FORECASTING TEA PRICES AND THEIR VOLATILITY IN KENYA

Robin Ochieng Otieno
SCT-325-002/2019

Department of Mathematics
Multimedia University of Kenya
MSC(Statistics)

October 2022

Supervisors:
Dr. Wycliffe Cheruiyot
Dr. Antony Karanja

Introduction: Background

Tea in Kenya

- Tea has become an essential contributor to the economies of producing countries such as Kenya, China, India, and Sri Lanka. Tea originated in China, and it has now advanced to a multi-billion-dollar global industry.
- In 2019, the tea sector accounted for 25 percent of the foreign exchange earnings and 1.5 percent of the Gross Domestic Product in Kenya KIPPRA (2020).

Introduction: Background

Tea in kenya

- Globally, Kenya falls in third place after China and India in tea production with a market share of 9% KIPPRA (2020).
- Kenya is the largest exporter of tea in the world accounting to 20 percent of the world tea exports Monroy et al. (2019).
- The Kenya tea sector contributes 26 percent of Kenya's export earnings as Kenya exports tea to Pakistan, Egypt, the United Kingdom, and others Muthamia and Muturi (2015).

Introduction: Basic Concepts

Volatility Definition and Measurement

- We define volatility as the irregular rising and falling observed in some events over time.
- In forecasting literature it is referred to as the conditional variance of the underlying return.
- The sample standard deviation is used as its measure;

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \mu)^2}{N}} \tag{1}$$

Introduction: Basic Concepts continued..

Volatility Clustering

- Volatility clustering behavior refers to the tendency for large changes in price to be followed by large changes and small changes to be followed by small changes Rachev et al. (2011).
- It is determined by enumerating the Autocorrelation function.
- Suppose X_t is a stationary time series, the ACF for the series is defined as:

$$\rho_k = \frac{Corr(X_t, X_{t-k})}{\sqrt{Var(X_t)Var(X_{t-k})}} = \frac{\gamma(k)}{\gamma(0)} \tag{2}$$

Introduction: Basic Concepts continued..

Stationarity in Time Series

- Stationarity is a critical assumption in time series models.
- Stationarity implies homogeneity in the sense that the series behaves in a similar way regardless of time.
- This means that its statistical properties ie the mean, variance and the autocovariance of the series do not change over time that is independent of time.

Introduction: Basic Concepts continued..

Differencing in Time Series

- The most common way to make non-stationary time series data stationary is by differencing.
- Differencing is very useful in eliminating the trend and seasonal effects. For seasonal data, first order differencing is generally adequate to achieve stationarity.
- Let $X_t = X_1, X_2, ...X_n$ be a non staionary time series. Then stationary time series is obtained by:

$$\nabla X_t = X_t - X_{t-1} \tag{3}$$

- This is called a first order difference.

Introduction: Basic Concepts continued..

Returns

- Directly analyzing financial prices is difficult since they are correlated and their variances typically rise with time; hence, we need returns 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

$$r_t = \log \frac{Y_t}{Y_{t-1}} \tag{4}$$

Literature Review

Literature Review

- According to Gesimba et al. (2005), the tea industry is facing a number of issues. The most challenging problem, he argues, is the unpredictability of tea export prices, which may be overcome by regulating demand and supply in the global market.
- The ultimate purpose of volatility analysis is to explain why there is volatility. The GARCH models are intended to study volatility in order to make financial choices such as risk evaluations.

Literature Review continued..

Literature Review

- Chinomona (2009) modeled South African inflation data using the ARIMA and GARCH models; she picked the $ARIMA(1, 1, 0) * (0, 1, 1)_{12}$ model from the ARIMA models and GARCH (1, 1) from the ARCH-GARCH models.
- Because of its capacity to capture fluctuations in the series, GARCH (1, 1) was chosen as the best model over $SARIMA(1, 1, 0) * (0, 1, 1)_{12}$.
- Another research by Fwaga et al. (2017) employed both the GARCH and the EGARCH models to model Kenyan inflation. The model with the lowest AIC and BIC values was chosen as the best model.

Statement of the Problem

Problem Statement

- The ultimate goal of price volatility analysis is to explain why prices fluctuate. Price volatility can be caused by shocks to demand and supply, which can be caused by changes in climatic conditions, such as extreme cold or hot weather.
- Because they take volatility clustering into account, GARCH models are appropriate for forecasting volatility. Their ability to reliably estimate time-varying conditional variances also enables them to forecast return variances and covariances.

Objectives of the Study

The Main Objective

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

The 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.

Significance of the Study

Significance of the study

- The impetus for doing this research stems from the issues confronting the tea industry, notably the insecurity of growers' revenues.
- The Tea Board of Kenya (TBK) will utilize the findings of this study to make financial decisions regarding price risk analysis and to plan for procurement needs because they will be able to estimate prices and defend against the negative repercussions 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.

Methodology

Introduction

- In this work, the SARIMA and GARCH models are used to model and forecast tea prices and their volatility in Kenya using a time-series method.
- The method entails identifying the price of tea pattern using a time series plot, unit root testing for stationarity, determining the model order, fitting and estimating the parameters of the appropriate SARIMA and GARCH models, and lastly forecasting future prices.
- The data used in this study comprise the monthly price of tea in Kenya over the period January 1, 1990 to November 1, 2021.

Methodology continued..

Test for stationarity and autocorrelation

- The Augmented Dickey-Fuller (ADF) test will be used to determine stationarity; after stationarity is achieved, the autocorrelation and partial autocorrelation plots will be inspected to determine their correct structures.
- It will be conducted using the Ljung-Box Q-statistic test by Ljung and Box (1978), which is defined as:

$$Q = n(n+2)\sum_{j=1}^{m} \frac{r_j^2}{n-j} \tag{5}$$

Methodology continued..

Seasonal Auto-regressive 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 parameters, (P, D, Q) are the seasonal parameters, and s is the seasonal period.
- The SARIMA model can be expressed as;

$$\Phi_P(B^s)\phi(B)\nabla_s^D\nabla^d X_t = \delta + \Theta_Q(B^s)\theta(B)w_t \tag{6}$$

- The SARIMA coefficients are represented by $\Phi_p(B^s)$ and $\Theta_Q(B^s)$ of orders P and Q, while the ordinary and seasonal difference coefficients are represented by $\nabla^d = (1 - B)$ and $\nabla_s^D = (1-B^s)$ Shumway et al. (2000).

Methodology continued..

Generalized Auto-regressive Conditional Heteroscedasticity(GARCH) model

- To represent the variance at time t, a GARCH model employs past squared residuals and past variances.
- The GARCH (p, q) model combines the ARCH(q) model and incorporates p lags of conditional variance in the model, where p is known as the GARCH order.
- The combined model is called the GARCH(p,q) model given as:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i r_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \tag{7}$$

Methodology continued..

GARCH Model

- In this study we will use GARCH (1, 1) where the parameters p = 1 and q = 1 are given by:

$$\sigma_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \tag{8}$$

- 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

Methodology continued..

Model Selection

- The model selection will be based on the Akaike Information Criterion (AIC) proposed by Akaike (1974).
- The Akaike Information Criterion approximates the quality of each model given a set of data models, and the model with the least AIC value is considered the best model.
- The formula for the Akaike Information Criterion (AIC) is:

$$AIC = 2K - 2\ln(L) \tag{9}$$

- Where K is the number of estimated model parameters and L denotes the maximum likelihood function for the model.

Methodology continued..

Model Diagnostics

- Following model selection and parameter estimation, the model residuals will be used to perform model diagnostics, which will test the independence and normality of the fitted model's residuals.
- The Jaque Bera test will be used to determine normality, while the Ljung Box test will be used to determine residual independence.
- The histogram density and the normal q-q plots are also used to confirm if the standard residuals conform to a normal distribution.

Methodology continued..

Forecasting Performance Evaluation

- The Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) are used to evaluate the models' predicting ability.
- The formula for the MAE and RMSE are given as:

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|Y_t - F_t| \tag{10}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(Y_t - F_t)^2} \tag{11}$$

Results and Discussion.

Data

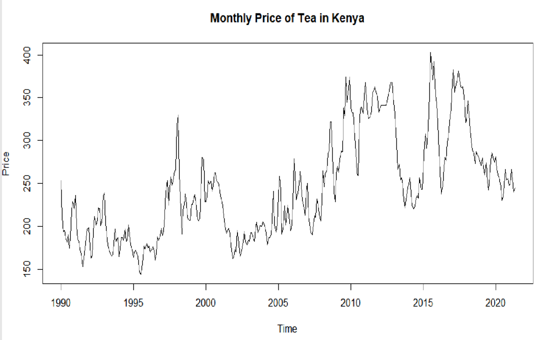
- To forecast the prices of tea and their volatility in Kenya, monthly prices per kilogram, in Kenyan shillings, data for the period January 1, 1990, to November 1, 2021, obtained from FRED.
- The average price of tea over the years is Ksh. 245.80 with a standard deviation of Ksh. 60.70. The series also shows positive skewness of 0.591, meaning that the distribution has a long right tail.

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

Results and Discussion.

A Plot of the data

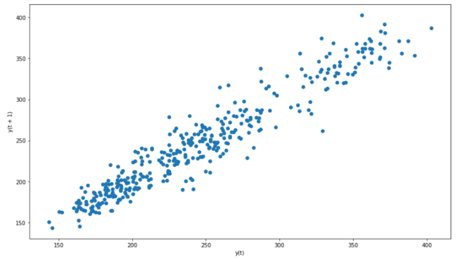
- The plot clearly shows that the data is non-stationary, with the series trend growing with time. Seasonal and cyclic patterns may also be seen throughout time.



Results and Discussion.

Test for Autocorrelation and Checking for Outliers

- Prior to the stationarity test, a lag plot was created to check for autocorrelation and the presence of outliers. The plot shows that there are no outliers in the data and that there is positive serial auto-correlation. As a result, an auto-regressive model is preferable.



Results and Discussion.

Test for Stationarity

- The stationarity test was carried out in two ways: inferentially and visually. It was done inferentially with the ADF test for stationarity and visually with the ACF plot. The ADF test statistic is (-1.8424), and the p-value is 0.3596, which is more than 0.05, confirming the presence of a unit root and suggesting that the data is non-stationary.

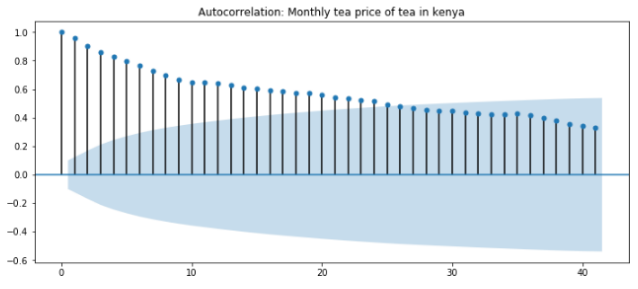
Table: Results of the ADF test

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

Test for Stationarity.

AutoCorrelation Function Plot

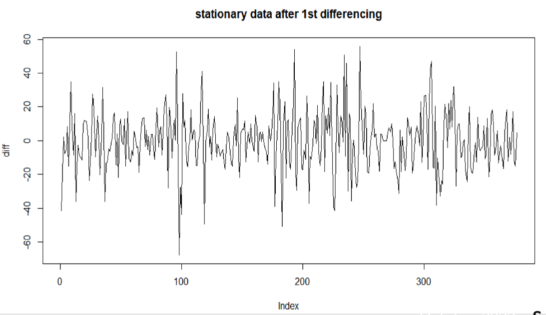
- The ACF plot shows a gradually decaying ACF, implying that previous values substantially impact future values of the monthly tea price in the series. This is an example of an ACF plot for non-stationary data.



Attaining Stationarity.

Differencing the Series

- Because we can't feed non-stationary data through our model, we make the data stationary by taking the first difference in tea prices. The plot of the stationary data is shown below, and it clearly shows a consistent mean and variance over time.



The SARIMA Model

SARIMA Model Parameter Identification

- The AR and MA parameters are identified using the Extended Autocorrelation Function (EACF) since the data is stationary. Based on the EACF results, an ARMA(3,3) model was selected as the best model to be fitted among other preliminary models.

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

The SARIMA Model

Model Selection

- In order to determine the optimal SARIMA model, several SARIMA models of different parameter orders are fitted on the monthly price of tea in Kenya. The optimal *SARIMA* model determined 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, which is given below.

Table: The AIC values of the fitted models

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

The SARIMA Model

Coefficient Estimation

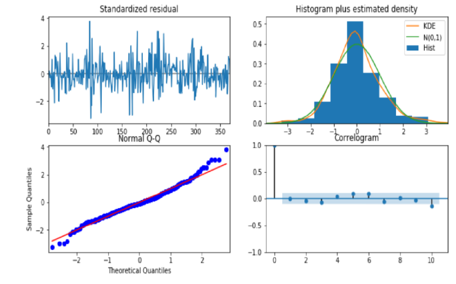
- The $SARIMA(3,0,3)(0,1,1)_{12}$ model was used to estimate the eight coefficients of the SARIMA model. It can be seen that the model coefficients are all significant since they are all less than 0.05, indicating that the $SARIMA(3,0,3)(0,1,1)_{12}$ model is significant.

	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

The SARIMA Model

Residual Diagnostics

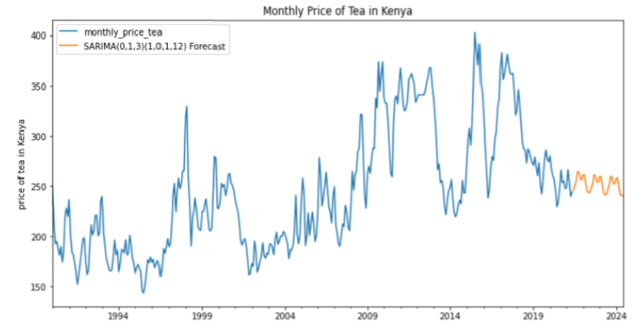
- The histogram shows that the standard residuals have a normal distribution. The normal probability plot showed that the data had a normal distribution, and the correlogram showed no substantial lag, indicating that the model was adequate.



The SARIMA Model

Forecasting Monthly tea prices to 2024

- The $SARIMA(3,0,3)(0,1,1)_{12}$ model was used to forecast future monthly prices of tea up to June 2024.



The SARIMA Model

Tea Price forecasts from December 2022 to June 2024

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

The SARIMA Model

Evaluating the SARIMA Model

- The SARIMA model's prediction capacity was evaluated using the Root Mean Squared Error and Mean Absolute Error indicated below. We may thus infer that the $SARIMA(3,0,3)(0,1,1)_{12}$ model performed well in terms of monthly tea prices forecasting.

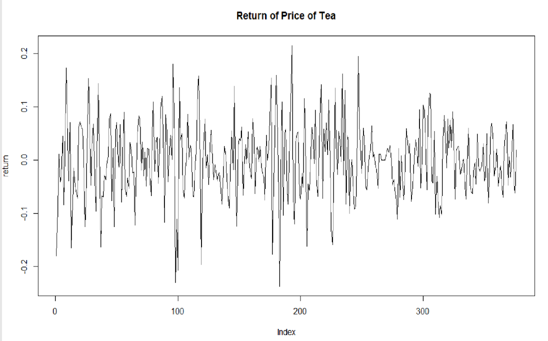
Table: Forecasting performance of the SARIMA Model

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

Volatility Modeling

Plot of the Returns

- It can be observed from the plot that the returns are stationary; that is, they have a constant mean and variance over time



Volatility Modeling

Descriptive Statistics of the Returns

- To further comprehend our return series, we computed descriptive statistics for tea's monthly returns. The lowest return was -21.038, the highest was 23.956, the mean was 0.346, and the standard deviation was 6.974. With a skewness of 0.224, the results are considerably skewed.

Table: 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

Volatility Modeling

Test for Heteroscedasticity

- The Heteroscedasticity test is used to determine the presence of arch effects in the squared residuals. It is calculated using the Lagrange Multiplier. We reject the null hypothesis since the ARCH-LM test statistic is 3.875 with a P-Value of 0.0490, and we conclude that there are significant Arch effects from January 1, 1990 to November 1, 2021. As a consequence, we can determine which GARCH model is optimal.

Returns	**Chi-square**	**P-Value**
Tea	3.875	0.0490

Volatility Modeling - The GARCH Model

Model Selection

- Different GARCH(p,q) models are simulated on the price of tea data, and the results show that GARCH(1,1) is the best model for fitting monthly tea price data, with an AIC of 2542.47 and a BIC of 2558.24. As a result, the GARCH(1,1) model coefficients are estimated for the monthly 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

Volatility Modeling - The GARCH Model

Coefficient Estimation

- The GARCH(1,1) model was used for coefficient estimation, and the p-values are all smaller than $\alpha = 0.05$. As a result, we conclude that all of the GARCH(1,1) model's coefficient are relevant and the GARCH(1,1) model is the optimal model.

	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

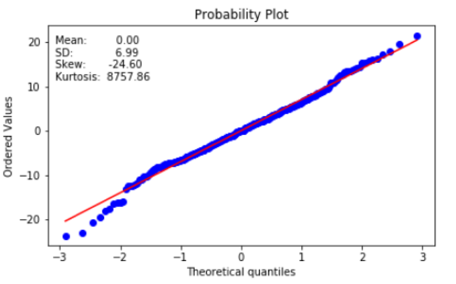
The volatility model GARCH(1,1) for the Monthly price of tea can thus be expressed as:

$$\sigma_t^2 = 2.9566 + 0.1527 r_{t-1}^2 + 0.7952 \sigma_{t-1}^2 \tag{12}$$

Volatility Modeling - The GARCH Model.

Residual Diagnostics - Probability Plot

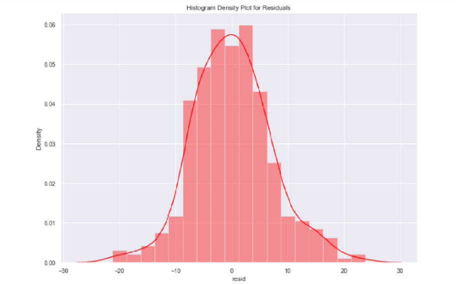
- The GARCH(1,1) model diagnostics were conducted on the standardized residuals, which are supposed to follow a normal distribution. The normality of the residuals is checked using the probability plot and the histogram density plot.



Volatility Modeling - The GARCH Model.

Residual Diagnostics - Histogram Density Plot

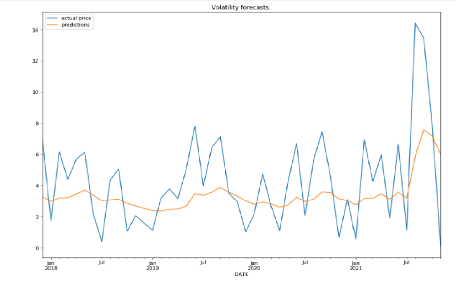
- The histogram density plot further demonstrates that the normalized residuals follow a normal distribution since the histogram is symmetrical in shape. The GARCH(1,1) model residuals are therefore proved to be significant.



Volatility Modeling - The GARCH Model.

Forecasting Volatility

- The train set is used to fit the GARCH(1) model and forecast for the next 36 months. The GARCH(1,1) model forecast data pattern and the actual data pattern are practically identical. As a result, we continued to evaluate the forecasts for the GARCH(1,1) model.



Volatility Modeling - The GARCH Model.

Evaluation of 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). The RMSE (5.903) and the MAE (5.089) values are significant. Thus we conclude that GARCH(1,1) model performed well in forecasting volatility.

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

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

Conclusion and Recommendation.

- The $SARIMA(3,0,3)(0,1,1)_{12}$ model was identified as the best model for predicting tea prices in Kenya using the EACF plot. The SARIMA model's eight parameters were all found to be significant, with p-values less than 0.05. The $GARCH(1,1)$ model was identified as the best GARCH model for forecasting tea price volatility.
- The long-run volatility persistence level determined using the $GARCH(1,1)$ model parameters $\alpha_1 + \beta_1 = 0.9479$, indicating lengthy durations of future volatility shock persistence.
- For further research, I would advocate using the SARIMA model residuals to estimate the parameters of the GARCH(1,1) model, resulting in a SARIMA-GARCH model that may be used to predict volatility.

References.

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

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

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.

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