



## Oil Price Predictors: Machine Learning Approach

Jaehyung An<sup>1</sup>, Alexey Mikhaylov<sup>2\*</sup>, Nikita Moiseev<sup>3</sup>

<sup>1</sup>College of Business, Hankuk University of Foreign Studies, Seoul, Korea, <sup>2</sup>Department of Financial Markets and Banks, Financial University Under the Government of the Russian Federation, Moscow, Russia, <sup>3</sup>Department of Mathematical Methods in Economics, Plekhanov Russian University of Economics, Moscow, Russia. \*Email: [ayumihajlov@fa.ru](mailto:ayumihajlov@fa.ru)

Received: 23 January 2019

Accepted: 24 June 2019

DOI: <https://doi.org/10.32479/ijeeep.7597>

### ABSTRACT

The paper proposes a machine-learning approach to predict oil price. Market participants can forecast prices using such factors as: US key rate, US dollar index, S and P500 index, Volatility index, US consumer price index. After analyzing the results and comparing the accuracy of the model first, we can conclude that oil prices in 2019-2022 will have a slight upward trend and will generally be stable. At the time of the fall in June 2012 the price of Brent fell to a minimum of 17 months. The reason for this was the weak demand for oil futures, which was caused by poor data on the state of the US labor market.

**Keywords:** Oil Price Shocks, Economic Growth, Oil Impact, Factors, Dollar Index, Inflation, Key Rate, Volatility Index; S and P 500 Index

**JEL Classification:** C51, C58, F31, G12, G15

## 1. INTRODUCTION

For many years, oil has remained one of the most important sources of energy. All countries are consumers of oil and oil products. The prices of oil and its derivatives are of interest to both producers and consumers.

The dynamics of oil prices affect the level of costs in all sectors of production. The economy of many countries is based on oil production and trade in oil and oil products, therefore, forecasting oil prices is an important task. It is also worth noting that some sectors of the economy are directly dependent on oil prices.

Oil prices affect political and economic processes that determine the value of oil companies' stocks, the rate of inflation in the countries that import oil, and the rate of economic growth. It is important to note the impact of oil prices on the pricing of alternative energy sources.

The purpose of this work is to identify factors affecting the price of oil and the creation of algorithm and machine learning based on a modified linear regression.

To achieve this goal it is necessary to complete a number of tasks:

- Study of factors affecting oil prices
- Consideration machine learning algorithm based on modified linear regression
- Descriptive data analysis.

The average annual volume of oil consumption in 2014-2019 is about 4.2 billion tons, which is 54% more than in 1974-1979. Thus, the average increase in oil consumption from the time of the oil shock was ~1%/year.

At the same time, after the economic crisis of 1973-1983, oil consumption rose steadily until the 2008 crisis. However, there is a widespread view that significant and unexpected fluctuations in oil prices have a negative impact on the well-being of both

oil-importing countries and the oil-producing countries (Backus and Crucini, 2000).

The price of oil is one of the key factors determining the Russian budget in terms of its revenues. The current practice of determining the forecast price for oil is based on the method of building a consensus forecast (Mikhaylov, 2018a).

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), World Bank, IHS Global Insight, Raiffeisen Bank, International Monetary Fund.

The following disadvantages should be attributed to this approach.

1. The closed nature of forecasting methods, on the basis of the results of which consensus forecasts are made.

Each prediction method has certain drawbacks; the closed nature of the methods used is not possible to evaluate the degree of the prediction error. Using the consensus and the results obtained from different sources can lead to “inherit” the drawbacks source about prognosis (Mikhaylov, 2018b).

2. On the other hand, when the initial estimates based on specific assumptions and hypotheses, allowing to obtain an acceptable prediction, the use of consensus prediction and actually eliminates result, distorting the original value of predictions obtained from other sources.

Analysis of the practice of building predictive estimates and forecasting methods used by various scientific organizations, s, government, companies and the pits showed that by far the most in demand approaches of machine learning based on econometric forecasting methods.

In this regard, as an alternative to consensus forecasting proposed to use the method of prediction based on machine learning (Nyangarika et al., 2018).

In addition, some sectors of the economy are directly dependent on the forecast of oil prices. For example, airlines that rely on price projections for flights, the auto industry, and simply homeowners who rely on oil price forecasts to model purchases of durable goods, such as cars or home heating systems.

Oil prices and the volatility of oil prices play a significant role in the global economy, although the effects are asymmetric depending on periods, regions, sectors, and causes (Nyangarika et al., 2019a; Nyangarika et al., 2019b).

## 2. LITERATURE REVIEW

Different opinions were expressed on the impact of changes in oil prices on the world economy.

During this discussion, several studies have found that higher oil prices have an adverse effect on the economy (Akpan, 2009; Amano and Van Norden, 1998).

In addition, the economic impact on oil of importing countries, such as South Korea, has been discovered by Baumeister and Peersman (2008).

In order to make appropriate decisions about the direction, it is therefore important to accurately predict future oil prices using effective models of Sadorsky (1999), Barsky and Kilian (2004), Kilian (2009), Segal (2011), Morana (2013) and Kilian and Murphy (2014).

In June 2008, world oil prices, which showed an upward trend since 2003, rose to \$ 134. Oil prices fell after the global economic crisis of 2008, but began to rise again in early 2009 (Farzanegan and Markwardt, 2009).

Buetzer et al. (2012) have suggested a possible explanation for such a predicted slowdown in oil demand growth, such as structural changes in the world economy, consumer response, and government policy.

After the OPEC decided to keep oil production in 2014, the price of oil fell to less than \$ 50 a barrel. The price remains familiarize Island at \$ 40/bbl spruce, despite sluggish demand for oil and shale in 2015 and 2016 (Tuzova and Qayum, 2016).

Presumably, since 2009, oil price volatility has been rising. In this context, the long-term trend oil prices are important for ensuring future economic stability in many countries, as changes in crude oil prices and unstable oil supplies can seriously affect their economies, which depend on the import and export of crude oil (Singer, 2007; Olomola and Adejumo, 2006).

Complex models are able to reliably predict long-term oil prices and provide updated information based on changing market conditions to all interested parties, thereby contributing to sensible decision-making by politicians and heads of companies (Morana, 2013).

To reflect this time series was used machine learning algorithm based on standard linear regression. The study used weekly data prices for Brent crude from December 2013 to December 2018.

In addition, this approach uses the relationship between the futures price and the spot price for oil in the short-term forecast (Hooker, 1996).

Estimating oil prices is necessary in order to increase producers' incomes in a completely competitive environment using dynamic optimization.

Many researchers explain so changes in oil prices, often s approach is difficult to apply to the actual data. Often considered are factors contributing to fluctuations in oil prices, which depend on manufacturers.

The Delphi approach, which repeatedly collects opinions to obtain jointly subjective expert opinion, can also be used to predict oil prices.

As a forecasting methodology proposed to use the prices determined on the fth chersnom oil market. This approach checks whether the futures price is an objective predictor of the spot price at maturity. The researchers used WTI spot and futures prices from July 2000 to June 2004 as sample data (Mikhaylov et al., 2018).

In the selection of the forecast period, which would give the most accurate forecasts, by comparing quarterly forecasts, based on futures prices, ranges from daily to six months for WTI oil prices.

To evaluate the model accuracy It is likely to compare futures prices (1, 2, 3 and 4 months) with a by spot prices from 1991 to 2018.

It is possible to use futures prices from a certain point in time to forecast a spot prices by the Granger test between WTI spot and futures prices.

Previous studies on oil pricing models generally suggested that the current oil price trend will continue in the future, and thus factors affecting oil will have the same consequences in the future (Narayan and Sharma 2011).

However, factors affecting oil prices have changed structurally over time. In the 1960s, supply-side factors determined the price of crude oil, and this trend continued until the collapse of the mid-1980s.

Therefore, the price of crude oil can be determined by both demand and supply. In the 1990s such emerging markets as China and India, has led to an increase in oil prices.

Since 2000 financial factors, including the penetration of speculative forces, the weakening dollar and the financial crisis if attention was drawn to the quality factors impact on the world's oil prices.

Researchers have found that financial turmoil contributed significantly to the rise in oil prices in the early 200 's and, to a much greater degree, from the mid-2000s (Nandha and Faff, 2008).

Among several financial factors, the expectations of market participants were indicated as an important determining factor in the price of goods (Eltony and Al-Awadi, 2001).

However, the role of speculation in the event of significant changes in the price of oil, still, is controversial (Huang and Guo, 2007),

A number of studies do not confirm that speculation is an important factor in determining real oil prices. Despite the fact that the paradigm of the world oil market is constantly changing, previous forecasts are known models rarely reflect such structural changes (Ferraro et al., 2015).

Thus, this study can contribute to the preparation of fast and accurate measures on the oil market by predicting short-term oil prices. The model of this study is highly applicable.

The oil price forecast can be used to make informed decisions by the government and the private sector.

### 3. METHODS

In this work, a modified linear regression-based machine learning algorithm was used, which gives more accurate predictions of prices in the future.

The linear regression model has the following parameters. Below, basic information is displayed to check the adequacy of the data in the established model (Table 1).

For example, fitting a linear model to data constructed with two out of five missing predictors and without interception:

$$X = \text{randn}(100,5) \quad (1)$$

$$y = X * [1; 0; 3; 0; -1] + \text{randn}(100,1) \quad (2)$$

$$\text{mdl} = \text{fitlm}(x,y) \quad (3)$$

$$y \sim 1 + x_1 + x_2 + x_3 + x_4 + x_5 \quad (4)$$

Where randn - function of creating an array of random variables, x1 - US dollar index, x2 - S and P 500 index, x3 - US Federal Reserve rate, x4 -, Volatility index (VIX), x5 - consumer price index (CPI).

Further It is necessary to create an array of floating-point random numbers obtained from the normal distribution with an average value of 100 and a variance of 1.

The general theory of random variables states that if x is a random variable, then:

$$y = ax + b \quad (5)$$

$$\mu y = a\mu x + b \quad (6)$$

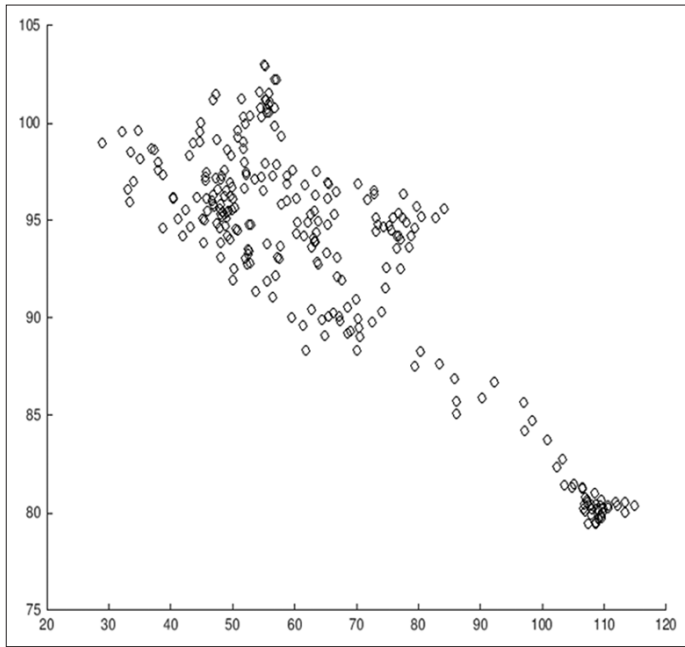
The best factors, which market participants can use for machine learning approach, are US key rate, US dollar index, S and P500 index, VIX index, US CPI (Figure 1-5).

**Table 1: Estimated coefficients**

Parameter	Estimate	Standard error	t stat	P value
(Intercept)	0.038164	0.099458	0.38372	0.70205
x1	0.92794	0.087307	10.628	8.5494e-18
x2	-0.075593	0.10044	-0.75264	0.45355
x3	2.8965	0.099879	29	1.1117e-48
x4	0.045311	0.10832	0.41831	0.67667
x5	-0.99708	0.11799	-8.4504	3.593e-13

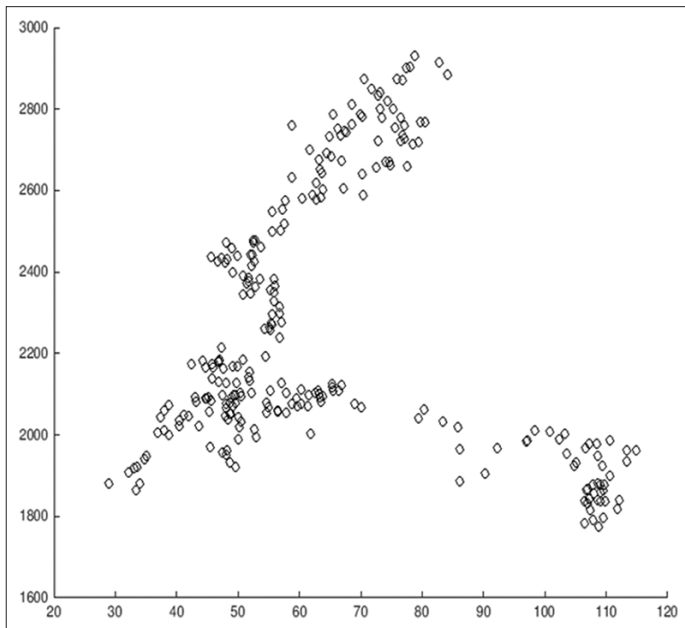
Source: Matlab, author's calculations

**Figure 1:** Distribution of oil prices and US dollar index



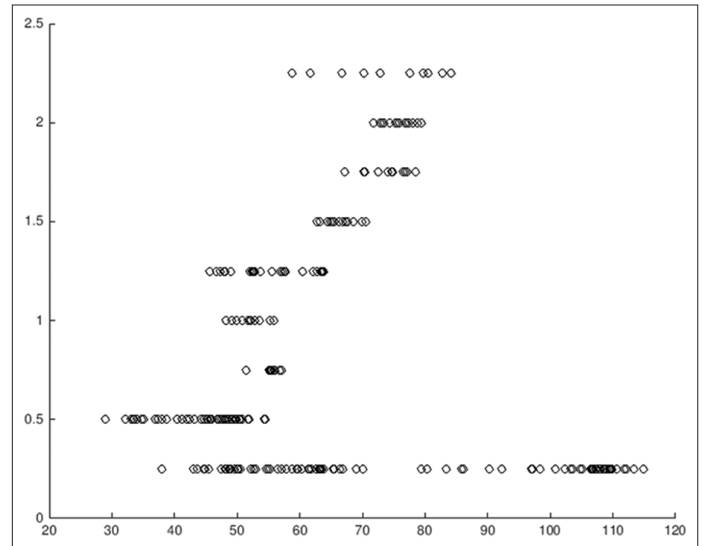
Source: Thomson Reuters Datastream

**Figure 2:** Distribution of oil prices and US key rate



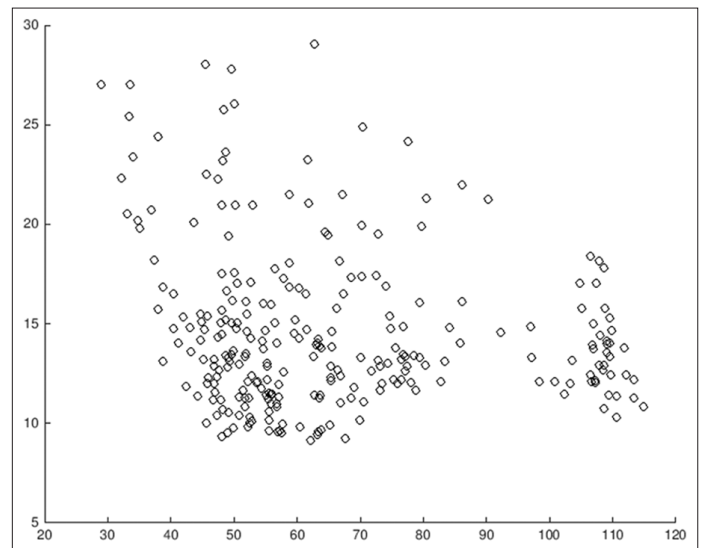
Source: Thomson Reuters Datastream

**Figure 3:** Distribution of oil prices and, Volatility index



Source: Thomson Reuters Datastream

**Figure 4:** Distribution of oil prices and US consumer price index



Source: Thomson Reuters Datastream

**Table 2. The simulation results of a linear regression algorithm**

Root mean square error	994.38
R-Square	0.92
MEE	9.888e+05
Mean absolute error	581.49
Prediction speed	~4200
Study time	2.3465 sec

Source: Matlab, author's calculations

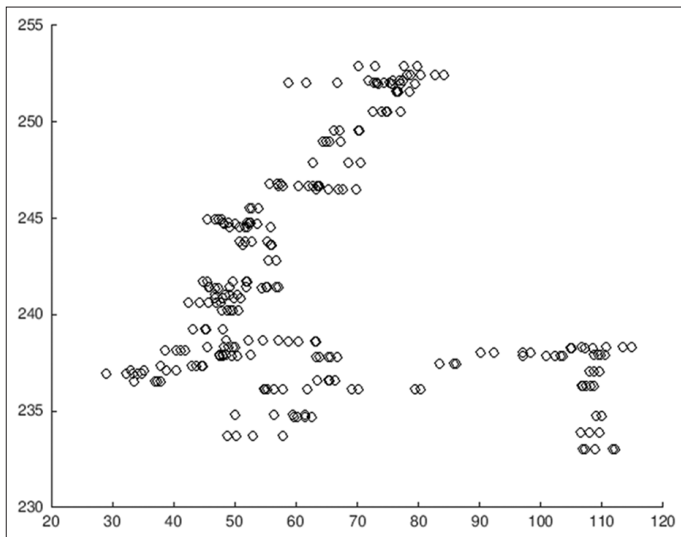
modified linear regression-based machine learning algorithm, it is advisable to use the following set of parameters (Figure 6).

As a result, the modification was able to improve the value of RMSE c 994.38, characteristic of standard linear regression (Table 2) to 927.76 (Table 3), as well rms error and mean absolute error. At the same time, the indicator R - square for the modified

### 3.1. Model Assumptions

- The display shows the estimated values of each coefficient in the evaluation column. These values are close to the true values [0; 1; 0; 3; 0; -1].
- There are standard errors for coefficient estimates.
- The reported values of P value (obtained from t statistics under the assumption of normal errors) for predictors 1, 3, and 5 are extremely small. These are the three predictors used to create the y data series.
- PValue for x2 and x4 much more e than 0.01. These three predictors are not used to create the y data series In the

**Figure 5:** Distribution of oil prices and US CPI



Source: Thomson Reuters Datastream

**Figure 6:** Parameters of the modified mode and linear regression

<input type="checkbox"/>	SplitCriterion	'mse'	1×3	char
<input type="checkbox"/>	MinParent	10	1×1	double
<input type="checkbox"/>	MinLeaf	4	1×1	double
<input type="checkbox"/>	MaxSplits	260	1×1	double
<input type="checkbox"/>	NVarToSample	'all'	1×3	char
<input type="checkbox"/>	MergeLeaves	'on'	1×2	char
<input type="checkbox"/>	Prune	'on'	1×2	char
<input type="checkbox"/>	PruneCriterion	'mse'	1×3	char
<input type="checkbox"/>	QEToler	1.0000e-06	1×1	double
<input type="checkbox"/>	NSurrogate	'off'	1×3	char
<input type="checkbox"/>	MaxCat	10	1×1	double
<input type="checkbox"/>	AlgCat	'auto'	1×4	char
<input type="checkbox"/>	PredictorSelection	'allsplits'	1×9	char
<input checked="" type="checkbox"/>	UseChisqTest	1	1×1	logical
<input type="checkbox"/>	Stream	[]	0×0	double
<input type="checkbox"/>	Version	2	1×1	double
<input type="checkbox"/>	Method	'Tree'	1×4	char
<input type="checkbox"/>	Type	'regression'	1×10	char

Source: Matlab, author's calculations

**Table 3: The simulation results of a modified linear regression algorithm**

Root mean square error	927.76
R-square	0.93
MEE	8.6074e+05
Mean absolute error	398.92
Prediction speed	~30000
Study time	0.29919 s

Source: Matlab, author's calculations

linear regression grew slightly from 0.92 to 0.93, which does not significantly affect the accuracy of the model.

The model also works when real Brent prices fall. Although it does not take into account the influence of external factors, such as a crisis in the market.

At the time of the fall in June 2012 the official price of Brent crude fell to a minimum of 17 months. The reason for this was the weak demand for oil futures, which was caused by poor data on the state of the US labor market.

On the accuracy of approximation indicates an index of R2, which reaches we thought and whose importance for this model is 0.97, indicating good predictive ability of the model.

## 4. CONCLUSION

In this paper, two main aspects were considered: Factors affecting the price of oil and methods for its price forecasting, using machine learning algorithms based on standard linear regression and modified linear regression.

The analysis revealed that the price of oil is most affected by the US Federal Reserve rates and the US dollar index.

It is also worth mentioning the factors that turned out to be insignificant in this model: The financial crisis, conflicts in Iran, Afghanistan, Syria, and the terrorist attacks that occurred in the Middle East and in the United States. Analyzing the results and scientific articles on this topic, we came to the following conclusion.

At the beginning of our study, we assumed that the price of gold would fall on the price of oil, as an alternative source of investment. But it did not confirm. This is explained by the fact that the popularity of investing in precious metals does not affect investments in oil stocks to the company.

The conflicts in the Middle East (in the areas of oil production) affect the price of oil is only short term.

Now we need to move on to the next aspect discussed in this paper – the projection of oil prices using machine learning algorithm based on a modified linear regression model.

The retro forecast also turned out to be close to real oil prices. The only thing that was not taken into account by the model when building the forecast, is about the sharp fall in prices, caused by instability in the market of oil futures. However, this model cannot take into account the influence of external factors, such as a crisis or market conditions.

After analyzing the results and comparing the accuracy of the model first, we can conclude that oil prices in 2019-2022 will have a slight upward trend and will generally be stable.

In this paper, not all the problems arising in forecasting oil prices were considered, therefore it is advisable to continue to consider various forecasting methods in the future so that the obtained values of the model parameters become as close to real as possible.

One of the directions for further research may be the use of other machine learning algorithms or the identification of new factors of influence for obtaining more accurate predictions.

## 5. ACKNOWLEDGEMENTS

This work was supported by Hankuk University of Foreign Studies Research Fund.

## 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.J. (2000), Oil prices and the terms of trade. *Journal of International Economics*, 50, 185-213.
- Barsky, R.B., Kilian, L. (2004), Oil and macroeconomy 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. In *American Economic Journal: Macroeconomics*, 59(4), 1-29.
- Buetzer, S., Habib, M., Stracca, L. (2012), Global Exchange Rate Configurations: Do Oil Shocks Matter? Working Paper, European Central Bank.
- Eltony, M.N., Al-Awadi, M. (2001), Oil price fluctuations and their impact on the macroeconomic variables of Kuwait: A case study using a VAR model. *International Journal of Energy Research*, 25(11), 939-959.
- 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.
- 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 Reviews*, 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.
- Mikhaylov, A. (2018a), Pricing in oil market and using probit model for analysis of stock market effects. *International Journal of Energy Economics and Policy*, 8(2), 69-73.
- Mikhaylov, A. (2018b), Volatility spillover effect between stock and exchange rate in oil exporting countries. *International Journal of Energy Economics and Policy*, 8(3), 321-326.
- Mikhaylov, A., Sokolinskaya, N., Nyangarika, A. (2018), Optimal carry trade strategy based on currencies of energy and developed economies. *Journal of Reviews on Global Economics*, 7, 582-592.
- Morana, K. (2013), The oil price-macroeconomy relationship since the mid-1980s: A global perspective. *Energy Journal*, 34, 153-190.
- Nandha, M., Faff, R. (2008), Does oil move equity prices? A global view. *Energy Economics*, 30, 986-997.
- Narayan, P.K., Sharma, S.S. (2011), New evidence on oil price and firm returns. *Journal of Banking and Finance*, 35, 3253-3262.
- Nyangarika, A., Mikhaylov, A., Richter, U. (2019b), Oil price factors: Forecasting on the base of modified auto-regressive integrated moving average model. *International Journal of Energy Economics and Policy*, 9(1), 149-160.
- Nyangarika, A., Mikhaylov, A., Richter, U. (2019a), Influence oil price towards economic indicators in Russia. *International Journal of Energy Economics and Policy*, 9(1), 123-129.
- Nyangarika, A., Mikhaylov, A., Tang, B.J. (2018), Correlation of oil prices and gross domestic product in oil producing countries. *International Journal of Energy Economics and Policy*, 8(5), 42-48.
- 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 macroeconomy. *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.