

Research Article

A Lead-Lag Relationship and Forecast Research between China's Crude Oil Futures and Spot Markets

Chi Zhang¹, Dandan Pan², Mingyan Yang³, and Zhengning Pu¹³

¹School of Economics, Nanjing University of Posts and Telecommunications, 9# Wenyuan Road, Nanjing, Jiangsu, China

²School of Business, Nanjing University, #16 Jinyin Street, Nanjing, Jiangsu, China

³School of Economics & Management, Southeast University, 2# Sipailou, Nanjing, Jiangsu, China

Correspondence should be addressed to Zhengning Pu; puzhengning@seu.edu.cn

Received 2 December 2021; Revised 25 March 2022; Accepted 7 April 2022; Published 21 April 2022

Academic Editor: Paulo Jorge Silveira Ferreira

Copyright © 2022 Chi Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The integration of the global economy has led to an increasingly strong connection between the futures and spot markets of commodities. First, based on one-minute high-frequency prices, this paper applies the thermal optimal path (TOP) method to examine the lead-lag relationship between Chinese crude oil futures and spot from March 2018 to December 2021. Second, we apply the Mixed Frequency Data Sampling Regression (MIDAS) model and indicators such as deviation degree to test the degree of prediction of high-frequency prices in the futures market to the spot market. The experimental results show that the futures markets lead the spot market most of the time, but the lead effect reverses when major events occur; 60-minute futures high-frequency prices are the most predictive of daily spot data; crude oil futures' predictive power declined after the Covid-19 outbreak and is more predictive when night trading is available. This study has important implications, not only to guide investors but also to provide empirical evidence and valid information for policy makers.

1. Introduction

Historically, financial experts and investors have devoted considerable attention to the relationships between international commodity markets and whether they exhibit similar price characteristics and converge over time or indeed are fully integrated. Concepts such as lead-lag relationships and comovement are often used to study efficiency in financial markets. The leading-lag relationship between futures and spot markets for the same commodity is always reflected by the transaction price, thus presenting the respective information efficiency. In the commodity market, crude oil futures have not only commodity attributes, but also financial attributes. As an important strategic resource, not only is crude oil related to energy security and people's livelihood development, but it also is an important hedging tool in the crude oil market, which in turn affects crude oil pricing. Therefore, it is of great practical value to study the leading-lag relationship between crude oil futures and crude oil spot prices and to forecast them. On the one hand, if investors can accurately

predict the trend of future liquidity in the market, then they can adjust their investment strategies and improve their investment returns in a timely manner; on the other hand, regulators can focus on monitoring possible future anomalies and take relevant measures to avoid abnormal market fluctuations and promote the healthy development of the market. China's crude oil futures have been traded since March 2018 and have now become one of the four major crude oil futures trading varieties in the world and have also played an important role in the international crude oil market. Then, what kind of leading-lag relationship exists between China's crude oil futures and China's crude oil spot and whether the movement of its price information can predict the price change of China's crude oil spot are the focus of our research concern. In this paper, we will use TOP and MIDAS to investigate the lead-lag relationship between the two markets. In addition, the outbreak of the Covid-19 epidemic suspended trading of Chinese crude oil futures for three months, giving us an additional window to observe the trend of crude oil spot prices during this period.

The rest is arranged as follows: Section 2 is literature review, Section 3 introduces the data sources and methods, Section 4 gives empirical results, and Section 5 concludes the paper.

2. Literature Review

Since this paper explores the lead-lag and forecasting relationship between Chinese crude oil futures and spot using both the TOP in Econophysics and MIDAS, we provide an overview of the literature in terms of both research subjects and research methods to paint a profile of the current research on Chinese crude oil futures.

2.1. Relationship between Futures and Spot Market. The relationship between futures and spot prices has been studied for a long time. At present, there are many related literature pieces, mainly focusing on the leading-lag relationship and price discovery. The research on the guiding relationship of spot prices was initially empirically tested and analyzed by Garbade and Silver, and a dynamic analysis model showing the relationship was established for the spot prices [1]. At present, the relationship between futures and spot research covers almost all financial derivatives markets, including index futures [2, 3] and energy futures [4]. Academia has also conducted a lot of research on the international mainstream oil futures and explored its impact on the stock market [5, 6], industrial chain [7, 8], and national economic conditions [9, 10]. With the diversification of financial instruments, oil futures are widely used in the risk hedging mechanism to control the risks of spot oil transactions. Once the oil futures market price changes, the price fluctuations will directly spread to the spot market.

Prior to 2018, most studies in this area focused on the spillover effects between the Chinese crude oil spot market and the international crude oil futures market, finding that global financial markets such as crude oil show a trend of global integration [11] and that there are also significant spillover effects between Chinese crude oil spot and WTI crude oil futures [12–14]. A study of the relationship between fuel oil futures and other energy financial derivatives also finds that the correlation between Chinese fuel oil spot, fuel oil futures, and energy equity markets is weaker than that of the US market, and the strength of the correlation has weakened after the financial crisis [15]. With the listing of Chinese crude oil futures after 2018, academics began to focus on its relationship with global benchmark crude oil futures, trying to analyze the impact it brings to the global energy finance market from several perspectives.

In terms of research on the relationship between Chinese crude oil futures and spot, using a complex network model, Ji et al. found a higher degree of integration with the global market during the nighttime trading hours but a regional fractional effect during the daytime hours [16]. By analyzing the correlation between Chinese crude oil futures and spot, Jie et al. found that Chinese crude oil futures achieved their main functions and establishment goals [17]. Most of these studies analyze the correlation between Chinese crude oil

futures and international crude oil futures. Yang et al. examined the transmission characteristics of returns and volatility between Chinese crude oil futures and international crude oil futures based on intraday data and found that there is a cointegration relationship between crude oil futures [18]. Huang et al. used wavelet analysis to investigate the return linkage characteristics between Chinese crude oil futures and the two major international benchmark crude oil futures (WTI and Brent). The results show that the linkage between Chinese crude oil futures and international crude oil futures is relatively weak and directionally variable, and the linkage with international crude oil prices is very different, and the price fluctuations of Chinese crude oil futures lag behind those of international crude oil futures [19]. Palao et al. found that Chinese crude oil futures do not have influence on WTI and Brent crude oil futures prices, but they are sensitive to Brent crude oil futures price fluctuations [20]. Similarly, Yang et al. constructed a VaR correlation network between Chinese crude oil futures and international crude oil futures for both upside and downside risks [21].

In 2020, with the outbreak of COVID-19, the changes regarding the crude oil futures market during the epidemic have also attracted the interest of many scholars. Most scholars have uncovered the linkages and spillovers between world benchmark crude oil futures and other financial derivative markets during the epidemic, including gold, stock markets, cryptocurrencies, and crude oil spot [22–25]. Jefferson analyzed the impact of the epidemic on the oil industry in different countries and found that, despite a partial recovery, many oil industries will still suffer [26]. Wang investigated the impact of multifractal cross-correlation between crude oil and agricultural futures under the epidemic and found that multifractal correlation between crude oil and agricultural futures increases after the epidemic [27]. Wang et al. found that the COVID-19 disrupted the oil supply and demand balance, reducing US oil by 18.14% [28]. Peng et al. found significant risk spillovers from the stock market to the oil market during the COVID-19 pandemic; the oil market was subject to high-risk spillovers from the second board market; and bidirectional risk spillovers between the Chinese stock market and the oil market increased rapidly [29]. Regarding Chinese crude oil futures, Lin and Su found that the driving effect of macro factors in China on future prices has oscillated after the Covid-19 outbreak [30]. In addition, the outbreak exacerbated the spillover effects of the commodity markets to the US and Chinese stock markets [31]. At the same time, the coflow of Chinese crude oil futures with international benchmark crude oil futures weakened, thus providing investors with a way to hedge their risks [16].

2.2. TOP and MIDAS

2.2.1. Thermal Optimal Path (TOP). Sornette and Zhou use the GDP growth rate, the unemployment rate, and the inflation rate in the United States and proposed TOP method to verify their time-varying linkage relationship and further explain that TOP has advantages over cross-correlation

analysis [32, 33]. Compared with traditional static research methods, TOP can accurately describe the time-varying leading and lagging relationship between two time series. This method is not limited by the general parameter distribution assumptions and can well show the structural changes of the nonlinear relationship in the leading-lag relationship and can determine the leading-lag direction and leading-lag order at each moment.

TOP has been currently widely used in real economic research and finance. Guo et al. studied the leading and lagging relationship between S & P and bond yields and found that the stock market can significantly guide short-term bonds and have a weaker effect on long-term bonds [34]. In addition, they also found that the Chinese stock market has a barometer effect on the economy [35]. Xu et al. used the thermal optimal path method to analyze the intraday and interday leading and lagging relationship between the onshore and offshore RMB exchange rates [36]. Mend et al. innovatively used self-consistent testing to test the rationality and effectiveness of the TOP method results and further improved the TOP method [37].

2.2.2. Mixed Frequency Data Sampling Regression Models (MIDAS). The data selected by the traditional econometric model when analyzing time series data must have the same frequency; otherwise, the model cannot be identified. Therefore, when the variable frequencies are different, the academic community usually adopts two types of processing methods: one is to use interpolation to make the low-frequency data high-frequency data; the other is to use the simple average method to make the high-frequency sample data low-frequency sample data. However, these methods ignore the timeliness of high-frequency data, and the processing results may lose the effective information of high-frequency data, which will lead to serious consequences such as distortion of quantitative analysis and reduced prediction accuracy.

To solve this problem, Ghysels et al. first proposed the MIDAS model in 2004 [38], which allows the frequency of variables on both sides of the model equation to be different. High-frequency variables are summed by a parameterized polynomial weight function to become variables at the same frequency as the low-frequency variables. MIDAS was originally used for volatility prediction [39, 40]. Recent work mostly uses monthly data to improve quarterly macro forecasts through regression [41–43] or use daily financial data to improve quarterly and monthly macroeconomic forecasts [44].

From another perspective, we can regard MIDAS as a simplified representation of the linear projection generated from the state-space model. In this simplified form, MIDAS regression does not require a complete state-space equation. Bai et al. pointed out that, in some cases, MIDAS regression is an accurate representation of the Kalman filter [45]. From another perspective, MIDAS can be seen as a simplified representation of the linear projection generated from the state-space model. Although the Kalman filter is optimal when performing linear prediction, it requires a complete

measurement system and state equations, and it is more prone to setting errors, so it requires more parameters, which leads to computational complexity. In contrast, when involving large datasets, MIDAS regression is computationally easy to implement but is more prone to setting errors [44].

3. Data and Methods

3.1. Data. Data used in the paper derives from two sources. The futures prices, including daily data, 60 min, 30 min, and 15 min high-frequency data, come from Tongdaxin database. The spot prices, based on the data availability and integrity, use the daily data of Shengli crude oil spot selected from Wind database. Considering that China crude oil futures were officially listed and traded from March 26, 2018, the data period of futures and spot in this article is from March 27, 2018, to December 31, 2021. Aiming at the collected data, this paper uses two methods for research. The first is the TOP, used to find the lead-lag relationship between two sets of time series of equal length. In this article, we use daily data of futures and spot for TOP analysis and both sets of series included 890 observations after matching. The second method is MIDAS. Due to the impact of the COVID-19 epidemic, the China Securities Regulatory Commission canceled night trading of crude oil futures on February 3 and resumed on May 6, which gave us an opportunity to study the control sample. In addition, due to the cancellation of the night market, the frequency of high-frequency data was not uniform throughout the sample period. Therefore, in the MIDAS analysis, we divide the sample period into three parts: the first part is with night trading, from March 27, 2018, to January 23, 2020, the second part is without night trading, from February 3 to April 30, 2020, and the third is with night trading, from May 8, 2020, to December 31, 2021. The price observations of one day of different frequencies of crude oil futures in two periods are shown in Table 1.

3.2. Methods

3.2.1. TOP. The main algorithm of TOP method is recursive operation by using the classic model in physics: probability transfer model. TOP introduces the distance matrix and estimates the parameter without parameters, then uses the classic partition function for recursing, and finally obtains the dynamic leading-lag relationship between the two sequences. In this article, we use Python to complete this algorithm calculation and the steps of TOP method are as follows.

First, normalizing two sequences and their distance matrix E_{xy} can be obtained by the following equation, and the simulated result is shown in Figure 1:

$$\varepsilon(t_1, t_2) = |X(t_1) - Y(t_2)|, \quad (1)$$

where $\{X\}, \{Y\}$ indicates the time series has been standardized. As shown in (1), matrix element $\varepsilon(t_1, t_2)$ represents the distance between the value of time series $\{X\}$ at

TABLE 1: High-frequency observations in different periods.

	With night trading	Without night trading
Trading days	828	62
60 min obs/day	10	5
30 min obs/day	19	8
15 min obs/day	37	15

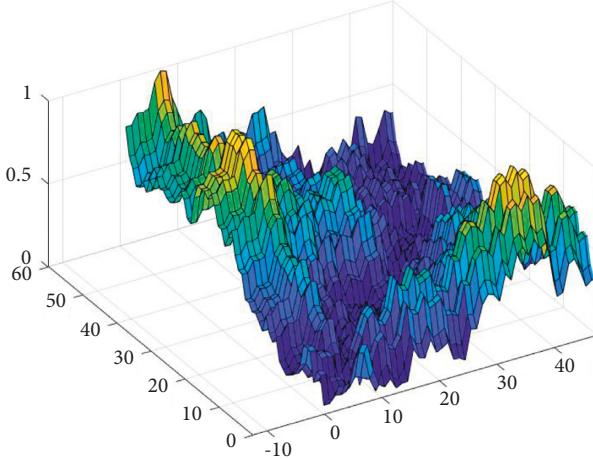


FIGURE 1: Distance matrix of simulation data.

time t_1 and the value of time series $\{Y\}$ at time t_2 , which can also be called the energy value. TOP takes the lowest energy as the objective function and achieves this by finding the correct path.

Rotating the coordinate system in (1),

$$\begin{cases} t_1 = 1 + \frac{(t-x)}{2}, \\ t_2 = 1 + \frac{(t+x)}{2}. \end{cases} \quad (2)$$

By the coordinate rotation of (2), t -axis of the new coordinate system is the direction of the main diagonal of the original coordinate system, x -axis of the new coordinate system is the subdiagonal of the original coordinate system, and y -axis is perpendicular to t -axis.

Secondly, the partition function relies on physical ideas. Zhou and Sornette use the Boltzmann factor as the weight of the partition function, and plenty of studies have shown that the leading-lag relationship of time series can be more accurate in the two-level algorithm. The algorithm only follows three different directions, vertical, horizontal, and diagonal, to approach the target and minimizes the energy value. Therefore, its corresponding partition function can be obtained in the following equation:

$$G(x, t+1) = [G(x-1, t) + G(x+1, t) + G(x, t-1)]e^{-\varepsilon(x,t)/T}. \quad (3)$$

T in (3) represents the allowable temperature, indicating the probability that the path deviates from the maximum value of the target increases with the allowable temperature. When $T = 0$, equation (3) means to find the corresponding

minimum value of x at any time t . However, the actual data is often random and unstable. The time series will overfit when T is too small and will lose massive information if T is too high. It can be seen that choosing an appropriate temperature is very important to determine a more accurate thermal optimal path.

Finally, calculating the thermal average position $x(t)$, at each time t , the probability that each x on the partition function can be taken is $G(x, t)/G(t)$, which is inversely proportional to energy and $G(t) = \sum_x G(x, t)$. The average leading-lag order is obtained by the following equation:

$$x(t) = \sum_x \frac{xG(x, t)}{G(t)}. \quad (4)$$

After the above steps, we can find the thermal average position at different time, that is, the leading-lag order. Although the calculation results will fluctuate to some extent when the order changes, the TOP method can still confirm the order well. Coordinate system transformation and iterative format of transfer matrix of TOP are shown in Figure 2.

3.2.2. MIDAS. MIDAS is a method of estimating and forecasting from models where the dependent variable is recorded at a lower frequency than one or more of the independent variables. Unlike the traditional aggregation approach, MIDAS uses information from every observation in the higher frequency space. In this paper, we run MIDAS in Matlab.

The single variable MIDAS model is the simplest and most basic MIDAS model. This model establishes the regression model by using the weight function between the high-frequency data and the low-frequency data and then uses the nonlinear least squares estimation to estimate the parameters, so as to realize the prediction ability of the high-frequency dependent variable to the low-frequency independent variable. The basic form of the univariate MIDAS (m, K) model is as shown in the following equation:

$$y_t = \beta_0 + \beta_1 W(L^{1/m}; \theta)x_t^{(m)} + \varepsilon_t^{(m)}, \quad (5)$$

where y_t is the independent variable of low frequency and $x_t^{(m)}$ is the explained variable of high frequency. m represents the ratio between the high-frequency and low-frequency variable. For example, when y_t is the daily data and $x_t^{(m)}$ is the 60 min data, $m = 24$ when the trading continues throughout the day. $W(L^{1/m}; \theta) = \sum_{k=0}^K w(k; \theta)L^{k/m}$, $L^{1/m}x_t^{(m)} = x_{t-1/m}^{(m)}$, $W(L^{1/m}; \theta)$ is the weighted hysteresis polynomial, $L^{1/m}$ is the high-frequency hysteresis operator, and K is the maximum hysteresis order of the high-frequency variable. Since the lag of this model is the weight lag of high-frequency explanatory variable, it is impossible to make out-of-sample forecast in this process.

In the economic system, the indicator data often has inertia, and there is also an autocorrelation relationship between previous and subsequent data in the time series. Therefore, it is necessary to add an autoregressive term to the mixed frequency data model. Clements and Galvao argue

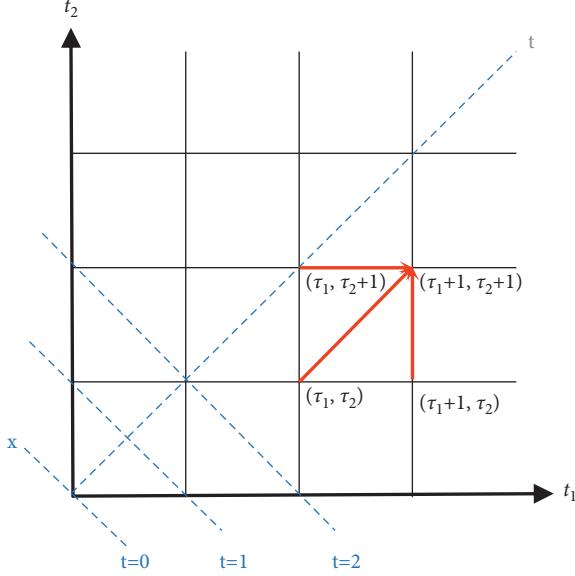


FIGURE 2: Coordinate system transformation and iterative format of transfer matrix of TOP.

that the inertia can only be further addressed by adding dynamic autoregression to the model [46]. In this paper, we add the lag of China's crude oil spot price as an expectation and obtain MIDAS-AR(1) model as in the following equation:

$$y_t = \beta_0 + \lambda y_{t-1} + \beta_1 W(L^{1/m}; \theta) x_t^{(m)} + \varepsilon_t^{(m)}. \quad (6)$$

At present, various parsimonious polynomial specifications $W(L^{1/m}; \theta)$ have been considered, including (1) Almon lag polynomial specifications, (2) exponential Almon lag polynomial specifications, (3) beta polynomial, and (4) step functions [47].

(1) Almon lag polynomial specifications are as follows:

$$w(k; \theta) = \frac{\theta_0 + \theta_1 k + \theta_2 k^2 + \dots + \theta_p k^p}{\sum_{k=0}^p (\theta_0 + \theta_1 k + \theta_2 k^2 + \dots + \theta_p k^p)}. \quad (7)$$

(2) Exponential Almon lag polynomial specifications are as follows:

$$w(k; \theta) = \frac{\exp(\theta_0 + \theta_1 k + \theta_2 k^2 + \dots + \theta_p k^p)}{\sum_{k=0}^p \exp(\theta_0 + \theta_1 k + \theta_2 k^2 + \dots + \theta_p k^p)}. \quad (8)$$

Exponential Almon hysteresis polynomial is the most frequently used weight function nowadays, because it can not only establish various types of weight functions but also ensure that the weight of the given variables is greater than zero, which makes the model have a good property of approximating residual to zero.

(3) Beta polynomial is as follows:

$$w(k; \theta_1; \theta_2) = \frac{f(k/K; \theta_1; \theta_2)}{\sum_{k=1}^K f(k/K; \theta_1; \theta_2)}, \quad (9)$$

where

$$f(x; a; b) = \frac{x^{\alpha-1} (1-x)^{b-1} \Gamma(a+b)}{\Gamma(a) \Gamma(b)}, \quad (10)$$

$$\Gamma(a) = \int_0^\infty e^{-x} x^{a-1} dx.$$

(4) Step functions are as follows:

$$w_i(\theta_1, \dots, \theta_p) = \theta_1 I_{ie}([a_0, a_1]) + \sum_{p=2}^P \theta_p I_{ie}(a_{p-1}, a_p), \quad (11)$$

where

$$a_0 = 1 < a_1 < \dots < a_p = N, \quad (12)$$

$$I_{ie}(a_{p-1}, a_p) = \begin{cases} 1, & a_{p-1} \leq i \leq a_p, \\ 0, & a_{p-1} \geq i, i \geq a_p. \end{cases}$$

4. Empirical Results

4.1. Leading-Lag Relationship. Figure 3 shows the changes in the futures closing price and spot price of the sample period with 890 observations, respectively. The main vertical axis represents the daily closing price of the futures in RMB per barrel, and the secondary vertical axis represents the daily spot price in USD per barrel. Since we are examining the leading-lag relationship of price changes, the effects of exchange rate changes are not considered here. For comparison, we have removed dates that do not overlap. It can be seen from the figure that although there are differences in the trading methods and trading hours between spot and futures, the price trend of the two shows a certain consistency, and the change in futures prices is slightly ahead, indicating that oil futures have a price discovery function. Regarding how many days the futures price changes lead, we will use the TOP method for further analysis.

Figure 4 shows the leading and lagging relationship at three different temperatures ($T = 0.5, 0.8, 1.0$), with positive values on the vertical axis representing futures ahead of the spot [$x(t) > 0$], and vice versa [$x(t) < 0$]. We found that, before the first delivery of the crude oil futures contract SC1809 on September 7, 2018, changes in temperature would cause a significant jump in the lead order, and the change in temperature will cause the leading order to jump significantly. The lower the temperature, the larger the leading order. This is because the optimal path is too sensitive to noise due to low temperature, and the path sensitivity will decrease as the temperature rises, but too high temperature will also cause information loss. The change in the order also shows from the other hand that, before the first delivery, the price of oil futures did not assume the function of price discovery and market stabilization, and its price formation was more affected by speculative factors. After the first delivery, the leading-lag order at different temperatures is basically the same, with only a small span at some time points, which indicates that, from this time on,

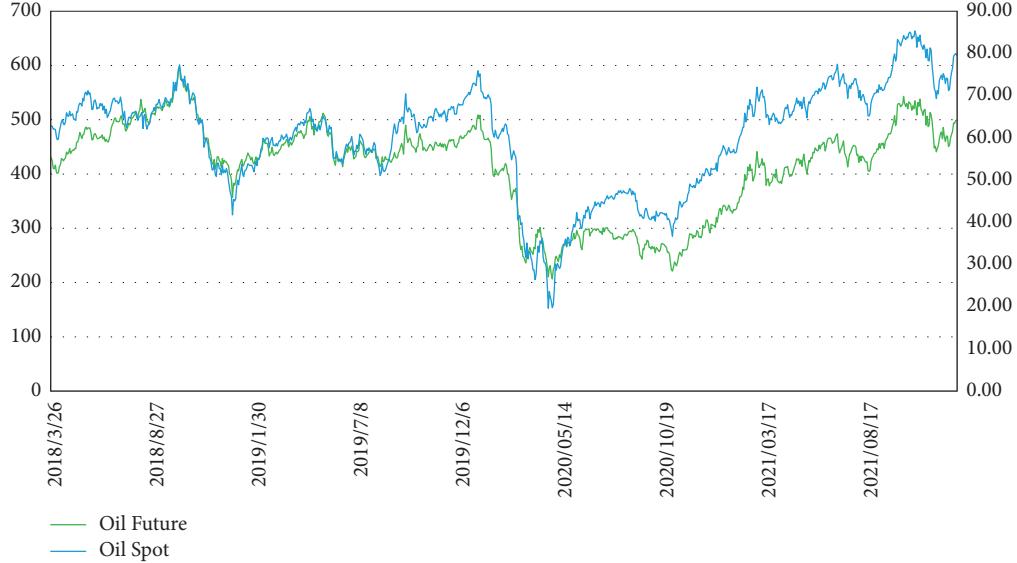


FIGURE 3: Crude oil futures (RMB per barrel) and spot (USD per barrel) from March 26, 2018, to December 31, 2021.

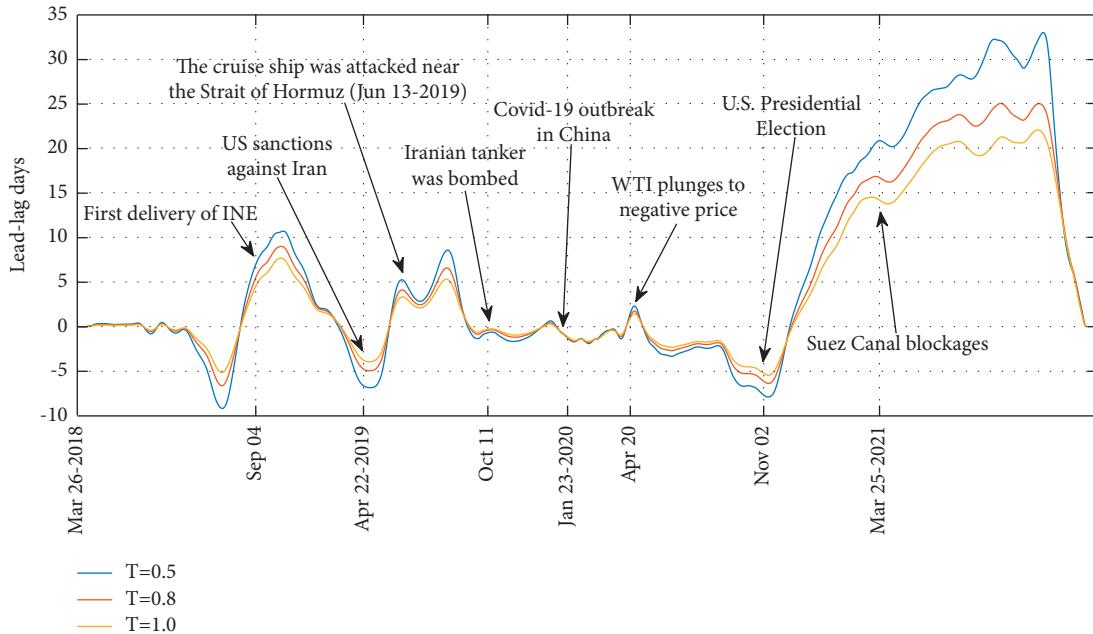


FIGURE 4: Leading-lag relationship between futures and spot.

the futures market trading mechanism is gradually complete and slowly begins to assume its due functions.

In Figure 4, we additionally label the timing of events that may have had a significant impact on the world crude oil market during the sample period. At the first INE settlement date, futures prices are ahead of spot prices by about five trading days. In April 2019, the US announced the continuation of sanctions against Iran, at which point the futures market changes lagged the spot market. In June 2019, tankers were bombed in the Strait of Hormuz, but the market, particularly in China, may have taken such events for granted and did not react violently, as was also seen in the bombing of Saudi crude oil facilities in October of the same

year. In 2020, futures prices moved slightly behind spot prices under the influence of a number of factors such as COVID-19, a trend that continued through the end of the year. The leading relationship between futures and spot prices begins to be significant as market sentiment changes on November 2, 2020, when the US presidential election is held, and is again strengthened by the Suez Canal in 2021. As can be seen from Figure 3, during the sample period, futures lead spot price changes at most of the time, and in a few periods the leading-lag relationship changes due to epidemic factors, but the lag order is less than 10 days at all three temperatures, indicating that, to a certain extent, INE guides domestic oil spot prices better.

4.2. Forecasting Results. In this part, six regression models, Beta-MIDAS, Beta-Nonzero-MIDAS, Exp-Almon-MIDAS, Stepfun-MIDAS, and Almon-MIDAS, were used to simulate the influence and prediction of different high-frequency explanatory variables on daily spot price, and the lag order of high-frequency variable changes from 1 to 40th. The parameter is estimated based on the minimum value of RMSE (root mean square error) and the largest value of R² (goodness of fit) to determine the optimal model. And the deviation degree and deviation matrix are used to measure the prediction accuracy when performing intrasample prediction. The estimation results of the integrated model and the prediction accuracy within the sample determine the optimal weight function form and the optimal lag order of the high-frequency explanatory variables. By comparing the fitting results and prediction accuracy of the model with autoregressive terms and without autoregressive terms, it was found that, after adding autoregressive terms, the model fitting results and the intrasample prediction accuracy were greatly improved. Tables 2, 3, and 4 show the actual and predicted values of the data for different frequencies with night trading before COVID-19, without night trading during the COVID-19 outbreak, and with night trading resumed after the COVID-19 outbreak, respectively. The corresponding line graphs are shown in Figures 5, 6, and 7.

By comparing Figures 5, 6, and 7, we can find that, regardless of whether there is a night trading market, futures prices at different frequencies can predict the spot daily price well. However, the prediction with night market is more accurate in observation. To verify this conjecture, we use the deviation degree and deviation distance to measure the prediction accuracy and the relevant results are shown in Table 5. The deviation degree is a method of measuring the degree of correlation between variables X and Y , and the range of deviation degree is $[-1, 1]$. The greater the absolute value of the deviation, the higher the correlation between X and Y . When they are linearly correlated, the deviation is 1 (positive linear correlation) or -1 (negative linear correlation). Therefore, the prediction sequence with the highest accuracy can be determined according to the maximum absolute value of the deviation degree and the minimum deviation distance. It can be found that, regardless of whether there is a night market, the prediction accuracy of the 60 min data is the highest, indicating that the information contained in the hour data can more accurately predict the price of the day. In addition, in the period of no night market, degree and distance are both lower than those in the period with night market, indicating that the night market also contains valuable information and improves the prediction accuracy. It is also worth noting that, in the third stage of forecasting, the accuracy is weaker than that of the first stage. This phenomenon may be explained by the fact that international crude oil prices produced an unexpectedly large increase in late 2021, which led to a weakening of the forecast accuracy of the futures.

4.3. Discussion of Results. The lead-lag and forecast analysis reveals, first, that the thermal optimal path indicates that the lead-lag relationship between the futures and spot markets

TABLE 2: Actual and forecast price (USD per barrel) in different frequencies with night trading (before February, 2020).

Date	Actual	Forecast 15	Forecast 30	Forecast 60
2019.12.26	70.99	70.72	70.56	70.52
2019.12.27	71.92	71.40	71.31	71.43
2019.12.30	72.84	72.51	72.49	72.44
2019.12.31	72.97	72.88	72.63	72.92
2020.1.2	72.69	71.98	71.97	71.83
2020.1.3	74.21	73.83	74.15	73.60
2020.1.6	75.89	75.59	76.12	75.76
2020.1.7	74.52	75.17	74.50	74.51
2020.1.8	75.17	76.51	75.82	75.44
2020.1.9	71.09	71.44	71.24	71.22
2020.1.10	70.83	70.56	70.51	70.46
2020.1.13	70.48	70.38	70.35	70.38
2020.1.14	69.45	68.53	68.74	69.15
2020.1.15	69.61	69.09	68.90	69.06
2020.1.16	69.22	69.64	69.56	69.29
2020.1.17	69.6	69.20	69.11	69.45
2020.1.20	70.23	69.87	69.75	69.80
2020.1.21	69.47	69.51	69.42	69.57
2020.1.22	69.4	69.25	69.21	69.14
2020.1.23	67.86	67.37	67.50	67.55

TABLE 3: Actual and forecast price (USD per barrel) in different frequencies without night trading (February, 2020–May, 2020).

Date	Actual	Forecast 15	Forecast 30	Forecast 60
2020.4.1	26.43	28.88	28.69	28.36
2020.4.2	27.6	29.58	29.51	30.03
2020.4.3	30.92	29.48	29.63	29.47
2020.4.7	34.34	33.29	32.82	33.87
2020.4.8	32.92	34.63	34.35	34.16
2020.4.9	35.58	34.52	34.32	35.20
2020.4.13	34.98	34.98	34.49	34.77
2020.4.14	36.15	34.05	33.53	34.56
2020.4.15	31.97	34.87	34.42	35.09
2020.4.16	30.51	30.70	30.38	31.08
2020.4.17	30.35	28.99	28.44	28.99
2020.4.20	29.47	29.04	28.55	29.59
2020.4.21	25.42	27.71	27.45	28.21
2020.4.22	19.64	22.68	22.32	22.62
2020.4.23	23.6	21.76	21.61	23.09
2020.4.24	22.31	24.46	24.53	24.39
2020.4.27	21.46	21.31	20.95	20.73
2020.4.28	19.75	20.57	20.26	21.08
2020.4.29	20.58	21.40	21.31	22.38
2020.4.30	24.38	22.55	22.53	22.97

alternates over the sample period, and that there are no instances where either the futures or spot markets dominate throughout the sample period. Moreover, some of the periods with strong lead-lag signals coincide with significant changes in oil and stock markets and geopolitical influences, suggesting that futures markets can reflect real global economic, financial, and geopolitical regime changes, which is consistent with the findings of Shao et al. [48]. Given the current sample period used, it can be assumed that the leading-lag relationship between oil futures and spot markets exists only temporarily and will change in the long run due to shocks from other factors. Second, using the MIDAS

TABLE 4: Actual and forecast price (USD per barrel) in different frequencies without night trading (after May, 2020).

Date	Actual	Forecast 15	Forecast 30	Forecast 60
2021/12/03	71.17	70.3474	70.421	70.3439
2021/12/06	70.41	71.1894	71.1003	71.1911
2021/12/07	73.5	72.447	72.5163	72.4415
2021/12/08	74.32	73.5189	73.7181	73.5195
2021/12/09	74.85	74.3533	74.4607	74.3512
2021/12/10	73.8	74.882	74.5993	74.1794
2021/12/13	75.21	74.8352	74.4897	74.8299
2021/12/14	74.5	75.2404	75.4638	75.239
2021/12/15	72.92	74.524	74.6919	74.5217
2021/12/16	74.17	74.9475	74.0677	74.9442
2021/12/17	74.08	74.1903	74.0489	74.1902
2021/12/20	71.16	71.1096	71.1407	71.1063
2021/12/21	71.66	71.1991	71.2304	71.3323
2021/12/22	73.75	73.6825	73.7492	73.6815
2021/12/23	75.45	73.7697	73.8251	73.7683
2021/12/24	76.52	75.4617	75.401	75.4615
2021/12/28	79.33	79.5291	79.5747	79.5289
2021/12/29	79.7	79.3305	79.317	79.3332
2021/12/30	79.99	79.7051	79.7667	79.7043
2021/12/31	79.62	79.996	79.0398	79.997

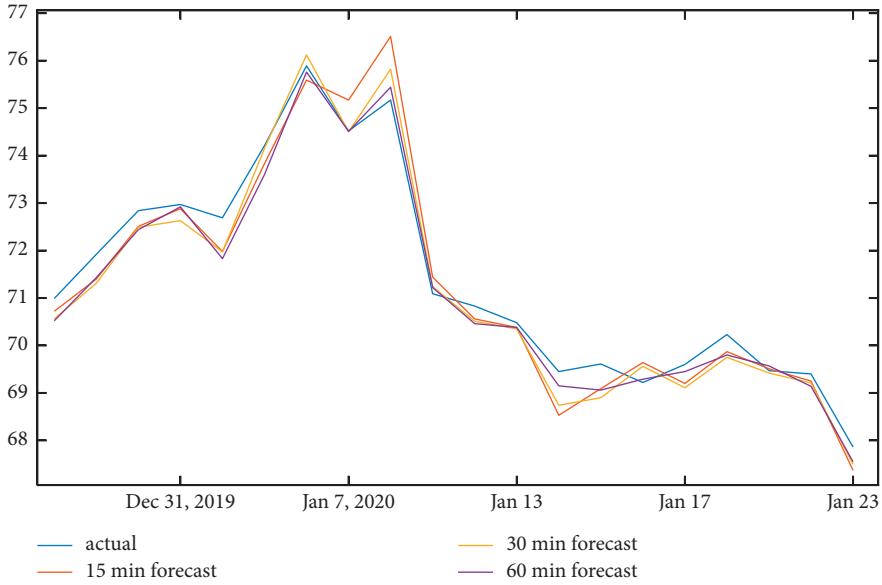


FIGURE 5: Actual and forecast price (USD per barrel) in different frequencies with night trading (before February, 2020).

model, we find that INE plays a role in the discovery of domestic crude oil spot prices, reflecting their dominant influence on the domestic market, which is similar to the findings of wang et al. [49]. Before and after the launch of INE, the relationship between domestic and global oil prices has changed significantly, with the addition of China's own

crude oil futures to what was previously only a hedge against domestic crude oil spot risk through international crude oil futures. Based on the results of the empirical analysis, investors can use 60 min futures data with overnight trading to hedge their investment risk by forecasting spot price movements for 5–10 days in the short term.

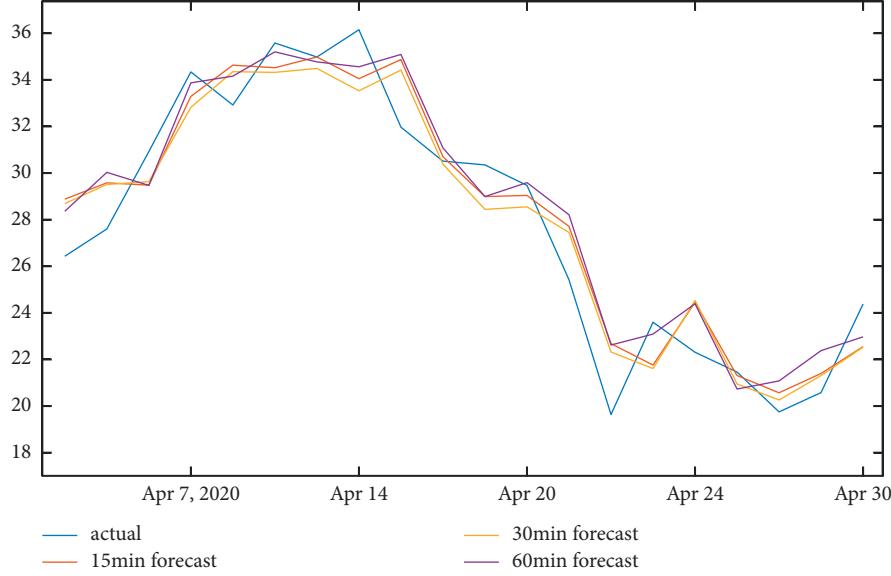


FIGURE 6: Actual and forecast price (USD per barrel) in different frequencies without night trading (February, 2020–May, 2020).

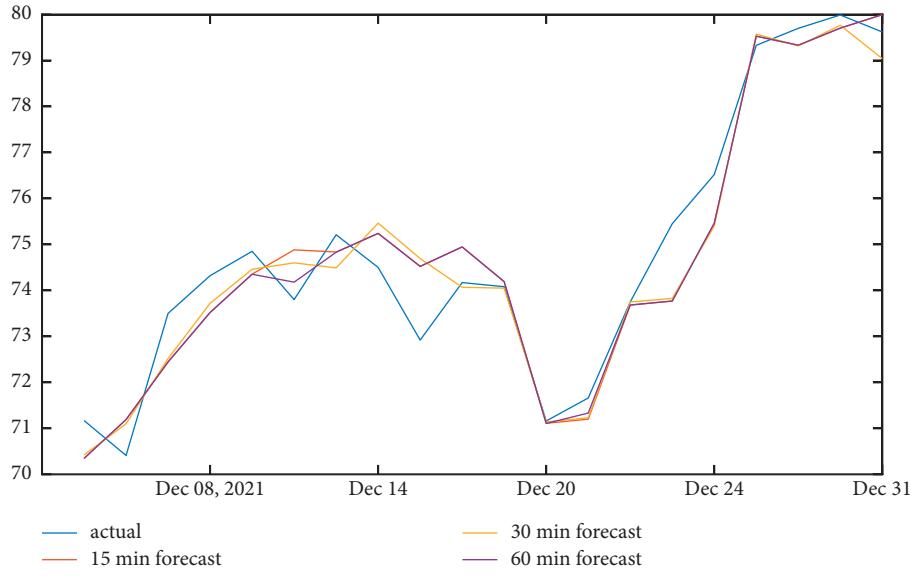


FIGURE 7: Actual and forecast price (USD per barrel) in different frequencies with night trading (after May, 2020).

TABLE 5: Comparison of deviation degree and deviation distance in different periods.

		15 min	30 min	60 min
With night trading (before February, 2020)	Deviation degree	0.9814	0.9902	0.9921
	Deviation distance	0.0186	0.0098	0.0079
No night trading (February, 2020–May, 2020)	Deviation degree	0.9503	0.949	0.9574
	Deviation distance	0.0497	0.051	0.0426
With night trading (after May, 2020)	Deviation degree	0.9610	0.9631	0.9647
	Deviation distance	0.0390	0.0370	0.0352

5. Conclusion

First, China crude oil futures have gradually played a good price guidance role since delivery and have generated immediate reactions to world events affecting crude oil supply.

The extent to which crude oil futures prices are ahead of domestic crude oil spot prices can reach 5–10 days in some time periods, providing an early warning of risk to relevant investors. Chinese crude oil spot pricing has gradually transitioned from only being able to refer to foreign

benchmark crude oil futures to jointly refer to foreign and domestic crude oil futures, providing a more comprehensive pricing basis for the Chinese crude oil market. After the COVID-19 outbreak, the price discovery function of crude oil futures was temporarily missing, and the lead-lag relationship between them and spot prices changed rapidly from leading to lagging, once reaching a lag of half a month, most likely due to the preexisting global industrial production and supply chain system being paralyzed by the impact of the COVID-19 outbreak. According to the results presented in the thermal optimal path chart, the impact of the epidemic on Chinese crude oil futures continues until the end of April 2020, after which Chinese crude oil futures will continue to guide the spot price as the global economy gradually recovers. As crude oil futures and spot prices have continued to climb since November 2020, the predictive function of futures for spot has continued to strengthen. Considering that this upward situation continues, the discussion of it can be left for future studies.

Second, futures trading price data at different frequencies are good predictors of spot daily prices, with the 60-minute data having the highest predictive accuracy, suggesting that higher frequency is not better when using high-frequency markets to predict low-frequency markets. The most likely conjecture for this phenomenon is that too high a frequency will trap a lot of white noise, while too low a frequency will lose some important information. 60-minute data is the most informative data for investors to forecast prices and hedge risks.

Finally, the changes in the price discovery function of futures when only day trading is available are explored with the help of a natural event, the temporary suspension of night trading in crude oil futures. The deviation degree and deviation distance confirm that futures prices are less accurate in predicting spot prices when there is no night trading than futures prices in periods that include night trading, suggesting that night trading in Chinese crude oil futures contains important market information. Further, by comparing the forecast accuracy when night trading was available both before and after COVID-19, the forecast accuracy decreased after the epidemic, probably due to the unanticipated impact of the unforeseen world geopolitical situation on prices, which caused significant price fluctuations and thus affected the forecast accuracy. Chinese crude oil futures can ensure connectivity with overseas markets through the night market and avoid situations such as being unable to close positions due to unexpected conditions that bring significant losses.

This study has important theoretical value and practical significance. On the one hand, the relevant research results can effectively guide relevant enterprises to take advantage of changes in the crude oil futures market and effectively avoid the adverse effects of crude oil price fluctuations. On the other hand, while developing the crude oil futures market and improving the efficiency of the crude oil futures market, the regulatory authorities should pay attention to the price signals emitted by the crude oil futures market and take measures to resolve abnormal situations when they occur, to provide help for better prevention of financial risks.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

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

Authors' Contributions

C. Zhang contributed to formal analysis, investigation, writing of the original draft, visualization, and data curation. D. Pan contributed to investigation and reviewed and edited the manuscript. M. Yang contributed to methodology, software, and validation and reviewed and edited the manuscript. Z. Pu contributed to conceptualization, methodology, investigation, supervision, and funding acquisition and reviewed and edited the manuscript.

Acknowledgments

This work was supported by the National Social Science Foundation, China (no. 20&ZD110), Social Science Fund, Nanjing University of Posts and Telecommunications (no. NYY219004).

References

- [1] K. D. Garbade and W. L. Silber, "Price movements and price discovery in futures and cash markets," *The Review of Economics and Statistics*, vol. 65, no. 2, pp. 289–297, 1983.
- [2] J. H. Min and M. Najand, "A further investigation of the lead-lag relationship between the spot market and stock index futures: early evidence from korea," *Journal of Futures Markets*, vol. 19, no. 2, pp. 217–232, 1999.
- [3] Q. C. Chu, W.-l. G. Hsieh, and Y. Tse, "Price discovery on the s&p 500 index markets: an analysis of spot index, index futures, and spdr's," *International Review of Financial Analysis*, vol. 8, no. 1, pp. 21–34, 1999.
- [4] D. Rittler, "Price discovery and volatility spillovers in the European Union emissions trading scheme: a high-frequency analysis," *Journal of Banking & Finance*, vol. 36, no. 3, pp. 774–785, 2012.
- [5] G. A. Z. Gourène and P. Mendy, "Oil prices and african stock markets co-movement: a time and frequency analysis," *Journal of African Trade*, vol. 5, pp. 55–67, 2018.
- [6] W. Mensi, S. Hammoudeh, S. J. H. Shahzad, and M. Shahbaz, "Modeling systemic risk and dependence structure between oil and stock markets using a variational mode decomposition-based copula method," *Journal of Banking & Finance*, vol. 75, pp. 258–279, 2017.
- [7] Q. Li, K. Cheng, and X. Yang, "Response pattern of stock returns to international oil price shocks: from the perspective of China's oil industrial chain," *Applied Energy*, vol. 185, pp. 1821–1831, 2017.
- [8] E. Elyas, M. Iqbal, and B. Odusami, "Oil price shocks and industry stock returns," *Energy Economics*, vol. 33, 2011.
- [9] G. Filis and I. Chatziantoniou, "Financial and monetary policy responses to oil price shocks: evidence from oil-importing and

- oil-exporting countries," *Review of Quantitative Finance and Accounting*, vol. 42, no. 4, pp. 709–729, 2014.
- [10] L. Yang, Y. Wang, and C. Wu, "Oil price shocks and stock market returns: evidence from oil-importing and oil-exporting countries," *SSRN Electronic Journal*, vol. 41, 2012.
 - [11] W. Mohti, A. Dionísio, I. Vieira, and P. Ferreira, "Regional and global integration of asian stock markets," *Research in International Business and Finance*, vol. 50, pp. 357–368, 2019.
 - [12] R. Li and G. C. K. Leung, "The integration of China into the world crude oil market since 1998," *Energy Policy*, vol. 39, no. 9, pp. 5159–5166, 2011.
 - [13] X. Jia, H. An, W. Fang, X. Sun, and X. Huang, "How do correlations of crude oil prices co-move? A grey correlation-based wavelet perspective," *Energy Economics*, vol. 49, pp. 588–598, 2015.
 - [14] B. Zhang, "Are Chinese and international oil markets integrated?" *International Review of Economics & Finance*, vol. 62, pp. 41–52, 2019.
 - [15] L. Ping, Z. Ziyi, Y. Tianna, and Z. Qingchao, "The relationship among China's fuel oil spot, futures and stock markets," *Finance Research Letters*, vol. 24, pp. 151–262, 2017.
 - [16] Q. Ji, D. Zhang, and Y. Zhao, "Intra-day co-movements of crude oil futures: China and the international benchmarks," *Journal of Operations Research*, vol. 5, pp. 1–27, 2021.
 - [17] L. I. Jie, L. Huang, and L. I. Ping, "Are Chinese crude oil futures good hedging tools?" *Finance Research Letters*, vol. 38, Article ID 101514, 2020.
 - [18] J. Yang and Y. Zhou, "Return and volatility transmission between China's and international crude oil futures markets: a first look," *Journal of Futures Markets*, vol. 40, no. 6, pp. 860–884, 2020.
 - [19] X. Huang and S. Huang, "Identifying the comovement of price between China's and international crude oil futures: a time-frequency perspective," *International Review of Financial Analysis*, vol. 72, Article ID 101562, 2020.
 - [20] F. Palao, Á. Pardo, and M. Roig, "Is the leadership of the brent-wti threatened by China's new crude oil futures market?" *Journal of Asian Economics*, vol. 70, Article ID 101237, 2020.
 - [21] Y. Yang, Y. R. Ma, M. Hu, D. Zhang, and Q. Ji, "Extreme risk spillover between Chinese and global crude oil futures," *Finance Research Letters*, vol. 40, Article ID 101743, 2021.
 - [22] A. Cg, B. Mw, and C. Sbj, "The bubble contagion effect of covid-19 outbreak: evidence from crude oil and gold markets - sciencedirect," *Finance Research Letters*, vol. 38, Article ID 101703, 2020.
 - [23] Y. Sakurai and T. Kurosaki, "How has the relationship between oil and the us stock market changed after the covid-19 crisis?" *Finance Research Letters*, vol. 37, Article ID 101773, 2020.
 - [24] V. V. Xuan and N. T. Hung, "Directional spillover effects and time-frequency nexus between oil, gold and stock markets: evidence from pre and during covid-19 outbreak," *International Review of Financial Analysis*, Article ID 101730, 2021.
 - [25] Y. Liu, C. Zhang, and W. Zhang, "A causal and correlation analysis between China energy futures and China energy-related companies stock market," *Complexity*, vol. 2021, Article ID 3459427, 2021.
 - [26] M. Jefferson, "A crude future? Covid-19's challenges for oil demand, supply and prices," *Energy Research & Social Science*, vol. 68, Article ID 101669, 2020.
 - [27] J. Wang, W. Shao, and J. Kim, "Analysis of the impact of covid-19 on the correlations between crude oil and agricultural futures," *Chaos, Solitons & Fractals*, vol. 136, Article ID 109896, 2020.
 - [28] Q. Wang, S. Li, M. Zhang, and R. Li, "Impact of covid-19 pandemic on oil consumption in the United States: a new estimation approach," *Energy*, vol. 239, Article ID 122280, 2022.
 - [29] P. Zhu, Y. Tang, Y. Wei, and T. Lu, "Multidimensional risk spillovers among crude oil, the us and Chinese stock markets: evidence during the covid-19 epidemic," *Energy*, vol. 231, Article ID 120949, 2021.
 - [30] B. Lin and T. Su, "Do China's macro-financial factors determine the shanghai crude oil futures market?" *International Review of Financial Analysis*, vol. 78, Article ID 101953, 2021.
 - [31] W. Mensi, J. C. Reboredo, and A. Ugolini, "Price-switching spillovers between gold, oil, and stock markets: evidence from the USA and China during the covid-19 pandemic," *Resources Policy*, vol. 73, Article ID 102217, 2021.
 - [32] W.-X. Zhou and D. Sornette, "Non-parametric determination of real-time lag structure between two time series: the "optimal thermal causal path" method with applications to economic data," *Journal of Macroeconomics*, vol. 28, no. 1, pp. 195–224, 2006.
 - [33] D. Sornette and W.-X. Zhou, "Non-parametric determination of real-time lag structure between two time series: the "optimal thermal causal path" method," *Quantitative Finance*, vol. 5, no. 6, pp. 577–591, 2005.
 - [34] G. Kun, W. Xing Zhou, S. W. Cheng, and Sornette, "The us stock market leads the federal funds rate and treasury bond yields," *PLoS One*, vol. 6, 2011.
 - [35] K. Guo and W. X. Zhou, "Economy barometer analysis of China stock market: A dynamic analysis based on the thermal optimal path method," *Journal of Management Sciences in China*, vol. 1, 2012.
 - [36] H.-C. Xu, W.-X. Zhou, and D. Sornette, "Time-dependent lead-lag relationship between the onshore and offshore renminbi exchange rates," *Journal of International Financial Markets, Institutions and Money*, vol. 49, pp. 173–183, 2017.
 - [37] H. Meng, W.-X. Zhou, and D. Sornette, "Symmetric thermal optimal path and time-dependent lead-lag relationship: novel statistical tests and application to UK and us real-estate and monetary policies," *SSRN Electronic Journal*, vol. 17, 2014.
 - [38] E. Ghysels, P. Santa-Clara, and R. Valkanov, "The midas touch: mixed data sampling regression models," *Cirano Working Papers*, vol. 5, pp. 512–517, 2004.
 - [39] C. Emre Alper, S. Fendoglu, and B. Saltoglu, "Midas volatility forecast performance under market stress: evidence from emerging stock market," *Economics Letters*, vol. 117, 2012.
 - [40] M. P. Clements and A. Beatriz Galvão, "Forecasting us output growth using leading indicators: an appraisal using midas models," *Journal of Applied Econometrics*, vol. 24, 2009.
 - [41] M. T. Armesto, M. T. Owyang, and J. Piger, "Measuring the information content of the beige book: a mixed data sampling approach," *Journal of Money, Credit, and Banking*, vol. 41, no. 1, pp. 35–55, 2009.
 - [42] M. Libero and G. Moretti, "Real-time forecasts of inflation: the role of financial variables," *Journal of Forecasting*, vol. 767, 2013.
 - [43] M. Marcellino and C. Schumacher, "Factor MIDAS for nowcasting and forecasting with ragged-edge data: a model comparison for German GDP," *Oxford Bulletin of Economics & Statistics*, vol. 72, no. 4, pp. 518–550, 2010.
 - [44] E. Andreou, E. Ghysels, and A. Kourtellos, "Should macroeconomic forecasters use daily financial data and how?"

- Journal of Business & Economic Statistics*, vol. 31, no. 2, pp. 240–251, 2013.
- [45] J. Bai, E. Ghysels, and J. H. Wright, “State space models and midas regressions,” *Econometric Reviews*, vol. 32, no. 7, pp. 779–813, 2013.
- [46] M. P. Clements and A. B. Galvão, “Macroeconomic forecasting with mixed-frequency data,” *Journal of Business & Economic Statistics*, vol. 26, no. 4, pp. 546–554, 2008.
- [47] G. Eric, S. Arthur, and R. Valkanov, “Midas regressions: further results and new directions,” *Econometric Reviews*, vol. 26, 2007.
- [48] Y.-H. Shao, Y.-H. Yang, H.-L. Shao, and H. E. Stanley, “Time-varying lead-lag structure between the crude oil spot and futures markets,” *Physica A: Statistical Mechanics and Its Applications*, vol. 523, pp. 723–733, 2019.
- [49] J. Wang, S. Qiu, and H. Y. Yick, “The influence of the shanghai crude oil futures on the global and domestic oil markets,” *Energy*, vol. 245, Article ID 123271, 2022.