Energy Economics 72 (2018) 321–330

Contents lists available at ScienceDirect

Energy Economics

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

Forecasting the oil futures price volatility: Large jumps and small jumps

Jing Liu [a,1], Feng Ma [b,1], Ke Yang [c,*,1], Yaojie Zhang [b,1]

[a] Business School of Sichuan University, Chengdu, China
[b] School of Economics & Management, Southwest Jiaotong University, Chengdu, China
[c] School of Economics and Commerce, South China University of Technology, China

ARTICLE INFO

Article history:
Received 2 September 2017
Received in revised form 4 April 2018
Accepted 11 April 2018
Available online 18 April 2018

JEL classification:
C22
C32
C52
C53

Keywords:
Volatility forecasting
oil futures price
Large and small jumps
Predictive evaluation

ABSTRACT

Macro news drives jumps, however, a jump does not seem to improve the predictability of the simple heterogeneous autoregressive realized volatility model (HAR-RV) in the oil futures market. This paper provides a new insight and seeks to investigate whether truncated jumps can help improve the forecasting ability compared to that achieved using the HAR-RV model and its various extensions with jumps. Our results provide strong evidence that the models incorporating both large and small jumps gain a significantly superior forecasting ability. Specifically, including small jumps in a high-frequency model significantly improves the forecast accuracy at the 1-day forecasting horizon, while including both large and small jumps can achieve a higher forecast accuracy at the weekly and monthly horizons. These findings reveal that considering the decomposed jumps with a certain threshold can increase the forecast accuracy of the corresponding model.

1. Introduction

Oil futures price fluctuations are crucial for both the global economy and the financial markets. However, the oil futures market is affected by numerous economic and political factors, and various information flows of these factors cause jumps in the oil futures price. Hence, the jumps result in high volatility which is an essential input in option pricing and value at risk (Naik and Lee, 1990; Liu et al., 2003). Accurately modelling and forecasting the oil futures price volatility by using jumps has been a crucial issue for market participants and policy makers.

In the earliest studies, the GARCH-class models using low-frequency data are widely used to forecast oil volatility. Interestingly, high-frequency data are more information-rich than are low-frequency data such as daily data and monthly data. In particular, the HAR-RV model proposed by Corsi (2009), is a representative and popular model based on the high-frequency data, that have developed rapidly in recent years. The HAR-RV and its variants (HAR-RV-type) are popularly applied and the general consensus is that HAR-RV-type models are more suitable for forecasting the volatility of high-frequency data than GARCH-class models are (e.g., Andersen et al., 2003; Hansen and Lunde, 2005; Wei et al., 2010; Çelik and Ergin, 2014; Ma et al., 2015).

The jumps that occur are associated with specific macroeconomic news announcements (Andersen et al., 2007), and modelling and forecasting volatility with jumps form an important branch of research. Based on Corsi (2009), several studies consider the jump components and add them as additional variables in volatility models, such as the HAR-RV-J, HAR-RV-CJ (Andersen et al., 2007), and HAR-RV-TJ models (Corsi et al., 2010). Overall, these models with jumps are constructed by using significant jump tests, such as the Z statistic test proposed by Huang and Tauchen (2005) and the C-Tz test proposed by Corsi et al. (2010). However, their forecasting performance remains questioned. Specifically, some studies claim that the jumps cannot produce better forecasts; thus, the inclusion of jump components does not help improve the forecast accuracy of the simple HAR-RV model in the crude oil market (Haugom et al., 2014; Sévi, 2014; Prokopczuk et al., 2016).

Although the aforementioned literature has considered the role of jump components in forecasting volatility, most studies concentrate on the stock and exchange market. To date, fewer than ten studies have concentrated on forecasting oil volatility by using the intraday data (e.g., Haugom et al., 2014; Sévi, 2014; Prokopczuk et al., 2016; Wen et al., 2016; Liu et al., 2017; Ma et al., 2017a; Degiannakis and Filis, 2017), and most simply consider the impacts of jumps on oil volatility.

In particular, using the S&P500 futures index and Dow 30 individual stock datasets, Duong and Swanson (2015) present evidence on the predictive content of truncated jump variables by using a threshold jump size. Their results show that past large jumps help less in

* Corresponding author.
E-mail addresses: jingliu666@gmail.com, (J. Liu), mafeng2016@swjtu.edu.cn, (F. Ma), yangkdc@scut.edu.cn, (K. Yang), yaojie_zhang@126.com. (Y. Zhang).
[1] All the authors contribute to the paper equally.

https://doi.org/10.1016/j.eneco.2018.04.023
0140-9883/

predicting future stock volatility than small jumps do. We conjecture that the magnitudes of large jumps are dependent on the truncation levels and decomposing jumps with certain thresholds filter out more "noise". Moreover, small jumps caused by macro news or information in the oil futures market are reflected so quickly in the market that they may be helpful for forecasting short-term volatility. In contrast, large jumps caused by unexpected huge shock huge shocks require a longer time to reflect in the oil market and could thus be useful for forecasting long-term volatility.

To the best of our knowledge, no studies have investigated the oil futures volatility by further decomposing the jumps of the HAR-RV-type models based on high-frequency data. Inspired by Duong and Swanson (2015), simultaneously considering that jumps are both highly prevalent and distinctly less persistent, this study reconsiders the role of jumps and provides an easy-to-implement way to improve the forecast accuracy of oil futures volatility by decomposing jump components. We further consider the decomposed jumps and propose new specifications with "significant" small jumps and large ones in oil futures prices. In addition, we conduct an analysis at three horizons thus following the very detailed analysis which we are interested in.

Therefore, this study adds to the scarce literature on oil futures' realized volatility forecasting using HAR-RV-type models and contributes to the field in three dimensions. First, in the framework of decomposed jumps, several popular jump detections are employed and compared including the Lagrange multiplier (LM) test (Lee and Mykland, 2008), the Z statistic test (Huang and Tauchen, 2005), and the C-Tz statistics (Corsi et al., 2010). Second, we enrich HAR-RV models with jump components by modelling with decomposed jumps. Based on the HAR-RV-JLM, three extensions are constructed by adding large jumps, small jumps and both. Similarly, we extend other selected volatility models by incorporating large jumps, small jumps, and both. In total, twelve new extensions are constructed based on the five existing models. Third, we assess the importance of decomposed jumps and jumps itself in forecasting oil futures volatility by using the advanced model confidence set (MCS) test (Hansen et al., 2011).

Our empirical results demonstrate that our proposed the HAR-RV-TJ model with small jumps performs the best for short-term forecasting, while the HAR-RV-JLM model with both large and small jumps is the most powerful for mid- and long-term forecasting. In particular, small jumps contribute greatly at 1-day-ahead forecasting, while both small and large jumps are recommended at 5- and 22-day-ahead forecasting. We check the robustness of our results using the MCS test with different estimation window sizes and using an alternative actual volatility-realized kernel (RK). Compared to volatility forecasting models with jumps alone, decomposed jumps yield a substantial improvement in forecasting oil futures volatility because decomposed jumps help filter more effective information. In brief, this study provides a further perspective of jumps in analyzing the oil futures market which is affected by an enormous number of uncertain factors.

The rest of this paper is structured as follows. Section 2 reviews the related literature. Section 3 presents the realized volatility, the jump tests and the econometric models. Section 4 describes the data. Section 5 analyzes the empirical results and Section 6 concludes this study.

2. Review of the literature

Financial asset volatility is a very important quantity for option pricing and value-at-risk (VaR) calculations in financial risk management. Due to the complexity of the oil futures market, improving the forecast accuracy of oil price volatility is an extremely empirical challenge. In this section, we revisit the techniques in modeling and forecasting oil price volatility.

In general, the earliest studies use the squired returns as a proxy of volatility and most of them are extended basing on the GARCH model (Bollerslev, 1986). GARCH-class models are commonly used for forecasting oil volatility (e.g., Lautier and Riva, 2004; Agnolucci, 2009; Kang et al., 2009; Wei et al., 2010; Nomikos and Andriosopoulos, 2012; Efimova and Serletis, 2014). However, these models are based on low-frequency data and perform poorly at intraday frequencies. In recent years, volatility models using intraday data have received great attention from researchers and market participants. Andersen and Bollerslev (1998) propose the realized volatility (RV), which is defined as the sum of all intraday squared returns to measure the unobserved volatility. Using the RV as a proxy, Corsi (2009) propose a simple heterogeneous autoregressive RV (henceforth HAR-RV) model, which is the classic one and is powerful for capturing "stylized facts" of financial asset volatility, such as long memory and multi-scaling behavior (Andersen et al., 2007; Corsi et al., 2010; Bekaert and Hoerova, 2014; Duong and Swanson, 2015; Bollerslev et al., 2016).

In particular, some studies recognize the role of jumps in financial economics. Modelling and forecasting the realized volatility using the jump comments are an important branch of such work (Andersen et al., 2007; Sévi, 2014; Prokopczuk et al., 2016). Andersen et al. (2007) propose the HAR-RV-J model, in which jumps are added. This model is measured by the bi-power variation. Moreover, to investigate the contribution of jumps, Andersen et al. (2007) first decompose the RV into continuous sample path and significant jump components, and then add the treatment of significant jumps to the simple HAR-RV model, namely the HAR-RV-CJ model. In addition, Corsi et al. (2010) combine the C-Tz statistics with the HAR-RV model to build the HAR-RV-TJ model. They find that the HAR-RV-TJ model significantly outperforms both the HAR model and the HAR-CJ model after a jump.

However, the contribution of jumps in forecasting remains controversial in terms of different markets and applied models. In an application to exchange rates, equity index returns, and bond yields, Andersen et al. (2007) find that nearly all the volatility predictability at the daily, weekly, and monthly horizons originates from the non-jump component. In contrast, Corsi et al. (2010) conduct an empirical analysis of the S&P500 index, individual stocks and US bond yields, and their results suggest both that jumps are highly important and that separating a rough jump from smooth continuous moves can lead to significant forecast improvements. Corsi et al. (2010) conclude that jumps have a positive and mostly significant impact on future volatility.

Regarding the investigation of jumps on forecasting oil volatility, Sévi (2014) forecasts the oil futures realized volatility by applying nine extensions of the HAR model including five models with jumps such as the HAR-RV-J, HAR-RV-CJ, and HAR-RV-SJ model. The results demonstrate that the simple HAR-RV model often performs significantly better than the more complicated models do in terms of considered jump components, leverage effects and asymmetries. In line with the conclusion of Sévi (2014), Prokopczuk et al. (2016) examine whether jumps matter for volatility forecasting in four prominent energy markets. They conclude that explicitly modeling jumps cannot significantly improve the forecast accuracy. However, Wen et al. (2016) conclude that the historical realized volatility, signed jump, signed and semi-jump all contain substantial and salient information for forecasting crude oil futures volatility. Moreover, crude oil futures volatility is substantially affected by both short- and long-term information.

In particular, Elder et al. (2013) note that the largest jumps tend to be preceded by identifiable economic news and that the arrival of new economic information has a strong correspondence to high-frequency jumps in oil prices volatility. Duong and Swanson (2015) set a threshold jump and obtain evidence for the predictability of truncated jumps. They claim that past large jumps help less in predicting future stock volatility than small jumps do. Furthermore, Duong and Swanson (2015) provide a new exploration of modelling with jumps. Due to the specialization of the oil market and its distinct reaction to news information, whether setting a threshold for jumps is effective in forecasting oil volatility remains questionable and is worth further investigation.

In summary, these aforementioned techniques provide a rich framework for forecasting asset volatility. In contrast to numerous studies on forecasting stock and bond markets, there are insufficient studies on forecasting the oil futures volatility and there is no literature investigating the role of decomposed jumps. Using high-frequency data, this paper extends the previous contributions of oil volatility forecasting by considering decomposed jump components and implementing new volatility forecasting models. We mainly use the LM test to detect jumps and decompose the jumps into large and small jumps. Additionally, the Z statistic test and C-Tz statistics are employed for comparison.

3. Realized volatility, jump tests and econometric models

In this section, we introduce the popular jump tests, such as the Z, C_Tz and LM tests. Moreover, we provide the formulations of the HAR-RV model and its various extensions.

3.1. Realized variance

We first divide the time interval [0, 1] into n subintervals and determine the sampling frequency, $\Delta = 1/M$. Realized variance (volatility)[2] is defined as the sum of all available intraday high-frequency squared returns:

$$RV_t = \sum_{i=1}^{M} r_{t,i}^2, \tag{1}$$

where r is the intraday return. When $\triangle \to 0$, RV can be satisfied:

$$RV_t \to \int_0^t \sigma^2(s)ds + \sum_{0<s\le t} k^2(s), \tag{2}$$

where $\int_0^t \sigma^2(s)ds$ is the integrated variance. This is a continuous component and can be computed using the realized bi-power variation (BPV), defined as follows:

$$BPV_t = \mu_1^{-2} \sum_{i=2}^{M} |r_{t,i}||r_{t,(i-1)}|, \tag{3}$$

where $\mu_1 = (2/\pi)^{0.5} \approx 0.7979$. $\sum_{0<s\le t} k^2(s)$ is the discontinuous portion of the quadratic variation and is called a jump component.

3.2. Jump tests and jump decompositions

Since the performance of volatility models is greatly affected by different jump detection methods, the main methods are briefly listed in this section.

3.2.1. Daily tests

Based on the Eq. (2), we have the equation as below:

$$\mathrm{RV}_t - \mathrm{BPV}_t \to \sum_{0<s\le t} \kappa^2(s), \tag{4}$$

To ensure that all of the daily jump variation estimations are non-negative, following Barndorff-Nielsen and Shephard (2004), the jump component is defined as:

$$\mathrm{J}_t = \max(\mathrm{RV}_t - \mathrm{BPV}_t, 0). \tag{5}$$

To determine the significant jump components, we use the C_Tz proposed by Corsi et al. (2010). There are two main reasons. First, the C_Tz test reduces the finite sample bias in estimating the integral of the second and fourth powers of continuous volatility in the presence of jumps. Second, Corsi et al. (2010) use simulation data to evaluate the performance of the two statistics, and find that in the presence of jumps the C_Tz statistic has substantially more power than the Z statistic does proposed by Huang and Tauchen (2005), particularly when jumps are consecutive, which occurs very frequently for high-frequency data (Sévi, 2014; Pu et al., 2016; Ma et al., 2017c). The C_Tz statistic is defined as:

$$\mathrm{C\_Tz}_t = \Delta^{-\frac{1}{2}} \frac{(\mathrm{RV}_t - \mathrm{C\_TBPV}_t)\cdot \mathrm{RV}_t^{-1}}{\sqrt{\left(\frac{\pi^2}{4}+\pi-5\right)\max\left\{1, \frac{\mathrm{C\_TTriPV}_t}{(\mathrm{C\_TTriPV})^2}\right\}}}, \tag{6}$$

The corrected realized threshold multi-power (C_TMPV) is defined as:

$$\mathrm{C{-}TMPV}_t{}^{[\gamma_1,\dots,\gamma_M]} = \Delta^{1-\frac{1}{2}(\gamma_1,\dots,\gamma_M)} \sum_{j=M}^{1/\Delta} \prod_{k=1}^{M} z_{\gamma_k}\left(r_{j-k+1}, \vartheta_{j-k+1}\right). \tag{7}$$

where $\gamma_1, \dots, \gamma_M > 0$, ϑ_t is the stochastic threshold, and $\vartheta_t = c_\vartheta^2 \hat{V}_t$, $\hat{V}_t$ is an auxiliary estimator of latent volatility, in which c_ϑ is a scale-free constant and is equal to 3 in this paper. Thus, $\mathrm{C\_TBPV}_t$ and $\mathrm{C\_TTriPV}_t$ from Eq. (6) are equal to $\mathrm{u}_1^{-2}\mathrm{C} - \mathrm{TMPV}_t^{[1,1]}$ and $\mathrm{u}_1^{-3}\mathrm{C\_TMPV}_t^{[4/3,4/3,4/3]}$, respectively. Therefore, we have the "significant" jump component from the C_Tz test in excess of a critical value, Φ_α, which is the standard normal cumulative distribution function,

$$\mathrm{TJ}_t = \mathrm{I}(\mathrm{C\_TZ}_t > \Phi_\alpha)\cdot[\mathrm{RV}_t - \mathrm{C\_TBPV}_t]. \tag{8}$$

where $\mathrm{I}(\cdot)$ is an indicator function.

3.2.2. Intraday tests

In this study, we introduce one of the popular intraday jump tests, the LM test (Lee and Mykland, 2008), which is constructed under the null hypothesis that at a certain time there is no jump in the realization process. Dumitru and Urga (2012) find that this test has the overall best performance when using the Monte Carlo and real high-frequency data for five stocks listed on the New York Stock Exchange. The LM test has received much attention from scholars (Clements and Liao, 2017; Gilder et al., 2014; Liu et al., 2016b). Ma et al. (2017b) and Liu et al. (2016b) confirm that the jump components calculated by the LM test are helpful in forecasting. Specifically, Liu et al. (2016a) find that the LM test is helpful in forecasting real oil price volatility.

The LM test computes a local volatility estimate that is robust to jumps at time t_i, $\hat{\sigma}_{ti}^2$, and then standardizes the intraday returns as follows:

$$\varsigma_{t,i} = \frac{r_{t,i} - \hat{m}_i}{\hat{\sigma}_{t,i}}, \qquad \hat{m}_i = \frac{1}{K-1}\sum_{j=i-K+1}^{i-1} r_{t,j} \tag{9}$$

where $\hat{\sigma}_{t,i} = \frac{1}{K-2}\sum_{j=i-K+2}^{i-1} |r_{t,j}||r_{t,j-1}|$ and $K = \sqrt{M \times 252}$. Lee and Mykland (2008) show that the distribution of $\varsigma_{t,\,i}$ is asymptotically standard normal in the absence of jumps. A jump is detected if

$$|\varsigma_{t,i}| > \upsilon = \frac{\xi}{\mathrm{c}\sqrt{2\,\ln\mathrm{M}}} + \frac{\sqrt{2\,\ln\mathrm{M}}}{\mathrm{c}} - \frac{\ln 4\pi + \ln(\ln(\mathrm{M}))}{2\mathrm{c}\sqrt{2\,\ln\mathrm{M}}} \tag{10}$$

where $\mathrm{c} = \sqrt{2/\pi}$ and $\xi = -\ln(-\ln(1-\alpha))$. When a jump occurs, its size dominates r_i; i.e., when $|\varsigma_{t,\,i}|$ is larger than the critical value, a significant jump is established at time i with size r_i. Thus, we can calculate the significant jump component on day t, $JLM_t = \sum r_i^2$, which is defined as the sum of intraday squared returns when jumps occur.

3.2.3. Jump decompositions

Generally, small shocks of macro news or information are reflected quickly in the financial market and cause small jumps in asset price fluctuations, while huge shocks cause large jumps, and the markets take much longer to digest the news. After jump detection, it is

[2] In our study, the realized volatility and realized variance are interchangeable.

important to choose the truncation level, γ, because the truncation level directly determines the magnitudes of large and small jumps. Duong and Swanson (2015) note that arbitrarily choosing large truncation levels will result in finding no evidence of large jumps. An appropriate truncation level is expected to filter out more "noise" and make the decomposed jumps more predictive.

Fortunately, Duong and Swanson (2011) and Duong and Swanson (2015) provide a threshold method to choose the appropriate value of γ, which can divide the jump component into large and small jumps. In accordance with Duong and Swanson (2015), the threshold values can be acquired by the monthly maximum absolute increments of the oil future price intraday returns with a certain procedure. We set $I = [\gamma_1, \gamma_{50}]$ and set the number of grid points to 50, where γ_1 and γ_{50} correspond to the median and 95th percentiles of the monthly maximum absolute increments of oil future price returns. Thus, the threshold values, γ, can be calculated by $\gamma = \{\gamma_1, (\gamma_{50} - \gamma_1)/50 * l\}_{l=1}^{50}$.

We can define the significant large and small jump components as follows, respectively:

$$\mathrm{TJL}_{t,\gamma} = \min\left(\mathrm{TJ}_t, \left(\sum_{i=1}^{M} r_{t,i}^2 I(|r_{t,i}|>\gamma) * I(|\mathrm{C\_Tz}|>\Phi_\alpha)\right)\right) \tag{11}$$

$$\mathrm{TJS}_{t,\gamma} = \mathrm{TJ}_t - \mathrm{TJL}_{t,\gamma} \tag{12}$$

Inspired by Duong and Swanson (2011) and Duong and Swanson (2015), we first divide the jump components based on the LM intraday test into large and small jump components, which is defined as:

$$\mathrm{JLML}_{t,\gamma} = \min\left(\mathrm{JLM}_t, \left(\sum_{i=1}^{M} r_{t,i}^2 I(|\varsigma_{t,i}|>\upsilon) I(|r_{t,i}|>\gamma)\right)\right) \tag{13}$$

$$\mathrm{JLMS}_{t,\gamma} = \mathrm{JLM}_t - \mathrm{JLML}_{t,\gamma} \tag{14}$$

In general, there is little literature on forecasting the oil futures price volatility by considering the effects of the jump components, particularly on further decomposing the jumps into large and small components. Thus, this study investigates whether those decomposed jumps can improve the predict ability of the volatility models and help in forecasting the oil futures volatility.

3.3. Prediction models

In recent years, a simple heterogeneous autoregressive RV model proposed by Corsi (2009) has received attention from many scholars (e.g., Andersen et al., 2007; Chen and Ghysels, 2011; Sévi, 2014; Patton and Sheppard, 2015; Prokopczuk et al., 2016); this is termed the HAR-RV model. This model easily estimates and includes three components, the lagged daily, weekly and monthly realized volatility, and its formulation is as follows:

$$RV_{t+h} = c + \beta_d RV_t + \beta_w RV_{t-4,t} + \beta_m RV_{t-21,t} + \omega_{t+h}, \tag{15}$$

where $RV_{t-4,\ t}$ is the average RV from day t-4 to day t and $RV_{t-21,\ t}$ is the average RV from day t-21 to day t. The subscript h is equal to 1 (5 and 22), which represents the forecasting short (medium and long) horizon realized volatility.

The HAR-RV-J model can be obtained by adding the jump component, J_t, to the simple HAR-RV model and its specification can be written as:

$$RV_{t+h} = c + \beta_d RV_t + \beta_w RVW_t + \beta_m RVM_t + \beta_j J_t + \varepsilon_{t+h}, \tag{16}$$

Following Andersen et al. (2007), the HAR-RRV-CJ model can be extended from the HAR-RV model by incorporating continuous sample path and jump components. The HAR-RV-CJ model is expressed as

$$RV_{t+h} = c + \beta_d CRV_{t,\alpha} + \beta_w CRVW_{t,\alpha} + \beta_m CRVM_{t,\alpha} + \beta_j CJ_{t,\alpha} + \beta_{jw} CJW_{t,\alpha} + \beta_{jm} CJM_{t,\alpha} + \varepsilon_{t+h}, \tag{17}$$

where $CRV_{t,\ \alpha}$, $CRVW_{t,\ \alpha}$ and $CRVM_{t,\ \alpha}$ are the daily, weekly, and monthly averages of the continuous sample path, respectively, and similarly $CJ_{t,\ \alpha}$, $CJW_{t,\ \alpha}$, and $CJM_{t,\ \alpha}$, refer to daily, weekly, and monthly jumps.

Following the work of Andersen et al. (2007), Corsi et al. (2010) divide the quadratic variation into continuous and discontinuous components by the C_Tz test and construct a new model, termed HAR-RV-TJ,

$$RV_{t+h} = c + \beta_d TCRV_{t,\alpha} + \beta_w TCRVW_{t,\alpha} + \beta_m TCRVM_{t,\alpha} + \beta_j TJ_{t,\alpha} + \varepsilon_{t+h}, \tag{18}$$

where the $TCRV_{t,\ \alpha}$ is the continuous sample path component and is defined as, $TCRV_{t,\ \alpha} = I(C\_Tz_t \le \Phi_\alpha) \cdot RV_t + I(C\_Tz_t > \Phi_\alpha) \cdot C\_TBPV_t$, $TCRVW_{t,\ \alpha}$ and $TCRVM_{t,\ \alpha}$ are the weekly and monthly continuous components. In our empirical study, α equals 99%.

Liu et al. (2016b) find that the jump components calculated by LM test (henceforth, JLM) can help in forecasting. We therefore use the JLM as additional variable in the HAR-RV model and propose a new model for forecasting the oil futures volatility, called HAR-RV-JLM, which is expressed as follows:

$$RV_{t+h} = c + \beta_d RV_t + \beta_w RVW_t + \beta_m RVM_t + \beta_j JLM_t + \varepsilon_{t+h}, \tag{19}$$

Based on the appropriate value of γ, as calculated in Section 3.2.3, we divide the jump components into large and small jumps. In the following, we present our new proposed models and their jump compositions and compare them with the existing HAR-RV-type models considering the out-of-sample forecasts.

The HAR-RV-TJL model. We use only the significant large jump components, $TJL_{t,\ \alpha}$, to replace the jumps of the HAR-RV-TJ, implying that a new specification can be defined as:

$$RV_{t+h} = c + \beta_d TCRV_{t,\alpha} + \beta_w TCRVW_{t,\alpha} + \beta_m TCRVM_{t,\alpha} + \beta_{jl} TJL_{t,\alpha} + \varepsilon_{t+h}, \tag{20}$$

This model only contains the large jumps, so their impact on forecasting future volatility can be directly examined. In addition, Duong and Swanson (2015) find that large jumps are often considered abnormal events that arise at a frequency of one in several months or even years. Thus, we utilize the HAR-RV-TJL model to explore whether the large jumps can significantly help volatility forecasting.

The HAR-RV-TJS model. Similar to the HAR-RV-TJL model, we replace the jump components of the HAR-RV-TJ model with small jumps $TJL_{t,\ \alpha}$, and construct a new model,

$$RV_{t+h} = c + \beta_d TCRV_{t,\alpha} + \beta_w TCRVW_{t,\alpha} + \beta_m TCRVM_{t,\alpha} + \beta_{js} TJS_{t,\alpha} + \varepsilon_{t+h}, \tag{21}$$

From the specification of the HAR-RV-TJS, it is clear that we investigate only the effects of those small jumps. We can easily evaluate the significant differences between this model and the benchmark model.

The HAR-RV-TJLS model. We decompose the significant jumps into large and small jump components using the threshold method and replace the jump components with both large and small jumps ($TJLS_{t,\ \alpha}$). The formulation of the new model is given as below:

$$RV_{t+h} = c + \beta_d TCRV_{t,\alpha} + \beta_w TCRVW_{t,\alpha} + \beta_m TCRVM_{t,\alpha} + \beta_{js} TJLS_{t,\alpha} + \varepsilon_{t+h}, \tag{22}$$

Table 1
Descriptive statistics of all variables.

	Mean	St. dev.	Skewness	Kurtosis (excess)	Jarque-Bera	ADF	Q(5)
RV	0.298	0.452	4.999***	36.407***	129,538.221***	−4.043**	5200.389***
TCRV	0.277	0.417	4.711***	30.589***	92,712.308***	−3.647**	5378.039***
TJ	0.021	0.102	13.827***	290.495***	7,738,205.971***	−10.859***	51.179***
TJL	0.008	0.091	18.716***	456.661***	19,078,352.179***	−16.453***	36.904***
TJS	0.014	0.043	4.681***	26.347***	71,046.791***	−12.740***	11.938**
LM	0.118	0.355	6.949***	69.734***	459,461.571***	−5.927**	2646.860***
LML	0.004	0.092	27.194***	875.439***	69,914,752.480**	−46.781***	3.511
LMS	0.113	0.323	5.938***	50.066***	240,604.483	−6.215**	2988.628***

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 the 1% and 5% significance levels. The Mean and St. dev. series are multiply 1000, respectively.

The HAR-RV-TJLS model that considers the impacts of large and small jumps on future volatility differs from both Model 1 and Model 2.

The HAR-RV-JLML model. Similar to Model 1, we use the large jump components tested by the LM test to replace the jump components ($JLML_{t,\alpha}$), and propose a new model,

$$RV_{t+h} = c + \beta_d RV_t + \beta_w RVW_t + \beta_m RVM_t + \beta_{jl} JLML_t + \varepsilon_{t+h}, \tag{23}$$

In particular, this model is the first to explore the impacts of large intraday jumps on predicting oil futures price volatility.

The HAR-RV-JLMS model. In this model, we replace all intraday jump components with small intraday components, and construct a new volatility model,

$$RV_{t+h} = c + \beta_d RV_t + \beta_w RVW_t + \beta_m RVM_t + \beta_{js} JLMS_t + \varepsilon_{t+h}, \tag{24}$$

It is clear that we only care about the impact of small jump components on the future volatility.

The HAR-RV-JLMLS model. In this model, we naturally use the decomposition of large and small jump components ($JLMLS_t$) and combine these with the HAR-RV model to build a new model,

$$RV_{t+h} = c + \beta_d RV_t + \beta_w RVW_t + \beta_m RVM_t + \beta_{jls} JLMLS_t + \varepsilon_{t+h}, \tag{25}$$

The HAR-RV-JLMLS model includes both large and small jump components so we can investigate the impacts of the two jump components in the HAR-RV framework. Similarly, the HAR-RV-J and HAR-RV-CJ models have three extensions that are derived by adding large, small, and both jump components. Based on Eqs. (16) and (17), these extensions can be expressed as below.

The HAR-RV-JL model:

$$RV_{t+h} = c + \beta_d RV_{t,\alpha} + \beta_w RVW_{t,\alpha} + \beta_m RVM_{t,\alpha} + \beta_{jl} JL_{t,\alpha} + \varepsilon_{t+h}, \tag{26}$$

The HAR-RV-JS model:

$$RV_{t+h} = c + \beta_d RV_{t,\alpha} + \beta_w RVW_{t,\alpha} + \beta_m RVM_{t,\alpha} + \beta_{js} JS_{t,\alpha} + \varepsilon_{t+h}, \tag{27}$$

The HAR-RV-CJL model:

$$\begin{aligned} RV_{t+h} = c &+ \beta_d CRV_{t,\alpha} + \beta_w CRVW_{t,\alpha} + \beta_m CRVM_{t,\alpha} + \beta_{jl} CJL_{t,\alpha} \\ &+ \beta_{jwl} CJWL_{t,\alpha} + \beta_{jml} CJML_{t,\alpha} + \varepsilon_{t+h}, \end{aligned} \tag{28}$$

The HAR-RV-CJS model:

$$\begin{aligned} RV_{t+h} = c &+ \beta_d CRV_{t,\alpha} + \beta_w CRVW_{t,\alpha} + \beta_m CRVM_{t,\alpha} + \beta_{js} CJS_{t,\alpha} \\ &+ \beta_{jws} CJWS_{t,\alpha} + \beta_{jms} CJMS_{t,\alpha} + \varepsilon_{t+h}, \end{aligned} \tag{29}$$

The HAR-RV-CJLS model:

$$\begin{aligned} RV_{t+h} = c &+ \beta_d CRV_{t,\alpha} + \beta_w CRVW_{t,\alpha} + \beta_m CRVM_{t,\alpha} + \beta_{jls} CJLS_{t,\alpha} \\ &+ \beta_{jwls} CJWLS_{t,\alpha} + \beta_{jmls} CJMLS_{t,\alpha} + \varepsilon_{t+h}. \end{aligned} \tag{30}$$

Therefore, in addition to the simple HAR-RV model and four original models with jumps, we extend them to create twelve new models, for a total of seventeen models. In this analysis, we can gain new insights to demonstrate the asymmetric effect of the large and small jump components and evaluate whether those decompositions of jumps can increase the forecast accuracy.

4. Data description

Following Haugom et al. (2014), Sévi (2014), Liu et al. (2015) and Ma et al. (2017c), we use the 5-min high-frequency data from one front month of the Light Sweet Crude Oil (WTI) oil futures to investigate the forecasting performance of our new proposed models with decomposed jumps. Compared with other month contracts, the front-month contracts are the most active. Each contract expires on the third business day before the 25th calendar day of the month preceding the delivery month. The contract unit is for 1000 barrels and the price is quoted in U.S. dollars. Moreover, 5-min high-frequency data have been found to be the optimal choice of sampling frequency because they have the best trade-off between measurement accuracy and market microstructure noise (Andersen and Bollerslev, 1998; Andersen et al., 2007; Corsi et al., 2010; Sévi, 2014; Liu et al., 2015). The data comes from the Thomson Reuters Tick History Database and cover the period[3] from January 3, 2007 to September 18, 2015. After removing days with a shortened trading session or too few transactions, we obtain 2181 observations. Our entire data sample of the continuous oil futures series is divided into two subgroups: a) in-sample data that are used to estimate the models and cover the first 1400 trading days and b) out-of-sample data that are used for model evaluation and cover the remaining 781 trading days.

Table 1 reports the descriptions of realized volatility, continuous sample path, jumps and its decompositions, such as large and small jump components. Under the null hypothesis of “skewness = 0” and “kurtosis = 3”, we find that all the variables in Table 1 are skewed to the right and exhibit high kurtosis at the 99% confidence level. Combined with the results of the Jarque-Bera test, we can further confirm that all series are non-Gaussian distributions. The ADF tests reject the null hypothesis of a unit root at the 1% significance level, which implies that all of the variables are stationary. The Ljung–Box tests show

[3] Studies on volatility forecasting differ from data period and data length. Our sample data covers from January 3, 2007 to September 18, 2015. Because this period is representative, since it contains the U.S. subprime mortgage crisis and the European debt crisis, as well as a cycle of bull market and bear market for oil. Moreover, the conclusions are robust with different data period.

Table 2
The estimation results of the HAR-RV-TJ and its extended models based on the Newey-West correction ($h = 1$).

	HAR-RV -TJ	HAR-RV -TJL	HAR-RV -TJS	HAR-RV -TJLS	HAR-RV -JLM	HAR-RV -JLML	HAR-RV -JLMS	HAR-RV -JLMLS
c	0.018	0.018	0.001	0.001	0.011	0.013	0.026	0.012
	(0.014)	(0.014)	(0.015)	(0.011)	(0.017)	(0.015)	(0.018)	(0.016)
β_d	0.167*	0.169*	0.182*	0.183*	0.206	0.200*	0.073	0.214
	(0.104)	(0.103)	(0.102)	(0.102)	(0.146)	(0.103)	(0.158)	(0.140)
β_w	0.437**	0.447***	0.425***	0.427***	0.393**	0.401***	0.421***	0.398***
	(0.163)	(0.164)	(0.162)	(0.165)	(0.157)	(0.150)	(0.156)	(0.153)
β_m	0.420***	0.416***	0.419***	0.417***	0.402***	0.375***	0.410***	0.374***
	(0.140)	(0.141)	(0.141)	(0.143)	(0.127)	(0.123)	(0.126)	(0.121)
β_j	0.082				−0.070			
	(0.123)				(0.123)			
β_{jl}		−0.026		−0.029		−0.363**		−0.373**
		(0.112)		(0.111)		(0.123)		(0.113)
β_{js}			0.650**	0.651**			0.090	−0.153
			(0.320)	(0.320)			(0.178)	(0.153)
Adj. R^2	0.647	0.634	0.638	0.637	0.646	0.638	0.634	0.638

Notes: All models are estimated by OLS method with Newey-West correction. The numbers of parenthesis represent the Newey-West standard error of each parameter. The coefficients of constant and Newey-West standard error are multiply 1000, respectively. The asterisk ***, ** and * denotes rejections of null hypothesis at the 1%, 5% and 10% significance levels. *Adj.* R^2 is the adjust R-squared.

that nearly all variables are serial autocorrelations to a maximum of the 5th order in the crude oil futures market.

5. Empirical results

5.1. In-sample estimation results

Tables 2, 3 and 4 exhibit the estimation results[4] of the HAR-RV-TJ and other extended models that were obtained using the ordinary least squares method with the Newey-West correction, which allows for correlation to a maximum order of 4. From the results of Table 2, we find that all of the weekly and monthly realized volatilities have significantly impact on 1-step-ahead volatility at the 1% significance level. However, the impact of daily realized volatility on next-day volatility changes across different models. In general, those three components have a significantly positive influence on future short-term volatility at the 90% confidence level. The jump components calculated by different jump tests perform differently in the models. We further find that the large and small jumps based on different jump components also exhibit different behavior. We use the Wald test to check the null hypotheses, $\beta_{jl} = \beta_{js}$, and find that the impacts of large and small jumps on future 1-step-ahead volatility are significantly mixed. Notably, the adjusted R-squares of the benchmark and seven new models are larger than 0.63, implying that these models have good explanatory power due to the strong serial correlation of the volatility series.

From the results of Table 3, it is clear that the coefficients of daily and monthly RV are significant, except for that of the weekly RV. The jumps and those decompositions show unsystematic performance. For example, the large jump components, according to the LM test, have a significantly negative impact on 5-step-ahead forecasting, which is consistent with the abovementioned results for 1-step-ahead forecasting. Table 4 demonstrates the results for the 22-step-ahead volatility, which indicate that only the coefficients of weekly realized volatility are significantly rejecting the null hypothesis, $\beta_w = 0$, and show that the weekly RV can substantially influence the long-term volatility. The adjusted R-squared are larger than 0.52. We can find that the adjusted R-squared is larger when forecasting 1-step-ahead oil volatility, suggesting that these models are better for short-term forecasting.

[4] To save space, only the estimations for the extensions of the HAR-RV-TJ and HAR-RV-LMJ models are listed in Tables 2-4. The estimation results for other models are available upon request.

5.2. Forecasting evaluation

To evaluate the forecasting performance of the benchmark and extended models, we apply the rolling window method to gain the future volatility and choose the two robust loss functions HMSE and QLIKE as our evaluation criteria. The two loss functions are defined as below:

$$\mathrm{HMSE} = M^{-1}\sum_{m=1}^{M}\left(1-\hat{\sigma}_m^2/RV_m\right)^2, \tag{24}$$

$$\mathrm{QLIKE} = M^{-1}\sum_{m=1}^{M}\left(\ln\left(\hat{\sigma}_m^2\right)+\frac{RV_m}{\hat{\sigma}_m^2}\right). \tag{25}$$

where $\hat{\sigma}_m^2$ denotes the out-of-sample volatility forecast obtained from different models, RV_m is a proxy for actual market volatility in the out-of-sample period, and M is the length of forecasting days, and in our case, $M = 781$. Pesaran and Timmermann (2005) claim that the choice of window size depends on the nature of the possible model instability and the timing of the possible breaks. Specifically, a large window is preferable if the data are more stationary while a shorter window may be more robust to structural breaks. To make our conclusions robust and reliable, we choose different estimation window sizes[5] of 1300 and 1500 according to our data set.

To determine whether the benchmark and extended models are statistically significant, loss functions, superior predictive ability (SPA) test, and advanced model confidence set (MCS) test are widely employed, particularly the SPA test and the MCS test (Brailsford and Faff, 1996; Hansen, 2005; Hansen et al., 2011). Since Hansen et al. (2011) first constructed the MCS test, it has been widely applied as a main criterion for model evaluation (e.g., Martens et al., 2009; Liu et al., 2015, 2016b, 2017; Tao et al., 2018) due to its great advantages. First, acknowledging the limitations of the data, uninformative data yield an MCS with many models, while informative data can yield an MCS with only a few models. Second, the MCS test does not need to provide a benchmark model, which is very helpful in applications. Third, the MCS test result allows more than one "best" model. Hence, in this research, the novel MCS test is applied to choose a subset of models containing all possible superior models from the initial set of models.

In this study, we set the 90% confidence level, implying that the better forecasting models of the MCS p-values are larger than 0.1. Following Hansen et al. (2011), the statistics and p-values of the MCS test are

[5] More details and discussions about the optimization of estimation window can be seen in Rossi and Inoue (2012).

Table 3
The estimation results of the HAR-RV-TJ and its extended models based on the Newey-West correction ($h = 5$).

	HAR-RV -TJ	HAR-RV -TJL	HAR-RV -TJS	HAR-RV -TJLS	HAR-RV -JLM	HAR-RV -JLML	HAR-RV -JLMS	HAR-RV -JLMLS
c	0.029	0.031	0.024	0.025	0.003	0.021	0.029	0.003
	(0.020)	(0.020)	(0.021)	(0.020)	(0.032)	(0.020)	(0.029)	(0.031)
β_d	0.140*	0.138*	0.151**	0.146*	0.390**	0.238**	0.149	0.400**
	(0.072)	(0.071)	(0.077)	(0.074)	(0.198)	(0.101)	(0.167)	(0.190)
β_w	0.206	0.214	0.211	0.203	0.102	0.143	0.150	0.109
	(0.150)	(0.148)	(0.150)	(0.151)	(0.142)	(0.142)	(0.132)	(0.138)
β_m	0.639***	0.637***	0.633***	0.638***	0.598***	0.576***	0.626***	0.561***
	(0.163)	(0.164)	(0.164)	(0.164)	(0.156)	(0.146)	(0.151)	(0.149)
β_j	0.169				−0.259			
	(0.274)				(0.196)			
β_{jl}		0.131		0.130		−0.537**		−0.663**
		(0.290)		(0.289)		(0.144)		(0.198)
β_{js}			0.377	0.375			0.001	−0.185
			(0.292)	(0.288)			(0.175)	(0.192)
Adj. R^2	0.580	0.579	0.581	0.582	0.584	0.589	0.578	0.591

Notes: All models are estimated by OLS method with Newey-West correction. The numbers of parenthesis represent the Newey-West standard error of each parameter. The coefficients of constant and Newey-West standard error are multiply 1000, respectively. The asterisk ***, ** and * denotes rejections of null hypothesis at the 1%, 5% and 10% significance levels. *Adj.* R^2 is the adjust R-squared.

Table 4
The estimation results of the HAR-RV-TJ and its extended models based on the Newey-West correction ($h = 22$).

	HAR-RV -TJ	HAR-RV -TJL	HAR-RV -TJS	HAR-RV -TJLS	HAR-RV -JLM	HAR-RV -JLML	HAR-RV -JLMS	HAR-RV -JLMLS
c	0.051***	0.050***	0.052***	0.051***	0.026	0.046	0.043	0.027
	(0.019)	(0.019)	(0.019)	(0.018)	(0.027)	(0.020)	(0.028)	(0.027)
β_d	0.110	0.111	0.108	0.110	0.307	0.127	0.154	0.312
	(0.086)	(0.087)	(0.085)	(0.087)	(0.204)	(0.091)	(0.188)	(0.205)
β_w	0.676***	0.674***	0.672***	0.676***	0.535***	0.577***	0.564***	0.539***
	(0.158)	(0.158)	(0.159)	(0.158)	(0.183)	(0.164)	(0.169)	(0.181)
β_m	0.136	0.137	0.139	0.138	0.163	0.164	0.188	0.147
	(0.156)	(0.156)	(0.154)	(0.156)	(0.152)	(0.146)	(0.153)	(0.148)
β_j	−0.055				−0.243			
	(0.201)				(0.205)			
β_{jl}		−0.056		−0.055		−0.272		−0.417
		(0.236)		(0.236)		(0.261)		(0.283)
β_{js}			−0.057	−0.056			−0.095	−0.212
			(0.274)	(0.275)			(0.205)	(0.215)
Adj. R^2	0.525	0.524	0.524	0.524	0.523	0.521	0.518	0.523

Notes: All models are estimated by OLS method with Newey-West correction. The numbers of parenthesis represent the Newey-West standard error of each parameter. The coefficients of constant and Newey-West standard error are multiply 1000, respectively. The asterisk ***, ** and * denotes rejections of null hypothesis at the 1%, 5% and 10% significance levels. *Adj.* R^2 is the adjust R-squared.

calculated by using number of simulations $B = 10{,}000$ and the block length[6] $l = 2$, which affect the autocorrelation and heteroscedasticity of the sampling data. Technical details and more in-depth discussions of the MCS test can be found in Hansen et al. (2011).

5.3. Out-of-sample evaluation

To assess the performance of the forecasting models, HMSE and QLIKE are widely applied in the forecast losses. Patton and Sheppard (2009) provide further motivation for using QLIKE in volatility forecasting applications. In particular, Patton (2011) claims that among the nine loss functions that are commonly used to compare volatility forecasts, only the MSE and QLIKE loss functions are robust to noise in the volatility proxy. Therefore, to assess the performance of the HAR-RV-type models, we do not use the loss functions HMSE and QLIKE directly but apply them as criteria in the MCS procedure. Tables 5–7 report the MCS test results for comparing the seventeen HAR-RV-type models over estimation windows of 1400, 1500, and 1300 days, respectively. Table 5 shows that the most significant models are the newly constructed ones with decomposed jumps.

Specifically, for 1-step-ahead forecasting, only the HAR-RV-TJS and HAR-RV-TJLS models significantly outperform the benchmark and other extended models under the two criteria of HMSE and QLIKE. In particular, the HAR-RV-TJS model has the largest MCS p-values and performs the best. The results imply that decomposing jumps are effective in forecasting daily oil futures volatility and that small jumps in particular can significantly improve the forecast accuracy. Thus, traders in the oil market could focus more on the common small shocks of macro news when analyzing their daily trading.

For 5-step-ahead forecasting, only the HAR-RV-JLM and HAR-RV-JLMLS models perform significantly better than do the others under the two criteria. The HAR-RV-JLM model with both large and small jumps is the best one at the weekly forecasting horizon. The results highlight the importance of decomposing jumps in forecasting the daily oil futures volatility. Both huge and small shocks in the oil market should be considered in managing oil futures risk.

Similarly, for 22-step-ahead forecasting, the HAR-RV-JLM and HAR-RV-JLMLS models also outperform the competing models under both loss functions. This finding implies that the jump components calculated by the LM test as well as its large and small components are helpful in forecasting both the 5- and 22-step-ahead volatility. Furthermore, the HAR-RV-JLMLS model shows the highest p-values and is the best at the monthly forecasting horizon. The findings also confirm the significance of decomposing jumps, although the contribution of jumps remains

[6] More details about block length selection can be seen in Bühlmann and Künsch (1999).

Table 5
The MCS test results of the benchmark and volatility models (Estimation window of 1400).

Model	HMSE	QLIKE	HMSE	QLIKE	HMSE	QLIKE
	$h=1$		$h=5$		$h=22$	
HAR-RV	0.000	**0.500**	0.000	0.012	0.000	0.000
HAR-RV-J	0.000	**0.198**	0.000	0.004	0.000	0.000
HAR-RV-JL	0.001	**1.000**	0.000	0.023	0.000	0.000
HAR-RV-JS	0.001	**0.265**	0.008	**0.246**	0.000	0.004
HAR-RV-JLS	0.056	**0.760**	0.000	**0.234**	0.000	**0.298**
HAR-RV-CJ	0.000	**0.206**	0.000	0.005	0.000	0.000
HAR-RV-CJL	0.000	**0.169**	0.000	0.001	0.000	0.000
HAR-RV-CJS	0.000	**0.265**	0.000	0.001	0.000	0.000
HAR-RV-CJLS	0.000	**0.396**	0.000	0.001	0.000	0.000
HAR-RV-TJ	0.001	**0.996**	0.000	**0.234**	0.000	0.000
HAR-RV-TJL	0.001	**0.996**	0.000	0.096	0.000	0.000
HAR-RV-TJS	**1.000**	**0.912**	0.000	**1.000**	0.000	0.001
HAR-RV-TJLS	**0.547**	**0.497**	0.000	**0.246**	0.000	0.010
HAR-RV-JLM	0.001	0.085	**0.519**	**0.246**	**0.576**	**0.664**
HAR-RV-JLML	0.001	**0.996**	0.000	**0.625**	0.000	0.000
HAR-RV-JLMS	0.000	0.008	0.000	0.001	0.000	0.000
HAR-RV-JLMLS	0.001	**0.114**	**1.000**	**0.625**	**1.000**	**1.000**

Notes: The MCS p-values which are larger 0.1 are marked in bold and underline, implying that the corresponding models significantly outperform the competing models.

Table 7
The MCS test results of the benchmark and volatility models (Estimation window of 1300).

Model	HMSE	QLIKE	HMSE	QLIKE	HMSE	QLIKE
	$h=1$		$h=5$		$h=22$	
HAR-RV	0.000	**0.538**	0.000	0.042	0.000	0.000
HAR-RV-J	0.000	**0.630**	0.000	0.003	0.000	0.000
HAR-RV-JL	0.001	**0.630**	0.000	0.050	0.000	0.001
HAR-RV-JS	**0.112**	**0.630**	0.000	**0.200**	0.000	0.001
HAR-RV-JLS	**0.375**	**0.630**	0.000	0.094	0.000	0.045
HAR-RV-CJ	0.000	**0.245**	0.000	0.006	0.000	0.000
HAR-RV-CJL	0.000	**0.389**	0.000	0.013	0.000	0.000
HAR-RV-CJS	0.000	**0.141**	0.000	0.002	0.000	0.000
HAR-RV-CJLS	0.000	**0.186**	0.000	0.003	0.000	0.000
HAR-RV-TJ	0.016	**1.000**	0.000	**0.252**	0.000	0.001
HAR-RV-TJL	0.025	**0.630**	0.000	**0.320**	0.000	0.001
HAR-RV-TJS	**0.375**	**0.538**	0.000	**0.643**	0.000	0.003
HAR-RV-TJLS	**0.375**	**0.378**	0.000	**0.122**	0.000	0.014
HAR-RV-JLM	**0.375**	**0.630**	0.072	**1.000**	0.010	0.069
HAR-RV-JLML	0.001	**0.630**	0.000	**0.625**	0.000	0.000
HAR-RV-JLMS	0.001	**0.103**	0.000	0.073	0.000	0.000
HAR-RV-JLMLS	**1.000**	**0.630**	**1.000**	**0.975**	**1.000**	**1.000**

Notes: The MCS p-values which are larger 0.1 are marked in bold and underline, implying that the corresponding models significantly outperform the competing models.

controversial. Both huge and small shocks in the oil market should be considered in both weekly and monthly trading.

Overall, the HAR-RV-TJ model with small jumps is the best at 1-step-ahead forecasting; the HAR-RV-JLM model with both large and small jumps performs the best at 5- and 22-step-ahead forecasting. Therefore, we can conclude that the best models are our proposed new models with decomposed jumps for all forecasting horizons. The results highlight the significance of decomposing jumps to improve the accuracy of oil volatility forecasts.

5.4. Robustness check

As Rossi and Inoue (2012) and Inoue et al. (2017) point out, a model's forecasting performance is sensitive to the choice of window size. However, there is no consensus on how to choose the right forecasting windows in academia. Thus, to make our conclusions robust and reliable, we choose different estimation window sizes of 1300 and 1500. From the empirical results of Tables 6 and 7, we find that the HAR-RV-TJS and HAR-RV-TJLS models achieve a significantly higher forecast accuracy for daily forecasting and that the HAR-RV-JLMLS model has better performance that do the competing models at all the forecasting horizons. Moreover, the original HAR-RV-J and HAR-RV-CJ models remain inferior to the HAR-RV-TJ and HAR-RV-JLM models, while the HAR-RV-J model with decomposed jumps, the HAR-RV-JLS shows significantly better performance for daily forecasting.

In sum, not all the HAR models with jumps can beat the simple HAR-RV; however, decomposed jumps further help improve the forecast accuracy in the oil futures market. Therefore, our conclusions are generally consistent and robust to different estimation window sizes.

Therefore, market microstructure noise cannot be ignored when estimating and forecasting the realized volatility. We use an alternative actual volatility, the realized kernel (RK), which is robust to noise (Barndorff-Nielsen et al., 2008; Wang et al., 2016), to examine whether our results are robust. The RK is defined as below:

$$RK_t = \sum_{j=-H}^{H} k\left(\frac{j}{j+1}\right)\eta_j, \eta_j = \sum_{i=|j|+1}^{n} r_{t,i} r_{t,i-|j|}. \tag{26}$$

where $k(x)$ is the Parzen kernel function:

$$k(x) = \begin{cases} 1-6x^2+6x^3, & 0 \le x \le 1/2 \\ 2(1-x)^3, & 1/2 \le x \le 1 \\ 0, & x > 1. \end{cases} \tag{27}$$

It is necessary for H to increase with the sample size to estimate the increments of quadratic variation consistently in the presence of noise. We follow precisely the bandwidth choice of H that was described by Barndorff-Nielsen et al. (2009).

Table 8 reports the MCS test results with the RK as the benchmark volatility. The HAR-RV-TJS and HAR-RV-TJLS models exhibit a higher ability in predicting the daily oil futures volatility. The HAR-RV-JLM and HAR-RV-JLMLS models have better performance in forecasting the weekly and monthly oil futures volatility. The better forecasting models all are our proposed models, and the results coincide with the conclusions drawn from Tables 5–7.

6. Concluding remarks

Among the growing body of literature on modelling and forecasting oil volatility, few studies consider the further application of jump components. This paper investigates the contribution of decomposed jumps in forecasting oil futures volatility and construct twelve new models with decomposed jumps. There is strong evidence that the HAR-RV-TJS model performs best at 1-step-ahead forecasting, while the HAR-RV-JLMLS model incorporating both small and large jump

Table 6
The MCS test results of the benchmark and volatility models (Estimation window of 1500).

Model	HMSE	QLIKE	HMSE	QLIKE	HMSE	QLIKE
	$h=1$		$h=5$		$h=22$	
HAR-RV	0.000	0.025	0.000	0.000	0.000	0.000
HAR-RV-J	0.000	0.021	0.000	0.000	0.000	0.000
HAR-RV-JL	0.000	**0.743**	0.000	0.000	0.000	0.000
HAR-RV-JS	0.001	**0.307**	0.000	0.025	0.000	0.000
HAR-RV-JLS	**0.597**	**0.743**	0.000	0.003	0.000	0.000
HAR-RV-CJ	0.000	0.054	0.000	0.000	0.000	0.000
HAR-RV-CJL	0.000	0.028	0.000	0.000	0.000	0.000
HAR-RV-CJS	0.000	0.035	0.000	0.000	0.000	0.000
HAR-RV-CJLS	0.000	0.042	0.000	0.000	0.000	0.000
HAR-RV-TJ	0.000	**0.107**	0.000	0.000	0.000	0.000
HAR-RV-TJL	0.000	**0.142**	0.000	0.001	0.000	0.000
HAR-RV-TJS	**0.597**	**0.142**	0.000	0.006	0.000	0.000
HAR-RV-TJLS	**1.000**	**0.128**	0.000	0.000	0.000	0.000
HAR-RV-JLM	**0.158**	**0.132**	**1.000**	0.036	**1.000**	0.013
HAR-RV-JLML	0.000	**1.000**	0.000	0.036	0.000	0.000
HAR-RV-JLMS	0.000	0.003	0.000	0.000	0.000	0.000
HAR-RV-JLMLS	0.001	**0.743**	**0.191**	**1.000**	**0.189**	**1.000**

Notes: The MCS p-values which are larger 0.1 are marked in bold and underline, implying that the corresponding models significantly outperform the competing models.

Table 8
The MCS test results of the benchmark and volatility models as the actual volatility, RK.

Model	HMSE	QLIKE	HMSE	QLIKE	HMSE	QLIKE
	$h=1$		$h=5$		$h=22$	
HAR-RV	0.0000	0.0020	0.0000	0.0000	0.0000	0.0000
HAR-RV-J	0.0000	0.0020	0.0000	0.0000	0.0000	0.0000
HAR-RV-JL	0.0000	**0.4220**	0.0000	0.0000	0.0000	0.0000
HAR-RV-JS	0.0000	**0.2470**	**0.2033**	**0.9620**	0.0000	0.0000
HAR-RV-JLS	0.0330	**0.4220**	0.0015	**0.7310**	0.0010	**0.3260**
HAR-RV-CJ	0.0000	0.0110	0.0000	0.0000	0.0000	0.0000
HAR-RV-CJL	0.0000	0.0020	0.0000	0.0000	0.0000	0.0000
HAR-RV-CJS	0.0000	0.0040	0.0000	0.0000	0.0000	0.0000
HAR-RV-CJLS	0.0000	0.0060	0.0000	0.0000	0.0000	0.0000
HAR-RV-TJ	0.0000	**0.3940**	0.0000	0.0350	0.0000	0.0000
HAR-RV-TJL	0.0000	**0.4220**	0.0000	0.0110	0.0000	0.0000
HAR-RV-TJS	**1.0000**	**1.0000**	0.0004	**0.9620**	0.0000	0.0000
HAR-RV-TJLS	**0.9160**	**0.4220**	0.000	0.0610	0.0000	0.0090
HAR-RV-JLM	0.0000	0.0640	**1.0000**	**1.0000**	**1.0000**	**0.9310**
HAR-RV-JLML	0.0000	**0.3830**	0.000	0.0350	0.0000	0.0000
HAR-RV-JLMS	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
HAR-RV-JLMLS	0.0000	0.0270	**0.6873**	**0.9620**	**0.7710**	**1.0000**

Notes: The MCS p-values which are larger 0.1 are marked in bold and underline, implying that the corresponding models significantly outperform the competing models.

components outperforms the others in both 5- and 22-step-ahead forecasting. This finding implies that the use of decomposing jumps in forecasting oil futures volatility can lead to performance improvements. A potential reason is that decomposed jumps contain more effective information on various uncertain factors in the oil futures market. Small jumps reflect information in a short period, however, large jumps require a longer period to digest the information in the oil futures market.

The findings from this paper indicate that oil market traders and practitioners could focus on huge macro news shocks when managing long-term risk and focus on small shocks in the oil market when managing short-term risk. Using HAR models with decomposed jumps to forecast oil futures volatility, one can obtain accurate results regarding market risk. In particular, modelling oil futures volatility with small jumps helps improve the forecast accuracy at the 1-day forecasting horizon, while modelling with both large and small jumps helps forecast the weekly and monthly volatility. Moreover, our findings are important for policy makers, who use oil market volatility as a barometer of the fragility of the energy markets and the economy.

In addition, our results indicate some interesting directions for future research. First, further investigation of the use of jumps such as decomposing jumps could be considered in forecasting the oil futures market. Second, decomposing jumps with dynamic thresholds could be an interesting direction for future research.

Acknowledgement

We are thankful for the chief-in-editor, Professor Richard S.J. Tol, and several referees, who provided us with more invaluable comments to improve the quality of this paper. Feng Ma is grateful for the financial support from the Natural Science Foundation of China [71671145, 71701170, 71701171], the humanities and social science fund of the Ministry of Education [17YJC790105, 17XJCZH002], and the Fundamental Research Funds for the Central Universities [26816WCX02, 2682017WCX01]. Ke Yang is supported by the financial support from the Natural Science Foundation of China [71673089], the Fundamental Research Funds for the Central Universities [2017MS114], and the National Social Science Foundation of China (17ZDA073).

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., Labys, P., 2003. Modeling and forecasting realized volatility. Econometrica 71 (2), 579–625.
Andersen, T.G., Bollerslev, T., Diebold, F.X., 2007. Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. Rev. Econ. Stat. 89 (4), 701–720.
Barndorff-Nielsen, O.E., Shephard, N., 2004. Power and bipower variation with stochastic volatility and jumps. J. Financ. Economet. 2 (1), 1–37.
Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A., Shephard, N., 2008. Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. Econometrica 76 (6), 1481–1536.
Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A., Shephard, N., 2009. Realized kernels in practice: trades and quotes. Econ. J. 12 (3), C1–C32.
Bekaert, G., Hoerova, M., 2014. The VIX, the variance premium and stock market volatility. J. Econ. 183 (2), 181–192.
Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. J. Econ. 31 (3), 307–327.
Bollerslev, T., Patton, A.J., Quaedvlieg, R., 2016. Exploiting the errors: a simple approach for improved volatility forecasting. J. Econ. 192 (1), 1–18.
Brailsford, T.J., Faff, R.W., 1996. An evaluation of volatility forecasting techniques. J. Bank. Financ. 20 (3), 419–438.
Bühlmann, P., Künsch, H.R., 1999. Block length selection in the bootstrap for time series. Comput. Stat. Data Anal. 31 (3), 295–310.
Çelik, S., Ergin, H., 2014. Volatility forecasting using high frequency data: evidence from stock markets. Econ. Model. 36, 176–190.
Chen, X., Ghysels, E., 2011. News—good or bad—and its impact on volatility predictions over multiple horizons. Rev. Financ. Stud. 24 (1), 46–81.
Cheung, Y.W., Lai, K.S., 1995. Lag order and critical values of the augmented Dickey–Fuller test. J. Bus. Econ. Stat. 13 (3), 277–280.
Clements, A., Liao, Y., 2017. Forecasting the variance of stock index returns using jumps and cojumps. Int. J. Forecast. 33 (3), 729–742.
Corsi, F., 2009. A simple approximate long-memory model of realized volatility. J. Financ. Economet. 7 (2), 174–196.
Corsi, F., Pirino, D., Reno, R., 2010. Threshold bipower variation and the impact of jumps on volatility forecasting. J. Econ. 159 (2), 276–288.
Degiannakis, S., Filis, G., 2017. Forecasting oil price realized volatility using information channels from other asset classes. J. Int. Money Financ. 76, 28–49.
Dumitru, A.-M., Urga, G., 2012. Identifying jumps in financial assets: a comparison between nonparametric jump tests. J. Bus. Econ. Stat. 30 (2), 242–255.
Duong, D., Swanson, N.R., 2011. Volatility in discrete and continuous-time models: a survey with new evidence on large and small jumps. Adv. Econ. 27, 179.
Duong, D., Swanson, N.R., 2015. Empirical evidence on the importance of aggregation, asymmetry, and jumps for volatility prediction. J. Econ. 187 (2), 606–621.
Efimova, O., Serletis, A., 2014. Energy markets volatility modelling using GARCH. Energy Econ. 43, 264–273.
Elder, J., Miao, H., Ramchander, S., 2013. Jumps in oil prices: the role of economic news. Energy J. 34 (3).
Gilder, D., Shackleton, M.B., Taylor, S.J., 2014. Cojumps in stock prices: empirical evidence. J. Bank. Financ. 40, 443–459.
Hansen, P.R., 2005. A test for superior predictive ability. J. Bus. Econ. Stat. 23 (4), 365–380.
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, X., Tauchen, G., 2005. The relative contribution of jumps to total price variance. J. Financ. Economet. 3 (4), 456–499.
Inoue, A., Jin, L., Rossi, B., 2017. Rolling window selection for out-of-sample forecasting with time-varying parameters. J. Econ. 196 (1), 55–67.
Jarque, C.M., Bera, A.K., 1987. A test for normality of observations and regression residuals. Int. Stat. Rev. 163–172.
Kang, S.H., Kang, S.-M., Yoon, S.-M., 2009. Forecasting volatility of crude oil markets. Energy Econ. 31 (1), 119–125.
Lautier, D., Riva, F., 2004. Volatility in the American crude oil futures market. Cahier de recherche du CEREG n, 08.
Lee, S.S., Mykland, P.A., 2008. Jumps in financial markets: a new nonparametric test and jump dynamics. Rev. Financ. Stud. 21 (6), 2535–2563.
Liu, J., Longstaff, F.A., Pan, J., 2003. Dynamic asset allocation with event risk. J. Financ. 58 (1), 231–259.
Liu, L.Y., Patton, A.J., Sheppard, K., 2015. Does anything beat 5-minute RV? A comparison of realized measures across multiple asset classes. J. Econ. 187 (1), 293–311.
Liu, L., Wang, Y., Wu, C., Wu, W., 2016a. Disentangling the determinants of real oil prices. Energy Econ. 56, 363–373.
Liu, Z., Ma, F., Wang, X., Xia, Z., 2016b. Forecasting the realized volatility: the role of jumps. Appl. Econ. Lett. 23 (10), 736–739.
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.
Ljung, G.M., Box, G.E., 1978. On a measure of lack of fit in time series models. Biometrika 65 (2), 297–303.
Ma, F., Liu, L., Liu, Z., Wei, Y., 2015. Forecasting realized range volatility: a regime-switching approach. Appl. Econ. Lett. 22 (17), 1361–1365.
Ma, F., Liu, J., Huang, D., Chen, W., 2017a. Forecasting the oil futures price volatility: a new approach. Econ. Model. 64, 560–566.
Ma, F., Wahab, M., Huang, D., Xu, W., 2017b. Forecasting the realized volatility of the oil futures market: a regime switching approach. Energy Econ. 67, 136–145.

Ma, F., Wei, Y., Chen, W., He, F., 2017c. Forecasting the volatility of crude oil futures using high-frequency data: further evidence. Empir. Econ. 1–26.
Martens, M., Van Dijk, D., De Pooter, M., 2009. Forecasting S&P 500 volatility: long memory, level shifts, leverage effects, day-of-the-week seasonality, and macroeconomic announcements. Int. J. Forecast. 25 (2), 282–303.
Naik, V., Lee, M., 1990. General equilibrium pricing of options on the market portfolio with discontinuous returns. Rev. Financ. Stud. 3 (4), 493–521.
Nomikos, N., Andriosopoulos, K., 2012. Modelling energy spot prices: empirical evidence from NYMEX. Energy Econ. 34 (4), 1153–1169.
Patton, A.J., 2011. Volatility forecast comparison using imperfect volatility proxies. J. Econ. 160 (1), 246–256.
Patton, A.J., Sheppard, K., 2009. Evaluating volatility and correlation forecasts. Handbook of Financial Time Series. Springer, pp. 801–838.
Patton, A.J., Sheppard, K., 2015. Good volatility, bad volatility: signed jumps and the persistence of volatility. Rev. Econ. Stat. 97 (3), 683–697.
Pesaran, M.H., Timmermann, A., 2005. Small sample properties of forecasts from autoregressive models under structural breaks. J. Econ. 129 (1), 183–217.
Prokopczuk, M., Symeonidis, L., Wese Simen, C., 2016. Do jumps matter for volatility forecasting? Evidence from energy markets. J. Futur. Mark. 36 (8), 758–792.
Pu, W., Chen, Y., Ma, F., 2016. Forecasting the realized volatility in the Chinese stock market: further evidence. Appl. Econ. 48 (33), 3116–3130.
Rossi, B., Inoue, A., 2012. Out-of-sample forecast tests robust to the choice of window size. J. Bus. Econ. Stat. 30 (3), 432–453.
Sévi, B., 2014. Forecasting the volatility of crude oil futures using intraday data. Eur. J. Oper. Res. 235 (3), 643–659.
Tao, Q., Wei, Y., Liu, J., Zhang, T., 2018. Modeling and forecasting multifractal volatility established upon the heterogeneous market hypothesis. Int. Rev. Econ. Financ. 54, 143–153.
Wang, Y., Ma, F., Wei, Y., Wu, C., 2016. Forecasting realized volatility in a changing world: a dynamic model averaging approach. J. Bank. Financ. 64, 136–149.
Wei, Y., Wang, Y., Huang, D., 2010. Forecasting crude oil market volatility: further evidence using GARCH-class models. Energy Econ. 32 (6), 1477–1484.
Wen, F., Gong, X., Cai, S., 2016. Forecasting the volatility of crude oil futures using HAR-type models with structural breaks. Energy Econ. 59, 400–413.