

Research on the Prediction of the Global Price of WTI Crude

Fei Xie*

School of Accounting, The Australian National University, Canberra, Australia

*Corresponding author: xf2108314@ldy.edu.rs

Abstract. As the cornerstone of industrial development, crude oil prices have attracted international attention. However, due to the frequency of price fluctuations and the sensitivity to the international economic situation, crude oil prices are difficult to predict. Due to significant differences in prices before and after 2007, this will increase the difficulty of prediction and decrease its accuracy. The ARIMA model, as a typical prediction model, is highly favored by scholars from various countries due to its accuracy and adaptability. This article will use the ARIMA model to forecast the one-year future prices. The conclusion is as follows: in the two months, the prices will have a slight upward trend, while in the next year, the overall trend will show a relatively obvious downward trend. Therefore, this article believes that it is not recommended to invest in crude oil in the short term. At the same time, this article believes that ensuring domestic economic stability can reduce the impact of international economic shocks and thus gain an advantageous position in international trade.

Keywords: Prediction; WTI crude; ARIMA model.

1. Introduction

From the perspective of the international situation, the world today is in a situation of great change that has not been seen in a century. COVID-19 and the situation in Russia and Ukraine have all made the price of crude oil fluctuate frequently, further increasing the uncertainty of its price changes. A good international economic environment requires a relatively stable price. Therefore, how to grasp the trend of price fluctuations and accurately predict it has become a focus of research in various countries.

As the cornerstone of industrial development, the changed price will have a profound impact on the global political and economic situation. Zhang stated that there is a strong connection between prices and stock revenue in China's traditional energy industry [1]. In addition, Irwin et al. believe that fluctuations in the price of crude oil have the potential to impact other commodity prices, which in turn can impact the actual economy's growth [2]. What is more, the price will also be affected by various factors. Shaikh found through the crude oil investor sentiment index that during the COVID-19 outbreak, the epidemic increased investor panic and anxiety, thereby damaging the global crude oil market [3]. Pellejero pointed out that due to the increase in US crude oil inventories, the future upward trend of prices may be hindered [4].

The fundamental purpose of analyzing the correlation between many variables and the price is to make the predicted price more accurate and reasonable. In the research process of many scholars, the price has been subjected to several models. Kadhem and Thajel used the hidden Markov process to analyze and model prices and found that the price movements show significant swings during various moments of market turbulence [5]. Pan and Li used the Long Short-Term Memory (LSTM) method to apply it to short-term price prediction and the finding is that using daily data from the first 24 months to predict daily data from the next 3 months was the best [6]. At the same time, more and more scholars have found that a single model is not ideal for predicting oil prices in most cases, and they are more willing to use more complex combination models to predict oil prices. Pan utilized an enhanced Fruit Fly Optimization Algorithm (FOA) model to forecast prices and found that the prediction model combining the Modified Fruit Fly Optimization Algorithm (MFOA) and Diagonal Recurrent Neural Networks (DRNN) possessed the greatest capacity to predict closing oil and gold prices [7]. Wu et al. used a combination of Ensemble Empirical Mode Decomposition (EEMD) and LSTM to predict and found that the model can still predict prices accurately despite varying decomposition outcomes [8]. Huang et al. creatively used Empirical Mode Decomposition (EMD),

which breaks down the initial information into a few Intrinsic Mode Functions (IMF) and rest components at different frequencies [9]. The Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) model used by María et al. can avoid modal confusion while excluding residual noise in the IMF [10].

In summary, the research on the important data of oil prices has attracted numerous scholars. This article will mainly use the ARIMA model to predict and analyze oil prices and provide corresponding suggestions to investors based on the predicted results.

2. Methods

2.1. Data Sources

The data is taken from Fred's global price of WTI crude. This data is the monthly average of each barrel of crude oil calculated in US dollars, with a total of 405 observations from January 1990 to September 2023.

2.2. Variable Selection

Crude oil prices are the result of changes in international economic development, reflecting the overall trend of global development and being influenced by major global events. The occurrence of major events is irregular and unpredictable, therefore, changes in prices can be very frequent, and difficult to determine the magnitude of the changes, as illustrated in Fig. 1:

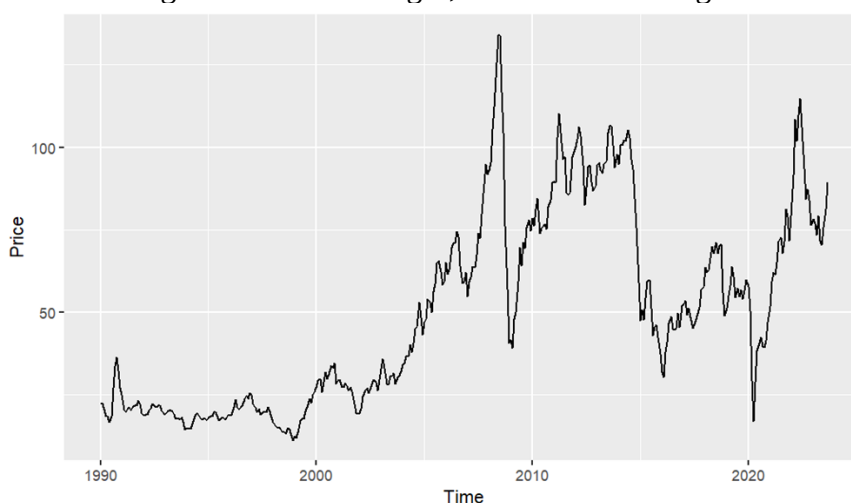


Fig. 1 Global WTI crude price

From Fig. 1, it can be concluded that before June 2007, crude oil prices had entered an upward phase from a stable stage, and then experienced a significant increase and decrease. Its highest value is almost 3.5 times the lowest value, and the period is only 9 months. This can be seen as a microcosm of the 2007 subprime mortgage crisis. Afterward, there was a significant increase or decrease every approximately 5 years, and during these periods, crude oil prices showed a slow upward trend.

2.3. Model Selection

This article selects the Autoregressive Integrated Moving Average (ARIMA) model to process the time series data of crude oil prices. ARIMA is a simple and wide model when processing time series data, consisting of an autoregressive (AR), integrated (I), and moving average (MA). The AR part is used to process the autoregressive part of the time series, which considers the impact of observations from past periods on the current value. By using differential processing to remove trend and seasonal components from the time series, the I part is utilized to bring non-stationary time series into stationarity. The MA part is used to process the moving average part of the time series, which considers the impact of past prediction errors on the current value.

3. Results and Discussions

3.1. Data Processing

The data used for ARIMA needs to ensure its stability. From the ACF graph in Fig. 2, it can be seen that the price data have strong autocorrelation and are a non-stationary sequence. Therefore, differential processing is needed to ensure data stability.

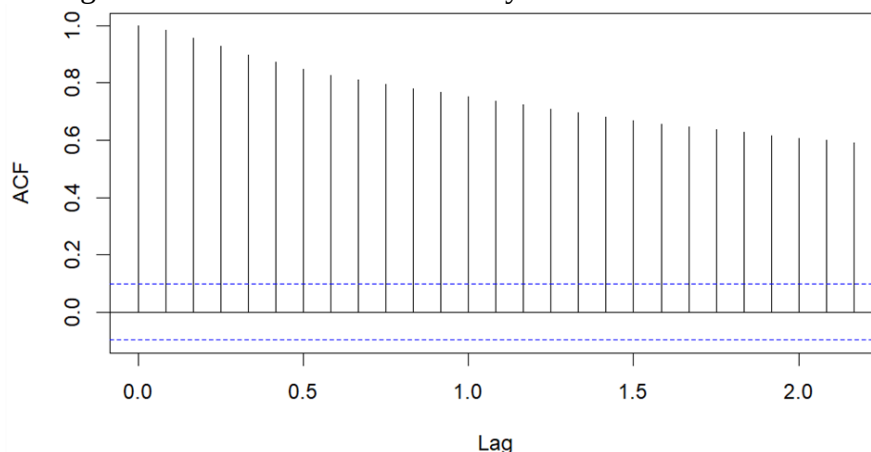


Fig. 2 The ACF plot of the global price of WTI crude

According to the results of the first-order difference, its ACF image (Fig. 3) and PACF image (Fig. 4) meet the requirements. Perform an ADF test on it, and the outcomes are displayed in Table 1. The p-value is 0.01. Therefore, it can be considered that the data obtained after the first-order difference are stationary.

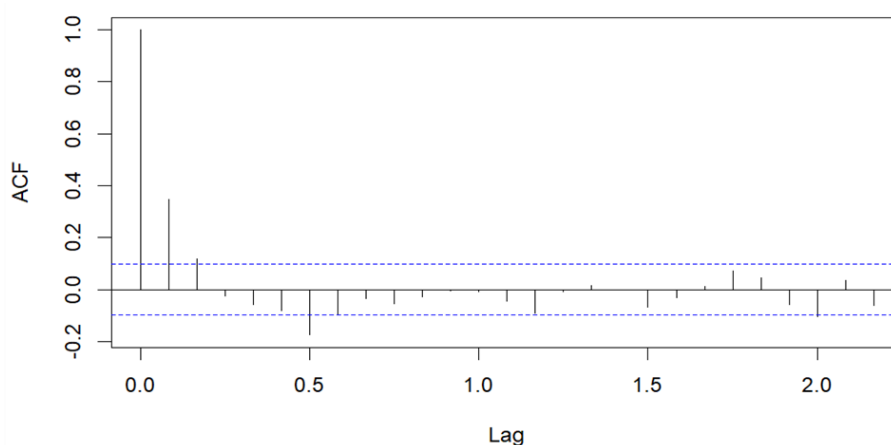


Fig. 3 The ACF plot

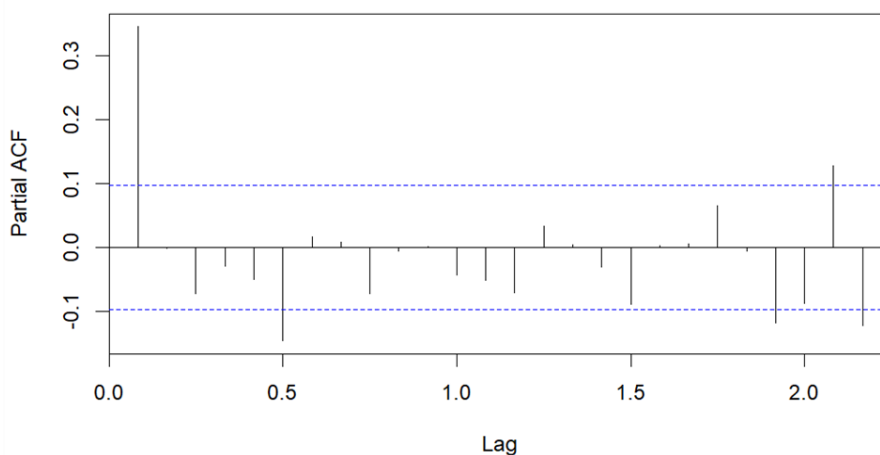


Fig. 4 The PACF plot

Table 1. The ADF test

Dickey-Fuller	Lag order	p-value
-7.671	7	0.01

3.2. Model Evaluation

The accuracy of the model is reflected in the deviation in the relationship between the fitted and real values. The less the variance, the better the fitting degree, and the more accurate the model. The I part has already been determined as 1 in the previous text. The following is to confirm the AR and MR parts.

This article selects two indicators, using root mean square error (RMSE) and Akaike information criterion (AIC) to estimate it. RMSE is a commonly used indicator of the difference between measured values, used to measure the deviation between fitted values and actual values. At the same time, AIC can be used to select models with the least number of free parameters in the data to avoid overfitting situations. These two show that the model fits the data better in a smaller size.

Table 2. Model Evaluation

ARIMA Model	RMSE	AIC
(1,1,0)	4.633	2390.52
(1,1,1)	4.633	2392.52
(2,1,0)	4.633	2392.52
(2,1,1)	4.632	2394.32
(3,1,0)	4.621	2392.44
(3,1,1)	4.579	2387.12
(3,1,2)	4.578	2389.04
(4,1,0)	4.619	2394.07
(4,1,1)	4.620	2396.18

From Table 2 above, it can be seen that the RMSE values of ARIMA (3,1,1) and ARIMA (3,1,2) are both small, but the difference is not significant, while the AIC value of ARIMA (3,1,1) is lower. Therefore, this article believes that ARIMA (3,1,1) is most appropriate given the statistics on oil prices. The following text will only use this model for prediction and analysis.

3.3. Forecasting

This article will use ARIMA (3,1,1) to predict 12 numerical values for oil price data, and the results are shown in Fig. 5 and Table 3:

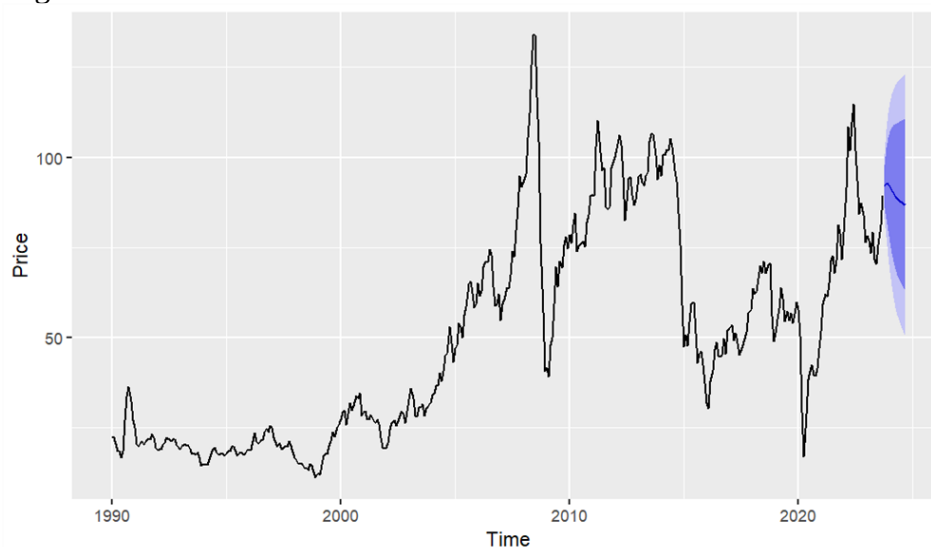


Fig. 5 Forecasts from ARIMA(3,1,1)

Table 3. Forecasts from ARIMA(3,1,1)

Time	Point Forecast	L 80	H 80	L 95	H 95
2023/10	92.023	86.148	97.899	83.038	101.009
2023/11	92.849	83.089	102.608	77.923	107.775
2023/12	92.573	79.648	105.498	72.806	112.340
2024/01	91.815	76.404	107.227	68.245	115.385
2024/02	90.899	73.546	108.253	64.359	117.440
2024/03	89.996	71.116	108.877	61.121	118.871
2024/04	89.185	69.086	109.284	58.446	119.924
2024/05	88.495	67.403	109.588	56.238	120.753
2024/06	87.928	66.008	109.849	54.404	121.454
2024/07	87.474	64.845	110.103	52.866	122.083
2024/08	87.116	63.867	110.366	51.559	122.674
2024/09	86.838	63.033	110.643	50.432	123.245

The blue line shows the model's anticipated outcomes and the black line shows the actual data. From the findings above, it is evident that the predicted oil price experienced a slight increase in the first two months, followed by a downward trend, and the magnitude decreased over time.

3.4. Check Residuals

To ensure that this model effectively interprets the information of the data, The next step is to detect whether its residual is white noise. The inspection results (Fig. 6) and QQ plot (Fig. 7) are as follows:

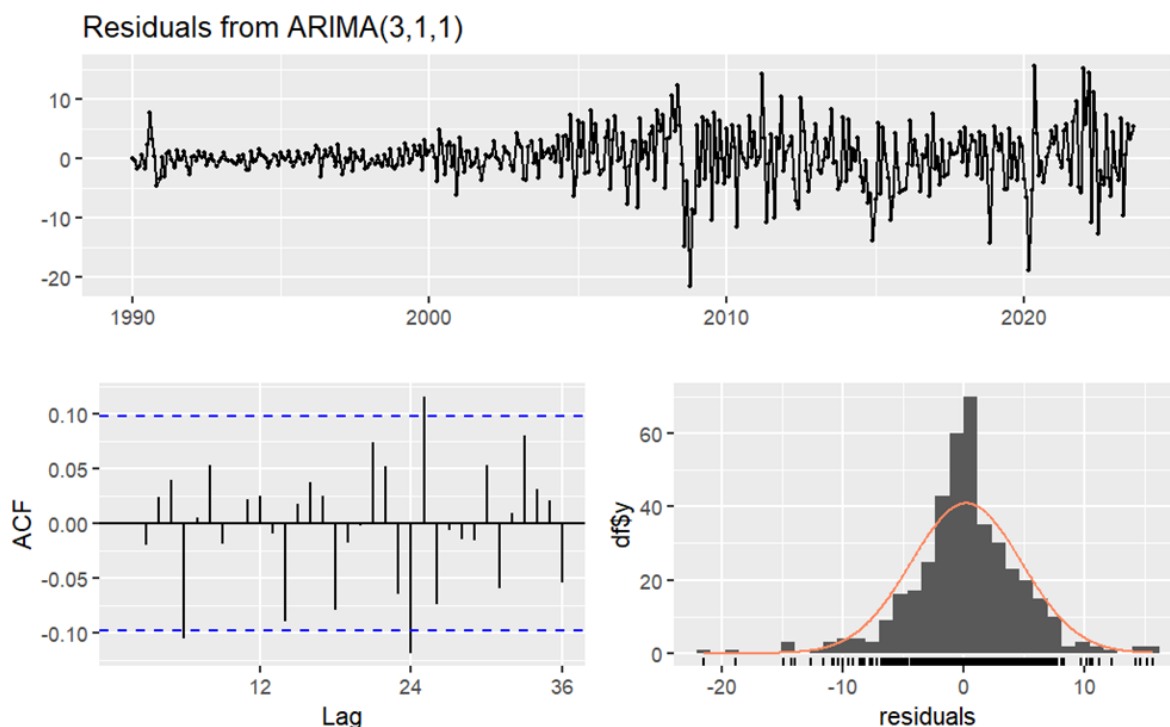


Fig. 6 Residual testing

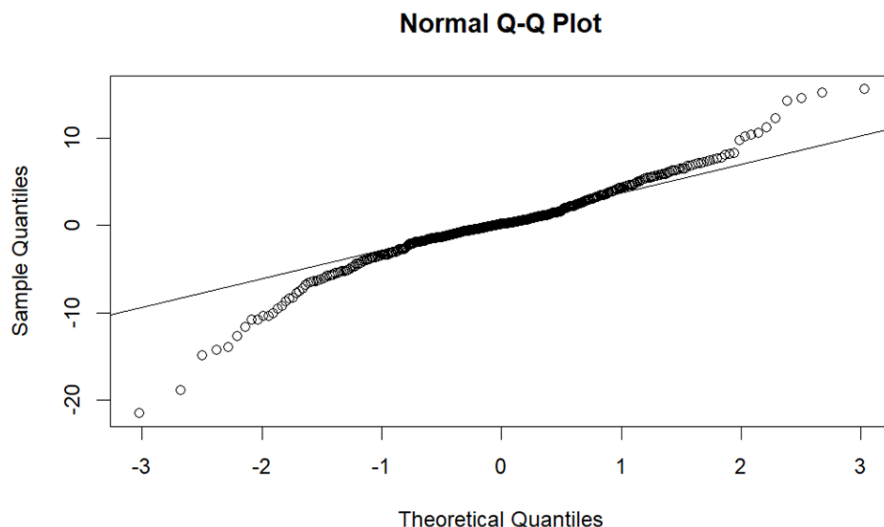


Fig. 7 ARIMA(3,1,1) QQ plot

In Fig. 6, it can be seen from the above figure that the variation amplitude of its residual after 2005 has significantly improved compared to before, which conforms to the distribution of white noise. In the ACF graph, three points exceeded the critical value. In the lower right figure, the residual is shown to have a normal distribution. The QQ diagram in Fig. 7 is a straight line, indicating that the residual can be considered as white noise. In addition, this article adopts the Ljung Box test as the final criterion for determining whether the residual is white noise, and Table 4 displays the outcomes:

Table 4. Ljung-Box test

Q^*	df	p-value
25.889	20	0.170

It is evident from the Ljung Box test findings that its p-value turns out with 0.170, bigger than 0.05. Therefore, this article believes its residual value is white noise and that ARIMA (3,1,1) is a reliable fitting model.

3.5. Critical Thinking

Although the above conclusions have passed their respective tests and the predicted results have a certain degree of credibility. The statement means that the given data is in line. There are still certain problems that cannot be solved in this article. Firstly, according to the image, the price changed significantly before and after 2007. Before 2007, prices were relatively stable and did not show significant changes, while after that, prices fluctuated frequently and could fluctuate more than three times. This article believes that if previous data is not taken into account, the predicted results will be more accurate and the test results will be better, but it will also undermine the integrity of time series data, leading to the loss of included information. Secondly, through the images, it can be seen that the state of the global economy significantly affects the price, and the predictions in this article clearly cannot predict when these global economic fluctuations will occur. Therefore, there are still certain limitations to the predictions in this article.

4. Conclusion

Through the research in this article, it can be concluded that although the price of crude oil may slightly increase in the short term, the overall trend of change in the next year will still decrease. However, due to the uncertainty of the international economic situation, predictions made further away will become more inaccurate. Meanwhile, in recent years, the world has been affected by the pandemic, and the economies of various countries have been impacted and are in a recovery period, which will make oil prices more difficult to predict.

Through the predictions and research in this article, some suggestions can be provided. From an individual perspective, it is not a reasonable choice for crude oil investors to make relevant investments shortly. From a national perspective, the government should try to minimize the impact of the international economic situation on the domestic market. This can not only maintain a stable development trend of the domestic economy, protect domestic enterprises from international economic shocks, and maintain social stability, but also serve as feedback, and more intuitively and accurately trade various international goods while ensuring domestic stability. This not only benefits the international crude oil market but also puts the country as a whole in an advantageous position in international trade.

References

- [1] Zhang Y J, Wu Y B. The dynamic information spill-over effect of WTI crude oil prices on China's traditional energy sectors. *China Agricultural Economic Review*, 2018, 10(3): 516-534.
- [2] Irwin S H, Sanders D R, Merrin R P. Devil or angel? The role of speculation in the recent commodity price boom (and bust). *Journal of Agricultural and Applied Economics*, 2009, 41(2): 377-391.
- [3] Shaikh I. On the relation between the crude oil market and pandemic Covid-19. *European Journal of Management and Business Economics*, 2021, 30(3):331-356.
- [4] Pellejero S. Oil prices fall as rising COVID-19 cases prompt demand concerns. *Investors Also Eye Rising Crude*, 2020.
- [5] Kadhem S, Thajel H. Modelling of crude oil price data using hidden Markov model. *Journal of Risk Finance*, 2023, 24(2): 269-284.
- [6] Pan S, Li H, Wang Y, Cai W. Crude oil price prediction with LSTM neural networks. *Computer Technology and Development*, 2021, 31(5): 180-185.
- [7] Pan W. Mixed modified fruit fly optimization algorithm with general regression neural network to build oil and gold prices forecasting model. *Kybernetes*, 2014, 43(7): 1053-1063.
- [8] Wu Y X, Wu Q B, Zhu J Q. Improved EEMD-based crude oil price forecasting using LSTM networks. *Physica A: Statistical Mechanics and its Applications*, 2019, 516(15): 114-124.
- [9] Huang N E, Shen Z, Long S R, et al. The empirical mode decomposition and the Hilbert spectrum for non-linear and non-stationary time series analysis. *Proceedings of The Royal Society A*, 1998, 454: 903-995.
- [10] María E T, Marcelo A C, Gaston S, et al. A complete ensemble empirical mode decomposition with adaptive noise. Prague, Czech Republic, 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2011: 4144-4147.