LITTERA SCRIPTA

Economics

Management

Corporate Finance

Finance and Valuation

1/2025

VŠTE

Littera Scripta

(Economics, Management, Corporate Finance, Finance and Valuation)

Ing. Jakub HORÁK, MBA, PhD. (Editor-in-chief)

Address Editor:

Institute of Technology and Business in České Budějovice

Okružní 517/10

370 01 České Budějovice, Czech Republic

Tel.: +420 387 842 183

e-mail: journal@littera-scripta.com

ISSN 1805-9112 (Online)

Date of issue: June 2025

Periodicity: Twice a year Since 2010

The Journal is indexed in:

- ERIH PLUS (European Reference Index for the Humanities and Social Sciences) – in 2015
- CEJSH (Central European Journal of Social Sciences and Humanities) – in 2015
- EZB (Elektronische Zeitschriftenbibliothek) – in 2017
- GOOGLE SCHOLAR – in 2017
- DOAJ (Directory of Open Access Journals) – in 2019

Content

Oil price trends and their prediction using the ARIMA model

Jakub Horák[1], Aneta Soběslavská[2]

[1] Institute of Technology and Business in Presov, Department of Economics, Management and Marketing, Slovakia
[2] Institute of Technology and Business in Ceske Budejovice, School of Expertness and Valuation, Czech Republic

Abstract

The price of oil is an extremely volatile commodity, the prediction of which is complex, but crucial for the stability of the global economy and the energy market. The aim of this work was to evaluate the development of the oil price over the past 25 years and to create a prediction model for estimating its price until 2030. For this purpose, the ARIMA method was used, applied to the time series of WTI oil prices obtained from the EIA database. Statistical tests of stationarity, autocorrelation and subsequent differentiation of the data were performed to correctly set the model. The result was a successful prediction of the future price development of oil with an expected range of 60–90 USD per barrel, with the ARIMA (2,1,1) model demonstrating a good ability to capture both historical trends and cyclical fluctuations. The contribution of the work was the creation of a reliable tool for economic planning in the energy sector. The research is limited by the fact that it does not take into account all external factors, such as geopolitical events or technological progress.

Keywords: Prediction, oil, ARIMA method, data stationarity, volatility, economics, energy

Introduction

Oil plays a key role in the global economy, accounting for one-third of global energy consumption. Despite its importance for policymaking and economic development, oil price prediction remains challenging due to its complexity and erratic price trends. Although a significant amount of research has been conducted to improve forecasts using external factors, machine learning, and deep learning, only a few studies have used hybrid models to improve prediction accuracy. (Kim & Jang, 2023) . Investments in the oil industry are usually associated with a high level of risk due to uncertainty caused by economic factors. Typical factors include oil and gas prices, interest rates, operating costs, and capital expenditures. In addition, investment risk increases with increasing offshore exploration, production, and production activities. Therefore, accurate prediction of economic factors is crucial for the oil and gas industry to make better strategic decisions with minimized risk (Naderi et al., 2019).

OPEC+, composed of the Organization of the Petroleum Exporting Countries (OPEC) and non-OPEC oil-producing countries, has a significant influence on the global oil market. However, the current literature lacks a comprehensive application of this factor in oil price prediction, mainly due to the complexity of measuring such political evolutions (Li et al., 2024).

High inflation targets continue to pose a challenge for macroeconomic stabilization policies in developing economies. Oil prices are considered a significant factor influencing inflation. Given the high and volatile international oil prices, the question of the relationship between inflation and oil prices, and its impacts on economic welfare, has become a crucial empirical issue (Ayisi, 2021) . Oil product prices respond to changes in oil prices with varying speed or intensity. The public perceives this as a relatively rapid response of oil product prices to increases in oil prices, as opposed to their adjustment when oil prices fall. If this asymmetric price transmission is present, it is a direct consequence of the behavior of entities in the energy sector that exploit oil price fluctuations to the detriment of customers. Although the existing literature is quite extensive, it is still predominantly focused on developed countries (primarily the US), with research findings showing significant differences (Cipcic, 2021).

The ongoing Russian-Ukrainian conflict has also had a significant impact on global energy dynamics, which has subsequently weakened the functional capacity of economies around the world (He, 2024) . Conflict caused worry regarding the stability of the global energy sector , given that Russia is a large player in this industry , produces and exports oil and natural gas gas all over world (Kot et al., 2024).

The aim of the thesis is to evaluate the development of oil prices over the last 10 years and to predict oil prices until 2030.

In connection with the objective, the following research questions are set:

The price of oil has experienced significant fluctuations since 2015, affecting the global economy. Examining these developments helps to understand the factors influencing the oil market and makes it easier to predict future prices.

VO1: How has the price of oil developed from 2000 to the present?

Oil price prediction is crucial for the economy, energy and investment decisions. Given the market instability, it is important to evaluate current economic indicators to better estimate the development of oil prices over the next 5 years.

VO2: How will the price of oil develop over the next 5 years?

The price of oil is one of the most important indicators of the global economy. Understanding these factors is crucial as it allows us to better predict price changes and make strategic decisions.

VO3: What factors most significantly influence the development of oil prices?

Methods and Data

The data used in this research comes from the Energy database Information Administration (EIA, 2024), which provides historical oil price data. Specifically, this will be a time series of data from January 15, 2000 to December 15, 2024. This data is monthly and contains historical WTI (West Texas Intermediate) crude oil prices in US dollars per barrel. The data will be used to apply the ARIMA method to predict future oil price developments based on historical values. All data are freely available and were checked for missing or anomalous values before being used for analysis.

AutoRegressive Inverse Model (method will be used for data analysis. Integrated Moving Average), which is often used for time series prediction. The goal is to use historical oil price data to create a model that will be able to predict future oil price developments. Before using the ARIMA model, basic statistics such as mean, median, mode, standard deviation, and variance will be calculated, which will allow a better understanding of the data structure.

The ARIMA model will be applied using RStudio, a powerful data analysis environment in the R language. Using the forecast package, an ARIMA model will be created that will be optimized for the best results based on historical data. The result will be a model capable of predicting oil prices based on a time series. The ARIMA model is defined by the following formula: (Hyndman & Athanasopoulos, 2018):

$$Y_t = \alpha + \sum_{i=1}^{p} \phi_i Y_{t-i} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} + \epsilon_t \tag{1}$$

where:

- Y_t is the value of the time series at time ttt,
- α is a constant (intercept),
- $\phi_i$are the parameters of the autoregressive part of AR,
- θ_j are the parameters for the moving average MA,
- ϵ_t is the error (residual) at time ttt,
- $p$is the order of the autoregressive part,
- $q$is the order of the moving average.

Before using the ARIMA model, a check for stationarity of the data will be performed, which is a key condition for the correct application of this model. In case the data are not stationary, they will be subjected to differentiation. The Dickey-Fuller test is commonly used to test for stationarity. It has the following hypothesis: (Hamilton, 1994)

- H_0 (null hypothesis): The time series has a unit root (is not stationary).
- H_1 (alternative hypothesis): The time series does not have a unit root (it is stationary).

If the p-value of this test is greater than 0.05, the data is considered nonlinear and will need to be adjusted (differentiated).

Another important formula in time series analysis is the differentiation model for achieving stationarity. If the time series is not stationary, differentiation is used to remove trends: (Hyndman & Athanasopoulos, 2018):

$$\Delta Y_t = Y_t - Y_{t-1} \tag{2}$$

where:

- ΔY_t is the first differentiation of the time series in time$t$
- Y_t is the value of the time series at time
- Y_{t-1} is the value of the time series at the previous time point$t - 1$

If the first differentiation is not sufficient to achieve stationarity, a second differentiation can be performed: (Hyndman & Athanasopoulos 2018):

$$\Delta^2 Y_t = \Delta Y_t - \Delta Y_{t-1} \tag{3}$$

where:

- $\Delta^2 Y_t$ je $d$reverse differentiation of a time series in time$t$
- ΔY_t je $p$smooth differentiation of a time series in time$t$
- ΔY_{t-1} je $první$ differentiation of a time series in the previous time$t - 1$

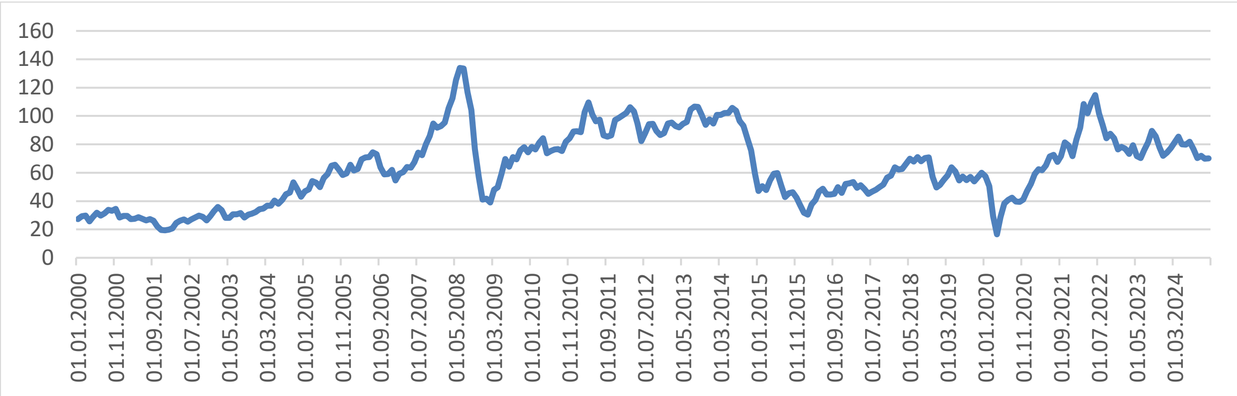
Next, the model will be optimized automatically using the auto.arima () function, which will select the best parameters for the model. Subsequently, oil prices for the next 5 years (until 2030) will be predicted and displayed on the graph.

The analysis is expected to result in a robust model that is capable of providing reliable oil price predictions based on historical trends. The results of this analysis will then be used to answer the research questions and fulfill the research objectives.

Results

A table was created to process historical data, from which the following graph (Figure 1) was created in RStudio, which shows the development of the oil price in the period between 2000 and 2024. As this Graph 1 shows, the price of oil has undergone significant changes over the past few decades. After a stable period in the early 2000s, when the price of oil fluctuated around 20-30 USD per barrel, there was a significant increase in the years 2004-2008, when the price rose to more than 140 USD per barrel. This growth was followed by a significant decrease in 2008 and a subsequent recovery in the years 2010-2014. After 2014, the price of oil fell below 30 USD per barrel again. In 2020, the price of oil experienced a historic decline, when short-term contracts even had negative values. In recent years, since approximately 2021, the price of oil has stabilized between $50 and $100 per barrel, with occasional fluctuations.

Figure 1: WTI oil price development in 2000-2024

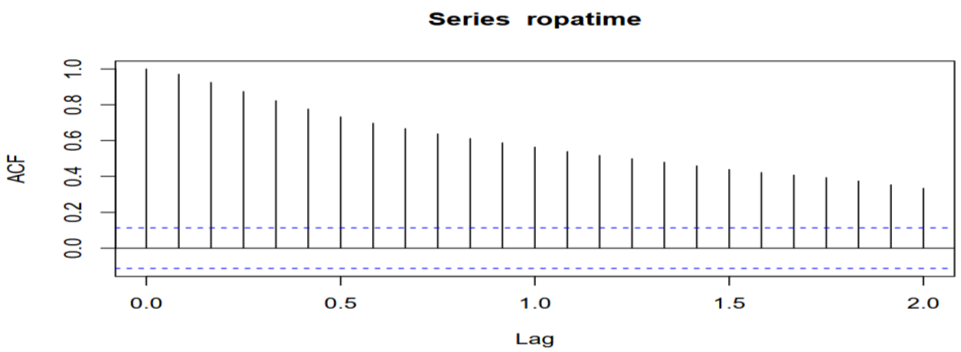


Source: Own processing based on data from Energy Information Administration (EIA,2024).

Next, the data was entered into RStudio, where ACF (autocorrelation) and PACF (partial autocorrelation) testing was performed.

The ACF (Figure 2) shows how the values of a time series depend on their previous values (lags). It is evident from this chart that for lower lags the autocorrelation is significant, but as lags increase the correlation decreases rapidly. This result suggests that the data show some dependence between the values, but this dependence weakens rapidly, which could mean that the data are not fully stationary, but that they are gradually leveling out.

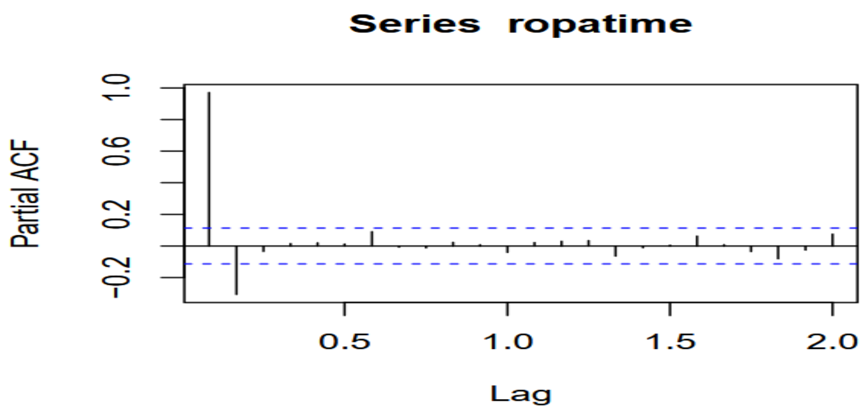
Figure 2: ACF of the original time series



Source: Own processing using RStudio.

The PACF (Figure 3) chart shows the partial autocorrelation, which measures the correlation between values at a time and their lagged values after removing the effect between previous lags. In this chart, we see that the values are mostly below the statistical significance threshold (blue lines), which means that there is no strong correlation between the values after cleaning up the effect of previous values.

Figure 3: PACF of the original time series



Source: Own processing using RStudio.

Based on these plots, it can be concluded that the data shows some autocorrelation, but after removing it (in the PACF), the values are not strongly correlated. This suggests that the data is not ideally stationary, but shows some weak autocorrelation.

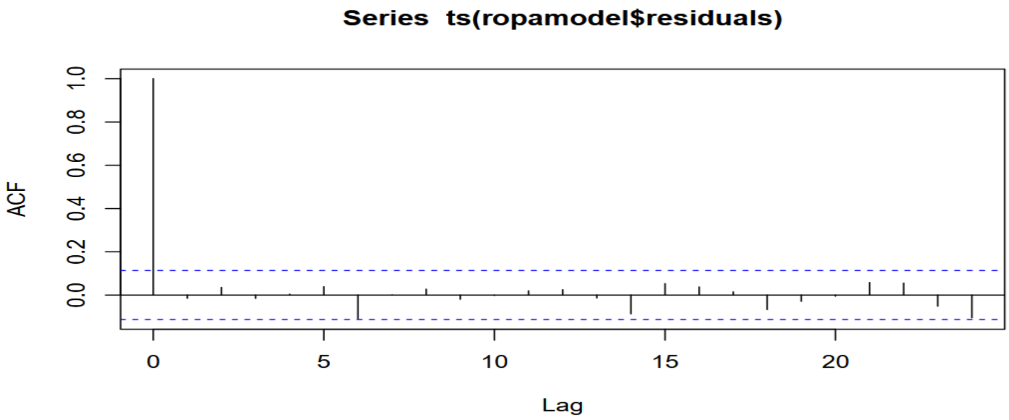
It was also carried out Augmented Dickey-Fuller (ADF) test. The ADF result indicates that the data is not stationary. The ADF test statistic was -2.4843, which indicates some degree of dependence between the values in the time series. However, the p-value of the test is 0.3721, which is higher than 0.05, and means that we cannot reject the null hypothesis that the time series has a unit root. This indicates that the data is non-stationary.

Thanks to these results, we know that the time series shows some trend or structure that causes a dependence between values over time. Therefore, to use this data in modeling, it is necessary to make adjustments so that the time series becomes stationary and can be properly analyzed and used for predictions.

After applying differentiation to the data to make it stationary, further stationarity checks were performed using ACF and PACF.

The ACF plot (Figure 4) shows that after differentiation, the autocorrelation is very weak and most of the values are below the statistical significance threshold (blue lines). This indicates that the model has estimated the structure of the time series well and the residuals do not contain any strong patterns.

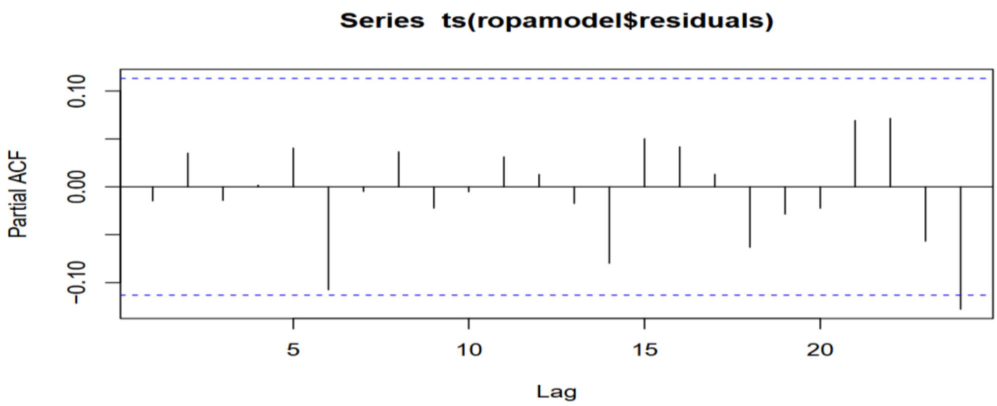
Figure 4: ACF adjusted time series



Source: Own processing using RStudio.

The PACF plot (Figure 5) shows that after removing the influence of previous lags, only a weak partial autocorrelation remains, confirming that the model adequately captures the structure of the data and no strong patterns remain. The data are stationary after differentiation. Based on these analyses, the ARIMA (2,1,1) model was recommended.

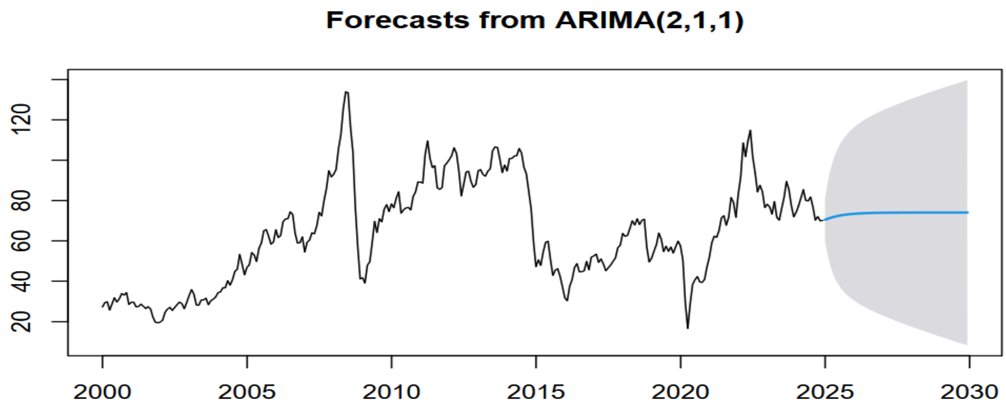
Figure 5: PACF adjusted time series



Source: Own processing using RStudio.

The final forecast for the WTI oil price was then made using the ARIMA (2,1,1) model. Figure 6 shows the historical oil price values along with the predicted future development up to 2030. The blue line shows the predicted values, while the grey area around this line shows the confidence interval of the prediction, which means the range in which the actual values are expected to move. The ARIMA (2,1,1) model is able to capture the main trend and cyclical fluctuations in the historical data, but the prediction into the future is burdened with some uncertainty, as seen in the form of a grey shadow. The model provides an estimate of the future development of the oil price, which should, however, be taken with caution due to the possibility of changes in external factors affecting the oil market.

Figure 6: Oil price forecast until 2030



Source: Own processing using RStudio.

Finally, a Box- Ljung test was performed to check whether the residuals show autocorrelation. This test is used to determine whether the remaining values after applying the model are random or whether there are any patterns in them that would indicate that the model was not properly fitted. The first result, where X- squared = 0.99601, df = 5 and p-value = 0.9629, indicates that the p-value is very high, which means that there is no significant autocorrelation in the residuals. Therefore, it can be said that the residuals are random and the model captures the data structure well. The second result with X- squared = 8.5349, df = 15 and p-value = 0.9006 again confirms that the p-value is very high, which means that the residuals do not show any significant autocorrelation.

Overall, these results confirm that the ARIMA (2,1,1) model is well-fitted and that the residuals are random. This suggests that the model captures the dynamics of the data well and can be considered reliable for predictions.

Discussion

VO1: How has the price of oil developed from 2000 to the present?

The development of the price of oil from 2000 to the present shows significant fluctuations, which is clearly visible from the historical price trends, as shown in the graph of the price of WTI oil. In the period between 2000 and 2008, the price of oil was stable, moving in the range of 20-30 USD per barrel. This stable development was characterized by relatively low demand and the absence of major geopolitical or economic shocks that could fundamentally affect prices on the oil market. However, in the years 2004-2008, there was a significant increase in the price, which rose to more than 140 USD per barrel. This growth was supported by strong global economic growth, especially in emerging economies such as China, which demanded more and more oil, and also contributed to the growth of speculative activities in commodity markets. After the peak in 2008, the price fell dramatically, which was related to the global financial crisis. The decline in oil prices at the beginning of the financial crisis was the result of a sharp economic slowdown, which led to a decline in demand for oil. This decline continued in 2020, when the world faced the COVID-19 pandemic, which led to an unprecedented decline in demand and oil prices. This year, the price of oil even briefly fell into negative values on futures markets due to excess supply and loss of demand. This development clearly shows the high volatility in the oil market and significant sensitivity to global economic and political factors. After a period of crises, from 2016 to the present, the price of oil has gradually increased, stabilizing between 50-100 USD per barrel, with occasional fluctuations. This trend is mainly attributed to the recovery of the global economy after the crisis and the growth in demand, as well as the decision of OPEC and other producers to limit production, which helped stabilize the market. However, this stabilization is subject to a certain degree of uncertainty, not only due to geopolitical factors (such as OPEC+ policy decisions), but also due to environmental and technical changes in the way of extraction, such as the increasing extraction of shale oil in the USA. This development is in line with the findings of Naderi et al. (2019), who observed similar fluctuations and trends when using the ARIMA model to analyze historical oil price data. Their study shows that even with the help of traditional statistical methods such as ARIMA, it is possible to capture long-term cyclical fluctuations and major crisis events that affect oil prices, which confirms the validity of the ARIMA model for this type of prediction.

VO2: How will the price of oil develop over the next 5 years?

The forecast of oil prices for the next 5 years using the ARIMA (2,1,1) model suggests that oil prices will continue to show volatility. The predicted price range is between 60-90 USD per barrel, reflecting the historical fluctuations that have been recorded in previous years. This price interval is consistent with the trend of the last decade, when the oil price has stabilized between 50-100 USD per barrel, with occasional fluctuations caused by geopolitical factors, changes in demand and OPEC+ decisions. This prediction is consistent with the research of Naderi et al. (2019), who used a similar ARIMA model to predict oil prices on a monthly basis. Their study showed that forecasts based on the ARIMA model are able to accurately capture long-term trends, although there is some uncertainty in the case of unexpected changes in the market, such as geopolitical conflicts, pandemics or new extraction technologies. As in their research, our 5-year forecast shows that price volatility will continue to be present, but the ARIMA model provides a solid framework for estimating these price changes if the market continues to trend similarly to the past. It is important to note that while the ARIMA model provides useful predictions, it is still sensitive to changes in external factors. Forecasts for longer time horizons are therefore subject to a higher degree of uncertainty.

VO3: What factors most significantly influence the development of oil prices?

Several key factors influencing the price of oil can be identified, including geopolitical factors (such as the Russia-Ukraine conflict), economic indicators (inflation, interest rates) and OPEC+ decisions. For example, in 2008, the oil price fell sharply after the outbreak of the global financial crisis. The same thing happened again in 2020, when the oil price even went negative. This decline was, on the contrary, a consequence of reduced demand caused by the global pandemic and worldwide lockdown. In the literature, authors such as Kim & Jang (2023) focus on the complex effects of various factors using a hybrid model, which also includes OPEC political factors. The analysis focused on historical data and macroeconomic indicators, but in the future, the accuracy of the prediction could be improved by including specific political and economic factors, as shown by Kim & Jang and Li et al. (2024). The prediction could therefore be enriched with factors that influence the price of oil but which were not included in this research.

This research used the ARIMA model, which is a widely used method for time series analysis. The results in terms of modeling historical data and predicting oil prices show similar developments to the studies of Naderi et al. (2019) and others who used ARIMA or other machine learning methods. A significant finding was that even using the ARIMA method, it was able to capture historical trends and cyclical fluctuations, which is consistent with similar studies. When comparing this prediction with machine learning methods such as LSTM (Abdulrahim et al., 2025), it can be said that even ARIMA provides solid predictions, although modern machine learning models may show better performance for some types of data.

Conclusion

The aim of this work was to analyze the development of oil prices over the last 25 years and to predict its price until 2030. For this purpose, the ARIMA (AutoRegressive Inverse Matrix) method was used. Integrated Moving Average), which is known for its ability to model time series and predict future values based on historical data.

The oil price forecast to 2030 suggests that the price will continue to show significant volatility, ranging between 60 and 90 USD per barrel. This price range reflects not only historical trends, but also current uncertainties and fluctuations in the global oil market. The ARIMA model was able to capture the main long-term trend as well as cyclical fluctuations that are typical for the market for this commodity.

Although the model shows good predictive ability based on historical data, its results are subject to a significant degree of uncertainty. This uncertainty mainly arises from external factors that the model cannot directly include or accurately predict. These factors include, in particular, geopolitical events, such as the ongoing war in Ukraine, which has significantly disrupted global oil supply chains and caused instability in the markets. This conflict has an impact on oil production, exports and prices, and although it is not possible to simply incorporate it into the ARIMA model, it should be taken into account as a key element influencing future price developments. Other key factors are OPEC+ decisions on production volumes, which have a direct impact on the supply of oil on the market and can cause sharp price fluctuations. In addition, economic factors such as changes in global demand, inflation, interest rates or technological innovations in mining and alternative energy sources also play a role. These factors often act in a complex way and can significantly shift the price trajectory.

The ARIMA model is therefore an effective tool for monitoring and predicting oil prices based on available historical data and allows capturing basic patterns in the data. However, due to the inability to include some dynamic and unpredictable external influences, it is necessary to interpret the results with caution and in the context of current events on the world market.

Overall, it can be stated that the objective of the work was met, as it was possible to build a reliable statistical model that provides valuable insight into the likely development of oil prices in the coming years. This model can serve as a useful tool for decision-making in the energy sector and for planning, however, to increase the accuracy of the prediction, it would be appropriate to incorporate a wider range of factors in the future, including geopolitical indices, macroeconomic indicators and technological changes, or to use more modern machine learning methods.

However, it should be mentioned that the main limitation of this study may be that the ARIMA model, although effective for time series prediction, does not take into account all external factors that can fundamentally affect oil prices, such as political events or changes in extraction technology. Another limitation is that predictions for a longer time horizon are always burdened with a higher degree of uncertainty, which is clearly visible in the prediction graph, where the gray area represents the confidence interval.

In the future, it would be useful to explore other predictive methods, such as machine learning methods, and compare them with classical statistical models such as ARIMA. Furthermore, it would be possible to incorporate a wider range of external factors into the model that can better explain fluctuations in the oil market.

Acknowledgement

This research was funded by Specific University Research under Institute of Technology and Business in České Budějovice, grant number 05SVV2306.

References

AYISI, R. K. 2021. The asymmetry effect of oil price changes on inflation, and the welfare implication for Ghana. *African Journal of Economic and Management Studies*, 12(1), 55–70. https://doi.org/10.1108/AJEMS-01-2020-0009

GAO, S., LEI, Y. 2017. A new approach for crude oil price prediction based on stream learning. *Geoscience Frontiers*, 8(1), 183–187. https://doi.org/10.1016/j.gsf.2016.08.002

HE, K., E., J. 2024. Potential formation mechanism and prediction of crude oil price based on underdetermined independent component analysis. *Petroleum Science and Technology*, 1–22. https://doi.org/10.1080/10916466.2024.2408437

HE, Y. 2024. Unraveling the economic echoes: The Russo-Ukrainian conflict's influence on South Korean macroeconomic stability – Insights from the energy sector. *Sage Open*, 14(4), 21582440241287296. https://doi.org/10.1177/21582440241287296

JHA, N., KUMAR TANNERU, H., PALLA, S., HUSSAIN MAFAT, I. 2024. Multivariate analysis and forecasting of the crude oil prices: Part I – Classical machine learning approaches. *Energy*, 296, 131185. https://doi.org/10.1016/j.energy.2024.131185

KIM, G. I., JANG, B. 2023. Petroleum price prediction with CNN-LSTM and CNN-GRU using skip-connection. *Mathematics*, 11(3), 547. https://doi.org/10.3390/math11030547

KOT, S., PIONTEK, B., KHALID, B. 2024. The economic shockwaves of the Russia-Ukraine conflict and energy sources prices—The impact on international energy markets. *Acta Montanistica Slovaca*, 29(2), 353–366. https://doi.org/10.46544/AMS.v29i2.10

LI, J., HONG, Z., YU, L., ZHANG, C., REN, J. 2024. Do OPEC plus policies help predict the oil price: A novel news-based predictor. *Helion*, 10(14), e34437. https://doi.org/10.1016/j.heliyon.2024.e34437

LIN, H., SUN, Q. 2020. Crude oil prices forecasting: An approach of using CEEMDAN-based multi-layer gated recurrent unit networks. *Energies*, 13(7), 1543. https://doi.org/10.3390/en13071543

LIU, J., ZHAO, X., LUO, R., TAO, Z. 2024. A novel link prediction model for interval-valued crude oil prices based on complex network and multi-source information. *Applied Energy*, 376, 124261. https://doi.org/10.1016/j.apenergy.2024.124261

NADERI, M., KHAMEHCHI, E., KARIMI, B. 2019. Novel statistical forecasting models for crude oil price, gas price, and interest rate based on meta-heuristic bat algorithm. *Journal of Petroleum Science and Engineering*, 172, 13–22. https://doi.org/10.1016/j.petrol.2018.09.031

SHIN, H., HOU, T., PARK, K., PARK, C.-K., CHOI, S. 2013. Prediction of movement direction in crude oil prices based on semi-supervised learning. *Decision Support Systems*, 55(1), 348–358. https://doi.org/10.1016/j.dss.2012.11.009

SOHRABI, P., DEHGHANI, H., RAFIE, R. 2022. Forecasting of WTI crude oil using combined ANN-Whale optimization algorithm. *Energy Sources, Part B: Economics, Planning, and Policy*, 17(1), 2083728. https://doi.org/10.1080/15567249.2022.2083728

US ENERGY INFORMATION ADMINISTRATION (EIA). Monthly crude oil prices. [online]. [cit. 2025-05-20]. Available from: https://www.eia.gov/

YANG, H., ZHANG, Y., JIANG, F. 2019. Crude oil prices forecast based on EMD and BP neural network. *2019 Chinese Control Conference (CCC)*, 8944–8949. https://doi.org/10.23919/ChiCC.2019.8866586

Contact address of the authors:

Jakub Horák, Institute of Technology and Business in Presov, Department of Economics, Management and Marketing, Slovakia, email: horak@vste.sk, ORCID: 0000-0001-6364-9745

Aneta Soběslavská, student, Institute of Technology and Business in Ceske Budejovice, School of Expertness and Valuation, Czech Republic

How to cite this article:

HORÁK, J., SOBĚSLAVSKÁ, A., 2025. Oil price trends and their prediction using the ARIMA model. *Littera Scripta*, 18(1), pp. 38-50. ISSN 1805-9112.