Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
The role of coronavirus news in the volatility forecasting of crude oil futures markets: Evidence from China

Zibo Niu a,b, Yuanyuan Liu b, Wang Gao c,d, Hongwei Zhang b,e,*

a School of Business, Central South University, Changsha, 410083, China
b School of Mathematics and Statistics, Central South University, Changsha, 410083, China
c School of Statistics, Renmin University of China, Beijing, 100872, China
d School of Social Sciences, Tsinghua University, Beijing, 100084, China
e Institute of Metal Resources Strategy, Central South University, Changsha, 410083, China

ABSTRACT

Based on the high-frequency heterogeneous autoregressive (HAR) model, this paper investigates whether coronavirus news (in China and globally) contains incremental information to predict the volatility of China’s crude oil, and studies which types of coronavirus news can better forecast China’s crude oil volatility. Considering the information overlap among various coronavirus news items and making full use of the information in various coronavirus news items, this paper uses two prevailing shrinkage methods, lasso and elastic nets, to select coronavirus news items and then uses the HAR model to predict China’s crude oil volatility. The results show that (i) coronavirus news can be utilized to significantly predict China’s crude oil volatility for both in-sample and out-of-sample analyses; (ii) the Panic Index (PI) and the Country Sentiment Index (CSI) have a greater impact on China’s crude oil volatility; and (iii) global coronavirus news provides more incremental information than China’s coronavirus news for predicting the volatility of China’s crude oil. Additionally, China’s crude oil volatility forecast; and (iii) global coronavirus news provides more incremental information than China’s coronavirus news for predicting the volatility of China’s crude oil market, which indicates that global coronavirus news is also a key factor to consider when predicting the market volatility of China’s crude oil.

1. Introduction

Over the past few months, the coronavirus pandemic has severely affected the dynamics of global financial markets (Adekoya and Oliyide, 2021; Baker et al., 2020; Guerrieri et al., 2020; Zhang et al., 2020). The notable impact is the large decline and subsequent rebound of the stock and oil markets. Investors appear to be paying more attention to oil markets as oil demand may signal whether and how the global economy will recover from the devastation caused by the coronavirus pandemic. Chinese crude oil futures were formally launched and traded in the Shanghai International Energy Exchange (INE) on March 26, 2018. For more than two years, the INE has been the world’s third largest crude oil futures contract, followed by Brent and West Texas Intermediate (WTI) (Ji and Zhang, 2019). The influence of China’s oil market on global markets has been increasing as China becomes the world’s largest crude oil importer. However, there are few studies on the volatility of China’s crude oil market.

In view of the impact of the coronavirus pandemic on the crude oil market, this paper mainly studies the impact of coronavirus news on the volatility of China’s crude oil futures market and whether it can be used to predict the volatility of China’s crude oil. In this article, we study the volatility of China’s crude oil futures market from the perspective of behavior – public news sentiment – which has attracted much interest. A large number of studies show that news is an important source of information used to predict stock returns and volatility (Garcia, 2013; Gurun and Butler, 2012; Narayan, 2019; Tetlock et al., 2008). For example, Narayan (2019) studied the role of financial news related to oil in driving stock returns. The results show that oil price news reported by leading global newspapers and news agencies predicts stock returns. In addition, in recent years, scholars have been increasingly concerned...
about the effect of public news sentiment on stock returns. A growing number of theoretical and empirical studies have focused on exploring whether stock market price changes driven by political or economic news (Broadstock and Zhang, 2019; Shi and Ho, 2021; Smales, 2014). Furthermore, information provided through social media channels has a substantial influence on financial market dynamics, especially during periods of political or economic uncertainty (Kindleberger and Aliber, 2011; Yang et al., 2020). For example, Kindleberger and Aliber (2011) reviewed global financial crises in the past few centuries. They show that panic sentiment is an important driving force for asset prices to plummet. In addition, Yang et al. (2020) find that investor panic has a significant influence on the risk of stock price crash, which indicates that the higher panic sentiment greatly increases the risk.

Some studies show that the effect of news, represented by online news, oil price news, and news stories, on oil prices is immediately apparent in the short term (Brandt and Gao, 2019; Grego, 2020; Gupta and Banerjee, 2019; Kilian and Vega, 2011; Li et al., 2019; Loughran et al., 2019). For example, Mashyuk-Escobedo et al. (2017) find that news sentiments related to energy comodities with the futures and spot earnings in energy commodies, such as heating oil, gasoline and crude oil. In addition, the results of this research show that aggregate sentiment indexes to a large extent affect the movements of crude oil prices. Li et al. (2019) propose a new crude oil price prediction method based on text, which uses deep learning techniques, topic extraction and sentiment analysis. The results show that news text adds meaningful information to crude oil price forecasts and that combining text information in news with financial market information can significantly improve accuracy. Using sentiment scores on global news on both macroeconomic and geopolitical issues, Brandt and Gao (2019) discover that news about macro fundamentals influences oil price heavily in the short term and is able to predict oil returns in the long run. Additionally, geopolitical news, although without showing predictability, exerts an even greater immediate effect.

Although the existing literature on news confirms its significant impact on oil prices, research focusing on the role of news in explaining crude oil market volatility is, surprisingly, very limited (Rosa, 2014; Schmidbauer and Rosch, 2012; Smales, 2017). For example, Rosa (2014) points out that crude oil futures prices were more volatile on days when the US macroeconomic news was released. Smales (2017) finds that volatility of commodity markets is heavily affected by Chinese and US macroeconomic news, which reveals information about the aggregate demand for commodities. However, the author of this article is not aware of any research on crude oil volatility based on coronavirus news.

This paper contributes to the literature in several ways. First, most previous articles focus on the impact of coronavirus news on returns (Gharib et al., 2021; Salisu et al., 2020a, 2020b), but there are few studies on the impact of this news on market volatility. There is sufficient evidence to show that news has a significant impact on modeling and predicting the volatility of the oil market (Andersen and Bollerslev, 1998b; Bolltm, 2003; Chua and Tsiaplias, 2019; Jones et al., 1998). Moreover, the existing literature shows that coronavirus news has a significant impact on market volatility (Ambros et al., 2020; Baek et al., 2020; Bakas and Triantafyllou, 2020; Cepoi, 2020). Under the influence of the pandemic, the stock and crude oil prices have experienced severe short-term fluctuations (Narayan, 2020; Zhang and Hamori, 2021). Obviously, the situation and impact of the pandemic is transmitted to the crude oil market through news information about the pandemic, which further reflects market volatility. Therefore, it is necessary to use coronavirus news to predict China’s crude oil volatility. In addition, previous articles use daily data or lower-frequency data to study the impact of news on crude oil volatility, but high-frequency volatility can more fully reflect the volatility information of one day; that is, there is a large amount of information that is very useful for market traders and investors seeking to make decisions faster. Moreover, the timeliness and short-term effectiveness of coronavirus news should be considered. To that end, this paper uses high-frequency crude oil futures data from China to fully extract coronavirus news information to predict China’s crude oil volatility, which is a characteristic that cannot be accurately predicted by low-frequency data. Thus far, the literature studying the effects of coronavirus news on oil market volatility is silent about the use of high-frequency data. Therefore, examining the role of coronavirus news in oil market volatility based on high-frequency data is urgent and highly beneficial.

Second, we develop four new HAR-type models that can obtain more predictive values for the volatility of crude oil. To test the effect of coronavirus news with incremental information on volatility predictions during the COVID-19 pandemic, we introduce six coronavirus news indexes into the HAR model proposed by Corsi (2009) to predict the volatility of China’s crude oil market. Additionally, to choose which type of coronavirus news has more predictive effects on China’s crude oil, this paper uses shrinkage methods to select coronavirus news variables during the entire sampling period and then uses the HAR model to predict China’s crude oil volatility. Motivated by previous studies, our study determines which popular coronavirus news predictor (the Panic Index (PI), the Media Hype Index (HY), the Fake News Index (FNI), the Country Sentiment Index (CSI), the Contagion Index (CTI), and the Media Coverage Index (MCI)) is more powerful for forecasting China’s crude oil volatility during this period of extreme fluctuation. This paper attempts to provide investors, researchers and policy makers with new and valuable information that can be used to forecast the volatility of crude oil markets by focusing on coronavirus news during the coronavirus pandemic.

Third, on the one hand, most previous works focuses on the volatility of international oil markets, such as the Brent and WTI, while the study of China’s oil market is very sparse. However, due to the improvement of China’s international status, the role of China’s crude oil market has become increasingly important, and its crude oil market cannot be ignored. On the other hand, a few existing studies provide evidence that China’s crude oil prices are sensitive to coronavirus news. It is not yet clear, however, how broad measures of coronavirus news affect China’s crude oil volatility. Therefore, this article studies the volatility of China’s crude oil market. Then, we perform an empirical study of the volatility of China’s crude oil considering six coronavirus news indexes. To our knowledge, this is the first paper to analyze the effects of coronavirus news on China’s crude oil volatility. By considering different coronavirus news items, this research provides rich insights into the determinants of China’s crude oil volatility.

The major empirical results arising from our empirical analyses are as follows. First, coronavirus news can be utilized to significantly predict China’s crude oil volatility under the high-frequency framework in both in-sample and out-of-sample analyses. Second, regarding China’s coronavirus news, the PI, FNI, and CSI have a greater impact on China’s crude oil volatility during the coronavirus pandemic. Third, regarding global coronavirus news, the PI and CSI have a significant influence on the volatility forecasting of China’s crude oil market during the coronavirus pandemic. Fourth, we find that global coronavirus news provides more incremental information than China’s coronavirus news that can be used to predict the volatility of China’s crude oil. Fifth, this article uses the realized kernel (RK) function and recursive window method to forecast China’s crude oil volatility and then tests the robustness of the results through out-of-sample analysis.

The paper is structured as follows. The related literature review is introduced in Section 2. Section 3 introduces the construction of the volatility measures, the prediction models and the model confidence set (MCS) for the prediction evaluations. Section 4 describes oil prices and coronavirus news data. Section 5 summarizes the results of the study. Section 6 summarizes the results of robustness tests. Finally, Section 7 concludes.

2. Related literature review

The earliest research on oil price volatility prediction is the study by Sadorsky (2006). It shows that out-of-sample predictions are evaluated
using prediction accuracy tests and finds that the GARCH model is suitable for unleaded gasoline and oil volatility and that the threshold GARCH (TGARCH) model is very applicable to heating oil volatility. Since then, there has been an endless stream of literature on oil volatility prediction research (Arouni et al., 2012; Kang and Yoon, 2013; Narayan and Narayan, 2007; Sevi, 2014). However, these uses study only the data of the crude oil market itself to study the volatility and do not study the influence of external factors on the volatility of the oil price market. Social events or unexpected political events that play considerable roles in oil market volatility can be reflected in news text. Therefore, it is scientifically reasonable to consider the news text.

In the past few decades, researchers have been paying attention to the impact of news on financial and commodity markets volatility (Engle and Ng, 1993; Jiang et al., 2012; Marshall et al., 2012; Rangel, 2011; Shi and Ho, 2021), especially in the crude oil market (Schmidbauer and Rösch, 2012). Recently, a large amount of literature has shown that news can influence market volatility based on daily data and GARCH specifications (Demirer and Kutan, 2010; Ewing and Malik, 2017; Mensi et al., 2014; Schmidbauer and Rösch, 2012). For example, Mensi et al. (2014) use GARCH-type models to study the impact of information content of OPEC on the volatility of WTI and Brent markets. The results show that news announcements have significant but differential impacts on the persistence of volatility. Ewing and Malik (2017) use an asymmetric GARCH model to study oil price volatility and find that both good and bad news have an impact on oil price volatility.

Although the main empirical findings indicate that HAR-type models provide better results than the stochastic volatility (SV) and GARCH-type models in out-of-sample predictions (Andersen et al., 2003, 2004; Sevi, 2014; Zhu et al., 2017b), few scholars have studied the impact of news on market volatility prediction based on HAR-type models. Chen and Ghysels (2011) find that both very good news and bad news increase volatility, with the latter having a more severe impact. Liang et al. (2020) find that news sentiment in social and Internet media can improve the accuracy of prediction based on the HAR-type models. These two studies based on the HAR model both study the impact of news on the volatility of the stock market, and there is no literature on the impact of news on the volatility of the crude oil market. To test the effect of coronavirus news with incremental information on volatility predictions during the COVID-19 pandemic, we introduce six coronavirus news indexes into the HAR model proposed by Corsi (2009) to predict the volatility of China’s crude oil market.

In summary, in terms of the existing literature, several apparent features, which motivated us to conduct this study, must be highlighted here. First, previous articles employed daily or lower-frequency data to study the impact of news on crude oil volatility, but high-frequency volatility can more fully reflect the volatility information of one day. Additionally, high-frequency data involve a considerable amount of information, which can enable market traders, and investors to make decisions faster. Second, using data on only China’s high-frequency crude oil futures may not provide accurate information for forecasting China’s crude oil volatility. Social events or unexpected political events that play considerable roles in China’s crude oil market volatility can be reflected in news text. Therefore, it is a scientific method to consider the news text. Thus, this article makes a crucial contribution to the literature by investigating the response of China’s crude oil market fluctuations to news of the coronavirus pandemic. The questions considered in this article are as follows: (1) Does coronavirus news influence the volatility prediction of China’s crude oil market? (2) What kind of coronavirus news has a large impact on the volatility forecasting of China’s crude oil futures market? (3) Of two sources of coronavirus news, Chinese and global coronavirus news, which kind of coronavirus news has a greater influence on the volatility forecasting of China’s crude oil market?

3. Methodology

3.1. Realized volatility and jump detection

According to Andersen and Bollerslev (1998a), realized variance is defined as the sum of the squares of intraday returns. In our article, a measure for realized volatility (RV) is used, and it can be written as:

$$RV_t = \sum_{j=1}^{M} r_{j,t}^2$$

where $r_{j,t}$ represents the $j^{th}$ intraday returns ($j = 1, 2, \ldots, M$) at day $t$ and $1/M$ represents the sampling frequency.

Barndorff-Nielsen and Shephard (2000) pointed out that when M tends to extend to infinity, RV can be written as:

$$RV_t^{M→∞} = \int_0^T \sigma_s^2 ds + \sum_{0<s<r} k_s^2,$$

as the sampling frequency, the realized bipower variance (RBV) converges uniformly in probability:

$$RBV_t^{M→∞} = \mu_1^2 \left( \frac{M}{M-2} \right) \sum_{j=1}^{M} r_{j,t-1} |r_{j,t}|^{2M-2} \int_0^T \sigma_s^2 ds,$$

where $\mu_1 = \sqrt{2/\pi}$.

The daily discontinuous jump variation $J_t$ can be written as:

$$RV_t - RBV_t^{M→∞} = J_t$$

To select statistically significant jumps, Huang and Tauchen (2005) construct the statistic as follows:

$$Z_t = \frac{(RV_t - RBV_t^{M→∞})/RV_t}{\sqrt{(\frac{1}{2} + \frac{1}{5}) \mu_1 \max \left( 1, \frac{RV_t^{M→∞}}{RBV_t^{M→∞}} \right)}} \sim N(0, 1),$$

where $RQV_t$ represents an estimator of fourth power variation that can be written as follows:

$$RQV_t = M\mu_2^2 \left( \frac{M}{M-4} \right) \sum_{j=1}^{M} r_{j,t-4} |r_{j,t}|^{4M-4}.$$
To improve forecasting accuracy, we adopt the jump test statistic $Z_{\text{med}}$ proposed by Andersen et al. (2012). The expression of test statistic $Z_{\text{med}}$ is defined by

$$Z_{\text{med}} = \frac{(RV_t - \text{Med}RV_t) / RV_t}{\sqrt{\left(\frac{1}{2} \times (\pi - 5)\right) \times \max(1, \text{Med}RV_t)}} \sim N(0, 1).$$

(9)

Hence, in our paper, the daily jump variation can be expressed by

$$J_t = I(Z_{\text{med}} > \Phi_{\alpha}) \cdot (RV_t - \text{Med}RV_t),$$

(10)

where $I(\cdot)$ is an indicator function, $\Phi_{\alpha}$ indicates the corresponding trigger value at the significance level of $\alpha$ is standard normal distribution.

3.2. Volatility models

We employ the HAR model defined by Corsi (2009); this model has been useful for describing RV dynamics. The HAR-type models are defined as:

3.2.1. HAR-RV and HR-RV-J models

$$RV_{t+1} = \alpha_0 + \alpha_1 RV_t + \alpha_2 RV_{t-5} + \alpha_3 RV_{t-22} + \epsilon_{t+1},$$

(11)

$$RV_{t+1} = \alpha_0 + \alpha_1 RV_t + \alpha_2 RV_{t-5} + \alpha_3 RV_{t-22} + \alpha_4 J_t + \epsilon_{t+1},$$

(12)

where $RV_{t+1}$ represents the 1-day future RV; $\alpha$ is the daily RV, as defined in Eq. (1), where $\lambda$ denotes the weekly and monthly RV; $RV_{t+1} = \alpha_0 + \alpha_1 RV_t + \alpha_2 RV_{t-5} + \alpha_3 RV_{t-22} + \alpha_4 J_t + \beta_1 \text{PI} + \beta_3 \text{FNI} + \beta_5 \text{CSI} + \epsilon_{t+1}$, is the J part defined in Eq. (10).

3.2.2. HAR-RV-CAN and HR-RV-J-CAN models

$$RV_{t+1} = \alpha_0 + \alpha_1 RV_t + \alpha_2 RV_{t-5} + \alpha_3 RV_{t-22} + \beta_1 \text{PI} + \beta_3 \text{HY} + \beta_5 \text{FNI},$$

$$+ \beta_7 \text{CSI} + \beta_9 \text{CTI} + \beta_6 \text{MCI} + \epsilon_{t+1},$$

(13)

$$RV_{t+1} = \alpha_0 + \alpha_1 RV_t + \alpha_2 RV_{t-5} + \alpha_3 RV_{t-22} + \alpha_4 J_t + \beta_1 \text{PI} + \beta_3 \text{HY} + \beta_5 \text{FNI},$$

$$+ \beta_7 \text{CSI} + \beta_9 \text{CTI} + \beta_6 \text{MCI} + \epsilon_{t+1},$$

(14)

where the six coronavirus news indicators are the China’s PI, the China’s Media Hype Index (HY), China’s FNI, China’s CSI, China’s CTI, and China’s MCI.

3.2.3. HAR-RV-CCN and HR-RV-J-CCN models

To choose which type of Chinese coronavirus news has a stronger predictive effect on the crude oil market and eliminate possible overlapping information, we use shrinkage methods to select coronavirus news indexes for the entire sample and then use the HAR model to predict crude oil volatility. One concern has to be addressed when performing this exercise. It is necessary to prevent volatility components from being filtered out by the lasso and elastic net methods. To address this concern, we control for volatility components in our regressions. This article uses the cross-validation (CV) method to choose $\lambda$ for the lasso method. For the elastic nets, the shrinkage parameters $\lambda$ and $\alpha$ and the tuning parameters $\lambda$ and $\alpha$ are selected via CV.

$$RV_{t+1} = \alpha_0 + \alpha_1 RV_t + \alpha_2 RV_{t-5} + \alpha_3 RV_{t-22} + \beta_1 \text{PI} + \beta_3 \text{FNI} + \beta_5 \text{CSI},$$

$$+ \epsilon_{t+1},$$

(15)

$$RV_{t+1} = \alpha_0 + \alpha_1 RV_t + \alpha_2 RV_{t-5} + \alpha_3 RV_{t-22} + \alpha_4 J_t + \beta_1 \text{PI} + \beta_3 \text{FNI},$$

$$+ \beta_5 \text{CSI} + \epsilon_{t+1},$$

(16)

where PI is China’s panic index, FNI is China’s fake news index, and CSI is China’s country sentiment index.

3.2.4. HAR-RV-WAN and HR-RV-J-WAN models

$$RV_{t+1} = \alpha_0 + \alpha_1 RV_t + \alpha_2 RV_{t-5} + \alpha_3 RV_{t-22} + \beta_1 \text{PI} + \beta_5 \text{HY} + \beta_3 \text{FNI},$$

$$+ \beta_7 \text{CSI} + \beta_9 \text{CTI} + \beta_6 \text{MCI} + \epsilon_{t+1},$$

(17)

$$RV_{t+1} = \alpha_0 + \alpha_1 RV_t + \alpha_2 RV_{t-5} + \alpha_3 RV_{t-22} + \alpha_4 J_t + \beta_1 \text{PI} + \beta_5 \text{HY} + \beta_3 \text{FNI},$$

$$+ \beta_7 \text{CSI} + \beta_9 \text{CTI} + \beta_6 \text{MCI} + \epsilon_{t+1},$$

(18)

where the six coronavirus news indicators are the global PI, the global Media Hype Index (HY), the global FNI, the global CSI, the global CTI, and the global MCI.

3.2.5. HAR-RV-WCN and HR-RV-J-WCN models

Furthermore, to determine which type of global coronavirus news has a stronger predictive effect on the crude oil market and eliminate the possibility of overlapping information, this paper uses shrinkage methods to select global coronavirus news indexes for the entire sample and then uses the HAR model to predict crude oil volatility. This article uses the CV method to choose $\lambda$. For the elastic nets, the shrinkage parameters $\lambda$ and $\alpha$ and the tuning parameters $\lambda$ and $\alpha$ are selected via CV.

$$RV_{t+1} = \alpha_0 + \alpha_1 RV_t + \alpha_2 RV_{t-5} + \alpha_3 RV_{t-22} + \alpha_4 J_t + \beta_1 \text{PI} + \beta_3 \text{CSI} + \epsilon_{t+1},$$

(19)

$$RV_{t+1} = \alpha_0 + \alpha_1 RV_t + \alpha_2 RV_{t-5} + \alpha_3 RV_{t-22} + \alpha_4 J_t + \beta_1 \text{PI} + \beta_3 \text{CSI} + \epsilon_{t+1},$$

(20)

where PI is global panic index and CSI is global country sentiment index.

3.3. Forecasting and evaluation

Policymakers and investors may be more concerned about prediction performance than the results for the sample because they may want to use the model to predict volatility in the future to guide further decisions. Therefore, below, we determine whether adding coronavirus news indexes to volatility models can optimize their ability to predict out-of-sample data. This paper uses two popular loss functions (heteroskedasticity-adjusted mean square error (HMSE) and heteroskedasticity-adjusted mean absolute error (HMAE)) to quantitatively compare the out-of-sample performance of the prediction model used in this article. Here, we introduce the two loss functions that can quantitatively measure the accuracy of forecasting:

$$\text{HMSE} = N^{-1} \sum_{t=1}^{N} \left(1 - \frac{RV_t}{\sigma_t^2}\right)^2,$$

(21)

$$\text{HMAE} = N^{-1} \sum_{t=1}^{N} \left|1 - \frac{RV_t}{\sigma_t}\right|,$$

(22)

where $RV_t$ represents the volatility prediction obtained from various HAR-type models; $\sigma_t$ is the real value of the volatility, for which we apply RV to replace it; and $N$ is the number of prediction days.

This paper uses the MCS proposed by Hansen et al. (2011) as a predictive evaluation method to compare the prediction performance of our proposed volatility models. The MCS test has been broadly used to assess the predictive performance of volatility models due to the advantages of not needing a prespecified benchmark model (Bauwens and Otranto, 2016; Koopman et al., 2016). According to the above literature, the semiquadratic (SemIQ) and range-based (Range) statistics are selected as MCS statistics, and their p-values are obtained using a
bootstrap program. In our research, every model has a p-value in an initial set of competition models, and the larger the p-value, the stronger the model’s predictive ability. In this paper, we choose this model as a good prediction performance model based on the criterion that the p-value of the volatility model is greater than 0.10. The Range and the model bootstrap program. In our research, every model has a p-value in an

good prediction performance model based on the criterion that the relative sample loss between the model and excessive jumps and matching the date with coronavirus news, we obtain 729 daily observations. To assess the prediction performance of various volatility models, the complete samples are divided into model estimation and out-of-sample prediction periods. The sample data in this paper are divided into two groups: 1) a group in which the estimation period is from March 27, 2018, to December 31, 2019, and 2) a group in which the out-of-sample forecast evaluation spans the period from January 1, 2020, to October 22, 2020.

4.1. China oil prices

On March 26, 2018, Chinese oil futures were officially listed on the INE. Therefore, this article uses data ranging from March 26, 2018, to October 22, 2020. Previous studies (Ma et al., 2017; Patton and Sheppard, 2015; Sévi, 2014; Yang et al., 2019; Zhang et al., 2018; Zhu et al., 2017a) use the 5-min sampling frequency to explore price volatility. Thus, we follow the previous research to select 5-min high-frequency data. The complete sampling period for our data is from March 27, 2018 to October 22, 2020. After removing the shortened trading days and excessive jumps and matching the date with coronavirus news, we obtain 729 daily observations. To assess the prediction performance of various volatility models, the complete samples are divided into model estimation and out-of-sample prediction periods. The sample data in this paper are divided into two groups: 1) a group in which the estimation period is from March 27, 2018, to December 31, 2019, and 2) a group in which the out-of-sample forecast evaluation spans the period from January 1, 2020, to October 22, 2020.

4.2. Coronavirus news

Information on variables related to coronavirus news is obtained from the RavenPack analysis tool. This platform offers information from real-time media analysis that analyzes announcements of important issues related to the COVID-19 pandemic, such as fake news, media hype, and panic. This tool includes sources such as the StockTwits, Wallstreet Journal, and Dow Jones Newswire, and others (Blitz et al., 2020). For instance, using this news monitor database, Smales (2014) and Shi and Ho (2021) explored the relationship between implied volatility and news sentiment. Additionally, Cepoi (2020) investigated the response of the stock markets of the six countries most affected by the pandemic to coronavirus news. The results indicate that the stock market is dependent on information related to the coronavirus pandemic, such as contagion, media coverage, or fake news. Following to the research of Cepoi (2020), we choose six coronavirus news indicators: the PI, the Media Hype Index (HY), the FNI, the CSI, the CTI, and the MCI. A detailed description of the data and the data sources is provided below.

The PI: This index calculates the extent to which news that mentions the coronavirus or hysteria and panic. The HY: This index represents the percentage of news occupied by reporting on the coronavirus. The FNI: This index represents the extent of media discussion about the new virus, which refers to fake news or misinformation that appears with coronavirus. The CSI: This index represents the emotional level of all entities mentioned in the news together with COVID-19. The CTI: This index measures the percentage of all entities (companies, places, etc.) reported with COVID-19 in the media. The MCI: This index represents the percentage of all news sources that cover the COVID-19 pandemic theme. Data for all the indexes were sourced from RavenPack https://coronavirus.ravenpack.com/.

4.3. Summary statistics

Table 1 outlines the descriptive statistics on the volatility components of China’s crude oil and various coronavirus news items. The skewness and kurtosis statistics results submit that the RV, discontinuous jump variation, PI, HY, FNI, CTI, and MCI are “right-skewed” and have “fat tails”. The descriptive statistics and test statistics produced by the Ljung-Box test for 10 and 20 lags are also presented in Table 1. The Ljung-Box Q-statistics (Q (20) and Q (10)) are significantly positive, indicating that both volatility components and the coronavirus news have a substantial dynamic dependence. Furthermore, the results of the augmented Dickey-Fuller test (ADF test) indicate that there is no unit root in every component, except for monthly volatility.

5. Results and discussion

5.1. In-sample estimation results

It is expected that coronavirus news provides a large amount of important information for volatility forecasts of China’s crude oil market. Considering the information overlap among various coronavirus news items and making full use of the information in these news items, we first use lasso and elastic net methods to select coronavirus news indexes for the entire sampling period and then use ordinary least squares (OLS) estimation throughout the sampling period to estimate the HAR-type models and these models with coronavirus news.

Table 2 displays the estimated results of the HAR-RV-type models, and it can be concluded from the results that most of the volatility parameters are significant. In Table 2, the results show that during the COVID-19 pandemic, the PI and CSI have a greater impact on China’s crude oil volatility. In addition, the FNI has a significant impact on the volatility forecasting of China’s crude oil during the coronavirus pandemic. Furthermore, we find that global coronavirus news has more incremental information than China’s coronavirus news that can be used to predict the volatility of China’s crude oil. Finally, we observe that the HAR-RV-type models have a higher R-squared than the HAR-RV model, indicating that for China’s crude oil, the HAR-RV-type models better capture and explain the dynamics of RV than the HAR-RV model. In conclusion, the results reveal that coronavirus news contains a large amount of in-sample information for volatility forecasting.

Table 3 exhibits the parameter estimates for the HAR-RV-J-type models. We find that most volatility parameters are significant. In Table 3, the results show that during the COVID-19 pandemic, the PI and CSI have a greater impact on the volatility forecasting of China’s crude oil futures. Moreover, we find that global coronavirus news has more incremental information than China’s coronavirus news that can be used to predict the volatility of China’s crude oil. The results in Table 3 show that these conclusions are similar to those in Table 2. In conclusion, the results reveal that coronavirus news contains a large amount of in-sample information for volatility forecasting.

The results of the volatility models presented in Tables 2 and 3 suggest that all the coefficients of the PI are positive. The results show that panic has a significant impact on the risk of crude oil price crash, which indicates that the panic sentiment considerably increases the risk, thereby increasing the volatility of China’s crude oil. The results of the volatility models presented in Tables 2 and 3 show that all the coefficients of the CSI are negative, indicating that country sentiment can mitigate the sharp fluctuations in crude oil volatility. The coefficient of the Chinese FNI is positive, indicating that the FNI can influence investors’ decision-making about crude oil, thereby affecting and increasing crude oil volatility.

5.2. Out-of-sample forecasting results

There is no doubt that policymakers and investors may be more concerned about prediction performance than the results for the sample
Note: Asterisks indicate statistical significance at the 1% (***) or 5% (**) levels, and 10% (*) levels. RV, CSI, and PI represent return volatility, China stock index, and price index, respectively.

because they may want to use the model to predict volatility in the future so that they can make further decisions. To address overlapping data, this paper applies the rolling window approach for the prediction and sets a specific length for the out-of-sample prediction. Furthermore, for this part, this article uses an estimation window of 511 observations, beginning on March 26, 2018; thus, the predicted sample includes 218 observations from January 1, 2020, to October 22, 2020.

Table 4 presents the HAR-type model estimation comparison of these models with models that include China’s coronavirus news and global coronavirus news indexes during the COVID-19 pandemic. All p-values of the HAR-type-CCN and HAR-type-WCN models are equal to 1 under different loss functions. The results show that HAR-type-CCN and HAR-type-WCN models outperform the other volatility models, which indicates that the PI and CSI significantly influence China’s crude oil volatility. Most notably, China’s FNI can influence the volatility forecasting of China’s crude oil during the COVID-19 pandemic. Additionally, these results suggest that China’s coronavirus news and global coronavirus news contain a large amount of short-term incremental information about China’s crude oil volatility. To further analyze the impact of the coronavirus news on China’s crude oil volatility, we conduct the analysis summarized in Table 5 and Table 6.

Table 5 displays the MCS results for the period of the coronavirus crisis. All p-values of the HAR-type-WAN models are equal to 1 under different loss functions. In addition, most HAR-type-WCN models are equal to 1 under different loss functions. It can be concluded from the results that HAR-type-WAN and HAR-type-WCN models outperform the other volatility models, which indicates that the global PI and CSI have a greater effect on China’s crude oil volatility. Additionally, these results
indicate that compared with China’s coronavirus news, global coronavirus news provides more incremental information to predict China’s crude oil volatility.

5.3. Further discussion

Inspired by Audrino and Knaus (2016), the framework of our article is different from the those of previous studies, and our model structure is similar to that of Wilms et al. (2021) in that we do not choose the components of volatility but fully control for the daily, weekly, and monthly volatility in the model and select variables for coronavirus news. The in-sample results of the volatility models suggest that all the coefficients of the PI are positive. The results show that panic has a significant impact on the risk of crude oil price crash, which indicates that the panic sentiment considerably increases the risk, thereby increasing the volatility of China’s crude oil. Moreover, the in-sample results of the volatility models show that all the coefficients of the CSI are negative, indicating that country sentiment can mitigate the sharp fluctuations in crude oil volatility. The results indicate that investors may have a certain understanding of the transparency of pandemic information and the development of pandemics; thus, investors have certain expectations for the development of the market so that they will not panic, thereby reducing the fluctuation of market volatility. The coefficient of the Chinese FNI is positive, indicating that the FNI can influence investors’ decision-making about crude oil, thereby affecting and increasing crude oil volatility. Through these findings, we can formulate effective measures to reduce market volatility to reduce investor risk. Specifically, market regulators should strengthen the monitoring of false news reports on the pandemic in the China’s crude oil markets and stabilize investor sentiment in the capital markets to reduce sharp fluctuations in volatility. In addition, the level of sentiment from all news about the coronavirus can reduce crude oil volatility, indicating that reports of coronavirus news can reduce China’s crude oil volatility. Although relevant departments should report more news about the pandemic, they should strengthen the control of false news reports on the pandemic.

The out-of-sample results can be summarized as follows. First, coronavirus news is helpful to optimize the out-of-sample prediction accuracy of traditional HAR models. Second, adding a large amount of coronavirus news to the HAR model may not significantly improve the volatility prediction. Due to the overlapping information among various coronavirus news items, the effect of studying various coronavirus news items on the improvement of volatility prediction performance is not obvious. Third, we find that the results are robust with the rolling window method and alternative realized measures analysis. These results show that global coronavirus news provides more incremental information than China’s coronavirus news in both in-sample and out-of-sample analysis, which can predict the volatility of China’s crude oil. Global coronavirus news has more predictive value for China’s crude oil market information than China’s coronavirus news. There may be two reasons. First, global coronavirus news covers China’s coronavirus news information, so global coronavirus news contains more information about the volatility of China’s crude oil. Second, we believe that from March 2020 to April 2020, the pandemic was well controlled in China, and as a result, the Chinese market was more sensitive to international pandemic information.

6. Robustness checks

This section uses two different methods to verify the robustness of out-of-sample prediction results. First, this article chooses an alternative volatility-RK to re-estimate the RV and jump component, evaluate the predictive power of these models, and then test the robustness of the results of this paper. Second, as in a previous study, this paper uses the various coronavirus news items, the effect of studying various coronavirus news items on the improvement of volatility prediction performance is not obvious. Third, we find that compared with China’s coronavirus news, global coronavirus news provides more incremental information to predict China’s crude oil volatility.

| Table 4 | MCS test results. |
|----------|------------------|
| Volatility models | Range | SemiQ |
|                | HMAE | HMSE | HMAE | HMSE |
| HAR-RV        | 0.3893 | 0.3527 | 0.3893 | 0.3527 |
| HAR-RV-CN     | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| HAR-RV-J      | 0.3893 | 0.3526 | 0.3893 | 0.3526 |
| HAR-RV-J-CN   | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| HAR-RV-J-WCN  | 0.5137 | 0.4095 | 0.5137 | 0.4095 |
| HAR-RV-J-WCN  | 1.0000 | 1.0000 | 1.0000 | 1.0000 |

Notes: The p-value of MCS is calculated based on the test statistics $T_R$ and $T_{SQ}$. The value of $p > 0.1$ is shown in bold.

| Table 5 | MCS test results. |
|----------|------------------|
| Volatility models | Range | SemiQ |
|                | HMAE | HMSE | HMAE | HMSE |
| HAR-RV        | 0.3842 | 0.3796 | 0.2795 | 0.1980 |
| HAR-RV-CN     | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| HAR-RV-J      | 0.3842 | 0.3796 | 0.2795 | 0.1980 |
| HAR-RV-J-CN   | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| HAR-RV-J-WAN  | 0.6341 | 0.4102 | 0.6328 | 0.3277 |
| HAR-RV-WCN    | 0.8239 | 1.0000 | 0.8239 | 1.0000 |
| HAR-RV-J-WAN  | 0.6420 | 0.4313 | 0.6421 | 0.3270 |
| HAR-RV-J-WCN  | 0.8260 | 1.0000 | 0.8260 | 1.0000 |

Notes: The p-value of MCS is calculated based on the test statistics $T_R$ and $T_{SQ}$. The value of $p > 0.1$ is shown in bold.

| Table 6 | MCS test results. |
|----------|------------------|
| Volatility models | Range | SemiQ |
|                | HMAE | HMSE | HMAE | HMSE |
| HAR-RV        | 0.4103 | 0.3740 | 0.6096 | 0.3821 |
| HAR-RV-CN     | 0.4784 | 0.3740 | 0.6096 | 0.3821 |
| HAR-RV-WCN    | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| HAR-RV-J      | 0.4117 | 0.3742 | 0.6108 | 0.3907 |
| HAR-RV-J-CN   | 0.4789 | 0.3742 | 0.6108 | 0.3907 |
| HAR-RV-J-WCN  | 1.0000 | 1.0000 | 1.0000 | 1.0000 |

Notes: The p-value of MCS is calculated based on the test statistics $T_R$ and $T_{SQ}$. The value of $p > 0.1$ is shown in bold.

suggest that all China’s coronavirus news and global coronavirus news contain a large amount of short-term incremental information about China’s crude oil market volatility, which is in line with the MCS test results shown in Table 4. Table 6 presents the HAR models and their comparison with the expansion of China’s coronavirus news index and the global coronavirus news index during the coronavirus crisis. All p-values of the HAR-type-WCN models are equal to 1 under different loss functions, indicating that HAR-type-WCN models outperform other volatility models, which is in line with the MCS test results in Table 5. Additionally, these results indicate that compared with China’s coronavirus news, global coronavirus news has more incremental information to predict China’s crude oil market volatility.

Our major empirical results can be summarized as follows. First, coronavirus news is helpful to optimize the out-of-sample prediction accuracy of traditional HAR models. Second, adding a large amount of coronavirus news to the HAR model may not significantly improve the volatility prediction. Due to the overlapping information provided by
rolling window method to obtain forecasts for China’s crude oil volatility. To confirm our results, we employ another method for calculating the forecasting window, the recursive window and reduct the empirical analysis (Neely et al., 2014).

6.1. Realized kernel

It is crucial for the choice of appropriate benchmark volatility in evaluations of the prediction performance of the model. In addition, as stated above, when estimating RV, market microstructural noise has a great influence. Therefore, another volatility-RK (Bardorff-Nielsen et al., 2008), which is robust to noise, is used to test whether the previous results are robust and reliable. The out-of-sample results in Tables 7-9 in this section indicate that these conclusions are similar to those obtained in Section 4.2.

6.2. Recursive window method

Tables 10-12 exhibit the prediction performance of the model. Table 10 displays the HAR-type models and a comparison of these models with models that include China’s coronavirus news and global coronavirus news indexes during the COVID-19 pandemic. We find that HAR-type-CCN and HAR-type-WCN models have the best predictive performance. These results suggest that HAR-type-CCN and HAR-type-WCN models contain a large amount of short-term incremental information about China’s crude oil market volatility. This result shows that China’s coronavirus news is helpful to optimize the out-of-sample prediction accuracy of traditional HAR models, which is in line with the MCS results shown in Table 4. Table 11 shows the MCS results for the period of the coronavirus crisis. All p-values of the HAR-type-WAN models are equal to 1 under different loss functions. Moreover, most HAR-type-WCN models are equal to 1 under different loss functions. The results indicate that HAR-type-WAN and HAR-type-WCN models outperform other volatility models. These results suggest that both China’s coronavirus news and global coronavirus news contain a large amount of short-term incremental information about China’s crude oil market volatility. Additionally, the results indicate that both China’s coronavirus news and global coronavirus news is helpful to optimize the out-of-sample prediction accuracy of the traditional HAR models, which is in line with the MCS test results shown in Table 5.

Table 12 presents the HAR models and their comparison with the models that include China’s coronavirus news and global coronavirus news indexes during the COVID-19 pandemic. We find that HAR-type-WAN and HAR-type-WCN models outperform other volatility models. These results indicate that HAR-type-WAN and HAR-type-WCN models contain a large amount of short-term incremental information about China’s crude oil market volatility. Additionally, the results indicate that both China’s coronavirus news and global coronavirus news is helpful to optimize the out-of-sample prediction accuracy of the traditional HAR models, which is in line with the MCS test results shown in Table 5.

Table 12 exhibits the prediction performance of the model. The value of p > 0.1 is shown in bold.

Table 8 MCS test results.

| Volatility models  | Range | SemiQ |
|--------------------|-------|-------|
|                    | HMAE  | HMSE  | HMAE  | HMSE  |
| HAR-RV             | 0.3840| 0.3798| 0.3759| 0.1990|
| HAR-RV-CAN         | 1.0000| 1.0000| 1.0000| 1.0000|
| HAR-RV-CCN         | 0.3840| 0.3798| 0.2742| 0.1989|
| HAR-RV-J           | 1.0000| 1.0000| 1.0000| 1.0000|
| HAR-RV-J-CAN       | 0.3840| 0.3798| 0.2742| 0.1989|
| HAR-RV-WCN         | 1.0000| 1.0000| 1.0000| 1.0000|
| HAR-RV-J-WCN       | 0.3810| 0.3472| 0.3810| 0.3472|
| HAR-RV-J-WAN       | 0.3854| 0.3441| 0.3870| 0.3581|
| HAR-RV-J-CCN       | 1.0000| 1.0000| 0.3814| 0.3608|

Notes: The p-value of MCS is calculated based on the test statistics T_R and T_Q. The value of p > 0.1 is shown in bold.

Table 9 MCS test results.

| Volatility models  | Range | SemiQ |
|--------------------|-------|-------|
|                    | HMAE  | HMSE  | HMAE  | HMSE  |
| HAR-RV             | 0.4854| 0.3926| 0.5029| 0.3881|
| HAR-RV-CCN         | 0.4854| 0.3926| 0.5029| 0.3881|
| HAR-RV-WCN         | 1.0000| 1.0000| 1.0000| 1.0000|
| HAR-RV-J           | 0.4891| 0.3928| 0.5073| 0.3882|
| HAR-RV-J-CCN       | 0.4891| 0.3928| 0.5073| 0.3882|
| HAR-RV-J-WCN       | 1.0000| 1.0000| 1.0000| 1.0000|

Notes: The p-value of MCS is calculated based on the test statistics T_R and T_Q. The value of p > 0.1 is shown in bold.

Table 10 MCS test results.

| Volatility models  | Range | SemiQ |
|--------------------|-------|-------|
|                    | HMAE  | HMSE  | HMAE  | HMSE  |
| HAR-RV             | 0.4105| 0.3528| 0.4105| 0.3528|
| HAR-RV-CCN         | 0.4114| 0.3528| 0.4114| 0.3528|
| HAR-RV-WCN         | 1.0000| 1.0000| 1.0000| 1.0000|
| HAR-RV-J           | 0.3743| 0.3509| 0.3743| 0.3509|
| HAR-RV-J-CCN       | 1.0000| 1.0000| 1.0000| 1.0000|
| HAR-RV-J-WCN       | 0.5216| 0.4372| 0.5216| 0.4372|
| HAR-RV-J-WAN       | 1.0000| 1.0000| 1.0000| 1.0000|

Notes: The p-value of MCS is calculated based on the test statistics T_R and T_Q. The value of p > 0.1 is shown in bold.

Table 11 MCS test results.

| Volatility models  | Range | SemiQ |
|--------------------|-------|-------|
|                    | HMAE  | HMSE  | HMAE  | HMSE  |
| HAR-RV             | 0.3846| 0.3796| 0.2809| 0.1980|
| HAR-RV-CCN         | 1.0000| 1.0000| 1.0000| 1.0000|
| HAR-RV-J           | 0.3846| 0.3796| 0.2791| 0.1982|
| HAR-RV-J-CCN       | 0.3846| 0.2791| 0.2791| 0.1982|
| HAR-RV-WCN         | 0.6345| 0.3537| 0.6333| 0.3273|
| HAR-RV-J-CCN       | 0.6345| 0.3537| 0.6333| 0.3273|
| HAR-RV-J-WCN       | 0.8238| 0.3602| 0.8259| 0.3608|
| HAR-RV-J-WAN       | 0.8238| 0.3602| 0.8259| 0.3608|
| HAR-RV-J-WCN       | 0.8238| 0.3602| 0.8259| 0.3608|

Notes: The p-value of MCS is calculated based on the test statistics T_R and T_Q. The value of p > 0.1 is shown in bold.

Table 7 MCS test results.

| Volatility models  | Range | SemiQ |
|--------------------|-------|-------|
|                    | HMAE  | HMSE  | HMAE  | HMSE  |
| HAR-RV             | 0.3819| 0.3702| 0.3819| 0.3702|
| HAR-RV-CCN         | 1.0000| 1.0000| 1.0000| 1.0000|
| HAR-RV-J           | 0.3820| 0.3702| 0.3820| 0.3702|
| HAR-RV-J-CCN       | 1.0000| 1.0000| 1.0000| 1.0000|
| HAR-RV             | 0.5296| 0.3853| 0.5296| 0.3853|
| HAR-RV-CCN         | 1.0000| 1.0000| 1.0000| 1.0000|
| HAR-RV-J           | 0.5319| 0.4416| 0.5319| 0.4416|
| HAR-RV-J-CCN       | 1.0000| 1.0000| 1.0000| 1.0000|

Notes: The p-value of MCS is calculated based on the test statistics T_R and T_Q. The value of p > 0.1 is shown in bold.
Appendix A

This paper follows some related works (Li et al., 2015; Wilms et al., 2021; Zhang et al., 2019a, 2019b) and uses the two prevalent shrinkage methods, namely, the lasso method defined by Tibshirani (1996) and the elastic net method pioneered by Zou and Hastie (2005). Therefore, the prediction of China’s crude oil volatility by lasso is as follows:

\[
\hat{RV}_{i+1} = \hat{\beta}_0 + \sum_{j=1}^{p} \hat{\beta}_j x_{ij},
\]  

where

\[
\hat{\beta} = \arg \min_{\beta} \left( \frac{1}{2(t-1)} \sum_{i=1}^{t-1} \left( RV_{i+1} - \hat{\beta}_0 - \sum_{j=1}^{p} \hat{\beta}_j x_{ij} \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right),
\]

\[
RV_{i+1} \text{ denotes the } RV \text{ of China’s crude oil on day } t + 1, N \text{ represents the number of all the used predictors, } x_{ij} \text{ is the } i\text{-th predictor variable on day } t, \hat{\beta} \text{ represents the shrinkage estimator of the regression coefficient, and } \lambda \text{ denotes the nonnegative regularization parameter. To use the lasso, we need to decide which value of } \lambda \text{ is best. This paper denotes the selected } \lambda \text{ as } \lambda'. \text{ Common methods for selecting } \lambda \text{ for lasso are } CV, \text{ adaptive lasso, and plugin estimators. This article uses the } CV \text{ method to choose } \lambda. \]

Next, we introduce elastic net, which performs automatic variables choice and continuous shrinkage simultaneously, and can choose related variable groups. The elastic net is an extension of the lasso, which overcomes the lasso restriction by combining the ridge and lasso regression methods. The regression coefficients can be estimated as:

\[
\hat{\beta} = \arg \min_{\beta} \left( \frac{1}{2(t-1)} \sum_{i=1}^{t-1} \left( RV_{i+1} - \hat{\beta}_0 - \sum_{j=1}^{p} \hat{\beta}_j x_{ij} \right)^2 + \lambda \sum_{j=1}^{p} \left( (1 - \alpha)\beta_j^2 + \alpha|\beta_j| \right) \right),
\]

where \(\alpha\) denotes a positive number, strictly between 0 and 1. In particular, the elastic net approaches the ridge regression when \(\alpha = 0\), and the elastic net reduces to the lasso when \(\alpha = 1\). Therefore, the shrinkage parameters \(\lambda\) and \(\alpha\) and the tuning parameters \(\lambda\) and \(\alpha\) are selected via CV.

Table 12
MCS test results.

| Volatility models        | Range HMAE | Range HMSE | SemiQ HMAE | SemiQ HMSE |
|--------------------------|------------|------------|------------|------------|
| HAR-RV                   | 0.4123     | 0.3753     | 0.6062     | 0.3841     |
| HAR-RV-CCN               | 0.4713     | 0.3753     | 0.6062     | 0.3841     |
| HAR-RV-WCN               | 1.0000     | 1.0000     | 1.0000     | 1.0000     |
| HAR-RV-J                 | 0.4142     | 0.3753     | 0.6067     | 0.3934     |
| HAR-RV-J-CCN             | 0.4714     | 0.3753     | 0.6067     | 0.3934     |
| HAR-RV-J-WCN             | 1.0000     | 1.0000     | 1.0000     | 1.0000     |

Notes: The p-value of MCS is calculated based on the test statistics \(T_R\) and \(T_Q\). The value of \(p > 0.1\) is shown in bold.

7. Conclusion

In this research, we empirically investigate whether coronavirus news plays a significant role in the predictability of China’s crude oil market volatility by introducing coronavirus news indicators (PI, HY, FNI, CSI, CTI, and MCI) in the traditional HAR models and their corresponding extended models. First, to test the effect of coronavirus news with other information on volatility predictions, we use elastic net and lasso methods to choose coronavirus news during the entire sampling period and then use OLS estimation throughout the sampling period to estimate HAR-type models and these models with coronavirus news. Second, we analyze the in-sample estimation and out-of-sample prediction capabilities of these models for a one-day horizon. Finally, this article uses the alternative RK function and recursive window method to reduce the panic of investors and reduce the fluctuation of China’s crude oil volatility. Furthermore, China’s FNI has a significant influence on the volatility forecasting of China’s crude oil during the coronavirus pandemic. Third, we find that global coronavirus news provides more incremental information than China’s coronavirus news that can be used to forecast the volatility of China’s crude oil market.

This research is of great value to policymakers, academics, and investors in China’s crude oil markets for three reasons. First, it should be pointed out that coronavirus news is a very rich source of information for predicting the volatility of the China’s oil market, and the inclusion of coronavirus news in the HAR model showed consistent out-performance. This finding is of great significance for investors who used Shanghai International Energy Exchange crude oil for portfolio hedging and trading strategies during the COVID-19 pandemic. Second, our findings not only are important for studying the future performance of the crude oil market, but also for those who wish to apply the method to other commodity markets. This method takes coronavirus news into account in volatility prediction and may provide new discoveries when applied to other markets. These findings may provide new research directions for improving the accuracy of market volatility forecasts. Finally, during the COVID-19 pandemic, the PI and CSI have a greater impact on China’s crude oil volatility. This result shows that the Panic index and the Country Sentiment index can significantly affect the volatility of China’s crude oil. Therefore, the relevant departments should increase the control of false news reports on the pandemic to reduce the panic of investors and reduce the fluctuation of China’s crude oil prices. Moreover, compared with China’s coronavirus news, global coronavirus news provides more incremental information that can be used to predict the volatility of China’s crude oil. In summary, the methods proposed and applied in this paper help make better use of the informational content of coronavirus news and hence, in many cases, can obtain better predictions of the variables of interest.
Author statement

Zibo Niu: Data Formal analysis, Writing – original draft, Writing – review & editing. Yuanyuan Liu: Supervision, Conceptualization, Writing – review & editing. Gao Wang: Data curation, Writing – review & editing. Hongwei Zhang: Conceptualization, Supervision, Methodology, Software, Writing – review & editing, Validation.

References

Adekoya, O.B., Oliyide, J.A., 2021. How COVID-19 drives connectedness among commodity and financial markets: evidence from TVP-VAR and causality-in-quantiles techniques. Resour. Pol. 70, 101898.
Ambros, M., Frenkel, M., Huynh, T.L.D., Klinke, M., 2020. COVID-19 pandemic news and stock market reaction during the onset of the crisis: evidence from high-frequency data. Appl. Econ. Lett. 1–4.
Andersen, T.G., Bollerslev, T., 1998a. Answering the skeptic: yes, standard volatility models do provide accurate forecasts. Int. Econ. Rev. 3885–95.
Andersen, T.G., Bollerslev, T., 1998b. Deutsche mark–dollar volatility: intraday activity patterns, macroeconomic announcements, and longer run dependences. J. Finance 53, 219–265.
Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P., 2003. The behavior of crude oil spot and futures prices around OPEC news announcements. Energy Econ. 25, 268–288.
Andersen, T.G., Bollerslev, T., Meddahi, N., 2004. Analytical evaluation of volatility forecasting. Int. Econ. Rev. 45, 1079–1110.
Andersen, T.G., Dobrev, D., Schaumburg, T., 2012. Jump-robust volatility estimation using nearest neighbor truncation. J. Econometrics 169, 75–93.
Arouri, M.E.H., Luan, A., Nguyen, D.K., 2012. Forecasting the conditional volatility of oil spot and futures prices with structural breaks and long memory models. Energy Econ. 34, 283–293.
Audrino, F., Knaus, S.D., 2016. Lassoing the HAR model: a model selection perspective on realized volatility dynamics. Econ. Rev. 35, 1485–1521.
Bae, S., Mohanty, S.K., Glambosky, M., 2020. COVID-19 and stock market volatility: an industry level analysis. Finance Res. Lett. 37, 101748.
Balak, D., Triantafyllou, A., 2020. Commodity price volatility and the economic unimportance of pandemics. Energy Econ. Lett. 19, 101552.
Baker, S.R., Bloom, N., Davis, S.J., Kost, K., Sammon, M., Viratyosin, T., 2020. The uncertainty of pandemics. Econ. Lett. 193, 109283.
Blitz, D., Huisman, R., Swinkels, L., van Vliet, P., 2020. Media attention and the volatility effect. Economics Rev. 35, 1485–1521.
Buenaventura, R., Oktay, E., 2016. Modeling the dependency of conditional correlations on market volatility. J. Bus. Econ. Stat. 34, 254–268.
Blitz, D., Huisman, R., Sviokla, J., van Vliet, P., 2020. Media attention and the volatility effect. Finance Res. Lett. 36, 101317.
Bonfim, A.N., 2003. Pre-announcement events, news effects, and volatility: monetary policy and the stock market. J. Bank. Finance 27, 133–151.
Brandt, M.W., Gao, L., 2019. Macro fundamentals or geopolitical events? A textual analysis of news events for crude oil. J. Empir. Finance 51, 64–94.
Broadstock, D.C., Zhang, D., 2019. Social-media and intraday stock returns: the pricing power of sentiment. Finance Res. Lett. 30, 116–123.
Cepoi, C.O., 2020. Asymmetric dependence between stock market returns and news during COVID19 financial turmoil. Finance Res. Lett. 36, 101552.
Chen, X., Ghysels, E., 2011. News—good or bad—and its impact on volatility predictions over multiple horizons. Rev. Financ. Stud. 24, 46–81.
Chua, C.L., Tsaiplas, S., 2019. Information flows and stock market volatility. J. Appl. Econ. 34, 129–148.
Corsi, F., 2009. A simple approximate long-memory model of realized volatility. Econometrica 77, 579–625.
Demos, A., Luen, A., Nason, J.M., 2015. 2011. The model confidence set. Econometrica 79, 653–697.
Zhang, D., Hu, M., Ji, Q., 2020. Financial markets under the global pandemic of COVID-19. Finance Res. Lett. 36, 101528.
Zhang, H., Zhu, X., Guo, Y., Liu, H., 2018. A separate reduced-form volatility forecasting model for nonferrous metal market: evidence from copper and aluminum. J. Forecast. 37, 754-766.
Zhang, W., Hamori, S., 2021. Crude oil market and stock markets during the COVID-19 pandemic: evidence from the US, Japan, and Germany. Int. Rev. Financ. Anal. 74, 101702.
Zhang, Y., Ma, F., Wang, Y., 2019a. Forecasting crude oil prices with a large set of predictors: can LASSO select powerful predictors? J. Empir. Finance 54, 97-117.
Zhang, Y., Wei, Y., Zhang, Y., Jin, D., 2019b. Forecasting oil price volatility: forecast combination versus shrinkage method. Energy Econ. 80, 423-433.
Zhu, X.-h., Zhang, H.-w., Zhong, M.-r., 2017a. Volatility forecasting in Chinese nonferrous metals futures market. Trans. Nonferrous Metals Soc. China 27, 1206-1214.
Zhu, X., Zhang, H., Zhong, M., 2017b. Volatility forecasting using high frequency data: the role of after-hours information and leverage effects. Resour. Pol. 54, 58-70.
Zou, H., Hastie, T., 2005. Regularization and variable selection via the elastic net. J. Roy. Stat. Soc. B 67, 301-320.