Forecast on Oil Futures Volatility Based on HAR-RV Model and HAR-RV-SW Model

Nuo Chen[1,*,†] Xinyi Chen[2,†]

[1]*Department of Mechanical, Aerospace and Civil Engineering, The University of Manchester, Manchester, M139PL, The United Kingdom*
[2]*Department of International Economics and Trade, University of International Business and Economics, Beijing, 100029, China*
*Corresponding author. amissnuochen@126.com, b201814004@uibe.edu.cn.
†These authors contribute equally.

ABSTRACT

Crude oil is an important chemical raw material and energy source, and its price is very crucial to the cost of industrial development. Therefore, the forecast of crude oil prices is also very meaningful. This paper studies the correlation between the actual daily, weekly, and monthly volatility and the forecast volatility by establishing a HAR-RV model. Through the establishment of this model, we study the influence of the weekday effect on the forecast of volatility and find that the forecast of daily volatility is positively correlated with weekly volatility, while daily, weekly and monthly volatilities have a positive correlation with the forecast of weekly volatility. Besides, we have also established the HAR-RV-SW model and obtained the forecast of the volatility of daily, weekly, and monthly has a day-to-the-week effect. This can help us predict crude oil futures prices through historical data.

Keywords: crude oil futures, HAR-RV model, HAR-RV-SW model, parameter estimation, volatility forecasting

1. INTRODUCTION

With the rapid development of China's economy and the continuous upgrading of industries, crude oil has received more and more attention as one of the important chemical raw materials. Crude oil is not only an indispensable basic raw material for industry and manufacturing but also closely related to people's daily lives. It has a vital impact on the national economy. Under the influence of the opening-up policy, the domestic market is in line with the international market. Thus, crude oil is increasingly affected by the international market. According to relevant data, China's crude oil dependence on foreign countries reached 70% in 2018 [1]. In 2019, crude oil imports exceeded 500 million tons, with a year-on-year increase of 9.5% [2]. With the continuous growth of China's economy, the demand for crude oil will further increase, and the impact on the economy will also increase.

As a widely used material, crude oil has many market participants and complex market conditions. It is affected by prices, supply, and demand in the upstream and downstream industries. Besides, because of its characteristics, crude oil is also subject to geopolitics, market speculation, and weather conditions. Due to the influence of many factors on the crude oil market, it has the characteristics of large price volatility and volatility clustering effect. Although a batch of crude oil futures has been launched in China to avoid the risk of price fluctuations, the Chinese market is still limited. And the relevant laws, regulations, and policies are not mature and complete. Therefore, the forecast of crude oil futures prices has important practical significance for the stable development of China's industrial sectors and enterprises, as well as the stable operation of the country's macro-economy.

Due to the important impact of crude oil on the economy, the forecast of crude oil futures prices has never stopped. In the early years, due to technical difficulties in the collection of high-frequency data, market research was mainly based on low-frequency data. Therefore, GARCH models were mostly used for analysis. However, due to the availability of high-frequency data and the impact of high-frequency data on the market A better degree of response makes the study of high-frequency data more meaningful, so the GARCH model is no longer suitable for current research. Andersen et al. constructed a VAR-RV model

by combining the conditional covariance matrix with the realized volatility and found that its prediction accuracy for future volatility is better than the traditional GARCH model; The advantage of the VAR model is that it establishes a forecasting system, and one of the advantages of the ARMA model is that it can predict future prices based on historical data only, but neither model takes into account the influence of price fluctuations. Heterogeneous autoregressive models based on high-frequency data can make up for the shortcomings of the above models; Two of the HAR models performed better [3].

The RV component can be added to the GARCH, SV, and ARFIMA models to construct a new model and use this to predict the daily changes of the Standard & Poor's 100 (S&P100) stock index series. They analyzed the transaction data of the past 7 years (one-by-one). Trading), the experimental results show that the new model's predictive ability for volatility is obviously higher than that of the traditional model, and it is found that the long-term memory model seems to provide more accurate predictions [4]. It was found that the HAR-RV model has better out-of-sample predictions than the FARIMA model through the modeling analysis of the volatility of the Shanghai stock index [5]. At the same time, the HAR-RV model can well describe the driving characteristics of long memory of volatility; The 5-minute high-frequency data of the Shanghai and Shenzhen 300 stock index futures simulation trading was used as an example to discuss and compare the prediction capabilities of various mainstream volatility models on their out-of-sample daily volatility and found that the prediction accuracy of HV was significantly lower than that of the RV model [6]. The GARCH-RV model cannot describe and predict the volatility of futures satisfactorily; The HAR-RV model is better than the ARFIMA-RV model by testing the long memory of the volatility of the Shanghai stock index return [7]; The ARCH family model and the HAR-RV model were used to compare the predicted volatility and actual volatility of four major global stock indexes in and out of the sample, and found that the HAR-RV model is better than the ARCH family model [8]; Numerous research results show that the analysis of high-frequency data can more accurately predict price changes in the futures market, and the RV and RV-SW models have higher accuracy and stability in the prediction of future price fluctuations. Besides, the current situation for the RV - SW model has less application analysis, so in this paper, we choose to use RV and RV-SW model to analyze historical high-frequency data and predict future changes in crude oil futures prices based on the correlation between historical prices and future prices.

This paper uses the baseline model HAR-RV and the improved HAR-RV-SW model with weekly effect to predict the price changes in the crude oil futures market.

Then, by comparing the original model and the adjusted model, the stability of the model prediction results is compared. The results of sample analysis show that compared with the traditional model, the HAR-RV-SW model can owe better prediction analysis on volatility, that is, the prediction results are more stable and more accurate. Finally, we show a new perspective. Previous studies mainly related to the silver wave prediction using VAR model based on low frequency data model, GARCH model and ARMA model, but this paper chose five minutes of high frequency data and HAR model is used to predict the volatility of crude oil futures price, taking into account factors such as Sunday effect.

In the second section, the RV value is calculated and analyzed by using 5-minute high-frequency data from April 11, 2018 to January 5, 2021. In the third section, we use the model HAR-RV and the model HAR-RV-SW to fit the data; In section 4, we analyze the data and discuss the influence of day-of-the-week effect on monthly, weekly and daily RV. Finally, we draw conclusions, suggestions, and limitations based on the results of the analysis.

2. DATA

In this paper, we select 5-minute frequency trading data from Global Financial Information Service (GFIS) to measure the price volatility of the oil futures. The sample period is from April 11, 2018 to January 5, 2021 and a total of 68,943 items of oil futures data are obtained from April 11, 2018 to December 18, 2020. After calculating RV and eliminating null values, a total of 767 items of data are obtained from April 11, 2018 to December 18, 2020.

In Figures 1 to 3, the RV value of crude oil futures prices from April 11, 2018 to December 11, 2020 is calculated and the null value is eliminated, and then the fluctuation chart of crude oil futures RV value is drawn on a daily, weekly, and monthly basis.
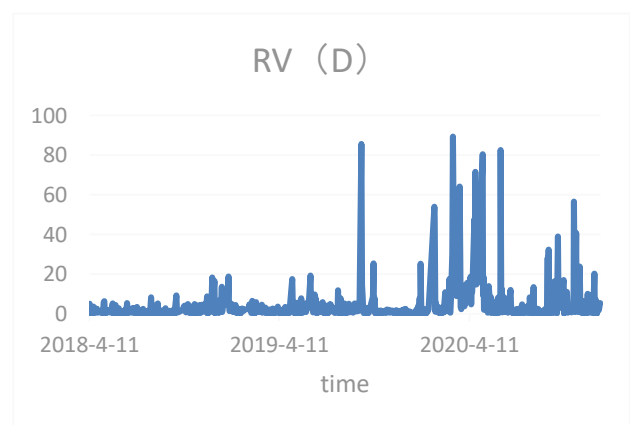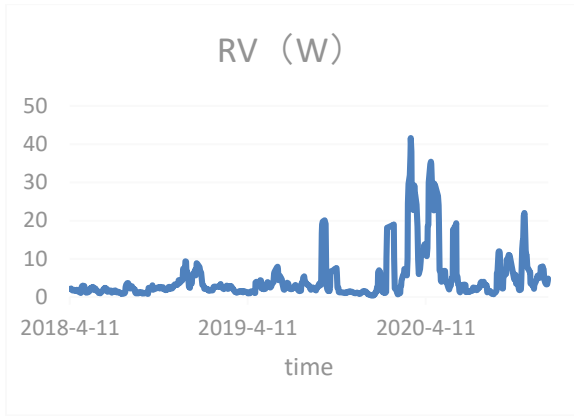

Figure 1 Daily RV of oil future
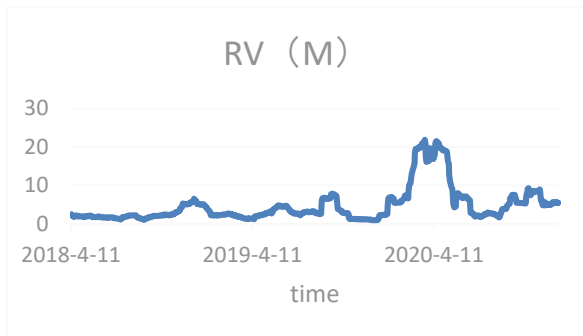
Figure 2 Weekly RV of oil future

Figure 3 Monthly RV of oil future

Figure 1 shows, from April 1, 2018 to August 11, 2019, crude oil prices were low and relatively stable, with the exception of November 11, 2018, solstice, January 11, 2019, and between April 11, 2019 and July 11, 2019, when the price of crude oil fluctuated significantly. Between September 11, 2019 solstice and July 11, 2020, with the exception of October 11, 2019 solstice and December 11, when prices were stable and low, crude oil prices were not only highly volatile and average high, but also peaked on March 11, 2020. In the following period, crude oil prices remained volatile, but the degree of volatility and prices were lower than the previous period.

As can be seen from Figure 2, the oil futures market has volatility aggregation and asymmetry, and volatility also exists. As the volatility of oil futures price is significantly different over time, in some periods, the volatility is significantly greater in some periods than in others.

In the third picture, it is obvious that the price of crude oil futures fluctuated the most from December 11, 2019 to June 11, 2020 and reached a peak. From October 11, 2018 to February 11, 2019, from April 11, 2019 to October 11, 2019, and from August 11, 2020 to December 11, 2020, there are relatively small fluctuations with similar volatility. In the rest of the time, the price of crude oil futures is relatively stable.

Next, we build a HAR regression model to analyze the realized volatility in detail.

3. ECONOMETRIC MODELS

This section will predict the volatility of the silver futures market. We will introduce two HAR-type models, including the baseline model HAR-RV, and the improved HAR-RV-SW model with weekly effects. In addition, this section also compares the original model and the Adjusted R-squares of HAR-RV and HAR-RV-SW model to compare the stability of the model's prediction results.

This section will predict the volatility of the silver futures market. We will introduce two HAR-type models, including the baseline model HAR-RV, and the improved HAR-RV-SW model with weekly effects. In addition, this section also compares the original model and the Adjusted R-squares of HAR-RV and HAR-RV-SW model to compare the stability of the model's prediction results.

3.1. HAR-RV model

HAR-RV model is better than ARFIMA-RV model in prediction [9]. To use it, we can divide a day into N part, and is the closing price in the trading day is . The logarithmic rate of return for is:

$$r_{t,i} \qquad (1)$$
$$= (lnP_{t,i}$$
$$- lnP_{t,i-1}) \times 10$$

The RV of the trading day t is , which can be calculated as:

$$RV_t^d = \Sigma_{i=1}^N r_{t,i}^2 \qquad (2)$$

Weekly RV of the trading day t is : which can be represented as:

$$RV_t^w \qquad (3)$$
$$= \frac{RV_t^d + RV_{t-1}^d + RV_{t-2}^d + RV_{t-3}^d + RV_{t-4}^d + RV_{t-}^d}{6}$$

Monthly RV of the trading day t is : which can be represented as:

$$RV_t^m = \frac{RV_t^d + RV_{t-1}^d + RV_{t-2}^d + \cdots + RV_{t-24}^d}{25} \qquad (4)$$

The average RV from day t to (t+ H) is , which can be calculated as:

$$\overline{RV}_{t+H} = \frac{1}{H} \Sigma_{i=1}^N RV_{t+i}^d \qquad (5)$$

Logarithmic form of RV is more stable, and that of HAR-RV model is stable, too. This model can be expressed as:

$$ln\overline{RV}_{t+H} = \beta_0 + \beta_d lnRV_t^d \qquad (6)$$
$$+ \beta_w lnRV_t^w$$
$$+ \beta_m lnRV_t^m$$

3.2. HAR-RV-SW model

So, according to the research, day-of-the-week effect is existed in future market. That is, the volatility of some specific dates has a significant correlation with the forecast results, and we can better improve the model through the day-of-the-week effect.

So, an HAR-RV-SW model can be established as follow:

$$ln\overline{RV}_{t+H} = \beta_0 + \beta_d lnRV_t^d + \beta_w lnRV_t^w \qquad (7)$$
$$+ \beta_m lnRV_t^m + \theta_1 D_1 + \theta_2 D_2$$
$$+ \theta_3 D_3 + \theta_4 D_4 + \theta_5 D_5$$

$D_1 D_2, D_3, D_4, D_5$ are dummy variables that represent Monday, Tuesday, Wednesday, Thursday and Friday, respectively.

If $D_1 = 1$ while $D_2, D_3, D_4, D_5 = 0$, then, it represents for Monday. If $D_1 = 1$ while $D_2, D_3, D_4, D_5 = 0$, then, it represents for Monday. And so on for the rest.

4. IN-SAMPLE ANALYSIS

For the empirical analysis, we used secondary data to fit the model. The scope of the data is crude oil futures data from April 1, 2018 to July 11, 2019. We first calculated the daily, weekly and monthly returns based on equations (2) (3) (4). The HAR-RV model and

HAR-RV-SW model were then fitted separately using Stata software. The fitted results were also analyzed.

4.1. Summary statistics

According to the chart below, from the descriptive statistical analysis of each variable (Table 1), it can be seen that the daily volatility of oil futures is between 0.05-89.38 and the standard deviation is the largest, that is, compared with the weekly volatility and the mean of monthly volatility is the largest and the volatility is the most severe. Moreover, the weekly volatility sequence of the futures has greater volatility and severity than monthly volatility.

Table1: Descriptive statistics of variables.

Variable	Mean	Std. dev.	Min	Max
RV_t^d	4.560415543	9.287477383	0.053080362	89.38089943
RV_t^w	4.554067797	5.976845667	0.38	41.62
RV_t^m	4.564046683	4.40521648	0.909853744	21.84174199

4.2. Parameter estimations

Table 2 shows the result of the parameter estimates of the HAR-RV model and HAR-RV-SW model. The results of HAR-RV model shows that the forecasts of daily and weekly volatility have a significant correlation with daily, weekly and monthly volatility. And the forecast of monthly volatility is significantly correlated with weekly and monthly volatility.

Table2: The result of parameter estimations of the HAR-RV model and HAR-RV-SW model

	HAR-RV			HAR-RV-SW		
	1-day	1-week	1-month	1-day	1-week	1-month
β_0	0.076637	-0.0230879	0.0259071	0.3325357	0.1622549	-0.0719833
$\ln RV_t^d$	-0.134711***	0.0447585***	-0.0006952	-0.1369577***	0.048168***	-0.0018369

$\ln RV_t^w$	1.302364***	0.8207817***	0.0099942*	1.302574**	0.8186963***	0.0109243*
$\ln RV_t^m$	-0.184120**	0.1387314***	0.9841611***	-0.1808798**	0.1388945***	0.9836172***
D_1				0.3939818	-0.2895279	0.1284165**
D_2				0.3279931	-0.4427202*	0.0644484
D_3				-0.6981607	-0.3985377*	0.1537103**
D_4				0.5290128	0.0421291	0.0756812
D_5				-2.118472**	-0.0294609	0.1713893***

Note: t statistics in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Analysis of parameter estimations of RV model

Daily and monthly volatility show a negative correlation with the forecast with daily volatility while weekly volatility shows a positive correlation with it. However, these three dependent variables are positively correlated with weekly volatility. And the forecast of the monthly figure is positively correlated with monthly volatility.

Analysis of parameter estimations of RV-SW model

We also find that Monday is positively correlated with the monthly volatility compared to Saturday, that is, there will be a significant increase in the monthly volatility of oil futures on Monday compared to Saturday. Similarly, Wednesday and Friday compared to Saturday and monthly volatility is also positively correlated. That is, Wednesday and Friday compared to Saturday's monthly volatility of oil futures will have a significant increase.

For weekly volatility, Tuesday and Wednesday are negatively correlated. Also, Friday is significant and negatively correlated with daily volatility.

For daily volatility, only Friday is significantly and negatively correlated.

4.3. Comparison of the in-sample fitting capacity

Table3: Adjusted R-squares of HAR-RV model and HAR-RV-SW model

	1-day	1-week	1-month
HAR-RV	0.4691	0.9058	0.9881
HAR-RV-SW	0.4757	0.9063	0.9882

The bigger the Adjusted R-square of the model is, the more stable the model is. That is, the prediction

result is more stable and more accurate, so we can know that the HAR-RV-SW model can predict and analyze the volatility better than the HAR-RV model.
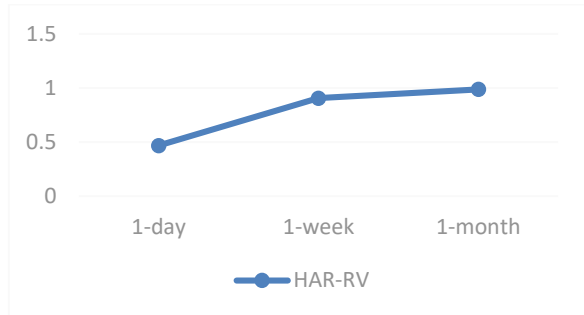

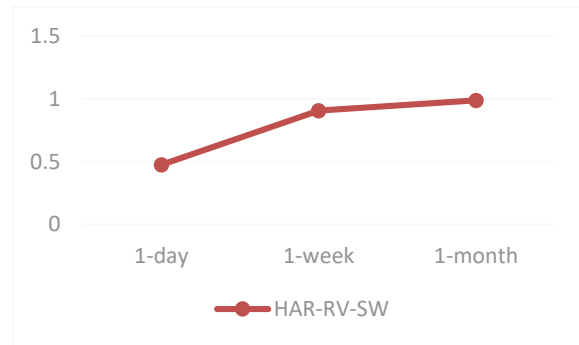**Figure 4.** Comparison of adj. R2 between HAR-RV

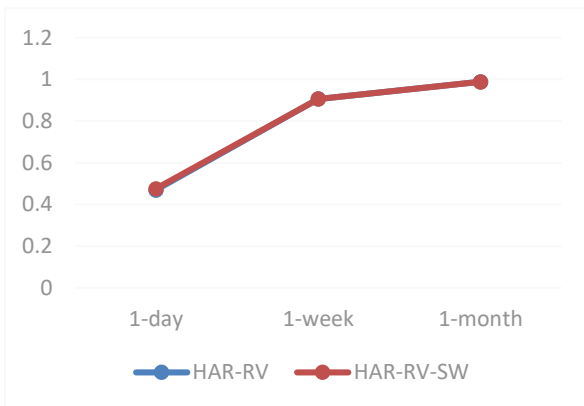**Figure 5.** Comparison of adj. R2 between HAR-RV-SW

Figure 6. Comparison of adj. R2 between HAR-RV, HAR-RV-SW

5. Conclusion

According to the efficient market theory, we can know that the futures market is theoretically effective. But in fact, many futures markets are invalid, that is, there is a correlation between the future prices and the historical price of oil. Judging from the high-frequency data of 5-min crude oil futures, it is characterized by high volatility, and the futures market is not effective. Therefore, we hope to study how to use historical data to predict the future price of oil.

We establish the RV model and the RV-SW model to study the correlation between daily, weekly, and monthly volatility and forecast of volatility and the impact of data on special days on the forecast of volatility. In conclusion, the forecast of daily volatility is positively correlated with weekly volatility, while daily, weekly and monthly volatilities have a positive correlation with the forecast of weekly volatility. For the forecast of monthly volatility, it is positively correlated with monthly figures.

For the empirical analysis, it can be seen that: 1. The predicted daily volatility and daily and monthly volatility are negatively correlated and positively correlated with weekly volatility. 2. The predicted weekly volatility and triple fold are both significantly positively correlated. 3. The predicted monthly volatility is positively correlated with weekly volatility and monthly volatility.

For the empirical analysis of day-of-the-week effects, it can be concluded that: 1. Predicted daily volatility is significantly negatively correlated with Friday. 2. Predicted weekly volatility is significantly negatively correlated with Tuesday and Wednesday. 3. Predicted monthly volatility is significantly negatively correlated with Monday, Wednesday and Friday. However, this method has limitations. The model is established based on their compound linear regression model, but they may not be linear, so in the future model establishment, the quadratic term can be added to fit and find the most suitable predictive model. Besides, the model only considers historic data and does not consider the impact of changes in fundamental information on forecasts. In the future, CJ statistics can be added to improve the model in order to make better predictions.

REFERENCES

[1]. Liu, P.P. & Zhang, X.S. (2019). Crude oil chemical futures price forecasting based on fuzzy time series model. Times Business & Economics (18), 10-11. doi:10.19463/j.cnki.sdjm.2019.18.003. (In Chinese)

[2]. Ge, X.B., Zhang, C & Zhou, X.M. (2020). A study on risk measurement of crude oil futures based on extreme value theory. Price Theory and Practice (08). doi:10.19851/j.cnki.CN11-1010/F.2020.08.322. (In Chinese)

[3]. Li, W.L., Cheng, Y.X. & Fang, Q. (2020). Forecast on silver futures linked with structural breaks and day-of-the-week effect. North American Journal of Economics and Finance (53), doi:10.1016/j.najef.2020.101192.

[4]. Siem Jan Koopman, Borus Jungbacker & Eugenie Hol.(2004).Forecasting daily variability of the S&P

100 stock index using historical, realised and implied volatility measurements. Journal of Empirical Finance (3), doi:10.1016/j.jempfin.2004.04.009.

[5]. Zhang, Bo, Zhong, YJ & Tian, JF. (2009). Analysis of long memory drivers of Shanghai index volatility based on high frequency data. Statistics and Information Forum (06), 21-26. doi:CNKI:SUN:TJLT.0.2009-06-003. (In Chinese)

[6]. Wei Y. (2010). A study on volatility forecasting model for CSI 300 stock index futures. Journal of Management Science (02), 66-76. doi:CNKI:SUN:JCYJ.0.2010-02-009. (In Chinese)

[7]. Tang, Y. & Chi, Y.G.(2010). Long memory analysis based on realized volatility. Journal of Fuzhou University (Philosophy and Social Science Edition) (05), 27-32+48. doi:CNKI:SUN:FZDS.0.2010-05-006. (In Chinese)

[8]. Mastro, D. (2014). Forecasting Realized Volatility: ARCH-type Models vs. the HAR-RV Model. Available at: SSRN 2519107.

[9]. Ma, F., Wei, Y., Huang, D., & Chen, Y. (2014). Which is the better forecasting model? A comparison between HAR-RV and multifractality volatility. Physica A: Statistical Mechanics and its Applications, 405, 171-180.