PRICING VOLATILITY DERIVATIVES UNDER LÉVY PROCESSES

A THESIS SUBMITTED TO AUCKLAND UNIVERSITY OF TECHNOLOGY
IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Supervisors

Dr. Wenjun Zhang

Prof. Jiling Cao

July 2020

By

Shu Su

School of Engineering, Computer and Mathematical Sciences

Copyright

Declaration

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person (except where explicitly defined in the acknowledgements), nor material which to a substantial extent has been accepted for the qualification of any other degree or diploma of a university or other institution of higher learning.

Signature of candidate

Acknowledgements

Foremost, I would like to express my sincere gratitude to my supervisors, Prof. Jiling Cao and Dr. Wenjun Zhang for the continuous support of my PhD study and research with their patience and enthusiasm. They sincerely conveyed the scientific learning methods and encouraged me to persistently learn new knowledge. Their guidance helped me throughout my PhD research and thesis writing. I could not have imagined having a better advisor and mentor for my PhD study.

Secondly, I would like to acknowledge the valuable input of Dr. Xinfeng Ruan, who contributed to many discussions and comments that helped to shape my PhD research projects and solve some difficult problems.

Thirdly, I am deeply grateful to Dr. José Da Fonseca and Assoc. Prof. Hui Zhao for sharing their knowledge and resources as well as their insightful comments and valuable advises. I also would like to thank the rest of AUT Financial Mathematics and Computation research group (AUT-FMC), Dr. Nuttanan (Nate) Wichitaksorn, Dr. Reza Moosavi Mohseni, and Gaurav Kapoor. The weekly group meetings and discussions are helpful for improving my presentation skills and motivating me to think more deeply about my research.

I also want to thank my husband and parents for supporting, understanding and encouraging me throughout my journey of the PhD study. Last but not the least, my sincere thanks go to the Auckland University of Technology for providing the learning opportunity, all research facilities and the financial supports.

Abstract

In this thesis, we study the pricing of the volatility derivatives, including VIX options, VIX futures, VXX options and S&P 500 variance futures, under Lévy processes with stochastic volatility. In particular, we investigate the role of different types of jump structures, such as finite-activity jump, infinite-activity jump and double jump structures, as well as the role of variance processes with time-varying mean in the valuation of volatility derivatives. In our models, we assume that the long-term mean of the variance process follows an Ornstein–Uhlenbeck process and specify the infinity-activity jump component of the main process in four cases: the variance gamma process (VG), the normal inverse Gaussian process (NIG), the tempered stable process (TS) and the generalized tempered stable process (GTS). Then, we apply the combined estimation approach of an unscented Kalman filter (UKF) and maximum log-likelihood estimation (MLE) to our models and make an extensive comparison analysis on the performance among the different models.

Our empirical studies reveal three important results. First, the models with infinite-activity jumps are superior to the models with finite-activity jumps, particularly in pricing VIX options and VXX options. Thus, the infinite-activity jumps cannot be ignored in pricing volatility derivatives. Second, both the infinite-activity jump and diffusion components play important roles in modelling the dynamics of the underlying asset returns for the volatility derivatives. Third, the mean of the variance process for the S&P 500 index returns varies stochastically toward to its long-term mean.

Journal Publications

[A] Cao, J., Ruan, X., Su, S. and Zhang, W. (2020). Pricing VIX derivatives with infinite-activity jumps, *Journal of Futures Markets*, *40*(3), 329-354.

[B] Cao, J., Ruan, X., Su, S. and Zhang, W. (2020). Pricing VXX options under Lévy processes, *Journal of Futures Markets*, to appear.

Conference Presentations

[A] Cao, J., Ruan, X., Su, S. and Zhang, W. Pricing VIX derivatives with infinite-Activity jumps. *Oral presentation at AUT Mathematical Sciences Symposium*, November 2018, Auckland, New Zealand.

[B] Cao, J., Ruan, X., Su, S. and Zhang, W. Pricing VIX derivatives with infinite-Activity jumps. *Oral presentation at the 2019 Derivative Markets Conference*, August 2019, Queenstown, New Zealand.

[C] Cao, J., Ruan, X., Su, S. and Zhang, W. Pricing VXX options under time-changed Lévy processes. *Oral presentation at AUT Mathematical Sciences Symposium*, November 2019, Auckland, New Zealand.

[C] Cao, J., Ruan, X., Su, S. and Zhang, W. Pricing VIX derivatives with infinite-Activity jumps. *Oral presentation at Quantitative Methods in Finance 2019 Conference*, December 2019, Sydney, Australia.

Contents

Copyright **2**

Declaration **3**

Acknowledgements **4**

Abstract **5**

Journal Publications **6**

1 Introduction **12**
1.1 Background 12
1.2 Literature Review 16
1.2.1 Early History of Financial Derivative Pricing 16
1.2.2 Literature Review on VIX and pricing volatility derivatives . . 21
1.3 Research Questions 24
1.4 Research Contributions and Thesis Organization 25

2 Preliminaries **27**
2.1 Mathematical Tools 27
2.1.1 Probability Measures 27
2.1.2 Random Variables 29
2.1.3 Conditional Expectation 34
2.1.4 Stochastic Processes 37
2.1.5 Poisson Process 40
2.1.6 Lévy process 42
2.1.7 Pure Jump Lévy Process 45
2.2 Fundamentals of Quantitative Finance 48
2.2.1 Futures and Options 48
2.2.2 Risk-Neutral Pricing 49
2.2.3 Implied Volatility 53
2.2.4 VIX and S&P 500 variance futures 54
2.3 Numerical Methods 57
2.3.1 Gauss-Laguerre Quadrature Method 58
2.3.2 4th-order Runge-Kutta Method 59

2.3.3 Unscented Kalman Filtering 61
2.3.4 Maximum Likelihood Estimation 63

3 Pricing VIX Derivatives with Infinite-Activity Jumps 65
3.1 Data 66
3.1.1 A Preliminary Analysis on VIX Data 66
3.1.2 Sample Construction and Sample Summary of VIX Derivatives Data 68
3.2 A Two-factor Model with Infinite-activity Jumps 70
3.2.1 Model Specifications 70
3.2.2 Characteristic Function 74
3.3 VIX Derivative Pricing 75
3.3.1 Futures Valuation 75
3.3.2 Options Valuation 76
3.4 Model Calibration 78
3.4.1 Estimation Procedure 78
3.4.2 Parameter Estimation Results 82
3.5 Model Performance Analysis 83
3.5.1 Criteria of Performance 84
3.5.2 Model Performance Comparison 85
3.5.3 Fit to VIX Futures and Options Pricing 86
3.5.4 Finite- vs. Infinite-activity Jumps 89

4 Pricing VXX Options under Lévy Process 95
4.1 Data 96
4.1.1 VXX 96
4.1.2 VXX Options 99
4.2 Model Specifications 102
4.3 Option Pricing Approach 103
4.4 Model Calibration 106
4.4.1 Calibration Procedure 106
4.4.2 Model Estimation Result Analysis 108
4.5 Model Performance Analysis 109
4.5.1 Performance Analysis on VXX Option Pricing 111
4.5.2 Fitting to the Term Structures of Options and Implied Volatility 112
4.5.3 Evolution of the Instantaneous Variance 115

5 Pricing S&P 500 Variance Futures 118
5.1 Data 118
5.2 Model Specifications 121
5.3 VIX Squared and Integrated Variance Premium 125
5.4 Variance Futures Valuation 128
5.5 Model Calibration and Estimation Results 130
5.6 Empirical Results Analysis 133

5.6.1 Criteria for Performance Measure 133
5.6.2 Model Performance Comparison 134

6 Conclusion and Future Work **138**
6.1 Conclusion 138
6.2 Future Work 141

References **145**

Appendices **151**

List of Tables

2.1 Characteristic exponent of the selected pure Lévy jump process 48

3.1 Descriptive statistics of VIX 1992-2017 67
3.2 Test statistics and critical values for the jump test 68
3.3 Summary statistics for VIX derivatives by time to maturity 71
3.4 VG and NIG processes 73
3.5 Model parameter estimates 82
3.6 Model performance comparison – log-likelihood 86
3.7 Model performance comparison pricing futures – RMSE 87
3.8 Model performance comparison in pricing options – RMSE 90
3.9 Parameter estimates - OU-CP 92
3.10 Model performance – OU-CP 93

4.1 VXX reverse split 97
4.2 Descriptive statistics of daily VXX 99
4.3 Summary for VXX options 101
4.3 Parameter estimates 110
4.4 Root mean square error (RMSE) 112
4.5 Pairwise model performance comparison 113
4.6 RMSEs of option prices by moneyness and maturity 114

5.1 Descriptive statistics of S&P 500 variance futures contracts 120
5.2 Descriptive statistics of the index levels and log levels 121
5.3 Parameter estimates 132
5.4 Root mean squared errors and mean absolute percentage errors 135

A.1 Parameters 155

List of Figures

1.1 Open interest of VIX futures and average daily trading volume, by six months, from March 2004 to October 2016 14
1.2 Daily average trading volume of VIX options between 2006 and 2019 15
1.3 Open interest and average daily trading volume of the VXX options, by quarters, from July 2010 to December 2017 17

3.1 CDF of the squared pricing errors for futures 88
3.2 Comparison of half-yearly average SEs 91

4.1 Daily close price evolution of the VXX and logarithmic VXX during the period between 25 May 2010 and 18 December 2017. 97
4.2 Annualized volatility of the VXX logarithmic returns, with 30-day rolling window, from 25 May 2010 to 28 December 2017. 100
4.3 Fit to the implied volatility 116
4.4 Historical evolution of the v_t from 25 May 2010 to 28 December 2017 117

5.1 Variance futures prices vs. the VIX squared 122
5.2 S&P 500 variance futures 129
5.3 Market data vs. estimated data 137

A.1 “True" m_t vs. Predicted m_t 156
A.2 “True" v_t vs. Predicted v_t 156

Chapter 1

Introduction

This chapter introduces the volatility indexes and the volatility derivatives (e.g. VIX options, VIX futures, VXX options and S&P variance futures). Section 1.1 provides the background on VIX and its derivatives. Section 1.2 gives a literature review on financial derivatives pricing. Based on the literature review, we raise our research questions in Section 1.3. In the end of this chapter, the major contributions of our research are stated and the thesis structure is outlined.

1.1 Background

It has been widely recognized that the time-varying volatility of an asset return is a prominent risk measurement for the financial market. Thus, several volatility indexes and their derivatives have evolved during the last three decades. One of the most influential volatility indexes, known as the Chicago Board Options Exchange (CBOE) Volatility Index (VIX), was first introduced in 1993 and its value was calculated based on the Black-Scholes model from the OEX 100 put and call options (Whaley, 1993; Carr & Lee, 2003). Ten years later, CBOE extended the underlying of VIX to a much broader index, S&P 500 composite stock price index (S&P 500), and proposed a more

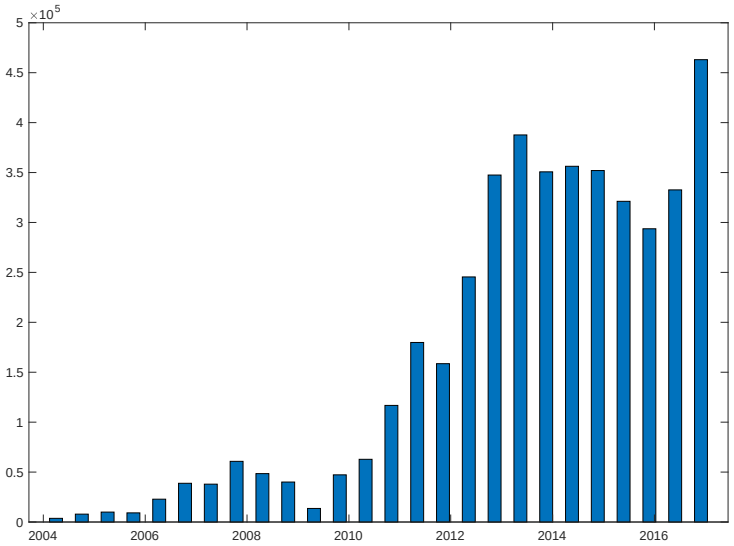
robust methodology for valuing the VIX, which is described in details in the CBOE white paper. It is calculated by the S&P 500 options market over the next 30-day period and indicates overall stock market fear (Carr & Lee, 2003) [1]. It also can be treated as a risk-neutral gauge of the future implied volatility in the US stock market. Thus, understanding the dynamics of volatility variation is crucial for market practitioners and academic researchers.

The VIX is not tradable but can be speculated through the VIX derivatives. Due to a negative correlation between the VIX and the S&P 500 index, market participants treat the VIX derivatives as an important trading vehicle for reducing exposure to risk. Reported by CBOE, the VIX derivatives include VIX futures and VIX options. In 2004, CBOE introduced VIX futures. Since then, VIX futures have been quickly accepted as trading instruments in the volatility market. According to Figure 1.1(a), the average daily trading volume of VIX futures on a half-year basis significantly increased to over $200,000$ contracts in October of 2016. From Figure 1.1(b), we observe that the open interest also has grown quickly since inception. Although it decreased for about one and half years after the peak in 2007, it bottomed out in the second half year of 2009 and reached over $460,000$ contracts in October 2016.

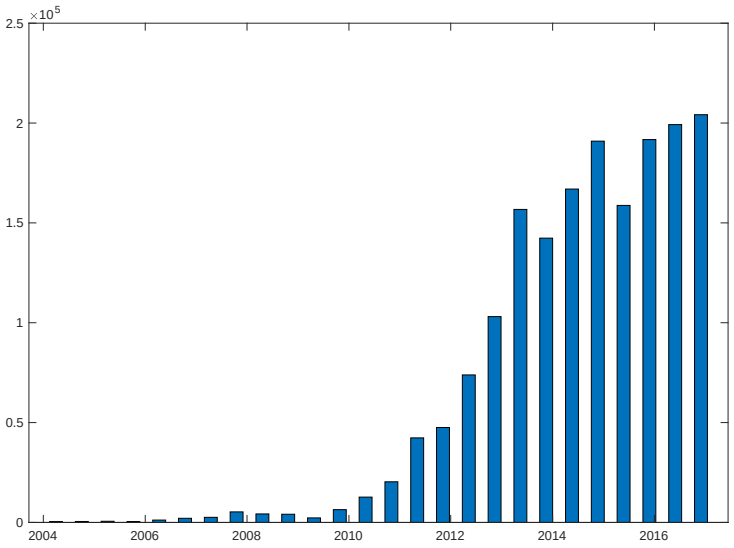
After the successful launch of VIX futures, CBOE introduced VIX options in 2006, which can provide an alternative way for hedging the market risk with relatively low costs. Since then, the VIX options have become popular financial tools in the risk management. Not surprisingly, the average daily volume of VIX options expanded over 25 times between 2006 and 2019, referring to Figure 1.2. In particular, the daily average trading volume reaches the highest level, over 700,000 per day, in 2007.

Due to the popularity of VIX futures, the first VIX futures exchanged-traded notes (ETNs), known as the iPath S&P 500 VIX short-term futures ETN (VXX), was introduced by Barclays Bank and listed on the New York Stock Exchange on 29 January

[1] See https://www.cboe.com/micro/vix/vixwhite.pdf.



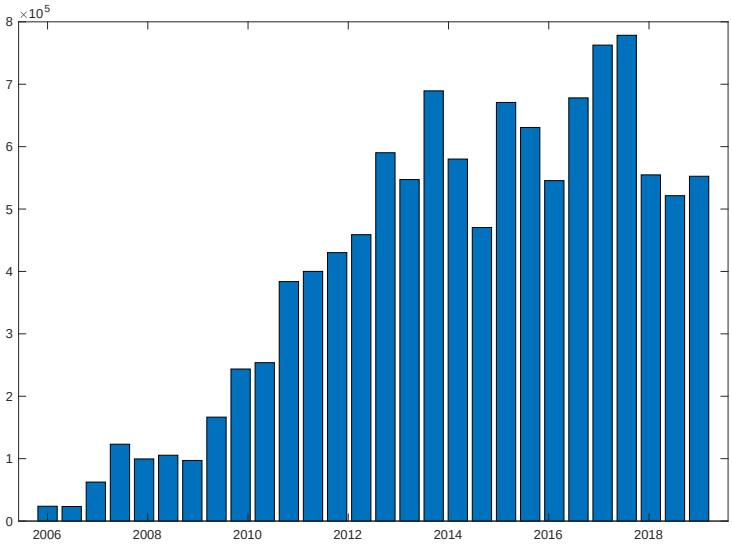
(a) Average daily open interest



(b) Average daily trading volume

Figure 1.1: Open interest of VIX futures and average daily trading volume, by six months, from March 2004 to October 2016

Figure 1.2: Daily average trading volume of VIX options between 2006 and 2019



Notes: The daily average trading volume of VIX options between 2006 and 2019, by six months. Source from: http://www.cboe.com/data/historical-options-data/volume-put-call-ratios.

2009 (Rhoads, 2011). This ETN is an unsecured debt security without principal protection, which reflects investment performance on short-term VIX futures. [2] The daily trading volume of the VXX has remained over 10 million and peaked at more than 109 million on 01 August 2019.[3] CBOE introduced the VXX option on the May of 2010. Each VXX option contract includes 100 VXX ETN. This option has quickly become a popular volatility trading vehicle. Figure 1.3 depicts the average daily trading volume and option interest by quarter for the VXX options. In Figure 1.3(a), we notice that the average daily open interest for the VXX options increases to a considerably high level (about $4,200,000$) within five years. During the same time, the average daily trading volume of VXX options displays an upward trend in the first few years and

[2] Short-term means 1-month average term VIX futures.

[3] https://www.nasdaq.com/market-activity/funds-and-etfs/vxx/historical. Please be aware that VXX matured on January 30 January, 2019. Due to the huge demand for VXX, Barclays Bank launched iPath Series B S&P 500 VIX Short-Term Futures ETN (VXXB) on 17 January 2018 which matures on 23 January 2048. VXXB was identical through to the maturity of VXX, so VXXB is regarded the same as VXX. See https://www.ipathetn.com/US/16/en/instruments.app#/details/341408.

then fluctuates over an impressive level (about $218,000$ contracts), which is shown in Figure 1.3(b). In particular, the trading volume reaches a new peak (more than $430,000$ contracts) in 2015. These are indications that the VXX option is a very actively traded product for hedging and speculating in the market.

S&P 500 variance futures (VA) is another type of volatility futures, which were listed on the CBOE and started for trading on 10 December 2012. The variance futures contracts are written on the realized variance of S&P 500. The final settlement price of a VA contract is calculated based on the S&P 500 realized variance from the initial listing date to the expiration date. The contracts expire at the third Friday of the expiring month[4]. In addition, VA are much less liquid than the VIX futures, as we can see that the trading volume of the VA is much smaller than that of VIX futures [5].

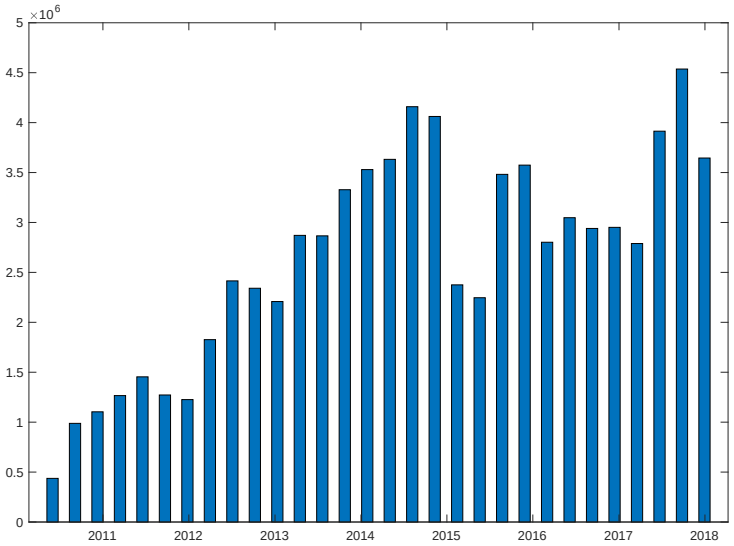
1.2 Literature Review

In this section, we give a comprehensive literature reviews on the history of the financial derivative pricing in Sub-section 1.2.1, and literature reviews on VIX and pricing volatility derivatives in Sub-section 1.2.2.

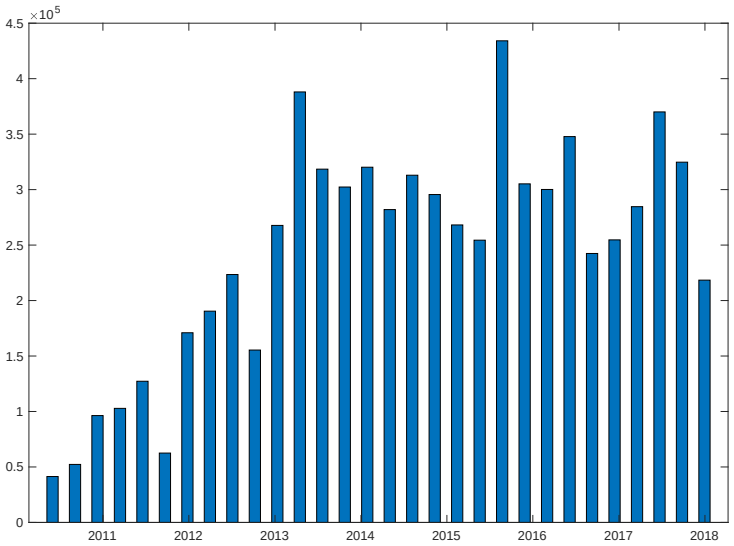
1.2.1 Early History of Financial Derivative Pricing

Research on derivative pricing has been developed over 50 years. In the derivative pricing theory, the fundamental assumption is that financial markets are efficient. An efficient market means that the change of asset prices is unpredictable and the asset prices are always determined by all information available at that time (Fama, 1965). During the same time period, Samuelson shows that the law of iterated expectations can be used in the security market analysis (Samuelson, 1965). A couple of years

[4]See http://www.cboe.com/products/futures/va-s-p-500-variance-futures/contract-specifications
[5]See https://markets.cboe.com/us/futures/market-statistics/historical-data/



(a) Average daily open interest



(b) Average daily trading volume

Figure 1.3: Open interest and average daily trading volume of the VXX options, by quarters, from July 2010 to December 2017

later, LeRoy gave more detailed reviews on this application in his paper (LeRoy, 1989). He implies that investing in an efficient market is a fair game. Furthermore, Malkiel provided a more explicit definition of the market efficiency based on the Fama's definition. With respect to the information set, the asset prices must not be influenced by releasing that information to the public. In addition, no profit can be earned from trading assets in accordance with that information set (Malkiel, 2003). Malkiel's definition provides two practical methods to examine whether a market is efficient or not. The first method is measuring the change of asset prices if the information is released to all participants in the market. The second method is measuring the change of profits if market participants trade assets on that information set. Those theories indicate that the process of asset returns has martingale property after adjusting in a proper way. The martingale property led to the random walk hypothesis in modelling asset prices.

Back to 1963, Samuelson connected the stochastic integration with warrant pricing (Samuelson, 1963). Later, in 1973, Black and Scholes constructed an explicit valuation formula for pricing European options under the efficient market assumption (Black & Scholes, 1973). Black and Scholes assumed that: (1) the short-term interest rate and variance of the stock are constant; (2) the stock market is efficient, so the stock returns move with a random walk in a continuous-time path and the distribution of the returns is log-normal; (3) no transaction costs and penalties occur during the trading process; (4) no other payments (e.g. dividends etc.) are made; (5) options are "European" type, which means it only can be exercised at maturity date; (6) participants can sell or buy any fractions of the products at allowed time. In the same year, Merton extended the equilibrium model of asset prices during the continuous-time under more dynamic conditions (Merton, 1976). These two papers have become the milestone of continuous-time pricing theory in the modern financial economics research.

The Black-Scholes model has been shown some obvious drawbacks since the American stock market crashed in 1987 (Bates, 2000). Many empirical studies on

the dynamics of stock returns put forward that: (1) the S&P 500 Index returns do not follow a log normal distribution but a leptokurtic distribution which has left skewness, higher kurtosis and fatter tails than a log-normal distribution (Kou, 2002, 2008); (2) the volatility of stock returns is time-varying (Merton, 1973) and moves randomly (e.g. Fama, 1965; Hull & White, 1987); (3) the implied volatility curve shows a smile shape and the smile shape changes at different maturities[6] (Kou, 2008); (4) the volatility of returns has clustering effect[7] because the volatility is auto-correlated but the stock returns have independent increments (Kou, 2008); and (5) jumps (large fluctuation) occur in continuous stock price movements due to risks, especially during the financial crisis period (Merton & Samuelson, 1974; Bates, 2000; Eraker, Johannes & Polson, 2003). Therefore, the constant volatility and log-normal distribution assumptions are not adequate to describe the features of the return dynamics.

A diffusion model with constant volatility does not agree with the observed facts on the dynamics of the asset returns, so stochastic volatility is taken account into the asset pricing theory. According to the researches on the dynamics of stock returns, we know that the high kurtosis of the stock return distribution is caused by the fluctuation of the volatility in the stock returns (Bates, 1996; Eraker, 2004). Further, the left skewness in the return distribution is due to the relationship between stock returns and variance (Bates, 1996). In addition, the stochastic volatility can cope with the clustering effect and smile effect (Kou, 2008).

The history of stochastic volatility models can be traced back to 1970s. In the middle of 1970s, Merton introduced the deterministic function of time to replace the

[6]The option prices which are calculated based on the Black-Scholes model do not exactly match the market data because the graph of implied volatility against strike price of the option with the same maturity date exhibits the smile curve (U-shape). The smile curve discloses that the prices calculated based on the Black-Sholes model with constant volatility are greater than those of out-of-money puts and those of in-the-money calls (Fouque, Papanicolaou & Sircar, 2000).

[7]The clustering property in asset price variation was initially observed by Mandelbrot. He clearly stated that the large changes in asset prices are likely to go after large changes and small changes likely to go after small changes (Mandelbrot, 1963).

constant volatility (Merton, 1973). In another paper of Merton, he added a finite jump process into the stock pricing diffusion model to handle large fluctuation of the stock price in the continuous sample path (Merton, 1976). In the 1980s, stochastic volatility was considered into stock pricing models (e.g. Hildebrand, 1987; Johnson & Shanno, 1987). However, it was very difficult to derive a closed-form option price formula from the model with stochastic volatility at that time and only numerical solutions could be obtained. In 1990s, Heston provided an approach of deriving analytical expression for pricing European style options from his stochastic volatility model by using the fast Fourier transform technique (Heston, 1993). Since then, this approach has been widely applied in quantitative finance field, refer to in Bates (1996); and Bakshi, Cao and Chen (1997).

Since late 1990s, stochastic volatility models have dominated the literature on the financial derivatives pricing because this type of models can provide a more flexible structure. In a majority of stochastic volatility models, people assume that the volatility follows an Ornstein-Uhlenbeck process or a CIR process (e.g. Stein & Stein, 1991; Heston, 1993; Scott, 2002, etc.). This kind of volatility specification is consistent with the empirical study result that the volatility of the asset returns exhibits mean reversion property throughout the time (e.g. Grünbichler & Longstaff, 1996; Scott, 2002; Kou, 2008; Psychoyios & Dotsis, 2010; Mencía & Sentana, 2013). In the meanwhile, the finite-activity jump processes are used for accommodating the sudden changes of the prices which are caused by the rare events (Kou, 2002; Eraker, 2004; Bates, 2000; Pan, 2002, etc.). In addition, Eraker (2004), Bates (2000), Pan (2002) and Lin (2007) figured out that the correlated jumps occur simultaneously in the return and volatility processes.

1.2.2 Literature Review on VIX and pricing volatility derivatives

First, as we know, the VIX is a key index for measuring the volatility of the U.S. stock market, thus pricing VIX can be treated as modelling the volatility of stock returns. The methodologies for pricing VIX derivatives can be classified into two directions. In the first direction, VIX derivatives pricing formulae are derived from the instantaneous volatilities of the S&P 500 index, which can provide a clear picture of the relationship among the S&P 500 index, VIX and VIX derivatives. Researchers in this direction believe that the instantaneous volatility process has a mean-reverting property. In addition, some of them add finite-activity jumps into the S&P 500 return process, or instantaneous volatility process, or both of them. However, it is challenging to derive an analytical VIX derivative pricing formula when more and more notable features are added to a model. For more information about research work in this direction, we refer the readers to Zhang and Zhu (2006), Lin (2007), Zhu and Lian (2012), Goutte, Ismail and Pham (2017), Luo, Zhang and Zhang (2019), etc.

In the other direction, researches directly model VIX dynamics and then derive the derivatives pricing formulae from the proposed models. In this direction, they employ different types of mean-reverting processes, such as the squared-root mean-reverting process, the arithmetic mean-reverting process, and the geometric mean-reverting process, in modeling the dynamics of the VIX. In addition, Psychoyios and Dotsis (2010); Mencía and Sentana (2013); Kaeck and Alexander (2013) and Goard and Mazur (2013) found clear evidence of jumps in the VIX return process, in which a compound Poisson process is employed to characterize finite-activity jumps triggered by influential financial events. As an advantage, the analytical pricing formulae for VIX derivatives can be obtained based on even more complex VIX models. For example, Mencía and Sentana (2013) extended previous models into a two-factor model with three major features of VIX dynamics: central tendency, stochastic volatility of VIX

and finite-activity jumps. They documented that their two-factor model works better than other existing models, in terms of pricing VIX options and futures.

Second, the growth in VXX ETN trading (the derivative of VIX futures) has caught the attention of researchers and some research papers regarding to the VXX have emerged in recent years (e.g. Bao, Li & Gong, 2012; Gehricke & Zhang, 2019, 2020; Grasselli & Wagalath, 2020, etc.). Bao et al. (2012) and Alexander, Kapraun and Korovilas (2015) revealed that the VXX highly correlates to the VIX at a positive level. Due to the close relationship between the VXX and the VIX, VXX modeling[8] can be inspired from the way of modeling VIX, which has been widely discussed in the literature (e.g., Zhang & Zhu, 2006; Zhu & Lian, 2012; Mencía & Sentana, 2013; Park, 2016; Cao, Ruan, Su & Zhang, 2019, etc.). Bao et al. (2012) modeled the VXX based on the dynamics of the logarithm of VXX returns. They considered the default risk factor and positive volatility skew in their models. They found that the mean-reverting model with correlated stochastic volatility and finite-activity jumps outperforms its competitors. This paper has become one of the most influential papers related to the VXX research.

Third, the S&P 500 variance futures are another relatively new volatility products. Currently, some papers have studied on variance futures contracts with the old specifications [9]. For example, Zhang and Huang (2010) derived the formula for pricing S&P 500 3-month variance futures based on its underlying model. In their research, they found that there is a linear relationship between the price of variance futures and the value of VIX squared. They also detected that volatility risk premium is negative. Chang, Jimenez-Martin, McAleer and Amaral (2013) found it is difficult to accurately model the dynamics of the returns of the 3-month and 12-month variance futures,

[8]Gehricke and Zhang (2018) propose a VXX model based on the underlying (SPX and VIX) dynamics. They provide more information about the relationship among the SPX, VIX and VXX.

[9]The old version variance futures contracts are classified into two kinds, the S&P 500 3-month and 12-month variance futures which are delisted on 10 December 2012 and 17 March 2011, respectively.

because of the small trading volume of the variance futures. Therefore, they applied three conditional volatility models (GARCH, GJR and EGARCH) with three different probability densities (Gaussian, Student-t and Generalized Normal distribution errors) for modelling volatility for the 3- and 12-month of variance futures. They showed that the values of the VIX are highly correlated to the prices of the 3-month variance futures and 12-month variance futures. However, pricing the variance futures with the new specifications has not been studied in the literature yet.

In the remainder part of this sub-section, we present the extensive literature review on applications of pure infinite-activity jump Lévy processes in the financial modelling field. There has been a growing interest in pricing stock options with infinite-activity jump processes (e.g. Clark, 1973; Geman, 2002; Carr, Geman, Madan & Yor, 2003; Carr & Wu, 2003, 2004; Lian, Zhu, Elliott & Cui, 2017, etc.). Some pieces of credible evidence show that infinite-activity jumps (a.k.a. high-frequency jumps) occur in the dynamics of financial asset returns. For instance, Wu (2011) figured out that small jumps frequently appear in the volatility dynamics based on the analysis of high-frequency S&P 500 returns. As another example, Yang and Kanniainen (2017) revealed that the models with infinite-activity jump processes can fit the S&P 500 option data better. Moreover, Wu (2011), Todorov and Tauchen (2011b) and Cao et al. (2019) discovered that small oscillations frequently appear in the dynamics of S&P 500 volatility as well. The major advantage of infinite-activity jump processes over finite-activity jump processes is that they are able to capture not only finite activities but also infinite activities within a finite time interval.

Several infinite-activity jump Lévy processes can be applied in financial modelling. For example, the generalized hyperbolic Lévy process with five parameters is one of the most fundamental Lévy processes (Barndorff-Nielsen, 1978). It has two impressive sub-classes, the hyperbolic Lévy process (Eberlein, Keller & Prause, 1988) and the normal inverse Gaussian process (NIG) (Barndorff-Nielsen, 1997). The main difference

between these sub-classes is the tail behaviour: the tails of the NIG are thicker than those of the hyperbolic Lévy process. The variance gamma process (VG) is a Lévy process with a finite variation sample path which is distinguishing from the NIG that has an infinite variation sample path (Cont & Tankov, 2004). Both of the VG and the NIG can be viewed as time-changed Wiener processes [10] but with different subordinators, i.e., the gamma process and the inverse Gaussian process, respectively. They can capture the asymmetric decay in the exponential tail behavior but behave differently in the activity rate of the small jumps (Geman, 2002). Another Lévy jump process with infinite divisibility, popularly applied in quantitative finance, is the Carr-Geman-Madan-Yor (CGMY) model introduced by Carr, Geman, Madan and Yor (2002). This is a generalization of the variance gamma process by adding a parameter Y permitting finite-activities or infinite-activities with a finite or infinite variation. When $Y < 0$, the process permits finite-activity like a compound Poisson process. On the other hand, if Y is between 0 and 2, the process permits infinite activities. The process has a finite variation when $0 < Y < 1$ and an infinite variation when $1 < Y < 2$ (Carr et al., 2002). Unlike the Lévy processes mentioned above, a generalized tempered stable process (GTS) does not have a Gaussian part and is softened by an exponential factor (Cont & Tankov, 2004). It exhibits a symmetrical local behavior, which can generate rich jump structures based on its parameter values.

1.3 Research Questions

In this research, we mainly focus on building mathematical models for accurately and efficiently valuing volatility derivatives (e.g. VIX options, VIX futures, VXX options and S&P 500 variance futures) under different types of Lévy processes. By following

[10] The time-changed Lévy process refers to a stochastic process with a randomness time clock, where the time process is a non-decreasing right-continuous process.

the literature review in the last section, we propose the following research questions.

Question 1: Dose a model with an infinite-activity jump structure capture the characteristics in the dynamics of the VIX or the price of volatility derivatives better than a model with a finite-activity jump structure?

Question 2: Are there any realistic factors with regard to market circumstances, such as stochastic volatility with time-varying mean, can be considered into our models? Is it worth putting these realistic features into our models? What kind of model specifications can perform better in volatility derivatives pricing?

Question 3: The cost of implementation increases as the models become more and more complicated. Therefore, it is necessary to investigate which calibration methods and numerical methods can be applied in our cases. What is the best calibration method for balancing the computational cost and accuracy of estimation results?

Question 4: How to derive closed-form pricing formulae for different types of volatility derivatives from the proposed models?

Question 5: To extend the first question, how does a model with double jumps perform in the pricing of volatility derivatives?

1.4 Research Contributions and Thesis Organization

The main contributions of this research provide critical evidence that infinite-activity jump models, particularly the jump processes with an infinite variation sample path, overall outperform finite-activity jump models. Hence infinite-activity jumps cannot be ignored in pricing volatility derivatives. The presence of infinite-activity jumps implies that not only big jumps but also frequent small jumps occur in the volatility derivatives market. From the economic aspect, infinite-activity Lévy jump processes can accommodate the high-frequency occurrences of small events and the microstructure of the VIX derivatives market, including some rare events (e.g. financial crisis, etc.).

We also notice that the advantage of models with infinite-activity jump structures is highlighted in the VIX and VXX options pricing. In addition, both Brownian motion and infinite-activity jumps are indispensable parts of the models. Furthermore, the stochastic mean is a crucial feature of the dynamics of the VIX index and the variance process of the S&P 500 index returns.

The remainder of this thesis is organized as follows. Chapter 2 illustrates the mathematical tools (e.g. mathematical concepts, methods, etc.) and financial modelling approaches and numerical methods, which will be used to solve our proposed questions. Chapter 3 constructs the two-factor models with infinite-activities jumps of the VIX and discusses the impacts of infinite-activities jump processes on pricing VIX derivatives . Chapter 4 discusses the performance of stochastic volatility models with different jump structures in the VXX options pricing to investigate the impact of the finite-activity and infinite-activity jumps on American-style option pricing. Chapter 5 derives the pricing formulae of variance futures based on models with double jumps and other model specifications. It figures out the features of the dynamics of the S&P 500 variance futures prices.

Chapter 2

Preliminaries

Modern probability theory has become a fundamental tool for quantifying various financial products. The risk-neutral valuation approach is a paramount technique for formulating volatility derivatives based on risk-neutral scenarios. Furthermore, numerical calculation methods are very effective mathematical tools for solving some mathematical problems, especially for problems which do not have explicit expression. This chapter presents theoretical foundation for our research. The key references for this chapter are Hildebrand (1987), Rossi (2018), Schoutens (2003), Cont and Tankov (2004), Jacod and Protter (2004), Øksendal and Sulem (2005), and Shreve (2008).

2.1 Mathematical Tools

In this section, we present important mathematical concepts and results in probability theory, which will be used in the subsequent chapters.

2.1.1 Probability Measures

Measures play a prominent role when we define a stochastic process such as a Lévy process or a jump process. In real life, the term “measure" can be interpreted as the

length of an cord, the area of a section, and so on. In probability theory, a measure on a set can be interpreted as the size of set (Cont & Tankov, 2004). In general, this theory includes three key elements, the state space, events and probability distribution. The following content in this sub-section is associated to these three elements. The state space contains all possible outcomes of an experiment, denoted by Ω. We use 2^{Ω} to express the collection of all subsets of Ω. Let $\mathcal{F}$ be a subset of 2^{Ω}. Elements of $\mathcal{F}$ are called events. Before giving the definition of a probability measure, we introduce the notion of a σ-algebra (Shreve, 2008, p.51).

Definition 2.1.1 (σ-algebra) *$\mathcal{F}$ is a σ-algebra if it satisfies*

1. $\varnothing \in \mathcal{F}$,
2. *if $A \in \mathcal{F}$ then $\Omega \setminus A \in \mathcal{F}$, where $\Omega \setminus A$ is the complement of A, and*
3. *if $A_1, A_2, A_3, ...$ is a sequence of events in $\mathcal{F}$, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.*

When $\mathcal{F}$ is a σ-algebra on Ω, according to the properties 2 and 3 in Definition 2.1.1. If $A_1, A_2, ...$ is a sequence of events in $\mathcal{F}$, we can conclude that $\bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$. In addition, the whole space Ω is also in $\mathcal{F}$ because Ω is the complement of $\varnothing$. We call $(\Omega, \mathcal{F})$ a measurable space. Now, we introduce the definition of a probability measure on the measurable space $(\Omega, \mathcal{F})$.

Definition 2.1.2 (Probability measure) *A probability measure defined on $(\Omega, \mathcal{F})$ is a function $\mathbb{P}$: $\mathcal{F} \to [0, 1]$ that satisfies:*

1. $\mathbb{P}(\Omega) = 1$, *and*
2. *(countable additivity) for every sequence $(A_i)_{i \geq 1}$ of pairwise disjoint elements of $\mathcal{F}$ (that is, $A_n \bigcap A_m = \varnothing$ whenever $n \neq m$), one has*

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i). \tag{2.1}$$

In Definition 2.1.2 the value of $\mathbb{P}(A)$ represents the probability of the event A. Then, $(\Omega, \mathcal{F}, \mathbb{P})$ is called a probability space. Based on the second condition in Definition 2.1.2, we have $\mathbb{P}(\varnothing) = 0$.

Assume that $\mathbb{P}$ and $\mathbb{Q}$ are two probability measures, defined on the same measurable space $(\Omega, \mathcal{F})$. The measure $\mathbb{P}$ is absolutely continuous with respect to $\mathbb{Q}$ (Cont & Tankov, 2004, p.25), if for any measurable set $A \in \mathcal{F}$,

$$\mathbb{Q}(A) = 0 \Rightarrow \mathbb{P}(A) = 0. \tag{2.2}$$

Theorem 2.1.1 (Radon-Nikodym theorem) *If $\mathbb{P}$ is absolutely continuous with respect to $\mathbb{Q}$ then there exists a $\mathbb{Q}$ measurable function Z: $\Omega \to [0, \infty)$ such that for any measurable set A,*

$$\mathbb{P}(A) = \int_A Z d\mathbb{Q}. \tag{2.3}$$

The function Z is called the density or Radon-Nikodym derivative of $\mathbb{P}$ with respect to $\mathbb{Q}$ and denoted as $d\mathbb{P}/d\mathbb{Q}$.

The integral in Eq. (2.3) is Lebesgue integral which is defined in terms of using the measurable functions. Now, two probability measures $\mathbb{P}$ and $\mathbb{Q}$ are said to be equivalent if they are absolutely continuous with respect to each other, that is,

$$\mathbb{P} \sim \mathbb{Q} \iff \left[\forall A \in \mathcal{F}, \mathbb{P}(A) = 0 \iff \mathbb{Q}(A) = 0\right]. \tag{2.4}$$

2.1.2 Random Variables

A random variable refers to an unknown value that fluctuates in accordance with which random event occurs. On a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a real-valued random variable refers to a measurable function X on Ω with the property that for each Borel subset

$B \subseteq \mathbb{R}$, the subset of Ω given by the form of,

$$\{X \in B\} = \{\omega \in \Omega; X(\omega) \in B\}, \tag{2.5}$$

is in the σ−algebra $\mathcal{F}$ [1] (Shreve, 2008, p.7). Let $\omega \in \Omega$ be a scenario of randomness. Then, $X(\omega)$ expresses the outcome of a random variable when the scenario ω occurs (Cont & Tankov, 2004, p. 27). The distribution measure of X (Shreve, 2008, p.9), denoted as μ_X, is defined as,

$$\mu_X(B) = \mathbb{P}\{X \in B\}, \tag{2.6}$$

where $B \subseteq \mathbb{R}$ and the set B can be a number or a set of real numbers.

The real-valued random variable X can be characterized by its cumulative distribution function (cdf), $F_X(x)$, which is given by:

$$F_X(x) = \mathbb{P}\{X \leq x\} = \mu_X(-\infty, x], \quad x \in \mathbb{R}. \tag{2.7}$$

So, if the cdf F is known, then we have $\mu_X(x, y] = F_X(y) - F_X(x)$ for any $x \in \mathbb{R}$, $y \in \mathbb{R}$ and $x < y$. On the other hand, the relationship between the cdf and the probability density function (pdf) of a random variable X is given by,

$$\mu_X[a, b] = \mathbb{P}\{a \leq X \leq b\} = F_X(b) - F_X(a) = \int_a^b f_X(x)dx, \tag{2.8}$$

where $-\infty < a \leq b < \infty$, $f_X(x)$ is the pdf which is a non-negative function.

The following properties hold for random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$(Schoutens, 2003, p.16).

[1] Note: the value range for a random variable can also be allowed to be from $+\infty$ to $-\infty$ in some situations.

Theorem 2.1.2 *Let X be a random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.*

(a) If X takes only finitely many values $y_0, y_1, \ldots, y_n$, then

$$\int_\Omega X(\omega) d\mathbb{P}(\omega) = \sum_{k=0}^{n} \mathbb{P}\{X = y_k\}. \tag{2.9}$$

*(b) (**Integrability**) X is integrable if and only if*

$$\int_\Omega X(\omega) d\mathbb{P}(\omega) < \infty. \tag{2.10}$$

Now let Y be another random variable on $(\Omega, \mathcal{F}, \mathbb{P})$.

*(c) (**Comparison**) If $X < Y$ almost surely (i.e., $\mathbb{P}\{X < Y\} = 1$) and if $\int_\Omega X(\omega) d\mathbb{P}(\omega)$ and $\int_\Omega Y(\omega) d\mathbb{P}(\omega)$ are defined, then*

$$\int_\Omega X(\omega) d\mathbb{P}(\omega) < \int_\Omega Y(\omega) d\mathbb{P}(\omega). \tag{2.11}$$

In particular, if $X = Y$ almost surely and one of the integrals is defined, then they are both defined and

$$\int_\Omega X(\omega) d\mathbb{P}(\omega) = \int_\Omega Y(\omega) d\mathbb{P}(\omega). \tag{2.12}$$

*(d) (**Linearity**) If α and β are real constants and X and Y are integrable, or if α and β are non-negative constants and X and Y are non-negative, then*

$$\int_\Omega (\alpha X(\omega) + \beta Y(\omega)) \, d\mathbb{P}(\omega) = \alpha \int_\Omega X(\omega) d\mathbb{P}(\omega) + \beta Y(\omega) d\mathbb{P}(\omega). \tag{2.13}$$

As stated in the beginning of this section, the value of a random variable is unknown because events happen randomly. However, we can calculate the average value for a

random variable, which is known as the expectation of a random variable.

Definition 2.1.3 (Expectation) *Let X be a random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The expectation of X is defined as*

$$\mathbb{E}^{\mathbb{P}}[X] = \int_{\Omega} X(\omega) d\mathbb{P}(\omega). \tag{2.14}$$

Note that if the expectation $\mathbb{E}^{\mathbb{P}}[X]$ is defined, then at least one of $\mathbb{E}^{\mathbb{P}}[X^+]$ or $\mathbb{E}^{\mathbb{P}}[X^-]$ is finite, where X^+ and X^- represent the positive and negative parts of the random variable X, such that

$$X^+(\omega) = \max\{X(\omega), 0\}, \quad X^-(\omega) = \max\{-X(\omega), 0\}. \tag{2.15}$$

The characteristic function of a random variable is the Fourier transform of its pdf (Cont & Tankov, 2004, p.29). It is always continuous and plays an important role for studying the behaviour of random variables.

Definition 2.1.4 (Characteristic function) *The characteristic function of a random variable X is the function $\varphi_X : \mathbb{R} \to \mathbb{C}$ defined by*

$$\varphi_X(\theta) = \mathbb{E}\left[e^{i\theta x}\right] = \int_{\mathbb{R}} e^{i\theta x} dF_X(x) = \int_{\mathbb{R}} e^{i\theta X} f_X(x) dx, \quad \forall \theta \in \mathbb{R}, \tag{2.16}$$

where F_X and f_X are the cumulative distribution function and probability density function of X, respectively.

If the characteristic functions of two random variables are the same, then their distributions are identical. If the random variables, $X_1, X_2, ..., X_n$ are independent, the characteristic function of the sum of these random variables is the product of the

characteristic functions of the individuals, that is,

$$\varphi_X(\theta) = \prod_{i=1}^{n} \varphi_{X_i}(\theta), \tag{2.17}$$

where $X = \sum_{i=1}^{n} X_i$. The n-th moment of a random variable X on $\mathbb{R}$ is defined by (Cont & Tankov, 2004, p.30)

$$m_n(X) = \mathbb{E}[X^n]. \tag{2.18}$$

Similarly, we have the n-th absolute moment of X as

$$m_n(|X|) = \mathbb{E}[|X|^n]. \tag{2.19}$$

The n-th centered moment μ_n is the n-th moment of $X - \mathbb{E}[X]$, that is

$$\mu_n(x) = \mathbb{E}\left[(X - \mathbb{E}[X])^n\right]. \tag{2.20}$$

At the end of this sub-section, we take a standard normal random variable X as an example. The pdf $f_X(x)$ of X is defined as

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \tag{2.21}$$

and its corresponding cdf $F_X(x)$ is

$$F_X(x) = \int_{-\infty}^{x} f_X(y)dy. \tag{2.22}$$

The cdf is strictly increasing, so we have $F_X(F_X^{-1}(y)) = y$. The Fourier transform of a standard normal density $f_X(x)$ in Eq. (2.21) is,

$$\hat{f}_X(\theta) = \int_{-\infty}^{\infty} f_X(x) e^{-i\theta x} dx = e^{-\frac{1}{2}(\sigma\theta)^2}. \tag{2.23}$$

Then, the characteristic function of this random variable is computed as

$$\varphi_X(\theta) = \mathbb{E}\left[e^{i\theta X}\right] = \hat{f}_X(-\theta) = e^{-\frac{1}{2}(\sigma\theta)^2}. \tag{2.24}$$

2.1.3 Conditional Expectation

If a random experiment is performed and the outcome of the experiment is set as ω, we may be given some information but not enough to obtain the exact vale of ω. However, we can narrow down the possibility of ω occurring based on this information. Now, let us consider $\sigma-$algebras, $\mathcal{F}_0, \mathcal{F}_1, \mathcal{F}_2, ..., \mathcal{F}_m, ..., \mathcal{F}_n$, indexed by time, where $m < n$. Since the time goes from time 0 to time n, we can gain more information. This implies that $\mathcal{F}_n$ contains more information than the other $\sigma-$algebras in this series, $\mathcal{F}_i$, for $i = 0, 1, ..., n-1$. This series of $\sigma-$algebras, $\mathcal{F}_0, \mathcal{F}_1, \mathcal{F}_2, ..., \mathcal{F}_n$, is called a filtration. The formal definition of the filtration is given below.

Definition 2.1.5 (Filtration) *Let Ω be a nonempty set. Let T be a fixed positive number, and assume that for each $t \in [0, T]$ there is a $\sigma-$algebra $\mathcal{F}_t$. Assume further that if $s \leq t$, then every set in $\mathcal{F}_s$ is also in $\mathcal{F}_t$. Then we call the collection of $\sigma-$algebras $\mathcal{F}_{0\leq t\leq T}$, a filtration.*

There is an extreme case that the information in $\mathcal{F}$ can not provide any clues which are used for evaluating the random variable X. In this case, we say X is independent of the information set $\mathcal{F}$. The formal definitions (Shreve, 2008, p.51) about independence are given below.

Definition 2.1.6 (Independence) *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $\mathcal{G} \subset \mathcal{F}$ and $\mathcal{E} \subset \mathcal{F}$ be two $\sigma-$algebras. We say $\mathcal{G}$ and $\mathcal{E}$ are independent if*

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B) \quad \textit{for all } A \in \mathcal{G},\ B \in \mathcal{E}. \tag{2.25}$$

Assume that X and Y are random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ and the $\sigma-$algebras generated by X and Y, are denoted by $\sigma(X)$ and $\sigma(Y)$, respectively. We say X and Y are independent if $\sigma(X)$ and $\sigma(Y)$ are independent. If the $\sigma(X)$ and $\mathcal{G}$ are independent, then we can say X is independent of the information $\mathcal{G}$.

Theorem 2.1.3 *Let X and Y be independent random variables. Then, the following conditions are equivalent.*

1 The joint distribution measure factors:

$$\mu_{X,Y}(A \times B) = \mu_X(A) \cdot \mu_Y(B) \quad \textit{for all } A \subseteq \mathbb{R}, B \subseteq \mathbb{R}. \tag{2.26}$$

2 The joint cumulative distribution function factors:

$$F_{X,Y}(a, b) = F_X(a) \cdot F_Y(b) \quad \textit{for all } a \in \mathbb{R}, b \in \mathbb{R}. \tag{2.27}$$

3 The joint moment-generating function factors:

$$\mathbb{E}[e^{uX+vY}] = \mathbb{E}[e^{uX}] \cdot \mathbb{E}[e^{vY}], \tag{2.28}$$

for all $u \in \mathbb{R}$, $v \in \mathbb{R}$ for which the expectations are finite.

4 The joint density factors:

$$f_{X,Y}(x, y) = F_X(x) \cdot F_Y(y) \quad \textit{for almost every } x \in \mathbb{R}, y \in \mathbb{R}. \tag{2.29}$$

5 The expectation factors:

$$\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y] \quad \textit{if } \mathbb{E}[|XY|] < \infty. \tag{2.30}$$

In the middle situation, the value of a random variable X can be estimated but not accurately in accordance with the information $\mathcal{G}$. This estimation is called the conditional expectation of X. So, we can define the conditional expectation (Shreve, 2008, p.68) as follows.

Definition 2.1.7 (Conditional expectation) *Let X be a random variable that is either non-negative or integrable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The conditional expectation of X given $\mathcal{G}$, denoted $\mathbb{E}[X|\mathcal{G}]$, is any random variable that satisfies*

1 ***(Measurability)*** *$\mathbb{E}[X|\mathcal{G}]$ is $\mathcal{G}$-measurable, and*

2 ***(Partial averaging)***

$$\int_A \mathbb{E}[X|\mathcal{G}](\omega)d\mathbb{P}(\omega) = \int_A X(\omega)d\mathbb{P}(\omega) \quad \textit{for all } A \in \mathcal{G}. \tag{2.31}$$

Now, we assume that X and Y are random variables which are integrable on $(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{G} \subseteq \mathcal{F}$. Then, we can have four fundamental properties of conditional expectations (Shreve, 2008, P.69), which are listed below:

1 **(Linearity of conditional expectations)** If c_1 and c_2 are constants, then

$$\mathbb{E}[c_1X + c_2Y|\mathcal{G}] = c_1\mathbb{E}[X|\mathcal{G}] + c_2\mathbb{E}[Y|\mathcal{G}]. \tag{2.32}$$

2 **(Taking out what is known)** If XY are integrable and X is $\mathcal{G}$-measurable, then

$$\mathbb{E}[XY|\mathcal{G}] = X\mathbb{E}[Y|\mathcal{G}]. \tag{2.33}$$

3 **(Iterated conditioning)** If $\mathcal{H}$ is a sub-σ-algebra of $\mathcal{G}$, then,

$$\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}] = \mathbb{E}[X|\mathcal{H}]. \tag{2.34}$$

4 **(Independence)** If X is independent of $\mathcal{G}$, then

$$\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X]. \tag{2.35}$$

2.1.4 Stochastic Processes

A stochastic process is defined as a series of random variables which is indexed by time, denoted by $\{X_t\}_{t\geq 0}$, where time t can be either continuous or discrete (Cont & Tankov, 2004, p.49). In other words, a stochastic process can be viewed as a function of time t and randomness ω. The sample path of a stochastic process is a function of time for every observed value of the randomness ω, that is,

$$X_{\cdot}(\omega) : t \to X_t(\omega).$$

Assume random variables are defined on a continuous space $C([0,T],\mathbb{R})$. These random variables can construct stochastic processes. The usual norm on $C([0,T],\mathbb{R})$ (Cont & Tankov, 2004, p.49) is expressed as

$$\|f\|_\infty = \sup_{t\in[0,T]} |f(t)|, \tag{2.36}$$

where $f : [0,T] \to \mathbb{R}$. The most classical example of the stochastic processes with continuous sample path is Wiener process. In some cases, the sample may be discontinuous functions. So, a space which allows for the discontinuous cases are needed to be defined.

Definition 2.1.8 (Cádlág function) *A cádlág function $f : [0,T] \to \mathbb{R}$ is right-continuous with left limits: for every $t \in [0,T]$ the limits,*

$$f(t-) = \lim_{s\to t^-} f(s) \qquad f(t+) = \lim_{s\to t^+} f(s), \tag{2.37}$$

exist and $f(t) = f(t+)$.

Note, a cádlág function can be either continuous or discontinuous function. If a stochastic process is right-continuous with a left limit everywhere in its sample paths, then this process is called a cádlág process. Now, we discuss about the martingale property (Cont & Tankov, 2004, p.54) of a cádlág process on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with respect to an information flow $\{\mathcal{F}_t\}_{t\geq 0}$.

Definition 2.1.9 (Martingale) *Let a cádlág process be* $\{X_t\}_{0\leq t\leq T}$ *on a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$. *The process is martingale if* $\mathbb{E}[|X_t|]$ *is finite for any* $t \in [0, T]$ *and*

$$\forall s > t, \qquad \mathbb{E}[X_s|\mathcal{F}_t] = X_t. \tag{2.38}$$

According to this definition, we know that the best estimation for a future value of a martingale is its present value. The martingale definition makes sense only if we specify the probability measure and the information flow.

Wiener process, $\{W_t\}_{t\geq 0}$, is a classical example of a martingale. The formal definition of Wiener process is given below.

Definition 2.1.10 (Wiener process) *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space. For every* $\omega \in \Omega$, *suppose there is a continuous function* $W(t)$ *of* $t \geq 0$ *that satisfies* $W(0) = 0$ *and that depends on* ω. *Then* $W(t)$, $t \geq 0$, *is a Wiener process if for all* $0 = t_0 < t_1 < \cdots < t_m$ *the increments*

$$W(t_1) = W(t_1) - W(t_0), W(t_2) - W(t_1), \ldots, W(t_m) - W(t_{m-1}), \tag{2.39}$$

are independent and each of these increments is normally distributed with

$$\mathbb{E}[W(t_i) - W(t_{i-1})] = 0, \tag{2.40}$$

$$Var[W(t_i) - W(t_{i-1})] = t_i - t_{i-1}. \tag{2.41}$$

A Wiener process in the literature is also called a Brownian motion. It has the following three important properties (Shreve, 2008, p.96).

- The increments of the process are independent and follow a normal distribution with mean as (2.40) and variance as (2.41).
- The random variables $W(t_1)$, $W(t_2)$, ..., $W(t_m)$ follow a joint normal distribution with zero mean and co-variance matrix which is given by

$$\begin{bmatrix} \mathbb{E}[W^2(t_1)] & \dots & \mathbb{E}[W(t_1)]\mathbb{E}[W(t_m)] \\ \mathbb{E}[W(t_2)]\mathbb{E}[W(t_1)] & \dots & \mathbb{E}[W(t_1)]\mathbb{E}[W(t_m)] \\ \vdots & \vdots & \vdots \\ \mathbb{E}[W(t_m)]\mathbb{E}[W(t_1)] & \dots & \mathbb{E}[W^2(t_m)] \end{bmatrix} = \begin{bmatrix} t_1 & t_1 & \cdots & t_1 \\ t_1 & t_2 & \cdots & t_2 \\ \vdots & \vdots & & \vdots \\ t_1 & t_2 & \cdots & t_m \end{bmatrix} \tag{2.42}$$

- The random variables $W(t_1)$, $W(t_2)$, ..., $W(t_m)$ have the joint moment-generating function , $\phi(\mu_1, \mu_2, \dots, \mu_m)$, and given by,

$$\begin{aligned} &\phi(\mu_1, \mu_2, \dots, \mu_m) \\ &= \mathbb{E}\left[\exp\{\mu_m W(t_m) + \mu_{m-1} W(t_{m-1}) + \cdots + \mu_1 W(t_1)\}\right] \\ &= \exp\left\{\frac{1}{2}(\mu_1 + \mu_2 + \cdots + \mu_m)^2 t_1 + \frac{1}{2}(\mu_2 + \mu_3 + \cdots + \mu_m)^2 (t_2 - t_1)\right. \\ &\quad \left. + \frac{1}{2}(\mu_{m-1} + \mu_m)^2 (t_{m-1} - t_{m-2})\right\} \end{aligned} \tag{2.43}$$

2.1.5 Poisson Process

A Poisson process is a fundamental stochastic process, which is a pure jump process. The exponential random variables play very important roles in the construction of a Poisson process. An exponential random variable τ has the probability density, such that (Shreve, 2008, p.462),

$$f_\tau(t) = \begin{cases} \lambda e^{-\lambda t}, & t \geq 0, \\ 0, & t < 0, \end{cases} \tag{2.44}$$

where λ is a positive constant.

Proposition 2.1.1 *If $\{\tau_i\}_{i\geq 1}$ are independent exponential random variables with parameter λ then, for any $t > 0$ the random variable*

$$N_t = \inf\{n \geq 1, \sum_{i=1}^{n} \tau_i > t\} \tag{2.45}$$

follows a Poisson distribution with parameter λt:

$$\mathbb{P}(N_t = n) = e^{-\lambda t}\frac{(\lambda t)^n}{n!} \tag{2.46}$$

Then, the Poisson process is constructed based on a sequence of independent exponential random variables (Cont & Tankov, 2004, p.48).

Definition 2.1.11 (Poisson process) *Let $\{\tau_i\}_{i\geq 1}$ be a sequence of independent exponential random variables with parameter λ and $T_n = \sum_{i=1}^{n} \tau_i$. The process $\{N_t\}_{t\geq 0}$ defined by*

$$N_t = \sum_{n\geq 1} \mathbf{1}_{t\geq T_n} \tag{2.47}$$

is called a Poisson process with intensity λ.

The Poisson process $\{N_t\}_{t\geq 0}$ is a counting process, so N_t measures the number of jumps from time 0 to time t:

$$N_t = \#\{i \geq 1, T_i \in [0,t]\}. \tag{2.48}$$

If $0 \leq t < s$, then the distribution of Poisson increment $N_s - N_t$ is given by (Shreve, 2008, p.466)

$$\mathbb{P}\{(N_s - N_t) = k\} = \frac{\lambda^k (s-t)^k}{k!} e^{-\lambda(s-t)}, \quad k = 0, 1, 2, ... \tag{2.49}$$

Definition 2.1.12 (Poisson random measure) *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $E \subseteq \mathbb{R}$ and μ a given (positive) Radon measure μ on $(E, \mathcal{E})$. A Poisson random measure on E with intensity measure μ is an integer valued random measure:*

$$M : \Omega \times \mathcal{E} \to \mathbb{N} \quad \textit{and} \quad (\omega, A) \mapsto M(\omega, A),$$

such that

1. *For (almost all) $\omega \in \Omega$, $M(\omega, .)$ is an integer-valued Radon measure on E: for any bounded measurable $A \subset E$, $M(A) < \infty$ is an integer valued random variable.*

2. *For each measurable set $A \subset E$, $M(., A) = M(A)$ is a Poisson random variable with parameter $\mu(A)$:*

$$\mathbb{P}(M(A) = k) = e^{-\mu(A)} \frac{(\mu(A))^k}{k!}, \quad \forall k \in \mathbb{N} \tag{2.50}$$

3. *For disjoint measurable sets $A_1, ..., A_n \in \mathcal{E}$, the variables $M(A_1), ..., M(A_n)$ are independent.*

2.1.6 Lévy process

A Lévy process is a continuous time stochastic process with independent stationary increments (Cont & Tankov, 2004, p.68). It is defined below.

Definition 2.1.13 (Lévy process) *A cádlág stochastic process $\{X_t\}_{t\geq 0}$ on $(\Omega, \mathcal{F}, \mathbb{P})$ with values in $\mathbb{R}$ such that $X_0 = 0$ is called a Lévy process if it possesses the following properties:*

1. *Independent increments: for every increasing sequence of times $t_0 ... t_n$, the random variables $X_{t_0}, X_{t_1} - X_{t_0}, ..., X_{t_n} - X_{t_{n-1}}$ are independent.*
2. *Stationary increments: the law of $X_{t+h} - X_t$ does not depend on t.*
3. *Stochastic continuity: $\forall \varepsilon > 0$, $\lim_{h\to 0} \mathbb{P}\left(|X_{t+h} - X_t| \geq \varepsilon\right) = 0$.*

Definition 2.1.14 (Lévy measure) *Let $\{X_t\}_{t\geq 0}$ be a Lévy process on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The measure ν on $\mathbb{R}$ defined by:*

$$\nu(A) = \mathbb{E}\left[\#\left\{t \in [0,1] : \Delta X_t \neq 0, \Delta X_t \in A\right\}\right], \quad A \in \mathcal{B}(\mathbb{R}) \tag{2.51}$$

is called the Lévy measure of X: $\nu(A)$ is the expected number, per unit time, of jumps whose size belongs to A.

The Lévy measure can be either a finite measure or an infinite measure. A process has a finite measure, which means that the process generates a finite number of jumps. On the other hand, if the measure is an infinite measure, then the process may have an infinite number of jumps on the finite time period $[0, T]$. For an infinite Lévy process, we can have the following theorem in regards to the decomposition of an infinite Lévy process (Øksendal & Sulem, 2005, p.3).

Theorem 2.1.4 (Lévy-Itô decomposition) *Let $\{X_t\}_{t\geq 0}$ be a Lévy process on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and ν its Lévy measure, given by Definition 2.1.14. Then X_t has the decomposition,*

$$X_t = bt + \sigma W_t + \int_{|x|<1} x\tilde{N}(t, dx) + \int_{|x|\geq 1} xN(t, dx), \tag{2.52}$$

for some constants $b \in \mathbb{R}$, $\sigma \in \mathbb{R}$ and $\{W_t\}_{t\geq 0}$ is a Wiener process. Here,

$$\tilde{N}(dt, dx) = N(dt, dx) - \nu(dx)dt$$

is the compensated Poisson random measure of X, where ν is a Radon measure on $\mathbb{R}\backslash\{0\}$ and satisfies:

$$\int_{|x|\leq 1} |x|^2 \nu(dx) < \infty, \qquad \int_{|x|\geq 1} \nu(dx) < \infty.$$

A Lévy process can be denoted by a triplet (b, σ^2, ν) which is called characteristic triplet. The next theorem, Lévy-Khintchine formula, provides a method for characterizing random variables with infinitely divisibility by using their characteristic functions (Øksendal & Sulem, 2005, p.4).

Theorem 2.1.5 (Lévy-Khintchine formula) *Let $\{X_t\}_{t\geq 0}$ be a Lévy process on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with the characteristic triplet (b, σ^2, ν), where $b \in \mathbb{R}$, $\sigma \in \mathbb{R}$, and a measure satisfying $\nu(\{0\}) = 0$ and $\int_{\mathbb{R}} \min(1, x^2)\nu(dx) < \infty$, then*

$$\varphi_{X_t}(\theta) = \mathbb{E}\left[e^{i\theta X_t}\right] = e^{t\psi(\theta)}, \quad \theta \in \mathbb{R}, \tag{2.53}$$

where ψ is called the characteristic exponent of X and is a continuous function. It is given by,

$$(\theta) \equiv ib\theta - \frac{1}{2}\sigma^2\theta^2 + \int_{\mathbb{R}_0^d} \left(e^{i\theta x} - 1 - i\theta x \mathbf{1}_{|x|\leq 1}\right) \nu(dx). \tag{2.54}$$

Theorem 2.1.5 is another important tool for Lévy processes. It can be used for finding the differential form for a Lévy process.

Theorem 2.1.6 (The 1-dimensional Itô formula) *Let $\{X_t\}_{t\geq 0}$ be a Lévy process on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ of the form,*

$$dX_t = \alpha(t,\omega)dt + \sigma(t,\omega)dW_t + \int_{\mathbb{R}} \gamma(t,x,\omega)\bar{N}(dt,dx), \tag{2.55}$$

where

$$\bar{N}(dt,dx) = \begin{cases} N(dt,dx) - \nu(dx)dt, & \text{if } |x| \leq 1, \\ N(dt,dx), & \text{if } |x| > 1. \end{cases} \tag{2.56}$$

Let $f \in C^2(\mathbb{R}^2)$ and define $Y_t = f(t, X_t)$. Then $\{Y_t\}_{t\geq 0}$ is again a Lévy process and

$$\begin{aligned} dY_t = {} & \frac{\partial f}{\partial t}(t,X_t)dt + \frac{\partial f}{\partial X}(t,X_t)\left[\alpha(t,\omega)dt + \sigma(t,\omega)dW_t\right] \\ & + \frac{1}{2}\sigma^2(t,\omega)\frac{\partial^2 f}{\partial X^2}(t,X_t)dt \\ & + \int_{|x|<1}\left\{f(t,X_{t^-}+\gamma(t,x,\omega)) - f(t,X_{t^-}) - \frac{\partial f}{\partial X}(t,X_{t^-})\gamma(t,x,\omega)\right\}\nu(dx)dt \\ & + \int_{\mathbb{R}}\{f(t,X_{t^-}+\gamma(t,x,\omega)) - f(t,X_{t^-})\}\bar{N}(dt,dx). \end{aligned} \tag{2.57}$$

Note that $N(dt,dx)$ counts the finite number of jumps whose magnitudes are in the set A ($A \subset \mathbb{R}$) within the time interval $[t, t+dt]$ and $\nu(A)$ is a Lévy measure which is defined as $\nu(A) = \mathbb{E}[N(1,A)]$ and is the compensation of $N(dt,A)$. When $|x| < 1$, the jump component, $N(dt,dx) - \nu(dx)dt$, is martingale. In addition, the jump part displays finite activity, if the following integral is finite (Carr & Wu, 2004),

$$\int_{\mathbb{R}\setminus\{0\}} \nu(dx) = \lambda < \infty, \tag{2.58}$$

where λ is known as jump intensity.

2.1.7 Pure Jump Lévy Process

Generally, a Lévy process is characterized by its Lévy triplet (b, σ^2, ν) which consists of the drift, variance and Lévy measure. A pure jump Lévy process refers to a process which only involves a drift and a jump component. In other words, the Lévy triplet for this type of processes is denoted by $(b, 0, \nu)$. A jump process can be either a finite-activity or an infinite-activity jump process, which is determined by its Lévy measure (Huang & Wu, 2004). A finite-activity jump process generates a finite number of jumps in each finite time interval and captures relatively large movements. On the other hand, an infinite-activity jump process can have infinitely many small jumps within a finite time interval. A pure Lévy jump process $\{J_t\}_{t\geq 0}$ can be written as,

$$dJ_t = bdt + \int_{y\in\mathbb{R}} x\bar{N}(dt, dx), \tag{2.59}$$

where x refers to the magnitude of a jump in $\{J_t\}_{t\geq 0}$, $\bar{N}(dt, dx)$ is specified in Eq. (2.56) and $\mathbb{E}[J_t] = bt$. Moreover, a sample path of a pure Lévy infinite-activity jump process may have either a finite variation or an infinite variation (Sato, 1999), which depends on whether $\int_{\mathbb{R}\setminus\{0\}} \min(1, |x|)\nu(dx)$ is bounded or not. In addition, $\min(1, |x|)\nu(dx)$ implies the jump behavior around the singular point at $y = 0$ (Applebaum, 2004).

There are a variety of pure jump Lévy processes which can be applied in the financial asset pricing theory. In this section, we introduce four types of jump processes, including compound Poisson jump process of Merton (1976) (MJ), variance gamma process of Madan and Seneta (1990) (VG), normal inverse Gaussian process of Barndorff-Nielsen (1997) (NIG) and generalized tempered stable process of Koponen (1995) (GTS).

The MJ is a finite-activity jump process, which is given by Cont and Tankov (2004).

Definition 2.1.15 (Compound Poisson Process of Merton (1976)) *A compound Poisson process with intensity $\lambda > 0$ is a stochastic process $\{X_t\}_{t\geq 0}$ defined as*

$$X_t = \sum_{i=1}^{N_t} Y_i, \tag{2.60}$$

where jump sizes $\{Y_i\}_{i\geq 1}$ are i.i.d. with Gaussian distribution and $\{N_t\}_{t\geq 0}$ is a Poisson process, independent from $\{Y_i\}_{i\geq 1}$.

The Lévy measure of the MJ is expressed as

$$\nu_{MJ}(dy) = \frac{\lambda_J}{\sigma_J\sqrt{2\pi}} e^{-\frac{(y-\theta_J)^2}{2\sigma_J^2}}, \tag{2.61}$$

where θ_J and σ_J^2 are the mean and variance of the jump size distribution.

The VG and the NIG are the normal tempered stable processes and also are infinite-activity jump processes. They are constructed by the time-changed Brownian motion with drift θ_J and volatility σ_J (Cont & Tankov, 2004, p.117). However, they have different sample path properties. The VG exhibits finite variation sample path but the NIG has an infinite variation sample path. The general form for normal tempered stable processes can be written as

$$J_t = \theta_J Z_t + \sigma_J W(Z_t), \tag{2.62}$$

where $\theta_J \in \mathbb{R}$, $\sigma_J > 0$ and $\{Z_t\}_{t\geq 0}$ is the subordinating process which is independent from the Wiener process, $W(\cdot)$. The subordinator is set as gamma process for the VG and inverse Gaussian process for the NIG. The Lévy measure of the VG and NIG are given in Eq. (2.63) and Eq. (2.64), respectively, by

$$\nu_{VG}(dx) = \frac{1}{\delta_J|x|} e^{\left(\frac{\theta_J}{\sigma_J^2}x - \frac{\sqrt{\theta_J^2+2\sigma_J^2/\delta_J}}{\sigma_J^2}|x|\right)}, \tag{2.63}$$

and

$$\nu_{NIG}(dx) = \frac{\sqrt{\theta_J^2\delta_J + \sigma_J^2}}{2\pi\sigma_J\delta_J|x|} K_1\left(\frac{\sqrt{\theta_J^2 + \sigma_J^2/\delta_J}}{\sigma_J^2}|x|\right) e^{\frac{\theta_J}{\sigma_J^2}x}, \tag{2.64}$$

where K_1 is the modified Bessel function of the second kind with the following form,

$$K_1(z) = \frac{z}{4}\int_0^\infty t^{-2} e^{\left(-t-\frac{z^2}{4t}\right)} dt.$$

The GTS is obtained by the multiplication of a stable process and a Lévy measure with a decreasing exponential on both sides of the real axis (Cont & Tankov, 2004, p.119). The Lévy measure of GTS is given by

$$\nu_{GTS}(dy) = \frac{\lambda_-}{|y|^{1+\alpha_-}} e^{-c_-|y|}\mathbf{1}_{y\in(-\infty,0)} + \frac{\lambda_+}{y^{1+\alpha_+}} e^{-c_+ y}\mathbf{1}_{y\in(0,\infty)}, \tag{2.65}$$

where the negative sign $(-)$ and positive sign $(+)$ represent downward jumps and upward jumps, respectively; $\lambda_- > 0$ and $\lambda_+ > 0$ measure the jump arrival frequency for the jumps; $c_- > 0$ and $c_+ > 0$ accommodate the decay rates of the tails and both α_+ and α_- should be less than 2. Moreover, the values of α parameters also determine the type of process sample path and jump behaviour.

The GTS can be either a finite-activity or an infinite-activity jump process, which is determined by the values of the parameters of α_+ and α_-. If $\alpha_+ < 0$ and $\alpha_- < 0$, it is a finite-activity process. The GTS is an infinite-activity process when at least one of the α parameters' value is greater than zero (Cont & Tankov, 2004; Kc̈hler & Tappe, 2013). Moreover, the properties of the sample path for the GTS are also determined by the values of α parameters. The sample path has a finite variation if $\alpha_- < 1$, $\alpha_+ < 1$ and at least one of them is larger than or equal to 0. It also can have an infinite variation if $\alpha_- \geq 1$, $\alpha_+ \geq 1$, or both of them are larger than or equal to 1. The GTS has a limiting case if $\alpha_- = \alpha_+ = 0$. In this case, it is an infinite-activity jump process and we call this

process as TS.

The characteristic exponents with a closed-form expression and the expectations of the four selected jump processes are summarized in Table 2.1.

Table 2.1: Characteristic exponent of the selected pure Lévy jump process

J_t	$\mathbb{E}[J_t]$	Characteristic Exponent $\Psi^J(u)$
MJ	$\lambda_J\theta_J t$	$\lambda_J\left[e^{\left(-\frac{\sigma_J^2u^2}{2}+iu\theta_J\right)}-1\right]$
VG	θ_J	$-\frac{1}{\delta_J}\ln\left(1+\frac{u^2\sigma_J^2\delta_J}{2}-iu\theta_J\delta_J\right)$
NIG	θ_J	$\frac{1}{\delta_J}\left(1-\sqrt{1+u^2\sigma_J^2\delta_J-2iu\theta_J\delta_J}\right)$
TS	0	$-\lambda_+\left\{\frac{iu}{c_+}+\ln\left(1-\frac{iu}{c_+}\right)\right\}-\lambda_-\left\{-\frac{iu}{c_-}+\ln\left(1+\frac{iu}{c_-}\right)\right\}$
GTS	0	$\Gamma(-\alpha_+)c_+^{\alpha_+}\lambda_+\left\{\left(1-\frac{iu}{c_+}\right)^{\alpha_+}-1+\frac{iu\alpha_+}{c_+}\right\}+\Gamma(-\alpha_-)c_-^{\alpha_-}\lambda_-\left\{\left(1+\frac{iu}{c_-}\right)^{\alpha_-}-1-\frac{iu\alpha_-}{c_-}\right\}$

2.2 Fundamentals of Quantitative Finance

In this section, we present the risk-neutral pricing theory and techniques for financial derivatives. The risk-neutral modelling is a fundamental approach in the derivative pricing theory. We derive the pricing formulae for volatility derivatives based on the no-arbitrage assumption.

2.2.1 Futures and Options

A financial derivative is a contract whose value depends on the underlying asset performance. The underlying assets can be commodities, currencies, market indexes and so on. Further, a volatility derivative is a derivative security whose underlying asset is VIX or a derivative of VIX. In this thesis, we focus on pricing two kinds of volatility derivatives, futures and options written on volatility products.

A futures contract is a legal agreement which allows the market practitioners to sell

or buy a financial product at a predetermined price and a specified future time (Capinski & Zastawniak, 2003, p.134). Assume the price of a financial asset at time t is denoted by S_t for $0 \leq t \leq T$, where T is the maturity date of the contract. The futures price, $\text{Fu}_{t,T}$ for each $t \leq T$, is unknown at current time ($t = 0$) and can be treated as a random variable. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space with respect to a filtration $\{\mathcal{F}_t\}_{t\geq 0}$. Then, the futures price at the maturity time can be calculated as

$$\text{Fu}_{t,T} = \mathbb{E}[S_T|\mathcal{F}_t]. \tag{2.66}$$

An option contract gives contract holder the right, but not the obligation, to buy or sell the underlying asset at a predetermined price before or on the expiration date (Cont & Tankov, 2004). To be specific, a call option gives the holder the right to buy and a put option gives the holders the right to sell. In this thesis, we valuate two different types of options, European options and American options. A European option can only be exercised at the expiration date, but an American option can be exercised at any time before the expiration date. The payoffs of a European call and a European put with strike price K at expiration date T are defined as,

$$H_C(S_T) = \max(S_T - K, 0), \tag{2.67}$$

and

$$H_P(S_T) = \max(K - S_T, 0), \tag{2.68}$$

where H_C and H_P are the payoff of call and put options, respectively.

2.2.2 Risk-Neutral Pricing

In this sub-section, we introduce Girsanov's theorem which tells how to change between two equivalent probability measures. We assume that a non-negative random variable,

Z, on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, satisfies the following equations,

$$\mathbb{E}^{\mathbb{P}}[Z] = 1 \quad \text{and} \quad \mathbb{P}\{Z > 0\} = 1, \tag{2.69}$$

where $\mathbb{P}$ is called physical measure or original probability measure. Now, we have another probability measure $\mathbb{Q}$ which is equivalent to the probability measure $\mathbb{P}$. The measure $\mathbb{Q}$ is defined as

$$\mathbb{Q}(A) = \int_A Z(\omega) dP(\omega) \quad \text{for all } A \in \mathcal{F}. \tag{2.70}$$

We say this probability measure a risk-neutral if the price of an asset is exactly equal to the discounted expectation of the asset price under this measure. Then, we denote the expectation of a random variable X under physical measure $\mathbb{P}$ as $\mathbb{E}^{\mathbb{P}}[X]$ and the expectation of X under risk-neutral measure $\mathbb{Q}$ as $\mathbb{E}^{\mathbb{Q}}[X]$. The relationship between the two expectations can be formulated as

$$\mathbb{E}^{\mathbb{Q}}[X] = \mathbb{E}^{\mathbb{P}}[XZ]. \tag{2.71}$$

We also say that Z is the Radon-Nikodým derivative of $\mathbb{Q}$ with respect to $\mathbb{P}$. In terms of the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with a filtration $\{\mathcal{F}_t\}_{0 \le t \le T}$, the Radon-Nikodým derivative process (Shreve, 2008) can be defined as

$$Z(t) = \mathbb{E}^{\mathbb{P}}[Z|\mathcal{F}_t], \qquad 0 \le t \le T. \tag{2.72}$$

If X is an $\mathcal{F}_t$-measurable random variable, then we have

$$\mathbb{E}^{\mathbb{Q}}[X] = \mathbb{E}^{\mathbb{P}}[XZ_t]. \tag{2.73}$$

Theorem 2.2.1 (Girsanov, 1-dimension) *Let $\{W^{\mathbb{P}}\}_{0\le t\le T}$ be a Wiener process on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with respect to a filtration $\{\mathcal{F}_t\}_{0\le t\le T}$, and Θ_t be an adapted process. Define*

$$Z_t = e^{\left(-\int_0^t \Theta_s dW_s^{\mathbb{P}} - \frac{1}{2}\int_0^t \Theta_s^2 ds\right)}, \tag{2.74}$$

$$W_t^{\mathbb{Q}} = W_t^{\mathbb{P}} + \int_0^t \Theta_s ds, \tag{2.75}$$

and assume that

$$\mathbb{E}^{\mathbb{P}}\left[\int_0^T \Theta_s^2 Z_s^2 ds\right] < \infty, \tag{2.76}$$

then, $\mathbb{E}^{\mathbb{P}}[Z_T] = 1$ and $\{W^{\mathbb{Q}}\}_{0\le t\le T}$ is a Wiener process under the probability measure $\mathbb{Q}$ which is given in Eq. (2.70)*.*

In the following part of this sub-section, we firstly consider an asset price model under the physical measure $\mathbb{P}$. Let $\{S_t\}_{t\ge 0}$ be a asset price process on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with respect to a filtration $\{\mathcal{F}_t\}_{t\ge 0}$. It has the form of,

$$S_t = S_0 \exp\left\{\int_0^t \sigma_s dW_s^{\mathbb{P}} + \int_0^t \left(\alpha_s - \frac{1}{2}\sigma_s^2\right) ds\right\}, \tag{2.77}$$

where the mean rate of return $\{\alpha_t\}_{t\ge 0}$ and volatility $\{\sigma_t\}_{t\ge 0}$ are the adapted processes, and $\{W_t^{\mathbb{P}}\}_{t\ge 0}$ is the Wiener process under the measure $\mathbb{P}$.

In addition, we have an interest rate process $\{R_t\}_{t\ge 0}$ which is an adapted process. Then, the discount process $\{D_t\}_{t\ge 0}$ can be defined as

$$D_t = \exp\left\{-\int_0^t R_s ds\right\}. \tag{2.78}$$

Thus, the discounted asset price process is given by,

$$D_t S_t = S_0 \exp\left\{\int_0^t \sigma_s dW_s^{\mathbb{P}} + \int_0^t \left(\alpha_s - R_s - \frac{1}{2}\sigma_s^2\right) ds\right\} \tag{2.79}$$

and the differential form of Eq. (2.79) is

$$\begin{aligned} d(D_t S_t) &= (\alpha_t - R_t)\, D_t S_t dt + \sigma_t D_t S_t dW_t^{\mathbb{P}} \\ &= \sigma_t D_t S_t \left[\frac{\alpha_t - R_t}{\sigma_t} dt + dW_t^{\mathbb{P}} \right]. \end{aligned} \tag{2.80}$$

We let Θ_t be equal to

$$\Theta_t = \frac{\alpha_t - R_t}{\sigma_t}. \tag{2.81}$$

According to Theorem 2.2.1, we can define the Wiener process $\{W_t^{\mathbb{Q}}\}_{t\geq 0}$ under the probability measure $\mathbb{Q}$, such that

$$W_t^{\mathbb{P}} = W_t^{\mathbb{Q}} - \int_0^t \Theta_s ds, \tag{2.82}$$

where Θ_s is given in Eq. (2.81). Therefore, the discounted asset price under the measure $\mathbb{Q}$ can be written as

$$D_t S_t = S_0 + \int_0^t \sigma_s D_s S_s dW_s^{\mathbb{Q}}, \tag{2.83}$$

and $\int_0^t \sigma_s D_s S_s dW_s^{\mathbb{Q}}$ is a martingale under $\mathbb{Q}$. The undiscounted asset price process $\{S_t\}_{t\geq 0}$ under the $\mathbb{Q}$ measure can be written as,

$$S_t = S_0 \exp\left\{ \int_0^t \sigma_s dW_s^{\mathbb{Q}} + \int_0^t \left(R_s - \frac{1}{2}\sigma_s^2 \right) ds \right\}. \tag{2.84}$$

Furthermore, the generalized pricing formula of a derivative under the measure $\mathbb{Q}$ can be defined as,

$$V_t = \mathbb{E}^{\mathbb{Q}} \left[e^{-\int_t^T R_s ds} V_T | \mathcal{F}_t \right], \qquad 0 \leq t \leq T, \tag{2.85}$$

where $\{V_t\}_{0\leq t\leq T}$ is a payoff process which is $\mathcal{F}_t$-adapted, and V_T denotes the payoff of a derivative at expiration time T.

2.2.3 Implied Volatility

The payoff of a European call option at expiration date is given in Eq. (2.67). Substituting the payoff function into Eq. (2.85), we can obtain the call option pricing formula with constant interest rate r, that is,

$$H_C(S_t) = \mathbb{E}^{\mathbb{Q}}\left[e^{-r(T-t)}\max\left(S_T - K, 0\right)|\mathcal{F}_t\right], \qquad 0 \le t \le T. \tag{2.86}$$

The process $\{S_t\}_{0\le t\le T}$ under $\mathbb{Q}$ is shown in Eq. (2.84), but the volatility and interest rate are assumed to be constants in the Black-Scholes model. Then the Black-Scholes price formula (Shreve, 2008, p.220) of a European call option can be derived as

$$C^{BS}(S_t, K, \tau, \sigma) = S_t N(d_1) - Ke^{-r\tau}N(d_2), \tag{2.87}$$

with

$$d_1 = \frac{\log\left(\frac{S_t}{K}\right) + \left(r + \frac{\sigma^2}{2}\right)\tau}{\sigma\sqrt{\tau}} \quad \text{and} \quad d_2 = d_1 - \sigma\sqrt{\tau},$$

where $\tau = T - t$, and

$$N(u) \equiv (2\pi)^{-1/2}\int_{-\infty}^{u} e^{-z^2/2}dz.$$

If all other parameters are set to be constants, then Eq. (2.87) is a function of σ. Assume the market price of a European call option is $C^M(K,T)$, then the implied volatility $\Sigma_t(T,K)$ can be gained by solving the following equation,

$$C^{BS}(S_t, K, \tau, \Sigma_t(T,K)) = C^M(K,T). \tag{2.88}$$

The function $\Sigma_t : (T,K) \to \Sigma_t(T,K)$ is known as the implied volatility surface at time t. Some features for implied volatility surfaces in exponential-Lévy models are stated below (Cont & Tankov, 2004, p.359):

1. Skew/smile features: a negatively skewed jump distribution gives rise to a skew in implied volatility and a strong variance of jumps causes smile pattern in the implied volatility.

2. Short term skew: contrarily to diffusion models which produce little skew for short maturities, the models with jumps in the price process lead to short term skew.

3. Flattening of the skew/smile with maturity: for a Lévy process with finite variance, the central limit theorem shows that when the maturity T is large, the distribution of $(S_T - \mathbb{E}[S_T])/\sqrt{T}$ becomes approximately Gaussian. In other words, the implied volatility smile will become flat as $T \to \infty$.

2.2.4 VIX and S&P 500 variance futures

This subsection presents the construction of VIX index valuation and the valuation of S&P 500 variance futures, based on the fair value of the contracts on S&P 500 realized variance. It is well known that squared VIX can be formulated based on a strip of S&P 500 options. The cash-and-carry arbitrage relationship does not exist because VIX is not tradable, however the call-put parity still holds (Rhoads, 2011). Assume the price process of S&P 500, $\{S_t\}_{t\geq 0}$, is a diffusion process under the risk-neutral measure $\mathbb{Q}$,

$$dS_t = (r - q)S_t dt + \sigma_t S_t dW_t^{\mathbb{Q}}, \tag{2.89}$$

where r and q are the risk-free interest rate and stock dividend respectively, $\{\sigma_t\}_{t\geq 0}$ is an adapted stochastic process and $\{W_t^{\mathbb{Q}}\}_{t\geq 0}$ is Wiener process. Applying Itô's lemma to Eq. (2.89), we gain the following equation,

$$d\ln S_t = (r - q - \frac{1}{2}\sigma_t^2)dt + \sigma_t dW_t^{\mathbb{Q}}. \tag{2.90}$$

If we let the forward price of S&P 500 at current time $t = 0$ be $F_{0,T}$, then we have the relationship between the forward price and the spot price as

$$F_{0,T} = S_0 e^{(r-q)T}, \tag{2.91}$$

where T is delivery date. The fair value of realized variance can be computed as

$$\frac{K_{var}}{100^2} = \mathbb{E}_0^{\mathbb{Q}}\left[\frac{1}{T}\int_0^T \sigma_t^2 dt\right], \tag{2.92}$$

where K_{var} is the fair variance strike. Combining with Eq. (2.90), the following equations can be obtained

$$\frac{K_{var}}{100^2} = \mathbb{E}_0^{\mathbb{Q}}\left[\frac{1}{T}\int_0^T \sigma_t^2 dt\right] = -\frac{2}{T}\mathbb{E}_0^{\mathbb{Q}}\left[\ln\left(\frac{S_T}{F_{0,T}}\right)\right]. \tag{2.93}$$

If the strikes of the S&P 500 options are assumed to be continuous, then the price formulae for S&P 500 call and put options with strike K and maturity T are given by Brennan (1979) as

$$C_{0,T}(K) = e^{-rT}\int_0^\infty (S_T - K)^+ q(S_T) dS_T, \tag{2.94}$$

$$P_{0,T}(K) = e^{-rT}\int_0^\infty (K - S_T)^+ q(S_T) dS_T, \tag{2.95}$$

where $C_{0,T}(K)$ and $P_{0,T}(K)$ represent the call and put option prices at time 0, respectively. By using the general pay-off function (Carr & Madan, 2001) [2], we obtain

$$\begin{aligned}\mathbb{E}_0^{\mathbb{Q}}\left[\ln\left(\frac{S_T}{F_{0,T}}\right)\right] = &-\ln\left(\frac{F_{0,T}}{K_0}\right)+\left(\frac{F_{0,T}}{K_0}-1\right)\\ &-e^{rT}\left\{\int_0^{K_0}\frac{P_{0,T}}{K^2}dK+\int_{K_0}^{\infty}\frac{C_{0,T}}{K^2}dK\right\}.\end{aligned} \tag{2.96}$$

Then, CBOE makes two approximations in Eq. (2.96). First, Taylor expansion is used for approximating $\ln\left(\frac{F_{0,T}}{K_0}\right)$. Second, the integrals in this equation are converted to sums of all existing S&P 500 options, excluding the options without bids. Finally, the discrete version of VIX valuation formula is obtained as

$$\left(\frac{VIX}{100}\right)^2 \approx -\frac{2}{T}\mathbb{E}_0^{\mathbb{Q}}\left[\ln\left(\frac{S_T}{F_{0,T}}\right)\right] = \frac{2}{T}\sum_i\frac{\triangle K_i}{K_i^2}e^{rT}q(K_i)-\frac{1}{T}\left(\frac{F_{0,T}}{K_0}-1\right)^2, \tag{2.97}$$

where $q(K_i)$ is the mid-point of the bid-ask spread, and $dK \to \triangle K_i = \frac{(K_{i+1}-K_{i-1})}{2}$.

According to the specification of the variance futures provided on CBOE website, we know that variance futures is quoted based on volatility points and vega notional. At the end of trading date, the realized variance of the current day is calculated by using the daily S&P 500 closing value. Then, the price is required to change volatility points to an adjusted futures price. Note that the closing price of the S&P 500 of that day should be treated as the realized observation. The conversion formula from volatility points to adjusted futures price, F_t, is given by,

$$F_t = DF_{t,T}\left(k-k_0\right) - \text{ARMVM} + 1000, \tag{2.98}$$

[2] Let the forward stock price be $F_{0,T} = \mathbb{E}_0[\omega(S_T)]$ and $\mathbb{E}_0^{\mathbb{Q}}[\omega(S_T)] = \int_0^{\infty}\omega(S_T)q(S_T)dS_T$. The general pay-off function is given by,

$$\begin{aligned}&\mathbb{E}_0^{\mathbb{Q}}[\omega(S_T)] = \omega(K_0)+\omega'(K_0)\left(F_{0,T}-K_0\right)\\ &+e^{rT}\left\{\int_0^{F_{0,T}}\omega(K)''P_{0,T}(K)dK+\int_{F_{0,T}}^{\infty}\omega(K)''C_{0,T}(K)dK\right\},\end{aligned}$$

where K_0 is strike price and preferred value for K_0 is set as $F_{0,T}$.

where $DF_{t,T}$ denotes the discount factor from current time t to maturity date T; ARMVM refers to accumulated return on modified variation margin; k_0 is the initial variance strike, which is set at 3:15 pm (Chicago time) on business day. In addition, k in Eq. (2.98) can be calculated as,

$$k = \frac{252}{N_e - 1} \times \left(\text{volatility strike}^2 \times \frac{(N_e - 1 - n)}{252} + \sum_{i=1}^{n} R_i^2 \times 10000 \right),$$

where N_e is the number of days from the time t to maturity date T; n is the number of returns to the current date; and $R_i = \ln(S_{i+1}/S_i)$ is the daily log return of the S&P 500. Moreover, the conversion formula from vega notional to variance units is given by,

$$\text{variance units} = \frac{\text{vega notional}}{2 \times \text{volatility strike}} \times \frac{N_e - 1}{N_e - 1 - n}. \tag{2.99}$$

From Eq. (2.98) and Eq. (2.99), the final settlement price of VA can be calculated as

$$F_T = \text{realized variance} - k_0 - \text{ARMVM} + 1000, \tag{2.100}$$

where realized variance is formulated as

$$\text{realized variance} = \left(252 \times \sum_{i=1}^{N_a - 1} \frac{R_i^2}{N_e - 1} \right),$$

where N_a represents actual number of days between the contract listed date and settlement date.

2.3 Numerical Methods

In this section, we introduce the numerical methods for evaluating the improper integrals, solving the ordinary differential equations, estimating the unobserved state variables

and estimating the values of parameters in the model.

2.3.1 Gauss-Laguerre Quadrature Method

The Gauss-Laguerre quadrature method (Burnett, 1937) is widely applied to approximate the value of a particular integral over $[0, \infty)$, which has the form of,

$$\int_0^\infty x^m e^{-x} f(x) dx, \tag{2.101}$$

where $m > 0$. According to Burnett (1937) and Concus, Cassatt, Jaehnig and Melby (1963), the above integral can be estimated by

$$\int_0^\infty x^m e^{-x} f(x) dx = \sum_{i=1}^{n} W_i f(a_i) + E_n, \tag{2.102}$$

where a_i is the ith zero of the Laguerre polynomial $L_n^m(a_i)$, W_i is the corresponding weight, and E_n is the error term. Here, $L_n^m(a_i)$ is given by,

$$L_n^m(a_i) = \sum_{k=0}^{n} \binom{n+m}{n-k} \frac{(-a_i)^k}{k!}.$$

Then, the weight is represented as

$$W_i = \frac{\Gamma(n+m+1)a_i}{n!\left[(n+1)L_{n+1}^m(a_i)\right]^2}, \quad i = 1, 2, ..., n,$$

and the error term is defined as

$$E_n = \frac{n!\Gamma(n+m+1)}{(2n)!} f^{(2n)}(\xi).$$

2.3.2 4th-order Runge-Kutta Method

The Runge-Kutta method is a relatively efficient numerical method for solving ordinary differential equation which is the form of,

$$\frac{dy}{dx} = f(x, y), \quad y(x_0) = y_0. \tag{2.103}$$

In the Runge-Kutta method, Eq. (2.103) can be effectively solved by substituting the following Taylor-series expansion (Hildebrand, 1987, p.285)

$$y_{n+1} = y_n + h\frac{dy}{dx} + \frac{h^2}{2}\frac{d^2y}{dx^2} + \frac{h^3}{6}\frac{d^3y}{dx^3} + \cdots, \tag{2.104}$$

by an approximation of y_{n+1}

$$\begin{aligned} y_{n+1} =& y_n + h\left[\alpha_0 f(x_n, y_n) + \alpha_1 f(x_n + \mu_1 h, y_n + b_1 h)\right. \\ &\left. + \alpha_2 f(x_n + \mu_2 h, y_n + b_2 h) + \cdots + \alpha_P f(x_n + \mu_p h, y_n + b_p h)\right], \end{aligned} \tag{2.105}$$

where α's, μ's and b's are deterministic. The advantage of this method is that the method can be employed when the function, $f(x, y)$, does not have a closed-form because it does not need to evaluate the derivatives of $f(x, y)$.

The classic Runge-Kutta method is also known as the 4th-order Runge-Kutta method which can be possible to approximate a function value with fourth-order accuracy. Consider a problem with initial condition which is specified in Eq. (2.103). Now choose a step size $h > 0$, then we have (Hildebrand, 1987, p.290)

$$\begin{aligned} y_{n+1} &= y_n + \frac{1}{6}(k_0 + 2k_1 + 2k_2 + k_3) + O(h^5), \\ x_{n+1} &= x_n + h, \end{aligned} \tag{2.106}$$

with

$$\begin{aligned} k_0 &= hf(x_n, y_n), \\ k_1 &= hf(x_n + \frac{1}{2}h, y_n + \frac{1}{2}k_0), \\ k_2 &= hf(x_n + \frac{1}{2}h, y_n + \frac{1}{2}k_1), \\ k_3 &= hf(x_n + h, y_n + k_2). \end{aligned} \tag{2.107}$$

This method also can be applied in a simultaneous system which has the form of,

$$\begin{aligned} \frac{dy}{dx} &= f(x, y, z), \quad y(x_0) = y_0, \\ \frac{dz}{dx} &= g(x, y, z), \quad z(x_0) = z_0. \end{aligned} \tag{2.108}$$

Similarly, we can get following estimations,

$$\begin{aligned} y_{n+1} &= y_n + \frac{1}{6}\left(k_0 + 2k_1 + 2k_2 + k_3\right) + O(h^5), \\ z_{n+1} &= z_n + \frac{1}{6}\left(m_0 + 2m_1 + 2m_2 + m_3\right) + O(h^5), \end{aligned} \tag{2.109}$$

with

$$\begin{aligned} k_0 &= hf(x_n, y_n, z_n), \\ k_1 &= hf(x_n + \frac{1}{2}h, y_n + \frac{1}{2}k_0, z_n + \frac{1}{2}m_0), \\ k_2 &= hf(x_n + \frac{1}{2}h, y_n + \frac{1}{2}k_1, z_n + \frac{1}{2}m_1), \\ k_3 &= hf(x_n + h, y_n + k_2, z_n + m_2), \end{aligned} \tag{2.110}$$

and

$$
\begin{aligned}
k_0 &= hg(x_n, y_n, z_n), \\
k_1 &= hg(x_n + \frac{1}{2}h, y_n + \frac{1}{2}k_0, z_n + \frac{1}{2}m_0), \\
k_2 &= hg(x_n + \frac{1}{2}h, y_n + \frac{1}{2}k_1, z_n + \frac{1}{2}m_1), \\
k_3 &= hg(x_n + h, y_n + k_2, z_n + m_2).
\end{aligned} \tag{2.111}
$$

2.3.3 Unscented Kalman Filtering

The unscented kalman filter (UKF) is widely used for estimating a nonlinear system based on the sigma-point sampling and weighting strategies (Grewal & Andrews, 2014). The unscented transform is the core of UKF. An abstract general-form of nonlinear system can be represented as

$$
\begin{aligned}
d\mathbf{X}_t &= h(\mathbf{X}_t, t)dt + L_t d\beta_t, \\
d\mathbf{Y}_t &= f(\mathbf{X}_t, t)dt + V_t d\eta_t,
\end{aligned} \tag{2.112}
$$

where $\mathbf{X}_t \in \mathbb{R}^n$ represents the state process; $\mathbf{Y}_t \in \mathbb{R}^m$ is the measurement process and f is the measurement function. The processes $\mathbf{X}_t$ and $\mathbf{Y}_t$ are independent, and β_t and η_t are independent Brownian motions. The UKF is applied in the above problem in the following two steps.

In the prediction step, we apply the linear forecast. Based on the information up to time $t-1$, we denote the predicted state mean by $\hat{\mathbf{X}}_{t|t-1}$ and predicted state covariance by $\hat{\Sigma}^{\mathbf{X}}_{t|t-1}$ (for detailed discussions on linear forecasts, see Grewal and Andrews (2015)).

In the update step, we employ scaled unscented transformation. As in Särkkä (2007), we generate $2n+1$ sigma points where n is the dimension of $\mathbf{X}_t$. These sigma points

are computed as follows:

$$\begin{aligned}
\mathscr{S}_t^0 &= \hat{\mathbf{X}}_{t|t-1}, \\
\mathscr{S}_t^i &= \hat{\mathbf{X}}_{t|t-1} + \left[\text{Chol}\left((n+\lambda)\,\hat{\Sigma}^{\mathbf{X}}_{t|t-1}\right)\right]_i, \qquad i = 1,\dots,n \\
\mathscr{S}_t^i &= \hat{\mathbf{X}}_{t|t-1} - \left[\text{Chol}\left((n+\lambda)\,\hat{\Sigma}^{\mathbf{X}}_{t|t-1}\right)\right]_{i-n}, \qquad i = n+1,\dots,2n
\end{aligned} \tag{2.113}$$

with the weights

$$W_0^{(m)} = \frac{\lambda}{n+\lambda}, \tag{2.114}$$

$$W_0^{(\Sigma)} = \frac{\lambda}{n+\lambda} + \left(1-\alpha^2+\beta\right), \tag{2.115}$$

$$W_i^{(m)} = W_i^{(\Sigma)} = \frac{1}{2\,(n+\lambda)}, \qquad i = 1,\dots,2n, \tag{2.116}$$

where Chol$(\cdot)$ means the Cholesky factorization, $[\cdot]_i$ denotes the ith column of the matrix, and $\lambda = \alpha^2\,(n+\beta)$. For the normal distribution, the optimal value of β is 2. We choose α to be 10^{-3}.

Then, the predicted mean of the measurements (for the measurement function see, Eq. (3.15)), variance of the measurements, and covariance of the latent variables and measurements can be computed by using the sigma points (2.113) as

$$\hat{\mathbf{Y}}_{t|t-1} = \sum_{i=0}^{2n} W_i^{(m)} \mathbf{f}(\mathscr{S}_t^i; x_t, \Theta), \tag{2.117}$$

$$\hat{\Sigma}^{\mathbf{Y}}_{t|t-1} = \sum_{i=0}^{2n} W_i^{(\Sigma)} \left(\mathbf{f}(\mathscr{S}_t^i; x_t, \Theta) - \hat{\mathbf{Y}}_{t|t-1}\right)\left(\mathbf{f}(\mathscr{S}_t^i; x_t, \Theta) - \hat{\mathbf{Y}}_{t|t-1}\right)^\top + \Sigma_t^\epsilon, \tag{2.118}$$

$$\hat{\Sigma}^{\mathbf{X},\mathbf{Y}}_{t|t-1} = \sum_{i=0}^{2n} W_i^{(\Sigma)} \left(\mathscr{S}_t^i - \hat{\mathbf{X}}_{t|t-1}\right)\left(\mathbf{f}(\mathscr{S}_t^i; x_t, \Theta) - \hat{\mathbf{Y}}_{t|t-1}\right)^\top. \tag{2.119}$$

Finally, we calculate the Kalman filter gain, the state mean, and state covaraince

conditional on the measurements as

$$K_t = \hat{\Sigma}^{\mathbf{X},\mathbf{Y}}_{t|t-1}\left(\hat{\Sigma}^{\mathbf{Y}}_{t|t-1}\right)^{-1}, \tag{2.120}$$

$$\hat{\mathbf{X}}_t = \hat{\mathbf{X}}_{t|t-1} + K_t\left(\mathbf{Y}_t - \hat{\mathbf{Y}}_{t|t-1}\right), \tag{2.121}$$

$$\hat{\Sigma}^{\mathbf{X}}_t = \hat{\Sigma}^{\mathbf{X}}_{t|t-1} - K_t\hat{\Sigma}^{\mathbf{Y}}_{t|t-1}K_t^{\top}. \tag{2.122}$$

2.3.4 Maximum Likelihood Estimation

Likelihood-based estimation methods (Rossi, 2018, p.234) are widely used for evaluating the estimators. Assume that a random sample X_i for $i = 1, ..., n$ has a joint probability density function $f(\cdot)$, such that

$$L(\theta) = f(X_1, ..., X_n; \theta), \tag{2.123}$$

and the log-likelihood function is defined as $l(\theta) = \ln[L(\theta)]$. If this sample, X_i for $i = 1, ..., n$, is i.i.d. random variables, then we have

$$L(\theta) = \Pi_{i=1}^{n} f(x_i; \theta) \tag{2.124}$$

Definition 2.3.1 (The Law of Likelihood) *Let X_i for $i = 1, ..., n$ be a sample of i.i.d. random variables with common pdf $f(x : \theta)$ and parameter space Θ. For $\theta \in \Theta$, the larger the value of $L(\theta)$ is, the more θ agrees with the observed data. Thus, the degree to which the information in the sample supports a parameter value $\theta_0 \in \Theta$, in comparison to another value $\theta_1 \in \Theta$, is equal to the ratio of their likelihoods $\Lambda(\theta_0, \theta_1) = L(\theta_0)/L(\theta_1)$. In particular, the information in the sample is in better agreement with θ_1 than θ_0 when $\Lambda < 1$, and the information in the sample agrees better with θ_0 than θ_1 when $\Lambda > 1$.*

The maximum likelihood estimation (MLE) method is one of the likelihood-based estimation methods. It is used for finding the estimators of unknown parameters by

maximizing the value of the likelihood function. The definition of maximum likelihood principle is given below (Rossi, 2018, p.223).

Definition 2.3.2 (The Maximum Likelihood Principle) *Given a random sample X_i for $i = 1, ..., n$ and a parametric model $f(x_1, ..., x_n : \theta)$, choose as the estimator of θ, say $\hat{\theta}(X)$, the value of $\theta \in \Theta$ that maximizes the likelihood function.*

Definition 2.3.3 *The MLE of θ, denoted by $\hat{\theta}$, is a solution to the maximization problem $\max_{\theta \in \Theta} L(\theta)$.*

Chapter 3

Pricing VIX Derivatives with Infinite-Activity Jumps

This chapter is joint work with J. Cao, X. Ruan and W. Zhang. Its earlier version was presented at AUT Mathematical Sciences Symposium, November 2018, AUT, Auckland, New Zealand; 2018 Derivative Markets Conference, August 2019, Queenstown, New Zealand; and Quantitative Methods in Finance 2019 Conference, December 2019, Sydney, Australia.

In this chapter, we derive the price formula for VIX options and futures based on the two-factor model with infinite-activity jumps and investigate the model performance in capturing the features of the dynamics for the VIX derivatives market data. This chapter is organized as follows. In Section 3.1, we conduct a preliminary analysis on the VIX and VIX derivatives data. In Section 3.2, we introduce our two-factor model with infinite-activity jumps. In Section 3.3, we derive the pricing formulae for VIX options and futures. In Section 3.4, we select market data to calibrate our two factor model and others. In Section 3.5, we conduct various goodness-of-fit tests and compare the performance of the models with different jump structures. Proofs and UKF method performance test are given in Appendix A.

3.1 Data

In this section, we conduct a preliminary analysis on VIX data by focusing on analyzing the features of the dataset and whether the jumps occur in the VIX dynamics. We also describe how we select the data of VIX option and VIX futures and present a summary on data statistics.

3.1.1 A Preliminary Analysis on VIX Data

In this subsection, we analyze the VIX daily data, which span over 26 years, from 1992 to 2017, with a total of 6,550 days. We also compare the characteristics of the full sample with the Wednesday sample which consists of all VIX close prices and their logarithm on Wednesdays during the sample period. Table 3.1 presents the descriptive statistics of the full sample and the Wednesday sample. From this table, we find the statistics of the full sample are considerably close to the statistics of the Wednesday sample. Particularly, the absolute differences of mean and standard deviation between the two samples under Log VIX panel are 0.09% and 0.01%, respectively. If we have a closer look at the Table 3.1, we can find that the kurtoses of VIX for the full sample and Wednesday sample are around 10.50 which is much bigger than 3, so it clearly shows that each of these two samples has positive excess kurtosis. However, the kurtosis of log VIX for each of the two samples is very close to 3, so it is important to do further investigation on the significance of excess kurtosis. Then, a normality test is applied to the log VIX data (for the details about this test, see (Duncan, 1998)). The results of this test on log VIX data are computed as 5.26 for the full sample and 2.15 for the Wednesday sample, which are greater than 1.96. Therefore, each of the two samples has positive excess kurtosis. Moreover, both of the samples have positive skewness. Based on the above observation, we figure out that the distributions of the VIX index and log VIX are asymmetry and leptokurtotic. There may have some jump parts in the

VIX dynamics.

Table 3.1: Descriptive statistics of VIX 1992-2017

	Mean	**Std.**	**Skew.**	**Kurt.**	**Min.**	**Med.**	**Max.**
				VIX			
Full Sample	19.2652	8.0508	2.1089	10.5107	9.1400	17.1950	80.8600
Wednesday	19.2441	8.0183	2.0973	10.5308	9.3100	17.2400	74.2600
				Log VIX			
Full Sample	2.8893	0.3569	0.6788	3.3176	2.2127	2.8446	4.3927
Wednesday	2.8884	0.3568	0.6645	3.2827	2.2311	2.8472	4.3076

This table presents the statistics summary of the VIX close values and logarithm of VIX close values for the full sample and the Wednesday sample during the period between 1992 and 2017. It includes the information about mean, standard deviation (Std.), skewness (Skew.), kurtosis (Kurt.), minimum value (Min.), median value (Med.) and maximum value (Max.) of the samples.

Alternatively, we examine whether jumps present in the log VIX return process by the statistical test proposed by Aït-Sahalia and Jacod (2009). Under a given test statistics $\hat{S}_T$ and time interval from 0 to T, if $\hat{S}_T$ is less than a critical value $c_{n,T}^{95\%}$, then the null hypothesis of no jumps should be rejected at the 95% confidence level. Theoretically, $\hat{S}_T$ approaches to 1 as $\Delta_n = T/n$ goes to 0, if jumps present in a process. Following the suggestions from Aït-Sahalia and Jacod (2009), we compute $c_{n,T}^{95\%}$ based on the thresholds $5\sigma\Delta_n^{\bar{\omega}}$ where $\bar{\omega} = 0.48$ and σ is the average standard deviation.

We conduct jump test on both VIX and log VIX data. Based on the above setting, the test statistics ($\hat{S}_T$) and their corresponding critical values ($c_{n,T}^{95\%}$) on the full sample and the Wednesday sample are calculated and shown in Table 3.2. From Table 3.2, we can see that the values of $\hat{S}_T$ under the VIX and log VIX column are much smaller than their corresponding critical values. Thus, we should reject the null hypothesis at the 95% confidence level for all the samples. So, we can conclude that there are jumps in the process of the daily level and daily log return as well as in the Wednesday sample.

According to the studies of Todorov and Tauchen (2011a) and Aït-Sahalia and Jacod (2011), infinite-activity jump test is required to be applied to high-frequency data set

Table 3.2: Test statistics and critical values for the jump test

	VIX		Log VIX	
	Test statistics ($\hat{S}_T$)	Critical Value ($c_{n,T}^{95\%}$)	Test statistics ($\hat{S}_T$)	Critical Value ($c_{n,T}^{95\%}$)
Full Sample	0.4771	1.8982	0.4270	1.8752
Wednesday	0.2940	1.7560	0.3607	1.7881

This table presents the test statistics and critical values which are calculated based on the jump test method of (Aït-Sahalia & Jacod, 2009) for the full sample and Wednesday sample.

(e.g. 5 seconds or 1 minute time frequencies) for the satisfied results. Therefore, it is difficult to detect whether infinite-activity jumps occur in the process, as we do not access the high-frequency VIX data. The purpose of this paper is to propose a new model which can improve the pricing performance of VIX derivatives compared with the existing models. Detecting whether jumps are finite- or infinite-activity in the VIX dynamics is not our focus. At this stage, we expect that infinite-activity jumps can improve the pricing performance of VIX derivatives.

3.1.2 Sample Construction and Sample Summary of VIX Derivatives Data

Our sample consists of VIX options and VIX futures data within the period between July 2006 and April 2016, with a total of 509 weeks. The data are collected from OptionMetrics and Bloomberg. The futures data set contains spot prices, the time-to-maturity, and the trading volume. The option data set contains strike prices, bid and ask prices, time-to-maturity, interest rates, market vega, implied volatility, and VIX index levels.

The massive data set causes a significant increase in the computational cost, thus we apply a series of filters on our data set to reduce the data size. In line with Yang and Kanniainen (2017) and Du and Luo (2019), we only use each Wednesday's data.

We filter out all anomalous data, which are the entries that do not contain either price value or implied volatility value, or contain other irrational values. We focus on the VIX options and futures with time-to-expiration being greater than two weeks. To address the illiquidity issues, we delete the futures whose trading volumes are less than 50 and the options whose trading volumes are less than 120. We choose the out-of-money (OTM) options because the OTM options tend to be more liquid. Also, we only use the VIX futures whose maturities are the same as those of options, so the total number of futures contracts in our full sample is 2,562. We take the average price of the best bid and best offer prices of each option as our option price. We select the options whose average prices are greater than or equal to 0.1 and the bid-ask spread less than or equal to 0.3. So it leaves us a total of 12,921 call options and a total of 5,249 put options.

We separate the full sample into two sub-samples. The first sub-sample is termed as in-sample and its period is from July 2006 to January 2013 (a total of 342 weeks) and the second sub-sample is termed as out-of-sample and its period is from February 2013 to April 2016 (a total of 167 weeks). Table 3.3 describes the statistics summary for VIX options and futures according to this split. Overall, the total numbers of futures and option contracts in the in-sample are greater than those in the out-of-sample, respectively.

In Table 3.3, the futures panel summarizes the number of futures contracts and average futures prices in the different maturity groups. The number of futures contracts increases as the time to maturity increases for both in-sample and out-of-sample. The in-sample average future prices are over 3 under all categories, while the out-of-sample average future prices are less than 3.

The options panel reports the number of option contracts, average option prices, average Black-Scholes implied volatility, and average Black-Scholes vega by time to maturity and moneyness. In terms of the moneyness, the average option prices vary from 0.95 to 2.78 for the in-sample data and from 0.71 to 2.14 for the out-of-sample

data. We find that the average implied volatility increases as the value of moneyness rises and decreases as the τ increases. We also observe that the average price and average vega of call options are less than those of put options, but the average implied volatility of call options is higher than that of put options in both samples.

3.2 A Two-factor Model with Infinite-activity Jumps

In this section, we introduce our two-factor model with infinite-activity jumps and derive the characteristic function from the proposed model.

3.2.1 Model Specifications

Psychoyios and Dotsis (2010) compare modelling VIX in the original index level and the logarithm of the original level by using mean reverting process. Their empirical results conclude that the modelling of VIX with logarithm can improve the goodness-of-fit from modelling of the original level. Later on, Mencía and Sentana (2013) conduct the same comparison but on a different time period and show that the VIX model in the logarithm outperforms that in the original level. Based on these researches, we directly model VIX dynamics in the logarithm.

Furthermore, Mencía and Sentana (2013) also find that the dynamics of the VIX exhibit highly persistent variations from the long-term mean. Thus, the VIX reverts stochastically towards the long-term mean, which means that an additional process of the long-term mean can help improve the performance of the model. Moreover, the instantaneous volatility has been widely studied in the option pricing models (Stein & Stein, 1991; Heston, 1993; Duffie, Pan & Singleton, 2000; Park, 2016, etc.). For example, Heston (1993) uses the CIR process to model instantaneous volatility.

In this paper, we model the VIX with stochastic volatility and time-varying mean in terms of log price, and allow the model to have the central tendency property. Following

Table 3.3: Summary statistics for VIX derivatives by time to maturity

	In-Sample				**Out-of-Sample**			
	$\tau \leq 30$	$30 < \tau \leq 90$	$\tau > 90$	All	$\tau \leq 30$	$30 < \tau \leq 90$	$\tau > 90$	All
				Futures				
No.	235	604	804	1643	113	308	498	919
Average prices	3.0999	3.1688	3.2278	3.1878	2.7783	2.8350	2.9017	2.8642
				Options				
				The number of options				
All	2236	5395	2316	9947	1411	4188	2624	8223
All calls	1576	3687	1570	6833	1070	3130	1888	6088
All puts	660	1708	746	3114	341	1058	736	2135
K/F ≤ 0.9	319	1066	479	1864	129	559	442	1130
0.9 < K/F ≤ 0.98	278	501	206	985	172	409	223	804
0.98 < K/F ≤ 1.02	125	260	132	517	75	176	158	409
1.02 < K/F ≤ 1.1	257	536	241	1034	165	385	215	765
K/F > 1.1	1257	3032	1258	5547	870	2659	1586	5115
				Average prices				
All	0.9396	1.3830	1.8957	1.4027	0.6616	1.0229	1.3721	1.0724
All calls	0.9033	1.3708	1.9128	1.3875	0.6425	1.0295	1.4015	1.0586
All puts	1.0263	1.4093	1.8598	1.4360	0.7214	1.0566	1.3715	1.1116
K/F ≤ 0.9	0.6531	0.9155	1.2129	0.9470	0.4339	0.6340	0.8771	0.7063
0.9 < K/F ≤ 0.98	1.2836	2.0662	2.8022	1.9992	0.8217	1.4086	1.9899	1.4443
0.98 < K/F ≤ 1.02	1.8282	2.7871	3.6839	2.7842	1.3507	2.0726	2.5921	2.1409
1.02 < K/F ≤ 1.1	1.6226	2.4568	3.2426	2.4326	1.0917	1.7964	2.4263	1.8214
K/F > 1.1	0.7082	1.1242	1.5616	1.1292	0.5228	0.8639	1.1588	0.8973
				Average implied volatility				
All	1.0798	0.8760	0.7010	0.8810	1.1246	0.9209	0.7199	0.8917
All calls	1.1664	0.9437	0.7453	0.9495	1.2166	0.9844	0.7604	0.9668
All puts	0.8729	0.7298	0.6078	0.7309	0.8357	0.6956	0.5788	0.6777
K/F ≤ 0.9	0.8815	0.7115	0.5898	0.7093	0.8527	0.6654	0.5516	0.6423
0.9 < K/F ≤ 0.98	0.8529	0.7560	0.6344	0.7579	0.8130	0.7188	0.6145	0.7100
0.98 < K/F ≤ 1.02	0.9191	0.7780	0.6539	0.7805	0.9104	0.7864	0.6428	0.7537
$1.02 < K/F \leq 1.1$	0.9798	0.8128	0.6641	0.8196	0.9592	0.8153	0.6781	0.8078
K/F > 1.1	1.2166	0.9732	0.7662	0.9814	1.2763	1.0299	0.7950	0.9990
				Average vega				
All	1.7087	2.9579	4.6707	3.0759	1.1898	2.1327	3.4157	2.3803
All calls	1.6526	2.9443	4.7510	3.0615	1.1571	2.1691	3.5487	2.3513
All puts	1.8428	2.9872	4.5017	3.1075	1.2927	2.2125	3.3654	2.4630
K/F ≤ 0.9	1.6533	2.6483	3.9989	2.8251	1.1249	1.9320	3.0040	2.2592
0.9 < K/F ≤ 0.98	2.0211	3.4831	5.2754	3.4453	1.3835	2.4761	3.8678	2.6284
0.98 < K/F ≤ 1.02	2.0803	3.7292	5.7809	3.8544	1.5279	2.7150	4.1311	3.0444
1.02 < K/F ≤ 1.1	2.1736	3.7854	5.7947	3.8531	1.4863	2.6844	4.1805	2.8464
K/F > 1.1	1.5217	2.7675	4.4956	2.8771	1.0758	2.0036	3.2919	2.2453

The futures prices are obtained from Bloomberg. The option prices, implied volatility, and vega are obtained from OptionMetrics. Wednesday futures and options are used. In this table, time to maturity is denoted as τ and the moneyness is defined as the ratio of strike price (K) to the future price (F). Under the options panel, “All" refers to all options including both call and put options.

Mencía and Sentana (2013), we assume that stochastic mean process $\{m_t\}_{t\geq 0}$ is mean-reverting. However, we assume that the dynamics of volatility process $\{v_t\}_{t\geq 0}$ follows the CIR process, which is one of differences from the best model in Mencía and Sentana (2013) (for more details about this model, refer to Eq. (3.12)). Let x_t denote the logarithm of VIX at time t in a fixed filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{Q})$. In order to investigate the influence of infinite-activity Lévy jumps on the price change of VIX derivatives, we allow infinite-activity jumps in the log VIX process and correlation between log VIX and volatility process. These are another two differences from the Mencía and Sentana (2013) model. We specify the dynamics of the log VIX under the risk-neutral measure $\mathbb{Q}$ as follows:

$$\begin{aligned} dx_t &= \kappa(m_t - x_t)dt + \rho\sqrt{v_t}dW_t^v + \sqrt{(1-\rho^2)v_t}dW_t^x + dJ_t, \\ dm_t &= \kappa_m(\theta_m - m_t)dt + \sigma_m dW_t^m, \\ dv_t &= \kappa_v(\theta_v - v_t)dt + \sigma_v\sqrt{v_t}dW_t^v, \end{aligned} \tag{3.1}$$

where W_t^x, W_t^m and W_t^v are independent and $\mathcal{F}_t$-adapted Wiener processes; ρ is the correlation coefficient of the process $\{x_t\}_{t\geq 0}$ and $\{v_t\}_{t\geq 0}$; κ, κ_m and κ_v are constants; θ_m and θ_v are the long-term means for $\{m_t\}_{t\geq 0}$ and $\{v_t\}_{t\geq 0}$ respectively; σ_m and σ_v are the volatility of the process $\{m_t\}_{t\geq 0}$ and $\{v_t\}_{t\geq 0}$, respectively, and $\{J_t\}_{t\geq 0}$ denotes the infinity-activity jump process which has been described in Section 2.1.7. The jump compensator, $\theta_J dt$, allows the jump part to be a martingale (Madan & Milne, 1991), where $\mathbb{E}^{\mathbb{Q}}[J_1] = \theta_J$. In this chapter, the jump process in Eq. (3.1) is chosen to be either the VG process or the NIG process. The model with the VG process is referred to as OU-VG, and the model with NIG process is referred to as OU-NIG. For the simplicity, we introduce a new set of parameters for the jump processes, see Table 3.4. Additionally, in this model, the correlation coefficient, ρ, can control the leverage effect (Duffie et al., 2000; Kiesel & Rahe, 2017; Du & Luo, 2019). The constants κ, κ_m and

κ_v accommodate the speed of reverting to their long-term means, respectively.

Table 3.4: VG and NIG processes

Process	**VG Process**	**NIG Process**
Lévy measure	$\nu(y) = \frac{\lambda}{\lvert y\rvert} e^{\frac{\alpha y - \sqrt{\alpha^2+2\beta^2}\lvert y\rvert}{\beta^2}}$	$\nu(y) = \frac{\lambda\alpha}{\pi\lvert y\rvert} e^{\beta y} K_1(\alpha\lvert y\rvert)$
	$\alpha = \theta_J \delta_J$ $\beta = \sqrt{\sigma_J^2 \delta_J}$ $\lambda = 1/\delta_J$	$\alpha = \sqrt{1/\delta_J \sigma_J^2 + \theta_J^2/\sigma_J^4}$ $\beta = \theta_J/\sigma_J^2$ $\lambda = \sigma_J$
$\Psi^J(u)$	$-\lambda \ln\left(1 - i\alpha u + \frac{1}{2}\beta^2 u^2\right)$	$\lambda\left[\sqrt{\alpha^2-\beta^2} - \sqrt{\alpha^2-(\beta+iu)^2}\right]$
$\mathbb{E}^{\mathbb{Q}}[J_1]$	$\alpha\lambda$	$\beta\lambda^2$

In this table, the characteristic exponent of a Lévy process is denoted by $\Psi^J(u)$. The general form of the normal tempered stable process is shown in the Eq. (2.62).

There are two main economic reasons why we select the NIG and VG processes. Firstly, both of them are the subordinate Brownian motion with exponential tails which tally with the features of VIX data. According to (2.62), each unit of calendar time may be viewed as having an economically relevant time length given by an independent random variable that has a gamma or inverse Gaussian density with unit mean and positive variance. Under the VG or NIG process, the unit period continuously compounded return is normally distributed, conditional on the realization of a random time which has a gamma or inverse Gaussian density. The parameters in the VG or NIG process could separately control tail probabilities of the return distribution: tail heaviness of steepness (i.e., kurtosis) and symmetry (i.e., skewness). Secondly, the VG process has a finite variation with a moderately low activity rate of small jumps, whereas the NIG process has an infinite variation with a stable arrival rate of small jumps (Cont & Tankov, 2004). For the VG process, the sum of the absolute log price changes is finite (i.e., a finite variation). So it can be written as two gamma processes, accounting for the price increases and the price decreases, respectively. For the NIG

process, it has a infinite variation, which well captures log prices if trading is continuous and frictionless, according to Harrison, Pitbladdo and Schaefer (1984).[1]

3.2.2 Characteristic Function

A tractable pricing formula for a VIX derivative can be derived from the characteristic function of the log VIX by using the inverse Fourier transform (for more details, see Heston, 1993; Carr & Madan, 1999). Our characteristic function can be obtained by solving a system of ordinary differential equations which is derived from the two-factor model (3.1) and the results are summarized in following proposition.

Proposition 3.2.1 *Under (3.1), we assume that the characteristic function of* $\{x_t\}_{t\geq 0}$ *has an exponential-affine form. Then it can be given by*

$$f(t,T,x_t,m_t,v_t;u) = \mathbb{E}^{\mathbb{Q}}\left[e^{iux_T}|\mathcal{F}_t\right] = e^{\psi_0(\tau)+\psi_x(\tau)x_t+\psi_m(\tau)m_t+\psi_v(\tau)v_t}, \tag{3.2}$$

where $\tau = T - t$, $\psi_0(\tau)$ *and* $\psi_v(\tau)$ *satisfy the following system of ordinary differential equations*

$$\begin{cases} \psi_0'(\tau) = \kappa_m\theta_m\psi_m(\tau) + \kappa_v\theta_v\psi_v(\tau) + \frac{1}{2}\sigma_m^2\psi_m^2(\tau) - \theta_J\psi_x(\tau) + \Psi^J(-i\psi_x(\tau)), \\ \psi_v'(\tau) = -\kappa_v\psi_v(\tau) + \rho\sigma_v\psi_x(\tau)\psi_v(\tau) + \frac{1}{2}\psi_x^2(\tau) + \frac{1}{2}\sigma_v^2\psi_v^2(\tau), \end{cases} \tag{3.3}$$

where $\psi_0(0) = 0$, $\psi_v(0) = 0$, *a and* $\Psi^J(x)$ *are specified in Table 3.4; and* $\psi_x(\tau)$ *and* $\psi_m(\tau)$ *have the closed-form expressions*

$$\begin{aligned} \psi_x(\tau) &= iue^{-\kappa\tau}, \\ \psi_m(\tau) &= \frac{iu\kappa}{\kappa_m - \kappa}\left(e^{-\kappa\tau} - e^{-\kappa_m\tau}\right). \end{aligned}$$

[1] This is because log prices with bounded variation admit arbitrage opportunities.

Proof. See Appendix A.1.

Generally, the system (3.3) in Proposition 3.2.1 cannot be solved analytically, but it can be solved by applying the 4th-order Runge–Kutta method (for the 4th-order Runge–Kutta method, see Section 2.3.2, and for the application in financial research, see Kiesel and Rahe (2017)).

3.3 VIX Derivative Pricing

In this section, we derive the formulae for VIX options and VIX futures by using the characteristic function of the log VIX in Section 3.2.2, and the risk-neutral pricing approach is applied.

3.3.1 Futures Valuation

The price of a VIX futures contract is the expectation of the VIX value at a specified trading time in the future (CBOE, 2019). Let $F(t;T_j,\Theta)$ be the price of jth futures contract at the current time t and the expiry time T_j under $\mathbb{Q}$, where $j = 1,2,3,\ldots,N_F$. By definition, we have

$$F(t;T_j,\Theta) = \mathbb{E}^{\mathbb{Q}}[\mathrm{VIX}_{T_j}|\mathcal{F}_t], \tag{3.4}$$

where VIX_{T_j} is the VIX value at time T_j and $\mathrm{VIX}_{T_j} = e^{x_{T_j}}$. According to Proposition 3.2.1, we know that $f(t,T_j,x_t,m_t,v_t;u)$ is the expectation of $e^{x_{T_j}}$, conditional on the current information when the value of u is set as $-i$. So, we can get the valuation formula for a futures contract in terms of the characteristic function, that is,

$$F(t;T_i,\Theta) = f(t,T_j,x_t,m_t,v_t;-i). \tag{3.5}$$

3.3.2 Options Valuation

VIX options are European-style. We assume that the value of the VIX at maturity time T_j is VIX_{T_j} with the strike price K_j. The price of the jth call option is denoted by $C(t;T_j,K_j,\Theta)$, where $j = 1,2,3,\ldots,N_C$. Similarly, the price of the jth put option is denoted by $P(t;T_j,K_j,\Theta)$, where $j = 1,2,3,\ldots,N_P$. We value options under the discounted cash flow with the risk-free interest rate, r_t, at t.[2] Then we have

$$\begin{aligned} C(t;T_j,K_j,\Theta) &= e^{-r_t(T_j-t)}\mathbb{E}^{\mathbb{Q}}\big[(\text{VIX}_{T_j}-K_j)\mathbf{1}_{\text{VIX}_{T_j}>K_j}|\mathcal{F}_t\big] \\ &= e^{-r_t(T_j-t)}\Big[\mathbb{E}^{\mathbb{Q}}\big[\text{VIX}_{T_j}\mathbf{1}_{\text{VIX}_{T_j}>K_j}|\mathcal{F}_t\big]-\mathbb{E}^{\mathbb{Q}}\big[K_j\mathbf{1}_{\text{VIX}_{T_j}>K_j}|\mathcal{F}_t\big]\Big], \end{aligned} \tag{3.6}$$

and

$$\begin{aligned} P(t;T_j,K_j,\Theta) &= e^{-r_t(T_j-t)}\mathbb{E}^{\mathbb{Q}}\big[(K_j-\text{VIX}_{T_j})\mathbf{1}_{K_j>\text{VIX}_{T_j}}|\mathcal{F}_t\big] \\ &= e^{-r_t(T_j-t)}\Big[\mathbb{E}^{\mathbb{Q}}\big[K_j\mathbf{1}_{K_j>\text{VIX}_{T_j}}|\mathcal{F}_t\big]-\mathbb{E}^{\mathbb{Q}}\big[\text{VIX}_{T_j}\mathbf{1}_{K_j>\text{VIX}_{T_j}}|\mathcal{F}_t\big]\Big]. \end{aligned} \tag{3.7}$$

In order to reduce the complexity of calculation, following Garman and Kohlhagen (1983), we introduce the new measure $\mathbb{Q}_1$ with VIX_t as numeraire for the first term in Eq. (3.6) and the second term in Eq. (3.7). Based on Girsanov's theorem, we can get the Radon-Nikodym derivative $d\mathbb{Q}_1/d\mathbb{Q}$,

$$\frac{d\mathbb{Q}_1}{d\mathbb{Q}} = \frac{\text{VIX}_{T_j}}{e^{-r_t(T_j-t)}\text{VIX}_{T_j}} = \frac{e^{x_{T_j}}}{\mathbb{E}^{\mathbb{Q}}\big[e^{x_{T_j}}|\mathcal{F}_t\big]}, \tag{3.8}$$

where the expectation of $d\mathbb{Q}_1/d\mathbb{Q}$ under $\mathbb{Q}$ always equals to 1. According to the

[2]The interest rate at time t is denoted as r_t because we use the daily interest rate. We assume that the interest rate is not a constant over the whole period and varies daily.

definition of the characteristic function, we have

$$\begin{aligned} f_1(t,T,x_t,m_t,v_t;u) &= \mathbb{E}^{\mathbb{Q}_1}[e^{iux_{T_j}}|\mathcal{F}_t] \\ &= \int_{\infty}^{\infty} e^{iux_{T_j}} d\mathbb{Q}_1 \\ &= \int_{\infty}^{\infty} e^{iux_{T_j}} \frac{d\mathbb{Q}_1}{d\mathbb{Q}} d\mathbb{Q} \\ &= \mathbb{E}^{\mathbb{Q}}[e^{iux_{T_j}} \frac{d\mathbb{Q}_1}{d\mathbb{Q}}|\mathcal{F}_t]. \end{aligned}$$

Then we plug Eq. (3.8) into above equation and get the new characteristic function, $f_1(t,T_j,x_t,m_t,v_t;u)$, under the new measure $\mathbb{Q}_1$ is

$$f_1(t,T_j,x_t,m_t,v_t;u) = \frac{f(t,T_j,x_t,m_t,v_t;u-i)}{f(t,T_j,x_t,m_t,v_t;-i)}. \tag{3.9}$$

In line with Bakshi and Madan (1999), the price formulae of a call option and a put option, which can be derived in a standard process, are expressed as

$$\begin{aligned} C(t;T_j,K_j,\Theta) &= e^{-r_t(T_j-t)}\left[f(t,T_j,x_t,m_t,v_t;-i)\Phi_1(t;T_j,K_j,\Theta) - K\Phi_2(t;T_j,K_j,\Theta)\right], \\ P(t;T_j,K_j,\Theta) &= e^{-r_t(T_j-t)}[K(1-\Phi_2(t;T_j,K_j,\Theta)) \\ &\qquad - f(t,T_j,x_t,m_t,v_t;-i)(1-\Phi_1(t;T_j,K_j,\Theta))], \end{aligned} \tag{3.10}$$

where $\Phi_1(t;T_j,K_j,\Theta) = \mathbb{E}^{\mathbb{Q}_1}[\mathbf{1}_{\mathrm{VIX}_{T_j}>K_j}|\mathcal{F}_t]$ and $\Phi_2(t;T_j,K_j,\Theta) = \mathbb{E}^{\mathbb{Q}}[\mathbf{1}_{textVIX_{T_j}} > K_j|\mathcal{F}_t]$, which are given by

$$\begin{aligned} \Phi_1(t;T_j,K_j,\Theta) &= \frac{1}{2} + \frac{1}{\pi}\int_0^{\infty} \mathrm{Re}\left[\frac{e^{-iu\ln K_j} f_1(v_t,\mu_t,w_t,t,T_j;u)}{i\phi}\right] du, \\ \Phi_2(t;T_j,K_j,\Theta) &= \frac{1}{2} + \frac{1}{\pi}\int_0^{\infty} \mathrm{Re}\left[\frac{e^{-iu\ln K_j} f(v_t,\mu_t,w_t,t,T_j;u)}{i\phi}\right] du. \end{aligned} \tag{3.11}$$

In the next section, we use the Gauss-Laguerre quadrature method with an order of 10

to calculate numerically the integrals in Eq. (3.11).

3.4 Model Calibration

In this section, we outline the estimation procedure for our two-factor models and the best two-factor model (OU-VJ) of Mencía and Sentana (2013), by using the joint data set of VIX and VIX derivatives. Mencía and Sentana (2013) employ the log-normal OU process to model the VIX under the risk-neutral measure. They consider stochastic volatility and central tendency in their best model. Also, they specify the jump process in the volatility process, that is,

$$\begin{aligned} dx_t &= \kappa(m_t - x_t)dt + \sqrt{v_t}dW_t^x, \\ dm_t &= \kappa_m(\theta_m - m_t)dt + \sigma_m dW_t^m, \\ dv_t &= \lambda_J v_t dt + dJ_t, \end{aligned} \tag{3.12}$$

where $\{J_t\}_{t\geq 0}$ is a compound Poisson process (CP) with jump intensity λ_J, and the jump size follows an exponential distribution with mean δ_J. There is no correlation between the jump process and the Wiener processes.

3.4.1 Estimation Procedure

Pan (2002) and Eraker (2004) suggest that a joint model estimation approach can be used in estimating more complicated models and this approach can generate more accurate estimation results than a simple estimation method. Kandepu, Foss and Imsland (2008) notice that the optimal filter can help to significantly reduce the computational cost and improve the performance for the large data set. Based on their suggestions, we estimate our models by applying the UKF on our latent state variables and combining it with the MLE.

Kalman filters are a group of filters (e.g. basic Kalman filter, extended Kalman filter, unscented Kalman filter, and Kalman-Bucy filter) for estimating every instantaneous state of a process (Grewal & Andrews, 2015). The UKF is one of the most popular estimation techniques for the processes with continuous-time state space and can deal with non-linear stochastic systems (e.g. Julier & Uhlmann, 1997; Wan & Merwe, 2000; Särkkä, 2007; Kiesel & Rahe, 2017, etc.). The core of the UKF is the unscented transformation (known as deterministic sampling method) which is proposed by (Julier & Uhlmann, 1997). The UKF has two steps, prediction step and updating step. The two steps link the latent state space and the observation status. Also, we can apply different transformations in each step because the two steps are independent. Based on the characteristics of our models, we use linear transformation in the prediction step and unscented transformation in the updating step. We give the details of applying the UKF in Section 2.3.3 and how it performs for our two-factor model in Appendix A.2.

In order to apply the UKF for estimating the values of latent variables in the calibration process, we have to specify the observation equation (measurement function, see Eq. (3.15)) and the state equations which describe the dynamics of latent variables under the physical condition (Wehn, Hoppe & Gregoriou, 2013). Therefore, we think it is necessary to re-state our state processes in the physical measure, rather than the risk-neutral measure. In our estimation procedure, we treat the processes $\{m_t\}_{t\geq 0}$ and $\{v_t\}_{t\geq 0}$ as latent processes. We denote them in the vector form $\mathbf{X}_t = [m_t, v_t]^\top$. All the parameters in the models are stacked in Θ. Firstly, we discretize the latent processes by the Euler discretization method under the physical measure. To do so, we introduce new parameters for risk premiums which are related to the Wiener processes and jump

process. Then, the discrete version for our models can be written as

$$\begin{aligned}\mathbf{X}_t = &\begin{pmatrix} \kappa_m\theta_m\Delta t \\ \kappa_v\theta_v\Delta t \end{pmatrix} + \begin{pmatrix} 1+(\eta_m-\kappa_m)\Delta t & 0 \\ 0 & 1+(\eta_v-\kappa_v)\Delta t \end{pmatrix}\mathbf{X}_{t-1} \\ &+ \begin{pmatrix} \sigma_m\sqrt{\Delta t} & 0 \\ 0 & \sigma_v\sqrt{v_{t-1}\Delta t} \end{pmatrix}\mathbf{e}_t,\end{aligned} \tag{3.13}$$

where $\mathbf{e}_t = [e_t^m, e_t^v]^\top$, $\{e_t^m\}_{t\geq 0}$ and $\{e_t^m\}_{t\geq 0}$ follow different standard normal distributions, η_m and η_v are the risk premia of the volatility and log-term mean processes, respectively.

Similarly, the discretized version for OU-VJ model can be written as

$$\begin{aligned}\mathbf{X}_t = &\begin{pmatrix} \kappa_m\theta_m\Delta t \\ \eta_\lambda\lambda_J(\delta_J+\eta_\delta)\Delta t \end{pmatrix} + \begin{pmatrix} 1+(\eta_m-\kappa_m)\Delta t & 0 \\ 0 & 1-\eta_\lambda\lambda_J\Delta t \end{pmatrix}\mathbf{X}_{t-1} \\ &+ \begin{pmatrix} \sigma_m\sqrt{\Delta t} & 0 \\ 0 & 2(\delta_J+\eta_\delta)\sqrt{2\eta_\lambda\lambda_J\Delta t} \end{pmatrix}\mathbf{e}_t,\end{aligned} \tag{3.14}$$

where $\mathbf{e}_t$ and η_v have the same specification as those in Eq. (3.13), and η_δ and η_λ are the risk premia related to the jump size and jump jump intensity, respectively.

Secondly, the measurement errors between our estimated prices and market prices are considered in the procedure. For this purpose, we let the measurement function be

$$\mathbf{Y}_t = \mathbf{f}(\mathbf{X_t}; x_t, \Theta) + \epsilon_t, \tag{3.15}$$

where the vector $\mathbf{Y}_t$ represents the log market prices of futures and market prices of options, the vector $\mathbf{f}$ denotes the proposed VIX derivatives' pricing formulae including the futures pricing formula ($\ln F(t; T_i, \Theta)$) and the option pricing formulae ($C(t; T_i, K_i, \Theta)$ and $P(t; T_i, K_i, \Theta)$), and ϵ_t is the term for measurement error. Following Yang and

Kanniainen (2017), we adjust the measurement errors for options with their market vegas ($\mathcal{V}(t;T_i,K_i)$). If there are a total of k_t derivatives' contracts at time t, then ϵ_t is a k_t-dimensional random vector, which follows an independently multivariate normal distribution ($\mathcal{N}(\mathbf{0},\Sigma_t^\epsilon)$), where Σ_t^ϵ is the covariance matrix,

$$\Sigma_t^\epsilon = \begin{bmatrix} \sigma_\epsilon^2 & 0 & 0 & \dots & 0 \\ 0 & \sigma_\epsilon^2 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \sigma_\epsilon^2 \end{bmatrix}, \tag{3.16}$$

with $\sigma_\epsilon = \sigma_F$ for futures and $\sigma_\epsilon = \mathcal{V}(t;T_i,K_i)\sigma_O$ for options. We assume that σ_F and σ_O are constants for all futures and options, respectively.

Finally, in line with Christoffersen, Dorion, Jacobs and Karoui (2014) and Du and Luo (2019), we construct the daily log-likelihood function of VIX derivatives at time t as

$$l_t(\Theta) = -\frac{1}{2}\left[k_t \ln 2\pi + \ln|\Sigma_t^{\mathbf{Y}}| + \epsilon_t^\top (\Sigma_t^{\mathbf{Y}})^{-1}\epsilon_t\right], \tag{3.17}$$

where $\Sigma_t^{\mathbf{Y}}$ is the variance matrix of $\mathbf{Y}_t$ at time t. The likelihood function for our whole sample period can be computed as

$$l(\Theta) = \sum_{t=1}^{T} l_t(\Theta), \tag{3.18}$$

where T is the total number of weeks. Our parameters are estimated by minimizing negative log-likelihood ($-l(\Theta)$). This is equivalent to the maximum log-likelihood estimation (MLE). In optimization, we employ a quasi-Newton algorithm and the Hessian matrix updated by Broyden-Fletcher-Goldfarb-Shanno (BFGS) method (see Broyden, 1956; Fletcher, 1987; Lewis & Overton, 2013).

3.4.2 Parameter Estimation Results

Table 3.5 presents the estimated results for the OU-VJ model and our two models with infinite jumps, that is, the OU-VG and OU-NIG models. We notice that the standard errors for all parameters are relatively small, except for the parameters η_m and η_v. This indicates that the majority of parameters in the models are accurately estimated. Please note that η_m and η_v are not our key parameters, so it does not matter whether they have slightly high standard errors. In the following, we will analyze and compare our estimates.

Table 3.5: Model parameter estimates

	OU-VJ	**OU-VG**	**OU-NIG**		**OU-VJ**	**OU-VG**	**OU-NIG**
ρ		0.5584	0.6350	λ_J	3.4332		
		(0.0216)	(0.0177)		(0.2420)		
κ	7.6175	7.5579	7.8276	δ_J	0.2684		
	(0.0373)	(0.0797)	(0.0737)		(0.0000)		
κ_m	0.9108	0.7194	0.7314	η_δ	-0.0010		
	(0.0049)	(0.0252)	(0.0242)		(0.0000)		
κ_v		1.2510	1.1977	η_λ	1.1347		
		(0.1010)	(0.0612)		(0.0800)		
θ_m	2.9495	3.2421	3.2729	α		0.3875	2.3837
	(0.0004)	(0.0090)	(0.0087)			(0.0554)	(0.0208)
θ_v		2.0366	2.2883	β		0.3681	0.8479
		(0.1269)	(0.1449)			(0.0623)	(0.0262)
σ_m	0.3412	0.4355	0.3940	λ		0.9717	0.8689
	(0.0005)	(0.0188)	(0.0122)			(0.1168)	(0.0125)
σ_v		2.7902	3.0377				
		(0.0720)	(0.0609)				
η_m	-0.0065	-0.0815	-0.0560	σ_O	0.0847	0.0653	0.0652
	(0.0001)	(0.0472)	(0.0300)		(0.0000)	(0.0005)	(0.0005)
η_v		-2.1331	-2.1267	σ_F	0.0327	0.0284	0.0285
		(0.4263)	(0.7345)		(0.0006)	(0.0006)	(0.0006)

The model parameters are estimated through the procedure as described in Section 3.4.1 by using the in-sample data from July 2006 to January 2013 (total 342 weeks). The standard errors of the estimators are presented in parentheses.

The parameters κ and κ_m measure the reverting speed towards their means. Intuitively, the smaller value of the reverting speed parameter suggests that the process is more persistent. The value of κ varies between 7.56 and 7.83 across the three models, which does not exhibit a significant difference among them. However, the value of

κ_m in the OU-VJ model is relatively higher than the values of our two infinite-activity jump models. This fact implies that the long-term mean process in the OU-VJ model is less persistent and has faster reverting speed than those processes in the OU-VG and OU-NIG models. Moreover, the parameter σ_m, denoting the volatility of the long-term mean process, is estimated at 0.34, 0.44, and 0.39 for the OU-VJ, OU-VG and OU-NIG, respectively. This indicates that the long-term mean process in the infinite-activity models is more volatile than that in the OU-VJ model. In addition, the risk premium (η_m) related to the long-term process has a negative value in all models. This means that the long-term mean reverts to the long-term level under the physical measure, which is lower than that under the risk-neutral measure.

Our estimates of variance risk premium (η_v) in our two infinite-activity jump models are negative values, which is consistent with the previous literature (Park, 2015). The estimated value of η_v for the OU-VG model is -2.13 which is very close to the η_v value (-2.13) of the OU-NIG model. This result supports the economic interpretation that the volatility of volatility has a negative market price. The correlation coefficient ρ is estimated to be greater than 0.5 for both models. The positive ρ value implies that there is a positive interaction between the logarithmic VIX process and its variance process. The larger ρ is, the stronger interaction is. In our estimated results, the value of ρ in the OU-NIG model is greater than that in the OU-VG model. Thus, different jump structures can affect the degree of the interaction between the two processes.

3.5 Model Performance Analysis

In this section, we investigate what kind of model specification (e.g., jump structure and jump type) have the best pricing performance in pricing VIX derivatives and how the different jump structures affect both in-sample and out-of-sample model performance. To answer these questions, we compare the performance of the models, based on

comparison criteria such as root mean squared errors and log-likelihood values.

3.5.1 Criteria of Performance

We use the root mean square error (RMSE) as our primary model comparison criterion. A smaller value of the RMSE indicates the model fits the data better on average. Firstly, we define the squared pricing error (SE) of a futures contract at time t as

$$\text{SE}^F(t,i;\Theta) = \left(\ln \hat{F}(t;T_i,\Theta) - \ln F(t;T_i)\right)^2, \tag{3.19}$$

where $\hat{F}(t;T_i,\Theta)$ is the estimated futures price by the model and $F(t,T_i)$ is the market price of futures. Then the RMSE for the futures is calculated as

$$\text{RMSE}^F(\Theta) = \sqrt{\frac{1}{N^F}\sum_{t=1}^{N_T}\sum_{i=1}^{N_t^F} SE^F(t,i;\Theta)}, \tag{3.20}$$

where N_T is the total sample period, N_t^F is the total number of future contracts at time t, and $N^F = \sum_{t=1}^{N_T} N_t^F$.

Following Kanniainen, Lin and Yang (2014), we scale the squared errors of call and put options by the market vegas. The SE of an option contract at time t is defined as

$$\text{SE}^O(t,i;\Theta) = \left(\frac{\hat{O}(t;T_i,K_i,\Theta) - O(t;T_i,K_i)}{\mathcal{V}(t;T_i,K_i)}\right)^2, \tag{3.21}$$

where $\hat{O}(t;T_i,K_i,\Theta)$ is the estimated price of the ith option contract at time t by the model, $O(t;T_i,K_i)$ is the market price of this option, and $\mathcal{V}(t;T_i,K_i)$ is its corresponding market vega. The RMSE for call and put options are calculated as

$$\text{RMSE}^O(\Theta) = \sqrt{\frac{1}{N^O}\sum_{t=1}^{N_T}\sum_{i=1}^{N_t^O} SE^O(t,i;\Theta)}, \tag{3.22}$$

where N_T is the same as that in Eq. (3.20), N_t^O is the total number of options at time t, and $N^O = \sum_{t=1}^{N_T} N_t^O$.

To compare the pairwise models, we apply the comparison procedure in Kaeck, Rodrigues and Seeger (2018) and Hansen, Lunde and Nason (2011). We define the performance criteria between Model i and Model j at time t as $d_t^{i,j} = l_t^i(\Theta) - l_t^j(\Theta)$ where l_t^i is the negative log-likelihood function of Model i. The log-likelihood function is defined in Eq (3.17). We let $u^{i,j}$ be the expectation of $d_t^{i,j}$, which is

$$u^{i,j} = \mathbb{E}[d_t^{i,j}] = \frac{\sum_{t=1}^{T} d_t^{i,j}}{T}, \tag{3.23}$$

where T is total number of weeks in the sample. Note that $u^{i,j}$ measures the performance loss of Model i compared with Model j. The negative value of u_{ij} indicates an average performance loss of Model i against Model j. Conversely, the positive value implies an average performance gain of Model i against Model j.

3.5.2 Model Performance Comparison

In this section, we compare the overall performance of the OU-VG, OU-NIG and OU-VJ models in terms of log-likelihood values. Table 3.6 reports the model performance ranking based on this criterion. Panel A in this table presents the total log-likelihood values of each model, and Panel B provides the pairwise model comparison based on the average daily log-likelihood value. According to Panel A, we can explicitly see that the performance of our two-factor models (OU-VG and OU-NIG) is significantly better than that of the OU-VJ model because both the in-sample and out-of-sample log-likelihoods of our models are three times more than that of the OU-VJ model. Among these three models, the OU-NIG model has the highest log-likelihood values (5,391.84 for in-sample and 5,120.83 for out-of-sample); thus this model is superior to the others in pricing VIX derivatives.

In the pairwise model comparison, we focus on analyzing the model performance loss ($u^{i,j}$) which can provide more information about the comparison of model performance. In Panel B, we find that there are negative entries $u^{i,j}$ in the upper triangular parts of pairwise performance measure matrices, which leads to the same conclusion as that from Panel A. In contrast to the OU-VG and OU-NIG models, the OU-VJ model has a performance loss more than 20 during the in-sample period and more than 40 during the out-of-sample period. This implies that the OU-VG and OU-NIG models are more reliable than the OU-VJ model in forecasting. Furthermore, the OU-NIG model is slightly superior to the OU-VG model.

Table 3.6: Model performance comparison – log-likelihood

	In-Sample			**Out-of-Sample**		
	OU-VJ	OU-VG	OU-NIG	OU-VJ	OU-VG	OU-NIG
	Panel A: Total log-likelihood values					
$l(\Theta)$	1677.04	5377.20	5391.84	1564.30	5046.01	5120.83
	Panel B: Pairwise model comparison - performance loss ($u^{i,j}$)					
OU-VJ	—	-22.81	-22.89	—	-41.70	-42.59
OU-VG	22.81	—	-0.09	41.70	—	-0.90
OU-NIG	22.89	0.09	—	42.59	0.90	—

This table shows the total log-likelihood values of the models and performance loss matrices which are computed based on the estimated parameters in Table 3.5. The in-sample and out-of-sample periods are from July 2006 to January 2013 and from February 2013 to April 2016, respectively. The total log-likelihood values are evaluated by Eq. (3.18) are presented in Panel A. Panel B presents pairwise model comparison evaluated by Eq. (3.23), where Model i is listed vertically and Model j listed horizontally.

3.5.3 Fit to VIX Futures and Options Pricing

In this subsection, we examine the model performance in pricing futures and options in order to investigate which model specification can better capture the dynamics of VIX futures and options prices, in terms of the SE and the RMSE.

Figure 3.1 illustrates the empirical CDFs of the squared futures pricing errors. In these CDF graphs, the OU-VJ curve (blue) is below the OU-NIG curve (yellow) and the OU-VG curve (red). Also, the yellow curve is slightly above the red curve. This

indicates that the OU-VG and OU-NIG models significantly outperform the OU-VJ model in pricing futures.

Table 3.7 compares the model performance in terms of the RMSE. Overall, the OU-VJ model has the largest in-sample RMSE (0.0318) and out-of-sample RMSE (0.0474). The differences between the OU-VG and the OU-NIG models are 0.0001 for the in-sample and 0.0002 for the out-of-sample, respectively. These differences are insignificant.

Panel B presents the RMSE of each model by time to maturity. We classify time to maturity into three categories: short-term ($\tau \leq 30$), middle-term ($30 < \tau \leq 90$) and long-term ($\tau > 90$). In comparing with the OU-VJ model, we observe that our jump models take more advantages in pricing middle-term futures than other term futures for both in-sample and out-of-sample because our models improve the RMSE over 24% in-sample and over 50% out-of-sample, respectively. Furthermore, the OU-VG model can best capture the features of long-term futures price data. In contrast, the OU-NIG model outperforms in the long-term category.

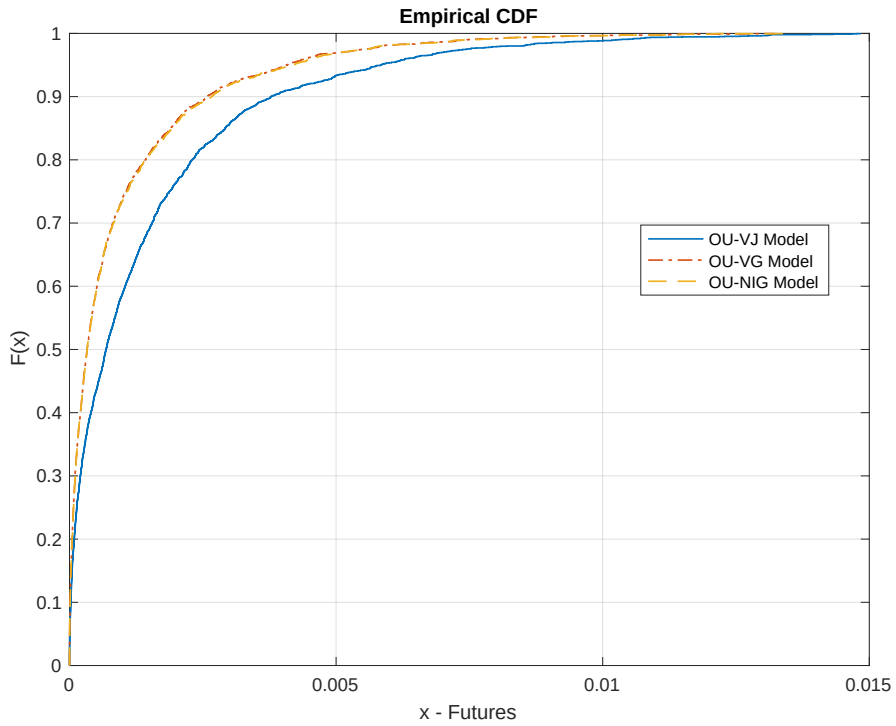
Table 3.7: Model performance comparison pricing futures – RMSE

	In-Sample			**Out-of-Sample**		
	OU-VJ	OU-VG	OU-NIG	OU-VJ	OU-VG	OU-NIG
			Panel A: Overall			
	0.0318	0.0278	0.0279	0.0474	0.0347	0.0349
			Panel B: Sorting by time to maturity			
$\tau \leq 30$	0.0275	0.0240	0.0239	0.0245	0.0230	0.0222
$30 < \tau \leq 90$	0.0277	0.0209	0.0205	0.0398	0.0186	0.0190
$\tau > 90$	0.0356	0.0329	0.0333	0.0551	0.0434	0.0438

This table shows the average RMSE of futures prices. The in-sample and out-of-sample periods are from July 2006 to January 2013 and from February 2013 to April 2016, respectively. The RMSE is computed from Eq. (3.20) based on the estimated parameters in Table 3.5. In Panel B, the time to maturity is denoted by τ.

Figure 3.2 compares the difference of the half-yearly average SE for the paired models from 2006 to 2016 (full-sample period). From the subplots (a) and (b), we find

Figure 3.1: CDF of the squared pricing errors for futures



Notes: CDF is the abbreviation of empirical cumulative distribution function. The squared pricing errors are calculated from Eq. (3.19) by using the full sample data from July 2006 to April 2016, based on the estimated parameters in Table 3.5.

that the half-yearly average SE of OU-VJ is significantly larger than that of OU-VG and OU-NIG through the whole sample period. It is not surprising that the pattern in the subplot (a) is very similar to that in the subplot (b). This is because the difference between squared pricing errors of the OU-VG model and squared pricing errors of the OU-NIG model is incredibly tiny, which can be observed from the subplot (c). Therefore, we can conclude that the OU-VG and OU-NIG models perform significantly better than the OU-VJ model in pricing options.

Table 3.8 reports the RMSE under different categories: time to maturity, moneyness, and volume. Overall, we can find that the RMSE of the OU-VJ model is two times more than those of the OU-VG and OU-NIG models. Thus, the OU-VG and OU-NIG models are dramatically superior to the OU-VJ model during both in-sample and out-of-sample

periods. More specifically, the OU-VG and OU-NIG models can fit both call and put option price dynamics significantly better than the OU-VJ model during the whole sample period because the OU-VG and OU-NIG models decrease the errors by over 50% in pricing both call and put options. Then, we further investigate how the models perform under the different categories. Under Panel B, the OU-VJ model underperforms our two models in every case. The OU-NIG model is the best model in most cases, but the OU-VG model has the best performance when the time to maturity is greater than 90 days. Under Panel D, we observe that the RMSE of the OU-VG and OU-NIG models increase for both in-sample and out-of-sample periods, whereas the RMSEs of the OU-VJ model decreases when the trading volume increases. The in-sample RMSEs of the OU-VG and OU-NIG models increase by 17.64% and 17.18%, respectively and the out-of-sample RMSEs increase by 24.80% and 22.22%, respectively. The in-sample and out-of-sample RMSEs of the OU-VJ model decrease by 9.33% and 4.25%, respectively.

3.5.4 Finite- vs. Infinite-activity Jumps

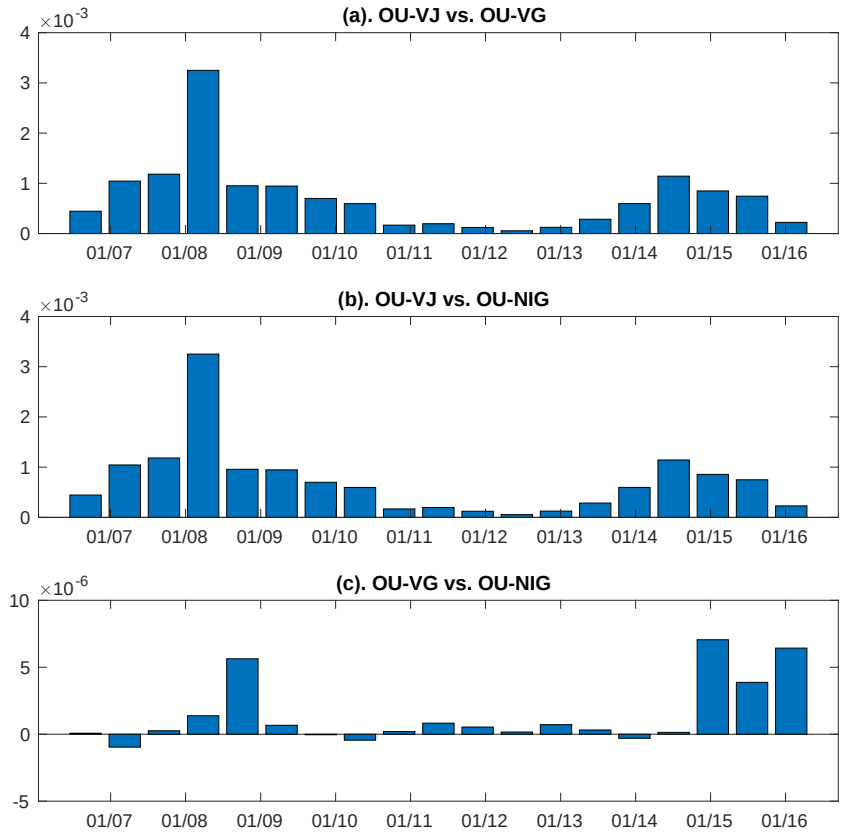
To further investigate the evidence of infinite-activity jumps in pricing VIX derivatives, we examine the performance of the finite-activity jump model which has the same specifications of long-term mean and volatility process as our models (OU-VG and OU-NIG), but the jump component is assumed to be a compound Poisson process (CP). This finite-activity jump model is the best model in Park (2016), in which the jump part in Eq. (3.1) can be further assumed as $dL_t^J = dJ_t^+ + dJ_t^- - \lambda_+\delta_+dt - \lambda_-\delta_-dt$, where J_t^+ and J_t^- follow compound Poisson processes with positive and negative jump intensity λ_+ and λ_-, respectively. The two jump processes are also independent from each other and from the Wiener processes in Eq. (3.1). Their jump sizes follow different exponential distributions. But the positive jumps have a positive jump size with mean δ_+. The negative jumps have a negative jump size with mean δ_-.

Table 3.8: Model performance comparison in pricing options – RMSE

	In-Sample			**Out-of-Sample**		
	OU-VJ	OU-VG	OU-NIG	OU-VJ	OU-VG	OU-NIG
			Panel A: Overall			
All	0.1435	0.0662	0.0658	0.1571	0.0739	0.0725
Calls	0.1320	0.0638	0.0634	0.1433	0.0657	0.0634
Puts	0.1660	0.0713	0.0708	0.1910	0.0935	0.0937
			Panel B: Sorting by time to maturity			
$\tau \leq 30$	0.1530	0.0891	0.0887	0.1911	0.1230	0.1200
$30 < \tau \leq 90$	0.1400	0.0576	0.0569	0.1471	0.0632	0.0615
$\tau > 90$	0.1420	0.0587	0.0590	0.1523	0.0511	0.0517
			Panel C: Sorting by moneyness			
K/F ≤ 0.9	0.1833	0.0639	0.0632	0.2127	0.0890	0.0908
0.9 < K/F ≤ 0.98	0.1382	0.0798	0.0796	0.1708	0.0982	0.0969
0.98 < K/F ≤ 1.02	0.1620	0.0914	0.0915	0.1944	0.0901	0.0898
1.02 < K/F ≤ 1.1	0.1648	0.0834	0.0838	0.2076	0.0635	0.0635
K/F > 1.1	0.1214	0.0573	0.0567	0.1250	0.0653	0.0625
			Panel D: Sorting by volume			
Vol ≤ 500	0.1501	0.0618	0.0617	0.1601	0.0653	0.0648
500 < Vol ≤ 3000	0.1430	0.0652	0.0647	0.1596	0.0728	0.0715
3000 < Vol ≤ 8000	0.1400	0.0693	0.0685	0.1567	0.0764	0.0748
Vol > 8000	0.1361	0.0727	0.0723	0.1533	0.0815	0.0792

This table lists the average RMSEs of option prices in terms of time to maturity (Panel B), moneyness (Panel C), and volume (Panel D) for the in-sample and out-of-sample periods from July 2006 to January 2013 and from February 2013 to April 2016, respectively. The RMSEs are computed from Eq (3.22) based on the estimated parameters in Table 3.5. In Panel B, the time to maturity is denoted by τ.

Figure 3.2: Comparison of half-yearly average SEs



Notes: The period for comparison is between 2006 and 2016 (full-sample period). The y-axis represents the differences between the average SEs of Model A and the average SEs of Model B. For example, in the subplot (a), the magnitude of each bar equals to average SE of the OU-VJ model subtracts the average SE of the OU-VG model. The x-axis represents the sample period. The squared pricing errors are calculated by Eq. (3.21) based on estimated parameters in Table 3.5.

We apply the same model estimation procedure to calibrate the OU-CP model. The estimation results are reported in Table 3.9. In this table, we find that the OU-CP model has a positive ρ and a negative η_v. This fact is consistent with the economic phenomenon discussed in Section 3.4.2. In comparison with the results in Table 3.5, the values of κ, κ_m, κ_v and σ_v in the OU-CP model are smaller than those in our models. Therefore, the infinite-activity jump structure makes the model more flexible and the variance process of the model fluctuates more.

Table 3.9: Parameter estimates - OU-CP

Para.	ρ	κ	κ_m	κ_v	θ_m	θ_v	σ_m	σ_v
Est.	0.5887	6.6532	0.6297	0.8860	3.2436	2.0028	0.3819	2.1640
Sd. err.	0.0256	0.0758	0.0274	0.0807	0.0110	0.1250	0.0161	0.0677
Para.	η_m	η_v	σ_O	σ_F	λ_+	δ_+	λ_-	δ_-
Est.	-0.0557	-2.1484	0.0677	0.0285	2.5948	0.2675	2.0418	-0.1414
Sd. err.	0.0432	0.3974	0.0006	0.0006	0.2675	0.0104	0.2766	0.0036

The model parameters are estimated through the procedure in Section 3.4.1. by using the in-sample data from July 2006 to January 2013 (total 342 weeks). Para. is the abbreviation of parameter. The estimated results are listed in the Est. row, and standard errors of the estimators are listed in the Sd. err. row.

Moreover, we compare the performance of the OU-CP model with our infinite-activity jump models regarding the log-likelihood and the RMSE. Firstly, comparing with the log-likelihood values of our infinite-activity jump models in Table 3.6, the log-likelihood value of the OU-CP model, in Table 3.10, drops by at least 200 in-sample from 5,377.20 (OU-VG) or 5,391.84 (OU-NIG) and at least 350 out-of-sample from 5,046.01 (OU-VG) or 5,120.83 (OU-NIG), respectively.

From Table 3.10 and Table 3.7, we do not observe big differences among the overall RMSEs of future prices of OU-CP, OU-VG, and OU-NIG. The infinite-activity jump structures seem not to improve substantially the overall model performance in pricing VIX futures. However, the infinite-activity jumps can achieve improvement in pricing VIX futures with relatively short-term maturities which are less than or equal to 90 days.

Comparing Table 3.10 with Table 3.8 , we find that the VG jump structure reduces the RMSE in pricing VIX options by 0.0012 from 0.0674 (OU-CP) to 0.0662 (OU-VG), and NIG jump structure reduces the RMSE by 0.0016 from 0.0674 (OU-CP) to 0.0658 (OU-NIG) during the in-sample period. For the out-of-sample period, the RMSE drops by 0.0006 from 0.0745 (OU-CP) to 0.0739 (OU-VG) and by 0.0020 from 0.0745 (OU-CP) to 0.0725 (OU-NIG). To test whether the improvements are significant or not, we apply the t-test on the daily option RMSE of the whole sample. The p-value of t-test

is 0.0146 between the OU-CP and OU-VG models, and 0.0001 between the OU-CP and OU-NIG models. Both of p-values are less than 0.0500, so we can conclude that the VG and NIG jump structures make statistically significant improvement in pricing VIX options. Particularly, the infinite-activity jump structures noticeably improve the accuracy in pricing put options.

Table 3.10: Model performance – OU-CP

Properties		**In-Sample**		**Out-of-Sample**	
Log-likelihood	$l(\Theta)$	5154.55		4693.61	
		Options	Futures	Options	Futures
			Overall		
	All	0.0674	0.0277	0.0745	0.0350
	Calls	0.0635	—	0.0659	—
	Puts	0.0752	—	0.0948	—
			Sorting by time to maturity		
RMSE	$\tau \leq 30$	0.0942	0.0249	0.1193	0.0242
	$30 < \tau \leq 90$	0.0579	0.0215	0.0661	0.0210
	$\tau > 90$	0.0562	0.0322	0.0525	0.0431
			Sorting by time to moneyness		
	K/F ≤ 0.9	0.0669		0.0915	
	0.9 < K/F ≤ 0.98	0.0844		0.0978	
	0.98 < K/F ≤ 1.02	0.0934		0.0949	
	1.02 < K/F ≤ 1.1	0.0821		0.0669	
	K/F > 1.1	0.0575		0.0647	

This table reports the OU-CP performance in terms of log-likelihood values and RMSEs for the in-sample and out-of-sample period from July 2006 to January 2013 and from February 2013 to April 2016, respectively. The log-likelihoods are evaluated by Eq. (3.18), and the RMSEs are computed from the Eq. (3.22) based on the estimated parameters in Table 3.5. The time to maturity is denoted by τ.

In summary, in addition to that the pricing errors can be reduced by introducing an infinite-activity jump structure instead of a finite-activity jump structure to our model, we also find that our model has two additional advantages over the OU-CP model. First, the number of parameters for describing the jump structure in the OU-VG and OU-NIG models is less than that in the OU-CP model. As a result, for the OU-VG and OU-NIG models, the computational cost can be reduced and the calibration accuracy can be increased. Although the VG and NIG processes have complex Lévy measures, their characteristic exponents have simple representations. Second, the OU-VG and

OU-NIG models have better out-of-sample prediction performance, which is considered as an important factor in the model comparison. By using Eq. (3.23), we can calculate the average daily model performance loss against the OU-CP model as 4.22 and 5.12, respectively, for the out-of-sample period.

Chapter 4

Pricing VXX Options under Lévy Process

This chapter is joint work with J. Cao, X. Ruan and W. Zhang. Its earlier version was presented at AUT Mathematical Sciences Symposium, November 2019, AUT, Auckland, New Zealand.

Since the models with an infinite-activity jump process show better performance in pricing VIX derivatives, we think that infinite-activity jumps may also occur in the VXX ETN returns. In this chapter, we introduce more types of infinite-activity jump structures, including the normal tempered stable processes and generalized tempered stable processes, in modelling the VXX ETN returns, to figure out the "real" dynamics of the returns. First, we conduct a comprehensive analysis on the VXX and VXX options data in Section 4.1. In Section 4.2, we construct the model for the VXX ETN returns. In Section 4.3, we derive the VXX option pricing formulas. In Section 4.4, we calibrate the models, analyze the estimation results and compare the performance of different models. In the last section of this chapter, we compare and analyze the model performance. The measure change of a Lévy process with different jump structures and the characteristics function derivation by using the time-changed method are given in

Appendix B.

4.1 Data

Our sample includes daily VXX and VXX options data. The selected sample period starts on 25 May 2010, and ends on 28 December 2017, including a total of 1,912 trading days, which covers a much wider time frame than that in Bao et al. (2012).[1] We equally split the full sample period into an in-sample period from 28 May 2010 to 27 March 2014 and an out-of-sample period from 28 March 2014 to 28 December 2017. In the following sub-sections, we focus on analyzing the VXX and VXX options data for both the full sample and the two sub-samples.

4.1.1 VXX

The VXX consists of the first-month and the second-month VIX futures contracts and their weights are adjusted daily (Rhoads, 2011). In the contango market, VIX futures prices decrease slowly, resulting in a reverse split on the stock. Thus, Barclays Bank splits VXX reversely with a ratio of 1-to-4 to avoid the share price of VXX dropping to an extremely low level. Here, the split ratio means that the shareholder owns four shares before the split and then owns one share after the split. There is a total of five reverse-splits on the VXX since the inception of VXX trading. The detailed information on the reverse-splits is shown in Table 4.1. During the model calibration procedure, the price of VXX should be adjusted in accordance with the information in Table 4.1.

Figure 4.1 demonstrates the evolution of the daily VXX close price (in graph (a)) and its logarithm (in graph (b)), with the reverse split dates, between 28 May 2010 and 28 December 2017. According to graph (a) in this figure, we find that the VXX

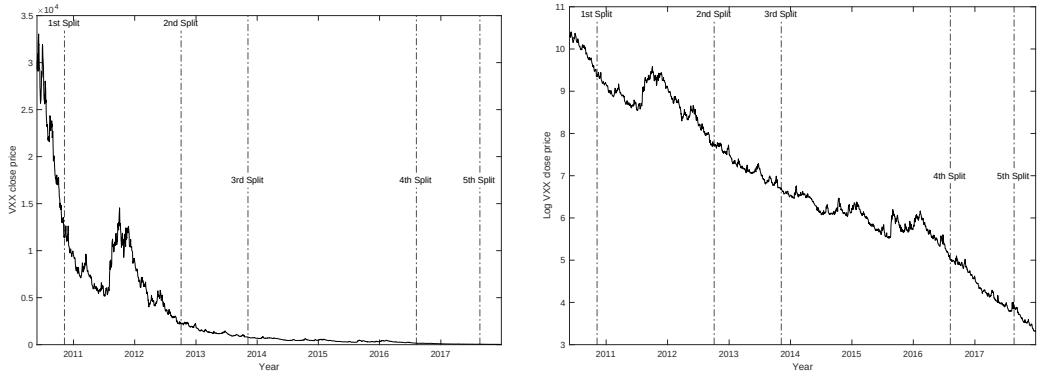
[1] VXX was launched on 29 January 2009. Our sample period starting on 25 May 2010 is due to the availability of the VXX options data. The data end in 2017 when this work was undertaken.

Table 4.1: VXX reverse split

Event	Split Date	Split Ratio	Close price	Months
1st Rev. Split	09 November 2010	4:1	11,563.52	21
2nd Rev. Split	05 October 2012	4:1	2,214.40	23
3rd Rev. Split	08 November 2013	4:1	812.80	13
4th Rev. Split	09 August 2016	4:1	148.80	34
5th Rev. Split	23 August 2017	4:1	47.36	12

This table shows the VXX reverse splits information from 28 May 2010 to 28 December 2017. The price under the fourth column is the close price right before the reverse split. These prices are not adjusted in accordance with the reverse splits. The number in the last column refers to the total number of months since inception or the previous split date.

price drops from a considerably high value $(29,265.92)$ to a relatively low value (only 27.47, which is the lowest price level during the whole sample period). Due to the large difference in the prices, we apply a logarithmic transformation to the VXX daily close prices to smooth the data and reduce the skewness of the original data; see graph (b) in Figure 4.1. Overall, both graphs display a downward trend during the whole sample period, although it has fluctuated all the time.



(a) VXX (b) The logarithm of VXX

Figure 4.1: Daily close price evolution of the VXX and logarithmic VXX during the period between 25 May 2010 and 18 December 2017.

We also notice that the VXX close price is considerably volatile between the first and second reverse-splits in Figure 4.1. During this period, it drops to a relatively lower

point at $5,148.16$ on 07 July 2011, then rises to a local maximum point at $14,551.04$ on 03 October 2011. After reaching the local maximum, it decreases again to a much lower price level $(2,193.92)$ just before the second split. This period takes a total of about 23 months. From 2012 to 2013, we do not observe big rises or falls in Figure 4.1, and the period between the second and third reverse-splits only takes about 13 months. Furthermore, the fourth reverse-split is conducted in about 34 months after the third reverse-split. This is because the volatility market is relatively stable in this time period, with the result that the price of VXX decays at a lower rate[2] (less than 5% per month) than in the other periods. After the fourth reverse-split, the low volatility of the stock market accelerates the speed of splitting. Thus, the fifth reverse-split occurs in 12 months after the fourth reverse-split.

The statistics of VXX and logarithmic VXX series are summarized in Table 4.2. The range of VXX is wide, from 27.47 to 33044.48. After taking the logarithm of VXX, the range becomes much more narrow, between 3.31 and 10.41. The higher moments confirm that the distribution of VXX is highly skewed to the left with tails heavier than those of the normal distribution. In the logarithm scale, the distribution of the VXX ETN returns is more symmetric with lighter tails and the kurtoses value is very close to zero. Then, we investigate the statistical properties of their first differences. From Table 4.2, we can observe that the kurtosis of ΔVXX and $\Delta\log(\text{VXX})$ are much greater than 3 and the skewness of them are also larger than zero, so we can conclude that the distributions have negative mean values, heavier tails and left skewness. In Table 4.2, the auto-correlations of VXX and $\log(\text{VXX})$ at lag one and lag two are very close to 1 and decay slowly, which indicates these series behave with high persistence in the

[2]The decay rate can be obtained by solving the following equation:

$$f = a(1-r)^t,$$

where f is the final VXX value of the period, a is the VXX value at the beginning of the period, t is the total time, and r is the decay rate.

sample period.

Table 4.2: Descriptive statistics of daily VXX

	VXX	$\log$(VXX)	ΔVXX	$\Delta \log$(VXX)
Mean	3482.3857	6.7809	-15.3483	-0.0037
Standard Deviation	5824.4943	1.8025	267.9216	0.0386
Skewness	2.5407	0.0800	1.2291	0.7051
Kurtosis	9.9195	2.0902	29.6603	6.0197
Min	27.4700	3.3131	-1843.2000	-0.1424
Median	674.4800	6.5139	-1.3200	-0.0058
Max	33044.4800	10.4056	2959.3600	0.2177
$\rho(1)$	0.9937	0.9978	0.0817	-0.0204
$\rho(2)$	0.9866	0.9956	0.0286	0.0000
$\rho(50)$	0.6950	0.8942	0.0000	-0.0198

This table statistically summarizes the VXX, the logarithmic VXX (denoting $\log$(VXX)), and the first differences of VXX (ΔVXX) and the first differences of logarithmic VXX (denoting $\Delta \log$(VXX)) from 25 May 2010 to 28 December 2017. $\rho(n)$ is the auto-correlation at lag n.

To further investigate the characteristics of the dynamics of the logarithmic VXX return (i.e., $\Delta \log(\text{VXX})$), we plot the annualized mean and volatility of the log return, with a rolling window of 30 days, for the whole sample period in Figure 4.2. Interestingly, from Figure 4.2, we can see that the volatility of the logarithmic return is characterised by stochastical swings around the blue dash line which represents the average values of the volatility process. To be more specific, this process tends to drive back to the long-term mean level when the values are higher or lower than the mean level, so it displays the mean-reverting property.

4.1.2 VXX Options

Following Bao et al. (2012), we only consider the VXX options with time-to-maturity from one-week to one-year. To avoid getting abnormal results, we delete all options with no values in implied-volatility. To avoid an illiquid problem, we only select the

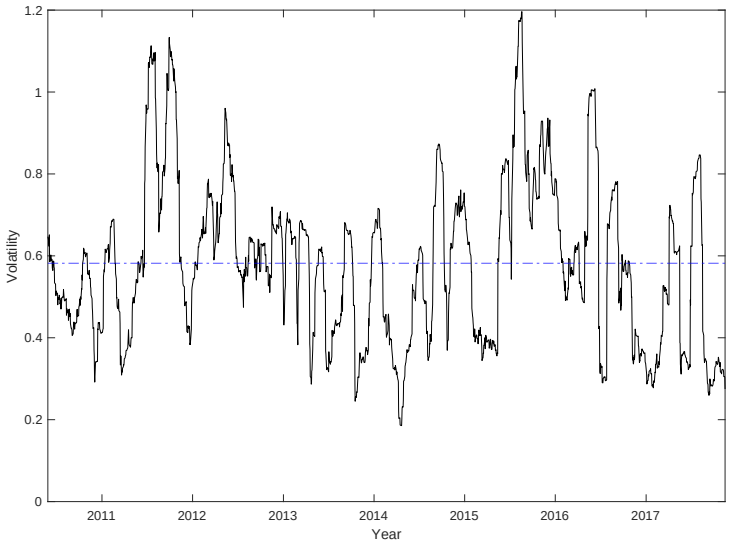


Figure 4.2: Annualized volatility of the VXX logarithmic returns, with 30-day rolling window, from 25 May 2010 to 28 December 2017.

options whose middle price[3] is greater than 0.1 and moneyness[4] is between -0.17 and 0.17 (Park, 2016; Yang & Kanniainen, 2017). We focus on analyzing more actively traded options, so we filter out the options whose daily trading volume is less than 100. Finally, we randomly select five options in each trading day [5] in the selected sample period. We note that the number of option contracts in some days is less than five after data pre-processing. In such cases, we use all the options in those days.

Table 4.3 summarizes the VXX options data by time-to-maturity and moneyness. Our sample consists of a total of 9,528 option contracts, including 4,803 and 4,725 contracts in the in-sample and out-of-sample, respectively. Panel A of Table 4.3 reveals that the number of option contracts distributes unevenly in different splits of time-to-maturity due to the random data selection. In particular, the number of option contracts with a long time-to-maturity ($\tau > 90$) is much smaller than that of option contracts with

[3]Middle price refers to the average price of the bid and ask price of each day.

[4]Based on Mencía and Sentana (2013), the log moneyness is defined as $\ln(\text{Strike}/\text{Forward price})$.

[5]We conduct several experiments. We find that using three observations (options) in each day can get desirable estimation results. We believe that five options per day should be sufficient for the model estimation. Using a larger number of options will increase the computation cost.

a relatively short time-to-maturity ($\tau \leq 90$). From Panel B, we find that the overall average price of the options fluctuates between 1.99 and 2.58 for the in-sample data, and between 1.61 and 1.90 for the out-of-sample data. In addition, the options with longer time-to-maturity tend to have a higher average price. Moreover, in Panel C, the overall average implied volatilities exhibit a volatility smirk (reverse skew) for both of the two sub-samples. A good option pricing model should be able to capture these characteristics in the VXX options data.

Table 4.3: Summary for VXX options

	In-sample				Out-of-sample			
	$\tau \leq 30$	$30-90$	$\tau > 90$	All	$\tau \leq 30$	$30-90$	$\tau > 90$	All
				Panel A: Number of option contracts				
Call	1102	1091	382	2575	1524	763	223	2510
Put	973	966	289	2228	1281	701	233	2215
All	2075	2057	671	4803	2805	1464	456	4725
				Panel B: Average option prices				
$M \leq -0.1$	1.3152	2.0462	3.2669	1.9913	0.9008	1.8344	3.2205	1.6207
$-0.1 < M \leq -0.06$	1.4532	2.3682	4.6020	2.2426	1.0913	1.8584	3.3167	1.6126
$-0.06 < M \leq -0.02$	1.4558	2.4439	4.1024	2.2129	1.1452	2.1662	3.7089	1.6376
$-0.02 < M \leq 0.02$	1.5348	2.7512	4.5093	2.4870	1.2781	2.5597	3.9418	1.9040
$0.02 < M \leq 0.06$	1.7345	2.8873	4.0880	2.5769	1.3267	2.4770	3.8078	1.8678
$0.06 < M \leq 0.1$	1.5390	2.5899	4.1256	2.3263	1.3018	2.3082	4.0291	1.8583
$M > 0.1$	1.6027	2.4902	4.5042	2.3547	1.0644	2.3225	3.7915	1.7163
				Panel D: Average implied volatilities				
$M \leq -0.1$	0.8006	0.7762	0.7647	0.7854	0.8081	0.7697	0.7412	0.7900
$-0.1 < M \leq -0.06$	0.7371	0.7399	0.7431	0.7391	0.7172	0.7165	0.7271	0.7179
$-0.06 < M \leq -0.02$	0.7087	0.7258	0.7422	0.7210	0.6622	0.6946	0.6870	0.6734
$-0.02 < M \leq 0.02$	0.6477	0.7070	0.7296	0.6849	0.6140	0.6676	0.7151	0.6392
$0.02 < M \leq 0.06$	0.6019	0.6729	0.7320	0.6487	0.5800	0.6471	0.6838	0.6076
$0.06 < M \leq 0.1$	0.6020	0.6655	0.7342	0.6457	0.5788	0.6180	0.6747	0.6033
$M > 0.1$	0.6159	0.6512	0.7100	0.6485	0.6040	0.6025	0.6756	0.6132

The VXX options data are obtained form OptionMetrics and filtered in accordance with the rules in Section 4.1.2. The second row in the table is the information about the time-to-maturity, denoted by τ. In this row, 30-90 means that τ is within the range between 30 days and 90 days, which is $30 < \tau \leq 90$. The symbol “M" in the first column refers to log moneyness which is defined in Section 4.1.2. The implied volatility is computed from the Black-Scholes formula by using the market data.

4.2 Model Specifications

In this section, we construct the models for the dynamics of the logarithmic VXX returns under Lévy processes with respect to a filtered probability space $(\Omega, \mathcal{F}, \mathbb{Q})$ with a complete filtration $(\mathcal{F}_t)_{t\geq 0}$. Here, $\mathbb{Q}$ refers to a risk-neutral probability measure. An analytical expression for the characteristic function of the logarithmic VXX returns can be derived from the given model. The VXX option pricing formulae can be converted from the characteristic functions via the inverse Fourier transform.

As stated in Section 4.1.1, the logarithmic transformation not only narrows down the range of the VXX data but also makes the distribution more symmetric. In most cases, the modeling techniques can cope with the data with a narrow range and symmetric distribution better. Therefore, we directly model the dynamics of the logarithmic VXX time series, not the original VXX. The VXX ETN is tradeable. So, we clarify that the discounted price of VXX under risk-neutral measure should remain a martingale unconditionally (e.g. Shreve, 2008, Carr & Wu, 2010). In terms of modelling the dynamics of the VIX under the risk-neutral measure, a mean-reverting process is often used, e.g., Mencía and Sentana (2013), Park (2016), Cao et al. (2019). Because the VIX is a non-tradable volatility index, the VIX process does not need to satisfy the martingale condition. For modelling the VXX, however, the literature also often uses the same mean-reverting process motivated by the VIX studies, for example Bao et al. (2012) and Tan et al. (2021). This mean-reverting setting makes the VXX fail to satisfy the martingale condition. As we know that the VXX ETN is an exchange-traded debt security, unlike the VIX (a non-tradable volatility index), the discounted price of the VXX, as a tradable asset, under the risk-neutral measure, should remain a martingale unconditionally (e.g., Carr and Wu (2004)).

Thus, under $\mathbb{Q}$, we denote the process of the logarithmic VXX by $\{x_t\}_{t\geq 0}$. Incorporating the stochastic volatility feature in the logarithmic VXX (See Figure 4.2), we

assume that the dynamics of $\{x_t\}_{t\geq 0}$ follow:[6]

$$\begin{aligned} dx_t &= \left(r_t - \Psi^J(-i) - \frac{1}{2}v_t\right)dt + \sqrt{v_t}dW_{t,1}^{\mathbb{Q}} + dJ_t, \\ dv_t &= \kappa_v\left(\theta_v - v_t\right)dt + \sigma_v\sqrt{v_t}dW_{t,2}^{\mathbb{Q}}, \end{aligned} \tag{4.1}$$

where $W_{t,1}^{\mathbb{Q}}$ and $W_{t,2}^{\mathbb{Q}}$ are Wiener processes with correlation ρ; κ_v and σ_v captures the reverting speeds and the volatility level of v_t, respectively; θ_v represents the long-term average for v_t; r_t is a risk-free interest rate and J_t refers to a jump process. In this chapter, we select four different jump processes (MJ, NIG, TS, and GTS) from Section 2.1.7 based on the structure of the jumps. Our selections include a finite-activity jump process (MJ) and three infinite-activity jump processes with finite or infinite variation (NIG, TS and GTS). Furthermore, the jump compensator, $\Psi^J(-i)$, is calculated from the characteristic exponent of the jump process, referring to Table 2.1. In order to investigate the role of the jumps in pricing VXX options, we also consider the models without jumps. Under the no jump assumption, we set the terms dJ_t and $\Psi^J(-i)$ to be zero. Thereafter, we call the model without jumps as the SV model, and the models with the different types of jumps as the SV-MJ, SV-NIG, SV-TS, and SV-GTS models, respectively.

4.3 Option Pricing Approach

Denote the prices of the VXX call and put options at time t with the strike price K and the maturity T as $c(t; x_t, v_t, T, K, \Theta)$ and $p(t; x_t, v_t, T, K, \Theta)$ respectively, where Θ is the vector of the model parameters. We set time-to-maturity as $\tau = T - t$. Based on the information available upon time t, the theoretical European call option price with an

[6]The relationship between the VIX and the VXX has been discussed in Appendix B.1. Our model (4.1) is consistent with (B.8).

instantaneous interest rate r_t under risk-neutral measure $\mathbb{Q}$ is defined by

$$c(t; x_t, v_t, T, K, \Theta) = \exp\{-r_t\tau\}\mathbb{E}^{\mathbb{Q}}\left[\max(\exp(x_t) - K, 0)|\mathcal{F}_t\right]. \tag{4.2}$$

In the same manner, the European put option price is defined by

$$p(t; x_t, v_t, T, K, \Theta) = \exp\{-r_t\tau\}\mathbb{E}^{\mathbb{Q}}\left[\max(K - \exp(x_t), 0)|\mathcal{F}_t\right]. \tag{4.3}$$

In analogy to Heston (1993), the European call and put option price formulae can be derived from the characteristic function $f(u; t, T, x_t, v_t, \Theta)$ of the logarithmic VXX return. The option price formulae for the specified models can be expressed as,

$$c(t; x_t, v_t, T, K, \Theta) = \exp\{-r_t\tau\}\left[f(-i)\Phi_1(t) - K\Phi_2(t)\right], \tag{4.4}$$

$$p(t; x_t, v_t, T, K, \Theta) = \exp\{-r_t\tau\}\left[K(1 - \Phi_2(t)) - f(-i)(1 - \Phi_1(t))\right], \tag{4.5}$$

where $f(u)$ is the short-form of $f(u; t, T, x_t, v_t, \Theta)$, and $\Phi_1(t)$ and $\Phi_2(t)$ are the short-form of $\Phi_1(t; T, x_t, v_t, K, \Theta)$ and $\Phi_2(t; T, x_t, v_t, K, \Theta)$ which can be interpreted as the probabilities of $x_t > \ln K$ under the two different risk-neutral measures, respectively. [7] The formulae for $\Phi_1(t)$ and $\Phi_2(t)$ are given as follows:

$$\Phi_1(t) = \frac{1}{2} + \frac{1}{\pi}\int_0^\infty \mathbf{Re}\left[\frac{\exp\{-iu\ln K\}f(u-i)}{iuf(-i)}\right]du, \tag{4.6}$$

$$\Phi_2(t) = \frac{1}{2} + \frac{1}{\pi}\int_0^\infty \mathbf{Re}\left[\frac{\exp\{-iu\ln K\}f(u)}{iu}\right]du. \tag{4.7}$$

It is difficult to derive the exact expressions for the integrals in Eq. (4.6) and (4.7), so we apply the Gauss–Laguerre quadrature method to evaluate these integrals. In this numerical evaluation, the higher the degree of Laguerre polynomial is used, the more

[7] In line with Appendix A in Park (2016), a change of measure from $\mathbb{Q}$ to $\mathbb{Q}_1$ is based on a Radon-Nikodym derivative $\frac{d\mathbb{Q}_1}{d\mathbb{Q}} = \frac{\exp(v_T)}{\mathbb{E}_t^{\mathbb{Q}}[\exp(v_T)]}$.

accurate is the result gained, but a longer time is taken for the calculation. To balance the computational cost and the results accuracy, we choose the degree of the Laguerre polynomial as 10.

As in Cont and Tankov (2004), the spot characteristic function of the logarithmic VXX returns, at time t, under $\mathbb{Q}$ is defined by

$$f(u;t,T,x_t,v_t,\Theta) = \mathbb{E}_t^{\mathbb{Q}}[\exp(iux_T)|\mathcal{F}_t]. \tag{4.8}$$

Under the Model 4.1, we assume Eq. (4.8) has an affine form, that is,

$$f(u;t,T,x_t,v_t,\Theta) = \exp\left[iu\tau(r_t - \Psi^J(-i)) - \psi_0(\tau) - \psi_v(\tau)v_t + iux_t\right]. \tag{4.9}$$

Then, we assume that the jump components are independent of the continuous parts in the model, so the characteristic function can be derived in two steps. First, we calculate the characteristic function of the continuous part of x_t, which we denote by x_t^c. Second, we multiply the characteristic function of x_t^c by the characteristic function of jump process described in Section 2.1.7. Thus, the characteristic function for our models can be expressed as

$$f(u;t,T,x_t,v_t,\Theta) = \exp\left[iu\tau(r_t - \Psi^J(-i)) - \psi_0(\tau) - \psi_v(\tau)v_t + iux_t + \tau\Psi^J(u)\right], \tag{4.10}$$

where $\Psi^J(u)$ is specified in Table 2.1, $\tau = T - t$ and $\psi_0(\tau)$ and $\psi_v(\tau)$ are given by

$$\psi_v(\tau) = \frac{u^2 + iu}{\xi\coth\left(\frac{\tau\xi}{2}\right) + b}, \tag{4.11}$$

$$\psi_0(\tau) = -\frac{\kappa_v\theta_v b\tau}{\sigma_v^2} + \frac{2\kappa_v\theta_v}{\sigma_v^2}\ln\left[\cosh\left(\frac{\tau\xi}{2}\right) + \frac{b}{\xi}\sinh\left(\frac{\tau\xi}{2}\right)\right], \tag{4.12}$$

where $b = \kappa_v - iu\rho\sigma_v$ and $\xi = \sqrt{b^2 + \sigma_v^2(u^2 + iu)}$. In the SV model, both $\Psi_J(-i)$

and $\Psi_J(u)$ are set to be zero. We use the time changed method for the derivation of $f(u;t,T,x_t,u_t,\Theta)$, see Appendix B.3 for more details.

4.4 Model Calibration

We use both the VXX option price data and the VXX data to calibrate our models. The prices of VXX options include the premium of early exercise, since the VXX options are American-style. Thus, by following Carr and Wu (2010), we convert the option prices into European-style options prices by using quoted implied volatility via the Black and Scholes (1973) option pricing model.

4.4.1 Calibration Procedure

In the calibration procedure, we apply the MLE to estimate the values of parameters in the models. For the models which contain latent processes, we use the UKF for estimating the values of latent variables to accelerate the speed of calibration. In order to search the maximum value of the total log-likelihood function, we employ the Nelder-Mead simplex algorithm which is a derivative-free method. We use this algorithm because it can be applied in the non-linear objective function which can be either continuous or discontinuous (Wouk, 1987; Barati, 2011). According to Aït-Sahalia and Kimmel (2007) and Du and Luo (2019), the log-likelihood function at time $t+1$ conditional on time t is defined as,

$$l_{t+1|t}(\Theta) = -\frac{1}{2}\left\{n_t \ln(2\pi) + \ln|\Sigma^O_{t+1|t}| + \left(O_{t+1} - \hat{O}_{t+1}\right)^\top \left(\Sigma^O_{t+1|t}\right)^{-1} \left(O_{t+1} - \hat{O}_{t+1}\right)\right\} \tag{4.13}$$

where n_t denotes the total number of options contracts during that day, O_{t+1} is the vector of actual option prices in day $t+1$, $\hat{O}_{t+1}$ is the vector of estimated option prices in the same day, and $\Sigma^O_{t+1|t}$ refers to the covariance of option prices which is conditional on

time t. The total log-likelihood function is the sum of the daily log-likelihood function (4.13), that is,

$$L(\Theta) = \sum_{t=1}^{T} l_{t+1|t}(\Theta), \tag{4.14}$$

where T is the total sample period.

In order to apply the UKF, it is necessary to convert the latent process in our proposed models from $\mathbb{Q}$ to a physical measure $\mathbb{P}$ by introducing a risk premium, η_v (Wehn et al., 2013; Park, 2016). The latent process $\{v_t\}_{t\geq 0}$ under $\mathbb{Q}$ is given in Section 4.2. After converting to $\mathbb{P}$, the dynamics of the processes are expressed as

$$dv_t = \kappa_v \left(\theta_v - v_t\right) dt + \eta_v v_t dt + \sigma_v \sqrt{v_t} dW_{t,2}^{\mathbb{P}}, \tag{4.15}$$

where $W_{t,2}^{\mathbb{P}}$ is a Wiener process under $\mathbb{P}$. Next, we need to discretize Eq. (4.15) by using the Euler approximation. The discretized version of Eq. (4.15) can be expressed as

$$\Delta v_t = \kappa_v \left(\theta_v - v_t\right) \Delta t + \eta_v v_t \Delta t + \sigma_v \sqrt{v_t \Delta t} \epsilon_{t,2}, \tag{4.16}$$

where $\epsilon_{t,2}$ is a standard normal random variable.

Suppose there are N call option contracts and M put option contracts at the time t. According to Yang and Kanniainen (2017), the measurement equations are defined by

$$\begin{pmatrix} c_1(t) \\ \vdots \\ c_N(t) \\ p_1(t) \\ \vdots \\ p_M(t) \end{pmatrix} = \begin{pmatrix} \hat{c}_1(t) \\ \vdots \\ \hat{c}_N(t) \\ \hat{p}_1(t) \\ \vdots \\ \hat{p}_M(t) \end{pmatrix} + \sigma_0 \begin{pmatrix} \mathcal{V}_{c_1}(t) e_{c_1}(t) \\ \vdots \\ \mathcal{V}_{c_N}(t) e_{c_N}(t) \\ \mathcal{V}_{p_1}(t) e_{p_1}(t) \\ \vdots \\ \mathcal{V}_{p_M}(t) e_{p_M}(t) \end{pmatrix}, \tag{4.17}$$

where σ_0 measures the size of pricing error, $c_i(t)$, for $i = 1, \ldots, N$, and $p_j(t)$, for

$j = 1, \ldots, M$, are the market prices of call and put options respectively, the notations $\hat{c}$ and $\hat{p}$ refer to the option prices which are estimated by Formula (4.4) or (4.5), and $\mathcal{V}_k(t)$, for $k = c_1, \ldots, c_N, p_1, \ldots, p_M$, is their corresponding market vega. Moreover, $e_{c_1}, \ldots, e_{p_M}$ are the independent normal variables with zero mean and unit variance.

4.4.2 Model Estimation Result Analysis

Table 4.3 reports the parameters which are estimated by using the in-sample data across the proposed models and the standard errors which are computed from the estimation results. First, we analyze the variation among the common parameters estimates under the no jump specification and different jump structures. According to this table, after considering the jump processes in the model, the mean reverting speed of the variance process decelerates as the estimated value for κ_v decreases from 9.68 to under 8.00. The estimates of κ_v in the jump models vary from 6.69 to 7.45. Generally, the infinite-activity jump models with infinite variation sample paths come with the smaller κ_v values. The table also shows that the estimates for the long-term mean (θ_v) and volatility (σ_v) of $\{v_t\}_{t\geq 0}$ decrease slightly after adding the jump processes into the model. For example, θ_v and σ_v are 0.55 and 3.19 in the SV model, respectively, while they decrease to 0.46 and 2.76 in the SV-MJ model, respectively. In addition, the jump influences the value of the leverage factor parameter, ρ. The correlation between the VXX ETN return and its variance (i.e., ρ) becomes more positive in the SV model than those in the jump diffusion models. The positive correlation suggests the positive volatility skew in the VXX option market, which aligns with the observation in Bao et al. (2012).

Second, the estimated results for the jump processes reveal several features in the dynamics of VXX ETN returns. The key difference between MJ jump and other selected jumps is the jump frequency. MJ jump is a finite-activity jump and its frequency is measured by λ_J. The estimate for λ_J is 0.40 which indicates the big jump (a rare event)

occurs around every two and half year. Interestingly, we find a roughly similar big jump frequency in the VXX ETN returns and S&P 500 index returns processes, refering to Huang and Wu (2004). On the contrary, the NIG, TS and GTS jump structures in the SV-NIG, SV-TS and SV-GTS models can generate infinitely many jumps during a finite time period. Here, we think the GTS jump process in the SV-GTS model is an infinity-activity jump process with infinity variation because α_+ is estimated at 1.48 which is greater than 1.00. In addition, the estimates of jump parameters in the infinite-activity jump models imply an asymmetry tail behaviour in the jump process.

We also investigate whether the diffusion part is important when an infinite-activity jump component is included in the models. Some literature points out that it may be not essential to include a diffusion component if a pure infinite-activity jump process is adopted (Carr et al., 2002; Carr & Wu, 2003). Based on the analysis in Section 4.5.3, we can observe that the estimated latent process (variance process) substantially fluctuates above zero. Hence, we believe that both the infinite-activity jump and the diffusion component are essential in this case.

4.5 Model Performance Analysis

In this section, we compare and analyze the performance of the five proposed models. We use the average root mean square error (RMSE) as our key measurement because the RMSE accommodates the higher magnitude error with relatively higher weight. As in Kanniainen et al. (2014) and Park (2016), the VXX options pricing error needs to be scaled by the corresponding market vega ($\mathcal{V}$). Then, the RMSE is defined by

$$\text{RMSE} = \sqrt{\frac{1}{M}\sum_{t=1}^{M}\left[\frac{1}{N_t}\sum_{j=1}^{N_t}\left(\frac{O(t;T_{t_j},K_{t_j}) - \hat{O}(t;T_{t_j},K_{t_j})}{\mathcal{V}(t;T_{t_j},K_{t_j})}\right)^2\right]}, \tag{4.18}$$

Table 4.3: Parameter estimates

	SV	SV-MJ	SV-NIG	SV-TS	SV-GTS
			Common parameters		
κ_v	9.6834	7.3750	6.8753	7.4452	6.6937
	(0.2590)	(0.2833)	(0.2166)	(0.2224)	(0.2232)
θ_v	0.5504	0.4599	0.4388	0.4549	0.4202
	(0.0030)	(0.0083)	(0.0078)	(0.0075)	(0.0058)
σ_v	3.1909	2.7576	2.9801	2.8808	3.0213
	(0.0410)	(0.0699)	(0.0744)	(0.0729)	(0.0704)
η_v	0.1756	-0.0644	0.3058	0.0608	0.1746
	(0.3124)	(0.5871)	(0.2777)	(0.2683)	(0.1341)
ρ	0.7890	0.6627	0.7265	0.6862	0.7387
	(0.0007)	(0.0184)	(0.0224)	(0.0164)	(0.0215)
σ_o	0.0384	0.0372	0.0356	0.0357	0.0354
	(0.0005)	(0.0004)	(0.0004)	(0.0004)	(0.0004)
			Jump parameters		
λ_J		0.3959			
		(0.0536)			
λ_+				4.3320	1.5290
				(0.2654)	(0.3811)
λ_-				0.6893	4.8272
				(0.0942)	(0.5322)
θ_J		0.4027	0.5193		
		(0.0312)	(0.0010)		
σ_J		0.0872	0.2524		
		(0.0598)	(0.0157)		
δ_J			0.1749		
			(0.0158)		
c_+				1.3544	0.0875
				(0.1902)	(0.0204)
c_-				0.0175	0.2503
				(0.0027)	(0.0907)
α_+					1.4802
					(0.0543)
α_-					-1.8082
					(0.3498)

The parameters are estimated by maximizing the value of the log-likelihood function and the UKF for all the models. The data used include the VXX index and VXX option prices between the period from 25 May 2010 and 27 March 2014. The values in parentheses show the standard errors which are obtained by the inverse of the Hessian matrix at the optimal situation.

where $O(t;T_{t_j},K_{t_j})$ represents the market option price, $\hat{O}(t_{t_j};T_{t_j},K_j)$ represents the estimated price from our models, M denotes the total number of days in the selected sample, and N_t represents the total number of option contracts at time t.

We also examine the statistical significance of the models' performance difference by using the t-test which is applied in Huang and Wu (2004). This test is calculated through the daily average squared pricing error. The test statistic $\hat{S}$ is formulated as follows:

$$\hat{S} = \frac{\overline{\mathrm{MSE}}^{i} - \overline{\mathrm{MSE}}^{j}}{\mathrm{SE}(\overline{\mathrm{MSE}}^{i} - \overline{\mathrm{MSE}}^{j})}, \tag{4.19}$$

where superscripts i and j represent the model i and the model j respectively, and $\overline{\mathrm{MSE}}$ denotes the sample mean of squared errors. Similar to the notations in Eq. (4.18), the daily average squared pricing error, $\overline{\mathrm{MSE}_{\mathrm{t}}}$, can be computed as

$$\overline{\mathrm{MSE}_{\mathrm{t}}} = \frac{1}{N_t}\sum_{j=1}^{N_t}\left(\frac{O(t;T_{t_j},K_{t_j}) - \hat{O}(t;T_{t_j},K_{t_j})}{\mathcal{V}(t;T_{t_j},K_{t_j})}\right)^2.$$

4.5.1 Performance Analysis on VXX Option Pricing

Table 4.4 compares the in-sample and out-of-sample performances of the models with different jump structures in terms of the RMSE. In this section, we investigate the performance in all pricing options but also the performance in pricing call and put options, respectively. Overall, the RMSEs for both of the in-sample and out-of-sample periods are improved by at least 2.931% after considering jumps into the model. When the jump structure is changed from finite-activity to infinite-activity, the in-sample's and out-of-sample's RMSEs can be further improved. Moreover, the models with jumps perform better in pricing put than in pricing call options.

We conduct a further performance comparison among our jump models by applying the t-test (Eq. (4.19)) on the pairwise models since the performance difference between the two jump models seems to be minor in Table 4.4. The t-test statistics are reported

Table 4.4: Root mean square error (RMSE)

	In-sample RMSEs			Out-of-sample RMSEs		
	Overall	Call	Put	Overall	Call	Put
SV	3.5218%	3.5145%	3.5303%	4.4503%	4.5633%	4.3186%
SV-MJ	3.4378%	3.4715%	3.3983%	4.3170%	4.3767%	4.2483%
SV-TS	3.2934%	3.3354%	3.2442%	4.2364%	4.2738%	4.1937%
SV-NIG	3.2875%	3.3307%	3.2368%	4.2538%	4.2901%	4.2121%
SV-GTS	3.2742%	3.3088%	3.2337%	4.2516%	4.2678%	4.2331%

The RMSE is computed in accordance with Eq. (4.18) by using the estimated results in Table 4.3 in Section 4.4.2. The in-sample RMSE and out-of-sample RMSE are based on the period from 25 May 2010 to 27 March 2014 and the period from 28 March 2014 to 28 December 2017, respectively.

in Table 4.5. According to Table 4.5, we observe that all test statistics are greater than the t-statistic critical value 1.6450 in Panel A, thus the infinite-activity jump models perform significantly better in pricing VXX options than the finite-activity jump model does. However, the performance difference between two infinite-activity jump models is not statistical significance for the in-sample period. During the out-of-sample period, we notice that the performance of the SV-TS model is significantly better than the SV-NIG model because the t statistics under MJ-NIG column in Panel B is -1.9305 which is less than -1.6450. In addition, the positive values for the t statistics in Panel C only indicate that the SV-GTS model slightly outperforms the SV-NIG model.

4.5.2 Fitting to the Term Structures of Options and Implied Volatility

In this sub-section, we investigate how our models perform under different degrees of moneyness and different ranges of maturities because model performance may vary under different conditions. Table 4.6 reports the RMSEs of our models by the moneyness and time to maturity based on the whole sample data. Overall, the full-sample RMSEs yielded by the infinite-activity jump models (SV-TS, SV-NIG and SV-GTS) are smaller than those yielded by the finite-activity jump model (SV-MJ) and the model without

Table 4.5: Pairwise model performance comparison

	In-sample			Out-of-sample		
	Panel A Finite-activity jumps vs. infinite-activity jumps					
	MJ-TS	MJ-NIG	MJ-GTS	MJ-TS	MJ-NIG	MJ-GTS
SV	5.9484	6.1165	6.5942	4.0848	3.0003	3.0004
	Panel B Finite variation vs. infinite variation					
	-	TS-NIG	TS-GTS	-	TS-NIG	TS-GTS
SV	-	0.0929	0.1155	-	-1.9305	-1.2909
	Panel C Comparing infinite-variation jumps with a infinite variation					
	-	-	NIG-GTS	-	-	NIG-GTS
SV	-	-	0.8542	-	-	0.3495

This table reports the test statistic of a t-test ($\hat{S}$), which is given by Eq. (4.19). The tests in different panels compare the performance of different jump structures in pricing VXX options. In this table, we compare five different types of jump structures, including finite-activity jumps (MJ), infinite-activity jumps with finite variation (TS), infinite-activity jumps with infinite-variation (NIG and GTS), each of which is under the given drift and volatility specifications. The first column lists the drift and volatility specification of the model. The first row in each panel gives information about the jump structure. For example, MJ-TS refers to the comparison between the compound Poisson jump process and the tempered stable jump process, and so on in a similar fashion. In the test for the model i and the model j, if the value for the test statistic $\hat{S}$ is greater than 1.6450,then the model j is superior to the model i at a 95% confidence interval, and vice versa. The in-sample covers the period from 25 May 2010 to 27 March 2014 and the out-of-sample covers the period from 28 March 2014 to 28 December 2017.

jumps (SV) across all degrees of moneyness and all ranges of maturities, so the infinite-activity jump models can capture more features in the term structures of the VXX options than the finite-activity jump model does. Moreover, the stochastic volatility models (including all proposed models) generally fit the market data better if the time to maturity of an option is between 30 and 90 calendar days.

In order to examine the ability of capturing the features in the implied volatility for different types of models, we calculate the average in- and out-of-sample model implied volatilities for different ranges of moneyness and time-to-maturity based on the Black-Scholes model. Figure 4.3 exhibits the comparison between the market implied volatilities and the model implied volatilities. According to Figure 4.3, the implied volatilities of our five models can fit the implied volatility smirk quite well across all the moneyness when the time-to-maturity is less than or equal to 90 days. However, the infinite-activity jump models can take more advantages than the model with a

Table 4.6: RMSEs of option prices by moneyness and maturity

Moneyness	SV	SV-MJ	SV-TS	SV-NIG	SV-GTS
		Panel A: $\tau \leq 30$			
$M \leq -0.1$	0.0568	0.0536	0.0538	0.0536	0.0529
$-0.1 < M \leq -0.06$	0.0402	0.0420	0.0419	0.0415	0.0401
$-0.06 < M \leq -0.02$	0.0350	0.0362	0.0345	0.0346	0.0345
$-0.02 < M \leq 0.02$	0.0331	0.0314	0.0313	0.0315	0.0314
$0.02 < M \leq 0.06$	0.0394	0.0380	0.0370	0.0375	0.0379
$0.06 < M \leq 0.1$	0.0392	0.0394	0.0380	0.0385	0.0390
$M > 0.1$	0.0471	0.0475	0.0463	0.0459	0.0462
		Panel B: $30 < \tau \leq 90$			
$M \leq -0.1$	0.0324	0.0319	0.0306	0.0309	0.0309
$-0.1 < M \leq -0.06$	0.0299	0.0294	0.0287	0.0289	0.0290
$-0.06 < M \leq -0.02$	0.0326	0.0319	0.0306	0.0306	0.0307
$-0.02 < M \leq 0.02$	0.0323	0.0310	0.0295	0.0298	0.0299
$0.02 < M \leq 0.06$	0.0326	0.0317	0.0298	0.0299	0.0300
$0.06 < M \leq 0.1$	0.0345	0.0328	0.0318	0.0326	0.0328
$M > 0.1$	0.0395	0.0375	0.0355	0.0362	0.0369
		Panel C: $\tau > 90$			
$M \leq -0.1$	0.0507	0.0462	0.0460	0.0461	0.0467
$-0.1 < M \leq -0.06$	0.0522	0.0495	0.0454	0.0456	0.0459
$-0.06 < M \leq -0.02$	0.0501	0.0461	0.0440	0.0432	0.0432
$-0.02 < M \leq 0.02$	0.0490	0.0468	0.0431	0.0428	0.0428
$0.02 < M \leq 0.06$	0.0548	0.0520	0.0485	0.0476	0.0473
$0.06 < M \leq 0.1$	0.0515	0.0492	0.0446	0.0440	0.0437
$M > 0.1$	0.0471	0.0465	0.0462	0.0460	0.0459

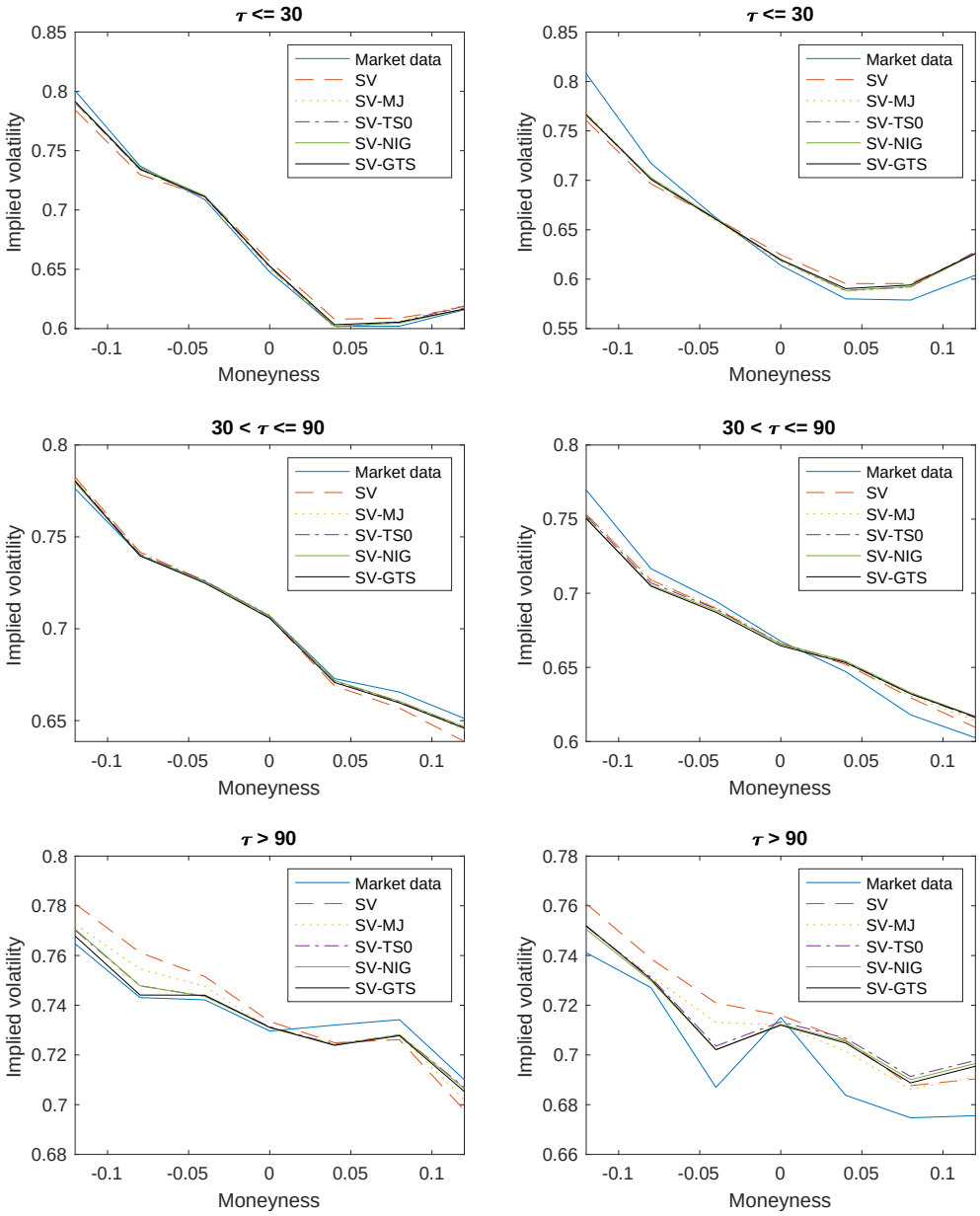
The RMSEs are computed in accordance with Eq. (4.18) by using the estimated results in Table 4.3 in Section 4.4.2. The full-sample data are used to calculate the RMSEs, including the period from 25 May 2010 to 28 December 2017. The moneyness refers to log moneyness which is defined in Section 4.1.2. The symbol τ denotes the time-to-maturity.

finite-activity jump or without jumps on capturing the dynamics of implied volatility for the options with the long-term time-to-maturity ($\tau > 90$).

4.5.3 Evolution of the Instantaneous Variance

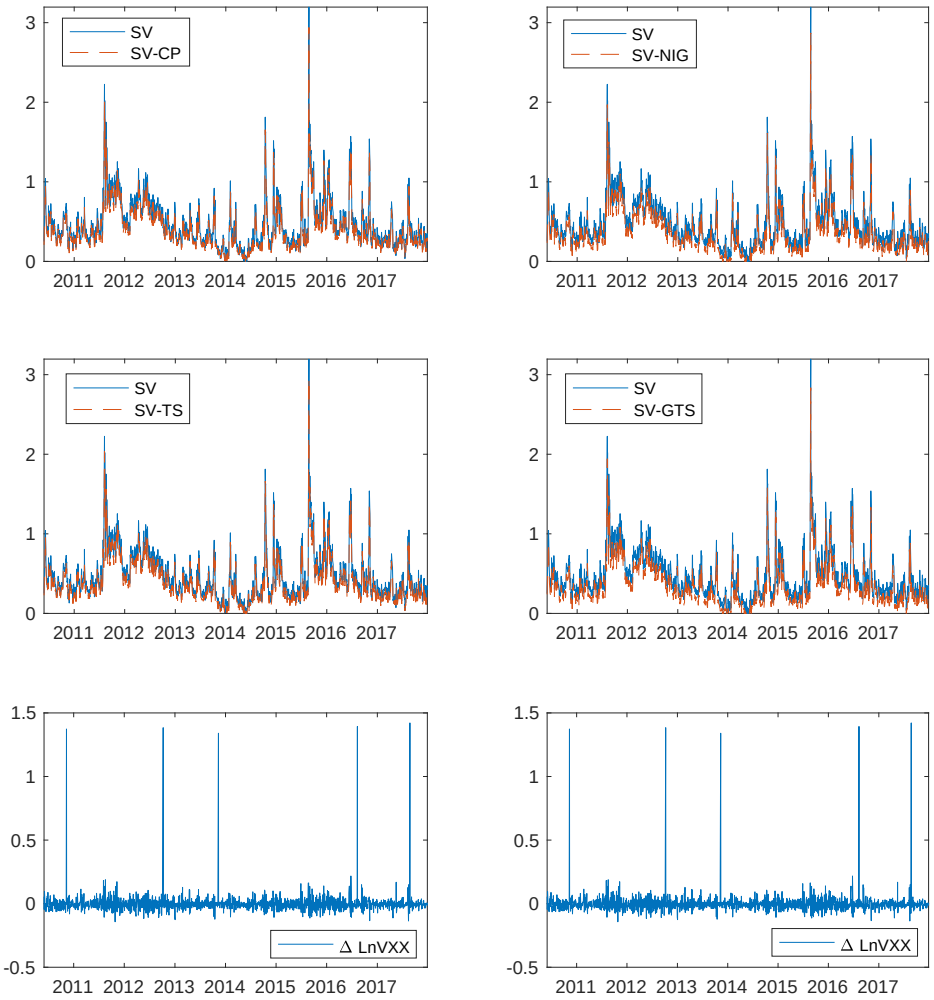
Figure 4.4 compares the historical evaluations of the instantaneous variance states which are estimated by using the in-sample data based on the different models. From the first four sub-figures, we notice that the patterns of the variance series for the SV, SV-MJ, SV-TS, SV-NIG and SV-GTS models are almost identical during the period between 25 May 2010 and 28 December 2017. However, the variance of the SV model is substantially larger than those of the jump models. Furthermore, these figures also indicate that the our stochastic volatility models are sensitive to the changes in the log VXX. In other words, the instantaneous variance becomes large when the one-day-ahead absolute difference of the log VXX is large, and vice versa. In addition, we can observe five significant changes in the last panel of Figure 4.4, but the estimated variance states did not increase significantly in order to capture suddenly big changes. This is because these five changes are corresponding to the five reverse-splits on the VXX, but not because of big events occurring in the market.

Figure 4.3: Fit to the implied volatility



Note: These curves in the figures display the average implied volatility for the different range of moneyness and time-to-maturity (τ). The left and right columns show the average implied volatilities for in-sample and out-of-sample respectively. The model implied volatility is calculated from the Black-Scholes model by using the model estimated price. The moneyness is given in Section 4.1.2. The in-sample covers the period from 25 May 2010 to 27 March 2014 and the out-of-sample covers the period from 28 March 2014 to 28 December 2017.

Figure 4.4: Historical evolution of the v_t from 25 May 2010 to 28 December 2017



Notes: The instantaneous variances (v_t) of different models are computed by using the estimated parameter values in Table 4.3. The sub-figures in the last row show the evolution of the one-day-ahead difference of the log VXX during the same period.

Chapter 5

Pricing S&P 500 Variance Futures

This chapter is joint work with J. Cao, X. Ruan and W. Zhang. In this chapter, we incorporate jumps not only in the process S&P 500 index returns but also in the variance process of our two-factor model. In addition, we allow the correlation between the two jump processes and the time-varying mean of the variance process. The model is described and explained in Section 5.2 and the subsequent section is about the derivation of a formula for the VIX squared and integrated variance premium. In Section 5.4, we derive a S&P 500 variance futures pricing formula based on the results of the previous section. We then calibrate the models and analyze the calibration results, in Section 5.5. At the end of this chapter, we conduct a comparison among performance of the models. The proofs of the formulae are given in Appendix C.

5.1 Data

We select the S&P 500 variance future contracts which expire between 17 January 2014 and 18 December 2020. The trading period of the selected contracts covers about seven and a half years from 24 June 2013 to 14 February 2020. Our sample consists of 83 variance future contracts in total, which is downloaded from the CBOE website. We

filter out the entries that the number of days between issue date and trade date is less than 90 days. It is very difficult to accurately estimate annualized realized variance of the S&P 500 index returns if the observed periods are not long enough. After filtering, 42 variance futures contracts remain in our sample. A descriptive statistics summary of the all selected contracts is reported in Table 5.1.

According to Table 5.1, the total number of business days between issue date and final settlement date fluctuates from one variance contract to another. The range is from 105 to 755 days. The contracts that are settled in December of each year have a trading period longer than those of contracts settled in the other periods of the year. The daily average settle price varies between 113.77 and 441.92. Generally, the contracts with a longer trading period have a higher daily average settle price, and vice versa. We also find that the average settle price of the contracts matured in 2016 reaches to a highest point at about 403.53.

Table 5.2 reports the descriptive statistics of the S&P 500 index, the VIX, the VIX squared and the settle price of S&P 500 variance futures in terms of index level in Panel A and in terms of the log levels in Panel B. In this table, only the logarithm of S&P 500 index displays negative skewness. In addition, the skewnesses of both of the index and the log levels are not close to zero. Moreover, except for the S&P 500 index, the kurtosises of others are greater than 3. The kurtosises of the VIX and the VIX squared are higher than those of S&P 500 index and S&P 500 variance futures prices. Therefore, we can conclude that distributions for the values of S&P 500 index, the VIX, the VIX squared and the variance futures are asymmetry.

Then, we investigate the relationship between the variance futures prices and the VIX squared across different years. In order to avoid cluttering in the figure, we select six variance futures contracts whose trading periods are longer than 600 business days. Figure 5.1 indicates that, overall, the S&P 500 variance futures prices of the selected six contracts (Z15,Z16, Z17, Z18, Z19, and Z20) increase as the value of the VIX

Table 5.1: Descriptive statistics of S&P 500 variance futures contracts

Code	No. of Obs.	No. of Days	Settle Price		Code	No. of Obs.	No. of Days	Settle Price	
			Mean	Std.				Mean	Std.
F14	54	144	188.9080	44.3120	H18	160	250	226.5259	103.8817
H14	97	187	235.8399	50.3069	M18	409	499	305.8135	87.4173
M14	250	369	289.3102	82.4285	U18	165	255	276.4878	109.9285
U14	160	250	231.2168	67.7068	Z18	666	757	370.5588	100.7465
X14	87	339	228.5410	85.4947	F19	244	335	357.8408	139.5112
Z14	377	496	319.3339	84.3104	H19	217	308	336.1308	117.5295
F15	14	105	297.5783	146.1982	M19	413	504	334.2610	84.4687
H15	159	250	313.5196	78.8818	N19	15	105	113.7740	59.3380
M15	120	211	293.6429	102.3710	Q19	35	125	228.5357	115.6702
U15	159	250	348.1140	181.1643	U19	232	323	329.7007	98.2931
Z15	537	628	381.2067	87.3538	V19	36	126	238.6673	79.0974
H16	160	250	441.9242	134.3280	X19	36	126	183.8255	100.6622
M16	410	501	417.7612	115.7661	Z19	664	755	335.7086	68.2689
U16	160	250	332.8204	145.5903	F20	15	105	122.0843	40.2887
Z16	661	752	421.6248	100.9430	G20	35	128	194.3406	51.9767
F17	246	336	348.9974	129.9748	H20	203	316	307.8875	57.2207
H17	160	250	265.2811	91.5074	M20	327	504	354.2449	59.8515
M17	411	501	377.3373	140.9098	U20	169	501	345.5185	42.3287
U17	160	250	192.1487	64.7934	V20	7	266	287.5786	6.7591
Z17	662	752	389.3416	148.0096	X20	7	291	304.5368	7.0007
F18	251	341	211.6624	69.4720	Z20	451	755	376.8651	42.1528

This table reports the descriptive statistics of the selected 42 variance futures contracts. Each code represents one variance futures contract and provides the information about the final settlement date of the contract. The letter refers to the month and the number refers to the year. For example, the code F14 means that the final settlement date of this contract is on the third Friday of January 2014. For the detailed information about the code of the month, see http://www.cboe.com/products/futures/trade-cfe/quote-vendor-symbols/va-s-amp-p-500-variance-futures. No. of obs. refers to the number of selected trading days for this contract. The no. of days counts the total number of days between the issue date and maturity date. Only the number of the business days is counted into the no. of obs. and the no. of days. All of the data sources are from the CBOE website https://markets.cboe.com/us/futures/market-statistics/historical-data/.

Table 5.2: Descriptive statistics of the index levels and log levels

	S&P 500	VIX	VIX^2	VF	S&P 500	VIX	VIX^2
	Panel A: Index levels				Panel B: log levels		
Mean	2330.7632	14.9101	237.3767	339.4294	0.0005	-0.0002	-0.0004
Median	2175.4399	13.9500	194.6025	333.9756	0.0006	-0.0054	-0.0109
Std.	423.3635	3.8826	143.7013	117.5299	0.0081	0.0791	0.1582
Skewness	0.3836	1.6664	3.0137	0.5671	-0.5149	1.2116	1.2116
Kurtosis	2.0283	7.2992	18.2327	5.4451	6.6986	10.8514	10.8514
Min	1573.0900	9.1400	83.5396	15.4056	-0.0418	-0.2998	-0.5997
Max	3379.4500	40.7400	1659.7478	1560.2500	0.0484	0.7682	1.5365

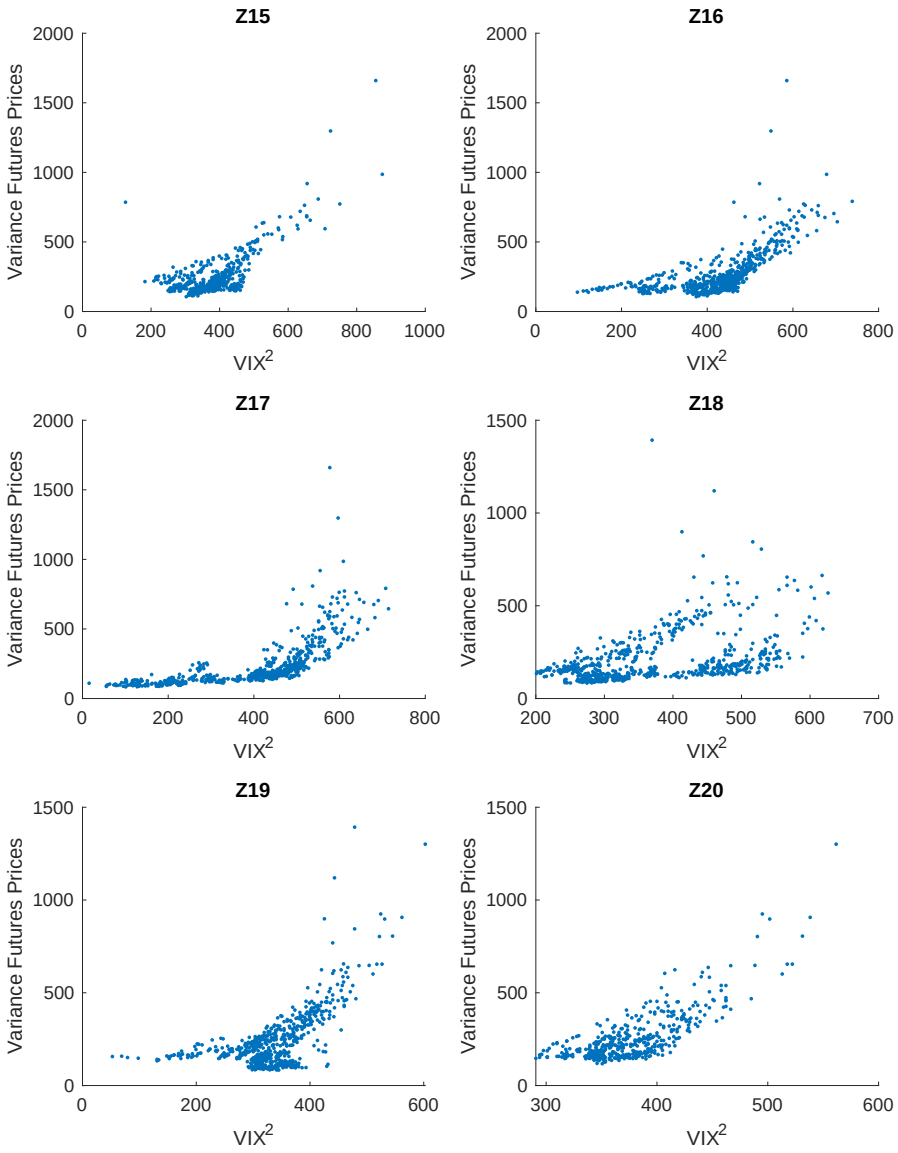
This table shows the descriptive statistics of the index level and the logarithm of the S&P 500 index, the VIX, the VIX squared and variance futures prices. Std. is the abbreviation of the standard deviation. The data are downloaded from the CBOE website: https://markets.cboe.com/us/futures/market-statistics/historical-data and the Yahoo Fiance website: https://nz.finance.yahoo.com.

squared goes up. We also notice that there seems to be a non-linear relationship between them. Next, we conduct a linear regression analysis to investigate further. It is well known that R-squared and root mean squared error are the important measures of the goodness of fit in this analysis. The values of R-squared for the most variance contracts range from 60% to 19% and most of them are less than 50%. Combining with the considerably large root mean squared errors and low value for R-squared, we believe that the relationship between them are not linear.

5.2 Model Specifications

We price the S&P 500 variance futures based on the instantaneous variance of the S&P 500 index. The instantaneous variance is related to the squared VIX. Thus, we need to build a model for the dynamics of the S&P 500 index first. Then, we can derive a formula for the instantaneous variance from the model. The prominent features for the dynamics of the S&P 500 index return have been discussed in the subsequent studies. First, Bakshi et al. (1997) found the empirical evidence that the volatility of the index returns changes stochastically and jumps occur in the index return process. Second,

Figure 5.1: Variance futures prices vs. the VIX squared



Notes: Every sub-figure shows the relationship between the prices of a particular variance futures contract and the VIX^2. The subtitles, Z15, Z16, Z17, Z18, Z19, and Z20, represent the codes of the contracts. For the descriptive statistics of each contract, see Table 5.1.

Eraker et al. (2003) observed that jumps contemporaneously appear in both of the return and volatility processes, and these two processes are correlated. Third, a number of literature (see Pan, 2002; Eraker, 2004; Lin, 2007; Andersen, Fusari & Todorov, 2015) pointed out that the jump intensity of these jump processes is time-varying and state-dependent. Forth, the volatilities of the index returns stochastically revert to their long-term mean level (e.g. Mencía & Sentana, 2013; Bardgett, Gourier & Leippold, 2019, etc.).

Inspired by the literature above, we allow correlated jumps in both of the return and the variance processes. We also assume the variance mean to be stochastic. So we call this model SVSMCJ. Now, let $\{\ln S_t\}_{t\geq 0}$ be the process of the S&P 500 index log return on a probability space with filtration $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$, where $\mathbb{P}$ refers to the physical measure, such that

$$\begin{aligned} d\ln S_t &= \left[a - \Lambda_t^{\mathbb{P}}\bar{k}^{\mathbb{P}} + \eta_S v_t - \frac{1}{2}v_t\right]dt + \sqrt{v_t}dW_{S,t}^{\mathbb{P}} + dJ_{S,t}^{\mathbb{P}}, \\ dv_t &= \kappa_v\left(m_t - v_t\right)dt + \sigma_v\sqrt{v_t}dW_{v,t}^{\mathbb{P}} + dJ_{v,t}^{\mathbb{P}}, \\ dm_t &= \left[\kappa_m\left(\theta_m - m_t\right) + \eta_m m_t\right]dt + \sigma_m\sqrt{m_t}dW_{m,t}^{\mathbb{P}}, \end{aligned} \tag{5.1}$$

where $\{v_t\}_{t\geq 0}$ is the process for variance of the return process; $\{m_t\}_{t\geq 0}$ represents the stochastic long-term mean level of varince process; $W_{S,t}^{\mathbb{P}}$ and $W_{v,t}^{\mathbb{P}}$ are correlated Wiener processes with covariance ρ; $W_{m,t}^{\mathbb{P}}$ is also a Wiener process which is independent of $W_{S,t}^{\mathbb{P}}$ and $W_{v,t}^{\mathbb{P}}$; η_S is the risk premium for the stock; η_m are the diffusion risk premium associated with $\{m_t\}_{t\geq 0}$; a is a constant; κ_v and κ_m accommodate reverting speeds; σ_v and σ_m measure the volatilities of $\{v_t\}_{t\geq 0}$ and $\{m_t\}_{t\geq 0}$, respectively; and θ_m refers to the long-term mean of $\{m_t\}_{t\geq 0}$. Morevoer, we specify the jump processes, $\{J_{S,t}^{\mathbb{P}}\}_{t\geq 0}$ and $\{J_{v,t}^{\mathbb{P}}\}_{t\geq 0}$ in Eq. (5.1), as compound Poisson processes with a jump intensity

$\Lambda_t^{\mathbb{P}} = \lambda_0^{\mathbb{P}} + \lambda_1^{\mathbb{P}} v_t + \lambda_2^{\mathbb{P}} m_t$, such that,

$$J_{S,t}^{\mathbb{P}} = \sum_{i=1}^{N_t} x_{S,i}^{\mathbb{P}}, \tag{5.2}$$

$$J_{v,t}^{\mathbb{P}} = \sum_{i=1}^{N_t} x_{v,i}^{\mathbb{P}}, \tag{5.3}$$

where the process N_t measures the number of jumps in the index return and variance processes; the jump size of variance process, $x_v^{\mathbb{P}}$, follows an exponential distribution with mean $\mu_v^{\mathbb{P}}$; and the jump size ($x_S^{\mathbb{P}}$) of $\{J_{S,t}^{\mathbb{P}}\}_{t\geq 0}$ follows a normal distribution, which is conditional on $x_v^{\mathbb{P}}$, that is, $x_S^{\mathbb{P}}|x_v^{\mathbb{P}} \sim N(\mu_J^{\mathbb{P}} + \rho_J x_v^{\mathbb{P}}, \sigma_J^2)$. According to Pan (2002), the mean jump size related to the process of S&P 500 index return is given by

$$\bar{k}^{\mathbb{P}} = \mathbb{E}\left[\exp(x_S^{\mathbb{P}}) - 1\right] = \exp\left(\mu_J^{\mathbb{P}} + \sigma_J^2/2\right)/(1 - \rho_J \mu_v^{\mathbb{P}}) - 1.$$

In addition, we have $\mathbb{E}[x_S^{\mathbb{P}}] = \mu_J^{\mathbb{P}} + \rho_J \mu_v^{\mathbb{P}}$ and $\text{Var}(x_S^{\mathbb{P}}) = \sigma_J^2 + \rho_J^2 (\mu_v^{\mathbb{P}})^2$. Then, the correlation between $x_S^{\mathbb{P}}$ and $x_v^{\mathbb{P}}$ is calculated as $\rho_J \mu_v^{\mathbb{P}} / \sqrt{\sigma_J^2 + \rho_J^2 (\mu_v^{\mathbb{P}})^2}$.

According to the financial derivatives pricing theory, we can find a risk-neutral measure $\mathbb{Q}$ whcih is equivalent to $\mathbb{P}$, if the market is efficient. The risk-neutral dynamics of $\{\ln S_t\}_{t\geq 0}$ is given by

$$\begin{aligned}
d\ln S_t &= \left[r - q - \Lambda_t^{\mathbb{Q}} \bar{k}^{\mathbb{Q}} - \frac{1}{2} v_t\right] dt + \sqrt{v_t} dW_{S,t}^{\mathbb{Q}} + dJ_{S,t}^{\mathbb{Q}}, \\
dv_t &= \left[\kappa_v (m_t - v_t) + \eta_v v_t\right] dt + \sigma_v \sqrt{v_t} dW_{v,t}^{\mathbb{Q}} + dJ_{v,t}^{\mathbb{Q}}, \\
dm_t &= \kappa_m (\theta_m - m_t)\, dt + \sigma_m \sqrt{m_t} dW_{m,t}^{\mathbb{Q}},
\end{aligned} \tag{5.4}$$

where η_v is the risk premium related to the diffusion component of $\{v_t\}_{t\geq 0}$. Similarly, the jump intensity can be expressed as $\Lambda_t^{\mathbb{Q}} = \lambda_0^{\mathbb{Q}} + \lambda_1^{\mathbb{Q}} v_t + \lambda_2^{\mathbb{Q}} m_t$. The distributions of the jump size for $\{J_{v,t}^{\mathbb{Q}}\}_{t\geq 0}$ and $\{J_{S,t}^{\mathbb{Q}}\}_{t\geq 0}$ are given by $x_v^{\mathbb{Q}} \sim \exp(\mu_v^{\mathbb{Q}})$ and $x_S^{\mathbb{Q}}|x_v^{\mathbb{Q}} \sim N(\mu_J^{\mathbb{Q}} + \rho_J x_v^{\mathbb{Q}}, \sigma_J^2)$, respectively. So, the expectation and variance of $x_S^{\mathbb{Q}}$ are $\mathbb{E}[x_S^{\mathbb{Q}}] = \mu_J^{\mathbb{Q}} + \rho_J \mu_v^{\mathbb{Q}}$

and $\text{Var}(x_S^{\mathbb{Q}}) = \sigma_J^2 + \rho_J^2(\mu_v^{\mathbb{Q}})^2$, respectively. The $\mathbb{Q}$-mean jump size associated with the index return process is of the form $\bar{k}^{\mathbb{Q}} = \exp\left(\mu_J^{\mathbb{Q}} + \sigma_J^2/2\right)/(1 - \rho_J\mu_v^{\mathbb{Q}}) - 1$. The jump-related risk premiums for the index return (η_J^S) and the variance (η_J^v) are defined by

$$\eta_J^S = \Lambda_t^{\mathbb{P}}\left(\mu_J^{\mathbb{P}} + \rho_J\mu_v^{\mathbb{P}}\right) - \Lambda_t^{\mathbb{Q}}\left(\mu_J^{\mathbb{Q}} + \rho_J\mu_v^{\mathbb{Q}}\right), \tag{5.5}$$

$$\eta_J^v = \Lambda_t^{\mathbb{P}}\mu_v^{\mathbb{P}} - \Lambda_t^{\mathbb{Q}}\mu_v^{\mathbb{Q}}. \tag{5.6}$$

5.3 VIX Squared and Integrated Variance Premium

The CBOE volatility index (VIX) is defined as a sum of stock call and put options' prices. To derive a formula for the VIX squared based on our model, we follow the methodology which is applied by Lin (2007) and Duan and Yeh (2010). Then, the VIX squared can be computed as the expectation of the log contract under $\mathbb{Q}$, that is,

$$\left(\frac{\text{VIX}_t}{100}\right)^2 = \frac{2}{\tau}\left[\int_0^F \frac{dY}{Y^2}P(Y) + \int_F^\infty \frac{dY}{Y^2}C(Y)\right] = -\frac{2}{\tau}\mathbb{E}_t^{\mathbb{Q}}\left[\ln\frac{S_T}{F}\right], \tag{5.7}$$

where Y represents the strike price of the options, $P(Y)$ and $C(Y)$ denote the forward prices of call and put options, respectively, the forward price of a stock is defined by $F = S_t\exp(r-q)\tau$ and $\tau = 30/365$. Thus, based on Eq. (5.4), the formula for the VIX squared can be derived as

$$\begin{aligned}\left(\frac{\text{VIX}_t}{100}\right)^2 = \frac{2}{\tau}&\left[\frac{1}{2} + \lambda_1^{\mathbb{Q}}\left(\bar{k}^{\mathbb{Q}} - \mu_J^{\mathbb{Q}} - \rho_J\mu_v^{\mathbb{Q}}\right)\right]\mathbb{E}_t^{\mathbb{Q}}\left[\int_t^{t+\tau} v_u du\right] \\ &+ 2\lambda_0^{\mathbb{Q}}\left(\bar{k}^{\mathbb{Q}} - \mu_J^{\mathbb{Q}} - \rho_J\mu_v^{\mathbb{Q}}\right) + \frac{2\lambda_2^{\mathbb{Q}}}{\tau}\left(\bar{k}^{\mathbb{Q}} - \mu_J^{\mathbb{Q}} - \rho_J\mu_v^{\mathbb{Q}}\right)\mathbb{E}_t^{\mathbb{Q}}\left[\int_t^{t+\tau} m_u du\right].\end{aligned} \tag{5.8}$$

Then, we substitute $\mathbb{E}_t^{\mathbb{Q}}[\int_t^{t+\tau} v_u du]$ and $\mathbb{E}_t^{\mathbb{Q}}[\int_t^{t+\tau} m_u du]$ into the above equation, and obtain the following proposition.

Proposition 5.3.1 *Under the given model* (5.4)*, the VIX squared at time* t *can be expressed as a linear function of* m_t *and* v_t*:*

$$\left(\frac{VIX_t}{100}\right)^2 = \alpha^{\mathbb{Q}}_{VIX} m_t + \beta^{\mathbb{Q}}_{VIX} v_t + \gamma^{\mathbb{Q}}_{VIX}, \tag{5.9}$$

where $\alpha^{\mathbb{Q}}_{VIX}$*,* $\beta^{\mathbb{Q}}_{VIX}$ *and* $\gamma^{\mathbb{Q}}_{VIX}$ *are given by*

$$\begin{aligned}
\alpha^{\mathbb{Q}}_{VIX} &= \frac{\alpha^{\mathbb{Q}}_v}{\tau} + \frac{2}{\tau}\left(\bar{k}^{\mathbb{Q}} - \mu^{\mathbb{Q}}_J - \rho_J \mu^{\mathbb{Q}}_v\right)\left(\alpha^{\mathbb{Q}}_v \lambda^{\mathbb{Q}}_1 + \alpha^{\mathbb{Q}}_m \lambda^{\mathbb{Q}}_2\right), \\
\beta^{\mathbb{Q}}_{VIX} &= \frac{\beta^{\mathbb{Q}}_v}{\tau}\left[1 + 2\lambda^{\mathbb{Q}}_1\left(\bar{k}^{\mathbb{Q}} - \mu^{\mathbb{Q}}_J - \rho_J \mu^{\mathbb{Q}}_v\right)\right], \\
\gamma^{\mathbb{Q}}_{VIX} &= 2\left(\bar{k}^{\mathbb{Q}} - \mu^{\mathbb{Q}}_J - \rho_J \mu^{\mathbb{Q}}_v\right)\left(\lambda^{\mathbb{Q}}_0 + \frac{\gamma^{\mathbb{Q}}_v}{\tau}\lambda^{\mathbb{Q}}_1 + \frac{\gamma^{\mathbb{Q}}_m}{\tau}\lambda^{\mathbb{Q}}_2\right) + \frac{\gamma^{\mathbb{Q}}_v}{\tau}.
\end{aligned}$$

where $\tau = 30/365$*,* $\alpha^{\mathbb{Q}}_m$*,* $\alpha^{\mathbb{Q}}_v$*,* $\beta^{\mathbb{Q}}_v$*,* $\gamma^{\mathbb{Q}}_m$ *and* $\gamma^{\mathbb{Q}}_v$ *are given as follows:*

$$\begin{aligned}
\alpha^{\mathbb{Q}}_m &= \left(1 - e^{\kappa_m (t-T)}\right)/\kappa_m, \\
\alpha^{\mathbb{Q}}_v &= \frac{\kappa_v + \lambda^{\mathbb{Q}}_2 \mu^{\mathbb{Q}}_v}{b^{\mathbb{Q}} - \kappa_m}\left(\alpha^{\mathbb{Q}}_m - \beta^{\mathbb{Q}}_v\right), \\
\beta^{\mathbb{Q}}_v &= \frac{1}{b^{\mathbb{Q}}}\left[1 - e^{b^{\mathbb{Q}}(t-T)}\right], \\
\gamma^{\mathbb{Q}}_m &= \theta_m\left[(T-t) - \alpha^{\mathbb{Q}}_m\right], \\
\gamma^{\mathbb{Q}}_v &= \left[\frac{\lambda^{\mathbb{Q}}_0 \mu^{\mathbb{Q}}_v}{b^{\mathbb{Q}}} - \frac{\kappa_m \theta_m\left(\kappa_v + \lambda^{\mathbb{Q}}_2 \mu^{\mathbb{Q}}_v\right)}{b^{\mathbb{Q}}\left(b^{\mathbb{Q}} - \kappa_m\right)}\right]\left(T - t - \beta^{\mathbb{Q}}_v\right) + \frac{\gamma^{\mathbb{Q}}_m\left(\kappa_v + \lambda^{\mathbb{Q}}_2 \mu^{\mathbb{Q}}_v\right)}{b^{\mathbb{Q}} - \kappa_m}.
\end{aligned}$$

Proof. See Appendix C.1.

The variance risk premium (VRP) is the difference of the quadratic variation (QV) for the log index return process under $\mathbb{P}$ and $\mathbb{Q}$ (Bardgett et al., 2019). The VRP from time t to T is defined by

$$\mathrm{VRP}_{t,T} = \frac{1}{T-t}\left(\mathbb{E}^{\mathbb{P}}_t\left[\mathrm{QV}_{t,T}\right] - \mathbb{E}^{\mathbb{Q}}_t\left[\mathrm{QV}_{t,T}\right]\right). \tag{5.10}$$

According to Carr and Wu (2009), the QV of the jump-diffusion model within the time interval $[t, T]$ is given by

$$\mathrm{QV}_{t,T} = \int_t^T v_u du + \sum_{i=N_t}^{N_T} (x_{S,i})^2. \tag{5.11}$$

Then, the expectations of $\mathrm{QV}_{t,T}$ under $\mathbb{P}$ and $\mathbb{Q}$ can be calculated as

$$\mathbb{E}_t^{\mathbb{P}}\left[\mathrm{QV}_{t,T}\right] = \alpha_{\mathrm{QV}}^{\mathbb{P}} m_t + \beta_{\mathrm{QV}}^{\mathbb{P}} v_t + \gamma_{\mathrm{QV}}^{\mathbb{P}}, \tag{5.12}$$

$$\mathbb{E}_t^{\mathbb{Q}}\left[\mathrm{QV}_{t,T}\right] = \alpha_{\mathrm{QV}}^{\mathbb{Q}} m_t + \beta_{\mathrm{QV}}^{\mathbb{Q}} v_t + \gamma_{\mathrm{QV}}^{\mathbb{Q}}, \tag{5.13}$$

where the coefficients $\alpha_{\mathrm{QV}}^{\mathbb{Q}}$, $\beta_{\mathrm{QV}}^{\mathbb{Q}}$, $\gamma_{\mathrm{QV}}^{\mathbb{Q}}$, $\alpha_{\mathrm{QV}}^{\mathbb{P}}$, $\beta_{\mathrm{QV}}^{\mathbb{P}}$ and $\gamma_{\mathrm{QV}}^{\mathbb{P}}$ are given as follows:

$$\begin{aligned}
\alpha_{\mathrm{QV}}^{\mathbb{Q}} &= \left[\sigma_J^2 + \rho_J^2(\mu_v^{\mathbb{Q}})^2 + \left(\mu_J^{\mathbb{Q}} + \rho_J\mu_v^{\mathbb{Q}}\right)^2\right]\left(\lambda_1^{\mathbb{Q}}\alpha_v^{\mathbb{Q}} + \lambda_2^{\mathbb{Q}}\alpha_m^{\mathbb{Q}}\right) + \alpha_v^{\mathbb{Q}},\\
\beta_{\mathrm{QV}}^{\mathbb{Q}} &= \beta_v^{\mathbb{Q}}\left\{1 + \lambda_1^{\mathbb{Q}}\left[\sigma_J^2 + \rho_J^2(\mu_v^{\mathbb{Q}})^2 + \left(\mu_J^{\mathbb{Q}} + \rho_J\mu_v^{\mathbb{Q}}\right)^2\right]\right\},\\
\gamma_{\mathrm{QV}}^{\mathbb{Q}} &= \left[\sigma_J^2 + \rho_J^2(\mu_v^{\mathbb{Q}})^2 + \left(\mu_J^{\mathbb{Q}} + \rho_J\mu_v^{\mathbb{Q}}\right)^2\right]\left[\lambda_0^{\mathbb{Q}}(T-t) + \lambda_1^{\mathbb{Q}}\gamma_v^{\mathbb{Q}} + \lambda_2^{\mathbb{Q}}\gamma_m^{\mathbb{Q}}\right] + \gamma_v^{\mathbb{Q}},\\
\alpha_{\mathrm{QV}}^{\mathbb{P}} &= \left[\sigma_J^2 + \rho_J^2(\mu_v^{\mathbb{P}})^2 + \left(\mu_J^{\mathbb{P}} + \rho_J\mu_v^{\mathbb{P}}\right)^2\right]\left(\lambda_1^{\mathbb{P}}\alpha_v^{\mathbb{P}} + \lambda_2^{\mathbb{P}}\alpha_m^{\mathbb{P}}\right) + \alpha_v^{\mathbb{P}},\\
\beta_{\mathrm{QV}}^{\mathbb{P}} &= \beta_v^{\mathbb{P}}\left\{1 + \lambda_1^{\mathbb{P}}\left[\sigma_J^2 + \rho_J^2(\mu_v^{\mathbb{P}})^2 + \left(\mu_J^{\mathbb{P}} + \rho_J\mu_v^{\mathbb{P}}\right)^2\right]\right\},\\
\gamma_{\mathrm{QV}}^{\mathbb{P}} &= \left[\sigma_J^2 + \rho_J^2(\mu_v^{\mathbb{P}})^2 + \left(\mu_J^{\mathbb{P}} + \rho_J\mu_v^{\mathbb{P}}\right)^2\right]\left[\lambda_0^{\mathbb{P}}(T-t) + \lambda_1^{\mathbb{P}}\gamma_v^{\mathbb{P}} + \lambda_2^{\mathbb{P}}\gamma_m^{\mathbb{P}}\right] + \gamma_v^{\mathbb{P}}.
\end{aligned}$$

Thus, the VRP measures the average difference between the realized variance and the variance swap strike. In other words, an investor is willing to pay the difference to hedge the future variation of the index return. If we plug Eq. (5.12) and Eq. (5.13) into Eq. (5.10), we can gain the following formula for the VRP,

$$\mathrm{VRP}_{t,T} = \frac{1}{T-t}\left[\left(\alpha_{\mathrm{QV}}^{\mathbb{P}} - \alpha_{\mathrm{QV}}^{\mathbb{Q}}\right)m_t + \left(\beta_{\mathrm{QV}}^{\mathbb{P}} - \beta_{\mathrm{QV}}^{\mathbb{Q}}\right)v_t + \left(\gamma_{\mathrm{QV}}^{\mathbb{P}} - \gamma_{\mathrm{QV}}^{\mathbb{Q}}\right)\right]. \tag{5.14}$$

5.4 Variance Futures Valuation

In contrast to the old version of variance futures, the total time period for computing the variance of the S&P 500 index is not fixed at 3-month or 12-month. Now, we assume that an S&P 500 variance futures contract is issued at time t_0 and settled at time T. According to the contract specification of the S&P 500 variance futures on the CBOE website[1], we evaluate the variance futures at the current time t based on two parts: the historical part $(t_0, t]$ and future part (t, T), see Figure 5.2. The price for the historical part is calculated from the annualized realized variance (RV) up to the date, and the price for future part is calculated as the implied variance strike (IV) during the period between current date and the expire date. Then, the price of the variance futures can be expressed as (Biscamp & Weithers, n.d.),

$$\text{variance future price} = \frac{(N-1)\text{RV} \times 10,000 + (N_e - N)\text{IV}}{N_e - 1}, \tag{5.15}$$

where N is the total number of the S&P 500 index prices which are observed from t_0 to t; N_e is the total number of the days between the issue date and the settlement date; and RV is defined in Zhang and Huang (2010) as

$$\text{RV} = 252 \times \sum_{i=1}^{N-1} \frac{R_i^2}{N-1}, \tag{5.16}$$

where $R_i = \ln(S_{i+1}/S_i)$ is the log return of the S&P 500 index.

Let F_t^T be the price of variance future at time t. Then, it can be formulated as

$$\frac{F_t^T}{10,000} = \frac{1}{T-t_0}\left[(t-t_0)\,\text{RV} + \mathbb{E}_t^{\mathbb{Q}}\left[\text{QV}_{t,T}\right]\right]. \tag{5.17}$$

From Eq. (5.17), we can have the following proposition which establishes the price

[1] http://www.cboe.com/products/futures/va-s-p-500-variance-futures/contract-specifications

Figure 5.2: S&P 500 variance futures

formula for the S&P 500 variance futures and the relationship between the variance future and the VIX.

Proposition 5.4.1 *Consider a variance futures contract issued at t_0 and expired at T. Under the given model* (5.4)*, the price of the S&P 500 variance futures at current time t is formulated as*

$$\begin{aligned}\frac{F_t^T}{10,000} &= \left(\frac{t-t_0}{T-t_0}\right) RV + \frac{1}{T-t_0}\left(\alpha_{QV}^{\mathbb{Q}} m_t + \beta_{QV}^{\mathbb{Q}} v_t + \gamma_{QV}^{\mathbb{Q}}\right), \\ &= \left(\frac{t-t_0}{T-t_0}\right) RV + \frac{1}{T-t_0}\left[\xi^* \left(\frac{VIX_t}{100}\right)^2 + \left(\alpha_{QV}^{\mathbb{Q}} - \xi^* \alpha_{VIX}^{\mathbb{Q}}\right) m_t + \left(\gamma_{QV}^{\mathbb{Q}} - \xi^* \gamma_{VIX}^{\mathbb{Q}}\right)\right],\end{aligned} \tag{5.18}$$

where $(VIX_t/100)^2$ is given in Eq. (5.7)*; $\xi^* = \beta_{QV}^{\mathbb{Q}}/\beta_{VIX}^{\mathbb{Q}}$; $\alpha_{VIX}^{\mathbb{Q}}$, $\beta_{VIX}^{\mathbb{Q}}$ and $\gamma_{VIX}^{\mathbb{Q}}$ are given in Proposition 5.3.1; and $\alpha_{QV}^{\mathbb{Q}}$, $\beta_{QV}^{\mathbb{Q}}$, and $\gamma_{QV}^{\mathbb{Q}}$ have the forms of*

$$\alpha_{QV}^{\mathbb{Q}} = \left[\sigma_J^2 + \rho_J^2(\mu_v^{\mathbb{Q}})^2 + \left(\mu_J^{\mathbb{Q}} + \rho_J \mu_v^{\mathbb{Q}}\right)^2\right]\left(\lambda_1^{\mathbb{Q}} \alpha_v^{\mathbb{Q}} + \lambda_2^{\mathbb{Q}} \alpha_m^{\mathbb{Q}}\right) + \alpha_v^{\mathbb{Q}}, \tag{5.19}$$

$$\beta_{QV}^{\mathbb{Q}} = \beta_v^{\mathbb{Q}}\left\{1 + \lambda_1^{\mathbb{Q}}\left[\sigma_J^2 + \rho_J^2(\mu_v^{\mathbb{Q}})^2 + \left(\mu_J^{\mathbb{Q}} + \rho_J \mu_v^{\mathbb{Q}}\right)^2\right]\right\}, \tag{5.20}$$

$$\gamma_{QV}^{\mathbb{Q}} = \left[\sigma_J^2 + \rho_J^2(\mu_v^{\mathbb{Q}})^2 + \left(\mu_J^{\mathbb{Q}} + \rho_J \mu_v^{\mathbb{Q}}\right)^2\right]\left[\lambda_0^{\mathbb{Q}}(T-t) + \lambda_1^{\mathbb{Q}} \gamma_v^{\mathbb{Q}} + \lambda_2^{\mathbb{Q}} \gamma_m^{\mathbb{Q}}\right] + \gamma_v^{\mathbb{Q}}, \tag{5.21}$$

where $\alpha_m^{\mathbb{Q}}$, $\alpha_v^{\mathbb{Q}}$, $\beta_v^{\mathbb{Q}}$, $\gamma_m^{\mathbb{Q}}$ *and* $\gamma_v^{\mathbb{Q}}$ *are given as follows*

$$
\begin{aligned}
\alpha_m^{\mathbb{Q}} &= \left(1 - e^{\kappa_m (t-T)}\right)/\kappa_m, \\
\alpha_v^{\mathbb{Q}} &= \frac{\kappa_v + \lambda_2^{\mathbb{Q}} \mu_v^{\mathbb{Q}}}{b^{\mathbb{Q}} - \kappa_m} \left(\alpha_m^{\mathbb{Q}} - \beta_v^{\mathbb{Q}}\right), \\
\beta_v^{\mathbb{Q}} &= \frac{1}{b^{\mathbb{Q}}} \left[1 - e^{b^{\mathbb{Q}}(t-T)}\right], \\
\gamma_m^{\mathbb{Q}} &= \theta_m \left[(T-t) - \alpha_m^{\mathbb{Q}}\right], \\
\gamma_v^{\mathbb{Q}} &= \left[\frac{\lambda_0^{\mathbb{Q}} \mu_v^{\mathbb{Q}}}{b^{\mathbb{Q}}} - \frac{\kappa_m \theta_m \left(\kappa_v + \lambda_2^{\mathbb{Q}} \mu_v^{\mathbb{Q}}\right)}{b^{\mathbb{Q}} \left(b^{\mathbb{Q}} - \kappa_m\right)}\right] \left(T - t - \beta_v^{\mathbb{Q}}\right) + \frac{\gamma_m^{\mathbb{Q}} \left(\kappa_v + \lambda_2^{\mathbb{Q}} \mu_v^{\mathbb{Q}}\right)}{b^{\mathbb{Q}} - \kappa_m}.
\end{aligned}
$$

Proof. See Appendix C.2.

5.5 Model Calibration and Estimation Results

In this section, we employ the joint estimation method which combines the MLE with the UKF. This method has been proved to work efficiently with the jump diffusion model which contains the diffusion latent processes in Chapter 3 and Chapter 4. As mentioned in previous two chapters, we conduct the Euler approximation on the latent process $\{m_t\}_{t\geq 0}$ under $\mathbb{P}$, see Eq. (5.1), before applying the UKF. The discrete version of $\{m_t\}_{t\geq 0}$ is written as,

$$
\Delta m_t = \left[\kappa_m \left(\theta_m - m_t\right) + \eta_m m_t\right] \Delta t + \sigma_m \sqrt{m_t \Delta t} \epsilon_t, \tag{5.22}
$$

where ϵ_t denotes a standard normal random variable. Next, we define the measurement function of the time t as

$$\begin{pmatrix} F_t^{T_1} \\ F_t^{T_2} \\ \vdots \\ F_t^{T_M} \end{pmatrix} = \begin{pmatrix} \hat{F}_t^{T_1} \\ \hat{F}_t^{T_2} \\ \vdots \\ \hat{F}_t^{T_M} \end{pmatrix} + \sigma_e \begin{pmatrix} e_t^1 \\ e_t^2 \\ \vdots \\ e_t^M \end{pmatrix}, \tag{5.23}$$

where M counts the number of variance futures contracts which are traded at time t, $F_t^{T_i}$ and $\hat{F}_t^{T_i}$, for $i = 1, 2, \ldots, M$, represent the market data and estimated variance futures prices, respectively; $e_t^1, e_t^2, \ldots, e_t^M$ are independent standard normal random variables, and σ_e captures the average size of estimation errors. For the detailed information about the joint method implementation, see Section 2.3.3 and Section 3.4.1.

In the calibration, we use the joint data set, including the S&P 500 index, the VIX and the variance futures data. The summary statistics of the selected data has been described in Section 5.1. We split the data into in-sample and out-of-sample. The in-sample data cover the period from 24 June 2013 to 31 October 2016 and the out-of-sample data cover period from 01 November 2016 to 14 February 2020. The average variance futures prices of in-sample and out-of-sample are 380.24 and 306.23, respectively. The in-sample data are used for the parameter estimation while the out-of-sample data are used for the model performance validation.

We not only calibrate the most complicated model (SVSMCJ) but also the simplified version of the SVSMCJ model to investigate how the different model specifications impact on the model performance. First, by removing the long-term mean process $\{m_t\}_{t\geq 0}$ from the SVSMCJ model, we gain the SVSCJ model which is the best model in Lin (2007) and Eraker (2004). Second, by restricting $\lambda_2^{\mathbb{P}} = 0$ and $\lambda_2^{\mathbb{Q}} = 0$ in the SVSMCJ model, we obtain the state-dependent jump intensity as $\lambda_0^{\mathbb{P}} + \lambda_1^{\mathbb{P}} v_t$ and $\lambda_0^{\mathbb{Q}} + \lambda_1^{\mathbb{Q}} v_t$,

respectively. This model is called SVSMCJ-R. Third, by letting $\rho_J = 0$, we set the two jump processes (the one is in the return process and the other is in the variance process) in the SVSMCJ model to be uncorrelated, so we call this model SVSMJ. The estimation results of these four models are reported in Table 5.3.

Table 5.3: Parameter estimates

Parameters	SVSCJ	SVSMJ	SVSMCJ-R	SVSMCJ
ρ	-0.2913	-0.5990	-0.8232	-0.6402
	(0.0021)	(0.0035)	(0.0029)	(0.0057)
κ_v	5.0103	6.7582	8.5001	7.0064
	(0.1350)	(0.8508)	(0.5164)	(0.1848)
σ_v	0.1877	0.3049	0.0590	0.3520
	(0.0005)	(0.0020)	(0.0008)	(0.0007)
η_v	-2.9873	-1.9803	-2.4728	-5.6115
	(0.1350)	(0.8487)	(1.0318)	(0.1863)
κ_m		0.1849	4.8718	0.0822
		(1.0435)	(0.3594)	(0.0298)
θ_m		0.1334	0.0025	0.0003
		(0.0149)	(0.0473)	(0.0017)
σ_m		0.2422	0.2152	0.0047
		(0.4530)	(0.0642)	(0.0245)
η_m		-0.1161	-0.1716	0.2828
		(0.5626)	(1.9625)	(0.4482)
$\lambda_0^{\mathbb{Q}}$	0.0088	0.2212	0.0035	0.0000
	(0.0061)	(0.2218)	(0.0167)	(0.0035)
$\lambda_1^{\mathbb{Q}}$	6.1212	2.7393	2.5135	1.1716
	(0.1138)	(1.2604)	(0.5571)	(0.0990)
$\lambda_2^{\mathbb{Q}}$		3.9805		20.9468
		(3.1681)		(0.4534)
$\mu_v^{\mathbb{Q}}$	0.8631	-0.8674	1.3193	4.0525
	(0.0257)	(0.1253)	(0.0785)	(0.1727)
$\mu_J^{\mathbb{Q}}$	-0.1257	-0.1602	0.1096	-0.1136
	(0.0559)	(0.0855)	(0.4750)	(0.2238)
ρ_J	-0.9839		-0.9966	-0.9423
	(0.0338)		(0.0669)	(0.1558)
σ_J	0.0219	0.0893	0.0001	0.4264
	(0.3245)	(0.1629)	(0.2309)	(0.6774)
σ_e	0.0086	0.0082	0.0080	0.0072
	(0.0059)	(0.0029)	(0.0046)	(0.0033)

The model parameters are estimated through the joint estimation method by using the in-sample data from 24 June 2013 to 31 October 2016. The standard errors of the estimators are presented in parentheses.

According to Table 5.3, the estimated values of ρ for the models with mean factor (the SVSMJ, SVSMCJ-R, and SVSMCJ models) are smaller than those of the SVSCJ

model which does not contain this factor. So, the stochastic mean factor affects the degree of correlation between the S&P 500 index return and variance processes. In addition, this factor makes the mean-reverting speed for the $\{v_t\}_{t\geq 0}$ much faster, as we can see that the mean reversion coefficient κ_v increases from 5.01 in the SVSCJ model to over 6.7582 in the other three models. Further, the estimates of the premium parameter, η_v, are negative across all the models, which are consistent with the finding in Bardgett et al. (2019). The result implies that the half-life (between around 30 and 50 days) under $\mathbb{P}$ is shorten than that under $\mathbb{Q}$. In addition, there is a relatively large fluctuation in the estimated values for the volatility parameter σ, which is from 0.06 to 0.35.

Furthermore, we also find that the two jump processes in the SVSCJ, SVSMCJ-R and SVSMCJ models are significantly negatively correlated as the estimators of ρ_J are very close to -1. Finally, the mean-reverting parameter κ_m is estimated at 4.87 in the SVSMCJ-R model. In contrast, the estimators of κ_m are much smaller in the SVSMJ and SVSMCJ models.

5.6 Empirical Results Analysis

In this section, we focus on analyzing both the in-sample and out-of-sample performance to figure out the "true" behaviour of underlying asset returns in accordance with the selected performance measure criteria. The two selected performance measure criteria are the root mean squared error (RMSE) and the mean absolute percentage error (MAPE).

5.6.1 Criteria for Performance Measure

The RMSE has been frequently used for measuring the model performance in the literature (e.g. Park, 2016; Du & Luo, 2019; Bardgett et al., 2019, etc.). It is the

standard deviation of the estimates errors. In other words, it is the average squared difference of the real market value ($F_t^{T_i}$) and model estimated value ($\hat{F}_t^{T_i}$), that is,

$$\text{RMSEs} = \sqrt{\frac{1}{N}\sum_{t=1}^{N}\left[\frac{1}{M_t}\sum_{i=1}^{M_t}\left(F_t^{T_i} - \hat{F}_t^{T_i}\right)^2\right]}, \tag{5.24}$$

where N measures the total number of business days in the sample, and M_t is the total number of variance futures contracts which are traded on day t.

The MAPE is another important criterion for measuring the accuracy of the model estimation in terms of percentage. It is defined by

$$\text{MAPEs} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\left(F_t^{T_i} - \hat{F}_t^{T_i}\right)}{F_t^{T_i}}\right|, \tag{5.25}$$

where n is the total number of the entries in the sample, that is $n = \sum_{t=1}^{N} M_t$.

5.6.2 Model Performance Comparison

This sub-section focuses on comparing the in-sample and out-of-sample performance in pricing the S&P 500 variance futures based on the RMSE and MAPE. Table 5.4 reports the RMSEs and MAPEs of the SVSCJ, SVSMJ, SVSMCJ-R and SVSMC models, including the overall estimated errors and the estimated errors for the S&P 500 variance futures whose day-to-maturity is less than 180 business days or greater than or equal to 180 business days, respectively.

The results in Table 5.4 show that the most complicated model (SVSMCJ) leads to the smallest overall RMSEs and MAPEs for both in-sample data (RMSE 0.0072 and MAPE 11.78%) and out-of-sample data (RMSE 0.0116 and MAPE 34.69%), thus it can capture the most economic information of the variance futures prices among these four models. Further, the models with the mean factor outperform the model without this factor (SVSCJ) under all categories in Table 5.4. The implication of this

finding is that the stochastic mean of the volatility process is an important feature in the dynamics of the S&P 500 index returns. Besides, from this table, we also observe that the overall RMSE and MAPE drop by 0.0011 and 1.63%, respectively, for the in-sample and 0.0014 and 2.81%, respectively, for the out-of-sample if the correlation between two jump processes is not considered. Thus, we can conclude that the correlation between jumps also cannot be ignored in the model. Moreover, our four stochastic volatility models perform better for the S&P 500 variance futures with a longer time-to-maturity ($\tau > 180$) than for those with a relatively shorter time-to-maturity ($\tau \leq 180$), during both of the in-sample and out-of-sample periods.

Table 5.4: Root mean squared errors and mean absolute percentage errors

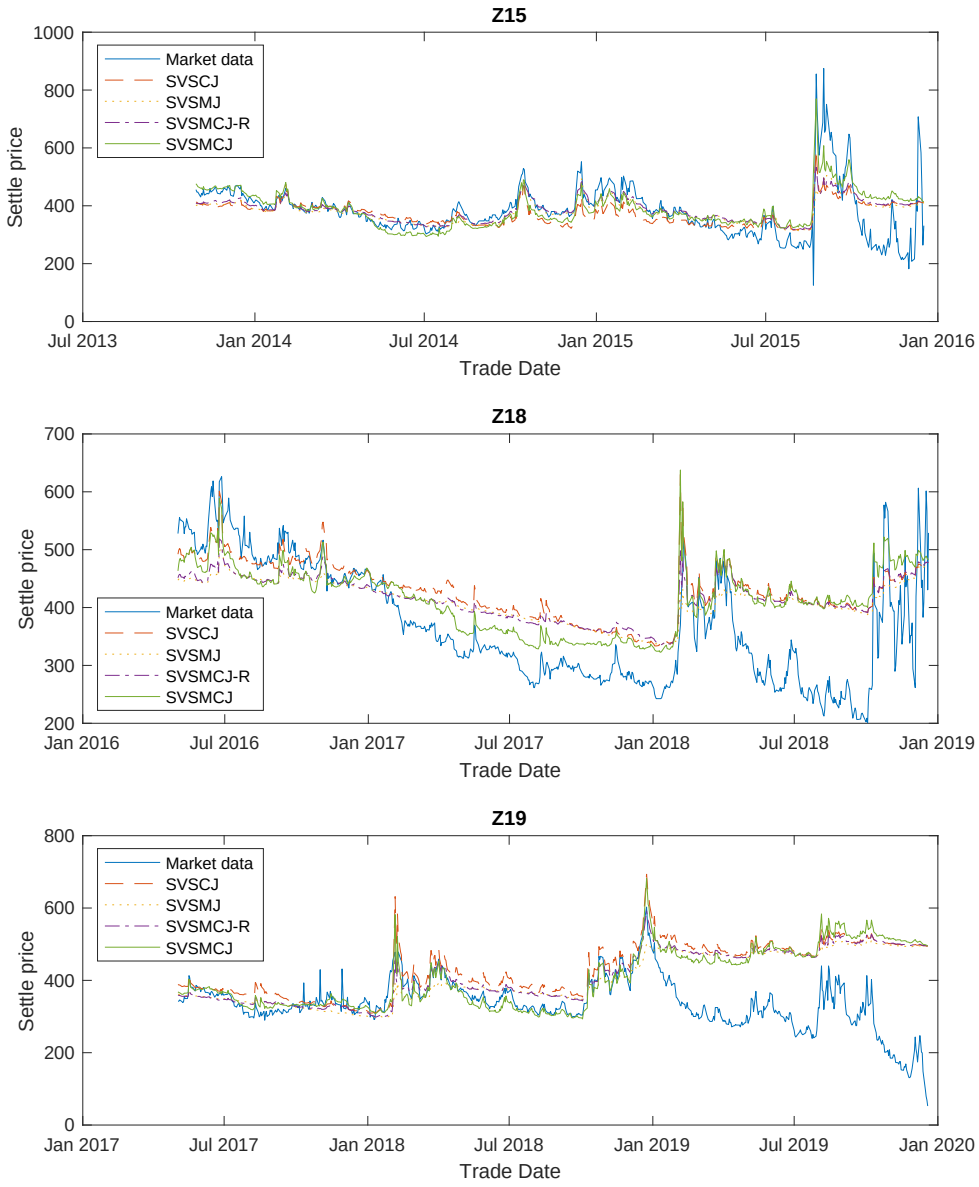
Measure criteria	**Model**	**Overall**	**Days to Maturity**	
			< 180	≥ 180
		In-sample		
RMSEs	SVSCJ	0.0086	0.0109	0.0051
	SVSMJ	0.0083	0.0108	0.0043
	SVSMCJ-R	0.0079	0.0103	0.0043
	SVSMCJ	0.0072	0.0095	0.0036
MAPEs	SVSCJ	0.1507	0.2528	0.0901
	SVSMJ	0.1341	0.2349	0.0698
	SVSMCJ-R	0.1271	0.2226	0.0664
	SVSMCJ	0.1178	0.2086	0.0591
		Out-of-sample		
RMSEs	SVSCJ	0.0130	0.0141	0.0117
	SVSMJ	0.0125	0.0143	0.0099
	SVSMCJ-R	0.0121	0.0138	0.0099
	SVSMCJ	0.0116	0.0138	0.0084
MAPEs	SVSCJ	0.3916	0.5074	0.2581
	SVSMJ	0.3750	0.5153	0.2132
	SVSMCJ-R	0.3628	0.4935	0.2122
	SVSMCJ	0.3469	0.4948	0.1763

The RMSEs and MAPEs are computed in accordance with Eq. (5.24) and Eq. (5.25), respectively, by using the estimated results in Table 5.3 in Section 5.5. The in-sample RMSE and MAPE are based on the period from 24 June 2013 to 31 October 2016. The out-of-sample RMSE and MAPE are based on the period from 01 November 2016 to 14 February 2020.

Next, we explore the model performance for a particular variance futures contract during the whole sample period. We select three variance contracts (Z15, Z18 and Z19)

which have relatively longer trading periods. Figure 5.3 provides an insight into the evolution of the daily estimated settle prices for the S&P 500 variance futures based on the four models during the period between 17 January 2014 and 18 December 2020. In each plot in Figure 5.3, the solid blue curve represents the market data. We observe that the four models have similar performance in general, but we still can tell that the estimated prices based on the SVSMCJ model fit the market data best. On the other hand, the four models do not capture the feature of rapid declines in the data, for the period between March 2018 and December 2018 in sub-figure Z18 as well as the period between 2019 and 2020 in sub-figure Z19, but they can capture the movement patterns of the market data.

Figure 5.3: Market data vs. estimated data



Notes: Each sub-figure shows the comparison of the variance futures market prices and estimated prices of a particular variance futures contract (Z15, Z18 and Z19) based on the SVSCJ, SVSMJ, SVSMCJ-R and SVSMCJ models and using the parameter estimators results in Table 5.3.

Chapter 6

Conclusion and Future Work

In this thesis, we aim to construct models which can capture characteristics of the dynamics of the indexes better and derive pricing formulae for the volatility derivatives from the proposed models. In this chapter, we summarize the major results of our research and give some suggestions which can be possible used to improve our works in the future.

6.1 Conclusion

We conduct an extensive analysis of the data about the VIX, VIX derivatives, the VXX, VXX options and variance futures. Then, we make a comprehensive performance comparison among different models to identify what kind of the model specifications can capture the characteristics of the market data best. To this end, we propose the models with different jump structures and specifications of the long-term mean process for the variance process. Then, we use the joint data set to calibrate our models by using the joint estimation method which combines the MLE and the UKF, to gain the desirable estimation results. The periods of the selected samples cover two serious financial crises, the European debt crisis and the Russian financial crisis, and fast economic growth

since the deep recession. In other words, the selected periods include both tranquility and turbulence periods, which is remarkably helpful to test how well the model can fit the real economic circumstances.

Although substantial progress has been made in applications of finite-activity jump processes in pricing volatility derivatives (e.g. Lin, 2007; Psychoyios & Dotsis, 2010; Mencía & Sentana, 2013; Park, 2016), it is unknown whether more generalized jump processes, infinite-activity jump Lévy processes, can be employed to improve the model performance in this field. It is interesting to investigate this because these types of processes can describe the phenomena where jumps frequently occur during a finite time interval. These phenomena have attracted more and more attentions in the current research. To fill this gap, we introduce four representative infinite-activity jump Lévy processes – variance gamma process, normal inverse Gaussian process and two tempered stable processes – into our models. In addition, we not only investigate the performance of infinite-activity jumps but also the double jumps in pricing volatility derivatives. Furthermore, we explore the importance of time-varying drift of the variance process.

In Chapter 3, we focus on modelling VIX index and compare the performance of our two-factor infinite-activity jump model with the OU-VJ model which is the best model in Mencía and Sentana (2013). In addition, we investigate whether having an infinite-activity jump is a feature of the VIX dynamics based on the performance analysis of the VIX derivatives pricing. Our empirical results provide critical evidence that the infinite-activity jump models (OU-VG and OU-NIG) overall outperform the finite-activity models (OU-VJ and OU-CP), hence infinite-activity jumps cannot be ignored in pricing the VIX derivatives. The presence of infinite-activity jumps implies that not only big jumps but also frequent small jumps occur in the VIX derivatives market. From the economic aspect, Lévy-type jump structures can accommodate the high-frequency occurrences of small events and the microstructure of the VIX derivatives market, including some rare events, such as a financial crisis. We also notice

that the advantage of the infinite-activity jump structures is highlighted in pricng VIX options but not in pricing the VIX futures.

Between two different infinite-activity models, we observe that the OU-NIG model is superior to the OU-VG model in general, which implies that small events tend to occur at a stable arrival rate. Notably, the OU-NIG model has better performance in pricing VIX derivatives with relatively short maturity which is less than or equal to 90 days, while the OU-VG model performs better in pricing VIX derivatives whose time to maturity is greater than 90 days.

In Chapter 4, more types of infinity-activity jumps, including a normal tempered stable process (NIG) and two tempered stable processes (TS and GTS), are examined in formulating the VXX options prices. According to the comparison of the model performance, we gain the following three main conclusions. First, we find that the infinite-activity jump structures (TS, NIG and GTS) are significantly superior to the finite-activity jump structure (MJ) in pricing VXX options across all the degrees of moneyness and the ranges of the maturities during the whole sample periods, which is similar to the conclusion obtained in Chapter 3. In addition, the infinite activity jump process with an infinite variation sample path (NIG or GTS) performs similarly to that with a finite variation sample path (TS) in this case. According to our observations, the in-sample pricing errors for the SV-NIG and SV-GTS models are smaller than that for the SV-TS model, but this is not the case for the out-of-sample pricing errors.

Second, the stochastic volatility models can capture the feature of the volatility smirk well when the time-to-maturity is less than or equal to 90 days. For estimating the prices of VXX options with a longer time-to-maturity, the infinite-activity jump models can capture the dynamics of the implied volatility better than the finite-activity model or models with no jumps for both of the in-sample and out-of-sample periods.

Third, our research results indicate that the diffusion component plays an important

role in modelling the dynamics of the VXX log returns, even if there is an infinite-activity jump component in the model. In addition, the jump components yield smaller variance and the variance series are sensitive to the variation of the VXX ETN price.

In Chapter 5, we investigate how a double jump structure affects the pricing performance of S&P 500 variance futures and what kind of the double jump specification can capture the dynamics of its underlying asset returns (S&P 500 index) best. Moreover, we allow the long-term mean of the variance process to be a mean-reverting process (CIR process) in the model. Our empirical studies show that the SVSMCJ model yields the smallest RMSEs and MAPEs, which are calculated based on the in- and out-of-sample data. Thus, we believe that jumps occur in both of the S&P 500 index returns and its variance process. In addition, the long-term mean of the variance process is time-varying. Moreover, the intensities of the jumps depend on all latent processes (the variance and long-term mean processes).

Furthermore, our variance futures pricing formula yields more accurate estimation results for the variance futures contracts with the time-to-maturity more than or equal to 180 days than those with time-to-maturity less than 180 days. However, none of our models can capture suddenly drops well in the prices of the S&P 500 variance futures.

6.2 Future Work

The research presented in this thesis has raised some new issues. Many different experiments and algorithm implementations have been left for the future due to the time and resource limitations. In the future research, it is worth to conduct deeper level analysis of the mechanisms in volatility derivatives. To address this point, we will try to use new theoretical methods and new available resources. The possible future works are listed below,

1. In Chapter 3, although our infinite-activity jump models do not show a distinct advantage in pricing VIX derivatives by using weekly data, more evidence indicates that their performance might be remarkably improved when high-frequency data are applied. For example, Eberlein and Özkan (2003) found that financial asset prices become more volatile on daily data or intraday data. In the future, we may use high-frequency data to conduct more empirical research on infinite-activity jump models.

2. The VXX, as an ETN, is a senior, unsecured, insubordinated debt security issued by Barclays Bank. We believe that our model can be extended to more sophisticated settings, for example, considering the instantaneous default rate following Carr and Wu (2010). If we assume that the arrival rate of a default event follows the CIR process on the filtered probability space $(\Omega, \mathcal{F}, \mathbb{Q})$, denoted by $\{\lambda_t\}_{t\geq 0}$, then Model (4.1) can be extended to

$$\begin{aligned} dx_t &= \left(r_t + \lambda_t - \frac{1}{2}v_t\right)dt + \sqrt{v_t}dW^{\mathbb{Q}}_{t,1}, \\ dv_t &= \kappa_v\left(\theta_v - v_t\right)dt + \sigma_v\sqrt{v_t}dW^{\mathbb{Q}}_{t,2}, \\ d\lambda_t &= \kappa_\lambda\left(\theta_\lambda - \lambda_t\right)dt + \sigma_\lambda\sqrt{\lambda_t}dW^{\mathbb{Q}}_{t,3}, \end{aligned} \tag{6.1}$$

where the process $\{\lambda_t\}_{t\geq 0}$ is independent of the other processes in the model. Model (6.1) can be rewritten in the form of

$$x_T = x_t + r_t\tau + \mathcal{T}^\lambda_t - \frac{1}{2}\mathcal{T}_t + W^{\mathbb{Q}}_{\mathcal{T}_t,1}, \tag{6.2}$$

where $\tau = T - t$; and $\mathcal{T}_t$ and $\mathcal{T}^\lambda_t$ are the random time clocks which are the non-decreasing functions. They are defined by

$$\mathcal{T}_t = \int_t^T v_s ds, \quad \text{and} \quad \mathcal{T}^\lambda_t = \int_t^T \lambda_s ds.$$

So, the characteristic function of Model (6.1) can be expressed in the form of

$$f(u;t,T,x_t,v_t,\Theta) = \mathbb{E}_t^{\mathbb{Q}}\left[\exp\left(-\int_t^T \lambda_s ds + iux_T\right)\right]. \tag{6.3}$$

Then, substituting Eq. (6.2) into the above characteristic function, we have

$$f(u) = \mathbb{E}_t^{\mathbb{Q}}\left[\exp\left(-\mathcal{T}_t^{\lambda} + iur_t\tau + iu\mathcal{T}_t^{\lambda} - \frac{1}{2}iu\mathcal{T}_t + iuW_{\mathcal{T}_t,1}^{\mathbb{Q}}\right)\right], \tag{6.4}$$

where $f(u)$ is the abbreviation of $f(u;t,T,x_t,v_t,\Theta)$. Now, we define a new complex-valued measure $\mathbb{M}$ such that

$$\frac{d\mathbb{M}}{d\mathbb{Q}} = \exp\left(iuW_{\mathcal{T}_t,1}^{\mathbb{Q}} + \frac{1}{2}u^2\mathcal{T}_t\right).$$

Under $\mathbb{M}$, the diffusion term of $\{\lambda_t\}_{t\geq 0}$ remains unchanged and the drift term of $\{v_t\}_{t\geq 0}$ is adjusted to $\beta^{\mathbb{M}} = \kappa_v\theta_v - (\kappa_v - iu\sigma_v\rho)v_t$. For simplicity, we let $b = \kappa_v - iu\sigma_v\rho$. For more details about this change, see Appendix B.3. We assume that the process $\{\lambda_t\}_{t\geq 0}$ is uncorrelated to the Wiener process in Model (4.1). Then, we obtain

$$f(u) = \exp\left(iux_t + iur_t\tau\right)\mathbb{E}_t^{\mathbb{M}}\left[-\int_t^T \alpha\mathbf{y}_s^{\top} ds\right], \tag{6.5}$$

where $\mathbf{y}_s = [\lambda_s, v_s]$, $\alpha = [\alpha_\lambda, \alpha_v]$, $\alpha_\lambda = 1 - iu$ and $\alpha_v = \frac{1}{2}\left(iu + u^2\right)$. Finally, we can gain the characteristic function $f(u)$ in the form of

$$f(u) = \exp\left(iux_t + iur_t\tau\right)\exp\left[-\psi_0(\tau) - \psi_\lambda(\tau)\lambda_t - \psi_v(\tau)v_t\right], \tag{6.6}$$

where $\psi_0(\tau)$, $\psi_v(\tau)$ and $\psi_\lambda(\tau)$ can be solved from the following ordinary differential equations,

$$\begin{aligned}
\psi'_\lambda(\tau) &= \alpha_\lambda - \kappa_\lambda\psi_\lambda(\tau) - \frac{1}{2}\sigma_\lambda^2\psi_\lambda^2(\tau),\\
\psi'_v(\tau) &= \alpha_v - b\psi_v(\tau) - \frac{1}{2}\sigma_v^2\psi_v^2(\tau),\\
\psi'_0(\tau) &= -\kappa_\lambda\theta_\lambda\psi_\lambda(\tau) - \kappa_v\theta_v\psi_v(\tau),
\end{aligned} \tag{6.7}$$

with initial condition $\psi_0(0) = 0$, $\psi_v(0) = 0$ and $\psi_\lambda(0) = 0$. Solving the Eq. (6.7), we get,

$$\psi_\lambda(\tau) = \frac{2\alpha_\lambda\left[\exp\left(-\sqrt{\Delta_\lambda}\tau\right) - 1\right]}{\left(\kappa_\lambda - \sqrt{\Delta_\lambda}\right)\left[\exp\left(-\sqrt{\Delta_\lambda}\tau\right) - 1\right] - 2\sqrt{\Delta_\lambda}}, \tag{6.8}$$

$$\psi_v(\tau) = \frac{2\alpha_v\left[\exp\left(-\sqrt{\Delta_v}\tau\right) - 1\right]}{\left(b - \sqrt{\Delta_v}\right)\left[\exp\left(-\sqrt{\Delta_v}\tau\right) - 1\right] - 2\sqrt{\Delta_v}}, \tag{6.9}$$

and

$$\begin{aligned}
\psi_0(\tau) =& \frac{\kappa_\lambda\theta_\lambda}{\sigma_\lambda^2}\left\{2\ln\left[1 - \frac{\kappa_\lambda - \sqrt{\Delta_\lambda}}{2\sqrt{\Delta_\lambda}}\left(\exp(-\sqrt{\Delta_\lambda}\tau) - 1\right)\right] - \left(\kappa_\lambda - \sqrt{\Delta_\lambda}\right)\tau\right\}\\
&+ \frac{\kappa_v\theta_v}{\sigma_v^2}\left\{2\ln\left[1 - \frac{b - \sqrt{\Delta_v}}{2\sqrt{\Delta_v}}\left(\exp(-\sqrt{\Delta_v}\tau) - 1\right)\right] - \left(b - \sqrt{\Delta_v}\right)\tau\right\},
\end{aligned} \tag{6.10}$$

where $\Delta_\lambda = \kappa_\lambda^2 + 2\sigma_\lambda\alpha_\lambda$ and $\Delta_v = \kappa_v^2 + 2\sigma_v\alpha_v$. Then, the option price formulae can be derived based on characteristic function.

In addition, we can further study whether jumps occur in the VXX return process, or its variance process or both of them.

3. The MLE conjoint with UKF can not be applied for estimating the latent processes with finite-activity jumps or infinite-activity jumps. to improve the calibration performance, we can try to employ the particle filter which has been used in Bardgett et al. (2019) to replace the UKF.

References

Aït-Sahalia, Y. & Jacod, J. (2009, February). Testing for jumps in a discretely observed process. *The Annals of Statistics*, *37*(1), 184–222.

Aït-Sahalia, Y. & Jacod, J. (2011, June). Testing whether jumps have finite or infinite activity. *The Annals of Statistics*, *39*(3), 1689–1719.

Aït-Sahalia, Y. & Kimmel, R. (2007, February). Maximum likelihood estimation of stochastic volatility models. *Journal of Financial Economics*, *83*(2), 413–452.

Alexander, C., Kapraun, J. & Korovilas, D. (2015, October). Trading and investing in volatility products. *Financial Markets, Institutions & Instruments*, *24*(4), 313–347.

Andersen, T., Fusari, N. & Todorov, V. (2015, September). The risk premia embedded in index options. *Journal of Financial Economics*, *117*(3), 558–584.

Applebaum, D. (2004). *Lévy Processes and Stochastic Calculus* (1st ed.). Cambridge, UK: Cambridge University Press.

Bakshi, G., Cao, C. & Chen, Z. (1997, December). Empirical performance of alternative option pricing models. *Journal of Finance*, *LII*(5), 2003–2049.

Bakshi, G. & Madan, D. (1999, May). Spanning and derivative-security valuation. *Journal of Financial Economics*, *55*(2), 205–238.

Bao, Q., Li, S. & Gong, D. (2012, November). Pricing VXX option with default risk and positive volatility skew. *European Journal of Operational Research*, *223*(1), 246–255.

Barati, R. (2011, November). Parameter estimation of nonlinear muskingum models using nelder-mead simplex algorithm. *Journal of Hydrologic Engineering*, *16*(11), 946–954.

Bardgett, C., Gourier, E. & Leippold, M. (2019, March). Inferring volatility dynamics and risk premia from the S&P 500 and VIX markets. *Journal of Financial Economics*, *131*(3), 593–618.

Barndorff-Nielsen, O. (1978). Hyperbolic distributions and distributions on hyperbolae. *Scandinavian Journal of Statistics*, *5*(3), 151–157.

Barndorff-Nielsen, O. (1997, November). Processes of normal inverse Gaussian type. *Finance and Stochastics*, *2*(1), 41–68.

Bates, D. (1996, Spring). Jumps and stochastic volatility: exchange rate processes implicit in deutsche market options. *The Review of Financial Studies*, *9*(1), 69–107.

Bates, D. (2000, January). Post-'87 crash fears in the S&P 500 futures option market.

Journal of Econometrics, *94*(1–2), 181–238.

Biscamp, L. & Weithers, T. (n.d.). *Variance swaps and CBOE S&P 500 variance futures* [Internet web page]. CBOE. Retrieved from `http://cfe.cboe.com/education/finaleuromoneyvarpaper.pdf`

Black, F. & Scholes, M. (1973, May-June). The pricing of options and corporate liabilities. *Journal of Political Economy*, *81*(3), 637–654.

Brennan, M. (1979, March). The pricing of contingent claims in discrete time models. *The Journal of Finance*, *34*(1), 53–68.

Broyden, C. (1956, October). A class of methods for solving nonlinear simultaneous equations. *Mathematics of Computation*, *19*(92), 577–593.

Burnett, D. (1937, July). The numerical calculation of $\int_0^\infty x^m e^{-x} f(x) dx$. *Cambridge Philosophical Society*, *33*(3), 359–362.

Cao, J., Ruan, X., Su, S. & Zhang, W. (2019, November). Pricing VIX derivatives with infinite-activity jumps. *Journal of Future Markets*, *40*(3), 329–354.

Capinski, M. & Zastawniak, T. (2003). *Mathematics for Finance: An Introduction to Financial Engineering*. London, United Kingdom: Springer.

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

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

Carr, P. & Lee, R. (2003, October). Volatility derivative. *Annual Review of Financial Economics*, *1*, 319–339.

Carr, P. & Madan, D. (1999). Option valuation using the fast Fourier transform. *Journal of Computational Finance*, *2*(4), 61–73.

Carr, P. & Madan, D. (2001). *Towards a Theory of Voltility Trading* (E. Jouini, J. Cvitanić & M. Musiela, Eds.). Cambridge, United Kingdom: Cambridge university press.

Carr, P. & Wu, L. (2003, March). Finite moment log stable process and option pricing. *Journal of Finance*, *LVIII*(2), 753–777.

Carr, P. & Wu, L. (2004, January). Time-changed Lévy processes and option pricing. *Journal of Financial Economics*, *71*(1), 113–141.

Carr, P. & Wu, L. (2009, March). Variance risk premiums. *The Review of Financial Studies*, *22*(3), 1311–1341.

Carr, P. & Wu, L. (2010, Fall). Stock options and credit default swaps: a joint framework for valuation and estimation. *Journal of Financial Econometrics*, *8*(4), 409–449.

CBOE. (2019). *White paper: CBOE volatility index* [Internet web page]. CBOE Exchange Inc. Retrieved from `https://www.cboe.com/micro/vix/vixwhite.pdf`

Chang, C., Jimenez-Martin, J., McAleer, M. & Amaral, T. (2013, August). The rise and fall of S&P 500 variance futures. *North American Journal of Economics and Finance*, *25*, 151–167.

Christoffersen, P., Dorion, C., Jacobs, K. & Karoui, L. (2014, May). Nonlinear Kalman filtering in affine term structure models. *Management Science*, *60*(9), 2248–2268.

Clark, P. (1973, January). A subordinated stochastic process model with finite variance for speculative prices. *Econometrica*, *41*(1), 135–155.

Concus, P., Cassatt, D., Jaehnig, G. & Melby, E. (1963, July). Tables for the evaluation of $\int_0^\infty x^\beta e^{-x} f(x)dx$ by gauss-laguerre quadrature. *Mathematics of Computation*, *17*(83), 245–256.

Cont, R. & Tankov, P. (2004). *Financial Modelling with Jump Processes*. London, United Kingdom: CRC press UK.

Du, D. & Luo, D. (2019, December). The pricing of jump propagation: evidence from spot and options markets. *Management Science*, *65*(5), 1949–2443.

Duan, J. & Yeh, C. (2010, November). Jump and volatility risk premiums implied by VIX. *Journal of Economic Dynamics & Control*, *34*(11), 2232–2244.

Duffie, D., Pan, J. & Singleton, K. (2000, November). Transform analysis and asset pricing for affine jump-diffusions. *Econometrica*, *68*(6), 1343–1376.

Duncan, C. (1998). *Fundamental Statistics for Social Research : Step-by-step Calculations and Computer Techniques Using SPSS for Windows* (1th ed.). Abingdon, United Kingdom: Routledge.

Eberlein, E., Keller, U. & Prause, K. (1988, July). New insights into smile, mispricing, and value at risk: the hyperbolic model. *The Journal of Business*, *71*(3), 371–405.

Eberlein, E. & Özkan, F. (2003, April). Time consistency of Lévy models. *Quantitative Finance*, *3*(1), 40–50.

Eraker, B. (2004, June). Do stock prices and volatility jump? reconciling evidence from spot and option prices. *Journal of Finance*, *59*(3), 1367–1403.

Eraker, B., Johannes, M. & Polson, N. (2003, May). The impact of jumps in volatility and returns. *Journal of Finance*, *58*(3), 1269–1300.

Fama, E. (1965, January). The behaviour of stock-market prices. *Journal of Business*, *38*(1), 34–105.

Fletcher, R. (1987). *Practical Methods of Optimization* (2nd ed.). West Sussex, UK: John Wiley & Sons.

Fouque, J., Papanicolaou, G. & Sircar, K. (2000). *Derivatives in Financial Markets with Stochastic Volatility*. Cambridge, UK: Cambridge University Press.

Garman, M. & Kohlhagen, S. (1983, December). Foreign currency options values. *Journal of International Money and Finance*, *2*(3), 231–237.

Gehricke, S. & Zhang, J. (2018, April). Modeling VXX. *Journal of Futures Markets*, *38*(8), 958–976.

Gehricke, S. & Zhang, J. (2019, August). The implied volatility smirk in the VXX options market. *Applied Economics*, *52*(8), 769–788.

Gehricke, S. & Zhang, J. (2020, June). Modeling VXX under jump-diffusion with stochastic long-term mean. *Journal of Futures Markets*, *40*(10), 1508–1534.

Geman, H. (2002). Pure jump Lévy processes for asset price modelling. *Journal of Banking & Finance*, *26*(7), 1297–1316.

Goard, J. & Mazur, M. (2013, February). Stochastic volatility models and the pricing of VIX options. *Mathematical Finance*, *23*(3), 439–458.

Goutte, S., Ismail, A. & Pham, H. (2017, May). Regime-switching stochastic volatility model: estimation and calibration to VIX options. *Applied Mathematical Finance*,

24(1), 38–75.

Grasselli, M. & Wagalath, L. (2020, June). VIX versus VXX: A joint analytical framework. *International Journal of Theoretical and Applied Finance*, *23*(5), 1–39.

Grewal, M. & Andrews, A. (2014). *Kalman Filtering: Theory and Practice Using MATLB* (4th ed.). Hoboken, New Jersey: John Wiley & Sons, Inc.

Grewal, M. & Andrews, A. (2015). *Kalman Filtering: Theory and Practice Using matlab* (4th ed.). Hoboken, New Jersey: John Wiley & Sons.

Grünbichler, A. & Longstaff, F. (1996, July). Valuing futures and options on volatility. *Journal of Banking & Finance*, *20*(6), 985–1001.

Hansen, P., Lunde, A. & Nason, J. (2011, February). The model confidence set. *Econometrica*, *79*(2), 453–497.

Harrison, J., Pitbladdo, R. & Schaefer, S. (1984, July). Continuous price processes in frictionless markets have infinite variation. *Journal of Business*, *57*(3), 353–365.

Heston, S. (1993, April). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies*, *6*(2), 327–343.

Hildebrand, F. (1987). *Introduction to Numerical Analysis* (2nd ed.). New York, USA: Dover Publications, Inc.

Huang, J. & Wu, L. (2004, June). Specification analysis of option pricing models based on time-changed Lévy processes. *Journal of Finance*, *59*(3), 1405–1439.

Hull, J. & White, A. (1987, June). The pricing of options on assets with stochastic volatilities. *Journal of Finance*, *42*(2), 281–300.

Ikeda, N. & Watanabe, S. (1992). *Stochastic Differential Equations and Diffusion Processes* (2st ed.). Tokyo: North Holland.

Jacod, J. & Protter, P. (2004). *Probability Essentials* (2th ed.). Berlin, Germany: Springer Verlag Berlin and Heidelberg GmbH & Co. KG.

Johnson, H. & Shanno, D. (1987, June). Option pricing when the variance is changing. *Journal of Financial and Quantitative Analysis*, *22*(2), 143–151.

Julier, S. & Uhlmann, J. (1997, July). New extension of the Kalman filter to nonlinear. *SPIE Proceedings, Signal Processing, Sensor Fusion, and Target Recognition IV*, *3068*(2), 182–193.

Kaeck, A. & Alexander, C. (2013, June). Continuous-time VIX dynamics: on the role of stochastic volatility of volatility. *International Review of Financial Analysis*, *28*, 46–56.

Kaeck, A., Rodrigues, P. & Seeger, N. (2018, May). Model complexity and out-of-sample performance: evidence from S&P 500 Index return. *Journal of Economic Dynamics & Control*, *90*(C), 1–29.

Kandepu, R., Foss, B. & Imsland, L. (2008, August-September). Spanning and derivative-security valuation. *Journal of Process Control*, *18*(7-8), 753–768.

Kanniainen, J., Lin, B. & Yang, H. (2014, June). Estimating and using GARCH models with VIX data for option valuation. *Journal of Banking & Finance*, *43*, 200–211.

Kc̈hler, U. & Tappe, S. (2013, December). Tempered stable distributions and processes. *Stochastic Processes and their Applications*, *123*(12), 4256–4293.

Kiesel, R. & Rahe, F. (2017, March). Option pricing under time-varying risk-aversion with applications to risk forecasting. *Journal of Banking & Finance*, *76*, 120–138.

Koponen, I. (1995, July). Analytic approach to the problem of convergence of truncated Lévy flights towards the gaussian stochastic process. *Physical Review E*, *52*(1), 1197–1199.

Kou, S. (2002, August). A jump-diffusion model for option pricing. *Management Science*, *48*(8), 1086–1101.

Kou, S. (2008). Jump-diffusion model for asset pricing in financial engineering. *Handbook in Operations Research & Management Science*, *15*, 73–116.

LeRoy, P. (1989, December). Efficient capital markets and martingales. *Journal of Economic Literature*, *27*(4), 1583–1621.

Lewis, A. & Overton, M. (2013, February). Nonsmooth optimization via quasi-Newton methods. *Mathematical Programming Series A*, *141*(1-2), 135–163.

Lian, G., Zhu, S., Elliott, R. & Cui, Z. (2017, February). Semi-analytical valuation for discrete barrier options under time-dependent Lévy processes. *Journal of Banking & Finance*, *75*, 167–183.

Lin, Y. (2007, December). Pricing VIX futures: evidence from integrated physical and risk-neutral probability measures. *Journal of Futures Markets*, *27*(12), 1175–1217.

Luo, X., Zhang, J. & Zhang, W. (2019, August). Instantaneous squared VIX and VIX derivatives. *Journal of Futures Markets*, *39*(10), 1193–1213.

Madan, D. & Milne, F. (1991, October). Option pricing with V.G. martingale components. *Mathematical Finance*, *1*(4), 39–55.

Madan, D. & Seneta, E. (1990, October). The variance gamma (V.G.) model for share market returns. *Journal of Business*, *63*(4), 511–524.

Malkiel, B. (2003, Winter). The efficient market hypothesis and its critics. *Journal of Economic Perspectives*, *17*(1), 59–82.

Mandelbrot, B. (1963, October). The variation of certain speculative prices. *Journal of Business*, *36*(4), 394–419.

Mencía, J. & Sentana, E. (2013, May). Valuation of VIX derivatives. *Journal of Financial Economics*, *108*(2), 367–391.

Merton, R. (1973, Spring). Theory of rational option pricing. *The Bell Journal of Economics and Management Science*, *4*(1), 141–183.

Merton, R. (1976, January-March). Option pricing when the underlying stock returns are discontinuous. *Journal of Financial Economics*, *3*(1-2), 125–144.

Merton, R. & Samuelson, P. (1974, May). Fallacy of the log-normal approximation to optimal portfolio decision-making over many periods. *Journal of Financial Economics*, *1*(1), 67–94.

Øksendal, B. & Sulem, A. (2005). *Applied Stochastic Control of Jump Diffusion* (2th ed.). New York, USA: Springer.

Pan, J. (2002, January). The jump-risk premia implicit in options: evidence from an integrated time-series study. *Journal of Financial Economics*, *63*(1), 3–50.

Park, Y. (2015, November). Volatility-of-volatility and tail risk hedging returns. *Journal of Financial Markets*, *26*, 38–63.

Park, Y. (2016, May). The effects of asymmetric volatility and jumps on the pricing of VIX derivatives. *Journal of Econometrics*, *192*(1), 313–328.

Psychoyios, D. & Dotsis, G. (2010, December). A jump diffusion model for VIX volatility options and futures. *Review of Quantitative Finance and Accounting*, *35*(3), 245–269.

Rhoads, R. (2011). *Trading VIX Derivatives: Trading and Hedging Strategies Using VIX Futures, Options and Exchange-traded Notes* (1st ed.). West Sussex, UK: John Wiley & Sons.

Rossi, R. (2018). *Mathematical Statistics: An Introduction to Likelihood Based Inference* (1st ed.). Hoboken, New Jersey: John Wiley & Sons, Inc.

Samuelson, P. (1963). Risk and uncertainty: a fallacy of large numbers. *Scientia*, *57*(98), 108–113.

Samuelson, P. (1965, Spring). Rational theory of warrant pricing. *Industrial Management Review*, *6*(2), 13–31.

Särkkä, S. (2007, October). On unscented Kalman filtering for state estimation of continuous-time nonlinear systems. *IEEE Transactions on Automatic Control*, *52*(9), 1631–1641.

Sato, K. (1999). *Lévy Processes and Infinitely Divisible Distributions* (1st ed.). Cambridge, UK: Cambridge University Press.

Schoutens, W. (2003). *Lévy Processes in Finance* (1th ed.). West Sussex, UK: John Wiley & Sons.

Scott, H. (2002, January). Pricing stock options in a jump-diffusion model with stochastic volatility and interest rates: application of fourier inversion methods. *Mathematical Finance*, *26*(6), 413–426.

Shreve, S. (2008). *Stochastic Calculus for Finance II: Continuous-Time Models* (8th ed.). New York, USA: Springer.

Stein, E. & Stein, J. (1991). Stock price distributions with stochastic volatility: an analytic approach. *The Review of Financial Studies*, *4*(4), 727–752.

Tan, X., Wang, C., Lin, W., Zhang, J., Li, S., Zhao, X. & Zhang, Z. (2021). The term structure of the VXX option smirk: pricing VXX option with a two-factor model and asymmetry jumps. *Journal of Futures Markets*, forthcoming.

Todorov, V. & Tauchen, G. (2011a, December). Volatility jumps. *Journal of Business & Economic Statistics*, *29*(3), 356–371.

Todorov, V. & Tauchen, G. (2011b, January). Volatility jumps. *Journal of Business & Economic Statistics*, *29*(3), 356–371.

Wan, E. & Merwe, R. (2000, October). The unscented Kalman filter for nonlinear estimation. *Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium*, 153–158.

Wehn, C., Hoppe, C. & Gregoriou, G. (2013). *Rethinking Valuation and Pricing Models: Lessons Learned from the Crisis and Future Challenges* (1st ed.). Waltham, USA: Elsevier.

Whaley, R. (1993, Fall). Derivatives on market volatility: hedging tools long overdue. *Journal of Derivatives*, *1*(1), 71–84.

Wouk, A. (1987). *New Computing Environments: Microcomputers in Large-scale Scientific Computing*. USA: Society for Industrial & Applied.

Wu, L. (2011, January). Variance dynamics: joint evidence from options and high-frequency returns. *Journal of Econometrics*, *160*(1), 280–287.

Yang, H. & Kanniainen, J. (2017, March). Jump and volatility dynamics for the S&P 500: evidence for infinite-activity jumps with non-affine volatility dynamics from stock and option markets. *Journal of Business & Economic Statistics*, *21*(2), 811–844.

Zhang, J. & Huang, Y. (2010, January). The CBOE S&P 500 three-month variance futures. *Journal of Futures Markets*, *30*(1), 48–70.

Zhang, J. & Zhu, Y. (2006, April). VIX futures. *Journal of Futures Markets*, *7*(4), 521–531.

Zhu, S. & Lian, G. (2012, February). Analytical formula for VIX futures and its applications. *Journal of Futures Markets*, *32*(2), 166–190.

Appendix A

Pricing VIX Derivatives with Infinite-Activity Jumps

A.1 Proof of Proposition 3.2.1

Under the given model in Eq. (3.1), according to Ito's lemma, the characteristic function $f(t,T,x_t,m_t,v_t;u)$ of $\{x_t\}_{t\geq 0}$ satisfies the following equation:

$$\begin{aligned} df =& \frac{\partial f}{\partial t}dt + \frac{\partial f}{\partial x}dx^c + \frac{\partial f}{\partial m}dm^c + \frac{\partial f}{\partial v}dv^c + \frac{\partial^2 f}{\partial x \partial m}dx^c dm^c + \frac{\partial^2 f}{\partial x \partial v}dx^c dv^c \\ &+ \frac{\partial^2 f}{\partial m \partial v}dm^c dv^c + \frac{1}{2}\left[\frac{\partial^2 f}{\partial x^2}(dx^c)^2 + \frac{\partial^2 f}{\partial m^2}(dm^c)^2 + \frac{\partial^2 f}{\partial v^2}(dv^c)^2\right] \\ &+ \int_{|y|<1}\left[f(t,T,x_{t^-}+y,m_t,v_t;u) - f(t,T,x_{t^-},m_t,v_t;u) - y\frac{\partial f}{\partial x}\right]\nu(dy)dt \\ &+ \int_{\mathbb{R}}\left[f(t,T,x_{t^-}+y,m_t,v_t;u) - f(t,T,x_{t^-},m_t,v_t;u)\right]\mu^L(dy,dt), \end{aligned} \tag{A.1}$$

where c represents the continuous part in a process. We allow the correlation between the process $\{x_t\}_{t\geq 0}$ and the process $\{v_t\}_{t\geq 0}$, so Eq. (A.1) can be converted to

$$\begin{aligned} df = &\left\{ \frac{\partial f}{\partial t} + \kappa\left(m-x\right)\frac{\partial f}{\partial x} + \kappa_m(\theta_m - m)\frac{\partial f}{\partial m} + \kappa_v(\theta_v - v)\frac{\partial f}{\partial v} + \rho\sigma_v v\frac{\partial^2 f}{\partial x \partial v} \right. \\ &\left. + \frac{1}{2}\left[v\frac{\partial^2 f}{\partial x^2} + \sigma_m^2\frac{\partial^2 f}{\partial m^2} + \sigma_v^2 v\frac{\partial^2 f}{\partial v^2}\right]\right\} dt + \sqrt{v}\frac{\partial f}{\partial x}dW_t^x + \sigma_m\frac{\partial f}{\partial m}dW_t^m + \sigma_v\sqrt{v}\frac{\partial f}{\partial v}dW_t^v \\ &+ \int_{|y|<1}\left[f(t,T,x_{t^-}+y,m_t,v_t;u) - f(t,T,x_{t^-},m_t,v_t;u) - y\frac{\partial f}{\partial x}\right]\nu(dy)dt \\ &+ \int_{\mathbb{R}}\left[f(t,T,x_{t^-}+y,m_t,v_t;u) - f(t,T,x_{t^-},m_t,v_t;u)\right]\mu^L(dy,dt). \end{aligned} \tag{A.2}$$

We assume that the characteristic function has an exponential-affine form, as mentioned in Eq. (3.2). Thus, we have

$$\begin{aligned} \frac{df}{f} = &\left[-\psi_0'(\tau) - x\psi_x'(\tau) - m\psi_m'(\tau) - v\psi_v'(\tau) + \kappa(m-x)\psi_x(\tau) \right. \\ &+ \kappa_m(\theta_m - m)\psi_m(\tau) + \kappa_v(\theta_v - v)\psi_v(\tau) + \rho\sigma_v v\psi_x(\tau)\psi_v(\tau) \\ &\left. + \frac{1}{2}\left(v\psi_x^2(\tau) + \sigma_m^2\psi_m^2(\tau) + \sigma_v^2 v\psi_v^2(\tau)\right) + \int_{|y|<1}\left(e^{y\psi_x(\tau)} - 1 - y\psi_x(\tau)\right)\nu(y)\right]dt \\ &+ \left(\sqrt{v}\psi_x(\tau)dW_t^x + \sigma_m\psi_m(\tau)dW_t^m + \sigma_v\psi_v(\tau)dW_t^v\right) \\ &+ \int_{\mathbb{R}}\left(e^{y\psi_x(\tau)} - 1\right)\mu^L(dy,dt). \end{aligned} \tag{A.3}$$

By taking expectation with respect to the risk-neutral measure $\mathbb{Q}$ for both sides of Eq. (A.3), we have,

$$\begin{aligned}\mathbb{E}_t^{\mathbb{Q}}\left[\frac{df}{f}\right] =& \Bigg[-\psi_0'(\tau) - x\psi_x'(\tau) - m\psi_m'(\tau) - v\psi_v'(\tau) + \kappa(m-x)\psi_x(\tau) \\ &+ \kappa_m(\theta_m - m)\psi_m(\tau) + \kappa_v(\theta_v - v)\psi_v(\tau) + \rho\sigma_v v\psi_x(\tau)\psi_v(\tau) \\ &+ \frac{1}{2}\left(v\psi_x^2(\tau) + \sigma_m^2\psi_m^2(\tau) + \sigma_v^2 v\psi_v^2(\tau)\right) + \int_{|y|<1}\left(e^{y\psi_x(\tau)} - 1 - y\psi_x(\tau)\right)\nu(y) \\ &+ \int_{\mathbb{R}}\left(e^{y\psi_x(\tau)} - 1\right)\nu(dy)\Bigg]dt.\end{aligned} \tag{A.4}$$

We know that $\mathbb{E}_t^{\mathbb{Q}}[df] = 0$, then we can get the following ordinary differential equations,

$$\psi_0'(\tau) = \kappa_m\theta_m\psi_m(\tau) + \kappa_v\theta_v\psi_v(\tau) + \frac{1}{2}\sigma_m^2\psi_m(\tau)^2 - \theta_J\psi_x(\tau) + \Psi^J(-i\psi_x(\tau)), \tag{A.5}$$

$$\psi_x'(\tau) = -\kappa\psi_x(\tau), \tag{A.6}$$

$$\psi_m'(\tau) = \kappa\psi_x(\tau) - \kappa_m\psi_m(\tau), \tag{A.7}$$

$$\psi_v'(\tau) = -\kappa_v\psi_v(\tau) + \rho\sigma_v\psi_x(\tau)\psi_v(\tau) + \frac{1}{2}\psi_x^2(\tau) + \frac{1}{2}\sigma_v^2\psi_v(\tau)^2, \tag{A.8}$$

where

$$\Psi^J(-i\psi_x(\tau)) = \theta_J\psi_x(\tau) + \int_{\mathbb{R}}\left(e^{y\psi_x(\tau)} - 1 - \psi_x(\tau)y\mathbf{1}_{|y|<1}\right)\nu(dy),$$

and $\theta_J = \alpha\lambda$ for the OU-VG model or $\theta_J = \beta\lambda^2$ for the OU-NIG model. Applying the initial conditions, $\psi_x(0) = iu$ and $\psi_m(0) = 0$, Eq. (A.6) and Eq. (A.7) can be solved analytically with closed form solutions as given in Proposition 3.2.1. Generally, Eq. (A.5) and Eq. (A.8) can only be solved numerically.

A.2 UKF Method and Test

We use the UKF method to estimate the latent state vector ($\mathbf{X}_t$) with the non-linear measurement equations. In terms of the models of the latent vectors under physical measure, see Eq. (3.13) for OU-CP, OU-VG, and OU-NIG and Eq. (3.14) for OU-VJ.

Furthermore, we test the performance of the UKF method by our artificial data. In this test, we perform the UKF on the following two-factor model without jumps,

$$\begin{aligned} dx_t &= \kappa(m_t - x_t)dt + \rho\sqrt{v_t}dW_t^v + \sqrt{(1-\rho^2)v_t}dW_t^x, \\ dm_t &= \kappa_m(\theta_m - m_t)dt + \sigma_m dW_t^m, \\ dv_t &= \kappa_v(\theta_v - v_t)dt + \sigma_v\sqrt{v_t}dW_t^v. \end{aligned} \tag{A.9}$$

We simulate time series data x_t, m_t and v_t, by using the Monte Carlo method under 10,000 simulated paths. The total sample length is 3,650 days. In the simulation procedure, we set the values of the model parameters which are shown in Table A.1. We fix the initial values for the processes x_t, m_t and v_t as 2.90, 2.50 and 0.90, respectively. Then, we use the simulated data set to compute the prices of VIX futures with three different maturities (30 days, 90 days and 270 days) through Eq. (3.5).

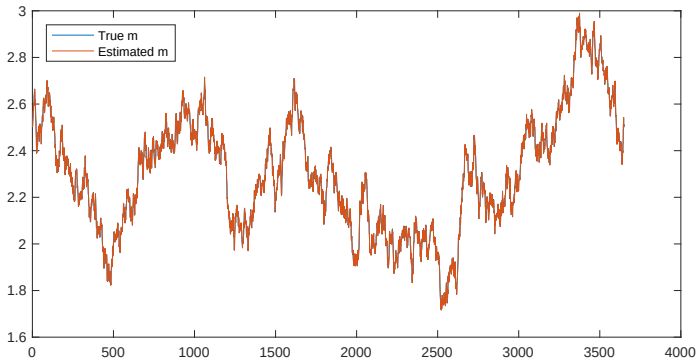
Table A.1: Parameters

Parameter	κ	κ_m	κ_v	θ_m	θ_v	σ_m	σ_v	ρ	r
Value	7.39	0.32	1.55	3.00	1.63	0.42	0.80	0.87	0.05

This table provides the setting values of the parameters which are used in the data simulation. The first eight parameters are for Model (A.9). The last parameter, r, referring to the fixed interest rate, is used to compute the futures prices.

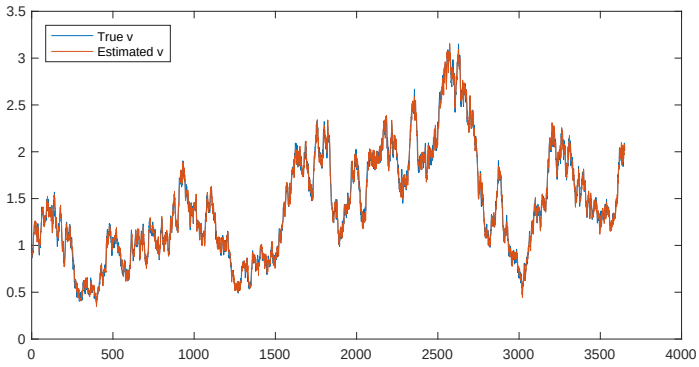
In Figures A.1 and A.2, we compare the “true" values of latent variables from the real data set with the estimated values by the UKF method. The blue and orange curves represent the real values and estimated values, respectively. We can observe that the orange curve highly overlaps with the blue curve in both figures, which means the difference between the real data and the estimated results is insignificant. Therefore, we

Figure A.1: “True" m_t vs. Predicted m_t



Notes: This graph compares the “true” m_t with the m_t which is estimated by the UKF method. We treat the Monte Carlo simulation results as “true” m_t.

Figure A.2: “True" v_t vs. Predicted v_t



Notes: Predicted v_t. This graph compares the “true” v_t with the v_t which is estimated by the UKF method. We treat the Monte Carlo simulation results as “true” v_t.

can conclude that the UKF method is an appropriate filtering method for estimating the latent variables in this case.

Appendix B

Pricing VXX Option under Lévy Processes

B.1 The relationship between the VIX and the VXX

To briefly build the relationship between the VIX and the VXX, we simply borrow the results from Cao et al. (2019). First we specify the dynamics of the log VIX under the risk-neutral measure $\mathbb{Q}$ as follows:

$$\begin{aligned} d\log \mathrm{VIX}_t &= \kappa(m - \log \mathrm{VIX}_t)dt + \sqrt{v_t}dW^{\mathbb{Q}}_{x,t} + dJ_t, \\ dv_t &= \kappa_v(\theta_v - v_t)dt + \sigma_v\sqrt{v_t}dW^{\mathbb{Q}}_{2,t}, \end{aligned} \tag{B.1}$$

where $W^{\mathbb{Q}}_{x,t}$ and $W^{\mathbb{Q}}_{2,t}$ are correlated with a correlation coefficient ϱ; κ and m are constants; other parameters and variables are the same as those in the model (4.1). Based on Proposition 1 in Cao et al. (2019), the characteristic function of $\{\log \mathrm{VIX}_t\}_{t\geq 0}$ is given in the form of

$$f(t, T, \log \mathrm{VIX}_t, m_t, v_t; u) = \mathbb{E}^{\mathbb{Q}}\left[e^{iu \log \mathrm{VIX}_T} | \mathcal{F}_t\right] = e^{\psi_0(\tau) + \psi_{\mathrm{VIX}}(\tau) \log \mathrm{VIX}_t + \psi_v(\tau) v_t}, \tag{B.2}$$

where $\tau = T - t$. According to Eq. (B.1)-(B.3) in Cao et al. (2019),

$$\frac{df}{f} = \text{Drift Term} \cdot dt + \psi_{VIX}(\tau)\sqrt{v}dW^{\mathbb{Q}}_{x,t} + \psi_v(\tau)\sigma_v\sqrt{v_t}dW^{\mathbb{Q}}_{2,t} + dJ_t - \Psi^J(-i)dt, \tag{B.3}$$

and $\psi_0(\tau)$, $\psi_{VIX}(\tau)$ and $\psi_v(\tau)$ could be solved from

$$\text{Drift Term} = 0, \tag{B.4}$$

using the martingale condition, i.e.,

$$\mathbb{E}^{\mathbb{Q}}\left[\frac{df}{f}\right] = 0. \tag{B.5}$$

Actually, the VIX futures price could be expressed as

$$F(t, T, \log \text{VIX}_t, m_t, v_t) := \mathbb{E}^{\mathbb{Q}}[\text{VIX}_T|\mathcal{F}_t] = \mathbb{E}^{\mathbb{Q}}[e^{\log \text{VIX}_T}|\mathcal{F}_t] = f(t, T, \log \text{VIX}_t, m_t, v_t; -i). \tag{B.6}$$

Based on the VXX Prospectus provided by Barclays Bank and (Gehricke & Zhang, 2018), the VXX tracks the performance of a position in the nearest and second-nearest maturing VIX futures contracts, which essentially creates a nearly constant 1-month maturity. Therefore, in continues-time, under the risk-neutral measure $\mathbb{Q}$,

$$\frac{d\text{VXX}_t}{\text{VXX}_t} = rdt + \frac{dF^{1M}_t}{F^{1M}_t}, \tag{B.7}$$

where $F^{1M}_t := F(t, t + 1/12, \log \text{VIX}_t, m_t, v_t)$ is the constant 1-month maturity VIX futures price.

Due to the fact that $\mathbb{E}^{\mathbb{Q}}\left[\frac{dF^{1M}_t}{F^{1M}_t}\right] = 0$, under the risk-neutral probability measure $\mathbb{Q}$, the discounted price of $\{\text{VXX}_t\}_{t\geq 0}$ is a martingale.

Finally, using Eq. (B.4) and Eq. (B.3) leads to

$$\begin{aligned}\frac{d\mathrm{VXX}_t}{\mathrm{VXX}_t} &= rdt + \psi_{\mathrm{VIX}}(1/12)\sqrt{v}dW^{\mathbb{Q}}_{x,t} + \psi_v(1/12)\sigma_v\sqrt{v_t}dW^{\mathbb{Q}}_{2,t} + dJ_t - \Psi^J(-i)dt,\\ dv_t &= \kappa_v\left(\theta_v - v_t\right)dt + \sigma_v\sqrt{v_t}dW^{\mathbb{Q}}_{t,2}.\end{aligned} \tag{B.8}$$

B.2 Change of Probability Measure

According to Proposition 9.8 in Cont and Tankov (2004) (Sato Theorems), the Radon-Nikodym derivative of our jump diffusion models is given by

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \exp\left\{\int_0^t \eta_x\sqrt{v_s}dW^{\mathbb{P}}_{s,1} - \int_0^t \frac{\eta_v\sqrt{v_s}}{\sigma_v}dW^{\mathbb{P}}_{s,2} - \frac{1}{2}\left[\int_0^t\left(\eta_x^2 v_s + \frac{\eta_v^2 v_s}{\sigma_v^2}\right)ds\right] + U_t\right\}, \tag{B.9}$$

where $\eta_x\sqrt{v_t}$ and $\frac{\eta_v\sqrt{v_t}}{\sigma_v}$ are the risk premia for the $dW_{t,1}$ and $dW_{t,2}$, respectively; and U_t is corresponding to the Lévy jump part. Assume that the characteristic triplets are (b,σ^2,ν) and $(b^{\mathbb{P}},(\sigma^{\mathbb{P}})^2,\nu^{\mathbb{P}})$ under $\mathbb{Q}$ and $\mathbb{P}$, respectively. Based on the Sato Theorems, if $\mathbb{Q}$ and $\mathbb{P}$ are equivalent, then σ^2 must be equal to $(\sigma^{\mathbb{P}})^2$. If $\sigma = \sigma^{\mathbb{P}} = 0$, then we must have,

$$b^{\mathbb{P}} - b = \int_{-1}^{1} y\left(\nu^{\mathbb{P}} - \nu\right)(dy). \tag{B.10}$$

In our selection (TS, NIG and GTS), the characteristic triplets are in the form of $(b,0,\nu)$ and $(b^{\mathbb{P}},0,\nu^{\mathbb{P}})$. Thus, U_t can be expressed as

$$U_t = -\eta_J b^{\mathbb{P}} t + \lim_{\epsilon\to 0}\left(\sum_{s\le t,|\Delta y_s|>\epsilon}\phi(\Delta x_s) - \int_0^t\int_{|x_s|>\epsilon}\left(\exp\left(\phi(y)\right) - 1\right)\nu^{\mathbb{P}}(dy)dt\right), \tag{B.11}$$

where $\phi(y) = \ln\left(\frac{d\mathbb{Q}}{d\mathbb{P}}\right)$, y denotes the jump size and x_t is the return process which is given in the model (4.1). For the compound Poisson process, U_t is given by

$$U_t = t\left(\lambda^{\mathbb{P}} - \lambda\right) + \sum_{s \le t} \phi(\Delta x_s), \tag{B.12}$$

where $\lambda^{\mathbb{P}}$ and λ are the jump intensities under $\mathbb{P}$ and $\mathbb{Q}$, respectively.

B.3 Derivation of Characteristic Function by Using the Time Changed Method

In this section, we only derive the characteristic function of the continuous part x_t^c of our models. According to Carr and Wu (2004), the continuous part of the model can be change to the form of

$$x_T = x_t + \left(r_t - \Psi^J(-i)\right)\tau - \frac{1}{2}\mathcal{T}_t + W^{\mathbb{Q}}_{\mathcal{T}_t,1}, \tag{B.13}$$

where $\tau = T - t$, and $\mathcal{T}_t$ is a random time clock which is a non-decreasing function. It can be defined as

$$\mathcal{T}_t = \int_t^T v_s ds, \tag{B.14}$$

where v_s is the instantaneous activity rate and given by the variance process in the model (4.1). Note that the conversion between $\sqrt{v_t} dW^{\mathbb{Q}}_{t,1}$ and $W^{\mathbb{Q}}_{\mathcal{T}_t,1}$ uses Theorem 4.3 in Ikeda and Watanabe (1992). Then, the characteristic function of the model without jumps is given by

$$f(u;t,T,x_t,v_t,\Theta) = \mathbb{E}_t^{\mathbb{Q}}\left[\exp\left(iux_t + iu\left(r_t - \Psi^J(-i)\right)\tau - \frac{1}{2}iu\mathcal{T}_t + iuW^{\mathbb{Q}}_{\mathcal{T}_t,1}\right)\right]. \tag{B.15}$$

Now, we introduce a new complex-valued measure $\mathbb{M}$ with respect to $\mathbb{Q}$, which is defined by the exponential martingale (Radon-Nikodym derivative) below:

$$\begin{aligned}\frac{d\mathbb{M}}{d\mathbb{Q}} &= \exp\left(iuW^{\mathbb{Q}}_{\mathcal{T}_t,1} - \frac{1}{2}iu\mathcal{T}_t + \frac{1}{2}\left(iu+u^2\right)\mathcal{T}_t\right)\\ &= \exp\left(iuW^{\mathbb{Q}}_{\mathcal{T}_t,1} + \frac{1}{2}u^2\mathcal{T}_t\right).\end{aligned} \tag{B.16}$$

Under $\mathbb{M}$, the diffusion term of v_t remains unchanged and the drift term of v_t, denoted by $\beta^{\mathbb{M}}$, is adjusted to $\beta^{\mathbb{M}} = \kappa_v(\theta_v - v_t) + iu\sigma_v\rho v_t$. For simplicity, we let $b = \kappa_v - iu\sigma_v\rho$. Then, the characteristic function ($f(u;t,T,x_t,v_t,\Theta)$) under $\mathbb{M}$ is given by

$$\begin{aligned}f(u;t,T,x_t,v_t,\Theta) &= \exp\left(iux_t + iu\left(r_t - \Psi^J(-i)\right)\tau\right)\mathbb{E}^{\mathbb{M}}_t\left[-\frac{1}{2}\left(iu+u^2\right)\mathcal{T}_t\right]\\ &= \exp\left(iux_t + iu\left(r_t - \Psi^J(-i)\right)\tau\right)\mathbb{E}^{\mathbb{M}}_t\left[-\int_t^T\frac{1}{2}\left(iu+u^2\right)v_s ds\right].\end{aligned} \tag{B.17}$$

Applying the Laplace transform on the random time $\mathcal{T}_t$, then we can get

$$g(u;t,T,v_t,\Theta) = \mathbb{E}^{\mathbb{M}}_t\left[-\int_t^T\frac{1}{2}\left(iu+u^2\right)v_s ds\right] = \exp\left[-\psi_0(\tau) - \psi_v(\tau)v_t\right], \tag{B.18}$$

where $\psi_0(\tau)$ and $\psi_v(\tau)$ can be obtained by solving the following ordinary differential equations:

$$\begin{aligned}\psi_v'(\tau) &= \frac{1}{2}\left(iu+u^2\right) - b\psi_v(\tau) - \frac{1}{2}\sigma_v^2\psi_v^2(\tau),\\ \psi_0'(\tau) &= -\kappa_v\theta_v\psi_v(\tau).\end{aligned} \tag{B.19}$$

Appendix C

S&P 500 Variance Futures

C.1 Derivation for the VIX Squared

We derive the formula of VIX squared in three steps under the given model (5.4). In the first step, we multiply the equation of m_t by $e^{\kappa_m t}$, and obtain

$$e^{\kappa_m t} dm_t + e^{\kappa_m t} \kappa_m m_t dt = e^{\kappa_m t}(\kappa_m \theta_m dt + \sigma_m \sqrt{m_t} dW^{\mathbb{Q}}_{m,t}),$$

which implies

$$de^{\kappa_m t} m_t = e^{\kappa_m t}(\kappa_m \theta_m dt + \sigma_m \sqrt{m_t} dW^{\mathbb{Q}}_{m,t}).$$

Integrating both sides of the last equation, we obtain

$$\int_t^T de^{\kappa_m s} m_s = \int_t^T e^{\kappa_m s} \kappa_m \theta_m ds + \int_t^T e^{\kappa_m s} \sigma_m \sqrt{m_t} dW^{\mathbb{Q}}_{m,s},$$

i.e.,

$$e^{\kappa_m T} m_T - e^{\kappa_m t} m_t = \theta_m(e^{\kappa_m T} - e^{\kappa_m t}) + \sigma_m \int_t^T e^{\kappa_m s} \sqrt{m_t} dW^{\mathbb{Q}}_{m,s},$$

where $T > t$. Now, m_T can be calculated as

$$m_T = e^{\kappa_m(t-T)}m_t + \theta_m\left(1 - e^{\kappa_m(t-T)}\right) + e^{-\kappa_m T}\sigma_m \int_t^T e^{\kappa_m s}\sqrt{m_t}dW^{\mathbb{Q}}_{m,s}. \tag{C.1}$$

Based on above Eq. (C.1), we can gain the mean of $\{m_t\}_{t\geq 0}$ as

$$\mathbb{E}^{\mathbb{Q}}_t[m_T] = e^{\kappa_m(t-T)}m_t + \theta_m\left(1 - e^{\kappa_m(t-T)}\right). \tag{C.2}$$

Thus, the expected integrated long-term mean over $[t, T]$ is calculated as

$$\mathbb{E}^{\mathbb{Q}}_t[\int_t^T m_u du] = \alpha^{\mathbb{Q}}_m m_t + \gamma^{\mathbb{Q}}_m, \tag{C.3}$$

where $\alpha^{\mathbb{Q}}_m = (1 - e^{\kappa_m(t-T)})/\kappa_m$ and $\gamma^{\mathbb{Q}}_m = \theta_m\left[(T-t) - \alpha^{\mathbb{Q}}_m\right]$.

In the second step, we apply Ito's lemma to $e^{\left(\kappa_v - \eta_v - \lambda^{\mathbb{Q}}_1\mu^{\mathbb{Q}}_v\right)t}v_t$ and obtain

$$\begin{aligned} d\left[e^{b^{\mathbb{Q}}t}v_t\right] &= \left(\kappa_v - \eta_v - \lambda^{\mathbb{Q}}_1\mu^{\mathbb{Q}}_v\right)e^{b^{\mathbb{Q}}t}v_t dt + e^{b^{\mathbb{Q}}t}dv_t \\ &= e^{b^{\mathbb{Q}}t}\left[\left(\kappa_v + \lambda^{\mathbb{Q}}_2\mu^{\mathbb{Q}}_v\right)m_t + \lambda^{\mathbb{Q}}_0\mu^{\mathbb{Q}}_v\right]dt + e^{b^{\mathbb{Q}}t}\sigma_v\sqrt{v_t}dW^{\mathbb{Q}}_{v,t} \\ &\quad + e^{b^{\mathbb{Q}}t}\left(dJ^{\mathbb{Q}}_{v,t} - \mu^{\mathbb{Q}}_v\Lambda^{\mathbb{Q}}_t dt\right), \end{aligned}$$

where $b^{\mathbb{Q}} = \kappa_v - \eta_v - \lambda^{\mathbb{Q}}_1\mu^{\mathbb{Q}}_v$. Integrating both sides of the above equation for $T > t$, we can get v_T as

$$\begin{aligned} v_T =& \frac{1}{b^{\mathbb{Q}}}\left[\lambda^{\mathbb{Q}}_0\mu^{\mathbb{Q}}_v - \frac{\kappa_m\theta_m\left(\kappa_v + \lambda^{\mathbb{Q}}_2\mu^{\mathbb{Q}}_v\right)}{b^{\mathbb{Q}} - \kappa_m}\right]\left[1 - e^{b^{\mathbb{Q}}(t-T)}\right] \\ &+ e^{b^{\mathbb{Q}}(t-T)}v_t + \sigma_m e^{-b^{\mathbb{Q}}T}\int_t^T e^{b^{\mathbb{Q}}s}dW^{\mathbb{Q}}_{m,s} + \sigma_v e^{-b^{\mathbb{Q}}T}\int_t^T e^{b^{\mathbb{Q}}s}\sqrt{v_s}dW^{\mathbb{Q}}_{v,s} \\ &+ e^{-b^{\mathbb{Q}}T}\int_t^T e^{b^{\mathbb{Q}}s}\left(dJ^{\mathbb{Q}}_{v,s} - \mu^{\mathbb{Q}}_v\Lambda^{\mathbb{Q}}_s ds\right) + \frac{\kappa_v + \lambda^{\mathbb{Q}}_2\mu^{\mathbb{Q}}_v}{b^{\mathbb{Q}} - \kappa_m}\left[m_T - m_t e^{b^{\mathbb{Q}}(t-T)}\right]. \end{aligned} \tag{C.4}$$

By substituting Eq. (C.2) into Eq. (C.4), we can compute the mean of the instantaneous

variance $\{v_t\}_{t\geq 0}$ as

$$\begin{aligned}\mathbb{E}_t^{\mathbb{Q}}[v_T] =& \frac{\kappa_v + \lambda_2^{\mathbb{Q}}\mu_v^{\mathbb{Q}}}{b^{\mathbb{Q}} - \kappa_m}\left[e^{\kappa_m(t-T)} - e^{b^{\mathbb{Q}}(t-T)}\right]m_t + \frac{\theta_m\left(\kappa_v + \lambda_2^{\mathbb{Q}}\mu_v^{\mathbb{Q}}\right)}{b^{\mathbb{Q}} - \kappa_m}\left[1 - e^{\kappa_m(t-T)}\right] \\ &+ e^{b^{\mathbb{Q}}(t-T)}v_t + \left[\frac{\lambda_0^{\mathbb{Q}}\mu_v^{\mathbb{Q}}}{b^{\mathbb{Q}}} - \frac{\kappa_m\theta_m\left(\kappa_v + \lambda_2^{\mathbb{Q}}\mu_v^{\mathbb{Q}}\right)}{b^{\mathbb{Q}}\left(b^{\mathbb{Q}} - \kappa_m\right)}\right]\left[1 - e^{b^{\mathbb{Q}}(t-T)}\right].\end{aligned} \tag{C.5}$$

The expected integrated variance over $[t, T]$ is calculated as

$$\mathbb{E}_t^{\mathbb{Q}}\left[\int_t^T v_u du\right] = \alpha_v^{\mathbb{Q}} m_t + \beta_v^{\mathbb{Q}} v_t + \gamma_v^{\mathbb{Q}}, \tag{C.6}$$

where $\alpha_v^{\mathbb{Q}}$, $\beta_v^{\mathbb{Q}}$ and $\gamma_v^{\mathbb{Q}}$ are given by

$$\begin{aligned}\alpha_v^{\mathbb{Q}} &= \frac{\kappa_v + \lambda_2^{\mathbb{Q}}\mu_v^{\mathbb{Q}}}{b^{\mathbb{Q}} - \kappa_m}\left(\alpha_m^{\mathbb{Q}} - \beta_v^{\mathbb{Q}}\right), \\ \beta_v^{\mathbb{Q}} &= \frac{1}{b^{\mathbb{Q}}}\left[1 - e^{b^{\mathbb{Q}}(t-T)}\right], \\ \gamma_v^{\mathbb{Q}} &= \left[\frac{\lambda_0^{\mathbb{Q}}\mu_v^{\mathbb{Q}}}{b^{\mathbb{Q}}} - \frac{\kappa_m\theta_m\left(\kappa_v + \lambda_2^{\mathbb{Q}}\mu_v^{\mathbb{Q}}\right)}{b^{\mathbb{Q}}\left(b^{\mathbb{Q}} - \kappa_m\right)}\right]\left(T - t - \beta_v^{\mathbb{Q}}\right) + \frac{\gamma_m^{\mathbb{Q}}\left(\kappa_v + \lambda_2^{\mathbb{Q}}\mu_v^{\mathbb{Q}}\right)}{b^{\mathbb{Q}} - \kappa_m}.\end{aligned}$$

From Eq. (5.7), we can calculate the VIX squared, denoted by $\left(\frac{\text{VIX}_t}{100}\right)^2$, by substituting Eq. (C.3) and Eq. (C.6) as

$$\begin{aligned}\left(\frac{\text{VIX}_t}{100}\right)^2 &= -\frac{2}{\tau}\mathbb{E}_t^{\mathbb{Q}}\left[\ln\frac{S_T}{F}\right] = 2(r-q) - \frac{2}{\tau}\mathbb{E}_t^{\mathbb{Q}}\left[\ln S_t - \ln S_t\right] \\ &= \frac{2}{\tau}\mathbb{E}_t^{\mathbb{Q}}\left[\int_t^T\left(\left(\lambda_0^{\mathbb{Q}} + \lambda_1^{\mathbb{Q}}v_u + \lambda_2^{\mathbb{Q}}m_u\right)\bar{k}^{\mathbb{Q}} + \frac{1}{2}v_u\right)du - \int_t^T\left(\sqrt{v_u}dW_{S,u}^{\mathbb{Q}} + dJ_{S,u}^{\mathbb{Q}}\right)\right] \\ &= 2\lambda_0^{\mathbb{Q}}\left(\bar{k}^{\mathbb{Q}} - \mu_J^{\mathbb{Q}} - \rho_J\mu_v^{\mathbb{Q}}\right) + \frac{2}{\tau}\left[\frac{1}{2} + \lambda_1^{\mathbb{Q}}\left(\bar{k}^{\mathbb{Q}} - \mu_J^{\mathbb{Q}} - \rho_J\mu_v^{\mathbb{Q}}\right)\right]\mathbb{E}_t^{\mathbb{Q}}\left[\int_t^T v_u du\right] \\ &\quad + \frac{2\lambda_2^{\mathbb{Q}}}{\tau}\left(\bar{k}^{\mathbb{Q}} - \mu_J^{\mathbb{Q}} - \rho_J\mu_v^{\mathbb{Q}}\right)\mathbb{E}_t^{\mathbb{Q}}\left[\int_t^T m_u du\right] \\ &= \alpha_{\text{VIX}}^{\mathbb{Q}} m_t + \beta_{\text{VIX}}^{\mathbb{Q}} v_t + \gamma_{\text{VIX}}^{\mathbb{Q}},\end{aligned} \tag{C.7}$$

where $\alpha^{\mathbb{Q}}_{\text{VIX}}$, $\beta^{\mathbb{Q}}_{\text{VIX}}$ and $\gamma^{\mathbb{Q}}_{\text{VIX}}$ are given by

$$\begin{aligned}
\alpha^{\mathbb{Q}}_{\text{VIX}} &= \frac{\alpha^{\mathbb{Q}}_v}{\tau} + \frac{2}{\tau}\left(\bar{k}^{\mathbb{Q}} - \mu^{\mathbb{Q}}_J - \rho_J\mu^{\mathbb{Q}}_v\right)\left(\alpha^{\mathbb{Q}}_v\lambda^{\mathbb{Q}}_1 + \alpha^{\mathbb{Q}}_m\lambda^{\mathbb{Q}}_2\right),\\
\beta^{\mathbb{Q}}_{\text{VIX}} &= \frac{\beta^{\mathbb{Q}}_v}{\tau}\left[1 + 2\lambda^{\mathbb{Q}}_1\left(\bar{k}^{\mathbb{Q}} - \mu^{\mathbb{Q}}_J - \rho_J\mu^{\mathbb{Q}}_v\right)\right],\\
\gamma^{\mathbb{Q}}_{\text{VIX}} &= 2\left(\bar{k}^{\mathbb{Q}} - \mu^{\mathbb{Q}}_J - \rho_J\mu^{\mathbb{Q}}_v\right)\left(\lambda^{\mathbb{Q}}_0 + \frac{\gamma^{\mathbb{Q}}_v}{\tau}\lambda^{\mathbb{Q}}_1 + \frac{\gamma^{\mathbb{Q}}_m}{\tau}\lambda^{\mathbb{Q}}_2\right) + \frac{\gamma^{\mathbb{Q}}_v}{\tau}.
\end{aligned}$$

Similarly, under the model (5.1), m_T is given in the following equation:

$$\begin{aligned}
m_T = e^{(\kappa_m-\eta_m)(t-T)}m_t + \frac{\kappa_m\theta_m}{(\kappa_m-\eta_m)}\left(1 - e^{(\kappa_m-\eta_m)(t-T)}\right)\\
+ e^{-(\kappa_m-\eta_m)T}\sigma_m\int_t^T e^{(\kappa_m-\eta_m)s}\sqrt{m_t}dW^{\mathbb{P}}_{m,s}.
\end{aligned} \tag{C.8}$$

Then, the expected integrated long-term mean over $[t,T]$ is expressed as

$$\mathbb{E}^{\mathbb{P}}_t\Big[\int_t^T m_u du\Big] = \alpha^{\mathbb{P}}_m m_t + \gamma^{\mathbb{P}}_m, \tag{C.9}$$

where $\alpha^{\mathbb{P}}_m = \left[1 - e^{(\kappa_m-\eta_m)(t-T)}\right]/(\kappa_m-\eta_m)$ and $\gamma^{\mathbb{P}}_m = \frac{\kappa_m\theta_m}{(\kappa_m-\eta_m)}\left[(T-t) - \alpha^{\mathbb{P}}_m\right]$. We also can get v_T as

$$\begin{aligned}
v_T =& \frac{1}{b^{\mathbb{P}}}\left[\lambda^{\mathbb{P}}_0\mu^{\mathbb{P}}_v - \frac{\kappa_m\theta_m\left(\kappa_v + \lambda^{\mathbb{P}}_2\mu^{\mathbb{P}}_v\right)}{b^{\mathbb{P}} - \kappa_m + \eta_m}\right]\left[1 - e^{b^{\mathbb{P}}(t-T)}\right]\\
&+ e^{b^{\mathbb{P}}(t-T)}v_t + \sigma_m e^{-b^{\mathbb{P}}T}\int_t^T e^{b^{\mathbb{P}}s}dW^{\mathbb{P}}_{m,s} + \sigma_v e^{-b^{\mathbb{P}}T}\int_t^T e^{b^{\mathbb{P}}s}\sqrt{v_s}dW^{\mathbb{P}}_{v,s}\\
&+ e^{-b^{\mathbb{P}}T}\int_t^T e^{b^{\mathbb{P}}s}\left(dJ^{\mathbb{P}}_{v,s} - \mu^{\mathbb{P}}_v\Lambda^{\mathbb{P}}_s ds\right) + \frac{\kappa_v + \lambda^{\mathbb{P}}_2}{b^{\mathbb{P}} - \kappa_m + \eta_m}\left[m_T - m_t e^{b^{\mathbb{P}}(t-T)}\right],
\end{aligned} \tag{C.10}$$

where $b^{\mathbb{P}} = \kappa_v - \lambda^{\mathbb{P}}_1\mu^{\mathbb{P}}_v$. The expected integrated variance over $[t,T]$ is expressed as

$$\mathbb{E}^{\mathbb{P}}_t\Big[\int_t^T v_u du\Big] = \alpha^{\mathbb{P}}_v m_t + \beta^{\mathbb{P}}_v v_t + \gamma^{\mathbb{P}}_v, \tag{C.11}$$

where $\alpha_v^{\mathbb{P}}$, $\beta_v^{\mathbb{P}}$ and $\gamma_v^{\mathbb{P}}$ are given by

$$\begin{aligned}
\alpha_v^{\mathbb{P}} &= \frac{\kappa_v + \lambda_2^{\mathbb{P}}\mu_v^{\mathbb{P}}}{b^{\mathbb{P}} - \kappa_m + \eta_m}\left(\alpha_m^{\mathbb{P}} - \beta_v^{\mathbb{P}}\right), \\
\beta_v^{\mathbb{P}} &= \frac{1}{b^{\mathbb{P}}}\left[1 - e^{b^{\mathbb{P}}(t-T)}\right], \\
\gamma_v^{\mathbb{P}} &= \left[\frac{\lambda_0^{\mathbb{P}}\mu_v^{\mathbb{P}}}{b^{\mathbb{P}}} - \frac{\kappa_m\theta_m\left(\kappa_v + \lambda_2^{\mathbb{P}}\mu_v^{\mathbb{P}}\right)}{b^{\mathbb{P}}\left(b^{\mathbb{P}} - \kappa_m + \eta_m\right)}\right]\left(T - t - \beta_v^{\mathbb{P}}\right) + \frac{\gamma_m^{\mathbb{P}}\left(\kappa_v + \lambda_2^{\mathbb{P}}\mu_v^{\mathbb{P}}\right)}{b^{\mathbb{P}} - \kappa_m + \eta_m}.
\end{aligned}$$

C.2 Derivation for Quadratic Variation

Based on the given model (5.1), the expectation of future realized variance from time t to T under $\mathbb{Q}$ is

$$\mathbb{E}_t^{\mathbb{Q}}\left[\mathrm{QV}_{t,T}\right] = \mathbb{E}_t^{\mathbb{Q}}\left[\int_t^T v_u du + \sum_{i=N_t}^{N_T}(x_{S,i}^{\mathbb{Q}})^2\right] = \mathbb{E}_t^{\mathbb{Q}}\left[\int_t^T v_u du\right] + \mathbb{E}_t^{\mathbb{Q}}\left[\sum_{i=N_t}^{N_T}(x_{S,i}^{\mathbb{Q}})^2\right].$$

From above equation, the expectation of jump part is derived as follows

$$\begin{aligned}
\mathbb{E}_t^{\mathbb{Q}}\left[\sum_{i=N_t}^{N_T}(x_{S,i}^{\mathbb{Q}})^2\right] = &\left\{\lambda_0^{\mathbb{Q}}(T-t) + \lambda_1^{\mathbb{Q}}\mathbb{E}_t^{\mathbb{Q}}\left[\int_t^T v_u du\right] + \lambda_2^{\mathbb{Q}}\mathbb{E}_t^{\mathbb{Q}}\left[\int_t^T m_u du\right]\right\} \\
&\times\left[\sigma_J^2 + \rho_J^2(\mu_v^{\mathbb{Q}})^2 + \left(\mu_J^{\mathbb{Q}} + \rho_J\mu_v^{\mathbb{Q}}\right)^2\right].
\end{aligned}$$

Thus, we have

$$\begin{aligned}
\mathbb{E}_t^{\mathbb{Q}}\left[\mathrm{QV}_{t,T}\right] = &\left[\sigma_J^2 + \rho_J^2(\mu_v^{\mathbb{Q}})^2 + \left(\mu_J^{\mathbb{Q}} + \rho_J\mu_v^{\mathbb{Q}}\right)^2\right]\left(\lambda_0^{\mathbb{Q}}(T-t) + \lambda_2^{\mathbb{Q}}\mathbb{E}_t^{\mathbb{Q}}\left[\int_t^T m_u du\right]\right) \\
&+ \left\{1 + \lambda_1^{\mathbb{Q}}\left[\sigma_J^2 + \rho_J^2(\mu_v^{\mathbb{Q}})^2 + \left(\mu_J^{\mathbb{Q}} + \rho_J\mu_v^{\mathbb{Q}}\right)^2\right]\mathbb{E}_t^{\mathbb{Q}}\left[\int_t^T v_u du\right]\right\}.
\end{aligned}$$

Then, plugging Eq. (C.3) and Eq. (C.6) into the last equation, we can get

$$\begin{aligned}\mathbb{E}_t^{\mathbb{Q}}\left[\mathrm{QV}_{t,T}\right] = &\left[\sigma_J^2 + \rho_J^2(\mu_v^{\mathbb{Q}})^2 + \left(\mu_J^{\mathbb{Q}} + \rho_J\mu_v^{\mathbb{Q}}\right)^2\right]\left[\lambda_0^{\mathbb{Q}}(T-t) + \lambda_1^{\mathbb{Q}}\gamma_v^{\mathbb{Q}} + \lambda_2^{\mathbb{Q}}\gamma_m^{\mathbb{Q}}\right] + \gamma_v^{\mathbb{Q}} \\ &+ \left\{\left[\sigma_J^2 + \rho_J^2(\mu_v^{\mathbb{Q}})^2 + \left(\mu_J^{\mathbb{Q}} + \rho_J\mu_v^{\mathbb{Q}}\right)^2\right]\left(\lambda_1^{\mathbb{Q}}\alpha_v^{\mathbb{Q}} + \lambda_2^{\mathbb{Q}}\alpha_m^{\mathbb{Q}}\right) + \alpha_v^{\mathbb{Q}}\right\} m_t \\ &+ \beta_v^{\mathbb{Q}}\left\{1 + \lambda_1^{\mathbb{Q}}\left[\sigma_J^2 + \rho_J^2(\mu_v^{\mathbb{Q}})^2 + \left(\mu_J^{\mathbb{Q}} + \rho_J\mu_v^{\mathbb{Q}}\right)^2\right]\right\} v_t.\end{aligned}$$

Thus, $\mathbb{E}_t^{\mathbb{Q}}\left[\mathrm{QV}_{t,T}\right]$ can be expressed as a linear function of m_t and v_t as

$$\mathbb{E}_t^{\mathbb{Q}}\left[\mathrm{QV}_{t,T}\right] = \alpha_{\mathrm{QV}}^{\mathbb{Q}} m_t + \beta_{\mathrm{QV}}^{\mathbb{Q}} v_t + \gamma_{\mathrm{QV}}^{\mathbb{Q}}, \tag{C.12}$$

where $\alpha_{\mathrm{QV}}^{\mathbb{Q}}$, $\beta_{\mathrm{QV}}^{\mathbb{Q}}$ and $\gamma_{\mathrm{QV}}^{\mathbb{Q}}$ are given by

$$\begin{aligned}\alpha_{\mathrm{QV}}^{\mathbb{Q}} &= \left[\sigma_J^2 + \rho_J^2(\mu_v^{\mathbb{Q}})^2 + \left(\mu_J^{\mathbb{Q}} + \rho_J\mu_v^{\mathbb{Q}}\right)^2\right]\left(\lambda_1^{\mathbb{Q}}\alpha_v^{\mathbb{Q}} + \lambda_2^{\mathbb{Q}}\alpha_m^{\mathbb{Q}}\right) + \alpha_v^{\mathbb{Q}}, \\ \beta_{\mathrm{QV}}^{\mathbb{Q}} &= \beta_v^{\mathbb{Q}}\left\{1 + \lambda_1^{\mathbb{Q}}\left[\sigma_J^2 + \rho_J^2(\mu_v^{\mathbb{Q}})^2 + \left(\mu_J^{\mathbb{Q}} + \rho_J\mu_v^{\mathbb{Q}}\right)^2\right]\right\}, \\ \gamma_{\mathrm{QV}}^{\mathbb{Q}} &= \left[\sigma_J^2 + \rho_J^2(\mu_v^{\mathbb{Q}})^2 + \left(\mu_J^{\mathbb{Q}} + \rho_J\mu_v^{\mathbb{Q}}\right)^2\right]\left[\lambda_0^{\mathbb{Q}}(T-t) + \lambda_1^{\mathbb{Q}}\gamma_v^{\mathbb{Q}} + \lambda_2^{\mathbb{Q}}\gamma_m^{\mathbb{Q}}\right] + \gamma_v^{\mathbb{Q}}.\end{aligned}$$

Similarly, the expectation of future realized variance from time t to T under $\mathbb{P}$ is in the form of

$$\mathbb{E}_t^{\mathbb{P}}\left[\mathrm{QV}_{t,T}\right] = \alpha_{\mathrm{QV}}^{\mathbb{P}} m_t + \beta_{\mathrm{QV}}^{\mathbb{P}} v_t + \gamma_{\mathrm{QV}}^{\mathbb{P}}, \tag{C.13}$$

where $\alpha_{\mathrm{QV}}^{\mathbb{P}}$, $\beta_{\mathrm{QV}}^{\mathbb{P}}$ and $\gamma_{\mathrm{QV}}^{\mathbb{P}}$ are given by

$$\begin{aligned}\alpha_{\mathrm{QV}}^{\mathbb{P}} &= \left[\sigma_J^2 + \rho_J^2(\mu_v^{\mathbb{P}})^2 + \left(\mu_J^{\mathbb{P}} + \rho_J\mu_v^{\mathbb{P}}\right)^2\right]\left(\lambda_1^{\mathbb{P}}\alpha_v^{\mathbb{P}} + \lambda_2^{\mathbb{P}}\alpha_m^{\mathbb{P}}\right) + \alpha_v^{\mathbb{P}}, \\ \beta_{\mathrm{QV}}^{\mathbb{P}} &= \beta_v^{\mathbb{P}}\left\{1 + \lambda_1^{\mathbb{P}}\left[\sigma_J^2 + \rho_J^2(\mu_v^{\mathbb{P}})^2 + \left(\mu_J^{\mathbb{P}} + \rho_J\mu_v^{\mathbb{P}}\right)^2\right]\right\}, \\ \gamma_{\mathrm{QV}}^{\mathbb{P}} &= \left[\sigma_J^2 + \rho_J^2(\mu_v^{\mathbb{P}})^2 + \left(\mu_J^{\mathbb{P}} + \rho_J\mu_v^{\mathbb{P}}\right)^2\right]\left[\lambda_0^{\mathbb{P}}(T-t) + \lambda_1^{\mathbb{P}}\gamma_v^{\mathbb{P}} + \lambda_2^{\mathbb{P}}\gamma_m^{\mathbb{P}}\right] + \gamma_v^{\mathbb{P}}.\end{aligned}$$