UNIVERSITY OF CAPE COAST

MULTIVARIATE TIME SERIES MODELLING OF EX-PUMP PRICES OF PETROLEUM PRODUCTS IN GHANA

THEOPHILUS ASAMOAH

2020

UNIVERSITY OF CAPE COAST

MULTIVARIATE TIME SERIES MODELLING OF EX-PUMP PRICES OF PETROLEUM PRODUCTS IN GHANA

BY

THEOPHILUS ASAMOAH

Thesis submitted to the Department of Statistics of School of Physical Sciences, College of Agriculture and Natural Sciences, University of Cape Coast, in partial fulfillment of the requirements for the award of Master of Philosophy degree in Statistics

OCTOBER, 2020

DECLARATION

Candidate's Declaration

I hereby declare that this thesis is the result of my own original research and that no part of it has been presented for another degree in this university or elsewhere.

Candidate's Signature:…………………………… Date:…………………………

Candidate's Name: Theophilus Asamoah

Supervisor's Declaration

We hereby declare that the preparation and presentation of this thesis were supervised in accordance with the guidelines on supervision of thesis laid down by the University of Cape Coast.

Principal Supervisor's Signature: ……………………..…… Date: ………………….

Name: Prof. Bismark Kwao Nkansah

Co-Supervisor's Signature: ………………..……………Date: ...…………………

Name: Dr. Alexander Boateng

ABSTRACT

The purpose of the study is to obtain a suitable model for the ex-pump prices of petroleum products in Ghana. It examines how changes in the prices of one product cause changes in the price of others in both the short and long terms. Data spanning January 2007 to June 2015 are obtained from the National Petroleum Authority of Ghana, covering four petroleum products; Premium Gasoline, Gas Oil, Kerosene, and Liquefied Petroleum Gas. The analysis is carried out using the technique of Vector Error Corrected (VEC) modelling. This technique is found suitable as the data obtained constitutes a time series that is multivariate in nature, and that the components are found to exhibit long-run relationships. The study reveals that there exists a general upward trend in the ex-pump prices of the products over the period. It also shows that changes in prices of some of the products influence others in the short run, but in the medium to long terms, stability is attained, with Premium Gasoline and Gas Oil prices accounting for most of the variations in changes in prices of the four products. A VEC model of order 1 is found suitable for the data and appears to perform better than the VAR model with the least R-squared of 90.3% for Gasoline prices and the highest of 96% for Gasoil prices. The results reflect the competitive usage of the two products that could serve as substitutes. Thus, to ensure economic prices in petroleum products, in general, it would be expedient to ensure effective management of price changes in these two products.

KEY WORDS

Ex-pump prices

Long-term relationship

Multivariate time series modelling

Petroleum product

Vector Autoregression modelling

Vector Error Corrected

ACKNOWLEDGEMENT

I would like to express my deepest gratitude to my principal supervisor, Prof. B.K. Nkansah of the Department of Statistics, University of Cape Coast, under whose guidance and supervision this study has become a success. By posing insightful questions and suggestions, comments and guidance, the work has been successfully completed. I am also grateful to Dr. Alexander Boateng and Dr. David K. Mensah, who also supported the supervision of this study in diverse ways. My profound appreciation also goes to Dr. N. Howard of the Department of Statistics, for his support and cooperation in pursuing this programme. Again, my appreciation goes to all the lecturers of the Departments of Mathematics and Statistics, University of Cape Coast for their support and cooperation during the programme. To the National Petroleum Authority who made data available, I say thank you. Thanks also go to all my colleagues and the non-academic staff of the Departments of Mathematics and Statistics, for their support and cooperation. I appreciate the diverse ways in which each of them helped me to complete this programme. I would also like to thank Miss Sandra Dede Odjer and Miss Vivian Abena Owusu (University of Cape Coast) for their unconditional support and words of encouragement; God richly bless you all, Dede and Abena. Finally, I am singularly thankful to my mum, Madam Esther Yookenn Addi, my brother, Daniel Ofotsu Agu, and all my family members for their cooperation and support throughout the period of the programme.

DEDICATION

To My Lovely Family

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACF	Auto-correlation Function
ADF	Augmented Dickey-Fuller
AR	Autoregressive
ARCH	Autoregressive Conditional Heteroscedascity
ARIMA	Autoregressive Integrated Moving Average
ARMA	Autoregressive Moving Average
FEVD	Forecast Error Variance Decomposition
GARCH	Generalized Autoregressive Conditional Heteroscedascity
IPS	Im, Perseran and Shin
IRF	Impulse Response Function
KPSS	Kwiatkowski, Phillips, Schmidt, and Shin
LLC	Levin, Lin, and Chu
LPG	Liquefied Petroleum Gas
MA	Moving Average
NPA	National Petroleum Authority
PACF	Partial Auto-correlation Function
PP	Philip Perron
VAR	Vector Autoregression
VEC	Vector Error Corrected

CHAPTER ONE

INTRODUCTION

Background to the Study

In the contemporary trends of mechanization and growth, energy has become one of the most important steering wheels of an economy, of which Ghana is no exception. This is an indication that economies in the world today are deeply dependent on energy. In other words, more energy consumption results in a corresponding increase in economic activity. Energy is important for economic development and its demand is connected to factors such as prices, income, population, degree of development, level of technological development, and the overall structure of an economy. The energy sector, therefore, is one indispensable sector for a country's socio-economic growth, production, and general improvement in the standard of living.

The major sources of energy are plants, coal, petroleum, electricity, sun, geothermal steam, and animals. The effective demand for commercial energy is related to economic conditions that influence the availability of and access to energy. According to the International Energy Agency Staff (2012), the foremost energy sources in Ghana comprise biofuels and waste (69.5%) (By rural homes), crude oil (24.1%), with the least being Hydropower (6.4%). In another study, Armah (2003) found out that crude oil is the major energy source for Ghana's productive sectors, accounting for 96.7%, 52%, and 92% of energy consumption in agricultural, formal manufacturing, and transport sectors, respectively. Although being the second-highest energy source used in the country, petroleum products have a serious influence on the economy due to the sectors that use them. The activities of the transport sector, for instance, cut across all the other sectors of the economy (commercial, domestic and industrial). The manufacturing sector depends on transportation for the transport of raw materials and finished products to various destinations. Therefore, any change in the price of petroleum products affects transportation which causes a corresponding increase in goods and services.

Even though Ghana has discovered oil in commercial quantities, she still imports oil and any change in the prices of oil products from her oil-exporting countries poses economic challenges. According to the African Development Bank (ADB), the changing aspects of the worldwide energy markets are particularly marked by high-pitched increases in global requests and severe supply shockwaves hitting overall economies. These trends are worrying due to their effects on economic performance (ADB, 2008). The supply-side effects produce direct economic misrepresentations hitting oil-producing sectors. This is due to other production inputs, particularly, non-declining labour per unit cost of production, leading to a fall in levels of output, and causing profit margins of oil production sectors to drop, causing harm to the economy as a whole.

It is a well-known fact that a rise in oil price leads to deterioration in trade of net oil importing countries, and subsequently, leads to a reduction in the purchasing power of firms and households (Dohner, 1981). This is essentially a transfer of wealth from net oil importing to net oil exporting countries. Nonetheless, some contend that the special effects of high oil prices can also be indirect, which works through the economies' trading counterparts. 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 the decision taken by net exporting countries to spend additional windfall purchasing power and their trade preferences. As most net oil importing African economies are not well-diversified, 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.

Consumption is affected through its positive relation to disposable income, and investment through increasing firm costs, if the oil supply shortage increases over a long period, they may lead to a change in the production structure in favour of non-oil intensive sectors, which may result in distortions. The subsequent re-allocation of labour and capital across sectors in response to the price increases can affect the unemployment situation in the long term (Loungani, 1986). Overall, net oil importing countries like Ghana remain susceptible to energy price shocks, particularly because it is more or less non-export-oriented. Since 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 rising and given the unique opportunity offered by discoveries of oil and gas fields, an exploratory study of the situation in the country is appropriate, especially in the face of emerging indication of the impact of high level and volatility of oil prices. For these reasons, consumers of petroleum products in the country and the government must be assisted with information on the trends of prices of these products to be able to make incisive decisions and to forecast, which will aid in preventing the unpredicted devastation of high and volatile oil prices bring. Hence, the need for the current study.

Statement of the Problem

Energy usage pattern controls the foundation of the worldwide economy. Without energy, any form of physical production and movement is completely impossible. Even services are not likely to be performed without energy. For instance, economic growth has been largely the same with increasing energy use, since the start of the industrial revolution, generally at declining real prices (Noreng, 2007). High-pitched rises in oil prices have significant effects on economic and macroeconomic policies. Precisely, high prices in the oil markets in the world are causes of concern of possible falls in economic performance. Ghana is a low-middle-income country and striving to become a high-middle-income one. Consequently, the pace of its growth has become one of the major aims of every successive government. This growth is driven partly by energy, mainly, electricity and crude fuel. Globally, consumption of petroleum products exceeds $500 billion and it is roughly 10% of the United States Gross Domestic Products [USGDP] (Roger, 1998). The topic of petroleum product prices has been contentious. Ghana is not let off from the problems of prices of petroleum products that are acceptable to the population and would not cripple the economy. This idea has accounted for all pricing systems for petroleum products and hence, causes volatility in the pricing of these products. Ghana endures choice-making due to political crises and a deficit in the balance of trade account which tends to weaken its currency.

In 2015 for instance, the average price for Gasoline was about GH¢ 199.45 per barrel (Odoi, Twumasi-Ankrah, & Al-Hassan, 2017). This variability in price also occurs in other products (e.g., Kerosene, Premix, Marine Gas Oil Local, Liquefied Petroleum Gas, Gasoline, and Gas Oil). The leading crude fuel used in the country includes petrol (super), diesel, premix fuel, liquefied petroleum gas, among others. Because these fuels have become the wheels on which the economy thrives, there is a need to have a system by which the government and consumers will be informed of the prices on these fuels to prevent shortage or excess supply as this has a significant effect on the economy at large. More often than not, a shortage of these products leads to volatile and high prices. Companies and individuals normally fold out of business when there are increases in the prices of petroleum products. This is because most of them rely on loans accessed from banks and are not able to pay back due to sudden price increases. Thus, when there is a price increment, firms move from energy-intensive to energy-efficient. Because these adjustments are not achieved in the short run, there is a tendency for growth in unemployment and underutilization (Pindyck & Rotemberg, 1983). In Davis and Haltiwanger (2001), a suggestion is offered on the special effects of oil price changes on manufacturing jobs of the US. Also, in the view of Lee and Ni (2002), oil shocks do not affect all areas in the same way. More importantly, they found the severity of the negative effect of oil shocks not interrelated to energy intensity. Others (Van Soest et al., 2000; Bernanke, 1983) underlined the significance of doubt since in periods of oil price instability, firms have a motivation to delay investment choices. Thus, there is the need to have a system that can monitor the price pattern for petroleum products to be able to plan successfully. Such a system should track how changes in the price of a particular product influence the price of others, and the periods in which such changes are likely to occur. The variations of these products, therefore, need to be addressed and urgent attention drawn to ensure the stability of the economy.

Regarding the above problem, Odoi, et al. (2017) analyzed and modelled price volatility in petroleum products in Ghana. The methods used were the trend analysis and Generalized Auto-Regression Conditional Heteroscedasticity (GARCH). Agyen (2012) also applied times series modelling and forecasting techniques to these products in Ghana. However, his study focused on the demand for the products rather than their prices. He applied Autoregressive Integrated Moving Average (ARIMA) and Seasonal Autoregressive Integrated Moving Average (SARIMA) models. Similarly, Tetteh and Xu (2015) forecast refined petroleum products' future price trends in Ghana by employing Benchmark techniques, cointegration, GARCH, and Artificial Neural Network (ANN). For all the studies so far reviewed, (Odoi, et. al., 2017; Tetteh & Xu, 2015; Agyen, 2012), there is scarcely used of Vector Error Corrected (VEC) processes in modelling and forecasting. The VEC examines the long run relationships among the products under consideration. But this was not examined in the previous studies. The present study attempts to explore other methodologies in characterizing the prices of these products.

Purpose of the Study

The objective of the study is to obtain a multivariate model for ex-pump prices of petroleum products. The following specific objectives guide the study. To:

i. Examine the trends of ex-pump prices of petroleum products.
ii. Examine how changes in the prices of one product cause changes in the prices of others.
iii. Obtain a model for the prices of the selected petroleum products.

Research Questions

The following research questions guided the study;

i. What are the trends of ex-pump prices of petroleum products in Ghana?
ii. How do changes in the prices of one product cause changes in the prices of others?
iii. What time series models may best describe the prices of petroleum products in Ghana?

Significance of the Study

The change in terms of trade for importers of oil reduces income, real consumption leads to a decline in the balance of trade and puts a downhill burden on exchange rates. Economic growth decelerates, higher costs, causing inflation to upsurge, and unemployment upshots. Forecasting petroleum products price is necessary for preparing the needed refining capacity to meet national consumption. It is an essential instrument for policymakers as they specify the extent of price increases necessary to limit future damages in almost all the economic sectors. By using new methods for the estimation of prices of these products, the study adds to the literature. The study will also provide models for forecasting price levels of the 4 products for planning purposes. This is to aid powers that be to manage and control the prices of the products to considerable levels. Additionally, this study aside from serving as a basis for future study in the area of predicting prices of these products will also arouse the attention of other academics to explore more on this new method of estimating prices as the method has proven much in the world of economics.

Delimitations of the Study

The study bordered on ex-pump prices of petroleum products in Ghana. There are many of these products being consumed in Ghana but the study is restricted to only four of these products due to the availability of data. The study is also investigating tends, relationships, and best models that may best describe the future prices of these products.

Limitations of the Study

Since the study is limited to these prices, its key outcomes cannot be generalized to all other countries. Moreover, the use of secondary data sometimes gives misrepresentative results, if the data-generating mechanism has challenges.

Organization of the Study

The study has five chapters. The first covered introduction and highlights the background, statement of the problem, purpose, research questions, and significance, as well as delimitations, and limitations. The second dealt with the review of related literature and this focused on methods adopted by previous investigators. The third discussed the mathematical and statistical methods and procedures used analysis of data. The fourth chapter also doled out the analysis of data. The discussions and summary of key findings were also presented. The final chapter covered the summary, conclusions, recommendations, as well as suggestions for further studies, references, and appendices.

Chapter Summary

In this chapter, the relevance of exploring a model for petroleum products in Ghana has been examined. It has highlighted the widespread use of petroleum products among energy consumption in the country and hence the effect on the economy of increases in these products. This intensive use of these products makes it important to explore various ways to examine its effect on the economy. The objectives of the study are therefore appropriately derived. In particular, the study seeks to examine the changes that are caused in other products' prices as a result of changes in a particular product's price. It has also been identified that several studies have already been carried out on petroleum products in Ghana. Some of these studies have made use of the demand for the products rather than prices. The related studies that are based on price volatility have made use of various multivariate time series techniques such as the GARCH models. Data on petroleum products in Ghana appears to generate diverse interests for research. The introductory chapter, therefore, shows the relevance of the current attempt to explore further on the suitable characterization of these petroleum products.

CHAPTER TWO

LITERATURE REVIEW

Introduction

The purpose of the study is to model the prices of petroleum products to predict the prices of these products. The chapter looks at studies by other scholars on time series forecasting techniques. The chapter is structured as follows; oil price instability and economic development, a theory of crude oil, causality from oil price to economic growth, and a brief review of time series methods.

Instability of Oil Prices and Development of the Economy

Characteristically, the concepts of growth have mainly focused on capital, labour, and land as the main inputs, failing to identify the cardinal importance of primary energy like crude oil. Nonetheless, there have been attempts to incorporate the role of oil price instability in economic growth, hence bridging the gap between the instability of oil prices and economic growth. Proponents of the symmetric link between economic growth and the price of oil such as (Laser, 1987) hypothesized that instability in the growth of Gross National Product (GNP) is caused by the volatility of oil prices. They based their theories on how the economies of countries that import and export oil reacted to occurrences in the market in 1948 and 1972. Laser (1987) later, confirmed the symmetric correlation between economic growth and volatility in oil prices. Laser (1987) opposed that increases in the price of oil lead to a fall in GDP while a reduction of oil price levels leads to an unambiguous result on GDP depending on the country (oil-exporting or importing). Another school of thought, the asymmetry-in-effect concept of growth of an economy, explains the asymmetric link between economic growth and oil price unpredictability by focusing on three potential ways: sectorial shock, counter-inflationary monetary policy, and vagueness. Proponents of this theory reveal that oil price rises and counter-inflationary responses are significantly related. Balke (1996) agrees with this submission and portends that oil price volatility effects on GDP are not sufficiently explained by monetary policy alone. The renaissance growth theory was whittled from the symmetric and asymmetric School of thought.

Lee (1998), a proponent of the renaissance growth theory, centered her theoretical work on differentiating between variations and instability in oil prices. She (1998) contends that oil price instability and oil price variation, influence the growth of an economy negatively and in diverse ways, influence oil price unpredictability on economic growth is immediate, meanwhile effects of the oil price change on economic growth are sensed a year after. Lee (1998) ended by affirming that "it is volatility in crude oil prices that has a significant effect on economic growth and not oil price levels". The analysis of Hamilton and Herrera (2004) for instance is congruent with that of other authors who show that counter-inflationary monetary policy was only partly accountable for the real effects of oil price shocks that hit the US during the past thirty years. Brown and Yucel (1999) for example, constructed a Vector Autoregressive (VAR) model of the US economy similar to the BGW model and established that after an oil price shock, the economy responds with a decrease in GDP, a rise in interest rates, and price level. Since the drop in GDP and the rise in deflator are similar in magnitude, so that nominal GDP remains comparatively constant, the finding that supports Robert Gordon's definition of monetary neutrality - the Federal Reserve seems to have been neutral to oil price increases. Since they perceived that if the federal funds rate is held constant after an oil price increase GDP. The price level and nominal GDP increase. They argued that the US monetary policy has perhaps had no role in worsening the effects of past oil price shocks. Following the real oil price shock of 1973-1974, which led to inflation and recession in the US and other industrialized countries, Darby (1982) contended that imported oil influences the cumulative production function.

A rise in oil prices leads to an opposing shift in the aggregate supply curve, causing an increase in aggregate prices and a reduction in output. With an increase in inflation, the local interest rate is likely to rise to cushion the effect of inflation. There is likely to be an inflow of foreign capital in response to a rise in the domestic interest rate, accounting for an appreciation of the domestic currency. Hamilton (1983) indicated a negative association between the growth of the United States (US) economy and crude oil prices. Later, Hamilton (1996) made known a non-linear transformation model and Granger causality. Their results indicated a negative connection between the price of oil and economic growth in New Zealand. In a similar study, Jin (2008) revealed that hikes in oil prices cause a decline in the economic growth of Japan and China, however, an increase in oil prices leads to the expansion of the Russian economy. Specifically, he concluded that a 10% rise in the price of crude oil leads to a 5.16% growth in the GDP of Japan. In related work, Glasure and Lee (1997) studied the interconnection between GDP and energy consumption for South Korea and Singapore employing the Granger causality test, together with cointegration and error corrected modelling. Their study discovered causality in both directions between income and energy for both countries. Contrary, the work showed no connection between energy consumption and GDP for South Korea. It also revealed a one-directional connection in the case of Singapore from energy consumption to GDP.

Chang and Wong (2003) concentrated on Singapore and examined how shocks in oil prices influence the economy. They established an insignificant negative relationship between oil price variations, inflation, GDP, and rate of unemployment. Contrastingly, a similar work by Farzanegan and Markwardt (2009) focusing on Iran showed a positive strong relationship between shocks in oil price and output. Another study by Rafiq, Salim, and Bloch (2009) in Thailand showed that instability in oil prices negatively affects investment in the short-run, but has an effect on unemployment in the long run. In a related study, Sadorsky (1999) strongly held that changes in prices of oil have a great influence on the economy. This was supported by Papapetrou (2001) when they assessed it from the view of Greece and some other European countries.

Jumah and Pastuszyn (2007) for the time spanning 1965 to 2004 assessed the linkage between shocks in oil price and monetary policy in Ghana for the period 1963 to 2004. The study examined the correlation between prices of crude oil and the global market and aggregate demand in Ghana by the use of cointegration analysis, through the interest rate channel. They established that crude oil prices on the global market have direct effects on general price levels which also affect output. The results additionally showed that the initial response in monetary policy is stagnant when oil prices go up but with high rates of inflation. The resulting increase in inflation causes an additional reduction in monetary policy. In essence, the output does not return that fast to where it was originally after a shock in oil price but then falls over a while. Similarly, Tweneboah and Adam (2008) assessed both the long and short-run relations between global oil prices and monetary policy between 1970 and 2006 in Ghana. Their findings showed the presence of long- and short-run linkages among global oil prices, price levels (domestic), exchange and interest rates, and GDP. They further showed that oil price shocks manifest in Ghana through a rise in the rate of inflation and a drop in output.

Akide (2007) also examined oil price volatility effects on economic growth in Nigeria and found that shocks in oil prices did not affect output. In supporting this neutrality hypothesis, Mulegeta, et al (2010) applied the panel cointegration method to Sub-Saharan African countries. The results of the study buttressed the neutrality hypothesis in the short run. Moreover, Aqeel and Butt (2001) between 1955 and 1956 applied the cointegration technique and Granger causality test and found that economic growth is the foundation of the consumption of petroleum in Pakistan, but gas consumption substantiates the neutrality hypothesis.

Theoretical Framework of Crude oil and Economic Growth

The theories on crude oil and economic growth are generally assembled into three groups. The first theory suggests that the consumption of energy is a requirement for economic growth, such that energy is a direct say in the production and complements labor and capital (Ebohon, 1996). This theory simply means that increasing the consumption of crude oil, activates the progress of an economy. This implies that energy conservation policies have damaging repercussions on economic growth. The second theory assumes a feedback link between the consumption of crude oil and economic development. The theory suggests the existence of a bidirectional linkage between the consumption of oil and economic development. This means an increase in the amount of crude oil consumed induces growth in the economy, likewise, growth in the economy also leads to growth in oil consumption. The third theory suggests a neutral connection between growth in the economy and the quantity of oil consumed. As such, it suggests that strategies introduced to preserve the use of energy will have no substantial consequence on the economy. The focus of this study is on Vector Autoregressive/ Vector Error Corrected (VAR/VEC) models, which can generate forecasts for two or more products at some time. These are discussed in detail in Chapter Three.

Use of Time Series Models for Forecasting Oil Products and Oil Prices

Lon-Mu (2006) investigate the dynamic relations concerning US gasoline, crude oil prices, and the stock of gasoline. By means of monthly data between January 1973 and December 1987, they established that the US gasoline price is mainly subjective to the price of crude oil. The stock of gasoline has little or no effect on the price of gasoline during the period before the second energy crisis and seems to have some effect during the period after. They also discover that the dynamic connection between the prices of gasoline and crude oil changes over time, shifting from a longer to a shorter lag response. Box-Jenkins Autoregressive Integrated Moving Average (ARIMA) and transfer function models are employed. These models are estimated using estimation techniques with and without outlier adjustment.

Liu et al. (1991) researched the consumption of natural gas in Taiwan and explore the dynamic associations among several potentially relevant time series variables to develop suitable models for forecasting. The temperature of service areas and the price of natural gas were significant factors in forecasting the residential consumption of natural gas. Because of the government price regulation policy, nonetheless, they found that the price variable employed in modelling and forecasting of natural gas consumption needs to be used judiciously. Otherwise, unsuitable models and poor forecasts may occur. They also studied the inclusion of the price variable using an intervention model and 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 were 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 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, based on 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 ANN is due to its 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).

A study by Volkan et al. (2006) aims at forecasting the most possible curve for domestic fossil fuel production of Turkey to help policymakers to develop policy implications for the 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 regression, ARIMA, and Seasonal Autoregressive Integrated Moving Average (SARIMA) methods 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 and 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 fossil fuel consumption and production is growing enormously and it reaches in 2030 to approximately twice of what it is in 2000.

A study by Kumar and Jain (2009) applies three-time series models, namely, the Grey-Markov model, Grey-Model with rolling mechanism, and singular spectrum analysis (SSA) to forecast the consumption of conventional energy in India. The Grey-Markov model has been employed to forecast crude-petroleum consumption while Grey-Model with a 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 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 is 97.9%, and 95.4% while for electricity consumption the prediction accuracy is 96.9% and 95.1%. Similarly, the prediction accuracy for crude-petroleum consumption is found to be 99.2% and 97.6% while for natural gas consumption values were 98.6% and 94.5%. The results obtained have also been compared with those of the Planning Commission of India's projection. The comparison points to the enormous potential that these time series models possess in energy consumption forecasting and can be considered as a viable alternative (Kumar & Jain, 2009).

Uri and Flanagan (2003) detailed 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 were made. The results indicate that natural gas production in 1978 increased by 2·8 percent over the 1977 level while crude petroleum production is likely or bound to fall by 4%.

Chapter Summary

The chapter considered various works done by other research and using time series techniques and other forecasting techniques. The chapter has reviewed the literature on oil price instability and economic development, theories on crude oil, causality from oil price to economic growth as well as a brief review of time series models applied to modelling petroleum prices and related statistical data. The literature is clear on the enormous impact of oil prices on economies around the world. There is thus a clear connection between oil prices and Gross Domestics Product/Gross National Product (GDP/GNP) of economies. Theories on crude oil and the growth of an economy were generally grouped into three categories. Multivariate time series models, like Autoregressive Conditional Heteroscedascity (ARCH), Generalized Autoregressive Conditional Heteroscedascity (GARCH), Vector Autoregressive Moving Average (VARMA), Vector Autoregressive Integrated Moving Average (VARIMA), and Error Correction Model (ECM) are used extensively in the literature on related studies.

CHAPTER THREE

RESEARCH METHODS

Introduction

The chapter is organized as follows: fundamentals of time series, components of time series, unit root test, cointegration test, as well as empirical strategy, model diagnostics. The methodology of this study follows a quantitative method. Furthermore, tests of unit root were applied. These tests include; Levin, Lin, and Chu t*, and Breitung t-stat; Im, Pesaran and Shin W-stat, ADF-Fisher Chi-square and PP-Fisher Chi-square. Others include; Augmented Dicky Fuller (ADF), Kwiatkowski, Phillips, Schmidt, and Shin (KPSS), and Phillips-Perron tests. Levin, Lin, and Chu t*, and Breitung t-stat; Im, Pesaran and Shin W-stat, ADF-Fisher Chi-square and PP-Fisher Chi-square are multivariate tests and ADF, and KPSS are univariate unit root tests. The univariate unit root tests are used as a confirmatory test to the multivariate unit root tests. Moreover, an appropriate lag length is selected by a lag length selection criterion. Therefore, Vector Autoregressive/Vector Error Corrected (VAR/VEC) model descriptions follow the procedure. Consequently, VAR/VEC models consist of the Granger causality test, Impulse Response Functions (IRFs), and Forecast Error Variance Decomposition (FEVD).

Data Description

Data on bimonthly ex-pump prices of four petroleum products, namely; Premium Gasoline, Gas Oil, Kerosene, and Liquefied Petroleum Gas, were obtained from the National Petroleum Authority (NPA). The data ranges from January 2007 to June 2015 and is multivariate in nature. There are many petroleum products consumed in Ghana. However, the choice of these 4 products for the study is based on the availability of complete data set for the selected start and end years on them. That is, complete and accurate data was available on these 4 products from 2007 to 2015, even though the data obtained spans 1989 to 2015.

Time Series

The gathering of observations on quantitative variables made in an unvarying set of a period, usually daily, weekly, monthly, among others are known as time series (TS). For instance, the quarterly Gross National Product (GNP) of a country for 20 years, etc, is TS. Time Series Analysis (TSA) encompasses procedures that break down a series into constituents and understandable portions that allow the identification, estimation, and forecasting of trends to be made. Fundamentally, TSA endeavors to know the underlying environment of data with the aid of a model by relating estimated values based on identified historical ones. Some TSA models include; ARIMA, VAR, VEC, and so on. However, the study focuses on VAR and VEC models to establish a multivariate relationship among ex-pump prices of 4 petroleum products.

Assumptions Underlying Time Series Analysis (TSA)

Normality

Before a model is established, the normality of the data is necessary for TSA. A normalized plot of residuals is normally examined before the assessment of an intervention. Transformation or standardization is normally applied if data is not normal. Jarque-Bara (JB) test of normality which considers the skewness and kurtosis of the data are normally used to ensure that the data is normalized.

Homogeneity of Variance

Before modelling, it is assumed that the variance between the data or observations is constant. Again, there is a contemplation for transforming if the homoscedasticity assumption is violated. McCleary, Hay, Meidinger, and McDowall (1980) recommend a logarithmic conversion as a way to solve the heterogeneity of variance problem.

Absence of Outliers

Outliers are data points that are extremely unreliable with the rest of the TS data. They can significantly disturb the results of an analysis and should necessarily be taken care of. There are no physical rules to regulate how different a case must be to be considered so in a TSA (Cryer, 1986). To deal with it, one checks the initial data for errors, delete data points, changing data points with imputed values, among others.

Unit Root

A unit root (UR) occurs when either Autoregressive, $\text{AR}(\text{p})$ or Moving Average, $\text{MA}(\text{q})$ polynomial of an Autoregressive Moving Average, $\text{ARMA}(\text{p,q})$ model has a root on or near the unit circle. Nevertheless, Vector Autoregressive/ Vector Error Corrected (VAR/VEC) processes are multivariate forms of TSA. So, precisely, an $\text{AR}(\text{p})$ TS $\{y_t\}$ is described as stationary if the roots of the polynomial,

$$m^p - \phi_1 m^{p-1} - \phi_2 m^{p-2} - \ldots - \phi_p = 0 \qquad (1)$$

are not equal to 1 (absolute terms). However, if they are equal to or almost 1, the TS is not stationary and is said to contain a unit root. Besides, a VAR process has a UR if its AR process is not stationary. A TS with a unit root is made stationary by transforming. On the contrary, an MA process with UR implies over differencing.

Unit Root Test (URT)

A significant part of TSA is making sure the series is stationary. A stationary TS is one with time-invariant statistical properties. Thus, the mean and covariance of the TS does not depend on time for any lag *k*, that is,

$$\gamma_y(k) = \operatorname{cov}(y_t \, , \, \mathrm{y}_{t+k}) \tag{2}$$

Several procedures are available for checking the stationarity of a TS (graphical and quantitative). The graphical method includes looking at the Autocorrelation Function (ACF) plots. A strong and slowly decaying ACF indicates deviance from stationarity. For this study, in addition to the ACF, univariate and multivariate techniques are used for checking the presence of UR. These are; (ADF; KPSS) URTs for the univariate case (Im, Pesaran & Shin, 2003; Levin, Lin & Chu, 2002) and URTs for the multivariate situation.

Univariate Unit Root (UUR) Tests

There are numerous UUR tests. The ones used in this study are ADF, PP, and KPSS tests as mentioned earlier. These tests are mostly used because of their ability to detect or identify stationarity or otherwise of a TS data set, as discussed below.

ADF Test

Dickey and Fuller (1979) suggested the ADF test. It was an improvement in the earlier test proposed by the same people. It is grounded on the hypothesis that the series follows a random walk. Think through the AR (1) below,

$$Y_t = \phi Y_{t-1} + \varepsilon_t \tag{3}$$

Where ε_t represents a successively uncorrelated white noise order with zero mean and constant variance. If "$\phi = 1$", (3) turns out to be a random walk model without drift -"non-stationary process". The rudimentary idea of the ADF test is merely regressing "Y_t" on its lagged value "Y_{t-1}" and determining if the estimated "ϕ" is equal to one or not in statistical terms. Equation (3) can be worked by deducting "Y_{t-1}" from both ends to get

$$\Delta Y_t = \delta Y_{t-1} + \varepsilon_t \tag{4}$$

Where,

$$\delta = \phi - 1 \text{ and } \Delta Y_t = Y_t - Y_{t-1} \tag{5}$$

Essentially, as an alternative of estimating (3), we reasonably approximate (4) and test for,

$$H_0 : \delta = 0$$

against (6)

$$H_1 : \delta \neq 0$$

Erdogdu (2007) stated that if $\delta = 0$, then, $\phi = 1$, which indicates the series non-stationary. Under the null hypothesis, the value of the estimated coefficient of "Y_{t-1}" does not have an asymptotic normal distribution. The choice of rejecting or failing to reject the null hypothesis is grounded on the DF critical values of the t-statistic. The DF test assumes the error terms are unrelated. The regression equation is presented as:

$$Y_t = \delta Y_{t-1} + \sum_{i=1}^{p} \gamma_i Y_{t-i} + \varepsilon_t \tag{7}$$

Where, $\sum_{i=1}^{p} \gamma_i Y_{t-i} + \varepsilon_t$ is the addition of the lagged values of the DV Y_t and p, the order of the process. The parameter of interest is δ. For $\delta = 0$, the series contains UR. The choice of preliminary augmentation order depends on data periodicity, the significance of γ_i approximations, and white noise. After the initial approximation, the non-significant parameter augmentation can be dropped to have more efficiency. The following hypotheses are tested;

$H_0 : \gamma = 0$ (The process is not stationary)

against (8)

$H_1 : \gamma < 0$ (The process is stationary)

The test statistic is stated as;

$$DF_\tau = \frac{\hat{\delta}}{SE\left(\hat{\delta}\right)} \tag{9}$$

Where $SE\left(\hat{\delta}\right)$ is the standard error of the least square estimate of $\hat{\delta}$. The null hypothesis is that "the series is not stationary" if the test statistic is less than the critical value, we will fail to reject the null hypothesis that the process is not stationary.

KPSS Test

A corresponding test for examining the order of integration of a series, "Y_t" is to test the null hypothesis that the data production process is stationary $\left[H_0 : Y_t \sim (0)\right]$ against the alternative, using the KPSS procedure. Kwiatkowski et al. (1992) derived a test for this pair of hypotheses. This test assumes that if there is no linear trend term, the point of departure is a data-producing process of the form,

$$Y_t = X_t + \varepsilon_t \tag{10}$$

Where, "Y_t" is a random walk, $X_t = X_{t\text{-}1} + vt,\ vt \sim iid\left(0, \sigma_v^2\right)$ and ε_t is a white noise order. In this situation, the preceding pair of hypotheses are equal to the pair,

$$H_0 : \sigma_u^2 = 0 \ \text{(The process is stationary)}$$

against (11)

$$H_1 : \sigma_u^2 > 0 \ \text{(The process is not stationary)}$$

If the null hypothesis is true, then "Y_t" is composed of a constant and a stationary process "ε_t". Thus, "Y_t" is also stationary. They suggested the following test statistic,

$$\text{KPSS} = \frac{\text{I}}{\text{T}^2} \sum_{t=1}^{T} \frac{\text{S}_t^2}{\hat{\sigma}_{¥}^2} \tag{12}$$

Where, "T" is the number of observations, $S_t = \sum_{j=1}^{t} \hat{\omega}_j$ with $\hat{\omega}_j = Y_t - \overline{Y}$ and σ_∞^2 an estimator of:

$$\sigma_\infty^2 = \lim_{T \to \infty} \text{T}^{-1} \text{Var}\left(\sum_{t=1}^{T} \varepsilon_t\right) \tag{13}$$

That is, σ_∞^2 is an estimator of the long-run variance of the process "ε_t". If "Y_t" is a stationary process, then "S_t" is an [I (1)] and the quantity in the denominator of the KPSS statistic estimates its variance, with a stochastic limit. The term in the denominator warrants that generally, the limiting distribution is free of unknown nuisance parameters. If "Y_t" is [I (1)], the numerator will grow boundless, causing the statistic to become large for larger samples. The null hypothesis is rejected for large values of KPSS. In other words, the decision rule is to reject the null hypothesis of stationarity if the computed value of the test statistic is greater than the critical value at a given level of significance.

PP Test

The PP test was developed by Phillips (1987) and Phillips and Perron (1988). The PPtests are based on the following ADF regression, and the critical values are the same as those used for the ADFtests:

$$Y_t = \delta Y_{t-1} + \sum_{i=1}^{p} \gamma_i Y_{t-1} + \varepsilon_t \quad (14)$$

Where "ε_t" is a random error.

The PPUR Test is used in preference to the AD F-test for the simple reason that: the test does not need the assumption of homoscedasticity of the error term (Phillips, 1987). Secondly, there is no loss of effective observations from the series since the lagged terms for the variable of interest are set to zero (Perron, 1988). This is especially valuable if data points are limited in number. The PPUR Test corrects the serial correlation and autoregressive heteroscedasticity of error terms. This provides UR Test results that are robust to serial correlation and time-dependent heteroscedasticity of errors. In both the PP and ADF, the null hypothesis is that "the series is non-stationary" and we fail to reject or reject by inspection of the t-ratio of the lagged term "Y_{t-1}" compared with a table value. We test:

$$H_0 : \gamma = 0 \ (\text{The process is not stationary})$$

against (15)

$$H_1 : \gamma < 0 \ (\text{The process is stationary})$$

If the t-ratio is greater than the table value, the null hypothesis is rejected and the series is deemed integrated of order zero, hence, considered to be stationary at levels. However, these univariate unit root tests were only used as confirmatory to the multivariate UR Tests.

Multivariate (URR) Tests

The econometric concept for dealing with panel data was fundamentally established for data sets where the number of TS observations (*T*) is small but the number of groups (*N*) is large. Here, the asymptotic statistical theory was derived by letting "*N*→" ∞, for fixed *T* in TSA which was done by letting "*T*→∞", for fixed *N*. One key feature of this data set is that sometimes both *T* and *N* are large and their extent of largeness is similar. This feature has varied consequences for theoretical and empirical analysis, and their understanding is very important for economists working on this kind of data. If we consider the basic model:

$$\Delta y_{it} = \rho y_{it-1} + u_{it}, \; i=1, 2, 3, ..., N \quad t=1,2,...,T \qquad (16)$$

in the case of a single equation, we are interested in testing $\rho_1 = 0$ against the alternative hypothesis $\rho_1 < 0$ and we apply a UR for the first TS. In the panel data case, the hypothesis we are interested in is:

$$H_0: \rho_i = \rho = 0$$

against (17)

$$H_1: \rho_i = \rho < 0$$

for $i=1, 2, 3, ..., N$

Some of these MUR tests are Levin, Lin, and Chu's (LLC), and Im, Pesaran, and Shin's (IPS) MUR tests.

LLC MUR Tests

Levin and Lin (1992, 1993) and Levin, Lin, and Chu (2002) [LLC] provided results on panel UR tests. They generalize Quah's model to allow for heterogeneity of individual deterministic effects and heterogeneous serial correlation feature of the error terms assuming homogeneous first-order autoregressive parameters. They assume that both *N* and *T* tend to boundlessness but *T* upsurges at a faster rate, such that "$N/T \to 0$". They came out with a process using pooled *t*-statistic of the estimator to measure the hypothesis that each TS contains UR. Thus, referring to the model in (17), LLC assume same autoregressive coefficients between individual, i.e. $\rho_i = \rho$, for all *i*, and tests;

$$H_0: \rho_i = \rho = 0$$

against (18)

$$H_1: \rho_i = \rho < 0$$

for all *i*.

Putting a cross-equation constraint on the first-order partial autocorrelation coefficients under the null, the method leads to a test of advanced influence than carrying out a distinct UR test for each series. The form of the LLC analysis may be specified as follows:

$$\Delta y_{it} = \rho y_{it-1} + \alpha_{0i} + \alpha_{1i} t + u_{it},\ i=1, 2, 3, ..., T \tag{19}$$

Where time trend $(\alpha_{1i} t)$ and separate effects (α_i) are combined. Note that the deterministic components are important sources of heterogeneity in this model since the coefficient of the lagged DV is constrained to be similar across all units in the panel. (u_{it}) is presumed to be individually dispersed across individuals and follow a stationary invertible ARMA process for each individual:

$$u = \sum_{j=1}^{\infty} \theta_{ij} u_{it-j} + \varepsilon_{it} \tag{20}$$

and the finite moment conditions are assumed to assure the weak convergence in (Phillips, 1987; Phillips & Perron, 1988) UR tests. LLC takes into account several subcases of models that are all projected by OLS as pooled regression models. Limiting distributions are derived by sequential limit theory $(T,N \rightarrow \infty)_{seq}$

LLC display that the asymptotic properties of the regressions estimators and test statistics are a combination of properties derived for stationary panel data, and properties derived in the TS literature on UR tests, in contrast to the non-standard distributions of UR test statistic for single TS (Phillips & Perron, 1988). The panel regression estimators and test statistics have limiting normal distributions, as in the case of stationary panel data (Hsiao, 2003). Nevertheless, the existence of a UR grounds the convergence degree of the estimators and t-statistics is higher when "$T\rightarrow\infty$" than when "$N\rightarrow\infty$". In the case of *i.i.d.* turbulences and no individual-specific fixed effects, under the null hypothesis, the panel regression UR t-statistic t_ρ based on the pooled estimator $\hat{\rho}$ converges to the standard N~(0,1) distribution when N and T lean towards infinity and $\sqrt{N/T} \rightarrow 0$.

Contrary, if there are individual-specific fixed effects in the turbulences, the resultant test statistic is not concentrated at zero, with a significant effect on the size of the UR test. Here, Levin and Lin recommended using an adjusted t-statistic:

$$t_{\rho}^{*} = \frac{t_{\rho=0} - N\tilde{T}\hat{S}_{NT}\hat{\sigma}_{\hat{\varepsilon}}RMSE\left(\hat{\rho}\right)\mu_{mT}^{*}}{\sigma_{mT}^{*}} \tag{21}$$

μ_{mT}^{*} and σ_{mT}^{*} being the mean and standard deviation adjustment terms which are gotten from Monte Carlo simulation (MCS) and tabularized in Levin and Lin (1992),

$$\hat{S}_{NT} = \frac{1}{N}\sum_{i=1}^{N}\frac{\sigma_{yi}}{\sigma_{ei}} \tag{22}$$

where σ_{yi}^{2} represents a Kernel estimator of the long-run variance for separate series. Applying sequential limit theory, $\left(\mathrm{T, N} \to \infty\right)_{\mathrm{seq}}$, the following restrictive distributions of $T\sqrt{N}\hat{\rho}$ and t_{ρ} are gotten:

$$\begin{aligned} N\hat{\rho} &\Rightarrow N\left(0,1\right) \\ t_{\rho} &\Rightarrow N\left(0,1\right) \end{aligned} \tag{23}$$

Levin and Lin (1993) are MCS outcomes showed that when there are not individual-specific fixed effects, the standard normal distribution is likely to provide a worthy estimate of the experiential distribution of the test statistic in relatively small samples and that the panel structure can provide vivid growths in power relative to performing a separate UR test for each TS. As According to Levin et al. (2002), panel-based UR tests are more applicable for panels of reasonable size (i.e., 10 <N <250; 25 <T <250). Existing UR test measures are suitable if T is very large, or T is very small but N is very large.

Nevertheless, for panels of reasonable size, standard multivariate procedures may not be computationally achievable and the LLC test seems to be more suitable. Regrettably, the LLC test has some restrictions. First, the test is crucially contingent on the independence assumption across individuals, and hence not appropriate if the cross-sectional association is present. But the main limitation is that the autoregressive parameters are considered as being identical across the panel:

$$H_0 = \rho_1 = \rho_2 = ... = \rho_N = \rho = 0$$

against (24)

$$H_1 = \rho_1 = \rho_2 = \ldots = \rho_N = \rho < 0$$

The null hypothesis is logical under some conditions, but as Maddala and Wu (1999) stated, the substitute is too robust to be held in any stimulating experimental cases. This constraint has been overcome by IPS (Im, Pesaran & Shin, 2003) which suggested a panel UR test deprived of the supposition of the identical first-order relationship under the alternative.

IPS Tests

IPS (2003), utilizing the likelihood outline, recommended a new more flexible, and computationally friendly UR testing technique for panels that allows for concurrent stationary and non-stationary series (i.e. ρ_i may differ individually). Additionally, the test permits residual serial correlation and heterogeneity of the dynamics and error variances through groups. IPS takes into consideration the mean of (A) DF statistics calculated for each cross-section unit in the panel when the error term "u_{it}" of the model (20) is serially connected, perhaps with dissimilar serial association features across cross-sectional units. That is,

$$\left(u_{it} = \sum_{j=1}^{p_i} \varphi_{ij} u_{it-j} + \varepsilon_{it} \right) \quad (25)$$

and *T* and *N* are adequately large. Replacing "u_{it}" in (20) and bearing in mind a linear trend for each of the *N* cross-section units, we obtain:

$$\Delta y_{it} = \alpha_{0i} + \rho_i y_{it-1} + \sum_{j=1}^{p_i} \varphi_{ij} \Delta y_{it-j} + \varepsilon_{it} \tag{26}$$

where, as usual, *i*=1, 2, ..., *N*, and *t* =1, 2, ..., *T*. The hypotheses are:

$$H_0 = \rho_i = 0 \text{ for all } i$$

against (27)

$$H_1 : \begin{cases} \rho_i < 0, \text{ for i=1,...,N}_1 \\ \rho_i = 0, \text{ for i=N}_1+1,...,\text{N}_2 \end{cases} \quad \text{with } 0<N_1<N$$

that permits (but not all) of separate series to have UR. IPS compute distinct UR for the *N* cross-section units and define their *t*-bar statistic as a simple average of the individual ADF statistics, t_{iT} for the null as:

$$\bar{t} = \frac{i}{N}\sum_{i=1}^{N} t_{iT} \tag{28}$$

IPS assumes that t_{iT} are *i.i.d.* and have finite mean and variance. Consequently, by the Lindeberg-Levy central limit theorem, the standardized *t*-bar statistic converges to standard normal variate as "*N*→∞" under the null hypothesis. To suggest a standardization of the *t*-bar statistic, the mean and variance have been calculated through MC approaches for dissimilar values of "T" and p_i's and tabularized by IPS (2003). It is imperative to note that in this process, only balanced panel data are measured. If unbalanced data are used, more replications have to be done to get critical values. In the case of serial correlation, IPS recommended using the ADF t-test for the distinct series. However, $E\left[t_{iT} \middle| \rho_i = 0\right]$ and $\text{Var}\left[t_{iT} \middle| \rho_i = 0\right]$ will differ as to the lag distance in the ADF regression differs. They tabulated $E\left[t_{iT} \left| \rho_i = 0\right.\right]$ and $\mathrm{Var}\left[t_{iT} \left| \rho_i = 0\right.\right]$ for diverse lag distances. In using their tables, it is essential to limit all the ADF regressions for separate series having the same lag length. IPS' replications indicate that, if the serial correlation is absent, the t-bar test has the precise size and is very influential, even for small values of "T" (T =10): its influence increases monotonically with "N" and "T". Replications display the significance of the right choice of the fundamental ADF regressions order, especially when the panel comprises deterministic time trends. When the disturbances in the dynamic panel are serially connected, the size and influence of the t-bar test are rationally acceptable, but "T" and "N" have to be adequately large. In this case, it is critically vital not to underestimate the order of the core ADF regressions.

An additional significant feature lies in the fact that the influence of the t-bar test is considerably more favorably impacted by an increase in "T" than by a corresponding upsurge in "N". Exceptional care ought to be taken when interpreting the outcomes of this panel UR test. Owing to the varied nature of the alternative hypothesis, rejection of the null hypothesis does not automatically suggest the unit root null is rejected for all *i*, but only that the null hypothesis is rejected for "$N_1 < N$" elements of the group such that as "$N \to \infty,\ \mathrm{N}_1 \left| N \to \delta > 0\right.$". The test does not offer directions as to the extent of δ, or the uniqueness of the particular panel element in which the null hypothesis is rejected.

Cointegration

For an extended time, it was a shared practice to approximate equations concerning non-stationary variables in macroeconomic models by conventional linear regression. It was not well known that testing hypotheses about coefficients using standard statistical inference might lead totally to spurious results. Granger and Newbold (1974) mentioned in their paper that such regression tests may often put forward a statistically significant association among variables where none exists. They arrived at their decision by generating independent non-stationary series, more exactly, random walks. The series is regressed on each other and observation of the value of the t-statistic of a coefficient estimate is estimated under the assumption that the true value of the coefficient equals zero. Even though the variables in the regression were independent, they found that the null hypothesis of a zero coefficient was rejected much more often than standard theory foretells. At the same time, they observed that the residuals of the projected equation showed very strong positive autocorrelation. These results showed that many of the seemingly noteworthy associations among non-stationary economic variables in current econometric models may well be spurious.

Cointegration in a vector TS has many consequences for workers in experiential macroeconomics. One supposed reward of identifying the cointegration rank in an integrated vector process is that it results in improved forecast performance. Engle and Yoo (1987) demonstrated that predictions taken from cointegrated systems are "tied together" because the cointegrating relations must "hold exactly in the long-run." They demonstrated in a series of MC experimentations that including cointegration into the estimating model can reduce mean squared prediction mistakes by up to 40% at medium to long prediction distances.

Cointegration Model Representations

This is used to analyze the combined movement of economic variables and deviations from equilibrium over time. It as well expresses the association between two non-stationary series for which the stochastic relationships are bounded. It establishes a connection between two non-stationary series by finding a linear grouping which gives integration of order zero (0) stationary and it helps specify the Error Correction Mechanism (ECM). For instance, consider two economic series, revenue (RVN) and inflation rate (INF):

$$\begin{aligned} RVN_{it} &= U_{it} + \Sigma_{it} \\ INF_{it} &= U_{zti} + \Sigma_{it} \end{aligned} \tag{29}$$

Where, RVN and INF are non-stationary, because they have stochastic trends, "U_{ti}" and "U_{zti}". "U_{ti}" represents the trend in the variables at period t and Σ_{it} represents stationary components in the variables a period t. EXG and INF are co-integrated of order $(1,1)$ if there are non-zero values of β_1, and β_2 in the linear combination,

$$\begin{aligned} &\beta_1 RVN_{it} + \beta_2 INF_{it} \\ &= \beta_1 \left(U_{it} + \Sigma_{it}\right) + \beta_2 \left(U_{zti} + \Sigma_{it}\right) \\ &= \beta_1 U_{it} + \beta_1 \Sigma_{it} + \beta_2 U_{zti} + \Sigma_{it} \end{aligned} \tag{30}$$

The essential and satisfactory condition for RVN and INF to be $CI(1,1)$ is that,

$$\beta_1 U_{it} + \beta_2 U_{zti} = 0 \tag{31}$$

vanishes where $CI(1,1)$ is a cointegrated vector. Suppose we have the series;

$$\begin{aligned} Y_t &= N_{yt} + E_{yt} \\ Z_t &= N_{zt} + E_{zt} \end{aligned} \tag{32}$$

Where, "Y_t" and "Z_t" are non-stationary, "N_{1t}" and "N_{2t}", "E_{yt}" and "E_{zt}" are error terms. A linear grouping of "Y_t" and "Z_t" is stationary if there are non-zero values of β_1 and β_2 in

$$\begin{aligned} \beta_1 Y_t + \beta_2 Z_t &= \beta_1 \left(N_{yt} + E_{yt}\right) + \beta_2 \left(N_{zt} + E_{zt}\right) \\ &= \beta_1 N_{yt} + \beta_1 E_{yt} + \beta_2 N_{zt} + \beta_2 E_{zt} \end{aligned} \tag{33}$$

For stationarity of the LHS, the term,

$$\beta_1 N_{yt} + \beta_2 N_{zt}$$

must vanish. Consequently, the essential and appropriate condition for "Y_t" and "Z_t" to be cointegrated of order (1), $CI(1,1)$ is;

$$\beta_1 N_{yt} + \beta_2 N_{zt} = 0 \tag{34}$$

Then, $CI(1,1)$ co-integrating vector, Y_t and Z_t must have the same stochastic trend since we preclude β_1 and β_2 from being equal to zero. That is, if:

$$Y_t = N_{yt} + E_{yt} \quad \text{and} \quad Z_t = N_{zt} + E_{zt} \tag{35}$$

then,

$$Y_t - Z_t = E_{yt} - E_{zt} \tag{36}$$

In summing up, the cointegration of two or more macroeconomic variables suggests there is a long run or equilibrium relationship among them.

Lag Length Selection Criteria

A VAR of order p is written as $\mathrm{VAR}(\mathrm{p})$, where p is the lag length. Therefore, a critical element in the description of a VAR model is determining its lag length. The lag length (p) must be stated long enough for the residuals not to be correlated serially. So many lag length selection criteria are defined by many authors like FPE, AIC suggested by Akaike (1974), and Hannan-Quinn Information Criterion (1979) just to mention a few. Ivanov and Kilian (2005) suggested the Akaike Information Criteria (AIC) for selecting lag length when data is monthly. But the lag length $\mathrm{VAR}(\mathrm{p})$ may be determined using any of the above-mentioned selection criteria. The broad methodology to VAR model estimation is fitting the VAR (p) with the order p = 0,..., $\mathrm{P}_{\max}$ and the value of p reduces some model selection criteria. Model selection criteria for $\mathrm{VAR}(\mathrm{p})$ models is:

$$\mathrm{IC}(\mathrm{p}) = \ln\left|\bar{\Sigma}(p)\right| + c_T\varphi(n, p) \qquad (37)$$

where,

$$\bar{\Sigma}(p) = T^{-1}\sum_{i=1}^{T}\hat{\varepsilon}_t\hat{\varepsilon}_t' \qquad (38)$$

is the residual covariance matrix without a df correction from the $\mathrm{VAR}(\mathrm{p})$ model, C_T is an order indexed by the sample size "T", and "$\varphi(n, p)$" is a penalty function which penalizes large $\mathrm{VAR}(\mathrm{p})$ models. The lag order is chosen by optimally balancing the term $\ln\left|\bar{\Sigma}(p)\right|$ which is a non-increasing function of order n and $\varphi(n, p)$ which increases with n. The three most common information criteria are AIC, Schwarz-Bayesian (BIC) and Hannan-Quinn (HQ):

$$AIC(p) = \ln\left|\sum(p)\right| + \frac{2}{T}pn^2 \qquad (39)$$

$$BIC(p) = \ln\left|\bar{\Sigma}(p)\right| + \frac{\ln[T]}{T} pn^2 \tag{40}$$

$$HQ(p) = \ln\left|\sum(p)\right| + \frac{2\ln[(\ln T)]}{T} pn^2 \tag{41}$$

Vector Autoregressive/Vector Error Corrected (VAR/VEC) Models

The VAR is one of the effective, flexible, and easy to apply models for the analysis of multivariate TS. VAR models in economics were made widespread by Sims (1980). It is an extension of the univariate autoregressive model. The VAR model is valuable for examining the dynamic behaviour of financial TS and forecasting. The application of VAR/VEC to examining dynamic associations among financial indicators has become shared in the literature (Barnhill et al., 2000). The attractiveness of these models has been associated with the understanding that interactions among financial indicators are so multifaceted that traditional TS models have been unsuccessful in completely capturing. Engle and Granger (1987) noted that for cointegrated systems, the VAR in first differences will be miss-specified and the VAR in levels will disregard significant restrictions on the coefficient matrices. Though these restrictions may be fulfilled asymptotically, effective gains and improvement in forecasts are probable to result from their nuisance. Hence, Engle and Granger (1987) suggested that if a TS system includes integrated variables of order 1 and fulfills the conditions of cointegration relationships, such a system is more appropriately specified as a VEC model rather than a VAR model. Consequently, Johansen (1991) and Ahn and Reinsel (1990) suggested various procedures for estimating cointegrating vectors in full-order VEC model, which contain all non-zero entries in the coefficient matrices.

VAR Representation

In its elementary form, VAR consists of a set of K endogenous variables,

$$y_t = \left(y_{1t}, \ldots, y_{kt}, \ldots, y_{Kt}\right), \text{ for } \text{k}=1,\cdots,\text{K} \tag{42}$$

The $\text{VAR}(\text{p})$ process is defined as;

$$y_t = A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + u_t \tag{43}$$

Where A_i are $(K \times K)$ coefficient matrices for $i = 1, \cdots, p$ and u_t is a K-dimensional process with,

$$E(u_t) = 0 \tag{44}$$

and time-invariant positive definite covariance matrix,

$$E(u_t, u_t') = \Sigma_u \text{, which is white noise} \tag{45}$$

A central distinguishing feature of a $\text{VAR}(\text{p})$ process is its constancy. This means that it produces stationary TS with time-invariant mean, variance, and covariance structure, given adequate initial values. This can be checked by assessing the polynomial:

$$\det\left(I_K - A_1 z - A_2 z^2 - \cdots - A_p z^p\right) \neq 0, \quad \text{for } |z| \leq 1 \tag{46}$$

If the answer of the above equation has a root, z = 1, then either some or all indicators in the $\text{VAR}(\text{p})$ process are integrated of order one. It might be the case that cointegration between the indicators exists. This occurrence can better be examined in the setting of the VEC model. Practically, the stability of an empirical $\text{VAR}(\text{p})$ process can be investigated by considering the companion form and estimating the eigenvalues of the coefficient matrix. For instance, for a $\text{VAR}(\text{p})$ process, where p=1, can be written as:

$$\omega_t = A\omega_{t-1} + v_t \tag{47}$$

Where,

$$\omega_t = \begin{bmatrix} y_t \\ \vdots \\ y_{t-p+1} \end{bmatrix}, \; A = \begin{bmatrix} A_1 & A_2 & \cdots & A_{p-1} & A_p \\ I & 0 & \cdots & 0 & 0 \\ 0 & I & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & I & 0 \end{bmatrix} \text{ and } v_t = \begin{bmatrix} u_t \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{48}$$

and the lengths of the stacked vectors, ω_t and v_t is $(Kp \times 1)$ and that of the matrix A is $(Kp \times Kp)$. If the moduli of the eigenvalues of A are less than one, then the $\text{VAR}(\text{p})$ is stable. For a given sample of the endogenous variables, “ $y_1, \ldots, y_T$: and adequate pre-sample values “ $y_{-p+1}, \ldots, y_0$ ”, the coefficients of a $\text{VAR}(\text{p})$ can be estimated efficiently by applying Ordinary Least Squares (OLS) separately to each of the equations.

Vector Error Corrected (VEC) Model and Regression

Engle and Granger (1987) defined variables that are separately driven by permanent shocks, but for which there are linear combinations that are mean reverting, as cointegrated variables. Johansen (1991) established in the Granger representation theorem that variables, independently driven by lasting shocks, are cointegrated if and only if there exists a VEC representation of the series. This general description has the gain of covering both long-run levels and short-run first differences of non-stationary variables. The cointegration in a vector TS (Engle & Granger, 1987) has many consequences for work in empirical macroeconomics.

Cointegration alters the linear combination of two non-stationary TS into a stationary one. In economics, cointegration is referred to as a long-run equilibrium relationship. The intuition is that non-stationary time series with a long-run equilibrium relationship cannot drift too far apart from the equilibrium because economic forces will act to restore the equilibrium relationship. Consequently, one of the professed rewards of cointegration in an integrated vector process is that it results in improved forecasting performance on a long horizon. In a very powerful and significant paper, Engle and Granger (1987) exhibited that cointegration implies the being of an error correction model (ECM) that defines the dynamic performance of two or more non-stationary series. The ECM links the long-run equilibrium relationship implied by cointegration with the short-run dynamic adjustment instrument that labels how the variables respond when they move out of long-run equilibrium. This ECM makes the concept of cointegration beneficial for modeling macroeconomic TS.

The VEC Model Representation

The VEC Model representation is deduced as follows:

Reconsider the VAR process from (43);

$$y_t = A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + u_t \tag{49}$$

The following VEC specifications exist by Pfaff 2006);

$$\Delta y_t = \alpha\beta' y_{t-p} + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1} \Delta y_{t-p+1} + u_t \tag{50}$$

Where,

$$\Gamma_1 = -\left(I - A_1 - \ldots - A_i\right),\ \text{i=1,…,p-1} \tag{51}$$

and

$$\Pi = \alpha\beta' = -\left(I - A_1 - \ldots - A_p\right) \tag{52}$$

The Γ_i matrices comprise the accumulative long-run influences. Hence, this VEC model specification is indicated by the "long-run" form. The other specification is given as;

$$\Delta y_t = \alpha\beta' y_{t-1} + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1} \Delta y_{t-p+1} + u_t \tag{53}$$

with

$$\Gamma_i = -\left(A_{i+1} + \ldots + A_{p+1}\right),\ \text{i=1,…,p-1} \tag{54}$$

Hence, the Π matrix is similar to the first description. However, Γ_i matrices now vary, in that they measure transitory effects. As a result, this description is shown as a "transitory" form. In a circumstance of cointegration, the matrix

$$\Pi = \alpha\beta' \tag{55}$$

is of reduced rank. The sizes of α and β is $(K \times r)$ and r is the cointegration rank. That is, how many long-run relationships among the variables "y_t" exist. The α is the loading matrix and the coefficients of the long-run relationships are contained in β.

Model Validations

In modelling, the robustness of the final model is key for forecasting. That is, model validation techniques are considered. Model validation and certification is a significant step in modelling. This method mostly helps stakeholders and industry players in minimizing the cost, time, and risk associated with comprehensive product testing. According to Salamanca, Krpo, Martilli, and Clappier (2010). They continued to say that these procedures are necessary practices for measuring and building trustworthiness in statistical models. In doing this, data is normally partitioned into desired proportions with the greater proportion used for modeling and the other percentage for validation purposes. The former is sometimes referred to as training data and the latter is validation data. In this study, the training data (70%) of the data points are used in the modelling and the validation is done with the remaining proportion.

Model Diagnostics

Once a VAR model is estimated, it is of key attention to check whether the residuals conform to the assumptions of the model. That is, one checks for the absence of serial correlation and Heteroscedasticity, and also verifies if the error process is normally distributed.

Heteroscedasticity

The multivariate ARCH-LM was used to test Heteroscedasticity. The test is based on the regression:

$$vech\left(\hat{u}_t\hat{u}_t'\right) = \beta_0 + \beta_1 vech\left(\hat{u}_{t-1}\hat{u}_{t-1}'\right) + \ldots + \beta_q vech\left(\hat{u}_{t-q}\hat{u}_{t-q}'\right) + v_t \quad (56)$$

where "v_t" allocates spherical error processes and "*vech*" serves as the column-stacking operator for symmetric matrices that stack the columns from the leading diagonal downward. The dimension of β_0 is;

$$\frac{1}{2}\left[K\left(K+1\right)\right] \quad (57)$$

The dimension for the coefficient matrices β_i with $i = 1, 2, \ldots, q$ is

$$\frac{1}{2}\left[K\left(K+1\right)\right] \text{ by } \frac{1}{2}\left[K\left(K+1\right)\right] \quad (58)$$

The hypotheses for this test are stated as:

$$H_o : \beta_1 = \beta_2 = ... = \beta_q = 0 \ (There \text{ is no heteroscedasticty})$$

against (59)

$$H_1 : \beta_1 \neq 0 \cap \beta_2 \neq 0 \cap ... \cap = \beta_q \neq 0 \ (\text{There is heteroscedasticty})$$

The test statistic is:

$$VARCH_{LM}(q) = \frac{1}{2} TK(K+1) R_m^2 \tag{60}$$

Where

$$R_m^2 = 1 - \frac{2}{K(K+1)} tr\left(\hat{\Omega}\hat{\Omega}_0^{-1}\right) \tag{61}$$

"$\hat{\Omega}$" denotes the covariance matrix of the above-defined regression model. The test statistic has a Chi-squared distribution of:

$$\chi^2 \left(\frac{qK^2 (K+1)^2}{4} \right) \tag{62}$$

Normality

This shows the frequency distribution of the series in a histogram. The histogram splits the series range into several equal-spaced intervals and displays the number of observations that fall into each interval. Complements of standard descriptive statistics are shown along with the histogram. Also, Jarque-Bera (JB)is a test statistic for testing whether the series is normally distributed or not. Under the null hypothesis, the JB statistic is distributed as χ^2 with 2 df. The hypothesis is stated as:

$$H_o : \beta_1 = \beta_2 = ... = \beta_q = 0 \ (There \text{ is normality})$$

against (63)

$$H_1 : \beta_1 \neq 0 \cap \beta_2 \neq 0 \cap ... \cap = \beta_q \neq 0 \ (\text{There is no normality})$$

The reported likelihood that a JB statistic surpasses the observed value under the null hypothesis. A small probability value hints at the rejection of the null hypothesis of a normal distribution.

Granger Causality

Causality is described as the relationship between cause and effect. Fundamentally, the term "causality" suggests a cause and effect relationship between two variables (Y and X). Conversely, recently, Granger causality modeling has received substantial consideration and use in several areas of study. Since the concept of Granger causality was presented, it has become a popular concept in econometrics and many other fields of human endeavor (Granger, 1969). One variable is said to Granger causes the other if it aids in making a more precise forecast of the other variable than had we only used the past of the latter as a predictor. Given two variables "X_t" and "Y_t", "X_t" Granger cause "Y_t" if "Y_t" can be better predicted using the past values of both "X_t" and "Y_t" than it can do by using the past values of Y_t alone. Several articles surfaced in the literature on the use of Granger causality tests to analyze TS data since its coming to being by Granger (1969). Some of the articles include (Entner &Hoyer, 2010) fairly to mention a few. Hence, it emphasizes the relationship between the variables, whether 0r unidirectional or bidirectional. The null:

$$H_o : \psi_{12} = 0 \ (\text{X does not Granger cause Y})$$

and (64)

$$H_0 : \psi_{21} = 0 \ (\text{Y does not Granger cause X})$$

against the alternative hypotheses

$$H_1 : \psi_{12} \neq 0 \ (\text{X does Granger cause Y})$$

and (65)

$$H_1 : \psi_{21} \neq 0 \ (\text{Y does Granger cause X})$$

Therefore, the null hypothesis is rejected when the p-value is less than 0.05. Once VAR a model has been approximated, the opportunity is wide opened for extra examination. A researcher might be concerned additionally about causal implications, and/or analyzing the observed model's dynamic behaviour. This is presented below.

Impulse Response Functions (IRFs)

In the univariate case, the ACF is enough to know how shocks degenerate. However, when examining vector data, this is no longer the case. A shock to a series has an instant consequence on itself and others in the system which, in turn, can feedback into the original variable. The IRF technique is used to study how one variable responds to a sudden change in other variables. These changes may be referred to as innovations in a variable (Harvey, 1994). This is commonly referred to as impulse. It reflects basically, the idea of a one-time shock happening at some point in time. The IRF of any "y_i" which is an element of "y", concerning a shock in "ε_j", an element of"ε", for any j and i, is the change in "y_{it+s}", for any $s \geq 0$, conforming to a unit shock in "ε_{jt}". The IRF can be demonstrated through a vector moving average (VMA). As long as " y_t " is covariance stationary, it must have a VMA representation of the form,

$$y_t = \mu + \varepsilon_t + \sum_1 \varepsilon_{t-1} + \sum_2 \varepsilon_{t-2} + \dots \tag{66}$$

Applying the VMA, the impulse response " y_t " regarding a shock in " ε_j " is simply $\{1,\ \Sigma_{1[ij]},\ \Sigma_{2[ij]},\dots\}$, if $i = j$ and $\{0,\ \Sigma_{1[ij]},\ \Sigma_{2[ij]},\dots\}$ otherwise. Confidence intervals are normally made to examine whether an impulse response is large in a statistically meaningful sense or not. Because the parameters of the VAR are asymptotically normal, the impulse responses are also asymptotically normal. In this study, the bootstrap method was used in calculating the confidence bands for the impulse response function. The bootstrap method operates based on the principle that "if the residuals are realizations of the actual error process, one can use them directly to simulate this distribution rather than making an arbitrary assumption about the error distribution (e.g. i.i.d. normal)". The procedure:

i. Compute " Φ " from the initial data and compute residuals " ε ".

ii. Using " ε ", compute a new series of residuals " ε_j " through sampling with replacement from the original residuals. The new residuals may be described as; $\{\varepsilon_{u1},\ \varepsilon_{u2},\dots,\varepsilon_{uT}\}$, where " u_i " are i.i.d. separate even random variables taking, 1, 2... T. In principle, the new set of residuals is just the old set reordered with some duplication and omission.

iii. Using " Φ " and " $\{\varepsilon_{u1},\ \varepsilon_{u2},\dots,\varepsilon_{uT}\}$ ", simulate a TS " $\tilde{y}_t$ " with as many data points as the original. These can be computed directly using the VAR;

$$\tilde{y}_t = \hat{\Phi}_0 + \hat{\Phi}_1 y_{t-1} + \dots + \hat{\Phi}_{t-p} + \hat{\varepsilon}_{ut} \tag{67}$$

iv. Using $\tilde{y}_t$ compute estimates of Φ_b from a VAR.

v. Using Φ_b compute the impulse responses $\{\sum\}$, where, $b = 1, 2, 3, \ldots, \mathrm{B}$. These values are saved.

vi. Return to the second step and compute a total of B impulse responses. Normally, B it is between 100 and 1000.

vii. For every impulse response for every horizon, sort the impulse responses. The 5th and 95th percentiles of this distribution are the confidence bands.

Forecast Error Variance Decomposition (FEVD)

FEVD is the breakdown of the forecast error variance into components due to shocks in the series. Fundamentally, variance decomposition tells a researcher the proportion of the variation in a TS accounted for by other variables at a selected time horizon. More exactly, the not correlatedness of the orthogonalized shocks " v_t 's "allow us to decompose the error variance of the s-step-ahead estimate of " y_{it} " into components explained by these shocks, or innovations. Let's look an orthogonalized VAR with m components in VMA representation,

$$y_t = \sum_{i=0}^{\infty} \psi^* (l) v_{t-1} \tag{68}$$

The s step-ahead forecast for y_t is then

$$E_t (y_{t+s}) = \sum_{i=s}^{\infty} \psi^* (l) v_{t+s-l} \tag{69}$$

Defining the s step-ahead forecast as

$$\mathrm{e}_{t+s} = \mathrm{y}_{t+s} - \mathrm{E}_t (\mathrm{y}_{t+s}) \tag{70}$$

we get:

$$\mathrm{e}_{t+s} = \sum_{i=0}^{s-1} \psi^* (l) v_{t+s-l} \tag{71}$$

and its i'th component is given by;

$$\begin{aligned} \mathrm{e}_{\mathrm{t+s}} &= \sum_{i=0}^{s-1}\sum_{j=1}^{m}\psi_{ij}^{*}(l)v_{j,t+s-l} \\ &= \sum_{j=1}^{m}\sum_{i=0}^{s-1}\psi_{ij}^{*}(l)v_{j,t+s-l} \end{aligned} \tag{72}$$

Because the shocks are both serially and contemporaneously uncorrelated, we obtain for the error variance;

$$\begin{aligned} \mathrm{Var}\left(\mathrm{e}_{\mathrm{i,t+s}}\right) &= \sum_{i=0}^{s-1}\sum_{j=1}^{m} Var\left[\psi_{ij}^{*}(l)v_{j,t+s-l}\right] \\ &= \sum_{j=1}^{m}\sum_{i=0}^{s-1}\psi_{ij}^{*}(l)^{2}\, Var\left(v_{j,t+s-l}\right) \end{aligned} \tag{73}$$

All shock components have unit variance, and this implies that

$$\mathrm{Var}\left(\mathrm{e}_{\mathrm{i,t+s}}\right) = \sum_{j=1}^{m}\left(\sum_{i=0}^{s-1}\left(\psi_{ij}^{*}(l)^{2}\right)\right) \tag{74}$$

Where $\left(\sum_{i=0}^{s-1}\left(\psi_{ij}^{*}(l)^{2}\right)\right)$ explains the error variance produced by innovations to " y_j ". Comparing this to the sum of innovation responses, we attain a relative degree how significant variable "j's" innovations are in explaining the disparity in variable "i" at different step-ahead forecasts, i.e,

$$R_{ij,s}^{2} = \left(\frac{\sum_{i=0}^{s-1}\left(\psi_{ij}^{*}(l)^{2}\right)}{\sum_{k=1}^{m}\sum_{i=0}^{s-1}\psi_{ik}^{*}(l)^{2}}\right)*100 \tag{75}$$

The forecast error covariance matrix is therefore given as;

$$Cov\left(\begin{bmatrix} y_{T+1} - y_{T+1}|T \\ \vdots \\ y_{T+h} - y_{T+h}|T \end{bmatrix}\right) = \begin{bmatrix} I & 0 & \cdots & 0 \\ \Phi_1 & I & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ \Phi_{h-1} & \Phi_{h-2} & \cdots & I \end{bmatrix} \left(\Sigma_u \otimes I_h\right) \begin{bmatrix} I & 0 & \cdots & 0 \\ \Phi_1 & I & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ \Phi_{h-1} & \Phi_{h-2} & \cdots & I \end{bmatrix}' \quad (76)$$

and the matrices Φ_i are the empirical coefficient matrices of the Wold moving average representation of a stable process. The operator $\otimes$ is the Kronecker product.

Forecasting

The h-step ahead of a VAR (p) process, $\tilde{y}_{t+h|t}$ is given by the formula;

$$E_t\left[\tilde{y}_{t+h}\right] = \sum_{j=0}^{h-1} \Phi_1^j \Phi_0 + \Phi_p^h y_t \quad (77)$$

Forecasts can be constructed by direct forward recursion beginning at h = 1, but it is often computed using the deviations from the VAR since it includes no intercept,

$$y_t = \Phi_1 \tilde{y}_{t-1} + \Phi_2 \tilde{y}_{t-2} + \ldots + \Phi_p \tilde{y}_{t-p} + \varepsilon_t \quad (78)$$

Using the deviances from h-step ahead forecasts can be computed using the recurrence;

$$E_t\left[\tilde{y}_{t+h}\right] = \Phi_1 E_t\left[\tilde{y}_{t+h-1}\right] + \Phi_2 E_t\left[\tilde{y}_{t+h-2}\right] + \ldots + \Phi_p E_t\left[\tilde{y}_{t+h-p}\right] + \varepsilon_t \quad (79)$$

starting at $E_t\left[\tilde{y}_{t+1}\right]$. Once the forecast of $E_t\left[\tilde{y}_{t+h}\right]$ has been computed the h-step ahead, forecast of "y_{t+h}" ahead is constructed by adding the long-run mean

$$E_t\left[y_{t+1}\right] = \mu + E_t\left[y_{t+h}\right] \quad (80)$$

Chapter Summary

The chapter was organized as follows: unit root tests, cointegration, VAR/VEC Model, Model Validations and model diagnostics, Granger Causality IRF, FEVD, and forecasting. The methodology of this study follows a quantitative one. The tests for unit root and stationarity include; LLC and IPS. All these are

Multivariate Unit Root Tests. The univariate Unit Root Tests considered are ADF, KPSS, and PP. These are used as confirmatory tests to the multivariate Unit Root Tests in this study. Moreover, the appropriate lag length is then looked at. Therefore, the VAR model description follows the procedure, after which some alternative tools are presented to help for the interpretation of the data. Consequently, VAR and VEC models consist of Granger causality, IRFs, and FEVDs.

CHAPTER FOUR

RESULTS AND DISCUSSION

Introduction

The chapter analyses and discusses the results. It presents results of the association between the prices of the products considered, namely, Gas Oil (GASOIL), Premium Gasoline (GASOLINE), Kerosene (KEROSENE), and Liquefied Petroleum Gas (LPG). The rest is organized as Descriptive Statistics, Estimation of Coefficients, Stationarity Tests, Lag Length Selection Criteria, and Cointegration, VAR/VEC model Estimation, Model Diagnostics, and validation, Granger Causality, IRFs, FEVD, and Forecasting. All associated tests and models are generated with R and Eviews software programs. In all, 204 observations are used (January 2007 to June 2015). Training data of 144 observations (January 2007 to December 2012) for modelling and 60 data points (January 2013 to June 2015) for model validations. The 144 observations and 60 data points are arrived at because the data is biweekly in nature. The descriptive statistics of the products are shown in Table 1.

Table 1: Summary Statistics

Statistics	GASOIL	GASOILNE	KEROSENE	LPG
Mean	122.445	123.570	82.989	94.766
Maximum	175.480	177.090	120.420	136.190
Minimum	11.600	49.170	6.470	58.500
Standard Deviation	32.306	31.817	27.186	20.609
Skewness	-0.201	0.1307	-1.988	0.413
Kurtosis	3.374	2.123	6.293	2.292
Number of Observations	144	144	144	144

Reference: Researcher's Computation (2020)

Table 1 presents the statistics for the prices of the products. The highest and lowest mean ex-pump prices are recorded for GASOLINE and KEROSENE respectively. The standard deviations for each variable show that there are greater deviations. The highest ex-pump price over the period was 177 and was recorded by GASOIL price whiles the least was 6.470, recorded by KEROSENE price. This means that GASOIL price recorded the highest ex-pump price whiles KEROSENE price is the least over the period. The variables have low kurtosis values that tend to have light tails or signified lack of outliers with values 2.760, 3.020, 3.640, and 2.760 respectively. We proceed to estimate slope coefficients for the time series models as presented below.

Estimate of Slope Coefficients

Many time series tend to grow over time. In estimating slope coefficients, it is important to carry out a unit root test. The hypothesis is that unit-roots exist. If that hypothesis is rejected, one may use Ordinary Least Squares (OLS) to estimate the slope coefficients. But, if the presence of a unit root is not rejected, then one applies to difference. If an additional test indicates the series is stationary, OLS is applied to estimate coefficients of the slope. Figures 1 and 2 present the time series plots, ACFs, and PACFs of the behaviour of the original series respectively. Figure 1 illustrates how the behaviour of the 4 products within equally spaced time intervals. It is observed that the series do not follow a specific systematic trend. It is also clear that there is no form of seasonality nor periodicity but only upward and downward moving trends over time. That is, the products exhibit an irregular swinging pattern. For GASOLINE price, the series decreases in the early part of the year 2007, reached an all-time low, and increases sharply in the same year and continued steadily throughout 2008 and decreases towards the middle of 2009. It increases from the second part of 2009 sharply and continues in a stepwise fashion between 2010 and 2011, and stabilized throughout 2012. For GASOIL price, the series decreased in the early part of the year 2007, reached its all-time low, and increases again steadily through to the middle of 2008, but drops again at the beginning of 2009 through the middle. It increases from the second part of 2009 and moves in a stepwise manner till 2012. For KEROSENE price, the series is stable in the first half of 2007. It rises steeply during the third quarter of the year 2007 and increases again steadily towards the end. It decreases steadily throughout 2008.

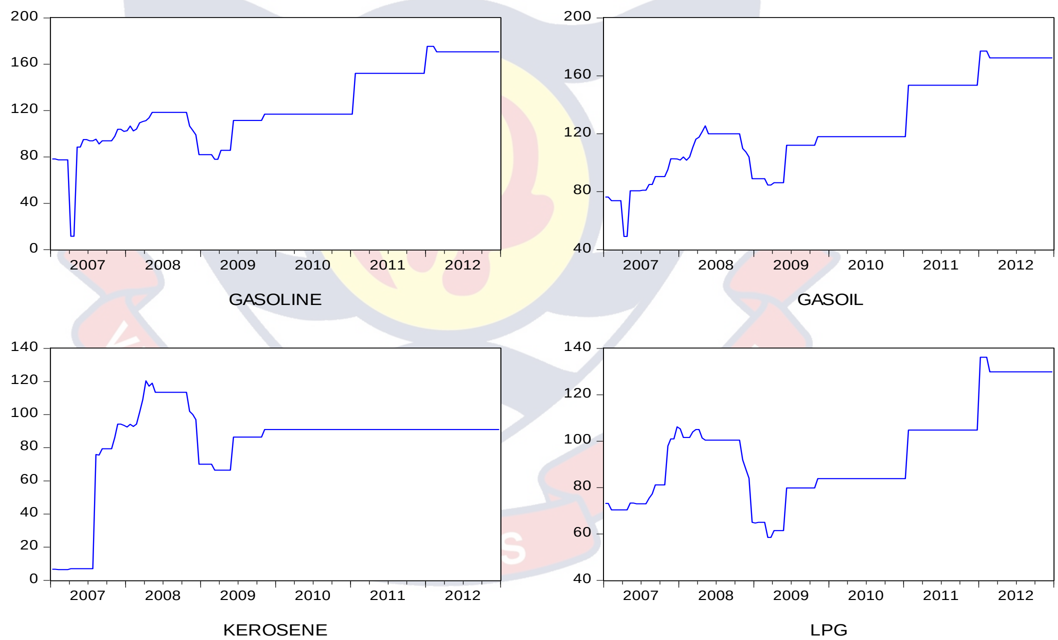


Figure 1: Time Series Plot of the Original Series

It decreases towards the beginning, moves up again, and appears to remain stable till 2012. For LPG price, the series increased steadily during the first three quarters of the year 2007 and increased sharply towards the end of the year. It dropped steadily in the first quarter of 2009 where it attained its all-time low figure in the same year. It then wound up and down from 2010 through to 2012. Thus, the magnitude of the price increases over time as time passed by.

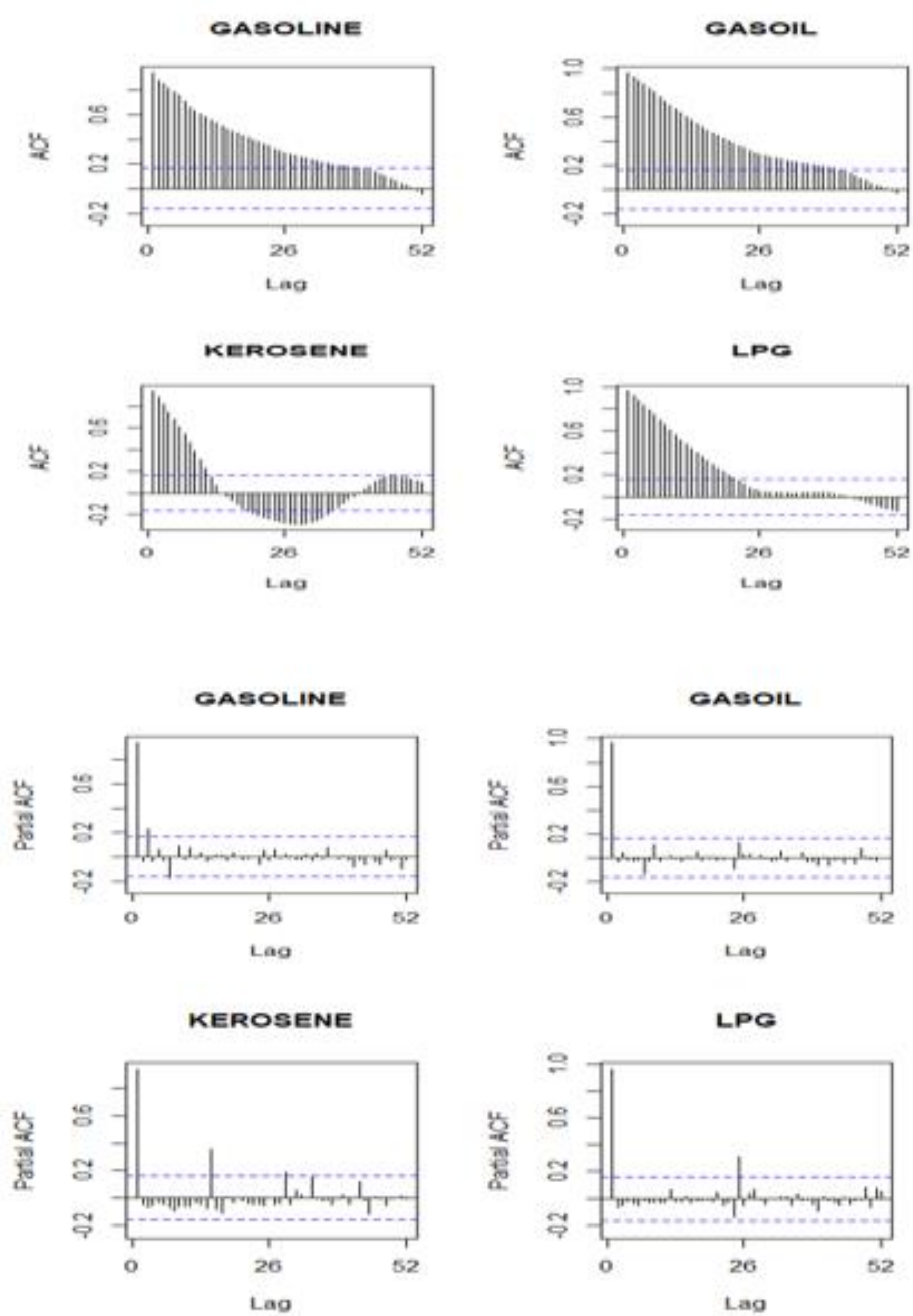


Figure 2: ACFs and PACFs

Thus, there is no consistent trend (upward or downward) over the entire period. Hence, the series appeared to slowly meander up and down. This is an indication that the series is not stationary. Figure 2 provides more information on the stationarity of the series. It shows that the autocorrelations did not die out quickly at higher lags, confirming the non-stationarity behaviour. Unit root tests are undertaken to confirm the stationarity or otherwise of the series. Refer to the next sub-section for the unit root procedures and analyses.

Stationarity Tests

In testing the stationarity of the series, unit root tests were conducted. When estimating VAR/VEC models, it is imperative to test the availability of unit roots. We have numerous ways of testing for the presence of a unit root. The multivariate and univariate methods. However, we focus our attention on the former and use the latter as confirmatory tests. We have chosen to apply, Levin, Lin, and Chu (LLC); Im, Pesaran, and Shin W-Stat (IPS), ADF-Fisher Chi-Square, and PP-Fisher Chi-Square tests. The LLC assumes a common unit root process for all series whiles IPS, ADF-Fisher Chi-square and PP-Fisher Chi-square assume individual unit roots. Others used are ADF, KPSS, and PP. The former, are all Multivariate Unit Root Tests (URTs) whiles the latter are Univariate URTs, which are only used as confirmatory tests. In the univariate case, the null hypotheses (except KPSS) state that the series assumes individual unit root processes. The results are presented in Tables 2 and 3. The URTs are conducted using Equations 2 to 29.

Table 2: Multivariate URTs of the Original Series

Method	Statistic	P-value	Sections	Observations
Null: Unit root (assumes common unit root process):				
Levin, Lin & Chu t*	-0.834	0.202	4	572
Null: Unit root (assumes individual unit root process):				
Im, Pesaran and Shin W-stat	-0.270	0.394	4	572
ADF - Fisher Chi-square	8.871	0.353	4	572
PP - Fisher Chi-square	8.113	0.423	4	572

Reference: Researcher's Computation (2020)

Table 2 presents information on multivariate URTs for the original series and tests that the series contains a unit root against the alternative hypothesis. All the variables have unit roots since we failed to reject the null hypothesis using p-values at 0.05 by the LLC test. Hence, the original series is not stationary. IPS computes distinct tests. This tests the hypothesis that the individual series contains a unit root against the alternative. Hence, the null hypothesis is rejected too since the significance values are all more than 0.05. The univariate URTs in Table 3 are used as confirmatory tests to the conclusions by the multivariate unit root tests. These are also carried out using equations: 7-13 for the ADF Test; for the KPSS test.

It is observed that for ADF, all the p-values are greater than 0.05 and this is an indication of the existence of unit roots. In the case of the KPSS, since the p-values are less than 0.05, there is also the presence of unit roots. There is a need for data differencing (p. 28). Figure 3 depicts the behaviour of the differenced series whiles Figure 4 presents the ACFs and PACFs. Consequently, the differenced method is applied and the first differencing (d=1) is carried out.

Table 3: Univariate URTs of the Original Series

		(Test Statistic)		(P-Values)	
Series	Lag Order	ADF	KPSS	ADF	KPSS
GASOLINE	5	-2.738	2.370	0.269	0.010
GASOIL	5	-2.450	2.437	0.389	0.010
KEROSENE	5	-3.106	0.709	0.116	0.010
LPG	5	-1.975	1.497	0.587	0.010

Reference: Researcher's Computation (2020)

The series appears to be stable after the difference. The ACFs and PACFs indicate no need for further differencing. This is confirmed by the multivariate and individual URTs in Tables 4 and 5.

Figure 3: Time Series Plot of the Differenced Series

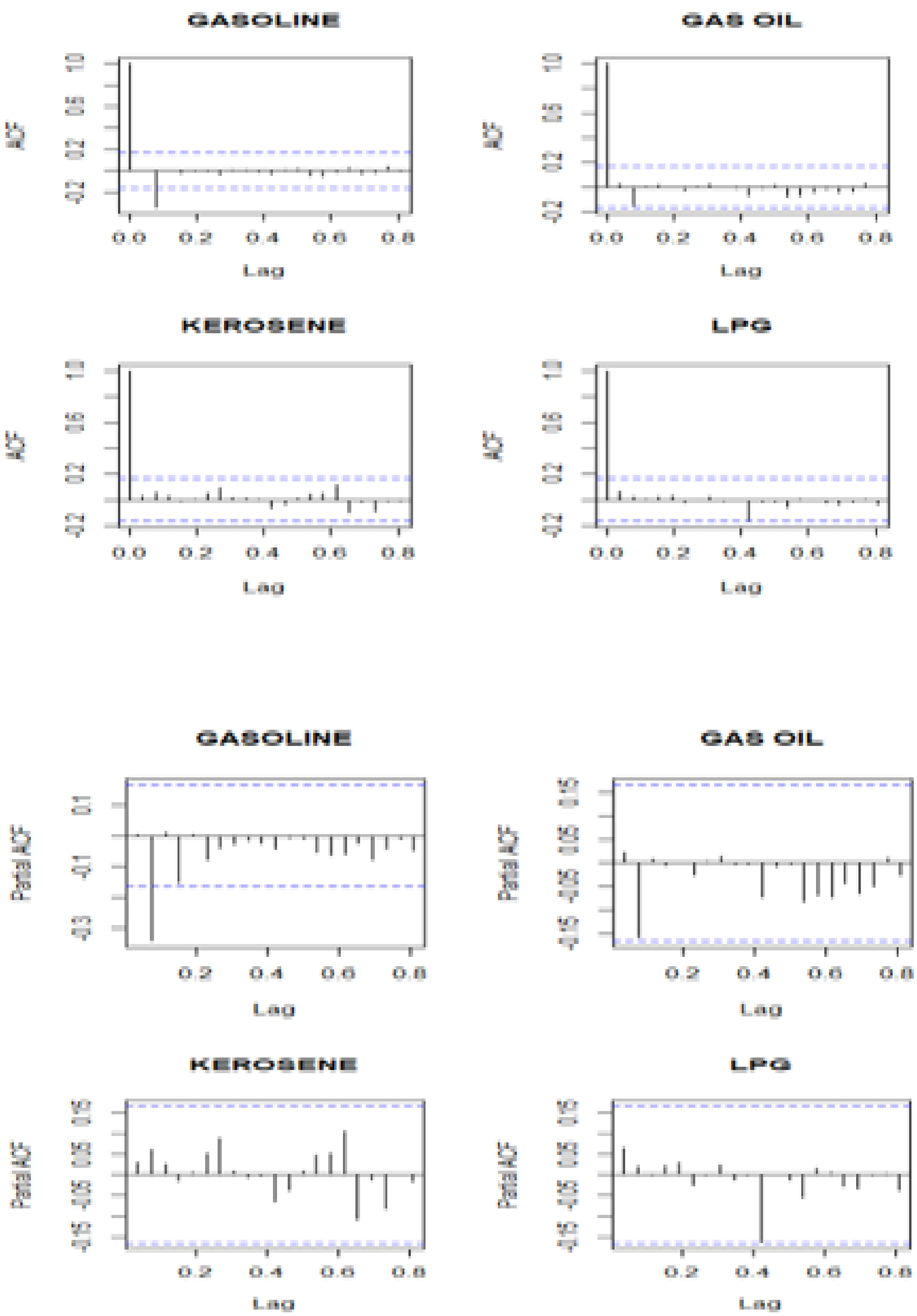


Figure 4: ACFs and PACFs of the Differenced Series

The URTs in Table 4 confirm the stationarity of the differenced series.

Table 4: MURTs of the Differenced Series

Method	Statistic	P-value	Sections	Observation
Null: Unit root (assumes common unit root process):				
Levin, Lin & Chu t*	-25.568	0.000	4	567
Null: Unit root (assumes individual unit root process):				
Im, Pesaran and Shin W-stat	-20.967	0.000	4	567
ADF - Fisher Chi-square	278.284	0.000	4	567
PP - Fisher Chi-square	323.662	0.000	4	568

Reference: Researcher's Computation (2020)

Table 4 shows the multivariate URT for the differenced series. All the variables have no unit roots since all the null hypotheses are rejected using p-values at 0.05 for both the common and the individual unit root processes for both multivariate and individual unit root tests respectively. A univariate URT was again conducted to serve as a confirmatory test and the result is presented in Table 5.

Table 5: URTs of the First Differenced Series

		(Test Statistic)		(P-Values)	
Series	Lag Order	ADF	KPSS	ADF	KPSS
GASOLINE	5	-7.781	0.031	0.010	0.100
GASOIL	5	-5.537	0.045	0.010	0.100
KEROSENE	5	-4.493	0.263	0.010	0.100
LPG	5	-4.473	0.063	0.010	0.100

Reference: Researcher's Computation (2020)

It is observed that for ADF, all the p-values of the series are less than 0.05 and this indicates the stationarity. The KPSS test also showed the same results. We now estimate the models since the series have attained stationarity.

Estimation of VAR/ VEC Models

Estimating parameters of VAR/VEC models require that variables are covariance stationary with their first two moments finite and time-invariant and I(1). VAR for instance cannot be used if the variables are not stationary. Also, if the data is non-stationary, the forecast cannot be done because VAR assumes stationarity. Thus, MLE (or OLS in this case) produces forecasts that mean revert quickly. The URTs confirm that the variables are I(1). We then test for the long run relationship using Johansen's cointegration test. That is if the result confirms that there is a long-run relationship among the variables, we can proceed to the VEC model. The first step involved in estimating is to first determine the lag Length or order. This is carried out using equations 38-41, as presented in Table 6.

Lag Lengths Selection (LLS) Criteria

LLS is significant for VAR/VEC models since selecting too few intervals to result in a cointegrated error and selecting too many intervals may lead to unnecessary loss of degrees of freedom. Three of the LLS criteria i.e., FPE, AIC, and SC support the inclusion of lag 1 as *italicized*, and *starred* in Table 6.

Table 6: Lag Length Selection Criteria

Lag	FPE	AIC	SC
0	1.03 x 10^9	32.107	32.192
*1**	*117944.1**	*23.029**	*23.454**
2	127300.7	23.105	23.869
3	142926.7	23.219	24.3224
4	149122.0	23.259	24.701
5	169942.3	23.385	25.167
6	156708.8	23.297	25.419

Reference: Researcher's Computation (2020)

From Table 6, we can rely on information criteria as only one of these three tests; FPE, AIC, and SC obtained minimum values at the indicated lag. FPE recorded a minimum value of (117944.1) at lag 1, whiles the AIC and SIC also recorded minimum values of 23.029 and 24.454 at lag 1. Thus, the test displays lag 1 as the optimum. Thus, the lag length for the estimation is 1. Once the unit roots and lag length selections are determined for a time series data, the next step is to inspect whether there exists a long-run equilibrium relationship among the variables or not, and this demands cointegration analysis which is important to dodge the risk of spurious regression.

Table 7: Determining the Number of Cointegrated Equations

Number of CE	Eigenvalues	Trace Statistic	p-value	Max-Eigen Statistic	p-value
None *	0.358	79.102	0.000	62.959	0.000
At most 1	0.070	16.143	0.702	10.258	0.720
At most 2	0.033	5.885	0.709	4.778	0.769
At most 3	0.008	1.107	0.293	1.107	0.293

Reference: Researcher's Computation (2020)

Cointegration analysis is important because, if two or more non-stationary variables are cointegrated, a VAR model in the first difference is mis-specified due to the effects of a common trend. The cointegration test determines the type of the regression model to be applied, i.e., the VAR or VEC model (Pradhan & Bagchi, 2013). But because the products serve as endogenous and exogenous variables at the same time, four different determinations are done. The hypothesis being test hypothesis is the null of non-cointegration against the alternative of the existence of cointegration using the Johansen maximum likelihood procedure. This is presented in Table 7. The hypothesis is stated as;

H_0: There is no cointegration equation

Against (81)

H_1: There is a cointegration equation

Conclusions

Remarkably, the Trace test and max-Eigen statistics suggest the existence of a cointegrated equation (CE). We shall take into account this fact at the next step. Since all the series are I(1) and cointegrated, the products ought to be modelled as a VEC model. As a result, a cointegration relationship is obtained. This throws more light on the long run relationships among the products. Consequently, the products; GASOLINE, GASOIL, KEROSENE, and LPG prices are linked by a long run equation. This is presented below.

Long-Run Relationships

The cointegrating (long-run) relationship is estimated to be;

$$GASOLINE = -0.232GASOIL + 0.073KEROSENE - 0.775LPG \qquad (82)$$

Thus, with GASOLINE price as the endogenous variable, the long-run relationship indicates that the ex-pump prices of the other products have long run effects. Specifically, the results indicate that the other products have a negative relation with GASOLINE price in the long run (except KEROSENE), all things being equal. Table 8 throws more light on the relationship established among the products concerning how imbalances in ex-pump prices in previous periods are corrected in the long run.

Long-Run Equilibrium

The coefficients of the error correction terms (*ECT*) [Table 8] show the speed of adjustments of disequilibrium in the period under study. The negative sign associated with the error term is simply a departure in one direction. These are satisfying as they imply convergence in the long run. That is, deviation from the long run is corrected.

Table 8: VEC Model Coefficients

Parameters	Coefficient	S.E	t-statistic
Gasoline Model:			
$(GASOLINE)_{t-1}$	0.691	0.189	3.650*
$(GASOIL)_{t-1}$	-0.602	0.386	-1.561
$(KEROSENE)_{t-1}$	0.027	0.091	0.294
$(LPG)_{t-1}$	-0.580	0.262	-2.211*
Constant	0.006	0.805	0.007
ECT	-1.613	0.145	-11.118*
Gasoil Model:			
$(GASOLINE)_{t-1}$	0.524	0.126	4.165*
$(GASOIL)_{t-1}$	-0.783	0.256	-3.059*
$(KEROSENE)_{t-1}$	-0.002	0.060	-0.030
$(LPG)_{t-1}$	-0.214	0.174	-1.227
Constant	0.017	0.535	0.032
ECT	-0.695	0.096	-7.215*
Kerosene Model:			
$(GASOLINE)_{t-1}$	0.058	0.163	0.359
$(GASOIL)_{t-1}$	-0.002	-0.085	-0.256
$(KEROSENE)_{t-1}$	-0.518	0.078	-6.652*
$(LPG)_{t-1}$	0.059	0.225	0.263
Constant	0.001	0.692	0.002
ECT	-0.039	0.125	-0.313
LPG Model:			
$(GASOLINE)_{t-1}$	0.054	0.106	0.505
$(GASOIL)_{t-1}$	-0.080	0.216	-0.370
$(KEROSENE)_{t-1}$	0.010	0.051	-0.197
$(LPG)_{t-1}$	-0.450	0.147	-3.058*
Constant	0.020	0.452	0.044
ECT	-0.036	0.081	-0.437

Reference: Researcher's Computation (2020) **significant coefficients*

The negative sign associated with the coefficients of the error term of GASOLINE price indicates that the models are stable dynamically. This suggests that the speed of adjustments is high. The magnitude of the correction of the imbalances, however, suggests for instance that, 61.3% of the imbalances in GASOLINE prices are corrected every two weeks. Concerning GASOIL prices, it indicates 69.5% of shocks in its prices are corrected every two weeks. For the KEROSENE price, 3.9% of such imbalances are corrected every two weeks. In the case of LPG price, only 3.6% of such imbalances are corrected.

Short-run Relationships

The short run relationships of the models are explained by the VEC model coefficients as presented in Table 8. Looking at the coefficients, it is observed that in the short-run, GASOLINE price [3.65] is significant. This is an indication that GASOLINE price exhibits an increment of 69.1% by itself and 2.21% reduction by GASOIL price whiles the others are not significant. Also, it is observed that GASOIL price [4.17] is significant. This is an indication that GASOIL price exhibits an increment of 52.4% by GASOLINE price with a 78.3% reduction by itself. The other products also exhibit both increment and reduction by themselves and/or other products. This is because the coefficients of these products are significant. The short-run results also indicate that the variables influence each other. Considering GASOLINE price as the dependent variable, it appears the ex-pump prices of the other products influence it. The consequence of this result is that increase ex-pump prices of one or more products are likely to influence others, as the R-squared, which describes the proportion of variability explained by the models is between 23.2% and 53.9%. Coefficients of the VEC models are presented in Table 8.

Now, having analyzed both the short and long-run relationships existing among the variables, the VEC models are estimated, diagnosed, and validated, and finally, forecasts are generated. Other areas looked at are Granger causality, Impulse Response Functions (IRFs) as well as Forecast Error Variance Decompositions (FEVDs).

Estimations of the VEC Models

The VEC models are estimated using equations 49-55, and the results of VAR are reported by the 4 equations below. The VEC models are computed with one lag. The models relating the products to their lags and that of others may best be described as;

$$\begin{bmatrix} w_t \\ x_t \\ y_t \\ z_t \end{bmatrix} = \begin{bmatrix} 0.006 \\ 0.017 \\ 0.001 \\ 0.020 \end{bmatrix} + \begin{bmatrix} 0.691 & 0.524 & 0.058 & 0.054 \\ -0.602 & -0.783 & -0.002 & -0.080 \\ 0.027 & -0.002 & -0.518 & 0.010 \\ -0.580 & -0.214 & 0.059 & -0.450 \end{bmatrix} \begin{bmatrix} w_{t-1} \\ x_{t-1} \\ y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} ECT_{pg} & 0 & 0 & 0 \\ 0 & ECT_g & 0 & 0 \\ 0 & 0 & ECT_k & 0 \\ 0 & 0 & 0 & ECT_l \end{bmatrix} \begin{bmatrix} -1.613 \\ -0.695 \\ -0.039 \\ -0.036 \end{bmatrix} \quad (83)$$

where, $\begin{bmatrix} w_t \\ x_t \\ y_t \\ z_t \end{bmatrix}$, represents the products (GASOLINE, GASOIL, KEROSENE, and LPG prices), at the time, t, $\begin{bmatrix} ECT_{pg} & 0 & 0 & 0 \\ 0 & ECT_g & 0 & 0 \\ 0 & 0 & ECT_k & 0 \\ 0 & 0 & 0 & ECT_l \end{bmatrix}$ refers to the error corrected terms (ECT) for each model, (pg, g, k, and L respectively representing GASOLINE, GASOIL, KEROSENE, and LPG prices) and $\begin{bmatrix} w_{t-1} \\ x_{t-1} \\ y_{t-1} \\ z_{t-1} \end{bmatrix}$ referring to the lags of the products (i.e. lag 1). The summary of the results of the VEC models is presented in Table 9.

Table 9: Summary Results of the Models

Statistics/Products	GASOLINE	GASOIL	KEROSENE	LPG
F-statistic	31.515	18.810	9.884	8.133
Prob (F-statistic)	0.000	0.000	0.000	0.000
S.E.	9.562	6.351	8.218	5.367
R-squared	53.9%	41.1%	26.8%	23.1%

Reference: Researcher's Computation (2020)

Table 9 is a summary of the statistics of the VEC models. The results indicate that the models perform creditably well. GASOIL appears to be the best in terms of the variability accounted for.

After modeling, some forecasts are normally estimated. But before the estimated model can be used to generate any forecast, it is imperative to undertake residual analysis or model diagnostics. The diagnostic test results include Q-statistics, residual portmanteau test, residual serial correlations, and white heteroscedasticity test. Refer to equations 56-63. Tables 10 to 12 provide information on the analysis of the residuals of the models.

Table 10: VEC Residual Portmanteau Tests for Autocorrelations

Lags	Q-Stat	Prob.	Adj Q-Stat	P-value	df
1	7.6189	NA*	7.677	NA*	NA*
2	21.603	0.157	21.879	0.147	16
3	29.843	0.576	30.312	0.552	32
4	44.304	0.625	45.228	0.587	48
5	50.472	0.891	51.641	0.867	64
6	111.265	0.012	115.351	0.006	80
7	131.458	0.010	136.684	0.004	96
8	188.126	0.000	197.038	0.000	112
9	197.007	0.000	206.575	0.000	128
10	202.285	0.001	212.289	0.000	144
11	216.121	0.002	227.393	0.000	160
12	230.102	0.004	242.784	0.001	176

Reference: Researcher's Computation (2020)

The null hypothesis is that there are no residual autocorrelations up to lag h. The test is valid only for lags larger than the selected lag order. We observe that the residual passes the white noise test since no autocorrelation is left in the VEC model after lag 1.

Table 11: VECM Residual Serial Correlation LM Tests

Lags	LM-Stat	P-values
1	18.022	0.323
2	18.924	0.273
3	8.073	0.947
4	14.278	0.578
5	6.001	0.988
6	66.795	0.000
7	20.731	0.189
8	69.676	0.000
9	8.623	0.928
10	5.030	0.996
11	14.290	0.577
12	14.019	0.597

Reference: Researcher's Computation (2020)

Table 12: VEC Residual Heteroskedasticity

Dependent	R-squared	F (34,163)	P-value	Chi-square (34)	P-value
res1*res1	0.292	3.415	0.001	38.236	0.290
res2*res2	0.093	0.850	0.615	12.184	0.093
res3*res3	0.216	2.280	0.009	28.266	0.216
res4*res4	0.065	0.573	0.881	8.474	0.065
res2*res1	0.215	2.273	0.009	28.198	0.215
res3*res1	0.155	1.516	0.116	20.262	0.155
res3*res2	0.181	1.835	0.041	23.753	0.181
res4*res1	0.064	0.569	0.884	8.414	0.064
res4*res2	0.049	0.425	0.964	6.387	0.049
res4*res3	0.265	2.980	0.001	34.654	0.264

Reference: Researcher's Computation (2020)

Table 12 shows the results of the test of Heteroscedasticity. This is used t o test for Heteroscedasticity in a linear regression model and assumes that the erro r terms are normally distributed. It tests the hypothesis that variances of the series are the same. It is concluded that the variances of the series are constant. This is b ecause the p-value for the Chi-Square test is more than 0.05. This assertion is supp orted by the p-values of both the F and Chi-square tests.

Model Validations

Model validation and certification is a significant step in modelling. This method mostly helps stakeholders and industry players in minimizing the cost, time, and risk associated with comprehensive product testing (Salamanca, Krpo, Martilli, & Clappier, 2010). They continued to say that these procedures are necessary practices for measuring and building trustworthiness in statistical models. As mentioned earlier, the data was portioned into two. The training and validation data. The training data (70%; 2007 to 2012) of the data points are used in the modelling and the validation is done with the remaining proportion of the data (2013 to 2015).

Out of sample forecasts covering the period, 2013 to 2015 is used in the model validation to generate products' prices and compared each to its original prices. Figure 10 throws more light on the original prices of the products and the forecasts. The Forecasts of the prices of each product has from the estimated model have been superimposed on the original prices (remaining 30% of the data) over the same time horizon (2013 to 2015) as shown in Figure 5. It is observed that the forecasts appear to mimic the price behaviours of the original products, which proves the effectiveness and accuracy of the models.

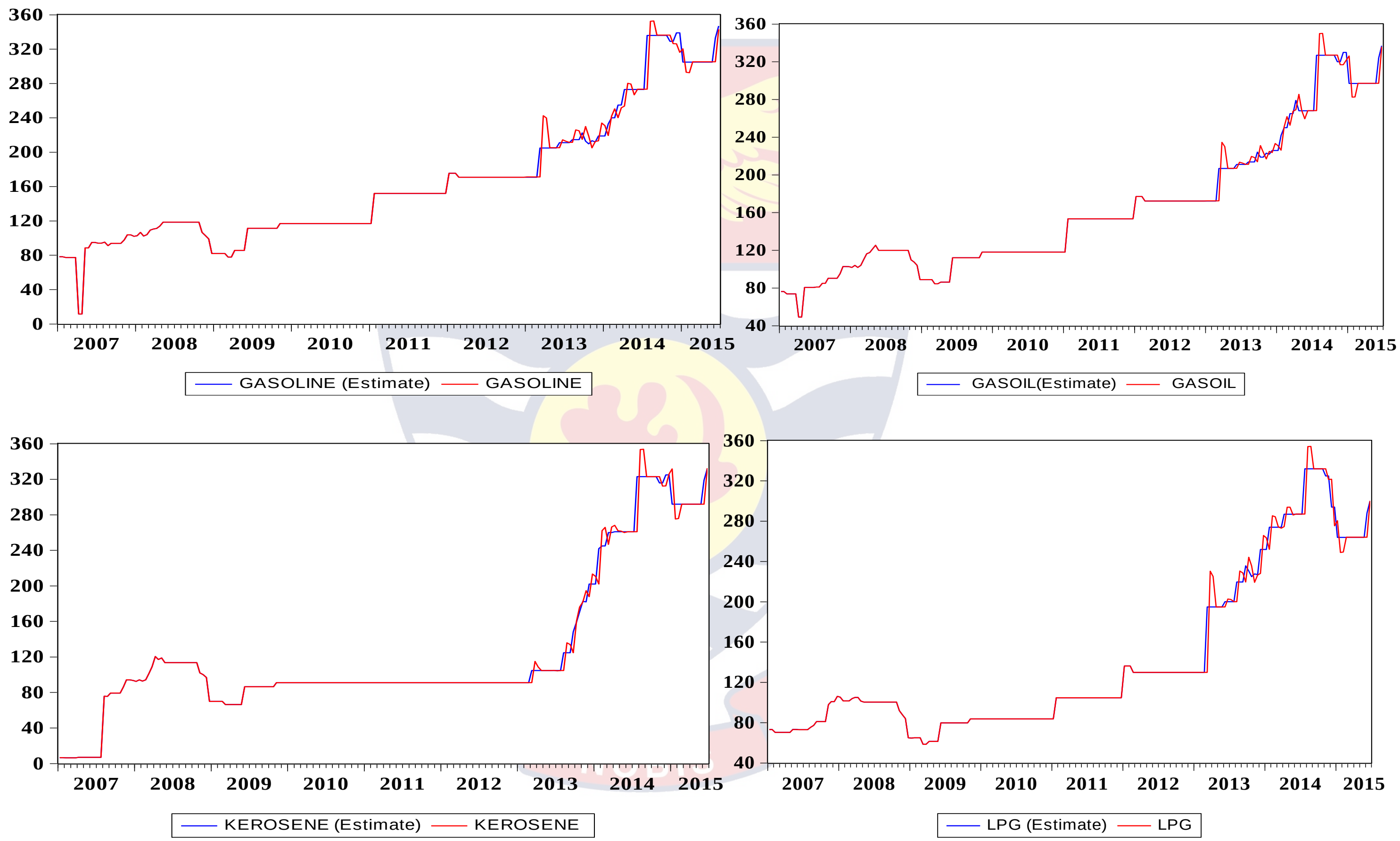


Figure 5: Model Validation

Forecasts of Ex-Pump Prices of the Products

After model diagnostics and validation, we, therefore, forecast the prices of the products for the next 12 periods (6 months) as shown in Table 13. This was also aided by equations 77 – 80.

Table 13: Forecasts of Ex-pump Prices for the four Products

Period	Premium Gasoline	Gas Oil	Kerosene	LPG
1	172.037	172.360	91.401	130.290
2	173.121	172.360	91.887	130.693
3	173.962	172.360	92.475	131.114
4	174.684	172.360	93.115	131.548
5	175.392	206.830	93.765	131.985
6	176.123	206.830	94.406	132.420
7	176.874	206.830	95.038	132.854
8	177.631	206.830	95.667	133.287
9	178.389	206.830	96.296	133.720
10	179.144	206.830	96.926	134.153
11	179.897	211.110	97.557	134.586
12	180.650	211.110	98.188	135.019

Reference: Researcher’s Estimation (2020)

It was observed that the forecasts for the series increased steadily over time. This conforms to the trend experienced in the original products (Figure 1). Thus, prices of petroleum products have always been on the ascendancy over the years and this is what is expected to be experienced over time.

Next, we look at causality among the products, hereafter referred to as Granger causality. This is presented in the next subsection.

Granger Causality Tests

After the diagnostic tests, the Granger causality between the variables is investigated with the aid of equations 64 and 65. The result is presented in Table 14.

Table 14: Granger Causality

Dependent variable: GASOLINE			
Excluded	Chi-square	df	P-value
GASOIL	4.889	1	0.027
KEROSENE	0.086	1	0.769
LPG	2.436	1	0.119
All	32.268	3	0.000
Dependent variable: GASOIL			
Excluded	Chi- Chi-square	df	P-value
GASOLINE	17.345	1	0.000
KEROSENE	0.001	1	0.976
LPG	1.506	1	0.220
All	33.009	3	0.000
Dependent variable: KEROSENE			
Excluded	Chi- Chi-square	df	P-value
GASOLINE	0.129	1	0.720
GASOIL	0.065	1	0.798
LPG	0.069	1	0.793
All	0.241	3	0.971
Dependent variable: LPG			
Excluded	Chi- Chi-square	df	P-value
GASOLINE	0.255	1	0.614
GASOIL	0.137	1	0.711
KEROSENE	0.039	1	0.844
All	0.349	3	0.951

Reference: Researcher’s Computation (2020)

Table 14 shows the Granger causality test for the variables. The null hypothesis being tested in each case is that the products do not granger cause one another, against the alternatives that they do. The hypothesis testing for GASOLINE price alongside all others respectively ganger cause each other. Thus, there is causality that at least one of the products granger causes GASOLINE price. For instance, GASOIL price granger causes GASOLINE price. However, KEROSENE and LPG do not granger cause it. Testing for the causality between GASOIL price and others, we observe that only GASOLINE price granger causes GASOIL price. This is a clear indication that GASOLINE and GASOIL prices Granger cause each other. Thus, the GASOLINE price granger causes GASOIL price and vice versa. Testing for causality between KEROSENE prices and others, we also observe none of the other product prices significantly granger cause KEROSENE price. For the test of causality between LPG and other series, we notice that none of the products' Granger causes LPG. But GASOLINE and GASOIL prices recorded the highest Chi-Squared values of (0.1288 and 0.2545) among the others for both KEROSENE and LPG prices, which means that they are the major contributors to explaining the variability in prices of these products to some extent. But, this contribution is just not significant. In summary, a bidirectional relationship is observed between GASOLINE and GASOIL prices.

We then move on to look at the effects of a shock to an endogenous variable on itself or another endogenous variable, known as Impulse Response Functions (IRFs).

Impulse Response Functions (IRFs)

The IRFs measure the effects of a shock to an endogenous variable on itself or another endogenous variable. Thus, the impulse response has been conducted to describe the reaction of variables to the shock caused by exogenous changes. It helps in identifying the reaction of the variable when a positive shock of one standard deviation is given to the error terms and to know how the variables react to each other. So, for each variable from each equation separately, a unit shock is applied to the error terms and the effects are noted. A standard decomposition is used to identify the shocks on levels of the endogenous variables. Figure 6 presents the results of the IRFs. This was also undertaken using equations 66 and 67.

When there is a change (shock) in GASOLINE price, its price decreases sharply (response negatively to changes in its prices) within the first five periods. Prices then increase slightly over the next five time periods, decreases slightly again, and then stabilizes till the end. This indicates that in the short to medium terms, GASOLINE price is affected by changes in its price. However, in the long term, it appears to have attained stability. When there is a change in GASOIL price, GASOLINE price increases sharply (response positively to changes in GASOIL price) within the first five periods (This is the period at which the effect of changes in GASOIL price on itself and that of GASOLINE price appears to be the same). Prices then increase slightly over the next five time periods and then stabilizes. This indicates that in the short to medium terms, GASOLINE price is affected by changes in GASOIL price. However, in the long term, it appears to have attained stability. This observations is not surprising as the Granger causality test carried out earlier also indicated a bidirectional relationship between GASOLINE and

GASOIL prices. However, shocks in the prices of KEROSENE and LPG do not significantly affect GASOLINE price as changes in the prices of these products appear to not influence GASOLINE prices, even though GASOLINE prices appears to have reacted slightly to changes in LPG prices in the short term period.

When there is a change (shock) in the price of GASOLINE, GASOIL prices decrease sharply (response negatively to changes in its GASOLINE prices) within the first four periods. Prices then decrease slightly over the next six time periods and then stabilizes. This indicates that in the short to medium terms, GASOIL price is affected by changes in GASOLINE price. However, in the long term, it appears to have attained stability. When there is a change in GASOIL price, its price increases steadily (response positively to changes in its prices) within the first four periods (This is the period at which the effect changes in GASOIL price on itself and that of GASOLINE price appears to be the same or be in equilibrium). Prices then increase slightly over the next six time periods and then stabilizes. This indicates also that in the short to medium terms, GASOIL price is affected by changes in GASOLINE price. However, in the long term, it also appears to have attained stability. This observation is not surprising as the Granger causality analysis indicates a bidirectional relationship between these two products. However, shocks in KEROSENE and LPG prices do not affect GASOIL price as changes in the prices of these products appear to have no influence

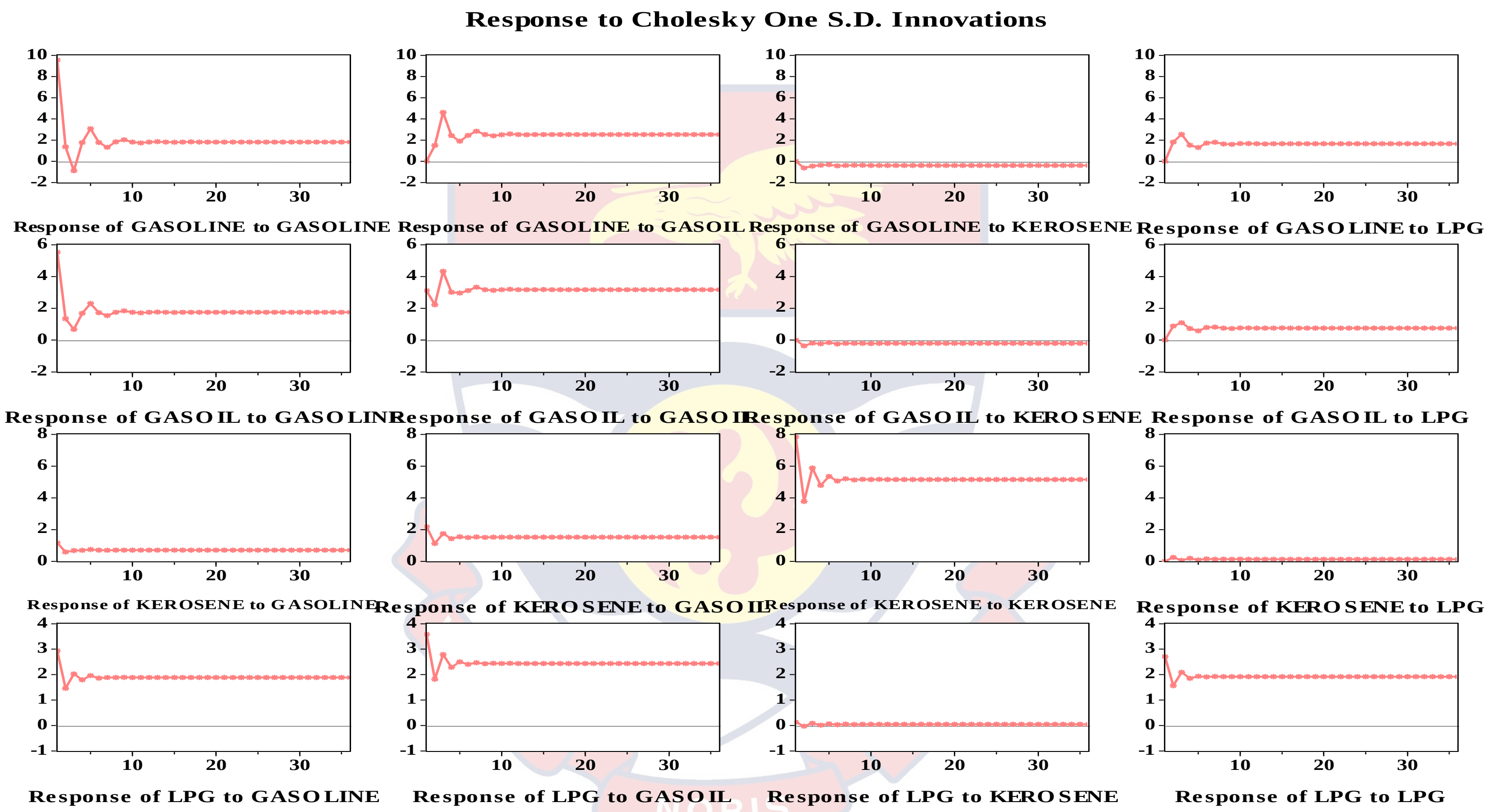


Figure 6: IRFs for GASOLINE, GASOIL, KEROSENE, and LPG Prices

When there is a change in the price of GASOLINE, KEROSENE price increases steadily (response positively to changes in GASOLINE prices) and continuously over the entire period, stabilizing towards the end. This indicates that in the short term, KEROSENE price is affected slightly by changes in GASOLINE price. However, in the medium to long term, it appears to have attained stability. Also, when there is a change in GASOIL price, KEROSENE price decreases steadily within the first six periods Prices then stabilizes over the remaining periods. This also shows that in the very short term, KEROSENE price is affected by changes in GASOIL price slightly. However, in the medium and long terms, it seems to have attained stability. This supports the Granger causality analysis earlier of no causality between GASOLINE and GASOIL, and KEROSENE prices. Also, shocks in LPG price do not affect KEROSENE price as changes in the prices of this product appear to have no influence, even though it also reacted positively within the first eight periods and stabilized afterward.

Finally, when there is a change in GASOLINE price, the LPG price increases progressively (response positively to changes in GASOLINE prices) within the four periods. This indicates that in the very short term, the price of LPG price is affected by changes in GASOLINE price Prices then decrease slightly throughout the remaining periods, with some stabilities experienced afterward. Also, when there is a change in GASOIL price, that of LPG (responds negatively) throughout the periods. This shows that in the short term, LPG price is affected by changes in GASOIL price. However, in the medium and long terms, it seems to have attained stability. Still, when there is a change in LPG price, its price decreases steadily (response negatively to changes in its price) throughout the periods.

The researcher now delves into the breaking down of the forecast error variance into components (Forecast Error Variance Decompositions, hereafter referred to as FEVD) due to shocks in the series, which fundamentally, explains the proportion of the variation in a time series accounted for by other variables at a selected time horizon.

Forecast Error Variance Decomposition (FEVD)

The results of FEVD are illustrated below. This analysis supplements the IFRs and granger causality analysis. These results show how much a variable's shock is explained by movements in its variance and that of other variables. It also shows how much shock is explained by a variable's movements in its variance and other variables. This was also done using Equations 68–76. Figure 7 gives more insights into the FEVDs of the four products.

The first image (top left) gives the FEVDs of the GASOLINE price. It shows that at the end of the first period, variability in GASOLINE price is exclusively due to innovations in itself and not to the others. This shows how exclusive GASOLINE is in terms of its use. But over time, the variation due to GASOLINE price is shared among itself and the others, particularly, GASOIL. In the long term, the proportion of variability in GASOLINE price explained by GASOIL increases steadily whiles that of GASOLINE decreases. Thus, in the short term, changes in GASOLINE price are accounted for by themselves, but in the medium to long terms, more of the variability is accounted for by GASOIL prices. This indicates that response to changes in GASOLINE price is reflected in general changes in prices of GASOIL. However, there does not appear to be a response to change in both products even in the long run. This means variability in GASOLINE price is much more explained by itself than the closest substitute, GASOIL even in the long short-run. This assertion supports the granger causality and IRFs and FEVD conclusions earlier of bidirectional relationships between GASOLINE and GASOIL prices.

For GASOIL price (top right), at the end of the first period, a very high proportion of the variation in its price is due to GASOLINE price changes, with the remaining proportion attributed to itself. Thus, unlike the case of GASOLINE where it exclusively explained variations in its prices by itself, the same cannot be said for GASOIL in the same period. KEROSENE and LPG price did not account for any variability. This means that GASOLINE price continues to account for an appreciable amount of variability in GASOIL price even in the long run. This is an indication of the competitiveness usage (advantage) of GASOLINE concerning GASOIL. This again shows how exclusive GASOLINE is in terms of its usage. But over time, the variation due to GASOIL prices is shared among itself and the others, particularly, GASOLINE price. This claim also supports the granger causality and IRFs conclusion earlier that GASOIL price granger causes GASOLINE price and vice versa. Thus, the bidirectional relationship between GASOIL and GASOLINE prices in the medium and long terms.

At the end of period one (bottom left), a very high proportion of the variation in KEROSENE price is due to itself, with the remaining proportions by GASOIL and GASOLINE.

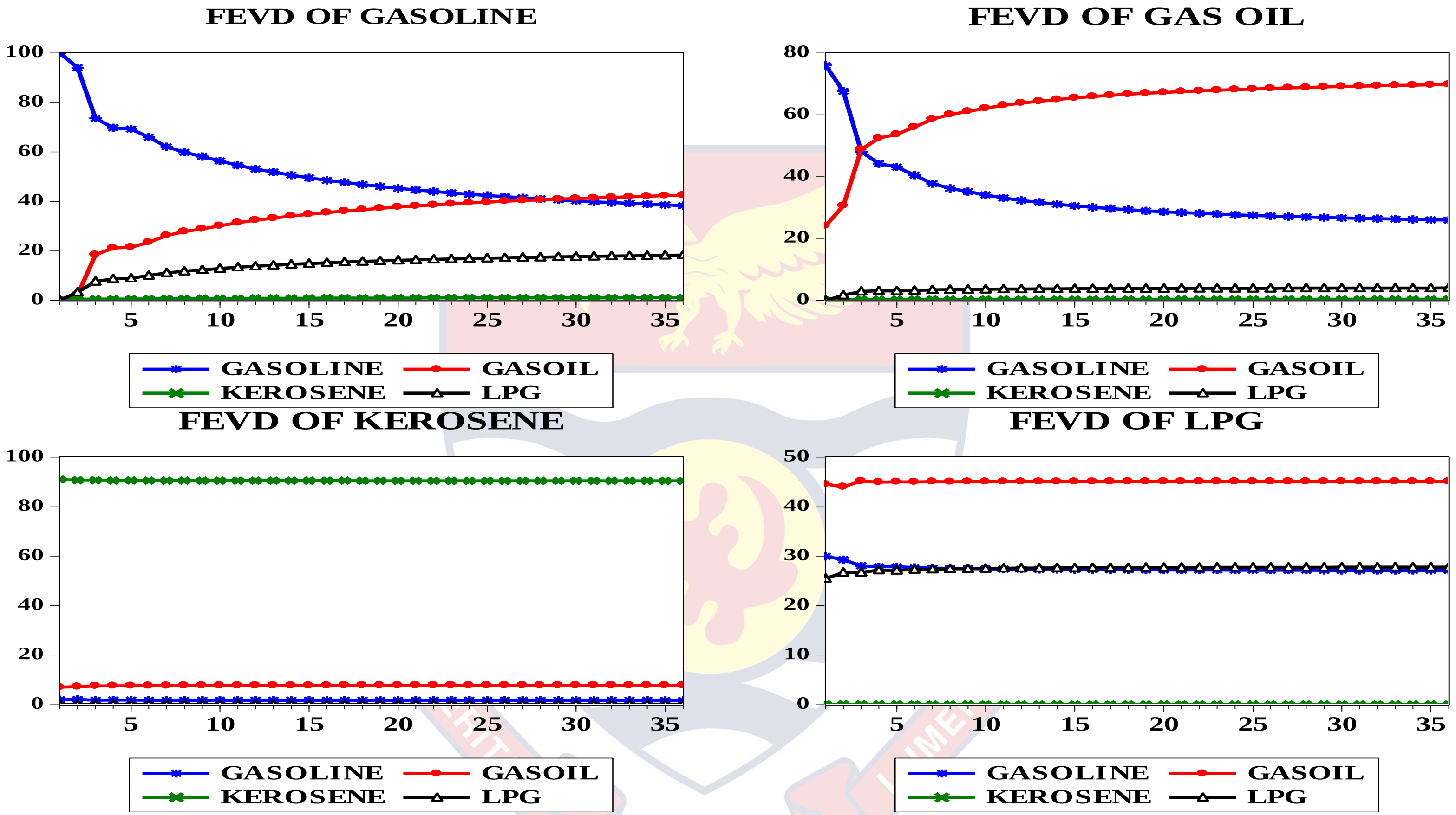


Figure 7: Plots of the FEVD of the Four Products

LPG price did not account for much of the variability in KEROSENE price. But over time, as the variation due to KEROSENE price on itself reduces, particularly, GASOLINE price begins to account for some variability in KEROSENE price. Generally, variability in KEROSENE price has a little contribution from variability in the other three products, particularly, from GASOLINE and LPG prices. This means that changes in the price of KEROSENE are not noticeably reflected in changes in LPG and GASOIL prices but GASOLINE prices to some extent. This further shows the other three products may not serve as good substitutes for KEROSENE. Thus, in the short term, variation in KEROSENE price is accounted for solely by KEROSENE itself, and it continued throughout the entire period even though it decreases, but with GASOLINE price also increased to some extent in the medium to long terms. This assertion supports the granger causality, IRFs, and FEVD conclusions earlier that GASOLINE and GASOIL prices affect the prices of others.

Unlike the other products accounting for the greatest proportions of variations in their prices in the first period, LPG price is not the same. LPG price decreases slowly in a linear fashion throughout the long term. The variability in its price is very highly accounted for by both GASOLINE and GASOIL prices competitively throughout the entire period. This shows that responses to changes in LPG price are rather much more reflected in response to GASOLINE and GASOIL prices.

In summary, the Granger causality, IRFs, and FEVD analyses indicate that in the short term, changes in GASOLINE price is accounted for by itself exclusively in the short term, but in the medium to long terms, the variability is shared between itself and GASOIL price, with GASOIL explaining more of it. Also, in the short term, changes in GASOIL price are accounted for by GASOLINE price, but in the medium to long term, the greater proportion is again explained by GASOIL price itself. Furthermore, variation in KEROSENE price is accounted for mainly by itself. Finally, even though the variability explained by LPG price by itself decreases generally, none of the other series meaningfully explained its variability well, as indicated by the granger, IRFs, and FEVD analysis. The findings of this study are discussed in other works, as presented in the next section.

Summary Results of VAR Model

In the previous section, the VEC model at lag 1 has been obtained. The performance of the model shows that the model closely fits the data. This is seen in the validation of the model. However, the model statistics show that the percentage variation in (GASOLINE, GASOIL, KEROSENE, and LPG prices) is 53.86%, 41.06%, 26.8%, and 23.15%, respectively. These figures are surprisingly much lower than depicted in the model fit graph in Figure 10. What is more surprising is that the performance statistics are much higher for a VAR model which are given in Table 15. In Table 15, it can be seen the lowest R-square (90.3%) is obtained for the model for GASOLINE prices. The fitted VAR graphs can be seen in Appendices A and B. The fits obtained in the graphs for VEC are better than those obtained in Appendices A and B. It is therefore anticipated that the performance statistics should be at least the same as in Table 15 since the VEC model is supposed to be an improvement over VAR.

Table 15: Summary Results of VAR Models

Statistics/Products	GASOLINE	GASOIL	KEROSENE	LPG
F-statistic	294.988	763.646	379.972	497.413
p-value	0.000	0.000	0.000	0.000
S.E.	9.468	5.928	0.366	4.573
R-square	0.903	0.960	0.923	0.940

Reference: Researcher’s Computation (2020)

Discussion of Findings

The study revealed there were dynamic relationships among the ex-pump prices of the series in the long run. Also, the models were seen to be dynamically stable. The magnitudes, however, suggest for instance that the imbalances in the prices of the products are corrected every two weeks. The short-run results also indicated that the prices influence each other. The consequence of this result is that in the short run, a change in the price of a particular series will consequently lead to a change in the price of the other, especially, GASOLINE and GASOIL prices in the long run. These relationships are confirmed by the IRFs, FEVD, granger causality analyses, as well as the forecasts for the various series. That is, changes in prices of some of the products have direct effects on themselves and others. These results show how much a products’ shock is explained by movements in its variance and that of others. For the forecasts, for instance, it was observed that the products increased steadily over time, mimicking the trends of the original products. This supports the fact that prices of petroleum products for some time now have been moving up as seen from the time series plot of the original series.

Thus, prices of these products have always been on the ascendancy over the years and this is what is expected to be experienced in the years ahead, even though the magnitude is not expected to be high. High-pitched upsurges in the price of oil (petroleum products) are generally recognized to have important effects on economic and macroeconomic policies. In particular, the very recent increases recorded in the world oil market are a cause of concern about a possible slowdown in the economic performance of most developed countries. Thus, not surprisingly, a considerable body of economic studies have looked at the canals through which oil price shocks impact economic variables. All other things being equal, these products are a source of energy in the country, and energy is the backbone of every economy, and an increase in these products affects the economy as a whole.

These assertions are seconded by the following researchers. According to Pindyck and Rotemberg, (1983), extra clarifications of the impact of oil price shocks include adjustment costs, investment under doubt, and the behavior of monetary powers that be. They continued that after a rise in energy prices, firms would move from being energy-intensive to energy-efficient, and since these adjustments cannot be achieved speedily, there would, therefore, be growth in unemployment and underutilization of resources. Furthermore, oil price shocks have the consequence of a temporary reduction in cumulative output because of sectoral shifts. Davis and Haltiwanger (2001) also offered evidence of the effects of oil price changes on U.S. manufacturing jobs.

However, the studies of some research are contrary to the conclusions already observed. In the study of Lee and Ni (2002), they assert that not all sectors of the economy are equally affected by oil price shocks. More significantly, they found that the severity of these adverse effects of oil shocks is not linked with the energy intensity of the sector. Other studies for instance, also underscore the significance of doubt since in periods characterized by oil price instability, firms have an enticement to delay investment choices (Van Soest et al., 2000; Bernanke, 1983).

It has been observed that the fitness of the model obtained as reflected in the model fit graph is not what is captured in the model statistics. It is anticipated that the least variation explained in any of the series components should be higher than 90%, a result obtained by the corresponding VAR model.

Chapter Summary

In this chapter, it has been observed that there exists a general upward trend in the ex-pump prices of the products over the period. It also shows that changes in prices of some of the products influence others in the short run, but in the medium to long terms, stability is attained, with Premium Gasoline and Gas Oil prices accounting for most of the variations in changes in prices. The Granger causality, IRF, and FEVD analyses all affirm this result. A VEC model of order 1 is found suitable for the data and appears to perform better than the VAR model with the least R-square of 90.3% for Premium Gasoline prices and the highest of 96% for Gasoil prices. The results point to the competitive nature of prices in both Premium Gasoline and Gas Oil and reflect the competitive usage of the two products.

CHAPTER FIVE

SUMMARY AND CONCLUSIONS

Overview

The chapter presents the summary, conclusion, and recommendations based on the findings of the study. In all the previous chapters, a summary of work done has been provided. All of these summaries are further presented in this chapter to provide a coherent and detailed summary for the entire work. From this summary, the conclusion of the study were drawn along with relevant recommendations.

Summary

The study is guided by the following objectives: examine trends of prices of ex-pump prices of the petroleum products, examine how changes in the price of one product influence the prices of others, taking into account the period over which these changes are significant and to obtain a multivariate time series model for the prices of these petroleum products. Data spanning January 2007 to June 2015 was collected from the National Petroleum Authority of Ghana and covered four petroleum products: Gasoline, Gas Oil, Kerosene, and Liquefied Petroleum Gas (LPG). Analysis of the data is carried out with the aid of R-Console and Eviews software programs, using the technique of Vector Error Correction modelling. This technique is found suitable as the data obtained constitutes a multivariate time series, and that the components are found to exhibit a long run relationship. In all, 204 observations are used. Data were partitioned into two. The first 70% of the data is used as training data for developing the model while the remaining proportion is used for modelling validation.

The study reveals that the highest and lowest mean prices were recorded for Gasoline and Gas Oil prices, and Kerosene price respectively. The highest price over the period was 177 and 175, being recorded by Gasoline and Gas Oil prices whiles the least was 6.47 and recorded by Kerosene price. This means that Gasoline and Gas Oil is the most expensive petroleum products, whiles Kerosene is the least. The time series plot of the various products shows that they all have trends and hence, are non-stationary. In particular, the plot shows a mixed relationship between them. Furthermore, the unit root test results formally confirmed the earlier conclusions that the series are non-stationary at a level using both the multivariate and univariate Unit Root Tests (URTs). However, they are all stationary after the first difference. Hence, multivariate time series models (VEC models) are fitted to examine these dynamic behaviours of the products. The lag length selection criteria reveal a lag order of 1, hence, a VEC model is fitted with a lag of 1. Since all the products are I(1) and cointegrated of order 1, both short and long-run dynamic behaviours are studied. Thus, it is demonstrated that though the products are non-stationary at the individual level, it is possible to obtain a linear combination of them that achieves stationarity. Hence, the Trace cointegration test indicates that there is at least one cointegrating relationship among the products. Thus, there exists a long-run relationship among the products. Having achieved cointegration among the products, we estimate the long and short-run relationships. Several diagnostics and validation techniques were performed on the models to determine their adequacy. The tests revealed that the models are free from serial correlation.

Specifically, the study revealed that there are dynamic relationships among the prices of the products. Also, the models are seen to be dynamically stable. Thus, the deviations of the models in previous periods are corrected in the long run. The magnitudes suggest for instance that the imbalances in the prices are corrected every two weeks. The short-run results also indicate that some of the products influence others (Gasoline, Gas Oil, and LPG prices). The consequence of this result is that in the short run, a rise in the price of a product will consequently lead to a growth in the product in question own price, and prices of the others, especially, Gasoline and Gas Oil prices. These relationships are confirmed by the IRFs, FEVD, Granger causality analyses, as well as the forecasts, which indicated that in the short term, changes in Gasoline and Gas Oil prices are accounted for by Gasoline, but in the medium to long terms, the variability is shared between Gasoline and Gas Oil prices, with Gas Oil slightly above Gasoline price in that regard. Furthermore, variation in Kerosene price is accounted for by itself, but with Gasoline price increasing to some extent in the medium to long terms. Finally, even though the variability explained by LPG price by itself decreases generally, none of the other products meaningfully explained its variability well. Thus, in the short term, changes in prices of the products account for changes in their prices and to some extent, prices of others, but in the medium to short term, these changes are not much more felt by the products. The consequence of this result is that increase ex-pump prices of one or more products are likely to influence others, as the coefficients of determinations (r-squared), which describe the predictive ability of the models are between 53.86% and 23.15%. That is, the coefficients of determination for

Gasoline, Gas Oil, Kerosene, and LPG ex-pump prices are 53.86%, 41.06%, 26.8%, and 23.15% respectively. It is observed that these performance statistics are in sharp contrast to the good fit of the model as the model fits the data closely and better than the fit obtained from the corresponding VAR. The VAR model obtains coefficients of determination values ranging from 90% to 96%, which points to the practical performance of the VEC model. The inconsistency in the performance of the VEC is worth investigating.

This supports the fact that prices of petroleum products appear to be been moving up most of the time. High-pitched surges in the prices of oil are generally recognized to have significant special effects on economic and macroeconomic policies. In particular, the very recent increases in prices in the world oil market are causing concerns about possible slowdowns in the economic performance of most industrialized countries. Consequently, not shockingly, a substantial body of economic research has studied the conduits through which oil price shocks affect economic variables. All other things being equal, these products are a source of energy in the country, and energy is the backbone of every economy, and increases in these products affect the economy as a whole. In a nutshell, changes in the price of the products have effects on themselves and similar products, and since the economy evolves around energy (petroleum products), they have direct effects on the economy and people at large.

Conclusions

Based on the summary of findings and discussion, some conclusions can be drawn. The study reveals that there exists a general upward trend in the ex-pump prices of the products over the period.

It also shows that changes in prices of some of the products influence others in the short run, but in the medium to long terms, stability is attained, with Premium Gasoline and Gas Oil prices accounting for most of the variations in changes in prices. The granger causality, IRF, and FEVD analyses all affirm this result.

A VEC model of order 1 is found suitable for the data. The model means, in particular, that the most recent past prices are what is most influential in determining current prices. The model appears to perform better than the corresponding VAR model with the least R-square of 90.3% for Premium Gasoline prices and the highest of 96% for Gas Oil prices. The results point to the competitive nature of prices in both Premium Gasoline and Gas Oil which reflect the competitive usage of the two products.

Recommendations

The study shows that prices of Gasoline, in particular, account for a large variation in prices of other petroleum products. It means that Gasoline could be used for various purposes and could be a suitable substitute for other products in Ghana. To effectively regulate the prices of petroleum products and the resultant shocks to the economy, it would be prudent to monitor and control the prices of Gasoline.

The study has also obtained a VEC model and also presented the corresponding VAR model, both of order 1. These models could serve as an important guide for policy decisions. In particular, most recent past prices are most influential in determining current prices.

Further studies could be carried out on the demand for these products over time to appropriately examine the relationship between them, and the periods within which such relationships are significant.

The study used biweekly data, obtained from the NPA website. The data spans January 2007 to June 2015. It is therefore recommended that an updated data is obtained so that the current trend in the ex-pump prices of these petroleum products can be studied for effective policies decisions.

The performance statistics of the VEC model do not reflect its practical significance since it fit the model better than the VAR model. This discrepancy is worth studying.

REFERENCES

Abeysinghe, T. (2001). Estimation of the direct and indirect impact of oil price on growth. *Economics letters*, *73*(2), 147-153.

African Developing Bank [ADB] (2008). African Developing Bank Annual Report: ADB.

Agyen, K. J. (2012). Time Series Modeling and Forecasting of the Demand of Some Petroleum Products in Ghana. Retrieved from Dspace.knust: http://dspace.knust.edu.gh/bitstream/123456789/4586/1/Agyen%20K.J.pdf accessed on 22 May, 2019.

Ahn, S. K., & Reinsel, G. C. (1990). Estimation for partially non-stationary multivariate autoregressive models. *Journal of the American statistical association*, 85(411), 813-823.

Akaike, H. (1974). A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control*, 19(6), 716-723...

Akide, A. (2007). Growth implications of oil price variations. A case study of Nigeria, 8(2), pp.20-27.

Apergis, N., & Rezitis, A. (2003b). Agricultural price volatility spillover effects: the case of Greece. *European Review of Agricultural Economics*, 30(3), 389-406.

Aqeel, A., & Butt, M. S. (2001). The relationship between energy consumption and economic growth in Pakistan. *Asia-Pacific Development Journal*, *8*(2), 101-110.

Armah, B. (2003). Contribution of Energy to the Ghanaian Economy. Economic Analysis of the Energy Sector, Ghana.

Bak, N. (2004). Completing your thesis: A practical guide. Pretoria: Van Schaik Publishers.

Balke, N. S., & Wynne, M. A. (1996). Are deep recessions followed by strong recoveries? Results for the G-7 countries. *Applied Economics*, 28(7), 889-897.

Barnhill, T. M., Joutz, F. L., & Maxwell, W. F. (2000). Factors affecting the yields on noninvestment grade bond indices: a cointegration analysis. *Journal of Empirical Finance*, 7(1), 57-86.

Bernanke, B. S. (1983). "Irreversibility, Uncertainty, and Cyclical Investment." *The Quarterly Journal of Economics*, 98(1), 85–106. From https://doi.org /10.2307/1885568

Brown, S. P. A., & Yücel, M. K. (1999). Federal Reserve Bank of Dallas, Economic and Financial Review.

Chamber of Mines (2008). Chamber of Mines Annual Report: Chamber of Mines.

Chang, Y., & Wong, J. F. (2003). Oil price fluctuations and Singapore economy. *Energy policy*, 31(11), 1151-1165.

Cryer, J. D. (1986). *Time series analysis* (Vol. 286). Boston: Duxbury Press.

Darby, M. (1982). The price of oil and world inflation and recession. *The American Economic Review*, 72, 738-751.

Davis, S. J., & Haltiwanger, J. (2001). "Sectorial Job Creation and Destluction Responses to Oil Price Changes." *Journal of Monetary Economics*, 48, 465-512.

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the Estimators for Autoregressive Time Series with a Unit-root. *Journal of the American Statistical Association*, 74, 427-431.

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.

Ebohon, O. J. (1996). "Energy, Economic Growth and Causality in Developing Countries: A case Study of Tanzania and Nigeria". *Energy Policy*, 24, 447–453.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society*, 987-1007.

Engle, R. F., & Yoo, B. S. (1987). Forecasting and testing in co-integrated systems. *Journal of econometrics*, *35*(1), 143-159.

Engle, R. F., & Granger, C. W. J. (1987). Co-integration and Error Correction Representation, Estimation and Testing, *Econometrica*, 55, pp. 251 – 276.

Entner, D., & Hoyer, P. O. (2010). On causal discovery from time series data using FCI. *Probabilistic graphical models*, 121-128.

Erdogdu, E. (2007). Electricity demand analysis using cointegration and ARIMA modelling: A case study of Turkey. *Energy policy*, 35(2), 1129-1146.

Farzanegan, M. R., & Markwardt, G. (2009). The effects of oil price shocks on the Iranian economy. *Energy Economics*, 31(1), 134-151.

Glasure, Y. U., & Lee, A. R. (1997). Cointegration, error correction and the relationship between GDP and energy: The case of South Korea and Singapore, *Resource and Energy Economics*, 20, 17–25.

Granger, C. W. & Newbold, P. (1974). Experience with forecasting univariate time series and the combination of forecasts. *Journal of the Royal Statistical Society: Series A (General)*, 137(2), 131-146.

Granger, C. W. J. (1969). "Investigating Causal Relations by Econometric Models and Cross Spectral Methods," *Econometrica*, 37, 424-438.

Hamilton, J. D. (1983). Oil and the macroeconomy since World War II. J. Polit. *Econ*. 92(2), 228-248.

Hamilton, J. D. (1996). Specification testing in Markov-switching time-series models. *Journal of econometrics*, 70(1), 127-157.

Hamilton, J. D., & Herrera, A. M. (2004). Comment: oil shocks and aggregate macroeconomic behavior: the role of monetary policy. *Journal of Money, credit and Banking*, 265-286.

Hannan, E., & Quinn, B. (1979). The Determination of the Order of an Autoregression. *J. Roy Statist. Soc. Ser.* B 41, 190–195.

Harvey, A. C., & Peters, S. (1984). Estimation procedures for structural time-series models. Discussion Paper No. A28, London School of Economics.

Hillmer, S. C., & Tiao, G. C. (1979). Likelihood Function of Stationary Multiple Autoregressive Moving Average Models. *Journal of the American Statistical Association*, 74, 652-660.

Hosking, J. (1981). Fractional Differencing. *Biometrika*, 68(1), 165–167.

Hsiao, C. (2003). Analysis of Panel Data. 2nd ed. New York: Cambridge University Press.

Hylleberg, S., Engle, R. F., Granger, C. W. J., & Yoo, B. S. (1990). Seasonal integration and cointegration, *Journal of Econometrxa*, 44, 215 238.

Im, K. S., Pesaran, M. H., & Shin, Y. (2003). Testing for unit roots in heterogeneous panels. *Journal of econometrics*, 115(1), 53-74. http://dx.doi.org/10.1016/S0304-4076 (03)00092-7

International Energy Agency Staff. (2012). *Energy Statistics of Non-OECD Countries*. Organization for Economic.

Ivanov, V., & Kilian, L. (2005). A practitioner's Guide to Lag Order Selection For VAR Impulse Response Analysis. *Studies in Nonlinear Dynamics & Econometrics*, 9(1), 1-36.

Jin, G. (2008). “The Impact of Oil Price Shock and Exchange Rate Volatility on Economic Growth: A Comparative Analysis for Russia Japan and China”, *Research Journal of International Studies*, 8, 98-111.

Johansen, S. (1991). Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models. *Econometrica*, 59, 1551–1580.

Johansen, S. (1995). Likelihood-based Inference in Cointegrated Vector Autoregressive Models, Oxford: Oxford University Press.

Jumah, A., & Pastuszyn, G. (2007). *Oil price shocks, monetary policy and aggregate demand in Ghana* (No. 212). Reihe Ökonomie/Economics Series.

Kumar, U., & Jain, V. K. (2010). Time series models (Grey-Markov, Grey Model with rolling mechanism and singular spectrum analysis) to forecast energy consumption in India. *Energy*, *35*(4), 1709-1716.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., & Shin, Y. (1992). Testing the Null Hypothesis of Stationarity against the Alternative of a Unit-root; How Sure are we that Economic Time Series have a unit-root? *Journal of Econometrics*, 54, 159-178.

Laser, Y. (1987). Interest, Inflation, Growth and the Direction of the Hong-Kong Economy, *Chinese Economic Reviews*, 120(8), 19-33.

Lee, K. (1998). Oil Price Changes and Volatility; A Correlation Analysis on the Economy of China, *Scholarly writers' publications,* 15(4), 44-49.

Lee, K., & Ni, S. (2002). On the dynamic effects of oil price shocks: a study using industry level data. *Journal of Monetary Economics*, 49(4), 823-852.

Levin, A. & Lin, C. F. (1992). `Unit Root Tests in Panel Data: Asymptotic and Finite Sample Properties, Department of Economics, University of California at San Diego, Discussion Paper No. 92-93 (revised 1993).

Levin, A. & Lin, C. F. (1993). Unit Root Tests in Panel Data: New Results', Blackwell Publishers 1999 Department of Economics, University of California at San Diego, Discussion Paper No. 93±56.

Levin, A., Lin, C. F., & Chu, C. S. J. (2002). Unit root tests in panel data: asymptotic and finite-sample properties. *Journal of econometrics, 108*(1), 1-24.

Liu, L. M., & Chen, C. (1991). Recent developments of time series analysis in environmental impact studies. *Journal of Environmental Science & Health Part A*, *26*(7), 1217-1252.

Lon-Mu, L. (2006). Time series analysis and forecasting [M]. Second Edition. U.S.A.: Scientific Computing Associates Corp, 2006, pp. 1-565

Loungani, P. (1986). "Oil Price Shocks and the Dispersion Hypothesis". *Review of Economics and Statistics*, 68, 536-539

Maddala, G. S., & Wu, S. (1999). A comparative study of unit root tests with panel data and a new simple test. *Oxford Bulletin of Economics and Statistics*, 61(1), 631-652. From: http://dx.doi.org/10.1111/1468-0084.61.S.13

McCleary, R., Hay, R. A., Meidinger, E. E., & McDowall, D. (1980). *Applied time series analysis for the social sciences*, 331. Beverly Hills, CA: Sage Publications.

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

Mulugeta, S. K, Nondo, C., Schaeffer, P. V., & Gebremedhin, T. G. (2010). "Does the level of income matter in the energy Consumption and GDP

Nexus: Evidence from Sub-saharan African Countries." Research Paper#7, Research Institute, West Virginia University.

Noreng, Q. (2007). *Crude Power, Politics and the Oil Market*, I. B. Tauris, and Co. Ltd, London, and New York, pp. 245-704.

National Petroleum Authority (2019). National Petroleum Authority. Retrieved from: http://npa.gov.gh/npa_new/index.php.

Odoi, B., Twumasi-Ankrah, S., & Al-Hassan, S. (2017). "Modeling Price Volatility in Petroleum Products in Ghana". *Ghana Journal of Technology*, 1(2), 40-44.

Pao, H. T. (2009). Forecasting energy consumption in Taiwan using hybrid nonlinear models. *Energy*, 34(10), 1438-1446.

Papapetrou, E. (2001). Oil price shocks, stock market, economic activity, and employment in Greece. *Energy Economics*, *23*(5), 511-532.

Perron, P. (1988). Trends and random walks in macroeconomic time series: Further evidence from a new approach. *Journal of economic dynamics and control*, *12*(2-3), 297-332.

Phillips, P. C. (1987). *Multiple regression with integrated time series* (No. 852). Cowles Foundation for Research in Economics, Yale University.

Phillips, P.C.B, & Perron, P. (1988). Testing for a Unit Root in Time Series Regressions. *Biometrika* 75, 335-346.

Pindyck, R. S., & Rotemberg, J. J. (1983). Dynamic Factor Demands and the Effects of Energy Price Shocks. *American Economic Review*, 73, 1066–79.

Pradhan, R. P., & Bagchi, T. P. (2013). Effect of transportation infrastructure on economic growth in India: The VECM approach. *Research in Transportation Economics*, 38, 139‐148.

Rafiq, S., Salim, R., & Bloch, H. (2009). Impact of crude oil price volatility on economic activities: An empirical investigation in the Thai economy. *Resources Policy*, *34*(3), 121-132.

Reinsel, G. C., & Ahn, S. K. (1992). Vector autoregressive models with unit roots and reduced rank structure: Estimation. Likelihood ratio test, and forecasting. *Journal of time series analysis*, *13*(4), 353-375.

Roger, F. (1998). “Energy Use in the United Kingdom”. *The Energy Journal*, Vol. 19(4), 1-42.

Sadorsky, P. (1999). Oil price shocks and stock market activity. *Energy Economics*, 21, 449-469.

Salamanca, F., Krpo, A., Martilli, A., & Clappier, A. (2010). A new building energy model coupled with an urban canopy parameterization for urban climate simulations-part I. formulation, verification, and sensitivity analysis of the model. *Theoretical and applied climatology*, *99*(3-4), 331.

Sims, C. A, (1980). Macroeconomics and Reality." *Econometrica*, 48, 1-48.

Tetteh, A., & Xu, Q. (2015). Forecasting Refined Petroleum Products Prices in Ghana. *Journal of Economics, Management, and Trade*, 94-107.

Tweneboah, G. & Adam, A. M., (2008). Macroeconomic factors and stock market movement: Evidence from Ghana. *Available at SSRN 1289842*.

Van Soest, D. P., Kuper, G. H., & Jacobs, J. (2000). Threshold Effects of Energy Price Changes. University of Groningen. Working Paper.

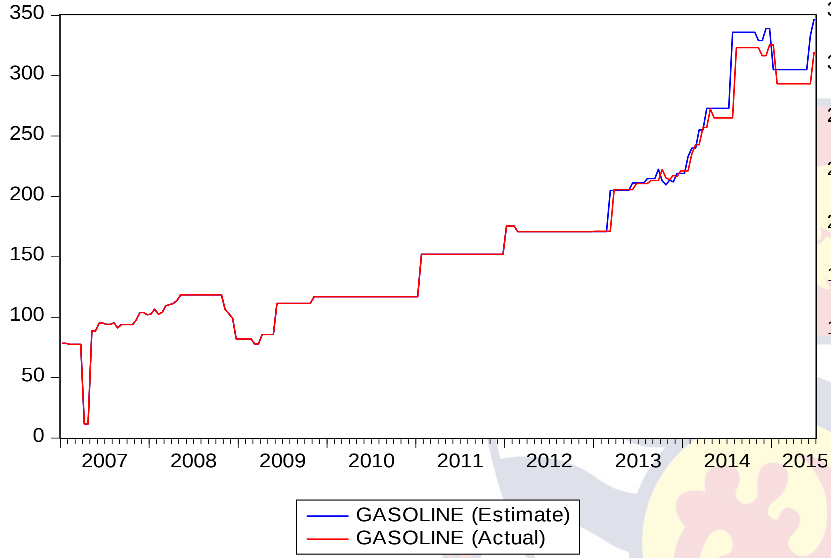
APPENDICES

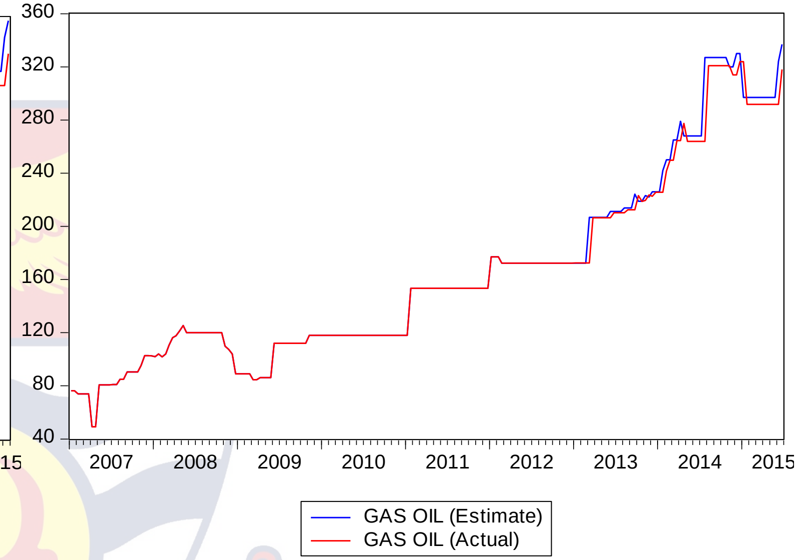
APPENDIX A: Model Fit of Var(1) Model

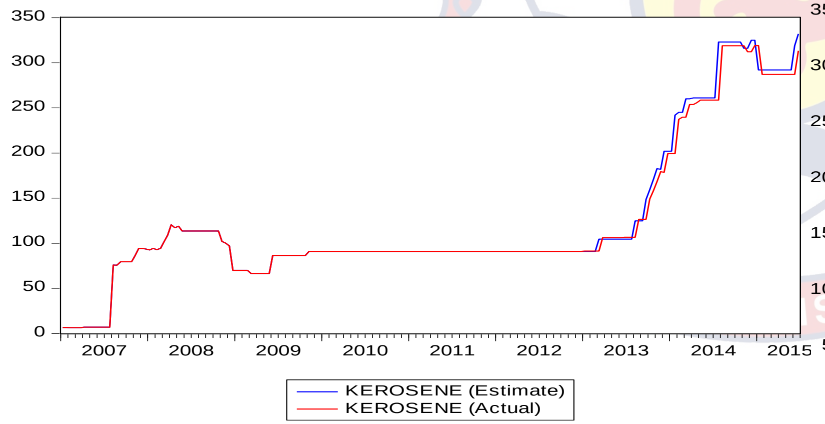
Parameters	Coefficient	S.E	t-statistic
Gasoline Model:			
$(GASOLINE)_{t-1}$	0.321	0.142	2.256*
$(GASOIL)_{t-1}$	0.643	0.155	4.142*
$(KEROSENE)_{t-1}$	-0.709	0.796	-0.890
$(LPG)_{t-1}$	0.048	0.097	0.498
Constant	6.1479	5.772	1.065
Gasoil Model:			
$(GASOLINE)_{t-1}$	-0.237	0.090	-2.662*
$(GASOIL)_{t-1}$	1.222	0.097	12.565*
$(KEROSENE)_{t-1}$	-0.463	0.499	-0.928
$(LPG)_{t-1}$	0.018	0.061	0.299
Constant	4.758	3.614	1.317
Kerosene Model:			
$(GASOLINE)_{t-1}$	0.006	0.005	1.143
$(GASOIL)_{t-1}$	-0.009	0.006	-1.454
$(KEROSENE)_{t-1}$	0.938	0.031	30.501*
$(LPG)_{t-1}$	0.005	0.004	1.348
Constant	0.405	0.223	1.816
LPG Model:			
$(GASOLINE)_{t-1}$	0.032	0.069	0.467
$(GASOIL)_{t-1}$	-0.011	0.075	-0.148
$(KEROSENE)_{t-1}$	-0.494	0.385	-1.286
$(LPG)_{t-1}$	0.975	0.047	20.832*
Constant	4.693	2.788	1.684

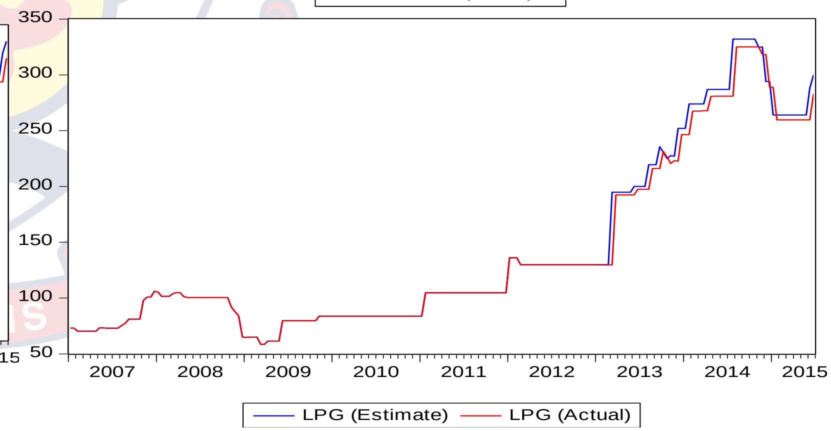
Reference: Researcher's Computations (2020) **significant coefficients*

APPENDIX B: Model Fit Graphs of Var(1) Model

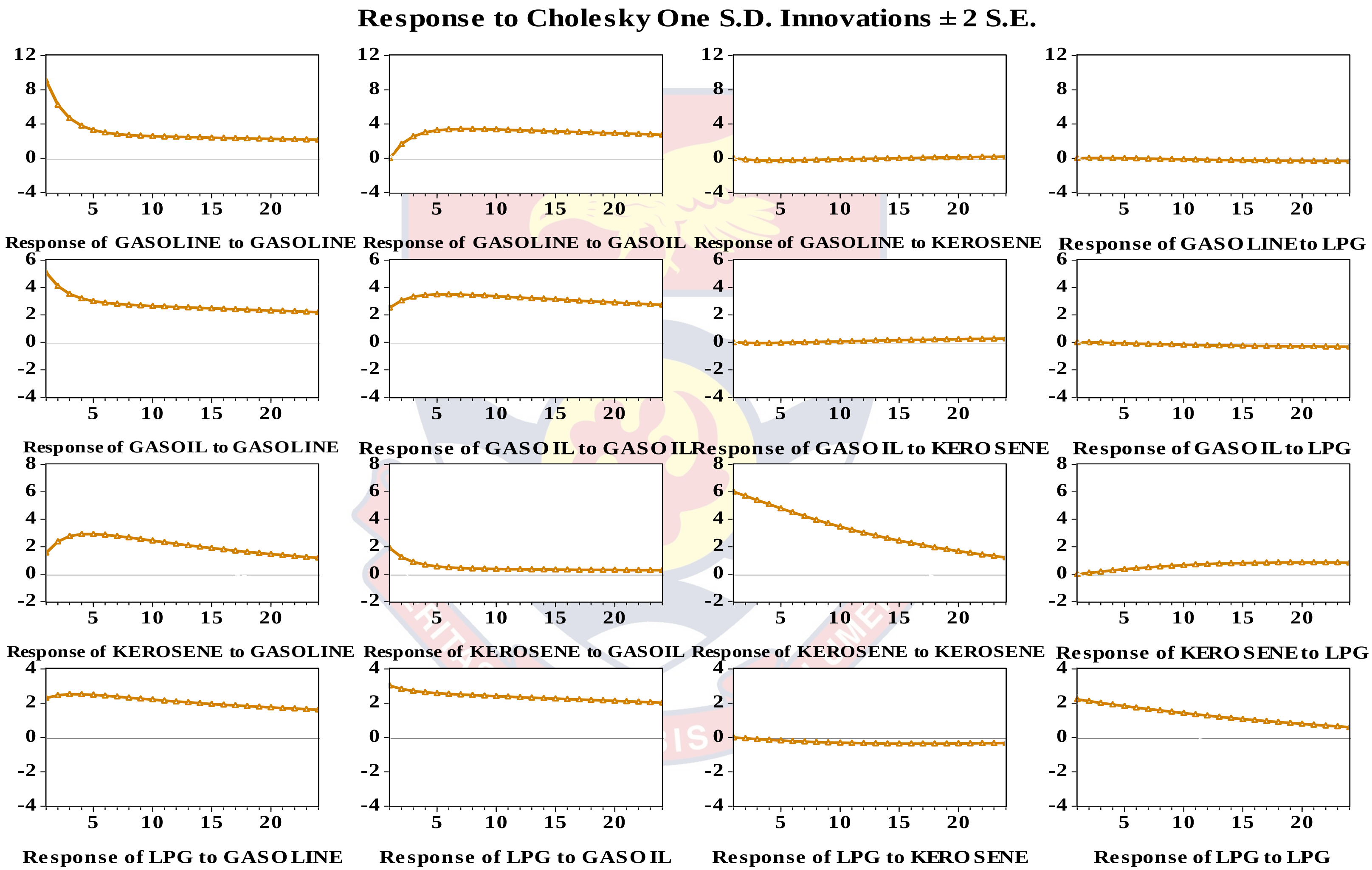






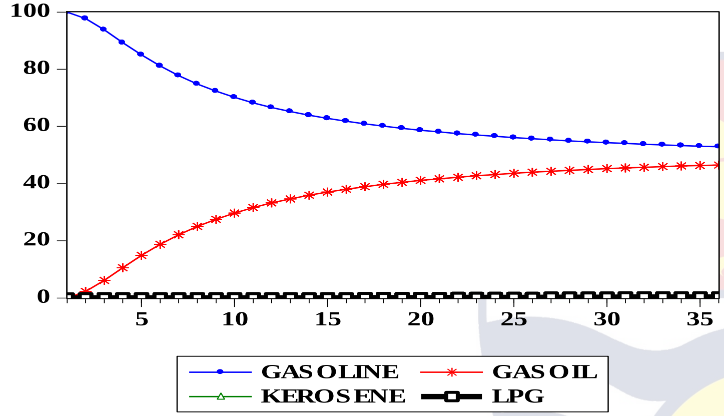


APPENDIX C: Var(1) Impulse Response Functions (IRF)

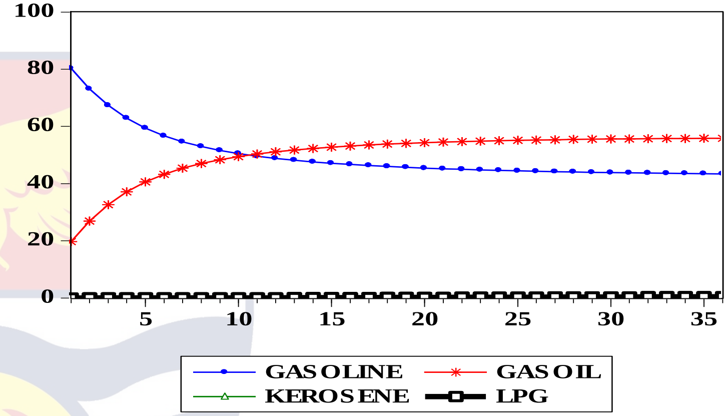


APPENDIX D: Var(1) Forecast Error Variance Decomposition (FEVD)

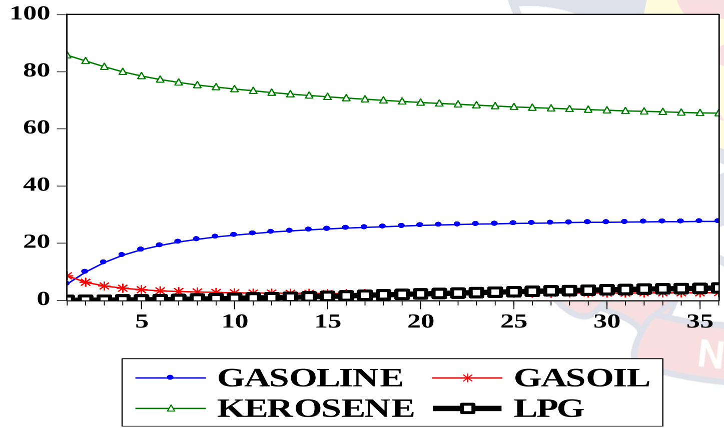
FEVD OF GASOLINE



FEVD of GASOIL



FEVD OF KEROSENE



FEVD OF LPG