ISSN 2004-2965 Energy Proceedings, Vol. 20, 2021

International Conference on Applied Energy 2021
Nov. 29 - Dec. 5, 2021, Thailand/Virtual
Paper ID: 586

Crude Oil Market Intraday Risk Prediction Based on Generalized Heterogeneity Autoregressive and Threshold Kernel Variation Method

Zhi-Yi Zheng[1], Lu-Tao Zhao[1,2,3*]

1 School of Mathematics and Physics, University of Science and Technology Beijing, Beijing 100083, China

2 Center for Energy and Environmental Policy Research, Beijing Institute of Technology, Beijing 100081, China

3 School of Management and Economics, Beijing Institute of Technology, Beijing 100081, China

ABSTRACT

In recent years, the advancement of trading technology and the acceleration of information transmission have intensified the intraday volatility of the oil market. To identify the volatility and risk of the intraday market accurately and effectively, this paper proposes a method for intraday risk prediction based on generalized heterogeneity autoregressive for high-frequency spot volatility modeling. First, use the threshold kernel variation method to separate jumps and characterize the spot volatility, then redefine the heterogeneity characteristics of the high-frequency intraday market to construct the optimal generalized heterogeneity autoregressive model, and finally predict and assess the intraday market risk. The results show that the intraday jumps of the high-frequency crude oil futures mostly occur in the event window of geopolitical news and EIA announcements, and there are short-term jump aggregations. Separating the jumping components can establish a more accurate prediction model for the fluctuation process. The model proposed takes into account the characteristics of intraday heterogeneity and finds that weekly fluctuations have no information contribution to high-frequency traders. Compared with the ARMA and GARCH models, it ensures the validity and accuracy of the results. With easy operation and scalability, it is an effective risk management tool for crude oil intraday market transactions.

Keywords: value at risk, high-frequency spot volatility, crude oil futures market, market heterogeneity, jump recognition

NONMENCLATURE

Abbreviations

EIA	Energy Information Administration
ARMA	Autoregressive Moving Average
GARCH	Generalized Autoregressive Conditional Heteroskedasticity
RV	Realized Volatility
HAR	Heterogeneity Autoregressive
KV	Kernel Variation
SV	Spot Volatility
TKV	Threshold Kernel Variation
GHAR	Generalized Heterogeneity Autoregressive
VaR	Value at Risk
MAE	Mean Average Error
MSE	Mean Square Error
MAPE	Mean Absolute Percentage Error
HMAE	Heteroscedasticity Mean Average Error
HMSE	Heteroscedasticity Square Average Error
LR	Likelihood Ratio

1. INTRODUCTION

In the financial market, volatility modeling and forecasting are the keys to applications such as asset allocation, risk management, asset pricing, and portfolio strategy selection [1, 2]. However, in recent years, with the advancement of trading technology and the acceleration of information dissemination, the intraday volatility of crude oil futures has intensified and risk

Selection and peer-review under responsibility of the scientific committee of the 13th Int. Conf. on Applied Energy (ICAE2021).

control has become more difficult. In the face of high-frequency and drastic changes in futures prices, accurately portraying intraday fluctuations and reasonably measuring risks are important bases for grasping intraday trading opportunities and making risk management decisions.

By accurately and effectively describing the volatility process, we can timely capture the impact of intraday news on the crude oil futures market, and even conduct risk analysis on the impact information with fixed periodicity, formulate targeted investment strategies, and seize opportunities in risks to obtain excess returns. For the existing methods of volatility modeling and forecasting, the current research [3-5] have mainly focused on low-frequency volatility and daily realized volatility (RV) modeling and forecasting. The GARCH models and HAR models are the most widely used. The GARCH models use historical information of the same frequency for modeling, and have clear advantage at low frequency. While, the HAR models target the realized volatility, simply using high-frequency intraday information. However, high-frequency data have outstanding differences in heteroscedasticity, memory ability, and sharp peak and thick tail characteristics. Therefore, these methods are poor to predict intraday fluctuations and lack guiding significance for intraday investment decision-making. With the development of the era of high-frequency information, we can obtain a large amount of high-frequency financial data. If we can find a suitable forecasting model for high-frequency crude oil futures market volatility, not only can we use massive historical fluctuation information to circumvent the problem of parameter uncertainty in low-frequency volatility modeling, but also improve the timeliness and effectiveness of intraday spot volatility(SV) prediction, to ensure the controllability of the risk in the intraday crude oil futures market.

As an important reference indicator for investment decision-making, the method of characterizing volatility is constantly evolving. High-frequency volatility agents, as a hot topic of current volatility research, can be subdivided into two categories: realized volatility and spot volatility. The RV can be regarded as the average value of the historical volatility over some time, focusing on the prediction of the daily realized volatility, so the estimated results still have a loss of information frequency. The high-frequency SV is relatively finer in time granularity and is more susceptible to the impact of market microstructure noise. According to literature review [6,7], the SV estimation based on the kernel variation (KV) method has a good proxy ability, superior to the RV method. Therefore, this paper chose to use KV and its improved method as the agent of high-frequency SV to fit the intraday volatility of crude oil futures prices, and separated continuous volatility and jump components.

Reasonably portraying market fluctuation and accurately predicting the intraday volatility process are the keys to judging market conditions and executing investment decisions. Aiming at the diffusion and jumping characteristics of the high-frequency price process, this paper combined the KV method and the generalized HAR(GHAR) method to characterize and predict the high-frequency SV, providing a new tool to measure the price risk of crude oil futures. Compared with the traditional methods, it has the characteristics of refinement, efficiency, and accuracy. By the adjustment of the threshold method, the continuous fluctuation process is more predictable, and the risk control of the real market can be closer to the ideal situation.

2. METHODS

2.1 High-frequency SV proxy based on KV

The method of kernel smoothing is often used in non-parametric estimation, using kernels to record objects such as density distribution and regression functions. The KV was proposed by Kristensen [8] as a method of realized SV proxy. The original variance can be filtered through the kernel to realize the non-parametric representation of the specific fluctuation process.

Define the intraday logarithmic price process $\{X_t\}$ of crude oil futures based on the diffusion process, namely:

$$dX_t = b_t dt + \sigma_t dW_t, t \in [0,T] \quad (1)$$

Where, W_t is the standard Brownian motion, b_t and σ_t are the stepwise measurable processes that make the equation (1) have the only strong solution in the interval.

For any $T > 0$, such as a trading day, $n+1$ discrete observation results of X_t can be obtained with the observation distance $\Delta_n = T/n$, denoted as: $X_{t_0}, X_{t_1}, \cdots, X_{t_{n-1}}, X_{t_n}$. For any $\tau \in [0,T]$, the kernel estimator of the SV at this time can be expressed as:

$$\hat{\sigma}_t^2 = KV_t = \sum_{i=1}^{n} K_h(t_i - \tau)(\Delta_i X)^2 \quad (2)$$

Where, $\Delta_i X = X_{t_i} - X_{t_{i-1}}$, $K_h(x) = K(x/h)/h$. To avoid using future information to model the realized volatility, the kernel function $K(x)$ takes the form of the unilateral kernel function:

$$K(x) = \begin{cases} 6\left(1+3x+2x^2\right), & -1 \le x \le 0 \\ 0, & otherwise \end{cases} \tag{3}$$

The estimation $\hat{\sigma}_t^2$ by the unilateral kernel method can be used in the modeling of forecasting future volatility.

Further consider that the impact of intraday information on the crude oil market may bring about abnormal price changes, as a jump process J_t, the jump-diffusion process is:

$$dX_t = b_t dt + \sigma_t dW_t + dJ_t, t \in [0,T] \tag{4}$$

By increasing the threshold function to separate the continuous process and the jump process, the threshold kernel variation (TKV) is defined as a proxy for the SV. The detailed theoretical basis can be found in [7], and the form is as follows:

$$\hat{\sigma}_t^2 = TKV_t = \sum_{i=1}^{n} K_h(t_i - \tau)(\Delta_i X)^2 I_{\left\{(\Delta_i X)^2 \le r(\Delta_n)\right\}} \tag{5}$$

Where, $I(\bullet)$ is an indicator function that indicates whether the variance exceeds the threshold function $r(\Delta_n) = \Delta_n^a$.

The proxy method of TKV can well separate continuous fluctuation and price jumps, and the choice of the threshold function is the key component. Too large a threshold will cause abnormal changes in the continuous fluctuation process, which will affect the subsequent prediction model fitting effect; too small a threshold will cause the normal fluctuation process to be classified as a jump, reducing the effect of investment returns. Therefore, to determine the best jump threshold that fits the crude oil futures price process, this paper adopts a grid optimization method to determine the threshold parameter, so that the proxy results of high-frequency SV are more scientific and objective.

2.2 Volatility prediction algorithm based on GHAR

The HAR algorithm is widely used in the prediction of realized volatility [6,7], because of its simple model, and the excellent forecasting effect considering the characteristics of heterogeneous markets. However, the high-frequency intraday market heterogeneity theory is insufficient, and the high-frequency SV has more prominent micro-market characteristics than the traditional realized volatility. Therefore, this paper considers the high-frequency SV as the research object, reconstructs the heterogeneous factors, and expands the generalized HAR prediction algorithm for intraday high-frequency SV prediction.

Taking minute-level SV as the research object, and considering the effects of heterogeneous characteristics of minutely, hourly, daily, weekly, and longer-term, we define a candidate set of heterogeneous factors:

$$\begin{cases} SV_t^M = SV_t \\ SV_t^H = \dfrac{1}{N_{\Delta_n}^H} \sum_{i=0}^{i=N_{\Delta_n}^M - 1} SV_{t-i\Delta_n}^M \\ SV_t^D = \dfrac{1}{N_{\Delta_n}^D} \sum_{i=0}^{i=N_{\Delta_n}^D - 1} SV_{t-i\Delta_n}^M \\ SV_t^W = \dfrac{1}{N_{\Delta_n}^W} \sum_{i=0}^{i=N_{\Delta_n}^W - 1} SV_{t-i\Delta_n}^M \\ \cdots \end{cases} \tag{6}$$

Where, $N_{\Delta_n}^{(\bullet)}$ is the number of the step Δ_n included in different frequency bands. This paper defaults the highest frequency heterogeneity factor to the minute factor. Accordingly, we construct the generalized heterogeneity regression (GHAR) prediction model as:

$$\begin{aligned} SV_{t+1}(k) &= SV_{t+\Delta_n}^M(k) \\ &= c + k_1\beta^M SV_t^M + k_2\beta^H SV_t^H + k_3\beta^D SV_t^D + \cdots + \omega_{t+1} \end{aligned} \tag{7}$$

Where k represents the number of heterogeneous factors selected into the model after being arranged from high frequency to low frequency. To determine the period which high-frequency traders refer to historical volatility information, this paper studies two cases of $k=3$ and $k=4$. In addition, considering the volatility aggregation characteristics of high-frequency data, the rolling modeling is adopted to predict the GHAR and provides the dynamic micro-volatility characteristics of the intraday crude oil futures market. The improvement of forecast accuracy can help high-frequency traders judge market patterns and make investment decisions.

2.3 High-frequency intraday value at risk measurement algorithm

The way to measure price risk by volatility is usually to adjust the quantile based on the conditional volatility of the GARCH model [9] to obtain the VaR. Based on the GHAR-TKV model, the intraday VaR measurement algorithm is proposed similarly.

Algorithm 1. High-frequency intraday risk measurement algorithm based on GHAR-TKV.

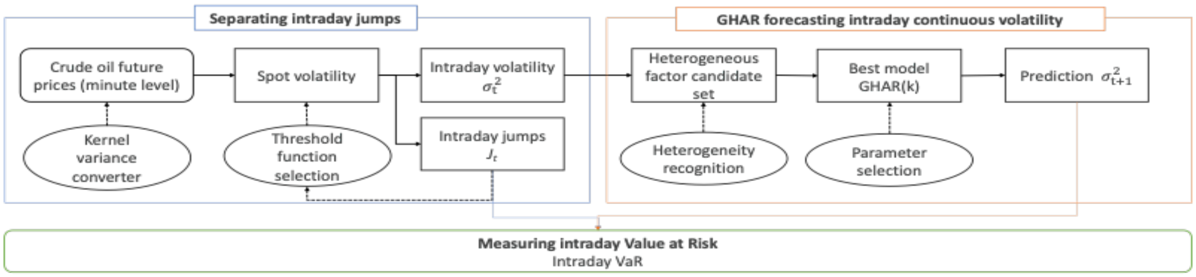


Fig. 1. Crude oil futures intraday risk measurement: flow chart

Step1. Process the high-frequency crude oil futures price data to obtain the intraday yield sequence suitable for jump-diffusion modeling.

Step2. Determine the parameters of the SV proxy method, and construct a proxy for the high-frequency SV of crude oil futures.

Step3. Check whether the jumping separation is reasonable. If it is unreasonable, re-select the threshold function and return to Step2; if it is reasonable, continue to Step4.

Step4. Construct a rolling GHAR model to predict TKV volatility.

Step5. Adjust the quantile based on the prediction result of TKV, and calculate $VaR_{t+1} = Z_\alpha \sqrt{TK\hat{V}_{t+1}}$. Where, Z_α is the α quantile of the corresponding distribution.

3. DATA SOURCE

In this paper, we collected the closing prices of New York crude oil future CL among 2015/1/1-2020/1/20 from WIND and processed them continuously. The price series spanning two weeks from 2020/1/5 to 2020/1/18, with a volume of 13,800, are used in empirical modeling and forecasting.

4. EMPIRICAL STUDY RESULTS AND DISCUSSION

Based on the intraday risk measurement algorithm, this paper measures and empirically analyzes the intraday risk of crude oil futures. It is mainly divided into three modules: intraday jumping separation in the crude oil futures market, characterization of intraday market heterogeneity, and crude oil futures market Intraday risk measurement. The specific process is shown in Fig. 1.

4.1 Intraday jumping separation in the crude oil futures market

Price jumps in the crude oil futures market are often impacted by short-term events such as news information and announcements. The existence of jumps makes it difficult for us to accurately measure the risk of normal return fluctuations. Before analyzing the heterogeneous characteristics of intraday continuous fluctuations, we need to separate jumps and continuous fluctuations from the price process.

Using the KV as a proxy for the SV, we further choose α=0.59 as the power of the threshold function, through the optimized algorithm. The recognition results of partial jumps are shown in Table 1.

Table 1. Recognition result of intraday jumps in the crude oil futures market

Date	Time	Jump Size/%	Reason
2020/1/7	18:22	0.58%	Geopolitics
2020/1/7	18:41	0.98%	Geopolitics
2020/1/7	18:42	-0.69%	Geopolitics
2020/1/7	20:40	-0.84%	Geopolitics
2020/1/8	10:30	-0.86%	EIA Announcement
2020/1/8	11:28	-0.72%	EIA Announcement
2020/1/8	15:46	0.58%	Geopolitics
2020/1/8	15:49	0.55%	Geopolitics

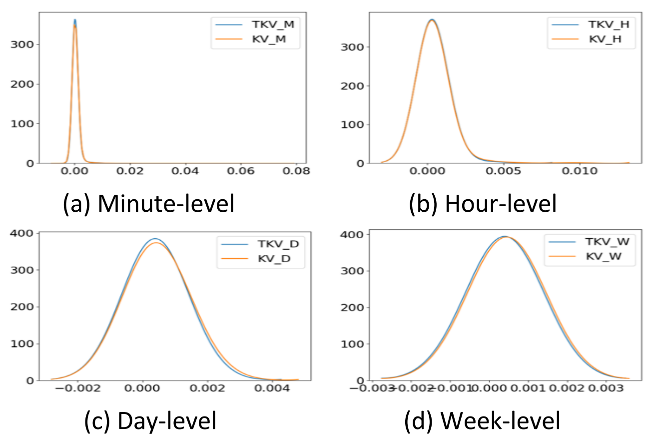
The identified intraday jumps can be attributed to the release of subject information such as geopolitics and EIA announcements according to the corresponding event window. The size of the jump is the degree of market response to the information shock. For the same event shock, the intensity of the price response has a certain decay over time. In addition, the concentration at the jumping moment reflects the significant feature of the crude oil futures market that jumps gathering in the process of asset prices. The same direction jumps are the underlying reason for the excess returns brought about by information shocks, while the reverse jumps require caution that good news do not materialize and instead causes the market to move downward.

4.2 Forecasting intraday continuous volatility by GHAR

After separating the jumping components, the analysis of high-frequency market heterogeneity should be given priority to forecast continuous volatility.

However, the traditional market heterogeneity hypothesis is for daily volatility. Investors will consider the volatility of the daily market, the weekly market, and the monthly market. In high-frequency intraday trading, we redefine the focus time band of heterogeneous investors. Considering the minute-level SV forecast, the traders may refer to previous minutely SV, hourly fluctuations, daily fluctuations, and weekly fluctuations.

Fig. 2 shows the distribution of high-frequency heterogeneity candidate factors. Intraday fluctuations have obvious tailing characteristics. With the extension of the heterogeneous period, the shape of the volatility distribution approaches the symmetrical distribution, and the tailing fluctuations are greatly reduced. Due to the presence of microstructure noise in the high-frequency intraday market, different heterogeneous factors are considered in the volatility modeling process to smooth historical fluctuations. The TKV method is adopted to maintain the stability of the overall distribution of volatility, and the abnormal tail fluctuations are truncated and separated, so that the continuity and concentration of the fluctuations are stronger, which is convenient for modeling.



(a) Minute-level (b) Hour-level

(c) Day-level (d) Week-level

Fig. 2. Distribution of Heterogeneous Factors of Crude Oil Futures in the Candidate Set

Because high-frequency traders are more inclined to short-term transactions, the generalized heterogeneity is mostly reflected in the consideration of daily and intraday market conditions, that is, weekly fluctuations have little reference value for high-frequency traders. Combined with fitting results, GHAR with k=3 is selected as the basis for subsequent risk measurement. In addition, the goodness of fit of the model with TKV as the research object is higher than that of KV, which further confirms that the separation of jump components can provide more suitable factor terms for model establishment.

By characterizing the heterogeneity of the crude oil futures intraday market, the GHAR model is constructed to predict high-frequency volatility. Compared with the ARMA model, GHAR has shown better predictive ability in different evaluation dimensions as shown in Fig. 3. The evaluation index has improved more significantly under the heterogeneity adjustment, such as more than 80% for HMSE and 60% for HMAE. Even without jumping separation, GHAR performs better than ARMA model, showing good memory ability for the high-frequency market. In addition, keeping a balance of volatility aggregation and memory in the high-frequency market, the rolling regression is more efficient and accurate.

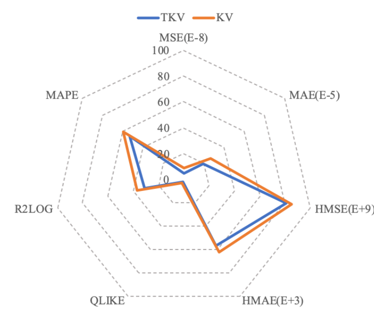


Fig. 3. GHAR improves the effect of the ARMA model

4.3 Crude oil futures market Intraday risk measurement

According to the sharp peak and thick tail of the high-frequency volatility distribution in the previous section, this paper uses the partial t-distribution to measure VaR, and the out-of-sample prediction results are shown in Fig. 4.

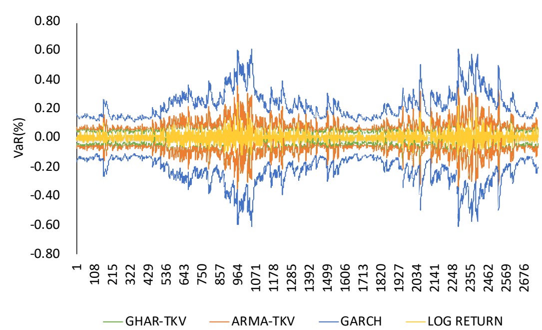


Fig. 4. Comparison of VaR measurement results (confidence level 95%)

Because the GHAR model takes into account the heterogeneous characteristics of the intraday market, which not only reflects the severe short-term impact of volatility clusters on returns, but also can quickly correct the normal fluctuations in returns after market sentiment is stabilized. Therefore, GHAR-TKV avoids overestimation of risks and loss of revenue opportunities with the GARCH model.

To distinguish the predictive effects of GHAR-TKV and ARMA-TKV, and to prove the effectiveness of GHAR-TKV for risk measurement, we back-tested VaR out of the sample. Although the number of failures of GHAR-TKV is higher than that of ARMA-TKV, the failure rate is more in line with the theoretical 5%, which is more effective in the LR test [10]. As shown in Fig. 5, GHAR-TKV has better prediction performance, which can isolate the jump risks, and prompt the trading opportunities in the jumps. While further measuring the risk of continuous volatility, it can quickly respond to volatility aggregation and measure more reasonably market risks, avoiding loss of opportunity and revenue for high-frequency investors.

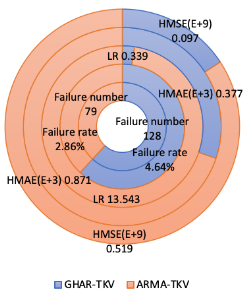


Fig. 5. VaR backtest performance outside the sample

5. CONCLUSIONS

Based on the unique characteristics of high-frequency crude oil intraday futures price, combined with intraday news crawling technology, this paper proposes the GHAR-TKV algorithm to measure intraday risk. Through empirical analysis, the following conclusions are supported:

(1) Using TKV to separate and estimate intraday jumps, an average intraday jump size of 0.73% with a clustering effect has been captured. It can be found that what triggers the jumps is the release of intraday information, including geopolitical news, EIA announcements, and other important crude oil market-related events.

(2) Similar to the general financial markets, investors in the high-frequency crude oil futures market are also heterogeneous. While high-frequency traders prefer intraday information, weekly fluctuations cannot provide effective information for them. The GHAR model based on the heterogeneity of high-frequency markets has a good predictive effect on high-frequency SV, and the most significant improvement in the evaluation indicators of HMSE and HMAE, that is, it can capture the heteroscedasticity of market price fluctuations promptly.

(3) The difficulty in measuring risk in the high-frequency market lies in the microscopic noise leading to thick tail, and rapid conversion of volatility aggregation. The GHAR-TKV provides a more effective risk warning than the GARCH model, which can avoid the severe overestimation of market risks and the loss of profit opportunities when the market is stable. For risk-averse people, when the estimated risk through GHAR-TKV exceeds the tolerance range, they should choose to lock in the income immediately; for risk-lovers, when we separate the jumps, immediately execute the trading strategy to obtain excess returns. It provides market participants with effective risk management tools, combined with the crawling of real-time information, which can further analyze market micro-changes and provide decision-making suggestions.

ACKNOWLEDGEMENT

The authors gratefully acknowledge the financial support from the National Natural Science Foundation of China under Grant No. 71871020.

REFERENCE

[1] Bollen B. What should the value of lambda be in the exponentially weighted moving average volatility model? APPL ECON. 2015;47:853-860.

[2] Ding J, Meade N. Forecasting accuracy of stochastic volatility, GARCH and EWMA models under different volatility scenarios. APPL FINANC ECON. 2010;20:771-783.

[3] Wei Y. Forecasting volatility of fuel oil futures in China: GARCH-type, SV or realized volatility models?. PHYSICA A. 2012;391:5546-5556.

[4] Corsi F. A simple approximate long-memory model of realized volatility. J FINANC ECONOMET. 2009;7:174-196.

[5] Clements A, Preve DP. A practical guide to harnessing the har volatility model. J BANK FINANC. 2021:106285.

[6] Figueroa-López JE, Li C. Optimal kernel estimation of spot volatility of stochastic differential equations. STOCH PROC APPL. 2020;130:4693-4720.

[7] Yu C, Fang Y, Li Z, Zhang B, Zhao X. Non-Parametric Estimation of High-Frequency Spot Volatility for Brownian Semimartingale with Jumps. J TIME SER ANAL. 2014;35:572-591.

[8] Kristensen D. Nonparametric filtering of the realized spot volatility: A kernel-based approach. ECONOMET THEOR. 2010;26:60-93.

[9] Angelidis T, Benos A, Degiannakis S. The use of GARCH models in VaR estimation. STAT METHODOL. 2004;1:105-128.

[10] Kupiec P. Techniques for verifying the accuracy of risk measurement models. J DERIV. 1995;3.

