American Journal of Multidisciplinary Research and Innovation (AJMRI)

ISSN: 2158-8155 (ONLINE), 2832-4854 (PRINT)

VOLUME 5 ISSUE 4 (2026)

PUBLISHED BY
E-PALLI PUBLISHERS, DELAWARE, USA

American Journal of Multidisciplinary Research and Innovation (AJMRI)

Volume 5 Issue 4, Year 2026
ISSN: 2158-8155 (Online), 2832-4854 (Print)
DOI: https://doi.org/10.54536/ajmri.v5i4.8261
https://journals.e-palli.com/home/index.php/ajmri

Hybrid Time-Series Approaches for Soybean Price Forecasting Using Linear and Non-Linear Models

Dharma Raj Katuwal[1*], Lila B. Karki[1], Yeon Nain Chi[1], Anuj Dhakal[1]

Article Information

Received: May 08, 2026

Accepted: July 22, 2026

Published: August 12, 2026

Keywords

Agricultural Price Forecasting, Hybrid ARIMA–NNAR Model, Machine Learning, Neural NetworkAutoregression, Time Series Forecasting

ABSTRACT

Soybean is an important agricultural commodity, essential for food security, livestock feed, and industrial uses. Therefore, understanding its price fluctuations is critical for informed decision-making. However, soybean prices show significant volatility due to fluctuations in global demand and supply, climate variability, and policy alterations, resulting in uncertainty for farmers, traders, and policymakers. This study aims to examine the time-series dynamics of soybean prices and establish an accurate forecasting framework using a monthly dataset from January 1960 to March 2026 obtained from the World Bank. The study uses various forecasting methodologies, comprising linear models (ARIMA, ETS, and Theta), a nonlinear model (NNAR), and hybrid models that combine linear and nonlinear frameworks, with model performance determined through accuracy metrics including Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). The findings reveal that soybean prices show considerable variability, averaging $292.59 per metric ton, ranging from $88.00 to $737.06, indicating high market volatility. Furthermore, while the ARIMA (1,1,3) model effectively identifies linear trends, the NNAR model shows superior performance among individual models by yielding reduced forecasting errors (RMSE = 162.86; MAPE = 36.22%). The hybrid ARIMA–NNAR model attains superior accuracy, significantly reducing forecast errors (RMSE = 98.03; MAPE = 21.66%). This highlights the benefits of integrating linear and nonlinear methods. The forecast indicates that soybean prices are expected to rise in the short-term, peak in 2027–2028, and then decline gradually toward 2030. However, the expanding prediction intervals show growing uncertainty over time. The findings confirm that hybrid models provide a more robust and reliable framework for forecasting soybean prices. Future studies should integrate exogenous variables such as climatic conditions, trade policies, and exchange rates, and should investigate new methodologies, including deep learning approaches, to improve forecasting precision and policy relevance.

INTRODUCTION

Soybean is an important crop in the global agricultural system due to their because of its diverse uses and strong linkages with food, energy, and industrial sectors (Siamabele, 2021; Pagano & Miransari, 2016). It serves as a principal protein source in cattle feed and is a significant component in the manufacturing of edible oils, making it important for both the food and feed sectors (Flachowsky *et al.*, 2017; Gunstone, 2011; Wilkinson & Lee, 2018). Moreover, soybean is increasingly utilized in biofuel production and diverse industrial applications, which are expanding their economic significance (Goldsmith, 2008). Soybean, as a globally traded commodity, significantly impacts global markets, with its price determined by production trends in leading exporting nations, global demand fluctuations, and trade regulations (Gottimukkala & Bhuram, 2025; Peng *et al.*, 2026). Thus, variations in soybean prices can significantly influence agricultural income, food prices, livestock production costs, and overall market stability.

However, soybean prices are highly volatile and are influenced by climatic conditions, global supply and demand, trade regulations, exchange rates, inventories, and unexpected market disruptions (Peri, 2017; Katuwal *et al.*, 2026). Adverse climatic occurrences may reduce crop production, while fluctuations in global demand can rapidly change price patterns (Haile *et al.*, 2017; Zhu *et al.*, 2024). Thus, these uncertainties prevent producers, traders, and policymakers from making rational decisions. Accurate price forecasting can help address this uncertainty by supporting production planning, marketing decisions, risk management, and policy responses. Furthermore, precise forecasts may enhance decision-making in fields such as inventory management, trade strategies, and price stabilization measures. Consequently, the application of complex forecasting methodologies is becoming vital in agricultural economics. According to the World Bank Commodity Markets Outlook, overall commodity prices are projected to increase by 16% in 2026, with energy prices expected to increase by 24% (World Bank, 2026). Such fluctuations significantly influence agricultural commodity markets, including soybeans, as input costs (e.g., fertilizers and energy) directly affect production decisions and market prices. Therefore, accurate forecasting of soybean prices is essential for informed policy decisions, risk management, and market planning.

In recent years, time series models have been extensively utilized to predict agricultural commodity prices.

[1] Department of Agriculture, University of Maryland Eastern Shore, Princess Anne, Maryland, USA
[*] Corresponding author's e-mail: drkatuwal@umes.edu

Conventional linear models, such as the autoregressive integrated moving average (ARIMA), effectively identify linear trends in data (Arumugam & Natarajan, 2023; Nelson, 1998). However, agricultural price series frequently have nonlinear characteristics due to complex market dynamics and external disturbances (Tripathi *et al.*, 2023). Thus, nonlinear models, including neural network autoregressive (NNAR) models, have drawn interest in their capacity to identify such patterns (Almarashi *et al.,* 2024; Khadka *et al.*, 2025; Limbu Sanwa *et al.*, 2025). Moreover, integrating linear and nonlinear methodologies through hybrid models has emerged as a viable strategy to enhance forecasting accuracy (Sina *et al.*, 2023; Duan *et al.*, 2022).

Despite these advances, there remains a necessity for more robust and dependable forecasting methodologies that can accurately capture the dynamic characteristics of soybean prices. Specifically, little attention has been paid to the comparative efficacy of various models and the prospective advantages of hybrid modeling in this area. This study examines the time-series dynamics of soybean prices and assesses the efficacy of diverse forecasting models, including linear, nonlinear, and hybrid methods. This study offers significant insights for several stakeholders. The findings can aid farmers in production planning, assist sellers in mitigating pricing risks, and enable policymakers to take informed, timely actions. Therefore, improving the accuracy of soybean price forecasting is not solely a methodological issue but also an essential development for strengthening decision-making within the agricultural sector.

LITERATURE REVIEW

Classical Time-Series Models for Soybean Price Forecasting

Statistical time-series models are widely utilized in agricultural price forecasting because they effectively capture temporal interdependence and organized patterns in data (Xia, 2024). The ARIMA model is one of the most widely used methods for predicting commodity prices because it effectively captures autoregressive and moving-average processes (Arumugam & Natarajan, 2023). In agricultural markets, ARIMA has served as a benchmark model for forecasting price fluctuations in commodities such as soybeans and cereals (Goodwin & Schnepf, 2000).

The Seasonal ARIMA (SARIMA) model integrates seasonality into monthly data, enabling identification of periodic variations linked to production cycles (Wei, 2006). Research indicates SARIMA is effective when price patterns follow consistent seasonal frameworks (Divisekara *et al.*, 2021). Exponential Smoothing (ETS) models are also utilized due to their adaptability in capturing trend and seasonal elements (Hyndman *et al.*, 2008). These methods assign greater weight to new observations while diminishing the influence of earlier values, thereby adapting to evolving data trends. Simple exponential smoothing is suitable for steady series, whereas Holt-Winters extensions are preferable for series with trend and seasonality (Ostertagová & Ostertag, 2012).

In the Box–Jenkins system, ARIMA modeling serves as both a forecasting equation and a systematic process including identification, estimation, diagnostic evaluation, and forecasting (Young, 1977). Applications typically involve transformation or differencing to achieve stationarity, followed by ACF and PACF analysis to detect lag structures (Young, 1977; Abdallah, 2019). This is important for monthly commodity prices, where seasonal dependency may arise at annual lags, making SARIMA a logical extension (Boniface & Martin, 2019). Stellwagen and Tashman (2013) argued that ARIMA should be regarded as a benchmark rather than universally superior. Decomposition methods such as STL isolate trend and seasonal components, enhancing interpretability (Cleveland *et al.*, 1990). TBATS models manage complex seasonal patterns in large datasets (De Livera *et al.*, 2012) and have shown efficacy in forecasting commodity prices with unpredictable seasonality (Athanasopoulos *et al.*, 2011). Despite widespread use, statistical models may inadequately represent nonlinear patterns and sudden price fluctuations (Granger, 2001).

Exponential Smoothing and Seasonal Adjustment Approaches

Exponential smoothing techniques provide a flexible framework for modeling time-series data by assigning greater weight to recent observations, allowing the model to adapt to evolving market conditions. Unlike ARIMA-based approaches, which rely on autocorrelation structures, smoothing methods emphasize level, trend, and seasonal components in a more data-driven manner (Ostertagová & Ostertag, 2012; Hyndman *et al.*, 2002). These approaches are widely applied in agricultural price forecasting due to their simplicity and strong empirical performance, particularly for short-term predictions.

In addition to smoothing techniques, seasonal adjustment methods are used to isolate underlying trends by removing periodic fluctuations from time-series data. Methods such as moving average-based seasonal indices and regression-based seasonal decomposition help distinguish between systematic seasonal effects and irregular variations (Jadhav *et al.*, 2018; Ostertagová & Ostertag, 2015). These approaches are particularly useful in agricultural markets, where seasonal production cycles may influence price behavior. However, in globally traded commodities such as soybeans, seasonality is often weaker and less stable, reducing the effectiveness of purely seasonal models.

Neural Network and Hybrid Forecasting Models

Hybrid models combine statistical and machine learning approaches to improve forecasting accuracy (Zhang, 2003). The ARIMA-ANN hybrid model uses ARIMA for linear components and neural networks for nonlinear residuals (Zhang, 2003).

Studies show hybrid models outperform individual models (Kumar & Thenmozhi, 2014). ARIMA-ANN models improve forecasting accuracy in financial and commodity markets (Babu & Reddy, 2014), and hybrid models enhance precision by capturing both linear and nonlinear patterns (Khandelwal *et al.*, 2015). Hybrid approaches are widely used in agricultural markets (Jaiswal *et al.*, 2023), especially decomposition-based models such as STL combined with machine learning (Panigrahi & Behera, 2017)

Model Selection, Diagnostics, and Forecast Evaluation

Model selection is essential, as models should not be chosen solely on theoretical grounds. The Box–Jenkins framework includes identification, estimation, diagnostic verification, and forecasting (Young, 1977). ACF and PACF are used to determine model structure (Young, 1977; Abdallah, 2019). Models are evaluated using AIC and BIC (Akaike, 1974; Schwarz, 1978; Abdallah, 2019), with preferred lower values.

Diagnostic verification ensures residuals approximate white noise. The Ljung-Box test assesses residual autocorrelation (Ljung & Box, 1978). Persistent autocorrelation indicates model inadequacy. Diagnostic checks should precede model evaluation (Jadhav *et al.*, 2017; Abdallah, 2019). Forecast accuracy is evaluated using out-of-sample performance. Metrics include MAPE, MSE, MAE, and RMSE (Jadhav *et al.*, 2017; Ostertagová & Ostertag, 2012; Abdallah, 2019). Model comparison should use consistent horizons and evaluation criteria.

Research Gap and Contribution of the Study

While prior research has utilized ARIMA, SARIMA, exponential smoothing, neural networks, and hybrid methodologies for commodities and economic time series, there is a lack of studies that integrate various approaches into a single model for an extensive historical monthly soybean price series. Jadhav *et al.* (2017) concentrated on ARIMA-based agricultural price forecasting, Ostertagová and Ostertag (2012) highlighted smoothing techniques, and Abdallah (2019) elaborated on the stages for constructing ARIMA models in a practical forecasting context. The majority of the current work thus focuses on a singular model family, a brief sample period, or a restricted set of evaluation criteria. This gap is particularly significant for soybean prices, as a monthly series from 1960 to 2026 is expected to capture long-term trend alterations, cyclical seasonal fluctuations, policy transformations, and substantial global disruptions. This study compares classical, smoothing-based, and nonlinear forecasting methods for an extensive soybean price series, explicitly combining structural break analysis, model diagnostics, and formal forecast validation.

MATERIALS AND METHODS

Data Description

The data used in this study were obtained from the World Bank Commodity Price Data, commonly known as the Pink Sheet (April 2026 edition) of the Commodity Markets Outlook. The dataset consists of monthly soybean prices from January 1960 to March 2026. The monthly soybean price series was converted into a time series with a 12-month frequency to reflect its monthly structure and maintain consistency with standard forecasting practices.

To examine the underlying structure of the series, it was decomposed into its fundamental components using an

$$Y_t=T_t+S_t+R_t \quad (1)$$

additive model as given by Dozie & Ijomah (2020):

Where,

Yt denotes the observed soybean price at time t,

Tt represents the trend component,

St captures seasonal variations, and

Rt corresponds to the irregular (random) component.

Forecasting Models

Three categories of forecasting models were utilized to summarize the various properties of the soybean price series:

Linear Models

Linear time series models have been used to identify trends and seasonal patterns within the data (Verbesselt *et al.*, 2010; Bandara *et al.*, 2025). The models encompass Autoregressive Integrated Moving Average (ARIMA), Seasonal ARIMA (SARIMA), Exponential Smoothing (ETS), and Theta models. These models are especially efficient at representing both linear dependencies and structured temporal patterns (Dubey *et al.*, 2021).

ARIMA Model

The ARIMA model is a popular linear time-series model that captures temporal dependencies and linear correlations in the data. It functions through three primary components: autoregression, differencing, and moving average (Bernal *et al.*, 2021; Lai & Dzombak, 2020; Osho, 2019). The ARIMA model uses historical data (autoregressive components), eliminates non-stationarity by differencing, and characterizes the residual errors with a moving average process (Momin & Chavan, 2018).

The standard representation of the ARIMA (p,d,q) model

$$\phi_p(B)(1-B)^d Y_t=C+\theta_q(B)\varepsilon_t \quad (2)$$

is expressed as follows, given by (Shumway & Stoffer, 2017):

Where,

B is the backshift operator,

Yt represents the observed value at time t,

C is a constant term, and

Ɛt denotes white noise.

Moreover, p represents the number of lagged observations included in the model (autoregressive order), d indicates the number of differencing steps required to get stationarity, and q denotes the number of lagged forecast errors (moving average order).

SARIMA Model

The Seasonal Autoregressive Integrated Moving Average (SARIMA) model extends the ARIMA framework by incorporating seasonal components, thereby capturing both short-term dynamics and periodic patterns in time-series data (Wei, 2006; Paul, 2014). The general SARIMA

$$\Phi_P(B^s)\,\phi_p(B)\,(1-B)^d\,(1-B^s)^D Y_t=\Theta_Q(B^s)\,\theta_q(B)\,\varepsilon_t \quad (3)$$

(p, d, q) (P, D, Q)s model is expressed as:

Where,

p, d, q denote the non-seasonal autoregressive, differencing, and moving average orders, and

P, D, Q represent the corresponding seasonal components with seasonal period s.

The operator B is the backshift operator, and ε_tis a white-noise error term.

Although SARIMA is theoretically appropriate for monthly data, empirical results from decomposition, autocorrelation analysis, and differencing tests indicate weak seasonal persistence in the soybean price series. Therefore, seasonal differencing was not required, and a non-seasonal ARIMA model was selected as the baseline specification for forecasting.

ETS Model

The Exponential Smoothing (ETS) model assigns greater weight to recent observations, making it effective for capturing evolving patterns in time series data (De Livera *et al.*, 2012).

Let Yt denote the observed series and “x” _”t” “= (“ “l” _”t” “,” “b” _”t” “,” “s” _”t” “,…) “ represent the state vector, where lt, bt, and st correspond to level, trend, and seasonality, respectively. The state space form given by

$$Y_t=Hx_{t-1}+\varepsilon_t, x_t=Fx_{t-1}+G\varepsilon_t \quad (6)$$

Hyndman *et al.* (2002) and Safi & Sanusi (2021) is:

Where,

ε_t~”NID”(0,σ^2)

TBATS Model

The TBATS model is used to identify complex and irregular seasonal patterns in time series data. It includes a Box–Cox transformation for variance stabilization, trigonometric components for seasonality, and ARMA errors (Andreozzi *et al.*, 2014, and Hutagalung & Sari, 2024).

$$yt(\lambda)=lt+st+\varepsilon t \quad (7)$$

The standard format is:

TBATS is particularly useful when multiple seasonal cycles or irregular patterns are present in the data.

Theta Model (Decomposition-Based)

$$Y_t=\theta_1 Y_t+\theta_2 Trend_t \quad (8)$$

The Theta model separates the series into components: It simplifies forecasting by separating trend and fluctuations, but it may be less effective for complex or highly nonlinear data.

Nonlinear Model

NNAR Model

The Neural Network Autoregression (NNAR) model was used to address potential nonlinear correlations and complex patterns found in the data (Daniyal *et al.*, 2022; Crone *et al.*, 2011). The NNAR model can learn nonlinear patterns that typical linear models may inadequately represent (Elseidi, 2024; Almarashi, 2025; Almarashi *et al.*, 2024).

The NNAR model uses lagged values of the time series as inputs to a feedforward neural network. The fundamental structure of the NNAR model given by Adenomon & Emmanuel (2024) and Zhang (2003) is:

$$Y_t=f(Y_{t-1},Y_{t-2},\ldots,Y_{t-p})+\varepsilon_t \quad (4)$$

For seasonal time series data, such as monthly soybean prices, the model can be extended to include seasonal lags:

$$Y_t=f(Y_{t-1},\ldots,Y_{t-p},Y_{t-12},\ldots)+\varepsilon_t \quad (5)$$

Where,

f(.) represents a nonlinear function approximated by a neural network, and Ɛt denotes the random error term.

3.2.3 Decomposition-Based Model

STL Model

The STL model decomposes a time series into three components and is given by (Spiliotis *et al.*, 2020):

$$Y_t=T_t+S_t+R_t \quad (9)$$

Where,

Yt is the observed series, Tt is the trend component, St is the seasonal component, and Rt is the remainder (irregular component).

After decomposition, forecasting is typically performed as:

$$\hat{Y}_{t+h}=\hat{T}_{t+h}+\hat{S}_{t+h} \quad (10)$$

Where,

- T ̂_(t+h)= forecasted trend component
- S ̂_(t+h)= forecasted seasonal component

Hybrid Models

Hybrid models were developed by integrating linear and nonlinear methodologies to improve forecasting accuracy, as they combine the strengths of distinct approaches to effectively capture both linear and nonlinear patterns present in time series data (Khashei & Bijari, 2011; Mamun *et al.*, 2020; Hajirahimi & Khashei, 2023; Khashei & Bijari, 2012; Chi, 2021).

General hybrid equation:

$$\widehat{Y}_t=\sum_{i=1}^{n} w_i\, \widehat{Y}_{i,t} \qquad (11)$$

Where,

Y ̂_(i,t)represents the forecast from the i^thmodel and w_idenotes the corresponding weight assigned to each model, with ∑w_i=1.

By integrating linear models (e.g., ARIMA, ETS) and nonlinear models (e.g., NNAR), hybrid models can capture a wider range of patterns and typically provide more robust and accurate forecasts.

Two weighting strategies were used to combine individual model predictions for the construction of hybrid forecasts.

Equal Weighted Hybrid Model

In the equal weighting approach, each model contributes uniformly to the final prediction. The weight allocated to each model is specified by:

$$w_i=\frac{1}{n} \qquad (12)$$

Where,

n represents the total number of models included in the hybrid. This approach assumes that all models perform equally well and provides a simple and unbiased combination.

Cross-Validation Weighted Hybrid Model

In the cross-validation weighting approach, weights are assigned based on the inverse of the forecasting errors derived from validation data. The weights are determined as follows:

The weights are calculated as:

$$w_i=\frac{1/\mathrm{Error}_i}{\sum_{i=1}^{n}(1/\mathrm{Error}_i)} \qquad (13)$$

Where,

Errori represents the forecasting error of the ith model. In this approach, models with lower errors receive higher weights, thereby improving overall forecasting accuracy.

Model Evaluation Metrics

The efficacy of the forecasting models was assessed utilizing a collection of generally recognized statistical accuracy metrics. These measures evaluate various dimensions of forecast errors, including bias, size, and relative deviation (Shcherbakov *et al.*, 2013; Dai *et al.*, 2025; Jaiswal *et al.*, 2023).

The Mean Error (ME) was used to measure the average bias in the forecasts and is defined as:

$$\mathrm{ME}=\frac{1}{n}\sum_{t=1}^{n}(Y_t-\widehat{Y}_t) \qquad (14)$$

The Mean Absolute Error (MAE) captures the average magnitude of forecast errors, regardless of direction (Wang & Lu, 2018; Willmott & Matsuura, 2005):

$$\mathrm{MAE}=\frac{1}{n}\sum_{t=1}^{n}|Y_t-\widehat{Y}_t| \qquad (15)$$

The Root Mean Square Error (RMSE) provides a measure of error magnitude that penalizes larger deviations more heavily (Wang & Lu, 2018; Willmott & Matsuura, 2005):

The Mean Absolute Percentage Error (MAPE) expresses

$$\mathrm{RMSE}=\sqrt{\frac{1}{n}\sum_{t=1}^{n}(Y_t-\widehat{Y}_t)^2} \qquad (16)$$

forecast accuracy as a percentage, allowing for relative comparison across models (Zhang, 2003):

Finally, the Mean Percentage Error (MPE) indicates the

$$\mathrm{MAPE}=\frac{100}{n}\sum_{t=1}^{n}\left|\frac{Y_t-\widehat{Y}_t}{Y_t}\right| \qquad (17)$$

average percentage bias in the forecasts (Kim and Kim, 2016):

These evaluation metrics provide a comprehensive

$$\mathrm{MPE}=\frac{100}{n}\sum_{t=1}^{n}\left(\frac{Y_t-\widehat{Y}_t}{Y_t}\right) \qquad (18)$$

assessment of model performance and a comparison of forecasting accuracy across different models.

Data Analysis

Data analysis was conducted in R version 4.5.1 with important packages including forecast, tseries, zoo, ggplot2, and forecastHybrid. The monthly soybean price data were first cleaned, arranged chronologically, and converted to a time-series format. Descriptive statistics, time-series plots, lag plots, decomposition, ADF and KPSS stationarity tests, differencing, and ACF/PACF diagnostics were used to examine the structure of the series. A fixed random seed using set.seed() was applied to ensure reproducibility, especially for neural network-based forecasting. Both individual models and hybrid models were estimated and compared using ME, RMSE, MAE, MPE, and MAPE. Finally, the best-performing model was used to forecast soybean prices through December 2030 with prediction intervals.

RESULTS AND DISCUSSIONS

Descriptive Statistics

The descriptive statistics reveal significant variation in monthly soybean prices during the study period. Analysis of 795 monthly records from January 1960 to March 2026 shows an average soybean price of $292.59 per metric ton, contrasted with a median price of $261.00 per metric ton, indicating that several elevated price periods have skewed the mean upward. Price levels varied from a minimum of $88.00 per metric ton to a maximum of $737.06 per metric ton. The standard deviation was 142.80, indicating significant volatility in soybean prices over time. The interquartile range ranged from $205.50 at the 25th percentile to $377.88 at the 75th percentile, indicating that the central 50% of recorded values spanned a wide range. The results indicate that soybean prices showed significant variability and marked increases over the sample period, highlighting the necessity for time-series methodologies capable of capturing evolving dynamics and price volatility.

Table 1: Descriptive statistics of monthly soybean prices ($/metric ton)

Statistic	Value
Number of observations	795
Minimum	88
Maximum	737.06
Mean	292.59
Median	261
Standard deviation	142.80
25th percentile (Q1)	205.5
75th percentile (Q3)	377.88

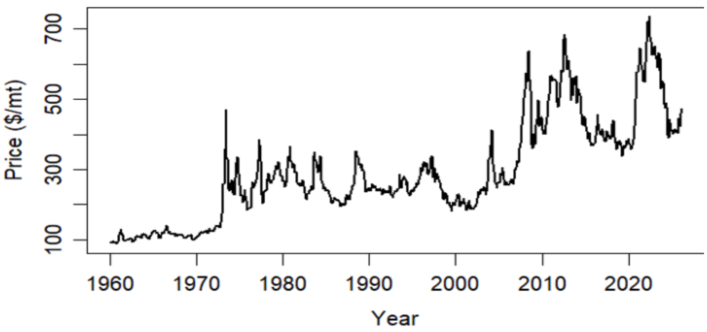

Figure 1: Monthly soybean price series

Time-series Behavior

The monthly soybean price series during the study period is illustrated in Figure 1. Generally, the series does not exhibit a uniform degree of fluctuation. Rather, it exhibits significant fluctuations over time, including periods of relative stability and periods of high price volatility. Prices were relatively stable and low in the prior decades. However, the quantity and variability of prices increased significantly from the mid-2000s onward. Specifically, the late 2000s, early 2010s, and early 2020s are characterized by numerous abrupt price increases. Consequently, the figure implies that the series possesses both time-varying volatility and a trend. Additionally, the pattern does not appear to be entirely random, which justifies further stationarity testing and decomposition in the subsequent sections.

The lag diagrams of monthly soybean prices up to lag 6 are depicted in Figure 1b. The soybean price series shows a strong serial dependence and persistence, as evidenced by the distinct positive association between current prices and their lagged values, particularly at shorter lags, as demonstrated by the plots

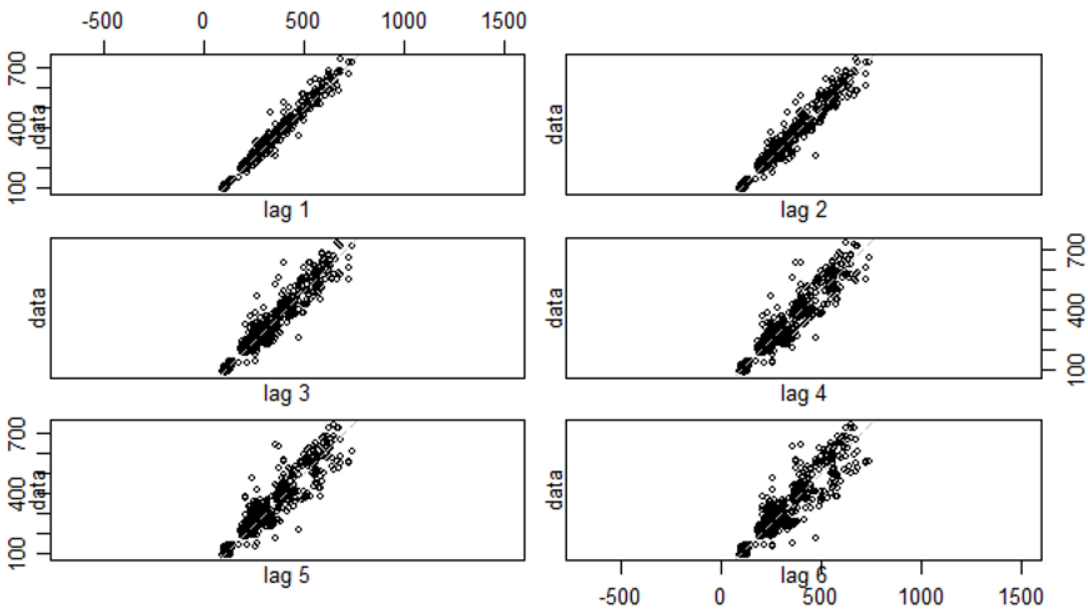

Figure 1b: Lag plots of monthly soybean prices (lags 1–6)

Decomposition and Seasonal Adjustment

An additive decomposition was implemented to better understand the internal pattern of the soybean price series. This approach divides the observed series into trend, seasonal, and random components. The series shows a stable seasonal pattern, irregular short-run fluctuations, and a varying long-term trend, as illustrated in Figure 2. The trend component, in particular, suggests that soybean prices have increased over time, despite the fact that the movement was not consistent throughout the sample period. At the same time, the seasonal component implies that monthly price fluctuations show a consistent pattern. Similarly, the random component captures short-term disruptions that are not accounted for by seasonality or trend. Thus, the decomposition further supports the existence of both systematic and irregular movements in the soybean price series.

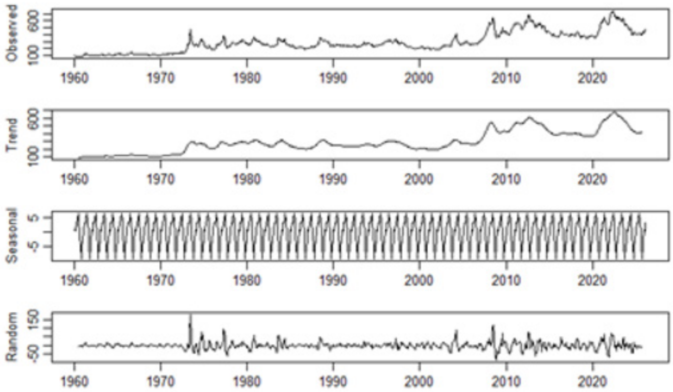

Figure 2: Additive decomposition of monthly soybean prices: observed, trend, seasonal, and random components

The soybean price series, which has been seasonally adjusted, is illustrated in Figure 3. The fundamental price movement becomes clearer after the regular seasonal effect is removed. The adjusted series continues to exhibit significant fluctuations and multiple sudden price increases, particularly in later years. Therefore, seasonality is simply a component of the overall movement; trend and irregular shocks also play significant roles.

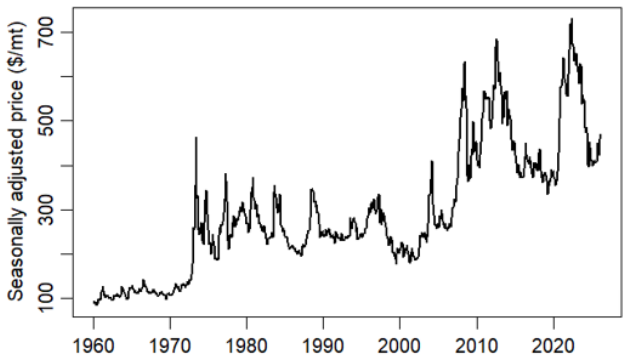

Figure 3: Seasonally adjusted monthly soybean prices

Stationarity and Differencing

The Kwiatkowski-Phillips-Schmidt-Shin test (KPSS) and the Augmented Dickey–Fuller (ADF) test have been used to provide an additional evaluation of the statistical properties of the seasonally adjusted series. The unit root null hypothesis was rejected at the 5% level, as indicated by the ADF test statistic of -3.5222 and a p-value of 0.04017. However, the null hypothesis of level stationarity was rejected, as the KPSS test statistic was 7.7534 with a p-value of 0.01. These findings provide mixed evidence and indicate that the seasonally adjusted soybean price series continues to fail to achieve stationarity. Therefore, differences were conducted in the subsequent stage prior to the identification and forecasting of the model.

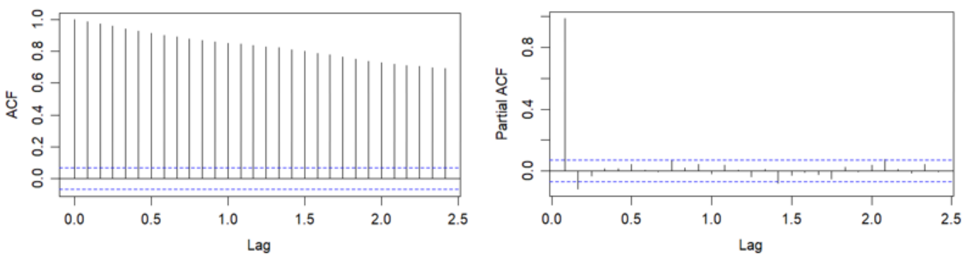

Figure 4: ACF and PACF plots of the original monthly soybean price series

The ACF and PACF plots of the original soybean price series are illustrated in Figure 4. Initially, the ACF shows a gradual decline and remains significant across multiple lags. This pattern suggests the series exhibits a high degree of persistence. In contrast, the PACF exhibits a significant rise at the initial lag, which is accompanied by significantly reduced values at subsequent delays. Overall, these findings suggest that the level series is non-stationary. Therefore, differencing was implemented before the identification and forecasting of the model.

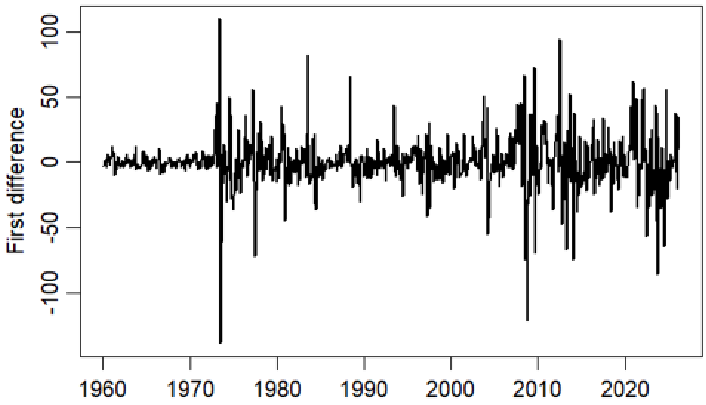

Figure 5: First-differenced soybean price series

Further, the first differenced series was analyzed. The differenced series shows a more consistent fluctuation around zero, as illustrated in Figure 5, despite a few substantial shocks. Furthermore, the ADF test for the first-differenced series provided a statistic of -10.436 with a p-value of 0.01, indicating stationarity after the initial differencing.

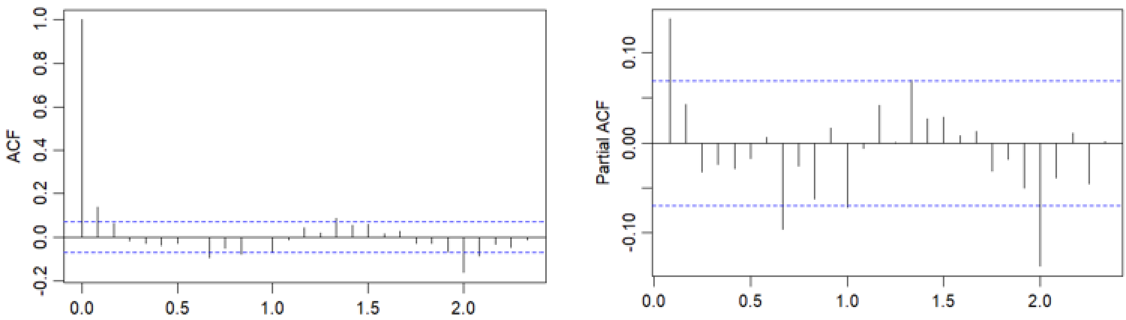

Figure 6: ACF and PACF plots of the first-differenced soybean price series

The first-differenced series' ACF and PACF diagrams are illustrated in Figure 6. The autocorrelations show a significantly faster decline following differencing than the level series. This implies that the data's persistence was significantly reduced by the initial differencing process. As a result, the ADF test result agrees with the graphical evidence.

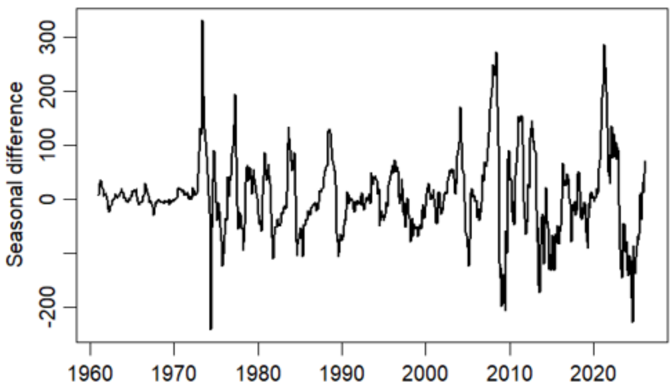

Figure 7: Seasonally differenced soybean price series

Seasonal differences were also taken into account due to the periodic nature of the data. The seasonally differenced series appears to be more stable than the original level series, as illustrated in Figure 7, despite noticeable fluctuations. Moreover, the ADF test for the seasonally differenced series yielded a statistic of -7.7056 with a p-value of 0.01, which further supports stationarity following differencing.

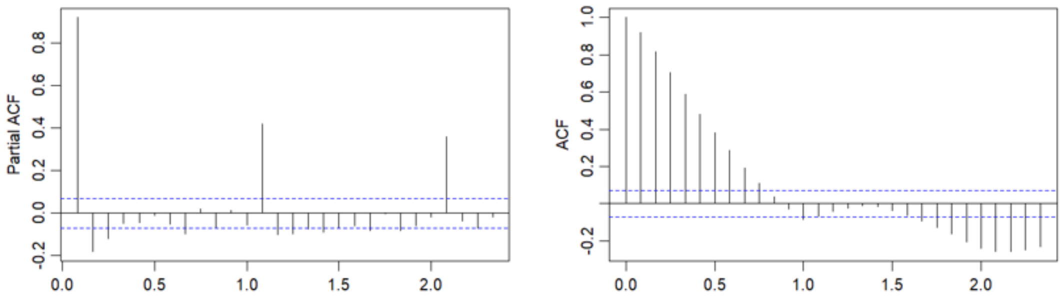

Figure 8: ACF and PACF plots of the seasonally differenced soybean price series

The ACF and PACF diagrams of the seasonally differenced series are shown in Figure 8. The ACF experiences a more rapid decline than the level series, while the PACF shows a few significant increases at specific lags.

The results of the Augmented Dickey–Fuller Test clearly confirm that the series is stationary after differencing. Specifically, the ADF statistics for the first-differenced series (-10.436, $p < 0.01$) and the seasonally differenced series (-7.7056, $p < 0.01$) indicate strong rejection of the unit root hypothesis. Therefore, the soybean price series becomes stationary after differencing and is suitable for subsequent model estimation.

Model Estimation and Diagnostics

The soybean price series was estimated using both manual and automatic ARIMA models. In addition to the non-seasonal ARIMA specification, a seasonal ARIMA model was also considered to capture potential monthly seasonal patterns in soybean prices. Specifically, a SARIMA (0,1,2)(2,0,0) [12] model was estimated. The residual diagnostics of this model indicate that the residuals behave as white noise, with no significant autocorrelation, suggesting an adequate fit. However, the differencing tests indicated that seasonal differencing was not required (nsdiffs = 0), implying that strong seasonal persistence is limited in the data. Therefore, while the SARIMA model provides a reasonable representation, the more parsimonious ARIMA (1,1,3) model was retained as the baseline linear specification for subsequent forecasting comparisons. First, a manually specified ARIMA (1,1,1) model was estimated as a benchmark. Next, the automatic selection procedure identified an ARIMA (1,1,3) model for the original monthly series. Although ARIMA (1,1,1) reported a lower AIC in this comparison, the auto ARIMA (1,1,3) specification was retained for forecasting because it was selected by the automatic model-selection procedure and produced acceptable residual diagnostics. Table 2 reports the estimated coefficients and model fit statistics for both specifications. The results show that the automatic procedure selected a model with a richer moving-average structure than the manual benchmark.

The fitted values from the auto ARIMA (1,1,3) model are compared to the actual soybean price series in Figure 9. The fitted values closely correspond to the observed series for the majority of the sample period. Specifically, the model effectively depicts the general upward and downward trends in soybean prices. However, a few abrupt spikes are not perfectly matched, a phenomenon that is common in commodity price series due to the difficulty of accurately predicting sudden shocks.

Table 2: Estimated ARIMA model coefficients and fit statistics

Model	**Coefficients**	**Sigma²**	**Log-likelihood**	**AIC**
ARIMA (1,1,1)	AR1 = 0.3235, MA1 = -0.1879	379.60	-3484.48	6974.95
ARIMA (1,1,3)	AR1 = 0.9439, MA1 = -0.8190, MA2 = -0.0586, MA3 = -0.0962	386.30	-3489.55	6989.09

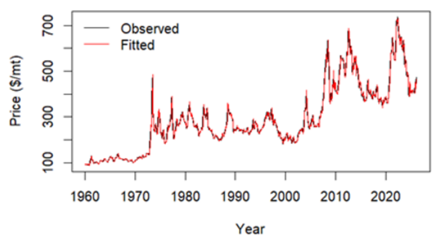

Figure 9: Fitted values from auto ARIMA over the actual series

The residual diagnostics of the auto ARIMA (1,1,3) model are illustrated in Figure 10. The standardized residuals show a lack of a distinct systematic pattern as they fluctuate around zero. Furthermore, the residual autocorrelation function does not show a robust remaining autocorrelation. At conventional significance levels, the Ljung–Box test yielded a p-value of 0.05925, which indicates that there is no significant residual autocorrelation issue. Thus, the residuals are broadly consistent with white noise.

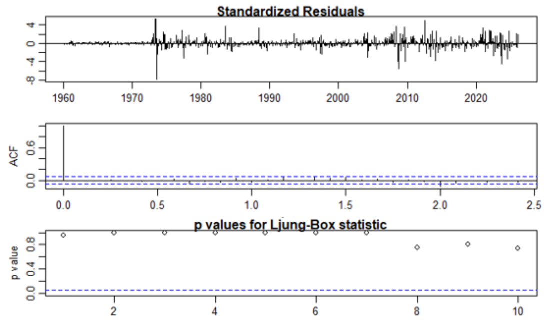

Figure 10: Residual diagnostics of the auto ARIMA (1,1,3) model

The auto ARIMA (1,1,3) model's forecast is illustrated in Figure 11. The point forecasts indicate that soybean prices will remain relatively stable, fluctuating within the $478–482 per metric ton range throughout the forecast period. Simultaneously, the prediction intervals progressively expand, suggesting that the forecast horizon expands, thereby increasing the level of uncertainty

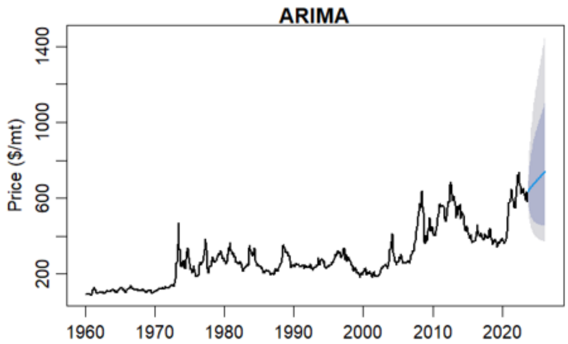

Figure 11: Forecasts from the auto ARIMA (1,1,3) model

Overall, the diagnostic and forecasting results indicate that the auto ARIMA (1,1,3) model is appropriate for forecasting the soybean price series and provides an acceptable fit.

Forecast Comparison across Models

The forecast performance was compared across six alternative models: ARIMA, ETS, NNAR, Theta, STLM, and TBATS. The comparison was conducted using out-of-sample accuracy measures, such as the mean error (ME), root mean squared error (RMSE), mean absolute error (MAE), mean percentage error (MPE), and mean absolute percentage error (MAPE). In general, forecasting efficacy becomes more accurate with lower RMSE, MAE, and MAPE values.

Table 3 illustrates that the NNAR model performed better than all other individual models. Specifically, it generated the lowest RMSE (162.8609), MAE (156.8184), and MAPE (36.21687). Therefore, NNAR was identified as the most suitable model among the individual forecasting approaches. The Theta and TBATS models were ranked second and third, respectively. In contrast, the forecast

Table 3: Forecast comparison across ARIMA, ETS, NNAR, Theta, STLM, and TBATS models

Model	ME	RMSE	MAE	MPE	MAPE
NNAR	-156.82	162.86	156.82	-36.22	36.22
THETA	-185.10	192.19	185.10	-42.71	42.71
TBATS	-194.58	200.90	194.58	-44.81	44.81
ARIMA	-240.98	251.84	240.98	-55.68	55.68
STLM	-267.31	281.80	267.31	-61.80	61.80
ETS	-275.74	295.75	275.74	-63.98	63.98

errors of ARIMA, STLM, and ETS were significantly larger. The ETS model had the weakest efficacy among all models, with the highest RMSE (295.7469) and MAPE (63.97945). Thus, the comparison indicates that the more adaptable nonlinear model, NNAR, captured soybean price dynamics more effectively than the alternative individual methods in this study.

Figure 12 presents the forecast comparison across the

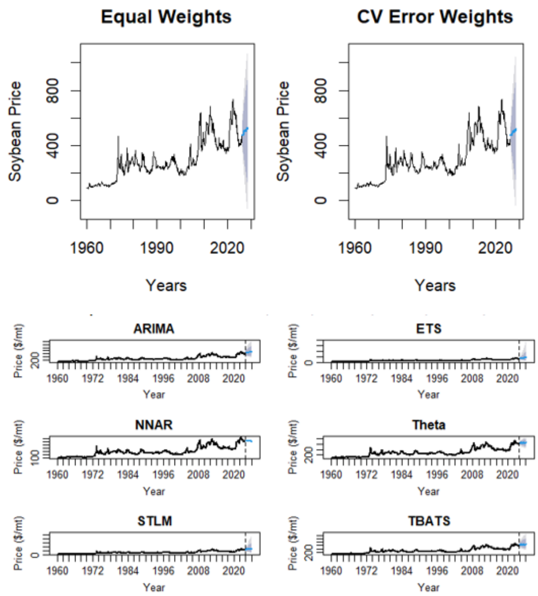

Figure 12: Forecast comparison across ARIMA, ETS, NNAR, Theta, STLM, and TBATS models

ARIMA, ETS, NNAR, Theta, STLM, and TBATS models. As shown in the figure, the models generate different forecast paths and uncertainty ranges. The ARIMA, ETS, STLM, and TBATS models show relatively wider forecast bands, indicating higher uncertainty. In contrast, the NNAR forecast is comparatively smoother and more stable, while the Theta model shows moderate fluctuation. These visual differences are consistent with the forecast accuracy results, which indicate that NNAR

Table 4: Forecast Accuracy Comparison of All Models

Model	ME	RMSE	MAE	MPE (%)	MAPE (%)
EQUAL (Simple Average Ensemble)	-210.644	219.873	210.644	-48.679	48.679
CV.ERRORS (Weighted Ensemble)	-202.078	210.48	202.078	-46.681	46.681
AUTO.ARIMA (1,1,3)	-240.976	251.836	240.976	-55.681	55.681
ETS (M,Ad,N)	-275.741	295.747	275.741	-63.979	63.979
NNAR (26,1,14)	-156.818	162.861	156.818	-36.217	36.217
THETA	-185.105	192.189	185.105	-42.712	42.712
STLM (M,Ad,N)	-267.31	281.803	267.31	-61.804	61.804
TBATS (0,0,0,-,-)	-194.582	200.897	194.582	-44.806	44.806
Hybrid 1 (AUTO.ARIMA (1,1,3) × ETS (M,Ad,N))	-177.258	184.621	177.258	-40.978	40.978
Hybrid 2 (AUTO.ARIMA (1,1,3) × NNAR (26,1,14))	-94.956	98.028	94.956	-21.662	21.662
Hybrid 3 (AUTO.ARIMA × TBATS)	-182.394	189.167	182.394	-42.094	42.094
Hybrid 4 (AUTO.ARIMA × STLM (M,Ad,N))	-182.689	189.834	182.689	-42.186	42.186
Hybrid 5 (ETS (M,Ad,N) × NNAR (26,1,14))	-102.891	105.987	102.891	-23.486	23.486
Hybrid 6 (ETS × TBATS)	-189.445	196.165	189.445	-43.691	43.691
Hybrid 7 (ETS × STLM (M,Ad,N))	-189.74	196.836	189.74	-43.783	43.783
Hybrid 8 (NNAR × TBATS)	-104.444	108.225	104.444	-23.869	23.869
Hybrid 9 (NNAR × STLM (M,Ad,N))	-103.034	105.574	103.034	-23.46	23.46

Hybrid 10 (TBATS × STLM (M,Ad,N))	-194.876	201.572	194.876	-44.898	44.898
Hybrid 11 (AUTO.ARIMA × ETS × NNAR)	-128.523	133.515	128.523	-29.612	29.612
Hybrid 12 (AUTO.ARIMA × ETS × STLM)	-183.229	190.393	183.229	-42.316	42.316
Hybrid 13 (AUTO.ARIMA × ETS × TBATS)	-183.032	189.954	183.032	-42.254	42.254
Hybrid 14 (Full Ensemble)	-152.926	157.413	152.926	-35.172	35.172

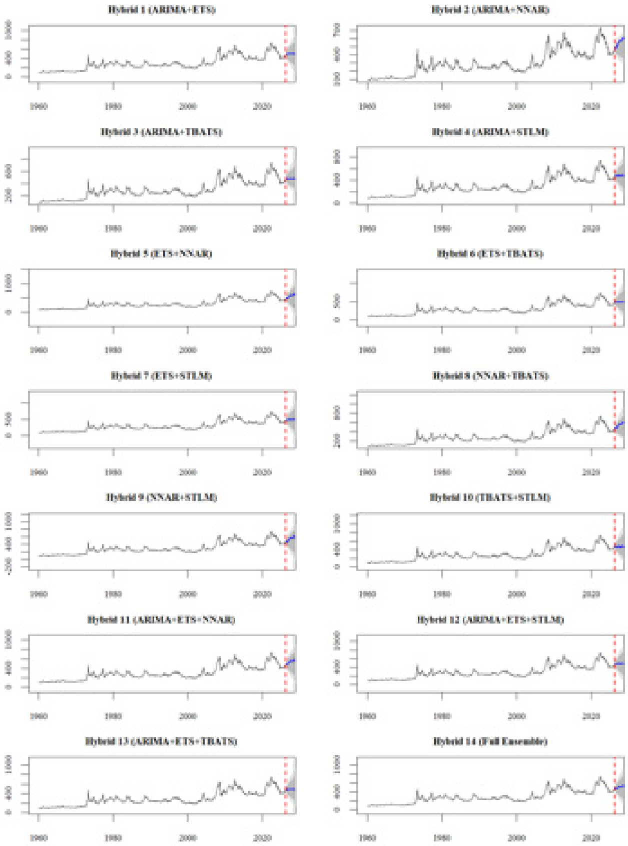

Figure 13: Forecast comparison across hybrid models

Final Forecast

The Hybrid 2 model, which combines AUTO.ARIMA (1,1,3) and NNAR (26,1,14) were selected as the final forecasting models due to their superior predictive efficacy. The predicted values derived from this model are shown in Table 5.

Moreover, the forecast increases from 470.67 $/mt in April 2026 to a peak of approximately 608.97 $/mt in mid-2028, followed by a decline to about 446.67 $/mt by December 2030. This pattern suggests an initial upward price trend, followed by a market correction in later periods.

Table 5: Selected forecast values from the Hybrid 2 model

Month	Forecast	Lo80	Hi80	Lo95	Hi95
2026-04	474.54	452.7	503.08	439.37	516.41
2026-12	512.36	399.02	561.27	356.07	604.22
2027-06	573.06	383.11	578.61	331.36	630.36
2027-12	566.98	372.91	589.82	315.49	647.24
2028-06	608.97	365.55	627.9	304.05	659.4
2028-12	596.19	359.78	604.17	295.1	668.85
2029-06	561.04	354.99	609.31	287.68	676.63
2029-12	515.6	350.83	613.73	281.24	683.32
2030-06	466.07	347.08	617.65	275.46	689.27
2030-12	446.67	343.63	621.23	270.15	694.71

Lo80 and Hi80 represent the lower and upper bounds of the 80% prediction interval, while Lo95 and Hi95 represent the 95% prediction interval.

The forecast from the Hybrid 2 model is presented in Figure 14. The figure shows the historical soybean price series, along with projected values through 2030. The prediction intervals expand as the forecast horizon expands, signifying increasing uncertainty over time. This trend corresponds to standard time-series forecasting behavior, wherein long-term projections are more unpredictable.

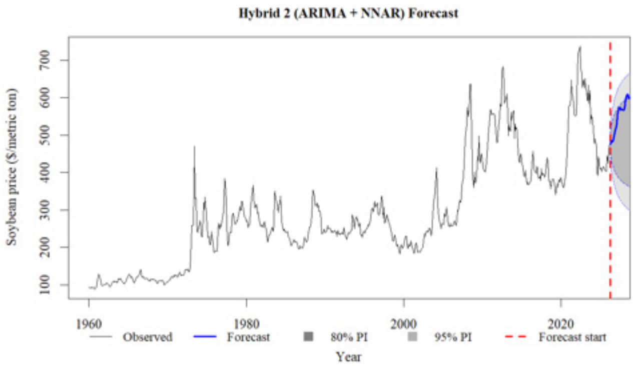

Figure 14: Forecast from the Hybrid 2 (ARIMA × NNAR) model

Discussion

The findings provide important insights into soybean price volatility and the relative performance of alternative forecasting models. The descriptive and statistical study indicates that soybean prices demonstrate significant variability, a long-term upward trend, and instances of increased volatility, especially following the mid-2000s. This volatility may be associated with biofuel expansion, trade changes, financial-market disruptions, and COVID-19-related shocks (Katuwal *et al.*, 2026). This pattern is consistent with previous research that emphasizes the impact of global demand-supply differences, climate changes, and policy measures on agricultural commodity prices (Goodwin & Schnepf, 2000; Sands *et al.*, 2014; Rosegrant *et al.*, 2012; Bathla, 2012). The observed volatility shows the necessity for resilient forecasting systems that can accommodate both structural changes and short-term fluctuations. The empirical findings from the model estimation indicate that the auto ARIMA technique identified a non-seasonal ARIMA (1,1,3) specification. This result indicates that the series dynamics are mostly influenced by trends and short-term dependencies rather than significant seasonal persistence. Similar findings have been reported in the literature, indicating that ARIMA models proficiently capture linear temporal patterns but may inadequately address nonlinear and unpredictable variations in agricultural prices (Kmytiuk *et al.*, 2024; Zevallos-Aquije *et al.*, 2025). Moreover, the good residual diagnostics and lack of significant autocorrelation confirm that the chosen ARIMA model effectively represents the linear structure of the series (Yu *et al.*, 2014; Xue & Hua, 2016).

Nevertheless, the comparative forecasting results indicate that linear models alone are insufficient to achieve high prediction accuracy. The NNAR model outperforms ARIMA, ETS, TBATS, and other comparable models by yielding reduced RMSE, MAE, and MAPE values (Perone, 2022). This outcome corresponds to prior research indicating that neural network models are more efficient at capturing nonlinear linkages and complex relationships in economic and agricultural time series (Jha & Sinha, 2013; Punyapornwithaya *et al.*, 2023). The exceptional efficacy of NNAR in this study emphasizes the nonlinear and volatile characteristics of soybean prices, which typical linear methods fail to adequately capture. Furthermore, the findings strongly support the efficacy of hybrid modeling methodologies. Although NNAR was identified as the best-performing individual model, the hybrid ARIMA–NNAR model (Hybrid 2) achieved the lowest overall forecast errors, outperforming all individual and ensemble models. The hybrid ARIMA–NNAR model (Hybrid 2) attains the minimal forecast errors compared to all other models, indicating a noticeable improvement in predictive accuracy. This discovery corresponds with the theoretical claim that hybrid models can incorporate the advantages of linear and nonlinear methodologies, therefore offering a more thorough depiction of time-series dynamics (Khashei & Bijari, 2011, 2012). Similar empirical evidence has been shown in agricultural and financial forecasting, indicating that hybrid models are consistently superior to individual models (Khandelwal *et al.*, 2015; Kumar & Thenmozhi, 2014).

The forecast comparison shows large differences in uncertainty among the models. Linear models like ARIMA and ETS yield broader prediction intervals, indicating greater forecast uncertainty, while NNAR and hybrid models produce more consistent and steady forecasts. This finding indicates that nonlinear and hybrid methodologies not only improve point forecast precision but also improve the reliability of forecast distributions (Parviz *et al.*, 2023; Haque *et al.*, 2014). These improvements are especially significant in agricultural markets, where uncertainty strongly influences decision-making processes. The ultimate forecast results suggest that soybean prices are projected to rise in the short term, peak between 2027 and 2028, and subsequently fall progressively towards 2030. This anticipated trend illustrates the cyclical characteristics of commodity markets and may be affected by expected changes in global supply, demand, and market dynamics. Similar cyclical patterns have been observed in prior research on agricultural pricing dynamics, wherein periods of price rise are frequently followed by market corrections (Goodwin & Schnepf, 2000). The expanding prediction intervals highlight the core uncertainty linked to long-term forecasts, a prevalent issue in time-series analysis (Makridakis *et al.*, 2018; Hyndman & Athanasopoulos, 2018).

Policy Implications

The findings of this study have significant policy implications aimed at improving decision-making in soybean markets. The extreme price volatility and nonlinear dynamics shown in the soybean price series suggest that traditional linear models alone may be insufficient for market planning under volatile price conditions. The superior performance of the NNAR and Hybrid (ARIMA–NNAR) models indicates that policymakers should implement advanced forecasting systems that integrate both linear and nonlinear components. These models can be integrated into national and regional market information systems to provide more precise and prompt pricing indications. This would facilitate improved planning for production, storage, and marketing strategies. Moreover, governments might leverage these forecasts to formulate more adaptive policies for buffer stock management, import-export regulation, and price stabilization, especially during expected price spikes or market adjustments.

The anticipated price trend, which indicates a short-term increase followed by a decrease towards 2030, shows the necessity for early and strategic governmental initiatives. For example, early warning systems that leverage accurate forecasts can help farmers and traders refine their strategies, including optimizing planting decisions, timing market entry, and managing storage. Additionally, investments in infrastructure, including storage facilities and supply chain optimization, are essential to mitigate the adverse effects of price volatility. These findings underscore the need to integrate data-driven technologies into agricultural policy frameworks. By implementing these measures, policymakers might improve resilience in soybean markets, increase farmers' income stability, and facilitate a more effective allocation of resources within the agricultural sector.

CONCLUSION

This study concludes that soybean price dynamics are mainly driven by long-term movement, volatility, and nonlinear variation rather than strong seasonal persistence. The comparison results indicate that dependence on linear models alone is insufficient, as nonlinear and hybrid methods, especially NNAR and ARIMA–NNAR combinations, provide significantly greater forecasting accuracy by effectively capturing complex market dynamics. The primary value of this study lies not only in determining the optimal model but also in establishing that a combination of linear and nonlinear information is essential to improving forecasts in volatile agricultural markets. The findings offer strong evidence that advanced and hybrid forecasting frameworks are better suited to facilitate decision-making in soybean markets.

Future studies should extend this work by incorporating external variables such as weather conditions, trade policies, exchange rates, stock levels, and demand-supply shocks. The utilization of advanced methodologies, such as deep learning models (e.g., Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU)) and regime-switching or structural break models, may improve forecasting performance by effectively capturing complex temporal patterns and unexpected market shifts. Extending the analysis to multivariate or panel data frameworks across nations would improve the generalizability and policy significance of the findings. Ultimately, the incorporation of real-time data and the assessment of models across various market conditions should yield more resilient and practical insights for farmers, traders, and policymakers.

Author Contributions: For research Conceptualization, D.R.K., L.K. and Y.N.C.; methodology, D.R.K., L.K. and A.D.; software, D.R.K. and Y.N.C.; validation, D.R.K., L.K., Y.N.C. and A.D.; writing—original draft preparation, D.R.K. and A.D.; supervision, D.R.K. and L.K.; project administration, D.R.K. and Y.N.C. All authors have read and agreed to the published version of the manuscript.

Funding: No funding was received for this research.

Data Availability Statement: The data supporting the findings of this study are available from public sources. The datasets used in this research were retrieved from the World Bank Commodity Markets database (available online at: https://www.worldbank.org/en/research/commodity-markets, accessed on May 5, 2026).

Conflicts of Interest: The authors declare no conflicts of interest.

REFERENCES

Abdallah, F. D. M. (2019). Role of time series analysis in forecasting egg production depending on ARIMA model. *Applied Mathematics, 9*(1), 1–5. https://doi.org/10.5923/j.am.20190901.01

Adenomon, M. O., & Emmanuel, P. (2024). Comparison of ARIMA, GARCH and NNAR models for modelling exchange rate in Nigeria. *Journal of the Royal Statistical Society Nigeria Group (JRSS-NIG Group), 1*(1), 110–130. Retrieved from JRSS-NIG article page

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705

Almarashi, A. M. (2025). Statistical modeling and forecasting of HIV and anti-retroviral therapy cases by time-series and machine learning models. Scientific *Reports, 15,* 27033. https://doi.org/10.1038/s41598-025-27033

Almarashi, A. M., Daniyal, M., & Jamal, F. (2024a). Modeling the GDP of KSA using linear and non-linear NNAR and hybrid stochastic time series models. *PLOS ONE, 19*(2), e0297180. https://doi.org/10.1371/journal.pone.0297180

Almarashi, A. M., Daniyal, M., & Jamal, F. (2024b). A novel comparative study of NNAR approach with linear stochastic time series models in predicting tennis player's performance. *BMC Sports Science, Medicine and Rehabilitation, 16*, 28. https://doi.org/10.1186/s13102-024-00815-7

Andreozzi, L., Blaconá, M. T., & Magnano, L. (2014). Time series models for different seasonal patterns. *In Proceedings of the 34th International Symposium on Forecasting (ISF 2014).* Retrieved from CONICET publication record

Arumugam, V., & Natarajan, V. (2023). Time series modeling and forecasting using autoregressive integrated moving average and seasonal autoregressive integrated moving average models. *Instrumentation, Mesure, Metrologie, 22*(4), 161–168. https://doi.org/10.18280/i2m.220404

Athanasopoulos, G., Hyndman, R. J., Song, H., & Wu, D. (2011). The tourism forecasting competition. *International Journal of Forecasting, 27*(3), 822–844. https://doi.org/10.1016/j.ijforecast.2010.04.009

Babu, C. N., & Reddy, B. E. (2014). A moving-average filter-based hybrid ARIMA–ANN model for forecasting time series data. *Applied Soft Computing, 23*, 27–38. https://doi.org/10.1016/j.asoc.2014.05.028

Bandara, K., Hyndman, R. J., & Bergmeir, C. (2025). MSTL: A seasonal-trend decomposition algorithm for time series with multiple seasonal patterns. *International Journal of Operational Research, 52*(1), 79–98. https://doi.org/10.1504/IJOR.2025.143957

Bathla, S. (2012). Volatility in agricultural commodity prices in India: Impact and macroeconomic and sector-specific policy responses. *In the 123rd EAAE Seminar, Dublin, Ireland.* https://doi.org/10.22004/ag.econ.122543

Bernal, J. L., Cummins, S., & Gasparrini, A. (2021). Interrupted time series analysis using autoregressive integrated moving average (ARIMA) models: A guide for evaluating large-scale health interventions. *BMC Medical Research Methodology, 21,* 58. https://doi.org/10.1186/s12874-021-01235-8

Boniface, A., & Martin, A. (2019). Time series modeling and forecasting of the consumer price index in Ghana. J*ournal of Advances in Mathematics and Computer Science, 32(1), 1*–11. https://doi.org/10.9734/jamcs/2019/v32i130134

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2016). *Time series analysis: Forecasting and control* (5th ed.). John Wiley & Sons. As above

Chi, Y. N. (2021). Time series forecasting of the global price of soybeans using a hybrid SARIMA and NARNN model. *Data Science: Journal of Computing and Applied Informatics, 5*(2). https://doi.org/10.32734/jocai.v5.i2-5674

Cleveland, R. B., Cleveland, W. S., McRae, J. E., & Terpenning, I. (1990). STL: A seasonal-trend decomposition procedure. *Journal of Official Statistics, 6*(1), 3–73.

Crone, S. F., Hibon, M., & Nikolopoulos, K. (2011). Advances in forecasting with neural networks? Empirical evidence from the NN3 competition on time series prediction. *International Journal of Forecasting, 27*(3), 635–660. https://doi.org/10.1016/j.ijforecast.2011.04.001

Dai, X., Chen, L., Hou, Y., Ning, X., Zhao, W., Cui, Y., Liu, J., & Wang, M. (2025). A comparative study of neural network models for China's soybean futures price forecasting. *Agriculture, 15*(24), 2586. https://doi.org/10.3390/agriculture15242586

Daniyal, M., Tawiah, K., Muhammadullah, S., & Opoku-Ameyaw, K. (2022). Comparison of conventional modeling techniques with the neural network autoregressive model (NNAR): Application to COVID-19 data. *Journal of Healthcare Engineering, 2022*, 4802743. https://doi.org/10.1155/2022/4802743

De Livera, A. M., Hyndman, R. J., & Snyder, R. D. (2011). Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of the American Statistical Association, 106*(496), 1513–1527. https://doi.org/10.1198/jasa.2011.tm09771

Divisekara, R. W., Jayasinghe, G. J. M. S. R., & Kumari, K. W. S. N. (2021). Forecasting the red lentils commodity market price using SARIMA models. *SN Business & Economics, 1,* Article 20. https://doi.org/10.1007/s43546-020-00020-4

Dozie, K. C. N., & Ijomah, M. A. (2020). A comparative study on additive and mixed models in descriptive time series. *American Journal of Mathematical and Computer Modeling, 5*(1), 12–17. https://doi.org/10.11648/j.ajmcm.20200501.12

Dubey, A. K., Kumar, A., García-Díaz, V., Sharma, A. K., & Kanhaiya, K. (2021). Study and analysis of SARIMA and LSTM in forecasting time series data. *Sustainable Energy Technologies and Assessments, 47,* 101474. https://doi.org/10.1016/j.seta.2021.101474

Duan, J., Wang, P., Ma, W., Fang, S., & Hou, Z. (2022). A novel hybrid model based on nonlinear weighted combination for short-term wind power forecasting. *International Journal of Electrical Power & Energy Systems, 134,* 107452. https://doi.org/10.1016/j.ijepes.2021.107452

Elseidi, M. (2024). MSTL-NNAR: A new hybrid model of machine learning and time series decomposition for wind speed forecasting. Stochastic *Environmental Research and Risk Assessment, 38,* 2613–2632. https://doi.org/10.1007/s00477-024-02701-7

Flachowsky, G., Meyer, U., & Südekum, K.-H. (2017). Land use for edible protein of animal origin—A review. *Animals, 7*(3), 25. https://doi.org/10.3390/ani7030025

Goldsmith, P. D. (2008). Economics of soybean production, marketing, and utilization. In L. A. Johnson, P. J. White, & R. Galloway (Eds.), *Soybeans: Chemistry, production, processing, and utilization* (pp. 117–150). AOCS Press. https://doi.org/10.1016/B978-1-893997-64-6.50008-1

Goodwin, B. K., & Schnepf, R. (2000). Determinants of endogenous price risk in corn and wheat futures markets. *Journal of Futures Markets, 20*(8), 753–774. https://doi.org/10.1002/1096-9934(200009)20:8

Gottimukkala, S. R., & Bhuram, S. K. (2025). The impact of soybean price fluctuations on final product pricing. *In Proceedings of the 2025 10th International Conference on Energy Efficiency and Agricultural Engineering (EE & AE)*. IEEE. https://doi.org/10.1109/EEAE65901.2025.11273420

Granger, C. W. J. (2001). *Essays in econometrics: The collected papers of Clive W. J. Granger*. Cambridge University Press. https://doi.org/10.1017/CBO9780511753978

Gunstone, F. D. (Ed.). *(2011). Vegetable oils in food technology: Composition, properties and uses. Blackwell Publishing Ltd.* https://doi.org/10.1002/9781444339925

Hajirahimi, Z., & Khashei, M. (2023). Hybridization of hybrid structures for time series forecasting: A review. *Artificial Intelligence Review, 56,* 1201–1261. https://doi.org/10.1007/s10462-022-10199-0

Haile, M. G., Wossen, T., Tesfaye, K., & von Braun, J. (2017). Impact of climate change, weather extremes, and price risk on global food supply. *Economics of Disasters and Climate Change, 1,* 55–75. https://doi.org/10.1007/s41885-017-0005-2

Haque, A. U., Nehrir, M. H., & Mandal, P. (2014). A hybrid intelligent model for deterministic and quantile regression approach for probabilistic wind power forecasting. *IEEE Transactions on Power Systems, 29*(4), 1663–1672. https://doi.org/10.1109/TPWRS.2014.2299801

Hutagalung, A. A., & Sari, R. F. (2024). Forecasting Indihome users by using TBATS methods. Mathline: *Journal of Mathematics and Education, 9*(1). https://doi.org/10.31943/mathline.v9i1.579

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

Hyndman, R. J., Koehler, A. B., Snyder, R. D., & Grose, S. (2002). A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting, 18(*3), 439–454. https://doi.org/10.1016/S0169-2070(01)00110-8

Hyndman, R. J., Wang, E., & Laptev, N. (2015). Large-scale unusual time series detection. In *2015 IEEE International Conference on Data Mining Workshop (ICDMW)* (pp. 1616–1619). IEEE. https://doi.org/10.1109/ICDMW.2015.104

Jadhav, V., Reddy, B. V. C., & Gaddi, G. M. (2018). *Application of the ARIMA model for forecasting agricultural prices*. Jomo Kenyatta University of Agriculture and Technology Repository. http://hdl.handle.net/123456789/3583

Jaiswal, R., Jha, G. K., Kumar, R. R., & Choudhary, K. (2026). STL-LSTM hybrid model for forecasting seasonal agricultural price series. *Annals of Data Science, 13(*2), 279–302. https://doi.org/10.1007/s40745-025-00590-3

Jaiswal, R., Jha, G. K., Kumar, R. R., & Lama, A. (2023). Agricultural price forecasting using NARX model for soybean oil. *Current Science, 125*(1), 79–84. https://doi.org/10.18520/cs/v125/i1/79-84

Jha, G. K., & Sinha, K. (2013). Agricultural price forecasting using neural network model: An innovative information delivery system. *Agricultural Economics Research Review, 26*(2). https://doi.org/10.22004/ag.econ.162150

Katuwal, D. R., Thapa, B., Dhakal, A., Limbu Sanwa, R., & Pokhrel, A. (2026). Price, inventory, and trade dynamics in the U.S. soybean market with structural breaks and VAR analysis. *American Journal of Applied Statistics and Economics, 5*(1), 139–155. https://doi.org/10.54536/ajase.v5i1.6992

Khandelwal, I., Adhikari, R., & Verma, G. (2015). Time series forecasting using hybrid ARIMA and ANN models based on DWT decomposition. *Procedia Computer Science, 48,* 173–179. https://doi.org/10.1016/j.procs.2015.04.167

Khashei, M., & Bijari, M. (2011a). A new hybrid methodology for nonlinear time series forecasting. *Modeling and Simulation in Engineering, 2011,* Article 379121. https://doi.org/10.1155/2011/379121

Khashei, M., & Bijari, M. (2011b). A novel hybridization of artificial neural networks and ARIMA models for time series forecasting. *Applied Soft Computing, 11*(2), 2664–2675. https://doi.org/10.1016/j.asoc.2010.10.015

Khashei, M., & Bijari, M. (2012). A new class of hybrid models for time series forecasting. *Expert Systems with Applications, 39*(4), 4344–4357. https://doi.org/10.1016/j.eswa.2011.09.157

Kim, S., & Kim, H. (2016). A new metric of absolute percentage error for intermittent demand forecasts. *International Journal of Forecasting, 32*(3), 669–679. https://doi.org/10.1016/j.ijforecast.2015.12.003

Kmytiuk, T., Majore, G., & Bilyk, T. (2024). Time series forecasting of the price of agricultural products using data science. *Agricultural and Resource Economics, 10*(3), 5–33.

Kumar, M., & Thenmozhi, M. (2014). Forecasting stock index returns using ARIMA-SVM, ARIMA-ANN, and ARIMA-random forest hybrid models. *International Journal of Banking, Accounting and Finance, 5*(3), 284–308.

Khadka, R., Limbu Sanwa, R., Khadka, N., & Chi, Y. N. (2025). Forecasting the global price of corn using a neural network and a hybrid approach. *IEEE Access*. https://doi.org/10.1109/ACCESS.2025.3611320

Limbu Sanwa, R., Khadka, R., & Chi, Y. N. (2025).

Forecasting global monthly cotton prices: The superiority of NNAR models over traditional models. *Frontiers in Artificial Intelligence, 8,* 1628744. https://doi.org/10.3389/frai.2025.1628744

Ljung, G. M., & Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika, 65*(2), 297–303. https://doi.org/10.2307/2335207

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). The M4 competition: Results, findings, conclusion, and way forward. *International Journal of Forecasting, 34*(4), 802–808. https://doi.org/10.1016/j.ijforecast.2018.06.001

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2022). M5 accuracy competition: Results, findings, and conclusions. *International Journal of Forecasting, 38(*4), 1346–1364. https://doi.org/10.1016/j.ijforecast.2021.11.013

Mamun, A. A., Sohel, M., Mohammad, N., Sunny, M. S. H., Dipta, D. R., & Hossain, E. (2020). A comprehensive review of load forecasting techniques. *IEEE Access, 8*, 134911–134939. https://doi.org/10.1109/ACCESS.2020.3010702

Momin, B., & Chavan, G. (2018). Univariate time series models for forecasting. In S. Satapathy & A. Joshi (Eds.), *ICTIS 2017* (pp. 219–226). Springer. https://doi.org/10.1007/978-3-319-63645-0_24

Nelson, B. K. (1998). Time series analysis using ARIMA models. *Academic Emergency Medicine, 5*(7), 739–744. https://doi.org/10.1111/j.1553-2712.1998.tb02493.x

Osho, G. S. (2019). A general framework for a time series forecasting model using ARIMA. *International Journal of Statistics and Probability, 8*(6), 23–31. https://doi.org/10.5539/ijsp.v8n6p23

Ostertagová, E., & Ostertag, O. (2012). Forecasting using the simple exponential smoothing method. *Acta Electrotechnica et Informatica, 12*(3), 62–66. https://doi.org/10.2478/v10198-012-0034-2

Ostertagová, E., & Ostertag, O. (2015). Regression analysis and seasonal adjustment of time series. *Journal of Automation and Control, 3(*3), 118–121. https://doi.org/10.12691/automation-3-3-16

Pagano, M. C., & Miransari, M. (2016). The importance of soybean production worldwide. In M. Miransari (Ed.), *Abiotic and biotic stresses in soybean production* (Vol. 1, pp. 1–26). Academic Press. https://doi.org/10.1016/B978-0-12-801536-0.00001-3

Panigrahi, S., & Behera, H. S. (2017). A hybrid ETS–ANN model for time series forecasting. *Engineering Applications of Artificial Intelligence, 66,* 49–59. https://doi.org/10.1016/j.engappai.2017.07.007

Parviz, L., Rasouli, K., & Torabi Haghighi, A. (2023). Improving hybrid models for precipitation forecasting. *Water Resources Management, 37,* 3833–3855. https://doi.org/10.1007/s11269-023-03528-7

Paul, R. K., Yeasin, M., Kumar, P., Kumar, P., Balasubramanian, M., Roy, H. S., Paul, A. K., & Gupta, A. (2022). Machine learning techniques for forecasting agricultural prices: A case of brinjal in Odisha, India. *PLOS ONE, 17*(7), e0270553. https://doi.org/10.1371/journal.pone.0270553

Peng, D., Zhang, H., Zhang, Y., Li, X., Huete, A., & Zhang, B. (2026). Global soybean trade dynamics: Drivers, impacts, and sustainability. *The Innovation, 7(*2), 101124. https://doi.org/10.1016/j.xinn.2026.101124

Peri, M. (2017). Climate variability and the volatility of global maize and soybean prices. *Food Security, 9*(4), 673–687. https://doi.org/10.1007/s12571-017-0702-2

Perone, G. (2022). Comparison of ARIMA, ETS, NNAR, TBATS, and hybrid models. *European Journal of Health Economics, 23,* 917–940. https://doi.org/10.1007/s10198-021-01347-4

Perron, P. (1989). The great crash, the oil price shock, and the unit root hypothesis. *Econometrica, 57*(6), 1361–1401. https://doi.org/10.2307/1913712

Petropoulos, F., Apiletti, D., Assimakopoulos, V., Babai, M. Z., Barrow, D. K., Ben Taieb, S., Bergmeir, C., Bessa, R. J., Bijak, J., Boylan, J. E., Browell, J., Carnevale, C., Castle, J. L., Cirillo, P., Clements, M. P., Cordeiro, C., Oliveira, F. L. C., De Baets, S., Dokumentov, A., … Ziel, F. (2022). Forecasting: Theory and practice. *International Journal of Forecasting, 38*(2), 705–871. https://doi.org/10.1016/j.ijforecast.2021.11.001

Punyapornwithaya, V., Arjkumpa, O., Buamithup, N., Kuatako, N., Klaharn, K., Sansamur, C., & Jampachaisri, K. (2023). Forecasting of daily new lumpy skin disease cases in Thailand at different stages of the epidemic using fuzzy logic time series, NNAR, and ARIMA methods. *Preventive Veterinary Medicine, 217*, 105964. https://doi.org/10.1016/j.prevetmed.2023.105964

Rosegrant, M. W., Batka, M., Bhandary, P., Bryan, E., Cenacchi, N., De Pinto, A., Enahoro, D., Leroy, L., Li, M., Magnan, N., & The IMPACT Development Team. (2012). I*nternational model for policy analysis of agricultural commodities and trade (IMPACT): Model description.* International Food Policy Research Institute (IFPRI). https://hdl.handle.net/10568/154027

Safi, S. K., & Sanusi, O. I. (2021). Hybrid ANN, ETS, and ARIMA models. *Mathematical Problems in Engineering,* 2021. https://doi.org/10.3233/MAS-2105

Sands, R. D., Jones, C., & Marshall, E. P. (2014). *Global drivers of agricultural demand and supply* (Economic Research Report No. 174). U.S. Department of Agriculture, Economic Research Service. https://doi.org/10.22004/ag.econ.186137

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*(2), 461–464. Retrieved from: https://www.jstor.org/stable/2958889

Shcherbakov, M. V., Brebels, A., Shcherbakova, N. L., Tyukov, A. P., Janovsky, T. A., & Kamaev, V. A. (2013). A survey of forecast error measures. *World Applied Sciences Journal, 24* (Information Technologies in Modern Industry, Education & Society), 171–176. https://doi.org/10.5829/idosi.wasj.2013.24.itmies.80032

Shmueli, G., & Koppius, O. R. (2011). Predictive analytics in information systems research. *MIS Quarterly, 35*(3), 553–572. https://doi.org/10.2307/23042796

Shumway, R. H., & Stoffer, D. S. (2017). *Time series analysis and its applications* (4th ed.). Springer. https://doi.org/10.1007/978-3-319-52452-8

Siamabele, B. (2021). The significance of soybean production in changing climates. *Cogent Food & Agriculture, 7*(1), 1933745. https://doi.org/10.1080/23311932.2021.1933745

Sina, L. B., Secco, C. A., Blazevic, M., & Nazemi, K. (2023). Hybrid forecasting methods—A systematic review. *Electronics, 12*(9), 2019. https://doi.org/10.3390/electronics12092019

Spiliotis, E., Assimakopoulos, V., & Makridakis, S. (2020). Generalizing the Theta method. *European Journal of Operational Research, 284*(2), 550–558. https://doi.org/10.1016/j.ejor.2020.01.007

Stellwagen, E., & Tashman, L. (2013). ARIMA: The models of Box and Jenkins. Foresight: The *International Journal of Applied Forecasting, 30,* 28–33

Tripathi, P. K., Singh, C. K., Singh, R., & Deshmukh, A. K. (2023). Farmer-centric decision support system. Benchmarking: *An International Journal, 30*(10), 3925–3952. https://doi.org/10.1108/BIJ-12-2021-0780

Verbesselt, J., Hyndman, R., Newnham, G., & Culvenor, D. (2010). Detecting trend and seasonal changes. *Remote Sensing of Environment, 114*(1), 106–115. https://doi.org/10.1016/j.rse.2009.08.014

Wang, W., & Lu, Y. (2018). Analysis of MAE and RMSE. *IOP Conference Series, 324*, 012049. https://doi.org/10.1088/1757-899X/324/1/012049

Wei, W. W. S. (2006). *Time series analysis: Univariate and multivariate methods* (2nd ed.). Addison-Wesley.

Wilkinson, J. M., & Lee, M. R. F. (2018). Use of human-edible animal feeds. *Animal, 12*(8), 1735–1743. https://doi.org/10.1017/S175173111700218X

Willmott, C. J., & Matsuura, K. (2005). Advantages of MAE over RMSE. *Climate Research, 30,* 79–82. https://doi.org/10.3354/cr030079

World Bank. (2026). *Commodity markets outlook: April 2026*. World Bank Group. https://openknowledge.worldbank.org/entities/publication/3ede6f8c-ba03-4176-867c-40e604698570

Xia, C. (2024). Comparative analysis of ARIMA and LSTM models for agricultural product price forecasting. *Highlights in Science, Engineering and Technology, 85*, 1032–104

Xue, D., & Hua, Z. (2016). An ARIMA-based time series forecasting model. *Recent Advances in Electrical & Electronic Engineering, 9*(2), 93–98.

Young, W. L. (1977). The Box-Jenkins approach to time series analysis and forecasting: Principles and applications. RAIRO. *Recherche opérationnelle, 11*(2), 129–143.

Yu, L., Wang, S., & Lai, K. K. (2008). Forecasting crude oil price with an EMD-based neural network ensemble learning paradigm. *Energy Economics, 30*(5), 2623–2635. https://doi.org/10.1016/j.eneco.2008.05.003

Yu, L., Zhou, L., Tan, L., Jiang, H., Wang, Y., Wei, S., & Nie, S. (2014). Hybrid ARIMA–NARNN model. *PLOS ONE, 9*(6), e98241. https://doi.org/10.1371/journal.pone.0098241

Zevallos-Aquije, A., Zevallos-Aquije, A., Palomino-Salcedo, K., & Salas-Bolaños, R. A. (2025). Improving accuracy in agricultural forecasts: Evaluation of SARIMA and forecast models for artichoke prices. *Research on World Agricultural Economy, 6*(3), 273–290. https://doi.org/10.36956/rwae.v6i3.1880

Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing, 50,* 159–175. https://doi.org/10.1016/S0925-2312(01)00702-0

Zhu, W., Li, J., & Xie, T. (2024). Impact of climate change on soybean production. *Advances in Resources Research, 4*(3), 474–496. https://doi.org/10.50908/arr.4.3_474