Energy Economics 67 (2017) 136–145

Contents lists available at ScienceDirect

Energy Economics

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

Forecasting the realized volatility of the oil futures market: A regime switching approach

Feng Ma [a,*], M.I.M. Wahab [b], Dengshi Huang [a], Weiju Xu [a]

[a] School of Economics & Management, Southwest Jiaotong University, Chengdu, China
[b] Department of Mechanical and Industrial Engineering, Ryerson University, Toronto, Canada

ARTICLE INFO

Article history:
Received 11 November 2016
Received in revised form 9 July 2017
Accepted 8 August 2017
Available online 19 August 2017

JEL classification:
C22
C53
C58
E27
E37
G14

Keywords:
Volatility forecasting
HAR-RV-type models
Regime switching approach
Forecasting evaluation

ABSTRACT

Considering nonlinear and highly persistent dynamics of realized volatility, we introduce Markov regime switching models to the Heterogeneous Autoregressive model of the Realized Volatility (HAR-RV) models to forecast the realized volatility of the crude oil futures market. In-sample results demonstrate that the high volatility regime is short-lived. Out-of-sample results suggest that HAR-RV models with regime switching increase the forecasting ability significantly than those without regime switching. Moreover, these findings are robust for different actual volatility benchmarks, forecasting windows, and model settings.

1. Introduction

Oil is an important energy commodity that plays an essential role in the world economy. Oil price volatility has a significant macroeconomic influence to the real economy (Hamilton, 1983, 2003; Kilian and Park, 2009) and financial markets (Aloui and Jammazi, 2009; Kilian and Park, 2009). Oil price volatility is also an important issue for risk management, derivative pricing, portfolio selection, and many other financial activities. Thus, modeling and forecasting the volatility of crude oil price is critical for researchers, market participants, and policymakers.

Modeling and predicting oil price volatility are investigated based on the framework of the GARCH-class models (e.g., Agnolucci, 2009; Cheong, 2009; Kang et al., 2009; Mohammadi and Su, 2010; Wei et al., 2010; Nomikos and Pouliasis, 2011; Nomikos and Andriosopoulos, 2012; Wang and Wu, 2012; Efimova and Serletis, 2014). However, GARCH-class models are constructed for daily or even a lower frequency data, which can result in a substantial loss of intraday trading information. Because of the availability of abundant high-frequency (intraday) data in recent years, research on financial market volatility has taken new avenues. Moreover, high-frequency data contains a wealth of information that can help market participants to make quicker decisions. As a result, volatility measure based on high-frequency data has received much attention in academia.

The seminal work on measuring volatility using high-frequency data by Andersen and Bollerslev (1998) proposes the realized volatility or variance[1] (RV), which is robust to market microstructure effects. For a given fixed interval, RV is defined as the sum of squared returns over non-overlapping intervals. Thus, RV can directly be observed, and it enables researchers to gauge the level of RV and understand its dynamics. The early study on describing and predicting RV is based on the autoregressive fractionally integrated moving average (ARFIMA) model proposed by Andersen et al. (2003). Although the ARFIMA model achieves a higher forecast accuracy than GARCH-class models (e.g., Koopman et al., 2005; Liu and Wan, 2012), Corsi (2009) points out that the ARFIMA model is just a convenient mathematical trick, lacks a clear economic interpretation, and leads to the loss of information on a vast number of transactions. Corsi (2009) also constructs a simple

* Corresponding author.
E-mail addresses: mafeng2016@swjtu.edu.cn (F. Ma), wahab@ryerson.ca (M.I.M. Wahab).

[1] We will use the terms realized volatility and realized variation (variance) interchangeably.

http://dx.doi.org/10.1016/j.eneco.2017.08.004
0140-9883/

heterogeneous autoregressive model of the realized volatility (HAR-RV), which can capture "stylized facts" in the financial markets, such as long memory and multi-behavior. As a result, HAR-RV model is commonly employed to forecast the volatility using high-frequency data.

Studies on forecasting volatility using high-frequency data have mainly concentrated on the stock and exchange rate markets (e.g., Andersen et al., 2007a; Corsi et al., 2010; Bekaert and Hoerova, 2014; Bollersleva et al., 2015; Degiannakis, 2008; Duong and Swanson, 2015; Wang et al., 2016). Nevertheless, to the best of our knowledge, there are insufficient studies on forecasting the oil futures price volatility using high-frequency data. For example, Degiannakis and George (2016) point out that studies, such as Haugom et al. (2014), Sévi (2014), Prokopczuk et al. (2016), and Wen et al. (2016), try to forecast oil price volatility using ultra-high frequency data. Therefore, in this research, as a first step, we use the HAR-RV model and its various extensions to forecast the realized volatility of the oil futures price. To be precise, we label those twelve models as HAR-RV-type models, which include: HAR-RV (Corsi, 2009), HAR-RV-J and HAR-RV-CJ (Andersen et al., 2007a, 2007b), HAR-RV-TJ (Corsi et al., 2010), HAR-S-RV-J (Chen and Ghysels, 2011), HAR-RV-PS2 and HAR-RV-PS3 (Patton and Sheppard, 2015), HAR-CSJ and HAR-CSJd (Sévi, 2014), HAR-ARJ (Prokopczuk et al., 2016), HAR-RV-JLM and HAR-S-RV-J-JLM (Liu et al., 2016).

HAR-RV-type models are linear, and the estimated coefficients of those models are constant. However, Granger and Ding (1996) find out that persistence in volatility is usually non-constant over time. Previous studies (e.g., Longin, 1997; Raggi and Bordignon, 2012; Goldman et al., 2013; Ma et al., 2015) provide evidence that a higher level of persistence exists when volatility is low, implying the presence of nonlinearities. Moreover, it is well known that due to many factors, such as business cycle, major events, and economic policy, the statistical property of volatility (e.g., volatility persistence) always undergoes structural breaks (e.g., Banerjee and Urga, 2005; Wahab and Lee, 2009) or switches between different regimes (Hamilton and Susmel, 1994). Therefore, it is appropriate to use a model with regime switching to describe volatility dynamics. For example, Goldman et al. (2013) use threshold autoregressive fractionally integrated moving average (TARFIMA) models with regime switching and show that TARFIMA achieves a higher forecast accuracy than ARFIMA. Raggi and Bordignon (2012) find that introducing nonlinearities leads to a better prediction for several forecast horizons. Though it is the fact that considering nonlinear and highly persistent dynamics of realized volatility can significantly improve the forecasting performance, a little has been done on forecasting realized volatility by using HAR-RV-type models with regime switching. Therefore, we introduce regime-switching characteristics to the HAR-RV-type models and examine whether HAR-RV-type models with regime switching can forecast better than HAR-RV-type models without regime switching. In regime switching models, one of the aspects is deciding the number of regimes. Similar to the existing studies, such as Bekaert et al. (2015), Goldman et al. (2013), Ma et al. (2015), Raggi and Bordignon (2012), Shi and Ho (2015), and Wang et al. (2016), we also consider two regimes: low volatility regime and high volatility regime.

The aim of this paper is to forecast the realized volatility of the crude oil futures price using HAR-RV models and their extensions. This research contributes to the literature of modeling realized volatility in two ways: (a) to forecast the realized volatility of the oil futures market, we consider the nonlinear and highly persistent dynamics of realized volatility, combine HAR-RV-type models with the regime switching, and construct new volatility models, which can provide a new perspective to model and forecast the volatility of the oil futures. Furthermore, the proposed models are instrumental since Nomikos and Pouliasis (2011) point out that a regime-switching model may be more suitable for modeling volatility, particularly in energy markets, where structural breaks are quite common, because oil market volatility is characterized by different dynamics under different market conditions. For instance, Fong and See (2002, 2003) document a strong evidence of regime switching in the temporal volatility dynamics of oil futures, consistent with the theory of storage. Nomikos and Pouliasis (2011) also state that an increase in backwardation is more likely to increase regime persistence in the high volatility state due to low inventories; and (b) evaluating the forecasting ability of HAR-RV-type models with regime switching, we find that those models can gain greater accuracy in prediction and that further promotes the applications of those models to forecast the realized volatility using the ultra-high frequency data. Moreover, the proposed HAR-RV-type models with regime switching can also be applied to forecast the future volatility of the other markets, such as stock and exchange markets.

In this paper, we compare the forecasting performance HAR-RV-type models and their extensions with regime switching based on the model confidence set (MCS) test under HMSE and HMAE loss functions. In-sample results show that the negative semi-variation has a significantly positive impact on the realized volatility, implying that the negative semi-variance contributes more to the realized volatility. Also, the high volatility regime is short-lived. Out-of-sample empirical results indicate that introducing the regime-switching behavior of daily realized volatility in HAR-RV-type models leads to greater forecast accuracy. Our results also show that the same findings are valid for another volatility benchmark - realized kernel (RK) (e.g., Barndorff-Nielsen et al., 2008) and different forecasting windows. We further consider high and low volatility regimes for all variables in volatility models and warrant that regime switching can significantly help in forecasting.

The rest of the paper is organized as follows: Section 2 describes the volatility measures and models. The methodology of out-of-sample forecasting and the Model Confidence Set (MCS) test are discussed in Section 3. Section 4 provides the data and some preliminary analysis. The empirical forecasting results are presented in Section 5. Section 6 concludes the paper.

2. Volatility models

Section 2 briefly describes several popular volatility measures based on intraday high-frequency data and the corresponding extended models with regime switching capturing the volatility dynamics.

2.1. Realized volatility and realized bi-power variation measures

The primary interest is to measure the daily variance of oil futures returns, which will be estimated from the realized variance. For a given day t, we divide the time interval, which is considered as [0, 1], into n subintervals of length, where $n = 1/\Delta$ and Δ is the sampling frequency. Consequently, the realized volatility can be defined as the sum of all available intraday high-frequency squared returns and given by,

$$RV_t = \sum_{j=1}^{1/\Delta} r^2_{(t-1)+j^*\Delta,\Delta} \tag{1}$$

where $r_{(t-1)+j^*\Delta,\Delta}$ represents the intraday returns. According to Barndorff-Nielsen and Shephard (2004), when $\Delta \to 0$, RV can be expressed as:

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

where $\int_0^t \sigma^2(s)ds$ is called as the integrated variance computed by realized bi-power variation (BPV), which can be defined as:

$$BPV_t = u_1^{-2} \sum_{j=2}^{1/\Delta} \left| r_{(t-1)+j^*\Delta,\Delta} \right| \left| r_{(t-1)+(j-1)*\Delta,\Delta} \right| \tag{3}$$

where $u_1 \simeq 0.7979$. $\sum_{0<s\leq t} \kappa^2(s)$ is the discontinuous jump part of the quadratic variation (QV) process. Let $J_t = \sum_{0<s\leq t} \kappa^2(s)$ and that can be written as $J_t = \max(RV_t - BPV_t, 0)$ (Barndorff-Nielsen and Shephard, 2004; Andersen et al., 2007a, 2007b).

2.2. HAR-RV and its regime switching model

As far as we are aware, the heterogeneous autoregressive (HAR) model (Corsi, 2009) has become a popular model for describing the dynamics of RV. This model accommodates some of the stylized facts found in financial asset return volatility, such as long memory and multiscaling behavior. The model only contains three explanatory variables: daily realized volatility, weekly realized volatility, and monthly realized volatility. Those variables represent the behavior of short-term, medium-term, and long-term volatilities, respectively. The model is defined as:

$$RV_{t+1} = c + \beta_d RV_t + \beta_w RV_{t-4,t} + \beta_m RV_{t-21,t} + \omega_{t+1} \tag{4}$$

where RV_t is the average RV on day t, $RV_{t-4,t}$ is the average RV from day $t-4$ to day t, $RV_{t-21,t}$ is the average RV from day $t-21$ to day t, and ω_{t+1} is disturbance error.

Given the studies, such as Bekaert et al. (2015), Goldman et al. (2013), Ma et al. (2015), Nomikos and Pouliasis (2011), Raggi and Bordignon (2012), Shi and Ho (2015), and Wang et al. (2016), we find that the HAR-RV model with regime switching is a suitable model to forecast the crude oil futures price volatility. To the best of our knowledge, there is insufficient literature focusing on the realized volatility of crude oil futures market. It is worth to note that a popular idea in econometrics applications has been some form of regime-switching, which means that the data-generating process is viewed as a linear process that switches among different regimes according to some rules. In this study, we consider two regimes similar to Raggi and Bordignon (2012), Goldman et al. (2013), Ma et al. (2015), Shi and Ho (2015) and Wang et al. (2016). Consequently, the HRA-RV model with regime switching is termed as MS-HAR-RV and given by,

$$RV_{t+1} = c_{s_t} + \beta_{d,s_t} RV_t + \beta_w RVW_t + \beta_m RVM_t + \omega_{t+1} \tag{5}$$

where $\omega_{t+1} \sim N(0, \sigma_{s_t}{}^2)$. Let S_t be an unobserved state variable. $S_t = 0$ indicates the low volatility regime with a smaller conditional variance, meaning that the market is stable. Whereas, $S_t = 1$ indicates the high volatility regime with a larger conditional variance, meaning that the market is highly fluctuating. The unobserved state variable, S_t, is assumed to follow a two-state Markov process with a transition probability matrix given by,

$$P = \begin{bmatrix} p^{00} & 1-p^{00} \\ 1-p^{11} & p^{11} \end{bmatrix}, \tag{6}$$

where,

$$p^{00} = p(s_t = 0 | s_{t-1} = 0), \tag{7}$$

$$p^{11} = p(s_t = 1 | s_{t-1} = 1). \tag{8}$$

The MS-HAR-RV model can be estimated by the maximum likelihood function using the filtering procedure of Hamilton (1990) followed by the smoothing algorithm of Kim (1994). The log likelihood of this model is given by[2]:

$$\ln L = \sum_{t=1}^{T} \ln \left(\frac{1}{\sqrt{2\pi\sigma_{s_t}^2}} \exp\left(-\frac{RV_{t+1} - c_{s_t} - \beta_{d,s_t} RV_t - \beta_w RVW_t - \beta_m RVM_t}{2\sigma_{s_t}^2} \right) \right). \tag{9}$$

[2] Due to scarcity of the space, we do not report the procedures on estimation. Please see more details on the estimation method in Hamilton and Susmel (1994), Kim and Nelson (1999) and Perlin (2015). In particular, the estimation and forecasting procedures can clearly be found in Perlin (2015) from page 7 to page 9.

2.3. HAR-RV-type models with regime switching

The focus of this study is to examine whether introducing a regime switching model can help improve forecasting the realized volatility of crude oil futures market. To this end, we choose another eleven popular volatility models from the literature, for example, Andersen et al. (2007a, 2007b), Corsi (2009), Corsi et al. (2010), Sévi (2014), Patton and Sheppard (2015), Prokopczuk et al. (2016), and then extend them with regime switching. Each extended model has been indicated by an additional label "MS-".

Andersen et al. (2007a, 2007b) extend HAR-RV by adding the jump component (HAR-RV-J), and the MS-HAR-RV-J model is given by:

$$RV_{t+1} = c_{s_t} + \beta_{d,s_t} RV_t + \beta_w RV_{t-4,t} + \beta_m RV_{t-21,t} + \beta_j J_t + \omega_{t+1}. \tag{10}$$

where J_t is the jump component. Furthermore, RV can be decomposed into continuous sample path and jump component. Based on this decomposition, Andersen et al. (2007a, 2007b) propose the HAR-RV-CJ model, and hence the MS-HAR-RV-CJ model can be written as:

$$RV_{t+1} = c_{s_t} + \beta_{d,s_t} C_t + \beta_w C_{t-4,t} + \beta_m C_{t-21,t} + \beta_{jd} CJ_t + \beta_{jw} CJ_{t-4,t} + \beta_{jm} CJ_{t-21,t} + \omega_{t+1}. \tag{11}$$

where C_t is the continue sample path and CJ_t is the significant jump. They are determined by:

$$C_t = I(Z_t \le \Phi_\alpha) \cdot RV_t + I(Z_t > \Phi_\alpha) \cdot BPV_t, \tag{12}$$

$$CJ_t = I(Z_t > \Phi_\alpha) \cdot [RV_t - BPV_t]. \tag{13}$$

where $I(\cdot)$ is the indicator function; $C_{t-4,t}$ and $C_{t-21,t}$ are the weekly and monthly averages of the continuous sample paths, respectively; $CJ_{t-4,t}$ and $CJ_{t-21,t}$ are the weekly and monthly averages of the jumps, respectively, and Z_t is the statistics of jump test proposed by Huang and Tauchen (2005).

Corsi et al. (2010) propose a novel test for jump detection (C_TZ)[3] and combine the idea of HAR-RV model to construct a new model, HAR-RV-TJ. Consequently, MS-HAR-RV-TJ is given by:

$$RV_{t+1} = c_{s_t} + \beta_{d,s_t} TCRV_{t,\alpha} + \beta_w TCRVW_{t,\alpha} + \beta_m TCRVM_{t,\alpha} + \beta_j TJ_{t,\alpha} + \omega_{t+1}. \tag{14}$$

where $TCRV_{t,\alpha}$ is the continue sample path component, which is calculated by using Eq. (15), and $TJ_{t,\alpha}$ is the significant discontinuous jump component, which is defined in Eq. (16).

$$TCRV_{t,\alpha} = I(C\_TZ_t \le \Phi_\alpha) \cdot RV_t + I(C\_TZ_t > \Phi_\alpha) \cdot C\_TBPV_t, \tag{15}$$

$$TJ_t \equiv I(C\_TZ_t > \Phi_\alpha) \cdot [RV_t - C\_TBPV_t]. \tag{16}$$

Chen and Ghysels (2011) construct the HAR-S-RV-J model based on the realized semi-variance proposed by Barndorff-Nielsen et al. (2010). As a result, the specification of MS-HAR-S-RV-J model is given by:

$$RV_{t+1} = c_{s_t} + \beta_{dp,s_t} RS_t^+ + \beta_{wp} RS_{t-4,t}^+ + \beta_{mp} RS_{t-21,t}^+ + \beta_{dn,s_t} RS_t^- + \beta_{wn} RS_{t-4,t-1}^- + \beta_{mn} RS_{t-21,t}^- + \beta_j J_t + \omega_{t+1}. \tag{17}$$

where RS_t^- is defined as the sum of all available squared negative intraday returns, and RS_t^+ is defined as the sum of all available squared positive

[3] Corsi et al. (2010) used the simulating data to compare the performances of the two tests, and found that in the presence of jumps the C_TZ test has substantially more powerful than the Z test, especially when jumps are consecutive, a situation which is quite frequent for high-frequency data.

intraday returns. Moreover, Barndorff-Nielsen et al. (2010) prove that when the sampling interval $\Delta \to 0$, RS_t^- and RS_t^+ can be computed as follows:

$$RS_t^- = \sum_{j=1}^{1/\Delta} r^2_{(t-1)+j^*\Delta,\Delta} I\left(r_{(t-1)+j^*\Delta} < 0\right), \tag{18}$$

$$RS_t^+ = \sum_{j=1}^{1/\Delta} r^2_{(t-1)+j^*\Delta,\Delta} I\left(r_{(t-1)+j^*\Delta} > 0\right). \tag{19}$$

and $RS^-_{t-4,t}(RS^+_{t-4,t})$ and $RS^-_{t-21,t}(RS^+_{t-21,t})$ are weekly, and monthly average lagged negative (positive) realized semi-variances, respectively.

To capture the role of the "leverage effect" in volatility dynamics, Patton and Sheppard (2015) develop a series of models using signed realized measures. The first model capturing the "leverage effect" contains a signed jump variation and an estimator of the variation caused by the continuous part (bi-power variation). Hence, the first model, MS-HAR-RV-PS2, is given by:

$$RV_{t+1} = c_{s_t} + \varphi SJV_t + \beta_{bpv,s_t} BPV_t + \beta_w RV_{t-4,t} + \beta_m RV_{t-21,t} + \omega_{t+1}. \tag{20}$$

where $SJV_t = RS_t^+ - RS_t^-$. The last model, MS-HAR-RV-PS3, for the "leverage effect" disentangles the role of the positive and negative jumps and is given by:

$$RV_{t+1} = c_{s_t} + \varphi^+ SJV_t^+ + \varphi^- SJV_t^- + \beta_{bpv,s_t} BPV_t + \beta_w RV_{t-4,t} + \beta_m RV_{t-21,t} + \omega_{t+1}. \tag{21}$$

where $SJV_t^+ = SJV_t I(SJV_t > 0)$ and $SJV_t^- = SJV_t I(SJV_t < 0)$.

Sévi (2014) add the weekly and monthly average lagged of SJV_t, SJV_t^+ and SJV_t^- into models HAR-RV-PS2 and HAR-RV-PS3, and replaced $RV_{t-4,t}(RV_{t-21,t})$ with $C_{t-4,t}(C_{t-21,t})$. That leads to another two HAR-RV type models. Consequently, the first model, MS-HAR-CSJ, is:

$$RV_{t+1} = c_{s_t} + \varphi_d SJV_t + \beta_{bpv,s_t} BPV_t + \beta_w C_{t-4,t} + \beta_m C_{t-21,t} + \varphi_w SJV_{t-4,t} + \varphi_m SJV_{t-21,t} + \omega_{t+1}. \tag{22}$$

The second model, MS-HAR-CSJd, is:

$$RV_{t+1} = c_{s_t} + \varphi_d^+ SJV_t^+ + \varphi_d^- SJV_t^- + \beta_{bpv,s_t} BPV_t + \beta_w C_{t-4,t} + \beta_m C_{t-21,t} + \varphi_w^+ SJV^+_{t-4,t} + \varphi_w^- SJV^-_{t-4,t} + \varphi_m^+ SJV^+_{t-21,t} + \varphi_m^- SJV^-_{t-21,t} + \omega_{t+1}. \tag{23}$$

Based on Tauchen and Zhou (2011), Prokopczuk et al. (2016) propose HAR-ARJ model, which considers the sign of the jump that has an asymmetric impact on volatility. The MS-HAR-ARJ model is given by:

$$RV_{t+1} = c_{s_t} + \beta_{d,s_t} RV_t + \beta_w RV_{t-4,t} + \beta_m RV_{t-21,t} + \beta_{jp} RJ_t^+ + \beta_{jn} RJ_t^- + \omega_{t+1}. \tag{24}$$

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

Dumitru and Urga (2012) consider eight nonparametric jump tests and found that the ABD-LM jump test (e.g., Andersen et al., 2007b; Lee and Mykland, 2008) has the overall best performance. Moreover, Liu et al. (2016) find that jump component (henceforth JLM) calculated by ABD-LM jump test can help forecast performance and also pointed out that ABD-LM jump component's decomposition forms based on signed returns can significantly improve the models' forecasting performances. Thus, we also use those two models with regime switching, MS-HAR-RV-JLM, and MS-HAR-S-RV-J-JLM, which are expressed, respectively, as follows:

$$RV_{t+1} = c_{s_t} + \beta_{d,s_t} RV_t + \beta_w RVW_t + \beta_m RVM_t + \beta_{jp} JLM_t^+ + \beta_{jn} JLM_t^+ + \omega_{t+1}. \tag{25}$$

$$RV_{t+h} = c_{s_t} + \beta_{dp,s_t} RS_t^+ + \beta_{wp} RSW_t^+ + \beta_{mp} RSM_t^+ + \beta_{dn,s_t} RS_t^- + \beta_{wn} RSW_t^- + \beta_{mn} RSM_t^- + \beta_{jp} JLM_t^+ + \beta_{jn} JLM_t^+ + \omega_{t+h}. \tag{26}$$

Lee and Mykland (2008) assume that when a jump occurs, its size dominates r_i, i.e., when $|\varsigma_i| > c$, a jump is established at time i with size r_i[4]. Therefore, we can calculate the "significant" jump component, which is defined as the sum of intraday squared returns, as follows: $JLM_t = \sum r_i^2$. Furthermore, we decompose the JLM into positive and negative jumps based on signed returns. Consequently, we have the JLM^+ and JLM^-, and they are defined as:

$$JLM_t^+ = \sum I_{(r_{t,i}>0)} r^2_{t,\text{i}} \ \text{ and } \ JLM_t^- = \sum I_{(r_{t,i}<0)} r^2_{t,\text{i}}. \tag{27}$$

Finally, we have altogether twenty-four models including twelve HAR-RV-type models and their extended models with regime switching to model and forecast the oil futures price volatility.

3. Data description

Crude oil is an important natural source of energy. Market participants around the world not only buy and sell physical quantities of crude oil but also trade the crude oil based on the WTI oil futures contract. Moreover, most of the oil-based derivatives are priced on WTI oil futures contract, and hence it is the reference contract for most of the investors involve in energy commodities. With respect to the frequency of the data, it is found that 5-min high-frequency data has been confirmed the choice of optimal sampling frequency as it has the best trade-off between measurement accuracy and market microstructure noise (e.g., Andersen and Bollerslev, 1998; Andersen et al., 2007a, 2007b; Corsi et al., 2010; Sévi, 2014; Liu et al., 2015). Furthermore, Liu et al. (2015) compare some estimators with the RV and conclude that 5-minute RV is the best. Consequently, we select 5-minute high-frequency data of the Light Sweet Crude Oil (WTI) futures contract with a maturity of one month traded on NYMEX.[5]

Data from January 2, 2008 to April 31, 2014 are collected from the Tick DATA, and cleaned by removing the days with a shortened trading session or too few transactions, and that yields 1591 daily observations.[6] Table 1 shows different statistics of several measures. All measures including jumps and signed jump variations are significantly skewed and leptokurtic at the 1% significance level, suggesting that each measure has a distribution with a fat-tail. The Jarque-Bera statistic test further demonstrates that the normality null hypothesis is rejected at the 1% significance level. Thus, all measures are not normally distributed. The Ljung-Box statistic for correlation shows that the null hypotheses of no autocorrelation up to the 5th order are rejected for most of the series, indicating the existence of a correlation.

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

[5] Our cleaning of the data is based on Sévi (2014) and Wu et al. (2011).

[6] As detailed in Zivot and Wang (2005), for the WTI futures contract we apply a first filter to remove: (1) transactions outside the official trading period, (2) transactions with a variation of >5% in absolute value compared to the previous transaction, and (3) transactions not reported in chronological order. The same practices have been implemented in Chevallier and Sévi (2012) and Sévi (2014).

Table 1
Descriptive statistics of realized volatility, jump and signed jump variation series.

Variables	Mean	St. dev.	Skewness	Kurtosis	Jarque-Bera	Q(5)
RV	0.315	0.453	3.676***	16.607***	21,867.004***	4771.995***
CRV	0.303	0.442	3.662***	16.227***	20,892.964***	4712.197***
TCRV	0.291	0.417	3.581***	15.445***	19,214.865***	4892.561***
BPV	0.286	0.416	3.723***	17.533***	24,055.051***	4751.268***
J	0.034	0.082	7.022***	67.355***	313,821.901***	406.461***
CJ	0.012	0.061	12.155***	188.023***	2,382,769.869***	13.494**
TJ	0.024	0.101	8.135***	80.220***	444,146.317***	23.055***
JLM^+	0.051	0.178	6.579***	55.583***	216,281.962***	1721.265***
JLM^-	0.058	0.184	5.236***	32.840***	78,763.434***	2529.847***
SJV	−0.007	0.162	0.302***	27.370***	49,682.549***	5.318

Notes: We test the null hypothesis, "Skewness = 0" and "Excess Kurtosis = 3". The Jarque-Bera statistic (Jarque and Bera, 1987) tests are for the null hypothesis of normality for the distribution of the series. $Q(n)$ is the Ljung-Box statistic proposed by Ljung and Box (1978) for up to 5th order serial correlation. Asterisk *** and ** denote rejections of null hypothesis at 1% and 5% significance levels, respectively. All of realized measures, jumps, and signed jump variations series in Table 1 are multiplied by 1000.

4. Forecasting evaluation

In this section, we briefly introduce the forecasting steps for out-of-sample rolling method and forecast evaluation using the model confidence set (MCS) proposed by Hansen et al. (2011).

Our sample data are divided into two groups: 1) in-sample data for volatility modeling, covering 800 trading days from January 2, 2008 to March 3, 2011; and 2) out-of-sample data for model evaluation, covering 791 trading days from March 4, 2011 to April 30, 2014. The estimation period is 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 quantitatively evaluate the forecasting accuracy, we use two popular loss functions addressed in the literature:

$$\mathrm{HMSE} = M^{-1}\sum_{t=1}^{M}\left(1-\hat{\sigma}_t^2/RV_t\right)^2, \tag{28}$$

$$\mathrm{HMAE} = M^{-1}\sum_{t=1}^{M}\left|1-\hat{\sigma}_t^2/RV_t\right|. \tag{29}$$

where $\hat{\sigma}_t^2$ denotes the out-of-sample volatility forecast obtained from different HAR-type models, RV_t is a proxy for actual market volatility in the out-of-sample period, and M is the number of forecasting days, and in our case, $M = 791$.

The disadvantage of the abovementioned loss functions is that they do not provide any information on whether the differences of forecasting losses among models are statistically significant. Hence, we use an advanced statistical test, 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 set of models. The MCS test has several attractive advantages over conventional tests (Hansen and Lunde, 2005). First, the MCS test does not require a benchmark model to be specified. That is very useful in applications without a clear benchmark. Second, the MCS test acknowledges the limitations of the data. Third, the MCS test allows for the possibility of more than one "best" model. If some volatility models have the p-values larger than a critical value, α, those models are called "surviving" models, which have a better volatility forecasting performance than the rest of models "removed" from the MCS test. Furthermore, the larger the p-value is (closer to 1) the better the model performs. The technical details and more in-depth discussions of the MCS test can be found in Hansen et al. (2011).

5. Empirical analysis

In this section, we first present the estimations of the twelve HAR-RV-type models and their extensions with regime-switching using in-sample data. We then evaluate the forecasting performance of those models using the MCS test and robustness check.

5.1. In-sample estimation results

Table 2 exhibits the estimations of the twelve HAR-RV-type models over the in-sample period based on the Newey-West correction, which allows for correlation up to the order of 5. All of daily, weekly and monthly realized-volatility parameters of each model are significant at the 5% level, indicating a high persistence in the realized volatility dynamics. Comparing the coefficient values of different frequencies, we find that the behavior of the mid-term and long-term investors has a larger impact on the realized volatility of the crude oil market. Furthermore, different jump components have different effects on the realized volatility. For example, the jump component of HAR-RV-TJ model has a significant positive impact on the realized volatility, whereas jump components of HAR-RV-J and HAR-RV-CJ models have no significant effect. Thus, the impact of the jump components on the realized volatility is mixed. From the estimations of HAR-S-RV-J and HAR-S-RV-JLM models, we can see that the negative semi-variation has a significantly positive impact on the realized volatility, implying that the negative semi-variance contributes more to the realized volatility. The effect of the signed jump variation (SJV) on the realized volatility is significantly negative, indicating that the SJV leads to a lower future volatility. Furthermore, from the components of the SJV based on signed returns, we find that the coefficient of the negative SJV, φ_d^-, ("bad" jumps) is larger than that of positive SJV, φ_d^+, ("good" jumps). That confirms a greater effect of "bad" jumps on the realized volatility of the crude oil market. Finally, we find that the adjusted R-square of each model is >0.75, implying a higher explanation ability to the future realized volatility.

Table 3 shows the estimations of the HAR-RV-type models with regime switching. The empirical results show that the coefficient of β_{d,s_0} is slightly bigger than that of β_{d,s_1} in most of the volatility models, and all of them are significant. Moreover, estimations of semi-variances of MS-HAR-S-RV-J and MS-HAR-S-RV-JLM model reveal the asymmetric effect in different regimes, and negative semi-variances of the high volatility regime have a larger impact on future volatility. In addition, compared the values of p^{00} and p^{11} ($p^{00} > p^{11}$), we find that the high volatility regime is short-lived. Finally, we use the likelihood-ratio statistic (LR-test) to test the linearity of those models[7] and conclude that all realized volatility models significantly reject the linear hypothesis based on the linear LR-test, indicating the regime switching models are existing.

[7] Due to scarcity of the space, we do not present the procedure of the Linearity LR-test. More details of the procedure can be found in Oxmetrics 6.21 with PcGive help (§14.2.3 Testing linearity). PcGive reports a test for linearity for all outputs. The test is based on the likelihood-ratio statistic between the estimated model and the derived linear model. It also reports the approximate upper bound for the significance level of the LR statistic derived from Davies (1987) and Garcia and Perron (1996).

Table 2
In-sample period estimations of the linear HAR-type models.

	HAR-RV	HAR-RV-J	HAR-RV-CJ	HAR-RV-TJ	HAR-S-RV-J	HAR-RV-PS2	HAR-RV-PS3	HAR-CSJ	HAR-CSJd	HAR-ARJ	HAR-RV-JLM	HAR-S-RV-J-JLM
β_d	0.162***	0158***	0.141**	0.152***						0.156***	0.181**	
β_{bpv}						0.139 ***	0.124***	0.138**	0.107**			
β_{dp}					0.016							0.043
β_{dn}					0.112***							0.112***
β_w	0.486***	0.487***	0.481***	0.493***		0.498***	0.496***	0.455***	0.543***	0.485***	0.474***	
β_{wp}					0.194**							0.191**
β_{wn}					0.244***							0.239***
β_m	0.320***	0.320***	0.306***	0.301***		0.326***	0.323***	0.352***	0.284***	0.324***	0.331***	
β_{mp}					−0.132							−0.134***
β_{mn}					0.515***							0.0523***
β_j		57.446	164.821	281.884***	104.751							
β_{jp}										−30.702	−208.600**	−114.090
β_{jn}										188.385	97.017	16.563
β_{wj}			660.823									
β_{mj}			1227.823									
φ_d						−177.651***		−152.623**				
φ_w								123.134				
φ_m								−1037.345***				
φ_d^+							−27.505		101.143			
φ_w^+									−718.978*			
φ_m^+									−480.804			
φ_d^-							−332.789***		−431.529**			
φ_w^-									920.492			
φ_m^-									−1735.232***			
Normality test	16.052***	15.722***	15.829***	13.907***	15.799***	11.084***	13.490***	13.178***	15.648***	15.702***	9.845***	11.748***
Adj.R^2	0.750	0.750	0.751	0.748	0.754	0.751	0.751	0.753	0.754	0.750	0.751	0.755

Notes: We have the null hypothesis that coefficients are zeros, for example, "$\beta_d = 0$", "$\beta_w = 0$" and so on. Asterisk ***, ** and * denote rejections of null hypothesis at 1%, 5% and 10% significance levels, respectively. We do not report the constant coefficients in this table.

Table 3
In-sample period estimations of the HAR-type models with regime switching.

	HAR-RV	HAR-RV-J	HAR-RV-CJ	HAR-RV-TJ	HAR-S-RV-J	HAR-RV-PS2	HAR-RV-PS3	HAR-CSJ	HAR-CSJd	HAR-ARJ	HAR-RV-JLM	HAR-S-RV-J-JLM
β_{d,s_0}	0.159***	0.150***	0.127***	0.142***						0.154***	0.167***	
β_{d,s_1}	0.155***	0.146***	0.120***	0.127***						0.150***	0.162***	
β_{bpv,s_0}						0.130***	0.111**	0.127***	0.103***			
β_{bpv,s_1}						0.126***	0.106**	0.124***	0.101***			
β_{dp,s_0}					0.028							0.045
β_{dp,s_1}					−0.055							−0.041
β_{dn,s_0}					0.093**							0.094**
β_{dn,s_1}					0.169							0.175***
β_w	0.466***	0.468***	0.464***	0.495***		0.482***	0.475***	0.423***	0.491***	0.465***	0.457***	
β_{wp}					0.174**							0.172**
β_{wn}					0.249**							0.244***
β_m	0.350***	0.351***	0.342***	0.325***		0.358*	0.356***	0.398***	0.347***	0.353***	0.356***	
β_{mp}					−0.078							−0.086
β_{mn}					0.493***							0.500***
β_j		103.412	157.441	194.340	161.915							
β_{jp}										−35.631	−137.579	−37.494
β_{jn}										144.417	100.850	16.324
β_{wj}			998.656***									
β_{mj}			965.855***									
γ												
φ_d						- 179.500 ***		−172.302*				
φ_w								137.213				
φ_m								−1198.48***				
φ_d^+							−7.678		32.055			
φ_w^+									−405.062**			
φ_m^+									−863.94***			
φ_d^-							−387.995***		−405.062**			
φ_w^-									757.772**			
φ_m^-									−1796.08**			
σ_0	0. 374***	0. 374***	0. 372***	0. 383***	0. 371***	0. 368***	0. 368***	0. 366***	0. 365***	0. 373***	0. 370***	0. 371***
σ_1	0.633***	0.634***	0.619***	0.638***	0.624***	0.600***	0.608***	0.602***	0.609***	0.634***	0.612***	0.619***
P^{00}	0.929***	0.930***	0.922***	0.932***	0.930***	0.913***	0.915***	0.921***	0.922***	0.930***	0.920***	0.929***
P^{11}	0.282***	0.282***	0.284***	0.326***	0.286***	0.258***	0.273***	0.243***	0.258***	0.277***	0.270***	0.281***
AIC	1.190	1.192	1.195	1.202	1.181	1.192	1.191	1.186	1.186	1.191	1.192	1.183
Linearity LR-test	18.716***	18.733***	17.287***	16.384***	18.580***	16.171***	17.017***	16.899***	16.647***	19.086***	16.140***	17.264***

Notes: We have the null hypothesis that coefficients are zeros, for example, "$\beta_d = 0$", "$\beta_w = 0$" and so on. Asterisk ***, ** and * denote rejections of null hypothesis at 1%, 5% and 10% significance levels, respectively. We do not report the constant coefficients in this table. AIC is the Akaike information criterion.

Table 4
Forecasting performances of linear HAR-RV-type models with MCS test.

Models	HMSE		HMAE		QLIKE	
	T_R	T_{SQ}	T_R	T_{SQ}	T_R	T_{SQ}
HAR-RV	**0.8250**	**0.8414**	**0.9674**	**0.9912**	**0.3627**	**0.4549**
HAR-RV-J	**0.9306**	**0.9306**	**0.9978**	**0.9993**	**0.4775**	**0.7087**
HAR-RV-CJ	**0.8250**	**0.6408**	**0.9674**	**0.9912**	**0.7931**	**0.8110**
HAR-RV-TJ	**1.0000**	**1.0000**	**0.9995**	**0.9996**	**0.9811**	**0.9948**
HAR-S-RV-J	**0.8250**	**0.6408**	**0.9995**	**0.9996**	**1.0000**	**1.0000**
HAR-RV-PS2	**0.3736**	**0.5771**	**0.9995**	**0.9996**	**0.9907**	**0.9955**
HAR-RV-PS3	**0.3736**	**0.5771**	**0.9497**	**0.9592**	**0.9859**	**0.9955**
HAR-CSJ	**0.3736**	**0.5078**	**0.9978**	**0.9990**	**0.9859**	**0.9955**
HAR-CSJd	**0.8250**	**0.5771**	**1.0000**	**1.0000**	**0.9907**	**0.9955**
HAR-ARJ	**0.8761**	**0.8890**	**0.9995**	**0.9996**	**0.9570**	**0.8698**
HAR-RV-JLM	**0.3736**	**0.5458**	**0.6977**	**0.7758**	**0.4775**	**0.4921**
HAR-S-RV-J-JLM	**0.3736**	**0.4401**	**0.9308**	**0.9210**	**0.9724**	**0.9770**

Notes: MCS p-values are calculated according to the test statistics T_R and T_{SQ}. Values of $p > 0.25$ are indicated in bold.

5.2. Out-of-sample forecasting evaluation

5.2.1. Forecasting performances of individual linear models

Wang et al. (2016) find that the in-sample predictive relationships are not constant but change over time. Compared with the in-sample performance, the out-of-sample performance of a model (i.e., its predictive ability) is more important to market participants, because they are more concerned about the model's ability to predict the future than its ability to analyze past. Many studies (e.g., Andersen et al., 2003, Andersen et al., 2007a, 2007b; Corsi, 2009; Pu et al., 2016; Wang et al., 2015) have found that the realized volatility exhibits heavy tails, high peak, and non-normality. Also, Aït-Sahalia and Mancini (2008) point out that the distribution of log-RV can be closer to Gaussian than that of RV. Furthermore, to ensure the volatility values are nonnegative, the twelve models and their extended regime switching models are log transformed.

In our analysis, we use the forecasting evaluation—model confidence set (MCS) proposed by Hansen et al. (2011) —to assess the statistical significance of the twelve HAR-RV-type models and their extended regime switching models. Based on Martens et al. (2009), Hansen et al. (2011), and Laurent et al. (2012), we set the confidence level α of 0.25, which means that if the p-value obtained from the MCS test is smaller than 0.25, we can exclude the corresponding model from the set. In other word, the forecasting performance of that model is significantly worse than that of other models in MCS test. The p-values are obtained from 10,000 bootstraps with 2 block length.[8] Table 4 reports the MCS test results of the twelve HAR-RV-type models with forecasting window of 791 days. From both HMSE and HMAE loss functions, we can conclude that the p-values of the twelve HAR-RV-type models are larger than 0.25, showing that the twelve volatility models have better performances in forecasting the volatility of the crude oil.

5.2.2. Forecasting performances of Markov-regimes models

Table 5 reports the MCS test results from the twelve linear HAR-RV-type models and their extensions of including a two-regime switching model. Under the HMSE and HMAE loss functions, the results show that the linear models cannot survive in the MCS test, indicating that linear models have a worse predicting ability than the models including regime switching. In addition, MS-HAR-RV-PS2, MS-HAR-RV-PS3, MS-HAR-CSJ, MS-HAR-CSJd, MS-HAR-RV-JLM and MS-S-HAR-RV-JLM models cannot survive in MCS test at 75% significant level. However, under the HMAE loss function, the models with regime switching almost pass the MCS test, indicating that inclusion of a two-regime switching model can gain a higher accuracy. Finally, under the both loss functions, most of the HAR-RV-type models with regime switching

[8] We choose different block lengths and can get the same empirical results.

Table 5
Forecasting performances of linear models and their regime switching models with MCS test.

Models	HMSE		HMAE		QLIKE	
	T_R	T_{SQ}	T_R	T_{SQ}	T_R	T_{SQ}
HAR-RV	0.1207	0.0278	0.0042	0.0021	0.0042	0.0021
HAR-RV-J	0.1207	0.0278	0.0042	0.0095	0.0042	0.0095
HAR-RV-CJ	0.1207	0.0278	0.0042	0.0042	0.0042	0.0042
HAR-RV-TJ	0.1207	0.1371	0.0111	0.0610	0.0111	0.0610
HAR-S-RV-J	0.1207	0.0278	0.0042	0.0063	0.0042	0.0063
HAR-RV-PS2	0.1207	0.0278	0.0042	0.0008	0.0042	0.0008
HAR-RV-PS3	0.1207	0.0278	0.0042	0.0042	0.0042	0.0042
HAR-CSJ	0.1207	0.0278	0.0042	0.0008	0.0042	0.0008
HAR-CSJd	0.1207	0.0278	0.0082	0.0419	0.0082	0.0419
HAR-ARJ	0.1207	0.0278	0.0042	0.0009	0.0042	0.0009
HAR-RV-JLM	0.1207	0.0278	0.0042	0.0008	0.0042	0.0008
HAR-S-RV-J-JLM	0.1207	0.0278	0.0042	0.0008	0.0042	0.0008
MS-HAR-RV	**0.7763**	**0.8853**	**0.9294**	**0.8938**	**0.9294**	**0.8938**
MS-HAR-RV-J	**0.9931**	**0.9931**	**0.9830**	**0.9829**	**0.9830**	**0.9829**
MS-HAR-RV-CJ	**0.8586**	**0.8905**	**0.9830**	**0.9829**	**0.9830**	**0.9829**
MS-HAR-RV-TJ	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**
MS-HAR-S-RV-J	**0.5384**	**0.4715**	**0.9830**	**0.9829**	**0.9830**	**0.9829**
MS-HAR-RV-PS2	0.1207	0.0278	0.0111	0.0610	0.0111	0.0610
MS-HAR-RV-PS3	0.1207	0.0278	**0.4006**	**0.6630**	**0.4006**	**0.6630**
MS-HAR-CSJ	0.1207	0.0278	**0.4006**	**0.5703**	**0.4006**	**0.5703**
MS-HAR-CSJd	0.1207	0.0278	**0.9466**	**0.9279**	**0.9466**	**0.9279**
MS-HAR-ARJ	**0.8586**	**0.8905**	**0.9830**	**0.9829**	**0.9830**	**0.9829**
MS-HAR-RV-JLM	0.1207	0.0278	**0.4006**	**0.5240**	**0.4006**	**0.5240**
MS-HAR-S-RV-J-JLM	0.1207	0.0278	**0.4006**	**0.6195**	**0.4006**	**0.6195**

Notes: MCS p-values are calculated according to the test statistics T_R and T_{SQ}. Values of $p > 0.25$ are indicated in bold.

improve their forecasting performance significantly. Our results further convince the view of Nomikos and Pouliasis (2011) that point out that a regime-switching model may be more suitable for modeling volatility, particularly in energy markets, because of structural breaks in those market are quite common.

5.2.3. Robustness check

Market microstructure noise plays an important role. Although Andersen and Bollerslev (1998), Madhavan (2000), Biais et al. (2005), Liu et al. (2015), and Wang et al. (2016) have used 5-min return data to mitigate the effects of the noise, Zhang et al. (2005) show that this is not an adequate solution to the problem. Therefore, we take an alternative realized measurement called realized kernel (RK), which is robust to noise (Barndorff-Nielsen et al., 2008; Wang et al., 2015), to examine whether our results are robust. RK is defined as below:

$$RK_t = \sum_{j=-H}^{H} k\left(\frac{H}{H+1}\right)\eta_j, \eta_j = \sum_{i=|j|+1}^{n} r_{t,i} r_{t,i-|j|}. \tag{30}$$

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

$$k(x) = \begin{cases} 1-6x^2+6x^3, & 0 \le x \le 1/2 \\ 2(1-x)^3, & 1/2 \le x \le 1 \\ 0, & x > 1. \end{cases} \tag{31}$$

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 spelled out in Barndorff-Nielsen et al. (2009). Table 6 shows the MCS test results of each model according to RK benchmark. Under the two loss functions, the results indicate that almost all models with regime switching can gain a higher forecast accuracy than linear HAR-RV-type models. This indicates that inclusion of a two-regime switching model in HAR-RV-type models can significantly improve the forecasting performance.

Rossi and Inoue (2012) argue that arbitrary choices of window sizes have consequences about how the sample is split into in-sample and out-of-sample portions. Although choosing an appropriate forecasting

Table 6
Forecasting performances of linear models and their regime switching models with MCS test based on RK benchmark.

Models	HMSE		HMAE	
	T_R	T_{SQ}	T_R	T_{SQ}
HAR-RV	**0.6319**	0.1010	0.0297	0.0197
HAR-RV-J	**0.6494**	0.1807	0.0392	0.0815
HAR-RV-CJ	**0.6319**	0.1010	0.0297	0.0234
HAR-RV-TJ	**0.6494**	**0.2907**	0.0971	0.2291
HAR-S-RV-J	**0.6494**	0.1010	0.0392	0.0689
HAR-RV-PS2	**0.6319**	0.0970	0.0297	0.0113
HAR-RV-PS3	**0.6280**	0.0917	0.0297	0.0165
HAR-CSJ	**0.6315**	0.0961	0.0297	0.0183
HAR-CSJd	**0.6315**	0.0961	0.0392	0.0605
HAR-ARJ	**0.6319**	0.1010	0.0392	0.0289
HAR-RV-JLM	**0.6280**	0.0916	0.0297	0.0113
HAR-S-RV-J-JLM	**0.6319**	0.0970	0.0297	0.0184
MS-HAR-RV	**0.6494**	0.2907	0.9326	0.8552
MS-HAR-RV-J	**0.7573**	**0.5976**	**0.9326**	**0.8704**
MS-HAR-RV-CJ	**0.7573**	**0.5783**	**0.9326**	**0.8704**
MS-HAR-RV-TJ	**1.0000**	**1.0000**	**1.0000**	**1.0000**
MS-HAR-S-RV-J	**0.6319**	0.1010	**0.8096**	**0.8552**
MS-HAR-RV-PS2	**0.6319**	0.1010	**0.9326**	**0.8552**
MS-HAR-RV-PS3	**0.6319**	0.1010	**0.8096**	**0.8552**
MS-HAR-CSJ	**0.6319**	0.0970	**0.8096**	**0.8549**
MS-HAR-CSJd	**0.6319**	0.1010	**0.9326**	**0.8552**
MS-HAR-ARJ	**0.7573**	**0.5355**	**0.9326**	**0.8704**
MS-HAR-RV-JLM	**0.6319**	0.1010	**0.8096**	**0.8552**
MS-HAR-S-RV-J-JLM	**0.6319**	0.0970	**0.9326**	**0.8552**

Notes: MCS *p*-values are calculated according to the test statistics T_R and T_{SQ}. Values of $p > 0.25$ are indicated in bold.

window is crucial to the forecasting performance of the models, there seems to be no consensus on how to choose the right forecasting windows. Thus, we also report the results in Table 7 obtained from other forecasting windows, such as 691 and 591. The result further validates that our finding is robust, indicating that inclusion of a two-regime switching model in HAR-RV-type models provides a powerful forecasting ability in the crude oil markets. Therefore, we can draw a conclusion that, in forecasting of the crude oil market, HAR-RV-type models with nonlinear regime-switching models have a better performance than the linear realized volatility models alone.

Table 8
Forecasting performances of linear models and their regime switching models with MCS test.

Models	HMSE		HMAE	
	T_R	T_{SQ}	T_R	T_{SQ}
HAR-RV	**0.6319**	0.1010	0.0297	0.0197
HAR-RV-J	**0.6494**	0.1807	0.0392	0.0815
HAR-RV-CJ	**0.6319**	0.1010	0.0297	0.0234
HAR-RV-TJ	**0.6494**	**0.2907**	0.0971	0.2291
HAR-S-RV-J	**0.6494**	0.1010	0.0392	0.0689
HAR-RV-PS2	**0.6319**	0.0970	0.0297	0.0113
HAR-RV-PS3	**0.6280**	0.0917	0.0297	0.0165
HAR-CSJ	**0.6315**	0.0961	0.0297	0.0183
HAR-CSJd	**0.6315**	0.0961	0.0392	0.0605
HAR-ARJ	**0.6319**	0.1010	0.0392	0.0289
HAR-RV-JLM	**0.6280**	0.0916	0.0297	0.0113
HAR-S-RV-J-JLM	**0.6319**	0.0970	0.0297	0.0184
MSA-HAR-RV	**0.6494**	**0.2907**	**0.9326**	**0.8552**
MSA-HAR-RV-J	**0.7573**	**0.5976**	**0.9326**	**0.8704**
MSA-HAR-RV-CJ	**0.7573**	**0.5783**	**0.9326**	**0.8704**
MSA-HAR-RV-TJ	**1.0000**	**1.0000**	**1.0000**	**1.0000**
MSA-HAR-S-RV-J	**0.6319**	0.1010	**0.8096**	**0.8552**
MSA-HAR-RV-PS2	**0.6319**	0.1010	**0.9326**	**0.8552**
MSA-HAR-RV-PS3	**0.6319**	0.1010	**0.8096**	**0.8552**
MSA-HAR-CSJ	**0.6319**	0.0970	**0.8096**	**0.8549**
MSA-HAR-CSJd	**0.6319**	0.1010	**0.9326**	**0.8552**
MSA-HAR-ARJ	**0.7573**	**0.5355**	**0.9326**	**0.8704**
MSA-HAR-RV-JLM	**0.6319**	0.1010	**0.8096**	**0.8552**
MSA-HAR-S-RV-J-JLM	**0.6319**	0.0970	**0.9326**	**0.8552**

Notes: MCS *p*-values are calculated according to the test statistics T_R and T_{SQ}. Values of $p > 0.25$ are indicated in bold. We use "MSA" to represent the model having all variables with regime switching.

6. Further analysis

In this section, following previous studies (e.g., Ma et al., 2015; Wang et al., 2016), we consider high and low volatility regimes for all variables of volatility models, such as the average weekly and monthly RV, jump components, and signed jump variations. Because of insufficient space, we only report the forecasting results in Table 8. The empirical results reinforce our finding that HAR-RV-type models with switching regime can indeed improve the forecasting performance.

Table 7
Forecasting performances of linear models and their regime switching models with MCS test.

Models	$M = 691$				$M = 591$			
	HMSE		HMAE		HMSE		HMAE	
	T_R	T_{SQ}	T_R	T_{SQ}	T_R	T_{SQ}	T_R	T_{SQ}
HAR-RV	0.1463	0.0278	0.0005	0.0000	0.0043	0.0105	0.0001	0.0000
HAR-RV-J	0.1555	0.0408	0.0005	0.0001	0.0043	0.0088	0.0000	0.0000
HAR-RV-CJ	0.1463	0.0225	0.0232	0.0009	0.1524	0.0203	0.0441	0.0000
HAR-RV-TJ	**0.5422**	0.2033	0.1660	0.0999	**0.4197**	0.0761	**0.5159**	0.0580
HAR-S-RV-J	0.1463	0.0215	0.0232	0.0123	0.0043	0.0094	0.0441	0.0026
HAR-RV-PS2	0.1463	0.0215	0.0005	0.0000	0.0043	0.0088	0.0001	0.0000
HAR-RV-PS3	0.1555	0.0408	0.0232	0.0060	0.1524	0.0352	0.0441	0.0000
HAR-CSJ	0.1463	0.0213	0.0005	0.0001	0.0043	0.0088	0.0441	0.0001
HAR-CSJd	0.1463	0.0217	0.0232	0.0370	0.1524	0.0455	0.0441	0.0006
HAR-ARJ	0.1463	0.0217	0.0005	0.0000	0.0043	0.0086	0.0000	0.0000
HAR-RV-JLM	0.1463	0.0205	0.0005	0.0000	0.0043	0.0088	0.0001	0.0000
HAR-S-RV-J-JLM	0.1463	0.0211	0.0232	0.0028	0.1524	0.0158	0.0441	0.0105
MS-HAR-RV	**0.7300**	**0.3402**	**0.8628**	**0.7233**	**0.4197**	0.0761	**0.5159**	0.0580
MS-HAR-RV-J	**0.7300**	**0.3402**	**0.8907**	**0.7233**	**0.4197**	0.0761	**0.5159**	0.0580
MS-HAR-RV-CJ	**0.7300**	**0.3402**	**0.8907**	**0.7233**	**0.4197**	0.0761	**0.5159**	0.2083
MS-HAR-RV-TJ	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**
MS-HAR-S-RV-J	**0.5422**	0.2033	**0.8907**	**0.7233**	**0.4197**	0.0761	**0.5159**	0.2083
MS-HAR-RV-PS2	**0.5422**	0.2033	0.1660	0.0999	**0.4197**	0.0761	**0.5159**	0.2083
MS-HAR-RV-PS3	0.1463	**0.0278**	0.1660	0.0961	**0.4197**	0.0761	**0.5159**	0.0533
MS-HAR-CSJ	**0.5422**	**0.2033**	**0.8907**	**0.7233**	**0.4197**	0.0761	**0.5159**	0.2083
MS-HAR-CSJd	**0.5422**	**0.2033**	**0.8907**	**0.7233**	**0.4197**	0.0761	**0.5159**	0.0580
MS-HAR-ARJ	**0.7300**	**0.3402**	**0.8628**	**0.7233**	**0.4197**	0.0761	**0.5159**	0.2083
MS-HAR-RV-JLM	**0.7300**	**0.3402**	**0.8628**	**0.7111**	**0.4197**	0.0761	**0.5159**	0.2083
MS-HAR-S-RV-J-JLM	0.1463	0.0278	**0.8907**	**0.7233**	**0.4197**	0.0761	**0.5159**	0.2083

Notes: MCS *p*-values are calculated according to the test statistics T_R and T_{SQ}. Values of $p > 0.25$ are indicated in bold.

Therefore, incorporating regime-switching behavior of realized volatility, jump components, and other relevant variables of volatility models can help improve modeling and forecasting the future volatility in statistical aspect.

7. Concluding remarks

An increasing amount of available ultra-high frequency data encourages researchers, market participants, and policymakers to use such data in modeling and forecasting the realized volatility. In this study, we forecast the realized volatility of the crude oil futures using the 5-min "rule-of-thumb" data. To this end, we use HAR-RV-type models and their regime switching models. In-sample results show that the negative semi-variation has a significantly positive impact on the realized volatility, implying that the negative semi-variance contributes more to the realized volatility. Also, the high volatility regime is short-lived. Evaluation of forecasting using the MCS test shows that HAR-RV-type models with regime switching lead to gain greater forecast accuracy. The same result holds for another volatility benchmark and different forecasting windows. Moreover, incorporating regime-switching behavior to all variables in HAR-RV-type models, we find that those models also yield a higher forecast accuracy. In this research, we consider fixed transition probabilities in Markov regime-switching models. Hence we will explore time-varying transition probabilities in one of our future works.

Acknowledgement

The authors are grateful to the Editors and the three anonymous referees for providing thoughtful comments and suggestions that enhanced the quality of this paper. The authors are also grateful to the Natural Science Foundation of China [71371157, 71671145; 71701170], the humanities and social science fund of the ministry of education [17YJC790105], and Fundamental Research Funds for the Central Universities [26816WCX02, 2682017WCX01].

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.eneco.2017.08.004.

References

Agnolucci, P., 2009. Volatility in crude oil futures: a comparison of the predictive ability of GARCH and implied volatility models. Energy Econ. 31, 316–321.

Aït-Sahalia, Y., Mancini, L., 2008. Out of sample forecasts of quadratic variation. J. Econ. 147 (1), 17–33.

Aloui, C., Jammazi, R., 2009. The effects of crude oil shocks on stock market shifts behaviour: a regime switching approach. Energy Econ. 31 (5), 789–799.

Andersen, T.G., Bollerslev, T., 1998. Answering the skeptics: yes, standard volatility models do provide accurate forecasts. Int. Econ. Rev. 39, 885–905.

Andersen, T.G., Bollerslev, T., Diebold, F.X., et al., 2003. Modeling and forecasting realized volatility. Econometrica 71, 579–625.

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, 701–720.

Andersen, T.G., Bollerslev, T., Dobrev, D., 2007b. No-arbitrage semi-martingale restrictions for continuous-time volatility models subject to leverage effects, jumps and I.I.D. noise: theory and testable distributional implications. J. Econ. 138, 125–180.

Banerjee, A., Urga, G., 2005. Modelling structural breaks, long memory and stock market volatility: an overview. J. Econ. 129, 1–34.

Barndorff-Nielsen, O.E., Shephard, N., 2004. Power and bipower variation with stochastic volatility and jumps. J. Financ. Econ. 2, 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, 1481–1536.

Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A., Shephard, N., 2009. Realized kernels in practice: trades and quotes. Econ. J. 12, 1–32.

Barndorff-Nielsen, O.E., Kinnebrock, S.K., Shephard, N., 2010. Measuring downside risk: realised semivariance. In: Bollerslev, T., Russell, J., Watson, M. (Eds.), Volatility and Time Series Econometrics: Essays in Honor of Robert F. Engle. Oxford University Press, New York, pp. 117–137.

Bekaert, G., Hoerova, M., 2014. The VIX, the variance premium and stock market volatility. J. Econ. 183, 181–192.

Bekaert, G., Engstrom, E., Ermolov, A., 2015. Bad environments, good environments: a non-Gaussian asymmetric volatility model. J. Econ. 186 (1), 258–275.

Biais, B., Glosten, L., Spatt, C., 2005. Market microstructure: a survey of microfoundations, empirical results, and policy implications. J. Financ. Mark. 8, 217–264.

Bollersleva, T., Patton, A.J., Quaedvlieg, R., 2015. Exploiting the errors: a simple approach for improved volatility forecasting. J. Econ. 192, 1–18.

Chen, X., Ghysels, E., 2011. News—good or bad—and its impact on volatility predictions over multiple horizons. Rev. Financ. Stud. 24, 46–81.

Cheong, C.W., 2009. Modeling and forecasting crude oil markets using ARCH-type models. Energy Policy 37, 2346–2355.

Chevallier, J., Sévi, B., 2012. On the volatility–volume relationship in energy futures markets using intraday data. Energy Econ. 34, 1896–1909.

Corsi, F., 2009. A simple approximate long-memory model of realized volatility. J. Financ. Econ. 7, 174–196.

Corsi, F., Pirino, D., Reno, R., 2010. Threshold bipower variation and the impact of jumps on volatility forecasting. J. Econ. 159, 276–288.

Davies, R.B., 1987. Hypothesis testing when a nuisance parameter is present only under the alternative. Biometrika 74, 33–43.

Degiannakis, S., 2008. ARFIMAX and ARFIMAX-TARCH realized volatility modeling. J. Appl. Stat. 35, 1169–1180.

Degiannakis, S., George, F., 2016. Forecasting Oil Price Realized Volatility: A New Approach, 2016, Working Paper.

Dumitru, A., Urga, G., 2012. Identifying jumps in financial assets: a comparison between nonparametric jump tests. J. Bus. Econ. Stat. 30, 242–255.

Duong, D., Swanson, N.R., 2015. Empirical evidence on the importance of aggregation, asymmetry, and jumps for volatility prediction. J. Econ. 187, 606–621.

Efimova, O., Serletis, A., 2014. Energy markets volatility modelling using GARCH. Energy Econ. 43, 264–273.

Fong, W.M., See, K.H., 2002. A Markov-switching model of the conditional volatility of crude oil prices. Energy Econ. 24 (1), 71–95.

Fong, W.M., See, K.H., 2003. Basis variations and regime-shifts in the oil futures market. Eur. J. Financ. 9 (5), 499–513.

Garcia, R., Perron, P., 1996. An analysis of the real interest rate under regime shifts. Rev. Econ. Stat. 111–125.

Goldman, E., Nam, J., Tsurumi, H., et al., 2013. Regimes and long memory in realized volatility. Stud. Nonlinear Dyn. Econ. 17, 521–549.

Granger, C.W., Ding, Z., 1996. Varieties of long memory models. J. Econ. 73 (1), 61–77.

Hamilton, J.D., 1983. Oil and the macroeconomy since World War II. J. Polit. Econ. 91 (2), 228–248.

Hamilton, J.D., 1990. Analysis of time series subject to changes in regime. J. Econ. 45, 39–70.

Hamilton, J.D., 2003. What is an oil shock? J. Econ. 113 (2), 363–398.

Hamilton, J.D., Susmel, R., 1994. Autoregressive conditional heteroskedasticity and changes in regime. J. Econ. 64, 307–333.

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. Econ. 3, 456–499.

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, 119–125.

Kilian, L., Park, C., 2009. The impact of oil price shocks on the US stock market. Int. Econ. Rev. 50 (4), 1267–1287.

Kim, C.J., 1994. Dynamic linear models with Markov-switching. J. Econ. 60, 1–22.

Kim, C.J., Nelson, C.R., 1999. State-space Models With Regime Switching: Classical and Gibbs-sampling Approaches With Applications. MIT Press Books, p. 1.

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, 445–475.

Laurent, S., Rombouts, J.V., Violante, F., 2012. On the forecasting accuracy of multivariate GARCH models. J. Appl. Econ. 27, 934–955.

Lee, S.S., Mykland, P.A., 2008. Jumps in financial markets: a new nonparametric test and jump dynamics. Rev. Financ. Stud. 21, 2535–2563.

Liu, L., Wan, J., 2012. A study of Shanghai fuel oil futures price volatility based on high-frequency data: long-range dependence, modeling and forecasting. Econ. Model. 29, 2245–2253.

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, 293–311.

Liu, Z., Ma, F., Wang, X., et al., 2016. Forecasting the realized volatility: the role of jumps. Appl. Econ. Lett. 23, 736–739.

Ljung, G.M., Box, G.E., 1978. On a measure of lack of fit in time series models. Biometrika 65, 297–303.

Longin, F.M., 1997. The threshold effect in expected volatility: a model based on asymmetric information. Rev. Financ. Stud. 10 (3), 837–869.

Ma, F., Liu, L., Liu, Z., et al., 2015. Forecasting realized range volatility: a regime-switching approach. Appl. Econ. Lett. 22, 1361–1365.

Madhavan, A., 2000. Market microstructure: a survey. J. Financ. Mark. 3, 205–258.
Martens, M., Van, D.D., De, P.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, 282–303.
Mohammadi, H., Su, L., 2010. International evidence on crude oil price dynamics: applications of ARIMA-GARCH models. Energy Econ. 32, 1001–1008.
Nomikos, N., Andriosopoulos, K., 2012. Modelling energy spot prices: empirical evidence from NYMEX. Energy Econ. 34, 1153–1169.
Nomikos, N.K., Pouliasis, P.K., 2011. Forecasting petroleum futures markets volatility: the role of regimes and market conditions. Energy Econ. 33, 321–337.
Patton, A.J., Sheppard, K., 2015. Good volatility, bad volatility: signed jumps and the persistence of volatility. Rev. Econ. Stat. 97, 683–697.
Perlin, M., 2015. MS_Regress-The MATLAB Package for Markov Regime Switching Models. SSRN https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1714016.
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.
Raggi, D., Bordignon, S., 2012. Long memory and nonlinearities in realized volatility: a Markov switching approach. Comput. Stat. Data Anal. 56, 3730–3742.
Rossi, B., Inoue, A., 2012. Out-of-sample forecast tests robust to the choice of window size. J. Bus. Econ. Stat. 30, 432–453.
Sévi, B., 2014. Forecasting the volatility of crude oil futures using intraday data. Eur. J. Oper. Res. 235, 643–659.
Shi, Y., Ho, K.Y., 2015. Long memory and regime switching: a simulation study on the Markov regime-switching ARFIMA model. J. Bank. Financ. 61, 189–204.
Tauchen, G., Zhou, H., 2011. Realized jumps on financial markets and predicting credit spreads. J. Econ. 160, 102–118.
Wahab, M.I.M., Lee, C.-G., 2009. A lattice approach to pricing of multivariate contingent claims with regime switching. J. Deriv. 17, 49–61.
Wang, Y., Wu, C., 2012. Forecasting energy market volatility using GARCH models: can multivariate models beat univariate models? Energy Econ. 34 (6), 2167–2181.
Wang, X., Wu, C., Xu, W., 2015. Volatility forecasting: the role of lunch-break returns, overnight returns, trading volume and leverage effects. Int. J. Forecast. 31, 609–619.
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.S., 2010. Forecasting crude oil market volatility: further evidence using GARCH-class models. Energy Econ. 32, 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.
Wu, F., Guan, Z., Myers, R.J., 2011. Volatility spillover effects and cross-hedging in corn and crude oil futures. J. Futur. Mark. 31 (11), 1052–1075.
Zhang, L., Mykland, P.A., Ait-Sahalia, Y., 2005. A tale of two-time scales. J. Am. Stat. Assoc. 100, 1394–1411.
Zivot, E., Wang, J., 2005. Modelling Financial Time Series With S-PLUS. Second editions. Springer.