Article

Forecasting Oil Price Volatility in the Era of Big Data: A Text Mining for VaR Approach

Lu-Tao Zhao ^{1,2} , Li-Na Liu ¹, Zi-Jie Wang ¹ and Ling-Yun He ^{2,3,4,*}¹ School of Mathematics and Physics, University of Science and Technology Beijing, Beijing 100083, China² Center for Energy and Environmental Policy Research & School of Management and Economics, Beijing Institute of Technology, Beijing 100081, China³ School of Economics, JiNan University, Guangzhou 510632, China⁴ School of Economics and Management, Nanjing University of Information Science and Technology, Nanjing 210044, China

* Correspondence: lyhe@amss.ac.cn

Received: 18 June 2019; Accepted: 16 July 2019; Published: 17 July 2019

Abstract: The rapid fluctuations in global crude oil prices are one of the important factors affecting both the sustainable development and the green transformation of the global economy. To accurately measure the risks of crude oil prices, in the context of big data, this study introduces the two-layer non-negative matrix factorization model, a kind of natural language processing, to extract the dynamic risk factors from online news and assign them as weighted factors to historical data. Finally, this study proposes a giant information history simulation (GIHS) method which is used to forecast the value-at-risk (VaR) of crude oil. In conclusion, this paper shows that considering the impact of dynamic risk factors from online news on the VaR can improve the accuracy of crude oil VaR measurement, providing an effective tool for analyzing crude oil price risks in oil market, providing risk management support for international oil market investors, and providing the country with a sense of risk analysis to achieve sustainable and green transformation.

Keywords: oil price volatility; risk identification; VaR; big data; natural language processing; two-layer non-negative matrix factorization

1. Introduction

As a strategic resource, crude oil is the foundation of global economic development and global commodity markets [1,2]. Slight fluctuations in crude oil prices stimulate the development of the world economy. Abnormal fluctuations in crude oil prices, however, unleash clear signals for the economy to pinpoint and solve the problems as soon as possible. Therefore, oil prices are closely related to the sustainable development of the world economy. In recent years, based on an increasingly complex global crude oil market environment, the uncertainty of crude oil price volatility has been increased by many factors [3–5]. For example, emergent events and political and economic events, (e.g., oil workers' strike action, financial crises, and two Gulf Wars) have severely affected the supply–demand balance of crude oil markets, which has resulted in more complex, rapidly changing crude oil risks [6]. Therefore, the fluctuations of global oil prices have caused global concerns: how to improve the accuracy of VaR forecasting and how to conduct risk management have become the focus of scholars [7].

In the financial industry, a measure widely applied in market risk measurement [8–12], value-at-risk (VaR), is defined as the maximum possible loss of a portfolio over a certain fixed time at a certain probability [13]. It accurately measures risks, and has thus become an international standard for risk measurement [14–16]. The Basel Accord clearly states the importance of VaR for monitoring financial risks and determining capital amounts, requiring all financial institutions to implement VaR [17].

As a kind of an ordinary commodity, crude oil has normal attributes, and its risk value is initially determined by fundamental factors (supply and demand) [18–22]. At the same time, crude oil is also a financial product and basic strategic resource, with political and financial attributes additional to its normal ones [21,23], and its risk value is often severely affected by non-fundamental factors such as oil market speculation, climate, wars, and the environment. Non-fundamental factors often lead to psychological changes particularly in oil market investors, further leading the risk value to deviate from its fundamental value [5,24,25].

At present, research into VaR measurement methods for various financial assets has been extensively conducted, and the applicability of the VaR method is continuously being improved and expanded from three aspects: non-parametric calculation method research, better estimates of VaR parameters, and tail loss simulation [26]. At present, there are two main non-parametric methods of VaR assessment: historical simulation (HS) and the Monte Carlo (MC) methods. For instance, Hendricks [13] used historical data applied to the equal weight moving average method and HS method to calculate VaR, and the experiment results showed that both two methods had almost the same effect in market risk measurement. Some improvements have also been made by scholars to make the HS method suit financial markets better. For example, Richardson et al. [27] proposed the use of an exponential weight for historical time series data (EWHs), which gives more weight to observed data that are closer to the present. Combining the HS method with an autoregressive moving average model (ARMA) model, David Cabedo and Moya [28] developed a novel historical simulation method to calculate Brent spot oil VaR based on the autoregressive moving average model (HSAF). By comparing their data with those from the HS method and the variance–covariance method based on the autoregressive conditional heteroskedasticity model (ARCH) model, they found that the HSAF method had the best forecasting effect. Some scholars have studied its model effect: for instance, Sadeghi and Shavvalpour [29] calculated organization of the petroleum exporting countries (OPEC) oil VaR with the HSAF method and the generalized autoregressive conditional heteroscedasticity (GARCH) method based variance–covariance approach, finding that the HSAF method was more effective. To ensure that the VaR metrics are consistent, Giannopoulos and Tunaru [30] proposed a filtered historical simulation (FHS) method which measures loss, and the experiment results showed that the FHS method performed better. In addition to the HS method and its improved version, a series of research projects aimed at improvements to MC simulation methods have also been undertaken: Jamshidian and Zhu [31] proposed a scenario MC method. Based on principal component analysis (PCA), they used a polynomial distribution to discretize market factors and generate a finite number of scenarios with different probabilities, which greatly simplified the calculation. Skiadopoulos et al. [32] used two different sizes of sample to simulate Greek stock and bond prices through the HS method and the MC method, concluding that the MC method was more effective. To obtain characteristics of stocks, Panigirtzoglou and Skiadopoulos [33] used the PCA method and then proposed a no-arbitrage MC simulation method for VaR estimation. Tezuka et al. [34] used the MC method for forecasting VaR and then established a financial risk management system. Dionne et al. [35] used tick-by-tick data to study the intraday VaR of the Toronto Stock Exchange. It is of significance to active investors in the stock market to use the MC method in the study of the daily risk value of irregularly spaced high-frequency data, and to prove that the model passes the back-testing criteria. Tzeng et al. [36] developed a randomized quasi-MC method to compensate for the shortcomings of quasi-MC method. At present, most of these methods are simple and widely used, but they mainly measure risks from a static perspective, and cannot be applied to dynamic markets well [26].

The second aspect is how to estimate the parameters of the models that calculate VaR. The addition of the GARCH model greatly enriches this aspect. For example, Mitnik and Paoletta [37] modeled returns on East Asian currencies to the US dollar exchange rate. To fit the skewness and kurtosis in VaR models, they considered AR(1), AR(1)-GARCH(1, 1) and AR(1)-APARCH(1, 1) models with a generalized t-distribution and asymmetry. Finally, experiment results showed that the asymmetric power ARCH (APARCH) model performed better than other models. Angelidis et al. [38] calculated

the VaR of two stock portfolios and proved that a filtered historical simulation method combined with GARCH volatility (FHS-G) was the best model, according to the Kupiec back-testing results. Chen et al. [39] conducted VaR estimations of four Asia-Pacific stock market portfolios: the results showed that the GARCH model was superior to the stochastic volatility model. Krause and Paolella [40] developed a rapid method for short-term forecasting of VaR based on GARCH processes driven by non-central innovations, which showed that the method is accurate and suitable for small samples. Youssef et al. [41] used different GARCH models to simulate crude oil VaR, and reported that the APARCH model with fractional integration performed better when forecasting VaR. On the other hand, when fitting the GARCH model to financial time series, the forecasted volatility is used to adjust the weight of observed historical data [42]. For example, to improve the forecasting of the stock index, Hull and White [43] forecasted volatility to update the current level of volatility. Adesi et al. [44] fitted historical data into the GARCH model, and the errors generated in this process were considered as changes in the forecast distribution. Considering the advantage of CGARCH model, Karmakar and Paul [45] used the CGARCH-EVT-copula method to model the marginal distribution, and the Kupiec back-testing results indicated that this model performed better than other models. Ardia et al. [46] found that the MSGARCH model can provide better VaR, ES, and left-tail distribution forecasting than the single mechanism model.

The third direction is simulating the tail loss of time series. Since GARCH models do not fit tail loss well, some improvements have been made to this model. For example, Hung et al. [47] modeled the prices of various crude oil and proposed a GARCH model with a heavy tail (HT) distribution, and the empirical results showed that this model forecast better. However, Normal-GARCH model and the t-GARCH model overestimate and underestimate tail risk, respectively. Extreme value theory (EVT) method and the copula function are therefore introduced into VaR methods. For example, Gencay and Selçuk [48] concluded that the EVT method performed better under extreme market conditions. Chan and Gray [49] used EVT to simulate the tail of the returns distribution. Statistical tests showed that the EVT-based model performed well in forecasting out-of-sample VaR, and proposed that it is a useful technique for forecasting VaR in the electricity market. Zhao et al. [50] used the copula-VaR method and HS to calculate the energy portfolio VaR, and pointed out that the investment plan with a minimum VaR energy portfolio can help to reduce the risks of a single energy source. In recent years, many scholars have done a lot of research on quantile regression methods. Fuertes and Olmo [51] proposed a conditional quantile forecast encompassing (CQFE) reasoning as a new model with which to improve downward tail risk forecasting. Based on the volatility, the intra-day range and overnight returns, Meng and Taylor [52] proposed a new quantile regression model, the empirical results of which showed that the model improved the forecasting accuracy of VaR.

In addition to the above traditional methods of forecasting oil prices, many scholars have introduced textual information into oil price forecasts. A large number of studies have found that text mining is a powerful tool for studying oil prices and their volatility. Prusa et al. [53] proposed a new model for adding text features, which effectively improved the accuracy of oil price forecasting compared with models using only historical oil price data. Li et al. [54] pointed out that textual data can help analyze the future trend of crude oil prices, and proposed an oil price forecasting model based on investor sentiment. They empirically showed that the model had a good forecasting power for oil price trends in a statistical sense. Chuaykoblap et al. [55] pointed out that news text data mining is a widely used algorithm to forecast crude oil price fluctuations, and finally proposed an expert-based text mining model. The model can effectively filter out noisy data and the accuracy was greatly improved compared with the traditional use of historical oil price data. Oussalah and Zaidi [56] used text mining to analyze Twitter information on US foreign policy and oil companies in order to forecast weekly WTI crude oil price movements, with forecasting accuracy that exceeded the models listed in the existing literature. Zhao and Zeng [57] discovered the link between crude oil price trends and news texts, and finally used support vector machines to study the timeliness of oil-related news.

In summary, scholars have made a series of research achievements in risk measurement. However, most have mainly considered the fluctuation characteristics of returns, or used a large amount of historical data to forecast VaR; few have used factors in online oil-related news (oil risk pheromones) that influence the fluctuations of oil prices to forecast VaR. A large amount of literature has confirmed that oil-related events can easily lead to increasing fluctuations in oil risks [58–62]. Therefore, it is reasonable and necessary to consider online news to assist VaR forecasting. The structural figure for forecasting oil price fluctuations using text mining is shown in Figure 1.

Figure 1. Oil price text mining structure.

As shown in Figure 1, the structure consists of two branches. The left branch mainly performs text-related operations. Firstly, massive online oil texts are obtained through a Python program. Secondly, text pre-processing operations are performed. Finally, oil texts are mined to extract the oil risk topics and weight. The right branch mainly conducts statistics and tests on oil returns. Finally, oil returns and risk topic weight are combined to forecast VaR. This is the general idea behind the use of text mining to predict the volatility of oil prices.

Considering the characteristics of financial sequence fluctuations, the purpose of this work is to use natural language processing to explore the role of oil risk pheromones in oil VaR forecasting in a big data context, and to propose a novel model to improve the accuracy of oil risk forecasting. Our original points are as follows:

- (1) In identifying energy market risks, many scholars have conducted a series of studies on risks, however, the dynamic evolutions of risks have not been considered. Therefore, in this paper, text mining is used to extract risks of oil in order to identify the risk topics and evolution process of the oil market. Text mining could overcome the shortcomings of traditional risk identification, such as strong subjectivity and weak timeliness, and provides a new perspective for energy market researchers.
- (2) The information of oil market texts is introduced in this paper when energy market risks are measured. Considering the interaction between oil risks and the VaR of oil price, we propose a GIHS model that not only pays attention to the feedback effect of oil risks on the VaR of oil price, but also improves the weight of historical data affecting oil price VaR. The Kupiec back-testing results show that the proposed GIHS model has better prediction accuracy than others.

The remainder of paper is organized as follows: Section 2 introduces the research method construction, Section 3 presents data sources and data processing, Section 4 describes empirical research and analysis of the results, Section 5 presents conclusions and recommendations.

2. Methods

The giant information history simulation model (GIHS) we propose can be divided into three modules. The first module considers how to model massive online oil-related news, the second module is how to build a GIHS model, and we finally evaluate the performance of our proposed model.

2.1. Topic Modeling: Oil-Related News

The topic model, the concept of which originates from latent semantic analysis (LSA), is a method for mining semantic information in text. It reduces the dimensionality of the target data by mapping a collection of high-dimensional words to a low-dimensional topic space, and the dimensional reduction method has good interpretability. At present, the most widely used model is the Latent Dirichlet Allocation (LDA) model, but this is a static topic model with the premise that topics do not change with time, which obviously does not match reality. The Dynamic Topic Model (DTM) solves this problem. However, the number of topics in DTM under each time window cannot change with time, which also differs from reality. The non-negative matrix factorization model (NMF) can also model text to track topics, and Greene and Cross [63] developed a two-layer NMF topic model that identifies the dynamic processes of topics based on topic modeling. Compared with general topic models, the two-layer NMF model is more likely to produce a variety of semantically coherent topics. Therefore, this study applied the two-layer NMF to oil-related news to get the value of various types of risk factor that influence oil prices.

2.2. GIHS Model

Although many improvements have been made in VaR calculation, due to general problems and their complexity, the current calculation methods of VaR mainly include the variance–covariance method, MC method, and HS method, where the HS method is widely used because of its ease of operation and efficacy; however, despite many optimizations of the HS method having been made, the most applicable method remains the HSAF model. Considering the impact of oil-related news on oil price risks, we used the two-layer NMF model in natural language processing to mine a massive information set of oil-related news items and extract the risk factors therefrom. Taking values of risk factors as the weight of historical data, we forecast VaR based on an HSAF model and propose a giant information historical simulation (GIHS) model, for which the algorithm is as follows:

Input: Brent oil returns (r_t).

Output: Brent oil returns VaR.

Step 1. Calculate the absolute value of oil returns.

$$R_t = |r_t| \quad (1)$$

Step 2. Smooth oil returns.

$$R_t' = a \sum_{s=0}^{t-1} (1-a)^s R_{t-s} \quad (2)$$

where $a = 0.97$, which is the smoothing factor, and R_t' denotes returns after smoothing.

Step 3. Establish an ARMA model, then calculate the forecasted value f_t .

$$f_t = \beta_0 + \beta_1 R_{t-1}' + \dots + \beta_p R_{t-p}' + \varepsilon_t + \alpha_1 \varepsilon_{t-1} + \dots + \alpha_q \varepsilon_{t-q} \quad (3)$$

Step 4. Reconstruct the historical data sequence of error E_t . Assign the risk weights w_t at time t to the forecasted error e_t at time t , to construct a new error sequence E_t . The risk weights w_t are derived from the results of modeling the massive news dataset using the two-layer NMF model, and they represent the probability of risks occurring on day t . By running the two-layer NMF model, two kinds of output are presented: Output 1 includes a number of topics, each of which includes many words; Output 2 represents the total probability of all texts belonging to different topics at different times. For example, if 12 topics need to be analyzed in Table 4, and we calculate 204 time windows from October, 2001 to September, 2018, then Output 2 is a matrix of 204×12 , and each column of this matrix is the value of each topic change with time. Each row of this matrix is the probability of occurrence of each topic at each moment, with a total probability of w_t . For example, in October, 2001, the total probability of the occurrence of 12 topics is w_1 , and so on. We then re-establish the error sequence E_t as shown below:

$$\begin{aligned} E_1 &= e_1, E_2 = e_1, \dots, E_{w_1} = e_1, \\ E_{w_1+1} &= e_2, E_{w_1+2} = e_2, \dots, E_{w_1+w_2} = e_2, \\ E_{w_1+w_2+1} &= e_3, E_{w_1+w_2+2} = e_3, \dots, E_{w_1+w_2+w_3} = e_3, \\ &\dots \\ E_{w_1+w_2+\dots+w_{t-1}+1} &= e_t, E_{w_1+w_2+\dots+w_{t-1}+2} = e_t, \dots, E_{w_1+w_2+\dots+w_{t-1}+w_t} = e_t \end{aligned} \quad (4)$$

It is considered that the farther away from the current observations, the smaller the impact on future oil prices, which should have less weight. Therefore, we focus on the “frequency” of each observation. By changing the “frequency,” different emphases are put on different observations. After the risk weights of every moment are assigned to the residual error, the error sequence E_t is reconstructed so that the “frequency” of the error at each observation is changed. For example, if w_1 is 55.3, after adjustment, we will construct $55e_1$ in the error sequence to increase the impact of e_1 , and so on.

Step 5. Calculate VaR.

$$VaR = f_t + q_t \quad (5)$$

where f_t denotes forecasted value and q_t is the quantile of corresponding to error sequence E_t .

2.3. VaR Estimation Performance of the GIHS Model

To analyze the accuracy of the GIHS model, which measures oil price volatility, and to judge whether or not, the GIHS model fully forecasts the actual risks in the oil market, we introduce a likelihood ratio (LR) test, provided by Kupiec [64]. The core idea of the method is as follows: consider that T is the total number of Brent oil return observations, N is the excess number of VaR violations, α_0 is the specified VaR level, and $f = N/T$ and $1 - \alpha_0$ denote the empirical failure rate and the theoretical failure rate, respectively. Therefore, the LR statistic in the existence of the null hypothesis is calculated by:

$$LR = -2 \ln \{ \alpha_0^N (1 - \alpha_0)^{T-N} \} + 2 \ln \{ (f)^N (1 - f)^{T-N} \} \quad (6)$$

The LR statistic is used to test whether the empirical failure rate is statistically equal to theoretical failure rate or not, which has a $\chi^2(1)$ distribution under the null hypothesis. The smaller the LR value, the more precise the model.

3. Data Description

According to the two-layer NMF model and the GIHS model, our data were divided into two parts, the news text, and the oil returns.

Figure 3. Monthly Brent spot prices.

Figure 4. Monthly Brent spot returns.

As shown in Figure 3, the Brent oil prices have undergone significant fluctuations, and were extremely unstable during the period 2007–2008. Therefore, it is of great importance to forecast oil price fluctuations using an appropriate method.

It can be seen from Figure 4 that Brent oil returns are highly volatile, which also reflects the existence of heteroscedasticity. The ups and downs of positive and negative Brent oil returns indicate that it is essential to conduct oil market risk measurement.

The figures of histogram distribution against normal distribution for monthly Brent oil returns are depicted in Figure 5. As shown in Figures 5a and 4b, the distribution of monthly Brent oil returns differs from normal distribution significantly, and exhibits asymmetry and fat tails. Therefore, traditional VaR forecasting methods of assuming that oil returns follow the normal distribution are no longer applicable.

Figure 5. Brent histogram against normal distribution and returns distribution.

4. Empirical Analysis

4.1. Oil-Related News Clustering Results

In this section, to identify the risk factors in the oil market, we applied the two-layer NMF model to annual time-stamped news, because annual time-stamped news periods are long and the amount of information is large, which can better reflect the macro trend of one topic. Therefore, we used annual

time-stamped news to explore many risk topics affecting oil prices. According to the topic coherence, the number of dynamic topics was 12, and the results are summarized in Table 2:

Table 2. The annual results of NMF topic clustering.

Topic	Top Words
1	energy emission climate carbon power change greenhouse plant coal global
2	cent crude gallon price gasoline york oil inventory barrel average
3	oil company production field barrel drill exploration offshore energy shell
4	opec saudi oil arabia production output export cut market crude
5	russia russian ukrainemoscow putin ukrainian european kiev europe minister
6	iran nuclear iranian sanction tehran pakistan india korea program weapon
7	police kill fire report game attack shoot city force official
8	price rise stock fell dollar trade market gain rate yen
9	china chinese beijing korea japan south coal import trade north
10	iraq iraqi baghdad oil government war kurdish saddam unite force
11	vehicle car fuel diesel sale hybrid engine ford electric motor
12	gas pipeline natural project energy cubic azerbaijan stream shale foot

For example, as can be seen from Table 2, Topic 1 mainly discusses news about the environment and climate; Topic 3 mainly includes oil companies and mining news; Topic 4 mainly consists of some oil-related news related to supply; Topic 6 is mainly composed of news related to nuclear sanctions in Iran; Topic 8 is mainly about crude oil prices, the market, and economy; Topic 9 mainly discusses energy demand; and Topic 10 mainly discusses geopolitics. These topics include fundamental factors (supply and demand) in oil markets, as well as non-fundamental factors such as the environment, climate, market, economy, geopolitics, and oil companies, further indicating that oil price fluctuations are the result of many factors. Therefore, topic modeling using online news to identify oil market risk factors is reasonable.

Unlike the common LDA topic model, the two-layer NMF model can also identify the evolution of each topic over time, which is the timeliness of the oil risk pheromones. We tracked Topic 6, and the results are summarized in Table 3.

Table 3. The birth and death process of Topic 6.

year	Top Words
2002	pakistan afghanistan india government war iran military sudan force terrorist
2003	korea nuclear north korean weapon pyongyang iran program south unite
2004	iran nuclear weapon uranium iranian tehran program iaea enrichment pakistan
2005	iran nuclear india iranian pakistan tehran indian pipeline delhi weapon
2006	iran nuclear iranian tehran sanction council uranium weapon program unite
2007	nuclear korea north iran south china talk korean reactor seoul
2008	iran iranian tehran nuclear sanction par islamic israel deal official
2009	iran pakistan iranian tehran india gas pipeline par islamabad sanction
2010	iran iranian sanction tehran nuclear program energy sector pressure gasoline
2011	iran iranian tehran india sanction pakistan delhi nuclear bank indian
2012	iran iranian sanction nuclear tehran oil european program weapon strait
2013	iran iranian sanction nuclear tehran pakistan program pakistani india oil
2018	iran sanction iranian trump oil nuclear deal tehran president unite

Table 3 shows the evolution of Topic 6 over time, which is closely related to Iranian nuclear sanctions. In this table, the term ‘nuclear’ jumped to the second place in 2003. At this time, Iran announced that they had extracted uranium, therefore, the Iranian nuclear issue began to become of popular concern. This topic disappeared after 2013 and did not reappear until 2018, which was basically consistent with the reality. In 2013, Iran and six other countries reached a phased agreement on the Iranian nuclear issue in Geneva, which brought the nuclear issue to a certain balance, and its impact on oil prices had

also reached a certain balance. Therefore, news no longer reported this topic, and it can be considered to have faded from the headlines within the period of 2014–2017. In 2018, the USA announced its withdrawal from the Iranian nuclear agreement, so Topic 6 reappeared.

Here, topics related to oil prices are regarded as risks. According to Table 3, the aforementioned development process of Topic 6 is defined as the birth and death process of risks, which is a special, discrete-state continuous time Markov process. The state of risk factors in the oil market is limited or countable (i.e., alive or dead), and the state change must be between adjacent states.

The above results were generated by the two-layer NMF clustering of annual news. However, VaR is currently measured by some large financial institutions like JP Morgan as the risk value of the assets held, and is disclosed regularly. At the same time, due to the volatility of the financial market and uncertain factors, risk managers are more concerned about the risk value of the assets within one month. In particular, for some highly liquid trading positions, the annual data often do not reflect the characteristics of their high frequency. Therefore, considering the actual application of VaR, we put the monthly news into the two-layer NMF model for clustering to forecast monthly oil VaR. According to the topic coherence score, the best number of topics was identified as 12, and the results are shown in Table 4.

Table 4. The monthly results of NMF topic clustering.

Topics	Top words
1	oil company gas production field natural energy pipeline drill exploration
2	cent gallon price gasoline york crude average oil heat mercantile
3	gas russia russian pipeline ukraine natural energy moscow european europe
4	opec oil saudi output production arabia cut producer price meet
5	iraq iraqi oil war government bush unite attack force baghdad
6	game score goal win season play shoot lead team night
7	price rise market rate stock economy growth bank quarter increase
8	energy power climate fuel plant change carbon coal emission china
9	iran nuclear iranian sanction tehran korea china pakistan unite india
10	vehicle car diesel emission test german fuel engine scandal carmaker
11	crude oil brent barrel future inventory gasoline data price supply
12	stock rise dollar fell yen close york gain trade euro

To track the evolution of each topic over time, we plotted the probability of the above 12 topics occurring in 2001–2018, as shown in Figure 6. These risk changes with oil prices information over time are defined as oil risk pheromones, for which the corresponding values are defined as the probability of risks occurring at different times.

The changes of monthly topic probability are presented in Figure 6. Usually, the sum of probabilities of all topics in this figure fluctuates only a little, however in 2003, the total probabilities were the largest in the whole sample interval. The main reason is that the probability of Topic 5 was relatively large. During this period, Topic 5 mainly discussed the Iraq war, which caused a certain degree of impact on oil prices. In addition, the ranges of all the topics were more or less the same, but each topic fluctuated more sharply. For example, the value of Topic 5 during 2001–2005 was relatively large, this topic is related to regional conflicts, and had a greater impact on oil prices. In 2006–2018, the probability of Topic 1 was relatively large, this topic is related to oil exploitation, which directly affects oil supply and demand, so it has a certain impact on oil prices. We extracted the values of these topics as error weights, changed the frequency of occurrence of errors in order to reconstruct the error sequence, and finally predicted oil price fluctuations.

Figure 6. The heat stream of oil risks.

In short, we extracted topics from online news in the oil market, where oil-related topics are also the risks that people often pay attention to in the oil market. The greater the risk, the more relevant texts of this risk, which are hot spots in the media, and the more oil price volatility will be presented. In particular, if there are unexpected events in the energy market, the impact on oil prices is relatively severe, the duration is relatively short, and the oil prices fluctuate more sharply. Therefore, there will be a large number of reports on some online media, and they will form a relatively strong positive relationship. We carried out the Granger causality test for risk and returns, the results of which are as shown in Table 5.

Table 5. Granger causality test for the relationship between risks and returns.

Causality	Lags							
	1	2	3	4	5	6	7	8
Risks→Returns	0.18	0.12	0.17	0.22	0.43	0.46	2.08	2.11
(p-Value)	0.67	0.89	0.92	0.93	0.83	0.83	0.05 **	0.04 **
Risks←Returns	0.43	0.14	0.12	0.15	0.34	0.90	0.76	0.70
(p-Value)	0.51	0.87	0.95	0.96	0.89	0.49	0.63	0.69

“→” and “←” represent one-way causality from the latter to the former and the former to the latter, with respect to the lags of 1 to 8 by Granger causality test. *p*-Value represents the degree of acceptance of the null hypothesis: returns/risks do not Granger-cause risks/returns. The lower the *p*-Value, the more significantly returns/risks Granger-cause risks/returns. ** represents the significant level at 5%.

As shown in Table 5, it can be concluded that risks Granger-cause returns significantly with the lags of 7 and 8. However, returns do not Granger-cause risks significantly. Therefore, the use of risks to assist in predicting oil price volatility is reasonable.

4.2. Analysis of Oil Price Volatility

This study made a quantitative analysis on oil returns: the basic statistical analysis of Brent crude oil spot returns is summarized in Table 6.

Table 6. Summary of basic statistical tests for monthly Brent returns.

Statistics	Value	Statistics	Value
Descriptive statistics:			
Mean	0.562	Skewness	−0.957
Maximum	19.598	Kurtosis	4.490
Minimum	−31.096	J.B test ¹	50.278 ***
Standard deviation	8.944		
Autocorrelation test:			
Q(5)	18.315 ***	Q ² (5)	70.956 ***
Q(10)	30.285 ***	Q ² (10)	73.965 ***
Unit roots and stationarity test:			
ADF ²	−10.951 ***	PP ³	−10.945 ***
KPSS ⁴	0.134 ***		

¹ J.B test is a normality test statistic. ² ADF, ³ PP, and ⁴ KPSS are unit roots and stationarity test statistics. Q(n) and Q²(n) are the Ljung-Box Q-statistic of order n on the returns and squared returns, respectively. *** denotes significance at the 1% level.

As shown in Table 6, the average of Brent oil returns was 0.562, which is consistent with the fact that oil returns fluctuate around the zero-value horizon and the mean value is close to zero. The skewness of Brent oil returns was −0.957, from which it can be concluded that Brent oil returns are skewed toward the left. The kurtosis was 4.49, which is significantly more than 3, showing leptokurtosis and significant fluctuations. The Jarque-Bera test results rejected the null hypothesis of a normal distribution at the 1% confidence level. Therefore, the traditional normal distribution assumption will produce large errors, so we choose to improve the non-parametric calculation method in VaR forecasting. According to the autocorrelation test, the Ljung-Box Q statistics of both the 5 and 10 order were significant, indicating that the oil returns have strong autocorrelation. In addition, modeling with ARMA requires that the time series be stationary. The ADF, PP, and KPSS statistics show that the null hypothesis of a unit root in Brent oil returns was rejected at the 1% confidence level, from which it can be concluded that it is reasonable to apply an ARMA-type model to fit Brent oil returns.

4.3. VaR Forecasting Results

We modeled the Brent oil returns from October 2001 to December 2011 to forecast the out-of-sample VaR from January 2012 to October 2018. The results are shown in Figure 7.

The traditional HS method treats historical data with equal weight, therefore, it is conservative when estimating VaR, its estimated fluctuation range is small, and its forecasting ability is poor. Considering that historical data have a certain impact on current data, we chose to smooth returns, giving the historical data a certain weight on current data, and we finally used the HS method to forecast VaR. This method (EWHS) represented an improvement on the basis of the HS method, but there was still a certain gap between forecast and actual returns. The HSAF method is based on the ARMA model, its VaR is the sum of forecasting value with ARMA, and the quantile of the corresponding error sequence. This method is further improved over the EWHS method. However, the weight of error considered by the HSAF model is equal. Here, we first took the absolute value of returns and smoothed the data, and we then assigned weights to the error according to the probability of the occurrence of risks in the two-layer NMF model, and the error sequence was thus reconstructed. The model has been further improved on the basis of the HSAF model, and it is superior to both the HS method and its improved version at 95%, 97.5%, and 99% confidence levels. The GIHS model is therefore more able to identify the risk fluctuations of oil prices, and provides the best forecasting

effect. In addition, the improvement of VaR forecasting accuracy further proved that there is indeed an interaction between oil price returns and online oil-related news.

(a) VaR at 95% confidence level.

(b) VaR at 97.5% confidence level.

Figure 7. Cont.

(c) VaR at 99% confidence level.

Figure 7. Comparison of effect of four methods on forecasting VaR at different confidence levels.

To examine the forecasting effect of the GIHS model, we examined the out-of-sample data from 2018. As shown in Table 7, under the three confidence levels, VaR_{min} forecasted by GIHS was smaller than other methods, and its VaR_{max} and HSAF's VaR_{max} were at a comparable level, which were greater than other methods. The overall forecasting level of GIHS (VaR_{avg}) was the lowest, i.e., closer to the actual oil returns than the other three methods. Therefore, the VaR forecasting effect of GIHS was better than other methods.

Table 7. Comparison of VaR forecasting effect.

Methods	HS	EWHS	HSAF	GIHS
At 95% confidence level				
$VaR_{max}(\%)$	20.245	19.832	27.059	26.964
$VaR_{min}(\%)$	15.873	15.381	13.518	12.541
$VaR_{avg}(\%)$	17.725	17.288	17.013	16.202
At 97.5% confidence level				
$VaR_{max}(\%)$ ¹	27.223	27.338	30.045	29.923
$VaR_{min}(\%)$ ²	24.240	23.801	16.405	15.606
$VaR_{avg}(\%)$ ³	26.758	26.741	19.608	18.616
At 99% confidence level				
$VaR_{max}(\%)$	31.096	31.068	32.248	32.591
$VaR_{min}(\%)$	30.626	30.165	18.896	18.430
$VaR_{avg}(\%)$	31.067	31.013	21.922	21.693

¹ VaR_{max} , ² VaR_{min} , and ³ VaR_{avg} represent the maximum, minimum, and average values of VaR, respectively.

To measure whether or not the risk forecasting model is reasonable when applied to a real oil market, we compared the GIHS model with the HS method, as well as its improved version, according to a back-testing method proposed by Kupiec. The back-testing results under the three confidence levels are shown in Table 8.

Table 8. VaR back-testing results.

Methods	HS	EWHS	HSAF	GIHS
At 95% confidence level				
N ¹	5	5	3	4
f ²	6.098%	6.098%	3.659%	4.878%
LR ³	0.195 *	0.195 *	0.341 *	0.003 *
At 97.5% confidence level				
N	0	0	1	2
f	0%	0%	1.220%	2.440%
LR	4.152 **	4.152 **	0.678 **	0.001 **
At 99% confidence level				
N	0	0	1	1
f	0%	0%	1.220%	1.220%
LR	1.648 ***	1.648 ***	0.037 ***	0.037 ***

¹ N is the excess number of VaR violations, ² f represents the empirical failure rate, ³ LR test is provided by Kupiec (1995). ***, ** and * denote significance at 1%, 5% and 10% levels, respectively.

In order to further qualitatively analyze the VaR forecasting effect of various forecasting models, and to investigate the forecasting ability of the giant information history simulation (GIHS), this paper modeled the oil returns of 2001.01–2011.12 as a training set, and forecast oil returns of 2012.01–2018.10. In this paper, the mean square error, mean absolute error, and mean absolute percentage error were introduced to analyze the forward-step risk forecasting effect of various methods. These results are shown in Table 9.

Table 9. The forward-step forecasting effect of VaR.

Methods	HS	EWHS	HSAF	GIHS
At 95% confidence level				
MSE ¹	167.519	157.694	145.225	128.389
MAE ²	12.242	11.861	11.184	10.453
MAPE ³	853.861	828.096	795.569	746.702
At 97.5% confidence level				
MSE	458.622	458.217	207.893	182.439
MAE	20.645	20.628	13.587	12.628
MAPE	1392.770	1392.524	945.442	885.739
At 99% confidence level				
MSE	654.792	651.974	275.869	268.716
MAE	24.954	24.900	15.839	15.627
MAPE	1635.158	1630.499	1079.521	1059.492

¹ MSE is mean square error, ² MAE represents the mean absolute error, ³ MAPE is mean absolute percentage error.

It can be seen from Table 9 that no matter which index was used, GIHS had the smallest fitting error value, so its forecasted value was closer to the real oil returns. From a forecasting perspective, GIHS clearly provides the best forecasting effect. Therefore, combining all the analysis in this chapter, the GIHS model is more advantageous in terms of fitting data and forecasting effect.

5. Conclusions

In the context of the rapid fluctuations in global oil prices, based on oil market massive news in a big data context, we used the two-layer NMF topic model in natural language processing to form a risk-identification algorithm for use in oil markets, and propose a novel giant information historical simulation (GIHS) method. Based on the empirical data of Brent crude oil returns from October 2001 to October 2018, the well-known VaR, which measures risk, was applied for risk qualification. Several conclusions and implications from the study are summarized as follows:

(1) Using the two-layer NMF model in natural language processing to model more than 200,000 news items, we finally identified various risk factors in the oil market, including not only fundamental factors (supply and demand) therein, but also non-fundamental factors such as environment, climate, market, economy, geopolitics, and oil companies, further illustrating that oil price volatility is the result of many factors.

(2) Considering the timeliness of risks, we defined the concept of oil risk pheromones, and quantified it for the first time. It can be seen that the oil risk pheromones fluctuate greatly over time. Therefore, it is very important for the country to formulate an oil price risk mechanism, especially for oil demand countries. Countries affected by the oil price mechanism need to re-examine the impact of energy policies, and they can achieve green transformation by establishing corresponding energy conservation and consumption reduction, finding alternative clean energy, and establishing cooperation with energy international organizations. In addition, in order to avoid high-risk shock to prices, the state should adjust its economic model in a timely manner, reduce excessive dependence on oil, and ultimately achieve sustainable and green transformation.

(3) Using risk analysis can help financial institutions effectively avoid the credit default risks caused by energy, such as oil exploitation risks. In terms of oil extraction, uncontrolled oil exploitation may bring certain credit risks, which affects the establishment of the national credit system through credit transmission mechanisms. In addition, it will bring enormous pressure on regulatory agencies. Through the risk analysis of oil, we can effectively control the amount of oil extracted, and, while ensuring the maximization of resource efficiency, we can also achieve sustainable development of resources and the ability of the state to withstand risks. Governments and regulatory organizations should encourage financial institutions to actively conduct risk analysis, and for financial institutions, they should raise awareness of risk analysis to improve their economic structure to green.

(4) Using oil risk pheromones to forecast VaR based on the HSAF method: compared with the HS method as well as its improved version, we found that the GIHS model proposed in this paper had significant LR values at 95%, 97.5%, and 99% confidence levels, according to Kupiec-type back-testing results. The maximum value forecast by the GIHS model was the largest, the minimum is the smallest, the average was the lowest, and the forecast values were closer to the actual returns. Therefore, the novel model can effectively improve the accuracy of risk measurement. In addition, the improvement of VaR forecasting accuracy further proves that there is indeed an interaction between oil price returns and online oil-related news.

(5) Investors should choose a confidence level appropriate to their specific situation when using the VaR forecasting method. The forecasting results show that this model significantly overestimated the oil price risks at the 99% confidence level. Therefore, when conducting actual operations, investors should consider their own risk preferences, current operating conditions, development strategies of companies, the volatility of financial markets, etc. For example, if the strategy of a company is more conservative, a higher confidence level should be chosen.

In conclusion, we combined a big data background and natural language processing methods to propose a GIHS model with which to measure global oil price risk. The model fully considers the interaction between massive online oil-related news and oil price returns, and uses oil risk pheromones to assist in forecasting VaR, which improved the accuracy of risks measurement. This may help to measure risks and for risk capitalists and financial institutions.

Author Contributions: Z.L.T., L.L.L. and W.Z.J. performed the research; Z.L.T., H.L.Y. and L.L.L. co-wrote the paper. Conceptualization, Z.L.T., L.L.L. and W.Z.J.; Data curation, H.L.Y.; Formal Analysis, Z.L.T., L.L.L. and H.L.Y.; Methodology, Z.L.T. and L.L.L.; Software, Z.L.T., L.L.L. and W.Z.J.; All authors read and approved the final manuscript.

Funding: This research was funded by the National Natural Science Foundation of China, grant number 71871020, 71874070, 71573258, 71521002.

Acknowledgments: We all thank the respectable anonymous reviewers and seminar participants of Department of Information and Computation Science, University of Science and Technology Beijing, for their helpful suggestions according to which we improved the content.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Toon, V.; Alban, K.; Bert, S.; Kimon, K.; Luis, R.L.S.; Krzysztof, W. Economic Exposure to Oil Price Shocks and the Fragility of Oil-Exporting Countries. *Energies* **2018**, *11*, 827. [\[CrossRef\]](#)
2. Zhang, Y.J.; Wang, J.L. Do high-frequency stock market data help forecast crude oil prices? Evidence from the MIDAS models. *Energy Econ.* **2019**, *78*, 192–201. [\[CrossRef\]](#)
3. Deng, S.; Sakurai, A. Crude Oil Spot Price Forecasting Based on Multiple Crude Oil Markets and Timeframes. *Energies* **2014**, *7*, 2761–2779. [\[CrossRef\]](#)
4. Aboura, S.; Chevallerier, J. Spikes and crashes in the oil market. *Res. Int. Bus. Finance* **2016**, *36*, 615–623. [\[CrossRef\]](#)
5. Zhang, Y.J.; Yao, T. Interpreting the movement of oil prices: Driven by fundamentals or bubbles? *Econ. Model.* **2016**, *55*, 226–240. [\[CrossRef\]](#)
6. Zhang, Y.J.; Wu, Y.B. The dynamic information spill-over effect of WTI crude oil prices on China's traditional energy sectors. *China Agric. Econ. Rev.* **2018**, *10*, 516–534. [\[CrossRef\]](#)
7. Zhang, Y.J.; Zhang, J.L. Volatility forecasting of crude oil market: A new hybrid method. *J. Forecast.* **2017**, *37*, 781–789. [\[CrossRef\]](#)
8. Duffie, D.; Pan, J. An overview of Value at Risk. *J. Deriv. Spring* **1997**, *4*, 7–49. [\[CrossRef\]](#)
9. Engle, R.F.; Manganelli, S. CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles. *J. Bus. Econ. Stat.* **2004**, *22*, 367–381. [\[CrossRef\]](#)
10. Berkowitz, J.; Christoffersen, P.; Pelletier, D. Evaluating Value-at-Risk Models with Desk-Level Data. *Manag. Sci.* **2011**, *57*, 2213–2227. [\[CrossRef\]](#)
11. Du, J.H.; Li, Z.M.; Wu, L.J. Optimal Stop-Loss Reinsurance Under the VaR and CTE Risk Measures: Variable Transformation Method. *Comput. Econ.* **2019**, *53*, 1133–1151. [\[CrossRef\]](#)
12. Sun, E.W.; Wang, Y.J.; Yu, M.T. Integrated Portfolio Risk Measure: Estimation and Asymptotics of Multivariate Geometric Quantiles. *Comput. Econ.* **2017**, *52*, 1–26. [\[CrossRef\]](#)
13. Hendricks, D. Evaluation of Value-at-Risk Models Using Historical Data. *Soc. Sci. Electron. Publ.* **1996**, *2*, 39–69. [\[CrossRef\]](#)
14. Suárez, A.; Carrillo, S. Computational Tools for the Analysis of Market Risk. *Comput. Econ.* **2003**, *21*, 153–172. [\[CrossRef\]](#)
15. Samuel Baixauli-Soler, J.; Alfaro-Cid, E.; Fernandez-Blanco, M.O. Mean-VaR Portfolio Selection Under Real Constraints. *Comput. Econ.* **2011**, *37*, 113–131. [\[CrossRef\]](#)
16. Samuel Baixauli, J.; Alvarez, S. Implied Severity Density Estimation: An Extended Semiparametric Method to Compute Credit Value at Risk. *Comput. Econ.* **2012**, *40*, 115–129. [\[CrossRef\]](#)
17. Nieto, M.R.; Ruiz, E. Frontiers in VaR forecasting and backtesting. *Int. J. Forecast.* **2016**, *32*, 475–501. [\[CrossRef\]](#)
18. Zhang, J.L.; Zhang, Y.J.; Zhang, L. A novel hybrid method for crude oil price forecasting. *Energy Econ.* **2015**, *49*, 649–659. [\[CrossRef\]](#)
19. Mi, Z.F.; Wei, Y.M.; Tang, B.J.; Cong, R.G.; Yu, H.; Cao, H.; Guan, D. Risk assessment of oil price from static and dynamic modelling approaches. *Appl. Econ.* **2016**, *49*, 1–30. [\[CrossRef\]](#)
20. Bunn, D.; Chevallerier, J.; Pen, Y.L.; Sevi, B. Fundamental and Financial Influences on the Co-movement of Oil and Gas Prices. *Energy J.* **2017**, *38*, 201–228. [\[CrossRef\]](#)
21. Yao, T.; Zhang, Y.J.; Ma, C.Q. How does investor attention affect international crude oil prices? *Appl. Energy* **2017**, *205*, 336–344. [\[CrossRef\]](#)
22. Zhang, X.B.; Qin, P.; Chen, X.L. Strategic oil stockpiling for energy security: The case of China and India. *Energy Econ.* **2017**, *61*, 253–260. [\[CrossRef\]](#)
23. Yao, T.; Zhang, Y.J. Forecasting Crude Oil Prices with the Google Index. *Energy Procedia* **2017**, *105*, 3772–3776. [\[CrossRef\]](#)
24. Zhang, Y.J.; Zhang, L. Interpreting the crude oil price movements: Evidence from the Markov regime switching model. *Appl. Energy* **2015**, *143*, 96–109. [\[CrossRef\]](#)
25. Narayan, P.K.; Ranjeeni, K.; Bannigidadmath, D. New Evidence of Psychological Barrier from the Oil Market. *J. Behav. Finance* **2017**, *18*, 1–13. [\[CrossRef\]](#)

26. Li, Y.S.; Li, A.H.; Liu, Z.D. Two Ways of Calculating VaR in Risk Management-An Empirical Study Based on CSI 300 Index. *Procedia Comput. Sci.* **2018**, *139*, 432–439. [\[CrossRef\]](#)
27. Richardson, M.P.; Boudoukh, J.; Whitelaw, R. The Best of Both Worlds: A Hybrid Approach to Calculating Value at Risk. *Soc. Sci. Electron. Publ.* **1998**, *11*, 410–414. [\[CrossRef\]](#)
28. David Cabedo, J.; Moya, I. Estimating oil price 'Value at Risk' using the historical simulation approach. *Energy Econ.* **2003**, *25*, 239–253. [\[CrossRef\]](#)
29. Sadeghi, M.; Shavvalpour, S. Energy risk management and value at risk modeling. *Energy Policy* **2006**, *34*, 3367–3373. [\[CrossRef\]](#)
30. Giannopoulos, K.; Tunaru, R. Coherent risk measures under filtered historical simulation. *J. Bank. Finance* **2005**, *29*, 979–996. [\[CrossRef\]](#)
31. Jamshidian, F.; Zhu, Y. Scenario Simulation: Theory and methodology. *Finance Stoch.* **1996**, *1*, 43–67. [\[CrossRef\]](#)
32. Skiadopoulos, G.; Lambadiaris, G.; Papadopolou, L.; Zoulis, Y. VaR: History or Simulation? *Soc. Sci. Electron. Publ.* **2003**, *3*, 123–126. Available online: https://www.researchgate.net/publication/228183499_VaR_History_or_Simulation (accessed on 2 March 2019).
33. Panigirtzoglou, N.; Skiadopoulos, G. A new approach to modeling the dynamics of implied distributions: Theory and evidence from the S&P 500 options. *J. Bank Finance* **2004**, *28*, 1499–1520. [\[CrossRef\]](#)
34. Tezuka, S.; Murata, H.; Tanaka, S.; Yumae, S. Monte Carlo grid for financial risk management. *Future Gener. Comput. Syst.* **2005**, *21*, 811–821. [\[CrossRef\]](#)
35. Dionne, G.; Duchesne, P.; Pacurar, M. Intraday Value at Risk (IVaR) using tick-by-tick data with application to the Toronto Stock Exchange. *J. Empir. Finance* **2009**, *16*, 777–792. [\[CrossRef\]](#)
36. Tzeng, Y.Y.; Beaumont, P.M.; Ökten, G. Time Series Simulation with Randomized Quasi-Monte Carlo Methods: An Application to Value at Risk and Expected Shortfall. *Comput. Econ.* **2018**, *52*, 1–23. [\[CrossRef\]](#)
37. Mittnik, S.; Paolella, M.S. Conditional density and value-at-risk prediction of Asian currency exchange rates. *J. Forecast.* **2000**, *19*, 313–333. [\[CrossRef\]](#)
38. Angelidis, T.; Benos, A.; Degiannakis, S. A robust VaR model under different time periods and weighting schemes. *Rev. Quant. Finance Account.* **2006**, *28*, 187–201. [\[CrossRef\]](#)
39. Chen, C.W.S.; Gerlach, R.H.; Lin, E.M.H.; Lee, C.W. Bayesian Forecasting for Financial Risk Management, Pre and Post the Global Financial Crisis. *J. Forecast.* **2012**, *31*, 661–687. [\[CrossRef\]](#)
40. Krause, J.; Paolella, M.S. A Fast, Accurate Method for Value-at-Risk and Expected Shortfall. *Econometrics* **2014**, *2*, 98–122. [\[CrossRef\]](#)
41. Youssef, M.; Belkacem, L.; Mokni, K. Value-at-Risk estimation of energy commodities: A long-memory GARCH-EVT approach. *Energy Econ.* **2015**, *51*, 99–110. [\[CrossRef\]](#)
42. Fries, C.P.; Nigbur, T.; Seeger, N. Displaced relative changes in historical simulation: Application to risk measures of interest rates with phases of negative rates. *J. Empir. Finance* **2017**, *42*, 175–198. [\[CrossRef\]](#)
43. Hull, J.; White, A. Incorporating volatility updating into the historical simulation method for Value-at-Risk. *J. Risk* **1998**, *1*, 5–19. [\[CrossRef\]](#)
44. Adesi, G.B.; Giannopoulos, K.; Vosper, L. VaR without Correlations for Portfolios of Derivative Securities. *J. Futur. Mark.* **1999**, *19*, 583–602. [\[CrossRef\]](#)
45. Karmakar, M.; Paul, S. Intraday portfolio risk management using VaR and CVaR: A CGARCH-EVT-Copula approach. *Int. J. Forecast.* **2019**, *35*, 699–709. [\[CrossRef\]](#)
46. Ardia, D.; Bluteau, K.; Boudt, K.; Catania, L. Forecasting risk with Markov-switching GARCH models: A large-scale performance study. *Int. J. Forecast.* **2018**, *34*, 733–747. [\[CrossRef\]](#)
47. Hung, J.C.; Lee, M.C.; Liu, H.C. Estimation of value-at-risk for energy commodities via fat-tailed GARCH models. *Energy Econ.* **2008**, *30*, 1173–1191. [\[CrossRef\]](#)
48. Gencay, R.; Selcuk, F. Extreme value theory and Value-at-Risk: Relative performance in emerging markets. *Int. J. Forecast.* **2004**, *20*, 287–303. [\[CrossRef\]](#)
49. Chan, K.F.; Gray, P. Using extreme value theory to measure value-at-risk for daily electricity spot prices. *Int. J. Forecast.* **2006**, *22*, 283–300. [\[CrossRef\]](#)
50. Zhao, L.T.; Li, T.; Zhang, Y.J.; Wei, Y.M. Measuring the price risk of energy portfolio with Copula-VaR model. *Syst. Eng. Theory Pract.* **2015**, *35*, 771–779. [\[CrossRef\]](#)
51. Fuertes, A.M.; Olmo, J. Optimally harnessing inter-day and intra-day information for daily value-at-risk prediction. *Int. J. Forecast.* **2013**, *29*, 28–42. [\[CrossRef\]](#)

52. Meng, X.C.; Taylor, J.W. An approximate long-memory range-based approach for value at risk estimation. *Int. J. Forecast.* **2018**, *34*, 377–388. [[CrossRef](#)]
53. Prusa, J.D.; Sagul, R.; Khoshgoftaar, T.M.; Sterling, M. Extracting Knowledge from Technical Reports for the Valuation of West Texas Intermediate Crude Oil Futures. *Int. Conf. Inf. Reuse. Integr.* **2017**, *27*, 43–48. [[CrossRef](#)]
54. Li, J.; Xu, Z.J.; Xu, H.J.; Tang, L.; Yu, L. Forecasting Oil Price Trends with Sentiment of Online News Articles. *Procedia Comput. Sci.* **2016**, *91*, 1081–1087. [[CrossRef](#)]
55. Chuaykoblap, S.; Chutima, P.; Chandrachai, A.; Nupairoj, N. Expert-based text mining with Delphi method for crude oil price prediction. *Int. J. Ind. Syst. Eng.* **2017**, *25*, 545–546. [[CrossRef](#)]
56. Oussalah, M.; Zaidi, A. Forecasting Weekly Crude Oil Using Twitter Sentiment of U.S. Foreign Policy and Oil Companies Data. *Int. Conf. Inf. Reuse. Integr.* **2018**, *1*, 201–208. [[CrossRef](#)]
57. Zhao, L.T.; Zeng, G.R. Analysis of Timeliness of Oil Price News Information Based on SVM. *Energy Procedia* **2019**, *158*, 4123–4128. [[CrossRef](#)]
58. Kaiser, M.J.; Yu, Y. The impact of Hurricanes Gustav and Ike on offshore oil and gas production in the Gulf of Mexico. *Appl. Energy* **2010**, *87*, 284–297. [[CrossRef](#)]
59. McLaren, N.; Shanbhogue, R. Using Internet Search Data as Economic Indicators. *Bank Engl. Q. Bull.* **2011**, *51*, 134–140. [[CrossRef](#)]
60. Vosen, S.; Schmidt, T. A monthly consumption indicator for Germany based on Internet search query data. *Appl. Econ. Lett.* **2012**, *19*, 683–687. [[CrossRef](#)]
61. Ji, Q.; Guo, J.F. Oil price volatility and oil-related events: An Internet concern study perspective. *Appl. Energy* **2015**, *137*, 256–264. [[CrossRef](#)]
62. Wang, J.; Athanasopoulos, G.; Hyndman, R.J.; Wang, S.Y. Crude oil price forecasting based on internet concern using an extreme learning machine. *Int. J. Forecast.* **2018**, *34*, 665–677. [[CrossRef](#)]
63. Greene, D.; Cross, J.P. Exploring the Political Agenda of the European Parliament Using a Dynamic Topic Modeling Approach. *Polit. Anal.* **2016**, *25*, 77–94. [[CrossRef](#)]
64. Kupiec, P.H. Techniques for Verifying the Accuracy of Risk Management Models. *Soc. Sci. Electron. Publ.* **1995**, *3*, 73–84. [[CrossRef](#)]

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).