ESTIMATION AND PREDICTION OF DAILY NATURAL GAS PRICES BY APPLYING THE AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA) MODELS

George-Eduard GRIGORE
Bucharest University of Economic Studies
Piața Romană, 6, 010374, Bucharest, Romania
grigoregeorge15@yahoo.com

Simona NICOLAE
Politehnica University of Bucharest
Splaiul Independenței 313, 060042, Bucharest, Romania
nsim2005@yahoo.com

Ioana-Ruxandra IONIȚĂ (LIE)
Politehnica University of Bucharest
Splaiul Independenței 313, 060042, Bucharest, Romania
ruxandra.lie@gmail.com

Adriana-Mihaela IONESCU
Politehnica University of Bucharest
Splaiul Independenței 313, 060042, Bucharest, Romania
amihaela.ionescu@gmail.com

Abstract. *Natural gas continues to represent the main source of traditional energy that is increasingly essential and vital to all mankind. From this perspective, increasing dependence on this energy source and investigating the price movements and dynamics mechanism by developing statistical and econometric models become relevant topics of discussion in current empirical studies. Thus, we proposed to carry out/perform an estimation and forecasting analysis of daily natural gas prices over 1997-2022 using Henry Hub as benchmarks for the entire North American natural gas market and parts of the global liquid natural gas (LNG) market. Applying ARIMA Integrated Autoregressive Models and following the specifications of Box-Jenkins methodology, we could provide an up-to-date picture of the behavior of natural gas prices time series, highlighting its most relevant co-movements and fluctuations. Our results showed that AR(1) MA(2) MA(4) Adjusted ARIMA Model is the most robust and suitable/appropriate model, after which the forecast analysis of daily natural gas prices could be performed. Moreover, it was shown that the natural gas prices persist in fluctuating and oscillating in the analyzed period, generating a high level of volatility (the predicted value of this volatility is approximately 13%) especially during the last two years: 2020-2022. Certainly, we can confirm that the investigation of the time series of natural gas prices will represent an actual and important point of interest in international financial markets, investments, and risk management.*

Keywords: *autoregressive models; ARIMA models; Box-Jenkinks methodology; Henry Hub Natural Gas prices; predicted volatility; risk management.*

Introduction

Actually, natural gas represents an important and vital energy resource for mankind. The indispensable character of this globally strategic resource generates a series of effects and implications both on a social and political level, and above all on an economic and financial level.

For example, natural gas is an important commodity in most international transactions between consuming and net importing countries (especially in the case of developing and developed countries) and the producing and exporting ones (i.e. the Caspian countries, Arab countries, etc.).

Nowadays, new opportunities to diversify these energy resources are being sought, as well as the construction of pipelines and new access routes to them with an eye/in order to cover the energy needs or requirements in the medium and long term in the case of importing countries/importers. From this perspective, many researchers and academics are increasingly concerned with finding those methods and techniques for estimating and forecasting the consumption and production of natural gas, as well as predicting and estimating the time series of natural gas prices. For these reasons, the main objective of our study is to perform a forecasting analysis of the time series of natural gas prices to measure the degree of volatility associated with this asset and understand fluctuating movements and co-movements that are more and more pronounced and persistent.

Moreover, the practical value of our study is to determine and provide an appropriate framework for the dynamic trend/evolution of natural gas prices for various participants such as: financial and energy companies, financial investors, government authorities, or policymakers. Therefore, we use the time series of daily global natural gas prices in 1997-2022, which are obtained by applying the first difference. Also, we consider the Henry Hub Natural Gas Spot Prices as the benchmark, where the data source is U.S. Energy Information Administration.

From a methodological point of view, we employ the Box-Jenkins method in six steps according to the autoregressive ARIMA models: (1) descriptive statistics of the natural gas prices time series; (2) assessment of the time series stationarity using unit root tests verification; (3) identification the tentative ARIMA models based on visualization of the correlogram, ACF and PACF; (4) estimation the coefficients of each ARIMA model; (5) selection to the appropriate and approximate ARIMA model based on the application of robustness tests, and (6) forecasting analysis in our sample based on the most suitable ARIMA Model.

Our study is organized in the following order: Section 2 presents the most relevant literature review; Section 3 describes the Research Methodology, Data, Preliminary Analysis of natural gas price series, and the applied ARIMA Models; Section 4 discusses the main results and interpretation of these results and Section 5 presents the conclusion.

Literature review

In this section, we will present the main theoretical and empirical studies to investigate and forecast the dynamic behavior of the global natural gas prices time series by applying the Autoregressive Integrated Moving Average (ARIMA) Models.

From this perspective, we can say that more and more researchers are interested in finding these econometric models that provide the most faithful and appropriate prediction regarding the price dynamics of this globally traded asset. Thus, using forecasting models becomes an important and relevant condition, especially in the case of the investment process, risk management, international trade, and international financial and natural gas markets.

We believe that our study is practical and helpful to investors increasingly interested in understanding the fluctuating and dynamic behavior of the prices of financial and non-financial assets in their portfolios (in our case, the global prices of natural gas) by applying the ARIMA models.

Hosseinipoor and Hajirezaie's (2016) study applies the autoregressive ARIMA and GARCH models to determine a forecasting analysis of Henry Hub Natural Gas prices in the case of the U.S. natural gas market in the long-term period. The results showed that natural gas prices are fluctuating extremely volatile in the analyzed period 1996-2016, generating a high risk for consumers, producers, and investors in the U.S. natural gas market.

Moreover, the authors showed that the ARIMA (5,1,9) and GARCH(1,1) are appropriate models to forecast the time series of natural gas prices, both of them suggesting a slight increase in the global price of natural gas in the future, up to 3.20 USD/Million Btu (value expected in September 2016). Similarly, Mishra (2012) proposes a non-parametric approach to forecast the time series of natural gas, oil, and gold from 1975-2010. The results obtained have highlighted that the three applied univariate and non-parametric models, i.e. ARIMA (1,2,1), GARCH (1,1,) Alternating Conditional Estimation (ACE model) provide extremely rigorous and adequate foresight in the case of the analyzed time series, being among the first studies that have developed a new, sophisticated and innovative forecasting approach.

More recently, many forecasting models are used to ensure a good prediction of the natural gas price, as well as the estimation of the production and consumption of this commodity. In this sense, the study by Manigandan et al. (2021) uses complex and advanced autoregressive models (i.e. SARIMA and SARIMAX models) to forecast and measure the consumption and production of natural gas in the USA.Using monthly time series, the results showed that both consumption and natural gas production tends to increase until 2025.

Last but not least, the study has practical implications for the future decision-making mechanism of the US natural gas market. Guan et al. (2022) use advanced machine learning techniques to develop an appropriate forecasting model that captures the fluctuating movements of natural gas prices, political events, and news from 2012-2021. This study's main contribution and novelty is the new technique (BiLSTM) used to extract price information related to news features. The robustness of the model is over

79%, explaining that the future trend of the natural gas price is dependent and influenced by various news and events, and the performance is better than most traditional machine learning algorithms.

The two probabilistic models, The Day-Ahead and The Month-Ahead Models developed by Berrisch and Ziel (2021), were used to estimate the future natural gas prices from 2011 to 2020. The main evidence showed a strong nexus between the natural gas market and the electricity market at the European level, and also the conditional volatility or variance of the price of natural gas (estimated by the TGARCH model) tends to increase up to 34%. Moreover, according to these forecasting models, the authors found that the highly significant pattern in the data estimating a total price reduction of 0.35 per month these risk premiums should decrease over time because the uncertainty decreases as time gets closer to delivery.

In another approach, the study by Göncü et al (2013) investigates the interdependence relationship between residential and commercial consumption of natural gas and the temperature level/degree in the case of Istanbul (Turkey) in the period 2004-2011 by applying the panel data methodology. In this new dynamic model, the authors have performed the Monte Carlo simulation and checked the accuracy of forecasting analysis using the most relevant diagnostic tests. The main results showed that the sudden increases in the temperature have directly influenced natural gas consumption, which generates a significant increase in domestic demand and the price level of natural gas.

Moreover, the necessity of establishing the natural gas futures market in Turkey is observed in the near future where those participants can probably hedge against these climate and environmental risks. The main purpose of the study by Nyangarika and Tang (2018) is represented by forecasting analysis of the global price of oil and natural gas time series from 1991 to 2016. The authors applied the modified autoregressive integrated moving average models in this regard.

In the same study, a linear regression was applied to explain the interdependence relationships between the global oil price, the global natural gas price, and six other dummy variables (i.e. World financial crisis, The Military company in Iran, Syria, and Iraq, Afghanistan, and the U.S.Terror). The obtained results suggested that the global oil price is directly influenced by the global natural gas price level and by the armed conflict/military company that broke out in Iraq in 2004. Also, the Modified ARIMA model suggested a better forecast for global oil and natural gas prices so investors can better understand the increasing and persistent volatile movements.

On the other hand, the study by Tamba et al. (2018) highlights the most important empirical and theoretical studies to determine and investigate the forecasting analysis of natural gas prices. From this perspective, these authors emphasized the series of methods and tools used, the specifics and characteristics of the data used, and the possible future implications in natural gas forecasting.

Within this study, a series of published articles were presented in a chronological way/manner regarding the estimation and forecasting of natural gas time series in the interval 1949-2015.

At the same time, a diversity of estimation and forecasting methods was suggested, such as: ARIMA and GARCH autoregressive models, time series models, regression models, panel regression models, non-parametric models, advanced machine learning techniques, combined models (SARIMAX, SARIMA, ARX), etc. Also, each model was tested and verified based on robustness indicators (Root Mean Square Error, Mean Absolute Error, Mean Absolute Percent Error, R-Squared, Sigma) depending on the expected and analyzed time horizon. The value of this study is impressive given the rigorous, logical, and detailed presentation of the published studies in the field of forecasting natural gas time series, generating new research directions, and resolution of possible methodological inconveniences.

Another empirical research (Siddiqui, 2019) proposes an Autoregressive Neural Network (ARN) Model to ensure a better and more accurate forecasting analysis of the time series of natural gas prices. Using the Henry Hub daily natural gas Spot prices in 1997-2018, the results suggested that the ARN model generates a considerable improvement over the ARIMA models used by over 30%. This study employed ARIMA (3,1,3) and ARIMA (2,1,2). The same study emphasizes that the Autoregressive Neural Network Model represents an actual model for basing spot gas purchase decisions.

Similarly, Viacaba et al (2012) use an innovative data mining model to forecast the time series of natural gas prices. According to the Selective Support Vector Regression (SVR model), the authors could measure the specific volatility of natural gas in the U.S. natural gas market from 2003-2009. They confirmed that the volatility measured by the SVR models is approximately the same as that provided by the U.S. Energy Information Administration by their Short Term Energy Outlook (STEO) forecast, having a value of 9.90%. In the same direction, it also applies to the study by Hosseinipoor (2016).

To forecast the U.S. natural gas prices time series, the author used three modern techniques based on the Artificial Neural Networks (i.e. simple Artificial Neural Network Model, three-layer Artificial Neural Network Model with 6 hidden neurons, and NARX network model with 6 neurons). All these models captured the price spikes, offering a simulation as close as possible to the reality of the present natural gas prices time series. The study by Wong-Parodi et al. (2006) presents a comparative analysis of U.S. natural gas price forecasting from 1998-2003. Taking into account the forecasting models provided by Short Term Energy Outlook-STEO (U.S. Energy Information Administration) and Henry Hub Foreward Prices, the authors concluded that in most cases the futures market is a more accurate predictor of natural gas prices than STEO for a 24-month forecast period.

Busse et al. (2012) highlight a dynamic approach to forecasting the natural gas price fluctuations in the market area of NetConnect Germany. In order to capture the interdependence relationships between natural gas prices and other independent variables they used the NARX neural network model and sensitivity analysis. The results illustrated that the most important factors that directly influenced the future trend of natural gas prices are: temperature, the exchange rate between USD and EUR, the exchange rate between GBP and EUR, and the settlements of the gas hubs. At the same time, the authors conclude that autoregressive econometric models, in particular, ARIMA and GARCH are user-friendly in forecasting analysis compared to the complex methodology proposed by Artificial Neural Network Models, providing a high level of accuracy and robustness. Nevertheless, data mining techniques represent an important and up-to-date step in the case of forecasting non-parametric models.

In another approach, Zhao et al. (2018) use the autoregressive ARIMA to implement an adequate forecasting model for fuel cost in the case of Texas in 2013-2016. The ARIMA (2,1,1) and ARIMA (2,0,1) models are those models on which the forecast analysis for Henry Hub Natural Gas spot prices was employed. The results show the proposed forecasting algorithm performs better than the method that uses the three-month delayed data.

Considering that the state of Iran represents an important producer and exporter of energy products, especially natural gas, and oil, Farrokhi and Hassanzadeh (2017) use ARIMA models to forecast and estimate the monthly and annual domestic consumption of natural gas. The results of the study following the application of the Box-Jenkins J methodology, showed that the best model for forecasting the annual consumption of natural gas in Iran is ARIMA (0,1,0), while SARIMA (1,0,0) (1,1,0) is suitable to forecast the monthly domestic consumption.

Methodology

The main objective of our study is to estimate and forecast the daily natural gas prices time series in the period 1997-2022. From this perspective, the research questions are as follows: (1) How can we estimate the daily natural gas prices in the most appropriate way? and (2) How we can forecast the natural gas price time series by applying the autoregressive models?

According to the U.S. Energy Information Administration (EIA, 2021), natural gas prices are a market supply and demand function. Increases in natural gas supply generally result in lower prices, and decreases in supply tend to lead to higher prices. Increases in demand generally lead to higher prices, and decreases in demand tend to lead to lower prices.

Another aspect that is offered by the EIA is that the strength of the economy influences natural gas markets. For example, during economic growth, increased demand for goods and services from the commercial and industrial sectors may increase natural gas consumption. Also, because of natural gas supply infrastructure constraints and limitations in the ability of many natural gas consumers to switch fuels quickly, short-term increases in demand and/or reductions in supply may cause large changes in natural gas prices, especially during the wintertime.

Moreover, when competing fuels' prices rise relative to natural gas prices, switching from those fuels to natural gas may increase demand and prices. Nevertheless, natural gas prices on the spot market may increase sharply during high-demand periods if natural gas supply sources are relatively low or constrained.

On the other hand, European Union (ECB, 2022) considered that natural gas is the second most important primary energy resource in the euro area, after petroleum-based products, and also acts as the key marginal energy resource in electricity generation, given the flexibility of gas-fired power plants and the overall gas infrastructure in responding to fluctuations in electricity demand.

We will then use/apply autoregressive integrated moving average models (ARIMA models) for estimating and forecasting the dynamic and volatile behavior of natural gas prices from 1997-2022.

An ARIMA model can be considered a special type of regression model in which the dependent variable has been rationalized and the independent variables are all lags of the dependent variable and/or lags of the errors, so it is straightforward in principle to extend an ARIMA model to incorporate information provided by leading indicators and other exogenous variables: you simply add one or more regressors to the forecasting equation.
The ARIMA model is thought to provide more accurate predictions by removing these difficulties. ARIMA models have demonstrated their efficient ability to produce short-term predictions. In terms of short-term prediction, it consistently outperformed complicated structural models (Mashadihasanli, 2022).

In general, the ARIMA model is based on AR and MA models. While the AR model is used to show that the current observation is dependent on previous observations, the MA model is used to show that the current and previous residuals compose a linear function. ARIMA, as a time series forecasting method, can be employed to understand the data through time series analysis and study the sequence formed by the state of the variables at different times (Mashadihasanli, 2022).

It can also be used to fit the data and make forecasts, quantitatively describing the pattern of variables in the time series and future trends. The advantage of the ARIMA model over other forecasting models is that it only requires forecasting based on endogenous variables, without the need to acquire other relevant exogenous variables, and focuses more on the patterns and trends of the variables to be studied. As an effective time series analysis tool, an ARIMA model is implemented to forecast price signals. ARIMA models have been analyzed and evaluated to forecast natural gas prices (Zhao et al. 2018).

In our case, we used the ARIMA (p,d,q) models, where: p = number of lags of the dependent variable; q =number of lags of the error term; d = how many times the variables is differenced to become stationary. Because engaging an ARIMA model is to forecast a series, the Box- Jenkins methodology (Figure 1) comes in handy in answering the predicting question.

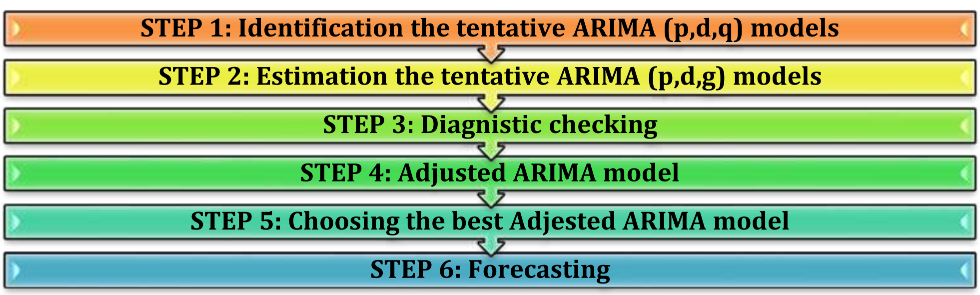

Figure 1. The Box-Jenkins methodology
Source: Authors' own research contribution

In step 1, identification, the question is "How can the appropriate model be identified?". We used the correlogram and the partial autocorrelation function (PACF) to answer. These tools measure the correlation between observations that are k times after controlling for correlations at intermediate lags. In other words, PACF is the correlation between $d(natural\ gas\ prices)_t$ and $d(natural\ gas\ prices)_{t-1}$ after removing the effect of the intermediate $d(natural\ gas\ prices)'s$. In this step, the identification procedure consists in: plot the series to visualize if stationary, or not; from the correlogram, calculate the ACF and PACF; check whether the series is stationary or not; take the first difference of the raw data and calculate ACF and PACF again, and visualize the graphs of the ACF and PACF and determine which models would be good. In step 2, we estimated the tentative models, using the Least Square and ARIMA Procedures. Moreover, in this step, we selected the appropriate model, which the significant coefficients, the less volatility (Sigma^2), the lowest Akaike Info Criterion (AIC) and Schwarz criterion (SC), and the highest R-Squared (R^2).
In step 3, we investigated the appropriate model, performing the diagnostic tests and/or re-estimating the adjusted ARIMA Models. Next, we performed the forecast analysis by plotting the forecast graphs and verifying the success of the forecast to predict the future values of the daily natural gas prices time series. Indeed, the fundamental idea of Box-Jenkin's methodology is that of parsimony.

Descriptive statistics of the natural gas prices time series used

The analyzed time series is that of natural gas prices globally, in 1997- 2022. The data source is U.S. Energy Information Administration (EIA), which provides the level of benchmark prices, namely Henry Hub Natural Gas Spot Prices. In our case, we used the daily natural gas prices time series in the period 1997-202 and the dynamic trend of these prices is shown in the Figure 2.

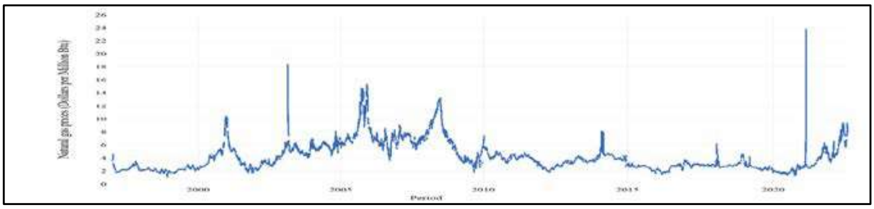

Figure 2. The plot of daily natural gas prices in the period of 1997-2022
Source: Authors' own research contribution

According to this figure, the evolution and fluctuating movements of natural gas prices over time can be observed. Also, we certainly say that natural gas is an important and strategic asset, highly traded in the international financial markets. Thus, we identify the main spikes and jumps since 2005, where the average level of natural gas price is about \$18 / Million BTU. Also, the negative effects of the Global Financial Crisis have been observed since 2008, when natural gas prices fell by 56%. Mostly, price fluctuations have continued until now, with 2020 recording an extremely high level of natural gas price of approximately \$24/Million BTU, mainly caused by the outbreak of the Coronavirus pandemic. We are also aware that fluctuating movements persist in the unpredictable and uncertain near future following the outbreak of the war between Russia and Ukraine (February 24, 2022), where the average natural gas price is approximately \$18/Million BTU.

Table 1. Descriptive statistics of the natural gas prices time series

Mean	4.1976
Median	3.5600
Maximum	23.8600
Minimum	1.0500
Standard Deviation	2.1827
Skewness	1.6516
Kurtosis	7.0142
Jarque-Bera/Prob.	7241.843 /0.0000
Observations	6431

Source: Authors' own research contribution

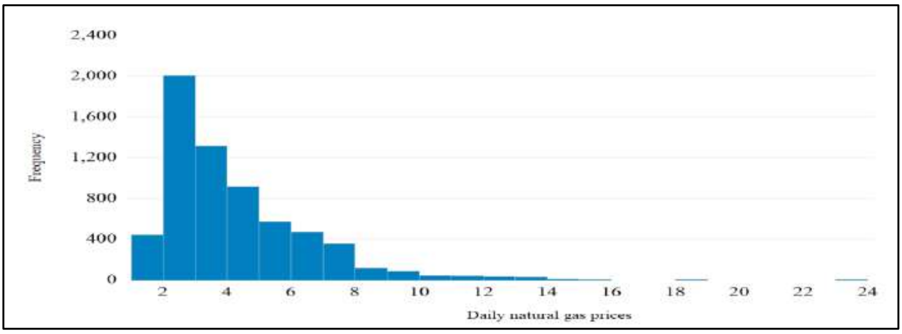

Figure 3. Histogram of daily natural gas prices in the 1997-2022 Source: Authors' research contribution

Table 1 presents a brief statistical description of the daily natural gas prices time series in our analyzed period. The time series contains 6431 observations, where the maximum value of the price is approximately \$24/Million BTU, while the minimum value of the natural gas price is about \$1 /Million BTU. In this period, the average value of the natural gas price is \$4.20/Million BTU, approximately \$3.60 representing the median value of the analyzed time series.

From a statistical point of view, we observe that the natural gas price deviates in average with \$2.18/Million BTU, which confirms a high risk associated with this globally traded asset. Regarding the distribution of natural gas price time series (Figure 3), the positive values of the Skewness (1.65) and Kurtosis (7.01) states highly present stylized facto in the analysis, namely the fat-tail property of distribution or the presence of a leptokurtic and asymmetric to the right distribution. At the same time, based on the histogram, the natural gas prices do not match and do not tend towards a normal or Gaussian distribution N (0,1) and this fact is visible from the extremely high value of the Jarque-Bera test.

On the other hand, our series is non-stationary according to the data provided by Figure 2 and Figure 3. This must be corrected before applying the ARIMA models. In this sense, the Augmented Dickey-Fuller (Dickey and Fuller, 1979) and Phillips Perron (Phillips and

Perron, 1988) tests were done to check for stationary of the natural gas prices time series. The results obtained can be seen in Table 2.

Table 2. The results of unit root tests

ADF Test	t-Statistic	Prob.	Test critical value at 1% level
None	-41.7636	0.0000	-2.5653
Intercept	-41.7612	0.0000	-3.4311
Trend and Intercept	-41.7602	0.0000	-3.9594
PP Test			
None	-105.2211	0.0001	-2.5653
Intercept	-105.2240	0.0001	-3.4311
Trend and Intercept	-105.2300	0.0001	-3.9594

Source: Authors' own research contribution

In our case, ADF and PP have the statistical test values approximately equal to -100.00 and the associated p-value at 0.0001. As the test values are lower than the critical values by choosing the 1% confidence level, it can be certainly confirmed that the null hypothesis is rejected. The results showed that the natural gas price series is stationary at the first difference, with an extremely high probability level (p-value is less than 1%). Therefore, the specific analysis of the autoregressive methodology continued on the new series, obtained by applying the first difference. In our case, the series is called $d(natural\ gas\ prices)$, which indicates that our time series is I(1). For this reason, we will apply the ARIMA models.

ARIMA Models

Using autoregressive (AR) and moving average (MA) processes to analyze and forecast the time series can be improved. Thus, by combining the two processes, a generalized model is obtained, called the Autoregressive Integrated Moving Average (ARIMA) Model. The ARIMA model combines the dependent variable's autoregressive lags and moving average process errors. ARIMA Model is popularly known as the Box-Jenkins (1976) methodology. It is a method among several used in forecasting variables and uses the information obtained from the variables themselves to forecast its trend.

The variable is regressed on its own past values, and from this perspective, the ARIMA model is based on univariate analysis. Also, it is designed to forecast future movements, knowing and analyzing the probabilities, or stochastic, properties of variables. ARIMA Model uses the following philosophy: "let the variable speak for itself".

ARIMA Model helps investors, government regulators, policymakers, and relevant stakeholders make informed decisions. For example, an investor before buying a financial asset will want to know if it is really worth buying and holding on to it. In the same way, policymakers and regulators will want to forecast the future trend of some economic series and formulate policies based on the previous realizations of such variables. The underlying assumptions of the ARIMA Model are: (i) stationarity (use unit root test), and (ii) invertibility (implicitly assumes that the series can be approximated by an autoregressive model).

The ARIMA Model specification are: (i) the BJ-type time series models allow the variable to be explained by past, or lagged values of the variable itself and stochastic error terms; (ii) ARIMA Models are sometimes called theoretical models because they are not derived from any economic or specialized theory; (iii) the series is simply explaining itself using its historical data; (iv) ARIMA is composed of two distinct models which explain the behavior of a series from two different perspectives: the autoregressive models (AR), and respectively, the moving average (MA) models.

The AR(p) model can be generalized to include more series lags, such that the values in brackets() indicate the number of lagged values of the regressand included in the model.

The formula of the Generalised AR(p) model is presented in the equation 1.
$d(natural\ gas\ price)_t = a + \sum_{i=1}^{p} \quad b_i d(natural\ gas\ price)_{t-i} + u_t$ *(1),* where: a= intercept; b_i= coefficients; $d(natural\ gas\ price)_{t-i}$= the past value of the natural gas price; and u_t= the error term; t=time; p= lags

The moving average (MA) model gas the following formula that is presented in Equation 2. Moreover, the generalized formula of the ARMA Model is illustrated in Equation 3.
$d(natural\ gas\ price)_t = \gamma + d_0 u_t + \sum_{j=1}^{q} \quad d_j u_{t-j}$ *(2),* where: $d(natural\ gas\ price)_t$ is explained by the value of the error term and the intermediate past error known at time t.
ARMA (p,q) => $d(natural\ gas\ price)_t = a + \sum_{i=1}^{p} \quad b_i d(natural\ gas\ price)_{t-i} + d_0 u_t + \sum_{j=1}^{q} \quad d_j u_{t-j}$ *(3)*
Distinction between ARMA and ARIMA models is the integration component which brings us back to the subject of stationarity.

The key to ARIMA modeling is employing the iterative identification, estimation, and diagnostic checking process. Thus, it is advisable not to choose the model a priori. Furthermore, ARIMA informs the researcher or reader that the series in question has been integrated before being used for any analysis. The main advantage of ARIMA forecasting is that it requires data on the time series in question. First, this feature is advantageous if one is forecasting many time series. Second, this avoids a problem sometimes with multivariate models (Castaneda et al., 2021; Mouchtaris et al., 2021).
Owing to purely statistical approaches, ARIMA models only need the historical data of a time series to generalize the forecast and increase prediction accuracy while keeping the model parsimonious. Potential cons of using ARIMA models are: difficult to predict turning points, there is quite a bit of subjectivity involved in determining (p,d,q) order of the model; computationally expensive; poorer performance for long-term forecasts, or cannot be used for seasonal time series (Castaneda et al., 2021; Mouchtaris et al., 2021; Bakar, Rosbi & Uzaki, 2018).

The entire methodological approach and the results obtained from estimating the ARIMA models were developed using the econometric software EViews12.

Results and discussions

In this section, we present and discuss the main results obtained from the analysis of the natural gas prices by applying the autoregressive integrated moving average (ARIMA) models. Specifically, the results will be presented according to the specific ARIMA autoregressive methodology, namely the Box-Jenkins method.

From this perspective, in the first part , we present how we identified and selected the tentative ARIMA models. We focus on the main results obtained from the estimation analysis of these tentative ARIMA models.

In the last part, we investigate and diagnose the applied autoregressive ARIMA models and choose the most suitable/reliable model to estimate the natural gas prices time series. Moreover, we present and discuss the main implications obtained from the forecast natural gas prices analysis in the period of 1997-2022.

The main results from the identification of the tentative ARIMA models

In order to use ARIMA models, the first step is to identify the tentative and experimental ARIMA models that explain and capture the dynamic evolution of the natural gas prices time series. Therefore, we applied the econometric method known as the correlogram.

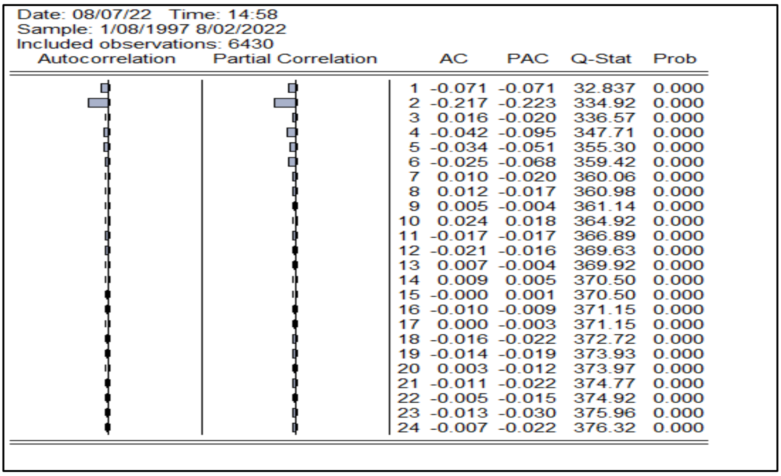

Figure 4. ACF and PACF to determine the tentative ARIMA Models

Source: Authors' own research contribution

The obtained results that are illustrated in Figure 4 show that both autocorrelation function (ACF) and partial autocorrelation function (PACF) present the same pattern and namely: the both of them decrease progressively. This means the estimation analysis will be done by applying the autoregressive integrated moving average (ARIMA) models. By visualising the correlogram, we observe that the most significant lags in which information can be obtained are Lag 1, Lag 2, Lag 4, Lag 5, and Lag 6. From this perspective, we construct and identify the ARIMA models formed from the random combination of these lags.

Table 3. The tentative ARIMA models

ARIMA (1,1,1)
ARIMA (2,1,2)
ARIMA (4,1,4)
ARIMA (5,1,5)
ARIMA (6,1,6)

ARIMA (1,1,2)
ARIMA (2,1,1)
ARIMA (2,1,4)
ARIMA (4,1,2)
ARIMA (1,1,4)
ARIMA (2,1,5)
ARIMA (2,1,6)
ARIMA (6,1,2)
ARIMA (5,1,2)
ARIMA (1,1,6)
ARIMA (1,1,5)
ARIMA (6,1,5)
ARIMA (6,1,4)

Source: Authors' own research contribution

In our case, 18 ARIMA models were created for estimation and forecasting analysis of the daily natural gas prices time series in the period 1997-2022 These models are shown in Table 3.

The main results from the estimation analysis of the tentative ARIMA models

The second section present the specific results for the 18 tentative autoregressive integrated moving average (ARIMA) models. Starting from the specific hypotheses in using the ARIMA models, many conditions are proposed in the literature that need to be fulfilled by each model applied.

In this regard, to select the most suitable ARIMA model, these models should have the least number of parameters, significant autoregressive (AR) and moving average (MA) parameters/coefficients, the lowest predicted value of volatility (Sigma^2), the highest R-Squared (R^2), lowest Schwarz Information Criteria (SC), lowest Akaike Information Criteria (AIC), and also no heteroskedasticity and no autocorrelation in the residual or errors terms series. In this sense, we resorted to creating a ranking/top regarding choosing the most suitable ARIMA model. These results can be viewed in Table 4 down below.

Table 4. The ranking of ARIMA tentative Models

MODEL	Significant coefficients	Sigma^2	R^2	AIC	SC	DW	Rank
ARIMA (1,1,2)	2	0.131625	0.064581	0.811344	0.815555	1.999902	**1**
ARIMA (2,1,2)	2	0.132556	0.057962	0.818394	0.822605	2.202283	2
ARIMA (4,1,2)	2	0.132645	0.057328	0.819066	0.823277	2.200469	3
ARIMA (2,1,4)	2	0.132656	0.057251	0.819148	0.823359	2.200313	4
ARIMA (5,1,2)	2	0.132755	0.056551	0.819891	0.824101	2.204511	5

MODEL	Significant coefficients	$Sigma^2$	R^2	AIC	SC	DW	Rank
ARIMA (6,1,2)	2	0.132791	0.056292	0.820165	0.824376	2.199942	6
ARIMA (2,1,1)	2	0.132932	0.055292	0.821218	0.822676	1.988970	7
ARIMA (2,1,6)	2	0.133939	0.048135	0.828765	0.832976	2.184811	8
ARIMA (2,1,5)	2	0.133975	0.047877	0.829035	0.833246	2.190658	9
ARIMA (1,1,1)	2	0.134093	0.047037	0.829921	0.831378	1.939554	10
ARIMA (1,1,4)	2	0.139736	0.006934	0.871124	0.875335	2.033281	11
ARIMA (1,1,5)	2	0.139784	0.006598	0.871463	0.875674	2.033088	12
ARIMA (1,1,6)	2	0.139884	0.005883	0.872182	0.876393	2.032088	13
ARIMA (6,1,4)	2	0.140306	0.002883	0.875197	0.879407	2.147425	14
ARIMA (6,1,5)	2	0.140440	0.001935	0.876146	0.880357	2.148023	15
ARIMA (4,1,4)	0	0.140438	0.001944	0.876137	0.880347	2.145394	16
ARIMA (5,1,5)	0	0.140486	0.001607	0.876474	0.880685	2.147409	17
ARIMA (6,1,6)	0	0.140547	0.001172	0.876910	0.881121	2.145343	18

Note: Significant coefficients = p-value is less than 5%; $Sigma^2$= volatility; R^2=R-Squared; AIC= Akaike info criterion; SC= Schwarz criterion; DW= Durbin-Waston stat

Source: Authors' own research contribution

From the start, we notice that the best ARIMA model in estimating analysis of the daily natural gas prices time series is ARIMA (1,1,2). Compared to the rest of the identified models, the selected ARIMA (1,1,2) model has the most high level of the R-Squared (R^2), which is approximately 6.5%). Also, this model has the lowest value of the predicted volatility ($Sigma^2$), respectively 13.65%.

Moreover, the lowest values of robustness indicators (Schwarz Information Criteria and Akaike Information Criteria) confirm that the ARIMA (1,1,2) model is the most suitable and appropriate for measuring natural gas prices. Also, the value of the Durbin-Waston test is extremely close to the value 2.00, which concludes that the errors are not autocorrelated, successfully validating this assumption.

It was interesting that in the majority of tentative ARIMA models, at least two coefficients were statistically significant at the 5% level of p-value. This aspect was not encountered in the case of the following ARIMA models: ARIMA (4,1,4); ARIMA (5,1,5), and ARIMA (6,1,6) models. Also, these three models have the highest values of predicted volatility, as well as the lowest values of the R-Squared.

In a comparative approach, we can state that the first 10 tentative ARIMA models are extremely close according to the conditions that should be met, respectively the values of the Schwarz Information Criteria (the values are between 0.82 to 0.83), Akaike

Information Criteria (these values are between 0.81 to 0.82), R-Squared (these values vary between 4,7% to 5,8%, and predicted variance (the values between 13,10% to 13,45%). On the other hand, the econometric analysis continued by performing diagnostic tests for the ARIMA (1, 1, 2) model, in our case being the best autoregressive model.

The main results from the diagnostic and forecasting analysis of the daily natural gas prices

As we have seen previously, the ARIMA (1,1,2) model is the most suitable to estimate the natural gas prices time series in the mentioned period. But to confirm this aspect and see if our model captures all the necessary information regarding the behavior of the natural gas prices time series, we performed several diagnostic tests based on the visualization of the correlogram, the autocorrelation function (ACF), and the partial autocorrelation function (PACF).

Date: 08/08/22 Time: 14:09
Sample: 1/08/1997 8/02/2022
Q-statistic probabilities adjusted for 2 ARMA terms

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.000	-0.000	5.E-06	
		2	0.015	0.015	1.4214	
		3	-0.021	-0.021	4.3084	0.038
		4	-0.050	-0.050	20.090	0.000
		5	-0.047	-0.047	34.295	0.000
		6	-0.038	-0.037	43.565	0.000
		7	-0.002	-0.003	43.589	0.000
		8	0.009	0.006	44.168	0.000
		9	0.004	-0.002	44.275	0.000
		10	0.022	0.016	47.291	0.000
		11	-0.016	-0.019	48.919	0.000
		12	-0.016	-0.017	50.518	0.000
		13	0.001	0.003	50.528	0.000
		14	0.002	0.005	50.565	0.000
		15	-0.002	-0.003	50.591	0.000
		16	-0.015	-0.017	52.097	0.000
		17	-0.008	-0.011	52.539	0.000
		18	-0.023	-0.024	55.956	0.000
		19	-0.023	-0.023	59.282	0.000
		20	-0.009	-0.011	59.798	0.000
		21	-0.023	-0.025	63.088	0.000
		22	-0.013	-0.018	64.190	0.000
		23	-0.021	-0.028	67.169	0.000
		24	-0.014	-0.021	68.481	0.000

Figure 5. The correlogram of ARIMA (1,1,2) Model
Source: Authors' own research contribution

From Figure 5, it appears that this ARIMA (1,1,2) model does not capture all the desired information, the correlogram is not flat and, consequently, we have to find other new ARIMA models.

Therefore, we will re-estimate other ARIMA models to contain the significant lags: Lag 4, Lag 5, and respectively, Lag 6. These are the Adjusted ARIMA models and they are illustrated in Table 5 down below.

Tale 5. Ranking and selection the most apporpiate Adjusted ARIMA Model

ADJ. ARIMA MODELS	Significant coefficients	Sigma^2	R^2	AIC	SC	DW	Rank
AR(1) MA(2) MA(4)	3	0.131193	0.067650	0.808371	0.813635	2.000983	**1**
AR(1) AR(4) MA(2)	3	0.131343	0.066582	0.809516	0.814779	2.004283	2
AR(1) MA(2) MA(5)	3	0.131349	0.066540	0.809561	0.814824	2.001003	3
AR(1) AR(5) MA(2)	3	0.131379	0.066328	0.809787	0.815051	2.003944	4
AR(1) AR(6) MA(2)	3	0.131439	0.065903	0.810242	0.815506	2.000703	5
AR(1) MA(2) MA(6)	3	0.131475	0.065647	0.810516	0.815780	2.000115	6
ARIMA (1,1,2)	2	0.131625	0.064581	0.811344	0.815555	1.999902	*7*

Note: Significant coefficients = p-value is less than 5%; Sigma^2= volatility; R^2=R-Squared; AIC= Akaike info criterion; SC= Schwarz criterion; DW= Durbin-Waston stat

Source: Authors' own research contribution

From the 6 adjusted ARIMA models, we will select the one that meets the conditions of robustness and accuracy, respectively: least number of parameters, significant autoregressive (AR) and moving average (MA) parameters/coefficients, the lowest predicted value of volatility (Sigma^2), the highest R-Squared (R^2), lowest Schwarz Information Criteria (SC), lowest Akaike Information Criteria (AIC), and also no heteroskedasticity and no autocorrelation in the residual or errors terms series.

Certainly, the AR (1) MA (2) MA (4) model represents the model that provides the better estimation, which captures any significant information about the natural gas prices time series.

Also, we can conclude that this adjusted ARIMA model and the correlogram of the residual is flat which indicates that all the information has been captured. The correlogram of AR (1) MA(2) MA(4) adjusted model is presented in Figure 6.

Date: 08/08/22 Time: 14:57
Sample: 1/08/1997 8/02/2022
Q-statistic probabilities adjusted for 3 ARMA terms

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.001	-0.001	0.0021	
		2	0.001	0.001	0.0044	
		3	-0.021	-0.021	2.9397	
		4	0.004	0.004	3.0275	0.082
		5	-0.047	-0.047	17.400	0.000
		6	-0.023	-0.024	20.937	0.000
		7	-0.003	-0.003	21.015	0.000
		8	0.013	0.011	22.027	0.001
		9	0.002	0.001	22.043	0.001
		10	0.021	0.019	24.965	0.001
		11	-0.017	-0.019	26.890	0.001
		12	-0.017	-0.018	28.710	0.001
		13	0.000	0.002	28.711	0.001
		14	0.003	0.002	28.754	0.002
		15	-0.005	-0.003	28.887	0.004
		16	-0.016	-0.017	30.623	0.004
		17	-0.010	-0.012	31.248	0.005
		18	-0.024	-0.025	34.905	0.003
		19	-0.024	-0.024	38.637	0.001
		20	-0.010	-0.011	39.325	0.002
		21	-0.023	-0.025	42.808	0.001
		22	-0.015	-0.018	44.220	0.001
		23	-0.023	-0.028	47.730	0.000
		24	-0.015	-0.020	49.155	0.000

Figure 6. The correlogram of AR(1) MA(2) MA(4) Adjusted ARIMA Model
Source: Authors' own research contribution

Inverse Roots of AR/MA Polynomial(s)
Specification: D_NATURAL_GAS_PRICES_ C AR(1) MA(2) MA(4)
Date: 08/13/22 Time: 15:56
Sample: 1/08/1997 8/02/2022
Included observations: 6430

AR Root(s)	Modulus	Cycle
-0.102317	0.102317	

No root lies outside the unit circle.
ARMA model is stationary.

MA Root(s)	Modulus	Cycle
-0.630206	0.630206	
0.630206	0.630206	
4.16e-17 ± 0.375309i	0.375309	4.000000

No root lies outside the unit circle.
ARMA model is invertible.

Figure 7. The stability of AR(1) MA(2) MA(4) Model Equation
Source: Authors' own research contribution

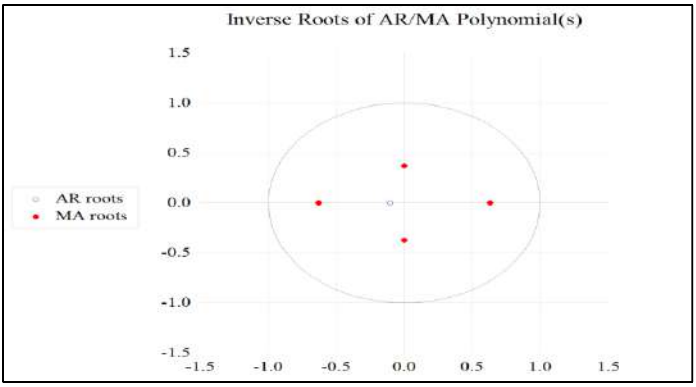

Figure 8. The representation of AR(1) MA(2) MA(4) equation stability
Source: Authors' own research contribution

Moreover, the forecast will be based on this adjusted ARIMA model. At the same time, according to the statistical results presented in Figure 7 and Figure 8, the modulus of the roots of the characteristic polynomial is less than 1, and therefore the equation of our selected adjusted ARIMA model is stable.

In our selected model, all the estimated coefficients are statistically significant at the 5% p-value, the predicted volatility is 13.11%, the R-Squared 6.75%) is the highest and the values, approximately equal to 2.00 of the Durbin- Waston Test confirmed the non-existence of the autocorrelation or serial correlations on the residual series.

A final aspect of our research consisted of performing the forecast analysis of the daily natural gas prices time series from 1997-2022. This step assumes the main essence of autoregressive integrated moving average models to provide a good prediction of future observations explained by past observations or historical data. In our case, the forecast was based on the AR(1) MA(2) MA(4) adjusted ARIMA model, and the results are presented in Figure 9.

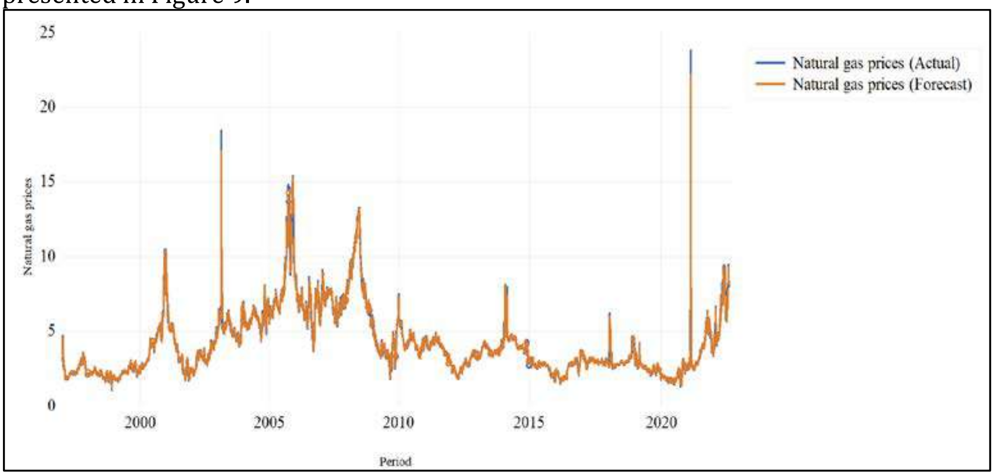

Figure 9. The forecast of natural gas prices in the period 1997-2022
Source: Authors' own research contribution

In general, we can state that this model provides a good prediction, and this model has the ability to capture any significant detail that influences the future value of natural gas prices. By visualizing Figure 9, we observe that the prediction from 2000 to 2018 is almost exact, but we notice a slight deviation in predicting the next period: 2019 to 2022. Overall, we can say that the forecast is good and appropriate (Bakar, Rosbi & Uzaki, 2018; Zhao et al., 2018).

At the same time, our model can capture the slight and predictable fluctuations in natural gas prices in the short and medium term. We must not omit the fact that there are other more advanced econometric methods, i.e. machine learning methods, and data mining techniques that are able to provide a more accurate forecasting analysis, especially in capturing the fluctuating movements in the long term of the time series.

Conclusions

The central aim of our study was to perform an estimation and forecasting analysis of the daily natural gas prices between 1997- 2022 by applying the ARIMA models.

In this sense, 18 ARIMA models were created in order to capture and determine the most statistically significant lags. In our case, these lags were: lag 1, lag 2, lag 4, lag 5, and lag 6. In the first phase/way, it was observed that the ARIMA (1,1,2) model is the most appropriate model regarding the fulfillment of robustness tests and diagnostics (i.e the significant number of coefficients, the highest value of the R-Squared, the lowest predicted volatility, and the lowest values of Akaike info criterion and Schwarz criterion). But, after performing other diagnostic tests of the selected ARIMA (1,1,2) model, we concluded that this model does not capture all the specific information from the lags, and consequently, the resulting correlogram was not flat and consistent.

Therefore, we resorted to the re-estimation analysis of another 6 Adjusted ARIMA Models, after which we could certainly say that the AR(1) MA(2) MA(4) Adjusted ARIMA Model is most appropriate in forecasting the daily natural gas prices time series.
Moreover, this final model could ensure a good forecast, producing better results than over-parameterized models. Also, we confirm that the price of natural gas continues to be volatile and fluctuating in the near future, with the predicted volatility ($Sigma^2$) reaching the value of approx. 13%.

Our results are almost similar to the studies of Busse et al.(2012); Farrokhi and Hassanzadeh (2017); Mishra (2012); Zhao et al. (2018) or Mashadihasanli (2022).

Also, according to the latest information provided by the U.S. Energy Information Administration, the Henry Hub price is expected to average $7.54/MMBtu in the second half of 2022 and then fall to an average of $5.10/MMBtu in 2023 amid rising natural gas production.

Consequently, our study provides an up-to-date picture of natural gas prices, having an informative and decision-making role in the case of financial investors, energy and financial companies, or other participants in the global financial markets and international trade.

Moreover, our study shows that autoregressive integrated moving average models prove efficient in estimating and measuring the future price of natural gas based on historical data.

The study results can set an example for researchers and practitioners working in the financial, equity, and stock market and can guide economic decision units and investors in these financial and economic areas. In modern financial markets, traders and practitioners have had trouble predicting the stock market price index, in our case the historical natural gas prices.

We are aware that there are other innovative and advanced econometric methods and models to offer a more precise and consistent forecasting analysis , such as: machine learning (for example: Neural Artificial Models, data mining techniques, etc.)

Another limitation of our study derives from the fact that ARIMA models do not have the capacity to capture those sudden and sharp natural gas price movements (i.e. spikes or jumps), being especially suitable for short-term forecast analysis. In further research, we aim to solve this aspect by applying machine learning tools, as well more sophisticated methodologies may be employed.

References

Bakar, N. A., Rosbi, S., & Uzaki, K. (2018). Evaluating Forecasting Method Using Autoregressive Integrated Moving Average (ARIMA) Approach for Shariah Compliant Oil and Gas Sector in Malaysia. *Journal of Mathematical and Computational Science , 1*(1), 19-33. https://jmcs.com.my/index.php/jmcs/article/view/30

Berrisch, J., & Ziel, F. (2022). Distributional modelling and forecasting of natural gas prices. *Journal of Forecasting , 41*, 1065-1086. https://doi.org/10.1002/for.2853

Busse, S., Helmholz, P., & Weinmann, M. (2012). Forecasting day ahead spot price movements of natural gas - An analysis of potential influence factors on basis of a NARX neural network. *Tagungsband der Multikonferenz Wirtschaftsinformatik (MKWI 2012 Conference)*, (pp. 1-13).

Castaneda, F., Schicks, M., Niro, S., & Hartmann, N. (2021). Forecasting natural gas price trends using random forest and support vector machine classifiers. *Journal of Energy Markets , 14*(4), 89-107. doi: 10.21314/JEM.2021.005

Farrokhi, A., & Hassanzadeh, R. (2017). Time Series Models to Predict the Monthly and Annual Consumption of Natural Gas in Iran. *Journal of Industral Startegic Management , 2* (2), 67-76.

Gonçu, A., Karahan, M. O., & Kuzubaș, T. U. (2019). Forecasting Daily Residential Natural Gas Consumption: A Dynamic Temperature Modelling Approach. *Boğaziçi Journal Review of Social, Economic and Administrative Studies , 33*(1), 25-46. doi: 10.21773/boun.33.1.3

Guan, R., Wang, A., Liang, Y., Fu, J., & Han, X. (2022). International Natural Gas Price Trends Prediction with Historical Prices and Related News. *Energies, 15*, 3573. https://doi.org/10.3390/en15103573

Hendikawati, P., Abdurakhman, S., & Tarno. (2020). A survey of time series forecasting from stochastic method to soft computing. *Journal of Physics: Conference Series Ahmad Dahlan International Conference on Mathematics and Mathematics Eduaction*, *1613*. https://doi.org/10.1088/1742-6596/1613/1/012019

Hosseinipoor, S. (2016). *Forecasting natural gas prices in the United States using Artificail Neural Networks.* A Degree Thesis of Master of Science in Natural Gas Engineering and Management, University of Oklahoma, Oklahoma. Doi:10.13140/RG.2.1.1284.8248

Hosseinipoor, S., & Hajirezaie, S. (2016, April). Application of ARIMA and GARCH Models in Forecasting the Natural Gas Prices. *Research*. Doi: 10.13140/RG.2.1.1925.5447

Manigandan, P., Alam, M. S., Alharthi, M., Khan, U., Alagirisamy, K., Pachiyappan, D., et al. (2021). Forecasting Natural Gas Production and Consumption in United States-Evidence from SARIMA and SARIMAX Models. *Energies, 14*, 6021. https://doi.org/10.3390/en14196021

Mashadihasanli, T. (2022). Stock Market Price Forecasting Using the Arima Model:an Application to Istanbul, Turkiye. *İktisat Politikası Araştırmaları Dergisi -Journal of Economic Policy Researches, 9*(2), 439-454. Doi: 10.26650/JEPR1056771

Mishra, P. (2012). Forecasting Natural Gas Price - Time Series and Nonparametric Approach. *Proceedings of the World Congress on Engineering*, *I*, 490-497. London, U.K.

Mouchtaris, D., Sofianos, E., Gogas, P., & Papadimitriou, T. (2021). Forecasting Natural Gas Spot Prices with Machine Learning. *Energies, 14*(18), 5782. https://doi.org/10.3390/en14185782

Nyangarika, A. M., & Tang, B.-j. (2018). Oil Price Factors: Forecasting on the Base of Modified ARIMA Model. In I. C. Science (Ed.), *2nd International Conference on Power and Energy Engineering (ICPEE 2018)*, *192*, 1-10. doi: 10.1088/1755-1315/192/1/012058

Siddiqui, A. W. (2019). Predicting Natural Gas Spot Prices Using Artificial Neural Network. *2nd International Conference on Computer Applications & Information Security (ICCAIS)*, (pp. 1-16). Doi: 10.1109/CAIS.2019.8769586

Tamba, J. G., Essiane, S. N., Sapnken, E. F., Koffi, F. D., Nsouandele, J. L., Soldo, B., et al. (2018). Forecasting Natural Gas: A Literature Survey. *International Journal of Energy Economics and Policy, , 8*(3), 216-249. https://www.econjournals.com/index.php/ijeep/article/view/6269

Viacaba, A., Poursaeidi, M. H., & Kundakcioglu, O. E. (2012). Natural Gas Price Forecasting via Selective Support Vector Regression. In G. Lim, & J. Herrmann (Eds.), *Proceedings of the 2012 Industrial and Systems Engineering Research Conference* (pp. 1-9). https://www.proquest.com/openview/6c3b482f807715176f6d825600c148e0/1?pq-origsite=gscholar&cbl=51908

Wong-Parodi, G., Dale, L., & Lekov, A. (2006). Comparing price forecast accuracy of natural gas models and futures markets. *Energy Policy, 34*(18), 4115-4122. https://doi.org/10.1016/j.enpol.2005.08.013

Zhao, Z., Fu, C., Wang, C., & Miller, C. J. (2018). Improvement to the Prediction of Fuel Cost Distributions Using ARIMA Model. *2018 IEEE Power & Energy Society General Meeting (PESGM)* (pp. 1-5). Doi:10.1109/PESGM.2018.8585984

U.S. Energy Information Administration (2022). *Short Term Energy Outlook.* https://www.eia.gov/outlooks/steo/

European Central Bank (2022). *Natural gas dependence and risks to euro area activity.*

https://www.ecb.europa.eu/pub/economic-bulletin/focus/2022/html/ecb.ebbox202201_04~63d8786255.en.html