Oil Price Factors: Forecasting on the Base of Modified ARIMA Model

Anthony Msafiri Nyangarika $^{1,2}$, Bao-jun Tang$^1$

$^1$School Management and Economy, Beijing Institute Technology, 5 South Zhongguancun Street, Beijing 100081, Peoples R China.
$^2$Institute of Adult Education, P.O. Box 20679 Dar sea Salaam, Tanzania.
E-mail: nyangarikatz@gmail.com

Abstract. The paper proposes the modification of ARIMA model for finding the parameters of estimation and forecasts using exponential smoothing. The study use data Brent crude oil price and gas prices in the period from January 1991 to December 2016. The result of the study showed an improvement in the accuracy of the predicted values, while the emissions occurred near the end of the time series. It has minimal or no effect on other emissions of this data series. The study suggests that investors can predict prices by analyzing the possible risks in oil futures markets.

1. Introduction
Over the years, oil has remained one of the most important sources of energy. All countries, one way or another, are consumers of oil and oil products. There are already more than 100 countries in the oil-producing countries. Prices for oil and its derivatives are of concern to both producers and consumers. The dynamics of oil prices affect the level of costs in all production sectors. The economy of many countries is based on oil production and trade in oil and oil products, the forecasting of oil prices is an urgent task. Sectors of the economy are directly dependent on oil price forecasts. Oil prices influence the political and economic processes that determine the value of oil companies' shares, the level of inflation in the oil importing countries, and the speed of economic growth. It is important to note the impact of oil prices on the formation of prices for alternative energy sources. The purpose of this work is to identify factors that affect the price of oil and to obtain a reliable forecast model of oil prices. To achieve this goal, it is necessary to perform a number of tasks:

- To study the factors influencing the price of oil;
- Consider the method of forecasting ARIMA data;
- Collect and conduct descriptive data analysis;
- Build a regression model and identify significant factors;
- Get forecasts on the methods outlined above, choose the best one and build on it a forecast for the future.

The total volume of oil consumption in 2014 was approximately 4.2 billion tons, which is 54% more than in 1973. Thus, the average increase in oil consumption over the years since the oil shock was $\sim 1\%$ per year. At the same time, after the economic crisis of 1973-1983, oil consumption steadily grew until the beginning of the 2008 crisis. However, there is a widespread opinion that significant and unexpected fluctuations in oil prices have a negative impact on the welfare of both oil importers and oil-producing countries. Oil prices and oil price volatility both play important roles in affecting the global economy, although the effects are asymmetric depending on periods, regions, sectors, the
reason for oil shock, and others. Several studies found that higher oil prices have an adverse impact on the global economy [1]. Moreover, [2],[3] found an economic impact on oil importing countries such as South Korea. In order to make appropriate decisions about the direction of economic policy, therefore, it is important to accurately forecast future oil prices with effective models.

The price of oil is one of the key factors determining the country's budget in terms of its revenues. The practice of determining the forecast price of oil is based on the method of constructing consensus forecasts. This method is based on forecasts of the largest players in the oil market, investment banks, international economic and financial organizations. These include the International Energy Agency, the Organization of Petroleum Exporting Countries (OPEC), the World Bank, IHS Global Insight, Raiffeisen Bank, the International Monetary Fund [4].

The following shortcomings attributed to this approach.

1. The closed nature of forecasting techniques, based on the results of which consensus forecasts are built. Since almost every method of forecasting has certain drawbacks, the closed nature of the applied methods does not allow us to estimate the degree of possible forecast error. Using in the construction of a consensus forecast the results obtained from various sources, each of which used different forecasting techniques, can lead to an "inheritance" of the deficiencies inherent in the initial projections.

2. On the other hand, the initial estimates were based on specific assumptions and assumptions, methodological approaches that allow us to obtain an acceptable forecast, the use of the consensus forecast will actually level the result, distorting the results of qualitative initial projections and introducing a share of erroneous forecasts estimates obtained from other sources.

Analysis of the practice of constructing forecast estimates and forecasting methods applied by various scientific organizations, state bodies, and commercial companies has shown that today the most popular approaches used by various financial organizations and institutions are econometric forecasting methods. In this regard, as an alternative to the consensus forecast method, [14] proposed to use the prediction method.

In addition, some sectors of the economy directly depend on the forecast of oil prices. For example, airlines that rely on air ticket price forecasts, the automotive industry and simply homeowners who rely on oil price forecasts (and prices for secondary products such as gasoline or heating oil) in modeling the purchase of long-term goods use such as cars or home heating systems.

2. Literature review

Several studies found that the Organization of the Petroleum Exporting Countries (OPEC) decided to maintain oil production in 2014, the crude oil price dropped to less than $50/Bbl. The price has stayed at mid-$40/Bbl on continued sluggish oil demand and strong shale supply in 2015 and 2016. [4],[5] proposed that consequently, oil price volatility and another oil crisis have been growing. In this context, knowing the long-term trend in crude oil prices is essential for ensuring future economic stability in many countries because significant changes in crude oil prices and unstable oil supplies may seriously affect their economies, which depend on crude oil imports and exports. The Auto-Regressive Integrated Moving Average (ARIMA) methodology was used time-series data to reflect the wild volatility of time-series data. Besides ARIMA models forecast oil prices by using the interrelationship between the future price and the spot price of crude oil in short-term forecasting. [6], [7] explained a conditional variance that changes over time, to forecasting the Brent oil price. [11] estimated the oil price needed to maximize the producer’s profit in a perfectly competitive and monopolistic market using dynamic optimization. In his results, oil prices followed a U- shape pattern in the case of a small initial reserve endowment but then showed a rise over time in the case of a large initial reserve endowment.

Since 2000, financial factors, including the penetration of speculative forces, a weakening dollar, and the financial crisis, have attracted attention as possible determinants of global oil prices. [15],[16] have also provided support for the role of speculation in the oil market, especially for its role in the rise of crude oil prices. For example, [17] found that financial shocks have considerably contributed to oil price increase since the early 2000s, and to a much larger extent since mid-2000s. Among several financial factors, the speculative expectation has been indicated as an important determinant of the price for a commodity.
Even though [18] explained the changing pattern in oil prices, his approach is difficult to apply to actual data and is limited in that it examines factors driving oil price fluctuations only from the supply side. Man research institutes have used EIA forecasts as credible data. Delphi approach, which repeatedly collects opinions to derive the joint subjective view of experts, can also be used to forecast oil prices. Using prices determined in the future oil market has been suggested as a forecasting methodology. [19],[20] analyzed if future prices from a certain time could be appropriately used to forecast spot prices by testing the Granger causality between WTI spot prices and future prices. While forecasting oil prices using future prices shows accurate performance in the short term. Such an approach tests if the future price is an unbiased predictor of the spot price at the maturity time [21],[22] evaluated forecasting accuracy by comparing future prices (1, 2, 3 and 4-month), future contracts with WTI spot prices from 1991 to 2016. [23] used WTI spot and future prices from July 2000 to June 2004 as sample data, selecting the forecasting period that yielded the most accurate forecasts by comparing quarterly forecasts based on future prices from the previous one to six months with the average of the quarterly WTI oil prices.

Previous research on oil price forecasting models has generally assumed that the current trend in oil prices will continue in the future and thus that factors influencing oil will have the same effects in the future. However, factors influencing oil prices have changed structurally over time. In the 1960s, supply-side factors determined the crude oil price, and this trend continued until the oil price collapse of the mid-1980s. Consequently, an oil pricing system linked to the oil market has existed since the late 1980s, and the crude oil price has been determined by demand as well as supply. In the 1990s, especially, emerging markets such as China and India led oil prices to rise. However, the role of speculation in causing the significant changes in oil prices is still debatable. Several studies are not supportive of speculation being an important determinant of the real oil prices and. Even though the global oil market paradigm has been changing continuously, previous forecasting models have rarely reflected such structural changes. As such, this study can contribute to preparing quick and accurate oil market countermeasures by forecasting short-term oil prices. This study’s model is highly applicable. The forecast oil prices reported here can thus be used to inform reasoned decision making by the government and the private sector.

3. Methodology

The auto-regressive integrated moving average (ARIMA) methodology was used time-series data, the wild volatility of time-series data. Besides time-series models such as ARIMA and GARCH models, ARIMA has also been employed to forecast oil prices by using the interrelationship between the future price and the spot price of crude oil, which explain a conditional variance that changes over time, to forecasting the Brent oil price which are used to prove the cointegration between the real (spot) oil prices and the prices of 1, 2, 3 and 4-month futures contracts. In this paper, we will consider the method of forecasting using the ARIMA model. Due to the constant changes occurring in the world, we found it prudent to build short-term and retro forecasts. In the framework of this work, we are primarily interested in such a method as ARIMA (Auto Regressive Integrated Moving Average). Despite the fact that this model belongs to the class of linear methods, it equally well describes stationary and non-stationary time series. In addition, independent variables are not used in this model, which means using only the information embedded in the data itself for forecasting. The autoregressive model (AR) of the order $p$ has the following form:

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \epsilon_t$$

Where, $Y_t$ - dependent variable at the time $t$; $\phi_0$, $\phi_1$, $\phi_2$, $\ldots$, $\phi_p$ - estimated coefficients; $\epsilon_t$ - an error describing the effects of variables that are not taken into account in the model. The moving average model (MA) of the order $q$ is described as follows:

$$Y_t = \mu + \epsilon_t - \omega_1 \epsilon_{t-1} - \omega_2 \epsilon_{t-2} - \ldots - \omega_q \epsilon_{t-q}$$
Where, $Y_t$ - dependent variable at the time $t$; $\mu$ - constant process average; $\varepsilon_t$ - an error at time $t$; $\omega_1, \omega_2, \ldots, \omega_q$ - estimated coefficients. Some non-stationary time series can be reduced to stationary ones using the operator of a consecutive difference. Assume that there is a time series $Y_t$, to which $d$ times applied this operator, after which the series became stationary $\Delta^d Y_t$ and satisfying the conditions of the model ARMA $(p, q)$. The model of auto regression and moving average will have the form:

$$
\Phi(L)y_t = \delta + \Theta(L)\varepsilon_t, \quad \varepsilon_t \sim iid(0, \sigma^2),
$$

(3)

Where

$$
\Phi(L) = 1 - \varphi_1 L - \varphi_2 L^2 - \cdots - \varphi_p L^p,
$$

$$
\Theta(L) = 1 - \theta_1 L - \theta_2 L^2 - \cdots - \theta_q L^q
$$

are polynomials from the shift operator. In this case $Y_t$ will be called the integrated process of auto regression and moving average or ARIMA $(p, d, q)$. This model allows you to build very accurate forecasts with a short forecasting range. It is also quite flexible and can be suitable for describing different time series. In addition, ARIMA models are simply checked for their adequacy. However, the disadvantages of this method include the need for a large number of initial data and the absence of a simple method of adjusting the parameters of the model.

4. Results and discussion

We use data on prices for Brent crude oil in the period from January 1991 to December 2016. We took this particular energy source as a substitute for oil because they are one of the most popular on the market today. The task was to see how much the price of oil depends on the price of alternative energy sources. The impact of armed clashes in the oil-producing countries is becoming less important in the formation of oil prices [8], [9] believe the opposite. Also, as a dummy variable, we included the global financial crisis - it was in 2008 that it had a significant impact on the price of oil, and caused one of the most significant falls. (Table.1) shows all the factors that we will include in the model - both in the form of time series (oil price, gold price, and gas price), and in the form of fictitious variables (World financial crisis, military conflicts of Iraq, Iran, Syria and Afghanistan, a terrorist attack in the United States). The right column of (Table.1) shows which designation for each variable we specify in the Eviews program.

We use the econometric models to identify the dependence between the variables, we need to look at the descriptive statistics for our variables, as well as check the data for the presence of emissions. All this must be done to obtain the most accurate model. Descriptive statistics for a number of oil prices in (Fig. 1).

Therefore, from this histogram, shown in Figure 1, the mathematical expectation for the OIL variable is 48.93, which means that the average value of oil prices fluctuates around $ 49 / bbl. The standard deviation of this variable is 34.93. Those the spread of individual values of OIL with respect to its mean value is 35. The following is the series for stationarity as shown below (Table. 2).

| Factor | Variable |
|--------|----------|
| Brent crude oil price, The price of gold | Oil, Gold, Gas |
| Price gas | |

| Dummy variables: | |
|------------------|------------------|
| World financial crisis | MFC |
| The Military company in Iraq | Iraq |
| The Military company in Iran | Iran |
| The Military company in Syria | Syria |
| The Military company in Afghanistan | Afghanistan |
| The US Terror | Terror |

Table 1. Description of variables

Source: Authors ‘calculation
The series is not stationary (Table.2), the probability value Prob = 0.6137, we cannot reject the hypothesis of the presence of a unit root; therefore, the series is not stationary. In order to get rid of no stationarity, we check the series for the first difference

**Table 2.** Test for stationarity of a number of oil prices

| Test                      | t-Statistic | Prob. |
|---------------------------|-------------|-------|
| Augment Dickey-Fuller test statistic | -1.298902   | 0.6137|

**Source:** Authors’ calculation

According to the results presented in (Table.3), the hypothesis of the presence of a unit root is rejected; we succeeded in bringing the series to a stationary form. In order to be convinced of the absence of emissions, a Boxplot graph should be constructed. Our graph for the OIL variable indicates no emissions (Fig. 2).
Boxplot graph carries out similar descriptive statistics for explanatory variables: gas and gold. We turn to a description of a number of gas prices.

From this histogram (Fig. 3) it can be seen that the mean for the GAS variable is 3.95, which indicates that the average value of coal prices fluctuates around 4. The standard deviation of this variable is 2.19. The spread of individual values of GAS with respect to its mean value is 2.2. By checking the series for stationarity, we again encountered the no stationarity of the data series (Table.4).

| Test                        | t-Statistic | Prob.   |
|-----------------------------|-------------|---------|
| Augment Dickey-Fuller test statistic | -1,855319   | 0.3467  |

**Source:** Authors’ calculation

The value Prob > 0.05, we cannot reject the null hypothesis about the presence of a unit root. Taking the first differences for a number of gas prices, the series to a stationary form (Table 5).

| Test                        | t-Statistic | Prob.   |
|-----------------------------|-------------|---------|
| Augment Dickey-Fuller test statistic | -6,023421   | 0.0000  |

**Source:** Authors’ calculation

The following test variable is the last of the series - gold prices.

From this histogram (Fig.4), it can be seen that the mathematical expectation for the variable GOLD is 685.25, which means that the average value of gold prices fluctuates around 685. The standard deviation of this variable is 457.02. The spread of individual values of GOLD with respect to its mean value is 457. A number of these gold prices were initially unsteady, so using the method of first differences already known to us; we bring the series to a stationary form.

The value of Prob < 0.05, therefore, we can reject the hypothesis of the presence of a unit root, thereby confirming the stationarity of the series (Table.7). Similarly, to complete the descriptive analysis, it is necessary to check the series for the presence of emissions. To do this, we built Boxplot graphics.

| Test                        | t-Statistic | Prob.   |
|-----------------------------|-------------|---------|
| Augment Dickey-Fuller test statistic | -5,524744  | 0.0002  |

**Source:** Authors’ calculation

According to the graphs (Fig. 5), we show that the gold variable GOLD has no emissions, which cannot be said about the variable that includes gas prices - GAS. Despite the presence of emissions from this variable, we will not get rid of them in order to get the most accurate and complete picture of the effect of gas prices on the price of oil. It will also be interesting to look at the correlograms for each of the series of data.

Analyzing the correlograms for each of the series of data (Fig. 6), we can say that all our series are stationary - the correlograms decreases from the germ k after the first values. In addition, there is no periodic component in each of the series of data, which tells us that there is no seasonality. In order not to encounter the phenomenon of multi collinearity in the future when constructing the regression, we will check our variables for the presence of a correlation between them.

In order to construct an econometric model, we will use fictitious variables, which include military conflicts and the global financial crisis. We created variable with value 1, in case of conflict, and otherwise 0. For example, the variable world financial crisis in our regression model will take
the value 1 in the period from 2008 to 2010, when during 2008 (year of the financial crisis), the value of oil prices assumed the lowest values, in other cases it will be zero, similar data will be made for other fictitious parameters.

![Figure 4. Histogram for gold prices](image)

![Figure 5. Pox Plot chart for gas and gold prices](image)

![Figure 6. Correlograms of the price of oil](image)

However, the construction of the regression model. As a dependent variable, we will use oil prices - OIL, as explanatory gas prices - GAS and gold - GOLD, as well as include dummy variables - CRISIS, IRAN, IRAQ, AFGHANISTAN, SYRIA, and TERROR. It is important to note that in order to construct the regression, we take all the data series in the differences. This is explained by the fact that initially, all our series were nonstationary, and we brought them to a stationary form by taking the first differences for each of the series of data. (Table 7) shows the values of the coefficients and probabilities for each of the variables included in the constructed model

| Variable        | Coefficient | Probability |
|-----------------|-------------|-------------|
| D(GAS)          | 6.336001    | 0.0001      |
| D(GOLD)         | 0.103214    | 0.0002      |
| Iran            | 3.262685    | 0.6778      |
| Iraq            | -11.17840   | 0.0092      |
| AFGHANISTAN     | 9.845998    | 0.2112      |
| SYRIA           | -7.859139   | 0.1227      |

Table 7. The value of the corresponding probabilities for the regression variables
From the above (Table.7), we can conclude that the variables D (gas), D (gold) and Iraq are significant (Table.7)- show that they have an effect on the price of oil. While the probabilities of the rest are greater than 0.05, which indicates their insignificance. There is no correlation between these variables and the oil price variable - OIL. A more detailed table obtained in the construction of the model (Fig.7).

The results show the variable GAS was significant, i.e. rising or falling in gas prices leads to changes in oil prices. This can be explained by the fact that each of these types of energy resources is very widely used and the volumes of their production and consumption are quite large, which leads to the influence on each other. Another explanation can be the fact that gas in some industries is a substitute for oil, therefore, in the case of an increase in oil prices, the demand for it will decrease and the transition to other, cheaper energy resources, for example, gas will be implemented, which will increase the demand for it and subsequently the price.

Therefore, the gold, we cannot reveal the effect of the change in gold prices on the price of oil. This is explained by the fact that despite the apparent popularity of investing in precious metals, they do not stop investing in shares of oil companies. Of all the fictitious variables, only IRAQ was significant, a conflict that began in December 2004. It can be said that the significance of military conflicts in the oil-producing countries has an ever-smaller and insignificant effect on the price of oil. Thus, we perceive that over time, in fact, one factor increases in importance, while others decrease. In order to correctly estimate the model constructed, we carry out the Ramsey test (Table.8).

| Test          | F-Statistic | Prob. F(1, 15) |
|---------------|-------------|----------------|
| Ramsey RESET Test | 2.123687   | 0.9697         |

Source: Authors’ calculation

According to the values of F-Statistic and Prob. Presented in (Table.8), we can conclude that the hypothesis of the acceptability of the functional form is adopted, that is, this model is correctly specified. To get a more accurate model, we conducted a test for extra variables (Fig.8). This test confirmed that the insignificant variables of our regression model, namely, IRAN, AFGHANISTAN, SYRIA, TERROR, and CRISIS are superfluous and we can exclude them from the model. After analyzing the correlograms (Fig.8) and eliminating the extra variables, we constructed the following model (Fig. 9). From all variables are significant, low probabilities tell us this (Prob) in (Fig.9). The value of the criterion Akaike info criterion decreased, which again indicates that this model has become better. F-statistics has assumed a higher value. In addition, when constructing the regression, we included the processes AR and MA to get rid of the autocorrelation, which we found in the analysis of the correlograms.
5. Conclusion
In this paper were considered: factors affecting the price of oil and ways to predict this price using different models. In the course of the analysis, it turned out that among all the factors we were considering, the value of oil prices is influenced by the price of gold (GOLD) and the armed conflict in Iraq that has occurred since 2004 (IRAQ). However, the factors that proved insignificant in this model: the financial crisis, the conflicts in Iran, Afghanistan, Syria, and the terrorist attacks that occurred in the Middle East and the United States. This can lead to increased demand for oil and, as a result, will lead to an increase in the price. In this paper, not all the problems that arise when forecasting oil prices were considered, therefore it would be advisable to continue to consider different forecasting methods in the future, so that the values obtained are as close as possible to the real ones. One of the directions for further research can be the application of a larger number of models of different types to obtain different forecasts of the series.

Reference
[1] Akpan, E. O. (2009), Oil price shocks and Nigeria’s macro economy. A Paper Presented at the Annual Conference of CSAE Conference. Economic Development in Africa. Retrieved from http://www.csae.ox.ac.uk/conferences/2009-EDIA/Papers/252-Akpan.Pdf.
[2] Amano, R. A. and Van Norden, S. (1998), Exchange Rates and Oil Prices. Review of International Economics, 6: 683–694. doi:10.1111/1467-9396.00136.
[3] Baumeister, C., & Peersman, G. (2008), Time-varying effects of oil supply shocks on the US economy.
[4] Backus, D. K. and M. J. Crucini (2000), Oil Prices and the Terms of Trade. Journal of International Economics, 50, 185-213.
[5] Barsky, R.B.; Kilian, L. (2004), Oil and macro economy since the 1970s. J. Econ. Perspect. 18, 115–134.
[6] Buetzer, S., Habib, M. and L. Stracca (2012), Global Exchange Rate Configurations: Do Oil Shocks Matter? Working Paper, European Central Bank.
[7] Farzanegan, M. R., & Markwardt, G. (2009), The effects of oil price shocks on the Iranian economy. Energy Economics, 31(1), 134-151. http://10.1016/j.eneco.2008.09.003.
[8] Ferraro D., Rogoff K. and B. Rossi (2015), Can Oil Prices Forecast Exchange Rates? Available at: http://www.econ.upf.edu/~brossi/OilAndExchangeRates.pdf.
[9] Huang, Y. and F. Guo (2007), The role of oil price shocks on China's real exchange rate. China Economic Review, 18: 403-416.
[10] Hooker, M. (1996), Oil prices and the rise and fall of the US real exchange rate. Journal of Monetary Economics, 25, 195-213.
[11] Hsu, T.-K.; Tsai, C.-C.; Cheng, K.-L. (2016), Forecast of 2013–2025 crude oil prices: Quadratic sine- curve trend model application. Energy Sources Part B 2016, 11, 205–211.
[12] Kilian, L. (2009), Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market. Am. Econ. Rev. 99, 1053–1069.
[13] Kilian, L.; Murphy, D.P. (2014), The role of inventories and speculative trading in the global market for crude oil. J. Appl. Econ., 29, 454–478.
[14] Mikhailov A. (2014), Faktory razvitiya ekonomiki Rossii v 2015 godu [Factors of development of the Russian economy in 2015]. Voprosy regulirovaniya ekonomiki = Questions of economic regulation, 4, pp. 62-69.
[15] Mikhailov A. (2018), Pricing in Oil Market and Using Probit Model for Analysis of Stock Market Effects. International Journal of Energy Economics and Policy, 4, pp. 43–53.
[16] Mikhailov, A. (2018), Volatility Spillover Effect between Stock and Exchange Rate in Oil Exporting Countries. International Journal of Energy Economics and Policy, 2018, 8(3), 321-326.
[17] Morana, C. A (2001), Semiparametric Approach to Short-term Oil Price Forecasting. Energy Econ, 23, 325–338.
[18] Li, H.; Wells, M.T.; Yu, C.L. (2008). Bayesian Analysis of Return Dynamics with Stochastic Volatility and Levy Jumps. Rev. Financ. Stud, 21, 2345–2378.
[19] Olomola, P. A., & Adejumo, A. V. (2006). Oil price shock and macroeconomic activities in Nigeria. International Research Journal of Finance and Economics, 3(1), 28-34.
[20] Sadorsky, P. (1999), Oil price shocks and stock market activity. Energy Econ. 21, 449-469.
[21] Segal, P. (2011), Oil price shocks and the macro economy. Oxf. Rev. Econ. Policy, 27, 169–185.
[22] Singer, E. (2007), Oil Price Volatility and the US Macroeconomy: 1983-2006. Working paper, Minnesota of USA: Carleton College.
[23] Tuzova, Y., & Qayum, F. (2016), Global oil glut and sanctions: The impact on Putin’s Russia. Energy Policy, 90, 140–151. https://doi.org/10.1016/j.enpol.2015.12.008