prior $p(x_k|x_{k-1}, y_{1:k-1})$, and then invoke Bayes' theorem using new acquired y_k to update the distribution of x_k , obtaining the posterior $p(x_k|y_{1:k})$.

- **Time Update** According to the theorem of total probability, the propagated distribution is given by Equation (4.3):

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

The second equality holds because of the Markov property, namely the information of $y_{1:k-1}$ has been included in x_{k-1} . If the uncertainty ν_k is normal and the propagating functions f_k and measurement functions h_k are linear, $p(x_k|y_{1:k-1})$ and $p(x_k|y_{1:k})$ are normal too. Only the propagation of mean and variance is needed and their analytic forms are available. This gives the Kalman filter method. Otherwise, the storage of the entire PDF in Equation (4.3) is equivalent to an infinite dimensional vector, and thus to obtain a tractable solution becomes impossible. If ν_k is normal, and f_k is nonlinear, we can use a set of points called sigma points to approximate the density of $p(x_{k-1}|y_{1:k-1})$. We propagate the sigma points instead of the density and use the propagated points to approximate the posterior density. This is the basic idea behind the unscented Kalman filter. If ν_k is not normal, and f_k is

nonlinear, the only general approach is to apply Monte-Carlo sampling techniques that essentially convert integrals to finite sums, which converge to the true solution at the limit. The particle filter discussed in a later section is an example of such an approach.

- **Measurement Update** After the observation y_k becomes available, Bayes' theorem is applied to update the distribution of x_k

$$p(x_k|y_{1:k}) = \frac{p(y_k|x_k)p(x_k|y_{1:k-1})}{\int p(y_k|x_k)p(x_k|y_{1:k-1})dx_k}. \quad (4.4)$$

For reasons similar to those described above, it is impossible to apply Equation (4.4) over the whole density. The Kalman filter applies when the measurement noise w_k is normal and measurement function h_k is linear. The UKF applies when w_k is normal and h_k is nonlinear. Otherwise we have to resort to a Monte Carlo method like the Particle Filter.

4.2 Kalman Filter

The Kalman Filter assumes the progress function f_k and the measurement function h_k are linear, and the noises ν and w are Gaussian.

- f_k, h_k are linear, $f_k(x) = F_k x, h_k(x) = H_k x;$
- ν_k and w_k are normal with variance Q_k and R_k .

Under the assumption above, it is easy to show that the probability $P(x_k|y_{1:k-1})$ and $P(x_k|y_{1:k})$ are normally distributed. Thus only the mean and covariance are needed

to recover the distribution of x_k . Let $\hat{x}_{k|k-1}$ and $P_{k|k-1}$ denote the prior estimate of the mean and covariance, and $\hat{x}_{k|k}$ and $P_{k|k}$ denote the posterior estimate of the mean and covariance after the newest update. The Kalman filter algorithm is carried out as follows:

$$\begin{aligned}\hat{x}_{k|k-1} &= F_k \hat{x}_{k-1|k-1} \\ P_{k|k-1} &= Q_{k-1} + F_k P_{k-1|k-1} F_k^T \\ \hat{x}_{k|k} &= \hat{x}_{k|k-1} + K_k (y_k - H_k \hat{x}_{k|k-1}) \\ P_{k|k} &= P_{k|k-1} - K_k S_k K_k^T\end{aligned}$$

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 called the Kalman gain. The last equation shows that the Kalman Filter helps reduce the covariance of the state 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 y_k through the Kalman gain.

4.3 Unscented Kalman Filter

The Kalman Filter is the only optimal solution under the aforementioned assumptions. However, in cases where the measurement functions are nonlinear, suboptimal solutions such as the Extended Kalman Filter (EKF) and the Unscented Kalman Filter (UKF) are needed. The Extended Kalman filter (EKF) applies Taylor series to approximate the nonlinear f_k , while the unscented Kalman (UKF) uses the unscented transformation (UT)(Julier 1997 [37]). The latter gives a derivative-

free alternative to the EKF and provides superior performance with an equivalent computational complexity (Wan 2001[53]).

4.3.1 Unscented Transformation

The Unscented Transformation (UT) calculates the statistics of a random variable undergoing a nonlinear transformation by propagating a set of sigma points. Let state x have a dimension d , mean μ_x and covariance P_x , where x undergoes a nonlinear transformation $y = f(x)$. To calculate the statistics of y , we first choose a set of sigma points $X = (x_i, W_i), i = 1, \dots, 2d + 1$ so that their mean and covariance match μ_x and P_x ,

$$\begin{aligned} X_0 &= \mu_x \\ X_i &= \mu_x + (\sqrt{(d + \lambda)P_x})_i, \quad i = 1, \dots, d \\ X_i &= \mu_x - (\sqrt{(d + \lambda)P_x})_i, \quad i = d + 1, \dots, 2d. \end{aligned}$$

where $\lambda = \alpha^2(d + \kappa) - d$ is a scaling parameter. The constant α determines the spread of the sigma points around the mean, κ and β are parameters to refine the prior distribution of x , and $(\sqrt{(d + \lambda)P_x})_i$ is the i th column of the matrix square root e.g., lower triangular Cholesky factorization) (Julier 1997[37]).

Propagating the sigma points X through $f(\cdot)$, we get the transformed sigma points. Then we approximate the mean μ_y and covariance P_y using a weighted

sample mean and covariance of the sigma points

$$\begin{aligned}
 Y_i &= f(X) \\
 \mu_y &\approx \sum_{i=0}^{2L} W_i^m Y_i \\
 P_y &\approx \sum_{i=0}^{2L} W_i^c (Y_i - \mu_y)(Y_i - \mu_y)^T.
 \end{aligned}$$

with weights W_i given by

$$\begin{aligned}
 W_0^m &= \lambda/(d + \lambda) \\
 W_0^c &= \lambda/(d + \lambda) + (1 - \alpha^2 + \beta) \\
 W_i^m &= W_i^c = 1/[2(d + \lambda)] \quad i = 1, \dots, 2d,
 \end{aligned}$$

The unscented approximations are accurate to the third order for Gaussian input for all nonlinearities. The constant κ is usually set to 0 or $3 - d$, and β is set to 2 for Gaussian. The Unscented Kalman Filter generally applies to Gaussian state variables. For non-Gaussian inputs, the skewness and kurtosis are needed beforehand to adjust the parameters α , β and κ . The approximations are accurate to at least the second order, with the accuracy of third and higher order moments determined by the choice of α and β (Wan 2001 [53]). In the multidimensional non-Gaussian case, all states have the same skewness and kurtosis, which may not be desirable.

4.3.2 Algorithm

The unscented Kalman Filter (UKF) recursively executes the unscented transformation. The algorithm initializes by choosing the sigma points according to initial

mean and covariance. The sigma points propagate through the nonlinear transformation, yielding a new set of points. The new estimated mean and covariance are then computed from the transformed point set and updated by the observation of y by Bayes' theorem. A set of new sigma points is then generated based on the updated mean and covariance.

Algorithm 1 (Unscented Kalman Filter)

Initialize with

$$\hat{x}_0 = E[x_0] \text{ and } P_0 = Cov[x_0]$$

For $k \in 1, \dots, T$

Calculate sigma points $X_{k-1} = [\hat{x}_{k-1}, \hat{x}_{k-1} + \gamma\sqrt{P_{k-1}}, \hat{x}_{k-1} - \gamma\sqrt{P_{k-1}} \]$

- **Time update:**

$$X_{k|k-1}^* = f_k(X_{k-1})$$

$$\hat{x}_k^- = \sum_0^{2d} W_i^m X_{k|k-1}^*$$

$$P_k^- = \sum_0^{2d} W_i^c [X_{i,k|k-1}^* - \hat{x}_k^-][X_{i,k|k-1}^* - \hat{x}_k^-]^T + Q$$

$$X_{k|k-1} = [\hat{x}_k^-, \hat{x}_k^- + \gamma\sqrt{P_k^-}, \hat{x}_k^- - \gamma\sqrt{P_k^-} \]$$

$$Y_{k|k-1} = h_k(X_{k|k-1})$$

$$\hat{y}_k^- = \sum_0^{2d} W_i^m Y_{k|k-1}$$

- **Measurement update:**

$$P_{y_k, y_k}^- = \sum_0^{2d} W_i^c [Y_{i,k|k-1} - \hat{y}_k^-][Y_{i,k|k-1} - \hat{y}_k^-]^T$$

$$P_{x_k, y_k}^- = \sum_0^{2d} W_i^c [X_{i,k|k-1} - \hat{x}_k^-][X_{i,k|k-1} - \hat{x}_k^-]^T$$

$$K_k = P_{x_k, y_k}^- * P_{y_k, y_k}^{-1}$$

$$\hat{x}_k = \hat{x}_k^- + K_k(y_k - \hat{y}_k^-)$$

$$P_k = P_k^- - K_k P_{y_k, y_k}^- K_k^T$$

where $\gamma = \sqrt{d + \lambda}$, λ is a composite scaling parameter.

4.3.2.1 Constrained UKF

In practical estimation problems, the states are often bounded or have constraints. For example, the interest rate is non-negative. However, the constraints are handled neither by the Kalman Filter nor by the UKF. The simplest way to incorporate the constraints in the KF is projecting the unconstrained KF estimates onto the boundary of the feasible region at each time step (See Simon 2002 [49], Ungarala 2007 [52]). The constraints information can be incorporated in the UKF algorithm in a simple way during the time-update step. After the propagation, the unconstrained transformed UKF estimate outside the feasible region are projected onto the boundary of the feasible region and we continue with the further steps. Let W_k be the weighting matrix at time k , $\hat{x}_{k|k}$ the unconstrained estimates and C be the feasible region. The projected estimates $\hat{x}_{k|k}^P$ are given by

$$\hat{x}_{k|k}^P = \arg \min_{x_k \in C} (x_k - \hat{x}_{k|k})^T W_k^{-1} (x_k - \hat{x}_{k|k}).$$

Figure 4.1 illustrates the estimation with state constraints. At $t = k$ the three unconstrained estimates which are outside the feasible region are projected onto the boundary. The mean and covariance of the constrained sigma points now represent the a priori UKF estimate, and they are further updated in the update step (Kandepu 2008 [39]).

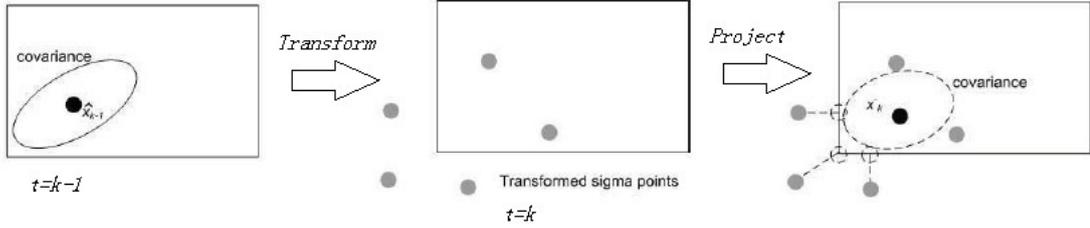


Figure 4.1: Illustration of estimation with state constraints

4.3.3 Mixed-Gaussian UKF

Lemma 1 (Anderson 1979 [2]) *Any density $p(x)$ associated with an n dimensional vector x can be approximated as closely as desired by a weighted combination of Gaussian densities of the form*

$$p_A(x) = \sum_{i=1}^N a_i \mathcal{N}(x; \nu_i, \Sigma_i)$$

for some integer N , and positive scalars a_i with $\sum_{i=1}^N a_i = 1$. Here $\mathcal{N}(\cdot)$ is normal PDF.

It can be shown that the density $p_A(x)$ converges uniformly to any density function of practical interest by letting the number of terms increase and each elemental covariance approach to zero (Anderson 1979 [2]). According to the lemma above, non-Gaussian noise densities can be approximated empirically by Gaussian sums. For a state-space model, it is possible to obtain both the predicted and posterior densities as Gaussian sums. The Unscented Kalman Filter generally only applies to the Gaussian state variables, especially in the multidimensional case. The mixed Gaussian model provides an approach to solve the non-Gaussian state estimation

problem. In this section, we assume that the progress noise ν is non-Gaussian and use a Gaussian mixture to approximate its density,

$$p(\nu_k) = \sum_{i=1}^J \omega_k^i \mathcal{N}(\nu_k; \mu_k^i, Q_k^i)$$

where $\sum_{i=1}^J \omega_k^i = 1$. It is easily extended to the cases in which the measurement noise is non-Gaussian too.

The sigma points $X_i, i = 1 : I$ at time $k-1$ are time-updated and measurement-updated based upon the mixture parameters $\{\omega_{k-1}^j, \mu_{k-1}^j, P_{k-1}^j\}_{j=1}^J$, and then the mixture weights are updated according to Bayes' theorem. At each step, each sigma point generates J new points due to J settings of the mixture parameters, so the size of the sigma points $(2d+1)J^k$ increases geometrically and so does the computational complexity. A clipping technique is applied to reduce the size of the sigma points. When the weight is lower than a preset threshold, the point is discarded (Arasaratnam 2007 [4]).

Algorithm 2 (Mixed-Gaussian UKF)

Initialize with: Choose \hat{x}_0 , ω_0^j , μ_0^j and Q_0^j according to the initial distribution

$p(x_0)$. **Generate sigma points** $X_0 = \bigcup_{j=1}^J [\hat{x}_0 + \mu_0^j + \gamma\sqrt{P_0^j}, \hat{x}_0 + \mu_0^j - \gamma\sqrt{P_0^j}]$

Assume there are n sigma points at time $k - 1$, $X_{k-1} = \{x_i, W_i\}_{i=1}^n$

Time update and Measurement update

Time update and measurement update X_{k-1} for each parameter set $\{\omega_k^j, \mu_k^j, Q_k^j\}$ and obtain sigma set X_k with $n \times J$ points according to UKF algorithm.

Weight update

$$W_{i,j} = \frac{\omega_k^j W_i \mathcal{N}(y_k; \bar{y}_k^j, P_{y_k, y_k}^j)}{\sum_{i,j} \omega_k^j W_i \mathcal{N}(y_k; \bar{y}_k^j, P_{y_k, y_k}^j)}$$

Clip Discard the points if $W_{i,j} < threshold$

Renormalize the weights.

4.4 Particle Filter

Particle filter (PF) provides an alternative algorithm to solve the non-Gaussian state estimation problem. Instead of updating only the mean and covariance, PF uses an ensemble of particles to empirically approximate the posterior distribution of interest. To approximate a continuous distribution by a finite number of particles is sub-optimal. We can increase the precision by using more particles at the cost of computational complexity. Given that computing power has improved rapidly in recent years, the particle filter technique offers great flexibility in estimating state-space models without restricted assumptions [57].

4.4.1 Importance Sampling

Let $\{X_k^i, W_k^i\}_{i=1}^{N_s}$ denote the particle set (and associated weights) that approximate the posterior density $p(x_k|y_{1:k})$

$$p(x_k|y_{1:k}) \sim \sum_{i=1}^{N_s} W_k^i \delta(x_k - X_k^i), \quad (4.5)$$

where $\delta(\cdot)$ is the delta function.

However, the posterior density $p(x_k|y_{1:k})$ is unknown. The particles $X_{i=1}^{N_s}$ are drawn from a proposal density $q(x)$ called the importance density. The principle of choosing the density and weights is called Importance Sampling [11][27]. The weights are defined by Equation (4.6):

$$W_k^{*i} \propto \frac{p(X_k^i|y_{1:k})}{q(X_k^i|y_{1:k})}. \quad (4.6)$$

Again, since the density $p(x_k|y_{1:k})$ is unknown, we cannot rely on Equation(4.6) to compute the weights W_k^{*i} . We first choose a suitable importance density, and arbitrary weights (usually equally weighted) to draw the particles, and then use Bayes' theorem to update the weights. Here we use W_k^i to denote the computed weight

$$\begin{aligned} W_k^i &\propto \frac{p(y_k|X_k^i)p(X_k^i|X_{k-1}^i)p(X_{k-1}^i|y_{1:k-1})}{q(X_k^i|X_{k-1}^i, y_{1:k})q(X_{k-1}^i|y_{1:k-1})} \\ &= W_{k-1}^i \frac{p(y_k|X_k^i)p(X_k^i|X_{k-1}^i)}{q(X_k^i|X_{k-1}^i, y_{1:k})}. \end{aligned}$$

The last equation holds because of the Markovian assumption. The importance density $q(\cdot)$ is chosen to minimize the variance of W_k^{*i} , the true weight. The greater the variance of the weights is, the more easily the particles run into degeneracy. This degeneracy implies that a large computational effort is devoted to updating

particles whose contribution to the approximation is almost zero. A suitable measure of degeneracy of the algorithm is the effective sample size introduced in Bergman 1999 [11] and defined as

$$N_{eff} = \frac{N_s}{1 + Var(W_k^{*i})}.$$

The quantity N_{eff} cannot be evaluated exactly; in practice, it is usually estimated by

$$\hat{N}_{eff} = \frac{1}{\sum_{i=1}^{N_s} (W_k^i)^2}. \quad (4.7)$$

Notice that $N_{eff} \leq N_s$, and small N_{eff} indicates severe degeneracy. In the Particle Filter, an often-used criteria is to maximize the estimated effective sample size. There are two approaches to reducing degeneracy: wise choice of importance density and use of resampling. The optimal choice of importance density may be either impossible or the computational cost may be too high. It is often convenient to choose the importance density to be the prior (Arulampalam 2002 [5])

$$q(x_k | X_{k-1}^i, y_k) = p(x_k | X_{k-1}^i).$$

Then Equation (4.7) becomes

$$W_k^i \propto W_{k-1}^i p(y_k | X_k^i). \quad (4.8)$$

Sequential Importance Sampling (SIS) is a simple algorithm for Particle Filter. We first draw particles from the prior distribution, propagate the particles through the dynamic system and finally evaluate the posterior weights according to Equation (4.8). We repeat the process through the time span. An unavoidable phe-

nomenon is that after a few iterations, all but one or two particles will have negligible weight (Doucet 1998[27]). This phenomenon is called degeneracy. Whenever a significant degeneracy is observed (i.e., when the variance of W_k^i falls below some threshold), the particles have to be resampled to avoid degeneracy. The resampling procedure is to split particles that carry large weights evenly into small particles, so technically it is to map a discrete measure $\{X_k^i, W_k^i\}$ into another discrete measure $\{X_k^{i*}, \frac{1}{N_s}\}$ of equal weights [57].

Algorithm 3 (Particle Filter)

$$[\{X_k^i, W_k^i\}_{i=1}^{N_s}] = PF[\{X_{k-1}^i, W_{k-1}^i\}_{i=1}^{N_s}, y_k]$$

- For $i = 1 : N_s$
 - Draw $X_k^i \sim q(X_k | X_{k-1}^i, y_k) = p(X_k^i | X_{k-1}^i)$
 - Calculate the weight

$$W_k^i = \frac{W_{k-1}^i p(y_k | X_k^i)}{\sum_{j=1}^{N_s} W_{k-1}^j p(y_k | X_k^j)} \text{ End for}$$

- Calculate \hat{N}_{eff}
- if $\hat{N}_{eff} < threshold$
 - Resample using

$$[\{X_k^i, W_k^i\}_{i=1}^{N_s}] = Resample[\{X_{k-1}^i, W_{k-1}^i\}_{i=1}^{N_s}]$$

Algorithm 4 (Resampling)

$$[\{X_k^i, W_k^i\}_{i=1}^{N_s}] = \text{Resample}[\{X_k^i, W_k^i\}_{i=1}^{N_s}]$$

- Initialize $c_1 = W_k^1$
- For $i = 2 : N_s$ - $c_i = c_{i-1} + W_k^i$
- End For Draw random number $n1 \sim \text{Unif}[0; 1/N]$, set $i = 1$
- For $i = 2 : N_s$
 - Let $u_j = u_1 + \frac{j-1}{N_s}$
 - While $u_j > c_i$
 - $i = i + 1$
 - Assign $X_k^{i*} = X_k^i$ and $W_k^i = \frac{1}{N_s}$ End for

Chapter 5

Multifactor Model

In interest rate modeling, there is an extensive literature on dynamic term structure models that estimates models on an entire time series of market prices of discount bonds and interest rate derivatives. As shown in Chapter 3, one-factor models have proved less effective at fitting the VIX option data across dates. This setup is in line with the conclusion of Heidari and Wu 2003 [32] which showed that one-factor models give poor result and three factors in the dynamics are adequate to explain more than 99% of the yield curve data historically. In this Chapter, we use multi-factor models to build a joint framework for consistently describing both VIX historical data and the VIX derivative data.

5.1 Data

Consistent records of time series data of the VIX index and its derivatives are available. We obtain the VIX historical data from Yahoo finance, VIX future data from COBE, and VIX option data from Optionmatrix. The VIX and VIX futures begin from March 30, 2004, VIX options begin from February 28, 2006 and all data end on October 28, 2009. To avoid weekday effects in the estimation, we only sample data weekly every Wednesday. The settle price is taken as the future price, and the average of ask and bid is taken as the option price. The future data include the

prices at different maturities for each date, and the option data include the option prices from different maturities and different strikes. We use the data from March 30, 2004 to December 16, 2008 as in-sample data to estimate model parameters. We test the model's out-of-sample performance on period from December 23, 2008 to October 27, 2009. If the model is well specified, we would expect the models out-of-sample performance to be similar to its in-sample performance.

5.2 The Basic Model Structure

We fix two filtered complete probability spaces with finite fixed time span satisfying the usual technical conditions (right continuity and P-completeness). The first space has the physical probability measure \mathbb{P} and the second has the risk-neutral probability measure \mathbb{Q} . Let x_t denote a vector Markov process in some state space $S_m \subset \mathbb{R}^m$, where m is the dimension of x_t . Assume the VIX process V_t is a linear function of x_t :

$$V_t = a + bx_t. \quad (5.1)$$

As x_t moves along the discrete time points according to the setting of the model, we have

$$x_k = f(x_{k-1}) + \nu(k), \quad (5.2)$$

where $f(x_{k-1})$ is the drift term or mean reverting trend, and $\nu(k)$ is the uncertainty due to the randomness of x_t process.

The VIX derivatives are priced under the risk-neutral measure \mathbb{Q} .

$$\begin{aligned}
F(t, V_t) &= E_t^{\mathbb{Q}} V_T \\
C(t, V_t) &= e^{-r\tau} E_t^{\mathbb{Q}} (V_T - K)^+ \\
P(t, V_t) &= e^{-r\tau} E_t^{\mathbb{Q}} (K - V_T)^+.
\end{aligned} \tag{5.3}$$

These prices are model predicted prices. We summarize the predicted level of VIX and the prices of the VIX derivatives at time k as the vector $\tilde{y}_t = h(x_t)$, and denote the observed value as y_t . We assume y_k is the sum of \tilde{y}_t and a noise term w_k as shown below:

$$y_k = h(x_k) + w(k). \tag{5.4}$$

Given the propagation Equation (5.2) and measurement Equation (5.4), we recursively employ appropriate filter methods to time-update and measurement-update the state, to calculate the model prediction \tilde{y}_t at each time step and the log likelihood defined in Equation (1.4). We estimate the parameters of the model by maximizing the sum of the log likelihood over all time steps .

5.3 GOU3 Model with UKF

We assume that the VIX process V_t is a linear combination of three independent factors $x_t = [x_{1t}, x_{2t}, x_{3t}]$

$$dV_t = a + bx_t = a + \sum_{i=1}^3 b_i x_{it} \tag{5.5}$$

where x_t is a three dimensional Gaussian OU process, governed by the following the SDE in the physical measure:

$$dx_t = -\kappa x_t dt + dW_t, \quad (5.6)$$

where W_t is a standard 3-dimensional Brownian motion and κ , a diagonal matrix, controls the mean-reverting property of the state process x_t . We normalize the state vector x_t to have zero long-run means and identity volatility matrices. We further assume an affine market price of risk

$$r(x_t) = c + f x_t \quad (5.7)$$

with $c \in \mathbb{R}^3$ and diagonal $f \in \mathbb{R}^{3 \times 3}$. Given the market price of risk specifications, x_t remains a Gaussian OU under the risk-neutral \mathbb{Q} , but with adjustments to the drift terms:

$$dx_t = -(c + \hat{\kappa} x_t) dt + dW_t^{\mathbb{Q}}, \quad (5.8)$$

with $\hat{\kappa} = f + \kappa$. Under this setting, in the risk-neutral measure conditional on x_0 at $t = 0$, x_t has a Gaussian distribution with mean μ and variance Σ . The i th element of μ and Σ are given by

$$\mu_i = E^{\mathbb{Q}} x_{it} = e^{-\hat{\kappa}_i t} x_0 - \frac{c_i}{\hat{\kappa}_i} (1 - e^{-\hat{\kappa}_i t}) \quad (5.9)$$

$$\Sigma_i = Var^{\mathbb{Q}} x_{it} = \frac{1}{2\hat{\kappa}_i} (1 - e^{-2\hat{\kappa}_i t}). \quad (5.10)$$

Here, κ_i is the i th diagonal element of $\hat{\kappa}$.

5.3.1 VIX Futures and Options

Let the present time be 0 and let the maturity be t . Under the model assumption, we can easily get the VIX future price,

$$F(0, V_0) = E^Q V_t = a + b\mu, \quad (5.11)$$

with μ given by Equation (5.9) and the price of a VIX call

$$C(0, V_0) = e^{-rt} E^Q (V_t - K)^+ \quad (5.12)$$

$$= e^{-rt} \sigma \{f_n(h) - h(1 - \mathbb{N}(h))\} \quad (5.13)$$

where $Var V_t = b^T \Sigma b$, $h = (K - \mu)/\sigma$, and $f_n(\cdot)$, $\mathbb{N}(\cdot)$ are the PDF, CDF of standard normal distribution. The price of the put option can be derived from the put-call parity

$$P(t, V_t) = C + e^{-r\tau} (K - F(0, V_0)). \quad (5.14)$$

5.3.2 Parameter Estimation

Since propagation noise ν is Gaussian in the Gaussian OU model, we apply the Unscented Kalman Filter (UKF) to estimate the parameters. We use an Euler approximation of the SDE Equation (5.6) as propagation equation. At time t

$$x_t = x_{t-1} e^{-\kappa \Delta t} + I \sqrt{\Delta t} \epsilon_t, \quad (5.15)$$

where $\Delta t = 1/52$ is the weekly discrete time interval, I is the identity matrix, $I \Delta t$ denotes the instantaneous covariance matrix and ϵ_t denotes an independent and identically distributed sequence of trivariate standard normal vectors. The model

predictions \tilde{y}_t are computed as follows: V_t are predicted by the linear combination of the mean of the factor x_t ,

$$V_t = a + \sum_{i=1}^3 b_i e^{-\kappa_i \Delta t} x_{t-1}, \quad (5.16)$$

and the prices of VIX futures, calls and puts are predicted by Equation (5.11), Equation (5.13) and Equation (5.14). The corresponding observation y_t are the historical VIX and VIX derivative data. We further assume that the forecasting errors on the measurement series are normally distributed and define the weekly log likelihood function (ignoring the constant term) as

$$l_t(y_t; \Theta) = -\frac{1}{2}(y_t - \tilde{y}_t)^T (A_t)^{-1} (y_t - \tilde{y}_t), \quad (5.17)$$

where A_t denotes the conditional covariance matrix of the forecasts of the measurement series. For convenience we take A_t equal to identity matrix I . The model parameters are estimated by maximizing the sum of log likelihoods defined in Equation (5.17) over all time steps $t = 1, \dots, T$:

$$\hat{\Theta} = \arg \max_{\Theta} \sum_{t=1}^T l_t(y_t; \Theta). \quad (5.18)$$

For the GOU3 model, the Parameters to be estimated and the results are a , b , c , κ , $\hat{\kappa}$, x_0 , total of 15 numbers, where x_0 is the initial state. The parameters estimated by MLE are tabulated in Table 5.1.

Table 5.1: In-sample maximum likelihood parameter estimates of GOU3 model with Unscented Kalman filter from March 30, 2004 to December 16, 2008

a	b	c	x_0
0.1100	$\begin{bmatrix} 9.9435 \\ 13.5958 \\ 8.5187 \end{bmatrix}$	$\begin{bmatrix} -0.0728 \\ -2.0197 \\ -0.7863 \end{bmatrix}$	$\begin{bmatrix} -0.2432 \\ 0.3370 \\ 0.6664 \end{bmatrix}$
κ			$\hat{\kappa}$
$\begin{bmatrix} 1.6506 & 0 & 0 \\ 0 & 0.1403 & 0 \\ 0 & 0 & 0.0016 \end{bmatrix}$	$\begin{bmatrix} 34.0849 & 0 & 0 \\ 0 & 5.1299 & 0 \\ 0 & 0 & 0.5141 \end{bmatrix}$		

5.3.3 In-Sample Test

The error statistics between model predictions \tilde{y}_t and historical data y_t from March 30, 2004 to December 16, 2008 are summarized in Table 5.2.

Figure 5.1 compares the historical VIX value and the GOU3 model predicted value, Figure 5.2 compares the historical VIX future price and the GOU3 model predicted price, and Figure 5.3 compares the historical VIX option price and the GOU3 model predicted price. The error statistics for the test are summarized in Table 5.2.

Table 5.2: In-sample error statistics (GOU3 model) for VIX and VIX derivatives from March 30, 2004 to December 16, 2008

	Mean	AAE	RMSE	ARPE	APE
VIX	0.5478	2.2878	3.6650	0.1014	0.1085
VIX Futures	0.0631	1.0757	1.62469	0.0502	0.0515
VIX Options	0.2120	0.8605	1.5053	0.8219	0.1588

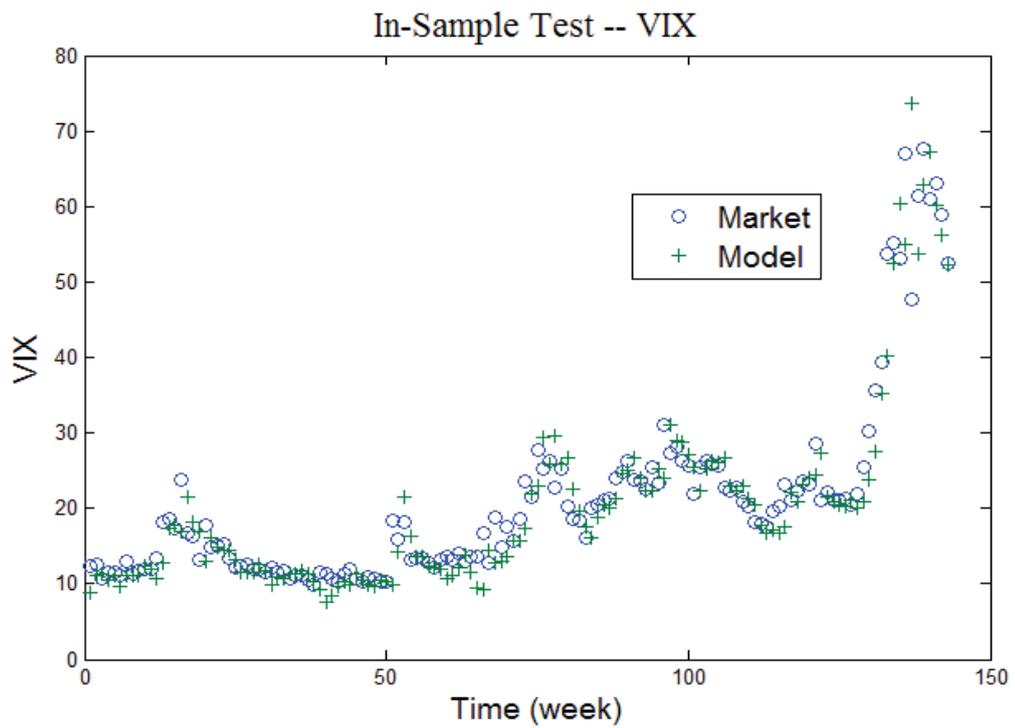


Figure 5.1: In-sample test (GOU3 model) for VIX from March 30, 2004 to December 16, 2008.

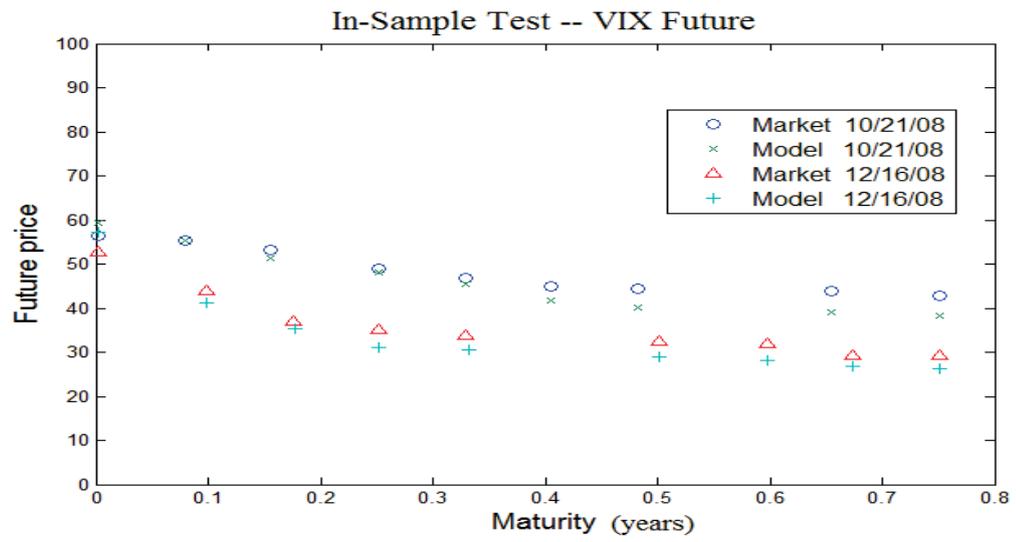


Figure 5.2: In-sample test (GOU3 model) for VIX Futures on October 21, 2008 and December 16, 2008.

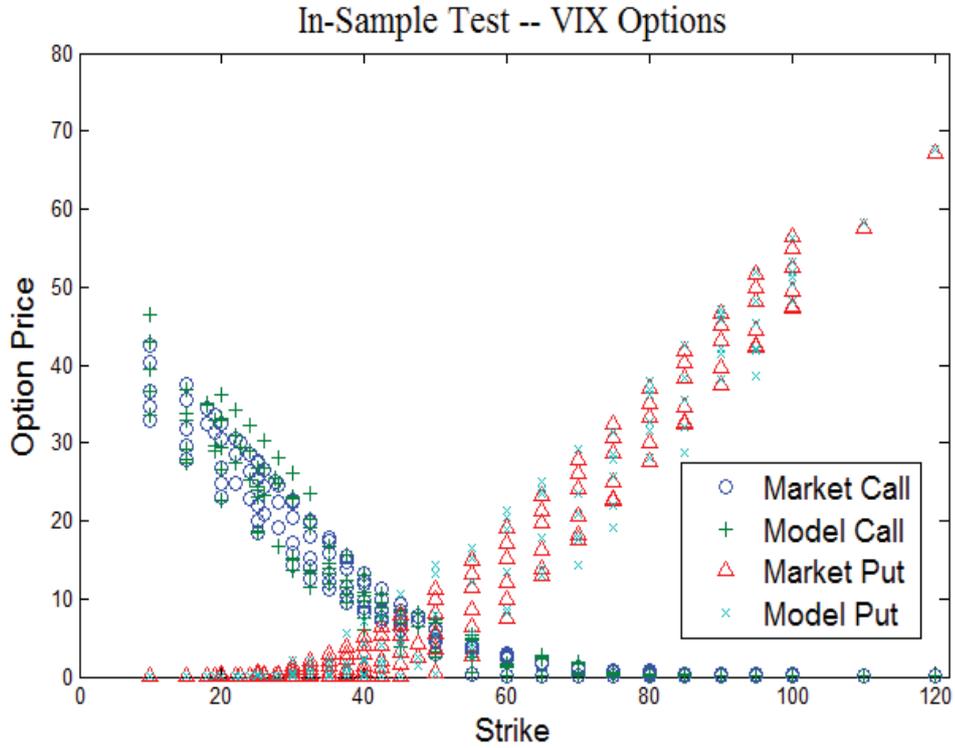


Figure 5.3: In-sample test (GOU3 model) for VIX calls and puts on October 21, 2008 with maturities 36, 64, 92, 127, 155 days.

5.3.4 Out-of-Sample Test

We test the model's out-of-sample performance in period from December 23, 2008 to October 27, 2009. Conditional on the estimated states x_t on December 16, 2008, we test the performance of the model during the out-of-sample period. The model prices, predicted from the UKF, are compared with the historic prices for both VIX and VIX derivatives. The out-of-sample test of VIX, VIX futures and VIX options are illustrated in Figure 5.4, Figure 5.5 and Figure 5.6 respectively. The error statistics of the out-of-sample tests are summarized in Table 5.3. From the

error statistics ARPE and APE, we can see the model's out-of-sample performance is similar to its in-sample performance, which proves that the model is well specified and the estimated parameters are stable.

Table 5.3: Out-of-sample error statistics (GOU3 model) for VIX and VIX derivatives for VIX from March 23, 2008 to October 27, 2009

	Mean	AAE	RMSE	ARPE	APE
VIX	0.4022	3.1038	4.0411	0.0882	0.0914
VIX Futures	0.1474	1.6744	2.4175	0.0459	0.0491
VIX Options	0.7022	1.2724	2.0344	0.5662	0.0989

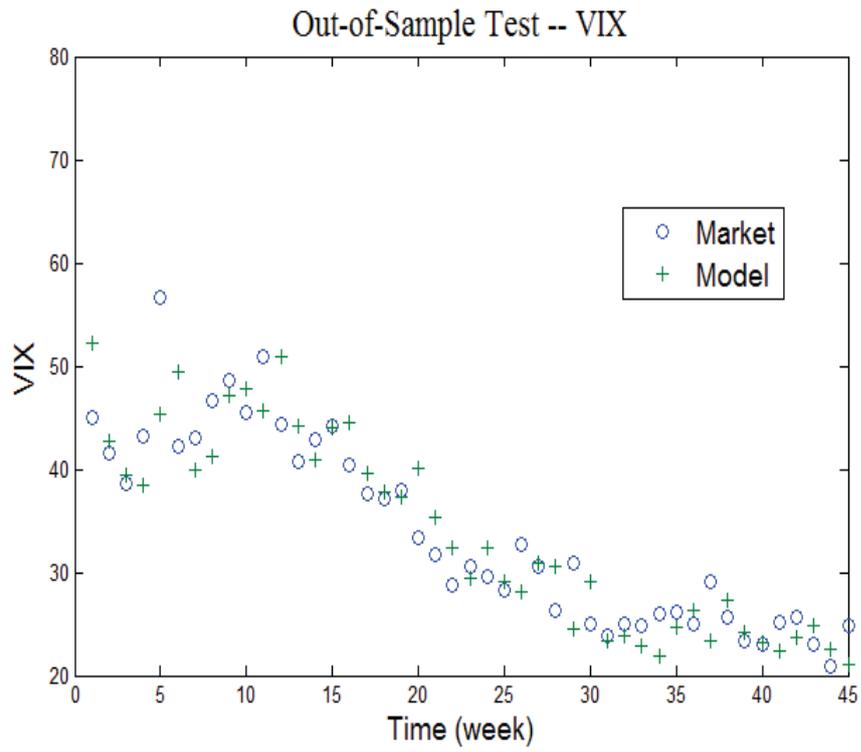


Figure 5.4: Out-of-sample test (GOU3 model)for VIX from March 23, 2008 to October 27, 2009.

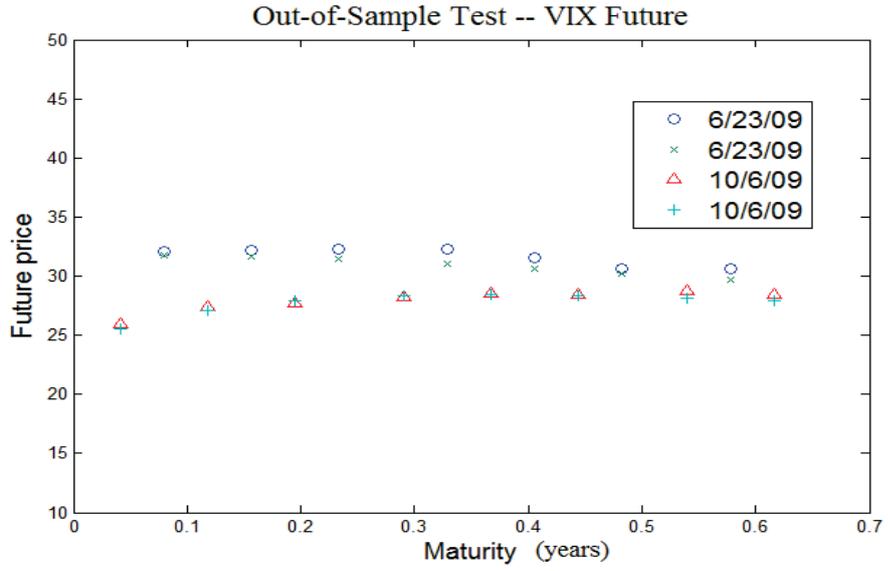


Figure 5.5: Out-of-sample test (GOU3 model) for VIX Futures on June 23, 2009 and October 6, 2009.

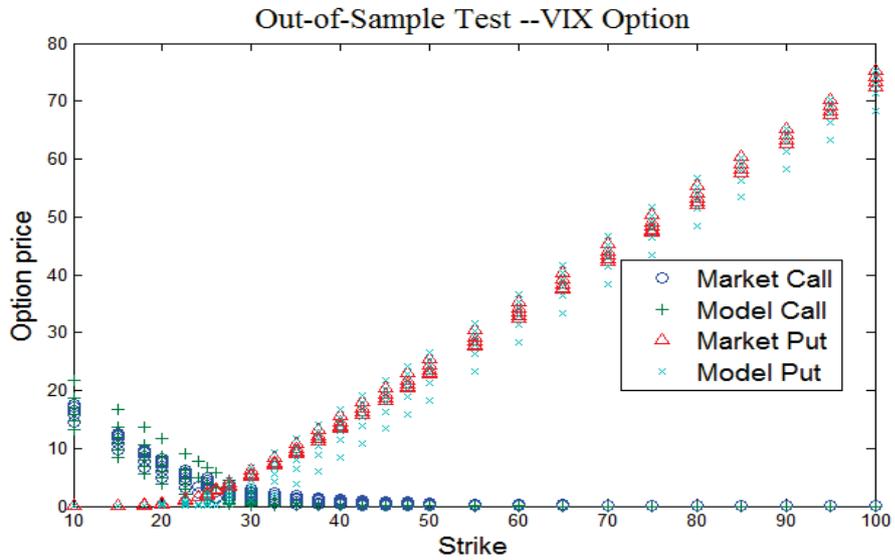


Figure 5.6: Out-of-sample test (GOU3 model) on VIX calls and puts on October 6, 2009 with maturities 15, 43, 71, 106, 134 days

The estimates of c reveal a negative market price for the driving random

factors. Due to Girsanov's Theorem,

$$\Lambda_t = \exp \left\{ - \int_0^t \gamma(x_s, s) dW_s - \frac{1}{2} \int_0^t \gamma(x_s, s)^2 ds \right\} \quad (5.19)$$

the Radon-Nikodym derivative Λ_t has a factor e^{-cW_t} . The investors assign high probability to the higher level of diffusion. This is consistent with the Heidari's [33] result for the mean reverting short rate and option market's negative volatility risk premium.

5.4 CIR3 Model with Constrained UKF

Our second model is a three-factor CIR model (or CIR3). In the CIR3 model, Equation (5.1) still holds, $x_t = [x_{1t}, x_{2t}, x_{3t}]$ and $x_{it}, i = 1, 2, 3$ are independent CIR factors. To simplify the calculation, we assume x_{1t}, x_{2t} , and x_{3t} are independent. To further reduce the number of parameters, we let $b = [1, 1, 1]$, which is not included in the estimation, so that

$$V_t = a + x_{1t} + x_{2t} + x_{3t}. \quad (5.20)$$

We assume that the x_{it} 's are CIR factors both in physical measure and risk-neutral measure, but with different parameters:

$$\begin{aligned} dx_{it} &= \kappa_i(m_i - x_{it}) + \sigma_i \sqrt{x_{it}} dw_{it} && \text{in physical measure,} \\ dx_{it} &= \hat{\kappa}_i(\hat{m}_i - x_{it}) + \hat{\sigma}_i \sqrt{x_{it}} d\hat{w}_{it} && \text{in risk-neutral measure.} \end{aligned} \quad (5.21)$$

The propagation equation uses an Euler approximation of the SDE, similar to the GOU3 model. Under the risk-neutral measure, as described in Section 2.3, conditional on the initial x_{i0} , x_{it} has a scaled noncentral Chi-square distribution, whose

characteristic function is analytically known. We calculate the option price by FFT as described in Section 3.2.2. The future price is given by

$$F(0, V_0) = a + \sum_{i=1}^3 \mu_i. \quad (5.22)$$

Where $\mu_i = e^{-\hat{\kappa}_i t} x_{i0} - (1 - e^{-\hat{\kappa}_i t}) \hat{\kappa}_i^{-1} \hat{m}_i$. The parameter estimation in CIR3 model is similar to that in GOU3 model. At each time step t , the model predictions \tilde{y}_t are computed and imputed to Equation (5.17) with historical data to obtain the likelihood. The model parameters are estimated by maximizing according to Equation (5.18).

Since the CIR factor x_i 's are assumed to be nonnegative, the nonnegative constraints on states need to be included in the estimation. We introduce a simple constrained UKF algorithm, during the time-update step, simply projecting the estimated state x_t into the space of nonnegative values.

5.4.1 Results

The parameters to be estimated are $a, \kappa_i, \hat{\kappa}_i, m_i, \hat{m}_i \sigma_i, \hat{\sigma}_i, x_0$ for $i = 1, 2, 3$ and the MLEs are presented in Table 5.4.

Table 5.4: In-sample maximum likelihood parameter estimates of CIR3 model with Constrained Unscented Kalman filter from March 30, 2004 to December 16, 2008

a	m	k	σ	\hat{m}	\hat{k}	$\hat{\sigma}$	x_0
	26.8231	0.3528	3.0590	2.6828	0.2787	3.1168	6.6154
2.0276	12.9768	0.3720	3.0785	24.0528	0.2616	2.8000	3.7761
	34.8199	0.3021	2.0476	22.3116	0.2017	3.2946	7.5197

We do the in-sample and out-of-sample tests for the same data as for CIR3.

Error statistics are summarized in Table 5.5 and Table 5.6 respectively.

Table 5.5: In-sample error statistics (CIR3) for VIX and VIX derivatives

	Mean	AAE	RMSE	ARPE	APE
VIX	0.4834	2.5886	3.7572	0.1067	0.1072
VIX Futures	0.0275	1.0561	1.5986	0.0498	0.0505
VIX Options	0.2199	0.8634	1.5121	0.7840	0.1593

Table 5.6: Out-of-sample error statistic (CIR3) for VIX and VIX derivatives

	Mean	AAE	RMSE	ARPE	APE
VIX	0.6501	3.1063	4.0155	0.0878	0.0915
VIX Futures	0.0719	1.7087	2.4415	0.0468	0.0501
VIX Options	0.7030	1.2787	2.0377	0.5586	0.0824

5.5 CGMOU2 with Mixed-Gaussian UKF

As in the CIR3 model, we assume $V_t = a + x_{1t} + x_{2t}$, we assume x_{it} 's are CGMOU processes. Since UKF does not apply to multi-dimensional non-Gaussian randomness, we get around it by using the mixed-Gaussian model to approximate the non-Gaussian propagation uncertainty, which is done in this section, or by using the Particle Filter, which is done in the next section. We choose a two-factor model for the Lévy OU, because its number of parameters (21) is comparable to the GOU3 model (16) and the CIR3 model (21):

$$x_{it} = \kappa_i(x_{it} - m_i)dt + dz_{it}, \quad i = 1, 2. \quad (5.23)$$

In the physical measure, we use a mixture of two normal distributions to approximate the driving randomness of x_{it} ,

$$p(\Delta Z_k) = \sum_{i=1}^2 \omega_k^i \mathcal{N}(x_k; \mu_k^i, Q_k^i),$$

where Δz_k is the discrete random change of z_t at time step k . We use an Euler approximation to obtain the propagation equation. Assuming the factors follow

CGMOU processes in the risk-neutral measure, the characteristic function for each factor is given by Equation (2.13). The characteristic function of V_t has an analytical form. The VIX options are priced by FFT. VIX is predicted by Equation (5.16) and VIX futures are priced by Equation (5.11). The parameters are estimated by MLE and the filter method is Mixed-Gaussian UKF.

The parameters to be estimated are $a, \kappa_i, p_i, \sigma_{i1}, \sigma_{i2}, \hat{\kappa}_i, m_i, C_i, M_i, x_0$ for $i = 1, 2$, and the results are presented in Table 5.7. Table 5.11 and Table 5.12 show the in-sample and out-of-sample error statistics.

Table 5.7: In-sample maximum likelihood parameter estimates of CGMOU2 model with Mixed-Gaussian UKF from March 30, 2004 to December 16, 2008

a	p_1	p_2	σ_1	σ_2	κ
0.1623	6.3140	2.3954	0.2311	6.5784	0.4123
			0.4420	7.4923	3.4051
$\hat{\kappa}$	m	\hat{C}	\hat{M}	\hat{G}	x_0
7.2122	6.3140	1.2201	0.2345	0.9818	15.2240
4.4175	2.3954	3.2024	0.4657	0.6745	3.1000

Table 5.8: In-sample error statistics (CGMOU2, Mixed Gaussian UKF) for VIX and VIX derivatives

	Mean	AAE	RMSE	ARPE	APE
VIX	0.5629	2.4400	3.9318	0.1038	0.1157
VIX Futures	0.0105	1.0572	1.5841	0.0502	0.0506
VIX Options	0.2004	0.8809	1.5438	0.8757	0.1425

Table 5.9: Out-of-sample error statistic (CGMOU2, Mixed Gaussian UKF) for VIX and VIX derivatives

	Mean	AAE	RMSE	ARPE	APE
VIX	-0.0467	3.1030	3.9816	0.0893	0.0914
VIX Futures	-0.2602	1.7293	2.4590	0.0475	0.0507
VIX Options	0.6911	1.2590	2.0090	0.5595	0.0879

5.6 CGMOU2 with Particle Filter

In this section we use a Particle Filter to estimate the CGMOU2 model. Both in the physical measure and the risk-neutral measure, V_t is assumed to be a constant a plus two CGMOU factors. We use an ensemble of particles to approximate the posterior distribution of the states of the x_i 's. We propagate the distribution by

propagating the particles. As discussed in Section 4.4, we can increase the precision by using more particles. However, an increase in the number of particles will significantly increase the computational time. By trial and error, we picked 150 as a practical number. According to the Euler approximation, the jumps of the state processes have a Variance Gamma distribution, which is simulated by the difference of two Gamma random variables. We choose the prior as importance density and resample the particles when the estimated effective sample size is less than a pre-set threshold. The VIX futures and VIX options are priced under the risk-neutral measure, as in the previous model.

In the CGMOU model, the source of the risk is the jumps in the Background Driving Lévy Process (BDLP). We are interested in how the market views those jumps. We assume that the physical measure and risk-neutral measure of the BDLP, which is Variance Gamma in this model, are equivalent. According to Chapter 2 Proposition 3, the probability measures generated by the paths of Variance Gamma processes are equivalent if and only if the parameter C 's of the two processes are the same. Thus we assume the same C for both the physical measure and risk-neutral measure.

The parameters to be estimated are $a, \kappa_i, p_i, \sigma_{i1}, \sigma_{i2}, \hat{\kappa}_i, m_i, C_i, M_i, x_0$ for $i = 1, 2$ and the results is showed in Table 5.10. Table 5.11 and Table 5.12 show the in-sample and out-of-sample error statistics.

Table 5.10: In-sample maximum likelihood parameter estimates of CGMOU2 model with Particle filter from March 30, 2004 to December 16, 2008

a	κ	C	M	G
1.2366	1.7342	4.3112	0.5542	0.7221
	1.0772	6.5431	1.2877	1.8688
$\hat{\kappa}$	m	\hat{M}	\hat{G}	x_0
5.6322	7.1004	0.2075	3.7259	10.1751
2.5487	4.6028	0.3697	5.3162	1.4963

Table 5.11: In-sample error statistic (CGMOU2, PF) for VIX and VIX derivatives

	Mean	AAE	RMSE	ARPE	APE
VIX	0.2801	2.2819	3.6469	0.1017	0.1082
VIX Futures	0.1283	1.1238	1.6914	0.0522	0.0538
VIX Options	0.1838	0.8676	1.5070	0.8689	0.1501

Table 5.12: Out-of-sample error statistic (CGMOU2, PF)for VIX and VIX derivatives

	Mean	AAE	RMSE	ARPE	APE
VIX	-0.0381	2.6110	3.1573	0.1278	0.1215
VIX Futures	0.0076	0.9095	1.1557	0.0354	0.0343
VIX Options	-0.3345	0.9632	1.1086	0.2521	0.0868

5.7 Tests on the Models

5.7.1 Radon-Nikodym Derivative

Much empirical work suggests that risk-neutral index volatility generally exceeds physical return volatility. The literature includes the researches of Canina and Figlewski (1993)[15], Lamoureux and Lastrapes (1993) [44], Bakshi et al. (2000) [7], and Christoffersen et al. (2005) [23] based on implied volatility; and that of Bakshi and Kapadia (2003) [9], Bollerslev et al. (2005) [13], Britten-Jones and Neuberger (2000) [14], Carr and Wu (2004) [20], Jiang and Tian (2005) [35], and Polimenis (2006) [46] on formal measures of risk-neutral volatility. Bakshi and Madan (2006) [8] formalize the departure between risk-neutral and physical index return volatilities, termed volatility spreads, and connect them to the higher-order physical return moments and the pricing kernel process. They theoretically and empirically prove the existence of positive volatility spreads when investors are risk averse and when

the physical index distribution is negatively skewed and leptokurtic. Theorem 1 in Bakshi and Madan (2006) [8] states:

Suppose that the aggregate investor behavior is modeled through a class of pricing kernels $m(R)$, satisfying the Taylor expansion around zero. The τ -period volatility spreads are theoretically determined as

$$\begin{aligned} \frac{\sigma_{rn}^2 - \sigma_p^2}{\sigma_p^2} &\approx \partial m / \partial R|_{R=0} (\sigma_p^2)^{1/2} \times \theta_p \\ &+ \frac{1}{2} \partial^2 m / \partial R^2|_{R=0} \sigma_p^2 \times \left(\kappa_p - 1 - 2 \frac{(\partial m / \partial R|_{R=0})^2}{\partial^2 m / \partial R^2|_{R=0}} \right) \end{aligned}$$

where R is the return of index, σ_p^2 , σ_{rn}^2 are the volatility in the physical measure and the risk-neutral measure respectively, and θ_p and κ_p are the skewness and the kurtosis of the physical distribution. The decreasing risk aversion assumption implies $\partial m / \partial R < 0$ and $\partial^2 m / \partial R^2 > 0$. Historical return data shows a mitigated skewness and a significant kurtosis. Thus the major cause of positive volatility spread is the interaction between the risk-aversion of the pricing kernel and the large kurtosis of the distribution of the return.

Risk-neutral probabilities are physical probabilities revised by investors risk preference as determined by the pricing kernel (Harrison and Kreps 1979). An economic interpretation of the positive volatility spreads is that Rational investors are averse to extreme loss and are willing to counteract these exposures by buying protection. The desire to cover these losses typically drives up the risk-neutral probability relative to the actual probability. Since the volatility itself is stochastic, stochastic volatility models such as Heston (1993) have become popular. In the stochastic volatility setting, the positive volatility spread means that the investor has

a twisted view on the distribution of the volatility. The investor's higher perception of higher volatility relative to the actual probability shifts the probability mass to risk-neutral tails. Bakshi and Kapadia (2003) [9] examine the statistical properties of delta hedged option portfolios (buy the option and hedge with stock) confirm that option prices support the negative volatility risk premium and thus the twist of probability measure.

In Chapter 3, we have showed that the statistics of the historical VIX have an outstanding kurtosis. Furthermore, we have shown that the risk-neutral measure behaves more volatile; in the risk-neutral measure, the VIX has more upward jumps than downward jumps. In this section, we investigate whether a similar distortion of probability measure exists in the VIX market.

Under the assumptions for the model, the probability measures generated by the path of Variance Gamma processes are equivalent, and the Radon-Nikodym Derivative depends on the terminal positive jump and negative jump. Here we use $x > 0$ to denote the absolute jump size.

Since the positive jumps and negative jumps are independent,

$$\frac{dQ^+}{dP} \Big|_{\mathcal{F}_t} = \frac{\exp\{-(M' - M)x\}}{\exp\{t \int_0^\infty (e^{-M'x} - e^{-Mx})/x dx\}} \quad (5.24)$$

$$\frac{dQ^-}{dP} \Big|_{\mathcal{F}_t} = \frac{\exp\{-(G' - G)x\}}{\exp\{t \int_{-\infty}^0 (e^{-G'x} - e^{-Gx})/x dx\}} \quad (5.25)$$

Figure 5.8 plots the Radon-Nikodym derivatives for positive jumps and negative jumps of the two factors x_{1t} and x_{2t} . For positive jumps of both factors, the Radon-Nikodym derivatives for positive jumps monotonically increase, while for the negative jumps, the Radon-Nikodym derivatives decrease. These effects can be ex-

plained by the market's greater concern about large positive jumps, which twists the probability space.

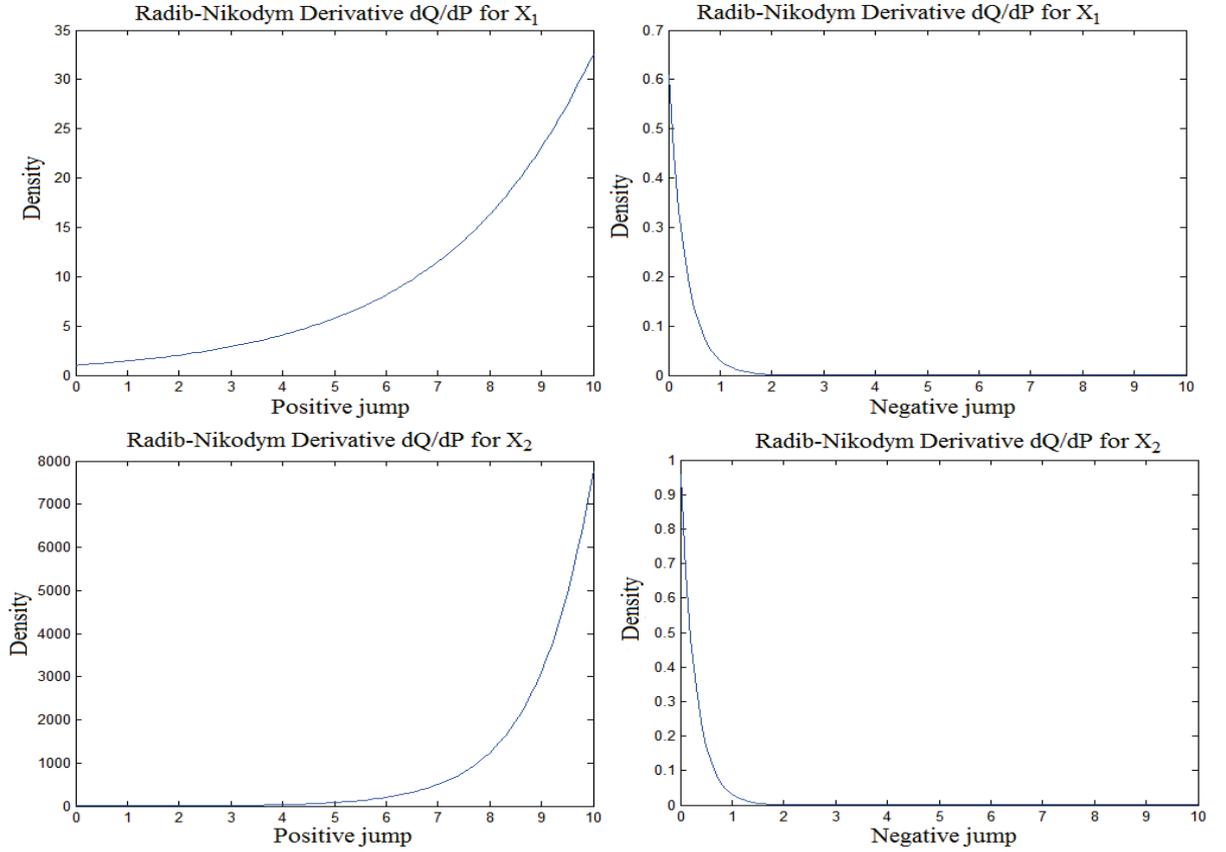


Figure 5.7: Radon-Nikodym derivatives for BDLP factor x_1 and x_2

Let us focus on the positive jump first. Denote by $X \in \mathcal{X}$ the accumulated positive jump from time 0 to t . The expectation of X under the risk-neutral measure is given by

$$E^Q X = E[\Lambda X].$$

The Radon-Nikodym derivative $\Lambda : [0, 1] \rightarrow [0, 1]$ is the distortion function of the physical probability P . From Eqn 5.24, we can see that Λ is an Esscher transform that is, $\Lambda = e^{-(M'-M)X} / E e^{-(M'-M)X}$. In our result, $M' - M < 0$, which means Λ is

an increasing exponential function; the effect of the distortion function is to assign a higher probability weighting to tail probability for positive jumps, or to the big jumps. This coincides with the prediction of positive volatility spreads in Bakshi and Madan (2006) and with the expectation that the agent is averse to the large jump in the VIX.

5.7.2 Trade Strategy Test

Dynamic trading strategies have been developed based on Heidari and Wu (2003) [33] 's Model in the financial market. In these strategies, historical data, including LIBOR, swap rates and interest caps, are used to estimate the both parameters of the physical measure and risk neutral measure, and the states. The estimated parameters and states are use to forecast the prices of bond and the interest rate derivatives. The investors make decisions to purchase or sell. Inspired by Heidari and Wu (2003), we develop a dynamic trading strategy. First estimate model parameters and end states from VIX and VIX derivatives every Wednesday during in-sample period. Then we use the parameters and states to forecast the future price daily in the out-of-sample period. We enter a position by selling a future if its market price is higher than the forecasted price and buying if lower. We close a outstanding position when it is outstanding for five days or at the first time the relationship between market price and forecasted price reverses. We want to check if this strategy can make money on our models. The in-sample period is from March 30, 2004 to May 5, 2009 and the out-of-sample period is from May 6, 2009 to

October 30, 2009. The cash flow and its daily gain are recorded. Unfortunately, the results show that the daily gain fluctuates and the cash flow has a random pattern, and the trade strategy does not make money. The daily gain and cash flow are graphed below.

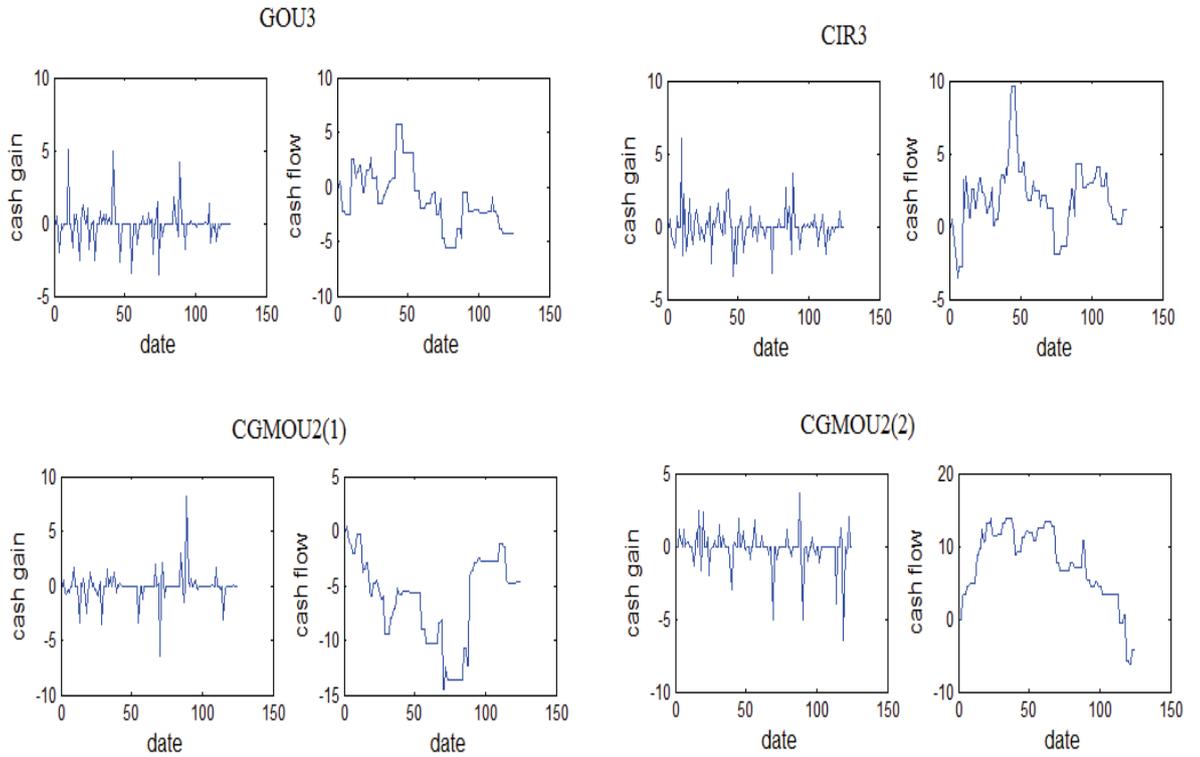


Figure 5.8: Radon-Nikodym Derivative for BGLV factor x_1 and x_2

5.8 Conclusion

From the comparison of the results of the different models, we find that all the multifactor models show quality fit to the historical data, and demonstrated certain forecasting power. With parameters estimated from the in-sample period,

all models give predictions with satisfactory error range to not only for VIX for quite a period, but also for VIX derivatives.

All error statistics are comparable among the five models either in-sample or out-of-sample. We cannot tell significant difference among the performance of the models.

Except for the GOU3 model, all models have intense computational complexity. The Fourier transformation is necessary for all and numerical integration is needed for the CGMOU2. Furthermore, the Particle filter requires large number of particles for the Monte Carlo simulation. The estimation time for the Gaussian optimization is on average about 2 hours, while the CIR3 optimization takes more than 20 hours and the CGMOU2 models take several days. With computational time in consideration, we recommend GOU3 for the joint estimation. The multi-factors Gaussian OU models show strength not only in fitting data across time but also in fitting the randomness with higher moments.

From the Radon-Nikodym Derivative tests, both GOU3 and CGMOU2 suggest that the market assigns higher probability weighting to a higher level of randomness. CGMOU2 model shows that market assigns higher probability weighting to big positive jumps.

The simple dynamic trading strategy does not make money on these multifactor models. There is an inconsistency in the procedures, since the parameter and state are estimated in weekly data, while the strategy are tested in daily data. We will use daily data in future research and make more tests.

Chapter 6

Conclusion and Future Study

In this dissertation, we explore how to employ different mean reverting processes to model the VIX movement, and how to use these models to price the VIX options. In particular, we introduce the Lévy OU models, and FFT with imaginary Gamma option in the option pricing. In the one-factor models, we find that the VIX process both in physical measure and risk-neutral is better presented as a Lévy OU processes with jump process rather than as a Gaussian OU process. Negative jumps need to be included in the Lévy OU process. In the risk-neutral measure, for the Lévy OU processes with jumps, positive jumps occur more frequently.

We use multi-factor mean reverting processes build a joint frame for both the physical measure and the risk-neutral measure. To estimate the various models, we employ several filter techniques, namely, Unscented Kalman Filter (UKF), constrained UKF, mixed-Gaussian UKF and Particle Filter. We find all the multi-factor models show fitting to the historical data, and demonstrated some forecasting power. We find no significant difference in the performance of the different models. Except for the GOU3 model, all models show intense computation complexity and take a long time. As present results suggest, we recommend the GOU3 model for the joint framework task. We also investigated the Radon-Nikodym derivative in the GOU3 and CGMOU2 models. The results show that the market assigns higher probability

weighting to positive shocks in VIX. We tested a dynamic trading strategy on those joint models, and got random pattern cash flow.

For future study, we can investigate how to improve the computation in the CIR3 and CGMOU2 model and reduce the complexity. We can improve the Gaussian Mixed UKF and particle filter implementation, to make these models take less time. We can study the effects of introducing more factors to the multifactor mean reverting model. We can try more trading strategy with the expectation to find a model applicable to the financial market.

A.1 The Derivation of VIX Call Price

Assuming the V_t (VIX) has a characteristic function $\Phi(\cdot)$ under risk neutral measure, The general Fourier transform of a call price is given by Equation (1).

$$\begin{aligned}\Psi_c(u) &= \int_0^\infty e^{iuK} C(K) dK \\ &= e^{-r\tau} \left\{ \frac{1}{u^2} [1 - \Phi(u)] + \frac{i}{u} F_\tau \right\},\end{aligned}\tag{1}$$

where $u = a + bi$, F_τ is future price with maturity τ , $a \geq 0$ is a constant and $b \in \Re$.

The derivation of Equation (1) is given as follows. let $f(x)$ be the PDF of VIX x at maturity.

$$\begin{aligned}\Psi_c(u) &= \int_0^\infty e^{iuK} C(K) dK \\ &= \int_0^\infty e^{iuK} e^{-r\tau} E^Q(X - K)^+ dK \\ &= e^{-r\tau} \int_0^\infty e^{iuK} \int_K^\infty (x - K) f(x) dx dK \\ &= e^{-r\tau} \int_0^\infty f(x) dx \int_0^x e^{iuK} (x - K) dK \\ &= e^{-r\tau} \int_0^\infty f(x) dx \frac{-e^{iux} + iux + 1}{u^2} \\ &= e^{-r\tau} \left\{ \int_0^\infty f(x) dx \frac{-e^{iux} + 1}{u^2} dx + \frac{i}{u} E^Q(X) \right\} \\ &= e^{-r\tau} \left\{ \frac{1}{u^2} [1 - \Phi(u)] + \frac{i}{u} F_t \right\}.\end{aligned}$$

Bibliography

- [1] D.L. Alpach and H.W. Sorenson, *Nonlinear Bayesian estimation using Gaussian sum approximations*. IEEE Transactions on Automatic Control, 17(4):439-448, August 1972.
- [2] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Englewood Cliffs, NJ:Prentice-Hall, 1979.
- [3] David Applebaum, *Lévy Processes and Stochastic Calculus*, CUP 2004
- [4] Ienkaran Arasaratnam, Simon Haykin, and Robert J. Elliott *Filtering algorithms using Gauss-Hermite quadrature* Proceedings of the IEEE, 953-977, 2007.
- [5] M. Sanjeev Arulampalam, Simon Maskell, Neil Gordon, and Tim Clapp, *A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking* IEEE Transactions on Signal Processing, VOL. 50, NO. 2, Feb 2002.
- [6] Bachelier, L., *Theorie de la speculation*, *Annales de l'Ecole Normale Supérieure*, 17(100), pp. 21-86.
- [7] Bakshi, G., C. Cao, Z. Chen. 2000, *Pricing and hedging long-term options*, *J. Econometrics* 94 277-318.
- [8] Bakshi, Gurdip S. and Madan, Dilip B., *A Theory of Volatility Spreads (January 2006)*, Robert H. Smith School Research Paper No. RHS 06-028.
- [9] Gurdip Bakshi and Nikunj Kapadia, *Delta-Hedged Gains and the Negative Market Volatility Risk Premium*, *The Review of Financial Studies* 2003, pp. 527-566.
- [10] Barndorff-Nielsen, O. E. and Shephard, N. (2001a) *Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics*, *Journal of the Royal Statistical Society B* 63, 1672-41.
- [11] N. Bergman, *Recursive Bayesian estimation: Navigation and tracking applications*, Ph.D. dissertation, Linköping Univ., Linköping, Sweden, 1999.
- [12] Black, F. and Scholes, M., *The pricing of options and corporate liabilities*, *Journal of Political Economy*, 3(1973)

- [13] Bollerslev, T., M. Gibson, H. Zhou. 2005, *Dynamic estimation of volatility risk premia and investor risk aversion from optionimplied and realized volatilities*, Working paper, Duke University, Durham, NC and Federal Reserve Board.
- [14] Britten-Jones, M., A. Neuberger. 2000. Option prices, *implied price processes, and stochastic volatility*, J. Finance 55 839866.
- [15] Canina, L., S. Figlewski. 1993. *The information content of implied volatilities* , Rev. Financial Stud. 6 659681.
- [16] Cao, L., and Fu, M. 2010, *Estimating Greeks for Variance-Gamma* . Proceedings of the 2010 Winter Simulation Conference.
- [17] Cappe, O.; Moulines, E.; Ryden, T. (2005). *Inference in Hidden Markov Models*. Springer
- [18] 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
- [19] Carr P and D Madan, 1998 *Towards a theory of volatility trading In Volatility*, edited by Robert Jarrow, Risk Books, London, pages 417427
- [20] Carr, P., L. Wu. 2004. *Variance risk premia*, Working paper, Baruch College.
- [21] Cont, R. (2001) *Empirical properties of asset returns: stylized facts and statistical issues*. Quantitive Finance 1, 223236
- [22] Rama Cont, Peter Tankov, *Financial modelling with jump processes* Chapman&Hall/CRC 2004
- [23] Christoffersen, P., S. Heston, K. Jacobs. 2006, *Option valuation with conditional skewness*, J. Econometrics, 131, 253-284.
- [24] COBE , *VIX White Paper*, <http://www.cboe.com/micro/vix/vixwhite.pdf>.
- [25] Cox, J. C., Ingersoll, J. E. Ross, S. A. , *An intertemporal general equilibrium model of asset prices*, Econometrica (1985b) 53, 363-384.
- [26] Dai, Q., and K. Singleton. *Specification Analysis of Affine Term Structure Models*, Journal of Finance, 55 (2000), 1943-1978.

- [27] A. Doucet, *On sequential Monte Carlo methods for Bayesian filtering*, Dept. Eng., Univ. Cambridge, UK, Tech. Rep., 1998.
- [28] Duffie, G. , *Term premia and interest rate forecasts in affine models*, Journal of Finance, 57, 405-443 (2002).
- [29] Michael C. Fu, *Variance-Gamma and Monte Carlo* Advances in Mathematical Finance, Birkhuser, 21-35, 2007.
- [30] Girsanov, I. V., "On transforming a certain class of stochastic processes by absolutely continuous substitution of measures", Theory of Probability and its Applications, 1960
- [31] Grunbichler, A. & Longstaff, F.A., *Valuing futures and options on volatility*, Journal of Banking and Finance, 20(6), 9851001 (1996).
- [32] Heidari, M., and L. Wu. *Are Interest Rate Derivatives Spanned by the Term Structure of Interest Rates?* Journal of Fixed Income 13 (2003), 7586.
- [33] Massoud Heidari, and Liuren Wu, *A Joint Framework for Consistently Pricing Interest Rates and Interest Rate Derivatives*, Journal of Financial and Quantitative Analysis, 2009, 44(3), 517-550
- [34] S. Howison, A. Rafailidis and H. Rasmussen, *On the Pricing and Hedging of Volatility Derivatives*, Applied Mathematical Finance J., 2004,
- [35] Jiang, G., Y. Tian. 2005, *The model-free implied volatility and its information content*, Rev. Financial Stud. 18 13051342.
- [36] S. J. Julier and J. K. Uhlmann, *A New Extension of the Kalman Filter to Nonlinear Systems*. In Proc. of AeroSense: The 11th Int. Symp. on Aerospace/Defence Sensing, Simulation and Controls., 1997
- [37] S. J. Julier, J. K. Uhlmann, H. F. Durrant-Whyte, *A New Approach for Filtering Nonlinear Systems*. In Proceedings of the American Control Conference, Seattle, Washington., pages 1628-1632, 1995.
- [38] R. E. Kalman, *A New Approach to Linear Filtering and Prediction Problems* Transaction of the ASME Journal of Basic Engineering, (March 1960) pp. 35-45.

- [39] Rambabu Kandepu, Lars Imsland and Bjarne A. Foss, *Constrained state estimation using the Unscented Kalman Filter* In proceedings of 16th Mediterranean conference on Control and Automation (MED), June 2008.
- [40] MacBeth, J.D. & Merville, L.J, *Tests of the Black-Scholes and Cox Call Option Valuation Models*, Journal of Finance, 35(2), 285-301 (2009).
- [41] Dilip Madan, Peter Carr, Eric Chang. *The Variance Gamma Process and Option Pricing*. European Finance Review 1998, 2, 79105.
- [42] Madan, D. B. and E. Seneta, *Chebyshev polynomial approximations and characteristic function estimation*, Journal of the Royal Statistical Society Series B 49 (2), 163-169 (1987).
- [43] Mandelbrot B. (1963) *The Variation of Certain Speculative Prices*. *Journal of Business*, vol. 36, 394
- [44] Lamoureux, C., W. Lastrapes. 1993, *The information content of implied volatilities*, Rev. Financial Stud. 6 659681.
- [45] Lin, Y. & Chang, C. (2009). *VIX option pricing*, Journal of Futures Market, 29, 523-543.
- [46] Polimenis, V. (2006) *Skewness correction for asset pricing*, Working paper, University of California, Riverside, CA.
- [47] K. Sato, *Lévy Processes and Infinitely Divisible Distribution*, Cambridge Press, 1999
- [48] Wim Schoutens, *Lévy Processes in Finance: Pricing Financial Derivatives*, Wiley, May 2003
- [49] D. Simon and T. L. Chia, *Kalman Filtering with State Equality Constraints*, IEEE Transactions on Aerospace and Electronic Systems, Vol. 38, 2002, pp. 128-136.
- [50] Peter Tankov, *Financial Modelling with Jump Processes*, Financial Mathematics Series Chapman & Hall/CRC (2003),
- [51] G. E. Uhlenbeck and L. S. Ornstein, *On the theory of Brownian motion*, Phys. Rev. 36 (1930) 823-841.

- [52] S. Ungarala, E. Dolence and K. Li, *Constrained Extended Kalman Filter for Nonlinear State Estimation*, In proceedings of 8th International IFAC Symposium on Dynamics and Control Process Systems, Cancun, Mexico, June 2007, vol. 2, pp. 63-68.
- [53] Eric A.Wan and Rudolph van der Merwe, *Kalman Filtering and Neural Networks* , Chapter 7 Wiley , Eds. S. Haykin, 2001.
- [54] M. C. Wang and G. E. Uhlenbeck, *On the theory of Brownian motion . II*, Rev. Modern Phys. 17 (1945) 323-342; MR0013266 (7,130e).
- [55] Whaley, R. E. *Derivatives on market volatility: Hedging tools long overdue*, Journal of Derivatives, 1, 7184(1993).
- [56] Wim Schoutens, *Lévy Processes in Finance, Pricing Financial Derivatives* Wiley, 2003
- [57] Bing Zhang, *A new Lévy based short-rate model for the fixed income market and its estimation with particle filter*, Thesis 2006, University of maryland.