FORECASTING RETAIL PRICES OF THE MOST COMMONLY USED RICE IN SRI LANKA

Warnakulasuriya Sudarshan Dushmantha Fernando

168854N

MSc in Financial Mathematics

Department of Mathematics

University of Moratuwa

Sri Lanka

November 2021

Declaration

I declare that this is my own work and this thesis does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Also, I hereby grant to University of Moratuwa the non-exclusive right to reproduce and distribute my thesis, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

…… ……

W. S. D. Fernando Date

The above candidate has carried out research for the Master of Financial Mathematics thesis/ dissertation under my supervision...

…… ……

Name of the Supervisor Date

Supervisor's signature: …….....

Acknowledgement

The timely and successful completion of the book could hardly be possible without the helps and supports from a lot of individuals.

I will take this opportunity to thank all of them who helped me either directly or indirectly during this important work.

First of all, I wish to express my sincere gratitude and due respect to my two supervisors Mr. T.M.J.A. Cooray. (B.Sc.), P.G.Diploma,(Mathematics), M.Sc.(Stat), M.Phil.(Stat) and Mr.Rohana Dissanayake, B.Sc. Mathematics(Colombo), M.Sc. (Pune). I am immensely grateful to them for their valuable guidance, continuous encouragements and positive supports which helped me a lot during the period of my work.

I would like to appreciate him for always showing keen interest in my queries and providing important suggestions. And I owe a lot to my family for their constant love and support. I also express whole hearted thanks to my friends and classmates for their care and moral supports.

Last but not the least I am also thankful to entire faculty staff and staff of department of mathematics for their unselfish help, I got whenever needed during the course of my work.

W. S. D. Fernando

Abstract

This thesis focuses on Modeling and forecasting Maximum Retail Price (MRP) of Samba, Nadu, Kekulu White and Kekulu Red rice in Sri Lanka using Univariate and Multivariate Time Series approaches. Sri Lanka is a developing country with population of 21.4 million as estimated in 2020. Rice is the most commonly used food in Sri Lanka. Thus, the finding a model for the forecasting prices is most economical advantage for Sri Lankan government. The fluctuations of the prices of rice making a great risk of investing, buffer stock maintaining, international trade and other associated actions. Thus, it is vital to forecast future prices for decision making purposes.

Our objective is to forecast the average weekly prices of selected four products. In this study, we consider weekly average retail prices of Samba, Nadu, Kekulu White and Kekulu Red from September 2017 to March 2019. Thus, each series consists of 93 data points. The missing values are estimated using expectation maximization algorithm. Data is collecting from Central Bank of Sri Lanka. First 83 data points are used to build the model and remaining 10 data points are used to validate the forecasting model. To select the best model, selection criteria based on the Akaike information criterion (AIC).

We observe that the best model for the Samba prices is exponential smoothing. Nadu price is ARIMA (2,1,0) and best model for Kekulu White and Kekulu Red are ARIMA (1,1,0) and ARIMA (1,1,0) respectively. Then, the testing data set is used to validate the prediction. Since there is a strong correlation between prices, we consider vector auto regression (VAR) model to improve the forecasts. Among several plausible models VAR of order 2 results in the best model. Nadu prices is independent of other three prices. VAR models provides better forecasts for prices of Nadu, Kekulu White, Kekulu Red.

Keywords: ADF, KPSS, PP, ARIMA, VAR, Cross correlation

TABLE OF CONTENTS

List of Figures Page

List of Tables

Page

List of Abbreviations

ACF	Autocorrelation function
ADF	Augmented Dickey-Fuller
AIC	Akaike information criterion
AR	Autoregressive
ARIMA	Autoregressive integrated moving average
BIC	Bayesian information criterion
CBSL	Central bank of Sri Lanka
HQIC	Hannan–Quinn information criterion
KPSS	Kwiatkowski–Phillips–Schmidt–Shin
MA	Moving average
MAE	Mean absolute error
MAPE	Mean absolute percentage error
PACF	Partial autocorrelation function
PP	Phillips Perron
RMSE	Root Mean Squared Error
VAR	Vector Auto Regression
VECM	Vector Error Correction Model

CHAPTER 01

INTRODUCTION

1.1 Background of the study

Sri Lanka is a developing country with an area of 65,610 km^2 and population of 21.8 million as estimated in 2020, [1]. Backbone of the Sri Lankan economy is agriculture and one-thirds of the population being dependent on agriculture. Rice is the main cultivating crop in Sri Lanka, which is cultivated in 1.1 million ha, contribution of 40% of the total cultivated area, which has been directed to self-sufficient country for rice. [3]

Rice has been one of the maple foods and the predominant dietary energy source over the world population including in Sri Lanka, since the written civilized ancient history.

The portion of 100g of rice contains 68.44g of water,28.17g of carbohydrates, 2.69g of protein as major components [2]. Rice produces 544kJ of energy per 100g portion and it contributes to 45% of total calorie and 40% of total protein requirement of an average Sri Lankan. [5]

Rice staple to the 21.8 million inhabitants in Sri Lanka where people used to consume minimum two rice-based meals per day. 4.5 million metric tons of raw rice currently produces by Sri Lanka to fulfil Sri Lanka's domestic requirement. The milled production output of raw rice is 3.09 million metric tons in 2019. The annuum consumption of rice is average of 3.0 million metric tons. Per capita rice consumption of single Sri Lankan is 107kg per year. [3] This figure varies yearly depending on the price of rice, price of bread and price of wheat flour. [4]

On average of 60% – 65% of rice production received during Maha (March/April) and 35% – 40% of production received during Yala (August / September) [3]. Eastern and North east provinces are the main rice intensive areas of the country.

In 2019, its contribution to GDP growth was 2.3% in, while agricultural contribution to the GDP 0.3%, in 2019 [1] The major consuming rice varieties in Sri Lanka are Samba, Kekulu - Red, Kekulu - White and Nadu. Department of agrarian services, Department of agriculture, Ministry of agriculture are the responsible units of rice cultivation, quality improvements, research and development and industrial aspects. The government of Sri Lanka supports though land provision, irrigation water supplies, fertilizer supply and fixed pricing.

Price of rice 1kg is being formulated and fixed by the board established for fixing raw rice price by considering the certified price for raw rice. The average rice price change percentage of all major varieties varies yearly as follows [2].

Rice Variety	Price change% 2018/2019	Price change% 2018/2019
Samba	9.4	-0.5
Kekulu – Red	-6.9	-1.2
Kekulu – White	-4.3	3.2
Nadu	-2.2	-0.2

Table 1.1: Average Rice Price Change Percentage.

The rice price does not maintain in the fixed price according to availability of rice in market which is governed by privet sector. [3]

Now a day's rice has been most worth fully crop in Sri Lanka. Then I would think to find useful forecasting model for above four variety of rice products. I think these models are most helpful for Sri Lanka government for their decision-making process to be optimizing. In most preference of the domestic production, I have selected to find useful forecasting models for Samba, Kekulu – Red, Kekulu – White and Nadu.

In past, there were some research about Rice production and Rice prices with various statistical instrument and different forecasting method such as regression analysis, exponential smoothing, and Box-Jenkins. The prediction of rice prices become a gradually more important tool for following purposes.

I. To develop appropriate prices scale
II. To decide future investments and decisions

1.2 Research Problem

My research problem is modelling and forecasting retail price of rice using Univariate and Multivariate Time Series approaches.

The rice is the most important crop in Sri Lankan people. The milled production output of raw rice is 3.09 million metric tons in 2019. The annum consumption of rice is average of 3.0 million metric tons [3]. Therefore, we can see the demand is quite lower than the supply and it drives the retail price of rice varies time to time. As well as MRP price of rice is most effective scenario for the Sri Lankan government in today. Thus, the finding a model for the forecasting prices is most economical advantage for Sri Lankan government.

This study answers the following situations

- Find appropriate Forecasting model
- Find independency of selected rice varieties.

1.3 Objective of the study

The objective of study are as follows

- To estimate appropriate ARIMA or exponential smoothing model for the selected four variables.
- To estimate VAR model for the selected four variables.
- Evaluate the forecasting performance of selected models.

1.4 Significance of the study

Content and analysis of this study is especially deal with retail prices of 4 rice varieties. This study emphases on Box-Jenkins models and Vector autoregressive model to prediction retail prices of rice. The study used 193 observations which are average weekly retail prices of Samba, Kekulu -White, Kekulu - Red and Nadu.

Find a suitable model for the forecasting process is the main significance work in this study. Also, we can find the interrelationship of each variable. It helps to determine prices movements are spread instantaneously other domestic markets. This study is also adding the facts to the understanding of the interrelationship among these four rice varieties and their returns on the GDP of the Sri Lankan economy.

1.5 Thesis Organization

This thesis is organized in fix chapters.

Chapter 1 discuss the background of the research area. It begins with background of study, research problem, objective and significance of the study.

Chapter 2 discuss review of related literature review for the study and their materials and methods and what are the significant method used in previous research as well as discussion of the results from previous studies.

Chapter 3 discuss the mathematical and statistical method were used for this study.

Chapter 4 discuss analysis part of this study. Fit ARIMA and the Exponential smoothing methods for individual variables as well as fit the Vector autoregressive model.

Chapter 5 discuss the discussion of Chapter 4.

Chapter 6 deal with the summarized and concludes whole of the study.

CHAPTER 02

LITERATURE REVIEW

One of the major aspects of time series is forecasting and it's a technique for prediction of event involved with a time component. It is a process of analyzing time series data using statistical based methods.

Kwadwo Agyei Nyantakyi, B. L. Peiris, L. H. P. Gunaratne (Kwadwo Agyei Nyantakyi, 2015) [6] design models for forecasting the Sri Lanka export prices of tea, rubber and coconut prices. They were utilized Box-Jenkins methodology in order to determine the best time series models for individual variables. They studied the interdependency of the prices of tea, rubber, and coconut production in Sri Lanka using Vector Auto Regression (VAR) Analysis. They even used the $l$ag−l cross-correlation matrix (CCM) to assess the strength of the linear interrelationship among the assets and, they construct a VAR model according to the selection range of standards based on the AIC, SIC and HQ. They looked at the individual behavior of each asset's individual prices and then analyses the combined effects of the prices. They found the best model out of all computing models for the price tea, rubber and coconut was ARIMA (0,1,0), ARIMA (3,1,1) and ARIMA (0,1,3) respectively. They looked at asset cointegration to determine if there was a long-run equilibrium and found that there was only one cointegration equation. As a result, they used the Vector Error Correction model (VECM), for the estimation. They also noticed that none of the coefficients between any variable were equal to zero. At all lags, the coefficient estimates between tea and rubber, tea and coconut, and rubber and coconut were not the same as between rubber and tea

G. Ramakrishna & R. Vijaya Kumari design ARIMA model for forecasting of rice production in India by using SAS [7]. They were used Box-Jenkins methodology for the find the suitable time series model for individual variables. Their primary goal in fitting the ARIMA model was to identify the time series' stochastic process and properly forecast future values. They used autocorrelation functions (ACF), partial autocorrelation functions (PACF), on the basis of minimum AIC, BIC, significance of AR and MA parameters, the ARIMA (011) model was selected. Parameter estimates along with corresponding standard errors of fitted ARIMA (011) model were reported, exhibit a non-significant pattern, this model was considered as valid and can be considered for forecasting.

B D P Rangoda, L M Abeywickrama, and M T N Fernando (B D P Rangoda') [8] design models for forecasting the Sri Lanka market prices of coconut oil, coconut wholesale prices and desiccated coconut prices. In this paper, the main goal of study was to fit the model that could forecast coconut price and related goods in Sri Lanka. Major behavioral patterns were discovered using time series plots. First order differences were used if the data series was non-stationary. Further, If the data series were non-stationary after first differences, log-transformed series of first order differences were selected. They have

used six conventional time series models which are General decomposition method, Winter's method, Single exponential smoothing method, Double exponential smoothing method, Moving average method, and Auto regressive moving average (ARIMA) method in order to fit the models and used Mean Absolute Deviation (MAD), Mean Squared Deviation (MSD) as well as Mean Absolute Percentage Error (MAPE) to test the model capability and precision. As the results, ARIMA & exponential methods are better than other models to prediction the prices of coconut and coconut products in Sri Lanka

Gurudeo Anand Tularam1, Tareq Saeed1,2 (Gurudeo Anand Tularam1, 2016,6) [9] design models for forecasting the oil prices using time series approaches. Autoregressive intergrade moving average (ARIMA), Holt-Winters (HW) and exponential smoothing (ES) are the three type of univariate time series models that they discussed in order to find the suitable model and they used six different approaches as selection range of standard to assess the accuracy of forecasting of these models. This comparison should aid policymakers and industry marketing strategists in determining the best oil market forecasting approach. The three models were evaluated by using them to the time series of regular oil prices for West Texas Intermediate (WTI) crude. According to the comparison, the HW model outperformed the ES model for a forecasting with a 95% confident interval. On the other hand, the ARIMA (2, 1, 2) model produced good results, indicating that this sophisticated and strong model surpassed other basic yet flexible models in the oil market.

K. NAVEENA, SANTOSHA RATHOD, GARIMA SHUKLA AND K.J. YOGISH (K. NAVEENA, 2014, April190-193) [10] implement a model for the coconut production in India. This research aims to develop a model that can predict coconut production in India. The researchers utilized time series data covering the period 61 years, from 1951 to 2012. Based on the least root mean square error values, the optimal model was chosen. It was discovered that the ARIMA (1, 1, 1) model is a suitable model for forecasting production. This study will also determine the future production of coconut in the next years. Further, Farmers, coconut-based companies, Governments and policymakers will benefit from the findings of this study.

Pasapitch Chujai*, Nittaya Kerdprasop, and Kittisak Kerdprasop (Pasapitch Chujai*, 2013, March 13 - 15) [11] implement model for the time series analysis for the household electric consumption. Individual household electric power usage was the time series data in their investigation, which ran from December 2006 to November 2010. The ARIMA (Autoregressive Integrated Moving Average) and ARMA (Autoregressive Moving Average) models were used to analyze the data. The most appropriate forecasting methods and the most appropriate forecasting period were chosen based on the minimum value of AIC (Akaike Information Criterion) and RMSE (Root Mean Square Error) values, respectively. The ARIMA model was shown to be the best model for determining the most appropriate forecasting period in monthly and quarterly forecasting. On the other hand, the ARMA model was the best model for determining the optimal forecasting period in daily and weekly forecasting. Then they have computed the best forecasting period, which

revealed that the approach was best for short-term forecasting (28 days, 5 weeks, 6 months, and 2 quarters, respectively).

CHAPTER 03

MATERIALS AND METHODS

3.1 Data sources

Data required for the study were collected from daily price report of Central Bank of Sri Lanka (CBSL) [12]. According to the data collected from daily price report of Central Bank of Sri Lanka during time period year 2016 to year 2020, observed that the daily price of rice on Sundays, Saturdays and government holidays were not reported during the time period year 2016 to year 2020. Therefore, we considered weekly average retail rice prices for this study. Sri Lankan weekly average Retail prices of Samba, Kekulu-White, Kekulu-Red and Nadu from the week 32 of 2016 to the week 13 of 2020 were considered for the analysis.

3.2 Types of Time series modeling methods

Mainly there are 2 types of time series modeling such as Univariate time series modeling and Multivariate time series modeling. Multivariate technique is extended version of univariate case. [13] [14]

3.2.1 Univariate Time series

Univariate time series consist of single observation recorded sequentially thought time. As an example, “Samba rice retail prices in Sri Lanka”. Auto regressive process, moving average process are examples of the univariate time series modeling methods.

3.2.2 Multivariate Time series

If we want to find relationship of two or more correlated variables, then we can model multivariate model for get suitable forecasting. Multivariate model is vector type of univariate modeling method. Domains contain at least two correlated variables. Vector Auto regressive, Vector Moving average methods are examples of Multivariate time series modeling.

3.3 Graphs

3.3.1 Time series plot

Time series plot is the usual line graph, but it is great way to detect patterns and behaviors in data over time. A time series plot represent observations in Y axis and X axis must be time. Using time series plot we can determine seasonal or annual variations, Trends and more things comparing with another series or before after effect of process. In financial sectors time series plot use very often for compare two or more-time related objects.

3.3.2 Autocorrelation Function

The autocorrelation function can be used for identifying the suitable time series model if data are not random. Given measurements, $\boldsymbol{Y_1, Y_2, \ldots, Y_n}$ at time $\boldsymbol{X_1, X_2, \ldots, X_n}$, the lag $\boldsymbol{k}$ autocorrelation function is defined as

$$\boldsymbol{\gamma_k = \frac{\sum_{i=1}^{n-k}(Y_i - \bar{Y})(Y_{i+k} - \bar{Y})}{\sum_{i=1}^{n-k}(Y_i - \bar{Y})^2}}$$

Autocorrelation is a correlation coefficient. The correlation is between two values of the same variable at times $\boldsymbol{X_i}$ and $\boldsymbol{X_{i+k}}$, rather than the correlation between two different variables. Usually the initial (lag 1) autocorrelation is generally of importance when using autocorrelation to detect non-randomness. When the autocorrelation is used to select a suitable time series model, the autocorrelations are generally plotted for many lags.

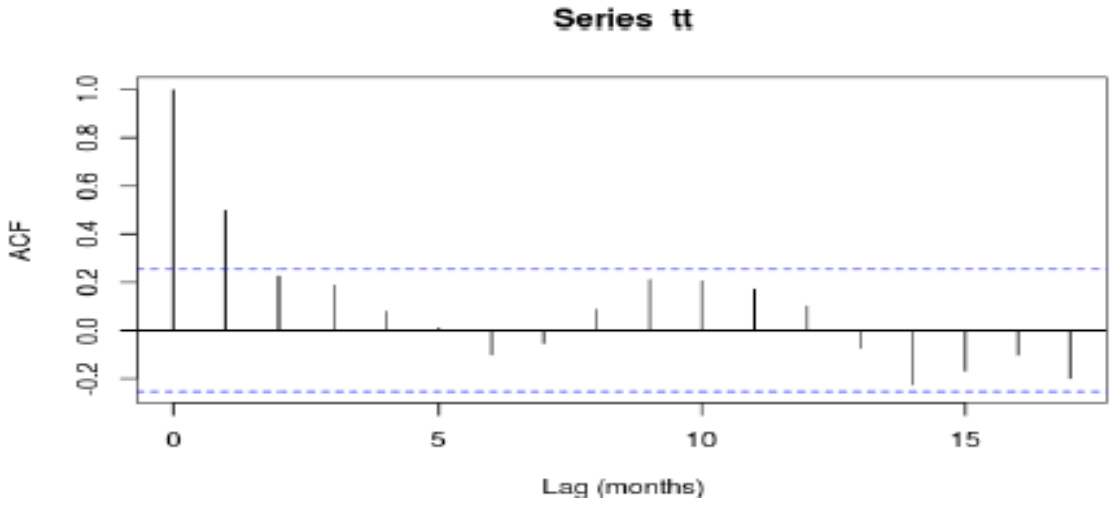

Figure 3.1: ACF graph

3.3.3 Partial Autocorrelation Function

In Box-Jenkins models, partial autocorrelation graphs are a frequent technique for model identification. The partial autocorrelation at lag k is the autocorrelation between $\boldsymbol{X_t}$ and $\boldsymbol{X_{t-k}}$ that is not accounted for by lags 1 through k−1. Specifically, Partial autocorrelations are particularly important for determining the order of an autoregressive model. At lag p+1 and higher, the partial autocorrelation of an AR(p) process is zero. If the sample autocorrelation plot suggests that an AR model is acceptable, the sample partial autocorrelation plot is evaluated to assist in determining the order. The point on the plot where the partial autocorrelations essentially become zero is what we're looking for. For

this purpose, placing a 95 % confidence interval for statistical significance is helpful and useful.

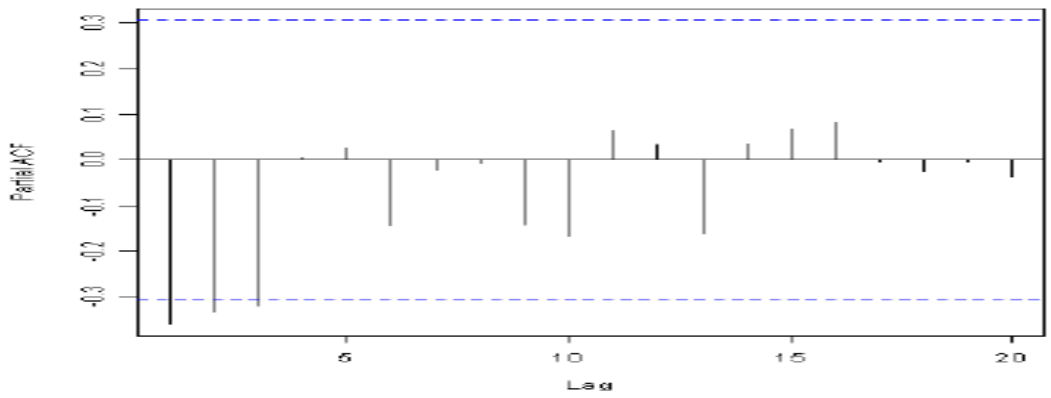

Figure 3.2: PACF graph

3.3.4 Quantile-Quantile plot

The quantile-quantile (q-q) plot is a graphical tool for assessing if two data sets are come from the populations with a common distribution. The benefit of the q-q plot is that sample sizes do not have to be equal. Many distributional characteristics can be evaluated at the same time. For instance, shifts in scale, shifts in location, changes in symmetry, and the existence of outliers can all be noticed from this plot. For instance. If the two data sets initiate from populations whose distributions change simply by a shift in location, the points should lie along a straight line that is displaced up or down from the 45-degree reference line.

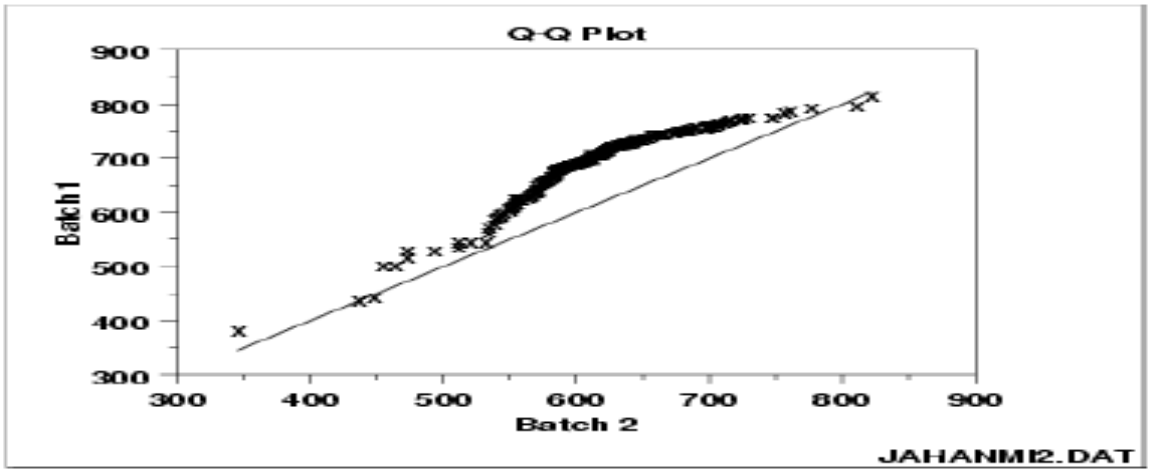

Figure 3.3: Q-Q plot

3.4 Model Selection Criteria

In applications are typically quite a few models for describing a population from a given sample of observations and one is thus challenged with the problem of model selection. Then the time series modelling one most important to select the suitable model in underlined data. Most common model selection criteria are the AIC and BIC. According to reviews most of researchers use AIC as the suitable selection criteria.

The **Akaike– information criterion** is computed as

$$\boldsymbol{AIC = -\frac{2l}{n} + \frac{2k}{n}}$$

Where l is the log likelihood.

3.5 Transform of Data

Data transformation is one on key thing in time series modeling. And most of time series are non-stationery and their variance are change over time. For literature review I have observed log transformation is the best transformation than other transformations for normalizing data and reduce heteroscedasticity. Therefore, all of this variable convert to their logarithmic transformation using following mathematical definition.

$$\boldsymbol{Y_t = log_{10}(X_t)}$$

$\boldsymbol{Y_t}$ is the log transformed version of $\boldsymbol{X_t}$

3.6 Stationery of time series data

A time series considered to be stationery if sample mean and sample variance and autocorrelation are not different over time period. A run sequence plot may generally be used to assess stationarity for practical reasons. If the time series is non stationary, we can usually convert it to stationarity using one of the approaches listed below.

i. Method of difference

There will be one point less in the differenced data than in the original data. Although the data might be differed several times, one difference is normally adequate.

ii. Taking the suitable technique such as logarithm or square root of the series may help to stabilize non-constant variance. Before applying the transformation on negative data, you can add an appropriate constant to make all the data positive. The anticipated (i.e.,

fitted) values and forecasts for future points may then be obtained by subtracting this constant from the model.

3.6.1 Unit root test for stationery

Statistical software provides various kind of tests for determine the existence of unit root such as **Augmented Dickey-Fuller** test (1979) and **Kwiatkowski-Phillips-Schmidt-Shin** test (1992) and **Phillips–Perron** tests. These tests are more helpful to determine there is a unit root (Non-stationery) or differencing method gives us stationery data.

3.6.1.1 Augmented Dickey-Fuller test

ADF test based on as follows

$$\boldsymbol{y_t = \alpha + \beta t + \gamma y_{t-1} + \varepsilon_t} \qquad\qquad \boldsymbol{\varepsilon_t \sim N(0, \sigma^2)}$$

The null hypothesis of ADF test is, $\boldsymbol{\gamma}$ **is unity,** which mean Data should be non-stationery Also, the alternative hypothesis is γ less than unity. Which mean data should be stationery.

3.6.1.2 Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test

The Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test tells that if a time series is stationary around a mean or linear trend or is non-stationary due to a unit root. A stationary time series is one where statistical property like the mean and variance are constant over time.

- The null hypothesis ($\mathbf{H_0}$) for the test is that the data is stationary.
- The alternate hypothesis ($\mathbf{H_1}$) for the test is that the data is not stationary.

3.6.1.3 Phillips–Perron (pp) test

PP test based on as follows

$$\boldsymbol{\Delta y_t = \rho y_{t-1} + u_t}$$

Δ is the first difference operator.

The null hypothesis is,

$\mathbf{H_0}$: $\boldsymbol{\rho = 1}$ which means data should be non-stationery.

Also, the alternative hypothesis gives the data should be stationery.

3.7 Forecast performance Measures

There are more than a few ways to determine the forecasting models performances. In this study, mean absolute error, root mean squared error and mean absolute percentage error were considered. Also, we were denoting actual value and forecasted value in period t as $\boldsymbol{Y_t}$ *and* $\boldsymbol{\widehat{Y}_t}$ respectively.

3.7.1 Mean Absolute Error (MAE)

The MAE measures the average magnitude of the error in the set of forecasts. And MAE is a linear score and that was calculated by following formula

$$\boldsymbol{MAE = \frac{\sum_{t=n+1}^{n+h} \left| \widehat{Y}_t - Y_t \right|}{h}}$$

3.7.2 Root Mean Squared Error

The RMSE is quadratic scoring rule which measure the average magnitude squared error. Calculating technique of the RMSE was the difference between forecast and actual values are squared and then get the average over the sample. RMSE frequently used to compare forecasts since the errors are squared before they are averaged.

The root mean squared error calculated by following formula

$$\boldsymbol{RMSE = \sqrt{\frac{\sum_{t=n+1}^{n+h} (\widehat{Y}_t - Y_t)^2}{h}}}$$

3.7.3 Mean Absolute Percentage Error

The mean absolute percentage error (MAPE) measures the size of the error in percentage term. Get the ratio of difference between forecast and corresponding observed values and the observed values of the all over the sample size.

MAPE are calculated by

$$\boldsymbol{MAPE = \frac{100 * \sum_{t=n+1}^{n+h} \left| \frac{(\widehat{Y}_t - Y_t)}{Y_t} \right|}{h}}$$

And also

MAPE	Judgment of Accuracy
< 10%	Highly Accurate
11% to 20%	Good Forecast
21% to 50%	Reasonable Forecast
> 51%	Inaccurate forecast

Table 3.1: Accuracy measurements criteria

3.7.4 Adjusted R squared

The proportion of total variation in Y explained by all X variable taking together (the model)

$$R^2 = \frac{SSR}{SST} = \frac{Regression\ sum\ of\ squares}{Total\ sum\ of\ squares}$$

Adjusted R squared is improvement of R squared

I. It values depend on number of explanatory variables

II. Imposes a penalty for adding additional explanatory variables.

$$Adjusted\ R^2 = \frac{k-1}{n-k}(1 - R^2)$$

If better model has highest Adjusted R squared values.

3.8 Box-Jenkins Methodology

The Box-Jenkins approach is better models than most of models because it does not assume any particular pattern in the historic data of the series to be forecast. This approach performs an iteratively for identifying a possible model. The chosen model is checked again historical data and see whether the accuracy of the series. If model fit well then residuals are generally small, randomly distributed. And specified model does not satisfy, the process is repeated using a new model designed to improve one the original one. The iterative procedure continuous until a useful model is found. At that model can use to forecasting.

3.8.1 Auto Regressive (AR) process

The current value of the series is linearly dependent upon its previous value with some error. Then we have linear relationship

$$\boldsymbol{Y_t = \emptyset_1 Y_{t-1} + \varepsilon_t}$$

Where $\boldsymbol{\varepsilon_t}$ is a white noise time series (this $\boldsymbol{\varepsilon_t}$ is a sequence of uncorrelated random variables (possibly normally distributed but not necessarily normal) with mean 0 and constant variance. This type of model called as AR model. Here foregoing model was an AR model with lag 1. Generally, we have an autoregressive (AR) model of order p.

AR(p) model was define by

$$\boldsymbol{Y_t = \sum_{i=1}^{p} \emptyset_i Y_{t-i} + \varepsilon_t}$$

3.8.2 Moving Average Process (MA)

This model is same as AR process, and the difference is value of time t is depending on weighted average of past values of the white noise series.

$$\boldsymbol{Y_t = \theta_1 \varepsilon_{t-1} + \varepsilon_t}$$

Where ε_t is a white noise process including zero mean and constant variance. Foregoing model is MA(1) process, because the lag is equal to 1. white noise term gives stochastically uncorrelated information which appear to each time step.

In generally MA (q) model was define as,

$$\boldsymbol{Y_t = \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}}$$

3.8.3 ARIMA model

The ARIMA is the short name of "Auto Regressive Integrated Moving Average" model. The function representing the ARIMA model is denoted ARIMA (p, d, q), which produces a stationary function ARMA (p, q) upon differentiation with respect to time. The ARMA model include the autoregressive AR model of order p. MA, the moving average of order q. This is the example of ARMA (p, q) model

$$Y_t = \delta + \sum_{i=1}^{p} \emptyset_i Y_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}$$

$$Y_t = \delta + \emptyset_1 Y_{t-1} + \emptyset_2 Y_{t-2} + \cdots \ldots + \emptyset_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots \ldots - \theta_q \varepsilon_{t-q}$$

$\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots$ are white noise terms and δ be the constant term and $\emptyset_1, \emptyset_2, \emptyset_3, \ldots$ and $\theta_1, \theta_2, \theta_3, \ldots..$ are autoregressive and moving average parameters respectively. Then ARIMA (p, d, q) is producing using ARMA process and d is the require number of differencing which data should be stationery. If d=0 we can get ARMA process. The linear difference operator **(Δ)**, is defined by,

$$\Delta Y_t = Y_t - Y_{t-1} = Y_t - BY_t = (1 - B)Y_t$$

The stationery series Y_t is obtained by $\mathbf{d^{th}}$ difference of Y_t

$$Y_t = \Delta^d Y_t = (1 - B)^d Y_t$$

ARIMA (p, d, q) general form,

$$\emptyset_p(B)Y_t = \delta + \theta_q(B)\varepsilon_t$$

Where B is back ward shift operator and important thing is ARIMA should be a non-seasonal time series model.

3.8.4 Seasonal ARIMA model

Time series data may sometimes exhibit strong periodic patterns. This is commonly referred to as a time series with seasonal behavior. This typically occurs when data is gathered in particular intervals-monthly, weekly, and so on. An additive model, in which the process is considered to be made of two components, is one approach to describe such data,

$$Y_t = S_t + N_t$$

Where S_t is seasonal component and N_t is the stochastic component. Generally seasonal ARIMA model is denoted by,

$$ARIMA(p, d, q) \times (P, D, Q)_{[s]}$$

Where P: Seasonal AR of order P

D: Number of differences required to achieving deseasonaling.

Q: Seasonal MA of order Q

Also, t is the periodic time.

3.8.5 Identification of Model

Box-Jenkins approach could not be used if data is non-stationery. important thing is check whether the data contain any seasonal pattern or trend. We can use ADF, KPSS, PP test or even look the pattern of ACF and PACF for check the data should in stationery. If a graph of ACF dies down extremely slowly, then time series values should be non-stationery. Hence input series needs to be stationery. That is input series has time depending mean, variance and autocorrelation.

If the series is not stationery, it can be transformed to a stationery by differencing. Differencing is done until, plot of data varies with fixed level or ACF graph dies down rapidly. The behavior of ACF and PACF helps to identify the which model describe most accuracy. Following table discuss the how to apply ACF and PACF to model identification.

Model	**ACF**	**PACF**
MA(q)	Cut off after lag **q**	Dies down
AR(p)	Dies down	Cut off after lag **p**
ARMA	Dies down	Dies down

Table 3.2: Accuracy measurements criteria

Identify model using ACF and PACF

The significance of ACF and PACF was determine by $\left[\frac{1.96}{\sqrt{n}}, -\frac{1.96}{\sqrt{n}}\right]$

3.8.6 Parameter Estimation

If we select the suitable model, next goal is estimating the parameters of appropriate model. The parameters in ARIMA models are estimated by minimizing sum of squares of forecasting error. Generally, parameters are estimated by nonlinear least square procedure. This procedure performs an algorithm that find the model which contain minimum error squared. If least square estimates and their standard errors are determined, then we can interpret it usual way. The parameters are significantly different from zero, then estimates retained the model. If parameters are not significant then they were drop in our model. If executed model has significant parameters and we want to move next step such as diagnostic checking. In practically we use to minimum AIC value for the get most powerful model than others.

3.8.7 Diagnostic Checking

If after parameter estimation and before forecasting, we must want to check whether model adequacy for the forecasting. Basically, a model is adequate residual cannot be affecting to forecast and residuals should be random. The data fitting more than one ARIMA models then estimate the parameters of each model and then perform the diagnostic checking. Then selected model could be more useful to forecasting. In following test are used frequently.

3.8.7.1 Ljung-Box test for residuals

Using residual autocorrelation plot we can see that clearly, if there exist relationship between residuals and fitted ARIMA model. The Ljung-Box test is based on that autocorrelation plot. This test in commonly used in ARIMA modelling.

Generally, the hypothesis are as follows

$\boldsymbol{H_0}$: There are no autocorrelation between residuals

$\boldsymbol{H_1}$: There are autocorrelation between residuals.

Test statistic of Ljung-Box test is

$$\boldsymbol{Q = n(n+2)\sum_{i=1}^{m} \frac{{\gamma_j}^2}{n-j}}$$

Where n is the sample size, $\boldsymbol{{\gamma_j}^2}$ is the autocorrelation at lag j, and m is number of lags being tested at corresponding significance level.

The hypothesis can be rejected if

$$\boldsymbol{Q > {x^2}_{1-\alpha,m-p}}$$

Here $\boldsymbol{x^2}$ is the chi squared distribution and p is the number of estimated parameters.

3.8.7.2 Ljung-Box test for squared residuals

When we use Ljung-box test for residual to check whether homoscedasticity of residuals. It is more helpful determine residuals have constant variance. Generally, Hypothesis of these test as follows,

$\boldsymbol{H_0}$: There is homoscedasticity between residuals

$\boldsymbol{H_1}$: there isn't homoscedasticity between residuals.

3.8.7.3 Jarque-Bera test

Jarque-Bera test were used to check whether the residuals are normally distributed or not. Also, we can use Quantile-Quantile plot to check whether the residuals follow a normal distribution. Test statistic of Jarque-Bera test is

$$\chi_{n-k}{}^2 = \frac{n}{6}\left\{(S)^2 + \frac{(K-3)^2}{4}\right\}$$

Corresponding hypotheses are

H_0: Residuals are normally distributed

H_1: residuals are not normally distributed.

3.9 Exponential Smoothing

Exponential smoothing linear method like as regression model. Typically, Exponential Smoothing is useful when the data pattern is almost horizontal, especially when no particular trend or seasonal variation existed in previous data sets. Exponential Smoothing is useful to short time forecasting process. [15], [16]

3.9.1 Model selection

Basically, we have

No trend + no seasonal series model → Single Exponential smoothing

Linear Trend+ no seasonal smoothing → Trend corrected exponential

Seasonal + trend → Holt Winters method.

3.9.1.1 Single exponential smoothing

More recent observations are given higher weights in exponential smoothing methods, and the weights fall exponentially as the observations move further apart. When the parameters representing the time, series change slowly over time, these approaches are most successful. When there is no trend or seasonal pattern, but the mean (or level) of the time series Y_t is steadily changing over time, the Simple Exponential Smoothing approach is used to forecast it. Generally, model defined as

$\boldsymbol{Y_t = \beta_0 + \varepsilon_t}$ Where $\boldsymbol{\varepsilon_t}$ is a normal distributed with zero mean and constant variance

3.9.1.2 Trend Estimated Exponential smoothing

Trend estimated exponential is another approach of time series modelling. The general equation as follows

$$\boldsymbol{Y_t = \beta_0 + \beta_1 t + \varepsilon_t}$$

the values of the parameters $\boldsymbol{\beta_0}$ and $\boldsymbol{\beta_1}$ are slowly changing over time, Holt's trend corrected exponential smoothing method can be applied to the time series observations.

3.9.2 Parameter estimation

Step 1 of single exponential smoothing method compute initial estimate of the mean (or level) of the series at time period t=0.And computed the updated estimates using the exponential equation. This equation can be used to get forecasting values.

$$\boldsymbol{l_0 = \overline{Y}}$$

$$\boldsymbol{l_t = \alpha Y_t + (1 - \alpha) l_{t-1}}$$

Where $\boldsymbol{\alpha}$ is smoothing constant between 0 and 1. $\boldsymbol{l_0}$ is the initial value and $\boldsymbol{l_t}$ denote the new value of time t.

Finally, the point forecast of single exponential smoothing model is given by

$$\boldsymbol{\widehat{Y}_{t+p} = l_t}$$

If we have a trend component, we use trend corrective exponential smoothing model

the level estimate is given by

$$\boldsymbol{l_t = \alpha Y_t + (1 - \alpha)(l_{t-1} + b_{t-1})}$$

And trend estimate

$$\boldsymbol{b_t = \gamma(l_t - l_{t-1}) + (1 - \gamma) b_{t-1}}$$

Where $\boldsymbol{\gamma}$ is the smoothing constant of trend and $0 < \boldsymbol{\gamma} < 1$.

Finally point forecast of this model is given by

$$\boldsymbol{\widehat{Y}_{t+p} = l + p b_t}$$

3.9.3 Diagnostic checking

Diagnostic test is most important session in forecasting process. We want to check whether the estimated exponential smoothing, model adequacy of forecasting. Following method are gives useful result of diagnostic checking.

3.9.3.1 Histogram of residuals and Quantile-Quantile plot

Histogram of residuals are one of checking method for normality of residuals. Reject the normality if the histogram departs dramatically from bell shape. And also, quantile-quantile plot gives the suitable determination of normality residuals. Following type of histogram and quantile-quantile plot shown as normality existence of residuals. Also, Box-test and Jarque-Bera test could be used to evaluate the other diagnostics.

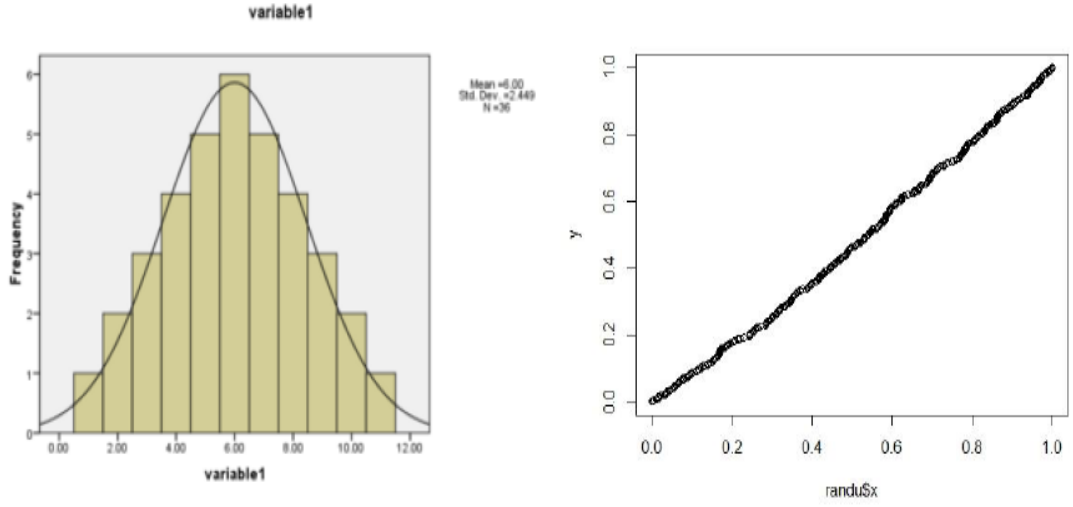

Figure 3.4: Histogram of residuals and Q-Q plot

3.10 Vector Auto Regressive model (VAR)

Since the seminal paper of Sims (1980) vector autoregressive models have become a key instrument in macroeconomic research. Vector auto regressive stochastic and multivariate time series model. Mostly it uses to check independency or dependency between correlated variables. Vector auto regressive model generalize the univariate autoregressive model by allocating more than one variable. And the VAR model has proven to be especially useful for describing the dynamic behavior of economic and financial time series and for forecasting. At this point the VAR approach comes in. A simple VAR model with lag 2 can be written as

$$\begin{pmatrix} \boldsymbol{Y_{1t}} \\ \boldsymbol{y_{2t}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{a_{11}} & \boldsymbol{a_{12}} \\ \boldsymbol{a_{21}} & \boldsymbol{a_{22}} \end{pmatrix} \begin{pmatrix} \boldsymbol{Y_{1t-1}} \\ \boldsymbol{y_{2t-1}} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\varepsilon_{1t}} \\ \boldsymbol{\varepsilon_{2t}} \end{pmatrix}$$

Basically, such a model implies that everything depends on everything. But as can be seen from this formulation, each row can be written as a separate equation, so that

$\boldsymbol{Y_{1t} = a_{11}Y_{1t-1} + a_{12}y_{2t-1} + \varepsilon_{1t}}$ and $\boldsymbol{y_{2t} = a_{21}Y_{1t-1} + a_{22}y_{2t-1} + \varepsilon_{2t}}$.

Where $\begin{pmatrix} \boldsymbol{a_{11}} & \boldsymbol{a_{12}} \\ \boldsymbol{a_{21}} & \boldsymbol{a_{22}} \end{pmatrix}$ are coefficient matrix of VAR model and $\begin{pmatrix} \boldsymbol{\varepsilon_{1t}} \\ \boldsymbol{\varepsilon_{2t}} \end{pmatrix}$ are multivariate normally distributed with 0 mean and covariance matrix ϵ.

3.10.1 Cointegration test

Nonstationary time series in $\boldsymbol{Y_t}$ are cointegrated if there is a linear combination of them that is stationary or $\boldsymbol{I(0)}$. If some elements of $\boldsymbol{\beta}$ are equal to zero, then only the subset of the time series in $\boldsymbol{Y_t}$ with non-zero coefficients are cointegrated. The linear combination $\boldsymbol{\beta_0 Y_t}$ is often motivated by economic theory and referred to as a long-run equilibrium relationship. The Johansson cointegration test was used for the check existence of cointegrating vectors. Hypothesis are as follows. If there is no cointegrating vectors, then we move VAR process and if there is significant cointegration then we want to move vector error correction model VECM.

3.10.2 Model Identification

Model identification is most of the important thing. Because we have different type of VAR model with different lag. Then the lag selection is the only one part of model selection procedure. Lag means, how many require number past terms we want to select for continue VAR process.

3.10.2.1 Lag Length Selection

The general approach is to fit VAR(p) models with orders p = 0, ..., p (max) and choose the value of p which minimizes some model selection criteria. Model selection criteria for VAR(p) models, the three most common information criteria are the Akaike (AIC), Schwarz-Bayesian (BIC) and Hannan-Quinn (HQ),

$$\boldsymbol{AIC(p) = \ln\left|\sum(p)\right| + \frac{2}{T}(number\ of\ freely\ estimated\ parameters)}$$

$$\boldsymbol{BIC(p) = \ln\left|\sum(p)\right| + \frac{\ln T}{T}(number\ of\ freely\ estimated\ parameters)}$$

$$\boldsymbol{HQ(p) = \ln\left|\sum(p)\right| + \frac{2\ln\ln T}{T}(number\ of\ freely\ estimated\ parameters)}$$

Where $\boldsymbol{\sum(p) = T^{-1}\sum_{t=1}^{T}\varepsilon_t\acute{\varepsilon}_t}$ and **T** be the sample size and $\boldsymbol{p}$ is the lag size.

3.10.3 Parameter Estimation

After selecting suitable lag then we want to estimate parameters of Vector auto regressive model. Generally, we assume the VAR model as

$$\boldsymbol{Y = AX + U}$$

Where $\boldsymbol{A}$ is coefficient matrix and $\boldsymbol{U}$ is error matrix.

We use multivariate least square estimation procedure and then we can estimate $\boldsymbol{A}$ as follows

$$\widehat{\boldsymbol{A}} = \boldsymbol{YX}^{/} + (\boldsymbol{XX}^{/})^{-\mathbf{1}}$$

As the explanatory variables are the same in each equation, the multivariate least squares estimator is equivalent to the ordinary least square estimator applied to each equation separately. Usually significant coefficients are considered.

3.10.4 Diagnostic checking

After parameter estimation, our goal is to check our estimated model adequacy of forecasting. There are several things to check for the VAR models.

3.10.4.1 Portmanteau Serial Test

This test originally developed by Box and Pierce for testing the residuals from a forecast mode. Any good forecast model should have forecast error which follows a white noise model.

The univariate Ljung-Box statistic $\boldsymbol{Q_m}$ has been generalized to the multivariate case by Hosking (1980) and Li and McLeod (1981) for a multivariate series. And the test would be visualized there are no auto correlation or cross correlation among the vector series $\boldsymbol{r_l}$

Hypothesis of portmanteau test are

$\boldsymbol{H_0}$: no autocorrelation of residuals

$\boldsymbol{H_1}$: significant autocorrelation of residuals.

Test statistic is $\boldsymbol{Q_k(m) = n\sum_{k=1}^{h} r_k{}^2}$

And $\boldsymbol{Q_k(m)}$ follows chi-squared distribution with where n is the sample size k is dimension of $\boldsymbol{r_k}$ and $\boldsymbol{r_k}$ is the residual auto correlation of lag k

3.10.4.2 Jarque-Bera test

As a univariate case, we use Jarque-Bera test for multivariate case. Corresponding hypotheses are

$\boldsymbol{H_0}$: Residuals are normally distributed

$\boldsymbol{H_1}$: residuals are not normally distributed.

3.10.4.3 Granger Causality

The Granger causality test is a statistical test used to assess if one time series may be used to foresee another. A time series X is said to Granger-cause Y if it can be demonstrated, often through a series of t-test and F-test on lagged values of X (and with lagged values of Y included), that those X values provide statistically significant information about future values of Y.

Corresponding hypotheses are

$\boldsymbol{H_0}$: X does not granger causal to Y.

$\boldsymbol{H_1}$: X granger causal to Y.

3.10.4.4 Instantaneous Causality

The formulation of Granger causality included no mention of the possibility of an instantaneous correlation between X and Y. We call this instantaneous causality if the innovation to Y and the innovation to X are correlated. The instantaneous causality test is useful for determining whether or not there is causality in a short time period.

Corresponding hypotheses are

$\boldsymbol{H_0}$: X does not instantaneous causal to Y.

$\boldsymbol{H_1}$: X instantaneous causal to Y.

3.10.4.5 Multivariate ARCH Test

ARCH test is used to test whether there exist ARCH effect or not. ARCH mean Autoregressive conditional heteroscedasticity. Corresponding hypothesis are

$\boldsymbol{H_0}$: There is no ARCH effect

$\boldsymbol{H_1}$: There is significant ARCH effect.

Test statistic is, $$\boldsymbol{\varepsilon_t^2 = \beta_0 + \sum_{i=1}^{q} \beta_i \varepsilon_{t-i}^2 + \delta_t}$$

Where ε is residual. This is a regression of the squared residuals on a constant and lagged up to order q.

CHAPTER 04

RESULTS

Introduction

This 4th chapter present the analysis of the retail prices of rice in Sri Lankan government. It consists of description on analysis of time series data using ARIMA, Exponential Smoothing and Vector Autoregressive process.

The R statistical software, SPSS and the Minitab were used for the analysis. Various types of models have been discovered in this situation and, select the suitable model using diagnostic checking and evaluating forecasting error.

4.1 Data Presentation

The retail prices were coming from daily price report of Central bank Sri Lanka. The data was available on average weekly retail prices of Samba, Nadu, Kekulu White and Kekulu Red including 3-year period (20017-2019). The Samba, Nadu, Kekulu White and Kekulu Red data were measured by rupees per one kilogram.

4.2 Data Aggregation

Each data was divided into two sets, 75% of them were used to formulate appropriate model and other 25% for forecasting evaluation process. These two sets are called training set and testing set. Testing set was used to get accuracy measurements of formulated model.

4.3 Missing Value Estimation of retail rice prices

Rice has been one of the maple foods and the predominant dietary energy source in Sri Lanka. There are four type of rice mainly consume in Sri Lanka such as Samba, Nadu, Kekulu Red and Kekulu White. Thus, the retail price of all of these categories are most important things for economics in Sri Lanka.

The data set has missing values for each category. To continue the analysis, estimating missing values are very important. Imputation is the process of replacing missing values with substituted values. Because missing values can create issues for analyzing data, imputation is seen as a way to avoid pitfalls involved with leastwise deletion of cases that have missing values.

We can use several methods to estimate missing values such as mean imputation, median imputation, mode imputation etc. Expectation-Maximization (EM) algorithm is a suitable method to use to find the missing values. Therefore, the estimation of missing values of each categories of data set is conducted by SPSS software.

4.3.1 Output of SPSS for Missing Value Estimation of retail rice prices

4.3.1.1 Missing value Analysis (MVA)

Run the EM algorithm for the data sets via SPSS in order to find the missing values.

EM(TOLERANCE=0.001 CONVERGENCE=0.0001 ITERATIONS=25

	N	Mean	Std. Deviation	Missing		No. of Extremes	
				Count	Percent	Low	High
Samba	84	106.57	5.071	9	9.7	8	0
Nadu	85	88.51	5.965	8	8.6	3	14
Kekulu White	87	85.76	6.798	6	6.5	0	0
Kekulu Red	87	80.63	5.225	6	6.5	0	0
a. Number of cases outside the range (Q1 - 1.5*IQR, Q3 + 1.5*IQR).							

Table 4.1: Univariate Statistics of missing values estimation

Here we can see that almost 10% of values are missing data on the variable of Samba. And other percentage of Nadu, Kekulu White, Kekulu Red are 8.6%, 6.5% and 6.5% respectively. Also, we can see that, the mean value of Samba is shown high value compared with other variables.

Below, Table 4.2 and Table 4.3 show the results of SPSS's EM procedure.

Summary of Estimated Means				
	Samba	Nadu	Kekulu White	Kekulu Red
All Values	106.57	88.51	85.76	80.63
EM	106.64	88.65	85.82	80.62

Table 4.2: Summary of Estimated Means of missing values estimation

Summary of Estimated Standard Deviations				
	Samba	Nadu	Kekulu White	Kekulu Red
All Values	5.071	5.965	6.798	5.225
EM	5.074	6.103	6.941	5.321

Table 4.3: Summary of Estimated Standard Deviation of missing values estimation

The algorithm leads to an estimated mean of 106.64 and standard deviation of 5.074 for Samba, not much different from the observed means for those prices on which we do have data. And also, we can see that, other three variables are not much different from the observed means.

4.3.1.2 EM Estimated Statistics

Here we have estimated means, covariance and correlation coefficients Table 4.4, 4.5, 4.6 respectively. Little's MCAR test the null hypothesis that the missing data are Missing Completely at Random. According to the Significant value 0.66, we conclude that the data are missing completely at random.

EM Means			
Samba	Nadu	Kekulu White	Kekulu Red
106.64	88.65	85.82	80.62
a. Little's MCAR test: Chi-Square = 11.327, DF = 14, Sig. = 0.660			

Table 4.4: EM Means of missing values estimation

EM Covariance				
	Samba	Nadu	Kekulu White	Kekulu Red
Samba	25.745			
Nadu	5.576	37.251		
Kekulu White	9.572	38.950	48.182	
Kekulu Red	8.388	25.309	32.433	28.318
a. Little's MCAR test: Chi-Square = 11.327, DF = 14, Sig. = 0.660				

Table 4.5: EM Covariance of missing values estimation

EM Correlations				
	Samba	Nadu	Kekulu White	Kekulu Red
Samba	1			
Nadu	0.180	1		
Kekulu White	0.272	0.919	1	
Kekulu Red	0.311	0.779	0.878	1
a. Little's MCAR test: Chi-Square = 11.327, DF = 14, Sig. = 0.660				

Table 4.6: EM Correlation of missing values estimation

Here we can see that EM correlation between Samba and other three variables are having a week correlation since the values are close to zero. The correlation between Nadu and Kekulu White, Nadu and Kekulu Red shows strong correlation as well as the correlation between Kekulu White and Kekulu Red.

Next step is to find suitable model for above four variable using completed data sets.

4.4 Retail Price of Samba

Including missing values, we were fit a model for forecast retail price of Samba. Fitted model contain 83 observations.

4.4.1 Samba retail prices forecast using ARIMA model.

4.4.1.1 Time Series Plot of Samba price

Figure 4.1: Time series plot of retail price of Samba

According to figure 4.1 we can see that retail prices of Samba vary with 98 and 110 LKR. We cannot clearly see that there is a trend. Also, cannot clearly say that there are have any seasonality pattern. But there is some constant value during small time periods throughout the years from 2017 to 2019.

4.4.1.2 Time series plot of Transformed data for the price of Samba

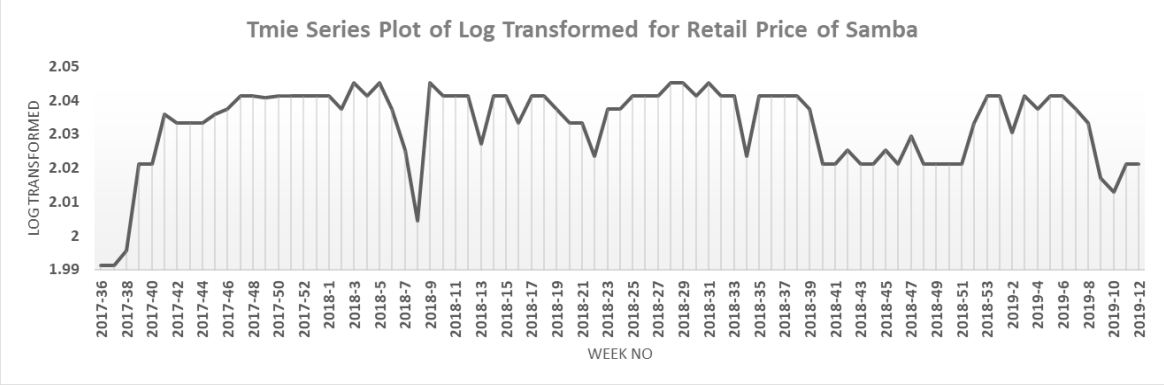

Figure 4.2: Time series plot of Transformed data for price of Samba

And most of time series are non-stationery and their variance are change over time. For literature review I have observed log transformation is the best transformation than other transformations for normalizing data and reduce heteroscedasticity. Figure 4.2 shows prices vary between 1.99 and 2.04. This implies transform should reduce heteroscedasticity.

4.4.1.3 Stationery Test

First of all, we want to check the original time series observations are stationery or not.

ADF.test (Samba Log Transformed)		
Dickey-Fuller = -3.9352	Lag order = 4	p – value = 0.01668
alternative hypothesis: stationary		

Table 4.7: Results of ADF test of price of Samba log transformed level data

According to ADF test of weekly average prices of Samba P-value is 0.01668 which is less than 0.05. Thus, the null hypothesis of non-stationery being rejected at 5% level of significance.

KPSS test also used to test stationery of prices of Samba

KPSS.test (Samba Log Transformed)		
KPSS Level = 0.19078	Truncation lag parameter = 3	p – value = 0.1
Warning message: In kpss.test(SambaLog) : p-value greater than printed p-value		

Table 4.8: Results of KPSS test of price of Samba log transformed level data

According to KPSS test of weekly average prices of Samba P value is 0. 1 which is greater than 0.05. Thus, the null hypothesis of stationery being not rejected at 5% level of significance.

PP test also used to test stationery of prices of Samba

PP.test (Samba Log Transformed)		
Dickey-Fuller Z(alpha) = -23.065	Truncation lag parameter = 3	p – value = 0.02592
alternative hypothesis: stationary		

Table 4.9: Results of PP test of price of Samba log transformed level data

According to PP test of weekly average prices of Samba P value is 0.02592 which is less than 0.05. Thus, the null hypothesis of non-stationery being rejected at 5% level of significance.

According to these 3 tests we can conclude that original weekly average prices of Samba are given stationery observations. Thus, we were not used differencing to achieve to stationery time series again. Thus, we were continued model selection procedure.

4.4.1.4 Model Identification

According to stationery results log transformed original observations of price Samba are stationery. Since we can find a suitable time series model after estimating significant lags.

Now we know that, original time series data is given Stationary observation and there is no need to go for the differencing method that implies d=0. Therefore, we can go for the Model ARMA (p, q) or ARIMA(p,0,q) .

4.4.1.4.1 ACF and PACF of transformed data

The ACF and PACF graph of original transformed data are computed as in Figure 4.3 and Figure 4.4 Then, the patterns of ACF & PACF were used to determine the parameter values of p and q.

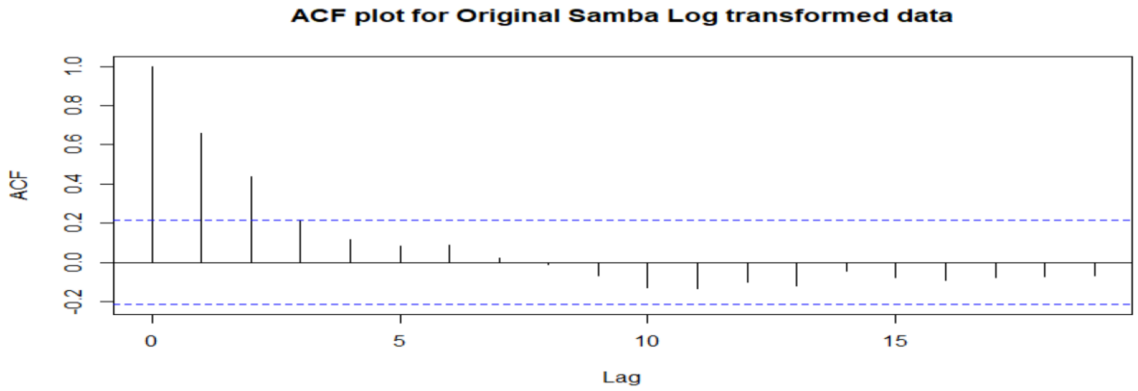

Figure 4.3: ACF plot of original Samba log transformed data

According to the Figure 4.3 lags decreasing. We cannot see there are significant spikes during the lags, thus MA (0) is the suitable parameters and q must be zero.

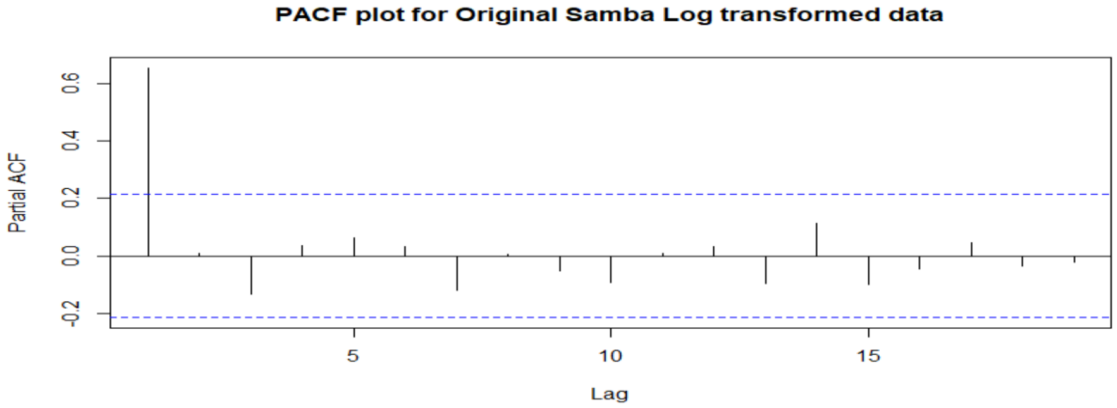

Figure 4.4: PACF plot of original Samba log transformed data

Figure 4.3 and 4.4 illustrate ACF and PACF for given stationary time series data. The ACF shows a gradually decreasing trend while the PACF cuts immediately after one lag. Thus AR (1) is the suitable parameters and p must be 1.

Therefore, the graphs suggest that an AR (1) model would be appropriate for the time series.

4.4.1.5 Model selection

After identifying the feasible values of p and q, the next step is an estimate the parameters under the most suitable model. We have an appropriate model for Samba price which is AR (1) or ARIMA (1,0,0). The parameter estimation of model is conducted by R software.

Thus ARIMA (1,0,0) is selected as suitable model for forecasting the weekly average retail prices of Samba

4.4.1.6 Parameter Estimation

ARIMA (Samba Log Transformed, order = c (1,0,0))		
Coefficients		
	ar1	intercept
	0.7679	2.0309
s.e.	0.0808	0.0039
t ratio	9.506547	521.978574
p value	1.970969e-21	0.000000e+00
σ^2 estimated as 0.0000697	Log likelihood = 278.96	AIC = -551.91

Table 4.10: Parameter estimation of ARIMA (1,0,0)

To evaluate if the relationship between the response and each term in the model is statistically significant, we may compare the p-value for the term to the level of significance to test the null hypothesis.

H_0: The term is not significantly different

H_1: The term is significantly different

According to the Table 4-10, the p-values of term is less than 0.05. Thus, we can conclude that the coefficient is statistically significant. So, we can fit the model with the term.

The model equation can be formulated as

$$y_t = 0.7679y_{t-1} + \varepsilon_t$$

Where y_t is the weekly average retail prices of Samba and ε_t is the error terms.

4.4.1.7 Diagnostic checking

After estimating the parameter for ARIMA (1,0,0) model, next step will be the diagnostic checking.

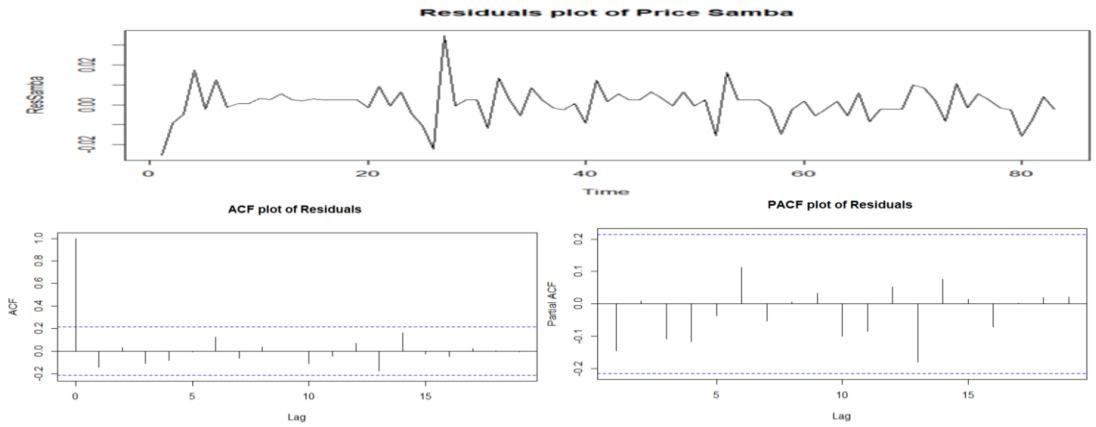

Figure 4.5: ACF and PACF plot of residuals and residual plot of ARIMA (1,0,0)

According to figure 4.5. Top of the box contain the time plot of standardized residuals. It shows that no pattern and looks like an independent identical distributed. And residual ACF and PACF shows all of lags are not significant. Since the residuals haven't significance autocorrelation.

Most important property of residual diagnostic is the residuals are uncorrelated. It means residuals are independent each other. To detect the correlation, we can study residual ACF and PACF plots. As well as Ljung-Box test used to detect correlation and can be used Ljung-box squared test for residual to check whether homoscedasticity of residuals.

Below are the results of test conducted for residuals for ARIMA (1,0,0) for Samba.

Type of Test	χ^2 Value	df	P Value
Ljung-Box test	6.6518	9	0.6733
Ljung-box squared test	7.5485	9	0.5802

Table 4.11: Box-test results of residuals for ARIMA (1,0,0) for Samba

For testing correlation of residuals for ARIMA (1,0,0) model, this test is on the null hypothesis of calming that there are no autocorrelations between residuals.

$\boldsymbol{H_0}$: There are no autocorrelations between residuals

$\boldsymbol{H_1}$: There are autocorrelations between residuals.

According to above R output, p-value of Ljung-Box test is 0.6733. Thus, the null hypothesis of Ljung-Box test isn't rejected at 5% level of significance. So, there is no correlation of residuals of the ARIMA (1,0,0) model.

Then we were used Ljung-box squared test for residual to check whether homoscedasticity of residuals. It is more helpful determine residuals have constant variance.

$\boldsymbol{H_0}$: There is homoscedasticity between residuals

$\boldsymbol{H_1}$: there isn't homoscedasticity between residuals.

According to above R output, p-value of Ljung-Box squared test is 0.5802. Thus, the null hypothesis of Ljung-Box squared test isn't rejected at 5% level of significance. So, there is no heteroscedasticity of residuals of the ARIMA (1,0,0) model. Thus, error term variance of ARIMA (1,0,0) model is constant.

Thus, the next thing is normality test of residuals it can be shows in following Histogram of residuals plots.

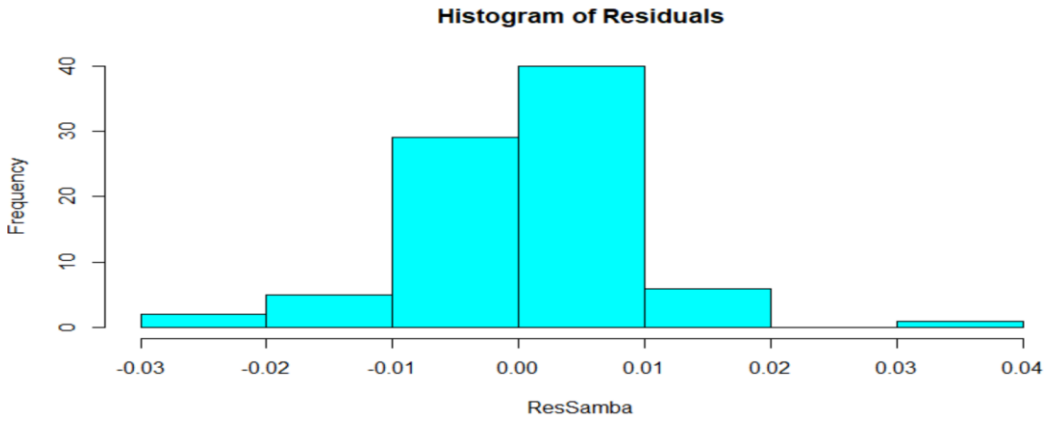

Figure 4.6: Histogram of residual of ARIMA (1,0,0) for Samba

Figure 4.6 shows the Histogram of residuals of the selected model. Histogram plot shows bell shape graph and its bit skewed to left side. Thus, this graphical scenario can for the normality condition in residuals.

Using Jarque-Bera test we were concluded this result as follows

Corresponding hypotheses are

$\boldsymbol{H_0}$: Residuals are normally distributed

$\boldsymbol{H_1}$: residuals are not normally distributed.

Jarque.Bera.test(Residuals of ARIMA (1,0,0) for Samba)		
$\boldsymbol{\chi^2}$ = 44.267	df = 2	p-value 2.44e-10=$2.44 * 10^{-10}$

Table 4.12: Jarque-Bera test results of ARIMA (1,0,0) for Samba

According to Table 4.12 shows the Jarque-Bera normality test results. Which p-value is less than 0.01. Thus, the null hypothesis of normality of residuals being rejected. Forecasting does not need the premise of normality. Rather, it refers to data stationarity. Forecasting can be done even if residuals are determined to be non-normally distributed (through, for example, the Jarque Bera Test). Forecasting should be avoided if the series is determined to be non-stationary (through, for example, the Augmented Dickey Fuller Test), because such forecasts would be inaccurate.

4.4.1.8 Forecasting using ARIMA (1,0,0)

The next step is forecasting using ARIMA (1,0,0).

Week	**Forecast**	**Actual**
2019-13	105.5470258	105
2019-14	105.96904	97
2019-15	106.2943058	95
2019-16	106.5447366	105
2019-17	106.7374935	95
2019-18	106.8855509	95
2019-19	106.999562	90

2019-20	107.0870611	95
2019-21	107.1543978	90
2019-22	107.2062241	95

Table 4.13: Forecast values of ARIMA (1,0,0)

According to this forecasting values, it gives difference value in entire space. Therefore, this implies a considerable forecasting value for the weekly average retail price of Samba.

Actual Vs forecast graph shows in bellow

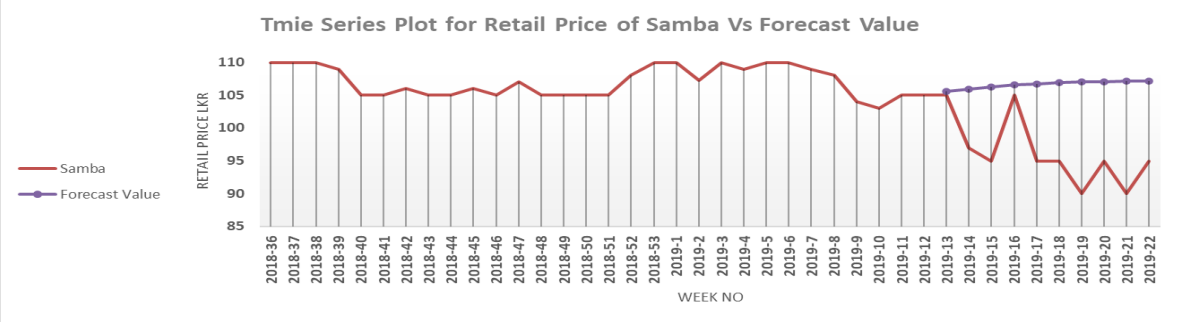

Figure 4.7: Actual vs Forecast graph of ARIMA (1,0,0) for Samba

According to the Figure 4.7, forecast values show the small upward trend and the actual value shows downward trend over the time period.

4.4.1.9 Accuracy of ARIMA (1,0,0)

After forecasting, we want to check validation of forecasting model. For the validation process we used to compute MAPE, RMSE and MAE of the selected model. These values called as accuracy measurements. Lowest accuracy measurements imply most appropriate forecasting model.

Measurement	**Value**
MAPE	11.2%
RMSE	11.69807763
MAE	10.4543022

Table 4.14: Accuracy measurements of ARIMA (1,0,0) for Samba

MAPE value is 11.2% which is greater than 10% and less than 20%. And also, we can see that the RMSE and MAE values are small considerably. According to these results, we can clearly say that the model ARIMA (1,0,0) gives Good accurate forecast.

4.4.2 Forecasting retail price of Samba Using Exponential Smoothing

According to the figure 4.1 of original time series plot, there is no seasonal pattern and there is no trend. And also, we can clearly see that the time series data are changing slowly over time. Therefore, Single exponential smoothing is more appropriate for this type.

For the optimum model parameters, the R output as follows

Holt-Winters exponential smoothing without trend and without seasonal component		
Call: HoltWinters(x = Samba, beta =FALSE, gamma = FALSE)		
Smoothing parameters:		
alpha: 0.7343156	beta: FALSE	gamma: FALSE
Coefficients: [,1]	a = 104.8992	

Table 4.15: Parameter estimation of Exponential smoothing for Samba

Where a is the intercept. Alpha is the smoothing constant for the level of the series. We can see that alpha value is 0.7343156. When alpha is close to 1, Thus, this model sets the current smoothed point to the current point means the smoothed series is the original series. Therefor we can use this model for short term forecasting.

4.4.2.1 Diagnostic checking of exponential smoothing

According to this below figure, residual ACF and PACF shows lags are not significant. Since the residuals haven't significance autocorrelation. Then residuals are uncorrelated. This implies the model is suitable for forecasting process.

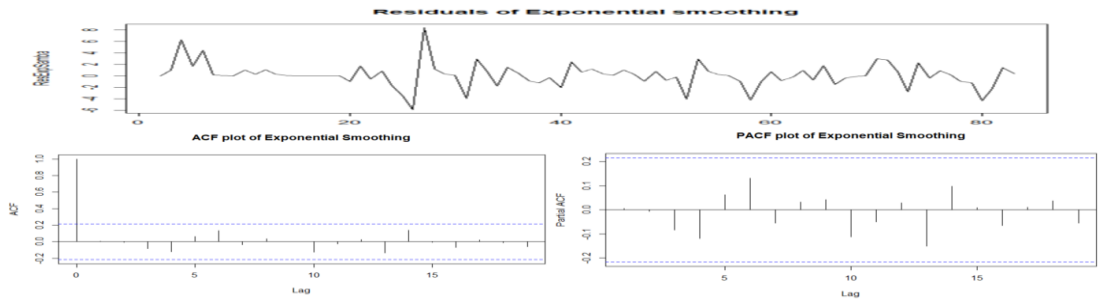

Figure 4.8: ACF and PACF of residuals and Residuals plot of exponential smoothing

Also, we were preformed Box test for the checking the residual autocorrelation and residual heteroscedasticity

Type of Test	χ^2 Value	df	P Value
Ljung-Box test	5.6485	10	0.8439
Ljung-box squared test	9.9652	10	0.4435

Table 4.16: Box test of residuals of exponential smoothing for Samba

According to above R output, p-value of Ljung-Box test is 0.8439 which is greater than 0.05. Thus, the null hypothesis of Ljung-Box test isn't rejected at 5% level of significance.so there is no correlation of residuals of the exponential smoothing model.

Now we are preformed Box test for the checking the residual heteroscedasticity.

According to above R output, p-value of Ljung-Box squared test is 0.4435 which is greater than 0.05. Thus, the null hypothesis of Ljung-Box squared test isn't rejected at 5% level of significance. So, there is no heteroscedasticity of residuals of the Exponential smoothing model.

Now we are preformed normality test for the residuals of Exponential smoothing model.

Also, residual histogram shows the normality assumption of fitted model.

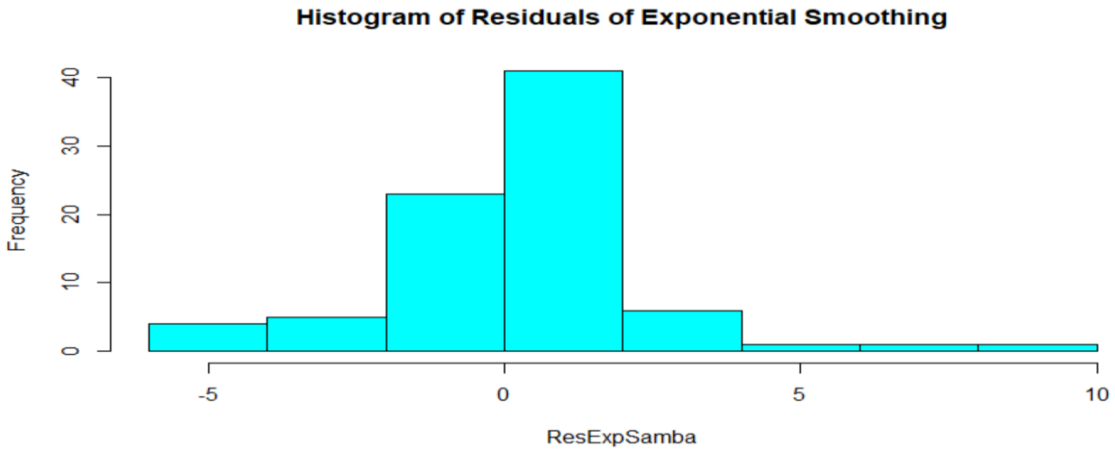

Figure 4.9: Histogram of residuals of exponential smoothing

Figure 4-9 shows the Histogram of residuals of the selected model. Histogram plot shows bell shape graph and its bit skewed to left side. Thus, this graphical scenario can for the normality condition in residuals.

Using Jarque-Bera test we were checked the normality condition of residuals.

Jarque.Bera.Test(Residuals of Exponential smoothing of Samba)		
χ^2 = 44.254	df = 2	p-value < 2.457e-10

Table 4.17: Jarque-Bera test results of exponential smoothing for Samba

According to Table 4.17 shows the Jarque-Bera normality test results which p-value is less than 0.05. Thus, null hypothesis of normality of residuals being rejected. As mentioned above 4.4.1.7, still we can do the forecasting.

4.4.2.2 Forecasting using exponential smoothing

Week	**Forecast**	**Actual**
2019-13	104.8992	105
2019-14	104.8992	97
2019-15	104.8992	95
2019-16	104.8992	105
2019-17	104.8992	95
2019-18	104.8992	95
2019-19	104.8992	90
2019-20	104.8992	95
2019-21	104.8992	90
2019-22	104.8992	95

Table 4.18: Forecast using exponential smoothing

We can see that, Exponential smoothing model gives the same forecasting values since the alpha is close to one.

4.4.2.3 Accuracy of exponential smoothing

Measurement	Value
MAPE	9.3%
RMSE	9.981737969
MAE	8.731122437

Table 4.19: Accuracy measurements of exponential smoothing for Samba

MAPE value is 9.3% which is less than 10%. And also, we can see that the RMSE and MAE values are small considerably. Then this exponential smoothing model gives highly accurate forecast.

4.4.3 Forecasting evaluation of Samba Price

Finally, we were found a model for forecasting process by using minimum MAPE, RMSE, MAE of foregoing performed models.

Model	MAPE	RMSE	MAE
ARIMA(1,0,0)	11.2%	11.69807763	10.4543022
Exponential Smoothing	**9.3%**	**9.981737969**	**8.731122437**

Table 4.20: Forecasting evaluation models of Samba Price

According to this figure we were found the both models exponential smoothing method and ARIMA(1,0,0) are the best method for the estimating forecast values. But MAPE value of ARIMA(1,0,0) is greater than MAPE value of Exponential smoothing. It implies Exponential smoothing forecasting model is better than ARIMA(1,0,0) model. As per the subsection 4.4.2 and the Table 4.15 mentioned above, we can use Exponential Smoothing for short-term forecasting. For the long-term forecasting, we can recommend the best model is ARIMA(1,0,0).

We were found the exponential smoothing method and ARIMA(1,0,0) both are the best method for the forecasting process. Since exponential smoothing method can perform a time series approach for the determine future values of Samba Retail price for short time and ARIMA(1,0,0) can perform a time series approach for the determine future values of Samba Retail price for long time.

4.5 Retail Price of Nadu

We were fit a new model for forecast retail price of Nadu. Fitted model contain 83 observations.

4.5.1 Nadu retail prices forecast using ARIMA model.

4.5.1.1 Time Series Plot of Nadu price

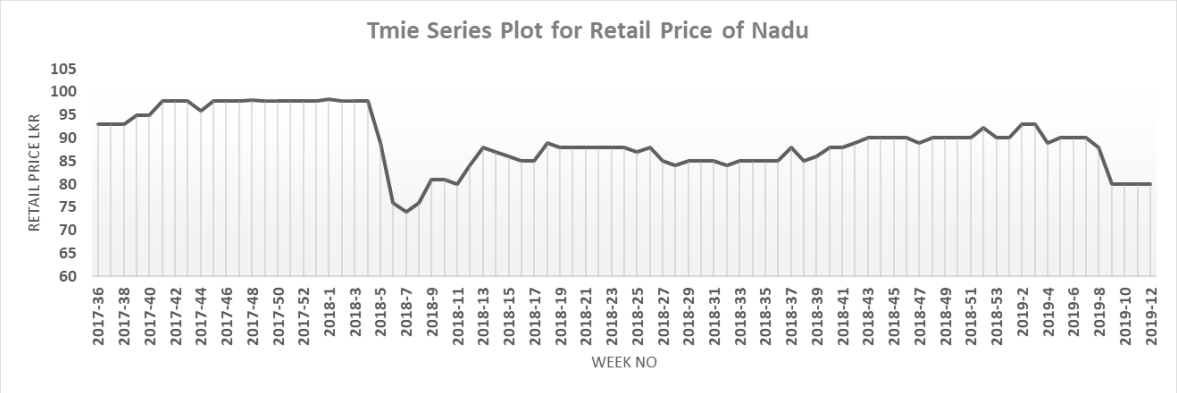

Figure 4.10: Time series graph of retail price of Nadu

According to figure 4.10 we can see that retail prices of Nadu vary with 75 and 100 LKR. Also clearly see that there is a downward trend. And cannot clearly say that there are have any seasonality pattern. There is some constant value during the week no 2017-45 to week no 2018-4

4.5.1.2 Time series plot of Transformed data for the price of Nadu

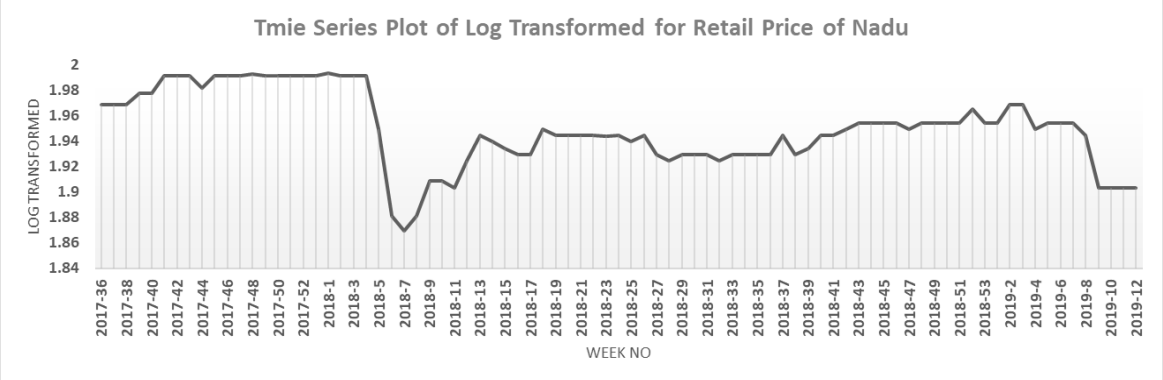

Figure 4.11: Time series graph of Log transformed price of Nadu

Figure 4.11 shows prices vary between 1.86 and 2.00. This implies transform should reduce heteroscedasticity.

4.5.1.3 Stationery Test

First of all, we want to check the original time series observations are stationery or not.

ADF.test (Nadu Log Transformed)		
Dickey-Fuller = -2.539	Lag order = 4	p – value = 0.3549
alternative hypothesis: stationary		

Table 4.21: Results of ADF test of price of Nadu log transformed level data

According to ADF test of weekly average prices of Nadu P_value is 0.3549 which is greater than 0.05. Thus, the null hypothesis of non-stationery being not rejected at 5% level of significance.

KPSS test also used to test stationery of prices of Nadu

KPSS.test (Nadu Log Transformed)		
KPSS Level = 0.54668	Truncation lag parameter = 3	p – value = 0.03115

Table 4.22: Results of KPSS test of price of Nadu log transformed level data

According to KPSS test of weekly average prices of Nadu P value is 0.03115 which is less than 0.05. Thus, the null hypothesis of stationery being rejected at 5% level of significance.

Both ADF and KPSS test given the non-stationary time series for the original observation. Hence there is no need to go for the PP test.

According to these 2 tests we can conclude that original weekly average prices of Nadu are given non-stationery observations. Thus, we were used differencing to achieve to stationery time series.

4.5.1.4 Stationery Test for log differenced data

According to foregoing results time series observations are non-stationery. Since we used to difference operation and again test the stationery condition of log transformed differenced data.

ADF.test (Nadu Log Transformed Difference)		
Dickey-Fuller = -3.6934	Lag order = 4	p – value = 0.03049
alternative hypothesis: stationary		

Table 4.23: Results of ADF test of price of Nadu log transformed 1st differenced data

According to ADF test of weekly average prices of Nadu P value is 0.03049 which is less than 0.05. Thus, the null hypothesis of non-stationery being rejected at 5% level of significance. Thus, according to ADF test the 1st differenced data should be stationery.

KPSS test also used to test stationery of prices of Nadu

KPSS.test (Nadu Log Transformed Difference)		
KPSS Level = 0.058807	Truncation lag parameter = 1	p – value = 0.1
Warning message: In kpss.test(NaduLogDiff): p-value greater than printed p value		

Table 4.24: Results of KPSS test of price of Nadu log transformed 1st differenced data

According to KPSS test of weekly average prices of Nadu P value is 0.1 which is greater than 0.05. Thus, the null hypothesis of stationery being not rejected at 5% level of significance. Thus, according to KPSS test the 1st differenced data should be stationery. According to these 2 tests we can conclude that differenced original weekly average prices of Nadu are given stationery observations. Thus, we were continued model selection procedure

4.5.1.5 Model Identification

According to stationery results log transformed 1st difference observations of price of Nadu are stationery. Since we can find a suitable time series model after estimating significant lags.

4.5.1.5.1 ACF and PACF of transformed & difference data

The ACF and PACF graph of difference transformed data are computed as in Figure 4.12 and Figure 4.13 Then, the patterns of ACF & PACF were used to determine the parameter values of p and q.

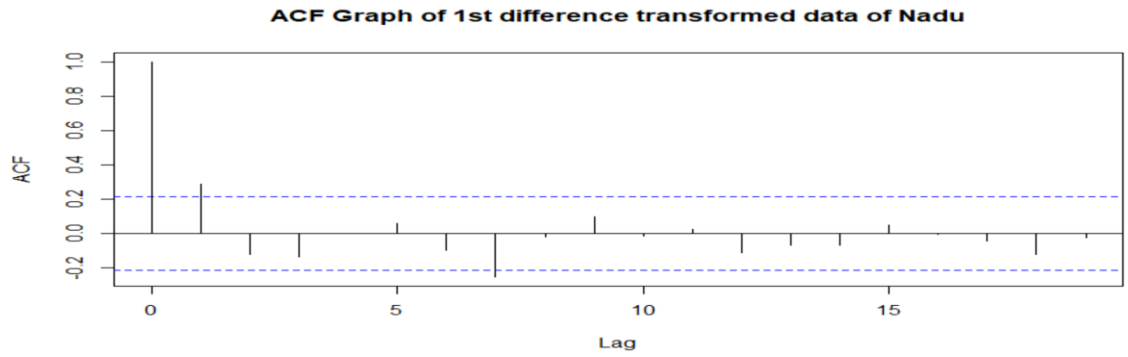

Figure 4.12: ACF plot of transformed 1st differenced observations of Nadu

According to the Figure 4.12 lags are dies after 1st lags. We cannot see there are significant spikes after the lag 1, thus MA(1) and MA(0) are the suitable parameters and q must be 1 or 0.

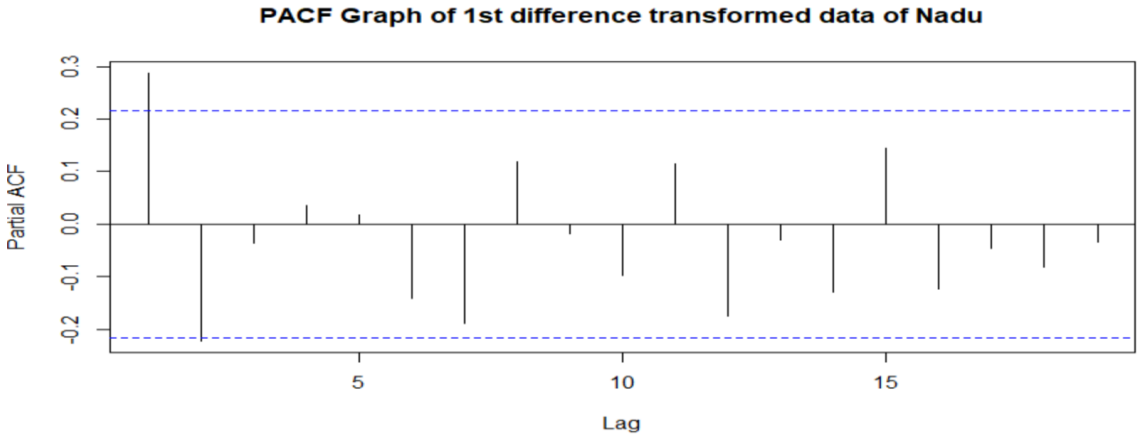

Figure 4.13: PACF plot of transformed 1st differenced observations of Nadu

According to the Figure 4.13 lags are dies after 2nd lags. We cannot see any significant spikes after lag 2, thus AR(2) and AR(1) are the suitable parameters and p must be 2 or 1.

Finally, we have selected different type of models and we were checked the most appropriate model for the continue forecasting procedure.

4.5.1.6 Model selection

After identifying the feasible values of p and q, the next step is an estimate the parameters under the most suitable model. The parameter estimation of model is conducted by R

software. ARIMA model have been selected using minimum AIC value. Table 4-27 gives the ARIMA models with AIC values.

Model	**AIC**
ARIMA(1,1,0)	-487.87
ARIMA(1,1,1)	-488.52
ARIMA(2,1,0)	**-489.86**
ARIMA(2,1,1)	-487.92

Table 4.25: Results of model selection of ARIMA (2,1,0) for Nadu

ARIMA (2,1,0) model has the minimum AIC value. Thus ARIMA (2,1,0) is selected as suitable model for forecasting the weekly average retail prices of Nadu.

4.5.1.7 Parameter Estimation

ARIMA (Nadu Log Transformed, order = c (2,1,0))		
Coefficients		
	ar1	ar2
	0.3513	-0.2152
s.e.	0.1072	0.1063
t ratio	3.277345	-2.023432
p value	0.001047884	0.043028645
σ^2 estimated as 0.0001382	Log likelihood = 247.93	AIC = -489.86

Table 4.26: Parameter estimation of ARIMA (2,1,0)

According to the Table 4.26, the p-values of terms are less than 0.05. Thus, we can conclude that the coefficient is statistically significant. So, we can fit the model with the terms.

The model equation can be formulated as

$$y_t = y_{t-1} + 0.3513(y_{t-1} - y_{t-2}) - 0.2152(y_{t-2} - y_{t-3}) + \varepsilon_t$$

Where y_t is the weekly average retail prices of Nadu in i[th] and ε_t is the error at i[th] time period.

4.5.1.8 Diagnostic checking

After estimating the parameter for ARIMA (2,1,0) model, next step will be the diagnostic checking.

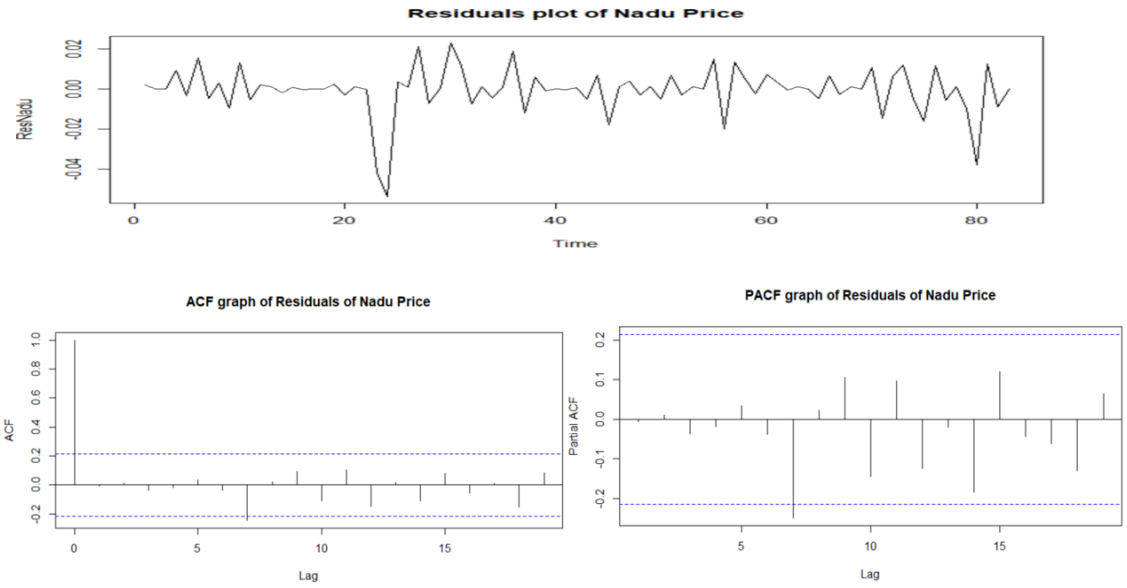

Figure 4.14: ACF and PACF plot of residuals and residual plot of ARIMA (2,1,0)

According to figure 4.14. Top of the box contain the time plot of standardized residuals. It shows that no pattern and looks like an independent identical distributed. And residual ACF and PACF shows all of lags are not significant. Since the residuals haven't significance autocorrelation.

Most important property of residual diagnostic is the residuals are uncorrelated. It means residuals are independent each other. To detect the correlation, we can study residual ACF and PACF plots. As well as Ljung-Box test used to detect correlation.

For testing correlation of residuals for ARIMA (2,1,0) model for Nadu

Type of Test	χ^2 Value	df	P Value
Ljung-Box test	8.1032	8	0.4235
Ljung-box squared test	11.306	8	0.185

Table 4.27: Box-test results of Residuals of ARIMA (2,1,0) for Nadu

According to above R output, p-value of Ljung-Box test is 0.4235. Thus, the null hypothesis of Ljung-Box test isn't rejected at 5% level of significance.so there is no correlation of residuals of the ARIMA (2,1,0) model.

Then we were used Ljung-box squared test for residual to check whether homoscedasticity of residuals. It is more helpful determine residuals have constant variance.

According to above R output, p-value of Ljung-Box squared test is 0.185. Thus, the null hypothesis of Ljung-Box test isn't rejected at 5% level of significance. So, there is no heteroscedasticity of residuals of the ARIMA (2,1,0) model. Thus, error term variance of ARIMA (2,1,0) model is constant.

Thus, the next thing is normality test of residuals it can be shows in following Histogram of residuals plots.

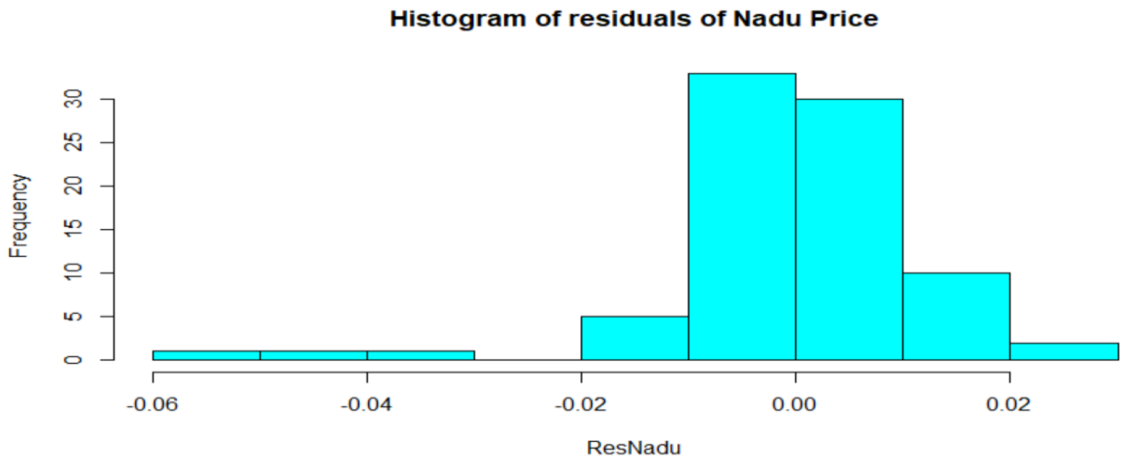

Figure 4.15: Histogram of residual of ARIMA (2,1,0)

Figure 4.15 shows the Histogram of residuals of the selected model. Histogram plot shows bell shape graph and, but it skewed to right side.

Using Jarque-Bera test we were checked the normality condition of residuals.

Jarque.Bera.Test(Residuals of ARIMA (2,1,0) for Nadu)		
χ^2 = 180.63	df = 2	p-value < 2.2e-16=$2.2* 10^{-16}$

Table 4.28: Jarque-Bera test results of Residuals of ARIMA (2,1,0) for Nadu

According to Table 4.28 shows the Jarque-Bera normality test results. Which p-value < $2.2* 10^{-16}$ is less than 0.05. Thus, the null hypothesis of normality of residuals being rejected. Forecasting does not need the premise of normality. Rather, it refers to data stationarity. Forecasting can be done even if residuals are determined to be non-normally distributed (through, for example, the Jarque Bera Test). Forecasting should be avoided if the series is determined to be non-stationary (through, for example, the Augmented Dickey Fuller Test), because such forecasts would be inaccurate.

4.5.1.9 Forecasting using ARIMA (2,1,0)

Week	**Forecast**	**Actual**
2019-13	80.0000024	80
2019-14	80.0000024	80
2019-15	80.0000024	80
2019-16	80.0000024	80
2019-17	80.0000024	80
2019-18	80.0000024	80
2019-19	80.0000024	90
2019-20	80.0000024	90
2019-21	80.0000024	90
2019-22	80.0000024	90

Table 4.29: Forecast values of ARIMA (2,1,0) for Nadu

According to the Table 4.29, we can see that the forecasting given the same value. This happen because of the selected model gives the value y_t based on the previous 3

observation which are y_{t-1}, y_{t-2} and y_{t-3}. According to our data sets last 3 observation of the price Nadu is 80. That is the reason why we got same forecasted values.

4.5.1.10 Accuracy of ARIMA (2,1,0)

After forecasting, we want to check validation of forecasting model. For the validation process we used to compute MAPE, RMSE and MAE of the selected model. These values called as accuracy measurements. Lowest accuracy measurements imply most appropriate forecasting model.

Measurement	**Value**
MAPE	4.5%
RMSE	6.327204331
MAE	4.080267988

Table 4.30: Accuracy measurements of ARIMA (2,1,0) for Nadu

MAPE value is 4.5% which is less than 10%. And also, we can see that the RMSE and MAE values are small considerably. According to these results, we can clearly say that the model ARIMA (2,1,0) gives highly accurate forecast.

4.5.2 Forecasting Retail Price of Nadu using Double Exponential Smoothing

According to the figure 4.10 of original time series plot, there is no seasonal pattern. But there is downward trend. And also, we can clearly see that the time series data are changing slowly over time. Therefore, double exponential smoothing forecast, also known as Holt's linear method, can be used for the series with trend.

For the optimum model parameters, the R output as follows

Holt-Winters exponential smoothing with trend and without seasonal component		
Call: HoltWinters(x = Nadu, gamma = FALSE, b.start=23)		
Smoothing parameters:		
alpha: 1	beta: 0.7520755	gamma: FALSE
Coefficients: [,1]	a = 80.00	b = -0.09736804

Table 4-31: Parameter estimation of Double Exponential smoothing for Nadu

Where a is the intercept and b is the trend component. Alpha is the smoothing constant for the level of the series. Beta is the smoothing constant for the trend. We can see that alpha value is 1 and beta value is close to 1. Thus, this model sets the current smoothed point to the current point means the smoothed series is the original series. Therefor we can use this model for short term forecasting.

4.5.2.1 Diagnostic checking of double exponential smoothing

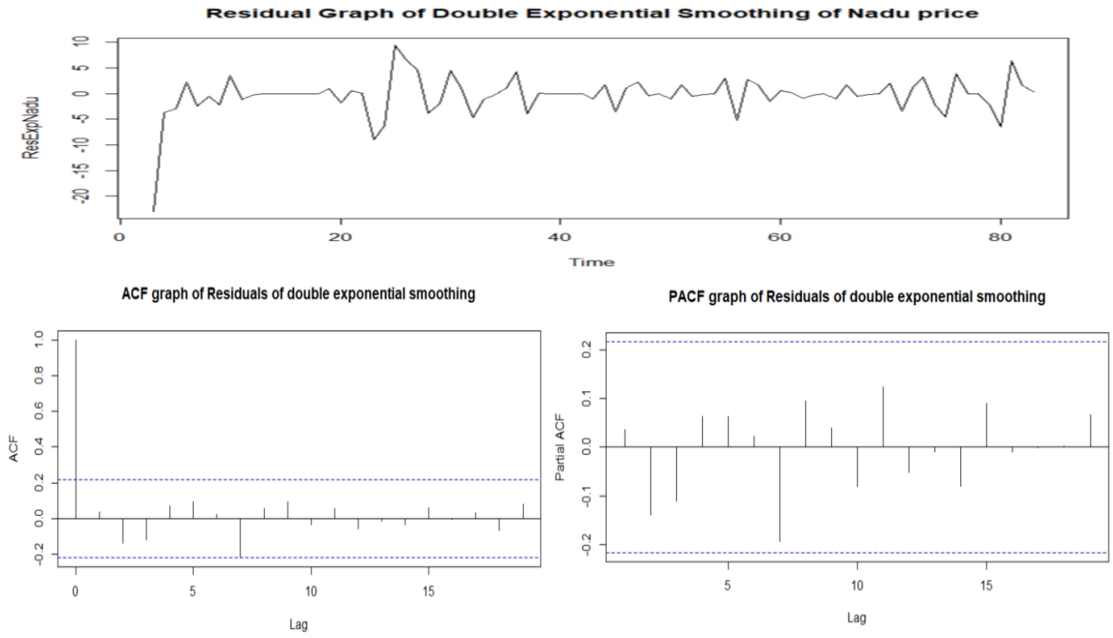

Figure 4.16: ACF and PACF of residuals and Residuals plot of double exponential smoothing

According to this figure, residual ACF and PACF shows lags are not significant. Since the residuals haven't significance autocorrelation. Then residuals are uncorrelated. This implies the model is suitable for forecasting process.

Also, we were preformed Box test for the checking the residual autocorrelation and residual heteroscedasticity.

Type of Test	χ^2 Value	df	P Value
Ljung-Box test	9.8759	10	0.4514
Ljung-box squared test	0.36333	10	1.0

Table 4.32: Box test of Residuals of double exponential smoothing for Nadu

According to above R output, p-value of Ljung-Box test is 0.4514 which is greater than 0.05. Thus, the null hypothesis of Ljung-Box test isn't rejected at 5% level of significance.so there is no correlation of residuals of the double exponential smoothing model.

Now we are preformed Box test for the checking the residual heteroscedasticity.

According to above R output, p-value of Ljung-Box squared test is 1 which is greater than 0.05. Thus, the null hypothesis of Ljung-Box squared test isn't rejected at 5% level of significance. So, there is no heteroscedasticity of residuals of the Double Exponential smoothing model.

Now we are preformed normality test for the residuals of Double Exponential smoothing model.

Figure 4.17 shows the Histogram of residuals of the selected model. Histogram plot shows bell shape graph and, but it skewed to right side.

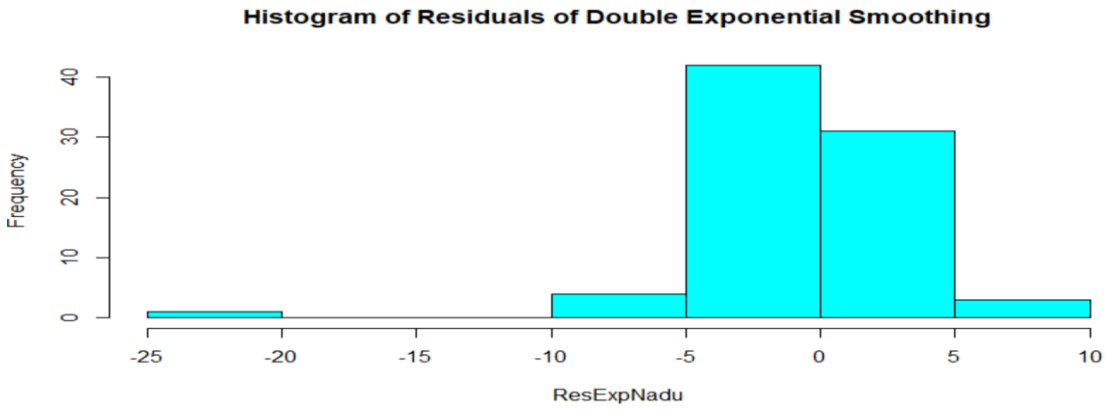

Figure 4.17: Histogram of residuals of double exponential smoothing

Using Jarque-Bera test we were checked the normality condition of residuals.

Jarque.Bera.Test(Residuals of Double Exponential Smoothing of Nadu)		
χ^2 = 709.41	df = 2	p-value < 2.2e-16=$2.2* 10^{-16}$

Table 4.33: Jarque-Bera test results of Residuals of double exponential smoothing

According to figure 4.33 shows the Jarque-Bera normality test results which p-value is less than 0.05. Thus, null hypothesis of normality of residuals being rejected. As mentioned above 4.4.1.7, still we can do the forecasting.

4.5.2.2 Forecasting using exponential smoothing

Week	Forecast	Actual
2019-13	79.90263	80
2019-14	79.80526	80
2019-15	79.7079	80
2019-16	79.61053	80
2019-17	79.51316	80
2019-18	79.41579	80
2019-19	79.31842	90
2019-20	79.22106	90
2019-21	79.12369	90
2019-22	79.02632	90

Table 4.34: Forecast using double exponential smoothing for weekly average retail price of Nadu

4.5.2.3 Accuracy of double exponential smoothing

Measurement	Value
MAPE	5.2%
RMSE	6.862099179
MAE	4.615792467

Table 4.35: Accuracy measurements of double exponential smoothing

MAPE value is 5.2% which is less than 10%. And also, we can see that the RMSE and MAE values are small considerably. Then this double exponential smoothing model gives highly accurate forecast.

4.5.3 Forecasting evaluation of Nadu Price

Finally, we were found a model for forecasting process by using minimum MAPE, RMSE, MAE of foregoing performed models.

Model	MAPE	RMSE	MAE
ARIMA(2,1,0)	4.5%	6.327204331	4.080267988
Double Exponential Smoothing	5.2%	6.862099179	4.615792467

Table 4.36: Forecasting evaluation models of Nadu Price

According to this figure we were found the both models double exponential smoothing method and ARIMA(2,1,0) are the best method for the estimating forecast values. But MAPE value of ARIMA(2,1,0) is less than MAPE value of Double Exponential smoothing. It implies ARIMA(2,1,0) model is better than double exponential smoothing model. Thus, we can recommend the best model is ARIMA(2,1,0).

We were found the ARIMA(2,1,0) is the best method for the forecasting process. Since ARIMA(2,1,0) can perform a time series approach for the determine future values of Nadu Retail price.

4.6 Retail Price of Kekulu White Rice

We were fit a new model for forecast retail price of Kekulu white rice. Fitted model contain 83 observations.

4.6.1 Kekulu White retail prices forecast using ARIMA model.

4.6.1.1 Time Series Plot of Kekulu White price

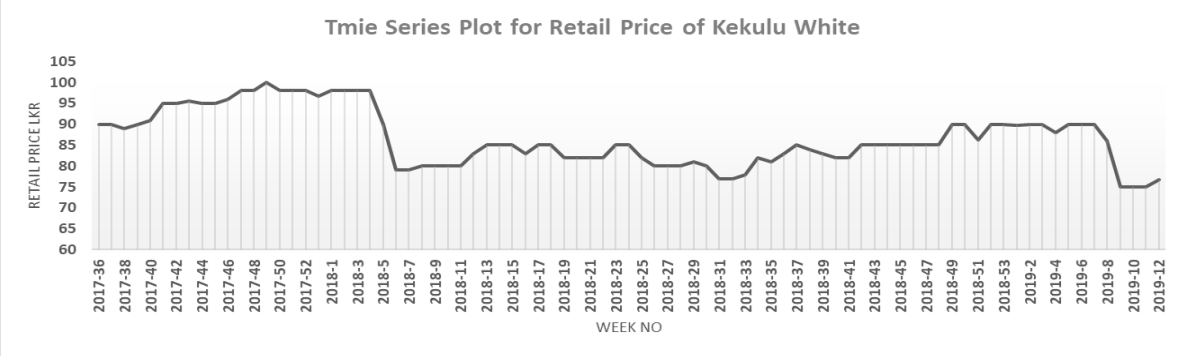

Figure 4.18: Time series graph of retail price of Kekulu White

According to figure 4.18 we can see that retail prices of Kekulu White vary with 75 and 100 LKR. Also clearly see that there is a downward trend. And cannot clearly say that there are have any seasonality pattern.

4.6.1.2 Time series plot of Transformed data for the price of Kekulu White

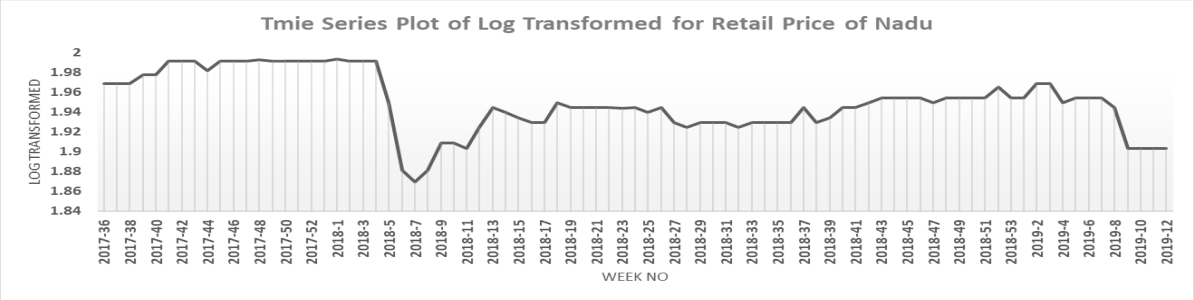

Figure 4.19: Time series graph of Log transformed price of Kekulu White

Figure 4.19 shows prices vary between 1.88 and 2.00. This implies transform should reduce heteroscedasticity.

4.6.1.3 Stationery Test

First of all, we want to check the original time series observations are stationery or not.

ADF.test (Kekulu White Log Transformed)		
Dickey-Fuller = -2.0271	Lag order = 4	p – value = 0.5647
alternative hypothesis: stationary		

Table 4.37: Results of ADF test of price of Kekulu White log transformed level data

According to ADF test of weekly average prices of Kekulu White P_value is 0.5647 which is greater than 0.05. Thus, the null hypothesis of non-stationery being not rejected at 5% level of significance.

KPSS test also used to test stationery of prices of Kekulu White

KPSS.test (Kekulu White Log Transformed)		
KPSS Level = 0.72102	Truncation lag parameter = 3	p – value = 0.01163

Table 4.38: Results of KPSS test of price of Kekulu White log transformed level data

According to KPSS test of weekly average prices of Kekulu White P value is 0.01163 which is less than 0.05. Thus, the null hypothesis of stationery being rejected at 5% level of significance.

Both ADF and KPSS test given the non-stationary time series for the original observation. Hence there is no need to go for the PP test.

According to these 2 tests we can conclude that original weekly average prices of Kekulu White are given non-stationery observations. Thus, we were used differencing to achieve to stationery time series.

4.6.1.4 Stationery Test for log differenced data

According to foregoing results time series observations are non-stationery. Since we used to difference operation and again test the stationery condition of log transformed differenced data.

ADF.test (Kekulu White Log Transformed Difference)		
Dickey-Fuller = -3.3453	Lag order = 4	p – value = 0.07011
alternative hypothesis: stationary		

Table 4.39: Results of ADF test of price of Kekulu White log transformed 1st differenced data

According to ADF test of weekly average prices of Kekulu White P value is 0.07011 which is greater than 0.05. Thus, the null hypothesis of non-stationery being not rejected at 5% level of significance. Thus, according to ADF test the 1st differenced data should be non-stationery.

KPSS test also used to test stationery of prices of Kekulu White

KPSS.test (Kekulu White Log Transformed Difference)		
KPSS Level = 0.074713	Truncation lag parameter = 3	p – value = 0.1
Warning message: In kpss.test(LogKekuluWhiteDiff): p-value greater than printed p value		

Table 4.40: Results of KPSS test of price of Kekulu White log transformed 1st differenced data

According to KPSS test of weekly average prices of Kekulu White P value is 0.1 which is greater than 0.05. Thus, the null hypothesis of stationery being not rejected at 5% level of significance. Thus, according to KPSS test the 1st differenced data should be stationery.

PP test also used to test stationery of prices of Kekulu White

PP.test (Kekulu White Log Transformed Difference)		
Dickey-Fuller Z(alpha) = -106.13	Truncation lag parameter = 3	p – value = 0.01
alternative hypothesis: stationary		
Warning message: In pp.test(LogKekuluWhiteDiff) : p-value smaller than printed p-value		

Table 4.41: Results of PP test of price of Kekulu White log transformed 1st differenced

According to PP test of weekly average prices of Kekulu White P value is 0.01 which is less than 0.05. Thus, the null hypothesis of non-stationery being rejected at 5% level of significance.

According to these 3 tests we can conclude that 1st differenced log transformation series are given stationery observations due to the results of 2 test out of 3 tests provide stationery series. Thus, we were continued model selection procedure.

4.6.1.5 Model Identification

According to stationery results log transformed 1st difference observations of price of Kekulu White are stationery. Since we can find a suitable time series model after estimating significant lags.

4.6.1.5.1 ACF and PACF of transformed & difference data

The ACF and PACF graph of difference transformed data are computed as in Figure 4.20 and Figure 4.21 Then, these two graphs were used to determine the parameter values of p and q.

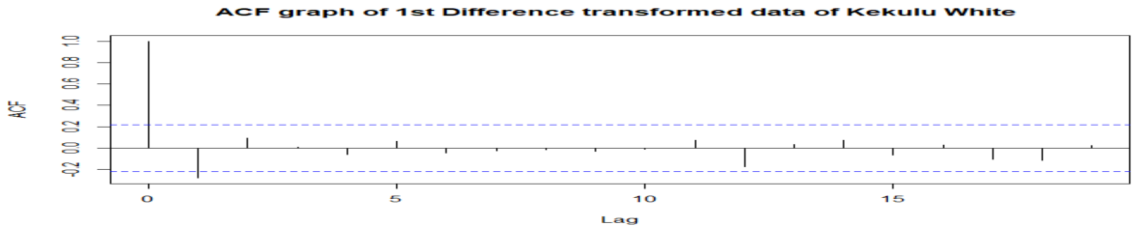

Figure 4.20: ACF plot of transformed 1st differenced observations of Kekulu White

According to the Figure 4.20 lags are dies after 1st lags. We cannot see there are significant spikes after the lag 1, thus MA(1) and MA(0) are the suitable parameters and q must be 1 or 0.

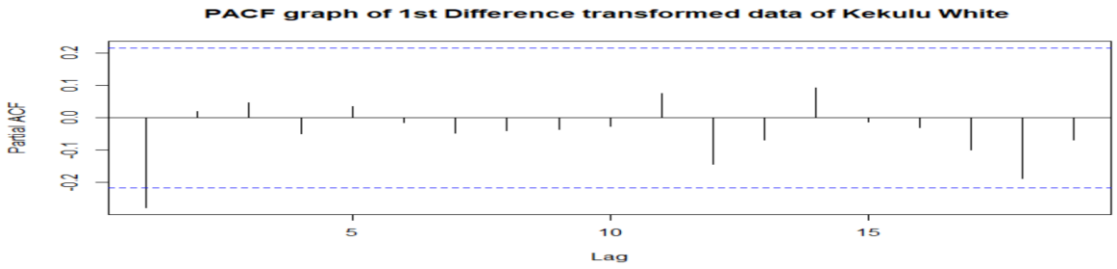

Figure 4.21: PACF plot of transformed 1st differenced observations of Kekulu White

According to the Figure 4.21 lags are dies after 1st lags. We cannot see any significant spikes after lag 1, thus AR(1) is the suitable parameters and p must be 1.

Finally, we have selected different type of models and we were checked the most appropriate model for the continue forecasting procedure.

4.6.1.6 Model selection

After identifying the feasible values of p and q, the next step is an estimate the parameters under the most suitable model. The parameter estimation of model is conducted by R software. ARIMA model have been selected using minimum AIC value. Table 4.42 gives the ARIMA models with AIC values.

Model	**AIC**
ARIMA(1,1,0)	**-439.32**
ARIMA(1,1,1)	-437.38

Table 4.42: Results of model selection of ARIMA (1,1,0) for Kekulu White

ARIMA (1,1,0) model has the minimum AIC value. Thus ARIMA (1,1,0) is selected as suitable model for forecasting the weekly average retail prices of Kekulu White.

4.6.1.7 Parameter Estimation

ARIMA (Kekulu White Log Transformed, order = c (1,1,0))		
Coefficients		
	ar1	
	-0. 2735	
s.e.	0.1056	
t ratio	- 2.588834	
p value	0.009630164	
σ^2 estimated as 0.0002625	Log likelihood = 221.66	AIC = -439.32

Table 4.43: Parameter estimation of ARIMA (1,1,0)

According to the Table 4.42, the p-values of terms are less than 0.05. Thus, we can conclude that the coefficient is statistically significant. So, we can fit the model with the terms.The model equation can be formulated as

$$y_t = y_{t-1} - 0.2735(y_{t-1} - y_{t-2}) + \varepsilon_t$$

Where y_t is the weekly average retail prices of Kekulu White in i[th] and ε_t is the error at i[th] time period.

4.6.1.8 Diagnostic checking

After estimating the parameter for ARIMA (1,1,0) model, next step will be the diagnostic checking.

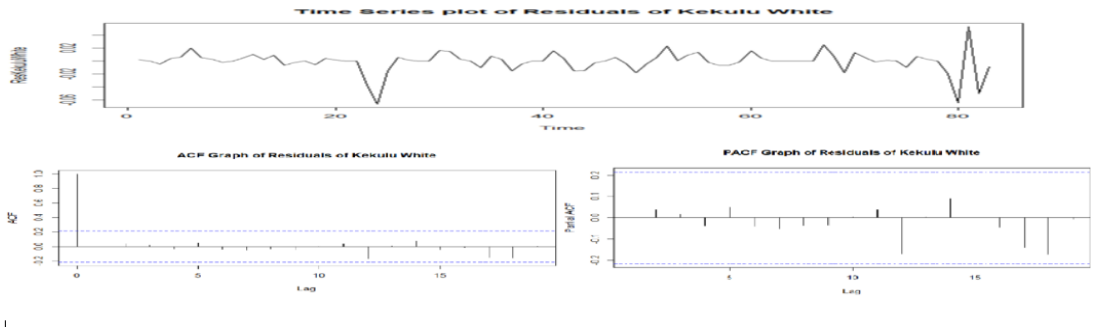

Figure 4.22: ACF and PACF plot of residuals and residual plot of ARIMA (1,1,0)

According to figure 4.22. Top of the box contain the time plot of standardized residuals. It shows that no pattern and looks like an independent identical distributed. And residual ACF and PACF shows all of lags are not significant. Since the residuals haven't significance autocorrelation.

To detect the correlation, we can study residual ACF and PACF plots. As well as Ljung-Box test used to detect correlation.

For testing correlation of residuals for ARIMA (1,1,0) model

Type of Test	χ^2 Value	df	P Value
Ljung-Box test	1.1495	8	0.9971
Ljung-box squared test	21.343	8	0.006291

Table 4.44: Box-test results of Residuals of ARIMA (1,1,0) for Kekulu White

According to above R output, p-value of Ljung-Box test is 0.9971. Thus, the null hypothesis of Ljung-Box test isn't rejected at 5% level of significance.so there is no correlation of residuals of the ARIMA (1,1,0) model.

Then we were used Ljung-box squared test for residual to check whether homoscedasticity of residuals. It is more helpful determine residuals have constant variance.

According to above R output, p-value of Ljung-Box squared test is 0.006291. Thus, the null hypothesis of Ljung-Box test is rejected at 5% level of significance. So, there is heteroscedasticity of residuals of the ARIMA (1,1,0) model. Thus, error term variance of ARIMA (1,1,0) model is not constant.

Thus, the next thing is normality test of residuals it can be shows in following Histogram of residuals plots.

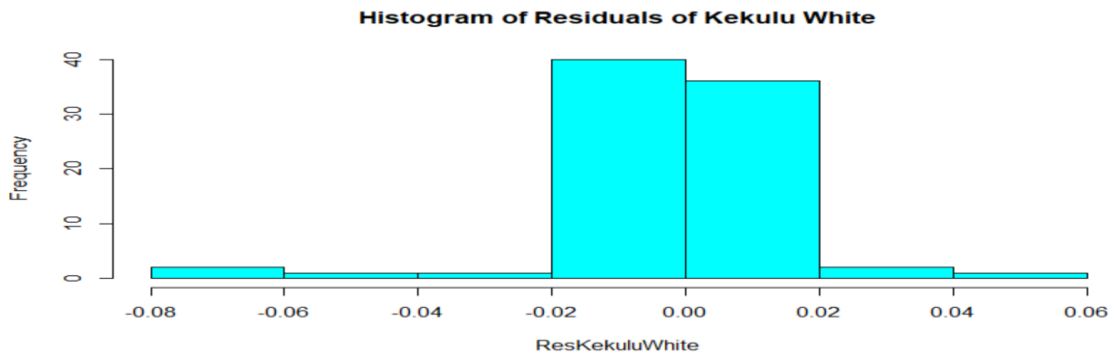

Figure 4.23: Histogram of residual of ARIMA (1,1,0)

Figure 4.23 shows the Histogram of residuals of the selected model. Histogram plot shows bell shape graph and, but it skewed to right side.

Using Jarque-Bera test we were checked the normality condition of residuals.

Jarque.Bera.Test(Residuals of ARIMA (1,1,0) for Kekulu White)		
χ^2 = 181.88	df = 2	p-value < 2.2e-16 =$2.2* 10^{-16}$

Table 4.45: Jarque-Bera test results of Residuals of ARIMA (1,1,0)

According to Table 4.44 shows the Jarque-Bera normality test results which p-value < $2.2* 10^{-16}$ is less than 0.05. Thus, the null hypothesis of normality of residuals being rejected. Forecasting does not need the premise of normality. Rather, it refers to data stationarity. Forecasting can be done even if residuals are determined to be non-normally distributed (through, for example, the Jarque Bera Test). Forecasting should be avoided if the series is determined to be non-stationary (through, for example, the Augmented Dickey Fuller Test), because such forecasts would be inaccurate.

4.6.1.9 Forecasting using ARIMA (1,1,0)

The next step is forecasting using ARIMA (1,1,0).

Week	**Forecast**	**Actual**
2019-13	76.24721758	75
2019-14	76.37689522	75
2019-15	76.34137889	75
2019-16	76.35104754	75
2019-17	76.34841052	75
2019-18	76.34911372	75
2019-19	76.34893792	85
2019-20	76.34893792	85
2019-21	76.34893792	85

2019-22	76.34893792	85

Table 4.46: Forecast values of ARIMA (1,1,0) for Kekulu White

According to the Table 4.45, According to this forecasting values, it gives difference value in entire space with increasing slightly. Therefore, this implies a considerable forecasting value for the weekly average retail price of Kekulu White.

Actual Vs forecast graph shows in bellow

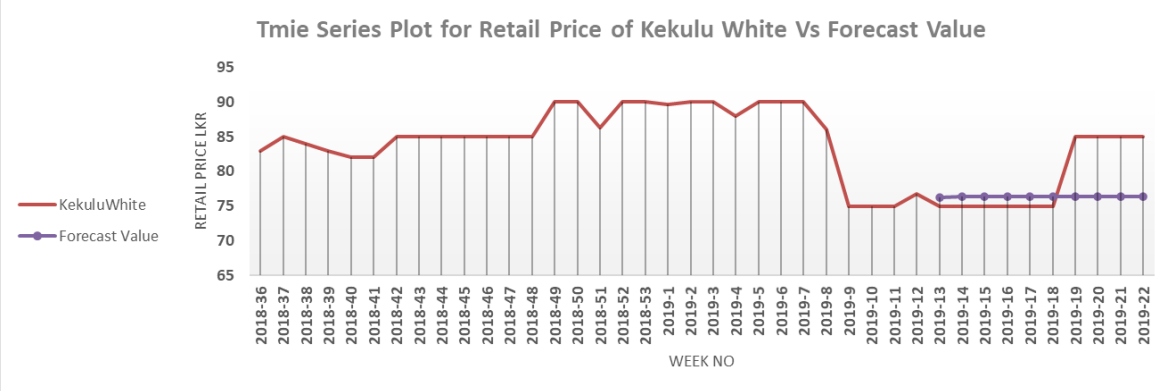

Figure 4.24: Actual vs Forecast graph of ARIMA (1,1,0)

4.6.1.10 Accuracy of ARIMA (1,1,0)

After forecasting, we want to check validation of forecasting model. For the validation process we used to compute MAPE, RMSE and MAE of the selected model. These values called as accuracy measurements. Lowest accuracy measurements imply most appropriate forecasting model.

Measurement	**Value**
MAPE	5.1%
RMSE	5.563566775
MAE	4.239822976

Table 4.47: Accuracy measurements of ARIMA (1,1,0) for Kekulu White

MAPE value is 5.1% which is less than 10%. And also, we can see that the RMSE and MAE values are small considerably. According to these results, we can clearly say that the model ARIMA (1,1,0) gives highly accurate forecast.

4.6.2 Forecasting Retail Price of Kekulu White using Double Exponential Smoothing

According to the figure 4.18 of original time series plot, there is no seasonal pattern. But there is downward trend. And also, we can clearly see that the time series data are changing slowly over time. Therefore, double exponential smoothing forecast, also known as Holt's linear method, can be used for the series with trend.

For the optimum model parameters, the R output as follows

Holt-Winters exponential smoothing with trend and without seasonal component		
Call: HoltWinters(x = KekuluWhiteData, gamma = FALSE, b.start=23)		
Smoothing parameters:		
alpha: 0.8573483	beta: 0.5937024	gamma: FALSE
Coefficients: [,1]	a = 76.403001	b = -1.895424

Table 4.48: Parameter estimation of Double Exponential smoothing

Where a is the intercept and b is the trend component. Alpha is the smoothing constant for the level of the series. Beta is the smoothing constant for the trend. We can see that alpha value is close to 1 and beta value is 0.5937024.

4.6.2.1 Diagnostic checking of double exponential smoothing

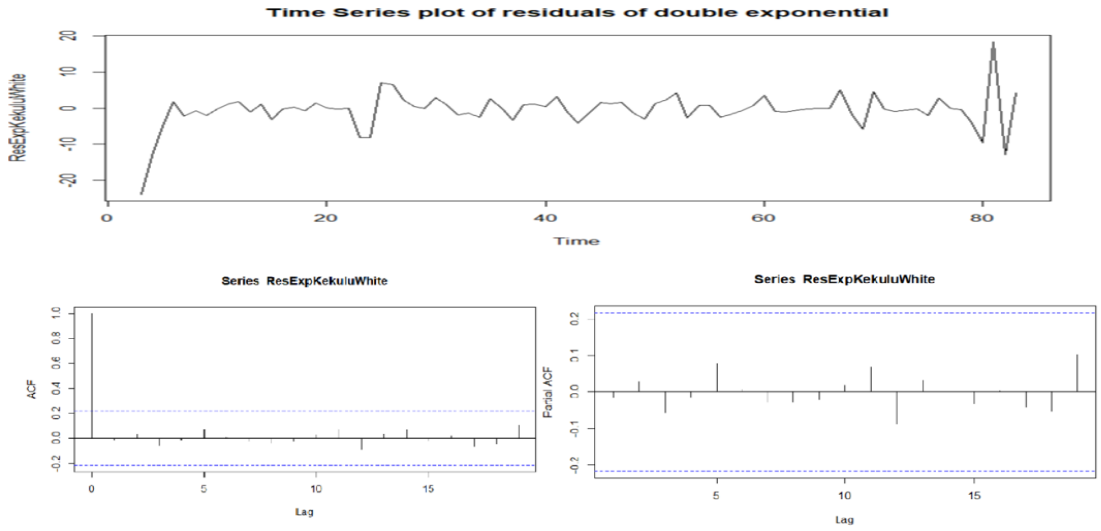

Figure 4.25: ACF and PACF of residuals and Residuals plot of double exponential smoothing for Kekulu White

According to this figure, residual ACF and PACF shows lags are not significant. Since the residuals haven't significance autocorrelation. Then residuals are uncorrelated. This implies the model is suitable for forecasting process.

Also, we were preformed Box test for the checking the residual autocorrelation and residual heteroscedasticity.

Type of Test	χ^2 Value	df	P Value
Ljung-Box test	1.1134	10	0.9997
Ljung-box squared test	11.647	10	0.3094

Table 4.49: Box test of Residuals of double exponential smoothing for Kekulu White

According to above R output, p-value of Ljung-Box test is 0.9997 which is greater than 0.05. Thus, the null hypothesis of Ljung-Box test isn't rejected at 5% level of significance.so there is no correlation of residuals of the double exponential smoothing model.

Now we are preformed Box test for the checking the residual heteroscedasticity.

According to above R output, p-value of Ljung-Box squared test is 0.3094 which is greater than 0.05. Thus, the null hypothesis of Ljung-Box squared test isn't rejected at 5% level of significance. So, there is no heteroscedasticity of residuals of the Double Exponential smoothing model.

Now we are preformed normality test for the residuals of Double Exponential smoothing model.

Figure 4.26 shows the Histogram of residuals of the selected model. Histogram plot shows bell shape graph and, but it skewed to right side.

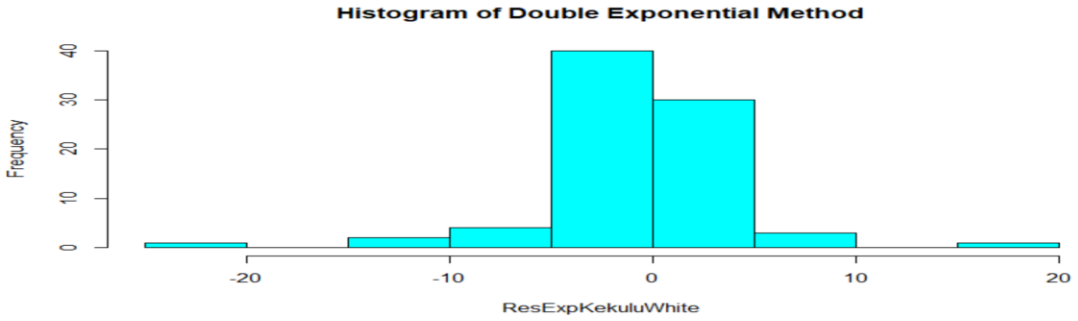

Figure 4.26: Histogram of residuals of double exponential smoothing

Using Jarque-Bera test we were checked the normality condition of residuals.

Jarque.Bera.Test(Residuals of Double Exponential Smoothing for Kekulu White)		
χ^2 = 260.33	df = 2	p-value < 2.2e-16=$2.2* 10^{-16}$

Table 4.50: Jarque-Bera test results of Residuals of double exponential smoothing

According to figure 4.49 shows the Jarque-Bera normality test results which p-value is less than 0.05. Thus, null hypothesis of normality of residuals being rejected. As mentioned above 4.4.1.7, still we can do the forecasting.

4.6.2.2 Forecasting using exponential smoothing

Week	**Forecast**	**Actual**
2019-13	74.50758	75
2019-14	72.61215	75
2019-15	70.71673	75
2019-16	68.82131	75
2019-17	66.92588	75
2019-18	65.03046	75
2019-19	63.13503	85
2019-20	61.23961	85
2019-21	59.34419	85
2019-22	57.44876	85

Table 4.51: Forecast using double exponential smoothing for weekly average retail price of Kekulu White

4.6.2.3 Accuracy of double exponential smoothing

Measurement	Value
MAPE	15.8%
RMSE	16.40297806
MAE	13.0438382

Table 4.52: Accuracy measurements of double exponential smoothing

MAPE value is 15.8% which is greater than 10% and less than 20%. And also, we can see that the RMSE and MAE values are small. Then this double exponential smoothing model gives good accurate forecast.

4.6.3 Forecasting evaluation of Kekulu White

Finally, we were found a model for forecasting process by using minimum MAPE, RMSE, MAE of foregoing performed models.

Model	MAPE	RMSE	MAE
ARIMA(1,1,0)	5.1%	5.563566775	4.239822976
Double Exponential Smoothing	15.8%	16.40297806	13.0438382

Table 4.53: Forecasting evaluation models of Kekulu White Price

According to this figure we were found the both models double exponential smoothing method and ARIMA(1,1,0) are the best method for the estimating forecast values. But MAPE value of ARIMA(1,1,0) is less than MAPE value of Double Exponential smoothing. It implies ARIMA(1,1,0) model is better than double exponential smoothing model. Thus, we can recommend the best model is ARIMA(1,1,0).

We were found the ARIMA(1,1,0) is the best method for the forecasting process. Since ARIMA(1,1,0) can perform a time series approach for the determine future values of Kekulu White Retail price.

4.7 Retail Price of Kekulu Red Rice

We were fit a new model for forecast retail price of Kekulu Red rice. Fitted model contain 83 observations.

4.7.1 Kekulu Red retail prices forecast using ARIMA model.

4.7.1.1 Time Series Plot of Kekulu Red price

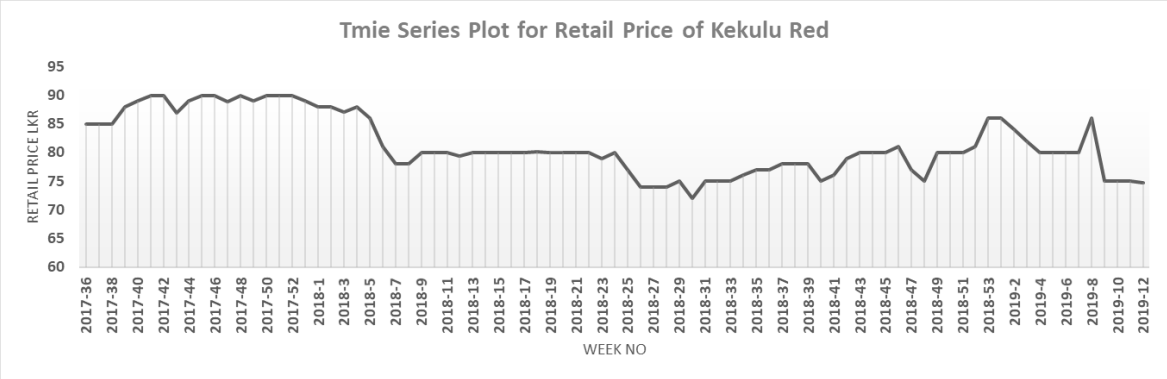

Figure 4.27: Time series graph of retail price of Kekulu Red

According to figure 4.27 we can see that retail prices of Kekulu Red vary with 72 and 90 LKR. Also clearly see that there is a downward trend. And cannot clearly say that there are have any seasonality pattern.

4.7.1.2 Time series plot of Transformed data for the price of Kekulu Red

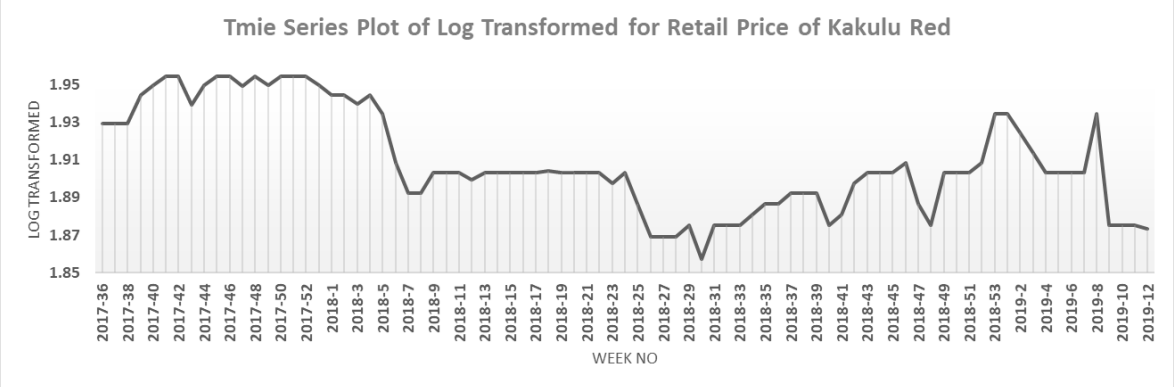

Figure 4.28: Time series graph of Log transformed price of Kekulu Red

Figure 4.28 shows prices vary between 1.86 and 1.96. This implies transform should reduce heteroscedasticity.

4.7.1.3 Stationery Test

First of all, we want to check the original time series observations are stationery or not.

ADF.test (Kekulu Red Log Transformed)		
Dickey-Fuller = -1.6664	Lag order = 4	p – value = 0.7126
alternative hypothesis: stationary		

Table 4.54: Results of ADF test of price of Kekulu Red log transformed level data

According to ADF test of weekly average prices of Kekulu Red P value is 0.7126 which is greater than 0.05. Thus, the null hypothesis of non-stationery being not rejected at 5% level of significance.

KPSS test also used to test stationery of prices of Kekulu Red

KPSS.test (Kekulu Red Log Transformed)		
KPSS Level = 1.0772	Truncation lag parameter = 3	p – value = 0.01
Warning message:In kpss.test(LogKekuluRedData) : p-value smaller than printed p-value		

Table 4.55: Results of KPSS test of price of Kekulu Red log transformed level data

According to KPSS test of weekly average prices of Kekulu Red P value is 0.01 which is less than 0.05. Thus, the null hypothesis of stationery being rejected at 5% level of significance.

Both ADF and KPSS test given the non-stationary time series for the original observation. Hence there is no need to go for the PP test.

According to these 2 tests we can conclude that original weekly average prices of Kekulu Red are given non-stationery observations. Thus, we were used differencing to achieve to stationery time series.

4.7.1.4 Stationery Test for log differenced data

According to foregoing results time series observations are non-stationery. Since we used to difference operation and again test the stationery condition of log transformed differenced data.

ADF.test (Kekulu Red Log Transformed Difference)		
Dickey-Fuller = -4.5863	Lag order = 4	p – value = 0.01
alternative hypothesis: stationary		
Warning message:In adf.test(LogKekuluRedDiff) : p-value smaller than printed p-value		

Table 4.56: Results of ADF test of price of Kekulu Red log transformed 1st differenced data

According to ADF test of weekly average prices of Kekulu Red P value is 0.01 which is less than 0.05. Thus, the null hypothesis of non-stationery being rejected at 5% level of significance. Thus, according to ADF test the 1st differenced data should be stationery.

KPSS test also used to test stationery of prices of Kekulu Red

KPSS.test (Kekulu Red Log Transformed Difference)		
KPSS Level = 0.11706	Truncation lag parameter = 3	p – value = 0.1
Warning message: In kpss.test(LogKekuluRedDiff): p-value greater than printed p value		

Table 4.57: Results of KPSS test of price of Kekulu Red log transformed 1st differenced data

According to KPSS test of weekly average prices of Kekulu Red P value is 0.1 which is greater than 0.05. Thus, the null hypothesis of stationery being not rejected at 5% level of significance. Thus, according to KPSS test the 1st differenced data should be stationery.

According to these 2 tests we can conclude that 1st differenced log transformation series are given stationery observations. Thus, we were continued model selection procedure.

4.7.1.5 Model Identification

According to stationery results log transformed 1st difference observations of price of Kekulu Red are stationery. Since we can find a suitable time series model after estimating significant lags.

4.7.1.5.1 ACF and PACF of transformed & difference data

The ACF and PACF graph of difference transformed data are computed as in Figure 4.29 and Figure 4.30 Then, these two graphs were used to determine the parameter values of p and q.

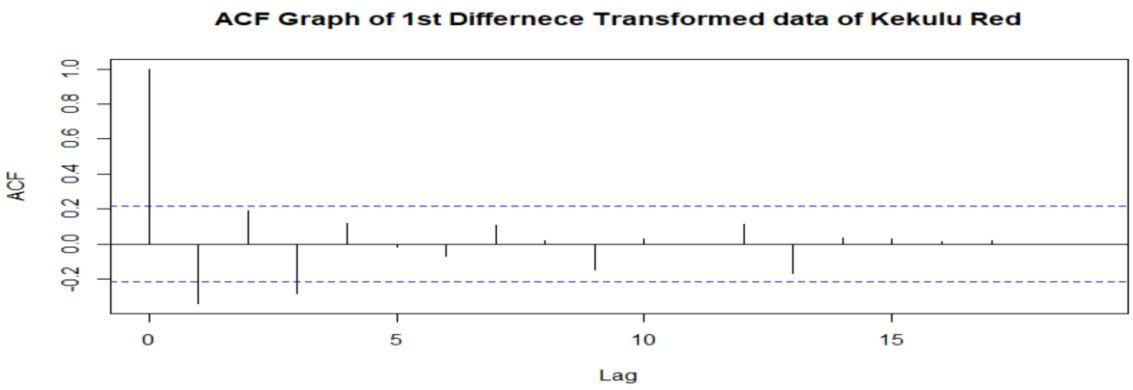

Figure 4.29: ACF plot of transformed 1st differenced observations of Kekulu Red

According to the Figure 4.29 lags are dies after 1st lags. We cannot see there are significant spikes after the lag 1, thus MA(1) and MA(0) are the suitable parameters and q must be 1 or 0.

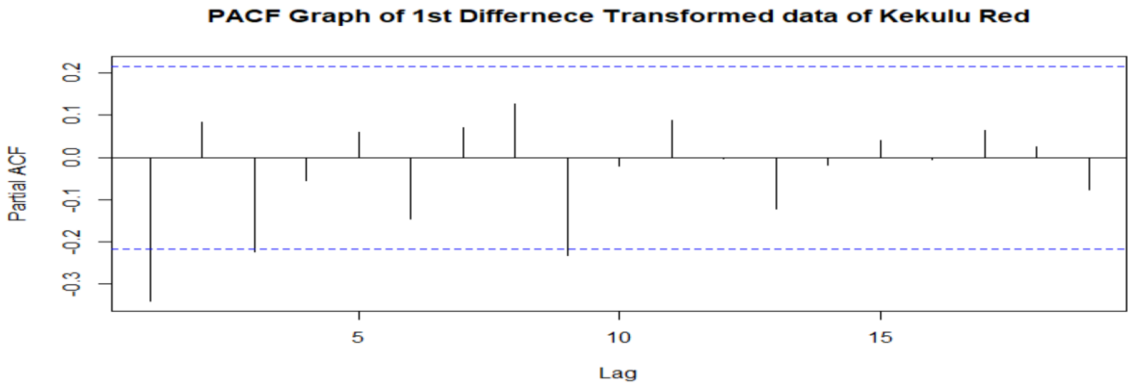

Figure 4.30: PACF plot of transformed 1st differenced observations of Kekulu Red

According to the Figure 4.30 lags are dies after 1st lags. We cannot see any significant spikes after lag 1, thus AR(1) or AR(0) is the suitable parameters and p must be 1 or 0.

Finally, we have selected different type of models and we were checked the most appropriate model for the continue forecasting procedure.

4.7.1.6 Model selection

After identifying the feasible values of p and q, the next step is an estimate the parameters under the most suitable model. The parameter estimation of model is conducted by R software. ARIMA model have been selected using minimum AIC value. Table 4.57 gives the ARIMA models with AIC values.

Model	**AIC**
ARIMA(0,1,0)	-470.93
ARIMA(0,1,1)	-479.15
ARIMA(1,1,0)	**-480.94**
ARIMA(1,1,1)	-480.48

Table 4.58: Results of model selection of ARIMA (1,1,0) for Kekulu Red

ARIMA (1,1,0) model has the minimum AIC value. Thus ARIMA (1,1,0) is selected as suitable model for forecasting the weekly average retail prices of Kekulu Red.

4.7.1.7 Parameter Estimation

ARIMA (Kekulu Red Log Transformed, order = c (1,1,0))		
Coefficients		
	ar1	
	-0.3998	
s.e.	0.1109	
t ratio	-3.606536	
p value	0.0003103124	
σ^2 estimated as 0.0001579	Log likelihood = 242.47	AIC = -480.94

Table 4.59: Parameter estimation of ARIMA (1,1,0) for Kekulu Red

According to the Table 4.58, the p-values of terms are less than 0.05. Thus, we can conclude that the coefficient is statistically significant. So, we can fit the model with the terms.

The model equation can be formulated as

$$y_t = y_{t-1} - 0.3998(y_{t-1} - y_{t-2}) + \varepsilon_t$$

Where y_t is the weekly average retail prices of Kekulu White in i[th] and ε_t is the error at i[th] time period.

4.7.1.8 Diagnostic checking

After estimating the parameter for ARIMA (1,1,0) model, next step will be the diagnostic checking.

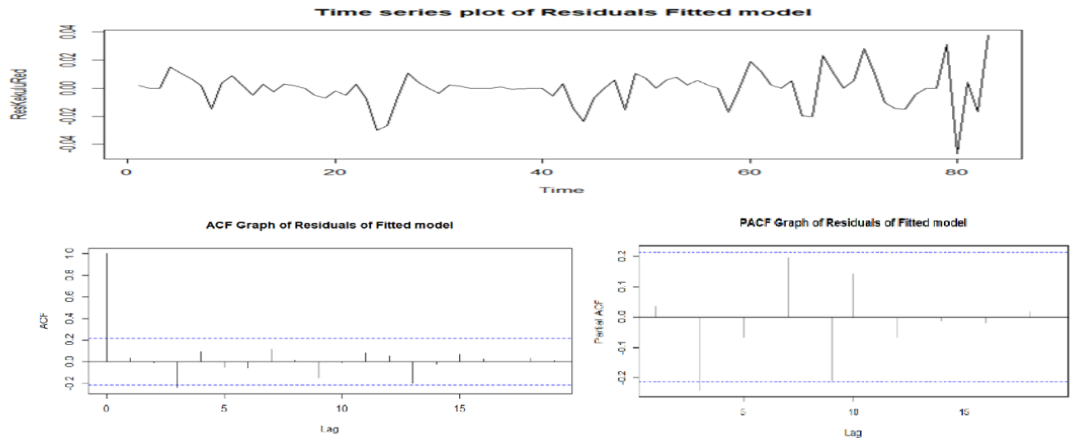

Figure 4.31: ACF and PACF plot of residuals and residual plot of ARIMA (1,1,0)

According to figure 4.31. Top of the box contain the time plot of standardized residuals. It shows that no pattern and looks like an independent identical distributed. And residual ACF and PACF shows all of lags are not significant. Since the residuals haven't significance autocorrelation.

To detect the correlation, we can study residual ACF and PACF plots. As well as Ljung-Box test used to detect correlation.

For testing correlation of residuals for ARIMA (1,1,0) model

Type of Test	χ^2 Value	df	P Value
Ljung-Box test	10.102	9	0.3423
Ljung-box squared test	15.394	9	0.08067

Table 4.60: Box-test results of Residuals of ARIMA (1,1,0) for Kekulu Red

According to above R output, p-value of Ljung-Box test is 0.3423. Thus, the null hypothesis of Ljung-Box test isn't rejected at 5% level of significance.so there is no correlation of residuals of the ARIMA (1,1,0) model.

Then we were used Ljung-box squared test for residual to check whether homoscedasticity of residuals. It is more helpful determine residuals have constant variance.

According to above R output, p-value of Ljung-Box squared test is 0.08067. Thus, the null hypothesis of Ljung-Box test is not rejected at 5% level of significance. So, there is no heteroscedasticity of residuals of the ARIMA (1,1,0) model. Thus, error term variance of ARIMA (1,1,0) model is constant.

Thus, the next thing is normality test of residuals it can be shows in following Histogram of residuals plots.

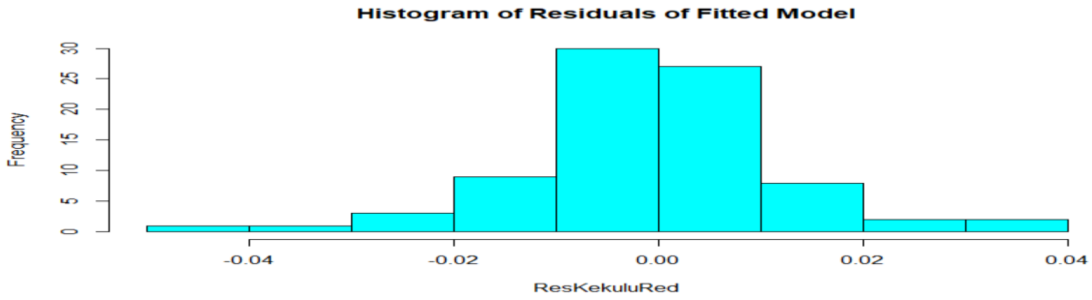

Figure 4.32: Histogram of residual of ARIMA (1,1,0)

Figure 4.32 shows the Histogram of residuals of the selected model. Histogram plot shows bell shape graph.

Using Jarque-Bera test we were checked the normality condition of residuals.

Jarque.Bera.Test(Residuals of ARIMA (1,1,0) for Kekulu Red)		
χ^2 = 26.324	df = 2	p-value = 1.922e-06=$1.922 * 10^{-6}$

Table 4.61: Jarque-Bera test results of ARIMA (1,1,0)

According to Table 4.60 shows the Jarque-Bera normality test results. Which p-value $1.922 * 10^{-6}$ is less than 0.05. Thus, the null hypothesis of normality of residuals being rejected. Forecasting does not need the premise of normality. Rather, it refers to data stationarity. Forecasting can be done even if residuals are determined to be non-normally distributed (through, for example, the Jarque Bera Test). Forecasting should be avoided if the series is determined to be non-stationary (through, for example, the Augmented Dickey Fuller Test), because such forecasts would be inaccurate.

4.7.1.9 Forecasting using ARIMA (1,1,0)

The next step is forecasting using ARIMA (1,1,0).

Week	Forecast	Actual
2019-13	80.2786388	75
2019-14	81.74646851	75
2019-15	81.15644223	75
2019-16	81.3916776	75
2019-17	81.2976515	73
2019-18	81.33528635	75
2019-19	81.32011797	75
2019-20	81.32611008	75
2019-21	81.32386299	75
2019-22	81.32479927	75

Table 4.62: Forecast values of ARIMA (1,1,0) f or Kekulu Red

According to the Table 4.61, According to this forecasting values, it gives difference value in entire space. Therefore, this implies a considerable forecasting value for the weekly average retail price of Kekulu Red.

Actual Vs forecast graph shows in bellow

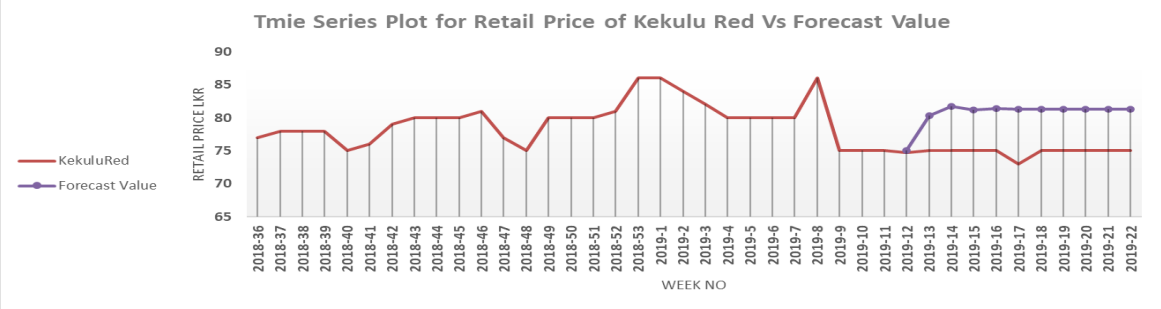

Figure 4.33: Actual vs Forecast graph of ARIMA (1,1,0) for Kekulu Red

4.7.1.10 Accuracy of ARIMA (1,1,0)

After forecasting, we want to check validation of forecasting model. For the validation process we used to compute MAPE, RMSE and MAE of the selected model. These values called as accuracy measurements. Lowest accuracy measurements imply most appropriate forecasting model.

Measurement	**Value**
MAPE	8.7%
RMSE	6.510465022
MAE	6.466706835

Table 4.63: Accuracy measurements of ARIMA (1,1,0) Kekulu Red

MAPE value is 8.7% which is less than 10%. And also, we can see that the RMSE and MAE values are small considerably. According to these results, we can clearly say that the model ARIMA (1,1,0) gives highly accurate forecast.

4.7.2 Forecasting Retail Price of Kekulu Red using Double Exponential Smoothing

According to the figure 4.27 of original time series plot, there is no seasonal pattern. But there is downward trend. And also, we can clearly see that the time series data are changing slowly over time. Therefore, double exponential smoothing forecast, also known as Holt's linear method, can be used for the series with trend.

For the optimum model parameters, the R output as follows

Holt-Winters exponential smoothing with trend and without seasonal component		
Call: HoltWinters(x = KekuluRedData, gamma = FALSE, b.start=18)		
Smoothing parameters:		
alpha: 0.8006609	beta: 0.7117236	gamma: FALSE
Coefficients: [,1]	a = 81.879069	b = 3.764678

Table 4.64: Parameter estimation of Double Exponential smoothing Kekulu Red

Where a is the intercept and b is the trend component. Alpha is the smoothing constant for the level of the series. Beta is the smoothing constant for the trend. We can see that alpha value is close to 1 and beta value is 0.7117236.

4.7.2.1 Diagnostic checking of double exponential smoothing

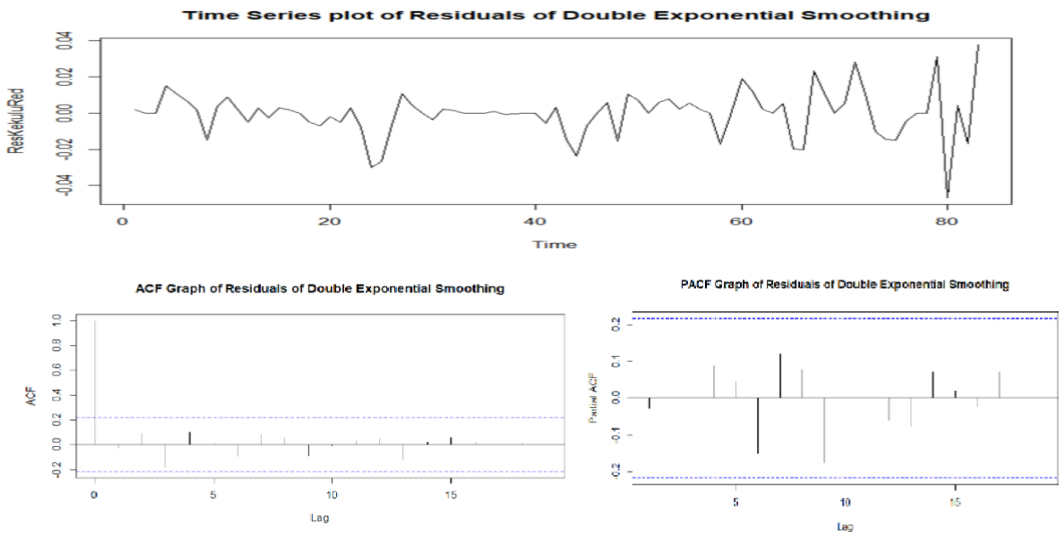

Figure 4.34: ACF and PACF of residuals and Residuals plot of double exponential smoothing for Kekulu Red

According to this figure, residual ACF and PACF shows lags are not significant. Since the residuals haven't significance autocorrelation. Then residuals are uncorrelated. This implies the model is suitable for forecasting process.

Also, we were preformed Box test for the checking the residual autocorrelation and residual heteroscedasticity.

Type of Test	χ^2 Value	df	P Value
Ljung-Box test	6.8177	10	0.7425
Ljung-box squared test	4.7175	10	0.9092

Table 4.65: Box test of Residuals of double exponential smoothing for Kekulu Red

According to above R output, p-value of Ljung-Box test is 0.7425 which is greater than 0.05. Thus, the null hypothesis of Ljung-Box test isn't rejected at 5% level of significance.so there is no correlation of residuals of the double exponential smoothing model.

Now we are preformed Box test for the checking the residual heteroscedasticity.

According to above R output, p-value of Ljung-Box squared test is 0.9092 which is greater than 0.05. Thus, the null hypothesis of Ljung-Box squared test isn't rejected at 5% level of significance. So, there is no heteroscedasticity of residuals of the Double Exponential smoothing model.

Now we are preformed normality test for the residuals of Double Exponential smoothing model.

Figure 4.35 shows the Histogram of residuals of the selected model. Histogram plot shows bell shape graph and, but it skewed to right side.

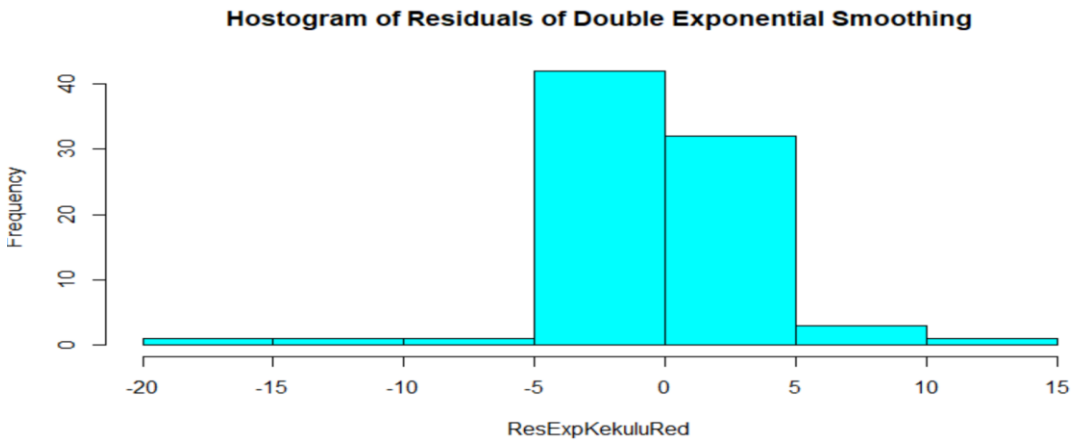

Figure 4.35: Histogram of residuals of double exponential smoothing

Using Jarque-Bera test we were checked the normality condition of residuals.

Jarque.Bera.Test(Residuals of Double Exponential Smoothing for Kekulu Red)		
χ^2 = 220.05	df = 2	p-value < 2.2e-16=$2.2* 10^{-16}$

Table 4.66: Jarque-Bera test results of double exponential smoothing for Kekulu White

According to figure 4.65 shows the Jarque-Bera normality test results which p-value is less than 0.05. Thus, null hypothesis of normality of residuals being rejected. As mentioned above 4.4.1.7, still we can do the forecasting.

4.7.2.2 Forecasting using exponential smoothing

Week	Forecast	Actual
2019-13	85.64375	75
2019-14	89.40843	75
2019-15	93.1731	75
2019-16	96.93778	75
2019-17	100.70246	73
2019-18	104.46714	75
2019-19	108.23182	75
2019-20	111.9965	75
2019-21	115.76117	75
2019-22	119.52585	75

Table 4.67: Forecast using double exponential smoothing for weekly average retail price of Kekulu Red

4.7.2.3 Accuracy of double exponential smoothing

Measurement	Value
MAPE	37.2%
RMSE	29.82365391
MAE	27.80140131

Table 4.68: Accuracy measurements of double exponential smoothing

MAPE value is 37.2% which is greater than 20% and less than 50%. And also, we can see that the RMSE and MAE values are not small considerably. Then this double exponential smoothing model gives Reasonable forecast.

4.7.3 Forecasting evaluation of Kekulu Red

Finally, we were found a model for forecasting process by using minimum MAPE, RMSE, MAE of foregoing performed models.

Model	MAPE	RMSE	MAE
ARIMA(1,1,0)	8.7%	6.510465022	6.466706835
Double Exponential Smoothing	37.2%	29.82365391	27.80140131

Table 4.69: Forecasting evaluation models of Kekulu Red Price

According to this figure we were found the both models double exponential smoothing method and ARIMA(1,1,0) are the best method for the estimating forecast values. But MAPE value of ARIMA(1,1,0) is less than MAPE value of Double Exponential smoothing. It implies ARIMA(1,1,0) model is better than double exponential smoothing model. Thus, we can recommend the best model is ARIMA(1,1,0).

We were found the ARIMA(1,1,0) is the best method for the forecasting process. Since ARIMA(1,1,0) can perform a time series approach for the determine future values of Kekulu Red Retail price.

4.8 Vector Autoregressive model

First of all, we want to find there exist strong correlation of these 4 variables

4.8.1 Correlation matrix of Samba, Nadu, Kekulu White and Kekulu Red prices

Variables	Samba	Nadu	Kekulu White	Kekulu Red
Samba	1	0.2115613	0.09283347	0.2253343
Nadu	0.2115613	1	0.58176079	0.3534846
Kekulu White	0.09283347	0.5817608	1	0.5548067
Kekulu Red	0.22533429	0.3534846	0.55480673	1

Table 4.70: Correlation matrix

According to Table 4.69 clearly see that there some strong correlation between some variable. Since we can perform VAR model for the check independency of variables and improve the forecasting process.

4.8.2 Selection of VAR

According to foregoing analysis retail prices of Nadu, Kekulu White and retail prices of Kekulu Red are stationery under 1st differences and retail prices of Samba is stationery under original data. Since all 4 variables are not non-stationery under original data with same number of observations.

Therefore, we can move to the VAR process analysis will only estimate short run relationships.

4.8.3 Lag selection of VAR

Lag	AIC	HQ	SC	FPE
1	6.563949	6.866065	7.322838	710.17
2	**6.465517**	**6.969043**	**7.730331**	**647.26**
3	6.587419	7.292356	8.358160	740.62
4	6.512291	7.418638	8.788957	703.17

Table 4.71: Lag selection criteria

According to Table 4.70 we can Cleary see that lag 2 provide the minimum values of AIC values. Since we can take the VAR (2) process for the continue the time series procedure.

4.8.4 Parameter estimation of VAR (2)

	Estimate	Std. Error	P value
Samba Lag1	-0.29289	0.10380	0.006075
Nadu Lag 2	0.35496	0.09744	0.000487
Kekulu Red Lag 2	-0.23894	0.10714	0.028647

Table 4.72: Parameter estimation of Samba as a dependent variable

Table 4.71 shows the estimates and their P value of estimates. All P values are less than 0.05. which means all are significance at 5% level of significance. Multiple R squared 20.35%

According to Multiple R squared value, selected model is explaining low variations.

H_0: Model isn't adequate

H_{01}: Model is adequate

P value = 0.0005237 which is less than 0.05. Then we can conclude that selected VAR (2) model is significant at 5% level of significance.

	Estimate	Std. Error	P value
Nadu Lag1	0.2883	0.1077	0.00906

Table 4.73: Parameter estimation of Nadu as a dependent variable

Table 4.72 shows the estimates and their P value of estimates. P values is less than 0.05. which means all are significance at 5% level of significance. Multiple R squared 8.31%

According to Multiple R squared value, selected model is explaining low variations.

H_0: Model isn't adequate

H_{01}: Model is adequate

P value = 0.009061 which is less than 0.05. Then we can conclude that selected VAR (2) model is significant at 5% level of significance.

	Estimate	Std. Error	P value
Kekulu Red Lag1	-0.5068	0.1485	0.00102
Kekulu Red Lag 2	0.3177	0.1530	0.04113

Table 4.74: Parameter estimation of Kekulu White as a dependent variable

Table 4.73 shows the estimates and their P value of estimates. All P values are less than 0.05. which means all are significance at 5% level of significance. Multiple R squared 22.05%

According to Multiple R squared value, selected model is explaining low variations.

H_0: Model isn't adequate

H_{01}: Model is adequate

P value = 6.046×10^{-5} which is less than 0.05. Then we can conclude that selected VAR (2) model is significant at 5% level of significance.

	Estimate	Std. Error	P value
Nadu Lag1	0.3663	0.1077	0.00107
Kekulu Red Lag 1	-0.5432	0.1165	1.27×10^{-5}

Table 4.75: Parameter estimation of Kekulu Red as a dependent variable

Table 4.74 shows the estimates and their P value of estimates. All P values are less than 0.05. which means all are significance at 5% level of significance. Multiple R squared 24.09%

According to Multiple R squared value, selected model is explaining low variations.

H_0: Model isn't adequate

H_{01}: Model is adequate

P value = 2.143×10^{-5} which is less than 0.05. Then we can conclude that selected VAR (2) model is significant at 5% level of significance.

4.8.5 Diagnostic checking of VAR (2)

After estimating parameters, we want to check the diagnostics of selected VAR (2) model. Thus, we test the VAR (2) residual has constant variance or residual has heteroscedasticity. Using multivariate ARCH test, we were found a conclusion

The R output as follows

H_0: There is no ARCH effect

H_{01}: There is significant ARCH effect

ARCH (multivariate)		
data: Residuals of VAR object varmodel		
Chi-squared = 700	df = 1000	p-value = 1

Table 4.76: Results of multivariate ARCH test VAR (2)

According to Table 4.75 P value is 1, which is greater than 0.05. Thus, we cannot reject the null hypothesis, there is no arch effect at 5 % level of significance.

Using portmanteau serial test, we can check the VAR (2) residuals has correlation or not.

The R output as follows

H_0: No autocorrelation of residuals

H_{01}: significant autocorrelation of residuals

Portmanteau Test (asymptotic)		
data: Residuals of VAR object varmodel		
Chi-squared = 137.01	df = 128	p-value = 0.277

Table 4.77: Results of Portmanteau serial test of VAR (2)

According to Table 4.76, the P value of portmanteau test is 0. 277.which is greater than 0.05. Thus, we cannot reject the null hypothesis at 5 % level of significance. Then we can conclude that VAR (2) residuals hasn't significance correlation.

Jarque Bera-Test (multivariate)		
data: Residuals of VAR object varmodel		
Chi-squared = 236.75	df = 8	p-value < 2.2×10^{-16}

Table 4.78: Jarque-Bera test(multivariate)

According to Table 4.77 shows the Jarque-Bera normality test results. Which p value < 2.2×10^{-16} is less than 0.01. Thus, null hypothesis of normality of residuals being rejected.

4.8.6 Causality Analysis

Granger causality

H_0: Nadu, Kekulu Red does not Granger-cause Samba

H_1: Nadu, Kekulu Red granger cause Samba

The R output as follows

causality (varmodel, cause = c("Nadu","Kekulu White","Kekulu Red"))			
data: VAR object varmodel			
F-test = 3.0624	df1 = 6	df2=280	P-value= 0.006428

Table 4.79: Granger causality test for Samba

Instantaneous causality

H_0: Nadu, Kekulu Red does not instantaneous-cause Samba

H_1: Nadu, Kekulu Red instantaneous-cause Samba

causality (varmodel, cause = c("Nadu","Kekulu White","Kekulu Red"))		
data: VAR object varmodel		
Chi-squared = 7.8485	df=3	P- value = 0.04925

Table 4.80: Instantaneous causality test for Samba

According to these 2 figures, P value of Table 4.78 is 0.006428 which is less than 0.05. Table 4.79 shows P value is 0. 04925 which is less than 0.05 thus the null hypothesis of no Granger-causality from Sambo to Nadu, Kekulu White and Kekulu Red must be rejected; whereas the null hypothesis of non-instantaneous causality can be rejected. This test outcome is economically plausible.

Granger causality

H_0: Samba, Kekulu White, Kekulu Red do not Granger-cause Nadu

H_1: Samba, Kekulu White, Kekulu Red granger cause Nadu

The R output as follows

causality (varmodel, cause = c("Samba","Kekulu White","Kekulu Red"))			
data: VAR object varmodel			
F-test = 1.2228	df1 = 6	df2=280	P-value= 0.2945

Table 4.81: Granger causality test for Nadu

Instantaneous causality

H_0: Samba, Kekulu White, Kekulu Red do not instantaneous-cause Nadu

H_1: Samba, Kekulu White, Kekulu Red instantaneous-cause Nadu

causality (varmodel, cause = c("Samba","Kekulu White","Kekulu Red"))		
data: VAR object varmodel		
Chi-squared = 22.391	df=3	P- value = 0.00005408

Table 4.82: Instantaneous causality test for Nadu

According to these 2 figures, P value of Table 4.80 is 0.2945 which is greater than 0.05. Table 4.81 shows P value is 0. 00005408 which is less than 0.05 thus the null hypothesis of no Granger-causality from Nadu to Samba, Kekulu White and Kekulu Red must be not rejected; whereas the null hypothesis of non-instantaneous causality can be rejected. Then we can say that there is causality between Nadu to Samba, Kekulu White and Kekulu Red and no significant instantaneous causality.

Granger causality and the R output as follows

H_0: Kekulu Red does not Granger-cause Kekulu White

H_1: Kekulu Red granger cause Kekulu White

causality (varmodel, cause = c("Samba","Nadu","Kekulu Red"))			
data: VAR object varmodel			
F-test = 3.2388	df1 = 6	df2=280	P-value= 0.004299

Table 4.83: Granger causality test for Kekulu White

Instantaneous causality

H_0: Kekulu Red does not instantaneous-cause Kekulu White

H_1: Kekulu Red instantaneous-cause Kekulu White

causality (varmodel, cause = c("Samba","Nadu","Kekulu Red"))		
data: VAR object varmodel		
Chi-squared = 24.41	df=3	P- value = 0.00002051

Table 4.84: Instantaneous causality test for Kekulu White

According to these 2 figures, P value of Table 4.82 is 0.004299 which is less than 0.05. Table 4.83 shows P value is 0. 00002051 which is less than 0.05 thus the null hypothesis of no Granger-causality from Kekulu White to Samba , Nadu and Kekulu Red must be rejected; whereas the null hypothesis of non-instantaneous causality can be rejected. This test outcome is economically plausible.

Granger causality

H_0: Nadu does not Granger-cause Kekulu Red

H_1: Nadu granger cause Kekulu Red

The R output as follows

causality (varmodel, cause = c("Samba","Nadu","Kekulu White"))			
data: VAR object varmodel			
F-test = 2.2796	df1 = 6	df2=280	P-value= 0.03644

Table 4.85: Granger causality test for Kekulu Red

Instantaneous causality

H_0: Nadu, does not instantaneous-cause Kekulu Red

H_1: Nadu instantaneous-cause Kekulu Red

causality (varmodel, cause = c("Samba","Nadu","Kekulu White"))		
data: VAR object varmodel		
Chi-squared = 14.472	df=3	P- value = 0.002328

Table 4.86: Instantaneous causality test for Kekulu Red

According to these 2 figures, P value of Table 4.84 is 0.03644 which is less than 0.05. Table 4.85 shows P value is 0. 002328 which is less than 0.05 thus the null hypothesis of no Granger-causality from Kekulu Red to Samba, Nadu and Kekulu White must be rejected, whereas the null hypothesis of non-instantaneous causality can be rejected. This test outcome is economically plausible.

4.8.7 Forecasting Using VAR (2)

Week No	Samba	Nadu	Kekulu White	Kekulu Red
2019-13	104.6964	79.86632	73.79121	76.79562
2019-14	103.1405	83.07755	81.49338	78.4913
2019-15	104.421	86.46736	81.8562	77.38867
2019-16	104.9221	88.12077	83.97614	79.61613
2019-17	105.4147	88.47269	83.62453	78.98337
2019-18	105.2743	88.58473	84.96787	79.43621
2019-19	105.434	88.64531	85.08815	79.12999
2019-20	105.4757	88.52684	85.43986	79.49758
2019-21	105.5517	88.2101	85.15009	79.40778
2019-22	105.4823	87.85294	85.09819	79.4791

Table 4.87: Forecasting values of VAR (2)

4.8.8 Forecasting evaluation of VAR (2)

Finally, we were found a model for forecasting process by using minimum MAPE, RMSE, MAE of foregoing performed models.

Model	MAPE	RMSE	MAE
Samba	9.5%	10.19933616	8.853752437
Nadu	5.0%	5.144371757	4.081890533
Kekulu White	5.7%	5.843203899	4.268311795
Kekulu Red	5.4%	4.203553296	4.039176305

Table 4.88: Accuracy measurements of VAR (2)

Table 4.87 shows the accuracy measurements of VAR (2). MAPE values of time series less than 10%. And also, we can see that the RMSE and MAE values are small considerably. According to these results, we can clearly say that the models VAR (2) gives highly accurate forecast.

Chapter 5

Discussion

This study made use of a unique set of data collecting analysis tool at a very fine physical and time scale. The objective of this study was the find the suitable model for the forecasting process and find the interrelationship among selected 4 markets. Generally, ARIMA, Exponential smoothing, Double Exponential smoothing, and VAR model have been used. This chapter contain 4 types of models fitted for Rice prices and subsequent improvements of this study. The following discussion focus on the modeling results, followed by challenges and limitation and further improvement.

5.1 ARIMA model

All of variables first modeled using time series ARIMA model. According to the analysis in chapter 4 it was seen that Samba not provide suitable ARIMA models. Because it was a part of random walk model. The forecast for a random walk is its last observed value, regardless of the forecast horizon. But Samba provide suitable ARIMA model for the long-term forecasting process.

5.2 Exponential method

According to the chapter 4, time series observations are modeled by using exponential smoothing. Basically, we were used trend corrective exponential smoothing except the time series of Samba. Reason for that all variables contained downward trend. Since we were ignored single exponential smoothing method and perform trend corrective exponential smoothing. Trend corrective exponential smoothing is extended version of simple exponential smoothing method.

5.3 VAR model

Finally, we were fitted vector autoregressive model for all 4 variables together. This study the cointegration test were not determine, there were one time series which was stationary at original stage. Since we move to the vector auto regression procedure. Vector auto regression is one of the multivariate time series models and it was used to determine interrelationship or independency among variables.

5.4 Challengers and Limitation

The accuracy of some of data input may have distorted results. Data were collected from daily price report of central bank and the data were not available on Saturday, Sunday as well as public holidays which lead for lot of missing values .The option was to go for the weekly average retail prices including less missing values in order to collect considerable data point around 93 to do the analysis for the rice retail prices.

It had been bit challenged to find the previous research papers because previous research papers help to determine the scope of thesis proper manner. And had to go through several research papers from different industries in order to achieve the set of specific objectives and the goal of this thesis.

Additionally, few R software related issues existed during the analyzing. Some packages were not available to get the required results of this study and had to write some codes in R language to get the results of analysis by self-studding the R language through internet.

5.5 Further Improvements

This study was used for the forecasting retail prices of rice and determine relationship among the variables. We were considered the weekly average retail prices of 4 categories of rice base products. The use of Box Jenkins approach to predict the maximum retail price of rice could be improved by incorporating covariates into models such as the introduction of product by outside parties, events that may be taking place in politics, natural disasters, pandemic situation like Covid-19 and speculations that are being introduce by companies in the market. A covariate is a secondary variable that can be affected the relationship between the dependent variable and other independent variable of key interest, and our main variable is time. By addressing the fact that there are outside components that influence the rice products, we can modify the model to try and anticipate the effect these events will have on the forecasts. Rice prices are affected for the Sri Lankan market. Since we can modify the model using effectiveness of these events will have on the forecasts.

Chapter 6

Conclusion

According to missing value estimation by using expectation maximization algorithm, exponential smoothing is not better than ARIMA model for the forecasting process except for time series of Samba. Thus, final model for the Samba retail price forecasting was exponential smoothing out of 2 models such as ARIMA and exponential smoothing. According to chapter 4 Nadu retail prices have highly accurate ARIMA model such as ARIMA (2,1,0). And Kekulu White retail prices and Kekulu Red retail prices have accurate ARIMA model such as ARIMA (1,1,0) and ARIMA (1,1,0) respectively. These models were determined by using minimum accuracy measurements.

As a multivariate time, series model, we were found VAR (2) model for the improve the forecast and check the independency of prices market. According to VAR results, we were determined Nadu retail prices independent from Samba, Kekulu White and Kekulu Red retail prices. And also, we were determined Kekulu White retail price was dependent only Kekulu Red retail price.

At all lags, the coefficient estimates between Samba, Nadu, and Kekulu Red are not the same as those between Samba, Nadu, and Kekulu Red. There is a feedback relationship between the four series, which can be seen in this. We may determine that the price movement of one market can readily and immediately spread to another market, and that the prices of the assets have a reasonably high correlation. As a result, essential markets are more or less interdependent on one another, implying that there is an interrelationship between variables.

The ARIMA and Exponential smoothing methodologies, as well as VAR as a multivariate model, were shown to be effective than other models in predicting prices of Samba, Nadu, Kekulu White, and Kekulu Red rice in Sri Lanka, with the lowest MAPE, MAE, and RMSE.

References

[1] Economic, price and financial system stability, outlook and policies report – CBSL (2019)

[2] Prices, wages, employment and productivity report – CBSL (2019)

[3] Global Agricultural Information Network report – USDA Foreign Agricultural Service (2019)

[4] D A M De Silva, M Yamao. Rice Pinch to war thrown nation: An overview of the rice supply chain of Sri Lanka and the consumer attitudes on government rice risk management (2009)

[5] USDA Nutrient database - USDA Agricultural Research Service (April 2020)

[6] Kwadwo Agyei Nyantakyi, B. L. (2015). Analysis of the Interrelationships between the Prices of Sri Lankan Rubber, Tea and Coconut Production Using Multivariate Time Series Horizon Research Publishing.

[7] G. Ramakrishna & R. Vijaya Kumari (2017). ARIMA model for forcasting of rice production in India by using SAS, Department of Statistics & Mathematics, College of Agriculture, Hyderabad, Telangana, India

[8] B D P Rangoda', L. M. (n.d.). An Analysis of different forecasting models for prices of coconut products in Sri Lanka. Proceedings of the Third Academic Sessions. Faculty of Agriculture, University of Ruhuna, Mapalana, Kamburupitiya 2 Coconut Research Institute, Lunuwila,

[9] Gurudeo Anand Tularam1, T. S. (2016,6). Oil-Price Forecasting Based on Various Univariate Time-Series Models. American Journal of Operations Research, 226-235.

[10] K. NAVEENA, S. R. (2014, April190-193). Forecasting of coconut production in India: A suitable time series model. International Journal of Agricultural Engineering.

[11] Pasapitch Chujai*, N. K. (2013, March 13 - 15). Time Series Analysis of Household Electric Consumption with ARIMA and ARMA Models. IMEC. Hong Kong.

[12] Daily Price Report of Central Bank of Sri Lanka (CBSL)

[13] Introduction to Time Series and Forecasting, Second Edition, Peter J. Brockwell Richard A. Davis, Department of Statistics, Colorado State University.

[14] Introduction to Time Series Analysis and Forecasting, DOUGLAS C. MONTGOMERY, Arizona State University, CHERYL L. JENNINGS, Bank of America, MURAT KULAHCI, Technical University of Denmark

[15] http://personal.cb.cityu.edu.hk/msawan/teaching/ms6215/Exponential%20Smoothing%20

[16] Methods.pdf

[17] http://a-little-book-of-r-for-time-series.readthedocs.io/en/latest/index.html

Study
Dealing with missing values and outliers
https://otexts.com/fpp2/missing-outliers.html

Appendix

Import data

```
library("forecast")

library('tseries')

Rice_Price_CompletedData_test<-read.csv("C:/Users/Ransika Peiris/OneDrive/Desktop/MSC/Rice_Price_CompletedData_test.csv")

Data_Set=Rice_Price_CompletedData_test

Samba=Data_Set[,3]

plot.ts(Samba , main= "Time Series plot for retail price of Samba")

SambaLog=Data_Set[,7]

plot.ts(SambaLog , main= "Time Series plot of Log Transformed for retail price of Samba")

adf.test(SambaLog)
```

```
kpss.test(SambaLog)
pp.test(SambaLog)
acf(SambaLog ,main="ACF plot for Original Samba Log transformed data")
pacf(SambaLog ,main="PACF plot for Original Samba Log transformed data")
arima(SambaLog, order = c(1,0,0))
arima0(SambaLog, order = c(1,0,0))
fitSamba=arima(SambaLog, order = c(1,0,0))
ResSamba=residuals(fitSamba)
plot.ts(ResSamba,main="Residuals plot of Price Samba")
acf(ResSamba,main="ACF plot of Residuals ")
pacf(ResSamba,main="PACF plot of Residuals ")
Box.test(ResSamba^2,lag= 10,type = c("Ljung-Box"),fitdf = 1)
hist(ResSamba,col = 5,main = "Histogram of Residuals")
jarque.bera.test(ResSamba)
plot.ts(SambaLog-ResSamba,col=2)
forecastSamba=predict(fitSamba,n.ahead = 10)
HoltWinters(X=Samba, bete=FALSE, gamma = FALSE)
fitExpSamba=HoltWinters(Samba, beta = FALSE, gamma = FALSE)
arima(SambaLog,order = c(1,0,0), include.driff=True)
ResExpSamba=residuals(fitExpSamba)
plot.ts(ResExpSamba,main="Residuals of Exponential smoothing")
pacf(ResExpSamba, main="PACF plot of Exponential Smoothing")
Box.test(ResExpSamba^2,lag=10,type = c("Ljung-Box"))
hist(ResExpSamba,col = 5, main = "Histogram of Residuals of Exponential Smoothing")
```

```
jarque.bera.test(ResExpSamba)

forecstExpSamba=predict(fitExpSamba,n.ahead = 10)

fitExpSamba


Nadu=Data_Set[,4]

NaduLog=Data_Set[,8]

ts.plot(Nadu)

plot.ts(NaduLog,ylab="Price",main="Time Series plot of Log Transformed Price of Nadu")

adf.test(NaduLog)

kpss.test(NaduLog)

NaduLogDiff=diff(NaduLog)

adf.test(NaduLogDiff)

kpss.test(NaduLogDiff)

pacf(NaduLogDiff,main="PACF Graph of 1st difference transformed data of Nadu")

fitNadu=arima(NaduLog, order = c(2,1,0))

ResNadu=residuals(fitNadu)

plot.ts(ResNadu,main="Residuals plot of Nadu Price")

pacf(ResNadu,main="PACF graph of Residuals of Nadu Price")

Box.test(ResNadu^2,lag = 10,type = "Ljung-Box",fitdf = 2)

hist(ResNadu,col = 5,main = "Histogram of residuals of Nadu Price")

jarque.bera.test(ResNadu)

ForecastNadu=predict(fitNadu,n.ahead = 10)

plot.ts(Nadu)
```

```
HoltWinters(Nadu,gamma = FALSE)

predict(HoltWinters(Nadu,gamma = FALSE, b.start = 23),n.ahead = 10)

fitExpNadu=HoltWinters(Nadu,gamma = FALSE, b.start = 23)

ResExpNadu=residuals(fitExpNadu)

plot.ts(ResExpNadu,main="Residual Graph of Double Exponential Smoothing of Nadu price")

pacf(ResExpNadu,main="PACF graph of Residuals of double exponential smoothing")

Box.test(ResExpNadu^2,lag = 10, type = c("Ljung-Box"))

hist(ResExpNadu,col = 5, main = "Histogram of Residuals of Double Exponential Smoothing")

jarque.bera.test(ResExpNadu)

forecastExpNadu=predict(fitExpNadu,n.ahead = 10)


Rice_Price_CompletedData_test<-read.csv("C:/Users/Ransika Peiris/OneDrive/Desktop/MSC/Rice_Price_CompletedData_test.csv")

KekuluWhiteData<- Rice_Price_CompletedData_test[,5]

LogKekuluWhiteData<- Rice_Price_CompletedData_test[,9]

ts.plot(KekuluWhiteData,main="Time Series plot of Kekulu White Price")

1.] ts.plot(LogKekuluWhiteData,main="Time Series plot of log transformation of Kekulu White")

adf.test(LogKekuluWhiteData)

kpss.test(LogKekuluWhiteData)

LogKekuluWhiteDiff <- diff(LogKekuluWhiteData)

adf.test(LogKekuluWhiteDiff)

kpss.test(LogKekuluWhiteDiff)
```

```
pp.test(LogKekuluWhiteDiff)

acf(LogKekuluWhiteDiff, main ="ACF graph of 1st Difference transformed data of Kekulu White ")

pacf(LogKekuluWhiteDiff,main ="PACF graph of 1st Difference transformed data of Kekulu White ")

arima(LogKekuluWhiteData, order = c(1,1,0))

fitKekuluWhite=arima(LogKekuluWhiteData, order = c(1,1,0))

ts.plot(ResKekuluWhite,main="Time Series plot of Residuals of Kekulu White")

pacf(ResKekuluWhite, main="PACF Graph of Residuals of Kekulu White")

Box.test(ResKekuluWhite,lag = 10,type = c("Ljung-Box"), fitdf = 2)

Box.test(ResKekuluWhite^2,lag = 10,type = c("Ljung-Box"), fitdf = 2)

hist(ResKekuluWhite, col=5 , main = "Histogram of Residuals of Kekulu White")

jarque.bera.test(ResKekuluWhite)

forecastKekuluWhite = predict(fitKekuluWhite,n.ahead = 10)

HoltWinters(KekuluWhiteData,gamma = FALSE, b.start = 23)

ExpFitKekuluWhite=HoltWinters(KekuluWhiteData,gamma = FALSE, b.start = 23)

ResExpKekuluWhite= residuals(ExpFitKekuluWhite)

ts.plot(ResExpKekuluWhite,main="Time Series plot of residuals of double exponential")

pacf(ResExpKekuluWhite,Main="PACF graph of residuals of double exponential ")

Box.test(ResExpKekuluWhite,lag = 10,type = c("Ljung-Box"))

Box.test(ResExpKekuluWhite^2,lag = 10,type = c("Ljung-Box"))

hist(ResExpKekuluWhite,col=5,main="Histogram of Double Exponential Method")

jarque.bera.test(ResExpKekuluWhite)

ForcastExpKekuluWhite= predict(ExpFitKekuluWhite,n.ahead = 10)
```

ForcastExpKekuluWhite

```
Rice_Price_CompletedData_test<-read.csv("C:/Users/Ransika Peiris/OneDrive/Desktop/MSC/Rice_Price_CompletedData_test.csv")

KekuluRedData<- Rice_Price_CompletedData_test[,6]

LogKekuluRedData<- Rice_Price_CompletedData_test[,10]

ts.plot(KekuluRedData,main ="Time Series plot of Kekulu Red Rice Price")

ts.plot(LogKekuluRedData,main ="Time Series plot of Log Transformation of Kekulu Red Rice Price")

adf.test(LogKekuluRedData)

kpss.test(LogKekuluRedData)

LogKekuluRedDiff= diff(LogKekuluRedData)

adf.test(LogKekuluRedDiff)

kpss.test(LogKekuluRedDiff)

acf(LogKekuluRedDiff,main="ACF Graph of 1st Differnece Transformed data of Kekulu Red")

pacf(LogKekuluRedDiff,main="PACF Graph of 1st Differnece Transformed data of Kekulu Red")

fitKekuluRed=arima(LogKekuluRedData, order = c(1,1,0))

ResKekuluRed=residuals(fitKekuluRed)

ts.plot(ResKekuluRed,main="Time series plot of Residuals Fitted model")

pacf(ResKekuluRed,main="PACF Graph of Residuals of Fitted model")

Box.test(ResKekuluRed, lag = 10, type = c("Ljung-Box"),fitdf = 1)

Box.test(ResKekuluRed^2, lag = 10, type = c("Ljung-Box"),fitdf = 1)

hist(ResKekuluRed,col = 5, main = "Histogram of Residuals of Fitted Model")
```

```
jarque.bera.test(ResKekuluRed)

forecastKekuluRed=predict(fitKekuluRed,n.ahead = 10)

forecastKekuluRed

ExpFitKekuluRed= HoltWinters(KekuluRedData, gamma = FALSE,b.start = 18)

ResExpKekuluRed=residuals(ExpFitKekuluRed)

ts.plot(ResKekuluRed,main="Time Series plot of Residuals of Double Exponential Smoothing")

pacf(ResExpKekuluRed, main="PACF Graph of Residuals of Double Exponential Smoothing")

Box.test(ResExpKekuluRed,lag = 10, type = c("Ljung-Box"))

Box.test(ResExpKekuluRed^2,lag = 10, type = c("Ljung-Box"))

hist(ResExpKekuluRed,col=5,main="Hostogram of Residuals of Double Exponential Smoothing")

jarque.bera.test(ResExpKekuluRed)

forecastExpKekuluRed= predict(ExpFitKekuluRed,n.ahead = 10)


dx1=diff(Rice_Price_CompletedData_test[,3],differences = 1)

dx2=diff(Rice_Price_CompletedData_test[,4],differences = 1)

dx3=diff(Rice_Price_CompletedData_test[,5],differences = 1)

dx4=diff(Rice_Price_CompletedData_test[,6],differences = 1)

dx=cbind(dx1,dx2,dx3,dx4)

cor(dx)

library("urca")

ca.jo(dx)

VARselect(dx, lag.max = 10, type = c("both"))
```

```
v=VAR(dx, p=2,type = c("both"))
varmodel=restrict(v,method =c("ser"),thresh=2)
summary(varmodel)
View(summary(test))
VARselect(dx,type = c("both"))
arch.test(varmodel, lags.multi = 10)
serial.test(varmodel,lags.pt = 10, type = c("PT.asymptotic"))
acf(residuals(varmodel))
causality(varmodel,cause = c("dx2","dx3","dx4"))
causality(varmodel,cause = c("dx1","dx3","dx4"))
causality(varmodel,cause = c("dx1","dx2","dx4"))
causality(varmodel,cause = c("dx1","dx2","dx3"))
residualVAR = residuals(varmodel)
residual
library("tseries")
jarque.bera.test(residualVAR)
normality.test(varmodel,multivariate.only = TRUE)
forcastVAR= predict(v, n.ahead = 10 , ci=0.95, dumvar = NULL)
forcastVAR
dx11=Rice_Price_CompletedData_test[,3]
dx22=Rice_Price_CompletedData_test[,4]
dx33=Rice_Price_CompletedData_test[,5]
dx44=Rice_Price_CompletedData_test[,6]
dxx=cbind(dx11,dx22,dx33,dx44)
```

```
v1=VAR(dxx, p=2,type = c("both"))

forcastVAR1= predict(v1, n.ahead = 10 , ci=0.95, dumvar = NULL)

forcastVAR1

View(forcastVAR1)

forcastVAR1[["fcst"]][["dx11"]]
```