Determining the Key Factors Affecting Global Pricing of Crude Oil Using Trend Analysis and Numerical Modelling*

[1]E. M. Amarfio, [1]M. Gyagri and [1]S. A. Marfo
[1]University of Mines and Technology, P. O. Box 237, Tarkwa, Ghana

Amarfio, E. M., Gyagri M., and Marfo, S. A. (2017), "Determining the Key Factors Affecting Global Pricing of Crude Oil Using Trend Analysis and Numerical Modelling" *Ghana Journal of Technology,* Vol. 2, No. 1, pp. 82 - 90.

Abstract

Crude Oil affects almost all activities of our modern day hydrocarbon society. The ever increasing importance and demand of oil globally, has led to its highly complex market and pricing system. This paper investigates the factors that determine the prices of crude oil and the impact of these individual factors using regression and trend analysis. Crude oil prices along with other factors such as global and Oil Production and Exporting Countries (OPEC's) production rates, global oil consumption and refinery rates where studied from 1965 to 2015. The analysis employed a log-log multiple regression method using ordinary least squares, with oil price as the response variable and the other factors mentioned as predictor variables. The results from the regression analysis were complemented with trend analysis of these factors with oil prices. It was revealed from the results that all explanatory variables examined save natural gas consumption and wars were found to have significant impact in the oil price determination. Global oil and OPEC's production rates have a negative (inverse) relationship with oil price while consumption, the presence of wars directly relate to oil prices. The study established that crude oil prices change over time cannot be attributed to one single market factor, or just the explanatory factors analyzed in this work.

Keywords: Complex Market, Crude Pricing System, Multiple Regression, Analysis, Predictor Variables

1 Introduction

The role of crude oil over the past two centuries for energy generation which is instrumental in the industrial, domestic, transportation, economic sector and to a large extent almost all activities of the modern day hydrocarbon society as a source of nonrenewable energy has undoubtedly been phenomenal. According to the BP statistical review of world energy in June 2016, oil remained the world's leading fuel, accounting for 32.9 % of global energy consumption (Anon, 2016a)

Crude oil market began around the 1860s following the successful drilling of the first oil well by Colonel Edwin Drake in 1859 at Oil Creek near Titusville in West Pennsylvania. Since then, the importance of oil has been on the ascendency (Yergin, 1991). The ever increasing demand and emergence of crude oil as a highly sought after global commodity has accounted for its market complexity and price volatility. Various researches have suggested a number of entities including oil production rates, the Organisation of Petroleum Exporting Countries (OPEC), speculators and large oil companies as determinants of crude oil prices. Other factors such as the constraints to the access of resources, increasing demand driven by economic and industrial growth, geopolitical and economic events, net importing oil countries, value of the dollar, other sources of energy, weather and natural disasters are also attributed to the pricing of the crude oil (King *et al*., 2011).

Because crude oil plays a major role in almost all modern day activities (transportation costs, prices of goods and services, energy provision etc.), there is the need to identify and understand the key factors influencing its pricing. This will aid governments, organisations and individuals in their planning, decision making and forecasting processes.

For this reason, this paper tries to identify the influential factors affecting the pricing of crude oil through the employment of trend analysis and multiple regression method.

2 Resources and Methods

2.1 Modelling Software

Minitab, a statistical analysis software package developed at the Pennsylvania State University in 1972 is very useful in statistical learning and research. It was therefore used in the generation of the regression models between oil price and its determinants. It was selected because of its high accuracy, reliability and user friendliness. It also provides a wide range of computation and graphical applications.

2.2 Data Acquisition

Secondary data on some of the factors outlined in the previous section were collected from the BP Statistical Review of World Energy 2016 workbook with the exception of oil wars that were obtained from Oil Wars and 1973 Oil Crisis (Anon, 2017a; Anon, 2017b). The choice of the sample size (period) was to determine the general influence of these factors on oil price in relatively recent years (50 years). Table 1 shows the sample period of the data collected and its units.

Table 1 Sample Period of Variables

Variable	Time Period
Spot Crude Oil Price of the day ($/bbl)	1965 - 2015
Crude Oil Production (Mbbl/d)	1965 - 2015
Crude Oil Consumption (Mbbl/d)	1965 - 2015
OPEC Production (Mbbl/d)	1965 - 2015
Natural Gas Consumption (Bcf/day)	1965 - 2015
Refinery Capacity (Mbbl/d)	1965 - 2015
Oil Wars	1965 - 2015

2.3 Methods

2.3.1 Regression Analysis

A Logarithmic Multiple Regression using Ordinary Least Squares (OLS) method was employed with oil price as the dependent variable and the other variables shown in Table 2 as independent variables. Least square approach was used to determine a line of best fit by minimising the sum of squares created by a mathematical function. A "square" is determined by squaring the distance between a data point and the regression line. The least square approach limits the distance between a function and the data points that a function is trying to explain (Ray, 2015).

Dependent Variable

The dependent (response) variable is global crude oil price. These prices represent the average annual money of the day prices of Arabian Light posted as Ras Tanura from 1965 – 1983 and Brent dated from 1984 – 2015, traded in U.S dollars per barrel ($/bbl.) (Anon, 2016b). The dependent Variable is denoted as (p).

Independent Variables

The independent variables of the regression equation were;

(i) Oil Production: The global yearly production of crude oil (pr) measured in a thousand barrels per day (Mbbl/d). The averaged values of production include crude oil, shale oil, oil sands NGLs (natural gas liquids - the liquid content of natural gas) and excludes liquid fuels from other sources such as biomass and derivatives of coal and natural gas.

(ii) Oil Consumption: Data collected on global crude oil consumption rates (cn) measured in thousand barrels per day (Mbbl/d). Consumption values include inland demand plus international aviation and marine bunkers and refinery fuel and fuel loss. Consumption of bio gasoline such as ethanol, biodiesel and derivatives of coal and natural gas are also included.

(iii) Natural Gas Consumption: Includes derivatives of coal and natural gas consumed in gas-to-liquid transformations, measured in billion cubic feet per day (Bcf/d). Represented as (ngc) in further analysis.

(iv) Refinery Capacity: To evaluate the effects of the refinery sector on crude oil prices, Data was collected on refinery capacity (rfc) measured in a thousand barrels per day (Mbbl/d).

(v) Data values for oil production, consumption, natural gas consumption and refinery capacity exclude that from Estonia, Latvia and Lithuania prior to 1985 and Slovenia before 1990.

(vi) OPEC Production: OPEC's influence is measured by production rates from its member countries. Denoted as (opd) in units of a thousand barrels of oil per day (Mbbl/d).

(vii) Oil Wars: War and conflict related to petroleum over the years is represented as a dummy variable. With (w_1) representing the presence of war.

Model Specification

The log-log multiple regression model specification applies; it is expressed mathematically as:

$$\ln(p_t) = \alpha + \sum_{i=1}^{6} \beta_i \ln(Xt_{ti}) + \varepsilon_t \qquad (1)$$

Where;
p_t = dependent variable changing with time
α = intercept
β_i = coefficients of predictor variables
X_{ti} = predictor variables changing with time
ε_t = error term

Hypotheses

The analysis of this work was done based on the following hypotheses. Where H_1: presents the research hypothesis and H_0: the null hypothesis. The discussion of the hypothesis of the various variables is shown in the Table 2.

Table 2 Summary of Hypotheses of Variables

Variable	Hypotheses
pr	H_1: Production is a determinant of crude oil price H_0: Production is not a determinant of crude oil price
cn	H_1: Consumption is a determinant of crude oil price H_0: Consumption is not a determinant of crude oil price
opd	H_1: OPEC's production influences oil price H_0: OPEC's production does not influence oil price
ngc	H_1: Natural gas consumption affects oil prices H_0: Natural gas consumption does not affect oil prices
rfc	H_1: Refinery capacities influence oil prices H_0: Refinery capacities do not influence oil prices
w_1	H_1: War affects crude prices H_0: War does not affect crude prices

2.4 Model Selection Criteria

Several models with different combinations of independent variables were run to generate the most appropriate and efficient model. The selection of the most appropriate method was based on the following limitations and goodness of fit statistics associated with multiple regression.

2.4.1 Limitations of Multiple Regression

Some limitations of multiple regression include:

Multicollinearity: Multiple regression requires that the model generated possess no or very little multicollinearity. This phenomenon occurs when the independent variables are not only potentially related to the dependent variable but are also potentially related to each other (Ray, 2015). Variables that are highly correlated were not combined in the same model in order to reduce redundancy and unstable coefficient estimates caused by multicollinearity.

Overfitting: Adding more independent variables to a multiple regression procedure does not mean the regression will be better or offer better predictions. It can lead to worse situations and this is called Overfitting (Ray, 2015).

2.4.2 Goodness of Fit Statistics

The following statistical characteristics were considered in selecting the right model:

F Test: It compares a model with no predictors (intercept-only model) to the model specified. The hypotheses for the F-test of the overall significance are as follows:

H0: The fit of the intercept-only model and one's model are equal.

H1: The fit of the intercept-only model is significantly reduced compared to one's model.

If the P value for the F-test of overall significance test is less than the most commonly used significance level (0.05), the null-hypothesis can be rejected and it can be concluded that the model generated provides a better fit than the intercept-only model.

T Test: This test compares the data of a variable to its null hypothesis by generating a t value. If the probability (p-value) of the t value falling within the rejection region of a T distribution is fairly low, the null hypothesis can be rejected using the common significance level of 0.05. Thus a low p-value (< 0.05) indicates the predictor variable has a meaningful impact on the response variable.

Standard Error of Regression (S): This statistic provides an overall measure of how well the model fits a data. It represents the average distance that the observed values fall from the regression line. Conveniently, it tells how wrong a regression model is on average using the units of the response variable. Smaller values are better because it indicates that the observations are closer to the fitted line.

R Squared: Measures closeness of the to the fitted regression line. It is the percentage of the response variable variation that is explained by a linear model. It is expressed mathematically as:

$$R^2 = \frac{\text{Explained variation}}{\text{Total variation}} \tag{2}$$

Generally, a higher R-squared, implies a better model fit. It is always between 0 and 100 %.

R Squared Adjusted: The adjusted R-squared compares the explanatory power of regression models that contain different numbers of predictors. The adjusted R-squared is a modified version of R-squared adjusted for the number of predictors in a model. The adjusted R-squared increases only if the new term improves the model more than would be expected by chance. It decreases when a predictor improves the model by less than expected by chance. It is always lower than the R-squared.

Residual Plots: In general, a model fits the data well if the differences between the observed values and the model's predicted values are small and unbiased. Residual plots should be checked before the analysis of goodness of fit statistics as they can reveal unwanted residual patterns that indicate biased results more effectively than numbers. In the Ordinary Least Square (OLS) context, random errors are assumed to produce residuals that are normally distributed. Therefore, a plot of the residual histogram should be bell-shaped or symmetrical, with a random residual versus fits plot. The normal probability plot of the residuals should also have a value greater than 0.05 which implies that the residuals are normally distributed (Frost, 2013).

2.4 Final Model Generated

Production and refinery capacity were excluded from the final regression model due to issues of multi-collinearity and overfitting. Production, consumption and refinery capacity were highly correlated (strong relationship amongst these variables). Table 3 shows the Pearson's correlation coefficient (R coefficient) for production, refinery capacity and consumption.

Table 3 Pearson's Correlation Coefficients

	lnpr	***lncn***	***lnrfc***
lnpr	1.0000		
lncn	0.9958	1.0000	
lnrfc	0.9666	0.9624	1.0000

3 Results and Discussion

3.1 Regression Model Generated

Based on the model selection process, the regression Equation 3 was developed;

$$lnp = -46.6 + 6.16\, lncn - 0.586\, lngc - 1.55\, lnopd + 0.029\, w_{-1} \quad (3)$$

Table 4 Model Statistics

Variable	Coefficient	T Value	P Value
lnpr	6.159	3.80	0.000***
lnngc	-0.5864	-0.68	0.501
lnopd	-1.5458	-2.90	0.006***
W_1	0.0291	0.11	0.911
α	-46.6	-4.17	0.000***
R^2 83.7% R^2 (adj) 82.3%		S 0.507131 F 59.24 (0.000***)	

Where ***, indicates the factors are significant at 1 % significance level.

3.2 Global Crude Oil Consumption

At a 1 % significance level, consumption rates are confirmed as a significant determinant of crude oil prices. With a coefficient of 6.159, it indicates that an increase in the consumption rates of crude oil by 1 % leads to an average increase in oil prices by 6.159 % all other factors held constant. This confirms with economic theory where an increase in the demand for a good drive it prices upward (demand and prices are directly related). According to Möbert (2007), the upward trend of oil prices at the spot market are as a result of increasing demand in emerging markets. The trend of oil price versus consumption (Fig. 1) shows a general increasing trend for both variables with more fluctuations in price which may be due to other determinants aside consumption causing price volatility.

3.3 Natural Gas Consumption

A substitute product is one that causes an increase in demand for another product when its price increases (Nicholson and Snyder, 2011). Natural gas and crude oil are relative substitutes as they can be used for similar energy provision activities. The price of crude oil and natural gas consumption are thus expected to have an inverse relationship as more gas consumption will mean reducing oil demand and consequently decreasing price. This is evident in the negative coefficient (-0.5864) of ngc in the regression equation. The null hypothesis which states gas consumption has no impact on crude prices cannot be rejected because, (p value is 0.501, > 0.05). Therefore, natural gas does not have a meaningful impact on global crude oil price. This is confirmed by Seth, (2015) who reported that gas and crude oil are close substitutes for each other in regions where both are supported by technology, infrastructure and markets.

3.4 OPEC's Production

OPEC's production has been statistically proven as a vital contributor in the determination of oil prices with a corresponding p value of 0.006. A rise in OPEC's production by 1 % implies an average price reduction by 1.5458 %. After applying a Vector Error Correction Method (VECM), King *et al*. in 2011, observed that oil price moved in opposite direction from OPEC deviations and quotas which represent the sum of individual OPEC production allocations and the difference between total OPEC production and quota respectively measured in millions of barrels per day.

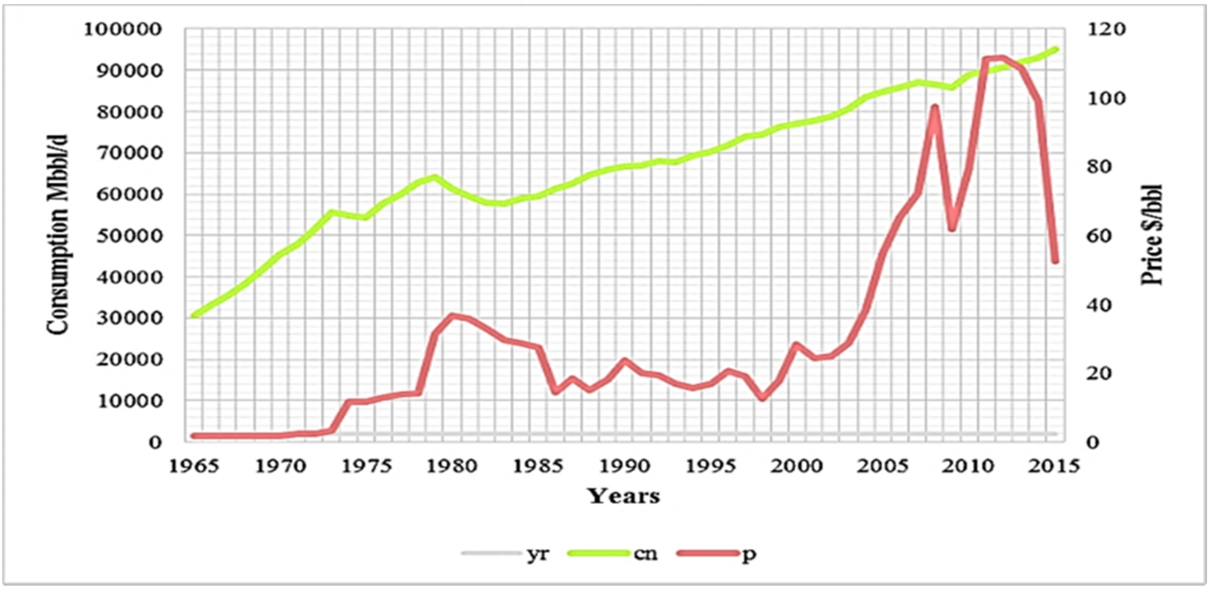


Fig. 1 Crude Oil Consumption against Prices

In 1973, OPEC raised the posted oil price by 70 % to $ 5.11 a barrel by gradually increasing production cut levels by 5 % per month (Anon, 2017b).

Fig. 2 monitors OPEC's impact on prices during the 1970's where it held its highest market dominance. OPEC's gradual plateau and decrease in production from 1972 to 1975 caused a sharp spike in oil prices. Prices became fairly stable when production rates began to increase after 1975 and the sharp price increase from 1979 responded to declining production rates from 1977.

These two steep increases were triggered by the Arab Oil embargo in 1973 and the outbreak of the Iranian Revolution in 1979 (Anon, 2017c).

3.5 Oil Wars

Some of the sharpest oil price increases in history have been associated with oil wars. The 1973 Oil Embargo in response to United States support for Israel during the Yom Kippur War caused a price spike from $ 3/bbl to nearly $ 12/bbl causing the first oil crises with many long and short term effects on the global economy (Anon, 2017b). Oil wars showed a positive relationship with oil prices from the model. This indicates that war times are associated with higher oil prices than non-war times, however, this was not significant at any of the conventional levels (p value of 0.911). This may be due to the fact that; wars were present in almost all the years considered under the study thus a good comparison between war presence and absence could not be established to prove wars' influential impact on oil prices.

3.6 Production

Since production and consumption have a very strong relationship, it can be inferred that global production rates also determine oil prices. These demand and supply factors work oppositely to keep a form of balance in the oil market, with increasing production affecting prices negatively which can be confirmed from OPEC's inverse relationship with price which is a subset of global oil production rates.

Fig. 3 shows the strong relationship between consumption and production rates. The trend shows that, when consumption rates began to exceed that of production, prices began to rise steeply. This can be explained by economic theory where prices of products and services escalate with increasing demand and production in the inverse direction.

According to Cummingham 2015, by the end of 2014, the largest oil producer, United States, had production levels of more than 9 million barrels of oil a day (9 MMbbl/d). This output accompanied with OPEC's refusal to cut down production levels went a long way in creating a glut of oil, which helped to send oil prices from 108 dollars per barrel in 2013 to 98 dollars per barrel and 52 dollars per barrel in 2014 and 2015 respectively.

Inferring from the strong correlation between refinery capacity and consumption rates, it can be concluded that refinery capacities also contribute to oil price determination. The relationship between refinery capacity and prices is explained using a trend analysis shown in Fig. 4.

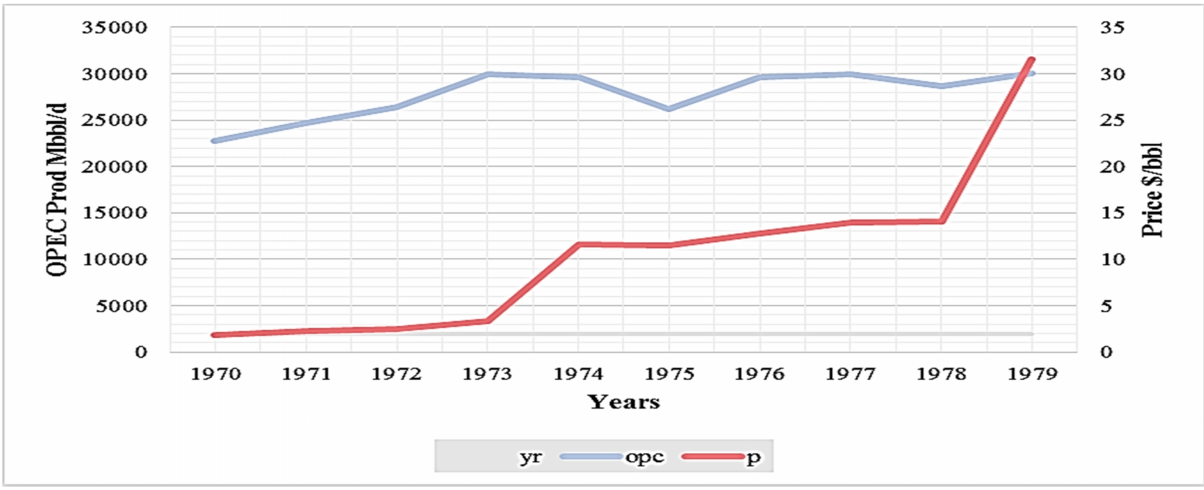


Fig. 2 OPEC's Production against Price in the 1970s

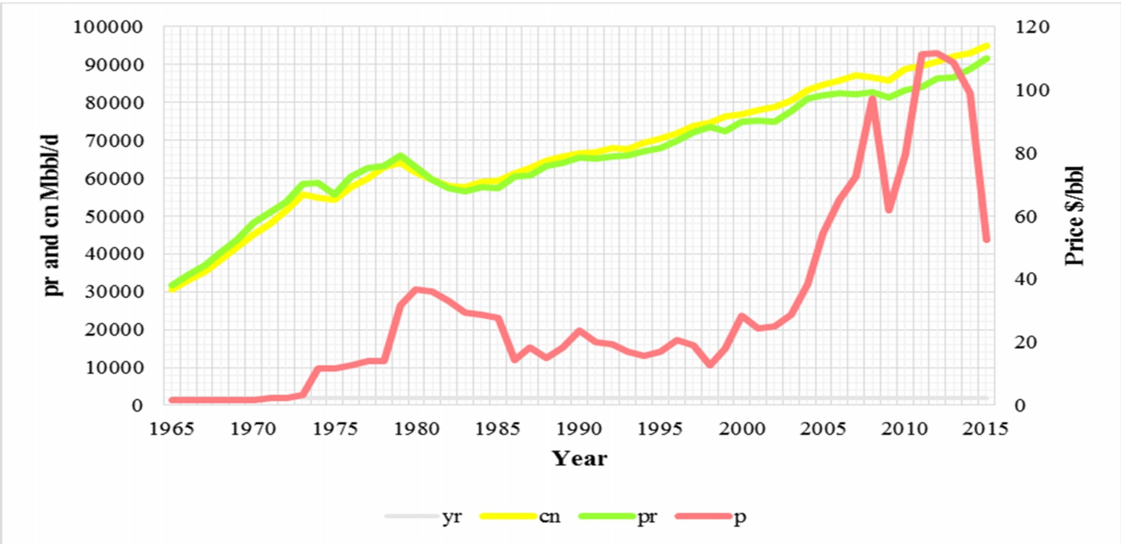


Fig. 3 Oil Production Consumption and Prices

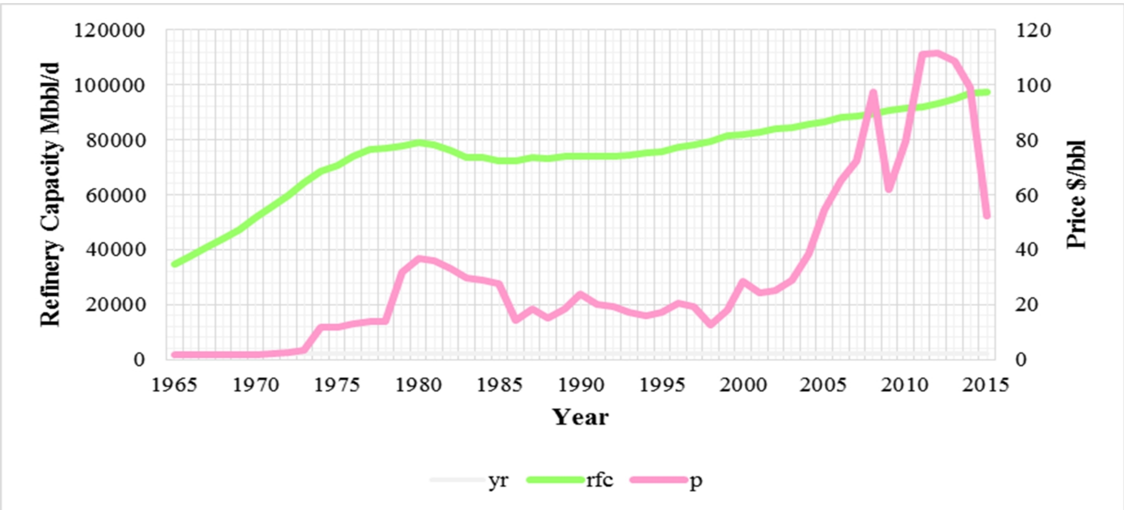


Fig. 4 Trend of Refinery Capacity and Price

3.7 Refinery Capacity

The trend of refinery capacity shows quite a sharp increase in earlier years followed by a gently downward movement and a steady, stable increase. The trend of price shows a general upward movement with a lot of fluctuations and instability within the years studied. The two trends are not identical as one is fairly stable and the other highly volatile. Although both trends tend to increase towards current years, there is no significant or concrete movement in refinery capacity's graph that can be attributed to the fluctuations in prices. Thus although refinery capacity rates play a role in pricing, its relationship with prices in the inverse or positive direction cannot be determined from this graph.

Olimb and Ødegård in 2010, established both negative and positive relationship of refinery capacity rates with oil prices using a time varying approach. Möbert (2007) found a positive relationship, while Kaufmann *et al.*, in 2005 found a negative relationship.

3.8 Diagnostic Check

Examining the goodness of fit statistics, the null hypothesis of the F test is rejected at 1 % significance level (p value of 0.000) proving the model generated is statistically significant. The results also indicated that about 82.3 % of the variation in oil price is caused by the factors present in the regression equation. With a relatively low p value of 0.507131, this model provides a very good fit for the available data values. The average distance of observed logged oil price values from that generated by the fitted plot is approximately 0.507131. The p value of the normality test for residuals was 0.045** ≈ 0.005, ** at 95 % confidence level as this approximation will have a negligible effect on the accuracy of the values generated. The residuals plots shown in Fig. 5 show the differences between the observed values and the models predicted values which should be minimal and unbiased. The normality plot reveals that the observed values lie close to those predicted and the histogram of residuals is also fairly bell-shaped. The residuals versus fits plot also shows a random distribution. Since the plots have these characteristics, the goodness of fit statistics can be depended upon as the residual plots reveal acceptable residual patterns. This regression model is thus appropriate in the analysis of the crude oil price determinants.

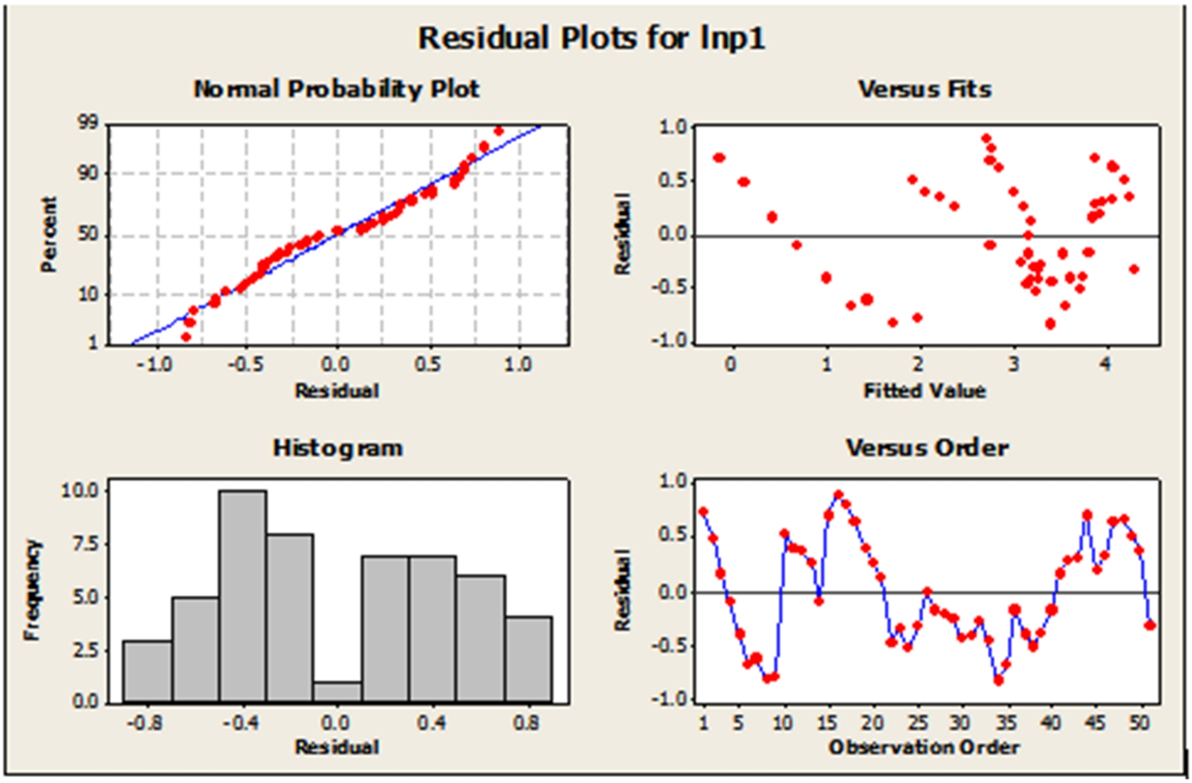


Fig. 5 Summary of Residual Plots of Regression Model

4 Conclusion

Six different factors were evaluated in this research using regression models and trend analysis. The evaluated factors were shown to behave in accordance with the hypotheses of the research even though a single method proved some factors such as oil wars to be insignificant in oil priced determination. From the analysis, it can be proven that:

(i) The explanatory variables examined except natural gas consumption and the presence of wars were found to be very significant in oil price determination.
(ii) Global oil and OPEC's production rates have a negative (inverse) relationship with oil price while consumption, the presence of wars directly related to oil prices.
(iii) That of refinery capacity however is not definitive as it can influence prices both directly and inversely.

In conclusion, crude oil prices change over time cannot be attributed to one single market factor, or just the explanatory factors analyzed in this work.

Acknowledgements

The authors wish to extend their gratitude first to the Almighty God for wisdom and understanding, and also a warm appreciation to the following individuals and institution for their technical support in bringing this paper to print: Prof Sunday Sunday Ikiensikimama, Mr Kofi Kamasa and University of Mines and Technology

References

Anon, (2017b), "1973 Oil Crisis", *https://en.wiki pedia.org/wiki /1973_oil_crisis*. Accessed: April 7, 2017.

Anon, (2017a), "Oil Wars", *https://en.wikipedia.o rg/wiki/Oil_war*. Accessed: February 17, 2017.

Anon, (2016a), "2015 in Review", *BP Statistical Review of World Energy Journal 2016*, 65th ed., pp. 1 – 43.

Anon, (2016b), "BP Statistical Review of World Energy June 2016 Workbook", *http://www.bp .com/statisticalreview*. Accessed: January 13, 2017.

Cunningham, N. (2015), "Top Five Factors Affecting Oil Prices in 2015", *http://oilprice .com/Energy/Energy-General/ Top-Five-Factors-Affecting-Oil-Prices-In-2015.html*. Accessed: January 25, 2017.

Frost, J. (2013), "Regression Analysis Tutorial and Examples", The Minitab Blog *http://blog .minitab.com/blog/adventures-in-statistics-2/regression-analysis-tutorial-and-examples*. Accessed: March 5, 2017.

Kaufmann, R., Dèes, S., Karadeloglou, P., Sànchez, M. (2005). "Modelling the World Oil Market: Assessment of a Quarterly Econometric Model", *Energy Policy*, pp. 178-191.

King, K., Deng, A., Metz, D. (2011), "An Economic Analysis of Oil Price Movements: The Role of Political Events and Economic News, Financial Trading, and Market Fundamentals", *Bates White Economic Consulting*, Washington DC, pp. 1 – 52.

Möbert, J. (2007), "Crude Oil Price Determinants", *Published MSc Thesis Report*, Germany, Darmstadt Discussion Papers in Economics, No. 186, 32 pp.

Nicholson, W., Snyder, C. (2011), *Microeconomic Theory: Basic Principles and Extensions*, Nelson Education. 10 ed, 763 pp.

Olimb, M., Ødegård, T. M. (2010), "Understanding the Factors Behind Crude Oil Price Changes A Time-varying Model Approach", *Unpublished MSc Thesis Report*, Norwegian University of Science and Technology, pp. 1 – 45.

Anon, (2017c), "Organisation of Petroleum Exporting Countries", OPEC webpage, *http://www.opec.org/opec_web/en/*. Accessed: February 8, 2017.

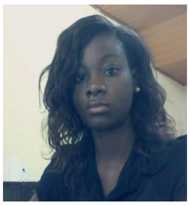
Ray, S. (2015), "7 Types of Regression Techniques you should know!", *https://www.scribd.com /document/313114056/7-types-of-regression-techniques-you-should-know*. Accessed: March 11, 2017.

Seth, S. (2015), "Do Oil and Natural Gas Prices Rise and Fall Together", Investopedia webpage, *http://www.investopedia.com/articles/active-trading/032515/do-oil-and-natural-gas-prices-rise-and-fall-together.asp*. Accessed: February 2, 2017.

Yergin, D. (1991), *The Prize: The Epic Quest for Oil, Money & Power*, Simon & Schuster, New York, pp. 1 – 78.

Authors

Eric Mensah Amarfio is a PhD candidate and a lecturer at the University of Mines and Technology, Tarkwa, Ghana. He obtained his MSc in Petroleum Engineering from African University of Science and Technology, Abuja, Nigeria and his BSc in Chemical Engineering from the Kwame Nkrumah University of Science and Technology, Kumasi, Ghana. He is a member of the Society of Petroleum Engineers International, USA, and the Editorial board of the International Journal of Petroleum and Petrochemical Engineering, India. His research areas of interest are; Nano-drilling fluids technology, Reservoir and production optimization, Enhanced oil recovery, and Petroleum Economics.



Mina Gyagri is a BSc graduate and former student of the University of Mines and Technology, Tarkwa, Ghana. She obtained her BSc in Petroleum Engineering from the University of Mines and Technology, Tarkwa, Ghana. She is a member of the Society of Petroleum Engineers International, USA. Her interest research areas include; Petroleum Economics, Fuel Quality and Enhanced oil recovery.

Solomon Adjei Marfo is a Lecturer at the Petroleum Engineering Department of University of Mines and Technology, Tarkwa, Ghana. He holds PhD in Petroleum Engineering from the University of Port Harcourt, Nigeria, MEng Degree in Mining (Petroleum Engineering) from the University of Belgrade, Serbia and BSc in Chemical Engineering from the Kwame Nkrumah University of Science and Technology, KNUST, Kumasi, Ghana. He is a member of the Society of Petroleum Engineers (SPE), a Registered Environmental Specialist (RES) with the National Registry of Environmental Professionals (NREP) of USA. His research interests include sand and water control in oilfields using chemicals; agro-waste and local materials evaluation for oilfields application; oil and gas pipeline design and modelling.