The predictive content of CBOE crude oil volatility index

Hongtao Chen[a], Li Liu [b]*, Xiaolei Li[c]

[a] School of Economics and Management, Southeast University, China

[b] School of Finance, Nanjing Audit University, China

[c] University of New South Wales, Sydney, Australia

*Corresponding author

Email: liuli840821@nau.edu.cn

Address: 86 West Yushan Road, Pukou District, Nanjing 211815, China

Abstract

Volatility forecasting is an important issue in the area of econophysics. The information content of implied volatility for financial return volatility has been well documented in the literature but very few studies focus on oil volatility. In this paper, we show that the CBOE crude oil volatility index (OVX) has predictive ability for spot volatility of WTI and Brent oil returns, from both in-sample and out-of-sample perspectives. Including OVX-based implied volatility in GARCH-type volatility models can improve forecasting accuracy most of time. The predictability from OVX to spot volatility is also found for longer forecasting horizons of 5 days and 20 days. The simple GARCH(1,1) and fractionally integrated GARCH with OVX performs significantly better than the other OVX models and all 6 univariate GARCH-type models without OVX. Robustness test results suggest that OVX provides different information from as short-term interest rate.

Keywords: CBOE oil volatility index; GARCH-class models; Out-of-sample forecasting; MCS test

1. Introduction

Crude oil is of great importance for the real economy. The impacts of oil price changes on the real economy and financial markets have been extensively investigated in the literature [1-6]. Volatility is a crucial input in option pricing, portfolio optimization and value at risk computing. Therefore, accurate forecasts of oil volatility are required by oil market participants, financial investors and policy makers. Under this background, it is not surprising that there exists quite a large number of studies on modeling and forecasting oil return volatility [7-17].

In the existing studies, the autoregressive conditional heteroscedasticity (ARCH) model initially proposed by Engle [18] and its various extensions are widely used to capture the evolution of oil volatility[1]. These models draw predictive information from historical volatility or return. In this paper, we contribute to the literature by using the option implied volatility represented by CBOE crude oil ETF volatility index (OVX). We find that OVX can predict oil spot volatility from both in-sample and out-of-sample perspectives. Our finding is helpful for understanding oil market volatility.

The OVX, obtained by applying "VIX" methodology to oil option, is available since May 2007. This index reflects market expectation of 30-day volatility of crude oil returns. Many studies investigate the predictive content of implied volatility of financial asset such as stock index [22-26], but few attentions are paid to implied

[1] Important extensions of ARCH include generalized ARCH (GARCH) [19], asymmetric GARCH [20] (GJR) and fractionally integrated GARCH (FIGARCH) [21].

volatility of crude oil. In particular, Agnolucci [27] takes implied volatility as the forecasts of WTI oil volatility and finds that they are less accurate than the forecasts obtained from GARCH-type models. Agnolucci [27]'s result indicates that implied volatility is a poor proxy of future oil volatility but it does not necessarily suggest that implied volatility is uninformative. Szakmary et al. [28] use data from 35 futures options markets including crude oil market and find that implied volatility outperforms historical volatility as a predictor of subsequent realized volatility although it is not a completely unbiased predictor. More recently, Haugom et al. [29] find that including the OVX-based implied volatility in realized volatility models can significantly improve daily and weekly volatility forecasts of crude oil futures returns. We extend existing papers because of three motivations.

First, the options underlying the OVX are options on the United States Oil fund (USO), an exchange traded fund (ETF) established to replicate the returns of the WTI price. USO is unable to completely replicate WTI spot price. The reason is that such fund is impossible to invest in physical oil, but uses futures contracts instead. Moreover, spot price is more fluctuated than futures price because the spot price is exposed to fundamental information completely whereas futures price is partly affected by people's expectation. For example, the standard deviations of returns of WTI oil spot and nearby futures contract during our sample period from May 10, 2007 through November 22, 2016 are 2.532 and 2.514, respectively. The more volatile spot price determines the fact that spot volatility is more difficult to be predicted. Therefore, although OVX is found to predict *futures* volatility successfully in

Haugom et al. [29], whether it can forecast *spot* volatility has not been investigated in the literature.

Second, the underlying asset of options underlying OVX is WTI oil, rather than Brent oil. WTI oil price reflects more information about U.S. domestic oil supply and demand, while Brent price is more global and is the benchmark price of oil traded in the Europe and Africa. Baumeister and Kilian [30] argue that the European central banks are more interested in Brent oil price, not WTI oil price. Therefore, accurate volatility forecasts of Brent oil are more required by policy makers in the European countries. Can OVX predict Brent oil volatility? This issue has not been addressed in the literature.

The third motivation comes from econometric modeling. Haugom et al. [29]'s econometric framework is built on the predictive regressions of realized volatility (RV), which is defined as the summation of the squared intradaily high-frequency returns. The use of RV treats daily volatility as an observed variable. However, the RV models are not suitable for modeling spot volatility because oil spot price is available only at daily frequency. In this case, the realized volatility is unavailable and the true volatility is a latent variable that is unknown for forecasters. We have to use the GARCH-type models to capture daily oil volatility. Does the predictability from OVX to oil volatility still hold when the true volatility is unobservable?

In this paper, we address aforementioned issues by forecasting WTI and Brent oil spot volatility using OVX. We add OVX-based implied volatility to each of 6 GARCH-type models. Our in-sample evidence suggests that the lagged OVX has significantly positive impacts on current volatility of both WTI and Brent oil returns, regardless of which GARCH-type specification is considered.

To investigate the predictive content of OVX out-of-sample, we compare the forecasting performances of a GARCH-type model with OVX and the same univariate model without OVX. We consider 6 GARCH-type specifications allowing for different stylized facts in oil volatility. Six different loss functions are employed to evaluate forecasting performance. We use the Diebold and Mariano [31] test to examine whether the difference between loss functions of an OVX model and the corresponding benchmark of univariate GARCH is significant. We find that including implied volatility in a GARCH model can significantly improve forecasting accuracy of oil spot volatility under most loss criteria.

To further understand the forecasting performance, we use a more advanced test, named model confidence set (MCS) [32], to compare the forecasting accuracy of six OVX models and six GARCH-type models without OVX one time. Our results show that the unvariate GARCH-type benchmarks are always excluded from the MCS, further confirming the superior performances of OVX models. Moreover, GARCH and FIGARCH models accounting for OVX information perform better than almost all the competitors, while the forecasting performances of themselves are equally well in the MCS sense. Our analysis is also extended to longer-horizon forecasting. The forecasting results consistently support the usefulness of OVX information for forecasting oil spot volatility.

The loss functions may be argued to be sensitive to extreme forecasts. Motivated by this fact, we employ an alternative evaluation criterion, success ratio (SR), which quantifies how often a GARCH-OVX model results in more accurate volatility forecasts than its corresponding benchmark model without OVX. Therefore, an SR higher than 0.5, the probability of tossing a coin, implies that adding OVX to a volatility model is more likely to obtain more accurate forecasts. We find that the success ratios of all OVX-based volatility models are significantly higher than 0.5 for both WTI and Brent volatility. Notably, when the FIGARCH specification is considered, the success ratio of OVX is as high as 0.6. Therefore, OVX can successfully predict oil volatility most of time. Further analysis shows that the event that the models with OVX display superior forecasting performance over the univariate benchmarks occurs persistently.

It is possible whether forecast improvement can be achieved by adding OVX to a volatility model is sensitive to omitted variables in this model. For robustness, we incorporate a macro variable, short-term interest, in both benchmark and alternative models, and perform forecasting exercise again. We find that the macro variable does not change the core finding that OVX can predict oil volatility, suggesting the robustness of our results. OVX and short-term interest provide different information regarding future oil volatility.

The remainder of this paper is organized as follows: Section 2 shows data and descriptive statistics. Section 3 describes the econometric methodology briefly. Section 4 gives in-sample and out-of-sample forecasting results. The last section concludes the paper.

2. Data and descriptive statistics

We collect daily closing price data of West Texas Intermediate (WTI) and Brent crude oils from Energy Information Administration (EIA) (www.eia.gov). Our data cover the period from May 10, 2007 through November 22, 2016. This starting date is consistent with the time when OVX was initially published. Based on the available price data, the returns can be computed using the following equation:

$$r_t = 100 \times (\log P_t - \log P_{t-1}), \tag{1}$$

where P_t is the oil price on day *t*. The first order differences of logarithmic oil prices are multiplied by 100 to denote the percent values. We show the graphical representations of WTI and Brent returns in Figure 1.

Figure 1 about here

CBOE crude oil ETF volatility index (OVX) data are obtained from the Federal Reserve Bank at Saint Louis (https://fred.stlouisfed.org/). This index measures the market's expectation of 30-day volatility of crude oil prices by applying the VIX methodology[2] to the United States Oil Fund[3], LP (Ticker-USO) options spanning a wide range of strike prices. Because OVX denotes the annualized standard deviation, we use the squared OVX divided by 252 to construct daily implied volatility. Figure 2 illustrates the evolution of implied volatility, as well as actual oil return volatility, proxied by the squared daily returns. It is evident that squared daily returns are noisier than implied volatility. This should be expected because OVX reflect expectation of

[2] For the details about VIX methodology, one can refer to the website of CBOE (http://www.cboe.com/micro/vix/part2.aspx).

[3] The United States Oil Fund is an exchange-traded security designed to track changes in crude oil prices. By holding near-term futures contracts and cash, the performance of the Fund is intended to reflect, as closely as possible, the spot price of West Texas Intermediate light, sweet crude oil, less USO expenses.

30-day volatility, while squared daily return is the instantaneous volatility.

Figure 2 about here

Table 1 shows the descriptive statistics of oil returns and volatilities. First, looking at the returns, the mean values of returns of WTI and Brent oil prices are rather close to zero. The Jarque-Bera statistics strongly reject the null hypothesis of normal distribution, indicating the fat-tailed distribution. The non-Gaussian distribution is also reinforced by the positive skewness and kurtosis higher than 3. The Ljung-Box Q statistics reject the null hypothesis of no serial correlation up to the 10^{th} lag order at 1% significance level, implying the existence of strong autocorrelation. This evidence motivates us to use AR(1) model to capture return dynamics. The augmented Dickey-Fuller statistics show rejections of null hypothesis that each return series contains a unit root at 1% significance level. The stationary time series enables us to apply GARCH models to returns directly with no need of any further transformation.

Then, turning to the three volatility series, the mean values of WTI spot volatility and implied volatility are rather close but the standard deviation of the latter series is much larger. This evidence is consistent with our intuition about actual and implied volatility. The descriptive statistics also suggests some stylized facts of oil volatility such as fat-tailed distribution, persistence and stationarity. Notably, the strong persistence makes the great appreciation of GARCH-class in modeling and forecasting volatility.

Table 1 about here

3. Econometric models

This section briefly shows the econometric methodology on how to capture the effects of OVX on future oil volatility. Prior to model the evolution of conditional volatility, we use a simple AR(1) specification to describe the changes in conditional mean to account for strong autocorrelation in oil returns. The AR(1) model is given by,

$$r_t = \mu + \lambda r_{t-1} + \varepsilon_t = \mu + \lambda r_{t-1} + \sigma_t z_t , \qquad (2)$$

where r_t is the return of crude oil spot, z_t is the standardized residuals which are assumed to be independent and normally distributed.

Based on the return innovations generated by above AR(1) model, a standard benchmark to forecast daily oil volatility is a GARCH-type model. This type of models assumes that the current expected volatility is a function of past squared residuals and past expected volatility. In general, the specification of GARCH-type models is given by,

$$\sigma_t^2 = f(\sigma_{t-1}^2, \varepsilon_{t-1}^2, \varepsilon_{t-1}) , \qquad (3)$$

To investigate the predictive content of OVX, we extend the benchmark by incorporating a lagged OVX-based implied volatility as an additional predictor:

$$\sigma_t^2 = f(\sigma_{t-1}^2, \varepsilon_{t-1}^2, \varepsilon_{t-1}) + \rho IV_{t-1} , \qquad (4)$$

where IV_{t-1} is the daily implied volatility defined by squared OVX divided by 252. The parameter ρ captures the effects of OVX on future oil volatility[4]. We obtain the parameter estimates in (4) using the maximum likelihood estimation (MLE). The null

[4] Please note that in the EGARCH specification, we use the logarithmic form of implied volatility.

hypothesis of no predictability, ρ =0, can be tested using the *t*-statistic.

The estimate of ρ depends on the specification of GARCH-type models. For robustness, we consider 6 popular GARCH-type models. The first is the simple GARCH(1,1) [19] which is given by[5],

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2, \tag{5}$$

where σ_t^2 is the conditional volatility with the sufficient conditions $\omega > 0$, $\alpha \geq 0$ and $\beta \geq 0$ to gauge σ_t^2 >0. The sufficient conditions for the existence of second moment for GARCH is $\alpha + \beta < 1$.

An extension of the GARCH is Glosten et al. [20] GJR-GARCH model that can capture the asymmetric relationships between return and volatility (i.e., leverage effect). The specification of GJR is given by,

$$\sigma_t^2 = \omega + [\alpha + \gamma\ I(\varepsilon_{t-1} < 0)]\,\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2, \tag{6}$$

where *I*(.) is an indicator function; i.e., when the condition in the parenthesis is satisfied, it takes the value of 1 and 0 otherwise. The parameter γ describes the volatility leverage effect. The sufficient condition of GJR is $\omega > 0$, $\alpha \geq 0$, $\beta \geq 0$ and $\alpha + \gamma \geq 0$ to ensure conditional volatility σ_t^2 >0. The sufficient condition for the existence of second moment for GJR is $\alpha + \beta + \gamma / 2 > 0$.

Another GARCH-type model allowing for leverage effect is the exponential GARCH (EGARCH) developed by Nelson [33]. Nelson [33] points out that the non-negative parameter restriction of GARCH is too strict. EGARCH can address this

---

[5] It may be argued that more lags should be used in GARCH. We use GARCH(1,1) due to two considerations. First, Hansen and Lunde [34] show that a GARCH model with longer lag lengths cannot significantly beat the simple GARCH(1,1) in forecasting return volatility. Second, the use of more lags will bring much higher calculation burden and takes much more time in computing out-of-sample forecasts.

problem by modeling the dynamics of logarithmic conditional volatility. The volatility equation of GARCH can be written as follows:

$$\log(\sigma_t^2) = \omega + \alpha(|z_{t-1}| - E|z_{t-1}|) + \gamma z_{t-1} + \beta \log(\sigma_{t-1}^2) , \tag{7}$$

where again the parameter γ describes the volatility leverage effect.

Aforementioned GARCH processes assume the short memory of autocorrelations. Therefore, they cannot capture the stylized fact of long memory in financial volatility [35]. Recent studies also point out that oil return volatility displays long memory property [36-37]. Due to this motivation, we use a fractionally integrated ARCH (FIGARCH) of Baillie et al. [21]. This model implies the hyperbolic decay of volatility autocorrelation, i.e., long memory behavior and the slow rate of decay after a volatility shock. The FIGARCH(1,d,1) process is given by,

$$\sigma_t^2 = \omega + \beta\sigma_{t-1}^2 + [1 - (1-\beta L)^{-1} (1-\varphi L)(1-L)^d]\, \varepsilon_t^2 , \tag{8}$$

where $0 \le d \le 1$, $\omega > 0$, $\varphi, \beta < 1$; d is the fractional integration parameter and L is the lag operator. For 0<d<1, FIGARCH is sufficiently flexible to consider the intermediate range of the degree of persistence, between the infinite persistence associated with d=1 and the geometric decay associated with d=0.

Another GARCH-class model allowing for long memory used in this paper is the hyperbolic GARCH (HYGARCH) developed by Davidson [38]. The application of HYGARCH to oil market data can be seen in Wei et al. [17]. The HYGARCH specification is given by,

$$\sigma_t^2 = \omega + \{1 - [1-\beta L]^{-1} \varphi L \{1 + k[(1-L)^d - 1]\}\}\, \varepsilon_t^2 , \tag{9}$$

where $0 \le d \le 1$, $\omega > 0$, $k \ge 0$, $\varphi, \beta < 1$. L is the lag operator. The HYGARCH process corresponds to FIGARCH for k=1 and stable GARCH for k=0.

We also consider a popular model accommodating both long memory and asymmetric effect. The FIEGARCH(1,d,1) is given by,

$$h_t = (1-L^d)(\ln\sigma_t^2 - \omega) = \beta h_{t-1} + g(z_{t-1}) + \varphi g(z_{t-2}), \tag{10}$$

where $g(z_t) = \theta z_t + \gamma(|z_t| - E|z_t|)$.

In summary, we use a total of 6 GARCH-type models, the standard GARCH(1,1), two asymmetric models (GJR and EGARCH), two long memory models (FIGARCH and HYGARCH) and a model allowing for both long memory and asymmetry (FIEGARCH).

4. Empirical results

This section constitutes in-sample and out-of-sample results about the predictive content of OVX for oil spot volatility. We pay more attention to the out-of-sample evidence by finding whether the incorporation of OVX in the GARCH volatility equation can improve the performance in forecasting oil volatility.

4.1. In-sample results

According to the argument of Inoue and Kilian [39], in-sample predictability is a necessary condition for out-of-sample predictability. It would be unreasonable to find out-of-sample predictability in the absence of in-sample predictability. Table 2 shows the estimated coefficients of OVX in 6 GARCH-type models. We do not report the other parameter estimates in each model because it is enough for us to analyze the in-sample predictive ability of OVX. Although the magnitude of OVX coefficient depends on the specification of GARCH-type models, we can find the consistent evidence that the coefficient of OVX in each model is significantly positive, indicating the strong predictability from OVX to oil spot volatility. We also give the change in log likelihood after adding OVX to each volatility equation. The result is clear that OVX can improve the in-sample performance of GARCH-type models.

Table 2 about here

4.2. Out-of-sample results

We use the technique of rolling window to generate out-of-sample volatility forecasts of WTI and Brent oil returns. In detail, the whole sample data with a total of *T* observations is divided into two subsamples: the first subsample for parameter estimation covering the initial *M* observations and the second subsample for forecast evaluation covering the remaining *T-M* observations. We use the first subsample to estimate parameters and obtain the out-of-sample forecast for the volatility on the (*M*+1)-th day. Then, the estimation window is moved forward by adding a new observation and dropping the earliest observation. That is, we use the observations from the 2^{nd} to (*M*+1)-th day to do parameter estimation again and generate volatility forecast on the (*M*+2)-th day accordingly. Going forward like this, the length of estimation window is fixed to *M* and one can finally obtain *T-M* volatility forecasts.

We set the length of rolling window to 1000. Figure 3 illustrates the graphical representations of volatility forecasts generated by GARCH(1,1) with OVX (GARCH-X) and the forecasts generated by GARCH(1,1) without OVX, as well as the true volatility[6]. During the 2011-2014 period when oil prices are less fluctuated,

[6] We do not plot the volatility forecasts from the other 5 GARCH-type models because their forecasts are close, but they are available upon request.

the GARCH-X leads to higher volatility forecasts than the benchmark GARCH, whereas the GARCH-X results in lower volatility forecasts during recent period when oil prices evolve more fiercely.

Figure 3 about here

4.2.1. Diebold and Mariano [31] test results

In order to evaluate the forecasting accuracy, we use following six criteria of loss functions:

$$MSE_1 = \frac{1}{T-M}\sum_{t=M+1}^{T}\left(\sigma_t^2 - \hat{\sigma}_t^2\right)^2, \tag{11}$$

$$MAE_1 = \frac{1}{T-M}\sum_{t=M+1}^{T}\left|\sigma_t^2 - \hat{\sigma}_t^2\right|, \tag{12}$$

$$MSE_2 = \frac{1}{T-M}\sum_{t=M+1}^{T}\left(\sigma_t - \hat{\sigma}_t\right)^2, \tag{13}$$

$$MAE_2 = \frac{1}{T-M}\sum_{t=M+1}^{T}\left|\sigma_t - \hat{\sigma}_t\right|, \tag{14}$$

$$R^2LOG = \frac{1}{T-M}\sum_{t=M+1}^{T}\left(\log(\sigma_t^2 / \hat{\sigma}_t^2)\right)^2, \text{ and,} \tag{15}$$

$$QLIKE = \frac{1}{T-M}\sum_{t=M+1}^{T}\left(\log(\hat{\sigma}_t^2) + \sigma_t^2 / \hat{\sigma}_t^2\right), \tag{16}$$

where σ_t^2 and $\hat{\sigma}_t^2$ are the true value and forecast of volatility on the *t*-th day, respectively. MSE and MAE are the mean squared error and mean absolute error, respectively. R^2LOG is similar to the R^2 of Mincer-Zarnowitz regression. Notably, Patton [40] demonstrates that MSE and QLIKE are robust to the imperfect volatility proxies.

To find whether adding OVX to the benchmark volatility model can improve forecasting performance, we calculate the ratios of loss functions of GARCH-X relative to the loss function of corresponding univariate GARCH-type model. A ratio smaller than 1 implies that the related GARCH-X model displays lower forecasting loss, therefore producing more accurate volatility forecasts under a pre-specified criterion. We use Diebold and Mariano [31] method to test whether the differences of loss functions of two models are significant.

Panel A of Table 3 shows the performances of OVX in forecasting WTI volatility. We can find that the loss ratios are all smaller than 1 when the benchmark model is the standard GARCH(1,1). The DM statistics reject the null hypothesis of the equal loss functions of GARCH(1,1) with OVX and the original benchmark model at 10% or lower significance levels. Therefore, the incorporation of OVX in the standard GARCH(1,1) can significantly improve the forecasting accuracy. The finding holds for two fractionally FIGARCH and FIEGARCH that the predictability of OVX is significant under almost all loss criteria. When using the benchmark of each of two asymmetric GARCH models (GJR and EGARCH), we also find the significant improvement of predictive ability under 4 out of 6 loss criteria. Under the other two criteria of loss functions (MSE_1 and MSE_2), EGARCH or GJR with OVX performs as well as the univariate benchmark model in the statistical sense. Thus, EGARCH or GJR specifications accounting for the information from OVX are better candidates than corresponding benchmark models in forecasting WTI volatility. The result is generally consistent when using HYGARCH. In summary, we can obtain more accurate volatility forecasts by taking advantage of information from OVX although the revealed predictability depends on the specifications of volatility models.

Table 3 about here

Panel B of Table 3 lists the loss function ratios of Brent oil volatility forecasts. The predictive content of OVX for Brent volatility is a bit weaker. For example, the predictability of OVX disappears when using the FIEGARCH model. The plausible explanation is that OVX itself reflects expectation of WTI oil volatility, not Brent oil volatility. More importantly, the significant predictability is still present under most loss criteria when each of the other 5 GARCH-type models is considered.

4.2.2. MCS results

We have shown that OVX can predict future oil spot volatility based on pairwise comparison between two individual models, i.e., a GARCH-type model and the same model with OVX-based implied volatility. The pairwise comparison results do not imply that an OVX model performs better than all GARCH-type models. The reason is that the predictive ability varies over different GARCH-type models. It is possible that after putting OVX in a volatility model, one can obtain more accurate volatility forecasts, but the forecasting performance is still worse than some of the other volatility models without OVX. Which volatility models perform best? This is an important question in the volatility forecasting literature. Due to this motivation, we use a test proposed by Hansen et al. [32], named model confidence set (MCS) to compare the forecasting performances of all 12 models at one time, in which 6 models take into account OVX information and the remaining 6 models do not. The idea behind this test is that the data available may be not informative enough to yield a single model that can dominate all of its competitors significantly. In such cases, one can only obtain a smaller set of models, called the model confidence set, which contains the best forecasting model at a given level of confidence. Therefore, the models in the MCS perform equally well at the given confidence level[7].

We use the confidence level of 90% and show the MCS results in Table 4. When forecasting WTI oil volatility, most of the tested models are included in the MCS under 2 out of 6 loss criteria (MSE_1 and QLIKE), indicating that these models perform equally well. Under each of the other 4 loss criteria, the MCS only contains GARCH-OVX and FIGARCH-OVX model. This evidence indicates that these two models significantly outperform their competitors. GARCH and FIGARCH with OVX can significantly beat the other GARCH-type models with OVX and all GARCH-type models without OVX, further confirming the usefulness of OVX in predicting WTI volatility. The MCS results for Brent oil volatility forecasting are quite similar. FIGARCH-OVX is included in MCS under all six loss criteria and most of its competitors are excluded from MCS under all criteria except MSE_1. Therefore, FIGARCH-OVX is the best choice in forecasting Brent volatility among the 12 models under consideration.

Table 4 about here

4.2.3. The success ratio results

It may be argued that the loss functions are sensitive to extreme values of volatility forecasts. For this consideration, we employ an alternative evaluation

[7] To save space, we do not give the detailed description of MCS test, but refer interested readers to Hansen et al. [32].

criterion, success ratio (SR), which measures how often a GARCH-OVX model results in better volatility forecasts than a corresponding GARCH-type model without OVX. This criterion is given by,

$$SR = \frac{1}{T-M} \sum_{t=M+1}^{T} l_t \,, \quad l_t = I\left(\left| \hat{\sigma}_{1,t}^2 - \sigma_t^2 \right| < \left| \hat{\sigma}_{0,t}^2 - \sigma_t^2 \right| \right), \tag{17}$$

where $\hat{\sigma}_{1,t}^2$ and $\hat{\sigma}_{0,t}^2$ are, respectively, the volatility forecasts from a OVX model and the corresponding univariate benchmark. The indicator function *I*(.) takes the value of 1 when the condition in the parenthesis is satisfied and zero otherwise. Obviously, SR quantifies the probability that the OVX model performs better than the benchmark. Therefore, an SR higher than 0.5, the probability of tossing a coin, implies that adding OVX to a volatility model is more likely to lead to more accurate forecasts. Compared with the loss function criteria, SR is less affected by outliers. We use the standard *t*-statistic to test whether the difference between an SR and 0.5 is significant.

Table 5 reports the success ratios of different OVX-based volatility models. We can find that the success ratios are significantly higher than 0.5 for all cases. When the FIGARCH model is considered, the success ratio of OVX is as high as 0.6, regardless of whether WTI or Brent oil volatility is predicted. Therefore, GARCH-type models with OVX outperform the univariate benchmark most of time. Current OVX can always provide useful predictive information regarding next day's oil spot volatility.

Table 5 about here

To further understand the predictability from OVX to oil spot volatility, it is interesting to find whether the time series of indicator function l_t in equation (17) exhibit persistence of some type or are independently distributed. This is important, because if the values of l_t are independently distributed it would be impossible to predict whether the inclusion of OVX in a GARCH-type model can improve the forecasting accuracy. To address this issue, we follow Sarno and Valente [41] by using a fairly general test for the null hypothesis that l_t series are independent and identically distributed against an unspecified form of dependence, namely, the Broock et al. [42] test. The results reported in Table 6 show that the null hypothesis is rejected at the 1% significance level for all cases, regardless of which model specification is considered. These results indicate that the event that the models with OVX display superior forecasting performance over the univariate benchmark occurs persistently.

Table 6 about here

4.2.4. Longer-horizon forecasting performance

Until now, we have found the predictability from OVX to oil volatility for the horizon of one day. In this section, we investigate the predictive content of OVX for the longer horizons because it is of importance for long-term risk management. We consider two forecasting horizons of 5 days (one week) and 20 days (one month). The results in Table 7 show strong predictability from OVX to oil volatility for longer horizons. Under each of 6 loss criteria, the ratios of loss functions are significantly lower than 1 for almost all GARCH-type specifications[8].

Table 7 about here

Table 8 shows the results of MCS test for the longer-horizon performances of 6

[8] To save space, we do not report longer-horizon forecasting results for Brent oil volatility.

GARCH-type models with OVX and 6 univariate GARCH-type models in forecasting WTI oil volatility. We can find that for the forecasting horizon of 5 days, the univariate GARCH-type models without OVX are excluded from MCS in almost all cases, indicating the significant inferior forecasting performance than the models in the MCS. This evidence implies that OVX models can significantly outperform the traditional univariate models for the horizon of 5 days. Not all OVX models perform equally well. For example, GARCH-OVX and FIGARCH-OVX models are incorporated in the MCS under all six loss criteria, but the other GARCH-type models with OVX are excluded from MCS under some specific criteria only.

The forecasting performances for the horizon of 20 days are consistent in quality. Again, we find that the GARCH-OVX models perform significantly better than univariate GARCH models. FIGARCH-OVX and GARCH-OVX are the better candidates for long-horizon volatility forecasting.

Table 8 about here

4.2.5. Robustness test

It is possible whether forecast improvement can be obtained by adding OVX to the benchmark model is sensitive to omitted variables in GARCH-type models. To examine this question, we also consider following benchmark model to do robustness check:

$$\sigma_t^2 = f(\sigma_{t-1}^2, \varepsilon_{t-1}^2, \varepsilon_{t-1}) + \theta x_{t-1}, \tag{18}$$

where $f(\sigma_{t-1}^2, \varepsilon_{t-1}^2, \varepsilon_{t-1})$ denotes a GARCH-type specification, x_t is a vector with non-OVX macro variables. As an alternative to this benchmark, we add the variable of OVX-based implied volatility:

$$\sigma_t^2 = f(\sigma_{t-1}^2, \varepsilon_{t-1}^2, \varepsilon_{t-1}) + \theta x_{t-1} + \rho IV_{t-1}. \tag{19}$$

where *IV* is the daily implied volatility proxied by squared OVX divided by 252.

The sets of potential macro variables x_t which have influences on oil volatility should be very large. Too many variables in x_t will cause estimation problem of volatility models. For simplicity, we only consider short-term interest rate to have quick glance at the robustness of the revealed predictability from OVX to oil volatility. We use the 3-month Treasury bill rate available at the website of Federal Reserve Bank at Saint Louis. We employ the volatility models (18) and (19) to do forecast exercise again and evaluate the forecasting performance using the MCS test. The MCS results for alternative m for WTI oil volatility is reported in Table 9[9]. We find highly consistent evidence that almost all the univariate GARCH-type models without OVX are excluded from MCS under five of six loss criteria. GARCH-OVX, FIGARCH-OVX and FIEGARCH-OVX models work significantly better than their competitors and the forecasting performances of themselves are equally well because they are included in the MCS under all criteria. These results confirm the robustness of our core finding that OVX can predict oil volatility. Also, we can say that OVX and the macro variables such as short-term interest provide different information regarding future oil volatility.

Table 9 about here

5. Conclusions

[9] To save space, we do not show the results for Brent volatility.

We examine the information content of OVX when forecasting spot volatility of WTI and Brent oil returns. The OVX-based implied volatility is taken as a predictor of oil volatility in GARCH-type models. Our in-sample evidence suggests that past OVX can significantly affect current oil spot volatility, regardless which GARCH-type model specification is considered.

We pay more attention to out-of-sample exercise by investigating whether including OVX in a GARCH-type volatility model can improve its forecasting performance. We use 6 different GARCH-type specifications to do out-of-sample forecasting analysis and employ 6 popular criteria of loss functions to evaluate forecasting performances. Generally, our results suggest that OVX can predict oil spot volatility out-of-sample and the predictability exists for all the model specifications under consideration. The success ratio that the GARCH-type model with OVX outperforms the univariate benchmark without OVX is as high as 60%. Our analysis is further extended to longer-horizon forecasting. We find the predictability from OVX to oil spot volatility for weekly and monthly forecasting horizons. The robustness analysis shows that the revealed predictability is not affected by the inclusion of macro variables such as short-term interest rate in volatility models.

Acknowledgements

Li Liu acknowledges the financial support from the National Science Foundation of China (Nos.71401077 and 71771124), Qing Lan Project in Jiangsu province, and the Top-notch Academic Programs Project of Jiangsu Higher Education Institutions

(PPZY2015B104). Hongtao Chen thanks the financial support from the National Science Foundation of China (No.71503039), the Jiangsu Social Science Foundation for Youth (No.16GLC002), Philosophy and Social Sciences Research Project of Colleges and universities in Jiangsu of China (No. 2017ZDIXM013)

References

[1] Baumeister, C., Peersman, G., 2013. Time-varying effects of oil supply shocks on the US economy. American Economic Journal: Macroeconomics 5, 1-28.

[2] Hamilton, J. D., 1983. Oil and the macroeconomy since World War II. The Journal of Political Economy 91, 228-248.

[3] Hamilton, J. D., 1996. This is what happened to the oil price-macroeconomy relationship. Journal of Monetary Economics 38, 215-220.

[4] Jones, C. M., Kaul, G., 1996. Oil and the stock markets. The Journal of Finance 51, 463-491.

[5] Kilian, L., 2009. Not All Oil Price Shocks Are Alike: Disentangling Demand and Supply Shocks in the Crude Oil Market. The American Economic Review 99, 1053-1069.

[6] Kilian, L., Park, C., 2009. The impact of oil price shocks on the US stock market. International Economic Review 50, 1267-1287.

[7] Arouri, M. E. H., Jouini, J., Nguyen, D. K., 2012a. On the impacts of oil price fluctuations on European equity markets: Volatility spillover and hedging

effectiveness. Energy Economics 34, 611-617.

[8] Arouri, M. E. H., Lahiani, A., Lévy, A., Nguyen, D. K., 2012b. Forecasting the conditional volatility of oil spot and futures prices with structural breaks and long memory models. Energy Economics 34, 283-293.

[9] Chkili, W., Hammoudeh, S., Nguyen, D. K., 2014. Volatility forecasting and risk management for commodity markets in the presence of asymmetry and long memory. Energy Economics 41, 1-18.

[10] Hou, A., Suardi, S., 2012. A nonparametric GARCH model of crude oil price return volatility. Energy Economics 34, 618-626.

[11] Kang, S.H., Kang, S.M., Yoon, S.M., 2009. Forecasting volatility of crude oil markets. Energy Economics 31, 119-125.

[12] Kang, S. H., Yoon, S. M., 2013. Modeling and forecasting the volatility of petroleum futures prices. Energy Economics 36, 354-362.

[13] Lux, T., Segnon, M., Gupta, R., 2016. Forecasting crude oil price volatility and value-at-risk: Evidence from historical and recent data. Energy Economics 56, 117-133.

[14] Nomikos, N.K., Pouliasis, P.K., 2011. Forecasting petroleum futures markets volatility: The role of regimes and market conditions. Energy Economics 33, 321-337.

[15] Sadorsky, P., 2006. Modeling and forecasting petroleum futures volatility. Energy Economics 28, 467-488.

[16] Wang, Y., Liu, L., Ma, F., Wu, C., 2016. What the investors need to know about

forecasting oil futures return volatility. Energy Economics 57, 128-139.

[17] Wei, Y., Wang, Y., Huang, D., 2010. Forecasting crude oil market volatility: Further evidence using GARCH-class models. Energy Economics 32, 1447-1484.

[18] Engle, R.F., 1982. Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. Econometrica 50, 987-1007.

[19] Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31, 307-327.

[20] Glosten, L.R., Jagannathan, R., Runkle, D.E., 1993. On the relation between the expected value and the volatility of the nominal excess return on stocks. Journal of Finance 48, 1779-1801.

[21] Baillie, R. T., Bollerslev, T., Mikkelsen, H. O., 1996. Fractionally integrated generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 74, 3-30.

[22] Canina, L., Figlewski, S., 1993. The informational content of implied volatility. Review of Financial Studies 6, 659-681.

[23] Cao, C., Yu, F., Zhong, Z., 2010. The information content of option-implied volatility for credit default swap valuation. Journal of Financial Markets 13, 321-343.

[24] Christensen, B. J., Prabhala, N. R., 1998. The relation between implied and realized volatility. Journal of Financial Economics 50, 125-150.

[25] Jiang, G. J., Tian, Y. S., 2005. The model-free implied volatility and its

information content. Review of Financial Studies 18, 1305-1342.

[26] Jorion, P., 1995. Predicting volatility in the foreign exchange market. The Journal of Finance 50, 507-528.

[27] Agnolucci, P., 2009. Volatility in crude oil futures: a comparison of the predictive ability of GARCH and implied volatility models. Energy Economics 31, 316-321.

[28] Szakmary, A., Ors, E., Kim, J. K., Davidson, W. N., 2003. The predictive power of implied volatility: Evidence from 35 futures markets. Journal of Banking & Finance 27, 2151-2175.

[29] Haugom, E., Langeland, H., Molnár, P., Westgaard, S., 2014. Forecasting volatility of the US oil market. Journal of Banking & Finance 47, 1-14.

[30] Baumeister, C., Kilian, L., 2014. What central bankers need to know about forecasting oil prices. International Economic Review 55, 869-889.

[31] Diebold, F. X., Mariano, R., 1995. Comparing predictive accuracy. Journal of Business and Economic Statistics 13, 253–265.

[32] Hansen, P. R., Lunde, A., Nason, J. M., 2011. The model confidence set. Econometrica 79, 453-497.

[33] Nelson, D. B., 1991. Conditional heteroskedasticity in asset returns: A new approach. Econometrica: Journal of the Econometric Society 59, 347-370.

[34] Hansen, P. R., Lunde, A., 2005. A forecast comparison of volatility models: does anything beat a GARCH (1, 1)?. Journal of Applied Econometrics 20, 873-889.

[35] Cont, R., 2001. Empirical properties of asset returns: stylized facts and statistical

issues. Quantitative Finance 1, 223–236.

[36] Wang, Y., Liu, L., 2010. Is WTI crude oil market becoming weakly efficient over time?: New evidence from multiscale analysis based on detrended fluctuation analysis. Energy Economics 32, 987-992.

[37] Wang, Y., Wu, C., 2013. Efficiency of Crude Oil Futures Markets: New Evidence from Multifractal Detrending Moving Average Analysis. Computational Economics 42, 393-414.

[38] Davidson, J., 2004. Moment and memory properties of linear conditional heteroscedasticity models, and a new model. Journal of Business & Economic Statistics 22, 16–29.

[39] Inoue, A., Kilian, L., 2004. In-sample or out-of-sample tests of predictability: Which one should we use? Econometric Reviews 23, 371-402.

[40] Patton, A. J., 2011. Volatility forecast comparison using imperfect volatility proxies. Journal of Econometrics 160, 246-256.

[41] Sarno, L., Valente, G., 2009. Exchange rates and fundamentals: Footloose or evolving relationship?. Journal of the European Economic Association 7, 786-830.

[42] Broock, W. A., Scheinkman, J. A., Dechert, W. D., LeBaron, B., 1996. A test for independence based on the correlation dimension. Econometric Reviews 15, 197-235.

Figures

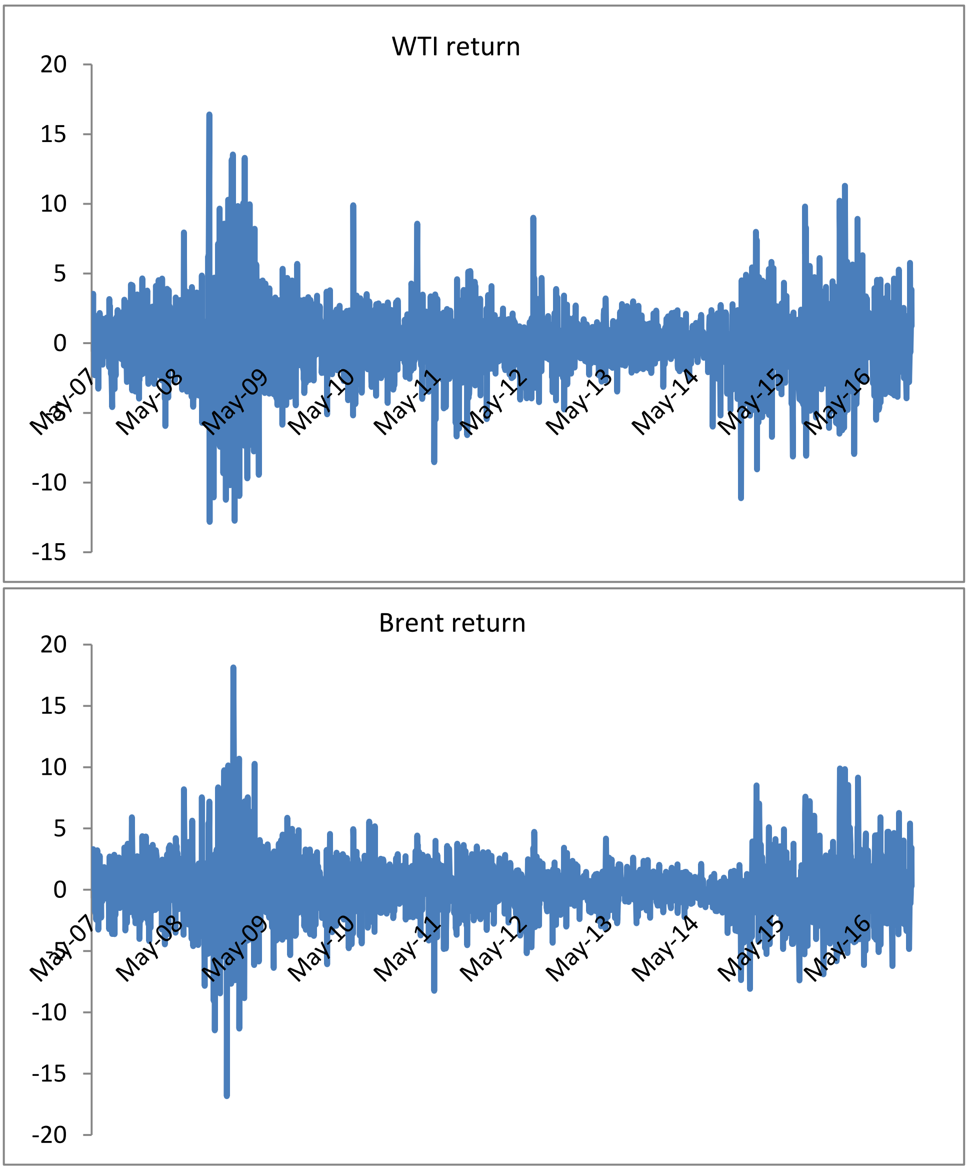

Figure 1 WTI and Brent oil returns

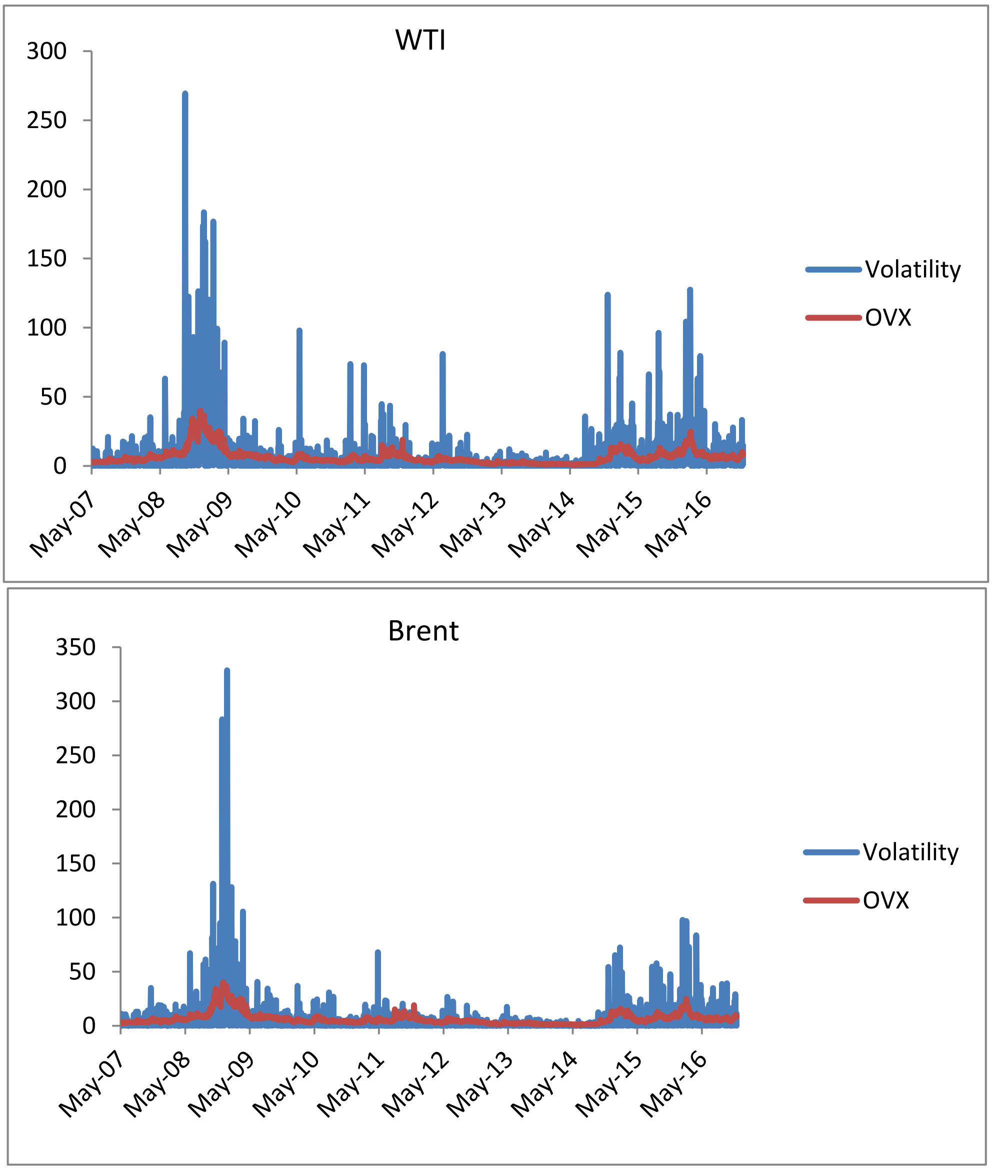

Figure 2 Oil spot volatility and OVX-based implied volatility

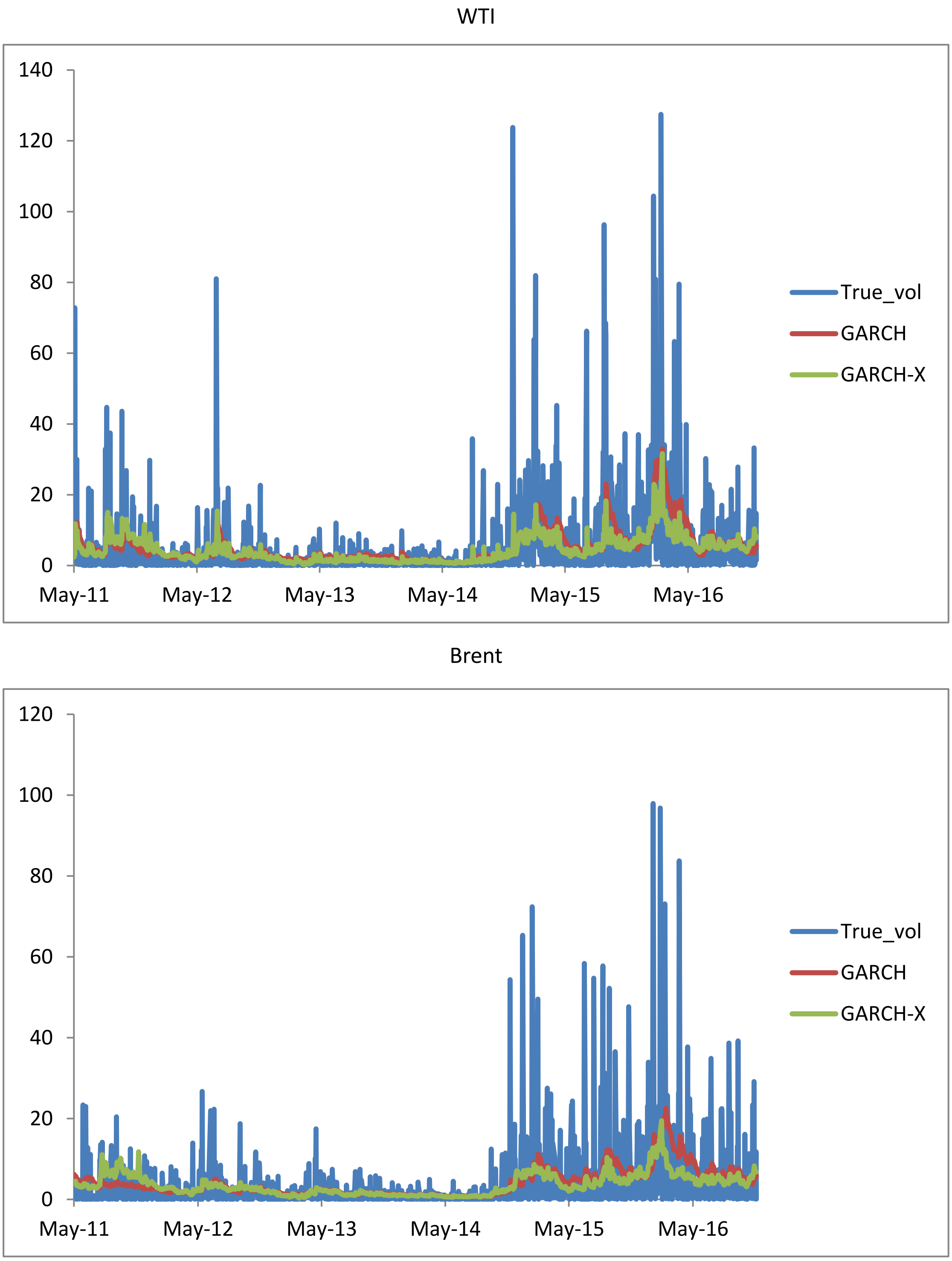

Figure 3 True values and forecasts of crude oil volatility

Tables

Table 1 Descriptive statistics of oil returns and volatilities

	Return		Volatility		
	WTI return	Brent return	WTI volatility	Brent volatility	Implied volatility
Mean	-0.011	-0.014	6.433	5.066	6.444
Median	0.021	0.035	1.599	1.202	4.861
Max.	16.41	18.130	269.4	328.7	40.017
Min.	-12.83	-16.832	0.000	0.000	0.834
Std.Dev.	2.537	2.251	16.26	13.99	5.468
Skewness	0.151	0.147	6.641	11.14	2.575
Kurtosis	7.390	8.627	65.90	203.52	11.290
JB	1934***	3125***	$>10^4$***	$>10^4$***	9512***
Q(10)	29.39***	26.24***	1367***	423.6***	$>10^4$***
ADF	-50.28***	-46.64***	-5.636***	-4.283***	-3.102**

Notes: This table provides the descriptive statistics of oil returns and volatilities. The implied volatility equals to the squared OVX divided by 252. JB denotes the Jarque-Bera statistics testing the null hypothesis of normal distribution. $Q(l)$ is the Ljung-Box Q statistics for the null hypothesis of serial correlation up to the lag order of l. ADF is the augmented Dickey-Fuller statistics which test the null hypothesis that the time series contain a unit root. The asterisks *, ** and *** denote rejections of null hypothesis at 10%, 5% and 1% significance levels, respectively.

Table 2 In-sample estimation results

	GARCH	GJR	EGARCH	FIGARCH	FIEGARCH	HYGARCH
WTI oil						
ρ	0.270***	0.062***	0.872***	0.338***	0.844***	0.406***
	(4.993)	(4.227)	(13.43)	(6.203)	(11.08)	(3.424)
ΔLog*L*	25	11	15	26	17	23
Brent oil						
ρ	0.705***	0.051***	1.127***	0.731***	0.939***	0.718***
	(3.498)	(4.376)	(26.87)	(4.584)	(14.24)	(5.645)
ΔLog*L*	33	18	33	33	37	36

Notes: This table reports in-sample estimation results of GARCH-type models with OVX-based implied volatility, the specifications of which are given by,

$$\sigma_t^2 = f(\sigma_{t-1}^2, \varepsilon_{t-1}^2, \varepsilon_{t-1}) + \rho IV_{t-1},$$

where $f(\sigma_{t-1}^2, \varepsilon_{t-1}^2, \varepsilon_{t-1})$ is a GARCH process. 6 different GARCH specifications are considered and their names are listed in the first row. The numbers in parentheses are the t-statistics. The changes in log likelihood of the model over the GARCH-type benchmark $\sigma_t^2 = f(\sigma_{t-1}^2, \varepsilon_{t-1}^2, \varepsilon_{t-1})$, $\Delta \text{Log}L$, are also given in the table. The asterisks *, ** and *** denote rejections of null hypothesis at 10%, 5% and 1% significance levels, respectively.

Table 3 1-day-ahead forecasting performance evaluated by loss functions

	MSE_1	MAE_1	R^2LOG	QLIKE	MSE_2	MAE_2
Panel A: Forecasting results for WTI oil volatility						
GARCH	**0.974***	**0.926*****	**0.916*****	**0.989***	**0.900*****	**0.925*****
	(-1.306)	(-5.920)	(-11.23)	(-1.361)	(-5.801)	(-8.161)
EGARCH	1.009	**0.981*****	**0.946*****	1.001	**0.981*****	**0.967*****
	(1.237)	(-3.120)	(-8.091)	(0.169)	(-2.019)	(-5.393)
GJR	**0.998**	**0.982*****	**0.951*****	**0.996**	**0.973*****	**0.968*****
	(-0.190)	(-2.488)	(-8.489)	(-0.524)	(-2.528)	(-5.254)
FIGARCH	**0.972***	**0.926*****	**0.928*****	**0.992**	**0.899*****	**0.933*****
	(-1.563)	(-6.092)	(-10.74)	(-0.928)	(-6.018)	(-7.506)
FIEGARCH	**0.969*****	**0.940*****	**0.943*****	**0.988****	**0.912*****	**0.946*****
	(-1.972)	(-5.518)	(-10.01)	(-1.888)	(-5.860)	(-6.880)
HYGARCH	**0.986**	**0.998**	**0.989*****	**0.999**	**0.992**	**0.994**
	(-0.777)	(-0.180)	(-1.982)	(-0.116)	(-0.573)	(-0.828)
Panel B: Forecasting results for Brent oil volatility						
GARCH	**0.970***	**0.949*****	**0.968*****	**0.995**	**0.927*****	**0.963*****
	(-1.357)	(-3.813)	(-4.718)	(-0.656)	(-3.632)	(-3.797)
EGARCH	**0.998**	**0.977*****	**0.947*****	**0.987*****	**0.971****	**0.964*****
	(-0.113)	(-2.474)	(-8.079)	(-2.009)	(-1.954)	(-4.877)
GJR	1.012	**0.988****	**0.972*****	**0.999**	**0.992**	**0.982*****
	(1.607)	(-1.729)	(-5.340)	(-0.141)	(-0.790)	(-3.090)
FIGARCH	**0.960*****	**0.938*****	**0.943*****	**0.979*****	**0.909*****	**0.945*****
	(-2.089)	(-5.314)	(-8.375)	(-2.322)	(-5.176)	(-6.200)
FIEGARCH	1.018	1.010	**0.994**	1.000	1.018	1.003
	(1.617)	(1.132)	(-1.139)	(-0.069)	(1.472)	(0.408)
HYGARCH	**0.955*****	**0.941*****	**0.962*****	**0.979*****	**0.912*****	**0.955*****
	(-2.298)	(-4.898)	(-5.937)	(-2.407)	(-4.942)	(-5.096)

Notes: This table shows the 1-day-ahead out-of-sample forecasting performance of GARCH-type models with OVX-based implied volatility, the specifications of which are given by,

$$\sigma_t^2 = f(\sigma_{t-1}^2, \varepsilon_{t-1}^2, \varepsilon_{t-1}) + \rho IV_{t-1},$$

where $f(\sigma_{t-1}^2, \varepsilon_{t-1}^2, \varepsilon_{t-1})$ is a GARCH process. 6 different GARCH specifications are considered and their names are listed in the first column. The numbers reported in the table are the ratio of loss functions of the GARCH-type models with implied volatility relative to the benchmark of univariate GARCH-type models $\sigma_t^2 = f(\sigma_{t-1}^2, \varepsilon_{t-1}^2, \varepsilon_{t-1})$.

The numbers in bolds denote that the corresponding GARCH-type model with implied volatility can beat the univariate benchmark. The Diebold and Mariano (1995) statistics testing for the equivalence of loss functions are given in the parentheses. The asterisks *, ** and *** denote rejections of null hypothesis at 10%, 5% and 1% significance levels, respectively.

Table 4 MCS results for 1-day-ahead forecasting performance

	MSE_1	MAE_1	R^2LOG	QLIKE	MSE_2	MAE_2
Panel A: MCS test results for WTI oil volatility						
GARCH	**0.185**	0	0	**0.034**	0	0
EGARCH	**0.569**	0	0	**0.363**	0.001	0
GJR	**0.941**	0	0	**0.689**	0	0
FIGARCH	**0.152**	0	0	0.090	0	0
FIEGARCH	**0.172**	0	0	0.044	0	0
HYGARCH	**0.328**	0	0	**0.648**	0.001	0
GARCH-X	**0.800**	**0.469**	**0.359**	**0.357**	**0.657**	**0.423**
EGARCH-X	**0.326**	0	0	**0.293**	0.005	0
GJR-X	**1.000**	0	0	**1.000**	0.015	0
FIGARCH-X	**0.765**	**1.000**	**1.000**	**0.334**	**1.000**	**1.000**
FIEGARCH-X	**0.929**	0.043	0.027	**0.404**	**0.402**	0.020
HYGARCH-X	**0.784**	0.001	0	**0.722**	0.009	0
Panel B: MCS test results for Brent oil volatility						
GARCH	0.046	0	0	0.012	0.002	0
EGARCH	**0.417**	0	0	0.026	0	0
GJR	**0.190**	0	0	**0.110**	0.004	0
FIGARCH	0.035	0.001	0	0.005	0.001	0
FIEGARCH	**1.000**	0	0	**0.706**	0.026	0
HYGARCH	0.023	0.001	0	0.006	0.001	0
GARCH-X	**0.234**	0	0	0.094	**0.261**	0
EGARCH-X	**0.578**	0.021	0.001	0.502	0.048	0.008
GJR-X	**0.105**	0.013	0	**0.134**	0.014	0.001
FIGARCH-X	**0.252**	**1.000**	**1.000**	**0.170**	**1.000**	**1.000**
FIEGARCH-X	**0.465**	0	0	**1.000**	0.005	0
HYGARCH-X	**0.240**	0	0	**0.152**	0.008	0

Notes: This table shows MCS test results for the 1-day-ahead out-of-sample forecasting performance of GARCH-type models with OVX-based implied volatility (GARCH-X) and the univariate benchmark without OVX. We use the confidence level of 90%. The numbers reported in this table are MCS p-values and the ones in bolds denote that the corresponding model is included in the MCS under a pre-specified loss criterion.

Table 5 Success ratios of implied volatility models

	GARCH	EGARCH	GJR	FIGARCH	FIEGARCH	HYGARCH
WTI	0.615***	0.561***	0.578***	0.603***	0.597***	0.528**
	(8.821)	(4.608)	(5.931)	(7.900)	(7.388)	(2.116)
Brent	0.577***	0.606***	0.608***	0.610***	0.550***	0.565***
	(5.731)	(8.014)	(8.188)	(8.305)	(3.731)	(4.837)

Notes: This table shows the success ratios of GARCH-type models with implied volatility. The numbers in parentheses are the t-statistics testing for the null hypothesis that the success ratio is equal to 0.5. The asterisks *, ** and *** denote the significance at 10%, 5% and 1% significance levels.

Table 6 Independence for occurrence of implied volatility model in beating univariate benchmark

	GARCH	EGARCH	GJR	FIGARCH	FIEGARCH	HYGARCH
m=3						
WTI	0.0114***	0.0101***	0.0100***	0.0101***	0.0065***	0.0041***
	(14.43)	(45.01)	(27.24)	(15.79)	(11.45)	(86.37)
Brent	0.0105***	0.0109***	0.0086***	0.0082***	0.0036***	0.0092***
	(29.28)	(15.91)	(12.05)	(11.16)	(23.47)	(35.68)
m=6						
WTI	0.0073***	0.0051***	0.0069***	0.0068***	0.0045***	0.0022***
	(18.79)	(51.64)	(29.21)	(22.22)	(17.14)	(108.7)
Brent	0.0067***	0.0061***	0.0056***	0.0048***	0.0015***	0.0055***
	(41.79)	(18.60)	(16.30)	(13.60)	(23.24)	(48.50)

Note: This table provides the BDS statistics of the Broock *et al.* (1996) test for the serial independence. The numbers in parentheses are the z-statistics. We test the independence for occurrence of the success of each implied volatility model in beating its univariate benchmark without OVX. The null hypothesis is that the series is independent against the unknown dependence. We report the results for the embedding dimensions m=3 and m=6. Asterisks denote rejection of the null hypothesis at 10%, 5% and 1% significance levels.

Table 7 Longer-horizon forecasting performance of WTI oil volatility evaluated by loss functions

	MSE_1	MAE_1	R^2LOG	QLIKE	MSE_2	MAE_2
Forecasting horizon h=5						
GARCH	**0.942*****	**0.892*****	**0.892*****	**0.968*****	**0.850*****	**0.896*****
	(-2..658)	(-7.687)	(-12.67)	(-3.699)	(-7.748)	(-10.03)
EGARCH	**0.997**	**0.967*****	**0.938*****	**0.993**	**0.960*****	**0.955*****
	(-0.355)	(-4.808)	(-8.934)	(-0.988)	(-3.761)	(-6.666)
GJR	**0.979***	**0.965*****	**0.934*****	**0.985*****	**0.946*****	**0.950*****
	(-1.513)	(-4.262)	(-9.385)	(-1.998)	(-4.305)	(-6.891)
FIGARCH	**0.943*****	**0.890*****	**0.894*****	**0.973*****	**0.849*****	**0.898*****
	(-2.785)	(-7.911)	(-13.37)	(-3.158)	(-7.860)	(-10.04)
FIEGARCH	**0.944*****	**0.908*****	**0.916*****	**0.975*****	**0.870*****	**0.915*****
	(-2.864)	(-7.146)	(-11.86)	(-3.217)	(-7.227)	(-9.097)
HYGARCH	**0.991**	**0.997**	**0.978*****	**0.988***	**0.984**	**0.989***
	(-0.697)	(-0.293)	(-3.858)	(-1.373)	(-1.078)	(-1.513)
Forecasting horizon h=20						
GARCH	**0.852*****	**0.812*****	**0.822*****	**0.929*****	**0.738*****	**0.821*****
	(-5.264)	(-11.74)	(-16.28)	(-6.653)	(-11.11)	(-15.15)
EGARCH	**0.996**	**0.926*****	**0.883*****	**0.980*****	**0.915*****	**0.907*****
	(-0.472)	(-9.205)	(-12.84)	(-2.382)	(-6.880)	(-11.81)
GJR	**0.940*****	**0.908*****	**0.869*****	**0.951*****	**0.859*****	**0.890*****
	(-3.561)	(-9.879)	(-13.96)	(-5.574)	(-9.254)	(-12.78)
FIGARCH	**0.863*****	**0.820*****	**0.837*****	**0.946*****	**0.751*****	**0.835*****
	(-5.040)	(-11.20)	(-15.62)	(-4.686)	(-10.46)	(-13.96)
FIEGARCH	**0.864*****	**0.840*****	**0.861*****	**0.937*****	**0.771*****	**0.855*****
	(-4.992)	(-10.25)	(-14.48)	(-6.157)	(-9.892)	(-12.99)
HYGARCH	**0.945*****	**0.964*****	**0.953*****	**0.970*****	**0.930*****	**0.960*****
	(-3.523)	(-3.407)	(-6.322)	(-3.797)	(-4.350)	(-4.784)

Notes: This table shows the 5-day-ahead and 20-day-ahead out-of-sample forecasting performance of GARCH-type models with OVX-based implied volatility, the specifications of which are given by,

$$\sigma_t^2 = f(\sigma_{t-1}^2, \varepsilon_{t-1}^2, \varepsilon_{t-1}) + \rho IV_{t-1},$$

where $f(\sigma_{t-1}^2, \varepsilon_{t-1}^2, \varepsilon_{t-1})$ is a GARCH process. 6 different GARCH specifications are considered and their names are listed in the first column. The numbers reported in the table are the ratio of loss functions of the GARCH-type models with implied volatility relative to the benchmark of univariate GARCH-type models $\sigma_t^2 = f(\sigma_{t-1}^2, \varepsilon_{t-1}^2, \varepsilon_{t-1})$.

The numbers in bolds denote that the corresponding GARCH-type model with implied volatility can beat the univariate benchmark. The Diebold and Mariano (1995) statistics testing for the equivalence of loss functions are given in the parentheses. The asterisks *, ** and *** denote rejections of null hypothesis at 10%, 5% and 1% significance levels, respectively.

Table 8 MCS results for longer-horizon forecasting performance of WTI oil volatility

	MSE_1	MAE_1	R^2LOG	QLIKE	MSE_2	MAE_2
Forecasting horizon *h*=5						
GARCH	0.004	0	0	0	0	0
EGARCH	0.051	0	0	0.004	0	0
GJR	**0.168**	0	0	0.044	0	0
FIGARCH	0.004	0	0	0	0	0
FIEGARCH	0.006	0	0	0.001	0	0
HYGARCH	**0.415**	0.001	0	**0.157**	0.002	0
GARCH-X	**0.184**	**0.101**	**0.381**	**0.109**	**0.500**	**0.381**
EGARCH-X	**0.101**	0.003	0	0.050	0.009	0.001
GJR-X	**1.000**	0.014	0.027	**0.660**	**0.139**	0.023
FIGARCH-X	**0.170**	**1**	**1**	**0.114**	**1**	**1**
FIEGARCH-X	**0.230**	0.035	**0.257**	**0.127**	**0.456**	**0.289**
HYGARCH-X	**0.702**	0	0	1	0.020	0
Forecasting horizon *h*=20						
GARCH	0	0	0	0	0	0
EGARCH	0.003	0	0	0	0	0
GJR	0.005	0	0	0	0	0
FIGARCH	0	0	0	0	0	0
FIEGARCH	0	0	0	0	0	0
HYGARCH	0.003	0	0	0.003	0	0
GARCH-X	**0.211**	**1**	**0.743**	**0.102**	**0.694**	**1**
EGARCH-X	0.009	0	0	0.002	0	0
GJR-X	0.055	0.059	0.050	**0.246**	0.047	**0.103**
FIGARCH-X	**0.220**	**0.653**	**1**	**0.105**	**1**	**0.684**
FIEGARCH-X	**0.136**	**0.415**	**0.241**	0.092	**0.290**	**0.342**
HYGARCH-X	1	0.001	0	**1**	0.004	0

Notes: This table shows MCS test results for the 5-day-ahead and 20-day-ahead out-of-sample forecasting performance of GARCH-type models with OVX-based implied volatility (GARCH-X) and the univariate benchmark without OVX. We use the confidence level of 90%. The numbers reported in this table are MCS p-values and the ones in bolds denote that the corresponding model is included in the MCS under a pre-specified loss criterion.

Table 9 MCS test results for 1-day-ahead forecasting performance of WTI oil volatility: alternative specification

	MSE_1	MAE_1	R^2LOG	QLIKE	MSE_2	MAE_2
GARCH	**0.251**	0	0	0.026	0.000	0
EGARCH	**0.449**	0	0	**0.226**	0.002	0
GJR	**1**	0	0	**0.608**	0.001	0
FIGARCH	**0.184**	0	0	0.059	0	0
FIEGARCH	**0.269**	0	0	0.035	0	0
HYGARCH	**0.575**	0	0	1	0.007	0
GARCH-X	**0.526**	**1**	**0.493**	**0.341**	**0.269**	**1**
EGARCH-X	**0.189**	0.009	0.029	**0.125**	**0.157**	0.021
GJR-X	**0.650**	0.002	0.000	**0.728**	0.020	0.000
FIGARCH-X	**0.535**	**0.571**	**0.486**	**0.352**	**0.254**	**0.590**
FIEGARCH-X	**0.536**	**0.451**	**1**	**0.353**	**1**	**0.512**
HYGARCH-X	**0.422**	0.000	0	**0.727**	0.003	0

Notes: This table shows MCS test results for 1-day-ahead out-of-sample forecasting performance of GARCH-type models with OVX-based implied volatility (GARCH-X)

$$\sigma_t^2 = f(\sigma_{t-1}^2, \varepsilon_{t-1}^2, \varepsilon_{t-1}) + \theta x_{t-1} + \rho IV_{t-1},$$

and the benchmark GARCH-type models,

$$\sigma_t^2 = f(\sigma_{t-1}^2, \varepsilon_{t-1}^2, \varepsilon_{t-1}) + \theta x_{t-1},$$

where $f(\sigma_{t-1}^2, \varepsilon_{t-1}^2, \varepsilon_{t-1})$ denotes a GARCH-type specification, x_t is short-term interest rate and IV_t is a OVX-based implied volatility. We use the confidence level of 90%. The numbers reported in this table are MCS p-values and the ones in bolds denote that the corresponding model is included in the MCS under a pre-specified loss criterion.