Ann. Data. Sci. (2015) 2(4):471–487
DOI 10.1007/s40745-015-0058-4

The Information Content of OVX for Crude Oil Returns Analysis and Risk Measurement: Evidence from the Kalman Filter Model

Yanhui Chen[1] · Kaijian He[2,3] · Lean Yu[2]

Received: 5 November 2015 / Revised: 4 December 2015 / Accepted: 6 December 2015 /
Published online: 17 December 2015

Abstract Crude oil volatility index (OVX) is a new index published by Chicago Board Option Exchange since 2007. In recent years it emerged as an important alternative measure to track and analyze the volatility of future oil prices. In this paper we firstly model and analyze the dynamic relationship between OVX changes and future crude oil price returns with time-varying coefficients, modeled using the Kalman filter, in the regression models. Empirical results show a weak negative relationship between OVX changes and future crude oil price returns movement, and extremely high/low levels of OVX cannot predict future positive/negative returns well. Secondly, this paper explores whether OVX can predict future realized volatility of crude oil price returns . The empirical findings suggest that OVX serves as an unbiased but not an efficient estimate of the future realized volatility and it includes information of the future realized volatility. Finally the incorporation of information of OVX in measuring market risk is analyzed. The empirical result indicates that Kalman filter based model provides the improved performance than the linear regression model in terms of forecasting accuracy for realized volatility prediction and the reliability for VaR estimate.

✉ Kaijian He
kaijian.he@my.cityu.edu.hk

Yanhui Chen
chenyh@shmtu.edu.cn

Lean Yu
yulean@amss.ac.cn

1 School of Economics and Management, Shanghai Maritime University, No. 1550 Haigang Avenue, Pudong 201306, Shanghai, People's Republic of China

2 School of Economics and Management, Beijing University of Chemical Technology, Beijing 100029, People's Republic of China

3 Department of Management Sciences, City University of Hong Kong, Tat Chee Avenue, Kowloon Tong, Kowloon, Hong Kong

Keywords Implied volatility · OVX · Crude oil price · Future realized volatility · Kalman filter

1 Introduction

The price of crude oil fluctuates severely over a wide range in the past ten years. From around \$60 per barrel in June 2006, the price of West Texas Intermediate (WTI) crude oil peaked at \$145.31 in July 2008, falled to \$30.28 6 months later, and then it fluctuated in the range of \$60 to \$110 for about 5 years. In the third and fourth quarter of 2014, the crude oil price plummeted to \$50. The instability of oil price increases the risk exposure for both investors and crude oil exploration and processing industries.

Chicago Board Option Exchange (CBOE) publishes the Crude Oil Volatility Index (OVX) which is calculated with the VIX methodology from May 2007. OVX is an up-to-the-minute market estimate of the expected 30-day volatility of crude oil prices and it has been calculated for options on the United States Oil Fund. It uses real-time bid/ask quotes of nearby and second nearby options with at least 8 days to expiration, and weights these options to derive a constant, 30-days measure of expected volatility. Empirical studies suggest that OVX provides a new measure to analyze the variance of future oil prices. In this paper, we examine the information content of OVX regarding the crude oil spot price returns and future realized volatility.

The relationship between implied volatility and underlying stock market returns has received significant interests in the literature. It has been well documented that there is a strong asymmetric negative contemporaneous relationship between the changes of the implied volatility index and the underlying index returns. Fleming, Ostdiek [1],Whaley [2,3], Simon [4], Giot [5] and Car and Wu [6] all identified the statistically significant negative and asymmetric contemporaneous relationship between VIX and S&P 100 (or S&P 500) returns. Furthermore, an analogous relationship has also been documented in other equity markets. For example, Simon [4] and Giot [5] studied Nasdaq, Dowling and Muthuswamy [7] focused on the Australian stock market, Siriopoulos and Fassas [8] studied FTSE 100 and later Chen and Lai [9] investigated the relationship between VHSI and HSI. Whether implied volatility can predict future asset returns or not is another research issue that attracted significant attentions in the literature, for example, Giot [5] investigated the information content of VIX/VXN levels and future S&P 100/Nasdaq 100 index returns; Banerjee et al. [10] assessed the relationship between future returns and current implied volatility levels and innovations.

How much incremental information the implied volatility brings to future realized volatility of underlying asset returns has also been discussed widely in the literature. Through a comprehensive review of ninety-three papers dealing with volatility forecast performance of several approaches including historical, stochastic and implied volatility based forecasts, Poon and Granger [11] documented the general conclusion that estimates of future realized volatility based on implied volatility often outperformed alternative methods. Subsequent to the endeavor of Poon and Granger, recent researches dealing with the US equity markets suggested again that, implied volatility is usually a superior estimator of future volatility. Specifically, Giot [12] assessed the information content of VIX and VXN as volatility estimators in a daily market risk evaluation framework. His findings suggested that their forecasts provided meaningful results. Jiang and Tian [13] demonstrated the model-free implied variance, reflected by the new VIX, was a more efficient forecast for future volatility. Likewise, Fleming, Ostdiek and Whaley [1] and Christenson and Prabhala [14] used regression model to test whether VIX (VXO) is a stock market volatility predictor. Corrado and Miller [15] showed that the CBOE implied volatility indices (VXO, VIX and VXN) outperformed historical volatility as estimators of future realized volatility of the corresponding underlying indices (S&P100, S&P500 and Nasdaq100). In contrast to the general conclusions, Canina and Figlewski [16] examined volatility implied in S&P 100 index options from 1983 to 1987 and suggested that implied volatility is a poor indicator of subsequently realized volatility. Becker and Clements [17] indicated that combinations of various model-based forecasts of S&P 500 realized volatility outperformed its implied volatility index-VIX. Outside of the US, recent empirical findings were mixed and suggested a slight lead of implied volatility. In particular, Dowling and Muthusswamy [7] found that their implied volatility index for Australian stock market under-performed historical volatility as a predictor of future realized volatility. Similarly, Siriopoulos and Fassas [8] suggested that VFTSE included information about future volatility beyond that contained in past volatility. Interested readers are referred to Siriopoulos and Fassas [18] for the comprehensive review on the incremental information content of all the publicly available implied volatility indices across the world to realized volatility of the corresponding underlying equity markets.

Many economic relationship is dynamic and time varying, which requires estimation of the state of a system that changes over time using a sequence of noisy measurements made on the system. Previous researches have analyzed the relationship either by using static linear regression model, or by dividing sample periods into several distinct sub-periods. This static approach cannot capture the time evolvement of the relationship. Also the static coefficients cannot reflect the immediate changes of two sides, which may reduce the forecasting accuracy. Many researchers have applied the Kalman filter to evaluate the dynamic evolvement (see [19–22] and [23] etc.). Using time-variant parameters obtained from Kalman filter can reveal the hidden dynamic relationship and evaluate it more accurately.

The research in this paper focuses on the information content of OVX for both crude oil spot price returns and future realized volatility using Kalman filter.The contributions of this paper are two folds. Firstly we found that OVX contains information for the future realized volatility of crude oil returns. The shocks of the information on the crude oil returns and risk movement are dynamic and time varying. We propose the regression model with the time varying coefficients to model them. Secondly we found that the extraction and modeling of the hidden information in the OVX, using the Kalman filter model, can help improve the forecasting performance of the future realized volatility and VaR estimated, in terms of reliability and accuracy.

The rest of the paper is structured as follows. Section 2 provides a brief introduction of Kalman filter. Section 3 examines the relationship between OVX and 1,5, 10, 20 and 60-day forward crude oil spot price returns. Section 4 investigates whether OVX can predict future realized volatility and the application in calculating VaR is also discussed in this section. Finally, a brief conclusion is provided in Sect. 5.

2 Methodology

Kalman filter defines and solves a recursive solution to the discrete linear filtering problem, with the underlying assumption that the posterior density at every time step is Gaussian. The mean and covariance can then be used to describe the underlying distribution [24]. The basic notion of Kalman filter model is to estimate the state of a process, in a way that minimizes the mean squared error [25].

The Kalman filter model is defined as follows:

$$x_t = A_t x_{t-1} + w_{t-1} \tag{1}$$

$$z_t = H_t x_t + v_t \tag{2}$$

where A_t and H_t are known matrices defining the linear functions. w_{t-1} and v_t refer to the random variables for the process and measurement noise respectively. They are assumed to be independent and normally distributed, and follow random walk process, i.e., $w_t \sim N(0, Q_t)$ and $v_t \sim N(0, R_t)$.

For the constant R_t, we can take some off-line sample measurements in order to determine the variance of the measurement noise. The maximum likelihood method is used in each step to estimate R_t [26].

If the elements of Q_t are set with very small numbers, it will require more data and longer time for the estimator to converge. The state variables cannot respond instantaneously to the new measurement. The maximum likelihood method is used to estimate Q_t [26].

We define $\hat{x}_t^- \in R^n$ as a prior state estimate at step t, given prior information of the process before step t. $\hat{x}_t \in R^n$ is defined as a posterior state estimate at step t given measurement z_t. We also define prior and posterior estimate errors as $e_t^- \equiv x_t - \hat{x}_t^-$ and $e_t \equiv x_t - \hat{x}_t$.

The prior estimate error covariance is given by $P_t^- = E\left[e_t^- e_t'\right]$. The posterior estimate error covariance is $P_t = E\left[e_t e_t'\right]$.

In order to derive the equations for the Kalman filter, we begin with the goal of finding an equation that computes posterior state estimate $\hat{x}_t$ as a linear combination of prior $\hat{x}_t^-$, a weighted difference between an actual measurement z_t and a measurement prediction $H\hat{x}_t^-$ as shown below.

$$\hat{x}_t = \hat{x}_t^- + K\left(z_t - H\hat{x}_t^-\right) \tag{3}$$

where $\left(z_t - H\hat{x}_t^-\right)$ is the measurement innovation, or the residual. The matrix K refers to the gain that minimizes the posterior error covariance. [27] and [28] shows one form K that minimizes P_t in Eq. (4):

$$K_t = P_t^- H' \left(H P_t^- H' + R\right)^{-1} \tag{4}$$

Ultimately the filtering process would lead to the conditional expectation of the state vector given the information set at time t.

3 OVX and Crude Oil Price Returns

3.1 Data Description

In this paper, daily data of OVX and crude oil spot prices, from 1 June 2007 to 31 July 2015, are used for the analysis of the contemporaneous relationship. OVX used in this study is obtained from Yahoo finance and crude oil spot prices are obtained from U.S. Energy Information Administration. Both data sets are made publicly available at the website of their respective publishers. Although the data set of crude oil spot prices includes WTI spot prices and Europe Brent spot prices, this research uses WTI spot prices only to represent the oil price, since these two data series moves synchronously.

We plot the OVX and crude oil spot price from 1 June 2007 to 31 July 2015 in Fig. 1. It shows that when oil prices were at a high level, the corresponding OVX values were usually very small. Secondly, when oil prices dropped suddenly, OVX values rocketed up during the same period. Thirdly, oil prices usually grew slowly for a period of time after it dropped, which indicates, the high level OVX values may forebode future positive oil price returns.

Following Simon's work [4], this study uses daily logarithmic changes for OVX and oil price. Correspondingly, we define $r_{OVX,t} = \ln(OVX_t) - \ln(OVX_{t-1})$ and $r^1_{Oil,t} = \ln(Oil\_Price_t) - \ln(Oil\_Price_{t-1})$ as the daily logarithmic returns. One advantage of using logarithms is that the logarithms can avoid negative estimation of volatility and oil price.

Descriptive statistics for OVX and crude oil spot price are calculated and listed in Table 1. The table reports that the daily logarithm returns of OVX are stationary time series with mean approximately equal to zero. As for OVX, the null hypothesis of Jarque–Bera test of normality is rejected.Table 1 also contains summary statistics of daily logarithm returns of oil prices. The means of logarithm returns are very close to zero. The Jarque–Bera statistic robustly rejects the normal distribution hypothesis.

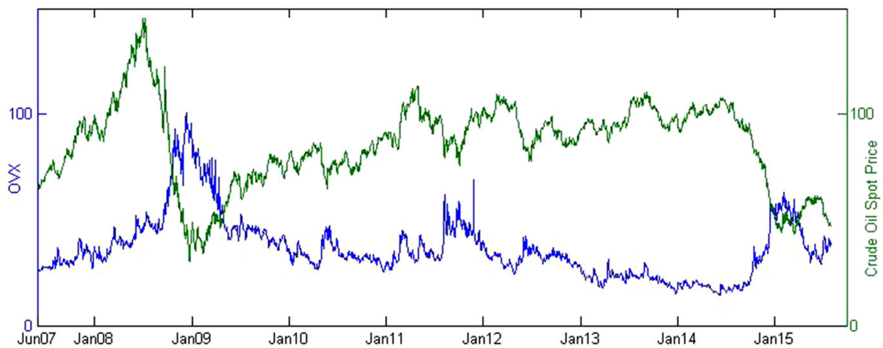

Fig. 1 OVX (*left axis*) versus crude oil spot price (*right axis*) from 1 June 2007 to 31 July 2015

Table 1 Descriptive statistics

Statistics	OVX	$r_{OVX,t}$	Oil	$r^1_{Oil,t}$
Obs.	2059	2058	2059	2058
Mean	36.472	0.000	85.631	−0.000
Max	100.420	0.425	145.310	0.164
Min	14.500	−0.440	30.280	−0.128
Median	33.450	−0.004	88.990	0.000
Std. Dev.	14.616	0.049	19.896	0.024
Skewness	1.471	0.847	−0.318	0.038
Kurtosis	5.849	14.190	3.107	8.484
Jarque–Bera	1438.744	10984.345	35.741	2579.074

3.2 Empirical Analysis

In the literature, the implied volatility index is usually regarded as "investor fear gauge" [3] since it spikes during periods when the market is in turmoil. Siriopoulos and Fassas [18] tested the relationship between stock market returns and implied volatility by a regression analysis of daily changes of the volatility index against the daily positive and negative returns of the corresponding underlying stock index. Giot[5] classified implied volatility levels with respect to their 2-year rolling history and then used the dummy variables to assess the relationship between implied volatility and forward looking stock index returns. He found extremely high (low) levels of the implied volatility indices indeed contained information to future returns.

In this paper, we analyze the relationship between logarithm returns of OVX and oil price with Kalman filter which assumes that the coefficients are time-varying instead of static. Since Giot [5] suggested positive (negative) forward looking returns of S&P100 or Nasdaq100 can be triggered by extremely high (low) levels of the implied volatility indices. We use two dummy variables ($D_{high,t}$ and $D_{low,t}$) to indicate extreme large and low levels of the implied volatility indices. Following Giot's [5] method, we also use the 20 equally spaced percentiles based on a rolling 2-year history of OVX to determine whether OVX is extremely large or low. Particularly, if OVX_t falls into the largest percentile $D_{high,t}$=1, and if OVX_t falls into the smallest percentile $D_{low,t}$=1. In order to distinguish the different impact of positive and negative previous returns, asymmetric effect is also considered in this research. The simple regression models are as follows:

$$r^1_{Oil,t} = \beta_1 r^+_{OVX,t-1} + \beta_2 r^-_{OVX,t-1} + \gamma_1 D_{high,t-1} + \gamma_2 D_{low,t-1} + \varepsilon_t \quad (5)$$
$$r^5_{Oil,t} = \beta_1 r^+_{OVX,t-5} + \beta_2 r^-_{OVX,t-5} + \gamma_1 D_{high,t-5} + \gamma_2 D_{low,t-5} + \varepsilon_t \quad (6)$$
$$r^{10}_{Oil,t} = \beta_1 r^+_{OVX,t-10} + \beta_2 r^-_{OVX,t-10} + \gamma_1 D_{high,t-10} + \gamma_2 D_{low,t-10} + \varepsilon_t \quad (7)$$
$$r^{20}_{Oil,t} = \beta_1 r^+_{OVX,t-20} + \beta_2 r^-_{OVX,t-20} + \gamma_1 D_{high,t-20} + \gamma_2 D_{low,t-20} + \varepsilon_t \quad (8)$$
$$r^{60}_{Oil,t} = \beta_1 r^+_{OVX,t-60} + \beta_2 r^-_{OVX,t-60} + \gamma_1 D_{high,t-60} + \gamma_2 D_{low,t-60} + \varepsilon_t \quad (9)$$

where $r^{+}_{OVX,t-i}$ equals the daily OVX return when the i-day previous OVX return is positive, and equals to zero when the i-day previous OVX return is negative. $r^{-}_{OVX,t-i}$ goes just the opposite. And $r^{1}_{Oil,t}$, $r^{5}_{Oil,t}$, $r^{10}_{Oil,t}$, $r^{20}_{Oil,t}$, $r^{60}_{Oil,t}$ are the forward looking 1-, 5-, 10-, 20- and 60-day relative changes in the level of the crude oil spot price, which means $r^{i}_{Oil,t} = \ln(Oil\_Price_t) - \ln(Oil\_Price_{t-i})$.

Time varying modification with Kalman filter, the transition function is given by Eq. (10).

$$\begin{bmatrix} \beta_{1,t+1} \\ \beta_{2,t+1} \\ \gamma_{1,t+1} \\ \gamma_{2,t+1} \end{bmatrix} = \begin{bmatrix} 1\ 0\ 0\ 0 \\ 0\ 1\ 0\ 0 \\ 0\ 0\ 1\ 0 \\ 0\ 0\ 0\ 1 \end{bmatrix} \begin{bmatrix} \beta_{1,t} \\ \beta_{2,t} \\ \gamma_{1,t} \\ \gamma_{2,t} \end{bmatrix} + w_t \tag{10}$$

The measurement function is given by Eq. (11).

$$r^{i}_{oil,t} = \left[r^{+}_{OVX,t-1}\ r^{-}_{OVX,t-1}\ D_{high,t-i}\ D_{low,t-i} \right] \begin{bmatrix} \beta_{1,t} \\ \beta_{2,t} \\ \gamma_{1,t} \\ \gamma_{2,t} \end{bmatrix} + v_t \tag{11}$$

As Harvey [29] indicated, when the process equation is non-stationary the initial distribution of the state variables should be specified in terms of diffuse prior. Here we consider the case when Q and R are determined by the first four hundred data points.

Estimates of β_1 and β_2 are described in Table 2. For the five sets of estimates the mean values of both β_1 and β_2 are negative, which indicates positive (negative) OVX changes associate with negative (positive) oil price returns, however, the standard deviations of β_1 and β_2 are larger and larger, which indicates as time passed by the impact of OVX changes is diminishing. Moreover, the differences between β_1 and β_2 are not consistent which cannot testify the asymmetric effect.

Table 2 also presents the estimates of γ_1 and γ_2. γ_1 indicates the impact of extremely high level OVX to oil price returns and γ_2 indicates the impact of extremely low level OVX to oil price returns. We can find that although γ_1 is positive and γ_2 is negative for all the five equations, the standard deviations are very large, which cannot provide the convincing evidence that extremely high level OVX is a signal of future positive returns and extremely low level OVX is a signal of future negative returns. This result is different from the result found by Giot [5] based on VIX/VXN. As Giot [5] summarized extremely high levels of implied volatility is a signal of 'buy' entry points for traders who want to take long positions in the underlying index. This result doesn't lend strong support for Giot's result when using OVX and crude oil spot prices since crude oil price is also affected by global demand, supply and inventory of world's leading countries (see [30,31] and [32] etc.) which is different from equity market. The outcome of taking extremely high/low level of OVX as trading signals is shown in Table 3. Results in Table 3 show that extremely high/low level of OVX doesn't indicate that crude oil market are oversold/overbuy at least using existing data.

Table 2 Coefficients estimation results

		β_1	β_2	γ_1	γ_2
1-Day forward looking returns					
Kalman	Mean	−0.1686	−0.1417	0.0189	0.0043
Filter	Std. Dev.	0.1600	0.1299	0.0155	0.0111
5-Day forward looking returns					
Kalman	Mean	−0.1281	−0.1683	0.0500	−0.0035
Filter	Std. Dev.	0.2698	0.3452	0.0469	0.0269
10-Day forward looking returns					
Kalman	Mean	−0.1074	−0.1845	0.0071	−0.0184
Filter	Std. Dev.	0.3987	0.6118	0.0484	0.0297
20-Day forward looking returns					
Kalman	Mean	−0.0510	−0.2531	0.0662	−0.0537
Filter	Std. Dev.	0.8033	1.2086	0.0749	0.0538
60-Day forward looking returns					
Kalman	Mean	−0.1396	−0.0381	0.1542	−0.0145
Filter	Std. Dev.	2.2810	3.4210	0.1046	0.0901

Table 3 Outcome of taking extremely high/low level of OVX as trading signals

		1-Day	5-Day	10-Day	20-Day	60-Day
high	r	−0.0009	−0.0111	−0.0376	−0.0563	−0.0277
#149	Std. Dev.	0.0399	0.0995	0.1138	0.1589	0.2747
low	r	−0.0007	−0.0028	−0.0048	−0.0109	0.0013
#264	Std. Dev.	0.0122	0.0303	0.0422	0.0662	0.1006

#149 and #264 indicate the trading numbers

4 OVX and Future Realized Volatility of Crude Oil

4.1 Data

Firstly we estimate the realized volatility to test the information content of implied volatility and evaluate how it can help forecast realized volatility. Following Siriopoulos and Fassas [18], we compute the squared return without mean-reversion assumption. We also use non-overlapping observations and compute the realized volatility, RV_m , separately for each calendar month. This method can improve the predictive power of implied volatility and avoid over estimation of past volatility. In particular, the ex-post realized volatility during the next calendar month is calculated according to the following equation:

$$RV_m = \sqrt{\frac{365}{n_m}\sum_{t=1}^{N_m}(r_{Oil,t})^2} \tag{12}$$

where R_t is the return of stock index on day t and n_m is the number of calendar days in month m. We annualize the volatility according to the actual 365 days counting convention, since the calculation of volatility index is based on calendar days instead of trading days [33].

According to the calculation method (see [33]) of implied volatility index, the volatility index recorded at the close of the last trading day in month $m-1$ can represent the market forecast of future volatility of the underlying index in month m, in essence. In this analysis we use OVX_m to denote the volatility index recorded at the close of the last trading day of month m .

Moreover, the information content of OVX at a relatively short time-horizon (5 days and 10 days) is examined as follows. By definition, the forward looking time horizon of implied volatility index is approximately equal to 30 calendar days and the implied volatility index is expressed in annualized terms. Therefore, we should transform annualized implied volatility index to the required 5- and 10-day interval by square root time rule. The forward looking implied volatility index is presented in Eq. (13).

$$IV_{i,t} = \sqrt{\frac{i}{365}} OVX_t \tag{13}$$

Thus $IV_{i,t}$ is the expected volatility over the next i days.

5-day (10-day) forward-looking realized volatility is computed by taking the square root of the sum of the (future) squared returns over $[t+1, t+5]$ ($[t+1, t+10]$) period as in Eq. (14).

$$RV_{i,t} = \sqrt{\sum_{j=1}^{i} r_{t+j}^2}\, , \; i = 5, 10 \tag{14}$$

As for the 5- and 10- day realized volatility, they are both defined from non-overlapping data. As noted by Christensen and Prabhala [14], the use of realized volatility calculated from overlapping data in regression analysis may potentially lead to strong auto-correlation problems in the regression's residuals.

Table 4 presents the descriptive statistics of volatility and log-volatility series. According to Christenson and Prabhala [14] and Hansen [34], log-transformed data brings skewness and kurtosis of their volatility data closer to a normal distribution. However, Fleming [35], Fleming, Ostdiek and Whaley [1] and other studies used untransformed volatility data. Corrado and Miller [15] performed parallel regressions using both the original volatility measures and the log-transformed volatility measures. But by the definition of implied volatility index, the original volatility measures can be interpreted directly. Our sample contains 411 non-overlapping 5-day observations, 205 non-overlapping 10-day observations and 98 non-overlapping monthly observations on realized volatility and implied volatility covering the period from June 2007 to July 2015. All distributions exhibit positive skewness and kurtosis and the null hypothesis of Jarque–Bera test of normality is rejected. This indicates the return deviates from the normal distribution.

Table 4 Descriptive statistics

Statistics	OVX_m	RV_m	$IV_{5,t}$	$RV_{5,t}$	$IV_{10,t}$	$RV_{10,t}$
Obs.	98	98	411	411	205	205
Mean	36.402	33.591	4.266	4.492	6.012	6.557
Max	88.930	119.330	11.579	21.664	15.402	27.902
Min	15.610	11.121	1.725	0.736	2.440	1.703
Median	33.640	27.844	3.907	3.700	5.504	5.469
Std. Dev.	14.671	19.403	1.728	3.093	2.407	4.071
Skewness	1.459	2.039	1.484	2.147	1.455	2.107
Kurtosis	5.483	7.704	5.866	8.880	5.690	8.449
Jarque–Bera	60.538	158.239	291.542	907.788	134.178	405.275

4.2 Empirical Analysis

We assess the relationship between implied volatility index and realized volatility based on a linear regression of the form:

$$RV_m = \alpha_0 + \alpha_1 IV_{m-1} + \varepsilon_m \tag{15}$$

Christensen and Prabhala [14] suggested three hypotheses that can be tested regarding Eq. 15. First, if IV contains at least some information about future realized volatility, coefficient α_1 should be statistically significant against a null hypothesis of $\alpha_1 = 0$. Second, if IV is an unbiased estimate of realized volatility, then the intercept of Eq. 15 should be zero and coefficient should equal one. This joint hypothesis can be tested using F statistic [36]. Lastly, if IV is indeed an efficient estimate, the residuals should be pure white noise and uncorrelated with any other variable.

In the first step, we use ordinary least square to estimate the coefficients in Eq. 15 and test three hypotheses proposed by Christensen and Prabhala [14]. The results are summarized in Table 5.

From results in Table 5, we find α_1 is statistically different from zero, which indicates that OVX contains information regarding future realized volatility of crude oil spot prices. Testing the null hypothesis of α_1=1, 5- and 10-day period data reject this hypothesis. The joint hypothesis of $\alpha_0 = 0$ and $\alpha_1 = 1$ are accepted for all the three data set, which suggests OVX is an unbiased estimation of future realized volatility. However, the Durbin–Watson (DW) statistic values indicates that OVX isn't an efficient predictor of future realized volatility, since for all the three sets of data are significantly different from two (indicating residuals are autocorrelated).

In the second step, we transform Eq. 15 with Kalman filter and analyze the parameters in a dynamic system:

$$\begin{bmatrix} \alpha_{0,k+1} \\ \alpha_{1,k+1} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha_{0,k} \\ \alpha_{1,k} \end{bmatrix} + w_k \tag{16}$$

Table 5 Descriptive statistics of the OLS estimation

	α_0	α_1	MSE	F stats	DW stats	R square
5-day	−0.013	1.348	0.415	536.095	1.760	0.567
	0.000	0.000		0.000	0.013	
10-day	−0.017	1.379	0.559	401.565	1.786	0.664
	0.000	0.000		0.000	0.108	
Monthly	−0.063	1.096	11.6	217.526	1.563	0.694
	0.034	0.000		0.000	0.016	

1.MSEs are multiplied by 1000
2. For 5-day data, H0: $\alpha_1 = 1$, p value is 0.000. The joint hypothesis H0: $\alpha_0 = 0$ and $\alpha_1 = 1$, p value is 0.000 and F statistic is 20.419.
3. For 10-day data, H0: $\alpha_1 = 1$, p value is 0.000. The joint hypothesis H0: $\alpha_0 = 0$ and $\alpha_1 = 1$, p value is 0.000 and F statistic is 20.588.
4. For monthly data, H0: $\alpha_1 = 1$, p value is 0.199. The joint hypothesis H0: $\alpha_0 = 0$ and $\alpha_1 = 1$, p value is 0.039 and F statistic is 4.102

Table 6 Descriptive statistics of the Kalman Filter estimation

		α_0	α_1	MSE
5-day	Mean	−0.015	1.380	0.359
	Std. Dev.	0.007	0.123	
10-day	Mean	−0.024	1.454	0.460
	Std. Dev.	0.009	0.098	
Monthly	Mean	−0.047	1.012	6.534
	Std. Dev.	0.057	0.161	

MSEs are multiplied by 1000

$$RV_{m,k} = [\,1 \;\; IV_{m-1,k}\,] \begin{bmatrix} \alpha_{0,k} \\ \alpha_{1,k} \end{bmatrix} + v_k \tag{17}$$

where w_k and v_k have the same meaning as Eqs. 1 and 2. We use the first quarter of the observations to initialize the covariance matrix of process noise Q, measurement noise R. These volatility series are synchronized, so that realized volatility in month m is aligned with implied volatility observed on the last trading day of month $m-1$.

The results of Kalman filter are presented and analyzed in Table 6.

Results in Table 6 indicate that Kalman filter achieves a smaller mean squared error. Moreover, the mean values of α_0 are very close to zero and the mean values of α_1 are around one. The results are very similar to the OLS estimates.

4.3 Quantification of Market Risk

In this paper, we quantify market risk based on the information content of OVX to future realized volatility. More specifically, we assess the added value of OVX based volatility forecasts when these forecasts are used to quantify and estimate short-term market risk level. We consider the widely used Value-at-Risk (VaR) framework which provides, at a given percentage level, the most likely loss for an investor. In this framework, we use the estimated RV calculated from OVX to substitute the predicted conditional standard deviation in the calculation of VaR as in Eq. 18.

$$VaR_{t+1|t} = F_{\alpha}(z_t;\theta)\sigma_{t+1|t} \tag{18}$$

where the distribution is assumed to be normal and $F_{\alpha}(z_t;\theta)$ is the relevant quantile at the $100 \cdot \alpha\%$ level from the normal distribution. $\sigma_{t+1|t}$ is the forecast of the conditional standard deviation at time $t+1$.

Furthermore we introduce RiskMetrics model, which can easily be used by market practitioners, as a competing model to evaluate the performance of the proposed model. The RiskMetrics model is defined in the following equation.

$$\sigma_{t|t-1} = \sqrt{\alpha_0 r_{t-1}^2 + \alpha_1 \sigma_{t|t-1}^2} \tag{19}$$

In which α_0 and α_1 are estimated based on the historical data.

In order to back-test the VaR results, we use Kupiec LR test in the paper [37]. Given the ex-post observed returns $\{r_{t+1}\}$ and ex-ante forecast $\{VaR_t\}$, the empirical failure rate $\hat{f}$ is expected to be equal to the number of returns smaller than the VaR. If the number of violations differs considerably from $\alpha \times 100\,\%$ of the sample, then the accuracy of the underlying risk model is called into question. The null hypothesis $H_0 : f = \alpha$ against $H_1 : f \neq \alpha$ can be tested with the LR statistics, which takes the form as follows.

$$LR = -2\ln[(1-\alpha)^{T-N}\alpha^{N}] + 2\ln[(1-N/T)^{T-N}(N/T)^{N}] \tag{20}$$

where N is the number of violations in the sample, T is the total number of observations. Under the null hypothesis, the test statistic is χ^2 distributed with one degree of freedom.

We apply both methods to estimate VaR of crude oil returns. The VaR for 5- , 10-day and 1-month are tested using the same period as in the last section. These VaR forecasts $\{VaR_t\}$, pertaining to returns defined on $[t, t+1]$, can then be back-tested against the observed returns $\{r_{t+1}\}$. For back-testing of VaR forecasts, we compute the empirical failure rates and Kupiec LR tests for both left and right quantile at 1 %, 2.5 % and 5 % since investors can hold short positions of crude oil. Empirical results for the two models based on OVX (estimated with Kalman filter and OLS) and RiskMetrics based on historical volatility, using time series of different frequencies, are shown in Table 7. The confidence level of LR test is 0.05. We find that in general Kalman filter based VaR model outperforms the OLS based VaR model and RiskMetrics model. The results of Kalman filter with $\alpha = 95\,\%$ are illustrated in Figs. 2, 3 and 4.

5 Conclusion

Implied volatility (IV) is a new measure of markets' expected risk derived from the price of a market traded option and it has attracted much attention in recent years

Table 7 Back-testing results with different and different time frequencies

Models	α (%)	Left side				Right side			
		VaR exceed	Kupiec stats	p Value	Model accept	VaR exceed	Kupiec stats	p Value	Model accept
5-Day									
OVX	1.0	5	0.182	0.330	Yes	1	3.417	0.935	Yes
(Kalman	2.5	12	0.282	0.405	Yes	7	1.203	0.727	Yes
filter)	5.0	21	0.010	0.081	Yes	23	0.297	0.414	Yes
OVX	1.0	12	10.090	0.999	No	7	1.695	0.807	Yes
(OLS)	2.5	20	7.429	0.994	No	15	1.956	0.838	Yes
	5.0	31	4.872	0.973	No	30	4.031	0.955	No
Risk	1.0	4	0.002	0.040	Yes	0	8.241	0.996	No
Metrics	2.5	7	1.187	0.724	Yes	2	10.133	0.999	No
	5.0	11	5.534	0.981	No	6	14.789	1.000	No
10-day									
OVX	1.0	3	0.389	0.467	Yes	2	0.001	0.028	Yes
(Kalman	2.5	8	1.417	0.766	Yes	3	1.059	0.697	Yes
filter)	5.0	13	0.718	0.603	Yes	11	0.056	0.188	Yes
OVX	1.0	8	10.062	0.998	No	1	0.670	0.587	Yes
(OLS)	2.5	10	3.739	0.947	Yes	4	0.274	0.399	Yes
	5.0	12	0.299	0.415	Yes	7	1.215	0.730	Yes

Table 7 continued

Models	α (%)	Left side				Right side			
		VaR exceed	Kupiec stats	p Value	Model accept	VaR exceed	Kupiec stats	p Value	Model accept
Risk	1.0	1	0.659	0.583	Yes	0	4.101	0.957	No
Metrics	2.5	2	2.504	0.886	Yes	1	5.026	0.975	No
	5.0	5	3.409	0.935	Yes	2	10.225	0.999	No
Monthly									
OVX	1.0	2	0.824	0.636	Yes	1	0.000	0.016	Yes
(Kalman	2.5	2	0.090	0.236	Yes	2	0.090	0.236	Yes
filter)	5.0	8	1.748	0.814	Yes	4	0.185	0.333	Yes
OVX	1.0	4	5.307	0.979	No	0	1.970	0.840	Yes
(OLS)	2.5	8	8.163	0.996	No	0	4.962	0.974	No
	5.0	9	2.927	0.913	Yes	3	0.895	0.656	Yes
Risk	1.0	4	5.371	0.980	No	2	0.846	0.642	Yes
Metrics	2.5	6	3.858	0.950	No	4	0.880	0.652	Yes
	5.0	7	0.888	0.654	Yes	7	0.888	0.654	Yes

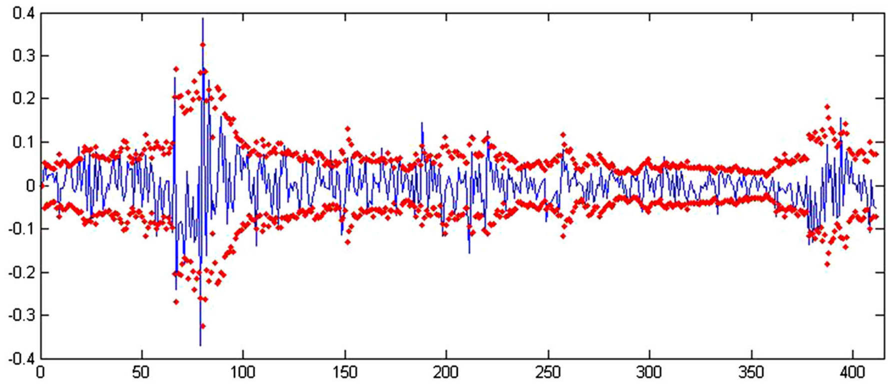

Fig. 2 5-day VaR estimates of the OVX based model with Kalman filter ($\alpha = 95\,\%$)

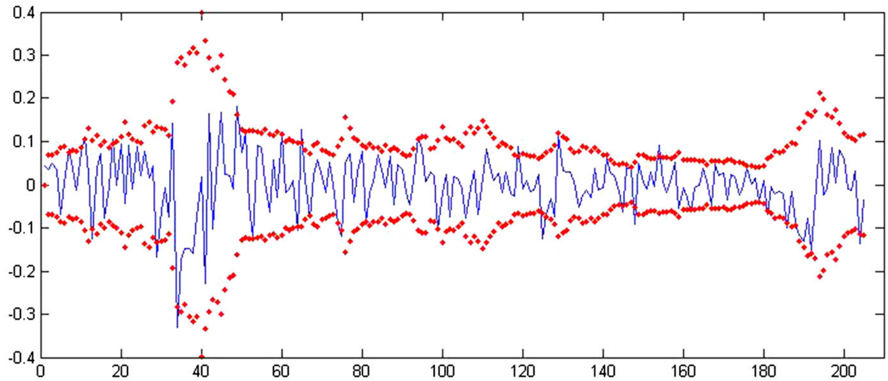

Fig. 3 10-day VaR estimates of the OVX based model with Kalman filter ($\alpha = 95\,\%$)

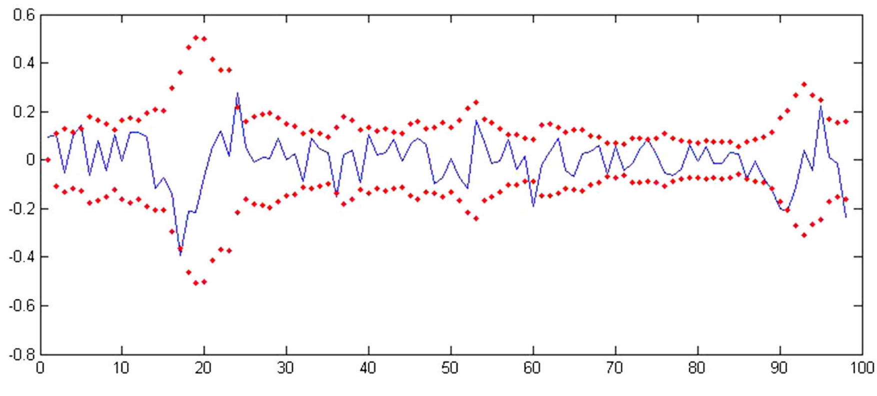

Fig. 4 Monthly VaR estimates of the OVX based model with Kalman filter ($\alpha = 95\,\%$)

because of its importance to financial markets. This research examines the relationship between OVX and crude oil spot returns. The results indicate that, firstly, the time-varying coefficients confirm the negative but weak relationship between OVX changes and crude oil spot returns, and the relationship diminishes as time goes on. Secondly the asymmetric effect is not significant and consistent with different time horizons. Finally, our results suggest neither extremely high or low level of OVX can predict long-term forward looking returns.

This paper also examines the relationship between OVX and realized volatility of crude oil spot returns with linear regression model and investigates the forecasting capability of OVX with Kalman filter. The t-test indicates OVX is an unbiased estimate of future realized volatility, but the DW test indicates OVX is not an efficient estimate of future realized volatility. But OVX contains information of future realized volatility since α_1 statistically is not equal to zero. A forecasting method involving Kalman filter provides better performance than linear regression model in forecasting future realized volatility based on OVX. Finally, application for quantifying market risk is investigated. Empirical results show it's feasible to use OVX to estimate VaR and Kalman filter method dominates other methods in these cases.

Acknowledgments This work is supported by the National Science Fund for Distinguished Young Scholars (NSFC No. 71025005), the National Natural Science Foundation of China (NSFC Nos. 71201054, 91224001), the Strategic Research Grant of City University of Hong Kong (No. 7004574), the Fundamental Research Funds for the Central Universities in BUCT, and the University Young Teacher Training Program of Shanghai. An earlier version of this paper with authors: Yanhui Chen and Yingchao Zou was awarded as the best paper and presented at the International Conference on Information Technology and Quantitative Management (ITQM 2015), Rio, Brazil, July 21-24, 2015 (see Procedia Computer Science, Volume 55, 2015, Pages 1359?1365, 3rd International Conference on Information Technology and Quantitative Management, ITQM 2015).

References

1. Fleming J, Ostdiek B, Whaley RE (1995) Predicting stock market volatility: a new measure. J Futur Mark 15(3):265–302
2. Whaley RE (2000) The investor fear gauge. J Portf Manag 26(3):12–17
3. Whaley RE (2009) Understanding the VIX. J Portf Manag 3:98–105
4. Simon DP (2003) The Nasdaq volatility index during and after the bubble. J Deriv 11(2):9–24
5. Giot P (2005) Relationships between implied volatility indexes and stock index returns. J Portf Manag 31(3):92–100
6. Carr P, Wu L (2006) A tale of two indices. J Deriv 13(3):13
7. Dowling S, Muthuswamy J (2005) The implied volatility of Australian index options. Rev Futur Mark 14(1):117–155
8. Siriopoulos C, Fassas A (2008) The Information Content of VFTSE
9. Chen Y, Lai KK (2013) Examination on the relationship between VHSI, HSI and future realized volatility with Kalman Filter. Eurasian Bus Rev 3(2):200–216
10. Banerjee P, Doran J, Peterson D (2007) Implied volatility and future portfolio returns. J Bank Financ 31(10):3183–3199
11. Poon SH, Granger CWJ (2003) Forecasting volatility in financial markets: a review. J Econ Lit 41(2):478–539
12. Giot P (2005) Implied volatility indexes and daily value at risk models. J Deriv 12(4):54–64
13. Jiang GJ, Tian YS (2005) The model-free implied volatility and its information content. Rev Financ Stud 18(4):1305–1342

14. Christensen BJ, Prabhala NR (1998) The relation between implied and realized volatility. J Financ Econ 50(2):125–150
15. Corrado CJ, Miller JTW (2005) The forecast quality of CBOE implied volatility indexes. J Futur Mark 25(4):339–373
16. Canina L, Figlewski S (1993) The informational content of implied volatility. Rev Financ Stud 6(3):659–681
17. Becker R, Clements AE (2008) Are combination forecasts of S&P 500 volatility statistically superior? Int J Forecast 24(1):122–133
18. Siriopoulos C, Fassas A (2009) Implied volatility indices-a review
19. Bhar R (1996) Modelling Australian bank bill rates: a Kalman filter approach. Account Financ 36(1):1–14
20. Manoliu M, Tompaidis S (2002) Energy futures prices: term structure models with Kalman filter estimation. Appl Math Financ 9(1):21–43
21. Bedendo M, Hodges SD (2009) The dynamics of the volatility skew: a Kalman filter approach. J Bank Financ 33(6):1156–1165
22. Kolodziej M et al (2014) Crude oil: commodity or financial asset? Energy Econ 46:216–223
23. Chen Y, Zou Y (2015) Examination on the relationship between OVX and crude oil price with Kalman filter. Proc Comput Sci 55:1359–1365
24. Kalman RE (1960) A new approach to linear filtering and prediction problems. J Basic Eng 82(1):35–45
25. Welch G, Bishop G (1995) An introduction to the Kalman filter. University of North Carolina at Chapel Hill, Chapel Hill, NC
26. Tsay RS (2005) Analysis of financial time series, vol 543. Johm Willey & Sons Inc., Hoboken
27. Brown RG, Hwang PYC (1992) Introduction to random signals and applied Kalman filtering. John Wiley & Sons Inc., New York
28. Jacobs OLR (1993) Introduction to control theory. Oxford University Press, Oxford
29. Harvey AC (1991) Forecasting, structural time series models and the Kalman filter. Cambridge University Press, Cambridge
30. Günter JHF (2014) How do oil producers respond to oil demand shocks? Energy Econ 44:1–13
31. Deeney P et al (2015) Sentiment in oil markets. Int Rev Financ Anal 39:179–185
32. Krichene N (2002) World crude oil and natural gas: a demand and supply model. Energy Econ 24(6):557–576
33. CBOE (2003) VIX White Paper. URL: http://www.cboe.com/micro/vix/vixwhite
34. Strunk Hansen C (2001) The relation between implied and realised volatility in the Danish option and equity markets. Account Finance 41(3):197–228
35. Fleming J (1998) The quality of market volatility forecasts implied by S&P 100 index option prices. J Empir Finance 5(4):317–345
36. Neter J, Wasserman W, Kutner MH (1983) Applied linear regression models. Richard D, Irwin Inc, Homewood Illinois
37. Kupiec P. (1995) Techniques for verifying the accuracy of risk measurement models. J Deriv 3(2): 73–84