APPLICATION OF AUTO-REGRESSION INTEGRATED MOVING AVAERAGE MODEL TO FORECAST TEA PRICES AT MOMBASA TEA AUCTION CENTER IN KENYA

BY

ODDAH ABALA OKUMU

A RESEARCH PROJECT SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE IN APPLIED STATISTICS

SCHOOL OF MATHEMATICS, STATISTICS AND ACTUARIAL SCIENCE

MASENO UNIVERSITY

© 2019

Declaration

I, Oddah Abala Okumu, duly declare that this project is my own work and has not been presented for a degree award in any other institution.

ODDAH ABALA OKUMU

Admission Number PG/MAT/00094/014

Signature:.....Date:.....

This project has been submitted for examination with my approval as the university supervisor.

PROF. FREDRICK ONYANGO

Department of Statistics and Actuarial Science

Maseno University

Signature:.....Date:.....

Acknowledgement

I acknowledge and offer my regards to all those who contributed to the success of my research project and subsequent preparation of this report.

I would like to thank my supervisor Prof. Fredrick Onyango for his guidance and valuable contribution in realization of the success of this work.

I sincerely would like to extend my gratitude to my lecturers at the university of Maseno, Dr.Otumba and Dr.Kaleli for their rich ideas through this research. I am also grateful to the University of Maseno for giving me an opportunity to further my studies thus an opportunity to write this project.

My family provided a solid rock of support for which am always grateful. I would also like to thank Evans Ongala for his continued input from the beginning to the end of this project.

Lastly, but most importantly I thank the Almighty God for His Grace throughout my study period.

Dedication

I dedicate this research project to my caring husband George Warera, for without his support my educational endeavours and career development would be futile. I again respectfully dedicate it to my lovely children, Tess, Gloria and Teddy.

Abstract

Tea production is the mainstay of many people in Kenya. The contribution of tea sector to the Kenyan economy is enormous as it accounts for about 4% of Gross Domestic Product (GDP). Favourable product prices act as a trigger for any commercial activity over time and also suggest sustainability of the activity. Tea is a very significant crop in Kenya and it is one of the major agricultural exports of the country. However, the industry has experienced many challenges such as stiff competition from other tea producing countries like Sri-Lanka, unfavourable climatic changes, land competition from other crops like sugar cane just to mention a few. These challenges have caused seasonal variation in tea pricing at the Mombasa tea auction which is the second largest black tea auction center in the world after Colombo in Sri-Lanka. There exist plethora of studies to justify that a careful and precise selection of Auto-regressive Integrated Moving Average (ARIMA) model can be fitted to time series data to forecast with better accuracy the future prices of tea auction. We have used time series ARIMA model to forecast tea pricing at Mombasa tea auction in Kenya. The general objective of this project is to study the trend and seasonality, and to forecast future tea prices. Monthly pricing data between 1995-2018 from International Tea Committee; African tea brokers limited was used for model validation. To identify the best ARIMA model R statistical software was used to analyse the tea pricing data. ARIMA (1, 1, 2) with Akaike Information criterion (AIC) and Bayesian information criterion in table 4.1. The model was then used to predict tea prices from 2019-2023 as shown in table 4.13.

TABLE OF CONTENTS

Title page i

Declaration ii

Acknowledgement iii

Dedication iv

Abstract v

Table of Contents vi

Abbreviations and Acronyms ix

List of Tables x

List of Figures xi

CHAPTER ONE: INTRODUCTION 1

1.1 Background information 1

1.1.1 Tea farming in Kenya 1

1.1.2 Tea marketing 2

1.1.3 The Mombasa Tea auction 2

1.2 Statement of the problem 2

1.3 Objectives of the study 3

1.4 Significance of the study 3

1.5 Basic concepts 3

1.5.1 Pricing 3

1.5.2 Pricing strategy 4

1.5.3 Auction 4

1.5.4 Time series data 4

1.5.5 Time series graph 4

1.5.6 Pure random process 4

1.5.7 Time series technique 5
1.5.8 Sample Autocorrelation Function (ACF) 5
1.5.9 Partial Autocorrelation Function(PACF) 5
1.5.10 Augmented Dickey Fuller (ADF) test 5

CHAPTER TWO: LITERATURE REVIEW 7

CHAPTER THREE: RESEARCH METHODOLOGY 13

3.1 Time series 13
3.1.1 Time series Analysis 13
3.1.2 Components of time series 13
3.1.3 The trend(T) 13
3.1.4 Seasonal Variation(S) 14
3.1.5 Cyclic variation 14
3.1.6 Random variation 14
3.2 Mean and auto co-variance functions 16
3.3 Stationary time series 16
3.4 Differencing time series 16
3.5 Auto-covariance function 16
3.6 Auto correlation 16
3.7 Auto correlation function(ACF) 17
3.8 Ljung-Box Q (LBQ)Statistic 17
3.9 Back shift operator 17
3.10 The Auto regressive Inter grated Moving Average(ARIMA) model 18
3.10.1 Auto regressive(AR)process 18
3.10.2 Moving Average process(MA) 18
3.10.3 Auto-regressive Moving Average series(ARMA) 18
3.10.4 ARIMA model 18
3.11 Model identification 19

3.11.1 Model estimation and selection 20
3.12 Model checking 21

CHAPTER FOUR: RESULTS AND DISCUSSIONS 23

4.1 Time series data for tea auction prices 23
4.2 Stages of building the ARIMA model 23
4.3 Model identification 24
4.4 Parameter estimation and selection 26
4.5 Model checking 27
4.6 Forecasting with ARIMA model 31

CHAPTER FIVE: CONCLUSIONS AND RECOMMENDATIONS 35

5.1 Conclusions 35
5.2 Recommendations 35
REFERENCES 40
APPENDIXES 41

ABBREVIATIONS AND ACRONYMS

AIC Akaike's Information Criterion

ARIMA Autoregressive integrated moving average

AR Autoregressive

AICc Corrected Akaike information criterion

ARMA Autoregressive moving average

ACF Autocorrelation function

BIC Bayesian Akaike information criterion

MA Moving average

PACF Partial autocorrelation function

ADF Augmented Dickey Fuller

KTDA Kenya Tea Development Authority

EATTA East African Tea Trade Association

MAE Mean Absolute Error

RMSE Root Mean Squared Error

MAPE Mean Absolute Percentage Error

LIST OF TABLES

4.1 Parameter estimates for ARIMA(1, 1, 2) 26

4.2 Model fit statistics 26

4.3 Ljung Box test for ARIMA(1,1,2)model 28

5.1 Data for tea auction prices at Mombasa tea auction center 43

LIST OF FIGURES

3.1 Box-plot for the Tea auction prices 15

3.2 Components time series plot of tea auction prices from 1995-2018 15

4.1 Non-stationary times series data for tea auction prices 24

4.2 First differenced series 25

4.3 ACF for the differenced series 27

4.4 PACF for the differenced series 27

4.5 Residuals for the ARIMA (1,1,2) 28

4.6 ACF plot of the ARIMA model residuals 28

4.7 PACF for the model ARIMA(1,1,2)residuals 29

4.8 Histogram of the model residuals 29

4.9 The normal Q-Q plot for the residuals of the ARIMA (1,1,2) model 30

4.10 Accuracy of ARIMA (1,1,2) model 30

4.11 Coefficients of ARIMA (1,1,2) 32

4.12 Forecast from ARIMA (1,1,2) 32

4.13 Upper and lower limits of the predicted Tea prices in 2019 33

4.14 Upper and lower limits of the predicted Tea prices in 2020 33

4.15 Upper and lower limits of the predicted Tea prices in 2021 33

4.16 Upper and lower limits of the predicted Tea prices in 2022 34

4.17 Upper and lower limits of the predicted Tea prices in 2023 34

4.18 The predicted values for the Tea prices from 2019 to 2023 34

CHAPTER ONE

INTRODUCTION

1.1 Background information

China, India, Kenya and Sri Lanka are the four biggest tea producers in the world [19]. China production share of tea is 28% but its market in the global share is 18%, since domestically, China consumes about 70% of tea it produces. Over the past decade China has shown an increase in tea consumption. India is the second largest producer of tea contributing to 25% share of production and market share export of about 12% [25].

1.1.1 Tea farming in Kenya

The history of tea growing in Kenya dates back in 1903 when a European settler Mr. G. W.Cain introduced the first tea in Limuru area of central Kenya. The early settlers and the colonial government restricted tea and coffee growing to large scale farmers and multinationals, ostensibly to maintain quality. However, the main reason was to lock out locals from the then very lucrative cash crop farming. After independence, land reform bills were passed . This enabled the locals to grow tea. The crop has since spread to various parts of the country and is currently an economic mainstay to many small scale farmers. Small scale farmers account for 60 % of all tea produced in Kenya. The tea industry operates under the Tea Act (Cap 343) and Agricultural Act (Cap 318) of the laws of Kenya. Small scale tea in Kenya is placed under the Kenya Tea Development Authority (KTDA). Kenya falls in the 3rd place with a 9% share of global production. It is the largest exporter of black tea in the world. The export share of Kenyan tea is 25% by volume and Kenyan export earnings is about 26% of the world's total by value. The contribution to the Kenyan agricultural sector's GDP is a bout 11% and over 60% of these production is contributed by small holders. Therefore tea is one of the largest employers in the private sector in the country [30].

1.1.2 Tea marketing

The emergence of global trade allows for liberalization towards demand driven economy. Kenyan tea is mainly destined to Pakistan, United Kingdom, Afghanistan and Sudan with individual organizations and Tea Board of Kenya being involved in the marketing of the product. Marketing of tea has become a fierce battle ground where both domestic and foreign competitors fight for market share. Kenyan tea has won an international acclaim for its aroma and taste. It is so famous that it is the beverage of choice served in many restaurants and cafeteria in many countries (Kenya information guide 2014). In its purest form, Kenyan tea is very superior to other tea produced in the world hence it is blended with other tea to produce another popular blend sold across many countries [24].

1.1.3 The Mombasa Tea auction

In November 1956, The export Auction System was initiated under the management of East African Tea Traders Association (EATTA) in Nairobi and most of the tea produced was consigned to the United Kingdom. In 1969 the auction center was moved to the port of Mombasa which was the nerve center of warehousing, handling and shipping. The Mombasa Tea Auction consist of main grades of auction and secondary grade auction and its held weekly on Mondays. The Mombasa Tea Auction has grown to be the second largest tea auction in the world after Colombo Tea Action in Sri Lanka. Tea brokerage is done at Mombasa tea auction by tea brokers who are appointed by tea producers and buyers. The brokers guarantee the sale of tea at the best possible price. Mombasa tea auction is a very important market for the commodity. The major determinants of the success of tea auction are the pricing and a well executed pricing strategy [30].

1.2 Statement of the problem

Tea industry plays an important role to Kenyan economy. However, tea industry in Kenya is faced with very many challenges in terms of production and export. These challenges include: stiff competition from other producing nations like Sri-Lanka, unfavourable climatic conditions such as long droughts and periods of frost, pest and diseases as well as pricing. Pricing of Kenyan black tea is mainly done at Mombasa tea auction. The prices

are quoted on weekly basis. The price quotes are in terms of per kilograms. The quotes reflect the demand and supply of tea. Due to the aforementioned challenges, disparity in prices and possible absence of efficient market exist. It is therefore significant to study characteristic of tea auction prices. In our study we have predicted tea prices at the Mombasa tea auction center using ARIMA model.

1.3 Objectives of the study

The general objective of this project is to study the trend and seasonality, and to forecast tea prices at the Mombasa tea auction center. The specific objectives include:

- To apply Auto-regressive Integrated Moving Average model to study the trend and seasonality of tea prices at the Mombasa tea auction centre in Kenya.
- To forecast the future tea prices at the Mombasa tea auction centre in Kenya.

1.4 Significance of the study

Pricing is very significant to marketers because it represents market's assessment of the value customers see in the product or service and are willing to pay. Pricing strategies must continually be reviewed and must consider that market is a dynamic entity operating within a very competitive environment[31]. This study is very significant in the sense that it can be used to sensitize the major players in the tea industry on the current market demands by using ARIMA model to study trend and seasonality of tea auction prices at Mombasa tea auction. This study can be used to advise farmers in East Africa to maintain the tea plants as well as plant more seedlings since they are likely to fetch better prices between 2020 and 2023.

1.5 Basic concepts

1.5.1 Pricing

Pricing is the process where by a business sets its price at which it will sell its products and services. This may be part of a firm's marketing plan. The pricing objectives takes

into account; the firm's financial goals, the fit with the contemporary market realities, market positioning and the market mix [44].

1.5.2 Pricing strategy

Pricing strategy takes into account the segments of ability to pay and the prevailing market conditions. The marketers develop the marketing strategies that is consistent with the organization's mission and values. The strategy is designed to provide broad guidance for price setters and ensures that the pricing strategy is consistent with the elements of pricing plan [44].

1.5.3 Auction

An auction is the process of buying and selling goods and services by offering them up for bids, taking bids and then selling them to the highest bidder. The open ascending is arguably the most common form of auction in use today. The participants bid openly against each other with each subsequent bid required to be higher than the previous one [36].

1.5.4 Time series data

Time series data is a set of observations on the values that a variable takes at different times i.e $X_1, X_2.....X_t\{X_t : t \in Z\}$[13].

1.5.5 Time series graph

Time series plot is a graph which displays information on the y-axis against equally spaced time intervals on the x-axis. The time series plot specifically consist of time scale on the x-axis, data scale on the y-axis and lines displaying each time series. The plots are used to detect trends and seasonality as well as comparing trends across all the groups.

1.5.6 Pure random process

Time series is said to be pure random walk process or white noise if it consists of random sequence e_t which are mutually independent and identically distributed with mean 0 and

variance σ^2 i.e X_t is a pure random process if $X_t = e_t \sim N(0\ \sigma^2)$ [11,12,13].

1.5.7 Time series technique

Time series analysis is a statistical technique that deals with time series data or trend analysis. It means that data is in a series of particular time periods or intervals.

1.5.8 Sample Autocorrelation Function (ACF)

The sample autocorrelation function (ACF) for a series gives correlations between the series X_t and lagged values of the series for lags of $1, 2, \ldots$.. It is defined as

$$\rho_k = \frac{\gamma_k}{\gamma_0}$$

Where ρ_k is the ACF at lag k, γ_k =Covariance(X_t, X_{t-k}) and γ_0 =Covariance(X_t, X_t) =Var(X_t). The estimate of the autocovariance is given by

$$c_k = \hat{\gamma_k} = \frac{1}{T}\sum_{t=1}^{T-k}(X_{t+k} - \bar{X})(X_t - \bar{X})$$

Where T is the total number of observations and $\bar{X}$ is the mean of the observations. The sample ACF is given by

$$r_k = \hat{\rho_k} = \frac{\hat{\gamma_k}}{\hat{\gamma_0}} = \frac{c_k}{c_0}$$

1.5.9 Partial Autocorrelation Function(PACF)

The partial autocorrelation between X_t and X_{t-k} is defined as the conditional correlation between X_t and X_{t-k}, conditional on $X_{t-k+1}, \ldots, X_{t-1}$ the set of observations that come between the time points t and $t-k$. For example a second order PACF is given by

$$\frac{Cov(X_t, X_{t-2}|X_{t-1})}{\sqrt{Var(X_t|X_{t-1})Var(X_{t-2}|X_{t-1})}}$$

1.5.10 Augmented Dickey Fuller (ADF) test

This is a formal test of stationarity of a time series. The hypothesis tested is

H_0 : There is a unit root

H_1 : The time series is stationary

Consider the AR(1) model

$$X_t = \phi X_{t-1} + \omega_t$$

where X_t is the value of the time series at time t, ϕ is the AR(1) parameter and ω_t is the residual at time t. If $\phi = 1$, then the AR(1) process is said to have a unit root.
Suppose we take a first difference, we have

$$\begin{aligned} \Delta X_t = X_t - X_{t-1} &= \phi X_{t-1} - X_{t-1} + \omega_t \\ &= (\phi - 1) X_{t-1} + \omega_t \\ &= \gamma X_{t-1} + \omega_t \end{aligned}$$

In this case testing for a unit root is equivalent to testing $\gamma = \phi - 1 = 0$. The ADF test allows for higher order AR processes.

A small p-value indicates a stationary series.

CHAPTER TWO

LITERATURE REVIEW

Mayler (1999) drew a frame work for ARIMA time series models for forecasting irish inflation. In their research , they emphasized heavily on optimizing forecast performance while forecasting more on minimizing out of sample forecast errors rather than minimizing in sample goodness of fit [31].

Rathnayaka,R.M (2017) applied Box-Jenkins ARIMA modelling approach to forecast stock prices under the volatility in the Colombo Stock Exchange(CSE), Sri-Lanka. The results suggested that ARIMA model is more suitable model for forecasting ASPI index under volatility [42].

Stargiou (1998) in his study used ARIMA model technique on a 17 years time series from 1964 to 1980 and 204 observations of pilchard Greek waters. They found a mean error of 14% which showed that ARIMA model was capable of forecasting the complex dynamics of pilchard fishery [45].

Raymond Y. C (1997) tested empirically the pattern of the real estate prices by employing ARIMA analysis. Results strongly showed that there exist cyclical trends in the office and industrial property prices in Hong Kong. The forecasting method can provide an indication of short term market direction a sense of whether or not the movement will be small or large and advance a warning well ahead of any turning points supplementary to investment strategy. The investor may wish to incorporate forecasts from an ARIMA model into his investment strategy for timing purposes [38].

Banerjee D. (2014) used ARIMA model to predict the future stock index of the Indian economy. They used monthly closing stock indices of Sensex for six years (2007-2012); they tried to develop an appropriate model which would help to forecast the future unobserved values of the Indian stock market indices. The study offered an application of ARIMA model which they used to predict the future stock indices which have a strong influence on the performance of the Indian economy. To validate establish the model they applied the validation technique with the observed data of Sensex of 2013 [7].

Burnham and Anderson (2004) shows that the Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC) time series analysis are based on the analysis of historical data. These analyses support the assumption that past patterns in data can be used to forecast future data points [14].

Liu(1991)used Box and Jenkins techniques to study the price dynamics in US crude oil, gasoline and stocks of gasoline with transferring function models [28].

Hong L. et al (2016), Studied India's Tea price Analysis based on ARMA model. ARMA model was used to analyse the sample of India's weekly tea auction price in the year 2013-2014. ARMA (1,1) was established then it was further used to predict the tea prices of the first two weeks in 2015 [29].

Saeed et al. (2003) suggested ARIMA (2, 2, 1) model to forecast wheat production in Pakistan. They used wheat production data series from 1947-1999. However, they did not mention the source of data [46].

Rahman (2010) applied ARIMA model to forecast Boro rice production in Bangladesh from 2008-09 to 2012-13. It appeared from the study that the local, modern and total Boro time series are first order homogeneous stationary. It was found from the study that the ARIMA (0,1,0), ARIMA (0,1,3) and ARIMA (0,1,2) are the best for local, modern and total Boro rice production respectively [41].

Badmus and Ariyo (2011) in their research used ARIMA model to forecast the cultivated area and production of maize in Nigeria. The result also shows maize production forecast for the year 2020 to be about 9952.72 tonnes with lower and upper limits of 6479.8 and 13425.64 thousands tonnes respectively [10].

Awal, M. A and Siddique M.A. B (2011) modelled rice production in Bangladesh using ARIMA model. The study was carried out to examine the best ARIMA model to efficiently forecast Aus, Aman and Boro rice production in Bangladesh. ARIMA (4,1,1), ARIMA (2,1,1) and ARIMA (2,2,3) to Aus, Aman and Boro rice production respectively [6].

P. Barber and R., Barber (2002) in their research paper used ARIMA model to study factors affecting gold prices in India. They listed different factors that affect gold prices and gave emphasis on the rise of gold prices between 2002 -2012 [37].

Aponsu and Jayasundra (2012) used polynomial regression to forecast tea prices and export in Sri-Lanka. They used time index of tea prices as explanatory variables [4].

Krishnarani (2013) applied ARIMA model to view the volatility of tea prices in terms of variations rainfall, drought, pest outbreaks and climatic variations. They used present and lagged prices as explanatory variables for modeling tea prices [25].

Hettiarachchi and Banneheka (2013) used time series regression with generalized least squares and Artificial Neural Network (ANN) approaches for forecasting tea prices. They performed one month ahead forecast for tea auction prices in Colombo, Kolkata, Cochin, Guwahatti, Chittagong, Mombasa and Jakarta. A Box-Cox transformation was used to deal with non-normality. Their results indicated a significant positive correlation between prices at Colombo centres and those of other auction centres. Moreover, they compared the results from both the models and concluded the (VNN) performed slightly better than time series regression approach. However, they were not able to validate their VNN due to the unavailability of weekly tea prices at auction centres with exception of Colombo [23].

Kumar M. et. al (2014) used Auto regressive Integrated Moving Average model to predict sugar cane prices in India [26].

Dentm and Swanson (1978), studied the forecasting with limited ARIMA models of the trailer and the flat car transportation markets and evaluated the appropriateness of ARIMA time series models in forecasting information-scarce environments, and a specific transportation sector examined. It is shown that the ARIMA model adequately characterised the alternative models and provided the information required for market efficiency. Specific attention is forecast on a procedure for isolating potential model misspecification and the need for post sample quality analysis [16].

Bessler (1982), reviewed the relationship between the exponential, weighted Moving Average and Optional univariate statistical predictors. He showed that the behavioural based adaptive expectation were a sub-class of both the exponentially weighted Moving Average, and the ARIMA(0,1,1) model. The applicability of the expectations model to 25 empirical price and quantity series was investigated. The adaptive expectations behaviour and optimal statistical forecasts were equivalent to 13 series 11 on yield and 2 on prices [9].

Rahul Tripathi et. al (2014) studied forecasts of rice area production of Odisha. Forecasts was made from historical data of 1950-51 to 2008-09 by using univariate ARIMA and was compared with the forecast of India data. The Autoregressive, p and Moving Average q, were identified based on the significant spikes in plots of ACF and PACF of different time series. ARIMA (2,1,0) was found suitable for all Indian rice productivity and production whereas ARIMA (1,1,1) was best fitted for forecasting rice productivity and production in Odisha [39].

Rajan and Palanivel (2017), Used ARIMA model to forecast cotton production of India by fitting of Univariate Autoregressive integrated moving average models. The data on cotton production collected between 1951-2011 are calculated based on the selected model. The study considered ARIMA model introduced by Box and Jenkins. Based on ARIMA p,d,q and its component ACF, PACF, normalized BIC, Box-Ljung Q-statistics and residuals estimated. ARIMA (0,1,0) was selected as the best model for forecasting cotton production in India [40].

Manoj and Madhu (2014) studied the forecast of sugar cane production in India. The best ARIMA model was found to be (2,1,0). Furthermore efforts were made to forecast as accurate as possible the future sugar cane production for the period upto five years by fitting ARIMA (2,1,0) model to the time series data. The forecast showed that the annual sugar cane production will grow in 2013 then it will take a sharp dip in 2014 and in subsequent years 2015 through 2017 and it will continuously grow with an average growth rate of approximately three percent every year [34].

Banhi G. et al (2016), studied the application of ARIMA time series model to forecast the future Gold price in Indian browser based on past data from November 2003 to January 2014 to mitigate the risk in purchase of Gold. ARIMA (1,1,1) was chosen from six different model parameters as it provided the best model which satisfied all the criterion of fit statistics [8].

Mohamad As'ad (2012), modelled the peak daily electricity demand using half hourly demand data. He coined for ARIMA models based on past three, six and twelve months of data and suggested that the ARIMA model build based on past three months data is the best model in terms of forecasting two to seven days ahead and the ARIMA model

based on the past six months data is the best model to forecast one day ahead [32].

Abdullah Lazim (2012) in his paper has addressed the forecasting of gold bullion coin prices through ARIMA model and concluded by suggesting that the gold bullion coin selling prices are in upward trends and could be considered as a worthy investment. The result suggested that ARIMA (2,1,0) was the most suitable model to be used for forecasting gold bullion coin prices [2].

Gupta G.S (1993), studied ARIMA model in details and developed an ARIMA model for tea production in India using monthly data for the period January 1979 through to July 1991 and applied the model to forecast tea production during the future twelve months [29].

Seneviratna,D. and Shuhua, M. (2013) used univariate time series ARIMA model to forecast government twelve month treasury bill rates in Sri-lanka over the period June 2008 to June 2013 [43].

Zakria and Muhammad (2009) studied population of Pakistan from 1951-2007 and modelled using Box-Jenkins ARIMA methodology. The population of Pakistan is also used to forecast for the next twenty years using the parsimonious ARIMA (1,2,0) model. If the growth rate trend continues the population of Pakistan would be approximately 230.7 in 2027 [50].

Nochai and Nochai (2006), studied model of forecasting oil palm price of Thailand in three types as farm price, wholesale price and pure oil price for a period of five years (2000-2004). The objective of the research was to find an appropriate ARIMA model for forecasting in three types of oil palm price by considering the minimum of mean absolute percentage error (MAPE). The results of forecasting were as follows: ARIMA model for forecasting farm price of oil palm is ARIMA (2,1,0), ARIMA model for forecasting wholesale price of oil palm is ARIMA (1,0,1), and ARIMA model for forecasting pure oil price of oil palm is ARIMA (3,0,0) or AR(3) [35].

Mondal et al (2014), conducted a study on the effectiveness of ARIMA model on fifty six Indian stocks from different sectors. The model was chosen because of it's simplicity and wide acceptability of the model. The comparison and parametrization of the ARIMA model have been done using Akaike information criterion (AIC) [33].

Ashwini D. et. al (2017) studied forecasting of common Paddy prices in India. The time series data on monthly average prices of Paddy from January 2006 to December 2016 was collected. ARIMA model was employed to predict the future prices of Paddy. Model parameters were estimated using the R programming software. The ARIMA model was the most representative model for the price forecast of Paddy in overall India [5].

Chinn, M.D., et. al (2005) studied the predictive content of energy futures. They examined the relationship between spot and future prices for energy commodities. An ARIMA (1,1,1) was used for crude oil prices forecast [15].

Wabomba, Musundi (2016) studied modelling and forecasting Kenyan Gross Domestic Product using ARIMA. GDP data was obtained from the Kenya National Bureau of Statistics for the years 1960 to 2012. R statistical software was used to build ARIMA model using the Box-Jenkins method to model the GDP. ARIMA (2,2,2) time series model was established as the best for modelling the Kenyan GDP according to the recognition rules and stationary test of time series under the AIC criterion. Finally, they used the model to forecast the GDP of Kenya for the next five years [48].

Wali et. al (2013) conducted a study on forecasting the cultivated area and production of cotton in India using Autoregressive Integrated Moving Average (ARIMA) model. Time series data covering the period of 1950-2010 was used for the study. The study revealed that ARIMA (0,1,0), ARIMA (1,1,4) and ARIMA (0,1,1) are the best fitted model for forecasting of cotton area, production and yield in India respectively [49].

CHAPTER THREE

RESEARCH METHODOLOGY

3.1 Time series

Time series is a collection of observations or measurements on quantitative variables made sequentially in a uniform set of time period, usually daily, weekly, monthly, quarterly ,annually and so on. An example of time series data are the weekly tea prices at Mombasa tea auction center.

3.1.1 Time series Analysis

Time series analysis comprises of methods or processes that break down time series data into components and explainable portions that allow trends to be identified and forecasts to be made. Basically time series analysis attempts to understand the underlying contexts of the data points through the use of a model to forecast future values based on past values. Such models include GARCH, EGARCH, ARIMA, e.t.c. The main focus of this study is the ARIMA model.

3.1.2 Components of time series

A vital step in choosing appropriate modelling and forecasting procedure is to consider the type of data patterns exhibited from the time series graphs of the time plots. The sources of variation in terms of patterns in time series data are mostly classified into four components. These components include seasonal variation, trend variation, cyclic changes and the remaining irregular fluctuations.

3.1.3 The trend(T)

The trend is the underlying long term behaviour or pattern of the data or series. The Australian Bureau of statistics (ABS 2008) defined trend as the long term movement in

a time series without a calendar related and irregular effects and is a reflection of the underlying level.

3.1.4 Seasonal Variation(S)

This is a cycle that repeats itself over time such us monthly or yearly. Multiple box plot can be used to detect seasonality in time series data. The box plot for tea auction prices is shown on fig 3.1.

3.1.5 Cyclic variation

These are the short term variations fluctuations i.e rises and falls that exist the Kenyan tea auction prices that are not of a fixed period. They may be due to unexpected or unpredictable events.

3.1.6 Random variation

They correspond to the movements that appear irregularly and generally during short periods. Irregular variations do not follow a particular model and are not predictable.

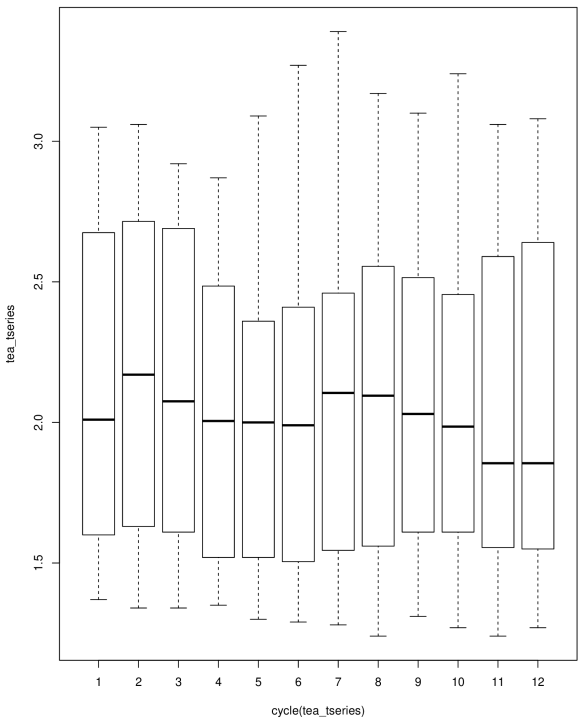

Figure 3.1: Box-plot for the Tea auction prices

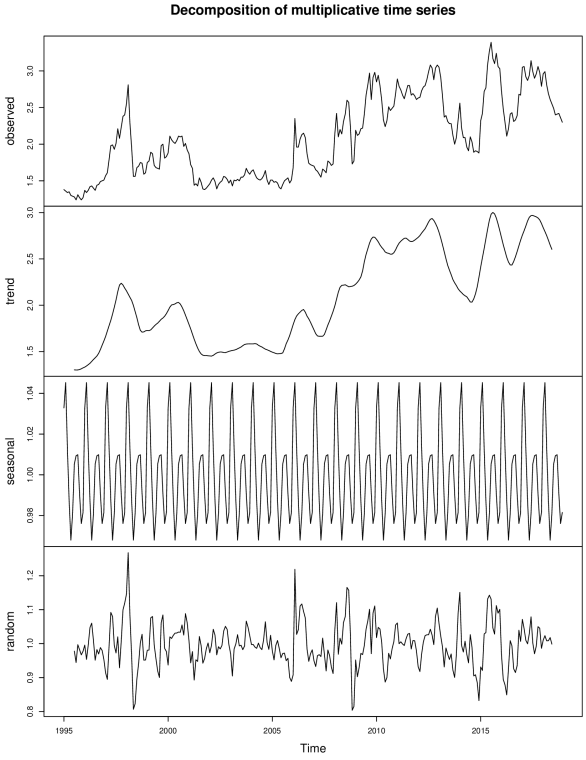

Figure 3.2: Components time series plot of tea auction prices from 1995-2018

3.2 Mean and auto co-variance functions

The mean function of a time series $\{X_t : t \in Z\}$ is defined by $\mu(t) = E(X_t)$ and the auto co-variance function denoted by $\delta(s,t) = cov(X_s, X_t)$ [11].

3.3 Stationary time series

A time series is said to be stationary if it has the following characteristics:

- The mean (μ) is constant $\forall t$
- The variance (σ^2) is constant $\forall t$
- The auto co-variance function between X_{t1} *and* X_{t2}only depends on the interval t_1 *and* t_2

  In a nut shell, a time series is stationary if its statistical properties such as mean, variance, covariance are constant over time otherwise non stationary.

3.4 Differencing time series

Difference in time series is the series of changes of time series from one period to the next. IfX_t denotes the value of time series X at time t, then the first difference is given by: $\Delta(X_t) = X_t - X_{t-1}$

3.5 Auto-covariance function

It is a function that gives the covariance of the process with itself at different pairs of time points. The auto-covariance function between the periods t_1 and t_2 is given as follows:

$$Cov(t_1, t_2) = E[(X_{t1} - \mu)(X_{t2} - \mu)]$$

3.6 Auto correlation

This is a characteristic of data in which the correlation between the values of the same variables is based on related objects. It measures the correlation between successive observations at different distances apart. e.g $(X_1, X_2), (X_2, X_3), ...(X_{n-1}, X_n)$ It occurs

mainly due to dependency within the data. In time series, it is defined as a delayed correlation of a given series by some number. Auto-correlation depicts types of curves which show certain kinds of patterns eg. a curve that shows: residual errors, cyclical patterns of upward or down ward movement, and so on [11].

3.7 Auto correlation function(ACF)

Auto correlation function is a measure of the correlation between observations of a time series that are separated by K time-units (y_t *and* y_{t-k}). It is one of the tools used to find patterns in data. The auto correlation function between t_1 *and* t_2 is given by:

$$Corr(X_{t1}, X_{t2}) = \frac{Cov(X_{t1}, X_{t2})}{\sqrt{VarX_{t1}, VarX_{t2}}} \tag{3.1}$$

3.8 Ljung-Box Q (LBQ)Statistic

LBQ statistic tests the null hypothesis that the autocorrelation up to lag K is equal to zero indicating that the data values are random and independent upto certain number of lags [13].

3.9 Back shift operator

Back shift operator denoted by B is an operator that shifts stochastic time series in a given time step.
Let the time series be an ARMA(p q) model i.e

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2}..... + \alpha_p X_{t-p} + e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2}..... + \theta_q e_{t-q}$$
$$= X_t - \alpha_1 X_{t-1} - \alpha_2 X_{t-2}..... - \alpha_p X_{t-p} = e_t + \theta_1 e_{t-1} + \theta_2 X_{t-2}..... + \theta_p X_{t-p}$$

With the back shift operator (3.2) becomes:

$$X_t - \alpha_1 B^1 X_t - \alpha_2 B^2 X_t.....\alpha_p B^p X_t = e_t + \theta_1 B^1 e_{t-1} + \theta_2 B^2 e_{t-2}..... + \theta_q B^q \tag{3.2}$$

3.10 The Auto regressive Inter grated Moving Average(ARIMA) model

ARIMA model is one of the methods of forecasting time series data. We intend to use this method to forecast tea pricing at Mombasa tea auction. It is therefore imperative to examine the components of ARIMA model. In full ARIMA is auto regressive integrated moving average [11,12]. AR is the auto regressive part, I (integrated) component and MA is the moving average part.

3.10.1 Auto regressive(AR)process

A time series X_t is said to be auto regressive if it satisfies the expression $X_t = \psi_1 X_{t-1} + \psi_2 X_{t-2}.....\psi_p X_{t-p} + e_t$ where e_t is white noise and ψ_i are constants, then X_t is said to be an auto-regressive model of order p denoted by $AR(P)$ [11].

3.10.2 Moving Average process(MA)

A time series X_t is said to be a moving average process of order q , $MA(q)$ if it satisfies the expression $X_t = e_t + \zeta_1 e_{t-1} + \zeta_2 e_{t-2}..... + \zeta_q e_{t-q}$ where e_t are white noise [11].

3.10.3 Auto-regressive Moving Average series(ARMA)

A time series X_t is said to auto regressive moving average process denoted by $ARM(p,q)$ if it satisfies the expression $X_t = \psi_1 X_{t-1} + \psi_2 X_{t-2}.....\psi_p X_{t-p} + e_t \zeta_1 e_{t-1} + \zeta_2 e_{t-2}..... + \zeta_q e_{t-q}$ where e_t are white noise. $ARMA$ model is a combination of $AR(p)$ *and* $MA(q)$ models [11].

3.10.4 ARIMA model

ARIMA models for forecasting time series are essentially agnostic. Unlike other methods they do not assume knowledge of any underlying economic model or structural relationships. It is assumed that past values of the series plus previous errors terms contain information for the purpose of forecasting. They have proven themselves to be relatively robust especially when generating short-run forecasts. They frequently outperform more sophisticated structural models in terms of short-run forecasting ability [31]. The ARIMA

model is used widely in medical research and it provides a comprehensive model in the domain of Time series analysis. Time series predictions are based on changes over time in historical data sets and they can produce mathematical models by using statistical data that can be extrapolated [51]. The auto regressive integrated moving average model denoted by $ARIMA(p, d, q)$ is obtained by differencing the $ARMA(p, q)$ d times. ARIMA models are a class of models that have the capabilities to represent stationary as well as non-stationary time series to produce accurate forecast based on a description of a historical data of a single variable. The ARIMA models do not assume any particular pattern hence making it different from other models [11,13]. Box-Jenkins approach of building ARIMA model is based on the following steps:

- Model identification
- Parameter estimation and selection
- Model validation
- Model use

3.11 Model identification

This involves determining the order of (p, d, q) of AR and MA components of the model in order to capture the silent dynamic features of the data. It seeks to check whether the data is stationary or non stationary and the order of differentiation d which makes the time series stationary. It mainly leads to use of graphical procedures . We have used monthly pricing data from 1995 -2018 to identify the model. If the graph of ACF of time series values either cuts off fairly quickly or dies down fairly quickly then the time series data is considered stationary and if the graph of A C F dies down extremely slowly then the time series is considered non stationary. To build the ARIMA model we have to make the non-stationary time series data to stationary. This is achieved through differencing time series data until the series vary about a fixed level [11,12,13].

3.11.1 Model estimation and selection

This involves the estimation of parameters of different models using identification process and proceeds the selection of the model using information criteria. We have used the auto.arima command in R statistical software to come up with the best ARIMA model with the lowest Akaike(AIC) Information Criteria and Bayesian Information Criteria(BIC). The Akaike Information Criteria(AIC) and Bayesian Information Criteria(BIC) are two measures of goodness of fit and they measure the trade off between the model fit and the complexity of the model.

R statistical language is free, open source computer language and software environment for statistical computing. R also provides quality plotting and graphing which are used when building an ARIMA model.

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

$$BIC = -2log(L) + log(N)K \tag{3.4}$$

Where:

L is the value of the likelihood function evaluated at the parameter estimates

N is the number of observations

K is the number of estimated parameters.

A lower AIC and BIC value indicate a better model fit [3].

The Bayesian Information Criterion is a criterion for model selection among a finite set of models. The BIC is widely used for model identification in time series and linear regression. When fitting models, it is possible to increase the likelihood by adding parameters, but doing so may result in over fitting. The BIC resolves this problem by introducing a penalty term for the number of parameters in the model. The penalty term is larger in BIC than AIC.

The AIC is an estimated measure of the quality of each of the available econometric models as they relate to one another for a certain set of data. A lower BIC value indicates lower penalty terms hence a better model. The only difference between BIC and AIC is that BIC considers the number of observations in the formula, which AIC does not.

3.12 Model checking

At this step we checked the residual properties of the model. We checked whether the residual properties from ARIMA model are random and have a normal distribution. We have used Ljung-Box Q statistic and the histogram of the residuals. The test statistic Q is given by:
$Q_m = n(n+2)\sum_{k=1}^{m} \frac{r^2 k(e)}{n-k}$
where $rk(e)$ is the residual auto-correlation at lag k
n is the number of the residuals
m is the number of time lags included in the test.
If the p-value associated with the Q statistic is small(p-value $< \alpha$)then the model is considered adequate [11].

- Forecasting with ARIMA model

In this step we used monthly data for Kenya tea exportation from the year 1995-2018 to build the model. The ARIMA model is given by:
$\phi_p(B)\Delta^d X_t = \theta_q(B)\xi_t$.

- Checking the Accuracy of the Forecast

In this step we used the Mean Absolute Error, Root Mean squared Error and the Mean Absolute Percentage Error to check the variation in the errors in a set of forecasts. Lower values of the errors indicate a better fit.
The Mean Absolute Error (MAE)
measures the average absolute deviations of forecast values from the original ones. It shows the magnitude of the errors in a set of forecasts without considering their direction [52]. It is given by :

$$\frac{\sum_{i=1}^{n} |y_i - x_i|}{n}$$

Where y_i is the prediction

x_i is the true value

n is the number of points.

The Root Mean Squared Error (RMSE)

is the square root of the variance of the residuals. It indicates the absolute fit of the model to the data-how close the observed data points are to the model's predicted values [52]. Lower values of RMSE indicate a better fit. It is given by:

$$RMSE = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_i - \acute{y_j})^2}$$

Where y_j is the predicted value

$\acute{y_j}$ is the true value

n is the number of points.

The Mean Absolute Percentage Error (MAPE)

is a statistical measure of how accurate of how accurate a forecast system is. It measures this accuracy as a percentage and can be calculated as the average absolute percent errors minus actual values divided by actual values [53]. It is given by:

$$MAPE = \frac{100}{n}\sum_{t=1}^{n}\frac{|A_t - F_t|}{A_t}$$

Where A_t is the actual value

F_t is the forecast value

n is the number of fitted points.

CHAPTER FOUR

RESULTS AND DISCUSSIONS

4.1 Time series data for tea auction prices

This chapter deals with the analysis of the tea prices at the Mombasa tea auction from 1995-2018 as shown in the appendix.
Time series plots display observations on the y-axis against equally spaced time intervals on the x- axis. They are used to evaluate patterns and behaviours in data over time as displayed in the figure below. The data has an upward trend and hence non stationary.

4.2 Stages of building the ARIMA model

The auto regressive integrated moving average model denoted by $ARIMA(p, d, q)$ is obtained by differencing the $ARMA(p, q)$ d times. ARIMA models are a class of models that have the capabilities to represent stationary as well as non-stationary time series to produce accurate forecast based on a description of a historical data of a single variable. The ARIMA models do not assume any particular pattern hence making it different from other models [11,12]. Box-Jenkins approach of building ARIMA model is based on the following steps:

- Model identification
- Parameter estimation and selection
- Model validation
- Model use

4.3 Model identification

This involves determining the order of (p, d, q) of AR and MA components of the model in order to capture the silent dynamic features of the data. It seeks to check whether the data is stationary or non- stationary and the order of differentiation d which makes the time series stationary. It mainly leads to use of graphical procedures . We have used monthly pricing data from 1995 -2018 for model validation.

The plot of the ACF data shows a slow decay hence the data is non- stationary. Therefore we have to transform the non- stationary data into stationary by differencing the series. The ACF plot of differenced series shows that the time series is stationary.

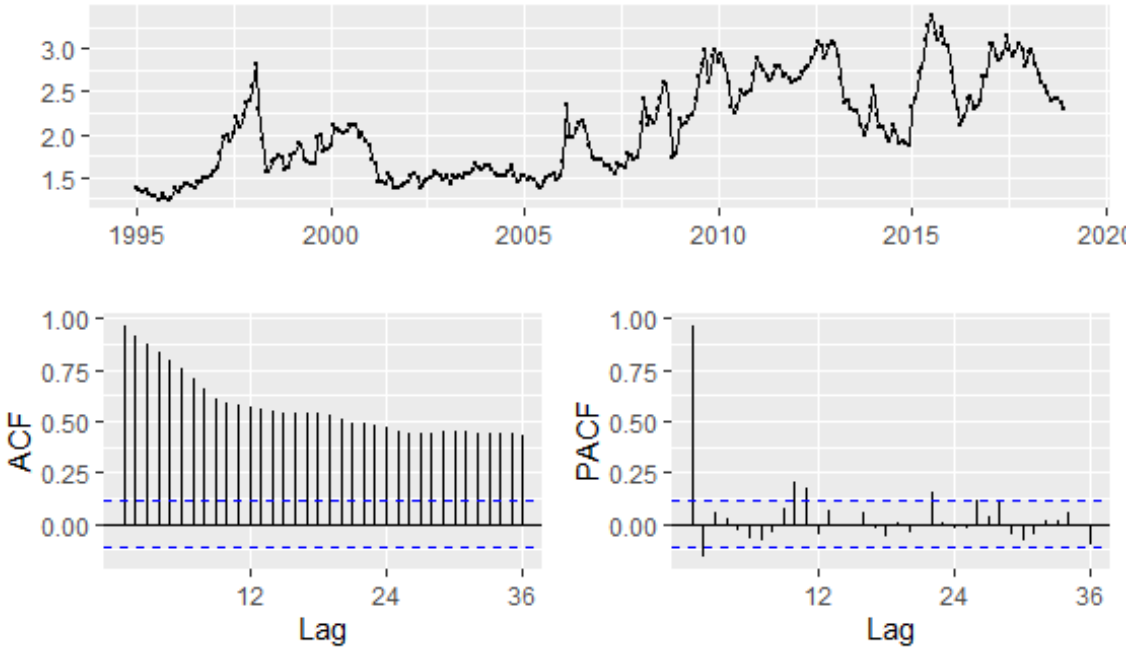

Figure 4.1: Non-stationary times series data for tea auction prices

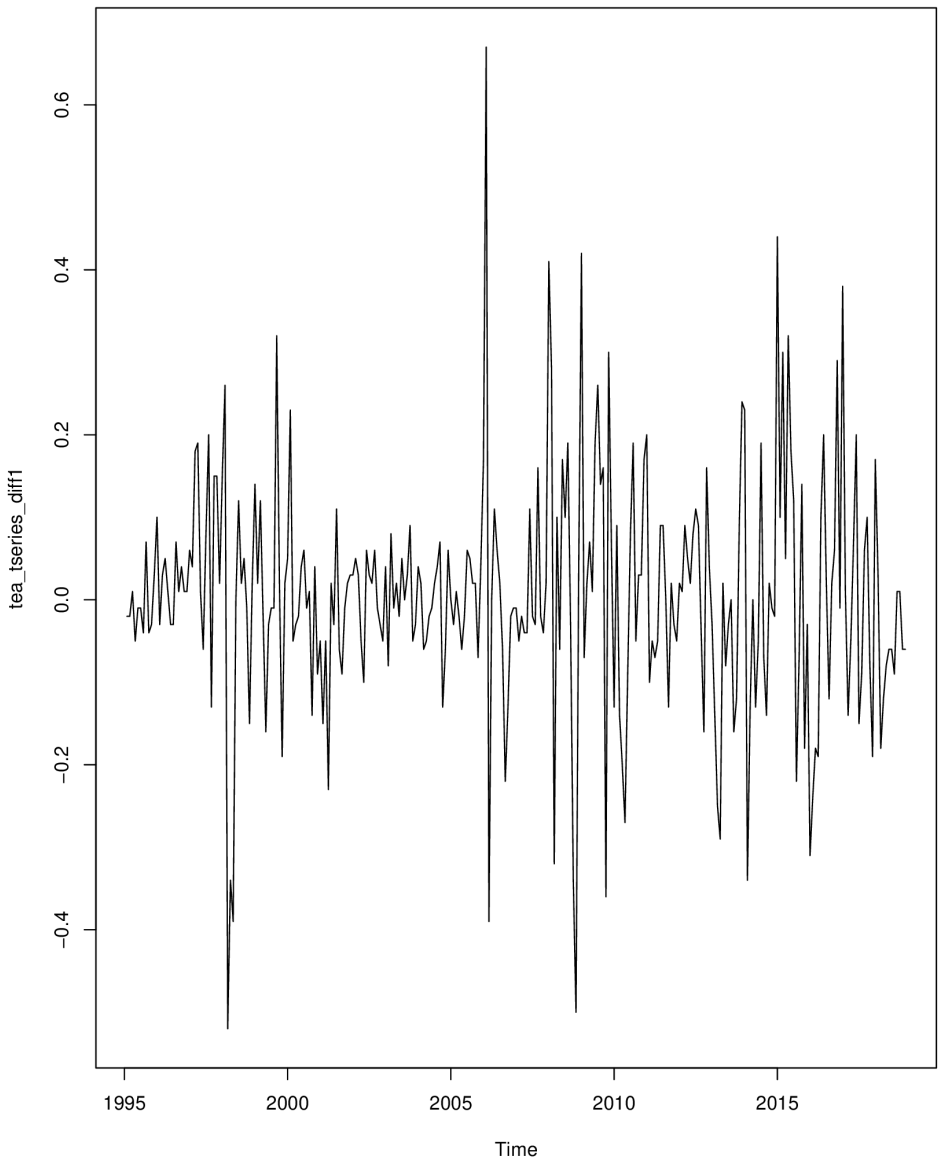

Figure 4.2: First differenced series

The time series of the first difference appears to be stationary in mean and variance and so an ARIMA (p,1,q) model is probably appropriate for the time series of the tea auction prices. The Augmented Dickey fuller check showed a p-value of 0.01 signifying that that tea price is stationary at first differencing.

4.4 Parameter estimation and selection

This involves the estimation of parameters of different models using identification process and proceeds to the selection of the model using information criteria. We have used the auto.arima command in R statistical software to come up with the best ARIMA model with the lowest Akaike(AIC) Information Criteria and Bayesian Information Criteria(BIC) as shown in table 4.2. The estimated parameters are as follows:

Table 4.1: Parameter estimates for ARIMA(1, 1, 2)

Coefficients	Estimate	Standard error
ar1	0.8911	0.0490
ma1	-0.7530	0.0713
ma2	-0.2055	0.0591

Table 4.2: Model fit statistics

AIC	AICc	BIC
-309.69	-309.55	-295.03

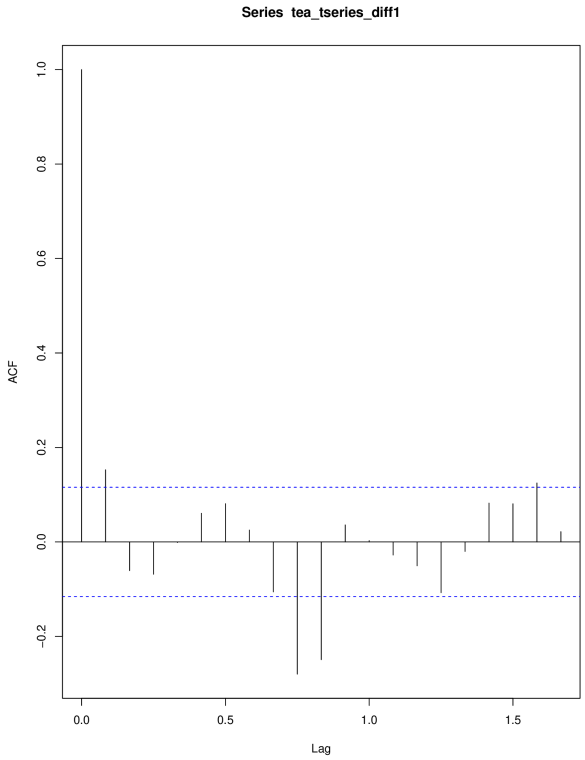

Figure 4.3: ACF for the differenced series

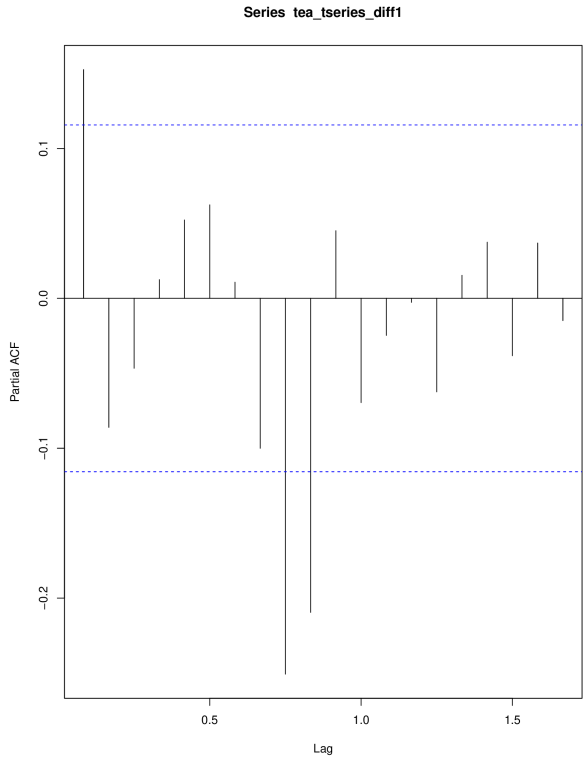

Figure 4.4: PACF for the differenced series

The above diagrams show plots of the ACF and PACF of the differenced tea series. The lags of ACF and PACF were generally displaying small autocorrelations, making it close to a white noise. The ACF tails off towards zero (exponential decay) and PACF died out slowly after lag 1.

4.5 Model checking

After having chosen ARIMA (1,1,2) as the best model, the model adequacy is further checked to draw empirical conclusion regarding the model as good fit hence its usage in estimation and forecasting. These tests were performed using Ljung Box test coupled with the ACF, PACF, the Q-Q plot and histogram plots of the residuals. The plots showed that the residuals from the model are similar to a white noise hence the model fits the data well. The p-value estimated by the Ljung Box test in the table 4.3 was greater than 0.05. The p-value was at 0.8834 which showed that the residuals were random.

Table 4.3: Ljung Box test for ARIMA(1,1,2)model

Test type	X -square(χ)	df	p-value
Ljung Box	1.743	5	0.8834

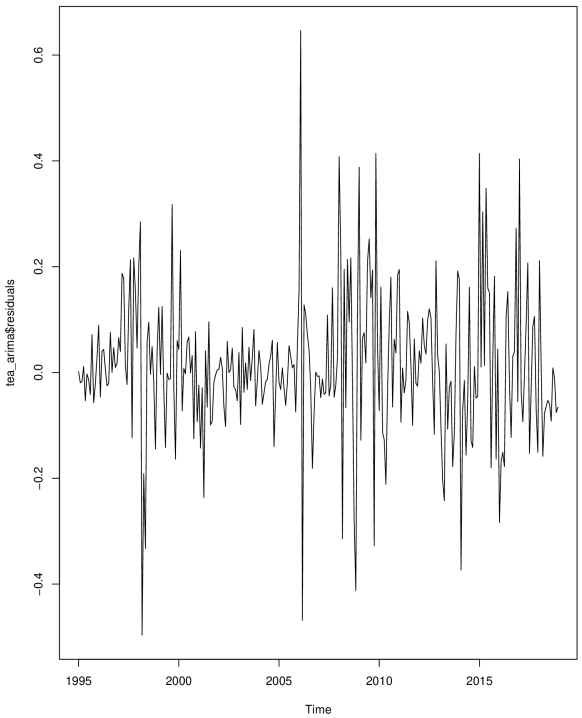

Figure 4.5: Residuals for the ARIMA (1,1,2)

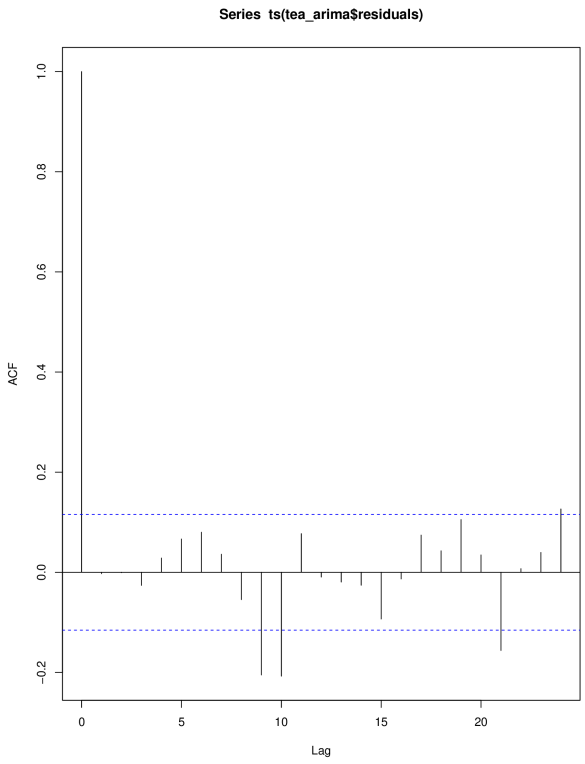

Figure 4.6: ACF plot of the ARIMA model residuals

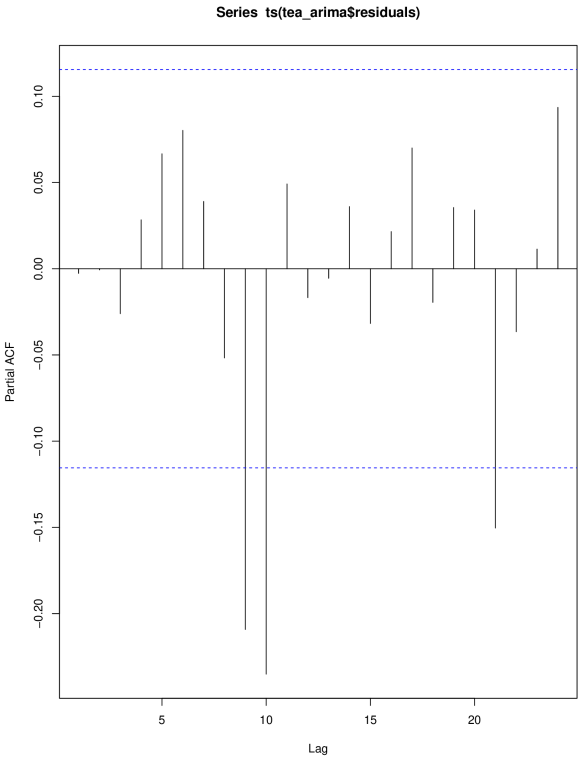

Figure 4.7: PACF for the model ARIMA(1,1,2)residuals

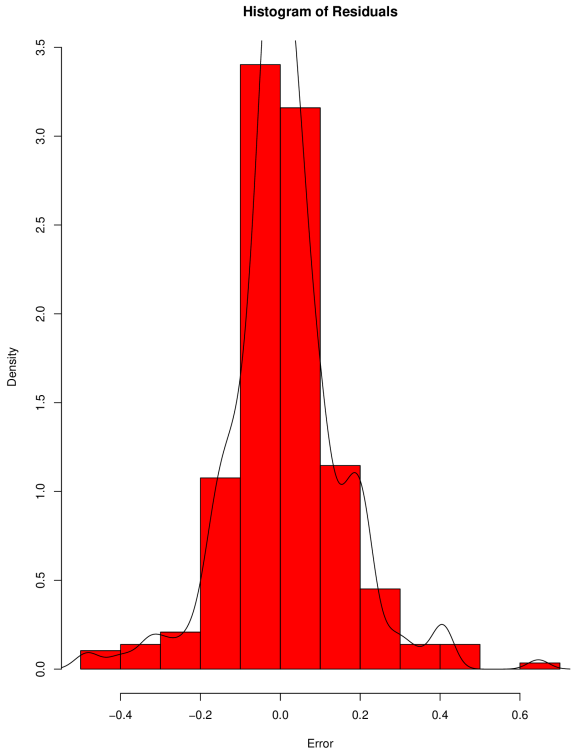

Figure 4.8: Histogram of the model residuals

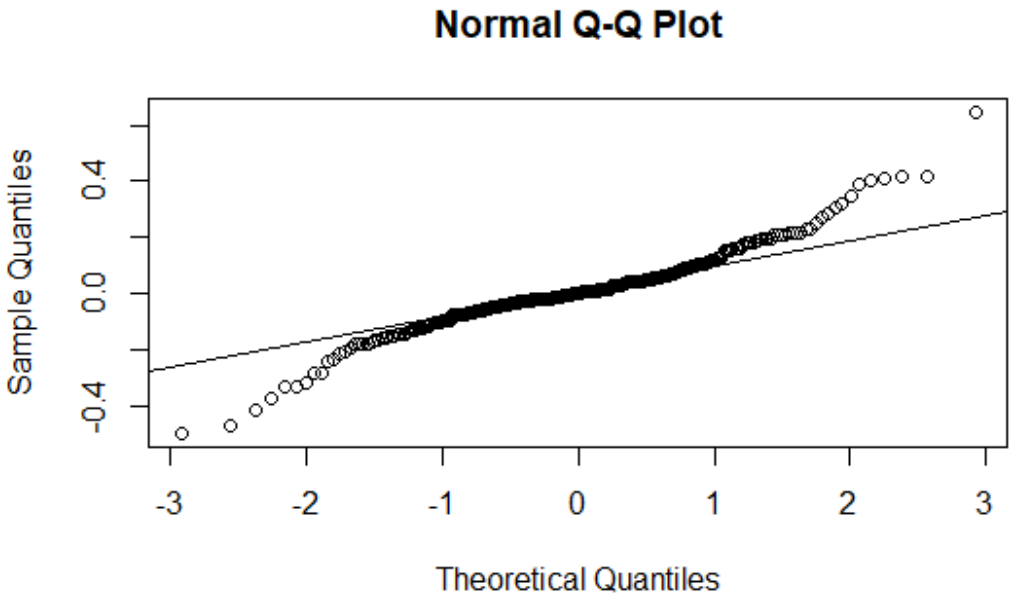

Figure 4.9: The normal Q-Q plot for the residuals of the ARIMA (1,1,2) model

	ME	RMSE	MAE	MPE	MAPE	MASE
Training set	0.0102652	0.1387933	0.09598332	0.2160826	4.469296	0.254872

Figure 4.10: Accuracy of ARIMA (1,1,2) model

Residuals are estimates of experimental error obtained by subtracting the observed responses from predicted responses. The overall pattern of the residuals should be similar to the bell- shaped pattern observed when plotting a histogram of normally distributed data.

The Q-Q plot points lie on a straight line, the deviations from the straight line are minimal which indicates a normal distribution. The residuals plots of the ARIMA (1,1,2) model shown above confirm that the residuals are random with no pattern hence we are sure that the model is fitting the data well and the statistical assumptions are all met, therefore we can use the model for forecasting.

4.6 Forecasting with ARIMA model

In this step we used monthly data for Kenya tea exportation from the year 1995-2018 to build the model. The ARIMA model is given by:

$\phi_p(B)\Delta^d X_t = \theta_q(B)\xi_t$ where:

p=The AR component

d=The differencing part

q=The MA component

For tea auction model,the values of p= 1 , d=1 and q=2. Hence our model becomes

$$\phi_1(B)\Delta^1 X_t = \theta_2(B)\xi_t \tag{4.1}$$

Using R statistical software we obtain the model parameters as follows: $\phi_1 = 0.8911, \theta_1 = -0.7530, \theta_2 = -0.2055$

Therefore (4.1) becomes:

$$X_t = 0.8911X_{t-1} - 0.7530\xi_{t-1} - 0.2055\xi_{t-2} + \xi_t \tag{4.2}$$

With (4.2) , we have used it to forecast the tea auction prices from 2020 to 2023 as shown in fig.4.13. The values for the forecast are represented graphically as shown in fig 4.11.

After selection of an appropriate model, future values of the time series are as shown in figure 4.13. The forecast values are shown graphically in fig 4.12 at a confidence level of 80% and 95% The actual data did not match the predicted data perfectly, but fell within

the predicted 80% confidence interval. The forecast prices and their lower and upper limits are shown in figure 4.13 to 4.18. The MAE, RMSE and MAPE are shown in table in table 4.10.

```
ARIMA(1,1,2)

Coefficients:
          ar1      ma1      ma2
      0.8911  -0.7530  -0.2055
s.e.  0.0490   0.0713   0.0591

sigma^2 estimated as 0.01953:  log likelihood=158.85
AIC=-309.69   AICc=-309.55   BIC=-295.05
```

Figure 4.11: Coefficients of ARIMA (1,1,2)

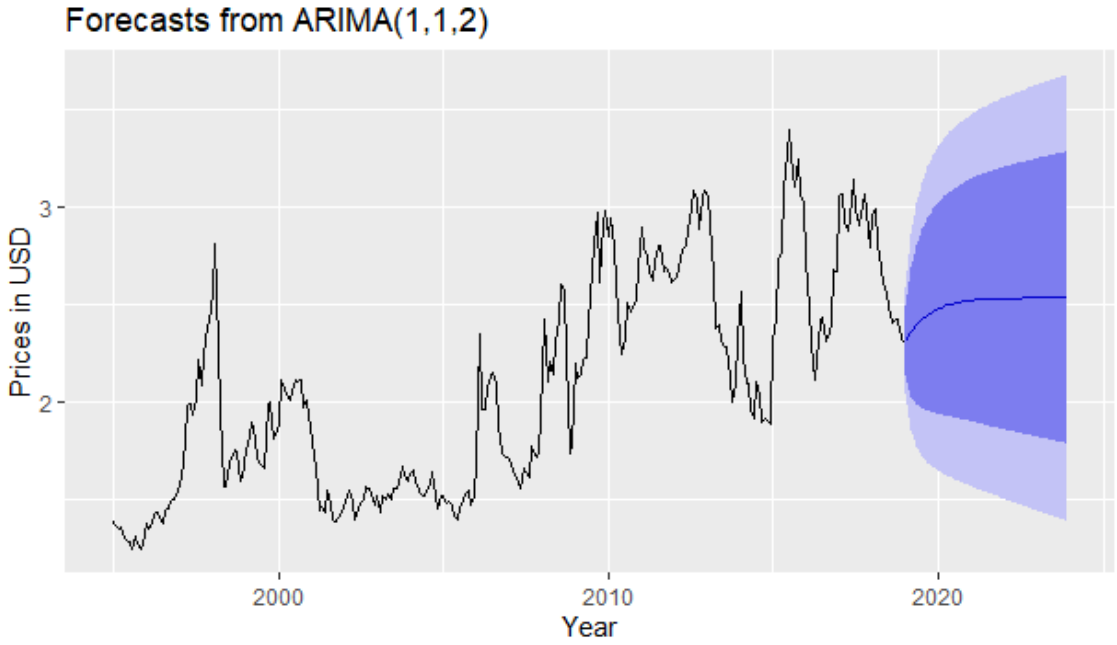

Figure 4.12: Forecast from ARIMA (1,1,2)

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Jan 2019	2.311657	2.132538	2.490776	2.0377183	2.585596
Feb 2019	2.335573	2.064198	2.606949	1.9205400	2.750606
Mar 2019	2.356885	2.026114	2.687656	1.8510145	2.862756
Apr 2019	2.375877	2.001209	2.750544	1.8028725	2.948881
May 2019	2.392801	1.983708	2.801894	1.7671476	3.018454
Jun 2019	2.407883	1.970826	2.844939	1.7394627	3.076302
Jul 2019	2.421322	1.960991	2.881653	1.7173069	3.125337
Aug 2019	2.433299	1.953232	2.913365	1.6991004	3.167497
Sep 2019	2.443971	1.946915	2.941028	1.6837888	3.204154
Oct 2019	2.453482	1.941609	2.965355	1.6706399	3.236324
Nov 2019	2.461957	1.937016	2.986897	1.6591298	3.264784
Dec 2019	2.469509	1.932926	3.006093	1.6488759	3.290143

Figure 4.13: Upper and lower limits of the predicted Tea prices in 2019

Jan 2020	2.476240	1.929187	3.023293	1.6395942	3.312885
Feb 2020	2.482237	1.925689	3.038785	1.6310708	3.333403
Mar 2020	2.487582	1.922356	3.052807	1.6231435	3.352020
Apr 2020	2.492344	1.919129	3.065559	1.6156877	3.369001
May 2020	2.496588	1.915969	3.077208	1.6086076	3.384569
Jun 2020	2.500371	1.912846	3.087896	1.6018287	3.398912
Jul 2020	2.503741	1.909739	3.097743	1.5952929	3.412189
Aug 2020	2.506744	1.906634	3.106855	1.5889549	3.424534
Sep 2020	2.509421	1.903522	3.115319	1.5827788	3.436063
Oct 2020	2.511806	1.900397	3.123215	1.5767367	3.446875
Nov 2020	2.513931	1.897255	3.130607	1.5708061	3.457056
Dec 2020	2.515825	1.894094	3.137556	1.5649696	3.466680

Figure 4.14: Upper and lower limits of the predicted Tea prices in 2020

Jan 2021	2.517513	1.890914	3.144111	1.5592132	3.475813
Feb 2021	2.519017	1.887716	3.150318	1.5535255	3.484508
Mar 2021	2.520357	1.884500	3.156214	1.5478979	3.492816
Apr 2021	2.521551	1.881268	3.161835	1.5423231	3.500780
May 2021	2.522616	1.878022	3.167209	1.5367954	3.508436
Jun 2021	2.523564	1.874764	3.172364	1.5313102	3.515818
Jul 2021	2.524409	1.871496	3.177323	1.5258639	3.522955
Aug 2021	2.525163	1.868219	3.182107	1.5204535	3.529872
Sep 2021	2.525834	1.864935	3.186733	1.5150765	3.536591
Oct 2021	2.526432	1.861647	3.191217	1.5097309	3.543133
Nov 2021	2.526965	1.858356	3.195574	1.5044152	3.549515
Dec 2021	2.527440	1.855063	3.199817	1.4991280	3.555752

Figure 4.15: Upper and lower limits of the predicted Tea prices in 2021

Jan 2022	2.527863	1.851770	3.203956	1.4938682	3.561858
Feb 2022	2.528240	1.848479	3.208002	1.4886348	3.567846
Mar 2022	2.528576	1.845190	3.211963	1.4834272	3.573726
Apr 2022	2.528876	1.841905	3.215847	1.4782446	3.579507
May 2022	2.529143	1.838625	3.219661	1.4730865	3.585199
Jun 2022	2.529381	1.835350	3.223411	1.4679524	3.590809
Jul 2022	2.529593	1.832082	3.227103	1.4628418	3.596343
Aug 2022	2.529782	1.828821	3.230742	1.4577544	3.601809
Sep 2022	2.529950	1.825567	3.234332	1.4526899	3.607210
Oct 2022	2.530100	1.822323	3.237877	1.4476479	3.612552
Nov 2022	2.530233	1.819087	3.241380	1.4426281	3.617839
Dec 2022	2.530353	1.815860	3.244845	1.4376302	3.623075

Figure 4.16: Upper and lower limits of the predicted Tea prices in 2022

Jan 2023	2.530459	1.812643	3.248275	1.4326540	3.628263
Feb 2023	2.530553	1.809436	3.251671	1.4276993	3.633407
Mar 2023	2.530638	1.806239	3.255036	1.4227657	3.638509
Apr 2023	2.530713	1.803053	3.258372	1.4178532	3.643572
May 2023	2.530780	1.799878	3.261682	1.4129613	3.648598
Jun 2023	2.530839	1.796713	3.264966	1.4080900	3.653589
Jul 2023	2.530892	1.793560	3.268225	1.4032390	3.658546
Aug 2023	2.530940	1.790417	3.271462	1.3984081	3.663471
Sep 2023	2.530982	1.787286	3.274678	1.3935971	3.668367
Oct 2023	2.531020	1.784166	3.277873	1.3888058	3.673233
Nov 2023	2.531053	1.781058	3.281049	1.3840340	3.678072
Dec 2023	2.531083	1.777960	3.284206	1.3792814	3.682885

Figure 4.17: Upper and lower limits of the predicted Tea prices in 2023

$pred

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
2019	2.311657	2.335573	2.356885	2.375877	2.392801	2.407883	2.421322	2.433299
2020	2.476240	2.482237	2.487582	2.492344	2.496588	2.500371	2.503741	2.506744
2021	2.517513	2.519017	2.520357	2.521551	2.522616	2.523564	2.524409	2.525163
2022	2.527863	2.528240	2.528576	2.528876	2.529143	2.529381	2.529593	2.529782
2023	2.530459	2.530553	2.530638	2.530713	2.530780	2.530839	2.530892	2.530940

	Sep	Oct	Nov	Dec
2019	2.443971	2.453482	2.461957	2.469509
2020	2.509421	2.511806	2.513931	2.515825
2021	2.525834	2.526432	2.526965	2.527440
2022	2.529950	2.530100	2.530233	2.530353
2023	2.530982	2.531020	2.531053	2.531083

Figure 4.18: The predicted values for the Tea prices from 2019 to 2023

CHAPTER FIVE

CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions

The final chapter of this study is based on the previous four chapters and it deals with the conclusion. Based on the objectives as well as the literature reviewed and from the analysis it is imperative to draw the following conclusion. First, with regard to the main objective, it can be concluded that ARIMA (1,1,2) was found to be significant and adequate. The Ljung Box test was further corroborated by diagnostics of the residual plots. It showed that the residuals of the ARIMA (1,1,2) were uncorrelated, hence the model was appropriate and working well. It can be concluded that the tea auction prices at the Mombasa tea auction are on an upward trend therefore tea farmers are advised to maintain the plants and even plant more since they likely to fetch better prices in future.

5.2 Recommendations

Tea is a major cash crop grown in Kenya, Kenyan tea has been the major foreign exchange earner in Kenya. The government of Kenya through the The Kenya Tea Development Authority(KTDA), should formulate and implement policies that will encourage farmers to maintain their tea plantations and even plant more since tea prices are on an upward trend.

REFERENCES

[1] Abdul T.K (1988). *How to choose the right forecasting technique.*Harward Business Rev July-August PP. 45-74.

[2] Abdullah L. (2012). *"ARIMA model for gold bullion coin selling prices forecasting."International Journal of Advances in Applied Sciences,Vol.1,no.4,pp.153-158.*

[3] Akaike, H. (1993).*Theory and extention of the maximum likelihood principal,In:Second International Symposium information;Budapest pp.25-129.*

[4] Aponsu G.M et. al (2012).*Time Flunctuation models to forecast tea production,Prices and Exports in Sri-Lanka, Paper presented at the 13th Annual Symposium,Faculty of graduate studies,University of Kaleniya Sri-Lanka.*

[5] Ashwini D.(2017).*Forecasting of common Paddy prices in India.National Institute of Agricultural Extension Management.Journal of Rice Research,Vol.10,No.1 pp 71-75.*

[6] Awal M.A and Sidique M.A.B(2011). *Rice production in Bangladesh by ARIMA model in Bangladesh journal of Agricultural research, 36(1) 51-62.*

[7] Banerjee,D.(2014).*Forecasting of Indian stock market using time series ARIMA model.International Journal Conference on Business and Information Management.ICBIM14.*

[8] Banhi, G. et. al (2016).*Gold price forecasting using ARIMA model.Journal of advance Management.*

[9] Bessler,David A.,(1982)*Adaptive Expectations,the Exponentially Weighted Forecast,and Optimal Statistical Predictors:Agricultural Economics Research,United States Department of Agriculture,Economic Research Service,vol.34(2),pages 1-8,April.*

[10] Badmus M.A and Ariyo, O.S(2011).*Forecasting cultivated area and production of maize in Nigeria using ARIMA model.Asian journal of Agricultural sciences,3(3) 171-176.*

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

[12] Box, G.E.P et. al(1994).*Time series Analysis:Forecasting and control,Third Edition,Engelwood Cliff,NJ Prentice Hall.*

[13] Box G.E.P and Ljung,J.(1978).*"On a measure of lack of Fit in Time series Models,"Biometrika.*

[14] Burnham and Anderson (2004).*Multimodel inference:Understanding AIC and BIC in model selection.*

[15] Chinn,M.D.,et. al (2005).*The predictive content of energy futures:An update on Petroleum,National Gas,Heating oil and Gasoline.First edition National Bureau of Economics Research,pp:17.*

[16] Dent,W. and Swanson,J.(1978).*Forecasting with limited information:ARIMA models of Trailer on Flatcar Transport Market.vol.73,pages 293-299.*

[17] Dickey D.A.,and Fuller W.A.(1979).*"Distribution of estimator for auto regressive time series with Unit Root."Journal of the American Statistical Association.*

[18] Dickey, D.A.,and Hasna D.P.and Fuller W.A(1984).*"Testing for unit roots in seasonal time series."Journal for The American Statistical Association.*

[19] FAOSTAT 2012, Fuller W.A(1976).*An introduction to statistical time series,New York:John wiley and sons Inc.*

[20] Gubnar, John A.(2006).*Probability and Random processes for electrical and Computer Engineers.Cambridge University press.*

[21] Guptar G.S (1993).*ARIMA model for forecasts on Tea production in India.Journal for Indian Institute of Management.*

[22] Gujarati, C.W et. al (1986). *Basic Econometrics Fourth edition, mecraw-hill.,Newyork.*

[23] Hettiarachchi H.A.C.K and Benneheka B.S.M.G(2013). *Time series regression and Artificial Neural Network approaches for forecasting unit tea prices at Colombo auction.Journal of National science foundation Sri-Lanka 41(1) 35-40*

[24] Hicks,A.(2009). *Current status and future development of global tea production and tea prices.AU.JT(12(4)251-264 Retrieved from Http.*

[25] Krishna,Vijaiy(2002). *Auction Theory San Diego USA academic press,ISBN 978-0-12-426297-3.*

[26] Kumar M. and Anand M,. 2014. *Application of time series ARIMA model for predicting sugarcane in India.*

[27] Kun II Park, *Fundamentals of probability and stochastic processes with Application to communications,Springer,2018,978-319-68074-3.*

[28] Liu L.M (1991). *Dynamic relationship Analysis of US crude oil and gasoline prices.J.Forecast.,521-547 DOI:10.1002.*

[29] Liu, H. et. al (2016). *India's Tea Price Analysis Based on ARMA model.Modern Economy,7,118-123.*

[30] Majumdar, B.A et.al (2012). *Tea statistics:Global scenario.Inc.J.Tea. Sci.,8(1)121-124.*

[31] Meyler et. al (1999). *Forecasting Irish inflation using ARIMA models,Central Bank and Financial Services Authority of Ireland Technical paper Series Vol. 1998. No.3/RT/1998(December 1998).*

[32] Mohamad As'ad (2012). *"Finding the best ARIMA model to forecast daily peak electricity demand",in the Fifth Annual ASEARC conference, 2012.*

[33] Mondal et. al (2014).*Study of the effectiveness of Time series modelling (ARIMA) in forecasting stock prices.Vol;4,International journal of computer science, engineering applications.*

[34] Manoj et. al.(2014).*An application of the series ARIMA forecasting model for predicting sugarcane production in India.*

[35] Nochai and Nochai,(2006).*ARIMA model forecasting oil palm price.*

[36] Panjankar,V.D,. and Thakare,A.B.(2009).*Consumption and export of tea in India.International journal of Agriculture and Statistical Sciences,5(1)65-71.*

[37] P.Baber and G.Thomas. (2002) *"Factors affecting Gold prices:A case study of India".*

[38] Raymond Y.C.Tse (1997),*An application of ARIMA model to real-estate prices in Hong Kong.Journal of property Finance Vol.8,No.2, pp 152-163*

[39] Rahul Tripathi et. al(2014).*Forecasting rice productivity and production of Odisha, Idnia using Autoregressive Intergrated Moving Average models vol.2014. Journal Advances in Agriculture, doi. 10.1155/2014/621313.*

[40] Rajan et. al (2017).*Forecasting of Area cotton in India.An application of ARIMA model.Vol:5,International Journal of Pure and Applied Bioscience.*

[41] Rahman,N. (2010).*Forecasting of Boro rice production in Bangladesh;An ARIMA approach.Journal of the Bangladesh Agricultural University.*

[42] Rathnayaka R.M et. al (2017).*Selection of Best ARIMA Modeling Approach for Forecasting Time Series Patterns;A Case Study on Colombo Stock Exchange.Vol.4,International Journal of Business,Economics and Management.*

[43] Senaviratna (2017).*Selection of the best ARIMA model for forecasting Time series pattern;A case study of Colombo stock exchange Vol.4, Journal of Business Economics and Management.*

[44] Smith P.(2011) *Pricing strategy: setting price levels,managing price discounts and establishing price structures,Cengage learning (2011) page 270-272.*

[45] Stergious K.I (1989),*Modelling and forecasting the fisheries in pilchard (Sardina Pichardus) in Geek waters Using ARIMA time series models ICES model of Marine Science,Volume 46 No.1 Pp 16-23.*

[46] Saeed N. et. al (2003).*Forecasting of wheat production in pakistan using ARIMA model,Int.J. Biol. 2:252-253.*

[47] *Tea Board of Kenya 2014.*

[48] Wabomba, Musundi (2016).*Modelling and Forecasting Kenyan GDP using ARIMA models.Vol.4,Science Journal of Applied Mathematics and Statistics.*

[49] Wali et. al (2017).*Forecasting of area and production of cotton in India: An application of ARIMA model.* vol.5 *International journal for pure and Applied Bioscience . dio.10.18782/2320-7051.5409*

[50] Zakria and Mohammad (2009).*Forecasting the Population of Pakistan Using ARIMA models.Vol. 46, Journal of Agric.Sci.*

[51] Zhan ying-han et. al (2012).*Application of an ARIMA model for predicting the incidence of Hemorrhagic Fever with Renal Syndrome. The American Journal of Tropical medicine and Hygiene.*

[52] Zhang G.P (2003).*"A neural network ensemble method with jittered training datd for time series forecasting. " Information sciences 177.*

[53] Zhang G.P (2003).*Time series forecasting using hybrid ARIMA and neural network model.Neurocomputing 50.*

Appendix

Table 5.1: Data for tea auction prices at Mombasa tea auction center

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1995	1.38	1.36	1.34	1.35	1.30	1.29	1.28	1.24	1.31	1.27	1.24	1.27
1996	1.37	1.34	1.37	1.42	1.43	1.40	1.37	1.44	1.45	1.49	1.50	1.51
1997	1.57	1.61	1.79	1.98	1.99	1.93	2.01	2.21	2.08	2.23	2.38	2.40
1998	2.55	2.81	2.29	1.95	1.56	1.56	1.68	1.70	1.75	1.74	1.59	1.61
1999	1.75	1.77	1.89	1.87	1.71	1.68	1.67	1.66	1.98	2.00	1.81	1.83
2000	1.88	2.11	2.06	2.03	2.01	2.05	2.11	2.10	2.11	1.97	2.01	1.92
2001	1.87	1.72	1.67	1.44	1.46	1.43	1.54	1.48	1.39	1.38	1.40	1.43
2002	1.46	1.51	1.54	1.49	1.39	1.45	1.48	1.50	1.56	1.55	1.52	1.47
2003	1.51	1.43	1.51	1.50	1.52	1.50	1.55	1.55	1.58	1.67	1.62	1.59
2004	1.63	1.65	1.59	1.54	1.52	1.51	1.53	1.57	1.64	1.51	1.45	1.51
2005	1.55	1.48	1.49	1.47	1.41	1.39	1.45	1.50	1.52	1.54	1.47	1.51
2006	1.68	2.35	1.96	1.96	2.07	2.13	2.15	2.09	1.87	1.74	1.72	1.71
2007	1.70	1.65	1.63	1.59	1.55	1.66	1.64	1.61	1.77	1.75	1.71	1.73
2008	2.14	2.42	2.10	2.20	2.14	2.31	2.41	2.60	2.57	2.23	1.73	1.77
2009	2.19	2.12	2.14	2.21	2.22	2.41	2.67	2.81	2.97	2.61	2.91	2.98
2010	2.85	2.94	2.80	2.60	2.33	2.24	2.32	2.51	2.46	2.49	2.52	2.69
2011	2.89	2.79	2.74	2.67	2.62	2.71	2.80	2.80	2.67	2.69	2.66	2.61
2012	2.63	2.64	2.73	2.78	2.80	2.88	2.99	3.08	3.04	2.88	3.04	3.08
2013	3.05	2.91	2.66	2.37	2.39	2.31	2.28	2.28	2.12	2.00	2.09	2.33
2014	2.56	2.22	2.09	2.09	1.96	1.91	2.10	2.03	1.89	1.91	1.90	1.88
2015	2.32	2.42	2.72	2.77	3.09	3.27	3.39	3.17	3.10	3.24	3.06	3.03
2016	2.72	2.48	2.30	2.11	2.21	2.41	2.43	2.31	2.33	2.39	2.68	2.67
2017	3.05	3.06	2.92	2.87	2.94	3.14	2.99	2.90	2.96	3.06	2.98	2.79
2018	2.96	2.99	2.81	2.69	2.61	2.55	2.49	2.40	2.41	2.42	2.36	2.30