The Crude Oil Effect on Major Energy Company Stock

Hua Mo

Introduction

The crude oil is the major resource to make the different products. There are three types of the companies. The upstream companies are the companies to take the crude oil out of the earth. The downstream companies which are called refinery, are the companies to separate the crude oil and fractionate to different fractions, such as gasoline, Diesel, Naphtha, etc..The petrochemical company are the companies to use the downstream products to make different chemical products. Some companies, such as XOM, Shell, integrate the upstream, downstream and petrochemical in the company portfolio. Some companies, such as Valero, Marathon, are only have downstream company with refinery. There are some standalone petrochemical companies, such as Lyondell. There are also some oil and gas service companies, such as Slumberger and Haliburton. These companies related to the crude oil.

Normally, when the crude oil price is high, there will be profits for the upstream company. Companies would make money by drilling the oil out of the ground. However the high crude oil price could potentially hurt the downstream and petrochemical business because the raw materials price will be high for these industry. The profit margin of downstream and petrochemical will shrink.

A company stock is normally reflecting the operation of the company. A stock of profitable company will be high because investor will like to buy the stock. A stock of poor operation company will be low because investor will lose interests.

In this capstone project, we will investigate the relationship between the crude oil price vs. a stock from a specific energy company. We will try to see the effect of crude oil on these energy companies.

Methodology

I. Data Wrangling

The stock data was streamed from alphavantage. The data were the past 24 months

The crude oil data was extracted from Yahoo Finance

2. Data Processing

Time series technique is used to remove the seasonal effect and residue.

The features comparison is used to investigate the effect of crude oil price on the energy companies

The ARIMA model is used to predict the potential future price of the crude oil and energy companies

Three batch programs are used to stream the data, feature comparison and predict the future price of energy companies. With the batch programs, it will be easy to process multi data at the same time.

Data Wrangling

Data wrangling is the first step to process data. There were two datasets needed:

I. Stock

1. Data Source

The source of the stock data is from alphavantage (<https://www.alphavantage.co>). This is a website to provide two years of the stock data without cost. To obtain the data, a free API key will be obtained after registering on the website. Due to a large amount of data, the 24 months data were sliced by individual month. To facilitate the data stream, a list was created. The list is shown below:

*Slice\_list=['year1month1','year1month2','year1month3','year1month4','year1month5','year1month6','year1month7','year1month8','year1month9','year1month10','year1month11','year1month12','year2month1','year2month2','year2month3','year2month4','year2month5','year2month6','year2month7','year2month8','year2month9','year2month10','year2month11','year2month12']*

As data analysis, it will be convenient to stream different company data at the same time. In order to achieve this goal, a company list was also created as following:

*symbol\_list=read\_list=['XOM','BP','RDS-B','CVX','VLO','MPC','LYB','SLB','HAL',"TOT"]*

In this project, ten stocks were used. The detail explanation of the stock is shown in

Table 1.

Table 1. The Stock

|  |  |  |
| --- | --- | --- |
| Stock | Company | Business |
| XOM | ExxonMobil | Upstream, Downstream, Petrochemical |
| BP | British Petroleum | Upstream, Downstream |
| RDS-B | Royal Dutch Shell | Upstream, Downstream, Petrochemical |
| CVX | Chervon | Upstream, Downstream |
| VLO | Valero | Downstream |
| MPC | Marathon Petroleum | Downstream |
| LYB | LyondellBassell | Petrochemical |
| SLB | Slumberger | Oil service for upstream |
| HAL | Halliburton | Oil service for upstream |
| TOT | Total | Petrochemical |

This project is an initial setup to fetch the data. Although more companies can be added to the symbol list, the runtime to fetch the data will be longer. The reviewer can add symbol data later. In this project, the time interval for the stock data was 60 min.

After the data has been fetched, the individual data was saved as individual symbol with filename. The output filename is stock\_symbol.csv. Therefore there would be 10 stock data.

2. Data Cleaning

After the stock data were obtained. The example of the stock data of XOM is shown in Table 2.

Table 2. XOM Stock Original Data

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | time | open | high | low | close | volume |
| 0 | 10/5/2020 20:00 | 33.80 | 33.80 | 33.60 | 33.64 | 22195 |
| 1 | 10/5/2020 19:00 | 33.74 | 33.80 | 33.74 | 33.78 | 30590 |
| 2 | 10/5/2020 18:00 | 33.73 | 33.75 | 33.70 | 33.74 | 22579 |

The stock data was averaged by daily through resample(‘D”).mean(). The open, high and low were dropped. Then the data form became Table 3.

Table 3. XOM Stock Cleaned Data

|  |  |  |
| --- | --- | --- |
| time | XOM close price | XOM volume |
| 11/12/2018 | 72.75968 | 719512.5 |
| 11/13/2018 | 71.27194 | 1053756 |
| 11/14/2018 | 70.26396 | 1168918 |

The cleaned data were saved as individual csv files with their stock symbol as the filename. For example, the file for the XOM cleaned data file was XOM.csv There were 10 data files corresponds to the stock symbol. These data files will be used for the future data exploration process. The detail data files were shown in the Appendix.

II. Crude Oil

The crude oil data was downloaded from the Yahoo finance. The data is shown in Table 4.

Table 4. Crude Oil Data

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Date | Open | High | Low | Close | Adj Close | Volume |
| 9/17/2018 | 68.93 | 69.72 | 68.53 | 68.91 | 68.91 | 422294 |
| 9/18/2018 | 68.80 | 70.42 | 68.53 | 69.85 | 69.85 | 234202 |

We used the Adj Close to evaluate the effect on the crude oil data. Therefore, the open, high, low and close were dropped. The data file was saved as crude.csv. This file would be used for the future data exploration. There is negative crude oil data. Since the SARIMAX model could not have any value equal or below zero, the negative value was manually set to 0.001.

Data Exploration

There are two sections of data exploration. The first one is the feature comparison. The second one is to focus the stock and crude future value with ARIMAX model.

I. Feature comparison

As we mentioned, the fluctuation of crude oil has impact on the energy industry. We will use the feature comparison to investigate the impact of the crude oil on the industry.

1. Import data

A batch files reader was created to import stock data through different data files. Then the data was stored in a dataframe (df) for the further processing. The crude oil data was also imported to the crude oil dataframe (df\_crude) for further processing. The example of the data is shown in Table 5.

XOM close price XOM volume

time

2018-11-12 72.759682 7.195125e+05

2018-11-13 71.271938 1.053756e+06

2018-11-14 70.263956 1.168918e+06

2018-11-15 70.104237 1.168161e+06

2018-11-16 71.184835 9.932945e+05

BP close price BP volume

time

2018-11-12 36.630352 412748.142857

2018-11-13 35.667207 673956.375000

2018-11-14 35.828752 544081.600000

2018-11-15 36.143733 499967.800000

2018-11-16 36.359525 405950.200000

….

The detail data will be shown in the Appendix. The stock data time range was from 11/18/2018-10/30/2020.

2. Data exploration

After the data has been imported, the data was organized by integration the crude data with stock data. The method was to use concat the two dataframes with the column.

The data in crude oil will merge to the stock data with some missing values. The missing values were dropped by dropna(). A new dataframe was created.

setting the time as index. Then the time was set to the timeseries. The stock close values and stock volume were separated into two dataframe (with volume or stock)

3. Initial heatmap

The initial heatmap of the crude oil vs. different stock close price is shown in Fig. 1

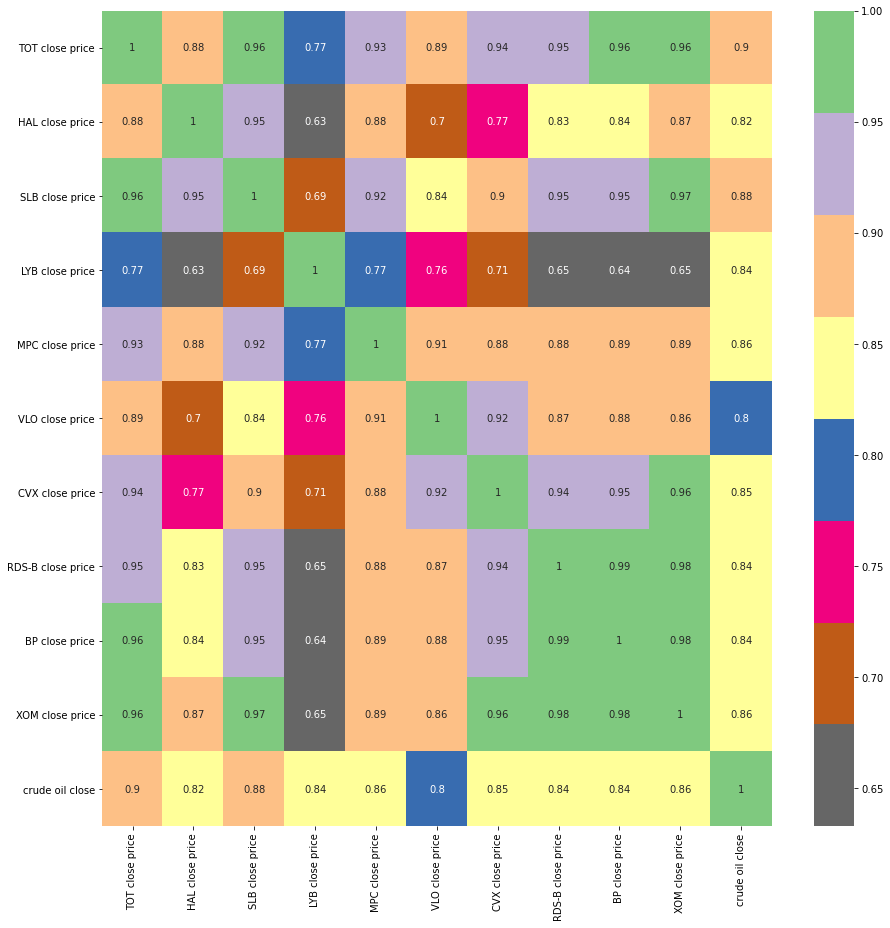


Fig.1. Heatmap of different features

The heatmap indicated that the all the energy stocks related to the crude oil price. The detail discussion will be presented in the following sections。

4. XOM stock price vs. crude oil close price

The XOM stock price vs. crude oil close price is shown in Fig. 2.

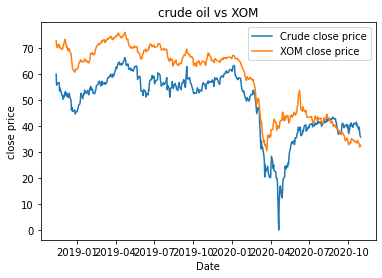


Fig. 2. Crude oil vs. XOM stock price

Before the Apr. 2020, the XOM stock price could trend crude oil close price. However, after the Apr. 2020, the XOM stock price still fell down even if the crude oil price became stable. The results indicated that other factors would affect the XOM stock price besides the crude oil price. Therefore other factors will be needed considered.

5. Crude oil price vs. BP stock price

The BP stock price vs. crude oil price is shown in Fig. 3.

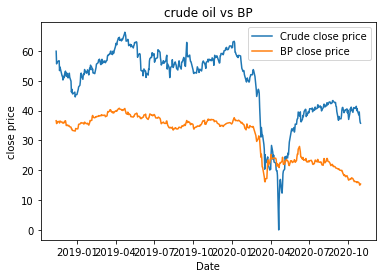


Fig.3. BP stock price vs. crude oil price

The graph showed similar trend to the XOM vs. crude oil price. The BP stock price trends well with crude oil price before Apr. 2020. Then both of them showed an opposite direction. The results indicated that other factors affect BP stock price as XOM.

6. RDS-B stock price vs. crude oil price

The RDS-B stock price vs. crude oil price is shown in Fig.4.

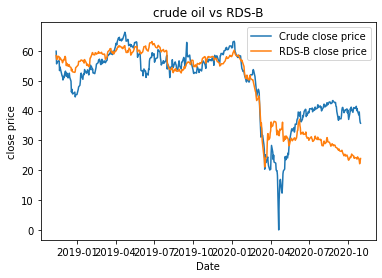


Fig. 4. The RDS-B stock price vs. crude oil price

The graph showed similar trend to the XOM vs. crude oil price. The RDS\_B stock price trends well with crude oil price before Apr. 2020. Then both of them showed an opposite direction. The results indicated that other factors affect RDS-B stock price as XOM.

7. CVX stock price vs. crude oil price

CVX stock price vs. crude oil price is shown in Fig. 5.

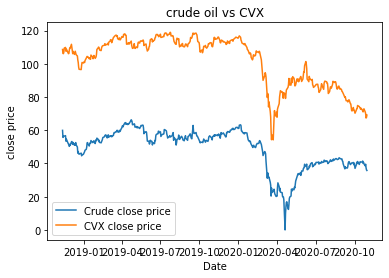


Fig. 5. CVX stock price vs. crude oil price

The graph showed similar trend to the XOM vs. crude oil price. The stock price trends well with crude oil price before Apr. 2020. Then both of them showed CVX an opposite direction. The results indicated that other factors affect CVX stock price as XOM.

7. VLO stock price vs. crude oil price

The Valero is a pure refinery company at downstream. There is no petrochemical section in this company.VLO stock price vs. crude oil price is shown in Fig. 6.

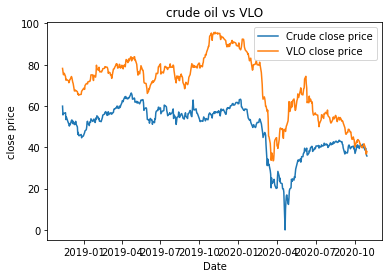


Fig. 6. VLO stock price vs. crude oil price

The graph showed similar trend to the XOM vs. crude oil price. The stock price trends well with crude oil price before Apr. 2020. Then both of them showed VLO an opposite direction. The results indicated that other factors affect VLO stock price as XOM.

8. MPC stock price vs. crude oil price

The Marathon Petrolum Corp. is a pure refinery company at downstream. There is no petrochemical section in this company. MPC stock price vs. crude oil price is shown in Fig. 7.

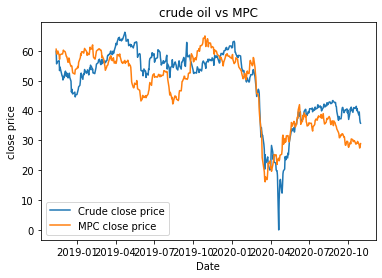


Fig. 7. MPC stock price vs. crude oil price

The graph showed similar trend to the MPC vs. crude oil price. The stock price trends well with crude oil price before Apr. 2020. Then both of them showed MPC an opposite direction. The results indicated that other factors affect MPC stock price as XOM.

9. LYB stock price vs. crude oil price

The LyondellBasell is a typical petrochemical company at downstream. The company is using products from refinery to produce different raw materials for the specialty chemicals. LYB stock price vs. crude oil price is shown in Fig. 8.



Fig. 8. LYB stock price vs. crude oil price

The graph showed similar trend to the LYB vs. crude oil price. The stock price trends well with crude oil price in these data. The results indicated that crude oil price really affect LYB stock price.

10. SLB stock price vs. crude oil price

The Stumberger is the largest oil field service company. It provided services for the oil field application. SLB stock price vs. crude oil price is shown in Fig. 9.

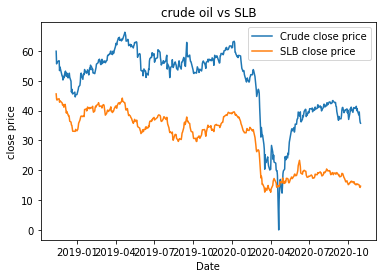


Fig. 9. SLB stock price vs. crude oil price

The graph showed similar trend to the SLB vs. crude oil price. The stock price trends well with crude oil price in these data. The results indicated that crude oil price really affect SLB stock price.

11. HAL stock price vs. crude oil price

The Haliburton is the second largest oil field service company. It provided services for the oil field application. HAL stock price vs. crude oil price is shown in Fig. 10.

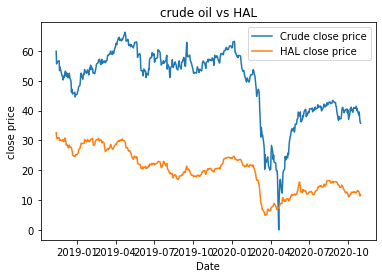


Fig. 10. HAL stock price vs. crude oil price

The graph showed similar trend to the HAL vs. crude oil price. The stock price trends well with crude oil price in these data. The results indicated that crude oil price really affect HAL stock price.

12. TOT stock price vs. crude oil price

The Total is a petrochemical company. It use the refinery products and making specialty chemicals. TOT stock price vs. crude oil price is shown in Fig. 11.

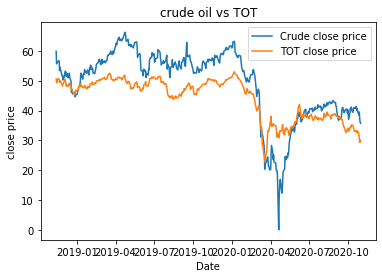


Fig. 11. TOT stock price vs. crude oil price

The graph showed similar trend to the LYB vs. crude oil price. The stock price trends well with crude oil price in these data. The results indicated that crude oil price really affect LYB stock price.

12. Conclusion

Although some experts said that the stock of energy company related to the crude oil price. Unfortunately, the data analysis did not support the statement. The result showed that the stocks of petrochemicals and oil field service company affected by the crude oil company. The stocks of the upstream companies and refinery companies did not appear to trend with the crude oil price, especially after April 2020. There would be some other key factors affect the stock price of these types of companies. I suspected that the nationwide lockdown and unemployment, which caused by COVID-19 pandemic, may be significant factors to affect the stock price. Therefore, more work will be needed to validate the hypothesis.

II. Stock and Crude Oil Forecast

The second part of this capstone project is to predict the future stock price and crude oil price. The program in this part was a separate program. Therefore, the data would be imported again through .csv file. A batch process was used to process all the data at the same program. A function, which contained seasonal decomposition, Dicky-Fuller test , was established to process signal data set. The model SARIMAX model. There were several pre-processing steps before this model could be applied.

1. Data import

The stock data and crude oil data were imported to the program through pandas. Then the time was set to the index and changed to the time series for further process. The stock time is from 2018-11-12 to 2020-10-30.

2. Seasonal effect

The time series data normally have the seasonal effect. In this project, the cycle of the data was five days period. Therefore we removed seasonal effect by setting frequency as 5. The data was decomposed to three parts:

* Trend
* Seasonal
* Residual

The trend data will be used for future simulation. The decomposed trend data are shown in the following figures. The trend and original data fits well in all the figures.

a. TOT stock price

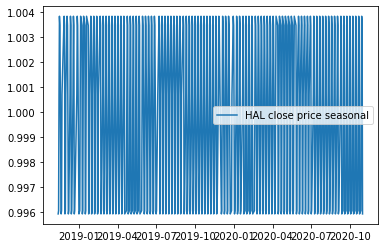
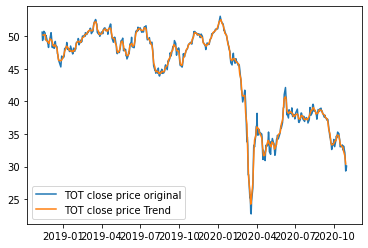


Fig. 12 a. TOT Trend vs. Original Fig. 12b. TOT Seasonal

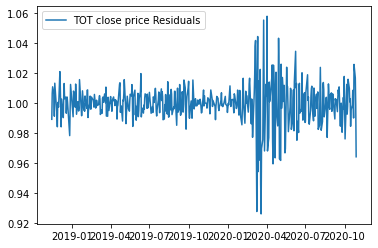


Fig. 12c. TOT Residuals

b. HAL stock price

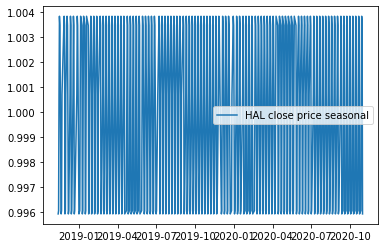
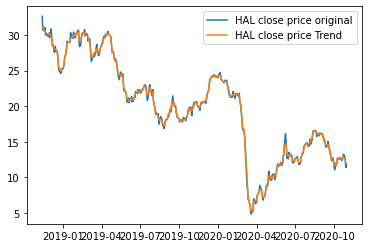


Fig. 13a. HAL Trend vs. Original Fig. 13b. HAL Seasonal

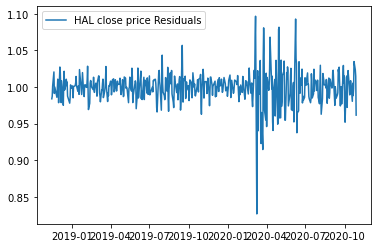


Fig. 13c. HAL Residuals

c. SLB stock price

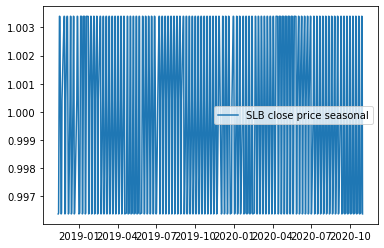
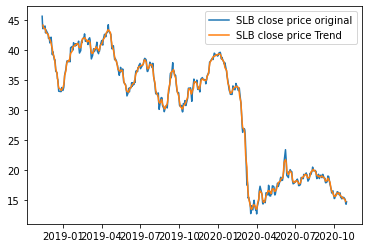


Fig. 14a. SLB Trend vs. Original Fig. 14b. SLB Seasonal



Fig. 14c. SLB Residuals

d. LYB stock price

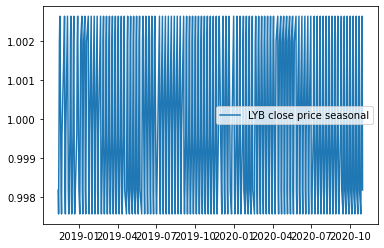
 

Fig. 15a. LYB Trend vs. Original Fig. 15b. LYB Seasonal

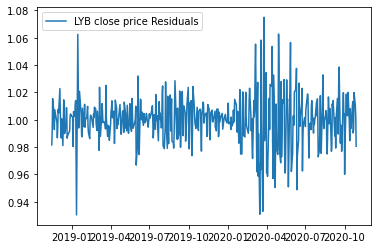


Fig. 15c. LYB Residuals

e. MPC stock price

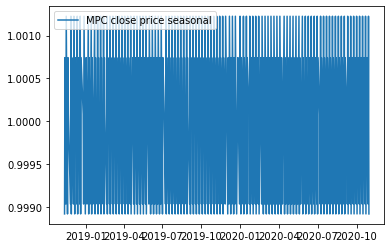
 

Fig. 16a. MPC Trend vs. Original Fig. 16b. MPC Seasonal

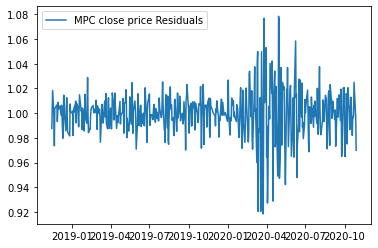


Fig. 16c. MPC Residuals

f. VLO stock price

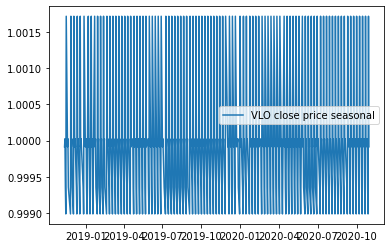
 

Fig. 17a. VLO Trend vs. Original Fig. 17b. VLO Seasonal

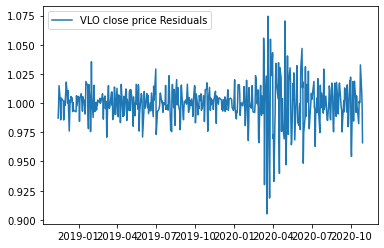


Fig. 17c. VLO Residuals

g. CVX stock price

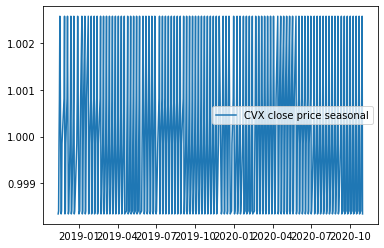
 

Fig. 18a. CVX Trend vs. Original Fig. 18b. CVX Seasonal

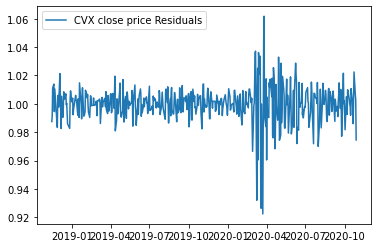


Fig. 18c. CVX Residuals

h. RDS-B stock price

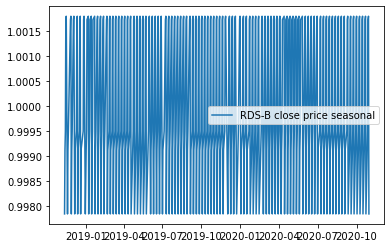
 

Fig. 19a. CVX Trend vs. Original Fig. 19b. CVX Seasonal

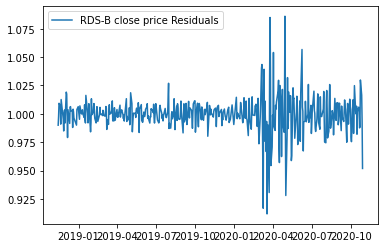


Fig. 19c. CVX Residuals

i. BP stock price

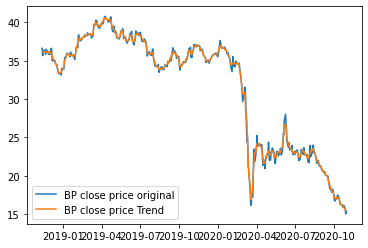
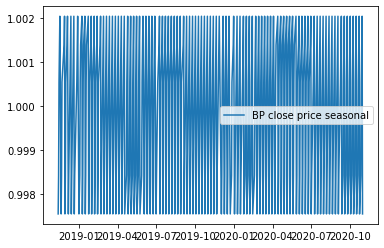
 

Fig. 20a. BP Trend vs. Original Fig. 20b. BP Seasonal

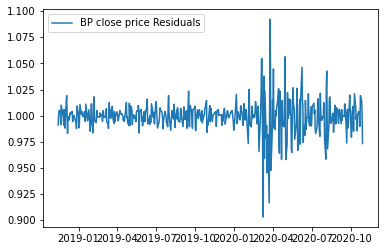


Fig. 20c. BP Residuals

j. XOM stock price

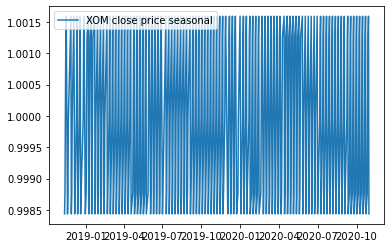
 

Fig. 21a. XOM Trend vs. Original Fig. 21b. XOM Seasonal

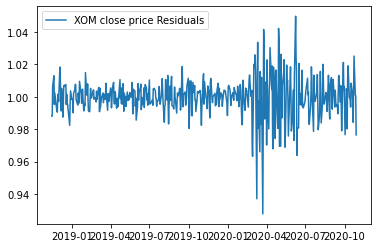


Fig. 21c. XOM Residuals

k. Crude oil price

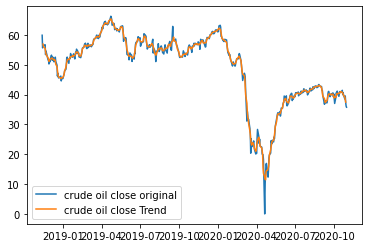
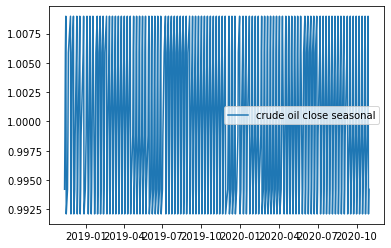
 

Fig. 22a. Crude Oil Trend vs. Original Fig. 22b. Crude Oil Seasonal

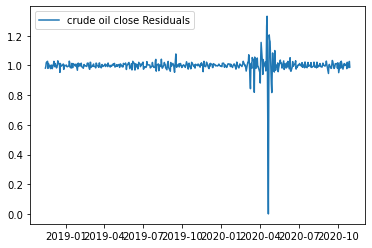


Fig. 22c. Crude Oil Residuals

l. Summary

After the data was decomposed by the seasonal frequency, the trend data still fit well with the original data. However significant fluctuations of the residuals were observed after April, 2020, except crude oil.

2. Stationary test

To better simulate the time series data, stationary data will be needed. There are several ways to test the stationary of the data. In this project, Dicky-Fuller method was used.

In [statistics](https://en.wikipedia.org/wiki/Statistics), the **Dickey–Fuller test** tests the [null hypothesis](https://en.wikipedia.org/wiki/Null_hypothesis) that a [unit root](https://en.wikipedia.org/wiki/Unit_root) is present in an [autoregressive](https://en.wikipedia.org/wiki/Autoregressive) model. The [alternative hypothesis](https://en.wikipedia.org/wiki/Alternative_hypothesis) is different depending on which version of the test is used, but is usually [stationarity](https://en.wikipedia.org/wiki/Stationarity_(statistics)) or [trend-stationarity](https://en.wikipedia.org/wiki/Trend-stationary_process).

The Dickey-Fuller function (adfuller) is used by passing individual trend data. If p< 0.05, then the data was stationary. Otherwise, it was no stationary. If the data was not stationary, a conventional way would be used by subtract two adjunct data to get the differences. The process would continue until p< 0.0005. The test results are shown in Table 6. The data showed that the preprocessed data became stationary after subtracting two adjunct data. After the data became stationary, the SARIMAX could be used to fit the data. Then the potential future data could be forecast based on the SARIMAX model.

Table 6. Dickey-Fuller Results

|  |  |  |
| --- | --- | --- |
| Name | Original p after Dickey-Fuller | difference p after Dickey-Fuller |
| TOT | 0.7121961596133362 | 0.00010949853259523735 |
| HAL | 0.5650465171470276 | 5.105473334645938e-05 |
| SLB | 0.8589146487819077 | 1.243188092010578e-06 |
| LYB | 0.31509906810099575 | 2.8558975649285265e-06 |
| MPC | 0.6648613976669153 | 1.715318981275725e-05 |
| VLO | 0.6151808383511976 | 0.0003475260930178213 |
| CVX | 0.8748669531299194 | 1.5161123143924243e-05 |
| RDS-B | 0.9389641476876494 | 0.00010988372090742102 |
| BP | 0.9537128114799808 | 0.0002254592115586212 |
| XOM | 0.9307096470018053 | 9.546850797996723e-05 |
| Crude Oil | 0.509538595868619 | 0.0003070874348606496 |

3. SARIMAX model

The SARIMAX model is used to fit and process time series data. It has three parameters p, d, q. These parameters needed to be optimized to get the best fit results. The optimization p,d,q was range from (0-10). The maximum AIC based on the model would be picked as the best parameters. In this project, a function was created for the batch processing of different data. The test results are shown in the Table 7.

Table 7. Optimized Parameters with Maximum AIC Results

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Name | p | d | q | AIC | BIC |
| TOT | 2 | 0 | 1 | -181.35 | -164.61 |
| HAL | 2 | 0 | 1 | -553.77 | -537.03 |
| SLB | 2 | 0 | 2 | -256.32 | -237.39 |
| LYB | 2 | 0 | 2 | 645.09 | 666.02 |
| MPC | 2 | 0 | 1 | 148.70 | 165.44 |
| VLO | 2 | 0 | 0 | 535.7 | 548.26 |
| CVX | 2 | 0 | 1 | 611.51 | 628.25 |
| RDS-B | 2 | 0 | 1 | -48.72 | -31.98 |
| BP | 2 | 0 | 1 | -424.52 | -407.77 |
| XOM | 2 | 0 | 1 | -56.81 | -40.06 |
| Crude Oil | 2 | 0 | 2 | 569.00 | 589.93 |

By using the optimized p, d , q, the preprocessed data was fit to the SARIMAX model, the fitting results are shown in the following figures.

a. TOT

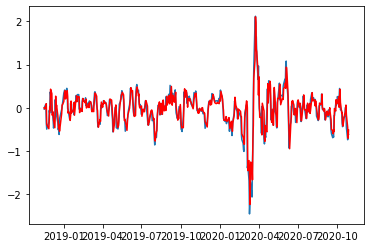


Fig. 23. TOT SARIMAX model fitting vs. original preprocess daa

b. HAL

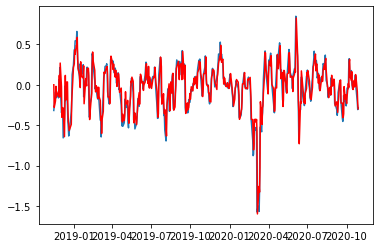


Fig. 24. HAL SARIMAX model fitting vs. original results

c. SLB

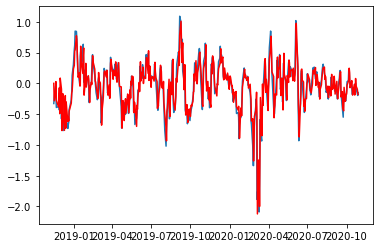


Fig. 25. SLB SARIMAX model fitting vs. original results

d. LYB

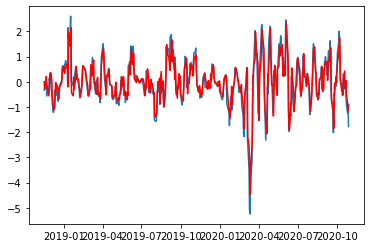


Fig. 26. LYB SARIMAX model fitting vs. original preprocess daa

e. MPC

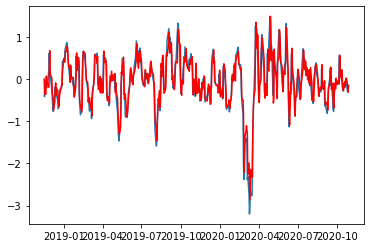


Fig. 27. MPC SARIMAX model fitting vs. original results

f. VLO

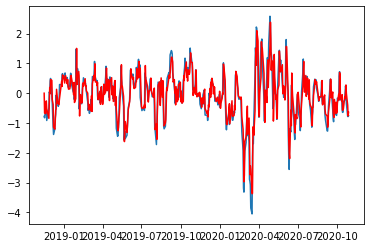


Fig. 28. VLO SARIMAX model fitting vs. original results

g. CVX

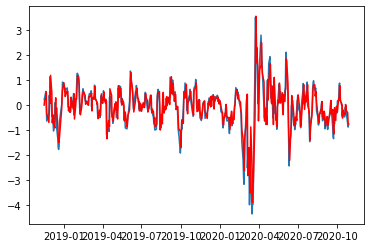


Fig. 29. CVX SARIMAX model fitting vs. original preprocess daa

h. RDS-B

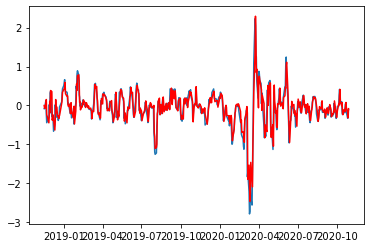


Fig. 30. RDS-B SARIMAX model fitting vs. original results

i. BP

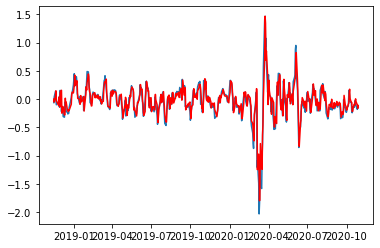


Fig. 31. BP SARIMAX model fitting vs. original results

j. XOM

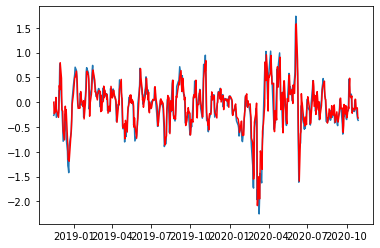


Fig. 32. XOM SARIMAX model fitting vs. original results

k. Crude Oil

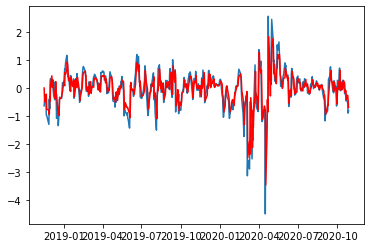


Fig. 33. Crude Oil SARIMAX model fitting vs. original results

l. Conclusion

The results showed that SARIMAX fit all the preprocessed data well, except crude oil.

4. Forecast of the stock price

The SARIMAX model provided the prediction function to predict the potential stock price change. In this project, we used 60 days prediction for the data. The predicted data was generated based on the preprocessed stationary data. Then the results were calculated back to the original stock price. Since the data did not contain any of the date information. A separate time series data array was created and merged to the forecast data. The results of individual data were shown in the following figures.

a. TOT

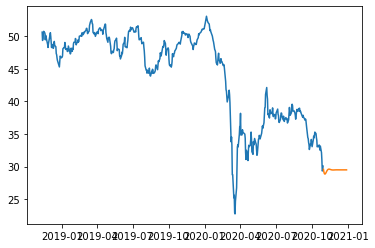


Fig.34. TOT Forecast Stock Price

b. HAL

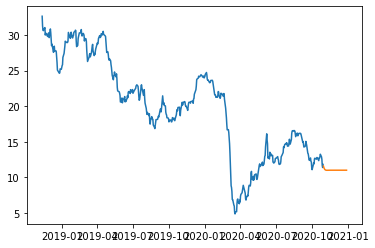


Fig.35. HAL Forecast Stock Price

c. SLB

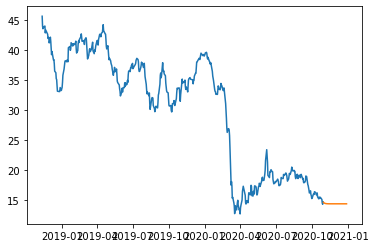


Fig.36. SLB Forecast Stock Price

d. LYB

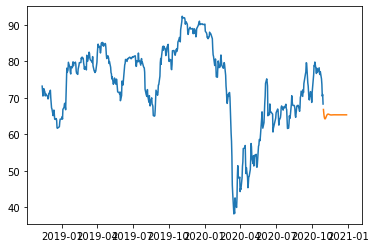


Fig.37. LYB Forecast Stock Price

e. MPC

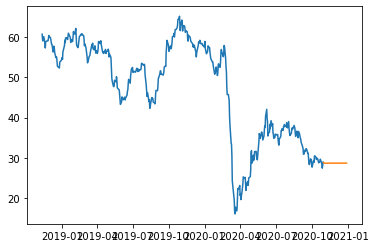


Fig.38. MPC Forecast Stock Price

f. VLO

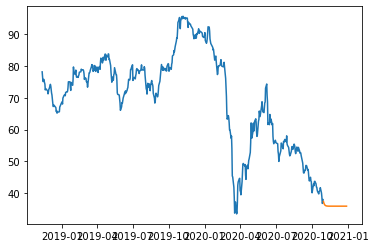


Fig.39. VLO Forecast Stock Price

g. CVX

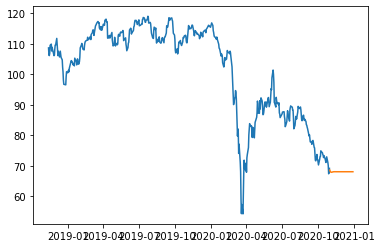


Fig.40. CVX Forecast Stock Price

h. RDS-B

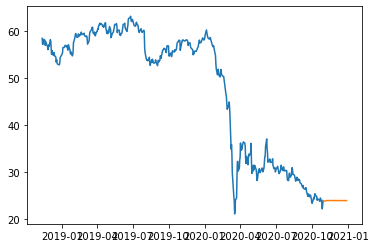


Fig.41. RDS-B Forecast Stock Price

i. BP

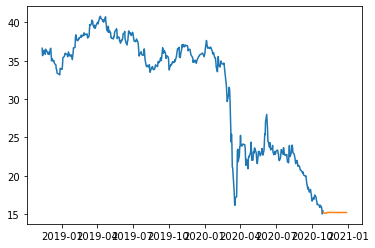


Fig.42. BP Forecast Stock Price

j. XOM

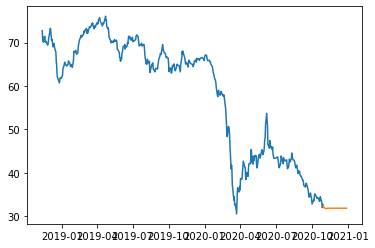


Fig.43. XOM Forecast Stock Price

k. Crude Oil

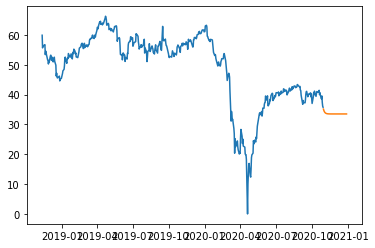


Fig.44. Crude Oil Forecast Stock Price

l. Conclusion

The data showed that all the stocks were unlikely to decrease within the next 60 days of period. It became stable after then.

Conclusion

In this project, several energy stocks were compared with crude oil price. The results indicated that the petrochemical companies and oil field service stocks trend the crude oil price. However, the upstream companies and refinery companies did not seem to trend the crude oil price. Therefore, some other factors also affected the stocks of these companies.

The SARIMAX model was used to fit the data. Then the forecast stock prices were predicted based on the model. The model showed that these stock prices would not decreased further in the next 60 days.