Volatility forecasting of crude oil futures based on a genetic algorithm regularization online extreme learning machine with a forgetting factor: The role of news during the COVID-19 pandemic

Futian Weng [a,b,c], Hongwei Zhang [d,e], Cai Yang [f,*]

[a] *School of Medicine, Xiamen University, Xiamen, Fujian, 361005, China*
[b] *National Institute for Data Science in Health and Medicine, Xiamen University, Xiamen, Fujian, 361005, China*
[c] *Data Mining Research Center, Xiamen University, Xiamen, Fujian, 361005, China*
[d] *School of Mathematics and Statistics, Central South University, Changsha, Hunan, 410083, China*
[e] *Institute of Metal Resources Strategy, Central South University, Changsha, 410083, China*
[f] *School of Business Administration, Hunan University, Changsha, Hunan, 410082, China*

ARTICLE INFO

ABSTRACT

The outbreak of news and opinions during the COVID-19 pandemic is unprecedented in this age of rapid dissemination of information. The ensuing uncertainty has led to the emergence of heightened volatility in prices of crude oil futures. Whether such news has predictive value for the volatility of crude oil futures during the COVID-19 pandemic is examined in this research. We proposed a modeling framework, genetic algorithm regularization online extreme learning machine with forgetting factor (GA-RFOS-ELM), to estimate the effects of news during the COVID-19 pandemic on the volatility of crude oil futures. GA-RFOS-ELM could learn block-by-block with fixed or varying block size when considering the block own valid period. The experimental results illustrate that news during the COVID-19 pandemic has more predictive information, which is crucial for short-term volatility forecasting of crude oil futures. The novel approach illustrates that online update learning ability is needed during the COVID-19 pandemic, which could be effective and efficient in volatility forecasting of crude oil futures. The contributions of our study are significant for investors and administrators to predict and understand the behavior of volatility during the COVID-19 pandemic.

1. Introduction

As information technology is upgraded substantially, flows of news and opinions are essentially instantaneous. Currently, the use of mobile devices to interact and make financial decisions based on the news and opinions has become a feasible and conventional trading strategy (Groß-Klußmann and). More recently, the outbreak of coronavirus (COVID-19) attracted widespread attention from media worldwide, which was heavily inclined towards emphasizing the severity (Blendon et al., 2004; Mairal, 2011; Young et al., 2013). The news and opinions related to coronavirus can cause public panic and influence investors' sentiments (Tetlock, 2007). In particular, the crude oil future market suffered huge losses due to the COVID-19 pandemic and overwhelming related news. For example, West Texas Intermediate (WTI) crude oil futures were trading at $25.16 per barrel, with Brent at $27.91 on March 19, 2020, a historic low since 2002. Even more shockingly, an unprecedented event occurred in which the crude oil prices plunged to zero for the first time and then to negative values at -$37.63. This was a rare occurrence in the global crude oil trading market and highlighted the acute imbalance between supply and demand and the imminent collapse of inventories, which exert enormous impacts on each participant in the market.

The impact of news reports on the crude oil market has attracted increasing attention of scholars in the last few years. The consensus is that news has predictive power and cannot be ignored when predicting crude oil future market dynamics, especially in times of economic uncertainty (Narayan, 2019). In fact, the price pressure hypothesis pointed out that individual investors do not have enough time, knowledge, experience, or energy to examine all crude oil futures, and that they generally buy oil futures that attracting their attention or are being widely discussed (Nofsinger and Sias, 1999; Barber and Odean, 2008). It can be reasonably concluded that news could attract investors' attention

and tend to generate abnormal returns (Takeda and Wakao, 2014). Furthermore, the rationale behind network analysis emphasized that individual investors rely primarily on news feedback strategies to judge the future prospects of stock (Bange, 2000).

Considerable news is generally related to major events, and the news caused by major events can break out in an instant. On the one hand, major events could affect the crude oil future market. Previous studies have identified crude oil market responses to disasters, political events and so on (Kowalewski and Śpiewanowski, 2020; Bash and Alsaifi, 2019; Shanaev and Ghimire, 2019). Furthermore, research on the effects of pandemic diseases on the crude oil future market is scant, especially with respect to the more recent COVID-19 contagious infectious disease. The COVID-19 pandemic has endured for a long time, and the investment and business environments are bombarded with a mass of public news related to the COVID-19 pandemic. A growing body of empirical and theoretical studies has proven that news during the COVID-19 pandemic was valuable in volatility forecasting of crude oil futures. However, to the best of our knowledge, the more challenging question of how much predictive power news exhibits is not yet satisfactorily answered.

For example, some scholars attempted to examine the relationship between the stock market returns and COVID-19 related news (Albulescu, 2020; Baker et al., 2020; Lopatta et al., 2020; Onali, 2020). Additionally, COVID-19 related news is associated with volatility in the stock markets (Haroon and Rizvi, 2020). (Cepoi, 2020) argued that the stock markets exhibited asymmetric dependencies with COVID-19 related news. However, earlier literature concentrated predominantly on the role of news during the COVID-19 pandemic with respect to the stock markets, the prediction ability and application of which remain to be further explored: especially the crude oil futures. This study works hard to close the gap by focusing on volatility forecasting of crude oil futures, which acts as a barometer of stress, financial risk, or uncertainty of financial investments. This study is profound for all stakeholders surrounding the crude oil market to understand the behavior of the crude oil market during the COVID-19 pandemic.

Various classic econometric and statistical models have been adopted to forecasting volatility in crude oil market, such as the GARCH model (Herrera et al., 2018), the heterogeneous autoregressive (HAR) model (Luo et al., 2019), the vector autoregression (VAR) model (Salisu and Oloko, 2015) and the Markov models (Zhang and Wang, 2015). Taking into account the non-linear and non-stationary patterns implicit in the crude oil price series, artificial intelligence models are used to predict crude oil volatility. For example (Huang et al., 2004), proposed the extreme learning machine (ELM) method, which performs well in predicting the non-linear and non-stationary time series. And it has been widely used in the study of short-term price fluctuations in the oil market (Wang et al., 2018). Then (Huang et al., 2006) proposed an online sequential extreme learning machine (OS-ELM), which is effective in non-linear and non-stationary time series forecasting (Pan and Zhao, 2013). However, the parameters in the two hidden-layers are randomly set (Zhang et al., 2020b), combines OS-ELM and genetic algorithm (GA) to search for the optimal parameters of the hidden layer, called genetic algorithm online sequential extreme learning machine (GA-OS-ELM). Afterwards, GA-OS-EM has been widely used to identify the crack behavior in concrete dam (Dai et al., 2019), fault diagnosis (Zhang et al., 2020b) and gold price forecasting (Weng et al., 2020).

Based on the analysis, we proposed a prediction modeling framework for forecasting the volatility of the crude oil futures market by using news during the COVID-19 pandemic. News during the COVID-19 pandemic comes from real-time media analytics, which capture panic, media hype, fake news and national sentiment. The timeliness aspect of news during the COVID-19 pandemic is considered: that is, each datum has its own valid period, and the proposed online sequential extreme learning machine (OS-ELM) could learn block-by-block with fixed or varying block size (Liang et al., 2006). Additionally, with the ongoing coronavirus outbreak, the nonlinear and nonstationary phenomenon of crude oil price exhibits critical features, as determined from Fig. 2. Therefore, the new data should increasingly contribute to representing the impact on recent investor behavior, and the old data should contribute less. In such cases, the validity of the outdated training datum used is lost after several units of time, and the datum should be abandoned in the following learning. Therefore, the forgetting factor is introduced in this work to gradually eliminate the outdated datum that could become potentially misleading news, which is more suitable for the timeliness aspect (Paleologu et al., 2008). Furthermore, according to the huge changes in the parameters of different data in different periods, there are some problems such as output instability and output matrix singularity (Man et al., 2011). In this research, the OS-ELM algorithm is added by a regularization factor, which could ensure and improve the stability and generalization ability of the OS-ELM algorithm (Huynh and Won, 2011; Weng et al., 2020; Chen et al., 2020). Therefore, the structural and empirical risk could be lessened and balanced. In order to confront these challenges, this paper proposes the novel genetic algorithm regularization online extreme learning machine with forgetting factor (GA-RFOS-ELM). In detail, in the initial stage, this algorithm could select the optimal weight matrix and hidden neuron threshold by adding the genetic algorithm. In the next stage, this algorithm improves the stability and generalization ability by using a regularization factor.

In order to investigate the effects and predictive power of news during the COVID-19 pandemic on the volatility of crude oil futures, this paper uses the COVID-19 news related variables of real-time media analytics about announcements describing essential coronavirus issues, such as the panic index (PI), the Media Hype Index (HY), the Fake News Index (FNI), and the Country Sentiment Index (CSI), which come from StockTwits, Dow Jones Newswire or The Wallstreet Journal, among others (Blitz et al., 2019). With illustration and verification, the empirical results indicate that the proposed GA-RFOS-ELM model statistically outperforms all considered benchmark models considered in terms of forecasting accuracy. The empirical findings indicate that news is valuable for volatility forecasting of crude oil futures, which is consistent with the existing literature (Haroon and Rizvi, 2020). In particular, the forgetting mechanism is appropriate and necessary for volatility forecasting of crude oil futures by using news during the COVID-19 pandemic.

In general, the main contributions of this research are as follows. First, we propose a novel OS-ELM model integrating regularization and genetic algorithm with forgetting factor by including news during the COVID-19 pandemic to further improve the accuracy of prediction for the future volatility. Offline learning ability was taken into consideration for some existing studies using the ELM model. However, news during the COVID-19 pandemic is seldom considered in the time-variant system due to the limited period of validity of each datum of news during the COVID-19 pandemic. Moreover, news during the COVID-19 pandemic may vary considerably on a daily basis. It is almost impossible to exactly predict the volatility when the model considers responses of time-varying nature. Furthermore, existing studies about OS-ELM models rarely consider other additional variables (i.e., news during the COVID-19 pandemic). Second, we evaluate the role of news during the COVID-19 pandemic on volatility forecasting of crude oil futures by respectively comparing the forecasting performances of the various models with those of different benchmarking prediction models, which highlights the superiority and robustness of the proposed GA-RFOS-ELM model. This study offers novel insight into the impact of news during COVID-19 pandemic volatility forecasting of the crude oil market and enlarges our understanding, which is in fact consistent with the finding of (Cepoi, 2020). Third, we extend the research about the responses of stock market returns to major unconventional emergencies, especially the interaction of the pandemic diseases with stock returns. The primary measure of the COVID-19 pandemic is the number of infected cases (Al-Awadhi et al., 2020). However, we suggest that COVID-19 related news would be a good measurement of the COVID-19 pandemic to estimate the effect on crude oil future market returns, because it may

affect the expectations of individual investors regarding future crude oil futures.

The remainder of this research is organized as follows. Section 2 presents the existing OS-ELM model and develops the GA-RFOS-ELM model. Section 3 provides the data of the measures of various variables, such as crude oil futures volatility and COVID-19 related news. Section 4 describes the empirical results and interpretive analysis about the forecasting performance of the GA-RFOS-ELM model. Section 5 discusses the conclusion, limitations, and future work.

2. Method

2.1. Genetic algorithm

GA is a search heuristic approach motivated by natural evolution theory, an operational model of global search based on probability conversion Holland et al. (1992). The notion of a natural selection process can be used for a search problem: that is, selecting the best one from a series of solutions for a problem Mitchell (1998). There are several phases in a genetic algorithm, including coding, initial population, fitness function, and genetic operator.

The chromosome, composed of gene arrangement, is the starting material for biological genetics research. The first step of a genetic algorithm is coding, which abstracts the issues into a string of specific symbols through some certain mechanism. The binary system is a common coding method for the genetic algorithm, where the corresponding real value in the problem interval is converted into a binary string $(b_i b_{i-1} \ldots b_0)$, where i denotes the number of binary encoding bits.

After coding, a genetic algorithm usually uses a random method to generate a set of individuals as the initial population. The quality of the genetic algorithm is estimated by the fitness function for the individual (solution). A larger fitness function value indicates a better solution quality. The fitness function is the driving force of the genetic algorithm evolution process, which is generally determined by combination with the requirements of solving, and can be set as the error function.

The genetic operator is adopted to select some individuals from the parent population and inherit them to the next-generation population. This operator utilizes the roulette selection approach, known as the proportional selection operator, which makes the probability of each individual being selected as the positive value of its fitness function. Supposing that the population size is n and F_i denotes the fitness function of individual i, then the probability P_i of individual i which will be selected to the next-generation population is:

$$P_i = F_i \bigg/ \sum_{i=1}^{n} F_i \tag{1}$$

Specifically, reproduction, crossover, and mutation are the basic operations of a genetic algorithm. The replication operation can select the best chromosome from the previous population, and it cannot be creative. The crossover operation is able to produce new excellent varieties through chromosome exchange. In addition, mutation simulates gene mutation caused by various accidental factors in natural genetics and randomly changes the value of the genetic gene with a certain probability. In the binary coding of chromosomes, mutation randomly transforms genes from 1 to 0, or from 0 to 1. Mutation can prevent the model from falling into the local optimum and terminating the course in the early stage of the operation process. Furthermore, it can obtain the optimal solution with high quality in the largest possible solution space.

2.2. Online sequential extreme learning machine

The online sequential extreme learning machine (OS-ELM) is a fast and effective algorithm which can update network parameters through an online learning mechanism (Liang et al., 2006). This algorithm primarily contains two parts. The first part is the initial stage, that is, to obtain the output weight β of the model using a small-scale dataset. The output weight β of the model learned in the initial part will be updated through a fixed or varying chunk size learning mechanism in the online learning stage.

Suppose that there are N random training datasets (X_i, t_i), $X_i = [x_{i1}, x_{i2}, \cdots, x_{in}]^T$ and $t_i = [t_{i1}, t_{i2}, \cdots, t_{im}]^T$. Thus, the OS-ELM model can be described as the following processes:

(1) Initialization stage.

Step 1 Determine the initial input weights $\mathbf{a}_i$ and hidden layer threshold $\mathbf{b}_i$, $i = 1, \cdots, L$ randomly based on an initial training sample set $S_0 = \{(\mathbf{x}_i, t_i)\}_{i=1}^{N_0}$.

Step 2 Obtain the initial matrix between the hidden and output layer H_0.

$$\mathbf{H}_0 = \begin{bmatrix} (a_1, b_1, X_1) & \cdots & g(a_L, b_L, X_1) \\ \vdots & & \vdots \\ (a_1, b_1, X_{N_0}) & \cdots & g(a_L, b_L, X_{N_0}) \end{bmatrix}_{N_0 \times L} \tag{2}$$

Step 3 Calculate output weight $\beta^{(0)}$.

$$\beta^{(0)} = P_0 H_0^T T_0 \tag{3}$$

where $\mathbf{P}_0 = \left(\mathbf{H}_0^T \mathbf{H}_0\right)^{-1}$, $\mathbf{T}_0 = [\mathbf{t}_1, \cdots, \mathbf{t}_{N_0}]^T$.

(2) Sequential update stage.

Step 4 Calculate the output matrix of hidden layer H_0 based on Equ.2

Step 5 Calculate the output of the model: $\hat{t}_k = h_k \beta_{k-1}$.

Step 6 Update the output weight $\beta^{(k)}$.

· Chuck-by-chuck learning mechanism:

$$\begin{aligned} \mathbf{P}_{k+1} &= \mathbf{P}_k - \mathbf{P}_k \mathbf{H}_{k+1}^T (\mathbf{I} + \mathbf{H}_{k+1} \mathbf{P}_k \mathbf{H}_{k+1^T})^{-1} \mathbf{H}_{k+1} \mathbf{P}_k \\ \beta^{(k+1)} &= \beta^{(k)} + \mathbf{P}_{k+1} \mathbf{H}_{k+1}^T \left(\mathbf{T}_{k+1} - \mathbf{H}_{k+1} \beta^{(k)}\right) \end{aligned} \tag{4}$$

· One-by-one learning mechanism:

$$\begin{aligned} \mathbf{P}_{k+1} &= \mathbf{P}_k - \frac{\mathbf{P}_k \mathbf{h}_{k+1} \mathbf{h}_{k+1}^T \mathbf{P}_k}{1 + \mathbf{h}_{k+1}^T \mathbf{P}_k \mathbf{h}_{k+1}} \\ \beta^{(k+1)} &= \beta^{(k)} + \mathbf{P}_{k+1} \mathbf{h}_{k+1} \left(\mathbf{t}_{k+1}^T - \mathbf{h}_{k+1}^T \beta^{(k)}\right) \end{aligned} \tag{5}$$

Step 7 Return to step 5.

2.3. GA-RFOS-ELM method for predicting the volatility of crude oil prices

The OS-ELM model involves the calculation of inversion in the update process, for which it has been proven that its generalization ability will be seriously reduced once a singular or ill-conditioned hidden layer matrix appears (Huynh and Won, 2011). To overcome this challenge, we develop a novel hybrid approach (GA-RFOS-ELM) to forecast the volatility of crude oil prices.

Previous researchers have shown that the stability of the extreme learning machine (ELM) can be effectively improved by obtaining a high-quality feature map in the first stage (Huang et al., 2015): the same is true for the OS-ELM model. Therefore, we introduce the genetic algorithm to determine optimal input weight and the threshold of the hidden layer in the initial stage of online learning.

In the beginning, the optimal input weight and threshold of the hidden layer are obtained through the genetic algorithm. In detail, the solution of crude oil futures volatility can be considered as a population in the genetic algorithm. We regard the input weight and hidden layer bias as the gene of the chromosome. In addition, we use the sum of

absolute errors to measure the fitness function.

$$F = \sum_{i=1}^{N_0} \left| X_i - \widehat{X}_i \right| \tag{6}$$

where N_0 is the number of an initial sample set in OS-ELM, and X_i and $\widehat{X}_i$ express the actual values and output of the model, respectively.

Therefore, the optimal solution of the input weight and hidden layer bias problem can be converted into the objective of reducing the fitness function and selecting the best chromosome. That is, we obtain the weights of input layer $\mathbf{a}_i$ and hidden layer bias $\mathbf{b}_i$ by genetic algorithm, not randomly.

More importantly, after adding the forgetting factor and regularization mechanism, the loss function of the model is as follows (Celaya and Agostini, 2015; Guo et al., 2018):

$$J(\beta_k) = \sum_{i=1}^{k} \lambda^{k-i} \left| t_i - h_i \beta_k \right|^2 + \delta \lambda^k \|\beta_k\|^2 \tag{7}$$

where l is the forgetting factor parameter, and δ denotes the regularization coefficient.

According to literature (Guo et al., 2018), β_k in the sequential update stage of the GA-RFOS-ELM algorithm can be deduced.

$$\mathbf{P}_k^* = \frac{1}{\lambda}\mathbf{P}_{k-1} - \frac{\delta(1-\lambda)}{\lambda^2}\mathbf{P}_{k-1}\left(I - \frac{\delta(1-\lambda)}{\lambda}\mathbf{P}_{k-1}\right)\mathbf{P}_{k-1}$$

$$\mathbf{P}_k = \mathbf{P}_k^* - \frac{\mathbf{p}_k^*\mathbf{h}_k^T\mathbf{h}_k\mathbf{P}_k^*}{1 + \mathbf{h}_k\mathbf{P}^*\mathbf{h}_k^l} \tag{8}$$

$$\beta_k = \beta_{k-1} + \mathbf{P}_k\mathbf{h}_k^T(t_k - \mathbf{h}_k\beta_{k-1}) - \delta(1-\lambda)\mathbf{P}_k\beta_{k-1}$$

Thus, suppose that there are N random training samples (X_i, t_i), where $X_i = [x_{i1}, x_{i2}, \cdots, x_{in}]^T$ and $t_i = [t_{i1}, t_{i2}, \cdots, t_{im}]^T$. The GA-RFOS-ELM algorithm can be described as the following steps:

(1) Initialization stage.

Given an initial training subset $S_{k-1} = \{(x_j, t_j)\}_{j=1}^{N_{k-1}}$, with the number of L hidden layer neurons.

Step 1 Using the genetic algorithm to calculate the input weights $\mathbf{a}_i$ and hidden layer threshold $\mathbf{b}_i$ based on an initial training dataset.
Step 2 Determine the output matrix of hidden layer H_{k-1}.

$$H_{k-1} = \left[\mathbf{h}_1^T\mathbf{h}_2^T \cdots \mathbf{h}_{k-1}^T\right]^T \tag{9}$$

Step 3 Obtain the output weight β_{k-1}.

$$\beta_{k-1} = P_{k-1}H_{k-1}^T T_{k-1} \tag{10}$$

while $P_0 = (H_0^T H_0 + \lambda I)^{-1}$, $T_{k-1} = [t_1 t_2 \cdots t_{k-1}]^T$.

(2) Sequential update stage.
Step 4 Calculate the output matrix H_k of the hidden layer.

$$\mathbf{H}_k = \begin{bmatrix} (a_1, b_1, X_1) & \cdots & g(a_L, b_L, X_1) \\ \vdots & \vdots & \\ (a_1, b_1, X_k) & \cdots & g(a_L, b_L, X_k) \end{bmatrix}_{k \times L} \tag{11}$$

Step 5 Calculate the predicted crude oil volatility output of the model: $\widehat{t}_k = h_k\beta_{k-1}$.
Step 6 Update the output weights β_k according to Equ.8.
Step 7 Return to step 5.

The forecasting framework of GA-RFOS-ELM is represented in Fig. 1. Previous price volatility and several indices that quantify the influence of the COVID-19 pandemic are considered as input variables to predict future crude oil price volatility. After correlation analysis, the ultimate input variables are determined. In the initialization stage, the genetic algorithm is exploited to optimize initial parameters and calculate initial output weight, and therefore the initial network could be determined. In the sequential update stage, the mechanism of the GA-RFOS-ELM model is to used predict while actually learning. Once the price volatility at time k is predicted, the variables and actual values before time $k + 1$ are used to update the network until all forecasts of price volatility are output.

2.4. Evaluation criteria for prediction accuracy

It is not certain which criteria are more appropriate to evaluate the predictions of volatility models (Lopez, 2001). We choose four different functions, including the root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE) (Wang et al., 2018), and median error (MdE).

$$\text{RMSE} = \frac{1}{N} \sqrt{\sum_{i=1}^{N} \left(r_i - \widehat{r}_i \right)^2} \tag{12}$$

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^{N} \left| r_i - \widehat{r}_i \right| \tag{13}$$

$$\text{MAPE} = \frac{1}{N} \sum_{i=1}^{N} \frac{\left| r_i - \widehat{r}_i \right|}{r_i} \tag{14}$$

$$\text{MdE} = \text{median}\left(\left\{ r_i - \widehat{r}_i \right\}_{t=1}^{N} \right) \tag{15}$$

3. Data descriptions and variable analysis

The effectiveness of the proposed forecasting model is validated with respect to the volatility forecasting of the crude oil dataset from WTI. We use the daily crude close prices from January 3, 2006, to July 24, 2020, which were obtained from the Wind website. The absolute returns of crude price r_t are employed as the volatility: that is, $r_t = 100 \times [\ln(P_t / P_{t-1})]$, where P_t indicates the close price on day t (Wang et al., 2018). Fig. 2 illustrates the volatility of crude oil prices during 2006–2020, and the red line denotes the node of the COVID-19 pandemic, which shows that huge fluctuations in the markets occur during the novel coronavirus pandemic.

It is obvious that some events would exert fluctuations in crude oil exchange, especially with respect to the impact of the novel coronavirus epidemic on crude oil volatility. Moreover, four 19-nCoV news-related variables from January 1, 2020, to July 24, 2020, are introduced to forecast the crude oil price volatility for the first time. All of the data can be obtained from RavenPack (https://coronavirus.ravenpack.com/).

Table 1 indicates the descriptive statistics for these four variables and crude oil volatility during the novel coronavirus, and these indices versus the future crude oil price volatility are represented in Fig. 3.

The panic index (PI) estimates the degree of discussions in the media that reference terror and coronavirus disease. A higher quota value indicates more news related to panic occurring in the news. Fig. 3(a) compares the crude oil futures price volatility and PI, which shows that the fluctuations of price volatility are coincident with PI. These findings demonstrate that the PI index is correlated with the future volatility of crude oil prices.

The extent of news discussing coronavirus disease is evaluated by the Media Hype Index (HY). The index values range between 0 and 100. A value of 5 means that 5 percent of news is discussing the novel virus.
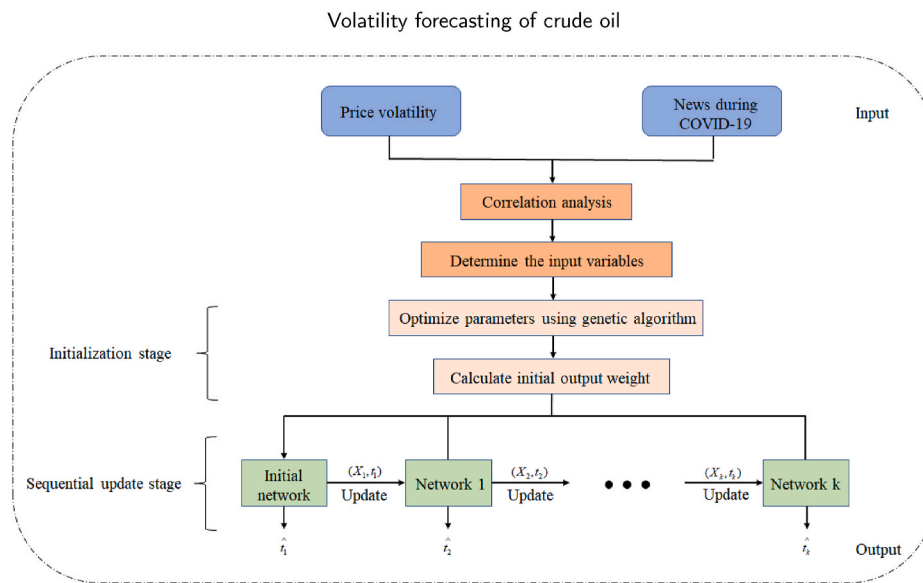
Volatility forecasting of crude oil

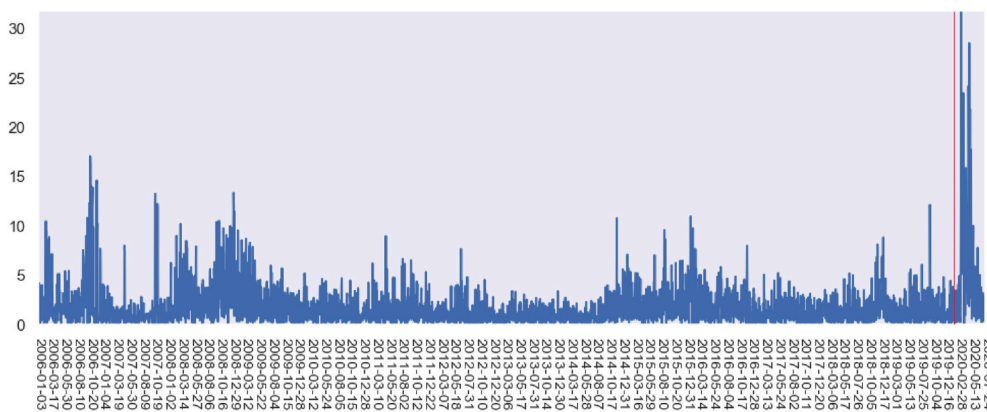**Fig. 1.** Forecasting framework of GA-RFOS-ELM.

.

Fig. 2. Volatility of crude oil prices.

Table 1
Descriptive statistical analysis on the variables during COVID-19 turmoil.

	PI	HY	FNI	CSI	Volatility
Minimum	0	0.020	0	−70	0.069
First quantile	1.685	12.550	0.42	−52.365	0.958
Median	2.920	38.830	0.62	−52.365	0.958
Third quartile	4.085	50.505	0.93	−11.62	5.247
Maximum	9.240	69.270	1.76	12.98	31.821
Mean	3.099	35.154	0.67	−31.01	4.705
Standard deviation	1.912	19.585	0.41	23.82	5.923

Fig. 3(b) represents the HY index versus the crude oil futures price volatility, which expresses greater coincidence between HY and the future oil price volatility during the middle and later periods.

The Fake News Index (FNI) estimates the extent of discussion in the media about COVID-19 that refers to 19-nCoV disinformation. Values are also defined between 0 and 100. The higher index indicates that more fake news appears. The FNI index is multiplied by 10 to make observation easier and enable comparison with the volatility of future crude oil prices. The high correlation between FNI and future crude oil price volatility is displayed in Fig. 3(c).

The degree of sentiment about substances referred to in the media

alongside COVID-19 is estimated by the Country Sentiment Index (CSI). Values range between 0 and 100, which indicates the extent of sentiment. Different from the previous three indices, the CSI index appears to exhibit a significantly consistent negative correlation with the future volatility (see Fig. 3(d)).

4. Simulation experiment

4.1. Implementation details

In this study, we removed abnormal data, such as the crude oil price on April 20, 2020, the day on which the crude oil price fell off a cliff. To make the result more reliable, the obtained dataset is divided into two parts. One part is from January 3, 2006 to December 31, 2019, which represents the dataset before the novel coronavirus, and the other indicates the epidemic period from January 1, 2020, to July 24, 2020.

Market participants generally care more about how well they can do in the future through the volatility forecasting approaches. Thus, the crude oil price volatility of the previous five days is regarded as the input variable to predict future volatility before January 1, 2020. During the 2019-nCoV, four quantitative indicators (PI, HY, FNI, and CSI) related to novel coronavirus pandemic news are appended to the input variables to
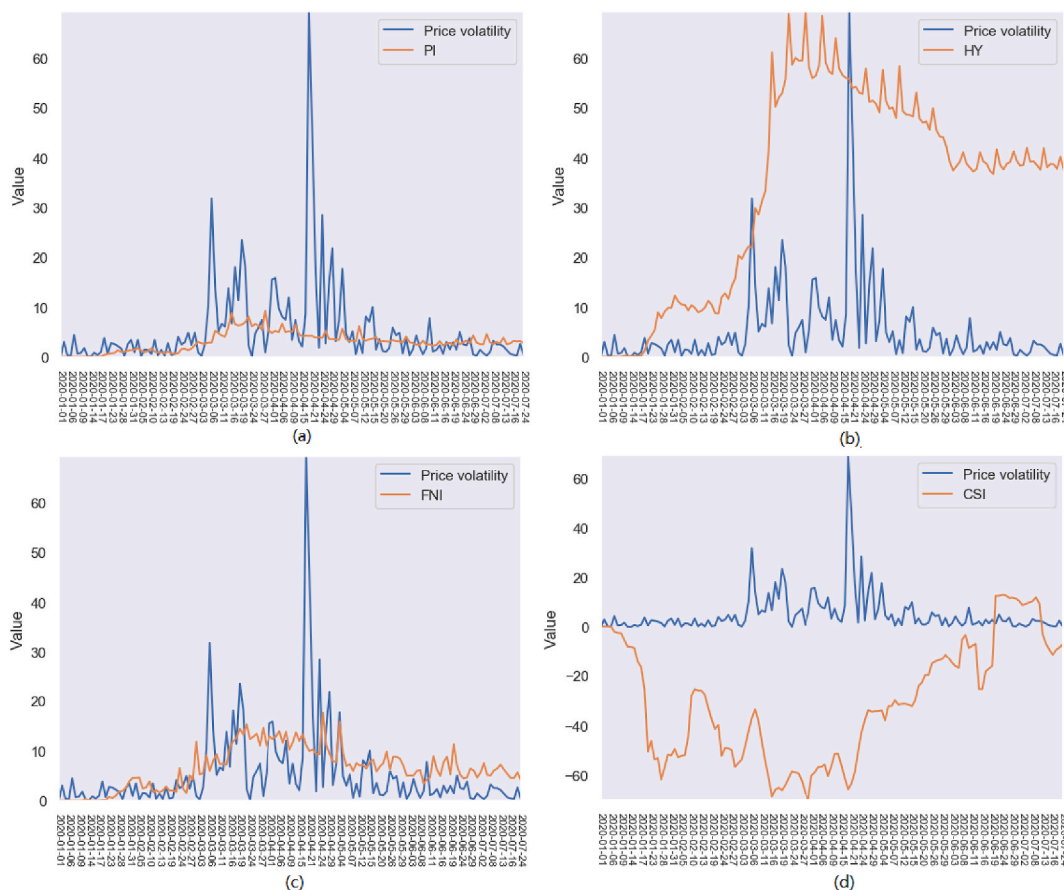
Fig. 3. Crude oil futures price volatility versus the indices regarding the novel coronavirus.

verify their roles in the future crude oil volatility forecast. To make the results more reliable, we split these datasets into two groups, with 70% forming the training set, and the rest used as the test set. It is common to divide the training and testing datasets as 7:3 (Alpaydin, 2020).

The input datasets are normalized to speed up the convergence, and the Sigmoid (Equ.16) is chosen as the activation function in our models.

$$S(x) = \frac{1}{1 + e^{-x}} \tag{16}$$

GA is introduced to obtain the optimal initial weights of input layer and hidden layer bias in the GA-RFOS-ELM model, the relevant parameters of which are illustrated in Table 2. Actually, there is no exact method to determine the parameters of the genetic algorithm. All parameters are determined through related literature (Leardi et al., 2002; Chung and Shin, 2020; Zhou et al., 2020) and experiments. We adopt binary encoding and set the size of the population as 20. Crossover and mutation probability are the most important parameters of the genetic algorithm, determined as 0.9 and 0.1, respectively. The terminal condition is the maximum iteration: that is, the optimization will be stopped when reaching the maximum iteration.

The number of hidden layer neurons is significant for the SLFN model and for the GA-RFOS-ELM crude oil forecasting model. Through cross-

Table 2
Parameters of genetic algorithm.

Name of parameters	Value
Size of population	20
Crossover probability	0.9
Mutation probability	0.1
Maximum number of iterations	100
Terminal condition	100

validation, this value is ultimately determined as 30. The forgetting factor parameter is set to 0.95 according to (Soares and Araújo, 2016).

In this paper, the GA-RFOS-ELM method proposed above is compared with several econometric models and machine learning algorithms, including autoregressive (AR), regression trees (RT), Bayesian regression (Bayes), support vector regression (SVR), extreme learning machine (ELM), online sequential extreme learning machine (OS-ELM) and genetic algorithm online sequential extreme learning machine (GA-OS-ELM). Among which, as an econometric model, AR is usually employed to solve time series problems. RT, Bayes, SVR are classic machine learning models based on different theories, which are commonly used as benchmark methods of machine learning. In addition, we consider ELM and OS-ELM, two significant algorithms in the development of the ELM algorithm family, which are also the basic methods of our proposed model in this paper. At the same time, GA-OS-ELM is introduced as a benchmark approach to compare the model we proposed. All of the above models are executed on a Dell server with 16 GB RAM and implemented in Python.

4.2. Results and analysis

To sum up, to verify the validity of models and the effect of COVID-19 news on crude oil price volatility, simulation experiments are implemented through three different datasets. Dataset 1 represents crude oil volatility from January 3, 2006, to December 31, 2019, which considers the crude oil price volatility of the previous 5 day as the input variables. Dataset 2 includes the crude oil volatility from January 1, 2020, to July 24, 2020, during coronavirus disease, which also regards the crude oil price volatility of the previous 5 day as input variables. Different from dataset 2, four quantitative indicators, PI, HY, FNI and CSI from January 1, 2020, to July 24, 2020, are considered in prediction

models. Therefore, a total of nine indicators are considered as input variables for dataset 3.

Figs. 4–6 show the evolution curves of the genetic algorithm for these three datasets. Error has been convergent within 100 iterations for the three different datasets through the genetic algorithm, which indicates that the optimal parameters are obtained. We consider the sum of L2 error as the fitness function; therefore, the error of dataset 1 is relatively larger than the other two due to its larger sample size (see Fig. 4). Actually, the genetic algorithm is used to obtain a better parameter instead of random determination, which mainly focuses on whether its error converges.

AR, RT, Bayes, SVR, ELM, OS-ELM and GA-OS-ELM are considered as benchmark models to compare with our proposed GA-RFOS-ELM model. The order of the AR model is determined by minimizing the Akaike information criterion (AIC) (Akaike, 1974). The minimal cost-complexity pruning algorithm Loh (2011) is used to prune a tree to avoid fitting for RT. Four hyperparameters α_1, α_2, λ_1 and λ_2 of the gamma prior distributions in the Bayes model are usually chosen to be noninformative. We select the RBF kernel in the SVR model, where the proper choice of c and γ is critical to its performance. The best pair of parameters are obtained by means of a grid search. As for ELM and OS-ELM methods, the number of hidden layers is selected as 30, which is the same as that for GA-OS-ELM and GA-RFOS-ELM. Moreover, the prediction accuracy of these three models, ELM, OS-ELM, GA-OS-ELM and our proposed GA-RFOS-ELM model, is obtained by running the models 200 times on average.

Table 3 illustrates the performances of the loss functions, such as RMSE, MAE, and MdE test, for predictive accuracy between the proposed GA-RFOS-ELM model with seven other benchmark models on the test samples of dataset 1. It is obvious that the proposed GA-RFOS-ELM algorithm performs better than the other six benchmark models in terms of the performance measurements of the RMSE, MAE, MAPE and MdE scores, yielding values of 1.1693, 0.8924, 2.1639 and 0.6414, respectively. In terms of RMSE scoring performance, GA-OS-ELM and Bayes models are second only to GA-RFOS-ELM, with scores of 1.3533 and 1.4262, respectively. As for MAE and MdE scores, the SVR model is actually better than the other six. If time-consuming is not taken into account, this model is generally considered to be a good choice for solving various problems.

It is observed that the GA-RFOS-ELM model also produces more accurate forecasts than the other seven comparison models for dataset 2 (see Table 4). Compared with AR, RT, Bayes, SVR and ELM models, the results of OS-ELM, GA-OS-ELM and GA-RFOS-ELM are relatively similar. As mentioned above, dataset 1 and dataset 2 have the same input-output structure, both of which used the crude oil price volatility of the previous 5 day as input variables to predict the future volatility. However,
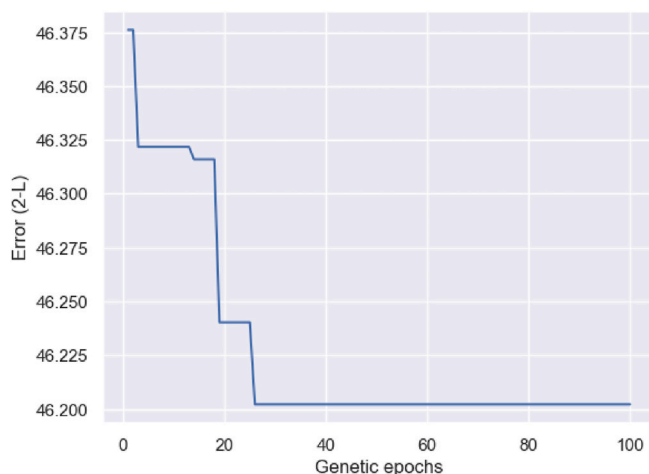

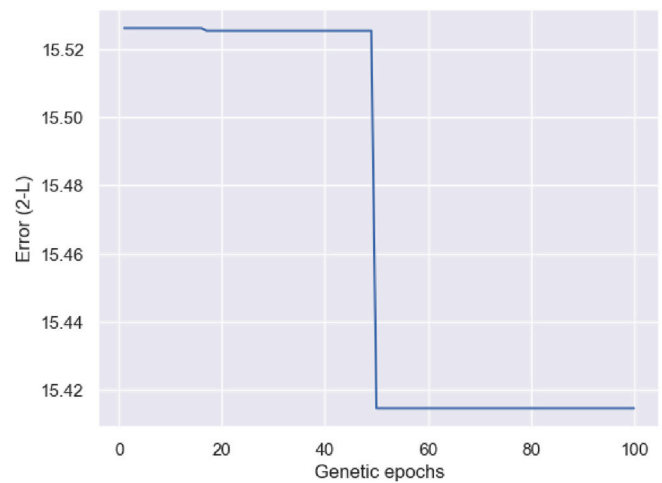**Fig. 5.** Evolution curve of genetic algorithm for dataset 2.
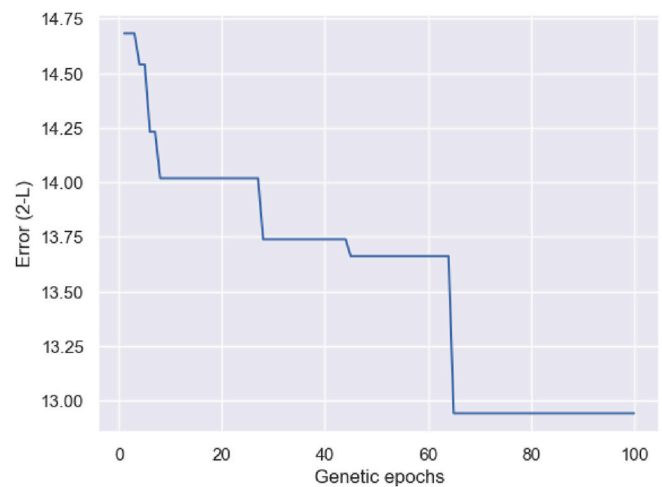

Fig. 6. Evolution curve of genetic algorithm for dataset 3.

Table 3
Predicting results of crude oil price volatility for dataset 1.

Model	RMSE	MAE	MAPE	MdE
AR	1.4898	1.1189	3.2019	1.0013
RT	2.0871	1.4887	3.2453	1.0486
Bayes	1.4262	1.0765	2.9336	0.9059
SVR	1.4657	1.0178	2.5039	0.7375
ELM	1.5734	1.1948	3.3247	0.9650
OS-ELM	1.4378	1.0857	2.6872	0.8157
GA-OS-ELM	1.3533	1.0324	2.3469	0.7846
GA-RFOS-ELM	1.1693	0.8924	2.1639	0.6414

Table 4
Predicting results of crude oil price volatility for dataset 2.

Model	RMSE	MAE	MAPE	MdE
AR	3.1155	2.8651	4.0479	3.0376
RT	4.2159	2.7427	2.8124	1.6067
Bayes	2.9548	2.6771	3.7840	2.7598
SVR	2.0675	1.5708	1.7722	1.3139
ELM	4.3781	3.4024	2.5649	3.0802
OS-ELM	1.8894	1.4244	1.6960	1.1707
GA-OS-ELM	1.6163	1.2806	1.4468	1.1304
GA-RFOS-ELM	1.4485	1.1759	1.3028	1.0560

Fig. 4. Evolution curve of genetic algorithm for dataset 1.

dataset 2 is a small sample dataset as compared with dataset 1, and the crude oil price fluctuated more sharply during the epidemic period.

Nononline learning machine learning (ML) algorithms, like RT, Bayes, SVR and ELM, may benefit from large data sets and consider global information to obtain relatively high prediction accuracy. Nevertheless, they are usually unable to capture the potential change of crude price volatility with time. They must retrain the models when the potential distribution of the data changes dramatically. However, the crude oil price fluctuated more sharply during the COVID-19 pandemic. At the same time, global learning may not be suitable in comparison with local learning. These may be the primary reasons why the online models, OS-ELM, GA-OS-ELM and GA-RFOS-ELM, perform better than the other models for dataset 2.

Table 5 demonstrates the forecasting and comparison results of the proposed GA-RFOS-ELM model and compared benchmark models, such as RT, Bayes, SVR, ELM, OS-ELM and GA-OS-ELM, for dataset 3. In terms of RMSE, MAE, MAPE and MdE scores, GA-RFOS-ELM yields the best performance for this task. In addition, OS-ELM, GA-OS-ELM and GA-RFOS-ELM are also better than four other nononline models. The AR model is more suitable for univariate prediction problems, and thus we did not consider it in this case.

Different from dataset 2, four indices about the news during the COVID-19 pandemic are combined with the input variable to predict the future price volatility of crude oil. The mark * in Table 5 denotes that the values of the loss function combined with news during the COVID-19 pandemic as input variables would achieve better performance. Obviously, combining novel coronavirus pneumonia data with the news, we confirm that models primarily outperform the basic models based on these four evaluation measures. This demonstrates that news about COVID-19 data could draw the attention and expectations of the crude oil futures market, thus serving an integral role in oil price volatility analysis and forecasting.

Table 6 displays the p values of the Diebold-Mariano (DM) testing of GA-RFOS-ELM and seven other methods based on different loss functions, like RMSE, MAE, MAPE and MdE. The DM test here is implemented by (DIEBOLD and MARIANO, 1995) with modification suggested by (Harvey et al., 1997). A small p value (usually less than 0.05) indicates that the proposed GA-RFOS-ELM model significantly outperforms the models named in the head-column.

The p values of DM testing of GA-RFOS-ELM and AR, RT, Bayes, SVR, ELM, OS-ELM and GA-OS-ELM are all below the significance level, which proves that the proposed GA-RFOS-ELM model outperforms other models in predicting the future price volatility of crude oil. These experiences show the superiority of the proposed model. Meanwhile, they prove the role of news during the COVID-19 pandemic.

To summarize, the 19-nCoV related news, which consists of a growing body of empirical and theoretical studies about public news sentiment, is more crucial for the volatility prediction of crude oil futures and was valuable with respect to stock returns. On the other hand, compared with the global learning, the prediction accuracy is lower when confronted with time-varying variables, which is more suitable for

Table 5
Predicting results of crude oil price volatility for dataset 3.

Model	RMSE	MAE	MAPE	MdE
RT	5.5843	2.7119*	2.5532*	1.3566*
Bayes	2.4563*	2.1265*	3.0967*	2.1248*
SVR	1.8648*	1.4205*	1.6362*	1.2858*
ELM	3.3186*	2.6460*	3.1073	2.3150*
OS-ELM	1.6958*	1.3527*	1.2472*	1.1727*
GA-OS-ELM	1.2831*	1.0356*	1.0290*	0.8768*
GA-RFOS-ELM	0.9767*	0.7260*	0.8477*	0.5418*

* denotes that the values of the loss function which consider news during the COVID-19 pandemic achieve better performance.

Table 6
DM statistic for different loss functions between GA-RFOS-ELM with benchmark models.

Model	RMSE	MAE	MAPE	MdE
AR	$1.43e^{-10}$	$1.55e^{-12}$	$1.56e^{-5}$	$5.52e^{-8}$
RT	$9.51e^{-3}$	$2.77e^{-4}$	$3.20e^{-2}$	$2.20e^{-2}$
Bayes	$8.06e^{-9}$	$3.28e^{-11}$	$1.47e^{-5}$	$1.54e^{-6}$
SVR	$1.38e^{-2}$	$2.02e^{-4}$	$3.13e^{-3}$	$4.24e^{-3}$
ELM	$1.73e^{-3}$	$8.96e^{-8}$	$8.87e^{-5}$	$7.81e^{-4}$
OS-ELM	$1.48e^{-2}$	$1.44e^{-4}$	$7.68e^{-3}$	$3.11e^{-3}$
GA-OS-ELM	$3.54e^{-2}$	$2.64e^{-3}$	$5.44e^{-3}$	$1.74e^{-2}$

local learning with long and short term memory. These empirical results show that news during the COVID-19 pandemic affects individual investors' decisions, but news exhibits timeliness due to information explosion, which is also consistent with the basic cognition of human beings.

5. Conclusion and future work

The influence of public news on the crude oil future market has attracted increasing attention over the last several years. There is a growing body of theoretical and empirical research on the relationship between social, economic or politically-driven news and changes in financial markets (Smales, 2014; Broadstock and Zhang, 2019; Shi and Ho, 2020). This paper further studies how news generated by coronavirus-related events is associated with the volatility of crude oil futures. While the current COVID-19 pandemic was related with great losses by investors all over the word, the existing studies about crude oil future market and the COVID-19 pandemic are limited. Therefore, weighting the predictive power of the COVID-19 pandemic becomes a profound and urgent issue.

This paper establishes a novel model and offers empirical evidence about volatility forecasting of crude oil futures by COVID-19 related news. Employing the GA-RFOS-ELM model shows that the crude oil futures present dependencies with contagion, media coverage, fake news and other information related to the COVID-19 pandemic. This result suggests that COVID-19 related news affects the price of crude oil futures, which has a certain explanatory power for the volatility of crude oil futures. Additionally, these results also are parallel to the dependence between news during COVID-19 financial turmoil and stock market returns (Cepoi, 2020). This is consistent with the study sentiment that COVID-19 related news is associated with volatility in the equity markets (Haroon and Rizvi, 2020). Therefore, we suggest a need for governments and social media to more intensively utilize positive information and applicable interactive communication platforms to mitigate ongoing coronavirus outbreak related financial market turmoil.

This paper offers two primary aspects of contribution. On the one hand, extending the rich literature of volatility forecasting, we introduce a novel GA-RFOS-ELM model. News during the COVID-19 pandemic exhibits timeliness, where new data attract more emphasis, while older data are gradually being forgotten. Considering the timeliness of COVID-19 related news, the GA-RFOS-ELM model strengthens the optimal search ability of the genetic algorithm. It may be effective and efficient to utilize a chunk-by-chunk learning mechanism with fixed or varying chunk size, which illustrates that online update learning ability is needed. Additionally, the crude oil price is nonlinear and nonstationary. Faced with these questions, this paper investigates the impact of the COVID-19 pandemic on volatility forecasting of crude oil futures by using the GA-RFOS-ELM model of news during the COVID-19 pandemic. These empirical results suggest that the forecasting performances of the model are superior to those of basic forecasting techniques. We demonstrate the power of news during the COVID-19 pandemic to improve the forecasting performance, which suggests that news during

the COVID-19 pandemic is a practical way to assist the prediction of volatility in the crude oil market and quantify investor emotion.

On the other hand, this paper adds to the research on commodity futures market response to the COVID-19 pandemic. There is plenty of evidence to suggest that news of major events generally contains important incremental predictive information on future market return (Albulescu, 2020; Zhang et al., 2020a; Al-Awadhi et al., 2020). This paper concludes that COVID-19 related news provides vital clues about the volatility of crude oil futures. The research has significant implications. With respect to policy insight, the regulator must consider the impact of news during the COVID-19 pandemic in formulating policy measures to mitigate the turbulence and instability of the crude oil market. Regarding market insight, investors should be conscious of the degree of news during COVID-19 and analyze the potency of the future crude oil futures market return.

Generally, this study is valuable and helpful for establishing more accurate forecasting models for crude oil futures with internet information. In addition, the GA-RFOS-ELM model could be applied to predictions of other temporal variables or other markets, such as corn, copper and gold, thus providing other more accurate models for volatility forecasting in future research, especially considering other variables. Additionally, constructing other indices by text mining of COVID-19 related news would be of particular interest.

Author statement

Futian Weng: Data analysis, Writing - original draft, Writing - review & editing.

Hongwei Zhang: Data curation, Supervision, Conceptualization, Writing - review & editing.

Cai Yang: Conceptualization, Supervision, Methodology, Software, Writing - review & editing, Validation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

References

Akaike, H., 1974. A new look at the statistical model identification. IEEE Trans. Automat. Contr. 19, 716–723.

Al-Awadhi, A.M., Al-Saifi, K., Al-Awadhi, A., Alhamadi, S., 2020. Death and contagious infectious diseases: impact of the covid-19 virus on stock market returns. J. Behav. Exp. Finance 100326.

Albulescu, C., 2020. Coronavirus and Financial Volatility: 40 Days of Fasting and Fear arXiv preprint arXiv:2003.04005.

Alpaydin, E., 2020. Introduction to Machine Learning. MIT press.

Baker, S.R., Bloom, N., Davis, S.J., Kost, K.J., Sammon, M.C., Viratyosin, T., 2020. The Unprecedented Stock Market Impact of COVID-19. National Bureau of Economic Research. Technical Report.

Bange, M.M., 2000. Do the portfolios of small investors reflect positive feedback trading? J. Financ. Quant. Anal. 239–255.

Barber, B.M., Odean, T., 2008. All that glitters: the effect of attention and news on the buying behavior of individual and institutional investors. Rev. Financ. Stud. 21, 785–818.

Bash, A., Alsaifi, K., 2019. Fear from uncertainty: an event study of khashoggi and stock market returns. J. Behav. Exp. Finance 23, 54–58.

Blendon, R.J., Benson, J.M., DesRoches, C.M., Raleigh, E., Taylor-Clark, K., 2004. The public's response to severe acute respiratory syndrome in toronto and the United States. Clin. Infect. Dis. 38, 925–931.

Blitz, D., Huisman, R., Swinkels, L., van Vliet, P., 2019. Media attention and the volatility effect. Finance Res. Lett. 101317.

Broadstock, D.C., Zhang, D., 2019. Social-media and intraday stock returns: the pricing power of sentiment. Finance Res. Lett. 30, 116–123.

Celaya, E., Agostini, A., 2015. Online em with weight-based forgetting. Neural Comput. 27, 1142–1157.

Cepoi, C.O., 2020. Asymmetric dependence between stock market returns and news during covid19 financial turmoil. Finance Res. Lett.

Chen, Y., Xie, X., Zhang, T., Bai, J., Hou, M., 2020. A deep residual compensation extreme learning machine and applications. J. Forecast. 39 (6), 986–999.

Chung, H., Shin, K.s., 2020. Genetic algorithm-optimized multi-channel convolutional neural network for stock market prediction. Neural Comput. Appl. 32, 7897–7914.

Dai, B., Gu, C., Zhao, E., Zhu, K., Cao, W., Qin, X., 2019. Improved online sequential extreme learning machine for identifying crack behavior in concrete dam. Adv. Struct. Eng. 22, 402–412.

Diebold, F.X., Mariano, R.S., 1995. Comparing predictive accuracy. J. Bus. Econ. Stat. 13.

Groß-Klußmann, A., Hautsch, N., 2011. When machines read the news: using automated text analytics to quantify high frequency news-implied market reactions. J. Empir. Finance 18, 321–340.

Guo, W., Xu, T., Tang, K., Yu, J., Chen, S., 2018. Online sequential extreme learning machine with generalized regularization and adaptive forgetting factor for time-varying system prediction. Math. Probl Eng. 2018.

Haroon, O., Rizvi, S.A.R., 2020. Covid-19: media coverage and financial markets behavior—a sectoral inquiry. J. Behav. Exp. Finance 100343.

Harvey, D., Leybourne, S., Newbold, P., 1997. Testing the equality of prediction mean squared errors. Int. J. Forecast. 13, 281–291.

Herrera, A.M., Hu, L., Pastor, D., 2018. Forecasting crude oil price volatility. Int. J. Forecast. 34, 622–635.

Holland, J.H., et al., 1992. Adaptation in Natural and Artificial Systems: an Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence. MIT press.

Huang, G., Huang, G.B., Song, S., You, K., 2015. Trends in extreme learning machines: a review. Neural Network. 61, 32–48.

Huang, G.B., Zhu, Q.Y., Siew, C.K., 2004. Extreme learning machine: a new learning scheme of feedforward neural networks. In: 2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No. 04CH37541). Ieee, pp. 985–990.

Huang, G.B., Zhu, Q.Y., Siew, C.K., 2006. Extreme learning machine: theory and applications. Neurocomputing 70, 489–501.

Huynh, H.T., Won, Y., 2011. Regularized online sequential learning algorithm for single-hidden layer feedforward neural networks. Pattern Recogn. Lett. 32, 1930–1935.

Kowalewski, O., Śpiewanowski, P., 2020. Stock market response to potash mine disasters. J. Commodity Markets 100124.

Leardi, R., Seasholtz, M.B., Pell, R.J., 2002. Variable selection for multivariate calibration using a genetic algorithm: prediction of additive concentrations in polymer films from fourier transform-infrared spectral data. Anal. Chim. Acta 461, 189–200.

Liang, N.Y., Huang, G.B., Saratchandran, P., Sundararajan, N., 2006. A fast and accurate online sequential learning algorithm for feedforward networks. IEEE Trans. Neural Network. 17, 1411–1423.

Loh, W.Y., 2011. Classification and regression trees. Wiley Interdisciplinary Reviews: Data Min. Knowl. Discov. 1, 14–23.

Lopatta, K., Alexander, E.K., Gastone, L.M., Tammen, T., 2020. To Report or Not to Report about Coronavirus? the Role of Periodic Reporting in Explaining Capital Market Reactions during the Global Covid-19 Pandemic. The Role of Periodic Reporting in Explaining Capital Market Reactions during the Global COVID-19 Pandemic.(April 3, 2020).

Lopez, J.A., 2001. Evaluating the predictive accuracy of volatility models. J. Forecast. 20, 87–109.

Luo, J., Klein, T., Ji, Q., Hou, C., 2019. Forecasting realized volatility of agricultural commodity futures with infinite hidden markov har models. Int. J. Forecast.

Mairal, G., 2011. The history and the narrative of risk in the media. Health Risk Soc. 13, 65–79.

Man, Z., Lee, K., Wang, D., Cao, Z., Miao, C., 2011. A new robust training algorithm for a class of single-hidden layer feedforward neural networks. Neurocomputing 74, 2491–2501.

Mitchell, M., 1998. An Introduction to Genetic Algorithms. MIT press.

Narayan, P.K., 2019. Can stale oil price news predict stock returns? Energy Econ. 83, 430–444.

Nofsinger, J.R., Sias, R.W., 1999. Herding and feedback trading by institutional and individual investors. J. Finance 54, 2263–2295.

Onali, E., 2020. Covid-19 and Stock Market Volatility. Available at: SSRN 3571453.

Paleologu, C., Benesty, J., Ciochina, S., 2008. A robust variable forgetting factor recursive least-squares algorithm for system identification. IEEE Signal Process. Lett. 15, 597–600.

Pan, F., Zhao, H.b., 2013. Online sequential extreme learning machine based multilayer perception with output self feedback for time series prediction. J. Shanghai Jiaot. Univ. 18, 366–375.

Salisu, A.A., Oloko, T.F., 2015. Modeling oil price–us stock nexus: a varma–bekk–agarch approach. Energy Econ. 50, 1–12.

Shanaev, S., Ghimire, B., 2019. Is all politics local? regional political risk in Russia and the panel of stock returns. J. Behav. Exp Finance 21, 70–82.

Shi, Y., Ho, K.Y., 2020. News sentiment and states of stock return volatility: evidence from long memory and discrete choice models. Finance Res. Lett. 101446.

Smales, L.A., 2014. News sentiment and the investor fear gauge. Finance Res. Lett. 11, 122–130.

Soares, S.G., Araújo, R., 2016. An adaptive ensemble of on-line extreme learning machines with variable forgetting factor for dynamic system prediction. Neurocomputing 171, 693–707.

Takeda, F., Wakao, T., 2014. Google search intensity and its relationship with returns and trading volume of Japanese stocks. Pac. Basin Finance J. 27, 1–18.

Tetlock, P.C., 2007. Giving content to investor sentiment: the role of media in the stock market. J. Finance 62, 1139–1168.

Wang, J., Athanasopoulos, G., Hyndman, R.J., Wang, S., 2018. Crude oil price forecasting based on internet concern using an extreme learning machine. Int. J. Forecast. 34, 665–677.

Weng, F., Chen, Y., Wang, Z., Hou, M., Luo, J., Tian, Z., 2020. Gold price forecasting research based on an improved online extreme learning machine algorithm. J. Ambient Intelligence and Humanized Computing 1–11.

Young, M.E., King, N., Harper, S., Humphreys, K.R., 2013. The influence of popular media on perceptions of personal and population risk in possible disease outbreaks. Health Risk Soc. 15, 103–114.

Zhang, D., Hu, M., Ji, Q., 2020a. Financial Markets under the Global Pandemic of Covid-19. Finance Research Letters, p. 101528.

Zhang, P., Huang, Y., Li, M., Wang, Y., Wen, P., Wang, S., Chen, S., 2020b. Fault diagnosis method of analog circuit based on ga-os-elm. In: 2020 International Conference on Sensing, Diagnostics, Prognostics, and Control (SDPC). IEEE, pp. 273–278.

Zhang, Y.J., Wang, J., 2015. Exploring the wti crude oil price bubble process using the markov regime switching model. Phys. Stat. Mech. Appl. 421, 377–387.

Zhou, Y., Wang, Y., Wang, K., Kang, L., Peng, F., Wang, L., Pang, J., 2020. Hybrid genetic algorithm method for efficient and robust evaluation of remaining useful life of supercapacitors. Appl. Energy 260, 114169.