

## RESEARCH ARTICLE

# Forecasting realized volatility of oil futures market: A new insight

Feng Ma<sup>1</sup>  | Yu Wei<sup>2</sup> | Li Liu<sup>3</sup> | Dengshi Huang<sup>1</sup>

<sup>1</sup> School of Economics and Management, Southwest Jiaotong University, Chengdu, China

<sup>2</sup> School of Finance, Yunnan University of Finance and Economics, Kunming, China

<sup>3</sup> School of Finance, Nanjing Audit University, Nanjing, China

## Correspondence

Yu Wei, School of Finance, Yunnan University of Finance and Economics, Kunming, China.

Email: weiyusy@126.com

## Funding information

Natural Science Foundation of China, Grant/Award Numbers: 71771124, 71671145, 71371157 and 71701170

## Abstract

In this study we propose several new variables, such as continuous realized semi-variance and signed jump variations including jump tests, and construct a new heterogeneous autoregressive model for realized volatility models to investigate the impacts that those new variables have on forecasting oil price volatility. In-sample results indicate that past negative returns have greater effects on future volatility than that of positive returns, and our new signed jump variations have a significantly negative influence on the future volatility. Out-of-sample empirical results with several robust checks demonstrate that our proposed models can not only obtain better performance in forecasting volatility but also garner larger economic values than can the existing models discussed in this paper.

## KEYWORDS

oil futures market, realized semi-variances, signed jump variations, statistic and economic evaluations, volatility forecasting

## 1 | INTRODUCTION

Crude oil plays an essential role in the world economy. Oil price uncertainty has important macroeconomic effects (Hamilton, 1983, 2003; Kilian, 2009) and effects on financial markets (Aloui & Jammazi, 2009; Kilian & Park, 2009). Oil price volatility is a key input for risk management, derivative pricing, portfolio selection, and many other financial activities. Therefore, modeling and forecasting the volatility of crude oil prices are critical for researchers, market participants, and policymakers.

There are many works on forecasting oil price volatility in the frameworks of generalized autoregressive conditional heteroskedasticity (GARCH) and its various extensions (see, e.g., Agnolucci, 2009; Charles & Darné, 2014; Efimova & Serletis, 2014; Nomikos & Poulisis, 2011; Wei, Wang, & Huang, 2010). However, these models are always applied to daily or lower-frequency data, which can result in a substantial loss of intraday trading information (Carnero, Peña, & Ruiz, 2004; Corsi, 2009).

Recently, with the availability of high-frequency data, research on financial market volatility has taken new avenues. The seminal works of Andersen and Bollerslev (1998) and Andersen, Bollerslev, Diebold, and Ebens (2001) propose the realized volatility or realized variance (RV), which are defined as the sum of all available intraday high-frequency squared returns. This volatility measure can enable researchers to better gauge the current level of volatility and understand its dynamics. Corsi (2009) propose a simple heterogeneous autoregressive model of realized volatility (HAR-RV) based on the heterogeneous market hypothesis, which can capture “stylized facts” in financial market volatility, such as long memory and multiscaling behavior, and have other advantages in forecasting. Thus this model has garnered wide focus in academia (see, e.g., Andersen, Bollerslev, & Diebold, 2007; Bekaert & Hoerova, 2014; Bollerslev, Patton, & Quaedvlieg, 2015; Busch, Christensen, & Nielsen, 2011; Corsi, Pirino, & Reno, 2010; Duong & Swanson, 2015; Wang, Ma, Wei, & Wu, 2016; Wang, Pan, & Wu, 2017).

Nevertheless, the HAR-RV model is of great interest to us here because there is an extremely limited strand of the literature focusing on the oil price-realized volatility forecasting using this model (Degiannakis & Filis, 2017; Haugom, Langeland, Molnár, & Westgaard, 2014; Haugom, Lien, Veka, & Westgaard, 2014; Haugom, Veka, Lien, & Westgaard, 2014; Liu & Wan, 2012; Ma, Wahab, Huang, & Xu, 2017; Prokopczuk, Symeonidis, & Wese Simen, 2016; Sévi, 2014). Recently, Patton and Sheppard (2015) used the realized semi-variances proposed by Barndorff-Nielsen, Kinnebrock, and Shephard (2010), which decompose the realized variance into a component that relates only to positive high-frequency returns (“good” volatility) and into a component that relates only to negative high-frequency returns (“bad” volatility). They demonstrate that the previous literature usually uses the square (absolute) intraday returns, which can result in losing the information that may be contained in the sign of these returns. They further find that those models that include the realized semi-variances can improve the forecasting performance. Owing to special information on signed intraday returns, realized semi-variances have been given attention by many scholars, such as Chen and Ghysels (2010), Baruník, Kočenda, and Vácha (2016), Duong and Swanson (2015), and Audrino and Hu (2016).

It is well known that the decomposition between the continuous and the jump components may help to obtain better forecast accuracy (Andersen et al., 2007; Corsi et al., 2010; Sévi, 2014). Thus, inspired by the above-mentioned reference, we further decompose the realized semi-variances into continuous and discontinuous jump components, which are similar to realized volatility. In other words, the positive (negative) realized semi-variance can divide the positive (negative) continued sample path realized semi-variance (hereafter denoted CRS) and positive jump components. In accordance with Patton and Sheppard (2015), we have new signed jump variations, which include the jump tests. We use those variables and combine them with HAR-RV and its various extensions to construct the new models, labeled as HAR-RV-type. In our study, we empirically investigate whether our proposed models have better performance than do the existing models in predicting the future volatility.

Our main contributions are as follows. The first contribution of this paper is to introduce new continuous realized semi-variances and signed jump variations. We can use those variables to gain new insights and first investigate the impact of those variables on forecasting oil price volatility. The second contribution is to propose several novel HAR-RV-type models based on those proposed variables and evaluate their forecasting performance using

the statistical and economic significance. Our papers are closely related to the studies by Sévi (2014) and Patton and Sheppard (2015); therefore, we compare our new HAR-RV-type models with their models by the statistical and economic significance. Regarding forecasting performance, the majority view is that statistical significance is not sufficient to prove the superiority of a specific model, because market investors are more interested in the economic value of volatility models. Therefore, we follow the literature by considering a mean-variance utility investor who allocates his or her assets between stock and the risk-free Treasury bill, where the optimal weight of stock in the portfolio is *ex ante* determined by volatility and mean forecasts of the stock return (see, e.g., Guidolin & Na, 2006; Neely, Rapach, Tu, & Zhou, 2014; Rapach, Strauss, & Zhou, 2010). To the best of our knowledge, the economic value of realized volatility forecasts has been considered in very few existing studies beyond the notable work of Fleming, Kirby, and OstDiek (2003). Therefore, in our study, we seek to answer the following question: Do our new HAR-RV-type models help investors obtain more economic benefits?

The major findings from this study are threefold. First, the new negative CRS has a stronger impact on the future volatility of the oil futures market than on that of the positive CRS. Second, the continuous signed jump variations have significant negative effects on future realized volatility, and the effect is larger than that of the existing sign jump variation proposed by Patton and Sheppard (2015). Depending on the Wald test, we find that our new continuous realized semi-variances and signed jump variations show a strong asymmetric “leverage effect” on future volatility. Third, based on those new variables, we extend several new HAR-RV-type volatility models. Our out-of-sample results demonstrate that the newly proposed variables and extended HAR-RV-type models not only attain better performance in forecasting the volatility of the oil futures market but also garner larger economic values than traditional HAR-RV-types. These conclusions are robust and reliable for different benchmark and forecasting windows.

The remainder of this paper is organized as follows. Section 2 provides the descriptions of the volatility measures and models. Section 3 presents the data and the preliminary analysis. The empirical forecasting results are presented in Section 4. Section 5 concludes the paper.

## 2 | VOLATILITY MEASURES, JUMPS AND HAR-TYPE MODELS

In this section we provide a brief description of several popular volatility measures using intraday data and the

corresponding extended models by capturing the dynamics of these volatility measures.

## 2.1 | Volatility measures and jumps

For a given day  $t$ , we divide the time interval  $[0, 1]$  into subintervals of length  $n = 1/\Delta$ , where  $\Delta$  is the sampling frequency. Therefore, realized volatility is defined as the sum of all available intraday high-frequency squared returns:

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

where  $r$  is the intraday return. As noted by Barndorff-Nielsen and Shephard (2004), when  $\Delta \rightarrow 0$ ,  $RV$  is satisfied:

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

where  $\int_0^t \sigma^2(s) ds$  is the integrated variance (IV) and  $\sum_{0 < s \leq t} \kappa^2(s)$  is the discontinuous part of the quadratic variation (QV). Integrated variance can be computed by realized bi-power variation (BPV), defined as below:

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

where  $u_1 = (2/\pi)^{0.5} \simeq 0.7979$ .

From Equations 1, 2, and 3, we can determine the jump component, which can be calculated as  $J_t = \max(RV_t - BPV_t, 0)$ . However, these jumps may be excessively small to be statistically significant. In this study, we use the C-Tz test proposed by Corsi et al. (2010) to identify whether the jumps are significant. The C-Tz test is basically a correction of the  $Z$  statistics of Barndorff-Nielsen and Shephard (2006), which soften the finite-sample bias in estimating the integral of the second and fourth powers of continuous volatility in the presence of jumps. Corsi et al. (2010) use the simulating data to compare the performance of the two tests and found that, in the presence of jumps, the C-Tz test has substantially more power than the  $Z$  test, particularly when jumps are consecutive—a situation that is very frequent for high-frequency data. The C-Tz test is defined as

$$C\text{-}Tz_t = \Delta^{-\frac{1}{2}} \frac{(RV_t - C\text{-}TBPV_t) \cdot RV_t^{-1}}{\sqrt{\left(\frac{\pi^2}{4} + \pi - 5\right) \max\left\{1, \frac{C\text{-}TTPV_t}{(C\text{-}TTPV)^2}\right\}}}. \quad (4)$$

The corrected realized threshold multi-power ( $C\_TMPV$ ) is defined as

$$C\text{-}TMPV_t^{[\gamma_1, \dots, \gamma_M]} = \Delta^{1-\frac{1}{2}} (\gamma_1, \dots, \gamma_M) \sum_{j=M}^{1/\Delta} \prod_{k=1}^M z_{\gamma_k} (r_{j-k+1}, \vartheta_{j-k+1}), \quad (5)$$

where  $\gamma_1, \dots, \gamma_M > 0$ ,  $\vartheta_t$  is the stochastic threshold, and  $\vartheta_t = c_\vartheta^2 \cdot \widehat{V}_t$ ,  $\widehat{V}_t$  is an auxiliary estimator of latent volatility;  $c_\vartheta$  is a scale-free constant ( $c_\vartheta = 3$ ). Thus Equation 4 of C - TBPV<sub>*t*</sub> and C - TTPV<sub>*t*</sub> are equal to  $u_1^{-2} C\_TMPV_{t[1,1]}$  and  $u_1^{-3} C\_TMPV_{t[4/3,4/3,4/3]}$ , respectively.

## 2.2 | Continuous realized semi-variations and signed jump variation

Barndorff-Nielsen et al. (2010) propose the new estimators (termed realized semi-variance) based on the sign of intraday returns that can identify downside and upside risk and capture the sign asymmetry of the volatility process; these are defined as follows:

$$RS_t^- = \sum_{j=1}^{1/\Delta} r_{(t-1)+j^*\Delta, \Delta}^2 I_{[r_{(t-1)+j^*\Delta} < 0]}, \quad (6)$$

$$RS_t^+ = \sum_{j=1}^{1/\Delta} r_{(t-1)+j^*\Delta, \Delta}^2 I_{[r_{(t-1)+j^*\Delta} > 0]}, \quad (7)$$

where  $I_{[\cdot]}$  is the indicator function and takes the value 1 if the argument in  $[\cdot]$  is true. From Equations 6 and 7 we find that  $RV$  is completely decomposed into  $RS^-$  and  $RS^+$  ( $RV = RS^- + RS^+$ ). Barndorff-Nielsen et al. (2010) prove that when the sampling interval  $\Delta \rightarrow 0$  these variables are satisfied under in-fill asymptotes:

$$RS_t^- \rightarrow \frac{1}{2} \int_0^t \sigma^2(s) ds + \sum_{0 < s \leq t} \kappa^2(s) I_{[\kappa_s < 0]}, \quad (8)$$

$$RS_t^+ \rightarrow \frac{1}{2} \int_0^t \sigma^2(s) ds + \sum_{0 < s \leq t} \kappa^2(s) I_{[\kappa_s > 0]}, \quad (9)$$

where  $RS^-$  ( $RS^+$ ) provides a new source of information that focuses on squared negative (positive) jumps. Interestingly, the continuous component can be removed by simply subtracting one  $RS$  from the other, and the remaining part is what we define as the signed jump variation (SJV):

$$SJV_t \equiv RS_t^+ - RS_t^- \rightarrow \sum_{0 < s \leq t} \kappa^2(s) I_{[\kappa_s > 0]} - \sum_{0 < s \leq t} \kappa^2(s) I_{[\kappa_s < 0]}. \quad (10)$$

Furthermore, we further decompose the signed jump variation into negative and positive components by utilizing the sign of signed jump variations:

$$SJV_t^- = (RS_t^+ - RS_t^-)I_{[(RS_t^+ - RS_t^-) < 0]}, \quad (11)$$

$$SJV_t^+ = (RS_t^+ - RS_t^-)I_{[(RS_t^+ - RS_t^-) > 0]}. \quad (12)$$

Inspired by Ceylan (2014), Patton and Sheppard (2015), and Audrino and Hu (2016), we first propose the continuous semi-variances using the C-Tz test. From Equations 3, 6, 7, 8, and 9, we can determine the magnitude of signed jumps (SJ):

$$SJ_t^- = RS_t^- - \frac{1}{2}BPV_t, \quad (13)$$

$$SJ_t^+ = RS_t^+ - \frac{1}{2}BPV_t. \quad (14)$$

Thus we can use the C-Tz test to identify significant negative and positive jumps. In accordance with the work of Corsi et al. (2010), we use the C-TBPV to replace the BPV in Equations 13 and 14. In addition, to ensure those variables are non-negative, we truncate the actual empirical measurements at zero and defined as

$$J_t^- = \max(0, SJ_t^-)I_{[C-TZ_t > \Phi_\alpha]}, \quad (15)$$

$$J_t^+ = \max(0, SJ_t^+)I_{[C-TZ_t > \Phi_\alpha]}. \quad (16)$$

Consequently, we can define the continuous semi-variances as below:

$$CRS_t^- = I_{[C-TZ_t \leq \Phi_\alpha]} \cdot RS_t^- + \frac{1}{2}I_{[C-TZ_t > \Phi_\alpha]} \cdot C-TBRV_t, \quad (17)$$

$$CRS_t^+ = I_{[C-TZ_t \leq \Phi_\alpha]} \cdot RS_t^+ + \frac{1}{2}I_{[C-TZ_t > \Phi_\alpha]} \cdot C-TBPV_t. \quad (18)$$

Based on the concepts of signed jump variations (Patton & Sheppard, 2015), we can have a new signed jump variation including the jumps test; this is termed the continuous signed jump variation (CSJV):

$$CSJV_t \equiv CRS_t^+ - CRS_t^-. \quad (19)$$

We further decompose the CSJV into its positive and negative components based on its sign, which are defined as

$$CSJV_t^- = (CRS_t^+ - CRS_t^-)I_{[(CRS_t^+ - CRS_t^-) < 0]}, \quad (20)$$

$$CSJV_t^+ = (CRS_t^+ - CRS_t^-)I_{[(CRS_t^+ - CRS_t^-) > 0]}. \quad (21)$$

In our analysis, we use new variables, such as  $CRS_t^-$ ,  $CRS_t^+$ ,  $CSJV_t$ ,  $CSJV_t^+$ , and  $CSJV_t^-$ , to gain new insights into

the empirical behavior of volatility as it relates to signed returns and the jump test.

## 2.3 | Prediction models

In recent years, the heterogeneous autoregressive model of Corsi (2009) has become the most popular model for describing the dynamics of RV (henceforth, HAR-RV). This model accommodates some of the stylized facts found in financial asset return volatility, such as long memory and multiscaling behavior. The HAR-RV model is simple to implement, as it only contains three explanatory variables: lagged average daily realized volatility, lagged average weekly realized volatility, and lagged average monthly realized volatility. To the best of our knowledge, there are many various extended HAR-RV models. However, in this paper, we consider only the extended HAR-RV model, which includes the realized semi-variances and signed jump variations components.

Chen and Ghysels (2010) construct the new model dependent on the realized semi-variance proposed by Barndorff-Nielsen et al. (2010); this is termed the CG model. This specification of model is

$$\begin{aligned} RV_{t+h} = & c + \beta_{dp}RS_t^+ + \beta_{wp}RS_{t-4,t}^+ + \beta_{mp}RS_{t-21,t}^+ \\ & + \beta_{dn}RS_t^- + \beta_{wn}RS_{t-4,t-1}^- + \beta_{mn}RS_{t-21,t}^- \\ & + \beta_j J_t + \omega_{t+h}. \end{aligned} \quad (22)$$

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 (PS1) adds a term that interacts the lagged realized variance with an indicator for the negative daily returns,  $RV_t I(r_t < 0)$ :

$$\begin{aligned} RV_{t+h} = & c + \beta_{dp}RS_t^+ + \beta_{dn}RS_t^- + \gamma RV_t I_{[r_t < 0]} \\ & + \beta_w RV_{t-4,t} + \beta_m RV_{t-21,t} + \omega_{t+h}. \end{aligned} \quad (23)$$

The second model to capture the “leverage effect” contains a signed jump variation and an estimator of the variation caused by the continuous part (bi-power variation) (PS2):

$$\begin{aligned} RV_{t+h} = & c + \varphi SJV_t + \beta_{BPV} BPV_t + \beta_w RV_{t-4,t} \\ & + \beta_m RV_{t-21,t} + \omega_{t+h}. \end{aligned} \quad (24)$$

The last model for the “leverage effect” disentangles the role of the positive and negative jumps (PS3):

$$\begin{aligned} RV_{t+h} = & c + \varphi^+ SJV_t^+ + \varphi^- SJV_t^- + \beta_{BPV} BPV_t \\ & + \beta_w RV_{t-4,t} + \beta_m RV_{t-21,t} + \omega_{t+h}. \end{aligned} \quad (25)$$

In the sequel, we estimate various HAR-RV models that incorporate all the jump variables discussed above.



Sévi (2014) add the lagged average weekly and monthly of  $SJV_t$ ,  $SJV_t^+$  and  $SJV_t^-$  into models PS2 and PS3 as well as other continuous components that are related to the Z test. Nevertheless, to reduce the numbers of models and consider the important variables, we use the C-Tz test to replace the Z test and obtain the continuous components, and we build two more new HAR-RV type models; the first is (CSJ):

$$RV_{t+h} = c + \varphi_d SJV_t + \beta_{BPV} C \cdot TBPV_t + \beta_w TC_{t-4,t} + \beta_m TC_{t-21,t} + \varphi_w SJV_{t-4,t} + \varphi_m SJV_{t-21,t} + \omega_{t+h}. \quad (26)$$

The second new one is (CSJd):

$$RV_{t+h} = c + \varphi_d^+ SJV_t^+ + \varphi_d^- SJV_t^- + \beta_{BPV} C \cdot TBPV_t + \beta_w TC_{t-4,t} + \beta_m TC_{t-21,t} + \varphi_w^+ SJV_{t-4,t}^+ + \varphi_w^- SJV_{t-4,t}^- + \varphi_m^+ SJV_{t-21,t}^+ + \varphi_m^- SJV_{t-21,t}^- + \omega_{t+h}. \quad (27)$$

In the sequel, we use these new variables, as discussed in Section 2.2, to replace  $RS_t^+$ ,  $RS_t^-$ ,  $SJV_t$ ,  $SJV_t^+$  and  $SJV_t^-$  in the existing HAR-type models discussed above. Therefore, we construct six new HAR-RV type models based on the new variables ( $CRS_t^+$ ,  $CRS_t^-$ ,  $CSJV_t$ ,  $CSJV_t^+$  and  $CSJV_t^-$ ) and investigate whether those new variables are more powerful in forecasting volatility. The specifications of the six new models are as follows:

#### New specification 1: CG-New

$$RV_{t+h} = c + \beta_{dp} CRS_t^+ + \beta_{wp} CRS_{t-4,t}^+ + \beta_{mp} CRS_{t-21,t}^+ + \beta_{dn} CRS_t^- + \beta_{wn} CRS_{t-4,t}^- + \beta_{mn} CRS_{t-21,t}^- + \beta_j TJ_t + \omega_{t+h}, \quad (28)$$

where  $CRS_{t-4,t}^+$  ( $CRS_{t-4,t}^-$ ) and  $CRS_{t-21,t}^+$  ( $CRS_{t-21,t}^-$ ) are the average weekly and monthly aggregate volatility of continuous realized semi-variations.

#### New specification 2: PS1-New

$$RV_{t+h} = c + \beta_{dp} CRS_t^+ + \beta_{dn} CRS_t^- + \gamma TC_t I(r_t < 0) + \beta_w TC_{t-4,t} + \beta_m TC_{t-21,t} + \beta_j TJ_t + \omega_{t+h}, \quad (29)$$

where  $TC_t$  is the continuous component using C-Tz test,  $TC_t = I[C - TZ_t \leq \Phi_\alpha] \cdot RV_t + I[C - TZ_t > \Phi_\alpha] \cdot C \cdot TBPV_t$ ,  $TC_{t-4,t}$  and  $TC_{t-21,t}$  are the lagged weekly and monthly average volatility, respectively.  $r_t$  is the daily return of day  $t$ .

#### New specification 3: PS2-New

$$RV_{t+h} = c + \varphi_{CSJV} CSJV_t + \beta_{BPV} C \cdot TBPV_t + \beta_w TC_{t-4,t} + \beta_m TC_{t-21,t} + \omega_{t+h}, \quad (30)$$

where  $CSJV_t$  is the continuous signed jump variations.

#### New specification 4: PS3-New

$$RV_{t+h} = c + \varphi_{CSJV}^+ CSJV_t^+ + \varphi_{CSJV}^- CSJV_t^- + \beta_{BPV} C \cdot TBPV_t + \beta_w TC_{t-4,t} + \beta_m TC_{t-21,t} + \omega_{t+h}, \quad (31)$$

where  $CSJV_t^+$  and  $CSJV_t^-$  are positive and negative continuous signed jump variations.

#### New specification 5: CSJ-New

$$RV_{t+h} = c + \varphi_{CSJV}^d CSJV_t + \beta_{BPV} C \cdot TBPV_t + \beta_w TC_{t-4,t} + \beta_m TC_{t-21,t} + \varphi_{CSJV}^w CSJV_{t-4,t} + \varphi_{CSJV}^m CSJV_{t-21,t} + \omega_{t+h}, \quad (32)$$

where  $CSJV_{t-4,t}$  and  $CSJV_{t-21,t}$  are the average of lagged weekly and monthly continuous signed jump variations.

#### New specification 6: CSJd-New

$$RV_{t+h} = c + \varphi_{TSJV}^+ CSJV_t^+ + \varphi_{TSJV}^- CSJV_t^- + \beta_{BPV} C \cdot TBPV_t + \beta_w TC_{t-4,t} + \beta_m TC_{t-21,t} + \varphi_{CSJV}^{w,+} CSJV_{t-4,t}^+ + \varphi_{CSJV}^{w,-} CSJV_{t-4,t}^- + \varphi_{CSJV}^{m,+} CSJV_{t-21,t}^+ + \varphi_{CSJV}^{m,-} CSJV_{t-21,t}^- + \omega_{t+h}, \quad (33)$$

where  $CSJV_{t-4,t}^+$  ( $CSJV_{t-4,t}^-$ ) and  $CSJV_{t-21,t}^+$  ( $CSJV_{t-21,t}^-$ ) are the average of lagged weekly and monthly positive (negative) continuous signed jump variations.

Table 1 provides a typology of the 22 models that are analyzed in the current study.

## 3 | DATA DESCRIPTION

Using the high-frequency financial data to construct the realized volatility, the impact of microstructure noise on the volatility estimate cannot be ignored. In addition, some studies (see, e.g., Andersen & Bollerslev, 1998; Biais, Glosten, & Spatt, 2005; Madhavan, 2000) found that the 5-minute sampling frequency is a tradeoff between measurement accuracy, which is theoretically optimized using the highest possible frequency, and microstructure noise that can arise through the bid-ask bounce, asynchronous trading, infrequent trading, and price discreteness, among other factors. Furthermore, Liu, Patton, and Sheppard (2015) showed that the standard 5-minute realized variance measure is difficult to beat in empirical forecasting exercises. Therefore, in this study, we use 5-minute data of the Light Sweet Crude Oil (WTI) futures contract with a maturity of 1 month traded in NYMEX.

As is well known, the WTI oil futures contract is one of the most heavily traded commodity futures throughout the world, and most oil-based derivatives are priced with

**TABLE 1** Typology of the empirical specifications used in the paper

Model number	Model name	Reference	Equation number in this paper
1	CG	Chen and Ghysels (2010), Equation 7	22
2	PS1	Patton and Sheppard (2015), Equation 17	23
3	PS2	Patton and Sheppard (2015), Equation 18	24
4	PS3	Patton and Sheppard (2015), Equation 19	25
5	CSJ	Sévi (2014), Equation 20	26
6	CSJd	Sévi (2014), Equation 22	27
7	CG-New		28
8	PS1-New		29
9	PS2-New	New specification in this paper	30
10	PS3-New		31
11	CSJ-New		32
12	CSJd-New		33

*Note.* We extend these six existing HAR-RV-type models (as stated in Equations 28–33) with our new proposed variables. The models with the symbol “-New” include our proposed variables, such as continuous realized semi-variance and signed jump variations including jump tests.

respect to this contract. In addition, the WTI oil futures are the reference for most investors involved in energy commodities. Furthermore, this contract is an important component asset of the Goldman Sachs Commodity Index (GSCI) and other commodity indices that are widely traded worldwide. Our data are collected from the TickDATA,<sup>1</sup> covering the period from January 2, 2007, to May 9, 2014. After removing days with a shortened trading session or excessively few transactions, we obtain 1,851 daily observations.<sup>2</sup> Figure 1(a) shows all the available data of the high-frequency oil futures prices of the whole sample. Figure 1(b) demonstrates the WTI spot oil prices of the same period to the futures market. We

<sup>1</sup><https://www.tickdata.com/>

<sup>2</sup>As detailed in Zivot and Wang (2005), for the WTI futures contract we apply an initial filter to remove (1) transactions outside the official trading period, (2) transactions with a variation of more than 5% in absolute value compared to the previous transaction, and (3) transactions not reported in chronological order. These are the same practices used by Chevallier and Sévi (2012) and Sévi (2014).

can observe that the two price series have the same tendency.

Table 2 shows the descriptive statistics of some relative time series. All realized measures, jumps, and signed jump variations series are significantly skewed and leptokurtic at the 1% level, suggesting that they are fat-tail distributed. The Jarque–Bera statistic further demonstrates that the null hypothesis of normality is rejected at the 1% significance level. The above tests' results tell us that all of the series that we have discussed are nonnormal distributed. The Ljung–Box statistic for serial correlation shows that the null hypothesis of no autocorrelation to the 5th order is rejected for most of the series, which indicate that there is evidence for series correlation in the realized measures, jumps, and signed jump variations series. Furthermore, at the 99% confidence level, the C-Tz statistics detect 367 significant jumps in the whole sample period, which is larger than that of the Z statistics (detecting 238 jumps).

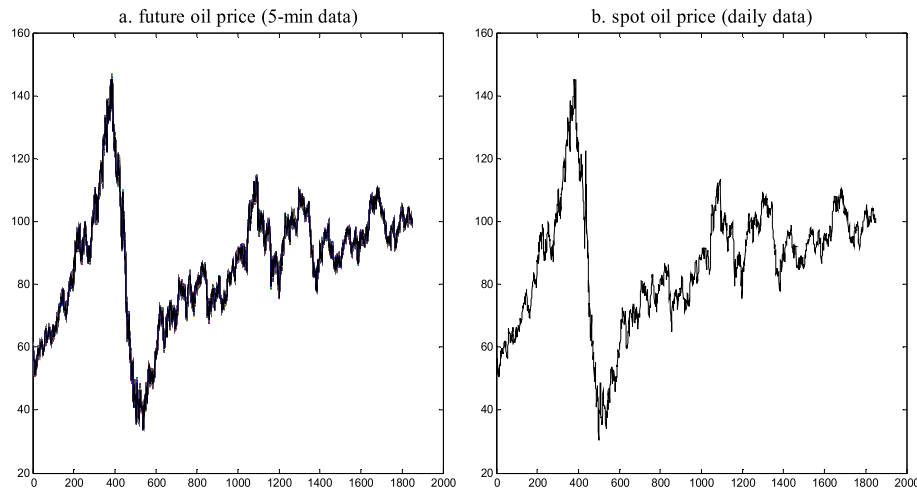
## 4 | EMPIRICAL ANALYSIS

In this section, we first provide the in-sample estimation results of the 12 volatility models in this study. Then, we evaluate the out-of-sample forecasting performance of those models, both statistically and economically.

### 4.1 | In-sample estimation analysis

Tables 3 and 4 only show the one-step-ahead estimation results of the six traditional and six newly proposed HAR-RV-type models over the in-sample period based on the Newey–West correction, which allows for series correlation to a maximum order of 5.<sup>3</sup> The signed jump variation and new signed jump variation proposed in this study have a significant negative impact on future realized volatility, indicating that those variables lead to lower future realized volatility. Comparing PS2 in Table 3 with the new model PS2-New in Table 4, we may find that the signed jump variation coefficient of the new model ( $\varphi_d$ ) is larger than that in PS2, implying that the new signed jump variations have greater impact on future volatility. The estimation results given in Tables 3 and Table 4 for PS3 and PS3-New demonstrate that the main influence of the signed jump variation on future volatility is the negative signed jump variation; this is the “asymmetry” effect. We further use the Wald test statistic to test the null hypothesis:  $\varphi_d^+ = \varphi_d^- = \varphi_d$ ; and all the results reject the hypothesis (the  $p$ -values are all equal (or close)

<sup>3</sup>They can request the estimation results of other horizons by the authors, such as  $h = 5$  (22).



**FIGURE 1** Oil futures price (5-minute) and spot oil prices (daily) over the whole sample period [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

**TABLE 2** Descriptive statistics of each volatility series

	Mean	SD	Skewness	Kurtosis	Jarque–Bera	Q(5)
RV	3.005	4.262	3.941***	19.335***	33225.704***	5453.734***
TC	2.772	3.921	3.842***	18.006***	29062.655***	5428.804***
BPV	2.726	3.906	3.988***	20.329***	36164.290***	5408.288***
C-TBPV	2.424	3.515	4.082***	22.096***	42285.452***	5370.023***
J	0.324	0.775	7.285***	73.919***	432583.731***	455.498***
CJ	0.113	0.579	12.325***	198.908***	3046388.890***	15.129***
TJ	0.232	0.961	8.360***	86.647***	590527.324***	25.295***
SJV	−0.060	1.529	0.299***	29.749***	67140.138***	5.217
CSJV	−0.041	1.105	0.571***	33.004***	83112.737***	16.323***

*Note.* Null hypothesis: “Skewness = 0” and “Kurtosis = 3”. Kurtosis is excess kurtosis. The Jarque–Bera statistic (Jarque & Bera, 1987) tests 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. Asterisks denote rejection of the null hypothesis at \*\*\*1%, \*\*5%, and \*10% significance levels. All of realized measures, jumps, and signed jump variations series in the table are multiplied by 10,000. RV is the realized variance, TC represents the continuous sample path and  $TC_t = I[C - TZ_t \leq \Phi_\alpha] \cdot RV_t + I[C - TZ_t > \Phi_\alpha] \cdot C - TBPV_t$ . BPV is the realized bi-power variation, C-TBPV =  $u_1^{-2} C - TMPV_{[1,1]}$ . J, CJ and TJ indicate the jump component. SJV is the signed jump variation; CSJV is the continuous signed jump variation using the C-Tz test.

to zero), thus providing convincing evidence of an asymmetric effect on future volatility by a negative and positive signed jump variation. In addition, the negative signed jump variations have a significant negative impact on the future volatility at the significance level of 1%; however, the positive signed jump variation has no influence on future volatility. Furthermore, the ranking of the absolute coefficient value is  $|\varphi_d^-| > |\varphi_d| > |\varphi_d^+|$ . Therefore, the future volatility is more strongly related to the past negative returns than that of positive returns.

The coefficient of the negative semi-variance ( $\beta_{dn}$ ) is positive and much larger than that of the positive semi-variance ( $\beta_{dp}$ ), implying that negative semi-variance

contributes more to future realized volatility. This finding reveals the existence of a strong “leverage effect.” Moreover, we examine the null hypothesis that positive and negative semi-variances have equal predictive power for realized volatility ( $\beta_{dn} = \beta_{dp}$ ) based on the model specifications of CG, PS1, CG-New and PS1-New. The chi-square statistics of the standard Wald test reject the null hypothesis at the 1% significance level, indicating that the “leverage effect” is significant.

From the results of Tables 3 and 4, we conclude that the effect of the jump components calculated by the C-Tz test on future realized volatility are significantly positive at the 10% confidence level. However, the jump

**TABLE 3** The estimation results of the existed HAR-type models in-sample period ( $h = 1$ )

	CG	PS1	PS2	PS3	CSJ	CSJd
$c$	0.000*	0.000	0.000	0.000	0.000**	0.000**
$\beta_d$					0.095**	
$\beta_{bpv}$			0.075**	0.024***		
$\beta_{dp}$	−0.074	−0.087				
$\beta_{dn}$	0.247***	0.270***				
$\beta_w$		0.382***	0.424***	0.394***	0.400***	0.377***
$\beta_{wp}$	0.238*					
$\beta_{wn}$	0.326**					
$\beta_m$		0.478***	0.474***	0.485***	0.708***	0.760***
$\beta_{mp}$	−0.438*					
$\beta_{mn}$	1.548***					
$\beta_j$	0.061					
$\beta_{wj}$						
$\beta_{mj}$						
$\gamma$		0.031				
$\varphi_d$			−0.205***		−0.212***	
$\varphi_w$					0.026***	
$\varphi_m$					−0.939***	
$\varphi_d^+$				0.073		0.248***
$\varphi_w^+$						−0.435**
$\varphi_m^+$						−1.251***
$\varphi_d^-$				−0.481***		−0.570***
$\varphi_w^-$						−0.028
$\varphi_m^-$						−0.356
Adj. $R^2$	0.725	0.720	0.719	0.723	0.715	0.723

Note. Asterisks denote rejection of the null hypothesis at \*\*\*1%, \*\*5% and \*10% significance levels. The CG, PS1, PS2, PS3, CSJ and CSJd models are all existing models, and estimated in Equations 28–33 over the in-sample period by using the OLS with Newey–West correction.

components are the difference of RV and BPV, which have no influence on future volatility.

## 4.2 | Forecasts evaluation

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 the past. Our sample data are divided into two subgroups: (1) in-sample data for volatility modeling, covering the first 1,222 trading days (January 2, 2007–November 1, 2011); and (2) out-of-sample data for model evaluation, covering the last 629 trading days (November 2, 2011–May 9, 2014). The

estimation period is then rolled forward by adding one new day and eliminating the most distant day. Thus the sample size that is used to estimate the models remains at a fixed length, and the forecasts do not overlap. In this section, a rolling window method is employed to obtain the volatility forecasting results of each model. The out-of-sample is set to be the last 629 days of the whole sample, indicating that each model is reestimated 629 times, and its parameters are time varying with different samples. Using the CG model (M4) as an example, Figure 2 shows the time-varying parameter estimates through the out-of-sample period, clearly exhibiting very volatile patterns for each parameter in this model, particularly in the last approximately 200 days.

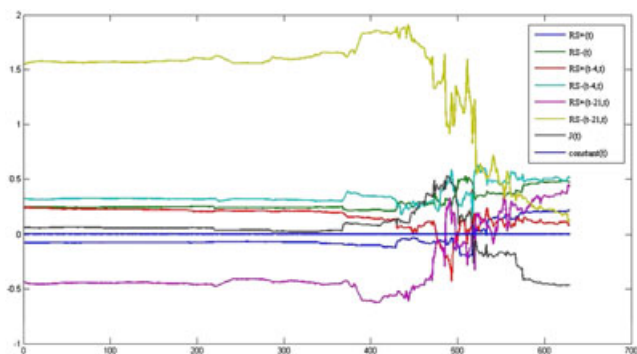
In our study, we first divide the six existing and six extended HAR-RV-type models into six groups: CG and CG-New; PS1 and PS1-New; PS2 and PS2-New; PS3 and



**TABLE 4** Estimation results of the new HAR-type models in-sample period ( $h = 1$ )

	CG-New	PS1-New	PS2-New	PS3-New	CSJ-New	CSJd-New
$c$	0.000	0.000	0.000	0.000	0.000	0.000**
$\beta_{bpv}$			0.122***	0.080***	0.087**	0.004*
$\beta_{dp}$	-0.171**	-0.229***				
$\beta_{dn}$	0.422***	0.486***				
$\beta_w$		0.407***	0.438***	0.465***	0.639***	0.645***
$\beta_{wp}$	0.219***					
$\beta_{wn}$	0.346**					
$\beta_m$		0.481***	0.484***	0.483***	0.816***	0.745***
$\beta_{mp}$	-0.787***					
$\beta_{mn}$	1.992***					
$\beta_j$	0.132*					
$\beta_{wj}$						
$\beta_{mj}$						
$\gamma$		0.022				
$\varphi_d$			-0.380***		-0.359***	
$\varphi_w$					-0.902***	
$\varphi_m$					-2.128***	
$\varphi_d^+$				-0.229**		0.064
$\varphi_w^+$						-1.135***
$\varphi_m^+$						-2.153***
$\varphi_d^-$				-0.505***		-0.669***
$\varphi_w^-$						-1.228***
$\varphi_m^-$						0.010
Adj. $R^2$	0.7272	0.719	0.7167	0.7161	0.7258	0.7342

Note. Asterisks denote rejection of the null hypothesis at \*\*\*1%, \*\*5% and \*10% significance levels. The CG, PS1, PS2, PS3, CSJ and CSJd models are all extended models, and estimated from Equations 22–27 over the in-sample period by using the OLS with Newey–West correction.



**FIGURE 2** Parameter estimation results of the CG model over the out-of-sample period [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

PS3-New; CSJ and CSJ-New; CSJd and CSJd-New. The benchmark volatility models of every group are CG, PS1, PS2, PS3, CSJ, and CSJd, respectively. Clark and

McCracken (2001) and McCracken (2007) show that the most popular evaluation methods—the Diebold and Mariano (1995) and West (1996) statistic—have nonstandard distributions when comparing forecasts from nested models. In accordance with the works of Campbell and Thompson (2008), Welch and Goyal (2008), and Rapach et al. (2010), we adopt the out-of-sample  $R^2$  statistic to evaluate the economic significance of differences in forecast performance. The statistic can be defined as

$$R_{\text{OOS}}^2 = 1 - \frac{\sum_{t=1}^M (RV_t - \widehat{RV}_{t,i})^2}{\sum_{t=1}^M (RV_t - \widehat{RV}_{t,0})^2} \quad (34)$$

where  $RV_t$  is the actual volatility. In accordance with previous studies (see, e.g., Corsi et al., 2010; Haugom, Langeland, et al., 2014; Koopman, Jungbacker, & Hol,

2005; Wang et al., 2016), we also use RV as a proxy for actual market volatility.  $\widehat{RV}_{t,i}$  is the out-of-sample forecasts. When  $R^2_{OOS}$  is larger than zero, the  $\widehat{RV}_{t,i}$  forecast outperforms the benchmark volatility model according to the MSPE metric. Thus, we use the *MSPE-adjusted* statistic proposed by Clark and West (2007) to assess the statistical significance between the benchmark and the extended models.

Table 5 shows the out-of-sample  $R^2$  statistic results. The main findings of the empirical results are as follows. Comparing the six groups, we find that for different

forecasting horizons the CG-New, PS1-New, PS2-New, PS3-New, and CSJd-New models all have significant differences in the forecasting performance of their benchmark models at the 10% significance level; these models are first proposed by this study. However, the CSJ-New model appears not to significantly beat the CSJ model. In general, our proposed models have better performance in forecasting, implying that our continuous realized semi-variances and signed jump variations are more powerful in forecasting the oil price volatility.

To evaluate the forecasting performance of the different groups and quantitatively evaluate the forecasting accuracy, we follow the literature by using the following four popular loss functions:

$$MSE = M^{-1} \sum_{m=H+1}^M (RV_m - \hat{\sigma}_m^2)^2, \quad (35)$$

$$HMSE = M^{-1} \sum_{m=1}^M (1 - \hat{\sigma}_m^2 / RV_m)^2, \quad (36)$$

$$HMAE = M^{-1} \sum_{m=1}^M |1 - \hat{\sigma}_m^2 / RV_m|, \quad (37)$$

$$QLIKE = M^{-1} \sum_{m=1}^M (\ln(\hat{\sigma}_m^2) + RV_m / \hat{\sigma}_m^2), \quad (38)$$

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

However, the above-mentioned loss functions provide no information on whether the differences of forecasting losses among models are statistically significant. For this consideration, we use an advanced statistical test, which is a model confidence set (MCS) test proposed by Hansen, Lunde, and Nason (2011), to choose a subset of models containing all possible superior models from the initial model set. The MCS method has several attractive advantages over conventional tests such as superior predictive ability (Hansen, 2005) and “reality check” tests. In our analysis, we do not provide further technical details of the MCS test; however, more in-depth discussions may be found in Hansen et al. (2011).

Following Martens, van Dyke, and de Pooter (2009), Hansen et al. (2011), and Laurent, Rombouts, and Violante (2012), among others, we set a 75% confidence level. This setting enables us to exclude (or cancel) a model from the initial model set with a  $p$ -value smaller than 0.25. The  $p$ -values of the MCS test are obtained based on 10,000 block bootstraps. Table 6 shows the results of the MCS test for superior volatility models tested by the out-of-sample  $R^2$  statistic. Bold entries in the table

**TABLE 5** Out-of-sample evaluation results using the out-of-sample  $R^2$  statistic

	$h = 1$	$h = 5$	$h = 22$
Benchmark: CG	CG-New	CG-New	CG-New
$R^2_{OOS}$ (%)	9.400***	15.90***	19.200***
MSPE-adjust	3.301	4.522	4.522
$p$ -value	0.000	0.000	0.000
Benchmark: PS1	PS1-New	PS1-New	PS1-New
$R^2_{OOS}$ (%)	4.600***	4.400***	4.500***
MSPE-adjust	2.446	2.569	2.369
$p$ -value	0.007	0.005	0.009
Benchmark: PS2	PS2-New	PS2-New	PS2-New
$R^2_{OOS}$ (%)	3.300***	3.400***	3.500***
MSPE-adjust	2.172	2.342	1.918
$p$ -value	0.015	0.010	0.028
Benchmark: PS3	PS3-New	PS3-New	PS3-New
$R^2_{OOS}$ (%)	6.500**	1.400***	6.500**
MSPE-adjust	1.649	2.042	1.649
$p$ -value	0.050	0.021	0.050
Benchmark: CSJ	CSJ-New	CSJ-New	CSJ-New
$R^2_{OOS}$ (%)	−1.600*	−9.400	−9.400
MSPE-adjust	1.465	−0.583	−2.007
$p$ -value	0.071	0.720	0.978
Benchmark: CSJd	CSJd-New	CSJd-New	CSJd-New
$R^2_{OOS}$ (%)	6.400*	1.100***	5.500***
MSPE-adjust	1.490	2.663	2.858
$p$ -value	0.068	0.004	0.002

*Note.* Asterisks denote rejection of the null hypothesis at \*\*\*1%, \*\*5% and \*10% significance levels. Values of  $R^2_{OOS}$  were obtained in Equation 34. If  $R^2_{OOS}$  is larger than zero, indicating that the corresponding model outperforms the benchmark model, we use the *MSPE-adjusted* statistic proposed by Clark and West (2007) to assess the statistical significance between the benchmark and the extended models. This table reports the evaluation results on 1-day, 5-day, and 22-day forecast horizons. The forecasting window is 629 days.

**TABLE 6** MCS of out-of-sample forecasts (the forecasting benchmark is RV)

	CSJ	CG-New	PS1-New	PS2-New	PS3-New	CSJ-New	CSJd-New
<i>h</i> = 1							
MSE	<b>0.627</b>	<b>1.000</b>	<b>0.800</b>	<b>0.647</b>	<b>0.627</b>	<b>0.532</b>	<b>0.627</b>
HMSE	<b>0.386</b>	<b>0.957</b>	<b>0.957</b>	<b>0.957</b>	<b>1.000</b>	0.000	0.015
HMAE	0.000	<b>1.000</b>	<b>0.321</b>	<b>0.321</b>	<b>0.321</b>	0.000	0.001
QLIKE	<b>1.000</b>	<b>0.611</b>	<b>0.611</b>	<b>0.588</b>	<b>0.494</b>	<b>0.470</b>	<b>0.611</b>
<i>h</i> = 5							
MSE	<b>0.891</b>	<b>0.891</b>	<b>1.000</b>	<b>0.891</b>	<b>0.288</b>	0.044	<b>0.349</b>
HMSE	<b>0.282</b>	<b>0.987</b>	<b>0.987</b>	<b>1.000</b>	<b>0.118</b>	0.000	0.004
HMAE	0.005	<b>1.000</b>	<b>0.469</b>	<b>0.469</b>	0.002	0.000	0.000
QLIKE	<b>0.666</b>	<b>0.666</b>	<b>1.000</b>	<b>0.666</b>	<b>0.479</b>	0.050	0.240
<i>h</i> = 22							
MSE	<b>1.000</b>	<b>0.307</b>	<b>0.950</b>	<b>0.950</b>	<b>0.670</b>	<b>0.011</b>	<b>0.457</b>
HMSE	<b>0.378</b>	<b>0.816</b>	<b>0.816</b>	<b>1.000</b>	<b>0.480</b>	0.000	<b>0.816</b>
HMAE	<b>0.238</b>	<b>1.000</b>	<b>0.859</b>	<b>0.859</b>	0.151	0.000	<b>0.859</b>
QLIKE	<b>0.785</b>	<b>0.448</b>	<b>1.000</b>	<b>0.785</b>	<b>0.448</b>	0.001	<b>0.584</b>

Note. Bold entries denote that the corresponding models are included in MCS for 75% confidence. We report the empirical results of the different forecast horizons. MSE, HMSE, HMAE, and QLIKE loss functions can be found in Equations 35–38. In this table, we use the realized volatility as actual market volatility. The forecasting window is 629 days.

indicate that the corresponding models are “surviving models” at the 75% confidence level.

The empirical results of Table 6 show that, in terms of forecasting one-step-ahead volatility, the MCS *p*-values of CSJ and our new models containing CG-New, PS1-New, PS2-New and PS3-New are larger than 0.25 under the criterion of HMSE, indicating that those models perform better than others. Furthermore, under the HMAE loss function, CSJ model have failed to pass the MCS test; only four new models—CG-New, PS1-New, PS2-New and PS3-New—can survive in the MCS test, which tell us that those models have better forecasting ability. Furthermore, under the MSE and QLIKE loss function, we find that all of the models have good predictive ability in the crude oil market. From the results of the four criteria, we conclude that, in general, CG-New, PS1-New, PS2-New, and PS3-New have better performance in forecasting than others; therefore, the best models are among our newly proposed models. Moreover, forecasting the middle-term volatility, CG-New, PS1-New, and PS2-New can survive in the four criteria. All the best models are also among our newly proposed models. Furthermore, regarding long-term horizons, CSJ, CG-New, PS1-New, and PS2-New can pass the MCS test. Similar to the aforementioned outcomes, most

of the best models belong to our newly proposed models, suggesting that we propose excellent HAR-RV-type models, particularly in forecasting the volatility of the oil futures market. In sum, forecasting the different horizons volatility, we find that CG-New, PS1-New, and PS2-New can attain better performance, implying that our newly presented factors, such as positive and negative continuous sample paths and signed jump variations, may contribute much to building better volatility forecasting models.

### 4.3 | Robustness check

It has been well documented in the literature that market microstructure noise plays a negative role in the measurement of realized volatility. Although several papers (e.g., Andersen & Bollerslev, 1998; Biais et al., 2005; Liu et al., 2015; Madhavan, 2000) have attempted to mitigate the effects of the noise by using 5-minute return data, Zhang, Mykland, and Ait-Sahalia (2005) show that this approach is not an adequate solution to the problem. Therefore, we use an alternative of the realized volatility-realized kernel (RK) that is robust to market microstructure noise (Barndorff-Nielsen, Hansen, Lunde, & Shephard, 2008) to reexamine whether our empirical results in Section 4.2 are robust. The realized kernel (RK) is defined as below:

$$RK_t = \sum_{j=-H}^H k\left(\frac{H}{H+1}\right) \gamma_j, \gamma_j = \sum_{i=|j|+1}^n r_{t,i} r_{t,i-|j|}, \quad (39)$$

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

$$k(x) = \begin{cases} 1-6x^2+6x^3, & 0 \leq x \leq 1/2 \\ 2(1-x)^3, & 1/2 \leq x \leq 1 \\ 0, & x > 1. \end{cases} \quad (40)$$

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 precisely follow the bandwidth choice of  $H$  that is delineated in the study by Barndorff-Nielsen, Hansen, Lunde, and Shephard (2009), to which we refer the reader for more details.

Table 7 shows the MCS test results of different volatility models using a realized kernel (RK) as forecasting criterion. Comparing the results of Tables 6 and 7, the results of Table 6 are very robust. First, when forecasting short-term volatility, all MCS *p*-values of CG-New, PS1-New, PS2-New, and PS3-New are larger than 0.25 under the MSE, HMSE, HMAE, and QLIKE criteria, respectively, which further implies that the best models are our newly proposed models. Second, in terms of middle-term horizons, the best model is CG-New under all loss functions. Third, for long-term horizons, CG-New,

**TABLE 7** MCS of out-of-sample forecasts (the forecasting benchmark is RK)

	CSJ	CG-New	PS1-New	PS2-New	PS3-New	CSJ-New	CSJd-New
<i>h</i> = 1							
MSE	<b>0.435</b>	<b>1.000</b>	<b>0.704</b>	<b>0.858</b>	<b>0.704</b>	0.094	0.233
HMSE	0.166	<b>0.747</b>	<b>0.488</b>	<b>0.747</b>	<b>1.000</b>	0.000	0.011
HMAE	0.000	<b>1.000</b>	<b>0.320</b>	<b>0.320</b>	<b>0.320</b>	0.000	0.001
QLIKE	<b>0.726</b>	<b>1.000</b>	<b>0.771</b>	<b>0.726</b>	<b>0.726</b>	0.048	<b>0.366</b>
<i>h</i> = 5							
MSE	<b>0.635</b>	<b>0.742</b>	<b>0.742</b>	<b>1.000</b>	<b>0.724</b>	0.002	0.085
HMSE	0.172	<b>0.780</b>	<b>0.746</b>	<b>1.000</b>	<b>0.384</b>	0.000	0.006
HMAE	0.002	<b>1.000</b>	0.229	0.229	0.052	0.000	0.000
QLIKE	<b>0.314</b>	<b>1.000</b>	<b>0.865</b>	<b>0.800</b>	<b>0.411</b>	0.001	0.068
<i>h</i> = 22							
MSE	<b>0.811</b>	<b>0.395</b>	<b>0.811</b>	<b>1.000</b>	<b>0.811</b>	0.003	0.225
HMSE	<b>0.531</b>	<b>0.580</b>	<b>0.580</b>	<b>1.000</b>	<b>0.580</b>	0.001	<b>0.580</b>
HMAE	0.156	<b>1.000</b>	<b>0.592</b>	<b>0.768</b>	<b>0.474</b>	0.000	<b>0.768</b>
QLIKE	<b>0.678</b>	<b>0.984</b>	<b>1.000</b>	<b>0.984</b>	<b>0.708</b>	0.000	<b>0.664</b>

Note. Bold entries denote that the corresponding models are included in MCS for 75% confidence. We report the empirical results of the different forecast horizons. MSE, HMSE, HMAE, and QLIKE loss functions can be found in Equations 35–38. In this table, we use the realized kernel as actual market volatility. The forecasting window is 629 days.

PS1-New, PS2-New, and PS3-New can survive in the MCS test, indicating that those models have superior performance in predicting long-term volatility. In short, the empirical superior performance of our newly proposed model and the volatility-related factors are very robust through different forecasting benchmarks.

Rossi and Inoue (2012) argue that arbitrary choices of window sizes have consequences regarding how the sample is split into in-sample and out-of-sample portions. Although choosing an appropriate forecasting window is crucial to the forecasting performance of the models, there appears to be no consensus on how to choose the proper forecasting windows. Therefore, we report the results of Table 8–10 obtained from other forecasting windows, such as 529 and 729. Although the results are not completely consistent with the above-mentioned results, the best models are all our newly proposed models in this paper. For example, choosing the length of the out-of-sample, 529, we find that the model with the better forecasting performance is the CG-New model, which is our proposed model in all loss functions and forecasting horizons. The results further validate that our findings are robust and reliable, indicating that our newly presented factors can help in forecasting.

**TABLE 8** Out-of-sample evaluation results using the out-of-sample  $R^2$  statistic ( $M = 529$ )

	<i>h</i> = 1	<i>h</i> = 5	<i>h</i> = 22
Benchmark: CG	CG-New	CG-New	CG-New
$R^2_{OOS}$ (%)	0.109***	0.182***	0.229***
MSPE-adjust	3.172	4.353	3.475
<i>p</i> -value	0.001	0.000	0.000
Benchmark: PS1	PS1-New	PS1-New	PS1-New
$R^2_{OOS}$ (%)	0.058***	0.053***	0.057***
MSPE-adjust	2.410	2.560	2.355
<i>p</i> -value	0.008	0.005	0.009
Benchmark: PS2	PS2-New	PS2-New	PS2-New
$R^2_{OOS}$ (%)	0.039**	0.039**	0.041**
MSPE-adjust	2.084	2.260	1.863
<i>p</i> -value	0.019	0.012	0.031
Benchmark: PS3	PS3-New	PS3-New	PS3-New
$R^2_{OOS}$ (%)	0.070**	0.026***	0.025***
MSPE-adjust	1.504	2.119	2.049
<i>p</i> -value	0.066	0.017	0.020
Benchmark: CSJ	CSJ-New	CSJ-New	CSJ-New
$R^2_{OOS}$ (%)	0.006**	−0.067	−0.081
MSPE-adjust	1.784	0.367	−1.173
<i>p</i> -value	0.037	0.357	0.880
Benchmark: CSJd	CSJd-New	CSJd-New	CSJd-New
$R^2_{OOS}$ (%)	0.079***	0.033***	0.089***
MSPE-adjust	1.412	2.962	3.072
<i>p</i> -value	0.079	0.002	0.001

Note. Asterisks denote rejection of the null hypothesis at \*\*\*1%, \*\*5% and \*10% significance levels. Values were obtained from Equation 34. If  $R^2_{OOS}$  is larger than zero, indicating that the corresponding model outperforms the benchmark model, we use the *MSPE-adjusted* statistic proposed by Clark and West (2007) to assess the statistical significance between the benchmark and the extended models. This table reports the evaluation results on 1-day, 5-day and 22-day forecast horizons. The forecasting window is 529 days.

#### 4.4 | Economic value evaluation

Compared to the statistical gains of volatility predictability, market investors care more about economic significance. Specifically, these investors are interested in how well these volatility forecasts do in asset allocation. To evaluate the economic value of volatility forecasts, we consider a mean–variance utility investor who allocates his or her assets between the stock index and the risk-free asset in accordance with the literature (see, e.g., Guidolin & Na, 2006; Neely et al., 2014; Rapach et al., 2010; Wang et al., 2016). Based on the certainty equivalence, Campbell and Thompson (2008) showed that, for a mean–variance

**TABLE 9** Out-of-sample evaluation results using the out-of-sample  $R^2$  statistic ( $M = 729$ )

	$h = 1$	$h = 5$	$h = 22$
Benchmark: CG	CG-New	CG-New	CG-New
$R^2_{OOS}$ (%)	0.026***	0.077***	0.090***
MSPE-adjust	2.670	4.493	3.249
$p$ -value	0.004	0.000	0.001
Benchmark: PS1	PS1-New	PS1-New	PS1-New
$R^2_{OOS}$ (%)	0.008**	0.010**	0.011**
MSPE-adjust	1.659	1.664	1.493
$p$ -value	0.049	0.048	0.068
Benchmark: PS2	PS2-New	PS2-New	PS2-New
$R^2_{OOS}$ (%)	−0.003**	−0.003*	0.006***
MSPE-adjust	1.507	1.329	1.424
$p$ -value	0.066	0.092	0.077
Benchmark: PS3	PS3-New	PS3-New	PS3-New
$R^2_{OOS}$ (%)	0.008*	0.013**	0.022**
MSPE-adjust	1.347	1.815	2.148
$p$ -value	0.089	0.035	0.016
Benchmark: CSJ	CSJ-New	CSJ-New	CSJ-New
$R^2_{OOS}$ (%)	−0.009*	−0.032	−0.043
MSPE-adjust	1.567	1.251	0.276
$p$ -value	0.059	0.106	0.391
Benchmark: CSJd	CSJd-New	CSJd-New	CSJd-New
$R^2_{OOS}$ (%)	−0.010	0.036**	0.014**
MSPE-adjust	1.067	4.260	2.411
$p$ -value	0.143	0.000	0.008

Note. Asterisks denote rejection of the null hypothesis at \*\*\*1%, \*\*5% and \*10% significance levels, respectively. Values were obtained from Equation 34. If  $R^2_{OOS}$  is larger than zero, indicating that the corresponding model outperforms the benchmark model, we use the *MSPE-adjusted* statistic proposed by Clark and West (2007) to assess the statistical significance between the benchmark and the extended models. This table report the evaluation results on 1-day, 5-days and 22-days forecasts horizons. The forecasting window is 729 days.

utility investor with a power utility function, the utility of a trading strategy is equal to the expected portfolio return minus  $(\gamma/2)$  times portfolio variance, where  $\gamma$  can be interpreted as the coefficient of relative risk aversion. A higher value of  $\gamma$  implies that the stock index (risky asset) is assigned a lower optimal weight in a portfolio. To measure the utility gain, the first step is to measure the average utility of the trading strategy based on the historical average. According to the CT framework, the investor will decide at the end of period  $t$  to allocate the following proportion of his or her portfolio to risky equities in period  $t + 1$ :

**TABLE 10** MCS of out-of-sample forecasts in different out-of-sample windows

	CSJ	CG-New	PS1-New	PS2-New	PS3-New	CSJ-New	CSJd-New
$M = 529$							
$h = 1$							
MSE	<b>0.494</b>	<b>1.000</b>	<b>0.708</b>	<b>0.533</b>	<b>0.494</b>	<b>0.533</b>	<b>0.643</b>
HMSE	0.003	<b>1.000</b>	<b>0.329</b>	<b>0.329</b>	<b>0.329</b>	0.000	0.000
HMAE	0.000	<b>1.000</b>	0.066	0.041	0.066	0.000	0.000
QLIKE	0.834	<b>0.970</b>	<b>1.000</b>	<b>0.542</b>	<b>0.588</b>	0.202	<b>0.675</b>
$h = 5$							
MSE	<b>0.663</b>	<b>0.891</b>	<b>1.000</b>	<b>0.516</b>	<b>0.254</b>	0.079	<b>0.416</b>
HMSE	0.000	<b>1.000</b>	0.150	0.150	0.018	0.000	0.000
HMAE	0.000	<b>1.000</b>	0.030	0.027	0.000	0.000	0.000
QLIKE	0.048	<b>1.000</b>	<b>0.846</b>	0.219	0.017	0.001	0.018
$h = 22$							
MSE	<b>0.628</b>	<b>0.628</b>	<b>1.000</b>	<b>0.628</b>	<b>0.457</b>	0.012	<b>0.628</b>
HMSE	0.000	<b>1.000</b>	<b>0.305</b>	<b>0.305</b>	0.083	0.000	0.240
HMAE	0.000	<b>1.000</b>	0.102	0.102	0.012	0.000	0.102
QLIKE	0.001	<b>1.000</b>	<b>0.342</b>	0.229	0.052	0.000	<b>0.342</b>
$M = 729$							
$h = 1$							
MSE	<b>0.874</b>	<b>0.465</b>	<b>1.000</b>	<b>0.526</b>	0.181	<b>0.874</b>	0.243
HMSE	<b>0.677</b>	<b>0.895</b>	<b>1.000</b>	<b>0.961</b>	<b>0.961</b>	0.004	0.024
HMAE	0.071	<b>1.000</b>	<b>0.826</b>	<b>0.665</b>	<b>0.665</b>	0.000	0.011
QLIKE	<b>1.000</b>	0.184	<b>0.371</b>	0.217	0.130	<b>0.371</b>	0.179
$h = 5$							
MSE	<b>0.880</b>	<b>0.405</b>	<b>0.972</b>	<b>0.501</b>	<b>1.000</b>	<b>0.485</b>	<b>0.405</b>
HMSE	<b>0.351</b>	<b>0.676</b>	<b>1.000</b>	<b>0.814</b>	0.149	0.004	0.002
HMAE	0.021	<b>1.000</b>	<b>0.980</b>	<b>0.908</b>	0.009	0.000	0.000
QLIKE	<b>0.604</b>	<b>0.408</b>	<b>1.000</b>	<b>0.589</b>	<b>0.604</b>	0.244	<b>0.491</b>
$h = 22$							
MSE	<b>0.943</b>	0.212	<b>0.455</b>	<b>0.399</b>	<b>1.000</b>	<b>0.360</b>	<b>0.279</b>
HMSE	<b>1.000</b>	0.166	<b>0.438</b>	<b>0.578</b>	0.166	0.004	0.215
HMAE	<b>1.000</b>	<b>0.349</b>	<b>0.447</b>	<b>0.578</b>	<b>0.349</b>	0.000	<b>0.433</b>
QLIKE	<b>1.000</b>	0.044	0.209	0.209	0.209	0.003	0.209

Note. Bold entries denote that the corresponding models are included in MCS for 75% confidence. We report the empirical results of the different forecasts horizons. The MSE, HMSE, HMAE and QLIKE loss functions can be found in Equations 35–38. In this table, we use the realized volatility as actual market volatility. The forecasting window is 529 (and 729) days.

$$w_t^* = \frac{1}{\gamma} \left( \frac{\bar{r}_{t+1}}{\hat{\sigma}_{t+1}^2} \right), \quad (41)$$

where  $\bar{r}_{t+1}$  is the mean of the all-sample data, and  $\hat{\sigma}_{t+1}^2$  is the rolling-window forecasting of the variance of asset excess returns, respectively. We follow the literature by restricting the optimal weight between 0 and 1.5 (i.e.,  $0 \leq w_t^* \leq 1.5$ ) to preclude short sales and prevent more than 50% leverage (Neely et al., 2014; Rapach et al., 2010).

Thus the portfolio return at day  $t + 1$  is given by

$$R_{p,t+1} = w_t^* r_{t+1} + r_{t+1,f}, \quad (42)$$

where  $r_{t+1}^*$  is the excess return; it can be obtained from  $r_{t+1}^* = r_{t+1} - r_{t+1,f}$ , and  $r_{t+1,f}$  is the risk-free return.



We employ the classical criterion of the Sharpe ratio (SR) to evaluate the performance of a portfolio constructed based on return and volatility forecasts, which is defined as

$$SR = \frac{\hat{\mu}_p}{\hat{\sigma}_p}, \quad (43)$$

where  $\hat{\mu}_p$  and  $\hat{\sigma}_p^2$  refer to the mean and volatility of excess returns, respectively.

Table 11 shows the average annualized average returns ( $R$ ) and SR of portfolios formed by out-of-sample realized volatility forecasts using the above-mentioned volatility models with the crude oil market. The empirical results of Table 11 show that in terms of one-step-ahead forecasting the new models with continuous realized semi-variances and signed jump variations can obtain higher economic benefits. The first and second largest portfolio returns are generated by all our new proposed models in the crude oil market. For example, when  $h = 1$ , CSJd-New as an investor strategy can attain the largest annualized portfolio return: approximately 3.364%; this value is significantly larger than other existing models. For the middle and long horizons, we find that the CSJd-New and CSJ-New strategies can obtain higher economic benefits than other strategies can. The best strategies are

all our proposed models. Although the two best models are not the same in different horizons, those models can attain higher economic benefits and are all our proposed models. These findings in Table 11 are very similar to the conclusions obtained in Sections 4.2 and 4.3, which further implies that our new volatility models can generate more accurate forecasts than can the existing models, both statistically and economically. Therefore, our findings are of immense importance for portfolio allocation and financial risk management.

## 5 | CONCLUDING REMARKS

In this study, we propose new variables, such as continuous realized semi-variance and signed jump variations, and use these new variables to build new HAR-type models. We forecast the realized volatility of the crude oil futures and the S&P 500 index to investigate whether our proposed models can attain higher forecast accuracy. Our findings are threefold. First, the in-sample results indicate that past negative returns have greater effects on future volatility than that of positive returns, and our new signed jump variations have a significantly negative influence on the future volatility. According to the result of the Wald test, negative and positive RS (CRS) and SJV (CSJV) present significant asymmetric effects on prospective realized volatilities. Second, forecasting the different horizons, out-of-sample results exhibit that the best forecasting models are nearly all our proposed models, implying that our models are superior to the existing HAR-RV-types in predicting future volatility. Third, our proposed models as investor strategies can also obtain higher economic benefits. Those conclusions are robust to different actual volatilities, forecasting windows, and markets. The contribution of this paper is that we further expand and enrich the traditional HAR-RV-type models and propose several new variables to build superior HAR-RV-types in modeling and forecasting volatility.

## ACKNOWLEDGMENTS

The authors are grateful to the Editor and the two anonymous referees for providing thoughtful comments and suggestions that enhanced the quality of this paper. The authors are also grateful for the financial support from the Natural Science Foundation of China through grant numbers 71701170 (Feng Ma), 71371157, and 71671145 (Yu Wei), 71771124 (Li Liu). Feng Ma is also supported by the program for Fundamental Research Funds for the Central Universities [2682017WCX01], and the Humanities and Social Science Fund of the Ministry of Education [17YJC790105].

**TABLE 11** Performance of portfolios formed by realized volatility forecasts (%)

Models	$h = 1$		$h = 5$		$h = 22$	
	$R$	SR	$R$	SR	$R$	SR
CG	−3.424	−1.190	1.950	0.717	5.481	2.377
PS1	2.563	0.885	0.507	0.185	4.659	2.053
PS2	2.853	0.986	0.728	0.265	4.713	2.074
PS3	2.831	0.988②	1.110	0.402	4.754	2.090
CSJ	0.470	0.160	2.142	0.785	5.936	2.646
CSJd	−0.608	−0.210	1.079	0.400	5.879	2.542
CG-New	2.091	0.700	2.183	0.774	4.764	2.085
PS1-New	0.286	0.096	−0.148	−0.053	4.555	1.988
PS2-New	2.293	0.773	0.081	0.029	4.593	2.004
PS3-New	2.917②	0.982	−0.184	−0.066	4.686	1.951
CSJ-New	2.602	0.920	3.752①	1.465①	7.406②	3.391①
CSJd-New	3.364①	1.156①	2.743②	1.027②	7.716①	3.277②

*Note.* This table shows the performance of portfolios formed by realized volatility forecasts. We give the mean excess return ( $R$ ) and Sharpe ratio (SR) of each portfolio. All numbers are multiplied by 100. We only report the risk aversion. All the values are annualized. In each column with superscript values, i.e., ① and ②, denote the largest and second largest  $R$  and SR, respectively. The length of forecasting windows is 629. Other windows' results are the similar to the length of 629 days.

## ORCID

Feng Ma  <http://orcid.org/0000-0002-6539-0965>

## REFERENCES

- Agnolucci, P. (2009). Volatility in crude oil futures: A comparison of the predictive ability of GARCH and implied volatility models. *Energy Economics*, 31, 316–321.
- Aloui, C., & Jammazi, R. (2009). The effect of crude oil shocks on stock market shifts behavior: A regime switching approach. *Energy Economics*, 31, 789–799.
- Andersen, T. G., & Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, 39, 885–905.
- Andersen, T. G., Bollerslev, T., & Diebold, F. X. (2007). Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. *Review of Economics and Statistics*, 89, 701–720.
- Andersen, T. G., Bollerslev, T., Diebold, F. X., & Ebens, H. (2001). The distribution of realized stock return volatility. *Journal of Financial Economics*, 61, 43–76.
- Audrino, F., & Hu, Y. (2016). Volatility forecasting: Downside risk, jumps and leverage effect. *Econometrics*, 4(1), 8.
- 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. *The Econometrics Journal*, 12, 1–32.
- Barndorff-Nielsen, O. E., Kinnebrock, S., & Shephard, N. (2010). Measuring downside risk: realised semi-variance. In T. Bollerslev, J. Russell, & M. Watson (Eds.), *Volatility and time series econometrics: Essays in honor of Robert F. Engle* (pp. 117–137). Oxford, UK: Oxford University Press.
- Barndorff-Nielsen, O. E., & Shephard, N. (2004). Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics*, 2, 1–37.
- Barndorff-Nielsen, O. E., & Shephard, N. (2006). Econometrics of testing for jumps in financial economics using bipower variation. *Journal of Financial Econometrics*, 4, 1–30.
- Baruník, J., Kočenda, E., & Vácha, L. (2016). Asymmetric connectedness on the US stock market: Bad and good volatility spillovers. *Journal of Financial Markets*, 27, 55–78.
- Bekaert, G., & Hoerova, M. (2014). The VIX, the variance premium and stock market volatility. *Journal of Econometrics*, 183, 181–192.
- Biais, B., Glosten, L., & Spatt, C. (2005). Market microstructure: A survey of microfoundations, empirical results, and policy implications. *Journal of Financial Markets*, 8, 217–264.
- Bollerslev, T., Patton, A. J., & Quaedvlieg, R. (2015). Exploiting the errors: A simple approach for improved volatility forecasting. *Journal of Econometrics*, 192, 1–18.
- Busch, T., Christensen, B. J., & Nielsen, M. Ø. (2011). The role of implied volatility in forecasting future realized volatility and jumps in foreign exchange, stock, and bond markets. *Journal of Econometrics*, 160, 48–57.
- Campbell, J. Y., & Thompson, S. B. (2008). Predicting the equity premium out of sample: Can anything beat the historical average? *Review of Financial Studies*, 21, 1509–1531.
- Carnero, M. A., Peña, D., & Ruiz, E. (2004). Persistence and kurtosis in GARCH and stochastic volatility models. *Journal of Financial Econometrics*, 2(2), 319–342.
- Ceylan, Ö. (2014). Time-varying volatility asymmetry: A conditioned HAR-RV (CJ) EGARCH-M model. *Journal of Risk*, 17(2), 21–49.
- Charles, A., & Darné, O. (2014). Volatility persistence in crude oil markets. *Energy Policy*, 65, 729–742.
- Chen, X., & Ghysels, E. (2010). News—good or bad—and its impact on volatility predictions over multiple horizons. *Review of Financial Studies*, 24(1), 46–81.
- Chevallier, J., & Sévi, B. (2012). On the volatility–volume relationship in energy futures markets using intraday data. *Energy Economics*, 34, 1896–1909.
- Clark, T. E., & McCracken, M. W. (2001). Tests of equal forecast accuracy and encompassing for nested models. *Journal of Econometrics*, 105(1), 85–110.
- Clark, T. E., & West, K. D. (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics*, 138(1), 291–311.
- Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7, 174–196.
- Corsi, F., Pirino, D., & Reno, R. (2010). Threshold bipower variation and the impact of jumps on volatility forecasting. *Journal of Econometrics*, 159, 276–288.
- Degiannakis, S., & Filis, G. (2017). Forecasting oil price realized volatility using information channels from other asset classes. *Journal of International Money and Finance*, 79, 28–49.
- Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13(3), 253–263.
- Duong, D., & Swanson, N. R. (2015). Empirical evidence on the importance of aggregation, asymmetry, and jumps for volatility prediction. *Journal of Econometrics*, 187(2), 606–621.
- Efimova, O., & Serletis, A. (2014). Energy markets volatility modelling using GARCH. *Energy Economics*, 43, 264–273.
- Fleming, J., Kirby, C., & Ostdiek, B. (2003). The economic value of volatility timing using “realized” volatility. *Journal of Financial Economics*, 67, 473–509.
- Guidolin, M., & Na, C. F. (2006). *The economic and statistical value of forecast combinations under regime switching: An application to predictable US returns* (Working Paper Series 2006-059). St. Louis, MO: Federal Reserve Bank of St. Louis.
- Hamilton, J. D. (1983). Oil and the macroeconomy since World War II. *Journal of Political Economy*, 91, 228–248.
- Hamilton, J. D. (2003). What is an oil shock? *Journal of Econometrics*, 113, 363–398.

- Hansen, P. R. (2005). A test for superior predictive ability. *Journal of Business and Economic Statistics*, 23, 365–380.
- Hansen, P. R., Lunde, A., & Nason, J. M. (2011). The model confidence set. *Econometrica*, 79, 453–497.
- Haugom, E., Langeland, H., Molnár, P., & Westgaard, S. (2014). Forecasting volatility of the US oil market. *Journal of Banking and Finance*, 47, 1–14.
- Haugom, E., Lien, G., Veka, S., & Westgaard, S. (2014). Covariance estimation using high-frequency data: Sensitivities of estimation methods. *Economic Modelling*, 43, 416–425.
- Haugom, E., Veka, S., Lien, G., & Westgaard, S. (2014). Estimating and evaluating Value-at-Risk forecasts based on realized variance: Empirical evidence from ICE Brent Crude oil futures. *OPEC Energy Review*, 38(4), 373–397.
- Jarque, C. M., & Bera, A. K. (1987). A test for normality of observations and regression residuals. *International Statistical Review/Revue Internationale de Statistique*, 55(2), 163–172.
- Kilian, L. (2009). Not all oil price shocks are alike: disentangling demand and supply shocks in the crude oil market. *American Economic Review*, 99, 1053–1069.
- Kilian, L., & Park, C. (2009). The impact of oil price shocks on the U. S. stock market. *International Economic Review*, 50, 1267–1287.
- 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. *Journal of Empirical Finance*, 12, 445–475.
- Laurent, S., Rombouts, J. V., & Violante, F. (2012). On the forecasting accuracy of multivariate GARCH models. *Journal of Applied Econometrics*, 27, 934–955.
- 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. *Economic Modelling*, 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. *Journal of Econometrics*, 187, 293–311.
- Ljung, G. M., & Box, G. E. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65, 297–303.
- Ma, F., Wahab, M. I. M., Huang, D., & Xu, W. (2017). Forecasting the realized volatility of the oil futures market: A regime switching approach. *Energy Economics*, 67, 136–145.
- Madhavan, A. (2000). Market microstructure: A survey. *Journal of Financial Markets*, 3, 205–258.
- Martens, M., van Dyke, D., & de Pooter, M. (2009). Forecasting S&P 500 volatility: Long memory, level shifts, leverage effects, day-of-the-week seasonality, and macroeconomic announcements. *International Journal of Forecasting*, 25, 282–303.
- McCracken, M. W. (2007). Asymptotics for out of sample tests of Granger causality. *Journal of Econometrics*, 140(2), 719–752.
- Neely, C. J., Rapach, D. E., Tu, J., & Zhou, G. (2014). Forecasting the equity risk premium: The role of technical indicators. *Management Science*, 60, 1772–1791.
- Nomikos, N. K., & Pouliasis, P. K. (2011). Forecasting petroleum futures markets volatility: The role of regimes and market conditions. *Energy Economics*, 33, 321–337.
- Patton, A. J., & Sheppard, K. (2015). Good volatility, bad volatility: Signed jumps and the persistence of volatility. *Review of Economics and Statistics*, 97, 683–697.
- Prokopczuk, M., Symeonidis, L., & Wese Simen, C. (2016). Do jumps matter for volatility forecasting? Evidence from energy markets. *Journal of Futures Markets*, 36(8), 758–792.
- Rapach, D. E., Strauss, J. K., & Zhou, G. (2010). Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *Review of Financial Studies*, 23, 821–862.
- Rossi, B., & Inoue, A. (2012). Out-of-sample forecast tests robust to the choice of window size. *Journal of Business and Economic Statistics*, 30(3), 432–453.
- Sévi, B. (2014). Forecasting the volatility of crude oil futures using intraday data. *European Journal of Operational Research*, 235, 643–659.
- Wang, Y., Ma, F., Wei, Y., & Wu, C. (2016). Forecasting realized volatility in a changing world: A dynamic model averaging approach. *Journal of Banking and Finance*, 64, 136–149.
- Wang, Y., Pan, Z., & Wu, C. (2017). Time-varying parameter realized volatility models. *Journal of Forecasting*, 36(5), 566–580.
- Wei, Y., Wang, Y., & Huang, D. S. (2010). Forecasting crude oil market volatility: Further evidence using GARCH-class models. *Energy Economics*, 32, 1477–1484.
- Welch, I., & Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, 21(4), 1455–1508.
- West, K. D. (1996). Asymptotic inference about predictive ability. *Econometrica*, 64(5), 1067–1084.
- Zhang, L., Mykland, P. A., & Ait-Sahalia, Y. (2005). A tale of two time scales. *Journal of the American Statistical Association*, 100, 1394–1411.
- Zivot, E., & Wang, J. (2005). *Modelling financial time series with S-PLUS* (2nd ed.). Berlin, Germany: Springer.

**Feng Ma** is an Assistant Professor in School of Economics & Management at Southwest Jiaotong University. His research interests include time series forecasting, empirical finance and energy economics.

**Yu Wei** is a Professor in School of Finance at Yunnan University of Finance and Economics. His research interests include time series forecasting, empirical finance and energy economics.

**Li Liu** is an Associate Professor in School of Finance at Nanjing Audit University. Her research interests lie in the area of time series forecasting, empirical finance and energy economics.

**Dengshi Huang** is a Professor in School of Economics & Management at Southwest Jiaotong University. His research interests lie in the area of time series forecasting, empirical finance and corporate finance.

**How to cite this article:** Ma F, Wei Y, Liu L, Huang D. Forecasting realized volatility of oil futures market: A new insight. *Journal of Forecasting*. 2018;1–18. <https://doi.org/10.1002/for.2511>

## APPENDIX

In this study, we use the S&P 500 and CAC 40 index as our research object and investigate whether our newly proposed model can attain higher forecast accuracy. Our data are collected by Thomson Reuters Tick History Database. To save space, we only use the MCS test to evaluate the forecasting performance between the existing and extended models. Tables A1 and A2 exhibit the results of the MCS test using the RV benchmark volatility and the technique of the out-of-sample method. The empirical results tell us that, when forecasting short-term horizons, the MCS  $p$ -values of CG-New and CSJd-New models are

larger than the 10% significance level under the MSE, HMSE, HMAE, and QLIKE loss functions. This finding shows that those models have better performance in forecasting the volatility for which the results are consistent with the crude oil market. The CSJd-New model has relatively better performance in forecasting the middle-term horizons' volatility of the S&P 500 under the four loss functions. In addition, from the CAC 40 index, we can observe that the  $p$ -values of all new models are larger than 10%, implying that our models can attain higher forecast accuracy. CG-New, PS1-New and PS2-New models are better in forecasting the oil volatility at the 10% significance level. We can observe that, although the best forecasting models are different in different markets, the best are among our new models. For the 22 steps ahead, the PS2-New model can pass the MCS test under all loss functions in the three markets, which shows that PS2-New is a satisfactory forecasting model of the S&P 500, CAC 40 and oil price. Finally, we further find that nearly all our proposed models are the best forecasting models; therefore, we provide new insights to increase the models' forecasting accuracy.

**TABLE A1** MCS of out-of-sample forecasts in different out-of-sample windows (S&P 500)

	$h = 1$				$h = 5$				$h = 22$			
	MSE	HMSE	HMAE	QLIKE	MSE	HMSE	HMAE	QLIKE	MSE	HMSE	HMAE	QLIKE
CG	<b>0.941</b>	0.064	<b>0.415</b>	<b>0.465</b>	0.216	0.099	0.038	<b>0.855</b>	0.004	0.084	0.008	0.004
PS1	<b>0.729</b>	0.064	0.011	<b>0.412</b>	<b>0.430</b>	0.099	0.038	<b>1.000</b>	0.020	0.060	0.003	0.172
PS2	<b>0.729</b>	0.064	0.011	<b>0.412</b>	<b>0.375</b>	0.099	0.036	<b>0.848</b>	0.124	0.079	0.139	<b>0.354</b>
PS3	<b>0.827</b>	0.043	0.011	0.156	<b>0.830</b>	0.099	0.038	<b>0.855</b>	0.124	<b>0.269</b>	<b>0.525</b>	0.172
CSJ	<b>0.707</b>	0.043	0.006	<b>0.290</b>	<b>0.430</b>	0.099	0.032	<b>0.638</b>	0.145	<b>0.269</b>	<b>0.525</b>	0.172
CSJd	<b>0.941</b>	0.064	0.050	0.248	0.217	0.099	0.057	<b>0.734</b>	0.032	0.078	0.005	0.036
CG-New	<b>0.783</b>	0.186	<b>0.415</b>	<b>0.290</b>	<b>0.830</b>	0.099	0.057	<b>0.855</b>	0.145	<b>0.269</b>	0.105	0.172
PS1-New	<b>1.000</b>	0.043	0.006	<b>1.000</b>	<b>0.830</b>	0.099	0.038	<b>0.855</b>	0.021	0.153	0.054	0.012
PS2-New	<b>0.707</b>	0.064	0.011	<b>0.284</b>	<b>0.830</b>	0.099	0.038	<b>0.855</b>	<b>1.000</b>	<b>0.572</b>	<b>1.000</b>	<b>1.000</b>
PS3-New	<b>0.723</b>	0.186	0.050	0.171	<b>0.830</b>	0.099	0.038	<b>0.944</b>	0.145	<b>0.298</b>	<b>0.525</b>	0.172
CSJ-New	<b>0.723</b>	0.186	0.050	0.193	<b>1.000</b>	0.099	0.038	<b>0.944</b>	0.145	<b>0.269</b>	<b>0.525</b>	0.172
CSJd-New	<b>0.941</b>	<b>1.000</b>	<b>1.000</b>	0.156	<b>0.592</b>	<b>1.000</b>	<b>1.000</b>	<b>0.944</b>	0.016	<b>1.000</b>	<b>0.525</b>	0.022

*Note.* Bold entries denote that the corresponding models are included in MCS for 75% confidence. We report the empirical results of the different forecasts horizons. The MSE, HMSE, HMAE, and QLIKE loss functions can be found in Equations 35–38. In this table, we use the realized kernel as actual market volatility. All models are forecasted using the 5-minute data of the S&P 500 index.

**TABLE A2** MCS of out-of-sample forecasts in different out-of-sample windows (CAC 40)

	$h = 1$				$h = 5$				$h = 22$			
	MSE	HMSE	HMAE	QLIKE	MSE	HMSE	HMAE	QLIKE	MSE	HMSE	HMAE	QLIKE
CG	<b>0.811</b>	0.219	<b>0.684</b>	<b>1.000</b>	<b>0.288</b>	0.127	0.045	<b>0.994</b>	<b>0.648</b>	0.232	0.140	<b>0.857</b>
PS1	<b>0.770</b>	<b>0.307</b>	<b>0.872</b>	<b>0.839</b>	<b>0.288</b>	<b>0.268</b>	<b>0.384</b>	<b>1.000</b>	<b>0.482</b>	<b>0.430</b>	<b>0.667</b>	<b>0.574</b>
PS2	<b>0.383</b>	0.161	0.100	<b>0.821</b>	0.087	0.098	0.014	<b>0.540</b>	<b>0.289</b>	<b>0.293</b>	0.105	<b>0.519</b>
PS3	<b>0.544</b>	0.142	0.226	<b>0.737</b>	0.089	0.120	0.018	<b>0.453</b>	<b>0.318</b>	<b>0.432</b>	<b>0.253</b>	<b>0.574</b>
CSJ	<b>0.258</b>	0.154	0.153	<b>0.839</b>	0.106	0.098	0.008	<b>0.574</b>	<b>0.353</b>	<b>0.384</b>	0.095	<b>0.574</b>
CSJd	<b>0.688</b>	0.174	<b>0.516</b>	<b>0.839</b>	0.026	0.199	0.027	0.183	0.039	<b>0.345</b>	0.054	0.087
CG-New	<b>0.934</b>	0.307	<b>0.888</b>	<b>0.839</b>	<b>1.000</b>	0.199	<b>0.384</b>	<b>1.000</b>	<b>0.849</b>	<b>0.278</b>	0.123	<b>0.709</b>
PS1-New	<b>0.811</b>	<b>1.000</b>	<b>1.000</b>	<b>0.795</b>	<b>0.565</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>0.962</b>	<b>0.687</b>	<b>0.689</b>	<b>1.000</b>
PS2-New	<b>1.000</b>	0.219	<b>0.888</b>	<b>0.795</b>	<b>0.311</b>	0.199	<b>0.384</b>	<b>0.574</b>	<b>0.962</b>	<b>0.460</b>	<b>0.327</b>	<b>0.857</b>
PS3-New	<b>0.845</b>	0.222	<b>0.888</b>	<b>0.839</b>	<b>0.311</b>	0.199	<b>0.384</b>	<b>0.634</b>	<b>0.962</b>	<b>0.687</b>	<b>0.689</b>	<b>0.949</b>
CSJ-New	<b>0.837</b>	0.222	<b>0.888</b>	<b>0.795</b>	<b>0.234</b>	0.199	0.217	<b>0.485</b>	<b>1.000</b>	<b>0.687</b>	<b>0.659</b>	<b>0.844</b>
CSJd-New	<b>0.934</b>	0.143	<b>0.870</b>	<b>0.839</b>	0.106	0.199	0.174	0.217	0.007	<b>1.000</b>	<b>1.000</b>	0.026

*Note.* Bold entries denote that the corresponding models are included in MCS for 75% confidence. We report the empirical results of the different forecasts horizons. The MSE, HMSE, HMAE, and QLIKE loss functions can be found in Equations 35–38. In this table, we use the realized kernel as actual market volatility. All models are forecasted using the 5-minute data of the CAC 40 index.