Norwegian University
of Life Sciences

Master's Thesis 2023 30 ECTS
Faculty of Social Sciences
School of Economics and Business

Crack Spread and its forecasting analysis. An empirical analysis using VAR and ARIMA model.

Md Mohsin Kabir
Master of Science in Economics

Contents

Acknowledgement

Firstly, I would like to express my boundless thanks to the almighty Allah for his favor and kindness throughout my life, without which I would not be able to come in this position (Alhamdulillah). I also would like to express my immense gratitude to my supervisor, Olvar Bergland (Associate Professor, The Norwegian Business School), for his invaluable guidance and supervision. Your intellectual views and expertise are much appreciated. I also appreciate Benedikt Goodman for his technical support in this paper. I am grateful to all my family members, relatives, friends and others who supported me to come to this stage. Finally, I remember my father who dreamed of the position where I am now.

May Allah give you all the best reward.

Abstract

In oil markets, crack spread refers to the theoretical refining margin, that is the difference between the price of crude oil and the price of refined products extracted from it. Also, this spread approximates the profit margin for the refinery companies. Even during the so-called golden age of refining, the crack spread remained remarkably steady from 1985 and 2021, averaging $10.50 per barrel. However, in recent times the difference just reached a record high of about $55. This study will analyze the historical price movements in the last decade, factors behind the recent price hike of the crack spread and its forecasting. It will investigate if Crack spread can influence WTI crude oil price movements. It will also examine whether crack spread can affect forecasting of WTI crude oil price or not. To examine the forecasting this study uses ARIMA and VAR model with their comparisons and respective performance measures. The result revealed that crack spread has an impact on predicting oil prices. The Granger Causality test found that Crack spread granger causes WTI crude oil price. Both models exhibit a well fitted model and are able to forecast with Crack spread and WTI crude oil price movements. Between these two methods VAR method fitted more well than ARIMA model as per their performance measure indicator (RMSE).

List of abbreviations

ARIMA = Autoregressive integrated moving average

AR = Autoregressive

MA = Moving Average

ADF = Augmented Dickey–Fuller test

VAR = Vector autoregression

MAE=Mean Absolute Error

MSE= Mean Square Error

RMSE = Root Mean Square Error

MAE = Mean-absolute Error

MLR = Multiple Linear Regression

TVP = Time Varying Parameter

TVP-VARs = Time Varying Parameter Vector Autoregression

AIC = Akaike Information Criteria

BIC = Bayesian Information Criteria

FPE = Akaike's Final Prediction Error criterion

HQIC = Hannan-Quinn information criterion

1. Introduction

Petroleum refiners work simultaneously on both sides of the market, acquiring crude oil while simultaneously selling gasoline, heating oil, and other distillation byproducts [1]. Crack spread is the difference between the purchase price of crude oil and the sale price of finished products, such as gasoline and distillate fuel, produced from crude oil by a refinery. Also, crack spreads are a measure of the short-term profit margin of oil refineries, since they compare the cost of crude oil inputs to the wholesale, or spot, price of outputs [2].

In recent years, the relationship between crude oil and refined product markets has been often discussed in energy economics literature. Several studies have specifically addressed the presence of a long-run equilibrium link between crude oil and refined product prices. (Asche, Gjolberg, & Völker(2003)). Recently, the price gap between these two has widened in 2022 as per Financial Times. Over the past many years, oil prices have fluctuated significantly, rising and falling sharply at various times (Bashiri et al (2013)).

The actual economy is experiencing a considerably more severe price shock than it looks, as fuel prices are growing significantly faster than crude prices. Typically, the prices of crude oil and processed products fluctuate fairly equally. The gap in between is known as the refining margin. From 1985 to 2021, the average crack spread was $10.50 per barrel. Even during the so-called golden period of refining, from 2004 to 2008, the crack spread never exceeded $30. However, in May 2022, the spread reached a record high of about $55. Crack margins for diesel and other petroleum products significantly increased (Bloomberg) [3].

This expanding crack spread and its underlying causes have been the subject of very few prior research. So, this study aims to discuss some factors which may affect this unusual behavior in crack spread. Geopolitical issues, impact of pandemic, distortions in supply and demand, and shutting down of refineries after pandemic are some of the remarkable factors (Marketplace). Bashiri & Manso studied different crude oil price forecasting methods along with its available literature. Since there are huge number of participants in the oil market and it is one of the biggest pricing factors for other commodities in the world, its forecasting analysis and finding more accurate method have been one of the most advantageous topics for them. For example, speculators and hedgers are using crack spread to predict the crude oil price. So, this study will additionally analyze the forecasting capability of different techniques (VAR and ARIMA models) to forecast crack spread and oil price. It will also strive to compare the methods for forecasting crack spread.

To perform the entire forecasting analysis, it will go through different steps such as finding the descriptive statistics, data visualization, stationarity test, auto-correlation analysis and model diagnostic tests. The study will also study model selection and validation by using different matrices such as RMSE and Mean values. It will also use AIC and BIC to ensure their accuracy and suitability for forecasting.

---

[1] Price co-movement and the crack spread in the US futures markets.

[2] Eia.gov Crack Spread.https://www.eia.gov/todayinenergy/includes/CrackSpread_Explain.php

[3] Bloomberg – "sorry, but for you, oil trades at $250 a barrel" WTI crude oil – 3:2:1 cracking margin

1.2 Research Question and objectives

There are relatively few studies that assess the current pattern of crack spread and its root causes. So, this study aims to seek the following questions:

1. Does crack spread can cause in price fluctuations of WTI crude oil?
2. Is Crack spread able to predict WTI crude oil price?
3. How does crack spread affect the price of WTI crude oil?
4. Which method can forecast crack spread better?
5. What are the factors that affect the widening crack spread?

1.3 Research Hypothesis

To get the response if there is any impact of crack spread on WTI crude oil price, Granger causality test has been used with the following hypothesis.

Hypothesis 1:

H_0: Crack spread has no effect on WTI crude oil price or crack spread does not granger causes WTI crude oil price.

H_1: Crack spread has effect on WTI crude oil price or crack spread granger causes WTI crude oil price.

To find if crack spread is able to predict WTI crude oil price, crack spread futures have been considered to as a predictor variable for the spot WTI crude oil price. Regression analysis will assist in finding the sufficient evidence against this hypothesis.

Hypothesis 2:

H_0: Crack spread cannot predict WTI crude oil price.

H_1: Crack spread can predict WTI crude oil price.

2 Literature Review

Crack spread, in the oil markets, is the link between the price of crude oil and the price of its refined product. The majority of oil market participants are subject to crack spread [4]. Silvapulle and Moosa (1999) examined the relationship between the spot and futures prices of WTI crude oil using a sample of daily data. The direction of causation is dependent on the linearity of the link, according to their findings. Non-linear causality testing finds a bidirectional effect, but linear causality testing demonstrates a causal relationship between futures and spot markets. Claudio M. (2001) utilized GARCH properties to forecast oil price distribution for short-term frame by using semiparametric methodology and bootstrap approach. The author found that an out-of-sample forecasting exercise indicates that the forecasting technique may be utilized to derive a performance measure for the future price.

According to Coppola, the indications of cointegration between the spot price and futures contracts for crude oil. Using a vector error correction model, he discovered that the information provided by the oil futures market can account for a substantial percentage of oil price fluctuations (Coppola, 2008). Murat and Tokat (2008), studied forecasting of oil price fluctuations with crack spread futures. In both the long- and short-term, they discovered the causal effect of crack spread futures on the spot oil market. They used random walk model (RWM) as a benchmark and assessed the forecasting power of crack spread futures against the crude oil futures. They found that both futures outperformed the RWM. M. Alimoradi & s. Mohajeri (2016) studied whether futures of crack spread can be a good predictor of oil price movements or not. They first examined the relationship between these two by using Vector Error Correction Model (VECM). The result showed that there is a causal relationship between crack spread futures and spot oil price in bull oil market, but in bear oil market the relationship becomes weaker.

2.1 Granger Causality

Although not identical, causality is intimately associated with the concept of cause and effect. Hume's foundational analysis states that if A (say, a billiard ball) strikes B (another ball) and causes it to move, then there is causal effect. Any analysis must have two features such as causes are asymmetrical, and causes are effective (Hoover (2006)).

However, Granger causality is a method for examining the causal relationship between two variables in a time series. It is a statistical concept of causality that is based on prediction. According to Granger Causality, if a signal X_1 granger-causes a signal X_2, then previous values of X_1 should contain information that helps to anticipate X_2 beyond what is contained in the past values of X_2 alone (Seth A. 2007). Its mathematical formulation relies on the linear regression modelling of stochastic processes (Granger 1969).

[4] M Alimoradi, S Mohajeri: Evaluation of Effectiveness of Crack Spread futures in crude oil price forecast.

In economics and econometrics, Granger (1969) causality has shown to be a valuable concept for describing dependent relationships between time series. Cross-spectral approaches are effective for expressing the relationship between two or more variables where one variable is the cause of the other variable(s) (C.W.J. Granger, 1969).

Troster, Shahbaz, and Uddin (2018) evaluated the causal link between renewable energy use, oil prices, and economic activity, taking the quantiles of the distribution into account. They used Granger-causality in quantiles analysis and discovered evidence of bi-directional causality between increases in renewable energy use and economic development at the bottom tail of the distribution.

Li, Zhang, and Yuan (2019) analyzed the correlations between investor attention and crude oil prices using the Fourier unit root test and the Granger causality test. They discovered that Granger causality occurs between investor attention and the future return of WTI crude oil. However, Obadi & Korcek (2018), investigated normal move of crude oil prices with its speculative trading in the future markets by using Granger causality method. They worked with 4 variables and found that there exists bidirectional granger causality between oil price and investment positioning of money managers.

However, this model assumes that only stationary series are involved.

In this study, it will be investigated whether there is causal relationship between Crack spread futures and WTI crude oil price based on prediction capability. Crack spread futures is generated from future contracts price where refiners sell output products and buy crude oil. Crack spread futures should Granger causes WTI crude oil price. To examine this Granger causality test have been followed.

2.2 ARIMA model for time series forecasting:

Time series data have been studied from 2005 to 2023 to do the forecasting analysis of crack spread. In the last three decades, the ARIMA method, commonly known as the Box-Jenkins method, has been one of the most extensively used linear frameworks for time-series forecasting (Zhang, 2003). Anand & Saeed (2016) studied and compared the performance of three time-series models for oil price forecasting. Compared models were Exponential Smoothing (ES), Holt-Winters (HW) and Auto Regressive Integrated Moving Average (ARIMA). To establish the best model, they used six distinct selection criteria to measure the forecasting accuracy of each model. They found ARIMA (2, 1, 2) model produced the best results, leading to infer that this sophisticated and robust model performed better in the oil market than other basic yet flexible models. The ARIMA forecasting model varies from other methods in that it presupposes no specific historical data-based forecasting trend (Gahirwal, 2013).

R. Nochai & T. Nochai (2006) examined the best model to forecast of the oil palm price of Thailand in three categories such as farm price, wholesale price and pure oil price from the period of 2000-04. They found three different combinations of ARIMA model appropriate for predicting these three categories of prices.

Conversely, Mostafa & Masry (2015), found that Gene Expression Programming (GEP) surpasses the Neural Network (NN) and ARIMA models in respect of the mean squared error, the root mean squared error and the mean absolute error.

Xie, Yu, Xu & Wang (2006), suggested an approach for crude oil price forecasting using support vector machine (SVM). They compared the performance of SVM to that of ARIMA and BPNN in order to assess its predicting capacity. However, the finding showed that SVM outperformed the other two approaches.

Ahmed & Shabri (2013) examined a study based on WTI to find a novel technique for forecasting crude oil based on RMSE and MAE to compare the proposed technique and that of ARIMA and GARCH methods. The finding showed that proposed technique surpasses both methods by their RMSE.

2.3 VAR model:

The vector autoregression (VAR) model is one of the most effective, versatile, and user-friendly techniques for analysing multivariate time series. It frequently produces more accurate forecasts than univariate time series models and complex theory-based simultaneous equation models. Forecasts derived from VAR models are very adaptable since they may be conditioned on the likely future courses of specified model variables (E Zivot, J Wang, 2006). Being linear models, they are generally straightforward to work with in theory and practice (H Lütkepohl, 2009).

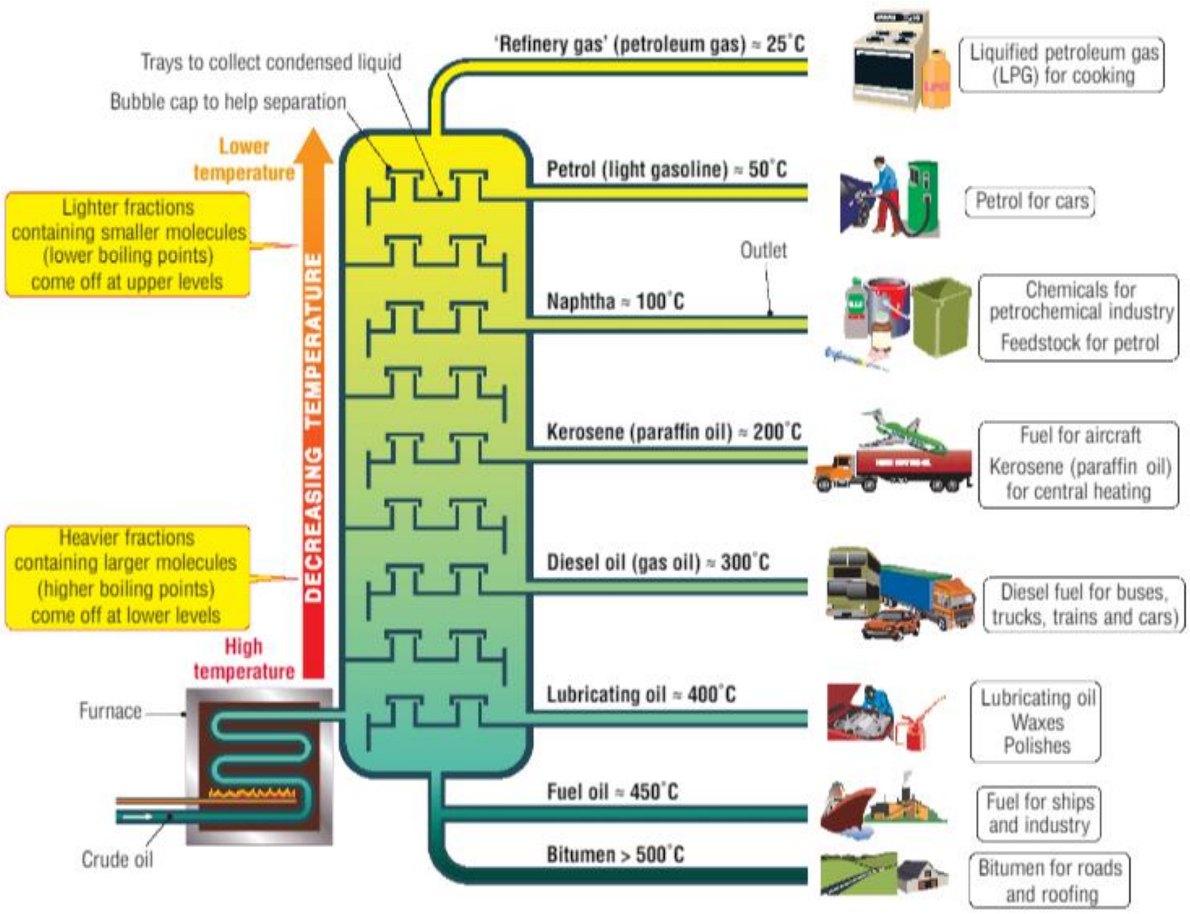
S.Bekiros, R Gupta & A Paccagnini (2015) compared the ability of VAR, standard Bayesian VARs, and time-varying parameter VARs, against random-walk and univariate AR models of real changes in oil prices and found that TVP-VARs surpassed the others in all perspectives.

However, S Mirmirani, HC Li compared VAR and neural networks to forecast U.S oil price movements. VAR-based forecast used three endogenous variables; lagged oil price, lagged oil supply and lagged energy consumption where neural networks model was made by using oil supply, energy consumption and money supply (M1). They found that Neural Networks with Genetic Algorithm remarkably outperformed VAR model where the evaluation criteria were RMSE and MAE.

2.4 Crack Spread

The spread, or gap, between the price of crude oil and the price of refined goods — gasoline and distillates — is directly proportional to refiners' earnings. This spread is referred to as a crack spread. Crack spread is a reference to the refining process that "cracks" crude oil into its principal refined products (CME).

The subsequent graph illustrates crude oil and its refined products.



2.5 Types of Crack spread

Simple crack spread: Simple crack spread, also known as 1:1 crack spread that reflects the profit margin between refined products (gasoline or diesel) and crude oil. The crack spread is accomplished by selling refined product futures (such as gasoline or diesel) and purchasing crude oil futures, effectively locking in the differential between refined products and crude oil.

Diversified Crack Spreads: This sort of spread involves numerous goods (often by-products such as gasoline, fuel oil, etc.) in a predetermined ratio, but this ratio might change based on the product mix and margin mix. Among the diversified crack spreads 3:2:1 and 5:3:2 crack spreads are very common.

3:2:1: - this ratio of crack spread is widely used by market participants where there are three crude oil futures contracts versus two gasoline futures contracts and one ULSD (heating oil) contracts.

5:3:2: - if the refiner has a lower yield of gasoline relative to distillate that will lead to this type of crack spread where refiners buy five crude oil futures contracts and sell three RBOB Gasoline futures contracts and two ULSD/heating oil contracts.

2.6 Calculation and Example of 3:2:1 Crack spread:

3:2:1 crack spread indicates Buying three Barrels of crude oil and selling two Barrels of gasoline and one Barrel of fuel oil. Let's assume the following:

Crude Oil: $60 per Barrel

Gasoline: $1.9 per Gallon

Fuel Oil: $1.2 per Gallon

Based on the above information and a 3:2:1 crack spread; at prevailing rates, this is coming out to:

Particulars	Price ($)	Units
Crude Oil	$ 60.00	Per Barrel
Gasoline	$ 1.90	Per Gallon
Fuel Oil	$ 1.20	Per Gallon
Crack Spread	$ 30.00	Per Barrel
1 Barrel is equivalent to 42 US Gallons		

Cracking margin is positive when the value of the refined product exceeds the price of the crude oil. The gross cracking margin is negative if the value of the refined product is less than that of crude oil.

3 Methodology

3.1 Stationarity:

Stationarity is a very important concept while analysing time series data. Many time series models have an assumption that the time series which will be used in those models are stationary. When examining the stationary of a series, it is one of the most often applied statistical tests. Dickey and Fuller (1979) established a method for determining if a variable has a unit root or, equivalently, if it follows a random walk. Using lag-selection methodologies, Hall. A. (2012) evaluated the influence of data-based lag-length estimates on the behaviour of the Augmented-Dicky-Fuller test for a unit root. The simulation findings showed that the performance of the ADF test is significantly enhanced when the lag duration is determined using data.

H_0: the variable contains a unit root or equivalently, the time series is non-stationary.

H_1: the variable does not contain a unit root, or equivalently, the time series is stationary.

Considering an AR (1) model: $Y_t = ØY_{t-1} + ε_1$

H0: $Ø = 1$ (non-stationary) vs H1: $Ø < 1$ (stationary)

3.1 Granger Causality:

In ordinary words, Granger Causality can be stated as X(t) granger causes Y(t), if the past values of X(t) help in predicting the future values of Y(t).

Hence, Y_t is a function of both lag of Y_t and lag of X_t.

A typical AR (1) model: $Y_t = α_0+ α_1Y_{t-1}+ε_1$

The ordinary model to test Granger Causality: $Y_t = α_0+ α_1Y_{t-1}+ α_2X_{t-1}+ ε_1$

If $α_2$ is significant that means the coefficient of X_{t-1} influences the model or improves the model, then we can say that X_t granger causes Y_t.

A generalized model: $Y_t = α_0+ α_1Y_{t-1}+ α_2Y_{t-2}+.... + α_pY_{t-p}$

$Y_t = α_0+ α_1Y_{t-1}+ α_2Y_{t-2}+.... + α_pY_{t-p}+β_1X_{t-1}+β_2X_{t-2}+.... +β_pX_{t-p}$

H_0: $β_1= β_2= β_3... β_p=0$, X_t doesn't granger causes Y_t

H_1: $β_1= β_2= β_3... β_p≠0$, X_t granger causes Y_t

Here, our WTI crude oil price is the dependent variable and crack spread futures is independent variable. Four lag variables have been considered to perform the Granger Causality test. Also, variables have been transformed as stationary, since the test assumes the data to be stationary.

Granger Causality test:

Null Hypothesis (H_0): Crack spread futures cannot predict WTI crude oil price or Crack spread futures does not granger causes WTI Crude oil price.

Null Hypothesis (H_1): Crack spread futures can predict WTI crude oil price or Crack spread futures granger causes WTI Crude oil price.

3.2 ARIMA Model:

Since its introduction by Box and Jenkins in 1976, the ARIMA model has become one of the most well-known techniques to forecasting. In this study, the Univariate Time Series Forecasting technique will be utilized because we will only be using past time series data to predict future values.

There are three standard terms of an ARIMA model, which are as follows:

p = The order of the AR term

d = The order of the MA term

q = The number of differences required to make the time series stationary.

Auto Regressive (AR) model: A Pure AR Model depends solely on its own lags. That means Y_t is a function of 'lags of Y_t'.

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \ldots \beta_p Y_{t-p} + \varepsilon_1$$

Where,

Y_t=dependent variable or the variable we want to forecast.

α = Constant term

Y_{t-1}= first lag period

Y_{t-2}=second lag period up to t-p period

β_1 and β_2 are the coefficients of lag periods up to p period

ε_1=error term

Moving Average (MR) model: A MR Model depends only on the errors of lagged forecast model.

$$Y_t = \alpha + \varepsilon_t + \emptyset_1 \varepsilon_{t-1} + \emptyset_2 \varepsilon_{t-2} + \ldots \emptyset_q \varepsilon_{t-q}$$

where, where the error terms represent the errors of the two autoregressive lag models. The errors $\boldsymbol{\varepsilon_1}$ and $\boldsymbol{\varepsilon_{t-1}}$ are the errors from the following equations:

$$Y_t = \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \ldots + \beta_0 Y_0 + \varepsilon_1$$

$$Y_{t-1} = \beta_1 Y_{t-2} + \beta_2 Y_{t-3} + \ldots + \beta_0 Y_0 + \varepsilon_{t-1}$$

ARIMA model is basically the combination of these two models which can also be stated as follows:

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \ldots + \beta_p Y_{t-p}\ \varepsilon_1 + \emptyset_1 \varepsilon_{t-1} + \emptyset_2 \varepsilon_{t-2} + \ldots\ \emptyset_q \varepsilon_{t-q}$$

We can also describe the ARIMA model as:

Predicated Y_t=Constant + Linear combination Lags of Y (up to p lags) + Linear combination of Lagged forecast errors (up to q lags).

So, to implement the ARIMA model, we are going to use Python with some stats models and other required packages. To do the analysis we first organized the data set and created lag variables along with required conversion of log values. One of our aims is to find the optimal value of p, d, and q to find the best ARIMA model. After that we fit the model to find the predicated mean values of crack spread.

3.3 VAR Model:

In the previous model, a method of univariate forecasting was applied, in which only one variable was examined. However, VAR is a multivariate technique for forecasting in which two or more time series interact. In this approach, crack spread futures and WTI crude oil price have been analyzed for forecasting both of these variables.

A typical AR model is like:

$$Y_t = C + \emptyset_1 Y_{t-1} + \emptyset_2 Y_{t-2} + \ldots\ \emptyset_p Y_{t-p} + \varepsilon_1$$

here, c is the intercept, Ø is the coefficient of lags of Y till order p and ε is the error term.

Since, we have two variables (Crack spread futures and WTI crude oil price), our VAR model would be like:

$$Y_{1,t} = C_1 + \emptyset_{11} 1Y1_{t-1} + \emptyset_{12} 1Y2_{t-1} + \varepsilon_{1,t}$$

$$Y_{2,t} = C_2 + \emptyset_{21} 1Y1_{t-1} + \emptyset_{22} 1Y2_{t-1} + \varepsilon_{2,t}$$

In this study, VAR model has been used where the max lags are five.

4 Data and Descriptive Statistics

Brent crude vs West Texas Intermediate (WTI) crude:

Brent crude, which originates from oil sources in the North Sea between the Shetland Islands and Norway, is the benchmark for the light oil market in Europe, Africa, and the Middle East. The benchmark for the U.S. light oil market, West Texas Intermediate is derived from U.S. oil fields (Source: Investopedia:1)

Brent Crude is more prominent, and most oil is priced using Brent Crude as the benchmark, equivalent to two-thirds of all oil pricing.

Since Brent Crude is produced in close proximity to the ocean, shipping expenses are considerably cheaper. In contrast, West Texas Intermediate is produced in landlocked regions, resulting in higher transportation expenses. (Source: Investopedia:2).

Brent and WTI crude oil price

Figure-1: Europe Brent Crude oil Spot price; 2023-2005

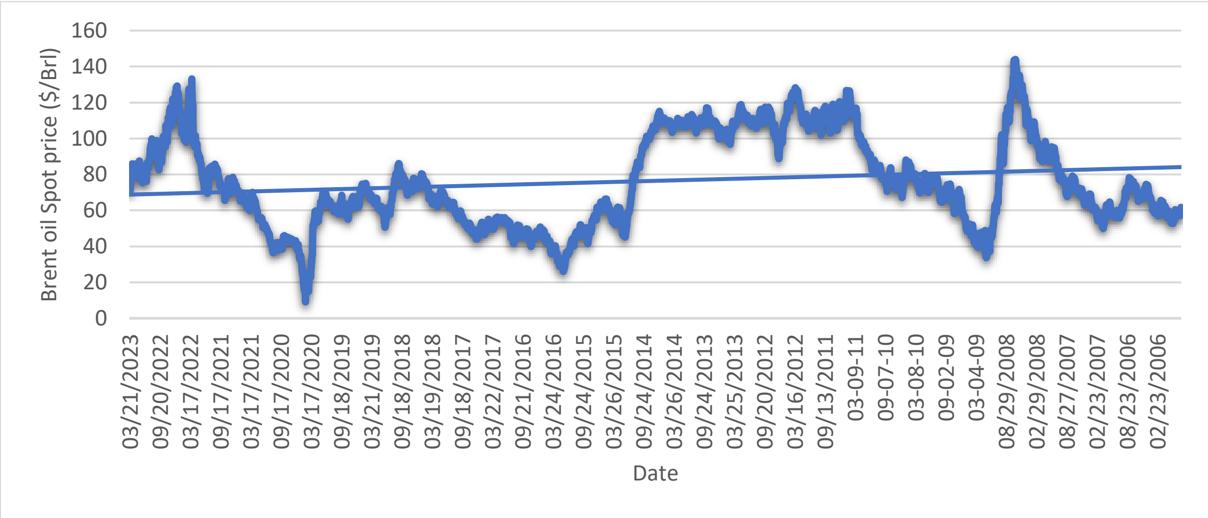
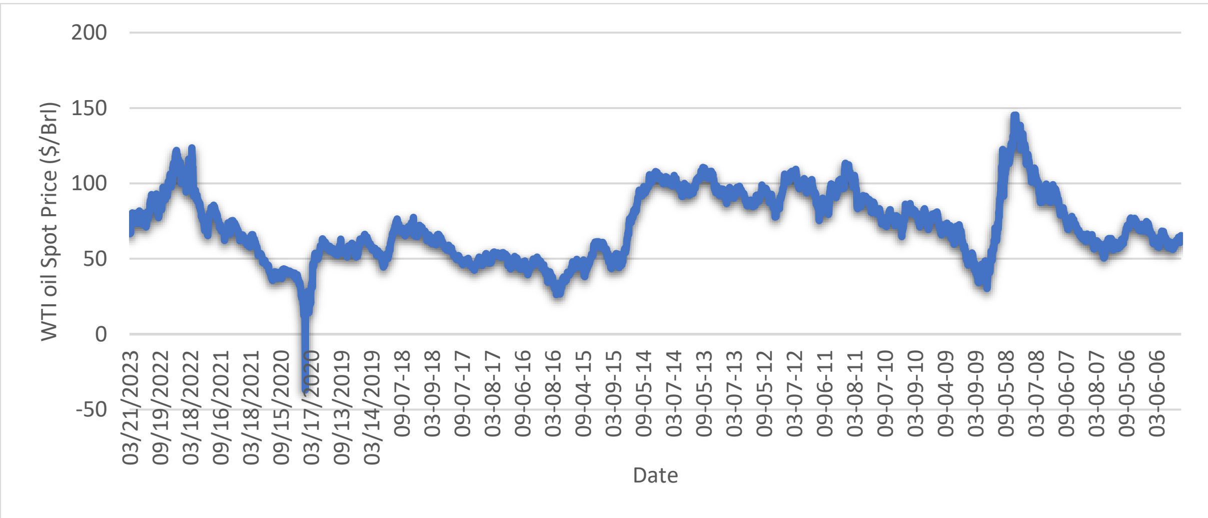


Figure-2: WTI Crude oil Spot price; 2023-2005



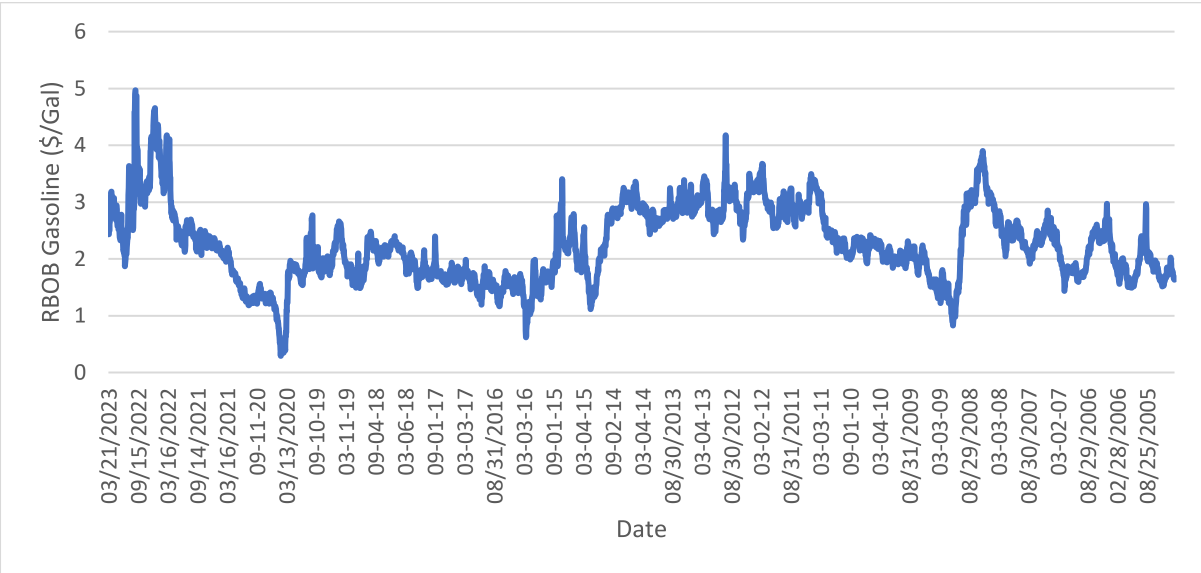
Both Brent and WTI crude oil price have almost similar movements. However, in the first quarter of 2020 both oil price had a very sharp fall. But WTI has a negative price fall due to landlocked and running out of storage in the Covid-19 time.

RBOB Gasoline:

Several distinct hydrocarbons, or long chains of molecules, make up crude oil. Longer chains result in heavier hydrocarbons with greater boiling points. By heating crude oil to various vaporization temperatures and then distilling the resultant vapors, oil refineries are able to separate the various chains. Gasoline is a combination of hydrocarbon chains with lower boiling points than water. Various quantities of these distinct chains are combined to create a consistent product for motor gasoline (Source: Investopedia:3).

Reformulated Blendstock for Oxygenate Blending (RBOB) is a component used to produce reformulated gasoline. Gasoline that has been reformulated creates less smog than other gasoline mixes. In this study RBOB Gasoline spot price has been applied from EIA database from 2005 to 2023 (Last 18 years). The data source was from RBOB Spot daily price has been used per US$ per gallon. One Gallon has been transformed into one barrel by multiplying with 42, since crude oil is marketed in barrel.

Figure-3: RBOB Gasoline Spot price; 2023-2005



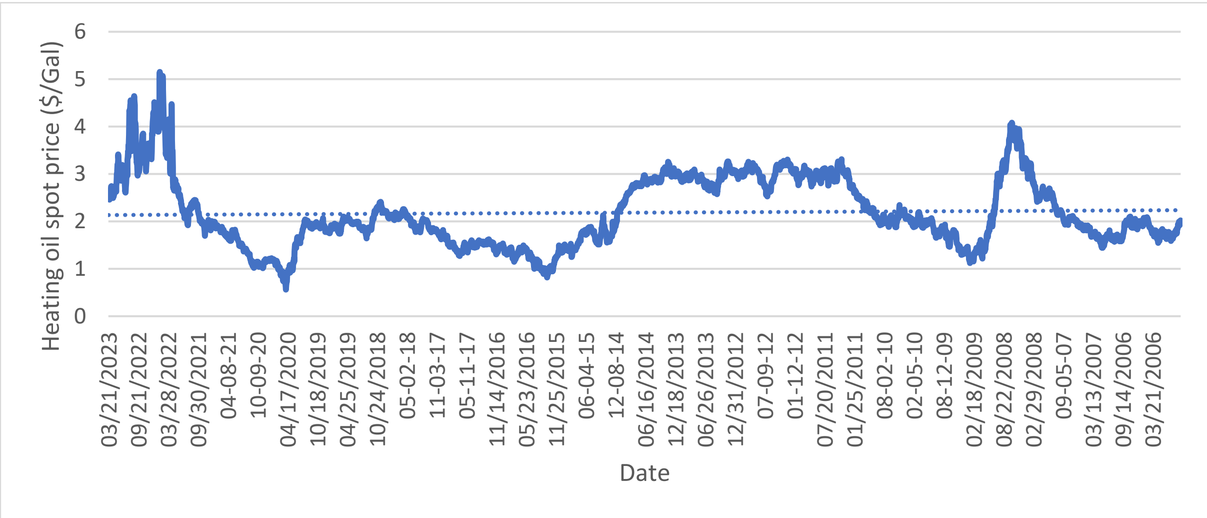
Throughout the entire time from 2005 to 2023, there are significant ups and downs in the price of RBOB gasoline. The average cost of fuel is about $2.50 per gallon. However, it had some sharp declines in 2009 and the first quarter of 2020, when the price was the lowest in the entire data set. However, after Covid's recovery, it continued to rise, reaching a high of $5 in the last quarter of 2022 before dropping to about $3.

Heating Oil:

Heating oil is mostly composed of petroleum and is utilized in furnaces, central heating systems, and industrial furnaces. Extra-light heating oil, which is an intermediate distillate, and heavy fuel oil, which is categorized as fuel oil, are the two most common forms of heating oil (Source: Oiltanking).

In this study Heating oil spot price has been applied from EIA database from 2005 to 2023.

Figure-4: Heating oil Spot price; 2023-2005



Compared to the price of gasoline, heating oil does not fluctuate as dramatically. However, throughout the duration of the entire series, it averaged out at about $2/Gallon. The price of heating oil increased after the first quarter of 2020, rising to more than $5 a gallon, just like the prices of the other two types of oil. After that, the price dropped to around $2 per gallon.

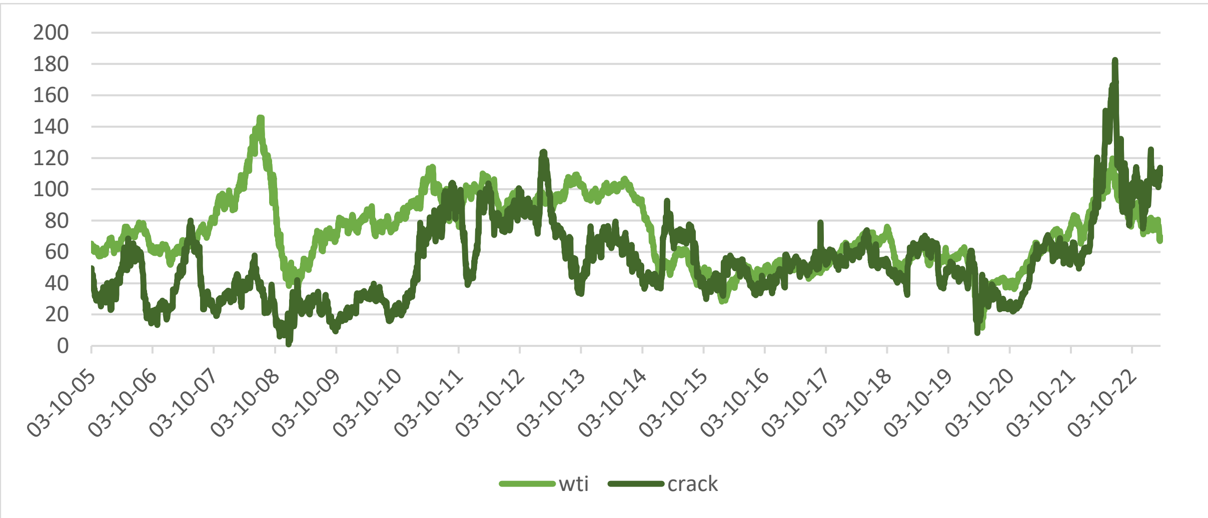
4.1 Futures contracts vs continuous futures contracts:

Futures markets have recently drawn more interest from both the trading public and finance academics (Gay, Hunter, 1983). Futures contracts and forward contracts are agreement to buy or sell an underlying asset on a specific price and a specific date in the future. These agreements allow buyers and sellers to lock in prices for physical transactions to mitigate the risk of price movements from the underlying assets through the date of delivery (CME).

Masteika, Ruthauskas & Alexandar (2012), examined a comparison of techniques for adjusting historical continuous futures data. In this study they clarified the difference between futures contracts and continuous futures contracts. The length of the futures contracts can be a day to weeks to moths or even longer, but usually are active for few months. But the problem arises when we try to analyse the historical futures data for different purposes. Contagno and backwardation factors are the primary causes of the price differences between different contracts which use the same underlying asset Pelletier, B. (2011). This price difference between two short-term futures contracts can result in inaccuracies when computing the fundamental numerical price indicators used in technical analysis. As an example, a moving average could fluctuate along with every gap which can generate a false reading. To avoid this problem, EIA database has been followed in this study where continuous futures contracts followed.

WTI Crude Oil and Crack Spread; 2005 to 2023

Figure-5: WTI Crude oil and Crack spread in futures price; 2005-2023



The graph above displays the crack spread and WTI crude oil trend from 2005 to 2023. When the crack spread reached close to the zero line, it indicated that the refiners were not making a profit or very a few, and it reached close to the zero almost two times throughout that time frame. The last downward trend, which was likely the strongest, occurred in first quarter of 2020. The remainder of the year, however, saw a record-breaking spike in crack spread. It was more than $180 per barrel.

During the first decade of the time frame which is around 2005-2015, the price movement of crack spread and WTI crude oil had a relatively higher gap even though both of these price had almost similar trend. After this period both prices had almost proportionately similar movement with small fluctuations. But the biggest gap between these two prices was occurred in first quarter of 2022.

4.2 Descriptive Statistics:

Descriptive statistics summarizes and shows the basic features of any dataset. For our dataset of crack spread and WTI crude oil has been presented below:

Table-1: Description of the data set

```
df.describe()
```

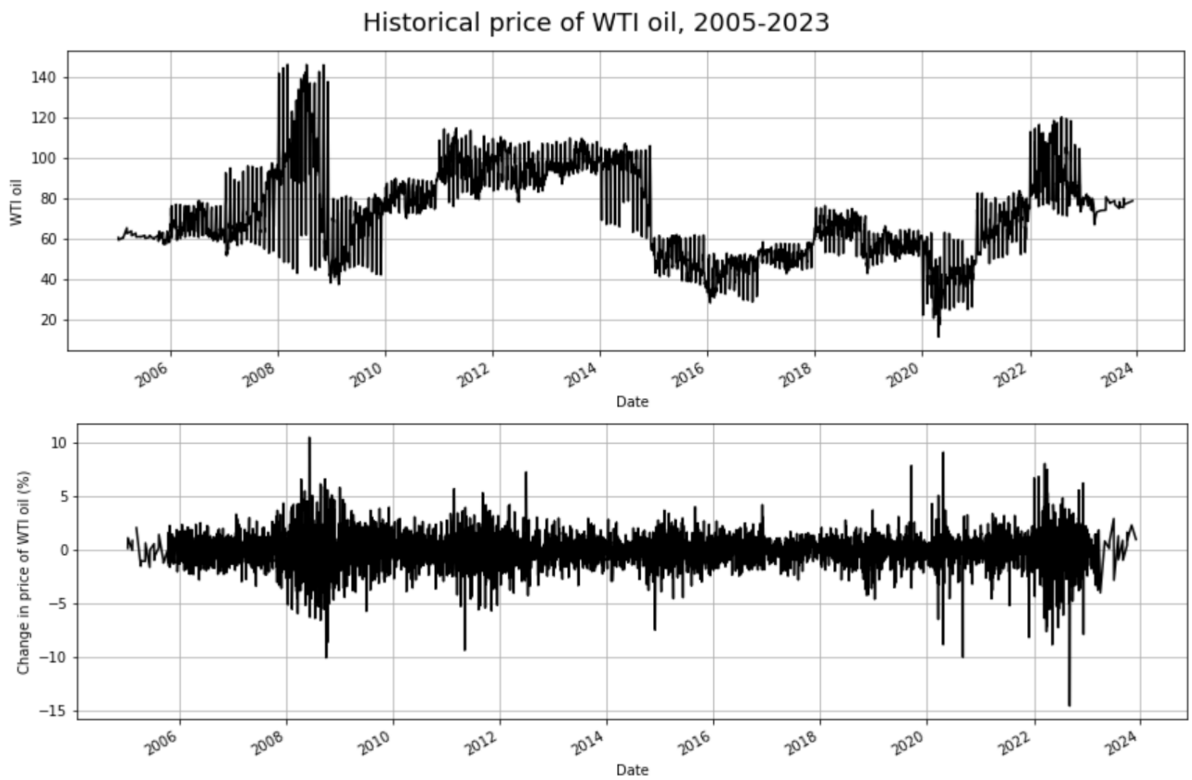
	wti	gasoline	heatoil	crack	lck	dlck	lwti	lgas	lheat
count	4396.000000	4396.000000	4396.000000	4396.000000	4396.000000	4395.000000	4396.000000	4396.000000	4396.000000
mean	72.443617	2.109841	2.231926	53.636630	3.862538	0.000181	4.233615	0.701630	0.754347
std	22.030663	0.616783	0.688064	25.705267	0.511793	0.081303	0.321656	0.306728	0.315450
min	11.570000	0.495000	0.704000	0.960000	-0.040822	-1.915083	2.448416	-0.703198	-0.350977
25%	54.367500	1.620000	1.739000	34.203000	3.532313	-0.027902	3.995767	0.482426	0.553310
50%	70.350000	2.038500	2.057000	51.081000	3.933413	0.000888	4.253483	0.712214	0.721249
75%	90.945000	2.621250	2.886000	67.107000	4.206288	0.028388	4.510255	0.963651	1.059871
max	145.860000	4.084000	4.418000	182.628000	5.207451	1.947694	4.982647	1.407077	1.485687

This study was analyzed based on total number of observations of 4396. The average price of WTI crude oil was around $72 per barrel where the max price was $145 per barrel and min price was about $11 per barrel. However, 75% of the WTI crude oil price lied below or equal to $90. The average crack spread price was around $53 per barrel where the max was approx. $182 and min value was about $0.96 per barrel. However, 75% of the crack spread price is located below or equal to $67. The mean price of Gasoline and Heating oil is $2.1 and $2.23 respectively. Standard deviation was also high for Crack spread at around $25 compared to $22 for the WTI crude oil which indicates the vulnerability of crack spread is higher than other variables.

4.3 Price of WTI crude oil and it's percentage change:

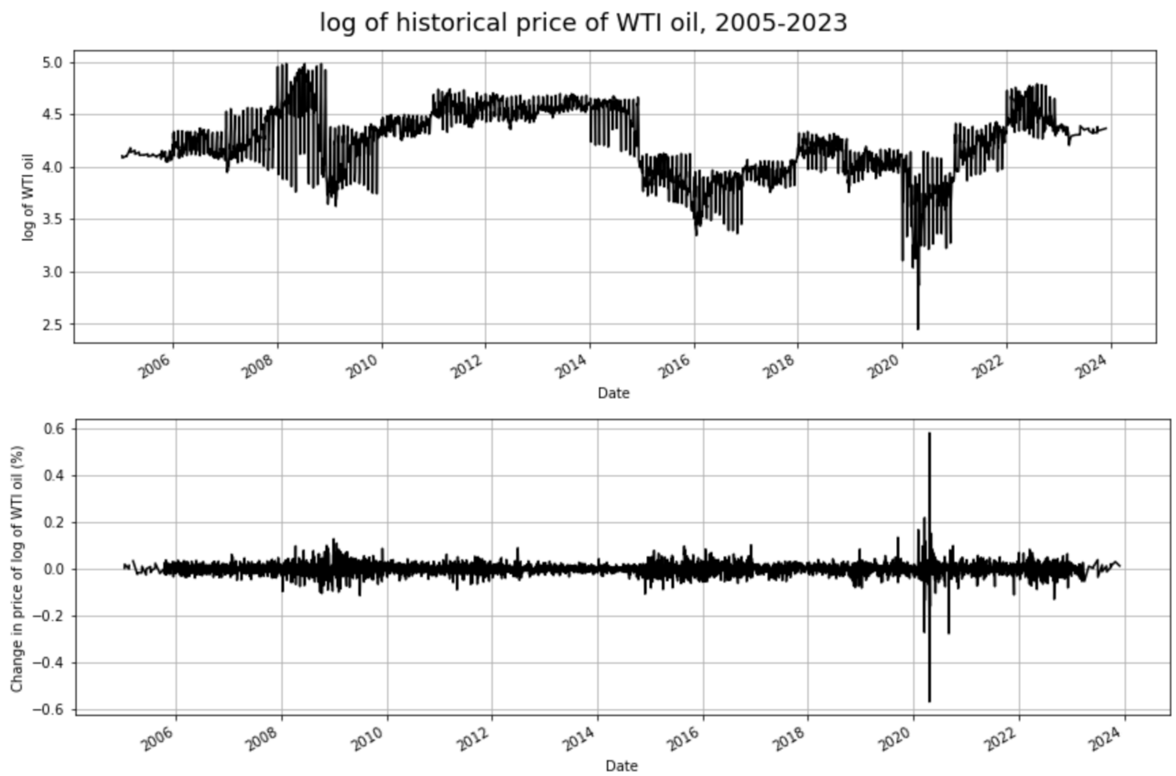
In the time of big economic recession throughout the world which is 2008-09, the price of WTI crude oil hit the peak point which was more than $140 per barrel. From 2010-2015 the price was quite stable and lower fluctuations and the change in price located between 5 to -10. But the highest spikes in change in price happened in the last quarter of 2022 and first quarter of 2023.

Figure-6: Price of WTI Crude oil and it's percentage change; 2005-2023



Log price of WTI crude oil and it's percentage change

Figure-7: Log Price of WTI Crude oil and it's percentage change; 2005-2023



However, as we are examining stock price charts, and especially comparing them over a long period of time, we should think about utilizing a log scale to lessen the impact of skewing towards large dollar values and more clearly display percentage changes.

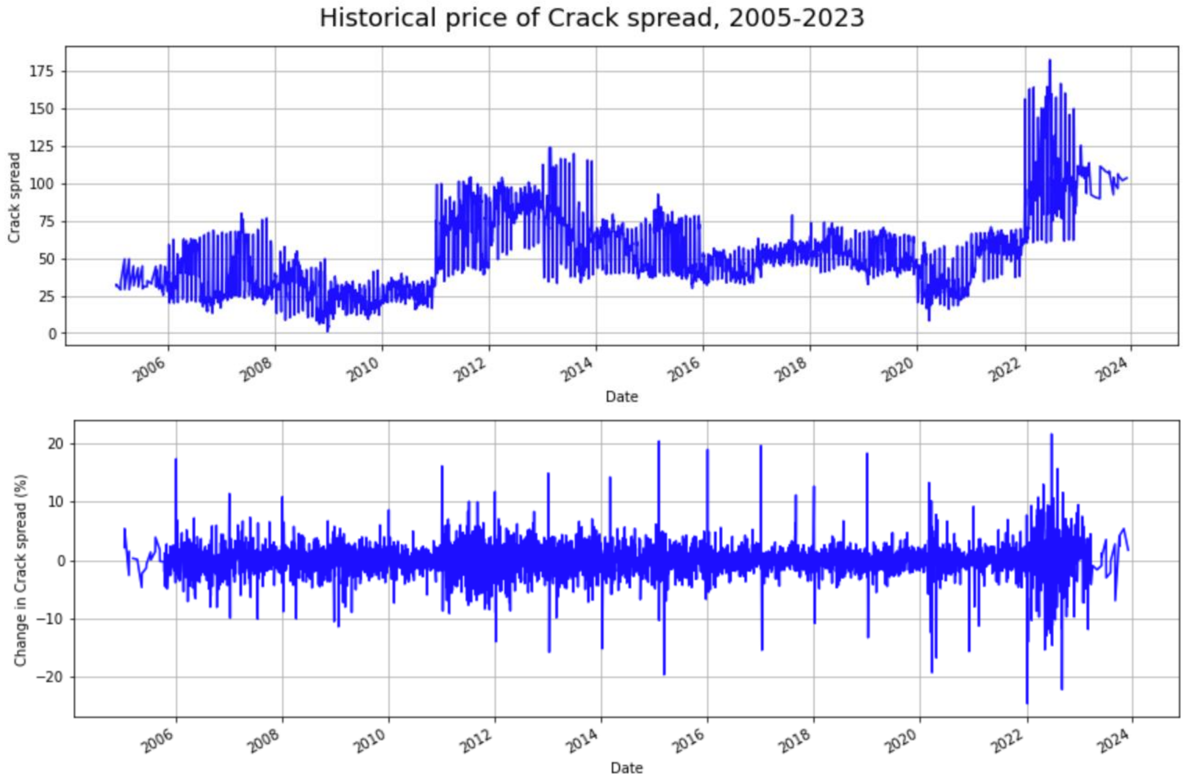
According to the above log charts, logarithmic change is quite stable compared to the previous change chart with less distortions and noise.

4.4 Price of Crack spread and it's percentage change:

The price of crack spread had quite steady fluctuations in the whole study period, except last year's price hike. From 2005 to 2011, the price was quite stable trend (around 25 to 50) which was also visible in change in price graph. The change in price was between around -10 to 10 except for one change in 2006. After that period, crack spread increased a bit which stood in between around 70-80 and couple of times above 100.

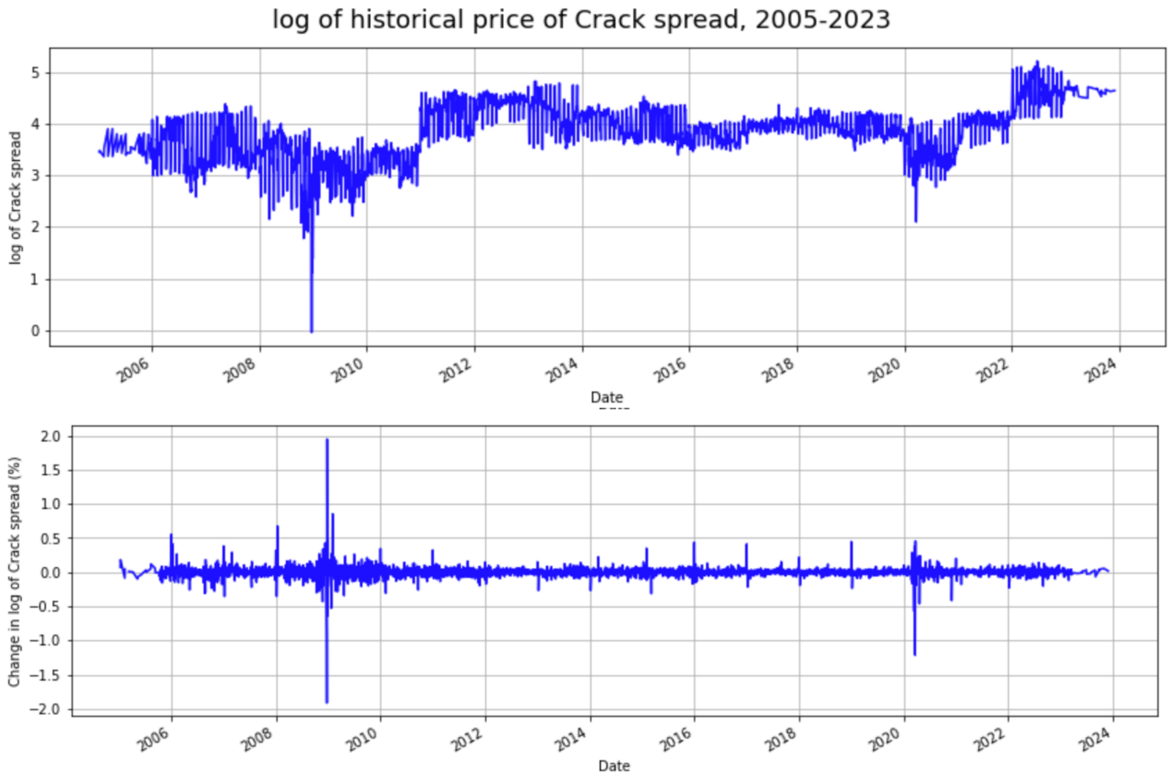
But after 2022, crack spread had couple of price hikes, and it went up to $180 till 2023. After that it showed a downward trend. It's also been focused on the change in price chart as well.

Figure-8: Price of Crack spread and it’s percentage change; 2005-2023



Log price of Crack spread and it’s percentage change

Figure-9: Log Price of Crack spread and it’s percentage change; 2005-2023



The logarithmic price and its change are shown in the graph above. The scenario for the logarithmic price movement and its change is slightly different from the previous price scenario. In this scenario, the highest price change occurred in 2009, whereas it did so in the earlier case in 2022–2023. In this instance, the price change scenario is remarkably stable.

4.5 Stationarity check of Crack Spread:

Testing data for stationarity is crucial in research when the underlying variables change over time, since otherwise the entire regression findings might be falsified. Stationary time series are those whose statistical features, such as mean, variance, and autocorrelation, remain constant across time. Most statistical forecasting techniques are predicated on the premise that time series may be made nearly stationary by mathematical modifications. A stationary series is comparatively simple to forecast.

In both ARIMA and VAR model also require the data set should be stationary in order to forecast fairly.

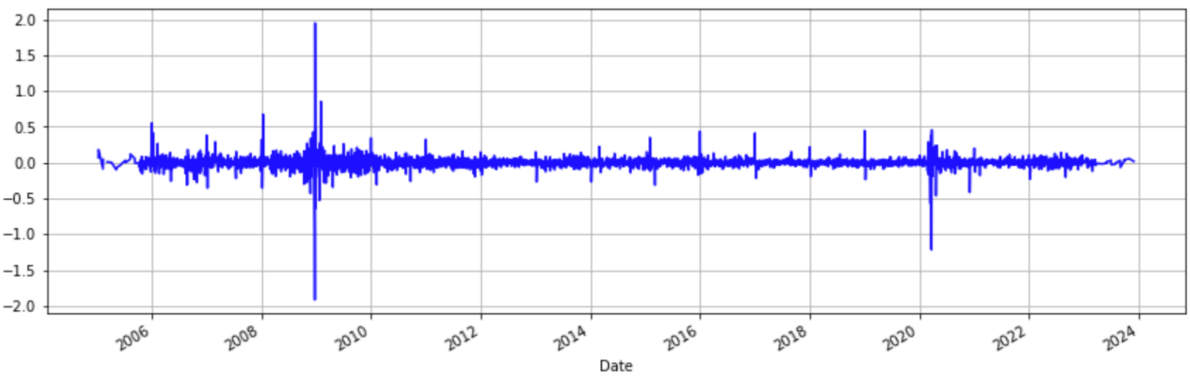
4.5.1 Stationarity check: Graphical analysis

There are numerous ways to determine if a time series is stationary or non-stationary, including direct observations, residuals, and other methods.

Graphical analysis:

We can visually examine a time series plot of the data to see if there are any noticeable trends or seasonality. As per Noyni (2008), we can determine whether a series is stationary or not by examining the time plot of that series.

Graph-10: First difference of Crack spread; 2005-2023



In this graphical presentation, it is easily visible that this data set is stationary, since it is not showing any increasing or decreasing trend over the time period.

4.5.2 Stationarity check: Dicky Fuller (ADF) test

After running a ADF test for the existing data set without doing any modification:

Result of ADF test for WTI crude oil

```
Observations of Dickey-fuller test for WTI Crude oil
ADF Statistic: -2.831342
P-value: 0.053934
Critical values:
        1%: -3.432
        5%: -2.862
        10%: -2.567
Failed to reject Ho - Time series is Non-stationary
```

We have found the result of ADF test with p-value of 0.053934 which is higher than 0.05 and we fail to reject the null hypothesis that the series is non-stationary. So, we need to do some mathematical modifications to the data set. So, first difference of WTI crude oil has been taken to make the data stationary and the result we've got as follows:

Result of ADF test for WTI crude oil after first differencing

```
Observations of Dickey-fuller test for WTI Crude oil after first differencing
ADF Statistic: -10.070991
P-value: 0.000000
Critical values:
        1%: -3.432
        5%: -2.862
        10%: -2.567
Reject Ho - Time series is Stationary
```

So, now our data set has become stationary where the p-value is almost 0.0 and it is fit for forecasting.

Similarly, we can do a similar testing process for Crack spread. So, the result of the ADF test before doing any modification to the data set is follows:

Result of ADF test for Crack spread without modification

```
Observations of Dickey-fuller test
ADF Statistic: -2.828947
P-value: 0.054256
Critical values:
        1%: -3.432
        5%: -2.862
        10%: -2.567
Failed to reject Ho - Time series is Non-stationary
```

We have found the result of ADF test with p-value of 0.054256 which is higher than 0.05 and we fail to reject the null hypothesis that the series is non-stationary. So, we need to do some mathematical modifications to the data set. So, first difference of WTI crude oil has been taken to make the data stationary and the result we've got as follows:

```
Observations of Dickey-fuller test after first differencing
ADF Statistic: -13.720592
P-value: 0.000000
Critical values:
        1%: -3.432
        5%: -2.862
        10%: -2.567
Reject Ho - Time series is Stationary
```

So, now our data set has become stationary where the p-value is almost 0.0 and it is fit for forecasting.

5 Results and discussion:

5.1 Granger Causality test: Hypothesis-1:

Granger Causality test: Crack spread granger causes WTI Crude oil:

There are many participants in the oil market and among those, crack spread can be a tool of speculation oil price movement. Specifically, it can be used by speculators and investors as a market indicator of crude oil and refined product price fluctuations.

As stated before, Granger causality test can be a good tool for testing the relationship between oil price and crack spread. To do the Granger causality test necessary modification have been done to make the data stationary. After that regression analysis has been performed where the dependent variable is WTI crude oil and independent variables are differentiated lags of WTI crude oil and lags of Crack spread. Regression analysis has been presented below:

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                   dwti   R-squared:                       0.105
Model:                            OLS   Adj. R-squared:                  0.103
Method:                 Least Squares   F-statistic:                     10.71
Date:                Wed, 19 Jul 2023   Prob (F-statistic):           3.60e-18
Time:                        13:33:12   Log-Likelihood:                -14833.
No. Observations:                4389   AIC:                         2.969e+04
Df Residuals:                    4378   BIC:                         2.976e+04
Df Model:                          10
Covariance Type:                  HAC
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      0.0065      0.102      0.064      0.949      -0.194       0.207
dwti_L1       -0.2086      0.030     -6.964      0.000      -0.267      -0.150
dwti_L2       -0.2114      0.032     -6.558      0.000      -0.275      -0.148
dwti_L3       -0.2327      0.030     -7.728      0.000      -0.292      -0.174
dwti_L4       -0.1909      0.026     -7.309      0.000      -0.242      -0.140
dwti_L5       -0.0904      0.018     -4.958      0.000      -0.126      -0.055
dcrack_L1      0.0031      0.014      0.219      0.826      -0.025       0.031
dcrack_L2     -0.0342      0.013     -2.600      0.009      -0.060      -0.008
dcrack_L3      0.0347      0.014      2.419      0.016       0.007       0.063
dcrack_L4      0.0119      0.014      0.880      0.379      -0.015       0.038
dcrack_L5      0.0081      0.011      0.766      0.444      -0.013       0.029
==============================================================================
Omnibus:                      935.512   Durbin-Watson:                   2.010
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            36521.202
Skew:                          -0.106   Prob(JB):                         0.00
Kurtosis:                      17.130   Cond. No.                         11.9
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using

5 lags and without small sample correction
```

Here, we can observe that F-statistic of the model is very high, and the p-value of the F-statistic is very close to zero, which are an indication of a good-fitted model. Also, the AIC and BIC score is also proximate to zero.

The intercept term is very close to zero which indicates a neutral or a very small positive relationship between WTI crude oil and crack spread, even though its statistically insignificant. However, all the parameters of lag coefficients of WTI crude oil which are up to 5 lag period have negative values which indicates a negative relationship. All coefficients are also statistically significant. Lag period parameters of crack spread are almost statistically insignificant except the second lag period which is negatively related and has the p-value of approx. 0.0.

However, if we run the Granger Causality test based on the above regression analysis, we get the following result:

```
Testing if crack spread Granger causes Wti oil:
F-stat : 3.1724086893959282
p-value: 0.007319698726738805
```

The test result clearly states that Crack spread Granger causes WTI crude oil, since the F-statistic is 3.17 and p-value of the F-statistic is nearly zero.

So, from our above test analysis, it can be asserted that we have enough evidence to reject the null hypothesis and accept the alternate hypothesis.

Granger Causality test: WTI crude oil Granger causes Crack spread:

In addition to the above hypothesis, we can find whether it can be explained in the other way. In other words, we can find **if WTI crude oil granger causes Crack spread**. We can follow the same methods to find that answer. So, the regression analysis result is given below:

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                 dcrack   R-squared:                       0.108
Model:                            OLS   Adj. R-squared:                  0.106
Method:                 Least Squares   F-statistic:                     24.36
Date:                Wed, 19 Jul 2023   Prob (F-statistic):           2.54e-45
Time:                        13:35:05   Log-Likelihood:                -16296.
No. Observations:                4389   AIC:                         3.261e+04
Df Residuals:                    4378   BIC:                         3.268e+04
Df Model:                          10
Covariance Type:                  HAC
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      0.0278      0.142      0.196      0.845      -0.251       0.307
dwti_L1       -0.0785      0.028     -2.820      0.005      -0.133      -0.024
dwti_L2       -0.0194      0.026     -0.746      0.456      -0.070       0.032
dwti_L3       -0.0979      0.023     -4.219      0.000      -0.143      -0.052
dwti_L4       -0.0207      0.019     -1.075      0.282      -0.058       0.017
dwti_L5        0.0187      0.018      1.067      0.286      -0.016       0.053
dcrack_L1     -0.1862      0.026     -7.130      0.000      -0.237      -0.135
dcrack_L2     -0.2442      0.025     -9.867      0.000      -0.293      -0.196
dcrack_L3     -0.1723      0.021     -8.042      0.000      -0.214      -0.130
dcrack_L4     -0.1859      0.021     -8.788      0.000      -0.227      -0.144
dcrack_L5     -0.1205      0.017     -7.024      0.000      -0.154      -0.087
==============================================================================
Omnibus:                      909.823   Durbin-Watson:                   2.013
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            20478.227
Skew:                          -0.406   Prob(JB):                         0.00
Kurtosis:                      13.551   Cond. No.                         11.9
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using

5 lags and without small sample correction
```

The above model has similarities with the previous model. It can also be stated as a good-fitted model as per F-statistic and p-value of F-statistic. AIC and BIC scores are also near to zero. Intercept term is showing positive relationship despite it is statistically insignificant.

Parameters of first and third lag difference coefficients of WTI crude oil are negatively related and also statistically significant. Parameters of second, fourth and fifth lag difference coefficients of WTI crude oil are statistically insignificant. However, all parameters of lag difference coefficients of crack spread are negatively related and statistically significant.

If we run the Granger causality test, we have the following result:

```
Testing if Wti Granger causes crack spread:
F-stat : 5.478503758984618
p-value: 4.949035181207948e-05
```

Since the F-statistic is 5.47 and the p-value of the F-statistic is almost zero, the test result clearly shows that WTI crude oil granger causes Crack spread.

5.2 Crack spread predictability: Hypothesis-2:

In order to find out whether crack spread can predict WTI crude oil price, regression has been performed where WTI crude oil price is the response variable and crack spread along with its lag variables, gasoline, and heating are the explanatory variables. This regression analysis is different from the previous regressions because of the new variables. Regression result is:

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                   lwti   R-squared:                       0.994
Model:                            OLS   Adj. R-squared:                  0.994
Method:                 Least Squares   F-statistic:                 1.028e+05
Date:                Tue, 25 Jul 2023   Prob (F-statistic):               0.00
Time:                        19:09:17   Log-Likelihood:                 9963.2
No. Observations:                4392   AIC:                        -1.991e+04
Df Residuals:                    4384   BIC:                        -1.986e+04
Df Model:                           7
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      4.1312      0.003   1322.036      0.000       4.125       4.137
lgas           0.7907      0.004    204.243      0.000       0.783       0.798
lheat          0.4101      0.004    114.381      0.000       0.403       0.417
lcrack        -0.1692      0.005    -35.822      0.000      -0.178      -0.160
lcrack_L1     -0.0142      0.006     -2.247      0.025      -0.027      -0.002
lcrack_L2     -0.0083      0.006     -1.312      0.189      -0.021       0.004
lcrack_L3      0.0069      0.006      1.098      0.272      -0.005       0.019
lcrack_L4     -0.0125      0.005     -2.652      0.008      -0.022      -0.003
==============================================================================
Omnibus:                     2966.181   Durbin-Watson:                   0.213
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            94807.082
Skew:                          -2.769   Prob(JB):                         0.00
Kurtosis:                      25.077   Cond. No.                         202.
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using 4 lags and without small sample correction
```

The value of R-squared and Adj. R-squared is 0.994 which tells us that almost 99% of the variation of the WTI crude oil price is explained by the prices of dependent variable. F-Statistic is very high, and the P-value of F-statistic is 0.00 which is definitely less than 0.05 which indicates that it is statistically significant. It also shows the overall significance of the regression model. This model's predictor variables are useful for explaining the variance in the response variable.

Our second hypothesis says that crack spread has no effect on WTI oil price. So, our coefficients of crack spread should be zero. But in our result showed it has negative coefficients of value -0.1692 with a p-value of 0.00 which is statistically significant. So, we can reject the null hypothesis and accept the alternate hypothesis. However, this crack spread coefficient states that a one percent increase in crack spread price will result about 17$ decrease in WTI crude oil price.

However, all the coefficients including the intercept term are statistically significant, as the p-value of all these coefficients are close to zero. Only exceptional is the second and third lag variable of the crack spread. The coefficients of these variables are not statistically insignificant which has p-value of higher than 0.05.

The average value of WTI crude oil price would be around 4 even if the value of the rest of the coefficients are zero. Gasoline and heating oil price have a positive impact on WTI crude oil price. But crack spread and its lag variables are negatively related to WTI crude oil price.

In the above regression model, the value of Durbin-Watson test (a test for determining the presence of autocorrelation) is 0.213 which indicates the presence of positive autocorrelation in the residuals. It implies that if the price of oil fell yesterday, it is likely to fall today as well.

5.3 Forecasting of WTI crude oil and Crack spread using ARIMA model:

5.3.1 ARIMA model for WTI crude oil forecasting:

The first step in completing the ARIMA model is to determine the best order for the ARIMA model. By using auto-Arima function we can complete this step very easily. Auto-Arima is going to try different combination of orders and for every combination it will assign a score which is called AIC. The lower the AIC, the better the fit. Eventually, it has showed us the best model for our data to forecast WTI crude oil, as follows:

```
Performing stepwise search to minimize aic
 ARIMA(2,1,2)(0,0,0)[0] intercept   : AIC=-19140.921, Time=1.63 sec
 ARIMA(0,1,0)(0,0,0)[0] intercept   : AIC=-19091.059, Time=0.48 sec
 ARIMA(1,1,0)(0,0,0)[0] intercept   : AIC=-19138.626, Time=0.28 sec
 ARIMA(0,1,1)(0,0,0)[0] intercept   : AIC=-19141.688, Time=0.80 sec
 ARIMA(0,1,0)(0,0,0)[0]             : AIC=-19093.055, Time=0.22 sec
 ARIMA(1,1,1)(0,0,0)[0] intercept   : AIC=-19143.323, Time=0.25 sec
 ARIMA(2,1,1)(0,0,0)[0] intercept   : AIC=-19139.446, Time=1.40 sec
 ARIMA(1,1,2)(0,0,0)[0] intercept   : AIC=-19139.938, Time=0.66 sec
 ARIMA(0,1,2)(0,0,0)[0] intercept   : AIC=-19142.538, Time=0.58 sec
 ARIMA(2,1,0)(0,0,0)[0] intercept   : AIC=-19140.766, Time=0.37 sec
 ARIMA(1,1,1)(0,0,0)[0]             : AIC=-19145.318, Time=0.39 sec
 ARIMA(0,1,1)(0,0,0)[0]             : AIC=-19143.683, Time=0.40 sec
 ARIMA(1,1,0)(0,0,0)[0]             : AIC=-19140.621, Time=0.14 sec
 ARIMA(2,1,1)(0,0,0)[0]             : AIC=-19143.457, Time=1.08 sec
 ARIMA(1,1,2)(0,0,0)[0]             : AIC=-19141.933, Time=0.39 sec
 ARIMA(0,1,2)(0,0,0)[0]             : AIC=-19144.533, Time=0.83 sec
 ARIMA(2,1,0)(0,0,0)[0]             : AIC=-19142.761, Time=0.20 sec
 ARIMA(2,1,2)(0,0,0)[0]             : AIC=-19142.916, Time=0.98 sec

Best model:  ARIMA(1,1,1)(0,0,0)[0]
Total fit time: 11.086 seconds
```

SARIMAX Results

Dep. Variable:	y	**No. Observations:**	4392
Model:	SARIMAX(1, 1, 1)	**Log Likelihood**	9575.659
Date:	Tue, 25 Jul 2023	**AIC**	-19145.318
Time:	21:29:58	**BIC**	-19126.156
Sample:	0	**HQIC**	-19138.557
	- 4392		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	0.2746	0.061	4.503	0.000	0.155	0.394
ma.L1	-0.3853	0.059	-6.578	0.000	-0.500	-0.270
sigma2	0.0007	2.47e-06	302.638	0.000	0.001	0.001

Ljung-Box (L1) (Q):	0.00	**Jarque-Bera (JB):**	1339223.44
Prob(Q):	0.98	**Prob(JB):**	0.00
Heteroskedasticity (H):	2.13	**Skew:**	-1.43
Prob(H) (two-sided):	0.00	**Kurtosis:**	88.51

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).

The first table performed a stepwise search to minimize AIC to find the best model. It is obvious that a SARIMAX order of (1,1,1) is the best fit order. SARIMAX includes seasonality, auto regression, integration, moving average and exogenous factors. However, auto ARIMA takes SARIMAX model as the backbone. Here, autoregressive order is 1, integration is 1, and moving average is also 1.

However, all the coefficients of the parameters like autoregressive, moving average and sigma are statistically significant, since the p-value is 0.00 for all of these.

According to the model summary, the model satisfies the criteria of residual independence (no correlation) since the p-value of the Ljung-Box test (Prob(Q)) is larger than 0.05. Therefore, we cannot reject the independence null hypothesis. Additionally, we can assert that the residual distribution is homoscedastic (constant variance) since the p-value of the Heteroskedasticity test (Prob(H)) is likewise greater than 0.05.

Now, if we fit the model into training data set to train the model and to check on the testing data set, if it’s a good fit or not. We have used the first 2100 data set as a training data set and the last 2100 as testing data set.

The result of training data set:

SARIMAX Results

Dep. Variable:	lwti	**No. Observations:**	2292
Model:	ARIMA(1, 1, 1)	**Log Likelihood**	5594.832
Date:	Tue, 25 Jul 2023	**AIC**	-11183.664
Time:	21:29:59	**BIC**	-11166.454
Sample:	0	**HQIC**	-11177.388
	- 2292		
Covariance Type:	opg		

	coef	**std err**	**z**	**P>|z|**	**[0.025**	**0.975]**
ar.L1	0.0974	0.285	0.342	0.732	-0.461	0.656
ma.L1	-0.1415	0.284	-0.498	0.618	-0.698	0.415
sigma2	0.0004	7.61e-06	58.198	0.000	0.000	0.000

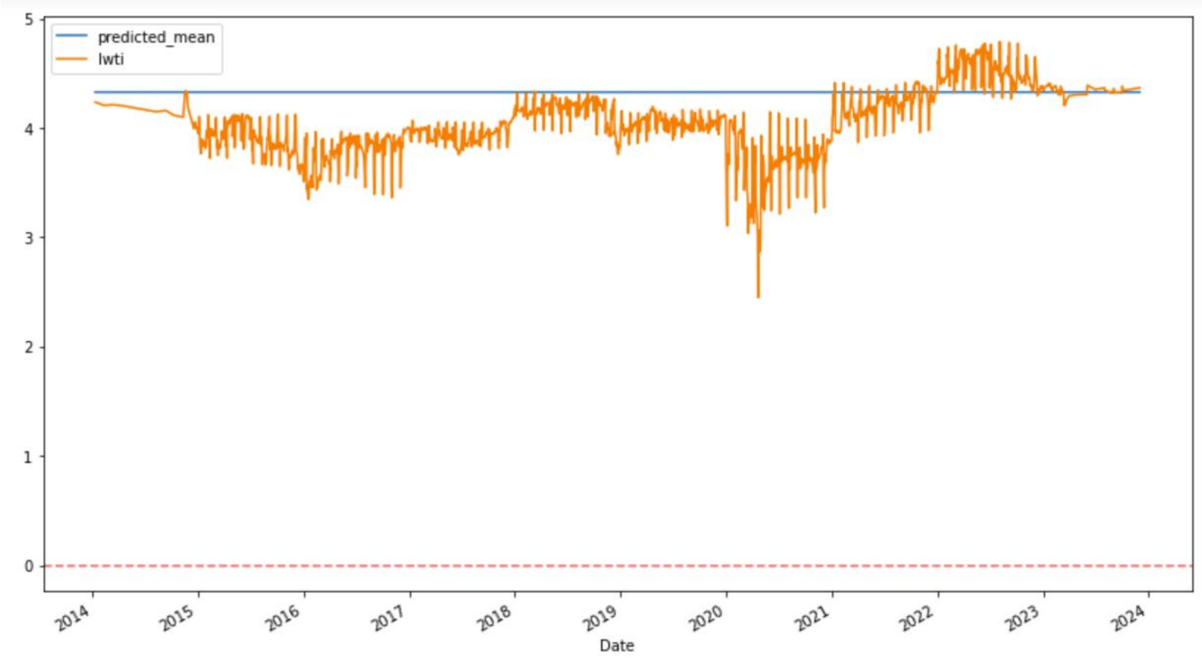
Ljung-Box (L1) (Q):	0.00	**Jarque-Bera (JB):**	1506.72
Prob(Q):	1.00	**Prob(JB):**	0.00
Heteroskedasticity (H):	0.42	**Skew:**	-0.15
Prob(H) (two-sided):	0.00	**Kurtosis:**	6.96

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).

In the above SARIMA training model regression result, auto regressive and moving average coefficients are not statistically significant, since the p-value of these coefficients are greater than 0.05. It indicates that the auto regression and moving average coefficients have less importance in forecasting WTI crude oil price. But the exogenous factor which is explained by sigma2 is statistically significant.

However, if we plot the predicted values for the testing data set, we get the following graph:

Graph-11: Predicted mean values vs actual WTI crude oil price.



From the above graph it is certain that the actual WTI crude oil price did not follow the predicted mean price at the beginning. But actual price trend was lower than the predicted price till 2022. After that actual price moved higher than the predicted price for the year first quarter of 2023. But in the end both predicted mean price and actual price have become parallel. So, by the graphical representation the model seems a good-fitted model. We can state how exactly the model is good or bad it is by calculating the following performance measure indicators.

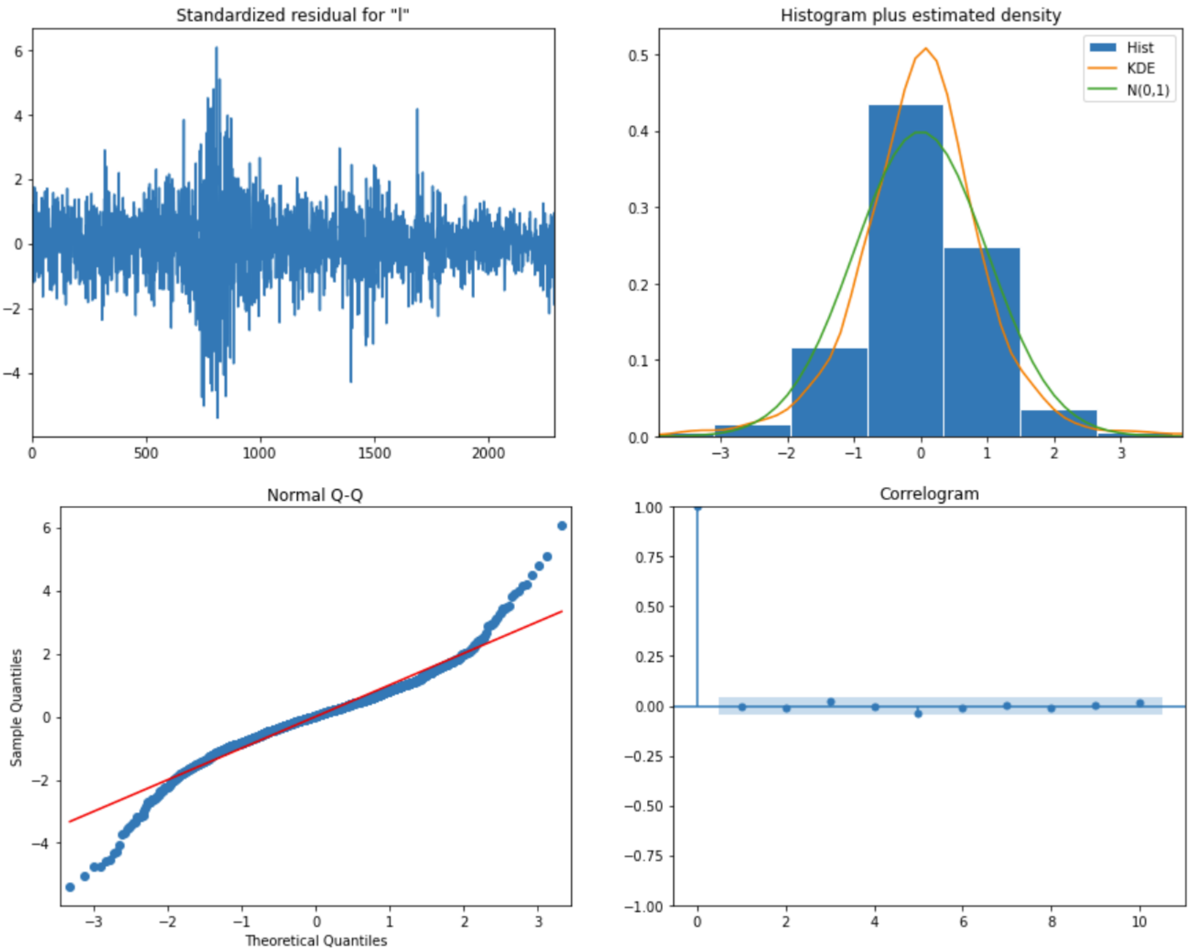
Testing if it's a good-fitted model	
Mean	4.03
MSE	0.168
RMSE	0.410
MAE	0.341

MSE assesses model error by calculating the average squared difference between observed and predicted values. It offers an estimate of the model's ability to accurately forecast the target value (accuracy).

The quality of a model increases as its Root Mean Square Error (RMSE), Mean Squared Error (MSE), and Mean Absolute Error (MAE) decreases. The lower the value of these performance measures (close to 0) the better fit the model is. Here, we have found the MSE of 0.168 which implies that around 16% of the error in predicting the testing observation. RMSE of 0.41 implies that 41% of the error and MAE of 0.341 implies around 34% of the error in predicting the testing observations. All of these evidence do not support it to be a highly good-fitted model rather moderately fitted model.

5.3.2 Diagnostic of ARIMA model for WTI crude oil:

If we plot the graphs about the diagnostic of the ARIMA model for WTI crude oil, we get the following figure:



Top left: The residual errors seem to fluctuate around a mean of zero and have a uniform variance. But it has numerous and big distortions which might affect the result.

Top right: The density map displays a zero-mean normal distribution. It does follow the normal distribution.

Bottom left: Each dot should be precisely aligned with the red line. Any major variances would indicate a skewed distribution. This dataset is almost aligned with the red line. But there are some points at the beginning and at the end which are not aligned with the red line.

Bottom Right: A correlogram (also known as an Autocorrelation Function (ACF) plot or an autocorrelation plot) is a graphical representation of serial correlation in time series data. Serial correlation (also known as autocorrelation) occurs when an error at one point in time reproduces to a later point in time. For instance, you may overestimate the value of your stock market assets in the first quarter, resulting in an overestimation of values in subsequent quarters.

The Correlogram indicates that there is no autocorrelation between the residual errors. Any autocorrelation would suggest that the residual errors exhibit a pattern that is not explained by the model. Therefore, additional X_s (predictors) must be added to the model.

5.3.3 ARIMA model for Crack spread forecasting:

At the first stage, we performed the stepwise search to minimize AIC, similar to what we did for the case of WTI crude oil, to find the best order to execute SARIMAX model. Here the best order found by auto ARIMA is:

```
Best model:  ARIMA(4,1,5)(0,0,0)[0]
Total fit time: 100.375 seconds
```

SARIMAX Results

Dep. Variable:	y	**No. Observations:**	4392
Model:	SARIMAX(4, 1, 5)	**Log Likelihood**	4893.853
Date:	Tue, 25 Jul 2023	**AIC**	-9767.706
Time:	21:31:40	**BIC**	-9703.833
Sample:	0	**HQIC**	-9745.172
	- 4392		
Covariance Type:	opg		

	coef	**std err**	**z**	**P>|z|**	**[0.025**	**0.975]**
ar.L1	-0.2995	0.113	-2.660	0.008	-0.520	-0.079
ar.L2	0.0230	0.071	0.323	0.747	-0.116	0.162
ar.L3	0.7293	0.043	17.044	0.000	0.645	0.813
ar.L4	0.0980	0.056	1.763	0.078	-0.011	0.207
ma.L1	0.2089	0.111	1.877	0.060	-0.009	0.427
ma.L2	-0.1583	0.062	-2.564	0.010	-0.279	-0.037
ma.L3	-0.7593	0.030	-25.456	0.000	-0.818	-0.701
ma.L4	-0.1310	0.065	-2.030	0.042	-0.258	-0.005
ma.L5	0.1138	0.015	7.531	0.000	0.084	0.143
sigma2	0.0063	3.01e-05	208.777	0.000	0.006	0.006

Ljung-Box (L1) (Q):	0.05	**Jarque-Bera (JB):**	3860245.51
Prob(Q):	0.82	**Prob(JB):**	0.00
Heteroskedasticity (H):	0.30	**Skew:**	-2.45
Prob(H) (two-sided):	0.00	**Kurtosis:**	148.17

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).

From the above table, it is obvious that the best order for running SARIMAX is (4,1,5) for crack spread where the autoregressive order is 4, integration is 1 and moving average is 5.

As per stepwise regression result, among the four lag autoregressive coefficients first and third coefficients are statistically significant. Also, all moving average coefficients excluding the first lag, are statistically significant. Integration which is indicated by sigma is also significant.

However, if we run the best order for the training data set, the following result has been found:

SARIMAX Results

Dep. Variable:	lcrack	**No. Observations:**	2292
Model:	ARIMA(4, 1, 5)	**Log Likelihood**	2156.568
Date:	Tue, 25 Jul 2023	**AIC**	-4293.137
Time:	21:31:42	**BIC**	-4235.769
Sample:	0	**HQIC**	-4272.217
	- 2292		
Covariance Type:	opg		

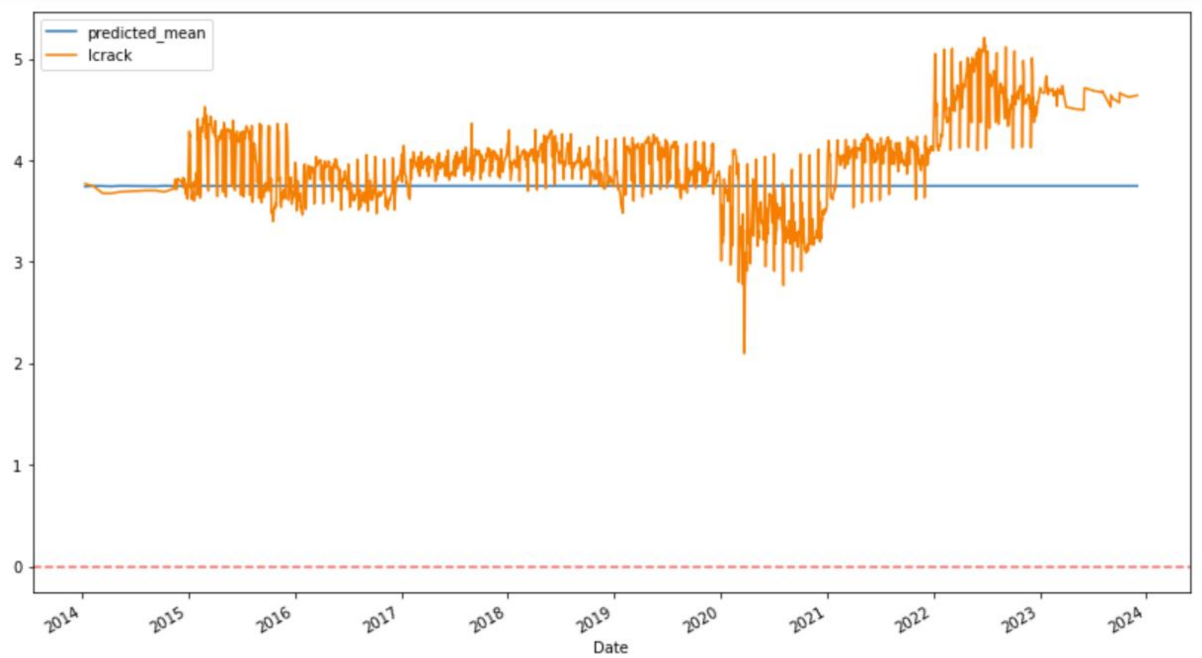
	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	-0.2245	0.284	-0.790	0.430	-0.782	0.333
ar.L2	0.0448	0.179	0.250	0.803	-0.306	0.396
ar.L3	0.6831	0.088	7.801	0.000	0.511	0.855
ar.L4	-0.0254	0.130	-0.196	0.845	-0.279	0.229
ma.L1	0.1281	0.283	0.453	0.650	-0.426	0.682
ma.L2	-0.1880	0.152	-1.240	0.215	-0.485	0.109
ma.L3	-0.6943	0.053	-13.185	0.000	-0.798	-0.591
ma.L4	-0.0149	0.154	-0.097	0.923	-0.317	0.287
ma.L5	0.0962	0.035	2.738	0.006	0.027	0.165
sigma2	0.0088	7.44e-05	117.956	0.000	0.009	0.009

Ljung-Box (L1) (Q):	0.01	**Jarque-Bera (JB):**	1524087.11
Prob(Q):	0.93	**Prob(JB):**	0.00
Heteroskedasticity (H):	0.37	**Skew:**	-1.89
Prob(H) (two-sided):	0.00	**Kurtosis:**	129.30

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).

In the above SARIMAX result table of the training data set, third lag auto-regressive coefficient, third lag moving average coefficient and exponential factor coefficients are statistically significant. Rest of the coefficients are not statistically significant. If we plot the predicted values into the graph, the following predicted mean price has been found for crack spread.

Graph-12: Predicted mean values vs actual Crack spread.



From the above graph it is certain that the actual crack spread did follow the predicted mean price at the beginning till up to 2020. But the actual price trend was lower than the predicted price till 2021. After that the actual price moved higher than the predicted price till the end of the time frame. But the actual price moved higher and higher specifically after the end of 2022. So, by the graphical representation the model seems a good-fitted model. We can state how exactly the model is good or bad it is by calculating the following performance measure indicators.

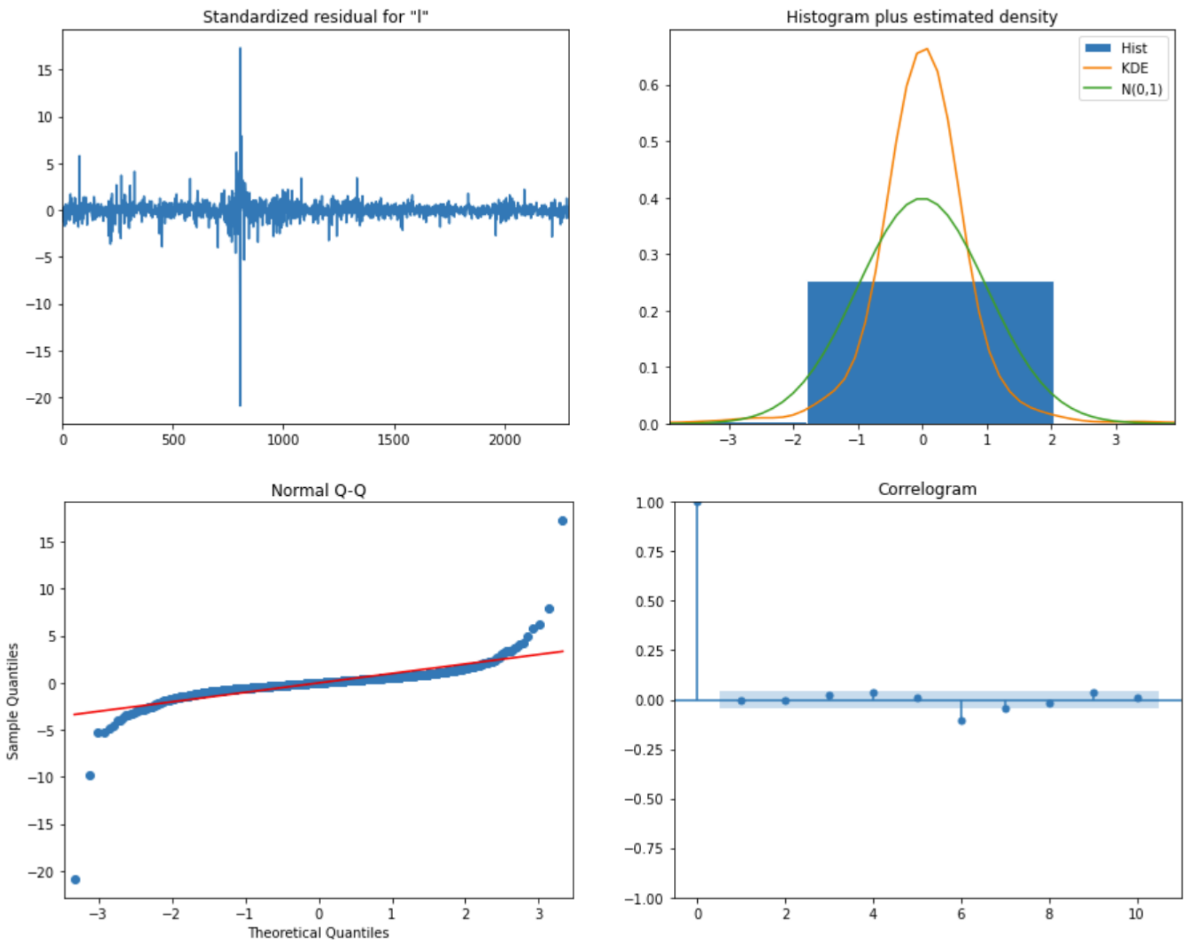
We can see how exactly the model good or bad it is by calculating the following performance measure indicators.

Testing if it's a good-fitted model	
Mean	3.99
MSE	0.263
RMSE	0.512
MAE	0.431

As described above, the lower the value of these performance measures (close to 0) the better fit the model is. Here, we have found the MSE of 0.263 which implies that around 26% of the error in predicting the testing observation. RMSE of 0.41 implies that 51% of the error and MAE of 0.431 implies around 43% of the error in predicting the testing observations. All of these evidence do not support it to be a highly good-fitted model rather moderately fitted model.

5.3.4 Diagnostic of ARIMA model for Crack spread:

If we plot the graphs about the diagnostic of the ARIMA model for Crack spread, we get the following figure:



Top left: The residual errors seem to fluctuate around a mean of zero and have a uniform variance. However, it has very few big distortions which might affect the result.

Top right: The density map displays a zero-mean normal distribution. But it doesn't appropriately follow the normal distribution.

Bottom left: Each dot should be precisely aligned with the red line. Any major variances would indicate a skewed distribution. This dataset is almost aligned with the red line. But there are some points at the beginning and at the ending point which are not aligned with the red line.

Bottom Right: A correlogram (also known as an Autocorrelation Function (ACF) plot or an autocorrelation plot) is a graphical representation of serial correlation in time series data. Serial correlation (also known as autocorrelation) occurs when an error at one point in time reproduces to a later point in time. For instance, you may overestimate the value of your stock market assets in the first quarter, resulting in an overestimation of values in subsequent quarters.

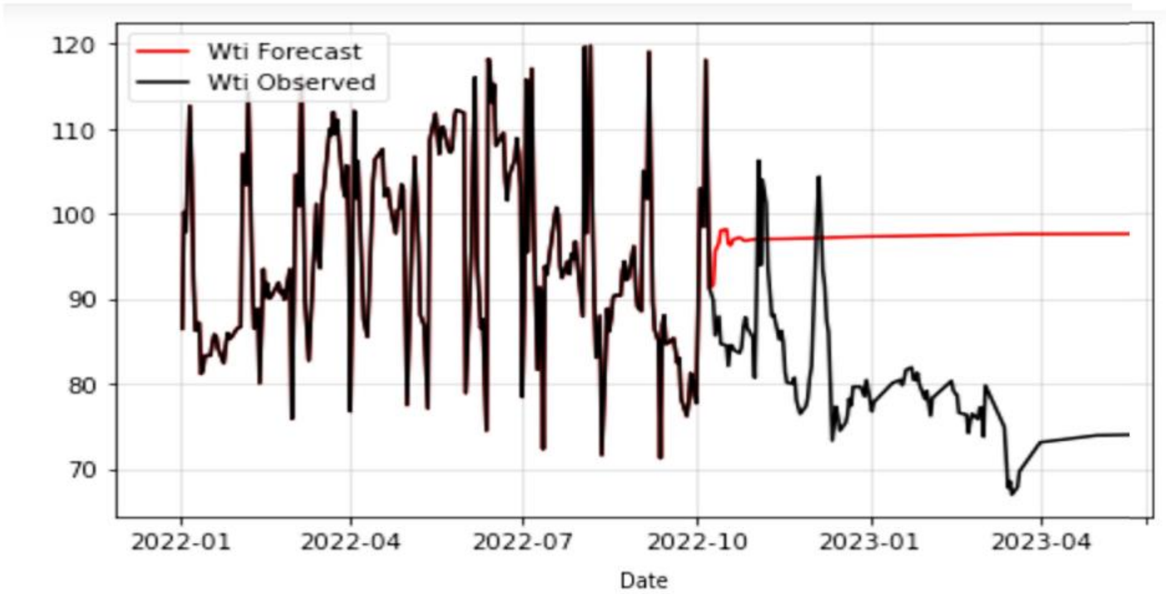
The Correlogram indicates that there is no autocorrelation between the residual errors. Any autocorrelation would suggest that the residual errors exhibit a pattern that is not explained by the model. Therefore, additional X_s (predictors) must be added to the model.

5.4 Forecasting with VAR model:

Vector Auto-regressive (VAR) model has been analysed and plotted based on five lag periods of WTI crude oil and Crack spread. The plotted diagram prediction was based on the regression analysis which were already explained in Granger Causality part.

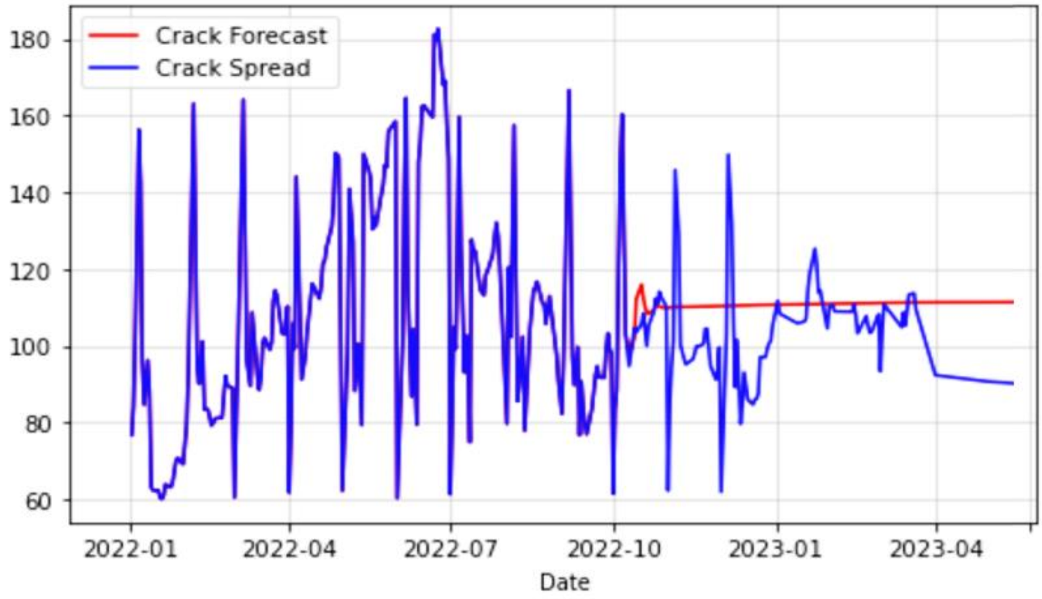
Graphical presentation of the VAR Model forecasting:

Graph-13: Predicted mean values vs actual values for WTI crude oil price.



In the above graph WTI observed and forecasted price have been displayed. The observed WTI price is lower than the forecasted price. The gap between these two prices are also wide.

Graph-14: Predicted mean values vs actual values for Crack spread.



In the above graph observed crack spread and forecasted crack spread price have been displayed. The observed crack price is almost aligned with the forecasted price for crack spread.

Root Mean Square Error and Mean Absolute Error:

Testing if it's a good-fitted model (VAR model)		
	WTI crude oil	**Crack Spread**
Mean price	72.44	53.63
RMSE	0.1805	0.1411

As, stated earlier, the quality of a model increases as its Root Mean Square Error (RMSE) decreases. The lower the value of these performance measures (close to 0) the better fit the model is. Here, we have found the RMSE of 0.1805 for WTI crude oil which implies that almost 18% of the error present in predicting the testing observation for crack spread. RMSE of 0.1411 which indicates that around 14% of the error present in the predicting the testing observation for crack spread.

Comparison between ARIMA and VAR model based on RMSE:

Root Mean Square Error (RMSE)

	ARIMA	**VAR**
WTI crude oil	0.41	0.18
Crack spread	0.51	0.14

From the above table it can be asserted that VAR method of forecasting has performed better according to the RMSE score. It is similar both for WTI crude oil and Crack spread. But if we compare between WTI crude oil and Crack spread, WTI crude oil has lower RMSE score in both methods of forecasting.

5.5 Forecasting models with lagged values:

The tests for Granger causality showed that lagged values do have predictive power. So, in this stage, we will supplement with forecasting models with lagged values for both WTI crude oil and crack spread, instead of basic ARIMA/SARIMAX models.

After performing all the steps of ARIMA/SARIMAX model including the lagged values for WTI crude oil, we get the following out-of-sample forecast measures.

Sample measures	Ordinary model	Including lagged values
RMSE	0.557	0.410
MAE	0.447	0.341

After including lagged values lowered both RMSE and MAE scores for WTI crude oil which indicates a better out of sample forecast than the ordinary model.

After performing all the steps of ARIMA/SARIMAX model including the lagged values for crack spread, we get the following out-of-sample forecast measures.

Sample measures	Ordinary model	Including lagged values
RMSE	0.512	0.578
MAE	0.431	0.435

However, after including lagged values lowered both RMSE and MAE scores have risen a little bit for crack spread which indicates that ordinary model had better out of sample forecast than the new model with including lagged values.

Nevertheless, VAR model has still better and lower score of sample measures even if lagged values have been included for both WTI crude oil and crack spread.

6 Factors Affecting Crack Spread Value:

Among underlying causes of widening crack spread such as soared demand for diesel, some strategic capping on reserves of US and its allies, declining refining capacity, Ukraine-Russia Invasion is one of the biggest triggers (Bloomberg).

According to the Chicago Mercantile Exchange (CME) analysis the following factors may affect crack spread and oil price [[5]].

1. **Geopolitical issues** such as Politics, geography, demography, economics and foreign policy affects crack spread initially because of higher crude oil prices relative to refined products. But after that crack gets strong as refineries respond to tighter crude oil supply and reduce product outputs.
2. **Winter seasonality** makes the crack strong because of higher demand. It becomes weaker in the summer season.
3. **Slower economic Growth** weakens the crack spread.
4. **Strong sustained product demand** boosts crack.
5. **Environmental regulation on tighter product specifications** tightens the product supply which boosts the crack in turn.
6. **Expiration of trading month;** crack values depend on the closing position of the trade.
7. **Tax increase after certain date:** crack gets weaker in front of the deadline and gets stronger after the deadline.
8. **Refinery maintenance** makes crack strong.
9. **Currency weakness:** The strength of crude oil is influenced by the currency's weakness, whereas the opposite is true for crack.
10. **If investors transfer capital into crude oil because of currency weakness, crude oil prices might climb rapidly, resulting in a decline in crack spreads.**
11. **The blending requirements of the U.S. Renewable Fuels Standard, which substitute refined hydrocarbon fuels with renewable products, influence crack spreads by bringing new supply sources to meet demand.**

[5] https://www.cmegroup.com/education/articles-and-reports/introduction-to-crack-spreads.html

Some other factors that might affect the crack spread which were analysed by EIA, such as:

- **Supply disruptions and rising demand:** In March 2021, imports of petroleum products, including gasoline, distillate, and other commodities, surged throughout the East Coast of the United States. Lower local supply, increased demand, and higher domestic petroleum product prices relative to European pricing led to an increase in imports [[6]].
- **Low transportation fuel demand:** During early March 2020, transportation fuel consumption has declined due to lower economic activity and stay-at-home restrictions intended to halt the spread of the 2019 new coronavirus sickness (COVID-19) [[7]]. Motor gasoline, Distillate fuel oil and jet fuel all went down due to the pandemic time. Since April 3, however, distillate product supplies have dropped to their lowest level in twenty-one years, 2,8 million barrels per day for the week of April 10.

Crack Spread as a Market Signal

Crack spreads are used by speculators and investors as a market indicator of crude oil and refined product price fluctuations. It gives a real-time indicator of the performance of the products and the profitability of the refinery as a whole. When the spread widens, it indicates that demand and prices for refined products are on the rise. Investors see this as an indication that crude oil prices will rise to match the demand for refined primary products such as gasoline and heating oil. If the spread narrows, refiners will reduce their output of refined goods in order to recover their profit margins (CFI).

[6] https://www.eia.gov/todayinenergy/detail.php?id=48316
[7] https://www.eia.gov/todayinenergy/detail.php?id=43595

7 Conclusions:

To conclude, it can be stated that for all participants of oil market for example refiners, speculators, and hedgers, crack spread can play a significant role in decision making. As, crack spread is highly correlated with oil prices, whether brent or WTI, it can influence any party to make/change decisions. So, speculators can make money by speculating it and understanding the movement of oil price through crack spread.

However, to find the answer if crack spread can cause in price fluctuations of WTI crude oil, we did the Granger causality test whether crack spread Granger causes WTI crude oil price movements. As per test result, crack spread granger causes WTI crude oil. In addition, the same test was also run to investigate if WTI crude oil granger causes crack spread. It also showed the same result. It has also observed that Crack spread can predict WTI crude oil price by running regression analysis on WTI crude oil, crack spread, heating and gasoline prices. The parameter of the crack spread coefficient revealed that there is inverse relationship between WTI crude oil and crack spread, where it was also statistically significant. Regression coefficient result indicated that a one percent increase in crack spread price will result about 17$ decrease in WTI crude oil price. In contrast, there was a positive relationship between WTI crude oil prices and gasoline & heating prices.

Moreover, both VAR and ARIMA model showed a fair way of forecasting both for crack spread and brent oil prices. To do the analysis of both models, data has been transformed to stationary according to the assumption. Even though the performance measure of both models does not support to be a very good-fitted model, rather a moderate fitted model. Both model showed an average predicted price. According to the performance measure calculation for both of models (RMSE), VAR model appeared to be a better forecasting model for the out of sample observations. It also hold true even if we added lag values (those had predicted power) for WTI crude oil and crack spread.

In addition, there are several factors which can affect the movement of both crack spread and oil prices. Some factors are related to geopolitical issues, some are related to domestic issues or conditions, and some are related with international oil market. Those factors can create distortions in the oil market which are sometimes difficult to forecast by using conventional techniques. As stated in EIA, supply disruptions and rising demand were responsible for the recent high crack spread.

This whole study was carried on based on the futures contracts from the US EIA database. A challenge with the WTI oil price is that the WTI was landlocked and running out of storage in the spring of 2022 leading to a brief period with negative/very low prices. Moreover, some factors of affecting crack spread have been presented in this study. However, a detailed study with reliable data for each factor can be carried forward to analyse the exact reason for such a sudden huge spike of crack spread.

8 References

Ahmed, R. A., & Shabri, A. B. (2014). Daily crude oil price forecasting model using arima, generalized autoregressive conditional heteroscedastic and support vector machines. American Journal of Applied Sciences, 11(3), 425.

Alimoradi, M., & Mohajeri, S. (2016). Evaluation of Effectiveness of crack spread futures in crude oil price forecast. Journal of Investment Knowledge, 5(17), 187-205.

ARIMA model in python, https://www.javatpoint.com/arima-model-in-python

ARIMA model in python, https://www.machinelearningplus.com/time-series/arima-model-time-series-forecasting-python/

Asche, F., Gjolberg, O., & Völker, T. (2003). Price relationships in the petroleum market: an analysis of crude oil and refined product prices. Energy Economics, 25 (3), 289-301.

Bashiri Behmiri, N., & Pires Manso, J. R. (2013). Crude oil price forecasting techniques: a comprehensive review of literature. Available at SSRN 2275428.

Bekiros, S., Gupta, R., & Paccagnini, A. (2015). Oil price forecastability and economic uncertainty. Economics Letters, 132, 125-128.

Blas, J. (2022). Sorry, But for You, Oil Trades at $250 aBarrel. *Bloomberg*, 2-3.
Box, G.E.P. and G. Jenkins, 1976. Time Series Analysis: Forecasting and Control. 1st Edn., Holden-Day, San Francisco, ISBN-10: 0816211043, pp: 575.

Brent Crude vs WTI crude: https://www.investopedia.com/ask/answers/052615/what-difference-between-brent-crude-and-west-texas-intermediate.asp#:~:text=Brent%20Crude%20is%20the%20benchmark,sourced%20from%20U.S.%20oil%20fields.

Coppola, A., 2008. Forecasting oil price movements: exploiting the information in the futures market. Journal of Futures Market 28, 34–56.

Correlogram; https://www.statisticshowto.com/correlogram/

Data source: Europe Brent Spot Price FOB; https://www.eia.gov/dnav/pet/hist/RBRTED.htm

Data source: WTI crude oil spot price FOB; https://www.eia.gov/dnav/pet/hist/RWTCD.htm

Crack spread as a Market signal: https://corporatefinanceinstitute.com/resources/commodities/crack-spread/

Diagnostic test result interpretation; https://www.alldatascience.com/time-series/forecasting-time-series-with-auto-arima/

Dickey, D. A., & Fuller, W. A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. Econometrica: journal of the Econometric Society, 1057-1072.

Dicky Fuller test in python: https://www.machinelearningplus.com/time-series/augmented-dickey-fuller-test/

Diks, C., & Panchenko, V. (2006). A new statistic and practical guidelines for nonparametric Granger causality testing. Journal of Economic Dynamics and Control, 30(9-10), 1647-1669.

Eia.gov. (2013). *Crack Spread*.https://www.eia.gov/todayinenergy/includes/Crack-Spread_Explain.php

Factors affecting crack spread: https://www.cmegroup.com/education/articles-and-reports/introduction-to-crack-spreads.html

Fink, J., & Fink, K. (2014). Do seasonal tropical storm forecasts affect crack spread prices?. Journal of Futures Markets, 34(5), 420-433.

Fousekis, P., & Grigoriadis, V. (2017). Price co-movement and the crack spread in the US futures markets. Journal of Commodity Markets, 7, 57-71.

Gahirwal, M. (2013) Inter Time Series Sales Forecasting. arXiv:1303.0117 [Citation Time(s):1]

Gay, G. D., Hunter, W. C., & Kolb, R. W. (1986). A comparative analysis of futures contract margins. The Journal of Futures Markets (1986-1998), 6(2), 307.

Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. Econometrica: journal of the Econometric Society, 424-438.

Hall, A. (1994). Testing for a unit root in time series with pretest data-based model selection. Journal of Business & Economic Statistics, 12(4), 461-470.

Heating oil futures data: https://www.investing.com/commodities/heating-oil-historical-data

Heating oil spot price: https://www.marketwatch.com/investing/future/ho00/download-data?startDate=10/3/2022&endDate=01/06/2023

Hoover, K. D. (2006). Causality in economics and econometrics.

Hyerczyk, J. (2011). The Data Game. Futures magazine, Aug.

Li, S., Zhang, H., & Yuan, D. (2019). Investor attention and crude oil prices: Evidence from nonlinear Granger causality tests. Energy Economics, 84, 104494.

Lopez, J. H. (1997). The power of the ADF test. Economics Letters, 57(1), 5-10.

Lütkepohl, H. (2009). Econometric analysis with vector autoregressive models. Handbook of computational econometrics, 281-319.

Low transportation fuel demand: https://www.eia.gov/todayinenergy/detail.php?id=43595

Masteika, S., Rutkauskas, A. V., & Alexander, J. A. (2012, February). Continuous futures data series for back testing and technical analysis. In Conference proceedings, 3rd international conference on financial theory and engineering (Vol. 29, pp. 265-269). IACSIT Press.

Mirmirani, S., & Li, H. C. (2004). A comparison of VAR and neural networks with genetic algorithm in forecasting price of oil. In Applications of artificial intelligence in finance and economics (Vol. 19, pp. 203-223). Emerald Group Publishing Limited.(Tularam & Saeed, 2016)

Morana, C. (2001). A semiparametric approach to short-term oil price forecasting. Energy economics, 23(3), 325-338.

Mostafa, M. M., & El-Masry, A. A. (2016). Oil price forecasting using gene expression programming and artificial neural networks. Economic Modelling, 54, 40-53.

Mushtaq, R. (2011). Augmented dickey fuller test.

Nochai, R., & Nochai, T. (2006, June). ARIMA model for forecasting oil palm price. In Proceedings of the 2nd IMT-GT Regional Conference on Mathematics, Statistics and applications (pp. 13-15).

Nyoni, T. (2018). Modeling and forecasting inflation in Kenya: Recent insights from ARIMA and GARCH analysis. Dimorian Review, 5(6), 16-40.

Obadi, S. M., & Korecek, M. (2018). The crude oil price and speculations: Investigation using granger causality test. International Journal of Energy Economics and Policy, 8(3), 275.

Oil tanking: https://www.oiltanking.com/en/news-info/glossary/heating-oil-extra-light-heating-oil-heavy-fuel-oil-bio-heating-oil.html

Pelletier, B. (2011). Computed contracts: their meaning, purpose and application. An Essay on Computed Contracts. // http://www.csidata.com/cgi-bin/getManualPage.pl?URL=essay.htm (referred on 17/09/2011).

RBOB Gasoline futures data: https://www.investing.com/commodities/gasoline-rbob

RBOB Gasoline spot price: https://www.marketwatch.com/investing/future/ho.1/download-data?startDate=1/3/2005&endDate=01/06/2023

Seth, A. (2007). Granger causality. Scholarpedia, 2(7), 1667.

Silvapulle, P., & Moosa, I. A. (1999). The relationship between spot and futures prices: evidence from the crude oil market. Journal of Futures Markets: Futures, Options, and Other Derivative Products, 19(2), 175-193.

Statistical Stationarity: https://people.duke.edu/~rnau/411diff.html

Supply disruptions and rising demand: https://www.eia.gov/todayinenergy/detail.php?id=48316

Tukaram, G. A., & Saeed, T. (2016). Oil-price forecasting based on various univariate time-series models. American journal of operations research, 6(03), 226.

U.S refiners doing well even if gasoline prices are down (2022), https://www.marketplace.org/2022/08/02/u-s-oil-refiners-are-doing-well-even-though-gasoline-prices-are-down/

Xie, W., Yu, L., Xu, S., & Wang, S. (2006, May). A new method for crude oil price forecasting based on support vector machines. In International conference on computational science (pp. 444-451). Springer, Berlin, Heidelberg.

Zhang, G.P. (2003) Time Series Forecasting Using a Hybrid ARIMA and Neural Network Model. Neurocomputing, 50, 159-175. http://dx.doi.org/10.1016/S0925-2312(01)00702-0 [Citation Time(s):1]

Zivot, E., & Wang, J. (2006). Vector autorsegressive models for multivariate time series. Modeling financial time series with S-PLUS®, 385-429.

Norges miljø- og biovitenskapelige universitet
Noregs miljø- og biovitskapelege universitet
Norwegian University of Life Sciences

Postboks 5003
NO-1432 Ås
Norway