BAREKENG: Journal of Mathematics and Its Applications
December 2025 Volume 19 Issue 4 Page 2647–2660
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol19no4pp2647-2660

PREDICTION OF NATURAL GAS PRICES ON THE NEW YORK MERCANTILE EXCHANGE BASED ON A PULSE FUNCTION INTERVENTION ANALYSIS APPROACH

Sediono [1*], **Toha Saifudin** [2], **Maria Setya Dewanti** [3], **Aurelia Islami Azis** [4]

[1,2,3,4]*Department of Mathematics, Faculty of Science and Technology, Universitas Airlangga*
Jl. Dr. Ir. H. Soekarno, Mulyorejo, Surabaya, Jawa Timur, 60115, Indonesia

Corresponding author's e-mail: * **sediono101@gmail.com**

Article History:

Received: 10th February 2025
Revised: 8th April 2025
Accepted: 2nd June 2025
Available online: 1st September 2025

Keywords:

ARIMA;
Intervention Analysis;
Nature Gas;
Pulse Function;
Russia-Ukraine Conflict.

ABSTRACT

Natural gas is a key energy commodity with significant global economic impact, and its pricing is influenced by factors like weather, energy policies, geopolitics, and supply-demand balance. The Russia-Ukraine conflict disrupted Russia's gas exports, causing price volatility and affecting global markets, including Indonesia. This has heightened the need for accurate price prediction to support policy and investment decisions. Previous studies show ARIMA-GARCH models predict well but need pulse function intervention for sudden shocks. This study aims to apply pulse function intervention analysis, which captures the immediate effects of external events on time-series data, to improve the precision of natural gas price forecasts, aiding government and industry decision-makers. The optimal intervention model for predicting natural gas prices on the New York Mercantile Exchange is the Probabilistic ARIMA (0,2,1) with a pulse function intervention order of b=0, r=2, and s=0. Using this model with the pulse function intervention approach yields consistent fluctuation patterns over time and achieves a MAPE value of 12.2586%, indicating that the model provides good predictive accuracy.

How to cite this article:

Sediono, T. Saifudin, M. S. Dewanti, and A. I. Azis, "PREDICTION OF NATURAL GAS PRICES ON THE NEW YORK MERCANTILE EXCHANGE BASED ON A PULSE FUNCTION INTERVENTION ANALYSIS APPROACH," *BAREKENG: J. Math. & App.*, vol. 19, iss. 4, pp. 2647-2660, December, 2025.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

Research Article · Open Access

1. INTRODUCTION

Natural gas is one of the main energy commodities that plays an important role in the world economy. The use of natural gas by utilizing technology for exploration, production, and distribution opens up opportunities for a country in the investment sector [1]. Fluctuations in natural gas prices not only affect companies engaged in the energy sector that are directly involved in the production and distribution of natural gas, but also affect the industrial sector whose industrial raw materials depend on natural gas. Natural gas prices on the New York Mercantile Exchange are influenced by many factors, such as weather, energy policy, the geopolitical situation, and the balance between supply and demand.

The prolonged conflict between Russia and Ukraine is one of the causes of the disruption of natural gas import-export activities from Russia, which is the world's second-largest exporter of natural gas, to the world market [2]. The conflict between Russia and Ukraine also impacted energy policy as well as the availability of fossil fuels in Indonesia. For instance, Nusantara Regas, an Indonesian company, plans to import up to 2 million metric tons of LNG annually from the United States. Analysts and market sources project that Indonesia may face a shortfall of around 10 LNG shipments in 2024 due to rising domestic demand, decreasing local supplies, and existing export commitments. This shortfall is expected to be addressed through additional imports. Since Russia's invasion of Ukraine, which culminated in 2022, Russia's natural gas exports have decreased by at least 40% from the previous year. Although Russian gas production has declined as a result of the invasion of Ukraine, Russia remains the owner of the largest natural gas reserves and is still the second largest gross gas exporter in the world, after the United States [3]. The geopolitical impact of the war has led to higher price fluctuations and added volatility to natural gas futures.

Given the uncertainty of natural gas futures prices in the global market due to the Russia-Ukraine conflict, there is a need for in-depth research to understand and predict these price fluctuations so that energy and economic policies can be designed appropriately. Choosing a good and appropriate prediction method will help in identifying and quantifying the impact of the Russia-Ukraine conflict on changes in natural gas futures prices. Accurate predictions also help many parties, such as governments, companies engaged in the energy sector, and investors, in making decisions such as investments.

There are several previous studies related to natural gas price prediction, such as [4], which predicts natural gas prices using the ARIMA and AR–GARCH models. Furthermore, [4] reports that the MAPE for the ARIMA (1,1,1) model does not provide a significant probability value and thus fails to fit the data properly. After detecting the presence of ARCH effects, the study applies an AR (1)–GARCH (1,1) model, which is found to be the best-fitted model with a highly significant probability value ($P < 0.0001$), along with low mean squared error (MSE) and root mean squared error (RMSE) values, indicating high prediction accuracy. Although the ARIMA–GARCH model provides accurate predictions under normal conditions, it lacks the ability to account for sudden shocks or unexpected structural changes in the data. Therefore, it is necessary to conduct further analysis using an intervention model, such as the pulse function, which is specifically designed to capture the effects of abrupt events on time series data like natural gas prices. Another relevant study is presented in [5], which applies the pulse function intervention analysis to predict the JCI. Based on this study, the best intervention model for forecasting JCI close price data is the ARIMA (3,1,0) model with $b = 0$, $r = 0$, $s = 11$, and a MAPE value of 0.98%, which means the model is able to do forecasting very well. The occurrence of extreme trend changes in August 2022 states that the data has intervened, so it is necessary to model the pulse function intervention. Intervention itself is a shock that arises due to internal and external factors [6]. While intervention analysis is a method in time series analysis that considers the impact of an intervention that results in changes to the average function or trend in the data. Box and Tiao (1975) revealed that the intervention function has two forms, namely the step function and pulse function. Intervention analysis with a pulse function is applied to overcome interventions that are temporary and only occur at a certain point in time, such as the Russia-Ukraine conflict [7].

This research was conducted to predict natural gas prices on the New York Mercantile Exchange using a method that has not previously been used in predicting them, namely the pulse function intervention analysis approach. The pulse function intervention analysis method was chosen due to its ability to identify and quantify the impact of external events on time series data, such as geopolitical events or economic policies, which are often difficult to capture by traditional models such as ARIMA. A key advantage of this method is its ability to assess the direct influence of interventions at any given time, which allows for more precise predictions and responsiveness to changes that occur. Therefore, this research provides accurate forecasts of natural gas price fluctuations on the New York Mercantile Exchange, enabling the government to design better energy policies, companies in the energy sector to optimize production and trading strategies, and investors to make informed decisions on market risks and investment timing.

2. RESEARCH METHODS

The research method contains explanations about the variables and data sources, literature review about methods, and research stages.

2.1 Data Sources and Variables

The data used in this study are weekly data of natural gas closing prices on the New York Mercantile Exchange obtained from the official investing.com website from the first week of January 2020 to the last week of October 2022, with a total of 147 data points. The data has gone through a Box-Cox transformation and differencing process to stabilize the variance before modeling. The data is divided into two parts, namely training data and testing data. Training data used to build the model, namely data for the first week of January 2020 to the first week of August 2022, with a total of 136 data points. Meanwhile, testing data is used to compare prediction results with actual data, namely, data from the third week of August 2022 to the last week of October 2022, with a total of 10 data points.

2.2 Literature Review about Methods

2.2.1 Autoregressive Integrated Moving Average (ARIMA)

ARIMA is a statistical method used in time series analysis and forecasting [8]. In forecasting, this method uses past data. ARIMA modeling is a time series analysis method that consists of three stages: identification, estimation, and diagnostic checks. In the identification stage, the data is analyzed to determine the appropriate parameters for the model. The estimation stage involves calculating those parameters using the available time series data. Finally, diagnostic checks are performed to ensure the model meets the key assumptions of ARIMA, which include stationarity, normality of residuals, whiteness (residuals behaving as white noise), and unbiasedness of the errors [9]. The ARIMA model itself consists of three main components: Autoregressive (AR), Moving Average (MA), and Differencing (I), which together capture different characteristics of the time series data [10]. The ARIMA (p, d, q) model is a forecasting model that applies differencing to transform non-stationary time series data into a stationary form before modeling it using autoregressive (AR) and moving average (MA) components. The model of ARIMA (p, d, q) is as follows:

$$\phi_p(B)(1-B)^d Z_t = \theta_q(B)a_t \tag{1}$$

The general model for ARIMA (p, d, q) with differencing 1 is as follows:

$$\begin{aligned} Z_t = \left(1+\phi_p\right)Z_{t-1} + (\phi_1+\phi_2)Z_{t-2} + \cdots + \left(\phi_p+\phi_{p-1}\right)Z_{t-p} - \phi_p Z_{t-p-1} \\ +a_t - \theta_1 a_{t-1} - \cdots - \theta_q\, a_{t-q}; \qquad a_t \sim N(0, \sigma_a^2) \end{aligned} \tag{2}$$

With Z_t represents the variable value at time t, while Z_{t-1}denotes the variable value before time t. The autoregressive coefficient is represented by ϕ_i, and the moving average coefficient is denoted by θ_i. The symbol a_t stands for white noise (residual) at time t, whereas a_{t-1} represents white noise (residual) before time t. The autoregressive polynomial of degree p is expressed as $\phi_p(B)$, and the polynomial moving average of degree q is denoted by $\theta_q(B)$.

In determining the best ARIMA model, it is necessary to identify the autocorrelation function (ACF) and partial autocorrelation function (PACF) plots. This is followed by parameter estimation using the ordinary least squares (OLS) method. The OLS method works by minimizing the sum of the squares of the difference between the observed value and the value predicted by the model [11]. Time series data requires classical assumptions, must be stationary, and unbiased. The ARIMA (p, q) model can be used if a_t meets the white noise assumption and is identically independent distributed (IID) $N(0, \sigma_a^2)$.

2.2.2 Best Model Criteria

In evaluating models, it is important to consider various criteria for both model selection and forecasting accuracy. Commonly used model selection criteria include the Akaike Information Criterion

(AIC) and the Schwarz Bayesian Criterion (SBC), while the Mean Squared Error (MSE) is frequently used to assess the model's forecasting performance [12].

Mean Squared Error (MSE) measures the average squared difference between the predicted and actual values of the dependent variable. This value is obtained by summing the squared difference between the prediction and the actual, then dividing it by the number of observations. The MSE equation is as follows [13]:

$$MSE = \frac{1}{n}\sum_{t=1}^{n}\left(Z_t - \hat{Z}_t\right)^2 \tag{3}$$

With:

Z_t : The actual value at time t

$\hat{Z}_t$: The predicted value (from the model) at time t

n : Total number of observations

Mean Absolute Percentage Error (MAPE) is a method to measure the accuracy of a prediction model by assessing the error as a percentage. It allows comparison of prediction errors between different models or methods. The main function of MAPE is to assess the quality of model predictions and identify the need for model improvement if necessary. To calculate MAPE, it is usually formulated in the form of the following equation:

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{Z_t - \hat{Z}_t}{Z_t}\right| \times 100\% \tag{4}$$

The lower the MAPE value, the more accurate the model is in predicting actual values. However, a MAPE value that is too low may indicate the risk of overfitting, so the optimal value is one that is low enough to indicate accuracy without overfitting [14]. For MAPE itself, there is a range of values that can be used as a measurement material regarding the ability of a forecasting model; the range of values can be seen in Table 1 [15]:

Table 1. Range Value of MAPE

MAPE Range	Description
$< 10\%$	Excellent Forecasting Model Ability
$10 - 20\%$	Good Forecasting Model Capability
$20 - 50\%$	Decent Forecasting Model Capability
$> 50\%$	Poor Forecasting Model Ability

2.2.3 Intervention Analysis

Time series intervention analysis is a statistical method used to study the impact of an intervention on time series variables, whether it is a deliberate change or an unexpected event that affects the time series pattern [16]. This method determines an intervention model with an intervention time T, which is determined through ARIMA modeling before the intervention, and an indicator variable is added to represent the intervention, where the indicator is either 1 or 0. In general, interventions are divided into two types: step and pulse. The step function describes a sudden and continuous change in the level of a variable, which can be mathematically written in the following equation:

$$I_t^{(T)} = S_t^{(T)} = \begin{cases} 1, & t < T \\ 0, & t \geq T \end{cases} \tag{5}$$

While the pulse function shows a sudden change that only lasts for a certain time, mathematically it can be written in the following equation:

$$I_t^{(T)} = P_t^{(T)} = \begin{cases} 1, & t = T \\ 0, & t \neq T \end{cases} \tag{6}$$

With:

$I_t^{(T)}$: An indicator function that changes value based on time t relative to a specific time T

$S_t^{(T)}$: A step function that is equal to 1 for all times before T, and 0 from time T onward. Represents a sudden drop at time T

$P_t^{(T)}$: A pulse function that is equal to 1 only at time T, and 0 at all other times. Represents a sudden spike at time T

The general form of the intervention model is as follows:

$$X_t = \frac{\omega_s(B)B^b}{\delta_r(B)} I_t^{(T)} + N_t \tag{7}$$

With X_t represents the response variable at time t, while I_t denotes the intervention variable. The parameter b indicates the delay time for the effect of the intervention I on X, and $\boldsymbol{s}$ represents the duration of the intervention's effect on the data after b periods. The pattern of the intervention effect after $b + s$ periods since the intervention event at time T is denoted by r. The terms $\omega_s = \omega_0 - \omega_1 B - \cdots - \omega_s B^s$ and $\delta_s = 1 - \delta_q B - \cdots - \delta_r B^r$ are used to describe the intervention effect mathematically. Finally, N_t refers to the best ARIMA model without the intervention effect. Mathematically, the ARIMA model without the effect of intervention can be written as follows:

$$N_t = \frac{\theta_q(B)}{\phi_p(B)(1-B)^d} a_t \tag{8}$$

The Cross Correlation function is used to determine the order of the intervention model parameters, namely b, s, and r, which serve as the basis for determining the transfer function. This transfer function describes the relationship between inputs and outputs in a system, with the aim of predicting outputs based on given inputs.

2.3 Research Stages

The steps in this research are as below:

1. Modeling natural gas price data on the New York Mercantile Exchange using an intervention analysis approach.
   a. Data exploration.
      i. Plot the time series of all the data and expose the extreme variables that allow intervention.
      ii. Divide the data into two groups, namely training data for data that occurs before an extreme event and testing data for data that occurs after an extreme event.
   b. Estimating the best ARIMA model for the training data.
      i. Checking the stationarity of the data in mean and variance based on the time series plot, ACF plot, PACF plot, and ADF test.
      ii. Perform a Box-Cox transformation or differencing process for testing data that does not meet the stationarity assumption.
      iii. Identify some candidate models based on the ACF and PACF plots of the testing data that have met the stationarity assumption.
      iv. Perform parameter estimation of the training data ARIMA model using the Ordinary Least Square (OLS) method.
      v. Perform diagnostic tests on the estimated model, namely parameter significance test, residual white-noise test, and data residual normality test.
      vi. Selecting the best model based on the lowest AIC, SIC, and MSE values.
      vii. Predicting the testing data based on the best ARIMA model of the testing data.

c. Modeling with an intervention analysis approach.

i. Determine the order of the intervention model (b, s, and r) by identifying the Cross Correlation Function (CCF) plot between the testing data and the predicted data of the selected ARIMA model.

ii. Perform parameter estimation of the intervention model using the Ordinary Least Square (OLS) method.

iii. Perform diagnostic tests on the estimated model, namely parameter significance, residual white-noise test, and data residual normality test.

2. Predict and analyze natural gas prices on the New York Mercantile Exchange market using an intervention analysis approach.

a. Write down the best model obtained.

b. Checking the accuracy of the model in making predictions using testing data.

c. Predicting the best model data.

3. Draw interpretations and conclusions from the model obtained in the previous stages.

3. RESULTS AND DISCUSSION

3.1 Descriptive Analysis of Natural Gas Price Data on the New York Mercantile Exchange

Descriptive analysis is carried out to describe the overall natural gas price data of the New York Mercantile Exchange as a research variable and to determine the extreme increases that occur in natural gas price data with a weekly period, starting from the first week of January 2020 to the last week of October 2022.

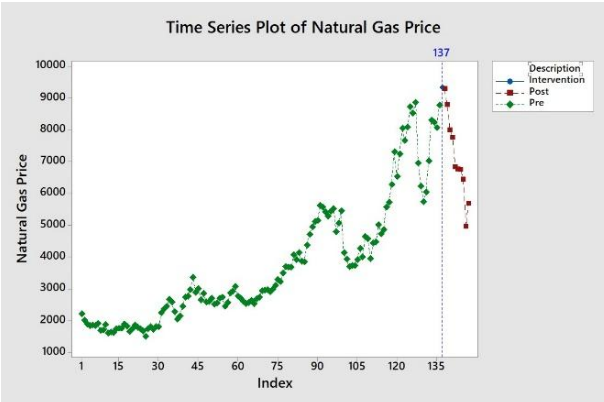

Figure 1. Time Series Plot of Natural Gas Price Data of the New York Mercantile Exchange

Figure 1 shows the pattern of natural gas price data on the New York Mercantile Exchange experiencing fluctuations and a sharp increase at the 137th point, namely on August 14, 2022, with a natural gas price of 9336 USD. Based on this increase, the data is divided into two, namely training data for data that occurred before the extreme increase and testing data for data that occurred after the extreme increase. The calculation of the amount of data, average, variance, median, minimum value, and maximum value is as in **Table 2**.

Table 2. Descriptive Analysis of Natural Gas Prices on the New York Mercantile Exchange

Variable	N	Mean	Variance	Minimum	Median	Maximum
Data	147	3,995	4,473,190	1,495	3,215	9,336
Training	136	3,726	3,706,130	1,495	2,966	8,850
Testing	10	7,128	1,801,986	4,959	6,797	9,296

Table 2 shows the price of natural gas on the New York Mercantile Exchange before the extreme increase, namely the first week of January 2020 to the first week of August 2022, has an average of 3,726 USD with a variance of 3,706,130 USD. Meanwhile, the average price of natural gas on the United States futures market after the extreme increase amounted to 7,128 USD with a variance of 1,801,986 USD. The lowest US natural gas futures price was obtained at 1,495 USD on June 21, 2020, or the third week of June 2020. The highest US natural gas futures price was 9,336 USD on August 14, 2022, or the second week of August in 2022, which is the point of extreme increase in this study.

3.2 Modeling Intervention Analysis of Natural Gas Price Data on the New York Mercantile Exchange

The first stage in forming an intervention model is to divide the data into two. Based on the time series plot in **Figure 1**, the intervention point occurs at point 137, which is on August 14, 2022, so that the data is divided into training data, namely, data from point 1 of the first week of January 2020 to point 136 of the first week of August 2022. The testing data is the data from the 138th point of the third week of August 2022 to the 147th point of the last week of October 2022.

3.2.1 ARIMA Modeling of Training Data

Training data is used to determine the best ARIMA model. This is to identify whether the data has been stationary in average, meaning it does not have a clear trend or pattern, and stationary in variance, meaning it does not have large fluctuations.

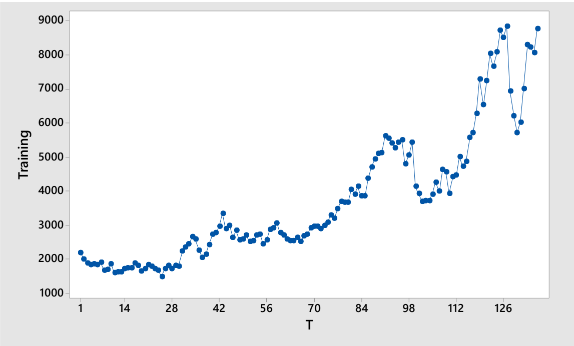

Figure 2. Plot of Time Series Data Testing Natural Gas Prices on the New York Mercantile Exchange

Based on **Figure 2**, the plot of natural gas prices on the New York Mercantile Exchange before the intervention shows an upward trend and uneven distribution of data. This indicates data non-stationarity in average and variance. To show the assumption that the training data has not been stationary in variance, it can be seen through the Box-Cox transformation in **Figure 3** (a). Then, checking the stationarity of the transformation results in average with the differencing process at lag 1 is shown in **Figure 3** (b).

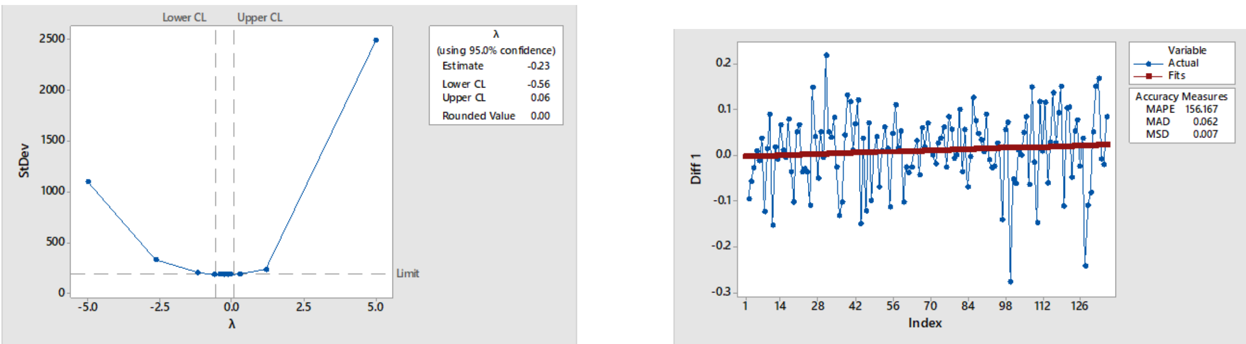

Figure 3. (a) Plot of Box-Cox Transformation Results and (b) Trend Analysis Plot of Differencing Data 1

Based on **Figure 3** (a), the rounded value (λ) of 0 is obtained, which proves that the data is not yet stationary in variance and needs to be transformed into $ln\ Z_t$. Based on **Figure 3** (b), the trend goes up slightly. This indicates that the data is still not stationary in average. To ensure that the data is still not stationary in average, it is necessary to identify the ACF and PACF plots as shown in **Figure 4** below.

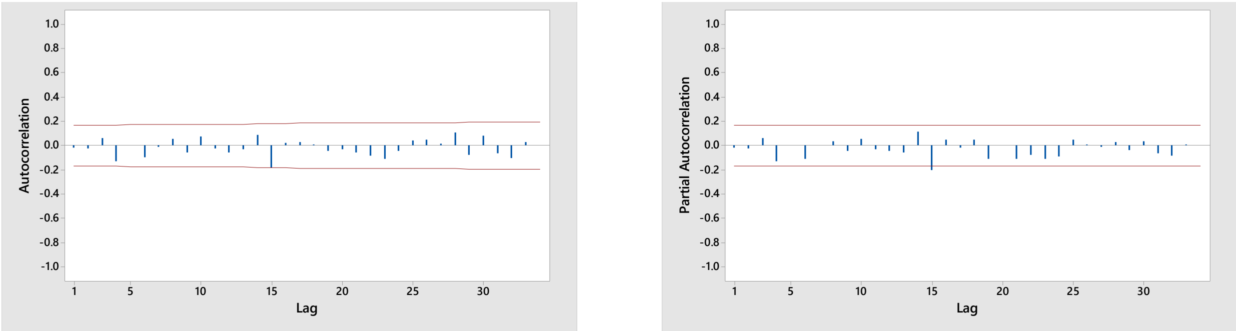

Figure 4. (a) Autocorrelation Function Plot and (b) Partial Autocorrelation Function Plot of Differencing 1 Data

Figure 4 shows that there is no lag out of the boundary line, so that differencing is required again with lag 1 to obtain optimal stationarity to the average. The results of the data plot that has undergone Box-Cox transformation and differencing 2 can be seen in Figure 5 below.

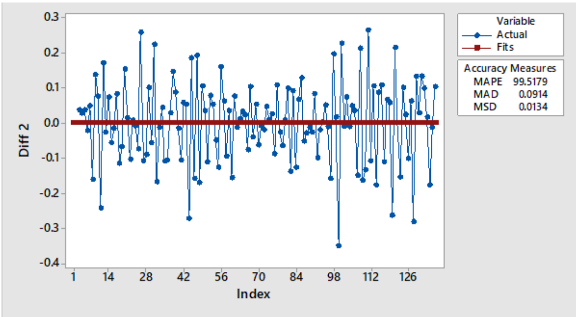

Figure 5. Trend Analysis Plot of Differencing 2 Data

Based on Figure 5, there is no significant upward or downward trend. This indicates that the data pattern tends to be stationary. To ensure that the data have been stationary in average and variance, it is necessary to conduct an ADF test, whose results are presented in Table 3.

Table 3. Augmented Dickey Fuller Test Results Differencing 2 Data

t-statistics	P-Value
-7.7863	0.01

Based on the ADF hypothesis, namely: $H_0 : Data\ non-stationary$ and $H_1 : Data\ stasionary$. If $P-value \leq 0.05$ then it is stated that it has failed to receive H_0. Based on Table 3, the p-value of 0.01 is less than 0.05, indicating that the data has reached stationarity. Furthermore, a temporary model estimation is carried out through identification on the ACF and PACF plots, as in Figure 6.

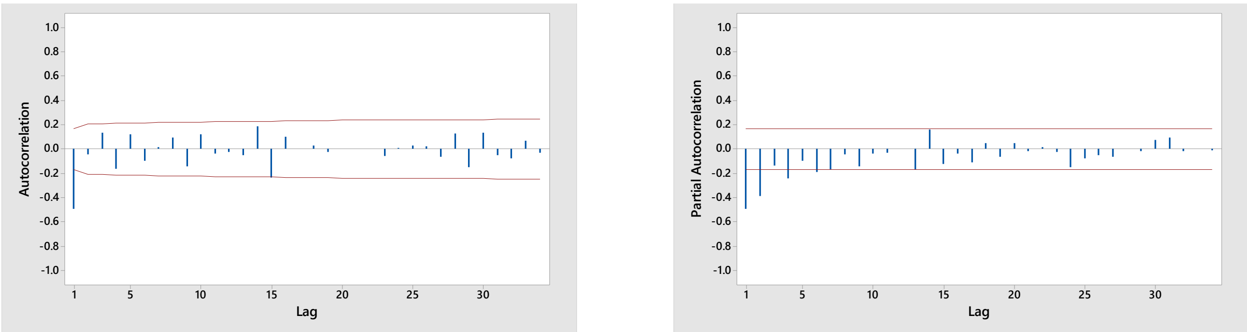

Figure 6. (a) Autocorrelation Function Plot and (b) Partial Autocorrelation Function Plot of Differencing 2 Data

The results of the ACF plot in Figure 6 (a) show that there is a lag that comes out of the boundary line, namely at lag 1. It shows that the estimated MA (1) model is obtained with a q order of 1. While the results of the PACF plot in Figure 6 (b) show that there is a lag that comes out of the boundary line, namely at lag 1 and lag 2. The results of the data stationarity test show that the data is stationary when given a differencing treatment 2 times, so that the order d is 2. The temporary conjecture models obtained are ARIMA (2,2,1), ARIMA (1,2,1), ARIMA (0,2,1), ARIMA (2,2,0), and ARIMA (1,2,0).

After obtaining the conjecture model, parameter estimati-on and parameter significance test of the conjecture model will be carried out. The significance test aims to show whether the parameters are feasible for use in modeling. The results of the presumptive model analysis are presented in Table 4 below.

Table 4. Conjectured ARIMA Model

Model		Parameter	Estimate	*P*-Value	Description
ARIMA (2,2,1)	Probabilistic	AR 1 AR 2 MA 1	0.025 -0.1349 1.00331	0.779 0.133 0.000	Not Significant
	Deterministic	AR 1 AR 2 MA 1 Constant	-0.0231 -0.0418 0.9877 0.000201	0.794 0.636 0.000 0.417	Not Significant
ARIMA (1,2,1)	Probabilistic	AR 1 MA 1	-0.0202 0.9885	0.816 0.000	Not Significant
	Deterministic	AR 1 MA 1 Constant	-0.0244 0.988 0.0001	0.781 0.000 0.663	Not Significant
ARIMA (0,2,1)	Probabilistic	MA 1	0.98712	0.000	**Significant**
	Deterministic	MA 1 Constant	0.9825 0.000206	0.000 0.521	Not Significant
ARIMA (2,2,0)	Probabilistic	AR 1 AR 2	-0.6913 -0.3902	0.000 0.000	**Significant**
	Deterministic	AR 1 AR 2 Constant	-0.6914 -0.3903 0.00159	0.000 0.000 0.844	Not Significant
ARIMA (1,2,0)	Probabilistic	AR 1	-0.4973	0.000	**Significant**
	Deterministic	AR 1 Constant	-0.4974 0.00147	0.000 0.866	Not Significant

Based on Table 4, it is found that ARIMA (0,2,1) Probabilistic, ARIMA (2,2,0) Probabilistic, and ARIMA (1,2,0) Probabilistic have a p-value of less than 0.05, so the model can be said to be significant. The next step is to test the residual assumptions on the estimated model that has been significant, namely the white noise test to prove there is a correlation between the residuals through the Ljung-Box test and the residual normality test to determine that the residual data has been normally distributed through the Kolmogorov-Smirnov test, whose results are presented in Table 5 below.

Table 5. Results of White Noise Test and Residual Normality Test

Model	Ljung-Box's P-Value				MSE	Normalities P-Value	Description
	Lag 12	Lag 24	Lag 36	Lag 48			
ARIMA (0,2,1) Probabilistic	0.714	0.683	0.842	0.942	0.0066955	0.135	Normal Distributed
ARIMA (2,2,0) Probabilistic	0.155	0.254	0.37	0.623	0.00868	0.144	Normal Distributed
ARIMA (1,2,0) Probabilistic	0.006	0.031	0.114	0.323	0.0101499	0.045	Not Normally Distributed

In Table 5, it is found that Probabilistic ARIMA (0,2,1) and Probabilistic ARIMA (2,2,0) have fulfilled both assumptions. Furthermore, in determining the best model, it is necessary to identify the smallest MSE value of the model that has been fulfilled. From Table 5, it can be seen that the smallest MSE value is found in the Probabilistic ARIMA (0,2,1) model, so that the model is the best model, which mathematically can be written as follows.

$$\phi_p(B)(1-B)^d\dot{Z}_t = \theta_q(B)a_t$$

With the value of $p = 0, d = 2,$ and $q = 1,$ then

$$\phi_0(B)(1-B)^2\dot{Z}_t = \theta_1(B)a_t$$

$$(1-B)^2\dot{Z}_t = (1-\theta_1 B)a_t$$

$$(1-2B-B^2)\dot{Z}_t = (1-\theta_1 B)a_t$$
$$\dot{Z}_t = \frac{(1-\theta_1 B)a_t}{1-2B-B^2}$$
$$\dot{Z}_t = (1-(2-\theta_1)B + (3-2\theta_1)B^2 + \cdots)a_t$$

Substitute the value of $\theta_1 = 0.98712$ obtained from **Table 5**.

$$\dot{Z}_t = \big(1-(2-0.98712)B + \big(3-2(0.98712)\big)B^2 + \cdots\big)a_t$$
$$\dot{Z}_t = (1-1.01279B + 1.02558B^2 + \cdots)a_t$$
$$\dot{Z}_t = a_t - (1.01279)a_{t-1} + (1.02558)a_{t-2} + \cdots$$

with $\dot{Z}_t = \ln Z_t - \ln Z_{t-1}$

The $\dot{Z}_t$ equation is the best ARIMA model without the influence of the intervention used as noise or N_t.

3.2.2 Pulse Function Intervention Analysis

After obtaining the best ARIMA model, the next step is to determine the value of the intervention parameters, namely b, r, and s through the cross-correlation plot on the testing data with the ARIMA model prediction data. The results of the cross-correlation plot can be seen in **Figure 7** below.

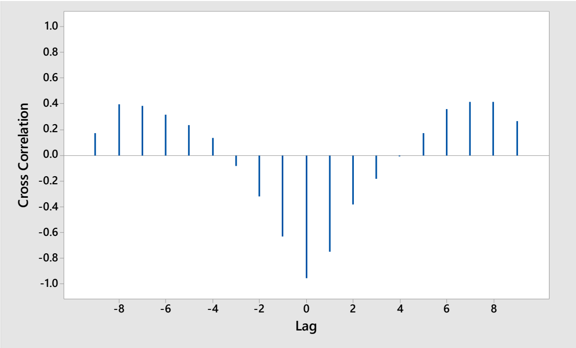

Figure 7. Plot of Cross Correlation Function

Based on **Figure 7,** it can be identified that the intervention order b is 0 because there is no delay when the intervention effect begins to occur, the intervention order r is 2 because it forms a wave pattern, and the intervention order s is 0 because the length of an intervention has no effect on the data after lag 0. After obtaining the intervention order, the parameter estimation of the pulse function intervention model is then carried out. The results of the pulse function intervention model parameter estimation are presented below.

Table 6. Intervention Model Significance Test Results

Intervention Model	Parameter	Estimate	P-Value	Description
ARIMA (0,2,1) with $b=0, r=2, s=0$	MA (1)	0.96920	< 0.0001	Significant
	ω_0	0.05737	0.0444	Significant
	δ_1	1.79964	< 0.0001	Significant
	δ_2	-1	0.0477	Significant

Based on **Table 6**, it can be seen that all p-values are less than 0.05, so the parameters of the ARIMA (0,2,1) model with $b=0$, $r=2$, and $s=0$ can be said to be significant. Next, the residual assumptions of white noise and normality are checked. The results of the white noise test and normality test can be seen in **Table 7** below.

Table 7. Intervention Model Residual Assumption Test Results

***P*-Value White Noise**				**Normalities *P*-Value**	**Description**
Lag 6	Lag 12	Lag 18	Lag 24		
0.7202	0.8704	0.7345	0.7867	0.0841	Normal Distributed

Based on **Table 7**, it is found that the intervention model has fulfilled the white noise assumption because the p-value of all lags that appear is more than 0.05. In addition, the intervention model residuals are normally distributed because the p-value of normality is more than 0.05.

3.3 Prediction and Analysis of US Natural Gas Futures Exchange Intervention Data

Based on the previous discussion, the intervention model built has met the assumptions, so the model can be used to predict natural gas prices on the New York Mercantile Exchange. The ARIMA (0,2,1) model with intervention parameters $b = 0, r = 2$, and $s = 0$ is written as follows.

$$\dot{Z}_t = \frac{\omega_s(B)B^b}{\delta_r(B)} P_t^T + N_t;\ N_t = (1 - 1.01279B + 1.02558B^2 + \cdots)a_t$$

With $b = 0, r = 2$, and $s = 0$ then

$$\dot{Z}_t = \frac{\omega_0(B)B^0}{\delta_2(B)} P_t^T + N_t$$

$$\dot{Z}_t = \frac{\omega_0(B)B^0}{(1 - \delta_1 B - \delta_2 B^2)} P_t^{(137)} + N_t$$

$$\dot{Z}_t = \omega_0 \frac{B}{(1 - \delta_1 B - \delta_2 B^2)} P_t^{(137)} + N_t$$

$$\dot{Z}_t = \omega_0(1 + (\delta_1 B + \delta_2 B^2) + (\delta_1 B + \delta_2 B^2)^2 + \cdots) P_t^{(137)} + N_t$$

$$\dot{Z}_t = \omega_0(1 + \delta_1 B + \delta_2 B^2 + \delta_1^2 B^2 + \delta_2^2 B^4 + 2\delta_1\delta_2 B^3 + \cdots) P_t^{(137)} + N_t$$

Substitute the value of $\omega_0 = 0.05737$, $\delta_1 = 1.79964$, and $\delta_2 = -1$ obtained from Table 6.

$$\dot{Z}_t = 0.05737(1 + 1.79964B + (-1)B^2 + (1.79964)^2 B^2 + (-1)^2 B^4 + 2(1.79964)(-1)B^3 + \cdots) P_t^{(137)} + N_t$$

$$\dot{Z}_t = 0.05737(1 + 1.79964B + 4.2387B^2 + B^4 - 3.59928B^3 + \cdots) P_t^{(137)} + (1 - 1.01279B + 1.02558B^2 + \cdots)a_t$$

$$\dot{Z}_t = (0.05737 + 0.103245B + 0.24317B^2 + 0.05737B^4 - 0.20649B^3 + \cdots) P_t^{(137)} + (1 - 1.01279B + 1.02558B^2 + \cdots)a_t$$

$$\dot{Z}_t = (0.05737)P_t^{(137)} + (0.103245)P_{t-1}^{(137)} + (0.24317)P_{t-2}^{(137)} + (-0.20649)P_{t-3}^{(137)} + (0.05737)P_{t-4}^{(137)} + \cdots + a_t - (1.01279)a_{t-1} + (1.02558)a_{t-2} + \cdots$$

From the results of the model that has been built, predictions will be obtained for the next 10 weeks. The prediction results can be seen in Table 8 below.

Table 8. Prediction Results of Natural Gas Prices on the United States Futures Exchange

Date	Predicted Results	Actual Data	APE
8/21/2022	9,930.837	9,296	0.068291
8/28/2022	9,666.291	8,786	0.100193
9/4/2022	8,877.72	7,996	0.11027
9/11/2022	7,850.245	7,764	0.011108
9/18/2022	7,465.891	6,828	0.093423
9/25/2022	6,475.622	6,766	0.042917
10/2/2022	6,428.522	6,748	0.047344
10/9/2022	6,498.327	6,453	0.007024
10/16/2022	6,344.858	4,959	0.279463
10/23/2022	4,965.653	9,296	0.465829
MAPE			12.25863

Based on Table 8, it can be seen that the prediction results have a MAPE value of 12.2586%, which, based on this value, indicates that the model's ability to predict is good. Furthermore, a comparison is made between the forecasting results and the testing data to ascertain whether the model built is good. The comparison graph between the forecasting results and the testing data can be seen in Figure 8.

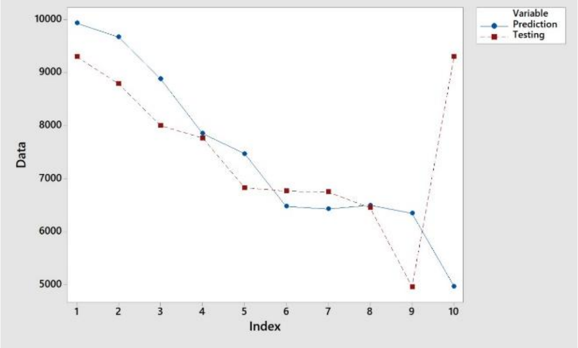

Figure 8. Comparison Plot of Actual and Predicted Data

Based on **Figure 8**, it can be concluded that the forecasting results for the third week of August 2022 to the last week of October 2022 do not have much difference from the actual data. This indicates that the developed model has good quality.

4. CONCLUSION

The best intervention model for natural gas prices in the New York Mercantile Exchange market is Probabilistic ARIMA (0,2,1) with pulse function intervention order $b = 0$, $r = 2$, and $s = 0$. The prediction results of this model show a price fluctuation pattern that is consistent with the actual data, with a MAPE value of 12.2586%. Consequently, this research offers valuable insights that enable the government to formulate more effective energy policies, assist companies in the energy sector to optimize their production and trading strategies, and support investors in making informed decisions regarding market risks and investment timing. However, this study has limitations, especially related to the very small size of the test data, which is only 10 data points. Therefore, the prediction results need to be further tested using larger data so that the reliability and generalizability of the model in various market conditions can be ensured.

AUTHOR CONTRIBUTIONS

Sediono: Supervision, Conceptualization, Validation. Toha Saifudin: Supervision, Validation. Maria Setya Dewanti: Writing - Original Draft, Writing - Review and Editing, Software. Aurelia Islami Azis: Writing - Original Draft, Writing - Review and Editing, Software. All authors discussed the results and contributed to the final manuscript.

FUNDING STATEMENT

This research has been funded by Airlangga University through the Faculty of Science and Technology.

ACKNOWLEDGMENT

The authors express deep gratitude to Harvard Dataverse and BPS RI for providing data and supporting this research.

CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest related to this research.

REFERENCES

[1] A. P. Pratama and Y. B. W. Tama, “PREDIKSI HARGA KOMODITAS GAS ALAM MENGGUNAKAN MODEL BROWNIAN MOTION DAN GEOMETRIC BROWNIAN MOTION,” *Proximal: Jurnal Penelitian Matematika dan Pendidikan Matematika*, vol. 6, no. 2, pp. 73–81, 2023, doi: https://doi.org/10.30605/proximal.v6i2.2655.

[2] C. R. Bakrie, M. O. Delanova, and Y. M. Yani, “PENGARUH PERANG RUSIA DAN UKRAINA TERHADAP PEREKONOMIAN NEGARA KAWASAN ASIA TENGGARA,” 2022. [Online]. Available: https://sridianti.com/konferensi-pbb-tentang-perdagangan-dan-pembangunan-

[3] A.-S. Corbeau and T. Mitrova, “RUSSIA’S GAS EXPORT STRATEGY: ADAPTING TO THE NEW REALITY,” 2024. [Online]. Available: https://energypolicy.

[4] Ambya, T. Gunarto, E. Hendrawaty, F. S. D. Kesumah, and F. K. Wisnu, “FUTURE NATURAL GAS PRICE FORECASTING MODEL AND ITS POLICY IMPLICATION,” *International Journal of Energy Economics and Policy*, vol. 10, no. 5, pp. 58–63, 2020, doi: https://doi.org/10.32479/ijeep.9676.

[5] E. D. Rosilawati, T. Tarno, and T. Wuryandari, “PERAMALAN INDEKS HARGA SAHAM GABUNGAN (IHSG) MENGGUNAKAN MODEL INTERVENSI FUNGSI PULSE,” *Jurnal Gaussian*, vol. 12, no. 3, pp. 382–391, Feb. 2024, doi: https://doi.org/10.14710/j.gauss.12.3.382-391.

[6] Lorensya, L. Ambarwati, and Y. Mahatma, “ANALISIS MODEL INTERVENSI MULTI INPUT FUNGSI PULSE DAN FUNGSI STEP DENGAN DETEKSI OUTLIER TERHADAP HARGA SAHAM,” *Jurnal EurekaMatika*, vol. 10, no. 1, pp. 61–72, 2022.doi: https://doi.org/10.17509/jem.v9i2.45387

[7] Muh. Qodri, U. Mukhaiyar, V. Ananda, and S. Maisaroh, “COMPARISON OF STOCK PREDICTION USING ARIMA MODEL WITH MULTIPLE INTERVENTIONS OF STEP AND PULSE FUNCTIONS,” *Jurnal Ilmiah Sains*, pp. 1–16, Mar. 2024, doi: https://doi.org/10.35799/jis.v24i1.51269.

[8] G. E. P. Box, G. M. Jennkins, G. C. Reinsel, and G. M. Ljung, *TIME SERIES ANALYSIS FORECASTING AND CONTROL*, 5th ed. Wiley, 2016.

[9] J. Kaur, K. S. Parmar, and S. Singh, “AUTOREGRESSIVE MODELS IN ENVIRONMENTAL FORECASTING TIME SERIES: A THEORETICAL AND APPLICATION REVIEW,” *Environmental Science and Pollution Research*, vol. 30, no. 8, pp. 19617–19641, Feb. 2023, doi: https://doi.org/10.1007/s11356-023-25148-9.

[10] J. D. Cryer and K.-S. Chan, *TIME SERIES ANALYSIS WITH APPLICATIONS IN R SECOND EDITION*, 2nd ed. Springer, 2008.doi: https://doi.org/10.1007/978-0-387-75959-3

[11] I. D. Syahputra and A. Karim, “PERBANDINGAN METODE ORDINARY LEAST SQUARE (OLS) DAN REGRESI ROBUST,” *Seminar Nasional Pendidikan*, 2017.

[12] D. C. Montgomery, C. L. Jennings, and M. Kulahci, *INTRODUCTION TO TIME SERIES ANALYSIS AND FORECASTING*, 2nd ed. Wiley, 2015.

[13] H. N. Tantika, N. Supriadi, and D. Anggraini, “METODE SEASONAL ARIMA UNTUK MERAMALKAN PRODUKSI KOPI DENGAN INDIKATOR CURAH HUJAN MENGGUNAKAN APLIKASI R DI KABUPATEN LAMPUNG BARAT,” *Jurnal Matematika*, vol. 17, no. 2, 2018, [Online]. Available: http://ejournal.unisba.ac.idDiterima:26/07/2018Disetujui:03/11/2018PublikasiOnline:30/11/2018

[14] J. J. Montaño Moreno, A. Palmer Pol, A. Sesé Abad, and B. Cajal Blasco, “USING THE R-MAPE INDEX AS A RESISTANT MEASURE OF FORECAST ACCURACY,” *Psicothema*, vol. 25, no. 4, pp. 500–506, 2013, doi: https://doi.org/10.7334/psicothema2013.23.

[15] M. A. Maricar, “ANALISA PERBANDINGAN NILAI AKURASI MOVING AVERAGE DAN EXPONENTIAL SMOOTHING UNTUK SISTEM PERAMALAN PENDAPATAN PADA PERUSAHAAN XYZ,” *Jurnal Sistem dan Informatika (JSI)*, vol. 13, no. 2, pp. 36–5, 2019.

[16] W. W. S. Wei, *TIME SERIES ANALYSIS: UNIVARIATE AND MULTIVARIATE METHODS*, 2nd ed. 2006.