Petroleum Science 19 (2022) 1433–1442

Contents lists available at ScienceDirect

Petroleum Science

journal homepage: www.keaipublishing.com/en/journals/petroleum-science

Spillover of international crude oil prices on China's refined oil wholesale prices and price forecasting: Daily-frequency data of private enterprises and local refineries

Xun-Zhang Pan [a], Xi-Ran Ma [a], Li-Ning Wang [b, c, *], Ya-Chen Lu [a, b], Jia-Quan Dai [b, c], Xiang Li [d]

[a] School of Economics and Management, China University of Petroleum, Beijing 102249, China
[b] Economics & Technology Research Institute, China National Petroleum Corporation, Beijing 100724, China
[c] Key Laboratory of Oil and Gas Market Simulation and Price Forecasting, China National Petroleum Corporation, Beijing 100724, China
[d] Institue of Energy, Peking University, Beijing 100080, China

ARTICLE INFO

Article history:
Received 3 December 2021
Received in revised form
10 March 2022
Accepted 10 March 2022
Available online 18 March 2022

Edited by Xiu-Qiu Peng

Keywords:
Own-monitored daily-frequency data
Refined oil wholesale prices
Spillover effects
VAR-BEKK-GARCH
PCA-BP neural Network

ABSTRACT

Compared with retail prices of state-owned companies used in almost all existing studies, China's refined oil wholesale prices of private enterprises and local refineries are more affected by the market and better reflect the real supply-demand situation. For the first time, this paper applies own-monitored daily-frequency wholesale prices of China's private enterprises and local refineries during 2013–2020 to derive spillover effects of international crude oil prices on China's refined oil prices through the VAR-BEKK-GARCH (vector autoregression-Baba, Engle, Kraft, and Kroner-generalized autoregressive conditional heteroscedasticity) model, and then tries to forecast wholesale prices through the PCA-BP (principal component analysis-back propagation) neural network model. Results show that international crude oil prices have significant mean spillover and volatility spillover effects on China's refined oil wholesale prices. Changes in crude oil prices are the Granger cause of changes in refined oil wholesale prices. With the improvement of China's oil-pricing mechanism in 2016, the volatility spillover from the international crude oil market to China's refined oil market gradually increases, and the BRENT price variation has an increasing impact on the refined oil wholesale price variation. The PCA-BP model could serve as a candidate tool for forecasting China's refined oil wholesale prices.

1. Introduction

In recent years, with the rapid growth of refining and chemical production capacities, the gap between refined oil production and demand has expanded in China. China's refined oil wholesale market has gradually changed from a seller's market to a buyer's market. The Chinese government is relaxing market access for refined oil products (NDRC, 2013). In this context, refined oil wholesale prices have basically achieved marketization in China. Refined oil uses crude oil as the primary raw material, so its prices are closely related to international crude oil prices. Assessing the spillover of international crude oil prices on China's refined oil prices could provide a reference for decision-makers to understand the relationship of China's oil market with the international market. Refined oil prices are also closely related to the interests of oil companies. Forecasting prices could help China's oil industry improve its production and operation management.

1.1. Literature review

In developed countries, refined oil prices have been marketized and adjusted flexibly. It has been proposed in the literature that the variation of refined oil prices in developed countries is driven by the variation of international crude oil prices, and there is a long-

* Corresponding author. School of Economics and Management, China University of Petroleum, Beijing, 102249, China.
E-mail address: wanglining87@cnpc.com.cn (L.-N. Wang).

https://doi.org/10.1016/j.petsci.2022.03.013

Abbreviations:

ARCH Autoregressive conditional heteroscedasticity
BP Back propagation
CCC-GARCH Constant conditional correlational-generalized autoregressive conditional heteroscedasticity
DCC-GARCH dynamic conditional correlational-generalized autoregressive conditional heteroscedasticity
GARCH Generalized autoregressive conditional heteroscedasticity
KMO Kaiser-Meyer-Olkin
LSTM Long short-term memory
PCA Principal component analysis
PCA-BP Principal component analysis-back propagation
VAR Vector autoregression
VAR-BEKK-GARCH Vector autoregression-Baba, Engle, Kraft, and Kroner-generalized autoregressive conditional heteroscedasticity

term equilibrium relationship between the two prices. For example, through a multivariate Johansen test, Asche et al. (2003) found that changes in crude oil prices would affect refined oil product prices in the Northwest European market in the long term. Gjolberg and Johnsen (1999) established a single-equation model with crude oil prices as the exogenous variable and concluded that crude oil prices determined refined oil prices in the United States (U.S.). Borenstein et al. (1997) established an asymmetric error correction model and found that the data from many developed countries supported the asymmetric relationship between crude oil prices and gasoline prices. Borenstein and Shepard (2002) showed that wholesale gasoline prices had a lagged response to the shock of crude oil prices. Chouinard and Perloff (2007) studied the influences of different factors on gasoline retail and wholesale prices in the U.S. and found that the crude oil price variation was the dominant factor driving the gasoline price variation. Compared with developed countries, China's refined oil market has been affected by governmental policies to a deeper degree and for a longer time. Some Chinese researchers used retail price data to explore the linkage of China's refined oil market with the international crude oil market. For example, Mi (2020) established an error correction model to analyze fluctuations of retail prices of refined oil products in China and international crude oil prices. It was found that crude oil prices had a guiding effect on China's refined oil retail prices. By using a cointegration analysis and an asymmetric error correction model, Cui and Li (2015) found that there was a long-term cointegration relationship between international crude oil prices and China's gasoline retail prices, and changes in crude oil prices preceded changes in gasoline retail prices in time. Han et al. (2017) adopted the VAR model and found that the shock of international crude oil prices had a significantly codirectional impact on China's refined oil retail prices. They concluded that the variation of refined oil retail prices in China followed the variation of crude oil prices in the international market, and China passively adjusted retail prices in response to the fluctuation of international prices.

The literature has mostly applied traditional econometrics or machine learning models (Costa et al., 2021) as the main method for forecasting energy prices including crude oil prices. For example, He et al. (2021) used an improved PCA method, Hsu et al. (2016) used a quadratic sinusoidal trend approach, and de Medeiros et al. (2022) used a mixed data sampling model to forecast crude oil prices, respectively. Di Sanzo (2018) proposed that the GARCH model had good analysis and prediction performance for the memory and dynamic switching mechanism of crude oil volatility. Urolagin et al. (2021) processed data using Markovian and Z-Score transformations, and then used the LSTM neural network to predict WTI (West Texas Intermediate) crude oil prices. Cen and Wang (2019) also applied the LSTM network to the WTI and BRENT crude oil price forecasting. Sun and Lei (2021) applied the empirical mode decomposition to time series, and then used the PCA to determine the input of the BP neural network to predict carbon prices. Liu et al. (2021) showed that the BP neural network had excellent performance in forecasting crude oil prices.

1.2. Contributions of this paper

For China, almost all existing studies in the literature have focused on retail prices whose frequency is 10 working days. Compared with retail prices of state-owned oil companies, wholesale prices of private enterprises and local refineries are more affected by the market and can better reflect the real supply-demand situation in the market. In China, the trading volume of refined oil products from private enterprises and local refineries has maintained rapid growth in recent years, reaching approximately one-third of the national total in 2020 (CNPC, 2021). China has been exploring the reform of oil-pricing mechanism (Chen S et al., 2020). On 13 January 2016, the National Development and Reform Commission (NDRC) of China improved its oil-pricing mechanism with the announcement “refined oil prices should be adjusted normally; when they should rise, they will rise, and when they should fall, they will fall” (www.ndrc.gov.cn/xxgk/zcfb/tz/201601/t20160113_963566.html). Since then, oil price adjustments in China have been sped up, and more entities have entered the market. In addition, in practice, Chinese oil companies are not only concerned about the trend of crude oil prices, but also the trend of refined oil prices, both of which are closely related to their benefits. However, as mentioned above, existing studies of price forecasting have mainly focused on crude oil.

This paper could contribute to the literature in the following aspects:

- For the first time, our own-monitored daily-frequency wholesale prices of China's private enterprises and local refineries will be released. The spillover effects of international crude oil prices on China's refined oil prices from these high-frequency data could better reflect price fluctuations and help decision-makers understand the real relationship between China's refined oil market and the international oil market.
- By dividing the original data into two time series, this paper will further compare the changes in spillover before and after the NDRC announcement, which could help decision-makers understand the effects of their improved mechanism and further promote China's oil-pricing mechanism and policy design.
- This paper will also try the PCA-BP neural network model to forecast China's refined oil wholesale prices. Initiating further studies on refined oil price forecasting could help provide additional price information to assist the production and operation management of Chinese oil companies.

The remainder of this paper is organized as follows. Section 2 describes the methods and data. Section 3 conducts the spillover analysis. Section 4 conducts the price forecasting. Lastly, Section 5 provides conclusions and policy suggestions.

2. Methods and data

2.1. Spillover analysis model

In economics, spillover means that the activity of one object has an effect on another object, which may be in a different market and seem unrelated to it. Spillover generally includes mean spillover and volatility spillover. The mean spillover is a first-order moment which reflects the interaction between markets by changes in means. Empirical studies often use the VAR model (Sims, 1980) to analyze the mean spillover between markets, and could use methods such as the Granger causality analysis (Granger, 1969) and the variance decomposition (Wickens, 1998) to further measure the temporal order of changes and factor contributions, respectively. The volatility spillover is a second-order moment which reflects the interaction between markets by changes in variances. Empirical studies often use the GARCH family models (Bollerslev, 1986), such as BEKK-GARCH, CCC-GARCH, and DCC-GARCH, to examine the volatility spillover between markets, and could use the Wald Test (Sims, 1980a) to verify the volatility spillover. The VAR(p) model is given as Eq. (1), where R_t is a stationary time sequence with a lag order of p, μ is the model constant, $\boldsymbol{\Phi}_j$ is the coefficient matrix, and ε_t is the residual term and normally distributed with a zero mean and constant variance.

$$R_t = \mu + \sum_{j=1}^{p} \boldsymbol{\Phi}_j R_{t-j} + \varepsilon_t \tag{1}$$

There are many nonlinear and random factors in crude oil and refined oil markets (Ouyang et al., 2021). Compared with other multivariate GARCH models, the BEKK-GARCH model (Engle and Kroner, 1995), which requires fewer parameters to be estimated and does not mandate a positive definiteness of the covariance matrix, has been widely used in the study of the volatility of energy finance time series (e.g., Thenmozhi and Maurya, 2020; Zankawah and Stewart, 2020; Zolfaghari et al., 2020). In particular, BEKK-GARCH(1,1) is most commonly used in the literature (Tan et al., 2019; Chen Y et al., 2020; Yu et al., 2020) to avoid the increasing number of estimated parameters. Therefore, this paper will use the daily-frequency price data to run the VAR(p)-BEKK-GARCH(1,1) model, as shown in Eq. (2), to quantify spillover effects of international crude oil (WTI and Brent) prices on China's diesel and gasoline wholesale prices, respectively. GARCH establishes a conditional variance equation for the residual sequence of the mean equation. In Eq. (2), $\boldsymbol{H}_t$ represents the conditional variance-covariance matrix ($h_{i,t}$ provides the variance of the i-th market variable and its conditional covariance with other markets on the t-th day), $\boldsymbol{A}$ is the ARCH coefficient matrix which reflects the volatility clustering, $\boldsymbol{B}$ is the GARCH coefficient matrix which reflects the volatility persistence, $\boldsymbol{C}$ is the lower triangular constant matrix, and $\sigma_{i,t}$ is the standard normal distribution residual.

$$\begin{cases} \varepsilon_{i,t} = \sigma_{i,t} h_{i,t} \\ h_{i,t} = c_i + a_i \varepsilon_{i,t-1}^2 + b_i h_{i,t-1} \\ \boldsymbol{H}_t = \boldsymbol{C}'\boldsymbol{C} + \boldsymbol{A}'(\varepsilon_{t-1}\varepsilon_{t-1}{}')\boldsymbol{A} + \boldsymbol{B}'\boldsymbol{H}_{t-1}\boldsymbol{B} \end{cases} \tag{2}$$

2.2. Price forecasting model

The PCA (Pearson, 1901) is a commonly used statistical method for linearly transforming correlated variables $\boldsymbol{X}=(X_1, X_2, \ldots, X_p)^T$ into uncorrelated composite variables $\boldsymbol{Y}=(Y_1, Y_2, \ldots, Y_n)^T$ ($n < p$; $Y_i = u_{1i}X_1 + u_{2i}X_2 + \ldots + u_{pi}X_p$). $\boldsymbol{Y}$ that satisfies the following conditions is called the principal component of $\boldsymbol{X}$: 1) $u_i{}'u_i = 1$; 2) there is no correlation between Y_i and Y_j; 3) among $X_1, X_2, \ldots, X_p$ linear combinations that are not correlated to $Y_1, Y_2, \ldots, Y_{i-1}$ and satisfy 1), Y_i has the largest variance. More information on the PCA can be found in studies such as Abdi and Williams (2010), He et al. (2021), and Lee and Jemain (2021).

The BP neural network (Rumelhart et al., 1986) is a multilayer feedforward neural network trained according to an error back propagation algorithm, which can learn a range of input-output mapping patterns. The basic structure of the BP network is illustrated in Fig. 1. Before training, the weights between the input layer and the hidden layer, and between the hidden layer and the output layer, are initialized to random values. The learning of the BP neural network is generally divided into two processes: forward propagation and error back propagation. In the forward propagation process, the input information is transmitted to the output layer through the calculation of the hidden layer. If the expected target (e.g., a preset error value) is not achieved, the error change of the output layer is calculated and transmitted back to the input layer, and the weights between layers are adjusted in the direction of error reduction. The above processes are repeated until the target is finally achieved. More information on the BP neural network can be found in studies such as Hecht-Nielsen (1992), Gupta and Nigam (2010), Sun and Huang (2020). In this paper, the PCA and the BP neural network are combined (the PCA-BP model) to try to forecast refined oil wholesale prices in China. Before running the BP, using the PCA to process raw data could help reduce potential deviations caused by autocorrelation and collinearity, and could also help the network avoid the curse of dimensionality and improve calculation efficiency.

2.3. Data and descriptions

This paper uses the daily-frequency wholesale prices of 0# diesel and 92# gasoline of private enterprises and local refineries during the period from 3 January 2013 to 31 December 2020 as China's refined oil prices. The data are obtained from the monitoring of the Key Laboratory of Oil and Gas Market Simulation and Price Forecasting of China National Petroleum Corporation (see Appendix; it's easy to find that refined oil wholesale prices fluctuate more greatly than retail prices in China). The daily WTI and BRENT crude oil spot prices during the same period, obtained from the WIND database (www.wind.com.cn), are used as international crude oil prices. WTI crude oil, the light intermediate base crude oil in U.S. West Texas, is the benchmark for global crude oil pricing. BRENT crude oil is the second largest crude oil pricing benchmark in the world and also plays an important role across the international community. The trends of the four price series are overall consistent (Fig. 2), which intuitively features a potential linkage between China's refined oil wholesale prices and international crude oil prices. At the beginning of 2020, due to COVID-19 and geopolitics, international crude oil prices plummeted, causing negative WTI prices. This paper converts price data to rate of return data for subsequent spillover analysis.

As shown in Table 1, the standard deviation of the return rate of international crude oil prices is much larger than that of China's refined oil wholesale prices, implying that crude oil prices fluctuate greatly and are susceptible to other factors, which is consistent with the findings of Nonejad (2020). The variation of China's refined oil wholesale prices is comparatively small, which is inseparable from China's national regulations on retail prices. China has imposed restrictions on gasoline and diesel retail prices, setting "ceiling prices" and "floor prices" and "when crude oil prices in the international market linked for domestic refined oil are higher than \$130 per barrel, the maximum retail prices of gasoline and diesel will not rise or only rise little; when crude oil prices are lower than \$40 per barrel, the maximum retail prices will no longer be

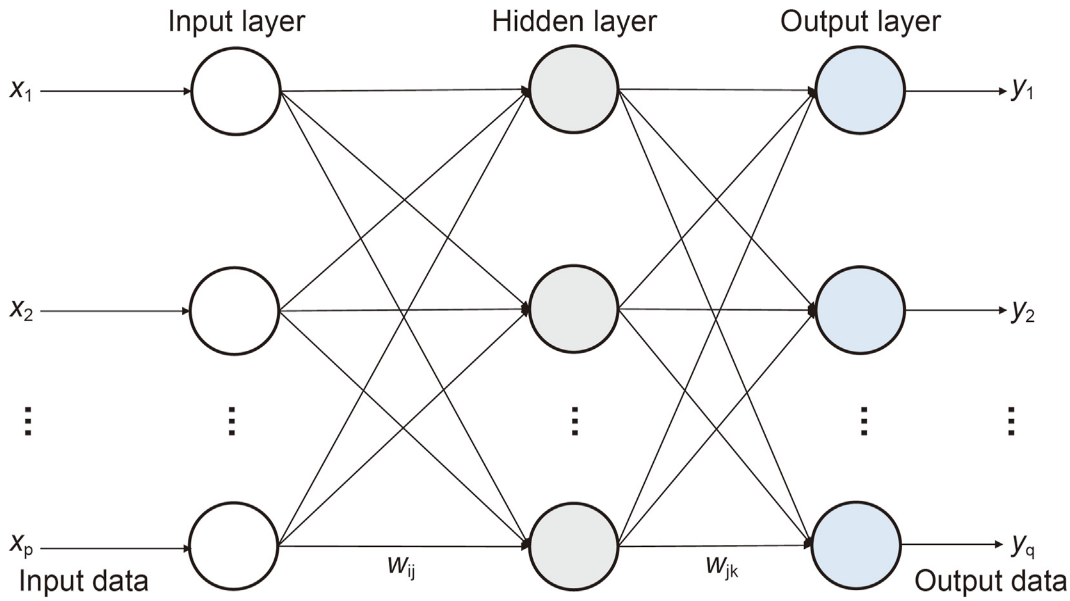

Fig. 1. Schematic diagram of a typical three-layer neural network structure. Notes: there might be several hidden layers in a neural network.

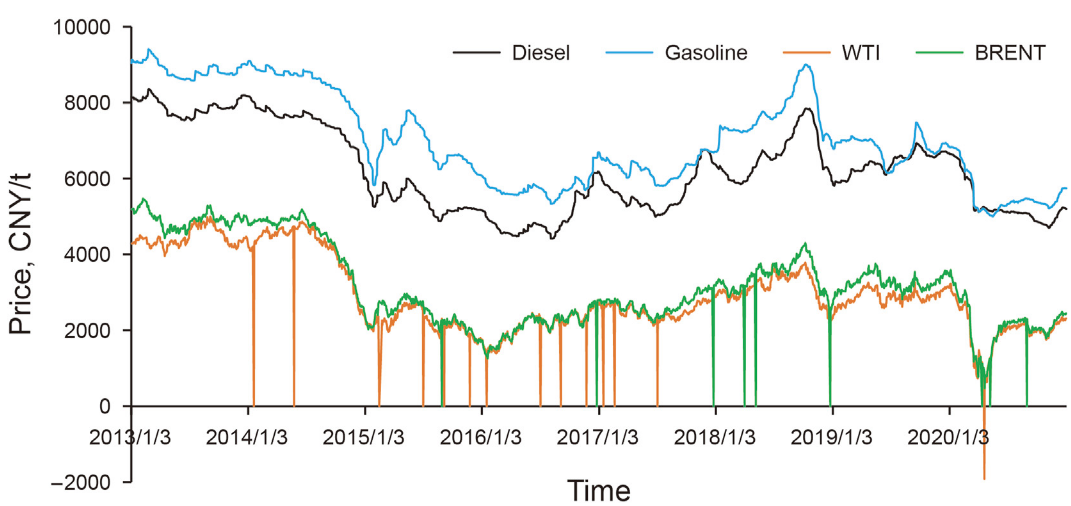

Fig. 2. Variations of China's refined oil wholesale prices and international crude oil prices.

Table 1
Descriptive statistics.

	Diesel	Gasoline	WTI	BRENT
Mean	−0.0022	−0.0083	0.0679	0.0022
Standard Error	0.5740	0.6137	3.2627	2.9462
Minimum	−6.0000	−8.0000	−28.0000	−26.0000
Maximum	5.0000	7.0000	43.0000	41.0000
Skewness	−0.7957	−0.4869	1.8282	1.6272
Kurtosis	35.5273	46.1944	41.2559	42.6676
Jarque-Bera	94749***	159937***	128513***	137181***
ADF	−31.7482***	−31.5961***	−44.2740***	−37.9680***

Notes: *** indicates the 1% significance level. Jarque-Bera is a normality test with the null hypothesis of JB = 0. ADF (augmented Dickey-Fuller) is a stationary test.

adjusted downwards" (www.gov.cn/zwgk/2009-05/08/content_1308543.htm). All time series pass the stationary test at the 1% significance level, indicating that they are stationary.

3. Spillover analysis

3.1. Mean spillover

In this paper, the VAR model is constructed according to a lag order of 8 (the optimal lag order according to the likelihood ratio test, final prediction error, Akaike information criterion, Hannan-Quinn information criterion, and Schwartz information criterion). It is found from Table 2 (the VAR model passes the stationary test) that China's refined oil (both diesel and gasoline) wholesale prices are greatly affected by its own previous values (lag terms), which is consistent with China's unique oil-pricing formation system. Meanwhile, the results of significance show that China's refined oil wholesale prices are also affected by international crude oil prices. It is preliminarily inferred that international crude oil prices have a mean spillover on wholesale prices of refined oil in China.

The Granger causality analysis is further used to evaluate the temporal order of changes in China's refined oil wholesale prices and international crude oil prices, and the variance decomposition

Table 2
Model estimation results.

	Diesel (DO)			Gasoline (GO)	
	Coefficient	T-value		Coefficient	T-value
L1.DO	0.1640***	(0.0234)	L1.GO	0.1810***	(0.0235)
L2.DO	0.0943***	(0.0238)	L2.GO	0.0795***	(0.0240)
L3.DO	0.0578**	(0.0239)	L3.GO	0.0706***	(0.0241)
L4.DO	0.0482**	(0.0240)	L4.GO	0.0477**	(0.0243)
L5.DO	0.0310	(0.0239)	L5.GO	0.0127	(0.0242)
L6.DO	0.0172	(0.0239)	L6.GO	−0.0145	(0.0241)
L7.DO	0.0091	(0.0238)	L7.GO	0.0182	(0.0241)
L8.DO	0.1290***	(0.0233)	L8.GO	0.0954***	(0.0235)
L1.WTI	0.0020	(0.0051)	L1.WTI	0.0027	(0.0055)
L2.WTI	−0.0067	(0.0053)	L2.WTI	−0.0069	(0.0058)
L3.WTI	−0.0084	(0.0055)	L3.WTI	−0.0108*	(0.0059)
L4.WTI	−0.0146***	(0.0056)	L4.WTI	−0.0180***	(0.0061)
L5.WTI	0.0018	(0.0055)	L5.WTI	−0.0008	(0.0060)
L6.WTI	−0.0091*	(0.0055)	L6.WTI	−0.0123**	(0.0060)
L7.WTI	0.0035	(0.0055)	L7.WTI	0.0051	(0.0060)
L8.WTI	0.0040	(0.0052)	L8.WTI	0.0016	(0.0057)
L1.BRENT	0.0196***	(0.0058)	L1.BRENT	0.0171***	(0.0063)
L2.BRENT	0.0223***	(0.0059)	L2.BRENT	0.0231***	(0.0065)
L3.BRENT	0.0172***	(0.0060)	L3.BRENT	0.0198***	(0.0066)
L4.BRENT	0.0215***	(0.0060)	L4.BRENT	0.0270***	(0.0066)
L5.BRENT	0.0030	(0.0061)	L5.BRENT	0.0042	(0.0066)
L6.BRENT	0.0138**	(0.0059)	L6.BRENT	0.0154**	(0.0065)
L7.BRENT	0.0092	(0.0058)	L7.BRENT	0.0140**	(0.0063)
L8.BRENT	0.0035	(0.0056)	L8.BRENT	0.0111*	(0.0061)
Constant	−0.0123	(0.0120)	Constant	−0.0150	(0.0131)

Notes: ***, ** and * indicate the 1%, 5% and 10% significance levels, respectively.

Table 3
Granger causality test results.

H_0	chi2	P-value
WTI is not the Grange cause of diesel	36.706	0.000
BRENT is not the Grange cause of diesel	16.471	0.046
WTI is not the Grange cause of gasoline	44.92	0.000
BRENT is not the Grange cause of gasoline	18.383	0.019

is used to explore the contribution of crude oil price variations to wholesale price variations. Table 3 clearly shows that WTI is the Granger cause of diesel and gasoline at the 1% significance level, and BRENT is the Granger cause of diesel and gasoline at the 5% significance level. This justifies that international crude oil prices are the Granger cause of China's refined oil wholesale prices. Changes in crude oil prices precede and might have a guiding effect on changes in wholesale prices, which is consistent with the findings for retail prices in Mi (2020), and Cui and Li (2015).

Fig. 3 further reflects that China's refined oil market does not respond immediately to external shocks from the international crude oil market (the decomposition result is 0 when the period is 1) and there is a certain time lag. Contributions of WTI and BRENT price variations to China's refined oil wholesale price variations increase gradually over period. For example, the contribution of the WTI (BRENT) price variation to the diesel price variation increases from 0.62% (0.61%) in period 2 to 4.30% (2.83%) in period 10, and increases from 0.47% (0.39%) in period 2 to 4.18% (3.05%) in period 10 for gasoline. Overall, the impact of the WTI price variation on China's refined oil wholesale price variation tends to be slightly greater than that of BRENT.

3.2. Volatility spillover

In Table 4, a_{21} and b_{21} represent the ARCH and GARCH volatility spillover of WTI crude oil prices to China's refined oil wholesale prices, which reflect the clustering and persistence of volatility, respectively. a_{31} and b_{31} correspond to BRENT. It is found that all coefficients except for a_{31} and b_{21} for diesel are significant at the 1% level. This indicates that WTI has a significant ARCH effect on both diesel and gasoline in China, and a significant GARCH effect on gasoline; BRENT has a significant GARCH effect on both diesel and gasoline, and a significant ARCH effect on gasoline. $|a_{21}|>|a_{31}|$ implies that WTI shows a stronger agglomeration shock to China's refined oil market, and $|b_{31}|>|b_{21}|$ implies that BRENT shows a stronger persistent volatility transmission. In summary, international crude oil prices have a significant volatility spillover on China's refined oil wholesale prices.

The Wald test further reflects (Table 5) a significant volatility spillover of WTI and BRENT to diesel and gasoline, verifying that the international crude oil market has an important impact on China's refined oil market.

3.3. Spillover over time

In this section, the original data are divided into two time stages by the NDRC's announcement on 13 January 2016 to explore the effects of the improved oil-pricing mechanism on the linkage of China's refined oil market with the international crude oil market.

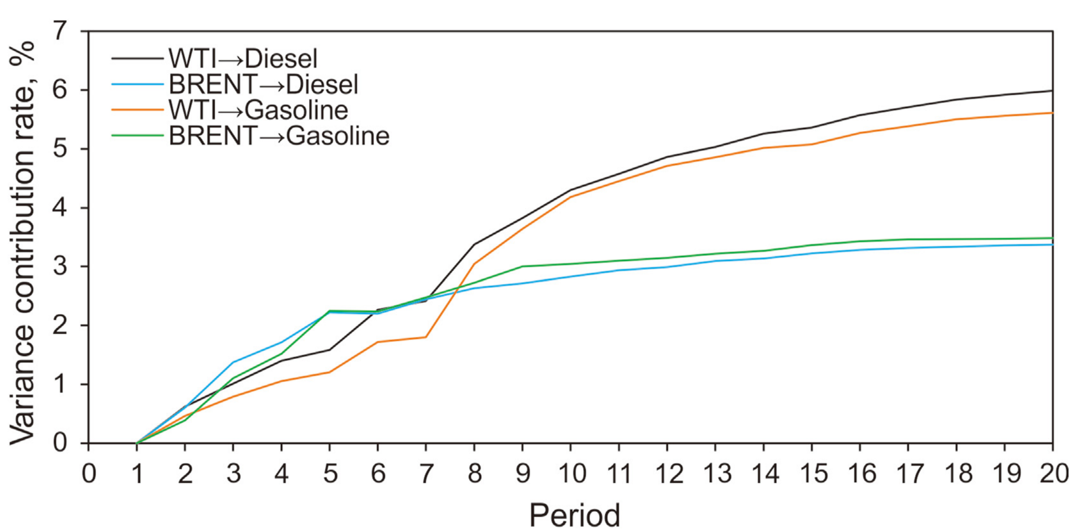

Fig. 3. Variance decomposition results.

Table 4
Volatility spillover for refined oil wholesale market.

	ARCH coefficient		GARCH coefficient	
	a_{21} (WTI)	a_{31} (BRENT)	b_{21} (WTI)	b_{31} (BRENT)
Diesel coefficient estimate	−0.0546***	0.0065	−0.0061	0.0360***
Diesel P-value	0.0000	0.2088	0.1731	0.0000
Gasoline coefficient estimated	−0.0515***	−0.0148***	−0.0349***	0.0611***
Gasoline P-value	0.0000	0.0000	0.0000	0.0000

Notes: *** indicates the 1% significance level.

Table 5
Wald test results.

H_0	Wald test	P-value
WTI $\nrightarrow$ Diesel	136.8728***	0.0000
BRENT $\nrightarrow$ Diesel	11.0390***	0.0040
WTI $\nrightarrow$ Gasoline	454.4114***	0.0000
BRENT $\nrightarrow$ Gasoline	191.9795***	0.0000

Notes: *** indicates the 1% significance level. '$\nrightarrow$'means that the associated ARCH and GARCH coefficients are both equal to 0.

Table 6
Granger causality test results by time stage.

H_0	Stage I		Stage II	
	chi2	P-value	chi2	P-value
WTI is not the Grange cause of diesel	34.994	0.000	43.808	0.002
BRENT is not the Grange cause of diesel	9.9599	0.104	29.817	0.026
WTI is not the Grange cause of gasoline	54.665	0.000	71.728	0.000
BRENT is not the Grange cause of gasoline	7.7171	0.043	33.709	0.009

The VAR models constructed in the two time stages (model details are not shown here due to the space limitation) verify that there is a relationship between China's refined oil wholesale prices and international crude oil prices within the entire time frame studied in this paper. Table 6 further shows that WTI is the Granger cause of diesel and gasoline at the 1% significance level in both two stages; BRENT is the Granger cause of gasoline at the 5% significance level in both two stages, and becomes the Granger cause of diesel at the 5% significance level in the second stage.

By comparing the two-stage variance decomposition results (Fig. 4), it is found that the contribution of the BRENT price variation in the first stage is lower than that in the second stage, reflecting that with the improvement of China's oil-pricing mechanism and the relaxation of the oil market, the impact of the BRENT price variation on the variation of China's refined oil wholesale prices is increasing. The contribution of the WTI price variation for diesel in the second stage is smaller than that in the first stage when the period is greater than 4, indicating that with the mechanism adjustment, the impact of the WTI price variation on China's diesel wholesale price variation is weakening to some extent.

The volatility spillover by time stage is given in Table 7. For diesel, WTI and BRENT have a significant GARCH effect in both stages, indicating that the international crude oil market has already formed a significant persistent volatility transmission to China's diesel wholesale market since 2013. In the second stage, WTI and BRENT also have a significant ARCH effect. For gasoline, in the first stage, WTI and BRENT have a significant ARCH effect but no significant GARCH effect, indicating that the international crude oil market did not form a significant persistent volatility transmission to China's gasoline wholesale market at that time. However, in the second stage, WTI and BRENT are shown to have both significant ARCH and GARCH effects on gasoline, implying that a significant

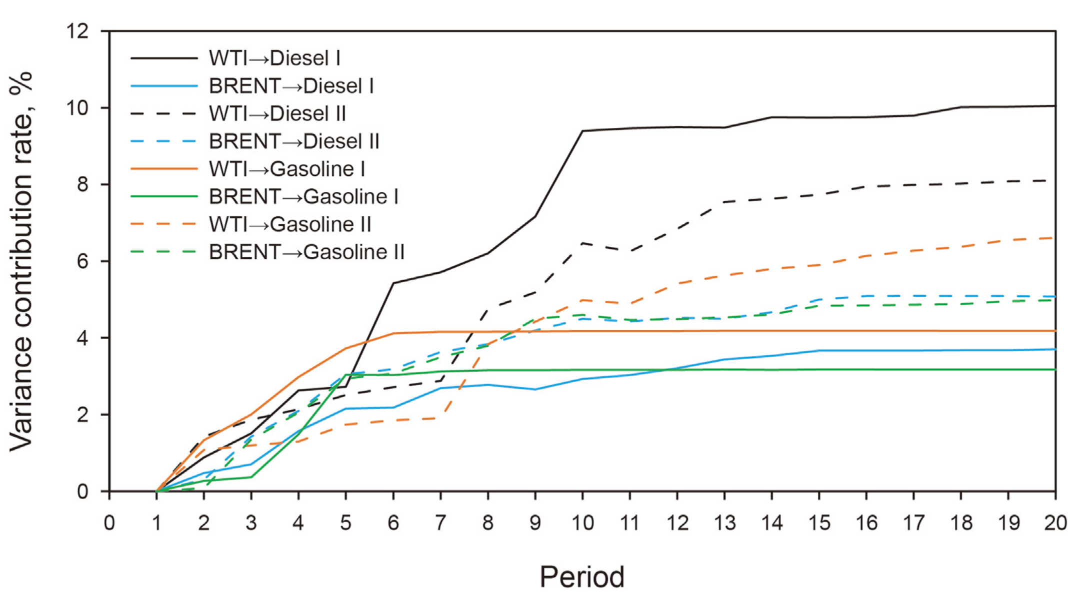

Fig. 4. Variance decomposition in two time stages.
Notes: 'I' indicates the first time stage (before 2016/1/13) and 'II' indicates the second time stage (after 2016/1/13).

Table 7
Volatility spillover for refined oil wholesale market by time stage.

	ARCH coefficient		GARCH coefficient	
	a_{21} (WTI)	a_{31} (BRENT)	b_{21} (WTI)	b_{31} (BRENT)
Diesel coefficient estimate I	−0.0130	0.0021	−0.1042***	0.2338***
Diesel P-value I	0.3982	0.9030	0.0000	0.0000
Diesel coefficient estimate II	−0.0124**	0.0163***	0.0285***	0.0997***
Diesel P-value II	0.0158	0.0066	0.0000	0.0000
Gasoline coefficient estimate I	0.2794***	−0.1912***	0.0012	−0.0007
Gasoline P-value I	0.0000	0.0000	0.8040	0.8484
Gasoline coefficient estimate II	0.0183***	0.0146***	−0.0186***	−0.5196***
Gasoline P-value II	0.0000	0.0014	0.0000	0.0000

Notes: *** and ** indicate the 1% and 5% significance levels, respectively. 'I' indicates the first time stage (before 2016/1/13) and 'II' indicates the second time stage (after 2016/1/13).

persistent transmission has been gradually formed since 2016. In summary, with the improved oil-pricing mechanism, the volatility spillover from the international crude oil market to China's refined oil market gradually increases.

4. Price forecasting

Wholesale prices provide useful information to assist the production and operation of oil companies. On the basic of the literature (e.g., Borenstein and Shepard, 2002; Zhang et al., 2018; Li et al., 2019; Yang et al., 2021), the spillover analysis above, the reality of China's refined oil market, and data availability, the original input variables selected here for forecasting diesel (gasoline) wholesale prices include WTI and BRENT crude oil spot prices, Daqing and Shengli crude oil spot prices in China, diesel (gasoline) productions in China, crude oil productions in China, and consumer price indices in China. Daqing and Shengli crude oil spot prices are obtained from the WIND database. China's crude oil, diesel and gasoline productions and consumer price indices are obtained from the EPS DATA platform (www.epsnet.com.cn/index.html#/Home). Since these data are monthly-frequency, this section provides a monthly forecast for wholesale prices, and uses the average daily prices for each month in Fig. 2. Overall, seven original input variables are used to forecast monthly diesel and gasoline wholesale prices in this section.

4.1. PCA data processing

For diesel-related data, the KMO and Butterlet tests find that the original input variables are correlated, so the PCA processing is performed first. As mentioned earlier, using PCA to preprocess data is helpful to reduce biases caused by potential issues such as data autocorrelation and collinearity. The correlation coefficient matrix of the original variables is calculated to derive the total variance explained table (Table 8). According to the principle that the cumulative percentage of variance is greater than 80%, the first three components are extracted as the principal components for diesel.

These three principal components are linear combinations of the original seven input variables, as given in Eq. (3). For gasoline-related data, two principal components are extracted following the same steps, as given in Eq. (4). In both equations, X_1 to X_7 represent WTI spot prices, BRENT spot prices, Daqing spot prices, Shengli spot prices, diesel (gasoline) productions, crude oil productions, and consumer price indices, respectively. Note that these components are based on monthly data between 2013 and 2020. Future applications could dynamically update according to prediction accuracy requirements and the latest available data.

$$\begin{cases} Y_1 = 0.487X_1 + 0.491X_2 + 0.493X_3 + 0.497X_4 - 0.142X_5 + 0.096X_6 + 0.062X_7 \\ Y_2 = 0.004X_1 - 0.015X_2 - 0.029X_3 - 0.072X_4 + 0.868X_5 + 0.472X_6 + 0.131X_7 \\ Y_3 = 0.024X_1 + 0.027X_2 - 0.003X_3 + 0.012X_4 + 0.168X_5 - 0.446X_6 + 0.878X_7 \end{cases} \tag{3}$$

$$\begin{cases} Y_1 = 0.485X_1 + 0.489X_2 + 0.483X_3 + 0.487X_4 - 0.204X_5 + 0.101X_6 + 0.065X_7 \\ Y_2 = 0.076X_1 + 0.093X_2 + 0.067X_3 + 0.093X_4 + 0.400X_5 - 0.629X_6 + 0.645X_7 \end{cases} \tag{4}$$

4.2. BP neural network forecasting

In this paper, a three-layer BP neural network is applied with an input layer, a hidden layer and an output layer, and the optimal number of nodes in the hidden layer is determined through debugging. Of the monthly samples between January 2013 and December 2020, 70% are used as the training set, 15% are used as the validation set, and the remaining 15% are used as the test set. The training target for error is set to 0.5%, the threshold of learning rate is set to 1%, the minimum allowable gradient value in training is set to 1×10^{-6}, and the maximum number of training times is set to 10,000. As long as one of these settings is reached, the training will stop.

For diesel, the input of the BP network is the three principal components identified in Eq. (3), and the output is predicted diesel wholesale prices. Forecasting results are illustrated in Fig. 5a–d. It is shown that predicted values of the training set are highly consistent with actual prices (the R-value reaches 0.97), indicating that the model has been well trained. Predicted values of the validation set are also consistent with actual prices, indicating that the model learned by the training set is reliable. The difference between predicted values of the test set and actual prices are small (the R-value reaches 0.99), further suggesting that the trained and validated BP neural network has a good capability to forecast China's diesel wholesale prices. For all sample data, predicted values can be considered to capture the fluctuation and magnitude

Table 8
Total variance explained for refined oil wholesale prices.

Component		Initial eigenvalues			Extract sums of squared loadings		
		Total	% of variance	Cumulative % of variance	Total	% of variance	Cumulative % of variance
Diesel	1	4.100	58.578	58.578	4.100	58.578	58.578
	2	1.168	16.690	75.268	1.168	16.690	75.268
	3	1.059	15.131	90.399	1.059	15.131	90.399
	4	0.642	9.178	99.577			
	5	0.021	0.303	99.880			
	6	0.006	0.082	99.962			
	7	0.003	0.038	100.000			
Gasoline	1	4.629	66.128	66.128	4.629	66.128	66.128
	2	1.257	17.958	84.086	1.257	17.958	84.086
	3	0.804	11.483	95.569			
	4	0.279	3.980	99.549			
	5	0.024	0.337	99.885			
	6	0.006	0.083	99.968			
	7	0.002	0.032	100.000			

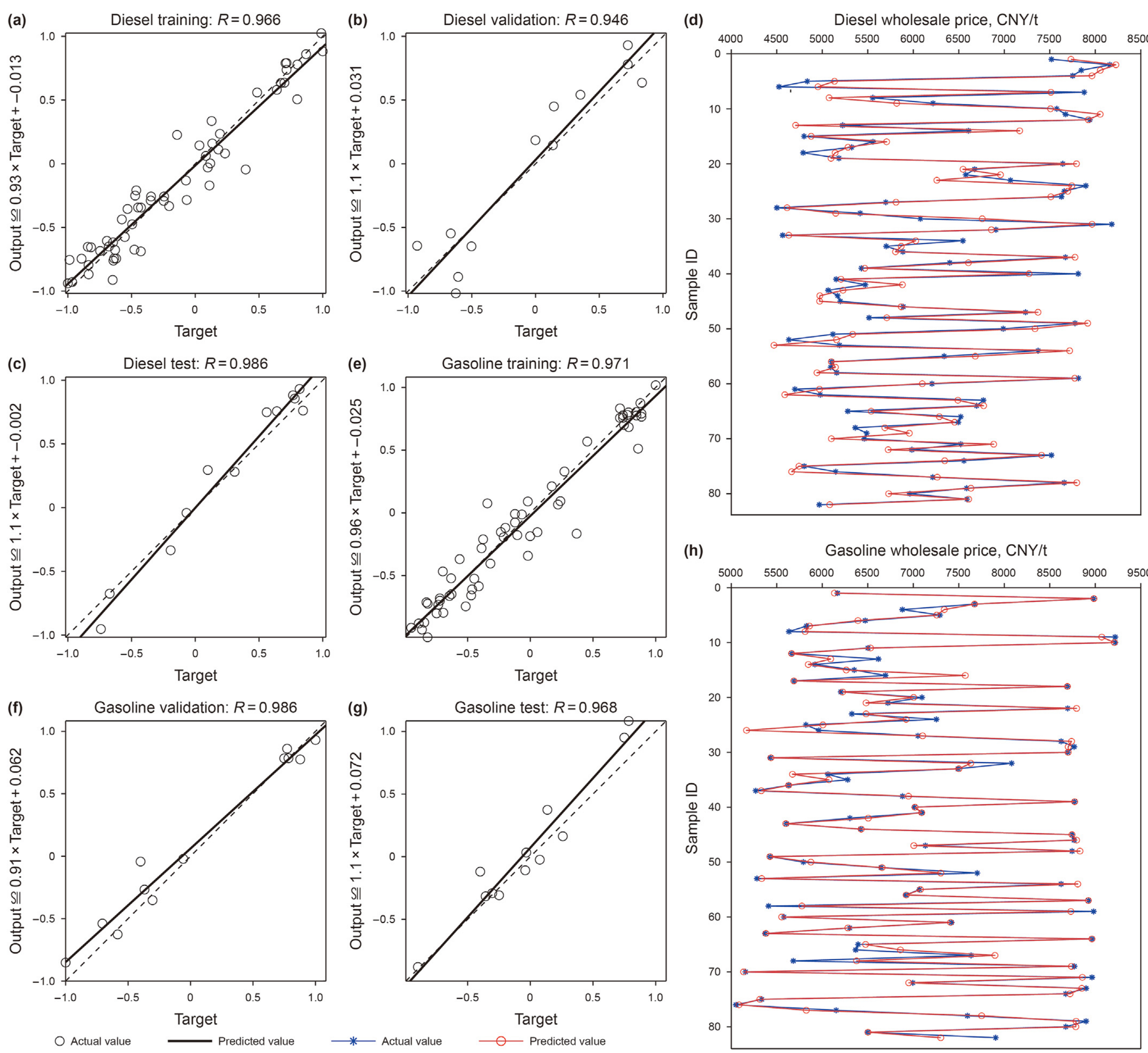

Fig. 5. Comparison of predicted values (Output) and actual values (Target) of wholesale prices.

of actual prices (the average error is only 3.92%). For gasoline, the input is the two principal components identified in Eq. (4), and the output is predicted gasoline wholesale prices. As presented in Fig. 5e–h, the R-values of the training set, verification set, and test set all exceed 0.96. Predicted values and their trend are highly consistent with actual prices (the average error is only 2.58%), suggesting that the trained and validated BP neural network also has a good capability to forecast gasoline wholesale prices. In summary, the PCA-BP neural network model is shown to be an applicable approach to forecasting China's refined oil wholesale prices.

5. Conclusions and policy suggestions

In this paper, for the first time, China's daily-frequency refined oil prices were contributed to the literature. The results obtained from the VAR-BEKK-GARCH model justified that international crude oil prices had significant mean spillover and volatility spillover effects on China's refined oil prices. With the improvement of China's oil-pricing mechanism, the volatility spillover from the international crude oil market was further enhanced. The PCA-BP model was competent for providing a reference for China's refined oil wholesale prices.

Oil will continue to play an important role in China's economy in the coming several decades (Pan et al., 2017, 2020). In the wave of marketization of refined oil prices, some suggestions can be made here. First, China could continue to improve its oil-pricing mechanism. The government could try to further shorten the oil price adjustment period and relax restrictions on the price adjustment range, which could promote a closer linkage of China's refined oil prices with international crude oil prices. Second, the government might also consider issuing refined oil futures. Refined oil futures could further increase the vitality of China's oil market, enable oil prices to accurately reflect the real supply-demand relationship, and drive the reform of the spot market. Third, the government could promote reasonable and orderly competition in the market, and gradually shift from a regulator to a supervisor, so that China's refined oil market can be steadily integrated with the international market and eventually fully marketized.

There are limitations in this paper. For example, due to the time difference between the large-scale outbreak of COVID-19 in China and across the world, this paper didn't assess the impact of COVID-19. Due to China's regulations on retail prices, the global outbreak of COVID-19 would likely have a greater impact on international crude oil prices than on China's refined oil prices. Although novel daily-frequency data of China's refined oil prices were contributed, the models used in this paper (while competent for research questions here) were not innovative. Future studies could consider developing improved or more advanced models to assess these new data or forecast prices.

Acknowledgements

The authors are grateful to the financial support from the Science Foundation of China University of Petroleum, Beijing (2462020YXZZ038).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.petsci.2022.03.013.

References

Abdi, H., Williams, L.J., 2010. Principal component analysis. WIREs Comput. Sci. 2 (4), 433–459. https://doi.org/10.1002/wics.101.
Asche, F., Gjølberg, O., Völker, T., 2003. Price relationships in the petroleum market: an analysis of crude oil and refined product prices. Energy Econ 25 (3), 289–301. https://doi.org/10.1016/S0140-9883(02)00110-X.
Bollerslev, T., 1986. Generalized autoregressive conditional heteroscedasticity. J. Econom. 31 (3), 307–327. https://doi.org/10.1016/0304-4076(86)90063-1.
Borenstein, S., Cameron, A.C., Gilbert, R., 1997. Do gasoline prices respond asymmetrically to crude oil price changes? Q. J. Econ. 112 (1), 305–339. https://doi.org/10.1162/003355397555118.
Borenstein, S., Shepard, A., 2002. Sticky prices, inventories, and market power in wholesale gasoline markets. Rand J. Econ. 33 (1), 116–139. https://doi.org/10.2307/2696378.
Cen, Z., Wang, J., 2019. Crude oil price prediction model with long short term memory deep learning based on prior knowledge data transfer. Energy 169, 160–171. https://doi.org/10.1016/j.energy.2018.12.016.
Chen, S., Zhang, Q., Mclellan, B., et al., 2020. Review on the petroleum market in China: history, challenges and prospects. Petrol. Sci. 17, 1779–1794. https://doi.org/10.1007/s12182-020-00501-6.
Chen, Y., Zheng, B., Qu, F., 2020. Modeling the nexus of crude oil, new energy and rare earth in China: an asymmetric VAR-BEKK (DCC)-GARCH approach. Resour. Pol. 65, 101545. https://doi.org/10.1016/j.resourpol.2019.101545.
Chouinard, H.H., Perloff, J.M., 2007. Gasoline price differences: taxes, pollution regulations, mergers, market power, and market conditions. BE J. Econ. Anal. Pol. 7 (1), 8. https://doi.org/10.2202/1935-1682.1599.
CNPC (China National Petroleum Corporation), 2021. 2020 Domestic and International Oil and Gas Development Report. Beijing. http://center.cnpc.com.cn/sysb/system/2021/04/22/030030794.shtml.
Costa, A.B.R., Ferreira, P.C.G., Gaglianone, W.P., et al., 2021. Machine learning and oil price point and density forecasting. Energy Econ 102, 105494. https://doi.org/10.1016/j.eneco.2021.105494.
Cui, Q., Li, H., 2015. Research on dynamic conduction effect between domestic gasoline prices and international crude oil prices: based on co-integration and non-symmetric ECM. Rev. Ind. Econ. 6, 104–111 (**in Chinese**).
De Medeiros, R.K., da Nóbrega Besarria, C., de Jesus, D.P., et al., 2022. Forecasting oil prices: new approaches. Energy 238, 121968. https://doi.org/10.1016/j.energy.2021.121968.
Di Sanzo, S., 2018. A Markov switching long memory model of crude oil price return volatility. Energy Econ 74, 351–359. https://doi.org/10.1017/S0266466600009063.
Engle, R., Kroner, K., 1995. Multivariate simultaneous generalized. ARCH. Econom. Theory. 11 (1), 122–150. https://doi.org/10.1016/S0140-9883(99)00023-7.
Gjolberg, O., Johnsen, T., 1999. Risk management in the oil industry: can information on long-run equilibrium prices be utilized? Energy Econ 21 (6), 517–527. https://doi.org/10.1016/S0140-9883(99)00023-7.
Granger, C.W.J., 1969. Investigating causal relations by econometric models and cross-spectral methods. Econometrica 37 (3), 424–438. https://doi.org/10.2307/1912791.
Gupta, N., Nigam, S., 2010. Crude oil price prediction using artificial neural network. Procedia Comput. Sci. 170, 642–647. https://doi.org/10.1016/j.procs.2020.03.136.
Han, S., Zhang, B., Tang, X., et al., 2017. The relationship between international crude oil prices and China's refined oil prices based on a structural VAR model. Petrol. Sci. 14, 228–235. https://doi.org/10.1007/s12182-016-0139-9.
He, M., Zhang, Y., Wen, D., et al., 2021. Forecasting crude oil prices: a scaled PCA approach. Energy Econ 97, 105189. https://doi.org/10.1016/j.eneco.2021.105189.
Hecht-Nielsen, R., 1992. III.3-Theory of the backpropagation neural network. Neural Network. Rercept 1, 65–93. https://doi.org/10.1016/B978-0-12-741252-8.50010-8.
Hsu, T.-K., Tsai, C.-C., Cheng, K., 2016. Forecast of 2013-2025 crude oil prices: quadratic sine-curve trend model application. Energy Sources Part B 11 (3), 205–211. https://doi.org/10.1080/15567249.2013.797053.
Lee, L.C., Jemain, A.A., 2021. On overview of PCA application strategy in processing high dimensionality forensic data. Microchem. J. 169, 106608. https://doi.org/10.1016/j.microc.2021.106608.
Li, J., Zhu, S., Wu, Q., 2019. Monthly crude oil spot price forecasting using variational mode decomposition. Energy Econ 83, 240–253. https://doi.org/10.1016/j.eneco.2019.07.009.
Liu, D., Liang, Y., Zhang, L., et al., 2021. Implied volatility forecast and option trading strategy. Int. Rev. Econ. Finance 71, 943–954. https://doi.org/10.1016/j.iref.2020.10.023.
Mi, J., 2020. An empirical analysis of the impact of international crude oil prices on domestic retail prices of refined oil. China's Prices 11, 92–94+98 (**in Chinese**).
NDRC (National Development and Reform Commission), 2013. Notice on Further Improving the Price Formation Mechanism of Refined Oil. Beijing. http://www.gov.cn/govweb/zwgk/2013-03/27/content_2363601.h.
Nonejad, N., 2020. Crude oil price volatility and equity return predictability: a comparative out-of-sample study. Int. Rev. Financ. Anal. 71, 101521. https://doi.org/10.1016/j.irfa.2020.101521.
Ouyang, Z., Qin, Z., Cao, H., et al., 2021. A spillover network analysis of the global crude oil market: evidence from the post-financial crisis era. Petrol. Sci. 18, 1256–1269. https://doi.org/10.1016/j.petsci.2021.05.003.
Pan, X., Chen, W., Clarke, L.E., et al., 2017. China's energy system transformation towards the 2°C goal: implications of different effort-sharing principles. Energy Pol. 103, 116–126. https://doi.org/10.1016/j.enpol.2017.01.020.
Pan, X., Wang, L., Dai, J., et al., 2020. Analysis of China's oil and gas consumption

under different scenarios toward 2050: an integrated modeling. Energy 195, 116991. https://doi.org/10.1016/j.energy.2020.116991.
Pearson, K., 1901. LIII. On lines and planes of closest fit to systems of points in space. London, Edinburgh, and Dublin. Philos. Mag. J. Sci. 2 (11), 559–572. https://doi.org/10.1080/14786440109462720.
Rumelhar, D.E., Hinton, G., Williams, R.J., 1986. Learning representations by back-propagating errors. Nature 323, 533–536. https://doi.org/10.1038/323533a0.
Sims, C.A., 1980. Macroeconomics and reality. Econometrica 48 (1), 1–47. https://doi.org/10.2307/1912017.
Sims, C.A., 1980a. Comparison of interwar and postwar business cycles: monetarism reconsidered. Am. Econ. Rev. 70 (2), 250–257. https://doi.org/10.3386/w0430.
Sun, W., Huang, C., 2020. A carbon price prediction model based on secondary decomposition algorithm and optimized back propagation neural network. J. Clean. Prod. 243, 118671. https://doi.org/10.1016/j.jclepro.2019.118671.
Sun, X., Lei, Y., 2021. Research on financial early warning of mining listed companies based on BP neural network model. Resour. Pol. 73, 102223. https://doi.org/10.1016/j.resourpol.2021.102223.
Tan, X., Zhang, H., Yang, N., et al., 2019. Transmission mechanism and spillover effects between RMB' sonshore and offshore market: empirical analysis based on VAR-GARCH-BEKK model. J. Manag. Sci. China 22 (7), 53–65 (**in Chinese**).
Thenmozhi, M., Maurya, S., 2020. Crude oil volatility transmission across food commodity markets: a multivariate BEKK-GARCH approach. J. Emerg. Mark. Finance 20 (2), 131–164. https://doi.org/10.1177/0972652720927623.
Urolagin, S., Sharma, N., Datta, T.K., 2021. A combined architecture of multivariate LSTM with Mahalanobis and Z-Score transformations for oil price forecasting. Energy 231, 120963. https://doi.org/10.1016/j.energy.2021.120963.
Wickens, T.D., 1998. Categorical data analysis. Annu. Rev. Psychol. 49, 537–558. https://doi.org/10.1146/annurev.psych.49.1.537.
Yang, Y., Guo, J., Sun, S., et al., 2021. Forecasting crude oil price with a new hybrid approach and multi-source data. Eng. Appl. Artif. Intell. 101, 104217. https://doi.org/10.1016/j.engappai.2021.104217.
Yu, L., Zha, R., Stafylas, D., et al., 2020. Dependences and volatility spillovers between the oil and stock markets: new evidence from the copula and VAR-BEKK-GARCH models. Int. Rev. Financ. Anal. 68, 101280. https://doi.org/10.1016/j.irfa.2018.11.007.
Zankawah, M.M., Stewart, C., 2020. Measuring the volatility spill-over effects of crude oil prices on the exchange rate and stock market in Ghana. J. Int. Trade Econ. Dev. 29 (4), 420–439. https://doi.org/10.1080/09638199.2019.1692895.
Zhang, H., Liu, C., Zhao, M., et al., 2018. Economics, fundamentals, technology, finance, speculation and geopolitics of crude oil prices: an econometric analysis and forecast based on data from 1990 to 2017. Petrol. Sci. 15, 432–450. https://doi.org/10.1007/s12182-018-0228-z.
Zolfaghari, M., Ghoddusi, H., Faghihian, F., 2020. Volatility spillovers for energy prices: a diagonal BEKK approach. Energy Econ 92, 104965. https://doi.org/10.1016/j.eneco.2020.104965.