IJEEP
INTERNATIONAL JOURNAL OF ENERGY ECONOMICS AND POLICY

International Journal of Energy Economics and Policy

ISSN: 2146-4553

available at http: www.econjournals.com

International Journal of Energy Economics and Policy, 2019, 9(1), 149-159.

Oil Price Factors: Forecasting on the Base of Modified Auto-regressive Integrated Moving Average Model

Anthony Nyangarika[1], Alexey Mikhaylov[2]*, Ulf Henning Richter[3]

[1]School of Management and Economics, Beijing Institute of Technology, Beijing, China, [2]Department of Financial Markets and Banks, Financial University under the Government of the Russian Federation, Moscow, Russia, [3]School of Economics and Management, Tongji University, Tongji, China. *Email: ayumihajlov@fa.ru

Received: 22 June 2018 **Accepted:** 13 November 2018 **DOI:** https://doi.org/10.32479/ijeep.6812

ABSTRACT

The paper proposes modification of auto-regressive integrated moving average 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 analyzing the possible risks in oil futures markets.

Keywords: Auto-regressive Integrated Moving Average Model; Econometric Model; Oil Price Forecast
JEL Classifications: C51, C58, F31, G12, G15

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 >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, so the forecasting of oil prices is an urgent task. It is also worth noting that some 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 in 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 auto-regressive integrated moving average (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 ~ 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.

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 of 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 and Backus and Crucini (2000).

The following shortcomings should be 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 was 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, Mikhaylov (2014) 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

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, reason of oil shock, and others.

Different views on the impact of changes in oil prices on the global economy have been suggested. For example, Sadorsky (1999), Barsky and Kilian (2004), Kilian (2009), Segal (2011), Morana (2001) and Kilian and Murphy (2014) present a good account of these different views.

Through this debate, several studies found that higher oil prices have an adverse impact on the global economy (Akpan, 2009). Moreover, Amano and Van Norden (1998) found 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.

In June 2008, global oil prices, which had been on an upward trend since 2003, surged to $134/Bbl (for West Texas Intermediate, WTI). Oil prices fell after the global economic recession of 2008 but started to rise in early 2009.

Studies have suggested a possible explanation for this projected slowdown in oil demand growth, such as structural changes of the global economy, consumer reactions and government policies, and shale gas development in the United States shown by Baumeister and Peersman (2008).

After the OPEC decided to maintain oil production in 2014, the crude oil price dropped to <$50/Bbl. The price has stayed at mid-$40/Bbl on continued sluggish oil demand and strong shale supply in 2015 and 2016.

Backus and Crucini (2000), Farzanegan and Markwardt (2009) 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.

Sophisticated forecasting models are able to reliably predict long-term crude oil prices and provide updated information based on fluctuating market conditions to all concerned parties, thereby contributing to reasonable decision-making by policymakers and company managers.

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. Buetzer et al. (2012) explained a conditional variance that changes over time, to forecasting the Brent oil price.

Hsu et al. (2016) 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.

Even though Li et al. (2008) 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.

Such an approach tests if the future price is an unbiased predictor of the spot price at the maturity time. Tuzova and Qayum (2016) 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 1-6 months with the average of the quarterly WTI oil prices.

Singer (2007) evaluated forecasting accuracy by comparing future prices (1, 2, 3 and 4-month), future contracts with WTI spot prices from 1991 to 2016.

Olomola and Adejumo (2006) 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.

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.

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. For example, Morana (2001) found that financial shocks have considerably contributed to oil price increase since early 2000s, and to a much larger extent since mid-2000s. Among several financial factors, speculative expectation has been indicated as an important determinant of the price for a commodity.

Mikhaylov (2018a), Mikhaylov (2018b) have also provided support for the role of speculation in the oil market, especially for its role in the rise of crude oil prices.

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

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 future contracts.

In the course of the work, the ASE and APE models including emissions were evaluated. The main idea of constructing this regression was that emissions in time series can influence the parameters of estimation and forecasts using exponential smoothing. The aim of the study was to show the way in which the necessary emissions can be included in linear models of innovation for the method of exponential smoothing. Researchers using this method emphasize the fact that attention should be paid to emissions at the end of the time series.

As a result of the study, the emission model showed an improvement in the accuracy of the predicted values, while the emissions occurred near the end of the time series, even considering the fact that they had minimal or no effect on other emissions of this series of data.

It is also worth mentioning studies in which forecasting models are used – ARIMA-GARCH, ARFIMA-GARCH and ARFIMA-FIGARCH. The main idea was to identify the best model for predicting the risks of three types of oil future contracts.

Studies suggest that none of the above forecasting models can be suitable for all three types of future contracts. For example, the price of WTI selects a simple ARIMA-GARCH model, while future prices for fuel oil and gasoline prefer ARFIMA-FIGARCH.

Hooker (1996) suggested that investors should be cautious; analyzing the possible risks in oil futures markets.

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.

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 has the following form:

$$Y_t = \varphi_0 + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \cdots + \varphi_p Y_{t-p} + \varepsilon_t \quad (1)$$

Where, Y_t - dependent variable at time t; φ_0, φ_1, φ_2, --- φ_p- estimated coefficients; ε_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:

$$Y_t = \mu + \varepsilon_t - \omega_1 \varepsilon_{t-1} - \omega_2 \varepsilon_{t-2} - \cdots + \omega_q \varepsilon_{t-q} \quad (2)$$

Where, Y_t- dependent variable at time t; μ - constant process average; ε_t - error at time t; $\omega 1$, $\omega 2$,---, ω_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 autoregression and moving average will have the form

$$\Phi(L) y_t = \delta + \Theta(L)\varepsilon_t, \varepsilon_t \sim \mathrm{iid}(0, \sigma^2) \quad (3)$$

Where, $\Phi(L) = I - \varphi_1 L_1 - \cdots - \varphi_p L_p$ and $\Theta(L) = I - \theta_1 L_1 - \cdots - \theta_q L_q$ - polynomials from the shift operator. In this case will be called the integrated process of autoregression 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. The quality of the obtained model will be determined by the following coefficients:

Determination coefficient R^2 (R-squared).

$$R^2 = 1 - \frac{\sum_t (y_t - \hat{y}_t)^2}{\sum_t (y_t - \bar{y}_t)^2} \quad (4)$$

Where y – sample mean of the dependent variable Y_t

The coefficient of determination shows how far the constructed model differs from the best constant. If the model has a free term (constant), the values of the determination coefficient vary from 0 to 1. In this case, it can be interpreted as the fraction of the variance of the dependent variable Y_t, explained by means of the independent and lag variables included in the model, in the form in which they are present in the model. Otherwise, the coefficient of determination may be negative.

Adjusted coefficient of determination R^2_{adj}

$$R^2_{adj} = 1 - (1 - R^2) \cdot \frac{T-1}{T-k} \quad (5)$$

Where T – length of time series; k – number of model parameters to be evaluated.

Since the ordinary determination coefficient R^2 It does not decrease when additional lags are included in the evaluated model, it can not serve as a good measure of the quality of the model. When calculating the corrected determination coefficient, a fine is introduced for additional regressors (lag variables), therefore the values of the adjusted determination coefficient do not exceed the corresponding values of the usual determination coefficient. R^2_{adj} can be reduced by including additional variables in the model, and may also be negative if the model is poorly specified.

Standard regression error *s.e.regr*

$$s.e.regr = \sqrt{\frac{\sum_{t=1}^{T} \hat{\varepsilon}_t^{\,2}}{T-k}} \quad (6)$$

Shows the variance of the time series relative to the constructed model.

Akaike information criterion AIC

$$\mathrm{AIC} = -2\frac{l}{T} + 2\frac{k}{T} \quad (7)$$

Where l logarithm of the likelihood function. The information criterion of Akaike, as well as the Schwartz information criterion, is used to select the best model from some set of alternative models - the smaller the criterion value, the better the model.

F-statistics. Using F-statistics, assuming that the model remains are normally distributed, the hypothesis of insignificance of the regression as a whole is verified. The null hypothesis is checked that the coefficients for all exogenous (independent and lagged) variables included in the model, except for the free term, are zero.

$$F = \frac{R^2 / (k-1)}{(1-R^2) / (T-k)} \quad (8)$$

Where F – the calculated value of F-statistics. It can be compared to a tabular F to accept or reject the null hypothesis at a significance level.

P-value (Prob (F-statistic)). The significance of F-statistics is the probability that for an arbitrary sample from the same population as our sample, the value of the F-statistic will be greater or equal to the calculated one (located farther from 1 than F calculated). Other owls, the probability of obtaining such a calculated value of F-statistics provided that the null hypothesis is true.

We use all time series from Thomson Reuters Datastream.

4. RESULTS AND DISCUSSION

For the practical part of this work, it was decided to use data on prices for Brent crude oil. We collected data on oil and gas prices 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. In addition, we took gold as one of the explanatory variables. This can be explained by the fact that the price of oil can depend on how many people invest in oil companies, gold in this case is an alternative form of investment, which is gaining increasing popularity.

We have introduced a number of fictitious variables, which were military conflicts in the Middle East and terrorist acts. The reason why we decided to consider military actions was a common opinion as to what impact they have on the price of oil. The impact of armed clashes in the oil-producing countries is becoming less important in the formation of oil prices, Huang and Guo (2007), Ferraro et al. (2015) 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.

In order to start analyzing the data and building 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 the Figure 1.

As can be seen 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.

We will check the series for stationarity.

The series is not stationary (Table 2), the probability value P = 0.6137, we can not reject the hypothesis of the presence of a unit root; therefore, the series is not stationary. In order to get rid of nonstationarity, we check the series for the first difference.

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 (Figure 2).

Now we will carry out similar descriptive statistics for explanatory variables: Gas and gold. We now turn to a description of a number of gas prices.

From this histogram, (Figure 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 nonstationarity of the data series (Table 4).

The value P > 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, we bring the series to a stationary form (Table 5).

And, finally, let's move on to the last of the series - gold prices.

Table 1: Description of variables

Factor	Variable
Brent crude oil price	Oil
The price of gold	Gold
Price gas	Gas
Dummy variables:	
World financial crisis	MFC
Military company in Iraq	Iraq
Military company in Iran	Iran
Military company in Syria	Syria
Military company in Afghanistan	Afghanistan
The US Terror	Terror

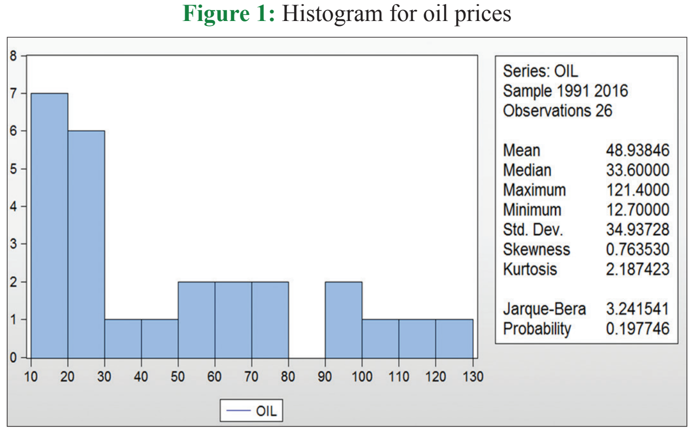
Source: Authors' calculation

Table 2: Test for stationarity of a number of oil prices

Test	t-Statistic	P
Augment Dickey-Fuller test statistic	−1.298902	0.6137

Source: Authors' calculation

Figure 1: Histogram for oil prices



Source: Authors' calculation

Table 3: Test for stationarity of a number of oil prices

Test	t-Statistic	P
Augment Dickey-Fuller test statistic	−4,121776	0.0041

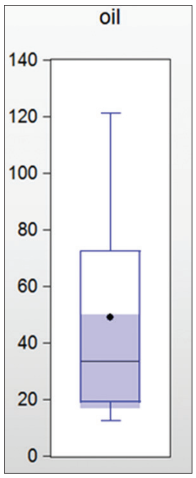
Source: Authors' calculation

Table 4: Stationary test of a number of gas prices

Test	t-Statistic	P
Augment Dickey-Fuller test statistic	−1,855319	0.3467

Source: Authors' calculation

Figure 2: Schedule BoxPlot for oil prices



Source: Authors' calculation

From this histogram (Figure 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 $P < 0.05$, therefore, we can reject the hypothesis of the presence of a unit root, thereby confirming the stationarity of the series (Table 6).

Also, to complete the descriptive analysis, it is necessary to check the series for the presence of emissions. To do this, we built BoxPlot graphics.

According to the graphs (Figure 5), we see that the gold variable GOLD has no emissions, which can not 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 (Figure 6), we can say that all our series are stationary - the correlogram decreases from the germ k after the first values. In addition, there is no periodic component in each of the series of

Table 5: Test for stationarity of a number of gold prices

Test	t-Statistic	P
Augment Dickey-Fuller test statistic	−5,524744	0.0002

Source: Authors' calculation

Table 6: Value of the corresponding probabilities for the regression variables

Variable	Coefficient	P
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
Crisis	−9,544256	0,1353
TERROR	−7,998058	0,4246

Source: Authors' calculation

data, which tells us that there is no seasonality. In order not to encounter the phenomenon of multicollinearity 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 them in such a way that in case of conflict our variable took the value 1, 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.

Now let's go directly to 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.

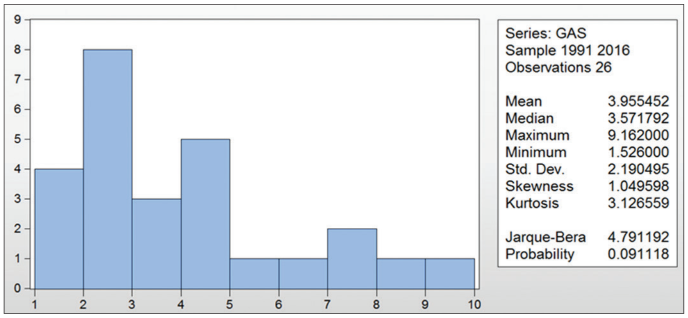
Table 6 shows the values of the coefficients and probabilities for each of the variables included in the constructed model.

From the above table, we can conclude that the variables D (gas), D (gold) and Iraq are significant (Table 7)- this tells us that they have an effect on our explained - the price of oil. While the probabilities of the rest are >0.05, which indicates their insignificance. That is, there is no correlation between these variables and the oil price variable - OIL. A more detailed table obtained in the construction of the model can be seen in the (Appendix 1).

The results obtained can be interpreted as follows:

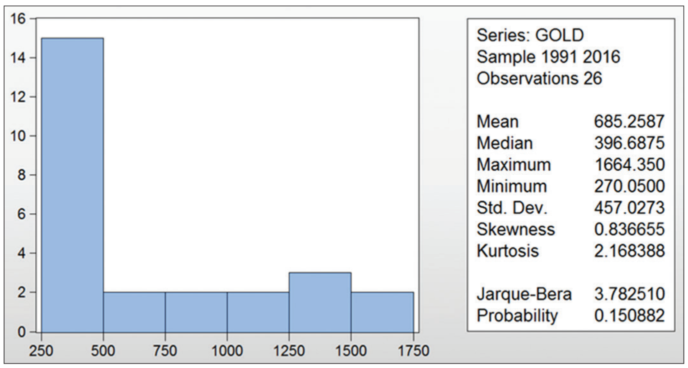
The variable GAS was significant, i.e., rising or falling in gas prices leads to changes in oil prices. This can be explained by the fact that each of these types of energy resources is very widely used and

Figure 3: Histogram for gas prices



Source: Authors' calculation

Figure 4: Histogram for gold prices



Source: Authors' calculation

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 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.

As for gold, here we can not 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.

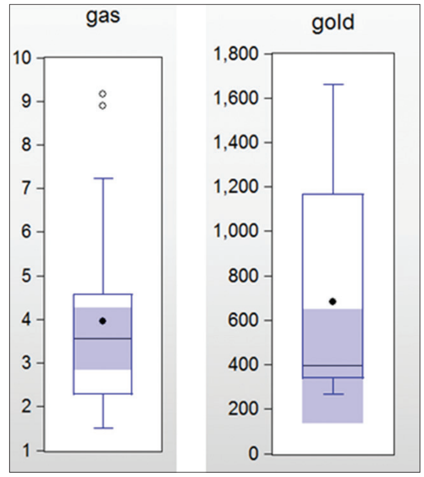
Of all the fictitious variables, only IRAQ was significant, a conflict that began in December 2004. It can be said that the significance of military conflicts in the oil-producing countries has an ever-smaller and insignificant effect on the price of oil. Thus, we see that over time, in fact, one factor increases in importance, while others decrease.

In order to correctly estimate the model constructed, we carry out the Ramsey test (Table 7).

According to the values of F-Statistic and Prob. Presented in Figure 8, we can conclude that the hypothesis of the acceptability of the functional form is adopted, that is, this model is correctly specified.

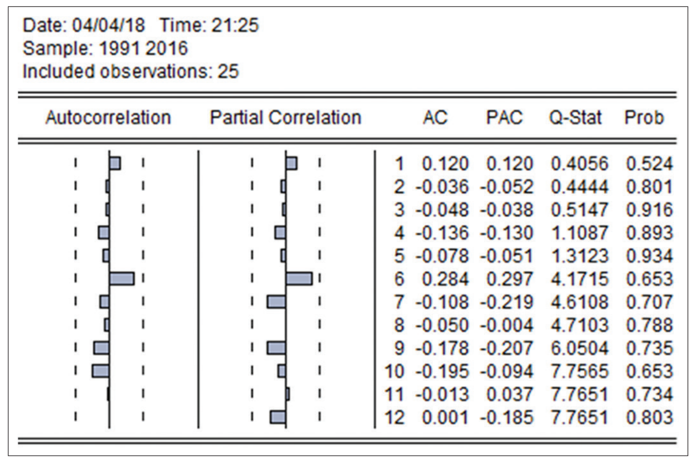
Now, to get a more accurate model, we conducted a test for extra variables see (Appendix 2).

Figure 5: Pox plot chart for gas and gold prices



Source: Authors' calculation

Figure 6: Correligram of the price of oil



Source: Authors' calculation

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 correlogram (Appendix 3) and eliminating the extra variables, we constructed the following model (Figure 7).

As can be seen from Figure 7, all variables are significant, low probabilities tell us this (Prob). 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 correlogram.

There are statistically significant effects of lags (4, 10) to reflect their influence; significant AR and MA processes - AR (4) and MA (10) - were included in the regression.

Now we will check the model for heteroscedasticity. To do this, we use the Breusch-Pagan test (Figure 8).

Probability value Prob. F = 0.6828 in Figure 8, which tells us that there is no heteroscedasticity. We also conducted a Glaser test (Figures 9 and 10).

In Figure 8, the probability value Prob.F = 0.4981, the hypothesis of homoskedasticity is confirmed.

An important point is the analysis of the residuals, that is, the deviations of the observed values from the theoretically expected ones. Having constructed a histogram (Appendix 4), we can conclude that the hypothesis that the remainders of this regression are subject to a normal distribution does not deviate by 5%.

For the construction of the forecast of oil prices, let us consider the time series OIL, which contains the prices for Brent crude oil in the period from 1991 to 2016. The series was initially nonstationary, but using the first-difference method, we brought it to a stationary form. To obtain the forecast for a number of oil price data, we have chosen the model of the mixed autoregressive moving average - ARIMA. This model for annual data will help to identify the time structure in an already existing series of these oil prices, and then it will be used to forecast prices for the next months.

Now let's proceed directly to the construction of the forecast. In this case, we built a retro forecast, in order to more accurately determine the accuracy of calculations and the correctness of the chosen model. To do this, we will reduce the number of observations by 2, that is, now our sample will be 2 years less (1991 and 1992 are removed years). And we will build a forecast for the period from 1991 to March 2016.

To determine the order of AR and MA, it is necessary to construct a correlogram from a number of data and analyze the knocking out lags.

Based on this correlogram in Figure 11, we can assume that there is a process AR (6) and MA (1). This is indicated by the lagging outbreaks in the PAs and AUs, respectively.

Now we construct the model ARIMA (6, 1, 1), the first value of 10 and the last 6 refer to the order of AR and MA processes respectively, while 1 in the middle indicates that we take the series in the differences.

From the above model, we can conclude that we were not wrong with the definition of the order of AR and MA processes. Zero probabilities Prob. = 0.0000 <0.05, indicate the significance of the variables of the model. Now it is necessary to check the obtained model for autocorrelation (Figures 12 and 13).

Analyzing the correlogram presented in Figure 13, one can say that there are no lobes out of the way, which indicates the correctness of the choice of orders for the AR and MA processes.

Thus, we can write the final model in the following form:

$$Y_t = -0.206567 + 0.437178 Y_{t-1} + 0.208182 Y_{t-2} \tag{9}$$

Table 7: Ramsey test

Test	F-Statistic	P. F (1, 15)
Ramsey RESET Test	2,123687	0.9697

Source: Authors' calculation

Figure 7: Regression model

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(GAS)	6.044392	1.506992	4.010899	0.0007
D(GOLD)	0.105575	0.020595	5.126218	0.0001
IRAQ	-6.189673	3.372827	-1.835159	0.0322
AR(4)	-0.569899	0.396904	-1.435861	0.0073
MA(10)	-0.635903	1.072926	-0.592681	0.0004
SIGMASQ	35.03806	23.60888	1.484105	0.1542

R-squared	0.865946	Mean dependent var	0.960000
Adjusted R-squared	0.830669	S.D. dependent var	16.50040
S.E. of regression	6.789897	Akaike info criterion	7.128503
Sum squared resid	875.9514	Schwarz criterion	7.421033
Log likelihood	-83.10628	Hannan-Quinn criter.	7.209638
Durbin-Watson stat	1.347336		

Source: Authors' calculation

Figure 8: Check for heteroscedasticity

F-statistic	0.682893	Prob. F(3,21)	0.5724
Obs*R-squared	2.222122	Prob. Chi-Square(3)	0.5276
Scaled explained SS	3.618894	Prob. Chi-Square(3)	0.3057

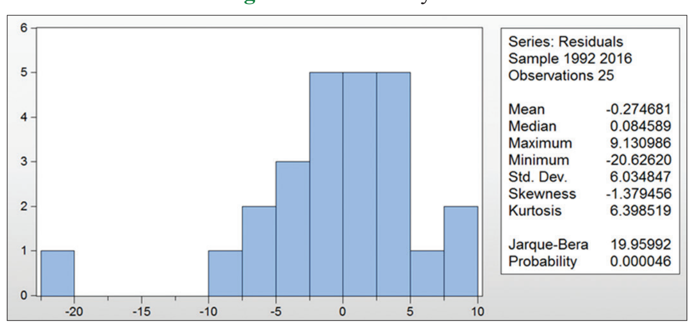
Source: Authors' calculation

Figure 9: Test for heteroscedasticity

F-statistic	0.498176	Prob. F(3,21)	0.6875
Obs*R-squared	1.660991	Prob. Chi-Square(3)	0.6456
Scaled explained SS	1.778427	Prob. Chi-Square(3)	0.6196

Source: Authors' calculation

Figure 10: Normality test

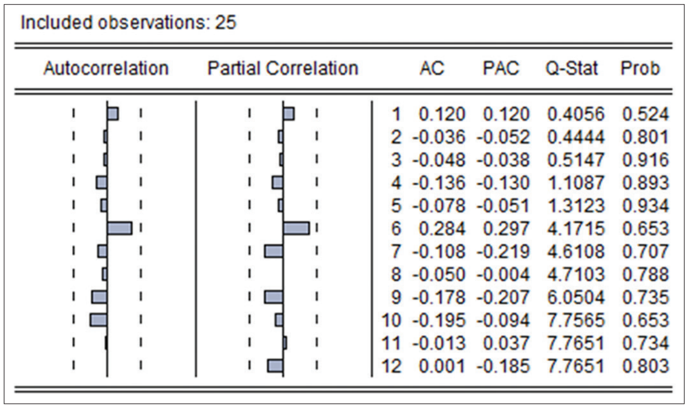


Source: Authors' calculation

The obtained values of the ARIMA model are close to the values of the initial series of oil prices. The obtained forecast values slightly exceed the initial ones, but are still close to them (Figure 14).

Figure 14 clearly shows that the constructed model follows the trend of the series. The forecast is also good except for one moment when the real prices for Brent crude are falling; the forecast does not give the same low values. This can be explained by the fact that this model does not take into account the influence of external factors, such as a crisis or position on the market.

Figure 11: Correlogram of oil time series in variances



Included observations: 25

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.120	0.120	0.4056	0.524
		2	-0.036	-0.052	0.4444	0.801
		3	-0.048	-0.038	0.5147	0.916
		4	-0.136	-0.130	1.1087	0.893
		5	-0.078	-0.051	1.3123	0.934
		6	0.284	0.297	4.1715	0.653
		7	-0.108	-0.219	4.6108	0.707
		8	-0.050	-0.004	4.7103	0.788
		9	-0.178	-0.207	6.0504	0.735
		10	-0.195	-0.094	7.7565	0.653
		11	-0.013	0.037	7.7651	0.734
		12	0.001	-0.185	7.7651	0.803

Source: Authors' calculation

Figure 12: Modified auto-regressive integrated moving average model (6, 1, 1)

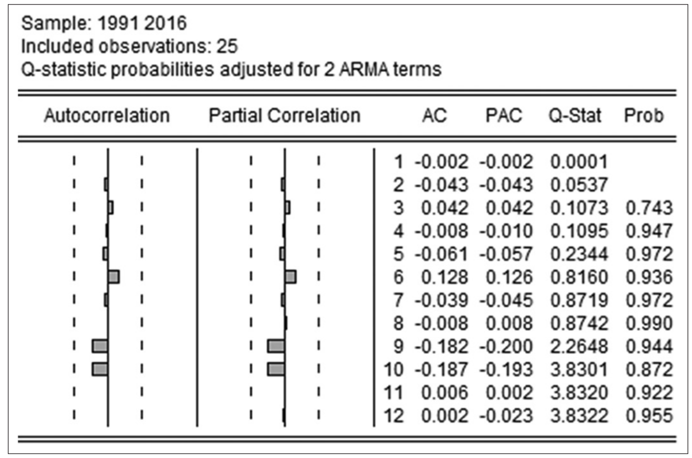
Included observations: 25
Convergence achieved after 41 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.206567	7.367148	-0.028039	0.9779
AR(6)	0.437178	0.177194	2.467225	0.0223
MA(1)	0.208182	0.152919	1.361387	0.0078
SIGMASQ	209.0974	65.48440	3.193088	0.0044

R-squared	0.970003	Mean dependent var	0.960000
Adjusted R-squared	0.085718	S.D. dependent var	16.50040
S.E. of regression	15.77737	Akaike info criterion	8.553353
Sum squared resid	5227.436	Schwarz criterion	8.748373
Log likelihood	-102.9169	Hannan-Quinn criter.	8.607444
F-statistic	1.750034	Durbin-Watson stat	1.983456
Prob(F-statistic)	0.187563		

Source: Authors' calculation

Figure 13: Correlation of residuals for modified auto-regressive integrated moving average model

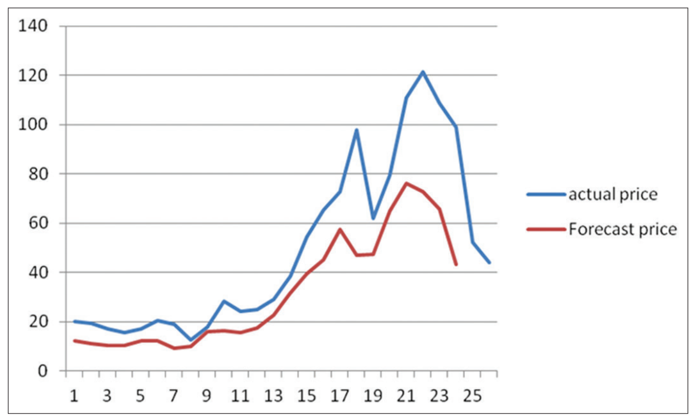


Sample: 1991 2016
Included observations: 25
Q-statistic probabilities adjusted for 2 ARMA terms

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.002	-0.002	0.0001	
		2	-0.043	-0.043	0.0537	
		3	0.042	0.042	0.1073	0.743
		4	-0.008	-0.010	0.1095	0.947
		5	-0.061	-0.057	0.2344	0.972
		6	0.128	0.126	0.8160	0.936
		7	-0.039	-0.045	0.8719	0.972
		8	-0.008	0.008	0.8742	0.990
		9	-0.182	-0.200	2.2648	0.944
		10	-0.187	-0.193	3.8301	0.872
		11	0.006	0.002	3.8320	0.922
		12	0.002	-0.023	3.8322	0.955

Source: Authors' calculation

Figure 14: Backtesting of modified auto-regressive integrated moving average model



Source: Authors' calculation

Namely, at the moment of falling real prices, which we see on the chart during our forecast - in June 2012 - the official price of Brent oil fell to a minimum in 17 months. The reason for this was the weak demand for oil futures, which was caused by poor data on the state of the labor market in the USA.

The accuracy of the approximation is indicated by the index R^2, which we calculated and whose value for this forecast model is 0.97, which indicates a good explanatory ability of the model.

5. CONCLUSION

In this paper, two main aspects 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).

It is also worth mentioning about 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. Analyzing the results and scientific articles on this topic, we came to the following conclusions. At the beginning of our study, we assumed that the price of gold will affect the price of oil, as an alternative source of investment of funds. And our hypothesis was not confirmed. This is explained by the fact that, the popularity of investing in precious metals, does not impact on investments in shares of oil companies.

And, finally, we hypothesized that military conflicts in the Middle East (in areas of oil production) affect the price of oil. But among the largest armed clashes that we have identified, only one was significant. This suggests that the significance of military operations in the oil-producing countries has an ever-smaller and insignificant effect on the price of oil. Now we need to move on to the next aspect, considered in this paper - the forecasting of oil prices. As the main methods of forecasting, we create the modification of ARIMA model.

The constructed retro-forecast, also turned out to be close to the real values of oil prices. The only point that did not take into account the model in constructing the forecast was a sharp drop in prices caused by instability in the oil futures market. However, this model can not take into account the influence of external factors, such as a crisis or position on the market.

Analyzing the results and comparing the accuracy of the models, we came to the conclusion to build the forecast for 2014 using the modification of ARIMA model.

This forecast showed that oil prices in 2014 will have a slight upward trend and will generally be stable. Looking at the forecasts for oil prices in 2014, which were already conducted by other researchers, we noted that analysts predict the growth of the economy of China and the US, the world's largest oil consumers. 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, so 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.

REFERENCES

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. Available from: http://www.csae.ox.ac.uk/conferences/2009-EDIA/Papers/252-Akpan.Pdf

Amano, R.A., Van Norden, S. (1998), Exchange rates and oil prices. Review of International Economics, 6, 683-694.

Backus, D.K., Crucini, M. (2000), Oil prices and the terms of trade. Journal of International Economics, 50, 185-213.

Barsky, R.B., Kilian, L. (2004), Oil and macro economy since the 1970s. Journal of Economic Perspectives, 18, 115-134.

Baumeister, C., Peersman, G. (2008), Time-Varying Effects of Oil Supply Shocks on the US economy. Available from: http://www.dx.doi.org/10.2139/ssrn.1093702.

Buetzer, S., Habib, M., Stracca L. (2012), Global Exchange Rate Configurations: Do Oil Shocks Matter? Working Paper, European Central Bank.

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

Ferraro, D., Rogoff, K., Rossi, B. (2015), Can Oil Prices Forecast Exchange Rates? Available from: http://www.econ.upf.edu/~brossi/OilAndExchangeRates.pdf.

Hooker, M. (1996), Oil prices and the rise and fall of the US real exchange rate. Journal of Monetary Economics, 25, 195-213.

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, 11, 205-211.

Huang, Y., Guo, F. (2007), The role of oil price shocks on China's real exchange rate. China Economic Review, 18, 403-416.

Kilian, L. (2009), Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market. American Economic Review, 99, 1053-1069.

Kilian, L., Murphy, D.P. (2014), The role of inventories and speculative trading in the global market for crude oil. Journal of Applied Economics, 29, 454-478.

Li, H., Wells, M.T., Yu, C.L. (2008), Bayesian analysis of return dynamics with stochastic volatility and levy jumps. The Review of Financial Studies, 21, 2345-2378.

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, 62-69.

Mikhailov, A. (2018), Pricing in oil market and using probit model for analysis of stock market effects. International Journal of Energy Economics and Policy, 8(3), 43-53.

Mikhaylov, A. (2018), Volatility spillover effect between stock and exchange rate in oil exporting countries. International Journal of Energy Economics and Policy, 8(3), 321-326.

Morana, C.A (2001), Semiparametric approach to short-term oil price forecasting. Energy Economics, 23, 325-338.

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.

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

Segal, P. (2011), Oil price shocks and the macro economy. Oxford Review of Economic Policy, 27, 169-185.

Singer, E. (2007), Oil Price Volatility and the US Macroeconomy: 1983-2006. Working paper, Minnesota of USA: Carleton College.

Tuzova, Y., Qayum, F. (2016), Global oil glut and sanctions: The impact on Putin's Russia. Energy Policy, 90, 140-151.

APPENDICES

Appendix 1

Variable	Coefficient	Standard Error	t-Statistic	P
C	−2.512660	2.574510	−0.975976	0.3436
D (GAS)	6.336001	1.250799	5.065564	0.0001
D (GOLD)	0.103214	0.021180	4.B73226	0.0002
IRAN	3262685	7.709218	0.423219	0.6778
IRAQ	−11.17840	6.589602	−1.696369	0.0092
AFGHANISTAN	9.845998	7.559718	1.302429	0.2112
CRISIS	−9.544256	6.067232	−1.573082	0.1353
TERRACT	−7.998053	9.760216	−0.819455	0.4246
SYRIA	−7.859139	4.823225	−1.629437	0.1227

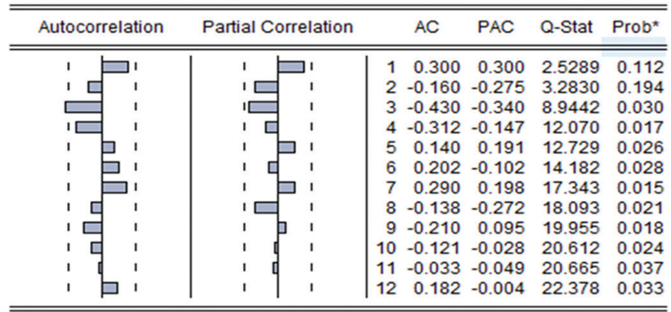
Appendix 2

Ramsey RESET test
Equation: UNTITLED
Specification: D (OIL>C D (GAS) D (GOLD>IRAN IRAQ AFGHANISTAN CRISIS TERRACT SYRIA
Omitted variables: Squares of fitted values

Variable.	Value	df	P
t-statistic	1.45728	1.5	0.1657
F-statistic	2.123637	(1,15)	0.9697
Likelihood ratio	3.310313	1	0.0688
F-test summary			Mean squares
Variable.	**Sum of sq.**	**df**	
Test SSR	122.791	1	122.7918
Restricted SSR	990.0932	16	61.88083
Unrestricted SSR	367.3014	15	57.32009
LR test summary			
Variable.	**Value**	**df**	
Restricted LogL	−81.46000	16	
Unrestricted LogL	79.30485	15	

Unrestricted test equation:
Dependent variable: D (OIL)
Method: Least squares
Date: 04M5TIB time: 10:23
Sample: 1992 2016
Included observations: 25

Appendix 3



Date: 04/05/18 Time: 11:18
Sample: 1991 2016
Included observations: 25
Q-statistic probabilities adjusted for 3 dynamic regressors

	AC	PAC	Q-Stat	Prob*
1	0.300	0.300	2.5289	0.112
2	-0.160	-0.275	3.2830	0.194
3	-0.430	-0.340	8.9442	0.030
4	-0.312	-0.147	12.070	0.017
5	0.140	0.191	12.729	0.026
6	0.202	-0.102	14.182	0.028
7	0.290	0.198	17.343	0.015
8	-0.138	-0.272	18.093	0.021
9	-0.210	0.095	19.955	0.018
10	-0.121	-0.028	20.612	0.024
11	-0.033	-0.049	20.665	0.037
12	0.182	-0.004	22.378	0.033

*Probabilities may not be valid for this equation specification.

Appendix 4

Redundant Variables Test
Null hypothesis: D (GAS) D (GOLD) IRAN IRAQ AFG AN I STAN CRISIS TER. Equation: UNTITLED
Specification: D (OIL) C D (GAS) D (GOLD) IRAN IRAQ AFGHANISTAN CRISIS TERRACT SYRIA
Redundant Variables: IRAN AFGHANISTAN CRISIS TERRACT SYRIA

Variable.	Value	df	Probability
F-statistic	11.19940	(8, 16)	0.0000
Likelihood ratio	47.17561	8	0.0000
F-test summary:			
Variable.	**Sum of Sq.**	**df**	**Mean squares**
Test SSR	5544.227	8	693.0283
Restricted SSR	6534.320	24	272.2633
Unrestricted SSR	990.0932	16	61.88033
LR test summary:			
Variable.	**Value**	**df**	
Restricted LogL	−105.0478	24	
Unrestricted LogL	−81.46000	16	

Restricted Test Equation:
Dependent Variable: D (OIL)
Method: Least Squares
Date: 04/05TIS Time: 10:24
Sample: 1992 2016
Included observations: 25