

Forecasting of oil and gas stock market in major US

Zhonghao Zuo

Faculty of Mathematical and Physical Sciences, University College London, Gower Street, London, WC1E 6BT, United Kingdom

ucakzzu@ucl.ac.uk

Abstract. Oil and gas are two crucial resources in modern life and the development of technology. Investors are not only concerned about their own price, but their stock price is also an important part of formulating investment plans. This article will focus on the adjusted closing price since it consists of adjustments for splits, dividends, and capital gain distributions, which can demonstrate the situation of oil and gas in the stock market better. After the occurrence of COVID-19, the economic market of the whole world including the US. encountered great changes in many aspects. Thus, this article considered applying the model with the best performance from ARIMA, exponential smoothing, and Holt-Winters' method to forecast the future adjusted closing prices of two large and representative companies in oil and gas from the US. According to the predicted results, the US oil and gas market is likely to experience a small but steady increase.

Keywords: Forecasting, ARIMA model, exponential smoothing, Holt-Winters' method.

1. Introduction

In the past few years, the global economy and market encountered a bunch of unexpected situations like the appearance of COVID-19 in 2019, and climate change. These unpredictable situations had a significant impact on the economic system from various aspects such as the stock market. According to the wide range of the applications of oil and gas, their stock price in major US companies is highly attractive to all investors interested in this field. From Basuony et al., the investigation of the stock market will benefit practitioners and policymakers by capturing the dynamic trend of the stock market [1]. Therefore, this research contributes to obtaining the prediction of future oil and gas stock adjusted closing price, which will help investors to make more reasonable and low-risk decisions since it can reveal the peripheral information of corresponding stock according to Wei et al. [2].

For any variables that vary with time, Chouksey stated that time series can be used to deal with it, and one of the illustrations is investigating the closing price of stock over time [3]. Thus, it is necessary to apply various time series models and choose the one with the best performance among all of them to gain more accurate forecasting values. Therefore, one key component in the research is choosing which models to fit the dataset and forecasting. According to Xue, the Autoregressive Integrated Moving Average (ARIMA) model is one of the most appropriate models to fit the dataset if it is a non-stationary time series [4]. Since there is no guarantee that the dataset is always stationary, applying the ARIMA model is necessary. Furthermore, there might exist seasonality in the dataset, the propose of Seasonal Autoregressive Integrated Moving Average (SARIMA) model can deal with it since Vishwakarma et al.

stated that the SARIMA model can directly handle time series data that contains seasonal components [5]. Panda also stated that the Holt-Winters forecasting algorithm plays an important role in forecasting and smoothing time series data [6]. Thus, it is essential to consider using the Holt-Winters forecasting algorithm to predict the future adjusted closing price of oil and gas in the US. From Kingsley Arum, it is suitable to apply the Autoregressive conditional heteroskedasticity (ARCH) model there since it is designed to deal with financial issues about the rate of change of stock price over time [7]. Moreover, it is inevitable to apply the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model after fitting the ARCH model to the dataset. The reason is that the GARCH model is an extension of the ARCH model and typically demonstrates better performance in fitting time series data from Almansour et al. [8].

After selecting these models for forecasting the dataset, it is important to determine the criteria for selecting the model with the best performance. Vijn et al. stated that the model with the lower value of Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) is more appropriate for forecasting stock closing price since it has higher efficiency when compared to other models [9]. In addition, Akaike's Information Criterion (AIC) also can be used as a judgment standard for selecting time series models.

Once the time series model with the best performance has been selected with specified criteria, it can be used to forecast future adjusted closing prices of two major oil and gas companies in the US. The forecasting of future stock adjusted closing prices there provides not only advice to stock investors but also some indication of possible development of the price of oil and gas. According to Jesus et al., the real scenario in the US is that the relationship between oil and stock price is not positively correlated in the long period [10]. This indicates that when US stock closing prices for oil and gas increase, it may be due to a decline in oil prices, and vice versa.

2. Methodology

2.1. Data sources

These two datasets are from Yahoo Finance, which are the historical monthly stock prices of Exxon Mobil Corporation (XOM) and Chevron Corporation (CVX) from January 1985 to July 2024. These two US local companies can somehow represent the feature of the stock price of oil and gas in major US. Thus, the adjusted closing price of these two companies has been chosen to investigate in this paper.

2.2. Data processing

From the time series plot of the adjusted closing price of XOM (Figure 1) and CVX (Figure 2), it is obvious that both have non-constant variation. Therefore, the transformation is needed for these two datasets to eliminate the influence of non-constant variation.

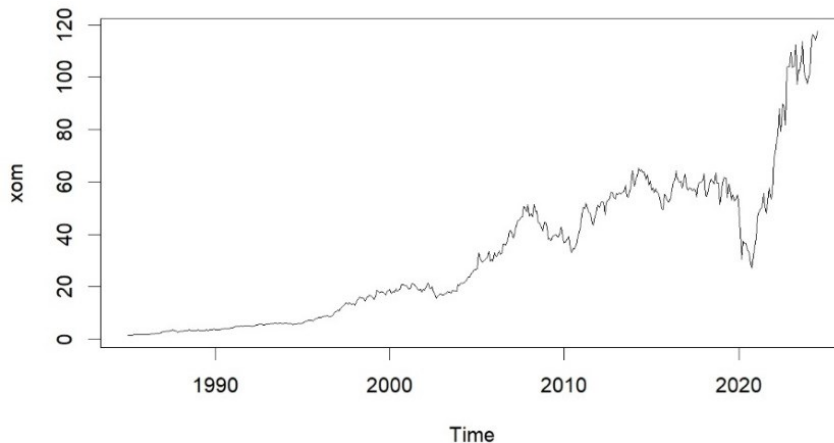


Figure 1. Time series plot of XOM (adjusted closing price)

For the stock market, the most used method is Log transformation. Furthermore, splitting the dataset into a train set and test set is necessary to decide the model and calculate the corresponding RMSE, MAPE, and AIC. The train set in this article for XOM and CVX is from the beginning of data to July 2022, and the test set is from August 2022 to July 2024.

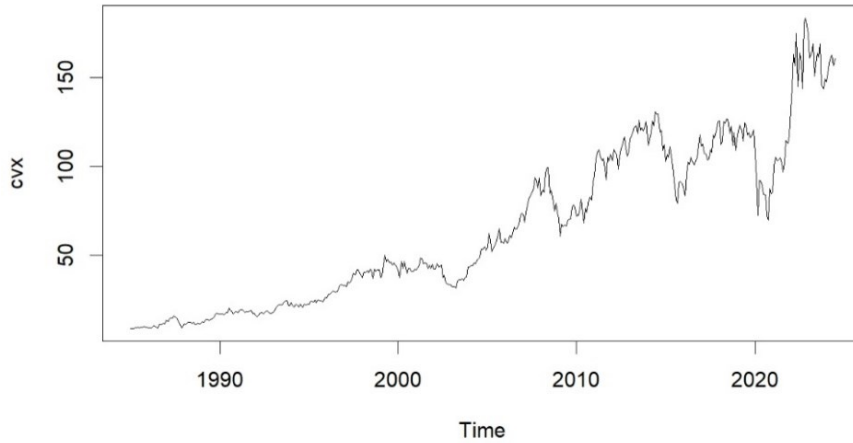


Figure 2. Time series plot of CVX (adjusted closing price)

2.3. Model selection

The ARIMA model is always appropriate for forecasting time series data such as stock price, so it should be considered first. If these two datasets show very strong patterns or seasonality, the SARIMA model needs to be included to deal with it. Moreover, exponential smoothing and Holt-Winters' method are also good at handling trend and seasonality, which is the consequent choice. To be more specific, this article used simple/double exponential smoothing and Holt-Winters' additive/multiplicative method to fit the dataset. The ARCH and GARCH models are commonly used in forecasting the volatility of financial time series data. Thus, the ARCH and GARCH models are needed since the change in stock adjusted closing price also is a kind of financial volatility. In this article, the proposed ARCH and GARCH models are ARCH (1) and GARCH (1, 1) since these two models are most commonly used in forecasting time series. The criteria for selecting a model there is based on the corresponding RMSE, MAPE, and AIC of each model. After the best model is decided for each dataset, the forecasting stock adjusted closing price can be obtained by the fitting model. Therefore, the feature of future major US stock adjusted closing price of oil and gas can be analyzed by the forecasting result of these two companies.

3. Results and discussions

3.1. Data processing for XOM

The initial step is splitting the dataset into a train set and test set, then figuring out appropriate models based on the train set and examining them with the test set. Since the presence of non-constant variation and increasing trend can be observed in the whole dataset including the train set, it indicated the existence of non-stationarity. Therefore, the application of log transformation is needed for to dataset to stabilize the variance. Moreover, the differencing should be considered to eliminate the non-stationarity. After log transformation and first-order differencing, the performance of an Augmented Dickey-Fuller (ADF) test can help determine if further differencing is required. Based on the p-value of the ADF test in Table 1, the train set after log transformation and first-order differencing is stationary since the p-value is smaller than 0.05, which accepts the alternative hypothesis. Additionally, the time series plot of the train set after operations (Figure 3) is also approximately stationary.

Table 1. ADF test for XOM train set after log transformation and first-order differencing.

Dickey-Fuller	Lag Order	p-value
-6.660	7	0.01
Alternative hypothesis: stationary		

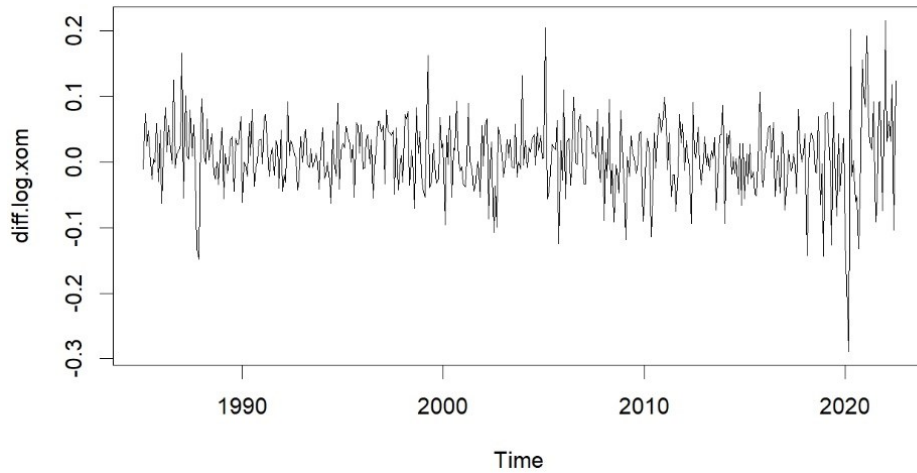


Figure 3. Time series plot of XOM train set (after log transformation and first-order differencing).

3.1.1. ARIMA model for XOM. For the ARIMA model, the parameters p , d , and q need to be decided manually according to corresponding ACF and PACF plots. The value of d is 1 since only first-order differencing is needed for a train set after log transformation. To determine the value of p and q , the lag that exceeded the confidence interval in the ACF and PACF plots is a critical factor, which is lag 7 for both plots. From the ACF plot (Figure 4) and the PACF plot (Figure 5), the proposed ARIMA models will be ARIMA (7, 1, 7). Moreover, there is no significant seasonal lag that exceeds the confidence interval in ACF and PACF plots, and the seasonal differencing is not applied there. Thus, the SARIMA model does not need to be considered there.

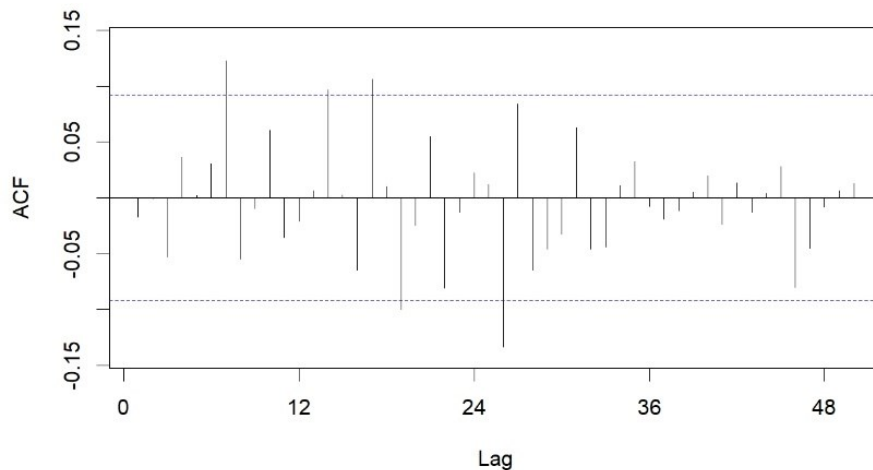


Figure 4. ACF plot for XOM train set (after log transformation and first-order differencing).

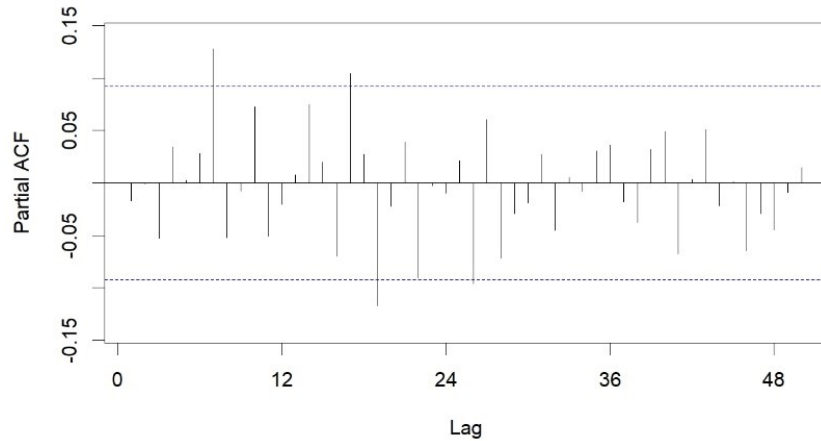


Figure 5. PACF plot for XOM train set (after log transformation and first-order differencing).

3.1.2. Other models for XOM. After generating the appropriate ARIMA model for the train set, comparing its performance to ARCH (1) and GARCH (1, 1) will be the next step. To fit the train set to ARCH (1) and GARCH (1, 1), it is important to ensure that the dataset is stationary. Thus, the dataset that is used to fit ARCH (1) and GARCH (1, 1) will still be a train set after log transformation and differencing.

For exponential smoothing and Holt-Winters' methods, log transformation is also needed for the train set even if these models can handle the non-stationary dataset. After eliminating the non-constant variance with log transformation, it can improve the performance of exponential smoothing and Holt-Winters' methods to obtain more accurate predictions.

3.1.3. Model Selections for XOM. According to the value of RMSE, MAPE, and AIC for each model in Table 2 below, the double exponential smoothing method will be the most suitable there. The reason is that it contained the lowest value of RMSE and MAPE, and its AIC is also relatively low when compared to other models.

Table 2. Values of RMSE, MAPE, and AIC for each model (XOM).

Model	RMSE	MAPE	AIC
ARIMA (7, 1, 7)	9.687	0.077	-1313.786
ARCH (1)	7.481	0.061	-1356.859
GARCH (1, 1)	7.640	0.064	-1377.681
Simple Exponential Smoothing	17.205	0.146	-1304.754
Double Exponential Smoothing	7.463	0.060	-1300.739
Holt-Winters' Addictive	8.565	0.072	-1233.734
Holt-Winters' Multiplicative	10.336	0.084	-1124.640

3.1.4. Predictions for XOM. The next procedure is forecasting the future monthly adjusted closing price based on the double exponential smoothing method. The key concept is transforming the forecasting value back to the original scale since the dataset that used to fit is after log transformation. Table 3 demonstrates the adjusted closing price over the next 12 months, and figure 6 combines the original dataset, predicted values, and prediction intervals together. Obviously, the forecasting adjusted closing price for the future 12 months follows the increasing trend. However, the range of the corresponding 95% prediction interval for each month increases as time increases, which means that the prediction is less accurate as time increases (Figure 6).

Table 3. Predicted adjusted closing price for XOM.

Date	Prediction	95% Prediction Interval
August 2024	118.846	[106.160, 133.047]
September 2024	120.149	[102.595, 140.705]
October 2024	121.466	[100.052, 147.463]
November 2024	122.797	[98.026, 153.828]
December 2024	124.143	[96.321, 160.003]
January 2025	125.504	[94.834, 166.093]
February 2025	126.880	[93.508, 172.162]
March 2025	128.271	[92.304, 178.253]
April 2025	129.667	[91.195, 184.398]
May 2025	131.099	[90.163, 190.620]
June 2025	132.536	[89.194, 196.939]
July 2025	133.989	[88.277, 203.370]

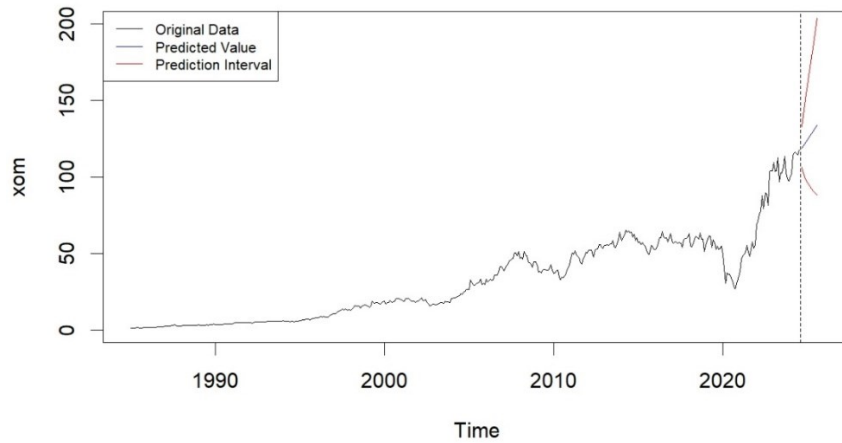


Figure 6. Times series plot of original data predicted values, and prediction interval (XOM).

3.2. Data processing for CVX

To determine the appropriate model for Chevron Corporation (CVX), the procedures will be similar to all operations above. The first movement is still separating the dataset into a train set and a test set. Since the adjusted closing price of CVX shows strong non-constant variance and increasing trend, which means transformation and differencing might be needed to handle these two non-stationary features. Based on the characteristics of this dataset, the application of log transformation will be appropriate. The result given by the ADF test in Table 4 indicated that the train set after operations is stationary since the p-value is less than 0.05. Moreover, through the time series plot of the train set after log transformation and differencing (Figure 7), it is stationary obviously.

Table 4. ADF test for CVX train set after log transformation and first-order differencing.

Dickey-Fuller	Lag Order	p-value
-7.409	7	0.01
Alternative hypothesis: stationary		

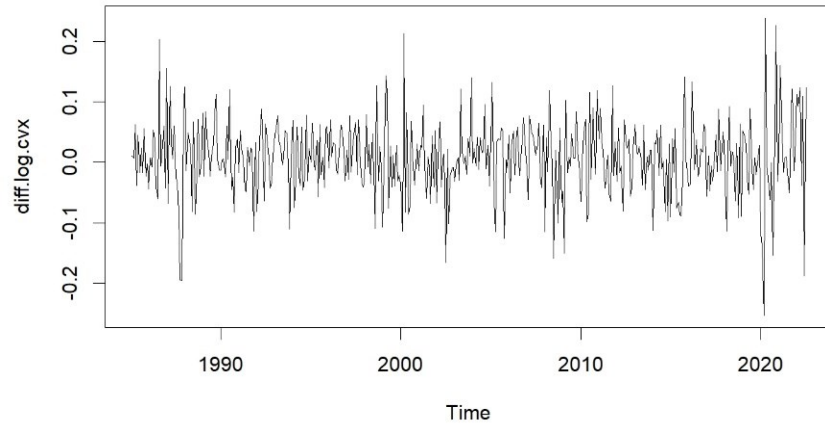


Figure 7. Time series plot of CVX train set (after log transformation and first-order differencing).

3.2.1. ARIMA model for CVX. The value of d is 1 in the ARIMA model since the first-order differencing is enough. To determine all parameters for the ARIMA model, the ACF plot and PACF plot of the train set after operations are necessary. According to lags in the ACF plot (Figure 8) and PACF plot (Figure 9) that exceed the confidence interval, the chosen value of p and q will both be 1. Therefore, the appropriate ARIMA model there will be ARIMA (1, 1, 1). Additionally, there is no significant spike in seasonal lags and seasonal differencing is not necessary there, so the SARIMA model is not essential for CVX.

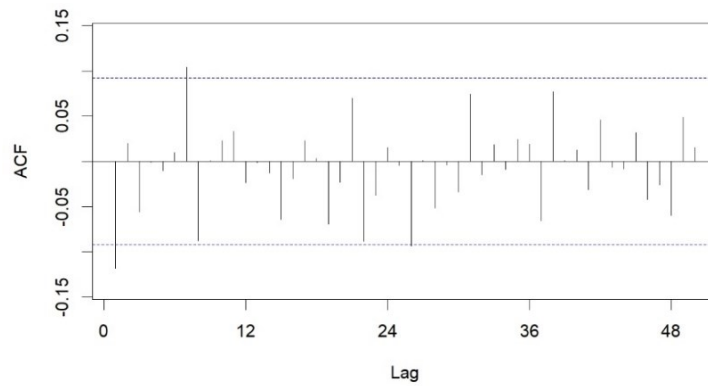


Figure 8. ACF plot for CVX train set (after log transformation and first-order differencing).

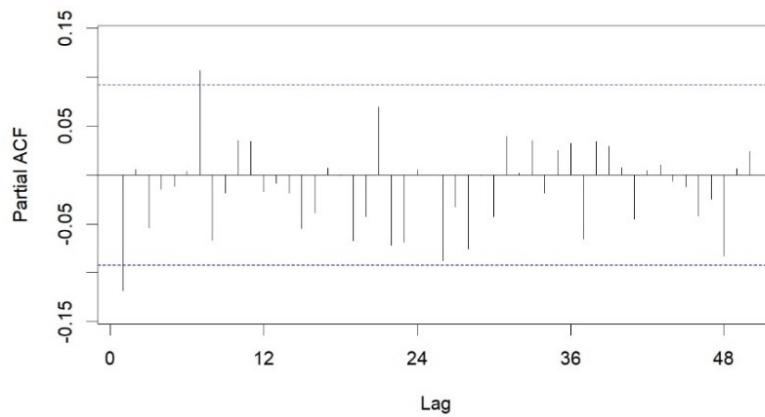


Figure 9. PACF plot for CVX train set (after log transformation and first-order differencing).

3.2.2. Other models for CVX. Before fitting the ARCH (1) and GARCH (1, 1) models, the train set needs to be stationary to satisfy the assumption. Thus, the train set after log transformation and differencing will be the target dataset. To fit exponential smoothing and Holt-Winters' method, log transformation is also needed to acquire more precise result.

3.2.3. Model selections for CVX. From Table 5, the ARIMA (1, 1, 1) will be the most suitable model there since it has the lowest values of RMSE and MAPE. Therefore, the ARIMA (1, 1, 1) should be used to forecast future monthly adjusted closing price for CVX over the next 12 months based on its performance.

Table 5. Values of RMSE, MAPE, and AIC for each model (CVX).

Model	RMSE	MAPE	AIC
ARIMA (1, 1, 1)	11.127	0.054	-1207.879
ARCH (1)	25.165	0.143	-1233.106
GARCH (1, 1)	28.784	0.163	-1243.897
Simple Exponential Smoothing	11.164	0.055	-1211.193
Double Exponential Smoothing	24.893	0.141	-1209.657
Holt-Winters' Addictive	26.102	0.145	-1145.729
Holt-Winters' Multiplicative	26.326	0.144	-1129.560

3.2.4. Predictions for CVX. To forecast the future monthly adjusted closing price based on the double exponential smoothing method. The critical process is transforming the forecasting value back to the original scale since the dataset that used to fit is after log transformation and first-order differencing. Table 6 below shows the adjusted closing price for CVX over the future 12 months, and Figure 10 contains the original dataset, predicted values, and prediction intervals for the final chosen model. Through Figure 10, the forecasting adjusted closing price for the future 12 months gradually approximates to stable. The range of the corresponding 95% prediction interval for each month also increases as time increases, which represents the appearance of increasing uncertainty.

Table 6. Predicted adjusted closing price for CVX.

Date	Prediction	95% Prediction Interval
August 2024	159.996	[141.331, 181.127]
September 2024	160.053	[135.617, 188.892]
October 2024	160.046	[131.072, 195.425]
November 2024	160.047	[127.335, 201.163]
December 2024	160.047	[124.108, 206.394]
January 2025	160.047	[121.249, 211.260]
February 2025	160.047	[118.6702, 215.851]
March 2025	160.047	[116.314, 220.223]
April 2025	160.047	[114.141, 224.416]
May 2025	160.047	[112.120, 228.462]
June 2025	160.047	[110.229, 232.381]
July 2025	160.047	[108.450, 236.193]

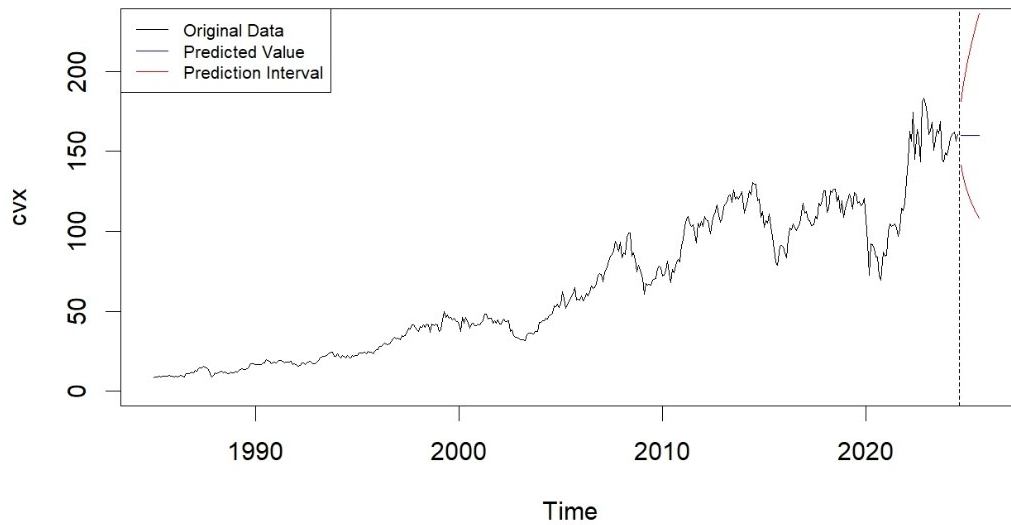


Figure 10. Times series plot of original data predicted values, and prediction interval (CVX).

4. Conclusion

After investigating in the adjusted closing price of Exxon Mobil Corporation (XOM) and Chevron Corporation (CVX), it provided a general analysis of the future major US stock market. To declare in advance, this article only researched the stock adjusted closing price of two iconic companies in the US, it can only give a rough analysis of the oil and gas stock market rather than a comprehensive analysis. Since there are many external factors that are not considered in this article, the results can only be seen as general suggestions to investors.

The stock price analyzed is an adjusted closing price, it reflects the overall value of stock including splits, dividends, and other business adjustments. According to the forecasting future adjusted closing price of XOM, it suggested that there might appear a steady upward trend in the stock market. For the forecasting future adjusted closing price of CVX, it exhibits that the stock market may usher in a very small increase and gradually stabilize. Based on these two forecasts, it is reasonable to assume that the major US oil and gas stock markets will encounter a relatively small percentage of the upward trend, but the overall outlook might likely be stable. With this finding, investors can make a general reference to how they should invest in the US oil and gas stock market next.

Moreover, the price of oil and gas might not exhibit as same as their stock price. According to the discovery from Jesus et al, the price of oil and gas in the US might fall slightly as the stock price rises, but it also might generally stabilize along with time. This is a general suggestion for investors who pay more attention to the price of oil and gas itself.

References

- [1] Basuony M A K, Bouaddi M, Ali H and EmadEldeen R 2021 The effect of COVID-19 pandemic on global stock markets: Return, volatility, and bad state probability dynamics. *J Public Aff*, 2761.
- [2] Wei J, Xu Q and He C 2022 Deep learning of predicting closing price through historical adjustment closing price. *Procedia Computer Science*, 202, 379-384.
- [3] Chouksey S 2019 Stock Price Prediction Using Time Series Models. MSc Data Analytics dissertation, Dublin Business School, 14.
- [4] Xue Q 2023 Stock Price Forecasting Based on ARIMA Model: An Example of Cheung Kong Hutchison Industrial Co. In: *Highlights in Business, Economics and Management*, 10, 425-430.

- [5] Vishwakarma A, Singh A, Mahadik A and Pradhan R 2020 Stock Price Prediction Using Sarima and Prophet Machine Learning Model. *International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)*, 9, 1-6.
- [6] Panda M 2020 Application of ARIMA and Holt-Winters forecasting model to predict the spreading of COVID-19 for India and its states. *medRxiv*.
- [7] Arum K C and Uche P I 2019 Volatility Modelling using Arch and Garch Models (A Case Study of the Nigerian Stock Exchange). *International Journal of Mathematics Trends and Technology (IJMTT)*, 65, 4.
- [8] Almansour B, Alshater M M and Almansour A Y 2021 Performance of ARCH and GARCH Models in Forecasting Cryptocurrency Market Volatility. *Industrial Engineering & Management Systems*, 20, 130-139.
- [9] Vijn M, Chandola D, Tikkiwal VA and Kumar A 2020 Stock Closing Price Prediction using Machine Learning Techniques. *Procedia Computer Science*, 167, 599-606.
- [10] Pitta de Jesus D, Souza Bezerra B F L and Da Nóbrega Besarria C 2020 The non-linear relationship between oil prices and stock prices: Evidence from oil-importing and oil-exporting countries. *Research in International Business and Finance*, 54, 101229.