The CBOE SKEW

Fang Zhen[1]
Department of Accountancy and Finance
Otago Business School, University of Otago
Dunedin 9054, New Zealand
Email: fang.zhen@otago.ac.nz

Jin E. Zhang
Department of Accountancy and Finance
Otago Business School, University of Otago
Dunedin 9054, New Zealand
Email: jin.zhang@otago.ac.nz

First Version: March 2014
This Version: August 2016

Keywords: The CBOE SKEW; Risk-neutral skewness; VIX; Jump-diffusion; Stochastic volatility

JEL Classification Code: G12; G13

---

[1]Corresponding author. Tel: 64 022 4214279. We are grateful to Timothy Crack, José Da Fonseca (our DMC discussant), seminar participants at University of Otago, 2014 Quantitative Methods in Finance Conference, 2015 Derivative Markets Conference (DMC) and 2016 New Zealand Finance Colloquium for helpful comments and suggestions. Jin E. Zhang has been supported by an establishment grant from University of Otago.

The CBOE SKEW

Abstract

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 (SPX) index for different maturities. In this paper, 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 SPX risk-neutral distribution and to price SPX options.

1 Introduction

The CBOE SKEW is an index launched by the Chicago Board Options Exchange (CBOE) in February 2011.[2] Its term structure tracks the risk-neutral skewness of the S&P 500 (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 paper 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.

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 logarithmic returns. SKEW is a complementary risk indicator to VIX and is calculated from SPX option

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

prices 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 (Harvey and Siddique, 2000), for idiosyncratic skewness (Boyer, Mitton and Vorkink, 2010), for 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).[3] However, these skewness measures are constructed by the 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 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 paper, we build a theory of the CBOE SKEW and look at the empirical performance of different option-pricing models.

Zhang et al. (2016) 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

[3] 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).

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 paper 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 returns' distribution.

In this paper, 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-Heston (Merton-extended Heston) model outperforms the Heston model by 14% (20%) 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%). 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 paper proceeds as follows. Section 2 introduces the CBOE SKEW index. Section 3 develops a theory of the CBOE SKEW. Section 4 discusses the data. Section 5 presents the estimation procedure and compares the empirical performance. Section 6 provides concluding remarks. All proofs are in the Appendix.

2 Definition of the CBOE SKEW

The CBOE SKEW is an index launched by the Chicago Board Options Exchange (CBOE) on February 23, 2011. Its term structure tracks the risk-neutral skewness of the S&P 500 (SPX) index 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

$$SKEW_t = 100 - 10 \times Sk_t, \tag{1}$$

where Sk_t is the risk-neutral skewness given by

$$Sk_t = E_t^Q \left[\frac{(R_t^T - \mu)^3}{\sigma^3} \right]; \tag{2}$$

R_t^T is the logarithmic return of SPX at time T, denoted as S_T, against the current forward price with maturity T, F_t^T; μ is the expected return and σ^2 is the variance in risk-neutral measure Q:

$$R_t^T = \ln \frac{S_T}{F_t^T}, \qquad \mu = E_t^Q(R_t^T), \qquad \sigma^2 = E_t^Q[(R_t^T - \mu)^2]. \tag{3}$$

Expanding equation (2) gives

$$Sk_t = \frac{E_t^Q[(R_t^T)^3] - 3E_t^Q(R_t^T)E_t^Q[(R_t^T)^2] + 2[E_t^Q(R_t^T)]^3}{E_t^Q[(R_t^T)^2] - [E_t^Q(R_t^T)]^2}. \tag{4}$$

Following Bakshi, Kapadia and Madan (2003), the CBOE evaluates the first three moments of log 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:

$$E_t^Q(R_t^T) = -e^{r\tau} \sum_i \frac{1}{K_i^2} Q(K_i)\Delta K_i + \ln \frac{K_0}{F_t^T} + \frac{F_t^T}{K_0} - 1, \tag{5}$$

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

$$E_t^Q[(R_t^T)^3] = e^{r\tau} \sum_i \frac{6\ln \frac{K_i}{F_t^T} - 3\ln^2 \frac{K_i}{F_t^T}}{K_i^2} Q(K_i)\Delta K_i + \ln^3 \frac{K_0}{F_t^T} + 3\ln^2 \frac{K_0}{F_t^T}\left(\frac{F_t^T}{K_0} - 1\right), \tag{7}$$

where F_t^T is the forward index level derived from SPX option prices using put-call parity; K_0 is the first listed price below F_t^T; K_i is the strike price of the ith OTM option (a call if $K_i > K_0$ and a put if $K_i < K_0$); ΔK_i is half the difference between strikes on either side of K_i, i.e., $\Delta K_i = \frac{1}{2}(K_{i+1} - K_{i-1})$, and for minimum (maximum) strike, ΔK_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 τ is the time to expiration as a fraction of a year. The reasoning behind equations (5), (6) and (7) are included in Appendix A.

The SKEW index refers to 30-day maturity, i.e., $\tau = \tau_0 \equiv 30/365$. 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^{near} and Sk_t^{next} as follows:

$$Sk_t = \omega Sk_{t,\text{near}} + (1-\omega) Sk_{t,\text{next}}, \tag{8}$$

where ω is a weight determined by

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

and τ_{near} and τ_{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, i.e., the third Friday of the month.

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 return. 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 paper 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, i.e., the Black-Scholes (1973) model.

Proposition 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 σ is volatility. The model-implied squared VIX and skewness at time t for time to maturity τ are as follows:

$$VIX_{t,\tau}^2 = \sigma^2, \qquad Sk_{t,\tau} = 0. \tag{9}$$

This is a trivial case. The log 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 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 λ. The model-implied squared VIX and skewness at time t for time to maturity τ are as follows:

$$VIX_{t,\tau}^2 = \sigma^2 + 2\lambda(e^x - 1 - x), \qquad Sk_{t,\tau} = \frac{\lambda x^3}{(\sigma^2 + \lambda x^2)^{3/2}}\ \tau^{-1/2}. \tag{10}$$

Proof. See Appendix B.1.

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

Remark 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) $Var_{t,\tau} = \sigma^2 + \lambda x^2$. 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 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

$$RSk_{t,\tau} = \sqrt{\tau}\ Sk_{t,\tau} = \sqrt{\tau} E_t^Q \left[\frac{(R_{t,\tau} - \mu(R_{t,\tau}))^3}{\sigma^3(R_{t,\tau})} \right], \tag{11}$$

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

[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)

In addition to jumps, skewness can also be created by using stochastic volatility with leverage effect, i.e., correlation between stock return 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, i.e., $\kappa = 0$. A full explicit formula was not available until the recent work of Zhang et al. (2016).

Proposition 3 *(Zhang et al., 2016) In the Heston (1993) model, the risk-neutral underlying stock is modelled by*

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

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

$$VIX_{t,\tau}^2 = (1-\omega)\theta + \omega v_t, \qquad \omega = \frac{1-e^{-\kappa\tau}}{\kappa\tau}, \tag{12}$$

$$Sk_{t,\tau} = \frac{TC^H}{(Var^H)^{3/2}}, \tag{13}$$

where the term variance and third cumulant are given by

$$Var^H = E_t(X_T^2) - E_t(X_TY_T) + \frac{1}{4}E_t(Y_T^2), \tag{14}$$

$$TC^H = E_t(X_T^3) - \frac{3}{2}E_t(X_T^2Y_T) + \frac{3}{4}E_t(X_TY_T^2) - \frac{1}{8}E_t(Y_T^3). \tag{15}$$

The integrated return uncertainty, X_T, and integrated instantaneous variance uncer-

tainty, 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

$$E_t(X_T^2) = \int_t^T E_t(v_u) du, \tag{16}$$

$$E_t(X_T Y_T) = \rho\sigma_v \int_t^T A_1 E_t(v_u) du, \qquad A_1 = \frac{1 - e^{-\kappa\tau^*}}{\kappa}, \qquad \tau^* = T - u, \tag{17}$$

$$E_t(Y_T^2) = \sigma_v^2 \int_t^T A_1^2 E_t(v_u) du. \tag{18}$$

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

$$E_t(X_T^3) = 3\rho\sigma_v \int_t^T A_1 \ E_t(v_u) du, \tag{19}$$

$$\begin{aligned} E_t(X_T^2 Y_T) &= \sigma_v^2 \int_t^T A_2 \ E_t(v_u) du, \\ A_2 &= \left(\frac{1 - e^{-\kappa\tau^*}}{\kappa}\right)^2 + 2\rho^2 \frac{1 - e^{-\kappa\tau^*} - \kappa\tau^* e^{-\kappa\tau^*}}{\kappa^2}, \end{aligned} \tag{20}$$

$$\begin{aligned} E_t(X_T Y_T^2) &= \rho\sigma_v^3 \int_t^T A_3 \ E_t(v_u) du, \\ A_3 &= 2\frac{1 - e^{-\kappa\tau^*} - \kappa\tau^* e^{-\kappa\tau^*}}{\kappa^2} \frac{1 - e^{-\kappa\tau^*}}{\kappa} + \frac{1 - e^{-2\kappa\tau^*} - 2\kappa\tau^* e^{-\kappa\tau^*}}{\kappa^3}, \end{aligned} \tag{21}$$

$$\begin{aligned} E_t(Y_T^3) &= 3\sigma_v^4 \int_t^T A_4 \ E_t(v_u) du, \\ A_4 &= \frac{1 - e^{-2\kappa\tau^*} - 2\kappa\tau^* e^{-\kappa\tau^*}}{\kappa^3} \frac{1 - e^{-\kappa\tau^*}}{\kappa}, \end{aligned} \tag{22}$$

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

Proof. See Zhang et al. (2016).

Remark 3.1. Through asymptotic analysis, Zhang et al. (2016) shows that for $\kappa > 0$, if τ 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 τ 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 τ for small τ, and approaches to a constant for large τ.

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 variance long-term mean, to retain the martingale property of the stochastic variance 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 4 *In an extended Heston model with stochastic long term mean, the risk-neutral underlying stock is modelled by*

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

[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 θ_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 θ_t follows a noncentral chi-squared distribution with zero degrees of freedom (see Siegel, 1979), which is a Possion 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 χ_0^2 distribution is identically zero), where λ is the noncentrality parameter. In this setup, the Feller condition is violated, so the variance long-term mean is not strictly positive and is accessible to zero.

where the new standard Brownian motion, B_t^θ, is independent of B_t^S and B_t^v, and σ_θ is the volatility of the long-term mean of variance. The model-implied squared VIX and skewness at time t for time to maturity τ are as follows

$$\begin{aligned} VIX_{t,\tau}^2 &= (1-\omega)\theta_t + \omega v_t, \qquad \omega = \frac{1-e^{-\kappa\tau}}{\kappa\tau}, \\ Sk_{t,\tau} &= \frac{TC^H + TC^{HM}}{[Var^H + Var^{HM}]^{3/2}}, \end{aligned} \tag{23}$$

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

$$Var^{HM} = \frac{1}{4}\sigma_\theta^2\theta_t C_1, \tag{24}$$

$$TC^{HM} = -\frac{3}{2}\sigma_\theta^2\theta_t C_1 + \frac{3}{2}\rho\sigma_v\sigma_\theta^2\theta_t C_2 - \frac{3}{8}\sigma_v^2\sigma_\theta^2\theta_t C_3 - \frac{3}{8}\sigma_\theta^4\theta_t C_4, \tag{25}$$

$$C_1 = \int_t^T \frac{(e^{-\kappa(T-u)} - 1 + \kappa(T-u))^2}{\kappa^2} du, \tag{26}$$

$$C_2 = \int_t^T \frac{1-e^{-\kappa(T-u)}}{\kappa} \int_t^u (1-e^{-\kappa(u-s)}) \frac{e^{-\kappa(T-s)} - 1 + \kappa(T-s)}{\kappa} dsdu, \tag{27}$$

$$C_3 = \int_t^T \frac{(1-e^{-\kappa(T-u)})^2}{\kappa^2} \int_t^u (1-e^{-\kappa(u-s)}) \frac{e^{-\kappa(T-s)} - 1 + \kappa(T-s)}{\kappa} dsdu, \tag{28}$$

$$C_4 = \int_t^T \frac{(e^{-\kappa(T-u)} - 1 + \kappa(T-u))^2}{\kappa^2} \int_t^u \frac{e^{-\kappa(T-s)} - 1 + \kappa(T-s)}{\kappa} dsdu, \tag{29}$$

and the term variance, Var^H, and third cumulant, TC^H, of the Heston model are given by Proposition 3.

Proof. See Appendix B.2.

Remark 4.1. In the extended Heston model, for $\kappa > 0$, if τ 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 τ 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 τ for both small τ and large τ.

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 τ. 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 τ. 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 5 *In a hybrid Merton-Heston model, i.e., 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 τ are as follows

$$\begin{aligned} VIX_{t,\tau}^2 &= (1-\omega)\theta + \omega v_t + 2\lambda(e^x - 1 - x), \qquad \omega = \frac{1 - e^{-\kappa\tau}}{\kappa\tau}, \\ Sk_{t,\tau} &= \frac{TC^H + \lambda x^3\tau}{[Var^H + \lambda x^2\tau]^{3/2}}, \end{aligned} \tag{30}$$

where the term variance, Var^H, and third cumulant, TC^H, of the Heston model are given by Proposition 3.

Remark 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 5.2. In the hybrid Merton-Heston model, for $\kappa > 0$, if τ 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 τ 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 τ and large τ.

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 paper); hence, we have to rely on the stochastic jump intensity, λ_t, to capture the time-varying SKEW. Retaining the affine structure, we propose the following three-factor model.

Proposition 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^λ *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* τ *are as follows:*

$$\begin{aligned}VIX_{t,\tau}^2 &= (1-\omega)\theta^V + \omega V_t,\\ Sk_{t,\tau} &= \frac{TC^H + TC^J}{[Var^H + Var^J]^{3/2}},\end{aligned} \tag{31}$$

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

$$Var^J = \Lambda_t^T x^2\tau + (e^x - 1 - x)^2 E_t(Z_T^2), \tag{32}$$

$$TC^J = \Lambda_t^T x^3\tau - 3x^2(e^x - 1 - x)E_t(Z_T^2) - (e^x - 1 - x)^3 E_t(Z_T^3), \tag{33}$$

the average jump intensity, Λ_t^T, *is given by*

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

[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 6.

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

$$E_t(Z_T^2) = \sigma_\lambda^2 \int_t^T \Big(\frac{1 - e^{-\kappa\tau^*}}{\kappa}\Big)^2 E_t(\lambda_u)du, \qquad \tau^* = T - u, \tag{34}$$

$$E_t(Z_T^3) = 3\sigma_\lambda^4 \int_t^T \frac{1 - e^{-2\kappa\tau^*} - 2\kappa\tau^* e^{-\kappa\tau^*}}{\kappa^3} \frac{1 - e^{-\kappa\tau^*}}{\kappa} E_t(\lambda_u)du, \tag{35}$$

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.

Proof. See Appendix B.3.

Remark 6.1. The mean-reverting speed of λ_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 6.2. As we can see from equations (32) and (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, Λ_t^T; the other one is due to the uncertainty, σ_λ, in jump intensity. There is no interaction term between stochastic volatility and jumps because they are independent.

Remark 6.3. In this three-factor model, for $\kappa > 0$, if τ 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 τ 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 τ and large τ.

As explained earlier, we need two variance factors, i.e., an instantaneous one, v_t, and a long-term one, θ_t^v, to capture the information of the time-varying VIX term structure. Similarly, in order the capture the information of the time-varying SKEW term structure, we also need two factors, i.e., instantaneous and long-term jump intensity, λ_t and θ_t^λ. The simplest five-factor model is presented as follows.

Proposition 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^λ and N_t. The model-implied squared VIX and skewness at time t for time to maturity τ are as follows:

$$\begin{aligned}VIX_{t,\tau}^2 &= (1-\omega)\theta_t^V + \omega V_t,\\ Sk_{t,\tau} &= \frac{TC^H + TC^J + TC^M}{[Var^H + Var^J + Var^M]^{3/2}},\end{aligned} \tag{36}$$

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

$$Var^M = Var^{HM} + (e^x - 1 - x)^2\sigma_2^2\theta_t^\lambda C_1, \tag{37}$$

$$TC^M = TC^{HM} - 3\sigma_2^2\theta_t^\lambda[x^2(e^x - 1 - x)C_1 + (e^x - 1 - x)^3(\sigma_\lambda^2 C_3 + \sigma_2^2 C_4)], \tag{38}$$

and the variance Var^H, third cumulant TC^H, of the Heston model are given by Proposition 3, Var^{HM}, TC^{HM}, C_1, C_3, C_4 are given by Proposition 4, Var^J and TC^J, of jump component are given by Proposition 6.

Remark 7.1. The result is built by combining those of the propositions 4 and 6 and including the additional term variance and third cumulant contributed from the uncertainty of long-term mean jump intensity, θ_t^λ.

Remark 7.2. In this five-factor model, for $\kappa > 0$, if τ 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 τ is large, then the skewness is given by

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

where

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

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

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 τ by using interpolation as follows:

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

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

where ω is a weight determined by

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

and τ_{last} and τ_{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 1. The average one-month SKEW is 118.28, which is equivalent to the skewness -1.828 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 a positive skewness 1.280.

Figure 1 shows the time evolution of the SKEW term structure for time to maturity from one month to 15 months, and Figure 2 shows the time series of three selected SKEWs with constant time to maturity. The sample period is from 2 January 1990 to 31 December 2015. As we can see from the figures, the SKEW term 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 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

[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 is used in computing the 30-day SKEW index. In this paper, we construct constant time to maturity SKEW by using the business day convention.

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.

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 paper, 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.

We use the term structure data of VIX and SKEW to estimate the models proposed in Section 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

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

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 σ^2. We fit the theoretical value of VIX, 100σ, with the VIX term structure data (1284 days with roughly 10 observations each day). The optimal value of σ^2 is 0.0432, with the VIX root mean squared error (RMSE) 5.4698. The theoretical value of SKEW in the Black-Scholes model is 100, and the SKEW RMSE is 26.6709. The estimation results are reported in Table 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 logarithmic returns[10]. Solving the following three equations

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

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), to estimate the model parameters (κ, θ) and latent variable $\{V_t\}$ using the term structure of the CBOE VIX. We follow the same procedure adopted by Luo and Zhang (2012) to estimate the mean-reverting speed κ, θ and daily realizations of instantaneous variance. Specifically, the estimation procedure is as follows:

Step 1 Given an initial value κ and θ, 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,

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

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

[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 ith logarithmic price and δ the first-order difference. We choose the period of a month, i.e., $\tau = 1/12$, to calculate the diffusive variance with a daily partition in our paper. 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 ith 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 and realized variance.

Step 2 Estimate κ and θ with $\{V_t\}$ obtained in step 1 by minimizing the overall objective function for VIX:

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

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). The estimation results are as follows:

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

with VIX RMSE 1.3696. However, the model is not fully estimated as we can only determine κ and θ at the VIX level. To obtain ρ and σ_v, we need the SPX and 30-day VIX data as well as the SKEW term structure data. Specifically, we estimate ρ with $Corr(d \log S_t, dVIX_t^2)$, and σ_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 maturity τ_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 SKEW RMSE 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 κ and two latent variables $\{V_t, \theta_t\}$. The estimation gives

$$\kappa = 1.96,$$

with VIX RMSE 0.5539. At the SKEW level, in addition to the unrestricted estimation results

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

with SKEW RMSE 9.122. Note that the average long-term mean θ (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 ρ 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 ρ 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, θ and θ^v denote their corresponding long-term mean levels. 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 SKEW RMSE 8.779.

In the Merton-extended Heston model, we estimate the parameters based on the VIX estimation results under the extended Heston model. We omit the impact of the stochastic long-term mean on the calculation of ρ, 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 SKEW RMSE 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 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 values in the Merton-Heston model at the SKEW-level estimation to see how the generalization improves the fitting performance. Given an initial value for σ_λ, 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 of the instantaneous VIX square $\theta = 0.149$ 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 of the diffusive variance $\theta^v = 0.142$ from the SKEW-level estimation for the Merton-Heston model, we can estimate 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 σ_λ, as the SKEW formula is insensitive to σ_λ, 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 σ_λ remains unchanged after each optimization for $\{\lambda_t\}$. The SKEW RMSE reduced 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 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 values are from the VIX-level estimation under the extended Heston model and the SKEW level estimation under the Merton-extended Heston model. 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 σ_λ, 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 σ_2, and then add an outer loop to estimate σ_λ 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 σ_λ. The SKEW RMSE reduced from 8.158 in the Merton-extended Heston model to 6.953 in the five-factor model. In Figure 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 are shown in

Figure 7.

In Table 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, and the Merton-Heston (Merton-extended Heston) model outperforms the Heston model by 14% (20%) 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%). 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 log returns; see Das and Sundaram (1999) and Zhang et al. (2016). Note that the nonzero correlation between the SPX logarithmic 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.

6 Conclusion

In this paper, 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 identify 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 returns, 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.

A The First Three Moments of Log Return

Bakshi, Kapadia and Madan (2003) propose a methodology of evaluating the first three moments of log 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

$$\begin{aligned} f(S_T) &= f(K_0) + f'(K_0)(S_T - K_0) + \int_0^{K_0} f''(K)\max(K - S_T, 0)dK \\ &\quad + \int_{K_0}^{+\infty} f''(K)\max(S_T - K, 0)dK, \end{aligned} \tag{39}$$

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 (39) to the power of log return gives

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

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 (5), (6) and (7).

B Model-Implied VIX and Skewness Formulas

B.1 The Merton Model

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

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

We define the logarithmic 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

$$\begin{aligned} E_t^Q[R_t^T - E_t^Q(R_t^T)]^2 &= E_t^Q\Big(\int_t^T \sigma dB_t + x dN_t - \lambda x dt\Big)^2 = (\sigma^2 + \lambda x^2)\tau, \\ E_t^Q[R_t^T - E_t^Q(R_t^T)]^3 &= E_t^Q\Big(\int_t^T \sigma dB_t + x dN_t - \lambda x dt\Big)^3 = \lambda x^3 \tau. \end{aligned}$$

Therefore, the skewness of logarithmic return is given by

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

B.2 The Heston Model with Stochastic Long-Term Mean

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

$$dv_t = \kappa(\theta_t - v_t)dt + \sigma_v \sqrt{v_t} dB_t^v, \tag{40}$$

$$d\theta_t = \sigma_\theta \sqrt{\theta_t} dB_t^\theta, \tag{41}$$

where B_t^θ is independent of B_t^v and B_t^S.

Converting equation (40) into the stochastic integral form and plugging equation (41) yields

$$v_s = E_t(v_s) + \sigma_v \int_t^s e^{-\kappa(s-u)}\sqrt{v_u}dB_u^v + \sigma_\theta \int_t^s (1-e^{-\kappa(s-u)})\sqrt{\theta_u}dB_u^\theta, \tag{42}$$

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. (2016), 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 (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 long-term mean variation explicitly, using the independency of B_s^θ and the martingale property of Y_s^H, Y_s^M and θ_s, we expand and simplify the variance and third cumulant of R_t^T in Zhang et al. (2016) 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. (2016), 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 θ_s as well as v_s are given by*

$$
E_t^Q(Y_s^M \theta_s) = \sigma_\theta^2 \theta_t \int_t^s \frac{e^{-\kappa(T-u)} - 1 + \kappa(T-u)}{\kappa} du, \tag{43}
$$

$$
E_t^Q(Y_s^M v_s) = \sigma_\theta^2 \theta_t \int_t^s (1 - e^{-\kappa(s-u)}) \frac{e^{-\kappa(T-u)} - 1 + \kappa(T-u)}{\kappa} du. \tag{44}
$$

Proof. See Appendix B.2.1.

Using Ito's Isometry and the martingale property of θ_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 $C_2 = \int_t^T \frac{1-e^{-\kappa(T-u)}}{\kappa} \int_t^u (1-e^{-\kappa(u-s)}) \frac{e^{-\kappa(T-s)}-1+\kappa(T-s)}{\kappa} ds du.$

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

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

B.2.1 Proof of Lemma 1

Using Ito's Lemma and martingale property of Y_u^M and θ_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 (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

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

Solving the ordinary differential equation (ODE) gives

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

Taking differentiation with respect to s gives

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

which is equivalent to equation (44). This completes the proof.

B.3 The Jump Diffusion Models

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

$$\frac{dS_t}{S_t} = rdt + \sqrt{v_t}dB_t^S + (e^x-1)dN_t - \lambda_t(e^x-1)dt, \tag{45}$$

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

Applying Ito's Lemma to equation (45) gives

$$d\ln S_t = \Big[r - \frac{1}{2}v_t - \lambda_t(e^x - 1 - x)\Big]dt + \sqrt{v_t}dB_t^S + xdN_t - \lambda_t x dt. \qquad (46)$$

The logarithmic return from current time t, to a future time, T, is defined by

$$R_t^T \equiv \ln\frac{S_T}{S_t} = \int_t^T \Big[r - \frac{1}{2}v_u - \lambda_u(e^x - 1 - x)\Big]du + \sqrt{v_u}dB_u^S + xdN_u - \lambda_u x du. \qquad (47)$$

The conditional expectation at time t is then given by

$$E_t^Q(R_t^T) = \int_t^T \Big[r - \frac{1}{2}E_t^Q(v_u) - E_t^Q(\lambda_u)(e^x - 1 - x)\Big]du. \qquad (48)$$

Following the notations in Zhang et al. (2016), 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 (48) from (47) yields

$$R_t^T - E_t^Q(R_t^T) = X_T - \frac{1}{2}Y_T + xI_T - (e^x - 1 - x)Z_T. \qquad (49)$$

Assuming that there is no jumps in variance, and using the incremental independency property of a Poisson process, 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. (2016) 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 the Heston model.

References

[1] 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.

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

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

[4] 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.

[5] 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.

[6] Black, Fischer, and Myron Scholes, 1973, The pricing of options and corporate liabilities, *Journal of Political Economy* 81(3), 637-654.

[7] Boyer, Brian, Todd Mitton, and Keith Vorkink, 2010, Expected idiosyncratic skewness, *Review of Financial Studies* 23(1), 169-202.

[8] Branger, Nicole, Holger Kraft, and Christoph Meinerding, 2016, The dynamics of crises and the equity premium, *Review of Financial Studies* 29(1): 232-270.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] Conrad, Jennifer, Robert F. Dittmar, and Eric Ghysels, 2013, Ex ante skewness and expected stock returns, *Journal of Finance* 68(1), 85-124.

[15] 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.

[16] 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.

[17] Dennis, Patrick, and Stewart Mayhew, 2002, Risk-neutral skewness: Evidence from stock options, *Journal of Financial and Quantitative Analysis* 37(3), 471-493.

[18] Duffie, Darrell, Jun Pan, and Kenneth Singleton, 2000, Transform analysis and asset pricing for affine jump-diffusions, *Econometrica* 68(6), 1343-1376.

[19] 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.

[20] Eraker, Bjørn, Michael Johannes, and Nicholas Polson, 2003, The impact of jumps in volatility and returns, *Journal of Finance* 58(3), 1269-1300.

[21] Eraker, Bjørn, 2004, Do stock prices and volatility jump? Reconciling evidence from spot and option prices, *Journal of Finance* 59(3), 1367-1403.

[22] Faff, Robert W., and Zhangxin Liu, 2014, Hitting SKEW for SIX, Working paper, University of Western Australia.

[23] 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.

[24] Han, Bing, 2008, Investor sentiment and option prices, *Review of Financial Studies* 21(1), 387-414.

[25] Harvey, Campbell R., and Akhtar Siddique, 2000, Conditional skewness in asset pricing tests, *Journal of Finance* 55(3), 1263-1295.

[26] 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.

[27] Liu, Jun, and Jun Pan, 2003, Dynamic derivative strategies, *Journal of Financial Economics* 69(3), 401-430.

[28] Luo, Xingguo, and Jin E. Zhang, 2012, The term structure of VIX, *Journal of Futures Markets* 32(12), 1092-1123.

[29] Merton, Robert C., 1976, Option pricing when underlying stock returns are discontinuous, *Journal of Financial Economics* 3(1-2), 125-144.

[30] 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.

[31] Siegel, Andrew F., 1979, The noncentral chi-squared distribution with zero degrees of freedom and testing for uniformity, *Biometrika* 66(2), 381-386.

[32] Wang, Zhiguang, and Robert T. Daigler, 2014, VIX and SKEW Indices for SPX and VIX Options, Working paper, South Dakota State University.

[33] Zhang, Jin E., Huimin Zhao, and Eric C. Chang, 2012, Equilibrium asset and option pricing under jump diffusion, *Mathematical Finance* 22(3), 538-568.

[34] Zhang, Jin E., Fang Zhen, Xiaoxia Sun, and Huimin Zhao, 2016, The skewness implied in the Heston model and its application, *Journal of Futures Markets*, Forthcoming.

[35] 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.

Table 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 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 return 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 the logarithmic return and the change of diffusive variance is calculated using the S&P 500 index (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	
	κ	ρ	x	σ_v	σ_λ	σ_1	σ_2	v	λ	θ^v	θ^λ	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 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.

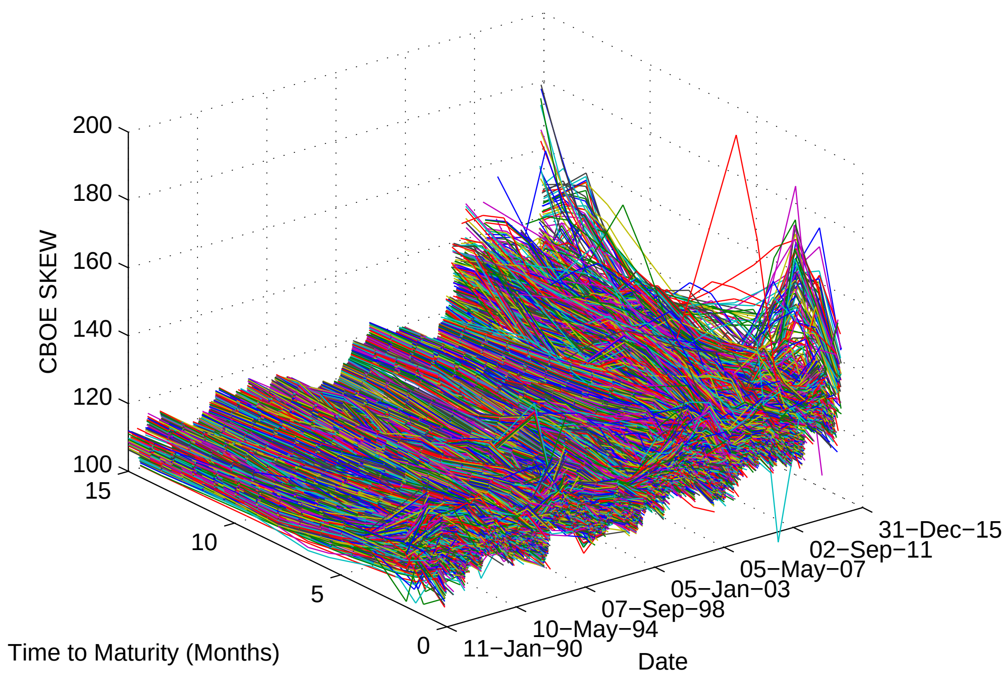


Figure 2: The Time Series of the CBOE SKEW with Fixed Times to Maturity
This graph shows the time series of the CBOE SKEW with one month, six months and 15 months to maturity. The sample period is from 2 January 1990 to 31 December 2015. We only consider the days when the maximum time to maturity is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

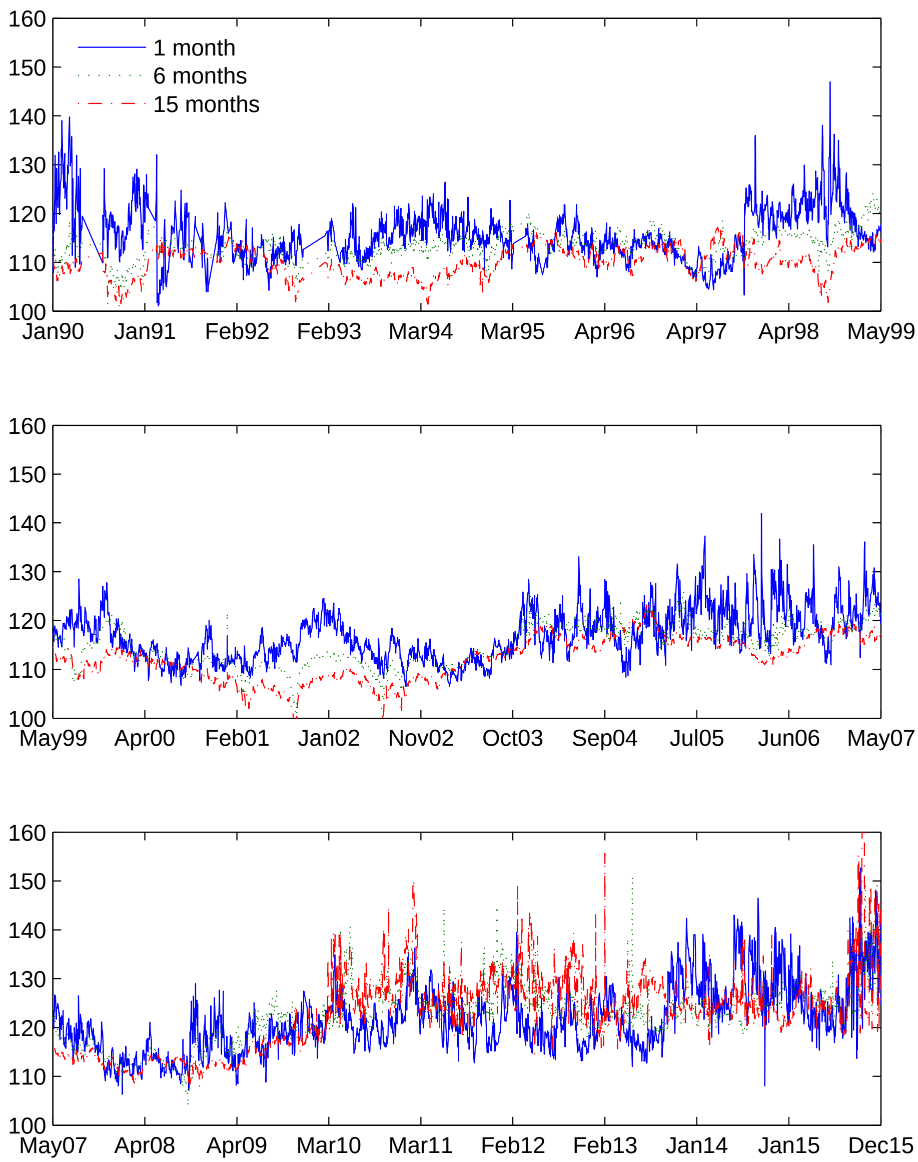


Figure 3: A few Samples of the CBOE SKEW Term Structure with Outliers
We choose the days when the CBOE SKEW is above 180 or below 90 for the period 2 January 1990 to 31 December 2015.

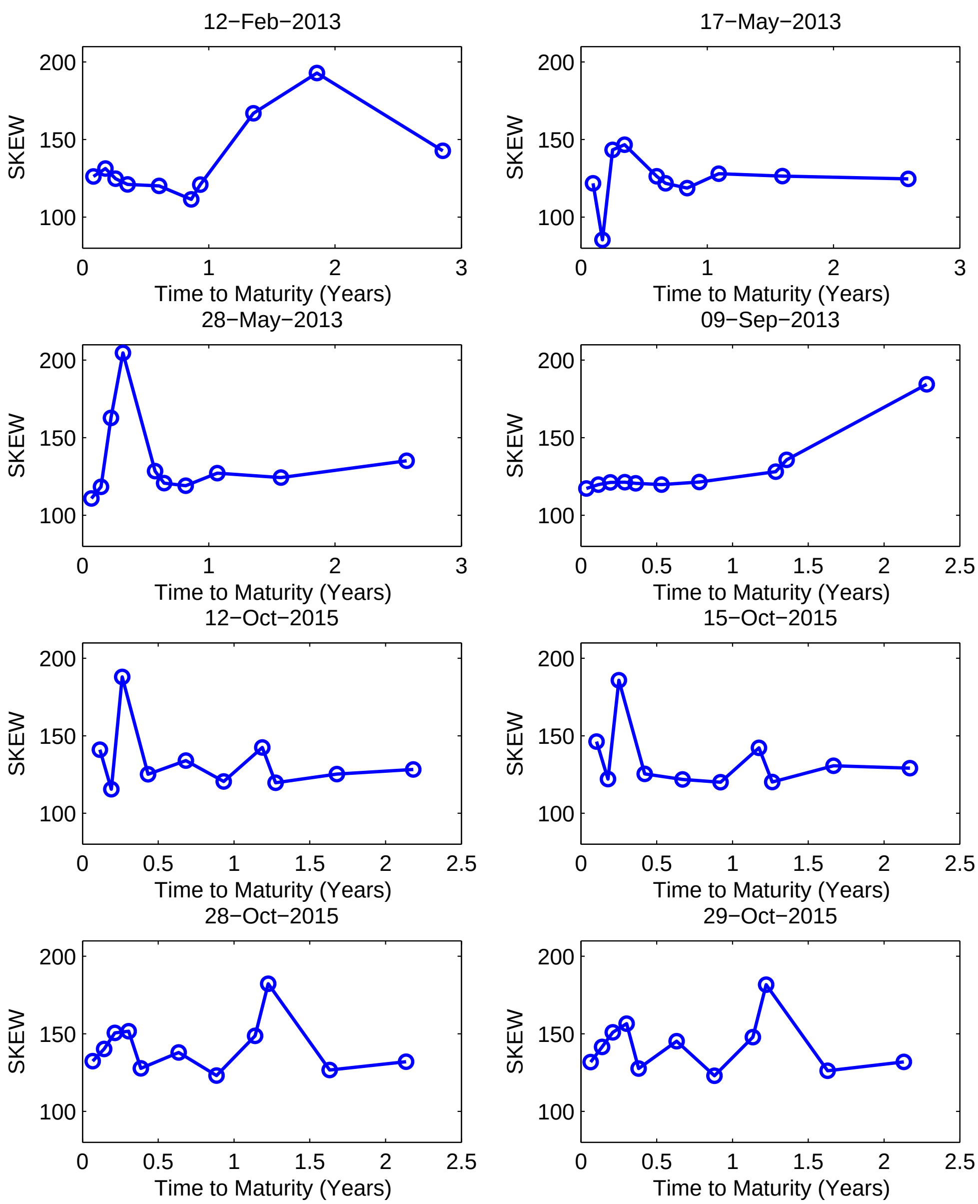


Figure 4: Model-Implied Aggregate Regularized Skewness.
The regularized skewness is defined as the product of the skewness and the square root of the time to maturity. The blue solid line represents the model-implied regularized skewness in the Merton-Heston model (upper) or the Merton-extended Heston 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 green dashed line represents the linear asymptote of the model-implied regularized skewness at zero and the red dotted line represents the Merton-implied regularized skewness. The sample period is from 24 November 2010 to 31 December 2015.

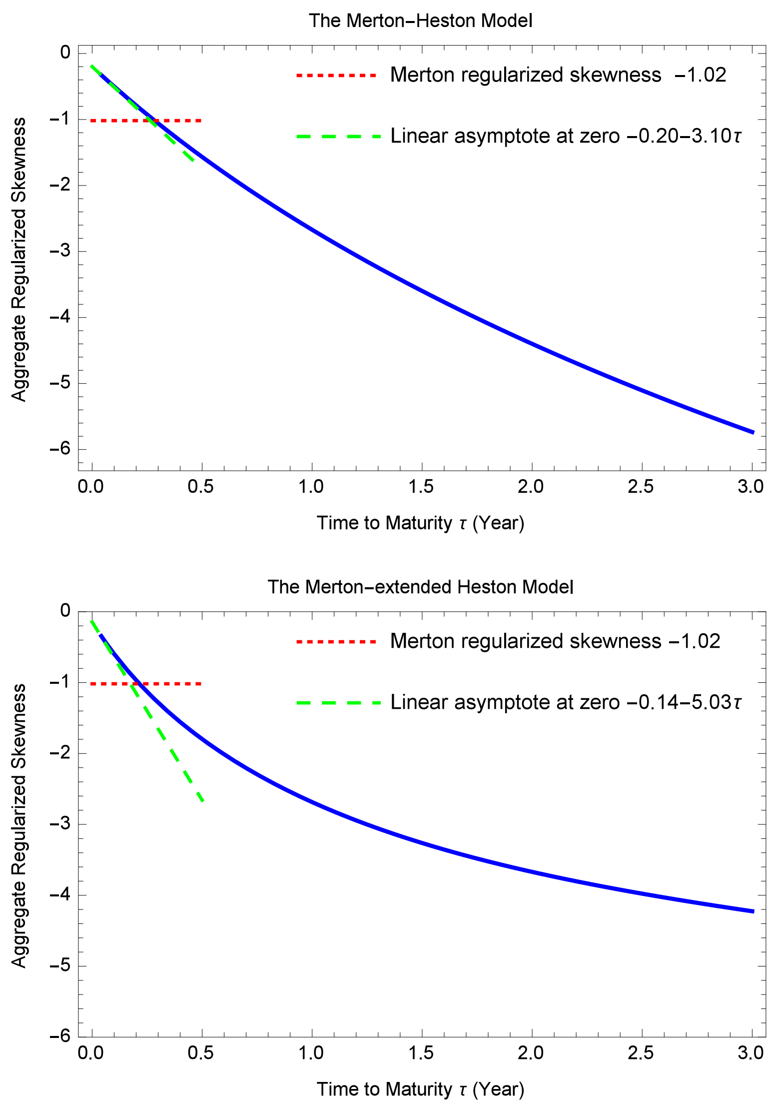


Figure 5: The Three-Factor Model

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 logarithmic returns and 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.

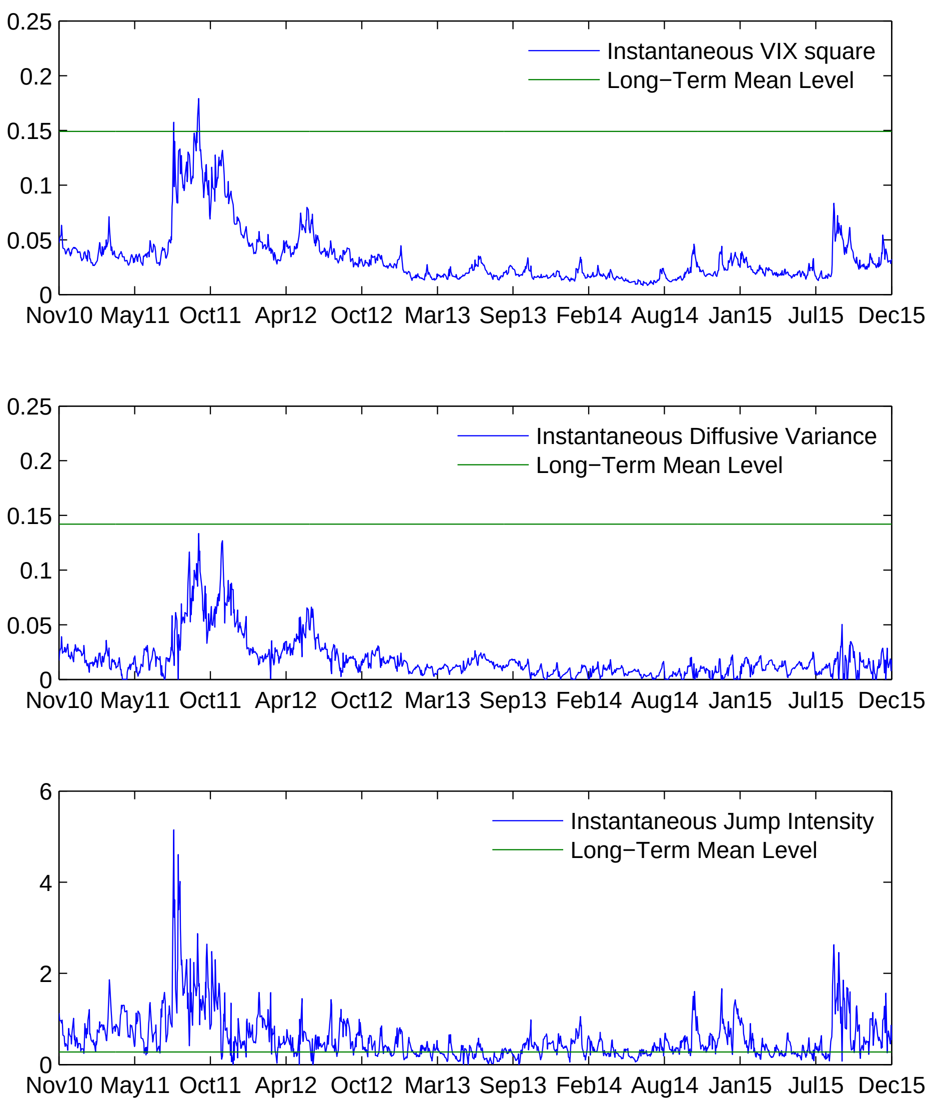
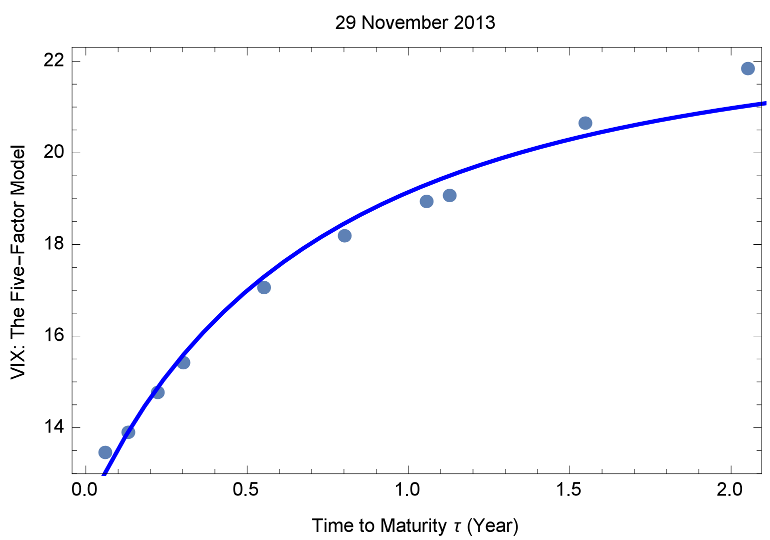


Figure 6: Fitting Performance on a Representative Date

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 logarithmic returns and 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 VIX RMSE 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 SKEW RMSE 1.442.



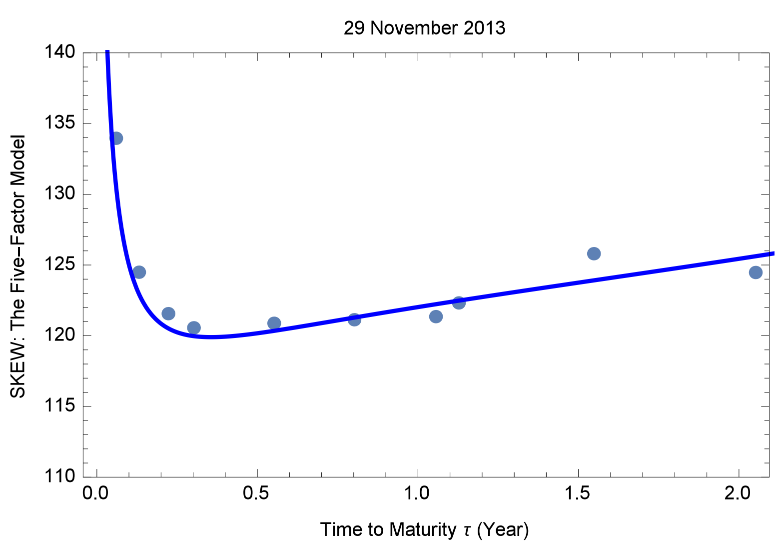


Figure 7: The Five-Factor Model

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 logarithmic returns and 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.