KWAME NKRUMAH UNNVERSITY OF SCIENCE AND TECHNOLOGY,

KUMASI-GHANA

COLLEGE OF SCIENCE

DEPARTMENT OF MATHEMATCS

TIME SERIES MODELING AND FORECASTING OF THE DEMAND OF SOME PETROLEUM PRODUCTS IN GHANA

By

Agyen Kwaku James

MAY, 2012

TIME SERIES MODELING AND FORECASTING OF THE DEMAND OF SOME PETROLEUM PRODUCTS IN GHANA

By

Agyen Kwaku James (MSc.)

A Thesis Submitted to Department of Mathematics in Partial Fulfilment of the Requirements for the Degree of

MASTERS OF PHILOSOPHY (APPLIED MATHEMATICS)

College of Science

May, 2012

DECLARATION

I hereby declare that this submission is my own work towards the MPHIL and that, to the best of my knowledge, it contains no material previously published by another person nor material, which has been accepted for the award of any other degree of the university, except where due acknowledgement has been made in the text.

AGYEN Kwaku James	....	....
(Student)	Signature	Date
Dr. F. T. Oduro	....	....
(Supervisor)	Signature	Date
Mr. F. K. Darkwa	....	....
(Head of Department)	Signature	Date

ABSTRACT

The study is an application of Box-Jenkins ARIMA modelling and forecasting to petroleum products demand in Ghana. Monthly data on demand levels from January, 1999 to December, 2010 of petroleum products namely; Gas Oil, Liquefied Petroleum Gas (LPG) and Premix Fuel were analysed and forecasts made 12 months ahead. Thus, after stationarity was established through differencing of the data, the sample ACF and the sample Partial Autocorrelation Function (PACF) of the differenced data were considered to generate possible models after which the AIC, AICc and BIC of the candidate models under the various data were examined and those candidate models with the smallest AIC, AICc and BIC were chosen as the best-fit models among the candidate models and used for forecasting. The best fit model for the National Gas Oil demand, National LPG demand and National Premix demand levels were found to be ARIMA(1,1,3), ARIMA(2,1,3) and SARIMA(3,1,0)$(2,0,0)_{[12]}$ respectively.

DEDICATION

This study is dedicated to my son Barfour Akoto Agyeman-Prempeh

ACKNOWLEDGMENTS

My deepest gratitude goes to the Almighty God for seeing me through this work successfully. I wish to express my sincere appreciation to all people who contributed to this master piece in one way or the other. Especially, I am very thankful to Dr. F.T. Oduro for giving me the chance to work on this topic. Secondly I am highly indebted to my main man and supervisor Mr. Emmanuel Harris whose relentless effort cooked the thesis topic and keenly monitored this work from the very beginning to the last end of it. I owe my gratitude to their supervision, helpful comments, discussions and support they readily offered me. My deepest appreciation goes to the lecturers and all the workers at the department of Mathematics.

I am also grateful to my family and friends for all the love, support, and humour that I have received from them, especially my wife, Ruth Grant-Essilfie for giving me the moral support and encouragement even in the most difficult times. My profound gratitude goes to the I T department of T.I. Ahmadiyya Senior High School.

I would also like to thank the management of the National Petroleum Authority for providing me with the data and helpful assistance for my work.

TABLE OF CONTENT

CHAPTER THREE

LIST OF TABLES

TITLE **PAGE**

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Background to the Study

In the modern trends of industrialisation and development, energy has become one of the most important wheels of every economy in the world of which Ghana is of no exception. This means that, the world economies are heavily reliant on energy as Alam (2006) puts it, "energy is the indispensable force driving all economic activities". In other words, the greater the energy consumption, the more the economic activity in the nation and as a result a greater economy emerges. Energy is important for economic development. Its demand is linked with factors such as energy prices, income, and population, degree of urbanization, level of technological development and the overall structure of the economy. The energy sector therefore is one indispensable sector for a country's socio-economic development, production, and better standard of living.

Energy is an essential element and has a decisive role in our daily life, agriculture, industry and social services. Plants, coal, petroleum, electricity, sun, geothermal steam, and animals are the main energy sources. The effective demand for commercial energy is, therefore, related to economic conditions which influence the availability and access to energy sources. Ghana's energy demand and end-consumption patterns are similar to those observed in other developing countries; (Mosse, 2002). Of all the types of energy, Ghana as a developing country uses three main energy; the biomass or wood fuel which comprises mainly of charcoal and fire wood this forms 60% of the total energy consumption in the country; electricity;11% and petroleum products;29%, (Wisdom, 2002). Petroleum products though the second highest of energy used in the country has a very serious impact on the economy due to the sectors that use the petroleum products. Transportation sector's activities for instance cuts across all the sectors of the economy that is, commercially, domestically and industrially. The manufacturing industries, depends on transportation for the transport of their raw materials and finished products to their various destinations. Therefore, any shortage in the supply of petroleum products affects transportation which causes corresponding shortage in the products and services that are directly affected by transportation.

Although Ghana has discovered oil in commercial quantities she is still importing oil and therefore any shortage in the supply and its corresponding high prices of oil products from her oil exporting countries can post a challenge in the country since the high and volatile oil prices is an obstacle to the growth of the economy. Dynamics of global energy markets have become distinctly marked by sharp increases in global demand and severe supply shocks that are hitting global economies. These trends are causes of concern as they affect economic performance, especially in oil-importing African countries, (African Developing Bank, 2008). The supply-side effect creates immediate economic distortions that hit oil-intensive production sectors. The supply-side effect refers to the reduced availability of a key production input (oil) when oil prices rise. Because the cost of other production inputs, notably labour, do not fall, the overall per unit cost of production rises, leading to reduced output levels. Since output prices do not necessarily rise with increasing oil prices, the profit margins of oil-intensive production sectors plummet and may have an overall negative effect on the macro-economy.

It is a well-established fact that a rise in oil prices leads to deterioration in terms of trade of net oil-importing countries, and, subsequently, to a fall in the purchasing power of firms and households in net oil-importing countries (Dohner, 1981). This is essentially a transfer of wealth from net oil-importing to net oil-exporting countries. However, some argue that the effects of high oil prices can also be indirect, which works through the economies' trading partners. Increased trade between net oil-importers and net oil-exporters, where oil windfall is used to import more manufactured products from net oil-importing countries, may have a positive effect on the economies of net oil-importers (Abeysinghe, 2001). Therefore, the net effect of oil shocks on net oil-importing economies depends on how net-exporting countries decide to spend extra windfall purchasing power, and their trade preferences. Since most net oil-importing African economies are not well diversified or industrialized, their effective supply response capacities are limited, even if net oil exporters choose to spend their windfalls on importing goods and services from them. Increasing oil supply shortage may lead to increased money demand from net oil-importing countries (Mork, 1994), and failure to meet this demand through increased money supply leads to higher interest rates and subsequently severe shortage in the supply of petroleum. This has negative effects on consumption and investment, leading to lower growth. Consumption is affected through its positive relation to disposable income, and investment through increasing firm costs, if oil supply shortage increases prevail over a long period, they may lead to a change in the production structure in favour of non-oil intensive sectors, which may lead to other distortions. The resulting reallocation of labour and capital across sectors in response to oil price increases can affect the unemployment situation in the long term (Loungani, 1986).Overall, therefore, net oil-importing African countries like Ghana remains vulnerable to energy price shocks, particularly because Ghana is more or less a non-export-oriented. Since economic diversification is still low in most of Ghana's economy, energy shocks have the potential to continue taking a toll on the country's economy. Given that the country's energy use efficiency is among the lowest in the world—precisely at a time when energy prices are sky rocketing and given the unique opportunity offered by discoveries of oil and gas fields on the country, an explorative study of the oil and gas situation in the country is timely, especially in the face of emerging evidence of the impact of the high level and volatility of oil prices. For these reasons consumers of petroleum products in the country and the government must be assisted with an information on the trend of the petroleum products for them to be able to make an incisive decision on the prices of petroleum products like petrol, kerosene, diesel, gas-oil and gas to forecast into the future to prevent the unexpected devastation high and volatile oil prices bring to them.

1.2 Statement of the Problem

Ghana is a low middle income country and striving to become a high middle income country. Thus, the rate of acceleration of its growth has become one of the prime aims of every government over the years; this growth is partly driven by energy mainly in the form of electricity and crude fuel. The dominant crude fuel used in the country includes petrol (super), diesel, premix fuel, liquefied petroleum gas, etc. Because these fuels have become the wheels on which the economy strives there is a need to have a forecasting system by which government will be informed of the demand pattern on these fuels in the country in order to prevent shortage or excess supply as these have significant effects on the economy and its growth. A more often than not shortage in the petroleum products leads to volatile and high prices of the products.

These volatile and high prices of petroleum products are so dangerous to government and private entities in that the fluctuation in price makes planning in to the future very uncertain and devastating thus yearly objectives and aims of companies and individuals are not achieved due to unplanned shocks making production and services very difficult and unreliable. Companies and individuals are folding out of business because, most of the companies and individual businesses thrive on loans accessed from both local and foreign banks and are not able to pay back the loans due to sudden shortages and the subsequent price increases in petroleum products accompanying such shortages. Thus we need to have a system that can predict accurately the demand patterns for petroleum products in order to be able to plan successfully into the future.

.

1.3 Objectives of the Study

The main objective of this study is to model the demand behavior for the Gas Oil, Liquefied Petroleum Product and Premix Fuel so as to be able to predict or forecast the quantity the nation would need of these products in the near future.

1.3.1 Specific Objectives

The specific objective of this research includes;

i. To determine ARIMA models that could best be used to predict future demand for Gas Oil, LPG and Premix Fuel in Ghana
ii. To interpret the results in the light of market conditions in Ghana

1.4 Proposed Methodology

The data for this study was collected from the National Petroleum Authority, Accra. The data covers the period from January, 1999 to October, 2010 and comprises monthly national demand for Gas Oil, LPG and premix fuel. Time series analysis by means of the R software is used to do all the analysis on the data obtained applying the Box-Jenkins ARIMA methodology. The best ARIMA models for fitting the data are checked from diagnostic tests made up of the standardised residual, normal Q-Q plot of standardised residuals and the p values for Ljung-Box statistics. The best model will be selected using the various AR and MA, and appropriate

SARIMA candidate models. The root means square percentage error (RMSPE) would be used to check how good the chosen models fit the data.

1.5 Justification

Sudden shortages in petroleum products supply leads to high increase in the prices of the petroleum products and subsequent increase in the products and services that depend on the petroleum products in the country and this has led to the collapse of some low-income oil-dependent companies with a lot more incurring high losses. The small scale mining sector for instance laid-off most of its workers, about 84% (Chamber of Mines Annual Report, 2008). Demand for petroleum products is a necessary parameter in projecting petroleum products prices and in planning the needed refining capacity to meet future domestic consumption. It is also indispensable tool for policy makers as they indicate the extent of price increase required to curtail future losses in almost all the economic sectors of the country. The adverse change (increase) in terms of trade for oil importers reduces incomes, lowers real consumption, causes deterioration in the balance of trade and puts downward pressure on exchange rates. Economic growth slows, higher costs causes inflation to rise and unemployment results

1.6 Thesis Organisation

The study is organised in five main chapters. The first chapter covers the introduction to the study and this highlights background of the study, problem statement, objectives of the study, methodology, justification of the study, scope and limitations of the study and organisation of the study. Chapter two deals with the review of relevant literature of the study and this review focuses on method that have been adopted by previous researchers and limitations of their methods, as well as a discussion of the results from previous studies.

The third chapter discusses vividly, the mathematical and statistical methods and procedures used in the analysis of the monthly data for some of the petroleum product demand in Ghana. The fourth chapter also deals with the analysis of monthly data for some of the petroleum product demand in Ghana over 12 year period comprising of data from Jnauary1999 to November 2010 that is 144 months. The interpretations and discussions are also presented in this chapter. The last chapter covers conclusion and recommendations.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This chapter takes into consideration various works done by other researchers and authors using time series techniques and other forecasting techniques. These include the Box-Jenkins ARIMA model technique, the exponential smoothing and the Holt-Winter Exponential Smoothing techniques, artificial neural network, etc.

2.2 Previous Works Using ARIMA and Other Models for Forecasting

Monthly Malaysia crude oil production data for the period of January 2005 to May 2010 were analyzed using time-series method called Autoregressive Integrated Moving Average (ARIMA) model. Autocorrelation and partial autocorrelation functions were calculated to examine the stationarity of the data. Then, an appropriate Box-Jenkins ARIMA model was fitted. Validity of the model was tested using Box-Pierce statistic and Ljung-Box statistic techniques. The predictability of future crude oil production as a forecast is measured for three leading months (Nazuha et al, 2011).

The oil industry has used decline curve analysis with limited success in estimating crude oil reserves and in predicting future behaviour of oil and gas wells. Ayeni (2003) explores the possibility of using the Autoregressive Integrated Moving Average (ARIMA) technique in forecasting and estimating crude oil reserves and compares this approach with the traditional decline method using real oil production data from twelve (12) oil wells in South Louisiana. The Box and Jenkins methodology is used to develop forecast functions for the twelve wells under study. These forecast functions are used to predict future oil production. The forecast values generated are then used to determine the remaining crude oil reserves for each well. The accuracy of the forecasts relative to the actual values for both ARIMA and decline curve methods is determined by various statistical error analyses. The conditions, under which one method gives better results than the other, are fully investigated. In almost all the cases considered, the ARIMA method is found to perform better than the decline curve method (Ayeni, 2003).

Over the past centuries, climate change has had a great influence on natural ecosystems and social economic, so studies on temperature have become increasingly important in recent years. Stockholm temperature has been recorded for a long time from 1756 to 2007. Li (2009) attempted to check whether the Stockholm monthly temperature series can be analyzed by a statistical method, and tried to build general linear (GLM) and ARIMA models to fit the data. Data used in that study has been adjusted by Anders Moberg and his colleagues Li (2009). Based on the features of the data, they divided the time period between the year of 1756 and that of 2007 into three periods: 1756-1925, 1926-1985 and 1986-2007, and tried to build GLM and ARIMA models to fit the data in the three periods. Then they forecasted the monthly temperature of 2008 and compared them with the true values. They compared the results and found that the Seasonal ARIMA (SARIMA) model for the series fitted the data better than the general linear model (Li, 2009).

Rangsan et al. (2006) studied a model for forecasting oil palm price of Thailand in three types as farm price, wholesale price and pure oil price for the period of five years: 2000 – 2004. The objective of the research was to find an appropriate ARIMA Model for forecasting three types of oil palm price by considering the minimum of mean absolute percentage error (MAPE). The results of forecasting were as follows: ARIMA Model for forecasting farm price of oil palm is ARIMA(2,1,0), ARIMA Model for forecasting wholesale price of oil palm is ARIMA (1,0,1) or RMA(1,1), and ARIMA Model for forecasting pure oil price of oil palm is ARIMA (3,0,0) or AR(3) (Rangsan et al, 2006).

Forecasting of energy demand in emerging markets is one of the most important policy tools used by the decision makers all over the world. In Turkey, most of the early studies used include various forms of econometric modeling. However, since the estimated economic and demographic parameters usually deviate from the realizations, time-series forecasting appears to give better results. Volkan et al. (2006) used the Autoregressive Integrated Moving Average (ARIMA) and seasonal ARIMA (SARIMA) methods to estimate the future primary energy demand of Turkey from 2005 to 2020. The ARIMA forecasting of the total primary energy demand appears to be more reliable than the summation of the individual forecasts. The results have shown that the average annual growth rates of individual energy sources and total primary energy will decrease in all cases except wood and animal–plant remains which will have negative growth rates. The decrease in the rate of energy demand may be interpreted that the energy intensity peak will be achieved in the coming decades. Another interpretation is that any decrease in energy demand will slow down the economic growth during the forecasted period. Rates of changes and reserves in the fossil fuels indicate that inter-fuel substitution should be made leading to a best mix of the country's energy system (Volkan et al., 2006).

Univariate Box-Jenkins time-series analysis has been used for modeling and forecasting monthly domestic electric energy consumption in the Eastern Province of Saudi Arabia (Aal et al, 1998). Autoregressive integrated moving average (ARIMA) models were developed using data for 5 yr and evaluated on forecasting new data for the sixth year. The optimum model derived is a multiplicative combination of seasonal and non-seasonal autoregressive parts, each being of the first order, following first differencing at both the seasonal and non-seasonal levels. Compared to regression and adductive network machine-learning models previously developed on the same data, ARIMA models require less data, have fewer coefficients, and are more accurate. The optimum ARIMA model forecasts monthly data for the evaluation year with an average percentage error of 3.8% compared to 8.1% and 5.6% for the best multiple-series regression and abductory induction mechanism (AIM) models, respectively; the mean-square forecasting error is reduced with the ARIMA model by factors of 3.2 and 1.6, respectively (Aal et al, 1998).

Liu et al. (1991) study the consumption of natural gas in Taiwan within the residential sector. In this study, the authors explore the dynamic relationships among several potentially relevant time series variables and develop appropriate models for forecasting. It is apparent that the temperature of service areas and the price of natural gas are important factors in forecasting the residential consumption of natural gas. Because of the government price control policy, however, they found that the price variable employed in modeling and forecasting of natural gas consumption needs to be used judiciously. Otherwise, inappropriate models and poor forecasts may occur. They also study the inclusion of the price variable using an intervention model and an outlier detection and adjustment method. They found that, both approaches provide more accurate forecasts and reveal useful information on the dynamics of the controlled variable. Both monthly and quarterly time series of the data are studied. It is easier to obtain appropriate models using quarterly data. However, the performance of quarterly models may not be as good as that of monthly models. However, the loss of performance efficiency in using quarterly data is not too great. This is probably due to the fact that the consumption of natural gas is subject to moving holiday effects and the use of quarterly data may conveniently avoid such systematic disturbances (Liu et al, 1991).

The total consumption of electricity and petroleum energies accounts for almost 90% of the total energy consumption in Taiwan, so it is critical to model and forecast them accurately. For univariate modeling, Pao (2009) proposes two new hybrid nonlinear models that combine a linear model with an artificial neural network (ANN) to develop adjusted forecasts, taking into account heteroscedasticity in the model's input. Both of the hybrid models can decrease round-off and prediction errors for multi-step-ahead forecasting. The results suggest that the new hybrid model generally produces forecasts which, on the basis of out-of-sample forecast encompassing tests and comparisons of three different statistic measures, routinely dominate the forecasts from conventional linear models. The superiority of the hybrid ANNs is due to their flexibility to account for potentially complex nonlinear relationships that are not easily captured by linear models. Furthermore, all of the linear and nonlinear models have highly accurate forecasts, since the mean absolute percentage forecast error (MAPE) results are less than 5%. Overall, the inclusion of heteroscedastic variations in the input layer of the hybrid univariate model could help improve the modeling accuracy for multi-step-ahead forecasting (Pao, 2009).

Lorek and Willinger (1995) provide evidence on the time-series properties and predictive ability of cash-flow data. It employs a sample of firms on which the accuracy of one-step-ahead cash-flow predictions is assessed during the 1989-1991 holdout period. They develop a new multivariate, time-series prediction model that employs past value of earnings, short-term accruals and cash-flows as independent variables in a time-series regression. The predictive results indicate that this model clearly outperforms firm-specific and common-structure ARIMA models as well as a multivariate, cross-sectional regression model popularized in the literature.

These findings are robust across alternative cash-flow metrics (levels, per-share, and deflated by total assets) and are considerate of earnings and accrual accounting data (Lorek & Willinger, 1995).

Apley and Shi (2007) propose an on-line Statistical Process Control (SPC) technique, based on a Generalized Likelihood Ratio Test (GLRT), for detecting and estimating mean shifts in auto-correlated processes that follow a normally distributed Autoregressive Integrated Moving Average (ARIMA) model. The GLRT is applied to the uncorrelated residuals of the appropriate time-series model. The performance of the GLRT is compared to two other commonly applied residual-based tests - a Shewhart individuals chart and a CUSUM test. A wide range of ARIMA models are considered, with the conclusion that the best residual-based test to use depends on the particular ARIMA model used to describe the autocorrelation. For many models, the GLRT performance is far superior to either a CUSUM or Shewhart test, while for others the difference is negligible or the CUSUM test performs slightly better. Simple, intuitive guidelines are provided for determining which residual-based test to use. Additional advantages of the GLRT are that it directly provides estimates of the magnitude and time of occurrence of the mean shift, and can be used to distinguish different types of faults, e.g., a sustained mean shift versus a temporary spike (Apley and Shi, 2007).

Bao (2006) reviews research that makes use of one of the most popular forecasting methods applied in accounting: time-series analysis using the Box-Jenkins methodology. It organizes the research in the area, surveys recent applications of time-series analysis in accounting, and discusses the potential for the methodology in addressing future research issues. The emphasis is on those aspects of the accounting system that possibly cause difficulties in applying time-series methods in accounting.

A study by Volkan et al. (2006) aims at forecasting the most possible curve for domestic fossil fuel production of Turkey to help policy makers to develop policy implications for rapidly growing dependency problem on imported fossil fuels. The fossil fuel dependency problem is international in scope and context and Turkey is a typical example for emerging energy markets of the developing world. Volkan et al. (2006) developed a decision support system for forecasting fossil fuel production by applying a regression, ARIMA and SARIMA method to the historical data from 1950 to 2003 in a comparative manner. The method integrates each model by using some decision parameters related to goodness-of-fit and confidence interval, behavior of the curve, and reserves. Different forecasting models are proposed for different fossil fuel types. The best result is obtained for oil since the reserve classifications used it is much better defined them for the others. Their findings show that the fossil fuel production peak has already been reached; indicating the total fossil fuel production of the country will diminish and theoretically will end in 2038. However, production is expected to end in 2019 for hard coal, in 2024 for natural gas, in 2029 for oil and 2031 for asphaltite. The gap between the fossil fuel consumption and production is growing enormously and it reaches in 2030 to approximately twice of what it is in 2000.

Statistical control chart is commonly used in the industry to help ensure stability of manufacturing process and it can also be used to monitor the environmental data, such as industrial waste or effluent of manufacturing process. However, control chart needs to be modified if the set of environmental data exhibits the property of long memory. In (Jen-Nan, 2007), a control chart for auto-correlated data using autoregressive fractionally integrated moving-average (ARFIMA) model is proposed to monitor the long-memory air quality data. Finally, Jen-Nan (2007) used the air quality data of Taiwan as examples to compare the difference between ARFIMA and autoregressive integrated moving-average (ARIMA) models. The results show that residual control charts using ARFIMA models are more appropriate than those using ARIMA models.

Bao et al. (2006) reviews research that makes use of one of the most popular forecasting methods applied in accounting: time-series analysis using the Box-Jenkins methodology. It organizes the research in the area, surveys recent applications of time-series analysis in accounting, and discusses the potential for the methodology in addressing future research issues. The emphasis is on those aspects of the accounting system that possibly cause difficulties in applying time-series methods in accounting.

Al-Zeaud (2011) presents the Box-Jenkins model as one of the forecasting techniques, which can be used on financial time series. The main aim is to predict the volatility for the bankjng sector. That is achieved by finding the tentative Autoregressive Integrated Moving Average (ARIMA) models that describe the equation of the forecasting for the banking sector. The data are accumulated weekly from the web site of Amman Stock Exchange (ASE) using the historical indices in the period from1/1/2005-1/4/2010. The number of the integrated equations is tested by using co- integration test, stationary test by using unit root, and then use a minimum mean square error(MSE), t-statistics value and p-statistics value to choose the best ARIMA models at 95% confidence interval. The results show that the best model for banks sector is ARIMA (2,0,2), since this model gives the minimum mean square error which is 0.0001003, then ARIMA (1,1,1).

Maia (2008) presents approaches to interval-valued time series forecasting. The first and second approaches are based on the autoregressive (AR) and autoregressive integrated moving average (ARIMA) models, respectively. The third approach is based on an artificial neural network (ANN) model and the last is based on a hybrid methodology that combines both ARIMA and ANN models. Each approach fits, respectively, two models on the mid-point and range of the interval values assumed by the interval-valued time series in the learning set. The forecasting of the lower and upper bounds of the interval value of the time series is accomplished through a combination of forecasts from the mid-point and range of the interval values. The evaluation of the models presented is based on the estimation of the average behaviour of the mean absolute error and mean squared error in the framework of a Monte Carlo experiment. The results demonstrate that the approaches are useful in forecasting alternatives for interval-valued time series and indicate that the hybrid model is an effective way to improve the forecasting accuracy achieved by any one of the models separately.

Air quality time series consists of complex linear and non-linear patterns and are difficult to forecast. Box–Jenkins Time Series (ARIMA) and multilinear regression (MLR) models have been applied to air quality forecasting in urban areas, but they have limited accuracy owing to their inability to predict extreme events. Artificial neural networks (ANN) can recognize non-linear patterns that include extremes. A novel hybrid model combining ARIMA and ANN to improve forecast accuracy for an area with limited air quality and meteorological data was applied to Temuco, Chile, where residential wood burning is a major pollution source during cold winters, using surface meteorological and PM10 measurements (Diaz-Robles et al, 2008). Experimental results indicated that the hybrid model can be an effective tool to improve the PM10 forecasting accuracy obtained by either of the models used separately, and compared with a deterministic MLR. The hybrid model was able to capture 100% and 80% of alert and pre-emergency episodes, respectively. This approach demonstrates the potential to be applied to air quality forecasting in other cities and countries (Diaz-Robles et al, 2008).

A study by (Kumar and Jain, 2009) applies three time series models, namely, Grey-Markov model, Grey-Model with rolling mechanism, and singular spectrum analysis (SSA) to forecast the consumption of conventional energy in India. Grey-Markov model has been employed to forecast crude-petroleum consumption while Grey-Model with rolling mechanism to forecast coal, electricity (in utilities) consumption and SSA to predict natural gas consumption. The models for each time series were selected by carefully examining the structure of the individual time series. The mean absolute percentage errors (MAPE) for two out of sample forecasts were obtained as follows: 1.6% for crude-petroleum, 3.5% for coal, 3.4% for electricity and 3.4% for natural gas consumption. For two out of sample forecasts, the prediction accuracy for coal consumption was 97.9%, 95.4% while for electricity consumption the prediction accuracy was 96.9%, 95.1%. Similarly, the prediction accuracy for crude-petroleum consumption was found to be 99.2%, 97.6% while for natural gas consumption these values were 98.6%, 94.5%. The results obtained have also been compared with those of Planning Commission of India's projection. The comparison clearly points to the enormous potential that these time series models possess in energy consumption forecasting and can be considered as a viable alternative (Kumar and Jain, 2009).

Uri and Flanagan (2003).detail the Box-Jenkins approach to forecasting time series and apply it to short-term natural gas marketed production and crude petroleum production in the United States. After establishing the efficacy of the approach for forecasting the two series of interest, monthly forecasts for 1978 are made. The results indicate that natural gas production in 1978 will increase by 2·8 per cent over the 1977 level while crude petroleum production will fall by 4·0 per cent.

In (Greenea and Chih-Kang, 1983), ARIMA time series model building techniques are used to construct fifty-one state gasoline demand models based on monthly data for the period of January, 1975 to July, 1960. Statistically satisfactory models are obtained for all states. Price elasticity estimates are >0 for all states. All but four are statistically significant at the 0.05 level. The significant price elasticity estimates range from −0.138 to −0.377, with most clustering about −0.2. Estimates of state gasoline supply shortages for May, June, and July, 1979 are also presented which range from 0 to 8 percent of normal consumption for the three-month-period.

Lon-Mu (2006) studies the dynamic relationships between US gasoline prices, crude oil prices, and the stock of gasoline. Using monthly data between January 1973 and December 1987, they found that the US gasoline price is mainly influenced by the price of crude oil. The stock of gasoline has little or no influence on the price of gasoline during the period before the second energy crisis, and seems to have some influence during the period after. Lon-Mu (2006) also finds that the dynamic relationship between the prices of gasoline and crude oil changes over time, shifting from a longer lag response to a shorter lag response. Box-Jenkins ARIMA and transfer function models are employed in this study. These models were estimated using estimation procedure with and without outlier adjustment. For model estimation with outlier adjustment, an iterative procedure for the joint estimation of model parameters and outlier effects is employed. The forecasting performance of these models is carefully examined. For the purpose of illustration, Lon-Mu (2006) also analyzes these time series using classical white-noise regression models. The results show the importance of using appropriate time-series methods in modeling and forecasting when the data are serially correlated. This paper also demonstrates the problems of time-series modeling when outliers are present.

Financial theory predicts that a change in an exchange rate should affect the value of a firm or an industry. To a large extent, past research has not supported this theory, which is surprising especially after considering the substantial exchange rate fluctuations over the three decades. A study by El Masry(2006) seeks to extend previous research on the foreign exchange rate exposure of UK non-financial companies at the industry level over the period of 1981-2001. In this study, exchange rate exposure was defined as the change in the value of the firm or industry due to the changes in exchange rates. This study differs from previous studies in that it considers the impact of the changes (actual and unexpected) in exchange rates on firms' or industries' stock returns. The approach employs OLS model to estimate foreign exchange rate exposure of 364 UK nonfinancial companies over the period 1981-2001. The findings indicated that a higher percentage of UK industries were exposed to contemporaneous exchange rate changes than those reported in previous studies. There was also evidence of significant lagged exchange rate exposure. This lagged exchange rate exposure is consistent with findings in previous studies that may exist in some market inefficiencies in incorporating exchange rate changes into the returns of firms and industries (El Masry, 2006).

The influence of economic and demographic variables on the annual electricity consumption in Italy has been investigated with the intention to develop a long-term consumption forecasting model. The time period considered for the historical data is from 1970 to 2007. Different regression models were developed, using historical electricity consumption, gross domestic product (GDP), gross domestic product per capita (GDP per capita) and population. (Bianco et al, 2009) first consider the estimation of GDP, price and GDP per capita elasticities of domestic and non-domestic electricity consumption. The domestic and non-domestic short run price elasticities are found to be both approximately equal to −0.06, while long run elasticities are equal to −0.24 and −0.09, respectively. On the contrary, the elasticities of GDP and GDP per capita present higher values. In the second part of (Bianco et al, 2009), different regression models, based on co-integrated or stationary data, are presented. Different statistical tests are employed to check the validity of the proposed models. A comparison with national forecasts, based on complex econometric models, such as Markal-Time, was performed, showing that the developed regressions are congruent with the official projections, with deviations of ±1% for the best case and ±11% for the worst. These deviations are to be considered acceptable in relation to the time span taken into account.

The accelerating use of fossil fuels since the Industrial Revolution and the rapid destruction of forests causes a significant increase in greenhouse gases. The increasing threat of global warming and climate change has been the major, worldwide, ongoing concern especially in the last two decades. The impacts of global warming on the world economy have been assessed intensively by researchers since the 1990s. Worldwide organizations have been attempting to reduce the adverse impacts of global warming through intergovernmental and binding agreements. Carbon dioxide (CO_2) is one of the most foremost greenhouse gases in the atmosphere. The energy sector is dominated by the direct combustion of fuels, a process leading to large emissions of CO_2. CO_2 from energy represents about 60% of the anthropogenic greenhouse gas emissions of global emissions. This percentage varies greatly by country, due to diverse national energy structures. The top-25 emitting countries accounted 82.27% of the world CO_2 emissions in 2007. In the same year China was the largest emitter and generated 20.96% of the world total. Trend analysis is based on the idea that what has happened in the past gives traders an idea of what will happen in the future. In this study, trend analysis approach has been employed for modelling to forecast of energy-related CO2 emissions. To this aim first, trends in

CO_2 emissions for the top-25 countries and the world total CO_2 emissions during 1971–2007 are identified (Kone and Buke, 2010). On developing the regression analyses, the regression analyses with R^2 values less than 0.94 showing insignificant influence in statistical tests have been discarded. Statistically significant trends are indicated in eleven countries namely, India, South Korea, Islamic Republic of Iran, Mexico, Australia, Indonesia, Saudi Arabia, Brazil, South Africa, Taiwan, Turkey and the world total. The results obtained from the analyses showed that the models for those countries can be used for CO_2 emission projections into the future planning. The calculated results for CO_2 emissions from fitted curves have been compared with the projected CO_2 emissions given in International Energy Outlook 2009 of U.S. Department of Energy calculated from "high economic growth case scenario", "reference case scenario" and "low economic growth case scenario" respectively. Agreements between calculated results and the projected CO_2 emissions from different scenarios are in the acceptable range (Kone and Buke, 2010).

The oil industry has used decline curve analysis with limited success in estimating crude oil reserves and in predicting future behaviour of oil and gas wells. Ayeni and Pilat (1991), therefore, explored the possibility of using the Autoregressive Integrated Moving Average (ARIMA) technique in forecasting and estimating crude oil reserves. The authors compared this approach with the traditional decline method using real oil production data from twelve (12) oil wells in South Louisiana. The Box and Jenkins methodology was used to develop forecast functions for the twelve wells under study. These forecast functions were used to predict future oil productions. The forecast values generated were then used to determine the remaining crude oil reserves for each well. The accuracy of the forecasts relative to the actual values for both ARIMA and decline curve methods is determined by various statistical error analyses. The conditions, under which one method gives better results than the other, were fully investigated. In almost all the cases considered, the ARIMA method is found to perform better than the decline curve method.

CHAPTER 3

METHODOLOGY

3.0 Introduction

This chapter deals with the methodology for this study and it looks at how the Box-Jenkins ARIMA models for time series are used to analyse past data and the effect they have on current and future values of such data.

3.1 Time Series

A time series is a set of statistics, usually collected at regular intervals. Time series data occur naturally in many application areas. Therefore, a time series is an ordered sequence of values of a variable at equally spaced time intervals.

3.2 Objectives of Time Series

An observed time series can be assumed as the realization of a stochastic process. Once we understand how the process operates, we can develop a mathematical model to predict the future values of the time series. Thus, there are two main objectives of time series analysis:

1. To understand the underlying structure of the time series by breaking it down to its components,

2. To fit a mathematical model and then proceed to forecast the future

3.3 Time Series Components

Original time series data made up of various patterns which are identified by time series analysis methods. There are two separate components of the basic underlying pattern that tend to characterize economics and business series; these are trend-cycle and the seasonal factors. The trend-cycle is sometimes separated into trend, cyclical or periodic components and an error component. In general, a time series data is in the form;

treated as the sum of a systematic part or trend and a random part or irregular. This model can be written as;

3.3.3.2 Seasonal variations

There could be periodic, repetitive variations in time-series which occur because of buying or consuming patterns and social habits, during different times of a year.

3.3.3.3 Cyclical variations

These refer to the variations in time series which arise out of the phenomenon of business cycles. A cycle refers to the periods of expansion followed by periods of contraction in a time series.

3.3.3.4 Random or irregular variations

These refer to the erratic fluctuations in the data which cannot be attributed to the trend, seasonal or cyclical factors. In many cases, the root cause of these variations can be isolated only after a detailed analysis of the data and the accompanying explanations, if any. Such variations can be due to a wide variety of factors like sudden weather changes, strike, and price hike in petroleum products.

3.4 Stationarity of Time Series Data

A time series is considered stationary if its sample mean and variance are not significantly different, in the statistical sense, for any major subsets of the series. That is to say that, a stationary time series is one whose statistical properties such as mean, variance, autocorrelation, etc. are all constant over time. A stationarized series is relatively easy to predict: you simply predict that its statistical properties will be the same in the future as they have been in the past. Therefore, one needs to only plot the time series and observe the following;

- If the mean of the plotted series varies over time, the series is considered non-stationary in mean. If there is no evidence of a change in mean level over time, then the series is considered mean-stationary.
- If the plotted series shows no obvious change in the variance over time, then the series is considered to be stationary in variance, otherwise it is considered to be non-stationary in variance.

3.4.1 Testing for non-stationarity

1**.** Autocorrelation function (Box-Jenkins approach)-if autocorrelations start high and decline slowly, then series is non-stationary, and should be differenced.

2. Dickey-Fuller test

to 1 indicates a strong, positive correlation; a value close to -1 indicates a strong negative correlation; and a value close to 0 indicates weak or no correlation. The sample ACF at lag k is the autocovariance of the observations at lag k normalized by the sample covariance of the time series

$E[z]$ is the expected value

solved for

persistence, in a time series is the percentage of the series variance that is reduced by fitting the series to an ARMA model (Anderson, O., 1976).

The graph of the sample autocorrelation function (ACF) and the sample partial autocorrelation function (PACF) can be used to determine the model whose processes can be summarized as follows:

Model	**ACF**	**PACF**
AR(p)	Dies down	Cut off after lag q
MA(q)	Cut off after lag p	Dies down
ARMA(p,q)	Dies down	Dies down

Table 3.1: How to determine the model by using ACF and PACF patterns

3.6.4 Autoregressive Average Integrated Moving Average (ARIMA) Models

Time series are naturally non-stationary though some of them are stationary, in order to induce stationarity in the non-stationary data, a concept called differencing is used. After modeling the d^{th} order differenced series with an appropriate ARMA model, to reclaim the modelled values corresponding to the original un-differenced series, it is necessary to reverse the differencing transformation and "integrate" d times. This is represented by "I" in the acronym ARIMA and the order of integration is same as the order of differencing. If p is the order of the AR model, q the order of the MA model and d the number of differencing needed to make a time series data stationary, then the ARIMA model involved is defined as ARIMA(p,d,q).

In terms of the lagged terms, L, involved in the time series itself and the residuals, the ARIMA(p,d,q) models can be defined as follows;

Let

3.6.5.1 Identification of Model

Identification stage consists of specifying the AR, I, and MA orders (p,d,q). That is, the autoregressive order, moving average order, or degree of differencing required to induce stationarity. It has been found that, in practice, adequate models rarely have values of p, d, and q greater than two. The basic tools for model identification are the graphs of the sample autocorrelation function (ACF) and sample partial autocorrelation function (PACF) obtained from the series. The ACF (correlogram) indicates the degree of correlation within the series for lags 1, 2, 3, . . . etc . Similarly, the PACF indicates the degree of correlation at a given lag after accounting for the correlation from the intervening lags (Pankratz, 1983). The ACF and PACF are plotted as spikes occurring at each lag order, if a spike lies outside the confidence limit lines, the correlation at that lag is significant. To determine the order of differencing d, the time series must be checked for non-stationarity. If non-stationarity is indicated, differencing or other transformations must be performed prior to further analysis. There are basically two methods currently in use by practitioners (Ali and Thalheimer, 1983). One is to simply inspect the plotted time series for shifts in level or increasing variability. The other involves examination of the ACF. If the ACF spikes fail to die out rapidly or remain statistically significant at high lag orders, differencing may be required. The required order of differencing determines d. To determine the AR and MA orders p and q, inspection of the ACF and PACF of the series (or differenced series, if called for) is performed. Theoretically, the number of successive ACF spikes at lags greater than zero equals the order of the moving average component, q and the number of significant PACF spikes at lag orders greater than zero indicates the order of the autoregressive component, p, thus making the ACF and PACF for a particular ARIMA(p,d,q) unique. In addition, other patterns in the ACF and PACF help validate these tentative indications.

This makes proponents of Box-Jenkins models proclaim that the methodology is superior to other modelling techniques because it "lets the data speak for themselves," rather than imposing a specific model form onto the data, (Reagan, 1984).

3.6.5.2 Estimation of Model

After a tentative model has been identified, the AR and/or MA parameters b and β are estimated from the time series data using an efficient nonlinear least-squares algorithm. The residuals, that is, the differences between the observed time series values and the model calculated or "fitted" values are also obtained at this stage. The least-squares estimates of b and β are those values which minimize the sum of the squared residuals. The model-calculated values are found by inserting initial estimates that can be used to solve for the AR and MA coefficients using nonlinear square estimation includes: the Maximum Likelihood Method, Unconditional Least Squares Method and the Conditional Least Squares Method In this work, the estimation of parameters was performed using the R software package.

If the tentative model has significant parameters, whose values lie within the bounds of stationarity and invertibility and are not highly correlated, then the analyst may proceed to the last stage, diagnostic checking. If not, the analyst must return to the identification stage and formulate an alternate model based on the information gained at the estimation stage.

3.6.5.3 Diagnostic Model

More than one tentative ARIMA model will be fitted to the data, estimate the parameters for each model and then perform a diagnostic check to test the validity of each model. The model which fits the best according to various statistical tests of fit is then selected for forecasting. To perform a diagnostic check the following will be considered.

a. A study of the residual series obtained after fitting the model to the data to see if any pattern remains unaccounted for and is fitted out of the autocorrelation structure, leaving uncorrelated residuals. The ACF and PACF plots of the residual series help in detecting any unaccounted pattern.

Hence the diagnostic checking stage consists of verifying that the residuals obtained at the estimation stage are white noise with mean zero and constant variance.

b. A study of the sampling statistics of the current optimum solution to check if any further simplification of the model is possible.

The following statistical tests for lack of fit were used in the work to check for the randomness of the residuals:

1. ACF and PACF plots of the residuals: The ACF of the residuals obtained after fitting a proper model to the data must show no significant autocorrelations at any lag order. Similarly, the PACF plot of the residuals must show no significant spikes at any lag order. Absence of any significant spikes in the residual ACF and PACF plots demonstrate proper fitting. However, there may be a few spikes which are close to significance.

2. Ljung-Box Chi-Square test: Another measure of check for the randomness of residuals is using the Ljung-Box Chi-Square test and this is used to test the normality of the residuals of the model and this must show a p-value greater than 0.05 otherwise the residuals by the model do are said to be not independent and identically distributed (i.i.d). The null hypothesis is that the set of autocorrelations for residuals is white noise. This statistic measures the significance of residual autocorrelations as a set and points out if they are collectively significant:

H_0: The data is random

H_1: The data is not random

Information Criteria (BIC). The AIC and BIC are used to compare competing models fit to the same series. The model with smaller AIC and BIC values is a statistically better fit.

1) Akaike's Information Criteria (AIC): It is a statistical tool for model selection and is grounded in the concept of randomness. It can be non-statistically described as a measure of trade-off between the precision and complexity of the model. The absolute value of AIC is not useful; the relative comparison of AIC values of different competing models can be used to infer the best model. The model with lowest AIC value is the best fit.

It is computed as:

For normally and independently distributed residuals, in the data, as evidenced from the fact that the autocorrelation coefficients at the seasonal lags of ACF plot will not die out rapidly, proper order of seasonal differencing (denoted by '*D* ') may be required to make the data seasonal stationary. Secondly, the presence of seasonal autoregressive and moving average coefficients in the data needs to be determined on similar lines as was discussed for the non-seasonal ARIMA model identification, but with using the autocorrelation coefficients of ACF and PACF plots at the seasonal lags. The general notation for seasonal ARIMA model is ARIMA (*P, D Q*), where '*P*' is the order of seasonal autoregressive component, '*Q*' is the order of seasonal moving average coefficient and '*D'* is the order of seasonal differencing used. In general, a time series often may contain both non-seasonal and seasonal components. Though the time series may be deseasonalized and a non-seasonal ARIMA model maybe fitted to the remainder, experience suggests that Box-Jenkins methodology provides good forecasts of periodic data series (Makridakis and Hibon, 2000). Thus, it may be advisable to leave the seasonal component in the data and fit a general class of ARIMA model which accounts for both seasonality and non-seasonality. Such a general ARIMA model can be represented by the form ARIMA $(p, d, q)(P,D,Q)^s$. This is commonly referred to as a seasonal ARIMA multiplicative model and it is represented by;

CHAPTER 4

DATA ANALYSIS AND RESULTS

4.0 Introduction

This chapter presents the analysis of some petroleum products consumption in the country. These include all quantities supplied from the Tema Oil Refinery (TOR) for distribution in the various regions of the country for industrial, domestic and commercial use. The R statistical software was used for the analysis and various tentative time series ARIMA models developed were fitted to each data and the suitable models were selected based on diagnostics of the residuals of the various models.

4.1 Data Presentation

The data which is made up of some of the petroleum products consumed in the country namely diesel, premix petrol (premix) and liquefied petroleum gas (LPG) was obtained from the National Petroleum Authority (NPA). It consists of quantities of the three products in litres for the diesel and the premix and kilograms (kg) for the LPG consumed by the nation from 1999 to November 201, compiled monthly. See the appendix for the data presentation.

4.2 Data Aggregation

Each of the three data was divided into two, an initialization set which was used to formulate the appropriate models for forecasting whiles the second set called the test set was used to check the validity of the chosen models. This is made to test the model for how accurate it is to help in the forecasting.

4.3 Computational Procedure

The R software was installed on HP 2000 Notebook PC

Processor: AMD E-350 Processor1.6Hz

Installed Memory (RAM): 3.00GB (2.60GB usable)

A programme was written in R language by the researcher to analyse the data collected from the National Petroleum Authority and using the R software, a time series analysis was conducted on the data.

4.4 Descriptive Analysis

Several plots were made using the R software on the diesel, the premix petrol and the LPG data. These plots involve time series plot of the three data sets, their ACF and PACF plots, their differencing plots and their diagnostics plots.

4..1 Time Series Plot of Data

The descriptive time series plots of the National Gas Oil Demand, National LPG Demand and Premix data are as follows:

4..2 Natioanal Gas Oil Data Analysis

The national gas oil data is analysed based on the monthly primary data obtained from the National Petroleum Authority. This is a compilation of gas oil supply or demand mainly by commercial and domestic users for transportation and agricultural purposes in the country.

4..2.1 Descriptive Analysis of the Gas Oil Demand In Ghana

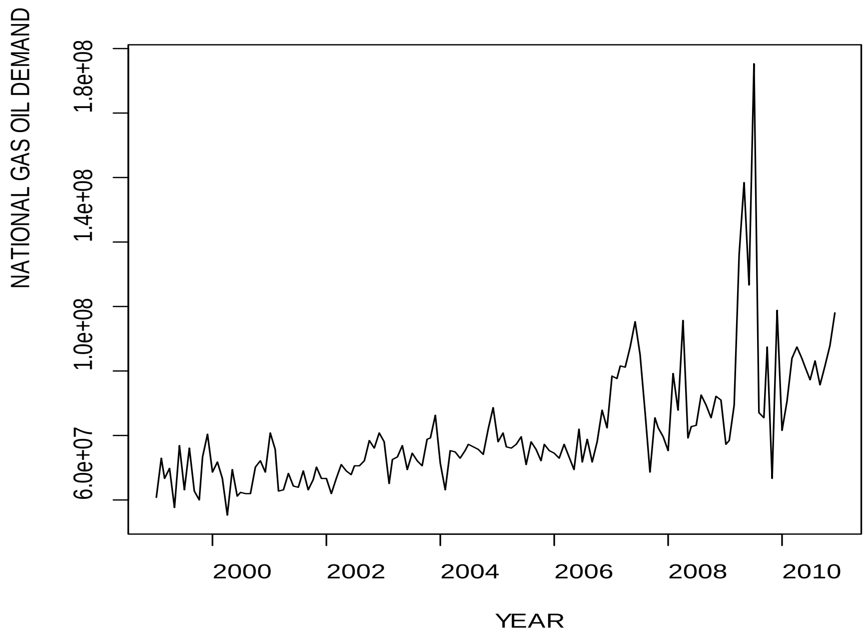


Figure 4.1: The time series plot of the national domestic gas oil consumption in Ghana from 1999 to 2010.

Figure 4.1 is the gas oil data time graph above shows a set of spikes and troughs which keeps rising though there are few downward surges in the pattern. These shows there are both slight seasonality and trend in the data. From 1999, there was an increase in the demand pattern from January to December 1999 followed by a sudden decrease that rose after the second month in 2000, the third month experienced almost a constant stability and this gave way to a rising pattern that continued till the around the beginning of 2001 and fell seriously and kept an undertone rise at that level till about the end of the first month in 2002 and a steady rise in the demand for gas oil till about the end of 2002, there was a little rise in demand for gas oil at the beginning of 2003 and this steadily rise-fall again with demand increasing at the latter part of 2003 but dropped seriously at the end of 2003 down to about the end of January 2004 and continued this part of rise and fall till about the end of 2006. Demand for gas oil in 2007 was very high as compared to the kind of rise experienced in the previous years. This rise continued through to the middle of 2007 and started reducing drastically till the beginning of the last quarter of 2007 and started its rise-fall pattern with the demand still being higher than that of the previous years. Around the beginning of 2009, there was a hooping rise in the demand of gas oil and this continued rapidly for the next four months and dropped considerably in the next two months of 2009 and rapidly fell at the end of the third quarter and gained a record high demand at around the beginning of the final quarter of 2009 and fell so low in the same quarter, a kind of drop in demand that has never been experienced with respect to the data at hand. There was a sharp increment in the demand for gas oil and a sharp drop down that gave one of the least in the last quarter of 2009. This was followed by a rise in demand and a continuous rise and fall pattern in the demand for gas oil in 2010.

4..2.2 ACF Of Natioanal Gas Oil Demand Data

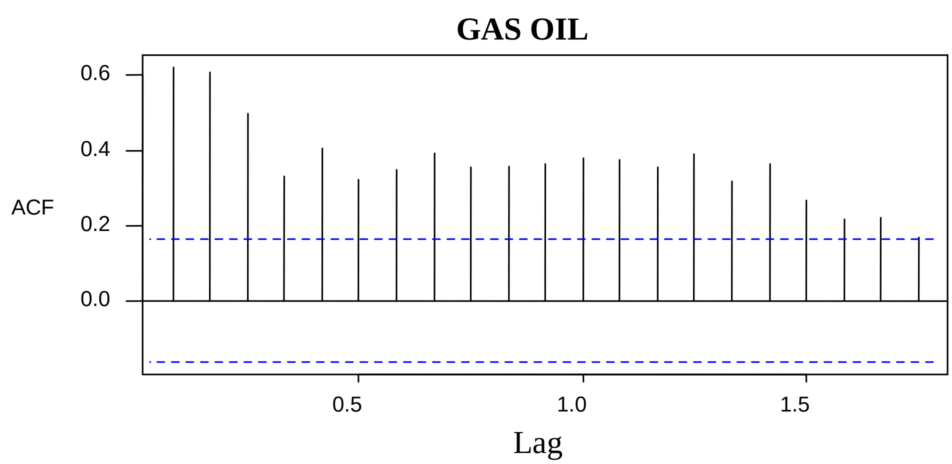


Figure 4.2: ACF of gas oil data

Figure 4.2 shows the autocorrelation function plot of the national domestic gas oil demand data. The autocorrelation function shows the correlation between the national domestic gas oil demand values which is a pattern of a set of decreasing and increasing spikes showing that there are both trend and seasonality in the national domestic gas oil consumption in Ghana. The autocorrelations do show a pattern of decreasing values but in the various decreasing values there are increases though not up to the initial ones and despite the variations, the autocorrelation seem to be decreasing to zero.

4..2.3 Plot of First Differencing

Figure 4.3 below is the time graph of the differenced domestic gas oil data, differencing was done to eliminate trend likely seasonality in the gas oil data.

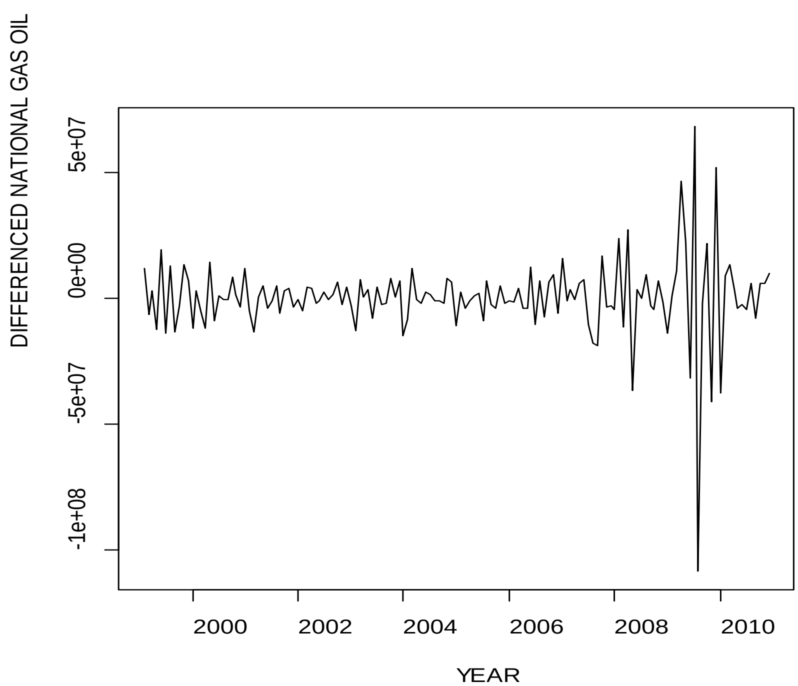


Figure 4.3: Plot of gas oil differenced data

4..2.4 Seasonal and Non-Seasonal Unit Root Test

The differenced national domestic gas oil demand data passes the Dickey-Fuller test for unit root since the Dickey-Fuller value of -7.0884 at lag order of 5 and a p-value of 0.01 which is less than 0.05 and by passing the Dickey-Fuller test the data can also be said to be not white noise, by this, we say there exists dependencies and this needs to be modeled. The KPSS test conducted on the differenced data indicated a p-value of 0.1 which is greater than 0.05, therefore, we do not reject the null hypothesis, the stationarity assumption holds for the series.

4..2.5 ACF and PACF Lags

To generate the candidate models from which the best model can be selected is determined using the ACF and PACF values at various lags, the ACF and PACF values for the first 22 lags are as shown in table 4.2 below;

A table of ACF and PACF Lags

LAG	ACF	PACF		LAG	ACF	PACF
[1,]	-0.49	-0.49		[12,]	-0.02	-0.08
[2,]	0.13	-0.14		[13,]	0.08	0.05
[3,]	0.09	0.13		[14,]	-0.12	-0.16
[4,]	-0.35	-0.31		[15,]	0.17	0.05
[5,]	0.23	-0.13		[16,]	-0.16	-0.09
[6,]	-0.16	-0.11		[17,]	0.08	-0.03
[7,]	-0.03	-0.15		[18,]	0	-0.06
[8,]	0.1	-0.12		[19,]	-0.11	-0.06
[9,]	-0.05	-0.01		[20,]	0.1	-0.11
[10,]	0.01	-0.1		[21,]	-0.07	-0.06
[11,]	0	-0.14		[22,]	-0.03	-0.16

Table 4.1: ACF and PACF values at various lags

4.4.2.6 ACF PLOT and PACF Plot

a.

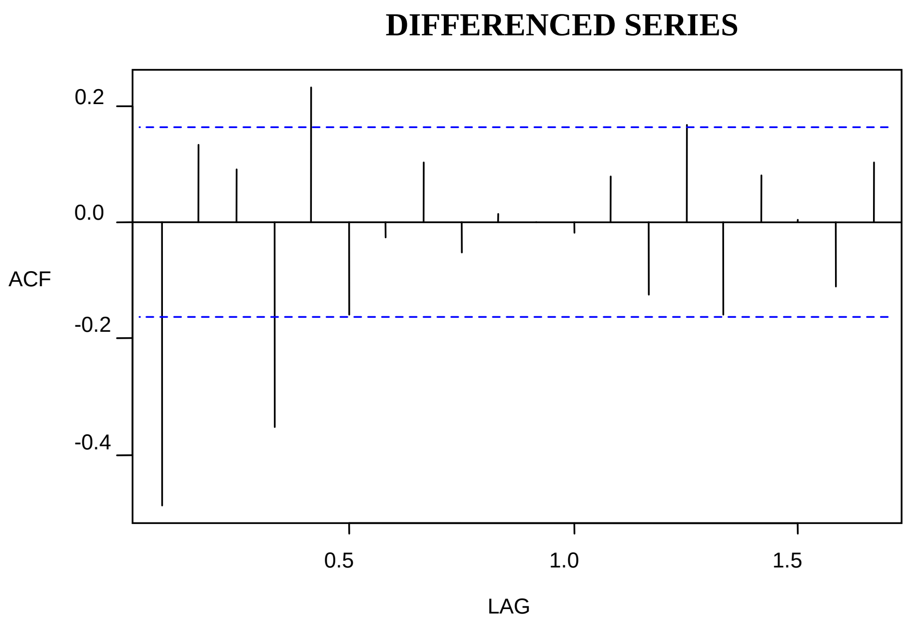


Figure 4.4: ACF plot at various

b.

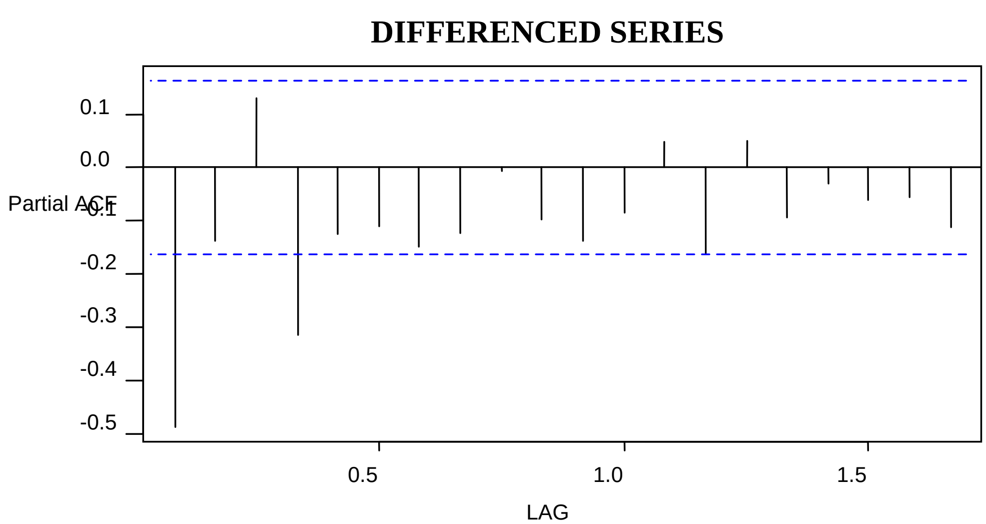


Figure 4.5: PACF plot at various

Figures 4.4 and 4.5 are respectively the ACF and PACF plots of the first differencing of the national domestic gas oil demand data. The top part of the figure is the plots of the autocorrelation function and the down plot is the partial autocorrelation function of the first differencing of the gas oil data at various lags. By comparing the error limits of the autocorrelations, it can be observed that the p value for the AR should not be less than 1. We can use q values starting from 1 upwards. Thus only the autocorrelations at lags 1, 4, 5, 6 and 14 are significant, indicating MA(1) or MA(4) or MA(5) or MA(6) and MA(14) behaviors respectively. Also, comparing the error limits of the partial autocorrelations the only significant partial autocorrelations are those at lags 1, 4, 13 and 21, indicating an AR(1), AR(4) and AR(14) behaviors respectively. By using the parsimony principle AR(1) and MA(1) are selected. The following models are recommended;

- ARIMA(1,1,1)
- ARIMA(0,1,1)
- ARIMA(1,1,0)

To select the best model for forecasting into the future we would need to examine the parameter estimates, diagnostics of the associated residuals and the three constants AIC, AICc and BIC for each model.

4.5 Model Selection for the National Domestic Gas Oil Demand Data

The time series analysis for the national domestic gas oil demand for the chosen candidate models is as shown below:

4.5.1 Parameter Estimates and Diagnostics of ARIMA (1, 1, 1) Model

Parameter estimates determine the coefficients of the time series equation that is generated from the data and the diagnostics test is used to check the correlation and significance of the residuals.

4.5.1.1 Parameter Estimates

The R result of the ARIMA(1,1,1) is as shown below;

```
Coefficients:
          ar1      ma1    intercept
      -0.3909   0.0926     99334.54
s.e.   0.1848   0.1905    152275.03
sigma^2 estimated as 5.356e+12:  log likelihood=-2298.58
AIC=4603.15   AICc=4603.44   BIC=4615
```

To further analyze the results we must appreciate how significant the parameters by taking the t-test on them. The t-test of the parameters are as follows;

The diagnostic test of ARIMA(1,1,1) is as shown in the figure 4.6 below;

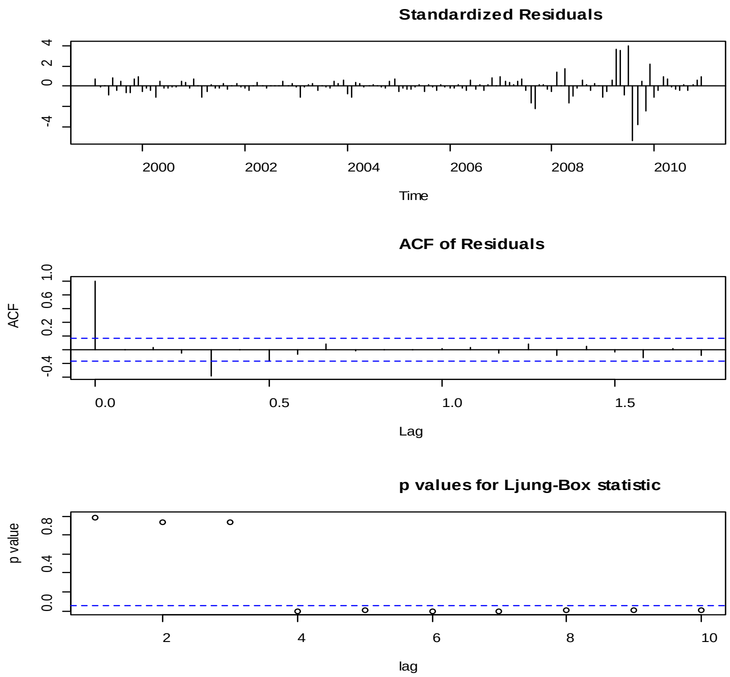


Figure 4.6: Diagnostic test plot of ARIMA(1,1,1)

a. The top box contains the time plot of the standardized residuals of the model which shows that no obvious pattern and looks like an independent identical distribution (i.i.d) sequence with mean zero and few outliers.

b. The middle part is the ACF plot of the residuals and this shows evidence of significance correlation at lag 3 probably due to outliers.

c. The bottom of the figure shows the Ljung-Box plot which suggests that the residuals of the LPG data by the ARIMA(1,1,1) are significant at any positive lag and the p-value of 0.0119 of the residuals by the model confirms the insignificance of the residuals and so, the residuals of the ARIMA(1,1,1) are not independent and identically distributed though the diagrams below shows otherwise.

The distribution of the errors by ARIMA(1,1,1) model is as shown in figure 4.7 below. From the diagram it can be observed that the residuals appear to be normally distributed since most of the data points are on the normal line and those which are not on the line deviate to the similar extent below and above the normal line.

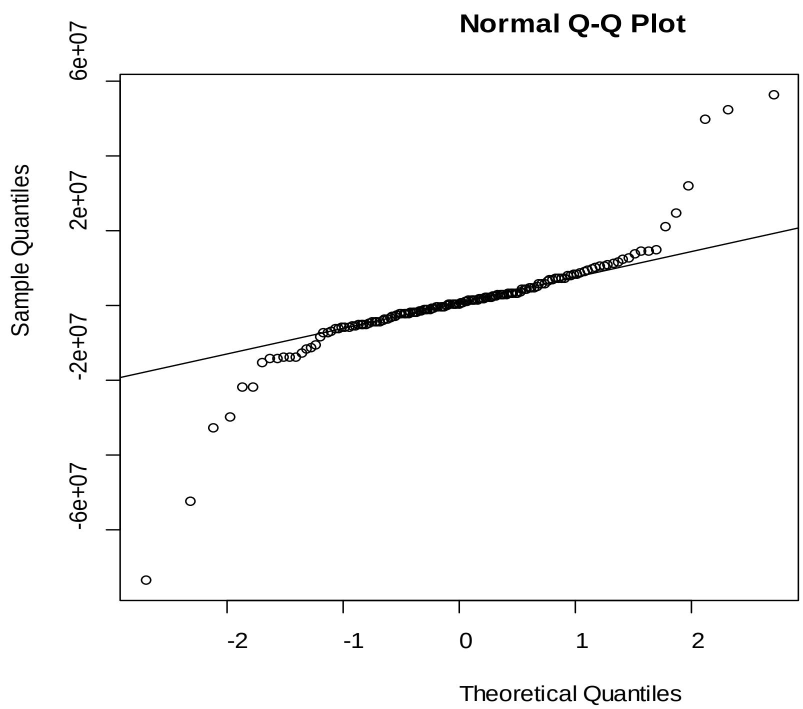


Figure 4.7: Normal Q-Q plot of the residuals of ARIMA(1,1,1)

The histogram plot fitted with the normal curve showing the normality of the residuals for ARIMA(1,1,1) model is as shown below and this confirms that the residuals are normally distributed.

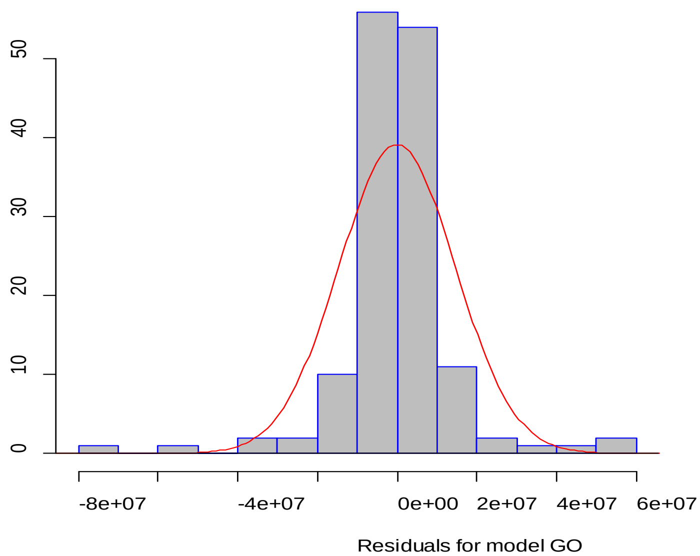


Figure 4.8: Plot of histogram fitted with normal curve of the residuals of ARIMA(1,1,1)

4.5.2 Parameter Estimates and Diagnostic of ARIMA(0,1,1) Model

Parameter estimates determine the coefficients of the time series equation that is generated from the data and the diagnostics test is used to check the correlation and significance of the residuals.

4.5.2.1 Parameter Estimates

The R result of ARIMA(0,1,1)

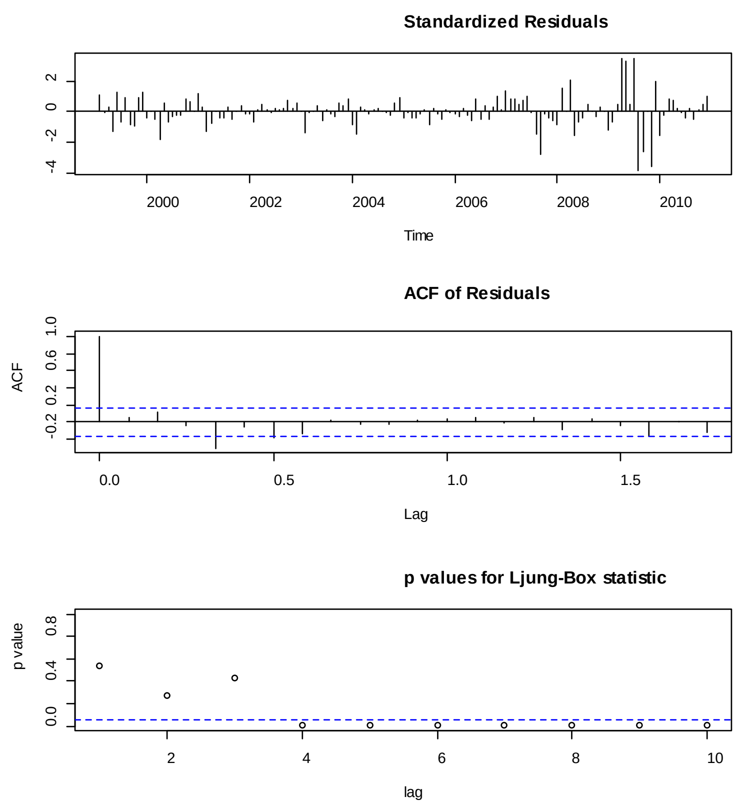
Coefficients:

	ma1	intercept
	-0.2737	100523.5
s.e.	0.0789	142087.5

sigma^2 estimated as 5.442e+12: log likelihood=-2299.69

AIC=4603.38 AICc=4603.55 BIC=4612.27

The t-test of the parameters are as follows;



a. The top box contains the time plot of the standardized residuals of the mode which shows that no obvious pattern and looks like an independent identical distribution (i.i.d) sequence with mean zero and few outliers.

b. The middle part is the ACF plot of the residuals and this shows evidence of significance correlation at lags 4 probably due to outliers.

c. The bottom of the figure shows the Ljung-Box plot which suggests that the residuals of the LPG data by the ARIMA(0,1,1) are significant at any positive lag and the p-value of 0.0059 of the residuals by the model confirms the insignificance of the residuals and so, the residuals of the ARIMA(0,1,1) are not independent and identically.

The distribution of the errors by ARIMA(0,1,1) model is as shown in figure 4.10 below. From the diagram it can be observed that the residuals appear to be normally since most of the data points are on the normal line and those which are not on the line deviate to a similar extent below and above the normal line.

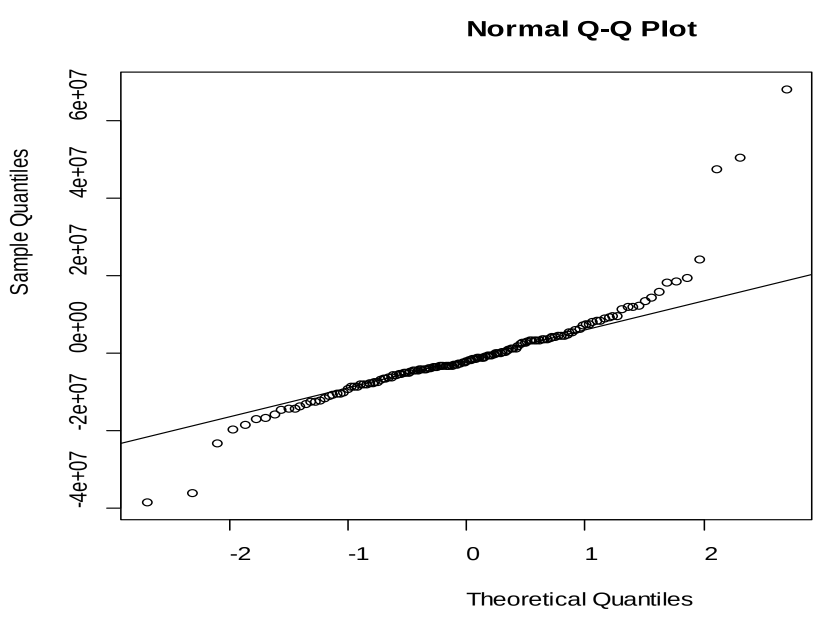


Figure 4.10: Normal Q-Q plot of the residuals of ARIMA(0,1,1)

The histogram plot fitted with the normal curve to show the normality of the residuals for ARIMA(0,1,1) model is as shown in figure 4.11 below and this confirms that the residuals for ARIMA(0,1,1) model are normally distributed.

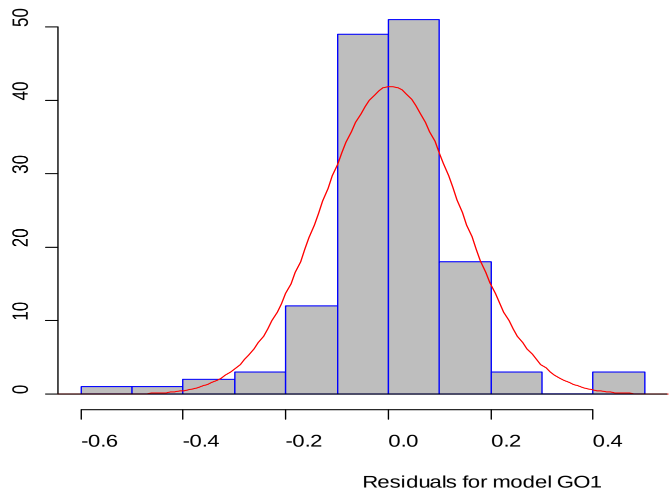


Figure 4.11: Plot of histogram fitted with normal curve of the residuals of ARIMA(0,1,1)

4.5.1 PARAMETER ESTIMATES AND DIAGNOSTICS OF ARIMA (1,1,0) MODEL

Parameter estimates determine the coefficients of the time series equation that is generated from the data and the diagnostics test is used to check the correlation and significance of the residuals.

4.5.3.1 Parameter Estimates

The R result of the ARIMA (1, 1, 0) is as shown below;

Coefficients:

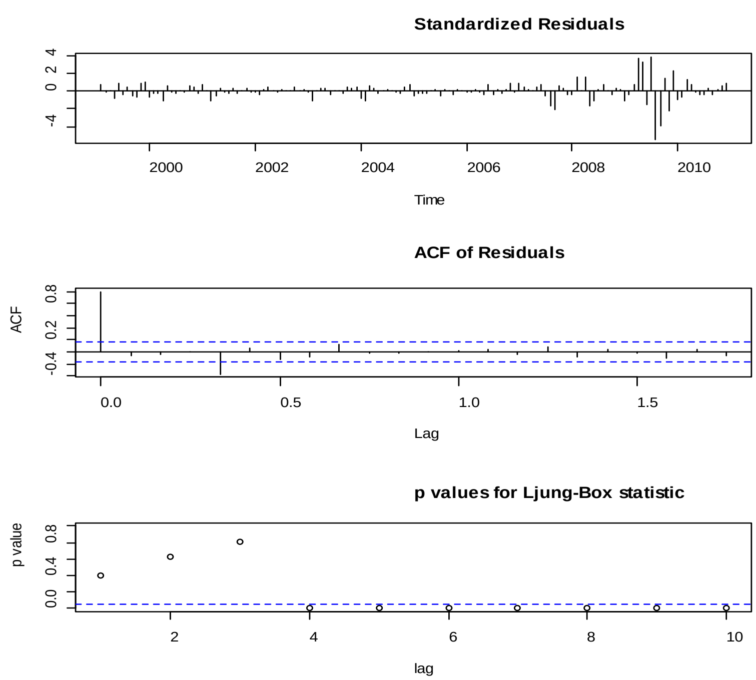
ar1	intercept
-0.3060	99526.5

s.e.   0.0796   148572.1

sigma^2 estimated as 5.365e+12: log likelihood=-2298.68

AIC=4601.37   AICc=4601.54   BIC=4610.26

The t-test of the parameters is as follows;



a. The top box contains the time plot of the standardized residuals of the mode which shows that no obvious pattern and looks like an independent identical distribution (i.i.d) sequence with mean zero and few outliers.

b. The middle part is the ACF plot of the residuals and this shows evidence of significance correlation at lags 4 probably due to an outlier.

c. The bottom of the figure shows the Ljung-Box plot which suggests that the residuals of the LPG data by the ARIMA(1,1,0) are significant at any positive lag and the p-value of 0.0089 of the residuals by the model confirms the insignificance of the residuals and so, the residuals of the ARIMA(1,1,0) are not independent and identically distributed.

The distribution of the errors by ARIMA(1,1,0) model is as shown below. From the diagram it can be observed that the residuals appear to be normally since most of the data points are on the normal line and those which are not on the line deviate to the similar extent below and above the normal line.

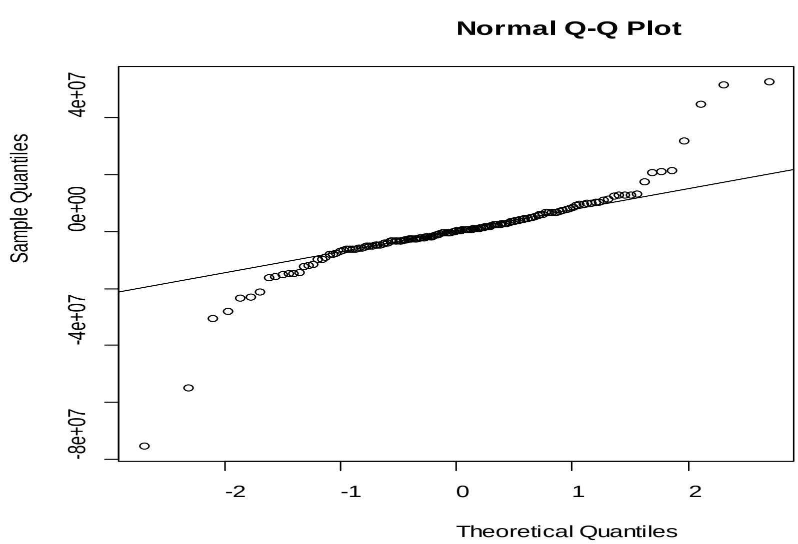


Figure 4.13: Normal Q-Q plot of the residuals of ARIMA(0,1,1)

The histogram plot fitted with the normal curve to show the normality of the residuals for

ARIMA(0,1,1) model is as shown in figure 4.14 below confirms that the residuals for ARIMA(0,1,1) model are normally distributed.

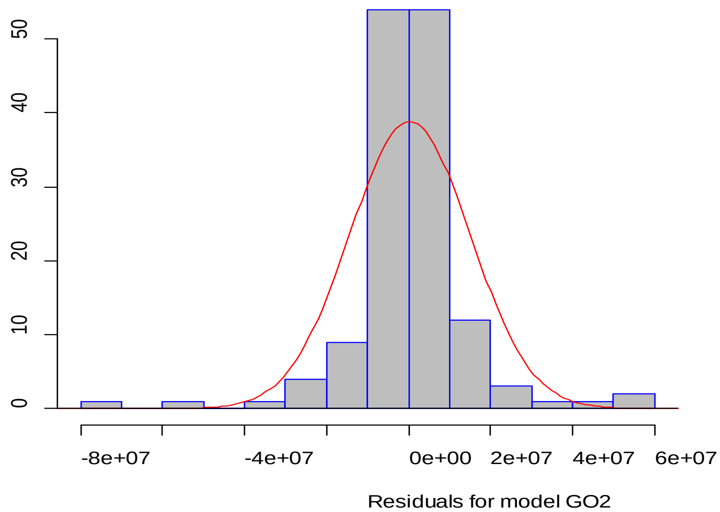


Figure 4.14: Plot of histogram fitted with normal curve of the residuals of ARIMA(0,1,1)

Because the candidate models chosen using the principle of parsimony fall short of some of the basic needs of a good model there is a need to try other models until a better model is obtain for predicting future values. After several trials ARIMA(1,1,3) model is of peculiar characteristics.

4.5.4 PARAMETER ESTIMATES AND DIAGNOSTICS OF ARIMA(1,1,3) FOR THE GAS OIL DATA

Parameter estimates determine the coefficients of the time series equation that is generated from the data and the diagnostics test is used to check the correlation and significance of the residuals.

4.5.4.1 Parameter Estimates

The R result of the ARIMA(1,1,3) is as shown below;

Coefficients:

	ar1	ma1	ma2	ma3	intercept
	0.5770	-1.0066	0.4806	-0.3939	97730.65
s.e.	0.1204	0.1193	0.1279	0.0838	37944.54

sigma^2 estimated as 4.624e+12: log likelihood=-2288.54

AIC=4587.07 AICc=4587.69 BIC=4604.85

To further analyze the results we must appreciate how significant the parameters are by taking the t-test on them. The t-test of the parameters are as follows;

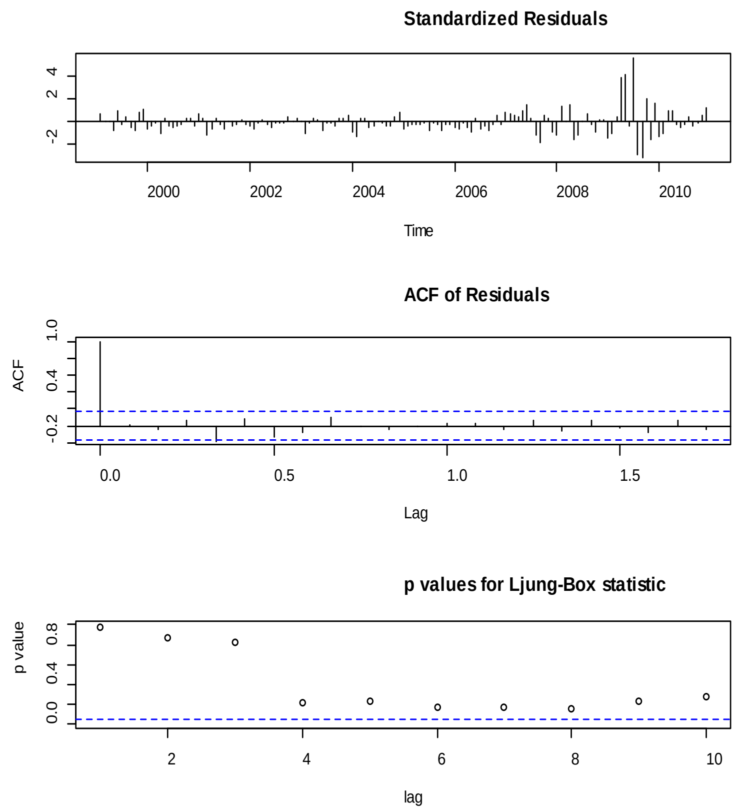


Figure 4.15: Diagnostic test for ARIMA(1,1,3)

a. The top box contains the time plot of the standardized residuals of the mode which shows that no obvious pattern and looks like an independent identical distribution (i.i.d) sequence with mean zero and few outliers.

b. The middle part is the ACF plot of the residuals and this shows evidence of significance correlation at lags 4 probably due to an outlier.

c. The bottom of the figure shows the Ljung-Box plot which suggests that residuals of the LPG data by the ARIMA(0,1,3) are not significant at any positive lag and the p-value of 0.6704 of the residuals by the model confirms the non-significance of the residuals since it is greater than 0.05 and so, the residuals of the ARIMA(1,1,3) are appear to be independent and identically distributed though the diagrams below show otherwise.

The distribution of the errors by ARIMA(1,1,3) model is as shown in figure 4.16 below. From the diagram it can be observed that the residuals appear to be normally distributed since most of the data points are on the normal line and those which are not on the line deviate to the similar extent below and above the normal line.

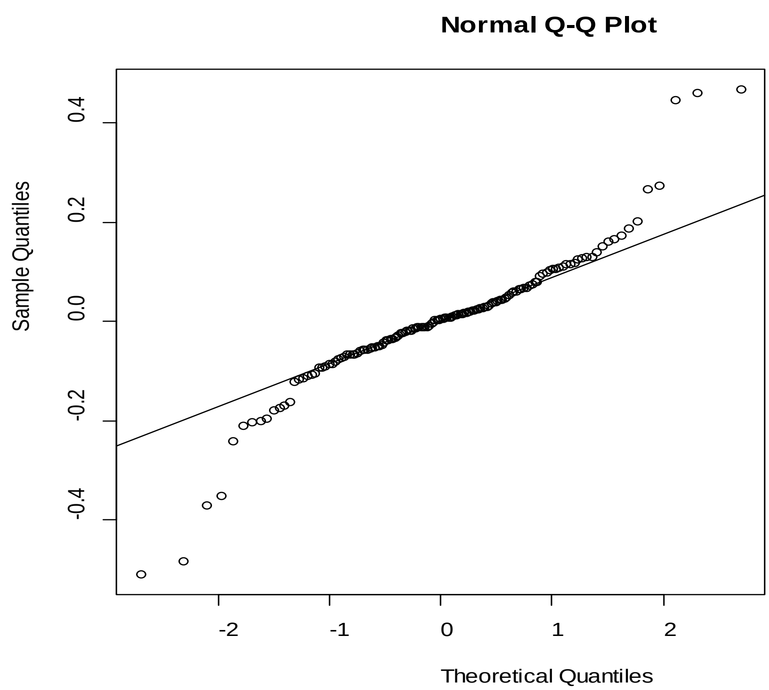


Figure 4.16: Normal Q-Q plot of the residuals of ARIMA(1,1,3)

The histogram plot fitted with the normal curve to show the normality of the residuals for ARIMA(1,1,3) model is as shown in figure 4.17 below and this confirms the normality of the residuals of ARIMA(1,1,3);

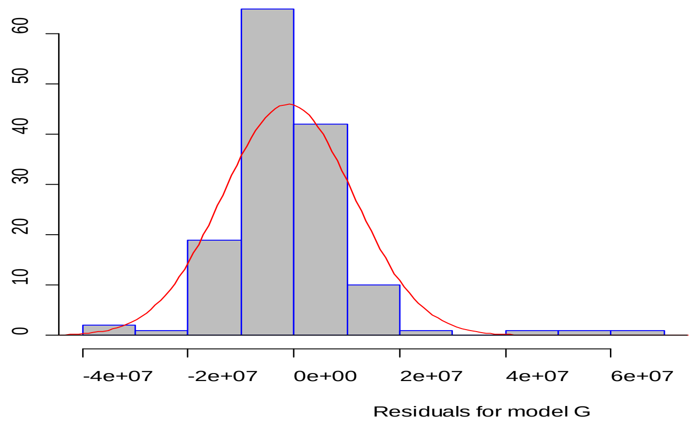


Figure 4.17: Plot of histogram fitted with normal curve of the residuals of ARIMA(0,1,1)

Best Model Selection

Comparing the AICs and BICs of the candidate models of the national gas oil demand data chosen using the principle of parsimony and the over fitted model, the AIC, AICc and BIC of ARIMA(0,1,3) model is better because it has the least of such values compared to the other models as shown in the table 4.3 below.

MODEL	AIC	AICc	BIC
ARIMA(1,1,1)	4603.15	4603.44	4615.00
ARIMA(0,1,1)	4603.38	4603.55	4612.27
ARIMA(1,1,0)	4601.37	4601.54	4610.26
ARIMA(1,1,3)	4587.07	4587.69	4604.85

Table 4.3: AIC, AICc and BIC of the candidate models

4.5.5 Fitting the National Domestic Gas Oil Demand Model

The best model for forecasting the National Gas Oil demand model is the ARIMA(1,1,3) model which has one AR term, a single differencing and three MA terms. The model in terms of the differenced series

4.5.6 Forecasting the National Domestic Gas Oil Demand Model

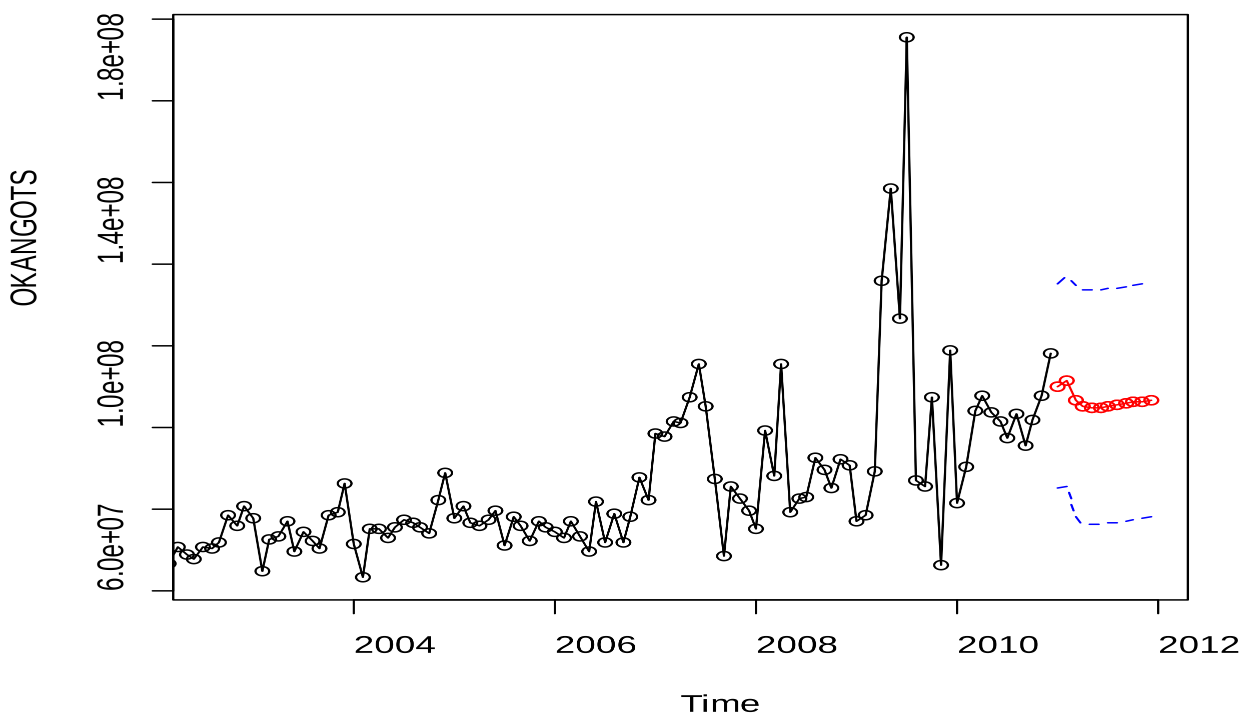


Figure 4.18: Graph of the national domestic Gas Oil demand data with its forecasts and confidence interval.

The forecast graph by ARIMA(1,1,3) is as shown in figure 4.18 below, the graph shows a visual representation of the primary data from the national petroleum authority which is represented by the black lines, the forecasted part represented by the red portion of the graph and the confidence interval represented by the blue dotted lines. Coupled with the forecasted values and the graph, it can be observed that the domestic gas oil demand will leap sharply for in the first quarter of 2011 and stabilize almost to a constant demand rise after the first quarter and will remain as such for the rest of the 2011 year.

A 12 point prediction into the future by the model is as shown in table 4.4 below;

Time Series:

Start=January, 2011

End=December, 2011

Frequency=12

Table of Forecasted Values for the Gas Oil Data in Litres

Month	Point Forecast	Lo 95	Hi 95
Jan 2011	110130409	81501871	140358130
Jan 2011	111445008	78093986	143766015
Mar 2011	106494180	74994945	146865056
Apr 2011	105107456	72130618	149729383
May2011	104868515	69452301	152407700
Jun 2011	194999207	66925969	154934032
Jul 2011	105248935	64526747	157333254
Aug 2011	105536997	62235800	159624201
Sep 2011	105837405	60038461	161821540
Oct 2011	106141788	57923048	163936953
Nov 2011	196447451	55880077	165979924
Dec 2011	106753527	53901725	167958276

Table 4.4: Predicted values 95% confidence interval showing the lowest (Lo) and highest (Hi) values for the intervals for the Gas Oil Data

4.5.7 Forecast Accuracy

The forecast values ARIMA(1,1,3) model for the gas oil data was tested to ascertain how close its predicted values are to the actual values left for testing the validity of the models by considering the error margins. We used the Root Mean Square Error (RMSPE). The RMSPE test gave a value of 8.59% also meaning that there is only 8.59% error in using ARIMA(1,1,3) to forecast into the future.

4.6 NATIONAL LPG DATA ANALYSIS

The national gas oil data is analysed based the R software on the monthly primary data obtained from the National Petroleum Authority. This is a compilation of gas oil supply or demand mainly by commercial and domestic users for transportation and agricultural purposes in the country.

4.6.4 Descriptive Analysis Of LPG Demand in Ghana

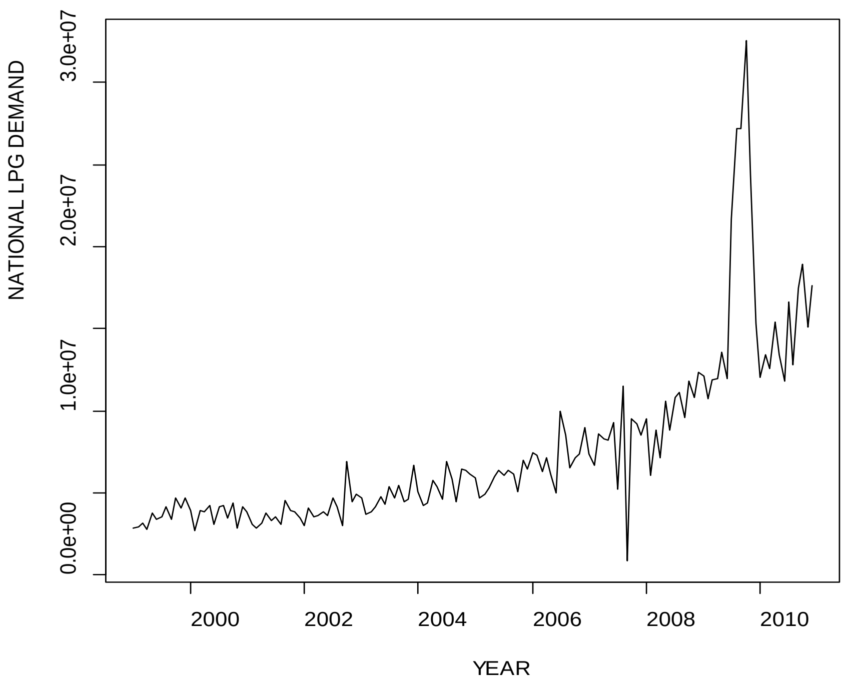


Figure 4.19: Time series plot of the LPG demand in Ghana.

Figure 4.19 above is the time graph of the LPG demand in the country and it shows the behavior of the demand pattern of LPG. The demand for LPG rose getting to the end of the 1999 which gave to a swift decrease in the demand by the middle of the first quarter of 2000 probably due to change in government in the country. By mid 2000, the demand had stabilize though less than the demand in of the commodity from January 1999. There was a sudden demand of the LPG by the end of 2002 and this continued steadily till mid 2006 probably due to the increase awareness of the danger charcoal and other fossil fuel is doing to our forest and environment as a whole and the pressure by international donors demanding the use of LPG for domestic and small scale company use (Energy Foundation Annual Report, 2005). By the last quarter of 2006, the demand was on the ascendancy and this kept rising until in the last quarter of 2007 when a fuel shortage hit the country for almost a month. This low demand is as never before, this quickly changed and the demand was on the ascendancy again and kept rising until in the first quarter of the of 2009 there was a sudden rush for gas as never before probably by taxi drivers as made known by a deputy minister of information (Meet the Press Conference, Ministry of information, March-2009). This hooping demand dropped sharply by the end of 2009 to a level still higher than the previous demand and increasing steadily. By the KPSS test, the LPG data is not stationary.

4.6.2 ACF of Natioanal LPG Demand Data

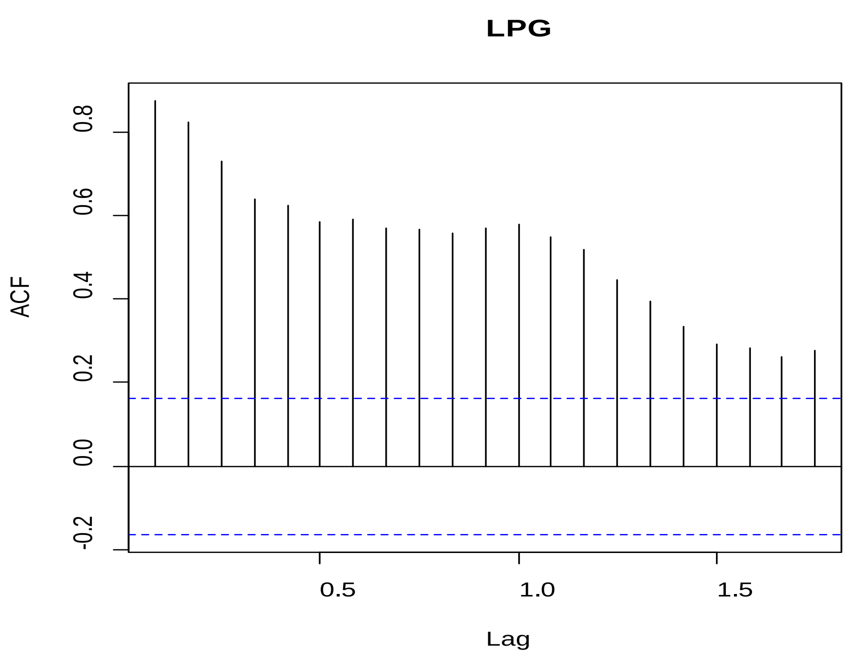


Figure 4.20: The ACF of the LPG primary data

The autocorrelation function of the LPG demand in the country is as shown in figure 4.20 which describes the correlation between the values of the LPG at different time, as a function of the time difference. The autocorrelation is decreasing systematically with time though there are rises in the decrease pattern and this shows there is a trend in the LPG demand which is important.

4.6.3 Plot of First Differencing

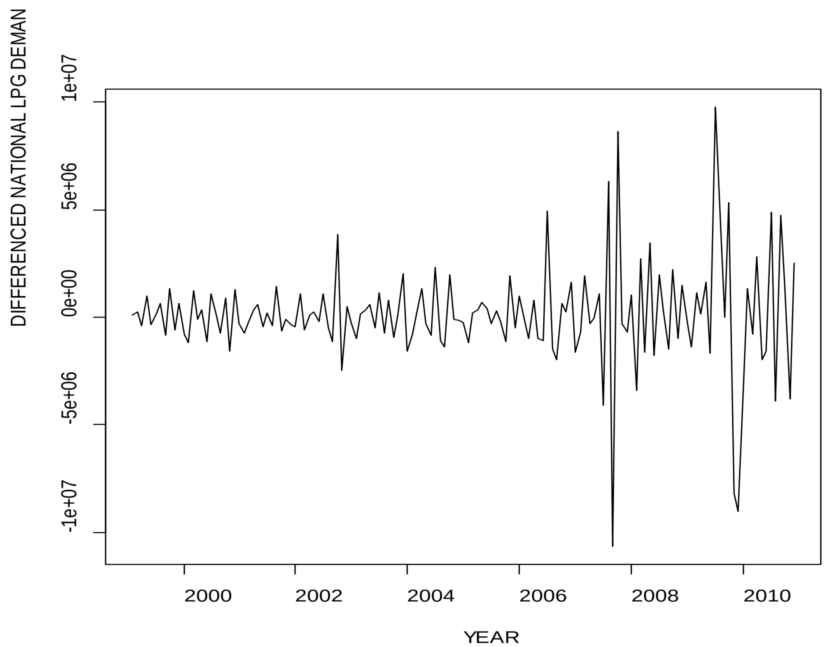


Figure 4.21: Plot of LPG differenced data

Figure 4.21 shows a graph of the differenced LPG data. Stationarity was achieved by differencing the LPG data and this was confirmed by the KPSS test given a p-value of 0.1. The differencing was done to remove the trend component of the in the LPG demand data. The observations now move irregularly but revert to its mean value and having a constant variance.

4..1.1 Seasonal and Non-Seasonal Unit Root Test

The differenced national domestic gas oil demand data passes the Dickey-Fuller test for unit root since the Dickey-Fuller value of -181.32 at lag order of 4 and a p-value of 0.01 which is less than 0.05 and by passing the Dickey-Fuller test the data can also be said to be not white noise, by this, we say there exists dependencies and this needs to be modeled. The KPSS test conducted on the differenced data indicated a p-value of 0.1 which is greater than 0.05, therefore, we do not reject the null hypothesis, the stationarity assumption holds for the series.

4.6.4 ACF and PACF Plot

To generate the candidate models from which the best model can be selected is determined using the ACF and PACF values at various lags, the ACF and PACF values for the first 22 lags are as shown in figure 4.5 below;

A table of ACF nad PACF values of lags

LAGS	ACF	PACF		LAGS	ACF	PACF
[1,]	-0.31	-0.31		[12,]	0.12	0.07
[2,]	0.14	0.05		[13,]	-0.12	-0.08
[3,]	0.01	0.07		[14,]	0.11	-0.13
[4,]	-0.3	-0.32		[15,]	-0.02	0.01
[5,]	0.08	-0.12		[16,]	0.02	0.04
[6,]	-0.16	-0.12		[17,]	0	-0.03
[7,]	0.1	0.04		[18,]	-0.04	-0.09
[8,]	-0.11	-0.18		[19,]	0.04	0
[9,]	0.07	-0.04		[20,]	-0.17	-0.18
[10,]	-0.12	-0.21		[21,]	0.26	0.22
[11,]	0.05	-0.03		[22,]	-0.15	-0.01

Table 4.5: ACF and PACF values at various lags

a. ACF Plot

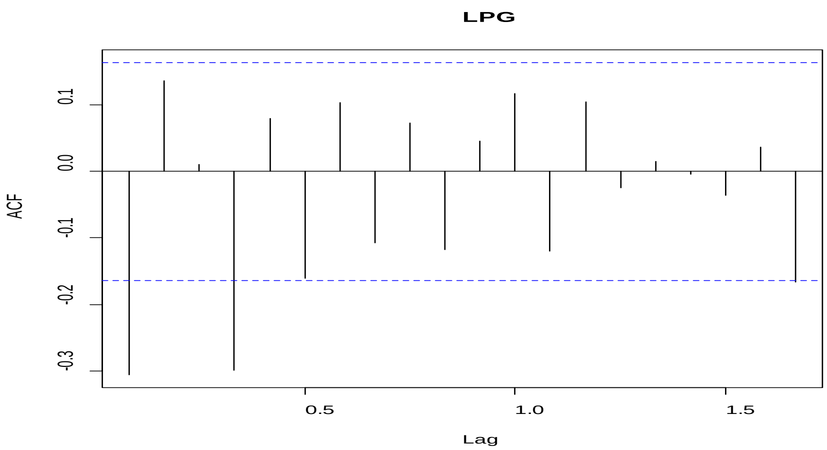


Figure 4.22: Autocorrelation function plot

Figure 4.22 is the ACF plot of the first differencing of the national LPG demand data at various lags. The ACF plot indicates significant error limits at lags 1, 4, and 20, meaning possible MA models are MA(1), MA(4) and MA(20) behaviors respectively. Because the error limit at lag 2 is not significant and the first to behave as such, we use a q value of 1 that is MA(1).

Figure 4.23 is the PACF plot of the first differencing of the national LPG demand data at various lags. The PACF plot indicates significant error limits at lags 1, 4, 8, 10 and 20 meaning possible AR models are AR(1), AR(4), AR(8), AR(10)and AR(20) behaviors respectively. Because the error limit at lag 2 in the PACF plot is not significant and the first to behave as such, we use a p value of 1 that is AR(1).

b. PACF Plot

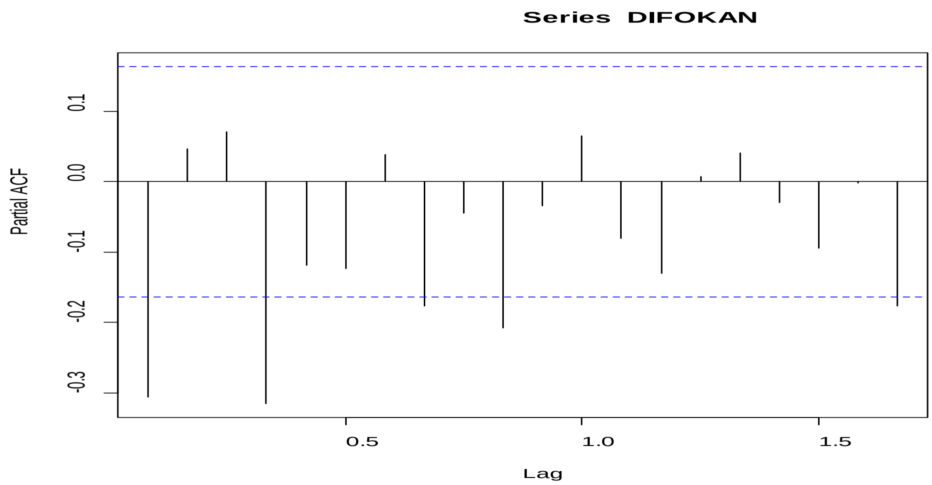


Figure 4.23: Partial Autocorrelation function plot

By using the parsimony principle AR(1) and MA(1) are selected. The following models are recommended;

- ARIMA(1,1,1)
- ARIMA(0,1,1)
- ARIMA(1,1,0)

4.5 MODEL SELECTION FOR THE NATIONAL DOMESTIC LPG DEMAND DATA

The time series analysis for the national domestic gas oil demand for the chosen candidate models is as shown below:

4.7.1 Parameter Estimates and Diagnostic of ARIMA(1,1,1) Model

Parameter estimates determine the coefficients of the time series equation that is generated from the data and the diagnostics test is used to check the correlation and significance of the residuals.

4.7.1.1 Parameter Estimates

The R result for the ARIMA(1,1,1) model is as shown below;

```
Coefficients:
          ar1      ma1    intercept
      -0.3909   0.0926    99334.54
s.e.   0.1848   0.1905   152275.03
sigma^2 estimated as 5.356e+12:  log likelihood=-2298.58
AIC=4603.15   AICc=4603.44   BIC=4615
```

To further analyze the results we must appreciate how significant the parameters are by taking the t-test on them. The t-test of the parameters is as follows;

4.7.1.2 Diagnostic Test

The diagnostic test for ARIMA(1,1,1) is as shown in figure 4.24 below;

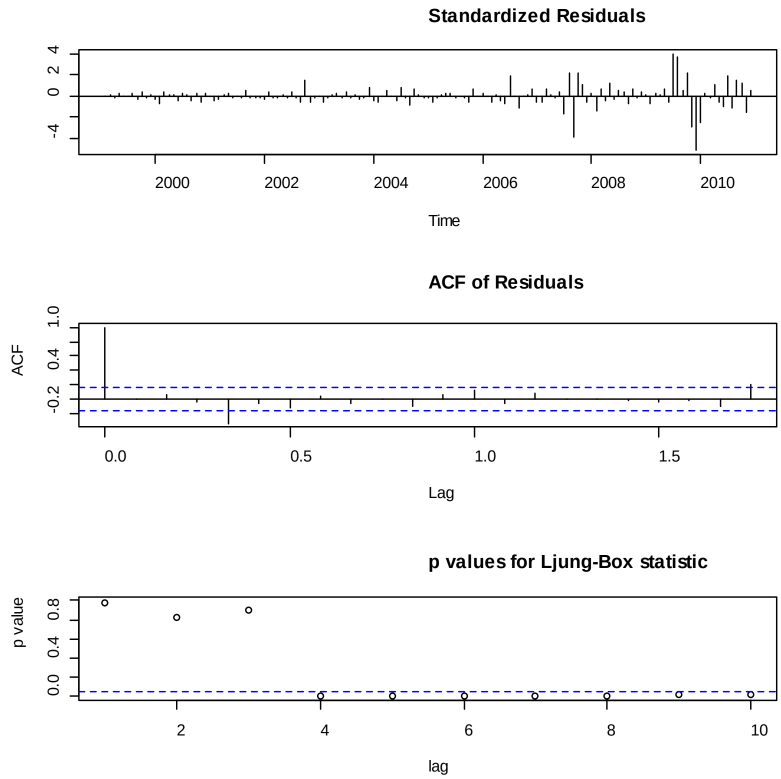


Figure 4.24: Diagnostic test for ARIMA(1,1,1)

a. The top box of figure 4.24 contains the time plot of the standardized residuals of the mode which shows that no obvious pattern and looks like an independent identical distribution (i.i.d) sequence with mean zero and few outliers.

b. The middle part of the figure 4.24 is the ACF plot of the residuals and this shows evidence of significance correlation at lags 3 and 18 probably due to outliers.

c. The bottom of the figure 4.24 shows the Ljung-Box plot which suggests that the residuals of the LPG data by the ARIMA(1,1,1) are significant at any positive lag and the p-value of 0.01192 of the residuals by the model confirms the insignificance of the residuals and so, the residuals of the ARIMA(1,1,1) are not independent and identically distributed and the diagrams below seem to show otherwise.

The distribution of the errors by ARIMA(1,1,1) model is as shown below. From the diagram it can be observed that the residuals appear to be normally distributed since most of the data points are on the normal line and those which are not on the line deviate to the similar extent below and above the normal line.

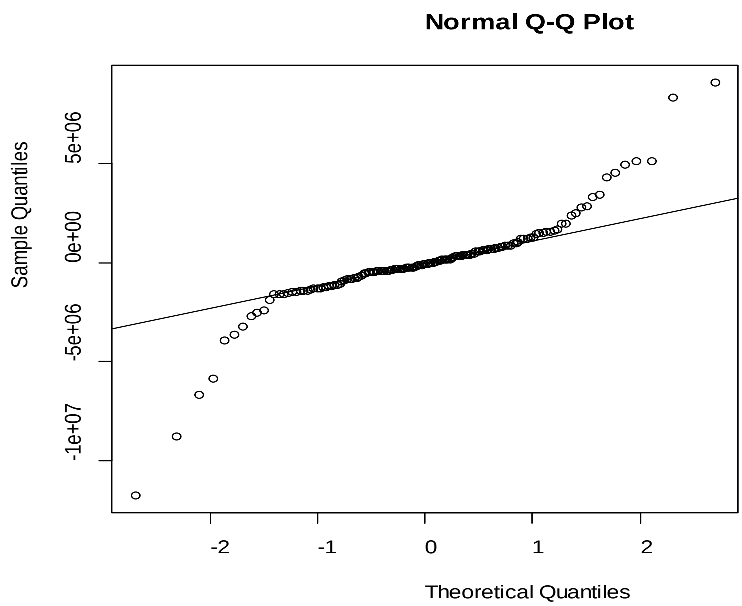


Figure 4.25: Normal Q-Q plot of the residuals of ARIMA(1,1,1)

The histogram plot fitted with the normal curve to show normality of the residuals for ARIMA(1,1,1) model is as shown in figure 4.26 below and this confirms the normality of the residuals of ARIMA(1,1,1) model;

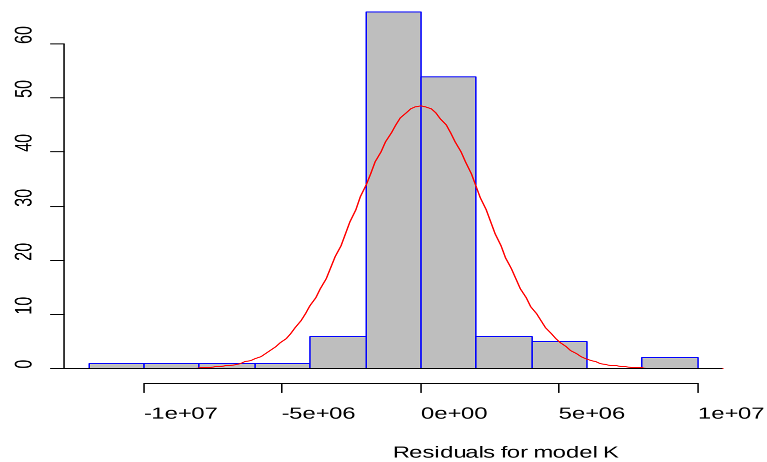


Figure 4.26: Plot of histogram fitted with normal curve of the residuals of ARIMA(1,1,1)

4.7.2 PARAMETER ESTIMATES AND DIAGNOSTICS OF ARIMA(0, 1, 1) MODEL FOR THE LPG DATA

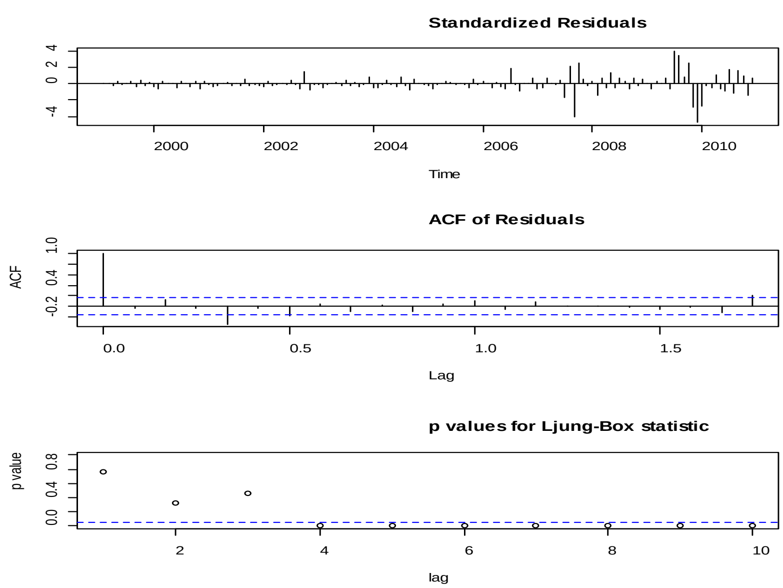
Parameter estimates determine the coefficients of the time series equation that is generated from the data and the diagnostics test is used to check the correlation and significance of the residuals.

4.7.2.1 Parameter Estimates

The R result for the ARIMA(0,1,1) model is as shown below;

```
Coefficients:
          ma1   intercept
      -0.2737    100523.5
s.e.   0.0789    142087.5
sigma^2 estimated as 5.442e+12:  log likelihood=-2299.69
AIC=4603.38   AICc=4603.55   BIC=4612.27
```

To further analyze the results we must appreciate how significant the parameters are by taking the t-test on them. The t-test of the parameters are as follows;



b. The middle part is the ACF plot of the residuals and this shows evidence of significance correlation at lags 4 and 21 probably due to outliers.

c. The bottom of the figure shows the Ljung-Box plot which suggests that the residuals of the LPG data by the ARIMA(0,1,1) are significant at any positive lag and the p-value of 0.0035 of the residuals by the model confirms the insignificance of the residuals and so, the residuals of the ARIMA(0,1,1) are not independent and identically distributed but the diagrams below seem to show otherwise.

The distribution of the errors by ARIMA(0,1,1) model is as shown below. From the diagram it can be observed that the residuals appear to be normally distributed since most of the data points are on the normal line and those which are not on the line deviate to the similar extent below and above the normal line.

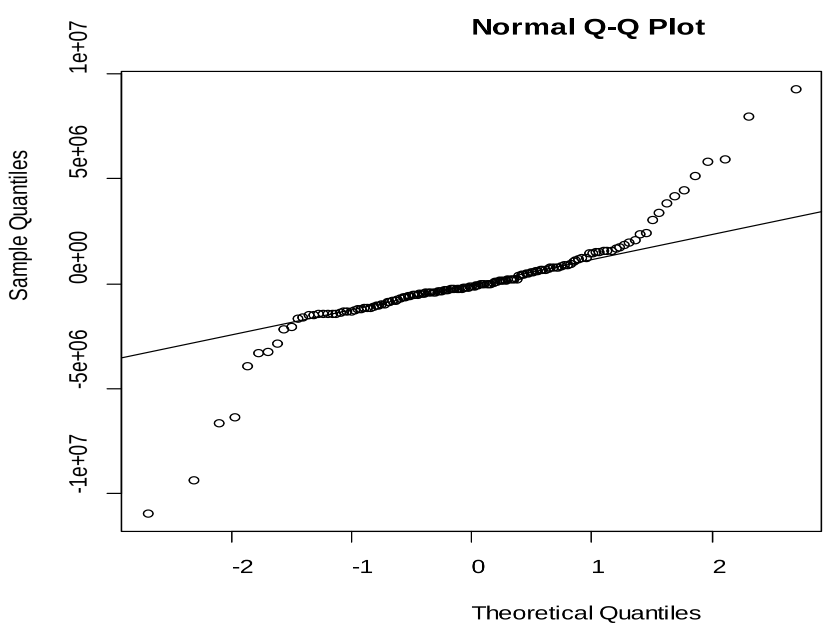


Figure 4.28:Normal Q-Q plot of the residuals of ARIMA(0,1,1)

The histogram plot fitted with the normal curve to show the normality of the residuals for ARIMA(0,1,1) model is as shown in figure 4.28 below and this confirms the normality of the residuals of ARIMA(0,1,1) model;

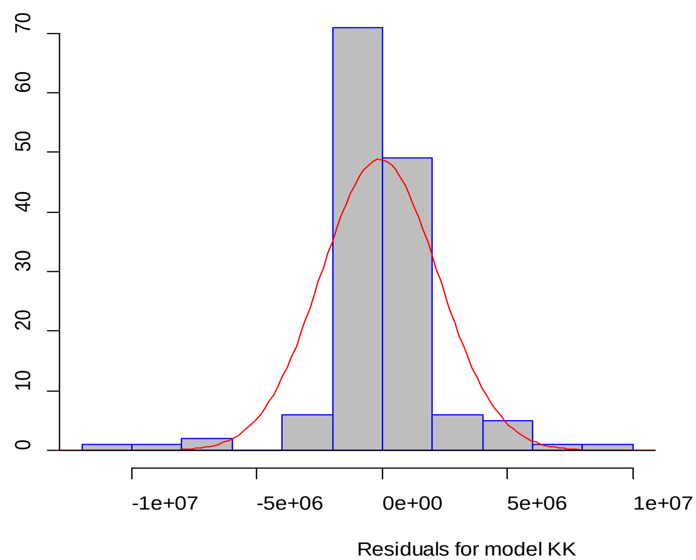


Figure 4.29: Plot of histogram fitted with normal curve of the residuals of ARIMA(0,1,1)

4.7.3 PARAMETER ESTIMATES AND DIAGNOSTICS OF ARIMA(1, 1, 0) MODEL FOR THE LPG DATA

Parameter estimates determine the coefficients of the time series equation that is generated from the data and the diagnostics test is used to check the correlation and significance of the residuals.

4.7.3.1 Parameter Estimates

The R result for the ARIMA(0,1,1) model is as shown below;

Coefficients:

ar1 intercept

-0.3060 99526.5

s.e. 0.0796 148572.1

sigma^2 estimated as 5.365e+12: log likelihood=-2298.68

AIC=4601.37 AICc=4601.54 BIC=4610.26

To further analyze the results we must appreciate how significant the parameters are by taking the t-test on them. The t-test of the parameters are as follows;

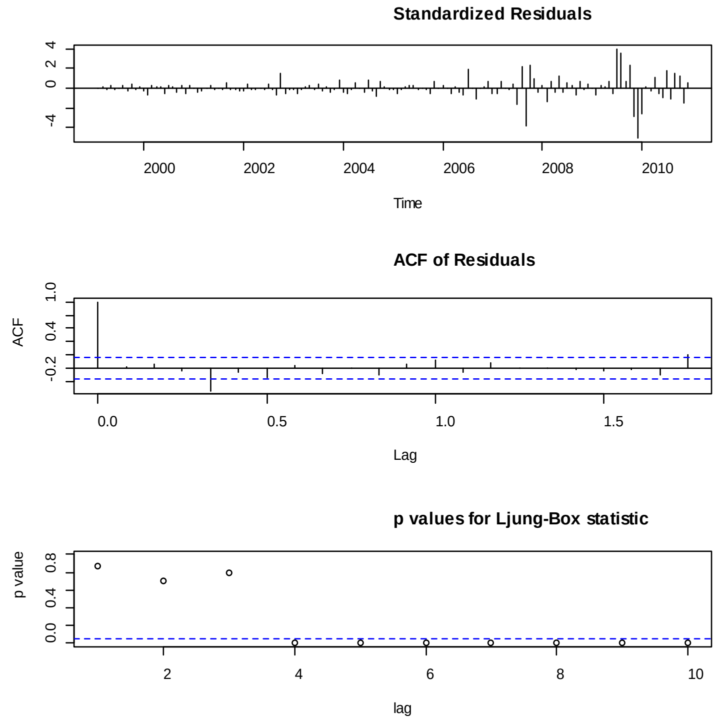


Figure 4.30: Diagnostic test for ARIMA(1,1,0)

a. The top box contains the time plot of the standardized residuals of the mode which shows that no obvious pattern and looks like an independent identical distribution (i.i.d) sequence with mean zero and few outliers.

b. The middle part is the ACF plot of the residuals and this shows evidence of significance correlation at lags 4 and 21 probably due to outliers.

c. The bottom of the figure shows the Ljung-Box plot which suggests that the residuals of the LPG data by the ARIMA(1,1,0) are significant at any positive lag and the p-value of 0.0089of the residuals by the model confirms the insignificance of the residuals and so, the residuals of the ARIMA(1,1,0) are not independent and identically distributed though the diagrams below show otherwise.

The distribution of the errors by ARIMA(1,1,0) model is as shown below. From the diagram it can be observed that the residuals appear to be normally distributed since most of the data points are on the normal line and those which are not on the line deviate to the similar extent below and above the normal line since most of the data points are on the normal line and those which are not on the line deviate to the similar extent below and above the normal line.

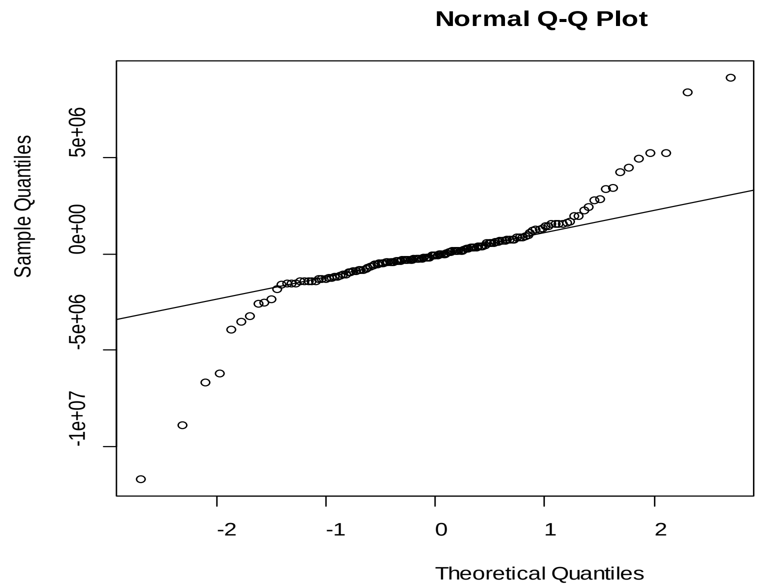


Figure 4.31: Normal Q-Q plot of the residuals of ARIMA(1,1,0)

The histogram plot fitted with the normal curve to show the normality of the residuals for ARIMA(0,1,1) model is as shown in figure 4.32 below and this confirms the normality of the residuals of ARIMA(0,1,1) model;

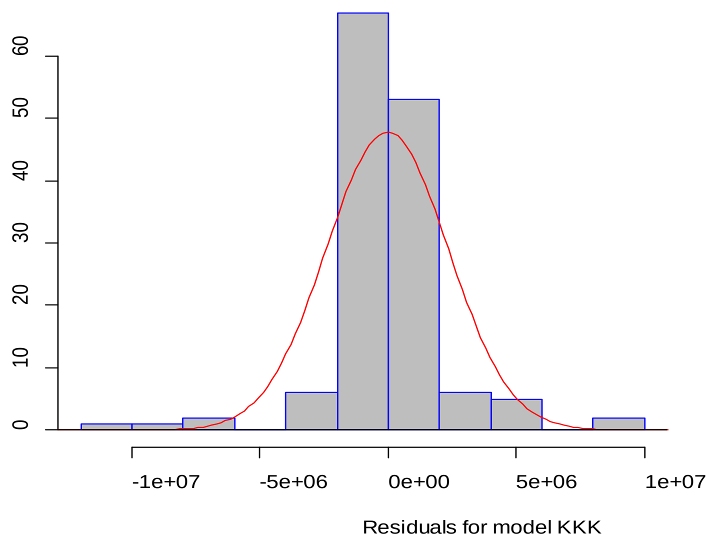


Figure 4.32: Plot of histogram fitted with normal curve of the residuals of ARIMA(1,1,0)

4.7.3.3 An Over Fit

Because none of the model candidates chosen by parsimony principle could have a p-value of its residuals greater than 0.05, we tried other models of the data using various values of p and q. After several trials ARIMA(2,1,3) model was chosen;

4.7.4 PARAMETER ESTIMATES AND DIAGNOSTICS OF ARIMA(2,1, 3)MODEL FOR THE LPG DATA

Parameter estimates determine the coefficients of the time series equation that is generated from the data and the diagnostics test is used to check the correlation and significance of the residuals.

4.7.4.1 Parameter Estimates

The R result for the ARIMA(2,1,3) model is as shown below;

Coefficients:

	ar1	ar2	ma1	ma2	ma3	intercept
	0.8408	-0.3718	-1.2564	0.9374	-0.5565	97542.60
s.e.	0.1365	0.1606	0.1176	0.2016	0.1073	44310.53

sigma^2 estimated as 4.51e+12: log likelihood=-2286.83

AIC=4585.66   AICc=4586.49   BIC=4606.4

To further analyze the results we must appreciate how significant the parameters are by taking the t-test on them. The t-test of the parameters are as follows;

4.7.4.2 Diagnostic Test

The diagnostic test plot of residuals of the ARIMA(2,1,3) is as shown below;

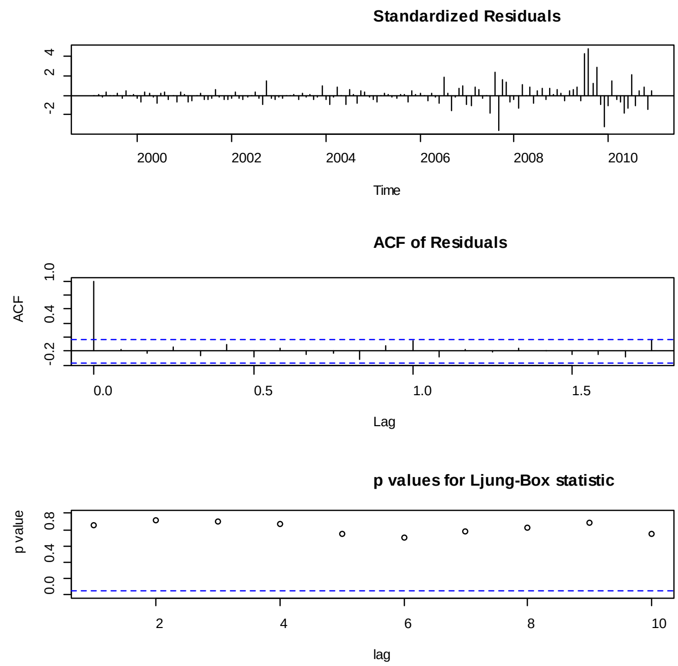


Figure 4.33: Diagnostic test for ARIMA(2,1,3)

a. The top box contains the time plot of the standardized residuals of the mode which shows that no obvious pattern and looks like an independent identical distribution (i.i.d) sequence with mean zero and few outliers.

b. The middle part is the ACF plot of the residuals and this shows evidence of significance correlation at lags 4 and 21 probably due to outliers.

c. The bottom of the figure shows the Ljung-Box plot which suggests that residuals of the LPG data by the ARIMA(2,1,3) are not significant at any positive lag and the p-value of 0.7482 of the residuals by the model confirms the non-significance of the residuals since it is greater than 0.05 and so, the residuals of the ARIMA(2,1,3) are independent and identically distributed and the diagrams below seem to confirm the significance of the residuals.

The distribution of the errors by ARIMA(2,1,3) model is as shown below. From the diagram it can be observed that the residuals appear to be normally distributed since most of the data points are on the normal line and those which are not on the line deviate to the similar extent below and above the normal line.

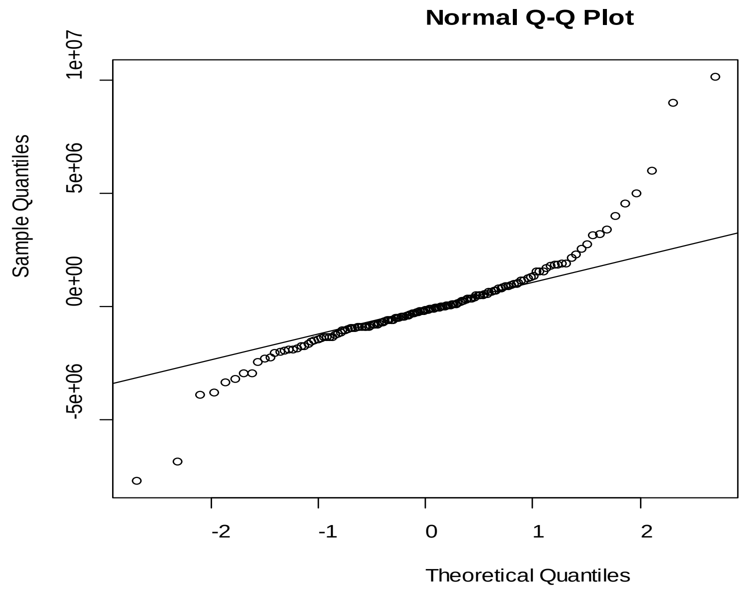


Figure 4.34: Normal Q-Q plot of the residuals of ARIMA(2,1,3)

The histogram plot fitted with the normal curve to show the normality of the residuals for ARIMA(2,1,3) model is as shown in figure 4.35 below and this confirms the normality of the residuals of ARIMA(2,1,3) model;

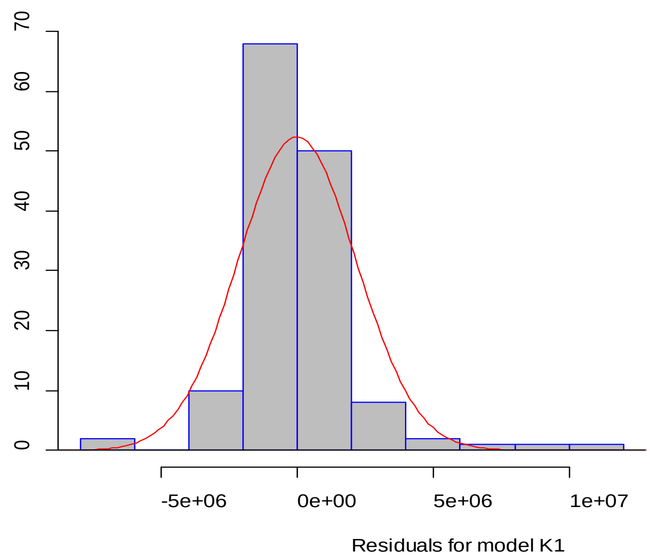


Figure 4.35: Plot of histogram fitted with normal curve of the residuals of ARIMA(2,1,3)

4.7.5 Best Model Selection

The AIC, AICc and BIC of the various candidate models for the LPG data is as shown in table 4.6 below. By comparing the AICs, AICcs and BICs of the candidate models, the AIC and BIC of ARIMA(2,1,3) model is better since it has the least of such values.

MODEL	AIC	AICc	BIC
ARIMA(1,1,1)	4603.13	4603.44	4615.00
ARIMA(0,1,1)	4603.38	4603.55	4612.27
ARIMA(1,1,0)	4601.37	4601.54	4610.25
ARIMA(2,1,3)	4585.66	4586.49	4606.40

Table 4.6: A table of models and their corresponding AIC, AICc and BIC for the LPG data

4.7.6 Fitting the National LPG Demand Model

The best model for forecasting the National LPG demand model is the ARIMA(2,1,3) model which has no AR term, a single differencing and three MA terms and so it is a non-seasonal integrated moving average with one level of differencing without AR terms. The model in terms of the differenced series

4.7.7 Forecasting the National LPG Demand Model

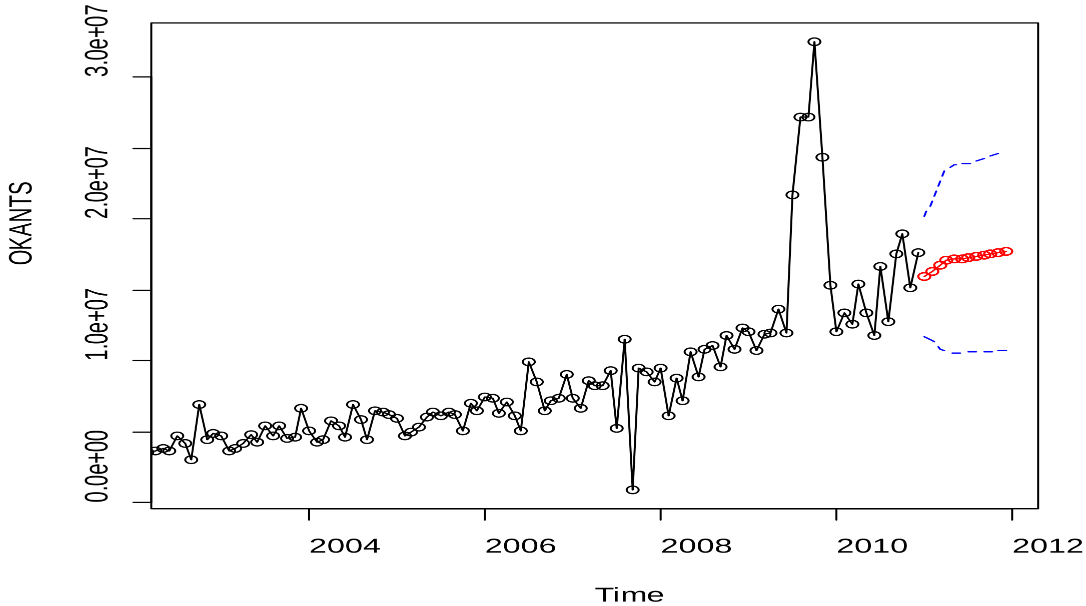


Figure 4.36: Graph of the national domestic LPG demand data with its forecasts and critical region

The graph shows a visual representation of the primary data from the national petroleum authority which is represented by the black lines, the forecasted part represented by the red portion of the graph and the confidence interval represented by the blue dotted lines. Coupled with the forecasted values and the graph, it can be observed that the domestic LPG demand will leap sharply for in the first quarter of 2011 and stabilize almost to a constant demand rise after the first quarter and will remain as such for the rest of the 2011 year.

A 12 point prediction into the future by the model is as shown table 4.7 below;

Table of Forecast Values for the LPG Data in Kilograms

Month	Point Forecast	Lo 95	Hi 95
Jan 2011	15331682	9301001	21362363
Jan 2011	13851785	7999019	19704552
Mar 2011	13793382	7586678	20000086
Apr 2011	13825234	7244491	20405977
May2011	15080583	7528133	22633033
Jun 2011	12992736	6176974	19808499
Jul 2011	17443929	7894082	26993776
Aug 2011	16845167	7251059	26439276
Sep 2011	15830313	6475661	25184965
Oct 2011	18806737	7302795	30310678
Nov 2011	15945816	5869843	26021789
Dec 2011	16792699	5851032	27734367

Table 4.7: Predicted values at 95% confidence intervals showing the lowest (Lo) and highest (Hi) values for the intervals for the LPG Data

4.7.8 Forecast Accuracy

The forecast values ARIMA(2,1,3) model for the gas oil data was tested to ascertain how close its predicted values are to the actual values left for testing the validity of the models by considering the error margins. We used the Root Mean Square Error (RMSPE). The RMSPE test gave a value of 1.09% also meaning that there is only 1.09% error in using ARIMA(1,1,3) to forecast into the future. The RMSPE gives a tolerable error percentage.

4.8 National Premix Fuel Data Analysis

The national gas oil data is analysed based on the monthly primary data obtained from the National Petroleum Authority. This is a compilation of gas oil supply or demand mainly by commercial and domestic users for transportation and agricultural purposes in the country.

4.8.1 Descriptive Analysis of the Premix Fuel Demand in Ghana

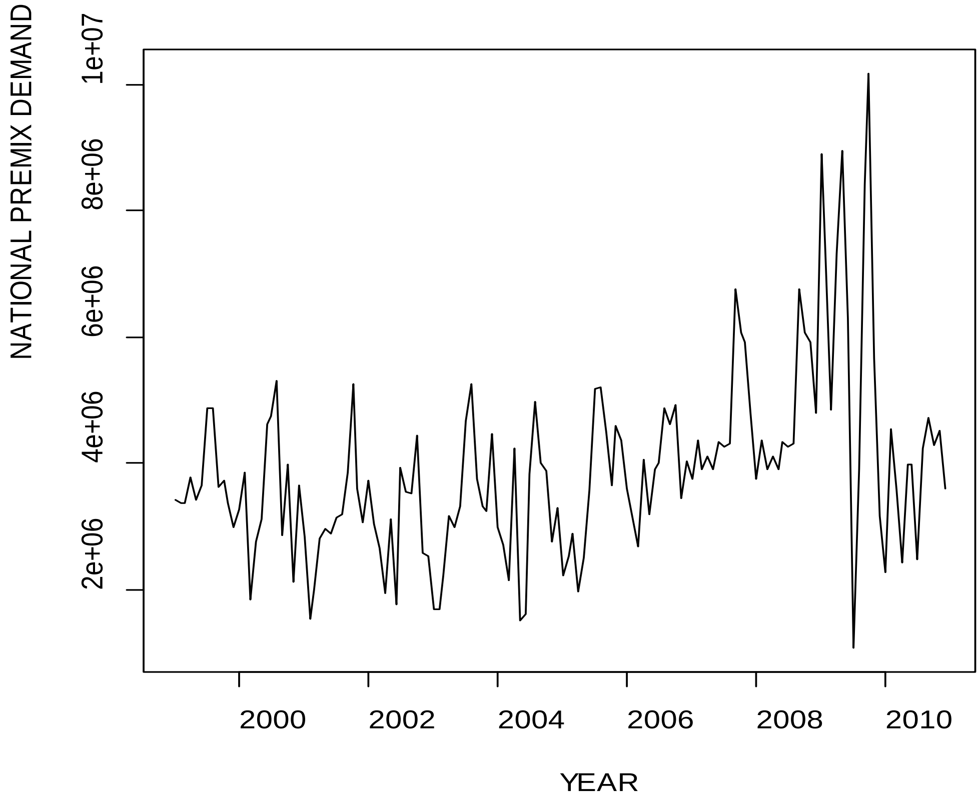


Figure 4.37: The time series plot of the Premix fuel demand data of Ghana

Figure 4.37 above is a time graph of the demand for premix fuel in the country. In 1999, there was fairly an increasing demand for premix fuel and getting to the latter part of the third quarter of 1999 year the premix fuel demand in Ghana was on the increase probably due to the sudden explosion in herring and other fishes in the Ghanaian seas, (Ministry of Food and Agriculture, Annual Report 1999), this demand decreased down into the first quarter of 2000 and there were fluctuations in the demand level till 2006 where a these demand fluctuations started increasing over time. In the third quarter in 2008 there was a hooping demand for Premix fuel and this persisted for about a year and dropped seriously to it ever minimum demand value in the latter part of the last quarter of 2009 and this continued through 2010.

4.8.2 ACF of Natioanal Premix Fuel Demand Data

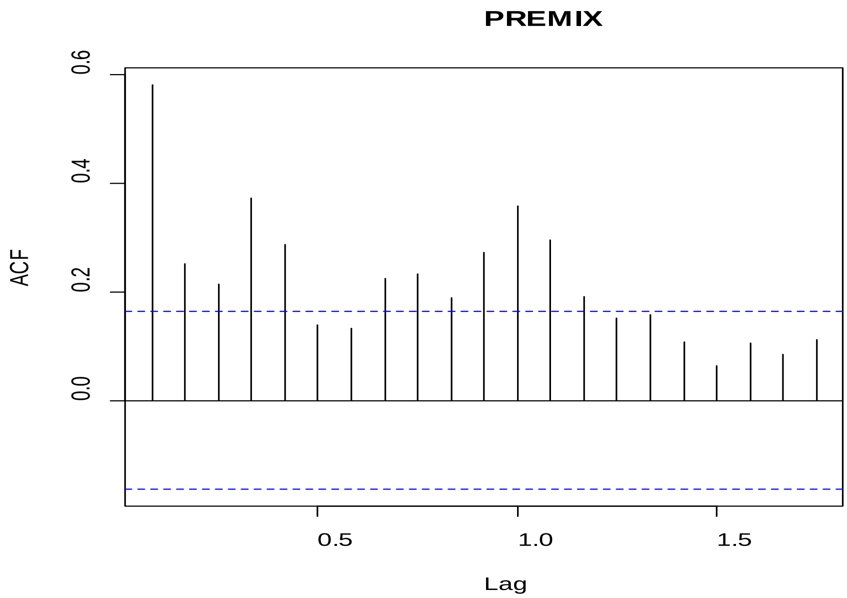


Figure 4.38: ACF plot of the Premix Fuel data

The autocorrelation function of the Premix fuel demand in the Ghana is as shown in figure 4.38 above and it describes the correlation between the values of the Premix fuel at different time, as a function of the time difference. The autocorrelation has a combination of decreasing and increasing spikes with the decreasing arranged at both sides of the higher spikes which suggests the presence of both trend and seasonality components in the Premix fuel demand data. The seasonality component was confirmed by the seasonality test using analysis of variance test, ANOVA, the Premix data passed the seasonality test with a p-value of 0.9320.

4.8.3 Plot of First Differencing

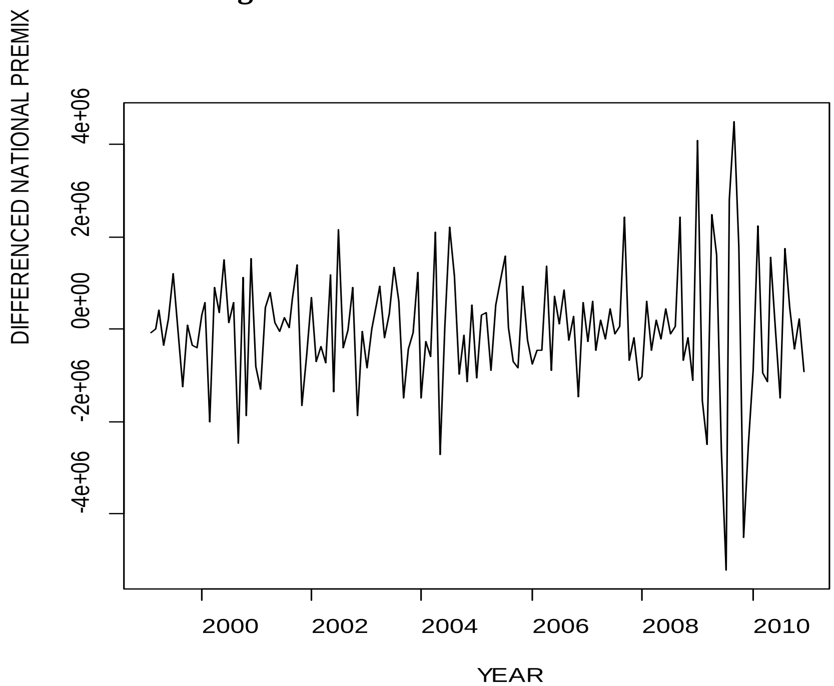


Figure 4.39: The plot of differenced Premix fuel data.

Because of the trend and seasonality components in the Premix data, there was a need to difference it and remove the components in order to be able to analyze the data correctly.

4.8.4 Seasonal and Non-Seasonal Unit Root Test

The differenced national Premix fuel demand data passes the Dickey-Fuller test for unit root since the Dickey-Fuller value of -7.0918 at lag order of 5 and a p-value of 0.01 which is less than 0.05 and by passing the Dickey-Fuller test the data can also be said to be not white noise, by this, we say there exists dependencies and this needs to be modeled. The KPSS test conducted on the differenced data indicated a p-value of 0.1 which is greater than 0.05, therefore, we do not reject the null hypothesis, the stationarity assumption holds for the series. Also the table below shows how significant the seasonal effect is and thus confirming that it must be part of the model since it shows an F value of 1.996 meaning the seasonality existence test shows a significant result.

ANOVA Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Month	11	4.364e+13	3.967e+12	1.996	0.0335 *
Residuals	132	2.624e+14	1.988e+12		

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Table 4.8: ANOVA table of seasonality test result

4.8.5 ACF and PACF PLOTS

Table 4.9 below shows the ACF and PACF values at various lags;

A table of ACF and PACF lags

LAGS	ACF	PACF				
[1,]	-0.11	-0.11		[12,]	0.18	0.02
[2,]	-0.35	-0.36		[13,]	0.05	-0.05
[3,]	-0.23	-0.38		[14,]	-0.08	0.05
[4,]	0.29	0.05		[15,]	-0.05	0.05
[5,]	0.07	-0.09		[16,]	0.06	0.02
[6,]	-0.16	-0.15		[17,]	-0.01	-0.02
[7,]	-0.12	-0.1		[18,]	-0.1	-0.09
[8,]	0.1	-0.09		[19,]	0.08	0.11
[9,]	0.06	-0.11		[20,]	-0.06	-0.14
[10,]	-0.15	-0.24		[21,]	0.01	-0.03
[11,]	-0.01	-0.12		[22,]	-0.12	-0.13

Table 4.9: Values of ACF and PACF at various lags

a. **ACF Plot**

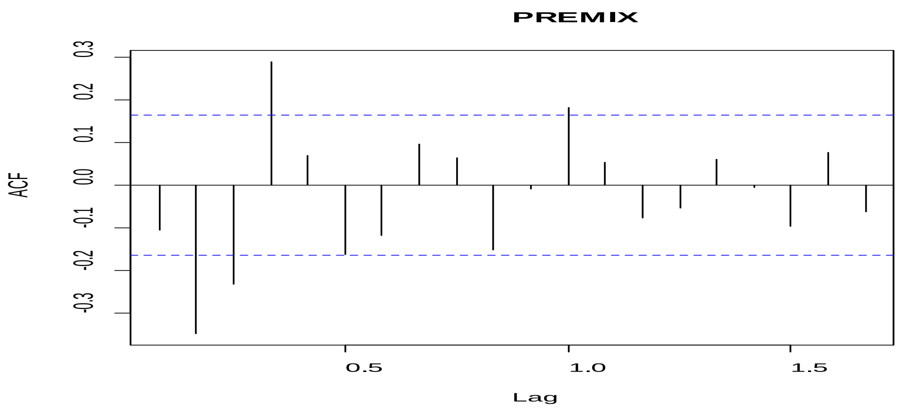


Figure 4.40: Autocorrelation function plot of the differenced Premix Fuel demand data

Figure 4.40 is the ACF plot of the first differencing of the national Premix demand data at various lags. The ACF plot indicates significant error limits at lags 2, 3, 4, 6, and 12, meaning possible MA models are MA(2), MA(3), MA(4), MA(6) and MA(12) behaviors respectively. Because the error limit at lag 5 is not significant and the first to behave as such, we use a q value of 4 that is MA(4).

b. PACF Plot

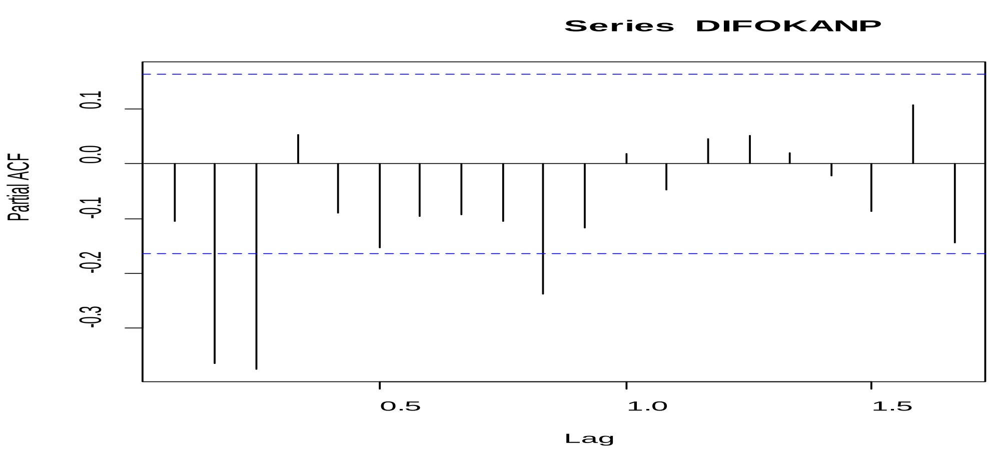


Figure 4.41: Partial Autocorrelation function plot

Figure 4.41 is the PACF plot of the first differencing of the Premix fuel demand data at various lags. The PACF plot indicates significant error limits at lags 2, 3, and 10 meaning possible AR models are AR(2), AR(3), and AR(10) behaviors respectively. Because the error limit at lag 4 in the PACF plot is not significant and the first to behave as such, we use a p value of 3 that is AR(3). By using the parsimony principle AR(3) and MA(4) are selected. The following models are recommended;

- ARIMA(3,1,4) $(2,0,0)_{[12]}$
- ARIMA(0,1,4)$(2,0,0)_{[12]}$
- ARIMA(3,1,0)(**2,0,0)**$_{[12]}$

4.9 MODEL SELECTION FOR THE NATIONAL DOMESTIC PREMIX FUEL DEMAND DATA

The time series analysis for the national domestic gas oil demand for the chosen candidate models is as shown below:

4.9.1 Parameter Estimates and Diagnostics of ARIMA(3, 1,4)(2,0,0)$_{[12]}$ Model For The Premix Fuel Data

Parameter estimates determine the coefficients of the time series equation that is generated from the data and the diagnostics test is used to check the correlation and significance of the residuals.

4.9.1.1 Parameter Estimates

The R result for the ARIMA(3,1,4)(2,0,0)$_{[12]}$ model is as shown below;

Coefficients:

	ar1	ar2	ar3	ma1	ma2	ma3	ma4	sar1	sar2	intercept
	-0.6485	-0.5520	-0.4088	0.2209	-0.0795	-0.3303	-0.1602	0.0947	0.4276	-10265.99
s.e.	0.2887	0.1936	0.1891	0.2944	0.1713	0.1588	0.2127	0.0790	0.1027	39682.92

sigma^2 estimated as 1.041e+12: log likelihood=-2184.68

AIC=4389.36 AICc=4391.37 BIC=4421.95

To further analyze the results we must appreciate how significant the parameters are by taking the t-test on them. The t-test of the parameters are as follows;

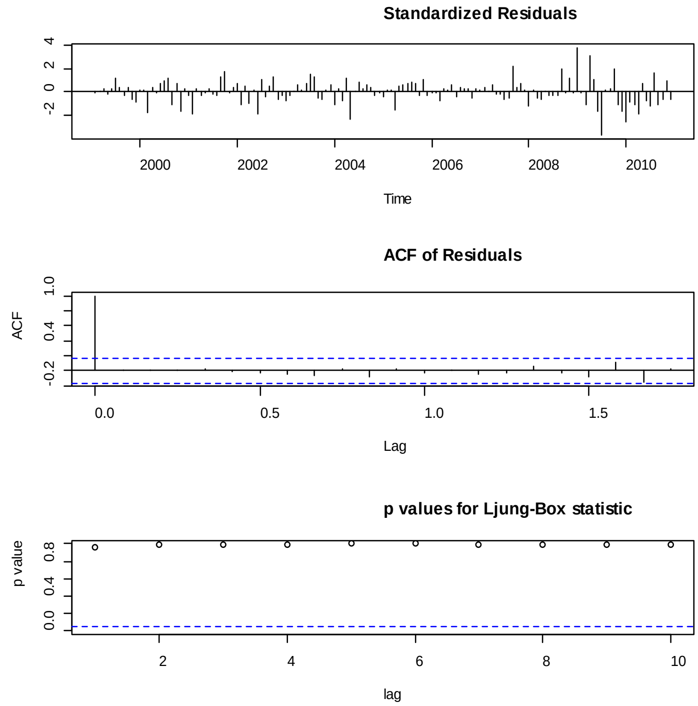


Figure 4.42: Diagnostic test plot for SARIMA(3,1,4)(2,0,0)$_{[12]}$

a. The top box of figure 4.42 above contains the time plot of the standardized residuals of the mode which shows that no obvious pattern and looks like an independent identical distribution (i.i.d) sequence with mean zero and few outliers.

b. The middle part of figure 4.42 is the ACF plot of the residuals and this shows evidence of significance correlation at lags 3 and 4 probably due to outliers.

c. The bottom part of figure 4.42 shows the residual plot for a model SARIMA(3,1,4)(2,0,0)$_{[12]}$, the box test of SARIMA(3,1,4)(2,0,0)$_{[12]}$ shows a p-value of 0.9858 and the p-value of 0.9858 of the residuals by the model confirms the non-significance of the residuals and so, the residuals of SARIMA(3,1,4)(2,0,0)$_{[12]}$ are independent and identically distributed.

The distribution of the errors by SARIMA(3,1,4)(2,0,0)$_{[12]}$ model is as shown in figure 4.43 below. From the diagram it can be observed that the residuals appear to be normally distributed since most of the data points are on the normal line and those which are not on the line deviate to the similar extent below and above the normal line.

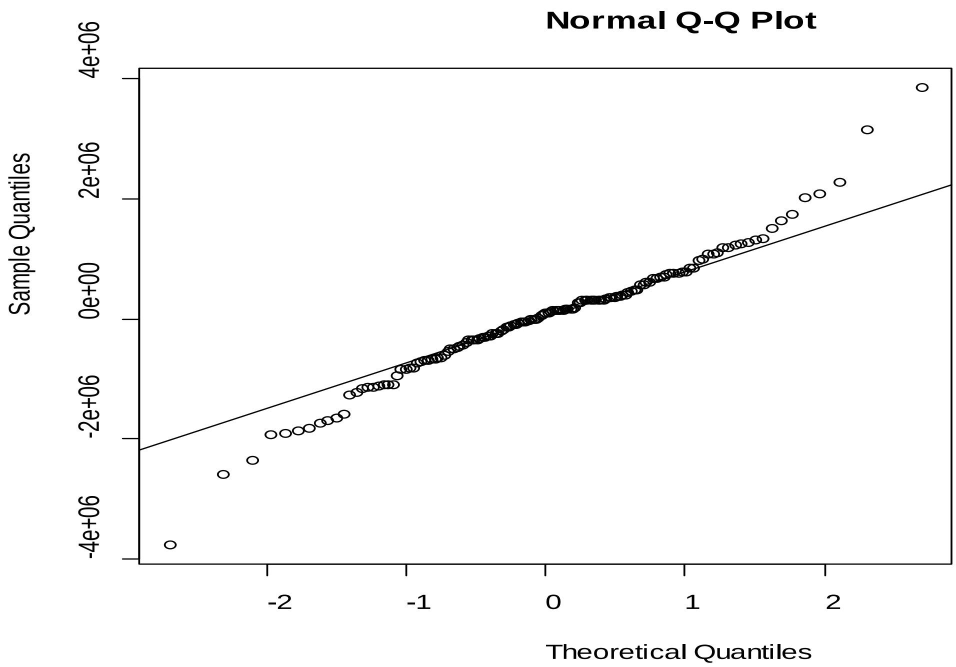


Figure 4.43: Normal Q-Q plot of the residuals of SARIMA(3, 1, 4)(2,0,0)$_{[12]}$

The histogram plot fitted with the normal curve to show the normality of the residuals for SARIMA(3, 1, 4)(2,0,0)$_{[12]}$ model is as shown in figure 4.35 below and this confirms the normality of the residuals of SARIMA(3, 1, 4)(2,0,0)$_{[12]}$ model;

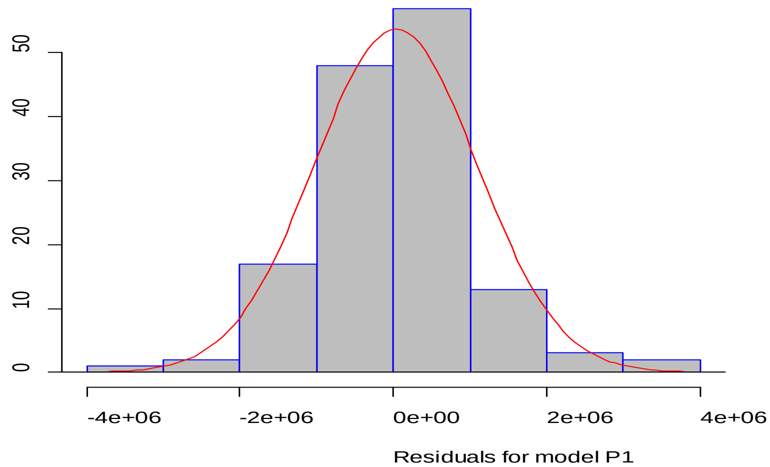


Fig.4.44:Plot of histogram fittedwith normal curve ofthe residuals of SARIMA(3,1,4) $(2,0,0)_{[12]}$

4.9.2 Parameter Estimates and Diagnostics of SARIMA(0, 1, 4)$(2,0,0)_{[12]}$ Model For The Premix Fuel Data

Parameter estimates determine the coefficients of the time series equation that is generated from the data and the diagnostics test is used to check the correlation and significance of the residuals.

4.9.2.1 Parameter Estimates

The R result for the SARIMA(0,1,4)$(2,0,0)_{[12]}$ model is as shown below;

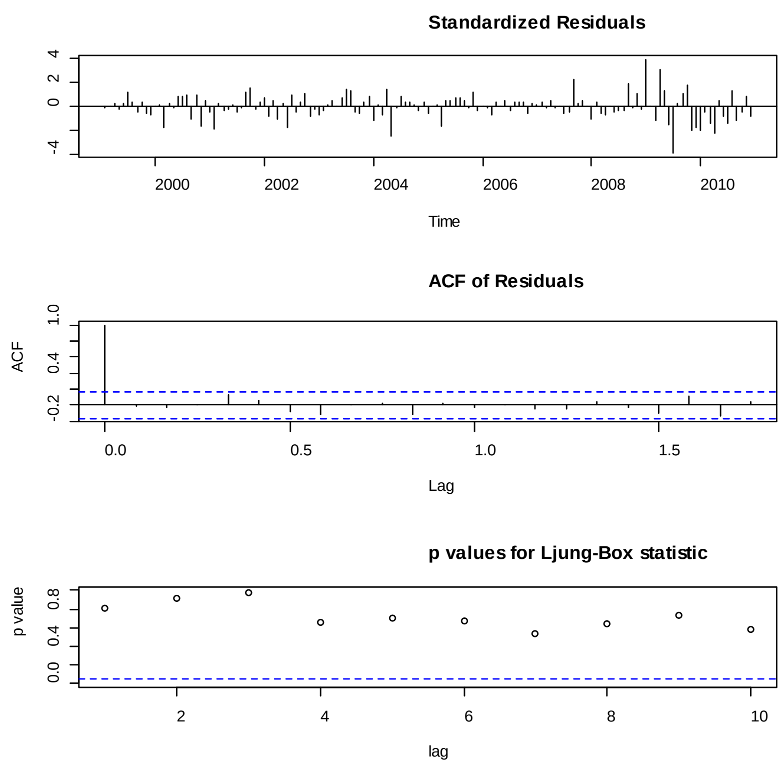
Coefficients:

	ma1	ma2	ma3	ma4	sar1	sar2	intercept
	-0.3850	-0.3601	-0.2939	0.2762	0.0881	0.3968	-9233.326
s.e.	0.0805	0.0896	0.0872	0.0776	0.0798	0.1084	36652.016

sigma^2 estimated as 1.085e+12: log likelihood=-2187.15

AIC=4388.31   AICc=4389.38   BIC=4412.01

To further analyze the results we must appreciate how significant the parameters are by taking the t-test on them. The t-test of the parameters is as follows;



Standardized Residuals
Time
ACF of Residuals
ACF
Lag
p values for Ljung-Box statistic
p value
lag

a. The top box of figure 4.45 contains the time plot of the standardized residuals of the mode which shows that no obvious pattern and looks like an independent identical distribution (i.i.d) sequence with mean zero and few outliers.

b. The middle part of the figure is the ACF plot of the residuals and this shows evidence of significance correlation at lags 3 and 4 probably due to outliers.

c. The bottom of the figure shows the Ljung-Box plot which suggests that the residuals of the LPG data by the SARIMA(0,1,4)(2,0,0)$_{[12]}$ are not significant at any positive lag and the p-value of 0.5802 of the residuals by the model confirms the non-significance of the residuals and so, the residuals of the SARIMA(0,1,4)(2,0,0)$_{[12]}$ appear to be independent and identically distributed.

The distribution of the errors by SARIMA(0,1,4)(2,0,0)$_{[12]}$ model is as shown in figure 4.46 below. From the diagram it can be observed that the residuals appear to be normally distributed since most of the data points are on the normal line and those which are not on the line deviate to the similar extent below and above the normal line.

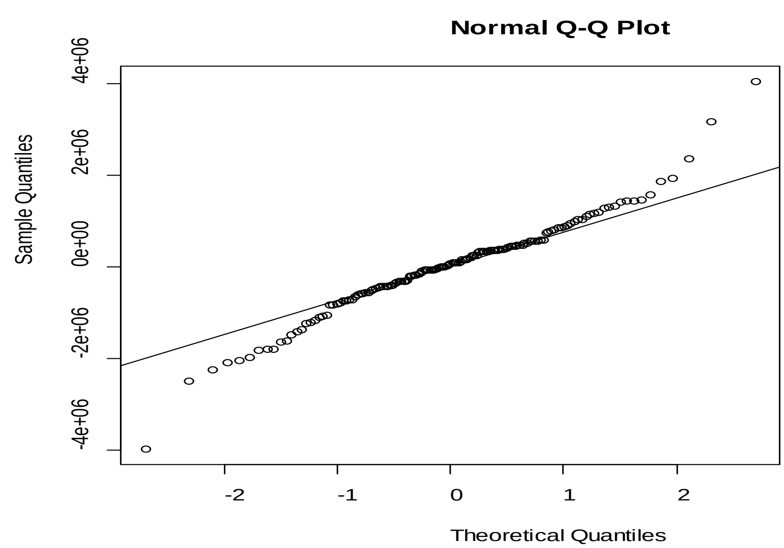


Figure 4.46: Normal Q-Q plot of the residuals of SARIMA(0, 1, 4)(2,0,0)$_{[12]}$

The histogram plot fitted with the normal curve to show the normality of the residuals for SARIMA(0, 1, 4)(2,0,0)$_{[12]}$ model is as shown in figure 4.47 below and this confirms the normality of the residuals of SARIMA(0, 1, 4)(2,0,0)$_{[12]}$ model;

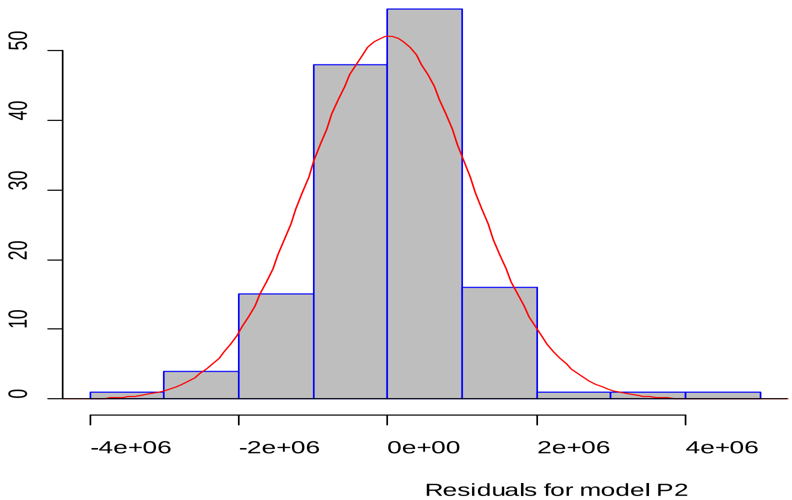


Figure4.47: Plot of histogram fitted with normal curve of the residuals for SARIMA(0,1,4)(2,0,0)$_{[12]}$

4.9.3 Parameter Estimates and Diagnostics of SARIMA(3, 1, 0)(2,0,0)$_{[12]}$ Model For The Premix Fuel Data

Parameter estimates determine the coefficients of the time series equation that is generated from the data and the diagnostics test is used to check the correlation and significance of the residuals.

4.9.3.1 Parameter Estimates

The R result for the SARIMA(3,1,0)(2,0,0)$_{[12]}$ model is as shown below;

Coefficients:

ar1 ar2 ar3 sar1 sar2 intercept

	-0.4014	-0.4535	-0.482	0.1377	0.4296	-13553.47
s.e.	0.0760	0.0707	0.074	0.0771	0.0997	73570.77

sigma^2 estimated as 1.078e+12: log likelihood=-2187.25

AIC=4386.5   AICc=4387.33   BIC=4407.24

To further analyze the results we must appreciate how significant the parameters are by taking the t-test on them. The t-test of the parameters are as follows;

4.9.3.2 Diagnostic Test

The diagnostic test plot is a shown in figure 4.48 below;

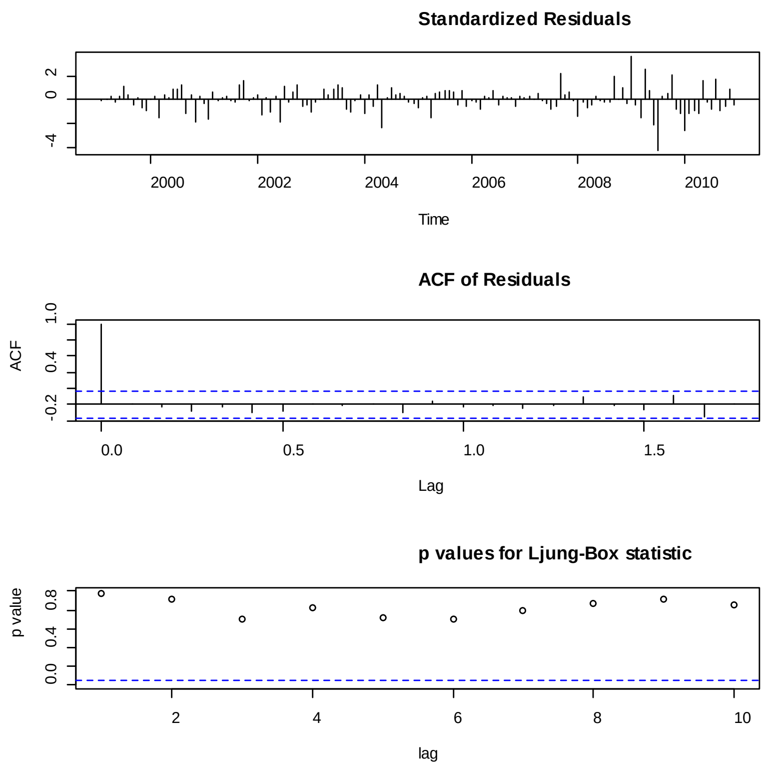


Figure 4.48: Diagnostic test for SARIMA(3,1,0)(2,0,0)$_{[12]}$

a. The top part of figure 4.48 box contains the time plot of the standardized residuals of the mode which shows that no obvious pattern and looks like an independent identical distribution (i.i.d) sequence with mean zero and few outliers.

b. The middle part of figure 4.48 is the ACF plot of the residuals and this shows evidence of no significant correlation at the various lags.

c. The bottom part of figure 4.48 shows the Ljung-Box plot which suggests that the residuals of the LPG data by the SARIMA(3,1,0)(2,0,0)$_{[12]}$ are not significant and the p-value of 0.8561 of the residuals by the model SARIMA(3,1,0)(2,0,0)$_{[12]}$ confirms the non-significance of the residuals and so, the residuals of the SARIMA(3,1,0)(2,0,0)$_{[12]}$ appear to be independent and identically distributed.

The distribution of the errors by SARIMA(3,1,0)(2,0,0)$_{[12]}$ model is as shown below. From the diagram it can be observed that the residuals appear to be normally distributed.

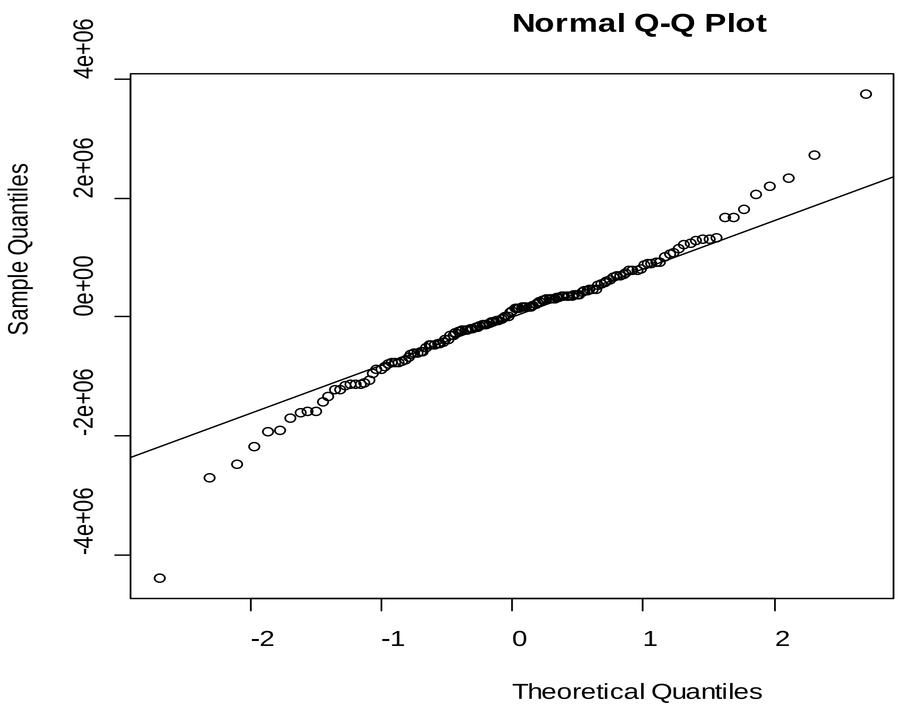


Figure 4.49: Normal Q-Q plot of the residuals of SARIMA(3,1,0)(2,0,0)$_{[12]}$

The histogram plot fitted with the normal curve to show the normality of the residuals for SARIMA(3,1,0)(2,0,0)$_{[12]}$ model is as shown in figure 4.50 below and this confirms the normality of the residuals of SARIMA(0, 1, 4)(2,0,0)$_{[12]}$;

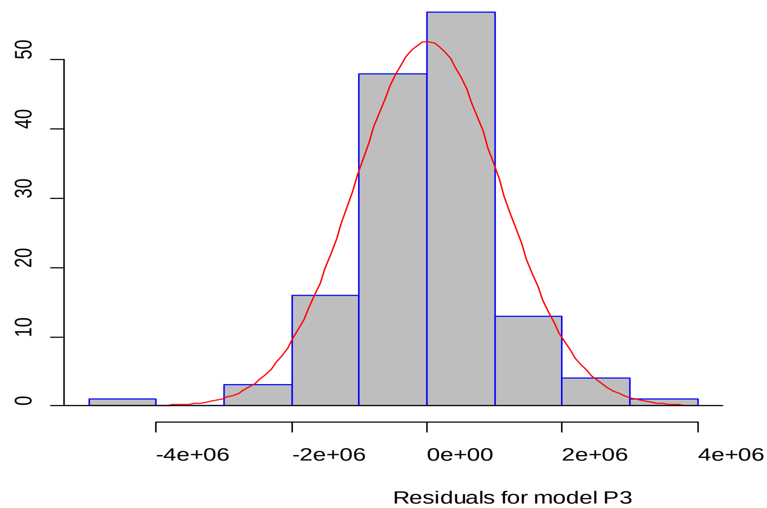


Fig.4.50: Plot of histogram fitted with normal curve of the residuals of SARIMA(3,1,0)(2,0,0)$_{[12]}$

4.9.4 Best Model Selection

Comparing the AIC, AICc and BICs of the candidate models chosen using the principle of parsimony and the over fit model, the AIC and BIC of ARIMA(3,1,0)(2,0,0)$_{[12]}$ model is better since it has the least of such values compared to the other models as shown in the table below.

MODEL	**AIC**	**AICc**	**BIC**
ARIMA(3,1,4)(2,0,0)$_{[12]}$	4389.36	4391.37	4421.95
ARIMA(0,1,4)(2,0,0)$_{[12]}$	4388.31	4389.38	4412.01
ARIMA(3,1,0)(2,0,0)$_{[12]}$	4386.50	4387.33	4407.24

Table 4.10: A table of models and their corresponding AIC, AICc and BIC for the LPG data.

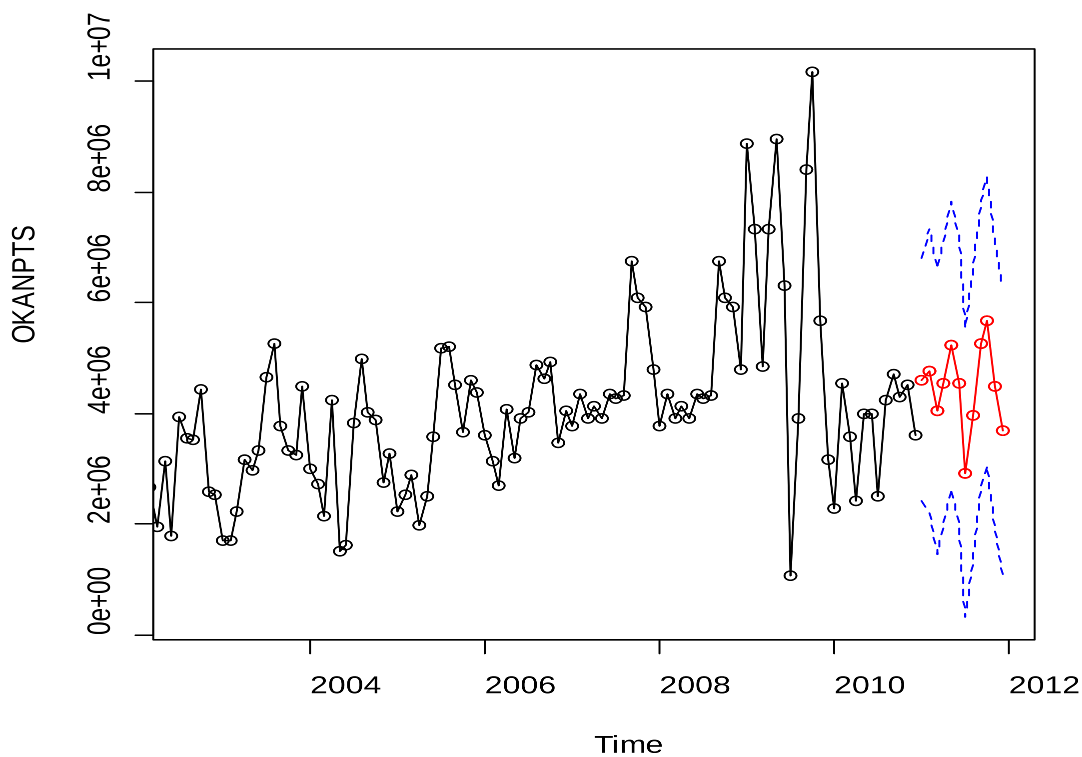
Though the residuals of the three models are all significant but due to the fact that the parameters of the SARIMA(3,1,4)(2,0,0)$_{[12]}$ and SARIMA(0,1,4)(2,0,0)$_{[12]}$ are not statistically significant but that of ARIMA(3,1,0)(2,0,0)$_{[12]}$ are all significant and so we use SARIMA(3,1,0)(2,0,0)$_{[12]}$ for forecasting.

4.9.5 Fitting the National Premix Fuel Demand Model

The best model for forecasting the National Premix Fuel demand model is ARIMA(3,1,0)(2,0,0)$_{[12]}$ which has three non-seasonal AR terms, a single differencing and no non-seasonal MA terms. Also, it has two seasonal AR terms with no seasonal MA terms.

In terms of the observed series

Therefore, the fitted equation for ARIMA(3,1,0)(2,0,0)$_{[12]}$ model for the National Domestic Gas Oil demand data from 1999 to 2010 is given by;



fuel demand will increase small in the first quarter of 2011 and experience a down drop in the first month of the second quarter and this will give rise to a sudden rise in the last month of the second quarter. The third quarter of 2011 will see a downward demand trend till about the middle of the quarter and rise again till the about the second month of the last quarter, the demand in the last quarter will then decrease to the end of 2011.

A 12 point prediction into the future by the model is as shown in table 4.11 below;

Table of Forecasted Values for the Premix Data in Litres

MODEL	POINT FORECAST	LOW 95%	HIGH 95%
January	3557027	1810470	5303583
February	3299784	1615196	4984371
March	2851015	1341734	4360296
April	3518732	1591562	5445902
May	3428266	1489633	5366899
June	3507287	1463173	5551401
July	4612523	1794629	7172504
August	4640095	1770019	7455027
September	4785007	1705567	7574624
October	4640095	1683029	7886986
November	3888795	1307385	6470205
December	3681470	1181530	6181411

Table 4.11: Predicted values at 95% confidence intervals showing the lowest (Lo) and highest (Hi) values for the intervals

4.9.7 Forecast Accuracy

The forecast values SARIMA(3,1,0)(2,0,0)$_{[12]}$ model for the gas oil data was tested to how close its predicted values are to the actual values left for testing the validity of the models by considering the error margins. We used the the Root Mean Square Percentage Error (RMSPE). The RMSPE test gave a value of 2.42% also meaning that there is only 2.42% error in using SARIMA(3,1,0)(2,0,0)$_{[12]}$ to forecast into the future. The RSMPE gives a tolerable percentage error.

CHAPTER 5

CONCLUSION AND RECOMMENDATIONS

5.1 Conclusion

The study embellishes the application of time series analysis using the Box-Jenkins approach to modelling petroleum products demand in Ghana. Thus, monthly data from 1999 to 2011 of petroleum products namely; Gas Oil, Liquefied Petroleum Gas (LPG) and Premix Fuel from the national petroleum authority were analysed and forecasted. By so doing the behaviour of the sample correlation function (ACF) of the monthly data of the three petroleum products reveal the existence of trend component in the Gas Oil and the Liquefied Petroleum Gas products data and the existence of seasonal component in the Premix Fuel data and these called for further analysis of the data. Therefore, a first order differencing was performed on the data to remove the trend and seasonality components in order to achieve stationarity in the data.

Each of the stationary data was examined by first differencing them and secondly considering their sample ACF and the sample Partial Autocorrelation Function (PACF) of the differenced petroleum product demand data. After considering the AIC, AICc and BIC of the candidate models under the various data those models with smallest AIC, AICc and BIC were chosen as the best-fit models among the candidate models. For the national gas oil demand data, ARIMA(1,1,3) model was selected as the best-fit model for predicting future demand values for the national gas oil demand. For the national LPG demand data, ARIMA(2,1,3) model was selected as the best-fit model for predicting future demand values for the national LPG demand. For the national Premix Fuel demand data, ARIMA(3,1,0)(2,0,0)$_{[12]}$ model was selected as the best-fit model for predicting future demand values for the national Premix Fuel demand.

To verify the goodness of fit of the models the root mean square percentage error and mean absolute percentage error were performed using the forecast values from each of the models. The RMSPE values for the forecast values of the gas oil by ARIMA(1,1,3) is 8.59% meaning there is an error of about 8% in the predicted values by ARIMA(1,1,3) for the gas oil data. The RMSPE values for the forecast values of the LPG by ARIMA(2,1,3) is 1.09% there is an error of about 1% in the predicted values by ARIMA(2,1,3) for the LPG data. The RMSPE values for the forecast values of the premix fuel by SARIMA(3,1,0)(2,0,0)$_{[12]}$ is 2.42% meaning there is an error of about 2% by the RMSPE method in the predicted values by SARIMA(3,1,0)(2,0,0)$_{[12]}$ for the premix fuel data. In effect the Box-Jenkins approach for time series analysis could find models that best fit the data of the three petroleum products. Interpreted the characteristics of the demand pattern for the chosen petroleum products in Ghana and in particular, the demand pattern for premix fuel was found to be seasonal and confirmed the prevailing market conditions in July, August and September.

5.2 Recommendations

The recommended models for predicting future values for the National Gas Oil demand, National LPG demand and National Premix demand are respectively ARIMA(1,1,3), ARIMA(2,1,3) and ARIMA(3,1,0)(2,0,0)$_{[12]}$. In implementing these models the following must be considered:

1. To allow for effective planning of the supply of petroleum products we recommend that the findings should be adopted by stake holders
2. We also recommend that further studies be undertaken using other time series analysis that incorporate exogenous factors to explain the macro economic implications

REFERENCES

Abdel-Aal, R. E., (2008). Computers and Industrial Engineering. Volume 54 Issue 4

Abeysinghe, T., Wilson, P., (2000). Could high oil prices derail Asian recovery? Press Release, Econometric Studies Unit, Department of Economics, National University of Singapore, http: / /www.fas.nus.edu.sg/ ecs / center /esu / oilprice.pdf.

African Developing Bank, (2008). Annual Report

Alexaandru, A., (2008). Modelling unemployment rate using Box-Jenkins procedure: Journal of applied quantitative methods.

Atger, F.,(2001). Verification of intense precipitation forecasts from single models and ensemble prediction systems. Nonlin. Proc. Geophys., 8, 401-417.

Ayeni, B.J., Mohn, S.W., and Rollins, D.K., (2001). A Time Series Investigation of Cast Web Caliper Instability, 3M Internal Publication.

Casati, B., Ross, D.B., Stephenson, (2004). A new intensity-scale approach for the verification of spatial precipitation forecasts, Meteorol. Appl., 11, pages 141-154.

Cryer, J. D., Chan, Kung. (2008).Time Series Analysis 2nd Edition. Springer Publications

DeGroot, M.H., and Fienberg, S.E., (1983). The comparison and evaluation of forecasters. The Statistician, 32, 14-22.

Dohner, R.S. (1981), “Energy prices, economic activity, and inflation: a survey of issues and results,” in K.A. Mork (ed.), *Energy Prices, Economic Activity, and Inflation,* Cambridge Mass): Ballinger.

Ehrendorfer, M., Murphy A.H., (1988). Comparative evaluation of weather forecasting systems: Sufficiency, quality, and accuracy. Mon. Wea. Rev., 116, 1757-1770.

Hamill, T.M., (2001): Interpretation of rank histograms for verifying ensemble forecasts. Mon. Wea. Rev., 129, 550-560.

Harris, D., Foufoula-Georgiou, E., Droegemeier, K.K., and Levit, J.J., (2001): Multiscale statistical properties of a high-resolution precipitation forecast. J. Hydromet., 2, 406-418.

Hamilton, J. D., (1994). Time Series Analysis, Princeton: Princeton University Press.

Jewson, S., (2003): Use of the likelihood for measuring the skill of probabilistic forecasts. http://arxiv.org/PS_cache/physics/pdf/0308/0308046v2.pdf

Lorek K.S., Willinger G.L., (2006). Multivariate time series prediction model for cash-flow data Illinois University Press.

Katz, R.W. and Murphy, A.H. (eds),. (1997). Economic Value of Weather and Climate Forecasts. Cambridge University Press, Cambridge.

Keil, C. and Craig, G.C., (2007). A displacement-based error measure applied in a regional ensemble forecasting system. Mon. Wea. Rev., 135, 3248-3259.

Koh, T. Y. and Ng, J. S., (2009). Improved diagnostics for NWP verification in the tropics. J. Geophys. Res., 114, D12102, doi:10.1029/2008JD011179.

Krzysztofowicz, R., (1992). Bayesian correlation score: A utilitarian measure of forecast skill. Mon. Wea. Rev., 120, 208-219.

Laio, F., Tamea, S., (2007): Verification tools for probabilistic forecasts of continuous hydrological variables. Hydrol. Earth Syst. Sci., 11, 1267-1277.

Loungani, Prakash, (1986).Oil Price Shocks and the Dispersion Hypothesis,. Review

Mason, I., (1982). A model for assessment of weather forecasts. Aust. Met. Mag., 30, 291-303.

Mason, S.J. and Graham, N.E., (1999). Conditional probabilities, relative operating characteristics, and relative operating levels. Weather Forecasting, 14, 713-725.

Mason, S.J., (2008). Understanding forecast verification statistics. Meteorol. Appl., 15-23.

Mork, K. A. (1994). Business Cycles and the Oil Market, Energy Journal 15 Special Issue 15-38.

Mosse, J.A. and Katz, R.W., ed., (1985). Probability, Statistics, and Decision Making in the Atmospheric Sciences. Westview Press, Boulder, CO.

N. Rangsan & Titida, *ARIMA Model For Forecasting Oil Palm Price*, Proceedings of the 2nd Indonesia-Malaysia-Thailand Growth Triangle (IMT-GT) Regional Conference on Mathematics, Statistics, and Applications, (2008). Universities Saints Malaysia, Penang, Malaysia, Vol. 3 (Statistics), pp. 16 – 21.

Pankratz, A., (1983). Forecasting with Univariate Box-Jenkins Model, (New York: John Wiley and Sons).

Shehiel A., (2009). Does Energy consumption contribute to economic performance? Empirical evidence from Nigeria. Journal of Economics and Business Vol.XII,:

Thornes, J.E. and Stephenson D.B., (2001). How to judge the quality and value of weather forecast products. Meteorol. Appl., 8, 307-314.

Volkan, S., Edigera, Sertac- Akar, (2007) ARIMA forecasting of primary energy demand by fuel in Turkey

APPENDIX

R Codes For LPG Data

```
library(forecast)
library(tseries)
library(TSA)
OKAN=read.csv("C://Users/Ras Wailer/Desktop/The Three/INITIALIZATION SET/LPG.csv")
OKAN
OKANTS=ts(OKAN,start=1999,frequency=12)
OKANTS
AA=sqrt(OKANTS)
AA
acf(OKANTS)
plot(OKANTS,xlab="YEAR",ylab="NATIONAL LPG DEMAND")
DIFOKAN=diff(OKANTS)
DIFOKAN
plot(DIFOKAN,xlab="YEAR",ylab="DIFFERENCED NATIONAL LPG DEMAND")
kpss.test(DIFOKAN)
pp.test(DIFOKAN)
acf(DIFOKAN,20)
pacf(DIFOKAN,20)
K=arima(DIFOKAN,order=c(1,0,1))
K
tsdiag(K)
Box.test(K$resid,type="Ljung-Box",lag=10)
qqnorm(resid(K));qqline(resid(K))
```

```
normal.freq(hist(resid(K),br=10,border="Blue",col="Grey",main="",xlab="Residuals for model K",ylab=""),col="red")

KK=arima(DIFOKAN,order=c(0,0,1))

KK

tsdiag(KK)

Box.test(KK$resid,type="Ljung-Box",lag=10)

qqnorm(resid(K));qqline(resid(KK))

normal.freq(hist(resid(KK),br=10,border="Blue",col="Grey",main="",xlab="Residuals for model KK",ylab=""),col="red")

KKK=arima(DIFOKAN,order=c(1,0,0))

KKK

tsdiag(KKK)

Box.test(KKK$resid,type="Ljung-Box",lag=10)

qqnorm(resid(KKK));qqline(resid(K))

normal.freq(hist(resid(KKK),br=10,border="Blue",col="Grey",main="",xlab="Residuals for model KKK",ylab=""),col="red")

K1=arima(DIFOKAN,order=c(2,0,3))

K1

tsdiag(K1)

Box.test(K1$resid,type="Ljung-Box",lag=10)

qqnorm(resid(K));qqline(resid(K1))

normal.freq(hist(resid(K1),br=10,border="Blue",col="Grey",main="",xlab="Residuals for model K1",ylab=""),col="red")

K1F=sarima.for(OKANTS,12,2,1,3)

K1F

PLPG=predict(OKANTS,h=12)

PLPG
```

R Codes For Gas Oil Data

```
library(forecast)

library(tseries)

library(TSA)

OKANGO=read.csv("C://Users/Ras Wailer/Desktop/The Three/INITIALIZATION SET/GAS_OIL.csv")

OKANGO

OKANGOTS=ts(OKANGO,start=1999,frequency=12)

OKANGOTS

acf(OKANGOTS)

plot(OKANGOTS,xlab="YEAR",ylab="NATIONAL GAS OIL DEMAND")

DIFOKANGO=diff(OKANGOTS)

DIFOKANGO

plot(DIFOKANGO,xlab="YEAR",ylab="DIFFERENCED NATIONAL GAS OIL DEMAND")

kpss.test(DIFOKANGO)

pp.test(DIFOKANGO)

adf.test(DIFOKANGO)

acf(DIFOKANGO,20)

pacf(DIFOKANGO,20)

GO=arima(DIFOKANGO,order=c(1,0,1))

GO

tsdiag(GO)

Box.test(GO$resid,type="Ljung-Box",lag=10)

qqnorm(resid(GO));qqline(resid(GO))

normal.freq(hist(resid(GO),br=10,border="Blue",col="Grey",main="",xlab="Residuals for model GO",ylab=""),col="red")

GO1=arima(DIFOKANGO,order=c(0,0,1))
```

```
GO1

tsdiag(GO1)

Box.test(GO1$resid,type="Ljung-Box",lag=10)

qqnorm(resid(GO));qqline(resid(GO1))

normal.freq(hist(resid(GO1),br=10,border="Blue",col="Grey",main="",xlab="Residuals for model GO1",ylab=""),col="red")

GO2=arima(DIFOKANGO,order=c(1,0,0))

GO2

tsdiag(GO2)

Box.test(GO2$resid,type="Ljung-Box",lag=10)

qqnorm(resid(GO2));qqline(resid(GO2))

normal.freq(hist(resid(GO2),br=10,border="Blue",col="Grey",main="",xlab="Residuals for model GO2",ylab=""),col="red")

G=arima(DIFOKANGO,order=c(1,0,3))

G

tsdiag(G)

Box.test(G$resid,type="Ljung-Box",lag=10)

qqnorm(resid(G));qqline(resid(G))

normal.freq(hist(resid(G),br=10,border="Blue",col="Grey",main="",xlab="Residuals for model G",ylab=""),col="red")

GG=sarima.for(OKANGOTS,12,1,1,3)

GG

FG=predict(OKANGOTS,h=12)

FG
```

R Code For Premix Oil Data

```
library(forecast)

library(tseries)

library(TSA)

OKANP=read.csv("C://Users/Ras Wailer/Desktop/The Three/INITIALIZATION SET/PREMIX1.csv")

OKANP

OKANPTS=ts(OKANP,start=1999,frequency=12)

OKANPTS

acf(OKANPTS)

plot(OKANPTS,xlab="YEAR",ylab="NATIONAL PREMIX DEMAND")

DIFOKANP=diff(OKANPTS)

DIFOKANP

plot(DIFOKANP,xlab="YEAR",ylab="DIFFERENCED NATIONAL PREMIX DEMAND")

kpss.test(DIFOKANP)

pp.test(DIFOKANP)

adf.test(DIFOKANP)

acf(DIFOKANP,20)

pacf(DIFOKANP,20)

P1=arima(DIFOKANP,order=c(3,0,4),seasonal=c(2,0,0))

P1

tsdiag(P1)

Box.test(P1$resid,type="Ljung-Box",lag=10)

qqnorm(resid(P1));qqline(resid(P1))

normal.freq(hist(resid(P1),br=10,border="Blue",col="Grey",main="",xlab="Residuals for model P1",ylab=""),col="red")

P2=arima(DIFOKANP,order=c(0,0,4),seasonal=c(2,0,0))

P2
```

```
tsdiag(P2)

Box.test(P2$resid,type="Ljung-Box",lag=10)

qqnorm(resid(P2));qqline(resid(P2))

normal.freq(hist(resid(P2),br=10,border="Blue",col="Grey",main="",xlab="Residuals for model P2",ylab=""),col="red")

P3=arima(DIFOKANP,order=c(3,0,0),seasonal=list(order=c(2,0,0),period=12))

P3

predict(P3,n.ahead=12)

tsdiag(P3)

Box.test(P3$resid,type="Ljung-Box",lag=10)

qqnorm(resid(P3));qqline(resid(P3))

normal.freq(hist(resid(P3),br=10,border="Blue",col="Grey",main="",xlab="Residuals for model P3",ylab=""),col="red")

SS=sarima.for(OKANPTS,24,3,0,0,2,0,0,12)

PPREF=predict(OKANPTS,h=12)

PPREF
```

R Code Anova Test For Seasonality

```
premixdata=read.table(file.choose(),header=T)
names(premxdata)
mod=aov(Premix~Month,data=premixdata)
summary(mod)
TukeyHSD(mod, "Month", ordered = TRUE)
```

R Code For RMSPE

Gas Oil

```
ACTUAL=read.csv("C://Users/Ras Wailer/Desktop/ACTUAL.csv")
ACTUAL
PREDICTED=read.csv("C://Users/Ras Wailer/Desktop/PREDICTED.csv")
PREDICTED
sqrt(sum( ((ACTUAL-PREDICTED)/ACTUAL)^2 )*(100/11))
```

LPG

```
ACTUAL=read.csv("C://Users/Ras Wailer/Desktop/ACTUALLPG.csv")
ACTUAL
PREDICTED=read.csv("C://Users/Ras Wailer/Desktop/PREDICTEDLPG.csv")
PREDICTED
sqrt(sum( ((ACTUAL-PREDICTED)/ACTUAL)^2 )*(100/11))
```

PREMIX FUEL

```
ACTUAL=read.csv("C://Users/Ras Wailer/Desktop/ACTUALPREMIX.csv")
ACTUAL
PREDICTED=read.csv("C://Users/Ras Wailer/Desktop/PREDICTEDPREMIX.csv")
PREDICTED
sqrt(sum( ((ACTUAL-PREDICTED)/ACTUAL)^2 )*(100/11))
```