FORECASTING TEA PRICES AND THEIR VOLATILITY IN KENYA

ROBIN OCHIENG

SCT-325-002/2019

A RESEARCH PROJECT SUBMITTED TO THE DEPARTMENT OF MATHEMATICS IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF A MASTER'S DEGREE IN STATISTICS AT THE MULTIMEDIA UNIVERSITY OF KENYA

JULY 2022

DECLARATION

This project is my authentic exertion and has not been given to any other institution for academic credit apart from the Multimedia University of Kenya:

Signature.............................. Date..................................

NAME: Robin Ochieng Otieno

REG NO: SCT-325-002/2019

This project has been submitted for examination with my approval as the University Supervisor:

Signature.............................. Date..................................

Dr. Wycliff Cheruiyot.

Signature.............................. Date..................................

Dr. Anthony Karanjah.

ACKNOWLEDGEMENT

This paper is dedicated to my parents.

ABSTRACT

The purpose of this study was to forecast monthly tea prices and their volatility in Kenya. The data used in this study was obtained from Federal Reserve for Economic Data (FRED) from January 1, 1990, to November 1, 2021. The approach consisted of identifying the pattern for the monthly price of tea in Kenya using a time series plot, Stationarity tests using the ACF plot, PACF plot, and the ADF test where the non-stationary data was made stationary by obtaining the first difference of the series. The appropriate SARIMA and GARCH models were fitted and used for forecasting future values. The $SARIMA(3,0,3)(0,1,1)_{12}$ was identified as the best model using the EACF plot, which identified the autoregressive and moving average terms and had an AIC of 3167.082. $SARIMA(3,0,3)(0,1,1)_{12}$ was further used for forecasting monthly tea prices for 2 years ahead up to June 2024. The P-Value 0.0490 was obtained from the ARCH-Lagrange multiplier, and it was concluded that significant ARCH effects exist; hence, volatility modelling was conducted. The $GARCH(1,1)$ model was identified as the best model with an AIC and BIC of 2542.47 and 2558.24, respectively. Coefficient estimation for the $GARCH(1,1)$ model indicated they were all significant. The coefficient α_1 was 0.1527, indicating stable term volatility, while the persistent level $\alpha_1 + \beta_1 = 0.9479$, indicating that in the future, there will be long periods of volatility. Thus the tea sector was recommended to prepare to deal with persistent volatility in the monthly price of tea in Kenya. The analysis was conducted using python, E-VIEWS, and R.

TABLE OF CONTENTS

List of Tables

List of Figures

LIST OF ABBREVIATIONS

ACF - Autocorrelation Function.

ADF - Augmented Dickey-Fuller.

AIC - Akaike Information Criterion.

ARCH - Autoregressive Conditional Heteroskedastic.

ARIMA - Autoregressive Integrated Moving Average.

BIC - Bayesian Information Criterion.

GARCH - Generalized Autoregressive Conditional Heteroskedastic.

KPSS - Kwiatkowski-Phillips-Schmidt-Shin.

MAE - Mean Absolute Error.

MAPE - Mean Absolute Percentage Error

RMSE - Root Mean Squared Error.

SARIMA - Seasonal Autoregressive Integrated Moving Average.

SBC - Schwartz Bayesian Criterion.

PACF - Partial Autocorrelation Function.

EACF - Extended Auto-Correlation Function.

1. INTRODUCTION

1.1 Background of the Study

The economy of tea-producing nations like Kenya, China, India, and Sri Lanka now rely heavily on tea exports. Tea was first cultivated in China and it has advanced into a multi-billion dollar international industry. In 1903, British colonial settler G.W.I. Caine brought tea from India to Kenya as an ornamental plant. Today, tea is Kenya's most significant and lucrative export crop (Swainson, 1980). The tea industry in Kenya contributed to 1.6% of the country's Gross Domestic Product and 26% of its foreign exchange profits in 2019 (KIPPRA, 2020) The Tea Board of Kenya has registered 54 factories. Under the Kenya Tea Development Agency (KTDA), the smallholder tea sector contributes around 60% of tea output. KTDA's objective is to offer effective management services to the smallholder tea sector for effective production, manufacturing, and distribution of high-quality tea.(Mwaura & Muku, 2008). At the same time, the large-scale growers account for the remaining 40% of the total tea production (Gesimba et al., 2005). Tea is produced at high altitudes between 1500 and 2700 meters above sea level, with annual rainfall ranging between 1200 and 2700mm; g rowing tea requires a temperature range of 12 degrees Celsius to a maximum of 28 degrees Celsius, as well as a soil pH range of 4.5 to 6.5.(Cheserek et al., 2015). Tea is grown in several areas around Kenya, including Kericho, Bomet, Kiambu, Nandi, Limuru, and Trans-Nzoia, where it enjoys 85% favorable weather patterns and supports more than 600,000 growers (Karlsson, 2022).Kenya ranks third internationally after China and India in tea production with a market share of 9% as shown in Figure 1.1.he tea sector in Kenya also accounts for 20% of the tea exports globally, as shown in Figure 2, making it the leading exporter of tea globally(Monroy et al., 2019).

Kenya, Uganda, Malawi, Tanzania, and Zimbabwe are among the African tea-producing coun-

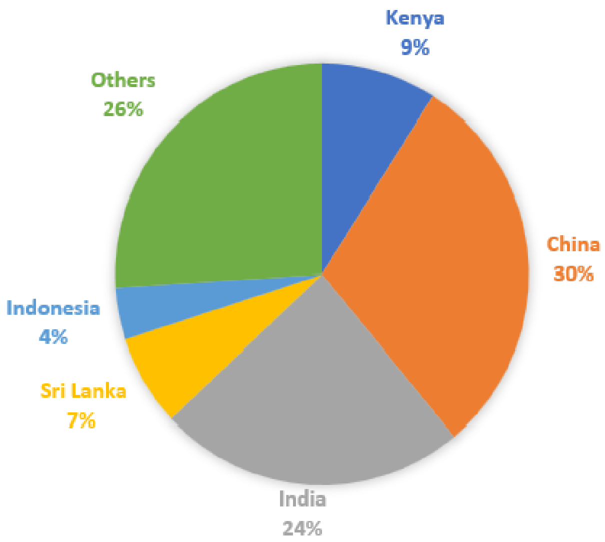

Figure 1.1: World Tea Production
Source: UN Comtrade

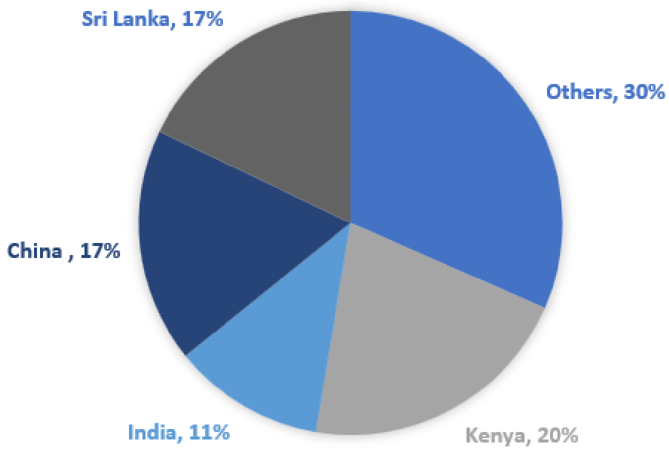

Figure 1.2: World Tea Exports
Source: FAOSTAT, KTB, and UN Comtrade.

tries, accounting for around 30 percent of global exports. Kenya's tea business accounts for 26 percent of Kenya's export revenues, with tea sent to Pakistan, Egypt, the United Kingdom, and other countries.as shown in Figure 1.3 (Muthamia & Muturi, 2015).

The government controls the tea industry through the Tea Board of Kenya (TBK) which must register tea growers, license factories that manufacture tea, register buyers and brokers, and tea consumption promotion (Kamunya et al., 2012). The government controls the tea industry through the Tea Board of Kenya (TBK), whose directors are elected directly by the stakeholders in the industry. The board must register tea growers, license factories that manufacture tea, register buyers and brokers, and tea consumption promotion (Kamunya et al., 2012). The purpose

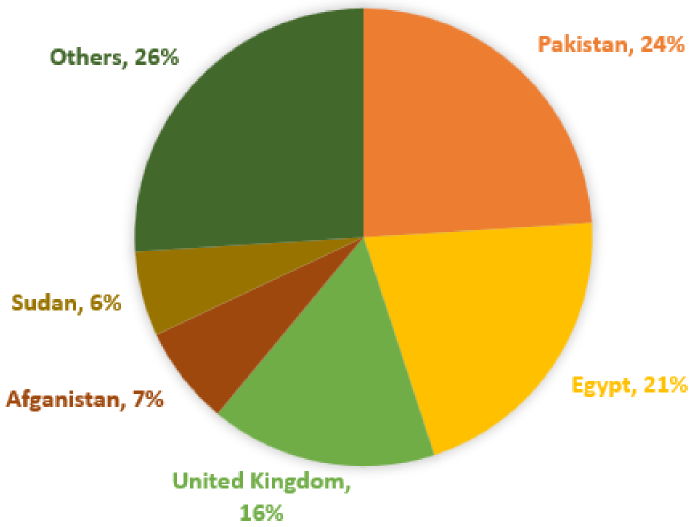

Figure 1.3: Kenya Tea Patners
Source: UN Comtrade

of this study is to model and forecast tea prices and their volatility in Kenya. Price fluctuation is one of the most significant concerns in the agricultural industry. Changing climatic factors such as hot and dry weather, political and economic instability in the nations that buy Kenyan tea, and competition from other tea exporting countries such as Sri Lanka and China are some of the causes of price volatility in Kenya's tea business (KIPPRA, 2020). GARCH models have shown to be adequate in modelling and forecasting volatility since it considers volatility clustering and time-varying conditional variances thus it gives reliable estimates of return variances and covariances (R. Engle, 2001). 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.

1.2 Basic Concepts

1.2.1 Volatility Definition and Measurement

Volatility is the irregular rising and falling observed in some events over time. It is referred to as conditional variance of the underlying return in modelling and forecasting literature. The sample standard deviation is used as its measure;

$$\sigma = \sqrt{\frac{\sum (r_i - \mu)^2}{N}} \tag{1.1}$$

Where σ denotes the standard deviation, r_i is the price return, and μ is the average price return over the N-period.

1.2.2 Volatility Clustering

Volatility clustering behaviour refers to the potential for huge changes in price to be accompanied by huge changes and low changes to be accompanied by low changes (Rachev et al., 2011). It is determined by enumerating the Autocorrelation function. Suppose X_t is a stationary series, the Autocorrelation function for the series is defined as:

$$\rho_k = \frac{Corr(X_t, X_{t-k})}{\sqrt{Var(X_t)Var(X_{t-k})}} = \frac{\gamma(k)}{\gamma(0)} \tag{1.2}$$

Where the value k denotes the lag.

1.2.3 Returns

Analyzing financial prices directly is challenging since prices are correlated, and their variances frequently increase with time; thus, we use returns or price changes for analysis.
Let Y_t and Y_{t-1} be the current and previous monthly prices of tea; then, the returns can be defined as:

$$r_t = \log \frac{Y_t}{Y_{t-1}} \tag{1.3}$$

1.2.4 Stationarity in Time Series

Stationarity implies homogeneity, which implies the series behaves similarly regardless of time. This means that its statistical properties of the series, do not change with time, i.e., independent of time. Therefore, in time series models, the stationarity assumption is very critical.

1.2.5 Differencing in Time Series

The method of making a non-stationary time series data stationary is by differencing. Differencing is particularly important in removing a trend and seasonal effect. For seasonal data, 1st order differencing is usually sufficient to obtain stationarity in the mean.
Let $X_t = X_1, X_2, ...X_n$ be a nonstationary time series, stationary time series can be obtained by:

$$\nabla X_t = X_t - X_{t-1} \tag{1.4}$$

The second-order difference is obtained by the operator ∇^2 where:

$$\nabla^2 X_t = \nabla X_t - \nabla X_{t-1} \tag{1.5}$$

$$= X_t - X_{t-1} - X_{t-1} + X_{t-2} \tag{1.6}$$

$$= X_t - 2X_{t-1} + X_{t-2} \tag{1.7}$$

However, if we are dealing with seasonal data where d is the period of the season, then:

$$\nabla^d X_t = X_t - X_{t-d} \tag{1.8}$$

1.2.6 Seasonal Autoregressive Integrated Moving Average (SARIMA) model

A multiplicative SARIMA model X_t can be denoted as $SARIMA(p, d, q)(P, D, Q)_s$; where (p, d, q) are the non-seasonal orders, (P, D, Q) are the seasonal orders. S is the seasonal period; for example, $s = 12$ represents seasonal fluctuations that occur every 12 month.

1.2.7 Generalized Autoregressive Conditional Heteroscedasticity (GARCH) Model

Bollerslev (1986) extended the $(ARCH)$ Model by coming up with the $GARCH$ model. 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 1.9.** The integrated model is known as $GARCH(p, q)$ model as given in **Equation 1.10**.

$$r_t = \sigma_t \epsilon_t \tag{1.9}$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i r_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \tag{1.10}$$

σ_t^2 on **Equation 1.10** is the conditional volatility, r_{t-1}^2 are 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 coefficients α_0, α_j and β_j are supposed to be non-negative.

2. LITERATURE REVIEW

2.1 Introduction

This part is broken down into three sections. The theoretical literature review, which deals with tea price volatility, followed by the empirical literature review, which discusses prior investigations on SARIMA and GARCH models. Finally, the overview summarizes the literature review.

2.2 Theoretical Literature Review

According to Gesimba et al. (2005), the tea industry encounters a number of issues. The main challenge is the volatility of tea's export price, which he suggests can be overcome by regulating the demand and supply in the global market. The global prices of tea, similar to almost all agricultural commodities, are noticeable by relatively increased volatility. The price volatility affects decisions for investment, food security, and the income of households. China, India, Sri Lanka, and Kenya are the major tea-producing countries employing more than 13 million farmers, whereas around 9 million are small-scale farmers (Qiao et al., 2016). Some of the factors that are responsible for price volatility include shocks to demand and supply, which can are brought by changes in the climatic conditions, i.e., cold or hot and dry weather, political and economic instabilities by the countries that import tea, competition from other exporting countries and finally macroeconomic factors that have cross-sectional impacts such as population growth rates and currency exchange rates also lead to price volatility (KIPPRA, 2020).

The least-squares model asserts that the mean value of the squared error component is constant at every given point. In cases where the error term variance is not constant, it is described as heteroscedastic. The models ARCH and GARCH consider heteroskedasticity as a modelable variance. In contrast, heteroskedasticity in ordinary least squares renders the regression coefficients unbiased, while the standard errors and confidence intervals are estimated to be too narrow. Thus, ARCH and GARCH models correct the least-square regression deficiencies and compute the prediction for each error term (R. Engle, 2001).

The ultimate purpose of volatility analysis is to explain the reasons of volatility. The purpose of the ARCH and GARCH models is to assess volatility so that it may be used in financial choices including risk analysis. The GARCH (1,1) model is the most basic and powerful of the volatility models; GARCH (p, q) is the generalized model for GARCH (1,1), which means it has extra lag terms. Such high-order models are frequently utilized when dealing with long-term data sets, such as daily data for several decades or hourly data for a year. (R. Engle, 2001).

2.3 Empirical Literature Review

Kibunja et al. (2014) examines a linear model and defines a SARIMA model for precipitation forecasts in the Mount Kenya region. From the Autocorrelation function and the Partial Autocorrelation function plots of precipitation, it was clear that there was stationarity in the data, and a SARIMA model was fitted. He obtained the best model by selecting the model that produced minimum AIC and BIC, and in addition, MAE, RMSE, and ME were also used in model selection. The model $SARIMA(1,0,1) * (1,0,0)_{12}$ turned out to be the finest model of the three tentative models. From the analysis the model yielded the following evaluation statistics ME = -0.005369, RMSE = 0.9838 and MAE = 0.7520. As a result, the SARIMA model was determined to be a good model for forecasting precipitation in the Mt. Kenya region.

Another research by Mutwiri (2019) used the SARIMA model to predict the wholesale price of tomatoes in Nairobi, Kenya. The series was stationary after the ADF test was conducted, indicating that the series did not have a unit root. $SARIMA(2,1,1) * (1,0,1)_{12}$ was picked as the best model since it yielded the least Akaike Information Criterion. The normality test and residual Q-Q plots were also conducted to see if the residuals followed a white noise procedure. The Q-Q plot, which was straight, proved that normality was good, and the Autocorrelation function (ACF) plot, whose first 20 lags were all in the 95% confidence interval bounds, indicated that the residuals were random. MAPE = 125.251, RMSE = 32.063, and MAE = 22.3 were obtained from predictive ability tests, proving that the SARIMA model was suitable for forecasting tomatoes in Nairobi, Kenya. The study's findings suggested that the series is stable, which means that tomato prices in Nairobi County do not fluctuate.

The considerable use of linear models is not surprising. They give a good first-order approximation to most processes. However, linear time series models are not powerful in describing certain features of a volatility series. Because the assumption of homoscedasticity is not suitable when utilizing financial data, they cannot explain numerous major phenomena common to various financial data, such as volatility clustering and leptokurtosis. Thus, looking at patterns that allow the variance to depend upon the past is preferable. Thus, such challenges of linear models have prompted many researchers to consider nonlinear alternatives. R. F. Engle (1982) proposed the ARCH model, and the GARCH model was proposed by Bollerslev (1986). Nelson (1991) introduced the Exponential Generalized Autoregressive Conditional Heteroscedasticity (EGARCH) model. These are some of the commonly used nonlinear models used in financial modelling.

In another study by Chinomona (2009), the modelling of South African inflation data was conducted. The $SARIMA(1,1,0) * (0,1,1)_{12}$model was selected from the ARIMA models, and GARCH (1, 1) was selected from the ARCH-GARCH models. The AIC for $SARIMA(1,1,0)* (0,1,1)_{12}$ was 411.03, and the BIC was 423.7, while GARCH (1, 1) had an AIC of 351.5 and a BIC of 376.7. GARCH (1, 1) had the least value of AIC and BIC, which shows that it explained the variance in the series better than $SARIMA(1,1,0) * (0,1,1)_{12}$. GARCH (1, 1) was thus selected to be the best model compared to $SARIMA(1,1,0) * (0,1,1)_{12}$ due to its ability to capture variations in the series.

Another study by Fwaga et al. (2017) used both GARCH and the EGARCH model to model the inflation in Kenya. The model with the least value of AIC and BIC was selected as the best model. EGARCH (1,1) was picked as the model with the least values of AIC and BIC, followed by the GARCH (1,1). EGARCH (1,1) was then used in conducting the model diagnostic test to test for the goodness of fit. Ljung Box test showed an absence of correlation in the model on the residuals test. The Q-Q plot for normality produced a fairly straight line indicating that they are normally distributed. EGARCH was thus selected as the best model for predicting future inflation.

Because they allow for a logical interpretation of nonlinearities, Markov-Switching models have also been used to understand the volatility dynamics of a financial time series. Markov-Switching models were originally proposed by Turner et al. (1989). Maiyo (2018), employed the Markov Regime Switching GARCH model instead of univariate GARCH models in modeling and forecasting tea price volatility. The results indicated that the Markov Regime Switching GARCH models may reduce the GARCH models' high persistence and outperform the univariate GARCH model in modeling volatility.

The developed countries have conducted many studies on volatility. Al-Jafari (2012), for example, investigated the day of the week influence on the Muscat securities market using two asymmetric models, TARCH(1,1) and EGARCH(1,1), and one symmetric model, GARCH(1,1). The data revealed that the day of the week effect did not exist.

Volatility modeling has received little attention in Sub-Saharan Africa. Among the investigations is Tendai and Chikobvu (2017), which simulated worldwide visitor arrivals and volatility to Zimbabwe's Victoria Falls rainforest. Under the normal distribution of errors, the $SARIMA(1,0,0)(0,1,1)$ $GARCH(1,0)$ model proved to be the best for modeling international tourist arrival volatility, and the model quantifies uncertainty in the future. In another research, Winrose (2018), used the autoregressive distributed lag (ARDL) model in the error correction form to evaluate the impact of price and exchange rate volatility on Kenya's black tea export. The study discovered that in the long and short term, increase in foreign income, price fluctuations, and exchange rate volatility were significant.

2.4 Conclusion

GARCH models are adequate in modelling and forecasting volatility since they consider volatility clustering. Their capacity to predict time-varying conditional variances allows them to forecast variances and covariances of returns accurately. GARCH is also widely used because it minimizes forecasting errors by taking care of errors in the previous forecasts and further improving the accuracy of ongoing predictions; therefore, in situations where return volatility is a critical threat, the application of GARCH models is widespread. When attempting to anticipate prices, financial professionals frequently choose the GARCH process since it gives more real-world context than other models. SARIMA and GARCH models are employed in this work to predict the seasonal and heteroscedastic character of volatility in the market price of tea in Kenya.

2.5 Statement of the problem

According to Gesimba et al. (2005), the tea sector is facing a number of challenges. The main threat is the volatility or fluctuations of the export price of tea, which he suggests can be overcome by regulating the demand and supply of tea in the world market. As a result, price volatility is one of the most significant challenges in the agriculture sector. Volatility in commodities is not readily visible, but it exhibits features in the returns of a time series. Some of the causes of price volatility in Kenya's tea sector are changing climatic conditions such as cold or hot and dry weather, which lead to a decrease in the production and supply of tea in Kenya, and political and economic instabilities in the countries that import Kenyan tea and competition the tea exporting countries such as Sri Lanka, China (KIPPRA, 2020). GARCH models are adequate in modelling and forecasting volatility since they consider volatility clustering. Their capacity to predict time-varying conditional variances allows them to forecast variances and covariances of returns with high accuracy (R. Engle, 2001). The 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. When attempting to anticipate prices, financial professionals frequently choose the GARCH process since it gives more real-world context than other models.

2.6 Research Objective

2.6.1 Main Objective

The main aim of this study was to forecast tea prices and their volatility in Kenya.

2.6.2 Specific Objectives

1. To determine the optimal SARIMA and GARCH models on the monthly price and returns of tea in Kenya.

2. To diagnose the goodness of fit of the residuals from the SARIMA and GARCH models on the price of tea in Kenya.

3. To predict monthly tea prices and volatility in Kenya using the SARIMA and GARCH models, respectively.

2.7 Significance of the Study

The agriculture industry is critical to achieving Kenya's Vision 2030. The tea sub-sector contributes significantly to the Kenyan economy's external performance through foreign exchange profits and the employment of over 500,000 small-scale farmers who produce this crucial commodity. As a result, the success of this sector is critical to the economy's overall growth. The impetus for doing this research stems from the issues confronting the tea industry, notably the insecurity of growers' revenues. The results of this study will be used by the Tea Board of Kenya (TBK) in financial choices about price risk analysis and in planning for procurement needs because they will be able to estimate prices and defend against the negative consequences of price volatility. This will contribute to increased tea output, resulting in more tea being exported and hence more money for tea-producing farmers and an increase in the country's foreign exchange profits and GDP.

3. METHODOLOGY

3.1 Introduction

A time-series technique is used to model and forecast global price of tea in Kenya. The method comprises of cleaning the data, charting the series to discover patterns in the monthly price of tea in Kenya, unit root testing for stationarity, defining the order of the model, fitting the necessary SARIMA and GARCH models, and lastly forecasting future tea price values. Python, E-VIEWS, and R were used for the analysis.

3.2 Data

The data used for this research was acquired from the Federal Reserve Economic Data and consists of 381 observations of the monthly price for a Kilogram of tea in Kenya from January 1, 1990 to November 11, 2021.

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 3.1: A Sample of the Data

Table 3.1 is the sample of the monthly price of tea per kilogram in Kenya recorded in Kenyan shillings.

3.2.1 Data cleaning

The data is cleaned to make it easier to read and employ during modeling and analysis. The procedure consists of four steps: first, the date column is set as the index, then the data type of the date column is transformed from object type to DateTime, and its frequency is set to monthly start. The tea prices column is then renamed, and any missing values are examined.

3.2.2 Test for Autocorrelation and Checking for Outliers

The lag plot is plotted 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; if they exist, they are dropped. Autocorrelation indicates that an auto-regressive model is more suitable for the monthly price of tea.

3.2.3 Tests for Stationarity.

The test for stationarity is conducted visually and inferentially. Visually is done using the plot of Auto-correlation and Partial Autocorrelation plots. Inferentially it is conducted using the Augmented Dickey-Fuller (ADF) Test. A time series with a unit root is nonstationary, and the ordinary least squares estimator of the series is not normally distributed. If the test statistic is greater than the crucial value, the null hypothesis is rejected. Following stationarity, autocorrelation and partial autocorrelation are investigated to determine their suitable structures. It is carried out using the Ljung-Box Q-statistic test Ljung and Box (1978), which is defined as:

$$Q = n(n+2)\sum_{j=1}^{m}\frac{r_j^2}{n-j} \tag{3.1}$$

From **Equation 3.1** r_j is the accumulated sample autocorrelations, n is the sample size.The sample size is n, and the maximum lag time is m.

3.2.4 Differencing the Series

Differencing is important in removing a trend and seasonal effect in the data before presenting it to the model. Our series is made stationary by taking the first order differencing since it is seasonal data.

3.2.5 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, season, and residuals which are the irregular variations in the series. The decomposition of the monthly price of tea in Kenya is conducted, and the observed series, trend, seasonal and random components are observed.

3.3 SARIMA Model

3.3.1 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 follows;

$$\Phi_p(B^s)\phi(B)\nabla_s^D\nabla^d X_t = \delta + \Theta_Q(B^s)\theta(B)w_t \tag{3.2}$$

From **Equation 3.2** X_t is the monthly price of tea in Kenya, w_t is the white noise process. The ordinary autoregressive and moving average components are represented by the polynomials $\phi(B)$ and $\theta(B)$ of orders p and q, 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 $\nabla_s^D = (1 - B^s)$ (Shumway et al., 2000).

3.3.2 Model Identification

Model identification and selection are based on the Akaike Information Criterion (AIC) proposed by Akaike (1974). Akaike Information Criterion approximates each model's quality given a collection of models for data, and the model with the most negligible AIC value is considered the best model. The formula for the Akaike Information Criterion (AIC) is:

$$AIC = 2K - 2\ln(L) \tag{3.3}$$

From **Equation 3.3**, K represents the number of model parameters estimated, while L is the value of the maximum likelihood function for the model.

3.3.3 Coeffiecient Estimation

The SARIMA model coefficients are estimated from the best model with the least value AIC. The coefficients are further evaluated using their respective p-values to check for significance.

3.3.4 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. The normal Q-Q plot, histogram, correlogram, and density plot are used to visually assess the residuals' normality. The Jacque-Bera test is used to execute the normalcy test inferentially. In this scenario, the Jarque-Berra Test is used to determine if the data includes skewness and kurtosis, with the null hypothesis asserting that the residual is normally distributed. The Jarque-Bera test statistic formula by Jarque and Bera (1980) is:

$$JB = \frac{n}{6}(S^2 + \frac{(K-3)^2}{4}) \tag{3.4}$$

From **Equation 3.6**, n is the number of observations or the degrees of freedom, S is the sample skewness, and K is the sample Kurtosis.

3.3.5 Forecasting for future periods

The model fitted to the training data forecasts future periods of equal size as in the test set. The forecasted values are plotted against the actual values in the test set to check for the model's performance. The full data is then fitted to the full data and used in forecasting for ten years ahead

3.3.6 Evaluating the SARIMA model

The Mean Absolute Error (MAE) and the 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 the MAE is given in **Equation 3.4** :

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|Y_t - F_t| \tag{3.5}$$

The formula for the RMSE is given **Equation 3.5**:

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(Y_t - F_t)^2} \tag{3.6}$$

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

3.4 Volatility Modelling

3.4.1 Returns

Let X_t denote the current month's price of tea and X_{t-1} denote the previous month's price of tea; the monthly return series were obtained using **equation 3.7**

$$r_t = \ln\left(\frac{X_t}{X_{t-1}}\right) \tag{3.7}$$

The descriptive statistics for the returns are also conducted to have a better understanding of the return series and its distribution.

3.4.2 Test for Heteroscedasticity

The heteroscedasticity test is used to determine the presence of ARCH effects, which occur when the variance is not constant over time. The ARCH-Lagrange Multiplier test proposed by R. F. Engle (1982) is conducted to determine 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 states that no arch effects exist, while the alternative hypothesis states that an arch effect exists. The Lagrange Multiplier test statistics are computed using **Equation 3.8**:

$$\mathcal{L}(x, \lambda) = f(x) - \lambda g(x) \tag{3.8}$$

3.5 GARCH Model

To estimate the variance at time t, a $GARCH$ model employs previous squared residuals and variances. The $GARCH(p,q)$ model combines the $ARCH(q)$ model and adds p lags of conditional variance to the model, where p is known as the $GARCH$ order. In a GARCH model, the return is provided by equation (3.12). The combined model is known as the generalized autoregressive conditional heteroscedasticity, or $GARCH(p,q)$ model, as expressed in equation (3.13):

$$r_t = \sigma_t \epsilon_t \tag{3.9}$$

$$\sigma_t^2 = \alpha_0 + \sum_{j=1}^{p} \alpha_j r_{t-j}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \tag{3.10}$$

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.

To ensure that the conditional volatility $\sigma_t^2 > 0$ all coefficients 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 3.14**:

$$\sigma_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \tag{3.11}$$

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 Coefficients α_0, α_1, and β_1 are non negative this ensures that the conditional volatility $\sigma_t^2 > 0$ at all times

3.5.1 Model Identification

After the Arch effects are proven to be present in the series, the GARCH model is identified. The $GARCH(p,q)$ model's maximum parameter value for p and q is 2. The parameter q, however, cannot be zero. All the possible combinations of the $GARCH(p,q)$ model that is simulated on the data are; $GARCH(0,1)$, $GARCH(0,2)$, $GARCH(1,1)$, $GARCH(1,2)$, $GARCH(1,2)$, $GARCH(2,1)$, and $GARCH(2,2)$. The optimal model is found by examining the model with the lowest AIC and BIC values, and it is used to estimate the model coefficients for the monthly tea price in Kenya.

3.5.2 Coefficient Estimation

The model with the lowest AIC and BIC value is fitted to the monthly tea returns. Under the normal distribution, the maximum Likelihood Method calculates the model's coefficients. Each model coefficient's p-value is also evaluated for significance. Significant coefficients have a p-value of less than 0.05. The resulting GARCH model coefficient values are replaced in the model. **Equation 3.14**.

3.5.3 Standard Residuals and Conditional Volatility

The Standard Residuals and Conditional Volatility are the coefficients of our GARCH model, and they are plotted alongside each other to observe their behavior over time. The Conditional Volatility plot should also exhibit volatility modelling features such as volatility clustering, which is turbulent periods followed by turbulent periods and vice versa. The standard residuals and conditional volatility are expected to behave the same; periods with high standard residuals should also have high volatility and vice versa.

3.5.4 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 Jacque-Bera test is used to determine the normality of the data. In this scenario, the Jarque-Berra Test is used to determine if the data includes skewness and kurtosis, with the null hypothesis asserting that the residual is normally distributed. The Jarque-Bera test statistic formula by Jarque and Bera (1980) is:

$$JB = \frac{n}{6}(S^2 + \frac{(K-3)^2}{4}) \tag{3.12}$$

In **Equation 3.12**, n represents the number of observations or degrees of freedom, S represents the sample skewness, and K represents the sample Kurtosis.

3.5.5 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 is used to forecast 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 the volatility of tea in Kenya. To evaluate the model's performance, the estimated values are plotted against the actual values in the test set.

3.5.6 Evaluation of Forecasts

The GARCH(1,1) model's predictions on the monthly price of tea in Kenya are assessed using the Mean Absolute Error (MAE) and the Root Mean Squared Error (RMSE). The formula for the MAE is given in **Equation 3.4** while the formula for the RMSE is given in **Equation 3.5**.

4. RESULTS AND DISCUSSION

4.1 Data

To model and forecast the prices of tea and their volatility in Kenya, monthly prices per kilogram, in Kenyan shillings, data for the period January 1, 1990, to November 1, 2021, obtained from Federal Reserve for Economic Data (FRED), is used. A sample of the data is shown on **Table 4.1** with only ten observations 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 4.1: Sample of the Data

4.2 Descriptive statistics of the data

Descriptive statistics of the series are shown in **Table 4.2** where the average price of tea over the years is Ksh. 245.80 with a standard deviation of Ksh. 60.70. The series also shows positive skewness of 0.591, meaning that the distribution has a long right tail and kurtosis of 2.370. The minimum price for tea that has been sold is Ksh. 143.37 per kilogram on July 1, 1995, while the maximum price tea has been sold is Ksh. 403.03 per Kilogram, which was on July 1, 2015.

Table 4.2: Descriptive statistics of the price of tea

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

4.3 Plot of the data

Figure 4.1 depicts the data plot from 1990 to 2021; it is clear that the series trend has been increasing over time. Seasonal and cyclic patterns can also be observed over time. There is a sudden rise and fall of tea prices.

It can also be observed from **Figure 4.1** that the data is non-stationary, that is, the statistical properties are not constant over time thus differencing will be done to make the data stationary.

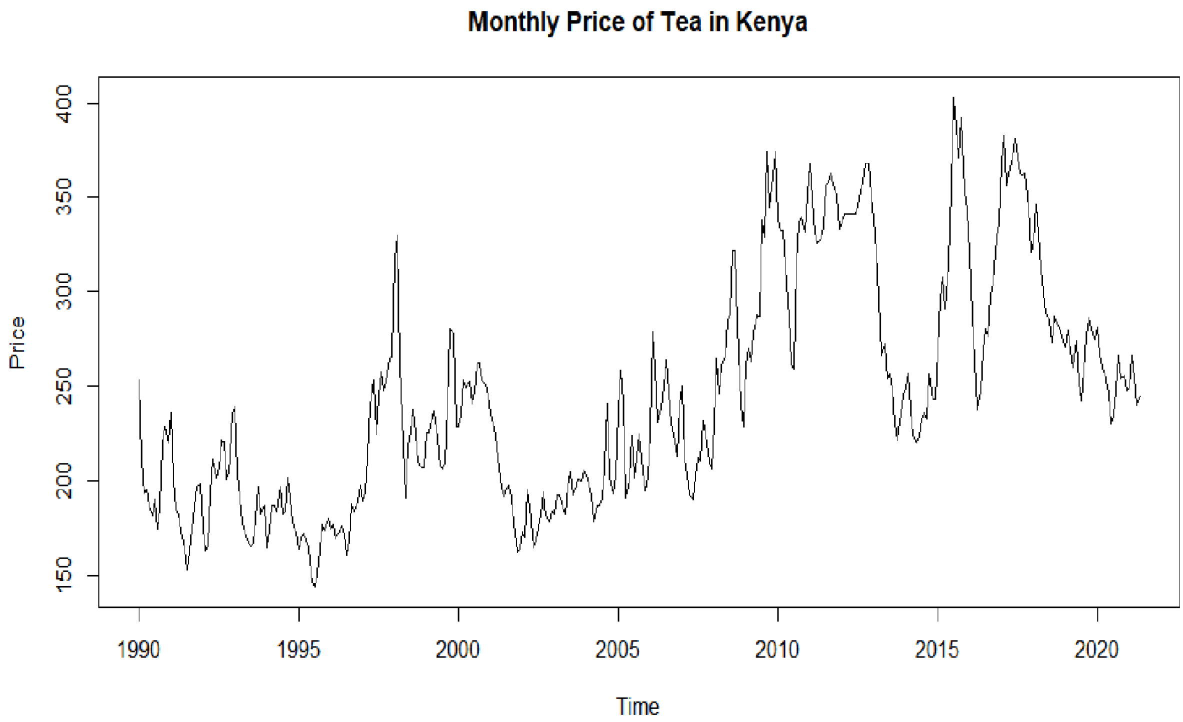

Figure 4.1: Time series plot 1990-2021

4.4 Test for Autocorrelation and Checking Outliers

A lag plot, as illustrated in **Figure 4.2**, was generated prior to the stationarity test to assess for autocorrelation and the existence of outliers.

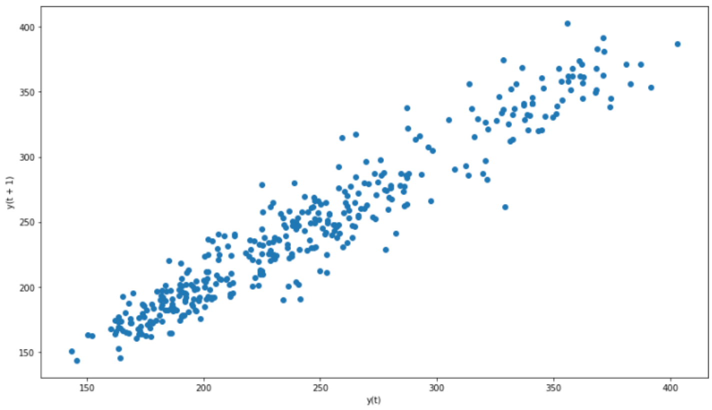

Figure 4.2: Lag plot of the data

It is evident from **Figure 4.2** that no outliers are present in the data, and there is a positive serial auto-correlation presence. Thus an auto-regressive model is more suitable.

4.5 Test for Stationarity

The stationarity test was performed in two ways: inferentially and visually. It was carried out inferentially using the ADF test for stationarity, and visually using the ACF and PACF plots.

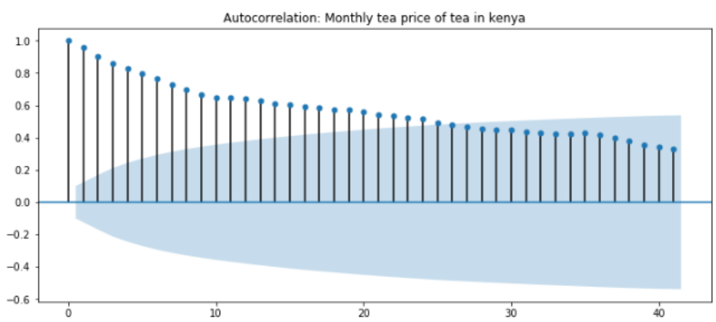

Figure 4.3: ACF plot of the data

Figure 4.3 is the ACF plot indicating 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 4.4**, it can be seen that the first and second lags are significant correlations confirming the presence of auto-regressive terms in the series.

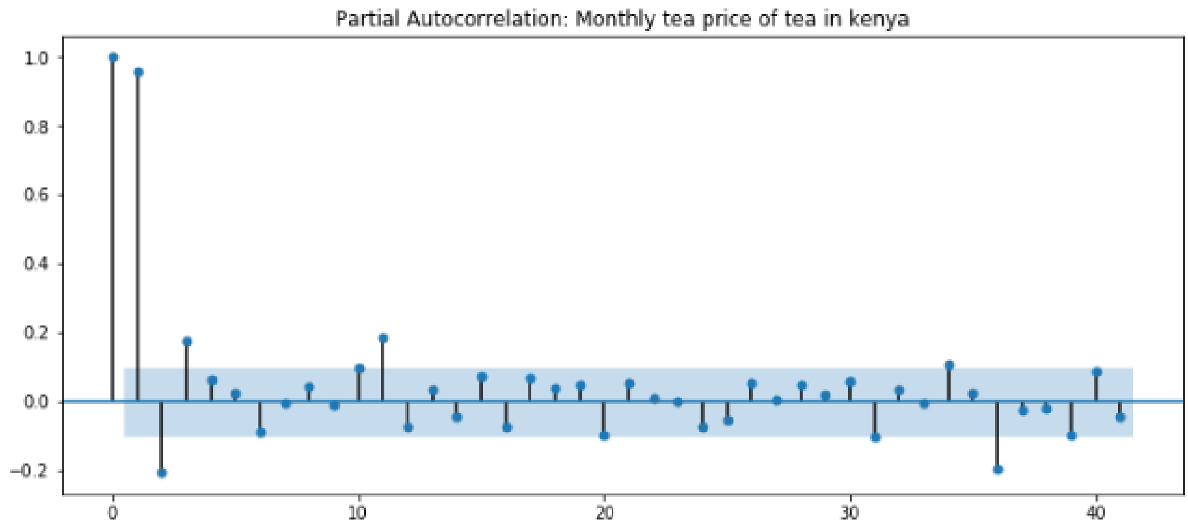

Figure 4.4: PACF plot of the data

According to **Table 4.3**, the ADF test statistic is (-1.8424) and the p-value is 0.3596, which is larger than 0.05, confirming the presence of a unit root and indicating that the data is non-stationary.

Table 4.3: Results of the ADF test

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

4.6 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 4.5** is the decomposition of the monthly price of tea in Kenya into observed, trend, seasonal, and random components.

Based on **Figure 4.5**, we can see a general upward trend in the price of tea. Seasonality is also observed since this is monthly data. The residuals also show volatility clustering, which is defined as periods of high and low volatility.

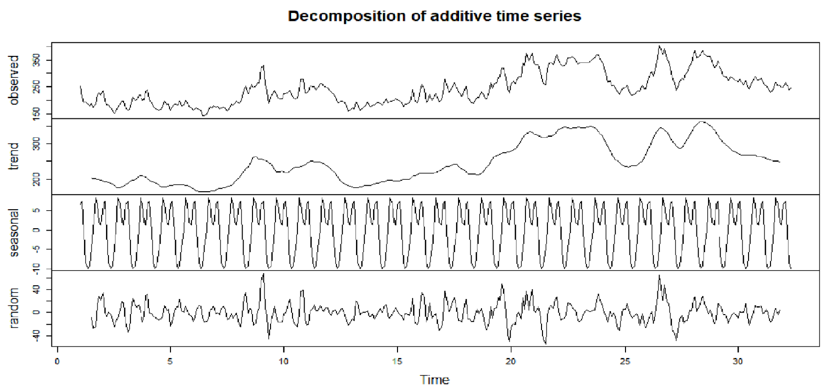

Figure 4.5: Decomposition of the series

4.7 The SARIMA model

4.7.1 SARIMA 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 4.6** is the plot of the stationary data, and clearly, it exhibits constant mean and variance over time.

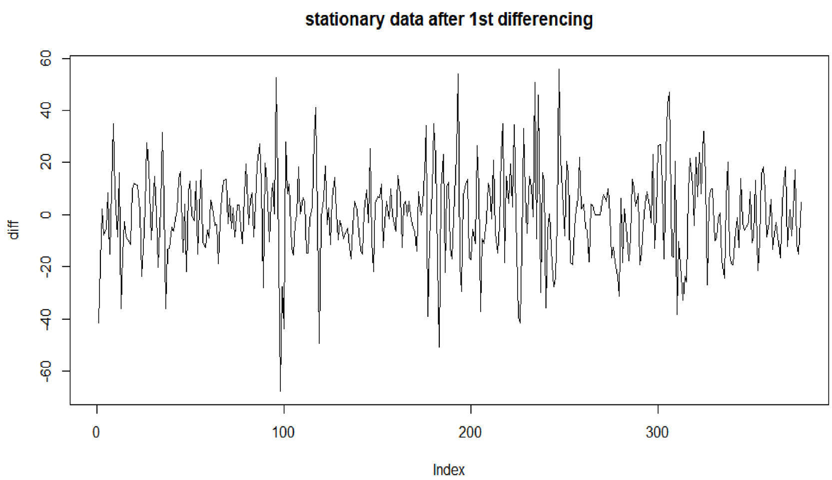

Figure 4.6: Stationary data

Since the data is stationary, the Extended Autocorrelation Function (EACF) is used to identify the AR and MA parameters. The results of EACF are shown on **figure 4.7**

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 4.7: Extended ACF

From the results in **Figure 4.7**, an ARMA(3,3) model was chosen as the best model which will be fitted along with other tentative models.

4.7.2 SARIMA Model Identification.

The trial and error method was used in estimating the best SARIMA model. Some of the potential SARIMA models are shown on **Table 4.4**

Table 4.4: The AIC of the models fitted

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

Based on **Table 4.4** 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.

4.7.3 SARIMA Model Coefficient Estimation

Before fitting the model, the data was split into 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, and the test set will be used during forecasting.

Table 4.5: Sarima model on the train set

	Coefficients	**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

From the results of **Table 4.5**, it can be see that all the model coefficients are significant since they are all less than 0.05 thus the model is significant.

4.7.4 SARIMA Model Residual Diagnostics

To further assess the goodness of fit of our model, residual diagnostics were performed. The normalized residuals, histogram plus estimated density, normal q-q plot, and correlogram in **Table 4.8** were used to test the SARIMA model's adequacy.

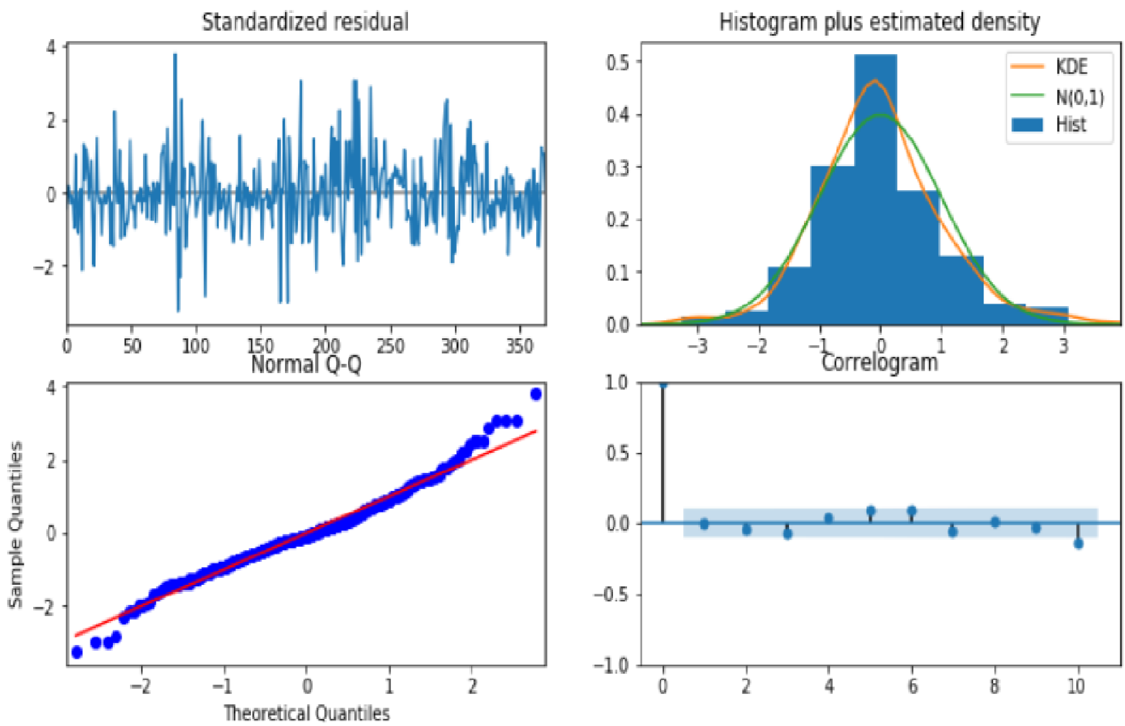

Figure 4.8: SARIMA Model Residual Diagnostic plot

Based on **Figure 4.8**, the standard residuals have a constant mean, and the variance follows a normal distribution, as seen in the histogram. The normal probability plot demonstrated that the data has a normal distribution, and the correlogram revealed no significant lag, indicating that the model was appropriate.

4.7.5 Forecasting using the test set

The test set was used in validating the forecasts made for the monthly price of tea for the 36 months using the SARIMA model created using the training set.

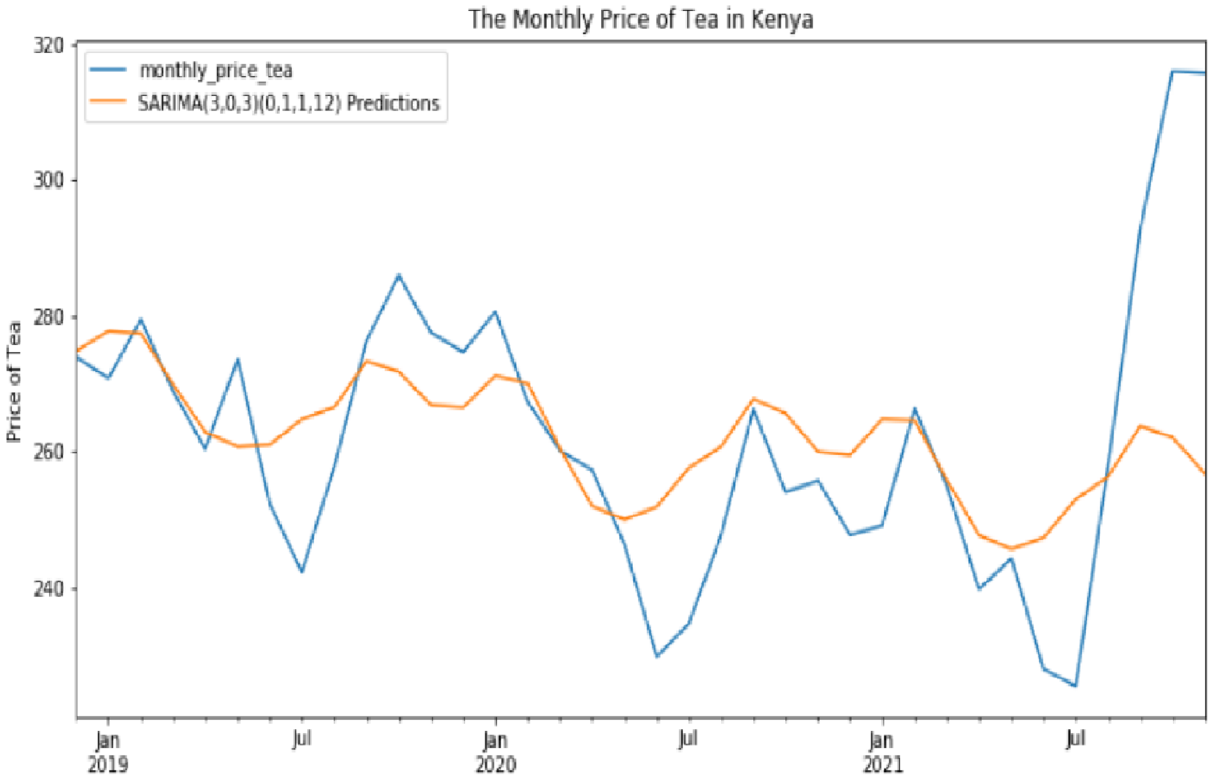

Figure 4.9: Forecasts for 36 months ahead

The actual values in the test set are shown against the forecasted values from the model in **Figure 4.9**.

4.7.6 Forecasting monthly tea prices to June 2024

The SARIMA model was fitted to the entire dataset and used to forecast monthly tea prices for the next two years, 2022-2024.

Figure 4.10 was the plot of the two-year forecasts for the monthly price of tea in Kenya. **Table 4.6** is the table containing the predicted monthly price of tea from December 2022 to June 2024

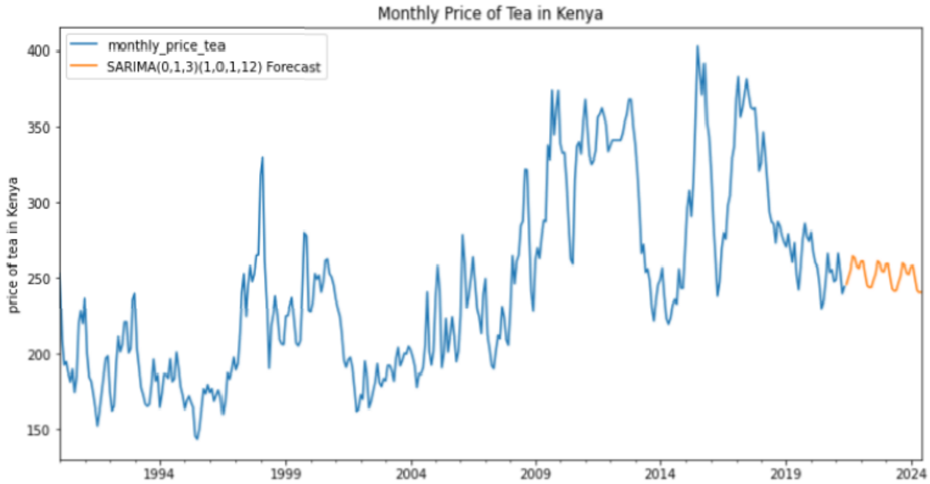

Figure 4.10: Forecasts for 10 years ahead

Date	Prediced price
2022-12-01	259.24
2023-01-01	259.33
2023-02-01	254.25
2023-03-01	250.32
2023-04-01	242.72
2023-05-01	241.38
2023-06-01	241.79
2023-07-01	247.01
2023-08-01	251.47
2023-09-01	259.84
2023-10-01	258.23
2023-11-01	253.23
2023-12-01	252.68
2024-01-01	257.70
2024-02-01	258.01
2024-03-01	249.53
2024-04-01	241.71
2024-05-01	240.36
2024-06-01	240.69

Table 4.6: Tea Price forecasts from December 2022 to June 2024

4.7.7 Evaluating the SARIMA Model

The Root Mean Squared Error and Mean Absolute Error were used to assess the SARIMA model's predicting ability on the test data.

Table 4.7: Forecasting performance of the SARIMA Model

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

Based on the forecast performance metrics obtained on **Table 4.7** the $SARIMA(3,0,3)(0,1,1)_{12}$ model's prediction had a RMSE of 17.90, MAE of 11.90, and MSE of 320.66. We can thus conclude that the SARIMA model performed well on the monthly price of tea.

4.8 Volatility Modelling

4.8.1 Plot of the Returns

Figure 4.11 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 4.11: Plot of the return series

4.8.2 Descriptive statistics of Returns

The descriptive statistics of the monthly returns of tea were calculated so as to have a better understanding of our return series.

Table 4.8: Descriptive statistics of the monthly Returns

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

Based on **Table 4.8** the minimum return was -21.038 while the maximum was 23.956, the mean was 0.346 and the standard deviation 6.974. The skewness of 0.224 indicates that the returns are moderately skewed.

4.8.3 Test for Heteroscedasticity

The test for Heteroscedasticity is done to check for the presence of arch effects in the squared residuals. It is based on the methodology's Lagrange Multiplier of **Equation 3.8**.

Returns	Chi-square	P-Value
Tea	3.875	0.0490

Table 4.9: ARCH-LM Test for Heteroscedasticity

According to **Table 4.9**, the ARCH-LM test statistic is 3.875 with a P-Value of 0.0490, therefore we reject the null hypothesis and conclude that there are substantial Arch effects from January 1, 1990 to November 1, 2021. As a result, we can figure out the best GARCH model.

4.9 GARCH Model

4.9.1 Model Identification

The maximum parameter value p and q, a $GARCH(p, q)$ model can take is 2, although q cannot be zero. The best model is identified by simulating all the possible combinations of the model and picking the one with the least value of both AIC and BIC.

Table 4.10: AIC and BIC of GARCH models

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

Based on **Table 4.10** different GARCH(p,q) models are simulated and from the results, GARCH(1,1) was the best model on the monthly price of tea data, with an AIC of 2542.47, and a BIC of 2558.24. The parameters of GARCH(1,1) model are thus estimated for the price of tea in Kenya.

4.9.2 Coefficient Estimation

Coefficient estimation was conducted as shown on **Table 4.11**. It can be seen that the p-values are all less than $\alpha = 0.05$. Thus we conclude that the coefficients of the GARCH(1,1) model are significant; thus, GARCH(1,1) model is the best model.

Table 4.11: GARCH(1,1) Coefficient estimation

	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

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

$$\sigma_t^2 = 2.9566 + 0.1527 r_{t-1}^2 + 0.7952 \sigma_{t-1}^2 \tag{4.1}$$

From **Equation 4.1** r_{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. $\alpha_1 = 0.1527$ is low indicating stable long-term volatility and $\beta_1 = 0.7952$ indicating that the tea's conditional variance will not last before decaying. Therefore the tea sector in Kenya needs to deal with persistent volatility.

4.9.3 Squared Residual and Conditional Volatility

The squared residuals and conditional volatility were plotted as shown in **Figure 4.12** from the GARCH(1,1) model.

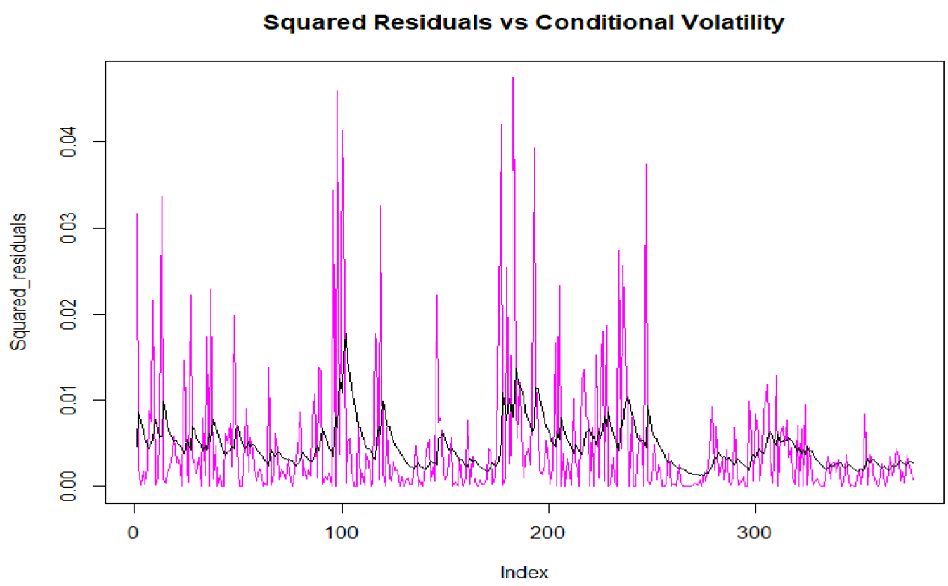

Figure 4.12: Squared Residuals vs Conditional Volatility

Based on **Figure 4.12**, 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 while periods of high volatility are followed by periods of high volatility.

4.9.4 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.

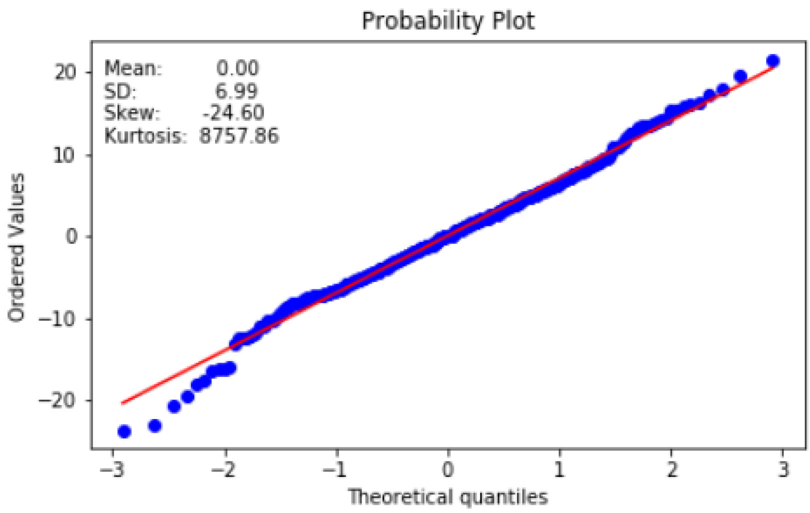

Figure 4.13: Probability Plot of the Residuals

From **Figure 4.13**, we can conclude that the standardized residuals follow a normal distribution despite the small divergence at the start of the plot.

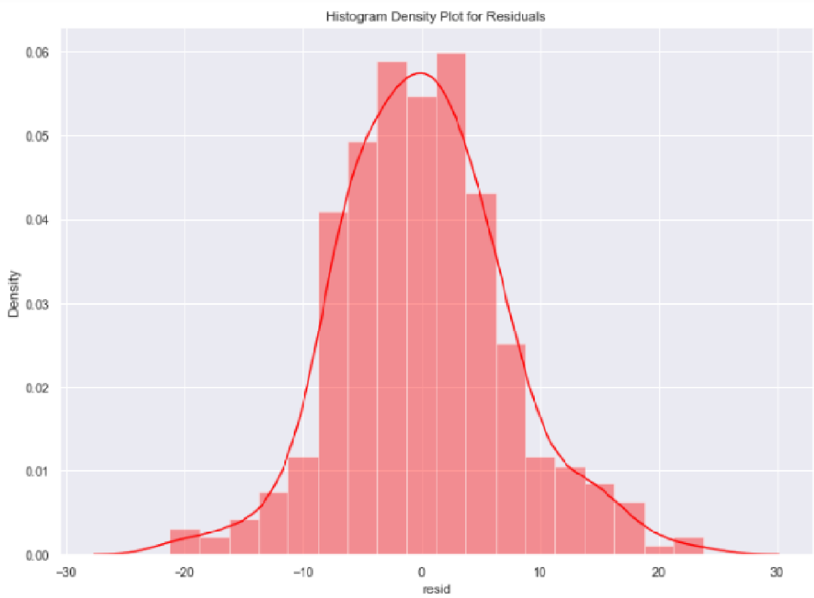

Figure 4.14: Histogram Density Plot of the Residuals

The histogram density plot in **Figure 4.14** also confirms that the standardized residuals follow a normal distribution. Since the residuals were significant, we proceeded to volatility forecasting.

4.9.5 Volatility Forecasting

Before volatility forecasting was conducted, the data was split into the train set and test set. The train set contained 80% of the data while the test set 20% of the data. The train set was used to fit the model and forecast 36 months ahead. The test set is used to evaluate the model forecasts.

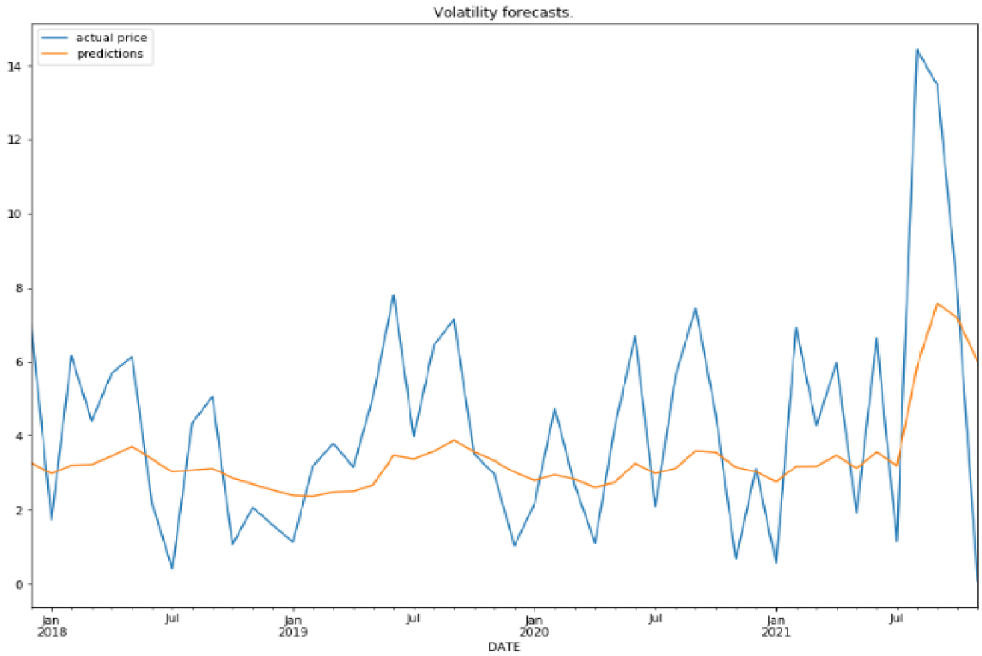

Figure 4.15: Actual data vs. Forecasts for Volatility

From **Figure 4.15**, we see that the forecast data pattern and the actual data pattern are almost similar. Thus we proceeded to evaluate the forecasts for the GARCH(1,1) model.

4.9.6 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).

Table 4.12: Forecasting performance of the GARCH(1,1) Model

Model	RMSE	MAE
$GARCH(1,1)$	5.903	5.089

Based on **Table 4.12**, the RMSE (5.903) and the MAE (5.089) values are significant. Thus we conclude that GARCH(1,1) model performed well.

5. Conclusion and Recommendation

The SARIMA model was used to forecast the monthly prices of tea in Kenya up to june 2024, while the GARCH model was used to estimate and forecast price volatility using the monthly return series derived from the monthly price of tea. The $SARIMA(3,0,3)(0,1,1)12$ model was chosen as the best model for forecasting tea prices in Kenya using the EACF plot, which supplied the auto-regressive and moving average parameters for the model. The SARIMA model's eight coefficients were estimated to be significant, with P values less than 0.05. The monthly price of tea was predicted by the $SARIMA(3,0,3)(0,1,1)12$ for 36 months ahead using the test set and then the forecasts were evaluated using the RMSE and MAE. The test for heteroscedasticity was conducted using the ARCH-LM 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 coefficient $\alpha_1 = 0.1527$ is close to 1 indicating stable long term volatility, while $\beta_1 = 0.7952$ indicating decaying conditional variance. The long-run volatility persistence level obtained from the $GARCH(1,1)$model coefficients $\alpha_1 + \beta_1 = 0.9479$, 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

Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, *19*(6), 716–723.

Al-Jafari, M. K. (2012). An empirical investigation of the day-of-the-week effect on stock returns and volatility: Evidence from muscat securities market. *International Journal of Economics and Finance*, *4*(7), 141–149.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, *31*(3), 307–327.

Cheserek, B. C., Elbehri, A., & Bore, J. (2015). Analysis of links between climate variables and tea production in the recent past in kenya. *Donnish Journal of Research in Environmental Studies*, *2*(2), 005–017.

Chinomona, A. (2009). *Time series modelling with application to south african inflation data* (Doctoral dissertation).

Engle, R. (2001). Garch 101: The use of arch/garch models in applied econometrics. *Journal of economic perspectives*, *15*(4), 157–168.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the econometric society*, 987–1007.

Fwaga, S. O., Orwa, G., & Athiany, H. (2017). Modelling rates of inflation in kenya: An application of garch and egarch models. *Mathematical Theory and Modelling*, *7*(5), 75–83.

Gesimba, R., Langat, M., Liu, G., & Wolukau, J. (2005). The tea industry in kenya; the challenges and positive developments. *Journal of Applied Sciences*, *5*(2), 334–336.

Jarque, C. M., & Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics letters*, *6*(3), 255–259.

Kamunya, S. M., Wachira, F. N., Pathak, R. S., Muoki, R. C., & Sharma, R. K. (2012). Tea improvement in kenya. In *Global tea breeding* (pp. 177–226). Springer.

Karlsson, B. G. (2022). The imperial weight of tea: On the politics of plants, plantations and science. *Geoforum*, *130*, 105–114.

Kibunja, H. W., Kihoro, J. M., Orwa, G. O., & Yodah, W. O. (2014). Forecasting precipitation using sarima model: A case study of mt. kenya region.

KIPPRA. (2020). *Fluctuations in market earnings for tea in kenya: What could be the cause and remedy?* https://kippra.or.ke/fluctuations-in-market-earnings-for-tea-in-kenya-what-could-be-the-cause-and-remedy/

Ljung, G. M., & Box, G. E. (1978). On a measure of lack of fit in time series models. *Biometrika*, *65*(2), 297–303.

Maiyo, M. K. (2018). *Efficiency of the markov regime switching garch model in modelling volatility for tea prices* (Doctoral dissertation). Strathmore University.

Monroy, L., Mulinge, W., & Witwer, M. (2019). Analysis of incentives and disincentives for tea in kenya. *Gates Open Res*, *3*(586), 586.

Muthamia, A. K., & Muturi, W. (2015). Determinants of earnings from tea export in kenya. *Journal of World Economic Research*, *4*(1), 15–22.

Mutwiri, R. M. (2019). Forecasting of tomatoes wholesale prices of nairobi in kenya: Time series analysis using sarima model.

Mwaura, F., & Muku, O. (2008). *Tea farming enterprise contribution to smallholders' well being in kenya* (tech. rep.).

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the econometric society*, 347–370.

Qiao, Y., Halberg, N., Vaheesan, S., & Scott, S. (2016). Assessing the social and economic benefits of organic and fair trade tea production for small-scale farmers in asia: A comparative case study of china and sri lanka. *Renewable Agriculture and Food Systems*, *31*(3), 246–257.

Rachev, S. T., Kim, Y. S., Bianchi, M. L., & Fabozzi, F. J. (2011). *Financial models with lévy processes and volatility clustering*. John Wiley & Sons.

Shumway, R. H., Stoffer, D. S., & Stoffer, D. S. (2000). *Time series analysis and its applications* (Vol. 3). Springer.

Swainson, N. (1980). *The development of corporate capitalism in kenya, 1918-77*. Univ of California Press.

Tendai, M., & Chikobvu, D. (2017). Modelling international tourist arrivals and volatility to the victoria falls rainforest, zimbabwe: Application of the garch family of models. *African Journal of Hospitality, Tourism and Leisure*, *6*(4), 1–16.

Turner, C. M., Startz, R., & Nelson, C. R. (1989). A markov model of heteroskedasticity, risk, and learning in the stock market. *Journal of Financial Economics*, *25*(1), 3–22.

Winrose, C. (2018). Effect of price and exchange rate volatility on kenyaâ€™ s black tea export demand: A pooled mean group estimation. *Journal of Development and Agricultural Economics*, *10*(3), 71–78.

A. APPENDIX

A.1 Data

The monthly price of tea data used for analysis is attached on the link below;

https://drive.google.com/file/d/18ow9gR7ZVVrSElsENYIxFkecTwbDrtnh/view

A.2 Codes

```
df = pd.read_csv("monthly_Price_of_tea.csv", index_col=0, parse_dates=True)
df.head()
df.rename(columns ="monthly_Price_of_tea":"monthly_price_tea", inplace = True)
df.index.freq = "MS"
df["monthly_price_tea"].loc["2011/01/01": "2021/11/01"].plot(figsize = (16,8))
lag_plot(df["monthly_price_tea"])
plot_acf(df["monthly_price_tea"], title = title, lags = lags,ax=ax)
plot_pacf(df["monthly_price_tea"], title = title1, lags = lags,ax=ax)
adf_test(df["monthly_price_tea"])
from statsmodels.tsa.seasonal import seasonal_decompose
result = seasonal_decompose(df["monthly_price_tea"], model= "add")
result.plot()
sarima_model = auto_arima(df["monthly_price_tea"], start_p=1, start_q=1, max_p=3, max_q=3,
m=12, start_P=0, seasonal=True, d=None, D=1, trace=True, error_action='ignore', suppress_warnings=True,
stepwise=True)
print(sarima_model.summary())
model = SARIMAX(train["monthly_price_tea"],order = (3,0,3), seasonal_order = (0,1,1,12))
model_fit = model.fit()
```

```
start = len(train)
end = len(train)+len(test)-1
predictions = model_fit.predict(start = start, end=end, dynamic=True,typ=’levels’).rename(’SARIMA(3,0,3)(0,
Predictions’)
ax = test[’monthly_price_tea’].plot(legend=True,figsize=(12,6),title=title)
predictions.plot(legend=True)
ax.autoscale(axis=’x’,tight=True)
ax.set(xlabel=xlabel, ylabel=ylabel)
from sklearn.metrics import mean_absolute_error,mean_squared_error
mse_error = mean_squared_error(test[”monthly_price_tea”], predictions)
print(f’SARIMA(3,0,3)(0,1,1,12) MSE Error: mse_error’)
rmse_error = sqrt(mean_squared_error(test[”monthly_price_tea”], predictions))
model_sarima = SARIMAX(df[”monthly_price_tea”],order = (3,0,3), seasonal_order = (0,1,1,12))
model_sarima_fit = model_sarima.fit()
forecast = model_sarima_fit.predict(len(df), len(df)+11,typ=”levels”).rename(’SARIMA(0,1,3)(1,0,1,12)
Forecast’)
forecast10 = model_sarima_fit.predict(len(df), len(df)+119,typ=”levels”).rename(’SARIMA(0,1,3)(1,0,1,12)
Forecast_10years’)
ax = df[’monthly_price_tea’].plot(legend=True,figsize=(12,6),title=title)
df[”returns”] = (df[”monthly_price_tea”].pct_change())*100 df[’Log_Return’] = np.log(df[’monthly_price_tea’]
df = df.dropna()
train_garch =garch_df.iloc[:333]
test_garch = garch_df.iloc[333:]
garch_model = arch_model(garch_df[”returns”], p=1, q=1,mean = ”constant”, vol = ”GARCH”,dist
= ”normal”)
garch_model_train = garch_model.fit(last_obs = test_garch.index[0],update_freq = 4)
garch_predictions = test_garch.copy()
garch_predictions[”Predictions”] = garch_model_train.forecast().residual_variance.loc[test_garch.index]
garch_predictions[”Predictions/10”]=garch_predictions[”Predictions”]/10
garch_predictions[”returns”].abs().plot(label=”actual price”,figsize=(14,10),title = ”Volatility fore-
```

```
casts.”)
garch_predictions[”Predictions/10”].plot(label = ”predictions”)
plt.legend()
plt.figure(figsize=(14,10))
ax=garch_model_train.plot()
mse_error_garch = mean_squared_error(test_garch[”returns”], garch_predictions[”Predictions/10”])
print(f’GARCH MSE Error: mse_error_garch’)
rmse_error_garch = sqrt(mean_squared_error(test_garch[”returns”], garch_predictions[”Predictions/10”]))
```

A.3 Plagarism Report

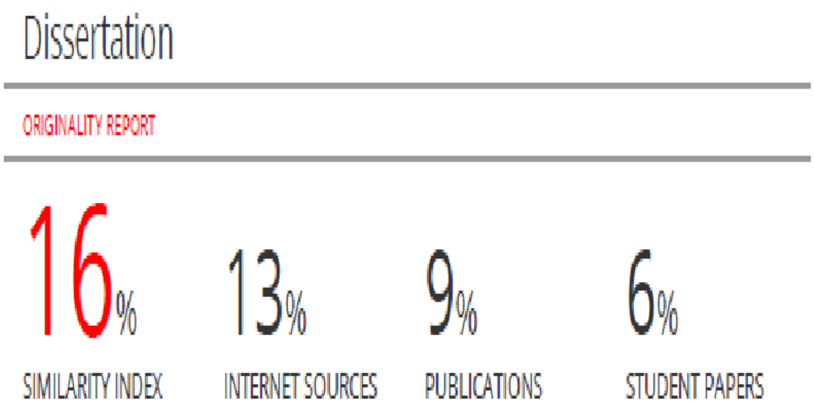

Figure A.1: Plagarism Report conducted on Turnitin