Eastern Africa Journal of
Contemporary Research (EAJCR)

The Power of First Differencing in Addressing Non-Stationarity Time Series: Empirical Evidence from the Price of Rice using SARIMA Model in Tanzania

Bahati Ilembo and Rahimu Kassim

Article information:

To cite this article:

Ilembo, B., & Kassim, R. (2025). The power of first differencing in addressing non-stationarity time series: Empirical evidence from the price of rice using SARIMA model in Tanzania. *Eastern Africa Journal of Contemporary Research, 2025*(1), 53–69.

For Authors

About Eastern Africa Journal of Contemporary Research (EAJCR)

The Eastern Africa Journal of Contemporary Research (EAJCR) is both an online (ISSN: 2663-7367) and print (ISSN: 2663-7359) double-blind peer-reviewed quarterly journal published by the Directorate of Research and Publications of Gretsa University, Kenya.

EAJCR aims at advancing the frontiers of knowledge by publishing contemporary multidisciplinary conceptual/ theoretical and empirical research articles as well as case studies and book reviews.

Vol. 5, Issue 1, 2025, ISSN: 2663-7367 (Online) & ISSN: 2663-7359

The Power of First Differencing in Addressing Non-Stationarity Time Series: Empirical Evidence from the Price of Rice using SARIMA Model in Tanzania

Bahati Ilembo
Department of Mathematics and Statistics Studies
Mzumbe University, Tanzania
Email: bmilembo@mzumbe.ac.tz

Rahimu Kassim
Department of Mathematics and Statistics Studies
Mzumbe University, Tanzania
Email: rahkas06@gmail.com

Abstract

The main purpose of this study was to forecast the price of rice in Tanzania using a secondary univariate time series from January 2005 to September 2024. We used the seasonal ARIMA (3, 1, 2) (2, 0, 0) [12] model with seasonal data. The analysis revealed that the price of rice in Tanzania will exhibit a consistent upward trend from October 2024 to a peak in March 2027 reflecting a total increase of approximately 121.4% over this period. From April 2027, prices will begin to decline steadily reaching low in August 2029, a reduction of 44.3% from the March 2027 peak. Prices typically dip in late summer, with the sharpest decline observed in August 2025, when prices fall by 32.8% from June 2025. The decline beginning in April 2027 reflects a significant market shift, with an average monthly decrease of approximately 2.1% from April 2027 to August 2029. This may be attributed to increased production or effective market interventions. The policy issues should focus on the areas of price stabilization programmes, strengthening trade and export policies putting in place mechanisms to support farmers by giving them access to agro inputs and ensuring the technical services are availed to farmers through agricultural extension officers. This will increase productivity and address the problem of demand which will in turn solve the entire problem of price volatility. This shall be achieved by influencing consumer behaviour and moderating demand, the interventions that help reduce the volatility that arises from sudden spikes in demand, contributing to more stable pricing over time.

Keywords: *ARIMA, Seasonal, Oryza sativa, SARIMA*

1. Introduction

Rice (Oryza sativa) is a significant staple food and plays a paramount role in global food security as well as the primary source of carbohydrates for the human population (Septiana, 2024; Aulia & Platama, 2024; Nkuba et. al, 2016; Hassan et. al, 2011). Many countries produce rice which has been used entirely as human food (Garov & Awasthi, 2023; Nkuba et.al, 2016; FAO, 2015; Abadan & Shabri, 2014). In Tanzania, rice is the third most important food crop after maize and cassava (FAO, 2015). According to Sekiya et.al (2020), rice consumption in Tanzania has greatly increased since the 1960s and it is predicted to continue to increase owing to urbanization and changes in consumer preferences which calls for the importance of multidisciplinary approaches in improving rice production in Tanzania. In the intricate landscape of agricultural economics, forecasting prices of staple commodities like rice hold significant implications in decision-making and planning for the government or communities like farmers, traders and consumers (Septiana, 2024). Price prediction is more acute with rice crops particularly due to its seasonality. Prediction of rice prices can provide critical and useful information to rice growers in making production and marketing decisions (Kundu & Sharma, 2022). It is further emphasized that forecasting rice prices is an action to help the government monitor and control it. Price monitoring is conducted to achieve the good development of domestic trade. In this case, price monitoring is done to maintain price stability so that it will not harm producers and consumers (Ohyver & Pudjihastuti, 2018).

Historical literature indicates that the prices of rice keep changing from time to time thereby creating risks to producers, suppliers, consumers, and other parties involved in the marketing and production of rice. Rice prices often surge due to uncontrolled demand, requiring government monitoring to maintain its stability. For that reason, forecasting on rice prices becomes significant (Aulia & Platama, 2024; Abadan & Shabri, 2014).

Understanding the dynamics of cereal prices is valuable for many economic agents. For the agriculture sector, the sale of cereals is one of the main sources of revenue. According to Kwas, Paccagnini and Rubaszek (2021), data from the report of the OECD and FAO (2020) indicate that in 2019 the combined global value of wheat, maize and rice production amounted to over USD 500bn. For many other industries, cereals are an important input in the production processes and for the selected countries, cereals constitute a significant fraction of exports, hence their prices influence key macroeconomic variables, for instance, the current account balance, terms of trade, and exchange rates.

Different forecasting methods have been used to forecast rice prices and serve various purposes. For example, The autoregressive integrated moving average (ARIMA) dominated the literature (Mubarak & Sundara, 2024; Ramadhani, Sukiyono & Suryanty, 2020; Ohyver & Pudjihastuti, 2018; Garov & Awasthi, 2023). However, other models such as autoregressive fractionally integrated moving average (ARFIMA) were found to outperform the ARIMA (Mitra & Paul, 2021), especially when attempting to capture the observed long-run persistency. In some instances, the Holt-Winters exponential smoothing model was used to predict rice prices (Septiana, 2024), however, the Holt-Winters is capable of handling single seasonality effectively and works well with simple linear or exponential trends. This paper used the potential of the seasonal ARIMA (SARIMA) to forecast rice prices due to its ability to handle complex seasonal patterns (such as multiple seasonal periods) and its ability to model complex trends, both seasonal and non-seasonal. The paper is organized as follows; section 1 carries the introduction, section 2 is on the Materials and Methods, and section 3 is on the results and discussion. The paper concludes with section 4 and provides the specific policy implications in section 5.

2. Methodology

2.1 Data type and source

The study used monthly seasonal time series data for the period January 2005 to September 2024. These are domestic prices (TZS/100kgs) available at: https://www.bot.go.tz/Publications/Economic%20Statistics/Monthly/en/2025010915173382.xlsx. The study employed a quantitative method of research to test the collected data, analyze and to build a forecast of the price of rice. As said earlier, rice is among the top three staple foods in Tanzania and thus, price forecasting plays an important role in efficient planning and formulation of executive decisions.

2.2 The conceptual framework

The seasonal Autoregressive Moving Average alias SARIMA is an extension of the Autoregressive Moving Average (ARIMA) with seasonality in addition to the non-seasonal components. While the ARIMA models are used in time series analysis and forecasting, the SARIMA models are precisely designed to manage the time series data with seasonal patterns. The nature of the time series data can be recorded daily, weekly, monthly, quarterly, semiannually or annually all of which depend on whether it is non-seasonal or seasonal. When data are seasonal, ARIMA models become appropriate and it is given by:

ARIMA $(p, d, q) \qquad (P, D. Q)_m$

whereby; the (p, d, q) is the non-seasonal part and the $(P, D. Q)_m$ is the seasonal part of the model. The other terms in the models are defined as follows:

$m = number\ of\ observations\ per\ year$

$P = number\ of\ seasonal\ AR\ terms$

$D = number\ of\ seasonal\ differences$

$Q = number\ of\ seasonal\ MA\ terms$

The upper case notations and lower case ones represent seasonal and non-seasonal parts of the model.

The seasonal component of SARIMA models adds the following three components:

i. **Seasonal Autoregressive (P)**: This component captures the relationship between the current value of the series and its past values, specifically at seasonal lags.

ii. **Seasonal Integrated (D)**: Similar to the non-seasonal differencing, this component accounts for the differencing required to remove seasonality from the series.

iii. **Seasonal Moving Average (Q)**: This component models the dependency between the current value and the residual errors of the previous predictions at seasonal lags.

ARIMA (1,0,1)(2,1,0)12 models, for example, is mathematically expressed as;

$$y_t - y_{t-12} = W_t$$

$$W_t = \mu + \vartheta_1 W_{t-1} + \vartheta_2 W_{t-12} + \vartheta_3 W_{t-24} + \alpha_1 \varepsilon_{t-1} + \varepsilon_t$$

This paper used the strengths of the Box- Jenkins methodology in forecasting, the process which was described by statisticians George Box and Gwilym Jenkins back in 1970. The methodology involves three common steps namely identification of an appropriate model process, fitting it to data and then using the fitted model for forecasting. The Box and Jenkins (1970) have generalized the ARIMA model to deal with seasonality and define a general multiplicative seasonal model in the form:

The procedure consists of fitting a mixed autoregressive integrated moving average (ARIMA) model to a given set of time series data and then taking conditional expectations. The main stages in setting up a Box-Jenkins forecasting model are as outlined below:

Model identification: Examine the data to reveal which member of the class of ARIMA processes appears to be the most appropriate.

Estimation: Estimate the parameters of the chosen model by least squares

Diagnostic checking: Examine the residuals from the fitted model to see if it is adequate

Alternative model consideration: Consider an alternative model if the first model appears to be inadequate for some reason, then other ARIMA models may be tried until a satisfactory model is found. The ARIMA models are powerful as they consider the Stationarity and are well suited for linear time series, unlike machine learning methods which do not require the series to be stationary.

2.2.1 Test for stationarity of time series

In testing for stationarity, the Augmented Dickey fuller (ADF) test has been used. The method is a convenient and effective tool for testing the stationarity of a time series. It is also simple and able to account for trends and autocorrelation and has been commonly applied in the field of time series forecasting and analysis in testing the null hypothesis that there is a unit root or time series data is not stationary (Dickey & Fuller, 1979). Rejecting the null hypothesis implies that there is strong evidence against the presence of a unit root, and therefore the series is stationary. If the null hypothesis is rejected, the practical implication is that a stationary series has constant mean, variance, and autocorrelation over time, which makes it suitable for time series modelling methods like ARIMA (AutoRegressive Integrated Moving Average). It also implies that one can proceed with modelling the series without needing to transform it (e.g., differencing) to make it stationary. Failing to reject the null hypothesis, suggests that there is insufficient evidence to conclude that the series is stationary. This means the series likely has a unit root and is non-stationary. The implication is that a non-stationary series often has trends or seasonality, and it may need to be transformed to make it stationary before modelling. Non-stationarity can lead to unreliable and misleading results in time series forecasting models. Therefore, further steps are needed to transform the data into a stationary form before applying models like ARIMA.

According to the unit-root-based test associated with the first lag of the time series variable, if the coefficient ($\gamma = 1$) has a unit root, the time series behaves similarly to the random walk model which is non-stationary and if the coefficient $|\gamma| < 1$ then, there is no unit root. Hence, we can test statistically whether the coefficient (γ) is equal to one or not. The Dickey-Fuller test adopts this procedure by carefully manipulating the equation, given as:

$$z_t = \alpha + \beta t + \vartheta\, z_{t-1} + e_t \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 1$$

Also, written as

$$\Delta z_t = z_t - z_{t-1} = \alpha + \beta t + \gamma\, z_{t-1} + e_t \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots .2$$

In the Dickey-Fuller test, we test the hypothesis:

$H_0: \vartheta = 1\ (called\ Null\ hypothesis)$ against $H_1: \vartheta\ \neq 1\ (called\ alternative\ hypothesis)$

Correlograms

In analyzing the relationship between a time series and its lagged values, the Autocorrelation Function (ACF) and the Partial Autocorrelation Function (PACF) are used as prominent statistical measures. When these are plotted, they yield very important results, especially in determining the values of p and q for Autoregressive (AR) and Moving Average (MA) models.

Autocorrelation Function (ACF)

The autocorrelation at lag 0 is always 1 because a series is always perfectly correlated with itself. The ACF values for lags greater than 0 show how correlated the series is with its previous observations, at different time intervals (lags). If the ACF shows significant spikes at specific lags (outside the confidence interval), it suggests there is a relationship between the series and the corresponding lags. These lags are important for model building. If the ACF gradually decays toward zero, it suggests that the series might follow a Moving Average (MA) process. If the ACF cuts off abruptly after a certain number of lags (e.g., significant at lag 1, but not at lag 2 and beyond), it suggests an AR (AutoRegressive) process, where the series is influenced by only a few previous observations. In practice, for a stationary time series (constant mean and variance over time), the ACF should decrease relatively quickly to zero as the lag increases. For a non-stationary series, the ACF may show slow decay, or it may remain significant at many lags, indicating trends or seasonality in the data.

In determining the association between a time series and its lagged values, the ACF is used to also assess how much the current value of a time series depends on its past values. Autocorrelation is fundamental in time series analysis, helping identify patterns and dependencies within the data. The statistical representation of the correlation between the current observation (z_t) and the previous observation (z_{t-k}) is given as:

$$\rho_i = \text{Corr}\,(z_t, z_{t-i}) = \frac{\text{Cov}\,(z_t, z_{t-i})}{\sqrt{\text{Var}\,(z_t)\cdot\,\text{Var}(\,z_{t-i})}} = \frac{\gamma_i}{\gamma_0} \quad \dots\dots\dots\dots\dots 3$$

Where, $i = 1,2,3, \dots$

Partial Autocorrelation Function (PACF)

Unlike Autocorrelation, partial Autocorrelation focuses on the direct correlation at each lag while removing the influence of intermediate lags to provide a clearer picture of the direct relationship between a variable and its past values. The PACF shows how much of the correlation at a particular lag remains after accounting for the effect of the previous lags. For example, if you have a lag of 3 and the PACF at lag 3 is high, it means that the value at time t is significantly correlated with the value at time $t-3$, after removing the influence of the values at $t-1$ and $t-2$. The partial Autocorrelation function at lag j for time series is given as:

$$\vartheta_{11} = \text{Corr}\,(Y_{t+1}, Y_t\,) = \rho_1$$

$$\vartheta_{jj} = \text{Corr}\,(Y_{t+j} - \hat{Y}_{t+j}, Y_t - \hat{Y}_t),\ \ j \geq 2 \dots\dots\dots\dots\dots\dots\dots\dots\dots .4$$

The appropriate values of p and q will be selected by observing the autocorrelation function (ACF) and partial autocorrelation function (PACF) of the time series data. As suggested by Hyndman and Athanasopoulos (2018), the appropriate ARIMA models will be selected by observing the behaviour of ACF and PACF spikes based on the order identified.

Seasonality Test

Seasonality can be detected by visually examining the seasonal component for regular patterns that repeat at a fixed interval. If the seasonal component exhibits regular patterns, this indicates the presence of seasonality in the time series. The method has been to conduct seasonal decomposition using an additive model which decomposes time series into its trend, seasonal, cyclical and regular components by assuming that, the time series can be modelled by using those components.

2.3 Model Estimation

The parameters of the selected SARIMA model with the specific values of (p, d, q) × (P, D, Q) $_S$ need to be estimated. The maximum likelihood estimation (MLE) estimates the coefficients of the suggested models at the identification stage. The selection of the best model was based on Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC).

2.3.1 Akaike Information Criterion (AIC)

Akaike Information Criterion values indicate a better-fitting model, as AIC penalizes for the complexity of the model while rewarding goodness of fit. However, it's important to acknowledge that models with lower AIC values may not always be the best choice, as they can sometimes overfit the data. Overfitting occurs when a model captures noise or random fluctuations in the data rather than the underlying pattern, leading to poor generalization on new, unseen data. Kullback *et al*., (1951) developed a measure to capture the information that is lost when approximating reality. Kullback & Leibler measure is a criterion for a good model that minimizes the loss of information. Two decades later, Akaike established a relationship between the Kullback-Leibler measure and the maximum likelihood estimation (MLE) method that was used in many statistical analyses for model selection (Akaike, 1974). This criterion referred to as Akaike Information Criterion (AIC), is generally considered the first model selection criterion that should be used in practice. The AIC is given as:

$$\text{AIC} = -2\log\text{L}\ (\hat{\theta}) + 2k \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots. 5$$

Where; θ is the set of model parameters, L $(\hat{\theta})$ is the likelihood of the candidate model given the data when evaluated at the maximum likelihood estimate of θ and k is the number of estimated parameters in the candidate model.

Since AIC does not consider the effect of sample size, for small sample sizes, the second-order equation of the Akaike information criterion (AIC_C) is defined as:

$$\text{AICc} = -2\log\text{L}\left(\hat{\theta}\right) + 2\text{k} + \frac{(2\text{k} + 1)}{(\text{n} - \text{k} - 1)} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots..6$$

where n denotes the total number of observations.

A small sample size is when $n/_k < 40$, also that when the number of observations increases, the third term in AIC_C approaches zero and will therefore give the same result as AIC in equation 6

2.3.2 Bayesian information criterion (BIC)

Bayesian information criterion is another model selection criterion based on information theory but set within a Bayesian context. The difference between the BIC and AIC is the greater penalty imposed for the number of parameters

$$\text{BIC} = -2\log L(\hat{\theta}) + k\log n \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 7$$

where n denotes the total number of observations.

The BIC strongly penalizes the number of involved parameters. High values of AIC mean that the observed data does fit the model, while lower values indicate strong evidence that the observed data fit the models. Similarly, lower values of BIC indicate better fitting of the models.

2.3.3 Diagnostic Checking

Diagnostic check evaluating the acceptance of the fitted SARIMA model by examining the residuals, which are the difference between observed and predicted values. The aim is to ensure the residuals are random and do not contain any patterns or structures. The diagnostic checks involved the use of the Ljung-Box test and the forecasting and forecasting accuracy.

- *Ljung-Box Test*

The Ljung-Box test helps to check whether the errors or residuals in our model have any pattern or correlation. The Ljung-Box test is defined as;

H_0 = Residuals are independently distributed, correlation in the population from which the sample is taken is 0

H_1 = Residuals are not independently distributed; they exhibit serial correlation.

The Ljung-Box test statistic is given as;

$$Q = n(n+2)\sum_{k=1}^{h}\frac{\hat{\rho}_k^2}{n-k} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 8$$

Where n is the sample size, $\hat{\rho}_k^2$ is the sample autocorrelation at lag k, and h is the number of lags being tested. Under the null hypothesis, the test statistic Q asymptotically follows a $\mathcal{X}_{(h)}^2$, for the significant level α, critical region is rejected if;

$$Q > \mathcal{X}_{(1-\alpha),h}^2$$

Where $\mathcal{X}_{(1-\alpha),h}^2$ is the (1−α) quantile of Chi-square distribution with h degrees of freedom.

After diagnostic checking, the fitted model will used in forecasting future values if the model is adequate. Otherwise, we need to repeat the selection and estimation method. Try with another potential candidate model (Ramasubramanian, 2007).

- *Forecasting and Forecasting Accuracy*

Once the selected model has been verified, the model will be then used to predict the price of rice for the prescribed future time. After the forecasting, this study employs the Mean Percentage Error (MAE) and Mean Absolute Percentage Error (MAPE) to evaluate the forecasting accuracy of the selected model. The MAE and MAPE are given by:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|Z_i - \widehat{Z}_\iota\right| \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots 9$$

and

$$MAPE = \frac{100}{n}\sum\left|\frac{(p_i - o_j}{o_j}\right| \qquad \ldots\ldots\ldots\ldots\ldots .10$$

Where p_i is the predicted value for the i^{th} observations, o_i is the observed value for the j^{th} observation, n is the number of non-missing residuals.

3. Results and Discussion

3.1 Time Series Plot for the price of rice

The figure below shows the fluctuation of rice prices from January 2005 to September 2024. In the beginning, the price was increasing and reached its highest peak in April 2012 and started decreasing and reached its lowest peak between Jan and February 2021. The fluctuations then continued to go up for quite some time before reaching the highest peak of Rice Price in the history of the country in April 2023 and decreased abruptly thereafter. It is evident from the graph that, the generation tends to increase and decrease around the centre similarly repeatedly. This pattern of movement demonstrates that the series appeared to be non-stationary.

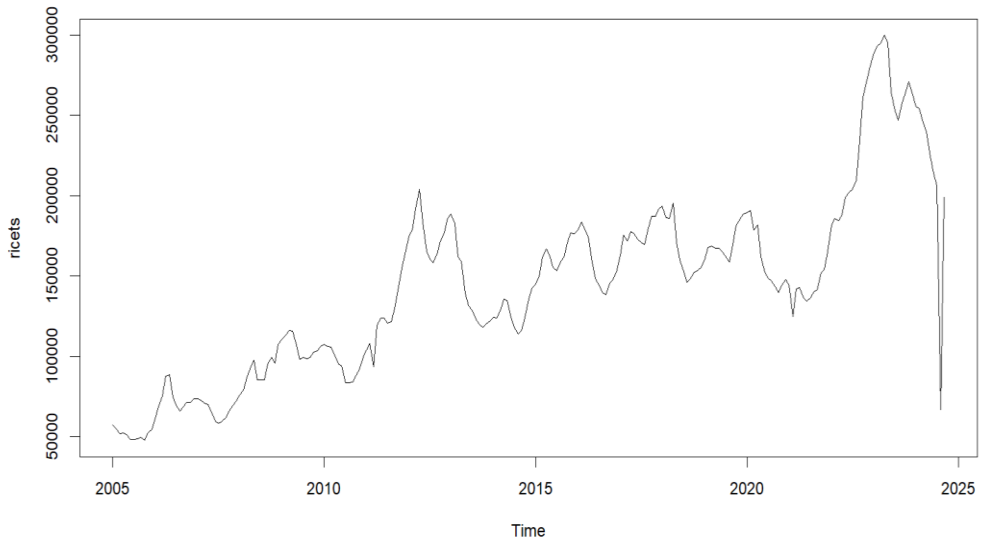

Figure 1.: Time series plot of price of rice in Tanzania.
Source(s): Created by authors

3.2 Testing for Stationarity: Augmented Dickey-Fuller (ADF) test

The figure below shows the fluctuation of rice prices from January 2005 to September 2024. In the beginning, the price was increasing and reached its highest peak in April 2012 and started decreasing and reached its lowest peak between Jan and February 2021. The fluctuations then continued to go up for quite some time before reaching the highest peak of Rice Price in the history of the country in April 2023 and decreased abruptly thereafter. It is evident from the graph that, the generation tends to increase and decrease around the centre similarly repeatedly. This pattern of movement demonstrates that the series appeared to be non-stationary.

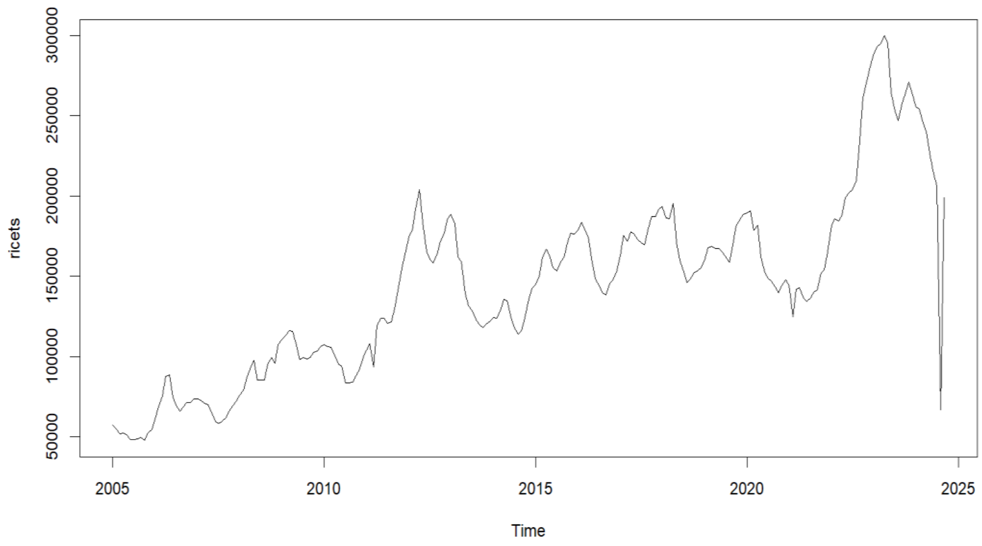

Figure 1.: Time series plot of price of rice in Tanzania.
Source(s): Created by authors

3.2 Testing for Stationarity: Augmented Dickey-Fuller (ADF) test

The results in Table 1 show that, the p-value of the test statistic in the ADF exceeds 5% (0.05) significance level, which suggests that we fail to reject the null hypothesis. This implies that the data contains a unit root hence the series is non-stationary. Table 2 shows the ADF test after performing the first differencing and that the results were significant to warrant rejecting the null hypothesis that data were not stationary.

Table 1: ADF Test before differencing

Augmented Dickey-Fuller (ADF) Test		
Dickey-Fuller = -3.1157	Lag order = 6	P Value = 0.1068

Source(s): Created by authors

Table 2: ADF Test after first differencing

Augmented Dickey-Fuller (ADF) Test		
Dickey-Fuller = -5.1012	Lag order = 6	P Value = 0.01

Source(s): Created by authors

3.3 Seasonality Test

Detecting Stationarity is paramount to identify the appropriate model which fits the data. Seasonal decomposition using an additive model serves this purpose as indicated in figure 2 below. It can be shown from the graph that the seasonal component displays a regular pattern which implies that there exists seasonality in time series rice price data.

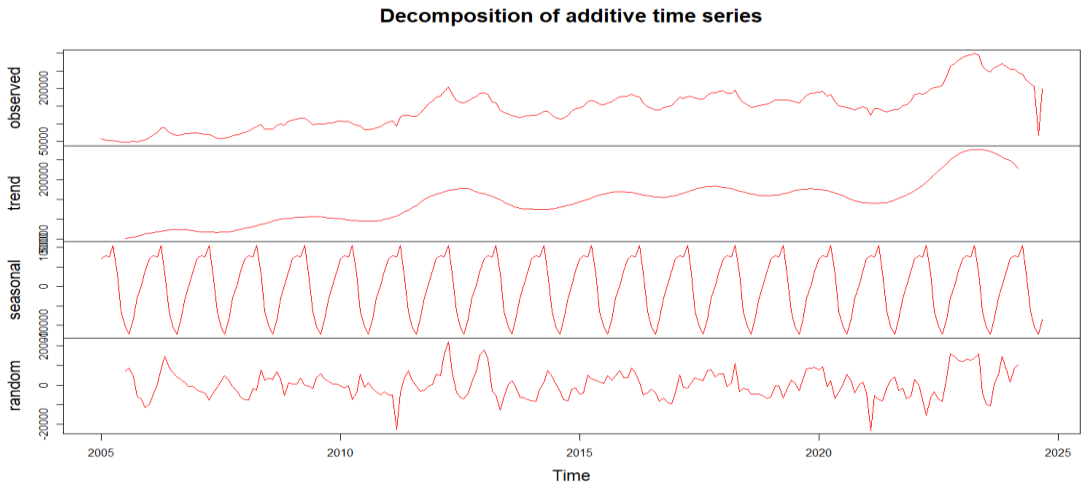

Figure 2: Decomposition of additive time series plot of the price of rice in Tanzania.
Source(s): Created by authors

3.4 Diagnostic checks

In doing the diagnosis checks, the study plotted the time plot, ACF and histogram for the residuals together with the Ljung-Box test. The idea was to see if the data fit the model under examination. The results of the plots and the Ljung-Box test are shown in Figure 3 and Table 3 below.

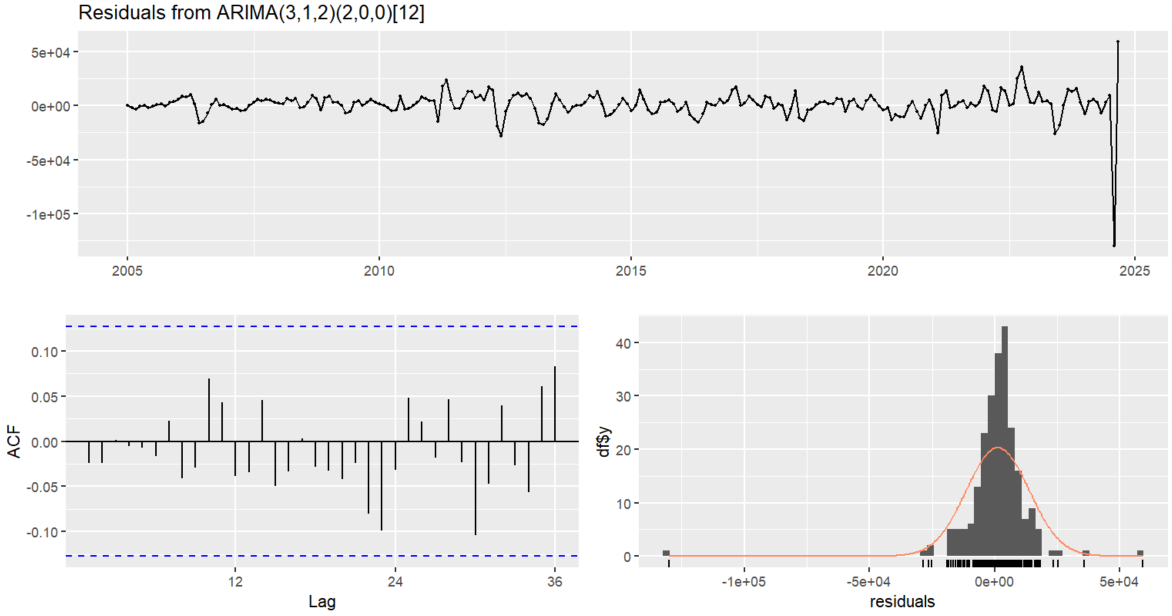

Figure 3: Time plot, ACF and Histogram for the residuals.
Source(s): Created by authors

Figure 3 shows the time plot, ACF plot, and the histogram of the residuals of the ARIMA Seasonal ARIMA (3,1,2) (2,0,0) [12]. The time plot shows the residuals are white noise meaning zero mean and constant variance. The ACF plot indicates that the residuals are approximately uncorrelated. Lastly, the histogram revealed that the residuals are approximately normally distributed with a mean zero. The adequacy test result is also confirmed by the formal test Ljung- Box test in Table 3.

Ljung – Box Test

Table 3: Test result of Ljung-Box test

Q^*	df	p-value
10.611	17	0.8761

H_0 = A series of residuals has no autocorrelation for a fixed number of lags.

Source(s): Created by authors

Table 3 above shows the results of the Ljung-Box test on the residuals of the model. The p-value of the test statistic is 0.8761 which is greater than the 0.05 significance level, signifying that residuals are uncorrelated and they are pure randomly.

3.5 Model selection and estimation

Since the time series tends to be stationary after the first differencing, we have to identify the order of Seasonal ARIMA (p, d, q) (P, D, Q)[12]. To achieve this the function auto. arima () in R software was used to select the required model for forecasting the price of rice with the minimum value of AIC and BIC and Maximum log-likelihood. The study arrived at a SARIMA model which is the Seasonal ARIMA (3,1,2) (2,0,0) [12] and was selected due to the lowest AIC of 5151.14, BIC of 5178.85 and the largest log-likelihood of −2567.57 among the other models and it was considered as the best model for forecasting price of rice.

3.6 Model Validation

To check if the model is best for forecasting, its forecasted predictions for the validation set are plotted against the observed values. Figure 5 below shows the graph of the fitted predicted values using the original data. The results revealed that the model is best for the data set.

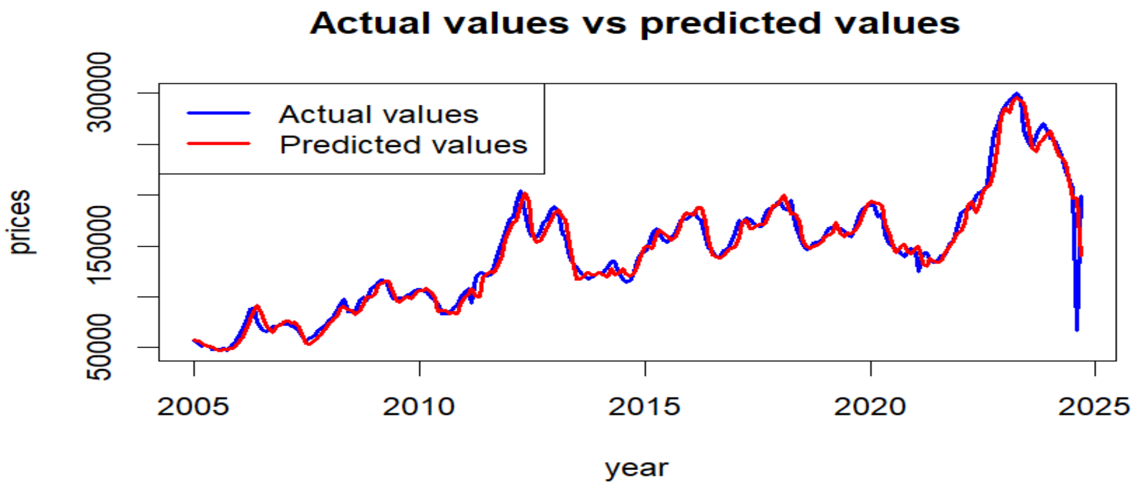

Figure 4: Model Validation of Rice Price Using SARIMA (3,1,2) (2,0,0)[12]
Source(s): Created by authors

3.7 Forecasting

Table 4 and Figure 5 below display the forecasted values of the price of rice for the coming 60 months using the Seasonal ARIMA (3,1,2) (2,0,0) [12] model.

Table 4: Forecasted Values for the TOTAL Variable

Point	Forecast	Point	Forecast
Oct-24	122317.3	Apr-27	272113.0
Nov-24	169659.0	May-27	271416.6
Dec-24	135368.9	Jun-27	265461.2
Jan-25	150348.5	Jul-27	263547.6
Feb-25	139592.4	Aug-27	227291.8
Mar-25	144459.7	Sep-27	262953.7
Apr-25	142760.0	Oct-27	243435.8
May-25	142480.6	Nov-27	255125.9
Jun-25	132621.6	Dec-27	242765.4
Jul-25	134189.9	Jan-28	242604.7
Aug-25	95817.2	Feb-28	236171.0
Sep-25	151579.5	Mar-28	232065.1
Oct-25	140071.7	Apr-28	226354.7
Nov-25	169085.3	May-28	219061.3
Dec-25	167381.6	Jun-28	209894.3
Jan-26	182165.6	Jul-28	203923.3
Feb-26	190914.3	Aug-28	166420.6
Mar-26	201997.6	Sep-28	196710.8
Apr-26	211855.8	Oct-28	176305.9
May-26	218186.2	Nov-28	186368.2
Jun-26	222688.2	Dec-28	175712.6
Jul-26	230975.7	Jan-29	177514.7
Aug-26	178891.2	Feb-29	174030.4
Sep-26	251764.4	Mar-29	174101.5
Oct-26	228168.6	Apr-29	173302.5
Nov-26	259975.1	May-29	172366.6
Dec-26	251795.4	Jun-29	169992.0
Jan-27	264840.1	Jul-29	171009.2
Feb-27	265423.7	Aug-29	150760.6
Mar-27	270830.9	Sep-29	178203.8

Source(s): Created by authors

From October 2024 to June 2027 there is an overall increase in the price of rice, this shows that, in this period there is a positive increase in trend. When measured as the rate of change, it is equivalent to a 5129.9 increase and shows that for each year from 2024 to 2027, the price of rice increases by 5129.9 units or over the three years, the price of rice would increase by approximately 5129.9 units per year. After July 2027, the forecasts show a consistent decline until mid-2028, after which they stabilize at lower values through 2029. The drop in the price of rice is likely to be a result of a combination of factors, including increased supply (from improved harvests or imports), decreased demand (due to changing preferences or economic conditions), and potential government policies or global market trends. From August 2029 the price forecast stabilizes and the price of rice increases again.

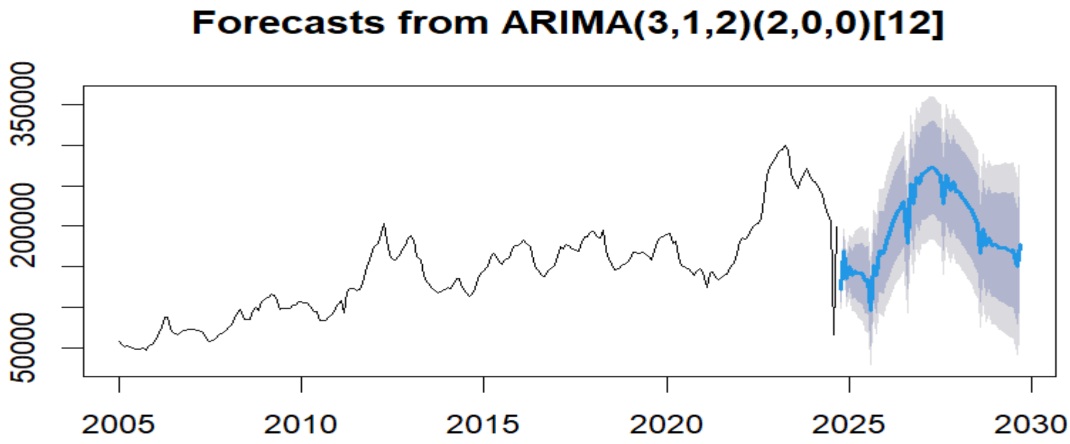

Figure 5: The Time Series Plot of the Forecasted rice price
Source(s): Created by authors

4. Conclusion

This study forecasted the price of rice in Tanzania using the SARIMA model. Precisely, we used the seasonal ARIMA (3, 1, 2) (2, 0, 0) [12] model with seasonal data. The analysis revealed that the price of rice in Tanzania exhibits a consistent upward trend from October 2024 (122,317.3 TZS) to a peak in March 2027 (270,830.9 TZS), reflecting a total increase of approximately 121.4% over this period. From April 2027, prices begin to decline steadily, reaching a low of 150,760.6 TZS in August 2029, a reduction of 44.3% from the March 2027 peak.

Prices are consistently highest, particularly in February–April. For example, from January to March 2027, prices increase by 2.2% from 265,423.7 TZS to 270,830.9 TZS.

Prices typically dip in late summer, with the sharpest decline observed in August 2025, when prices fall to 95,817.2 TZS, a decrease of 32.8% from June 2025 (132,621.6 TZS).

The decline beginning in April 2027 reflects a significant market shift, with an average monthly decrease of approximately 2.1% from April 2027 to August 2029. This may be attributed to increased production or effective market interventions.

The methodology that has been used in this paper (SARIMA) serves to help in determining the future price of Rice in the country and will enable the policymakers in Tanzania and government officials to make a well-informed decision to improve rice prices which shows no significant change for the coming months. However, due to the fluctuation in the data series, this research needs to be extended by applying other methodologies such as the Autoregressive Integrated Moving Average with exogenous variable (SARIMAX), Simple Exponential Smoothing (SES) or the Holt-Winters Exponential Smoothing (HWES), the methodologies that may improve the results from this study but also widening the scope on how the forecasting of the rice prices generation can yield best results.

5. Policy Recommendations

The policy recommendations should look at both short-term and long-term interventions. These altogether will help the government through its policymakers to create a reliable and stable rice market that will later on favour both producers and consumers. The policy focus should precisely look into the areas of price stabilization programmes by establishing national rice reserves. To ensure a feasible and sustainable national rice reserve system, the government may allocate funds from the national budget including one-time setup costs for building storage facilities and ongoing operational costs for procurement and storage, building and maintaining warehouses, silos, and climate-controlled storage spaces as well as plan for funding and maintaining reserves during times of economic recession. The government may also strengthen trade and export policies by encouraging the free movement of rice during shortages and tax relief thereto, which will attract supply from neighbouring countries experiencing different production conditions while at the same time protecting local producers. Further, the government may consider putting in place mechanisms to support farmers by giving them access to agro-inputs (appropriate seed, fertilizer and pesticides) as well as ensuring the technical services are availed to farmers through agricultural extension officers. There is a relationship between the demand for rice and a subsequent increase in rice prices. Therefore, when the rice yield increases or is stable, it will address the problem of demand that will in turn solve the entire problem of price volatility. Further, given its potential, as both food for humans and a source of foreign currency, the government may consider putting the systems in place to monitor rice production, supply, demand, and price data. This can be achieved through appropriate and timely collection of the data, analysis and reporting methods across all levels. By doing so, there will always be accurate and timely data which can improve forecasting accuracy and subsequently, enable better policy responses to price fluctuations and early resilience mitigation strategies

References

Akaike, Hirotugu. "A new look at the statistical model identification." IEEE Transactions on Automatic Control 19, no. 6 (1974): 716-723.

Aulia, I. D., & Pratama, I. (2024). Analysis of Forecasting Methods on Rice Price Data at Milling Level According to Quality. *Edu Komputika Journal*, *11*(1), 1-10.

Box, G., & Jenkins, G. (1970). Time series analysis: Forecasting and control. San Francisco: Holden-Day.

Dickey, David A., and Wayne A. Fuller. "Distribution of the estimators for autoregressive time series with a unit root." Journal of the American Statistical Association 74, no. 366a (1979): 427-431.

FAO. (2015). *The State of Food and Agriculture 2015: Social protection and agriculture: Breaking the cycle of rural poverty*. Food and Agriculture Organization of the United Nations. http://www.fao.org/3/a-i4910e.pdf

Garov, A. K., & Awasthi, A. K. (2023). Time series Arima-based forecasting of rice price. *International Journal of Agricultural & Statistical Sciences*, *19*(1).

Hassan, M. F., Islam, M. A., Imam, M. F., & Sayem, S. M. (2011). Forecasting coarse rice prices in Bangladesh. *Progressive Agriculture*, *22*(1-2), 193-201.

Kullback, Solomon, and Richard A. Leibler. "On information and sufficiency." The Annals of Mathematical Statistics 22, no. 1 (1951): 79-86.

Kundu, R., & Sharma, A. (2022). Development of Seasonal ARIMA Model to Predict Wholesale Price of Rice in Delhi Market. *Current Journal of Applied Science and Technology*, *41*(48), 155-161.

Kwas, M., Paccagnini, A., & Rubaszek, M. (2021). Common factors and the dynamics of industrial metal prices. A forecasting perspective. *Resources Policy*, *74*, 102319.

Mitra, D., & Paul, R. K. (2021). Forecasting of the price of rice in India using a long-memory time-series model. *National Academy Science Letters*, *44*(4), 289-293.

Mubarak, F., & Sundara, V. Y. (2024). Forecasting the Price of Rice in Banda Aceh after COVID-19. *arXiv preprint arXiv:2411.15228*.

Nkuba, J., Ndunguru, A., Madulu, R., Lwezaura, D., Kajiru, G., Babu, A., ... & Ley, G. (2016). Rice value chain analysis in Tanzania: identification of constraints, opportunities and upgrading strategies. *African Crop Science Journal*, *24*(1), 73-87.

Ohyver, M., & Pudjihastuti, H. (2018). Arima model for forecasting the price of medium-quality rice to anticipate price fluctuations. *Procedia Computer Science*, *135*, 707-711.

Ramadhani, F., Sukiyono, K., & Suryanty, M. (2020). Forecasting of Paddy Grain and Rice's Price: An ARIMA (Autoregressive Integrated Moving Average) Model Application. *SOCA: Jurnal Sosial, Ekonomi Pertanian*, *14*(2), 224.

Ramasubramanian, V. "Time series analysis." New Delhi (2007).

Sekiya, N., Oizumi, N., Kessy, T. T., Fimbo, K. M., Tomitaka, M., Katsura, K., & Araki, H. (2020). Importance of market-oriented research for rice production in Tanzania. A review. *Agronomy for Sustainable Development*, *40*, 1-16.

Septiana, D. (2024). Forecasting Rice Prices with Holt-Winter Exponential Smoothing Model. *Hanif Journal of Information Systems*, *1*(2), 62-67.

.