Proceedings of the 10th International Conference on Economic Management and Green Development
DOI: 10.54254/2754-1169/2026.33125

# *Natural Gas Price Forecasting in the Energy Sector: A Time Series Approach Based on Henry Hub Data*

**Zhenqing Xi**

*Ferguson College of Agriculture, Oklahoma State University, Stillwater, USA*
*dennis.xi@okstate.edu*

***Abstract.*** The energy sector has experienced significant changes in recent years, with natural gas becoming an important transitional energy source in the shift toward renewable energy. As natural gas plays a key role in electricity generation and industrial production, accurately forecasting its price has become increasingly important. This study applies time series analysis methods to examine and predict Henry Hub natural gas prices using historical data. Specifically, naïve, drift, exponential smoothing (ETS), and ARIMA models are implemented following standard procedures, including stationarity testing, differencing, and model identification using ACF and PACF. The results show that the ARIMA model outperforms other models in terms of forecasting accuracy, indicating that natural gas prices exhibit certain predictable patterns rather than purely random behavior. These findings suggest that time series models can effectively capture the dynamics of energy prices and provide useful insights for short-term forecasting. The study contributes to a better understanding of energy market behavior and offers practical implications for investors and policymakers.

***Keywords:*** Natural Gas Price, Time series forecasting, ARIMA model, Energy market, Henry hub

## 1. Introduction

There have been many changes in the energy sector in the last few years due to the increased focus on sourcing energy in cleaner and more sustainable ways around the world. While energy sources like solar and wind, are gaining more publicity, natural gas is still very relevant as an energy source. Natural gas has many applications in the production of electricity and in industry. It also has a relatively lower carbon emission footprint as compared to other fossil fuels. These factors contribute to how the natural gas markets become the foundation of measuring the overall energy markets. Furthermore, the energy sector is impacted by the natural gas markets in terms of pricing, how energy companies invest, and the regulatory policies that are adopted. Therefore, the need to study changes in the natural gas markets is critical. Natural gas price fluctuation is less volatile and more predictable compared to equity prices. The main factors include seasonal demand and supply. These factors are the main motivation of the research.

The use of different time series models to predict energy prices has been the focus of numerous studies. As an example, the Autoregressive Integrated Moving Average (ARIMA) model is a popular

DOI: 10.54254/2754-1169/2026.33125

model to use to account for autocorrelation and is also effective for modeling non-stationary data [1]. Studies have found that prices of energy resources, like natural gas and oil, can be predicted by ARIMA techniques as these prices have identifiable trends and cycles [2]. Alternatively, exponential smoothing, like in the ETS model, is also utilized for energy price forecasting as it considers both the levels and trends of the data [3]. Some studies have even found the naive and drift forecasting models to be useful as a benchmark in forecasting [4]. Despite the introduction of more sophisticated approaches, like machine learning, classical time series forecasting is still favored for their simplicity and the ability to easily interpret the results [5].

The primary goal of this research is to study natural gas price prediction, specifically the Henry Hub Natural Gas Spot Price, using time series analysis methods. For this purpose, this study uses time series analysis methods, i.e. naive, drift, ETS and ARIMA, and employs a standard time series analysis such as stationary testing, differencing, and ACF and PACF to design and apply these models. RMSE and MAE are calculated and compared in this study for training and testing datasets. The main goal of this research study is to determine the natural gas price forecasting model and find out if natural gas prices have a predictable price pattern. The research gaps this study addresses is an empirical study to energy prices forecasting methodology with an emphasis on ARIMA model. Practically, the results of this study would help investors and policymakers to understand and mitigate uncertainties in the energy market.

## 2. Methodology

This paper uses time series forecasting techniques to compute and forecast the prices of natural gas using the Henry Hub Natural Gas Spot Price as data. The reason why time series models are frequently employed in energy economics is that they can be used to model the temporal dependence of price movements as well as dynamic trends [6].

The data is sourced through the Federal Reserve Economic Data (FRED) database, offering trusted and popular macroeconomic and energy-related time series. In this research, the Henry Hub Natural Gas Spot Price is taken as the key variable, which will be the daily spot price of natural gas. The sample time will be January 7, 2015, to December 31, 2025, which will encompass various market conditions, such as steady conditions, price surges, and volatile conditions due to supply-demand shocks. The natural gas prices have more systematic trends that are affected by the economic fundamentals compared to stock price data. The Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) are used to identify the model and assist in identifying the right lag structures. They are then estimated using four forecasting models: the naive model, the drift model, the exponential smoothing (ETS) model and the Autoregressive Integrated Moving Average (ARIMA) model. ARIMA model is a popular model in time series analysis because it integrates autoregressive (AR) terms, differencing (I) and moving average (MA) terms to model non-stationary data, and model time-dependent dependencies.

A number of preprocessing steps are carried out before model estimation. To maintain continuity of the time series, the dataset is first filled forward and backward to address the missing values. Second, the data is transformed into time series object to be further analyzed. Third, the Augmented Dickey-Fuller (ADF) test is used to test stationarity. Augmented Dickey-Fuller (ADF) test indicates that the original series is stationary at the 5% level of significance. Nevertheless, the appearance shows that there are volatility and structural changes. Thus, first-order differencing (see Figure 1) remains used to enhance the stability of the models.

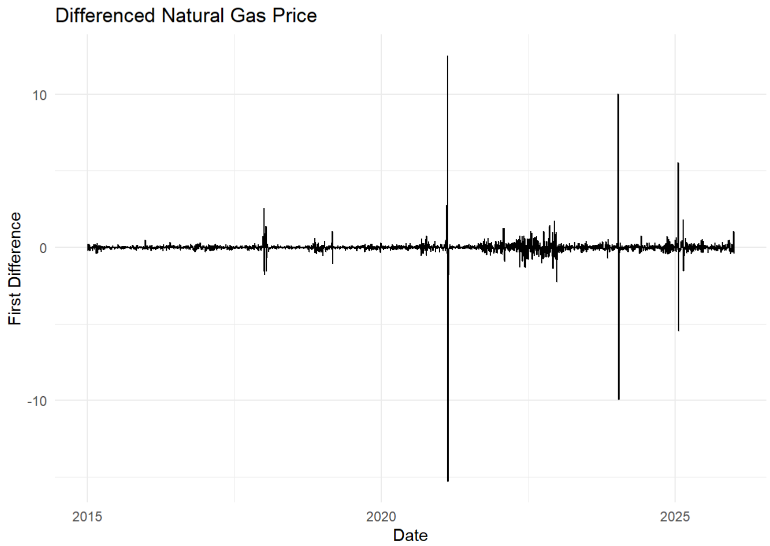


Figure 1. Differenced natural gas price

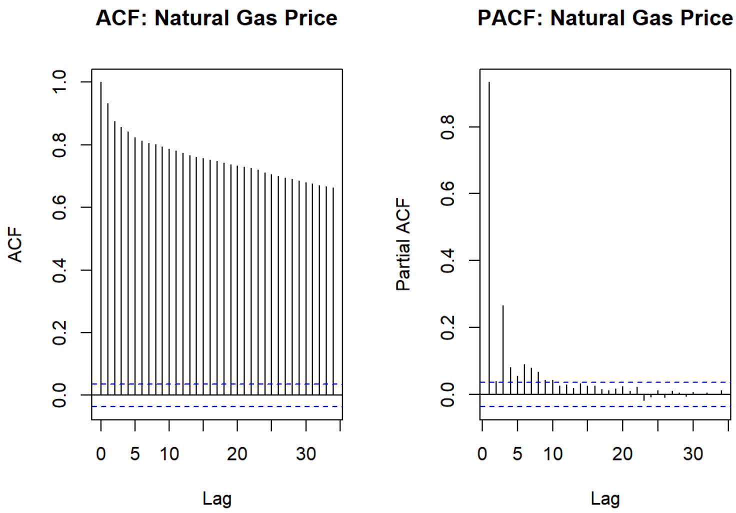


Figure 2. ACF and PACF of the natural gas price

The Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) (see Figure 2) are used to identify the model to identify the right lag structure. Four forecasting models are estimated: naïve model, drift model, exponential smoothing (ETS) model and ARIMA model. Naive model is the assumption that the future value is the last observation whereas the drift model is characterized by a linear trend through time. The ETS model breaks down the level and trend elements through the exponential smoothing method. Both autoregressive and the moving average structures in the data are modeled using the ARIMA model. Naive model is utilized as a point of reference and ARIMA is the main model since it is capable of capturing the autoregressive and moving average components [7].

To evaluate forecast performance, the dataset is divided into training and testing sets (see Figure 3). To evaluate forecasting performance, the dataset is divided into training and testing sets. The training set spans from January 7, 2015 to January 13, 2025, while the testing set covers the period

DOI: 10.54254/2754-1169/2026.33125

from January 14, 2025 to December 31, 2025. The last 252 observations are used as the test set, representing approximately one year of data.

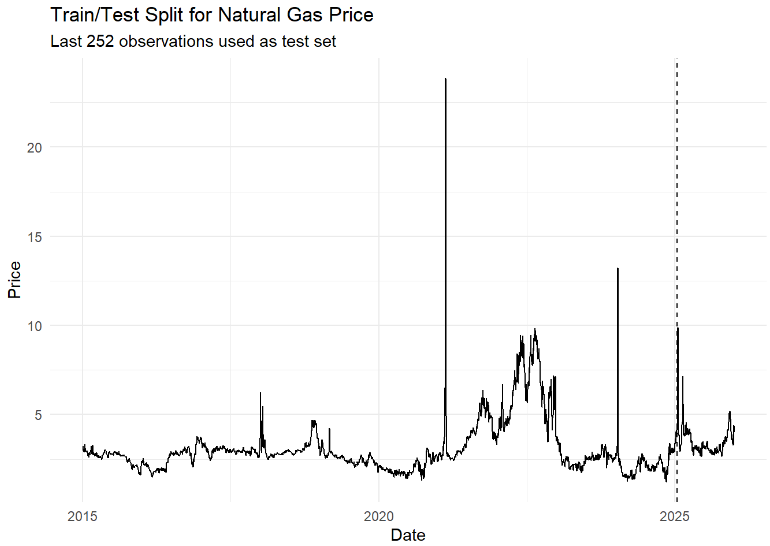


Figure 3. Train/test split

Model accuracy is assessed using Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE), and the best model is selected based on the lowest RMSE value.

## 3. Results

### 3.1. Stationarity analysis

The time series plot of Henry Hub natural gas prices (see Figure 4) shows clear fluctuations over time, with periods of both sharp increases and declines.

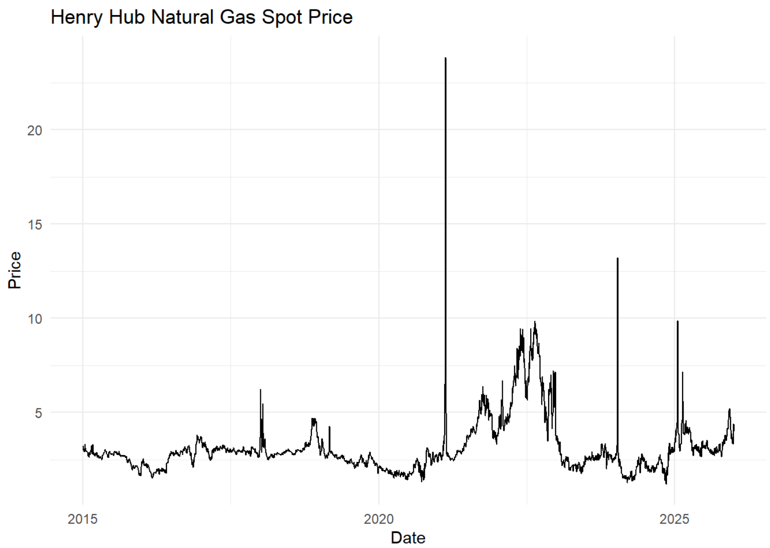


Figure 4. Henry hub natural gas spot price

DOI: 10.54254/2754-1169/2026.33125

The series can be seen to have a greater observable structure as compared to stock prices, probably due to underlying economic factors of seasonal demand, supply shocks, and energy market conditions.

Table 1 reports the results of the Augmented Dickey-Fuller (ADF) test. In the case of the original series, the p-value is 0.0205 that is less than the 5 percent level and this implies that the series is stationary. Nonetheless, the eye can concur that the series continuing to have both ups and downs and even structural alterations with time.

Table 1. ADF test results for natural gas price series

| Series | Dickey-Fuller Statistic | p-value | Conclusion |
|---|---|---|---|
| Original series | -3.7691 | 0.0205 | Stationary |
| First-differenced series | -18.848 | < 0.01 | Stationary |

The ADF test statistics are a lot more negative after making first-order differencing and the p-value is less than 0.01 which further shows that the differenced series is stationary. Thus, differencing is used to enhance the stability of the model and guarantee effective estimation.

The ACF and PACF of the differenced series (see Figure 5) indicate that there is considerable autocorrelation at low lags, indicating that past values can be useful in prediction. Following first-order differencing, the time series would be more stable. Differenced series is also oscillating around a fixed mean and ADF test confirms that the series is stationary.

The ACF and PACF plots of the differenced series indicate that there is a high degree of autocorrelation at low lags implying that the past values can be used to predict the future.

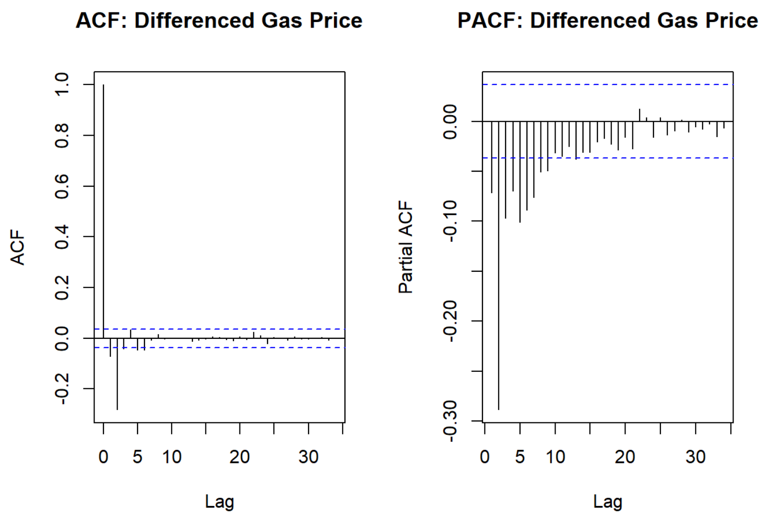


Figure 5. ACF and PACF of the differenced gas price

This indicates that ARIMA models are appropriate for capturing the underlying dynamics of the data.

## 3.2. Model estimation

The ARIMA model obtained with the automatic procedure (auto.arima) is denoted as ARIMA(p, d, q) where p is the autoregressive order, d is the differencing order and q is the moving average order

DOI: 10.54254/2754-1169/2026.33125

[8]. This model is a good fit to the training data. The estimated parameters are statistically significant meaning that the model represents the underlying structure of the series.

The model is also adequate as indicated by residual diagnostics. The residuals vary randomly around zero and no obvious trend is observed and the autocorrelation function (ACF) of the residuals do not have any noticeable spikes. Besides, the results of the Ljung-Box test show that there are no significant autocorrelations between the residuals. Therefore, the residuals can be considered approximately white noise, which is a key assumption for a well-specified ARIMA model, suggesting that the model has successfully captured the main structure of the time series [9].

Relative to stock price modeling where ARIMA generally tends to degenerate to a random walk model, the natural gas price series shows more complicated dynamics capable of being modeled successfully.

### 3.3. Forecasting performance

The forecasting performance of the four models is evaluated using RMSE and MAE. The results of Table 2 shows that the ARIMA model achieves the lowest RMSE among all candidate models, indicating that it provides the most accurate predictions.

Table 2. Forecasting performance of different models

| Model | RMSE | MAE |
|---|---|---|
| ARIMA | 0.8797 | 0.6347 |
| ETS | 1.2187 | 1.0443 |
| Naive | 1.2187 | 1.0443 |
| Drift | 1.2728 | 1.0997 |

According to Table 2, ARIMA model has the lowest RMSE (0.8797) and MAE (0.6347) when compared to all candidate models, thus giving the most accurate forecasts on the price of natural gas.

Comparatively, both the ETS and naive models exhibit the same performance with RMSE of 1.2187 and MAE of 1.0443 indicating that the models are not able to describe the underlying structure of the data. The drift model has the lowest ranking, the largest RMSE (1.2728) and MAE (1.0997), which implies that a simple lineal trend assumption cannot be used to explain the natural gas prices dynamics.

All these findings imply that the ARIMA model would better represent the time dependence and variation of the data, hence it will be the most appropriate model to predict the future prices of natural gas in this case.

### 3.4. Forecast results

The forecast plot (see Figures 6-9) shows that the ARIMA model closely follows the actual values in the test period.

DOI: 10.54254/2754-1169/2026.33125

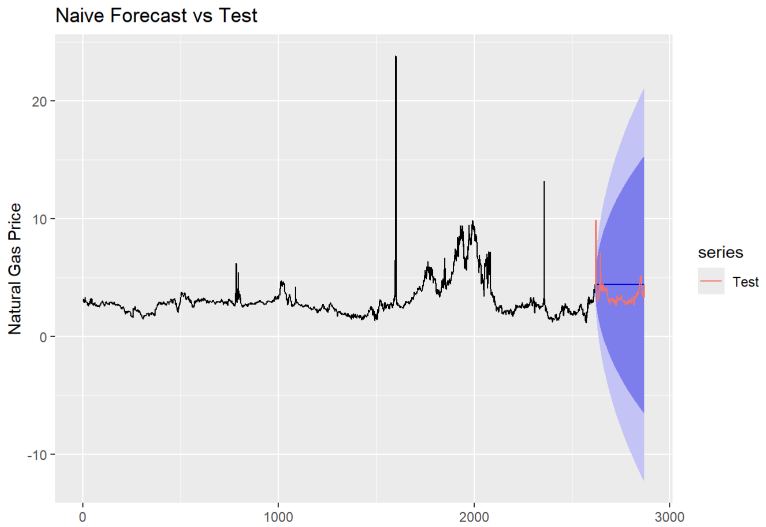


Figure 6. Naïve forecast

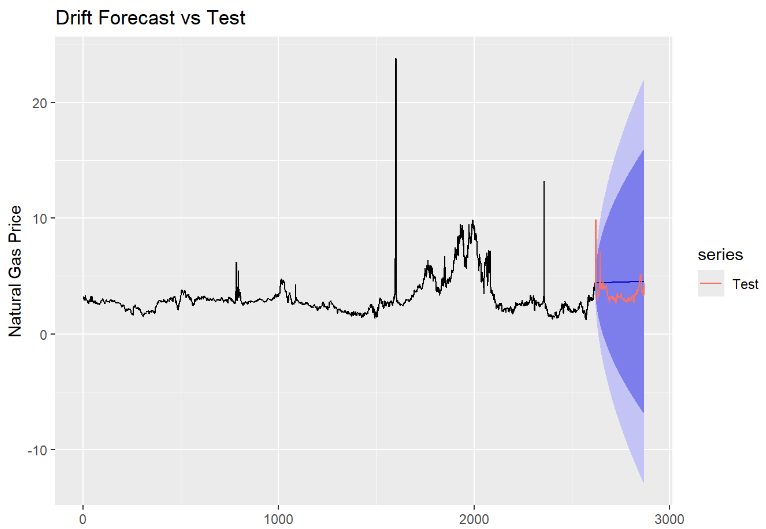


Figure 7. Drift forecast

DOI: 10.54254/2754-1169/2026.33125

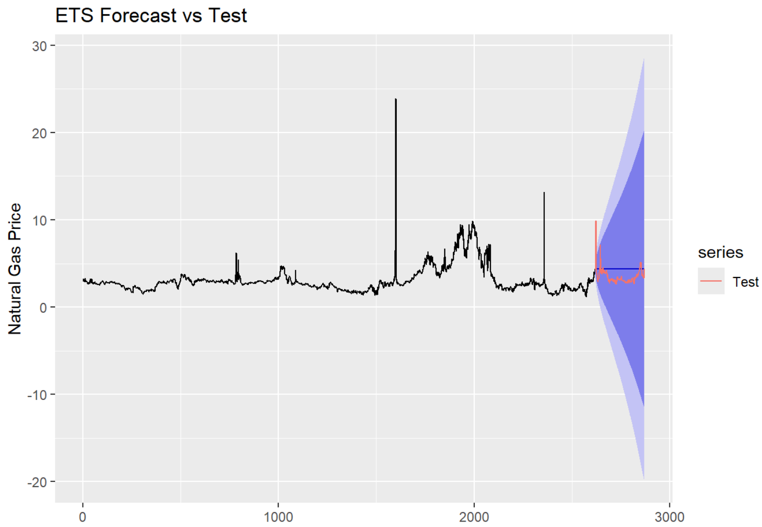


Figure 8. ETS forecast

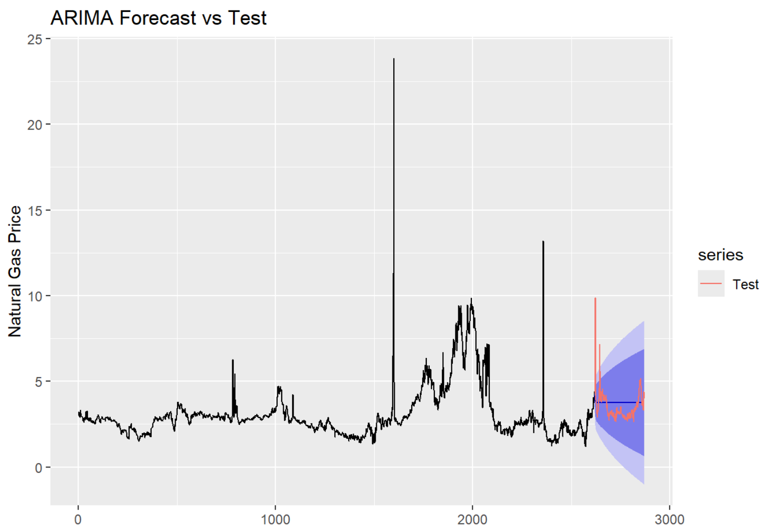


Figure 9. ARIMA forecast

Although some deviations occur due to short-term volatility, the model is able to capture both the general trend and short-term fluctuations.

As the forecast horizon increases, the prediction intervals become wider, reflecting the increasing uncertainty associated with long-term forecasting.

## 4. Discussion

The empirical findings are that natural gas prices are structured time series which can be modeled successfully with ARIMA. In contrast to the stock prices, which tend to be on a random walk and are not easy to predict, the natural gas prices are affected by economic fundamentals like supply, demand and external shocks. This renders them more appropriate to the conventional time series models.

DOI: 10.54254/2754-1169/2026.33125

The high accuracy of the ARIMA model suggests that the past value and error terms also have valuable information to predict the future prices. This observation shows how crucial it is to choose the right datasets in the implementation of time series models. In this paper, substituting stock price data with natural gas prices drastically leads to a better performance of the model.

Economically speaking, natural gas prices are influenced by a variety of factors, such as the time of the year (i.e. heating during winter), the amount of production and the market conditions of the entire energy industry. Such factors bring in trends that can be modelled using ARIMA models.

To apply the findings practically, the findings indicate that the ARIMA models can be effective energy price predictors, but they might not perform well with highly nonlinear or complex dynamics [10]. Such models can be used by investors to make short term trading decisions and by policymakers to make energy policy forecasts and supply risk management.

## 5. Conclusion

This paper examines the dynamics and prediction of natural gas prices using time series, which is based on the Henry Hub Natural Gas Spot Price. The data is analyzed using a typical time series model, which encompasses data preprocessing, stationarity test with the Augmented Dickey-Fuller (ADF) test, first-order differencing, identification of the model with the help of ACF and PACF plots, and estimation of the model with the assistance of various forecasting methods. The empirical findings indicate that, as much as the original price series seems to be relatively stable based on the ADF test, additional differencing enhances stability and reliability of the model. The ARIMA model shows the best forecasting performance by RMSE and MAE compared to the other models, including the naïve model, drift model, exponential smoothing (ETS), and ARIMA model. The diagnostics Residual diagnostics show that the model effectively describes the underlying structure of the time series, with the residuals acting more or less like white noise. On the whole, the findings indicate that natural gas prices have some predictable trends and can be modelled successfully with the help of ARIMA methods.

There are limitations to this study, however, as well. The model is univariate and it does not incorporate any external explanatory variables like the oil prices, weather conditions and geopolitical forces. These factors in fact can influence the prices of natural gas immensely. Further studies might build on this analysis by adding multivariate models or more sophisticated machine learning algorithms like LSTM or GARCH to more effectively model complicated dynamics.

## References


[1] Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). Time series analysis: Forecasting and control (5th ed.). Wiley.

[2] Hamilton, J. D. (1994). Time series analysis. Princeton University Press.

[3] Hyndman, R. J., & Athanasopoulos, G. (2021). Forecasting: Principles and practice (3rd ed.). OTexts.

[4] Makridakis, S., Wheelwright, S. C., & Hyndman, R. J. (1998). Forecasting methods and applications (3rd ed.). Wiley.

[5] Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing, 50, 159–175.

[6] U.S. Energy Information Administration (EIA). (2023). Natural gas explained. Retrieved from https: //www.eia.gov/energyexplained/natural-gas/.

[7] Federal Reserve Bank of St. Louis. (2025). Henry Hub Natural Gas Spot Price (DHHNGSP). FRED database. Retrieved from https: //fred.stlouisfed.org/series/DHHNGSP.

[8] Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for R. Journal of Statistical Software, 27(3), 1–22.

DOI: 10.54254/2754-1169/2026.33125

[9] Ljung, G. M., & Box, G. E. P. (1978). On a measure of lack of fit in time series models. Biometrika, 65(2), 297–303.
[10] Pindyck, R. S. (1999). The long-run evolution of energy prices. The Energy Journal, 20(2), 1–27.