Kelvin kiplagat kigen

maasai mara university

**APPLICATION OF AUTO-REGRESSIVE INTERGRATED MOVING AVERAGE MODEL TO FORECAST PETROLEUM PRICES IN KENYA.**

A RESEARCH SUBMITTED AS A PARTIAL REQUIREMENT FOR THE BACHELORS OF SCIENCE DEGREE IN APPLIED STATISTICS WITH COMPUTING IN THE SCHOOL OF PURE, HEALTH AND APPLIED SCIENCES

April 2023

**DECLARATION**

I declare that this research is my original work and has never been presented or submitted for examination, either in part or as a whole, in any other university.

**Signed**: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ **Date**: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Name**: Kelvin Kiplagat kigen

**Registration Number**: SB07/SR/MN/7759/2018

**Signed**: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ **Date**: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Supervisor:** Mr Otieno Kevin

**ACKNOWLEDGEMENT**

I am deeply grateful to my supervisor, Mr. Otieno Kevin, for providing invaluable guidance and support throughout the preparation of this project. His unwavering patience, scholarly expertise, and insightful feedback were instrumental in shaping the development of this project. I would also like to express my appreciation to Maasai Mara University for providing me with the opportunity to pursue a Bachelor of Science degree in Applied Statistics with Computing without forgetting my classmates and friends whose positive encouragement and motivation kept me focused. This program has not only broadened my knowledge and skills but also given me a new perspective on various issues and situations in life. I am thankful for the resources and facilities that the university has provided, which have contributed to the successful completion of this project. I also want to say a big thank for my lecturers

**DEDICATION**

.

I dedicate this project to my parents, whose unwavering support and dedication to my education have been a source of inspiration throughout my life journey. They have instilled in me a strong work ethic, encouraged me to pursue my dreams, and strengthened my faith in the future. I am grateful for their sacrifices, guidance, and love, which have laid the foundation for my academic and personal growth

# **ABSTRACT**

The modelling and prediction of petroleum prices is an important factor for economic growth in developing countries such as Kenya. It is crucial to understand the current and future prices of petroleum as it can help users make informed decisions about buying and selling patterns. Additionally, it assists the government in implementing appropriate policies to maintain stable prices. The study utilized the Auto-regressive integrated moving average model (ARIMA) to account for the volatility in petroleum prices and predict future prices. The study discovered that the monthly petroleum prices in Kenya were non-stationary, indicating instability within the petroleum market. The ARIMA (3,2,2) model was deemed suitable and valid for predicting volatility and forecasting petroleum prices. To enhance economic growth and assist users in making informed decisions, the study recommends that the Kenyan government should implement a stable and low petroleum pricing system. This would minimize production costs as petroleum is one of the inputs in production. The ARIMA model was then used to predict petroleum prices for the period between January 2011 and December 2021. Which was necessary in forecasting the future prices in coming years

# **TABLE OF CONTENTS**

Contents

[**ABSTRACT** 5](#_Toc101295637)

[**TABLE OF CONTENTS** 6](#_Toc101295638)

[**CHAPTER ONE** 8](#_Toc101295639)

[**INTRODUCTION** 9](#_Toc101295640)

[**1.1 Introduction** 9](#_Toc101295641)

[**1.2 Background of the study** 9](#_Toc101295642)

[**1.3 Statement of the problem** 10](#_Toc101295643)

[**1.4 Objectives of the study** 11](#_Toc101295644)

[1.4.1 General Objective 11](#_Toc101295645)

[1.4.2 Specific Objectives 11](#_Toc101295646)

[**1.5 Research hypothesis** 11](#_Toc101295647)

[**1.5** **Statement of hypothesis** 12](#_Toc101295648)

[**1.6 Significance of the study** 12](#_Toc101295649)

[**1.7 Definition of terms** 12](#_Toc101295650)

[**CHAPTER TWO**  14](#_Toc101295651)

[**2.1 Introduction** 14](#_Toc101295652)

[**2.3** **THEORITICAL REVIEW** 15](#_Toc101295653)

[**CHAPTER THREE** 16](#_Toc101295654)

[**RESEARCH DESIGN AND METHODOLOGY** 16](#_Toc101295655)

[**3.1 INTRODUCTION** 16](#_Toc101295656)

[**3.2 Data and data source** 16](#_Toc101295657)

[**3.2Autoregressive Integrated Moving Average (ARIMA) model** 16](#_Toc101295658)

[**3.3 Stationarity Analysis** 17](#_Toc101295659)

[**3.4 Fitting ARIMA model** 17](#_Toc101295660)

[**4.0 CHAPTER FOUR:** 20](#_Toc101295661)

[**RESULTS AND FINDINGS** 20](#_Toc101295662)……………………………………………………………………………………………………..20

4.1 Time Series Data For Petroleum Prices…………………………………………………………………………….20

Table 1: Summary Of Descriptive Statistics ……………………………………………………………………………20

4.1.1 Stages Of Building The ARIMA Model……………………………………………………………………………20 petroleum prices from 2011 to 2021………………………………………………………………………………………….21

price series……………………………………………………………………………………………………………………….25

Series ts(kilo$residuals) ……………………………………………………………………………………………………26

Normal Q-Q Plot……………………………………………………………………………………………………………….27

4.2 FORCASTING USING ARIMA (3,2,2)…………………………………………………………………………….27 Forecasts from ARIMA(3,2,2)………………………………………………………………………………………………..28

**CHAPTER FIVE**………………………………………………………………………………………………………………….30 CONCLUSSION AND RECOMMENDATIONS……………………………………………………………………30

5.1 INTRODUCTION …………………………………………………………………………………………………………..30

5.2 CONCLUSIONS……………………………………………………………………………………………………………..30

5.3 RECCOMENDATIONS.  [30](#_Toc101295664)

# 

# **LIST OF ABBREVIATIONS**

ERC – The Energy Regulatory Commission

AIC – The Akaike Information Criterion

BIC – The Bayesian Information Criterion

KPC – The Kenya Pipeline Company

AR – Autoregressive

MA – Moving Average

ARMA – Autoregressive Moving Average

ARIMA – Autoregressive Integrated Moving Average

SARIMA – Seasonal Autoregressive Integrated Moving Average

ADF – The Augmented Dickey Fuller test

ACF – The Autocorrelation Function

PACF – The Partial Autocorrelation Function.

**CHAPTER ONE**

# **INTRODUCTION**

## **1.1 Introduction**

In this chapter, we provide the basis for the study, we give an overview of the petroleum prices to provide background for the study. Then, we outline the statement of the problem, objective of the study, research question, and significance of the study.

* 1. **Background of the study**

Petroleum is a flammable liquid that consists of a complex mixture of hydrocarbons and other liquid organic compounds found beneath the earth’s surface. In today’s uncertain environment, governments and organizations must plan and make decisions that affect their future by forecasting. Forecasting is essential to plan and meet the conditions of the future, which we have imperfect knowledge of (Hank et al., 2003). Organizations that cannot react quickly to changing conditions and cannot foresee the future accurately are likely to become extinct. Moreover, forecasts play a crucial role in the planning process of businesses and organizations by enabling managers to anticipate the future and reducing uncertainties, making the planning process easier, and enabling them to develop meaningful plans. A forecast is a statement about the future (Stevenson, 1999).

In recent times, the prices of petroleum products, including kerosene, petrol, gas, and diesel, have risen sharply, significantly impacting almost all major sectors of the economy. These products are obtained from crude oil, which is a naturally occurring oil composed of hydrocarbons and other organic materials mined from underground and then processed in oil refineries. Kenya’s oil refineries are located on the coast. After refining, these products are transported through a pipeline network managed by the Kenya Pipeline Company (KPC).

The Kenya Pipeline Company is responsible for transporting, storing, and delivering petroleum products to consumers in Kenya via its pipeline system and oil depot network. Petroleum products are transported from Mombasa to Nairobi and western Kenya via Nakuru Eldoret, and Kisumu. KPC works closely with the National Oil Corporation of Kenya and operates five storage and distribution depots in the towns mentioned above.

The pricing of oil products is a complex process controlled by the relevant government department, which in Kenya is the Energy Regulatory Commission (ERC). The commission sets the maximum and minimum acceptable prices for the products, protecting the general public against exorbitant prices charged by retailers and other oil marketers. The price capping is reviewed every month by the ERC. In 2010, the ERC instituted a pricing formula to be used by oil marketing firms, which takes into account transportation cost from Mombasa to the nearest wholesale depot, allowed losses in the pipeline and in the depot, allowed gross wholesale and retail margins, among other considerations, when determining the maximum retail pump prices for petrol, kerosene, and diesel products. Actual oil retail prices for various oil products will vary from one oil marketing firm to the other based on the unique aspects of each of the firms. Since Kenya mostly relies on imported crude oil from the Middle East region, the prices of petroleum products are affected by fluctuations in global crude oil prices and how the Kenyan shilling performs against the US dollar (foreign exchange rate).

## **1.3 Statement of the problem**

In today’s world, the rise in global oil prices is having a significant impact on economic activities worldwide, and Kenya is no exception. Petroleum products are a crucial source of energy for the country, and fluctuations in oil prices have been observed over the years. It is highly likely that these changes will continue in the foreseeable future, affecting every individual and organization in Kenya. When petroleum product prices increase, it affects the cost of all basic commodities, including transportation. It is crucial for individuals and organizations to have prior information about these changes to manage and budget their cash flow effectively. Unfortunately, such situations create many challenges, forcing people to spend beyond their budget and leading to losses for companies, which can ultimately result in job losses for workers. Therefore, we conducted a study to investigate past changes in petroleum product prices and forecast future prices. Our aim is to help the government and consumers manage their cash flows effectively.

## **1.4 Objectives of the study**

The study’s objectives were split into two categories;

### 1**.4.1 General Objective**

The primary aim of this project was to predict the annual prices of petroleum products in Kenya.

### 1.4.2 **Specific Objectives**

The study has two specific objectives which is;

1. To forecast the monthly prices of petroleum in Kenya from January 2022 to December 2023.
2. To test the stationarity of the data by using the Autocorrelation function (ACF) and the Partial Autocorrelation Function (PACF), as well as conducting a unit root test.

## **Research hypothesis**

1. Do yearly prices exhibit a trend?
2. Is the data stationary?
3. Which ARMA model is the most appropriate for forecasting petroleum prices?
4. What is the forecasted value for petroleum prices?
5. Does the data exhibit any seasonal patterns?

## **Statement of hypothesis**

1. H0: The monthly prices of petroleum in Kenya do not exhibit stationarity.

H1: The monthly prices of petroleum in Kenya exhibit stationarity.

1. H0: The monthly prices of petroleum in Kenya are not seasonal.

H1: The monthly prices of petroleum in Kenya are seasonal

## **1.6 Significance of the study**

The research aimed to forecast changes in petroleum prices, which is significant for both the energy industry and consumers. The results of the study can be used by industry players for their procurement needs and by consumers to plan and budget appropriately, mitigating the negative effects of price volatility and cost increases. By managing cash flow effectively, individuals and organizations can avoid adverse reactions to price changes. Additionally, the findings of the study can inspire further research on petroleum prices and other petroleum products in Kenya.

## **1.7 Definition of terms**

**Time series**: A time series is a collection of measurements obtained through equally spaced observations over time.

**A stationary time series**: has constant statistical properties such as mean and variance over time, while a non-stationary time series exhibits trend and seasonality.

**Autoregressive (AR) models;** Autoregressive (AR) models predict future values based on past values,

**Moving average (MA) models:** use a weighted average of current and past observations as uncorrelated, mean-zero random noises.

**Autoregressive Integrated Moving Average (ARIMA) models**: statistical models that use time series data to either understand the dataset better or predict future trends.

**ARIMA model:** is a discrete-time linear equation with AR order p, MA order q, and differencing order d.

**The autocorrelation function (ACF):** measures correlations between a time series and its lagged values,

**partial autocorrelation function (PACF):** measures the partial correlation of a stationary time series with its own lagged values, taking into account values at all shorter lags.

**1.5.7 Augmented Dickey Fuller (ADF) test**

statistical method used to determine whether a time series data set is stationary or not by testing the presence of a unit root.

**H0:** There is a unit root in the data,

**H1**: Data is stationary.

# **CHAPTER TWO**

**LITERATURE REVIEW**

## **2.1 Introduction**

This chapter provides a summary of previous research on the patterns, trends, and prevalence of petroleum. The studies conducted by other researchers, their duration and personnel are highlighted. The rising prices of petroleum products have become a major concern in recent years, leading to microeconomic issues. Various forecasting techniques have been utilized in this study, but conflicting results have been obtained. In Banks’ (2004) research, he examined the economic theory associated with the aggregate supply of oil, with a focus on the exhaustible resource theory. Coimbra and Estevez (2004) found it challenging to identify any systematic behavior in oil prices as they appeared to follow a random-walk process. However, Chinn et al. (2005) observed that a random-walk characterization of energy commodity prices is not particularly accurate. Some studies have focused on individual petroleum products. For instance, Sanders and Manfred (2004) used heating oil futures to forecast the on-highway retail diesel prices, and they found that the basis could be used to prepare forecasts for the cash price if futures contract prices were available for a particular commodity. However, such studies have not been conducted in Kenya to forecast the prices of petroleum Products. According to Harari (2014), commodities such as oil cannot be predicted or explained deterministically because they are chaotic, with so many forces at work, and their interactions are so complex that small variations in the strength of the forces and the way they interact produce significant differences in outcomes.

Petroleum products are referred to as a “level two” chaotic system, meaning that they do not respond to predictions about them. The weather, for example, is an example of level one chaos. Banks (2004) investigated the economic theory involved in the aggregate supply of oil, focusing on the theory of exhaustible resources. Coimbra and Estevez (2004) noted that oil prices exhibited a random walk process with no systematic behavior. However, Chinn et al. (2005) observed that a random walk characterization of energy commodity prices was not particularly accurate.

Other studies have focused on individual petroleum products. For instance, Sanders and Manfred (2004) used heating oil futures to forecast on-highway retail diesel prices. The study found that if futures contract prices were available for a particular commodity, forecasts for the cash price could be easily prepared by utilizing the basis. However, such studies have not been conducted in Kenya as a way of forecasting petroleum product prices.

According to Harari (2014), petroleum products, such as oil, cannot be predicted deterministically because they are chaotic. Petroleum products are ‘level two’ chaotic systems, meaning that they do not react to predictions made about them. Chaotic systems come in two shapes. Level one chaos does not react to predictions, whereas level two chaos may be influenced by predictions. The weather, for example, is a level one chaotic system

## **2.3 THEORITICAL REVIEW**

The idea of the ARMA and ARIMA models was introduced in Peter Whittle’s 1951 thesis titled “Hypothesis Testing in Time Series Analysis,” and it became more widely known after George E.P. Box and Gwilym Jenkins’ book in 1976. These models allow the dependent variable to depend on past or lagged values of the independent variable, as well as its own past or lagged values. Forecasting can be achieved if the present can be modeled using only past or lagged values of the independent variable.

ARMA and ARIMA models are widely used in predicting second-order stationary processes (Box and Jenkins, 1976; Chan, 2011; Montgomery et al., 2011; Henrik, 2007; Wei and Wei, 1990). The ARMA model can be further divided into the autoregressive (AR) and moving average (MA) parts. The AR part involves regressing the variable on its own lagged values, while the MA part models the error term as a linear combination of error terms occurring contemporally and at various times in the past

# **CHAPTER THREE**

# **RESEARCH DESIGN AND METHODOLOGY**

## **3.1 INTRODUCTION**

A time series data refers to a set of observations or measurements on quantitative variables that are recorded sequentially at a uniform time interval. These time intervals can be daily, weekly, monthly, quarterly, annually, etc. For instance, one example of time series data could be the monthly petroleum prices at the ERC

## **3.2 Data and data source**

In this study, secondary data was collected from reputable online sources such as [www.erc.go.ke](http://www.erc.go.ke) for past petroleum product prices. Time series analysis relies on the assumption that the factors that influenced the activity being studied in the past and present will continue to do so in a similar manner in the future. The primary objectives of time series analysis are to identify and isolate the patterns that affect the activity for predictive purposes. Hence, more recent time series data is considered more relevant for forecasting the immediate future. The sample period for this study was selected to capture the most recent data over the past 11 years. To forecast the prices, the study utilized the ARIMA methodology. The data analysis was performed using R software and MS excel with descriptive statistics using SPSS

## **3.2Autoregressive Integrated Moving Average (ARIMA) model**

The ARIMA model is designed to analyze data that exhibits patterns rather than random events, with a focus on non-seasonal factors. This model utilizes data collected at regular intervals to identify patterns and make predictions. The model is denoted as ARIMA (p, d, q), where p represents the non-seasonal AR order, q represents the non-seasonal MA order, and d represents the non-seasonal differencing order. The non-seasonal AR order, p, is used to model the dependence between an observation and a certain number of lagged observations. The non-seasonal MA order, q, is used to model the dependence between an observation and a certain number of lagged errors. The non-seasonal differencing order, d, is used to remove any trends or seasonality from the data, making it stationary and easier to model. By identifying the appropriate values for p, d, and q, the ARIMA model can be used to forecast future values based on past patterns.

The ARIMA (p, d, q) model is represented as a linear equation with discrete time and noise, in the form of:

Φ(B)(1 − B) dXt = Θ(B)ωt 3.1

A specific type of ARIMA (0, 1, 0) model is known as a random walk [21]. This model supposes that in each period, the variable moves randomly away from its prior value by a certain step size, and that these steps are independently and identically distributed. The model is expressed as follows

Xt = Xt−1 + ωt 3.2

## **3.3 Stationarity Analysis**

The characteristic equation of an AR(p) process is represented as Φ(B) = 0, where ωt = Φ(B) is a function of B (the backward shift operator) and the coefficients ω1, ω2, ..., ωp. To ensure stationarity in an AR(p) process, all the roots of the characteristic equation must be outside the unit circle. If the AR (1) process is represented as Xt = δ + ϕ1Xt−1 + ωt, where δ is a constant, then it is stationary if the absolute value of ϕ1 is less than 1

## **3.4 Fitting The ARIMA model**

To apply this model, the following procedures will be implemented:

**The first stage in fitting ARIMA model is model identification**.

To make the data stationary, it may be necessary to perform pre-processing. The following techniques can be used to achieve stationarity

* Observe the data
* Rescale the data by using logarithmic or exponential transformation
* Eliminate deterministic components
* Perform differencing repeatedly until the data becomes stationary.

The stationarity is observed by the autocorrelations decrease exponentially and quickly towards zero. After making the data stationary, we can proceed to fit an ARMA(p, q) model.

**The next step in the ARIMA model fitting process is the estimation of the model parameters**

**AR process**

**To fit** a pure AR(p) ,i.e. Xt= by using the **Yule-Walker equation.**

**.** fit by solving

**The final stage in fitting an ARIMA model is the verification stage**

This involves testing the fitted model to see how well it fits the original data and to assess its performance in making future forecasts. Different diagnostic tests may be conducted to evaluate the model, such as checking for autocorrelation in the residuals, checking for normality of the residuals, and examining the forecast errors. If the model does not fit the data well or does not perform well in forecasting, further modifications or a different model may need to be considered.

This includes;

* Overfitting: adding unnecessary parameters to the model that may lead to overfitting. Instead, use statistical tests such as the likelihood ratio test or t-test to verify if the additional parameters are significant or not.
* Residual analysis: Calculate the residuals from the model and examine them to ensure that they resemble white noise by checking the autocorrelation functions (ACFs), partial autocorrelation functions (PACFs), then plot them.

The verification stage of fitting an ARIMA model involves the following procedures:

Examine the ACF of the residuals to ensure that they are non-significant for a good model.

• Plot a normal Q-Q plot to check if the residuals follow a normal distribution.

• Plot the p-values of the Ljung-Box statistic. This statistic is a function of accumulate sample autocovariances, rj, up to a specified time lag m. It is determined as

Q(m) = n (n + 2) Xm j=1 r 2 j n − j

where n is the number of usable data points after any differencing. Q(m) has a chi-square distribution with m degrees of freedom.

n refers to the count of data points that can be utilized after any necessary differencing. The statistic Q(m) follows a chi-square distribution and has ‘m’ degrees of freedom.

The Ljung-Box tests the null hypothesis that the residuals are random i.e. they are independently distributed against the alternative hypothesis that the residuals are not random. If the p-value are greater than 0.05 then the model is a good fit.

The final step in fitting an ARIMA model is to use it for forecasting future values of a time series based on the data collected up to the present. The forecasting process involves predicting future values and calculating the difference between the predicted values and the observed values, which is called the forecast error. To forecast future values, I followed the procedure below

* for any ωj with 1 ≤ j ≤ n, use the sample residue for time point j;
* for any ωj with j > n, use 0 as the value of ωj;
* for any Xj with 1 ≤ j ≤ n, use the observed value of Xj;
* for any Xj with j > n, use the forecasted value Xj.

The accuracy of the forecast was determined by the degree of closeness between the predicted quantity and the actual value. The effectiveness of the model was evaluated using Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Root Mean Square Error (RMSE). These measures of accuracy were used to determine the magnitude of the overall error occurred due to forecasting. A small MAE indicates a good forecast, as it measures the average absolute deviation of forecasted values from the original ones.

# 

# **4.0 CHAPTER FOUR:**

# **RESULTS AND FINDINGS**

## **4.1 Time Series Data for Petroleum Prices**

This chapter deals with the analysis of petroleum prices in Kenya from 2011 to 2021. The time series plot display observations on the Y-axis against equally spaced time interval on the X-axis. They are used to evaluate patterns and behavior of data over time as displayed in the figures below. The data has both upward and downward trend and hence non-stationary.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 132 | 101.44 | 80.00 | 89.08 | 114.83 | 100.30 | 13.43197 | 122.00 |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **N** | **MEAN** | **MINIMUM** | **1ST**  **QUARTILE** | **3RD**  **QUARTILE** | **MEDIAN** | **STANDARD DEVIATION** | **MAX** |
| 132 | 101.44 | 80.00 | 89.08 | 114.83 | 100.30 | 13.43197 | 122.00 |

### Table 1: Summary Of Descriptive Statistics

The data consisted of 132 monthly observations from January 2011 to December 2021. The highest petroleum price was recorded in the year 2015 during the months of August. The mean of the data was found to be 101.44 using the R software.

### 4.1.1 Stages of Building The ARIMA Model

The Auto Regressive Intergraded Moving Average denoted by ARIMA (p, d ) was obtained by differencing the ARMA (p,q) d-times. ARIMA models are the class of models that have the capabilities to represent the stationary as well as non-stationary time series to produce accurate forecast based on a description of historical data of a single variable. The ARIMA models do not assume any particular pattern hence making it different from other models. Box-Jenkins approach of building ARIMA model is based on the following steps:

* Model identification
* Parameter estimation and selection
* Model validation
* Model use

#### **4.1.1.1 Model Identification**

This involved determining the order of (p, d, q) of AR and MA components of the model in order to capture the silent dynamic features of the data. It seeks to check whether the data is stationary or non- stationary and the order of differentiation d which makes the time series stationary. It mainly leads to use of graphical procedures. We have used monthly pricing data from 2011 -2021 for model validation. The plot of the ACF data shows a slow decay hence the data is non- stationary. Therefore we have to transform the non- stationary data into stationary by differencing the series. The ACF plot of differenced series shows that the time series is stationary

## petroleum prices from 2011 to 2021

petroleum prices

2012

2014

2016

2018

2020

2022

80

90

100

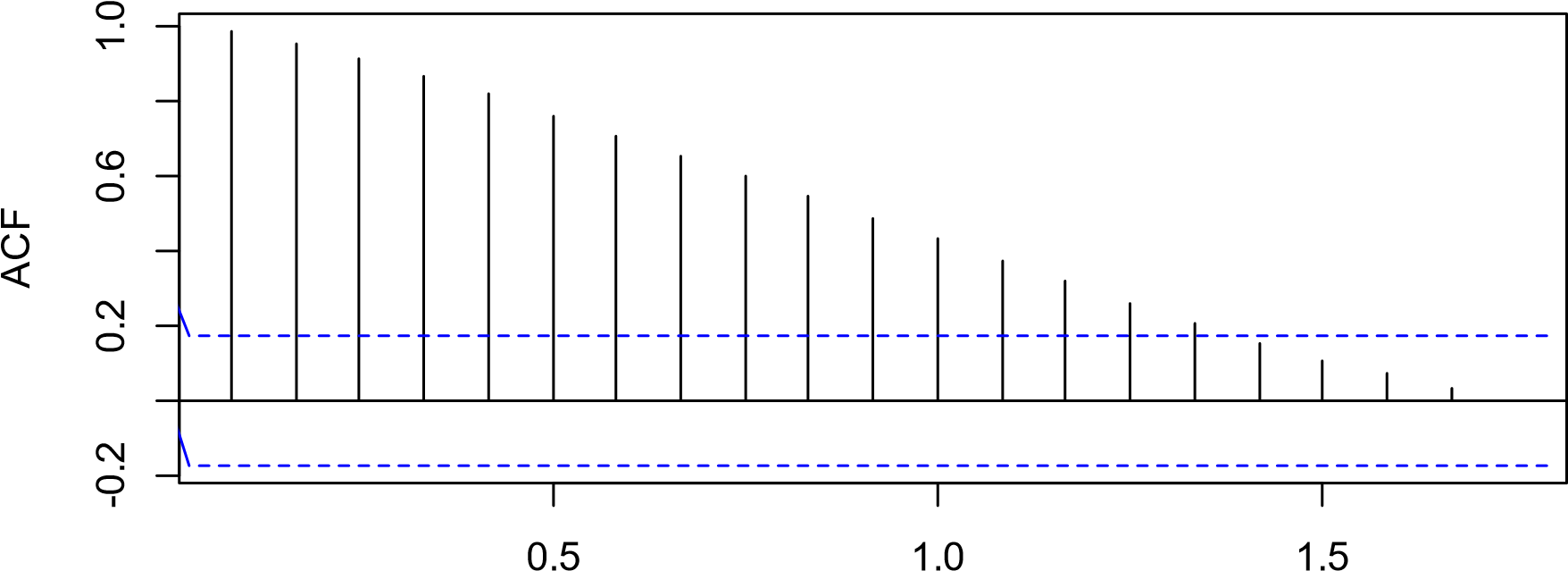
110

120

year

**Figure 3: Plot of non-stationary time series data for petroleum price**

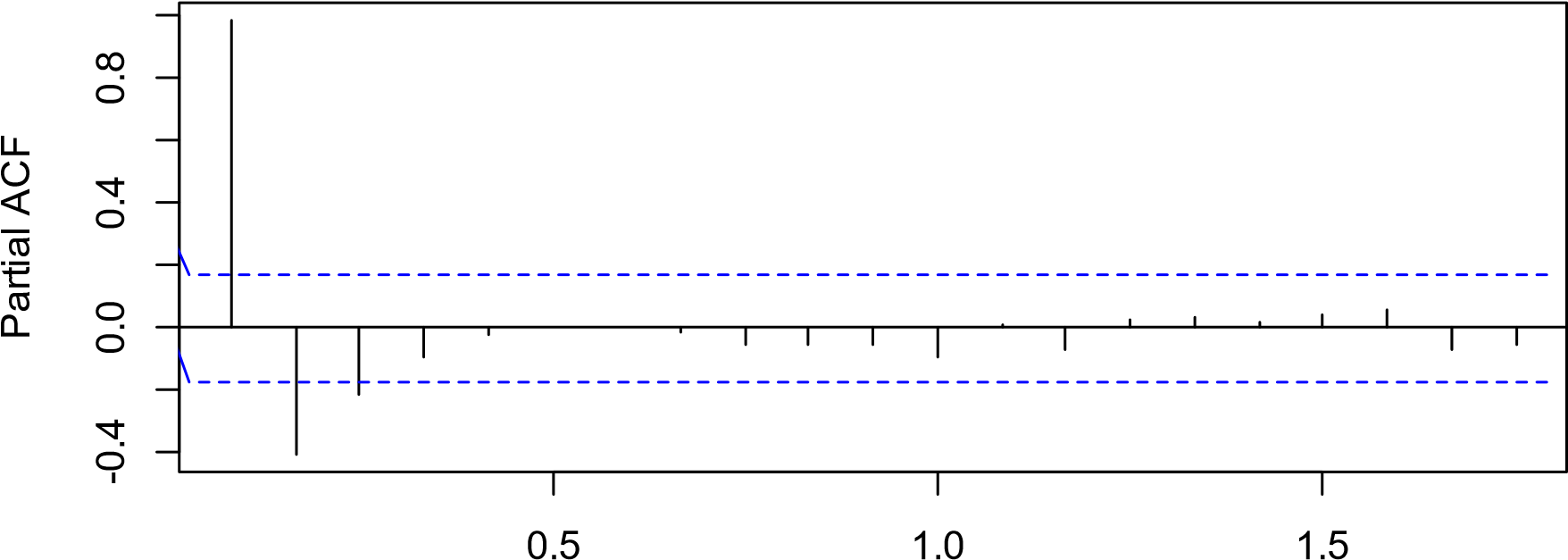
### price series



Lag

**Figure 4: ACF of non-stationary time series data for petroleum prices**

### price series



Lag

#### **Figure 5: PACF of non-stationary time series data for petroleum prices**

petroleum prices

2012

2014

2016

2018

2020

2022

-6

-4

-2

0

2

4

Time

#### **Figure 6: Plot of second differenced series**

The time series of the second difference appears to be stationary in mean and variance and so an ARIMA (p,2,q) model is probably appropriate for the time series of the petroleum prices. The Augmented Dickey fuller check showed a p-value of 0.01 signifying that the petroleum price is stationary at second differencing**.**

#### **4.1.1.2 Parameter Estimation and Selection**

The model used in this study model is of the form ARIMA (p, d, q). From figure 7 below, we saw that the mean was stationary and so our d=2. The values of p, q are derived from the ACF and the PACF of the time series with both stationary mean and variance.

The ARIMA models are, ARIMA (0,2,1), ARIMA (1,2,0), ARIMA (0,2,2), ARIMA (3,2,0) and ARIMA (3,2,2).

Using AIC and BIC, ARIMA models of different orders above were tested. This enabled the best model with lowest Akaike information criterion to be chosen. A summary of the result was presented in table below.

|  |  |  |
| --- | --- | --- |
| MODEL | AIC | BIC |
| ARIMA(1,2,0) | 614.73 | 620.44 |
| ARIMA(0,2,2) | 463.17 | 417.73 |
| ARIMA(3,2,0) | 517.78 | 529.19 |
| ARIMA(3,2,2) | 451.74 | 468.86 |

#### **Table 2: Summary of ARIMA Models**

Based on AICs of the models, ARIMA (3,2,2) was chosen as the best model. The model parameter was significant from table above, hence our proposed model was justified. After considering the four models, the model ARIMA (3,2,2) had significant parameters and lowest AIC and BIC values.

Series: diff2

ARIMA (3,2,2)

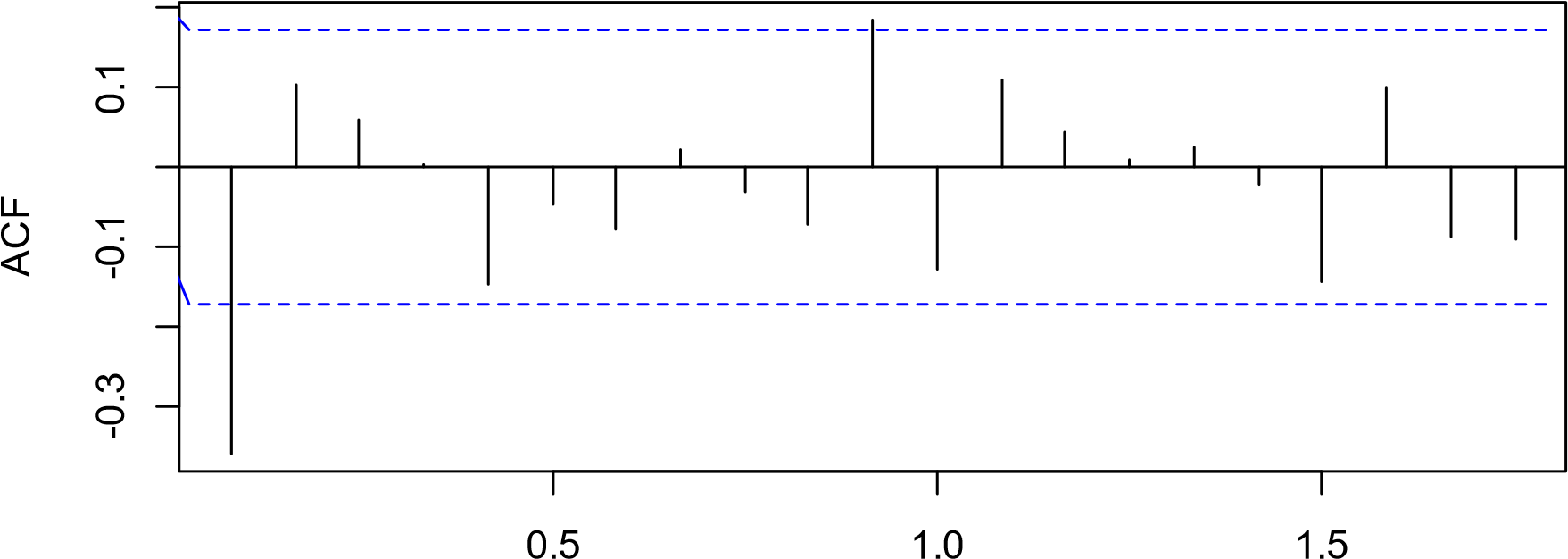
|  |  |  |
| --- | --- | --- |
| Coefficients: | Estimate | Standard error |
| ar1 | -0.3436 | 0.0906 |
| ar2 | 0.0404 | 0.0977 |
| ar3 | 0.1437 | 0.0932 |
| ma1 | -1.9980 | 0.0747 |
| ma2 | 0.9995 | 0.0746 |

#### **Table 3: Summary of The Best Model**

|  |  |  |
| --- | --- | --- |
| AIC | AICc | BIC |
| 451.74 | 452.44 | 468.86 |

#### **Table 4: AIC and BIC of Best Model**

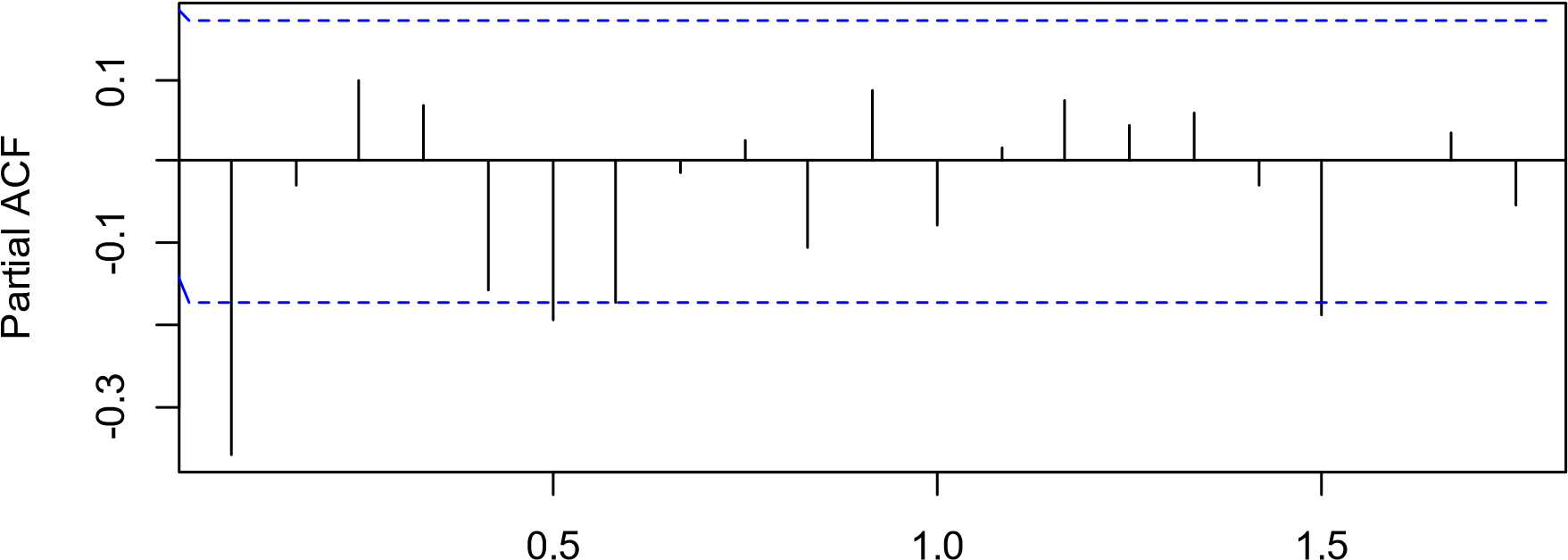
### price series



Lag

**Figure 7: ACF OF ARIMA (3,2,2)**

### price series



Lag

#### **Figure 8: PACF OF ARIMA (3,2,2)**

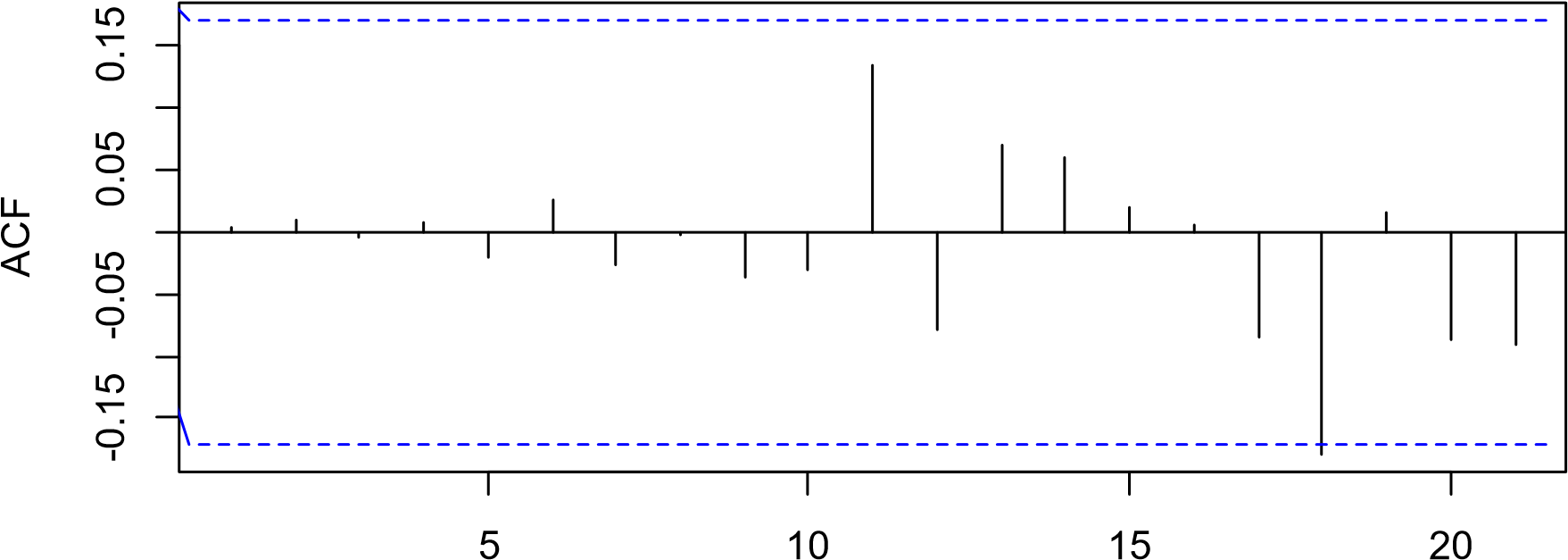
#### **4.1.1.3 Model Validation**

After having chosen ARIMA (3,2,2) as the best model, the model adequacy is further checked to draw empirical conclusion regarding the model as good fit hence its usage in estimation and forecasting. These tests were performed using Ljung Box test coupled with the ACF and the Q-Q plot of the residuals. The plots showed that the residuals from the model are similar to a white noise hence the model fits the data well. The p-value estimated by the Ljung Box test in the table was greater than 0.05. The p-value was 0.9577 which showed that the residuals were random.

|  |  |  |  |
| --- | --- | --- | --- |
| **Test type** | **X-square(x)** | **df** | **p-value** |
| **Ljung Box** | **0.0028165** | **1** | **0.9577** |

**Table 5: Ljung Box Test Results**

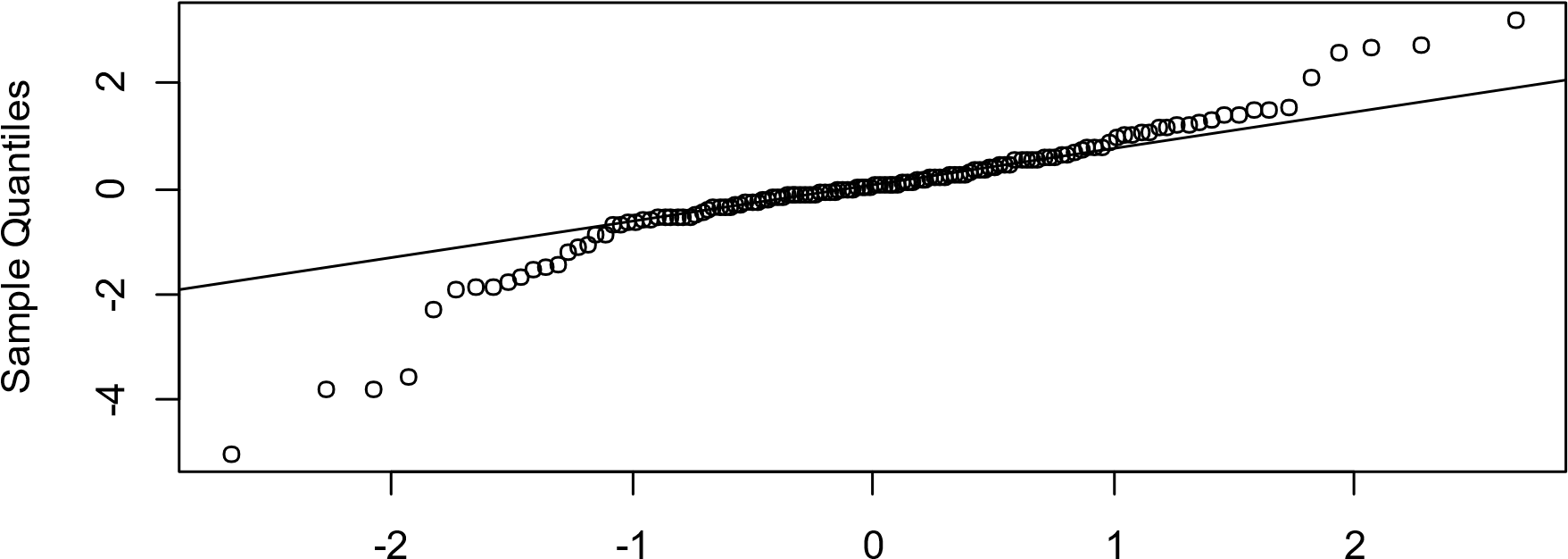
### Series ts(kilo$residuals)



Lag

**Figure 9: ACF Of Residuals**

### Normal Q-Q Plot



Theoretical Quantiles

#### **Figure 10: Graph of Normal Q-Q Plot Of Residuals**

Residuals are estimates of experimental error obtained by subtracting the observed responses from predicted response. The Q-Q plot points lie on a straight line, the deviations from the straight line are minimal which indicates a normal distribution. The residuals plot of the ARIMA (3,2,2) model shown above confirm that the residuals are random with no pattern hence we are sure that the model is fitting the data well and the statistical assumptions are all met, therefore we can use the model for forecasting.

## **4.2 FORCASTING USING ARIMA (3,2,2)**

## 

In this step the study utilized monthly data for Kenya tea exportation from the year 2012-2018 to build the model. The ARIMA model is given by:

φp(B)∆dXt = θq(B)ξt where:

p=The AR component d=The differencing part q=The MA component

For petroleum price model, the values of p= 3, d=2 and q=2. Hence our model becomes ∅3(𝐵)∆2𝑋𝑡 = 𝜃2(𝐵)𝑒𝑡 (14)

Using R statistical software, the model parameters obtained was as follows:

φ1=-0.3436, φ2=0.0404, φ3=0.1437, θ1 = −1.9980, θ2 = 0.9995

Therefore

𝑋𝑡 = −0.3436𝑋𝑡−1 + 0.0404𝑋𝑡−2 + 0.1437𝑋𝑡−3 − 1.9980𝑒𝑡−1 + 0.9995𝑒𝑡−2 + (15)

### Forecasts from ARIMA (3,2,2)

2012

2014

2016

2018

2020

2022

2024

40

60

80

100

#### **Figure 11: Forecasts from ARIMA (3,2,2)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| MONTH | YEAR | POINT  FORECAST | L0WER 95%  CI | HIGHER 95%  CI |
| JAN | 2023 | 98.33613 | 96.01524 | 100.65703 |
| FEB | 2023 | 95.03086 | 90.82246 | 99.23927 |
| MAR | 2023 | 92.34316 | 85.78925 | 98.89706 |
| APR | 2023 | 90.47878 | 81.12277 | 99.83476 |
| MAY | 2023 | 89.30999 | 77.01846 | 101.60152 |
| JUNE | 2023 | 88.56635 | 73.47364 | 103.65907 |
| JULY | 2023 | 88.04052 | 70.37129 | 105.70975 |
| AUG | 2023 | 87.63991 | 67.59935 | 107.68047 |
| SEPT | 2023 | 87.33898 | 65.08773 | 109.59023 |
| OCT | 2023 | 87.12565 | 62.79323 | 111.45808 |
| NOV | 2023 | 86.98100 | 60.68149 | 113.28050 |
| DEC | 2023 | 86.88298 | 58.72151 | 115.04445 |
| JAN | 2024 | 86.81428 | 56.88696 | 116.74157 |
| FEB | 2024 | 88.76465 | 55.15761 | 118.37169 |
| MARCH | 2024 | 86.72874 | 53.51823 | 119.93925 |
| APRIL | 2024 | 86.70325 | 51.95705 | 121.44946 |
| MAY | 2024 | 86.68549 | 50.46432 | 122.90622 |
| JUNE | 2024 | 86.67314 | 49.03187 | 124.31442 |
| JULY | 2024 | 86.66448 | 47.65284 | 125.67612 |
| AUG | 2024 | 86.65832 | 46.32157 | 126.99507 |
| SEPT | 2024 | 86.65393 | 45.03336 | 128.27451 |
| OCT | 2024 | 86.6503 | 43.78424 | 129.51741 |
| NOV | 2024 | 86.64864 | 42.57083 | 130.72645 |
| DEC | 2024 | 86.64711 | 41.39019 | 131.90403 |

**Table 6: Summary of Forecast Data**

**CHAPTER 5**

## **CONCLUSSION AND RECOMMENDATIONS**.

## **5.1 INTRODUCTION**

This chapter presents the conclusions and recommendations based on the analysis of the study. The chapter draws upon the previous four chapters and provides an overview of the main findings.

## **5.2 CONCLUSIONS.**

Based on the objectives of the study and the literature review, as well as the results obtained from the analysis, the following conclusions can be drawn with regard to the main objective of the study, it can be concluded that the price of petroleum is expected to increase in the coming two years. This conclusion is based on the analysis of the data and the identified patterns in the data. it was found that the data was stationary after being differentiated twice. This conclusion was corroborated by the Ljung Box test and the diagnostics of the residual plots. It was found that the residuals of the ARIMA (3,2,2) were uncorrelated, indicating that the model was appropriate and working well.

It can be concluded that there was no seasonal variation or specific trend in the levels of petroleum price in Kenya during the period under study. This conclusion was based on the analysis of the data and the identified patterns in the data. The final chapter of this study is based on the previous four chapters and it deals with the conclusion. Based on the objectives as well as the literature reviewed and from the analysis it is imperative to draw the following conclusion. with regard to the main objective, it can be concluded that price of petroleum will increase in the coming if all the factors remain constant. It can also be concluded that the data was stationary after differentiating twice.

The Ljung Box test was further corroborated by diagnostics of the residual plots. It showed that the residuals of the ARIMA (3,2,2) were uncorrelated, hence the model was appropriate and working well. It can be concluded that there was no seasonal variation and no specific trend in the levels of petroleum price in Kenya during the period under study.

## **5.3 RECCOMENDATIONS.**

Due to the factors that affect petroleum prices, such as weather, political instability, and wars, it is recommended that the model should be continuously updated every two years for accurate forecasting. This recommendation is based on the understanding that the factors affecting petroleum prices are constantly changing and evolving, and therefore the model should be regularly updated to reflect these changes. it is recommended that further research be conducted using more advanced models like artificial neural networks. These models have become increasingly popular among researchers in recent years, and are known to be more accurate in forecasting than traditional models. This recommendation is based on the understanding that the use of more advanced models can improve the accuracy of forecasting and provide a more comprehensive understanding of the factors affecting petroleum prices.

This study concludes that the price of petroleum is expected to increase in the coming two years and that there was no seasonal variation or specific trend in the levels of petroleum price in Kenya during the period under study. The study recommends that the model should be updated every two years for accurate forecasting and that further research should be conducted using more advanced

#### **REFERENCES**

George. E. P .Box; Gwilym M .Jenkins (1976). Time Series and Analysis: Forecasting and Control.

Published by Holden. Day, 1976.

Peter Whittle (1927) Hypothesis testing and Time Series analysis. Published by Almqvist and Wiksells: Statistics Uppsala, Vol.4.

Shumway, Robert H, Stouffer, David S. (2010) Time Series Analysis and Its application with R examples. Published by Springer (2010).

Jonathan D Cryer; Kung-Sik Chan (2010). Time Series Analysis with application in R 2𝑛𝑑 edition. Published by Springer

Douglas C Montgomery; Lynwood A. Johnson; John S. Gardiner (2011) Forecasting and Time Series Analysis 2𝑛𝑑 Edition. Published by McGraw-Hill; Subsequent edition (July1, 1990).

Henrik Madsen (2007) Time Series Analysis 1𝑠𝑡 edition Published by eBook (28𝑡ℎNov 2007).

William W. S. Wei (1990) Time Series Analysis Univariate and Multivariate Methods 2𝑛𝑑 edition.

Publisher Greg Tobln 2006.

Menzie D. Chinn; Michael Le Blance; Oliver Coibion (2005) The Predictive Content of Energy Futures:

An update on Petroleum, Natural gas, Heating oil and Gasoline, Thesis.

Harari, Yuval Noah (2015) A Brief History of Humankind, Published by Random House, 2015.

Sthorsten Beck and Ross Levine (2004). Stock market, banks and growth: panel evidence, journal of banking and finance, vol. 28, issue 3 page 423-442.

Carlo Coimbra, Paulo soares Esteves (2004) oil assumption in macro economic forecasts: should we follow future market expectations? Vol. 28 issue 2, page 87- 106.

Gathondu. E.K. (2014). Modelling of wholesale prices for selected vegetables using time series model in kenya (Masters thesis, university of Nairobi).