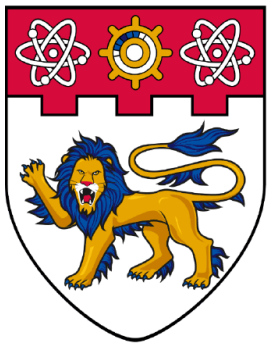


NANYANG TECHNOLOGICAL UNIVERSITY
SINGAPORE

AY 2023/24 Semester 1
BC2406 Analytics 1: Visual and Predictive Techniques

Group Project
Business Analytics Proof of Concept (POC) for Aramco

Seminar Group: S03
Group: Team 6
Date of Submission: 5/11/2023
Tutor: Prof. Josephine Zhou

Name	Matriculation Number
Tan Yue Hui	U2221209K
Kong Jie Hao	U2222806L
Aung Kaung Kaung	U2221195J
Lim Jia Jie, Isaac	U2222066A
Leong See Neng Shannon	U2222007D
Gao Wenjie	U2221990H

Executive Summary

Opportunity Statement

The global energy landscape is inextricably linked to the price of crude oil, which accounts for one-third of worldwide energy consumption (EIA, 2023). Volatility in crude oil prices can have far-reaching implications for global economies, making accurate price forecasting a crucial tool for mitigating financial risks. In this context, we are focusing on creating predictive analytics for oil price. The opportunity lies in leveraging data-driven insights to enhance Aramco's decision-making processes, optimise resource allocation, and navigate the dynamic oil market with increased agility.

Proof of Concept

Our team addressed several issues within the dataset such as missing data, redundant variables, and outliers. Redundant variables such as GSCI, were removed from the dataset, and missing data and outliers were replaced with the variable's monthly means, ensuring consistency and reliability data.

In the data exploration stage, the team uncovered valuable insights about the WTI closing price. Notably, the team observed a positive correlation between WTI closing price and the variable Inflation_5Y_BE, while a negative correlation was found with the variable DXY (US Dollar Index). Additional observations revealed that variables like Federal_Funds_Effective, Economic_Uncertainty_Index, US_3M_Treasury, and US_1Y_Treasury exhibited positive skewness, while US_Oil_Demand and Demand_Less_Supply showed negative skewness. These insights played a pivotal role in understanding the underlying data dynamics.

To evaluate the model's overall accuracy and variability, the team compared the Random Forest model with linear regression, CART, and ARIMA models. After a comprehensive evaluation, our team concluded that the random forest model yielded the most optimal results in terms of accuracy and variability, making it the most suitable model for predicting the closing price of WTI crude oil spot. The key variables of highest importance for the random forest model in predicting the closing price of WTI crude oil spot were identified as follows: Inflation_5Y_BE, SP500, NASDAQ, DJIA, DXY, Demand_Less_Supply, and 10Y_Less_2Y. These variables played a significant role in the model's performance and highlighted the critical factors influencing the WTI spot price.

Insights & Solutions

Our project team recommends the adoption of the Random Forest model to Aramco due to its superior accuracy in predicting the closing price of West Texas Instrument (WTI) crude oil spot. However, it is essential for Aramco to maintain and recondition the model regularly with updated data. This is crucial to sustain the model's accuracy in predicting the closing price of WIT crude oil spot. Over time, the quality of predictions can fluctuate due to ever changing data patterns and external factors such as geopolitical events.

By leveraging on the capability to predict the closing price of WTI crude oil spot, Aramco can enhance their business operations through the implementation of strategies such as i) Optimising oil inventory management by utilising a dynamic inventory system for better resource allocation and minimise operation costs, ii) Predicting risk-hedging investments to facilitate the judicious utilisation of profitable future contracts, securing advantageous price swaps and extensive capitalisation of options trading, and iii) Proactive diversification for market expansion during stable oil prices environment and investing in renewable energy during volatile oil price seasons to contribute to environmental sustainability.

1. Introduction

1.1 Background on Saudi Aramco

Saudi Arabian Oil Company, also known as Saudi Aramco, is the leading oil producer in the world based in Dhahran, Saudi Arabia. Surpassing companies such as Apple and Alphabet's Google, Aramco is the most profitable company in the world. (DELVENTHAL, 2022) Pumping oil from the ground and selling it to the global export market, where the largest customers include countries such as the United States, India and China, is Aramaco's primary line of business. As an independent trader, Aramco also searches for crude oil, refines it into goods like chemicals and gasoline, and purchases and sells the petroleum of other businesses. The company is also responsible for the maintenance of Saudi Arabia's spare capacity. Generating approximately 2 million to 3 million barrels a day for the past few years has given Aramco the ability to stabilise prices and adapt to the demand quickly by controlling their supply easily. (Barnett et al., 2017)

1.2 Opportunity Statement

With a second-quarter net profit of $30 billion, Aramco has suffered a drop of nearly 40% compared to the same period in 2022. This decline is primarily attributed to the fall in hydrocarbon prices. The net income for the second-quarter of 2023 represents a 38% decrease from the $48.4 billion earned in the same quarter in 2022. During that period, the results for the second-quarter of 2022 was up 90% compared to 2021. This significant increase was due to the spike in energy prices caused by the Russian war in Ukraine. (Murphy & Turak, 2023)

Through incorporation of data-driven technology into its business operations, there is substantial opportunity for Aramco to significantly increase its earnings and ultimately improve the company's bottom line. By strategically using data-analytic methodologies, we aim to help Aramco revolutionise big data into actionable insights which can facilitate its efforts in achieving more reliable predictions of market fluctuations, earlier forecasting of future business patterns, smoother optimization of its supply chain and more effective implementation of tactical pricing decisions. Therefore, by taking advantage of data-driven decision-making, Aramco can seize more lucrative opportunities to maximise its expected profits and strengthen its financial foothold in the competitive global oil market.

1.3 Key Business Questions

In this report, we will aim to solve several key business questions related to demand-supply balance, market dynamics and risk management.

Demand-Supply Balance:

- How can Aramco optimise the balance between oil production and market demand to maximise profitability?

Market Dynamics:

- What are the leading indicators of oil price fluctuations, and how can they be incorporated into predictive models?

Risk Management:

- How can predictive analytics help in identifying and mitigating risks associated with volatile oil prices?
- What strategies can be employed to manage uncertainties and ensure business continuity?

2. Data Cleaning and Preprocessing

Our dataset was obtained from Kaggle, and it contains 2555 rows and 16 columns.

There are 16 continuous variables in the dataset. The variable *WTI_Spot* represents the current spot price for WTI crude oil, a crucial benchmark for crude oil pricing. This variable, *WTI_Spot*, serves as the independent variable that we aim to predict accurately. The remaining variables encompass the date, 5 stock prices and 9 economic indicators (Refer to Appendix A).

2.1 Removal of GSCI

The column *GSCI* was removed as *WTI_Spot*, which is the variable we aim to predict, is one of the components that makes up *GSCI* (Refer to Table B.1 in Appendix B). This leads to multicollinearity and redundancy in our model. If we do not remove *GSCI*, it may lead to inflated importance of *GSCI*. This could result in a misleading representation of the true relationships between variables.

2.2 Missing values

We found that the columns with missing data are: WTI_Spot, SP500, NASDAQ, DJIA, DXY, Federal_Funds_Effective, Economic_Uncertainty_Index, Inflation_5Y_BE, US_3M_Treasury, US_1Y_Treasury and X10Y_Less_2Y.

2.2.1 Replacing missing values

We decided to replace the missing data with the **mean of each month (Month-wise Averaging).** This way, we can **preserve the seasonality and trends** as it accounts for the potential monthly variations and is more contextually relevant in a time series dataset.

To perform month-wise averaging, we grouped our data by months and calculated the mean of the data points within that month, and replaced the missing values with the mean.

2.3 Outlier Detection and replacing them

The **Interquartile Range (IQR)** is a robust statistical method commonly used for identifying outliers in a dataset. Outliers are values that fall significantly below Q1 or above Q3, this indicates that they are unusually high or low when compared to the majority of the data.

- Lower Bound: Q1 - (1.5 * IQR)
- Upper Bound: Q3 + (1.5 * IQR)

We define the outliers as any data points that fall below the lower bound, or above the upper bound. After identifying the outliers, we replaced them with the **calculated monthly mean for that month**, similar to how we replaced the missing values.

2.4 Ensuring no duplicate date values

We ensured that there are no duplicates for *DATE* values by removing the duplicates and keeping the first occurrence of each *DATE* value. The dates should be unique as there should only be one entry per day for each market closing price.

3. Exploratory Data Analysis

3.1 Summary Statistics

For this project, we will be predicting the closing price of West Texas Intermediate (WTI) Crude Oil Spot (WTI_Spot), using price indexes, interest rate measures and economic indicator variables as seen in Table A.1. We will be performing machine learning techniques such as Linear Regression, Classification and Regression Tree (CART), Random Forest Models and Time-Series Analysis (ARIMA).

3.2 Univariate Analysis

Plotting boxplots, histograms and kernel density estimate (KDE) plots in Fig. A.4, A.5, A.6 visually illustrated each of the variables. We noticed a few variables with a long tail on the KDE plots, where we then calculated the skewness of our variables.

In Fig. A.7, we can observe that a few of our variables, namely Federal_Funds_Effective, Economic_Uncertainty_Index, US_3M_Treasury and US_1Y_Treasury were positively skewed while US_Oil_Demand and Demand_Less_Supply were negatively skewed.

3.3 Bivariate Analysis

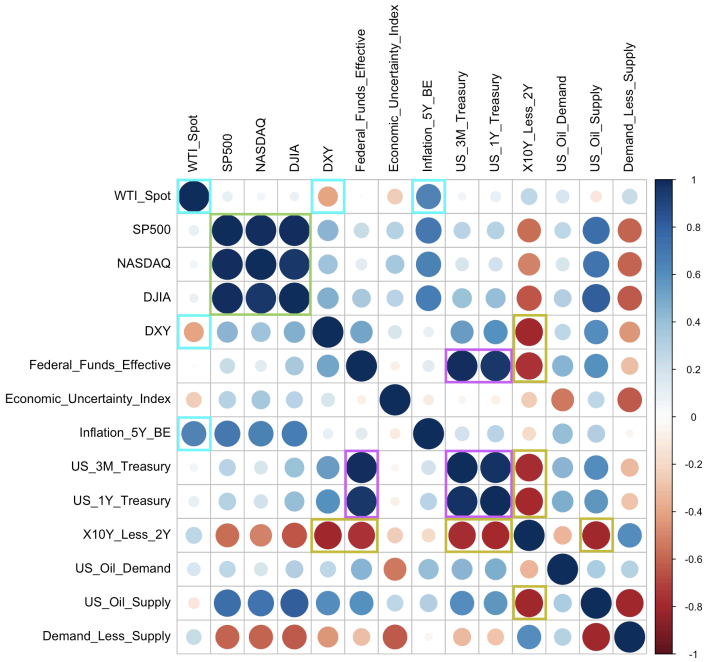


Fig. 1: Correlation Matrix

Correlation between WTI_Spot, Inflation_5Y_BE and DXY (Blue)

In Fig. 1, WTI_Spot has a noticeable positive correlation with Inflation_5Y_BE, and a negative correlation with DXY.

For the former, higher WTI_Spot, crude oil prices, can lead to inflation through the increase of direct energy costs or indirectly by raising the costs of goods manufactured or transported with using oil. The Federal Reserve estimates that a $10 increase in the price of crude oil can raise inflation by 0.2% and lower economic growth by 0.1% (Lioudis, 2023).

As for the latter, when DXY, or the US dollar, strengthens, WTI_Spot, or crude oil prices, tend to fall. This inverse correlation is largely due to the petrodollar system, where global oil transactions are conducted in US dollars (Barnes & Cingari, 2022).

Correlation between SP500, NASDAQ and DJIA (Green)

There are also a few prominent relationships, such as the high positive correlation between SP500, NASDAQ and DJIA. All three price indexes are measures of market performance on any given day, and generally move in the same direction (Sharma, 2022).

Correlation between US_3M_Treasury, US_1Y_Treasury and Federal_Funds_Effective (Purple)

The US_3M_Treasury and US_1Y_Treasury is also observed to have a high positive correlation between each other and Federal_Funds_Effective primarily due to interest rate expectations. Federal funds rate is the short-term interest rate that is directly influenced by the Federal Reserve, where changes in the federal funds rate are often seen as signals of the central bank's monetary policy stance. As a result, the federal funds rate and the rate on short-term treasuries often track each other closely (Martin, 2001).

Correlation between X10Y_Less_2Y, DXY, Federal_Funds_Effective, US_3M_Treasury, US_1Y_Treasury and US_Oil_Supply (Yellow)

X10Y_Less_2Y has a high negative correlation with multiple variables: DXY, Federal_Funds_Effective, US_3M_Treasury, US_1Y_Treasury, US_Oil_Supply.

When the spread between the 10-Year and 2-Year US Treasuries yields (X10Y_Less_2Y) widens, it could signal to investors about concerns regarding future economic growth. This leads to investors moving to other currencies and assets, leading to a decrease in DXY (Chen, n.d.).

On the other hand, the negative correlation with Federal_Funds_Effective is a reflection of the market expectations with regards to the economy and the Federal Reserve's monetary policy, which is reflected by the Federal_Funds_Effective. When the market anticipates an economic recession the spread between 10-year and 2-year treasury yields tend to decrease, while the Federal Funds Rate is adjusted in response to these changing economic conditions (Chen, 2023).

The high negative correlation of X10Y_Less_2Y and short-term treasury rates (US_3M_Treasury, US_1Y_Treasury), suggests that changes in short-term interest rates are affecting the slope and shape of the yield curve. If short-term rates rise significantly compared to long-term rates (increase in X10Y_Less_2Y), it could imply the market's expectation of a tightening monetary policy or a potential economic slowdown, influencing the yield curve's shape and impacting various financial markets.

Finally, the supply of oil (US_Oil_Supply), can impact the broader economy and financial markets. A significant increase in oil supply might be associated with lower inflation expectations, which can affect interest rates. These lower inflation expectations could lead to lower long-term interest rates, such as the 10-year treasury rate, resulting in a fall in X10Y_Less_2Y.

4. Data Modelling & Performance Evaluation

There are also other relationships with high correlations present but they are, however, not as significant as they are affected by numerous external variables as well and the high correlation does not indicate any form of direct causation.

4.1 Methodologies and Considerations

We will be using the Linear Regression, CART, Random Forest Models and ARIMA (Time-series Analysis) in our project to predict the outcome variable *WTI_Spot*. The dataset was split into 70-30, for training and testing. The purpose of the train-test split is to evaluate the performance of our machine learning model, as we want to know how well our model performs on new, unseen data.

To evaluate the accuracy of each of our models, the RMSE (Root mean squared error) and R-Squared (Goodness of Fit) was compared between the regression models. Our team then evaluated and selected the appropriate model that will help us predict oil prices with higher accuracy.

4.2 Linear Regression Model

4.2.1 Optimisation of Linear Model via AIC minimisation

We initially fitted the linear model on the training set to predict WTI_Spot using all X-variables, obtaining a full linear regression model. Afterwards, we used stepwise backwards elimination to choose our final optimal linear model, which was the one with lowest Akaike's Information Criterion (AIC) to minimise the estimated information loss of the model. (Refer to Fig. B.1 in Appendix B)

4.2.2 Model Performance Evaluation

4.2.2.1 Model Complexity

Fig. B.2 (Appendix B) shows a summary of our optimal linear regression model, where a total of 12 X-variables were used in the final equation. Our adjusted R-squared value, which is a measure of the model's goodness of fit that adjusts for degrees of freedom and penalises for high model complexity, has a high value (0.8772), and it is only slightly lower than its ordinary R-squared value, which shows that our linear model's complexity is still acceptable and not too excessive.

4.2.2.2 Overall Statistical Significance of Entire Model

The F-test is a form of statistical hypothesis testing in which the null hypothesis is that all regression coefficients are equal to zero. As seen in Fig. B.2 (Appendix B), our F-test's p-value is extremely small (<0.001), meaning we have sufficient evidence, even at a strict significance level of 1%, to reject the null hypothesis and conclude that there exists at least one non-zero regression coefficient. This means that our entire linear model is statistically significant, implying that there is a high probability that WTI_Spot price is indeed significantly affected by some of the X-variables.

4.2.2.3 Visual Analysis of Model's Accuracy and Goodness of Fit

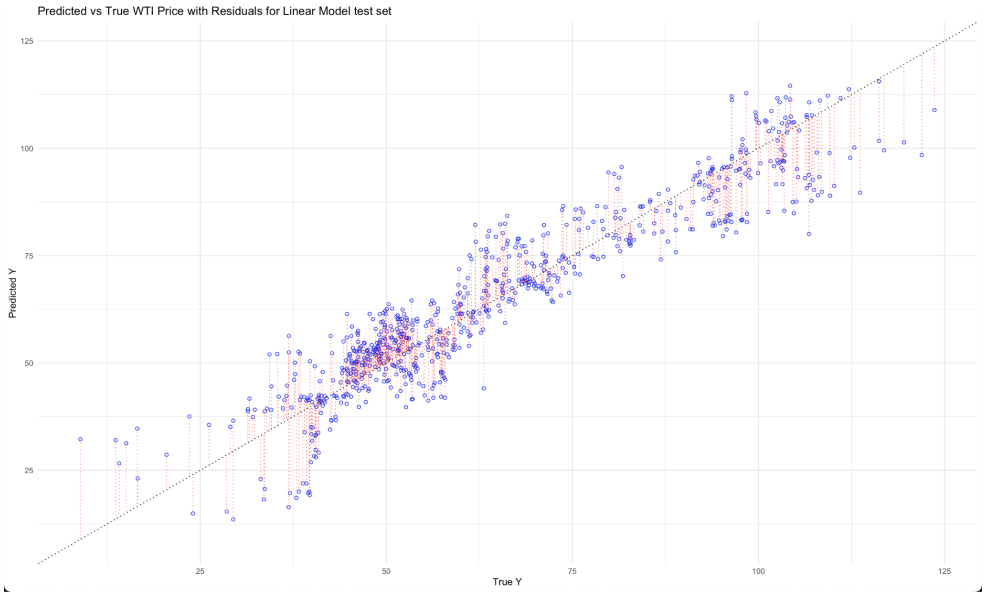


Fig. 2: Linear Regression Model Prediction with Scatter Plot

In the scatter-plots of predicted against true values for both training (Appendix B Fig B.3) and testing set (Fig. 2 above), we observe that all points are distributed evenly around the dotted control line denoting y=x, which represents the ideal-case scenario where the predicted value is exactly equal to the true value, and they are all within fairly close proximity to the control line, with the length of dotted red lines denoting the residuals, which is the difference between the predicted and true values. This shows that there is a fairly good fit, with moderately small magnitude of residuals and thus relatively small prediction error.

Fig. B.4 & B.4.5 (Appendix B) show line charts of the predicted and true WTI_Spot price over time for the training and testing sets respectively. We see that although the predicted values (red line) largely follow the general trend of true values (blue line), there are still some misalignments and it is not a perfect match. This suggests that our linear model has moderate prediction accuracy but there is definitely room for improvement.

4.2.3 Variable Importance

The t-test p-values of all regression coefficients are very small (<0.001) thus statistically significant. Fig. B.5 (Appendix B) shows a bar graph of the importance of each X-variable in our linear model, as measured by the absolute value of its regression coefficient's t-statistic, where higher magnitude indicates greater statistical significance of the variable. As highlighted by our model, DXY, SP500 and Inflation_5Y_BE are likely to be the most important factors determining WTI_Spot price.

4.3 Classification and Regression Tree (CART) Model

In our CART analysis, we used the rpart and rpart.plot packages to visualise the decision tree and identify key decision points.

4.3.1 Optimisation of CART model via complexity parameter pruning

The CART model offers the distinct advantage of automated 10-fold cross-validation, a technique that allows a robust estimation of our model. This approach involves dividing the dataset into 10 subsets and consecutively training and testing our model on various combinations of these subsets. Complexity parameter pruning of a fully grown regression tree enables us to choose the optimal tree that has close to the minimum cross-validation error, ultimately yielding a robust performance estimate. We also set the minimum split to be 100 to avoid overfitting of our model. After pruning the tree, the optimal tree has 19 terminal nodes, which is of acceptable model complexity.

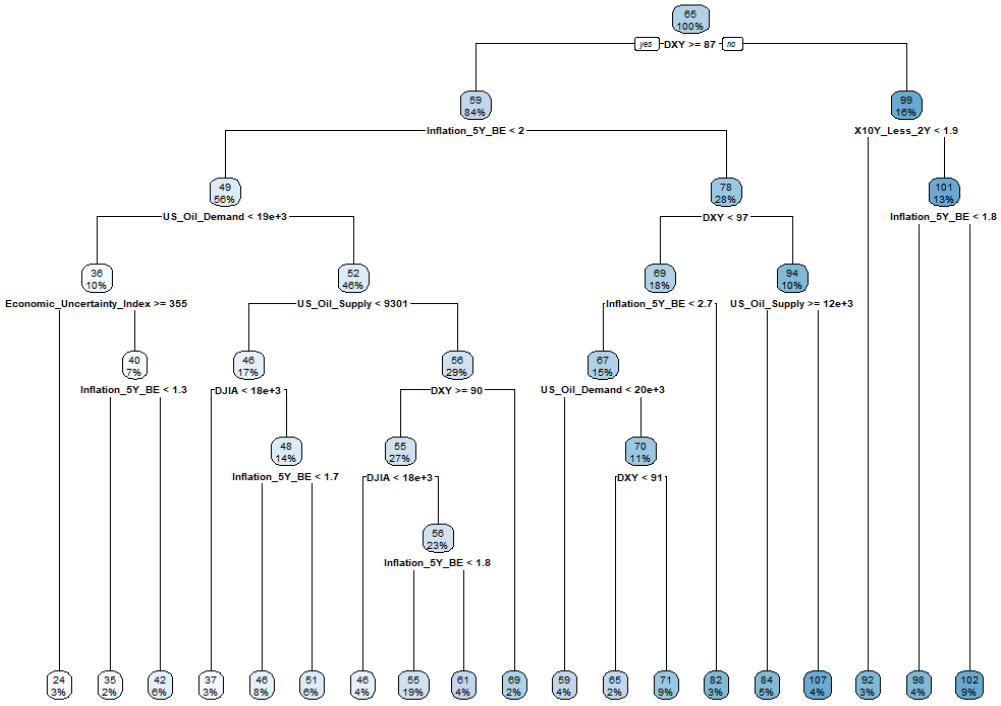


Fig. 3: Pruned Tree for Train Set

4.3.2 Variable Importance

In CART, importance of a variable is measured by the total sum of relative reduction in MSE (Mean Squared Error) due to splits by that variable. Hence, those variables that contribute more to MSE reduction are considered more important. Our optimal tree showed that the variables: NASDAQ, DXY, SP500, DJIA, X10Y_Less_2Y and US_Oil_Supply were the most important factors in predicting WTI_Spot Price. (see Table B.2 in Appendix B)

4.3.3 CART's Performance

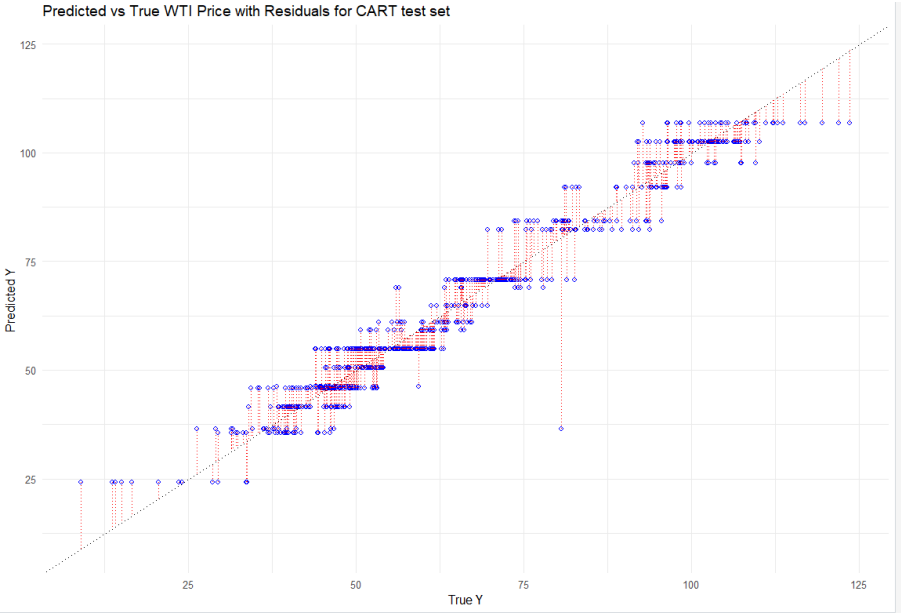


Fig. 4: CART Model Prediction with Scatter Plot

In a similar scatter-plot of results (Fig. 4 above and Fig. B.10 (Appendix B)), we can visually see that the data points are scattered more closely with the CART's control line compared to the Linear model's control line. This indicates that the residuals are relatively small and that the model is well-fitted. Similar line chart plots as retrieved from Fig. B.11 & B.11.5 (Appendix B) shows that the predicted values follow the true values more closely as compared to the Linear model.

Both plots affirm that the CART model is significantly **better-performing** than our linear model, with a much better fit and higher prediction accuracy.

4.4 Random Forest

4.4.1 Introduction of Random Forest

Random Forest is an ensemble learning method, which means it combines the predictions of multiple decision trees. Random Forest models tend to be more robust and less prone to overfitting because they combine the result from multiple trees, reducing the impact of variance of predictions (IBM, 2022) and providing more accurate predictions as compared to single decision trees. We used 500 trees.

4.4.2 Random Forest Performance

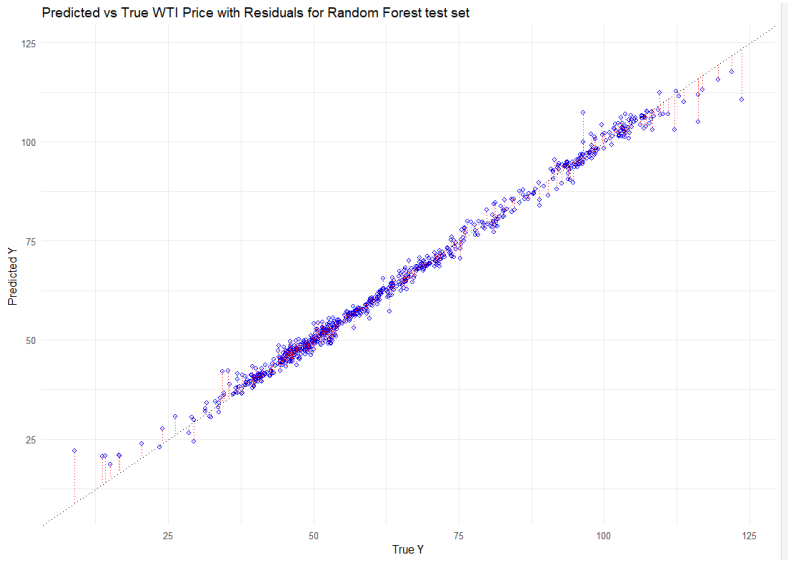


Fig. 5: Random Forest Model Prediction with Scatter Plot

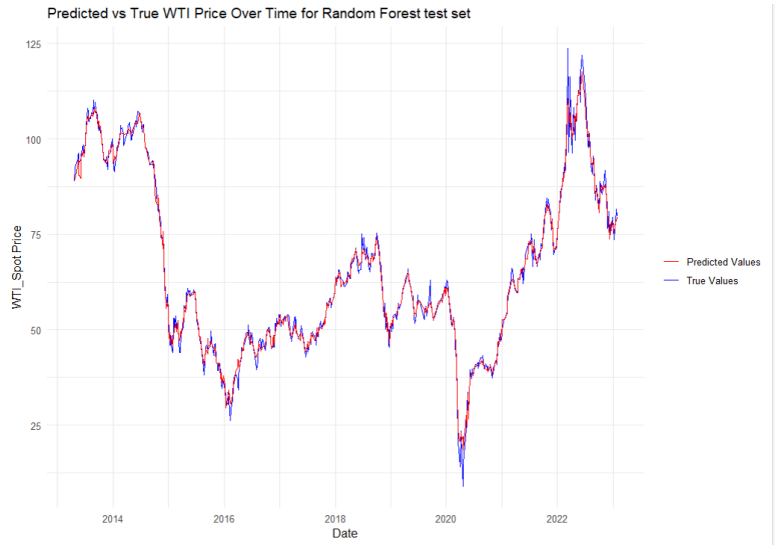


Fig. 6: Random Forest Model Predicted vs True Line Chart

In the scatter-plot (Fig. 5, above), the difference between the performance of Random Forest and the other 2 models is striking. The data points are much more tightly clustered around the predicted line, appearing to even wrap around it. Similarly, for the line charts (Fig. 6, above & Fig. B.12 (Appendix B)) the predicted values are almost perfectly overlapping the true values.

Such close alignment of predicted and true values suggests that the Random Forest model is performing exceptionally well in capturing the underlying patterns and relationships within the data, providing outstanding fit. This underscores its superior predictive accuracy, making it the strongest candidate for accurately modelling and forecasting WTI Spot price.

4.5 ARIMA Model — Time Series Analysis

4.5.1 Introduction of ARIMA Model

ARIMA, short for Auto-Regressive Integrated Moving Average, is used to predict data that varies across time. As the term autoregression implies, ARIMA aims to predict future data according to its past behaviour.

ARIMA is separated into three components. Firstly, autoregression of the model is represented by the order of the autoregressive model, p. Secondly, integration is applied which makes a time series stationary by subtracting from the present value by the previous values a number of times. Lastly, the process is regulated by degree of differencing d. Moving average computes the average residual errors from the previous sections. This removes noise from the data. It is determined by the order of the moving average, q. (Utkarsh, 2020)

The parameters of the model (p, q, d) can be determined by autocorrelation and partial autocorrelation plots. For our model, we automated the model by using the in-built R function auto-arima() which computes the process by selecting the best model after iterating through many graphs.

4.5.2 Prediction of the ARIMA Model

For the ARIMA model, we installed the forecast package. The auto.arima() function identifies the best combination of p, d and q. The graph can be retrieved from Appendix B. (Fig. 15) The blue line represents the mean stock price of the WTI stock price. The dark blue and grey regions represent the 85% and 90% probabilities of how the stock price data is likely to behave respectively. For greater accuracy and reliability, we used *WTI_Price* data three times to predict data for other time periods present in the dataset. (Fig. 19 to Fig. 23)

Company stock prices are affected by news, real-time events and company actions that influence how investors feel towards the stock. (Yang & Pan, 2022) Since people tend to gauge stock prices from events from recent years, we aim to include only stock price data from the past one to two years in our training set. According to industry norms, ARIMA is used to predict a maximum of six months in the future. (Hayes, 2023) Adhering to this rule, we only trained the model to predict 183 days ≈ six months.

4.5.3 Evaluation of Performance

For the evaluation of performance, we used the *checkResidual()* function that has in-built portmanteau tests, which are tests of accuracy that permits a wider range of errors.

(Fig. 16) Taking Ljung-Box chi-square statistics, we observe the p-value is large (> 0.05). The residual error of the data is insignificant and results from the fluctuations in stock prices. The error is described as ‘background noise’ of the dataset. (Minitab, 2023)

(Fig. 17) Since ARIMA relies on the lag errors as predictors, it works best when the predictors are not independent of each other. (EDUCBA, 2023) The autocorrelation (ACF) plot is able to measure the dependency of the residual errors. From the plot, most errors fall within the acceptable threshold lines. While there are one or two errors that fall outside of the threshold they can be attributed to random error. (Hyndman & Athanasopaulos, 2018)

Furthermore, from the topmost graph of residual errors, the errors are mostly uniform except towards the end. The last graph shows that errors are normally distributed.

Some limitations of the ARIMA model include assuming that past oil prices will have effect on future oil prices, being unable to predict sharp rises or falls in oil prices and the inability to consider seasonal trends. (Hayes, 2023) In fact, seasonal trends being present in the stock price data negatively affects the ARIMA model. This is mitigated through using the seasonal ARIMA model (SARIMA) into which we will not delve into further detail in this project.

4.6 Evaluation of All Models

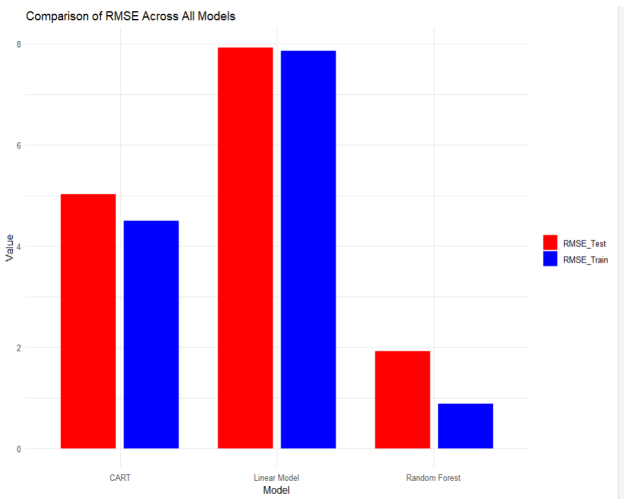
The evaluation metrics used are Root Mean Squared Error (RMSE) and R-Squared.

Model (Train Set)	RMSE	R-Squared
Linear Regression	7.856	0.878
CART	4.496	0.960
Random Forest	0.880	0.998

Table 1: Performance indices for test set of different models

Model (Test Set)	RMSE	R-Squared
Linear Regression	7.923	0.873
CART	5.027	0.949
Random Forest	1.917	0.993

Table 2: Performance indices for test set of different models



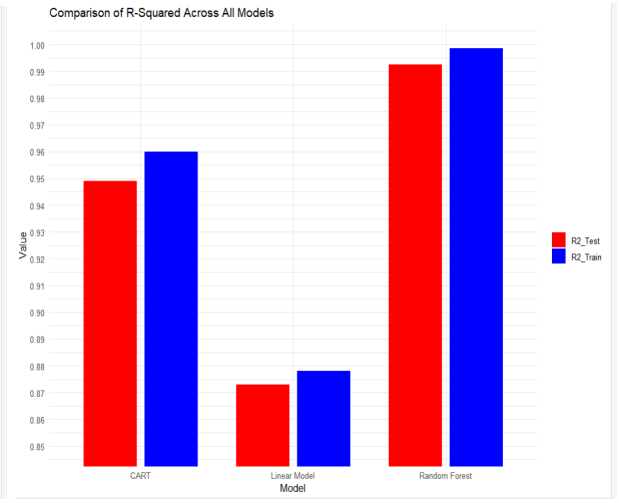


Fig. 7 & 8: Comparison of RMSE and R-Squared Across all Models

From the tables and figures above, we have concluded that Random Forest is the most suitable to predict WTI Crude Oil Spot Price, given that for both train and test set, it has the lowest RMSE and highest R-Squared.

5. Business Insights & Solutions

5.1 Business Insights

Using our best performing model, which is the Random Forest, we were able to identify the variables that are of highest importance and would most likely have a significant impact on WTI Crude Oil Spot Price (See Fig. B.13 & B.14 in Appendix B).

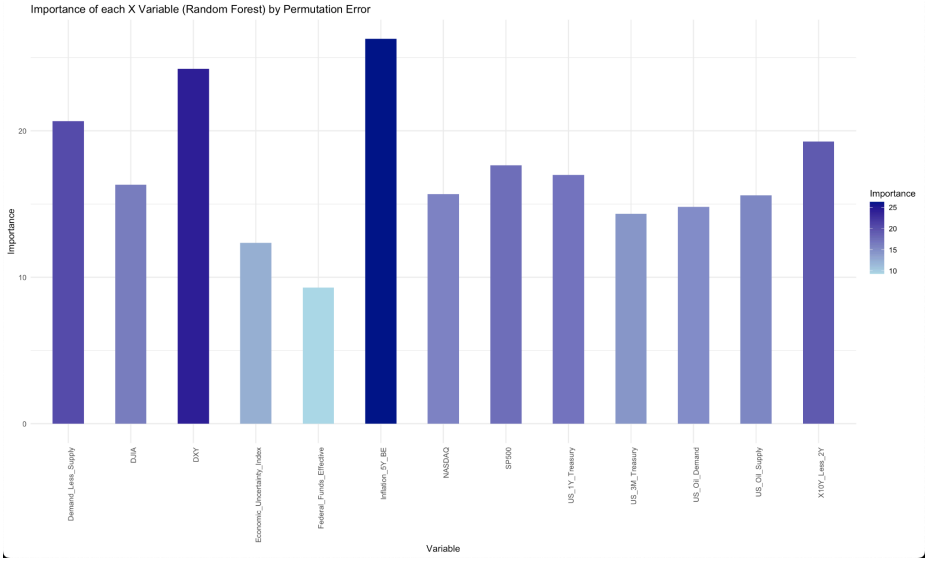


Figure 9: Variable Importance for Random Forest by Permutation Error

Armed with this newfound knowledge, Aramco can make more well-informed business decisions by monitoring the fluctuations in these variables. Hence, this information serves as crucial business insights for Aramco as it empowers them to be more prepared in anticipation of volatile oil prices.

Important Variables	Importance to Aramco
5 Year Breakeven Inflation rate	Anticipated inflation rates can impact oil prices. Higher inflation expectations may lead investors to buy commodities like oil as a hedge against currency devaluation, increasing oil demand and prices.
SP500, NASDAQ, DJIA	These stock markets reflect overall economic conditions and investor sentiment. A bullish stock market suggests economic growth and

	increased business activity. Hence higher stock market indices can potentially lead to higher oil demand and prices.
DXY (US Dollar Index)	The value of the US dollar plays a significant role in oil prices. Higher DXY signifies a stronger dollar which can make oil more expensive for international buyers, potentially reducing global demand and oil prices. Vice versa, a weaker dollar can have the opposite effect.
Demand Less Supply	When demand for crude oil surpasses the available supply, a shortage occurs. In this case, oil prices tend to rise as suppliers may increase prices to capitalise on strong demand.
US 10-2 Year Treasury Yield Spread	The spread between long-term and short-term Treasury rates provides insights into the state of the economy. When the yield curve inverts, it often signals economic downturns and potential recessions. An economic downturn can lead to reduced oil demand, and putting downward pressure on oil prices. Higher yield spread may raise economic expectations, subsequently leading to increased oil demand and prices.

Table 3: Important variables and significance to Aramco

5.2 Solutions for Aramco's Future Success

5.2.1 Dynamic Inventory Approach

This is a forward-thinking solution to optimise Aramco's oil inventory management in the volatile oil industry. We aim to help Aramco achieve a balance between retaining and selling inventory, responding to the predicted price surges and decline.

Data Integration

Aramco has to ensure that they have a real-time or near-real-time data feed to keep the oil price data up to date.

Decision Framework

Price Threshold refers to specific price levels that trigger oil inventory management actions. The exact threshold values will depend on Aramco's objectives and risk tolerance, as well as other factors such as holding costs and market conditions.

Model Prediction	Actions Taken
Exceed Price Threshold	Retain Oil to sell in the future
Falls below Price Threshold	Sell more Oil

Table 4: Showing actions for Aramco for exceeding/falling below price threshold

If our model anticipates a price surge, it offers a lucrative opportunity for Aramco to retain its oil inventory, subsequently capitalising on selling at a higher price point, thereby enhancing revenue. Conversely, when our model foresees a decline in oil prices, Aramco can strategically increase its oil sales to reduce inventory levels. This can help avoid potential losses from holding the inventory during a period of declining prices.

In addition, overstocking during forecasted price downturns, can result in unwelcoming holding costs and potential losses when disposing oil at reduced rates. Likewise, understocking during projected price upswings can result in missed revenue potential. Therefore, achieving and sustaining this equilibrium is of paramount importance.

Short-Term Inelasticity

In the short term, demand for crude oil is often considered **price inelastic**. This means that changes in price of the crude oil have a relatively small impact on the quantity demanded. This is because many consumers and industries have limited options for quickly switching to alternative energy sources, so they will continue to use oil despite the price increases.

This supports our framework, because if we retain more oil inventory to sell during higher prices, Aramco's consumers will still continue to purchase from them.

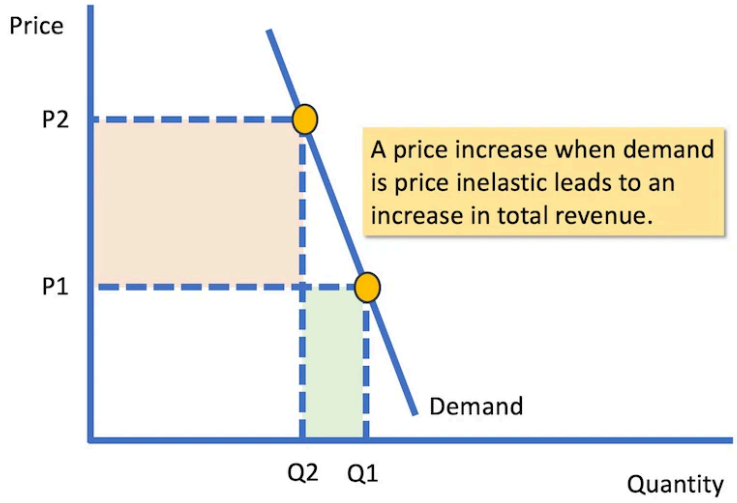


Figure 10: Inelastic demand curve

From the figure above, we can infer that if demand is inelastic, an increase in price leads to a larger increase in total revenue. This is because the percentage decrease in quantity demanded is smaller than the percentage increase in price, resulting in a net positive effect on revenue.

Investment in Infrastructure

Our models can also advise Aramco on when to increase their investments in infrastructure. Predictions of a prolonged period of high prices can signal Aramco to expand production, refining capabilities, and inventory infrastructure, ensuring they are in a position to handle surging demand. This would allow Aramco to maximise its revenue potential and ensure they remain at the forefront of the oil industry.

5.2.2 Predictive Risk-Hedging Investment in Financial Instruments

This strategy demonstrates how Aramco can harness the power and guidance of our model to effectively and robustly take advantage of various financial instruments to revolutionise its risk-hedging investment plans and, in turn, secure its financial well-being.

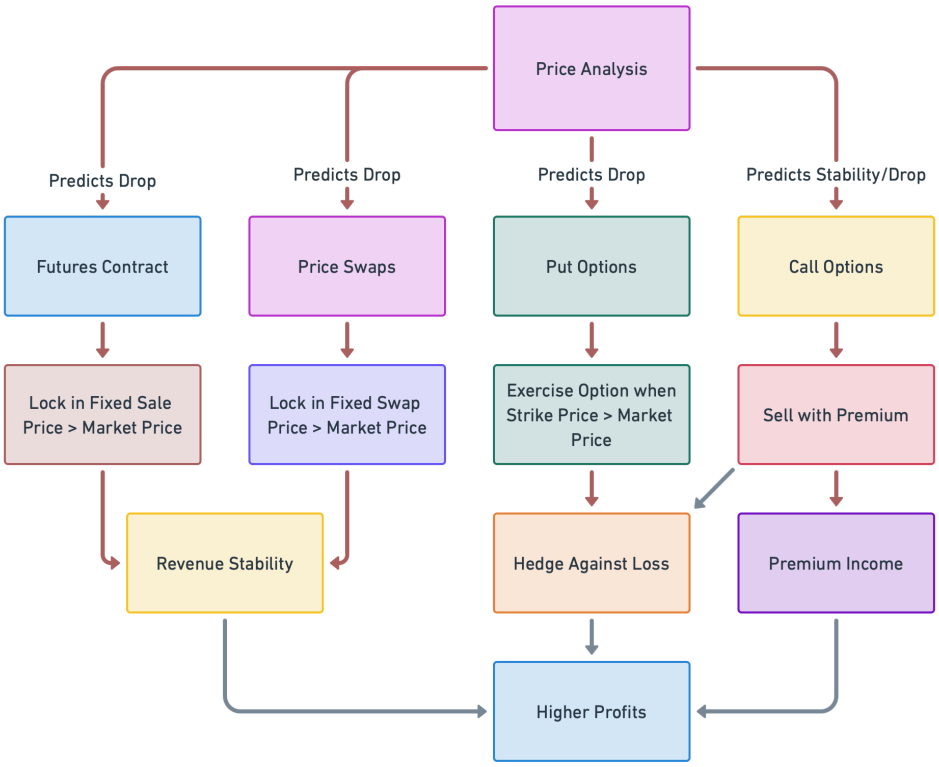


Figure 11: Risk-Hedging Investment Strategy

5.2.2.1 Utilising Lucrative Futures Contracts

Aramco's adoption of predictive risk-hedging investment begins with leveraging futures contracts.
A futures contract is a standardised financial agreement that allows Aramco to buy or sell oil at a predetermined price in future. These contracts are typically traded on organised commodity futures exchanges. When our model predicts a decline in oil prices due to increased global production or poor economic outlook, Aramco can initiate a strategic response by proactively locking in strategic futures contracts that enable the sale of oil at higher predetermined prices, irrespective of market fluctuations, ensuring that their revenue stream remains stable and profitable, even in turbulent times. Unknown to the other party of the contract, Aramco will actually have the upper hand and benefit more from the contract than them since market prices are going to drop lower than the contract price as accurately predicted by our reliable model. By securing prices in advance, Aramco can mitigate the risk of revenue erosion caused by falling market rates. This proactive approach not only protects revenues but also provides financial stability and predictability.

5.2.2.2 Securing Profitable Price Swaps

Swaps involve the exchange of fixed and floating oil prices with another party, ensuring a degree of price stability over a defined timeframe. When our model predicts dropping oil prices, Aramco can consider

entering into an oil price swap agreement to lock in a higher predetermined price for a set volume of oil it plans to produce and sell during that period. If the actual market price exceeds the predetermined price, Aramco will make a payment to the counterparty, but if it falls below that, Aramco will receive a payment from the counterparty to compensate for the difference, effectively acting as a hedge against potential revenue loss. Again Aramco will be at the higher ground as compared to the counterparty since, as accurately predicted by our reliable model, it is much more likely that the market price will decrease below the predetermined price thus Aramco will still receive its expected revenue and this price swap will be more profitable to Aramco than to the counterparty. Hence, this strategy provides price predictability and revenue stability for a defined timeframe, mitigating the risks associated with unfavourable price movements.

5.2.2.3 Extensive Capitalisation of Options Trading

Incorporating options contracts is the next layer of fine-tuning risk management. Put options will grant Aramco the right to sell oil at a predetermined price, also known as the strike price. If our model forecasts a decrease in oil prices due to a global economic downturn or oversupply, Aramco can respond by purchasing put options at a strike price that guarantees a profitable sale, even in a declining market. This means that if the market price of oil drops below the strike price, Aramco can exercise the put option, selling oil at the higher strike price, which effectively hedges against the potential loss in revenue due to falling prices. To take this strategy a step further, Aramco can also consider selling call options, which is a contract that gives Aramco's customers the right to buy a specified amount of oil at the strike price within a specific time frame. In exchange, Aramco will receive a premium payment from its customers as a compensation for agreeing to potentially sell oil at a strike price which may be lower than the market price.

Backed with newfound valuable insights from our excellent statistical model, Aramco can proactively tailor their options trading strategy. Unbeknownst to its customers and competitors, the ability to anticipate future oil prices empowers Aramco to make informed decisions about when and which options to buy and sell. For example, when our model signals a potential price decline, Aramco can acquire put options at optimal spot prices with advantageous strike prices to lock in enormous profits even in a bearish market where its competitors will be struggling. Additionally, when the forecast is stable or dropping, Aramco can consider selling attractive call options to lure customers, who are fumbling in the dark, with strike prices similar to current market prices. Little do they know, they will never get a good chance to exercise their call options since market prices are not increasing, so essentially Aramco will be able to generate premium income at virtually no cost.

By leveraging on our predictive model's forecasts, Aramco can strategically position itself in the options market to capitalise on price movements by thinking multiple steps ahead of its economic stakeholders and thus gaining an edge over them. This dynamic approach ensures that Aramco maximises its profits from options trading, making the most of market opportunities and minimising the impact of price volatility on their revenue. In essence, the integration of options trading as a risk management strategy, guided by our predictive model, allows Aramco to not only protect its bottom line but also potentially enhance its profitability in the complex and ever-changing world of the oil market.

5.2.3 Proactive Diversification

Proactive diversification capitalises on opportunities presented by stable oil prices and volatile market conditions to explore new markets, expand existing ones, and invest in renewable energy.

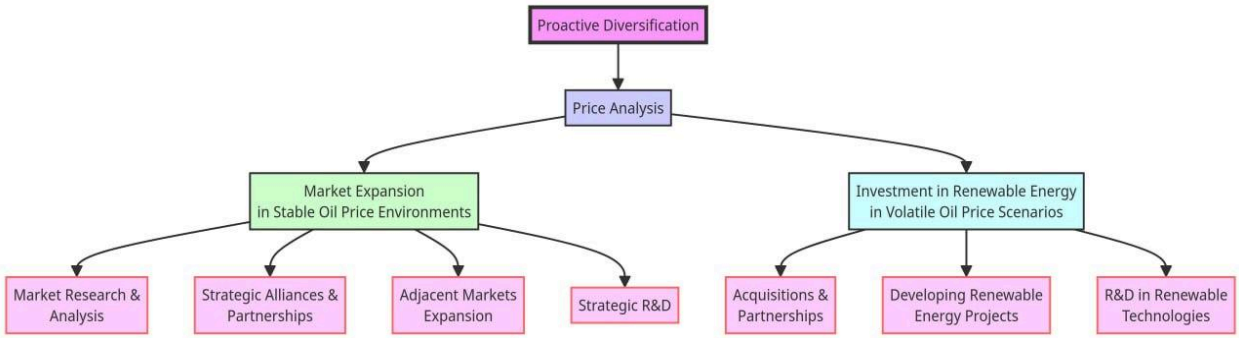


Figure 12: Proactive Diversification Strategy

5.2.3.1 Market Expansion in Stable Oil Price Environments

During periods of stable oil prices predicted by our models, Aramco can seize the opportunity to diversify its revenue streams by entering untapped markets or fortifying its presence in existing ones. This forward-looking approach reduces the company's reliance on oil as its sole revenue source.

In times of price stability, Aramco can conduct comprehensive market research and analysis to identify untapped markets with growth potential. This includes assessing demand, competition, regulatory conditions, and market dynamics. Additionally, collaborating with local or global partners can provide Aramco with market insights, established networks, and shared resources. Strategic alliances and partnerships facilitate market entry and reduce risks. Moreover, Aramco can explore and expand into adjacent markets related to its core business, such as petrochemicals or refining. Diversifying within existing markets leverages expertise and assets while reducing risk. This should be combined with strategic research and technological development like improving efficiency and safety of oil refineries and oil wells, as such investments can lead to product innovation, market differentiation and most importantly higher output productivity.

Entering new markets and strengthening existing ones diversifies Aramco's revenue sources. This effectively mitigates the risks associated with fluctuations in the global oil market and ensures a more financial stable performance. Furthermore, expanding into new markets allows Aramco to establish a broader footprint in the global energy landscape thereby strategically reinforcing the company's competitive position and creating long term growth opportunities.

5.2.3.2 Investment in Renewable Energy in Volatile Oil Price Scenarios

In times of volatile oil prices, Aramco can take a strategic leap by investing in alternative energies and entering the renewable energy market sector. This proactive approach positions Aramco as a forward-thinking leader in sustainable and green energy solutions, reducing its dependence on oil and securing a future-proofed portfolio.

Aramco can consider strategically acquiring or partnering with established renewable energy companies to fast-track its entry into the sector. Acquisitions provide access to existing projects and expertise. Apart from that, developing renewable energy projects, such as solar or wind farms, allows Aramco to build a presence in the sector. These projects offer alternative long-term revenue streams, while supporting

sustainability goals at the same time. Besides this, investing in research and development of renewable technologies and energy storage solutions can lead to innovation and competitive advantages in the renewable energy sector.

Investing in renewable energy diversifies Aramco's operations, reducing its reliance on oil and the inherent vulnerabilities associated with oil price fluctuations. The renewable energy sector offers substantial growth potential, so in times of volatile oil prices, Aramco can tap into this burgeoning market, generating additional revenue streams and ensuring long-term profitability.

5.3 Analysis of Effectiveness

5.3.1 Addressing the Opportunity Statement

The incorporation of data-driven decision-making is at the core of each proposed solution. Aramco's adoption of these solutions equips the company with the tools to transform big data into actionable insights. This empowers Aramco to make well-informed business decisions based on the analysis of various variables and market dynamics. Furthermore, the solutions enable Aramco to anticipate and react swiftly to market fluctuations as the use of our predictive model and risk management strategies ensures more reliable predictions of oil price trends and early forecasting of business patterns. By promptly managing inventory and securing future prices, Aramco can optimise its supply chain and pricing decisions, so these solutions also assist in achieving more effective implementation of tactical pricing strategies. Most importantly, the strategies presented are not only risk-mitigating but also profit-enhancing so Aramco can maximise its expected profits by locking in favourable prices and expanding revenue sources based on the insights provided by our predictive model, thereby strengthening Aramco’s financial position. By minimising Aramco financial risks, diversifying operation as well as responding proactively to market conditions, Aramco will be in a good position to maintain a strong financial footprint in the competitive global oil market.

5.3.2 Answering the Key Business Questions

5.3.2.1 Demand-Supply Balance

The dynamic inventory strategy optimises the balance between demand and supply. Aramco can efficiently manage its oil inventory based on predictions of price surges and declines. When the model anticipates price surges, Aramco retains its inventory for selling at higher prices, maximising revenue. Conversely, during predicted price declines, Aramco strategically increases sales to avoid potential losses from holding inventory during a period of falling prices. This approach ensures that Aramco strikes the right balance, avoiding costly overstocking or understocking, ultimately maximising profitability.

5.3.2.2 Market Dynamics

The use of predictive risk-hedging investment strategy incorporates market dynamics and leading indicators of oil price fluctuations into Aramco's decision-making. By leveraging financial instruments such as futures contracts, price swaps, and options trading, Aramco can respond quickly to market changes. For example, when the model predicts a decline in oil prices due to factors such as increased global production or poor economic outlook, Aramco can proactively lock in futures contracts, securing the sale of oil at higher predetermined prices. This approach allows Aramco to insulate itself from market fluctuations and ensure a stable and profitable revenue stream.

5.3.2.3 Risk Management

The predictive risk-hedging investment strategy focuses on risk management, allowing Aramco to hedge against volatile oil prices using financial instruments. For example, in anticipation of a price decline, Aramco can secure profitable sales with put options. With the ability to make informed decisions about options trading, Aramco will effectively mitigate risk and ensure long term revenue stability. In the context of risk management, the proactive diversification strategy offers an innovative approach whereby during uncertain periods of volatility, Aramco can reduce its risk by investing in more renewable energy as well as diversifying its operations. This forward-thinking strategy minimises vulnerability to oil price fluctuations and ensures long-term risk resilience and profitability.

5.4 Model Limitations

While our models are able to predict oil prices accurately, their accuracy decreases when measured over longer time periods. This decrease is due to external factors such as geopolitical events like OPEC decisions, global pandemics such as COVID-19 and growing environmental concerns. These factors are unquantifiable and therefore unpredictable, and they may impact our models in ways our current models can't predict. As a result, our models are limited to short to medium time horizons and are less effective for longer time horizons spanning years.

Aramco is one of the world's largest oil producers with stakes in many unique regions. While Aramco operates on a global scale, the intricacies of each regional market can differ significantly. Our model, which predicts global oil prices, may not be suitable for all of Aramco's situations. Aramco operates in many regions of the world with dynamically changing and volatile markets, facing challenges such as supply chain issues, trade regulations, and varying competition. A global oil price prediction model may therefore be unpredictable and limited.

5.5 Conclusion

Saudi Aramco stands at the intersection of possibilities and opportunities. Our ambition for Aramco is to lead, not follow—to navigate the industry into new realms of achievement. With our Dynamic inventory management, predictive risk-hedging investment and proactive diversification secure financial stability, Aramco will secure robust growth and a leading market position. Envision Aramco not just as a participant but as a trendsetter, a titan of innovation whose very name is synonymous with enduring success. With our proposal, Aramco will transform its operations and solidify its status at the vanguard of the industry, equipped to navigate the challenges of tomorrow with confidence and strategic agility.

References

Barnes, A., & Cingari, P. (2022, December 15). Crude and dollar correlation: Why oil and the USD move in different directions?. Online Trading with Smart Investment App. https://capital.com/crude-dollar-correlation

Barnett, A., French, J., & Said, S. (2017, December 12). *Aramco: How The World's Most Important Energy Company Works*. The Wall Street Journal. https://www.wsj.com/graphics/what-is-aramco/

Chen, J. (n.d.). Treasury yield: What it is and factors that affect it. Investopedia. https://www.investopedia.com/terms/t/treasury-yield.asp

DELVENTHAL, S. (2022, March 27). *What Is Saudi Aramco? Its History, IPO, and Financials*. Investopedia. https://www.investopedia.com/what-is-saudi-aramco-4682590#:~

Great expectations: How Treasury bonds and interest rate changes impact currencies. OFX (SG). (2022, February 7). https://www.ofx.com/en-sg/forex-news/articles/how-interest-rates-impact-currencies/

Hayes, A. (2023, September 29) *Autoregressive Integrated Moving Average (ARIMA) Prediction Model.* Investopedia. https://www.investopedia.com/terms/a/autoregressive-integrated-moving-average-arima.asp

Hyndman, R. and Athanosopoulos, G. (2018) *Arima Modelling in R.* Forecasting: Principles and Practice. https://otexts.com/fpp2/arima-r.html

IBM. (2022). *What is Random Forest?* IBM. https://www.ibm.com/topics/random-forest#:~:text=Random%20forest%20is%20a%20commonly,both%20classification%20and%20regression%20problems.

Lioudis, N. (2023, August). What is the relationship between oil prices and inflation?. Investopedia. https://www.investopedia.com/ask/answers/06/oilpricesinflation.asp

Martin, F. M. (2021, December 9). How might increases in the Fed funds rate impact other interest rates?. Saint Louis Fed Eagle. https://www.stlouisfed.org/on-the-economy/2017/october/increases-fed-funds-rate-impact-other-interest-rates

Minitab. (2023) *Interpret the Key Results for ARIMA.* SupportMinitab. https://support.minitab.com/en-us/minitab/21/help-and-how-to/statistical-modeling/time-series/how-to/arima/interpret-the-results/key-results/?SID=117600&SID=117600

Murphy, D., & Turak, N. (2023, August 7). *Oil giant Saudi Aramco posts 38% drop in second-quarter profit as lower prices bite*. CNBC. https://www.cnbc.com/2023/08/07/oil-giant-saudi-aramco-posts-sharp-drop-in-second-quarter-profit-as-lower-prices-bite.html

Sharma, R. (n.d.). NASDAQ composite vs. S&P 500 vs. Dow: What's the difference? Investopedia. https://www.investopedia.com/nasdaq-vs-sp500-vs-dow-5213252#toc-the-bottom-line

Utkarsh, K. (2020, July 08) *Time Series Analysis using ARIMA Model in R-programming.* GeeksforGeeks. https://www.geeksforgeeks.org/time-series-analysis-using-arima-model-in-r-programming/

US, EIA. (2023). *U.S. Energy Information Administration - EIA - independent statistics and analysis*. Oil prices and outlook - U.S. Energy Information Administration (EIA). https://www.eia.gov/energyexplained/oil-and-petroleum-products/prices-and-outlook.php#:~:text=Many%20countries%20also%20rely%20primarily,of%20total%20world%20energy%20consumption.

Varshney, P. (2020, October 31) *Measure Performance for a Time Model.* TowardsDataScience. https://towardsdatascience.com/measures-performance-for-a-time-series-model-ets-or-arima-18b0a3e91e83

Appendix A

1. Summary Statistics
Dataset Link: https://www.kaggle.com/datasets/filteredtaph2o/wti-crude-oil-price-prediction

Variable	Data Type	Description
Date	Character	Date
WTI_Spot	Numeric	Closing Price of WTI Crude Oil Spot
GSCI	Numeric	Closing Price of GSCI Commodities Index
SP500	Numeric	Closing Price of S&P500
DJIA	Numeric	Closing Price of DJIA
NASDAQ	Numeric	Closing Price of NASDAQ
DXY	Numeric	Closing price of DXY. This is an index that measures the relative strength of the USD versus a basket of other relevant currencies
Federal_Funds_Effective	Numeric	Effective Federal Funds Rate
Economic_Uncertainty_Index	Numeric	Metric for Economic Policy Uncertainty
Inflation_5Y_BE	Numeric	5-Year Break even inflation rate
US_3M_Treasury	Numeric	Closing price of US 3-month treasury
US_1Y_Treasury	Numeric	Closing price of US 1-year treasury
10Y_Less_2Y	Numeric	US 10-year treasury rate minus US 2-year treasury rate
US_Oil_Demand	Numeric	Monthly US Oil Demand for the observation month
US_Oil_Supply	Numeric	Monthly US Oil Supply for the observation month
Demand_Less_Supply	Numeric	Monthly US Oil Demand minus Monthly US Oil Supply for the observation month

Table A.1: Data Dictionary

2. Univariate Analysis

WTI_Spot	SP500	NASDAQ	DJIA	
Min. : 8.91	Min. :1562	Min. : 3234	Min. :14567	Min.
1st Qu.: 48.33	1st Qu.:2069	1st Qu.: 4894	1st Qu.:17723	1st
Median : 59.12	Median :2642	Median : 7024	Median :23996	Medi
Mean : 65.06	Mean :2789	Mean : 7784	Mean :23795	Mear
3rd Qu.: 81.24	3rd Qu.:3351	3rd Qu.:10558	3rd Qu.:28384	3rd
Max. :123.64	Max. :4797	Max. :16057	Max. :36800	Max.

Inflation_5Y_BE	US_3M_Treasury	US_1Y_Treasury	X10Y_Less_2Y
Min. :0.6745	Min. :-0.0500	Min. :0.040	Min. :-0.8400
1st Qu.:1.5300	1st Qu.: 0.0500	1st Qu.:0.130	1st Qu.: 0.3000
Median :1.7500	Median : 0.2700	Median :0.550	Median : 0.8900
Mean :1.8345	Mean : 0.8348	Mean :1.025	Mean : 0.9359
3rd Qu.:2.0400	3rd Qu.: 1.5400	3rd Qu.:1.730	3rd Qu.: 1.4200
Max. :3.4109	Max. : 4.5375	Max. :4.512	Max. : 2.6600

Fig. A.1: Summary statistics of Variables

DXY	Federal_Funds_Effective	Economic_Uncertainty_Index
Min. : 79.86	Min. :0.0400	Min. : 3.32
1st Qu.: 91.59	1st Qu.:0.0900	1st Qu.: 61.38
Median : 95.40	Median :0.3600	Median : 89.75
Mean : 94.15	Mean :0.8348	Mean :111.22
3rd Qu.: 97.85	3rd Qu.:1.5500	3rd Qu.:134.00
Max. :111.90	Max. :4.3300	Max. :529.29

2Y	US_Oil_Demand	US_Oil_Supply	Demand_Less_Supply
8400	Min. :14690	Min. : 7279	Min. : 2776
3000	1st Qu.:19260	1st Qu.: 9101	1st Qu.: 8292
8900	Median :19830	Median :10085	Median : 9845
9359	Mean :19699	Mean :10275	Mean : 9424
4200	3rd Qu.:20330	3rd Qu.:11634	3rd Qu.:10495
6600	Max. :21630	Max. :13000	Max. :11737

Fig. A.2: Summary statistics of Variables (cont.)

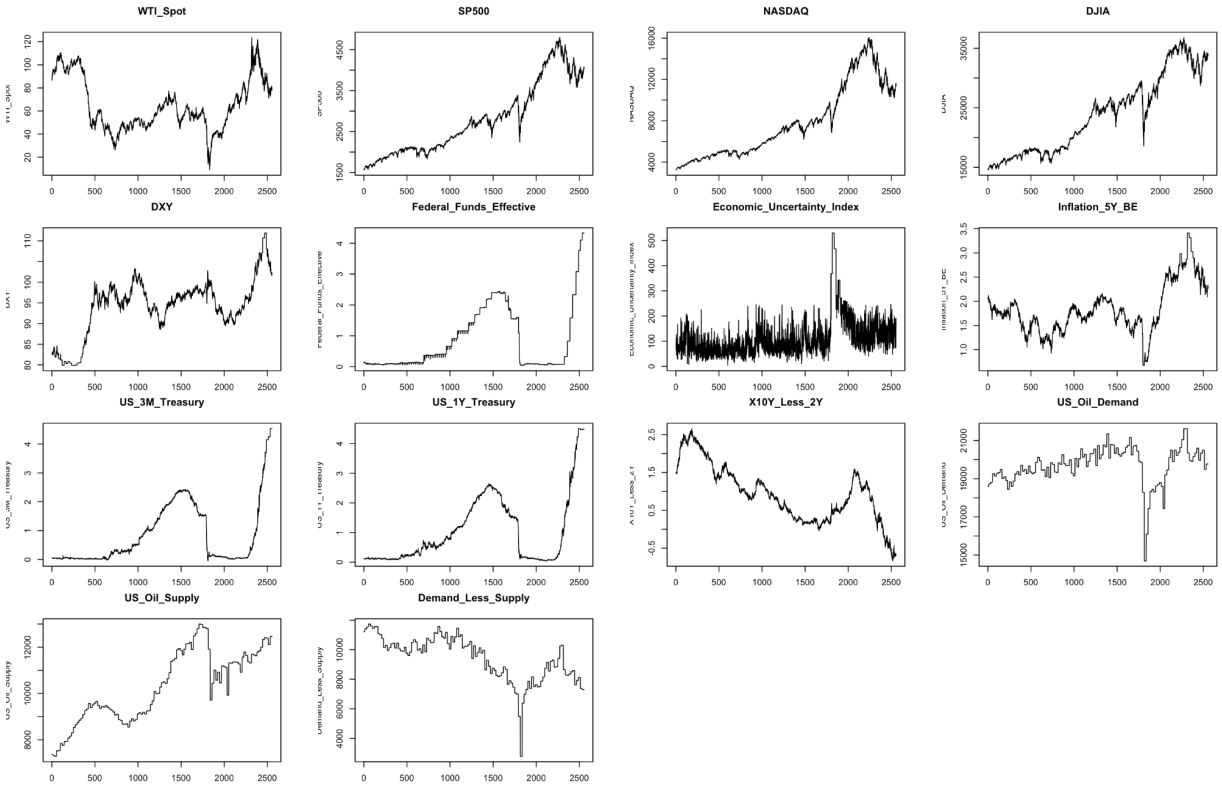


Fig. A.3: General Trends of Variables

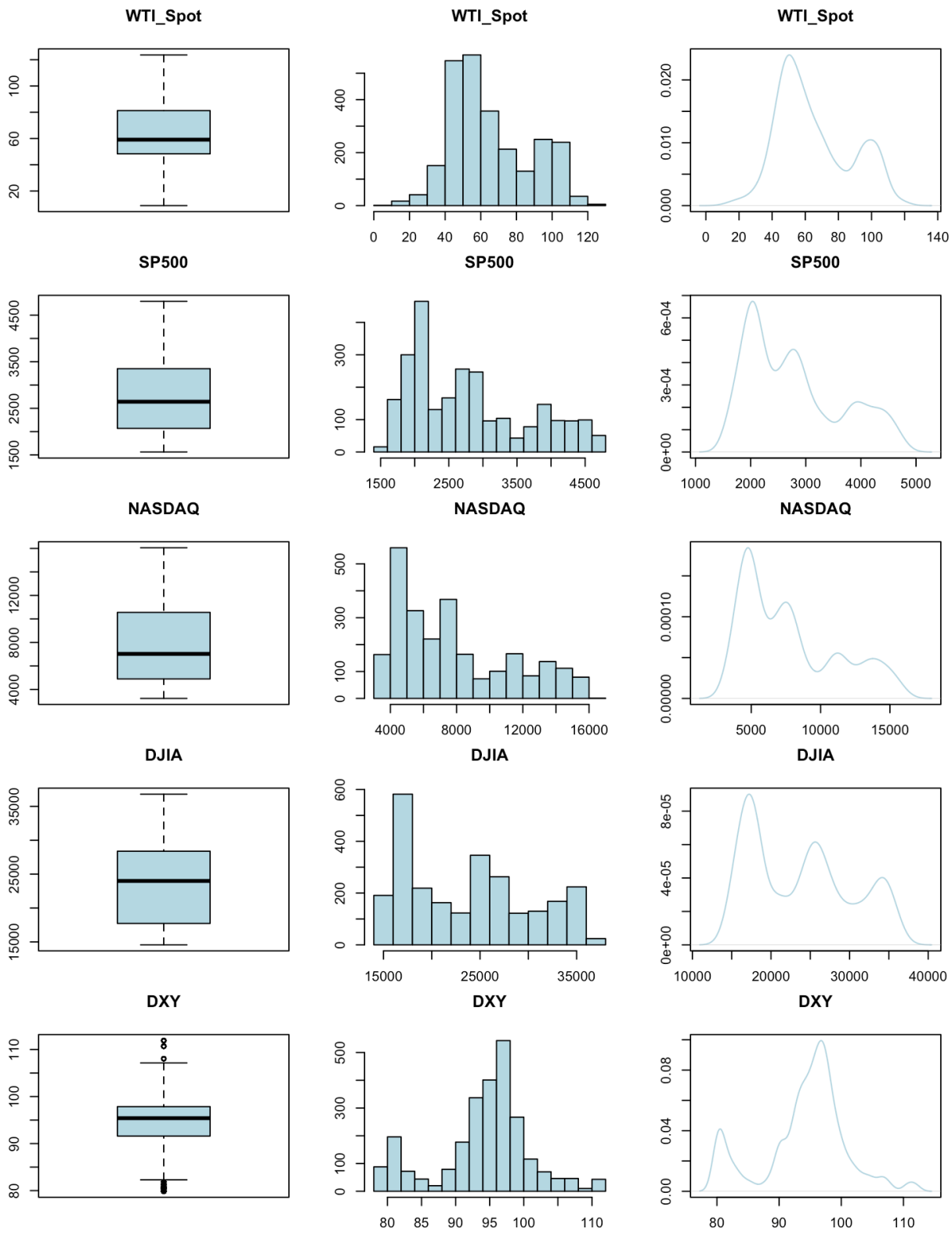


Fig. A.4: Box-plots, Histograms and KDE Plots of Variables

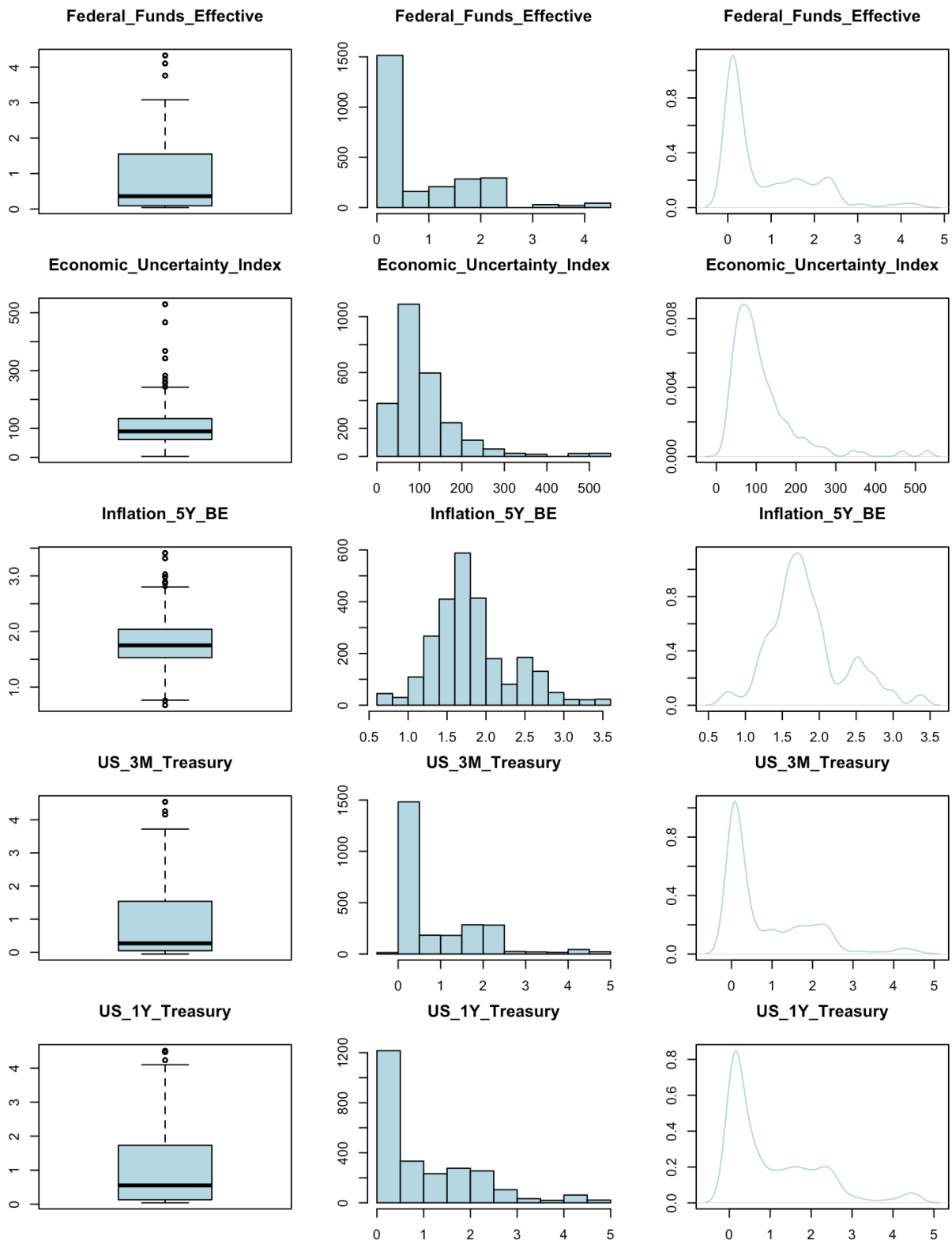


Fig. A.5: Box-plots, Histograms and KDE Plots of Variables

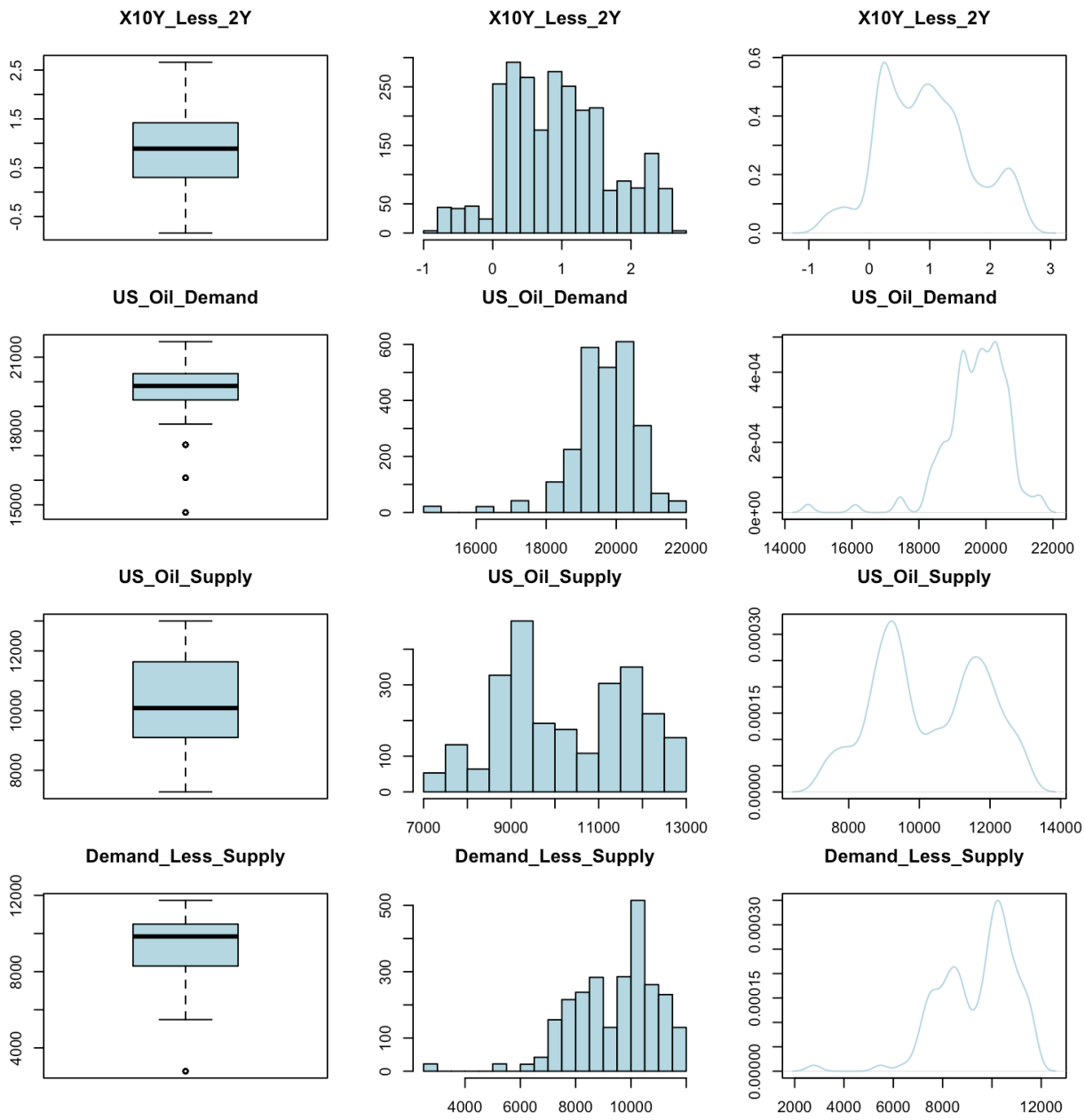


Fig. A.6: Box-plots, Histograms and KDE Plots of Variables

WTI_Spot	SP500	NASDAQ
0.503814143	0.659431454	0.753480207
DJIA	DXY	Federal_Funds_Effective
0.346697870	-0.429929919	1.321767959
Economic_Uncertainty_Index	Inflation_5Y_BE	US_3M_Treasury
2.517128846	0.705943731	1.399502479
US_1Y_Treasury	X10Y_Less_2Y	US_Oil_Demand
1.285317213	0.240828538	-1.652931471
US_Oil_Supply	Demand_Less_Supply	
0.001754281	-0.959753253	

Fig. A.7: Skewness of Continuous Variables

3. Bivariate Analysis

	WTI_Spot	SP500	NASDAQ	DJIA	DXY	Federal_Funds_Effective	Economic_Uncertainty_Index
WTI_Spot	1.000000000	0.1020736	0.05764585	0.07569115	-0.4009226	0.004884758	-0.24316658
SP500	0.102073631	1.0000000	0.98854638	0.98664069	0.4488399	0.232159526	0.30159556
NASDAQ	0.057645851	0.9885464	1.00000000	0.96641942	0.3844293	0.130964535	0.35376979
DJIA	0.075691151	0.9866407	0.96641942	1.00000000	0.4655665	0.349680610	0.28848543
DXY	-0.400922579	0.4488399	0.38442932	0.46556646	1.0000000	0.513773250	0.18831736
Federal_Funds_Effective	0.004884758	0.2321595	0.13096454	0.34968061	0.5137733	1.000000000	-0.07753074
Economic_Uncertainty_Index	-0.243166578	0.3015956	0.35376979	0.28848543	0.1883174	-0.077530736	1.00000000
Inflation_5Y_BE	0.644621350	0.6987324	0.65782186	0.67203271	0.1003640	0.133867420	-0.10581502
US_3M_Treasury	0.059805799	0.2855496	0.18234247	0.39533745	0.5410458	0.988328418	-0.04999987
US_1Y_Treasury	0.109998788	0.3086233	0.19750707	0.40422892	0.5973074	0.950118694	-0.07063278
X10Y_Less_2Y	0.277562653	-0.5757706	-0.50736628	-0.64803756	-0.7909489	-0.758093726	-0.24770759
US_Oil_Demand	0.178351895	0.2722754	0.18818474	0.31764972	0.2606171	0.453698179	-0.53136216
US_Oil_Supply	-0.122629280	0.7563192	0.71322872	0.81544434	0.6014948	0.591161373	0.27723034
Demand_Less_Supply	0.238655850	-0.5914878	-0.60178728	-0.62227196	-0.4421255	-0.307778927	-0.62175879

	Inflation_5Y_BE	US_3M_Treasury	US_1Y_Treasury	X10Y_Less_2Y	US_Oil_Demand	US_Oil_Supply	Demand_Less_Supply
WTI_Spot	0.64462135	0.05980580	0.10999879	0.2775627	0.1783519	-0.1226293	0.23865585
SP500	0.69873244	0.28554960	0.30862325	-0.5757706	0.2722754	0.7563192	-0.59148783
NASDAQ	0.65782186	0.18234247	0.19750707	-0.5073663	0.1881847	0.7132287	-0.60178728
DJIA	0.67203271	0.39533745	0.40422892	-0.6480376	0.3176497	0.8154443	-0.62227196
DXY	0.10036400	0.54104584	0.59730744	-0.7909489	0.2606171	0.6014948	-0.44212546
Federal_Funds_Effective	0.13386742	0.98832842	0.95011869	-0.7580937	0.4536982	0.5911614	-0.30777893
Economic_Uncertainty_Index	-0.10581502	-0.04999987	-0.07063278	-0.2477076	-0.5313622	0.2772303	-0.62175879
Inflation_5Y_BE	1.00000000	0.20678545	0.28989888	-0.1827461	0.4181667	0.3169949	-0.05283578
US_3M_Treasury	0.20678545	1.00000000	0.97879605	-0.7790229	0.4463558	0.6060104	-0.32753231
US_1Y_Treasury	0.28989888	0.97879605	1.00000000	-0.7894610	0.4755442	0.5791514	-0.28159629
X10Y_Less_2Y	-0.18274613	-0.77902294	-0.78946101	1.0000000	-0.3344977	-0.8073185	0.60323070
US_Oil_Demand	0.41816666	0.44635579	0.47554418	-0.3344977	1.0000000	0.3368497	0.30034765
US_Oil_Supply	0.31699487	0.60601045	0.57915138	-0.8073185	0.3368497	1.0000000	-0.79691451
Demand_Less_Supply	-0.05283578	-0.32753231	-0.28159629	0.6032307	0.3003476	-0.7969145	1.00000000

Fig. A.8: Correlation Matrix of Variables

Appendix B

Model 1. Linear Regression Model

Components and weights [edit]

S&P GSCI Components and Dollar Weights as of May 07, 2020[3]

Energy	61.71%	Industrial Metals	10.65%	Precious Metals	4.50%	Agriculture	15.88%	Livestock	7.25%
WTI Crude Oil	25.31%	LME Aluminium	3.69%	Gold	4.08%	Chicago Wheat	2.85%	Live Cattle	3.90%
Brent Crude Oil	18.41%	LME Copper	4.36%	Silver	0.42%	Kansas Wheat	1.25%	Feeder Cattle	1.30%
RBOB Gasoline	4.53%	LME Lead	0.68%			Corn	4.90%	Lean Hogs	2.05%
Heating Oil	4.27%	LME Nickel	0.80%			Soybeans	3.11%		
Gasoil	5.95%	LME Zinc	1.12%			Cotton	1.26%		
Natural Gas	3.24%					Sugar	1.52%		
						Coffee	0.65%		
						Cocoa	0.34%		

Table B.1: Components and weights of GSCI Index

```
Step:  AIC=7409.71
WTI_Spot ~ SP500 + NASDAQ + DJIA + DXY + Federal_Funds_Effective +
    Economic_Uncertainty_Index + Inflation_5Y_BE + US_3M_Treasury +
    US_1Y_Treasury + X10Y_Less_2Y + US_Oil_Demand + US_Oil_Supply

                           Df Sum of Sq    RSS    AIC
<none>                                  110548 7409.7
- US_1Y_Treasury            1       973 111521 7423.4
- US_3M_Treasury            1      1221 111769 7427.4
- Federal_Funds_Effective   1      1303 111851 7428.7
- US_Oil_Demand             1      2048 112596 7440.6
- Economic_Uncertainty_Index 1     4401 114949 7477.6
- X10Y_Less_2Y              1      8569 119117 7541.4
- US_Oil_Supply             1     19257 129805 7695.3
- NASDAQ                    1     26783 137332 7796.3
- DJIA                      1     32399 142947 7868.0
- Inflation_5Y_BE           1     47444 157992 8047.3
- SP500                     1     53500 164048 8114.6
- DXY                       1     75439 185987 8339.4
```

Fig. B.1: Optimisation of Linear Model via AIC minimisation

```
Call:
lm(formula = WTI_Spot ~ SP500 + NASDAQ + DJIA + DXY + Federal_Funds_Effective +
    Economic_Uncertainty_Index + Inflation_5Y_BE + US_3M_Treasury +
    US_1Y_Treasury + X10Y_Less_2Y + US_Oil_Demand + US_Oil_Supply,
    data = training_data[, -1])

Residuals:
    Min      1Q  Median      3Q     Max
-23.896  -4.948  -0.505   3.905  29.021

Coefficients:
                            Estimate Std. Error t value Pr(>|t|)
(Intercept)                1.643e+02  8.408e+00  19.542  < 2e-16 ***
SP500                      1.009e-01  3.439e-03  29.334  < 2e-16 ***
NASDAQ                    -1.334e-02  6.429e-04 -20.755  < 2e-16 ***
DJIA                      -7.794e-03  3.414e-04 -22.827  < 2e-16 ***
DXY                       -2.026e+00  5.817e-02 -34.833  < 2e-16 ***
Federal_Funds_Effective   -7.955e+00  1.738e+00  -4.578 5.03e-06 ***
Economic_Uncertainty_Index 3.072e-02  3.651e-03   8.413  < 2e-16 ***
Inflation_5Y_BE            3.051e+01  1.105e+00  27.624  < 2e-16 ***
US_3M_Treasury             1.129e+01  2.547e+00   4.432 9.91e-06 ***
US_1Y_Treasury             6.410e+00  1.620e+00   3.956 7.92e-05 ***
X10Y_Less_2Y               1.036e+01  8.822e-01  11.740  < 2e-16 ***
US_Oil_Demand             -1.829e-03  3.188e-04  -5.739 1.12e-08 ***
US_Oil_Supply              5.566e-03  3.163e-04  17.599  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.885 on 1778 degrees of freedom
Multiple R-squared:  0.878,     Adjusted R-squared:  0.8772
F-statistic:  1066 on 12 and 1778 DF,  p-value: < 2.2e-16
```

Fig. B.2: Overall Summary of Optimal Linear Model

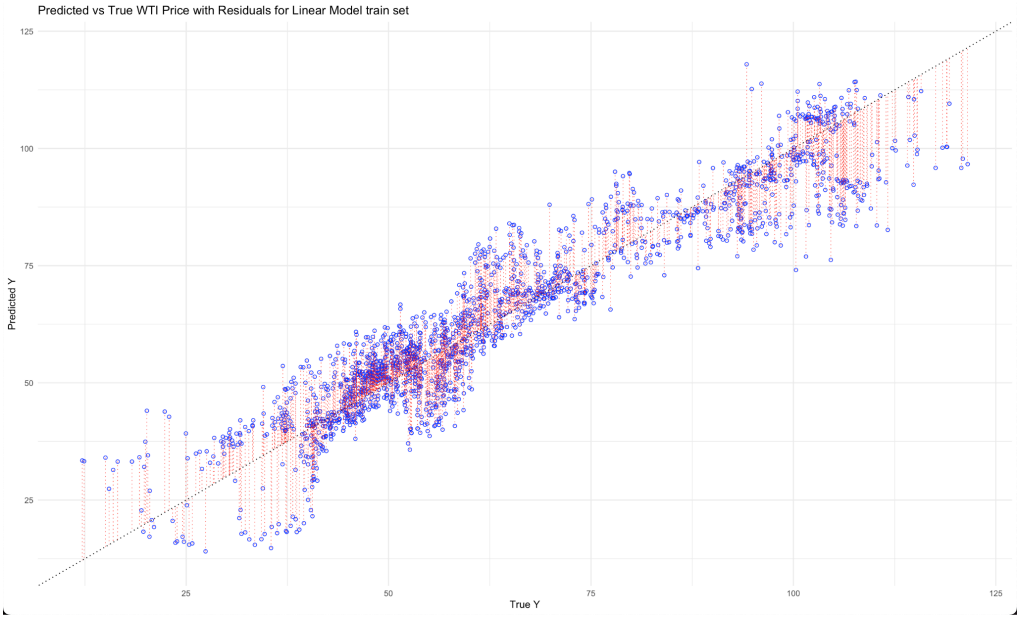


Fig. B.3: Predicted vs True WTI Price with Residuals for Linear Model Train Set

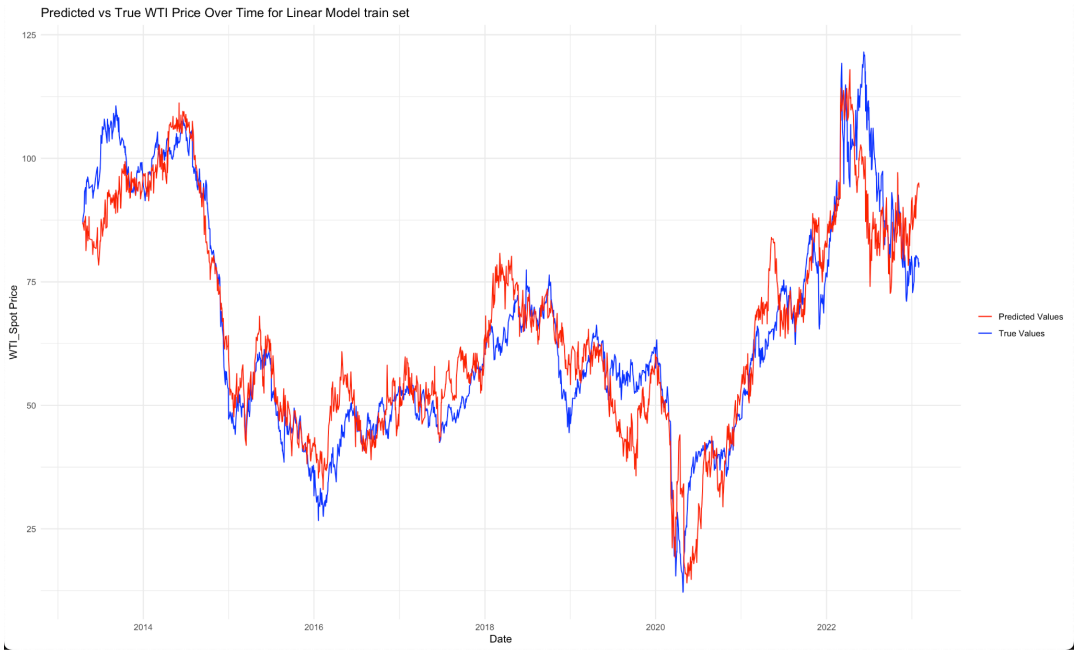


Fig. B.4: Predicted vs True WTI Price with over Time for Linear Model Train Set

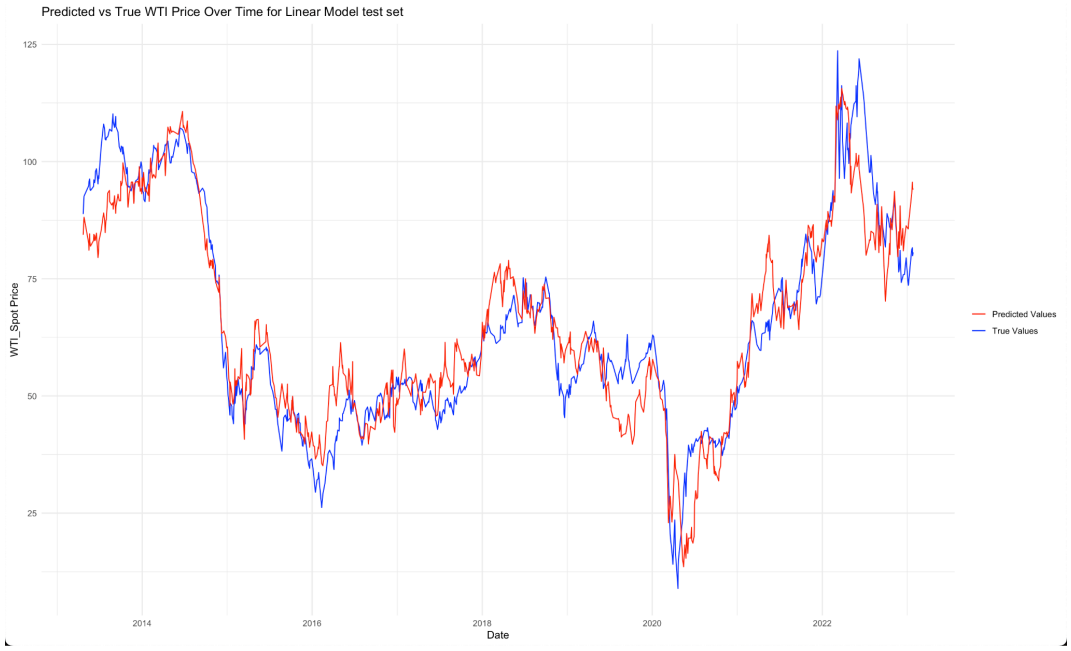


Fig. B.4.5: Predicted vs True WTI Price with over Time for Linear Model Test Set

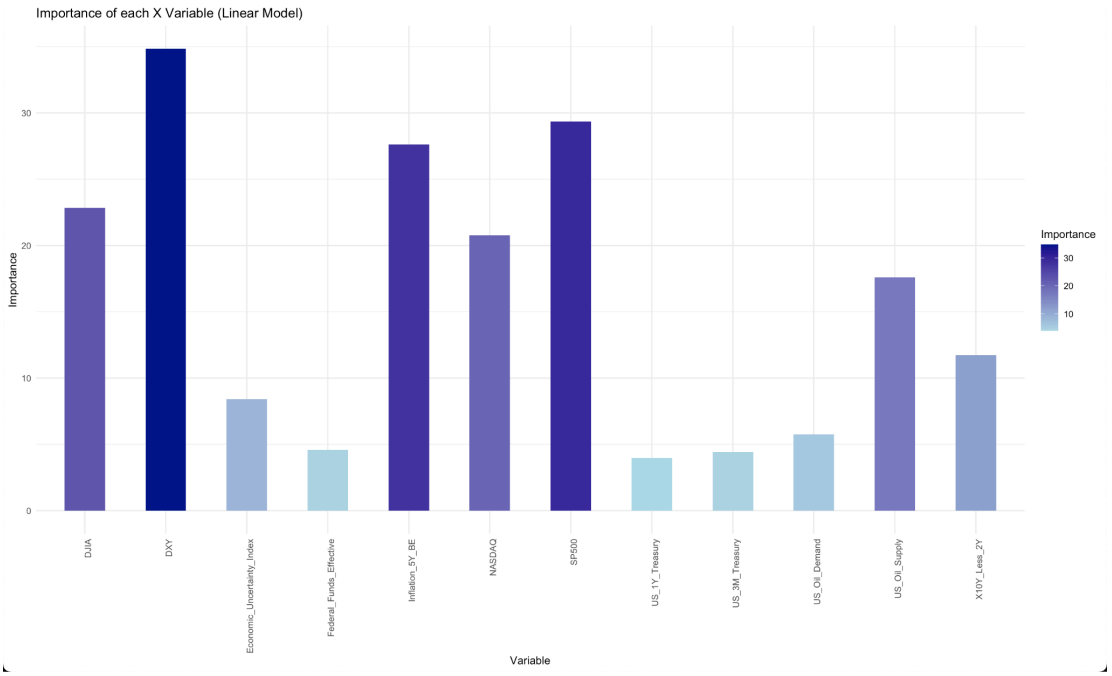


Fig. B.5: Importance of each X Variable (Linear Model)

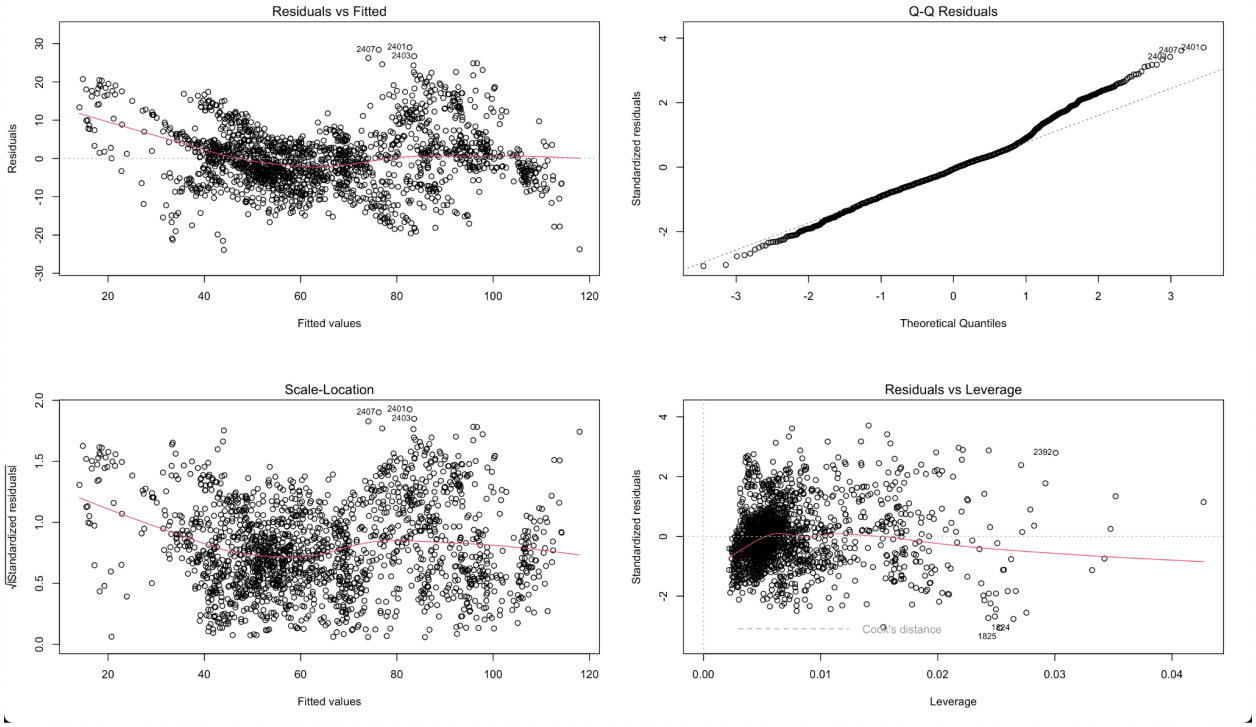


Fig. B.6: Linear Model Diagnostic Plot

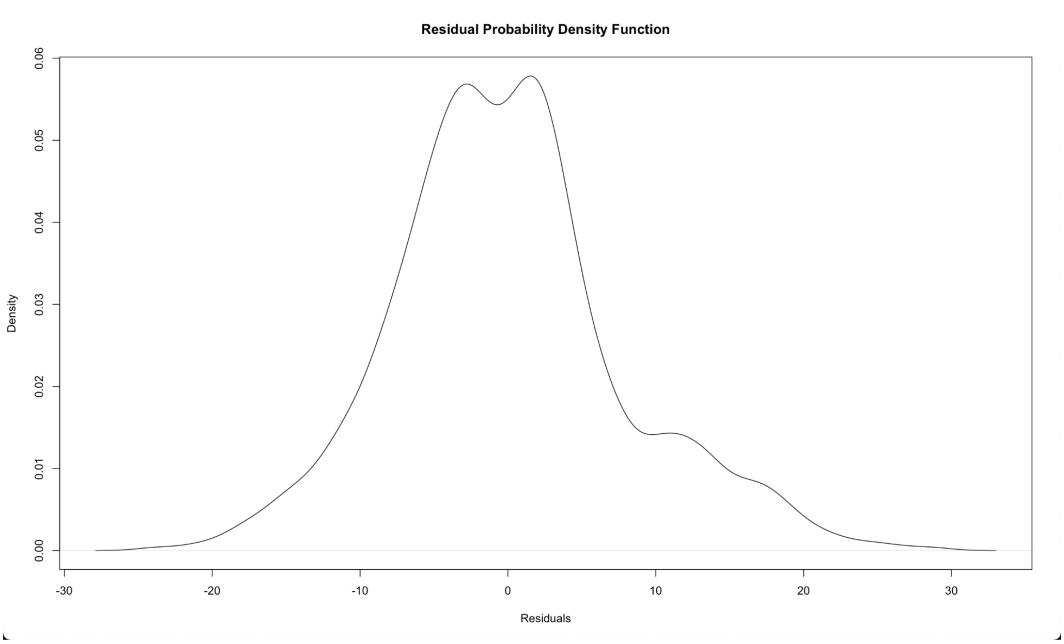


Fig. B.7: Residual Probability Density Function

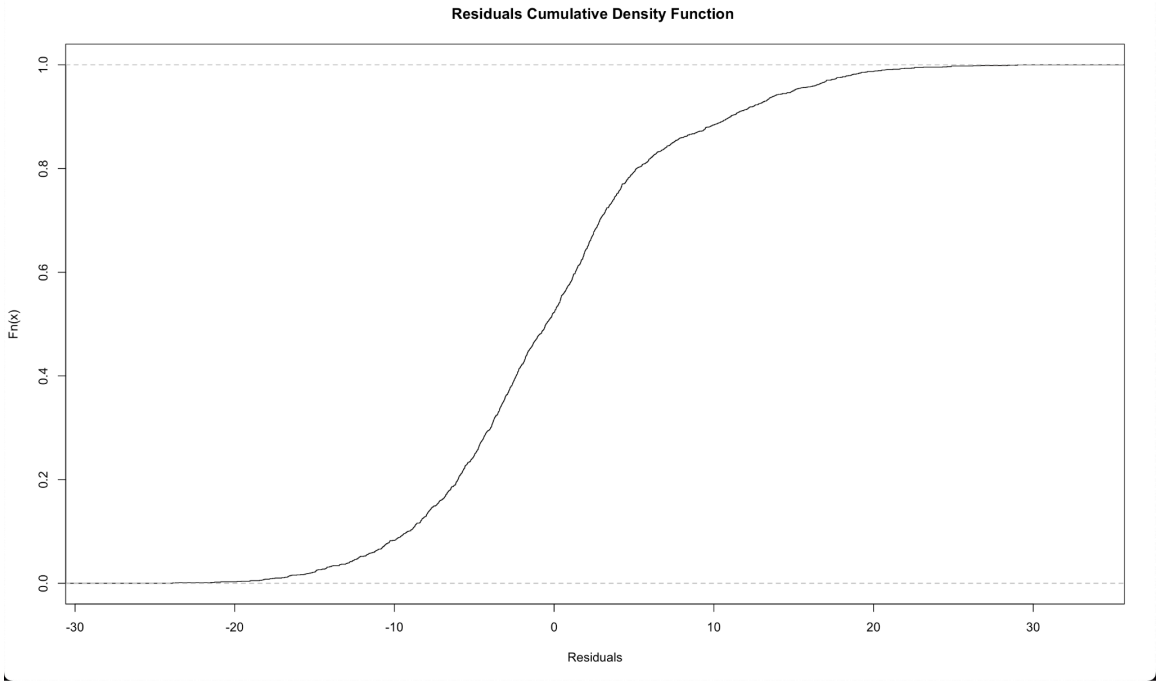


Fig. B.7.5: Residual Cumulative Density Plot

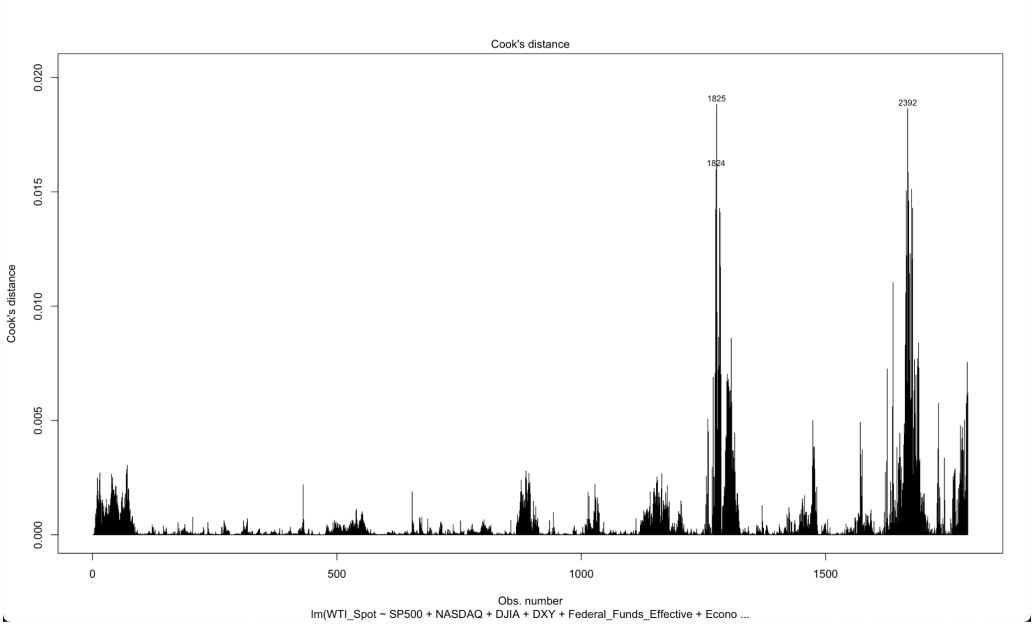


Fig. B.8: Cook's Distance (Linear Regression)

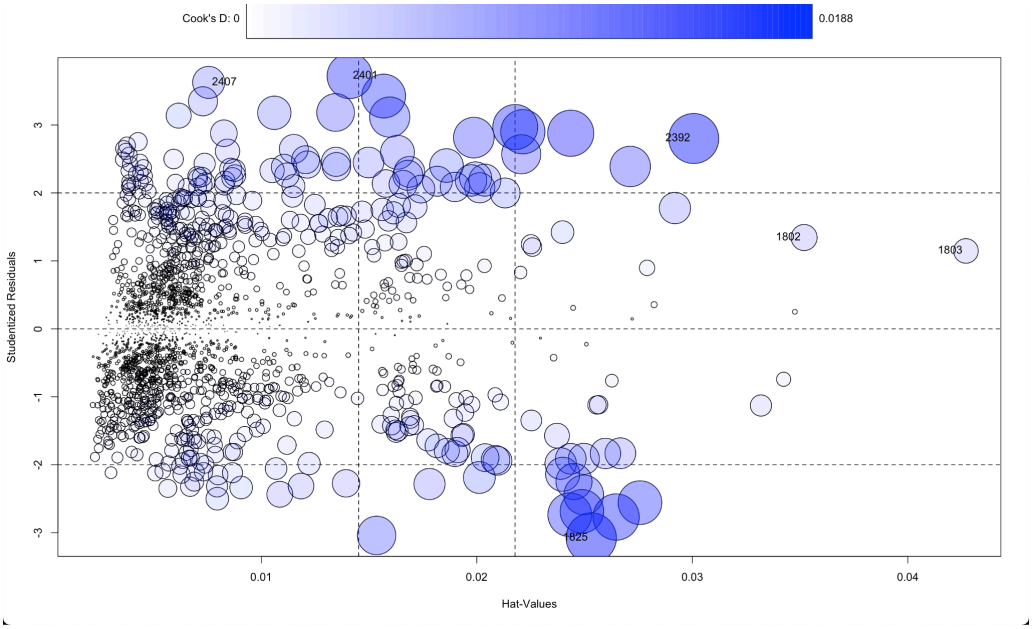


Fig. B.9: Linear Model Influence of Cook's Distance Plot

Model 2. CART Model

	Importance
NASDAQ	553330.59
DXY	548996.05
SP500	529576.93
DJIA	526484.10
X10Y_Less_2Y	435627.92
US_Oil_Supply	364002.72
Inflation_5Y_BE	309959.59
US_1Y_Treasury	178230.92
Federal_Funds_Effective	97234.62
US_3M_Treasury	66092.66
Demand_Less_Supply	57982.41
US_Oil_Demand	54473.18
Economic_Uncertainty_Index	34196.72

Table B.2: Variable Importance of CART

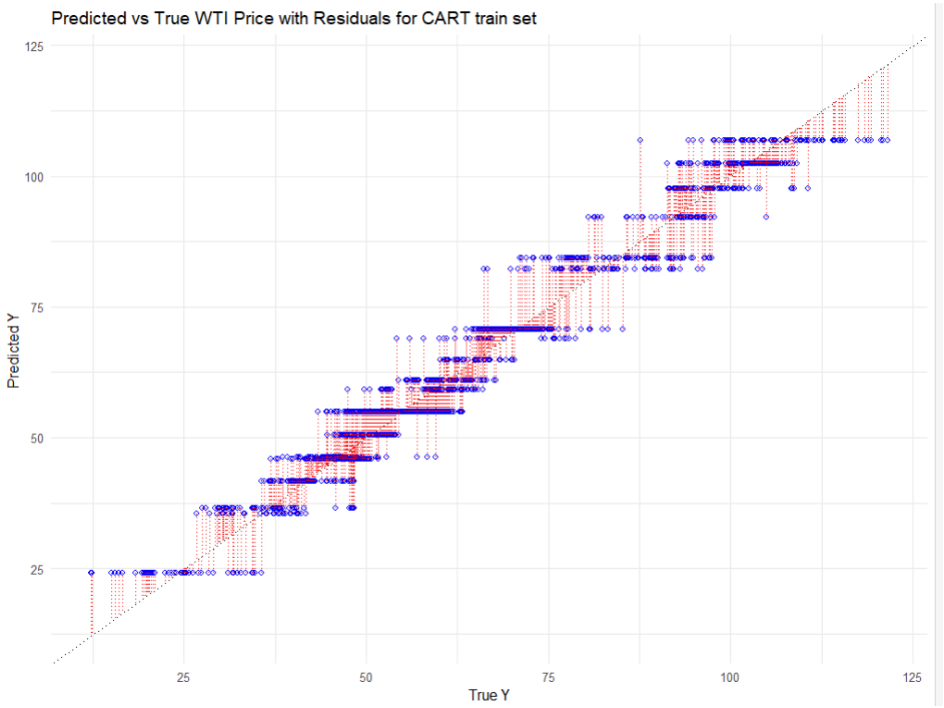


Fig. B.10: Predicted WTI_Spot vs True WTI_Spot with Residuals for CART Train set

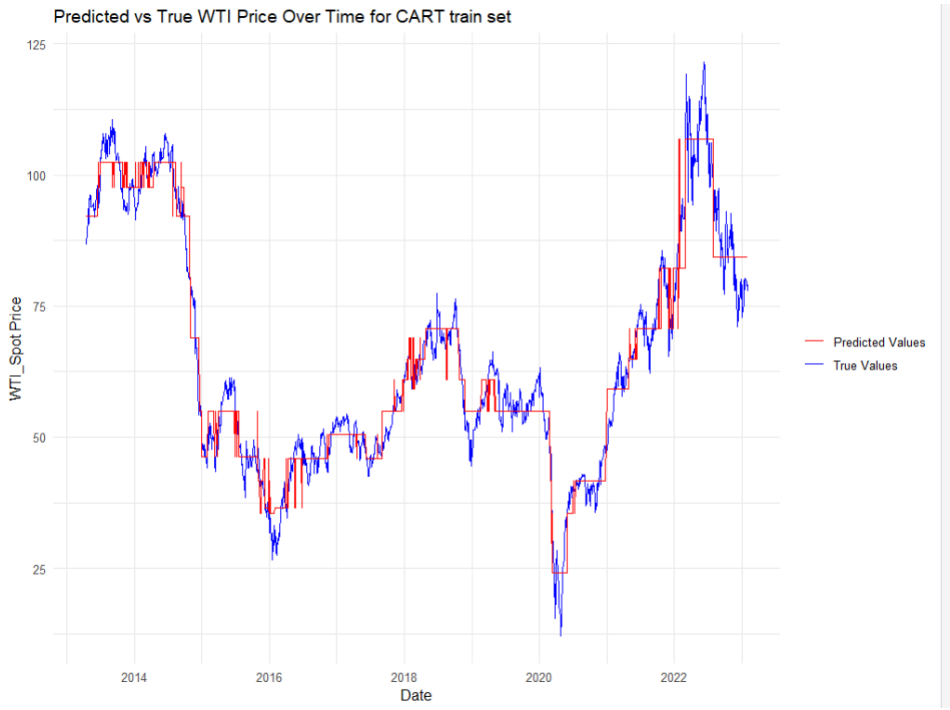


Fig. B.11: Predicted WTI_Spot vs True WTI_Spot over Time for CART Train set

Fig. B.11.5: Predicted WTI_Spot vs True WTI_Spot over Time for CART Test set

Model 3. Random Forest

Fig. B.12: Predicted vs True WTI_Spot over Time for Random Forest Train set

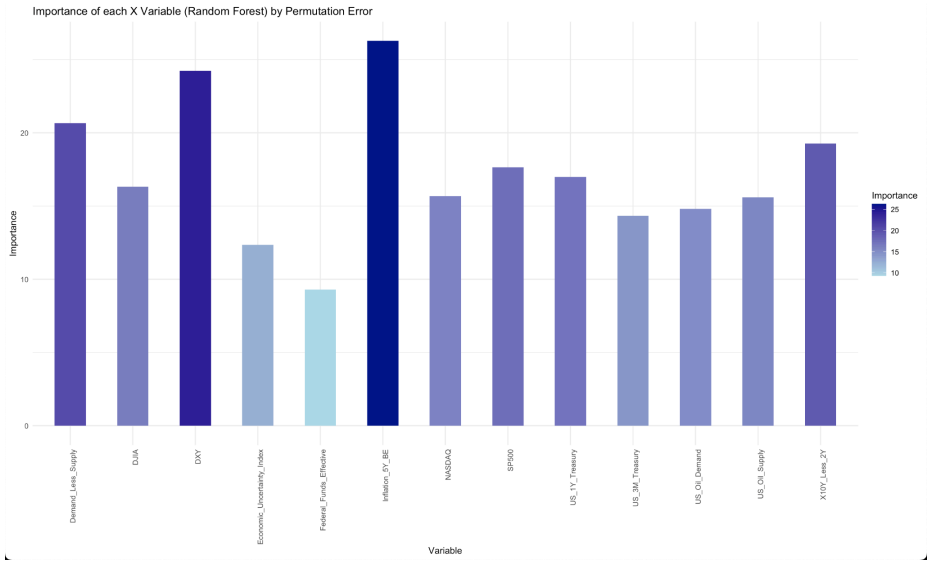


Fig. B.13: Variable Importance for Random Forest by Permutation Error

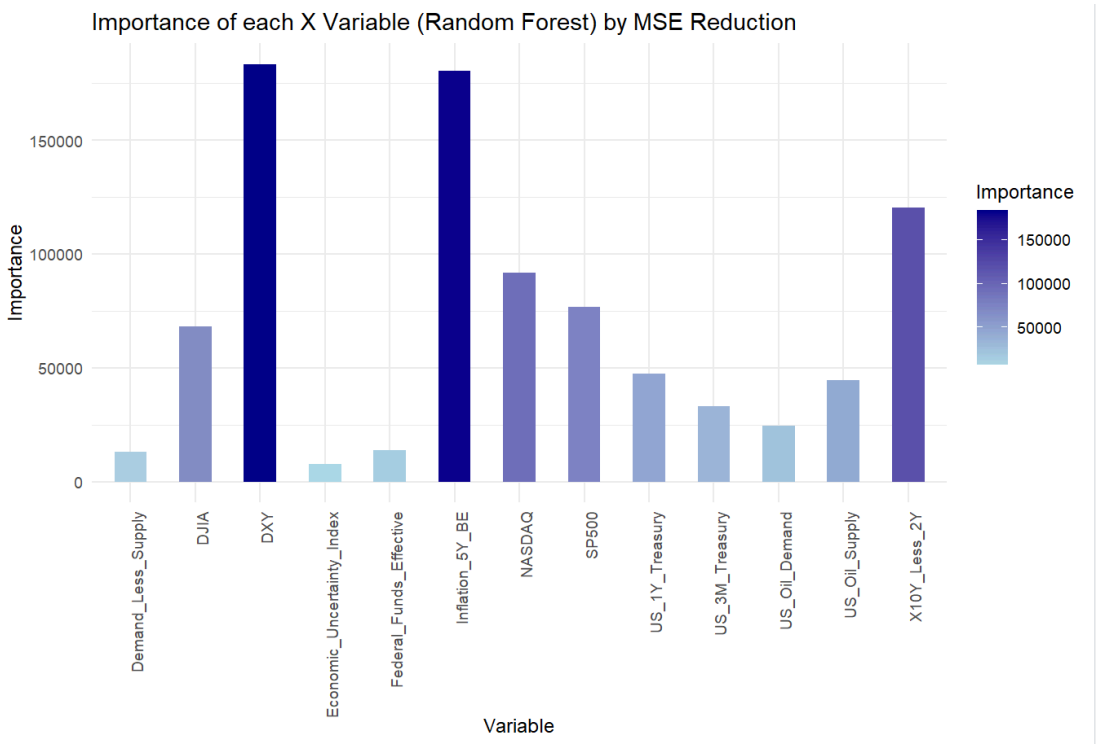


Fig. B.14: Variable Importance for Random Forest by MSE Reduction

Model 4. ARIMA Model

First Model

Dates: January 2018 ⟶ June 2019 **used to predict** June 2019 ⟶ Dec 2019

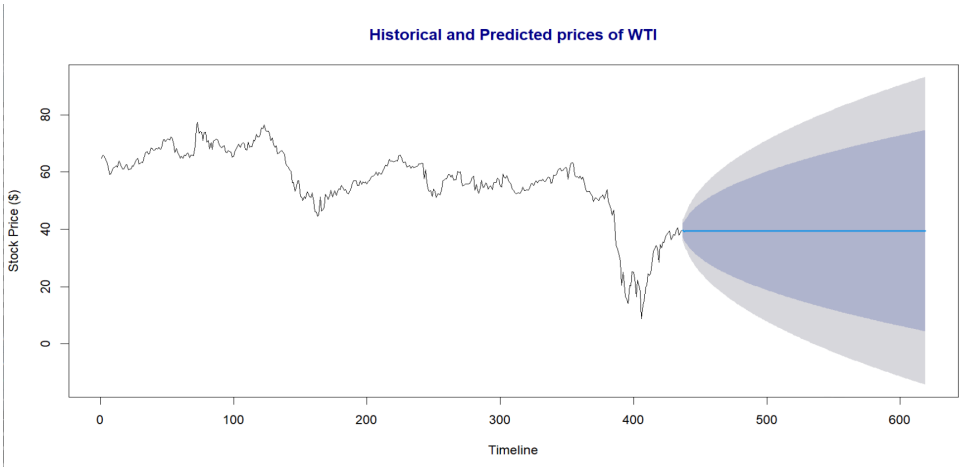


Fig. B.15: Graph of Predicted and Historical Stock Pricings

```
        Ljung-Box test

data:  Residuals from ARIMA(0,1,4)
Q* = 9.1641, df = 6, p-value = 0.1646

Model df: 4.   Total lags used: 10
```

Fig. B.16: Ljung-Box test from R-Script

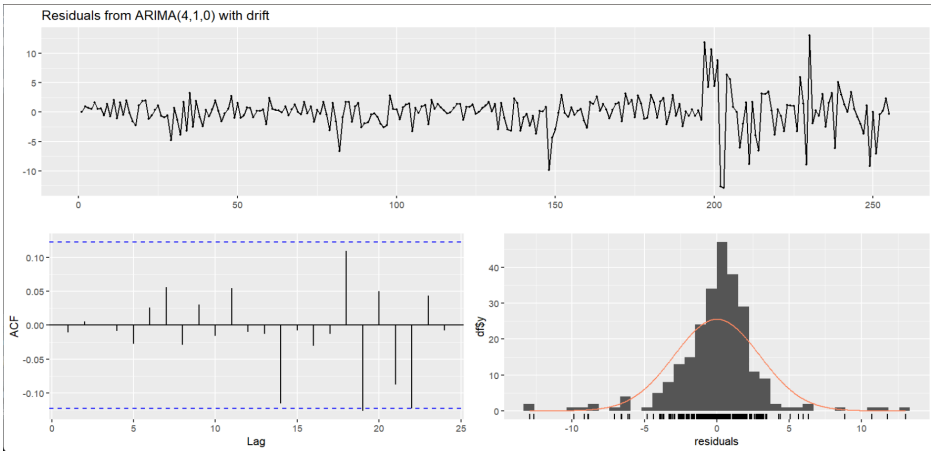


Fig. B.17: Graph to Analyse Residual Errors of Predicted Values

Second Model

Dates January 2019 ⟶ June 2020 **used to predict** June 2020 ⟶ Dec 2020

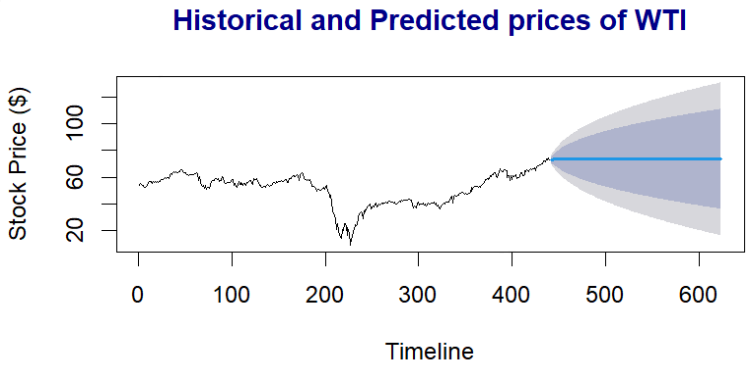


Fig. B.18: [Second Model] Graph of Stock Pricings

```
        Ljung-Box test

data:  Residuals from ARIMA(1,1,2)
Q* = 6.9371, df = 7, p-value = 0.4355

Model df: 3.   Total lags used: 10
```

Fig. B.19: [Second Model] (p-value remains greater than 0.05)

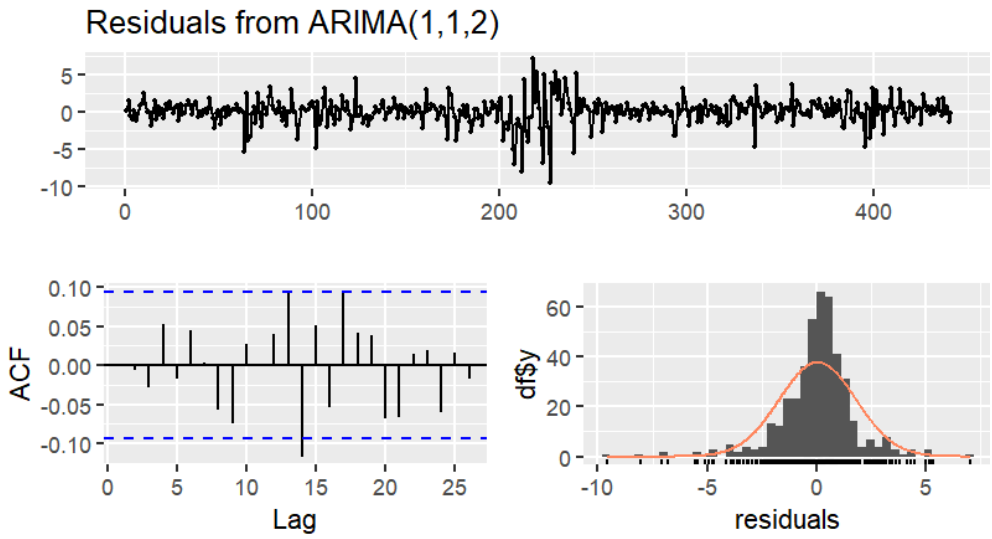


Fig. B.20: [Second Model] (Residual errors are uniform; ACF Plot has few errors beyond thresholds; Errors are normally distributed)

Third Model

Dates January 2017 → June 2018 **used to predict** June 2018 → Dec 2018

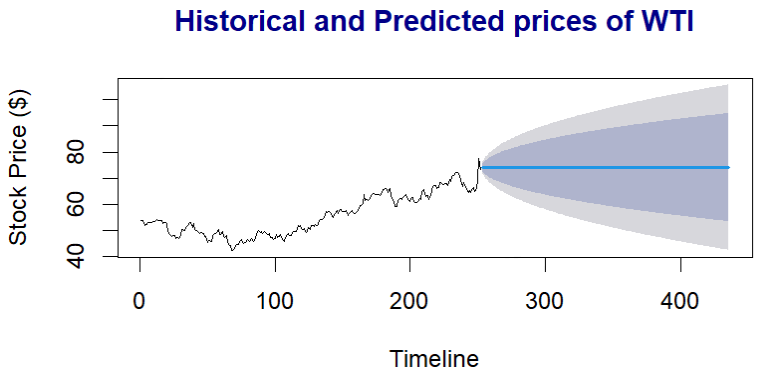


Fig. B.21: [Third Model] Graph of Stock Pricings

```
        Ljung-Box test

data:  Residuals from ARIMA(0,1,0)
Q* = 3.9438, df = 10, p-value = 0.9498

Model df: 0.   Total lags used: 10
```

Fig. B.22: [Third Model] (p-value is still greater than 0.05)

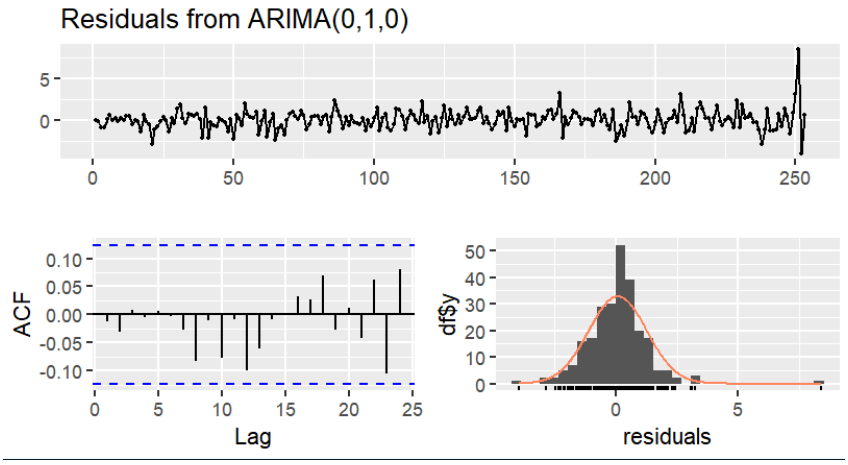


Fig. B.23: [Third Model]

Model 5. Combination of All Models

varimptlmdf

Filter

	Importance
SP500	29.333853
NASDAQ	20.755062
DJIA	22.827377
DXY	34.832891
Federal_Funds_Effective	4.577668
Economic_Uncertainty_Index	8.413154
Inflation_5Y_BE	27.623671
US_3M_Treasury	4.431950
US_1Y_Treasury	3.955914
X10Y_Less_2Y	11.739654
US_Oil_Demand	5.739268
US_Oil_Supply	17.598791

Table B.3: Linear Model Variable Importance

vifdf

Filter

	VIF
SP500	251.726965
NASDAQ	141.769402
DJIA	139.569902
DXY	4.629339
Federal_Funds_Effective	85.996155
Economic_Uncertainty_Index	2.561604
Inflation_5Y_BE	9.104567
US_3M_Treasury	210.277064
US_1Y_Treasury	94.202731
X10Y_Less_2Y	12.813947
US_Oil_Demand	2.818163
US_Oil_Supply	6.748945

Table B.4: Linear Model VIF

varimptrfdf1

Filter

	Importance
SP500	17.638851
NASDAQ	15.678167
DJIA	16.328340
DXY	24.228165
Federal_Funds_Effective	9.303287
Economic_Uncertainty_Index	12.356488
Inflation_5Y_BE	26.273207
US_3M_Treasury	14.333078
US_1Y_Treasury	16.993537
X10Y_Less_2Y	19.279195
US_Oil_Demand	14.817260
US_Oil_Supply	15.594689
Demand_Less_Supply	20.663288

Table B.5: Random Forest Variable Importance by Permutation Error

varimptrfdf2

Filter

	Importance
SP500	76898.216
NASDAQ	91621.382
DJIA	68044.932
DXY	183367.134
Federal_Funds_Effective	13741.104
Economic_Uncertainty_Index	7709.479
Inflation_5Y_BE	180426.814
US_3M_Treasury	32996.123
US_1Y_Treasury	47295.751
X10Y_Less_2Y	120187.615
US_Oil_Demand	24466.024
US_Oil_Supply	44483.051
Demand_Less_Supply	13000.689

Table B.6: Random forest Variable Importance by MSE reduction

sortedinfl

Filter

	StudRes	Hat	CookD
1825	-3.076907	0.025312361	0.018823018
2392	2.800795	0.030065447	0.018632705
2401	3.720029	0.014099944	0.015115029
2407	3.626888	0.007558864	0.007654522
1802	1.343876	0.035188146	0.005064449
1803	1.146107	0.042695953	0.004505753

Table B.7: Summary of Linear Model Influential Points

rmsedf

Filter

	Model	RMSE_Train	RMSE_Test
1	CART	2.6355647	3.701449
2	Linear Model	7.8564742	7.922597
3	Random Forest	0.8803811	1.916670

Table B.8: Comparison of RMSE Across all Models

rsqdf

Filter

	Model	R2_Train	R2_Test
1	CART	0.98627	0.97229
2	Linear Model	0.87799	0.87306
3	Random Forest	0.99847	0.99257

Table B.9: Comparison of R-Squared Across all Models