American Journal of Economics and Business Innovation (AJEBI)

ISSN: 2831-5588 (ONLINE), 2832-4862 (PRINT)

VOLUME 5 ISSUE 2 (2026)

PUBLISHED BY
E-PALLI PUBLISHERS, DELAWARE, USA

e palli

American Journal of Economics and Business Innovation (AJEBI)

Volume 5 Issue 2, Year 2026
ISSN: 2831-5588 (Online), 2832-4862 (Print)
DOI: https://doi.org/10.54536/ajebi.v5i2.6298
https://journals.e-palli.com/home/index.php/ajebi

Forecasting Tomato Prices in the United States Using a Seasonal ARIMA Model

Rumita Limbu Sanwa[1*], Raksha Khadka[1], Lila Karki[2]

Article Information

Received: February 02, 2026

Accepted: April 04, 2026

Published: July 28, 2026

Keywords

Arima, Box-Jenkins, Methodology, Price Forecasting, Time Series Analysis, Tomato Prices

ABSTRACT

The price of tomatoes often changes with seasonal patterns, supply conditions, and shifting consumer demands. As a globally important crop, these variations can create uncertainty for producers and market planners, making price forecasting important. To better understand how prices might change in the near future, we examined the monthly retail prices of field-grown tomatoes in the U.S. (May 2020 - August 2025) using the data reported in the Federal Reserve Economic Data (FRED). Applying the Box-Jenkins methodology found that a SARIMA (1,0,1) (0,0) [12] model was the best-fit model based on the Akaike (AIC) and the Bayesian (BIC) Information Criteria. Model diagnostics showed that the residuals behaved like white noise (Box-Ljung test, p = 0.9734), confirming that the model was appropriate. Also, the model demonstrated high predictive accuracy (RMSE = 0.0448; MAPE = 1.79%). The price was forecasted from September 2025 to July 2027, which indicated a relatively stable price ranging from $1.90 to $1.94 per pound. The results of our study demonstrated that the SARIMA model can effectively capture short-term patterns in tomato prices, which are valuable for market planning and risk management. However, it is important to note that these predictions indicate the expected values and should be seen as guidance rather than exact figures.

INTRODUCTION

Tomatoes (Solanum lycopersicum L.) are rich in vitamins, carotenoids, and phenolic compounds, which support nutrition and overall health. They are grown widely around the world, both for fresh consumption and for making processed products (Quinet *et al.*, 2019). Their market prices can change frequently in response to supply and demand, as they are highly perishable. Hence, accurate price forecasting during the harvest period is very important for growers to make informed and timely production decisions and plan accordingly (Reddy, 2018). Timely and accurate information on agricultural commodity prices is very important in today's rapidly changing market. Predicting these prices is difficult because of uncertainties in production, demand, and market conditions. Time series models improve forecast accuracy and decision-making (Prathilothamai *et al.*, 2025). Various external factors, like adverse weather, natural disasters, global demand, and economic uncertainties, frequently affect agricultural commodities, leading to frequent price fluctuations and market volatility. Reliable forecasting of price trends can help policymakers and farmers to make informed decisions and promote stability within the agricultural sector (Bernard *et al.*, 2025). For perishable products like tomatoes, which are harvested seasonally, informed expected prices can serve as a valuable tool for making better decisions and handling price risks effectively (Mathenge Mutwiri, 2019). Frequent price fluctuations can lead to losses for farmers and higher costs for consumers. Therefore, price forecasting of horticultural crops is essential to guide farmers' decisions and enhance the profitability of agribusinesses. The Autoregressive Integrated Moving Average (ARIMA) model is one of the most widely used linear approaches for price forecasting in time series analysis (Prakash & Farzana, 2020).

LITERATURE REVIEW

The change in weather patterns, production level, market demand, and economic uncertainties all together create hardship for farmers and stakeholders, especially in regions where agriculture plays a major economic role. Consequently, reliable information on price is considered valuable for guiding market planning, risk management, and policy decisions (Manogna *et al.*, 2025). Crop prices depend on several interrelated factors, including land use, consumer demand, and climate variability. As a result, accurate price prediction is challenging. However, developing reliable forecasting models is essential to support market stability and help participants make informed decisions. If the farmers have a better understanding of the expected prices, they can choose the most favorable time to sell their produce to maximize earnings, while the government can use this information to plan post-harvest storage and management strategies that help maintain a stable market throughout the year

[1] Department of Agriculture, Food, & Resource Sciences, School of Agricultural and Natural Sciences, University of Maryland Eastern Shore (UMES), Princess Anne, MD 21853, USA
[2] Department of Agriculture, UMES Extension, School of Agricultural and Natural Sciences, University of Maryland Eastern Shore, Princess Anne, MD 21853, USA
[*] Corresponding author's e-mail: rlimbusanwa@umes.edu

(Purohit *et al.*, 2021).

As all the people depend on agriculture directly or indirectly, informed commodity prices can help maintain financial stability, guiding investment choices and ensuring food security. (Tammali & Mishra, 2024). The Autoregressive Integrated Moving Average (ARIMA), often used in time series research, acts as a practical approach to explore the price behavior and generate forecasts. Numerous studies have shown that it can effectively capture price trends, seasonal patterns, and short-term fluctuations across different crops and regions. For instance, Guzma (2024) highlights how ARIMA can effectively capture price volatility, providing valuable insights that help shape sustainable agricultural policies, guide resource allocation, support the advancement of Sustainable Development Goals, and strengthen economic stability for smallholder farmers. Sahoo (2020) found that using estimated prices can help farmers avoid setting uncompetitive rates and better follow market trends. In the study, they compared actual and forecasted prices to evaluate the effectiveness of forecasting tools. However, actual prices may still fluctuate due to various other factors. Although ARIMA has been focused on a single crop, it can also be utilized to examine price trends in the future for multiple commodities. The predicted price for the chosen commodity closely aligns well with the observed values, highlighting the model's robustness, which proves the model to be a reliable forecasting method, capable of accommodating diverse time series patterns without the need to define parameter values in advance (Jadhav *et al.*, 2017).

Table 1: Application of ARIMA/SARIMA Models for Forecasting Agricultural Commodity Price across Countries.

Arima Model	Country	Commodity	Reference
ARIMA (5,1,4), (6,1,3), & (6,1,1)	Global	Rice, Corn, & Soybeans,	Bernard *et al.*, 2025
ARIMA (4,1,1) (0,0,2)	United States	Cotton	Sanwa & Chi, 2025
ARIMA (1,1,0)	Mexico	Corn	Guzma, 2024
ARIMA (4,0,3)	Philippines	Wheat	Rubillos *et al.*, 2024
ARIMA (1,0,1), (2,0,1), & (3,0,1)	India	Potato	Badal *et al.*, 2022
SARIMA (2, 1, 1) (1, 0, 1)	Kenya	Tomato	Mathenge Mutwiri, 2019

Price forecasting of Tomato using the ARIMA model

Among the countries producing fresh tomatoes, the United States is recognized as a major producer, with an annual production of 12.37 million tonnes in 2023 (FAOSTAT, 2023). The country's significant contribution to global tomato production is highlighted by FAOSTAT data alongside other major producers, such as China, India, Turkey, and Egypt. These statistics highlight the contribution of the United States in meeting the domestic and international demand for fresh tomatoes. However, recent interruptions in the global economy have exposed vulnerabilities within the food supply chain, drawing attention to the need for a thorough understanding of price dynamics and market behavior for perishable commodities such as tomatoes (Cui *et al.*, 2022). The statistical time series models that account for trends and seasonal effects, including ARIMA and its seasonal extension, SARIMA, are commonly used to forecast tomato prices across regions. Reddy (2018) studied monthly tomato prices from major producing states in India using a SARIMA model, and significant regional price variations were observed, largely because of limited cold storage and transportation facilities. The model captured price movements during the harvest season well, although reduced accuracy was observed in the market with high volatility.

Similarly, seasonal price patterns in Turkey's wholesale tomato markets were examined by Adanacıoğlu and Yercan (2012) using a SARIMA(1,0,0)(1,1,1) model with 12-month seasonality. Their results indicated that tomato prices reached a peak in October, indicating strong seasonal effects. Hence, the sustainable tomato production depends on maintaining fair prices for farmers, managing price risks, and adopting good agricultural practices (GAP) to ensure food safety and meet consumer demand. Likewise, wholesale tomato prices in India were evaluated by Puri *et al.* (2023) using the ARIMA (1, 0, 0) (0, 1, 1) 12 model and found that incorporating seasonality improved the model's predictive accuracy. In addition, the study highlighted that reliable price forecasts can guide farmers in making informed decisions regarding storage, marketing, and value-added activities, supporting more stable incomes despite price changes. Overall, these studies demonstrate that ARIMA and SARIMA models are effective for short-term forecasting in the tomato market, provided that consistent time series data, seasonal variation, and a stable market infrastructure are available.

MATERIALS AND METHODS

DATA: The monthly tomato price data from May 2020 to August 2025 were obtained from the Federal Reserve Economic Data (FRED) maintained by the Federal Reserve Bank of St. Louis, with the source being the International Monetary Fund (IMF) (https://fred.

stlouisfed.org/series/APU0000712311). The analysis was carried out in R software (version 4.4.3). The dataset consists of the average retail price of field-grown tomatoes in U.S. city markets, reported in U.S. dollars per pound (453.6 grams) and not seasonally adjusted. The data were transformed into a time series format in R for modeling and forecasting. The predictive ability of the model was examined using established error measures, including Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE), following the approach used by Farzana and Prakash (2025) in their tomato price forecasting study.

Box-Jenkins Methodology

A structured procedure for time series analysis, according to Gujarati (2003), the Box-Jenkins methodology (Box & Jenkins, 1976), establishes a stepwise procedure for building ARIMA models. It consists of four key steps:

1. **Model identification**: The initial step involves checking the stationarity of the time series data, meaning its mean and variance stay stable over time. If the series is non-stationary, differencing is applied (one or more times) until stationarity is achieved (this is the "I" or integration part). Then, the autocorrelation (ACF) and partial autocorrelation function (PACF) plots are checked to see patterns in the data and help decide orders of AR (p) and MA (q). This process helps narrow down a set of candidates (p, d, q) models for further analysis.
2. **Parameter estimation:** After choosing the possible models, the next step is to calculate their parameters, including the AR (φ), MA (θ) coefficients, along with any constant terms. This is usually done using methods like Maximum Likelihood Estimation (MLE) or non-linear least squares, aiming to find parameter values that best represent the historical data, ensuring that the model fits the observed patterns as closely as possible while maximizing the likelihood or minimizing the residual errors (Hyndman, 2001).
3. **Diagnostic checking:** After the model has been fitted, it is essential to evaluate its accuracy. The residuals should have an average of zero, constant variability, and ideally follow a normal distribution. The Ljung-Box test is applied to detect correlation among the errors. If these conditions are not met, the model is adjusted by modifying the values of p, q, or the degree of differencing until a satisfactory fit is achieved.
4. **Forecasting**: Once a model has successfully passed diagnostic checks, it can provide upcoming observations. The ARIMA model uses its estimated structure along with past observations to generate forecasts over different time periods.

Although ARIMA models effectively capture trends and patterns, they may sometimes lag in responding to sudden or unexpected changes (Zhang & Cao, 2025).

ARIMA Model

Originally proposed by Box and Jenkins (1976), the Autoregressive Integrated Moving Average (ARIMA) model provides a standard framework for analyzing and predicting time series that consists of a single variable. The model integrates autoregression (AR), integration or differencing (I) for stationarity, and moving average (MA) to identify trends and patterns in historical observations.

According to Gujarati (2003), the Box-Jenkins methodology offers a systematic procedure for developing ARIMA models, including stages such as identifying the model, estimating parameters, performing diagnostic checks, and generating forecasts.

Autoregressive (AR) model

The Autoregressive (AR) model is a fundamental tool in time series forecasting. Originally introduced by Yule in 1927 and later refined within the Box-Jenkins framework (Box & Jenkins, 1976), the model captures the dependence of the current values of a variable on its previous observations. For AR (p) model, each observation is represented as a linear combination of the last p observations, with an associated random error term. This structure lets the model detect ongoing patterns and the influence of past values on current ones. AR models are widely used across fields, such as economics, finance, and environmental science, to analyze trends and generate forecasts based on historical behavior.

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p \qquad (1)$$

The general representation of the AR(p) polynomial:
where:

ϕ(B)defines the structure of the AR process.

ϕ_iare the autoregressive coefficients, representing the influence of past observations.

B^ishifts the series back i time steps.

Moving Average (MA) Model

The Moving Average (MA) model is used to analyze time series by focusing on past errors rather than past values. It shows the current value of a series as a linear combination of past errors or random shocks. It was originally introduced by Slutsky (1937) and later expanded within the Box-Jenkins methodology (Box & Jenkins, 1976). It helps to reveal how these recent disturbances shape the behavior of a time series. By doing so, it smooths out short-term noise and provides a clearer view of underlying trends.

$$\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q \qquad (2)$$

The moving average (MA) polynomial is expressed as:

θ(B) represents the moving average polynomial operator.

θ_i(for i = 1, 2, …, q) are the parameters of the moving average model that measure the influence of past error terms on the current value.

B is the backshift (or lag) operator, where B^k ϵ_t=ϵ_(t-k).

q denotes the order of the moving average process, indicating the number of lagged error terms included in the model.

The general ARIMA (p, d, q) model can be expressed mathematically using the backshift operator B, as described by Tebbs (2013):

$$\phi_p(B)(1 - B)^d X_t = \theta_q(B)\epsilon_t \qquad (3)$$

where :
X_t is the observed time series value at time t,
ϕ_p (B) is the autoregressive polynomial (AR) of order p,
θ_q (B) is the moving average polynomial (MA) of order q,
□d=(1 - B)^d represents the d-th order differencing operator,
ϵ_t is a white noise (i.i.d. error term).
Differencing is performed as:

$$(1 - B)^d X_t = \Delta^d \qquad (4)$$

The Seasonal Autoregressive Integrated Moving Average (SARIMA) model
The SARIMA model extends the traditional ARIMA framework by adding both seasonal and non-seasonal components, allowing it to capture complex temporal structured patterns in time series data (Brockwell & Davis, 2002). Using the backshift operator B, the general multiplicative form of the SARIMA model is expressed as: Using the backshift operator B, the SARIMA model can be expressed as:

$$\Phi_P(B^s)\phi_p(B)(1 - B)^d(1 - B^s)^D X_t = \Theta_Q(B^s)\theta_q(B)\epsilon_t + C \qquad (5)$$

Where:
p and q are the non-seasonal autoregressive (AR) and moving average (MA) orders,
P and Q the seasonal AR and MA orders (seasonal lags are multiples of s),
ϕ_p (B) and θ_q (B): non-seasonal polynomial AR and MA polynomials,
Φ_P (B^s) and Θ_Q (B^s): seasonal AR and MA polynomials,
(1 - B)^d: non-seasonal differencing,
(1 - B^s)^D: seasonal differencing,
C : constant term (intercept)
ϵ_t : white noise error term

This representation highlights how SARIMA combines seasonal and non-seasonal components multiplicatively to capture both short-term and long-term patterns in the data (Hyndman & Athanasopoulos, 2018; Box *et al.*, 2015).

RESULTS AND DISCUSSION

Identification of the Seasonal ARIMA Model

Table 2: Summary Statistics of monthly tomato prices in the United States (May 2020 - August 2025).

Statistic	Value (U.S. dollars per pound)
Minimum	1.705
1st Quartile	1.836
Median	1.892
Mean	1.906
3rd Quartile	1.949
Maximum	2.228
Variance	0.01
Standard Deviation	0.1

Source: Authors' compilation based on data from the Federal Reserve Economic Data (FRED).

Table 1 displays the summary of monthly tomato prices in the United States from May 2020 to August 2025. During this period, the average retail price was $1.91 per pound, with prices ranging from $1.71 to $2.23 per pound. The small standard deviation (0.10) and variance (0.01) suggest that prices did not fluctuate much, indicating that tomato prices remained stable throughout the observed period. The median price of $1.89 is almost identical to the mean, reflecting a symmetric distribution of prices with no significant skewness. Figure 1 illustrates how the monthly prices of tomatoes in the United States changed between May 2020 and August 2025. The chart shows clear ups and downs over the years, reflecting seasonal trends and market shifts. Prices generally reach their peak during late summer and early autumn, likely due to lower supply or higher demand, while lower prices typically occur in early spring. The highest recorded price was about $2.23 per pound in August 2022, followed

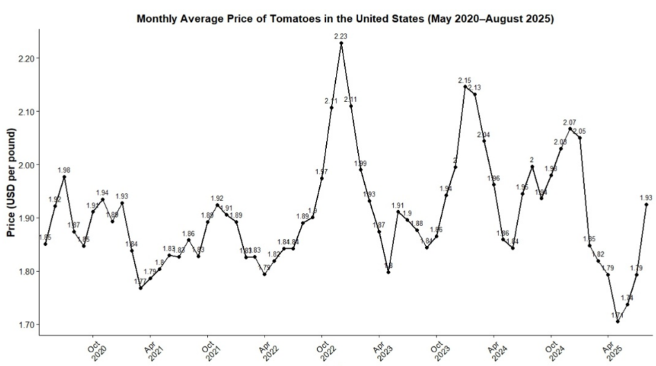

Figure 1: Time series chart of monthly tomato prices in the United States from May 2020 to August 2025.
Source: Authors' compilation based on data from the Federal Reserve Economic Data (FRED).

by another peak of \$2.15 per pound in July 2023. In contrast, the lowest price, around \$1.74 per pound, was observed in March 2025. Overall, the pattern shows that tomato prices tend to exhibit short-term fluctuations and seasonal trends, consistent with the production cycles and market supply-demand dynamics in the market. Because of these predictable shifts, time series approaches such as ARIMA are valuable for capturing and forecasting price behavior agricultural market.

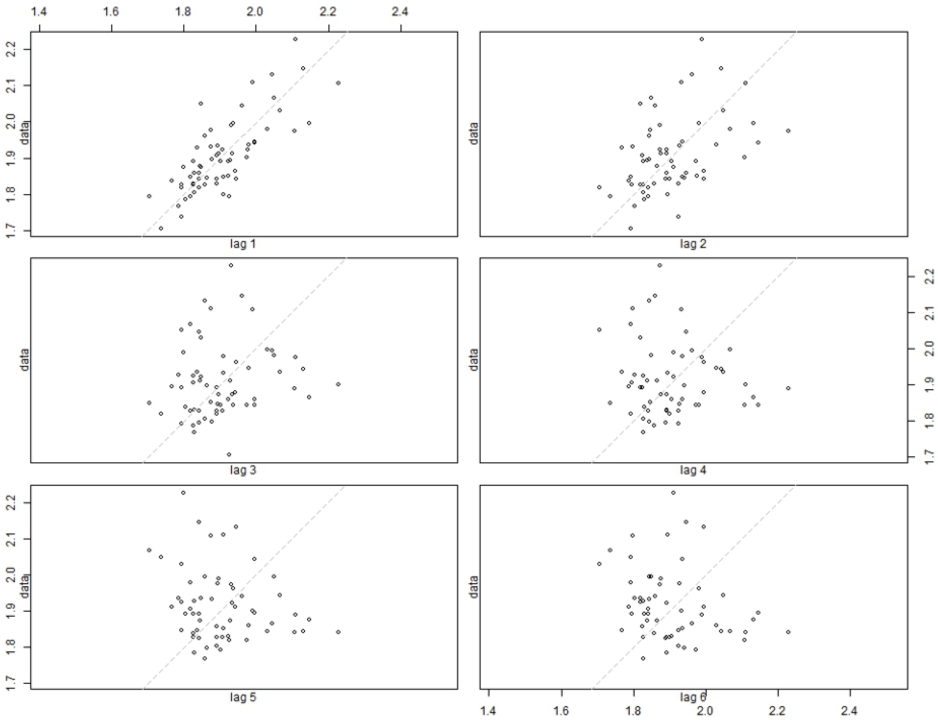

Figure 2: Lag plots of monthly tomato prices in the U.S. from May 2020 to August 2025.

Figure 2 presents lagged scatter plots of monthly tomato prices in the U.S. from May 2020 to August 2025. Each panel compares the price in a given month with prices from one to six months earlier (lags 1-6). A clear positive relationship is observed at lag 1 and lag 2, indicating that tomato prices in one or two consecutive months generally follow the same trend. However, as the lag increases beyond two months, the relationship becomes weaker and more dispersed, suggesting that the influence of past prices diminishes over time. This pattern suggests short-term dependence on tomato prices characterized by mild autocorrelation that fades as the time gap increases. The Augmented Dickey-Fuller (ADF) test was applied to check if the tomato price series is stationary. The test produced statistics of -4.11 with a lag of 3 and a p-value of 0.0105, indicating that the series is stationary at the 5% level.

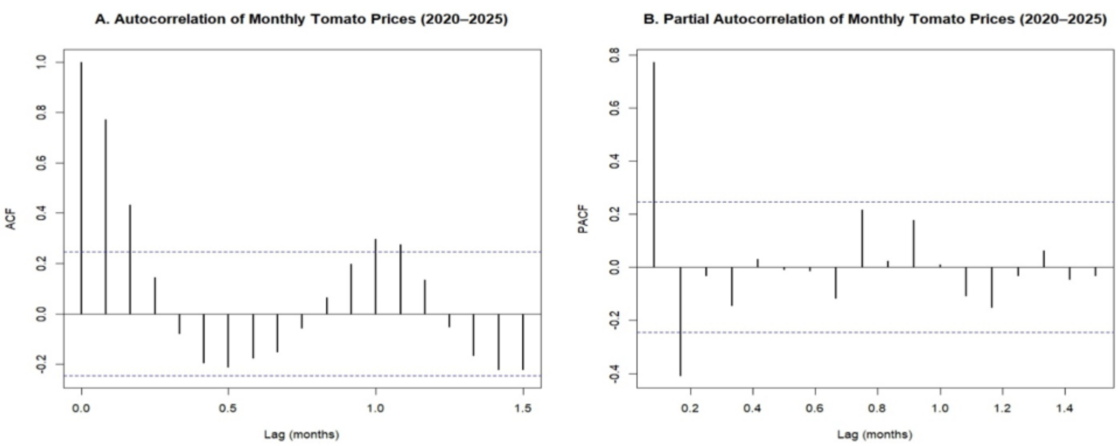

Figure 3: ACF and PACF plots of monthly tomato prices in the United States from May 2020 to August 2025 before differencing.

Figure 3 displays the autocorrelation (ACF) and partial autocorrelation (PACF) plots of monthly tomato prices before differencing. In Panel A, the ACF declines slowly across multiple lags, showing that the series is not stationary and that earlier prices continue to influence later prices. This pattern suggests the presence of a trend or seasonal behavior. Panel B shows the PACF, where only the first lag stands out as significant, while the remaining lags fall within the confidence bands. This indicates short-term dependence mainly at lag 1. The overall pattern of both plots indicates that the observed tomato price series is not stable over time and needs differencing or seasonal adjustment before applying ARIMA or SARIMA models for forecasting.

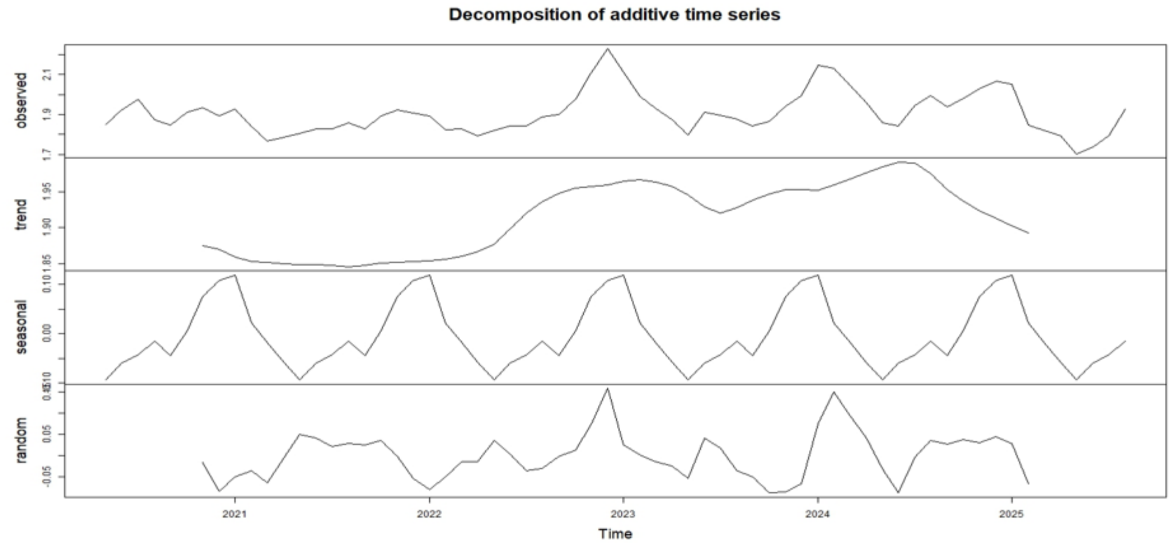

Figure 4: Decomposition of monthly tomato prices in the U. S. from May 2020 to August 2025.

Figure 4 presents the classical decomposition of monthly tomato prices in the U.S. from May 2020 to August 2025 into trend, seasonal, and residual components. This breakdown helps clarify the long-term movement of prices, repeating seasonal patterns, and the irregular fluctuations, giving a clearer picture of how the time series behaves over time. To check whether the seasonally adjusted series is stationary, the Augmented Dickey-Fuller (ADF) test was conducted. The test gave a statistic of 3.1949 with a lag order of 3 and a p-value of 0.0964. Although the negative statistics suggest some tendency toward stationarity, the p-value is above the 0.05 threshold, implying that the null hypothesis of non-stationarity cannot be rejected, and the series is still lacking a constant mean and variance.

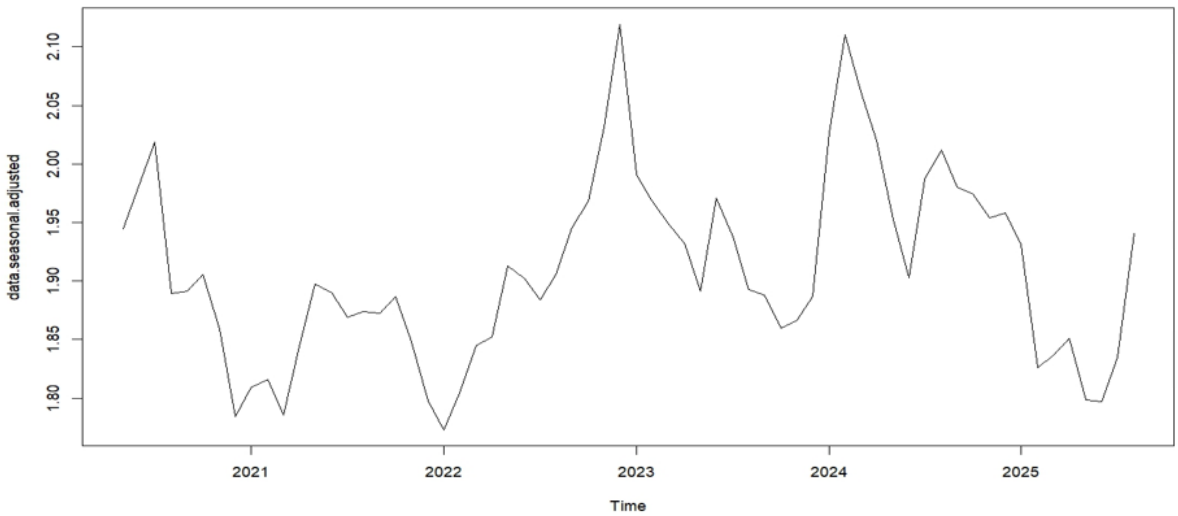

Figure 5: Seasonally adjusted monthly tomato prices in the U.S. from May 2020 to August 2025.

Figure 5 shows the seasonally adjusted monthly tomato price from May 2020 to August 2025. With the seasonal component removed, the remaining fluctuations reflect underlying trends and irregular market variations. The adjusted series does not display any clear long-term upward or downward trend. Instead, these fluctuations may reflect changes in supply conditions, production costs, and market demand rather than seasonal effects. The seasonally adjusted pattern confirms that, after removing seasonality, price variations are mainly driven by non-seasonal market factors.

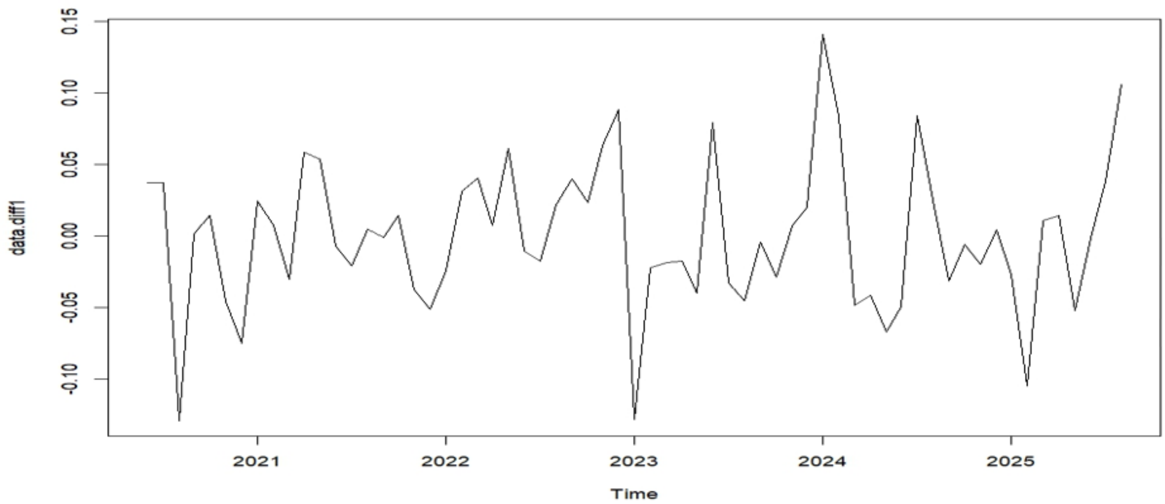

Figure 6: First-difference seasonally adjusted monthly tomato prices in the U.S. from May 2020 to August 2025.

After removing the seasonal component, the series was transformed to further address instability over time. Applying the Augmented Dickey-Fuller (ADF) test to the transformed data produced a statistic of -4.3732, lag order of 3, and a p-value of 0.01. This result indicates that the first-differenced series is stationary, confirming that differencing successfully removed the trend and stabilized the mean, making the series suitable for time series modeling and forecasting.

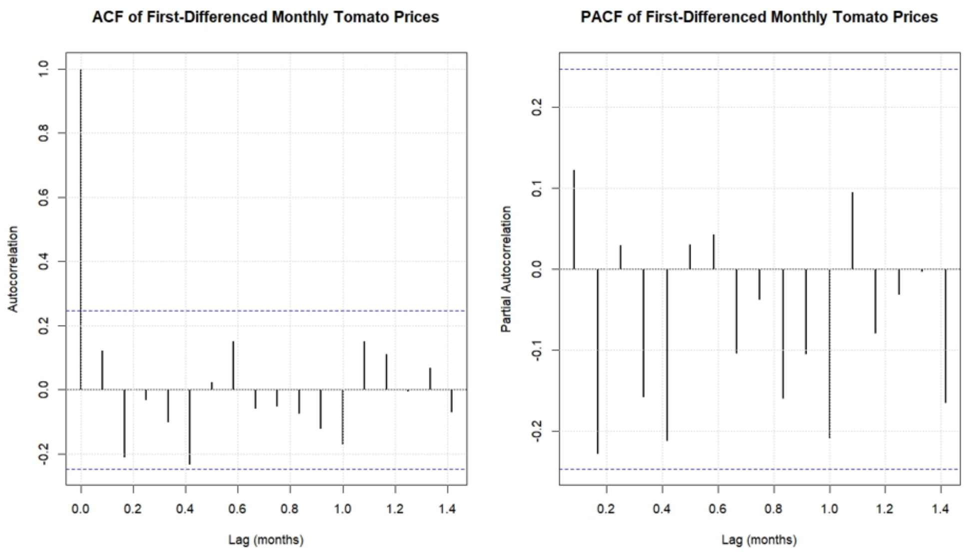

Figure 7: ACF and PACF plots of monthly tomato prices after first-order differencing.

Figure 7 shows the autocorrelation (ACF) and partial autocorrelation (PACF) plots of the first-differenced monthly tomato prices. After differencing, the autocorrelations drop sharply and stay within the confidence limits, indicating that the series has become stationary. The absence of significant spikes in both plots confirms that the non-stationarity present in the original data has been removed. These results suggest that the first-order differencing was sufficient to stabilize the mean and eliminate trends in the tomato price series, making it suitable for fitting an ARIMA model and forecasting.

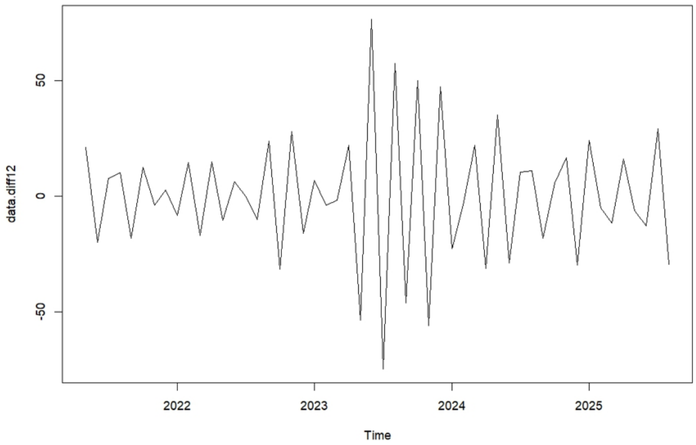

Figure 8: Time series plot of the seasonally adjusted monthly tomato prices after seasonal differencing in the U.S. from May 2020 to August 2025.

Figure 8 shows the seasonally adjusted monthly tomato price after applying the 12th difference. This differencing removes trend and seasonal effects, making the data stable for analysis. The series fluctuates around zero, indicating that the data has become stable after removing seasonal and trend effects. The sharp movements around 2023 suggest short-term market disturbances, while the overall pattern confirms that the series is now stationary and suitable for ARIMA modeling. The Augmented Dickey-Fuller (ADF) test on the transformed series yielded a test statistic of -17.668 with a lag order of 3 and a p-value of 0.01. These results provide strong evidence against the null hypothesis of non-stationarity, confirming that the 12th-differenced series is stationary at the 5% significance level.

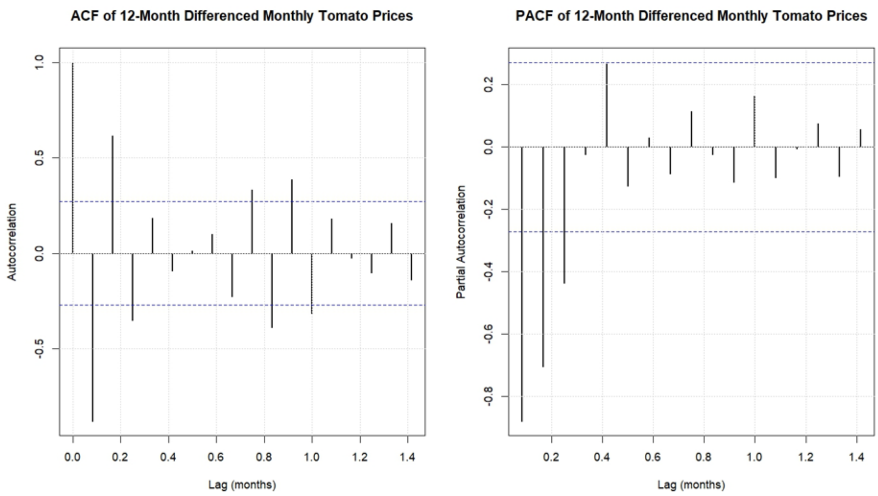

Figure 9: The autocorrelation function (ACF) and partial autocorrelation function (PACF) plots of monthly tomato prices in the U.S. from May 2020 to August 2025 after seasonal differencing.

Figure 9 displays the autocorrelation function (ACF) and partial autocorrelation function (PACF) plots of the seasonally differenced global monthly cotton price series. These plots reveal the autocorrelation structure in the data, guiding the selection of appropriate ARIMA model components. The seasonally differenced series was tested for stationarity with the Augmented Dickey-Fuller (ADF) test. It gave a statistic of -17.668 with a lag order of 3 and a p-value of 0.01. These results indicate that the null hypothesis of non-stationarity cannot be rejected, confirming that the transformed series is stationary at the 5% level.

Estimation of the Seasonal ARIMA Model Parameter

The parameter values generated from the ARIMA (1,0,1)(1,0,0,0) model applied to monthly tomato prices are shown in Table 2. The model includes an autoregressive term (AR1 = 0.6447), indicating that current prices are positively influenced by the prices of the previous month, and a moving average term (MA1 = 0.4303), reflecting the impact of past random shocks on current prices. The seasonal autoregressive term (SAR1 = -0.2366) captures seasonal dependencies at a 12-month lag. The mean of the series is 1.9146. Diagnostics results show a residual variance (Sigma²) of 0.002141 and a log-likelihood of

Table 3: Summary of the parameter estimates for the fitted ARIMA

Parameters			Coefficients		Std. Error	
Differencing			1			
AR1			0.6447		0.1127	
MA1			0.4303		0.1412	
SAR1			-0.2366		0.1355	
mean			1.9146		0.0183	
Sigma2 = 0.002141		log likelihood = 106.98				
AIC = -203.95		AICc = -202.92		BIC = -193.16		
ME	**RMSE**	**MAE**	**MPE**	**MAPE**	**MASE**	**ACF1**
-0.00067	0.0448	0.0344	-0.0877	1.7938	0.3723	0.0041

Note: AR: Auto-Regressive, MA: Moving Average, SAR: seasonal Autoregressive, S.E.: Standard Error, AIC: Akaike Information Criterion, AICc: Corrected Akaike Information Criterion, BIC: Bayesian Information Criterion, ME: Mean Error, RMSE: Root Mean Squared Error, MAE: Mean Absolute Error, MPE: Mean Percentage Error, MAPE: Mean Absolute Percentage Error, MASE: Mean Absolute Scaled Error, and ACF1: Autocorrelation of residuals at lag 1.
Source: Authors' compilation based on data obtained from the Federal Reserve's FRED online archive.

106.98. Model selection criteria suggest a good fit, with AIC = -203.95, AICc = -202.92, and BIC = -193.16. The forecast accuracy metrics show that the models perform very well. Bias is minimal (ME = -0.00067) and low overall errors (RMSE = 0.0448, MAE = 0.0344), with small percentage deviations from actual prices (MPE = -0.0877, MAPE = 1.7938). The MASE value of 0.3723 demonstrates that the model predicts more accurately than a simple naive approach, and the low autocorrelation at lag 1 (ACF1 = 0.0041) suggests that model residuals are nearly uncorrelated.

Applying Equation (5), the model that best fits the seasonally adjusted monthly tomato prices in the U.S. is expressed as:
Together, these outcomes indicate that the model tracks

$$(1-0.6447B)(1+0.2366B12)X_t = (1+0.4303B)\epsilon_t + 1.9146 \quad (6)$$

both immediate variations and seasonal shifts in tomato prices, supporting its usefulness for forecasting.

Diagnostic Checking of the Seasonal ARIMA Model

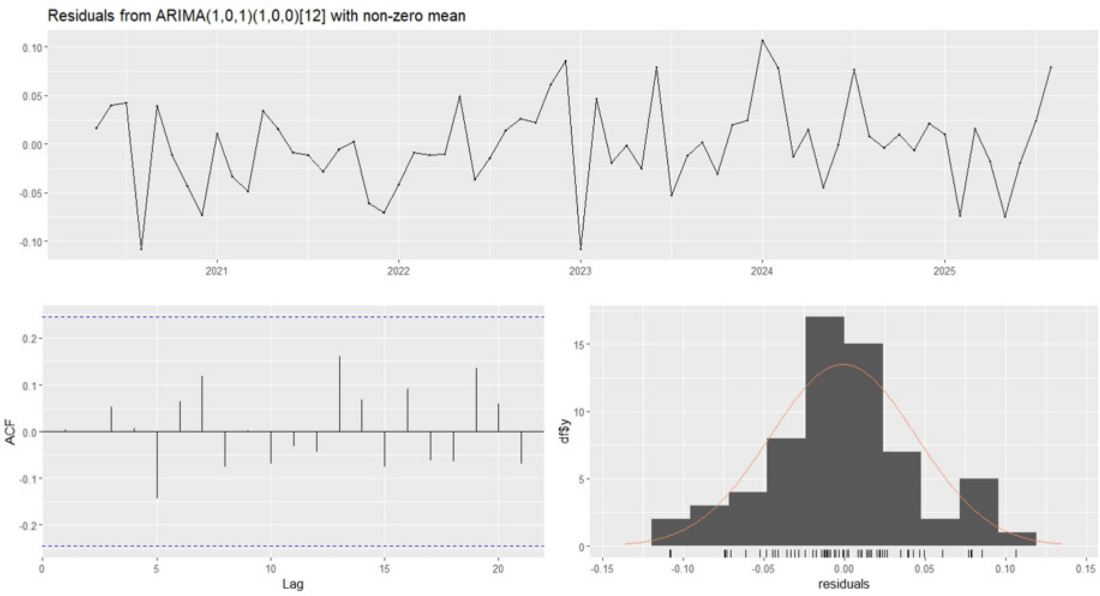

Figure 10: Results of the Ljung-Box test applied to the seasonally adjusted tomato price data in the U.S. (May 2020-August 2025).

Using the Ljung-Box test, Figure 10 displays the result for evaluating the residuals of the ARIMA (1,0,1) (1,0,0) [12] model for seasonally adjusted monthly tomato prices in the United States from May 2020 to August 2025 are free of autocorrelation. With a Q statistic of 6.1621(10 degrees of freedom) and a p-value of 0.8015, the Ljung-Box test shows no evidence of autocorrelation in the residuals. The high p-value indicates that the residuals are independent and random, showing that the model captures the time-dependent patterns.

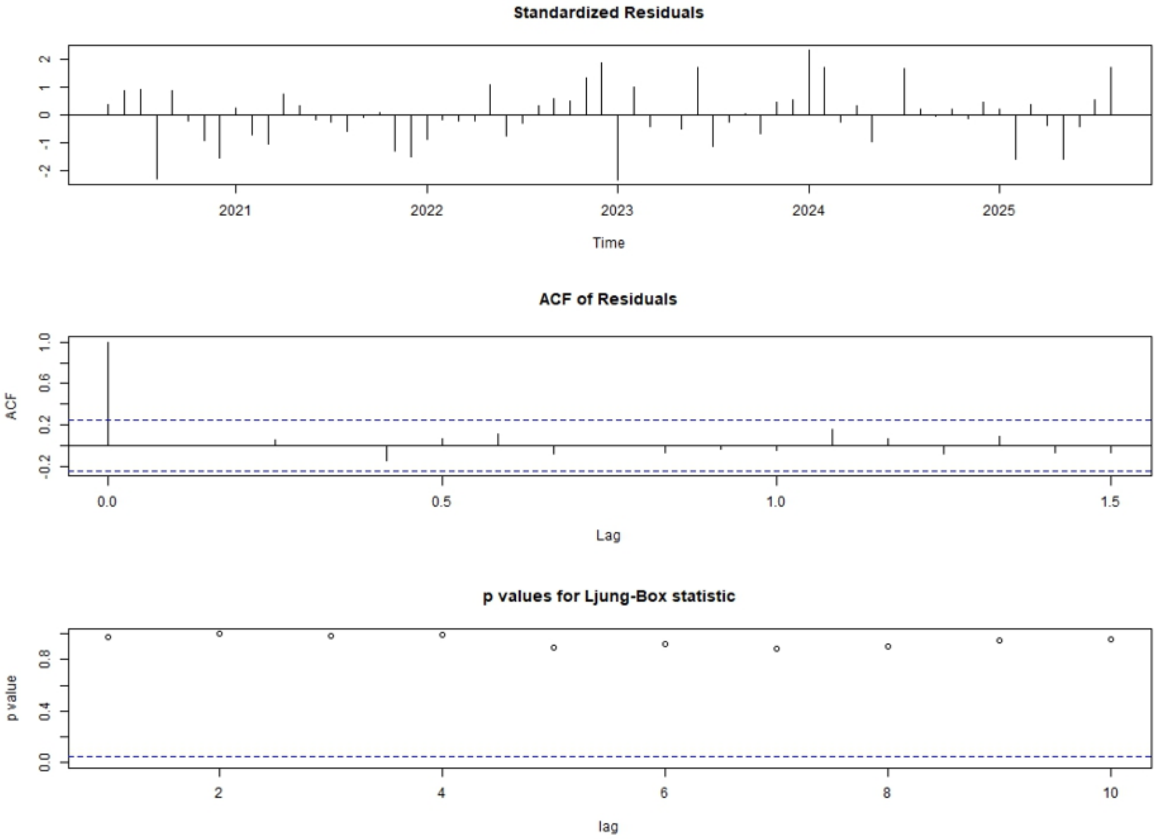

Figure 11: Residual plots of the seasonally adjusted monthly tomato prices in the U.S. (May 2020- August 2025).

In Figure 11, the first panel shows that the standardized residuals vary randomly around zero, indicating that the residuals do not exhibit any obvious pattern and suggesting that the model captures the main trends in the data. As shown in the second panel, the ACF of the residuals has most spikes within the 95% confidence limits, implying that the residuals are largely uncorrelated and resemble white noise. The third panel shows the p-values of the Ljung-Box test, all of which are above the 0.05 significance level, confirming that there is no significant autocorrelation remaining in the residuals. Overall, these diagnostic results confirm that the fitted model is statistically sound and well-suited for forecasting purposes.

D. Using the Seasonal ARIMA Model in Forecasting

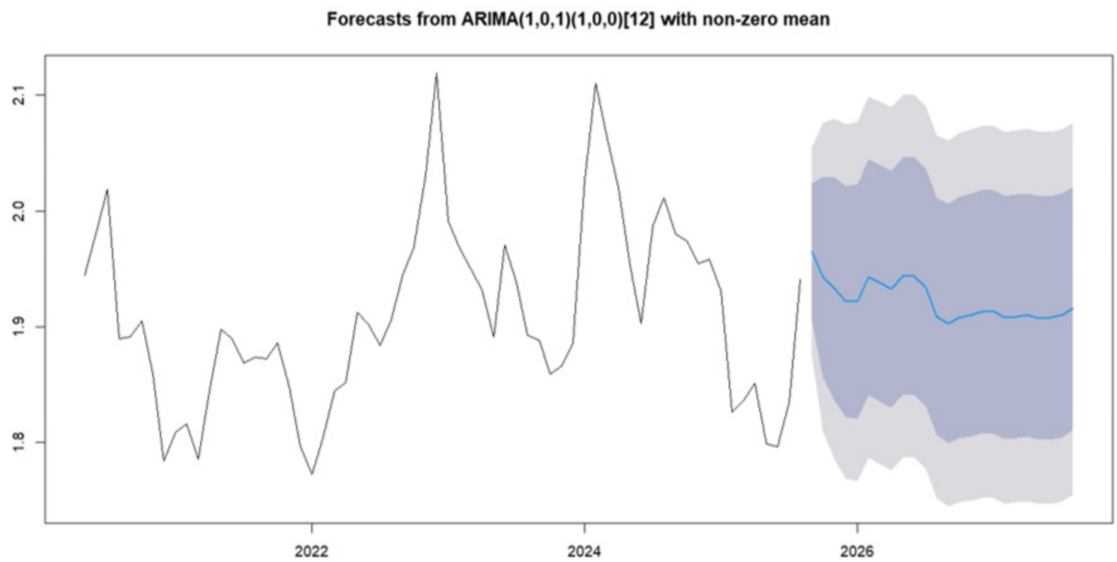

Figure 12: Forecast of monthly tomato prices in the United States using an ARIMA model.

Figure 12 presents the forecast of seasonally adjusted monthly tomato prices in the United States from May 2020 to August 2025, generated using the ARIMA (1,0,1)(1,0,0) [12] model with a non-zero mean. The black line represents the observed prices, while the blue line shows the model's forecast. The shaded areas indicate the 80% and 95% confidence intervals, reflecting the level of uncertainty in the projections. The forecast indicates that tomato prices are likely to remain relatively stable with minor fluctuations over the forecast period. As expected, the confidence interval widens over time, indicating greater uncertainty in longer-term predictions. Overall, the model demonstrates a good fit to the data and offers a reliable short-term outlook for U.S. tomato prices.

Table 4: Forecasted monthly tomato price in the U.S. (September 2025 - July 2027) (U.S. dollars per pound)

Month/Year	Point Forecast	Confidence interval	
		Lower 95%	Upper 95%
September 2025	1.9647	1.874	2.0554
October 2025	1.9428	1.8096	2.0759
November 2025	1.9325	1.7852	2.0797
December 2025	1.9218	1.7691	2.0745
January 2026	1.9221	1.7671	2.077
February 2026	1.9428	1.7869	2.0986
March 2026	1.9377	1.7815	2.094
April 2026	1.9327	1.7763	2.0891
May 2026	1.944	1.7875	2.1004
June 2026	1.9437	1.7872	2.1001
July 2026	1.9342	1.7777	2.0907
August 2026	1.9089	1.7524	2.0654
September 2026	1.903	1.7452	2.0609
October 2026	1.9081	1.7487	2.0676
November 2026	1.9105	1.7503	2.0706
December 2026	1.9129	1.7525	2.0733
January 2027	1.9129	1.7523	2.0734
February 2027	1.9079	1.7474	2.0685
March 2027	1.9091	1.7485	2.0697
April 2027	1.9103	1.7497	2.0709
May 2027	1.9076	1.747	2.0682
June 2027	1.9077	1.7471	2.0683
July 2027	1.9099	1.7493	2.0705

No significant autocorrelation was found in the residuals as confirmed by the Box-Ljung test ($\chi^2 = 0.0011$, df = 1, p = 0.9734), indicating that the SARIMA model effectively represents key patterns in the tomato price series.

Discussion

The selected ARIMA (1,0,1) (1,0,0) [12] model effectively captures both the short-term variations and the recurring annual seasonal pattern in seasonally adjusted monthly tomato prices. These findings are consistent with previous studies showing that Seasonal ARIMA (SARIMA) models are well-suited for agricultural commodity prices that exhibit clear seasonality and autocorrelation structures. For instance, K *et al.* (2024) demonstrated that a SARIMA model provided a good fit and reasonable forecasting accuracy for tomato prices in India, although wider confidence intervals reflected market changes. Likewise, Tamilselvi and Naidu (2020) reported that the SARIMA (2,2,2) (1,0,1) model delivered the most effective in forecasting tomato prices. The forecasted values closely matched the original prices, showing that the model was reliable for short-term forecasting. These forecasts can help policymakers design effective strategies to stabilize tomato markets and support both producers and consumers. In the context of U.S. tomato prices, the strong performance of the ARIMA (1,0,1) (1,0,0) model indicates that the price series is relatively stable, characterized by a consistent 12-month seasonal cycle and residuals that approximate white noise. The Mean Absolute Scaled Error (MASE) of 0.3723 shows that the model's one-step-ahead forecasts are approximately 63% more accurate than those from a naïve seasonal random-walk model, highlighting its strong predictive ability.

From a practical standpoint, accurate monthly price forecasts offer valuable guidance for producers, wholesalers, and supply chain managers. These forecasts help in making informed decisions regarding harvest scheduling, storage, and marketing strategies. The negative seasonal autoregressive coefficient further suggests a mild corrective effect, where higher prices in a given month tend to be followed by slightly lower prices in the same month of the subsequent year. This pattern likely reflects cyclical supply-demand adjustments or recurring planting and harvest schedules. Recognizing this relationship allows stakeholders to anticipate seasonal fluctuations and plan production and marketing activities more effectively. While traditional models like ARIMA work well for stable and seasonal price series, they may struggle to capture sudden shocks and nonlinear trends. In such cases, machine learning approaches offer a promising alternative. For example, Khadka *et al.* (2025) found that the Neural Network Auto Regressive (NNETAR) model outperformed conventional econometric models in forecasting global corn prices, effectively capturing complex temporal patterns and adapting to abrupt market changes.

Future research could investigate hybrid modeling approaches that combine linear and nonlinear methods to improve forecasting accuracy. For example, Ghiyal and Kumar (2025) showed that a hybrid SARIMA+ANN model produced more precise forecasts for monthly wholesale tomato prices compared to a traditional SARIMA model. Adopting such hybrid frameworks could enable policymakers and market participants to make better-informed decisions on procurement, distribution, storage, trade, and inventory management across both domestic and international markets. In addition, advanced deep learning models offer promising opportunities for further improvement. Studies by Farzana and Prakash (2025) have shown that methods like stacked LSTM networks can effectively handle nonlinear and complex temporal relationships in price series. Future studies could benefit from applying these advanced techniques alongside larger, more diverse datasets that incorporate broader market and environmental variables. Applying such methods has the ability to improve reliability, adaptability, and overall performance of forecasting models for agricultural commodities, such as tomatoes.

CONCLUSION

The ARIMA model successfully captured both seasonal patterns and short-term fluctuations in U.S. tomato prices, providing reliable forecasts that are more accurate than simple baseline methods. These findings offer valuable guidance for farmers, wholesalers, and policymakers, helping them make more informed decisions regarding harvest scheduling, storage, and marketing effectively. Future research could use hybrid or deep learning models to further improve predictive accuracy, particularly when using larger and more detailed datasets. Overall, this study highlights the value of data-driven approaches that can support more informed decisions and mitigate risks within the tomato supply chain.

REFERENCES

Adanacıoğlu, H., & Yercan, M. (2012). An analysis of tomato prices at the wholesale level in Turkey: An application of the SARIMA model. *Custos e @gronegócio online, 8,* 52-75.

Badal, P. S., Kamalvanshi, V., Goyal, A., Kumar, P., & Mondal, B. (2022). Forecasting potato prices: Application of the ARIMA model. *Economic Affairs, 67*(4), 491-496. https://doi.org/10.13140/RG.2.2.12974.48966

Bernard, B., Francois, L., & Renville, D. S. (2025). Forecasting the international market prices for rice, corn, and soybeans using ARIMA time series modelling. *International Journal of Agricultural Economics, 10*(4), 170-182. https://doi.org/10.11648/J.IJAE.20251004.13

Box, G. E. P., & Jenkins, G. M. (1976). *Time series analysis: Forecasting and control.* San Francisco: Holden-Day.

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

Brockwell, P. J., & Davis, R. A. (2002). I*ntroduction to time series and forecasting* (2nd ed.). Springer.

Cui, X., Guan, Z., Morgan, K. L., Huang, K. M., & Hammami, A. M. (2022). Multitiered fresh produce supply chain: The case of tomatoes. *Horticulturae, 8*(12), 1204. https://doi.org/10.3390/horticulturae8121204

FAOSTAT. (2023). Crops and livestock products: Tomatoes (production quantity). Food and Agriculture Organization of the United Nations. https://www.fao.org/faostat/en/#data/QCL

Food and Agriculture Organization of the United Nations (2025) - with major processing by Our World in Data. "Tomato production - UN FAO" [dataset]. Food and Agriculture Organization of the United Nations, "Production: Crops and livestock products" [original data]. Retrieved October 26, 2025, from https://archive.ourworldindata.org/20251014-161652/grapher/tomato-production.html (archived on October 14, 2025).

Ghiyal, P., & Kumar, J. (2025). Forecasting of monthly average wholesale prices of tomato: An application of hybrid time series models.

Gujarati, D. N. (2003). *Basic econometrics* (4th ed.). McGraw-Hill.

Guzma, L. (2024). Application of the ARIMA model to forecast corn prices in Mexico. *Scientia et PRAXIS, 4*(08), 63-95. https://doi.org/10.55965/setp.4.08.a3

Hyndman, R. J. (2001). Box-Jenkins modelling.

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice* (2nd ed.). OTexts.

Jadhav, V., Chinnappa, R. B., & Gaddi, G. M. (2017). Application of the ARIMA model for forecasting agricultural prices.

K, L. P., Suresh, P., Yeturu, N., Bhargavi, N., Mathew, M., J, P. M., & A. N. (2024). Price behaviour and price forecasting of tomato in Koyambedu market, Tamil Nadu. *International Journal of Religion, 5(*10), 2954-2962.

https://doi.org/10.61707/XPERM644

Khadka, R., Sanwa, R. L., Khadka, N., & Chi, Y. N. (2025). Forecasting the global price of corn using a neural network and hybrid approach. *IEEE Access, 13,* 167424-167438. https://doi.org/10.1109/ACCESS.2025.3611320

Manogna, R. L., Dharmaji, V., & Sarang, S. (2025). Enhancing agricultural commodity price forecasting with deep learning. *Scientific Reports, 15*(1), 1-24. https://doi.org/10.1038/s41598-025-05103-z

Mathenge Mutwiri, R. (2019). Forecasting of tomatoes wholesale prices of Nairobi in Kenya: Time series analysis using the SARIMA model. *International Journal of Statistical Distributions and Applications, 5*(3), 46. https://doi.org/10.11648/J.IJSD.20190503.11

Prakash, N., & Farzana, G. S. (2020). Short-term price forecasting of horticultural crops using long short-term memory neural network. *Learning and Analytics in Intelligent Systems, 16,* 111–118. https://doi.org/10.1007/978-3-030-46943-6_12

Prathilothamai, M., Vinay, K., Vishnu Vardhan, K., Rama Vamsidhar Reddy, G., & Nithin, U. (2025). Enhancing agricultural price forecasting with time series models: A case study on tomato markets. In S. J. Nanda (Eds.), *Data science and applications: ICDSA 2024* (Lecture Notes in Networks and Systems, Vol. 1237). Springer. https://doi.org/10.1007/978-981-96-1185-0_16

Puria, P., Singh, K. M., & Parihar, L. (2023). Decomposition and forecasting of tomato price in Bhopal district, Madhya Pradesh. *Agricultural Economics Research Review, 36,* 111-122.

Purohit, S. K., Panigrahi, S., Sethy, P. K., & Behera, S. K. (2021). Time series forecasting of price of agricultural products using hybrid methods. *Applied Artificial Intelligence, 35*(15), 1388-1406. https://doi.org/10.1080/08839514.2021.1981659

Quinet, M., Angosto, T., Yuste-Lisbona, F. J., Blanchard-Gros, R., Bigot, S., Martinez, J. P., & Lutts, S. (2019). Tomato fruit development and metabolism. *Frontiers in Plant Science, 10,* 1554. https://doi.org/10.3389/fpls.2019.01554

Reddy, A. A. (2018). Price forecasting of tomatoes. *International Journal of Vegetable Science, 25(*2), 176-184. https://doi.org/10.1080/19315260.2018.1495674

Rubillos, L. E., Mangubat, A. F., Seno, M. A., Caragay, J. C., Salazar, N., Caladcad, J. A., & Villeta, R. (2024). Time series analysis of wheat moisture content variations for grain storage systems. *Journal of Stored Products Research, 108,* 102395. https://doi.org/10.1016/j.jspr.2024.102395

Sahoo, M. (2020). Agricultural product price forecasting using the ARIMA model. *International Journal, 8*(5), 5203-5207.

Limbu Sanwa, R., & Nain Chi, Y. (2025). Global monthly cotton price trends: Insights from ARIMA time-series forecasting. *International Journal of Business & Economics, 10*(1), 104–124. https://doi.org/10.58885/ijbe.v10i1.104.rls

Slutsky, E. (1937). The summation of random causes as the source of cyclic processes. *Econometrica, 5*(2), 105-146.

Sumaiya Farzana, G., & Prakash, N. (2025). Tomato price forecasting using a stacked LSTM network with global scaling CNN and RC layer. *International Journal of Advanced Technology and Engineering Exploration, 12,* 2394-7454. https://doi.org/10.19101/IJATEE.2024.111100298

Sunanda, N., Elakkiya Nagappan, S. B., Ashok, J., & Padmaja, B. (2024). A statistical study on tomato prices in Andhra Pradesh.

Tamilselvi, C., & Naidu, G. M. (2020). Forecasting vegetable arrivals and wholesale prices at Koyambedu market in Chennai using the SARIMA approach. *International Journal of Current Microbiology and Applied Sciences, 9*(10), 3059–3068. https://doi.org/10.20546/ijcmas.2020.910.368

Tammali, R., & Mishra, P. K. (2024). Forecasting agricultural commodity prices with ARIMA models. In *Responsible production and consumption* (pp. 58-65). CRC Press.

Tebbs, J. M. (2013). STAT 5*20: Forecasting and time series.* University of South Carolina, Department of Statistics, 80, 306.

Yule, G. U. (1927). On a method of investigating periodicities in disturbed series. *Philosophical Transactions of the Royal Society of London. Series A, 226*, 267-298.