# Essays on Skewness

by

Fang Zhen

A thesis submitted in fulfillment of the requirements
for the degree of Doctor of Philosophy in Finance

at the

Department of Accountancy and Finance
Otago Business School, University of Otago
Dunedin, New Zealand

March 2017

# Essays on Skewness


## *Abstract*

This thesis is devoted to the study of the higher-moment risk, in particular, the skewness risk.

In Chapter 2, we provide an exact formula for the skewness of stock returns implied in the Heston (1993) model by using a moment-computing approach. We compute the moments of Itô integrals by using Itô's Lemma skillfully. The model's affine property allows us to obtain analytical formulas for cumulants. The formulas for the variance and the third cumulant are written as time-weighted sums of expected instantaneous variance, which are neater and more intuitive than those obtained with the characteristic function approach. Our skewness formula is then applied in calibrating Heston's model by using the market data of the CBOE VIX and SKEW.

The CBOE SKEW is an index launched by the Chicago Board Options Exchange (CBOE) in February 2011. Its term structure tracks the risk-neutral skewness of the S&P 500 index (SPX) for different maturities. In Chapter 3, we develop a theory for the CBOE SKEW by modelling SPX using a jump-diffusion process with stochastic volatility and stochastic jump intensity. With the term structure data of VIX and SKEW, we estimate model parameters and obtain the four processes of variance, jump intensity and their long-term mean levels. Our results can be used to describe the risk-neutral distribution of the SPX returns and to price SPX options.

In Chapter 4, We measure the jump magnitude of the SPX index by using sum of cubed returns (i.e., realized cubic variation). We further detect the existence of jumps if the jump magnitude is higher than a given threshold. Both option-implied and time-series information are used to forecast future one-month jump magnitude and jump existence likelihood. Our results show that option-implied information, coupled with past diffusive variance, is more efficient in forecasting jump magnitude than is time-series information, whereas past realized variance outperforms option-implied information in forecasting jump existence likelihood. We also find that the realized cubic variation and its risk-neutral expectation are significantly negatively correlated.

# *Acknowledgements*

First and foremost, I would like to give thanks to my primary supervisor, Professor Jin Zhang, for his patience and insightful guidance. Professor Zhang has inspired me so much on how to become a competent scholar. He is not only my thesis advisor but also my life mentor. I have benefited a lot from Jin's carefully designed training on research and teaching. I would not have gotten this far without his great effort in helping me.

Second, I am grateful to my co-supervisor, Professor Timothy Crack, for his helpful comments and instructions, especially on the paper titled "Jump risk: a cubic-variation approach". His conscientious attitude always motivates me to think logically and critically. I am also thankful to my co-supervisor, Dr. Eric Tan, for his continuous encouragement.

Third, I would like to thank all the seminar participants at the University of Otago, including Associate Professor I.M. Premachandra, Dr. Helen Roberts, Dr. Scott Chaput, Associate Professor Ivan Diaz-Rainey and Dr. Xing Han; our group meeting members, Nhu Nguyen, Xinfeng Ruan, Sebastian Gehricke and Tian Yue; and our proof editor, Marianne Lown.

I also want to thank Dr. Jing (Annie) Zhang, for organising wonderful PhD gatherings. An opportunity to network with experienced lecturers and other PhD students always reminds me to refresh myself and strive for my academic career.

I am indebted to my parents for their unconditional love and support. Last but not least, I would like to thank my friends in Dunedin for enriching my life with joyful and unforgettable experiences.

# Contents

# List of Tables

# List of Figures

# Chapter 1

# Introduction

After obtaining a Master's degree in Probability and Mathematical Statistics from the Chinese Academy of Sciences, I started my research journey as a PhD student at the University of Otago in November 2013 with an expectation to contribute in the area of quantitative finance.

After some discussions with my supervisor, Professor Jin Zhang, I started a systematic study of the literature on skewness published in top finance journals in the last 40 years. I realized that skewness is an indispensable higher-moment risk factor in asset pricing studies. Its empirical implication has been examined extensively in the literature. Hence, I decided to work on the higher-moment risk, in particular skewness risk, as the topic of my PhD thesis.

## 1.1 Literature on Skewness

Skewness risks play an important role in the asset-pricing literature. Kraus and Litzenberger (1976) propose a three-moment capital asset pricing model (CAPM). They find that systematic skewness is relevant to market valuation and investors have a preference for positive skewness. Harvey and Siddique (2000) find that conditional co-skewness helps explain the cross-sectional variation of expected returns across assets and systematic skewness commands a risk premium of 3.60 percent per year on average. Boyer, Mitton and Vorkink (2010) find that expected idiosyncratic skewness and returns are negatively correlated. Chang, Christoffersen and Jacobs (2013) find that stocks with high exposure to innovations in implied market skewness exhibit low returns on average. Conrad, Dittmar and Ghysels (2013) find

that more ex-ante negatively (positively) skewed returns yield subsequent higher (lower) returns. Amaya, Christoffersen, Jacobs and Vasquez (2015) find that buying stocks in the lowest realized skewness decile and selling stocks in the highest realized skewness decile generates an average return of 19 basis points the following week with a t-statistic of 3.70. Additionally, Bali and Murray (2013) find a strong negative relation between the risk-neutral skewness and the returns of their created delta-neutral and vega-neutral skewness assets.

The equity or index return skewness has also been analysed in the literature. Harvey and Siddique (1999) show that the evidence of asymmetric variance is consistent with conditional skewness. Chen, Hong and Stein (2001) find that negative skewness is most pronounced in stocks that have experienced (1) an increase in trading volume relative to trend over the prior six months and (2) positive returns over the prior 36 months. Dennis and Mayhew (2002) find that beta, market volatility and other firm-specific factors, including firm size and trading volume, help explain the cross-sectional variation in the risk-neutral skewness implied by individual stock option prices. Friesen, Zhang and Zorn (2012) find that belief differences may be one of the unexplained firm-specific components affecting skewness. Bakshi, Kapadia and Madan (2003) show that individual risk-neutral distributions differ from that of the market index by being far less negatively skewed. Han's (2008) findings reveal that the risk-neutral skewness of monthly index returns is more (less) negative when market sentiment becomes more bearish (bullish). Neuberger (2012) proposes a definition of realized third moment and demonstrates that the skewness of equity index returns, far from diminishing with horizons, actually increases with horizons up to a year. Kozhan, Neuberger and Schneider (2013) develop a new method for measuring moment risk premiums and find that skew risk is tightly related to variance risk.

Moreover, skewness in equilibrium models has been studied. Hong and Stein (2003) develop a model with differences of opinion among investors and short-sales constraints. Their model makes a distinctive new prediction that returns will be more negatively skewed conditional on high trading volume. Xu (2007) develops a model which predicts that short-sales constraints and disagreement about information precision increase skewness. Mitton and Vorkink (2007) develop a one-period model where investors have heterogeneous preference for skewness, which allows

the investors to under-diversify in equilibrium. Barberis and Huang (2008) study the asset pricing implication of cumulative prospect theory with a particular focus on its probability weighting component and show that a security's own skewness can be priced: a positively skewed security can be "overpriced" and can earn a negative average excess return. Albuquerque (2012) provides a unified theory that reconciles the negative skewness of aggregate stock market returns and the positive skewness of firm stock returns by explicitly modelling firm-level heterogeneity.

## 1.2 The CBOE SKEW

In February 2011, the CBOE launched the SKEW index to track the skewness of the distribution of the SPX returns by using Bakshi, Kapadia and Madan's (2003) model-free methodology. This publicly available information provides a standard measure of the skewness risk of aggregate stock market returns. It is a combination of weighted sums of out-of-the-money option prices. Hence, it alleviates the impact of the idiosyncratic errors in individual options (see Li and Zhang, 2013). The CBOE SKEW is built to complement the volatility index VIX. The latter has already become a benchmark for practitioners and academics in volatility gauging and trading. Analogously, SKEW is likely to be a skewness benchmark. Moreover, given the popularity of the VIX futures and options (see Mencía and Sentana, 2013), there is a potential market for skewness trading and SKEW derivatives may become as successful as VIX derivatives. However, the literature on SKEW is sparse. Therefore, this thesis focuses on the study of the CBOE SKEW.

## 1.3 Structure of this PhD thesis

This thesis comprises three independent but related papers, as shown in Table 1.1.

First, we calculate the skewness implied in the Heston (1993) model in Chapter 2, and further extend the Heston model to more realistic affine jump-diffusion models in Chapter 3. The model parameters and latent variables are estimated by using the term-structure data of the CBOE SKEW and VIX plus the SPX and 30-day VIX data. Finally, in Chapter 4, we explore the relation between the realized cubic variation

and its risk-neutral expectation using the 30-day SKEW and VIX as well as the SPX data.

Inspired by its importance and potential future applications, I aimed to develop a model for the CBOE SKEW under affine jump-diffusion models for the SPX index. After an exhaustive search of the related literature, I found that an explicit formula for the skewness in the pure-diffusion Heston (1993) model was not available. Only a working paper written by my supervisor Professor Jin Zhang and his co-authors Xiaoxia Sun and Huimin Zhao is on this topic. Nevertheless, their formula is lengthy and incomplete. After some intensive discussions with them, we derived jointly a new exact formula for the skewness implied in the Heston model. In December 2014, we finished a working paper mainly on theoretical results and started submitting it to a journal for publication. In August 2015, we received a positive report from the *Journal of Futures Markets* with a request for more research on applications. I completed the application part under the guidance of Professor Zhang. The paper was accepted for publication in June 2016. It is now presented here as Chapter 2.

Noticing the inability of the Heston model to capture short-term skewness and stochastic long-term variance, in Chapter 3, we further derive the skewness formulas in more realistic and sophisticated option-pricing models. The CBOE launched two important characteristics (VIX in 1993 and SKEW in 2011, respectively) of the SPX returns' distribution in the risk-neutral measure. This public information provides two standard layers of risks and could be used for different purposes. We use VIX and SKEW for model estimation and analyse skewness in various jump-diffusion models.

After modelling the CBOE SKEW in the risk-neutral measure, we explore how to use it in the physical world in Chapter 4. The SKEW is extracted from option prices, and option-implied information reflects investors' expectation about the underlying stock market. Hence, given the forward-looking nature of the CBOE SKEW, we use it to forecast the SPX future realized non-standardized skewness, which is defined as the sum of cubed daily returns.

TABLE 1.1: Thesis Chapters

| Chp | Paper title | Authors | Contribution of candidate | Journal | Status |
|---|---|---|---|---|---|
| 2 | The skewness implied in the Heston model and its application | Jin E. Zhang, Fang Zhen, Xiaoxia Sun & Huimin Zhao | Undertook formula derivation, simulation and model estimation. Coauthors provided guidance and support on research, in writing, submitting and revising the paper. | *Journal of Futures Markets* | Volume 37, Issue 3, March 2017, Pages 211-237 |
| 3 | The CBOE SKEW | Fang Zhen & Jin E. Zhang | Undertook skewness formulas derivation and model estimation. Coauthor provided guidance on literature, theoretical techniques, research and writing skills. | N/A | N/A |
| 4 | Jump risk: a cubic-variation approach | Fang Zhen & Jin E. Zhang | Undertook theoretical and empirical analyses. Coauthor provided support on literature and research. | N/A | N/A |

# Chapter 2

# The Skewness Implied in the Heston Model and Its Application

This chapter is a joint work with Jin E. Zhang, Xiaoxia Sun and Huimin Zhao. It has been published in the *Journal of Futures Markets* 37(3), 2017, 211-237.

## 2.1 Introduction

Option-pricing literature is well-developed, with all kinds of affine jump-diffusion models. The estimation or calibration of (often many) model parameters is still a challenging problem. Recently, CBOE launched new volatility and skewness indexes, VIX and SKEW, and their term structures, which are model-free measures of the second and third moments computed from SPX options.[1] By making good use of the market data of the CBOE VIX and SKEW term structures, we are able to calibrate parameters in a continuous-time model efficiently and effectively. However, in order to achieve this goal, we need the VIX and SKEW formulas implied in the model.

Heston's (1993) model is often used by finance researchers to describe the dynamics of stock prices with stochastic volatility.[2] Hence, it is important to obtain

[1] It has been well-documented that market skewness risk is one of the important factors that drive future stock returns, see e.g., Chang, Christoffersen and Jacobs (2013) and Conrad, Dittmar and Ghysels (2013) among others.

[2] Heston (1993) has been cited by 6,971 papers in Google Scholar as of 3 March 2017. This provides evidence of the popularity of the model.

an analytical formula for the skewness of stock returns implied in the model.[3] Unfortunately, a closed-form formula for the skewness has never been presented. This chapter fills the gap by providing an exact formula for the skewness of stock returns implied in Heston's model by using a moment-computing approach.[4]

In Heston's model (1993), the price of a stock, $S_t$, is modelled as a stochastic process as follows:

$$\frac{dS_t}{S_t} = \mu dt + \sqrt{v_t} dB_t^S, \tag{2.1}$$

where $\mu$ is the expected return, $B_t^S$ is a standard Brownian motion and $v_t$ is the stochastic *instantaneous variance* that follows a mean-reverting squared root process:

$$dv_t = \kappa(\theta - v_t)dt + \sigma_v \sqrt{v_t} dB_t^v, \tag{2.2}$$

where $\kappa$ is a measure of the mean-reverting speed, $\theta$ is the long-term mean level of the variance, $\sigma_v$ is the volatility of variance and $B_t^v$ is another standard Brownian motion that is correlated with $B_t^S$ with a constant coefficient, $\rho$.[5]

Applying Itô's Lemma to Equation (2.1) gives

$$d \ln S_t = \left(\mu - \frac{1}{2} v_t\right) dt + \sqrt{v_t} dB_t^S. \tag{2.3}$$

Integrating from current time, $t$, to a future time, $T$, yields an expression for the *continuously compounded return* as follows:

$$R_t^T \equiv \ln \frac{S_T}{S_t} = \int_t^T \left[\left(\mu - \frac{1}{2} v_u\right) du + \sqrt{v_u} dB_u^S\right]. \tag{2.4}$$

[3]The *return* here stands for *term-return*, which is defined as a return over a finite period (term) from $t$ to $T$, see Equation (2.4) for definition.

[4]Both terms *moment* and *cumulant* are used in this chapter. Their meanings are different; see e.g., Zhao, Zhang and Chang (2013) for their definitions.

[5]The risk-neutral skewness is different from observable skewness. Because under the risk-neutral measure and the physical measure, the variance process $v_t$ in Equation (2.2) has a similar structure but a different mean-reverting speed $\kappa$ and long-term mean level $\theta$.

The conditional expectation of the continuously compounded return is then given by

$$E_t(R_t^T) = \int_t^T \left[\mu - \frac{1}{2}E_t(v_u)\right] du. \tag{2.5}$$

The higher-order cumulants of stock returns implied in the Heston model are of interest in asset and option pricing. Das and Sundaram (1999) (referred to as DS hereafter) obtain closed-form solutions of the skewness and kurtosis for a similar stock price model with stochastic volatility, but their results do not apply in the Heston model, as we will see in Section 2.3. Zhao, Zhang and Chang (2013) provide analytical formulas of the variance and partial results of the third and fourth cumulants for the special case of $\kappa = 0$. In principle, one is able to derive moment formulas by using the characteristic function available in Heston (1993), but the resulting formulas are lengthy and lack intuition. A complete analysis of the third cumulant or skewness is not available in the literature.[6]

The purpose of this chapter is to provide an analytical formula for the third cumulant, i.e.,

$$E_t\left[R_t^T - E_t(R_t^T)\right]^3,$$

or the skewness[7]

$$\frac{E_t\left[R_t^T - E_t(R_t^T)\right]^3}{\left\{E_t\left[R_t^T - E_t(R_t^T)\right]^2\right\}^{3/2}}.$$

In this chapter, we compute the moments of Itô integrals by using Itô's Lemma skillfully. The model's affine property allows us to obtain analytical formulas for cumulants. The formulas for the variance and the third cumulant are written as time-weighted sums of expected instantaneous variance, which are neater and more

[6]Dufresne (2001) studies the moments of (integrated) instantaneous variance process, but he does not study the co-moments between the stock return and integrated variance processes. Drăgulescu and Yakovenko (2002) study the probability density function of stock returns implied in the Heston model, but they do not study the higher moments, including skewness of returns.

[7]In this chapter, the term *skewness* means *conditional skewness* unless otherwise specified.

intuitive than those obtained with the characteristic function approach. Our skewness formula is then applied in calibrating Heston's model by using the market data of the CBOE VIX and SKEW.

This chapter makes four contributions to the literature. First, we point out that the DS formula in their Proposition 2 does not apply in the Heston model. Second, we provide exact formulas for the third cumulant and skewness implied in the Heston model in a neat and intuitive form. Third, we provide a new moment-computing method that can be applied in studying the skewness and higher moments in more general affine jump-diffusion models. Fourth, we provide an efficient and effective way to calibrate the Heston model.

This chapter proceeds as follows. In Section 2.2, we review existing results on the variance of the continuously compounded return, illustrate our methodology of computing the moments of Itô integrals and examine the limiting case of mean-reverting speed $\kappa = 0$. In Section 2.3, we present our main results on the third cumulant and skewness. The details of mathematical proof are included in Appendix 2.7.1. We also examine the relationship between the DS formula and ours, and study their asymptotic limits for small and large $\tau \equiv T - t$. Section 2.4 examines the difference between the DS formula and ours numerically. Section 2.5 applies our theoretical formula in calibrating the Heston model. Section 2.6 concludes.

## 2.2 The Variance of the Continuously Compounded Return

In this section, we review existing results on the variance of the continuously compounded return implied in the Heston model.

To facilitate our presentation, we introduce following notations

$$X_T \equiv \int_t^T \sqrt{v_u} dB_u^S, \qquad Y_T \equiv \int_t^T [v_u - E_t(v_u)] du, \tag{2.6}$$

where $X_T$ measures the cumulative uncertainty of asset returns, and $Y_T$ measures the uncertainty of integrated variance process over the period from $t$ to $T$. We notice that $X_u$ is an Itô integral. Hence, it is a martingale, while $Y_u$ is not.

With this notation, subtracting (2.5) from (2.4) gives

$$R_t^T - E_t(R_t^T) = X_T - \frac{1}{2}Y_T. \tag{2.7}$$

We now present a list of useful results on the instantaneous variance

$$v_s = \theta + (v_t - \theta)e^{-\kappa(s-t)} + \sigma_v \int_t^s e^{-\kappa(s-u)}\sqrt{v_u}dB_u^v, \tag{2.8}$$

$$E_t(v_s) = \theta + (v_t - \theta)e^{-\kappa(s-t)}, \tag{2.9}$$

$$v_s - E_t(v_s) = \sigma_v \int_t^s e^{-\kappa(s-u)}\sqrt{v_u}dB_u^v, \tag{2.10}$$

which can be derived easily from the variance process (2.2).

The de-meaned integrated variance is then written as

$$\begin{aligned} Y_T &= \int_t^T [v_s - E_t(v_s)]ds = \sigma_v \int_t^T \int_t^s e^{-\kappa(s-u)}\sqrt{v_u}dB_u^v ds \\ &= \sigma_v \int_t^T \int_u^T e^{-\kappa(s-u)}ds\ \sqrt{v_u}dB_u^v = \sigma_v \int_t^T \frac{1-e^{-\kappa(T-u)}}{\kappa}\sqrt{v_u}dB_u^v, \end{aligned} \tag{2.11}$$

where Equation (2.10) has been used in deriving the second equality and a technique of interchanging the order of integration has been used in deriving the third equality.

**Proposition 2.1** *Zhao, Zhang and Chang (2013): The variance of the continuously compounded return, $R_t^T$, is given as follows:*

$$E_t\left[R_t^T - E_t(R_t^T)\right]^2 = E_t\left(X_T - \frac{1}{2}Y_T\right)^2 = E_t(X_T^2) - E_t(X_T Y_T) + \frac{1}{4}E_t(Y_T^2), \tag{2.12}$$

*where $X_T$ and $Y_T$ are defined by*

$$X_T \equiv \int_t^T \sqrt{v_u}dB_u^S, \qquad Y_T \equiv \int_t^T [v_u - E_t(v_u)]du = \sigma_v \int_t^T \frac{1-e^{-\kappa(T-u)}}{\kappa}\sqrt{v_u}dB_u^v,$$

*and the variance and covariance of $X_T$ and $Y_T$ are given by*

$$E_t(X_T^2) = \int_t^T E_t(v_u)du, \tag{2.13}$$

$$E_t(X_T Y_T) = \rho\sigma_v \int_t^T \frac{1-e^{-\kappa(T-u)}}{\kappa} E_t(v_u)du, \tag{2.14}$$

$$E_t(Y_T^2) = \sigma_v^2 \int_t^T \frac{(1-e^{-\kappa(T-u)})^2}{\kappa^2} E_t(v_u)du, \tag{2.15}$$

*and $E_t(v_u) = \theta + (v_t - \theta)e^{-\kappa(u-t)}$ is the expected instantaneous variance.*

**Remark 2.1.1** The results of the variance and covariance of $X_T$ and $Y_T$ in Equations (2.13), (2.14) and (2.15) are presented in terms of the weighted sum of the expected instantaneous variance over the period from $t$ to $T$, where the weights are increasing functions of time to maturity, $T-u$. It is quite straightforward carrying out these integrations to obtain the final formulas in Zhao, Zhang and Chang (2013). However, we choose to present them (and the third cumulant in the next section) this way so that the results are neater and more intuitive. We will come back to this point in Remark 2.3.2, after we present an explicit formula for the skewness. Figure 2.1 demonstrates how to obtain final formulas by computing these integrations straightforwardly with *Mathematica*.

**Remark 2.1.2** The result of Equation (2.13) can be obtained by using Itô's Isometry, or alternatively by using Itô's Lemma as follows:

$$\begin{aligned} E_t(X_T^2) &= E_t \int_t^T d(X_u^2) = E_t \int_t^T 2X_u dX_u + (dX_u)^2 \\ &= E_t \int_t^T (dX_u)^2 = E_t \int_t^T v_u du = \int_t^T E_t(v_u)du. \end{aligned}$$

We define a new process, $Y_s^*$, as follows:

$$Y_s^* \equiv \sigma_v \int_t^s \frac{1-e^{-\kappa(T-u)}}{\kappa}\sqrt{v_u}dB_u^v, \qquad dY_s^* = \sigma_v \frac{1-e^{-\kappa(T-s)}}{\kappa}\sqrt{v_s}dB_s^v.$$

Recall the definition of $Y_s$

$$Y_s \equiv \sigma_v \int_t^s \frac{1-e^{-\kappa(s-u)}}{\kappa}\sqrt{v_u}dB_u^v, \qquad dY_s = [v_s - E_t(v_s)]ds.$$

Notice that $Y_T^* = Y_T$. The weight function in $Y_s^*$ is $\frac{1}{\kappa}(1-e^{-\kappa(T-u)})$, which is independent of $s$; while the weight function in $Y_s$ is $\frac{1}{\kappa}(1-e^{-\kappa(s-u)})$, which depends on $s$. This weight difference determines that $Y_s^*$ is an Itô process (martingale) and $Y_s$ is not. The martingale property of $Y_s^*$ allows us to perform the following calculations:

$$\begin{aligned}
E_t(X_TY_T) = E_t(X_TY_T^*) &= E_t\int_t^T d(X_uY_u^*) = E_t\int_t^T Y_u^*dX_u + X_udY_u^* + dX_udY_u^* \\
&= E_t\int_t^T dX_udY_u^* = \rho\sigma_v\int_t^T \frac{1-e^{-\kappa(T-u)}}{\kappa}E_t(v_u)du, \\
E_t(Y_T^2) = E_t(Y_T^{*2}) &= E_t\int_t^T d(Y_u^{*2}) = E_t\int_t^T 2Y_u^*dY_u^* + (dY_u^*)^2 \\
&= E_t\int_t^T (dY_u^*)^2 = \sigma_v^2\int_t^T \frac{(1-e^{-\kappa(T-u)})^2}{\kappa^2}E_t(v_u)du.
\end{aligned}$$

The new process, $Y_s^*$, can be regarded as a *shadow* of $Y_s$. The design of the shadow process is an innovation in computing the moments of $Y_T$ and its co-moments with $X_T$. It is regarded as one of technical contributions of this chapter. The martingale property of the newly introduced shadow process, $Y_s^*$, dramatically simplifies the process of deriving the expectations. This technique will be frequently used in deriving the third cumulant in the next section.

**Remark 2.1.3** For the limiting case of $\kappa = 0$, the process of $v_u$ is a martingale, i.e., $E_t(v_u) = v_t$.[8] The results in Equations (2.13), (2.14) and (2.15) are reduced to

$$\begin{aligned}
E_t(X_T^2) &= \int_t^T v_t du = v_t \tau, \qquad \tau = T - t, \\
E_t(X_T Y_T) &= \rho\sigma_v \int_t^T (T-u) v_t du = \frac{1}{2}\rho\sigma_v v_t \tau^2, \\
E_t(Y_T^2) &= \sigma_v^2 \int_t^T (T-u)^2 v_t du = \frac{1}{3}\sigma_v^2 v_t \tau^3.
\end{aligned}$$

and

$$\begin{aligned}
E_t\left[R_t^T - E_t(R_t^T)\right]^2 &= v_t\tau - \frac{1}{2}\rho\sigma_v v_t \tau^2 + \frac{1}{12}\sigma_v^2 v_t \tau^3 \\
&= v_t \tau \left(1 - \frac{1}{2}\rho\sigma_v \tau + \frac{1}{12}\sigma_v^2 \tau^2\right),
\end{aligned}$$

which is always positive due to the fact that $\Delta = \frac{1}{4}\rho^2\sigma_v^2 - \frac{1}{3}\sigma_v^2 < 0$. For small $\tau$, the variance of $X_T$ is much larger than the covariance between $X_T$ and $Y_T$, which is much larger than the variance of $Y_T$. This result is consistent with our intuition that the risk in returns mainly comes from $X_T$, which is a weighted cumulation of $dB_t^S$, rather than $Y_T$, which is a weighted cumulation of $dB_t^v$.

## 2.3 Main Results

In this section, we present our main results on the third cumulant implied in the Heston model.

[8] For the limiting case of $\kappa = 0$, the drift of variance process is zero. The limiting case is important. The reasons are as follows. 1) Some popular models such as SABR (stochastic alpha beta rho) have zero drifts in both return and volatility processes. They are widely used by practitioners in the financial industry. 2) The limiting case allows us to have a quick and intuitive estimation on the order of magnitude of each term for a small time to maturity.

**Proposition 2.2** *The third cumulant of the continuously compounded return, $R_t^T$, is given as follows:*

$$\begin{aligned} E_t\left[R_t^T - E_t(R_t^T)\right]^3 &= E_t\left(X_T - \frac{1}{2}Y_T\right)^3 \\ &= E_t(X_T^3) - \frac{3}{2}E_t(X_T^2 Y_T) + \frac{3}{4}E_t(X_T Y_T^2) - \frac{1}{8}E_t(Y_T^3), \end{aligned} \tag{2.16}$$

*where $X_T$ and $Y_T$ are defined by*

$$X_T \equiv \int_t^T \sqrt{v_u} dB_u^S, \qquad Y_T \equiv \int_t^T [v_u - E_t(v_u)]du = \sigma_v \int_t^T \frac{1-e^{-\kappa(T-u)}}{\kappa}\sqrt{v_u}dB_u^v,$$

*and the third and co-third cumulants of $X_T$ and $Y_T$ are given by*

$$\begin{aligned} E_t(X_T^3) &= 3\rho\sigma_v \int_t^T A_1\, E_t(v_u)du, \\ A_1 &= \frac{1-e^{-\kappa\tau^*}}{\kappa}, \qquad \tau^* = T-u, \end{aligned} \tag{2.17}$$

$$\begin{aligned} E_t(X_T^2 Y_T) &= \sigma_v^2 \int_t^T A_2\, E_t(v_u)du, \\ A_2 &= \left(\frac{1-e^{-\kappa\tau^*}}{\kappa}\right)^2 + 2\rho^2 \frac{1-e^{-\kappa\tau^*} - \kappa\tau^* e^{-\kappa\tau^*}}{\kappa^2}, \end{aligned} \tag{2.18}$$

$$\begin{aligned} E_t(X_T Y_T^2) &= \rho\sigma_v^3 \int_t^T A_3\, E_t(v_u)du, \\ A_3 &= 2\frac{1-e^{-\kappa\tau^*} - \kappa\tau^* e^{-\kappa\tau^*}}{\kappa^2}\frac{1-e^{-\kappa\tau^*}}{\kappa} + \frac{1-e^{-2\kappa\tau^*} - 2\kappa\tau^* e^{-\kappa\tau^*}}{\kappa^3}, \end{aligned} \tag{2.19}$$

$$\begin{aligned} E_t(Y_T^3) &= 3\sigma_v^4 \int_t^T A_4\, E_t(v_u)du, \\ A_4 &= \frac{1-e^{-2\kappa\tau^*} - 2\kappa\tau^* e^{-\kappa\tau^*}}{\kappa^3}\frac{1-e^{-\kappa\tau^*}}{\kappa}, \end{aligned} \tag{2.20}$$

*and $E_t(v_u) = \theta + (v_t - \theta)e^{-\kappa(u-t)}$ is the expected instantaneous variance.*

*Proof.* See Appendix 2.7.1.

**Remark 2.2.1** The results of the third and co-third cumulants of $X_T$ and $Y_T$ in Equations (2.17), (2.18), (2.19) and (2.20) are also presented in terms of the weighted sum

of the expected instantaneous variance over the period from $t$ to $T$. The weights, $A_i$s, $i = 1,\ 2,\ 3$ and $4$, are increasing functions of $\tau^* \equiv T - u$. As a result, the third cumulant of $R_t^T$ is a linear combination of $\theta$ and $v_t$. This interesting result is a consequence of the affine property of the Heston model. Once again, Figure 2.1 provides the *Mathematica* code of computing the integrations and resulting exact formulas.

**Remark 2.2.2** For the limiting case of $\kappa = 0$, $E_t(v_u) = v_t$. The results in Equations (2.17), (2.18), (2.19) and (2.20) are reduced to

$$A_1 = \tau^*, \qquad A_2 = (1+\rho^2)\tau^{*2}, \qquad A_3 = \frac{4}{3}\tau^{*3}, \qquad A_4 = \frac{1}{3}\tau^{*4},$$

$$\begin{aligned}
E_t(X_T^3) &= 3\rho\sigma_v \int_t^T (T-u)v_t du = \frac{3}{2}\rho\sigma_v v_t \tau^2, \\
E_t(X_T^2 Y_T) &= \sigma_v^2 \int_t^T (1+\rho^2)(T-u)^2 v_t du = \frac{1}{3}(1+\rho^2)\sigma_v^2 v_t \tau^3, \\
E_t(X_T Y_T^2) &= \rho\sigma_v^3 \int_t^T \frac{4}{3}(T-u)^3 v_t du = \frac{1}{3}\rho\sigma_v^3 v_t \tau^4, \\
E_t(Y_T^3) &= 3\sigma_v^4 \int_t^T \frac{1}{3}(T-u)^4 v_t du = \frac{1}{5}\sigma_v^4 v_t \tau^5.
\end{aligned}$$

and

$$E_t\left[R_t^T - E_t(R_t^T)\right]^3 = \frac{3}{2}\rho\sigma_v v_t \tau^2 - \frac{1}{2}(1+\rho^2)\sigma_v^2 v_t \tau^3 + \frac{1}{4}\rho\sigma_v^3 v_t \tau^4 - \frac{1}{40}\sigma_v^4 v_t \tau^5.$$

For small $\tau$, the contribution to the third cumulant of $R_t^T$ mainly comes from $E_t(X_T^3)$, followed by $E_t(X_T^2 Y_T)$, $E_t(X_T Y_T^2)$ and $E_t(Y_T^3)$. Once again, this result is consistent with our intuition.

Once we have the formulas for the variance and the third cumulant from Proposition 2.1 and 2.2, it is easy to obtain an exact formula for the skewness.

**Proposition 2.3** *The skewness of the continuously compounded return, $R_t^T$, is given as follows:*

$$\text{Skewness} \equiv \frac{E_t(X_T^3) - \frac{3}{2}E_t(X_T^2 Y_T) + \frac{3}{4}E_t(X_T Y_T^2) - \frac{1}{8}E_t(Y_T^3)}{\left[E_t(X_T^2) - E_t(X_T Y_T) + \frac{1}{4}E_t(Y_T^2)\right]^{3/2}} = -\sigma_v \frac{\sqrt{2}A}{\sqrt{\kappa}B^{3/2}}, \tag{2.21}$$

*where*

$$\begin{aligned}
A =& 6\,\mathrm{e}^{3\kappa\tau} v_t\,{\sigma_v}^3 - 22\,\mathrm{e}^{3\kappa\tau}{\sigma_v}^3\theta + 3\,\mathrm{e}^{2\kappa\tau} v_t\,{\sigma_v}^3 + 15\,\mathrm{e}^{2\kappa\tau}{\sigma_v}^3\theta + 24\,\mathrm{e}^{\kappa\tau}\kappa^2 v_t\,\rho\,{\sigma_v}^2\tau \\
&- 12\,\mathrm{e}^{\kappa\tau}\kappa^2\rho\,{\sigma_v}^2\tau\,\theta - 12\,\mathrm{e}^{\kappa\tau}\kappa\,v_t\,{\sigma_v}^3\tau + 6\,\mathrm{e}^{\kappa\tau}\kappa\,{\sigma_v}^3\tau\,\theta + 36\,\mathrm{e}^{\kappa\tau}\kappa\,v_t\,\rho\,{\sigma_v}^2 \\
&- 24\,\mathrm{e}^{\kappa\tau}\kappa\,\rho\,{\sigma_v}^2\theta - 6\,\mathrm{e}^{\kappa\tau} v_t\,{\sigma_v}^3 + 6\,\mathrm{e}^{\kappa\tau}{\sigma_v}^3\theta - 3\,v_t\,{\sigma_v}^3 + {\sigma_v}^3\theta - 24\,\mathrm{e}^{\kappa\tau}\kappa^2 v_t\,\sigma_v \\
&+ 12\,\mathrm{e}^{\kappa\tau}\kappa^2\sigma_v\,\theta - 48\,\mathrm{e}^{3\kappa\tau}\kappa^3 v_t\,\rho + 96\,\mathrm{e}^{3\kappa\tau}\kappa^3\rho\,\theta + 24\,\mathrm{e}^{3\kappa\tau}\kappa^2 v_t\,\sigma_v - 60\,\mathrm{e}^{3\kappa\tau}\kappa^2\sigma_v\,\theta \\
&+ 48\,\mathrm{e}^{2\kappa\tau}\kappa^3 v_t\,\rho - 96\,\mathrm{e}^{2\kappa\tau}\kappa^3\rho\,\theta + 48\,\mathrm{e}^{2\kappa\tau}\kappa^2\sigma_v\,\theta - 6\,\mathrm{e}^{2\kappa\tau}\kappa^2 v_t\,{\sigma_v}^3\tau^2 + 6\,\mathrm{e}^{2\kappa\tau}\kappa^2{\sigma_v}^3\tau^2\theta \\
&+ 48\,\mathrm{e}^{3\kappa\tau}\kappa^2 v_t\,\rho^2\sigma_v - 144\,\mathrm{e}^{3\kappa\tau}\kappa^2\rho^2\sigma_v\,\theta + 6\,\mathrm{e}^{3\kappa\tau}\kappa\,{\sigma_v}^3\tau\,\theta - 36\,\mathrm{e}^{3\kappa\tau}\kappa\,v_t\,\rho\,{\sigma_v}^2 \\
&+ 120\,\mathrm{e}^{3\kappa\tau}\kappa\,\rho\,{\sigma_v}^2\theta - 48\,\mathrm{e}^{2\kappa\tau}\kappa^2 v_t\,\rho^2\sigma_v + 144\,\mathrm{e}^{2\kappa\tau}\kappa^2\rho^2\sigma_v\,\theta - 6\,\mathrm{e}^{2\kappa\tau}\kappa\,v_t\,{\sigma_v}^3\tau \\
&+ 18\,\mathrm{e}^{2\kappa\tau}\kappa\,{\sigma_v}^3\tau\,\theta - 48\,\mathrm{e}^{3\kappa\tau}\kappa^4\rho\,\tau\,\theta - 96\,\mathrm{e}^{2\kappa\tau}\kappa\,\rho\,{\sigma_v}^2\theta + 24\,\mathrm{e}^{3\kappa\tau}\kappa^3\sigma_v\,\tau\,\theta \\
&+ 48\,\mathrm{e}^{2\kappa\tau}\kappa^4 v_t\,\rho\,\tau - 48\,\mathrm{e}^{2\kappa\tau}\kappa^4\rho\,\tau\,\theta - 48\,\mathrm{e}^{2\kappa\tau}\kappa^3 v_t\,\sigma_v\,\tau + 48\,\mathrm{e}^{2\kappa\tau}\kappa^3\sigma_v\,\tau\,\theta \\
&- 24\,\mathrm{e}^{2\kappa\tau}\kappa^4 v_t\,\rho^2\sigma_v\,\tau^2 + 24\,\mathrm{e}^{2\kappa\tau}\kappa^4\rho^2\sigma_v\,\tau^2\theta + 48\,\mathrm{e}^{3\kappa\tau}\kappa^3\rho^2\sigma_v\,\tau\,\theta + 24\,\mathrm{e}^{2\kappa\tau}\kappa^3 v_t\,\rho\,{\sigma_v}^2\tau^2 \\
&- 24\,\mathrm{e}^{2\kappa\tau}\kappa^3\rho\,{\sigma_v}^2\tau^2\theta - 36\,\mathrm{e}^{3\kappa\tau}\kappa^2\rho\,{\sigma_v}^2\tau\,\theta - 48\,\mathrm{e}^{2\kappa\tau}\kappa^3 v_t\,\rho^2\sigma_v\,\tau + 96\,\mathrm{e}^{2\kappa\tau}\kappa^3\rho^2\sigma_v\,\tau\,\theta \\
&+ 48\,\mathrm{e}^{2\kappa\tau}\kappa^2 v_t\,\rho\,{\sigma_v}^2\tau - 96\,\mathrm{e}^{2\kappa\tau}\kappa^2\rho\,{\sigma_v}^2\tau\,\theta, \\
B =& -8\,\mathrm{e}^{2\kappa\tau}\kappa^2\rho\,\sigma_v\,\tau\,\theta + 8\,\mathrm{e}^{2\kappa\tau}\kappa^3\tau\,\theta + 2\,\mathrm{e}^{2\kappa\tau}\kappa\,{\sigma_v}^2\tau\,\theta + 8\,\mathrm{e}^{\kappa\tau}\kappa^2 v_t\,\rho\,\sigma_v\,\tau - 8\,\mathrm{e}^{\kappa\tau}\kappa^2\rho\,\sigma\,\tau\,\theta \\
&- 8\,\mathrm{e}^{2\kappa\tau}\kappa\,v_t\,\rho\,\sigma_v + 16\,\mathrm{e}^{2\kappa\tau}\kappa\,\rho\,\sigma_v\,\theta - 4\,\mathrm{e}^{\kappa\tau}\kappa\,v_t\,{\sigma_v}^2\tau + 4\,\mathrm{e}^{\kappa\tau}\kappa\,{\sigma_v}^2\tau\,\theta + 8\,\mathrm{e}^{2\kappa\tau}\kappa^2 v_t \\
&- 8\,\mathrm{e}^{2\kappa\tau}\kappa^2\theta + 2\,\mathrm{e}^{2\kappa\tau} v_t\,\sigma^2 - 5\,\mathrm{e}^{2\kappa\tau}{\sigma_v}^2\theta + 8\,\mathrm{e}^{\kappa\tau}\kappa\,v_t\,\rho\,\sigma_v - 16\,\mathrm{e}^{\kappa\tau}\kappa\,\rho\,\sigma\,\theta - 8\,\mathrm{e}^{\kappa\tau}\kappa^2 v_t \\
&+ 8\,\mathrm{e}^{\kappa\tau}\kappa^2\theta + 4\,\mathrm{e}^{\kappa\tau}{\sigma_v}^2\theta - 2\,v_t\,{\sigma_v}^2 + {\sigma_v}^2\theta.
\end{aligned}$$

*Proof.* Substituting the results of Propositions 2.1 and 2.2 into the definition of skewness yields the results.

**Remark 2.3.1** We have done an exercise of deriving the skewness by using the characteristic function available in Heston (1993) and obtained the same formula as that presented in Proposition 2.3. Christoffersen, Heston and Jacobs (2009) use the moment generating function to compute skewness in the Heston (1993) model. In their

online appendix, they note that "Unfortunately, the analysis of conditional skewness and kurtosis is rather complex because no simple expressions are available for the cumulants. We use closed-form expressions for conditional cumulants that are derived using Mathematica. These expressions are rather lengthy and are available from the authors on request."

**Remark 2.3.2** The lengthy formula in Proposition 2.3 contains less information and is less convenient to use than our formulas for the variance and the third cumulant presented in Propositions 2.1 and 2.2. Our design of presenting the main results as the time-weighted sum of expected instantaneous variance can be regarded as another contribution of this chapter.

**Remark 2.3.3** The long-term mean level $\theta$ can be regarded as an *unconditional mean* of the instantaneous variance, $v_t$. By setting $v_t = \theta$ in the formulas for $A$ and $B$ above, we obtain a new set of formulas as follows:

$$\begin{aligned}
A = &- 2\,{\sigma_v}^3\theta - 12\,\mathrm{e}^{\kappa\,\tau}\kappa^2\sigma_v\,\theta - 6\,\mathrm{e}^{\kappa\,\tau}\kappa\,{\sigma_v}^3\tau\,\theta + 12\,\mathrm{e}^{\kappa\,\tau}\kappa\,\rho\,{\sigma_v}^2\theta + 12\,\mathrm{e}^{\kappa\,\tau}\kappa^2\rho\,{\sigma_v}^2\tau\,\theta \\
&+ 6\,\mathrm{e}^{3\,\kappa\,\tau}\kappa\,{\sigma_v}^3\tau\,\theta + 84\,\mathrm{e}^{3\,\kappa\,\tau}\kappa\,\rho\,{\sigma_v}^2\theta + 96\,\mathrm{e}^{2\,\kappa\,\tau}\kappa^2\rho^2\sigma_v\,\theta + 12\,\mathrm{e}^{2\,\kappa\,\tau}\kappa\,{\sigma_v}^3\tau\,\theta \\
&- 48\,\mathrm{e}^{3\,\kappa\,\tau}\kappa^4\rho\,\tau\,\theta - 96\,\mathrm{e}^{2\,\kappa\,\tau}\kappa\,\rho\,{\sigma_v}^2\theta + 24\,\mathrm{e}^{3\,\kappa\,\tau}\kappa^3\sigma_v\,\tau\,\theta - 96\,\mathrm{e}^{3\,\kappa\,\tau}\kappa^2\rho^2\sigma_v\,\theta \\
&+ 48\,\mathrm{e}^{3\,\kappa\,\tau}\kappa^3\rho^2\sigma_v\,\tau\,\theta - 36\,\mathrm{e}^{3\,\kappa\,\tau}\kappa^2\rho\,{\sigma_v}^2\tau\,\theta + 48\,\mathrm{e}^{2\,\kappa\,\tau}\kappa^3\rho^2\sigma_v\,\tau\,\theta - 48\,\mathrm{e}^{2\,\kappa\,\tau}\kappa^2\rho\,{\sigma_v}^2\tau\,\theta \\
&+ 48\,\mathrm{e}^{3\,\kappa\,\tau}\kappa^3\rho\,\theta - 36\,\mathrm{e}^{3\,\kappa\,\tau}\kappa^2\sigma_v\,\theta - 48\,\mathrm{e}^{2\,\kappa\,\tau}\kappa^3\rho\,\theta + 48\,\mathrm{e}^{2\,\kappa\,\tau}\kappa^2\sigma_v\,\theta \\
&- 16\,\mathrm{e}^{3\,\kappa\,\tau}{\sigma_v}^3\theta + 18\,\mathrm{e}^{2\,\kappa\,\tau}{\sigma_v}^3\theta, \\
B = &- 8\,\mathrm{e}^{2\,\kappa\,\tau}\kappa^2\rho\,\sigma_v\,\tau\,\theta + 8\,\mathrm{e}^{2\,\kappa\,\tau}\kappa^3\tau\,\theta + 2\,\mathrm{e}^{2\,\kappa\,\tau}\kappa\,{\sigma_v}^2\tau\,\theta + 8\,\mathrm{e}^{2\,\kappa\,\tau}\kappa\,\rho\,\sigma_v\,\theta - 3\,\mathrm{e}^{2\,\kappa\,\tau}{\sigma_v}^2\theta \\
&- 8\,\mathrm{e}^{\kappa\,\tau}\kappa\,\rho\,\sigma_v\,\theta + 4\,\mathrm{e}^{\kappa\,\tau}{\sigma_v}^2\theta - {\sigma_v}^2\theta,
\end{aligned}$$

which are for the *unconditional skewness* given by Equation 2.21.

With some observation and analysis, we obtain the following result on the DS (1999) formula.

**Proposition 2.4** *The skewness formula presented by DS (1999) is given by*

$$\text{Skewness\_DS} = \frac{E_t(X_T^3)}{[E_t(X_T^2)]^{3/2}}, \tag{2.22}$$

*where*

$$E_t(X_T^2) = \int_t^T E_t(v_u)du. \tag{2.23}$$

$$E_t(X_T^3) = 3\rho\sigma_v \int_t^T \frac{1-e^{-\kappa(T-u)}}{\kappa} E_t(v_u)du, \tag{2.24}$$

*Proof.* Substituting $E_t(v_u) = \theta + (v_t - \theta)e^{-\kappa(u-t)}$ into (2.23) and (2.24) gives

$$\begin{aligned}
E_t(X_T^2) &= \int_t^T \left[\theta + (v_t-\theta)e^{-\kappa(u-t)}\right] du = \theta\tau + (v_t-\theta)\frac{1-e^{-\kappa\tau}}{\kappa} \\
&= \frac{e^{-\kappa\tau}}{\kappa}\left[\theta(1-e^{\kappa\tau}+\kappa\tau e^{\kappa\tau}) + v_t(e^{\kappa\tau}-1)\right], \\
E_t(X_T^3) &= 3\rho\sigma_v \int_t^T \frac{1-e^{-\kappa(T-u)}}{\kappa}\left[\theta + (v_t-\theta)e^{-\kappa(u-t)}\right] du \\
&= 3\rho\sigma_v\left[\theta\frac{\kappa\tau-1+e^{-\kappa\tau}}{\kappa^2} + (v_t-\theta)\frac{-e^{-\kappa\tau}+1-\kappa\tau e^{-\kappa\tau}}{\kappa^2}\right] \\
&= 3\rho\sigma_v\frac{e^{-\kappa\tau}}{\kappa^2}\left[\theta(2-2e^{\kappa\tau}+\kappa\tau+\kappa\tau e^{\kappa\tau}) - v_t(1+\kappa\tau-e^{\kappa\tau})\right].
\end{aligned}$$

Substituting these two equations into (2.22) gives

$$\frac{E_t(X_T^3)}{[E_t(X_T^2)]^{3/2}} = \frac{3\rho\sigma_v e^{\frac{1}{2}\kappa\tau}}{\sqrt{\kappa}} \times \frac{\theta(2-2e^{\kappa\tau}+\kappa\tau+\kappa\tau e^{\kappa\tau}) - v_t(1+\kappa\tau-e^{\kappa\tau})}{\left[\theta(1-e^{\kappa\tau}+\kappa\tau e^{\kappa\tau}) + v_t(e^{\kappa\tau}-1)\right]^{3/2}}, \tag{2.25}$$

which is the formula presented by DS (1999) in Equation (16) in Proposition 2 on page 221 in their paper.

**Remark 2.4.1** DS (1999) study the term structure of skewness by using a stock price model with stochastic volatility as follows:

$$\begin{aligned}
d\ln S_t &= \alpha dt + \sqrt{v_t}dB_t^S, \\
dv_t &= \kappa(\theta - v_t)dt + \sigma_v\sqrt{v_t}dB_t^v,
\end{aligned}$$

where $\alpha$ is a constant. This model is not the *full* Heston model given by Equations (2.1) or (2.3) and (2.2), where $\alpha = \mu - \frac{1}{2}v_t$ is a stochastic process. When computing skewness, DS completely ignore the contribution from the uncertainty of $Y_T$ that

comes from the stochastic nature of $\alpha$. This can also be observed in DS's characteristic function in their Proposition B.1., which is different from that of Heston (1993). Hence, the result of DS's Proposition 2 does not apply in the Heston model studied in this chapter.[9]

**Remark 2.4.2** For the limiting case of $\kappa = 0$, the exact formula of the skewness is given by

$$\text{Skewness} = \frac{\frac{3}{2}\rho\sigma_v v_t \tau^2 - \frac{1}{2}(1+\rho^2)\sigma_v^2 v_t \tau^3 + \frac{1}{4}\rho\sigma_v^3 v_t \tau^4 - \frac{1}{40}\sigma_v^4 v_t \tau^5}{\left(v_t\tau - \frac{1}{2}\rho\sigma_v v_t \tau^2 + \frac{1}{12}\sigma_v^2 v_t \tau^3\right)^{3/2}},$$

which behaves like $\dfrac{3}{2}\rho\dfrac{\sigma_v}{\sqrt{v_t}}\sqrt{\tau}$ for small $\tau$, and $-\dfrac{3\sqrt{3}}{5}\dfrac{\sigma_v}{\sqrt{v_t}}\sqrt{\tau}$ for large $\tau$. However, the skewness given by the DS formula is

$$\text{Skewness\_DS} = \frac{\frac{3}{2}\rho\sigma_v v_t \tau^2}{(v_t\tau)^{3/2}} = \frac{3}{2}\rho\frac{\sigma_v}{\sqrt{v_t}}\sqrt{\tau},$$

which agrees with the exact formula for small $\tau$, but has an incorrect sign for large $\tau$.

**Proposition 2.5** *For the general case $\kappa > 0$, if $\tau$ is small, then the skewness is given by*

$$\text{Skewness} = \frac{3}{2}\rho\frac{\sigma_v}{\sqrt{v_t}}\sqrt{\tau} + O(\tau).$$

*If $\tau$ is large, then the skewness is given by*

$$\text{Skewness} = \frac{a}{\sqrt{b}}\frac{1}{\sqrt{\theta\tau}} + o\left(\frac{1}{\sqrt{\tau}}\right), \tag{2.26}$$

[9] It is not shown in the literature that the result in DS's Proposition 2 does not apply in the Heston model. For example, Singleton (2001) implicitly assumes that the DS model is the Heston model by equating its characteristic function with DS one. Pan (2002) treats DS as a special case in her model setting, which implicitly implies that DS study the Heston model. Han (2008) quotes the DS skewness formula as the skewness implied in the Heston model. Todorov (2011) also regards the DS model as the Heston model. Park (2015) cites the DS formula and regards it as the result of the Heston model.

*where*

$$a = 3\rho\frac{\sigma_v}{\kappa} - \frac{3}{2}\frac{\sigma_v^2}{\kappa^2}, \qquad b = 1 - \rho\frac{\sigma_v}{\kappa} + \frac{1}{4}\frac{\sigma_v^2}{\kappa^2}.$$

*Proof*. Examining the asymptotic limits of the variance and the third cumulant in Propositions 2.1 and 2.2 for small or large $\tau$ gives us the results directly.

**Remark 2.5.1** For small $\tau$, the DS skewness formula shares the same leading order term with our exact one. However, for large $\tau$, the DS skewness formula is given by

$$\text{Skewness\_DS} = 3\rho\frac{\sigma_v}{\kappa}\frac{1}{\sqrt{\theta\tau}} + o\left(\frac{1}{\sqrt{\tau}}\right),$$

which has a systematic difference with our exact one in Equation (2.26).

Figure 2.2 shows the term structure of skewness implied in the Heston model for a set of parameters: $\rho = -0.25$, $\kappa = 5$, $\theta = 0.1$, $a = v_t/\theta = 1.25$ and $\sigma_v = 0.4$. As we can see from this figure, the DS formula has a systematic difference quantitatively for a wide range of $\tau$ from one year up to 20 years, even though it has the same shape of term structure as our exact formula. The asymptotic formulas give a good approximate values for $\tau$ less than two months or $\tau$ larger than 10 years.

## 2.4 The Difference between the DS Formula and Ours

To further justify the correctness of our skewness formula numerically, we conduct Monte Carlo simulations to calculate the skewness implied by the Heston model with different times to maturity, and compare them with the DS formula and ours in Table 2.1. We set parameters to be $\mu = 0.05$, $\kappa = 5$, $\theta = 0.1$, $\sigma_v = 0.4$, $\rho = -0.25$ with starting points $S_0 = 2000$, $v_0 = 0.0125$, and select an intra-daily partition $\mathcal{P}$ with the norm $||\mathcal{P}|| = \frac{1}{2520}$ for the calculation of the third cumulant and variance with times to maturity from 1 month (21 days) to 24 months (504 days).[10] The results show that

[10]The parameters $\kappa = 5, \sigma_v = 0.4, \rho = -0.25, v_0 = 0.0125$ are chosen from the parameter sets used in Das and Sundaram (DS, 1999). However, we set $\theta = 0.1$ in order to highlight the difference between DS's skewness formula and ours. As it is shown in the skewness formula in Equation (2.21), $\mu$ and $S_0$ will not affect the results. These two parameters are chosen to conduct the simulation. Overall, different parameter sets will not alter our conclusion drawn from the simulation.

the skewness computed through simulation converges to the value calculated by our analytical formula when the simulation times increase from 10, $10^2$, $\cdots$ to $10^6$.

The DS formula can be treated as an approximate one for the skewness implied in the Heston model. By comparing it with the values produced by our exact formula in Proposition 2.3, we are able to examine the differences in the numerical results produced by the DS formula. In Table 2.1, we also present numerical results of DS formula and compare them with those of ours. As we can see from the table, the skewness computed through the simulations converges to our formula instead of DS' one.

In Table 2.2, we present the numerical results of the skewness implied in the Heston model computed by using our exact formula and DS. For an easy comparison, we use the same set of parameters as those used in Table 2 of DS (1999). The right columns present the differences of the DS formula for the same set of parameters. We have the following observations: 1) If the stock return and variance processes are independent, the DS formula gives zero skewness. However, the values from our exact formula are not zero due to the contribution of the co-third cumulant between $X_T$ and $Y_T$, $E_t(X_T^2 Y_T)$, and the third cumulant of $Y_T$, $E_t(Y_T^3)$. 2) The difference of the DS formula depends on the volatility of the variance process, $\sigma_v$. The higher the volatility of the variance is, the larger the differences are. 3) The difference of the DS formula also depends on the time to maturity or the length of the return period. The longer the return period is, the larger the differences are. For example, for the set of parameters, $\rho = -0.25$, $\kappa = 1$, $\theta = 0.01$, $a = v_t/\theta = 0.75$, $\sigma_v = 0.4$, and $\tau = 3$ months, the relative difference of the DS formula is roughly 10% of the value given by our exact formula.[11]

Table 2.3 presents the numerical values of the term structure of the skewness implied in the Heston model computed by our exact and DS approximate formulas for a set of parameters that are the same as those in Figure 2.2: $\rho = -0.25$, $\kappa = 5$, $\theta = 0.1$, $a = v_t/\theta = 1.25$ and $\sigma_v = 0.4$. As we can see from this table, the relative difference of the DS formula is around 12% of the value given by our exact formula

[11]This set of parameters is chosen from those of DS. We thank an anonymous referee for pointing out the fact that it violates the Feller's condition ($\kappa\theta = 0.05$, $\frac{1}{2}\sigma_v^2 = 0.08$, hence $\kappa\theta < \frac{1}{2}\sigma_v^2$). In order to have an easy comparison with the numerical values of DS, we decide to keep the parameters the same as those in DS.

for a wide range of times to maturity from one year to 20 years.[12]

The significant difference of the DS formula suggests that one should use our exact formula in the study of skewness by using a Heston-type model.

## 2.5 Application

Our theoretical results on the variance and skewness implied in the Heston model can be applied in calibrating the model.

The Heston (1993) model is one of the most popular models after the Black-Scholes (1973) model, because it is able to capture the stochastic nature of variance with a negative correlation with returns, and it has a closed-form option-pricing formula. However, the literature on calibrated parameters in the Heston model is not conclusive. The estimates of the mean-reverting speed from different authors are very different, as shown in Table 2.4, where the risk-neutral mean-reverting speed ranges from -5.46 to 7.16. Pan (2002) and Duan and Yeh (2010) even obtain negative mean-reverting speeds, which indicates that the variance process is not mean-reverting. Aït-Sahalia and Kimmel (2007) obtain a speedy mean-reverting variance process with a positive market price of variance risk, while Bakshi, Cao and Chen (1997), Eraker (2004) and Garcia et al. (2011) obtain a moderate mean-reverting speed. Recently, Lee (2016) has estimated physical parameters by using a generalized method of moments that is developed with Choe and Lee's (2014) idea of high moment variations. Regarding the estimate of $\rho$, Lee (2016) suggests that the result by the simple method is more reasonable, that is $\rho = -0.6189$.

In the literature, one often calibrate an option-pricing model by minimizing the distance (e.g., the root of mean squared errors) between model and market prices of all available options, see e.g., Bakshi, Cao and Chen (1997). There are three problems with this traditional method. 1) The solution to the optimization problem is often not unique especially when one has a large number of model parameters, which

[12]In an empirical study, one often studies skewness for a couple of months. In this case, the DS formula provides reasonable approximate values, as its difference is only around 5%. Furthermore, what matters in empirical studies is often a ranking of stock return skewness instead of precise numerical values. That is why finance researchers are not aware of the fact that the DS formula in fact does not apply in the Heston model. We contribute to the literature by pointing out this fact and providing an exact skewness formula implied in the Heston model.

is the case for some advanced affine jump-diffusion models. The problem of non-uniqueness makes the calibrated model parameters unstable and unreliable. 2) The method requires full set of market options price data, which is large and inconvenient for some users to acquire. 3) The computation of solving optimization problem with many unknowns is time-consuming.

Here in this chapter, we provide an efficient and effective way to calibrate the parameters in Heston model by making good use of the market data of the VIX and SKEW term structures. With a careful design, we obtain model parameters sequentially to avoid solving optimization problem with many unknowns. In particular, we determined the mean-reverting speed, $\kappa$, the instantaneous variance, $v_t$, and long-term mean level of variance, $\theta$ (or $\theta_t$), from the VIX term structure. The correlation coefficient, $\rho$, is determined from the time series of SPX and VIX. Finally, we determine the volatility of variance, $\sigma_v$, from the SKEW term structure. With this procedure, we solve optimization problems with at most two unknowns each time, and obtain unique solutions very quickly. Given the fact that the VIX and SKEW term structures are now public information, anyone can acquire the data and replicate our calibration procedure easily. We will have a chance to unify the calibrated model parameters and make them standardized. Besides, the VIX and SKEW term structures are not trivial information. They are carefully designed and calculated by the CBOE from the SPX options prices with all strikes and maturities to measure the levels and slopes of implied volatility curve, i.e., risk-neutral volatility and skewness at different times to maturity. Making good use of these valuable information is certainly what we should do in the future.

We will demonstrate our sequential method by calibrating the Heston model. The idea of calibrating option-pricing model using the VIX and SKEW has been extended by Zhen and Zhang (2014) to handle more advanced affine jump-diffusion models with many more parameters.

### 2.5.1 Calibrating the Heston Model on One Particular Day

We now demonstrate our estimation procedure by using a randomly chosen date, 22 June 2015. We choose the maturity dates when both the closing quotes of VIX

and SKEW are available, and the term structure data of VIX and SKEW on this day are as follows:

| Maturity Date | VIX | SKEW |
|---|---|---|
| 17-Jul-15 | 12.40 | 120.32 |
| 21-Aug-15 | 13.77 | 121.65 |
| 18-Sep-15 | 14.83 | 122.35 |
| 19-Dec-15 | 15.50 | 128.34 |
| 15-Jan-16 | 16.87 | 121.69 |
| 18-Mar-16 | 16.93 | 121.51 |
| 17-Jun-16 | 17.83 | 124.45 |
| 16-Dec-16 | 18.97 | 128.76 |
| 16-Jun-17 | 20.42 | 130.85 |

The VIX formula under the Heston (1993) model is given by

$$VIX_{t,T} = 100\sqrt{\frac{1}{T-t}E_t^Q(X_T^2)},$$

where $E^Q$ denotes the expectation under the risk-neutral measure. Using a couple of lines of *Mathematica* code, we can easily fit the market quotes of the VIX term structure with the theoretical formula and obtain

$$v_t = 0.0143, \qquad \theta = 0.0531, \qquad \kappa = 1.45,$$

with a root of mean squared error (RMSE) of 0.3603.

The SKEW, defined by the CBOE, is a scaled skewness in Equation (2.21) under the risk-neutral measure. Its formula is given by

$$\mathrm{SKEW_{t,T}} = 100 - 10\frac{E_t^Q(X_T^3) - \frac{3}{2}E_t^Q(X_T^2Y_T) + \frac{3}{4}E_t^Q(X_TY_T^2) - \frac{1}{8}E_t^Q(Y_T^3)}{\left[E_t^Q(X_T^2) - E_t^Q(X_TY_T) + \frac{1}{4}E_t^Q(Y_T^2)\right]^{3/2}}.$$

Given the estimate of $\{v_t, \theta, \kappa\}$, we can easily fit the market quotes of the SKEW term structure with the theoretical formula and obtain

$$\sigma_v = 0.7325, \qquad \rho = -0.7177, \qquad \Longrightarrow \qquad \rho\sigma_v = -0.5257,$$

with a RMSE of 3.4847. For the comparison with the DS model, we fit the SKEW term structure data with the following DS formula

$$\mathrm{SKEW\_DS_{t,T}} = 100 - 10\frac{E_t^Q(X_T^3)}{\left[E_t^Q(X_T^2)\right]^{3/2}}.$$

As the VIX level estimation are the same for both the Heston model and the DS model, noting that $\rho$ and $\sigma_v$ appear as a product in the DS skewness formula in Equation (2.25), we can only obtain ($\rho$ and $\sigma_v$ cannot be separated)

$$\rho\sigma_v = -0.5556,$$

(slightly different from $-0.5257$) with a RMSE of 3.5566, which is slightly higher than 3.4847 using our skewness formula.

Both the market data and fitted curves are presented in Figure 2.3. As we see from this figure, the fitting performance for the VIX is reasonably good, but that for the SKEW is poor for the short term. This is because the skewness implied in the Heston model goes to zero as the time to maturity becomes very short. In order to produce a reasonable value for the short-term skewness, one has to introduce jumps in the stock returns in the model. The importance of jumps in prices has been noticed and examined by, among others, Press (1967), Ball and Torous (1983, 1985), Andersen, Benzoni and Lund (2002) and Yan (2011), empirically using returns data, options data or the joint data.

From the lower panel of Figure 2.3, we can see that the calibrated DS skewness formula is slightly different from the calibrated our skewness formula for the Heston model. In terms of fitting performance to the market skewness data, neither is doing a reasonably good job due to the lack of jump components in the Heston model.

### 2.5.2 Calibrating the Heston Model for a Period of Time

We further calibrate the Heston model by using the market data for a period from 24 November 2010 through 30 June 2015.

We adopt a two-step iterative procedure used by Christoffersen, Heston and Jacobs (2009) and Luo and Zhang (2012). First, for a fixed set of structural parameters

$\{\theta, \kappa\}$, we estimate $\{v_t\}$ on each day $t$ by using the daily data of the VIX term structure. Second, with the estimated time series of $\{v_t\}$, we solve the optimization problem to obtain an optimal set of $\{\theta, \kappa\}$. The two-step process is iterated until there is no further improvement in RMSE.

The estimated structural parameters are as follows:

$$\theta = 0.1651, \qquad \kappa = 0.28,$$

and the daily realizations for instantaneous variance $\{v_t\}$ are presented in Figure 2.4 with an average of 0.0342. The RMSE for the whole period is 1.353, which is three times larger than that (0.3603) for daily calibration in the previous section. Due to the restriction of the fixed $\theta$, one cannot fit the VIX term structure trend for some days. The large error mainly comes from this restriction. In order to enhance model performance, we have to modify the Heston model by introducing a new factor of stochastic long-term mean.

Our theoretical analysis in the previous sections shows that $\rho$ and $\sigma_v$ often appear together as a product, which makes the solution of the estimation not unique. In the DS formula in Equation 2.25, we can see that $\rho\sigma_v$ appears together as a product, so one cannot separate them. In our formula in Proposition 2.3, even though $\rho$ and $\sigma_v$ can be separated, but the key contribution comes from DS formula. Therefore our formula is not sensitive in separating $\rho$ and $\sigma_v$. Therefore we propose to estimate $\rho$ by computing the correlation coefficient between the SPX returns and the changes of the 30-day VIX square for the sample period from 24 November 2010 through 30 June 2015, and obtain

$$\rho = -0.7872.$$

Specifically, we get the SPX daily closing data from Bloomberg, and 30-day VIX index from the CBOE web site, then obtain the changes of SPX and VIX square to calculate the unconditional correlation coefficient of the two time series with a standard method.

Because $\rho\sigma_v$ as a product is the most important parameter to determine the skewness, we have to determine one of them through some other ways and the other one

using the SKEW. We chose to estimate $\rho$ by using some simple method, i.e., computing the aggregate correlation coefficient of two time series, SPX daily returns and the changes of VIX square.

Given the optimal values $\{v_t, \theta, \kappa\}$ estimated from VIX term structure data and the correlation coefficient $\rho$, we fit the SKEW term structure with our formula and obtain

$$\sigma_v = 0.5914$$

with a RMSE of 8.8953.

To have some feeling about the importance of changing volatility of volatility, $\sigma_v$, we have done a numerical experiment by allowing $\sigma_v$ to be floating. We obtain daily $\sigma_v$ and present them in Figure 2.4. The average $\sigma_v$ is 0.6101, which is very close to the unconditional estimation of 0.5914. The RMSE for SKEW is 8.2364, which is around 10% smaller than the previous error of 8.8953 with a fixed $\sigma_v$. The numerical fitting exercise shows that the floating volatility of volatility does not reduce the error much, hence it is not critical in capturing the fluctuations of the SKEW.

To justify the importance of the stochastic long-term mean, we have done a calibration exercise by using floating $\theta_t$. We adopt the same estimation procedure as above. The calibration results are presented in Table 2.5 and Figure 2.5. The long-term mean level is higher than the average of the instantaneous variance for both fixed and floating long-run volatility cases. This is consistent with the phenomenon of upward-sloping VIX term structure observed in the literature, see e.g., Zhang, Shu and Brenner (2010). Due to the negative variance risk premium, the long-term mean level of variance in the risk-neutral measure is larger than that in the physical measure, see e.g., Zhang and Huang (2010) for a discussion on the issue. As we can see from Table 2.5, the RMSE of fitting the VIX term structure has been dramatically reduced to be less than half, from 1.3531 for the fixed $\theta$ to 0.5473 for the floating $\theta_t$. The RMSE of fitting the SKEW term structure has also been reduced from 8.8953 to 7.9561. The numerical evidence of fitting performance shows that the stochastic long-term mean is important in capturing both the VIX and SKEW term structures. This observation is consistent with the literature on modeling the term structure of volatility, see, among others, Christoffersen et al. (2008), Egloff et al. (2010), Luo

and Zhang (2012). Their research shows that the fitting performance of the models with long-run volatility component is much beter than that of the models with single short-run volatility component.

## 2.6 Conclusion

In this chapter, we study the skewness of stock returns implied in the Heston (1993) model. The problem has been studied by Das and Sundaram (DS 1999) for a similar stock price model with stochastic volatility, but the DS model is not the Heston model. In this chapter, we show that DS's formula in Proposition 2 on page 221 in their paper does not apply in the Heston model.

By working with the expectation of co-moments of integrated stock return and volatility processes, denoted as $X_T$ and $Y_T$ in this chapter, we derive a skewness formula in closed form. We determine the reason why the DS formula does not apply in the Heston model. By treating the DS formula as an approximate one, we study the difference between its numerical results and ours. Numerical exercises show that the DS formula could have more than 10% difference for some reasonable set of parameters. Our research suggests that one should use our exact formula in the study of skewness by using the Heston model.

Our skewness formula can be used to calibrate the Heston model by using the market data of the CBOE VIX and SKEW term structures. Our calibration exercise shows that the Heston model is not able to capture finite short-term skewness due to its lack of jumps in stock returns. We also show that the long-term mean level of variance is time-varying. In order to enhance the performance of the Heston model, it is important to incorporate this additional factor into the model. The skewness implied in an affine jump-diffusion model with stochastic long-term mean is a topic for further research.

The method presented in this chapter can be applied to study the kurtosis implied in the Heston model, but the result is lengthy. We have decided to present it in a subsequent research paper.

## 2.7 Appendix

### 2.7.1 Proof of Proposition 2.2

Our strategy of proving Proposition 2.2 is as follows. Using Itô's Lemma skillfully, we reduce the expectations $E_t(X_T^3)$, $E_t(X_T^2 Y_T)$ and $E_t(X_T Y_T^2)$ into expectations of time-weighted sums of $X_s v_s$ and $Y_s^* v_s$ in Equations (2.29), (2.30) and (2.31) respectively. We further reduce them into time-weighted sums of expected instantaneous variance by using the result of Lemma 1, which is derived by using Itô's Lemma again. The affine property of the variance process is crucial during the derivation of Lemma 1.

We need the following results in proving Proposition 2.2.

**Lemma 1** Given the definition of $X_s$, $Y_s^*$ and $v_s$ as follows

$$
\begin{aligned}
&X_s = \int_t^s \sqrt{v_u} dB_u^S, \qquad dX_s = \sqrt{v_s} dB_s^S, \\
&Y_s^* = \sigma_v \int_t^s \frac{1-e^{-\kappa(T-u)}}{\kappa} \sqrt{v_u} dB_u^v, \qquad dY_s^* = \sigma_v \frac{1-e^{-\kappa(T-s)}}{\kappa} \sqrt{v_s} dB_s^v. \\
&dv_s = \kappa(\theta - v_s) ds + \sigma_v \sqrt{v_s} dB_s^v, \qquad dB_t^S dB_t^v = \rho dt,
\end{aligned}
$$

we have

$$
\textbf{L1.1} \quad E_t(X_s v_s) = \rho \sigma_v \int_t^s e^{-\kappa(s-u)} E_t(v_u) du, \tag{2.27}
$$

$$
\textbf{L1.2} \quad E_t(Y_s^* v_s) = \sigma_v^2 \int_t^s e^{-\kappa(s-u)} \frac{1-e^{-\kappa(T-u)}}{\kappa} E_t(v_u) du, \tag{2.28}
$$

*Proof*. See Appendix 2.7.2.

Using Itô's Lemma and the martingale property of $X_s$, we have

$$
E_t(X_T^3) = E_t \int_t^T d(X_s^3) = E_t \int_t^T 3X_s^2 dX_s + 3X_s (dX_s)^2 = 3 \int_t^T E_t(X_s v_s) ds. \tag{2.29}
$$

Substituting Equation (2.27) in Lemma 1 into this equation gives

$$\begin{aligned}E_t(X_T^3) &= 3\int_t^T \rho\sigma_v \int_t^s e^{-\kappa(s-u)} E_t(v_u)duds = 3\rho\sigma_v \int_t^T \left[\int_u^T e^{-\kappa(s-u)}ds\right] E_t(v_u)du \\ &= 3\rho\sigma_v \int_t^T \frac{1-e^{-\kappa(T-u)}}{\kappa} E_t(v_u)du,\end{aligned}$$

which is equivalent to Equation (2.17).

Similarly using Itô's Lemma and the martingale property of $X_s$ and $Y_s^*$, we have

$$\begin{aligned}E_t(X_T^2 Y_T) &= E_t(X_T^2 Y_T^*) = E_t \int_t^T d(X_s^2 Y_s^*) \\ &= E_t \int_t^T 2X_s Y_s^* dX_s + X_s^2 dY_s^* + Y_s^*(dX_s)^2 + 2X_s dX_s dY_s^* \\ &= \int_t^T E_t(Y_s^* v_s)ds + 2\rho\sigma_v \int_t^T \frac{1-e^{-\kappa(T-s)}}{\kappa} E_t(X_s v_s)ds \qquad (2.30)\end{aligned}$$

Substituting Equations (2.27) and (2.28) in Lemma 1 into this equation gives

$$\begin{aligned}E_t(X_T^2 Y_T) &= \int_t^T \sigma_v^2 \int_t^s e^{-\kappa(s-u)} \frac{1-e^{-\kappa(T-u)}}{\kappa} E_t(v_u)duds \\ &\quad +2\rho\sigma_v \int_t^T \frac{1-e^{-\kappa(T-s)}}{\kappa} \rho\sigma_v \int_t^s e^{-\kappa(s-u)} E_t(v_u)duds \\ &= \sigma_v^2 \int_t^T \left[\int_u^T e^{-\kappa(s-u)}ds\right] \frac{1-e^{-\kappa(T-u)}}{\kappa} E_t(v_u)du \\ &\quad +2\rho^2\sigma_v^2 \int_t^T \left[\int_u^T \frac{1-e^{-\kappa(T-s)}}{\kappa} e^{-\kappa(s-u)}ds\right] E_t(v_u)du \\ &= \sigma_v^2 \int_t^T \left[\frac{(1-e^{-\kappa(T-u)})^2}{\kappa^2} + 2\rho^2 \frac{1-e^{-\kappa(T-u)} - \kappa(T-u)e^{-\kappa(T-u)}}{\kappa^2}\right] E_t(v_u)du,\end{aligned}$$

which is equivalent to Equation (2.18).

Using Itô's Lemma and the martingale property of $X_s$ and $Y_s^*$, we have

$$\begin{aligned}
E_t(X_T Y_T^2) = E_t(X_T Y_T^{*2}) &= E_t \int_t^T d(X_s Y_s^{*2}) \\
&= E_t \int_t^T Y_s^{*2} dX_s + 2X_s Y_s^* dY_s^* + 2Y_s^* dX_s dY_s^* + X_s (dY_s^*)^2 \\
&= 2\rho\sigma_v \int_t^T \frac{1-e^{-\kappa(T-s)}}{\kappa} E_t(Y_s^* v_s) ds + \sigma_v^2 \int_t^T \frac{(1-e^{-\kappa(T-s)})^2}{\kappa^2} E_t(X_s v_s) ds
\end{aligned} \tag{2.31}$$

Substituting Equations (2.27) and (2.28) in Lemma 1 into this equation gives

$$\begin{aligned}
E_t(X_T Y_T^2) &= 2\rho\sigma_v \int_t^T \frac{1-e^{-\kappa(T-s)}}{\kappa} \sigma_v^2 \int_t^s e^{-\kappa(s-u)} \frac{1-e^{-\kappa(T-u)}}{\kappa} E_t(v_u) du ds \\
&\quad + \sigma_v^2 \int_t^T \frac{(1-e^{-\kappa(T-s)})^2}{\kappa^2} \rho\sigma_v \int_t^s e^{-\kappa(s-u)} E_t(v_u) du ds \\
&= 2\rho\sigma_v^3 \int_t^T \left[ \int_u^T \frac{1-e^{-\kappa(T-s)}}{\kappa} e^{-\kappa(s-u)} ds \right] \frac{1-e^{-\kappa(T-u)}}{\kappa} E_t(v_u) du \\
&\quad + \rho\sigma_v^3 \int_t^T \left[ \int_u^T \frac{(1-e^{-\kappa(T-s)})^2}{\kappa^2} e^{-\kappa(s-u)} ds \right] E_t(v_u) du \\
&= 2\rho\sigma_v^3 \int_t^T \frac{1-e^{-\kappa(T-u)} - \kappa(T-u)e^{-\kappa(T-u)}}{\kappa^2} \frac{1-e^{-\kappa(T-u)}}{\kappa} E_t(v_u) du \\
&\quad + \rho\sigma_v^3 \int_t^T \frac{1-e^{-2\kappa(T-u)} - 2\kappa(T-u)e^{-\kappa(T-u)}}{\kappa^3} E_t(v_u) du,
\end{aligned}$$

which is equivalent to Equation (2.19).

Using Itô's Lemma and the martingale property of $Y_s^*$, we have

$$\begin{aligned}
E_t(Y_T^3) = E_t(Y_T^{*3}) = E_t \int_t^T d(Y_s^{*3}) &= E_t \int_t^T 3Y_s^{*2} dY_s^* + 3Y_s^* (dY_s^*)^2 \\
&= 3\sigma_v^2 \int_t^T \frac{(1-e^{-\kappa(T-s)})^2}{\kappa^2} E_t(Y_s^* v_s) ds
\end{aligned}$$

Substituting Equation (2.28) in Lemma 1 into this equation gives

$$\begin{aligned}
E_t(Y_T^3) &= 3\sigma_v^2 \int_t^T \frac{(1-e^{-\kappa(T-s)})^2}{\kappa^2} \sigma_v^2 \int_t^s e^{-\kappa(s-u)} \frac{1-e^{-\kappa(T-u)}}{\kappa} E_t(v_u) du ds \\
&= 3\sigma_v^4 \int_t^T \left[ \int_u^T \frac{(1-e^{-\kappa(T-s)})^2}{\kappa^2} e^{-\kappa(s-u)} ds \right] \frac{1-e^{-\kappa(T-u)}}{\kappa} E_t(v_u) du \\
&= 3\sigma_v^4 \int_t^T \frac{1-e^{-2\kappa(T-u)} - 2\kappa(T-u)e^{-\kappa(T-u)}}{\kappa^3} \frac{1-e^{-\kappa(T-u)}}{\kappa} E_t(v_u) du,
\end{aligned}$$

which is equivalent to Equation (2.20).
**Q.E.D.**

### 2.7.2 Proof of Lemma 1

**L1.1**: Using Itô's Lemma and the martingale property of $X_s$, we have

$$\begin{aligned}
E_t(X_s v_s) &= E_t \int_t^s d(X_u v_u) = E_t \int_t^s v_u dX_u + X_u dv_u + dX_u dv_u \\
&= E_t \int_t^s X_u \kappa(\theta - v_u) du + \rho\sigma_v v_u du \\
&= -\kappa \int_t^s E_t(X_u v_u) du + \rho\sigma_v \int_t^s E_t(v_u) du.
\end{aligned}$$

As we can see, the affine property of the variance process is crucial in obtaining the ordinary differential equation (ODE). Solving the ODE gives

$$\begin{aligned}
\int_t^s E_t(X_u v_u) du &= \rho\sigma_v \int_t^s e^{-\kappa(s-m)} \int_t^m E_t(v_u) du dm \\
&= \rho\sigma_v \int_t^s \frac{1-e^{-\kappa(s-u)}}{\kappa} E_t(v_u) du.
\end{aligned}$$

Taking differentiation with respect to $s$ gives Equation (2.27)

$$E_t(X_s v_s) = \rho\sigma_v \int_t^s e^{-\kappa(s-u)} E_t(v_u) du.$$

**L1.2**: Using Itô's Lemma and the martingale property of $Y_s^*$, we have

$$\begin{aligned}
E_t(Y_s^* v_s) &= E_t \int_t^s d(Y_u^* v_u) = E_t \int_t^s v_u dY_u^* + Y_u^* dv_u + dY_u^* dv_u \\
&= E_t \int_t^s Y_u^* \kappa(\theta - v_u) du + \sigma_v^2 \frac{1 - e^{-\kappa(T-u)}}{\kappa} v_u du \\
&= -\kappa \int_t^s E_t(Y_u^* v_u) du + \sigma_v^2 \int_t^s \frac{1 - e^{-\kappa(T-u)}}{\kappa} E_t(v_u) du.
\end{aligned}$$

Solving the ODE gives

$$\begin{aligned}
\int_t^s E_t(Y_u^* v_u) du &= \sigma_v^2 \int_t^s e^{-\kappa(s-m)} \int_t^m \frac{1 - e^{-\kappa(T-u)}}{\kappa} E_t(v_u) du dm \\
&= \sigma_v^2 \int_t^s \frac{1 - e^{-\kappa(s-u)}}{\kappa} \frac{1 - e^{-\kappa(T-u)}}{\kappa} E_t(v_u) du.
\end{aligned}$$

Taking differentiation with respect to $s$ gives Equation (2.28)

$$E_t(Y_s^* v_s) = \sigma_v^2 \int_t^s e^{-\kappa(s-u)} \frac{1 - e^{-\kappa(T-u)}}{\kappa} E_t(v_u) du.$$

TABLE 2.1: The Results of Monte Carlo Simulations for the Heston (1993) Model with a Given Set of Parameters

The results of Monte Carlo simulations for the Heston (1993) model with a set of parameters $\mu = 0.05$, $\kappa = 5$, $\theta = 0.1$, $\sigma_v = 0.4$, $\rho = -0.25$ and starting points $S_0 = 2000$, $v_0 = 0.0125$. We choose an intra-daily partition $\mathcal{P}$ with the norm $||\mathcal{P}|| = \frac{1}{2520}$ for the calculation of the third cumulant and variance with times to maturity from 1 month (21 days) to 24 months (504 days). The last row reports the root mean squared error (RMSE) between the value in the corresponding column and the skewness calculated using our formula.

| | Simulation Times | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Maturities | 10 | $10^2$ | $10^3$ | $10^4$ | $10^5$ | $10^6$ | DS | Ours |
| 1 | -1.6670 | -0.6072 | -0.3274 | -0.2323 | -0.2010 | -0.1895 | -0.1882 | -0.1941 |
| 2 | -1.6270 | -0.9699 | -0.2139 | -0.1843 | -0.2190 | -0.2072 | -0.1940 | -0.2043 |
| 3 | 0.5155 | -0.8617 | -0.1947 | -0.1894 | -0.2170 | -0.2050 | -0.1934 | -0.2068 |
| 4 | 0.1241 | -0.6627 | -0.2425 | -0.2006 | -0.2201 | -0.2017 | -0.1909 | -0.2066 |
| 5 | -0.9266 | -0.6036 | -0.1041 | -0.2082 | -0.2254 | -0.2056 | -0.1877 | -0.2050 |
| 6 | -1.4280 | -0.6240 | -0.2227 | -0.1966 | -0.2186 | -0.1957 | -0.1840 | -0.2025 |
| 7 | -1.9250 | -0.4719 | -0.2745 | -0.2167 | -0.2191 | -0.1966 | -0.1800 | -0.1993 |
| 8 | -2.5830 | -0.3873 | -0.3106 | -0.2144 | -0.2205 | -0.1955 | -0.1760 | -0.1958 |
| 9 | -2.4120 | -0.6164 | -0.2730 | -0.2052 | -0.2050 | -0.1924 | -0.1719 | -0.1920 |
| 10 | -0.6370 | -0.5244 | -0.1281 | -0.1719 | -0.1875 | -0.1870 | -0.1678 | -0.1881 |
| 11 | -0.5683 | -0.5968 | -0.2019 | -0.1671 | -0.1925 | -0.1845 | -0.1639 | -0.1841 |
| 12 | -0.7109 | -0.6681 | -0.2909 | -0.1768 | -0.1826 | -0.1808 | -0.1600 | -0.1802 |
| 13 | -0.6125 | -0.4361 | -0.2585 | -0.1580 | -0.1783 | -0.1756 | -0.1563 | -0.1764 |
| 14 | -1.1200 | -0.4871 | -0.2362 | -0.1372 | -0.1730 | -0.1701 | -0.1528 | -0.1726 |
| 15 | -1.4150 | -0.4761 | -0.2666 | -0.1088 | -0.1636 | -0.1672 | -0.1494 | -0.1690 |
| 16 | -0.9909 | -0.3503 | -0.2857 | -0.1124 | -0.1665 | -0.1644 | -0.1461 | -0.1655 |
| 17 | -0.3583 | -0.2947 | -0.2834 | -0.1115 | -0.1562 | -0.1613 | -0.1430 | -0.1622 |
| 18 | -0.3183 | -0.3745 | -0.3373 | -0.1030 | -0.1612 | -0.1588 | -0.1401 | -0.1590 |
| 19 | 0.0479 | -0.3972 | -0.3521 | -0.0981 | -0.1524 | -0.1576 | -0.1372 | -0.1559 |
| 20 | 0.1296 | -0.1924 | -0.3507 | -0.0999 | -0.1522 | -0.1556 | -0.1346 | -0.1530 |
| 21 | 0.4031 | -0.1340 | -0.2980 | -0.0987 | -0.1468 | -0.1532 | -0.1320 | -0.1502 |
| 22 | 0.1081 | 0.0441 | -0.3034 | -0.1085 | -0.1460 | -0.1479 | -0.1296 | -0.1475 |
| 23 | 0.2747 | 0.0654 | -0.2234 | -0.0837 | -0.1364 | -0.1441 | -0.1273 | -0.1449 |
| 24 | 0.1392 | 0.1121 | -0.1920 | -0.0675 | -0.1320 | -0.1418 | -0.1251 | -0.1425 |
| RMSE | 1.0250 | 0.3598 | 0.1090 | 0.0393 | 0.0107 | 0.0025 | 0.0181 | |

TABLE 2.2: A Comparison of Skewness Implied in the Heston (1993) Model Computed by Using Das and Sundaram's (1999) Formula and Our Exact One

A comparison of skewness implied in the Heston (1993) model computed by using Das and Sundaram's (DS 1999) formula and our exact one (Ours). For an easy comparison, the parameters values are set to be the same as those in Table 2 of DS (1999). The long-term mean level of variance is set to be $\theta = 0.01$, and $a$ is defined as $a = v_t/\theta$. The Difference is defined as $\text{Difference} = \text{DS} - \text{Ours}$.

| Parameters | | | | Ours | | | DS | | | Difference | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\rho$ | $\kappa$ | $a$ | $\sigma_v$ | 1 W | 1 M | 3 Ms | 1 W | 1 M | 3 Ms | 1 W | 1 M | 3 Ms |
| 0 | 1 | 0.75 | 0.1 | -0.0002 | -0.0013 | -0.0058 | 0 | 0 | 0 | 0.0002 | 0.0013 | 0.0058 |
| 0 | 1 | 1.00 | 0.1 | -0.0001 | -0.0011 | -0.0052 | 0 | 0 | 0 | 0.0001 | 0.0011 | 0.0052 |
| 0 | 1 | 1.25 | 0.1 | -0.0001 | -0.0010 | -0.0048 | 0 | 0 | 0 | 0.0001 | 0.0010 | 0.0048 |
| 0 | 5 | 0.75 | 0.1 | -0.0001 | -0.0010 | -0.0028 | 0 | 0 | 0 | 0.0001 | 0.0010 | 0.0028 |
| 0 | 5 | 1.00 | 0.1 | -0.0001 | -0.0009 | -0.0027 | 0 | 0 | 0 | 0.0001 | 0.0009 | 0.0027 |
| 0 | 5 | 1.25 | 0.1 | -0.0001 | -0.0008 | -0.0026 | 0 | 0 | 0 | 0.0001 | 0.0008 | 0.0026 |
| -0.25 | 1 | 0.75 | 0.1 | -0.0596 | -0.1215 | -0.1989 | -0.0595 | -0.1203 | -0.1936 | 0.0001 | 0.0012 | 0.0053 |
| -0.25 | 1 | 1.00 | 0.1 | -0.0518 | -0.1063 | -0.1776 | -0.0517 | -0.1053 | -0.1728 | 0.0001 | 0.0010 | 0.0048 |
| -0.25 | 1 | 1.25 | 0.1 | -0.0464 | -0.0958 | -0.1619 | -0.0463 | -0.0948 | -0.1575 | 0.0001 | 0.0010 | 0.0044 |
| -0.25 | 5 | 0.75 | 0.1 | -0.0576 | -0.1053 | -0.1379 | -0.0574 | -0.1044 | -0.1354 | 0.0002 | 0.0009 | 0.0025 |
| -0.25 | 5 | 1.00 | 0.1 | -0.0505 | -0.0955 | -0.1313 | -0.0504 | -0.0947 | -0.1288 | 0.0001 | 0.0008 | 0.0025 |
| -0.25 | 5 | 1.25 | 0.1 | -0.0455 | -0.0880 | -0.1253 | -0.0454 | -0.0872 | -0.1229 | 0.0001 | 0.0008 | 0.0024 |
| 0 | 1 | 0.75 | 0.4 | -0.0024 | -0.0206 | -0.0927 | 0 | 0 | 0 | 0.0024 | 0.0206 | 0.0927 |
| 0 | 1 | 1.00 | 0.4 | -0.0021 | -0.0181 | -0.0832 | 0 | 0 | 0 | 0.0021 | 0.0181 | 0.0832 |
| 0 | 1 | 1.25 | 0.4 | -0.0019 | -0.0163 | -0.0761 | 0 | 0 | 0 | 0.0019 | 0.0163 | 0.0761 |
| 0 | 5 | 0.75 | 0.4 | -0.0023 | -0.0156 | -0.0448 | 0 | 0 | 0 | 0.0023 | 0.0156 | 0.0448 |
| 0 | 5 | 1.00 | 0.4 | -0.0020 | -0.0142 | -0.0433 | 0 | 0 | 0 | 0.0020 | 0.0142 | 0.0433 |
| 0 | 5 | 1.25 | 0.4 | -0.0018 | -0.0132 | -0.0418 | 0 | 0 | 0 | 0.0018 | 0.0132 | 0.0418 |
| -0.25 | 1 | 0.75 | 0.4 | -0.2403 | -0.5000 | -0.8588 | -0.2380 | -0.4811 | -0.7744 | 0.0023 | 0.0189 | 0.0844 |
| -0.25 | 1 | 1.00 | 0.4 | -0.2086 | -0.4378 | -0.7669 | -0.2067 | -0.4212 | -0.6912 | 0.0019 | 0.0166 | 0.0757 |
| -0.25 | 1 | 1.25 | 0.4 | -0.1869 | -0.3943 | -0.6992 | -0.1852 | -0.3793 | -0.6300 | 0.0017 | 0.0150 | 0.0692 |
| -0.25 | 5 | 0.75 | 0.4 | -0.2318 | -0.4318 | -0.5826 | -0.2298 | -0.4175 | -0.5414 | 0.0020 | 0.0143 | 0.0412 |
| -0.25 | 5 | 1.00 | 0.4 | -0.2033 | -0.3917 | -0.5548 | -0.2015 | -0.3786 | -0.5150 | 0.0018 | 0.0131 | 0.0398 |
| -0.25 | 5 | 1.25 | 0.4 | -0.1833 | -0.3609 | -0.5298 | -0.1816 | -0.3488 | -0.4915 | 0.0017 | 0.0121 | 0.0383 |

TABLE 2.3: The Term Structure of Skewness Implied in the Heston (1993) Model for a Given Set of Parameters

The term structure of skewness implied in the Heston (1993) model for a set of parameters: $\rho = -0.25, \kappa = 5, \theta = 0.1, a = v_t/\theta = 1.25$ and $\sigma_v = 0.4$ by using our exact formula (Ours) and DS approximate formula (DS). The difference ($\mathrm{D} = \mathrm{DS} - \mathrm{Ours}$) and relative difference ($\mathrm{RD} = \mathrm{D}/|\mathrm{Ours}|$) are also presented for different maturities.

| Maturity | Ours | DS | D | RD |
|---|---|---|---|---|
| 1/365 | -0.0222 | -0.0221 | 0.0001 | 0.1% |
| 1/52 | -0.0580 | -0.0574 | 0.0006 | 0.9% |
| 1/12 | -0.1141 | -0.1103 | 0.0038 | 3.5% |
| 1/6 | -0.1489 | -0.1405 | 0.0084 | 5.6% |
| 1/4 | -0.1675 | -0.1554 | 0.0121 | 7.2% |
| 1/3 | -0.1775 | -0.1626 | 0.0149 | 8.4% |
| 0.5 | -0.1837 | -0.1656 | 0.0181 | 9.8% |
| 0.75 | -0.1789 | -0.1592 | 0.0196 | 11.0% |
| 1 | -0.1694 | -0.1497 | 0.0196 | 11.5% |
| 1.5 | -0.1509 | -0.1327 | 0.0182 | 12.1% |
| 2 | -0.1363 | -0.1196 | 0.0168 | 12.3% |
| 3 | -0.1160 | -0.1015 | 0.0145 | 12.5% |
| 5 | -0.0928 | -0.0811 | 0.0117 | 12.7% |
| 10 | -0.0672 | -0.0587 | 0.0086 | 12.8% |
| 15 | -0.0553 | -0.0483 | 0.0071 | 12.8% |
| 20 | -0.0481 | -0.0420 | 0.0062 | 12.8% |

TABLE 2.4: A Comparison of Estimation Results for the Heston (1993) Model

A comparison of estimation results for the Heston (1993) model. This table shows the physical parameters $\{\kappa^P$, $\kappa\theta$, $\sigma_v$, $\rho\}$ and risk-neutral mean-reverting speed $\kappa$ estimated by Bakshi, Cao and Chen (BCC1997), Pan (P2002), Eraker (E2004), Aït-Sahalia and Kimmel (AK2007), Duan and Yeh (DY2010), Garcia, Lewis, Pastorello and Renault (GLPR2011), Lee (L2016) and this chapter. The parameters $\kappa\theta$, $\rho$ and $\sigma_v$ are the same in the two measures. The parameters $\{\kappa^P$, $\kappa\theta$, $\kappa\}$ estimated by Eraker (2004) and Garcia, Lewis, Pastorello, and Renault (2011) have been annualized so that they are comparable with the results of others.

| Authors | Data Types | Sample Periods | Estimation Results | | | | |
|---|---|---|---|---|---|---|---|
| | | | $\kappa^P$ | $\kappa$ | $\kappa\theta$ | $\sigma_v$ | $\rho$ |
| BCC1997 | Options | Jun1988-May1991 | | 1.15 | 0.04 | 0.39 | -0.64 |
| P2002 | SPX&Options | Jan1989-Dec1996 | 7.1 | -0.5 | 0.0973 | 0.32 | -0.53 |
| E2004 | SPX&Options | Jan1970-Dec1990 | 4.79 | 2.27 | 0.2332 | 0.22 | -0.57 |
| AK2007 | SPX&VIX | Jan1990-Sep2003 | 5.13 | 7.16 | 0.2237 | 0.52 | -0.75 |
| DY2010 | SPX&VIX | Jan1990-Aug2007 | 5.23 | -5.46 | 0.1387 | 0.39 | -0.67 |
| GLPR2011 | 5-min SPX&Options | Jan1996-Dec2005 | 6.80 | 2.52 | 0.1447 | 0.19 | -0.22 |
| L2016 | SPX | 2003-2007 | 10.05 | | 0.1969 | 0.85 | -0.62 |
| This chapter | SPX&VIX&SKEW | Nov2010-Jun2015 | | 0.28 | 0.0462 | 0.59 | -0.79 |

TABLE 2.5: The Results of Our Calibration on the Heston (1993) Model and Its Extensions by Using the Market Data of the VIX and SKEW Term Structures

The results of our calibration on the Heston (1993) model and its extensions by using the market data of the VIX and SKEW term structures. This table shows the risk-neutral parameters fitted with the term structures of VIX and SKEW. The sample period is from 24 November 2010 to 30 June 2015. The correlation coefficient are calculated using the SPX returns and the changes of the VIX square. RMSE stands for the root mean squared error. $Std(v)$ stands for the standard deviation of the time series $v_t$.

Panel A: Estimation using VIX

| | $\kappa$ | $\bar{v}$ | $Std(v)$ | $\bar{\theta}$ | $Std(\theta)$ | RMSE |
|---|---|---|---|---|---|---|
| Fixed $\theta$ | 0.28 | 0.0342 | 0.0264 | 0.1651 | - | 1.3531 |
| Floating $\theta$ | 1.91 | 0.0299 | 0.0289 | 0.0729 | 0.0253 | 0.5473 |

Panel B: Estimation using SPX and VIX, $\rho = -0.7872$

Panel C: Estimation using SKEW

| | | $\bar{\sigma}_v$ | $Std(\sigma_v)$ | RMSE |
|---|---|---|---|---|
| Fixed $\theta$ | Fixed $\sigma_v$ | 0.5914 | - | 8.8953 |
| Fixed $\theta$ | Floating $\sigma_v$ | 0.6101 | 0.0905 | 8.2364 |
| Floating $\theta$ | Fixed $\sigma_v$ | 0.8829 | - | 7.9561 |
| Floating $\theta$ | Floating $\sigma_v$ | 0.9440 | 0.1988 | 6.2783 |

FIGURE 2.1: The Mathematica code to compute the integrals in the variance and third cumulant and resulted exact formulas after execution.

In[1]:=

$$\mathbf{Etvu} = \theta + (\mathbf{vt} - \theta)\, e^{-\kappa\,(u-t)}$$

$$\mathbf{EtX2T} = \int_t^T \mathbf{Etvu}\, du$$

$$\mathbf{EtXTYT} = \rho\,\sigma\mathbf{v} \int_t^T \frac{1 - e^{-\kappa\,(T-u)}}{\kappa}\, \mathbf{Etvu}\, du$$

$$\mathbf{EtY2T} = \sigma\mathbf{v}^2 \int_t^T \frac{\left(1 - e^{-\kappa\,(T-u)}\right)^2}{\kappa^2}\, \mathbf{Etvu}\, du$$

$$\mathbf{EtX3T} = 3\,\rho\,\sigma\mathbf{v} \int_t^T \frac{1 - e^{-\kappa\,(T-u)}}{\kappa}\, \mathbf{Etvu}\, du$$

$$\mathbf{EtX2TYT} = \sigma\mathbf{v}^2 \int_t^T \left( \frac{\left(1 - e^{-\kappa\,(T-u)}\right)^2}{\kappa^2} + 2\,\rho^2\, \frac{1 - e^{-\kappa\,(T-u)} - \kappa\,(T-u)\, e^{-\kappa\,(T-u)}}{\kappa^2} \right) \mathbf{Etvu}\, du$$

$$\mathbf{EtXTY2T} = \rho\,\sigma\mathbf{v}^3 \int_t^T \left( 2\, \frac{1 - e^{-\kappa\,(T-u)} - \kappa\,(T-u)\, e^{-\kappa\,(T-u)}}{\kappa^2}\, \frac{1 - e^{-\kappa\,(T-u)}}{\kappa} + \frac{1 - e^{-2\,\kappa\,(T-u)} - 2\,\kappa\,(T-u)\, e^{-\kappa\,(T-u)}}{\kappa^3} \right) \mathbf{Etvu}\, du$$

$$\mathbf{EtY3T} = 3\,\sigma\mathbf{v}^4 \int_t^T \left( \frac{1 - e^{-2\,\kappa\,(T-u)} - 2\,\kappa\,(T-u)\, e^{-\kappa\,(T-u)}}{\kappa^3}\, \frac{1 - e^{-\kappa\,(T-u)}}{\kappa} \right) \mathbf{Etvu}\, du$$

Out[1]=

$$e^{-(-t+u)\,\kappa}\,(\mathbf{vt} - \theta) + \theta$$

Out[2]=

$$\frac{\mathbf{vt} - \theta + e^{(t-T)\,\kappa}\,(-\mathbf{vt} + \theta) - t\,\theta\,\kappa + T\,\theta\,\kappa}{\kappa}$$

Out[3]=

$$\frac{e^{-T\,\kappa}\left(e^{t\,\kappa}\,(-\mathbf{vt} + 2\,\theta + (t - T)\,(\mathbf{vt} - \theta)\,\kappa) + e^{T\,\kappa}\,(\mathbf{vt} + \theta\,(-2 - t\,\kappa + T\,\kappa))\right)\rho\,\sigma\mathbf{v}}{\kappa^2}$$

Out[4]=

$$\frac{e^{-2\,T\,\kappa}\left(e^{2\,t\,\kappa}\,(-2\,\mathbf{vt} + \theta) + 4\,e^{(t+T)\,\kappa}\,(\theta + (t - T)\,(\mathbf{vt} - \theta)\,\kappa) + e^{2\,T\,\kappa}\,(2\,\mathbf{vt} + \theta\,(-5 - 2\,t\,\kappa + 2\,T\,\kappa))\right)\sigma\mathbf{v}^2}{2\,\kappa^3}$$

Out[5]=

$$\frac{3\,e^{-T\,\kappa}\left(e^{t\,\kappa}\,(-\mathbf{vt} + 2\,\theta + (t - T)\,(\mathbf{vt} - \theta)\,\kappa) + e^{T\,\kappa}\,(\mathbf{vt} + \theta\,(-2 - t\,\kappa + T\,\kappa))\right)\rho\,\sigma\mathbf{v}}{\kappa^2}$$

Out[6]=

$$\frac{1}{2\,\kappa^3}\Big(-e^{-(t+2\,T)\,\kappa}\left(2\,e^{3\,t\,\kappa}\,(\mathbf{vt} - \theta) + e^{3\,t\,\kappa}\,\theta - 2\,e^{(t+2\,T)\,\kappa}\,(\mathbf{vt} - \theta)\left(1 + 2\,\rho^2\right) + 2\,e^{(t+2\,T)\,\kappa}\,t\,\theta\,\kappa\left(1 + 2\,\rho^2\right) + 2\,e^{(2\,t+T)\,\kappa}\,t\,(\mathbf{vt} - \theta)\,\kappa\left(-2 + (-2 + t\,\kappa - 2\,T\,\kappa)\,\rho^2\right) + 4\,e^{(2\,t+T)\,\kappa}\,\theta\left(-1 + (-2 + t\,\kappa - T\,\kappa)\,\rho^2\right)\right) + e^{-T\,\kappa}\left(-2\,e^{t\,\kappa}\,(\mathbf{vt} - \theta)\left(2\,\rho^2 + T^2\,\kappa^2\,\rho^2 + 2\,T\,\kappa\left(1 + \rho^2\right)\right) + e^{T\,\kappa}\,\theta\left(-3 - 8\,\rho^2 + 2\,T\left(\kappa + 2\,\kappa\,\rho^2\right)\right)\right)\Big)\sigma\mathbf{v}^2$$

Out[7]=

$$\frac{1}{\kappa^4}\,e^{-2\,T\,\kappa}\Big(e^{2\,t\,\kappa}\,(-3\,\mathbf{vt} + 2\,\theta + (t - T)\,(2\,\mathbf{vt} - \theta)\,\kappa) + 2\,e^{(t+T)\,\kappa}\,(-2 + t\,\kappa - T\,\kappa)\,(-2\,\theta + (t - T)\,(-\mathbf{vt} + \theta)\,\kappa) + e^{2\,T\,\kappa}\,(3\,\mathbf{vt} + \theta\,(-10 - 3\,t\,\kappa + 3\,T\,\kappa))\Big)\rho\,\sigma\mathbf{v}^3$$

Out[8]=

$$\frac{1}{2\,\kappa^5}\Big(e^{-T\,\kappa}\left(2\,e^{T\,\kappa}\,\theta\,(-8 + 3\,T\,\kappa) - 3\,e^{t\,\kappa}\,(\mathbf{vt} - \theta)\left(-1 + 2\,T\,\kappa + 2\,T^2\,\kappa^2\right)\right) + e^{-(t+3\,T)\,\kappa}\left(e^{4\,t\,\kappa}\,(-3\,\mathbf{vt} + \theta) + 6\,e^{(t+3\,T)\,\kappa}\,(\mathbf{vt} - \theta\,(1 + t\,\kappa)) + 6\,e^{(3\,t+T)\,\kappa}\,(\theta - t\,\theta\,\kappa + T\,\theta\,\kappa + \mathbf{vt}\,(-1 + 2\,t\,\kappa - 2\,T\,\kappa)) + 6\,e^{2\,(t+T)\,\kappa}\left(t\,\mathbf{vt}\,\kappa\,(1 - t\,\kappa + 2\,T\,\kappa) + \theta\left(3 + 2\,T\,\kappa + t^2\,\kappa^2 - t\,\kappa\,(3 + 2\,T\,\kappa)\right)\right)\right)\Big)\sigma\mathbf{v}^4$$

FIGURE 2.2: The Term Structure of Skewness Implied in the Heston (1993) Model for a Given Set of Parameters

The term structure of skewness implied in the Heston (1993) model for a set of parameters: $\rho = -0.25$, $\kappa = 5$, $\theta = 0.1$, $a = v_t/\theta = 1.25$ and $\sigma_v = 0.4$. The solid line is from our exact formula. The dashed line is from Das and Sundaram's (1999) formula. The asymptotic lines for small and large $\tau$ are also shown on the graph for $\tau \in (0, 0.2)$ and $(10, 20)$ respectively.

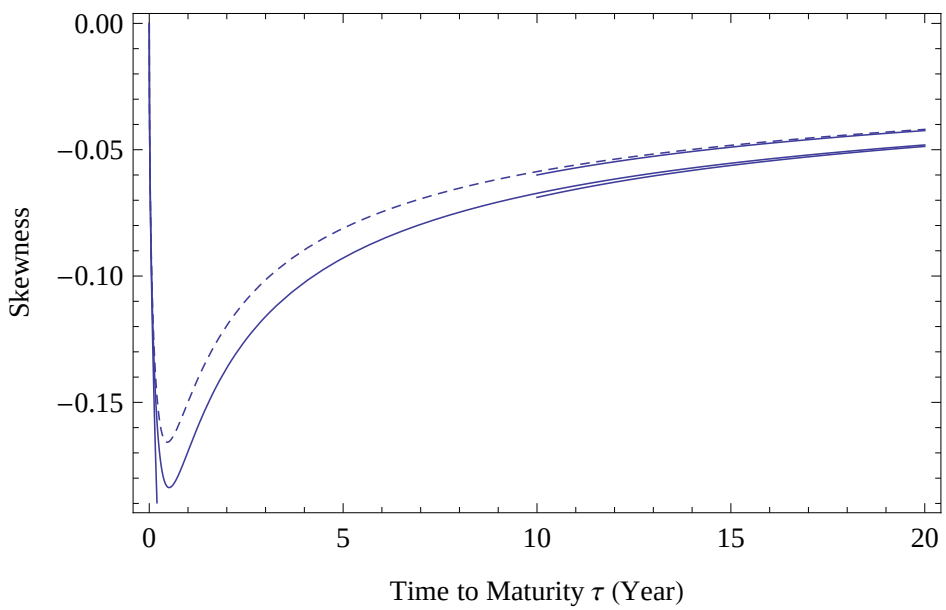


FIGURE 2.3: The Fitting Performance on a Randomly-chosen Day

The fitting performance on a randomly-chosen day. The upper and lower graphs show the fitting performance of VIX and SKEW on 22 June 2015, respectively. The lines are model implied theoretical values and the dots represent market data. The solid line in the upper graph shows the VIX fitting performance of both the Heston (1993) model and the Das-Sundaram (1999) model with the optimal parameters $v_t = 0.0143$, $\theta = 0.0531$, $\kappa = 1.45$, and a root mean squared error (RMSE) of 0.3603. The solid line in the lower graph shows the SKEW fitting performance of the Heston model with the optimal parameters $\sigma_v = 0.7325$, $\rho = -0.7176$ and a RMSE of 3.4847. The dashed line in the lower graph represents the SKEW fitting performance of the DS model with the optimal parameters $\rho\sigma_v = -0.5556$ and a RMSE of 3.5566.

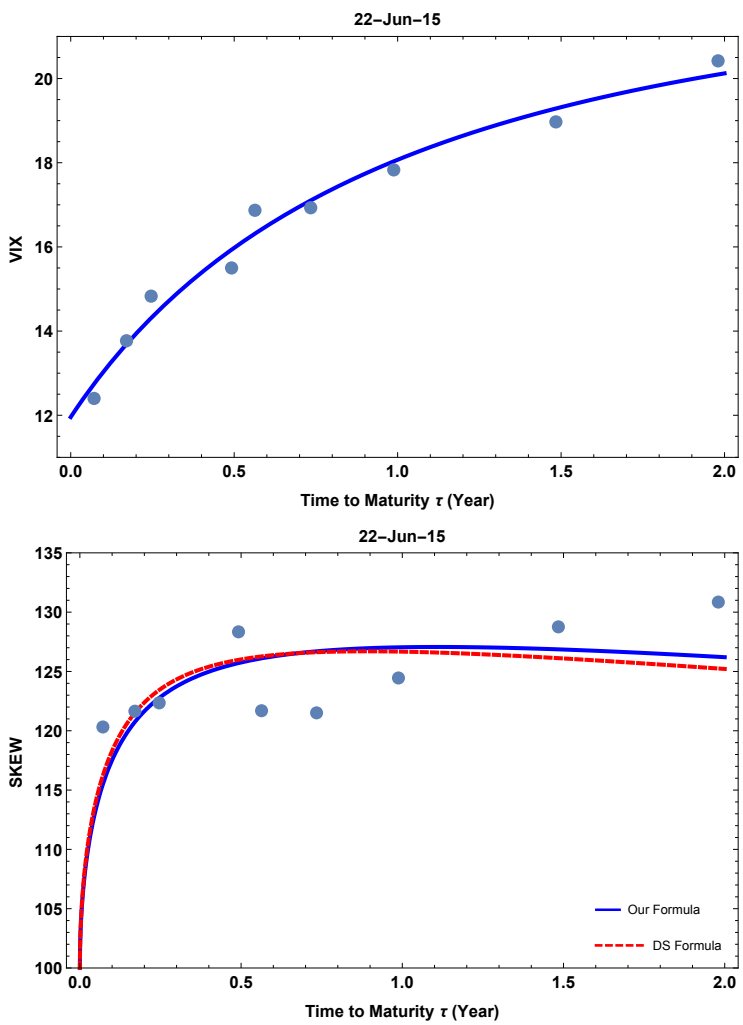


FIGURE 2.4: The Estimation of the Heston (1993) Model

The estimation of the Heston (1993) model for the sample period from 24 November 2010 to 30 June 2015. The estimated risk-neutral mean-reverting speed is $\kappa = 0.28$. The upper graph shows the daily estimated instantaneous variance, $v_t$, and fixed long-term mean level, $\theta = 0.1651$. The correlation coefficient that is computed from the time series of SPX and VIX is $\rho = -0.7872$. The lower graph shows the fixed volatility of volatility, $\sigma_v = 0.5914$, which is estimated over the whole sample period and floating one that is fitted daily.

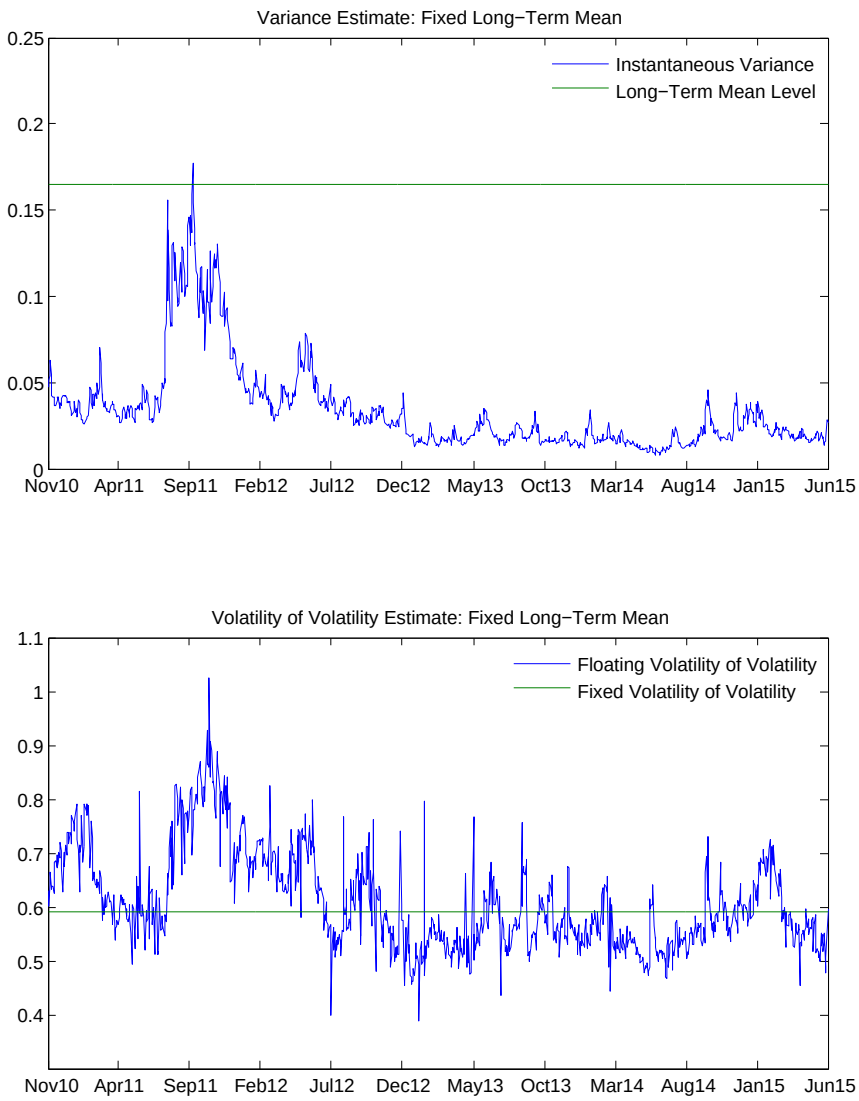
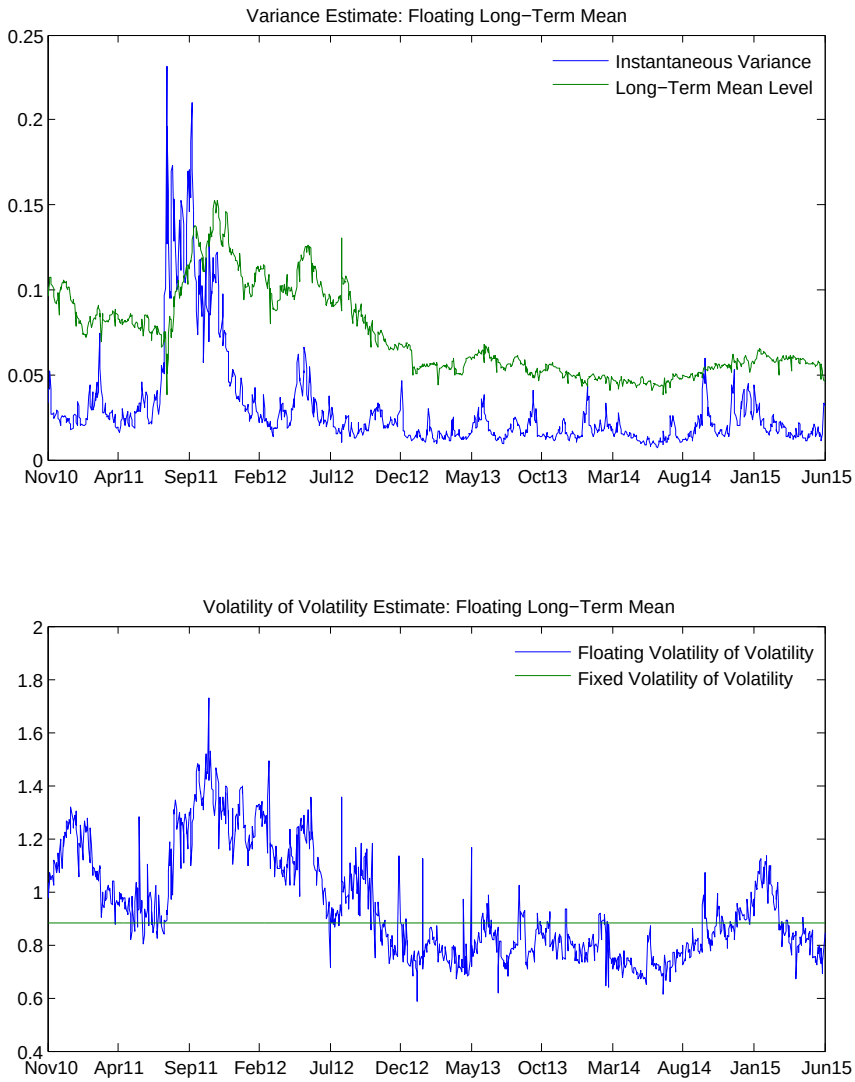


FIGURE 2.5: The Estimation of an Extended Heston (1993) Model with Floating Long-term Mean

The estimation of an extended Heston (1993) model with floating long-term mean for the sample period from 24 November 2010 to 30 June 2015. The estimated risk-neutral mean-reverting speed is $\kappa = 1.91$. The upper graph shows the daily estimated instantaneous variance, $v_t$, and instantaneous long-term mean level, $\theta_t$. The correlation coefficient that is computed from the time series of SPX and VIX is $\rho = -0.7872$. The lower graph shows the fixed volatility of volatility, $\sigma_v = 0.8829$, which is estimated over the whole sample period and floating one that is fitted daily.



# Chapter 3

# The CBOE SKEW

This chapter is a joint work with Jin E. Zhang. Its earlier version was presented at the Quantitative Methods in Finance 2014 Conference, 17–20 December 2014, UTS, Sydney, Australia; the 2015 Derivative Markets Conference, 13–14 August 2015, AUT, Auckland, New Zealand; the 20th New Zealand Finance Colloquium, 11–12 February 2016, University of Otago, Queenstown, New Zealand (by my co-author); and the 2016 Auckland Finance Meeting, 16–18 December 2016, AUT, Auckland, New Zealand (by my co-author). This chapter was previously circulated as "A Theory of the CBOE SKEW".

## 3.1 Introduction

The CBOE SKEW is an index launched by CBOE in February 2011.[1] Its term structure tracks the risk-neutral skewness of the SPX index for different maturities. The observable public information of the SKEW could be useful in forecasting future stock returns, even market crashes. However, the literature on SKEW is sparse, and a theory of the CBOE SKEW has not been developed. This chapter fills the gap by establishing a theory for the SKEW index by modelling SPX using a jump-diffusion process with stochastic volatility and stochastic jump intensity. We propose a new sequential procedure to estimate model parameters and latent variables explicitly by fitting the market data with the model implied term structure of the CBOE volatility index (VIX) and SKEW.

[1] The CBOE SKEW index is referred to as SKEW hereafter.

The CBOE has been dedicated to developing new markets for higher-moment trading, including volatility (second moment), over the last 20 years. First introduced in 1993, VIX serves as the premier benchmark for U.S. stock market volatility. VIX was revised in 2003 by averaging the weighted prices of out-of-the-money SPX puts and calls over a wide range of strike prices, and enhanced by using the weekly options in 2014. Moreover, CBOE made it possible for market participants to directly invest in the VIX index by launching the VIX futures in 2004 and the VIX options in 2006. These VIX derivatives are among the most actively traded products with total trading volumes of 51.6 million contracts and 144.4 million contracts, respectively, in 2015. In addition to VIX, CBOE launched SKEW in 2011, which is a global, 30-day forward-looking, strike-independent measure of the slope of the implied volatility and reflects the asymmetry of the risk-neutral distribution of the SPX returns. SKEW is a complementary risk indicator to VIX and is calculated from SPX option prices with the methodology similar to VIX. It is the first observable information regarding the skewness risk across different maturities, and provides the foundation for the creation of SKEW derivatives and potential skewness trading.

The importance of the skewness risk in explaining the cross-sectional variation of stock returns has been tested by a large strand in the empirical asset pricing literature; see, among others, for coskewness (Kraus and Litzenberger, 1976; Harvey and Siddique, 2000), for idiosyncratic skewness (Boyer, Mitton and Vorkink, 2010), for implied market skewness (Chang, Christoffersen and Jacobs, 2013), for ex-ante risk-neutral skewness (Conrad, Dittmar and Ghysels, 2013), and for realized skewness (Amaya, Christoffersen, Jacobs and Vasquez, 2015).[2] However, these skewness measures are constructed by academics either from historical returns or option prices. The public availability and observability of SKEW give rise to an easily accessible measure of aggregate skewness. This unified public measure of skewness points towards a standardized risk gauge. Nevertheless, the literature on SKEW is sparse. Faff and Liu (2014) propose an alternative measure of market asymmetry against the SKEW index. Wang and Daigler (2014) find evidence supporting the bidirectional information flow between SPX and VIX options markets by analyzing the SKEW

[2]The determinants of cross-sectional skewness have also been tested, such as the trading volume and past returns (see Chen, Hong and Stein, 2001), the systematic and firm-specific factors (see Dennis and Mayhew, 2002), the investor sentiment (see Han, 2008), and the heterogeneous beliefs (see Friesen, Zhang and Zorn, 2012).

and VIX indices for SPX and VIX. Their articles focus on the informational aspect of SKEW. The intrinsic nature of SKEW has not yet been explored. In this chapter, we build a theory of the CBOE SKEW and look at the empirical performance of different option-pricing models.

Zhang et al. (2017) point out the inability of the Heston (1993) model to capture short-term skewness and stochastic long-term variance. To further calibrate more realistic and sophisticated option-pricing models, we provide analytical formulas for the CBOE SKEW in various affine jump-diffusion models (see Duffie, Pan and Singleton, 2000), and obtain the latent diffusive volatility and jump-intensity variables in a five-factor model using the term structure data of the CBOE VIX and SKEW. The SKEW and VIX theoretical calculation formulas provide a channel to estimate the unobserved diffusive variance, jump intensity as well as other fixed parameters, which are crucial factors for option pricing and hedging.

The option-pricing literature is well-established. However, the model estimation methodologies are not unified, and are challenging when the state variables are unobservable. The estimation approaches include the simulated method of moments (SMM) used by Bakshi, Cao and Chen (2000), the efficient method of moments (EMM) adopted by Chernov and Ghysels (2000) and Anderson, Benzoni and Lund (2002), the implied-state generalized method of moments (IS-GMM) proposed by Pan (2002), and the Markov Chain Monte Carlo (MCMC) method applied by Eraker, Johannes and Polson (2003) and Eraker (2004). These estimators impose severe computational burdens, and are not easy to implement. Moreover, the types of data used for estimation are different. For instance, Bakshi, Cao and Chen (2000) use options data, Chernov, Ghysels (2000), Pan (2002) and Eraker (2004) use joint returns and options data, whereas Andersen, Benzoni and Lund (2002) and Eraker, Johannes and Polson (2003) use returns data only. In contrast, given the public information of VIX and SKEW, we propose an easily implemented sequential estimation procedure, which consistently obtains the model parameters and the latent variables across the second and third cumulants (or equivalently central moments) of the risk-neutral distribution of the SPX returns. This chapter is the first to uncover the diffusive variance and jump-intensity processes explicitly using combined VIX and SKEW data, which are carefully designed by CBOE using Bakshi, Kapadia and Madan's (2003) model-free methodology and reflect the most important features of

the SPX returns' distribution.

In this chapter, we find that the Merton-extended Heston model improves the fitting performance of the VIX term structure by 60% compared with the Merton-Heston model. The Merton-extended Heston (Merton-Heston) model outperforms the Heston model by 20% (14%) when fitting the SKEW term structure. The performance of the five-factor (three-factor) model in fitting the SKEW term structure exceeds that of the Merton-extended Heston (Merton-Heston) model by 15% (2%).[3] Our estimation results suggest that the five-factor model is superior when modelling the term structure of the CBOE VIX and SKEW.

The remainder of this chapter proceeds as follows. Section 3.2 introduces the CBOE SKEW index. Section 3.3 develops a theory of the CBOE SKEW. Section 3.4 discusses the data. Section 3.5 presents the estimation procedure and compares the empirical performance. Section 3.6 provides concluding remarks. All proofs are in the Appendix.

## 3.2 Definition of the CBOE SKEW

The CBOE SKEW is an index launched by CBOE on February 23, 2011. Its term structure tracks the risk-neutral skewness of the SPX returns for different maturities. The SKEW is computed from all of the out-of-the-money (OTM) SPX option prices by using Bakshi, Kapadia and Madan's (2003) methodology.

At time $t$, the SKEW is defined as

$$SKEW_t = 100 - 10 \times Sk_t, \tag{3.1}$$

where $Sk_t$ is the risk-neutral skewness given by

$$Sk_t = E_t^Q \left[ \frac{(R_t^T - \mu)^3}{\sigma^3} \right]; \tag{3.2}$$

[3]The five factors are referred to as $\ln S_t$, $v_t$, $\lambda_t$, $\theta_t^v$ and $\theta_t^\lambda$, where the notations will be defined later in this chapter. The long-term mean levels are constant in the three-factor model, and the three factors are referred to as $\ln S_t$, $v_t$ and $\lambda_t$.

$R_t^T$ is the continuously compounded return of SPX at time $T$, denoted as $S_T$, against the current forward price with maturity $T$, $F_t^T$; $\mu$ is the expected return and $\sigma^2$ is the variance in the risk-neutral measure $Q$:

$$R_t^T = \ln \frac{S_T}{F_t^T}, \qquad \mu = E_t^Q(R_t^T), \qquad \sigma^2 = E_t^Q[(R_t^T - \mu)^2]. \tag{3.3}$$

Expanding Equation (3.2) gives

$$Sk_t = \frac{E_t^Q[(R_t^T)^3] - 3E_t^Q(R_t^T)E_t^Q[(R_t^T)^2] + 2[E_t^Q(R_t^T)]^3}{\left(E_t^Q[(R_t^T)^2] - [E_t^Q(R_t^T)]^2\right)^{3/2}}. \tag{3.4}$$

Following Bakshi, Kapadia and Madan (2003), the CBOE evaluates the first three moments of the continuously compounded return $E_t^Q(R_t^T)$, $E_t^Q[(R_t^T)^2]$ and $E_t^Q[(R_t^T)^3]$ in the risk-neutral measure by using the current prices of European options with all available strikes as follows:

$$E_t^Q(R_t^T) = -e^{r\tau} \sum_i \frac{1}{K_i^2} Q(K_i)\Delta K_i + \ln \frac{K_0}{F_t^T} + \frac{F_t^T}{K_0} - 1, \tag{3.5}$$

$$E_t^Q[(R_t^T)^2] = e^{r\tau} \sum_i \frac{2 - 2\ln \frac{K_i}{F_t^T}}{K_i^2} Q(K_i)\Delta K_i + \ln^2 \frac{K_0}{F_t^T} + 2\ln \frac{K_0}{F_t^T}\left(\frac{F_t^T}{K_0} - 1\right), \tag{3.6}$$

$$E_t^Q[(R_t^T)^3] = e^{r\tau} \sum_i \frac{6\ln \frac{K_i}{F_t^T} - 3\ln^2 \frac{K_i}{F_t^T}}{K_i^2} Q(K_i)\Delta K_i + \ln^3 \frac{K_0}{F_t^T} + 3\ln^2 \frac{K_0}{F_t^T}\left(\frac{F_t^T}{K_0} - 1\right), \tag{3.7}$$

where $F_t^T$ is the forward index level derived from SPX option prices by using put-call parity; $K_0$ is the first listed price below $F_t^T$; $K_i$ is the strike price of the $i$th OTM option (a call if $K_i > K_0$ and a put if $K_i < K_0$); $\Delta K_i$ is half the difference between strikes on either side of $K_i$, that is, $\Delta K_i = \frac{1}{2}(K_{i+1} - K_{i-1})$, and for minimum (maximum) strike, $\Delta K_i$ is simply the distance to the next strike above (below); $r$ is the risk-free interest rate; $Q(K_i)$ is the midpoint of the bid-ask spread for each option with strike $K_i$; and $\tau$ is the time to expiration as a fraction of a year. The reasoning behind Equations (3.5), (3.6) and (3.7) are included in Appendix 3.7.1.

The SKEW index corresponds to the returns with 30 days to maturity, that is, $\tau = \tau_0 \equiv 30/365$. In general, 30-day options are not available. The current 30-day skewness $Sk_t$ is derived by inter- or extrapolation from the current risk-neutral skewness at adjacent expirations, $Sk_t^{\text{near}}$ and $Sk_t^{\text{next}}$ as follows:

$$Sk_t = \omega Sk_{t,\text{near}} + (1-\omega) Sk_{t,\text{next}}, \tag{3.8}$$

where $\omega$ is a weight determined by

$$\omega = \frac{\tau_{\text{next}} - \tau_0}{\tau_{\text{next}} - \tau_{\text{near}}},$$

and $\tau_{\text{near}}$ and $\tau_{\text{next}}$ are the times to expiration (up to minute) of the near- and next-term options, respectively. The near- and next-term options are usually the first and second SPX contract months. "Near-term" options must have at least one week to expiration in order to minimize possible close-to-expiration pricing anomalies. For near-term options with less than one week to expiration, the data roll to the second SPX contract month. "Next-term" is the next contract month following near-term. While calculating time to expiration, the SPX options are deemed to expire at the open of trading on the SPX settlement day, that is, the third Friday of the month.

## 3.3 Theory

Option-pricing models have been developed by using different kinds of stochastic processes for the underlying stock, such as jump-diffusion with stochastic volatility and stochastic jump intensity. The purpose of making volatility and jump intensity stochastic is to capture the time-varying second and third moments of stock returns. Traditionally, these models are usually estimated by using some numerical approaches, with volatility and jump intensity being latent variables. These numerical estimation procedures are often highly technical and very time-consuming.

With the observable information of two term structures of VIX and SKEW from the CBOE, it is now possible to estimate the risk-neutral underlying process explicitly. The volatility and jump intensity processes are not latent any more. In fact, in this chapter we will make them semi-observable. In order to achieve this goal, we

need some theoretical results on the term structures of VIX and risk-neutral skewness implied in different models.

To intuit the result, we begin with the simplest case, that is, the Black-Scholes (1973) model.

**Proposition 3.1** *In the Black-Scholes (1973) model, the risk-neutral underlying stock is modelled by*

$$\frac{dS_t}{S_t} = rdt + \sigma dB_t,$$

*where $B_t$ is a standard Brownian motion, $r$ is the risk-free rate and $\sigma$ is volatility. The model-implied squared VIX and skewness at time $t$ for time to maturity $\tau$ are as follows:*

$$VIX^2_{t,\tau} = \sigma^2, \qquad Sk_{t,\tau} = 0. \tag{3.9}$$

This is a trivial case. The return is normally distributed; hence, the volatility term structure is flat and the skewness is zero.

In order to create skewness, we need to include jumps, e.g., the Poisson process, into the model. Merton's (1976) jump-diffusion model is a pioneer along this direction.

**Proposition 3.2** *In the Merton (1976) model, the risk-neutral underlying stock is modelled by*

$$\frac{dS_t}{S_t} = rdt + \sigma dB_t + (e^x - 1)dN_t - \lambda(e^x - 1)dt,$$

*where $N_t$ is a Poisson process with constant jump size $x$ and jump intensity $\lambda$. The model-implied squared VIX and skewness at time $t$ for time to maturity $\tau$ are as follows:*

$$VIX^2_{t,\tau} = \sigma^2 + 2\lambda(e^x - 1 - x), \qquad Sk_{t,\tau} = \frac{\lambda x^3}{(\sigma^2 + \lambda x^2)^{3/2}}\,\tau^{-1/2}. \tag{3.10}$$

*Proof*. See Appendix 3.7.2.

**Remark 3.2.1.** Here we present a result with a constant jump size in order to make the formulas of the VIX and SKEW term structures simple and intuitive.[4] It can be

[4]The assumption of constant jump size is also used in the literature, see, among others, Liu and Pan (2003) and Branger, Kraft and Meinerding (2016)

extended to an arbitrary distribution without much difficulty; see e.g., Zhang, Zhao and Chang (2012).

**Remark 3.2.2.** The VIX term structure is again flat with a squared VIX being $\sigma^2 + 2\lambda(e^x - 1 - x)$, which is different from the annualized term variance (variance swap rate) $\sigma^2 + \lambda x^2$. The difference is due to the CBOE definition of VIX. It is small for small $x$; see Luo and Zhang (2012) for a detailed discussion.

**Remark 3.2.3.** Noticing that $\sqrt{\tau}\ Sk_{t,\tau}$ is a constant in this model, we obtain a simple criterion as follows: *If $\sqrt{\tau}\ Sk_{t,\tau}$ is not a constant, then Merton's (1976) jump-diffusion model does not apply in the options market.*

The returns in the Merton model are independent and identically distributed (i.i.d.). Furthermore, for any i.i.d. variable, its variance and third cumulant are additive and $Sk_{t,\tau}$ is proportional to $\frac{1}{\sqrt{\tau}}$, resulting in infinite instantaneous skewness and zero long term skewness. Note that $\sqrt{\tau}\ Sk_{t,\tau}$ is regular for i.i.d. returns; we introduce a new concept of the regularized skewness.

**Definition** The regularized skewness, $RSk_{t,\tau}$, at time $t$, is given by

$$RSk_{t,\tau} = \sqrt{\tau}\ Sk_{t,\tau} = \sqrt{\tau} E_t^Q \left[ \frac{(R_{t,\tau} - \mu(R_{t,\tau}))^3}{\sigma^3(R_{t,\tau})} \right], \tag{3.11}$$

where $R_{t,\tau}$ denotes the continuously compounded return over the period $[t, t+\tau]$, $\mu(R_{t,\tau})$ and $\sigma^2(R_{t,\tau})$ denote its mean and variance, respectively.

In addition to jumps, skewness can also be created by using stochastic volatility with leverage effect, that is, correlation between stock returns and volatility. Along this direction, Heston's (1993) model has become a standard platform partially because of its analytical tractability due to an affine structure. Das and Sundaram (1999) describe analytically the skewness implied in a stochastic volatility model which is similar to Heston's (1993) setup; unfortunately, their closed-form formula does not apply in the Heston model. Zhao, Zhang and Chang (2013) provide a partial result for a special case of zero mean-reverting speed, that is, $\kappa = 0$. A full explicit formula was not available until the recent work of Zhang et al. (2017).

**Proposition 3.3** *(Zhang et al., 2017) In the Heston (1993) model, the risk-neutral underlying stock is modelled by*

$$\frac{dS_t}{S_t} = rdt + \sqrt{v_t}dB_t^S,$$
$$dv_t = \kappa(\theta - v_t)dt + \sigma_v\sqrt{v_t}dB_t^v,$$

*where two standard Brownian motions, $B_t^S$ and $B_t^v$, are correlated with a constant coefficient $\rho$. The model-implied squared VIX and skewness at time $t$ for time to maturity $\tau$ are as follows:*

$$VIX_{t,\tau}^2 = (1-\omega)\theta + \omega v_t, \qquad \omega = \frac{1-e^{-\kappa\tau}}{\kappa\tau}, \tag{3.12}$$

$$Sk_{t,\tau} = \frac{TC^H}{(Var^H)^{3/2}}, \tag{3.13}$$

*where the term variance and third cumulant are given by*

$$Var^H = E_t(X_T^2) - E_t(X_TY_T) + \frac{1}{4}E_t(Y_T^2), \tag{3.14}$$

$$TC^H = E_t(X_T^3) - \frac{3}{2}E_t(X_T^2Y_T) + \frac{3}{4}E_t(X_TY_T^2) - \frac{1}{8}E_t(Y_T^3). \tag{3.15}$$

*The integrated return uncertainty, $X_T$, and integrated instantaneous variance uncertainty, $Y_T$, are defined by*

$$X_T \equiv \int_t^T \sqrt{v_u}dB_u^S, \qquad Y_T \equiv \int_t^T [v_u - E_t(v_u)]du = \sigma_v\int_t^T \frac{1-e^{-\kappa(T-u)}}{\kappa}\sqrt{v_u}dB_u^v.$$

*The variance and covariance of $X_T$ and $Y_T$ are given by*

$$E_t(X_T^2) = \int_t^T E_t(v_u)du, \tag{3.16}$$

$$E_t(X_TY_T) = \rho\sigma_v\int_t^T A_1E_t(v_u)du, \qquad A_1 = \frac{1-e^{-\kappa\tau^*}}{\kappa}, \qquad \tau^* = T-u, \tag{3.17}$$

$$E_t(Y_T^2) = \sigma_v^2\int_t^T A_1^2E_t(v_u)du. \tag{3.18}$$

*The third and co-third cumulants of $X_T$ and $Y_T$ are given by*

$$E_t(X_T^3) = 3\rho\sigma_v \int_t^T A_1 \; E_t(v_u)du, \tag{3.19}$$

$$E_t(X_T^2 Y_T) = \sigma_v^2 \int_t^T A_2 \; E_t(v_u)du,$$

$$A_2 = \left(\frac{1-e^{-\kappa\tau^*}}{\kappa}\right)^2 + 2\rho^2 \frac{1-e^{-\kappa\tau^*} - \kappa\tau^* e^{-\kappa\tau^*}}{\kappa^2}, \tag{3.20}$$

$$E_t(X_T Y_T^2) = \rho\sigma_v^3 \int_t^T A_3 \; E_t(v_u)du,$$

$$A_3 = 2\frac{1-e^{-\kappa\tau^*} - \kappa\tau^* e^{-\kappa\tau^*}}{\kappa^2}\frac{1-e^{-\kappa\tau^*}}{\kappa} + \frac{1-e^{-2\kappa\tau^*} - 2\kappa\tau^* e^{-\kappa\tau^*}}{\kappa^3}, \tag{3.21}$$

$$E_t(Y_T^3) = 3\sigma_v^4 \int_t^T A_4 \; E_t(v_u)du,$$

$$A_4 = \frac{1-e^{-2\kappa\tau^*} - 2\kappa\tau^* e^{-\kappa\tau^*}}{\kappa^3}\frac{1-e^{-\kappa\tau^*}}{\kappa}, \tag{3.22}$$

*and $E_t(v_u) = \theta + (v_t - \theta)e^{-\kappa(u-t)}$ is the expected instantaneous variance.*

*Proof*. See Zhang et al. (2017).

**Remark 3.3.1.** Through asymptotic analysis, Zhang et al. (2017) shows that for $\kappa > 0$, if $\tau$ is small, the skewness is given by

$$Sk_{t,\tau} = \frac{3}{2}\rho\frac{\sigma_v}{\sqrt{v_t}}\sqrt{\tau} + o(\sqrt{\tau}).$$

If $\tau$ is large, then the skewness is given by

$$Sk_{t,\tau} = \frac{3\rho\frac{\sigma_v}{\kappa} - \frac{3}{2}\frac{\sigma_v^2}{\kappa^2}}{\sqrt{1-\rho\frac{\sigma_v}{\kappa} + \frac{1}{4}\frac{\sigma_v^2}{\kappa^2}}}\frac{1}{\sqrt{\theta\tau}} + o\left(\frac{1}{\sqrt{\tau}}\right).$$

Hence, the regularized skewness $\sqrt{\tau}\; Sk_{t,\tau}$ behaves linearly with $\tau$ for small $\tau$, and approaches to a constant for large $\tau$.

When modelling the VIX term structure, Luo and Zhang (2012) observe that both short and long ends of the term structure are time-varying. Hence, it is necessary to

include a new factor of stochastic long-term mean into the standard Heston model in order to enhance its performance in fitting the term structure of variance. Egloff, Leippold and Wu (2010) show the high persistence (low mean reverting speed) of the long-term mean level of variance, to retain the martingale property of the stochastic long-term mean, we set its mean-reverting speed to be zero.[5] The impact of the new factor on skewness is presented in the next proposition.

**Proposition 3.4** *In an extended Heston model with stochastic long term mean, the risk-neutral underlying stock is modelled by*

$$\begin{aligned}
\frac{dS_t}{S_t} &= rdt + \sqrt{v_t}dB_t^S \\
dv_t &= \kappa(\theta_t - v_t)dt + \sigma_v\sqrt{v_t}dB_t^v. \\
d\theta_t &= \sigma_\theta\sqrt{\theta_t}dB_t^\theta,
\end{aligned}$$

*where the new standard Brownian motion, $B_t^\theta$, is independent of $B_t^S$ and $B_t^v$, and $\sigma_\theta$ is the volatility of the long-term mean level of variance. The model-implied squared VIX and skewness at time t for time to maturity $\tau$ are as follows*

$$\begin{aligned}
VIX_{t,\tau}^2 &= (1-\omega)\theta_t + \omega v_t, \qquad \omega = \frac{1-e^{-\kappa\tau}}{\kappa\tau}, \\
Sk_{t,\tau} &= \frac{TC^H + TC^{HM}}{[Var^H + Var^{HM}]^{3/2}},
\end{aligned} \tag{3.23}$$

[5]We assume that the stochastic long-term mean follows a square root process with zero drift. We regard this setup as a limiting case of the CIR model (see Cox, Ingersoll and Ross, 1985) when the mean-reverting speed approaches zero. Thus, its time-$t$ conditional mean goes to the current long-term mean level $\theta_t$, and its conditional variance to $\sigma_\theta^2\theta_t(T-t)$, where $T$ is a future time. Note that the conditional variance is infinite for an infinitely large future time, but as we show later in the data, the conditional variance is finite, for the longest time to maturity in the data is three years. The standard deviation factor $\sigma_\theta\sqrt{\theta_t}$ avoids the possibility of negative values, but a zero long-term mean is not precluded. The stochastic long-term mean $\theta_t$ follows a noncentral chi-squared distribution with zero degrees of freedom (see Siegel, 1979), which is a Poisson mixture of central chi-square distributions with even degrees of freedom. Specifically, the noncentral $\chi_0^2(\lambda)$ distribution is a mixture of central $\{\chi_{2K}^2\}$ distributions with Poisson weights $P(K=k) = \frac{(\lambda/2)^k}{k!}e^{-\lambda/2}$, $k = 0, 1, 2, \cdots$ (the central $\chi_0^2$ distribution is identically zero), where $\lambda$ is the noncentrality parameter. In this setup, the Feller condition is violated, so the long-term mean level of variance is not strictly positive and is accessible to zero.

*where the contributions of the stochastic long-term mean to the variance and the third cumulant are given by*

$$Var^{HM} = \frac{1}{4}\sigma_\theta^2\theta_t C_1, \tag{3.24}$$

$$TC^{HM} = -\frac{3}{2}\sigma_\theta^2\theta_t C_1 + \frac{3}{2}\rho\sigma_v\sigma_\theta^2\theta_t C_2 - \frac{3}{8}\sigma_v^2\sigma_\theta^2\theta_t C_3 - \frac{3}{8}\sigma_\theta^4\theta_t C_4, \tag{3.25}$$

$$C_1 = \int_t^T \frac{(e^{-\kappa(T-u)} - 1 + \kappa(T-u))^2}{\kappa^2} du, \tag{3.26}$$

$$C_2 = \int_t^T \frac{1 - e^{-\kappa(T-u)}}{\kappa} \int_t^u (1 - e^{-\kappa(u-s)}) \frac{e^{-\kappa(T-s)} - 1 + \kappa(T-s)}{\kappa} dsdu, \tag{3.27}$$

$$C_3 = \int_t^T \frac{(1 - e^{-\kappa(T-u)})^2}{\kappa^2} \int_t^u (1 - e^{-\kappa(u-s)}) \frac{e^{-\kappa(T-s)} - 1 + \kappa(T-s)}{\kappa} dsdu, \tag{3.28}$$

$$C_4 = \int_t^T \frac{(e^{-\kappa(T-u)} - 1 + \kappa(T-u))^2}{\kappa^2} \int_t^u \frac{e^{-\kappa(T-s)} - 1 + \kappa(T-s)}{\kappa} dsdu, \tag{3.29}$$

*and the term variance, $Var^H$, and third cumulant, $TC^H$, of the Heston model are given by Proposition 3.3.*

*Proof*. See Appendix 3.7.3.

**Remark 3.4.1.** In the extended Heston model, for $\kappa > 0$, if $\tau$ is small, the asymptotic skewness is the same as that in the Heston model, given by

$$Sk_{t,\tau} = \frac{3}{2}\rho\frac{\sigma_v}{\sqrt{v_t}}\sqrt{\tau} + o(\sqrt{\tau}).$$

If $\tau$ is large, then the skewness is given by

$$Sk_{t,\tau} = -\frac{3\sqrt{3}}{5}\frac{\sigma_\theta}{\sqrt{\theta_t}}\sqrt{\tau} + o(\sqrt{\tau}).$$

Hence the regularized skewness $\sqrt{\tau}\ Sk_{t,\tau}$ behaves linearly with $\tau$ for both small $\tau$ and large $\tau$.

With the Merton (1976) jump-diffusion model, we are not able to create a flexible

SKEW term structure because the model-implied skewness times $\sqrt{\tau}$ is a constant across different $\tau$. With the Heston (1993) model, we are not able to produce a large short-term skewness because the model-implied skewness goes to zero for small $\tau$. Hence, it is necessary to combine these two models in order to create a SKEW term structure flexible enough to fit market data. The result of a hybrid Merton-Heston model is presented in the next proposition.

**Proposition 3.5** *In a hybrid Merton-Heston model, that is, a jump-diffusion model with stochastic volatility, the risk-neutral underlying stock is modelled by*

$$\begin{aligned}\frac{dS_t}{S_t} &= rdt + \sqrt{v_t}dB_t^S + (e^x - 1)dN_t - \lambda(e^x - 1)dt,\\ dv_t &= \kappa(\theta - v_t)dt + \sigma_v\sqrt{v_t}dB_t^v.\end{aligned}$$

*The model-implied squared VIX and skewness at time $t$ for time to maturity $\tau$ are as follows*

$$\begin{aligned}VIX_{t,\tau}^2 &= (1-\omega)\theta + \omega v_t + 2\lambda(e^x - 1 - x), \qquad \omega = \frac{1-e^{-\kappa\tau}}{\kappa\tau},\\ Sk_{t,\tau} &= \frac{TC^H + \lambda x^3\tau}{[Var^H + \lambda x^2\tau]^{3/2}},\end{aligned} \tag{3.30}$$

*where the term variance, $Var^H$, and third cumulant, $TC^H$, of the Heston model are given by Proposition 3.3.*

**Remark 3.5.1.** Due to the independence between the jump and diffusion processes, the term variance and third-cumulant of the hybrid Merton-Heston model is simply the sum of the contributions from each model. The short-term skewness is no longer zero due to the third cumulant contributed from jumps.

**Remark 3.5.2.** In the hybrid Merton-Heston model, for $\kappa > 0$, if $\tau$ is small, the asymptotic skewness is given by

$$Sk_{t,\tau} = \frac{\lambda x^3}{(v_t + \lambda x^2)^{3/2}}\frac{1}{\sqrt{\tau}} + c\sqrt{\tau} + o(\sqrt{\tau}),$$

where

$$c = \frac{3\rho\sigma_v v_t}{2(v_t + \lambda x^2)^{3/2}} + \frac{3[\rho\sigma_v v_t + \kappa(v_t - \theta^v)]\lambda x^3}{4(v_t + \lambda x^2)^{5/2}}.$$

If $\tau$ is large, then the skewness is given by

$$Sk_{t,\tau} = \frac{a}{b\sqrt{b}}\frac{1}{\sqrt{\tau}} + o\left(\frac{1}{\sqrt{\tau}}\right),$$

where

$$\begin{aligned} a &= \lambda x^3 + \theta\left[3\rho\frac{\sigma_v}{\kappa} - \frac{3}{2}(1+2\rho^2)\frac{\sigma_v^2}{\kappa^2} + \frac{9}{4}\rho\frac{\sigma_v^3}{\kappa^3} - \frac{3}{8}\frac{\sigma_v^4}{\kappa^4}\right], \\ b &= \lambda x^2 + \theta\left(1 - \rho\frac{\sigma_v}{\kappa} + \frac{1}{4}\frac{\sigma_v^2}{\kappa^2}\right). \end{aligned}$$

Hence, the regularized skewness $\sqrt{\tau}\, Sk_{t,\tau}$ approaches to a constant for both small $\tau$ and large $\tau$.

The time-varying feature of VIX has been picked up by stochastic instantaneous variance, $v_t$, in the continuous-time models presented before. The time-varying feature of SKEW has to be picked up by jump-related variables. The jump size is usually assumed to follow a static distribution (in particular, a constant in this chapter); hence, we have to rely on the stochastic jump intensity, $\lambda_t$, to capture the time-varying SKEW. Retaining the affine structure, we propose the following three-factor model.

**Proposition 3.6** *In a jump-diffusion model with stochastic volatility and stochastic jump intensity with the same mean-reverting speed, the risk-neutral underlying stock is modelled by*

$$\begin{aligned} \frac{dS_t}{S_t} &= rdt + \sqrt{v_t}dB_t^S + (e^x - 1)dN_t - \lambda_t(e^x - 1)dt, \\ dv_t &= \kappa(\theta^v - v_t)dt + \sigma_v\sqrt{v_t}dB_t^v, \\ d\lambda_t &= \kappa(\theta^\lambda - \lambda_t)dt + \sigma_\lambda\sqrt{\lambda_t}dB_t^\lambda, \end{aligned}$$

*where $B_t^\lambda$ is independent of $B_t^S$, $B_t^v$, $N_t$.*[6] *The model-implied squared VIX and skewness at time $t$ for time to maturity $\tau$ are as follows:*

$$\begin{aligned} VIX_{t,\tau}^2 &= (1-\omega)\theta^V + \omega V_t, \\ Sk_{t,\tau} &= \frac{TC^H + TC^J}{[Var^H + Var^J]^{3/2}}, \end{aligned} \tag{3.31}$$

*where*

$$\theta^V = \theta^v + 2\theta^\lambda(e^x - 1 - x), \qquad V_t = v_t + 2\lambda_t(e^x - 1 - x),$$

*the contributions of the jump component to variance and the third cumulant are given by*

$$Var^J = \Lambda_t^T x^2 \tau + (e^x - 1 - x)^2 E_t(Z_T^2), \tag{3.32}$$

$$TC^J = \Lambda_t^T x^3 \tau - 3x^2(e^x - 1 - x)E_t(Z_T^2) - (e^x - 1 - x)^3 E_t(Z_T^3), \tag{3.33}$$

*the average jump intensity, $\Lambda_t^T$, is given by*

$$\Lambda_t^T = \frac{1}{T-t}\int_t^T E_t(\lambda_u)du = (1-\omega)\theta^\lambda + \omega\lambda_t.$$

*$Z_T$ is defined by*

$$Z_T \equiv \int_t^T [\lambda_u - E_t(\lambda_u)]du = \sigma_\lambda \int_t^T \frac{1 - e^{-\kappa(T-u)}}{\kappa}\sqrt{\lambda_u}dB_u^\lambda,$$

*and the variance and third cumulant of $Z_T$ are given by*

$$E_t(Z_T^2) = \sigma_\lambda^2 \int_t^T \left(\frac{1 - e^{-\kappa\tau^*}}{\kappa}\right)^2 E_t(\lambda_u)du, \qquad \tau^* = T - u, \tag{3.34}$$

$$E_t(Z_T^3) = 3\sigma_\lambda^4 \int_t^T \frac{1 - e^{-2\kappa\tau^*} - 2\kappa\tau^* e^{-\kappa\tau^*}}{\kappa^3}\frac{1 - e^{-\kappa\tau^*}}{\kappa} E_t(\lambda_u)du, \tag{3.35}$$

[6] In the literature, the stochastic jump intensity is usually assumed to depend on variance; see, among others, Andersen, Benzoni and Lund (2002), Pan (2002) and Eraker (2004). In their setup, the jump intensity also shares the same mean-reverting speed with the diffusive variance. Chen, Joslin and Tran (2012) assume the same process for the stochastic jump intensity as that in our Proposition 3.6.

*and* $E_t(\lambda_u) = \theta^\lambda + (\lambda_t - \theta^\lambda)e^{-\kappa(u-t)}$ *is the expected jump intensity. The term variance,* $Var^H$*, and third cumulant,* $TC^H$*, of the Heston model are given by Proposition 3.3.*

*Proof*. See Appendix 3.7.4.

**Remark 3.6.1.** The mean-reverting speed of $\lambda_t$ is designed to be the same as that of $v_t$, so that the resulting *instantaneous squared VIX* defined by Luo and Zhang (2012), $V_t = v_t + 2\lambda_t(e^x - 1 - x)$, follows a mean-reverting process with the same speed:

$$dV_t = \kappa(\theta^V - V_t)dt + \sigma_v\sqrt{v_t}dB_t^v + 2(e^x - 1 - x)\sigma_\lambda\sqrt{\lambda_t}dB_t^\lambda.$$

The VIX term structure model of Luo and Zhang (2012) can be directly applied.

**Remark 3.6.2.** As we can see from Equations (3.32) and (3.33), the term variance, $Var^J$, and third cumulant, $TC^J$, contributed from jumps consist of two components. One of them is due to the average jump intensity, $\Lambda_t^T$; the other one is due to the uncertainty, $\sigma_\lambda$, in jump intensity. There is no interaction term between stochastic volatility and jumps because they are independent.

**Remark 3.6.3.** In this three-factor model, for $\kappa > 0$, if $\tau$ is small, the asymptotic skewness is given by

$$Sk_{t,\tau} = \frac{\lambda_t x^3}{(v_t + \lambda_t x^2)^{3/2}}\frac{1}{\sqrt{\tau}} + c\sqrt{\tau} + o(\sqrt{\tau}),$$

where

$$c = \frac{3\rho\sigma_v v_t}{2(v_t + \lambda_t x^2)^{3/2}} + \frac{(3\rho\sigma_v v_t\lambda_t + \kappa v_t\lambda_t + 2\kappa v_t\theta^\lambda - 3\kappa\theta^v\lambda_t)x^3}{4(v_t + \lambda_t x^2)^{5/2}} + \frac{\kappa\lambda_t(\lambda_t - \theta^\lambda)x^5}{4(v_t + \lambda_t x^2)^{5/2}}.$$

If $\tau$ is large, then the skewness is given by

$$Sk_{t,\tau} = \frac{a}{b\sqrt{b}}\frac{1}{\sqrt{\tau}} + o\left(\frac{1}{\sqrt{\tau}}\right),$$

where

$$
\begin{aligned}
a &= \theta^{\lambda}\left[x^3 - 3x^2(e^x - 1 - x)\frac{\sigma_\lambda^2}{\kappa^2} - 3(e^x - 1 - x)^3\frac{\sigma_\lambda^4}{\kappa^4}\right] \\
&\quad + \theta^{v}\left[3\rho\frac{\sigma_v}{\kappa} - \frac{3}{2}(1 + 2\rho^2)\frac{\sigma_v^2}{\kappa^2} + \frac{9}{4}\rho\frac{\sigma_v^3}{\kappa^3} - \frac{3}{8}\frac{\sigma_v^4}{\kappa^4}\right], \\
b &= \theta^{\lambda}\left[x^2 + (e^x - 1 - x)^2\frac{\sigma_\lambda^2}{\kappa^2}\right] + \theta^{v}\left(1 - \rho\frac{\sigma_v}{\kappa} + \frac{1}{4}\frac{\sigma_v^2}{\kappa^2}\right).
\end{aligned}
$$

Hence the regularized skewness $\sqrt{\tau}\, Sk_{t,\tau}$ approaches to a constant for both small $\tau$ and large $\tau$.

As explained earlier, we need two variance factors, that is, the instantaneous diffusive variance, $v_t$, and its long-term mean level, $\theta_t^v$, to capture the time-varying VIX term structure. Similarly, in order to capture the time-varying SKEW term structure, we also need two factors, which are the instantaneous and long-term jump intensities, $\lambda_t$ and $\theta_t^\lambda$. The simplest five-factor model is presented as follows.

**Proposition 3.7** *In a jump-diffusion model with stochastic volatility and stochastic jump intensity as well as stochastic corresponding long term mean levels, the risk-neutral underlying stock is modelled by*

$$
\begin{aligned}
\frac{dS_t}{S_t} &= rdt + \sqrt{v_t}dB_t^S + (e^x - 1)dN_t - \lambda_t(e^x - 1)dt, \\
dv_t &= \kappa(\theta_t^v - v_t)dt + \sigma_v\sqrt{v_t}dB_t^v, \\
d\lambda_t &= \kappa(\theta_t^\lambda - \lambda_t)dt + \sigma_\lambda\sqrt{\lambda_t}dB_t^\lambda, \\
d\theta_t^v &= \sigma_1\sqrt{\theta_t^v}dB_{1,t}^v, \\
d\theta_t^\lambda &= \sigma_2\sqrt{\theta_t^\lambda}dB_{2,t}^\lambda,
\end{aligned}
$$

*where the Brownian motions $B_{1,t}^v$, $B_{2,t}^\lambda$ are independent of each other and $B_t^S$, $B_t^v$, $B_t^\lambda$ and $N_t$. The model-implied squared VIX and skewness at time $t$ for time to maturity $\tau$ are as follows:*

$$
\begin{aligned}
VIX_{t,\tau}^2 &= (1 - \omega)\theta_t^V + \omega V_t, \\
Sk_{t,\tau} &= \frac{TC^H + TC^J + TC^M}{[Var^H + Var^J + Var^M]^{3/2}},
\end{aligned}
\tag{3.36}
$$

*where*

$$\theta_t^V = \theta_t^v + 2\theta_t^\lambda(e^x - 1 - x), \qquad V_t = v_t + 2\lambda_t(e^x - 1 - x),$$

*the contributions of long term mean variation to variance and third cumulant are given by*

$$Var^M = Var^{HM} + (e^x - 1 - x)^2\sigma_2^2\theta_t^\lambda C_1, \tag{3.37}$$

$$TC^M = TC^{HM} - 3\sigma_2^2\theta_t^\lambda[x^2(e^x - 1 - x)C_1 + (e^x - 1 - x)^3(\sigma_\lambda^2 C_3 + \sigma_2^2 C_4)], \tag{3.38}$$

*and the variance $Var^H$, third cumulant $TC^H$, of the Heston model are given by Proposition 3.3, $Var^{HM}$, $TC^{HM}$, $C_1$, $C_3$, $C_4$ are given by Proposition 3.4, $Var^J$ and $TC^J$, of jump component are given by Proposition 3.6.*

**Remark 3.7.1.** The result is built by combining those of the Propositions 3.4 and 3.6 and including the additional term variance and third cumulant contributed from the uncertainty of the long-term mean level of jump intensity, $\theta_t^\lambda$.

**Remark 3.7.2.** In this five-factor model, for $\kappa > 0$, if $\tau$ is small, the asymptotic skewness is given by

$$Sk_{t,\tau} = \frac{\lambda_t x^3}{(v_t + \lambda_t x^2)^{3/2}} \frac{1}{\sqrt{\tau}} + o\left(\frac{1}{\sqrt{\tau}}\right).$$

If $\tau$ is large, then the skewness is given by

$$Sk_{t,\tau} = \frac{a}{b\sqrt{b}}\sqrt{\tau} + o(\sqrt{\tau}),$$

where

$$a = -\frac{1}{5}(e^x - 1 - x)^3\theta_t^\lambda\sigma_2^4 - \frac{1}{40}\theta_t^v\sigma_1^4,$$
$$b = \frac{1}{3}(e^x - 1 - x)^2\theta_t^\lambda\sigma_2^2 + \frac{1}{12}\theta_t^v\sigma_1^2.$$

Hence, the regularized skewness $\sqrt{\tau}\, Sk_{t,\tau}$ approaches to a constant for small $\tau$, and behaves linearly with $\tau$ for large $\tau$.

## 3.4 Data

Our daily data on the term structure of VIX are from 24 November 2010 to 31 December 2015, and the data on the term structure of SKEW range from 2 January 1990 to 31 December 2015, provided by the CBOE website.[7] On each day, we have VIX and SKEW data for up to 12 maturity dates. The time to maturity becomes one day shorter as we move forward by one day. We are particularly interested in times to maturity, $\tau$ = 1, 2, $\cdots$ 15 months. Following the practice of the CBOE, we compute the SKEW at time $t$ for the desirable time to maturity $\tau$ by using interpolation as follows:

$$SKEW_{t,\tau} = \omega SKEW_{t,\text{last}} + (1-\omega)SKEW_{t,\text{next}},$$

where $\omega$ is a weight determined by

$$\omega = \frac{\tau_{\text{next}} - \tau}{\tau_{\text{next}} - \tau_{\text{last}}},$$

and $\tau_{\text{last}}$ and $\tau_{\text{next}}$ are the times to maturity (up to minute) of the last and next available data, respectively.[8] We select the days with the maximum time to maturity no less than 15 months (6183 trading days in total), and the summary statistics for the interpolated SKEW term structure are reported in Table 3.1. The average one-month SKEW is 118.28, which is equivalent to the skewness of -1.828 of the distribution of the SPX monthly returns. The average SKEW slightly decreases as the time to maturity increases, but it is still around 115 for the 15-month time to maturity. The maximum of the interpolated SKEW is 201.15 for the four-month time to maturity, and the minimum is 87.20 for the two-month time to maturity, which corresponds to the positive skewness of 1.280.

Figure 3.1 shows the time evolution of the SKEW term structure for time to maturity from one month to 15 months, and Figure 3.2 shows the time series of three selected SKEWs with constant times to maturity. The sample period is from 2 January 1990 to 31 December 2015. As we can see from the figures, the SKEW term

[7] The 30-day VIX and SKEW indices are available from 2 January 1990 to 31 December 2015.

[8] There is an inconsistency in the CBOE practice on calculating the number of days for the time to maturity. The business-day convention is used in computing the term structure of the SKEW; however, the calendar-day convention is used in computing the 30-day SKEW index. In this chapter, we construct constant time to maturity SKEW by using the business-day convention.

structure has a relative stable shape before the financial tsunami. It has become erratic since 2010. The level of SKEW significantly increases in recent years, which indicates that option traders expected higher tail risk after the financial crisis.

Figure 3.3 shows a few samples of SKEW term structures with outliers. They either have a very low minimum SKEW (smaller than 90) or a very high maximum SKEW (higher than 180). The minimum SKEW value (85.28) occurred on 17 May 2013 with expiry date 20 July 2013, and the maximum value (204.72) on 28 May 2013 with expiry date 21 September 2013. The index CBOE put/call ratio changed from 1.35 to 0.96 on 17 May 2013, which implies a bullish sentiment shift. Followed by 0.71, the index put/call ratio is roughly 1.1 on 28 May 2013, which exhibits a bearish sentiment.[9] Therefore, the outliers in May 2013 might be caused by market sentiment. The other two outliers in 2013 could be induced by the low liquidity of long-term options. The four examples in October 2015 indicate option market participants anticipate more negative skewed mid-term or long-term returns due to the concerns about the stock market instability.

## 3.5 Model Estimation and Empirical Performance

The latent state variables in the stochastic volatility models have placed a big challenge for the estimation. Several approaches have been proposed in the literature; see, for example, Bakshi, Cao and Chen (2000) for the simulated method of moments (SMM), Chernov and Ghysels (2000) and Anderson, Benzoni and Lund (2002) for the efficient method of moments (EMM), Pan (2002) for the implied-state generalized method of moments (IS-GMM), and Eraker, Johannes and Polson (2003) and Eraker (2004) for the Markov Chain Monte Carlo (MCMC) method. These estimators, using options and/or returns data, impose severe computational burdens and are not easy to implement. In this chapter, with the carefully processed and publicly available term structure data of the CBOE VIX and SKEW, we develop a new easily implemented estimation procedure, which learns about the model parameters sequentially with feasible computation steps.

[9] The index put/call ratio data are obtained from the CBOE website.

We use the term structure data of VIX and SKEW to estimate the models proposed in Section 3.4. Due to the availability of the CBOE VIX term structure data, we adjust all other data to 24 November 2010 to 31 December 2015 (1284 trading days), including SPX (with extra one-month data before the starting date), the 30-day VIX index and the SKEW term structure data. We apply the sequential estimation method starting from the Black-Scholes (1973) model.

In the Black-Scholes model, there is just one constant unknown $\sigma^2$. We fit the theoretical value of VIX, $100\sigma$, with the VIX term structure data (1284 days with roughly 10 observations each day). The optimal value of $\sigma^2$ is 0.0432, with a VIX root mean squared error (RMSE) of 5.4698. The theoretical value of SKEW in the Black-Scholes model is 100, and a SKEW RMSE is 26.6709. The estimation results are reported in Table 3.2.

In the Merton model, we can confirm two values: the total variance (denoted as $Var$) and the regularized skewness (denoted as $RSk$). The estimation at the VIX stage, which refers to the estimation using the VIX term structure data, is the same as that in the Black-Scholes model. The estimation at the SKEW stage is as follows: We fit the theoretical value of $SKEW = 100 - 10\frac{RSk}{\sqrt{\tau}}$ with the SKEW term structure data. The optimal value for RSk=-1.02, and the SKEW RMSE is 16.655. To further estimate the model, we need an extra condition, which is the diffusive variance proportion 0.88 (denoted as $Prop$), computed as the ratio of the average monthly diffusive variance to the average monthly realized variance of the SPX daily returns.[10] Solving the following three equations

$$Var = \sigma^2 + 2\lambda(e^x - 1 - x), \quad RSk = \frac{\lambda x^3}{(\sigma^2 + \lambda x^2)^{\frac{3}{2}}}, \quad Prop = \frac{\sigma^2}{\sigma^2 + \lambda x^2},$$

[10]Barndorff-Nielsen and Shephard (2004) define the discrete version of the diffusive (integrated) variance as $DV_{t-\tau,t} = \frac{\pi}{2}\sum^{\mathcal{P}} |\delta s_i||\delta s_{i+1}|$, where $\mathcal{P}$ denotes a partition of the interval $[t-\tau, t]$, $s_i$ denotes the $i$th logarithmic price and $\delta$ the first-order difference. We choose the period of a month, that is, $\tau = 1/12$, to calculate the diffusive variance with a daily partition in this chapter. The diffusive $DV_t$ approaches the integrated variance in the probability limit sense, $\underset{||\mathcal{P}||\to 0}{plim}\ DV_t = \int_{t-\tau}^{t} v_s ds$, and is robust to rare jumps, whereas the realized variance $RV_t = \sum^{\mathcal{P}} (\delta s_i)^2$ approaches the integrated variance plus the jump variance, $\underset{||\mathcal{P}||\to 0}{plim}\ RV_t = \int_{t-\tau}^{t} v_s ds + \sum_{i=1}^{N_t} (x_i)^2$, where $N_t$ denotes the total number of jumps in $[t-\tau, t]$ and $x_i$ denotes the jump size given the $i$th jump arrival. We calculate the diffusive variance and realized variance on each trading day from 24 November 2010 to 31 December 2015, and take the proportion of the average diffusive variance to the average realized variance.

gives

$$\sigma^2 = 0.0399, \quad \lambda = 0.00167, \quad x = -180.5\%.$$

The resulting jump size is unreasonably high in magnitude.

In the Heston model, we use a two-step iterative approach, which is also used by Christoffersen, Heston and Jacobs (2009), Luo and Zhang (2012) and Zhang et al. (2017), to estimate the model parameters $(\kappa, \theta^V)$ and latent variable $\{V_t\}$ using the term structure of the CBOE VIX.[11] Specifically, the estimation procedure is as follows:

*Step 1*: Given an initial value $\kappa$ and $\theta^V$, obtain the daily realizations of the instantaneous variance $\{V_t\}$, $t = 1, 2, \cdots, T$, where $T$ is the total number of trading days in the sample. In this step, we are required to solve $T$ optimization problems,

$$\{V_t\} = argmin \sum_{j=1}^{n_t} (VIX_{t,\tau_j} - VIX_{t,\tau_j}^{Mkt})^2,$$

where $VIX_{t,\tau_j}$ is the model-implied value of VIX with time to maturity $\tau_j$ on day $t$, $VIX_{t,\tau_j}^{Mkt}$ is the corresponding market value and $n_t$ is the total number of maturities for the VIX term structure on day $t$.

*Step 2*: Estimate $\kappa$ and $\theta^V$ with $\{V_t\}$ obtained in step 1 by minimizing the overall objective function for VIX:

$$\{\kappa, \theta^V\} = argmin \sum_{t=1}^{T} \sum_{j=1}^{n_t} (VIX_{t,\tau_j} - VIX_{t,\tau_j}^{Mkt})^2.$$

Step 1 and Step 2 are repeated until no further significant decrease exists in the overall objective (i.e., the aggregate sum of the VIX squared errors). We stop the iteration if the change of the overall objective is less than $10^{-4}$. The estimation results are as follows:

$$\kappa = 0.312, \quad \theta^V = 0.149,$$

with a VIX RMSE of 1.3696. However, the model is not fully estimated as we can

[11] Note that there is no jump component in the Heston model, thus the diffusive variance $v_t$ is equal to the instantaneous VIX square $V_t$, and their long-term mean levels, $\theta^v$ and $\theta^V$, are also equal. The same argument holds for the extended Heston model.

only determine $\kappa$ and $\theta^V$ at the VIX level. To obtain $\rho$ and $\sigma_v$, we need the SPX and 30-day VIX data as well as the SKEW term structure data. Specifically, we estimate $\rho$ with $Corr(d\log S_t, dVIX_t^2)$, and $\sigma_v$ with

$$\{\sigma_v\} = argmin \sum_{t=1}^{T} \sum_{j=1}^{m_t} (SKEW_{t,\tau_j} - SKEW_{t,\tau_j}^{Mkt})^2,$$

where $SKEW_{t,\tau_j}$ is the theoretical value of SKEW with time to maturity $\tau_j$ on date $t$, $SKEW_{t,\tau_j}^{Mkt}$ is the corresponding market value, $m_t$ is the total number of maturities for the SKEW term structure on day $t$, and $T$ is the total number of trading days in the sample. Note that $m_t$ might be different from $n_t$, but $T$ is the same for the two-level estimation: the VIX level and the SKEW level. The optimization gives

$$\rho = -0.784, \quad \sigma_v = 0.609,$$

with a SKEW RMSE of 10.201.

In the extended Heston model, where the long-term mean level of the variance is an additional stochastic variable, the estimation procedure is similar to that in the Heston model. Instead, we estimate one parameter $\kappa$ and two latent variables $\{V_t, \theta_t^V\}$. The estimation gives

$$\kappa = 1.96,$$

with a VIX RMSE of 0.5539. At the SKEW level, we obtain

$$\rho = -0.784, \quad \sigma_v = 0.899, \quad \sigma_1 = 0.240,$$

with a SKEW RMSE of 9.122. Note that the average long-term mean $\theta$ (0.071) is half of the previous Heston estimation (0.149). The VIX RMSE reduced by more than half in the extended Heston model. We further check the estimation results using

$$var(dV_t) \approx E(var_t(dV_t)) = \sigma^2 E(V_t)dt,$$

where $var(dV_t)$ denotes the unconditional variance of the change of the instantaneous variance $V_t$ and $E(V_t)$ denotes its expected value. The rough estimation results in $\sigma_v = 0.831$, $\sigma_1 = 0.235$, confirming the consistency of the estimation using the

SKEW term structure data. We use the expected conditional variance $E(var_t(dV_t))$ to approximate the unconditional variance $var(dV_t)$, for the difference between them $var(E_t(dV_t)) = var(dV_t) - E(var_t(dV_t))$ is of order $(dt)^2$.

In the Merton-Heston model, which includes a jump component, we adjust the correlation coefficient $\rho$ using the diffusive variance proportion (0.88), which is the same as the additional condition in the calibration of the Merton model. Our estimation for $\rho$ is based on the following relationship

$$Corr(d\log S_t, dVIX_t^2) \approx \rho\sqrt{\frac{E(v_t)}{E(v_t + x^2\lambda_t)}},$$

where we use the expected conditional variance or covariance to approximate the unconditional one. We substitute $v_t = V_t - 2\lambda(e^x - 1 - x)$ and $\theta^v = \theta^V - 2\lambda(e^x - 1 - x)$ into the SKEW formula, where $V_t$ and $v_t$ denote the instantaneous VIX square and diffusive variance, respectively, $\theta^V$ and $\theta^v$ denote their corresponding long-term mean levels. The mean-reverting speed $\kappa$, the long-term mean level $\theta^V$ and the daily realizations of $\{V_t\}$ are obtained from the VIX-level estimation. They are the same as those obtained under the Heston model. The estimation result for the Merton-Heston Model is as follows

$$\rho = -0.836, \quad \sigma_v = 0.633, \quad x = -16.7\%, \quad \theta^v = 0.142, \quad \lambda = 0.274,$$

with a SKEW RMSE of 8.779.

In the Merton-extended Heston model, we estimate the parameters based on the VIX-level estimation results under the extended Heston model. We omit the impact of the stochastic long-term mean on the calculation of $\rho$, substitute $v_t = V_t - 2\lambda(e^x - 1 - x)$ and $\theta_t^v = \theta_t^V - 2\lambda(e^x - 1 - x)$ into the SKEW formula, and the estimation result is as follows

$$\rho = -0.836, \quad \sigma_v = 0.983, \quad x = -10.6\%, \quad \sigma_1 = 0.173, \quad \lambda = 0.641,$$

with a SKEW RMSE of 8.158. We are interested in the behaviour of the regularized skewness, for it is a constant for the independent and identically distributed returns, as the case in the Merton model. The incorporation of the stochastic volatility

changed this feature. We show the aggregate regularized skewness by substituting the optimal estimated parameters and the averages of the latent variables in Figure 3.4 for the Merton-Heston and Merton-extended Heston models.

In the three-factor model, which is a generalization of the Merton-Heston model by relaxing the constraints $\sigma_\lambda = 0$, $\lambda_t = \theta^\lambda$, we take all the optimal parameters (excluding $\lambda$) from the SKEW-level estimation in the Merton-Heston model and the daily realizations of $\{V_t\}$ from the VIX-level estimation in the Heston model to see how the generalization improves the fitting performance. Given an initial value for $\sigma_\lambda$, the daily estimation using the SKEW term structure data is as follows: With the optimal mean-reverting speed $\kappa = 0.312$, the long-term mean level of the instantaneous VIX square $\theta^V = 0.149$ and time series $\{V_t\}$ obtained from the VIX-level estimation as well as the correlation coefficient $\rho = -0.836$, the volatility of volatility $\sigma_v = 0.633$, the jump size $x = -16.7\%$ and the long-term mean level of the diffusive variance $\theta^v = 0.142$ from the SKEW-level estimation for the Merton-Heston model, we can get the daily diffusive variance $v_t$ by solving $T$ optimizations

$$\{v_t\} = argmin \sum_{j=1}^{m_t} (SKEW_{t,\tau_j} - SKEW_{t,\tau_j}^{Mkt})^2,$$

with the constraint $v_t < V_t$. Note that we substitute $\lambda_t = \frac{V_t - v_t}{2(e^x - 1 - x)}$ and $\theta^\lambda = \frac{\theta^V - \theta^v}{2(e^x - 1 - x)}$ into the SKEW formula. For the estimation of $\sigma_\lambda$, as the SKEW formula is insensitive to $\sigma_\lambda$, we estimate it using $\sqrt{\frac{var(d\lambda_t)}{E(\lambda_t)dt}}$ with the optimal daily ($dt = \frac{1}{252}$) realizations of the jump intensity $\{\lambda_t\}$, computed as $\frac{V_t - v_t}{2(e^x - 1 - x)}$, where the instantaneous VIX square $\{V_t\}$ is the optimal realization from the VIX-level estimation, and the diffusive variance $\{v_t\}$ is from the SKEW-level estimation. We repeat this two-step iterative procedure until the value of $\sigma_\lambda$ remains unchanged after each optimization for $\{\lambda_t\}$. The SKEW RMSE reduced slightly from 8.779 in the Merton-Heston model to 8.604 in the three-factor model. The daily realizations of the instantaneous VIX square, the diffusive variance and the jump intensity are shown in Figure 3.5.

In the five-factor model, which is a generalization of the Merton-extended Heston model by relaxing the constraints $\sigma_\lambda = 0$, $\sigma_2 = 0$, $\lambda_t = \theta_t^\lambda$, the daily estimation procedure is similar to that in the three-factor model, but all the optimal parameters (excluding $\lambda$) are from the SKEW-level estimation in the Merton-extended Heston

model and the daily realizations of $\{V_t, \theta_t^V\}$ are from the VIX-level estimation in the extended Heston model.[12] We substitute $\lambda_t = \frac{V_t - v_t}{2(e^x - 1 - x)}$ and $\theta_t^\lambda = \frac{\theta_t^V - \theta_t^v}{2(e^x - 1 - x)}$ into the SKEW formula, and adopt a three-step iterative procedure for the estimation of the five-factor model. Specifically, given an initial value of $\sigma_\lambda$, we use a two-step procedure, which is the same as that for the VIX-level estimation, to obtain $\{v_t\}$, $\{\theta_t^v\}$ and $\sigma_2$, and then add an outer loop to estimate $\sigma_\lambda$ using $\sqrt{\frac{var(d\lambda_t)}{E(\lambda_t)dt}}$ with the optimal realizations of the jump intensity $\{\lambda_t\}$ until we get an unchanged $\sigma_\lambda$. The SKEW RMSE reduced from 8.158 in the Merton-extended Heston model to 6.953 in the five-factor model. In Figure 3.6, the date 29 November 2013 is chosen to demonstrate the fitting performance of the five-factor model. From this figure, we see that the jumps are essential to produce the nonzero short-term skewness, and the instantaneous skewness becomes infinity at point zero. The daily realizations of the instantaneous VIX square, diffusive variance and the jump intensity and their corresponding long-term mean levels are shown in Figure 3.7.

In Table 3.2, we report all the parameter estimates and corresponding RMSEs for the aforementioned eight models. We measure and compare the models' fitting performances in terms of RMSEs. We find that the Merton-extended Heston model improve the fitting performance of the VIX term structure by 60% compared with the Merton-Heston model with the VIX RMSE changed from 1.3696 to 0.5539, and the Merton-extended Heston (Merton-Heston) model outperforms the Heston model by 20% (14%) when fitting the SKEW term structure with the SKEW RMSE changed from 10.201 to 8.158 (8.779). The performance of the five-factor (three-factor) model in fitting the SKEW term structure exceeds that of the Merton-extended Heston (Merton-Heston) model by 15% (2%) with the SKEW RMSE changed from 8.158 (8.779) to 6.953 (8.604). Our estimation results suggest that the five-factor model is superior when modelling the term structure of the CBOE VIX and SKEW.

We show in the model section that the Black-Scholes (1973), Merton (1976) and Heston (1993) models exhibit zero skewness, constant regularized skewness and zero instantaneous skewness, respectively, which are implausible to explain the stylized skewness of the distribution of the SPX returns; see Das and Sundaram (1999)

[12]The Merton-extend Heston model could be regarded as a special case of the five-factor model with the constraints $\sigma_\lambda = 0$, $\sigma_2 = 0$, $\lambda_t = \theta_t^\lambda$, which basically indicate that the jump intensity is constant. Our estimation results show that the five-factor model exhibits better fitting performance than the Merton-extend Heston model because it relaxes these constraints.

and Zhang et al. (2017). Note that the nonzero correlation between the SPX returns and the daily changes of the VIX square is dominated by the diffusive variance term. It is essential to incorporate the Heston stochastic volatility into the model setup. The estimation result in the Merton model confirms the importance of the stochastic volatility in measuring skewness, as the jump component is extremely large without the correlation term in the skewness formula. Therefore, a model (e.g., the Merton-Heston model) with both a jump component and a return-correlated diffusive variance component, is basic for capturing the time-changing skewness. However, to better model the behaviour of skewness, we further propose a five-factor model, where the diffusive variance, jump intensity and their long-term mean levels are changing over time. Luo and Zhang (2012) show the advantages of modelling the term structure of the VIX square using the extended Heston model, where the long-term mean level follows a martingale process. We find additionally that the extended Heston model is advantageous in modelling the term structure of SKEW compared with the standard Heston model in terms of the lower root mean squared error.

## 3.6 Conclusion

In this chapter, we derive skewness formulas under various affine jump-diffusion models, proposed by Duffie, Pan and Singleton (2000), which are typically used in option pricing. Given the VIX formulas in Luo and Zhang (2012), the skewness formulas provide a new perspective to estimate the model parameters as well as the latent variables using the combined CBOE VIX and SKEW term structure data, which are daily updated on the CBOE website, as opposed to the option cross-sectional data, which are only available in subscribed databases. We also analyse the asymptotic behaviours of the skewness formulas.

To model skewness more accurately, we propose an affine jump diffusion model with five factors, which are logarithmic price, diffusive variance, jump intensity and the long-term mean levels of diffusive variance and jump intensity. We compare the empirical performances of different models, and find that the five-factor model is the most powerful one to fit both the VIX and SKEW term structure in terms of

the lowest root mean squared errors. As the VIX and SKEW term structure data are extracted from option prices, the parameters and latent variables are estimated under the risk-neutral measure, and can be directly applied to option pricing.

## 3.7 Appendix

### 3.7.1 The First Three Moments of the Continuously Compounded Return

Bakshi, Kapadia and Madan (2003) propose a methodology of evaluating the first three moments of the continuously compounded return, $E_t^Q(R_t^T)$, $E_t^Q[(R_t^T)^2]$ and $E_t^Q[(R_t^T)^3]$ by using current prices of European options as follows.

For any twice differentiable function $f(S_T)$, following equality holds

$$\begin{aligned} f(S_T) = f(K_0) + f'(K_0)(S_T - K_0) + \int_0^{K_0} f''(K)\max(K - S_T, 0)dK \\ + \int_{K_0}^{+\infty} f''(K)\max(S_T - K, 0)dK, \end{aligned} \tag{3.39}$$

where $K_0$ is a reference strike price that could take any value. This mathematical equality has a profound financial meaning: A European-style derivative with an arbitrary payoff function, $f(S_T)$, can be decomposed into a portfolio of bonds with a face value $f(K_0)$, $f'(K_0)$ amount of forward contract and $f''(K)$ amount of European options, with strikes between $0$ and $K_0$ for puts and between $K_0$ and $+\infty$ for calls.

Applying Equation (3.39) to the power functions of $\ln \frac{S_T}{F_t^T}$ gives

$$\begin{aligned} \ln \frac{S_T}{F_t^T} &= \ln \frac{K_0}{F_t^T} + \frac{S_T}{K_0} - 1 - \int_0^{K_0} \frac{1}{K^2}\max(K - S_T, 0)dK \\ &\quad - \int_{K_0}^{+\infty} \frac{1}{K^2}\max(S_T - K, 0)dK, \\ \ln^2 \frac{S_T}{F_t^T} &= \ln^2 \frac{K_0}{F_t^T} + 2\ln \frac{K_0}{F_t^T}\left(\frac{S_T}{K_0} - 1\right) + \int_0^{K_0} \frac{2 - 2\ln \frac{K}{F_t^T}}{K^2}\max(K - S_T, 0)dK \\ &\quad + \int_{K_0}^{+\infty} \frac{2 - 2\ln \frac{K}{F_t^T}}{K^2}\max(S_T - K, 0)dK, \end{aligned}$$

$$\ln^3 \frac{S_T}{F_t^T} = \ln^3 \frac{K_0}{F_t^T} + 3\ln^2 \frac{K_0}{F_t^T}\left(\frac{S_T}{K_0} - 1\right) + \int_0^{K_0} \frac{6\ln\frac{K}{F_t^T} - 3\ln^2\frac{K}{F_t^T}}{K^2} \max(K - S_T, 0)dK$$
$$+ \int_{K_0}^{+\infty} \frac{6\ln\frac{K}{F_t^T} - 3\ln^2\frac{K}{F_t^T}}{K^2} \max(S_T - K, 0)dK.$$

Applying conditional expectation to the three equations in the risk-neutral measure, we notice that

$$E_t^Q[\max(S_T - K, 0)] = e^{r\tau} c_t(K), \qquad E_t^Q[\max(K - S_T, 0)] = e^{r\tau} p_t(K),$$

where $c_t(K)$ ($p_t(K)$) is call (put) option price at the current time. Evaluating the integration approximately using discretization gives the first three moments in Equations (3.5), (3.6) and (3.7).

### 3.7.2 The Merton Model

In the Merton (1976) model, the risk-neutral logarithmic process of the underlying stock is as follows

$$d\ln S_t = \left[r - \frac{1}{2}\sigma^2 - \lambda(e^x - 1 - x)\right]dt + \sigma dB_t + x dN_t - \lambda x dt.$$

We define the continuously compounded return from time $t$ to $T$ as $R_t^T \equiv \ln\frac{S_T}{S_t}$, then the variance and third cumulant of $R_t^T$ are given by

$$E_t^Q[R_t^T - E_t^Q(R_t^T)]^2 = E_t^Q\Big(\int_t^T \sigma dB_u + x dN_u - \lambda x du\Big)^2 = (\sigma^2 + \lambda x^2)\tau,$$
$$E_t^Q[R_t^T - E_t^Q(R_t^T)]^3 = E_t^Q\Big(\int_t^T \sigma dB_u + x dN_u - \lambda x du\Big)^3 = \lambda x^3 \tau.$$

Therefore, the skewness of the continuously compounded return is given by

$$Sk_t = \frac{E_t^Q[R_t^T - E_t^Q(R_t^T)]^3}{\{E_t^Q[R_t^T - E_t^Q(R_t^T)]^2\}^{3/2}} = \frac{\lambda x^3}{(\sigma^2 + \lambda x^2)^{3/2}}\tau^{-1/2}.$$

### 3.7.3 The Heston Model with Stochastic Long-Term Mean

We generalize the Heston variance process by adding another stochastic component, the long term mean $\theta_t$, as follows

$$dv_t = \kappa(\theta_t - v_t)dt + \sigma_v\sqrt{v_t}dB_t^v, \tag{3.40}$$

$$d\theta_t = \sigma_\theta\sqrt{\theta_t}dB_t^\theta, \tag{3.41}$$

where $B_t^\theta$ is independent of $B_t^v$ and $B_t^S$.

Converting Equation (3.40) into the stochastic integral form and plugging Equation (3.41) yields

$$v_s = E_t(v_s) + \sigma_v\int_t^s e^{-\kappa(s-u)}\sqrt{v_u}dB_u^v + \sigma_\theta\int_t^s(1-e^{-\kappa(s-u)})\sqrt{\theta_u}dB_u^\theta, \tag{3.42}$$

where the expectation of $v_s$ at time $t$ ($t < s$) is given by

$$E_t(v_s) = \theta_t + (v_t - \theta_t)e^{-\kappa(s-t)}.$$

Following the procedure in Zhang et al. (2017), we define two integrals as

$$X_T \equiv \int_t^T \sqrt{v_u}dB_u^S, \qquad Y_T \equiv \int_t^T [v_u - E_t^Q(v_u)]du.$$

Substituting Equation (3.42) and interchanging the order of integrations gives

$$\begin{aligned} Y_T &= \sigma_v\int_t^T\int_t^s e^{-\kappa(s-u)}\sqrt{v_u}dB_u^v ds + \sigma_\theta\int_t^T\int_t^s(1-e^{-\kappa(s-u)})\sqrt{\theta_u}dB_u^\theta ds, \\ &= \sigma_v\int_t^T\frac{1-e^{-\kappa(T-u)}}{\kappa}\sqrt{v_u}dB_u^v + \sigma_\theta\int_t^T\frac{e^{-\kappa(T-u)}-1+\kappa(T-u)}{\kappa}\sqrt{\theta_u}dB_u^\theta. \end{aligned}$$

We introduce new martingale processes, $Y_s^H$ and $Y_s^M$, as follows

$$\begin{aligned} Y_s^H &= \sigma_v\int_t^s\frac{1-e^{-\kappa(T-u)}}{\kappa}\sqrt{v_u}dB_u^v, \\ Y_s^M &= \sigma_\theta\int_t^s\frac{e^{-\kappa(T-u)}-1+\kappa(T-u)}{\kappa}\sqrt{\theta_u}dB_u^\theta. \end{aligned}$$

Therefore, at time $T$, we have $Y_T = Y_T^H + Y_T^M$.

To express the contribution of the stochastic long-term mean explicitly, using the independency of $B_s^\theta$ and the martingale property of $Y_s^H$, $Y_s^M$ and $\theta_s$, we expand and simplify the variance and third cumulant of $R_t^T$ in Zhang et al. (2017) as follows:

$$\begin{aligned}
E_t^Q[R_t^T - E_t^Q(R_t^T)]^2 &= E_t^Q(X_T^2) - E_t^Q(X_T Y_T) + \frac{1}{4}E_t^Q(Y_T^2) \\
&= Var^H - E_t^Q(X_T Y_T^M) + \frac{1}{4}E_t^Q[2Y_T^H Y_T^M + (Y_T^M)^2] \\
&= Var^H + \frac{1}{4}E_t^Q[(Y_T^M)^2],
\end{aligned}$$

$$\begin{aligned}
E_t^Q[R_t^T - E_t^Q(R_t^T)]^3 &= E_t^Q(X_T^3) - \frac{3}{2}E_t^Q(X_T^2 Y_T) + \frac{3}{4}E_t^Q(X_T Y_T^2) - \frac{1}{8}E_t^Q(Y_T^3) \\
&= TC^H - \frac{3}{2}E_t^Q(X_T^2 Y_T^M) + \frac{3}{4}E_t^Q[2X_T Y_T^H Y_T^M + X_T(Y_T^M)^2] \\
&\quad -\frac{1}{8}E_t^Q[3(Y_T^H)^2 Y_T^M + 3Y_T^H(Y_T^M)^2 + (Y_T^M)^3] \\
&= TC^H - \frac{3}{2}E_t^Q(X_T^2 Y_T^M) + \frac{3}{2}E_t^Q(X_T Y_T^H Y_T^M) \\
&\quad -\frac{1}{8}E_t^Q[3(Y_T^H)^2 Y_T^M + (Y_T^M)^3],
\end{aligned}$$

where the variance $Var^H$ and third cumulant $TC^H$ are of the original forms in Zhang et al. (2017), with no impact of stochastic long-term mean.

We need the following results in expressing the extra terms arising from the stochastic long-term mean.

**Lemma 1** *The correlations between $Y_s^M$ and $\theta_s$ as well as $v_s$ are given by*

$$E_t^Q(Y_s^M \theta_s) = \sigma_\theta^2 \theta_t \int_t^s \frac{e^{-\kappa(T-u)} - 1 + \kappa(T-u)}{\kappa} du, \tag{3.43}$$

$$E_t^Q(Y_s^M v_s) = \sigma_\theta^2 \theta_t \int_t^s (1 - e^{-\kappa(s-u)}) \frac{e^{-\kappa(T-u)} - 1 + \kappa(T-u)}{\kappa} du. \tag{3.44}$$

*Proof.* Using Ito's Lemma and martingale property of $Y_u^M$ and $\theta_u$, we have

$$\begin{aligned} E_t^Q(Y_s^M\theta_s) &= E_t^Q\int_t^s d(Y_u^M\theta_u) = E_t^Q\int_t^s \theta_u dY_u^M + Y_u^M d\theta_u + dY_u^M d\theta_u \\ &= \sigma_\theta^2\theta_t\int_t^s \frac{e^{-\kappa(T-u)}-1+\kappa(T-u)}{\kappa}du, \end{aligned}$$

which is equivalent to Equation (3.43).

Furthermore, we have the integral

$$\begin{aligned} \int_t^s E_t^Q(Y_u^M\theta_u)du &= \sigma_\theta^2\theta_t\int_t^s\int_t^u \frac{e^{-\kappa(T-r)}-1+\kappa(T-r)}{\kappa}drdu \\ &= \sigma_\theta^2\theta_t\int_t^s (s-u)\frac{e^{-\kappa(T-u)}-1+\kappa(T-u)}{\kappa}du. \end{aligned}$$

Using Ito's Lemma and martingale property of $Y_u^M$, we have

$$\begin{aligned} E_t^Q(Y_s^M v_s) &= E_t^Q\int_t^s d(Y_u^M v_u) = E_t^Q\int_t^s v_u dY_u^M + Y_u^M dv_u + dY_u^M dv_u \\ &= -\kappa E_t^Q\int_t^s E_t^Q(Y_u^M v_u)du + \kappa\int_t^s E_t^Q(Y_u^M\theta_u)du \\ &= -\kappa E_t^Q\int_t^s E_t^Q(Y_u^M v_u)du + \sigma_\theta^2\theta_t\int_t^s \kappa(s-u)\frac{e^{-\kappa(T-u)}-1+\kappa(T-u)}{\kappa}du. \end{aligned}$$

Solving the ordinary differential equation (ODE) gives

$$\begin{aligned} \int_t^s E_t^Q(Y_u^M v_u)du &= \sigma_\theta^2\theta_t\int_t^s e^{-\kappa(s-r)}\int_t^r \kappa(r-u)\frac{e^{-\kappa(T-u)}-1+\kappa(T-u)}{\kappa}dudr \\ &= \sigma_\theta^2\theta_t\int_t^s \frac{e^{-\kappa(s-u)}-1+\kappa(s-u)}{\kappa}\frac{e^{-\kappa(T-u)}-1+\kappa(T-u)}{\kappa}du. \end{aligned}$$

Taking differentiation with respect to s gives

$$E_t^Q(Y_s^M v_s) = \sigma_\theta^2\theta_t\int_t^s (1-e^{-\kappa(s-u)})\frac{e^{-\kappa(T-u)}-1+\kappa(T-u)}{\kappa}du,$$

which is equivalent to Equation (3.44). This completes the proof.

Using Ito's Isometry and the martingale property of $\theta_u$, we have

$$E_t^Q[(Y_T^M)^2] = E_t^Q\Big(\sigma_\theta \int_t^T \frac{e^{-\kappa(T-u)} - 1 + \kappa(T-u)}{\kappa}\sqrt{\theta_u}dB_u^\theta\Big)^2 = \sigma_\theta^2\theta_t C_1,$$

where $C_1 = \int_t^T \frac{(e^{-\kappa(T-u)}-1+\kappa(T-u))^2}{\kappa^2}du$.

Using Ito's Lemma and martingale property of $X_u$ and $Y_u^M$, we have

$$\begin{aligned}
E_t^Q(X_T^2 Y_T^M) &= E_t^Q \int_t^T d(X_u^2 Y_u^M) \\
&= E_t^Q \int_t^T 2X_u Y_u^M dX_u + X_u^2 dY_u^M + Y_u^M (dX_u)^2 + 2X_u dX_u dY_u^M \\
&= \int_t^T E_t^Q(Y_u^M v_u)du = \sigma_\theta^2\theta_t C_1.
\end{aligned}$$

Using Ito's Lemma and martingale property of $X_u$, $Y_u^H$ and $Y_u^M$, we have

$$\begin{aligned}
E_t^Q[X_T Y_T^H Y_T^M] &= E_t^Q \int_t^T d(X_u Y_u^H Y_u^M) \\
&= E_t^Q \int_t^T Y_u^H Y_u^M dX_u + X_u Y_u^M dY_u^H + X_u Y_u^H dY_u^M \\
&\qquad + Y_u^M dX_u dY_u^H + Y_u^H dX_u dY_u^M + X_u dY_u^H dY_u^M \\
&= \rho\sigma_v \int_t^T \frac{1-e^{-\kappa(T-u)}}{\kappa} E_t^Q(Y_u^M v_u)du = \rho\sigma_v\sigma_\theta^2\theta_t C_2,
\end{aligned}$$

where $C_2 = \int_t^T \frac{1-e^{-\kappa(T-u)}}{\kappa}\int_t^u (1-e^{-\kappa(u-s)})\frac{e^{-\kappa(T-s)}-1+\kappa(T-s)}{\kappa}dsdu$.

Using Ito's Lemma and martingale property of $Y_u^H$ and $Y_u^M$, we have

$$\begin{aligned}
E_t^Q[(Y_T^H)^2 Y_T^M] &= E_t^Q \int_t^T d[(Y_u^H)^2 Y_T^M] \\
&= E_t^Q \int_t^T 2Y_u^H Y_u^M dY_u^H + (Y_u^H)^2 dY_u^M + Y_u^M (dY_u^H)^2 + 2Y_u^H dY_u^H dY_u^M \\
&= \sigma_v^2 \int_t^T \frac{(1-e^{-\kappa(T-u)})^2}{\kappa^2} E_t^Q(Y_u^M v_u)du = \sigma_v^2\sigma_\theta^2\theta_t C_3,
\end{aligned}$$

where $C_3 = \int_t^T \frac{(1-e^{-\kappa(T-u)})^2}{\kappa^2}\int_t^u (1-e^{-\kappa(u-s)})\frac{e^{-\kappa(T-s)}-1+\kappa(T-s)}{\kappa}dsdu$.

Using Ito's Lemma and martingale property of $Y_u^M$, we have

$$\begin{aligned} E_t^Q[(Y_T^M)^3] &= E_t^Q \int_t^T d(Y_u^M)^3 = E_t^Q \int_t^T 3(Y_u^M)^2 dY_u^M + 3Y_u^M (dY_u^M)^2 \\ &= 3\sigma_\theta^2 \int_t^T \frac{(e^{-\kappa(T-u)} - 1 + \kappa(T-u))^2}{\kappa^2} E_t^Q(Y_u^M \theta_u) du = 3\sigma_\theta^4 \theta_t C_4, \end{aligned}$$

where $C_4 = \int_t^T \frac{(e^{-\kappa(T-u)}-1+\kappa(T-u))^2}{\kappa^2} \int_t^u \frac{e^{-\kappa(T-s)}-1+\kappa(T-s)}{\kappa} ds du$.

### 3.7.4 Affine Jump-Diffusion Models

In a jump-diffusion model, the risk-neutral underlying stock is modelled by

$$\frac{dS_t}{S_t} = r dt + \sqrt{v_t} dB_t^S + (e^x - 1) dN_t - \lambda_t (e^x - 1) dt, \tag{3.45}$$

where $r$ is the risk-free rate, the jump size $x$ is constant, and the diffusive variance $v_t$ and jump intensity $\lambda_t$ are stochastic. We do not specify the processes for $v_t$ and $\lambda_t$ to derive generic expressions for the second and third central moments of the continuously compounded return in jump-diffusion models.

Applying Ito's Lemma to Equation (3.45) gives

$$d \ln S_t = \left[ r - \frac{1}{2} v_t - \lambda_t (e^x - 1 - x) \right] dt + \sqrt{v_t} dB_t^S + x dN_t - \lambda_t x dt. \tag{3.46}$$

The continuously compounded return from current time $t$, to a future time, $T$, is defined by

$$R_t^T \equiv \ln \frac{S_T}{S_t} = \int_t^T \left[ r - \frac{1}{2} v_u - \lambda_u (e^x - 1 - x) \right] du + \sqrt{v_u} dB_u^S + x dN_u - \lambda_u x du. \tag{3.47}$$

The conditional expectation at time $t$ is then given by

$$E_t^Q(R_t^T) = \int_t^T \left[ r - \frac{1}{2} E_t^Q(v_u) - E_t^Q(\lambda_u)(e^x - 1 - x) \right] du. \tag{3.48}$$

Following the notations in Zhang et al. (2017), we introduce another two integrals

$$Z_T \equiv \int_t^T [\lambda_u - E_t^Q(\lambda_u)]du, \qquad I_T \equiv \int_t^T dN_u - \lambda_u du.$$

With those notations as well as $X_T$ and $Y_T$, subtracting (3.48) from (3.47) yields

$$R_t^T - E_t^Q(R_t^T) = X_T - \frac{1}{2}Y_T + xI_T - (e^x - 1 - x)Z_T. \tag{3.49}$$

Assuming that there is no jumps in variance and jump intensity is governed by a stochastic component that is independent of any random variables in variance and price, we obtain the variance and third cumulant of $R_t^T$ as follows

$$\begin{aligned}
E_t^Q[R_t^T - E_t^Q(R_t^T)]^2 &= E_t^Q(X_T^2) - E_t^Q(X_T Y_T) + \frac{1}{4}E_t^Q(Y_T^2) + x^2 E_t^Q(I_T^2) \\
&\quad + (e^x - 1 - x)^2 E_t^Q(Z_T^2), \\
E_t^Q[R_t^T - E_t^Q(R_t^T)]^3 &= E_t^Q(X_T^3) - \frac{3}{2}E_t^Q(X_T^2 Y_T) + \frac{3}{4}E_t^Q(X_T Y_T^2) - \frac{1}{8}E_t^Q(Y_T^3) \\
&\quad + x^3 E_t^Q(I_T^3) - 3x^2(e^x - 1 - x)E_t^Q(Z_T^2) - (e^x - 1 - x)^3 E_t^Q(Z_T^3),
\end{aligned}$$

where the variance of $X_T$, the variance and third cumulant of $I_T$ are given by

$$E_t^Q(X_T^2) = \int_t^T E_t^Q(v_u)du, \qquad E_t^Q(I_T^2) = E_t^Q(I_T^3) = \int_t^T E_t^Q(\lambda_u)du.$$

To express the terms in variance and third cumulant of $R_t^T$ explicitly, we need to specify the processes for the diffusive variance and jump intensity. See Zhang et al. (2017) for the case of the Heston model. Furthermore, if the jump intensity also follows a square root process, we obtain similar results for $E_t^Q(Z_T^2)$ and $E_t^Q(Z_T^3)$ to those in the Heston model.

TABLE 3.1: Descriptive Statistics of the CBOE SKEW Term Structure

We interpolate the daily SKEW term structure to constant time to maturity, $\tau = 1, 2, \cdots, 15$ months. The sample period is from 2 January 1990 to 31 December 2015. We only consider the days when the maximum time to maturity is not less than 15 months.

| Maturity | Mean | Std Dev | Skewness | Kurtosis | Minimum | Maximum |
|---|---|---|---|---|---|---|
| 1 | 118.2838 | 6.4471 | 0.8605 | 4.1506 | 101.0830 | 152.6511 |
| 2 | 117.7163 | 6.5946 | 1.9604 | 11.9646 | 87.1974 | 176.4242 |
| 3 | 117.5648 | 7.1675 | 2.4682 | 16.3426 | 96.8083 | 185.8500 |
| 4 | 117.6166 | 5.9607 | 1.7114 | 13.4535 | 102.1300 | 201.1458 |
| 5 | 117.4569 | 5.7951 | 0.9476 | 5.7038 | 101.9557 | 176.1263 |
| 6 | 116.9714 | 6.0167 | 0.7759 | 3.8101 | 99.5674 | 151.1067 |
| 7 | 116.4618 | 6.3072 | 0.7679 | 3.4917 | 99.8546 | 147.7979 |
| 8 | 115.9695 | 6.5611 | 0.8015 | 3.5354 | 98.3849 | 153.6933 |
| 9 | 115.4839 | 6.7418 | 0.8370 | 3.8238 | 99.2181 | 165.8313 |
| 10 | 115.0946 | 6.8667 | 0.7507 | 3.4183 | 95.4138 | 152.2994 |
| 11 | 114.9395 | 7.0655 | 0.6711 | 3.1955 | 93.1753 | 147.1083 |
| 12 | 115.0983 | 7.3712 | 0.8476 | 3.6436 | 97.0592 | 152.6782 |
| 13 | 115.3490 | 7.7741 | 1.1685 | 4.9493 | 97.8423 | 158.9494 |
| 14 | 115.4172 | 7.9974 | 1.2156 | 5.0365 | 97.1610 | 164.1166 |
| 15 | 115.2539 | 8.1480 | 1.3494 | 6.4342 | 96.4798 | 179.0453 |

TABLE 3.2: Estimation Results

This table shows the estimation results for eight models: BS (Black-Scholes), M (Merton), H (Heston), eH (extended Heston), MH (Merton-Heston), 3F (three-factor), MeH (Merton-extended Heston) and 5F (five-factor). The estimation comprises two levels: the VIX level and the SKEW level. For the M model, the VIX and SKEW term structure data are not sufficient to determine the parameters. Therefore, we use the diffusive variance proportion (0.88) computed as the ratio of the average monthly diffusive variance to average monthly realized variance with the SPX daily returns data as an additional condition. For the 3F and 5F models, we use the optimal parameters estimated from the MH and MeH models, respectively, and separate the diffusive variance from total variance indicated by instantaneous VIX square, which also includes the jump variance. The correlation coefficient between returns and changes of diffusive variance is calculated using the SPX and the 30-day VIX index, and adjusted with the diffusive variance proportion for the models where jumps exist in prices. The average values of the latent variables ($v, \lambda, \theta^v, \theta^\lambda$) are reported with the corresponding standard deviations in parentheses. The columns labelled VIX and SKEW report the root mean squared errors (RMSE) for the estimations using the VIX and SKEW term structure data, respectively. The sample period is from 24 November 2010 to 31 December 2015.

| | | | | | | | | | | | | RMSE | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\kappa$ | $\rho$ | $x$ | $\sigma_v$ | $\sigma_\lambda$ | $\sigma_1$ | $\sigma_2$ | $v$ | $\lambda$ | $\theta^v$ | $\theta^\lambda$ | VIX | SKEW |
| BS | - | - | - | - | - | - | - | 0.0432 | - | - | - | 5.4698 | 26.671 |
| M | - | - | -180.5% | - | - | - | - | 0.0399 | 0.00167 | - | - | 5.4698 | 16.655 |
| H | 0.312 | -0.784 | - | 0.609 | - | - | - | 0.0343(0.0258) | - | 0.149 | - | 1.3696 | 10.201 |
| eH | 1.96 | -0.784 | - | 0.899 | - | 0.240 | - | 0.0303(0.0283) | - | 0.0709(0.0245) | - | 0.5539 | 9.122 |
| MH | 0.312 | -0.836 | -16.7% | 0.633 | - | - | - | 0.0270(0.0258) | 0.274 | 0.142 | - | 1.3696 | 8.779 |
| 3F | 0.312 | -0.836 | -16.7% | 0.633 | 6.010 | - | - | 0.0187(0.0190) | 0.589(0.504) | 0.142 | 0.274 | 1.3696 | 8.604 |
| MeH | 1.96 | -0.836 | -10.6% | 0.983 | - | 0.173 | - | 0.0233(0.0283) | 0.641 | 0.0639(0.0245) | - | 0.5539 | 8.158 |
| 5F | 1.96 | -0.836 | -10.6% | 0.983 | 9.748 | 0.173 | 5.269 | 0.0247(0.0295) | 0.508(0.598) | 0.0504(0.0395) | 1.88(1.77) | 0.5539 | 6.953 |

FIGURE 3.1: The Time Evolution of the CBOE SKEW Term Structure.

This graph shows the time evolution of the interpolated SKEW term structure for the time to maturity, $\tau = 1, 2, \cdots, 15$ months. The sample period is from 2 January 1990 to 31 December 2015. We only consider the days when the maximum time to maturity is not less than 15 months.

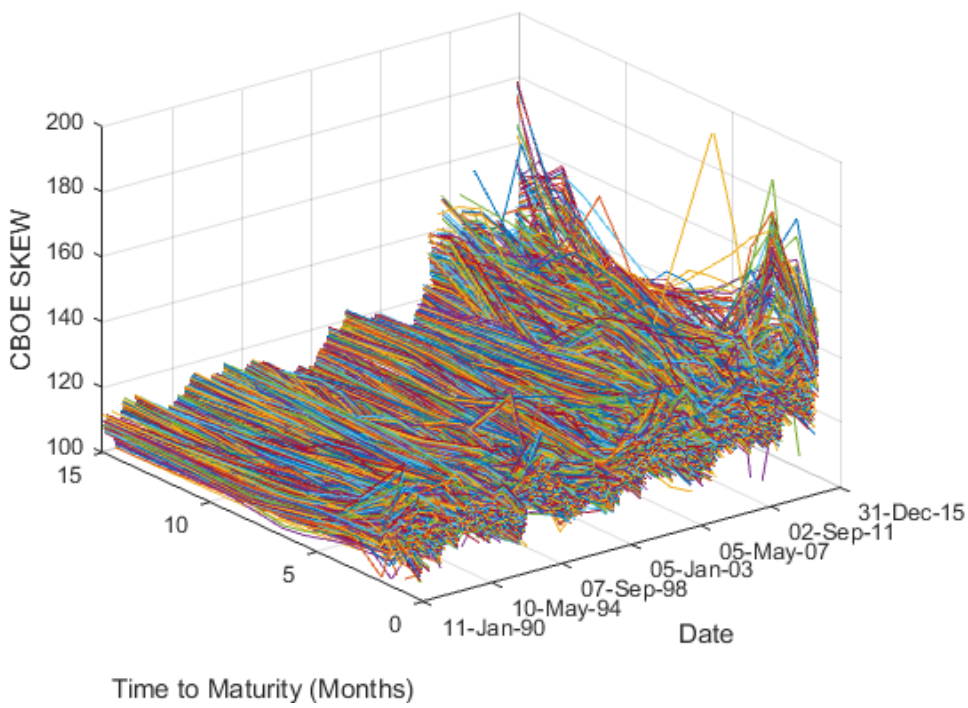


FIGURE 3.2: The Time Series of the CBOE SKEW with Fixed Times to Maturity

This graph shows the time series of the CBOE SKEW with one month, six months and 15 months to maturity. The sample period is from 2 January 1990 to 31 December 2015. We only consider the days when the maximum time to maturity is not less than 15 months.

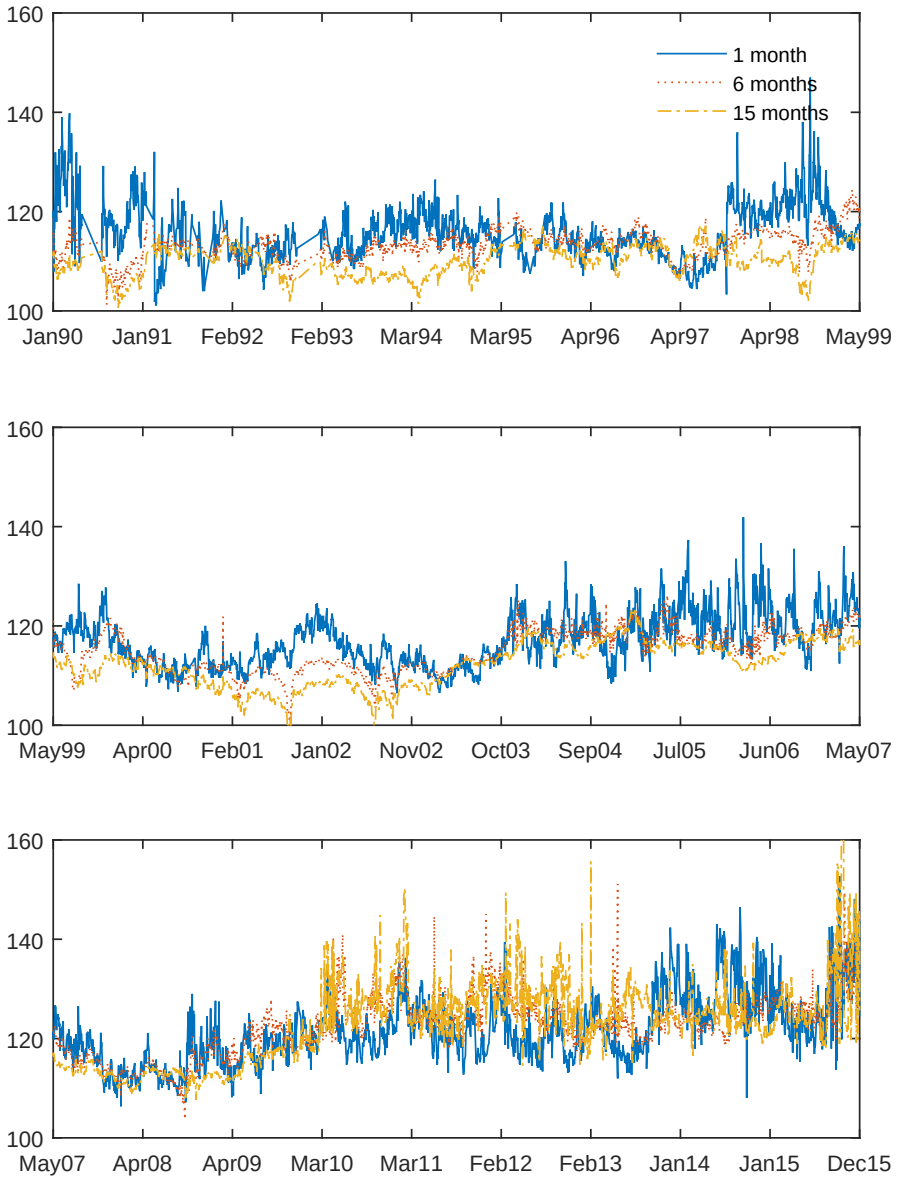


FIGURE 3.3: A few Samples of the CBOE SKEW Term Structure with Outliers

We choose the days when the CBOE SKEW is above 180 or below 90 for the period 2 January 1990 to 31 December 2015.

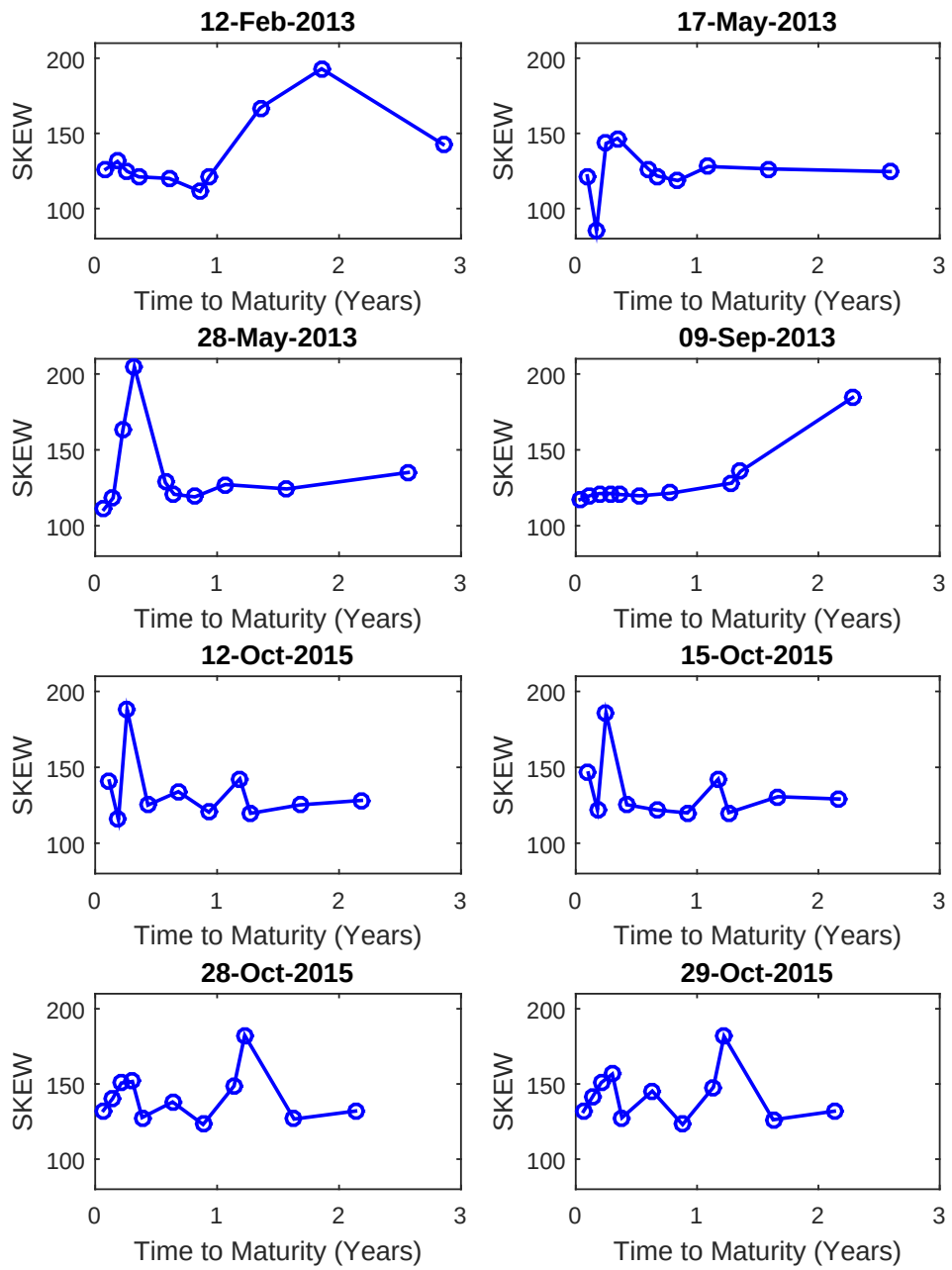


FIGURE 3.4: Model-Implied Aggregate Regularized Skewness.

The regularized skewness is defined as the product of skewness and the square root of time to maturity. The solid line represents the model-implied regularized skewness in the Merton-Heston model (upper) or the Merton-extended Heston model (lower), respectively, where the inputs are the optimal estimated parameters and averages of the latent variables: $\kappa = 0.312$, $\rho = -0.836$, $x = -16.7\%$, $\sigma_v = 0.633$, $\lambda = 0.274$, $\theta^V = 0.149$, $\theta^v = 0.142$, $\bar{V}_t = 0.0343$ and $\bar{v}_t = 0.0270$ for the Merton-Heston model and $\kappa = 1.96$, $\rho = -0.836$, $x = -10.6\%$, $\sigma_v = 0.983$, $\sigma_1 = 0.173$, $\lambda = 0.641$, $\bar{\theta}_t^V = 0.0709$, $\bar{\theta}_t^v = 0.0639$, $\bar{V}_t = 0.0303$ and $\bar{v}_t = 0.0233$ for the Merton-extended Heston model. The dashed line represents the linear asymptote of the model-implied regularized skewness at zero and the dotted line represents the Merton regularized skewness. The sample period is from 24 November 2010 to 31 December 2015.

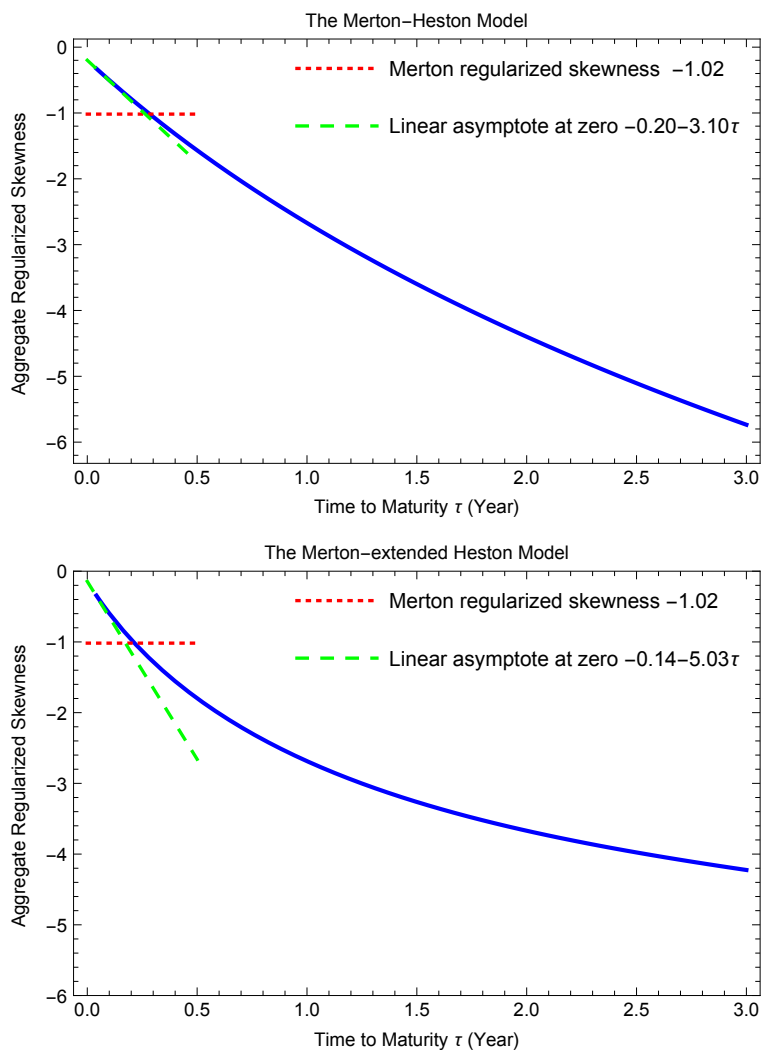


FIGURE 3.5: The Three-Factor Model

This graph shows the daily optimal realizations of the instantaneous squared VIX, diffusive variance, jump intensity and their long term mean levels during the sample period from 24 November 2010 to 31 December 2015. The averages of the instantaneous squared VIX, diffusive variance and jump intensity are 0.0343, 0.0187 and 0.589 with standard deviations 0.0258, 0.0190 and 0.504, respectively. Their long-term mean levels are 0.149, 0.142 and 0.274, respectively. The other optimal parameters include $\rho = -0.836$ from the correlation between the SPX returns and the changes of the 30-day VIX square, and $\kappa = 0.312$, $x = -16.7\%$, $\sigma_v = 0.633$ and $\sigma_\lambda = 6.010$ from the sequential estimation using the term structure data of the VIX and SKEW.

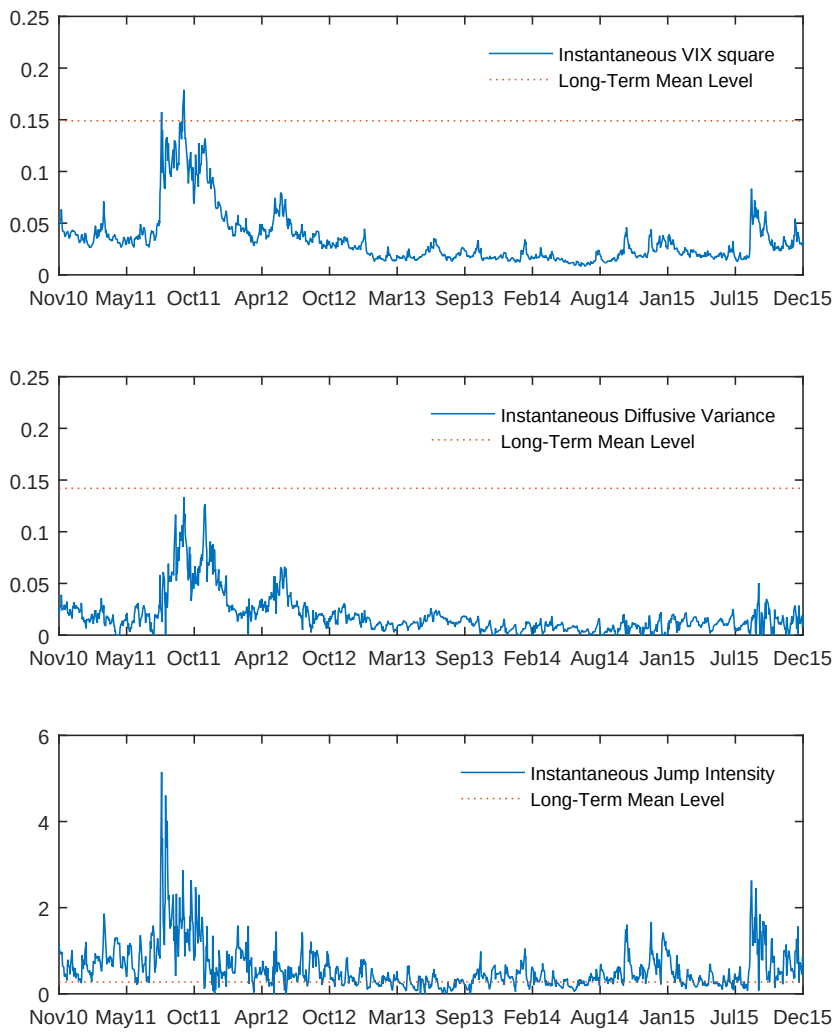


FIGURE 3.6: Fitting Performance on a Representative Day

This graph shows the fitting performance on 29 November 2013 in the five-factor model. The sample period for estimation is from 24 November 2010 to 31 December 2015. The solid lines represent the theoretical values given the optimal estimate of the model parameters and latent variables, and the dots represent the market values. We obtain $\rho = -0.836$ using the correlation between the SPX returns and the changes of the 30-day VIX square. The estimation using the VIX term structure data results in $\kappa = 1.96$, $V_t = 0.0147$ and $\theta_t^V = 0.0537$ with a VIX RMSE of 0.3772, and the estimation using the SKEW term structure data results in $x = -10.6\%$, $\sigma_v = 0.983$, $\sigma_\lambda = 9.748$, $\sigma_1 = 0.173$, $\sigma_2 = 5.269$, $v_t = 0.0026$, $\theta_t^v = 0.0209$, $\lambda_t = 1.11$ and $\theta_t^\lambda = 3.00$ with a SKEW RMSE of 1.442. RMSE stands for the root mean squared error.

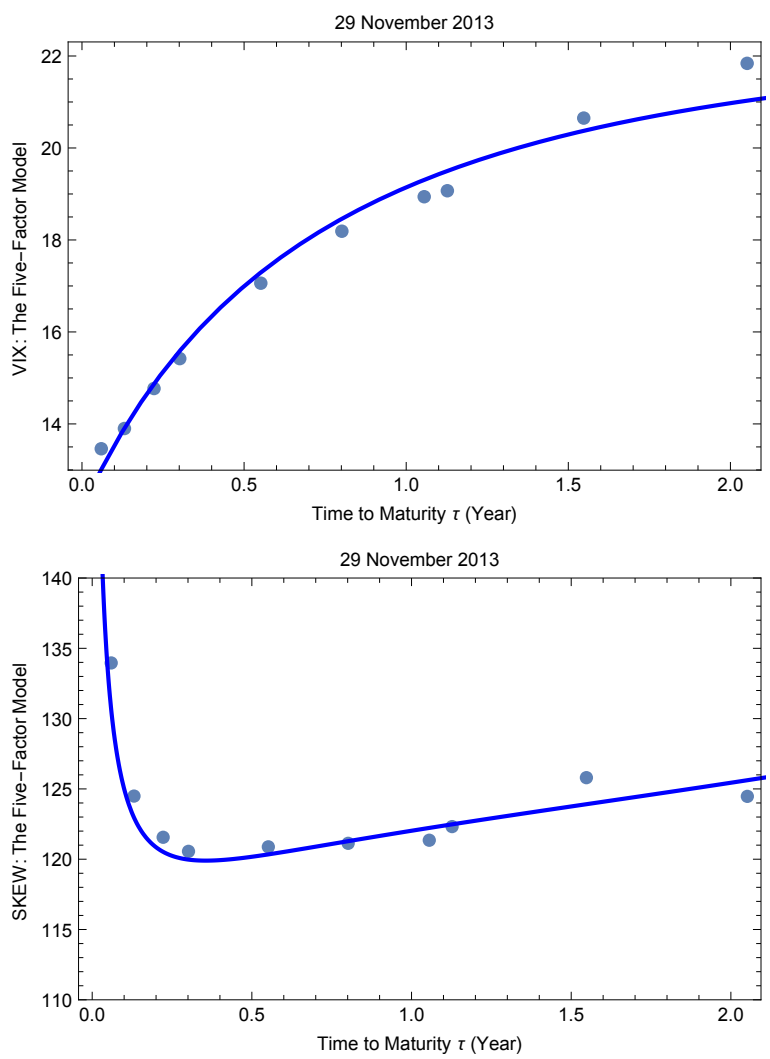
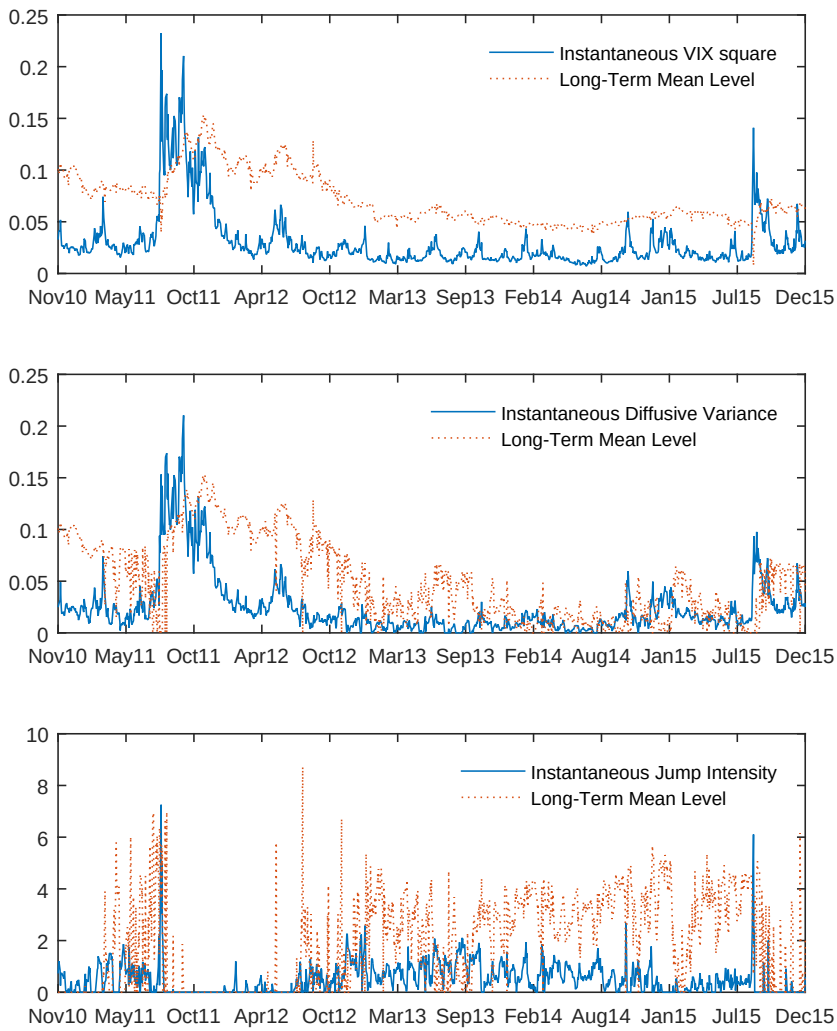


FIGURE 3.7: The Five-Factor Model

This graph shows the daily optimal realizations of the instantaneous squared VIX, diffusive variance, jump intensity and their long term mean levels during the sample period from 24 November 2010 to 31 December 2015. The averages of the instantaneous squared VIX, diffusive variance and jump intensity are 0.0303, 0.0247 and 0.508 with standard deviations 0.0283, 0.0295 and 0.598, respectively. The averages of their long-term mean levels are 0.0709, 0.0504 and 1.88 with standard deviations 0.0245, 0.0395 and 1.77, respectively. The other optimal parameters include $\rho = -0.836$ from the correlation between the SPX returns and the changes of the 30-day VIX square, and $\kappa = 1.96$, $x = -10.6\%$, $\sigma_v = 0.983$, $\sigma_\lambda = 9.748$, $\sigma_1 = 0.173$, $\sigma_2 = 5.269$ from the sequential estimation using the term structure data of the VIX and SKEW.



# Chapter 4

# Jump Risk: A Cubic-Variation Approach

This chapter is a joint work with Jin E. Zhang. Its earlier version was presented at the 2016 Auckland Finance Meeting, 16–18 December 2016, AUT, Auckland, New Zealand.

## 4.1 Introduction

Jump risk is an important risk factor. It signals instability and crashes of financial markets. However, jump-identification methods are not unified in the literature, and the understanding of the dynamics of jumps is far from complete. We propose a simple and intuitive approach to measure jump magnitude and detect jump existence, and we forecast these in the SPX index with option-implied and lagged time-series information.

Jumps play an important role in financial modelling; see, among others, Press (1967), Merton (1976), Jarrow and Rosenfeld (1984), Pan (2002), Yan (2011), Chen, Joslin and Tran (2012) and Branger, Kraft and Meinerding (2016). The discontinuous jumps are different from diffusion in risk management and asset allocation; see, for instance, Liu, Longstaff and Pan (2003) and Liu, Pan and Wang (2005). Therefore, disentangling jumps from diffusion is a necessity for decision makers. In the literature, the jump component can be identified through different methods. Barndorff-Nielsen and Shephard (2004) use the standardized difference between the realized variance and the realized bi-power variation. Jiang and Oomen (2008) use

the standardized difference between the swap variance and realized variance. Lee and Mykland (2008) use the standardized high-frequency returns. In contrast, we propose a simpler way to detect jumps by using the sum of cubed daily returns, that is, realized cubic variation.[1]

The option market reflects participants' expectations about the underlying stock market, and option-implied volatility has shown forecasting ability superior to that of historical volatility; see, for instance, Christensen and Prabhala (1998) and Jiang and Tian (2005). Since the launch of SKEW in 2011 and VIX in 1993, such public option-implied information has been easily obtainable from the Chicago Board Options Exchange (CBOE) website. SKEW and VIX track the risk-neutral skewness and volatility of the future 30-day returns of SPX, respectively. Zhen and Zhang (2014) show that SKEW contains both jump and variance information. Hence, in this chapter, we use the option-implied third cumulant extracted from SKEW and VIX to forecast the future realized cubic variation of the SPX returns, and compare its informational efficiency with the past realized cubic variation.[2] The past diffusive variance is used to control the impact of the return-correlated variance embedded in the implied third cumulant. We find that option-implied information coupled with past diffusive variance is more efficient in forecasting future the SPX jump magnitude, measured by realized cubic variation over a future month using daily returns. We also find that the realized cubic variation is negatively correlated with its risk-neutral expectation. Furthermore, we detect monthly jump existence if the realized cubic variation exceeds a given threshold, and explore its property using both option-implied and time-series information. We find that the past realized variance shows the best performance in forecasting future jump existence likelihood.

Our results shed some light on the specifications of modelling SPX jump risks, including jump size, jump intensity and the existence of jumps in variance. The

[1]The cubic variation with respect to a given partition is the sum of absolute cubed returns $\sum_{i=1}^{n} |\delta R_{t_i}|^3$ (see Nourdin, 2008), where $\delta R_{t_i}$ denotes the continuously compounded return. However, in finance, the cubic variation without the absolute value is more meaningful. This is because the limit of its expectation is the third cumulant of the long-horizon returns with independent and identically distributed sub-period returns when the norm of partition approaches zero. Thus, in this chapter, we drop the absolute value and use the sum of cubed returns as the realized cubic variation.

[2]The $n$th cumulant $\kappa_n$ of a random variable $X$ can be derived through the cumulant-generating function $K(t)$, which is the natural logarithm of the moment-generating function, that is, $K(t) = \log E(e^{tX}) = \sum_{n=1}^{\infty} \kappa_n \frac{t^n}{n!}$. Only the second and third cumulants are the same as the second and third central moments, respectively.

change of the price-jump-size distributions under the physical and risk-neutral measures explains the negative correlation between the realized cubic variation and its risk-neutral expectation. The price and variance co-jumps make the past diffusive variance contain information for future jump existence likelihood. Our empirical findings suggest that the jump intensity is time-varying, and there exist variance jumps which happen contemporaneously with price jumps.

The remainder of this chapter is organized as follows. Section 4.2 defines jump magnitude and proposes a jump-detecting approach. Section 4.3 shows the properties of the realized cubic variation in a parsimonious jump-diffusion model. Section 4.4 discusses the empirical methodology. Section 4.5 presents the results. Section 4.6 concludes. Proofs are provided in the Appendix.

## 4.2 Jump Magnitude and Jump Existence

In the literature, Bandi and Renò (2016) adopt a jump-detecting approach with the same logic as that proposed by Lee and Mykland (2008), Jiang and Yao (2013) use the variance swap approach developed in Jiang and Oomen (2008) for jump identification, and Todorov (2010) applies Barndorff-Nielsen and Shephard's (2006) method to separate jump variance from diffusive variance. These articles either use return or difference of variance measured in different ways to capture jumps. However, in a jump-diffusion setup, both return and variance contain diffusion terms. By raising the power of moments to three, in this chapter, we propose a new jump-risk barometer based on the sum of cubed returns, which solely captures the jump term.

Given a partition $\mathcal{P}$: $t = t_0 < t_1 < \cdots < t_n = T$ of the interval $[t, T]$, the annualized realized cubic variation, $RCV$, is defined as

$$RCV_{t,T} = \frac{1}{T-t}\sum_{i=0}^{n-1}(\delta R_{t_i})^3, \tag{4.1}$$

where $t$ is the starting time, $T$ is the ending time, $n$ is the number of sub-intervals of the partition, $R_{t_i} = \ln\frac{S_{t_i}}{S_{t_0}}$ denotes the continuously compounded return from time $t_0$ to time $t_i$ ($i = 1, 2, \cdots, n$), $\delta$ denotes the difference operator in a sub-interval,

and $\delta R_{t_i}$ denotes the continuously compounded return over $[t_i, t_{i+1}]$.[3] Suppose the partition is equally divided, then the length of each sub-interval is given by $h = \frac{T-t}{n}$. The $RCV$ is a natural filter for jumps, as the diffusion term is of order $h^{3/2}$ and the jump term is of order $h$ in jump-diffusion models, which will be used in the following section. Thus, we use $RCV$ as a measure of *jump magnitude*.

We apply our jump-risk barometer $RCV$ defined in Equation (4.1) to SPX to capture the sum of its daily jumps in one month. The daily close SPX data are obtained from the Bloomberg Professional Service. The sample period is from 2 January 1990 to 31 December 2015. Table 4.1 reports the monthly $RCV$ values that are greater than five basis points ($5/10^4$). The monthly $RCV$ (or the analogous future term $RCV^f$) is computed with the daily data over the past (future) month (30 calendar days, which comprises 22 trading days at most) on the fourth Wednesday, which has fewest holidays among all weekdays, and the calculation period is matched with the 30-day SKEW and VIX. We choose the third trading day following the monthly option expirations (third Friday) to mitigate the interpolation errors in the 30-day VIX and SKEW indices. If SKEW or VIX is not available on the fourth Wednesday, we first use the following Thursday instead, then the preceding Tuesday and then the nearest trading day when both SKEW and VIX are available.

Furthermore, noting that jumps in prices are accompanied by an extremely large $RCV$ value in magnitude, we identify the *jump existence* in prices if the absolute value of $RCV$ is greater than five basis points, that is,

$$J_t = \mathbb{1}_{|RCV_t|>5/10^4}, \tag{4.2}$$

where $\mathbb{1}$ denotes an indicator function.[4] We determine the cutoff value by the rule of thumb. Suppose that the volatility is 0.2, and the daily increment of a Brownian motion is $\sqrt{1/252}$ and keeps positive (or negative) for the whole year, then $|RCV| = 252 \times (0.2/\sqrt{252})^3 \approx 5/10^4$. However, it is extremely unlikely for the increment of a Brownian motion to stay positive (or negative) for a whole month. Thus, if there is

[3] As the calculation period is fixed in this chapter, we omit one subscript, for example, $RCV_t$ or $RCV_t^f$, where the former is computed with daily returns over the past month and the latter is over the future month.

[4] If $RCV$ is calculated with future returns, then we put a superscript on $RCV$ and $J$, which are $RCV^f$ and $J^f$, respectively.

no jumps in prices, the $RCV$ is not likely to reach five basis points. After applying our jump-detecting approach, we identify 39 jump months ($12.5\%$) out of 312 total months.

Table 4.1 reports the jump months with extreme $RCV$ values and the corresponding financial events. We split the jump months into six groups: 1990–1991, 1997–1998, late 1999–2002, 2007–2008, 2011 and 2015, which correspond to early 1990s recession, Asian financial crisis, dot-com bubble burst, 2008 sub-prime mortgage crisis, 2011 stock market fall and 2015 stock market selloff, respectively. Overall, the jump size could be positive or negative. We identified three and four jump months for the first two groups, respectively. There are 13 jumps during the dot-com bubble, including the peak in March 2000, the 9/11 attacks in 2001 and the stock market downturn in 2002. During the 2008 financial crisis, SPX jumped four times before the consecutive nine-month run of monthly jumps starting from the bankruptcy of large financial institutions, for example, the collapse of Lehman Brothers on 15 September 2008. SPX also jumped for five consecutive months starting from late July 2011 triggered by the European Debt Crisis and only jumped once in 2015.

Lee and Mykland (2008) use the standardized high-frequency return as the statistic to test jumps. However, their method is not applicable in low-frequency data (e.g., daily data), especially when the return is not standardized by its volatility. Barndorff-Nielsen and Shephard (2006) construct the jump-testing statistic as the standardized difference of realized variance and bi-power variation. Nevertheless, the difference of realized variance and bi-power variation is not always positive when using daily data, which indicates a contradicting fact that the variance caused by jumps is occasionally negative. Jiang and Ooman's (2008) statistic, computed as the standardized difference of swap variance and realized variance, is the most similar one to ours. But their test exploits the third and higher-order moments of asset returns, whereas our jump-risk barometer only concentrates on the cubed returns. Hence, it is more intuitive to determine the threshold for jump existence when using our method.

## 4.3 A Jump-Diffusion Model

The Poisson mixture of normal distributions has been widely used in the literature to model security prices (see, e.g., Press, 1967; Merton, 1976; Jarrow and Rosenfeld, 1984; Pan, 2002; Yan, 2011; Chen, Joslin and Tran 2012; Branger, Kraft and Meinerding, 2016). In this section, we adopt a jump-diffusion model to explain the meaning of the realized cubic variation $RCV$ defined in Equation (4.1).

In a model with stochastic jump intensity and co-jumps in prices and variance, under the physical measure $P$, assume the price of the underlying asset follows a jump diffusion process, and the diffusive variance and jump intensity follow mean-reverting processes modelled by

$$\frac{dS_t}{S_t} = \mu_t dt + \sqrt{v_t} dB_t^S + (e^x - 1)dN_t - E(e^x - 1)\lambda_t dt, \tag{4.3}$$

$$dv_t = \kappa_1(\theta_{1,t} - v_t)dt + \sigma_v \sqrt{v_t} dB_t^v + y dN_t - \mu_y \lambda_t dt, \tag{4.4}$$

$$d\lambda_t = \kappa_2(\theta_2 - \lambda_t)dt + \sigma_\lambda \sqrt{\lambda_t} dB_t^\lambda, \tag{4.5}$$

where $S_t$ is the price of the underlying asset, $\mu_t$ is the risky rate of return, $\kappa_1$ and $\kappa_2$ denote the mean-reverting speeds of the diffusive variance $v_t$ and jump intensity $\lambda_t$, respectively, $\theta_{1,t}$ and $\theta_2$ denote their long-term mean levels, respectively, $B_t^S$ and $B_t^v$ denote two Brownian motions with the correlation coefficient $corr(dB_t^S, dB_t^v) = \rho$, $N_t$ is a Poisson counter with stochastic intensity $\lambda_t$: Prob$(dN_t = 1) = \lambda_t dt$, Prob$(dN_t = 0) = 1 - \lambda_t dt$, and given the arrival of a jump, the price jump size $x$ is normally distributed with mean $\mu_x$ and variance $\sigma_x^2$, and the variance jump size $y$ is exponentially distributed with mean $\mu_y$. The risky rate of return is given by

$$\mu_t = r + \eta^s v_t + [E(e^x - 1)\lambda_t - E^Q(e^x - 1)\lambda_t^Q], \tag{4.6}$$

where the superscript $Q$ denotes the risk-neutral measure, $E^Q$ means the expectation in $Q$-measure, $\lambda_t^Q$ is the jump intensity in $Q$-measure, $r$ denotes the risk-free rate, $\eta^s v_t$ is the risk premium associated with the diffusive shock in prices, and $E(e^x - 1)\lambda_t - E^Q(e^x - 1)\lambda_t^Q$ is the risk premium associated with jumps in prices. The long-term

mean level of diffusive variance is given by

$$\theta_{1,t} = \theta_1 + \frac{\mu_y}{\kappa_1}(\lambda_t - \lambda_t^Q). \tag{4.7}$$

For simplicity, we assume that $x$ and $y$ are independent. The discontinuous Poisson-driven jump component is independent of the continuous diffusive terms, and the Brownian motion $B_t^\lambda$ is independent of $B_t^S$ and $B_t^v$. The random jump sizes $x$ and $y$ are assumed to be independent of $B_t^S$, $B_t^v$, $B_t^\lambda$ and $N_t$, and independent across different jump times. The co-jumps in prices and variance have been broadly accepted in the literature (see, e.g., Eraker, Johannes and Polson 2003; Broadie, Chernov and Johannes, 2007; Bandi and Renò, 2016). The model described by Equations (4.3)-(4.5) belongs to the affine jump-diffusion class in Duffie, Pan and Singleton (2000). In particular, we set the jump intensity $\lambda_t$ to be stochastic to accommodate its time-varying feature.

**Proposition 4.1** *Under the model setup described in Equations (4.3)-(4.5), as the norm of the partition approaches zero, in $P$-measure, the time-$t$ conditional expectation of the realized cubic variation defined in Equation (4.1) is given by*

$$\lim_{||\mathcal{P}||\to 0} E_t(RCV_{t,T}) = E(x^3)\frac{1}{T-t}\int_t^T E_t(\lambda_u)du, \tag{4.8}$$

*where $E(x^3) = 3\mu_x\sigma_x^2 + \mu_x^3$.*

*Proof*. See Appendix 4.7.1.

To link the physical processes with the risk-neutral ones, we propose a pricing kernel, $\pi$, which is a combination of the pricing kernels adopted by Pan (2002) and Liu, Pan and Wang (2005), of the following form:

$$\pi_t = e^{-rt}\pi_t^D\pi_t^J, \tag{4.9}$$

where $\pi_t^D$ and $\pi_t^J$ are given by

$$\pi_t^D = e^{-\int_0^t \zeta_s^{(1)}dW_s^{(1)} - \frac{1}{2}\int_0^t(\zeta_s^{(1)})^2ds - \int_0^t \zeta_s^{(2)}dW_s^{(2)} - \frac{1}{2}\int_0^t(\zeta_s^{(2)})^2ds - \int_0^t \zeta_s^{(3)}dW_s^{(3)} - \frac{1}{2}\int_0^t(\zeta_s^{(3)})^2ds}, \tag{4.10}$$

$$\pi_t^J = e^{\int_0^t(g+hx-h\mu_x-\frac{1}{2}h^2\sigma_x^2)dN_s-(e^g-1)\int_0^t\lambda_s ds}, \tag{4.11}$$

$W_t^{(1)}$, $W_t^{(2)}$ and $W_t^{(3)}$ denote three independent Wiener processes, given by

$$W_t^{(1)} = B_t^S, \quad W_t^{(2)} = \frac{B_t^v - \rho B_t^S}{\sqrt{1-\rho^2}}, \quad W_t^{(3)} = B_t^\lambda, \tag{4.12}$$

the market prices $\zeta$ of diffusive shocks in prices, volatility and jump intensity are defined as

$$\zeta_t^{(1)} = \eta^s \sqrt{v_t}, \quad \zeta_t^{(2)} = -\frac{\rho\eta^s + \frac{\eta^v}{\sigma_v}}{\sqrt{1-\rho^2}}\sqrt{v_t}, \quad \zeta_t^{(3)} = -\frac{\eta^\lambda}{\sigma_\lambda}\sqrt{\lambda_t}. \tag{4.13}$$

**Proposition 4.2** *The state price density $\pi_t$ defined in Equations (4.9)-(4.13) converts the physical processes described by Equations (4.3)-(4.5) to the risk-neutral processes, given in the following equations:*

$$\frac{dS_t}{S_t} = rdt + \sqrt{v_t}dB_t^S(Q) + (e^x - 1)dN_t - E^Q(e^x - 1)\lambda_t^Q dt, \tag{4.14}$$

$$dv_t = \kappa_1^Q(\theta_1^Q - v_t)dt + \sigma_v\sqrt{v_t}dB_t^v(Q) + ydN_t - \mu_y\lambda_t^Q dt, \tag{4.15}$$

$$d\lambda_t^Q = \kappa_2^Q(\theta_2^Q - \lambda_t^Q)dt + \sigma_\lambda^Q\sqrt{\lambda_t^Q}dB_t^\lambda(Q), \tag{4.16}$$

*where $\kappa_1^Q = \kappa_1 - \eta^v$, $\theta_1^Q = \frac{\kappa_1\theta_1}{\kappa_1^Q}$, $\lambda_t^Q = e^g\lambda_t$, $\kappa_2^Q = \kappa_2 - \eta^\lambda$, $\theta_2^Q = e^g\frac{\kappa_2\theta_2}{\kappa_2^Q}$, $\sigma_\lambda^Q = e^{g/2}\sigma_\lambda$. Under the risk-neutral measure $Q$, the price jump size $x$ is normally distributed with mean $\mu_x + h\sigma_x^2$ and variance $\sigma_x^2$, the distribution of the variance jump size $y$ is not changed, and the three Brownian motions $B_t^S(Q)$, $B_t^v(Q)$ and $B_t^\lambda(Q)$ are given by*

$$dB_t^S(Q) = dW_t^{(1)} + \zeta_t^{(1)}dt, \tag{4.17}$$

$$dB_t^v(Q) = \rho(dW_t^{(1)} + \zeta_t^{(1)}dt) + \sqrt{1-\rho^2}(dW_t^{(2)} + \zeta_t^{(2)}dt), \tag{4.18}$$

$$dB_t^\lambda(Q) = dW_t^{(3)} + \zeta_t^{(3)}dt. \tag{4.19}$$

*Proof.* See Appendix 4.7.2.

**Proposition 4.3** *Under the model setup in Equations (4.14)-(4.16), when the time to maturity $T - t$ is small and $R_T = \ln\frac{S_T}{S_t}$, the annualized implied third cumulant $ITC_{t,T}$, defined*

*as*

$$ITC_t = \frac{1}{T-t} E_t^Q[R_T - E_t^Q(R_T)]^3, \tag{4.20}$$

*is approximated as shown in the following equation:*

$$ITC_{t,T} \approx \int_t^T \frac{E^Q(x^3)E_t^Q(\lambda_u^Q) + 3A_1(\kappa_1^Q, u)[\rho\sigma_v E_t^Q(v_u) + E^Q(xy)E_t^Q(\lambda_u^Q)]}{T-t} du, \tag{4.21}$$

*where* $A_1(\kappa_1^Q, u) = \frac{1-e^{-\kappa_1^Q(T-u)}}{\kappa_1^Q}$.

*Proof.* See Appendix 4.7.3.[5]

**Remark 4.3.1.** Equation (4.21) shows that the implied third cumulant is mainly determined by three components: the jumps in prices, $\int_t^T \frac{E^Q(x^3)E_t^Q(\lambda_u^Q)}{T-t} du$, the return-correlated variance, $3\rho\sigma_v \int_t^T \frac{A_1(\kappa_1^Q,u)E_t^Q(v_u)}{T-t} du$, and the co-jumps in prices and variance, $3E^Q(xy) \int_t^T \frac{A_1(\kappa_1^Q,u)E_t^Q(\lambda_u^Q)}{T-t} du$. The price jump component is also the risk-neutral expectation of the realized cubic variation.

**Remark 4.3.2.** Since the physical processes Equations (4.3)-(4.5) and risk-neutral processes Equations (4.14)-(4.16) are of the same structure with different parameters, the risk-neutral expected cubic variation is similar to that in Equation (4.8). We assume that the daily partition is sufficient to measure the limit of the expected $RCV$. For brevity, we set

$$\Psi^P = E(x^3), \quad \Psi^Q = E^Q(x^3), \quad \Phi^Q = \mu_x^Q \mu_y, \quad \mu_x^Q = \mu_x + h\sigma_x^2,$$
$$\Lambda_t^P = \frac{1}{T-t}\int_t^T E_t(\lambda_u)du = \omega_1^P \lambda_t + (1-\omega_1^P)\theta_2, \quad \omega_1^P = \frac{1-e^{\kappa_2(T-t)}}{\kappa_2(T-t)},$$
$$\Lambda_t^Q = \frac{1}{T-t}\int_t^T E_t^Q(\lambda_u^Q)du = \omega_1^Q \lambda_t^Q + (1-\omega_1^Q)\theta_2^Q, \quad \omega_1^Q = \frac{1-e^{\kappa_2^Q(T-t)}}{\kappa_2^Q(T-t)},$$
$$\Gamma_t^Q = 3\int_t^T \frac{1-e^{-\kappa_1^Q(T-u)}}{\kappa_1^Q(T-t)} E_t^Q(\lambda_u^Q)du = \omega_2^Q \lambda_t^Q + \omega_3^Q \theta_2^Q,$$
$$\omega_2^Q = 3\int_t^T \frac{1-e^{-\kappa_1^Q(T-u)}}{\kappa_1^Q(T-t)} e^{-\kappa_2^Q(u-t)}du, \quad \omega_3^Q = 3\int_t^T \frac{1-e^{-\kappa_1^Q(T-u)}}{\kappa_1^Q(T-t)}(1-e^{-\kappa_2^Q(u-t)})du,$$

[5]The exact formula is also provided in the Appendix.

then the covariance of the realized cubic variation and its risk-neutral expectation is given by

$$cov(RCV^f_{t,T}, \Psi^Q \Lambda^Q_t) = cov(\Psi^P \Lambda^P_t, \Psi^Q \Lambda^Q_t) = \Psi^P \Psi^Q e^g \omega^P_1 \omega^Q_1 var(\lambda_t),$$

the covariance of the realized cubic variation and the co-jump component in the third cumulant $ITC$ is given by

$$cov(RCV^f_{t,T}, \Phi^Q \Gamma^Q_t) = cov(\Psi^P \Lambda^P_t, \Phi^Q \Gamma^Q_t) = \Psi^P \Phi^Q e^g \omega^P_1 \omega^Q_2 var(\lambda_t),$$

where every element in $\{\mu_y, e^g, \omega^P_1, \omega^Q_1, \omega^Q_2, var(\lambda_t)\}$ is positive, $\Psi^P = \mu_x(3\sigma^2_x + \mu^2_x)$, $\Psi^Q = \mu^Q_x[3(\sigma_x)^2 + (\mu^Q_x)^2]$, and the covariance of the realized cubic variation and the variance component in $ITC$ is zero. Note that $var(\lambda_t)$ is positive only when $\lambda_t$ is stochastic. Therefore, the sign of the correlation between the future realized cubic variation $RCV^f$ and the implied third cumulant $ITC$ is the same as that of $RCV^f$ and its risk-neutral expectation. Both of them are determined by the signs of the physical and risk-neutral expectations of the price jump size $x$, which are $\mu_x$ and $\mu_x + h\sigma^2_x$, respectively. Note that the price-jump-size distributions changed under the physical and risk-neutral measures, and the difference of its expectations is $h\sigma^2_x$.

## 4.4 Methodology

As option-implied information reflects market participants' expectation about future returns' movements. It is widely believed to be more efficient than historical time-series information in forecasting; see, for instance, Christensen and Prabhala (1998) and Jiang and Tian (2005) on forecasting future realized volatility. Therefore, in this chapter, we outline our approach to forecasting the future jump magnitude and the future jump existence likelihood by using option-implied information as compared with past time-series information.

### 4.4.1 Predictor Variables

Zhen and Zhang (2014) show that the CBOE SKEW is a nested outcome of jumps and return-correlated variance and it could be useful in forecasting future jump

magnitude. We extract the annualized option-implied third cumulant, defined in Equation (4.20), from the CBOE VIX and SKEW indices by using the following approximation formula

$$ITC_t \approx \frac{1}{T-t}\frac{SKEW_{t,T}-100}{-10}\Big[\Big(\frac{VIX_{t,T}}{100}\Big)^2(T-t)\Big]^{3/2}, \tag{4.22}$$

where the risk-neutral variance is approximated by scaled VIX square, that is,

$$RNV_t \approx \Big(\frac{VIX_{t,T}}{100}\Big)^2, \tag{4.23}$$

time $t$ is the current date, $T$ is the expiry date and the time to maturity $T-t=\frac{1}{12}$ is fixed. The CBOE SKEW is the scaled skewness of the risk-neutral distribution of the SPX returns, defined as,

$$SKEW_{t,T} = 100 - 10 \times Sk_{t,T}, \tag{4.24}$$

where $Sk_{t,T}$ denotes the risk-neutral skewness, which is a dimensionless statistic and given by the third cumulant divided by the variance to the power of $\frac{3}{2}$, that is,

$$Sk_{t,T} = \frac{E_t^Q[R_T - E_t^Q(R_T)]^3}{[E_t^Q[R_T - E_t^Q(R_T)]^2]^{3/2}}. \tag{4.25}$$

The VIX is defined as

$$VIX_{t,T} = 100\sqrt{\frac{E_t^Q[\sum^{\mathcal{P}} 2(e^{\delta R}-1-\delta R)]}{T-t}}, \tag{4.26}$$

where $\mathcal{P}$ denotes a partition of the interval $[t,T]$, and $Q$ denotes the risk-neutral measure. In $Q$-measure, Zhang et al. (2017) show that the main contribution of the risk-neutral variance, $E_t^Q[R_T - E_t^Q(R_T)]^2$, where $R_T = \sum^{\mathcal{P}}(\delta R)$, comes from the variance swap rate, $E_t^Q[\sum^{\mathcal{P}}(\delta R)^2]$, which is the risk-neutral expected value of the floating leg of a variance swap. Carr and Wu (2009) use the scaled VIX square to approximate the variance swap rate and state that the approximation error is of the order of cubed returns, which is negligible compared with squared returns. Hence, the difference between the risk-neutral variance and the scaled VIX square

is neglected when calculating the option-implied third cumulant.

We obtain the 30-day VIX and SKEW indices data from the CBOE website. Figure 4.1 shows the dynamics of $SK$ and $ITC$ for the sample period 2 January 1990 to 31 December 2015. As the skewness is erratic over the entire sample period, whereas there are several spikes in $ITC$, one could not visually tell the 2008 financial crisis from $Sk$ but rather from $ITC$. These spikes in $ITC$ correspond to the early 1990s recession, 1997-1998 Asian crisis, dot-com bubble, 2008 financial crisis, 2010 flash crash, 2011 stock market fall and 2015 stock market selloff (see Table 4.2). The large price jumps are usually accompanied with volatility jumps (see Bandi and Renò, 2016). This phenomenon could make the skewness rather small during a financial crisis when the jump components in the third cumulant, which are shown in Remark 4.3.1, is diluted by the jumps in variance, and consequently leads skewness to be a poor indicator of market jump risk. Hence, we choose $ITC$ rather than $SK$ as a predictor variable.

Nevertheless, the implied third cumulant $ITC_t$ in Equation (4.20) is a noisy predictor for the realized cubic variation, $RCV_t$, which is caused solely by jumps in prices.[6] To control the impact of the return-correlated diffusive variance, we use the realized bi-power variation proposed by Barndorff-Nielsen and Shephard (2004), given by

$$RBV_t = \frac{\pi}{2(t-s)} \sum_{i=0}^{n-2} |\delta R_{t_i}||\delta R_{t_{i+1}}|, \tag{4.27}$$

where $n$ is the number of sub-intervals of the partition of $[s, t]$, $s$ is a past time and $t$ is the current time. Barndorff-Nielsen and Shephard (2004) show that the realized bi-power variation $RBV_t$ converges in probability to the annualized integrated diffusive variance and is robust to rare price jumps, whose impact is contaminated in

[6]This fact can be seen in Neuberger (2012). For simplicity, we assume that the logarithmic price $\ln S_u$ is a martingale. The third cumulant comprises two components: the sums of cubed returns and the covariances of returns and changes of conditional variance, that is, $E_t[R_T^3] = E_t[\sum_{u=t}^{T}((\delta R_u)^3 + 3\delta R_u \delta V_u)]$, where $V_u = var_u(R_T)$ is the time-$u$ conditional variance of $R_T$. The realized variance is also an unbiased estimator of the second cumulant, that is, $E_t[\sum_{u=t}^{T}(\delta R_u)^2] = E_t[R_T^2]$. See Zhen and Zhang (2014) for the case when $R_T$ is not a martingale.

the risk-neutral variance defined in Equation (4.23) and the realized variance defined as

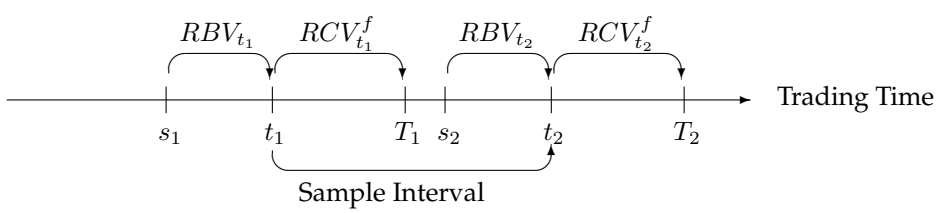
$$RV_t = \frac{1}{(t-s)} \sum_{i=0}^{n-1} (\delta R_{t_i})^2. \tag{4.28}$$

Thus, we use $RBV$, rather than the other two variance proxies, $RNV$ and $RV$, as a control variable.

### 4.4.2 Regression Analyses

We use the future realized cubic variation $RCV^f$ as a barometer for the market jump risk, as it captures only the jumps in prices. The predictor variables are the implied third cumulant $ITC$ defined in Equation (4.22) and the realized bi-power variation $RBV$ defined in Equation (4.27). The calculation timeline for $RBV$ and $RCV^f$ is shown in the following figure. The calculation period for $RBV$ is the past 30 calendar days, whereas $RCV^f$ is calculated with the future 30 calendar days.

This graph shows the calculation timeline for the realized cubic variation $RCV_t$ in Equation (4.1) and the realized bi-power variation $RBV_t$ in Equation (4.27). Let $t_1$ and $t_2$ denote two consecutive monthly observation dates. The realized cubic variation $RCV^f_{t_1}$ ($RCV^f_{t_2}$) is calculated with future returns until time $T_1$ ($T_2$), and the realized bi-power variation $RBV_{t_1}$ ($RBV_{t_2}$) is constructed with past absolute returns starting from time $s_1$ ($s_2$).




The following regression is used to investigate the forecasting power of $ITC$:

$$RCV_t^f = \alpha + \beta_1 ITC_t + \beta_2 RBV_t + \varepsilon_t, \tag{4.29}$$

where $RBV$ controls for the diffusive variance term embedded in $ITC$. To deal with the multi-collinearity in Equation (4.29), we can first regress $ITC$ on $RBV$, extract the residuals to be $ITCO$ (orthogonal $ITC$), and then replace $ITC$ with $ITCO$. However, this orthogonalization does not help, as the estimated $ITC$ slope and residuals in Equation (4.29) are not changed (Mitchell, 1991).[7] For comparison, we also run the univariate regression using $ITC$ or $RBV$. The realized variance ($RV$) and risk-neutral variance ($RNV$) are used as alternatives to $RBV$ for a robustness check, and one-month lagged $RCV$ is used to replace $ITC$ to explore the mean-reverting behaviour of the jump magnitude.

Furthermore, we use a monotonic transform of the implied third cumulant $ITC$ as a predictor variable to forecast the future jump existence likelihood. The response variable is binary with two outcomes, zero or more than one jump, and the number of jumps is assumed to follow a Poisson distribution; thus we use the binomial regression to forecast the likelihood of the jump existence defined in Equation (4.2) with the complementary log-log link function as follows:

$$P(J_t^f = 1|ITC_t) = \phi_B(\alpha + \beta \ln(|ITC_t|)), \quad \phi_B(x) = 1 - e^{-e^x}, \tag{4.30}$$

where $e^x$ represents jump intensity. Note that we implicitly assume that the jump intensity is time-varying by using this regression. The difference between our binomial regression and the logistic regression lies in the link function. The latter has been used by Kumar, Moorthy and Perraudin (2003) and Berger and Pukthuanthong (2012). Nevertheless, the complementary log-log link function is more appropriate in this chapter as the occurrence of jumps is assumed to be controlled by a Poisson process. The difference between Equations (4.29) and (4.30) is that the former focuses on the total jump effect which is a compound outcome of jump size and jump intensity, whereas only the jump intensity matters in the latter regression. We also replace $ITC$ with the realized bi-power variation $RBV$, the realized variance $RV$, the risk-neutral variance $RNV$, the lagged cubic variation $RCV$ and the lagged jump existence $1 + J$ to compare their predictive performances.

[7] This fact could also be seen in Section 7 in Fama and French (2015).

## 4.5 Empirical Results

Table 4.3 provides the summary statistics of the realized cubic variation ($RCV$, $RCV^f$), the implied third cumulant ($ITC$), the realized bi-power variation ($RBV$), the realized variance ($RV$) and the risk-neutral variance ($RNV$). All of these variables are stationary, as indicated by an augmented Dickey-Fuller test. The statistics of $RCV$ and $RCV^f$ are similar. The average future $RCV^f$ only accounts for 1.4% of the average $ITC$. The small proportion confirms that $ITC$ contains a diffusive variance term, and also indicates that the change of the price-jump-size distributions under the physical and risk-neutral measures might be large in magnitude. The average $RBV$ of 0.0285 is less than the average $RV$ of 0.0324 because the former is robust to the jumps and simply captures the integrated diffusive variance. The high average $RNV$ of 0.0440 implies the existence of the variance risk premium. The implied third cumulant $ITC$ and $RCV^f$ are also negatively correlated, which confirms the existence of the change of price-jump-size distributions across the physical and risk-neutral measures. Moreover, $ITC$, which includes both jumps and variance terms, is negatively correlated with the three different measures of variance ($RBV$, $RV$ and $RNV$) due to the leverage effect, which is the negative correlation between returns and changes of diffusive variance.

Table 4.4 reports the forecasting performance of both univariate and bivariate linear regressions. The baseline regression in Equation (4.29) shows the highest predictive power with a R-squared of 25.3%, followed by the univariate regression of one-month lagged $RCV$ with a R-squared of 9.3%. This result shows that option-implied information is indeed more efficient than historical time-series information in forecasting future jump magnitude, as measured by the future realized cubic variation. We split the sample into two sub-periods, from January 1990 to December 2002, which includes the Asian financial crisis and dot-com bubble, and from January 2003 to December 2015, which includes the 2008 financial crisis. Our empirical results show that the predictive power mainly comes from the second sub-sample with a R-squared of 38.6%. This result is not surprising, as the 2008 financial crisis comprises nine consecutive jump months in our sample and it is broadly considered to be the worst financial crisis since the Great Depression of the 1930s. As jumps are rare events, we split the sample into only two sub-periods, which comprise 20

and 19 jump months, respectively. The findings regarding informational efficiency do not alter in the second period. However, for the first period, the time-series information $RCV$ shows a mild predictive power with a R-squared of 3.1%. The univariate regression of $ITC$ exhibits a lower R-squared of 1.8%, due to the noisy component caused by the return-correlated variance. The univariate regression of $RBV$ also exhibits a minor predictive power with a R-squared of 3.7%. In addition, the significantly negative slope of $ITC$ in the baseline regression specified in Equation (4.29) indicates that the change of the price-jump-size distributions induces the opposite signs of the physical and risk-neutral expectations of the price jump size. The predicted and observed values in the baseline regression are shown in Figure 4.2.

Table 4.5 presents the results of the binomial regressions with the complementary log-log link function. Three pseudo R-squared values, Efron, Macfadden and Adjusted Count, are reported for the goodness-of-fit evaluation. The realized variance ($RV$) shows the best forecasting performance with an Adjusted Count R-squared of 51.28%, whereas the one-month lagged jump indicator exhibits the worst performance with zero Adjusted Count R-squared. These results suggest that past realized variance information outperforms option-implied information in forecasting future jump existence likelihood. The realized bi-power variation also exhibits some predictive power with an Adjusted Count R-squared of 25.64%, which implies the existence of co-jumps in prices and variance so that the diffusive variance contains price jump information. Moreover, the significantly positive coefficient $\beta$ implies that an increase of volatility or absolute cubic variation signals a higher possibility of market jumps, that is, a higher jump intensity. The predicted and observed values in the binomial regression specified in Equation (4.30) against the predictor $\ln(|ITC|)$ (upper panel) and against the observation date (lower panel) are shown in Figure 4.3.

## 4.6 Conclusion

In this chapter, we measure the SPX jump magnitude by the sum of cubed daily returns, that is, realized cubic variation, and forecast it with the option-implied third

cumulant and lagged time-series information. We find that option-implied information coupled with past diffusive variance is more efficient in forecasting future jump magnitude than is the time-series information. The change of the price-jump-size distributions under the physical and risk-neutral measures induces a negative relation between the realized cubic variation and its risk-neutral expectation. Moreover, we detect the existence of jumps when the realized cubic variation is larger than a given threshold in magnitude. Both time-series and option-implied data have predictive power and the former outperforms the latter in forecasting future jump existence likelihood. Our empirical findings suggest that the jump intensity is time-varying, and there exist variance jumps which happen contemporaneously with price jumps. Our jump-detecting approach could be easily applied to individual stocks or other financial markets. Further research could look at the intraday pattern of the realized cubic variation.

## 4.7 Appendix

### 4.7.1 Proof of Proposition 4.1

Supposing the partition is equidistant, the unit length is $h$ and the number of observations is $n$, we split the time-$t_i$ unit continuously compounded return into three parts, that is,

$$\delta R_{t_i} = A_{t_i} + M^c_{t_i} + M^d_{t_i},$$

where the drift is $A_{t_i} = \mu_{t_i} h$, and the continuous and discontinuous martingales are respectively given by

$$\begin{aligned} M^c_{t_i} &= \sqrt{v_{t_i}} \delta B^S_{t_i}, \quad \delta B^S_{t_i} = B^S_{t_i+h} - B^S_{t_i}, \\ M^d_{t_i} &= x \delta N_{t_i} - \mu_x \lambda_{t_i} h, \quad \delta N_{t_i} = N_{t_i+h} - N_{t_i}. \end{aligned}$$

In $P$-measure, the time-$t_i$ conditional third cumulants of the terms related with $M^c_{t_i}$ are zero, that is,

$$E_{t_i}[(M^c_{t_i})^3] = E_{t_i}[(M^c_{t_i})^2 M^d_{t_i}] = E_{t_i}[M^c_{t_i}(M^d_{t_i})^2] = 0,$$

and the conditional third cumulant of $\{M^d_{t_i}\}$ is as follows

$$\begin{aligned}
&E_{t_i}[(M^d_{t_i})^3] = E_{t_i}[(x\delta N_{t_i} - \mu_x \lambda_{t_i} h)^3] \\
&= E_{t_i} E_{t_i}\{[(x-\mu_x)\delta N_t + \mu_x(\delta N_{t_i} - \lambda_{t_i} h)]^3 | \delta N_{t_i}\} \\
&= E_{t_i}\{E_t[(x-\mu_x)^3]\delta N_t + 3\mu_x E_t[(x-\mu_x)^2](\delta N_{t_i} - \lambda_{t_i} h)^2 + \mu_x^3(\delta N_{t_i} - \lambda_{t_i} h)^3\} \\
&= 3\sigma_x^2 \mu_x E_{t_i}(\delta N_{t_i} - \lambda_{t_i} h)^2 + \mu_x^3 E_{t_i}(\delta N_{t_i} - \lambda_{t_i} h)^3 = E(x^3)\lambda_{t_i} h.
\end{aligned}$$

Thus, we have

$$E_{t_0}\Big[\sum_{i=0}^{n-1}(\delta R_{t_i})^3\Big] = \sum_{i=0}^{n-1} E_{t_0}\Big[E_{t_i}[(\delta R_{t_i})^3]\Big] = \sum_{i=0}^{n-1} E_{t_0}\Big[E(x^3)\lambda_{t_i} h + O(h^2)\Big],$$

where $O(h^2)$ means in the order of $h^2$ and it equals $3\mu_{t_i}(v_{t_i} + E(x^2)\lambda_{t_i})h^2 + \mu_{t_i}^3 h^3$. As $h \to 0$, by the definition of Riemann integral, we obtain Equation (4.8). This completes the proof.

### 4.7.2 Proof of Proposition 4.2

Given a state price density

$$\pi_t = e^{-rt}\exp\Big(-\int_0^t \zeta_s dW_s - \frac{1}{2}\int_0^t \zeta_s^2 ds\Big),$$

the corresponding Radon-Nikodym derivative is

$$\Big(\frac{dQ}{dP}\Big)_t = \exp\Big(-\int_0^t \zeta_s dW_s - \frac{1}{2}\int_0^t \zeta_s^2 ds\Big),$$

where $W_s$ is a 1-dimensional Brownian motion, $Q$ denotes the risk-neutral measure and $P$ denotes the physical measure. Define $B_t(Q) = W_t + \int_0^t \zeta_s ds$. Calculating its

moment-generating function in risk-neutral measure gives

$$\begin{aligned} E^Q[e^{cB_t(Q)}] &= E^P\Big[\exp\Big(c(W_t + \int_0^t \zeta_s ds)\Big)\exp\Big(-\int_0^t \zeta_s dW_s - \frac{1}{2}\int_0^t \zeta_s^2 ds\Big)\Big] \\ &= E^P\Big[\exp\Big(\int_0^t (c-\zeta_s)dW_s + c\int_0^t \zeta_s ds - \frac{1}{2}\int_0^t \zeta_s^2 ds\Big)\Big] = \exp\Big(\frac{1}{2}c^2 t\Big). \end{aligned}$$

Therefore, $B_t(Q)$ is a Brownian motion under the risk-neutral measure. It is straightforward to generalize the proof to the case of state price density with $n$-dimensional independent Brownian motions. Rewriting the three risk-neutral Brownian motions $B_t^S(Q)$, $B_t^V(Q)$ and $B_t^\Lambda(Q)$ in a differentiation form gives Equations (4.17)-(4.19).

Under the risk-neutral measure, the jump intensity of the Poisson counter $N_t$ is $e^g\lambda_t$ and the jump size $x \sim N(\mu_x + h\sigma_x^2, \sigma_x^2)$. This can be seen through the following conditional moment generating function of the compound Poisson process $\int_0^t x dN_u$,

$$\begin{aligned} &E^Q[e^{c\int_0^t x dN_u}|\lambda_{u\in[0,t]}] \\ &= E^P\Big[\exp\Big(\int_0^t [g + (c+h)x - h\mu_x - \frac{1}{2}h^2\sigma_x^2]dN_u - (e^g - 1)\int_0^t \lambda_u du\Big)\Big|\lambda_{u\in[0,t]}\Big] \\ &= \exp\Big((e^{c(\mu_x + h\sigma_x^2) + \frac{1}{2}c^2\sigma_x^2} - 1)\int_0^t e^g\lambda_u du\Big) = \exp\Big((\psi_x^Q(c) - 1)\int_0^t \lambda_u^Q du\Big), \end{aligned}$$

where $\psi_x^Q(c)$ denotes the moment generating function of $x$ and $\lambda_u^Q$ denotes the jump intensity at time $u$ under $Q$-measure. This completes the proof.

### 4.7.3 Proof of Proposition 4.3

Applying Ito's Lemma to Equation (4.14) gives

$$d\ln S_t = \Big[r - \frac{1}{2}v_t - \lambda_t^Q E^Q(e^x - 1 - x)\Big]dt + \sqrt{v_t}dB_t^S(Q) + x dN_t - \lambda_t^Q \mu_x^Q dt. \quad (4.31)$$

The long-horizon return from current time $t$, to a future time, $T$, is defined by

$$R_t^T \equiv \ln\frac{S_T}{S_t} = \int_t^T \Big[r - \frac{1}{2}v_u - \lambda_u^Q \mu_{f(x)}^Q\Big]du + \sqrt{v_u}dB_u^S(Q) + x dN_u - \lambda_u^Q \mu_x^Q du, \quad (4.32)$$

where $f(x) = e^x - 1 - x$ and $\mu^Q_{f(x)}$ denotes its mean under $Q$-measure. The conditional expectation at time $t$ is then given by

$$E^Q_t(R^T_t) = \int_t^T \left[ r - \frac{1}{2} E^Q_t(v_u) - E^Q_t(\lambda^Q_u)\mu^Q_{f(x)} \right] du. \tag{4.33}$$

Adopting the notations in Zhang et al. (2017), we have

$$X_T \equiv \int_t^T \sqrt{v_u} dB^S_u(Q), \qquad Y_T \equiv \int_t^T [v_u - E^Q_t(v_u)] du.$$

and we introduce another two integrals

$$I_T \equiv \int_t^T (x dN_u - \mu^Q_x \lambda^Q_u du), \qquad Z_T \equiv \int_t^T [\lambda^Q_u - E^Q_t(\lambda^Q_u)] du.$$

With those notations as well as $X_T$ and $Y_T$, subtracting (4.33) from (4.32) yields

$$R^T_t - E^Q_t(R^T_t) = X_T - \frac{1}{2} Y_T + I_T - \mu^Q_{f(x)} Z_T. \tag{4.34}$$

The diffusive variance is modelled by

$$dv_t = \kappa^Q_1(\theta^Q_1 - v_t)dt + \sigma_v \sqrt{v_t} dB^v_t(Q) + y dN_t - \mu_y \lambda^Q_t dt, \tag{4.35}$$

where $B^v_t(Q)$ and $B^S_t(Q)$ have correlation coefficient $\rho$. Converting Equation (4.35) into the stochastic integral form yields

$$v_s = E^Q_t(v_s) + \sigma_v \int_t^s e^{-\kappa^Q_1(s-u)} \sqrt{v_u} dB^v_u(Q) + \int_t^s e^{-\kappa^Q_1(s-u)} (y dN_u - \mu_y \lambda^Q_u du), \tag{4.36}$$

where $E^Q_t(v_s) = \theta^Q_1 + (v_t - \theta^Q_1) e^{-\kappa^Q_1(s-t)}$ is the conditional expected diffusive variance. Substituting Equation (4.36) and interchanging the order of integrations gives

$$Y_T = \sigma_v \int_t^T A_1(\kappa^Q_1, u) \sqrt{v_u} dB^v_u + \int_t^T A_1(\kappa^Q_1, u)(y dN_u - \mu_y \lambda^Q_u du),$$

where $A_1(\kappa_1^Q, u) = \frac{1-e^{-\kappa_1^Q(T-u)}}{\kappa_1^Q}$. We introduce new martingale processes, $Y_s^H$ and $Y_s^J$, as follows

$$Y_s^H \equiv \sigma_v \int_t^s A_1(\kappa_1^Q, u)\sqrt{v_u} dB_u^v, \quad Y_s^J \equiv \int_t^s A_1(\kappa_1^Q, u)(y dN_u - \mu_y \lambda_u^Q du),$$

and at time $T$, we have $Y_T = Y_T^H + Y_T^J$. Hence, Equation (4.34) becomes

$$R_t^T - E_t^Q(R_t^T) = X_T - \frac{1}{2}Y_T^H + I_T - \frac{1}{2}Y_T^J - \mu_{f(x)}^Q Z_T. \tag{4.37}$$

The jump intensity follows a square root process modelled by

$$d\lambda_t^Q = \kappa_2^Q(\theta_2^Q - \lambda_t^Q)dt + \sigma_\lambda^Q \sqrt{\lambda_t^Q} dB_t^\lambda(Q),$$

where $B_t^\lambda(Q)$ is independent of $B_t^S(Q)$, $B_t^v(Q)$ and $N_t$. Similarly, we obtain

$$\lambda_s^Q = E_t^Q(\lambda_s^Q) + \sigma_\lambda^Q \int_t^s e^{-\kappa_2^Q(s-u)}\sqrt{\lambda_u^Q} dB_u^\lambda(Q),$$
$$Z_T = \sigma_\lambda^Q \int_t^T A_1(\kappa_2^Q, u)\sqrt{\lambda_u^Q} dB_t^\lambda(Q),$$

where $E_t^Q(\lambda_s^Q) = \theta_2^Q + (\lambda_t^Q - \theta_2^Q)e^{-\kappa_2^Q(s-t)}$ is the conditional expected jump intensity. Noting the independency of $\{B_t^S(Q), B_t^v(Q)\}$ and $\{B_t^\lambda(Q), N_t\}$, the co-third cumulant $E_t^Q[(X_T - \frac{1}{2}Y_T^H)(I_T - \frac{1}{2}Y_T^J - \mu_{f(x)}^Q Z_t)^2]$ is zero. Hence, the third cumulant of $R_t^T$ is as follows

$$E_t^Q[R_t^T - E_t^Q(R_t^T)]^3 = TC^H + TC^J + TC^C, \tag{4.38}$$

where $TC^H$, $TC^J$ and $TC^C$ denote the third cumulant of that in the Heston model, the third cumulant of jump terms and the co-third cumulant of diffusion and jump

terms, respectively, that is,

$$
\begin{aligned}
TC^H &= E_t^Q\Big[\Big(X_T - \frac{1}{2}Y_T^H\Big)^3\Big],\\
TC^J &= E_t^Q\Big[\Big(I_T - \frac{1}{2}Y_T^J - \mu_{f(x)}^Q Z_t\Big)^3\Big],\\
TC^C &= 3E_t^Q\Big[\Big(X_T - \frac{1}{2}Y_T^H\Big)^2\Big(I_T - \frac{1}{2}Y_T^J - \mu_{f(x)}^Q Z_t\Big)\Big].
\end{aligned}
$$

The Heston-type third cumulant $TC^H$ is given by Zhang et al. (2017), as follows

$$
TC^H = E_t^Q[X_T^3] - \frac{3}{2}E_t^Q[X_T^2 Y_T^H] + \frac{3}{4}E_t^Q[X_T(Y_T^H)^2] - \frac{1}{8}E_t^Q[(Y_T^H)^3], \tag{4.39}
$$

where the third and co-third cumulants of $X_T$ and $Y_T^H$ are given by

$$
E_t^Q[X_T^3] = 3\rho\sigma_v \int_t^T A_1(\kappa_1^Q, u)\; E_t^Q(v_u)du, \tag{4.40}
$$

$$
E_t^Q[X_T^2 Y_T^H] = \sigma_v^2 \int_t^T A_2(\kappa_1^Q, u)\; E_t^Q(v_u)du, \tag{4.41}
$$

$$
E_t^Q[X_T(Y_T^H)^2] = \rho\sigma_v^3 \int_t^T A_3(\kappa_1^Q, u)\; E_t^Q(v_u)du, \tag{4.42}
$$

$$
E_t^Q[(Y_T^H)^3] = 3\sigma_v^4 \int_t^T A_4(\kappa_1^Q, u)\; E_t^Q(v_u)du, \tag{4.43}
$$

where the weights are given by

$$
\begin{aligned}
A_1(\kappa_1^Q, u) &= \frac{1 - e^{-\kappa_1^Q \tau^*}}{\kappa_1^Q}, \qquad \tau^* = T - u,\\
A_2(\kappa_1^Q, u) &= \Big(\frac{1 - e^{-\kappa_1^Q \tau^*}}{\kappa_1^Q}\Big)^2 + 2\rho^2 \frac{1 - e^{-\kappa_1^Q \tau^*} - \kappa_1^Q \tau^* e^{-\kappa_1^Q \tau^*}}{(\kappa_1^Q)^2},\\
A_3(\kappa_1^Q, u) &= 2\frac{1 - e^{-\kappa_1^Q \tau^*} - \kappa_1^Q \tau^* e^{-\kappa_1^Q \tau^*}}{(\kappa_1^Q)^2}\frac{1 - e^{-\kappa_1^Q \tau^*}}{\kappa_1^Q} + \frac{1 - e^{-2\kappa_1^Q \tau^*} - 2\kappa_1^Q \tau^* e^{-\kappa_1^Q \tau^*}}{(\kappa_1^Q)^3},\\
A_4(\kappa_1^Q, u) &= \frac{1 - e^{-2\kappa_1^Q \tau^*} - 2\kappa_1^Q \tau^* e^{-\kappa_1^Q \tau^*}}{(\kappa_1^Q)^3}\frac{1 - e^{-\kappa_1^Q \tau^*}}{\kappa_1^Q}.
\end{aligned}
$$

The co-third cumulant $E_t^Q[(I_T - \frac{1}{2}Y_T^J)Z_T^2] = E_t^Q[Z_T^2 E_t^Q[(I_T - \frac{1}{2}Y_T^J)|\lambda_{u\in[t,T]}]]$ is zero. Therefore, the third cumulant of jump terms is given by

$$\begin{aligned} TC^J = E_t^Q[I_T^3] - \frac{3}{2}E_t^Q[I_T^2 Y_T^J] + \frac{3}{4}E_t^Q[I_T(Y_T^J)^2] - \frac{1}{8}E_t^Q[(Y_T^J)^3] \\ -3\mu_{f(x)}^Q E_t^Q\Big[\Big(I_T - \frac{1}{2}Y_T^J\Big)^2 Z_T\Big] - (\mu_{f(x)}^Q)^3 E_t^Q[Z_T^3], \end{aligned} \tag{4.44}$$

where the (co-) variance and (co-) third cumulants of $I_T$ and $Y_T^J$ are given by

$$E_t^Q(I_T^3) = E^Q(x^3)\int_t^T E_t^Q(\lambda_u^Q)du, \tag{4.45}$$

$$E_t^Q[I_T^2 Y_T^J] = E^Q(x^2 y)\int_t^T A_1(\kappa_1^Q, u)E_t^Q(\lambda_u^Q)du, \tag{4.46}$$

$$E_t^Q[I_T(Y_T^J)^2] = E^Q(xy^2)\int_t^T A_1(\kappa_1^Q, u)^2 E_t^Q(\lambda_u^Q)du, \tag{4.47}$$

$$E_t^Q[(Y_T^J)^3] = E^Q(y^3)\int_t^T A_1(\kappa_1^Q, u)^3 E_t^Q(\lambda_u^Q)du. \tag{4.48}$$

We obtain a similar result for $E_t^Q(Z_T^3)$ to that in the Heston model, that is, the third cumulant of $Z_T$ is given by

$$E_t(Z_T^3) = 3(\sigma_\lambda^Q)^4\int_t^T A_4(\kappa_2^Q, u)\ E_t^Q(\lambda_u^Q)du, \tag{4.49}$$

and the co-third cumulant of $I_T - \frac{1}{2}Y_T^J$ and $Z_t$ is given by

$$\begin{aligned} &E_t^Q\Big[\Big(I_T - \frac{1}{2}Y_T^J\Big)^2 Z_T\Big] = E_t^Q\Big[Z_T E_t^Q\Big[\Big(I_T - \frac{1}{2}Y_T^J\Big)^2\Big|\lambda_{u\in[t,T]}\Big]\Big] \\ &= E_t^Q\Big[\int_t^T(\lambda_u^Q - E_t^Q(\lambda_u^Q))du\int_t^T E^Q\Big[\Big(x - \frac{1}{2}yA_1(\kappa_1^Q, s)\Big)^2\Big]\lambda_s^Q ds\Big] \\ &= E_t^Q\Big[\sigma_\lambda^Q\int_t^T A_1(\kappa_2^Q, u)\sqrt{\lambda_u^Q}dB_u^\lambda(Q) \\ &\qquad\times\int_t^T E^Q\Big[\Big(x - \frac{1}{2}yA_1(\kappa_1^Q, s)\Big)^2\Big]\sigma_\lambda^Q\int_t^s e^{-\kappa_2^Q(s-u)}\sqrt{\lambda_u^Q}dB_u^\lambda(Q)ds\Big] \\ &= (\sigma_\lambda^Q)^2\int_t^T A_1(\kappa_2^Q, u)\int_u^T E^Q\Big[\Big(x - \frac{1}{2}yA_1(\kappa_1^Q, s)\Big)^2\Big]e^{-\kappa_2^Q(s-u)}dsE_t^Q(\lambda_u^Q)du. \end{aligned} \tag{4.50}$$

Given that $B_t^{\lambda}(Q)$ is independent of $\{B_t^v(Q), N_t\}$, $E_t^Q[(X_T - \frac{1}{2}Y_T^H)^2 Z_t]$ is zero. Thus, the co-third cumulant of diffusion and jump terms is given by

$$\begin{aligned} TC^C = {} & 3E_t^Q[X_T^2 I_T] - 3E_t^Q[X_T Y_T^H I_T] + \frac{3}{4}E_t^Q[(Y_T^H)^2 I_T] \\ & -\frac{3}{2}E_t^Q[X_T^2 Y_T^J] + \frac{3}{2}E_t^Q[X_T Y_T^H Y_T^J] - \frac{3}{8}E_t^Q[(Y_T^H)^2 Y_T^J], \end{aligned} \tag{4.51}$$

where the co-third cumulants of $X_T$, $Y_T^H$ and $I_T$ or $Y_T^J$ are given by

$$E_t^Q[X_T^2 I_T] = E^Q(xy)\int_t^T A_1(\kappa_1^Q, u)E_t^Q(\lambda_u^Q)du, \tag{4.52}$$

$$E_t^Q[X_T Y_T^H I_T] = \rho\sigma_v E^Q(xy)\int_t^T A_5(\kappa_1^Q, u)E_t^Q(\lambda_u^Q)du, \tag{4.53}$$

$$E_t^Q[(Y_T^H)^2 I_T] = \sigma_v^2 E^Q(xy)\int_t^T A_6(\kappa_1^Q, u)E_t^Q(\lambda_u^Q)du, \tag{4.54}$$

$$E_t^Q[X_T^2 Y_T^J] = E^Q(y^2)\int_t^T A_1(\kappa_1^Q, u)^2 E_t^Q(\lambda_u)du, \tag{4.55}$$

$$E_t^Q[X_T Y_T^H Y_T^J] = \rho\sigma_v E^Q(y^2)\int_t^T A_7(\kappa_1^Q, u)E_t^Q(\lambda_u^Q)du, \tag{4.56}$$

$$E_t^Q[(Y_T^H)^2 Y_T^J] = \sigma_v^2 E^Q(y^2)\int_t^T A_4(\kappa_1^Q, u)E_t^Q(\lambda_u^Q)du, \tag{4.57}$$

and the weights are given by

$$\begin{aligned} A_5(\kappa_1^Q, u) &= \frac{1 - e^{-\kappa_1^Q\tau^*} - \kappa_1^Q\tau^* e^{-\kappa_1^Q\tau^*}}{(\kappa_1^Q)^2}, \\ A_6(\kappa_1^Q, u) &= \frac{1 - e^{-2\kappa_1^Q\tau^*} - 2\kappa_1^Q\tau^* e^{-\kappa_1^Q\tau^*}}{(\kappa_1^Q)^3}, \\ A_7(\kappa_1^Q, u) &= \frac{1 - e^{-\kappa_1^Q\tau^*} - \kappa_1^Q\tau^* e^{-\kappa_1^Q\tau^*}}{(\kappa_1^Q)^2}\frac{1 - e^{-\kappa_1^Q\tau^*}}{\kappa_1^Q}. \end{aligned}$$

Noting that $E_t^Q[X_T^3]$, $E_t^Q(I_T^3)$ and $3E_t^Q[X_T^2 I_T]$ are the main terms, and all the other terms are related with drift in Equation (4.31), which is negligible for short-term returns, we obtain the approximate formula in Equation (4.21). This completes the proof.

TABLE 4.1: Monthly Jump Magnitude

This table shows the monthly future jump magnitude, which is measured by the realized cubic variation, $RCV^f$, calculated with daily returns over one future month (30 calendar days) starting from the fourth Wednesday (Data: January 1990 to December 2015). The calculation period is matched with the 30-day SKEW and VIX. If SKEW or VIX is not available on the fourth Wednesday, we first use the following Thursday instead, then the preceding Tuesday and then the nearest trading day when both SKEW and VIX are available. We report the starting date of the sample month when the absolute value of $RCV^f$ is greater than five basis points (0.0005), and the corresponding financial events.

| Starting Date | $RCV^f * 10^3$ | Events |
|---|---|---|
| 25-Jul-1990 | -1.0848 | Early 1990s Recession |
| 26-Dec-1990 | 0.5169 | |
| 23-Oct-1991 | -0.6255 | |
| 27-Aug-1997 | 0.5576 | 1997-98 Asian Financial Crash |
| 22-Oct-1997 | -2.6807 | |
| 22-Jul-1998 | -0.6891 | |
| 26-Aug-1998 | -2.5441 | |
| 27-Oct-1999 | 0.6269 | Dot-Com Bubble |
| 26-Jan-2000 | -0.6422 | |
| 23-Feb-2000 | 1.6277 | (Dot-com Bubble Peak) |
| 22-Mar-2000 | -2.0994 | |
| 24-May-2000 | 0.5274 | |
| 27-Dec-2000 | 0.9524 | |
| 28-Feb-2001 | -1.5214 | |
| 28-Mar-2001 | 1.3092 | |
| 22-Aug-2001 | -2.3063 | (September 11 attacks) |
| 24-Apr-2002 | 0.5164 | |
| 24-Jul-2002 | 2.2815 | |
| 28-Aug-2002 | -1.7283 | (2002 Stock Market Downturn) |
| 25-Sep-2002 | 2.1081 | |
| 26-Dec-2007 | -0.7618 | 2007-08 Financial Crisis |
| 27-Feb-2008 | 0.8634 | |
| 26-Mar-2008 | 0.5968 | |
| 28-May-2008 | -0.7273 | |
| 27-Aug-2008 | -2.3516 | (Lehman Brothers Bankruptcy) |
| 24-Sep-2008 | -13.7539 | |
| 22-Oct-2008 | 5.9481 | |
| 26-Nov-2008 | -7.1596 | |
| 24-Dec-2008 | -1.5327 | |
| 28-Jan-2009 | -3.1088 | |
| 25-Feb-2009 | 5.3737 | |
| 25-Mar-2009 | -0.5206 | |
| 22-Apr-2009 | 0.7018 | |
| 27-Jul-2011 | -4.9485 | 2011 Stock Market Fall |
| 24-Aug-2011 | -0.5853 | |
| 28-Sep-2011 | 0.7754 | |
| 26-Oct-2011 | -0.7287 | |
| 23-Nov-2011 | 1.3889 | |
| 22-Jul-2015 | -0.5125 | 2015 Stock Market Selloff |

TABLE 4.2: Implied Third Cumulant

This table shows the extreme values of the implied third cumulant, $ITC$ (Data: January 1990 to December 2015 from the CBOE website). The monthly sample is selected on the fourth Wednesday with several exceptions. If SKEW or VIX is not available on the fourth Wednesday, we first use the following Thursday instead, then the preceding Tuesday and then the nearest trading day when both SKEW and VIX are available. We report the date when the value of $ITC$ is less than -0.01, and the corresponding financial events.

| Starting Date | $ITC$ | Events |
|---|---|---|
| 24-Jan-1990 | -0.0136 | Early 1990s Recession |
| 22-Aug-1990 | -0.0103 | |
| 24-Oct-1990 | -0.0106 | |
| 26-Dec-1990 | -0.0111 | |
| 25-Nov-1997 | -0.0172 | 1997-98 Asian Financial Crash |
| 24-Dec-1997 | -0.0140 | |
| 26-Aug-1998 | -0.0219 | |
| 23-Sep-1998 | -0.0226 | |
| 28-Oct-1998 | -0.0249 | |
| 27-Jan-1999 | -0.0123 | |
| 26-May-1999 | -0.0110 | |
| 22-Nov-2000 | -0.0103 | Dot-com Bubble |
| 26-Sep-2001 | -0.0183 | |
| 24-Oct-2001 | -0.0171 | |
| 24-Jul-2002 | -0.0185 | |
| 28-Aug-2002 | -0.0163 | |
| 25-Sep-2002 | -0.0165 | |
| 23-Oct-2002 | -0.0149 | |
| 24-Sep-2008 | -0.0146 | 2008 Financial Crisis |
| 22-Oct-2008 | -0.1886 | |
| 26-Nov-2008 | -0.0636 | |
| 24-Dec-2008 | -0.0307 | |
| 28-Jan-2009 | -0.0403 | |
| 25-Feb-2009 | -0.0385 | |
| 25-Mar-2009 | -0.0310 | |
| 22-Apr-2009 | -0.0203 | |
| 27-May-2009 | -0.0196 | |
| 24-Jun-2009 | -0.0120 | |
| 28-Oct-2009 | -0.0115 | |
| 26-May-2010 | -0.0266 | 2010 Flash Crash |
| 23-Jun-2010 | -0.0112 | |
| 24-Aug-2011 | -0.0312 | 2011 stock market fall |
| 28-Sep-2011 | -0.0364 | |
| 26-Oct-2011 | -0.0174 | |
| 23-Nov-2011 | -0.0190 | |
| 26-Aug-2015 | -0.0196 | 2015 stock market selloff |
| 23-Sep-2015 | -0.0115 | |

TABLE 4.3: Summary Statistics

This table presents the summary statistics of the realized cubic variation ($RCV$), the implied third cumulant ($ITC$), the realized bi-power variation ($RBV$), the realized variance ($RV$) and the risk-neutral variance ($RNV$). The realized variables, $RCV$ (or $RCV^f$), $RV$ and $RBV$, are computed using the past (or future) SPX daily returns over 30 calendar days. The sample period is from from January 1990 to December 2015 (312 months), and the monthly sample points are on the fourth Wednesday with several exceptions. If SKEW or VIX is not available on the fourth Wednesday, we first use the following Thursday instead, then the preceding Tuesday and then the nearest trading day when both SKEW and VIX are available. The asterisk $^*$ indicates that the correlation coefficient is significantly different from zero at the 1% level.

| | Mean | Std Dev | Skewness | Kurtosis | Min | Max |
|---|---|---|---|---|---|---|
| RCV | -0.00005 | 0.0012 | -2.6807 | 72.2281 | -0.0133 | 0.0103 |
| RCV$^f$ | -0.00008 | 0.0011 | -5.8730 | 77.1740 | -0.0138 | 0.0059 |
| ITC | -0.00576 | 0.0125 | -10.6033 | 147.8363 | -0.1886 | -0.0004 |
| RBV | 0.02854 | 0.0507 | 6.7867 | 64.7378 | 0.0025 | 0.5983 |
| RV | 0.03238 | 0.0586 | 7.2150 | 69.0014 | 0.0026 | 0.6631 |
| RNV | 0.04396 | 0.0435 | 4.8124 | 40.5915 | 0.0087 | 0.4851 |

| Correlation | RCV | RCV$^f$ | ITC | RBV | RV | RNV |
|---|---|---|---|---|---|---|
| RCV | 1.0000 | -0.3093$^*$ | 0.4766$^*$ | -0.0768 | -0.2280$^*$ | -0.3154$^*$ |
| RCV$^f$ | | 1.0000 | -0.1455$^*$ | -0.1997$^*$ | -0.0717 | 0.0353 |
| ITC | | | 1.0000 | -0.7682$^*$ | -0.8800$^*$ | -0.9085$^*$ |
| RBV | | | | 1.0000 | 0.9675$^*$ | 0.8551$^*$ |
| RV | | | | | 1.0000 | 0.9033$^*$ |
| RNV | | | | | | 1.0000 |

TABLE 4.4: Forecasting Jump Magnitude

We use the implied third cumulant ($ITC$) and realized bi-power variation ($RBV$) to forecast future realized cubic variation ($RCV^f$). The regression is specified by

$$RCV_t^f = \alpha + \beta_1 ITC_t + \beta_2 RBV_t + \varepsilon_t.$$

The realized variance ($RV$) and risk-neutral variance ($RNV$) are used as alternatives to $RBV$ for a robustness check, and one-month lagged $RCV$ is used to replace $ITC$ to explore the mean-reverting behaviour of the jump magnitude. We choose the monthly sample on the fourth Wednesday. If SKEW or VIX is not available on the fourth Wednesday, we first use the following Thursday instead, then the preceding Tuesday and then the nearest trading day when both SKEW and VIX are available. We report the results for the full sample, FS: January 1990 - December 2015, and two equally-divided sub-samples, S1: January 1990 - December 2002 and S2: January 2003 - December 2015. The t-statistics, which are reported in parentheses, are adjusted for heteroscedasticity and serial correlation.

| | $\alpha * 10^3$ | | $\beta_1 * 10^3$ | | $\beta_2 * 10^3$ | | Adj $R^2$ |
|---|---|---|---|---|---|---|---|
| FS ITC | -0.156 | (-1.904) | -13.236 | (-2.607) | | | 0.018 |
| FS RBV | 0.049 | ( 1.209) | | | -4.494 | (-2.189) | 0.037 |
| FS RV | -0.035 | (-0.702) | | | -1.397 | (-0.946) | 0.002 |
| FS RNV | -0.120 | (-1.710) | | | 0.926 | ( 0.544) | -0.002 |
| FS RCV | -0.093 | (-1.353) | -293.975 | (-3.309) | | | 0.093 |
| S1 RCV | -0.018 | (-0.443) | -228.400 | (-1.851) | | | 0.031 |
| S2 RCV | -0.170 | (-1.331) | -302.970 | (-3.096) | | | 0.101 |
| FS ITC+RBV | 0.026 | ( 0.651) | -66.345 | (-4.153) | -17.103 | (-3.082) | 0.253 |
| FS ITC+RV | -0.006 | (-0.123) | -84.125 | (-2.440) | -17.255 | (-2.101) | 0.193 |
| FS ITC+RNV | 0.220 | ( 0.734) | -59.083 | (-1.557) | -14.565 | (-1.101) | 0.069 |
| S1 ITC+RBV | -0.068 | (-1.468) | 2.605 | ( 0.114) | 2.425 | ( 0.782) | -0.001 |
| S2 ITC+RBV | 0.022 | ( 0.429) | -82.799 | (-5.191) | -22.657 | (-3.935) | 0.386 |

TABLE 4.5: Forecasting Jump Existence Likelihood

We use the implied third cumulant $\ln(|ITC|)$ to forecast future jump existence likelihood. The regression is specified by

$$P(J_t^f = 1|ITC_t) = \phi_B(\alpha + \beta \ln(|ITC_t|)), \quad \phi_B(x) = 1 - e^{-e^x}.$$

The realized bi-power variation $\ln(RBV)$, the realized variance $\ln(RV)$, the risk-neutral variance $\ln(RNV)$, the lagged cubic variation $\ln(|RCV|)$ and the lagged jump existence $\ln(1 + J)$ are are used as alternatives to $\ln(|ITC|)$ for comprison. We choose the monthly sample on the fourth Wednesday. If SKEW or VIX is not available on the fourth Wednesday, we first use the following Thursday instead, then the preceding Tuesday and then the nearest trading day when both SKEW and VIX are available. We report three pseudo R-squareds: Efron, Macfadden and Adjusted Count. The t-statistics are reported in parentheses.

| | | | | | Pseudo $R^2$ | | |
|---|---|---|---|---|---|---|---|
| | $\alpha$ | | $\beta$ | | Efron | Macfadden | Adj Count |
| ITC | 4.3773 | ( 5.1797) | 1.2069 | (6.9359) | 0.2368 | 0.2370 | 0.1795 |
| RBV | 3.0572 | ( 5.0273) | 1.4051 | (7.4863) | 0.3046 | 0.3025 | 0.2564 |
| RV | 5.9005 | ( 6.2087) | 2.3952 | (7.4359) | 0.5318 | 0.5388 | 0.5128 |
| RNV | 4.2761 | ( 5.5635) | 2.0822 | (7.3663) | 0.2878 | 0.2942 | 0.2821 |
| RCV | 4.3703 | ( 5.0363) | 0.7346 | (6.7920) | 0.2127 | 0.2117 | 0.0513 |
| J | -2.5214 | (-11.5513) | 2.8954 | (6.1972) | 0.1448 | 0.1367 | 0.0000 |

FIGURE 4.1: Implied Third Cumulant and Skewness

This graph shows the evolutions of the option-implied third cumulant ($ITC$ magnified by 10) and skewness ($Sk$) from 02 January 1990 to 31 December 2015. The solid line represents $ITC$ multiplied by 10, and the dotted line represents $Sk$.

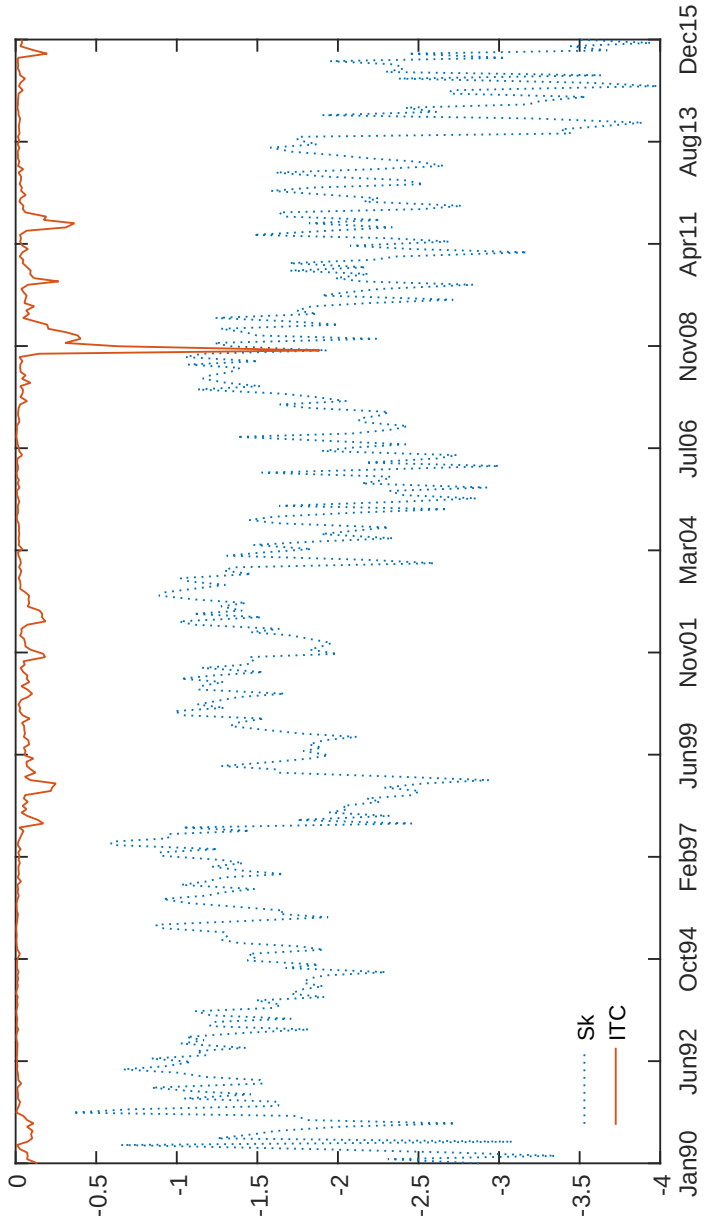


FIGURE 4.2: Forecasting Jump Magnitude

This graph shows the observed and forecasted values from the multivariate regression specified by

$$RCV_t^f = \alpha + \beta_1 ITC_t + \beta_2 RBV_t + \varepsilon_t.$$

The sample period is from 2 January 1990 to 31 December 2015.

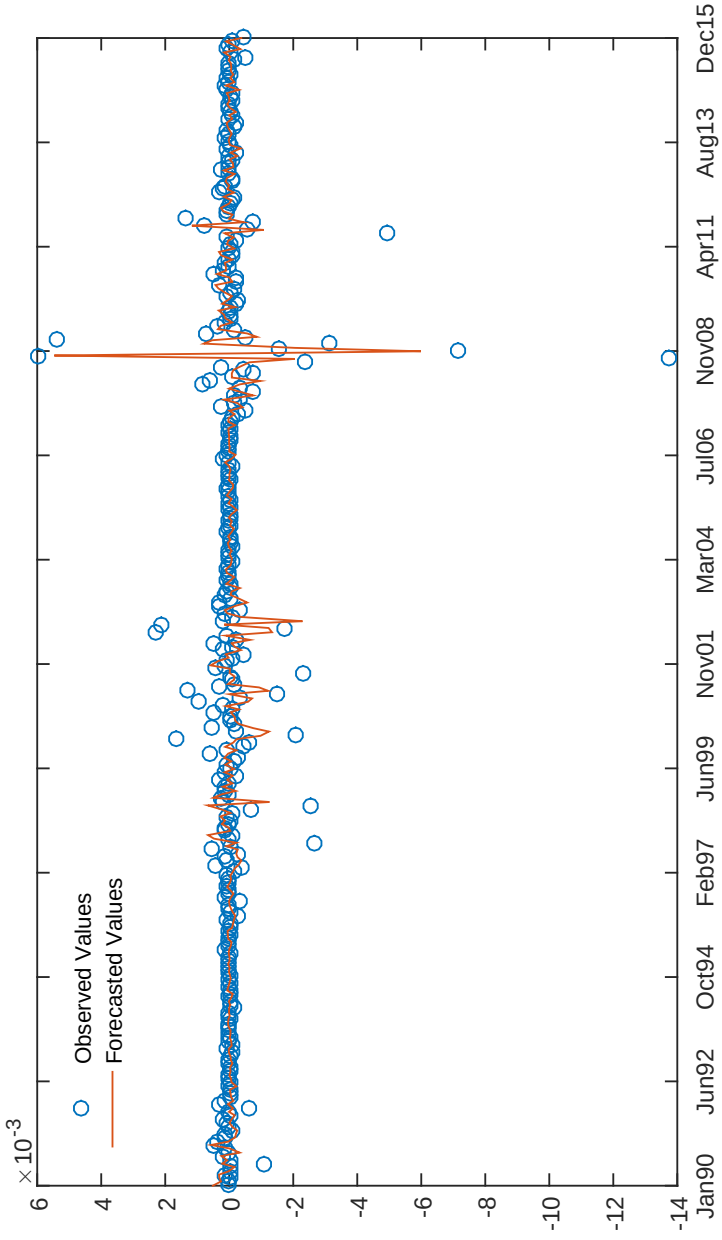
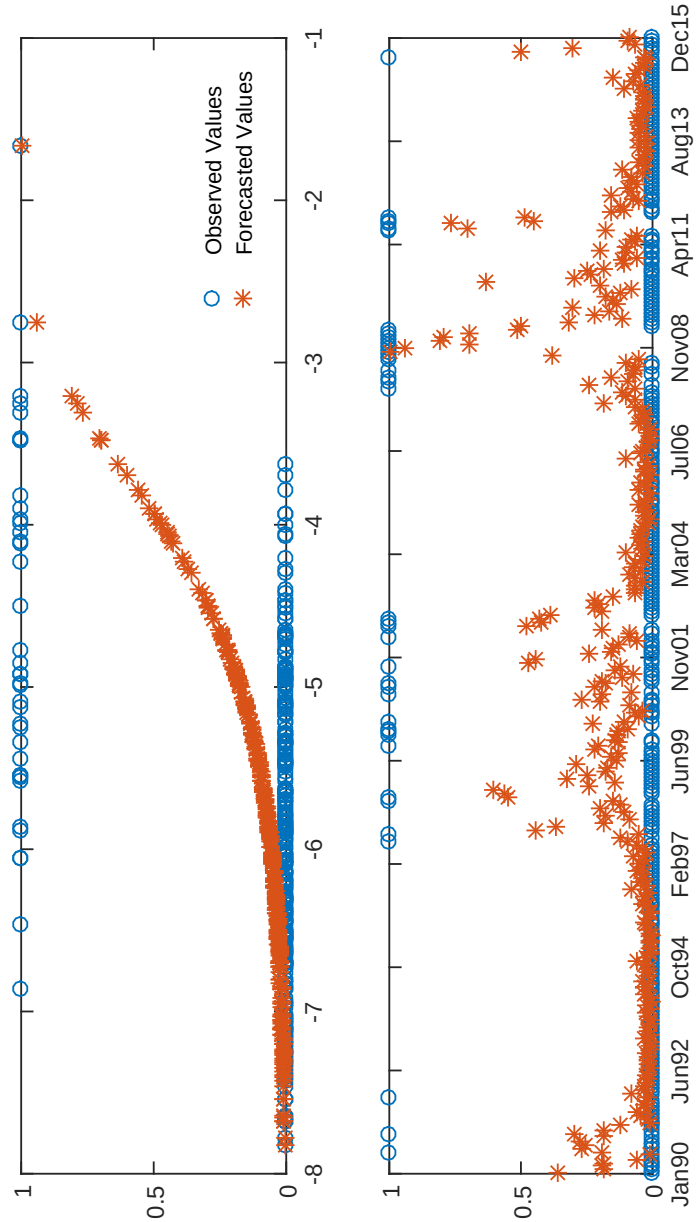


FIGURE 4.3: Forecasting Jump Existence Likelihood

This graph shows the observed and forecasted values against the predictor variable $\ln(|ITC|)$ (upper panel) and against the observation date (lower panel) from the binomial regression specified by

$$P(J_t^f = 1|ITC_t) = \phi_B(\alpha + \beta \ln(|ITC_t|)), \quad \phi_B(x) = 1 - e^{-e^x}.$$

The sample period is from 2 January 1990 to 31 December 2015.



# Chapter 5

# Conclusion

In Chapter 2, we obtain a closed-form formula of the skewness implied in the Heston (1993) model. In the literature, Das and Sundaram's (1999) skewness formula in a stochastic-volatility model, which is similar to but different from the Heston model, has misled scholars to perceive it as the skewness in the Heston model. We point out the difference between them theoretically and numerically. We further estimate the parameters in the Heston model using the term structure data of the CBOE VIX and SKEW plus the SPX and 30-day VIX data. Our estimation results show that the Heston model is unable to capture the non-zero short-term skewness, and the time-varying behavior of the stochastic long-term mean level of variance. In order to enhance the performance of the Heston model, it is important to incorporate these additional factors into the model.

In Chapter 3, we derive skewness formulas under various affine jump-diffusion models (Duffie, Pan and Singleton, 2000), and compare the empirical performance of different models in fitting the term structure data of the CBOE VIX and SKEW. Our estimation results show that the five-factor model with stochastic variance, stochastic jump intensity and their corresponding stochastic long-term mean levels exhibits the best fitting performance. As the VIX and SKEW term structure data are extracted from option prices, the parameters and latent variables are estimated under the risk-neutral measure, and can be directly applied to option pricing.

In Chapter 4, we test the relation between the implied third cumulant extracted from the 30-day CBOE VIX and SKEW and the SPX future realized cubic variation, which is used to measure the SPX jump magnitude. We find that the opposite signs of the expectations of the price-jump-size under the risk-neutral and physical measures induce a negative relation between the realized cubic variation and

its risk-neutral expectation. We also find that option-implied information coupled with past diffusive variance is more efficient in forecasting future jump magnitude than is time-series information. Moreover, we detect the existence of jumps if the absolute value of the realized cubic variation exceeds a given threshold. Time-series information outperforms option-implied information in forecasting future jump existence likelihood. Our empirical findings support the time-varying behavior of the jump intensity and the existence of variance jumps which happen contemporaneously with price jumps.

# Bibliography


[1] Aït-Sahalia, Yacine, and Robert Kimmel, 2007, Maximum likelihood estimation of stochastic volatility models, *Journal of Financial Economics* 83(2), 413-452.

[2] Albuquerque, Rui, 2012, Skewness in stock returns: Reconciling the evidence on firm versus aggregate returns, *Review of Financial Studies* 25(5), 1630-1673.

[3] Amaya, D., Christoffersen, P., Jacobs, K. and Vasquez, A., 2015, Does realized skewness predict the cross-section of equity returns? *Journal of Financial Economics* 118(1), 135-167.

[4] Andersen, Torben G., Luca Benzoni, and Jesper Lund, 2002, An empirical investigation of continuous-time equity return models, *Journal of Finance* 57(3), 1239-1284.

[5] Bakshi, Gurdip, Charles Cao, and Zhiwu Chen, 1997, Empirical performance of alternative option pricing models, *Journal of Finance* 52(5), 2003-2049.

[6] Bakshi, Gurdip, Charles Cao, and Zhiwu Chen, 2000, Pricing and hedging long-term options, *Journal of Econometrics* 94(1), 277-318.

[7] Bakshi, Gurdip, Nikunj Kapadia, and Dilip Madan, 2003, Stock return characteristics, skew laws, and the differential pricing of individual equity options, *Review of Financial Studies* 16(1), 101-143.

[8] Bali, Turan G., and Scott Murray, 2013, Does risk-neutral skewness predict the cross-section of equity option portfolio returns? *Journal of Financial and Quantitative Analysis* 48(4), 1145-1171.

[9] Ball, Clifford A., and Walter N. Torous, 1983, A simplified jump process for common stock returns, *Journal of Financial and Quantitative analysis* 18(1), 53-65.

[10] Ball, Clifford A., and Walter N. Torous, 1985, On jumps in common stock prices and their impact on call option pricing, *Journal of Finance* 40(1), 155-173.

[11] Bandi, Federico M., and Roberto Renò, 2016, Price and volatility co-jumps, *Journal of Financial Economics* 119(1), 107-146.

[12] Barberis, Nicholas, and Ming Huang, 2008, Stocks as lotteries: The implications of probability weighting for security prices, *American Economic Review* 98(5), 2066-2100.

[13] Barndorff-Neilsen, Ole E., and Neil Shephard, 2004, Power and bipower variation with stochastic volatility and jumps, *Journal of Financial Econometrics* 2(1), 1-37.

[14] Barndorff-Nielsen, Ole E., and Neil Shephard, 2006, Econometrics of testing for jumps in financial economics using bipower variation, *Journal of Financial Econometrics* 4(1), 1-30.

[15] Berger, Dave, and Kuntara Pukthuanthong, 2012, Market fragility and international market crashes, *Journal of Financial Economics* 105(3), 565-580.

[16] Black, Fischer, and Myron Scholes, 1973, The pricing of options and corporate liabilities, *Journal of Political Economy* 81(3), 637-654.

[17] Boyer, Brian, Todd Mitton, and Keith Vorkink, 2010, Expected idiosyncratic skewness, *Review of Financial Studies* 23(1), 169-202.

[18] Branger, Nicole, Holger Kraft, and Christoph Meinerding, 2016, The dynamics of crises and the equity premium, *Review of Financial Studies* 29(1), 232-270.

[19] Broadie, Mark, Mikhail Chernov, and Michael Johannes, 2007, Model specification and risk premia: Evidence from futures options, *Journal of Finance* 62(3), 1453-1490.

[20] Carr, Peter, and Liuren Wu, 2009, Variance risk premiums, *Review of Financial Studies* 22(3), 1311-1341.

[21] Chang, Bo Young, Peter Christoffersen, and Kris Jacobs, 2013, Market skewness risk and the cross section of stock returns, *Journal of Financial Economics* 107(1), 46-68.

[22] Chen, Joseph, Harrison Hong, and Jeremy C Stein, 2001, Forecasting crashes: Trading volume, past returns, and conditional skewness in stock prices, *Journal of Financial Economics* 61(3), 345-381.

[23] Chen, Hui, Scott Joslin, and Ngoc-Khanh Tran, 2012, Rare disasters and risk sharing with heterogeneous beliefs, *Review of Financial Studies* 25(7), 2189-2224.

[24] Chernov, Mikhail, and Eric Ghysels, 2000, A study towards a unified approach to the joint estimation of objective and risk neutral measures for the purpose of options valuation, *Journal of Financial Economics* 56(3), 407-458.

[25] Choe, Geon Ho, and Kyungsub Lee, 2014, High moment variations and their application, *Journal of Futures Markets* 34(11), 1040-1061.

[26] Christensen, B.J., and N.R. Prabhala, 1998, The relation between implied and realized volatility, *Journal of Financial Economics* 50(2), 125-150.

[27] 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(12), 1914-1932.

[28] Christoffersen, Peter, Kris Jacobs, Chayawat Ornthanalai, and Yintian Wang, 2008, Option valuation with long-run and short-run volatility components, *Journal of Financial Economics* 90(3), 272-297.

[29] Conrad, Jennifer, Robert F. Dittmar, and Eric Ghysels, 2013, Ex ante skewness and expected stock returns, *Journal of Finance* 68(1), 85-124.

[30] Cox, John C., Jonathan E. Ingersoll Jr, and Stephen A. Ross, 1985, A theory of the term structure of interest rates, *Econometrica* 53(2): 385-407.

[31] Das, Sanjiv Ranjan, and Rangarajan K. Sundaram, 1999, Of smiles and smirks: A term-structure perspective, *Journal of Financial and Quantitative Analysis* 34(2), 211-240.

[32] Dennis, Patrick, and Stewart Mayhew, 2002, Risk-neutral skewness: Evidence from stock options, *Journal of Financial and Quantitative Analysis* 37(3), 471-493.

[33] Drăgulescu, Adrian A, and Victor M Yakovenko, 2002, Probability distribution of returns in the Heston model with stochastic volatility, *Quantitative Finance* 2, 443-453.

[34] Duan, Jin-Chuan, and Chung-Ying Yeh, 2010, Jump and volatility risk premiums implied by VIX, *Journal of Economic Dynamics and Control* 34(11), 2232-2244.

[35] Duffie, Darrell, Jun Pan, and Kenneth Singleton, 2000, Transform analysis and asset pricing for affine jump-diffusions, *Econometrica* 68(6), 1343-1376.

[36] Dufresne, Daniel, 2001, The integrated square root process, Working paper, University of Melbourne.

[37] Egloff, Daniel, Markus Leippold, and Liuren Wu, 2010, The term structure of variance swap rates and optimal variance swap investments, *Journal of Financial and Quantitative Analysis* 45(5), 1279-1310.

[38] Eraker, Bjørn, 2004, Do stock prices and volatility jump? Reconciling evidence from spot and option prices. *Journal of Finance* 59(3), 1367-1403.

[39] Eraker, Bjørn, Michael Johannes, and Nicholas Polson, 2003, The impact of jumps in volatility and returns, *Journal of Finance* 58(3), 1269-1300.

[40] Faff, Robert W., and Zhangxin Liu, 2014, Hitting SKEW for SIX, Working paper, University of Western Australia.

[41] Fama, Eugene F., Kenneth R. French, 2015, A five-factor asset pricing model, *Journal of Financial Economics* 116(1), 1-22.

[42] Friesen, Geoffrey C., Yi Zhang and Thomas S. Zorn, 2012, Heterogeneous beliefs and risk-neutral skewness, *Journal of Financial and Quantitative Analysis* 47(4), 851-872.

[43] Garcia, René, Marc-André Lewis, Sergio Pastorello, and Éric Renault, 2011, Estimation of objective and risk-neutral distributions based on moments of integrated volatility, *Journal of Econometrics* 160(1), 22-32.

[44] Han, Bing, 2008, Investor sentiment and option prices, *Review of Financial Studies* 21(1), 387-414.

[45] Harvey, Campbell R., and Akhtar Siddique, 1999, Autoregressive conditional skewness, *Journal of Financial and Quantitative Analysis* 34(4), 465-487.

[46] Harvey, Campbell R., and Akhtar Siddique, 2000, Conditional skewness in asset pricing tests, *Journal of Finance* 55(3), 1263-1295.

[47] 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(2), 327-343.

[48] Hong, Harrison, and Jeremy C. Stein, 2003, Differences of opinion, short-sales constraints, and market crashes. *Review of Financial Studies* 16(2), 487-525.

[49] Jarrow, Robert A., and Eric R. Rosenfeld, 1984, Jump risks and the intertemporal capital asset pricing model, *Journal of Business* 57(3), 337-351.

[50] Jiang, George J., and Roel CA Oomen, 2008, Testing for jumps when asset prices are observed with noise - a “swap variance" approach, *Journal of Econometrics* 144(2), 352-370.

[51] Jiang, George J., and Yisong S. Tian, 2005, The model-free implied volatility and its information content, *Review of Financial Studies* 18(4), 1305-1342.

[52] Jiang, George J., and Tong Yao, 2013, Stock price jumps and cross-sectional return predictability, *Journal of Financial and Quantitative Analysis* 48(5), 1519-1544.

[53] Kozhan, Roman, Anthony Neuberger, and Paul Schneider, 2013, The skew risk premium in the equity index market, *Review of Financial Studies* 26(9), 2174-2203.

[54] Kraus, Alan, and Robert H. Litzenberger, 1976, Skewness preference and the valuation of risk assets, *Journal of Finance* 31(4), 1085-1100.

[55] Kumar, Mohan, Uma Moorthy, and William Perraudin, 2003, Predicting emerging market currency crashes, *Journal of Empirical Finance* 10(4), 427-454.

[56] Lee, Kyungsub, 2016, Probabilistic and statistical properties of moment variations and their use in inference and estimation based on high frequency return data, *Studies in Nonlinear Dynamics and Econometrics* 20(1), 19-36.

[57] Lee, Suzanne S., and Per A. Mykland, 2008, Jumps in financial markets: A new nonparametric test and jump dynamics, *Review of Financial Studies* 21(6), 2535-2563.

[58] Li, Gang, and Chu Zhang, 2013, Diagnosing affine models of options pricing: Evidence from VIX, *Journal of Financial Economics* 107(1), 199-219.

[59] Liu, Jun, Francis A. Longstaff, and Jun Pan, 2003, Dynamic asset allocation with event risk, *Journal of Finance* 58(1), 231-259.

[60] Liu, Jun, and Jun Pan, 2003, Dynamic derivative strategies, *Journal of Financial Economics* 69(3), 401-430.

[61] Liu, Jun, Jun Pan, and Tan Wang, 2005, An equilibrium model of rare-event premia and its implication for option smirks, *Review of Financial Studies* 18(1), 131-164.

[62] Luo, Xingguo, and Jin E. Zhang, 2012, The term structure of VIX, *Journal of Futures Markets* 32(12), 1092-1123.

[63] Mencía, Javier, and Enrique Sentana, 2013, Valuation of VIX derivatives, *Journal of Financial Economics* 108(2), 367-391.

[64] Merton, Robert C., 1976, Option pricing when underlying stock returns are discontinuous, *Journal of Financial Economics* 3(1-2), 125-144.

[65] Mitton, Todd, and Keith Vorkink, 2007, Equilibrium underdiversification and the preference for skewness, *Review of Financial Studies* 20(4), 1255-1288.

[66] Mitchell, Douglas W., 1991, Invariance of results under a common orthogonalization, *Journal of Economics and Business* 43(2), 193-196.

[67] Neuberger, Anthony, 2012, Realized skewness, *Review of Financial Studies* 25(11), 3423-3455.

[68] Nourdin, Ivan, 2008, Asymptotic behavior of weighted quadratic and cubic variations of fractional Brownian motion, *Annals of Probability* 36(6), 2159-2175.

[69] Pan, Jun, 2002, The jump-risk premia implicit in options: evidence from an integrated time-series study, *Journal of Financial Economics* 63(1), 3-50.

[70] Park, Yang-Ho, 2015, Volatility-of-volatility and tail risk hedging returns, *Journal of Financial Markets* 26, 38-63.

[71] Press, S. James, 1967, A compound events model for security prices, *Journal of Business* 40(3), 317-335.

[72] Siegel, Andrew F., 1979, The noncentral chi-squared distribution with zero degrees of freedom and testing for uniformity, *Biometrika* 66(2), 381-386.

[73] Singleton, Kenneth J., 2001, Estimation of affine asset pricing models using the empirical characteristic function, *Journal of Econometrics* 102(1), 111-141.

[74] Todorov, Viktor, 2010, Variance risk-premium dynamics: The role of jumps, *Review of Financial Studies* 23(1), 345-383.

[75] Todorov, Viktor, 2011, Econometric analysis of jump-driven stochastic volatility models, *Journal of Econometrics* 160(1), 12-21.

[76] Wang, Zhiguang, and Robert T. Daigler, 2014, VIX and SKEW Indices for SPX and VIX Options, Working paper, South Dakota State University.

[77] Xu, Jianguo, 2007, Price convexity and skewness, *Journal of Finance* 62(5), 2521-2552.

[78] Yan, Shu, 2011, Jump risk, stock returns, and slope of implied volatility smile, *Journal of Financial Economics* 99(1), 216-233.

[79] Zhang, Jin E., and Yuqin Huang, 2010, The CBOE S&P 500 Three-Month Variance Futures, *Journal of Futures Markets*, 30(1), 48-70.

[80] Zhang, Jin E., Jinghong Shu, and Menachem Brenner, 2010, The New Market for Volatility Trading, *Journal of Futures Markets*, 30(9), 809-833.

[81] Zhang, Jin E., Huimin Zhao, and Eric C. Chang, 2012, Equilibrium asset and option pricing under jump diffusion, *Mathematical Finance* 22(3), 538-568.

[82] Zhang, Jin E., Fang Zhen, Xiaoxia Sun, and Huimin Zhao, 2017, The skewness implied in the Heston model and its application, *Journal of Futures Markets* 37(3), 211-237.

[83] Zhao, Huimin, Jin E. Zhang, and Eric C. Chang, 2013, The relation between physical and risk-neutral cumulants, *International Review of Finance* 13(3), 345-381.

[84] Zhen, Fang, and Jin E. Zhang, 2014, The CBOE SKEW, Working paper, University of Otago.

[85] Zhen, Fang, and Jin E. Zhang, 2016, Jump risk: a cubic-variation approach, Working paper, University of Otago.