

Time-Varying Skew in VIX Derivatives Pricing

Peixuan Yuan

Rutgers Business School

Rutgers University*

This version: March, 2020

Abstract

This paper proposes a new reduced-form model for the pricing of VIX derivatives that includes an independent stochastic jump intensity factor and co-jumps in the level and variance of VIX, while allowing the mean of VIX variance to be time-varying. I fit the model to daily prices of futures and European options from April 2007 through December 2017. The empirical results indicate that the model significantly outperforms all other nested models and improves on benchmark by 21.6% in-sample and 31.2% out-of-sample. The model more accurately portrays the tail behavior of VIX risk-neutral distribution for both short and long maturities, as it successfully captures the time-varying skew found to be largely independent of the level of the VIX smile.

Keywords: VIX derivatives; Co-jumps; Jump intensity; Central tendency; Implied volatility surface

JEL classification: G12, G13.

*Piscataway, NJ, 08854. Phone: (609)997-1398. Email: peixuan.yuan@rutgers.edu

1. Introduction

Since their introduction, VIX derivatives have become important for hedging portfolio volatility risk and speculating on the future direction or movement of volatility.¹ Therefore, it is necessary to achieve an accurate valuation of VIX derivatives. Although current models can generate stochastic variance of VIX, they struggle with capturing the tail behavior of VIX risk-neutral distribution, mainly because they fail to allow for independence of the level and slope of the VIX smile.² In this paper, I develop a new reduced-form model that more accurately portrays the underlying dynamics of VIX derivatives and thus delivers the lowest root mean square relative errors (RMSRERs) among all other popular models tested.³

To capture the independence feature of both short- and long-term VIX options, I primarily include two more risk factors, in addition to the level central tendency and stochastic variance factors introduced in Mencía and Sentana (2013).⁴ The first is the jump intensity factor which independently governs the occurrence rate of jumps, while the second is the variance central tendency factor which controls the stochastic mean level of VIX variance. I show that the flexible combination of risk factors with distinct yet complementary impacts on the implied volatility (IV) surface can capture the independence feature. Specifically, the jump intensity factor, together with the stochastic variance factor, can capture the independence feature of short-term options and, together with the variance central tendency factor, describe the independence feature of long-term

¹In 1993, the Chicago Board Options Exchange (CBOE) introduced the VIX index which was originally designed to measure the market's expectation of 30-day volatility implied by at-the-money S&P 100 Index option prices. Ten years later in 2004, it was expanded to use options based on the more popular index, namely, the S&P 500. CBOE launched VIX futures on March 26, 2004, and then European-style options on February 24, 2006. According to the CBOE website, in 2019 VIX futures had an average daily volume of 255,000 contracts, and average daily open interest of 390,000 contracts. VIX options had an average daily volume of 534,000 and average daily open interest of 8.4 million contracts. More details are available in <http://www.cboe.com/micro/vix/pdf/VIXfactsheet2019.pdf> and <https://www.cboe.com/micro/vix/vixwhite.pdf>.

²This type of feature has also been documented in equity index options (see, e.g., Christoffersen, Heston, and Jacobs (2009)). However, I find this feature to be much more pronounced in VIX options, as the correlation between short-term implied skew and at-the-money (ATM) volatility is only 0.0832 in their levels and -0.0285 in their changes.

³This paper focuses on the reduced-form model, which directly provides specific dynamics of the volatility index (e.g., Mencía and Sentana (2013), Park (2016)). Another kind of model, namely, consistent model, involves the volatility index derived by specifying the dynamics of stock market returns (e.g., Cheng, Ibraimi, Leippold, and Zhang (2012), Branger, Kraftschik, and Vlkert (2016), Bardgett, Gourier, and Leippold (2019)). Although these models take into account the joint dynamics of the VIX and S&P 500 index returns, they require a much more complicated structure to capture stylized features. From a practical point of view, VIX options claim on VIX futures and thus can be more easily evaluated by postulating the dynamics of VIX itself.

⁴The level central tendency factor characterizes the time-varying long-run mean of VIX. The stochastic variance factor is the stochastic variance of the continuous part in VIX. They both are crucial in pricing VIX derivatives (see, e.g., Mencía and Sentana (2013), Park (2016), Bardgett et al. (2019) and Branger et al. (2016)).

options.

The model design is also based on the finding that there exists a highly persistent risk factor governing the skew both in the short- and long-end term structure of options IV. Andersen, Fusari, and Todorov (2015) develop a model with a separate tail factor controlling the negative jump intensity and then estimate it on S&P 500 index options. They find that the tail factor is highly persistent and is identified by the relatively more expensive out-of-the-money (OTM) long maturity puts. Motivated by their study, I propose using an independent risk factor to drive the stochastic jump intensity in order to generate different dynamics from those of stochastic variance. The jump intensity is identified as a relatively persistent factor by VIX OTM long maturity calls.

Furthermore, I note that the serial correlations of both ATM IV and skew increase with the option maturity, suggesting that low persistence factors dominate short-term contracts, while high persistence factors dominate long-term contracts.⁵ Moreover, VIX options exhibit a stochastic term structure of both IV and skew. Therefore, the stochastic variance central tendency can also improve the pricing performance along the time to maturity dimension. Importantly, by introducing central tendency factors, we can distinguish transitory shocks and persistent shocks from the general variations in the VIX market, and then observe how market participants perceive different risks (e.g., variance and variance of variance risks) for different horizons.

The IV surface of VIX options exhibits much more fluctuation than that of equity index options. Following Christoffersen et al. (2009) and Andersen, Fusari, and Todorov (2015), I first investigate the IV surface by Principal Components Analysis (PCA) and find that the first three principal components are responsible for 97.75% of the total variations, with each explaining 79.58%, 12.95%, and 4.94%, respectively. The results of PCA indicate that the second and third PCs account for much higher proportions than those in the PCA results of equity options reported by Christoffersen et al. (2009) and Andersen et al. (2015). As such, it is crucial to include more risk factors to improve the performance of the valuation model for VIX options.

Based on the seminal paper by Duffie, Pan, and Singleton (2000), I develop a general affine jump-diffusion pricing model for VIX derivatives, which includes a wide range of nested models. The

⁵This feature is also found in other markets, for example, stock options (Christoffersen et al. (2009)), currency options (Carr and Wu (2007)), and commodity derivatives (Schwartz and Smith (2000)). However, almost all VIX option pricing models in the literature only focus on short-end features, thus causing inaccuracy in IV term structures.

preferred model also accommodates contemporaneous jumps both in VIX level and VIX variance.⁶ The empirical results imply that the preferred model outperforms other nested models along the term structure dimension as well as the moneyness dimension of VIX options. Importantly, I show that in addition to level central tendency and stochastic variance factors, variance central tendency and jump intensity factors are both indispensable to the valuation of VIX options. Because the model that includes only one of these two risk factors can only slightly reduce out-of-sample RMSREs, the model with both factors can significantly reduce the errors by 30.83%.

To illustrate how the preferred model captures the independence feature in VIX options, I first show the distinct roles of the risk factors in pricing VIX options. I construct a simulated IV surface by changing levels of different risk factors with estimated model parameters. I show that each risk factor has a distinct but complementary role in pricing VIX derivatives. This pricing feature makes the model flexible enough to accurately portray the tail behavior of VIX risk-neutral distribution.

Next, I demonstrate the capability of the model to capture the time-varying skew and the independence feature from multiple perspectives. First, I conduct two simulation studies to show that the model can generate stochastic skew which is independent of the level of the overall variance. In the first study, I calculate the model implied ATM IV and skew by using randomly generated states. I then draw a scatter plot of the level and slope, and compare it with the empirically observed one. The simulated result shows that the model can capture the independence feature reasonably well. Moreover, the model matches the ranges of ATM IV and skew observed from the market data. In the second study, I investigate the integral influence of risk factors on IV slope while controlling for IV level. I show that, given the same ATM IV level, the model is able to generate a wide range of IV skew by using desirable combinations of the risk factors. This enables the model to capture the independence of the level and slope of the VIX smile.

Finally, I provide evidence from the model fit performance. I compare the market IV smirk with calibrated IV smirk of four typical days (e.g., low-level and low-skew; low-level and high-skew; high-level and high-skew; high-level and low-skew). The results show that the model fits both the level and skew in all situations very well.

This paper is most closely related to Mencía and Sentana (2013), my work, however, differs from

⁶The preferred model is the pricing model that has the lowest RMSRE for both in- and out-of-sample and lowest Akaike and Bayesian information criteria (AIC and BIC).

their study in the following aspects. First, the models in their study fail to capture the independence between the level and slope of the VIX smile, whereas my models armed with distinct risk factors can capture the independence feature of VIX options for both short and long maturities. Second, the authors provide evidence that the stochastic central tendency of VIX level improves the model fit of VIX futures. Based on their findings, I further show that the stochastic central tendency of VIX variance is crucial in pricing VIX options. Third, I show that the contemporaneous jumps in both VIX level and VIX variance improve the fits of high-strike options along the moneyness dimension and long-term options along the time to maturity dimension. Last but not least, the authors study a completely unspanned volatility model in which it is hard to explain the positive relationship between changes in the VIX index and VVIX index.⁷ However, I allow a nonzero correlation between Brownian motions in VIX level and VIX variance (see, e.g., Park (2016)). Furthermore, I include co-jumps to capture the common jumps in VIX and VVIX. I find that the positive correlation and contemporaneous upward jumps help generate the positive slope of the VIX smile.

Based on the work of Mencía and Sentana (2013), Park (2016) proposes a diffusion-jump model to accommodate positive correlations between changes in the VIX level and VIX variance, and further allows upward discontinuous movement in VIX, resulting in an upward sloping skew. The single variance factor in the Park (2016) model governs both stochastic variance and skew, thus contradicting the fact that their evolution is largely independent. In contrast, my model can successfully capture the behavior of the time-varying skew of the VIX smile.

To my knowledge, this paper is the first to investigate the effect of co-jumps in VIX options. By analyzing the VVIX index, Zang, Ni, Huang, and Wu (2017) document the existence of co-jumps in VIX level and VIX variance. However, OTM options are more informative in the structure of jumps than VVIX as they capture more tail features of risk-neutral distribution. Furthermore, Zang et al. (2017) assume that the underlying of VVIX is VIX itself. In contrast, I emphasize that the underlying of VIX derivatives is VIX futures.

The present work is also related to the study by Christoffersen et al. (2009). They focus on the multifactor stochastic volatility models, while I study the models that incorporate the jump

⁷The CBOE VVIX Index (the VVIX) represents a volatility of volatility in the sense that it measures the expected volatility of the 30-day forward price of VIX. This forward price is the price of a hypothetical VIX futures contract that expires in 30 days.

process. Most importantly, I provide a different model structure to capture the independence feature of the VIX smirk. Christoffersen et al. (2009) show that the multifactor stochastic volatility models display stochastic correlation which is independent of the level of variance. Therefore, this feature enables the model to capture the independence of the level and slope of the IV smirk, as stochastic correlation can potentially explain the stochastic skew. Different from their work, my model captures the independence feature by primarily including three risk factors: variance, variance central tendency, and jump intensity. I show that the flexible combination of the different risk factors with distinct yet complementary impacts on the IV surface can capture the independence of the level and slope of the IV smirk for both short and long maturities.

The rest of this paper is organized as follows. Section 2 describes the VIX derivatives data and provides empirical findings that motivate the parametric modeling. Section 3 lays out the pricing model and introduces the estimation methodology. Section 4 discusses the parameter estimates and pricing performance across the different model specifications. Section 5 investigates the properties of the pricing model. Section 6 concludes the paper. The Appendix contains technical derivations.

2. Data and Preliminary Analysis

2.1. Data description

There are both futures and options written on the VIX in the derivatives market. Daily settlement prices of VIX futures and options quotes are obtained from the CBOE’s website and OptionMetrics, respectively. The VIX futures were introduced on March 26, 2004, and the VIX options market started about two years later on February 24, 2006. Because the trading of VIX options was inactive in the first year after they were introduced, the sample spans the period from April 2, 2007, to December 29, 2017, for a total of 2701 business days.

I restrict the analysis to time to maturities between 7 and 160 days and apply several standard filters to eliminate inaccurate or illiquid options (see, e.g., Bakshi, Cao, and Chen (1997)). Specifically, I eliminate option quotes that do not satisfy standard no-arbitrage conditions and report zero trading volume on a given date. I further delete options with negative bid-ask spreads and options for which the IV cannot be calculated. Moreover, because OTM options tend to be more liquid than in-the-money (ITM) ones, I only work with liquid OTM call and put options. These

adjustments leave a total of 156,088 options. The number of VIX options in the sample on a given date increases with time, with around 20 options at the beginning of the data set and around 90 options at the end. Although VIX options only started trading in 2006, the total volume per day is about one-half of the total volume of S&P 500 options traded at the end of the sample. Note that VIX option prices do not satisfy no-arbitrage relations with respect to the VIX index, but they do with respect to the VIX futures value. Thus, I obtain VIX futures that are paired with VIX options on a given day by interpolation of the future curve. This gives a total of 14,203 futures contracts.

[Insert Figure 1 near here]

Panel A in Figure 1 plots temporal variation of the number of VIX options for different days to maturity (DTM) bins. We can notice that the number of contracts for all DTM bins has dramatically increased since 2007. Furthermore, the trading of short-term contracts is relatively more active at the beginning of the sample period. However, the number of long-term options increases to a similar level as that of short-term contracts at the end of the sample.

Panel B in Figure 1 further illustrates the time variation of the number of options contracts for different moneyness bins. I define moneyness as $K/F(t, T)$, where $F(t, T)$ is VIX futures price with time to maturity T at t , and K denotes strike price.⁸ Furthermore, I define the OTM call (put) options as the options that have moneyness $K/F > 1.8 (< 0.8)$. As expected, we continue to observe an upward trend for the number of contracts in all moneyness bins. Importantly, the number of OTM call options has significantly increased, and even dominates all other types of options in 2017. Furthermore, the deep OTM put and call options show distinct time variations, motivating further investigation of how the IV smile changes over time.

Panel C in Figure 1 plots time variation of the term structure of VIX IV. In general, the IV term structure is downward sloping. Furthermore, both the short- and long-end of IV term structure show substantial time variations, with short-end increasing dramatically during crises. Panel D in Figure 1 shows that IV of deep OTM call options is generally larger than that of deep OTM put options, and that the IV smile significantly changes over time. Therefore, a successful option pricing model should capture these features well.

⁸For simplicity, hereafter, I use F instead of $F(t, T)$ to present futures price with time to maturity T at t .

[Insert Table 1 near here]

Table 1 presents descriptive statistics for the option quotes by moneyness and maturity. Consistent with previous studies (e.g., Vlkert (2015) and Park (2016)), pronounced upward sloping patterns of option IV is evident from Panel C across all maturities, with short-term options exhibiting the steepest volatility smirk. VIX call options are more heavily traded than put options. This is due to the leverage effect, namely, negative changes in market returns are associated with increases in volatility, indicating that OTM call options on the VIX may provide disaster insurance for the overall equity market. For example, market participants use OTM VIX call options to protect their portfolios against sharp increases in volatility.

Bardgett et al. (2019) point out that, compared with S&P 500 options, VIX options are likely to contain more information on the future density of S&P 500 return variance, as they characterize the whole conditional density of future VIX levels. In contrast, S&P 500 options only provide us with a point estimate of the conditional variance of returns at each traded maturity. As such, the VIX derivatives market can provide more insights about market participants' perceptions of aggregate stock market uncertainty (e.g., Vlkert (2015)).

Under the classic hypothesis, the increments of $\log(\text{VIX})$ are independently and identically distributed (iid) with non-normal distribution, and the non-normality disappears rapidly as the time horizon increases (e.g., Carr and Wu (2007)). Nevertheless, Panel C shows that the average smirk flattens out slowly as the option maturity increases. This pattern indicates that the risk-neutral distribution remains highly positively skewed as the time interval increases. As a result, the VIX option pricing model demands a more persistent driving factor to slow down the convergence of the distribution of (\log) VIX to normality.

2.2. Time-varying skew and other features of VIX IV surface

I compute IV using the standard Black-Scholes formula to identify several essential features that an option pricing model should accommodate. Following Christoffersen et al. (2009) and Andersen et al. (2015), I first investigate the evolution of IV characteristics and their different properties among different maturities. Then, I apply PCA to explore the number of factors that drive the VIX IV surface.

In Figure 2, I plot the VIX index and its futures prices. The plots show that, historically, the VIX and its futures have experienced significant variations, but VIX futures prices are relatively less volatile due to mean-reverting property of the VIX. VIX futures prices measure the first moment or the mean of the risk-neutral distribution of VIX. Panel B of Figure 2 also reveals that the term structure of VIX futures varies over time. In general, during periods of market tranquility, the price of long-term VIX futures is higher than that of short-term futures. In contrast, during periods of market turmoil, the price of short-term futures responds quickly and even exceeds that of long-term futures.

[Insert Figure 2 near here]

Figure 3 plots the characteristics of the VIX IV surface. Panels A and B show the evolution of IV for short- and long-term ATM options, respectively. ATM options have moneyness K/F between 0.9 and 1.1. Furthermore, short-term (long-term) options are those with days to maturity $< 45 (> 120)$ days. First, we can notice that ATM IVs fluctuate dramatically along with time, indicating that a stochastic volatility model is required to accommodate such features (e.g., Mencía and Sentana (2013) and Park (2016)). Panel B also shows that, compared with the short-term ATM IV, the long-term one fluctuates around a relatively lower level, showing an overall downward-sloping term structure of VIX variance. This is in contrast to a general upward-sloping term structure of VIX futures. Moreover, the long-term ATM IV manifests a much more persistent evolution. Although the short-term ATM IV swings along a horizontal line, the long-term one fluctuates along an upward-sloping line, thus illustrating an increasing uncertainty about long-term market volatility perceived by investors. The distinct features of the long-term ATM IV thus require that the option pricing model include a high persistent driving factor to control the common behavior of short- and long-term options contracts, and a low persistent factor to capture the transitory variation among short-term options contracts (e.g., Carr and Wu (2007), Christoffersen et al. (2009) and Trolle and Schwartz (2009)).

[Insert Figure 3 near here]

Panel C of Figure 3 represents the time series of short-term IV skew that provides an intuitive measure of the asymmetry of the IV along the moneyness dimension. Specifically, IV skew is the

difference between the IV of OTM call and OTM put options. OTM call (put) options have monogeneity $K/F > 1.8$ (< 0.8). Similar to ATM IV, short-term skew exhibits a remarkable fluctuation. In contrast to the negative skew of S&P 500 index options IV, the skew of VIX options IV is consistently positive, with the sample average of 0.54, implying that OTM call options are more expensive than corresponding OTM put options. Another distinct feature is that the correlation between skew and ATM IV is only 0.0832 in levels and -0.0285 in changes. Thus, the slope and the level of IV are largely independent.

The solid line in Panel C represents the moving average of IV skew. When compared with plots in Panels A and C, we can notice that the IV skew evolves quite independently from the IV level. This feature of IV skew challenges the most popular single stochastic volatility model in VIX options pricing literature (e.g., Mencía and Sentana (2013), Park (2016); Bardgett et al. (2019) and Branger et al. (2016)).

Panel D shows the dynamic of the long-term IV skew over time, along with its moving average. Long-term skew also exhibits a significant time variation. The moving average of long-term skew has shown an upward trend since the financial crisis in 2008. Thus, market participants believe that large upward movements in market volatility (VIX) are more likely to happen. The correlation between skew and ATM IV of long-term options is 0.47, therefore higher than that of short-term options. The higher correlation is due to an upward trend in both time series. However, the correlation between their first differences turns out to be only -0.0571.

Panels E and F plot the term structure of ATM IV and skew, respectively. Consistent with previous findings, the term structures of both ATM IV and skew are generally negative. Furthermore, the evolution of skew term structure is quite distinct from that of ATM IV term structure, with a correlation of 0.0061. Comparing Panel E with Panel A, we can notice a considerable degree of commonality. However, ATM IV term structure spikes relatively more often than short-term ATM IV. Finally, the skew term structure fluctuates dramatically within a relatively stable range.

Turning to the PCA, I investigate the number of factors in the VIX IV surface. Following Christoffersen et al. (2009), I perform the PCA on an interpolated IV surface with fixed delta and maturity. Table 3 shows that the first three principal components capture about 97.75%. The first principal component explains about 79.58% of the total variation of the IV surface and captures the level of the IV, while the second principal component explains an additional 12.95%

and captures the slope of the IV in the maturity dimension, and the third component accounts for 4.94%, capturing the slope of the IV in the moneyness dimension. Importantly, the second and the third components totally explain a significant part, summing up to 15.8% of the total variation. In contrast, they account in total for only 7.64% and 2.8% variation of S&P 500 options reported in Christoffersen et al. (2009) and Andersen et al. (2015), respectively. This finding suggests that the structure of VIX options IV surface is much more complex than that of SP 500 options, and that each of the first three principal components plays an essential role in explaining the variation of VIX options.

[Insert Tables 2 and 3 near here]

In summary, preliminary analysis shows that the VIX options have several salient features in the term structure dimension as well as the moneyness dimension. First and foremost, both short- and long-term IV skew are stochastic and exhibit low correlations to ATM IVs, revealing that the traditional VIX option pricing model with a single stochastic variance factor is not flexible enough. Second, the pronounced positive skew in long-term options reflects a persistent factor slowing down the convergence of the distribution to normality. Third, Table 2 shows that the serial correlations of all time series increase with the option maturity. To take into account these features, I propose including multiple risk factors with different persistence. It is well known that a higher value of volatility factor that follows a CIR process, e.g., Heston (1993), always results in a flatter smirk (Du and Luo (2019)). As a result, the option pricing model requires an additional factor with high persistence to accommodate an upward trend in the IV level and IV skew for the long-term contracts. In Section 4, I will show that the contemporaneous jumps in both VIX level and VIX variance with a highly persistent stochastic jump arrival rate can generate such a pattern. More detailed analysis of model properties is in Section 5

3. Model and Estimation

3.1. Model specification

In this section, I introduce a tractable framework for pricing VIX derivatives in the presence of a comprehensive stochastic variance factor with time-varying central tendency and an independent

jump intensity factor. This model is novel in VIX derivatives pricing literature and able to capture essential facts in the VIX derivatives market. I show that, although the stochastic central tendency factor in VIX level (hereafter called level central tendency factor) is able to describe the term structure of VIX futures, the one in VIX variance (hereafter called variance central tendency factor) is crucial in order to capture the term structure of VIX options. In contrast to Branger et al. (2016), who maintain that adding jumps is of only minor importance for the pricing of VIX derivatives, I illustrate that the level and variance co-jumps with stochastic arrival rate are essential to capture the stochastic skew of the VIX IV smile, which is critical to infer time-varying fear sentiment of investors.

I assume that the logarithm of VIX (vix_t) follows an affine jump-diffusion process. This assumption is motivated by the well-known empirical finding that a logarithmic model does a better job of describing the volatility dynamics of a stock index than the square root model of Heston (1993). Furthermore, Mencía and Sentana (2013) compare logarithmic and square root models with respect to the pricing of VIX derivatives and find that the former outperforms the latter.

I adopt a general-to-specific approach. The proposed model subsequently serves as the basis for the analysis of the VIX derivatives. Recently, Mencía and Sentana (2013) and Branger et al. (2016) point out that the time-varying long-run mean of vix_t , denoted by m_t , is crucial for pricing VIX futures. This is in line with the studies of Duffie et al. (2000) and Balduzzi, Das, and Foresi (1998). The above studies also find that the stochastic variance of the continuous part in vix_t , denoted by V_t , is important to match the higher order risk-neutral moments and thus provides a better fit for VIX options. Even though this two-latent factor model fits VIX derivatives data well at a given point, it is not flexible enough to capture the time-varying nature of the cross section in the VIX options market. In other words, m_t is crucial for futures but has a negligible effect on options once we condition on the current futures price. As a result, only one latent factor is left to capture the complicated time-varying behavior of the IV surface of VIX options. When the model is calibrated to capture a steep smirk, it cannot capture a flat smirk, and vice versa. Thus, the model fails to capture the time-varying skew and thus requires more factors to describe the IV surface.

Therefore, based on the model in Park (2016), I further introduce two risk factors: variance central tendency C_t which controls stochastic mean level of V_t , and stochastic jump intensity λ_t which independently controls the arrival rate of co-jumps in vix_t and V_t .

I use $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{Q})$ to denote a complete stochastic basis defined on the risk-neutral measure \mathbb{Q} , under which the most generous model for the (logarithmic) VIX dynamics is given by

$$dvix_t = \kappa (m_t - vix_t) dt + \sqrt{V_t} dW_t + J dN_t - \lambda_t \mu dt, \quad (1)$$

$$dm_t = \kappa_m (\theta_m - m_t) dt + \omega_m \sqrt{m_t} dW_t^m, \quad (2)$$

$$dV_t = \kappa_v (C_t - V_t) dt + \omega_v \sqrt{V_t} dW_t^v + J_v dN_t, \quad (3)$$

$$dC_t = \kappa_c (\theta_c - C_t) dt + \omega_c \sqrt{C_t} dW_t^c, \quad (4)$$

$$d\lambda_t = \kappa_\lambda (\theta_\lambda - \lambda_t) dt + \omega_\lambda \sqrt{\lambda_t} dW_t^\lambda, \quad (5)$$

where $(W_t, W_t^m, W_t^v, W_t^c, W_t^\lambda)$ is a five-dimensional Brownian motion. I allow the Brownian motions W_t in the vix_t process and Brownian motions W_t^v in V_t to be correlated, $\rho dt = E[dW_t dW_t^v]$. All other Brownian motions are mutually independent. N_t is the risk-neutral Poisson process with stochastic intensity λ_t following Equation (5). I assume that the random jump sizes are independent and identically distributed. The jump sizes in vix_t and V_t are assumed to be exponentially distributed with means μ and δ . For parsimony, I only consider upward jumps based on the result in Park (2016) that downward jumps provide little improvement in the pricing of VIX derivatives. Thus, the positive ρ and upward co-jumps capture the fact that there is a positive relation between changes in VIX and VVIX indices of the CBOE. Hence, the above model specification prices only jump upwards, whereas any downward movement is simply captured by the diffusion components in vix_t and V_t .

Equation (2) drives the dynamics of the stochastic level around which vix_t reverts to, which I refer as the level central tendency. Furthermore, Equation (4) characterizes the time-varying long-run mean of variance V_t , which I refer to as variance central tendency. The high persistence features of m_t and C_t identified by the data indicates that their effects dominate long-term contracts, whereas low persistent vix_t and V_t dominate short-term contracts. As a result, the model is better able to capture the term structure of both VIX futures and options. Finally, I assume that the jump intensity follows a CIR process. In Section 4, I show that allowing λ_t to independently control the jump arrival rate is crucial to capture the variation of implied skew. This assumption contrasts

with most previous work that usually assumes that jump intensity is either constant or proportional to the variance of diffusion, resulting in a tied link between the level and slope of IV.

3.2. Nested specifications

The model presented above is quite general and subsumes existing pricing models along many dimensions. First, the model allows for stochastic long-run mean in VIX variance to better capture the behavior of long-term options contracts. Mencía and Sentana (2013), Park (2016) and Branger et al. (2016) only allow stochastic central tendency in VIX level but not in VIX variance. Second, the model introduces an independent stochastic jump intensity to drive a wedge between the diffusion and jump process. Nearly all studies assume that jump intensity is constant, including, Eraker, Johannes, and Polson (2003) and Broadie, Chernov, and Johannes (2007), or proportional to the volatility, for example, Pan (2002), Eraker (2004) and Carr and Wu (2007). Finally, the model features contemporaneous jumps both in VIX level and VIX variance. The effect of co-jumps has been studied in different markets. Duffie et al. (2000) and Du and Luo (2019) apply a double jumps model in the S&P 500 index options market; Bakshi, Cao, and Zhong (2012) study it in individual stock options; and Aït-Sahalia, Karaman, and Mancini (2018) study it in variance swaps, while it has not yet been studied in VIX derivatives.

The most generous specification will be denoted by VVC-DJ (VIX level and VIX Variance Central tendency with Double Jumps).⁹ I have also experimented with a wide range of nested and more parsimonious alternatives. To streamline the exposition of the above model features, I will focus on the five most interesting specifications which will be denoted by VC (VIX level Central tendency without Jumps), VC-CJ (VIX level Central tendency with Constant Jumps in VIX level), VVC-CJ (VIX level and VIX Variance Central tendency with Constant Jumps in VIX level), VC-SJ (VIX level Central tendency with Stochastic Jumps in VIX level), and VVC-SJ (VIX level and VIX Variance Central tendency with Stochastic Jumps in VIX level). The details of model specifications are summarized in Table 4.

Specifically, the VC model is the most basic model and only includes VIX level central tendency factor m_t and VIX variance factor V_t . In the VC-CJ specification, I set C_t and λ_t as constant θ_v

⁹Considering the importance VIX level central tendency in pricing VIX futures, I include it in all models considered in this paper.

and λ , respectively. This specification has been well studied in Park (2016) and Branger et al. (2016) and is the benchmark in this paper. In the VVC-CJ specification, I set λ_t as constant λ to investigate the effect of stochastic jump intensity on VIX derivatives. In order to study the importance of central tendency in variance and double jumps in VIX level and VIX variance, I further introduce two more specifications, VC-SJ and VVC-SJ. They both allow jumps in VIX level only, and VC-SJ further restricts C_t to a constant.

3.3. Risk premium specification

In order to apply the estimation method introduced in Section 3.5, I need to specify the dynamics of the state vector under the actual probability measure \mathbb{P} .¹⁰ Based on model specification from Equations (1) to (5), the risks can be decomposed into diffusive risk and jump risk. I assume that the market prices of diffusive risks are proportional to the corresponding risk level, with the coefficient of proportionality given by¹¹

$$\Lambda_{vix} = \gamma, \quad (6)$$

$$\Lambda_m = \kappa_m - \kappa_m^{\mathbb{P}}, \quad (7)$$

$$\Lambda_v = \kappa_v - \kappa_v^{\mathbb{P}}, \quad (8)$$

$$\Lambda_c = \kappa_c - \kappa_c^{\mathbb{P}}. \quad (9)$$

This specification preserves the affine structure of the state vector under the change of measure.¹² Furthermore, following Pan (2002) and Eraker (2004), I assume that the arrive rate of jumps is the same under \mathbb{Q} and \mathbb{P} . However, I allow the means of jump sizes to be different under \mathbb{Q} and \mathbb{P} . Specifically, the local jump intensity λ_t is the same under \mathbb{Q} and \mathbb{P} , but the means of jump sizes in vix_t and V_t under \mathbb{P} are $\mu^{\mathbb{P}}$ and $\delta^{\mathbb{P}}$, respectively.¹³ Therefore, I obtain the following

¹⁰I estimate the models by maximum likelihood. However, a closed-form expression for the likelihood is not available. Therefore, following Carr and Wu (2007), among others, I cast the models into a state-space form and calibrate their parameters by quasi-maximum likelihood. Furthermore, I use both derivative prices and historical observations on the VIX itself since I want to estimate both physical and risk-neutral measures.

¹¹This specification has been wide applied in the literature (see, e.g., Ait-Sahalia et al. (2018), Li and Zinna (2018), Bardgett et al. (2019), Bakshi, Carr, and Wu (2008), and Bakshi and Wu (2010)).

¹²In the Internet Appendix, I provide an analysis of the risk premiums embedded in VIX derivatives. I apply a parametric method to construct the components of the risk premia at different horizons. The decomposition allows us to quantify how the various risks contribute to the risk premium and the corresponding term structure of risk premium.

¹³The primary motivation for this assumption is the well-known limited ability to estimate jump components in

model under the actual probability measure \mathbb{P}

$$dvix_t = \kappa(m_t - vix_t)dt + \lambda_t(\mu^{\mathbb{P}} - \mu)dt + \gamma V_t dt + \sqrt{V_t} dW_t^{\mathbb{P}} + J^{\mathbb{P}} dN_t - \lambda_t \mu^{\mathbb{P}} dt, \quad (10)$$

$$dm_t = \kappa_m^{\mathbb{P}}(\theta_m^{\mathbb{P}} - m_t)dt + \omega_m \sqrt{m_t} dW_t^{m,\mathbb{P}}, \quad (11)$$

$$dV_t = \kappa_v^{\mathbb{P}}\left(\frac{\kappa_v}{\kappa_v^{\mathbb{P}}}C_t - V_t\right)dt + \omega_v \sqrt{V_t} dW_t^{v,\mathbb{P}} + J_v^{\mathbb{P}} dN_t, \quad (12)$$

$$dC_t = \kappa_c^{\mathbb{P}}(\theta_c^{\mathbb{P}} - C_t)dt + \omega_c \sqrt{C_t} dW_t^{c,\mathbb{P}}, \quad (13)$$

$$d\lambda_t = \kappa_{\lambda}(\theta_{\lambda} - \lambda_t)dt + \omega_{\lambda} \sqrt{\lambda_t} dW_t^{\lambda,\mathbb{P}}, \quad (14)$$

where $\kappa_m^{\mathbb{P}}$, $\kappa_v^{\mathbb{P}}$ and $\kappa_c^{\mathbb{P}}$ capture the persistence of m_t , V_t and C_t , $\theta_m^{\mathbb{P}} = \kappa_m \theta_m / \kappa_m^{\mathbb{P}}$, $\theta_c^{\mathbb{P}} = \kappa_c \theta_c / \kappa_c^{\mathbb{P}}$ and $(W_t^{\mathbb{P}}, W_t^{m,\mathbb{P}}, W_t^{v,\mathbb{P}}, W_t^{c,\mathbb{P}}, W_t^{\lambda,\mathbb{P}})$ is a five-dimensional Brownian motion. $J^{\mathbb{P}}$ and $J_v^{\mathbb{P}}$ are magnitude of upward level jumps and variance jumps under \mathbb{P} , respectively.

3.4. VIX derivatives pricing

VIX futures and options are written on the forward VIX. Thus, the price of futures with a maturity of T at time t is equal to the risk-neutral expectation of the forward VIX

$$F(t, T) = E_t^{\mathbb{Q}}[VIX_T]. \quad (15)$$

The payoff of a call option on the VIX with maturity in T and strike price K is the maximum value between $VIX_T - K$ and 0. Based on the risk-neutral pricing theory, we can express the value of a European call option as

$$Call_t(K, T) = \exp(-r_t \tau) E_t^{\mathbb{Q}}[\max(VIX_T - K, 0)], \quad (16)$$

where $\tau = T - t$ and r_t is the risk-free rate at time t . Note that the actual underlying of VIX options is VIX futures. Therefore, I use futures that expire on the same date to price options. Similarly, European put prices can be obtained from the put-call-forward parity relationship

$$Put_t(K, T) = Call_t(K, T) - \exp(-r_t \tau)[F(T, t) - K]. \quad (17)$$

index returns and the corresponding risk premium using daily data. Thus, all price jump risk premium is absorbed by the price jump size risk premium.

The model specifications belong to the class of affine jump-diffusion state processes. As such, the conditional characteristic function of vix_T at time t , $E_t^{\mathbb{Q}} [\exp(i\phi vix_T)]$ can be quasi-analytically calculated by solving a system of Riccati ordinary differential equations (ODEs). Then, option pricing is performed using Fourier inversion techniques. For estimation, I also need to develop the log-likelihood function for the v_T . Similarly, I first derive the characteristic function of vix_T under the historical measure \mathbb{P} and then obtain the density via Fourier inversion. Appendix A summarizes the density of vix_T and semi closed-form of options pricing.

3.5. Maximum likelihood estimation

To estimate the dynamic models using an extensive joint data set of VIX, VIX futures, and options, I cast the models into a state-space form and infer the driving factors using a filtering technique.

I estimate the models using the joint MLE of daily in-sample data from April 2, 2007 to March 31, 2015.¹⁴ In the following, I use the most general model (VVC-DJ) to illustrate the estimation procedure. Similar approaches for the other models can be easily deduced as nested cases. Let $X_t = [m_t, V_t, C_t, \lambda_t]^{\top}$ denote the unobservable state variables. Following Carr and Wu (2017), I discretize the physical dynamics of X_t using the conventional Euler scheme at the daily interval. The unobservable state propagation equation can be written as

$$X_t = \mathbf{F}(X_{t-1}; \Theta) + \sqrt{\mathbf{Q}_{t-1}} \epsilon_t, \quad (18)$$

where Θ denotes the model parameters; ϵ_t denotes the standardized forecast error vector with mean zero. The forecasting function $\mathbf{F}(X_{t-1}; \Theta)$ and the forecasting error covariance matrix are given by

$$\mathbf{F}(X_t; \Theta) = \begin{bmatrix} m_t + \kappa_m^{\mathbb{P}} (\theta_m^{\mathbb{P}} - m_t) \Delta t \\ V_t + \kappa_v^{\mathbb{P}} \left(\frac{\kappa_p}{\kappa_v^{\mathbb{P}}} C_t - V_t \right) \Delta t + \lambda_t \delta^{\mathbb{P}} \Delta t \\ C_t + \kappa_c^{\mathbb{P}} (\theta_c^{\mathbb{P}} - C_t) \Delta t \\ \lambda_t + \kappa_\lambda (\theta_\lambda - \lambda_t) \Delta t \end{bmatrix}, \quad (19)$$

¹⁴The joint maximum likelihood estimation method has been widely applied in the literature (see, e.g., Bakshi et al. (2008); Bakshi and Wu (2010); Carr and Wu (2017); Du and Luo (2019); Bardgett et al. (2019)).

$$\mathbf{Q}(X_t; \Theta) = \begin{pmatrix} \omega_m^2 m_t \Delta t \\ \omega_v^2 V_t \Delta t + 2 (\delta^{\mathbb{P}})^2 \lambda_t \Delta t \\ \omega_c^2 C_t \Delta t \\ \omega_{\lambda}^2 \lambda_t \Delta t \end{pmatrix}, \quad (20)$$

with $\Delta t = 1/252$ denoting the daily frequency of the data and $\langle \cdot \rangle$ denoting a diagonal matrix. The observation equation takes the form

$$y_t = \mathbf{H}(X_t; vix_t; \Theta) + u_t, \quad (21)$$

where y_t denotes the time- t forward value of the VIX futures and options, which depends on the models' structural parameters Θ , the unobservable state X_t , and (logarithmic) VIX vix_t . Following Broadie et al. (2007) and Christoffersen, Jacobs, and Ornthalalai (2012), I assume that the measurement error u_t is normally distributed with covariance matrix \mathbf{R} . Instead of using original prices of futures and options, I take the logarithm of futures and scale the option prices by their corresponding vegas.

As in Carr and Wu (2007) and Bakshi et al. (2008), among others, I make three simplifying assumptions on the covariance matrices \mathbf{Q} and \mathbf{R} . First, \mathbf{R} is assumed to be constant and diagonal, for example, pricing errors are uncorrelated in the cross-section. Second, I set the volatility of the pricing error equal to σ_f for all futures and σ_o for all options. Furthermore, following Carr and Wu (2017) and Du and Luo (2019), I approximate the distribution of the jump process in VIX variance state V_t by its quadratic variation.¹⁵

To address nonlinearity in the state and measurement equations, I apply the unscented Kalman filter (UKF) to extract the latent states (see Wan and Merwe (2000)) and construct log likelihood of the joint data. Christoffersen, Dorion, Jacobs, and Karoui (2014) demonstrate the superior performance of UKF over extended Kalman filter and particle filter for a variety of pricing problems.

Let \bar{X}_t , \bar{y}_t and $\bar{\Sigma}_{y,t}$ be the time- $(t-1)$ ex-ante forecasts of time- t values of the latent state vector, the measurement series, and the covariance of the measurement series, respectively. The unscented Kalman filter is applied to the VIX derivatives data sequentially from the first day to the last day

¹⁵The more precise method is to use the particle filter. However, the particle filter will substantially increase the computation burden. In Appendix C, I quantify the filtering results from UKF by conducting the Monte Carlo simulation and show that deviations of the filtered states from the "true" ones are quite small.

of the observation. With the filtering results, I construct the log-likelihood for each day's VIX derivative observations by assuming that the forecasting errors are normally distributed

$$l_t^y(\Theta) = -\frac{n_t}{2} \log(2\pi) - \frac{1}{2} \log |\bar{\Sigma}_{yy,t}| - \frac{1}{2} (y_t - \bar{y}_t)^\top (\bar{\Sigma}_{yy,t})^{-1} (y_t - \bar{y}_t). \quad (22)$$

Furthermore, conditional on the state vector filtered from the VIX derivatives, I compute the conditional probability density of vix_T . Let $\phi(\varphi; \hat{X}_t, vix_t, \Theta, T)$ denote the characteristic function of the vix_T under the \mathbb{P} -measure, where \hat{X}_t denotes the latent states at time t filtered from the VIX futures and options. Given the closed-form valuation of $\phi(\cdot)$, I apply fast Fourier inversion to obtain the probability density of vix_T , $f(vix_T; \hat{X}_t, vix_t, \Theta)$. The conditional log likelihood of the vix_T is

$$l_t^{vix}(\Theta) = \log(f(vix_T; \hat{X}_t, vix_t, \Theta)). \quad (23)$$

To accurately capture the dynamics under \mathbb{P} measure, I use different time horizons, $T - t$, of the probability density, including one day, one week, one month, two months, and four months.

For model estimation, I assume the distribution of the forecasting and pricing errors as being normal and independent of the vix_t dynamics. Thus, we can write the joint likelihood of VIX derivatives and vix_T as the sum of their marginal log-likelihood values. This practice is purely for computational feasibility because deriving the joint density would involve multi-dimensional integrations that are computationally intensive for the model estimation. Using the product of marginal densities incurs some theoretical information loss, but provides significant gains in numerical stability and feasibility.

VIX derivatives are available in the time-series as well as in the cross section. As a result, the number of data points available for computing the log likelihood for fitting VIX derivatives, $l_t^y(\Theta)$, is significantly larger than the number of data points available for computing the log likelihood for vix_T , $l_t^{vix}(\Theta)$.

In order to ensure that the parameter estimates from the joint estimation are not dominated by the VIX derivatives data, I use the weighted likelihood function analyzed in Hu and Zidek (2002), and then choose model parameters by maximizing the following weighted joint likelihood

$$\begin{aligned}
l(\Theta) &= \sum_{t=1}^T \left(\frac{1}{N_{vix}} l_t^{vix}(\Theta) + \frac{1}{N_y} l_t^y(\Theta) \right) \cdot T \\
&= \left(\frac{1}{N_{vix}} \mathcal{L}^{vix}(\Theta) + \frac{1}{N_y} \mathcal{L}^y(\Theta) \right) \cdot T
\end{aligned} \tag{24}$$

where T is the total number of days in the sample; $N_{vix} = T \cdot n_{vix}$ represents the total number of probability density under \mathbb{P} measure; n_{vix} is the number of density under \mathbb{P} measure on t , e.g., 5 in the application; N_y denotes the total number of VIX derivatives; and \mathcal{L}^{vix} and \mathcal{L}^y represent the total log-likelihoods for vix_T and VIX derivatives, respectively. Note that the last term, T , in (24) serves as a scaling constant and does not impact the parameter estimates. Thus, I assign equal weight to each data point when computing the total log-likelihood. This procedure has been used by Ornthalalai (2014) and Bardgett et al. (2019).

In addition to the extensive daily data set of VIX index and VIX derivatives, the valuation of VIX options involves the use of inverse Fourier transform and solving ODEs, which dramatically increases the calculation burden. Therefore, I utilize parallel programming (i.e., CUDA) to offload the highly numerically demanding part to the Graphics Processing Unit (GPU).

4. Empirical Results

4.1. In-sample parameter estimates

Table 5 displays in-sample point estimates and standard errors resulting from the estimation of VC-C, VC-CJ, VC-SJ, VVC-CJ, VVC-SJ, and VVC-DJ specifications.

[Insert Table 5 near here]

I note that across different model specifications, vix_t is highly mean-reverting toward m_t ($9.0856 < \kappa < 11.3630$), which in turn mean-reverts rather more slowly to its long-run mean θ_m ($0.5958 < \kappa_m < 0.7428$). The parameter estimates for m_t are quite stable among different model specifications ($3.2522 < \theta_m < 3.3363$ and $0.1736 < \omega_m < 0.2803$). Therefore, including stochastic central tendency m_t in vix_t is crucial to describe the first moment of vix_t and thus capture well the features of VIX futures. However, different models yield very different dynamics

for other states. The dynamics of V_t differ depending on whether stochastic central tendency in variance, C_t , is included. First, V_t is moderately persistent under the risk-neutral measure in models VC ($\kappa_v = 2.6426$), VC-CJ ($\kappa_v = 1.1892$) and VC-SJ ($\kappa_v = 0.1142$). However, V_t is fast mean reverting in other model specifications when C_t is included. The speed κ_v in VVC-CJ, VVC-SJ and VVC-DJ is estimated at 7.6650, 6.1008, and 6.7418, respectively. As expected, C_t is very persistent as κ_c is 0.1002, 0.1155 and 0.2029 in VVC-CJ, VVC-SJ and VVC-DJ, respectively. Thus, V_t captures transitory shocks to vix_t , whereas C_t captures more persistent shocks to vix_t . As a result, V_t primarily affects prices of short-term contracts, while C_t affects prices of all contracts.

The correlation ρ between changes in VIX level and VIX variance is positive across all model specifications, which is consistent with the findings in Park (2016). The estimated results show that correlation ρ in VVC-type models depends on jump specification. The correlation is highest in the VVC-DJ model ($\rho = 0.9176$) and lowest in the VVC-SJ model ($\rho = 0.6307$). The VVC-CJ model ($\rho = 0.6427$) falls in between. Thus, a jump process would also have an impact on the pricing of VIX derivatives through the correlation.

The highly statistically significant estimates of the jump parameters support the presence of jumps both in vix_t and V_t . The parameter μ , which controls the average jumps size in vix_t under risk-neutral measure, is about 0.3368 across different models with standard deviation below 0.005. The parameter λ , which governs the occurrence rate of the jumps in VIX level in VC-type models, is around 1.3987. In the VVC-DJ specification, the mean size of jumps in V_t is 2.4723 (with standard deviation of 0.0252), which is much higher than that in vix_t ($\mu = 0.2216$). Thus, diffusion variance does jump upward in conjunction with positive jumps in vix_t ; for example, there is strong evidence of co-jumps in VIX level and VIX variance. An important finding is that the estimates of κ_λ , θ_λ and ω_λ are all statistically significant, which strongly supports the existence of a time-varying jump intensity. As pointed out in Li and Zinna (2018), the identification of the jump intensity parameters stems from the combined use of the underlying VIX index and VIX derivatives for a comprehensive set of maturities and moneyness.

Much more interestingly, the parameter estimates of λ_t also depend on whether C_t is included. In the VC-SJ specification, jump intensity displays the fastest mean reversion as κ_λ is estimated at 8.0740. This evidence, accompanied with the result of highly persistent V_t in the VC-SJ model, indicates that λ_t dominates short-term contracts, while V_t affects prices of all contracts when the

model is lacking in C_t . On the contrary, in the VVC-type specifications, λ_t is much more persistent as κ_λ is estimated at 1.5245 and 0.9297 in VVC-SJ and VVC-DJ, respectively. Hence, λ_t and C_t have the potential to impact the derivatives substantially across both short and long maturities, whereas V_t mainly affects short-term contracts in VVC-type specifications. In fact, VVC-type models have lower instantaneous volatility of jump intensity, ω_λ , than VC-type models, implying that λ_t is more stable and more important for all contracts when C_t is included.

Comparing the estimated results with the estimation of models reported in Mencía and Sentana (2013), Park (2016), and Branger et al. (2016), I find that the mean-reversion of the V_t in VVC-DJ is far stronger than reported in the above-mentioned papers. This is compensated by the presence of the very persistent C_t and λ_t . In the traditional VIX derivatives pricing model, the jump intensity is proportional to the volatility factor. In this paper, through the introduction of an independent factor λ_t , I drive a wedge between volatility and jump intensity, and the VIX derivatives data identify jump intensity as the more persistent factor. Intuitively, because λ_t controls the right jump tails, this result stems from the relative expensiveness of the OTM long maturity calls, for example, the comparatively slow flattening out of the IV skew at long maturities. In research on the S&P 500 options market, Andersen et al. (2015) also find a more persistent factor that controls negative jump intensity and thus left jump tails of stock return distribution.

The risk premium coefficient related to vix_t is negative across all models ($-0.0342 < \gamma < -0.0108$), which implies a negative diffusive risk premium of VIX. This result is consistent with the negative variance risk premium of the S&P 500 index found by Carr and Wu (2009), Bollerslev, Tauchen, and Zhou (2009), and Bekaert and Hoerova (2014), as VIX measures the expected variance of the S&P 500 index implied from its options, and investors believe that it reverts to a higher level in the risk-neutral world. The risk premium coefficients of m_t , V_t , C_t (Λ_m , Λ_v , Λ_c in Equation (7) - (9)) are always found negative for each of the models considered, implying that the market risk factors of VIX derivatives are negatively priced. Furthermore, the average magnitude of jumps in vix_t is consistently found smaller under the objective measure than risk-neutral measure, which induces a negative instantaneous jump risk premium in vix_t , $\lambda_t (\mu^{\mathbb{P}} - \mu) < 0$.

Inspection of the estimated pricing errors of futures and options (σ_f and σ_o), which roughly measure fitness of the model, suggests that VVC-DJ is substantially superior to all nested specifications examined. Additionally, the log-likelihood (Log(L)), AIC, and BIC across the all models

further support the superiority of the VVC-DJ specification. Interestingly, the log-likelihood results of VC-SJ and VVC-CJ suggest that stochastic central tendency in V_t is relatively more important than stochastic jump intensity in pricing VIX options (3657.37 vs. 3779.81). Comparing the pricing errors of VIX futures and options in VC-CJ with those in VVC-DJ, I find that σ_f decreases by 14.18% from 0.0275 to 0.0236, and σ_o decreases by 26.27% from 0.0628 to 0.0463. Thus, the advantage of VVC-DJ mostly stems from the improvement of the fit to options.

4.2. Pricing performances

4.2.1. Option panel fit and out-of-sample performance

Data from April 2, 2007, to March 31, 2015, are used for in-sample estimation, while the remaining sample data from April 1, 2015, to December 29, 2017, are used for out-of-sample checks. To assess the model performance, I first define option pricing error as Root Mean Square Relative Errors (RMSRE)¹⁶

$$\text{RMSRE} = \sqrt{\frac{1}{N} \sum_{n=1}^N \left(\frac{\text{IV}_n - \text{IV}_n^M}{\text{IV}_n} \right)^2}, \quad (25)$$

where IV and IV^M denote the market-implied and model-implied volatilities, respectively; N represents the total numbers of VIX options.

[Insert Tables 6 and 7 near here]

The first three panels in Tables 6 and 7 present the RMSRE of all model specifications for in- and out-of-sample, respectively. Consistent with the findings in Park (2016) and Branger et al. (2016), I find that the RMSREs of VC are the highest both for in- and out-of-sample periods, thereby confirming the importance of the jump in VIX level. Therefore, for comparative purposes, I use VC-CJ as the benchmark model. The results show that the overall RMSRE from VVC-DJ is 0.0536 in-sample and 0.0689 out-of-sample, and RMSRE from VVC-SJ is 0.0543 in-sample and 0.0693 out-of-sample. Other specifications result in a relatively poor fit, with the in-sample RMSRE values ranging from 0.0722 for VC to 0.0645 for VVC-CJ and out-of-sample RMSRE values ranging

¹⁶Since all specifications match VIX futures well, here I only focus on the comparison of VIX options fitness.

from 0.1077 for VC to 0.0901 for VVC-CJ. Thus, VVC-SJ and VVC-DJ substantially outperform other specifications. Importantly, the superiority of VVC-DJ over VC-CJ is more notable in the out-of-sample period. This is because VVC-DJ reduces the option pricing errors over VC-CJ by 21.6% in-sample and 30.2% out-of-sample, indicating that VVC-DJ captures the empirical features of VIX options well. The overall RMSREs from in- and out-of-sample indicate that including both stochastic jump intensity and variance central tendency factors can dramatically reduce pricing errors.

Next, I perform similar comparisons after sorting the option contracts by moneyness (K/F) and days to maturity (DTM). In all subcategories, VVC-SJ and VVC-DJ outperform other models with clear economic significance for both in- and out-of-sample periods. Furthermore, VVC-DJ strongly outperforms VVC-SJ at high strikes along the moneyness dimension and for long-term contracts along the maturity dimension. The performance differences mainly come from the two jump specifications: the dynamic of jump intensity and the correlation between innovations in vix_t and V_t . Specifically, the relatively high persistence of jump intensity factor improves the fitness of the long-end of IV surface, and contemporaneous upward jumps in vix_t and V_t can generate more positive IV skew. These two effects together can potentially slow down the convergence of the distribution of (log) VIX to normality.

To test the statistical significance of the performance difference between different models, I make pairwise model comparisons. Let $D_{i,j,t}$ be the difference of daily RMSRE on time t between models i and j ,

$$D_{i,j,t} = \text{RMSRE}_{i,t} - \text{RMSRE}_{j,t}. \quad (26)$$

Next, I define a test statistic between modes i and j as

$$z_{i,j} = \frac{\overline{D_{i,j,t}}}{\text{stdev}(D_{i,j,t})}, \quad (27)$$

where $\overline{D_{i,j,t}}$ denotes the sample average and $\text{stdev}(\cdot)$ denotes the standard error of the sample mean difference. I adjust the standard error calculation for serial dependence based on Newey and West (1987), with the number of lags optimally chosen based on Andrews (1991) and an AR(1)

specification. The test statistic in Equation (27) follows a standard normal distribution under the null hypothesis that there is no statistically significant difference between modes i and j .

Panel D in Tables 6 and 7 report in- and out-of-sample pairwise test statistics, respectively. The statistics are in a (6×6) matrix, with the (i, j) th element being the statistic on model i versus model j . Given the symmetry of the test, the diagonal terms are zero by definition and the lower triangular elements are equal to the negative of the upper triangular elements. Thus, I focus on the lower triangular entries and bold the elements whose absolute values are greater than 2.33, which corresponds to a 99% confidence level on a one-side test. A negative (positive) statistic in entries (i, j) indicates that model i outperforms (underperforms) model j . All of the lower triangular elements in the last two rows are negative and strongly significant, indicating that both variance central tendency and jump intensity factors are indispensable. The fourth element in the last row further supports contemporaneous jumps in both VIX level and VIX variance.

4.2.2. Fitting the option characteristics

Next, I investigate the models' success in capturing the dynamics of the options panel. Specifically, I analyze the success of the models in tracking the dynamics of the IV surface by focusing on their ability to fit the option characteristics.

[Insert Figures 4 and 5 near here]

In Figure 4, I plot the time series of option skew implied by VVC-DJ and VC-CJ. Panels A and B display the comparison between the VC-CJ implied and observed option skew for short and long maturities, respectively. Obviously, throughout the full sample period, VC-CJ generates a skew that is too stable to capture the big swings in short-term skew. Although VC-CJ performs relatively well for matching long-term skew, it generates a skew that is too flat during tranquil periods (e.g., in 2015 and 2017), while overshooting during periods of market turmoil (e.g., in 2008, 2012 and 2016). Compared with option implied skew in the stock market, options skew in the VIX market fluctuates much more dramatically, thus demanding a more flexible model. However, the IV surface generated by VC-CJ is only driven by the VIX variance factor V_t once we condition on the current futures price, implying that the level and slope of IV are tightly related. Therefore, it is hard to match the desired the slope once we control for the level of IV smile. Panels C and D

show the fitted results of VVC-DJ for short- and long-term options. As expected, VVC-DJ closely replicates its data counterpart during the whole sample period.

Similar results are obtained in Figure 5 where I plot term structures of model- and market-implied volatility surface. To fully illustrate the variation in term structure of the IV surface, I plot the term structure of ATM IV in Panels A and C, and the term structure of skew in Panels B and D. For both types of term structure, VVC-DJ significantly outperforms VC-CJ in fitting the market data.

5. Model Properties

5.1. The role of risk factors

Given the model parameter estimates in Table 5, we can construct the IV surface based on different risk levels to investigate how risk factors affect the pricing of VIX options. I first explore the impact of risk factors on IV along the moneyness dimension. In constructing the IV surface, I set the interest rate at its full-sample average and fix the moneyness, $K/F(t, T)$, from 0.8 to 1.8. Figure 6 plots the impact of each risk factor on IV for short maturity (15 days) and long maturity (150 days), respectively. The solid lines are the mean IV evaluated at the sample average of four risk factors. The dashed (dash-dotted) lines in each panel are generated by setting the specific risk factor at its 90th (10th) percentile value and fixing other factors at their sample mean.

[Insert Figure 6 near here]

Consistent with findings in Mencía and Sentana (2013), Panels A and B in Figure 6 show that risk factor m has a negligible effect on IV once we condition on the current futures price, especially for short maturity. Thus m is relatively more important for pricing VIX futures. The risk factor V dominates the short-term contracts, as shown in Panels C and D. Variations in V lead to slightly uniform shifts in the long-term IV curve but have a significant impact on short-term IV at low strikes. Interestingly, an increase in V elevates the level but flattens the slope of IV. Thus, a single stochastic variance model struggles with generating an IV curve with both high level and steep slope. In contrast, the impact of C is mainly on long-term contracts, as displayed in Panel F. Compared with Panel C, Panel E reflects that C plays a similar role in pricing short-term options

as V , but has a relatively smaller effect. Panels G and H show the impact of jump intensity, λ , on short- and long-term IV, respectively. Not surprisingly, high persistent λ is essential for pricing options across all maturities. Since λ controls positive jumps in vix_t and V_t , it is understandable that the jump intensity risk factor has its major impacts on deep out-of-the-money call options, especially for short-term contracts. Specifically, the increment in λ not only increases the level but also steepens the slope of IV. Therefore, including the λ factor results in a much more flexible pricing model.

Next, I exploit the impact of each risk factor on IV along the maturity dimension. I follow the same procedure described above, but plot the IV at different maturities (from 15 days to 150 days). The results are displayed in Figure 7. The left (right) panels plot the impact of each factor on IV (skew) term structure defined as the difference between ATM IV (skew) of short-term options and ATM IV (skew) of long-term options.

[Insert Figure 7 near here]

As shown in Panels A and B, the contribution of m to the term structures continues to be small. Panels C and D illustrate that an increment in V steepens the slope of IV term structure but flattens the slope of skew term structure. In contrast, C has a relatively higher effect on the long-end of IV term structure. More specifically, an increase in C elevates the level and flattens the slope of IV term structure. Furthermore, variations in C lead to relatively uniform shifts in the skew term structure. Finally, Panel G shows that the jump intensity factor λ has a uniform impact on IV term structure. However, as shown in Panel H, λ has a significantly higher impact on the short-end than the long-end of skew term structure. Therefore, an increment in λ not only increases the level but also steepens the slope of skew term structure.

To sum up, the risk factors have distinct yet complementary impacts on IV along the moneyness dimension as well as the term structure dimension. Thus, the flexible combination of risk factors enables the model to closely capture the fluctuations in the characteristics of the VIX IV surface.

5.2. The Independence of Implied Volatility Level and Slope

5.2.1. Evidence from model simulation I

In Section 2, I find that ATM IV and skew of VIX options fluctuate largely independently. This is the main stylized feature that the model needs to accommodate. In this section, I examine whether the preferred model (VVC-DJ) is capable of capturing this feature by conducting the Monte Carlo simulation.¹⁷ I first obtain the distribution of state variables under the physical measure by simulating the VVC-DJ model in Equations (10) - (14). Next, I calculate the model implied ATM IV and skew by feeding states that are randomly generated from the distributions into the VVC-DJ model with the parameters in Table 5.

[Insert Figure 8 near here]

The scatter plot of ATM IV and skew are presented in Figure 8. Panels A and B display the scatter plots of short-term options, and Panels C and D show the plots of long-term options. In Panel A, I first present the scatter plots based on the empirical market data. The plot confirms the previous finding that ATM IV and skew are highly uncorrelated. In other words, the values of implied skew do not depend on the level of ATM IV. In Panel B, I display the simulated result. We can observe that high values of implied skew are just as likely to occur with high levels of ATM IV as they are with low levels of ATM IV. Therefore, the model captures the independence feature reasonably well. Moreover, the plot indicates that the model can match the ranges of ATM IV and skew obtained from the market data.

As shown in Panels C and D, the model generates the similar ranges of ATM IV and skew as those of empirical data. Importantly, the model is able to capture the independence of ATM IV and skew for long-term options as well.

5.2.2. Evidence from model simulation II

In Section 5.1, I analyze the impact of each risk factor on the IV surface and find that risk factors V , C and λ have distinct roles in determining the shape of the IV surface. I next investigate the integral influence of risk factors V , C and λ on IV slope while controlling for IV level. However,

¹⁷I thank an anonymous referee for suggesting this study.

it is difficult to directly control the level of VIX IV as ATM IV is a nonlinear function of risk factors. Therefore, I approximate the overall level of VIX IV by using VVIX.¹⁸

I generate 1000 random points X in 3-dimensional space (V , C and λ) subject to several linear constraints

$$\begin{aligned} \text{lb} &\leq X \leq \text{ub}, \\ \text{Aeq} \cdot X &= \text{Beq}, \end{aligned} \tag{28}$$

where lb and ub denote lower and upper bound of X , respectively. I generate risk factors V , C and λ between their historical 5th and 95th percentiles. Beq is equal to $\overline{\text{VVIX}}^2(\tau) - A(\tau) - Q(\tau)\bar{m}$; $\overline{\text{VVIX}}^2(\tau)$ is calculated by setting risk factors at their respective mean values in Equation (B.7) Appendix B;¹⁹ and \bar{m} denotes the mean of risk factor m_t .²⁰ Finally, Aeq represents vector $[W(\tau), E(\tau), R(\tau)]$ in Equation (B.7). All parameters are calculated based on the results in Table 5.

[Insert Figure 9 near here]

I present the simulated results in Figure 9. Panels A and B plot the simulation for the short-term IV (15 days), and Panels C and D display the simulation for the long-term IV (150 days). Panel A shows a scatter plot of ATM IV (x -axis) and skew (y -axis). The dotted line denotes VVIX. The simulated ATM IV is very close to VVIX. As expected, ATM IV only deviates from VVIX when the skew is high. Importantly, the plot shows that, given the same overall level of IV, the model can generate a wide range of skew by varying different risk factors. This enables the model to capture the independence of the level slope of the IV smile.

¹⁸By definition, VVIX is a kind of mode-free volatility which is calculated by averaging the prices of out-of-the-money VIX calls and puts with weights inversely proportional to the squared strike price. As shown in Appendix B, squared VVIX can be approximated by the risk-neutral expectation of quadratic variation of VIX futures. Importantly, the squared VVIX can be decomposed into a linear contribution of risk factors

$$\begin{aligned} \text{VVIX}^2(t, T) &= \frac{1}{T-t} \int_t^T E^{\mathbb{Q}} [(d \log F(s, T))^2 | \mathcal{F}_t] ds \\ &= A(t, T) + Q(t, T)m_t + W(t, T)V_t + E(t, T)C_t + R(t, T)\lambda_t. \end{aligned}$$

Therefore, we can easily control for the squared VVIX, thus VVIX, by varying risk factors along the surface defined in the above equation.

¹⁹Specifically, $\overline{\text{VVIX}}^2$ is calculated by setting risk factors at their respective mean \bar{m} , \bar{V} , \bar{C} , and $\bar{\lambda}$

$$\overline{\text{VVIX}}^2(\tau) = A(\tau) + Q(\tau)\bar{m} + W(\tau)\bar{V} + E(\tau)\bar{C} + R(\tau)\bar{\lambda},$$

where τ represents time interval, I fix it at 15 days for the short term and 150 days for the long term.

²⁰As shown in Figure 6, risk factor m_t has a negligible effect on IV surface once we condition on the current futures price. Therefore, I fix risk factor m at its historical mean value and investigate the effects of other risk factors.

Panel B shows detailed results for the simulation in 3-D view. There are several noteworthy observations. First, all generated points are in the same plane. Second, the points with high skew appear in the high- λ and low- V space. Third, risk factor C has a little effect on the short-term skew.

Panel C shows the scatter plot of long-term ATM IV and skew. Simulated ATM IV almost overlaps with VVIX as the skew is relatively low for the long-term IV. Furthermore, given the same overall IV level, the model is also able to generate different levels of skew. Panel D shows the simulated skew for all generated points. Similar to the result discovered in Panel B, all points are in the same plane. As expected, the role of C and V reverses for the long-term IV. Specifically, the points with high skew exist in the high- λ and low- C space, whereas V has a negligible impact on the long-term skew.

Overall, the simulation results show that the VVC-DJ model is able to capture the independence of the level and slope of the VIX smile for the short and long maturities. This model property stems from the flexible combination of different risk factors with distinct yet complementary impacts on the IV surface.

5.2.3. Evidence from model fit performance

In addition to the simulation studies, I explore the empirical results of model fit in this section and provide more evidence of the overwhelming flexibility of the VVC-DJ model.

[Insert Figure 10 near here]

In Figure 10, I plot the calibrated volatility skew with both short and long time to maturity on four different days. November 6, 2008, represents a fairly typical day with a high level of volatility but a flat skew, while August 9, 2017, has a high level of volatility and a steep skew. The remaining two days include November 11, 2010, with low volatility level and a flat skew, and January 22, 2014, with low volatility level but a steep skew. Inspection of all panels indicates that both models do well at matching the level, but only VVC-DJ exhibits good fits to the slope of the IV curve. Therefore, the VVC-DJ model almost perfectly captures the independence of the level and slope of the VIX smile.

6. Conclusions

In this paper, I first perform an empirical analysis of VIX derivatives and find that the option-implied skew exhibits significant fluctuations and extremely low correlations with ATM IV. In addition, there exists a highly persistent factor governing both short- and long-term skew. Moreover, the serial correlations of ATM IV and skew increase with the option maturity.

Next, I propose a general affine jump-diffusion pricing model for VIX derivatives to accommodate the features as mentioned above. Mencía and Sentana (2013) show that the stochastic central tendency in VIX level and stochastic variance are crucial in pricing VIX derivatives. Based on their findings, I further introduce an independent jump intensity factor to control positive jumps to generate a wedge between the dynamics of the level and skew of the VIX smile. Furthermore, I include a variance central tendency factor to capture the persistent shocks to VIX variance.

I estimate the models with a comprehensive daily data set which includes VIX options, futures, and VIX index from April 2, 2007, to December 29, 2017. In addition to data abundance and complexity, the valuation of VIX derivatives also involves the use of inverse Fourier transform and solving ODEs, which dramatically increase the calculation burden. Therefore, during estimation, I utilize parallel programming, namely, CUDA, to offload the highly computationally demanding part to the Graphics Processing Unit (GPU). The estimated results show that the preferred model (VVC-DJ) provides the lowest pricing errors for both in- and out-of-sample periods and significantly outperforms all other nested models considering the test statistics. Most importantly, I illustrate that VVC-DJ can accommodate all stylized features in VIX derivatives, and that the variance central tendency and jump intensity factors are both indispensable to the valuation of VIX options.

Finally, I carry out an extensive analysis of the properties of the preferred model. I first show the distinct yet complementary roles of the risk factors in pricing VIX derivatives by conducting a simulation study. Specifically, I construct a simulated IV surface by changing levels of different risk factors with estimated model parameters. Next, I demonstrate the capability of the model to capture the independence feature from multiple perspectives. The results show that the jump intensity factor, together with the stochastic variance factor, can capture the independence feature of short- term options and, together with the variance central tendency factor, is able to describe the independence feature of long-term options.

Appendix A. VIX derivatives pricing

A.1. Valuation for VIX derivatives

I present how to derive VIX derivatives prices based on the most general specification, VVC-DJ.

All other nested models can be derived in a similar approach.

I use $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{Q})$ to denote a complete stochastic basis defined on the risk-neutral measure \mathbb{Q} , under which the VVC-DJ model for the log(VIX), vix_t , is given by

$$dvix_t = \kappa(m_t - vix_t)dt + \sqrt{V_t}dW_t + JdN_t - \lambda_t\mu dt, \quad (\text{A.1})$$

$$dm_t = \kappa_m(\theta_m - m_t)dt + \omega_m\sqrt{m_t}dW_t^m, \quad (\text{A.2})$$

$$dV_t = \kappa_v(C_t - V_t)dt + \omega_v\sqrt{V_t}dW_t^v + J_v dN_t, \quad (\text{A.3})$$

$$dC_t = \kappa_c(\theta_c - C_t)dt + \omega_c\sqrt{C_t}dW_t^c, \quad (\text{A.4})$$

$$d\lambda_t = \kappa_\lambda(\theta_\lambda - \lambda_t)dt + \omega_\lambda\sqrt{\lambda_t}dW_t^\lambda, \quad (\text{A.5})$$

where $(W_t, W_t^m, W_t^v, W_t^c, W_t^\lambda)$ is a five-dimensional Brownian motion. $\rho dt = E[dW_t dW_t^v]$. N_t is the risk-neutral Poisson process with stochastic intensity λ_t . The jump sizes in vix_t and V_t are assumed to be exponentially distributed with means μ and δ . For parsimony, I only consider upward jumps based on the result in Park (2016) that downward jump provides little improvement in the pricing of VIX derivatives.

Next, I define the characteristic function of vix_T under risk-neutral measure as

$$\phi_t(T, u) = \mathbb{E}_t [e^{iu \cdot vix_T}], \quad (\text{A.6})$$

where $u \in \mathbb{R}$ and i denotes imaginary number. Since the model belongs to the affine diffusion-jump specification, I apply the method by Duffie, Pan, and Singleton (2000) to obtain the closed-form solution to $\phi_t(T, u)$ as follows

$$\phi_t(T, u) = e^{-\alpha(\tau) - \beta_{vix}(\tau) \cdot vix_t - \beta_m(\tau) \cdot m_t - \beta_v(\tau) \cdot V_t - \beta_C(\tau) \cdot C_t - \beta_\lambda(\tau) \cdot \lambda_t}, \quad (\text{A.7})$$

where $\tau = T - t$, and the coefficients satisfy the following ODEs

$$\dot{\beta}_{vix}(\tau) = -\kappa\beta_{vix}(\tau), \quad (\text{A.8})$$

$$\dot{\beta}_m(\tau) = \kappa\beta_{vix}(\tau) - \kappa_m\beta_m(\tau) - \frac{1}{2}\omega_m^2\beta_m(\tau)^2, \quad (\text{A.9})$$

$$\dot{\beta}_v(\tau) = -\kappa_v\beta_v(\tau) - \frac{1}{2}\beta_{vix}(\tau)^2 - \beta_{vix}(\tau)\beta_v(\tau)\omega_v\rho - \frac{1}{2}\omega_v^2\beta_v(\tau)^2, \quad (\text{A.10})$$

$$\dot{\beta}_c(\tau) = \kappa_v\beta_v(\tau) - \kappa_c\beta_c(\tau) - \frac{1}{2}\omega_c^2\beta_c(\tau)^2, \quad (\text{A.11})$$

$$\dot{\beta}_\lambda(\tau) = -\mu\beta_{vix}(\tau) - \kappa_\lambda\beta_\lambda(\tau) - \frac{1}{2}\omega_\lambda^2\beta_\lambda(\tau)^2 - \left(\frac{1}{(1 + \mu\beta_{vix}(\tau))(1 + \delta\beta_v(\tau))} - 1 \right), \quad (\text{A.12})$$

$$\dot{\alpha}(\tau) = \kappa_m\theta_m\beta_m(\tau) + \kappa_c\theta_c\beta_c(\tau) + \kappa_\lambda\theta_\lambda\beta_\lambda(\tau), \quad (\text{A.13})$$

subject to $\alpha(0) = \beta_m(0) = \beta_v(0) = \beta_c(0) = \beta_\lambda(0) = 0$, and $\beta_{vix}(0) = -iu$. With the $\phi_t(T, u)$ derived in closed-form, I then use direction integration method to calculate option prices (see details in Appendix A.2).

Let $F(t, T)$ denotes the time- t VIX futures price with a maturity of T . Thus, $F(t, T)$ can be easily calculated by

$$\begin{aligned} F(t, T) &= \mathbb{E}_t [VIX_T] \\ &= \mathbb{E}_t [e^{vix_T}] \\ &= \phi_t(T, -i) \\ &= e^{-\tilde{\alpha}(\tau) - \tilde{\beta}_{vix}(\tau) \cdot vix_t - \tilde{\beta}_m(\tau) \cdot m_t - \tilde{\beta}_v(\tau) \cdot V_t - \tilde{\beta}_C(\tau) \cdot C_t - \tilde{\beta}_\lambda(\tau) \cdot \lambda_t}. \end{aligned} \quad (\text{A.14})$$

For estimation, I also need to develop the log likelihood function for vix_T under physical measure, \mathbb{P} . Based on the market price of risk specification in Equations (6) - (9), I can obtain the \mathbb{P} -dynamics in Equations (10) - (14). The characteristic function, $\phi_t^\mathbb{P}(T, u)$, can be derived in a similar way as $\phi_t(T, u)$

$$\phi_t^\mathbb{P}(T, u) = e^{-\alpha^\mathbb{P}(\tau) - \beta_{vix}^\mathbb{P}(\tau) \cdot vix_t - \beta_m^\mathbb{P}(\tau) \cdot m_t - \beta_v^\mathbb{P}(\tau) \cdot V_t - \beta_C^\mathbb{P}(\tau) \cdot C_t - \beta_\lambda^\mathbb{P}(\tau) \cdot \lambda_t}, \quad (\text{A.15})$$

where $\tau = T - t$, and the coefficients satisfy the following ODEs

$$\dot{\beta}_{vix}^{\mathbb{P}}(\tau) = -\kappa \beta_{vix}^{\mathbb{P}}(\tau), \quad (\text{A.16})$$

$$\dot{\beta}_m^{\mathbb{P}}(\tau) = \kappa \beta_{vix}^{\mathbb{P}}(\tau) - \kappa_m^{\mathbb{P}} \beta_m^{\mathbb{P}}(\tau) - \frac{1}{2} \omega_m^2 \beta_m^{\mathbb{P}}(\tau)^2, \quad (\text{A.17})$$

$$\dot{\beta}_v^{\mathbb{P}}(\tau) = \gamma \beta_{vix}^{\mathbb{P}} - \kappa_v^{\mathbb{P}} \beta_v^{\mathbb{P}}(\tau) - \frac{1}{2} \beta_{vix}^{\mathbb{P}}(\tau)^2 - \beta_{vix}^{\mathbb{P}}(\tau) \beta_v^{\mathbb{P}}(\tau) \omega_v \rho - \frac{1}{2} \omega_v^2 \beta_v^{\mathbb{P}}(\tau)^2, \quad (\text{A.18})$$

$$\dot{\beta}_c^{\mathbb{P}}(\tau) = \kappa_v \beta_v^{\mathbb{P}}(\tau) - \kappa_c^{\mathbb{P}} \beta_c^{\mathbb{P}}(\tau) - \frac{1}{2} \omega_c^2 \beta_c^{\mathbb{P}}(\tau)^2, \quad (\text{A.19})$$

$$\dot{\beta}_{\lambda}^{\mathbb{P}}(\tau) = -\mu \beta_{vix}^{\mathbb{P}}(\tau) - \kappa_{\lambda} \beta_{\lambda}^{\mathbb{P}}(\tau) - \frac{1}{2} \omega_{\lambda}^2 \beta_{\lambda}^{\mathbb{P}}(\tau)^2 - \left(\frac{1}{(1 + \mu^{\mathbb{P}} \beta_{vix}^{\mathbb{P}}(\tau)) (1 + \delta^{\mathbb{P}} \beta_v^{\mathbb{P}}(\tau))} - 1 \right), \quad (\text{A.20})$$

$$\dot{\alpha}^{\mathbb{P}}(\tau) = \kappa_m \theta_m \beta_m^{\mathbb{P}}(\tau) + \kappa_c \theta_c \beta_c^{\mathbb{P}}(\tau) + \kappa_{\lambda} \theta_{\lambda} \beta_{\lambda}^{\mathbb{P}}(\tau), \quad (\text{A.21})$$

subject to $\alpha^{\mathbb{P}}(0) = \beta_m^{\mathbb{P}}(0) = \beta_v^{\mathbb{P}}(0) = \beta_c^{\mathbb{P}}(0) = \beta_{\lambda}^{\mathbb{P}}(0) = 0$, and $\beta_{vix}^{\mathbb{P}}(0) = -iu$.

A.2. GPU Parallel Programming

I briefly introduce how I utilize the Graphics Processing Unit (GPU) with the tools provided by NVIDIA CUDA®. Although fast Fourier transform (FFT) (e.g., Carr and Madan (1999)) is the most popular method in the literature on option pricing, I prefer to use the direction integration (DI) approach to take advantage of GPU. Specifically, the call option can be calculated by

$$C_t(K, T) = \frac{\exp(-\alpha \log(K) - r(T-t))}{\pi} \int_0^\infty \exp(-iu \log(K)) I(u, \alpha) du, \quad (\text{A.22})$$

with an integrand function $I(u, \alpha)$

$$I(u, \alpha) = \frac{\phi_t(T, u - (\alpha + 1)i)}{\alpha^2 + \alpha - u^2 + (2\alpha + 1)iu}. \quad (\text{A.23})$$

In this paper, I set the smoothing parameter, α as 1.2, truncate the upper bound for the integration domain by 300, and use 128 integrand points, resulting in good results.

One advantage of DI over FFT is that I do not need to interpolate option prices over the strikes, leading to a more accurate result with dramatically reduced time. The most important is that with FFT, I have to employ a large number of strikes (e.g., $2^{12} = 4096$), which leads to a broad strike band. However, in most cases, I only need 100 strikes to do the calibration, which means that

only 100 of 4096 calculated option prices are really used. Therefore, with DI, I can use the strike vector method to accelerate computation as I only need to calculate characteristic function once for multiple option prices with different strikes.

To further accelerate the computation, I utilize parallel programming on GPU. The main effort is partitioning the computational requirement into thousands of small computations that can be executed simultaneously. These computations are assigned to thousands of threads of the GPU which are executed concurrently on different cores. The GPU hardware consists of a number of streaming multiprocessors which in turn consist of multiple cores. Threads are organized in blocks, where one or more block runs on a streaming multiprocessor (see, e.g., Cook (2012)).

Therefore, in each day, I map the model computation on M blocks of the GPU, with each block calculating the option price for one data point. The number of threads N in a block is determined by the number of terms in the Gaussian integration (e.g., 128 in this paper).

To speed up the density of vix_T under physical measure computation, I let each block calculates densities of different maturities. Thus the number of blocks is the same as the number of days (e.g., 2007 blocks for in-sample calibration). The number of threads N in a block is determined by the number of terms in the Gaussian integration.

Appendix B. VVIX Index

I present how to derive VVIX index of the CBOE (Chicago Board Options Exchange) based on the most general specification, VVC-DJ. All other nested models can be derived in a similar approach.

I emphasize that the underling of VIX options is VIX futures, not VIX itself. Therefore, squared VVIX can be approximated by the risk-neutral expectation of quadratic variation of VIX futures

$$\begin{aligned} \text{VVIX}^2(t, T) &= \frac{1}{T-t} E^{\mathbb{Q}} [QV_{t,T} | \mathcal{F}_t] \\ &= \frac{1}{T-t} E^{\mathbb{Q}} \left[\int_t^T (d \log F(s, T))^2 ds | \mathcal{F}_t \right] \\ &= \frac{1}{T-t} \int_t^T E^{\mathbb{Q}} \left[(d \log F(s, T))^2 | \mathcal{F}_t \right] ds, \end{aligned} \quad (\text{B.1})$$

where $F(s, T)$ is the VIX futures price at time s with maturity at time T .

Based on the result in Equation A.14, we can write

$$\begin{aligned} d \log F(t, T) &= d\tilde{\alpha}(\tau) + d\tilde{\beta}_v(\tau) \cdot vix_t + d\tilde{\beta}_m(\tau) \cdot m_t \\ &\quad + d\tilde{\beta}_V(\tau) \cdot V_t + d\tilde{\beta}_C(\tau) \cdot C_t + d\tilde{\beta}_\lambda(\tau) \cdot \lambda_t. \end{aligned} \quad (\text{B.2})$$

Based on Itô lemma, I get

$$\begin{aligned} d\tilde{\beta}_v(\tau) \cdot vix_t &\approx \frac{d}{dt} \tilde{\beta}_v(\tau) + \tilde{\beta}_v(\tau) dvix_t, \\ d\tilde{\beta}_m(\tau) \cdot m_t &\approx \frac{d}{dt} \tilde{\beta}_m(\tau) + \tilde{\beta}_m(\tau) dm_t, \\ d\tilde{\beta}_V(\tau) \cdot V_t &\approx \frac{d}{dt} \tilde{\beta}_V(\tau) + \tilde{\beta}_V(\tau) dV_t, \\ d\tilde{\beta}_C(\tau) \cdot C_t &\approx \frac{d}{dt} \tilde{\beta}_C(\tau) + \tilde{\beta}_C(\tau) dC_t, \\ d\tilde{\beta}_\lambda(\tau) \cdot \lambda_t &\approx \frac{d}{dt} \tilde{\beta}_\lambda(\tau) + \tilde{\beta}_\lambda(\tau) d\lambda_t. \end{aligned} \quad (\text{B.3})$$

This yields

$$\begin{aligned}
(d \log F(t, T))^2 &= \tilde{\beta}_v(\tau)^2 dvix_t^2 + \tilde{\beta}_m(\tau)^2 dm_t^2 + \tilde{\beta}_V(\tau)^2 dV_t^2 + \tilde{\beta}_C(\tau)^2 dC_t^2 + \tilde{\beta}_\lambda(\tau)^2 d\lambda_t^2 \\
&\quad + 2\tilde{\beta}_v(\tau)\tilde{\beta}_m dvix_t \cdot dm_t + 2\tilde{\beta}_v(\tau)\tilde{\beta}_V dvix_t \cdot dV_t \\
&\quad + 2\tilde{\beta}_v(\tau)\tilde{\beta}_C dvix_t \cdot dC_t + 2\tilde{\beta}_v(\tau)\tilde{\beta}_\lambda dvix_t \cdot dm_\lambda \\
&\quad + 2\tilde{\beta}_m(\tau)\tilde{\beta}_V dm_t \cdot dV_t + 2\tilde{\beta}_m(\tau)\tilde{\beta}_C dm_t \cdot dC_t + 2\tilde{\beta}_m(\tau)\tilde{\beta}_\lambda dm_t \cdot d\lambda_t \\
&\quad + 2\tilde{\beta}_V(\tau)\tilde{\beta}_C dV_t \cdot dC_t + 2\tilde{\beta}_V(\tau)\tilde{\beta}_\lambda dV_t \cdot d\lambda_t + 2\tilde{\beta}_C(\tau)\tilde{\beta}_\lambda dC_t \cdot d\lambda_t.
\end{aligned} \tag{B.4}$$

According to the risk-neutral dynamics in Equations (1) - (5), I obtain

$$\begin{aligned}
E_t^{\mathbb{Q}} \left[(d \log F(s, T))^2 | \mathcal{F}_t \right] &= (\beta_v^2(s) + \beta_V^2(s)\omega_V^2 + 2\beta_v(s)\beta_V(s)\rho\omega_V) E_t^{\mathbb{Q}} [V_s] \\
&\quad + (\beta_m^2(s)\omega_m^2) E_t^{\mathbb{Q}} [m_s] + (\beta_c^2(s)\omega_c^2) E_t^{\mathbb{Q}} [C_s] \\
&\quad + (2\beta_v^2(s)\mu^2 + 2\beta_V^2(s)\delta^2 + \beta_\lambda^2(s)\omega_\lambda^2) E_t^{\mathbb{Q}} [\lambda_s],
\end{aligned} \tag{B.5}$$

where

$$\begin{aligned}
E_t^{\mathbb{Q}} [m_s] &= e^{-\kappa_m(s-t)} m_t + (1 - e^{-\kappa_m(s-t)}) \theta_m, \\
E_t^{\mathbb{Q}} [C_s] &= e^{-\kappa_C(s-t)} C_t + (1 - e^{-\kappa_C(s-t)}) \theta_C, \\
E_t^{\mathbb{Q}} [\lambda_s] &= e^{-\kappa_\lambda(s-t)} \lambda_t + (1 - e^{-\kappa_\lambda(s-t)}) \theta_\lambda, \\
E_t^{\mathbb{Q}} [V_s] &= e^{-\kappa_v(s-t)} V_t + \frac{\kappa_v}{\kappa_v - \kappa_c} (e^{-\kappa_c(s-t)} - e^{-\kappa_v(s-t)}) C_t \\
&\quad + \left(1 - e^{-\kappa_v(s-t)} - \frac{\kappa_v}{\kappa_v - \kappa_c} (e^{-\kappa_c(s-t)} - e^{-\kappa_v(s-t)}) \right) \theta_c \\
&\quad + \frac{\delta}{\kappa_v - \kappa_\lambda} (e^{-\kappa_\lambda(s-t)} - e^{-\kappa_v(s-t)}) \lambda_t \\
&\quad + \left(1 - e^{-\kappa_v(s-t)} - \frac{\kappa_v}{\kappa_v - \kappa_\lambda} (e^{-\kappa_\lambda(s-t)} - e^{-\kappa_v(s-t)}) \right) \frac{\delta}{\kappa_v} \theta_\lambda.
\end{aligned} \tag{B.6}$$

Finally, I numerically integrate Equation (B.5) from t to T , resulting in

$$\begin{aligned}
\text{VVIX}^2(t, T) &= \frac{1}{T-t} \int_t^T E_t^{\mathbb{Q}} \left[(d \log F(s, T))^2 | \mathcal{F}_t \right] ds \\
&= A(t, T) + Q(t, T)m_t + W(t, T)V_t + E(t, T)C_t + R(t, T)\lambda_t.
\end{aligned} \tag{B.7}$$

Appendix C. The Filtering Procedure

I estimate the model parameters by maximizing the joint likelihood of the VIX derivatives (options and futures) and returns. To construct the likelihood function, I cast the model into a state-space form and extract the distributions of the states at each date using a filtering technique. In the application, the measurement equation in Equation (21) is nonlinear. Therefore, I use the unscented Kalman filter (UKF) (Wan and Merwe (2000)) to deal with the nonlinearity.

Let \bar{X}_t , $\bar{\Sigma}_{X,t}$, \bar{y}_t and $\bar{\Sigma}_{y,t}$ be the time- $(t-1)$ ex-ante forecasts of time- t values of the latent state, the covariance matrix of the state, the measurement series, and the covariance matrix of the measurement series, respectively, and let \hat{X}_t and $\hat{\Sigma}_{X,t}$ denote the ex post updates on the state vector and the state covariance based on observations y_t at time t . UKF uses a set of deterministically chosen sigma points and then propagate these sigma points through the nonlinear measurement equation.

Specifically, let p be the number of states, $\delta > 0$ be a control parameter, and Σ_i be the i th column of a matrix Σ . A set of $2p+1$ sigma vectors χ_i are generated based on the conditional mean and covariance forecasts on the state vector according the equations

$$\begin{aligned}\chi_{0,t-1} &= \bar{X}_{t-1} \\ \chi_{i,t-1} &= \bar{X}_{t-1} + \left(\sqrt{(p+\delta)(\bar{\Sigma}_{X,t-1})} \right)_i \quad i = 1, 2, \dots, p, \\ \chi_{i,t-1} &= \bar{X}_{t-1} - \left(\sqrt{(p+\delta)(\bar{\Sigma}_{X,t-1})} \right)_i \quad i = p+1, p+2, \dots, 2p,\end{aligned}\tag{C.1}$$

where $\delta = \omega^2(p+\nu) - p$ is a scaling parameter. ν is usually set at 0, and the usual choice for ω is a small positive number between 10^{-4} and 1. I propagate the states through propagation equation in Equation (18) (denoted by $\mathbf{F}(\cdot)$) using the sigma points and approximate the time-updated mean and covariance of the states as follows

$$\begin{aligned}\chi_{i,t} &= \mathbf{F}(\chi_{i,t-1}), \quad \bar{X}_t = \sum_{i=0}^{2p} W_i^{(m)} \chi_{i,t} \\ \bar{\Sigma}_{X,t} &= \sum_{i=0}^{2p} W_i^{(c)} (\chi_{i,t} - \bar{X}_t) (\chi_{i,t} - \bar{X}_t)^\top + \mathbf{Q}_{t-1}.\end{aligned}\tag{C.2}$$

The forecasted mean and covariance of the measurements are given by

$$\begin{aligned}\mathcal{Y}_{i,t} &= \mathbf{H}(\chi_{i,t}), \quad \bar{y}_t = \sum_{i=0}^{2p} W_i^{(m)} \mathcal{Y}_{i,t}, \\ \bar{\Sigma}_{y,t} &= \sum_{i=0}^{2p} W_i^{(c)} (\mathcal{Y}_{i,t} - \bar{y}_t) (\mathcal{Y}_{i,t} - \bar{y}_t)^\top + \mathbf{R}_{t-1}, \\ \bar{\Sigma}_{xy,t} &= \sum_{i=0}^{2p} W_i^{(c)} (\chi_{i,t} - \bar{X}_t) (\mathcal{Y}_{i,t} - \bar{y}_t)^\top.\end{aligned}\tag{C.3}$$

The weights W_i are given by

$$\begin{aligned}W_0^{(m)} &= \frac{\delta}{p+\delta}, \quad W_0^{(c)} = \frac{\delta}{p+\delta} + 1 - \omega^2 + \beta, \\ W_i^{(m)} &= W_i^{(c)} = \frac{1}{2(p+\delta)}, \quad i = 1, 2, \dots, 2p,\end{aligned}\tag{C.4}$$

where β is the tuning variable that incorporates prior information of the distribution of the states ($\beta = 2$ is optimal for Gaussian distributions). The sigma points can be regarded as discrete distributions with probabilities W_i . As in the Kalman filter, the measurement-update equation is

$$\hat{X}_t = \bar{X}_t + \mathcal{K}_t (y_t - \bar{y}_t), \quad \hat{\Sigma}_{X,t} = \bar{\Sigma}_{X,t} - \mathcal{K}_t \bar{\Sigma}_{y,t} \mathcal{K}_t^\top\tag{C.5}$$

where the Kalman gain is $\mathcal{K}_t = \bar{\Sigma}_{xy,t} (\bar{\Sigma}_{y,t})^{-1}$. As in Carr and Wu (2007), Bakshi et al. (2008) among others, I make three simplifying assumptions on the covariance matrices \mathbf{Q} and \mathbf{R} . First, \mathbf{R} is assumed to be constant and diagonal, e.g., pricing errors are uncorrelated in the cross-section. Second, I set the volatility of the pricing error equal to σ_f for all futures and σ_o for all options. Furthermore, following Carr and Wu (2017) and Du and Luo (2019), I approximate the distribution of jump process in VIX variance state (V_t) by its quadratic variation.

To quantify the filtering results, I perform similar Monte Carlo exercises as in Christoffersen et al. (2014). I simulate the VVC-DJ model for 2700 days' sample length and generate VIX derivatives mimicking the real data. I fix the starting values of the states at their respective unconditional means. The paths of the states from one typical simulation are shown in Figure C.1. we can see that the filtered states follow closely the unobservable "true" states.

To gauge the goodness of fit, I repeat the simulation for 500 times, and report in Table C.1 the

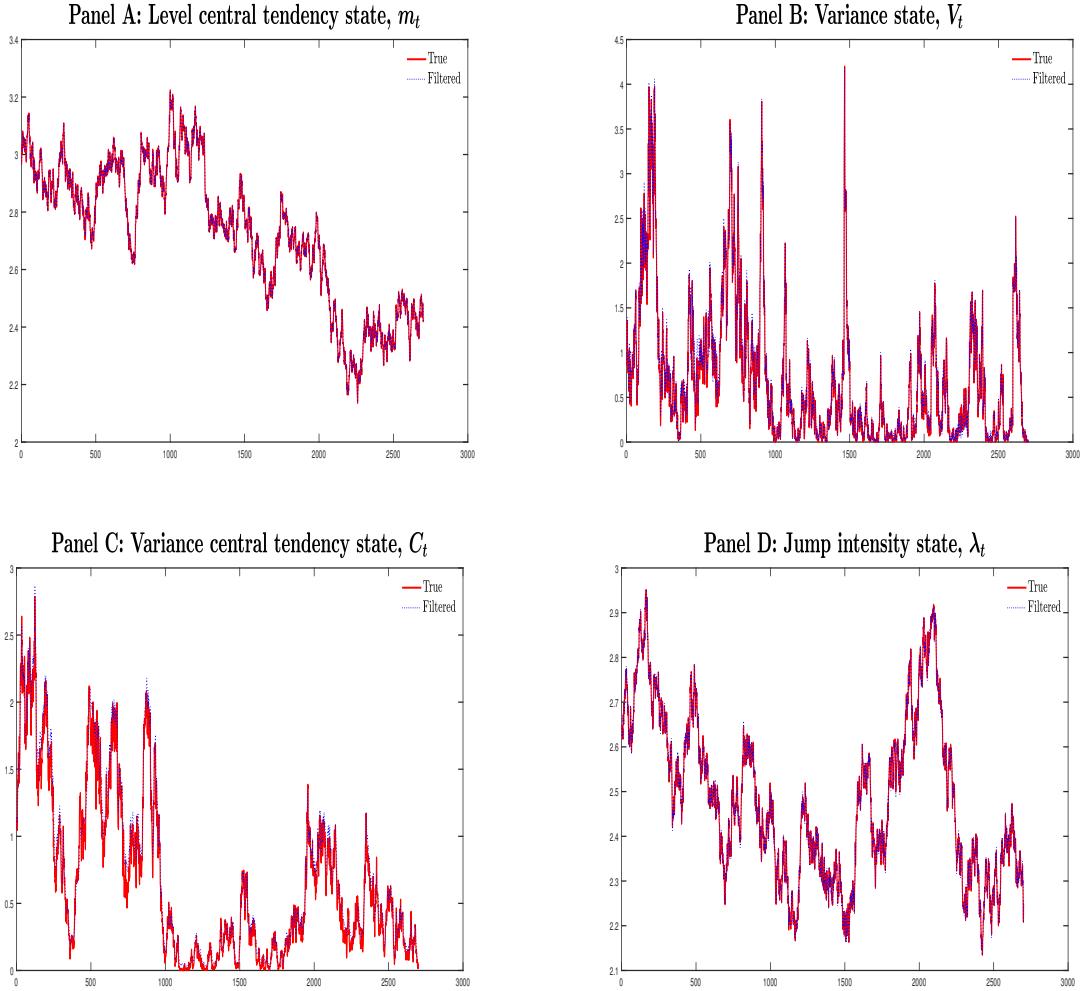


Fig. C.1. This figure shows the performance of the UKF. I use the VVC-DJ model for illustration. I treat the estimates in Table 5 as the "true" parameters, and simulate states' path for 2700 days using Equations (10)-(14). I also fix the starting values of the states at their respective unconditional means. Then, I generate VIX derivatives mimicking the real data. Finally, I apply UKF to extract unobservable states and compare them with "true" ones.

performance of the UKF using the root mean square error (RMSE) over 2700 days for each Monte Carlo simulation

$$\text{RMSRE}_k(i) = \sqrt{\frac{1}{2700} \sum_{t=1}^{2700} (x_{i,p,t} - \hat{x}_{i,k,t})^2}, \quad (\text{C.6})$$

where $x_{i,k,t}$ and $\hat{x}_{i,k,t}$ denote respectively the "true" and filtered state i in simulation k and at time t . After computing the RMSE for each simulation, I report the mean, median, and standard deviation of the RMSE for each state across the 100 simulated samples. Deviations of the filtered states from the "true" ones appear modest. Therefore, UKF works reasonably well in the application.

Table C.1: RMSEs of states filtered from VIX derivatives. I report the mean, median, and standard deviation of the RMSEs for 500 simulated paths under the VVC-DJ model. The "true" model parameters are provided in Table 5.

Factor	mean	median	Std
m	0.0016	0.0016	0.0002
V	0.0429	0.0430	0.0044
C	0.0596	0.0599	0.0053
λ	0.0672	0.0666	0.0064

References

- Aït-Sahalia, Yacine, Mustafa Karaman, and Loriano Mancini, 2018, The term structure of equity and variance risk premia, Working paper.
- Andersen, Torben G., Nicola Fusari, and Viktor Todorov, 2015, The risk premia embedded in index options, *Journal of Financial Economics* 117, 558–584.
- Andrews, Donald, 1991, Heteroskedasticity and autocorrelation consistent covariance matrix estimation, *Econometrica* 59, 817–858.
- Bakshi, Gurdip, Charles Cao, and Zhiwu Chen, 1997, Empirical performance of alternative option pricing models, *Journal of Finance* 52, 2003–2049.
- Bakshi, Gurdip, Charles Cao, and Zhaodong Zhong, 2012, Assessing models of individual equity option prices, Working paper.
- Bakshi, Gurdip, Peter Carr, and Liuren Wu, 2008, Stochastic risk premiums, stochastic skewness in currency options, and stochastic discount factors in international economies, *Journal of Financial Economics* 87, 132–156.
- Bakshi, Gurdip, and Liuren Wu, 2010, The behavior of risk and market prices of risk over the nasdaq bubble period, *Management Science* 56, 2251–2264.
- Baldazzi, Pierluigi, Sanjiv Ranjan Das, and Silverio Foresi, 1998, The central tendency: A second factor in bond yields, *Review of Economics and Statistics* 80, 62–72.
- Bardgett, Chris, Elise Gourier, and Markus Leippold, 2019, Inferring volatility dynamics and risk premia from the s&p 500 and vix markets, *Journal of Financial Economics* 131, 593–618.
- Bekaert, Geert, and Marie Hoerova, 2014, The vix, the variance premium and stock market volatility, *Journal of Econometrics* 183, 181–192.
- Bollerslev, Tim, George Tauchen, and Hao Zhou, 2009, Expected stock returns and variance risk premia, *Review of Financial Studies* 22, 4463–4492.

- Branger, Nicole, Alexander Kraftschik, and Clemens Vlkert, 2016, The fine structure of variance: Pricing vix derivatives in consistent and log-vix models, Working paper.
- Broadie, Mark, Mikhail Chernov, and Michael Johannes, 2007, Model specification and risk premia: Evidence from futures options, *Journal of Finance* 62, 1453–1490.
- Carr, Peter, and Liuren Wu, 2007, Stochastic skew in currency options, *Journal of Financial Economics* 86, 213–247.
- Carr, Peter, and Liuren Wu, 2009, Variance risk premiums, *Review of Financial Studies* 22, 1311–1341.
- Carr, Peter, and Liuren Wu, 2017, Leverage effect, volatility feedback, and self-exciting market disruptions, *Journal of Financial and Quantitative Analysis* 52, 2119–2156.
- Cheng, Jun, Meriton Ibraimi, Markus Leippold, and Jin E. Zhang, 2012, A remark on lin and chang’s paper consistent modeling of s&p 500 and vix derivatives, *Journal of Economic Dynamics and Control* 36, 708–715.
- Christoffersen, Peter, Christian Dorion, Kris Jacobs, and Lotfi Karoui, 2014, Nonlinear kalman filtering in affine term structure models, *Management Science* 60, 2248–2268.
- Christoffersen, Peter, Steven Heston, and Kris Jacobs, 2009, The shape and term structure of the index option smirk: Why multifactor stochastic volatility models work so well, *Management Science* 55, 1914–1932.
- Christoffersen, Peter, Kris Jacobs, and Chayawat Ornthanalai, 2012, Dynamic jump intensities and risk premiums: Evidence from s&p500 returns and options, *Journal of Financial Economics* 106, 447–472.
- Cook, S., 2012, *CUDA Programming: A Developer’s Guide to Parallel Computing with GPUs*, Applications of GPU Computing Series (Elsevier Science).
- Du, Du, and Dan Luo, 2019, The pricing of jump propagation: Evidence from spot and options markets, *Management Science* 65, 2360–2387.

- Duffie, Darrell, Jun Pan, and Kenneth Singleton, 2000, Transform analysis and asset pricing for affine jump-diffusions, *Econometrica*. 68, 1343–1376.
- Eraker, Bjrn, 2004, Do stock prices and volatility jump? reconciling evidence from spot and option prices, *Journal of Finance* 59, 1367–1403.
- Eraker, Bjrn, Michael Johannes, and Nicholas Polson, 2003, The impact of jumps in volatility and returns, *Journal of Finance* 58, 1269–1300.
- Heston, Steven L., 1993, A closed-form solution for options with stochastic volatility with applications to bond and currency options, *Review of Financial Studies* 6, 327–343.
- Hu, Feifang, and James V Zidek, 2002, The weighted likelihood, *Canadian journal of statistics* 30, 347–371.
- Li, Junye, and Gabriele Zinna, 2018, The variance risk premium: Components, term structures, and stock return predictability, *Journal of Business & Economic Statistics* 36, 411–425.
- Mencía, Javier, and Enrique Sentana, 2013, Valuation of vix derivatives, *Journal of Financial Economics* 108, 367–391.
- Newey, Whitney K., and Kenneth D. West, 1987, Hypothesis testing with efficient method of moments estimation, *International Economic Review* 28, 777–787.
- Ornthalalai, Chayawat, 2014, Levy jump risk: Evidence from options and returns, *Journal of Financial Economics* 112, 69–90.
- Pan, Jun, 2002, The jump-risk premia implicit in options: evidence from an integrated time-series study, *Journal of Financial Economics*. 63, 3–50.
- Park, Yang-Ho, 2016, The effects of asymmetric volatility and jumps on the pricing of vix derivatives, *Journal of Econometrics* 192, 313–328.
- Schwartz, Eduardo, and James E Smith, 2000, Short-term variations and long-term dynamics in commodity prices, *Management Science* 46, 893–911.
- Trolle, Anders B., and Eduardo S. Schwartz, 2009, Unspanned stochastic volatility and the pricing of commodity derivatives, *Review of Financial Studies* 22, 4423–4461.

Vlkert, Clemens, 2015, The distribution of uncertainty: Evidence from the vix options market, *Journal of Futures Markets* 35, 597–624.

Wan, E. A., and R. Van Der Merwe, 2000, The unscented kalman filter for nonlinear estimation, in *Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (Cat. No.00EX373)*, 153–158.

Zang, Xin, Jun Ni, Jing-Zhi Huang, and Lan Wu, 2017, Double-jump diffusion model for vix: evidence from vvix, *Quantitative Finance* 17, 227–240.

Table 1: VIX options data summary, April 2, 2007 - December 29, 2017 (2701 business days). Daily settlement prices of VIX options quotes are obtained from the OptionMetrics. Moneyness is defined as the strike to future price ratio, K/F . Days-to-maturity (DTM) is in the number of actual days. The implied volatilities are calculated using the Black-Scholes formula.

DTM	< 30	30 – 60	60 – 90	90 – 120	> 120	All
Panel A: Number of option contracts						
$K/F < 0.8$	490	1620	2137	1754	1831	7832
$0.8 < K/F < 0.9$	2077	4167	3772	2562	2181	14759
$0.9 < K/F < 1.0$	3979	4474	3648	2415	2008	16524
$1.0 < K/F < 1.2$	8164	7930	7547	5630	4888	34159
$1.2 < K/F < 1.4$	5582	6807	6374	4937	4603	28303
$1.4 < K/F < 1.6$	3214	5469	5282	4767	5080	23812
$1.6 < K/F < 1.8$	1470	3619	3543	3371	3482	15485
$1.8 < K/F$	734	3271	3745	3548	3916	15214
All	25710	37357	36048	28984	27989	156088
Panel B: Average option prices						
$K/F < 0.8$	0.46	0.49	0.59	0.66	0.73	0.59
$0.8 < K/F < 0.9$	0.60	0.82	1.11	1.28	1.46	1.05
$0.9 < K/F < 1.0$	1.11	1.70	2.12	2.37	2.63	1.99
$1.0 < K/F < 1.2$	1.14	1.77	2.20	2.44	2.68	2.04
$1.2 < K/F < 1.4$	0.64	1.05	1.38	1.59	1.76	1.28
$1.4 < K/F < 1.6$	0.41	0.68	0.93	1.08	1.20	0.86
$1.6 < K/F < 1.8$	0.30	0.48	0.66	0.78	0.86	0.62
$1.8 < K/F$	0.23	0.32	0.43	0.52	0.58	0.41
All	0.61	0.91	1.18	1.34	1.49	1.11
Panel C: Average option implied volatility						
$K/F < 0.8$	1.02	0.76	0.65	0.59	0.54	0.71
$0.8 < K/F < 0.9$	0.85	0.71	0.65	0.61	0.57	0.68
$0.9 < K/F < 1.0$	0.87	0.78	0.71	0.66	0.61	0.73
$1.0 < K/F < 1.2$	1.03	0.89	0.79	0.72	0.66	0.82
$1.2 < K/F < 1.4$	1.22	1.01	0.88	0.79	0.72	0.92
$1.4 < K/F < 1.6$	1.36	1.10	0.94	0.85	0.76	1.00
$1.6 < K/F < 1.8$	1.46	1.19	1.00	0.89	0.80	1.07
$1.8 < K/F$	1.57	1.30	1.10	0.97	0.86	1.16
All	1.17	0.97	0.84	0.76	0.69	0.89

Table 2: Summary statistics for interpolated options implied volatility (IV) surface. The surface data is formed based on five maturities: 30, 60, 90, 120, and 150 days, and five fixed strikes (measured by delta): -20, -35, 50, 35, and 20. Statistics include mean, standard deviation (Std), skewness (Skew), kurtosis (Kurt) and autocorrelation (AC) with different lags.

Maturity	Delta	Mean	Std	Skew	Kurt	AC(1)	AC(15)	AC(30)
30	-20	0.7152	0.1424	1.0928	6.0612	0.9057	0.4581	0.2620
	-35	0.7923	0.1490	1.2013	6.7262	0.9180	0.4466	0.2394
	50	0.8807	0.1565	1.0984	6.1991	0.9260	0.4634	0.2584
	35	0.9850	0.1666	0.7940	5.1052	0.9265	0.5138	0.3391
	20	1.1152	0.1802	0.3668	4.0306	0.9409	0.5935	0.4541
60	-20	0.6367	0.0940	0.5754	3.6099	0.9499	0.5994	0.3737
	-35	0.7098	0.0968	0.4034	3.3122	0.9574	0.6027	0.3897
	50	0.7858	0.1066	0.1683	2.9704	0.9647	0.6586	0.4769
	35	0.8675	0.1210	-0.1159	2.6597	0.9750	0.7374	0.5936
	20	0.9583	0.1397	-0.2536	2.7385	0.9830	0.8056	0.6950
90	-20	0.5844	0.0745	0.3660	3.1998	0.9626	0.6748	0.4509
	-35	0.6492	0.0767	0.0237	2.7256	0.9716	0.6918	0.4874
	50	0.7131	0.0863	-0.1908	2.5825	0.9779	0.7509	0.5857
	35	0.7777	0.0999	-0.3428	2.5846	0.9853	0.8145	0.6836
	20	0.8451	0.1149	-0.3597	2.8075	0.9891	0.8594	0.7553
120	-20	0.5510	0.0648	0.2722	3.0773	0.9679	0.7140	0.4916
	-35	0.6097	0.0665	-0.1049	2.6229	0.9768	0.7328	0.5256
	50	0.6654	0.0749	-0.2734	2.5738	0.9823	0.7852	0.6151
	35	0.7187	0.0860	-0.3758	2.6792	0.9873	0.8327	0.6938
	20	0.7724	0.0980	-0.3784	2.9223	0.9881	0.8642	0.7511
150	-20	0.5333	0.0609	0.1856	3.0131	0.9681	0.7421	0.5324
	-35	0.5887	0.0626	-0.1578	2.6583	0.9774	0.7580	0.5586
	50	0.6403	0.0701	-0.2776	2.6503	0.9821	0.8011	0.6341
	35	0.7034	0.0835	-0.3670	2.8185	0.9844	0.8452	0.7126
	20	0.7351	0.0911	-0.3689	2.9564	0.9848	0.8630	0.7476

Table 3: Principal Component Analysis. This table reports the factors loadings and variance explained by the first four principal components. The PCA is performed on an interpolated implied volatility surface. The surface data is formed based on five maturities: 30, 60, 90, 120, and 150 days, and five fixed strikes (measured by delta): -20, -35, 50, 35, and 20.

Maturity	Delta	Principal components			
		First	Second	Third	Fourth
30	-20	0.1540	0.3661	0.1816	-0.3060
	-35	0.1728	0.3033	0.2549	-0.2285
	50	0.1890	0.2107	0.3219	-0.1665
	35	0.1986	0.0859	0.3703	-0.1410
	20	0.1944	-0.0666	0.3754	-0.1489
60	-20	0.1786	0.3166	-0.0618	0.2420
	-35	0.2058	0.1981	0.0252	0.3071
	50	0.2176	0.0583	0.1049	0.3437
	35	0.2149	-0.0850	0.1561	0.3135
	20	0.1970	-0.2142	0.1765	0.2521
90	-20	0.1912	0.2421	-0.2306	0.1162
	-35	0.2164	0.0975	-0.1334	0.1637
	50	0.2210	-0.0438	-0.0386	0.1766
	35	0.2121	-0.1681	0.0314	0.1498
	20	0.1934	-0.2685	0.0742	0.0866
120	-20	0.1937	0.1914	-0.3122	-0.0534
	-35	0.2159	0.0390	-0.2123	-0.0330
	50	0.2172	-0.0939	-0.1103	-0.0324
	35	0.2079	-0.1999	-0.0358	-0.0502
	20	0.1910	-0.2832	0.0163	-0.1004
150	-20	0.1922	0.1497	-0.3505	-0.2014
	-35	0.2128	-0.0025	-0.2537	-0.1963
	50	0.2124	-0.1272	-0.1521	-0.2034
	35	0.1980	-0.2461	-0.0581	-0.2232
	20	0.1862	-0.2917	-0.0235	-0.2493
Explained by PCs:		79.5758	12.9478	4.9365	0.9108

Table 4: Summary of Model specifications. In the head of the table, CT refers to as the central tendency and JI refers to as the jump intensity. Columns represent the modeling features. The checkmarks indicate the model includes corresponding features.

	Level CT	Variance CT	Constant JI	Stochastic JI	Level Jumps	Variance Jumps
VC	✓					
VC-CJ	✓				✓	
VC-SJ	✓		✓		✓	
VVC-CJ	✓		✓		✓	
VVC-SJ	✓		✓	✓	✓	
VVC-DJ	✓		✓	✓	✓	✓

Table 5: In-sample model parameter estimates. This table reports the maximum likelihood parameter estimates and their standard errors (in parentheses). AIC and BIC stand for Akaike and Bayesian information criteria, respectively.

ϑ	VC	VC-CJ	VC-SJ	VVC-CJ	VVC-SJ	VVC-DJ
κ	10.4117 (0.0959)	9.1393 (0.0875)	9.0856 (0.1161)	11.2470 (0.1748)	11.3630 (0.1413)	10.4387 (0.0436)
κ_m	0.5986 (0.0154)	0.5467 (0.0139)	0.6892 (0.0158)	0.5958 (0.0128)	0.6854 (0.0139)	0.7428 (0.0136)
θ_m	3.2262 (0.0190)	3.2453 (0.0214)	3.2522 (0.0175)	3.3363 (0.0147)	3.2874 (0.0126)	3.2843 (0.0117)
ω_m	0.2519 (0.0043)	0.1736 (0.0056)	0.2803 (0.0047)	0.2181 (0.0083)	0.2113 (0.0046)	0.1880 (0.0050)
κ_v	2.6426 (0.1134)	1.1892 (0.0408)	0.1142 (0.0159)	7.6650 (0.1639)	6.1008 (0.0958)	6.7418 (0.0480)
θ_v	2.2006 (0.0347)	3.2112 (0.0659)	0.9131 (0.0858)			
ω_v	4.2520 (0.0589)	2.9978 (0.0588)	2.4507 (0.0717)	5.9195 (0.0889)	4.3947 (0.0894)	2.9319 (0.0212)
ρ	0.8835 (0.0151)	0.7644 (0.0174)	0.7404 (0.0297)	0.6427 (0.0157)	0.6307 (0.0170)	0.9176 (0.0082)
κ_c				0.1002 (0.0051)	0.1155 (0.0116)	0.2029 (0.0419)
θ_c				2.0929 (0.3669)	0.5071 (0.0830)	3.0253 (0.4135)
ω_c				1.1492 (0.0619)	1.1506 (0.1897)	1.3727 (0.1756)
κ_λ			8.0740 (3.2137)		1.5245 (0.1800)	0.9297 (0.1655)
θ_λ			2.9986 (0.0797)		0.5063 (0.0450)	2.3857 (0.1838)
ω_λ			2.2916 (0.3252)		0.1025 (0.0081)	0.1865 (0.0132)
λ		1.2267 (0.0449)		1.5707 (0.0628)		
μ		0.3490 (0.0047)	0.3255 (0.0029)	0.2889 (0.0038)	0.4990 (0.0026)	0.2216 (0.0049)
δ						2.4723 (0.0252)
κ_m^P	0.7616 (0.0173)	0.7025 (0.0155)	0.9356 (0.0180)	0.8300 (0.0160)	0.8587 (0.0167)	0.9011 (0.0171)
κ_v^P	4.4390 (0.4542)	2.5644 (0.1506)	5.4715 (0.6260)	11.1841 (1.7283)	8.7778 (0.5776)	9.0433 (0.7298)
κ_c^P				3.2796 (0.1421)	1.8773 (0.1812)	0.5044 (0.2830)
μ^P		0.2960 (0.0422)	0.2436 (0.0141)	0.1744 (0.0585)	0.3414 (0.1042)	0.1914 (0.0382)
δ^P						2.4326 (0.1350)
γ	-0.0108 (0.0066)	-0.0289 (0.0084)	-0.0342 (0.0190)	-0.0337 (0.0313)	-0.0338 (0.0224)	-0.0340 (0.0236)
σ_o	0.0661 (0.0011)	0.0628 (0.0011)	0.0573 (0.0010)	0.0529 (0.0009)	0.0497 (0.0008)	0.0463 (0.0008)
σ_f	0.0275 (0.0017)	0.0275 (0.0012)	0.0259 (0.0018)	0.0260 (0.0016)	0.0248 (0.0011)	0.0236 (0.0013)
Log(L)	3452.99	3562.51	3657.37	3779.81	3925.13	4016.45
AIC	-6879.98	-7093.02	-7278.73	-7521.62	-7808.26	-7986.90
BIC	-6902.68	-7121.72	-7311.43	-7556.32	-7846.96	-8029.60

Table 6: In-sample Root Mean Square Relative Errors (RMSREs) on the options implied volatilities. The statistics are computed from daily VIX derivatives data from April 2, 2007, to March 31, 2015 (2007 business days). Moneyness is measured as the strike to futures price ratio, K/F . DTM denotes days to maturity. In Panel D, I report the pairwise t -statistics (defined in Equation (27)) in a (6×6) matrix, with the (i, j) th element denoting the statistic on model i versus model j . A negative (positive) statistic in entries (i, j) indicates that model i outperforms (underperforms) model j . I bold the lower triangular elements whose absolute value is greater than 2.33, the critical value at 99% confidence level.

	VC	VC-CJ	VC-SJ	VVC-CJ	VVC-SJ	VVC-DJ
Panel A: Overall						
	0.0722	0.0684	0.0698	0.0645	0.0543	0.0536
Panel B: Sorting by moneyness						
$K/F < 0.8$	0.1062	0.0925	0.0852	0.0922	0.0967	0.0794
$0.8 < K/F < 0.9$	0.0916	0.0890	0.0831	0.0810	0.0595	0.0684
$0.9 < K/F < 1.0$	0.0811	0.0757	0.0750	0.0646	0.0505	0.0617
$1.0 < K/F < 1.2$	0.0598	0.0578	0.0613	0.0532	0.0489	0.0490
$1.2 < K/F < 1.4$	0.0679	0.0666	0.0715	0.0615	0.0531	0.0490
$1.4 < K/F < 1.6$	0.0677	0.0655	0.0705	0.0638	0.0513	0.0479
$1.6 < K/F < 1.8$	0.0634	0.0591	0.0621	0.0600	0.0468	0.0447
$1.8 < K/F$	0.0671	0.0606	0.0610	0.0631	0.0479	0.0457
Panel C: Sorting by time to maturity						
$DTM < 30$	0.0893	0.0822	0.0845	0.0763	0.0670	0.0745
$30 < DTM < 60$	0.0685	0.0675	0.0665	0.0654	0.0486	0.0509
$60 < DTM < 90$	0.0684	0.0651	0.0638	0.0582	0.0486	0.0427
$90 < DTM < 120$	0.0680	0.0643	0.0669	0.0588	0.0506	0.0454
$DTM > 120$	0.0685	0.0638	0.0694	0.0648	0.0591	0.0547
Panel D: Pairwise t -statistics						
VC	0.0000	7.5088	3.5248	6.7716	9.3069	10.2119
VC-CJ	-7.5088	0.0000	-1.6356	4.0658	7.0933	8.0375
VC-SJ	-3.5248	1.6356	0.0000	5.4735	7.9318	9.9975
VVC-CJ	-6.7716	-4.0658	-5.4735	0.0000	6.8134	8.0348
VVC-SJ	-9.3069	-7.0933	-7.9318	-6.8134	0.0000	3.4492
VVC-DJ	-10.2119	-8.0375	-9.9975	-8.0348	-3.4492	0.0000

Table 7: Out-of-sample Root Mean Square Relative Errors (RMSREs) on the options implied volatilities. The statistics are computed from daily VIX derivatives data from April 1, 2015, to December 29, 2017. Moneyness is measured as the strike to futures price ratio, K/F . DTM denotes days to maturity. In Panel D, I report the pairwise t -statistics (defined in Equation (27)) in a (6×6) matrix, with the (i, j) th element denoting the statistic on model i versus model j . A negative (positive) statistic in entries (i, j) indicates that model i outperforms (underperforms) model j . I bold the lower triangular elements whose absolute value is greater than 2.33, the critical value at 99% confidence level.

	VC	VC-CJ	VC-SJ	VVC-CJ	VVC-SJ	VVC-DJ
Panel A: Overall						
	0.1077	0.1002	0.1025	0.0901	0.0693	0.0689
Panel B: Sorting by moneyness						
$K/F < 0.8$	0.1250	0.1122	0.1057	0.1211	0.1359	0.1036
$0.8 < K/F < 0.9$	0.1315	0.1300	0.1153	0.1230	0.0821	0.1016
$0.9 < K/F < 1.0$	0.1278	0.1205	0.1152	0.1009	0.0709	0.0841
$1.0 < K/F < 1.2$	0.1029	0.0971	0.1069	0.0826	0.0697	0.0728
$1.2 < K/F < 1.4$	0.0843	0.0884	0.0970	0.0787	0.0615	0.0550
$1.4 < K/F < 1.6$	0.0906	0.0911	0.0974	0.0823	0.0578	0.0503
$1.6 < K/F < 1.8$	0.0974	0.0899	0.0940	0.0811	0.0511	0.0454
$1.8 < K/F$	0.1015	0.0865	0.0878	0.0770	0.0439	0.0443
Panel C: Sorting by time to maturity						
$DTM < 30$	0.1436	0.1332	0.1203	0.1182	0.0932	0.0941
$30 < DTM < 60$	0.1201	0.0942	0.0850	0.0874	0.0569	0.0640
$60 < DTM < 90$	0.1033	0.0938	0.0978	0.0832	0.0623	0.0601
$90 < DTM < 120$	0.1000	0.0935	0.1073	0.0823	0.0643	0.0609
$DTM > 120$	0.0917	0.0865	0.1067	0.0798	0.0720	0.0663
Panel D: Pairwise t -statistics						
VC	0.0000	8.5779	4.3986	6.5287	16.1786	17.3184
VC-CJ	-8.5779	0.0000	-1.2031	4.4551	15.1643	16.1526
VC-SJ	-4.3986	1.2031	0.0000	8.9049	15.7802	16.8733
VVC-CJ	-6.5287	-4.4551	-8.9049	0.0000	11.4638	12.8429
VVC-SJ	-16.1786	-15.1643	-15.7802	-11.4638	0.0000	3.6710
VVC-DJ	-17.3184	-16.1526	-16.8733	-12.8429	-3.6710	0.0000

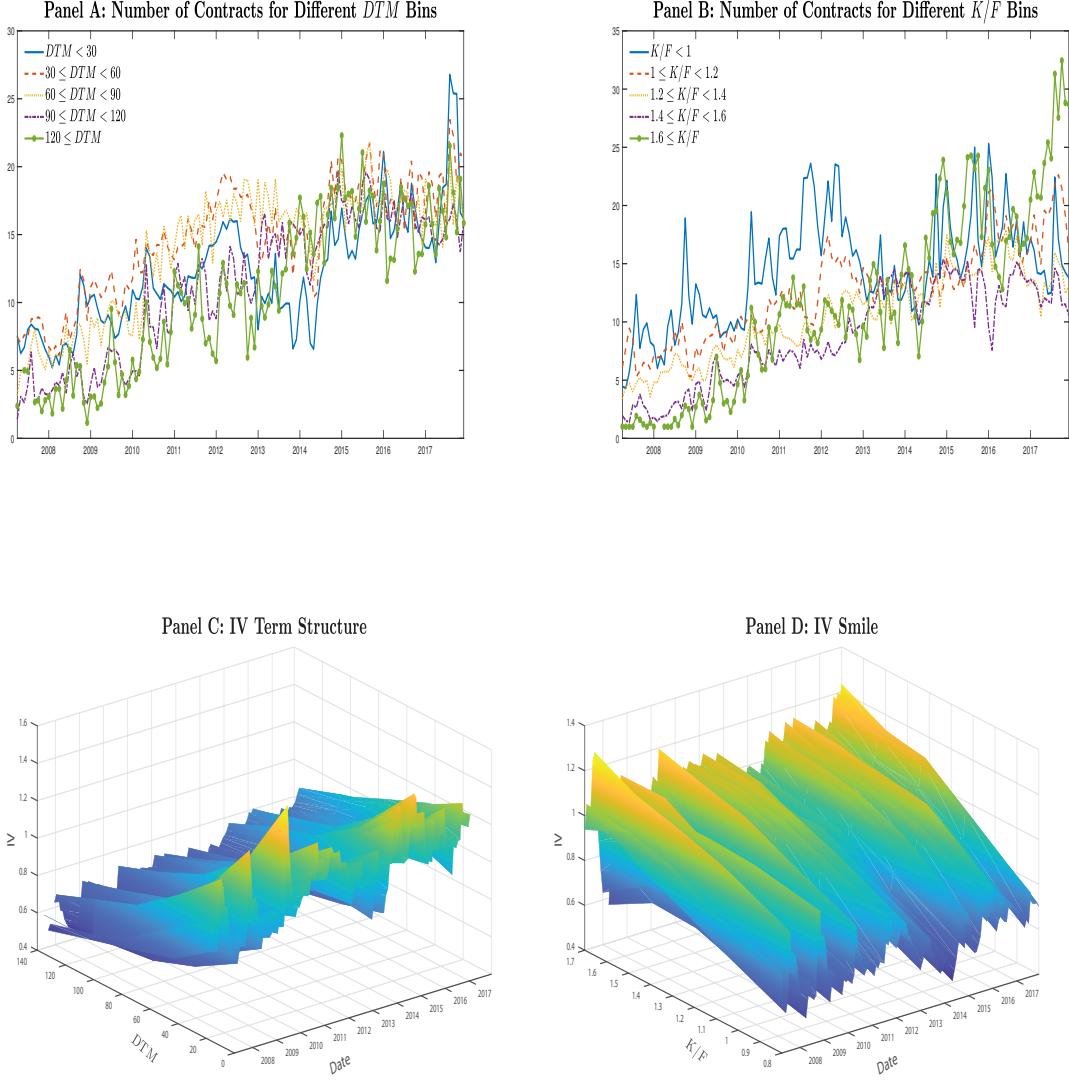


Fig. 1. This figure shows the time-series plots of VIX options data from April 2, 2007, to December 29, 2017, in monthly frequency. Panel A plots the number of contracts for different days-to-maturity bins. Panel B plots the number of contracts for different moneyness bins. Panels C and D plot VIX implied volatility (IV) slope along with the term structure and the moneyness dimension, respectively. Moneyness is defined as the strike to future price ratio, K/F . Days-to-maturity (DTM) is in the number of actual days. The implied volatilities are calculated using the Black-Scholes formula.

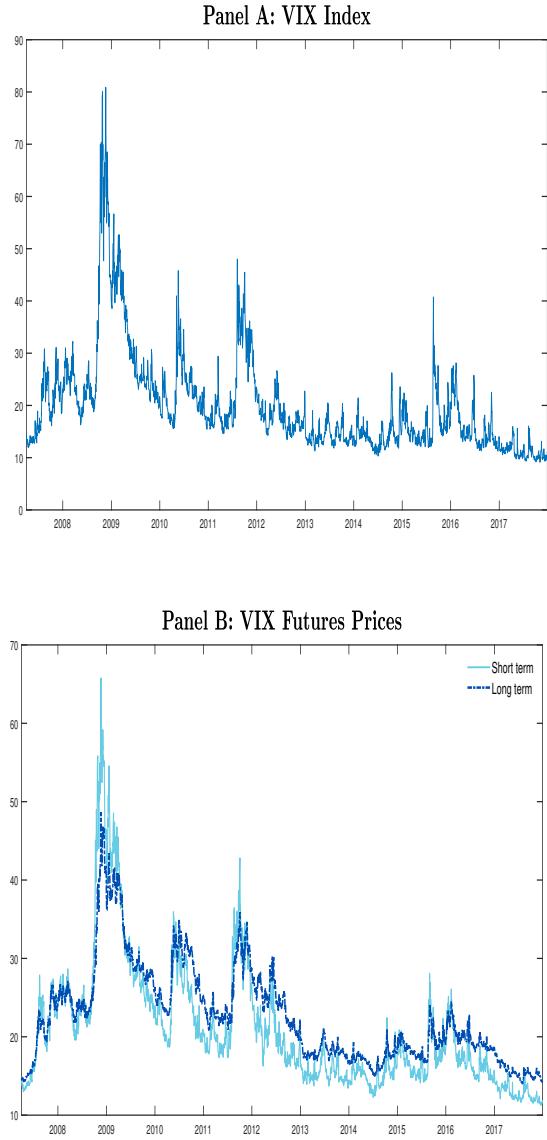


Fig. 2. This figure plots the VIX index and VIX futures prices from April 2, 2007, to December 29, 2017. Panel A plots the time series of VIX index. Panel B compares the dynamics of VIX short- and long-term futures prices. Short-term contracts are those with days to maturity < 45 days, and long-term contracts are those with days to maturity > 120 days.

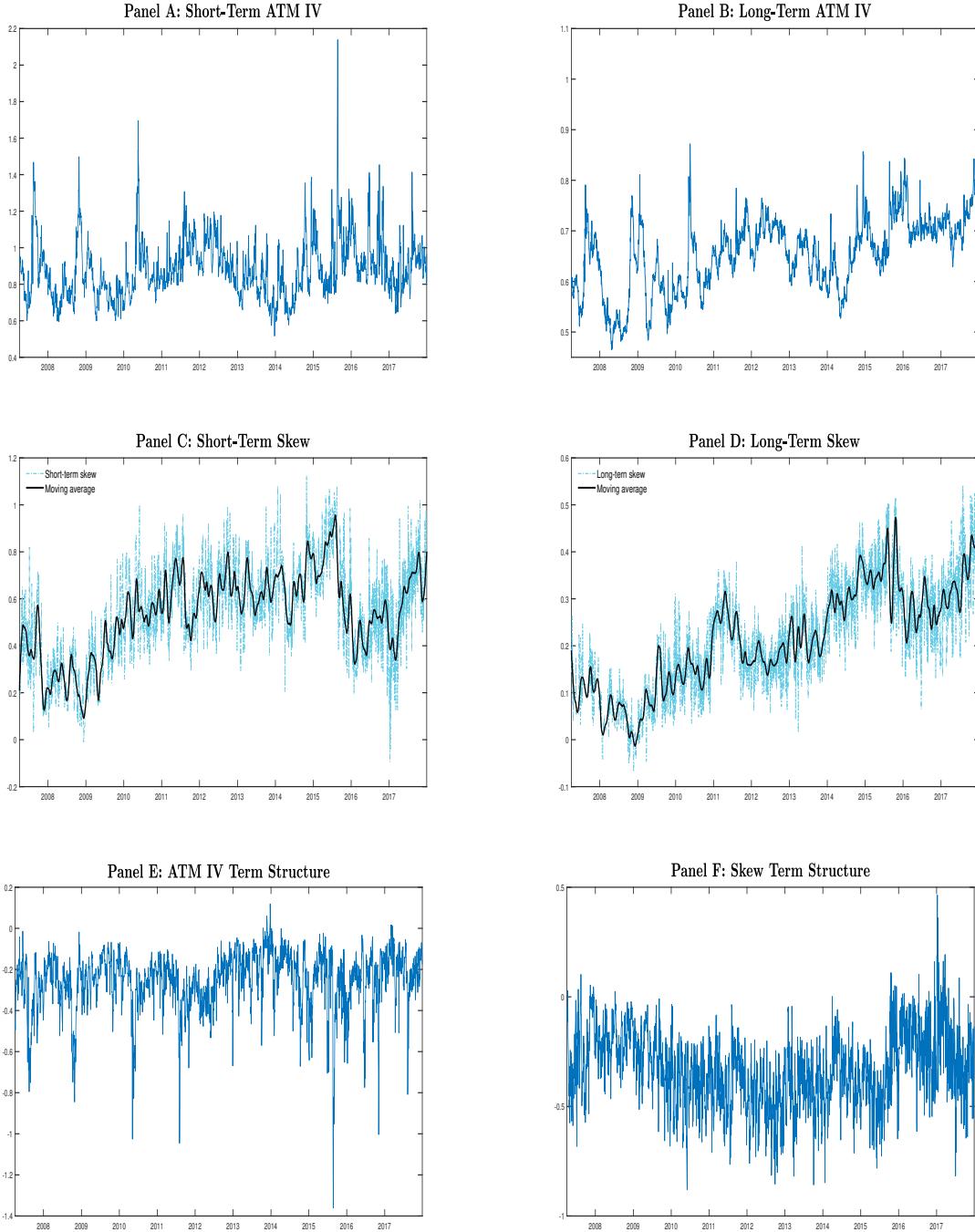


Fig. 3. This figure plots the characteristics of VIX implied volatility (IV) surface. Panels A and B plot the average IV for short-term and long-term at-the-money (ATM) options, respectively. Panels C and D show the IV skew for short and long maturities, respectively. IV skew is defined as the difference between the IVs of out-of-the-money (OTM) call and OTM put options. In Panel E, the ATM IV term structure is the difference between long- and short-dated ATM IVs. In Panel F, the skew term structure is the difference between the long- and short-dated IV skew. Short-term (long-term) options are those with days to maturity < 45 (> 120) days. ATM options have moneyness, K/F , between 0.9 and 1.1. OTM call (put) options have $K/F > 1.8$ (< 0.8).

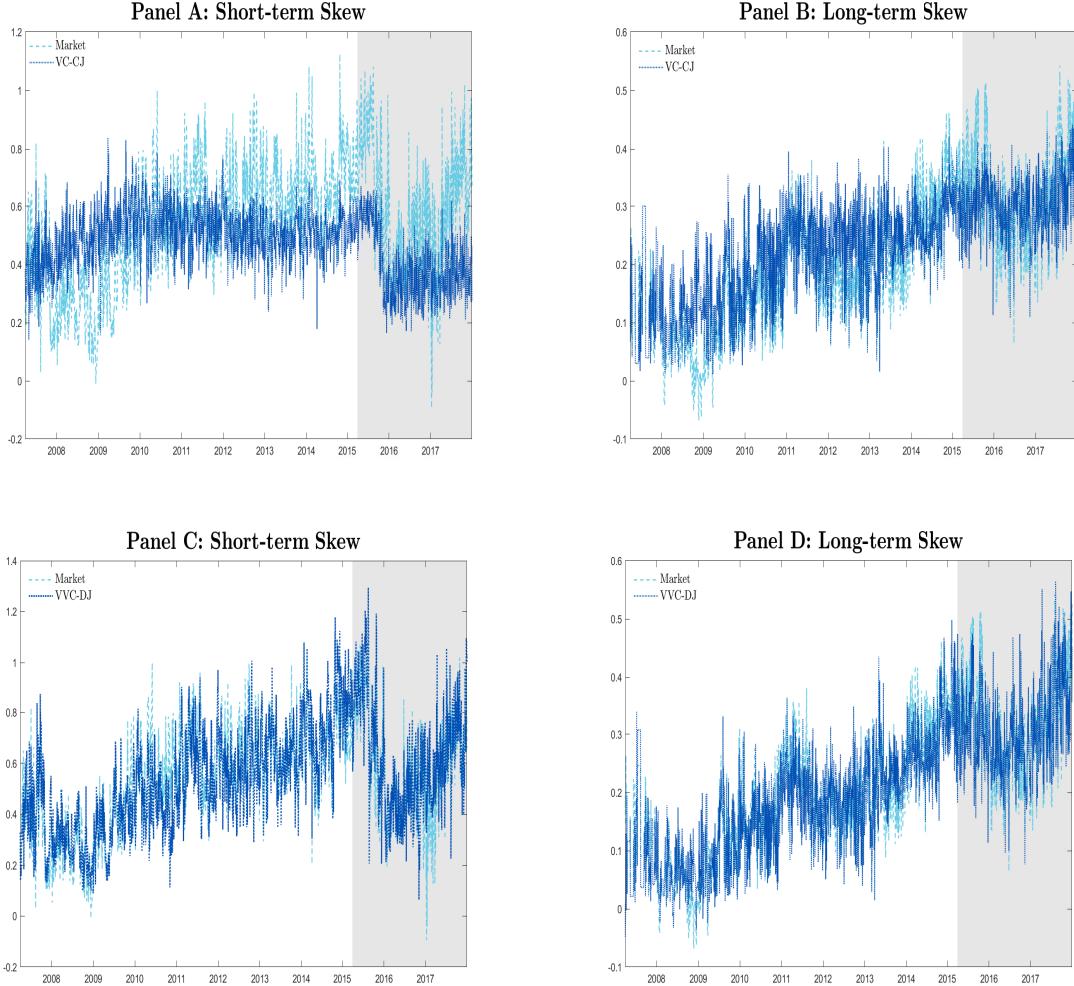


Fig. 4. This figure shows the model-implied volatility skew. The dashed lines correspond to the market-implied volatility skew, and the dotted lines refer to the fit by models with estimated parameters in Table 5. Panels A and B plot IV skew for the VC-CJ model, and Panels C and D show the results from the VVC-DJ model. The shaded part of the graph represents the out-of-sample period, from April 1, 2015, to December 29, 2017.

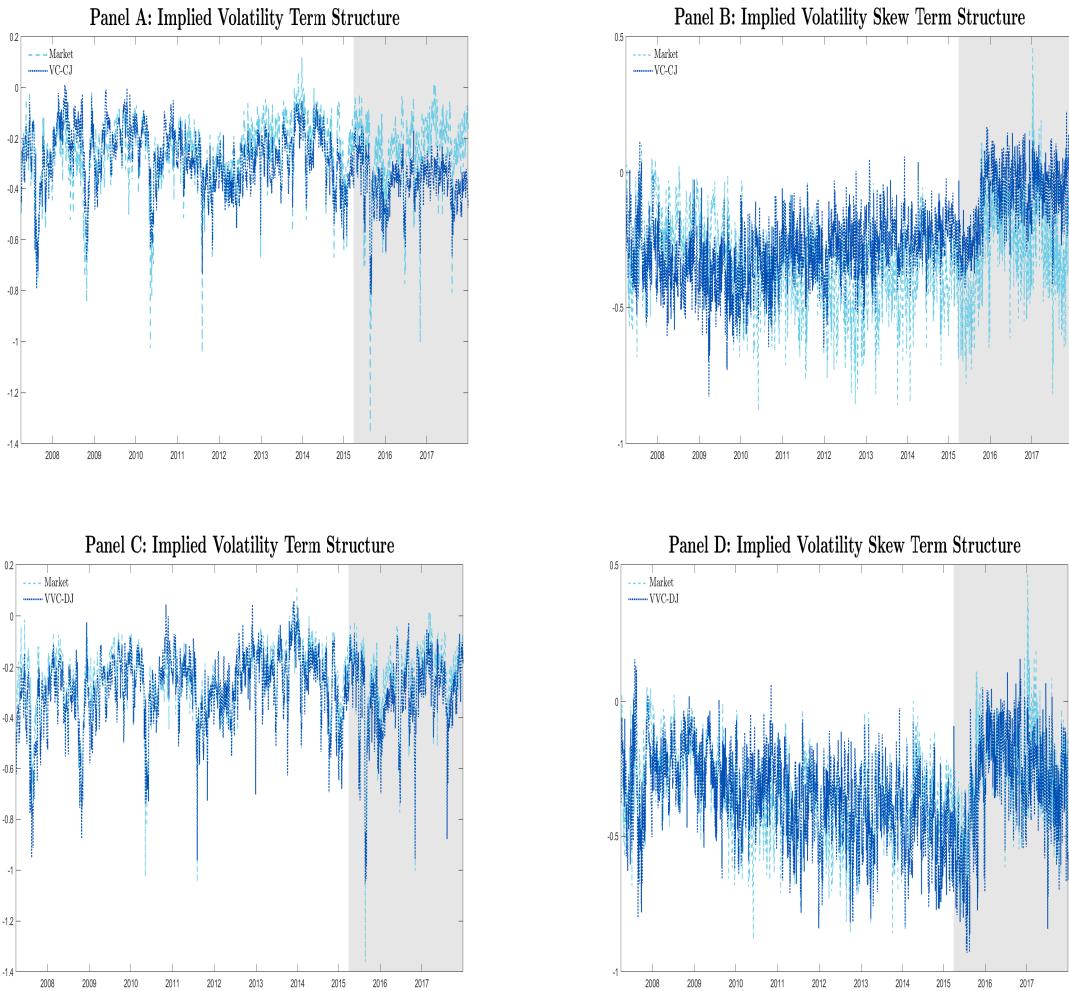


Fig. 5. This figure shows the model-implied volatility and skew term structures. The dashed lines correspond to the market-implied term structure, and the dotted lines refer to the fit by models with estimated parameters in Table 5. Panels A and B plot the fitted term structure for the VC-CJ model, and Panels C and D show the results from the VVC-DJ model. The shaded part of the graph represents the out-of-sample period, from April 1, 2015, to December 29, 2017.

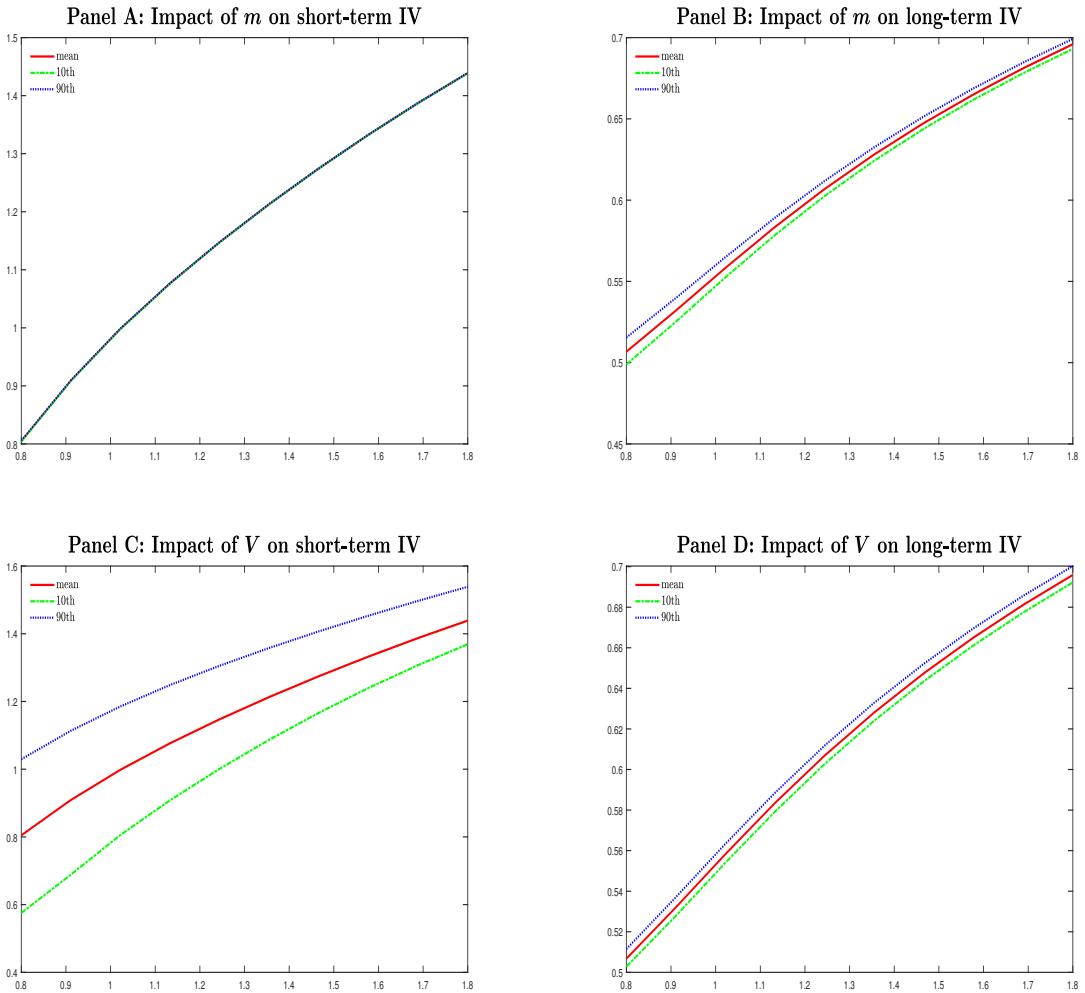


Fig. 6. This figure illustrates the impact of risk factors on implied volatility (IV) surface along the moneyness dimension. The simulated IV surface is calculated based on the VVC-DJ model with parameter estimates in Table 5. The solid lines are mean value of IVs computed by using the sample mean levels of the risk factors. Dotted lines and dashed lines are computed by setting the specific factor at its 90th and 10th percentile values and fixing other risk factors, respectively.

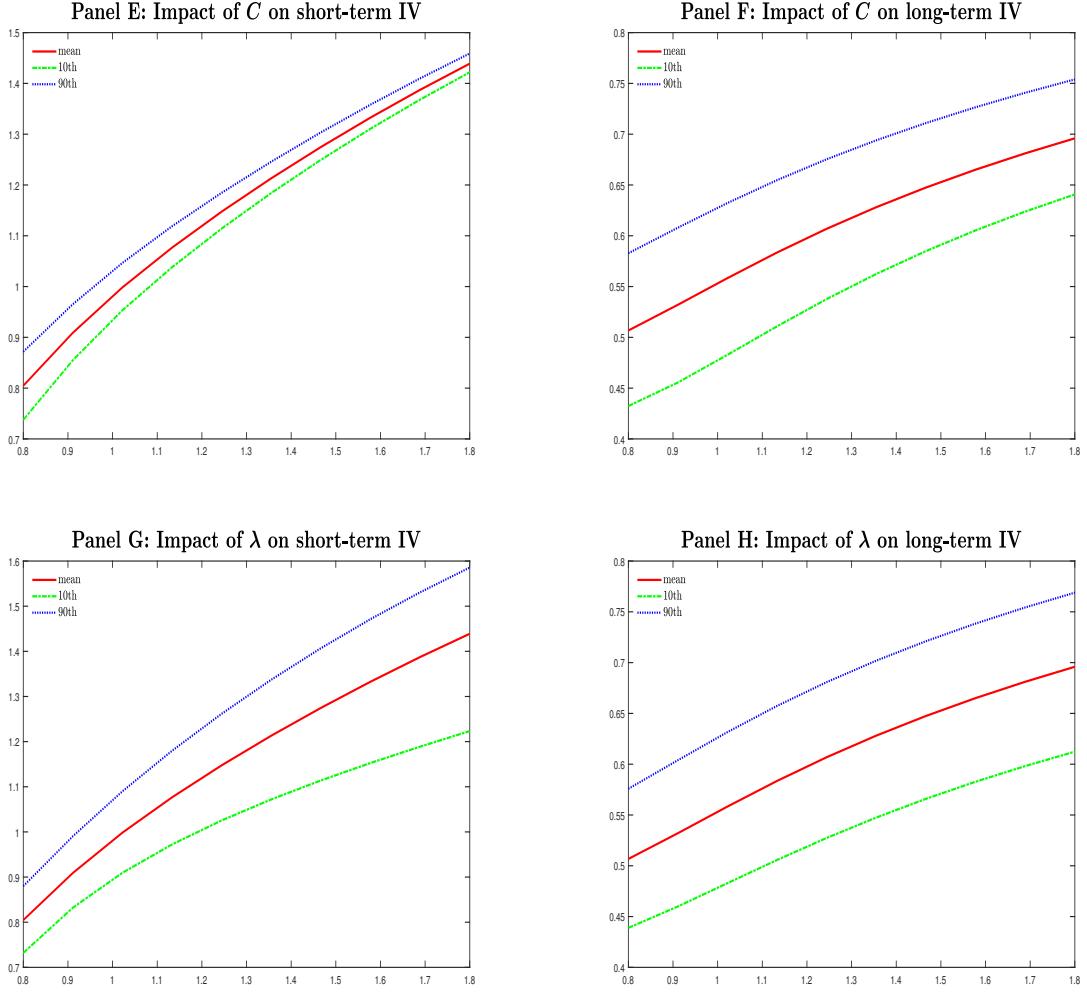


Fig. 6. Continued. This figure illustrates the impact of risk factors on implied volatility (IV) surface along the moneyness dimension. The simulated IV surface is calculated based on the VVC-DJ model with parameter estimates in Table 5. The solid lines are mean value of IVs computed by using the sample mean levels of the risk factors. Dotted lines and dashed lines are computed by setting the specific factor at its 90th and 10th percentile values and fixing other risk factors, respectively.

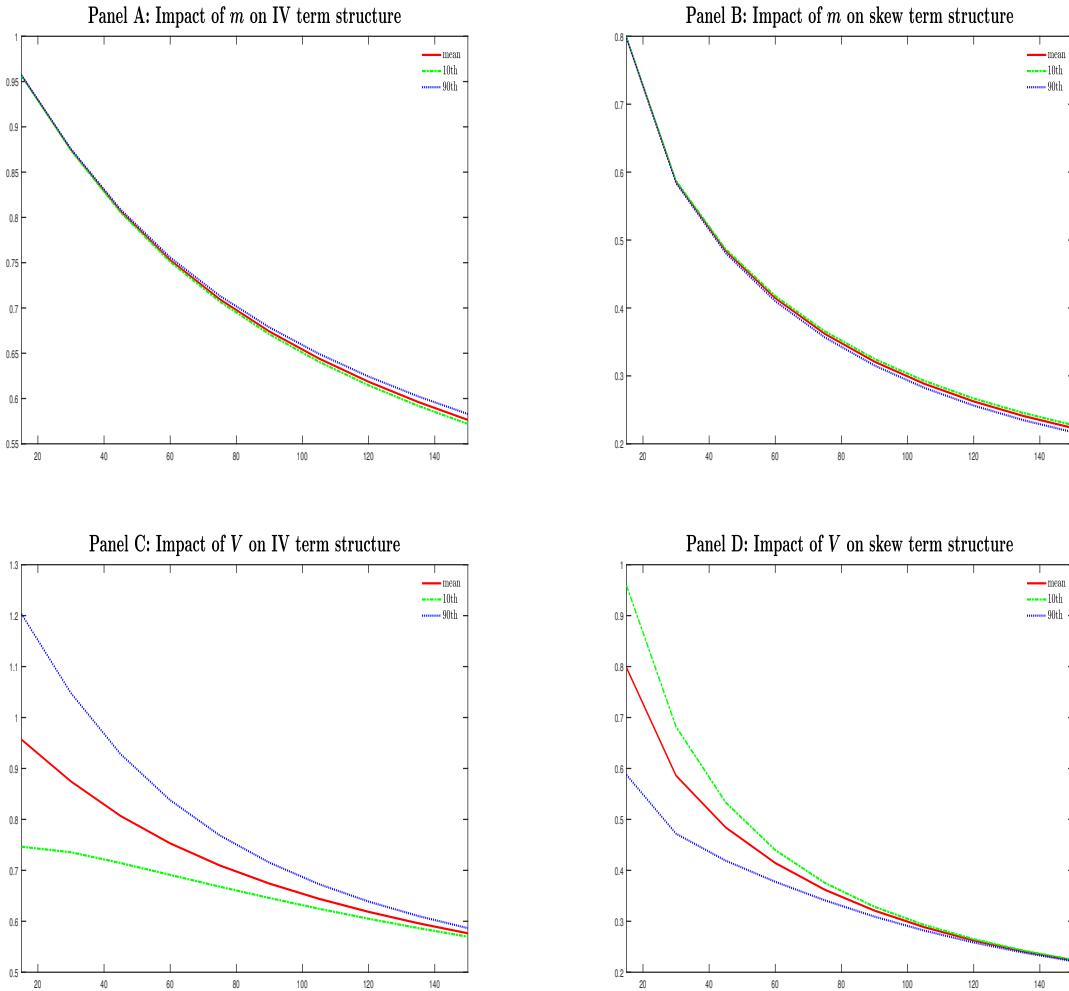


Fig. 7. This figure illustrates the impact of risk factors on implied volatility (IV) surface along the maturity dimension. The simulated IV surface is calculated based on the VVC-DJ model with parameter estimates in Table 5. The solid lines are mean value of IVs computed by using the sample mean levels of four risk factors. Dotted lines and dashed lines are computed by setting the specific factor at its 90th and 10th percentile values and fixing other risk factors, respectively.

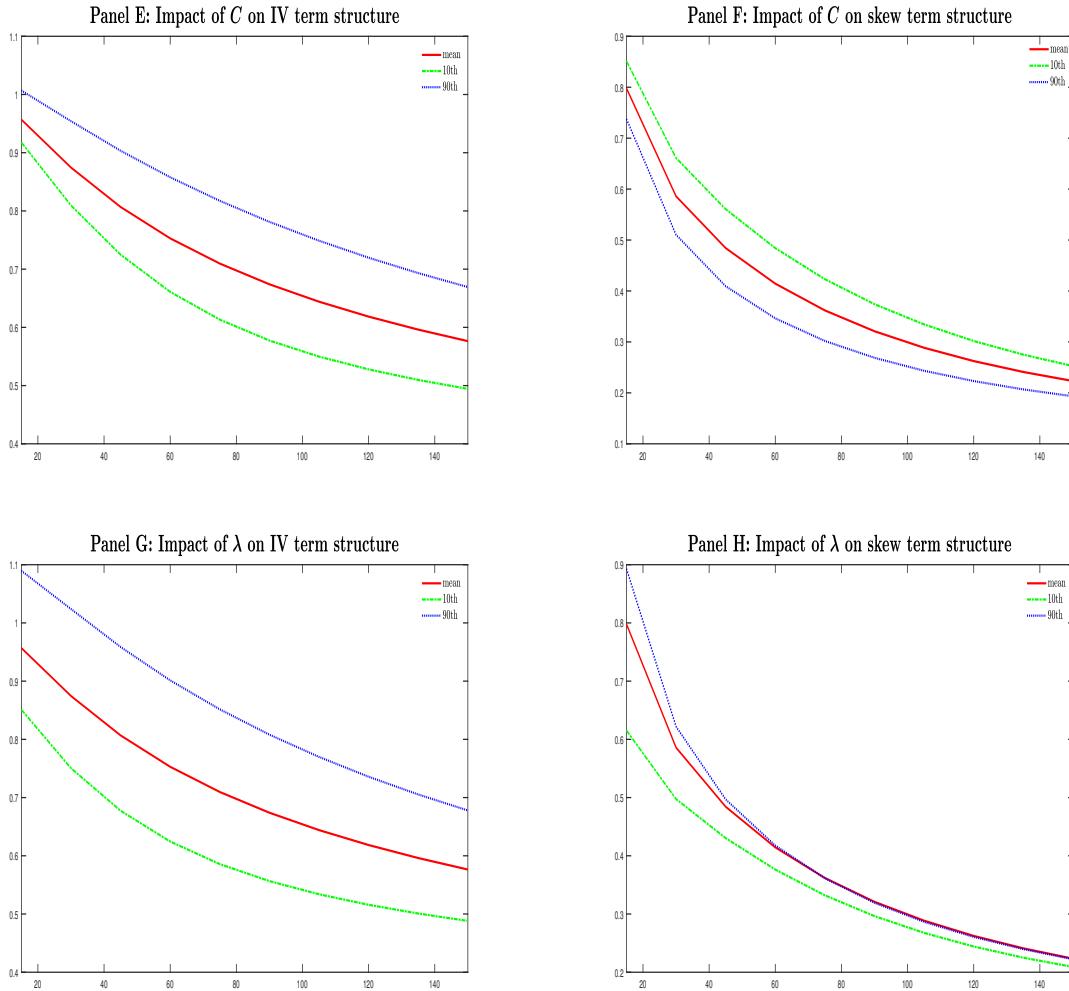


Fig. 7. Continued. This figure illustrates the impact of risk factors on implied volatility (IV) surface along the maturity dimension. The simulated IV surface is calculated based on the VVC-DJ model with parameter estimates in Table 5. The solid lines are mean value of IVs computed by using the sample mean levels of the risk factors. Dotted lines and dashed lines are computed by setting the specific factor at its 90th and 10th percentile values and fixing other risk factors, respectively.

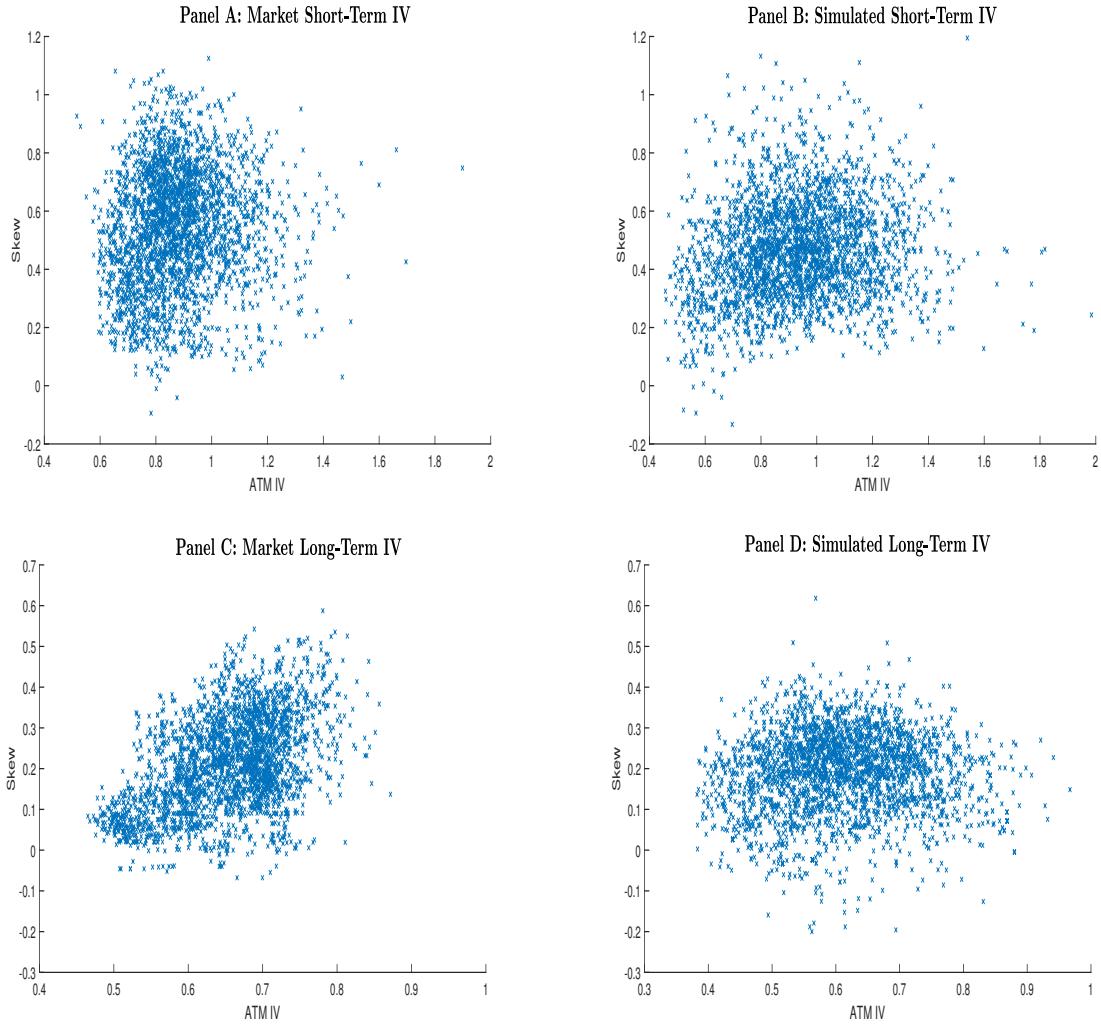


Fig. 8. This figure compares the simulated scatter plots of ATM IV and skew with empirically observed scatter plots. I first obtain the distribution of state variables under the physical measure by simulating the VVC-DJ model in Equations (10) - (14). Next, I calculate the model implied ATM IV and skew by feeding states that are randomly generated from the distributions into the VVC-DJ model with the parameters in Table 5. Panels A and B display the scatter plots for the short-term options, and Panels C and D show the plots of the long-term contracts.

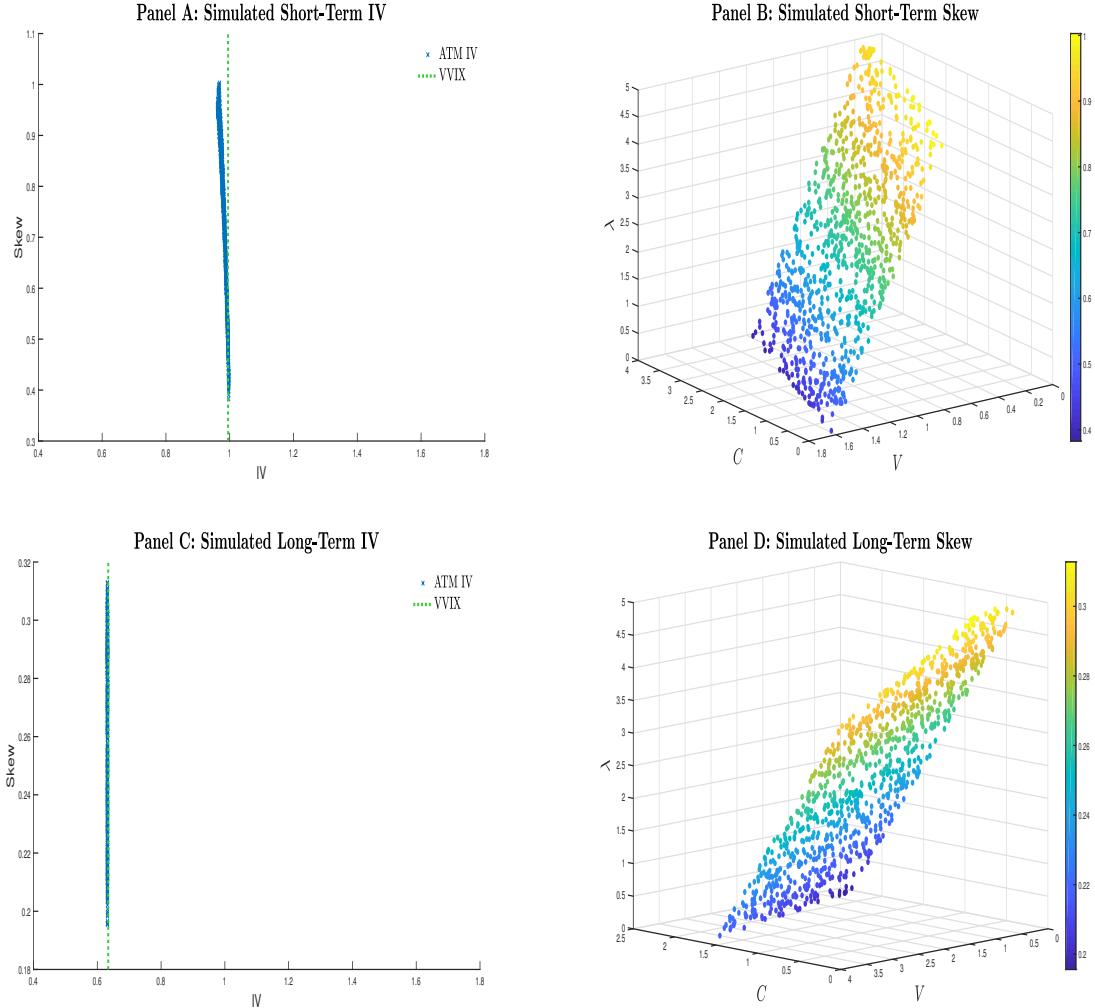


Fig. 9. This figure shows the simulated skew and ATM IV. I generate 1000 random points X in 3-dimensional space (V , C and λ) subject to several linear constraints in Equation (28). Specifically, I generate different values of risk factors V , C and λ such that the simulated VVIX stays at the mean value $\overline{\text{VVIX}}$. $\overline{\text{VVIX}}^2(\tau)$ is calculated by setting risk factors at their respective mean values in Equation (B.7) Appendix B, using the parameters in Table 5. Panels A and B plot the simulation for short-term IV (15 days), and Panels C and D display the simulation for long-term IV (150 days). Furthermore, Panels A and C show scatter plots of ATM IV (x-axis) and skew (y-axis). The dotted lines denote VVIX. Panels B and D present the simulated skew in the 3-D plot. The color bar denotes the level of simulated skew.

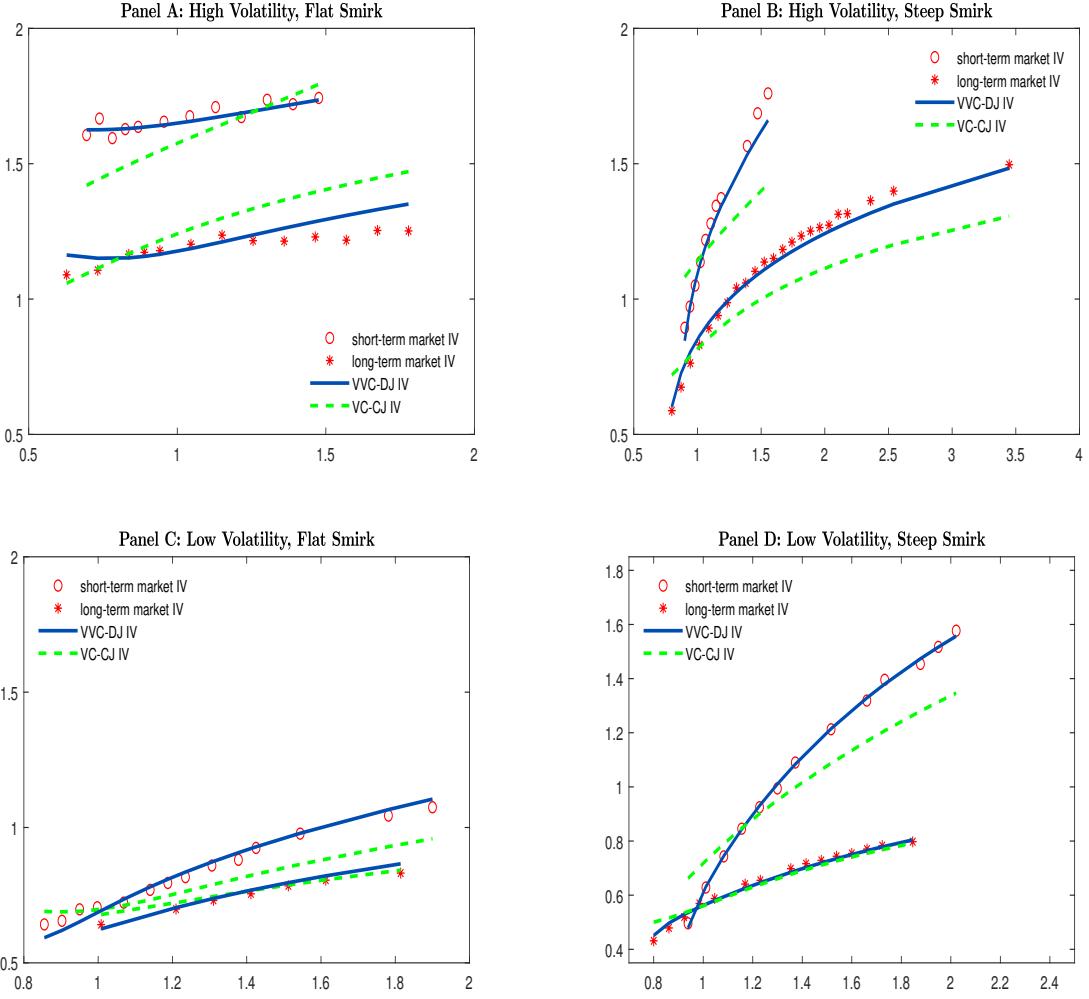


Fig. 10. This figure compares market and model implied volatility (IV) for short and long maturities on four different days: November 6, 2008, a high volatility day with a flat smirk, in Panel A; August 9, 2017, a high volatility day with a steep smirk, in Panel B; November 11, 2010, a low volatility day with a flat smirk, in Panel C; January 22, 2014, a low volatility day with a steep smirk, in Panel D. Circles and crosses represent the market-implied IV for short and long maturities, respectively. The solid and dashed lines correspond to the model implied IV for VVC-DJ and VC-CJ models, respectively.