IJRAR.ORG E-ISSN: 2348-1269, P-ISSN: 2349-5138
INTERNATIONAL JOURNAL OF RESEARCH AND ANALYTICAL REVIEWS (IJRAR) | IJRAR.ORG
An International Open Access, Peer-reviewed, Refereed Journal

A study on Analyzing and Forecasting Crude oil prices and BSE Oil & Gas Index using Forecasting Models

Dr. Amish Soni
Asst. Professor,
Faculty of Management,
GLS University,
Ahmedabad

Miss. Sakshi Mehulbhai Shah
Student, GLS University

Abstract

The purpose of this study is to study and forecast the crude oil prices in India using Autoregressive Integrated Moving Average (ARIMA) model of time series analysis. The report tables, charts, and ARIMA model are used to forecast the prices of crude oil in India. Also, in this study we have forecasted crude oil prices and BSE oil & Gas index accuracy of model is found to be 90% The forecast will help investors to take early bird decisions & help to make necessary strategies & planning of business.

Keywords: ARIMA, Crude Oil, Forecasting, BSE Oil & Gas Index

1. Introduction:

Crude oil is gaining its important as a lifeline to the world's economy in general and to some of the nation's economy in specific.

India imports about 70% of its total oil consumption and it makes exported petroleum products worth $86.21 billion during the first 11 months of the current financial year ending March 2023, which accounts for more than 21 per cent of India's total commodity exports.

This accounts for one third of its total imports. The fall in global oil prices may be beneficial to India, but it has its downsides It affects the exporters of petroleum producers in the country.

A fall in oil price may impact their economy, and hamper demand for Indian products This would indirectly affect India and its companies.

2. Literature Review:

Jessin Shah P, Dr. G Kiruthiga (2020) Crude oil is one of the most important chemical and energy resources. The crude oil and its price affect the economic and social activities, therefore crude effective crude oil forecast can help stabilize economic development and prevent energy crisis. The WTI crude oil data from the period of 1987 to 2020 were analyzed and showed the non-stationarity of data. We conducted data analysis and preprocessing prepared for ARIMA. To value the commodity, three crude oil benchmarks are set; they are Brent, West Texas Intermediate (WTI) and Dubai/Oman oil benchmarks. WTI crude oil price data refers to crude oil extracted in the U.S. is used in this study.

Talha, M., Sohail, M., Tariq, R., & Ahmad, M. T. (2021) Oil prices, energy consumption, and economic growth have made a significant effect on inflation rates in Malaysia. This study attempts to examine this effect through secondary data collected on specific macroeconomic indicators such as inflation and country's oil prices, energy consumption, and gross domestic product. For this purpose, year to year information from 1986 to 2019 series was utilized. This research study used the E- views regression model, correlation model, and descriptive analysis to break down the information about oil prices, energy consumption, economic growth, and inflation rate. It investigated how oil prices, energy consumption, and economic growth have a positive association with the inflation rate in Malaysia. It is premised in this study that the rate of oil and renewable energy consumption enhances the economic growth and also improves the inflation rate in the country. This finding helps the Malaysian government in entering into essential leadership deal with oil prices, and energy consumption to manage inflation rates.

Yadav, N., Tandon, P., Tripathi, R., & Shastri, R. K. (2020). The purpose of the study is to investigate the long-run and short-run dynamic relationship between crude oil prices and the movement of Sensex for the period of 2000–2018.The study uses the augmented Dickey– Fuller test for the presence of unit root, Johansen cointegration test for estimating the cointegration among the variables. Further, in the case of no cointegration found, the study employed the vector autoregression (VAR) model to estimate the long-run relationship and the Granger causality/Wald test for short-run relationship. The study also conducted tests for the prerequisites of the model: serial correlation, heteroskedasticity and normality of data.

Venkataraman, P., & Subramaniam, K. (2022). After the United States, China, and Japan, India was the world's fourth biggest consumer of oil and petroleum products. The nation is significantly reliant on crude oil imports, the majority of which come from the Middle East. The Indian oil and gas business is one of the country's six main sectors, with important forward links to the rest of the economy. More than two-thirds of the country's overall primary energy demands are met by the oil and gas industry. The industry has played a key role in placing India on the global map. India is now the world's sixth biggest crude oil user and ninth largest crude oil importer. In addition, the country's portion of the worldwide refining market is growing. India's refining industry is now the world's sixth biggest. With plans for Reliance Petroleum Limited to commission another refinery with a capacity of 29 MTPA next 16 to its 33 MTPA refinery in Jamnagar, Gujarat, this position is projected to be enhanced. As a consequence, the Reliance refinery would be the biggest single-site refinery in the world. Based on secondary data gathered from CMIE, the current research examines the ratios influencing the profitability of selected oil exploration and production businesses in India during a 10-year period.

Elsaraiti, M., Ali, G., Musbah, H., Merabet, A., & Little, T. (2021, April). Power consumption is a very important factor in smart grids for load management process. Forecasting energy consumption is the first step in dealing with load management. For forecasting time series, the ARIMA models are one of the widely used models which showing encouraging results. In this study, ARIMA models were proposed to predict future electricity consumption. The ACF and PACF plots were used as well as stationarity of the data to identify (p, d, q) values. The results showed the accuracy and efficiency of the models and their ability to compete with current techniques for forecasting electricity consumption based on the use of the Mean Absolute Percentage Error (MAPE) to measure the accuracy of the prediction, as the model was able to predict with an error of 4.332%

Lin, B., & Bai, R. (2021) The crude oil markets have seen unpredictability over petroleum markets' good and bad times worldwide in the past few years. Various facets of the natural energy industry discuss in this research paper. The effect of the Indian ecosystem in determining oil demand and supply is vital to understand price fluctuations. Researchers in this research paper concentrates on the ARIMA model and other regression models used in the post-1991 LPG reforms to determine crude petroleum values and their primary effect on the Indian ecosystem (GDP) through time-series data from 1991 to 2019.The ARIMA model is further evidence of the validation of datasets and the potential trend in the show of global oil rates.

Dritsaki, C. (2021). Oil is considered one of the most widely used commodity worldwide and one of the most important goods for a country's productivity. Even if the effect of renewable energy sources tries to replace the consumption of fossil fuels, such as oil, nonetheless the level of worldwide oil consumption hasn't changed. Forecasting oil consumption plays an important role on the designing of energy strategies for policy makers. This paper aims at modeling and forecasting oil consumption in Greece using Box-Jenkins methodology during 960-2020. Forecasting oil consumption was accomplished both with static and dynamic procedure, in and out-of-sample using various forecasting criteria. The results of our paper present a downturn in oil consumption for the following years due to two basic factors. The first is referred to Covid-19 pandemia where economic activity of the country decreased as well as business revenues. The second is the efforts made by the country to replace, oil consumption with other energy forms such as natural gas and mostly renewable sources like sun and wind. With these actions taken, the country – member of EU is consistent with the regulations signed to Kyoto protocol where there are commitments for CO2 reduction emissions and improvement of energy use.

Kumar, S., Choudhary, S., Singh, G., & Singhal, S. (2021) study investigates the nexus among natural gas price, crude oil price, gold price, exchange rate, and stock market index in Indian context using the Nonlinear Autoregressive Distributed Lag (NARDL) model on weekly data for the time period of January 1997 to June 2019. The result of the study provides empirical evidence about the presence of asymmetries in the short and long-run among these asset classes. The findings of the study confirm that gold, stock market and natural gas has an asymmetric effect on crude oil in the long-run and crude oil asymmetrically influence natural gas in the short-run. Exchange rate is observed to have no impact on crude oil and natural gas price and results indicate gold as statistically significant variable for both natural gas and crude oil in the short-run and long-run. This is the first study to delineate the dynamic simultaneous interaction among these asset markets and its findings can be extremely useful for investors, academicians, and policymakers to make financial decisions.

Sharma, A. (2018). This paper estimates the linear interdependencies between international crude oil prices and stock market indices of India using weekly data spanning from January 2010 to January 2017 in a vector autoregressive (VAR) framework. The time series used for the analysis are crude oil futures prices, nifty index, and BSE energy index. Augmented Dickey-Fuller and Philips-Perron unit root tests reveal that all the time series are non-stationary at level and stationary at first difference. Cointegration test reveals the absence of cointegrating factor i.e., absence of long run relationship. VAR model captures all the time series as endogenous variables and independent variables are studied at two lags. Result shows that the Energy Index is very well explained by the lagged values of crude oil futures prices, nifty index, and BSE energy Index. impulse response function reveals that crude oil prices are affected negatively when one standard deviation shocks are given to stock indices. Crude oil is gaining its important as a lifeline to the world’s economy in general and to some of the nation’s economy in specific. Crude oil is the most actively traded commodity in the world.

Soundarapandiyan, K., & Ganesh, M. (2017). This essay examines the effects of crude oil on the consumer price index and the GDP. Crude oil is essential in the volatile market to control inflation and maintain strong economic growth. India is the world's fourth-largest user of crude oil, importing 100 million tonnes annually, or 37% of the total.

3. Research Methodology:

3.1Research problem statement

“A study on analyzing and Forecasting Crude oil prices in India using ARIMA model”

3.2 Research Objectives

- To forecast the crude oil prices using time series analysis
- To study the impact of crude oil prices on BSE index.

3.3 Scope of study

The fundamental purpose of this study is to examine the changes in prices of crude oil and also how inter – related both BSE Index & crude oil prices are there.

3.4 Need of Study

Crude oil is one of the World’s most important tradable commodities and its prices have ripple effects through the broader economy. Rising oil prices mean higher gasoline prices at pump, higher gasoline prices at the pump, higher shipping costs and increased impute costs for producers and crude oil price fluctuations have a far-reaching impact on global economies and thus price forecasting can assist in minimizing the risks associated with volatility in oil prices.

3.5 Research Design

The report used descriptive research design because here the report includes the study about forecasting the crude oil prices in India.

3.6 Sample size

The report used data of crude oil prices of 5 years starting from June 2018 to May 2023

3.7 Data collection

Here the report uses secondary sources of data.

3.8 Hypothesis

H0: There is no increase in price of crude oil on basis of estimate value.

H1: There is increase in price of crude oil on basis of estimate value.

H0: There is no autocorrelations in variables

H1: There is auto correlations in variables

H0: There is no dependency of crude oil prices on BSE Oil & Gas Index

H1: There is a dependency of crude oil prices on BSE Oil & Gas Index

H0: There is no fit to forecast the BSE Oil & Gas Index value from crude oil prices

H1: There is a fit to forecast the BSE Oil & Gas Index value from crude oil prices.

3.9 Data analysis tools and techniques

In the report tables, charts and ARIMA model of time series analysis is used to study and forecast the crude oil prices in India.

3.10 Limitations

• Accuracy of available data

• Knowledge constraint

• Sometimes there is error containing in analyzing and forecasting prices through time series analysis

4.Analysis & Forecasting

4.1 Introduction to ARIMA Model

The main objective of ARIMA model is for forecasting (predicting future values of the time series). The model is generally referred to as ARIMA (p,d,q), where p,d, & q are nonnegative numerical values.

Auto Regressive Integrated Moving Average (ARIMA) model, an autoregressive integrated moving average, or ARIMA, is a statistical analysis model that uses time series data to either better understand the data set or to predict future trends. A statistical model is autoregressive if it predicts future values based on past values. For example, an ARIMA model might seek to predict a stock's future prices based on its past performance or forecast a company's earnings based on past periods. An autoregressive integrated moving average model is a form of regression analysis that gauges the strength of one dependent variable relative to other changing variables. The model's goal is to predict future securities or financial market moves by examining the differences between values in the series instead of through actual values.

Auto regression (AR) refers to a model that shows a changing variable that regresses on its own lagged or prior, values. Integrated (I) represents the differencing of raw observations to allow for the time series to become stationary (i.e., data values are replaced by the difference between the data values and the previous values). Moving average (MA) incorporates the dependency between an observation and a residual error from a moving average model applied to lagged observations.

Autoregressive integrated moving average (ARIMA) models predict future values based on past values year, could negatively affect the regression model. If a trend appears and stationarity is not evident, many of the computations throughout the process cannot be made with great efficacy.

The ARIMA model predicts a given time series based on its own past values. It can be used for any non-seasonal series of numbers that exhibits patterns and is not a series of random events. For example, sales data from a clothing store would be a time series because it was collected over a period of time. One of the key characteristics is the data is collected over a 47 series of constant, regular intervals. A modified version can be created to model predictions over multiple seasons.

Steps to be followed during ARIMA Model:

1. Identification
2. Estimation
3. Diagnostic Testing
4. Forecasting
5. Identification

The first stage in the ARIMA Model is to determine if the data is stationary or non-stationary.

- Plot/preview
- Correlogram to check stationarity of Time Series (gradual decline- Nonstationary)-ACF &PACF
- If non- stationarity, take d/n or simple First Difference of the series, recheck ACF & PACF- Count spikes out of CI (dotted lines)
- Determine order of ARIMA (p,d,q)- ACF for Q-terms and PACF for p-terms and d= First Difference

4.1.1 Sequence Plot

Table 1: Model Description

Model Name		MOD_1
Series or Sequence	1	Price
Transformation		None
Non-Seasonal Differencing		0
Seasonal Differencing		0
Length of Seasonal Period		No periodicity
Horizontal Axis Labels		Date
Intervention Onsets		None
Reference Lines		None
Area Below the Curve		Not filled

Applying the model specifications from MOD_1

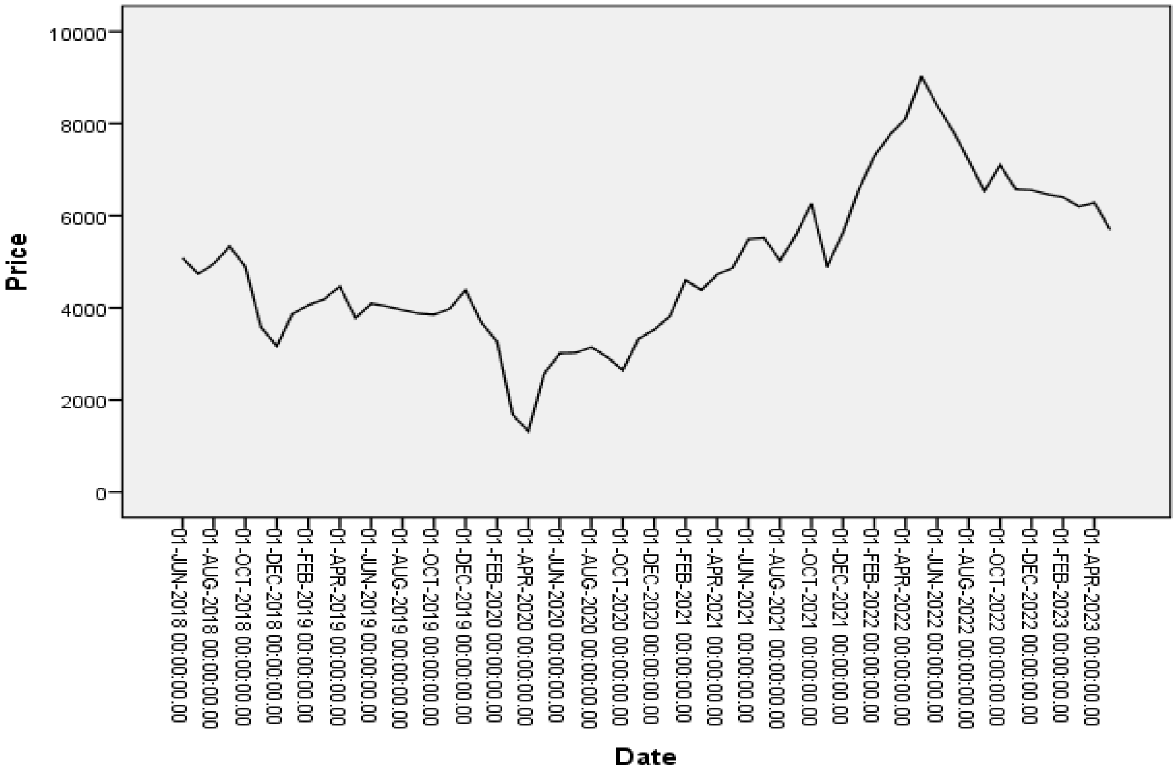
4.1.2. Case Processing Summary

Table 2

	Price
Series or Sequence Length	60
Number of Missing Values in the Plot:	
User-Missing	0
System-Missing	0

4.1.3 Sequence Plot - Identification

Chart -1



Interpretation: Model description explains the summary of ARIMA model and overall details of the model & chart where it discloses model name, which series is used in model on which ARIMA model is focussed similarly case processing sequence discloses the series or sequence length of data i.e in our case we have taken monthly wise closing price of crude oil for five years period (12*5).

The first stage in the ARIMA Model is to determine if the data is stationary or non-stationary. We may use this sequence plot chart to determine the trend and variations, since there are more upward trends and fluctuating crude oil prices that are non-seasonal and non-stationary, which is why we can use the ARIMA Model here. The ARIMA model is used to anticipate the future of non-stationary data. As a result of this chart, we may determine that the data is non-stationary and can apply the ARIMA Model.

Identification - Correlogram

ACF

Table 3: Model Description

Model Name		MOD_6
Series Name	1	Price
Transformation		None
Non-Seasonal Differencing		0
Seasonal Differencing		0
Length of Seasonal Period		12
Maximum Number of Lags		16
Process Assumed for Calculating the Standard Errors of the Autocorrelations		Independence(white noise)[a]

Display and Plot	All lags

Applying the model specifications from MOD_6

a. Not applicable for calculating the standard errors of the partial autocorrelations.

Table 4: Case Processing Summary

		Price
Series Length		60
Number of Missing Values	User-Missing	0
	System-Missing	0
Number of Valid Values		60
Number of Computable First Lags		59

Price

Table -5

Autocorrelations

Series: Price

Lag	Autocorrelation	Std. Error[a]	Box-Ljung Statistic		
			Value	df	Sig.[b]
1	.940	.126	55.763	1	.000
2	.865	.125	103.787	2	.000
3	.804	.124	145.944	3	.000
4	.742	.123	182.488	4	.000
5	.691	.122	214.747	5	.000
6	.639	.120	242.885	6	.000
7	.587	.119	267.108	7	.000
8	.528	.118	287.054	8	.000
9	.460	.117	302.517	9	.000
10	.396	.116	314.175	10	.000
11	.320	.115	321.961	11	.000
12	.245	.114	326.595	12	.000
13	.177	.112	329.063	13	.000
14	.112	.111	330.072	14	.000
15	.052	.110	330.298	15	.000
16	-.005	.109	330.299	16	.000

a. The underlying process assumed is independence (white noise).

b. Based on the asymptotic chi-square approximation.

Interpretation: The auto correlation function (ACF) is a statistical technique that we can use to identify how correlated the values in a time series are with each other. The ACF plots the correlation coefficient against lag.

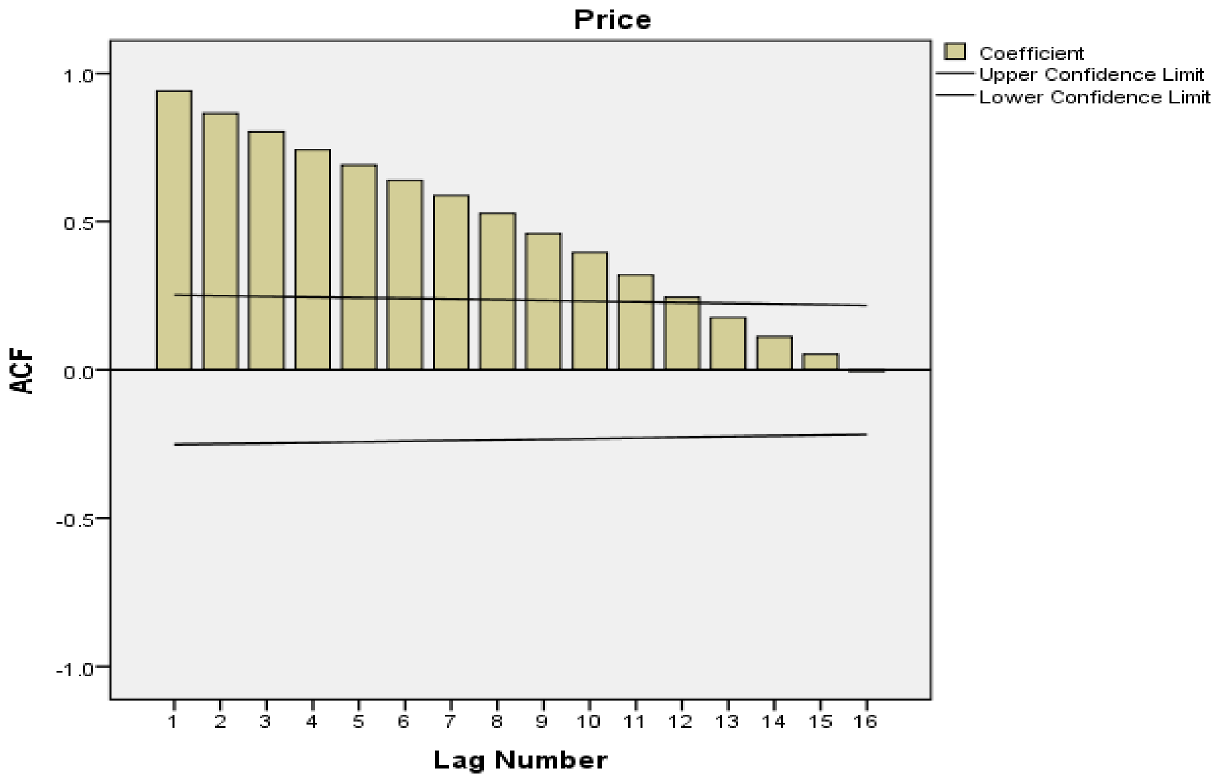
H0: There is no auto correlation

H1: There is auto correlation

If significance is less than 0.05 than will reject null hypothesis and accept the alternative hypothesis. Hence here significance level is less than 0.05 so will reject null hypothesis. So there is auto correlation.

4.1.4 Identification Graph- ACF

Chart-2 ACF



Interpretation: Here we can observe all lags in given time series analysis is beyond the UCl that indicates data is non – Stationary. Hence to make it Stationary Arima Model Is required.

Table 6: Partial Autocorrelations

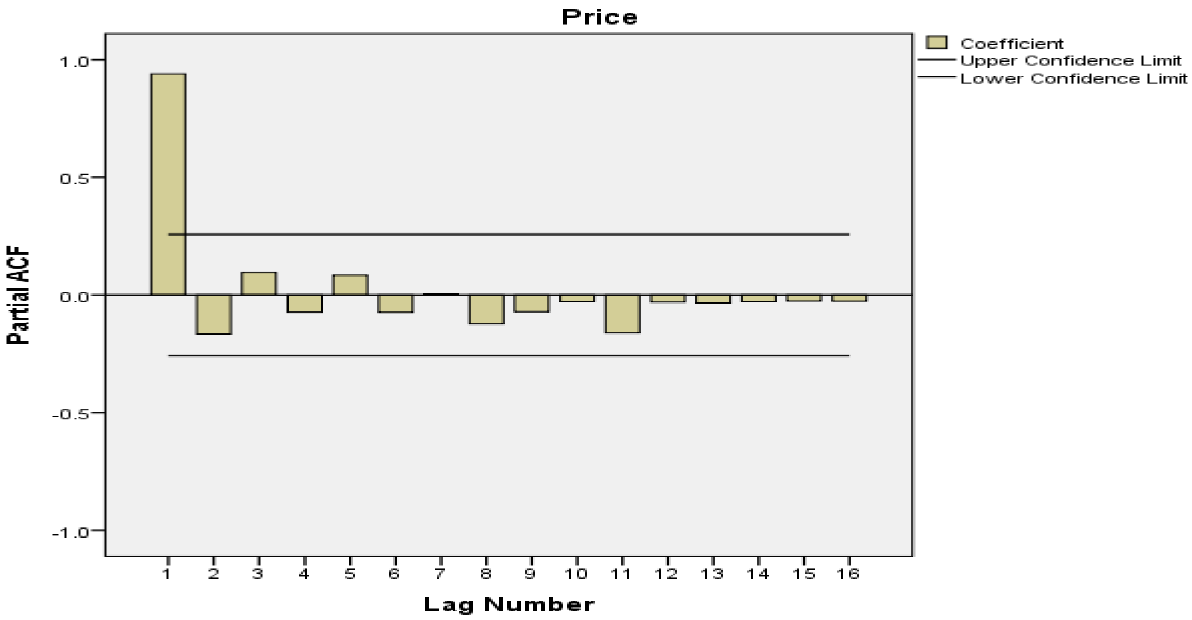
Series: Price

Lag	Partial Autocorrelation	Std. Error
1	.940	.129
2	-.165	.129
3	.097	.129
4	-.073	.129
5	.085	.129
6	-.074	.129
7	.003	.129
8	-.122	.129
9	-.071	.129
10	-.029	.129
11	-.161	.129
12	-.030	.129
13	-.034	.129

14	-.028	.129
15	-.026	.129
16	-.027	.129

4.1.5 Identification Graph -PACF

Chart -3 PCAF



Interpretation: In correlogram there are two graphs ACF&PACF, ACF known as auto correlation, it is graph which plots co- relation of current value and any period lag value against the lag k.

ACF- It determines current values relation with it past values

PACF- Relationship of two conductive values holding other variable constant, keeping aside the other values.

Hence it basically indicates relationship current values & lag values

4.1.6 Correlogram and its indictors:

- ACF is autocorrelation function. This graph plots cor (y t Yt-k) against k.
- Series is stationary if spikes decay exponentially.
- PACF graph plots estimated co-efficient of ytk form OLS estimated of AR(k) process against k.
- Series is stationary if spikes decay exponentially.
- PACF is high and significant up to p lags and zero for lags beyond p.
- P represents order of autoregressive while q indicates order of moving average,
- The solid lines in PACF & ACF indicates confidence level which is 95%. Whole data is analyzed between this confidence level upper & lower bounds.
- Here this correlogram are original form without any difference so if we analysis PACF chart than out of 16 lags spike of 1st lag is out of confidence level.
- Here in ACF graph all spikes are outside the confidence level and hence this indicates surety to apply ARIMA test as almost all spikes are out of confidence interval.
- If prices are stable (within confidence interval) than forecasting is done by any financial techniques. But in our case prices are not stable as all spikes are out of confidence level hence best suited

technique is ARIMA model. The gradual decline of ACF correlogram also indicates its non-stationary

- Here to bring prices in stable form will analysis Autocorrelations by taking difference 1

4.1.7 ACF & PACF when difference is 1 =

Table -7 Delta yt= yt- yt-1

Model Description

Model Name		MOD_7
Series Name	1	Price
Transformation		None
Non-Seasonal Differencing		1
Seasonal Differencing		0
Length of Seasonal Period		12
Maximum Number of Lags		16
Process Assumed for Calculating the Standard Errors of the Autocorrelations		Independence(white noise)[a]
Display and Plot		All lags

Applying the model specifications from MOD_7

a. Not applicable for calculating the standard errors of the partial autocorrelations.

Table -8

Case Processing Summary

		Price
Series Length		60
Number of Missing Values	User-Missing	0
	System-Missing	0
Number of Valid Values		60
Number of Values Lost Due to Differencing		1
Number of Computable First Lags After Differencing		58

Table -9

Autocorrelations

Series: Price

Lag	Autocorrelation	Std. Error[a]	Box-Ljung Statistic		
			Value	df	Sig.[b]
1	.113	.127	.795	1	.373
2	-.116	.126	1.647	2	.439
3	-.009	.125	1.652	3	.648
4	-.094	.124	2.231	4	.693
5	.010	.122	2.237	5	.815

6	.002	.121	2.237	6	.897
7	.039	.120	2.341	7	.939
8	.090	.119	2.915	8	.940
9	-.055	.118	3.133	9	.959
10	.065	.117	3.446	10	.969
11	-.017	.115	3.466	11	.983
12	-.099	.114	4.216	12	.979
13	.009	.113	4.222	13	.989
14	-.034	.112	4.317	14	.993
15	-.006	.111	4.319	15	.996
16	.189	.109	7.317	16	.967

a. The underlying process assumed is independence (white noise).

b. Based on the asymptotic chi-square approximation.

4.1.8 Estimation Graph – ACF (Difference-1)

Chart -4

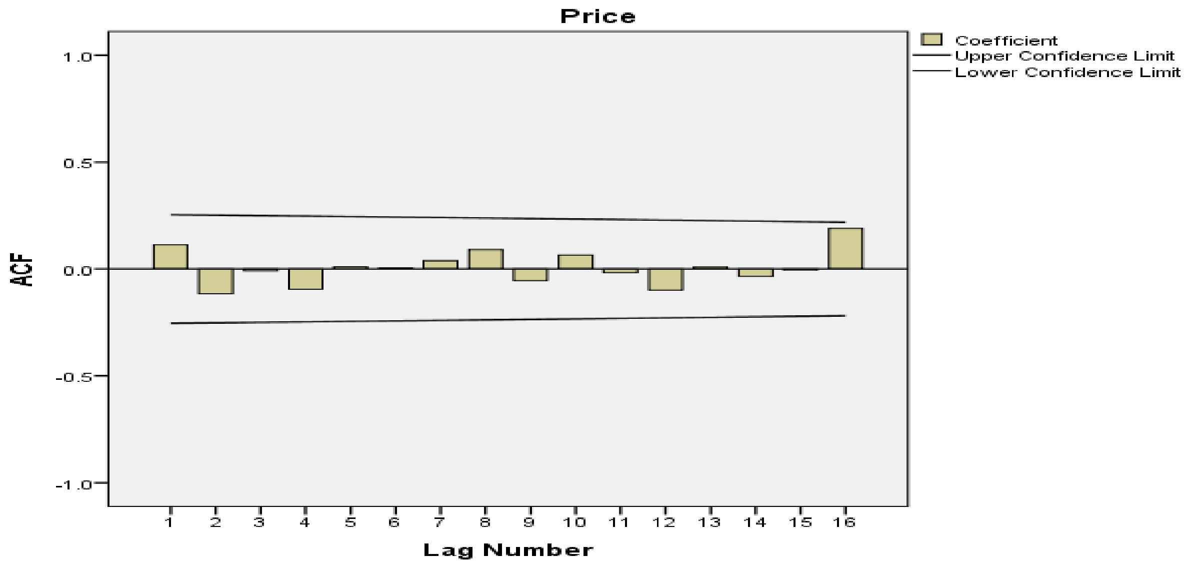


Table 9 Partial Autocorrelations

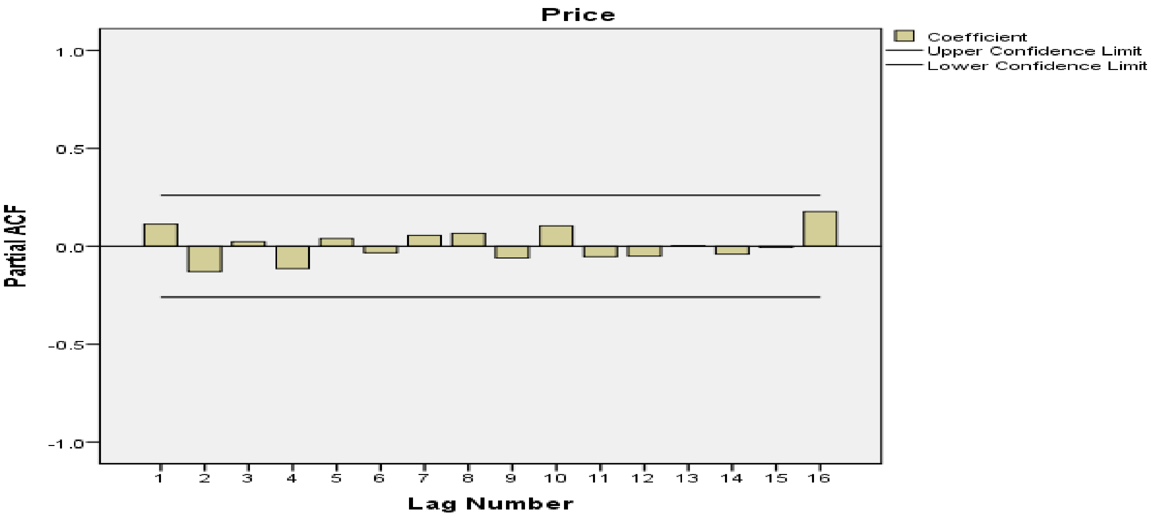
Series: Price

Lag	Partial Autocorrelation	Std. Error
1	.113	.130
2	-.131	.130
3	.022	.130
4	-.114	.130
5	.040	.130
6	-.033	.130
7	.055	.130
8	.065	.130

9	-.061	.130
10	.104	.130
11	-.055	.130
12	-.051	.130
13	.002	.130
14	-.040	.130
15	-.006	.130
16	.176	.130

4.1.9 Estimation Graph – PACF (Difference 1)

Chart -5



Interpretation: Stationarity – It means that mean, variance and autocorrelation structure do not change over time.

4.1.10 Stationary data:

When spikes of both ACF &PACF is within confidence level is known as stationary data. So, at first for identification, we have to check whether given data indicates stationary or not. If data is non- stationary than will analysis data by taking difference 1 and make it stationary. One can take difference up to data becomes stationary. So from above charts we can see data is within confidence level & become stationary so will estimate the model value ARIMA(p,d,q).

P= specify the AR model lags

D= differencing

Q= Moving average lags.

Here p value is determined from PACF graph while q value is determined from ACF graph.

How many spikes touches or out of confidence level will consider it as P value & similarly for q value by observing it in ACF chart

4.1.11Estimation

So here is our Possible ARIMA model ARIMA(P,d,q)= (0,1,0),

ARIMA (0,1,0)

AR (0), P=0

MA (0), Q=0

We will try different ARIMA model for same.

Normalized Bic value

Eg- ARIMA (0,1,0) 12.788

We will do permutation & combinations of model, whose Bic value will be lower will provide accurate forecast of prices. Will find Bic value by creating different models.

So, from above random trials of model, we find ARIMA (1,1,0) has lowest normalized BIC value. i.e. 12.788

Table-10 Time Series Modeler

Model Description

	Model Type
Model ID Price Model_1	ARIMA (0,1,0) (0,0,0)

Interpretations: The model type discloses Non -seasonal &Seasonal model, we have taken here non-seasonal model ARIMA (0,1,0).

Model Fit

Table -11

	Fit Statistic								
		Stationary R-squared	R-squared	RMSE	MAPE	MaxAPE	MAE	MaxAE	Normalized BIC
Mean		2.220E-016	.885	577.960	11.128	94.836	453.488	1593.237	12.788
SE		.	.	.	.	.	.	.	.
Minimum		2.220E-016	.885	577.960	11.128	94.836	453.488	1593.237	12.788
Maximum		2.220E-016	.885	577.960	11.128	94.836	453.488	1593.237	12.788
Percentile	5	2.220E-016	.885	577.960	11.128	94.836	453.488	1593.237	12.788
	10	2.220E-016	.885	577.960	11.128	94.836	453.488	1593.237	12.788
	25	2.220E-016	.885	577.960	11.128	94.836	453.488	1593.237	12.788
	50	2.220E-016	.885	577.960	11.128	94.836	453.488	1593.237	12.788
	75	2.220E-	.885	577.96	11.12	94.836	453.48	1593.23	12.788

		016		0	8		8	7	
	90	2.220E-016	.885	577.960	11.128	94.836	453.488	1593.237	12.788
	95	2.220E-016	.885	577.960	11.128	94.836	453.488	1593.237	12.788

Table -12

Model Statistics		
		Model
		Price-Model_1
Number of Predictors		0
Model Fit statistics	Stationary R-squared	2.220E-016
	Normalized BIC	12.788
Ljung-Box Q (18)	Statistics	10.978
	DF	18
	Sig.	.895
Number of Outliers		0

Hypothesis: H0: For model fit R square is not stationary

H1: For model fit R square is stationary.

Interpretation: Here R Square sig value is more than 0.05 hence will accept Null hypothesis while Normal BIC value is lower from hence it predicts that this

Model best suits for forecasting. In our case R-squared value is more than 0.7 hence we can consider there is high correlation between variables.

Note – R square represent as coefficient of determination which shows goodness of fit of the model. Ideally it should be more than 0.7 which indicates

High level of Correlation, whereas a measure below 0.4 would show a low correlation. However, it will depend on Specific Analysis also.

BIC (Bayesian Information Criteria) estimates the likelihood of a model to predict. There is no explicitly 'good' BIC value. BIC value needs to be compared

The best model for data is the one with lowest BIC Value.

Table -13

Residual ACF		
	Model	
	Price-Model_1	
	ACF	SE
1	.113	.130
2	-.116	.132
3	-.009	.134
4	-.094	.134

5	.010	.135
6	.002	.135
7	.039	.135
8	.090	.135
9	-.055	.136
10	.065	.136
11	-.017	.137
12	-.099	.137
13	.009	.138
14	-.034	.138
15	-.006	.138
16	.189	.138
17	-.005	.143
18	-.204	.143
19	-.136	.147
20	.135	.150
21	-.031	.152
22	.036	.152
23	-.089	.152
24	-.096	.153

Residual PACF

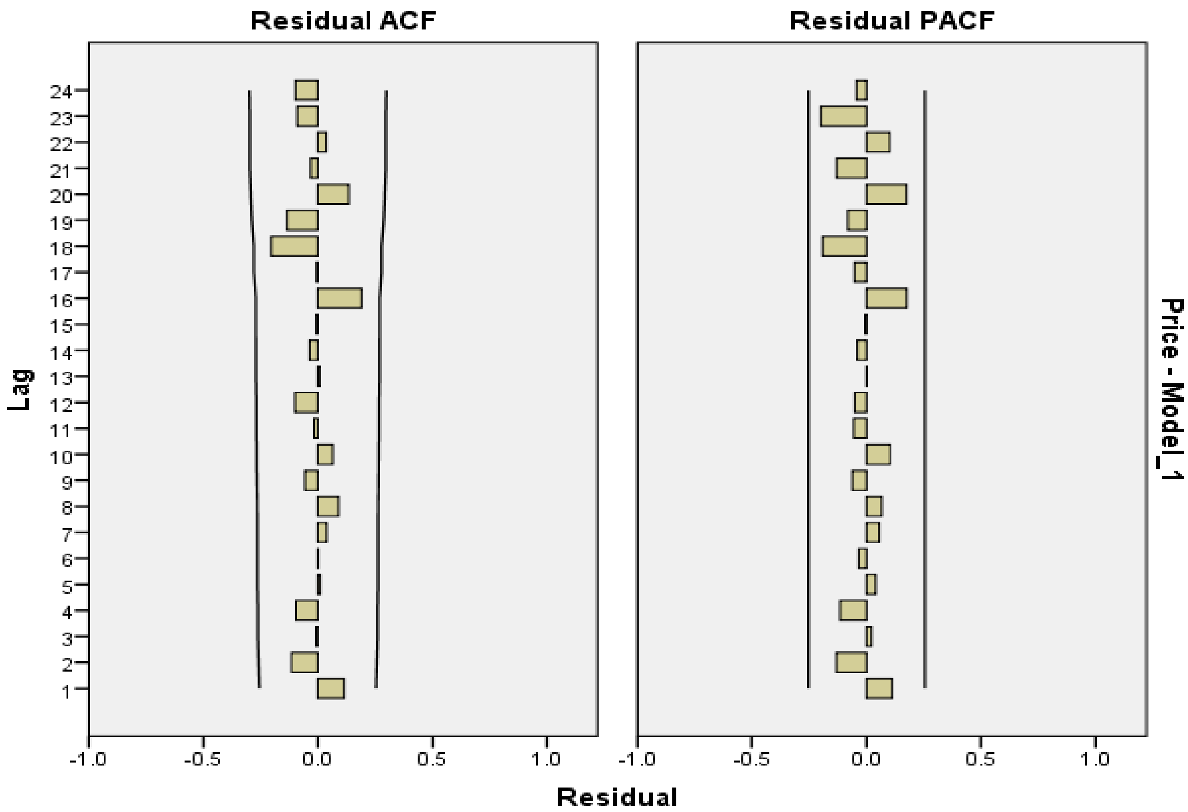
Table -14

Model		
	Price-Model_1	
	PACF	SE
1	.113	.130
2	-.131	.130
3	.022	.130
4	-.114	.130
5	.040	.130
6	-.033	.130
7	.055	.130
8	.065	.130
9	-.061	.130
10	.104	.130
11	-.055	.130
12	-.051	.130

13	.002	.130
14	-.040	.130
15	-.006	.130
16	.176	.130
17	-.052	.130
18	-.189	.130
19	-.081	.130
20	.175	.130
21	-.130	.130
22	.101	.130
23	-.197	.130
24	-.043	.130

4.1.12 Residual Graph – ACF & PACF

Chart-6



Interpretations:

By taking difference 1 model. we can conclude that all lags respectively for Residual ACF & Residual PACF falls in 95% confidence level. Between UCL & LCL. So, we can conclude our model is fit to run with appropriate forecast values.

Table-15

ARIMA Model Parameters

				Estimate	SE	t	Sig.
Price-Model_1	Price	No Transformation	Constant	10.237	75.244	.136	.892
			Difference	1			

Table-16

Model		Jun 2023	Jul 2023	Aug 2023	Sep 2023	Oct 2023	Nov 2023	Dec 2023
Price-Model_1	Forecast	5700	5710	5721	5731	5741	5751	5762
	UCL	6857	7347	7725	8045	8328	8585	8823
	LCL	4543	4074	3717	3417	3154	2918	2701

For each model, forecasts start after the last non-missing in the range of the requested estimation period, and end at the last period for which non-missing values of all the predictors are available or at the end date of the requested forecast period, whichever is earlier.

Interpretation: ARIMA model: y'(t) = c + ϕ1* y′(t−1) +⋯ + ϕp*y′(t−p) + θ1*ε(t−1) +⋯ + θq*ε(t−q) + εt.

Here,

C = constant

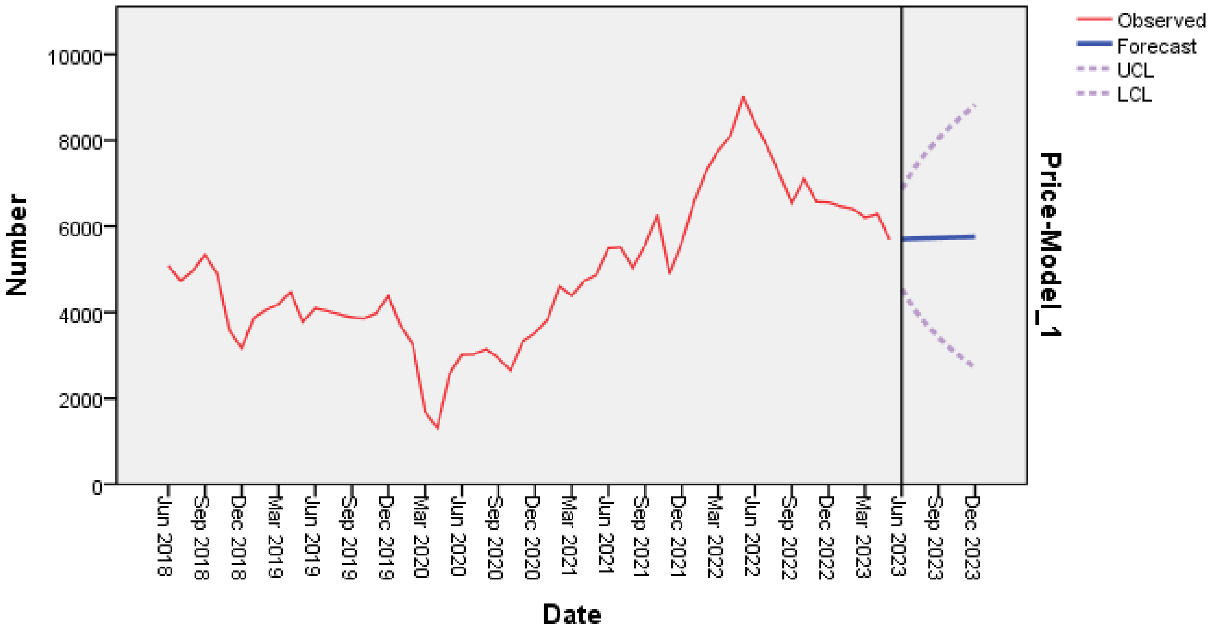
ϕ1= AR lag 1 value

(t-p) indicates interval of lagging like 1 lag, 2 lag and so on.

Where errors terms are error of auto regressive models of respective lags. The errors Et and E(t-1) are errors of following

4.1.13 Forecast graph of Crude oil prices

Chart-7



Interpretation: In this chart red line represents the observed value (historical data) and on basis of it by applying ARIMA model we get future forecast from June 2023 up to December 2023 with blue line. We can also see it in march 2020 the prices were lower than lower control limit. Which indicates major down fall of prices this is due to Corona emerging situation along with other external factors.

4.2 Forecast analysis of BSE Index Price

Sequence Plot

Table-17

Model Description

Model Name		MOD_1
Series or Sequence	1	bse index price
Transformation		None
Non-Seasonal Differencing		0
Seasonal Differencing		0
Length of Seasonal Period		12
Horizontal Axis Labels		YEAR, not periodic
Intervention Onsets		None
Reference Lines		None
Area Below the Curve		Not filled

Applying the model specifications from MOD_1

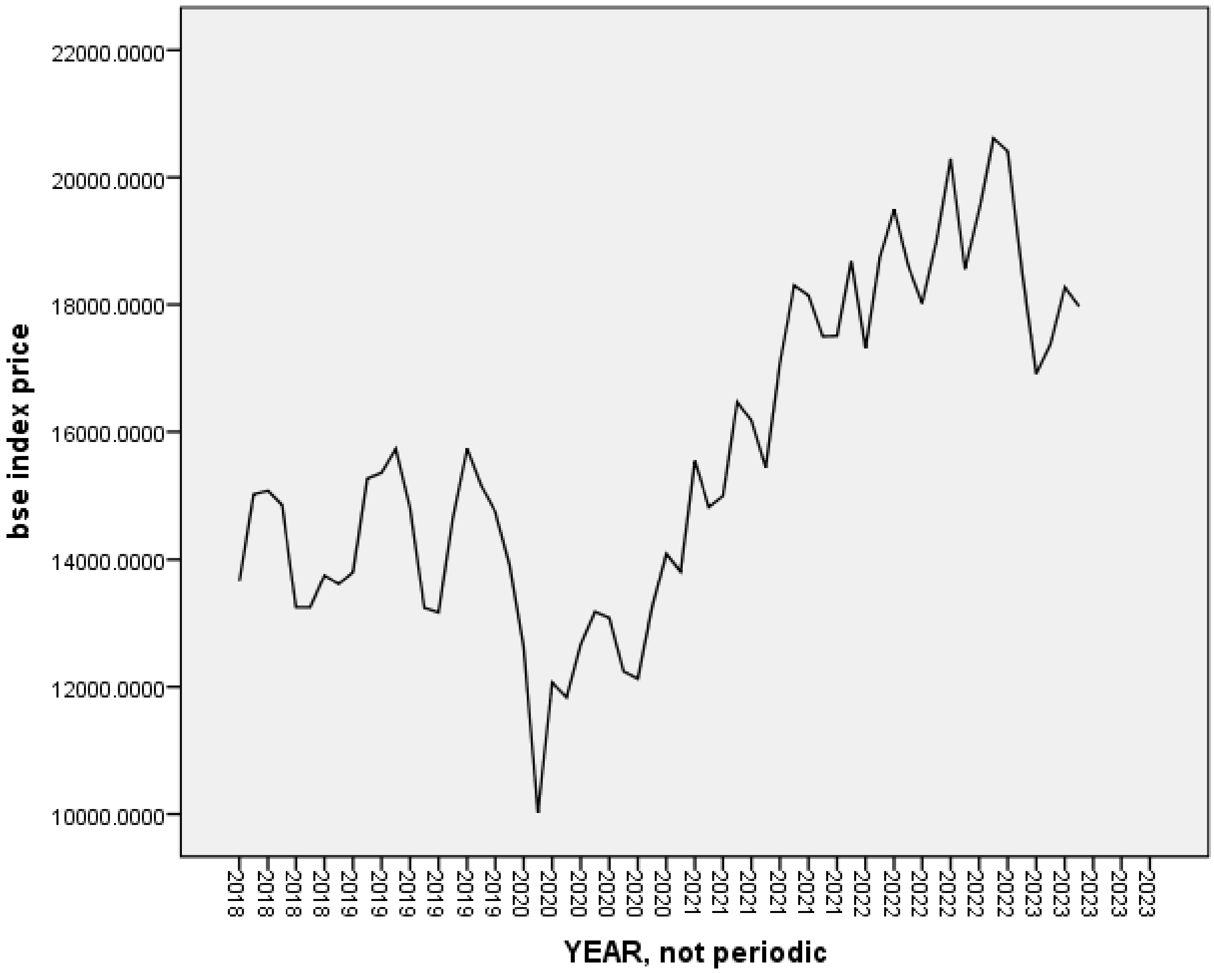
Table -18

Case Processing Summary

		bse index price
Series or Sequence Length		65
Number of Missing Values in the Plot	User-Missing	0
	System-Missing	5

4.2.2 Sequence plot for Identification

Chart-8



Interpretation: Model description explains the summary of ARIMA model and overall details of the model & chart where it discloses model name, which series is used in model on which ARIMA model is focused similarly case processing sequence discloses the series or sequence length of data i.e in our case we have taken monthly wise closing price of crude oil for five years period (12*5).

The first stage in the ARIMA Model is to determine if the data is stationary or non-stationary. We may use this sequence plot chart to determine the trend and variations, since there are more upward trends and fluctuating crude oil prices that are non-seasonal and non-stationary, which is why we can use the ARIMA Model here. The ARIMA model is used to anticipate the future of non-stationary data. As a result of this chart, we may determine that the data is non-stationary and can apply the ARIMA Model.

4.2.3 Identification - Autocorrelations

Table-19

Model Description

Model Name		MOD_3
Series Name	1	bse index price
Transformation		None
Non-Seasonal Differencing		1
Seasonal Differencing		0
Length of Seasonal Period		12

Maximum Number of Lags	16
Process Assumed for Calculating the Standard Errors of the Autocorrelations	Independence (white noise)[a]
Display and Plot	All lags

Applying the model specifications from MOD_3

a. Not applicable for calculating the standard errors of the partial autocorrelations.

Table-20

Case Processing Summary

		bse index price
Series Length		65
Number of Missing Values	User-Missing	0
	System-Missing	5[a]
Number of Valid Values		60
Number of Values Lost Due to Differencing		1
Number of Computable First Lags After Differencing		58

a. Some of the missing values are imbedded within the series.

4.2.4 BSE Index Price

Table -21

Autocorrelations

Series: bse index price

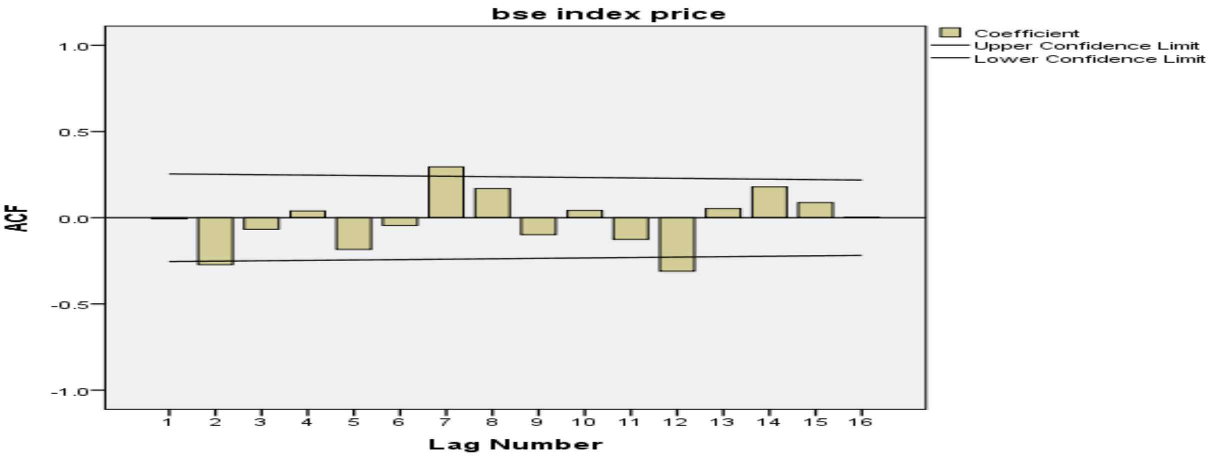
Lag	Autocorrelation	Std. Error[a]	Box-Ljung Statistic		
			Value	df	Sig.[b]
1	-.007	.127	.003	1	.959
2	-.272	.126	4.665	2	.097
3	-.067	.125	4.949	3	.176
4	.039	.124	5.049	4	.282
5	-.184	.122	7.302	5	.199
6	-.045	.121	7.439	6	.282
7	.295	.120	13.455	7	.062
8	.169	.119	15.473	8	.051
9	-.098	.118	16.171	9	.063
10	.042	.117	16.298	10	.091
11	-.125	.115	17.477	11	.095
12	-.310	.114	24.857	12	.016
13	.053	.113	25.076	13	.023
14	.179	.112	27.637	14	.016
15	.088	.111	28.264	15	.020
16	.002	.109	28.264	16	.029

a. The underlying process assumed is independence (white noise).

b. Based on the asymptotic chi-square approximation.

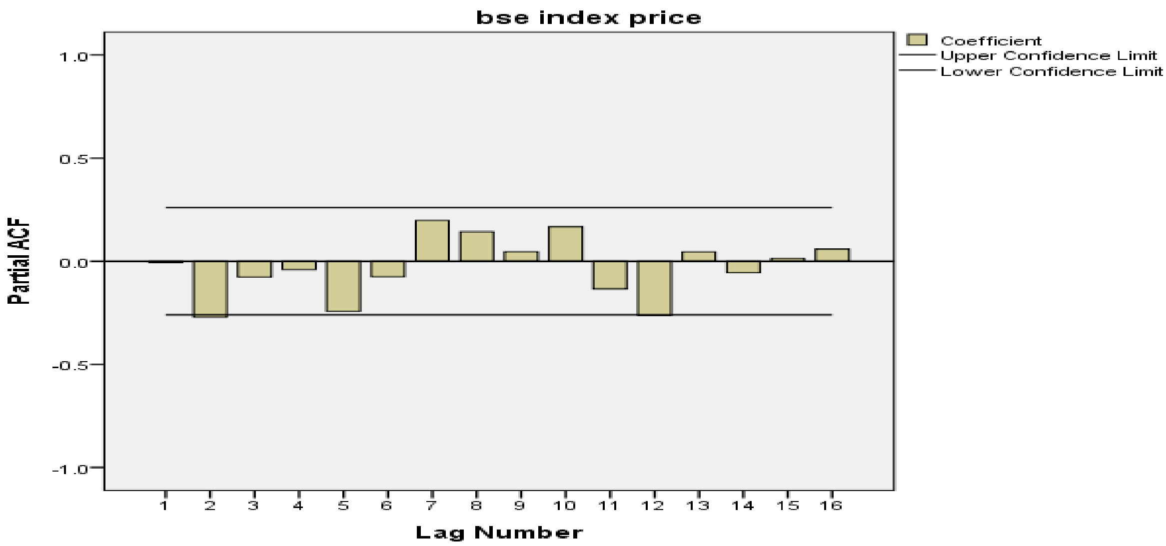
4.2.5 Estimation Graph - ACF (Difference -1)

Chart- 9



4.2.6 Estimation Graph -ACF (Difference -1)

Chart-10



Interpretation: After sequence plotting at first for identification, we have to check whether given data indicates stationary or not. If data is non- stationary than with help of autocorrelations will analysis data by taking difference 1 and make it stationary. One can take difference upto data becomes near to or to stationary. So from above charts we can see data is within confidence level & become stationary so will estimate the model value ARIMA(p,d,q).

Here by observing in PACF graph we have determined at lag 2 spikes touches confidence level also at lag 12, so p= (2,12). Similarly, by observing in ACF graph we can determine q value, q = (2,7,12) & d= difference = 1

Estimation:

So here is our Possible ARIMA model ARIMA(P,d,q)= (2,1,2)

ARIMA (2,1,2)

AR (0), P=2,12

MA (0), Q=2,7,12

We will try different ARIMA model for same.

Normalized BIC value

Example:

ARIMA (2,1,12)	14.833
ARIMA (2,1,7)	14.497
ARIMA (2,1,2)	14.237
ARIMA (12,1,12)	15.639
ARIMA (12,1,7)	15.189
ARIMA (12,1,2)	14.839

We will do permutation & combinations of model, whose Bic value will be lower will provide accurate forecast of prices. Will find BIC value by creating different models.

So, from above random trials of model, we find ARIMA (2,1,2) has lowest normalized BIC value. I.E 14.237.

Time Series Modeler

Table-22

Model Description

			Model Type
Model ID	bse index price	Model_1	ARIMA (2,1,2) (0,0,0)

Table -23

Model Fit									
		Fit Statistic							
		Stationary R-squared	R-squared	RMSE	MAPE	MaxAPE	MAE	MaxAE	Normalized BIC
Mean		.090	.843	1038.691	5.183	30.323	776.938	3038.671	14.237
SE		.	.	.	.	.	.	.	.
Minimum		.090	.843	1038.691	5.183	30.323	776.938	3038.671	14.237
Maximum		.090	.843	1038.691	5.183	30.323	776.938	3038.671	14.237
Percentile	5	.090	.843	1038.691	5.183	30.323	776.938	3038.671	14.237
	10	.090	.843	1038.691	5.183	30.323	776.938	3038.671	14.237
	25	.090	.843	1038.691	5.183	30.323	776.938	3038.671	14.237
	50	.090	.843	1038.691	5.183	30.323	776.938	3038.671	14.237
	75	.090	.843	1038.691	5.183	30.323	776.938	3038.671	14.237
	90	.090	.843	1038.691	5.183	30.323	776.938	3038.671	14.237
	95	.090	.843	1038.691	5.183	30.323	776.938	3038.671	14.237

Table-24

Model Statistics

		Model
		bse index price-Model_1
Number of Predictors		0
Model Fit statistics	Stationary R-squared	.090
	Normalized BIC	14.237
Ljung-Box Q (18)	Statistics	18.773
	DF	14
	Sig.	.174
Number of Outliers		0

Hypothesis: H0: For model fit R square is not stationary

H1: For model fit R square is stationary.

Interpretation: Here R Square sig value is more than 0.05 hence will accept Null hypothesis while Normal BIC value is lower from hence it predicts that this

Model best suits for forecasting. In our case R-squared value is less than 0.7 hence we can consider there is low correlation between variables.

Table-26

Residual ACF

	Model	
	bse index price-Model_1	
	ACF	SE
1	-.002	.130
2	-.006	.130
3	-.021	.130
4	.043	.130
5	-.126	.130
6	-.003	.133
7	.251	.133
8	.161	.140
9	-.084	.143
10	.029	.144
11	-.123	.144
12	-.287	.146
13	.034	.155
14	.086	.156

15	.050	.156
16	-.014	.157
17	-.046	.157
18	-.162	.157
19	-.125	.160
20	-.085	.161
21	.055	.162
22	-.209	.162
23	.159	.167
24	-.042	.169

Table-27

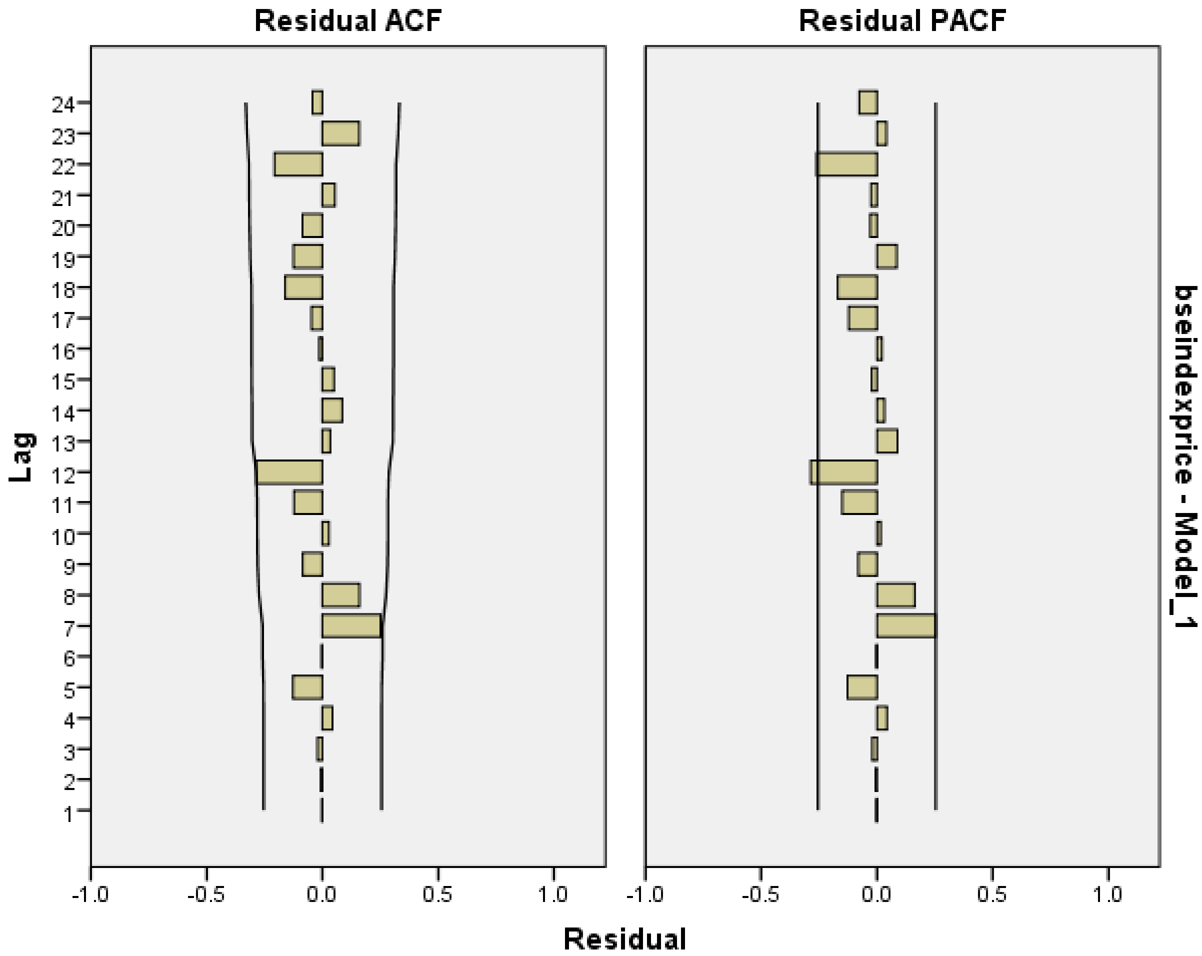
Residual PACF

	Model	
	bse index price-Model_1	
	PACF	SE
1	-.002	.130
2	-.006	.130
3	-.021	.130
4	.043	.130
5	-.127	.130
6	-.003	.130
7	.255	.130
8	.165	.130
9	-.081	.130
10	.016	.130
11	-.150	.130
12	-.286	.130
13	.088	.130
14	.032	.130
15	-.023	.130
16	.021	.130
17	-.122	.130
18	-.170	.130
19	.085	.130
20	-.029	.130
21	-.026	.130
22	-.261	.130

23	.041	.130
24	-.077	.130

4.2.7 Residual Graphs -BsE Index price

Chart-11



Interpretation: By taking difference 1 model. we can conclude that NOT all lags respectively for Residual ACF & Residual PACF falls in 95% confidence level. Between UCL & LCL but at level 1, minimum lags touch the UCL /LCL, so we can conclude our model is fit to run with appropriate forecast values.

Table-27

ARIMA Model Parameters

					Estimate	SE	t	Sig.
bse index price-Model_1	bse index price	No Transformation	Constant		69.453	84.907	.818	.417
			AR	Lag 1	.324	.532	.608	.546
				Lag 2	-.161	.496	-.325	.746
			Difference		1			

	MA	Lag 1	.368	.537	.685	.496
		Lag 2	.119	.530	.225	.823

Table-28

Forecast

Model		Jun 2023	Jul 2023	Aug 2023	Sep 2023	Oct 2023	Nov 2023	Dec 2023
bse index price-Model_1	Forecast	17971.0175	18141.2721	18254.7162	18322.1597	18383.8731	18451.1452	18521.1393
	UCL	20050.5160	21017.9071	21442.9359	21725.6185	22002.9478	22287.5258	22565.4796
	LCL	15891.5189	15264.6371	15066.4965	14918.7008	14764.7984	14614.7646	14476.7990

For each model, forecasts start after the last non-missing in the range of the requested estimation period, and end at the last period for which non-missing values of all the predictors are available or at the end date of the requested forecast period, whichever is earlier.

Interpretation: ARIMA model: y'(t) = c + ϕ1* y′(t−1) +⋯ + ϕp*y′(t−p) + θ1*ε(t−1) +⋯ + θq*ε(t−q) + εt.

Here,

C = constant

t-p = indicates the terms of lags at different lags,

e – indicates errors.

hence on basis of this formula we have forecasted this price.

We can analysis every month price is increased at estimation of 69.54. The forecast is ranges in between Ucl & Lcl.

<u>Benefits of forecast</u>

The test performed in analysis of forecast itself indicates the available forecast holds significance importance.

This forecast helps investors to take early bird decisions & help to make necessary strategies & planning of business.

Those business which are trading in crude oil or in import – export of Crude oil, or other petroleum products can do future purchase of crude oil along with ca

Purchase dollars. This forecast will helpfull in making decision how much fluctuations & variations can be estimated in crude oil prices.

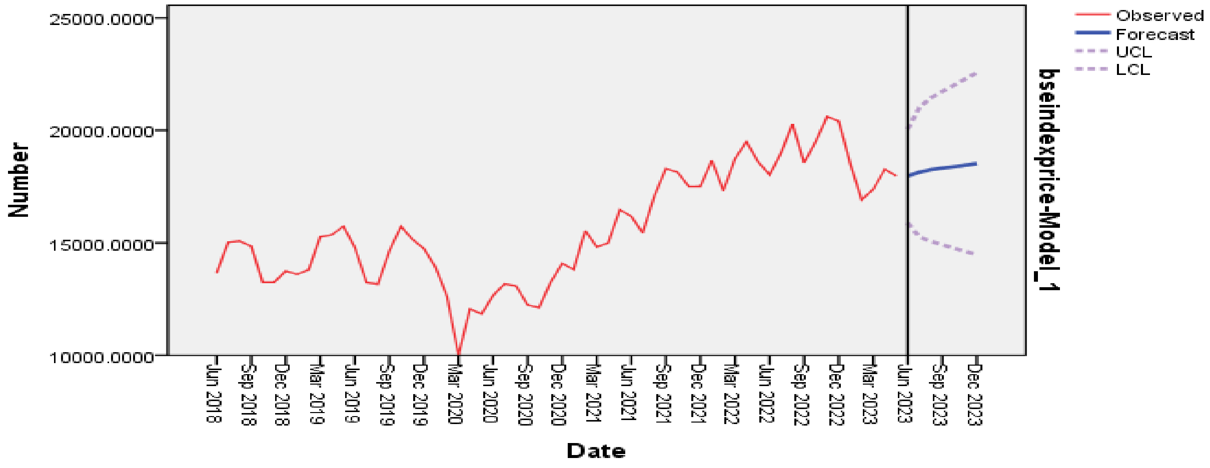
Eg – If currently crude oil price is reduced & on basis of forecast if price will be increased in future assuming other factors will remain constant then they investors/ companies can do future purchase of Crude oil along with dollars at today’s date.

Assumptions:

This forecast is truly based on Historical data and it is assumed that other external factors are constant.

4.3 BSE Index price Forecast Chart

Chart-12



Interpretation: In This chart red line represents the observed value (historical data) and on basis of it by applying ARIMA model we get

Future forecast from June 2023 up to December 2023 with blue line. We can also see it in march 2020 the prices were lower than lower control limit. Which indicates major down

Fall of prices this is due to Corona emerging situation along with other external factors.

Regression analysis between BSE Index of oil & gas industries & Crude oil prices.

month	Bse index forecast	crude oil oprice forecast
Jun-23	18047.62	5629
Jul-23	18120.59	5629
Aug-23	18266.587	5637
Sep-23	18193.59	5645
Oct-23	18339.583	5653
Nov-23	18412.57	5661
Dec-23	18485.57	5669

SUMMARY OUTPUT

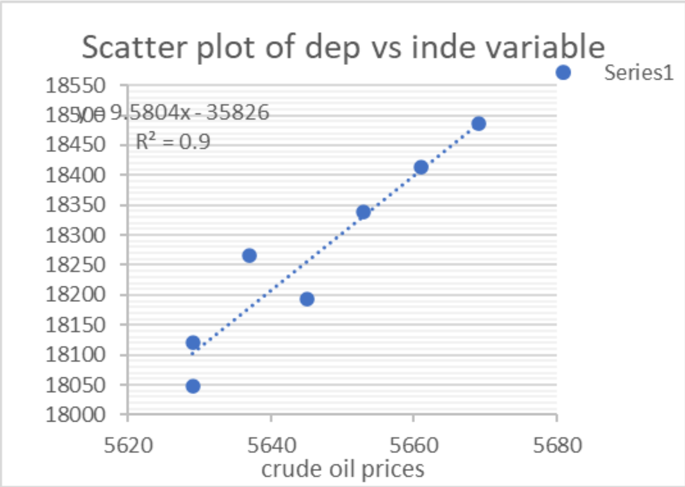
Regression Statistics	
Multiple R	0.948687
R Square	0.900008
Adjusted R	0.88001
Standard E	54.62077
Observatic	7

ANOVA

	df	*SS*	*MS*	*F*	*ignificance F*
Regressior	1	134266.1	134266.1	45.00396	0.001114
Residual	5	14917.14	2983.428		
Total	6	149183.2			

	Coefficients	*andard Erre*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
Intercept	-35825.5	8063.25	-4.44306	0.006746	-56552.8	-15098.3	-56552.8	-15098.3
X Variable	9.580369	1.428094	6.708499	0.001114	5.909336	13.2514	5.909336	13.2514

RESIDUAL OUTPUT			PROBABILITY OUTPUT	
Observation	*Predicted Y*	*Residuals*	*Percentile*	*Y*
1	18102.35	-54.7323	7.142857	18047.62
2	18102.35	18.23775	21.42857	18120.59
3	18179	87.5918	35.71429	18193.59
4	18255.64	-62.0482	50	18266.59
5	18332.28	7.3019	64.28571	18339.58
6	18408.92	3.64595	78.57143	18412.57
7	18485.57	0.003	92.85714	18485.57



Linear regression equation =y= a+ bx

Where,

X= independent variable

Y= dependent variable

a= y intercept (expected mean value of y when all x variables =0, basically point on regression graph that's point where line crosses the y axis)

b= slope of regression line (which is rate of change for y as x changes

Note- linear regression always have error term as because in real life predictors are never precise.

Here we have taken crude oil prices as independent variable whereas BSE Index prices as dependent variable.

So, in our case we can say there is strong positive relationship between index &crude oil prices i.e – 0.94

R square- It is called as Coefficient of determination which is used as indicator of goodness of fit. It basically shows how many points fall on Regression line

Here our R square is 0.9 means 90% of our value fit the regression analysis model. Higher the value higher the good fit.

Adjusted r square – It is the r square adjusted for the number of of independent variables in model.

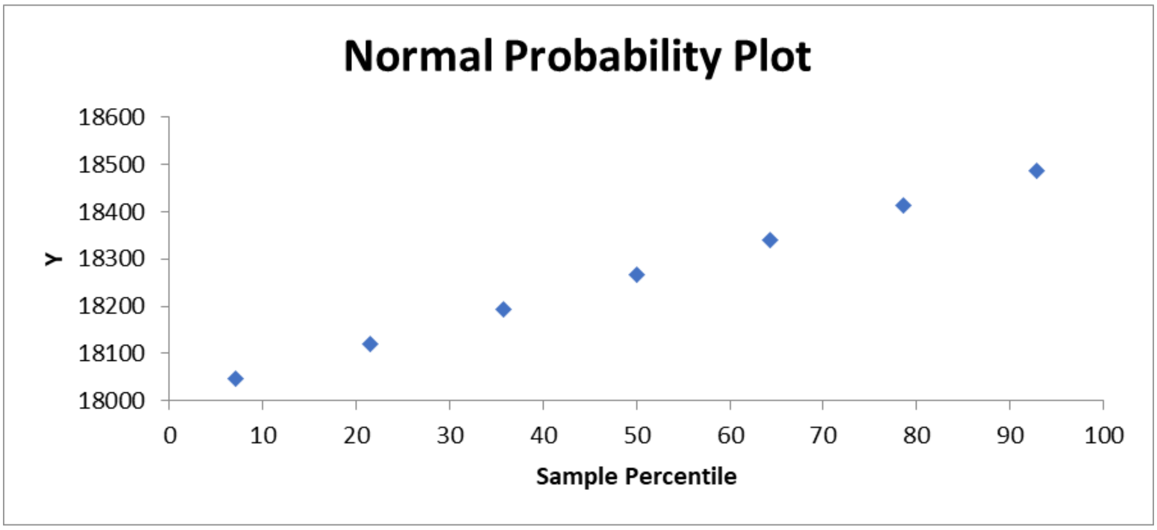
Standard error – it shows goodness of fit measure that basically shows the precision of your regression analysis so smaller the value the more certain you can be about your regression equation.

ANNOVA – It provides level of Variability with help of its components within your regression model F-It is called f test for null hypothesis which is used to measure overall significance of model Significance f- which is the p value of f which helps to determine how reliably & statistically significant our results are. If sign f is less than 0.05 that's 5% our model is okay. greater than 5 we will have to better choose other independent variables.

Coefficients -It is main important thing of Regression Analysis which enables you to build linear Regression equation in form of y=bx+a

Residual output – It helps to determine how many observations are there what is predicted forecast (accurate) of actual forecast values so in case of 1st observation we estimated forecast 18102 and actual forecast is 18047 which means residual(difference) is -54.53. lesser the residual more accurate is your actual forecast (dependent Variable). & so on for all observation. from all observation 7th observation has lesser residual which indicates there is more accuracy of actual forecast compared with estimate forecast. Hence residual helps us to predict how far actual forecast is there from estimated forecast.

Percentile value discloses probability output which is shown with help of graph.



5. Findings

There is an increase in price of crude oil on basis of estimate value.

There is an auto correlation between variables.

There is a dependency of crude oil prices on BSE Oil & Gas Index.

There is a fit to forecast the BSE Oil & Gas Index value from crude oil prices.

Apart from above hypothesis finding the model is tested with the real prices which shows more than 90% accuracy with the real prices in the last six months.

6. Conclusions

To Conclude we can say with the help of the ARIMA we can forecast the prices of the crude oil by applying different model values which is very helpful as we are not stick with one model values of p, q and d variables. With the help of the forecasted value, we can find the Index value or vice a versa. ARIMA is very useful to forecast the uni-variable without comparison with other variable and regression can be useful to forecast the value of two variable or more.

Here, in this study we applied ARIMA on Oil and Gas Sector and Crude Oil as commodity, one can use this model for different commodity and sectors for the forecasting the value of the same.

Bibliography

Sharma, A. (2018), Relationship between crude oil prices and stock market. In Relationship between crude oil prices and stock market: Sharma, Ankit.

Kumar, S., Choudhary, S., Singh, G., & Singhal, S. (2021). Crude oil, gold, natural gas, exchange rate and Indian stock market: Evidence from the asymmetric nonlinear ARDL model. Resources Policy, 73, 102194

Talha, M., Sohail, M., Tariq, R., & Ahmad, M. T. (2021). Impact of oil prices, energy consumption and economic growth on the inflation rate in Malaysia. *Cuadernos de Economía*, *44*(124), 26-32. https://cude.es/submit-a-manuscript/index.php/CUDE/article/view/144

Yadav, N., Tandon, P., Tripathi, R., & Shastri, R. K. (2020). A dynamic relationship between crude oil price and Indian equity market: an empirical study with special reference to Indian benchmark index Sensex. *Benchmarking: An International Journal*, *28*(2), 582-599. https://www.emerald.com/insight/content/doi/10.1108/BIJ-06-2020-0306/full/html?utm_campaign=Emerald_Strategy_PPV_November22_RoN

Venkataraman, P., & Subramaniam, K. (2022). An Empirical Analysis of Financial Performance of Selected Oil Exploration and Production Companies in India. *International Journal of Engineering and Management Research*, *e-ISSN*, 2250-0758. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4136624

Makala, D., & Li, Z. (2021, February). Prediction of gold price with ARIMA and SVM. In *Journal of Physics: Conference Series* (Vol. 1767, No. 1, p. 012022). IOP Publishing. https://iopscience.iop.org/article/10.1088/1742-6596/1767/1/012022/meta

Dritsaki, C. (2021). Oil consumption forecasting using ARIMA models: an empirical study for Greece. *International Journal of Energy Economics and Policy*. https://www.zbw.eu/econis-archiv/bitstream/11159/7770/1/1771767685_0.pdf

Menculini, L., Marini, A., Proietti, M., Garinei, A., Bozza, A., Moretti, C., & Marconi, M. (2021). Comparing prophet and deep learning to ARIMA in forecasting wholesale food prices. *F* Monge, M., & Infante, J. (2022). A fractional ARIMA (ARFIMA) model in the analysis of historical crude oil prices. *Energy Research Letters*, *3*(Early View). Forecasting, *3*(3), 644-662or Greece. *International Journal of Energy Economics and Policy*. https://www.zbw.eu/econis-archiv/bitstream/11159/7770/1/1771767685_0.pdf

Lin, B., & Bai, R. (2021). Oil prices and economic policy uncertainty: Evidence from global, oil importers, and exporters' perspective. *Research in International Business and Finance*, *56*, 101357. https://scholar.google.com/scholar?start=30&q=crude+oil+price+trends+in+global+market+2021&hl=en&as_sdt=0,5#d=gs_cit&t=1689715182607&u=%2Fscholar%3Fq%3Dinfo%3AsE8mqFeFg9wJ%3Ascholar.google.com%2F%26output%3Dcite%26scirp%3D36%26hl%3Den

Elsaraiti, M., Ali, G., Musbah, H., Merabet, A., & Little, T. (2021, April). Time series analysis of electricity consumption forecasting using ARIMA model. In *2021 IEEE Green technologies conference (GreenTech)* (pp. 259-262). IEEE. https://ieeexplore.ieee.org/abstract/document/9458543