ARTICLE IN PRESS

Economic Modelling xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Economic Modelling

journal homepage: www.elsevier.com/locate/econmod

Forecasting the oil futures price volatility: A new approach

Feng Ma[a], Jing Liu[a], Dengshi Huang[a], Wang Chen[b,*]

[a] School of Economics & Management, Southwest Jiaotong University, Sichuan, China
[b] College of Finance and Economics, Yangtze Normal University, Chongqing, China

ARTICLE INFO

JEL:
C22
C52
C55

Keywords:
Volatility forecasting
Oil futures price
Realized range-based volatility
Volatility of realized range-based volatility
Model confidence set

ABSTRACT

This study provides a new perspective of modelling and forecasting realized range-based volatility (RRV) for crude oil futures. We are the first to improve the Heterogeneous Autoregressive model of Realized Range-based Volatility (HAR-RRV) model by considering the significant jump components, signed returns and volatility of realized range-based volatility. The empirical results show that the volatility of volatility significantly exists in the oil futures market. Moreover, our new proposed models with significant jump components, signed returns and volatility of volatility can gain higher forecast accuracy than HAR-RRV-type models. The results are robust to different forecasting windows and forecasting horizons. Our new findings are strategically important for investors making better decisions.

1. Introduction

High-frequency financial data contains more information and could help financial market investors gain higher returns and reduce risk in time. Special issues on high-frequency financial data have aroused great attention in recent years such as modelling and forecasting asset volatility. Specifically, improving the forecast accuracy of crude oil futures, which is strategically important for derivative pricing, portfolio selection, and risk management, has been attracted much concerns for the scholars and financial market participants. However, forecasting oil futures volatility is an intractable issue due to the complexion of crude oil market.

Since the realized volatility (RV) is proposed as a proxy for the latent volatility in Andersen and Bollerslev (1998), previous studies mainly model and forecast asset volatility based on the RV. Subsequently, a bunch of forecasting models are constructed such as the popular HAR-RV model and its various extensions (Andersen et al., 2007a; McAleer and Medeiros, 2008; Corsi, 2009; Corsi et al., 2010; Todorova and Souček, 2014; Huang et al., 2016). These HAR-RV-type models show their great superiority over the traditional GARCH-class models in forecasting asset volatility (e.g., Andersen et al., 2003; Koopman et al., 2005; Hansen and Lunde, 2005; Wei et al., 2010; Çelik and Ergin, 2014; Ma et al., 2015).

However, RV loses some information and fails to identify the daily integrated variance of the frictionless equilibrium price with market microstructure noise (Bandi and Russell, 2008). Martin and Dick (2007) and Christensen and Podolskij (2007) introduce another volatility measure, called the realized range-based volatility (RRV). The RRV is simply calculated by the difference between the maximum and minimum prices during a certain period, while the realized volatility (RV) is defined as the sum of non-overlapping squared returns within a fixed period. Christensen and Podolskij (2007) derive the asymptotic properties of the realized range and state that it is theoretically five times more efficient than the realized variance sampled at the same frequency and converges to the integrated variance at the same rate. Christensen and Podolskij (2012) further reveal that the range-based multi-power variations could be significantly efficient over comparable sub-sampled return-based estimators. Moreover, although rarely less literatures focus on the RRV, some studies find that the RRV is much more efficient and sufficient in modelling and forecasting volatility (Tseng et al., 2009; Tseng et al., 2012; Ma et al., 2015; Liu et al., 2017).

In addition, based on the realized range-based volatility, the HAR-RRV model and its extensions are constructed, and some of them are applied in forecasting the volatility of oil futures price (e.g., Tseng et al., 2009; Liu et al., 2017). For example, by considering leverage terms related to negative returns, we can extend HAR-RRV to the leverage HAR-RRV (LHAR-RRV) model by following Corsi et al. (2010). By considering the jump components, HAR-RRV-J and HAR-RRV-CJ model can be obtained. Particularly, Tseng et al. (2009) demonstrate that HAR-RRV-type models, which involve a realized range-based estimator, can successfully capture the long-term memory behavior

* Corresponding author.
E-mail addresses: mafeng2016@swjtu.edu.cn (F. Ma), liujing@my.swjtu.edu.cn (J. Liu), dengshi_huang@126.com (D. Huang), mrchenwang@126.com (W. Chen).

http://dx.doi.org/10.1016/j.econmod.2017.04.020
Received 7 February 2017; Received in revised form 12 April 2017; Accepted 22 April 2017
0264-9993/

in volatility of crude oil futures.

Nevertheless, some issues remain to be addressed to improve the forecasting models' performances. First, more accurate RRV forecasts are required in forecasting crude oil futures. Second, whether including the volatility of the RRV can improve the HAR-RRV-type models' performances remains unsolved. As far as we are aware, the GARCH-class models have also achieved great success for modelling low-frequency data (e.g., Glosten et al., 1993; Baillie et al., 1996; Bollerslev and Mikkelsen, 1996; Cheong, 2009; Agnolucci, 2009; Wei et al., 2010; Mohammadi and Su, 2010; Nomikos and Andriosopoulos, 2012; Charles and Darné, 2014; Efimova and Serletis, 2014; Lean and Smyth, 2015). Among them, the Fractionally Integrated EGARCH (FIEGARCH) can capture the heterogeneous leverage effects and the persistence (Bollerslev and Mikkelsen, 1996), while the GJR-GARCH model can account for heteroskedasticity in observed volatility sequences. Additionally, Corsi et al. (2008) propose a GARCH error process to account for the time varying conditional heteroscedasticity of the normally distributed HAR errors called volatility of realized volatility. Caporin and Velo (2015) use GJR-GARCH model to account for asymmetric effects in the volatility of volatility. Thus, GARCH-class models could provide a new path for modelling the volatility of RRV and gain higher forecast accuracy.

In this study, by using the RRV to measure oil futures volatility, we investigate whether the volatility of RRV help forecast the oil futures volatility and propose several new volatility models by considering the GARCH-type components. We use GARCH, GJR-GARCH and FIEGARCH model to describe the volatility of RRV and thus construct HAR-RRV-GARCH-type models. For instance, we can obtain the LHAR-RRV-CJ-GARCH model by combining the leverage, jumps and heterogeneity in RRV.

Our contributions to the previous literatures are threefold. First, although the RRV carries more information and its forecast precision is proved to be much greater than that of the RV, extant literature mainly modelling and forecasting the volatility based on the RV and the empirical analysis based on the RRV remains obscure. This paper models and forecasts oil futures price volatility in the framework of the RRV. Second, we are the first to forecast oil futures volatility by further modelling the volatility of the RRV. Third, we provide a new perspective on improving forecast accuracy of oil futures volatility by incorporating popular HAR-RRV-type models with GARCH components. Consequently, this study provides a comprehensive comparison of new proposed HAR-RRV-GARCH-type models with the popular ones. To forecast oil futures RRV, we construct eighteen HAR-RRV-type models with GARCH components. Their forecasting performances are compared with those of basic HAR-RRV-type models by the model confidence set (MCS) under different forecasting windows (Hansen et al., 2011). The results show that the LHAR-RRV-CJ-GARCH model proposed by this paper exhibits absolutely advantage in forecasting oil futures volatility. Moreover, the new models with significant jump components, signed returns and the volatility of RRV can gain higher forecasts accuracy at different forecast horizons. We find that using the GARCH-type models to describe the volatility of RRV can substantially help in forecasting the oil futures volatility.

The remainder of the paper is organized as follows. The next section briefly introduces the realized range-based volatility measurement and forecasting models, such as HAR-RRV and its various extensions. Section 3 describes the empirical data, for example, the mean and standard error of volatility estimators. In-sample estimation, out-of-sample evaluation, and robustness check are discussed in Section 4. Section 5 presents the study's conclusions.

2. Realized range-based volatility measurement and forecasting models

In this section, we will introduce the realized range-based volatility measurement jump test and the heterogeneous autoregressive model of realized range-based volatility (HAR-RRV) and its various extensions. Moreover, we will consider the volatility of volatility to combine those models to investigate the forecasting performance of the crude oil market.

2.1. Realized range-based volatility measurement

Consider the equidistant partition $0 < t_0 < t_1 < \ldots < t_n = 1$, where $t_i = i/n$ and $\Delta = 1/n$, for $i = 1, 2\ldots, n$.The intraday range at sampling times t_{i-1} and t_i is:

$$s_{p,\Delta_i} = \sup_{t_i \le t_s \le t_i} \{p_t - p_s\}, \tag{1}$$

where p is the log-price. The realized range-based volatility (henceforth RRV) estimator for the interval [0, 1] is defined as:

$$RRV_t^{\Delta} = \frac{1}{\lambda_{2,m}} \sum_{j=1}^{1/\Delta} s_{(t-1)+j*\Delta,\Delta}^2 \tag{2}$$

where $\lambda_{r,m} = E(s_{W,m}^r).\lambda_{r,m}$ is the r_{th} moment of the range of Brownian motion over a unit interval. To obtain $\lambda_{r,m}$ for intraday intervals of m equidistant prices each, we follow the works of Todorova (2012) and Todorova and Husmann (2012), a Brownian motion is simulated[1]:

$$B_t = B_{t-1} + \frac{1}{m}\varepsilon_i, \qquad i = 1,\ldots,m, \quad \varepsilon_i \sim N(0, 1). \tag{3}$$

Christensen and Podolskij, 2006) evidently find that when $\triangle \to 0$, realized range-based volatility (RRV) has satisfied as below:

$$RRV_t^{\Delta} \to \int_0^t \sigma^2(s)ds + \frac{1}{\lambda_{2,m}} \sum_{0<s\le t} \kappa^2(s). \tag{4}$$

where $\int_0^t \sigma^2(s)ds$ is the Integrated Volatility, which can be calculated by the realized range-based power variation (RBV) when $\triangle \to 0$:

$$RBV_{t,m} = \frac{1}{\lambda_{1,m}^2} \sum_{j=2}^{1/\Delta} |s_{(t-1)+j*\Delta}||s_{(t-1)+(j-1)*\Delta}|. \tag{5}$$

$\sum_{0<s\le t} \kappa^2(s)$ represents the jump component. Following the studies of the Christensen and Podolskij (2006), we have the new range-based estimator:

$$RRV_{t,m} = \lambda_{2,m} RRV_t^{\Delta} + (1-\lambda_{2,m}) RBV_{t,m} \to \int_0^t \sigma^2(s)ds + \sum_{0<s\le t} \kappa^2(s). \tag{6}$$

2.2. Jump test

Notably, we find that there are many jump tests in the framework of realized volatility, for example, Z test (Huang and Tauchen, 2005), C-Tz test (Corsi et al., 2010), LM test (Lee and Mykland, 2008). It is interesting to us that Christensen and Podolskij (2006) have also proposed the significant jump test based on the realized range-based volatility, and labeled as $Z_{t,m}$, which defined as below:

$$Z_{t,m} = \frac{\sqrt{n}\,(1 - RBV_{t,m}/RRV'_{t,m})}{\sqrt{\nu_m \max\left\{1/t, \frac{RQQ_{t,m}}{(RBV_{t,m})^2}\right\}}} \to N(0, 1), \tag{7}$$

where $\nu_m = \lambda_{2,m}^2(\Lambda_m^R + \Lambda_m^B - 2\Lambda_m^{RB}), \Lambda_m^B = (\lambda_{2,m}^2 + 2\lambda_{1,m}^2\lambda_{2,m} - 3\lambda_{1,m}^4)/\lambda_{1,m}^4$, $\Lambda_m^{RB} = (2\lambda_{3,m}\lambda_{1,m} - 2\lambda_{2,m}\lambda_{1,m}^2)/\lambda_{2,m}\lambda_{1,m}^2, \Lambda_m^B = (\lambda_{4,m} - \lambda_{2,m}^2)/\lambda_{2,m}^2$. In addition, Range-based Quad-power Quarticity ($RQQ_{t,m}$) can be calculated by:

$$RQQ_{t,m} = \frac{n}{\lambda_{1,m}^4} \sum_{j=4}^{1/\Delta} |s_{(t-1)+j*\Delta}||s_{(t-1)+(j-1)*\Delta}||s_{(t-1)+(j-2)*\Delta}||s_{(t-1)+(j-3)*\Delta}|. \tag{8}$$

[1] The simulation results can be obtained by the readers' request.

where the $\lambda_{1,m}$ can be simulated by the Eq. (3). To ensure the jump components are non-negative, we have the following formulations:

$$J_{t,m} = \max(RRV_{t,m} - RBV_{t,m},\ \ 0). \tag{9}$$

Inspired by Andersen et al. (2007a) and given the significance level, α,[2] we can naturally gain the significant jumps and defined as below:

$$CJ_{t,m} \equiv I(Z_{t,m} > \Phi_\alpha)\cdot[RRV_{t,m} - RBV_{t,m}]. \tag{10}$$

2.3. Forecasting models

Corsi (2009) propose the Heterogeneous Autoregressive model of Realized Volatility (HAR-RV) based on the Heterogeneous Market Hypothesis presented by Muller et al. (1993), which recognizes the presence of heterogeneity across traders. This model can capture some of the stylized facts found in financial asset return volatility, such as long memory and multi-scaling behavior. Moreover, the HAR-RV model can be estimated easily and includes three components: lagged average realized volatility, weekly realized volatility, and monthly realized volatility, which can be expressed as:

$$RV_{t+h} = c + \beta_d RV_t + \beta_w RV_{t-4,t} + \beta_m RV_{t-21,t} + \omega_{t+h}. \tag{11}$$

where $RV_{t-4,t} = (RV_t + RV_{t-1} + RV_{t-2} + RV_{t-3} + RV_{t-4})/5$,- $RV_{t-21,t} = (RV_t + RV_{t-1} + \cdots + RV_{t-21})/22$. The subscript h represents the forecast horizons, which usually set the values of 1, 5, 22, which correspond to one-day, one-week, and one-month ahead, respectively. Inspired by the Corsi (2009) and Muller et al. (1993), there are growing literatures to modelling and forecasting the realized range-based volatility using the Heterogeneous Autoregressive model of Realized Range-based Volatility (HAR-RRV), for example, Tseng et al. (2009), Todorova (2012), Todorova and Husmann (2012), Caporin and Velo (2015), Liu et al. (2017). The specifications of HAR-RRV model are as below:

$$RRV_{t+h} = c + \beta_d RRV_t + \beta_w RRV_{t-4,t} + \beta_m RRV_{t-21,t} + \omega_{t+h}. \tag{12}$$

where the $RRV_{t-4,t}$ and $RRV_{t-21,t}$ are the average of the lagged weekly and monthly realized volatility, respectively.

Add the jump components of Eq. (20) to the HAR-RRV model, we have the second model in this study, termed HAR-RRV-J, which is similar to HAR-RV-J proposed by Andersen et al. (2007a),

$$RRV_{t+h} = c + \beta_d RRV_t + \beta_w RRV_{t-4,t} + \beta_m RRV_{t-21,t} + \beta_j J_t + \omega_{t+h}. \tag{13}$$

Based on the Andersen et al. (2007a) and Tseng et al. (2009), we have the HAR-RRV-CJ model, which contains the continuous sample path and significant jump components and can be defined as:

$$\begin{aligned} RRV_{t+h} = {} & c + \beta_d CRRV_t + \beta_w CRRVW_{t-4,t} + \beta_m CRRVM_{t-21,t} + \beta_{jd} CJ_t \\ & + \beta_{jw} CJ_{t-4,t} + \beta_{jm} CJ_{t-21,t} + \omega_{t+h}. \end{aligned} \tag{14}$$

where $CRRV_{t,\alpha} = I(Z_t \le \Phi_\alpha)\cdot RRV_t + I(Z_t > \Phi_\alpha)\cdot RBV_t$, which the continuous components.$CRRVW_{t-4,t}$ and $CRRVM_{t-21,t}$ are the lagged weekly and monthly realized range-based volatility, respectively. CJ_t is the significant jump components, which can be obtained by Eq. (10).$CJ_{t-4,t}$ and $CJ_{t-21,t}$ are the average of lagged weekly and monthly significant jump components.

Corsi and Renò (2012) introduce the negative returns to the realized volatility model and capture the leverage effect, and find that those variables can gain higher forecast accuracy. Thus, we also add those variables to the HAR-RRV-CJ model, labeled as LHAR-RRV-CJ,

$$\begin{aligned} RRV_{t+h} = {} & c + \beta_d CRRV_t + \beta_w CRRVW_{t-4,t} + \beta_m CRRVM_{t-21,t} + \beta_{jd} CJ_t \\ & + \beta_{jw} CJ_{t-4,t} + \beta_{jm} CJ_{t-21,t} + \beta_{rd} r_t^- + \beta_{rw} r_{t-4,t}^- + \beta_{rm} r_{t-21,t}^- + \omega_{t+h}. \end{aligned} \tag{15}$$

where r_t^-=min(r_t,0), r_t is the daily return. $r_{t-4,t}^-$ and $r_{t-21,t}^-$ are the lagged average weekly and monthly returns.

Based on Tauchen and Zhou (2011), Prokopczuk et al. (2015) propose the HAR-RV-ARJ model, which considers the sign of the jumps that have an asymmetric impact on volatility. We first add the sign of the jumps to the HAR-RRV and construct a new model, termed HAR-RRV-ARJ,

$$\begin{aligned} RRV_{t+h} = {} & c + \beta_d RRV_t + \beta_w RRV_{t-4,t} + \beta_m RRV_{t-21,t} + \beta_{jp} RJ_t^+ \\ & + \beta_{jn} RJ_t^- + \omega_{t+h}. , \end{aligned} \tag{16}$$

where the $RJ_t = \mathrm{sign}(r_t)\cdot J_t . RJ_t^+ = \max(RJ_t, 0)$ and $RJ_t^- = min(RJ_t, 0)$.

Dumitru and Urga (2012) use eight nonparametric jump tests and find that the ABD-LM jump test (e.g., Andersen et al., 2007b; Lee and Mykland, 2008) has the overall best performance.[3] Moreover, Liu et al. (2016) find that jump component (henceforth JLM) calculated by ABD-LM jump test can help forecast performance and also point out that ABD-LM jump components' decomposition forms based on signed returns can significantly improve the models' forecasting performance. Therefore, we use the significant jump components to add the HAR-RRV model and have a new model, named as HAR-RRV-JLM, which are expressed as follows:

$$\begin{aligned} RRV_{t+h} = {} & c + \beta_d RRV_t + \beta_w RRV_{t-4,t} + \beta_m RRV_{t-21,t} + \beta_{jp} JLM_t^+ + \beta_{jn} JLM_t^+ \\ & + \omega_{t+h}. \end{aligned} \tag{17}$$

Following the works of Lee and Mykland (2008), we can calculate the "significant" jump components, which is defined as the sum of intraday squared returns and the returns occur the significant jumps ($JLM_t = \sum r_i^2$). Furthermore, we decompose the JLM into positive and negative jumps based on signed returns. Consequently, we have the JLM^+ and JLM^- defined as:

$$JLM_t^+ = \sum I_{(r_{t,i} > 0)} r_{t,i}^2 \quad \text{and} \quad JLM_t^- = \sum I_{(r_{t,i} < 0)} r_{t,i}^2 \tag{18}$$

Considered the abovementioned analysis, we have one HAR-RRV and five extended models, and termed as linear HAR-RRV-type models.

To the best of our knowledge, we find that there are extremely rare literatures on modelling and forecasting the volatility of volatility on the framework of the realized range-based volatility. Corsi et al. (2008) find that the residuals of the HAR-RV (ω_{t+h}, Eq. (1)) have strongly the observed volatility clustering. They use the generalized autoregressive conditional heteroskedasticity (GARCH) to describe those stylized facts and show that the improvement of considering the realized volatility of volatility can gain higher forecast accuracy and better density forecasts. Inspired by Corsi et al. (2008), we propose the HAR-RRV-type models and combine with the GARCH model, which can model the volatility of realized range-based volatility, termed as HAR-RRV-GARCH, HAR-RRV-J-GARCH, HAR-RRV-CJ-GARCH, LHAR-RRV-CJ-GARCH, HAR-RRV-ARJ-GARCH, HAR-RRV-JLM-GARCH. For example, the HAR-RRV-GARCH(p,q) model, which can be expressed by as below:

$$RRV_{t+h} = c + \beta_d RRV_t + \beta_w RRV_{t-4,t} + \beta_m RRV_{t-21,t} + \varepsilon_{t+h}, \tag{19}$$

$$\varepsilon_{t+h} = \sigma_{t+h} z_{t+h}, \tag{20}$$

$$\sigma_{t+h}^2 = \omega + \alpha(L)\varepsilon_{t+h-L}^2 + \beta(L)\sigma_{t+h-L}^2. \tag{21}$$

where $\alpha(L) = \alpha_1 L + \alpha_2 L^2 + \ldots + \alpha_q L^q$ and $\beta(L) = \beta_1 L + \beta_2 L^2 + \ldots + \beta_p L^p$ $.z_t \sim SkT(\nu, \lambda), SkT$ represents the Skew-t distribution.

[2] In this study, we will set, α=0.999. Considered the different values of α (0.99 and 0.95), the results are quite similar. You can available upon the request of the other results.

[3] Please see more details about ABD-LM jump in Lee and Mykland (2008) and Dumitru and Urga (2012).

Moreover, Caporin and Velo (2015) use the GJR-GARCH model to account for heteroskedasticity in observed volatility sequences and asymmetric effects in the volatility of volatility. In addition, in order to capture long range dependence in the volatility of realized volatility, Louzis et al. (2012) propose a Fractionally Integrated GARCH (FIGARCH) implementation for the conditional heteroscedasticity of the residuals and demonstrate that those improvement exhibits superior in sample fitting and out of sample volatility forecasting performance. As we are known, the FIEGARCH the Fractionally Integrated EGARCH (FIEGARCH) of Bollerslev and Mikkelsen (1996) can capture the heterogeneous leverage effects and the persistence. Given those issues, we further introduce the GJR-GARCH and FIEGARCH models to account for those stylized facts and combine with the HAR-RRV-type models to evaluate whether those extended models can gain higher forecast accuracy and whether those models including the leverage and long range dependence have better forecasting performance. The GJR-GARCH(p,q) and FIEGARCH(p,q) can be written as follows:

$$\sigma_{t+h}^2 = \omega + [\alpha(L) + \gamma(L)I(\varepsilon_{t+h-L} < 0)]\varepsilon_{t+h-L}^2 + \beta(L)\sigma_{t+h-L}^2. \quad (22)$$

$$\log(\sigma_{t+h}^2) = \omega + \varphi(L)^{-1}[1-L]^{-d}[(1 + \alpha(L))]g(z_{t-1}). \quad (23)$$

where $1-\beta(L) = \varphi(L)[1-L]^d$ and $g(z_t) = \theta z_t + \gamma[|z_t| - E|z_t|]$. $I(.)$ is an indicator function. The γ is the asymmetric leverage coefficient, which describes the volatility leverage effect. In addition, d is the fractional integration parameter and L is the lag operator. The parameter d characterizes the long-memory property of hyperbolic decay in volatility because it allows autocorrelations to decay at a slow hyperbolic rate. The θ generates an asymmetry in news impact on volatility. Therefore, we have another twelve HAR-RRV-GJR-type and HAR-RRV-FIEGARCH-type models. Our paper first introduces the volatility of RRV using the GARCH, GJR and FIEGARCH model to the HAR-RRV-type models and investigate the forecasting performances of those proposed models in the crude oil markets. Considered the logarithms of the realized range-based volatility estimators are closer to the Gaussian distribution, we analyze the logarithmic forms of the HAR-RRV and its extended models in this study.

3. Data description

Following the works of Haugom et al. (2014), Sévi (2014), Sévi (2015) we use the 5-min high-frequency data of one front month oil futures to investigate the forecasting performance of those existing and extended models in this study. Electronic trading begins at 6:00 p.m. and runs through 5:15 p.m. the next day on Sunday through Friday. Electronic trading breaks for only 45-min each day starting at 5:15 p.m. The contract unit is for 1,000 barrels and the price is quoted in U.S. dollars. The data comes from the Thomson Reuters Tick History Database, from the period that starts on January 1, 2007 to June 30, 2016. After removing days with a shortened trading session or too few transactions, we obtain 2369 observations daily.

From the results of Table 1, we find that all of series, such as realized volatility, continuous components, (signed) jump and "significant" jump have significantly right skewness and high excess kurtosis at the 99% confidence level and reject the null hypothesis: "skewness = 0" and "kurtosis-3=0". The Jarque-Bera statistics further imply that all of series are non-Gaussian distributions at the 1% significance level. Moreover, the Augmented Dickey–Fuller statistics show that all the time series are stationary at the 1% significance level, because of rejecting the null hypothesis of a unit root. Finally, the results of the Ljung–Box test show that the RRV, CRRV, JLM and signed JIM variables are serial autocorrelations up to the 20th order in the crude oil market. However, the rest of variables seem not to have significant autocorrelations.

4. Empirical results

4.1. In-sample fit analysis

Following the studies of Corsi et al. (2008), Caporin and Velo (2015) and Louzis et al. (2012), we just choose the GARCH(1,1), GJR-GARCH(1,1) and FIEGARCH(1,1) to describe the volatility of realized range-based volatility, respectively. To save space, we only show the results of the HAR-RRV and its extended models at different horizons (next day (h=1), next week (h=5) and next month (h=22)) using the OLS and Newey-West correction methods[4] in Table 2. The parameters of realized range-based volatility estimators, such as RRV, RRVW and RRVM, are all significant at the 99% confidence level. We find that the coefficients of RRV have more weights at short horizons, implying that the daily RRV has the strongest influence on the next day's volatility level in the crude oil futures market, which tell us that short-term traders are mainly concerned about short-term volatility. For the middle and long horizons, the investors pay more attention to the long-term volatility. The results of the Engle's LM ARCH test, assuming there is no ARCH effect of the residuals, strongly exhibit that the HAR-RRV model has a significant ARCH effect at the 99% confidence level. The asymmetric leverage coefficient, γ, are significant in different horizons, which show that the volatility of realized range-based volatility exist the asymmetric effect. However, the memory parameter d of the HAR-RRV-FIEGARCH model at different horizons seems not to reject the null hypothesis, "d=0", implying that the long-memory property is very weak in the volatility of RRV.

4.2. Evaluate the linear HAR-RRV-type and HAR-RRV-GARCH-type with MCS test

In this study, we use the popular method, termed rolling window, to obtain the future volatility. First, we divide our sample into two parts: (a) in-sample, covering the first 1569 trading days from January 2, 2007 to April 16, 2013; (b) out-of-sample, covering the last 800 trading days from April 17, 2013 to June 30, 2016. The estimation period is

Table 1
Descriptive statistics of all variables.

	Mean	St. dev.	Skewness	Kurtosis (excess)	Jarque-Bera	ADF	Q(5)
RRV	0.495	1.108	16.970***	414.498***	17072671.369***	−6.839***	896.885***
CRRV	0.373	0.456	3.062***	10.996***	15577.694***	−3.587**	8249.922***
J	0.122	0.922	24.757***	702.248***	48734442.525***	−13.923***	0.191
RJ^+	0.069	0.749	32.754***	1242.954***	152340539.947***	−48.329***	0.986
RJ^-	0.053	0.545	35.904***	1388.521***	190092757.731***	−3.599**	0.983
CJ	0.122	0.922	24.753***	702.082***	48711323.379***	−13.925***	0.194
JLM	0.341	1.016	18.920***	592.745***	34822124.400***	−5.365***	448.349***
JLM^+	0.184	0.824	29.911***	1156.158***	131794182.126***	−7.561***	448.349***
JLM^-	0.157	0.332	3.815***	17.602***	36193.363***	−3.255	7018.844***

Notes: St. dev. is standard error. The Jarque-Bera statistic (Jarque and Bera, 1987) tests for the null hypothesis of normality for the distribution of the series. ADF is the Augmented Dickey–Fuller statistic (Cheung and Lai, 1995) based on the least AIC criterion. Q(n) is the Ljung-Box statistic proposed by Ljung and Box (1978) for up to n-th order serial autocorrelation. The asterisk ***, **denotes rejections of null hypothesis at 1% and 5% significance levels. The Mean and St. dev. series are multiply 1000, respectively.

Table 2
The estimated parameter results of HAR-RRV-type models with OLS and Newey-West correction.

	c	β_d	β_w	β_m	ω	α	β	γ	d	ν	λ	θ	ARCH test
HAR-RRV	−0.473***	0.390***	0.254***	0.302***									189.470***
HAR-RRV-GARCH	−0.339***	0.386***	0.356***	0.220***	0.065***	0.205**	0.409***			4.598***	0.287***		
HAR-RRV-GJR	−0.340***	0.409***	0.350***	0.203***	0.059***	0.249***	0.474***	−0.157***		4.608***	0.294***		
HAR-RRV-FIEGARCH	−0.309***	0.436***	0.320***	0.209***	−1.186***	0.843***	0.241***	0.356***	−0.011	4.498***	0.303***	0.138***	
h=5													
HAR-RRV	−0.796***	0.241***	0.245***	0.423***									455.07***
HAR-RRV-GARCH	−0.727***	0.154***	0.260***	0.504***	0.097***	0.518***	0.170*			6.378***	0.314***		
HAR-RRV-GJR	−0.614***	0.131***	0.249***	0.550***	0.091***	0.253**	0.263***	0.509***		5.383***	0.447***		
HAR-RRV-FIEGARCH	−0.761***	0.113***	0.409***	0.385***	−0.760***	0.179***	0.358***	0.623***	0.085	5.541***	0.258***	−0.131	
h=22													
HAR-RRV	−1.506***	0.197***	0.178***	0.446***									164.92***
HAR-RRV-GARCH	−1.739***	0.066***	0.221***	0.513***	0.106***	0.590***	0.163***			12.115***	0.388***		
HAR-RRV-GJR	−1.710***	0.057**	0.231***	0.512***	0.095***	0.401***	0.254***	0.281***		9.947***	0.484***		
HAR-RRV-FIEGARCH	−1.797***	0.055**	0.187***	0.545***	1.454***	−0.217	0.792***	0.678***	−0.012	8.684***	0.520***	−0.115**	ARCH test

Notes: Engle's LM ARCH test (Engle, 1982) to check the presence of ARCH effects in a series, which the test statistic is distributed $\chi_2(p)$ under the null hypothesis of no ARCH effects. The asterisk ***, ** and * denote rejections of null hypothesis at 1%, 5% and 10% significance levels, respectively.

then rolled forward by adding one new day and dropping the most distant day. In this way, the sample size used to estimate the models remains at a fixed length and the forecasts do not overlap.

To evaluate the significant difference of the abovementioned volatility models' predictability, we use the two popular loss functions as below:

$$\text{HMSE} = M^{-1} \sum_{m=1}^{M} (1 - \hat{\sigma}_m^2 / RRV_m)^2. \tag{24}$$

$$\text{HMAE} = M^{-1} \sum_{m=1}^{M} |1 - \hat{\sigma}_m^2 / RRV_m|. \tag{25}$$

where $\hat{\sigma}_m^2$ denotes the out-of-sample volatility forecast obtained by different volatility models, RRV_m is a proxy for actual market volatility in out-of-sample period, and M is the number of forecasting days. However, the abovementioned loss functions do not provide any information on whether the differences of forecasting losses among models are statistical significant. Given those reasons, we use an advanced model confidence set (MCS) test, proposed by Hansen et al. (2011), to choose a subset of models containing all possible superior models from the initial model set. The more details on the procedures of MCS test can be seen in Hansen et al. (2011).

Following the works of Martens et al. (2009), Hansen et al. (2011), Laurent et al. (2012), Liu et al. (2017), we choose the 75% as the confidence level of MCS test, which enable us to exclude a model from the 'best' model set with a p-value smaller than 0.25. The p-values are obtained by the 10,000 bootstrap and 2 block length.[5] Table 3 exhibits the out-of-sample forecasting results of the MCS test. Considering the one-day horizon, we observe from Table 3 that, only the linear HAR-RRV-CJ and LHAR-RRV-CJ models can survive in MCS test, because their p-values are larger than 0.25. Compared with other models, we find that the significant jump components based on Z test (Christensen and Podolskij, 2006) and signed returns can help in forecasting the short-term volatility. The HAR-RRV-CJ and LHAR-RRV-CJ models including the GARCH-type components have also better performance in forecasting. In our study, we provide a new insight to forecast the oil futures price realized range-based volatility. When forecasting the one-week horizon volatility, the p-values of the HAR-RRV-CJ-GARCH and LHAR-RRV-CJ-GARCH models are bigger than 0.25 under the two loss functions, which strongly imply that both of them have more powerful performance in forecasting. We can further deserve that the volatility

Table 3
Out-of-sample forecasting evaluation results using the MCS test (M=800).

	h=1		h=5		h=22	
Models	HMSE	HMAE	HMSE	HMAE	HMSE	HMAE
HAR-RRV	0.092	0.006	0.018	0.000	0.000	0.000
HAR-RRV-J	0.097	0.006	0.018	0.000	0.000	0.000
HAR-RRV-CJ	**1.000**	**0.971**	0.020	0.000	0.087	0.031
LHAR-RRV-CJ	**0.907**	**0.971**	0.020	0.000	0.087	0.059
HAR-RRV-RJ	0.097	0.006	0.018	0.000	0.000	0.000
HAR-RRV-LMJ	0.097	0.006	0.020	0.000	0.001	0.000
HAR-RRV-GARCH	0.091	0.006	0.009	0.000	0.087	0.045
HAR-RRV-GJR	0.097	0.006	0.007	0.000	0.055	0.015
HAR-RRV-FIEGARCH	0.097	0.026	0.004	0.000	0.114	0.231
HAR-RRV-J-GARCH	0.097	0.006	0.009	0.000	0.087	0.031
HAR-RRV-J-GJR	0.097	0.006	0.004	0.000	0.046	0.013
HAR-RRV-J-FIEGARCH	0.150	0.011	0.003	0.000	0.175	0.104
HAR-RRV-CJ-GARCH	**0.907**	**0.971**	**1.000**	**0.574**	**1.000**	**0.380**
HAR-RRV-CJ-GJR	**0.524**	**0.951**	0.043	0.000	**0.804**	**1.000**
HAR-RRV-CJ-FIEGARCH	**0.643**	**0.971**	**0.443**	0.000	**0.804**	**0.693**
LRRV-CJ-GARCH	**0.624**	**1.000**	**0.443**	**1.000**	**0.804**	**0.693**
LHAR-RRV-CJ-GJR	**0.424**	**0.971**	0.020	0.001	0.116	0.104
LHAR-RRV-CJ-FIEGARCH	0.000	0.000	0.043	0.000	**0.283**	**0.380**
HAR-RRV-ARJ-GARCH	0.097	0.132	0.018	0.000	0.085	0.013
HAR-RRV-ARJ-GJR	0.097	0.051	0.009	0.000	0.009	0.002
HAR-RRV-ARJ-FIEGARCH	0.097	0.006	0.003	0.000	0.087	0.059
HAR-RRV-JLM-GARCH	0.090	0.006	0.009	0.000	0.015	0.001
HAR-RRV-JLM-GJR	0.095	0.006	0.004	0.000	0.005	0.000
HAR-RRV-JLM-FIEGARCH	0.097	0.006	0.003	0.000	0.023	0.003

Notes: The p -values with bold and underline are larger 0.25, implying that the corresponding models can survive in MCS test and have better performance in forecasting.

models including the significant jump components, signed returns and volatility of volatility can gain higher forecasts accuracy in the oil futures market. Finally, we find that the HAR-RRV-CJ-GARCH, HAR-RRV-CJ-GJR, HAR-RRV-CJ-FIEGARCH, LHAR-RRV-CJ-GARCH and LHAR-RRV-CJ-FIEGARCH models can survive in MCS test, which show that those models exhibit superior forecasting ability in forecasting one-month horizon volatility. Therefore, we conclude that using the GARCH-type models to describe the volatility of realized range-based volatility and considering the impacts of jump components and signed returns can substantially help in forecasting the oil futures volatility. To the best of authors' knowledge, we are the first to investigate the impacts of volatility of volatility on forecasting the oil futures RRV and propose some new volatility models including the GARCH-type com-

[4] They can available upon request by authors.
[5] We choose different block lengths and can gain the same empirical results.

Table 4
Out-of-sample forecasting evaluation results using the MCS test (M=600).

	h=1		h=5		h=22	
Models	HMSE	HMAE	HMSE	HMAE	HMSE	HMAE
HAR-RRV	0.089	0.009	0.011	0.000	0.011	0.001
HAR-RRV-J	0.089	0.009	0.011	0.000	0.011	0.001
HAR-RRV-CJ	**0.294**	0.125	0.011	0.000	0.011	0.002
LHAR-RRV-CJ	**1.000**	**1.000**	0.103	0.103	0.035	0.014
HAR-RRV-RJ	0.089	0.009	0.011	0.000	0.011	0.001
HAR-RRV-LMJ	0.088	0.009	0.011	0.000	0.011	0.002
HAR-RRV-GARCH	0.089	0.009	0.011	0.000	0.011	0.002
HAR-RRV-GJR	0.089	0.009	0.011	0.000	0.011	0.002
HAR-RRV-FIEGARCH	0.089	0.009	0.011	0.000	0.011	0.002
HAR-RRV-J-GARCH	0.089	0.009	0.011	0.000	0.011	0.002
HAR-RRV-J-GJR	0.089	0.009	0.011	0.000	0.011	0.002
HAR-RRV-J-FIEGARCH	0.089	0.009	0.011	0.000	0.011	0.002
HAR-RRV-CJ-GARCH	0.100	0.030	0.103	0.028	0.053	0.019
HAR-RRV-CJ-GJR	0.089	0.066	0.011	0.000	0.013	0.002
HAR-RRV-CJ-FIEGARCH	0.089	0.052	0.011	0.000	0.017	0.002
LRRV-CJ-GARCH	**0.277**	0.205	**1.000**	**1.000**	**1.000**	**1.000**
LHAR-RRV-CJ-GJR	**0.524**	**0.802**	0.011	0.000	0.015	0.004
LHAR-RRV-CJ-FIEGARCH	**0.524**	**0.802**	0.011	0.000	0.053	0.066
HAR-RRV-ARJ-GARCH	0.087	0.009	0.011	0.000	0.011	0.002
HAR-RRV-ARJ-GJR	0.088	0.009	0.011	0.000	0.011	0.002
HAR-RRV-ARJ-FIEGARCH	0.087	0.009	0.011	0.000	0.011	0.002
HAR-RRV-JLM-GARCH	0.089	0.009	0.011	0.000	0.011	0.002
HAR-RRV-JLM-GJR	0.089	0.009	0.011	0.000	0.011	0.002
HAR-RRV-JLM-FIEGARCH	0.089	0.009	0.011	0.000	0.011	0.002

Notes: The p -values with bold and underline are larger 0.25, implying that the corresponding models can survive in MCS test and have better performance in forecasting.

ponents, and our empirical results suggest that our new models with significant jump components, signed returns and volatility of volatility (i.e., HAR-RRV-CJ-GARCH, LHAR-RRV-CJ-GARCH,) can gain higher forecasts accuracy at different forecast horizons.

4.3. Robustness check

As far as we are aware, choosing the appropriate forecasting window is crucial to the forecasting performance of the models. However, there is no consensus on how to choose the right forecasting windows in academia. Thus, we choose another alternative out-of-sample forecasting windows, 600, as our robustness check. Table 4 demonstrates the empirical results of different forecasting windows. We find that the p-value of LHAR-RRV-CJ and its extended models with GARCH, GJR-GARCH and FIEGARCH model can survive in the MCS test, implying that those models have better performance in forecasting one-day volatility. Regarding the one-week and one-month horizon, the LHAR-RRV-CJ-GARCH model exhibits absolutely advantage in forecasting oil futures volatility. From the results of Tables 3 and 4, we find that our new proposed models with significant jump components, signed return and volatility of volatility modeled by GARCH can provide more accurate volatility forecasts in the oil futures market.

5. Conclusions

In this study, we first explore the oil futures realized range-based volatility by considering the significant jump, signed returns and volatility of realized range-based volatility are modeled by the GARCH, GJR-GARCH and FIEGARCH models. This paper uses the rolling window to deserve the future volatility of different models and utilize the MCS test to evaluate those models' predictability. First, we find that the linear HAR-RRV-type models exhibit strong volatility clustering effect by ARCH test. Given those empirical results, we construct some new realized range-based volatility models including the GARCH-type components to investigate the oil futures volatility, which provide new insights to forecast oil futures volatility for the researchers, market participants, and policy makers. Second, out-of-sample results suggest that our new realized range-based volatility models with significant jump components, signed returns and volatility of volatility can gain higher forecasts accuracy than those of other models at one-day, one-week and one-month horizons. Third, our robustness results also support that considering the volatility of RRV can help in forecasting the oil futures price volatility. Our new findings are very important to risk management, derivative pricing, and portfolio selection.

Acknowledgement

The authors are grateful to the Co-Editor and the anonymous referees for providing thoughtful comments and suggestions that enhanced the quality of this paper. The authors are also grateful to the financial support from the Natural Science Foundation of China [71371157, 71401077, 71671145]; the Humanities and Social Science Fund of Ministry of Education [14YJC790073, 15YJA790031 and 16YJA790062]; the Young Scholar Fund of Science & Technology Department of Sichuan Province [2015JQO010]; and Fundamental Research Funds for the Central Universities [26816WCX02, 2682017WCX01].

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.econmod.2017.04.020.

References

Agnolucci, P., 2009. Volatility in crude oil futures: a comparison of the predictive ability of GARCH and implied volatility models. Energy Econ. 31 (2), 316–321.
Andersen, T.G., Bollerslev, T., 1998. Answering the skeptics: yes, standard volatility models do provide accurate forecasts. Int. Econ. Rev. 39 (4), 885–905.
Andersen, T.G., Bollerslev, T., Diebold, F.X., 2007a. Roughing it up: including jump components in the measurement, modeling, and forecasting of return volatility. Rev. Econ. Stat. 89 (4), 701–720.
Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P., 2003. Modeling and forecasting realized volatility. Econometrica 71 (2), 579–625.
Andersen, T.G., Bollerslev, T., Dobrev, D., 2007b. No-arbitrage semi-martingale restrictions for continuous-time volatility models subject to leverage effects, jumps and iid noise: theory and testable distributional implications. J. Econ. 138 (1), 125–180.
Baillie, R.T., Bollerslev, T., Mikkelsen, H.O., 1996. Fractionally integrated generalized autoregressive conditional heteroskedasticity. J. Econ. 74 (1), 3–30.
Bandi, F.M., Russell, J.R., 2008. Microstructure noise, realized variance, and optimal sampling. Rev. Econ. Stud. 75 (2), 339–369.
Bollerslev, T., Mikkelsen, H.O., 1996. Modeling and pricing long memory in stock market volatility. J. Econ. 73 (1), 151–184.
Caporin, M., Velo, G.G., 2015. Realized range volatility forecasting: dynamic features and predictive variables. Int. Rev. Econ. Financ. 40, 98–112.
Çelik, S., Ergin, H., 2014. Volatility forecasting using high frequency data: evidence from stock markets. Econ. Model. 36, 176–190.
Charles, A., Darné, O., 2014. Volatility persistence in crude oil markets. Energy Policy 65, 729–742.
Cheong, C.W., 2009. Modeling and forecasting crude oil markets using ARCH-type models. Energy Policy 37 (6), 2346–2355.
Christensen, K., Podolskij, M., 2006. Range-based estimation of quadratic variation. International Conference on High-Frequency Finance, Technical Report, No. 2006, 37.
Christensen, K., Podolskij, M., 2007. Realized range-based estimation of integrated variance. J. Econ. 141 (2), 323–349.
Christensen, K., Podolskij, M., 2012. Asymptotic theory of range-based multipower variation. J. Financ. Econ. 10 (3), 417–456.
Corsi, F., 2009. A simple approximate long-memory model of realized volatility. J. Financ. Econ. 7 (2), 174–196.
Corsi, F., Mittnik, S., Pigorsch, C., Pigorsch, U., 2008. The volatility of realized volatility. Econ. Rev. 27 (1-3), 46–78.
Corsi, F., Pirino, D., Reno, R., 2010. Threshold bipower variation and the impact of jumps on volatility forecasting. J. Econ. 159 (2), 276–288.
Efimova, O., Serletis, A., 2014. Energy markets volatility modelling using GARCH. Energy Econ. 43, 264–273.

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. J. Financ. 48 (5), 1779–1801.

Hansen, P.R., Lunde, A., 2005. A forecast comparison of volatility models: does anything beat a GARCH (1, 1)? J. Appl. Econ. 20 (7), 873–889.

Hansen, P.R., Lunde, A., Nason, J.M., 2011. The model confidence set. Econometrica 79 (2), 453–497.

Haugom, E., Langeland, H., Molnár, P., Westgaard, S., 2014. Forecasting volatility of the US oil market. J. Bank. Financ. 47, 1–14.

Huang, Z., Liu, H., Wang, T., 2016. Modeling long memory volatility using realized measures of volatility: a realized HAR GARCH model. Econ. Model. 52, 812–821.

Koopman, S.J., Jungbacker, B., Hol, E., 2005. Forecasting daily variability of the S & P 100 stock index using historical, realised and implied volatility measurements. J. Empir. Financ. 12 (3), 445–475.

Lean, H.H., Smyth, R., 2015. Testing for weak-form efficiency of crude palm oil spot and future markets: new evidence from a GARCH unit root test with multiple structural breaks. Appl. Econ. 47 (16), 1710–1721.

Liu, J., Wei, Y., Ma, F., Wahab, M., 2017. Forecasting the realized range-based volatility using dynamic model averaging approach. Econ. Model. 61, 12–26.

Ma, F., Liu, L., Liu, Z., Wei, Y., 2015. Forecasting realized range volatility: a regime-switching approach. Appl. Econ. Lett. 22 (17), 1361–1365.

Martin, M., Dick, V.D., 2007. Measuring volatility with the realized range. J. Econ. 138 (1), 181–207.

McAleer, M., Medeiros, M.C., 2008. Realized volatility: a review. Econ. Rev. 27 (1-3), 10–45.

Mohammadi, H., Su, L., 2010. International evidence on crude oil price dynamics: applications of ARIMA-GARCH models. Energy Econ. 32 (5), 1001–1008.

Nomikos, N., Andriosopoulos, K., 2012. Modelling energy spot prices: empirical evidence from NYMEX. Energy Econ. 34 (4), 1153–1169.

Sévi, B., 2014. Forecasting the volatility of crude oil futures using intraday data. Eur. J. Oper. Res. 235 (3), 643–659.

Sévi, B., 2015. Explaining the convenience yield in the WTI crude oil market using realized volatility and jumps. Econ. Model. 44, 243–251.

Todorova, N., Souček, M., 2014. The impact of trading volume, number of trades and overnight returns on forecasting the daily realized range. Econ. Model. 36, 332–340.

Tseng, T.-C., Chung, H., Huang, C.-S., 2009. Modeling jump and continuous components in the volatility of oil futures. Stud. Nonlinear Dyn. Econ. 13 (3), 1–30.

Tseng, T.-C., Lai, H.-C., Lin, C.-F., 2012. The impact of overnight returns on realized volatility. Appl. Financ. Econ. 22 (5), 357–364.

Wei, Y., Wang, Y., Huang, D., 2010. Forecasting crude oil market volatility: further evidence using GARCH-class models. Energy Econ. 32 (6), 1477–1484.