**CHAPTER 4**

**ANALYSIS OF DATA**

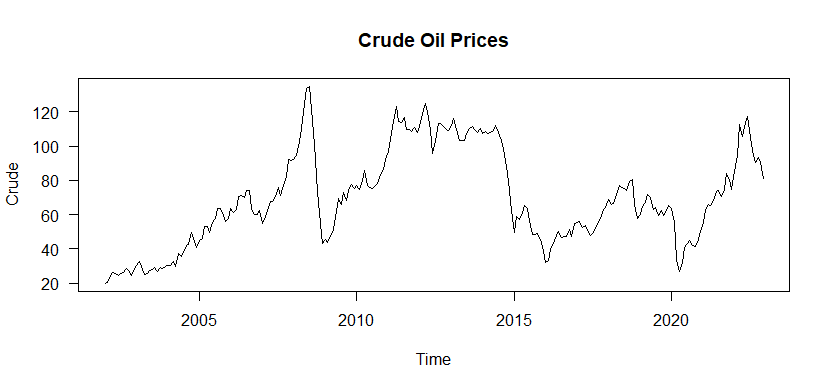
**4.0 Introduction**

Monthly Brent Crude Oil prices from the Bank of Ghana website was analysed from 2002 to, 2024 as stipulated in the previous chapter for the purposes of this study. Three different time series models (Autoregressive Integrated Moving Average (ARIMA); Singular Spectrum Analysis (SSA); and Prophet) were adopted to model this data in an effort to predict crude prices for 2023 and further compare which model effectively predicts crude oil prices.

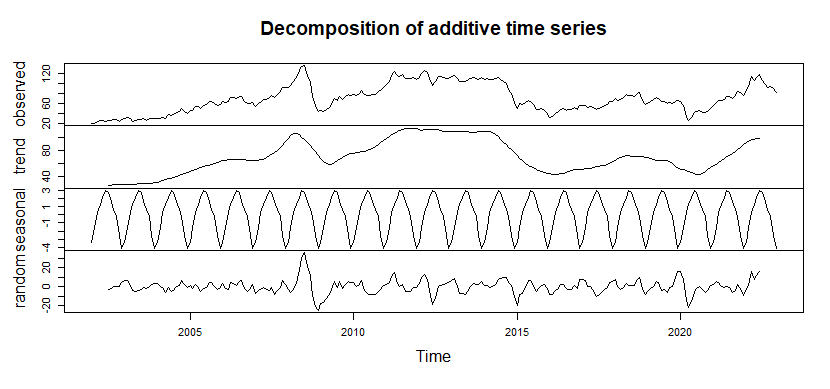
**4.1 Preliminary Analysis**

In the preliminary analysis, brent crude oil prices from 2002 to 2022 was visualized in order to ascertain whether or not the data was stationary, informing the researchers as to the level of differencing to adopt and further analysis to undertake. The retrieved crude oil data was imported into R and analysed accordingly.

**4.1.1 Crude Oil Price Visualization**



**Figure 4.1: A plot of Crude oil prices from 2002 – 2022**

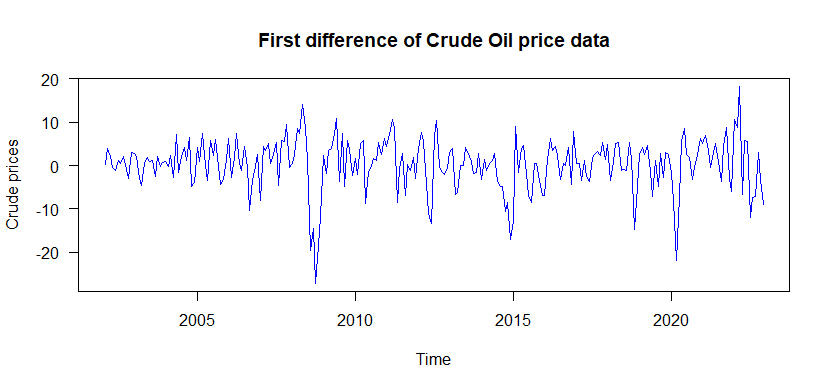
As depicted in Figure 4.1, there existed various components within the time series plot. From the plot, there existed an uptrend from 2002 to 2008 and a sharp downtrend in 2009. Similarly, crude oil prices recorded an uptrend from mid-2015 through till 2022. Clearly, the data exhibits variety of components and patterns such as the unusual and random fluctuations across different times. There is therefore the need to perform a decomposition of the Brent crude oil prices data so as to unearth the various patterns present in the data. 

**Figure 4.2: A decomposed plot of Crude oil prices from 2002 – 2022**

The various time series components showed in the decomposed plot includes the trend, seasonal effect, random effect as well as the observed effect. It can clearly be seen from Figure 4.2 that, there appears to be a wave-like up and down trend present in the crude oil price data with no seasonal effect. However there seem to be a random effect present in the data. Hitherto, there is the need to an appropriate differencing in order to stationaries the data.

**4.1.2 Applying Appropriate Differencing to Crude Oil Price Data**

In order to build effective and efficient time series models with the Crude oil price data and further predict further prices, there is the need to achieve stationarity in the data by applying the appropriate differencing (I) so as to stationaries the data before further techniques can be applied to the data. An Augmented Dickey-Fuller test (ADF) test performed on the crude oil data returned a p-value of 0.4411 (Dickey-Fuller = -2.32, Lag order = 6) which suggests that the data is not stationary and therefore the is the need to difference the data.

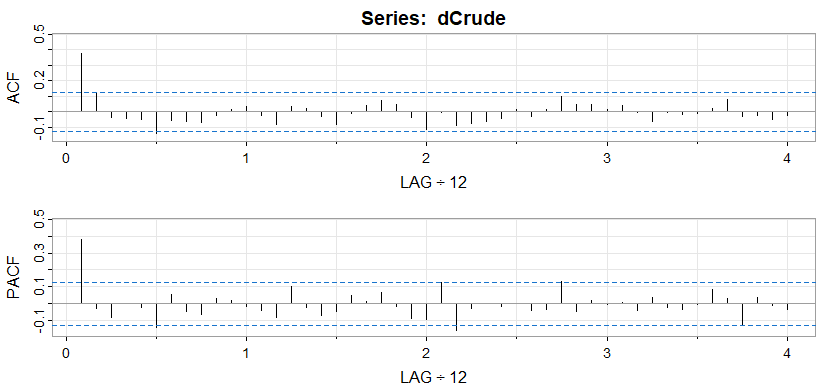


**Figure 4.3: A plot of the first difference of Crude oil prices.**

As shown in Figure 4.3, it can clearly be seen that applying the first differencing to prices of crude oil from 2002 to 2022 results in a stationarization. A further ADF test performed on the differenced data returned a p-value of 0.01 indicative that the data is stationary at the first difference.

**4.2 Further Analysis**

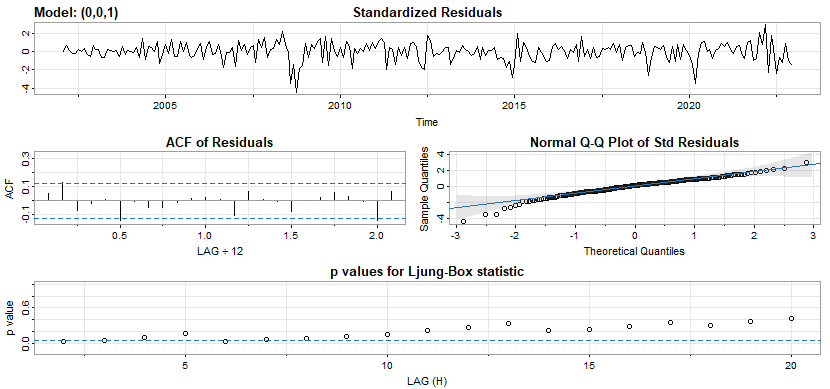
**4.2.1 Autoregressive Integrated Moving Average (ARIMA) Model Building**

In building the ARIMA model, the acf2 function within the Applied Statistical Time Series Analysis (astsa) package was adopted to unearth the various competing models so as to derive the most effective and suitable model for forecasting 2023 crude oil prices. 

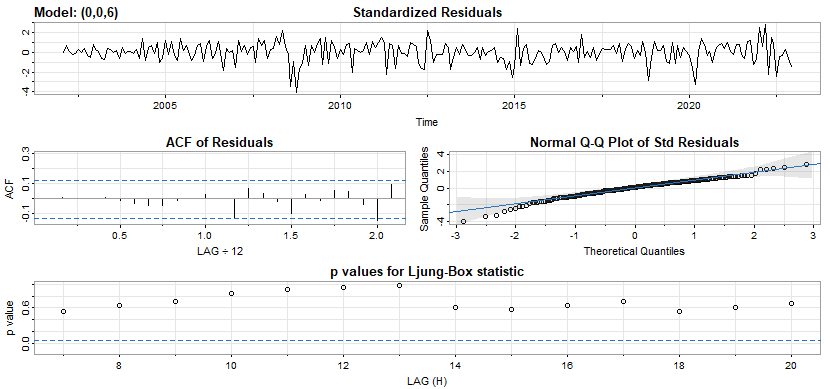
**Figure 4.4: ACF and PACF plot of differenced data**

By critically observing the plot in Figure 4.4 above, it can clearly be observed that, for the ACF, there exist a significant lag spike in Lags 1 and lag 6 which falls outside the significant region, which suggest a Moving Average, MA (1) and MA (6) process. However, since there is no lag spike quarterly, bi-quarterly and annually, we can conclude that there is no seasonal component present in the crude oil price data. The lack of seasonal variations was similarly shown in Figure 4.2 in the decomposition plot. Considering the PACF plots in Figure 4.4, we realize that there is a significant lag spike in Lag 1, and lag 5 suggestive of an Autoregressive, AR (1) and AR (5) process. Having ascertained this information from both ACF and PACF plots, various models were built by the combination of the Autoregressive (AR) and Moving Average (MA) components so as to arrive at the most suitable model with the lowest AIC, AICc and BIC value.

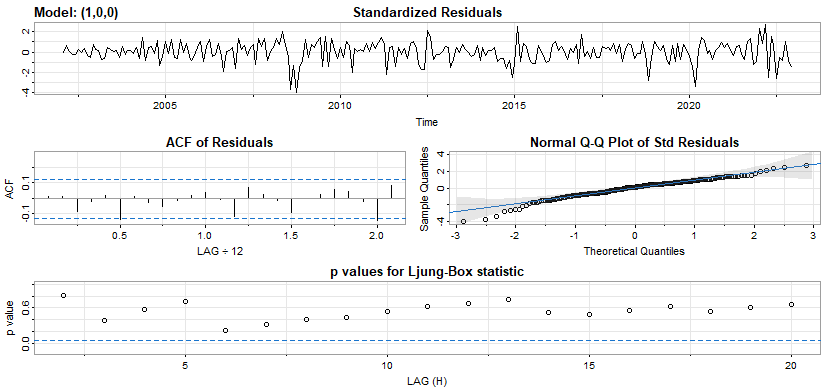
The study went ahead and fitted various models and their residual plots were compared together with their AICc value.



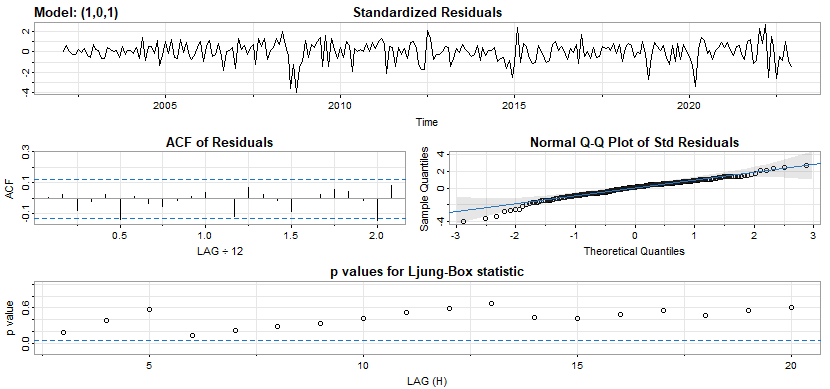
**Figure 4.5: Residual plot of ARIMA Model (0,0,1)**



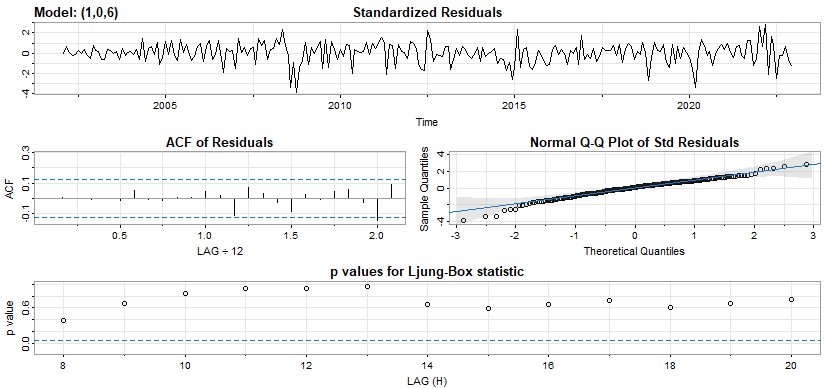
**Figure 4.6: Residual plot of ARIMA Model (0,0,6)**



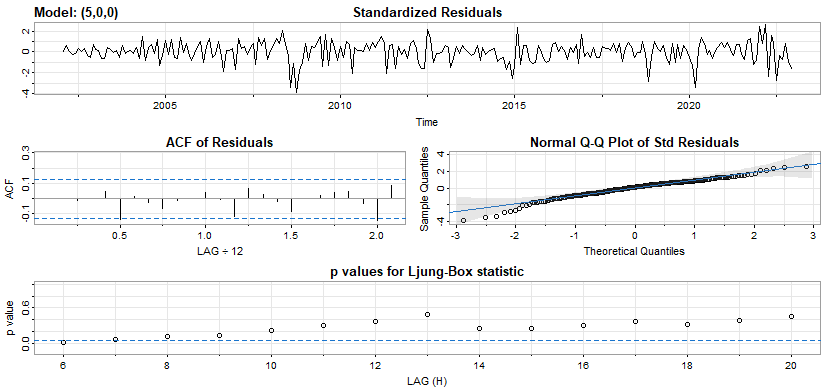
**Figure 4.7: Residual plot of ARIMA Model (1,0,0)**



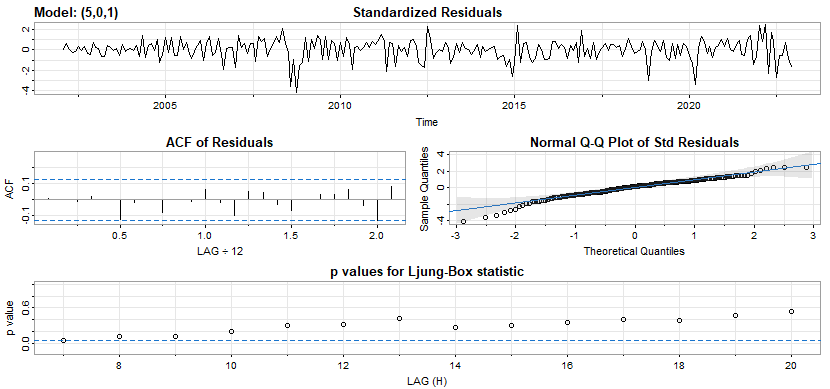
**Figure 4.8: Residual plot of ARIMA Model (1,0,1)**



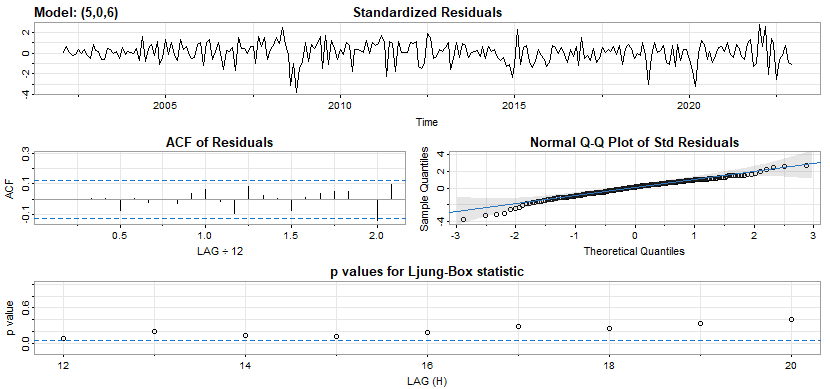
**Figure 4.9: Residual plot of ARIMA Model (1,0,6)**



**Figure 4.10: Residual plot of ARIMA Model (5,0,0)**



**Figure 4.11: Residual plot of ARIMA Model (5,0,1)**

****

**Figure 4.12: Residual plot of ARIMA Model (5,0,6)**

**Table 4.1: Fitted Models and their respective AICc values.**

|  |  |  |  |
| --- | --- | --- | --- |
| **FIT** | **MODEL** | **AICc** | **P-VALUE** |
| Fit 1 | ARIMA (0,0,1) | 6.294753 | 0.0000 |
| Fit 2 | ARIMA (0,0,6) | 6.285031 | 0.0000 |
| Fit 3 | ARIMA (1,0,1) | 6.282985 | 0.0110 |
| Fit 4 | ARIMA (1,0,6) | 6.291099 | 0.0397 |
| Fit 5 | ARIMA (5,0,1) | 6.284281 | 0.0000 |
| Fit 6 | ARIMA (5,0,6) | 6.301782 | 0.2464 |
| Fit 7 | ARIMA (1,0,0) | 6.275358 | 0.0000 |
| Fit 8 | ARIMA (5,0,0) | 6.299578 | 0.0000 |

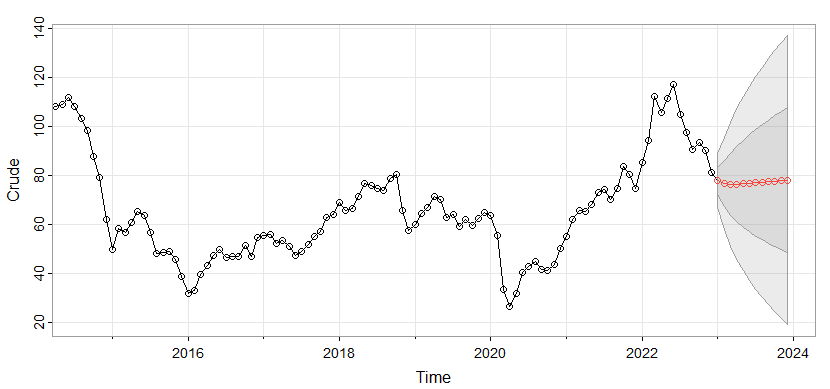
Careful consideration and selection from the 12 different fitted models revealed the ARIMA (1, 1, 0) model to be the most efficient model, having the lowest Akaike Information Criterion (AICc) value of 6.275358 amongst all 8 selected and fitted models. The selected model was then used to forecast crude oil prices for 2023.

**4.2.1.1 ARIMA Forecasting**

Having obtained the most appropriate/preferred model, a forecast of Crude oil price for the 2023 year was performed using this model. The predicted crude oil prices were deduced from the model and displayed in Table 4.2 below.

**Table 4.2: 2023 ARIMA-Predicted Crude oil prices**

|  |  |
| --- | --- |
| **Month** | **ARIMA Forecast** |
| January | 78.02410 |
| February | 76.89465 |
| March | 76.60036 |
| April | 76.62508 |
| May | 76.77165 |
| June | 76.96477 |
| July | 77.17566 |
| August | 77.39335 |
| September | 77.61363 |
| October | 77.83490 |
| November | 78.05655 |
| December | 78.27835 |



**Figure 4.13: A plot of 2023 Arima predicted crude oil prices.**

**4.2.1.2 ARIMA Model Accuracy**

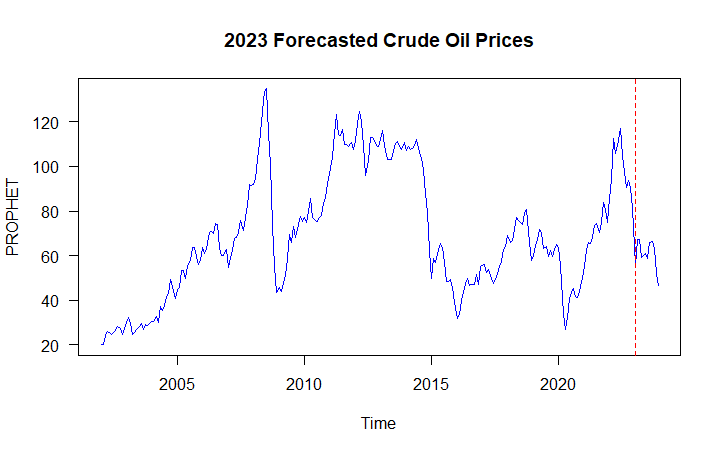
**Table 4.3: Model accuracy for ARIMA**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **MAE** | **RMSE** | **MAPE** | **SMAPE** |
| **ARIIMA Forecast** | 21.448 | 24.32 | 20.67 | 23.66 |

Considering the 2023 ARIMA-predicted crude oil prices in Table 4.2 and its visualized plot in Figure 4.13, it can clearly be deduced that the ARIMA model predicted a downtrend in crude prices taking into consideration the prices of crude oil in previous years. The Model prices of crude oil to slightly increase in the first two quarters of 2023 and a subsequent decline in oil prices in the second half of 2023.

**4.2.2 Prophet Model Building**

In the model building process using Prophet, we started by fitting the model over the entire training and testing dataset after which the forecast object was created. The predicted dataframe contained the dates and predicted crude oil prices for the 2023 financial. This predicted dataframe was extracted from the yhat column of the future predicted model.



**Figure 4.14: A plot of Prophet predicted crude oil prices**

The figure above illustrates the prophet models prediction of prices of crude oil for the year 2023. From the above graph, the prophet model predicted a decline in crude oil prices to approximately 50 at the end of 2023. We can observed that there was up upward trend (increase) in crude oil prices from 2000 through till mid-2021 before a sharp drop in crude oil prices to 2022. Building on this data, the prophet model predicted a further drop in crude oil prices till 2023 as seen in Figure 4.14.

**Table 4.4: 2023 PROPHET-Predicted Crude oil prices**

|  |  |
| --- | --- |
| **Month** | **PROPHET Forecast** |
| January | 55.73 |
| February | 67.42 |
| March | 67.08 |
| April | 59.12 |
| May | 60.07 |
| June | 60.93 |
| July | 58.51 |
| August | 65.87 |
| September | 66.13 |
| October | 63.32 |
| November | 51.76 |
| December | 46.47 |

**4.2.2.1 Prophet Model Accuracy**

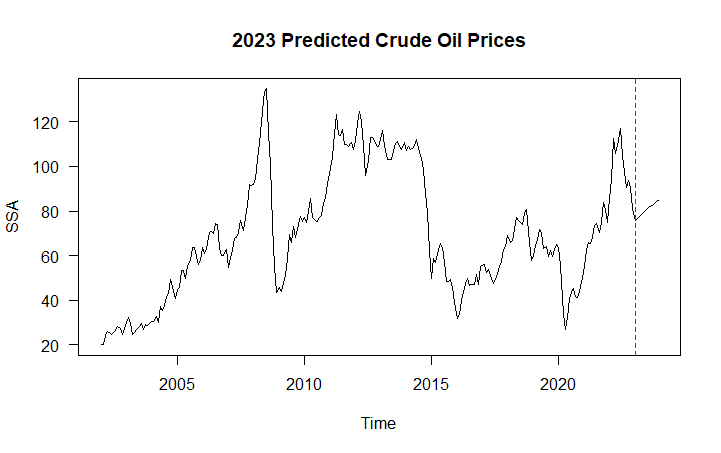
**Table 4.3: Model accuracy for PROPHET**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **MAE** | **RMSE** | **MAPE** | **SMAPE** |
| **ARIIMA Forecast** | 24.70179 | 29.78582 | 40.58452 | 38.06291 |

The accuracy of the model’s prediction was measured by way of the Mean Absolute Error (MAE), and Root Mean Square error (RMSE) together with the Mean Absolute Percentage Errors (MAPE). As shown in Table 4.3, the prophet models MAE of 24.70 and RMSE of 29.78 is slightly higher than that of the ARIMA model which respectively are 21.448 and 24.32. This simply suggests that the ARIMA model in the prediction of crude oil prices is more effective that the Prophet model.

**4.2.3 Singular Spectrum Analysis Model Building**

In forecasting the crude oil prices using Singular Spectrum Analysis, the testing part of the data which comprised from data from 2002 to 2021 was embedded into the insample set. Further, Singular Value Decomposition (SVD) was executed by way of Hankel matrix, and by transforming it to a transpose of its matrix, a trajectory was then created and the eigenvalues and eigenvectors of the trajectory matrix computed. The model then used the training data set to test using data from 2021 to 2022 and further predicted crude oil prices for 2023. The result of the forecast is displayed in the figure below.



**Figure 4.15: A plot of SSA predicted crude oil prices**

**Table 4.5: 2023 SSA-Predicted Crude oil prices**

|  |  |
| --- | --- |
| **Month** | **SSA Forecast** |
| January | 75.98 |
| February | 76.81 |
| March | 77.66 |
| April | 78.51 |
| May | 79.35 |
| June | 80.20 |
| July | 81.03 |
| August | 81.85 |
| September | 82.65 |
| October | 83.43 |
| November | 84.18 |
| December | 84.90 |

Comparing the SSA-model forecast of Crude oil prices for 2023 as against that of the ARIMA model, it can clearly be seen in Figure 4.15 that the SSA model predicted a sharp rise in crude oil prices from 2022 to 2023. This somewhat contradicts the Prophet models prediction of a further continuation in decline or fall of crude oil prices in 2023. The accuracy of the SSA models’ prediction was estimated and displayed in Table 4.4 below.

**4.2.3.1 SSA Model Accuracy**

**Table 4.6: Model accuracy for SSA**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **MAE** | **RMSE** | **MAPE** | **SMAPE** |
| **ARIIMA Forecast** | 18.85091 | 22.0116 | 18.07709 | 20.4454 |

Based on the model accuracy for the SSA forecast as depicted in Table 4.6, we observed that, from all three models adopted for this study, the SSA model has the lowest model accuracy values. That is to say the MAE of 18.85 and RMSE of 22.01 is the least compared to the Prophet model (MAE = 24.70179, RMSE = 29.78582) and that of the ARIMA model (MAE = 21.448, RMSE = 24.32).

This suggests that from all three models adopted for this study, the Singular Spectrum Analysis model had the most effective and efficient model for predicting the Crude oil prices for the 2023 financial year.

**4.3 Discussion**

The main objective of the study was to discover the appropriate time series model which can be used to effectively forecast the crude oil prices for 2023. Three different time series models were adopted to predicting the Crude oil prices, and these models are; the Autoregressive Integrated Moving Average (ARIMA) model, the Prophet model and Singular Spectrum Analysis (SSA). Upon fitting these models and further ascertaining their accuracy through their predictive ability and error margins, the Singular Spectrum Analysis model emerged as the most effective model due to its lower errors. As shown in Table 4.6, the accuracy test of the SSA model as compared to other models revealed that, it had a smaller Mean Absolute Error (MAE = 18.85091), Root Mean Square Error (RMSE = 22.0116), and also a smaller Mean Absolute Percentage Error (MAPE = 18.07709) as compared to both ARIMA and Prophet models. This suggests that the SSA model is the most appropriate model for forecasting crude oil prices and should be adopted and explored by regulatory institutions in Ghana. Furthermore, upon critically studying and understanding the characteristics and patterns of the crude oil prices, investors and institutions alike can take advantage and make informed decisions about the purchase and sale of the crude oil prices.

Similarly, there appears to be some form of similarity between the ARIMA model and the Singular Spectrum Analysis Model. In that, their accuracy diagnostics were almost similar. A careful comparison between the two models (see Figure 4.13 and Figure 4.15) revealed that, both models predicted an upward change/turn in crude oil prices in 2023. This also means that, the ARIMA model predicted crude oil prices better that that of the Prophet model.

For the Prophet model’s forecast of crude oil prices, as shown in Table 4.3 and in Figure 4.14, it can be observed that, the prophet-model also predicted a fall in crude prices from 2022 to 2023, however there seem to be a huge disparity between the ARIMA, SSA models against the Prophet model. As these other two models predicts an upward move in prices, the Prophet model predicted a downward fall in crude oil prices for the same period.

**CHAPTER 5**

**Conclusion and Recommendation**

The main purpose of this study was to build a time series model effective and efficient to predicts crude oil prices for the year 2023 in Ghana. Three (3) different time series models were used specifically for purpose of this study. These models include the Singular Spectrum Analysis (SSA) model, the Autoregressive Integrated Moving Average (ARIMA) model and the Prophet model. This study identified the SSA model to be the most appropriate and efficient model for the prediction 2023 crude oil prices in Ghana. Based on the accuracy of the predictive model for all the three models (i.e., MAE, MAPE, SMAPE, etc), the SSA model had the lowest model errors. This led to the selection of the SSA model as the most suitable model. The researcher therefore recommends the Singular Spectrum Analysis model to regulators, oil companies and individuals alike to adopt the model in forecasting prices so as to make informed decisions.

The use of Singular Spectrum Analysis in forecasting Crude oil prices can to a large extent improve upon the accuracy of predictions and speculations made by some organizations and further assist decision-making bodies like the Government to strategize and build strategic and long-term plans when dealing with the prices of crude oil and its related commodity. Subsequently, the use and well-development of this SSA model can help generate accurate forecast of the commodity allowing for a better anticipation of future patterns and trend in the prices.