Forecasting of Monthly Crude Oil Prices in Kenya Using Comparative Time Series Models

Abstract —

Crude oil is one of the most vital products that ever existed and variation in prices affects all sectors of the economy and variation in its prices is very crucial. Therefore, without an accurate and appropriate predictive model for crude oil prices, it has proven difficult to predict future oil prices. Therefore, appropriate modeling is crucial for the oil companies to adjust strategies used in the production and supply as well as structural optimization.This study sought to fit several time series forecasting models namely; ARIMA, Naïve, Seasonal Naïve, Time Series Linear Model (TSLM),and Exponential Smoothing (ETS) to model and forecast crude oil prices. The study used Kenya's monthly crude oil prices from Jan 2003 to Dec 2023, giving a sample of 252 observations. The selection of the best model was based on the minimization of the information criteria, where ARIMA, ETS, and Seasonal Naïve attained an AIC of 1294.193, 1846.780, and 1403.821, respectively. Similarly, the models attained the BIC of 1304.991, 1863.071, and 1450.489, respectively. Since the Naïve and TSLM could not provide the AIC and BIC, model selection was entirely based on the forecast accuracy measures. From the results, the Naive model reported the lowest RMSE (19.6329), indicating that it has the smallest average squared error, with the lowest MAE (15.5212), suggesting it has the smallest average absolute error. Besides, the Naive model reported an ME (1.9819) which is relatively low but not the lowest. The automatic ETS model has a slightly lower ME (3.478211). The naive model reported lower MPE and MAPE values (-3.9525 and 26.8022, respectively) compared to most models, indicating less percentage error. Similarly, a lower MASE and RMSSE were reported by the naïve model, (0.7343624) and RMSSE (0.7228792), respectively, indicating that the model performs well relative to forecasting crude oil prices in Kenya. The naïve model demonstrated a higher consistency and reliability in forecasting crude oil prices in Kenya.

***Keywords*—**Crude oil, Time Series, Forecasting, Stationarity, Oil prices

1. Introduction

Oil is the main source of energy for most industrial processes [1] and it has maintained its significance in the industrial and manufacturing replacing the traditional source of energy, coal. Even though other sources of energy were discovered and other sources of energy are coming up, oil has remained the most significant source of energy globally. Globally, oil and oil products are used directly or indirectly and thus the variation in oil price affects nearly every commodity [2]. As a result, governments and policymakers are more concerned with the variation in oil prices. It is worth noting that these variations many times are caused and influenced by the global increase in demand for oil as well as geopolitical tensions such as Russian Russian-Ukrainian war [3]. During the Russian-Ukrainian war, there was a 70.72% and 73.62% fluctuation in Brent Crude Oil andWest Texas Intermediate (WTI) oil prices, respectively. Due to these variations in oil prices, an appropriate forecasting mechanism is required to model and predict oil prices for an appropriate cop-up mechanism. Developing a forecasting model that can appropriately predict oil prices is a priority for a government that cares about fluctuations in commodity prices and the general standard of living.

Oil and its products are vital products whose variations affect technically every commodity, and thus, developing a model that accurately, easily, and quickly predicts the prices of a commodity is vital for the government [4]. Traditionally and for the longest time, researchers have been relying on ARIMA and SARIMA, the classical models for time series analysis, with little or no consideration to other time series models such as Time Series Linear Model (TSLM), Exponential Smoothing (ETS), Naïve, Seasonal Naïve, and average model. In a study to model crude oil prices using SARIMA, Ariyanti, and Yusnitasari notes that oil prices and economic health are intertwined. Since economic growth expands and contracts in cycles, the swing in crude oil prices can significantly impact this pattern. Predicting future oil prices is crucial for government agencies, oil companies, and policymakers. An ideal model should be quick, easy to use, and accurate, allowing policymakers and governments to make informed decisions. In their study, Ariyanti and Yusnitasari explored ARIMA and SARIMA, a time series predictive algorithm, for crude oil price prediction. Daily data from Yahoo Finance, spanning January 27th, 2020 to January

25th, 2023, was used to build the models. Both ARIMA and SARIMA achieved an evaluation score (RMSE) of 1.905. The 7-day forecast results were very similar, with ARIMA predicting a price of $86.23 and SARIMA predicting $86.26.

The rising complexity and volatility of global crude oil prices pose a growing threat to economic stability, and therefore, accurate oil price forecasts can help mitigate these risks [5]. In their study, the authors analyzed the impact of both market and non-market factors on oil prices. The study then utilizes a Vector Autoregression (VAR) model to quantify these relationships. Building upon the VAR model's insights, the paper proposes a novel VAR-SVM model that combines VAR with Support Vector Machines (SVM) for enhanced prediction accuracy. Additionally, a Genetic Algorithm (GA) was employed to optimize the model's parameters. Empirical results demonstrate the VAR-SVM model's superiority in accuracy and effectiveness compared to traditional methods like VAR, CGARCH, Artificial Neural Networks (ANN), and autoregression SVM.

In their studies on forecasting crude oil prices by using the ARIMA Model, evidence from Tanzania [6] aimed at predicting oil prices during unusual circumstances. The authors applied the Box-Jenkins methodology in developing the ARIMA for forecasting crude oil prices, where the selection of the best model was based on AIC and BIC values. The results showed that among the three ARIMA models, ARIMA (0, 1, 0), ARIMA (0, 1, 1) and ARIMA (1, 1, 0) estimated, ARIMA (0, 1, 1) emerged the least AIC and BIC values of -460.94 and -453.95, respectively. In a relatively similar study, [7] compared the performance of ARIMA and SARIMA in modeling and forecasting crude oil prices. The results of the study indicated that ARIMA and SARIMA performed relatively well with both having the RMSE value of 0.0195% and with seven days of 86.230003 and 86.260002 for ARIMA and SARIMA. The researcher was left with the option to pick either of the prediction models. The available literature leaves a gap to investigate a comparative analysis of various time series models shifting the focus from conventional models such as ARIMA and SARIMA.

On the other hand, [8] investigated the effectiveness of various ARIMA and GARCH models in forecasting Nigerian crude oil prices. The analysis utilizes 189 monthly data points from January 1998 to September 2013.In their study, the data's stationarity was first assessed using visual inspection of autocorrelation (ACF) and partial autocorrelation (PACF) plots, followed by statistical tests like KPSS and ADF. The results indicated non-stationarity in the raw data. Differencing the logarithms of the data series achieves stationarity, confirmed by similar plots and tests.Besides, the optimal ARIMA and GARCH models were chosen based on information criteria like AIC, HQC, and SIC. The model with the lowest criteria value is considered the best. Consequently, ARIMA (3,1,1) and GARCH (2,1) were identified as the most suitable models for forecasting Nigerian crude oil prices. However, the study did not consider other benchmark models like ETS, TSLM, and naïve among other models, creating a gap in modeling and forecasting crude oil prices.

The variation in crude oil prices is time-dependent [9]. In their study,the authors investigated the effectiveness of Time-Varying Vector Autoregression (TVP-VAR) models in forecasting real crude oil prices specifically exploring model averaging and selection techniques across multiple TVP-VAR models. These approaches address the challenge of fluctuating uncertainty in oil price determinants and the evolving strength of their relationships. The study builds upon the success of Dynamic Model Averaging and extends it for TVP-VAR models and incorporates geopolitical risk as an internal variable within each model. The model combination scheme was designed to capture the joint forecasting ability for both real crude oil prices and geopolitical risk.The findings demonstrate that the Vector Autoregression approach using model averaging or selection outperforms single-equation Time-Varying Regression and standard Dynamic Model Averaging. Additionally, combining multiple Vector Autoregression models yields superior results compared to a single model. To assess forecast accuracy, the study employs novel techniques like Giacomini-Rossi fluctuation tests and Murphy diagrams, which capture the significance of time-varying predictive ability along with various scoring functions.

The government of Kenya in early 2023 signed a government-to-government kind of deal famously known as the g-to-g oil deal, which presumably meant to curb the skyrocketing dollar against the shilling, with the ultimate goal of lowering crude oil prices [10]. According to the government report, the importation of petroleum products from government to government was initiated by the government of Kenya to help avoid the economic shutdown caused by supply constraints due to the US dollar liquidity problem. Even though g-to-g oil was signed between the government of Kenya and Saudi Arabia, to bring about macroeconomic stability, the deal did not deliver its mandated and intended purpose and later failed due to the distortion of the currency market [11]. The government's fiscal policy of the g-to-g oil deal failed to contain the fluctuation of crude oil prices, which formed part of the deal's ultimate objective. In other words, without an accurate and appropriate predictive model for crude oil prices, it has proven difficult to predict future oil prices. Therefore, appropriate modeling is crucial for the oil companies to adjust strategies used in the production and supply as well as structural optimization [4].

ARIMA and SARIMA in many cases are preferred models in forecasting future values of a series due to their ability to capture seasonality and trend in the data [12]. However, it is important to consider the general structure of the data when coming up with the appropriate prediction model. For instance, the naïve model tends to be generally better than ARIMA and SARIMA when the time series data follows a random walk without any clear trend or seasonal patterns[13]. Besides, when the data shows strong and stable seasonal patterns, seasonal naïve tends to perform better than SARIMA. That is, when the time series exhibits strong and stable seasonal patterns, the seasonal naive model can provide accurate forecasts. Besides, in cases where the seasonal component of the series does not change over time, the seasonal naïve model can outperform more complex models that attempt to capture unnecessary dynamics. Lastly, when a times series has changing trends and seasonality, components that may not be well captured by ARIMA and SARIMA, an automatic exponential smoothing tends to outperform the traditionally considered best models like ARIMA and SARIMA. Therefore, this study sought to perform a comparative analysis of times models with average, and automatic ETS as the benchmark models. The study compared automatic ARIMA, automatic ETS, Naïve and Seasonal Naïve, Average model, and Time Series Linear Model (TSLM) with trends and seasonality components included. The overall best model was selected based on various performance metrics after various types of cross-validation applied to the model and test set, a concept of machine learning. The following are the objectives guiding this study.

I. To fit various time series models to crude oil price data.
II. To evaluate the predictive accuracy of the fitted times series models
III. To forecast Kenya crude oil prices data using the overall best-fitted model.

The study focused on modeling monthly crude oil prices in Kenya from Jan 2003 to Dec 2023, using several time series models including ARIMA/SARIMA, Naïve, Seasonal Naïve, Exponential Smoothing, and a combination of the four models. The main objective and the focus of the study was to evaluate the time series models estimated and pick the overall best model for forecasting monthly crude oil prices in Kenya. However, like any other study, the study faces one main limitation regarding the data. The study used monthly crude oil price data which did not capture intra-month crude oil price volatility. Secondly, the study focused on Kenya and the results may not be generalized to countries that do not have similar economic structures and oil pricing mechanisms.

2. METHODS AND MATERIALS

2.1. Research Design

The research design employed in this study is descriptive. Within this descriptive research design, the primary objective is to investigate and elucidate the inherent characteristics of time series data [14].

2.2. Data Collection

Secondary data was extracted from Kenya National Bureau of Statistics (KNBS)[15] available on their website from January 2003 to December 2023. The monthly data was downloaded in Excel format with both the date series and a series containing the monthly crude oil prices in Kenyawith 252 observations.

2.3. Data Analysis

Data analysis was done using R -Statistical Software [16]. The preliminary stage involved checking for stationarity of the data. The data exploration was done by examining the visual plot to check for any underlying patterns of behaviors in the series such as trend or seasonality.The stationarity of the data was tested using the Augmented Dickey-Fuller (ADF) [17]. If the ADF statistic is greater than the critical value at a 5% significance level, then the null hypothesis would be accepted, and the series thus rendered to be non-stationary. If the series is found to be non-stationary, differencing at different degrees is done until a stationary series of a given degree of differencing is obtained.

2.4. Models Fitting

2.4.1. SARIMA Process

Let B^S denote the operator such that

$B^S X_t = (X_t - X_{t-s})$ [1]

Then the seasonal differencing is written as,

$(1 - B^S)X_t = (X_t - X_{t-s})$ [2]

Given that our data is monthly with 12 months per year (s=12) the seasonal difference is obtained as;

$(1 - B^{12})X_t = (X_t - X_{t-12})$ [3]

The SARIMA model with non-seasonal terms of order (p, d, q) and seasonal terms of order (P, D, Q) is abbreviated as $\text{SARIMA}(p, d, q)(P, D, Q)_S$ model and may be written as,

$\varphi(B)\Phi(1 - B^S)^D X_t = \theta(B)\Theta(B^S)Z_t$ [4]

In their study, [18] confirmed that the multiplicative seasonal autoregressive integrated moving average model, or SARIMA model, of Box and Jenkins (1970) is given by;

3. $\varphi_P(B^S)\Phi(B){\nabla_S}^D\nabla^d x_t = \alpha + \Theta_Q(B^S)\theta(B)Z_t$ [5]

The general model is denoted as ARIMA (p, d, q)(P, D, Q)$_{12}$. The ordinary autoregressive and moving average components are represented by polynomials $\Phi(B)$ and $\theta(B)$ of orders p and q, respectively, and the seasonal autoregressive and moving average components by $\varphi_P(B^S)$ and $\Theta_Q(B^S)$ of orders P and Q and ordinary and seasonal difference components by:

$\nabla^d = (1 - B)^D$ and ${\nabla_S}^D = (1 - B^S)^D$ [6]

Since the data of interest in this study is monthly data, s = 12, hence our equation becomes,

$(1 - B)(1 - B^{12})X_t = (1 + \theta B)\Theta(B^{12})Z_t,$ [7]

2.4.2. Naïve Process

The mathematical expression for the naïve process for time series is given as shown in equation 8;

$\hat{Y}_{(T+1|T)} = \widehat{Y_T}$

[8]

Where:

$\hat{Y}_{(T+1)}$ is the forecasted value for the next period (time T+1).

$\widehat{Y_T}$ is the actual value at the current period (time T).

Naïve process is a simple approach to model a times series data. In this process, the previous value acts as the present value. Naïve method is recommended for model comparison by [19] and, [20]. The method is considered the best when benchmarking is needed.

2.4.3. Seasonal Naïve Process

The seasonal naïve is an extension of the naïve process. The seasonal naïve process assumes that the forecasts are equal to the last value from the same season. Consider the mathematical expression shown in Equation 9

$\hat{y}_{(T+h|T)} = y_{T+h-m(k+1)}$ [9]

where m = seasonal periodand $k$is the integer part of $(h - 1)/m$.

2.4.4. Times Series Linear Model (TSLM)

The TSLM approach is a linear technique used to forecast a times series data incorporating trends and seasonality present in the data[21]. The mathematical expression is given in Equation 10

$$\hat{Y}_T = \beta_0 + \beta_{1T} + \beta_2 D_{Jan} \beta_2 D_{Feb} + \cdots + \beta_{12} D_{Dec} + \epsilon_T \quad [10]$$

Where;

β_0 is the baseline level of crude oil prices

β_{1T}captures the time trends present in the crude oil prices

$\beta_1, \beta_2, , , \beta_{12}$ are the seasonal effects of each month

2.4.5. Exponential Smoothing (ETS)

The exponential smoothing process was developed between the 1950s and 1960s,[22]. The primary goal of this model was to produce a point forecast. The process is mathematically expressed in three different forms, additive form, multiplicative and combined form as expressed in equations 11, 12, and 13, respectively.

$$Additive\ ETS = y_t = l_{t-1} + b_{t-1} + s_{t-m} + \epsilon_t \quad [11]$$

$$Multiplicative\ ETS = y_t = l_{t-1} \times b_{t-1} \times s_{t-m}(1 - \epsilon_t) \quad [12]$$

$$Combined\ ETS = (l_{t-1} + b_{t-1})(s_{t-m} + \epsilon_t) \quad [13]$$

In the three processes above, the error term has a mean of zero (0) and a variance of σ^2 expressed in equation 14;

$$\epsilon_t \sim NID(0, \sigma^2) \quad [14]$$

2.5. Model Evaluation and Selection

2.5.1. Model Selection

The selection of the modelwasbased on the minimization of the Akaike Information Criterion [23], Schwarz Criterion [21], and Hannan-Quinn Criterion [24]. On the other hand, the performance of the best model was evaluated using criteria such as the Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE), Akaike's Information Criteria (AIC), Bayesian Information Criteria (BIC).Akaike (1974) and Schwarz (1978) state that the AIC function is denoted by

$$AIC = -2\log(L) + 2\log(p + q) \quad [15]$$

BIC is obtained as shown in equation 16 as shown in Gideon E. Schwarz (1978)

$$BIC = -2\log(L) + k\log(n) \quad [16]$$

2.5.2. Model Evaluation

During the valuation of the model's forecast accuracy,[25]used the following formulas to calculate the RMSE, MAE, and MAPE.

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}(\hat{y}_t - y_t)^2}{n}} = \sqrt{\frac{\sum_{t=1}^{n} e_t{}^2}{n}} \quad [17]$$

$$MAE = \frac{\sum_{t=1}^{n}|Y_t - \hat{Y}_t|}{n} \quad [18]$$

$$MAPE = \left\{\frac{\sum_{t=1}^{n}|Y_t - \hat{Y}_t|}{n}\right\} \times 100 \quad [19]$$

2.5.3. Residual Diagnostic

The evaluation of models used in forecasting is an important step in the process of time series modeling. Ensuring that a model effectively captures the underlying patterns and dynamics in time series data is paramount for making informed decisions and predictions [26]. Ideally, the residuals should exhibit a random scatter around zero, devoid of any discernible patterns. Departures from this randomness can indicate model deficiencies. Another critical aspect of residual analysis is assessing the normality assumption. For normally distributed residuals, the density curve in a histogram should take on a bell-shaped appearance. Deviations from the bell-shaped curve may suggest deviations from normality, which could impact the model's validity.

2.5.4. Forecasting

Forecasting was done on the final overall best model. The accuracy of the forecast was evaluated using measures and criteria of evaluating the accuracy of the forecast which include but are not limited to RMSE, MAPE, and MAE.

3. RESULTS AND DISCUSSION

3.1. Trend Analysis

From Figure 1, the time series plot of crude oil prices from 2003 to 2023 illustrates significant trends and fluctuations driven by global economic conditions, geopolitical events, and market dynamics. From the plot, there is a rise in crude oil prices between 2003 and 2008 due to the increasing global demand for crude oil. The global financial crisis of 2008 disrupted the market and led to a drop in crude oil prices. However, the post-crisis, which occurred between 2009 and 2014 led to a rise in crude oil prices following an increase in global demand. The impact of the COVID-19 pandemic is seen in the plot as indicated by a decrease in crude oil prices following the disrupted market demand and supply, however, the post-covid-19 pandemic was following an increase in crude oil prices due to the rising global demand for crude oil [27].

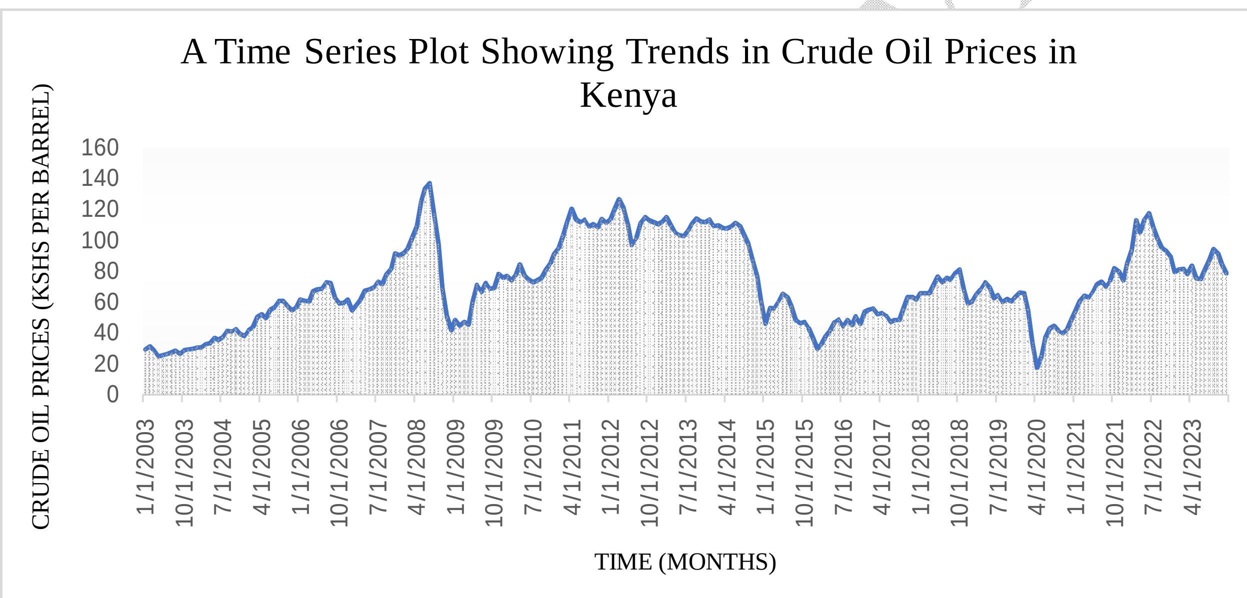

Figure 1: A Plot of Crude Oil Prices Over Time

3.2. Stationarity Test

For many years, non-stationary time series data has faced a lot of criticism for being unfit for modeling due to its susceptibility to yield misleading results. In this paper, stationarity was tested at level and first difference using the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test, and the results are presented below in Table 1.

Table 1: KPSS Stationarity Test

KPSS Test at Level		KPSS Test at First Difference	
KPSS Stat	KPSS P-value	KPSS Stat	KPSS P-value
0.4906358	0.04377571	0.0199052	0.1

The results indicate that the crude oil prices data was non-stationary at level and later became stationary after taking the first difference. The seasonal difference was not necessary in the study to make the series stationary. The plot of the stationary series is shown in Figure 2 below.

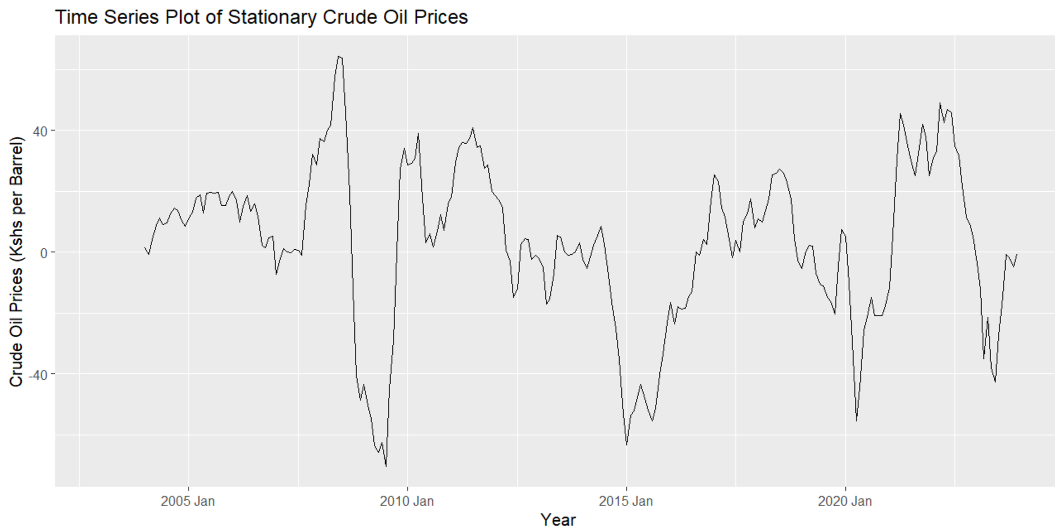

Figure 2: Stationary Time Series Plot of Crude Oil Prices in Kenya

3.3. Descriptive Statistics

The results in Table 2 show the descriptive statistics for the monthly crude oil prices in Kenya between 2003 and 2023. The skewness statistics indicate that the data do not seem to follow a normal distribution as indicated by the Jerque Bera test statistic with a p-value of less than 0.05.

Table 2: Descriptive Statistics

Descriptive Statistics									Normality test	
Variable	N	Mean	Min	Max	Range	Kurtosis	SE	Skewness	Jerque Bera	P-value
Crude Oil Prices	252	71.9600	17.66	137.35	119.69	-0.8573	1.6981	0.2661	10.42	0.0055

3.4. Time Series Models Estimation and Evaluation

Table 3: Model Selection Metrics

Models	**AIC**	**AICc**	**BIC**
Automatic ARIMA	1294.193	1294.333	1304.991
Exponential Smoothing	1846.780	1847.077	1863.072
Naive	N/A	N/A	N/A
Seasonal Naive	1403.821	1406.018	1450.489
Time Series Linear Model (TSLM)	N/A	N/A	N/A

The results in Table 3 show the metric for model selection. In this study, model selection entirely relied on measures of forecast accuracy since some models like Naïve, and TSLM could not provide AIC and BIC values.

3.5. Model Evaluation for the Forecast Accuracy

The selection of the best model was subject to the forecast evaluation. The best-performing model was evaluated based onminimizing forecast accuracy measures reported in Table 4.

Table 4: Forecast Accuracy Measures

Models	Sample	ME	RMSE	MAE	MPE	MAPE	MASE	RMSSE
Automatic ARIMA	Test	1.8279	20.0051	15.5733	-3.5606	26.8635	0.7604	0.7539
Automatic ETS	Test	0.0676	23.6186	16.6694	-5.5308	29.3969	0.8139	0.8900
Combination	Test	2.5925	20.4426	16.2180	-3.9649	28.0825	0.7919	0.7704
Naive	Test	1.9819	19.6329	15.5212	-3.9525	26.8022	0.7579	0.7398
Seasonal Naive	Test	5.8819	26.1602	21.9897	-2.8032	36.8029	1.0737	0.9858

From the results in Table 4, the Naive model has the lowest RMSE (19.6329), indicating it has the smallest average squared error, with thelowest MAE (15.5212), suggesting it has the smallest average absolute error. Besides, the Naive model reported an ME (1.9819) which is relatively low but not the lowest. The automatic ETS model has a slightly lower ME (0.0676).The naive model reported lower MPE and MAPE values (-3.9525 and 26.8022, respectively) compared to most models, indicating less percentage error.Similarly, a lower MASE and RMSSE were reported by the naïve model, (0.7343624) and RMSSE (0.7228792), respectively, indicating that the model performs well relative to forecasting crude oil prices. On the other, seasonal naïve reported a higher ME of 5.8819 and a higher RMSE of 26.1602. Similarly, higher values of MPE, MAPE, MASE, and RMSEE were reported by seasonal naïve, -2.8032, 36.8029, 1.0737, and 0.9858, respectively. These results indicate that seasonal naïve is the least-performing model, with naïve as the best-performing model. Therefore, the study recommended the naive model in predicting crude oil prices for its consistent and strong performance across multiple evaluation criteria. The model is given by the equation 20 below.

$$\hat{Y}_{(T+1|T)} = 79\hat{Y}_T$$
(20)

In a typical Naive time series model, there is no traditional coefficient like one would have in a regression model because the forecast is solely based on the most recent observed value. Equation 20 shows the Naïve times series model with a coefficient of 79 showing the most recent crude oil price as of 31st December of 2023.

3.6. Forecasting

Forecasting was done using the Naïve model which attained a higher forecast accuracy as compared to all other candidate models.Besides, the Naïve model outperformed the combined model which came third in a scale of best-performing models. The forecast from the Naïve model is given in Table 5 below.

Table 5:Twelve Months Forecast of Crude Oil Prices

Model	Months	95% Confidence Interval	Forecasted Crude Oil Price
naive	2024 Jan	N(39, 79)	79
naive	2024 Feb	N(70, 90)	82
naive	2024 Mar	N(79, 118)	90
naive	2024 Apr	N(79, 157)	79
naive	2024 May	N(55, 197)	68
naive	2024 Jun	N(69, 236)	79
naive	2024 Jul	N(79, 275)	89
naive	2024 Aug	N(79, 314)	90
naive	2024 Sep	N(79, 354)	110
naive	2024 Oct	N(79, 393)	96
naive	2024 Nov	N(79, 432)	100

naive	2024 Dec	N(79, 472)	120

Graphically, the forecasted crude oil prices above can be visualized as shown in Figure 3 below.

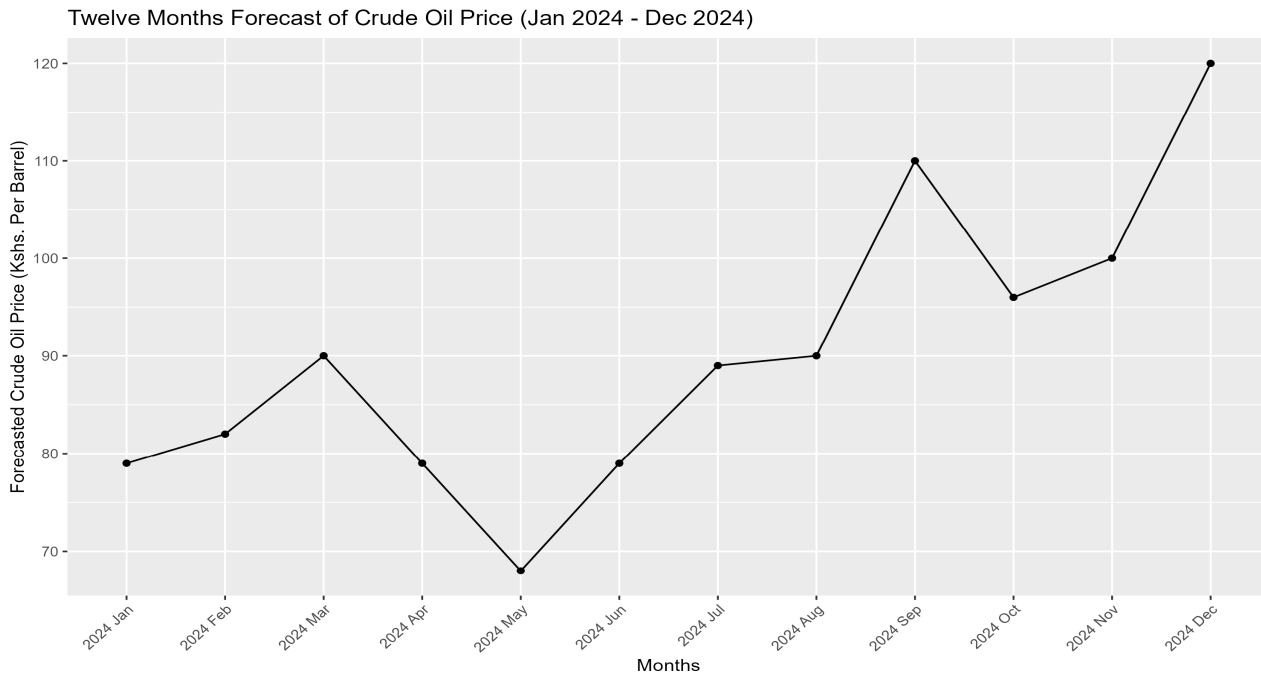

Figure 3: A Plot of Twelve Months Forecasts

Figure 4 shows the 12-monthdistribution of the forecast and the mean crude oil forecast (2024 Jan – 2024 Dec) using the Naïve model. From the plot, the study found that the stability and certainty of the forecast decreased with an increased forecast horizon as shown in Figure 4.

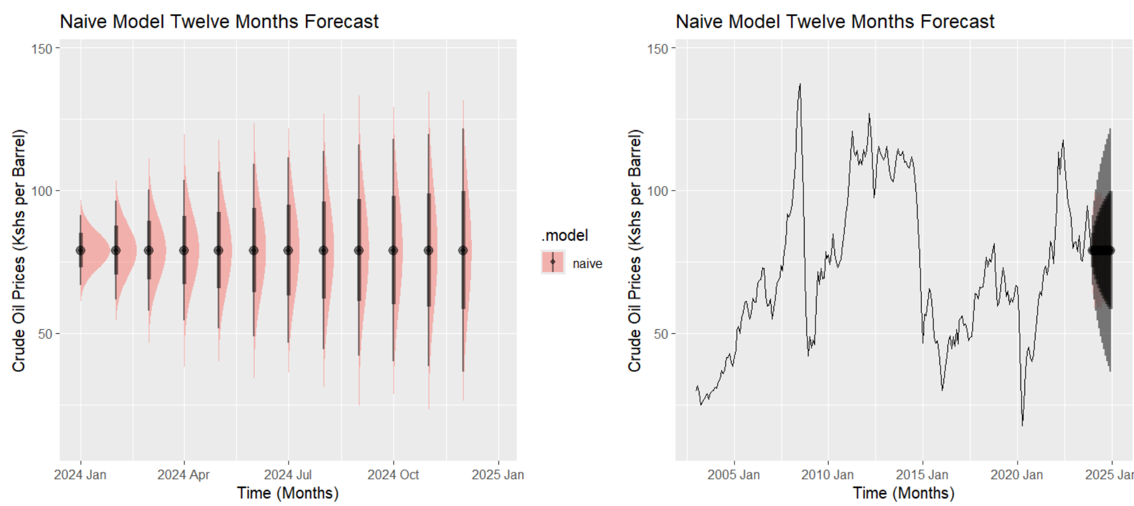

Figure 4: Distribution ofForecast for Crudes Oil Prices in Kenya

4. CONCLUSION

Oil is the main source of energy for most industrial processes and it has maintained its significance in the industrial and manufacturing after replacing the traditional source of energy, coal. However, the rising complexity and volatility of global crude oil prices pose a growing threat to economic stability [5]. Therefore, accurate oil price forecasts can help mitigate these risks. The goal of any forecasting problem is to obtain the model that can best capture the variation in a time series data to be able to forecast future values. This study sought to perform a comparative analysis of times models with TSLM, and automatic ETS as the benchmark models. The study compared automatic ARIMA, automatic ETS, Naïve, and Seasonal Naïve models. The overall best model was selected based on various performance metrics after cross-validation was applied to the models using the test set, a concept of machine learning. While the automatic ARIMA model has slightly lower ME and MAE, the Naïve model'sperformance across multiple metrics, especially its lower RMSE and MAPE, suggests it has a more accurate and reliable forecasting ability overall. The study made a twelve-month forecast (2024 Jan – 2024 Dec) as shown in Figure 3. The study found that increasing the forecast horizon makes the forecast unstable and uncertain, and therefore, recommends a short-term forecast horizon. The results showed that the absence of seasonality in the data made the naïve model outperform the conventional time series model (ARIMA/SARIMA).

ACKNOWLEDGMENT

We would like to express our sincere gratitude to the Kenya National Bureau of Statistics (KNBS) for providing recent relevant and accurate data to help address the research objectives.

FUNDING

The research conducted for this study did not receive any external funding.

CONFLICT OF INTEREST

There were no conflicts of interest associated with this research.

DISCLAIMER ON ARTIFICIAL INTELLIGENCE

Authors hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc) and text-to-image generators have been used during the writing or editing of manuscripts.

References

1. Kılıç, N. Ö., & Öztürk, S. (2020, June). *SOURCES OF ENERGY AND THE SIGNIFICANCE OF OIL*. ResearchGate; unknown. https://www.researchgate.net/publication/348265348_SOURCES_OF_ENERGY_AND_THE_SIGNIFICANCE_OF_OIL
2. Wang, Y. (2023). Oil Price Forecasting based on Improved SARIMA Model. *Highlights in Science, Engineering and Technology*, *39*, 238–245. https://doi.org/10.54097/hset.v39i.6535
3. Zhang, Q., Hu, Y., Jiao, J., & Wang, S. (2024). The impact of Russia–Ukraine war on crude oil prices: an EMC framework. *Humanities & Social Sciences Communications*, *11*(1). https://doi.org/10.1057/s41599-023-02526-9
4. Ariyanti, V. P., &Yusnitasari, N. T. (2023). *Comparison of ARIMA and SARIMA for Forecasting Crude Oil Prices*. *7*(2), 405–413. https://doi.org/10.29207/resti.v7i2.4895
5. Zhao, L.-T., Cheng, L., Wan, Y., Zhang, H., & Zhang, Z. (2015). A VAR-SVM model for crude oil price forecasting. *International Journal of Global Energy Issues*, *38*(1/2/3), 126–126. https://doi.org/10.1504/ijgei.2015.069485
6. Gasper, L., &Mbwambo, H. (2023). Forecasting Crude Oil Prices By Using ARIMA Model: Evidence From Tanzania Journal of Accounting, Finance and Auditing Studies 9/2 (2023): 158-175. *Journal of Accounting, Finance and Auditing Studies*. https://doi.org/10.32602/jafas.2023.017
7. Suleiman, S., Alabi, M. A., Usman, U., & Muhammad Adamu Umar. (2023, August 4). *Modeling and Forecasting the Crude Oil Price in Nigeria*. ResearchGate; unknown. https://www.researchgate.net/publication/372899162_Modeling_and_Forecasting_the_Crude_Oil_Price_in_Nigeria
8. Ariyanti, V. P., &Yusnitasari, N. T. (2023). *Comparison of ARIMA and SARIMA for Forecasting Crude Oil Prices*. *7*(2), 405–413. https://doi.org/10.29207/resti.v7i2.4895
9. Drachal, K. (2021). Forecasting crude oil real prices with averaging time-varying VAR models. *Resources*

Policy, *74*, 102244. https://doi.org/10.1016/j.resourpol.2021.102244
10. The National Treasury and Economic Planning-Kenya. (2023). *Press Release on the Importation of Petroleum Products Through a Government to Government Deal*. The Republic of Kenya.
11. MBURU, P. (2024, January 20). *How Kenya oil deal is distorting currency market*. The East African; The East African. https://www.theeastafrican.co.ke/tea/business/how-kenya-oil-deal-is-distorting-currency-market-4497914
12. Valipour, M. (2015). Long-term runoff study using SARIMA and ARIMA models in the United States. *Meteorological Applications*, *22*(3), 592–598. https://doi.org/10.1002/met.1491
13. Iqbal, M., & Naveed, A. (2016). Forecasting Inflation: Autoregressive Integrated Moving Average Model. European Scientific Journal, ESJ, 12(1), 83. https://doi.org/10.19044/esj.2016.v12n1p83
14. Aggarwal, R., & Ranganathan, P. (2019). Study designs: Part 2 – Descriptive Studies. Perspectives in Clinical Research, 10(1), 34–36. NCBI. https://doi.org/10.4103/picr.PICR_154_18
15. KNBS. (2021, December 23). *Kenya National Bureau of Statistics - Kenya's Top Data Site*. Kenya National Bureau of Statistics. https://www.knbs.or.ke/
16. R Core Team. (2022). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. https://www.r-project.org/
17. Dickey, D., & Fuller, W. A. (1979). "Distribution of the Estimators for Autoregressive Time Series with a Unit Root", Journal of the American Statistical Association, 74 (366), 427-431.
18. Shumway, R. H., & Stoffer, D. S. (2017). *Time series analysis and its applications : with R examples*. Springer.
19. Makridakis, S., Wheelwritht, S.C., & Hyndman, R. J. (1998). Forecasting methods and application. (3rd ed.). John Wiley & Sons, Inc.
20. Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting : Principles and Practice (2nd ed.). Otexts. https://otexts.com/fpp2/
21. Hyndman, R.J., & Athanasopoulos, G. (2016). Forecasting: Principles and practice, a online text book. Retrieved from https://www.otexts.org/fpp/2/3 Accessed on 12.11.2016
22. Akaike, H. (1977). On entropy maximization principle, Applications of Statistics. Krishnaiah PR.
23. Schwarz, G. (1978). Estimating the dimension of a model. The annals of statistics, 6(2), 461-464.
24. Hannan, E. J., & Quinn, B. G. (1979). The determination of the order of an autoregression. Journal of the Royal Statistical Society: Series B (Methodological), 41(2), 190-195
25. Wanjuki, T. M., Wagala, A., & Muriithi, D. K. (2022). Evaluating the Predictive Ability of Seasonal Autoregressive Integrated Moving Average (SARIMA) Models using Food and Beverages Price Index in Kenya. European Journal of Mathematics and Statistics, 3(2), 28–38. https://doi.org/10.24018/ejmath.2022.3.2.100
26. Wanjuki, T. M., Wagala, A., & Muriithi, D. K. (2021). Forecasting Commodity Price Index of Food and Beverages in Kenya Using Seasonal Autoregressive Integrated Moving Average (SARIMA) Models. European Journal of Mathematics and Statistics, 2(6), 50–63. https://doi.org/10.24018/ejmath.2021.2.6.80
27. Gharib, C., Mefteh-Wali, S., Serret, V., & Ben Jabeur, S. (2021). Impact of COVID-19 pandemic on crude oil prices: Evidence from Econophysics approach. Resources Policy, 74, 102392. https://doi.org/10.1016/j.resourpol.2021.102392