Pricing VIX Derivatives with Infinite-Activity Jumps

Jiling Cao
Department of Mathematical Sciences
School of Engineering, Computer and Mathematical Sciences
Auckland University of Technology
Private Bag 92006, Auckland 1142, New Zealand
Email: jiling.cao@aut.ac.nz

Xinfeng Ruan
Department of Accountancy and Finance
Otago Business School, University of Otago
Dunedin 9054, New Zealand
Email: xinfeng.ruan@otago.ac.nz

Shu Su
Department of Mathematical Sciences
School of Engineering, Computer and Mathematical Sciences
Auckland University of Technology
Private Bag 92006, Auckland 1142, New Zealand
Email: sus.bearwith@gmail.com

Wenjun Zhang
Department of Mathematical Sciences
School of Engineering, Computer and Mathematical Sciences
Auckland University of Technology
Private Bag 92006, Auckland 1142, New Zealand
Email: wenjun.zhang@aut.ac.nz

Current version: 03 April 2019

Pricing VIX Derivatives with Infinite-Activity Jumps

Abstract

In this paper, we investigate a two-factor VIX model with infinite-activity jumps, which is more realistic to reduce the errors in pricing VIX derivatives, comparing with Mencía and Sentana (2013). Our two-factor model features infinite-activity pure jump Lévy processes, central tendency, and stochastic volatility. We specify infinite-activity pure jump Lévy processes as two cases: the variance gamma (VG) process and the normal inverse Gaussian (NIG) process. We apply the combined estimation approach of unscented Kalman filter (UKF) and quasi-maximum log-likelihood estimation (QMLE) to our model and compare extensive performance among types of two-factor models with different jump processes. We find empirical evidence that the model with infinite-activity jumps superior to the models with finite-activity jumps, particularly in pricing VIX options. As a result, infinite-activity jumps should not be ignored in terms of pricing VIX derivatives.

Keywords: Infinite-activity jumps; VIX derivatives; unscented Kalman filter; quasi-maximum log-likelihood estimation.

JEL Classifications: G12; G13.

1. Introduction

Time-varying volatility is a key risk factor in financial markets, thus understanding the dynamics of volatility variation is crucial for market practitioners and academic researchers. Chicago Board Options Exchange (CBOE) Volatility Index, known as its ticker symbol VIX, is a risk-neutral gauge of the future implied volatility in the US stock market. It is calculated by S&P 500 index options market over the next 30-day period and indicates overall stock market fear.[1]

VIX can be traded through the VIX derivatives. Because of a negative correlation between the return of VIX and the return of S&P 500 index, market participants treat the VIX derivatives as the important trading vehicles for reducing exposure to the risk. Reported by CBOE, the VIX derivatives include VIX futures and VIX options. In 2004, CBOE introduced VIX futures. According to the CBOE website, the average daily volume of VIX futures rose more than 135-fold from 2004 to 2017, with a significant growth after 2010.[2] 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 risk management. Not surprisingly, the average daily volume of VIX options expanded over 28 times between 2006 and 2017.[3] Therefore, it is important to find an accurate valuation model of the VIX derivatives.

The methodologies for pricing VIX derivatives can be classified into two directions. In the first direction, the 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 the mean-reverting property. In addition, some of them add finite-activity jumps into the S&P 500 return process, 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

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

[2] See http://www.cboe.com/blogs/options-hub/2017/03/01/eight-charts-highlighting-growth-in-options-and-vix-futures.

[3] See http://www.cboe.com/blogs/options-hub/2017/08/25/record-days-for-vix-futures-and-options-volume-and-open-interest-this-month.

about research work in this direction, we refer the reader to Zhang and Zhu (2006); Zhu and Zhang (2007); Lin (2007); Lin and Chang (2009); Zhu and Lian (2012); Goutte, Ismail, and Pham (2017) and others.

In the other direction, researches directly model VIX dynamics and then derive the derivatives pricing formulae from the proposed models. In this direction, they acknowledge that volatility varies stochastically and reverts towards long-term mean (Grünbichler and Longstaff, 1996; Psychoyios and Dotsis, 2010; Mencía and Sentana, 2013, etc.). Thus, different type of mean-reverting processes, such as the square-root mean-reverting process, the arithmetic mean-reverting process and the geometric mean-reverting process, are applied for modeling VIX. In addition, Psychoyios and Dotsis (2010); Mencía and Sentana (2013); Kaeck and Alexander (2013); Goard and Mazur (2013) find clear evidence of jumps in VIX return process, in which the compound Poisson process is employed to characterize finite-activity jumps triggered by the influential financial events. As an advantage, analytical VIX derivatives pricing formulae can ne obtained based on even more complex VIX models. For example, Mencía and Sentana (2013) extend previous models into a two-factor model with three major features of VIX dynamics: central tendency, stochastic volatility of VIX and finite-activity jumps and they document that their two-factor model works better than other existing models, in terms of pricing VIX derivatives. Our research is along this direction and propose a new two-factor model with infinite-activity jumps, which can significantly improve the pricing performance of VIX derivatives.

There has been a growing interest in pricing stock options with infinite-activity jump processes (Clark, 1973; Geman, 2002; Carr, Geman, Madan, and Yor, 2003; Carr and Wu, 2003, 2004; Lian, Zhu, Elliott, and Cui, 2017, etc.). Some pieces of credible evidence show that infinite-activity jumps (a.k.a. high-frequency jumps) occur in dynamics of financial asset return. For instance, Wu (2011) figures 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) reveal that the models with infinite-activity jump processes can fit the S&P 500 option data better. 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. Through non-parametric analysis of

high-frequency VIX data, Todorov and Tauchen (2011) documents that VIX dynamics have infinite-activity jumps with infinite variation. However, there is little literature pricing VIX derivatives with infinite-activity jump processes. In order to fill this gap, we propose a new two-factor model with infinite-activity jumps to price VIX derivatives.

Several infinite-activity jump Lévy processes can be applied in 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, and Prause, 1988) and the normal inverse Gaussian (NIG) process (Barndorff-Nielsen, 1997). The main difference between these sub-classes is the tail behaviour: the tails of the NIG process are thicker than those of hyperbolic Lévy process. Another Lévy jump process with infinitely divisibility, popularly applied in finance, is the CGMY is introduced by Carr, Geman, Madan, and Yor (2002). It can generate both symmetric and asymmetric distributions through different parameter values. The variance gamma (VG) process (Madan, Carr, and Chang, 1999) is a special case of the CGMY, which also belongs to the generalized hyperbolic Lévy process class.

We select the two most representative Lévy processes as our jump process: the NIG process and the VG process. There are two main reasons why we do this. Firstly, both of them are the subordinate Brownian motion with exponential tails which tally with the features of VIX data. Secondly, there are differences between these two processes. 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 and Tankov, 2004). From the comparison, we can see how different type Lévy processes perform in the VIX derivatives pricing. Todorov and Tauchen (2011) find that VIX dynamics have infinite-activity jumps. Our paper is different to Li, Li, and Zhang (2017), in which a pure jump semimartingale (one type of time-changed Lévy process in Carr and Wu (2004)) is generated by an additive subordination. Our model considers time-varying mean level of the log VIX and is estimated using much longer data sample.

To this end, we apply an appropriate estimation procedure to calibrate models and various measurement criteria to test the performance of our model. In this paper, we apply a combined estimation approach of the unscented Kalman filter (UKF) and the quasi-maximum

log-likelihood estimation method (QMLE) to our model. Christoffersen, Dorion, Jacobs, and Karoui (2014) find that the UKF is a good method to estimate affine factor models. Following the literature, we use the root mean squared errors (RMSEs) as our major performance measurement criterion (Mencía and Sentana, 2013; Park, 2015; Yang and Kanniainen, 2017, etc.).

The rest of this paper is organized as follows. In Section 2, we first conduct a preliminary analysis on the VIX data and then introduce our two-factor VIX model with infinity-activity jumps. In Section 3, we derive the pricing formulae for VIX options and futures. In Section 4, we select market data to calibrate our two factor model and others. In Section 5, we conduct various goodness-of-fit tests and compare the performance of different models. Last section concludes.

2. A two-factor model with infinite-activity jumps

In this section, we conduct a preliminary analysis on the VIX data and introduce the specification of our two-factor model with infinite-activity jumps.

2.1 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 6550 days. We also compare the characteristics of the full sample with the Wednesday sample, which consist of all log VIX close prices on Wednesdays during the sample period. Table 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 are 0.09% and 0.01%, respectively. Both of them have a positive skewness (around 0.67) and a positive excess kurtosis (around 0.30). This means that the distribution of the log VIX is asymmetry and leptokurtotic. There may be some jump parts in the VIX dynamics, so that the distribution becomes non normal.

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 $\hat{S}_T$

Table 1: Descriptive statistics of log VIX index 1992-2017

	Mean	**Std.**	**Skew.**	**Kurt.**	**Min.**	**Med.**	**Max.**
Full Sample	2.8893	0.3569	0.6788	3.3176	2.2127	4.3927	2.8446
Wednesday	2.8884	0.3568	0.6645	3.2827	2.2311	4.3076	2.8472

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

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 95% confidence level. Theoretically, $\hat{S}_T$ approaches to 1 as $\Delta_n = T/n$ goes to 0, if jumps present in a process. By 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.

Based on the above setting, the test statistics ($\hat{S}_T$) of full sample and Wednesday sample are calculated as 0.4270 and 0.3607, respectively. Their corresponding critical values equal to 1.8752 and 1.7881, respectively (for more calculation details, see Section 5 in Aït-Sahalia and Jacod (2009)). We should reject the null hypothesis at 95% confidence level for both the full sample and Wednesday sample because $\hat{S}_T$ is less than the critical value. Therefore, statistically, there are jumps in the daily log VIX return process and the Wednesday log VIX return process. However, it is difficult to detect whether infinite-activity jumps occur in the process as we do not access the high-frequency VIX data (Todorov and Tauchen, 2011). 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.

2.2 Infinite-activity Lévy jump specification

A Lévy process has a stochastic continuous path with left limits, and independent and stationary increments (Cont and Tankov, 2004). Generally, a Lévy process contains three components: drift, Brownian motion and jump components. The jump component can be either finite-activity or infinite-activity jumps. Finite-activity jumps mean that the process gen-

erates finite a number of jumps in each finite time interval and captures relatively large movements. On the other hand, infinite-activity jumps refer to infinity jumps appear in a finite time interval. A Pure jump Lévy processes refers to a process which only involves drift and a jump component.

Let $\{L_t\}_{t\geq 0}$ be an $\mathcal{F}_t$-adapted Lévy process with $L_0 = 0$ on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{Q})$. It is well known that the differential version for a Lévy process can be expressed as

$$dL_t = adt + bdW_t + \int_{y\in\mathbb{R}} y\mu^L(dy, dt), \tag{2.1}$$

where $a \in \mathbb{R}$, $b \geq 0$ and $\{W_t\}_{t\geq 0}$ is a standard Brownian motion. The last term in (2.1) is the jump component of the process. It can be decomposed into two parts as follows,

$$\mu^L(dy, dt) = \begin{cases} \mu(dy, dt) - \nu(dy)dt & \text{if } |y| < R, \\ \mu(dy, dt) & \text{if } |y| \geq R. \end{cases} \tag{2.2}$$

Here, $R \in [0, +\infty]$, y refers to the magnitude of a jump in $\{L_t\}_{t\geq 0}$, $\mu(A, dt)$ counts the finite number of jumps whose magnitudes are in the set A ($A \subset \mathbb{R}$) within the $[t, t + dt]$, and $\nu(A)$ is a Lévy measure which is defined as $\nu(A) = \mathbb{E}[\mu(A, 1)]$ which is the compensation of $\mu(A, dt)$. When $|y| < R$, the jump component, $\mu(dy, dt) - \nu(dy)dt$, is martingale. We always set R in (2.2) as 1 in accordance with Theorem 1.8 in (Øksendal and Sulem, 2005) because $\mathbb{E}^{\mathbb{Q}}[y_t]$ is finite. Correspondingly, the characteristic function of $\{L_t\}_{t\geq 0}$ is given by the Lévy-Khintchine formula (Øksendal and Sulem, 2005)

$$\mathbb{E}[e^{iuL_t}] = e^{t\Psi(u)}, \quad t \geq 0 \text{ and } u \in \mathbb{R},$$

where $\Psi(u)$ is characteristic exponent and takes the form of

$$\Psi(u) = iau - \frac{1}{2}b^2u^2 + \int_{\mathbb{R}} \left(e^{iuy} - 1 - iuy\mathbf{1}_{|y|<1}\right)\nu(dy). \tag{2.3}$$

For a pure Lévy jump process, $b = 0$. Therefore, a pure Lévy jump process with infinite-

activity can be written as

$$dL_t^J = adt + \int_{y\in\mathbb{R}} y\mu^L(dy, dt), \tag{2.4}$$

where $\mu^L(dy, dt)$ is specified in (2.2) and $a = \mathbb{E}[L_1^J]$ because the jump term in (2.4) is a martingale. The jump intensity is $\nu(\mathbb{R})$ which equals to ∞ because this process has infinite-activity jumps (Sato, 1999). Although a pure Lévy infinite-activity jump process has an infinite number of jumps, the sample paths may have either a finite variation or an infinite variation (Sato, 1999), which depend on whether the sample paths have finite variation if $\int_{|y|<1} |y|\nu(dy)$ is finite or infinite.

The characteristic exponent ($\Psi^J(u)$) of a pure jump process with infinite activities can be extracted from (2.3) as follows

$$\Psi^J(u) = iau + \int_{\mathbb{R}} \left(e^{iuy} - 1 - iuy\mathbf{1}_{|y|<1}\right) \nu(dx). \tag{2.5}$$

Carr and Wu (2004) list a variety of examples of pure jump Lévy processes with infinite-activity jumps, such as the variance gamma process (VG), the normal inverse Gaussian process (NIG), the generalized hyperbolic process, and so on. In this paper, the GV process (Madan et al., 1999) and the NIG process (Barndorff-Nielsen, 1997) are chosen. It is well known that both of them belong to the subordinate Brownian motion. The general form of the processes in this class can be written as (Geman, 2002)

$$L_t^J = \theta S_t + \sigma W(S_t), \tag{2.6}$$

where θ is the drift term; σ is the volatility of the process; and S_t is the subordinating process. For detailed information of S_t, see Table 2 below (Geman, 2002; Cont and Tankov, 2004).

2.3 Model specifications

Mencía and Sentana (2013) show that the log Ornstein-Uhlenbeck (OU) process outperforms the square root mean-reverting process in the modelling of VIX. They also find that the dynamics of the VIX index exhibit the highly persistent variations from the long-term mean. Thus, the VIX reverts stochastically towards the long-term mean, which means that an

Table 2: The comparison of VG and NIG

Process	**VG Process**	**NIG Process**
Subordinating process	Gamma process	Inverse Guassian process
Properties	Finite variation	Infinite variation
Parameters	Variance $= w$ and Mean $= 1$	
Lévy measure	$\nu(y) = \frac{\lambda}{\lvert y\rvert} e^{\frac{\alpha y - \sqrt{\alpha^2 + 2\beta^2}\lvert y\rvert}{\beta^2}}$	$\nu(y) = \frac{\lambda\alpha}{\pi\lvert y\rvert} e^{\beta y} K_1(\alpha\lvert y\rvert)$
	$\alpha = \theta w$	$\alpha = \sqrt{1/w\sigma^2 + \theta^2/\sigma^4}$
	$\beta = \sqrt{\sigma^2 w}$	$\beta = \theta/\sigma^2$
	$\lambda = 1/w$	$\lambda = \sigma$
$\Psi^J(u)$	$-\lambda \ln\left(1 - i\alpha u + \frac{1}{2}\beta^2 u^2\right)$	$\lambda\left[\sqrt{\alpha^2 - \beta^2} - \sqrt{\alpha^2 - (\beta + iu)^2}\right]$
$\mathbb{E}[L_1^J]$	θ	θ

This table summaries the VG and NIG processes. The 'Parameters' row shows the parameters in the subordinating processes. In order to simplify the models, we introduce new parameters, α, β, and λ. The characteristic exponent of a Lévy process is denoted by $\Psi^J(u)$. In the NIG process column, K_1 is the Bessel function of the second kind. For the formula of K_1, refer to (Cont and Tankov, 2004, P.499).

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 and Stein, 1991; Heston, 1993; Duffie, Pan, and Singleton, 2000; Park, 2016, etc.). For example, Heston (1993) uses the CIR process to model the instantaneous volatility.

In this paper, we directly give the VIX model with stochastic volatility and time-varying mean in term of log price, and allow the model to have the central tendency property. We assume that the dynamics of volatility $\{v_t\}_{t\geq 0}$ and mean $\{m_t\}_{t\geq 0}$ follow the CIR process and the mean-reverting process respectively. Let $\{x_t\}_{t\geq 0}$ 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 capture the influence of infinite-activity Lévy jumps on the price change of VIX derivatives, we add this type of jump processes into our model. We specify the dynamics of the log VIX under the risk-neutral

measure $\mathbb{Q}$ as follows

$$\begin{aligned} dx_t &= \kappa(m_t - x_t)dt + \rho\sqrt{v_t}dW_t^v + \sqrt{(1-\rho^2)v_t}dW_t^x + dL_t^J, \\ dm_t &= \kappa_m(\theta_m - m_t)dt + \sigma_m dW_t^m, \\ dv_t &= \kappa_v(\theta_v - v_t)dt + \sigma_v\sqrt{v_t}dW_t^v, \end{aligned} \tag{2.7}$$

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 dL_t^J denotes the Lévy-type jump which is specified in Formula (2.4) in the Section 2.2. The jump compensator, adt, allows the jump part to be a martingale (Madan and Milne, 1991) where $a = \mathbb{E}^{\mathbb{Q}}[L_1^J]$. Additionally, in this model, the correlation coefficient, ρ, can control the leverage effect (Duffie et al., 2000; Kiesel and Rahe, 2017; Du and Luo, 2017). The constants κ, κ_m and κ_v accommodate the speed of reverting to their long-term means respectively.

In this paper, the Lévy jump processes are chosen as either the VG process or the NIG process. The model with VG process is referred as OU-VG and the model with NIG process is referred as OU-NIG.

2.4 Characteristic function

A tractable pricing formula of the VIX derivatives can be derived from the characteristic function of the log VIX by using the inverse Fourier transform (for more details, see Heston, 1993; Carr and Madan, 1999). Our characteristic function can be obtained by solving a system of ordinary differential equations which is derived from the two-factor model (2.7) and the results are summarized in Proposition 1.

Proposition 1 *Under (2.7), we assume that the characteristic function $f(t, T, x_t, m_t, v_t; u)$ of $\{x_t\}_{t\geq 0}$ has an exponential-affine form and then f can be given by*

$$f(t, T, x_t, m_t, v_t; u) = \mathbb{E}^{\mathbb{Q}}[e^{iux_T}|\mathcal{F}_t] = e^{\psi_0(\tau)+\psi_x(\tau)x_t+\psi_m(\tau)m_t+\psi_v(\tau)v_t}, \tag{2.8}$$

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

$$\begin{cases} \psi_0'(\tau) = \kappa_m\theta_m\psi_m(\tau) + \kappa_v\theta_v\psi_v(\tau) + \frac{1}{2}\sigma_m^2\psi_m^2(\tau) - a\psi_x(\tau) + \Psi^J(-i\psi_x(\tau)), \\ \psi_v'(\tau) = -\kappa_v\psi_v(\tau) + \rho\sigma_v\psi_x(\tau)\psi_v(\tau) + \frac{1}{2}\psi_x^2(\tau) + \frac{1}{2}\sigma_v^2\psi_v^2(\tau), \end{cases} \tag{2.9}$$

where $\psi_0(0) = 0$, $\psi_v(0) = 0$, *a and* $\Psi^J(x)$ *are specified in Table 2; and* $\psi_x(\tau)$ *and* $\psi_m(\tau)$ *have the closed-form expressions*

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

Proof. See Appendix A.

Generally, the system (2.8) in Proposition 1 cannot be solved analytically, but it can be solved by 4th-order Runge–Kutta method (RK4) (for RK4 method, see Butcher (2016), and for the application in financial research, see Kiesel and Rahe (2017)).

3. VIX derivative pricing

In this section, we derive the formulae of VIX options and VIX futures by using the characteristic function of the log VIX in Subsection 2.4, and the risk-neutral pricing approach is applied.

3.1 Futures valuation

The price in a VIX futures contract is the expectation of VIX index price at specified trading time in the future (CBOE, 2018). 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

$$F(t; T_j, \Theta) = \mathbb{E}^{\mathbb{Q}}[VIX_{T_j}|\mathcal{F}_t], \tag{3.1}$$

where VIX_{T_j} is the VIX index price at time T_i and $VIX_{T_j} = e^{x_{T_j}}$. According to Proposition 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 of a future contract in term of the characteristic function, that is,

$$F(t;T_i,\Theta) = f(t,T_j,x_t,m_t,v_t;-i). \tag{3.2}$$

3.2 Options valuation

VIX options are the European-style options. We assume that the underlying asset price of a VIX option is VIX_{T_j} at maturity T_j and strike price is K_j. The price of 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 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.[4] Then we have

$$\begin{aligned} C(t;T_j,K_j,\Theta) &= e^{-r_t(T_j-t)}\mathbb{E}^{\mathbb{Q}}[(VIX_{T_j}-K_j)\mathbf{1}_{VIX_{T_j}>K_j}|\mathcal{F}_t] \\ &= e^{-r_t(T_j-t)}\left[\mathbb{E}^{\mathbb{Q}}[VIX_{T_j}\mathbf{1}_{VIX_{T_j}>K_j}|\mathcal{F}_t] - \mathbb{E}^{\mathbb{Q}}[K_j\mathbf{1}_{VIX_{T_j}>K_j}|\mathcal{F}_t]\right], \end{aligned} \tag{3.3}$$

and

$$\begin{aligned} P(t;T_j,K_j,\Theta) &= e^{-r_t(T_j-t)}\mathbb{E}^{\mathbb{Q}}[(K_j-VIX_{T_j})\mathbf{1}_{K_j>VIX_{T_j}}|\mathcal{F}_t] \\ &= e^{-r_t(T_j-t)}\left[\mathbb{E}^{\mathbb{Q}}[K_j\mathbf{1}_{K_j>VIX_{T_j}}|\mathcal{F}_t] - \mathbb{E}^{\mathbb{Q}}[VIX_{T_j}\mathbf{1}_{K_j>VIX_{T_j}}|\mathcal{F}_t]\right]. \end{aligned} \tag{3.4}$$

In order to reduce the complexity of calculation following by Garman and Kohlhagen (1983), we introduce new measure $\mathbb{Q}_1$ with VIX_t as numeraire for the first term in (3.3) and the second term in (3.4). Based on Girsanov's theorem, we can get the Radon-Nikodym derivative $\frac{d\mathbb{Q}_1}{d\mathbb{Q}}$,

$$\frac{d\mathbb{Q}_1}{d\mathbb{Q}} = \frac{VIX_{T_j}}{e^{-r_t(T_j-t)}VIX_t} = \frac{e^{x_{T_j}}}{\mathbb{E}^{\mathbb{Q}}[e^{x_{T_j}}|\mathcal{F}_t]}, \tag{3.5}$$

where the expectation of $\frac{d\mathbb{Q}_1}{d\mathbb{Q}}$ under $\mathbb{Q}$ always equals to 1. Then the new characteristic function, $f_1(t,T_j,x_t,m_t,v_t;u)$, under the new measure $\mathbb{Q}_1$ is

$$f_1(t,T_j,x_t,m_t,v_t;u) = \frac{f(t,T_j,x_t,m_t,v_t;u-i)}{f(t,T_j,x_t,m_t,v_t;-i)}. \tag{3.6}$$

[4]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.

In line with Bakshi and Madan (2000), the price formulae of a call option and a put option, which can be derived in a standard process, are expressed as

$$\begin{aligned} C(t;T_j,K_j,\Theta) &= e^{-r_t(T_j-t)}\left[f(t,T_j,x_t,m_t,v_t;-i)\Phi_1(t;T_j,K_j,\Theta) - K\Phi_2(t;T_j,K_j,\Theta)\right], \\ P(t;T_j,K_j,\Theta) &= e^{-r_t(T_j-t)}[K(1-\Phi_2(t;T_j,K_j,\Theta)) \\ &\qquad - f(t,T_j,x_t,m_t,v_t;-i)(1-\Phi_1(t;T_j,K_j,\Theta))], \end{aligned} \tag{3.7}$$

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

$$\begin{aligned} \Phi_1(t;T_j,K_j,\Theta) &= \frac{1}{2} + \frac{1}{\pi}\int_0^\infty \mathrm{Re}\left[\frac{\mathrm{e}^{-\mathrm{iu}\ln \mathrm{K_j}}\mathrm{f_1}(\mathrm{v_t},\mu_\mathrm{t},\mathrm{w_t},\mathrm{t},\mathrm{T_j};\phi)}{\mathrm{i}\phi}\right]\mathrm{d}\phi, \\ \Phi_2(t;T_j,K_j,\Theta) &= \frac{1}{2} + \frac{1}{\pi}\int_0^\infty \mathrm{Re}\left[\frac{\mathrm{e}^{-\mathrm{iu}\ln \mathrm{K_j}}\mathrm{f}(\mathrm{v_t},\mu_\mathrm{t},\mathrm{w_t},\mathrm{t},\mathrm{T_j};\phi)}{\mathrm{i}\phi}\right]\mathrm{d}\phi. \end{aligned} \tag{3.8}$$

For (3.8), we use the Gauss-Laguerre quadrature method with an order of 20 to calculate the numerical integration.

4. Data selection and model calibration

In this section, we outline the estimation procedure for our two-factor models, and the best two-factor model (OU-VJ) in Mencía-Sentana's paper (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 into the their best model. Also, they specify the jump process in the volatility process. That is

$$\begin{aligned} dx_t &= \kappa(m_t - x_t)dt + \sqrt{v_t}dW_t^x, \\ dm_t &= \kappa_m(\theta_m - m_t)dt + \sigma_m dW_t^m, \\ dv_t &= \lambda_J v_t dt + dJ_t, \end{aligned} \tag{4.1}$$

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 average δ_J. There is no correlation between the jump process and the Wiener processes.

4.1 Sample construction and statistics summary

We calibrate our model parameters by using the joint data of VIX, VIX options and VIX futures. The full sample includes the period between July 2006 and April 2016, with a total of 509 weeks. The data is 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. By following Du and Luo (2017) and Yang and Kanniainen (2017), we only use each Wednesday data. We filter out all anomalous data which are the entries that does 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 the 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 total 12,921 call options and 5,249 put options.

We separate the full sample into two sub-samples. The first sub-sample is in-sample from July 2006 to January 2013 (total 342 weeks) and the second sub-sample is out-of-sample from February 2013 to April 2016 (total 167 weeks). Table 3 describes the statistics summary for VIX options and futures according to this split. Overall, the total number of futures and option contracts in the in-sample is greater than that in the out-of-sample.

In Table 3, the futures panel summaries the number of futures and average futures prices in the different maturity groups. The number of futures 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 options, average options prices, average Black-Scholes implied volatility and average Black-Scholes vega by time to maturity and moneyness. In term of the moneyness, the average option prices vary from 0.9470 to 2.7842 for the in-sample data and from 0.7063 to 2.1409 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 increase. We also observe that the average prices and average vega of call options are less than put options, but the average implied volatility of call options is higher than that of put options in both samples.

4.2 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 filtering 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 unscented Kalman filter (UKF) on our latent state variables and combining with quasi-maximum log-likelihood estimate method (QMLE).

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 and Andrews, 2015). The UKF is one of the most popular estimation technique for the processes with continuous-time state space and can deal with non-linear stochastic systems (e.g. Julier and Uhlmann, 1997; Wan and Merwe, 2000; Särkkä, 2007; Kiesel and Rahe, 2017, etc.). The core of the UKF is the unscented transformation (known as deterministic sampling method) which is proposed by Julier and 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 transformation in each step

Table 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 $\leq$ 0.9	319	1066	479	1864	129	559	442	1130
0.9 $<$ K/F $\leq$ 0.98	278	501	206	985	172	409	223	804
0.98 $<$ K/F $\leq$ 1.02	125	260	132	517	75	176	158	409
1.02 $<$ K/F $\leq$ 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 $\leq$ 0.9	0.6531	0.9155	1.2129	0.9470	0.4339	0.6340	0.8771	0.7063
0.9 $<$ K/F $\leq$ 0.98	1.2836	2.0662	2.8022	1.9992	0.8217	1.4086	1.9899	1.4443
0.98 $<$ K/F $\leq$ 1.02	1.8282	2.7871	3.6839	2.7842	1.3507	2.0726	2.5921	2.1409
1.02 $<$ K/F $\leq$ 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 $\leq$ 0.9	0.8815	0.7115	0.5898	0.7093	0.8527	0.6654	0.5516	0.6423
0.9 $<$ K/F $\leq$ 0.98	0.8529	0.7560	0.6344	0.7579	0.8130	0.7188	0.6145	0.7100
0.98 $<$ K/F $\leq$ 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 $\leq$ 0.9	1.6533	2.6483	3.9989	2.8251	1.1249	1.9320	3.0040	2.2592
0.9 $<$ K/F $\leq$ 0.98	2.0211	3.4831	5.2754	3.4453	1.3835	2.4761	3.8678	2.6284
0.98 $<$ K/F $\leq$ 1.02	2.0803	3.7292	5.7809	3.8544	1.5279	2.7150	4.1311	3.0444
1.02 $<$ K/F $\leq$ 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 provided by Bloomberg. The option prices, implied volatility and vega are provided by 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) and the future price (F). Under the options panel, "All" refers to all call options and put options.

because the two steps are independent. Based on the characteristics of our models, we use the linear transformation in the prediction step and unscented transformation in the updating step. We give the details of applying the UKF and its performance in Appendix B.

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 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 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

$$\begin{aligned}\mathbf{X}_t = \begin{pmatrix} \kappa_m\theta_m\Delta t \\ \kappa_v\theta_v\Delta t \end{pmatrix} + \begin{pmatrix} 1+(\eta_m-\kappa_m)\Delta t & 0 \\ 0 & 1+(\eta_v-\kappa_v)\Delta t \end{pmatrix} \mathbf{X}_{t-1} \\ + \begin{pmatrix} \sigma_m\sqrt{\Delta t} & 0 \\ 0 & \sigma_v\sqrt{v_{t-1}\Delta t} \end{pmatrix} \mathbf{e}_t,\end{aligned} \tag{4.2}$$

where $\mathbf{e}_t = [e_t^m, e_t^v]$, and $\{e_t^m\}_{t\geq 0}$ and $\{e_t^m\}_{t\geq 0}$ follow different standard normal distributions.

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

$$\begin{aligned}\mathbf{X}_t = \begin{pmatrix} \kappa_m\theta_m\Delta t \\ \eta_\lambda\lambda_J(\delta_J+\eta_\delta)\Delta t \end{pmatrix} + \begin{pmatrix} 1+(\eta_m-\kappa_m)\Delta t & 0 \\ 0 & 1-\eta_\lambda\lambda_J\Delta t \end{pmatrix} \mathbf{X}_{t-1} \\ + \begin{pmatrix} \sigma_m\sqrt{\Delta t} & 0 \\ 0 & 2(\delta_J+\eta_\delta)\sqrt{2\eta_\lambda\lambda_J\Delta t} \end{pmatrix} \mathbf{e}_t,\end{aligned} \tag{4.3}$$

where $\mathbf{e}_t$ has the same specification as in Equation (4.2).

Secondly, the measurement errors between our estimated prices and market prices are considered into the procedure. Then, we let the measurement function be,

$$\mathbf{Y}_t = \mathbf{f}(\mathbf{X_t}; x_t, \Theta) + \epsilon_t, \tag{4.4}$$

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

$$\Sigma_t^\epsilon = \begin{bmatrix} \sigma_\epsilon^2 & 0 & 0 & \dots & 0 \\ 0 & \sigma_\epsilon^2 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \sigma_\epsilon^2 \end{bmatrix}, \tag{4.5}$$

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

Finally, by following Trolle and Schwartz (2009), we construct the daily log-likelihood function of VIX derivatives at time t as

$$l_t(\Theta) = -\frac{1}{2}\left[k_t \ln 2\pi + \ln|\Sigma_t^{\mathbf{Y}}| + \epsilon_t^\top (\Sigma_t^{\mathbf{Y}})^{-1}\epsilon_t\right], \tag{4.6}$$

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

$$l(\Theta) = \sum_{t=1}^{T} l_t(\Theta), \tag{4.7}$$

where T is the total number of weeks. Our parameters are estimated by minimizing negative log-likelihood ($-l(\Theta)$). In optimization, we employ quasi-Newton algorithm and the Hessian matrix updated by Broyden-Fletcher-Goldfarb-Shanno (BFGS) method (see Broyden, 1965; Fletcher, 1987; Lewis and Overton, 2013).

4.3 Results of parameter estimation

Table 4 presents the estimated results for OU-VJ model, and our two models with infinite jumps, OU-VG and OU-NIG models. We notice that the standard errors for all parameters

are relatively small, except the parameters, η_m and η_v. It indicates that the most parameters are accurately estimated. Please note that η_m and η_v are not our key parameters, so it does not matter that they have slightly high standard errors. In the following, we will analyze and compare our estimates.

The parameters κ and κ_m measure the reverting speed towards their means. Intuitively, the smaller value of reverting speed parameter suggests that the process is more persistent. The values of κ vary between 7.5579 and 7.8276 across the three models, which did not exhibit the 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 0.3412, 0.4355 and 0.3940 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 is 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.

Then, we analyze the estimation results of our two infinite-activity jump models. Our estimates of variance risk premium (η_v) in these two models are negative values which consistent with the previous literature (Park, 2015). The estimated value of η_v for OU-VG model is -2.1331 which is very close to the η_v value (-2.1267) for 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 there is a positive interaction between 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.

Table 4: 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 in Section 5.3. 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.

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 of VIX derivatives and how the different jump structures affect both in-sample and out-of-sample model performance. To answer these questions , we will compare the performance of the models based on comparison criteria, root mean squared errors and log-likelihood values.

5.1 Criteria of performance

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

$$SE^F(t,i;\Theta) = \left(\ln \hat{F}(t;T_i,\Theta) - \ln F(t;T_i)\right)^2 \tag{5.1}$$

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 RMSEs for the futures is calculated as

$$RMSEs^F(\Theta) = \sqrt{\frac{1}{N^F}\sum_{t=1}^{N_T}\sum_{i=1}^{N_t^F} SE^F(t,i;\Theta)}, \tag{5.2}$$

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

By 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

$$SE^O(t,i;\Theta) = \left(\frac{\hat{O}(t;T_i,K_i,\Theta) - O(t;T_i,K_i)}{\mathcal{V}(t;T_i,K_i)}\right)^2, \tag{5.3}$$

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. Then, the RMSEs for call and put options is calculated as

$$RMSEs^O(\Theta) = \sqrt{\frac{1}{N^O}\sum_{t=1}^{N_T}\sum_{i=1}^{N_t^O} SE^O(t,i;\Theta)}, \tag{5.4}$$

where N_T is the same as that in Equation (5.2), 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 Equation (4.6). We let $u^{i,j}$ be the expectation of $d_t^{i,j}$, which is

$$u^{i,j} = \mathbb{E}[d_t^{i,j}] = \frac{\sum_{t=1}^{T} d_t^{i,j}}{T}, \tag{5.5}$$

where T is total number of weeks in the sample. It measures the performance loss of Model i comparing 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.

5.2 Model performance comparison

In this section, we compare the overall performance of OU-VG, OU-NIG and OU-VJ model in terms of log-likelihood values. Table 5 reports the model performance ranking based 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) are significantly better than that of the OU-VJ model because the both 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 (5391.84 for in-sample and 5120.83 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 conclusions same as that from the Panel A. In contrast to OU-VG and OU-NIG models, the OU-VJ model has performance loss more than 20 during the in-sample period and more than 40 during the out-of-sample period. This implies that infinite-activity jump models are more reliable than finite-activity jump model in forecasting. Furthermore, the infinite-activity jump model with an infinite variation (OU-NIG) is slightly superior to the model with a finite variation (OU-VG).

5.3 Fit to VIX futures and options pricing

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

Figure 1 illustrates the empirical CDFs of the square futures pricing errors. In this CDF graph, the OU-VJ curve (blue) is below the OU-NIG curve (yellow) and the OU-VG curve

Table 5: 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 4. The in-sample and out-of-sample periods are from July 2006 to January 2013 and from February 2013 to April 2016, respectively. In Panel A, the total log-likelihood values are evaluated by Equation (4.7). Panel B presents the performance loss (evaluated by Equation (5.5)) of models in pairwise comparison. In this panel, the models listed in the first column are in the set of Model i. Correspondingly, the models listed in second row in this table are in the set of Model j. We do not compare the model with itself and "—" represents no value in the cell.

(red). Also, the yellow curve is slightly above the red curve. This indicates that infinite-activity jump models significantly outperform the OU-VJ model in the futures pricing.

Table 6 compares the model performance in terms of RMSEs. Overall, the OU-VJ model has the largest in-sample RMSEs (0.0318) and out-of-sample RMSEs (0.0474) and are higher than those of the OU-VG model and the OU-NIG model. The differences between OU-VG and OU-NIG are insignificant, which are 0.0001 for the in-sample and 0.0002 for the out-of-sample.

Panel B presents the RMSEs 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($\tau > 90$). To compare with OU-VJ model, we observe that our jump models take more advantage on pricing middle-term futures than other term futures in both in-sample and out-of-sample because our models improve the RMSEs over 24% in-sample and over 50% out-of-sample. 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.

Figure 2 compares the difference of the half-yearly average SEs for the paired models from 2006 to 2016 (full-sample period). From the subplot a and b, we find that the half-yearly average SEs of OU-VJ are significantly larger than those 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

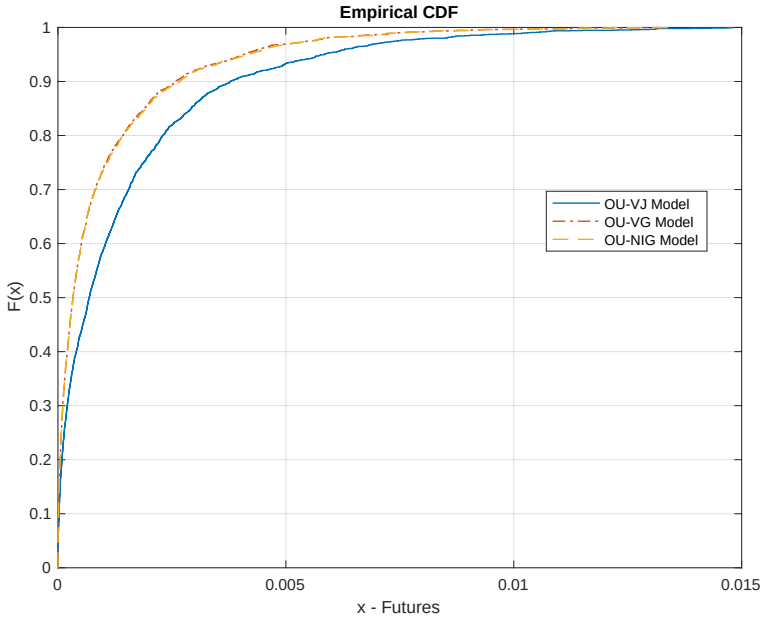

Figure 1: Empirical cumulative distribution function (CDF) of the square pricing errors for futures. The square pricing errors are calculated from Equation (5.1) by using full sample data from July 2006 to April 2016, based on estimated parameters in Table (4).

to that in the subplot (b). This is because the difference between square pricing errors of OU-VG model and square pricing errors of OU-NIG model is incredibly tiny, which we can observe 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 the option pricing.

Table 7 reports the RMSEs under different categories: time to maturity, moneyness and volume. Overall, we can find that the RMSEs of the OU-VJ model are 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 both call and put option pricing. 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 RMSEs

Table 6: Model performance comparison in future pricing – RMSEs

	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 RMSEs of future prices. The in-sample and out-of-sample period from July 2006 to January 2013 and from February 2013 to April 2016 respectively. The RMSEs is computed from the Equation 5.2 based on the estimated parameters in Table (4). In the Panel B, the time to maturity is denote as τ.

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.

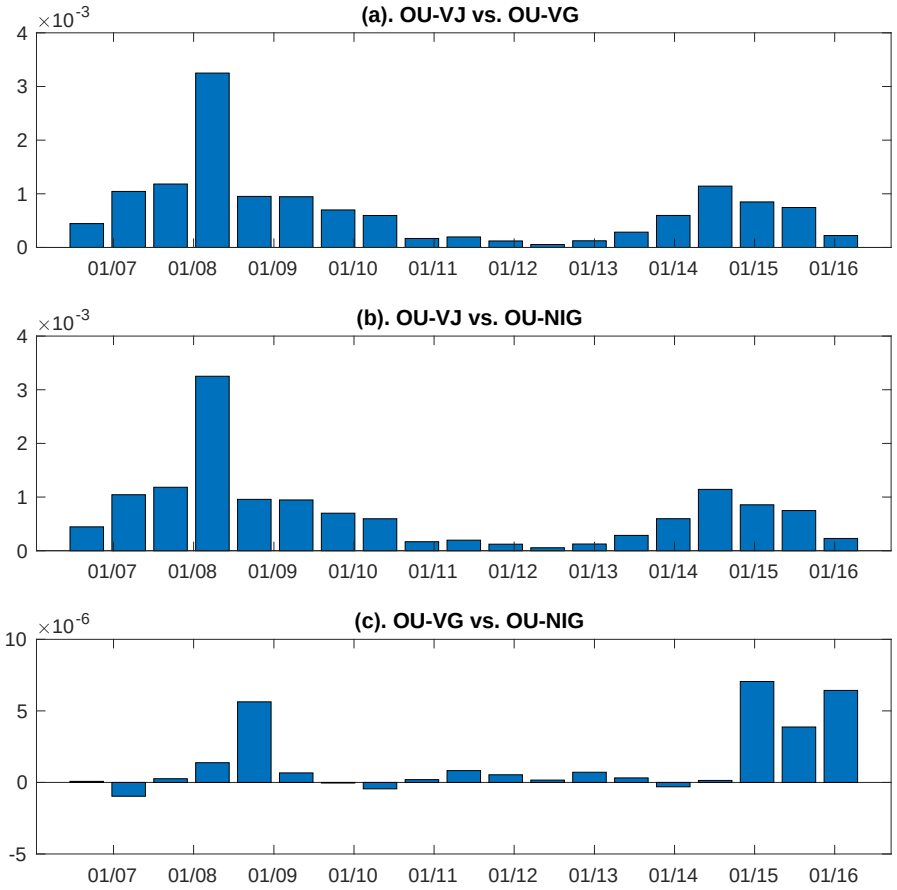

Figure 2: The comparison of half-yearly average SEs during the period between 2006 and 2016 (full-sample period). The y-axis represents the differences between the average SEs of the Model A and the average SEs of Model B. For example, in the subplot a, the magnitude of the each bar equals to average SE of OU-VJ model subtracts the average SE of OU-VG model. The x-axis represents the sample period. The squared pricing errors are calculated by Equation (5.3) based on estimated parameters in Table (4).

5.4 Finite- vs. infinite-activity jumps

To further investigate the evidence of infinite-activity jumps in the VIX derivatives pricing, 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 as a compound Poisson process (CP), against our two models. This finite-activity jump model is the best model in Park (2016), in which the jump part in

Table 7: Model performance comparison in option pricing – RMSEs

	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			
$K/F \leq 0.9$	0.1833	0.0639	0.0632	0.2127	0.0890	0.0908
$0.9 < K/F \leq 0.98$	0.1382	0.0798	0.0796	0.1708	0.0982	0.0969
$0.98 < K/F \leq 1.02$	0.1620	0.0914	0.0915	0.1944	0.0901	0.0898
$1.02 < K/F \leq 1.1$	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			
$Vol \leq 500$	0.1501	0.0618	0.0617	0.1601	0.0653	0.0648
$500 < Vol \leq 3000$	0.1430	0.0652	0.0647	0.1596	0.0728	0.0715
$3000 < Vol \leq 8000$	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 average RMSEs of option prices in terms of time-to-maturity (Panel B), moneyness (Panel C), and volume (Panel D). The in-sample and out-of-sample period from July 2006 to January 2013 and from February 2013 to April 2016 respectively. The RMSEs is computed from the Equation 5.4 based on the estimated parameters in Table (4). In the Panel B, the time to maturity is denote as τ.

(2.7) 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 Wiener processes. Their jump sizes follow different exponential distributions. But, the positive jumps have positive jump size with mean δ_+. The negative jumps have negative jump size with mean δ_-.

We apply the same model estimation procedure to calibrate the OU-CP model. The estimation results are reported in Table 8. In this table, we find that OU-CP has positive ρ and negative η_v. This factor is consistent with the economic phenomenon discussed in Section 4.3. In comparing with the results in Table 4, we find that the values of κ, κ_m, κ_v and σ_v in OU-CP are smaller than those in our models. Therefore, the infinite-activity jump structure makes model more flexible and the variance process of the model more fluctuate.

Table 8: 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 5.3. 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 log-likelihood and RMSEs. Firstly, to compare with log-likelihood of our infinite-activity jump models in Table 5, the log-likelihood of the OU-CP, in Table 9, drops by at least 200 in-sample from 5377.20 (OU-VG) or 5391.84 (OU-NIG) and at least 350 out-of-sample from 5046.01 (OU-VG) or 5120.83 (OU-NIG).

From Table 9 and Table 6, we can not notice the 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 substantially improve the overall model performance in VIX future pricing. 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 7 and Table 9, we find that VG jump structure reduces the RMSEs of VIX options by 0.0012 from 0.0674 (OU-CP) to 0.0662 (OU-VG) and NIG jump structure reduces RMSEs by 0.0016 from 0.0674 (OU-CP) to 0.0658 (OU-NIG) during the in-sample period. For the out-of-sample, the RMSEs 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 RMSEs of the whole sample. The p-value of t-test is 0.0146 for OU-CP and OU-VG and 0.0001 for the OU-CP and OU-NIG. Both p-values are less than 0.0500, so we can conclude that VG and NIG jump structures make statistically significant improvements in the pricing of VIX options. Particularly, the infinite-activity jump structures noticeably improve the pricing accuracy of put options.

Table 9: 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		
RMSEs	$\tau \leq 30$	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		
	$K/F \leq 0.9$	0.0669		0.0915	
	$0.9 < K/F \leq 0.98$	0.0844		0.0978	
	$0.98 < K/F \leq 1.02$	0.0934		0.0949	
	$1.02 < K/F \leq 1.1$	0.0821		0.0669	
	$K/F > 1.1$	0.0575		0.0647	

This table reports the OU-CP performance in terms of log-likelihood and RMSEs. 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 Equation (4.7) and the RMSEs is computed from the Equation (5.4) based on the estimated parameters in Table (4). The time to maturity is denote as τ.

6. Conclusion and discussion

Although substantial progress has been made in developing more realistic VIX models to improve the pricing accuracy of VIX derivatives (Psychoyios and Dotsis, 2010; Mencía and Sentana, 2013; Park, 2016, etc.), it is unknown whether more generalized jump processes (e.g., infinite-activity jump Lévy process) can improve the performance of VIX derivatives pricing. It is interesting to investigate this because this type of processes can describe the phenomenon where jumps frequently occur during a finite time interval. This phenomenon has attracted more and more attentions in the current research. To fill this gap, we introduce two representative infinite-activity pure jump Lévy process, variance gamma (VG) process and normal inverse Gaussian (NIG) process, into our two-factor models.

Then, we investigate what kind of model specification have the best pricing performance of VIX derivatives. To answer this question, firstly, we obtain accurate model calibration by using the joint estimation approach which combines UKF with QMLE on the joint data of VIX and VIX derivatives. Then, we carry out an extensive comparison among several two-factor models to figure out the best model configuration.

The empirical results provide the critical evidence that the infinite-activity jump models (OU-VG and OU-NIG) overall outperform the finite-activity models (OU-VJ and OU-CP) and infinite-activity jumps cannot be ignored in the VIX derivatives pricing. 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, the Lévy-type jump structure 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 structure is highlighted in VIX option pricing but not in the VIX future pricing.

In comparing two different infinite-activity models, we figure out that the OU-NIG model is superior to 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 perform better in pricing VIX derivatives which time to maturity is greater

than 90 days.

Finally, 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) find that the finical asset prices become more volatile on the daily data or intraday data. In the future, we may use high-frequency data to conduct more empirical research on infinite-activity jump models.

A. Appendix: Proof of Proposition 1

Under the given model (2.7) and characteristic function, $f(t,T,x_t,m_t,v_t;u)$, we can get the following equation by applying Ito's lemma.

$$\begin{aligned} df =& \frac{\partial f}{\partial t}dt + \frac{\partial f}{\partial x}dx^c + \frac{\partial f}{\partial m}dm^c + \frac{\partial f}{\partial v}dv^c + \frac{\partial^2 f}{\partial x \partial m}dx^c dm^c + \frac{\partial^2 f}{\partial x \partial v}dx^c dv^c \\ &+ \frac{\partial^2 f}{\partial m \partial v}dm^c dv^c + \frac{1}{2}\left[\frac{\partial^2 f}{\partial x^2}(dx^c)^2 + \frac{\partial^2 f}{\partial m^2}(dm^c)^2 + \frac{\partial^2 f}{\partial v^2}(dv^c)^2\right] \\ &+ \int_{|y|<1}\left[f(t,T,x_{t^-}+y,m_t,v_t;u) - f(t,T,x_{t^-},m_t,v_t;u) - y\frac{\partial f}{\partial x}\right]\nu(dy)dt \\ &+ \int_{\mathbb{R}}\left[f(t,T,x_{t^-}+y,m_t,v_t;u) - f(t,T,x_{t^-},m_t,v_t;u)\right]\mu^L(dy,dt), \end{aligned} \tag{A.1}$$

where c represents the continuous part in a process. We only allow the correlation between the process $\{x_t\}_{t\geq 0}$ and the process $\{v_t\}_{t\geq 0}$, so the Equation (A.1) can be computed as

$$\begin{aligned} df =& \left\{\frac{\partial f}{\partial t} + \kappa\left(m-x\right)\frac{\partial f}{\partial x} + \kappa_m(\theta_m - m)\frac{\partial f}{\partial m} + \kappa_v(\theta_v - v)\frac{\partial f}{\partial v} + \rho\sigma_v v\frac{\partial^2 f}{\partial x \partial v}\right. \\ &\left. + \frac{1}{2}\left[v\frac{\partial^2 f}{\partial x^2} + \sigma_m^2\frac{\partial^2 f}{\partial m^2} + \sigma_v^2 v\frac{\partial^2 f}{\partial v^2}\right]\right\}dt + \sqrt{v}\frac{\partial f}{\partial x}dW_t^x + \sigma_m\frac{\partial f}{\partial m}dW_t^m + \sigma_v\sqrt{v}\frac{\partial f}{\partial v}dW_t^v \\ &+ \int_{|y|<1}\left[f(t,T,x_{t^-}+y,m_t,v_t;u) - f(t,T,x_{t^-},m_t,v_t;u) - y\frac{\partial f}{\partial x}\right]\nu(dy)dt \\ &+ \int_{\mathbb{R}}\left[f(t,T,x_{t^-}+y,m_t,v_t;u) - f(t,T,x_{t^-},m_t,v_t;u)\right]\mu^L(dy,dt). \end{aligned} \tag{A.2}$$

We assume that the characteristic function has an exponential-affine form, as mentioned in Equation (2.8). We have

$$
\begin{aligned}
\frac{df}{f} =& \Big[-\psi_0^{'}(\tau) - x\psi_x^{'}(\tau) - m\psi_m^{'}(\tau) - v\psi_v^{'}(\tau) + \kappa(m-x)\psi_x(\tau) \\
&+ \kappa_m(\theta_m - m)\psi_m(\tau) + \kappa_v(\theta_v - v)\psi_v(\tau) + \rho\sigma_v v\psi_x(\tau)\psi_v(\tau) \\
&+ \frac{1}{2}\left(v\psi_x^2(\tau) + \sigma_m^2\psi_m^2(\tau) + \sigma_v^2 v\psi_v^2(\tau)\right) + \int_{|y|<1}\left(e^{y\psi_x(\tau)} - 1 - y\psi_x(\tau)\right)\nu(y)\Big]dt \\
&+ \left(\sqrt{v}\psi_x(\tau)dW_t^x + \sigma_m\psi_m(\tau)dW_t^m + \sigma_v\psi_v(\tau)dW_t^v\right) \\
&+ \int_{\mathbb{R}}\left(e^{y\psi_x(\tau)} - 1\right)\mu^L(dy, dt).
\end{aligned} \tag{A.3}
$$

By taking expectation for the both sides of Equation (A.3), we have,

$$
\begin{aligned}
\mathbb{E}_t^{\mathbb{Q}}\left[\frac{df}{f}\right] =& \Big[-\psi_0^{'}(\tau) - x\psi_x^{'}(\tau) - m\psi_m^{'}(\tau) - v\psi_v^{'}(\tau) + \kappa(m-x)\psi_x(\tau) \\
&+ \kappa_m(\theta_m - m)\psi_m(\tau) + \kappa_v(\theta_v - v)\psi_v(\tau) + \rho\sigma_v v\psi_x(\tau)\psi_v(\tau) \\
&+ \frac{1}{2}\left(v\psi_x^2(\tau) + \sigma_m^2\psi_m^2(\tau) + \sigma_v^2 v\psi_v^2(\tau)\right) + \int_{|y|<1}\left(e^{y\psi_x(\tau)} - 1 - y\psi_x(\tau)\right)\nu(y) \\
&+ \int_{\mathbb{R}}\left(e^{y\psi_x(\tau)} - 1\right)\nu(dy)\Big]dt.
\end{aligned} \tag{A.4}
$$

We know that $\mathbb{E}_t^{\mathbb{Q}}[df] = 0$, then we can get the following ordinary differential equations,

$$
\psi_0^{'}(\tau) = \kappa_m\theta_m\psi_m(\tau) + \kappa_v\theta_v\psi_v(\tau) + \frac{1}{2}\sigma_m^2\psi_m(\tau)^2 - a\psi_x(\tau) + \Psi^J(-i\psi_x(\tau)), \tag{A.5}
$$

$$
\psi_x^{'}(\tau) = -\kappa\psi_x(\tau), \tag{A.6}
$$

$$
\psi_m^{'}(\tau) = \kappa\psi_x(\tau) - \kappa_m\psi_m(\tau), \tag{A.7}
$$

$$
\psi_v^{'}(\tau) = -\kappa_v\psi_v(\tau) + \rho\sigma_v\psi_x(\tau)\psi_v(\tau) + \frac{1}{2}\psi_x^2(\tau) + \frac{1}{2}\sigma_v^2\psi_v(\tau)^2, \tag{A.8}
$$

where $\Psi^J(-i\psi_x(\tau)) = a\psi_x(\tau) + \int_{\mathbb{R}}\left(e^{y\psi_x(\tau)} - 1 - \psi_x(\tau)y\mathbf{1}_{|y|<1}\right)\nu(dy)$, and $a = \alpha\lambda$ for the OU-VG model or $a = \beta\lambda^2$ for the OU-NIG model. Applying the initial conditions, $\psi_x(0) = iu$ and $\psi_m(0) = 0$, the Equation (A.6) and Equation (A.7) can be solved analytically.

B. Appendix: UKF method and test

We use UKF to estimate the latent state vector ($\mathbf{X}_t$) with the non-linear measurement equations. The models of the latent vectors under physical measure see Equation (4.2) for OU-CP, OU-VG, and OU-NIG and Equation (4.3) for OU-VJ. In the prediction step, we apply the linear forecast. Based on the information up to time $t-1$, we denote that the predicted state mean as $\hat{\mathbf{X}}_{t|t-1}$ and predicted state covariance as $\hat{\Sigma}^{\mathbf{X}}_{t|t-1}$ (For detailed discussions on linear forecast, 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

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

with the weights

$$W_0^{(m)} = \frac{\lambda}{n+\lambda}, \tag{B.2}$$

$$W_0^{(\Sigma)} = \frac{\lambda}{n+\lambda} + \left(1-\alpha^2+\beta\right), \tag{B.3}$$

$$W_i^{(m)} = W_i^{(\Sigma)} = \frac{1}{2\,(n+\lambda)}, \qquad i = 1,\ldots,2n, \tag{B.4}$$

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 chose α to be small positive number 10^{-3}.

Then, the predicted mean of the measurements (measurement function see Equation (4.4)), variance of the measurements and covariance of the latent variables and measurements

can be computed by using the sigma points (B.1) as,

$$\hat{\mathbf{Y}}_{t|t-1} = \sum_{i=0}^{2n} W_i^{(m)} \mathbf{f}(\mathscr{S}_t^i; x_t, \Theta), \tag{B.5}$$

$$\hat{\Sigma}^{\mathbf{Y}}_{t|t-1} = \sum_{i=0}^{2n} W_i^{(\Sigma)} \left(\mathbf{f}(\mathscr{S}_t^i; x_t, \Theta) - \hat{\mathbf{Y}}_{t|t-1}\right) \left(\mathbf{f}(\mathscr{S}_t^i; x_t, \Theta) - \hat{\mathbf{Y}}_{t|t-1}\right)^\top + \Sigma_t^\epsilon, \tag{B.6}$$

$$\hat{\Sigma}^{\mathbf{X},\mathbf{Y}}_{t|t-1} = \sum_{i=0}^{2n} W_i^{(\Sigma)} \left(\mathscr{S}_t^i - \hat{\mathbf{X}}_{t|t-1}\right) \left(\mathbf{f}(\mathscr{S}_t^i; x_t, \Theta) - \hat{\mathbf{Y}}_{t|t-1}\right)^\top. \tag{B.7}$$

Finally, we calculate the Kalman filter gain, the state mean and state covaraince conditional on the measurements as,

$$K_t = \hat{\Sigma}^{\mathbf{X},\mathbf{Y}}_{t|t-1} \left(\hat{\Sigma}^{\mathbf{Y}}_{t|t-1}\right)^{-1}, \tag{B.8}$$

$$\hat{\mathbf{X}}_t = \hat{\mathbf{X}}_{t|t-1} + K_t \left(\mathbf{Y}_t - \hat{\mathbf{Y}}_{t|t-1}\right), \tag{B.9}$$

$$\hat{\Sigma}^{\mathbf{X}}_t = \hat{\Sigma}^{\mathbf{X}}_{t|t-1} - K_t \hat{\Sigma}^{\mathbf{Y}}_{t|t-1} K_t^\top. \tag{B.10}$$

Furthermore, we test the performance of UKF on our artificial data. In this test, we perform the UKF on the following 2-factor model without jumps,

$$\begin{aligned} dx_t &= \kappa(m_t - x_t)dt + \rho\sqrt{v_t}dW_t^v + \sqrt{(1-\rho^2)v_t}dW_t^x, \\ dm_t &= \kappa_m(\theta_m - m_t)dt + \sigma_m dW_t^m, \\ dv_t &= \kappa_v(\theta_v - v_t)dt + \sigma_v\sqrt{v_t}dW_t^v. \end{aligned} \tag{B.11}$$

We simulate time series data x_t, m_t and v_t, by using Monte Carlo method under 10,000 simulated paths. The total sample length is 3650 days. In the simulation procedure, we set the values of the model parameters which are shown in Table 10. 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 the Equation (3.2).

Table 10: 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 the Model (B.11) parameters. The last parameter, r, refers to the fixed interest rate, which are used to compute the futures prices.

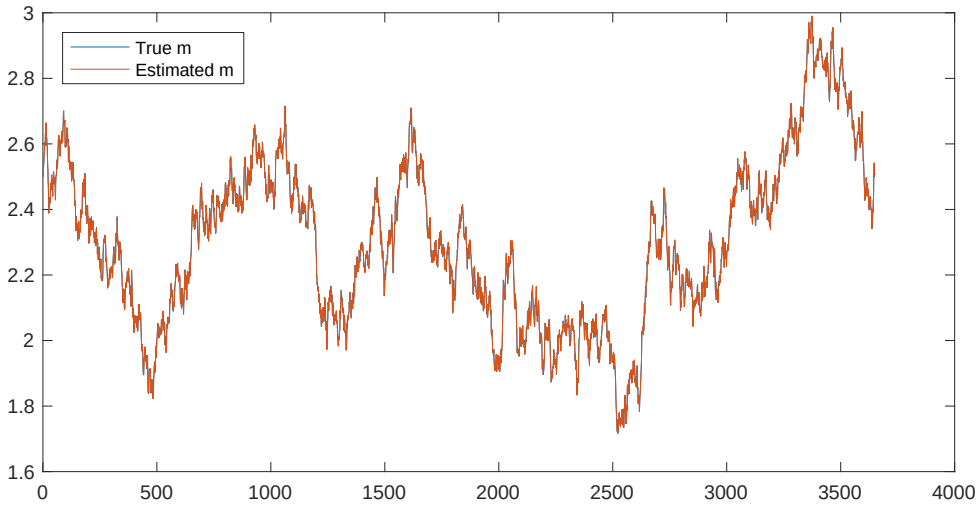

Figure 3: "True" m_t vs. Predicted m_t. This graph compares the "true" m_t with the estimated m_t by UKF. We treat the Monte Carlo simulation results as "true" m_t.

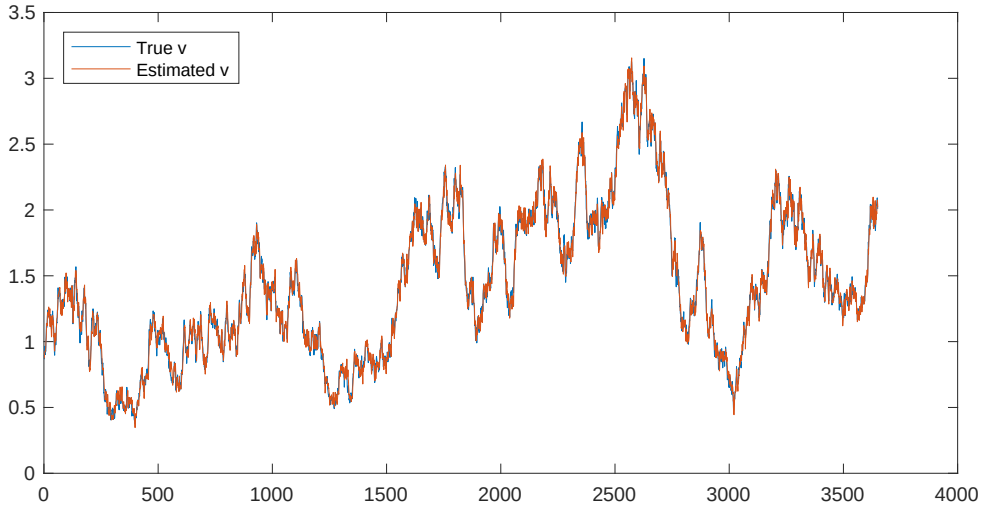

Figure 4: "True" v_t vs. Predicted v_t. This graph compares the "true" v_t with the estimated v_t by UKF. We treat the Monte Carlo simulation results as "true" v_t.

In Figures 3 and 4, we compare the "true" value of latent variables from real data set with the estimated values by UKF. The blue lines and orange lines represent the real values and estimated values respectively. We can observe that the orange line highly overlaps with the blue line in both figures, which means the difference between the real data and the estimated results is insignificant. Therefore, we can conclude that the UKF is appropriate filtering method for estimating the latent variables in this case.

References

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

Bakshi, Gurdip, and Dilip Madan, 2000, Spanning and derivative-security valuation, *Journal of Financial Economics* 55, 205–238.

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

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

Broyden, C.G., 1965, A class of methods for solving nonlinear simultaneous equations, *Mathematics of Computation* 19, 577–593.

Butcher, John Charles, 2016, *Numerical methods for ordinary differential equations* (John Wiley & Sons).

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

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

Carr, P., and D.B. Madan, 1999, Option valuation using the fast Fourier transform, *Journal of Computational Finance* 2, 61–73.

Carr, P., and L. Wu, 2004, Time-changed Lévy processes and option pricing, *Journal of Financial Economics* 71, 113–141.

Carr, Peter, and Liuren Wu, 2003, The finite moment log stable process and option pricing, *Journal of Finance* 58, 753–777.

CBOE, 2018, White paper: CBOE volatility index .

Christoffersen, Peter, Christian Dorion, Kris Jacobs, and Lotfi Karoui, 2014, Nonlinear kalman filtering in affine term structure models, *Management Science* 60, 2248–2268.

Clark, P.K., 1973, A subordinated stochastic process model with finite variance for speculative prices, *Econometrica* 41, 135–155.

Cont, R., and P. Tankov, 2004, *Financial Modelling with Jump Processes* (CRC press).

Du, D., and D. Luo, 2017, The pricing of jump propagation:evidence from spot and options markets, *Management Science, forthcoming* .

Duffie, D., J. Pan, and K. Singleton, 2000, Transform analysis and asset pricing for affine jump-diffusions, *Econometrica* 68, 1343–1376.

Eberlein, E., U. Keller, and K. Prause, 1988, New insights into smile, mispricing, and value at risk: the hyperbolic model, *The Journal of Business* 71, 371–405.

Eberlein, E., and F. Özkan, 2003, Time consistency of Lévy models, *Quantitative Finance* 3, 40–50.

Eraker, B., 2004, Do stock prices and volatility jump? reconciling evidence from spot and option prices, *Journal of Finance* 59, 1367–1403.

Fletcher, R., 1987, *Practical Methods of Optimization*, second edition (Wiley).

Garman, M.B., and S.W. Kohlhagen, 1983, Foreign currency options values, *Journal of International Money and Finance* 2, 231–237.

Geman, H., 2002, Pure jump Lévy processes for asset price modelling, *Journal of Banking & Finance* 26, 1297–1316.

Goard, J., and M. Mazur, 2013, Stochastic volatility models and the pricing of VIX options, *Mathematical Finance* 23, 439–458.

Goutte, S., A. Ismail, and H. Pham, 2017, Regime-switching stochastic volatility model: estimation and calibration to VIX options, *Applied Mathematical Finance* 24, 38–75.

Grewal, Mohinder S., and Angus P. Andrews, 2015, *Kalman Filtering: Theory and Practice Using MATLAB*, fourth edition (John Wiley & Sons, Inc.).

Grünbichler, A., and F.A. Longstaff, 1996, Valuing futures and options on volatility, *Journal of Banking & Finance* 20, 985–1001.

Hansen, P.R., A. Lunde, and J.M. Nason, 2011, The model confidence set, *Econometrica* 79, 453–497.

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

Julier, S.J., and J.K. Uhlmann, 1997, A new extension of the Kalman filter to nonlinear, *Proc. SPIE 3068, Signal Processing, Sensor Fusion, and Target Recognition IV* .

Kaeck, A., and C. Alexander, 2013, Continuous-time VIX dynamics: on the role of stochastic volatility of volatility, *International Review of Financial Analysis* 28, 46–56.

Kaeck, A., P. Rodrigues, and N.J. Seeger, 2018, Model complexity and out-of-sample performance: evidence from S&P 500 Index return, *Journal of Economic Dynamics & Control* 90, 1–29.

Kandepu, R., B. Foss, and L. Imsland, 2008, Applying the unscented Kalman filter for nonlinear state estimation, *Journal of Process Control* 18, 753–768.

Kanniainen, J., B. Lin, and H. Yang, 2014, Estimating and using GARCH models with VIX data for option valuation, *Journal of Banking & Finance* 43, 200–211.

Kiesel, R., and F. Rahe, 2017, Option pricing under time-varying risk-aversion with applications to risk forecasting, *Journal of Banking & Finance* 76, 120–138.

Lewis, A.S., and M.L. Overton, 2013, Nonsmooth optimization via quasi-Newton methods, *Mathematical Programming Series A* 141, 135–163.

Li, Jing, Lingfei Li, and Gongqiu Zhang, 2017, Pure jump models for pricing and hedging vix derivatives, *Journal of Economic Dynamics and Control* 74, 28–55.

Lian, Guanghua, Song-Ping Zhu, Robert J Elliott, and Zhenyu Cui, 2017, Semi-analytical valuation for discrete barrier options under time-dependent lévy processes, *Journal of Banking & Finance* 75, 167–183.

Lin, Y.N., 2007, VIX futures: evidence from integrated physical and risk-neutral probability measures, *Journal of Futures Markets* 27, 1175–1217.

Lin, Y.N., and C.H. Chang, 2009, VIX option pricing, *Journal of Futures Markets* 29, 523–543.

Madan, D.B., P. Carr, and E.C. Chang, 1999, The variance gamma process and option pricing, *European Finance Review* 2, 79–105.

Madan, D.B., and F. Milne, 1991, Option pricing with V.G. martingale components, *Mathematical Finance* 1, 39–55.

Mencía, J., and E. Sentana, 2013, Valuation of VIX derivatives, *Journal of Financial Economics* 108, 367–391.

Øksendal, B., and A. Sulem, 2005, *Applied Stochastic Control of Jump Diffusion*, second edition (Springer).

Pan, J., 2002, The jump-risk premia implicit in options: evidence from an integrated time-series study, *Journal of Financial Economics* 63, 3–50.

Park, H., 2015, Volatility-of-volatility and tail risk hedging returns, *Journal of Financial Markets* 26, 38–63.

Park, Y.H., 2016, The effects of asymmetric volatility and jumps on the pricing of VIX derivatives, *Journal of Econometrics* 192, 313–328.

Psychoyios, D., and G. Dotsis, 2010, A jump diffusion model for VIX volatility options and futures, *Review of Quantitative Finance and Accounting* 35, 245–269.

Särkkä, S., 2007, On unscented Kalman filtering for state estimation of continuous-time nonlinear systems, *IEEE Transactions on Automatic Control* 52, 1631–1641.

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

Stein, E.M., and J.C. Stein, 1991, Stock price distributions with stochastic volatility: an analytic approach, *The Review of Financial Studies* 4, 727–752.

Todorov, Viktor, and George Tauchen, 2011, Volatility jumps, *Journal of Business & Economic Statistics* 29, 356–371.

Trolle, A.B., and E.S. Schwartz, 2009, A general stochastic volatility model for the pricing of interest rate derivatives, *Review of Financial Studies* 22, 2007–2057.

Wan, E.A., and R.V. Merwe, 2000, The unscented Kalman filter for nonlinear estimation, *Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium* 153–158.

Wu, L., 2011, Variance dynamics: joint evidence from options and high-frequency returns, *Journal of Econometrics* 160, 280–287.

Yang, H., and J. Kanniainen, 2017, Jump and volatility dynamics for the S&P 500: evidence for infinite-activity jumps with non-affine volatility dynamics from stock and option markets, *Review of Finance* 21, 811–844.

Zhang, J. E., and Y. Zhu, 2006, VIX futures, *Journal of Futures Markets* 26, 521–531.

Zhu, S.P., and G.H. Lian, 2012, Analytical formula for VIX futures and its applications, *Journal of Futures Markets* 32, 166–190.

Zhu, Y., and J.E. Zhang, 2007, Variance term structure and VIX future pricing, *International Journal of Theoretical and Applied Finance* 10, 111–127.