International Journal of Recent Engineering Research and Development (IJRERD)
ISSN: 2455-8761
www.ijrerd.com || Volume 07 – Issue 09 || September 2022 || PP. 09-22

Modelling and Forecasting Tea Prices and their volatility in Kenya

Robin Ochieng Otieno
Department of Mathematics
Multimedia University of Kenya
P.O. Box 50598, Nairobi Kenya

Abstract: This study aimed to model and forecast monthly tea prices and their volatility in Kenya. The data used in this study was obtained from FRED for the period January 1, 1990, to November 1, 2021. The series was non-stationary and was made stationary by obtaining the first difference. The $SARIMA(3,0,3)(0,1,1)_{12}$ was then identified using the EACF plot that obtained the auto-regressive and moving average terms, and had an AIC of 3167.082. The P-Value 0.0490 was obtained from the ARCH-Lagrange multiplier test, and it was concluded that significant ARCH effects exist; hence, volatility modelling was conducted. $GARCH(1,1)$ model was identified as the best volatility model with an AIC and BIC of 2542.47 and 2558.24, respectively. The parameter α_1 was 0.1527, indicating stable term volatility, while the persistent level $\alpha_1 + \beta_1 = 0.9479$ indicated that in the future, there would be long periods of volatility. Thus, the tea sector in Kenya was recommended prepare to deal with persistent volatility in the monthly price of tea.

Keywords: Volatility; GARCH model; Returns; Heteroscedasticity; SARIMA model.

1. Introduction

Tea has become essential to the economies of producing countries such as Kenya, China, India, and Sri Lanka. In 2019, 25% of the foreign exchange earnings, and 1.5% of the Gross Domestic Product in Kenya were accounted for by the tea sector [5]. Globally, Kenya falls in third place after China and India in tea production with a market share of 9% as shown in Figure 1. The tea sector in Kenya also accounts for 20% of the tea exports globally, as shown in Figure 2, making it the largest exporter of tea in the world[7]. In Africa, the tea-producing countries include Kenya, Uganda, Malawi, Tanzania, and Zimbabwe, producing about 30% of the world's exports.

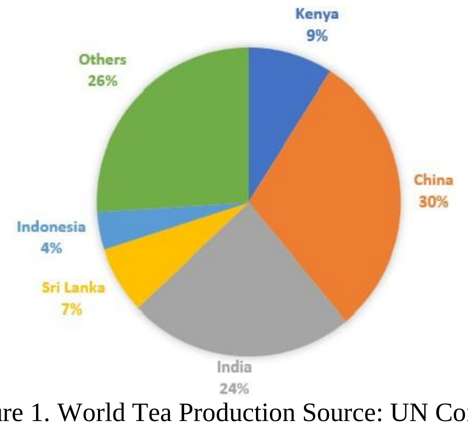

Figure 1. World Tea Production Source: UN Comtrade

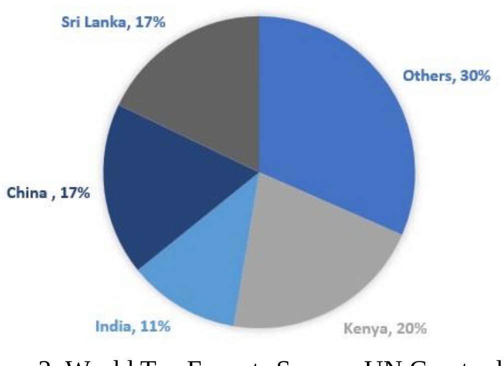

Figure 2. World Tea Exports Source: UN Comtrade.

The tea sector in Kenya contributes to 26% of the export earnings as Kenya exports tea to Pakistan, Egypt, the United Kingdom, and others, as shown in Figure 3 [8]. The main aim of this study was to model and forecast tea prices and their volatility in Kenya. Price volatility is one of the main risks in the agribusiness sector. Volatility in commodities is not directly observable, although it exhibits features in the returns. Some causes of price volatility in Kenya's tea sector are changing climatic conditions such as hot and dry weather, political and economic instabilities in the countries that import Kenyan tea, and competition from other tea exporting countries such as Sri Lanka and China [5]. GARCH models have shown to be adequate in modelling and forecasting volatility since it considers volatility clustering. Their ability to also model time-varying conditional variances provides accurate forecasts

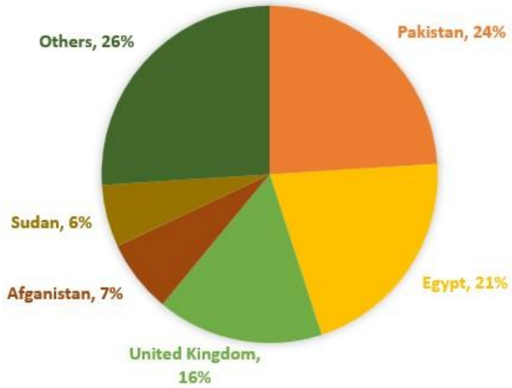

Figure 3 Kenya Tea Patners

Source: UN Comtrade of variances and covariances of returns [2]. GARCH model is used in this study because it minimizes forecasting errors by taking care of errors in the previous forecasts and further improving the accuracy of ongoing predictions.

2. Methodology

2.1. Introduction

A time-series approach is used to model and forecast the global price of tea in Kenya. The data used in this study was obtained from Federal Reserve Economic Data and consists of the monthly price for a Kilogram of tea in Kenya for the period January 1, 1990, to November 11, 2021, and consists of 381 observations.

Date	Price of tea(Ksh.)
1990-01-01	252.929993
1990-02-01	211.289993
1990-03-01	192.600006
1990-04-01	194.809998
1990-05-01	187.020004
1990-06-01	181.490005
1990-07-01	189.789993
1990-08-01	174.550003
1990-09-01	185.250000
1990-10-01	220.199997

Table 1 A Sample of the Data

Table 1 is the sample of the monthly price of tea in Kenya data recorded in Kenyan shillings.

2.2. Data Preprocessing

Data preprocessing is conducted to make it easy to interpret and use during modelling and analysis. The process involves four steps; firstly, the date column is set as the index, then the date column data type is converted from object type to Date Time, and its frequency is set to monthly start. Finally, the prices column is renamed, and any missing values in the data are checked. The descriptive statistics for the data are checked to better have a good understanding of the data.

2.3. Test for Autocorrelation and Checking for Outliers

The lag plot is used to test for autocorrelation presence and check for outliers in Kenya's monthly price of tea. Outliers are the data points that differ from the other data points, and if they exist, they are dropped. Autocorrelation presence indicates that an auto-regressive model is more suitable for the monthly price of tea.

2.4. Tests for Stationarity.

The test for stationarity is conducted visually and inferentially. Visually it is conducted using Auto-correlation and Partial Autocorrelation plots. Inferentially it is conducted using the Augmented Dickey-Fuller (ADF) test. A unit root in a time series indicates that the series is non-stationary, and the ordinary least squares estimator is not normally distributed. The null hypothesis is rejected if the test statistic is larger than the critical value. After attaining stationarity, autocorrelation and partial autocorrelation are examined to identify their proper structures. It is conducted using the Ljung-Box Q-statistic test [6], which is defined as:

$$Q = n(n+2)\sum_{j=1}^{m}\frac{r_j^2}{n-j} \tag{2.1}$$

From Equation 2.1, r_j is the accumulated sample autocorrelations, n is the sample size, and m is the maximum lag length. Our series is made stationary by taking the first order differencing since it is seasonal data.

2.5. SARIMA Model

2.5.1. Model Identification: Stationary data is attained by taking the first difference of the data; the Extended Autocorrelation Function (EACF) is used to determine the autoregressive and moving average components. The model identified from the EACF parameters is also used in estimating the best SARIMA model by trial and error on potential models. The SARIMA model can be expressed as;

$$\Phi_p(B^s)\phi(B)\nabla_s^D\nabla^d X_t = \delta + \Theta_Q(B^s)\theta(B)w_t \tag{2.2}$$

From Equation 2.2, X_t is the monthly price of tea in Kenya, w_t is the white noise process, and the polynomials $\phi(B)$ and $\theta(B)$ of orders p and q represent the ordinary autoregressive and moving average components, respectively.

The seasonal auto-regressive integrated moving average elements are represented by $\Phi_p(B^s)$ and $\Theta_Q(B^s)$ of orders P and Q, while the ordinary and seasonal difference components are represented by $\nabla^d = (1 - B)$ and $\nabla_s^D = (1 - B^s)$ [9].

2.5.2. Model Selection: Model selection is conducted based on the Akaike Information Criterion (AIC) proposed by [1]. Akaike Information Criterion approximates each model's quality given a collection of models for data, and the model with the most negligible value of AIC value is considered the best model. The formula for the Akaike Information Criterion (AIC) is:

$$AIC = 2K - 2\ln(L) \tag{2.3}$$

From Equation 2.3, K represents the number of model parameters that have been estimated, while L is the value of the maximum likelihood function for the model.

2.5.3. Model fitting: The data is split into the train and test sets, where 80 percent of the data is used as the train set to train the model, while 20 percent of the data is used as the test set to evaluate the model forecasting performance. The train set is then fitted on the best model estimated and is used for forecasting.

2.5.4. Forecasting for future periods: The fitted model was used to forecast 36 months ahead, which are the same prices as in the test set. The forecasted values are plotted against the actual values in the test set to check the model's forecasting performance. The model is then fitted to the full data and used in forecasting for ten years ahead

2.5.5. Evaluating the SARIMA model: The Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) are used to evaluate the forecasting performance of the SARIMA model on the monthly price of tea in Kenya. The formula for MAE is given in Equation 2.4:

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|Y_t - F_t| \tag{2.4}$$

The formula for RMSE is given Equation 2.5:

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(Y_t - F_t)^2} \tag{2.5}$$

where Yt = Observed value of time t, Ft = Forecasted value of time t, and n is the number of observations.

2.5.6. Residual Diagnostics: The Residual diagnostics are conducted to assess further our model's goodness of fit, where the independence and normality of the residuals from the fitted model are checked. The test for normality is conducted to test if the residuals follow the normality assumption. Visually, the normal Q-Q plot, histogram, correlogram, and density plot are used to test the normality of the residuals. Inferentially, the normality test is conducted using the Jacque-Bera test. The Jarque-Berra Test is used in this case to test whether the data has skewness and kurtosis marching a normal distribution. The null hypothesis states that the residuals are normally distributed. The formula for the Jarque-Bera test statistic by [4] is:

$$JB = \frac{n}{6}(S^2 + \frac{(K-3)^2}{4}) \quad (2.6)$$

From Equation 2.6, n is the number of observations or the degrees of freedom, S is the sample skewness, and K is the sample Kurtosis.

2.6. Volatility Modelling

2.6.1. Returns: Let X_t and X_{t-1} denote the current month and the previous month's price of tea, respectively. The monthly return series were obtained by using Equation 2.7:

$$r_t = \ln(\frac{X_t}{X_{t-1}}) \quad (2.7)$$

2.6.2. Test for Heteroscedasticity: The test for heteroscedasticity is done to check for the presence of ARCH effects, that is if the variance is non-constant over time. The ARCH-Lagrange Multiplier test proposed by [3] determines the ARCH effects' significance. The test is done with the assumption that the residuals are heteroscedastic, while the square of the residuals is autocorrelated. The null hypothesis is that there are no arch effects in the series, while the alternative hypothesis is that there is the presence of arch effects in the series. The test statistics of the Lagrange Multiplier is calculated using Equation 2.8:

$$L(x,\lambda) = f(x) - \lambda g(x) \quad (2.8)$$

2.7. GARCH Model

A *GARCH* model uses values of the past squared residuals and past variances to model the variance at time t. The *GARCH*(*p*,*q*) model is a combined model of the *ARCH*(*q*) model and introduces p lags of the conditional variance in the model, where p is known as the *GARCH* order. The return in a GARCH model is given in Equation 2.9. The combined model is called the generalized autoregressive conditional heteroscedasticity, *GARCH*(*p*,*q*) model as given in Equation 2.10 as:

$$r_t = \sigma_t \epsilon_t \quad (2.9)$$

$$\sigma_t^2 = \alpha_0 + \sum_{j=1}^{p} \alpha_j r_{t-j}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \quad (2.10)$$

Where σ_t^2 is the conditional volatility, r_{t-1}^2 is the previous months squared returns, and σ_{t-1}^2 is the previous months volatility. To ensure that the conditional volatility $\sigma_t^2 > 0$ all parameters α_0, α_j and β_j are supposed to be non-negative In this study *GARCH*(1,1) is used where the coefficients $p = 1$ and $q = 1$ are as given in Equation 2.11:

$$\sigma_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \quad (2.11)$$

Where σ_t^2 is the conditional volatility, r_{t-1}^2 are the previous months squared returns, and σ_{t-1}^2 is the previous months volatility. All parameters α_0, α_1, and β_1 are non negative this ensures that the conditional volatility $\sigma_t^2 > 0$ at all times.

2.7.1. Parameter Estimation: The maximum Likelihood Method estimates the model's parameters under the normal distribution. The p-value associated with each model parameter is checked for significance. Parameters with a pvalue less than 0.05 are considered significant. The GARCH model parameter values obtained are substituted to the model given in Equation 2.11.

2.7.2. Diagnostic checking: The best GARCH model that is fitted is checked for the goodness of fit. The standard residuals of the GARCH model are used for diagnostic checking. The standard residuals, by assumption, are supposed to follow a normal distribution. The probability and histogram density plots are used to conduct the diagnostic check. The normality test is conducted using the Jacque-Bera test. The Jarque-Berra Test is used in this case to test whether the data has skewness and kurtosis marching a normal distribution, with the null hypothesis stating that the residual is normally distributed. The formula for the Jarque-Bera test statistic by [4] is:

$$JB = \frac{n}{6}(S^2 + \frac{(K-3)^2}{4}) \quad (2.12)$$

From Equation 2.12, n is the degrees of freedom, S is the sample skewness, and K is the sample Kurtosis.

2.7.3. Volatility Forecasting: The series is split into the train set 80%, and the test set 20%. The best GARCH model is fitted to the train data. The train data model forecasts for the period in the test set. The test set, which contains the actual values, is thus used to evaluate the GARCH model's performance in forecasting volatility in Kenya. The forecasted values are plotted against the values in the test set to evaluate the model's performance.

2.7.4. Evaluation of Forecasts: The forecast of the GARCH (1,1) model is evaluated using the Mean Absolute Error(MAE) and the Root Mean Squared Error (RMSE) on the monthly price of tea in Kenya. The formula for the MAE is given in Equation 2.4, while the formula for the RMSE is given in Equation 2.5.

3. Results and Discussion

3.1. Data

To model and forecast tea prices and their volatility in Kenya, the monthly prices of tea per kilogram were obtained from the Federal Reserve for Economic Data (FRED) for the period January 1, 1990, to November 1, 2021. A sample of ten observations from the data are shown in Table 4.1 from January 1990 to October 1990.

Date	Price of tea (Ksh.)
1990-01-01	252.929993
1990-02-01	211.289993
1990-03-01	192.600006
1990-04-01	194.809998
1990-05-01	187.020004
1990-06-01	181.490005
1990-07-01	189.789993
1990-08-01	174.550003
1990-09-01	185.250000
1990-10-01	220.199997

Table 2 Sample of the Data

3.1.1. Descriptive statistics of the series: The series had 381 observations with a mean price of Ksh. 245.80, and a standard deviation of Ksh. 60.70. The series has a positive skewness of 0.591, and a kurtosis of 2.370, which means that the distribution has a long right tail. The minimum price for tea in the series is Ksh. 143.37 which was sold on July 1, 1995, while the maximum price is Ksh. 403.03 per Kilogram, which was sold on July 1, 2015.

Observations	381
Mean	245.8216
Median	236.6700
Maximum	403.0323
Minimum	143.3701
Std. Dev.	60.6912
Skewness	0.5909
Kurtosis	2.3701

Table 3 Descriptive statistics of the price of tea

3.1.2. Plot of the data: Figure 4 represents the plot of the data from 1990 - 2021, it can be observed that the data is non-stationary, that is, the mean and variance are not constant over time, thus differencing will be conducted to make the data stationary. Seasonal and cyclic patterns can be observed over time. Periods of high and low volatility can also be observed.

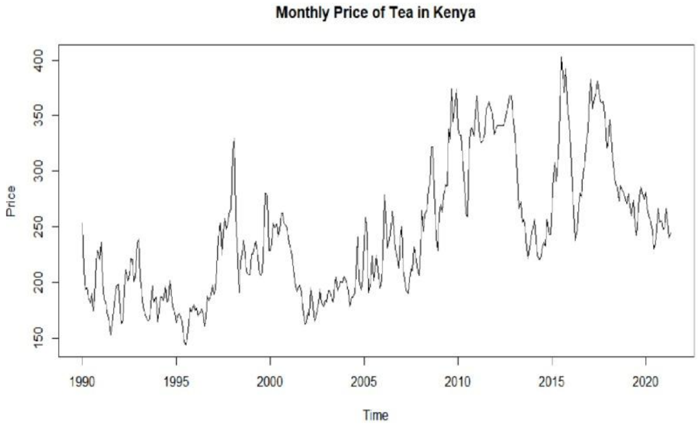

Figure 4 Time series plot 1990-2021

3.1.3. Test for Autocorrelation and Checking Outliers: The test for autocorrelation and checking for outliers were conducted using a lag plot. From Figure 5 no outliers are present in the series, and there is a positive serial auto-correlation presence. Thus, an auto-regressive model is more suitable.

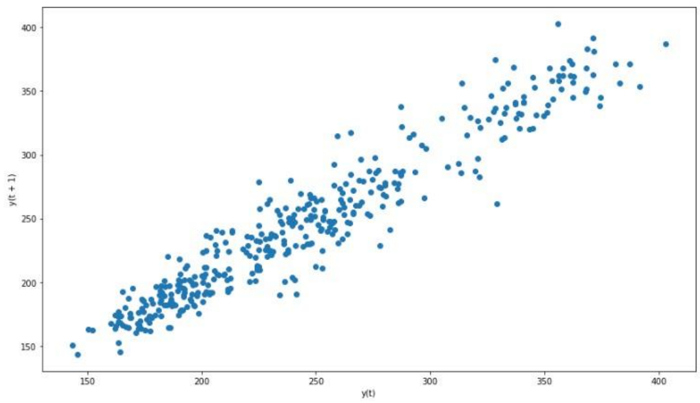

Figure 5 Lag plot of the data

3.2. Test for Stationarity

The test for stationarity was conducted both inferentially and visually. Inferentially it was performed using the Augmented Dickey-Fuller (ADF) test, while visually, it was conducted using the Auto-correlation (ACF) and Partial Autocorrelation (PACF) plots.

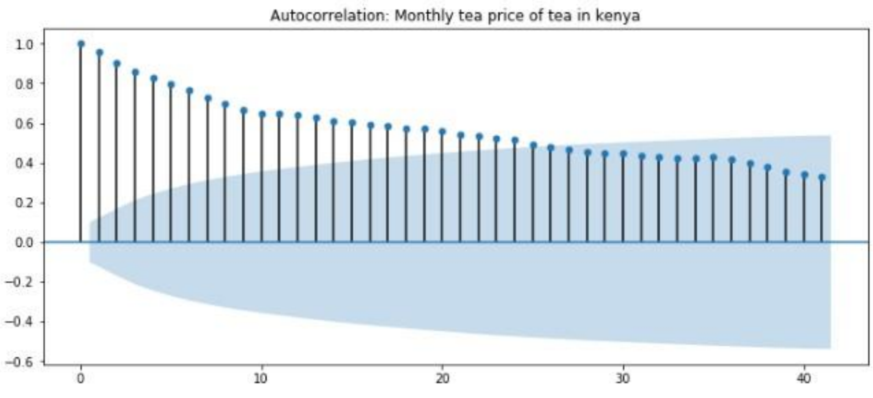

Figure 6 ACF plot of the data

Figure 6 indicates a slowly decaying ACF, which implies that the past values heavily influence the future values of the monthly price of tea in the series. It is a typical ACF plot for non-stationary data, with lags on the horizontal axis and correlations on the vertical axis. From Figure 7 which is the PACF plot; it can be seen that the first and second lags are significant correlations confirming the presence of auto-regressive terms in the series.

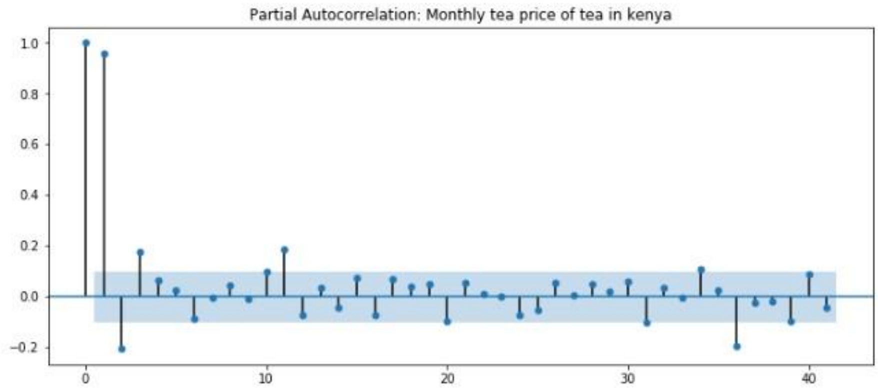

Figure 7 PACF plot of the data

In Table 4, the ADF test statistic is (−1.8424), with a corresponding p-value of 0.3596, which is greater than 0.05. Hence the null hypothesis is rejected, confirming the presence of a unit root. Thus we conclude that the data is non-stationary.

Test Statistics	-1.8424
P-Value	0.3596
Lags Used	10.0000
Critical Value(1%)	-3.4481
Critical Value(5%)	-2.8693
Critical Value(10%)	-2.5709

Table 4 Results of the ADF test

3.3. Decomposition of the time series

Time series data can display a diversity of patterns and can be split into several components such as the trend, seasonality, and residuals which are the irregular variations in the series. Figure 8 is the decomposition of the monthly price of tea in Kenya into observed, trend, seasonal, and random components. Based on Figure 8, we can see a general upward trend in the price of tea. Seasonality is also observed since this is monthly data. The residuals exhibit periods of high volatility and low volatility i.e volatility clustering.

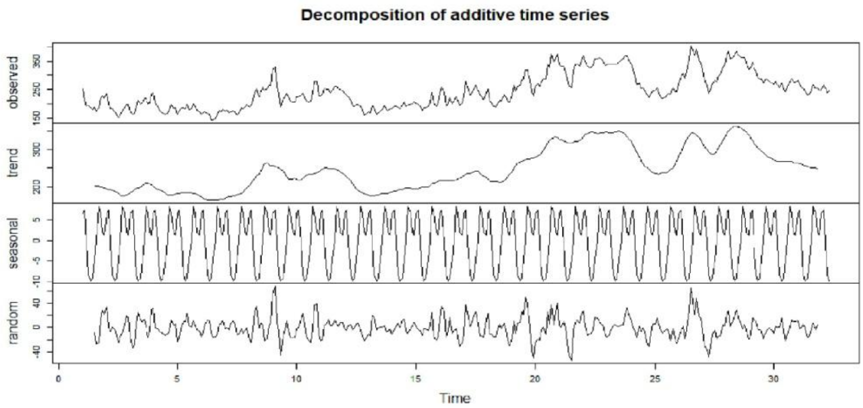

Figure 8 Decomposition of the series

3.4. The SARIMA model

3.4.1. Model Identification: Since we cannot pass non-stationary data on our model, the data is made stationary by taking the first difference in the price of tea. Figure 10 is the plot of the stationary data; clearly, it exhibits constant mean and variance over time.

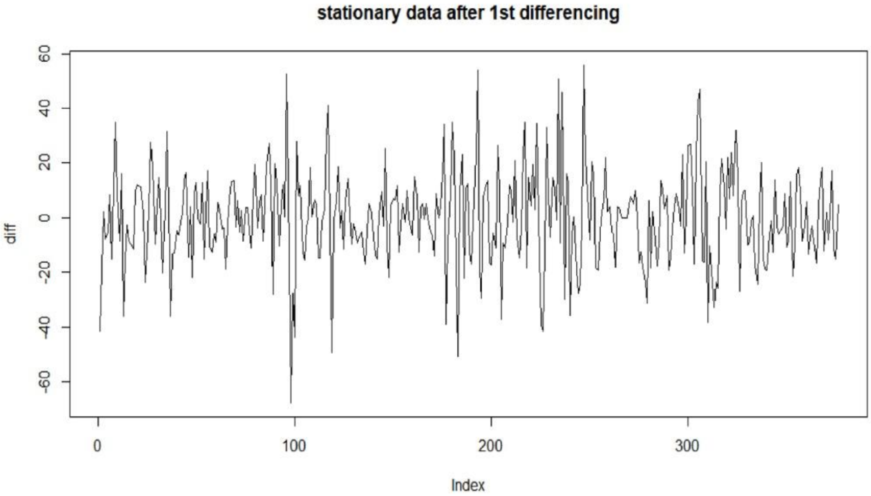

Figure 9 Stationary data

The Extended Autocorrelation Function (EACF) is used to identify the autoregressive and moving average components of the stationary data obtained. The results of EACF are shown in Figure 10.

AR/MA	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	x	x	x	x	x	x	x	x	x	x	x	x	x	x
1	x	x	x	o	x	o	o	o	o	x	o	o	o	o
2	x	x	x	o	x	o	o	o	o	x	x	o	o	o
3	x	o	o	o	o	o	o	o	o	x	x	o	o	o
4	o	o	x	o	o	o	o	o	o	x	o	o	o	o
5	o	x	x	o	o	o	o	o	o	x	o	o	o	o
6	o	x	x	o	o	o	o	o	o	x	o	o	o	o
7	x	x	o	x	o	o	o	o	o	x	o	o	o	o

Figure 10 Extended ACF

From the results obtained in figure 10, an ARMA (3, 3) model was chosen as the best model, which was fitted along with other tentative models.

3.4.2. Estimation of the SARIMA model: The trial and error method was used in estimating the best SARIMA model. Some of the potential SARIMA models along with their AIC value are shown in Table 5. The best *SARIMA* model for fitting the monthly price of tea in Kenya is *SARIMA* (3, 0, 3) (0, 1, 1)$_{12}$ with an AIC value of 3167.082.

Model	AIC
SARIMA(1,0,1)(0,1,1)$_{12}$	3168.693
SARIMA(1,0,0)(1,1,0)$_{12}$	3274.834
SARIMA(2,0,1)(0,1,1)$_{12}$	3169.513
SARIMA(2,0,3)(0,1,1)$_{12}$	3172.886
SARIMA(1,0,2)(0,1,1)$_{12}$	3168.906
SARIMA(3, 0, 3)(0, 1, 1)$_{12}$	3167.082
SARIMA(1,0,1)(1,1,2)$_{12}$	3172.675

Table 5. The AIC of the models fitted

3.4.3. Fitting the SARIMA Model: Before fitting the model, the data was split into the train and test sets to validate the model's performance. The train set had 347 months, while the test set had 36 months. *SARIMA* (3,0,3) (0,1,1)$_{12}$ was fitted on the train set using the SARIMAX function from Stats models in python. From the results of Table 6, we can see that the model parameters are all significant since they are all less than 0.05. Thus, the model is significant.

Parameters	Coef	P-Value	Conclusion
AR(1)	2.130	0.000	Significant
AR(2)	-1.913	0.000	Significant
AR(3)	0.753	0.000	Significant
MA(1)	-1.025	0.000	Significant
MA(2)	0.497	0.000	Significant
MA(3)	0.205	0.000	Significant
MA.S(1)	-0.971	0.000	Significant
Sigma	267.178	0.000	Significant

Table 6 Sarima model on the train set

3.4.4. Forecasting using the test set: The test set was used to validate the forecasts made for the monthly price of tea for 36 months ahead using the SARIMA model created from the training set. From Figure 11, we see the actual values in the test set plotted against the predicted values from the model.

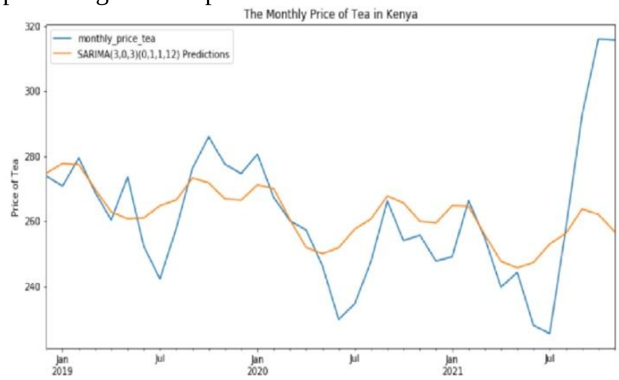

Figure 11 Forecasts for 36 months ahead

3.5. Forecasting for the next ten years

The SARIMA model was fitted to the full data and was used to forecast the monthly tea price for the next ten years, 2021-2031. Figure 12 is the plot for ten-year forecasts for the monthly price of tea in Kenya. We can see that the tea price trend in Kenya is decreasing with time.

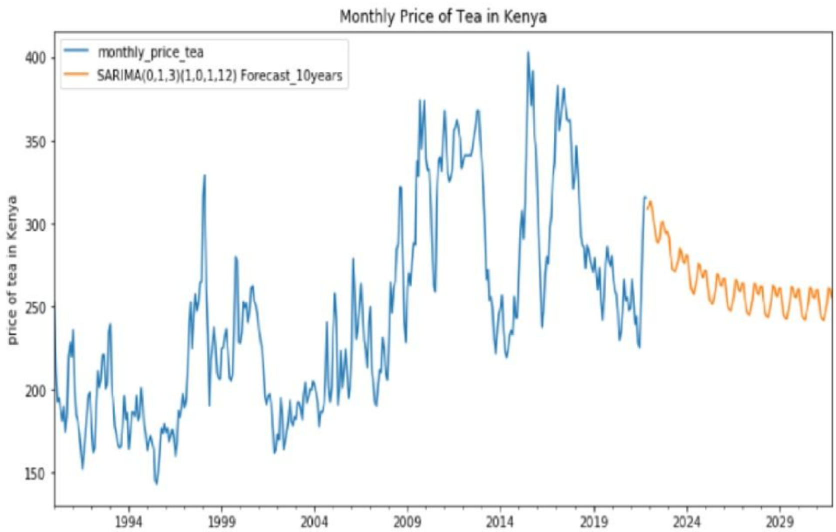

Figure 12 Forecasts for 10 years ahead

3.5.1. Evaluating the SARIMA Model: The Root Mean Squared Error and Mean Absolute Error was used to evaluate the forecasting performance of the SARIMA model on the test data.

Model	RMSE	MAE	MSE
$SARIMA(3,0,3)(0,1,1)_{12}$	17.90	11.90	320.66

Table 7 Forecasting performance of the SARIMA Model

Based on the forecast performance metrics that are obtained in Table 7, the $SARIMA(3,0,3)(0,1,1)_{12}$ model's prediction had a Root Mean Squared Error of 17.90, Mean Absolute Error of 11.90, and Mean Squared Error of 320.66. We can thus conclude that the SARIMA model performed well on the monthly price of tea.

3.5.2. SARIMA Model Residual Diagnostics: Residual diagnostics were conducted to assess the goodness of fit of our model. Figure 13 contains the standardized residuals, histogram plus estimated density, normal q-q plot, and correlogram for checking the SARIMA model adequacy.

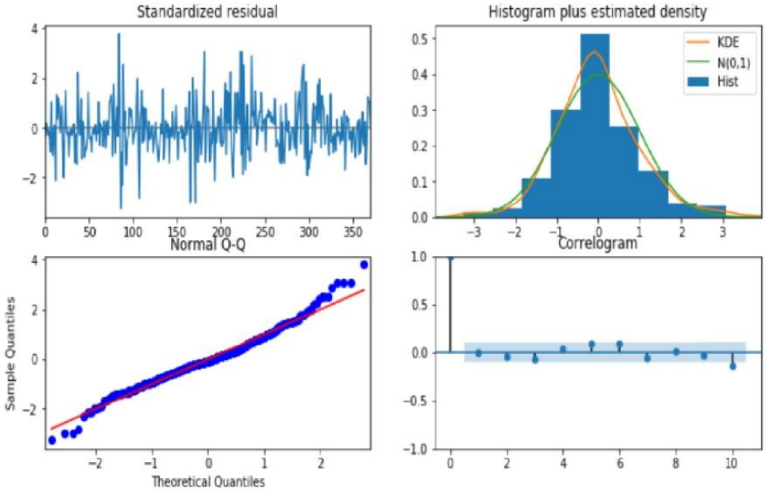

Figure 13 SARIMA Model Residual Diagnostic plot

Based on Figure 13, we see that the standard residuals have a constant mean, and the variance was confirmed to follow a normal distribution as observed in the histogram. The normal probability plot confirmed that the data follows a normal distribution, and the correlogram showed no significant lag. Thus, the model was appropriate.

3.6. Volatility Modelling

3.6.1. Plot of the Returns: Figure 14 is the plot of the monthly returns of the price of tea in Kenya. It can be observed that the returns are stationary; that is, they have a constant mean and variance over time

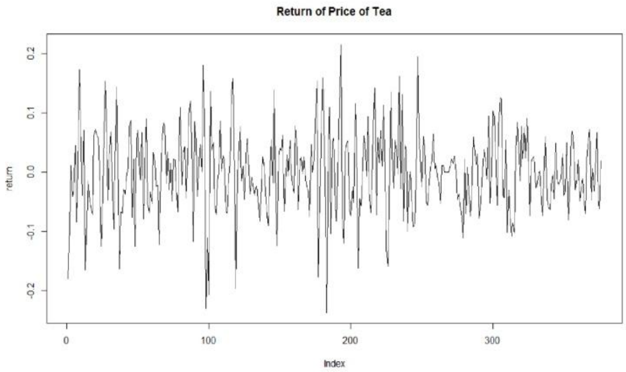

Figure 14 Plot of the return series

3.6.2. Descriptive statistics of Returns: The descriptive statistics of tea's monthly returns were calculated to understand our return series better. Based on Table 8, the minimum return is -21.038, the maximum return

23.956, the mean 0.346, and the standard deviation 6.974. The skewness of 0.224 indicates that the returns are moderately skewed.

Observations	Minimum	Mean	Std. Dev	Maximum	Skewness	Kurtosis
381	−21.038	0.346	6.974	23.956	0.224	3.617

Table 8 Descriptive statistics of the monthly Returns

3.6.3. Test for Heteroscedasticity: The test for Heteroscedasticity is conducted to check for the presence of arch effects in the squared residuals. It is based on the methodology's Lagrange Multiplier of Equation 2.8. Based on Table 9, the ARCH-LM test statistic is 3.875 with an associated P-Value of 0.0490. Thus, we reject the null hypothesis and conclude that significant Arch effects exist for the period January 1, 1990, to November 1, 2021. Thus we can identify the best GARCH model.

Returns	Chi-square	P-Value
Tea	3.875	0.0490

Table 9 ARCH-LM Test for Heteroscedasticity

3.7. GARCH Model

3.7.1. Model Identification: The best model is identified by simulating all the possible combinations of the models and picking the one with the least value of both AIC and BIC. Based on Table 10, different GARCH(p,q) models are simulated, and from the results, GARCH(1,1) was the best model for the monthly price of tea data, with an AIC of 2542.47 and a BIC of 2558.24. The parameters of the GARCH(1,1) model are thus estimated for the price of tea in Kenya.

Model	AIC	BIC
GARCH(0,1)	2563.33	2575.15
GARCH(0,2)	2545.57	2561.34
GARCH(1,1)	2542.47	2558.24
GARCH(1,2)	2543.97	2563.68
GARCH(2,1)	2544.47	2564.19
GARCH(2,2)	2543.79	2567.45

Table 10 AIC and BIC of GARCH models

3.7.2. Parameter Estimation: Parameter estimation was conducted, and the results are shown in Table 11. It can be seen that the p-values are all less than α = 0.05. Thus we conclude that the GARCH(1,1) model is the best model since all its parameters are significant.

Parameters	Coefficient	P-Value	Conclusion
α_0	2.9566	$9.135e-02$	Significant
α_1	0.1527	$1.022e-02$	Significant
β_1	0.7952	$5.531e-34$	Significant

Table 11 GARCH(1,1) Model Parameters

The volatility model GARCH (1, 1) for the Monthly price of tea can thus be expressed as:

$$\sigma_t^2 = 2.9566 + 0.1527e_{t-1}^2 + 0.7952\sigma_{t-1}^2 \quad (3.1)$$

From Equation 3 .1, e_{t-1}^2 is the previous months error term and σ_{t-1}^2 is the previous months volatility. The persistent level ($\alpha_1 + \beta_1 = 0.9479$) is almost equal to one showing that in the future, there will be long periods of the persistence of shocks to the volatility. α_1 = 0.1527 is low indicating stable long-term volatility and β_1 = 0.7952 indicates that the tea's conditional variance will not last before decaying. Therefore the tea sector in Kenya needs to deal with persistent volatility.

3.7.3. Squared Residual and Conditional Volatility: The squared residuals and conditional volatility were plotted as shown in Figure 15. We can see that when the squared residuals are high, the conditional volatility is also high and vice versa. We can also observe volatility clustering; that is, periods of low volatility are followed by periods of low volatility. In contrast, periods of high volatility are followed by periods of high volatility.

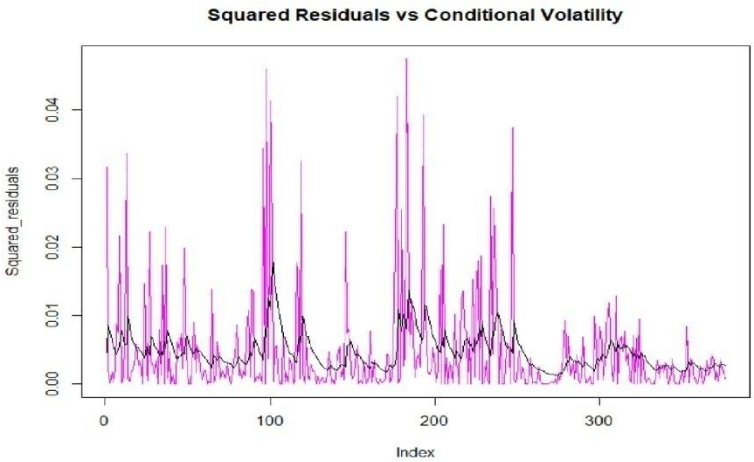

Figure 15 Squared Residuals vs. Conditional Volatility

3.7.4. GARCH model Diagnostic test: The GARCH (1, 1) model diagnostics were conducted on the standardized residuals, which are supposed to follow a normal distribution. The normality of the residuals is checked using the probability plot and the histogram density plot. From Figure 16, the standardized residuals follow a normal distribution despite the small divergence at the start of the plot. The histogram density plot in Figure 17 also confirmed that the standardized residuals follow a normal distribution. Therefore, volatility forecasting is conducted.

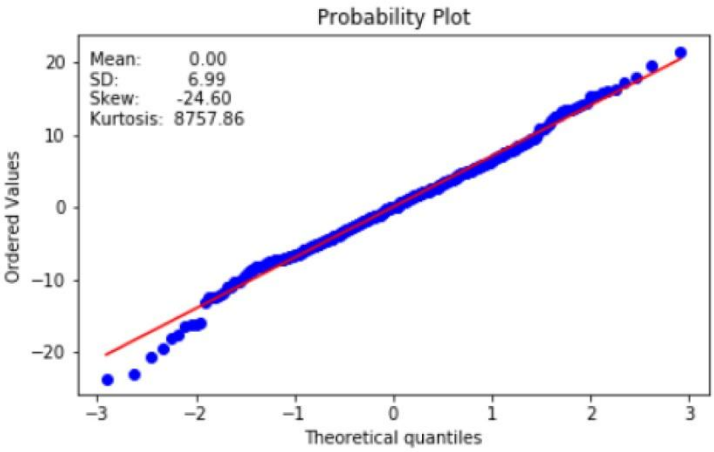

Figure 16 Probability Plot of the Residuals

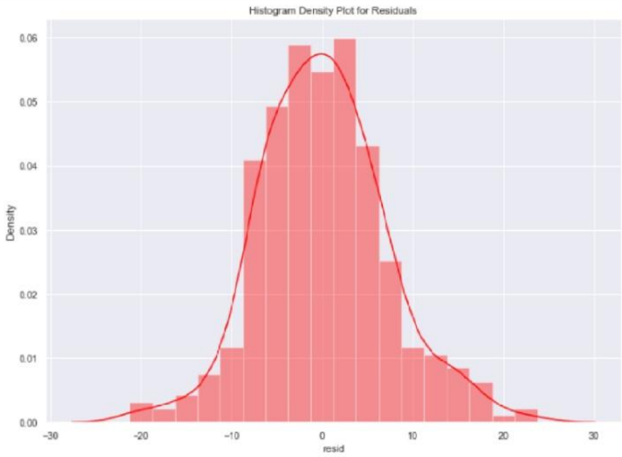

Figure 17 Histogram Density Plot of the Residuals

3.7.5. Volatility Forecasting: The data was split into the train set, and test set before volatility forecasting was conducted. The train set contained 80% of the data while the test set 20% of the data. The train set fitted the GARCH (1, 1) model and forecast 36 months ahead. The test set is used to evaluate the model forecasts. Figure 18 shows that the forecast and actual data patterns are almost similar. Thus, we evaluated the estimates for the GARCH (1, 1) model.

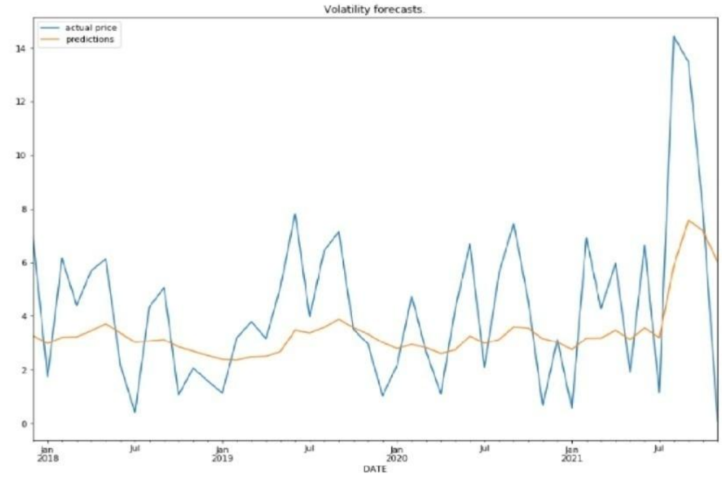

Figure 18 Actual data vs. Forecasts for Volatility

3.7.6. Evaluation of the forecasts: Evaluation of the forecast of the GARCH (1, 1) model is conducted using the Mean Absolute Error (MAE) and the Root Mean Squared Error(RMSE). Based on Table 12, the RMSE (5.903) and the MAE (5.089) values are significant. Thus, we conclude that GARCH (1, 1) model performed well.

Model	RMSE	MAE
GARCH(1,1)	5.903	5.089

Table 12 Forecasting performance of the SARIMA Model

4. Conclusion

This study's objectives have been largely achieved. The SARIMA model was used to forecast the monthly tea price in Kenya. In contrast, the GARCH model was used to model and forecast price volatility using the monthly return series obtained from the monthly price of tea. *SARIMA* (3,0,3) $(0,1,1)_{12}$ was identified as the best model for forecasting the price of tea in Kenya using the EACF plot that provided the auto-regressive and moving average terms for the model. The test for heteroscedasticity was conducted using the ARCHLM on the squared residuals; the p-value of 0.0490 from the test led to the conclusion that the series had significant ARCH effects. The *GARCH* (1, 1) model was identified as the best GARCH model for forecasting tea price volatility. The fitted *GARCH* (1, 1) model indicated volatility persistence on the monthly price of tea series. The parameter $\alpha_1 = 0.1527$ is close to 1 indicating stable long term volatility, while $\beta_1 = 0.7952$ indicates a decaying conditional variance. The long-run volatility persistence level obtained from the *GARCH* (1, 1)model parameters $\alpha_1 + \beta_1 = 0.9479$, which indicates long periods of persistence of shocks to the volatility in the future. Thus the tea sector in Kenya needs to deal with persistent volatility.

References

[1] Hirotugu Akaike. "A new look at the statistical model identification". In: *IEEE transactions on automatic control* 19.6 (1974), pp. 716–723.

[2] Robert Engle. "GARCH 101: The use of ARCH/GARCH models in applied econometrics". In: *Journal of economic perspectives* 15.4 (2001), pp. 157–168.

[3] Robert F Engle. "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation". In: *Econometrica: Journal of the econometric society* (1982), pp. 987–1007.

[4] Carlos M Jarque and Anil K Bera. "Efficient tests for normality, homoscedasticity and serial independence of regression residuals". In: *Economics letters* 6.3 (1980), pp. 255–259.

[5] KIPPRA. *Fluctuations in Market Earnings for Tea in Kenya: What Could be the Cause and Remedy?* 2020. url: https://kippra.or. ke/fluctuations-in-market-earnings-for-tea-in-kenya-whatcould-be-the-cause-and-remedy/.

[6] Greta M Ljung and George EP Box. "On a measure of lack of fit in time series models". In: *Biometrika* 65.2 (1978), pp. 297–303.

[7] L Monroy, W Mulinge, and M Witwer. "Analysis of incentives and disincentives for tea in Kenya". In: *Gates Open Res* 3.586 (2019), p. 586.

[8] Agnes Kinya Muthamia and Willy Muturi. "Determinants of earnings from tea export in Kenya". In: *Journal of World Economic Research* 4.1 (2015), pp. 15–22.

[9] Robert H Shumway, David S Stoffer, and David S Stoffer. *Time series analysis and its applications*. Vol. 3. Springer, 2000.